DISSERTATION

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of
Heidelberg, Germany

for the degree of
Doctor of Natural Sciences

presented by

Diplom-Physicist: Andreas Breidenassel
born in: Bergisch Gladbach

Oral examination: May 18, 2005

A High Dynamic Range CMOS Image Sensor with Adaptive Integration Time Control

Referees: Prof. Dr. Karlheinz Meier
Prof. Dr. Volker Lindenstruth

Zusammenfassung

Hochdynamischer CMOS Bildsensor mit adaptiver Integrationszeitregelung - Im Rahmen der vorliegenden Arbeit wurde ein auf CMOS-Technologie basierender Bildsensor entwickelt und getestet. Die Bildaufnahme vollzieht sich durch Integration des durch einfallendes Licht generierten Photostromes im jeweiligen Pixel. Die Implementierung eines Konzeptes zur lokal-adaptiven Regelung der Integrationszeit ermöglicht die Aufnahme optisch hoch dynamischer Szenerien ohne Informationsverlust aufgrund von Über- oder Unterbelichtung. Das Konzept gestattet es, abhängig von der Größe des implementierten Speichers, eine frei verschiebbare Bildregion zu bestimmen, innerhalb der die Integrationszeitregelung stattfinden soll. Bei einer gewählten maximalen Integrationszeit von 33 ms beträgt der dynamische Bereich des Sensors 134 dB und umfasst einen Intensitätsbereich von $1\,\mathrm{mW/m^2}$ bis $5\,\mathrm{kW/m^2}$. Der realisierte Prototyp besitzt eine Auflösung von 170×170 Pixeln mit einer hochdynamischen Region von 85×85 Pixeln. Die eingebaute Möglichkeit der Mittelung benachbarter Pixel erlaubt eine Ausdehnung der hochdynamischen Region über das gesamte Pixelarray. Eine zusätzlich implementierte Schaltung zur Doppelabstastung ermöglicht die Reduktion der durch Prozessschwankungen verursachten Pixel-zu-Pixel Variationen.

Abstract

A High Dynamic Range CMOS Image Sensor with Adaptive Integration Time Control - The scope of this thesis encompasses the development and testing of a CMOS based image sensor. The imaging process consists of the integration of the photocurrent generated by incident light in each pixel. The implementation of a concept for adaptive regulation of the local integration time allows imaging of high dynamic range scenes without loss of information due to over- or underexposure. Depending on the size of the integrated memory, the proposed concept allows the specification of a freely movable image region within which the integration time is regulated. At a chosen maximum integration time of 33 ms the dynamic range of the sensor amounts to 134 dB and covers a range of intensities from $1\,\mathrm{mW/m^2}$ to $5\,\mathrm{kW/m^2}$. The prototype consists of 170×170 pixels with a high dynamic region of 85×85 pixels. The additionally implemented ability to average neighboring pixels allows an expansion of the high dynamic range region over the entire extent of the sensor. An on-chip double sampling circuitry reduces the fixed pattern noise caused by unavoidable device-to-device mismatch.

Contents

Introduction

In recent years a rapidly growing market for vision systems has led to an increasing demand for high performance image sensors. Applications range from automotive systems, security cameras or consumer electronics such as video and digital still cameras to industrial and scientific imaging. Future developments will lead to an ever increasing pressure for system integration especially with regard to portable systems where a reduction of power consumption and overall system size is essential.

A major problem for artificial vision systems is the extraordinarily wide range of scene illuminations found in everyday life. The natural world exhibits a range of light intensities spanning approximately 9 orders of magnitude from bright sunlight down to star-lit scenes. The human visual system, equipped with the unique ability to adapt to the predominant intensity, is capable of covering nearly the whole range of occurring intensities[1]. In contrast to the human eye, most artificial image sensors are not able to capture scenes under extreme illumination conditions without loss of information. The quantity that describes the ability of an optical sensor to cope with such situations is the optical dynamic range. It is defined as the ratio of the maximum non-saturating input signal to the standard deviation of the read noise under dark conditions.

Currently, the dominant technology for image sensors is the charge-coupled device (CCD). While offering low noise operation, high quantum efficiencies and high fill-factors[2] there are significant drawbacks of this technology, which derive from the fundamental principle of its operation. In CCDs the signal is read out by a sequential transfer of charge packets across macroscopic distances through a semiconductor. In order to reach the high transfer efficiency necessary, specialized process technologies, multiple supply and bias voltages and clocking signals with large amplitudes (5-10 V) are required [BLA01] [FOS93]. The consequences are an increased system complexity and power consumption as well as incompatibility with deep-submicron VLSI[3] technology. Furthermore, conventional CCDs typically offer an optical dynamic range of approximately 3-3.5 decades, which reveals their inability to cope with high dynamic range (HDR) situations [YAD04-1].

In recent years image sensors based on CMOS[4] technology have entered competition offering several advantages over conventional CCDs. The usage of standard CMOS processes, originally developed for high volume production of standard CMOS logic and memory chips, permits the realization of imagers that exhibit low voltage operation and low power consumption. In contrast to CCDs, individual read-out of regions of interest is possible. This is due to the fact that the photoreceptor signal is driven over metallic wires. The major advantage of CMOS technology is the possibility of integrating electronic circuits and optical sensors on the same microchip. The integration of control logic, signal chain and individual pixel processing offers the opportunity to reduce the overall system size and at the same time to realize systems-on-a-chip with capabilities not achievable with CCDs.

[1] Thereof approximately 4 decades in a single view [DEV02].

[2] The term ”fill-factor” stands for the ratio of the photosensitive area to the total pixel area.

[3] Very large Scale Integration

[4] Complementary Metal Oxide Semiconductor

This includes the realization of most concepts for the expansion of the optical dynamic range.

A standard CMOS active pixel sensor has a dynamic range of approximately three decades, usually even smaller than a CCD sensor. Several approaches for the realization of CMOS HDR imagers have been investigated so far. Total dynamic ranges of up to 6 decades were reported [LOO01] [SCH00]. The different concepts range from sensors with logarithmic response or capacity adjusting schemes to global or local multiple sampling techniques and others. They will be presented in chapter 2 of this work.

Depending on the concept used different disadvantages have to be accepted. The major drawback of CMOS technology is the mismatch of identical structures due to process variations, which can lead to an excessive fixed pattern noise. The effect on the imager is determined by the photoreceptor circuit and the compensation techniques used for the reduction of this noise. Both are dependent on the chosen method for HDR expansion. Many implementations explored so far suffer from a reduced signal-to-noise ratio or large pixel pitches resulting from a complex pixel circuitry. Another drawback of some concepts is a complex image reconstruction from the recorded data.

The concept presented in this thesis aims at the realization of a HDR imager, which does not suffer from these problems. The dynamic range expansion scheme is based on a pixel-wise adaptive regulation of the respective integration time. A prototype image sensor which offers the opportunity of a freely movable high dynamic range region has been realized and tested. For the reduction of fixed pattern noise caused by unavoidable device-to-device mismatch an on-chip circuitry for double-sampling is used. A total dynamic range of 134 dB (6.7 decades) was measured using a maximum integration time of 33 ms.

The thesis is structured as follows: In the first chapter an introduction to the fundamentals of CMOS imaging is given. Important quantities for the characterization of an image sensor are defined and the CMOS process is discussed regarding its opto-electronic properties and photosensitive devices. Sources of noise are identified and device matching properties are discussed. The second chapter will present a short overview of existing concepts for high dynamic range CMOS imagers that can be found in literature, explaining the basic ideas and the achieved results.

In the third chapter the concept of adaptive integration time control is described, which was used for the sensor developed within this thesis. Critical properties of HDR imagers are identified that led to the development of this regulation scheme. The prototype imager is controlled by a programmable logic array, being part of the electronic test system. Its programming, which is essential for realization of the integration time control, is also explained in this chapter.

The implementation of the image sensor in CMOS technology is presented in chapter four. An overview of the logical chip architecture is given and the global data flow is illustrated. Afterwards, the individual components of the chip and their interplay with each other are described in terms of circuit diagrams and layout implementation.

The final chapter describes the measurements that have been carried out to characterize the image sensor. After a description of the optical and electronic test setup, the results of the measurements are presented, which permits a comparison with other sensors. Finally, sample images of several scenes recorded under varying illumination conditions are shown to illustrate the performance of the imager.

Chapter 1

Fundamentals of CMOS Imaging

In this chapter an introduction to the fundamentals of CMOS imaging is given. The first section describes the absorption of electromagnetic radiation in silicon and provides a basic understanding of the underlying principles of photon-to-charge conversion. The second section introduces different kinds of photodetectors that can be used in CMOS technology. Emphasis is placed on the discussion of the different types of photodiodes. Important parameters for the characterization of image sensors such as quantum efficiency, signal-to-noise ratio and dynamic range are defined. The last section addresses the subject of device-to-device mismatch, which is of high relevance for CMOS image sensors.

Today, complementary metal-oxide semiconductor technology is the dominant technology for VLSI digital and mixed-signal[1] designs. The base material used for the implementation of the circuits is crystalline silicon. Owing to the fact that silicon exhibits a strong absorption in the visible spectral range, the realization of photodetectors working in this range of wavelengths is possible.

The promising possibility to integrate photodetectors as well as additional electronics on the same die[2] has led to the investigation of different photodetectors and the development of numerous complex vision systems. However, a standard CMOS process is not optimized for the implementation of opto-electronic devices. A characterization of photodetectors is not provided by the silicon foundries. Device-to-device mismatch, which arises due to the unavoidable variation of process parameters, can significantly deteriorate the performance of an imager. Additionally, impinging photons will influence active devices in their functionality. In order to minimize the deterioration and to maximize the sensor performance, a basic understanding of the underlying correlations is essential.

1.1 Absorption of Radiation in Silicon

The operation of CMOS photodetectors is based on the absorption of electromagnetic radiation in crystalline silicon. In general, the absorption of a photon in a semiconductor causes the lift of an

[1] The term ”mixed-signal” describes circuits with both analog and digital circuitry on the same piece of silicon.

[2] The term ”die” denotes a piece of silicon that serves as substrate material for the implementation of the electronic devices. In the following it is used interchangeably with the term ”chip”.

electron from the valence band into the conduction band (inner photoeffect). The electron leaves an empty state behind in the valence band, which can be regarded as occupied by a positive charge known as a ”hole”. The hereby generated electron-hole pair has to be separated in the high electric field of a photodetector before it recombines again. The separated charge carriers contribute to a measurable current (”photocurrent”).

A basic understanding of the absorption behaviour of a material can be achieved by the consideration of the possible photoexcited electronic transitions. Depending on the initial and final electronic states, photoexcited transitions can be classified as interband, intraband or trap-to-band transitions. In figure 1.1 the different kinds of transition are illustrated by the example of a schematic band structure. In intraband transitions the initial and the final state of the carrier is located within the same band. Trap-to-band transitions are possible due to the existence of states in the bandgap caused by crystal defects that act as a trap for carriers. Interband transitions can be further subdivided into so called *direct* and *indirect* transitions, which can be explained as follows.

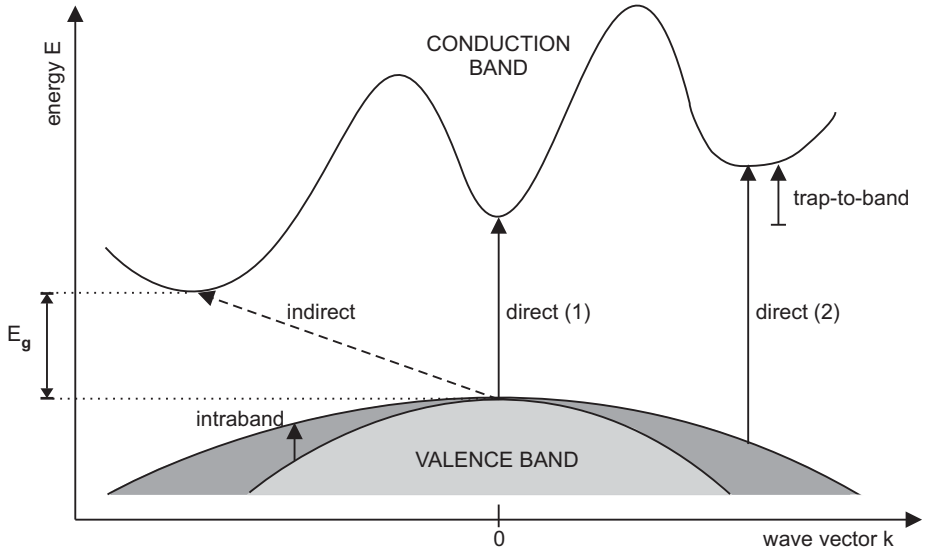


Figure 1.1: Schematic diagram for the illustration of different photoexcitation processes. The indirect transition is the only one where a phonon is involved.

A transition can only take place if the laws of conservation of momentum and energy are fulfilled. When an electron absorbs a photon of frequency ν its energy increases by the amount $h\nu$. For a transition from the valence band to the conduction band the minimum energy required is the bandgap energy E_g, which is the difference between the minimum conduction band and maximum valence band energies:

$$E_{ph} = h\nu \geq E_g \tag{1.1}$$

For silicon this energy amounts to E_g=1.12 eV (at T=300 K) [IBA95]. If the initial state in the valence and the final state in the conduction band have the same wave vector k the transition can take place by a simple absorption of a photon with an energy of $\geq E_g$. Such a transition is referred to as a *direct* one.

On the other hand, if the respective states have different wave vectors, the transition requires an additional momentum transfer to conserve the total momentum of the system. Usually, the momentum of the incident photon ($\hbar k$) is small compared to that of the electron, therefore it can be neglected here. The required momentum can be delivered by a phonon, i.e. a quantized lattice vibration. How-

ever, this requires the simultaneous interaction of an electron with a photon and a phonon, which results in a smaller probability for such a transition than for a direct one. A transition involving a phonon is called an *indirect* transition (compare with figure 1.1).

According to this naming convention, semiconductors can be divided into two groups: Those where the maximum energy level of the valence band and the minimum energy level of the conduction band are located at the same wave vector k are referred to as *direct* semiconductors (e.g. GaAs). Those where the respective levels lie at different values of k are called *indirect* semiconductors. Silicon is an *indirect* semiconductor. This classification has direct consequences for the absorption behaviour of the respective material.

Regarding photons with the same wavelength, each photon that impinges the semiconductor has the same probability of being absorbed. Consecutive layers of material with the same thickness will always absorb the same fraction of photons that impinge on their respective surface. In consequence, the loss of intensity $-dJ$ along the distance dx is proportional to the local intensity at the depth x:

$$-\frac{dJ}{dx} = \alpha J(x) \tag{1.2}$$

where the proportionality factor α is the absorption coefficient of the respective material. Integration of equation 1.2 yields in *Lambert-Beer's* law of absorption [BER93-1]:

$$J(x) = J_0\, e^{-\alpha x} \tag{1.3}$$

where J_0 is the intensity at $x = 0$. The absorption coefficient of silicon shows a strong dependency on the photon energy. This dependency is depicted in figure 1.2, it can be explained as follows.

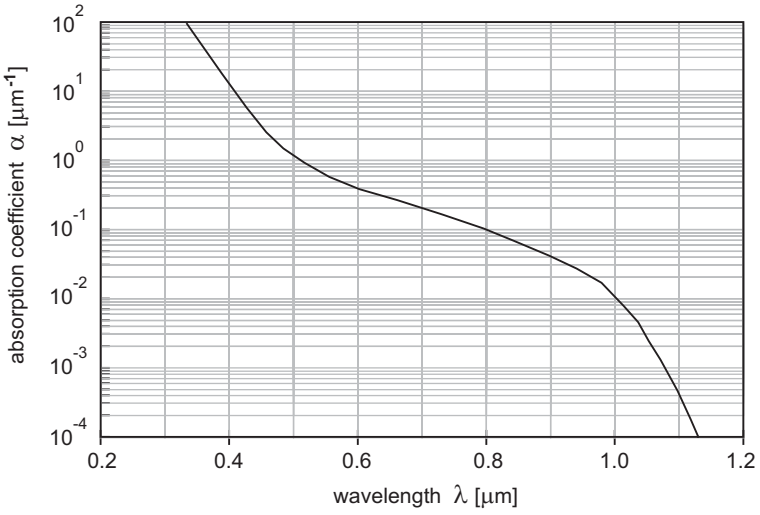


Figure 1.2: Absorption coefficient α of silicon as a function of the wavelength. The cutoff wavelength of silicon is $\lambda_g = 1.1\,\mu$m (data taken from [HAU99]).

From the discussion above it follows that in the case of silicon the absorption of photons close to the cutoff wavelength ($\lambda_g = 1.1\,\mu$m) requires the simultaneous interaction with a phonon. Therefore, α increases much slower with decreasing wavelengths than it is the case for a direct semiconductor such as GaAs. The increase itself can be explained by the fact that the number of possible energy states for excited electrons rises with increasing potential energy. For wavelengths smaller than about 500 nm a steeper increase of α can be observed. The reason for this is that here the transition

without a momentum change becomes possible, i.e. there is no need for an interaction with a phonon (transition direct(1) in figure 1.1).

Photons with an energy of less than the bandgap energy E_g, i.e. $\lambda > 1100\,\text{nm}$, cannot be absorbed through band-to-band transitions, therefore it should be $\alpha = 0$ for these wavelengths. However, a weak absorption is still measurable. This can be explained by the remaining possibility of trap-to-band transitions (see above). A second reason is the absorption by free electrons brought to the conduction band by thermal excitation.

Returning to Lambert-Beer's law of absorption, equation 1.3 expresses that the light intensity decreases exponentially with increasing depth of penetration. The knowledge of this relation permits the calculation of the number of photons that get absorbed in a layer of the thickness d. Let $N_{ph}(x)$ be the number of photons per unit time and unit area, then is $N_{ph}(x) = J(x)/h\nu$. Regarding equations 1.2 and 1.3 the number of absorbed photons per volume $g(x)$ now can be expressed as

$$g(x) = -\frac{dN_{ph}}{dx} = \alpha\, N_{ph}(x) = \alpha\, N_0\, e^{-\alpha x} \tag{1.4}$$

The integration of this equation from the surface of the material ($x = 0$) to a specific depth d results in the number of absorbed photons in a layer of thickness d per unit time:

$$G(d) = \int_0^d \alpha\, N_0\, e^{-\alpha x}\, dx = N_0 \cdot \left(1 - e^{-\alpha d}\right) \tag{1.5}$$

Taking the corresponding values of α from figure 1.2 the percentage of absorbed photons of a particular wavelength can be calculated in dependence of the depth d in silicon ($100 \cdot G(d)/N_0$). The resulting absorption curves for five exemplary wavelengths are shown in figure 1.3. It can be seen that the visible spectrum is nearly completely absorbed within the first 10 μm, whereas infrared light penetrates deeply into the substrate due to its weak absorption. Therefore, infrared detectors would have to be located deeply in the substrate to reach an acceptable efficiency.

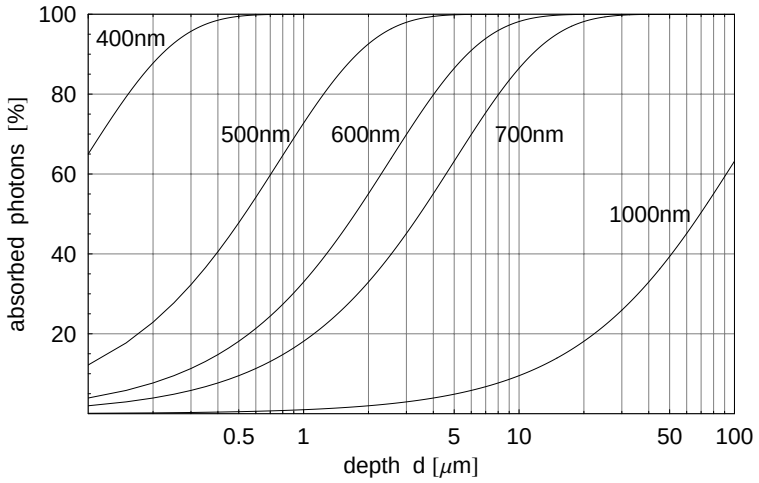


Figure 1.3: Percentage of absorbed photons in dependence of the penetration depth and the corresponding wavelength.

Quantum Efficiency

Regarding the total efficiency of the photon-to-charge conversion, further effects have to be taken into account. Even if the energy of the incident photons is larger than the bandgap energy E_g, not every photon will lead to the generation of an electron-hole pair that contributes to the measured photocurrent. First, only a fraction of the incident photons will lead to the generation of a carrier pair at all. A part of the light is absorbed in additional layers on top of the semiconductor (such as the passivation layer consisting of SiO_2) or reflected at the boundaries of these layers, especially at the surface of the substrate. The photons that reach the inner semiconductor will be either absorbed or cross the material without interaction (transmission). Second, even if an electron-hole pair is generated it will not contribute to the photocurrent if it recombines beforehand.

The total efficiency of the photon-to-charge conversion that includes all these effects determines the spectral sensitivity of the photoreceptor and is referred to as the external (measurable) quantum efficiency η. It is defined by the ratio of the number of electron-hole pairs per time that contribute to the photocurrent to the number of incident photons per time. The energy of an incident photon is hc/λ. Let P_{ph} be the incident light power, then the number of incident photons per time is $n_{ph}/t = P_{ph}\lambda/hc$. The number of generated electrons per time is $n_e/t = I_{ph}/q_e$, where I_{ph} is the measured photocurrent and q_e is the elementary charge. For the quantum efficiency this yields

$$\eta = \frac{n_e/t}{n_{ph}/t} = \frac{I_{ph}}{P_{ph}\lambda} \cdot \frac{hc}{q_e} \tag{1.6}$$

In order to determine the spectral quantum efficiency of a given detector, the photocurrent and the incident light power have to be measured in dependence of the wavelength.

1.2 Photodetectors in CMOS Technology

Different kinds of photodetectors are available in CMOS technology. The most popular of these is the *pn*-junction diode. Owing to the fact that a photodiode was also used in the image sensor realized within this thesis, the following section will place an emphasis on the description of the properties of this detector type. For completeness, at the end of the section also other existing types of photodetectors will be shortly described.

1.2.1 Photodiodes

In a photodiode a depleted semiconductor region with a high electric field is used to separate photogenerated electron-hole pairs. This way, incident light is converted to a measurable photocurrent. The *pn*-junction diode consists of two layers of opposite doping in a semiconductor material. The *n*-type layer is produced by the implantation of donor atoms (atoms with 5 valence electrons, such as phosphorus, arsenic or stannous) in the semiconductor (such as silicon with 4 valence electrons). The *p*-type layer is obtained by the implantation of acceptor atoms (atoms with 3 valence electrons, such as boron, aluminium or gallium). In thermal equilibrium there is a depletion layer between the two layers: The concentration gradient of the charge carriers (electrons and holes respectively) between the *n*-type and the *p*-type layer leads to a diffusion current. In turn, the carrier transfer through this current leads to an electrical field which causes a current to flow in the opposite direction. In equilibrium, the two currents compensate each other.

In figure 1.4 the corresponding space-charge distribution is shown for the assumption of an abrupt change of the doping concentrations. The electrical properties of the photodiode are determined by

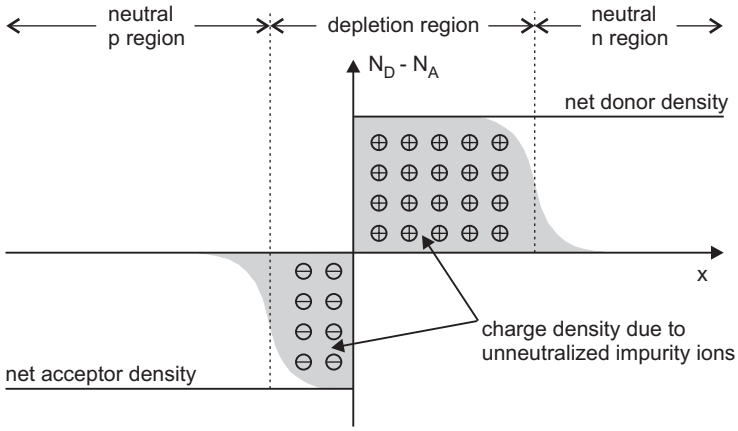


Figure 1.4: Space-charge distribution for an abrupt *pn*-junction in thermal equilibrium. The grey region indicates the majority-carrier distribution [SZE81-1].

the absolute value of the doping concentration and the change of this concentration in space (the doping profile). For the assumed abrupt junction the width W_d of the depletion layer can be calculated by [SZE81-1]

$$W_d = \sqrt{\frac{2\epsilon_{si}}{q_e}\left(\frac{1}{N_A}+\frac{1}{N_D}\right)(V_{bi}-V)} \tag{1.7}$$

where V is an externally applied voltage and ϵ_{si} is the permittivity of silicon. N_A and N_D are the acceptor and donor densities respectively. V_{bi} is the diffusion potential or build-in potential and can be calculated by

$$V_{bi} = \frac{kT}{q_e}\ln\left(\frac{N_A N_D}{n_i^2}\right). \tag{1.8}$$

where n_i is the intrinsic charge carrier density[3]. Doping concentrations used in current CMOS processes result in a maximum width of a few micrometer (most even lower). The small width leads to a high electric field in the depletion region, which is used for the separation of the photo-generated electron-hole pairs. As a matter of course, a wider depletion region results in a higher quantum efficiency. With regard to equation 1.7 it can be widened by increasing the external voltage V in reverse direction (V negative). Carriers generated in the bulk of the semiconductor can also contribute to the detected signal if they are less than a minority-carrier diffusion length away from the depletion region. The sensitive volume of the photodiode is therefore larger than its depletion region.

Another important property of the junction is its capacitance. First, it influences the frequency response of the diode, which will be discussed at the end of this section. Second, if the diode is used for photocurrent integration in an active pixel sensor, the junction capacitance will determine the charge-to-voltage conversion gain as it is the main contributor to the integration capacitance. The junction capacitance can be calculated by

$$C_j = A \cdot \sqrt{\frac{\epsilon_{si} q_e N_A N_D}{2(N_A+N_D)(V_{bi}-V)}} \tag{1.9}$$

[3] An "intrinsic" semiconductor is a perfect semiconductor crystal without any defects or impurities. The intrinsic carrier concentration of Si is $1.5 \cdot 10^{10}\,\mathrm{cm}^{-3}$ (at T=300 K) [IBA95].

where A is the area of the *pn*-junction [DRO99]. Typical capacitances of a $5 \times 5\,\mu\text{m}^2$ photodiode are in the range of some ten fF.

Photodiodes in CMOS Technology

By definition, in a CMOS process both types of MOS[4] transistors, the p-channel as well as the n-channel type, must be realizable. In consequence, different kinds of doping layers are available for the realization of different types of *pn*-junctions. Each of these junctions can be used as a photodiode. In practice, there is no abrupt change in the doping concentration as it was assumed above. The doping layers are produced by the diffusion of donor or acceptor atoms. Alternatively they are produced by ion implantation, where the ions of the doping atoms are accelerated by a high electric field and "shot" in the semiconductor. All of these techniques result in gradients in the doping concentration instead of sharp borders.

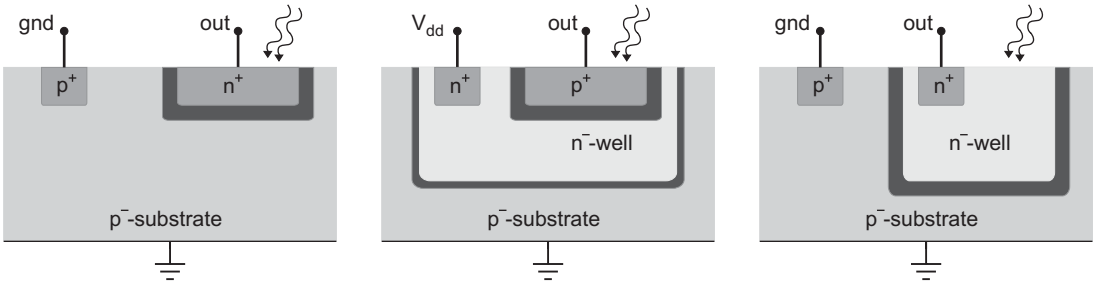


Figure 1.5: Cross sections of the three *pn*-junctions usable in a standard CMOS process. From left to right: n^+/substrate diode, p^+/nwell diode and nwell/substrate diode.

Figure 1.5 shows the cross sections of the different types of junctions that can be used as photodiodes in a standard CMOS process with a lightly doped p-substrate. The depletion regions are drawn in black. As can be seen, a photodiode not only consists of a vertical *pn*-junction (bottom junction), there are always also lateral junctions involved. For the calculation of the total capacitance both parts, the bottom as well as the sidewall capacitances, have to be taken into account. Owing to their different doping profiles, the junctions shown in figure 1.5 will differ in their electrical properties and in particular in their quantum efficiency. For standard CMOS processes the chip manufacturer does not deliver characterization data for the available photodetectors, hence the designer is forced to characterize the different junctions by measurements. However, with regard to the absorption process discussed in the previous section some general statements can be made concerning the quantum efficiencies of the junctions. The n^+-substrate-diode[5] is located close to the surface of the silicon wafer. Therefore, the maximum of its quantum efficiency η will be at shorter wavelengths. As most of the light is absorbed near the surface, its absolute value is comparatively high. The quantum efficiency of the p^+-nwell-diode will be low since it cannot collect charges from greater depths (shielded by the additional nwell-substrate junction). The third photodiode (nwell-substrate junction) will exhibit a large response time since charge carriers can diffuse to the junction from above and below (see the discussion of response speed later).

The spectral quantum efficiencies of the photodiodes available in the used CMOS process were measured in this thesis. The results of these measurements are presented in chapter 5.

[4] Metal Oxide Semiconductor

[5] The suffix +/- indicates the doping concentration (high/low).

Device Noise

In figure 1.6 the equivalent circuit diagram for the noise analysis of a photodiode that is used in reverse direction is shown. The load connected to the photodiode is represented by the resistance R_L. The electrical properties of the photodiode are given by the series resistance R_s, the junction capacitance C_j (cf. equation 1.9) and the junction resistance R_j. Compared to the other resistances, the series resistance is usually small so that it can be neglected. The junction resistance is very high. The photocurrent I_{ph} that is generated by the incident light can be taken from equation 1.6:

$$I_{ph} = q_e \eta P_{ph} \cdot \frac{\lambda}{hc} \tag{1.10}$$

To maintain consistency with the noise equations below, it is assumed here, that P_{ph} is the rms[6] incident light power and in consequence I_{ph} the rms photocurrent. Thermal generation of electron-hole pairs inside the depletion region will lead to a dark current I_d:

$$I_d = q_e A n_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) \tag{1.11}$$

where D_p and D_n are the carrier diffusion coefficients (D_p for holes and D_n for electrons) and L_p and L_n are the diffusion lengths of the minority carriers.

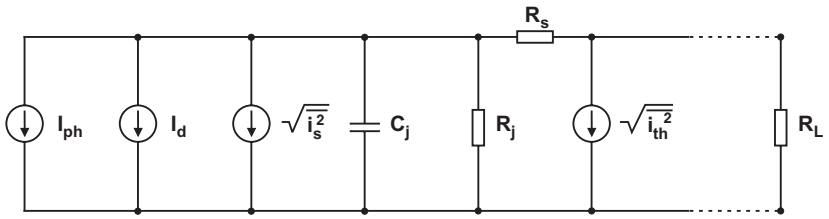


Figure 1.6: Equivalent circuit diagram of a photodiode for noise analysis [SZE81-2].

There are two sources of noise that have to be taken into account: thermal noise and shot noise. Neglecting the small series resistance R_s, the thermal noise is contributed by the junction resistance R_j and the load resistance R_L. Defining an equivalent resistance value R_{eq} by $1/R_{eq} = 1/R_L + 1/R_j$ the thermal noise is given by

$$\overline{i_{th}^2} = 4kT \frac{1}{R_{eq}} \cdot \Delta f \tag{1.12}$$

where Δf is the bandwidth of the system. The shot noise on the other hand occurs due to the randomness of the generation of photocurrent and dark current. It amounts to

$$\overline{i_s^2} = 2q_e(I_{ph} + I_d) \cdot \Delta f. \tag{1.13}$$

The equations above can be used to calculate the signal-to-noise ratio (SNR) of the system by dividing the signal power by the noise power:

$$\mathrm{SNR}_{power} = \frac{I_{ph}^2 R_{eq}}{(\overline{i_s^2} + \overline{i_{th}^2}) R_{eq}} = \frac{\left(q_e \eta P_{ph} \frac{\lambda}{hc}\right)^2}{2q_e \left(q_e \eta P_{ph} \frac{\lambda}{hc} + I_d\right) + 4kT/R_{eq}} \cdot \frac{1}{\Delta f}. \tag{1.14}$$

[6] root mean square

Therefore, in order to enhance the SNR the load resistance R_L should be increased, the system bandwidth should be decreased and a photodetector with a quantum efficiency as large as possible should be chosen.

Response speed

The response speed of a photodiode is limited by three factors: the drift time in the depletion region, the capacitance of the depletion region and the diffusion of carriers. In order to increase the amount of collected charge carriers, the depletion region should be sufficiently wide. However, it should not be too wide, this would result in an increased response time due to the limited transit-time of charge carriers. On the other hand, if the depletion layer is too thin, this will lead to an increased junction capacitance and in consequence result in a large $R_L C_j$ time constant (R_L is the load capacitance). Therefore, a tradeoff has to be found for the width of the depletion layer.

The last factor, the diffusion of carriers, results from the fact that carriers generated outside of the depletion region have to diffuse to the junction, which results in a time delay. This effect can be reduced by the selection of a junction lying close to the surface of the semiconductor.

1.2.2 Photogate

A photogate is a device which is closely related to a CCD[7] in its functionality, but implementable in a standard CMOS process. In principle, it is no more than a MOS capacitor that is exposed to light. In figure 1.7 a cross section of the photogate structure is shown. The photogate PG as well as the gate TX of the transfer device consist of polysilicon. Between these gates a diffusion is located to permit the realization of the separate transfer gate TX in such CMOS processes that offer only a single layer of polysilicon. In a dual polysilicon process this diffusion can be omitted by the use of a small overlap of the gates PG and TX [FOS97]. The output node is a floating diffusion.

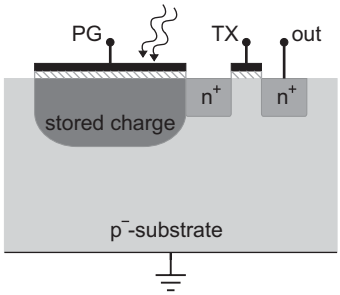


Figure 1.7: Cross section of the photogate structure.

The photogate operates by integrating the charges generated by the light that penetrates the polysilicon gate PG. During the integration process PG is held at V_{dd} and the photogenerated electrons are collected in the potential well, which is created by the voltage at PG. Gate TX is held at a lower voltage than PG, isolating the collected charge from the floating diffusion output node. When the integration time has passed, the signal is read out as follows: First, the output node is reset to a voltage close to V_{dd} by the connected pixel circuitry. The exact reset level is read out to permit a later offset correction. Subsequently, the electrons collected under the photogate are transferred to the output node by holding PG at ground level for a short time. Afterwards, the changed signal at the

[7] Charge Coupled Device

output node is read out and the voltage at PG is changed to V_{dd} in order to start the next integration period. It should be noted that the potential well drawn under the photogate has a small depth. For instance, in a 2 μm process the typical depth is less than 0.5 μm [MOI00-1].

The first advantage of the photogate is a comparatively small dark current. The reason for this is that there is no junction present in the device. The dark current is caused only by carrier recombination in the substrate [MOI00-2]. A second advantage arises from the fact that true correlated double-sampling can be easily realized, since the reset value can be read out directly before the signal is sampled. Nevertheless, it has to be taken into account that there is an additional amount of transfer noise due to fact that the conductance of the transfer channel contributes thermal noise.

The major drawback of the photogate is its low quantum efficiency. In order to reach the semiconductor the incident light first has to penetrate through the layer of polysilicon. Especially the response to blue light is very poor since the gate material absorbs this part of the spectrum. The quantum efficiency varies over the charge integration interval. The reason for this variation is that the depletion width under the photogate is a function of the surface potential, which changes as signal charges are collected at the surface [YAD04-2].

1.2.3 Phototransistor

A device which is rarely used in image sensor arrays is the bipolar phototransistor. In a CMOS process two types of bipolar transistors are available, which differ in the spatial arrangement of the diffusions: the lateral and the vertical type. The vertical type is always created when a PMOS[8]-FET[9] is implemented (assuming a nwell-process), therefore in a CMOS process it is called a parasitic device. In contrast to a conventional bipolar transistor a phototransistor is characterized by a large collector/base junction area which is used for the collection of photogenerated charge. Figure 1.8 shows the cross section of a vertical parasitic pnp transistor and the corresponding equivalent circuit diagram. The capacitance C_{CB} is the capacitance of the reverse-biased collector/base junction. Since the substrate in figure 1.8 is connected to ground the pnp transistor is used in a common-collector configuration.

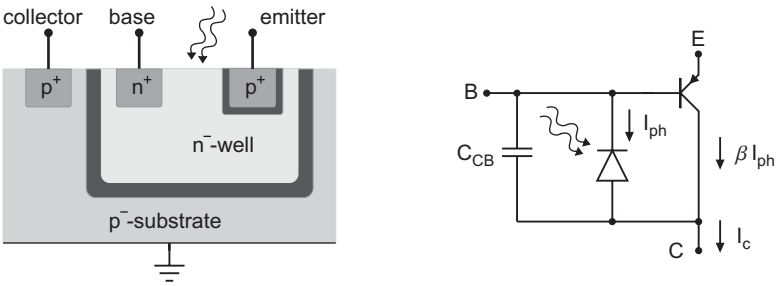


Figure 1.8: Cross section and equivalent circuit diagram of the vertical pnp-phototransistor.

In normal operation mode of a bipolar transistor, the base-emitter current is amplified by the factor h_{FE} (dc common-emitter current gain), which can be much larger than unity. Phototransistors are operated with a floating base terminal. Charge carriers generated by the incident light in the collector/base junction will lead to a photocurrent I_{ph}. This current in turn causes a decrease of the base potential. In addition to this current, electrons generated in the base region and swept from the

[8] P-channel MOS
[9] Field Effect Transistor

collector into the base give rise to a further decrease. The decrease of the base potential is equivalent to an increase of the emitter-base voltage, which leads to an injection of holes from the emitter across the base to the collector. In addition to the photocurrent I_{ph} there will always be a (thermally generated) dark current I_d also flowing through the photodiode. This current will contribute to a higher collector current in the same way like the photocurrent. The resulting total collector current is

$$I_c = (1 + h_{FE}) \cdot (I_{ph} + I_d) \tag{1.15}$$

Owing to the fact that the base/collector junction is used for the light detection, the course of the quantum efficiency of the phototransistor will be dominated by this junction. But as a consequence of the current amplification, the effective quantum efficiency is (1+h_{FE}) times larger. Unfortunately, the noise is amplified by the same factor.

There are two drawbacks connected with the use of a phototransistor. First, its response time is long due to the large base/collector capacitance and is increased further by the gain of the detector on account of the feedback effect. The typical increase in comparison with the photodiode is a factor of about 100 [SZE81-3], which disables the phototransistor for high frequency applications. Second, the amplified dark current severely limits the usage of the phototransistor in integration based photocircuits. The large dark current would lead to a quicker discharge of the integration capacitance and hence limit the maximum integration time due to the risk of saturation.

1.3 Dynamic Range and Signal-to-Noise Ratio

Two parameters often used for the description of the performance of an imager are the signal-to-noise ratio (SNR) and the dynamic range (DR). The signal that is to be measured by a photodetector is the incident light intensity J. In most cases, an electronic detection circuitry is used to convert the generated photocharge into a signal voltage, e.g. by means of a capacitance. The SNR is then defined as the ratio of the (mean) actual signal voltage $\overline{V_{sig}}$ to the standard deviation of the measured noise $\sigma(V_{sig})$. Absolute values are specified in decibel, therefore it is defined as:

$$\text{SNR} = 20 \cdot \log\left(\frac{\overline{V_{sig}}}{\sigma(V_{sig})}\right) \tag{1.16}$$

The DR[10] on the other hand, indicates the input signal range, where the sensor delivers a useful signal. It is defined as the ratio of the maximum light intensity J_{max}, which does not saturate the sensor, to the intensity J_{min}, which results in a signal-to-noise ratio of 1. Like for the SNR[11], the values for the DR are specified in decibel:

$$\text{DR} = 20 \cdot \log\left(\frac{J_{max}}{J_{min}}\right) \tag{1.17}$$

An illustration of these definitions is shown in figure 1.9. The signal does not start in the point of origin since the quantum efficiency is always smaller than one. Different kinds of noises influencing both values are shown schematically. Fixed pattern noise is caused by device-to-device mismatch, which will be discussed in the next section.

It should be noted here, that the definitions above are not consistent with usage elsewhere in optical physics: Owing to the fact that usually the signal voltage is proportional to the power of the incident light, a pre-factor of 10 is used instead of 20. On the other hand, in electrical engineering power is always associated with the square of voltage levels. For this reason the presented definitions are used throughout the imager related literature and in almost all sensor specifications [SEI99].

[10] Dynamic Range
[11] Signal-to-Noise Ratio

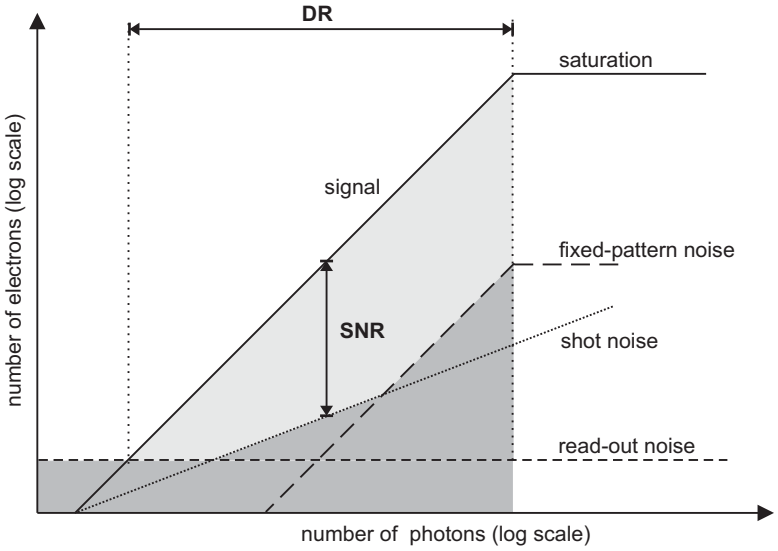


Figure 1.9: Schematic illustration of the definitions of SNR and DR in connection with the arising sources of noise (cf. [KNO92]).

1.4 Matching Considerations

The major advantage of CMOS technology is the possibility to integrate electronic circuits and optical sensors on the same microchip. In an image sensor this is used for the implementation of in-pixel circuitry, analog and digital signal processing, control logic and several other applications. However, the major disadvantage of CMOS technology is that devices, which are identically drawn in the layout do not show an identical behaviour in reality. The reasons for the observed mismatch can be found in the production process. Limited imaging quality of the photolithographic process as well as parameter variation across the wafer lead to the variation of the electrical properties of devices. Many analog building blocks like current mirrors or differential pairs rely on identical behaviour of their sub-devices. Mismatch can severely deteriorate their performance. In CMOS imagers, mismatch of the in-pixel circuitry leads to the commonly known problem of fixed pattern noise: Non-uniformities in the pixel response, which are time-invariant but randomly distributed over the pixel array. Passive as well as active devices are affected with mismatch. In general, most of the FPN[12] is not caused by mismatch of the photodiodes. The diodes are comparatively large which reduces their mismatch. The major part of the FPN arises from transistor mismatch in the connected pixel circuitry. Concerning the imager realized within this thesis transistor mismatch is also of particular relevance for the design of the comparators, output amplifiers, sample-and-hold stages and for the sense amplifiers.

1.4.1 Drain Current Mismatch

The effect of transistor mismatch depends on the way the respective transistor is operated. In the following, mismatch will be discussed at the example of a transistor operated in strong inversion. The results will be compared with the mismatch of transistors in weak inversion, which are used in sensors with logarithmic response.

A MOS transistor is operated in strong inversion or equivalently in ”saturation” if the gate-source-

[12] Fixed Pattern Noise

voltage V_{GS} is higher than the threshold voltage V_T and the drain-source voltage V_{DS} is higher than $V_{GS} - V_T$. The drain current I_D is then a quadratic function of the gate-source-voltage V_{GS}:

$$I_D = \frac{\beta}{2}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \quad \text{with} \quad \beta = \mu C_{ox}\frac{W_{eff}}{L_{eff}} \tag{1.18}$$

where μ is the channel mobility, C_{ox} the gate oxide capacitance and λ the channel length modulation parameter. W and L are the width and length of the transistor channel drawn in the layout by the designer. They can be chosen freely within the limits of the process. The effective width W_{eff} and length L_{eff} are smaller than the drawn values. Encroachment of the field oxide[13] into the channel shortens the width and lateral diffusion of the source and drain diffusions shortens the length of the transistor. Variation of I_D is caused by a variation of the transconductance parameter β and the threshold voltage V_T. The relative error of I_D caused by these variations can be calculated by the quadratic summation of the individual errors [LAK86]:

$$\frac{\sigma(I_D)}{I_D} = \sqrt{\frac{\sigma^2(\beta)}{\beta^2} + 4\frac{\sigma^2(V_T)}{(V_{GS} - V_T)^2}} \tag{1.19}$$

The error in I_D therefore decreases with increasing gate-to-source voltage V_{GS}. This behaviour is confirmed by measurements [LOV98].

When a transistor is operated in weak inversion or equivalently in "subthreshold region" the square-law in equation 1.18 is not valid any more. The transistor enters the subthreshold region if V_{GS} gets smaller than V_T. The $I_D - V_{GS}$ characteristic now changes to an exponential relation. Under the assumption of a bulk-source voltage $V_{BS} = 0\,V$ and $V_{DS} \gg kT/q_e$ it can be derived [GEI90-1]

$$I_D \simeq I_{D0}\frac{W_{eff}}{L_{eff}} \cdot e^{\frac{V_{GS} - V_T}{nV_{th}}} \quad \text{with} \quad I_{D0} \simeq \mu C_{ox}\frac{2(nV_{th})^2}{e^2} \tag{1.20}$$

where n is the *subthreshold slope factor*, which is determined by process parameters (typical values: $n = 1..2$), and $V_{th} = kT/q_e$ is the so called *temperature potential*. The parameter T is the device temperature, k is the Bolzmann's constant and q_e is the elementary charge. Owing to the fact that the drain current is exponentially dependent on the threshold voltage, the mismatch shows a high sensitivity for variations of V_T. The calculation of the relative error of I_D yields [LOO99-2]:

$$\frac{\sigma(I_D)}{I_D} = \sqrt{\frac{\sigma^2(\beta)}{\beta^2} + \frac{\sigma^2(V_T)}{(nV_{th})^2}} \tag{1.21}$$

In comparison with the error in the saturation region (equation 1.19) a much higher contribution of the V_T mismatch results. At room temperature the temperature potential is $V_{th} \approx 25\,\text{mV}$, which results in a small denominator of the second term in equation 1.21 and consequently a large influence of $\sigma(V_T)$. Transistors operating in weak inversion are used in image sensors with logarithmic response. The exponential current-voltage law in equation 1.20 is reversed into a logarithmic voltage-current law. The resulting mismatch $\sigma(V_{GS})$ for the measured gate-source-voltage directly depends on $\sigma(V_T)$ [LOO99-2].

Equations 1.19 and 1.21 describe how the error of I_D depends on $\sigma(V_T)$ and $\sigma(\beta)$. The variation of these parameters in dependence of the design parameters W and L has been extensively investigated [LOV98] [PEL89]. Variations in V_T are assumed to be caused by variations in the substrate

[13] Field oxide is a thick layer of silicon dioxide that is grown to achieve isolation between active regions.

doping. It was found that

$$\sigma(V_T) = \frac{A_{VT0}}{\sqrt{W_{eff} L_{eff}}}, \tag{1.22}$$

where A_{VT0} is a process dependent constant in first approximation. In consequence, devices with a larger effective area $W_{eff} \cdot L_{eff}$ exhibit a smaller threshold voltage mismatch. It should be noted here, that the drawn length L is stronger reduced than the drawn width W: $(W - W_{eff}) < (L - L_{eff})$. Therefore, while comparing two devices with equal layout area, a short channel transistor (small length) has a smaller effective area than a narrow channel transistor (small width) and consequently a poorer matching.

Mismatch of the transconductance parameter β is assumed to be caused mainly by variations in the mobility. The following expression could be derived for mismatch of β [LOV98]:

$$\frac{\sigma^2(\beta)}{\beta^2} = \frac{A^2_{L^2}}{L^2 W} + \frac{A^2_{W^2}}{W^2 L} + \frac{A^2_{\beta^2}}{WL} \tag{1.23}$$

where A_{W^2}, A_{L^2}, A_{β^2} are constants. The first term on the right side ($A^2_{L^2}/L^2W$) becomes significant for short channel devices and leads to an increase of β mismatch. The term $A^2_{W^2}/W^2L$ on the other hand, increases for narrow channel transistors but not to the same extend as the first term in the other case. The last term in equation 1.23 is responsible for the fact that β mismatch follows the same general trend as V_T mismatch, i.e. it is proportional to $1/\sqrt{WL}$. Only short channel devices will introduce a significant deviation from this trend.

1.4.2 Reduction of Mismatch

The designer of an electronic circuit has several possibilities to reduce the mismatch of critical devices or to minimize the effect of mismatch on the operating precision of the circuit. In principle, the possibilities for the reduction of mismatch can be divided into two different categories: active techniques and passive techniques.

Active techniques rely on the implementation of an additional circuitry, which is used to reduce the influence of existing mismatch. Examples for this can be found in this work: Double-sampling is used for the reduction of FPN, which is described in section 4.4.1. Charge injection compensation is used in the sample-and-hold circuits (cf. section 4.4.2).

Passive techniques rely on the realization of an optimized layout of the critical devices. Unfortunately, most of these techniques come at the expense of area and wiring. Owing to the demand for a pixel size as small as possible their usefulness for the design of the pixel array is limited. Nevertheless they play an important role for the design of precise analog circuits like sense amplifiers, comparators and operational amplifiers in this work. Some general guidelines can be given for the layout of critical devices [HAS01]:

Large active areas: According to equations 1.22 and 1.23 transistor mismatch decreases with increasing effective area. It was shown that mismatch increases for short channel devices, which should therefore be avoided in critical circuits. Care should be taken that the increase of the device size does not lead to an increase of other errors (e.g. charge injection).

Close proximity: Global variations of process parameters (e.g. gate oxide thickness) across the wafer will introduce additional variations of electrical properties (e.g. gate oxide capacitance C_{ox}). In order to minimize these effects matched devices should be placed with minimum distance from each other.

Multiple identical devices: Matched devices should be designed identically. If possible, each device should be divided into smaller ones (again identical for all matched devices) that can be interlocked (e.g. ”finger design” of transistors) with those of the other device (see *common-centroid layout*).

Common-centroid layout: Common-centroid is a method to arrange the subparts of the devices that should match in such a way that the influence of process parameter gradients is averaged out. Attention must be paid to the fulfillment of *same orientation* (see below) if this method is used. Examples for common-centroid layouts can be found in sections 4.3.2 (comparator design) and 4.3.8 (sense amplifier design).

Same orientation: Matched transistors should have the same orientation of their channels and the same direction of current flow. Stress- and tilt-induced mobility variations can otherwise lead to variations of several percent in their transconductance. Thus, even minimally matched transistors should follow this guideline. The mismatch considerations in the last section referred to devices which were neither rotated nor mirrored. Misalignment can further increase the mismatch. The requirement for the same direction of current flow can also be fulfilled by splitting the concerned transistors in a way that each transistor contains an equal number of segments oriented in each direction. The sensitivity of mirrored or rotated devices to mismatch can be of particular relevance for the design of device arrays since mirrored cells can hereby result in an extra fixed pattern noise component.

Same surrounding: In order to avoid edge effects in arrayed transistors dummy devices should be placed at the borders of the array to provide the same surrounding for each transistor. Dummy devices are not actively used, instead they are connected to a constant potential, which prevents channel formation beneath them. The spacing between the dummy gates and the real transistors must be equal to the spacing between the real transistors (or its segments). Attention has to be paid to metal traces which cross active areas: Distribution and potential of these traces should be the same for all matched devices.

Same temperature: Temperature has a significant influence on the behaviour of the individual devices (e.g. see equation 1.20). Consequently, transistors should be placed on the same isotherm to reduce mismatch caused by temperature differences of the concerned devices.

Chapter 2

Existing Concepts for HDR Sensors

In the following chapter a short overview of existing concepts for high dynamic range CMOS image sensors is presented. Examples for different concepts are given explaining the basic ideas and the achieved results. To begin with, the difference between passive and active pixel sensors is described. Each of the following sections is dedicated to a different class of HDR concepts. The first section presents sensors based on the control over integration time. Next, sensors with logarithmic response are discussed. The last two sections explain the well capacity adjustment scheme and photoreceptors which are based on pulse frequency modulation.

2.1 Active and Passive Pixel Sensors

Prior to a discussion of the individual concepts for widening the dynamic range of a sensor, the fundamental difference between the two main groups of CMOS pixel sensors, the active and the passive sensor, is explained in this section.

The passive pixel sensor was introduced by Weckler [WEC67]. The basic architecture of this sensor is depicted in figure 2.1. It consists of just two devices, a photodetector (usually a photodiode) and one transistor, which is used to connect the diode to the column read-out line. Mostly, the output signal is detected by a charge integration amplifier, which is connected to the read-out line. While offering a high fill-factor, the passive pixel sensor has some severe disadvantages. The large capacitive load at its output leads to a comparatively low read-out speed. Additionally, the read-out noise is typically high [YAD04-1].

The disadvantages of the passive pixel sensor led to the introduction of a sensor with an in-pixel active amplifier [NOB68]. An example for such an active pixel sensor is shown in figure 2.2. Transistor M_2 operates as a source-follower, which isolates the sense node from the capacitance of the read-out line. The load of this buffer is located outside the array. There is one active-current-source load for each column of pixels. Transistor M_3 connects the output to the read-out line, whereas transistor M_1 is used to reset the photodiode.

Commonly, all pixel sensors that employ an in-pixel amplifier are referred to as active pixel sensors (APS), independently of an existing reset transistor. In this sense, most of the sensor circuits

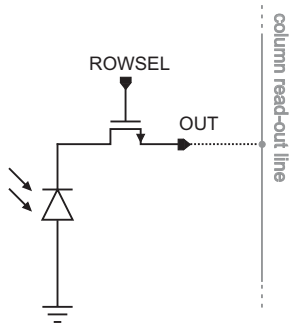


Figure 2.1: Circuit diagram of a passive pixel sensor.

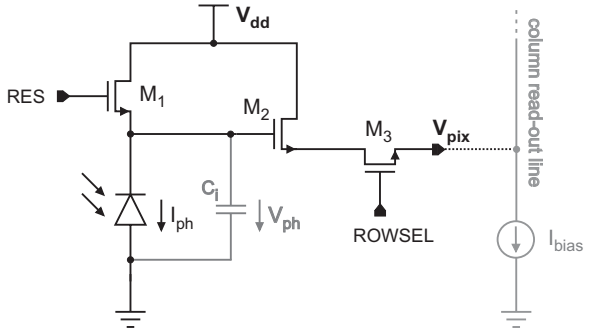


Figure 2.2: Circuit diagram of a standard 3 transistor active pixel sensor.

presented in this chapter are active pixels sensors.

The circuit shown in figure 2.2 represents a standard pixel circuit, which is used in many integration based sensors. The capacitance C_i is used for the integration of the photocurrent. It consists of the capacitance of the photodiode and the parasitic capacitances connected to this node. The operation of the circuit is as follows. First, the signal RES is set to high to reset the integration capacitance. After RES is set to low again, C_i starts to be discharged by the photocurrent I_{ph}. The resulting decrease of the output voltage V_{pix} is proportional to the product of photocurrent and integration time T_{int}:

$$V_{dd} - V_{pix} \propto \frac{I_{ph} \cdot T_{int}}{C_i} = V_{dd} - V_{ph} \tag{2.1}$$

where V_{ph} is the voltage at the capacitor. The integration time T_{int} is defined as the difference in time between the pixel reset and the selection for read-out by use of ROWSEL.

Double-Sampling

Mismatch of the pixel transistors, especially of M_2, will cause a certain amount of fixed pattern noise in the final image. In integration based imagers, the standard method for the removal of FPN is double-sampling. It is performed by the subtraction of two values from each other. The first value is the pixel output after the integration time T_{int}. The second is the pixel output directly after the following reset. Since both values are affected by the same offset, the result of the subtraction is offset-free. This method is also used for the sensor build within this thesis. The relevant circuits are explained in section 4.4.

In literature, the respective method used to remove the FPN is often referred to as ”correlated double-sampling” (CDS) although ordinary double-sampling is performed. Strictly spoken, the term ”correlated” means, that the reset value is sampled prior to the signal value while both values belong to the same frame. In this case, the initial signal value at the start of integration is identical with the reset value. This is important due to the existence of ”reset noise”: Thermal noise in the reset transistor leads to varying reset levels. The resulting error voltage at the photodiode is dependent on the temperature T and the capacitance C_i. It can be calculated by [FRE00]:

$$V_{kTC} = \sqrt{\frac{kT}{C_i}} \tag{2.2}$$

where k is Bolzmann’s constant. If no true CDS is performed, the read-noise will be limited by the reset noise. True CDS can only be performed by the introduction of an additional read-out node in the pixel.

2.2 Integration Time Control

2.2.1 Global Exposure Control

Global exposure control describes the adjustment of the same integration time for all pixels of an imager. A possible pixel for such an imager is the standard APS[1] depicted in figure 2.2. High dynamic range can be reached by a multiple read-out of the sensor array after different integration times. The captured images are stored in a digital off-chip frame memory. After the last frame is captured, the recorded frames are fused into a single high dynamic range image by collecting the non-saturated pixels with the longest integration time (for a high SNR). For each pixel, the corresponding integration time (frame indicator) has to be stored separately.

Depending on the complexity of the system used, a saturation detection circuit is employed to store only non-saturated pixel values and the corresponding integration time during the capturing of the individual frames. Thereby, the required amount of memory is reduced and the "fusion" of the frames is accomplished during the read-out. Examples for high dynamic range sensors using this approach can be found in [SCH02] [SCH00] [SAS04].

Advantages of the global exposure control are a small pixel size (3 transistor APS) and a dynamic range that is easily adaptable to the requirements of the application. The drawbacks are an increased system size and power consumption. Additionally, the required multiple read-out of the sensor leads to the need of high read-out speeds and a reduced frame rate (see section 3.1.2)

Yang et al. proposed a concept to overcome this speed problem [YAN99-2]. Globally controlled pixel-level ADC[2]s are used to perform an in-pixel analog-to-digital conversion of the signal for different integration times. The resulting digital data can be read out at fast SRAM[3] speeds. One ADC is shared by 4 pixels. This way, 22 transistors for four pixels were needed. Fabricated in a 0.35 μm CMOS process a pixel size of 10.5 μm $\times$10.5 μm was reported. However, the shared ADC leads to a reduction of the maximum possible integration time to $T_{max} = 0.25/f_R$, where f_R is the required frame rate of the imager. Additionally, the minimum possible integration time of the implementation is approximately 1 ms, which results in a limitation of the achievable dynamic range expansion.

2.2.2 Local Exposure Control

In-Pixel Exposure Control

One way to achieve local exposure control is the integration of a regulating circuitry in each pixel. A couple of implementations have been published so far that follow this approach. One of the first was a photoreceptor proposed by Miyagawa et al. [MIY95]. It uses multiple integration periods, which are chosen in dependence of the level of incident light intensity to avoid saturation. When the pixels run the risk to saturate, the integration is stopped and the integration time as well as the integration value are stored within the pixel.

The circuit diagram of the pixel used is depicted in figure 2.3. Two photodiodes are integrated in a single pixel. The first photodiode (*PDa*) is used for the detection of saturation while the second (*PDb*) is used for the storage of photogenerated charge. *PDa* is connected to an inverter, which thresholds the output voltage of this diode. The output of the inverter is latched to control the gate of the transistor M_3, which connects the second photodiode to the storage capacitor C_1. C_1 is used to integrate the signal charge. Thus, the output of the latch controls the integration periods. A second

[1] Active Pixel Sensor
[2] Analog-to-Digital Converter
[3] Static Random Access Memory

capacitor (C_2) stores a voltage level that indicates the corresponding integration time. Like the first capacitor, it is connected to a transistor (M_4) whose gate is controlled by the latch.

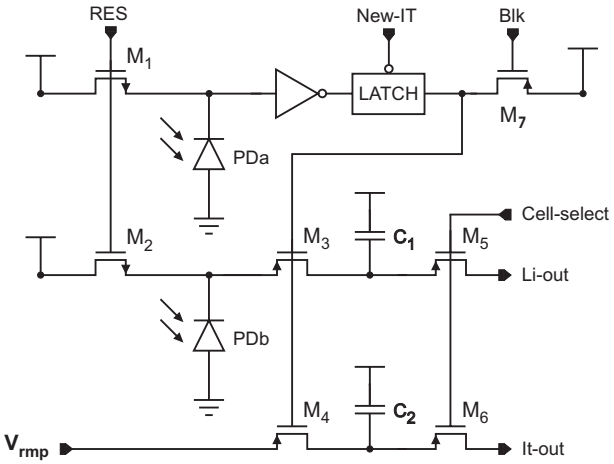


Figure 2.3: Circuit diagram of a photoreceptor with in-pixel local exposure control (cf. [MIY95]).

The operation of the circuit is as follows. First, both photodiodes are reset to a high voltage by use of the signal RES. When the reset is finished, the photocurrent leads to a decrease of the voltages at the two photodiodes. During the integration period, the voltage V_{rmp} is increased step-by-step for each new integration time adjustment. The time between two increases of V_{rmp} doubles from step to step. At the start of each new step the signal New-It (new integration time) allows the latch to change its state for a short time. Hence, the voltage indicating the finally applied integration time is stored on C_2 in parallel to the pixel value itself. At the end of the integration, a row of pixels is selected by the signal Cell-select and the two output values Li-out (light intensity output) and It-out (integration time output) are read out.

Owing to the complex in-pixel circuitry, the pixel size is rather large. The implementation of the photoreceptor shown in figure 2.3 required an area of 109 μm $\times$110 μm in a 2 μm process.

In a later implementation of an in-pixel exposure control by Lulé et al. [LUL00] only a single photodiode is employed. Similar to the concept described above, a ramped voltage is used for an in-pixel storage of a time stamp indicating the integration time used. A total dynamic range of 120 dB was reported with a pixel consisting of 17 transistors and 2 capacitors. Implemented in a 0.8 μm process the pixel size amounted to 40 μm $\times$ 38.3 μm.

On-Chip local exposure control

A few concepts were published, which try to implement on-chip local exposure control without integrating the complete control circuit in the pixel. Hamamoto et al. [HAM01] published an implementation of a sensor, where only parts of the integration time control are located in the pixel. However, the pixel circuit is also used for motion detection, which results in 17 transistors per pixel. A total dynamic range of 56 dB was reported.

Recently, the implementation of a prototype imager was published [YAD03], which uses an integration time control scheme similar to that of the imager presented in this thesis. The local integration time is adjusted in exponentially decreasing periods and the corresponding time stamp is stored in an on-chip memory. The chip has a resolution of 64 $\times$ 64 pixels. Neither a HDR[4] window nor the

[4]High Dynamic Range

possibility of on-chip averaging of neighboring pixels is available. There are no on-chip circuits for the reduction of FPN. A total dynamic range of 83 dB was reported.

2.2.3 Dual-Sampling

Dual-sampling is a technique to widen the dynamic range of a sensor by the use of two different integration times, a short one and a much longer one. It must not be mistaken with double-sampling, which is a fixed-pattern noise reduction technique. In conjunction with an APS dual-sampling was first used by the Jet Propulsion Laboratory [YAD97].

In conventional integration based CMOS APS sensors operating in normal mode, the rows of the sensor array are selected for read-out one after another. The data of the selected row is read-out onto sample-and-hold stages outside the array, which is done in parallel for all pixels that belong to the corresponding row. After the read-out the row is reset (next integration period is started) and the reset values are also sampled. Before the next row is selected, the sample-and-hold stages are serially read out by the off-chip circuitry. This procedure is repeated for every row. When the end of the array is reached the selection starts with the first row again. Hence, the integration time T_{int} of any pixel is given by the time between two accesses of the same row.

For dual-sampling, a different access sequence is used. The technique requires a second row of sample-and-hold stages, but the standard active pixel sensor can be used without modifications (only 3 transistors per pixel). The basic operation scheme is illustrated in figure 2.4. First, row number n is selected (”back curtain”) and the pixel values are copied onto the capacitors of the sample-and-hold stages on top of the pixel array. The row is reset and the reset values are also copied. Prior to the serial read-out of the sample-and-hold stages, a second row with number $n + \Delta$ is selected (”front curtain”). The contents are copied onto the capacitors of a the sample-and-hold stages at the bottom of the pixel array. Like row n also row $n + \Delta$ is reset and the reset values are sampled. Next, the sample-and-hold stages are both serially read out by the off-chip circuitry. When the read-out is finished, the procedure is repeated for the rows $n + 1$ and $n + \Delta + 1$ and the respective subsequent rows.

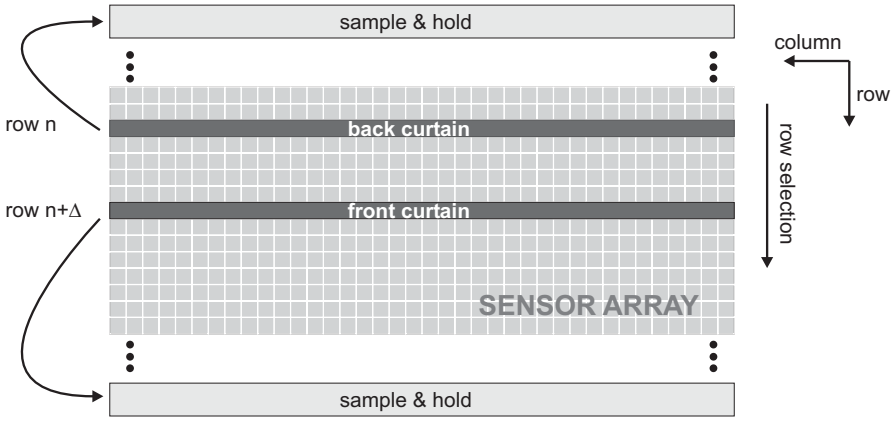


Figure 2.4: Basic operation scheme of the dual-sampling technique (cf. [YAD97]).

The consequence of this read-out scheme is that pixels which are read out in the back curtain always have a much shorter integration time than those in the front curtain. The time between reset

and read-out for a back curtain row is proportional to Δ, i.e. the number of rows between front and back curtain. For the front curtain row this time is proportional to $(N_{row} - \Delta)$, where N_{row} is the number of rows of the sensor.

In essence, the dynamic range is expanded by the factor $(N_{row} - \Delta)/\Delta$. For the 64×64 pixels prototype sensor of the Jet Propulsion Laboratory a total dynamic range of 109 dB was reported.

It should be noted here that the read-out of the array ends up with two frames recorded at different integration times. These have to be combined externally to get the final image with an extended dynamic range. Therefore, in spite of the different read-out scheme the result is the same as for a global control of integration time. However, the major drawback of dual-sampling is the use of only two integration times. Illumination levels that do not fit to one of these integration times will be recorded at a strongly reduced SNR [YAN99-1].

2.3 Logarithmic Compression

In sensors with logarithmic response, the dynamic range is expanded by the logarithmic encoding of the photocurrent, which is reached by the use of a transistor operating in subthreshold region. Figure 2.5 shows the basic architecture of a simple logarithmic sensor. The transistors M_2 and M_3 serve the same purpose as in the standard APS (signal amplification and row selection). M_1 replaces the reset transistor in the standard APS. The pixel shown does not require a reset since it operates continuously, i.e. the output signal will represent the actual illumination as long as the input frequency lies within the circuits bandwidth. The key elements of logarithmic sensors are the transistor M_1 and the photodiode: The photocurrent I_{ph} that is generated in the photodiode flows through M_1. Owing to the fact, that in general I_{ph} is very low[5], M_1 operates in subthreshold region. According to equation 1.20 the gate-source voltage of a transistor in weak inversion shows a logarithmic dependence on the drain-current ($I_D = I_{ph}$):

$$V_{GS} = V_T + 2nV_{th} + nV_{th} \cdot \ln\left(\frac{I_{ph}}{2\beta(nV_{th})^2}\right) \tag{2.3}$$

with identical definitions as in equation 1.20. Owing to the fact that the gate voltage of M_1 is fixed to V_{dd}, the source potential (V_{out}) shows the described logarithmic dependence on I_{ph}. Equation 2.3 is valid for a variation of I_{ph} over several decades, which enables the realization of more than six decades of dynamic range.

The simple circuit shown in figure 2.5 is advantageous with respect to pixels size and signal compression. Nevertheless, it has some severe disadvantages. First, the output shows a significant temperature dependence. Second, the circuit becomes slow at low light intensities, since the low photocurrent has to charge the capacity at the output node (photodiode and parasitic capacitances). The response speed can be enhanced by the inclusion of a feedback amplifier [MOI00-2]. Third, the compression of the signal leads to a reduction of image contrast. However, the major disadvantage of the circuit is an extremely high fixed pattern noise, which is caused by the high mismatch sensitivity of M_1 due to its operation in weak inversion. As described in section 1.4.1, this mismatch is caused by an increased sensitivity for threshold voltage variations.

The suggested solutions to overcome the problem of high FPN can be divided into digital off-chip correction and analog on-chip calibration methods.

[5] The photocurrent can be calculated by equation 1.6. For an assumed quantum efficiency of $\eta = 0.5$ and a wavelength of $\lambda = 600\,\text{nm}$ the photocurrent for a $5 \times 5\,\mu m^2$ diode that is illuminated with an intensity of $J_0 = 1\,\text{W/m}^2$ amounts to 6 pA. This is 3-6 decades below the typical operational currents of MOS transistors.

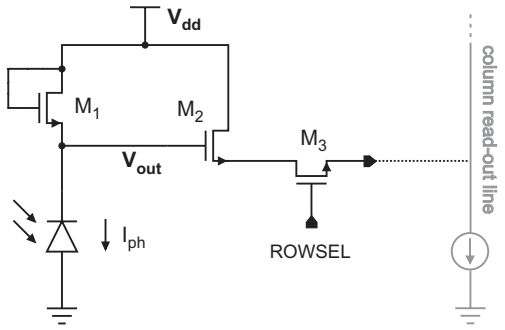


Figure 2.5: Basic architecture of a photoreceptor with logarithmic response.

2.3.1 Digital Calibration

Digital correction is realized by a subtraction of the fixed pattern noise during the read-out of the pixels. In the calibration procedure, the entire sensor array is homogeneously illuminated and the measured pixel signals (calibration pattern) are stored in a digital off-chip memory. During the ordinary operation of the chip the acquired pixel signals are digitized and the calibration pattern is used to remove the individual signal offsets. In principle, the calibration is not restricted to offset correction. With higher efforts for the correction, also the compensation of slope variations or higher order variations is possible. An example for a solution using off-chip digital correction can be found in [IMS00].

However, the digital correction has some drawbacks. First, the off-chip correction leads to a larger overall system size and to an increased power consumption due to the need of additional devices. In principle, the digital correction could be integrated on chip, but this requires the on-chip integration of large memories for the storage of the calibration patterns. This in turn would result in higher production costs due to larger chip sizes and a reduced yield. Second, the digital correction method does not account for variations in the FPN caused by aging or temperature effects. On-chip analog calibration methods try to overcome these disadvantages.

2.3.2 Analog Calibration

In order to calibrate a sensor system, it is necessary to stimulate the system with a reference signal and to measure its response. For off-chip digital calibration this is accomplished by the stimulation of the sensor with a certain light intensity. Analog calibration methods rely on the stimulation of the sensor circuit by an electronically generated reference current instead of the photocurrent generated in the diode. Different methods for the implementation of analog calibration were suggested. In the following two examples are presented and compared.

Calibration by Use of an In-Pixel Memory

Loose et al. [LOO01] presented an analog calibration method that relies on an in-pixel storage of the calibrated pixel state. In figure 2.6 the circuit diagram of the implemented self-calibrating photoreceptor is shown. The lines I_{ref}, V_{out1}, V_{out2}, and V_{corr} are common to all pixels lying on the same column, whereas the control signals CS (calibration select), $\overline{\text{CS}}$, and $\overline{\text{RS}}$ (read select) are common to all pixels lying on the same row. In order to reach a higher signal slope, two transistors in series (M_1 and M_2) are used to accomplish the logarithmic compression. Both work in subthreshold region. The transistor M_3 operates as a source follower and buffers the signal V_{log}. The calibration amplifier and

the current sources shown on the right side are located outside the pixel array. They are implemented once for each column. The circuit has two modes of operation:

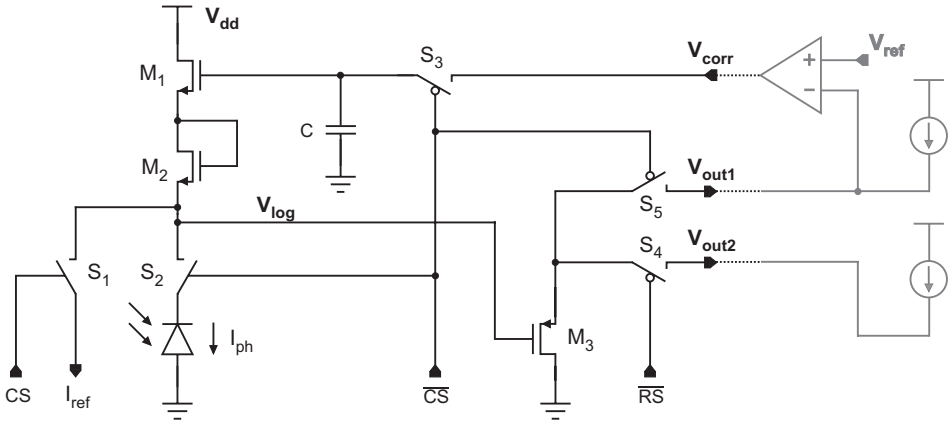


Figure 2.6: Circuit diagram of a self-calibrating photoreceptor with logarithmic response (cf. [LOO01]).

In normal operation mode, the signal CS is set low ($\overline{\text{CS}}$ is set high). Consequently, the switch S_2 is closed, whereas the switches S_1, S_3 and S_5 are open. The photocurrent I_{ph} is then converted into the logarithmic voltage V_{log}. Finally, the pixel can be read out by setting $\overline{\text{RS}}$ to low, which closes S_4 and connects M_3 to the read-out line V_{out2}.

The calibration mode is entered by setting CS to high ($\overline{\text{CS}}$ is set low). Now, the switches S_1, S_3, and S_5 are closed, whereas S_2 is open. As a result, the sensor circuit is no longer stimulated by the photocurrent but by the reference current I_{ref}. Via the closed switch S_5, the signal voltage V_{log} is led to the input of the calibration amplifier, where it is compared with the voltage V_{ref}. The amplifier generates a correction voltage V_{corr}, which in turn leads to a change of V_{log} (feedback loop). The calibration cycle is finished by the return to the normal operation mode (CS set to low). Switch S_3 is opened and V_{corr} is stored on the in-pixel capacitor C.

Charge injection caused by the switch-off of S_3 can result in an additional FPN. Therefore, a dummy transistor for the compensation of the injected charge is required for this switch (cf. section 4.4.2). The remaining switches are implemented as single transistor switches. In total, 9 transistors and 1 capacitor are needed for the realization of one pixel.

It can be shown that after the calibration, the output voltage V_{out2} no longer depends on mismatch of the transistors M_1, M_2, and M_3. Therefore the FPN is reduced to a large extend. At an intensity of 1 W/m^2 the remaining FPN amounts to 3.8 % of a decade (rms), which corresponds to 0.63 % (rms) of the achieved total dynamic range (120 dB).

Correction by Double-Sampling

A different calibration scheme was introduced by Kavadias et al. [KAV00]. The method is similar to the offset correction in integration based sensors since it also relies on a double-sampling technique. The first sampled value is the signal of the sensor in its normal operation mode. The second pixel value is obtained by the stimulation of the receptor circuit with a known current. The difference between these two values is free from offsets caused by threshold voltage variations.

In figure 2.7 the circuit diagram of the corresponding pixel is shown. The lines CAL and OUT are common to all pixels lying on the same column. Logarithmic compression is accomplished by M_1 operating in weak inversion. Transistor M_2 operates as a source follower and M_3 is used to connect

the signal to the read-out line OUT. M_4 enables the calibration and M_5 connects the sense node to the gate of M_2.

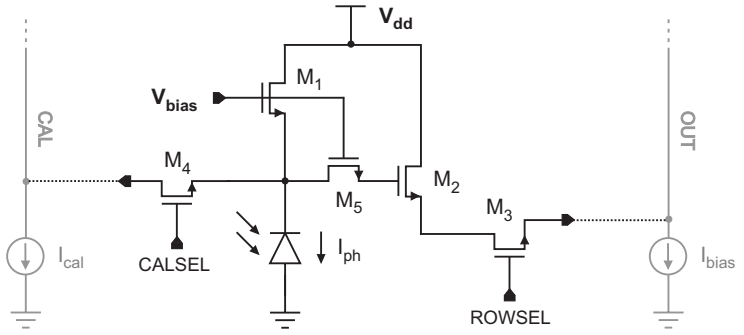


Figure 2.7: Circuit diagram of a photoreceptor with logarithmic response that can be calibrated by a double-sampling technique (cf. [KAV00]).

The operation of the circuit is as follows. First, the pixel signal is read out by setting ROWSEL to high, while keeping CALSEL low. The output of the pixel is sampled and stored in an amplifier connected to the respective column. Next, CALSEL is set to high with the consequence that an additional current I_{cal} flows through the transistor M_1. I_{cal} is much higher than the photocurrent so that M_1 operates in strong inversion. The resulting output voltage of the pixel is sampled again and subsequently subtracted from the value sampled first. It can be shown that the resulting value is independent from threshold voltage variations within the pixel. However, offsets arising from variations in other parameters are still present. By the use of this circuit a dynamic range of 120 dB was reached. A remaining FPN of 2.5 % of the total signal swing was reported.

In summary, the pixel in figure 2.7 is smaller than for the analog calibration method that was previously presented (5 transistors or 9 transistors and a capacitor respectively), but at the cost of a significantly higher remaining FPN (2.5 % of the total signal swing or 0.63 % respectively).

2.4 Well Capacity Adjusting

The adjustment of the well capacity is a dynamic range enhancement technique described by Sayag [SAY90] and implemented by Decker et al. [DEC98]. The scheme is based on the clipping of the sensor signal in bright regions of the image.

In figure 2.8 the circuit diagram of the pixel used is shown. Essentially, it is identical to a standard APS pixel design. The only difference consists in the introduction of the transistor M_1, the so called "charge spill gate". M_1 is not required for the execution of the wide dynamic range algorithm. It is used to increase the sensitivity of the pixel by employing the small (parasitic) capacitance C_S at its source diffusion for the sensing of the charges generated in the large photodiode. V_{spill} is held at a constant potential so that charge generated in the photodiode can flow into the sense node. The dynamic range enhancement is reached by a gradual decrease of the lateral overflow gate voltage $V_B(t)$ during the integration of the photogenerated charge.

Figure 2.9 illustrates the sequence for the adjustment of $V_B(t)$. A reduction of the lateral overflow gate voltage corresponds to an increase of the potential barrier height that separates the sense node from the supply voltage node. In the first step (a), the sense node is reset to a voltage close to V_{dd} by pulling $V_B(t)$ high. It should be noted here, that the photodiode itself is never reset. In the next step (b), the barrier is abruptly raised a small amount. The photogenerated charge now begins

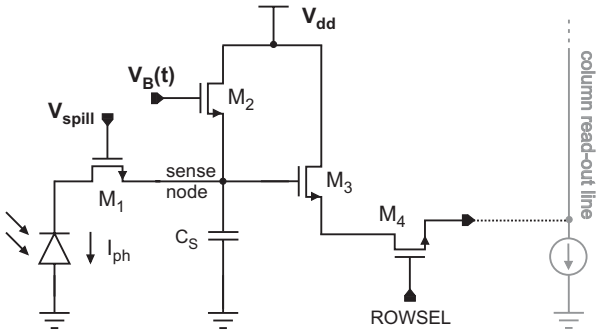


Figure 2.8: Circuit diagram of a pixel used in the well capacity adjusting scheme.

to accumulate on the sense node at a rate that is proportional to the incident light intensity. The light intensity is assumed to be constant during the period of integration. During the accumulation, the barrier rises at a rate that increases with time. Hence, the barrier height as a function of time determines the resultant compression of the signal. In picture (b) it is assumed that at the start of integration the charge accumulates faster than the barrier height is increasing (high light intensity). If the light intensity is high enough, there will be a period of time, where the charge accumulation is limited by the barrier (picture (c)) and the excess charge flows into V_{dd}. If the barrier rises fast enough (rising rate is increasing), after a certain time it will get ahead and lead to free charge accumulation again (picture (d)) until the integration stops and the sense node is read out.

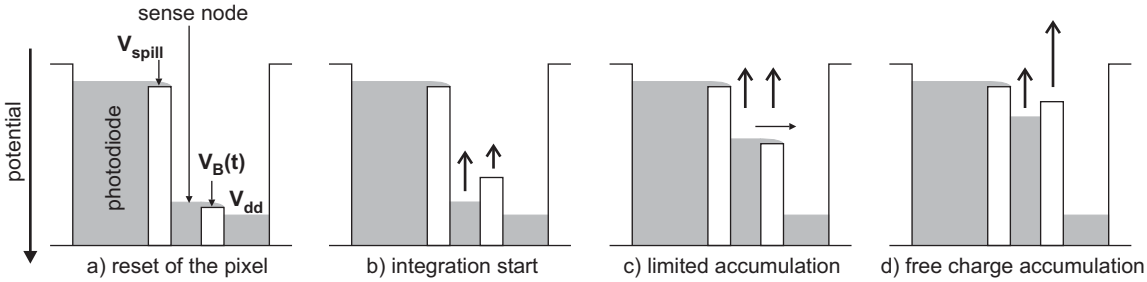


Figure 2.9: Sequence for the adjustment of the lateral overflow gate voltage $V_B(t)$. The length of the arrows in pictures b)-d) indicate the speed of the change of the corresponding potential (cf. [DEC98]).

$V_B(t)$ is monotonically decreased as an exponential like function of time. Consequently, the transfer characteristic of input intensity to output voltage is non-linear. According to the described sequence, the signal compression is a result of the fact that the charge accumulation is slowed down for high light intensities by the drain off of excess charge. Effectively, the adjustment of the barrier is an adjustment of the sense node capacity.

The technique was implemented by Decker et al. in an imager with a resolution of 256×256 pixels. The compression curve was realized by an decrease of $V_B(t)$ in eight steps. At a frame rate of 30 Hz, a total dynamic range of 96 dB was reached. The scheme permits the use of correlated double-sampling, therefore FPN can be effectively reduced.

There are two drawbacks of the well capacity adjusting scheme. First, as for sensors with logarithmic response, the near-logarithmic compression of the signal that is used here leads to a reduction of image contrast and consequently of image quality. Second, the step-by-step adjustment of the well capacity can result in strong dips in SNR, just as for the dual-sampling scheme [YAN99-1].

2.5 Pulse Frequency Modulation

Yang [YAN94] presented a method to widen the dynamic range by use of an oscillating pixel sensor. The analog output of the photodiode is converted into a pulse train with a frequency proportional to the incident light intensity.

Figure 2.10 shows the basic architecture of a pulse frequency modulating pixel. The photodiode acts as a variable current source controlled by the input light intensity. It is charged through the reset transistor M_1. The gate of the reset transistor is controlled by the output of an inverter chain, which is connected to the sense node of the photodiode.

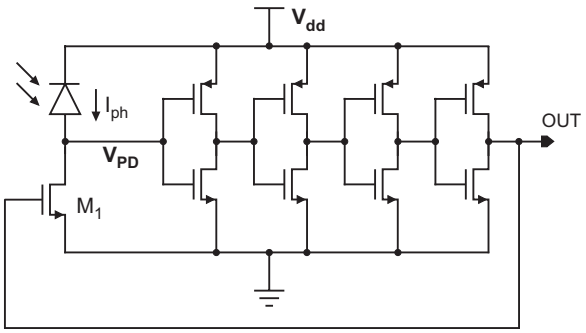


Figure 2.10: Basic architecture of a pixel using pulse frequency modulation (cf. [YAN94]).

The operation sequence of the circuit is as follows. Suppose the voltage at the photodiode V_{PD} is initially smaller than the threshold of the first inverter. Consequently, the outputs of the following four inverters expressed in logic levels are high, low, high, and low. Thus, M_1 is switched off. As light illuminates the photodiode, V_{PD} increases with time until it passes the threshold of the first inverter, which results in an inversion of the logic output levels. M_1 is switched on and resets the photodiode. The reset causes a decrease of V_{PD}. The reset continues until M_1 is switched off again due to the decreased voltage V_{PD} and a new integration cycle is started. The repetition of this sequence causes the generation of a pulse train at the output of the pixel. The frequency f_{out} of the oscillation is given by

$$f_{out} = \frac{I_{ph}}{C \cdot \Delta V} \tag{2.4}$$

where C is the integrating node capacitance and ΔV the difference between the reset voltage level and the threshold of the inverter. According to the variation of the input light intensity, the photocurrent I_{ph} and consequently the frequency f_{out} varies over several decades. Yang demonstrated an 32×32 pixel array with a dynamic range of five orders of magnitude. Drawbacks of the concept are a strongly increased pixel size and a FPN which results from the threshold voltage mismatch between different transistors.

The concept was used in a number of other implementations [ILL01] [CUL01] [OHT01]. McIlrath demonstrated an imager using pulse frequency modulation in conjunction with pixel-parallel analog-to-digital conversion [ILL01]. A dynamic range of 104 dB was reported. However, 19 transistors are required for each pixel resulting in a pixel size of $30\,\mu\text{m} \times 30\,\mu\text{m}$ in a $0.5\,\mu$m process.

Chapter 3

Adaptive Integration Time Control

This chapter introduces the concept of adaptive integration time control. In the first part, critical properties of HDR imagers are identified that led to the implementation of the proposed concept. Next, the operation of the integration time control is described independently of a specific sensor implementation. Different control schemes are discussed with regard to dynamic range enhancement and required memory. In the second part, the need for a well defined chronological order for the access of individual sensor rows is brought into focus. The row sequence is generated by a programmable logic array, which is part of the electronic test system. In the last section, the programming of this array is explained.

3.1 Critical Properties of HDR imagers

To judge the performance and quality of an imager, several parameters should be taken into account. For the acquisition of real world images, the dynamic range is a very important parameter, however depending on the application others may be of equal importance. The concepts presented in the previous chapter try to enhance the dynamic range, but this is often done at the cost of other parameters such as pixel size or overall image quality. The following subsections will focus on those properties of particular interest for the comparison of HDR concepts.

3.1.1 Image Quality

Often high dynamic range is regarded synonymous to high image quality, i.e. a sensor with high dynamic range is expected to produce images of higher quality than a low dynamic range imager. This assumption is justified for systems operating in integration mode with a fixed integration time, where the SNR increases monotonically with the signal. This is not necessarily true when dynamic range enhancement schemes are employed. Depending on the used scheme the SNR can be influenced in various ways.

Continuously working photoreceptors with logarithmic compression represent one branch of investigated HDR concepts. The key element of these sensors is a diode-connected MOS transistor that is operated in the subthreshold region. As a result of the subthreshold operation, there is an expo-

nential relation between the drain current and the gate-source voltage of this transistor, which finally leads to the desired logarithmic compression. An undesirable consequence of this relation is a high mismatch sensitivity especially for variations of V_T, which results in a strong fixed pattern noise. Off-chip as well as on-chip calibration methods have been established to cope with the resulting FPN. Off-chip solutions that accomplish a digital calibration by the use of a frame memory for the storage of the individual pixel offsets effectively reduce the FPN. Nevertheless, they require a post-processing of the recorded data and external facilities [IMS00] [MAR98]. On-chip calibration methods on the other hand rely on additional in-pixel structures. Depending on the complexity of the pixel circuit a stronger reduction of the FPN can be reached at the cost of an increased pixel size [LOO01] [KAV00]. The remaining FPN is still a few percent of a decade. However, a further consequence of the logarithmic compression curve is the decrease of image contrast, which is disadvantageous for many image-processing applications.

Integration based concepts on the other hand offer a higher SNR at a lower effort than those using a logarithmic response. There are different reasons for this. First, the impacts of device-to-device mismatch are much lower than for logarithmic sensors[1]. Additionally, the FPN resulting from the mismatch can easily be reduced on-chip by the proven method of double-sampling. This method does not lead to an excessive increase of the pixel size but it cannot be used in continuously working sensors since there is no reset level that could be subtracted. Furthermore, the integration of the photocurrent over a certain period of time basically acts as a low-pass filter, which removes high frequency components of the noise [MOI00-2] leading to a decreased temporal noise.

However, even the different concepts based on integration differ significantly in their achievable SNR. The specification of a single figure for the SNR (in literature, often only the maximum SNR is given) is not sufficient for the characterization of the performance. The course of the SNR as a function of the incident light intensity can exhibit strong dips resulting in a decreased image quality within certain ranges of light intensity. A comprehensive comparison of different integration based sensor concepts (well capacity adjusting, dual and multiple sampling) with respect to achievable SNR can be found in [YAN99-1]. In essence, it is shown that using the well adjusting scheme SNR degrades as DR is increased, whereas with multiple sampling the DR can be expanded without noticeably degrading SNR. For instance, the sharp decline of SNR observable in a single well capacity adjustment amounts up to -24 dB[2] with the assumed realistic conditions[3]. This is even stronger than for the dual sampling scheme under the same conditions. Multiple sampling on the other hand, can be used to reach a consistently high SNR. Here, for a dynamic range expansion by a factor of 256 the maximum decline of SNR amounts to only -7 dB. These dips in SNR occur at those light intensities where the integration time has to be changed to prevent saturation. In order to take advantage of the achievable SNR, the finally implemented HDR concept is based on an image acquisition using several different integration times.

3.1.2 Availability and Post-processing of Image Data

Ideally, the sensor output reflects changes in the light stimulus without latency (infinite bandwidth) so that the image data is available at any time. Furthermore, it should be possible that the data coming from the imager can directly be used by the following application without post-processing.

[1] In (uncalibrated) sensors with logarithmic response the peak-peak variation in the output signal often corresponds to 2-3 decades of light intensity [LOO99-2] [RIC95].

[2] Assuming a dynamic range enhancement factor of 32.

[3] Assumed conditions: max. integration time T_{max}=30 ms, amount of charge storable by the sensor $q_{max} = 1.23 \cdot 10^5\, e^-$, noise due to read-out circuits $\sigma_r = 20\, e^-$, dark current I_d=1 fA.

Unfortunately, this is rarely the case. Regarding on the used concept for the expansion of the dynamic range, different efforts have to be made for reconstruction and preparation of the image data for the intended application. These efforts may result in an increased overall power consumption, system size, and cost. Additionally, depending on the concept the availability of the image data is affected to a different extend.

With regard to some of these properties, sensors with logarithmic compression are advantageous for some applications. Without the necessity for post-processing, a single frame of data coming from the chip represents an image with the full recordable dynamic range. Additionally, the logarithmic response corresponds to the behaviour of the human visual system, which can be useful. An example for this is the representation of the image on a display device (such as CRT[4]/LCD[5] monitors). Thanks to the logarithmic response the recorded luminances always lie within the displayable range of the device, additional compression is not necessary. However, in other applications the nonlinear output can lead to the necessity of a complicated subsequent signal processing (e.g. for color processing).

The accessibility of the image data from logarithmic sensors is claimed to be good, since the concept relies on a continuously working sensor, i.e. the results can be read out anytime. In practice, however, the potential of scanning virtually anytime is not fully usable. The speed of adaptation is strongly dependent on the incident light intensity, i.e. for low intensities the pixel output cannot follow rapid changes of the light stimulus immediately. For typical room light levels roughly 1 ms is required for the adaptation [YAD99]. Anyway, this is an acceptable value.

In comparison with logarithmic sensors, the concept of reading out the full sensor array after the expiration of different integration times leads to several disadvantages. For the reconstruction of the original image, an off-chip memory has to be provided for the storage of the images taken at a particular integration time. Depending on the complexity of the implemented scheme the required amount of memory can be reduced (e.g. by exclusively storing unsaturated values). Nevertheless, the minimum amount that has to be provided is the number of pixels times the sampling resolution per pixel value. The second disadvantage arises from the time required for the acquisition of a complete HDR image. Assuming an exponential global regulation of the maximum integration time T_{max} in steps of $1/2^k$ (k=0,1,2,...k$_{max}$) a total acquisition time of

$$T_{acq} = T_{max} \cdot (1/2^0 + 1/2^1 + ... + 1/2^{k_{max}}) \tag{3.1}$$

is required. Neglecting the read-out times, this results in a minimum acquisition time of approximately 2 T_{max} for multiple sampling (high k) of complete frames. Therefore, a maximum frame rate of only $1/(2\ T_{max})$ can be reached. In addition the complete sensor array has to be read out k$_{max}$ times at full precision (long sampling times). Hence, either an effort has to be made to reach a higher read-out speed or the acquisition time T_{acq} has to be further enlarged.

Section 3.2 will show that multiple sampling can be used to realize an HDR concept, which does not suffer from these drawbacks. By the implementation of an on-chip local adaptive integration time control it is possible to achieve high dynamic range without the necessity of external image reconstruction. The array has only to be read out once at full precision and a frame rate of $1/T_{max}$ is reached. The final read-out is still row-by-row as required by many applications.

3.1.3 Pixel Size

A matter which is always related to the chosen HDR concept is that of pixel size. As described in the previous chapter, the different concepts result in various pixel circuits and different pixel

[4]Cathode Ray Tube
[5]Liquid Crystal Display

dimensions. In general, a pixel should be as small as possible, mainly for two reasons: First, to minimize the cost for the implementation of a given resolution. Second, to minimize the area that has to be illuminated by the mounted lens. This results in a smaller lens, which is especially important for mobile applications[6]. Unfortunately the reduction of the pixel size is reached at the expense of a reduced sensitivity due to the smaller photosensitive area.

This area is not identical with the total pixel area since only that part can be used for the integration of the photodiode, which is not occupied by the remaining circuit. In most designs this is about 30-60 % of the total pixel area (referred to as the "fill factor").

In summary it may be stated that the demand for small pixel size in conjunction with the fact that the photodiode and the remaining in-pixel devices compete for the same area leads to the requirement of a small pixel circuit to reach a certain sensitivity. The most simple circuit that is applied in CMOS HDR imagers is the three transistor APS. As described in the previous chapter, it is used in many integration based sensors as well as in receptors with logarithmic response. Chapter 4 will show that the concept proposed in the next section can be realized by adding a single extra transistor to the three transistor APS (neglecting extra devices for averaging).

3.2 Concept of Adaptive Integration Time Control

The proposed concept for the expansion of dynamic range is based on the integration of photocurrent generated by the incident light at the respective pixel. In order to prevent pixels saturation due to high illumination and to achieve an acceptable SNR for those lying in darker regions, different integration times are used for different pixels. Instead of adjusting a global integration time which is the same for all pixels and reading out the entire array after each adjustment, the integration time is individually adapted for each pixel in dependence of the predominant light intensity at the specific pixel location. A concise description of the integration time regulation scheme can be found in [BRE04].

The adaptation of the integration time is possible on the basis of a periodic check of the pixel output voltages. If individual pixels run the risk of getting saturated during integration, a reset of those pixels is performed. The new integration time, i.e. the remaining time until read-out, is successively reduced during subsequent checks until it is guaranteed that the pixel will not be saturated until read-out.

Apart from the pixel output value (integration result) the only information needed to reconstruct the value of the recorded illumination is the duration of the integration time, i.e. the time since the last reset. Therefore, this value has to be stored individually for each pixel in the course of regulation. A tradeoff has to be found for the number of different integration times, which are used in the regulation. On the one hand, the more different integration times are used, the higher the demand on memory and the effort for regulation. On the other hand, with a high number of integration times finer regulation and/or higher dynamic range expansion is reached. For the implemented control scheme, the integration time T_{int} is regulated exponentially by a factor of $1/2^z$:

$$T_{int} = \frac{T_{max}}{2^z} \quad z \in \{0, 1, 2...z_{max}\} \tag{3.2}$$

where T_{max} represents the maximum possible integration time and z is an exponent, which is varied to accomplish the regulation. In the following z will referred to as the "time stamp". Only the time

[6]The reasonable limit for a reduction of the pixel size is given by the diffraction limit of the lens system used. For current lens systems found in consumer cameras this amounts to a few micrometers. The pixel size of currently available sensors for consumer digital still cameras amounts to a minimum of 2.35 μm (non-HDR CCD sensor) [SON04].

stamp has to be stored to reconstruct the light intensity J at the respective pixel location from the measured pixel output voltage V_{pix}:

$$J = A \cdot V_{pix} \cdot 2^z \tag{3.3}$$

where the factor A is a constant which is identical for all pixels. A depends on the properties of the photodiode (size of photosensitive area, quantum efficiency), the pixel circuit (integration capacitance, source-follower gain), and the maximum integration time T_{max}. In chapter 5 the calibrated response of the sensor to absolute values of light intensity is discussed.

An important advantage of the regulation in steps of 2^z is the simplicity of image data reconstruction. The sampled analog value V_{pix} must simply be multiplied by 2^z, which is equivalent to a bit-shift of the digitized value by the amount of digits given by the time stamp. This can be easily integrated on chip.

In figure 3.1 the course of the integration time regulation is illustrated by the example of two pixels that are located in the same row but illuminated by different light intensities. In the upper diagram the output voltages $V_{pix1/2}$ of the two pixels are shown. At the beginning ($t = 0$), both pixels are reset by charging an in-pixel integration capacitance to the reset level. Assuming a constant light intensity during the integration time, a constant photocurrent will discharge this capacitance leading to a linear decrease of the pixel output voltage[7]. Depending on the light level at the respective pixel location, the capacitance will be discharged at different speeds. Pixel no.2 is illuminated with a higher intensity leading to a steeper decrease of its output voltage.

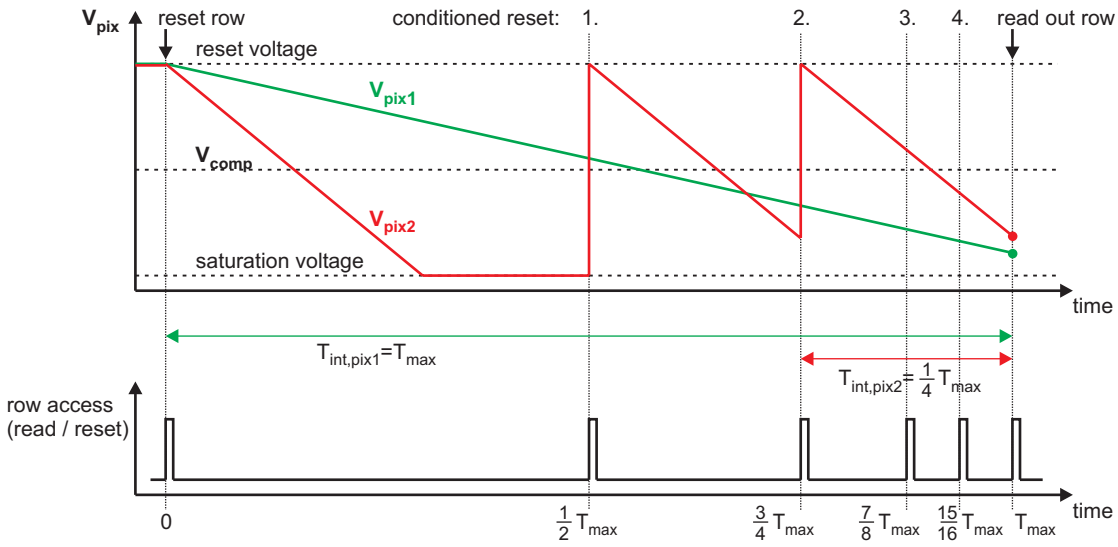


Figure 3.1: Course of the integration time regulation for two pixels within the same row but illuminated with different light intensities. For the sake of simplicity, the regulation is limited to four conditioned resets of individual pixels here. In the upper part the pixel output voltages are shown. The lower diagram indicates the points in time when the row is accessed for read-out or/and reset.

In the lower diagram of figure 3.1 the points in time are indicated when the row containing two considered pixels is accessed for read-out or/and reset. The illustration is limited to five different integration times, for the implemented sensor up to sixteen are possible. As mentioned above, during the first access ($t = 0$) the entire row is reset. In each of the following four accesses ("conditioned resets" z =1...4), all pixels of the selected row are simultaneously checked if some of them run the

[7] In reality, the pixel response will not be exactly linear (cf. section 4.2.1).

risk of being saturated. Those pixels, which fulfill the conditions for a reset, are reset and start a new integration period. Finally, in the last row access ($t = T_{max}$) the entire row is read out (acquisition of V_{pix}). Corresponding to equation 3.2 the conditional resets lead to a gradual reduction of the maximum integration time T_{max} in steps of $1/2^z$ if this is necessary for an individual pixel. For pixel no.1, which is only reset at $t = 0$ (no conditioned reset, $z = 0$), an integration time of $T_{int} = T_{max}$ results, whereas for pixel no.2 it is $T_{int} = T_{max}/4$.

The purpose of the conditioned resets is to allow for the longest possible integration time (high SNR) without resulting in a saturated signal ($V_{pix} <$ saturation voltage) at final read-out. In order to achieve this, the conditions that have to be fulfilled for a reset are as follows:

1. $V_{pix} < V_{comp}$

2. The pixel has been reset in the previous conditioned reset.

V_{comp} is a comparison voltage, which is chosen here as the average between the reset voltage and the minimum acceptable pixel output (saturation voltage). The pixel will only be reset if both conditions are fulfilled at the time when the pixel is checked.

The application of the conditioned resets can be observed during the course of the pixel output voltages: Pixel no.1 is never reset between the initial reset and the final read-out. At the first check for a reset V_{pix1} is still higher than V_{comp} (first condition not fulfilled), whereas at the following checks the second condition is not fulfilled. Pixel no.2 on the other hand, is reset at the first and at the second check. At the third check condition no.1 is not fulfilled anymore, whereas at the fourth check condition no.2 prevents a reset.

The selection of pixels for read-out (and for a reset check) is always row-by-row. Owing to the fact that there is only one read-out line per column of pixels (limitation of space) it is impossible to read-out two rows in parallel. An important consequence of the reset sequence is that all pixels of the selected row are ready for read-out at the same time namely at $t = T_{max}$. This guarantees that the final integration results of the entire sensor array can be read out one after another without the necessity to access any row two or more times due to pixels that are still integrating. This is an important property for applications which demand for a row-by-row input of image data (e.g. video). Furthermore it assures an optimal use of the available read-out speed since a continuous data flow can be reached.

An intuitive approach to the operation of the integration time control within the sensor array is given in figure 3.2. Several selections of rows are shown, distributed over the sensor array with exponentially decreasing distance. The integration time control can be thought as a pattern of ”reset curtains” that is shifted over the array with constant speed. Looking at a particular row, e.g. the row at the bottom of the shown array, the pattern will pass this row resulting in the same chronological sequence as was discussed in figure 3.1. First, the entire row is reset, then the conditioned resets follow with decreasing distance in time and finally the entire row is read out. When the pattern is shifted over the entire array, starting at the first row, each of the rows will be controlled by the same sequence as the one discussed.

In principle, the implementation of the moving selection pattern is sufficient for the realization of the adaptive integration time control. Unfortunately, the determination of the chronological sequence for the selection of the rows is not trivial since rows cannot be read out in parallel. Each row access needs a certain amount of time that will lead to a block up the read-out bus for this time. The determination of the required row sequence will be discussed in section 3.4.

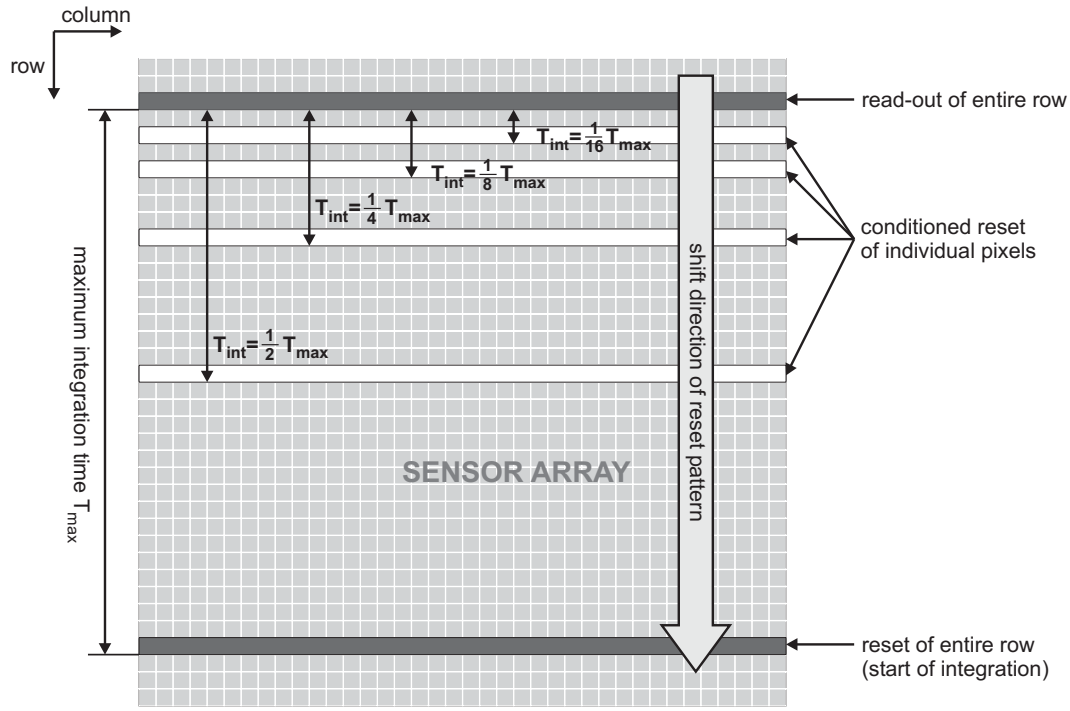


Figure 3.2: Visualization of the integration time control by the assumption of a pattern of reset curtains that is shifted over the sensor array with constant speed. As in figure 3.1 the illustration is limited to four curtains where a conditioned reset is executed.

3.3 Required Memory Size

In order to be able to reconstruct the image data correctly, it is necessary to know the integration time of the individual pixels. As a consequence of equation 3.3 only the time stamp of the last reset has to be stored for this purpose. The more different integration times are adjustable, the larger is the range of possible stamp values and consequently the higher is the amount of memory that has to be provided for the storage of a single stamp.

Depending on the definition of the regulation scheme, the maximum number of stamps z_{max} determines the dynamic range expansion that can be reached. Assuming the regulation defined by equation 3.2 this dynamic range expansion is $2^{z_{max}}$. For the imager build within this thesis, a memory of 4 bit per stamp value is reserved. This results in a maximum stamp value of $z_{max} = 15$ and in consequence in a maximum dynamic range expansion of $2^{15} = 32768 = 90\,\text{dB}$. In most cases this expansion will be sufficient: On the one hand a given maximum value of T_{max} has to be met (e.g. for video frame rate) and on the other hand shorter integration times than $T_{max}/2^{15}$ are mostly not reasonable due to the lack of high light intensities, which would require even shorter integration times. A change of the regulation factor (e.g. from $1/2^z$ to $1/4^z$) in conjunction with an adaptation of V_{comp} would lead to an increase of the dynamic range expansion, but this would be at the price of a lower SNR.

The entire sensor array consists of N_{pix} pixels. In any case, the maximum number of stamps that have to be stored is N_{pix}. In the course of the integration time regulation the content of the reserved memory for each pixel will be repeatedly updated with the current stamp value until the time T_{max} has run down and the final stamp value can be read out. If the stamp size is K bit the

required memory is

$$M_{max} = N_{pix} \cdot K \tag{3.4}$$

On closer inspection, it turns out that this is an overestimation. The reason for this is that only parts of the memory size of a stamp are used during certain periods of time. This can be explained by considering figure 3.2 once again. In the following it will be assumed that the integration scheme covers the entire sensor array (the reason for this will be explained in the next section). Furthermore we assume that the initial reset has passed the entire array and the read-out curtain has reached the first row, similar to the pattern position shown in figure 3.2 but with all rows lying within the pattern. In the following, the memory that is required at this point in time will be calculated.

Looking at the pattern of reset curtains it is obvious that pixels which have been reset the first time but have not passed the first conditioned reset curtain yet, do not need any memory until the arrival of this curtain. In principle, the first conditioned reset curtain does not need to check the second reset condition, since all concerned pixels have been reset previously. Pixels lying between the first and the second conditioned reset curtain require one bit per pixel, which indicates if they were reset in the first curtain or not. Those between the second and the third curtain need two bit per pixel to indicate whether the last reset took place in the 0th, 1st or 2nd curtain. Two bits are also sufficient for the pixels between the third and the fourth curtain but the remaining pixels require additional bits for the indication. To simplify the calculation of the total memory size and in particular of the required management of memory resources, it is assumed that pixels lying beyond the fourth curtain need the full stamp size of K bit per pixel.

The required memory size is now determined by the amount of pixels between the respective curtains. A quarter of the array is between the 1st and the 2nd reset curtain, $\frac{1}{8}$ between the 2nd and the 3rd, $\frac{1}{16}$ between the 3rd and the 4th and $\frac{1}{16}$ between the remaining curtains. For a stamp size of $K = 4\,\text{bit}$ the resulting total amount of memory M_{min} that is used in this snapshot is

$$M_{min} = N_{pix} \cdot (\frac{1}{4} \cdot 1\,\text{bit} + \frac{1}{8} \cdot 2\,\text{bit} + \frac{1}{16} \cdot 2\,\text{bit} + \frac{1}{16} \cdot 4\,\text{bit}) = 0.875\,\text{bit} \cdot N_{pix} \tag{3.5}$$

In principle, this consideration is applicable to any point in time during the regulation. Therefore it represents the minimum amount of memory that is needed to implement a regulation scheme with a stamp size of 4 bit. However, in order to take advantage of the reduced memory size in comparison to M_{max}, a complex access control has to be implemented, which assures the correct selection of the corresponding memory cells. A fixed assignment of a memory cell to a pixel at a certain position, as it is used in the current implementation, is not possible any more here.

There is a second possibility for the reduction of the required memory size. In many applications it is not necessary to realize HDR over the entire sensor array all the time. A HDR region located at the position of the a blinding light spot or at any region of interest could be sufficient. The image sensor that was built within this thesis offers the opportunity of a freely movable HDR window. The window has a size of one quarter of the entire sensor array and therefore also needs only one quarter of the memory size that would be needed for the regulation of the entire array. In addition, it is also possible to expand the window over the entire sensor array by averaging neighbouring pixels. This way, the whole scene can be inspected at a reduced resolution but with high dynamic range to detect regions of interest. As a matter of course, it is also possible to shift the window in four steps over the sensor array to get the full image at the full resolution in high dynamic range. This would be at the price of a reduced frame rate and the necessary effort to compose the final image from the recorded frames.

3.4 Determination of the Row Sequence

The integration time control explained in the previous chapter relies on a multiple selection of each row with an exponentially decreasing distance in time between two successive selections. In the visualization of the scheme by figure 3.2 the individual row accesses are treated as taking place instantaneously and needing no time. As a matter of fact, this is not the case. The read-out of the photoreceptors, the test of the reset conditions, and the execution of a reset require a certain amount of time. Since rows cannot be selected in parallel, a chronological order has to be determined, which indicates the row that has to be accessed next.

There are two different kinds of row accesses depending on the tasks which have to be worked off by the sensor. During the first one the reset conditions are checked for the selected row and those pixels which fulfill both conditions are reset. The stamp that indicates the current reset curtain has to be stored for these pixels. In the following, this type of row access will be named "ITC[8] cycle". Ignoring the details of what happens in the chip (cf. section 4.5), the cycle will need a certain amount of time T_{ITC} for its execution[9]. For a particular row the ITC cycle has to be applied in the chronological order of the conditioned resets shown in figure 3.1.

The second kind of row access is the "DS[10] cycle". It combines the read-out of the integration results (and reset values) from the entire row with the initial reset of the selected row. This combination is possible since these two actions directly follow each other in the final implementation of the scheme. The immediate reset after the read-out of the integration values results in the direct start of the next integration period without wasting precious time, which could be used for integration. With respect to figure 3.2 this corresponds to an expansion of the reset pattern over the entire area of the sensor. None of the pixels have to wait for the start of the next integration period. In general, the time T_{DS} necessary for the execution of the DS cycle will be larger than T_{ITC} due to the fact that a considerable time is needed for a precise read-out of the results. In the current implementation it is $T_{DS} = 2\,T_{ITC}$. At the end of the DS cycle, the memory cells storing the time stamps of the corresponding row are initialized.

In figure 3.3 the process of finding a suitable row sequence is illustrated by the simple example of two rows whose the integration time is regulated by three ITC cycles. The two sequences have to be combined to a single one while avoiding the overlap of accesses. This can be achieved by a reasonable choice of T_{row}, which is the time between the DS cycles of two successive rows, i.e. the time between the read-outs of these rows.

The determination of T_{row} turns out to be a complex task if many rows have to be controlled and several ITC cycles are used to reach a large expansion of the dynamic range. Even for the imager built within this work with 170 rows and a maximum of 15 ITC cycles (+ one DS cycle) a total amount of $170 \cdot 16 = 2720$ row accesses have to be coordinated within T_{max}. Larger arrays will result in an even higher effort since many more accesses have to be distributed over the same integration time. Therefore, a systematical approach has to be developed, which permits the determination of the row sequence for any array size and for a user defined number of ITC cycles and integration time T_{max}.

Owing to the fact that an ITC cycle cannot be followed by another ITC cycle (or a DS cycle) until the first one has passed off, the minimum realizable integration time is T_{ITC}. The problem of finding T_{row} can be simplified by the partitioning of T_{max} in units of ITC cycles which is equivalent to the condition that T_{row} is only allowed to be a multiple of T_{ITC}.

As mentioned above, the incidence of a row access by a DS cycle indicates the end of an integra-

[8] Integration Time Control

[9] For the current implementation it is T_{ITC}=800 ns.

[10] Double-Sampling

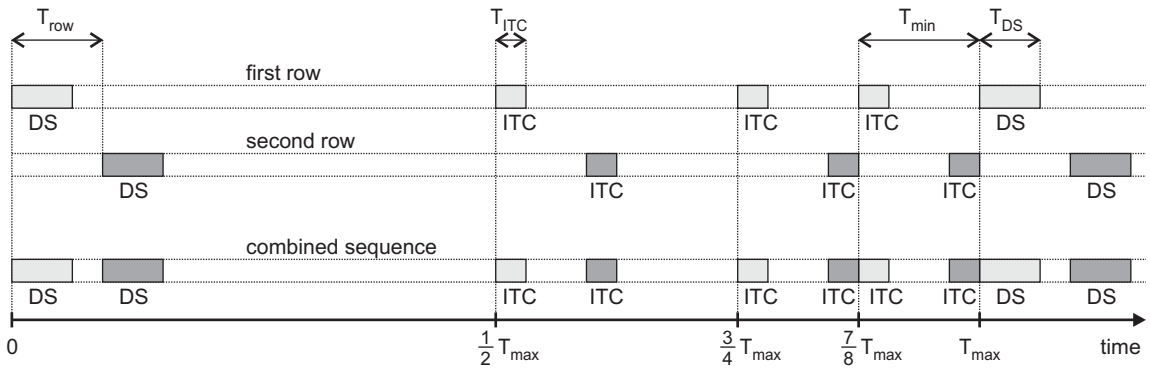


Figure 3.3: Illustration of a possible arrangement of ITC/DS cycles to reach a non-overlapping sequence of row accesses. The example is limited to two rows and three ITC cycles.

tion period and the beginning of the next period with a maximum length of T_{max}. Since the execution of the row accesses is a periodic process for each row, the accesses can be arranged on a circle with a circumference of T_{max}. For every row that has to be controlled, the corresponding DS and ITC cycles have to be added under avoidance of overlap. This is illustrated in figure 3.4 at the simple example of three rows, which can integrate for a maximum period of $T_{max} = 24\,T_{ITC}$. The integration time is regulated in three conditioned resets, i.e. three ITC cycles per row. Different values of T_{row} are possible for the arrangement of the cycles. The left circle shows the resulting sequence for $T_{row} = 5$, whereas for the right one a value of $T_{row} = 8$ is chosen.

Previous to a systematical determination of a suitable T_{row} the following parameters have to be chosen:

T_{max}**:** maximum integration time

T_{min}**:** minimum integration time

T_{ITC}**:** length of an ITC cycle

N_{ITC}**:** number of ITC cycles per row

N_{row}**:** number of rows in the sensor array

The parameters cannot be chosen independently from each other. For the realization of the described integration time control, the following relation has to be fulfilled:

$$T_{max} = T_{min} \cdot 2^{N_{ITC}} \tag{3.6}$$

It should be noted here, that this is only necessary for a precise realization of the points in time when the ITC cycles are executed. In practice, the execution of the first ITC cycle needs not to be exactly at $\frac{1}{2}T_{max}$ but only within a few T_{ITC} around that point in time, depending on the absolute value of T_{max}. The reason for this is the limited SNR of the sensor, which leads to the fact that small deviations from large integration times are not noticeable and therefore still acceptable. For example, assuming a maximum integration time of $T_{max} = 33\,\text{ms}$ and a very high SNR of 70 dB, a change of $\pm 5\,\mu\text{s}$ would still not be measurable. It is crucial here, that this deviation is restricted to the time between the initial reset of the pixel and the first conditioned reset. This way it is assured that shorter integration times, where the deviations could play a role, are not affected.

After the parameters above have been chosen, the determination of T_{row} can begin. Some criteria for the choice of T_{row} are:

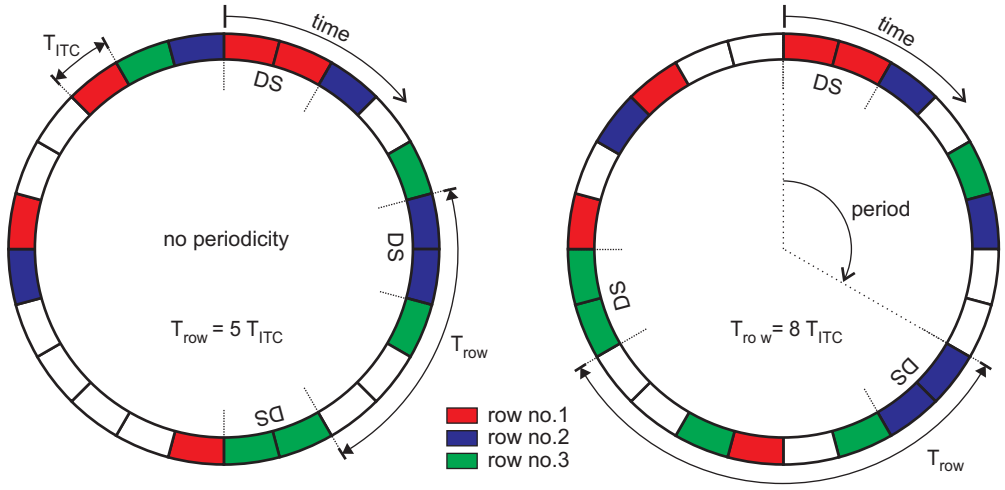


Figure 3.4: Illustration of the access distribution over the integration time T_{max} for three rows and three ITC cycles. The colored positions stand for an access of the corresponding row and indicate the execution of an ITC or DS cycle. White positions indicate time slots where no row access takes place. In the right circle a periodic solution was found.

1. $N_{row} \cdot T_{row} \leq T_{max}$
2. Maximize T_{row} in consideration of point 1.
3. Find a periodic solution.

The first criterion results from the demand for a frame rate of $1/T_{max}$. A lower frame rate is always possible. In that case, the available time is not entirely used for integration. Of course, another condition for this frame rate is that the off-chip read-out of the data from an entire row is possible within T_{row}, which is generally required for every sensor. In order to reach this with a low effort for the data acquisition system, T_{row} should be as large as possible, which is expressed by the second criterion.

The third criterion can be explained considering figure 3.4 once again. In the right circle a periodic solution was found, whereas the left circle has no periodicity. Periodicity here stands for the possibility to calculate the sequence of the remaining row accesses from the knowledge of the sequence in the first period. As can be seen, the first period in the right circle has a sequence ($T_{DS} = 2\,T_{ITC}$):

cycle:	DS	ITC	idle	ITC	ITC	idle	idle
row number:	1	2	idle	3	2	idle	idle

The following periods all have the same cycle sequence as the first one. The sequence of row numbers can be calculated by taking the value of the previous period and adding 1 to each row number. Row numbers above three are not possible here, for results above 3 the counting has to be started with row number 1 again.

Periodicity is important if the generation of the sequence is accomplished on-chip in order to be independent from an external control. If there was no periodicity, all accesses within T_{max} would have to be stored on-chip, which would require the implementation of a large additional memory. Even for the comparably low resolution of the realized imager a memory size of about

$N_{row} \cdot (N_{ITC} + 1) \cdot 9\,\text{bit} = 170 \cdot 16 \cdot 9\,\text{bit} = 24.5\,\text{kbit}$ would be required (8 bit for the row number, 1 bit for the kind of cycle).

Obviously, periodicity is reached T_{max}/T_{row} is a natural number. If this number is larger than N_{row} the resulting sequence can be understood as a solution, which is usable for all arrays with a row number of T_{max}/T_{row} or smaller (point 1, see above). Considering the right circle of figure 3.4 this is obvious, since $T_{row} = 8\,T_{ITC}$ can also be used to arrange the accesses of only two rows instead of the shown three.

T_{row} values with $T_{max}/T_{row} > N_{row}$ can also be interpreted as the use of an array with N_{row} real rows that is virtually expanded by a number of N_{virt} rows. Keeping this in mind, the three criteria for the determination of T_{row} mentioned above now can be summed up in a single equation:

$$T_{row} = \frac{T_{max}}{N_{row} + N_{virt}} \tag{3.7}$$

In order to fulfill the need for a large T_{row}, N_{virt} has to be minimized now. Values of $N_{virt} > N_{row}/2$ should be rejected to omit unnecessarily high off-chip read-out speeds. The virtual rows are treated like real rows, except for the final execution of a corresponding accesses: Since these rows do not exist, nothing has to be done (idle).

There are many solutions of equation 3.7. However, most of them are not suitable for the regulation scheme. It has to be additionally assured that the resulting cycle sequence is non-overlapping. This was tested by a simulation of the resulting sequence using *Mathematica*[11], an integrated environment for technical computing.

Based on the parameters given by the user, the simulation algorithm varies N_{virt} and calculates possible values for T_{row}. Each of these values is tested by the simulation of the full sequence of accesses. This sequence is generated by adding the required accesses row by row to the timing circle. As soon as an overlap of two row accesses is located, the corresponding value of T_{row} is rejected and the next value is tested. In addition to the variation of N_{virt} it is also possible to permit small deviations from T_{max} as was mentioned in the discussion of equation 3.6, in order to minimize N_{virt}.

For the implemented imager, the sequences for three different values of T_{max} were exemplary calculated. The maximum integration time finally employed can be chosen in the user interface (cf. section 5.1.2). The parameters resulting from the simulation permit a simple calculation of the next access and are shown in appendix D. The calculation itself and the execution of the access is accomplished by an FPGA[12], whose programming is described in the next section.

3.5 Programming of the FPGA

The implemented imager is controlled by a programmable logic array. The use of an FPGA provides the flexibility to test different control schemes in preparation of a later on-chip implementation. Furthermore, the programming of specific test modes simplifies the measurement of particular properties of the imager such as leakage currents of the analog memory cells or comparator offsets (cf. chapter 5).

For the programming of the FPGA it is necessary to write the source code in a hardware description language. For the presented test system this was done using VHDL[13] [VHD97]. The code is a description of the abstract behaviour of the desired logic. Simulations to assure the correctness of

[11] *Mathematica* is a registered trademark of Wolfram Research Inc.

[12] Field Programmable Gate Array

[13] VHDL=VHSIC HDL= Very High Speed Integrated Circuit Hardware Description Language

the code were accomplished employing the software tool *Modelsim*[14]. The code was subsequently converted into an electronic circuit that can be mapped to the internal FPGA structure. This was done using the software package *Xilinx ISE 5.2*.

In the following an overview of the programming is given. The decisive task of the FPGA is the calculation of the row access sequence and the corresponding cycles that have to be executed. However, as part of the electronic test system its scope of functions is not limited to the control of the imager. Additionally, it includes the management of the data acquisition and the interfaces for the communication with the PC.

Overview

A block diagram illustrating the functionality of the VHDL module is shown in figure 3.5. A description of the corresponding signals can be found in appendix A. The *main control* is responsible for the overall control and the communication between the individual components. Parameters for the operation of the imager chosen by the user are stored here and passed to the responsible components. At the beginning of each new frame the user settings are checked for any changes and the requested operation is executed. Settings which are not changed during the processing of a frame are directly led to the imager. These are the choice of the averaging mode (signals AV, AV4, AV8), the storage of a new position of the HDR window (signal STORESADR) and the activation of test settings (signal ARROFF). The main system clock is used to generate the clocks for the individual components.

The blocks *read-out control*, *SRAM control*, *DAC control* and *PCI I/O* stand for the interfaces to the corresponding devices, which are located on the printed circuit board hosting the FPGA. The devices are used for the data acquisition (on-board ADC/SRAM), the control of bias voltages (DAC[15]s) and the communication with the PC, which hosts the FPGA board. The properties of these devices and their applications in the test environment are explained in the description of the electrical test setup (cf. section 5.1.1).

The functional block *DS control* generates the control signals for the in-chip data acquisition. The acronym DS stands for ”double-sampling”. It describes the acquisition process which consists of a subsequent sampling of the integration result and the reset values in order to reduce the fixed pattern noise. The meaning of the output signals and the generated timing diagram are explained in section 4.4.1. The following will focus on the *sequence calculation* as this is the most important part of the imager control.

Sequence Calculation

In the previous section it was shown how suitable access sequences can be found for a given set of parameters, which determine the operation of the integration time control. The periodicity of the simulated sequences permits the calculation of the next access, i.e. the number of the next row and the kind of cycle that has to be executed, from the knowledge of the sequence in the first period. Therefore, only the first period has to be stored in the FPGA.

The period is stored in a compact format consisting of three one-dimensional arrays. The array *cflag* indicates the cycle that has to be executed (1=ITC, 0=idle). A period always starts with the only DS cycle within this period, therefore this cycle has not to be indicated. In the second array *num* the number of the row that has to be accessed in an ITC cycle is given, whereas for idle cycles the stored number is the number of idle cycles which have to be executed next. The third array *stamps* contains the corresponding stamp that has to be transferred to the imager. The stamp sequence stays

[14] Modelsim is a VHDL/Verilog simulator distributed by Mentor Graphics.

[15] Digital-to-Analog Converter

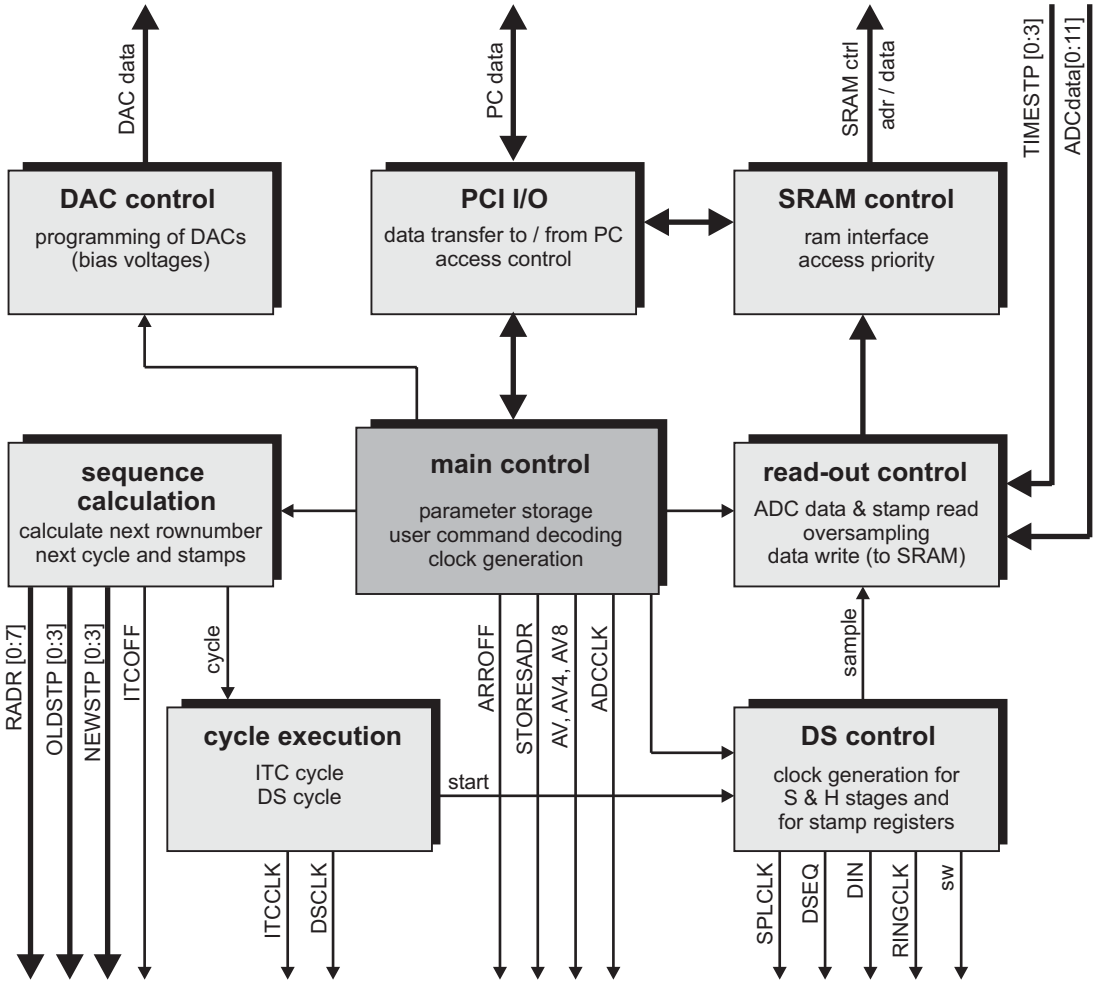


Figure 3.5: Block diagram showing the main functional blocks of the FPGA programming. A short description of the signals connected to the imager can be found in appendix A.

the same for every period. Table 3.1 shows the contents of the arrays for the simple example that was discussed in the previous section (compare with the right circle in figure 3.4).

index i:	1	2	3	4	5	6
cflag:	-	1	0	1	1	0
num:	1	2	1	3	2	2
stamps:	0	2	-	1	3	-

Table 3.1: Parameters for the row accesses within the first period by the example of the control sequence illustrated by the right circle of figure 3.4.

The parameters for the next access are the contents at a specific index i, which is successively increased. When the end of the array is reached, the index i is initialized (i=1) again. If the current access is a DS or an ITC cycle, i.e. index=1 or *cflag*=1, the cycle is executed and the content of *num* is increased by one to prepare the next period. Otherwise, i.e. if *cflag*=0, the number of idle cycles

given by *num* is executed and index i is increased without changing the content of *num*.

There are two points which complicate the determination of the next access from the stored sequence. The first one is the access of a virtual row. Owing to the fact that these rows do not exist in the real sensor, nothing has to be done. Strictly speaking, it has to be assured that no cycle is executed. For instance, the execution of an ITC cycle could lead to the storage of wrong time stamps that could overwrite correct values. Fortunately, the identification of virtual rows is simple since they have row numbers larger than the maximum row number of the real sensor.

The second point concerns the HDR window. The window is defined by the fact that the integration time is solely regulated for those pixels which are located within its region. Therefore, ITC cycles of rows located within the window have to be executed, whereas those of rows lying outside have to be excluded from execution[16]. Regarding the DS cycles there is another difference between them. Lines lying outside of the window have to be reset and when their fixed integration time T_{max} has passed they have to be read out. In consequence, DS cycles of these rows have to be executed as well. However, a DS cycle should not result in an initialization of stored stamp values here, since there are no stamps stored for these rows. The sensor chip provides the possibility to disable the read/write of stamps during a DS cycle by use of the signal ITCOFF. This signal has to be used if a DS cycle for a row lying outside of the HDR window is executed.

In figure 3.6 the procedure that was explained before is illustrated by a flow diagram. For every new value of the index i the diagram is evaluated and the actions given in the individual boxes are executed. The additional test if the maximum virtual row number is reached assures the correctness of the sequence for the next frame.

In appendix D the contents of the arrays *cflag*, *num* and *stamps* are given for the implemented maximum integration times. Other values of T_{max} can be used by a simple change of the contents of the arrays. The maximum value of the index i is 32. Each entry of *cflag* has a width of 1 bit and each stamp uses 4 bit. Depending on the value of T_{max}, the contents of *num* require up to 12 bit. Therefore, the required memory for the complete period sums up to $32 \cdot (1+4+12)\,\text{bit} = 544\,\text{bit}$.

[16] Notice that a line that crosses the HDR window will always also comprise pixels which are located outside the window. This is solved by an additional restriction of the regulation in horizontal direction, as will be explained in section 4.3.5.

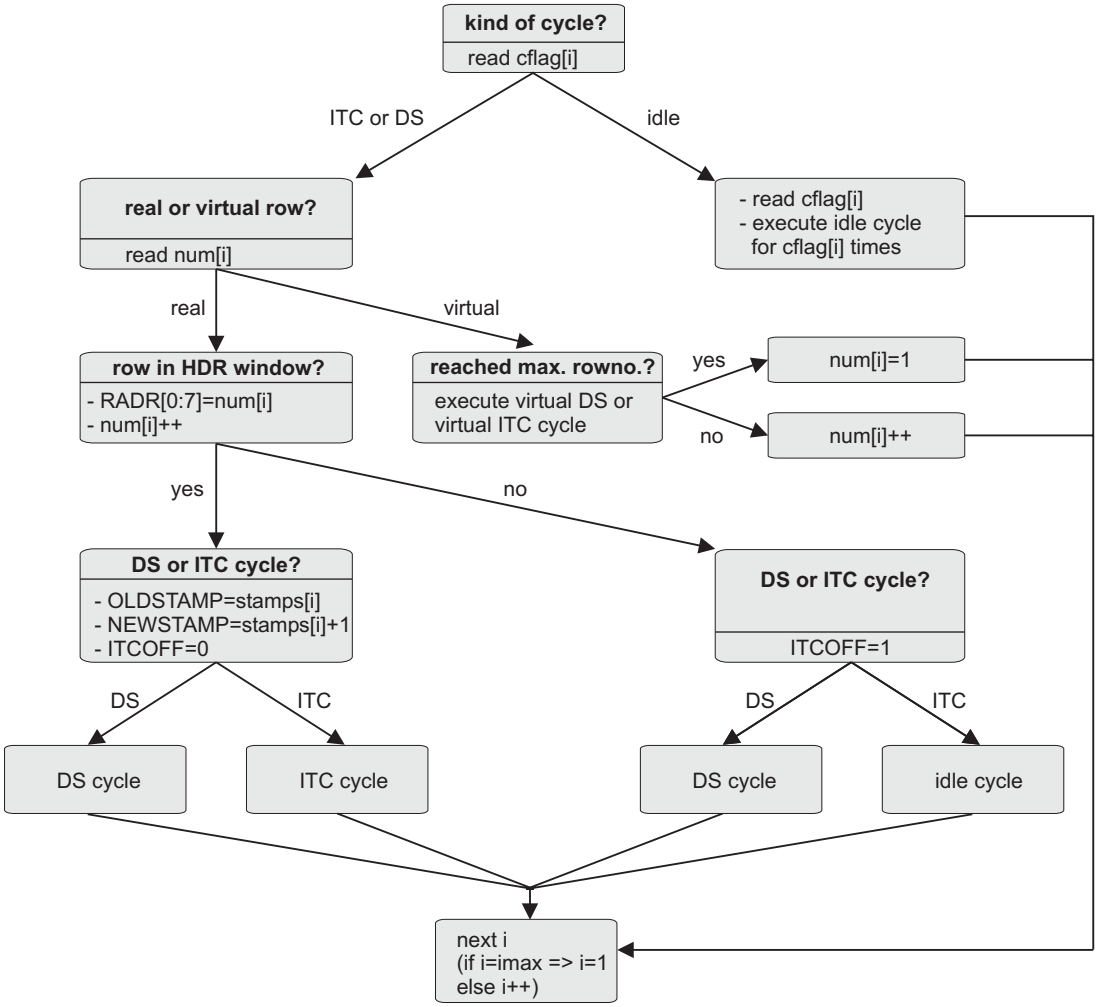


Figure 3.6: Flow diagram illustrating the determination of the parameters for the next row access from the stored sequence of the first period.

Chapter 4

Implementation

This chapter describes the implementation of the image sensor in CMOS technology. First, an overview of the logical chip architecture is presented to subdivide the chip into building blocks and to illustrate the global data flow. Next, the individual components of the chip and their interplay with each other are described in terms of circuit diagrams and layout implementation. Timing diagrams illustrate the internal course of an ITC cycle and a DS cycle, which were introduced in the previous chapter to explain the working of the integration time control. The chapter closes with a die photograph and a description of the global layout.

4.1 Chip Architecture

Before the basic architecture of image sensor is explained, there will be some introductory remarks regarding the history of development and the CMOS process the chip has been realized in. This background is necessary to understand the final architecture and the details of its realization.

4.1.1 Introductory Remarks

The image sensor realized within this thesis was build to prove the functionality of the adaptive integration time regulation and to demonstrate the attainable performance and flexibility of this concept. Initially it was planned to equip the chip with a promising technology referred to as thin-film-on-ASIC[1] (TFA) [SCH99]. This technology permits the integration of photodiodes on top of the chip by a post-processing of the wafers, which were produced in a standard CMOS process. This way it is possible to reach a fill factor of 100 %. The vertically integrated photodiodes consist of amorphous Si and are connected with the electronics underneath by an additional metal layer. Unfortunately, at a certain time during the course of this theses the TFA technology was no longer available due to the economical development of the company that provided this technology. The design phase of the imager was almost completed at this time and based on the integration of TFA photodiodes. However, since the concept of adaptive integration time regulation is independent of the used photodiodes, it was decided to redesign the sensor array for the use of the diodes provided by the CMOS process. The

[1] Application Specific Integrated Circuit

consequence of this was an increase of the pixel pitch. In order to keep the size of the sensor array within the planned area and still be able to connect it to the surrounding circuits the resolution of the sensor had to be decreased. The initially planned resolution of 256×256 pixels with a HDR window of 128 ×128 was changed to a resolution of 170 ×170 with a HDR window of 85×85. The design of the integration time control (i.e. the ITC entity, see section 4.3) was also completed at this time and was not changed on account of the replaced sensor array. In principle, the realized ITC entity is therefore able to process a window size of 128 ×128, but due to the necessary pitch adaptation between the entity and the sensor array only a window of 85×85 is available.

For the TFA post-processing whole wafers are needed to avoid border effects, which would otherwise arise due to inhomogeneous coating of single dies. The 0.25 μm CMOS process used was chosen on account of the availability of whole wafers without the need for an expensive production run[2]. The disadvantage of the process was the lack of an available library of standard cells (e.g. buffers, gates, flip-flops) except for the pad structures. As only radiation hard cells were available, which were not used due to their size, every implemented circuit in the presented chip had to be designed from scratch.

Another difficulty arose from the coating of the dies with a layer of polyimide. This layer partly blocks incident light. It could finally be removed by chemical post-processing. The details of the procedure are presented in appendix C.

4.1.2 Overview of the Chip Architecture

In this subsection the individual functional blocks of the chip are identified and their interplay with each other is explained. They are discussed in greater detail in the subsequent sections. Figure 4.1 depicts the basic architecture of the image sensor. As shown, the chip can be subdivided into four functional blocks: sensor array, ITC entity, DS entity, and the cycle control. The arrows between these blocks indicate the internal communication among them, whereas the other arrows stand for the external communication with the outer world via the I/O pads of the chip. A diagram in appendix A illustrates the pad distribution over the chip. In a subsequent table a description of the corresponding signals is given. Signals that are exclusively used for test purposes and with no relevance to the ordinary operation of the chip are omitted in figure 4.1 for sake of clarity. The only output of the chip is the sampled voltage V_{int} from the integration of the photocurrents, the sampled reset voltage V_{res}, and the digital bus TIMESTP[0:3] which provides the exponent of the integration time control (the time stamp) that can be read out serially for all pixels.

The largest block on the chip is made up of the sensor array itself. It consists of 178 ×178 pixels, of which 170 ×170 are addressable and can be used for the image acquisition. The remaining four columns/rows on each side are used as dummy structures to avoid border effects in the selectable array. The drivers for the selection of an individual row of pixels as well as the necessary decoder logic are located on the right side of the array. The connection circuits for the selection of the averaging scheme can be found on the left side and on the bottom. The load transistors for the in-pixel source followers are positioned on top of the array (one transistor per column). The only signals coming from the sensor array are the pixel output voltages V_{pix} [0:169] which are connected to the ITC entity as well as to the DS entity.

The double-sampling entity is responsible for the analog data acquisition and the serial read-out of the results from integration and reset of the pixels. It contains the sample-and-hold stages for the

[2] In a ”production run”, a wafer is exclusively used by a single chip design that is repeated multiple times to fill the area of the wafer. In contrast to this, in a so called ”multiproject run” designs from different chips are placed on the same wafer for economical reasons (e.g. shared costs for photomask production).

parallel signal acquisition from the pixels of the selected row. Each memory cell of these stages is connected to a column preamplifier which can be selected by a shift register for a serial read-out of the sampled values. The two analog read-out lines are connected to two-stage operational amplifiers which are used as off-chip drivers.

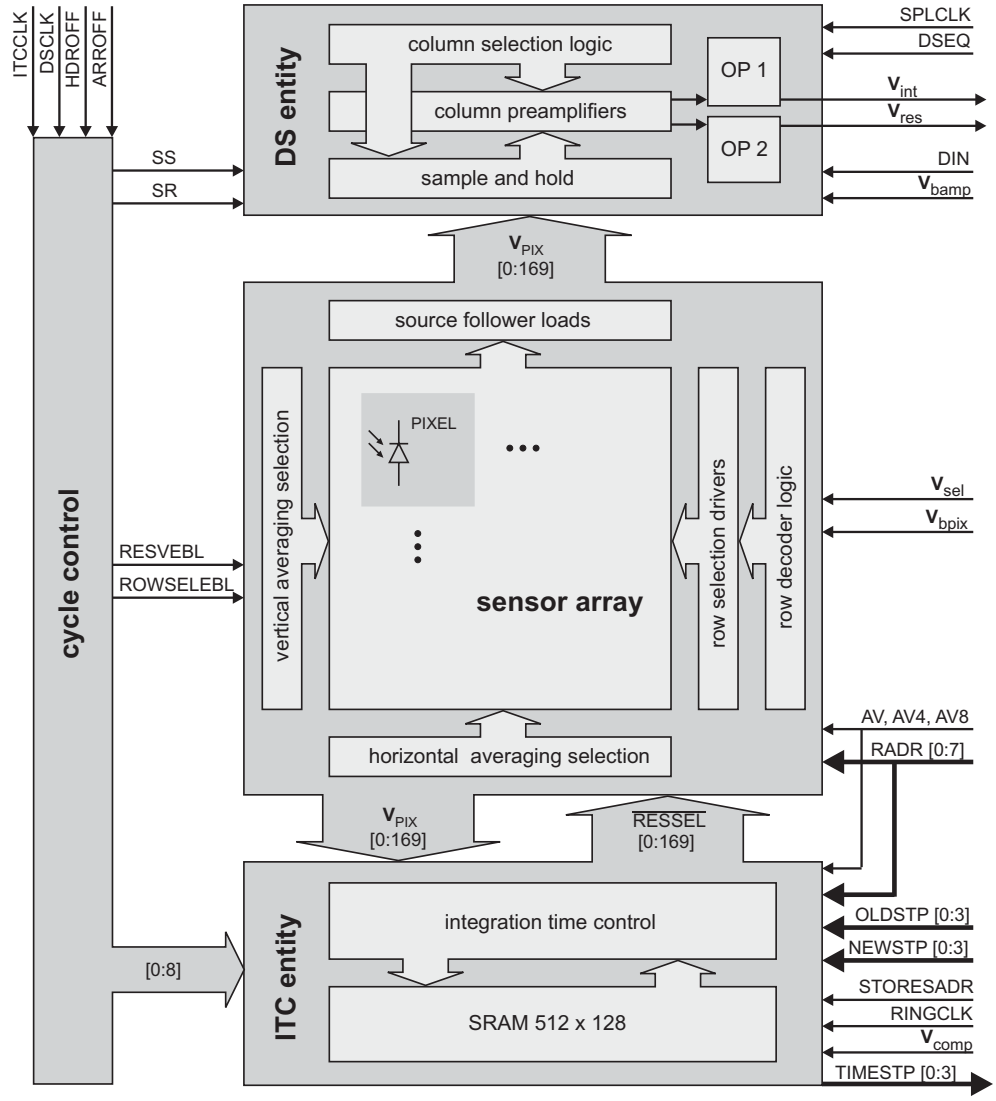


Figure 4.1: Overview of the basic image sensor architecture. Data paths that do not connect two blocks with each other indicate the communication of the chip with the outer world (compare with the pad diagram in appendix A). The internal architecture of the ITC entity is expanded in figure 4.10.

The adaptive integration time control is carried out by the third block, the ITC entity. The only input received from the sensor array are the pixel output voltages V_{pix} [0:169]. The only output lead back to the array is the digital bus $\overline{\text{RESSEL}}$ [0:169]. This bus indicates which pixels from the selected row have to be reset to avoid saturation. As discussed in chapter 3 the regulation of the integration time relies on the fulfillment of two reset conditions, which have to be verified in the ITC entity (cf. section 3.2). As mentioned above, this verification requires a local memory. It consists of a 512×128 bit SRAM, which occupies the largest area of the entity. The complete architecture of the

ITC entity is too complex to be included in figure 4.1. The full logic structure of it is shown in figure 4.10 of section 4.3 where it is discussed in detail.

The last block in figure 4.1 is the cycle control. It contains the digital logic for the execution of the DS cycle and the ITC cycle. As mentioned in the previous chapter these cycles result in the read-out and reset of a whole row or the selective reset of individual pixels in a row, respectively. Therefore the control signals lead to all of the other blocks, most to the ITC entity (see figure 4.10). The signals DSCLK and ITCCLK from outside of the chip determine the time when one of these cycles has to be executed. The signals RADR[7:0] determine the number of the row the respective cycle is applied to. The sequence of the cycles was discussed in the previous chapter.

In the following sections the four functional blocks of figure 4.1 are described in greater detail. First, the sensor array and in particular the design of a pixel is presented. Afterwards, the circuitry of the ITC entity is explained. The third section contains a description of the DS entity. Finally, the cycle control, which controls large parts of the previously introduced blocks, is presented.

4.2 Sensor Array

The sensor array is that part of the chip, where the conversion of incident light into a processable electronic signal is accomplished. It consists of 170 $\times$170 selectable photoreceptors (pixels), whose circuit, functionality, and layout implementation are described in this section. Besides the pixel itself, the address decoders for the selection of a particular row, the row drivers, and the averaging control will be presented.

4.2.1 Photoreceptor Implementation

In chapter 2 the functionality of the three transistor APS has been described, which is used as a photoreceptor in many integration based concepts. In contrast to continuous working sensors the APS is operated in two distinct phases: In the reset phase the junction capacitance of a connected photodiode is charged to a predefined voltage, whereas during the integration phase this capacitance is discharged by the photocurrent. Owing to the fact that there is only one transistor for the reset control, it is not possible to reset individual pixels within a particular row. Instead of this, only an entire row can be reset. As explained in chapter 3, the adaptive integration time control is based on the reset of individual pixels, therefore the three transistor APS had to be expanded.

Photoreceptor Circuit

Figure 4.2 shows the circuit diagram of the implemented photoreceptor. The channel dimensions (W/L in μm) of the individual transistors are written next to the corresponding identifier M. The devices M_2-M_4 serve the same purpose as their counterparts in the three transistor APS. M_3 is part of the source follower, which is used to amplify the signal connected to its gate. The load transistor of the source follower is located outside the array. There is one for each pixel column (drawn in grey in figure 4.2). A pixel is read out by connecting its output to a column read-out line, which is done by closing the single transistor switch M_4. When ROWSEL is pulled high, an entire row is connected to the read-out lines. Transistor M_2 is used to reset the capacitance at the integration node n1. In order to enable the reset of individual pixels of a row, the transistor M_1 has been added to the circuit. An individual reset can be accomplished by a simultaneous control of $\overline{\text{RES}}$ and $\overline{\text{RESSEL}}$. $\overline{\text{RES}}$ is common to all pixels of a particular row while $\overline{\text{RESSEL}}$ is common to all pixels of a particular column of the array. M_1 and M_2 are realized as PMOS transistors. This way, the dynamic range of the pixel is enhanced since a reset voltage of V_{dda} can be guaranteed. Additionally, the necessary reset time is

shortened and the risk of image lag[3] is eliminated. A purely NMOS[4] implementation would also be possible if $\overline{\text{RES}}$ is strongly increased to reach a complete reset in this case.

The bulk terminals of all PMOS transistors in the pixel are connected to V_{dda}. The devices M_5 and M_6, also realized as PMOS, were introduced to allow an averaging of photocurrents of adjacent pixels. V_{hor} is connected to the integration node of the pixel on the right side, whereas V_{ver} is connected to the pixel above. The details of the averaging will be discussed in section 4.2.3.

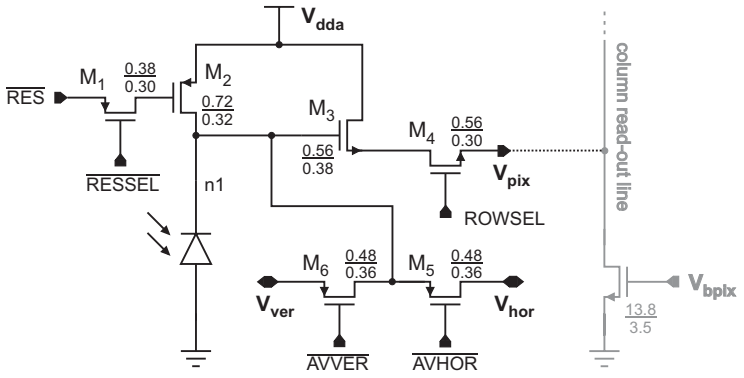


Figure 4.2: Circuit diagram of the implemented photoreceptor.

For the realization of a small pixel pitch and to have as much pixel area as possible available for the implementation of the photodiode, it is necessary to keep the transistor dimensions small. The reduction of the dimensions is limited by the fact that the effects of mismatch will increase drastically with approaching minimum channel size ($W_{min}/L_{min} = 0.30\,\mu$m/$0.24\,\mu$m) (cf. section 1.4). A factor of 1.3-3.0 above the process minimum was selected for most of the transistor dimensions. In particular this is important for M_3 since parameter variations of this transistor are responsible for a substantial part of the pixel FPN. Besides, an increase of its length leads to a decrease of the channel length modulation, which in turn results in an improvement of the linearity. For the reset transistor M_1 on the other hand, an increased length is necessary to reduce the leakage current flowing to the integration node.

Operation of the Pixel

The operation of the pixel is as follows. First, an entire row of pixels is reset by pulling $\overline{\text{RES}}$=0 and $\overline{\text{RESSEL}}$[0:169]=0. After the voltage at the integration node has reached V_{dda}, $\overline{\text{RES}}$ is pulled high ($\overline{\text{RESSEL}}$[0:169] stay low) to start the integration phase. The photocurrent generated by the incident light now begins to discharge the capacitance at the integration node. This current is assumed to be constant during the integration phase, it would therefore lead to a linear decrease of the voltage at the photodiode. Deviations from the linear decrease of the pixel output voltage V_{pix} arise on the one hand from the voltage dependency of the photodiode junction capacitance (cf. equation 1.9) and on the non-linearity of the output source follower on the other hand.

In figure 4.3 the simulated response curve for different photocurrents is shown. A photocurrent of 5 pA corresponds to an incident light intensity of approximately 1 W/m^2. No regulation of the

[3]Image lag is a consequence of the dependency of the final reset voltage from the photodiode voltage at the beginning of reset (previous integration result). In a NMOS implementation of the pixel circuit image lag can arise owing to the fact that the reset transistor is operating in weak inversion at the end of the reset period. The source of lag is therefore different from the source of image lag in CCDs where it is caused by incomplete charge transfer.

[4]N-channel MOS

integration time has been used here. In the case of I_{ph}=25 pA and 5 pA the integration capacitance is discharged until the stored voltage is too low to be amplified by the source follower (saturation). In figure 4.4 on the other hand, the simulated response curve for I_{ph}=5 pA with activated integration time control is shown. In order to prevent the photoreceptors from being saturated, their output voltages are checked periodically row by row as described in the previous chapter. One of the conditions for a reset of a particular pixel is the decrease of V_{pix} below a predefined value V_{comp}. This condition is checked here at the time $t = 4$ ms and the pixel is subsequently reset. After the maximum integration time T_{max}=8 ms the signal is read out. Due to the reduced real integration time of 4 ms the pixel is still unsaturated at this point in time.

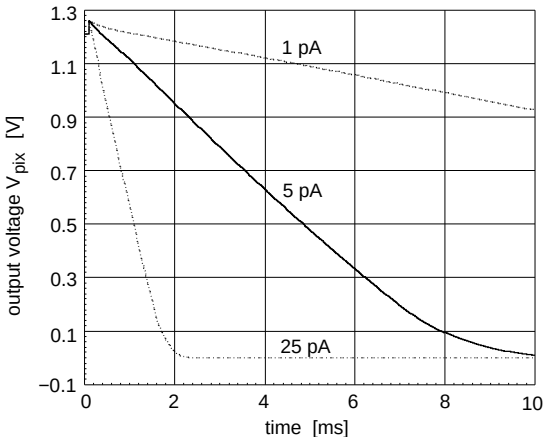


Figure 4.3: Simulated response curves for different photocurrents (without integration time control).

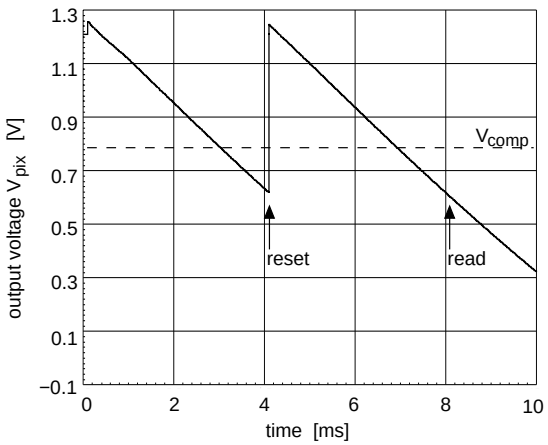


Figure 4.4: Simulated response curve for $I_{ph} =$ 5 pA (activated integration time control).

The procedure to exclusively reset individual pixels of a selected row is as follows: For pixels of a selected row that should not be reset the corresponding $\overline{\text{RESSEL}}$ signal is pulled high separating the gate of the reset transistor M_2 from the $\overline{\text{RES}}$ line. Subsequently, $\overline{\text{RES}}$ is pulled low to accomplish the reset of those pixels which are still connected. Finally, after $\overline{\text{RES}}$ has returned to high (end of reset phase), all signals $\overline{\text{RESSEL}}$ [0:169] are pulled low to reconnect the gate of the reset transistor M_2 to the $\overline{\text{RES}}$ line. The connection to $\overline{\text{RES}}$ throughout the long integration periods between the reset curtains (video mode: up to $T_{max}/2$=16.5 ms) is important to prevent the pixels from being partly reset. Within the short period when $\overline{\text{RESSEL}}$ is high (here 300 ns, cf. section 4.5.2), the gate voltage of M_2 is stored on the small parasitic capacitance at the drain of M_1 to prevent a reset. If this node is isolated too long from the $\overline{\text{RES}}$ line, leakage currents will lead to a decrease of the stored gate voltage of M_2 and initiate a reset current flowing from V_{dda} to the integration node (partial reset).

The slope of the response curve is determined by the capacitance at the integration node and the gain of the source follower. The capacitance of the photodiode (reverse biased at 2.5 V) and the parasitic capacitances sum up to 22 fF, which results in a conversion gain of 7.3 μV/e$^-$. Taking the gain of the source follower (g=0.77) into account a total conversion gain of 5.6 μV/e$^-$ is achieved at the column read-out lines.

The exact course of the illustrated pixel response curve also depends on the settings of the control voltages. For the simulations shown in this section and most of the measurements presented in chapter 5 the following settings were used.

Regarding the choice of the bias voltage of the source follower, on the one hand it should be as

low as possible to reach a large output voltage range, on the other hand it has to be high enough to meet the requirements for speed and noise immunity. Basing on simulations, it is adjusted to V_{bpix}=800 mV, which results in a current of about 15 pA flowing through the load transistor during the read-out of a pixel. The analog supply voltage is set to V_{dda}=2.5 V.

The voltage to switch off the reset transistor M_2, i.e. the 'high' level of $\overline{\text{RES}}$, has to be chosen considering the mismatch of M_2. The use of small transistor lengths can lead to high leakage currents, which will vary from device to device. Since M_2 isolates the integration node from the reset voltage for the full integration time T_{int}, leakage at this point can lead to an increase of the pixel FPN. It turned out that the leakage of M_2 is too strong to be switched off with a voltage of 2.5 V. Thus, a switch-off voltage of 2.7 V had to be used. This could only be reached by an increase of the digital power supply voltage V_{dd} to 2.7 V, which is still within the process specifications.

The comparison voltage V_{comp} is set to 800 mV. This guarantees that the final output voltage of a pixel stays above 400 mV and therefore lies within the linear range of the response curve.

Regarding the operation of the pixel under high illumination, it has to be taken into account that a pixel can be temporarily saturated, which will lead to charge overflow in the photodiode. A consequence of this could be an increase of crosstalk between neighbouring pixels. The integration time control guarantees that no saturation occurs between the final reset and the read-out of the pixel if the incident light intensity is not too high for the shortest possible integration time. Nevertheless, since the regulation consists in a successive decrease of integration time, it cannot be avoided that pixels are eventually saturated during some of the longer integration periods. The longest time between two checks of the reset conditions is half the maximum integration time T_{max} (cf. chapter 3). Saturation occurs if the time T_{sat} is shorter than T_{max}/2. T_{sat} is the time which the voltage at the integration node needs to reach the ground level. This time can be calculated by

$$T_{sat} = \frac{C \cdot V_0}{I_{ph}} \tag{4.1}$$

where C is the capacitance at the integration node, V_0 is the voltage at the begin of integration (2.5 V) and I_{ph} is the photocurrent. For a medium incident intensity of 1 W/m^2 this results in a saturation time of T_{sat}=11 ms. Therefore, even at medium intensity saturation results under the assumption of video frame rate ($T_{max}/2$=16.5 ms).

Provisions against crosstalk were made by the implementation of a substrate contact per pixel, which is connected to the ground line. In section 5.2.8 crosstalk measurements will be presented. Disturbances deriving from crosstalk are not visible in the images taken with the sensor.

Layout of the Photoreceptor

The imager has been realized in a 0.25 μm standard CMOS process, which provides three metal layers and one layer of polysilicon. Figure 4.5 contains a cutout from the layout of the implemented sensor array. Two complete pixels are shown together with parts of the adjacent ones.

The assignment of the individual layers is explained at the bottom of the figure. In this explanation *diff* stands for diffusion and *poly* for polysilicon. The layer *block* denotes areas which are blocked from the n-channel FET source/drain/gate implants, i.e. it indicates p-channel FETs, p^+-junctions and substrate contacts. The *via1/2* layers define connections between *metal1/2* or *metal2/3* respectively, whereas *contact* indicates a connection between *metal1* and *diffusion* or *poly*. While examining the layout it has to be kept in mind, that the production process permits to stack *contact*, *via1*, and *via2*, which therefore can be covered by each other in the presentation of the layout.

A single pixel has a size of 7.5 μm $\times$7.5 μm. The pixel layout is dominated by the photodiode itself, which occupies about 40 % of the total pixel area. It is implemented as a n^+-substrate-diode

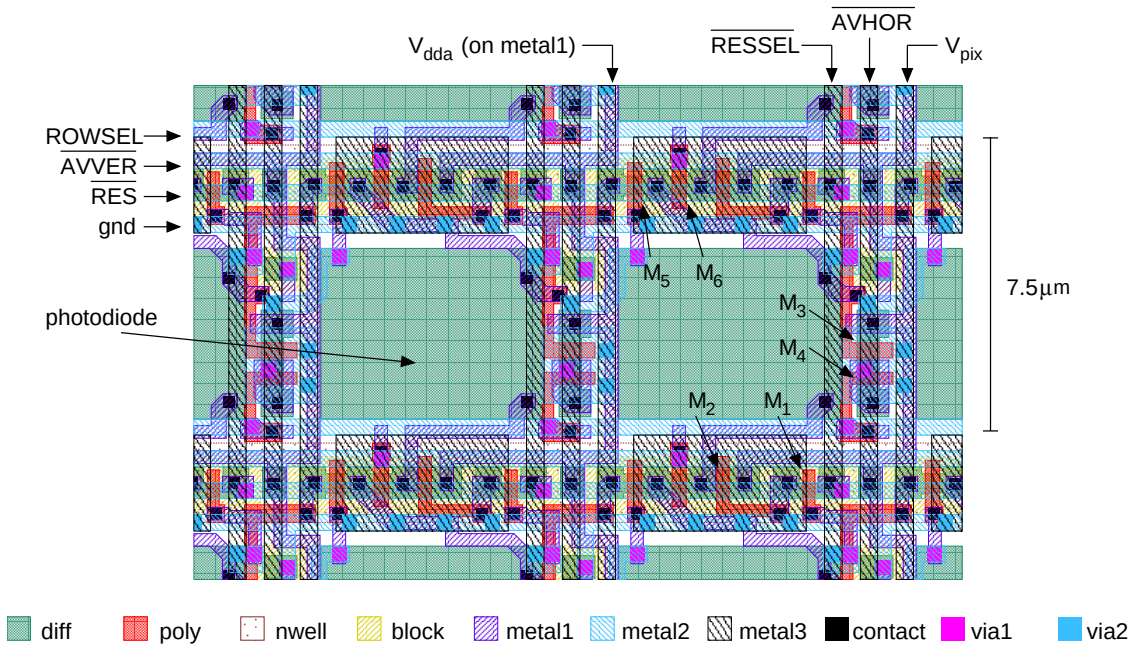


Figure 4.5: Section of the layout of the sensor array as it is implemented in the imager. The individual devices are indicated by their identifier M (cf. figure 4.2).

to take advantage of the comparatively high quantum efficiency (see section 5.2.1 for a comparison of the different diodes). In the used CMOS process, diffusions and polysilicon are automatically silicided[5] to reach low-resistance interconnects. In order to still be able to build resistors made of diffusions or polysilicon, an extra mask is available for the definition of regions where the formation of silicide should be blocked. The use of this mask made it possible to omit the silicidation for the photodiodes, which is essential for the penetration of the incident light.

The control and signal lines needed for the operation of the photoreceptor are indicated in the layout. Vertical running lines are realized in *metal3* (top metal layer), whereas horizontally aligned lines run on layer *metal2*. All these lines exhibit the minimum width and the minimum possible spacing. *Metal2* and *metal3* are also utilized for shielding the in-pixel transistors from the incident light. The layer *Metal1* is used for the in-pixel connections and for the analog supply voltage V_{dda}. Each pixel of a particular row has its own (vertically running) V_{dda} line. As a consequence of this, when a row is selected for read-out (or reset), the resulting load is distributed over many lines and the supply voltages of the pixels are decoupled from each other to a large degree.

Instead of a rectangular outline, adjacent pixels are interlocked, but the resulting pitch is 7.5 μm in horizontal as well as in vertical direction. The transistors indicated in figure 4.5 by their identifier M belong to the same pixel. The nesting is necessary for the realization of a compact layout.

As mentioned in section 4.1.1, in an earlier stage of the sensor development the use of TFA based photodiodes was intended instead of using the junctions provided by the CMOS process. In figure 4.6 a section of the layout of the corresponding sensor array is shown. Owing to the fact that the photodiodes are integrated on top of the die by post-processing of the wafer, only a small contact has to be provided for the connection of the photodiode lying above the pixel circuit. The connection

[5] In a silicide technology, a material with a low resistance (such as $TiSi_2$) is placed on top of the polysilicon/diffusions to strongly reduce the overall resistance [ALL02-1].

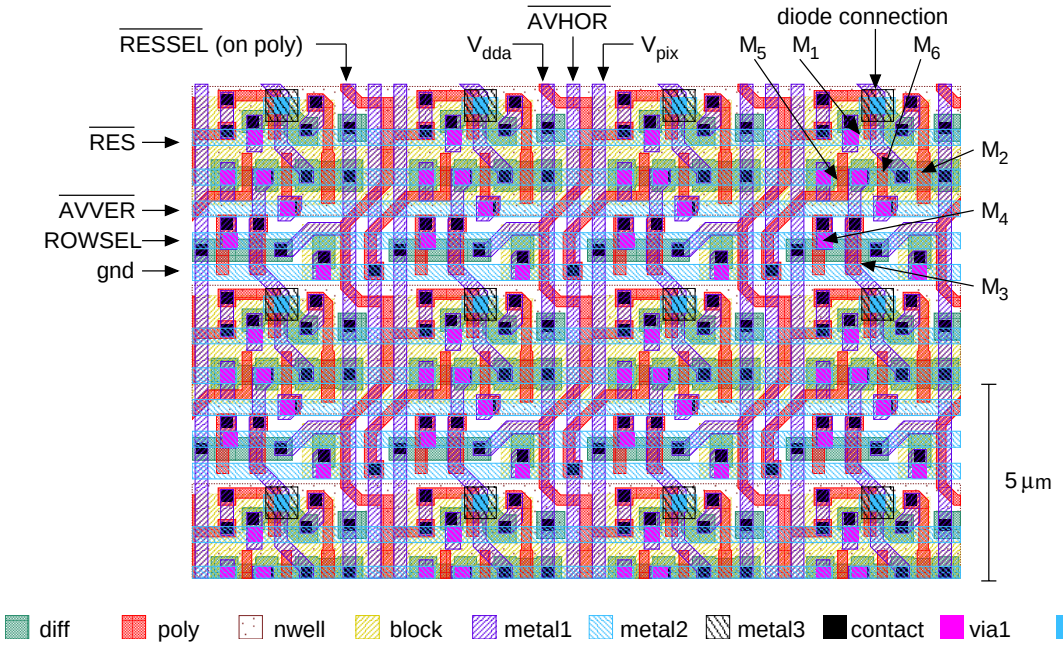


Figure 4.6: Cutout from the layout of a previously intended TFA based sensor implementation. The layout is shown on the same scale as that in figure 4.5.

itself is created by an additional patterned metal layer. On account of the saved area the pixel pitch can be reduced to 5 μm. The line for the signal $\overline{\text{RESSEL}}$ is implemented here in polysilicon. For larger arrays this would have to be changed due to the high resistance of this layer, which would otherwise lead to slow $\overline{\text{RESSEL}}$ signals.

4.2.2 Address Decoder

The application of the integration time control introduced in the previous chapter results in a non-successive selection of pixel rows. This requires the possibility to arbitrarily select rows within the pixel array. In order to make this possible, a decoder was implemented to generate the selection signal for the corresponding row. At first, a binary encoded address value is presented to the decoder. Subsequently, this input is converted to a one-hot output selecting one of N row select lines (ROWSEL in figure 4.2) by the decoder circuitry.

Such an address decoder is also required to access the SRAM. Owing to the fact that there is an unequivocal assignment of each pixel row to a particular row of SRAM cells, the address has to be decoded only once. As a result, parts of the decoder can be used by both arrays and must be implemented only once.

A difficulty arises in the implementation of the decoder. The decoder connects to the array cells, i.e. the pixels or the memory cells. Consequently, the layout of a part of the circuitry has to fit into the tight cell pitch. In particular this would result in narrow but very long selection circuits for the SRAM, which would run the risk of widening the chip dimensions. To overcome this problem, the address decoder is divided into two parts, a predecoder and a final decoder. The predecoder has to be implemented only once, it is shared by both arrays. The final decoder on the other hand is comprised by the circuitry located at the end of each row select line. It is kept simple to make a small layout possible.

The structure of the complete decoder can be explained by considering figure 4.7. The predecoder consists of 21 NAND gates and eight inverters. It is designed as a small compact building block which is located close to the cycle control on the chip photograph in figure 4.38. The incoming address bits RADR [0:6] are decoded into three sets of one-hot signals, comprising a total of 20 lines (8+8+4). These lines are lead to the SRAM on the one side and to the pixel array on the other side. They run in parallel to the columns of each array and are connected to the final decoder by perpendicular running lines.

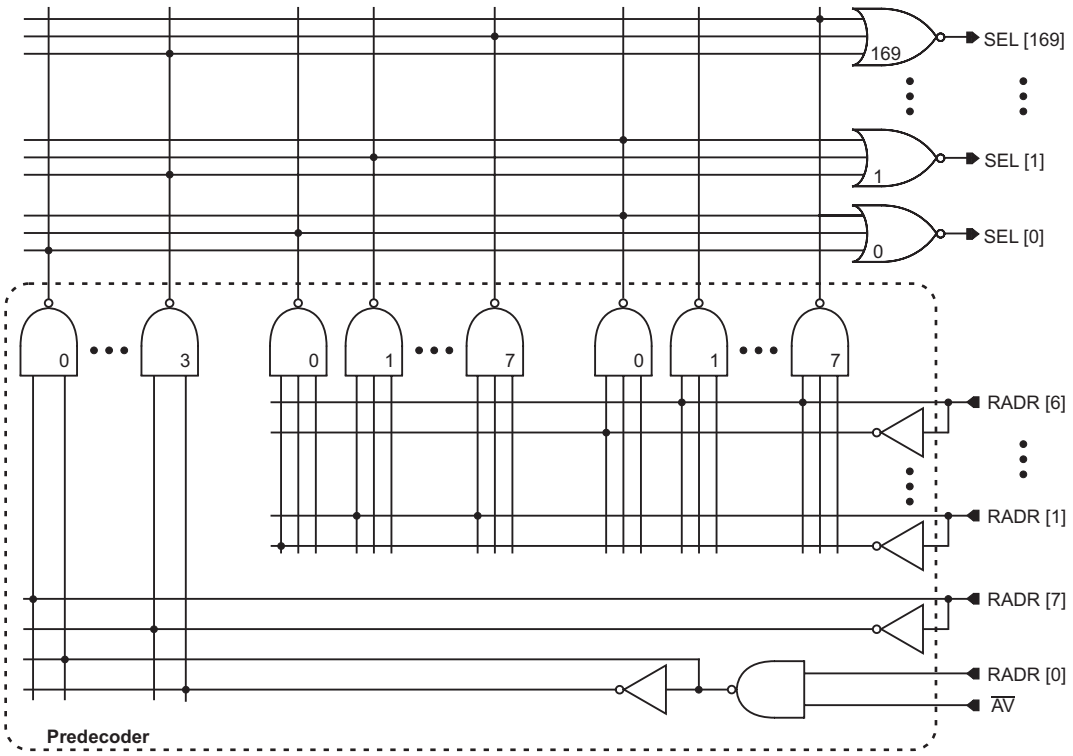


Figure 4.7: Circuit diagram to explain the basic function of the address decoder used for the row selection in the sensor array and the SRAM. The predecoder shown in the lower part is shared by both arrays. The final decoder above is for the sensor array. It is shown without the following selection logic (cf. figure 4.8).

The final decoder is responsible for the row selection. Each of its cells has to fit into the cell pitch of the corresponding array. A single decoder cell consists of a three-input NOR gate and a buffer to drive the corresponding select line. Each input of the NOR gate is connected to one signal of the three predecoded sets to perform the final decoding.

The implementation of the buffer is different for the SRAM and for the pixel array. In case of the SRAM it consists of two successive inverters. For the pixel array on the other side, the circuitry controlled by the selection signal is shown in figure 4.8. The selection of the row is done in the lower branch, it depends on the state of the signal ROWSELEBL (row selection enable). A lumped capacity of $\approx 350\,\mathrm{fF}$ has to be driven by the signal ROWSEL. The selection voltage V_{sel}, which is used to connect the pixels of the row to their read-out lines, can be adjusted externally. For most of the measurements in chapter 5 is was set to 2.7 V. In the upper branch, the reset signal $\overline{\mathrm{RES}}$ is generated in dependence of the signal RESVEBL.

In figure 4.7 only the connection of the predecoder to the final decoder of the pixel array is outlined. The SRAM is comprised by only 128 rows and therefore exhibits a somewhat different

connection scheme. As already mentioned it must be guaranteed that every row within the HDR window is assigned to a different row in the memory. In order to accomplish this, only the lower 7 address bits (RADR [0:6]) are used for the SRAM. Since an HDR window consists of 128 successive rows, each will be assigned to a particular SRAM row depending on the specific position of the window.

However, if averaging of neighboring pixels is applied, the HDR window is expanded over the whole sensor array. Depending on the used averaging method (2×2, 4×4 or 8×8), the integration time control is only applied to every 2nd, 4th or 8th sensor row. In order to realize the required unequivocal assignment of rows in this case, the LSB[6] of the row address is disabled for the pixel array ($\overline{\text{AV}}$=0 in figure 4.7). On the other hand, nothing changes for the SRAM: The LSB is directly led to its final decoder, so that $\overline{\text{AV}}$ does not affect the decoding here. Owing to the fact that a change of the LSB now solely leads to a change of the selected SRAM row, it can be used to avoid that the same row of memory cells is assigned to two different sensor array rows (namely no. x and no. x+128). The correct setting of RADR [0:7] is done by the FPGA which is used to control the imager.

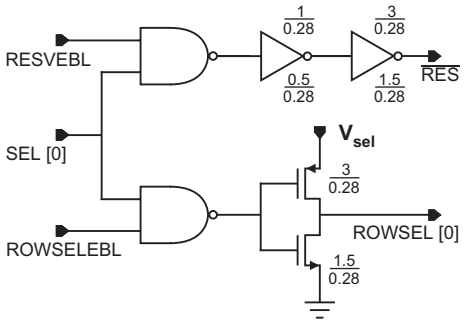


Figure 4.8: Circuit diagram of the selection logic for a single pixel row.

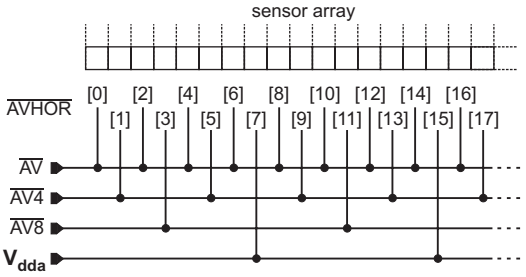


Figure 4.9: Connection diagram for the control of averaging. The same diagram is valid for the vertical connections.

4.2.3 Averaging of Neighboring Pixels

In many applications it is desirable to be able to process images at a reduced resolution on account of limited available computational power. The possibility of averaging neighboring pixels can be used here to get a first impression of the image structure, for a coarse feature extraction and to find regions of interest, which will be read out at a higher resolution in the next step.

In principle, the integration time control can be accomplished either before or after the pixel values are averaged. In the first case, a time control for every individual pixel is required, whereas in the second case only a single integration time for a cluster of averaged pixels values has to be regulated. Every integration time that has to be regulated requires a certain amount of memory for the storage of the time stamp. As a result, prior averaging is favorable unless the size of the implemented memory is sufficient for the entire array.

In general, there are two different possibilities to accomplish an averaging of pixel values. The first consists of an averaging of the corresponding photocurrents, whereas the second consists of an averaging of the pixel output voltages V_{pix}. While the first solution requires additional in-pixel switches to connect the neighboring photodiodes with each other, an averaging of the output voltages could be achieved by simply connecting the corresponding output lines with each other [LOO99-1].

[6] Least Significant Bit

However, the second method is not suitable if adaptive integration time control is applied, as will be shown in the following.

In case of voltage averaging, a regulation on the basis of the averaged output signal ($\overline{V_{pix}}$) is not possible. The value of $\overline{V_{pix}}$ does not indicate whether individual pixels run the risk of saturation or are even already saturated. Therefore, the final value of $\overline{V_{pix}}$ could be wrong, since it is not proportional to the average of the incident light intensities (including the recorded time stamp). The problem could be solved by a regulation of T_{int} for each pixel on the basis of the individual output voltages, but as mentioned above, this would lead to an increased demand for memory. Additionally, the final determination of $\overline{V_{pix}}$ would require the inclusion of the individual time stamps, which in turn requires an additional circuitry for the calculation of this value.

In contrast, if current averaging is applied, the output voltages of all pixels within the averaged cluster are identical and are proportional to the average of the incident light intensities (including the time stamp) at the pixel locations. Therefore a regulation on the basis of a single output signal V_{pix} for a cluster of pixels is possible. Memory for just a single time stamp per cluster has to be provided. A further processing for the calculation of the average is not necessary.

On account of these advantages, current averaging was chosen for the implementation of the sensor. The hereby reduced need for memory can be used for an expansion of the HDR region over the entire pixel array. The implemented memory of 64 kbit is sufficient for the regulation of all pixel-clusters if an averaging of 2×2 (or more) is applied.

As mentioned above, the current averaging is achieved by electrically connecting the corresponding photodiodes with each other. The result simply corresponds to a larger photodiode, whose area is the sum of the connected pixel diodes. There are two switches per pixel (M_5 and M_6, see figure 4.2) for the connection to the neighboring pixels lying above and to the right of the respective pixel. The activation of the signals $\overline{\text{AVHOR}}$ or $\overline{\text{AVVER}}$ results in a connection of the adjacent photodiodes of two neighboring rows or columns with each other. All signal lines ($\overline{\text{AVHOR}}$[0:169] and $\overline{\text{AVVER}}$[0:169]) are controlled by the external signals AV, AV4, and AV8. They are connected with each other in a way that permits the averaging of 2×2, 4×4 or 8×8 pixels ([AV, AV4, AV8]=[1,0,0],[1,1,0] or [1,1,1]). The diagram for the connection of the first lines of $\overline{\text{AVHOR}}$[0:169] is shown in figure 4.9. The same connection scheme is also used in the vertical direction for $\overline{\text{AVVER}}$[0:169].

The sample images shown in section 5.3.4 were taken with different averaging modes. They illustrate the interplay of integration time regulation and averaging.

4.3 ITC Entity

This entity is responsible for the adaptive integration time control. As discussed in chapter 3 the regulation is based upon a periodic check if individual pixels run the risk of being overexposed and therefore have to be reset. For a reset of an individual pixel, two conditions have to be fulfilled, which have to be verified by the ITC entity: First, the pixel output voltage V_{pix} must have decreased below a predefined level V_{comp}. Second, the pixel must have been reset in the previous curtain. A verification of the second condition requires the preceding storage of a mark which indicates the last reset curtain where a reset has been executed. This mark is referred to as the time stamp. The required storage implies the existence of a memory, which must be updated periodically by the entity. The stored data have to be accessible to the outer world since they finally permit the calculation of the integration time that has been used for a particular pixel.

Another task of the entity is the mapping of the data from the HDR region to the local memory and vice versa. The integration time control should work independently of the selected position of

the region. Additionally it should permit an expansion of the high dynamic region over the whole sensor array if averaging of neighboring pixels is applied.

In the beginning of the next subsection the architecture of the ITC entity is presented. The interplay of the individual components is described here. The details of these components are discussed in the subsections following this description.

4.3.1 Architecture of the ITC Entity

The architecture of the ITC entity can be explained by considering figure 4.10. As mentioned in 4.1.1, in principle the whole entity is able to process a HDR window with a width of 128 pixels in a sensor array with a width of 256 pixels. Therefore, most of the internal data busses exhibit a width of a multiple of 128. There are provided 256 outputs $\overline{\text{RESSEL}}$, but in reality only 170 of them are connected to the sensor array, corresponding to the pixel outputs delivered from the array.

Taking the sensor array as the starting point, the first components of the entity are 256 comparators. They are connected to the 170 pixel outputs V_{pix} in a pattern where every 5th and 6th comparator is left out. The purpose of the comparators is to verify the first reset condition, i.e. to test for each pixel whether $V_{pix} < V_{comp}$. The result of this comparison (COMP1 [0:255]) is led to the next component where the pixels that have to be reset are selected. This selection can only be made if the results from the verification of the second reset condition (COMP2 [0:255]) are also available.

As mentioned above, the test of the second reset condition requires the recall of the time stamps that indicate when a particular pixel was last reset. These time stamps are stored in a 512 $\times$128 bit SRAM that can be found at the bottom of figure 4.10. For the time stamp of each pixel a memory size of 4 bit is reserved. Each row of 512 bit corresponds to the 128 pixels of the selected row in the HDR window. Therefore all time stamps stored for a selected row can be recalled by the selection of a single row of the SRAM. Within a read access the signals BIT [0:511] and $\overline{\text{BIT}}$ [0:511] from the selected row are led to 512 sense amplifiers which transform them to a digital output. This output in turn is led to the component *stamp comparison* where the second reset condition is verified.

The test of this condition is done by a comparison of the time stamps delivered by the SRAM with the stamp indicating the previous reset curtain applied to the selected row. This stamp is delivered by the bus OLDSTP [0:3]. Of course, it is the same for all pixels of this row. The result of the comparison is lead via the *window control* to the *reset selection* component.

Which pixels of the selected row must be reset is decided within the *reset selection*. If both reset conditions are fulfilled for a particular pixel, the corresponding bit in $\overline{\text{RESSEL}}$ [0:169] and ALLRESSEL [0:255] is set. The only difference between the two busses is that $\overline{\text{RESSEL}}$ [0:169] consists of only those (inverted) signals of ALLRESSEL [0:255], which are really connected with the sensor array. The bus ALLRESSEL [0:255] is necessary for the update of the SRAM with the new time stamps. It is connected via the *window control* with the component *stamp write*. Here, the new time stamps are written to the SRAM. The information which stamps have to be updated is given by the indication of the pixels that are reset. The new time stamp itself is delivered by NEWSTP [0:3], which is the same for all stamps that are updated. In the case of an ordinary operation of the imager it is always NEWSTP=OLDSTP+1.

A component which was left aside up to this point is the *window control*. As mentioned above, the results from *stamp comparison* WINCOMP [0:127] are connected to the *window control*, where they are expanded to the bus COMP2 [0:255], which is lead to the *reset selection*. The incoming data which belong to the HDR window are mapped by the *window control* to the correct position of the window within the array width of 256 pixels. In the reverse case, the bus ALLRESSEL [0:255], which includes the whole width of the array, is reduced to the signals which belong to the window. This is

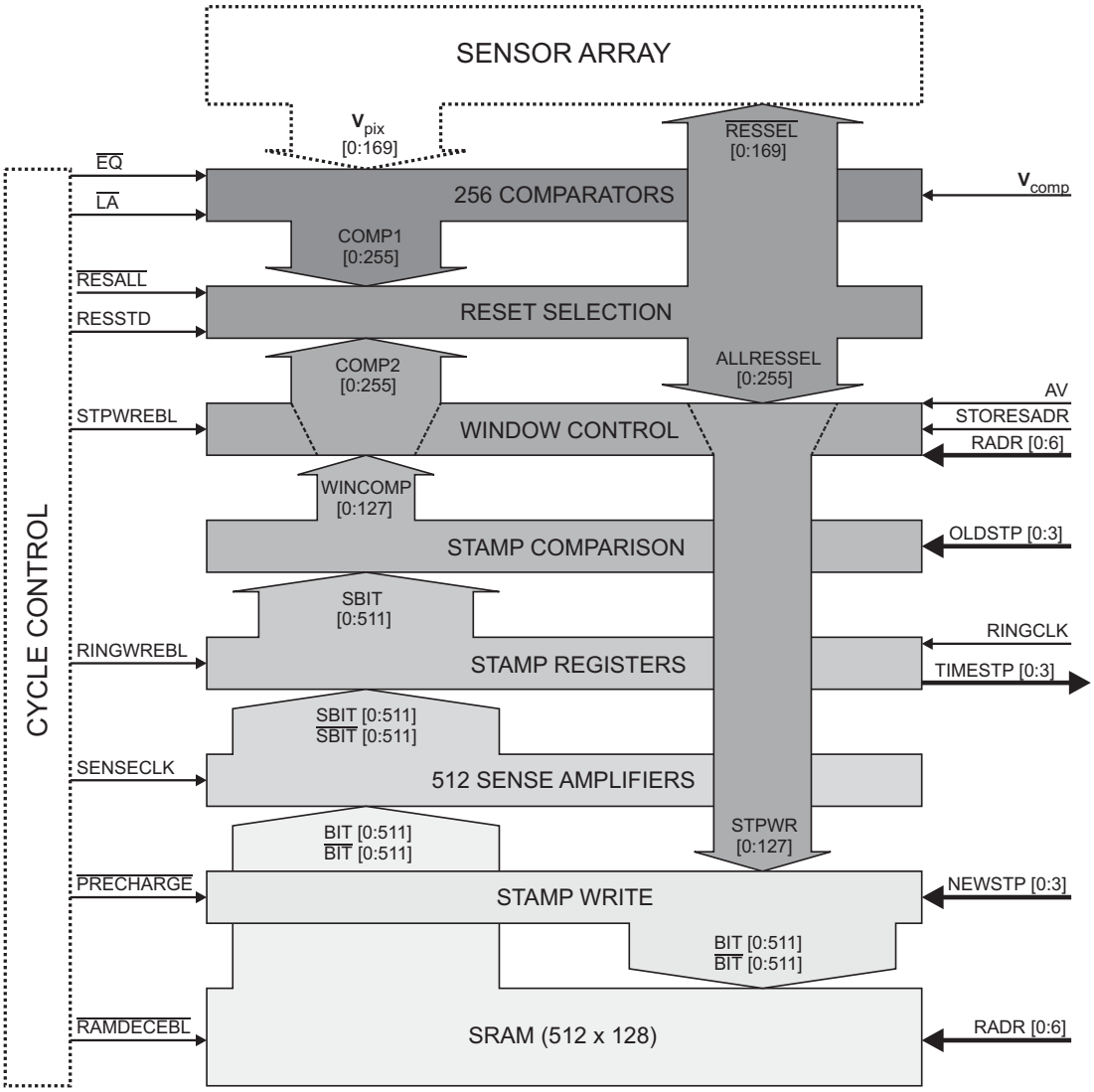


Figure 4.10: Overview of the architecture and the internal dataflow of the ITC entity. The signals on the right side lead to the I/O pads of the chip (compare with figure 4.1).

done since these are the only signals that have to (and can be) stored in the SRAM.

The last remaining component is *stamp registers*. Here, the data read from the SRAM is made accessible to the outer world. The data is copied to 512 shift registers, which can be read out serially via the bus TIMESTP [0:3].

In the following subsections the components introduced here, are presented in more detail.

4.3.2 Comparator Design

A comparator is a circuit that compares one analog signal with another and outputs the result of that comparison as a binary signal. The signals which have to be compared here are the pixel output voltage V_{pix} and the externally applied reference voltage V_{comp}. The result of the comparison represents the verification of the first reset condition (cf. chapter 3). Since the comparison has to be done for each pixel of a selected row and the results should be available simultaneously, 170 comparators are

required (respectively 256, the width the ITC entity was originally designed for, see section 4.1.1).

The requirements for the comparators are as follows: One comparison has to be accomplished per ITC cycle. Since the cycles do not follow one after another (cf. chapter 3) there will be periods of different lengths during which the comparators are not used. Within a single cycle the comparison has to be finished in a few 100 ns (including the sampling time). Due to the high amount of comparators, which increases linearly with the array size, a low power design is favorable. The sensitivity of the comparator, i.e. the minimum input-voltage difference needed to produce the correct binary output, should be comparable (or smaller) to the pixel FPN at the column lines they are connected to. This would guarantee that the influence of the comparators on the accuracy of the time regulation is of minor importance (cf. section 5.2.6). Therefore, a medium accuracy of $\leq 10\,\mathrm{mV}$ would be sufficient here.

There are many possible architectures for the realization of a comparator. They can be divided into three different types: open-loop architectures, regenerative architectures and combinations of them [ALL02-2]. While open-loop comparators basically can be regarded as operational amplifiers without frequency compensation, the regenerative type uses positive feedback to accomplish the comparison.

The requirements mentioned above can be fulfilled with a simple regenerative comparator, the switched CMOS latch. A grey region in figure 4.11 indicates the most important components. The CMOS latch itself consists of the transistors M_3-M_6. It has two modes of operation: The first mode (SAMPLE=1) disables the positive feedback and applies the input signal to the latch. The second mode (SAMPLE=0) closes the sample switches M_7, M_8 and enables the latch by connecting it to the power supplies (disregard M_1 and M_{10}). An asymmetry introduced by the input signal will be amplified by the positive feedback until the output nodes n1, n2 reach the binary values corresponding to the sign of the initial input difference. Compared with pure NMOS (or PMOS) latches and with open-loop comparators, the CMOS latch has the inherent advantage of low power dissipation. In both modes no static current flows through the latch, only during the short transition phase power is consumed. It is therefore appropriate for applications with longer suspend phases.

Unfortunately, the CMOS latch has two disadvantages which have to be taken into account: First, the generation of kickback noise at its inputs and secondly the limitation in sensitivity due to its input-offset voltage.

Kickback noise denominates the disturbance of the input levels when the sample switches M_7, M_8 are opened [RAZ92]. The disturbance is generated by the latch which has to adapt from its last state to the new input voltages. Unless the preceding circuit (the in-pixel source follower) has a low output impedance, it may take a long time to recover the input levels. The influence of the kickback could be virtually eliminated by the introduction of additional components (M_1, M_{10}, M_{14},M_{18}) to the latch, which can be seen in figure 4.13. M_{14} and M_{15} are used to equalize the latch with the reference voltage V_{comp} prior to a connection to the signal lines. Hereby, the influence of the kickback is minimized: The closer the input signal to V_{comp}, the more critical is the influence of noise to the decision of the comparator, but the smaller is the amount of kickback. Additionally, now the direction of the feedback is independent of the preceding state of the latch. Transistors M_1 and M_{10} are used to separate the latch from the power supplies during equalization. M_{11}-M_{13} and M_{16}-M_{18} make up the output inverters of the comparator. M_{12} and M_{17} are necessary to prevent a static current flow during the equalization. The control signals EQUAL, LATCH and SAMPLE are generated from the input signals $\overline{\mathrm{EQ}}$ and $\overline{\mathrm{LA}}$. The corresponding timing diagram can be seen in figure 4.12. The signal sampling is started at $t = 0\,\mathrm{ns}$. The equalization is switched off prior to the opening of the sample switches. After 300 ns the sampling ends and the latch gets activated. The result is hold for another 200 ns after which the equalization is restarted (suspend mode).

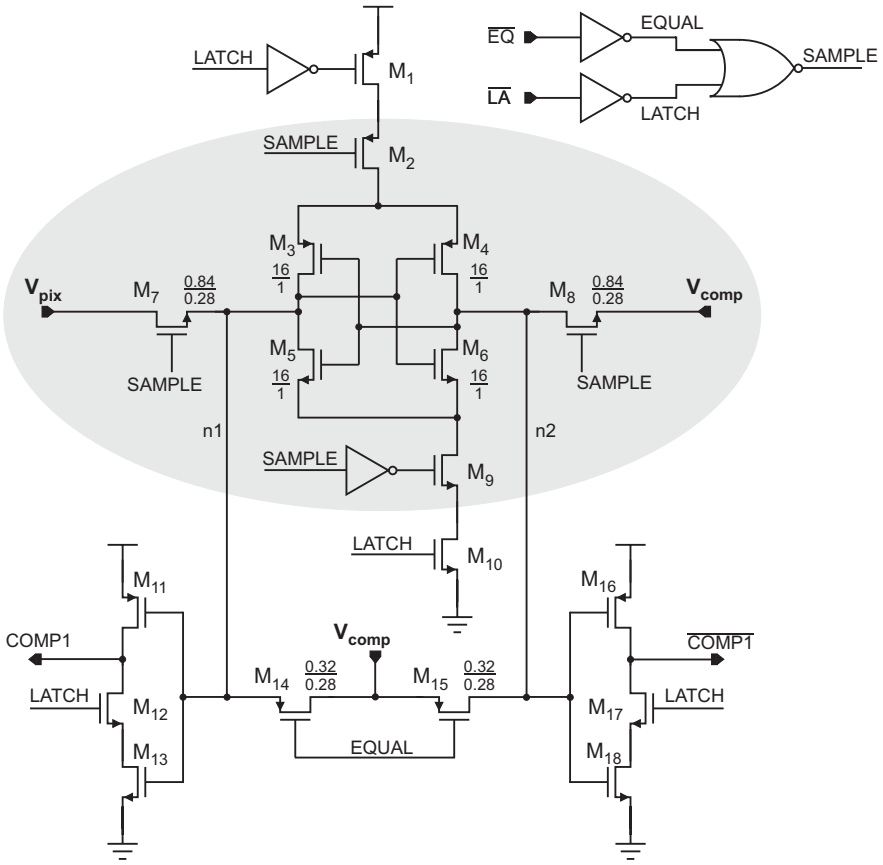


Figure 4.11: Circuit diagram of a single comparator. The decisive components of the dynamic latch are enclosed by the grey region.

As mentioned above, the second disadvantage of the latch is its limitation in sensitivity, which arises from the in principle unavoidable device-to-device mismatch (cf. section 1.4). This mismatch leads to an input-offset voltage which in turn limits the sensitivity of the comparator. The input-offset voltage V_{OS} itself would not be disturbing since the reference voltage V_{comp} can be chosen freely. But due to the fact that V_{OS} is different for each comparator and its value is not predictable, it is a decisive quantity. As will be shown in the following, proper device sizing and layout can be used to reduce V_{OS} and hereby reach the required sensitivity.

In order to avoid the introduction of offsets, the left side of the latch obviously has to match the right side, i.e. M_3 has to match M_4 and M_5 M_6 respectively. As described in section 1.4 the mismatches of the threshold voltage V_T and the transconductance β are proportional to $(W_{eff}\, L_{eff})^{-1}$ and can therefore be reduced by increasing W·L. Since devices with a length close to L_{min} exhibit a strongly increased mismatch [LOV98], a drawn length of L=1 μm was chosen. As can be seen in the circuit diagram, the chosen W/L ratios of the n- and p-channel devices are equal. Owing to the higher mobility in the n-channel devices, their transconductance will be higher than that of the p-channel devices (approximately by a factor of 2.5). The mismatch of the latch is therefore dominated by the mismatch of the two NMOS transistors [SAR91]. In view of this fact a coarse estimation of the sensitivity can be made by the application of an empirical formula for the calculation of the

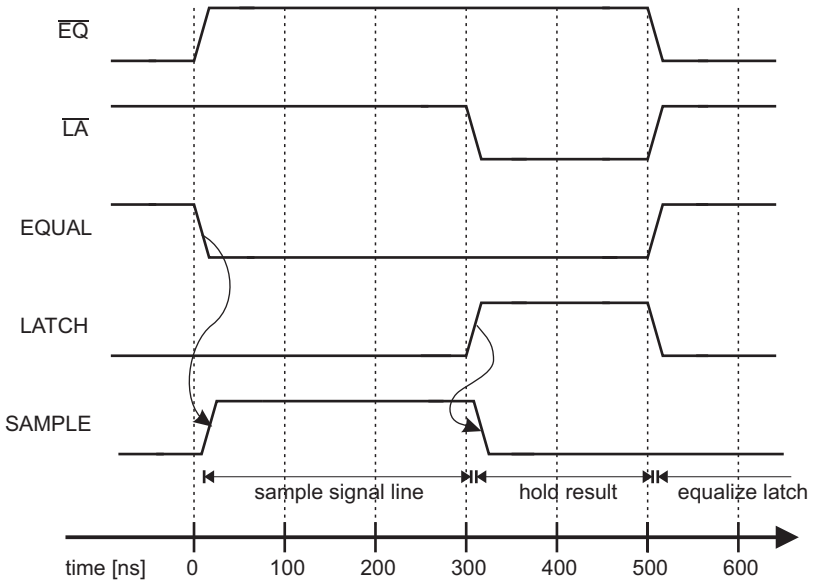


Figure 4.12: Timing diagram of the comparator control.

sensitivity S of a NMOS latch postulated in [IED78] and examined in [SAR91]:

$$\begin{aligned} S &= 2\,\delta V_T + \sqrt{\frac{m C_L}{\beta}} \cdot \frac{\delta\beta}{\beta} \\ &= 1.8\,\mathrm{mV} + \sqrt{\frac{1.8 \cdot 10^9\,\mathrm{V/s} \cdot 95 \cdot 10^{-15}\,\mathrm{F}}{0.95 \cdot 10^{-3}\,\mathrm{A/V^2}}} \cdot 0.0046 \;=\; 2.8\,\mathrm{mV} \end{aligned} \tag{4.2}$$

where m is the latching rate obtained by simulation and C_L is the load capacity, which is here nearly equal to the gate capacity of one NMOS transistor. The values for the threshold voltage mismatch δV_T and the transconductance mismatch $\delta\beta$ can be calculated from the process parameters. The original expression for S includes an additional term for the mismatch of C_L. This term has been omitted here, due to the fact that the CMOS latch is much less sensitive to mismatch in C_L than the NMOS latch. The reason for this is that changes in the voltages at node n1 or n2 are relatively indifferent to the size of C_L if charge is added to and removed from the node at the same time [SAR91].

Equation 4.2 is a quite conservative estimation since the individual sources of mismatch were not regarded as independent contributions (no quadratic addition of the errors). However, for an estimation of the sensitivity the charge injection of the switches M_7 and M_8 also has to be considered. The injected charge can lead to an additional error voltage of a few mV depending on the mismatch of the switches. In section 5.2.6 the results from the measurements of the comparator offset-distribution will be presented. A standard deviation of 2 mV could be observed.

Layout of the Comparator

In order to reach minimum mismatch a proper layout of the critical components is of high importance. Figure 4.13 contains the layout of a single comparator. Compared to figure 4.10 it is rotated by 90° for a compact presentation. The meaning of the individual layers is explained at the bottom of the figure.

Metal3, the top metal layer, is not shown here to enhance the recognizability of the other layers. It is used for power routing and for shielding sensitive circuits against incident light. The whole comparator is covered by it, except of two small slits (spacing=0.6 μm) located over *metal2*. Examining the layout it has to be taken into account that the production process permits to stack *contact*, *via1*, and *via2* which therefore could be covered by each other.

The complete layout of the 256 comparators consists of a repetition of the shown layout for 128 times in y-direction and two times in x-direction. Communication lines which connect the sensor array with the remaining components of the ITC entity run from left to right between the comparators. The most critical transistors are designed in a common-centroid geometry ABBA / BAAB, where each transistor is divided in 4 parts indicated by the corresponding letter. This pattern leads to an improved matching compared to the simpler AB / BA geometry [ISH94]. Due to the small size of the sample switches, the corresponding transistors M_7, M_8 are arranged in a simpler ABBA pattern with identical gate orientation. Dummy structures at the sides of each row of comparators provide a similar surrounding for the border devices.

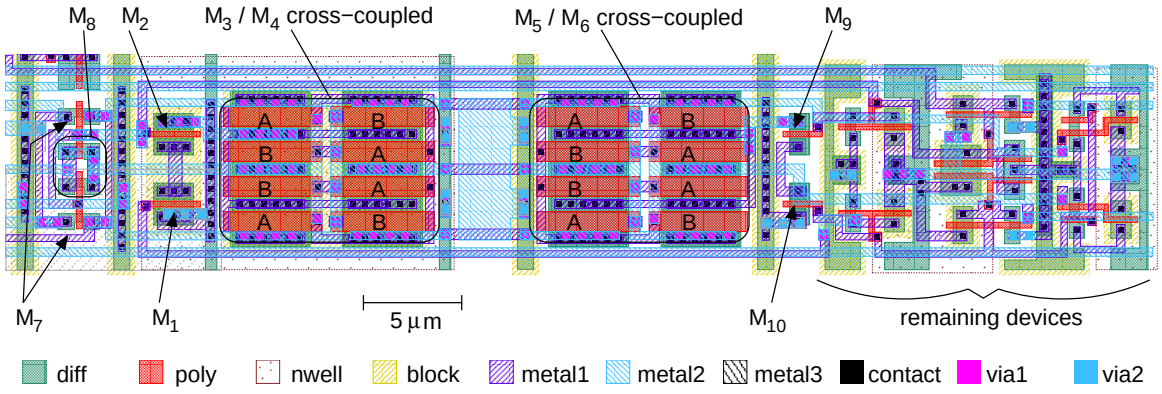


Figure 4.13: Layout of a single comparator. The size of the shown region is 11 μm $\times$ 58 μm. The layer *metal3* is not shown here.

4.3.3 Stamp Comparison

As mentioned above, this component is responsible for the verification of the second reset condition (cf. chapter 3), which is accomplished by a comparison of stamps. The time stamps that indicate when the pixels of the selected row have been reset the last time are read from the SRAM. Since a stamp consists of 4 bit per pixel and the HDR window has a width of 128 pixels the input has a width of 512 bit. Each of this stamps has to be compared with the stamp that indicates the previous reset curtain applied to the selected row. As described in chapter 3 this stamp is delivered by the FPGA via the I/O OLDSTP[0:3].

Figure 4.14 contains the circuit diagram of the component. A circuit to compare the four bits of a single time stamp with OLDSTP[0:3] consists of four XNOR gates whose inputs are connected to the corresponding bits. The outputs of these gates are lead to a single AND which produces the result of the comparison. As the output of a XNOR is only high if the two input values are identical, WINCOMP is only high if the stamps have the same value. This circuit is repeated 128 times to cover all inputs and produce the outputs WINCOMP[0:127].

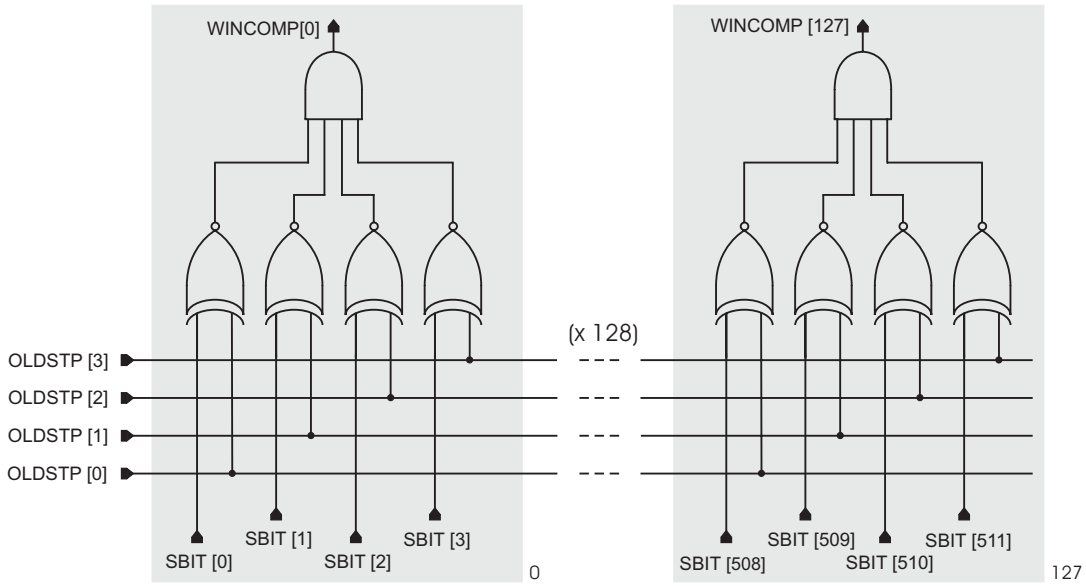


Figure 4.14: Circuitry for the comparison of time stamps that were read from the SRAM with those of the previous ITC cycle executed in the selected row.

4.3.4 Reset Selection

This component is used to select the pixels which have to be reset. The decision for a reset depends on the fulfillment of both reset conditions evaluated for each pixel. Since the outputs of the *reset selection* are directly connected with the sensor array, it has a width of 256 elements. The results from the two reset conditions are received via the two incoming busses COMP1 [0:255] and COMP2 [0:255]. In principle, a single NAND gate for each pixel would be sufficient to produce the desired output, i.e. the reset selections. These outputs would be directly dependent upon changes in the SRAM output or the comparator output. As can be seen in the timing diagrams of section 4.5 it would be desirable to decouple the selection lines from these components. This is done by adding a reset-set flip-flop (RS-FF) as shown in figure 4.15. Now the results can be stored in the meantime and the outputs ALLRESSEL [0:255] will only change synchronously to the signals RESSTD (for reset selected) and $\overline{\text{RESALL}}$ (for reset all). The signal $\overline{\text{RESALL}}$ is necessary to be enable reset of an entire row, independently of the fulfillment of the reset conditions.

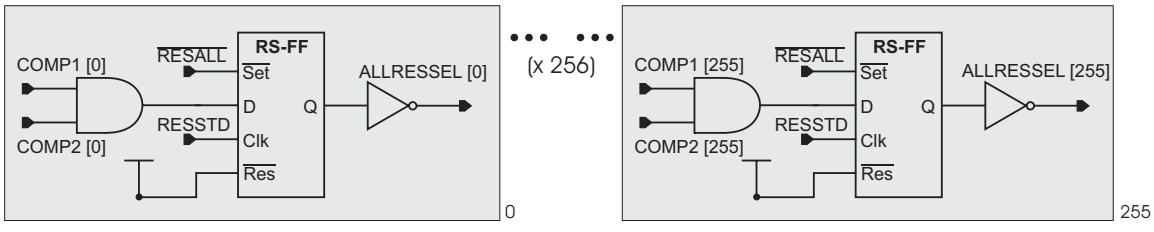


Figure 4.15: Circuitry for the selection of the pixels which have to be reset within the currently selected row.

As mentioned above, the connection to the sensor array is the bus $\overline{\text{RESSEL}}$ [0:169] which consists of only a subset of the (inverted) signals from ALLRESSEL [0:255]. It should be kept in mind, that the selection for a reset, which is done here, is not identical to the execution of a reset. As can be seen

in the circuit diagram of the pixel (figure 4.2) a low $\overline{\text{RESSEL}}$ signal only connects the reset transistor M_2 to the reset line $\overline{\text{RES}}$. The reset itself is caused by a change of $\overline{\text{RES}}$.

4.3.5 Window Control

In chapter 3 the ability of the image sensor to adaptively regulate the integration time in a selected region was introduced. This region, which is designated the HDR window, has a size of 85×85 pixels, i.e. one quarter of the sensor array. Its position can be chosen freely by the user within the array. In order to reach a time regulation within the intended region, ITC cycles are executed solely in the rows which belong to the window (fixing the y-position). This is controlled by the FPGA (cf. section 3.5).

The realization of the windows x-position is accomplished by the application of the time control solely to those pixels lying within the chosen x-position. The restriction of the time control to these pixels is done by the *window control*. It is realized by a selective pass of those lines to the SRAM, which carry information belonging to the columns lying within the x-range of the window. These lines transport the information if a particular pixel within the window is going to be reset or not (STPWR[0:127]). On the other hand, it has to be guaranteed that the data which is read back from the SRAM is used to execute a potential reset at the correct position. Additionally, a correct mapping must be reached if averaging of adjacent pixels is used: In this case, the HDR window should be expandable over the whole sensor area.

Figure 4.16 illustrates the basic operation of the window control. If no averaging of pixels is performed, the window control works similar to a barrel shifter. First, the window position is stored in a register. It can be changed with the begin of a new frame. The address lines RADR[0:6] coming from the register are connected to a predecoder similar to the one in figure 4.2. Its outputs are led to two final decoders, each for one half of the sensor array. The left one manages the first 128 lines from 0 to 127, the right one selects the next 128 lines in reverse order, i.e. decreasing from left to right. For a specific address the two corresponding selection lines, one in each half of the array, are pulled high. The following switches are implemented in a way that they are activated as if they are controlled by a thermometer code: All switches with a number lying above the selected one are closed. This way, only the signals which belong to the window are connected to closed switches.

Nevertheless there are two remaining difficulties: First, the amount of lines leaving the window control has to be reduced to those which belong to the window. Second, the pitch of these lines, which is determined by the pixel pitch, has to be adapted to the pitch of the following components that in turn is a consequence of the SRAM cell pitch. With respect to the illustrated signal selection this can be reached easily by a specific connection scheme, suggested in the lower part of figure 4.16. The lines which belong to the left side of the sensor array are connected to those from the right side 1:1. Therefore, there are always two lines which lead to a common horizontal line. For any window position, there is always only one of these two lines which is really connected by a closed switch, namely that one which belongs to the window. On the other hand, every horizontal line is connected to a single vertical line, which leads to the SRAM. Owing to the fact that the position of the connection can be chosen freely in the horizontal direction, the realization of a different line pitch is possible.

The left picture in figure 4.17 contains the complete routing pattern for the example of 16 lines with *pitch1* which are connected to 8 lines of *pitch2*. This represents the simplest possible solution. As a consequence of the limited available routing layers all horizontal lines will be on the same layer. In conjunction with the adjacent lines, every line makes up a capacitor whose capacitance C is determined by the distance from these lines and the length of the common run. With a signal change,

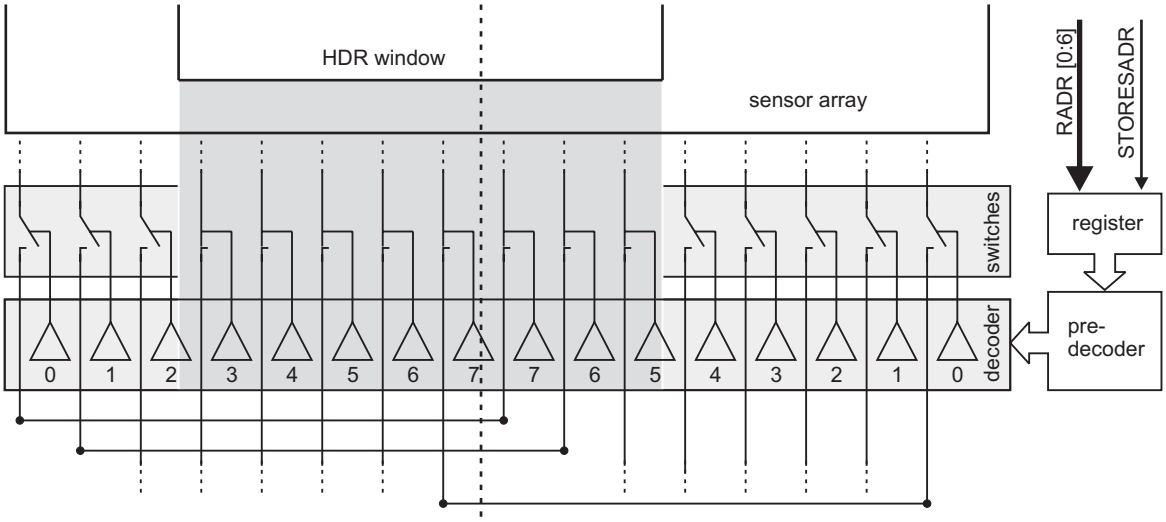


Figure 4.16: Illustration of the basic operation of the window control. For the sake of simplification the depicted sensor array consists of just 16 columns. The real circuit implementation is shown in figure 4.18. The connection shown the lower part are presented in greater detail in figure 4.17.

this capacitor is charged or discharged. The dynamic power dissipation P_{dyn} of a circuit is given by the relation

$$P_{dyn} = C \cdot V^2 f \tag{4.3}$$

where V is the change in potential and f is the frequency of the signal change [GEI90-2]. In order to reduce the power dissipation, C can be reduced by the choice of a geometrical distribution of the routing different from the one shown in the left part of figure 4.17. The routing pattern that was used is shown in the right part. The lines are shifted against each other in alternating order, which results in a reduction of the line capacitance by 30 %.

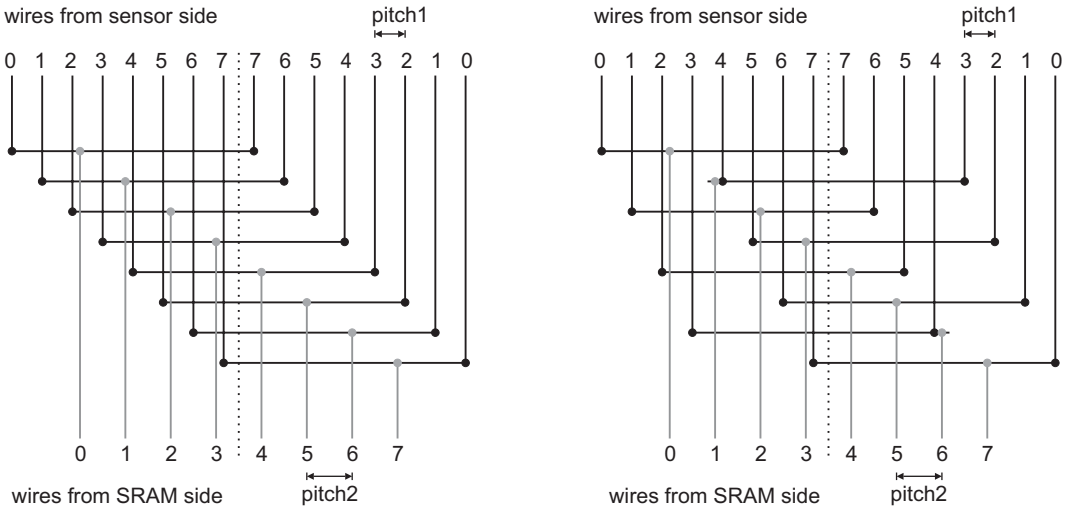


Figure 4.17: A simple routing pattern for the connection of two busses with different line pitch is shown on the left side. Two lines from the top bus are connected to a single line of the bottom bus. On the right side the realized pattern is shown. It provides an equivalent connectivity but with lower parasitic capacitance.

The implemented circuit to realize the basic operation scheme from figure 4.16 can be explained by considering figure 4.18. The input signals ALLRESSEL [0:255] with *pitch1* (from sensor array) are connected to the output signals STPWR [0:127] with *pitch2* (to SRAM). On the other hand, the input signals WINCOMP [0:127] with *pitch2* are connected to the output signals COMP2 [0:255] with *pitch1*. The outputs from the address decoder are indicated as SEL [0:63]. The box "connection distribution" stands for the routing scheme in figure 4.17 which has to be used for each of these two connections. In the upper half of the figure the circuit diagrams of the switches from figure 4.16 can be seen. Each grey box represents two switches and there are 64 of them for the first half of the sensor array (128 lines). The switches of the second half are not shown here. They have the same structure as the ones in the first half with the only difference that they are arranged in reverse order to reach an activation in the opposite direction.

The circuits/lines that are shown in the lower part of figure 4.18 are shared by both sides, i.e. each half of lines coming from the array is connected to the identical lines as explained by the routing scheme on the right side of figure 4.17.

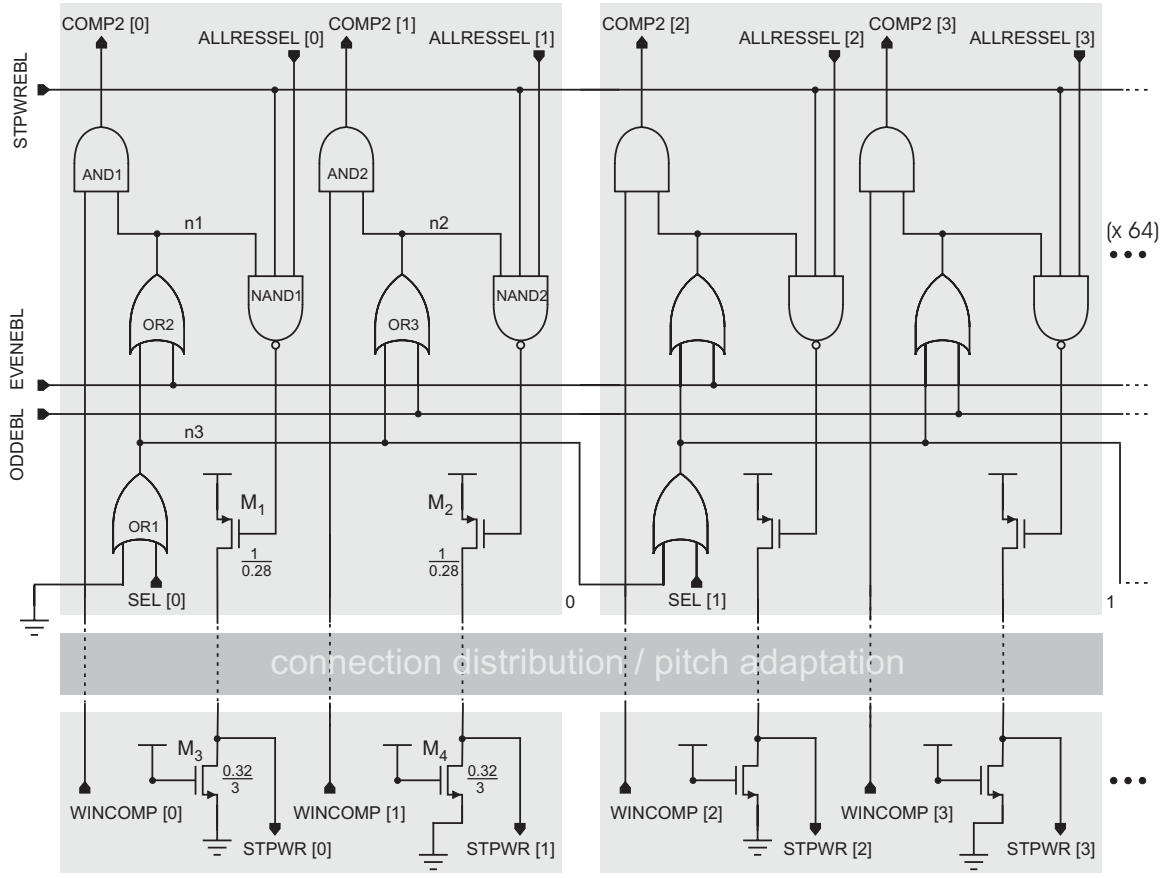


Figure 4.18: Circuit diagram of the implemented logic for the window control. The circuit enclosed by the grey box is repeated 64 times for the left half of the array. The right half consists of the same circuit blocks, but connected in reverse order as indicated in figure 4.16. The connection distribution can be found in figure 4.17.

The operation of the circuit can be explained as follows. Considering the connection from ALLRESSEL [0:1] to STPWR [0:1] the decisive components are NAND1, NAND2 and M_1-M_4. The transistors M_3 and M_4 are used as pull-down transistors for STPWR [0:1]. A low output of NAND1 and

NAND2 will result in a high STPWR[0:1] with a static current of 25 μA flowing through M_1,M_3 or M_2,M_4 respectively (for the short time of a high STPWREBL, cf. timing diagrams in section 4.5). The dimensions of M_3 and M_4 determine the magnitude of this current and therefore represent a trade-off between power consumption and rise time. Assuming that the signal STPWREBL (stamp write enable) is high and the nodes n1,n2 are also high, then STPWR will always represent the logic state of the corresponding ALLRESSEL. The same is valid for the connection from WINCOMP[0:127] to COMP2[0:255]: If node1 is high for instance, COMP2[0] will represent the logic state of WINCOMP[0]. This behavior is equivalent to a closed switch in figure 4.16.

The remaining gates OR1-3 are used for the selection of the window position, i.e. for the definition which of the switches are closed. The window position can be changed in steps of two pixels, therefore there is a new SEL input every second pixel. For example, if SEL[0] is high then node n3 is also high and as a consequence of this n1 and n2 too. The connection between each node n3 and an input of OR1 in the following grey box ensures that all switches lying above the selected one will also be closed. The activation signal propagates from switch to switch until it reaches the end of the chain, i.e. the half of the array width. At the same time, in the right half of the array the switches are closed one after another in reverse order. The maximum time needed to close all switches within the window region is the time that is necessary for the activation signal to propagate through all switches until it reaches the end of the chain (worst case: SEL[0] is high). A simulation of this propagation time resulted in 30 ns (worst case speed simulation).

Beside the possibility to define a HDR window it is also possible to use averaging of neighboring pixels for the expansion of the HDR region over the whole sensor array. In case of 2×2 averaging for example, the time stamp of every second pixel must be stored in the SRAM in order to be able to reconstruct the integration times for the whole sensor array. This can be reached by use of the control signal ODDEBL in figure 4.18 (the signal EVENEBL is only for test purposes). Again, the principle can be understood more easily by considering figure 4.16 first. If ODDEBL is high, all lines lying in the left half and having an odd index will be connected to the corresponding horizontal lines. As ODDEBL is in the right half connected to switches of the lines with an even index, here the lines having an even index will be connected to the corresponding lines. As can be seen in figure 4.16 this leads to a complete use of the following bus with *pitch2*, no information gets lost. The same method is also used for 4×4 and 8×8 averaging modes, the unnecessarily stored information (every second instead of every fourth line) is of no concern.

In the previous discussion about how the data is stored in case of an activated HDR window or averaging, it can be confusing that the order of the data at the output bus is different from the one at the input bus. For instance in case of a HDR window, the part of the window that lies on the right half of the input bus will appear on the left side of the output bus (cf. figure 4.16). It has to be kept in mind here, that the order how the data is stored is irrelevant since it is read back in the same way, i.e. it will be reordered and thus appear at the correct positions automatically. The only case where the order is of relevance is the storage of the stamps in the *stamp registers* (cf. section 4.3.9). Here, the registers are connected to form a ring, which permits the read-out of the stamp value synchronously to the corresponding analog value.

4.3.6 Static Random Access Memory

Chapter 3 describes how the time when a particular pixel has been reset, can be calculated from the exponent of the integration time regulation factor, which is referred to as the time stamp. In order to make the adaptive integration time control possible, it is necessary to provide a local memory, where these time stamps can be stored and read back in the course of regulation. This memory has been

realized as a static random access memory.

The circuit diagram of a single memory cell is depicted in figure 4.19. It consists of six transistors. Two cross-coupled inverters (M_1-M_4) serve as storage elements and two pass gates (M_5 and M_6) as a combination read/write port. In view of the fact that the memory cell will be repeated many times, it is important to minimize the cell area. The devices are typically dimensioned close to the process minimum. However, the size is restricted by the requirements for a stable cell as will be shown in the following. The architecture of the cell is horizontally symmetric. Therefore, the subsequent discussion of transistor sizes is always valid for both sides.

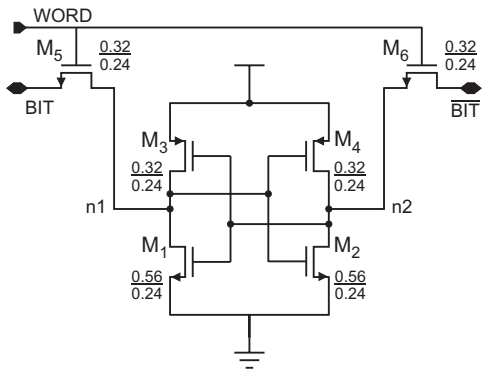


Figure 4.19: Circuit diagram of a SRAM cell.

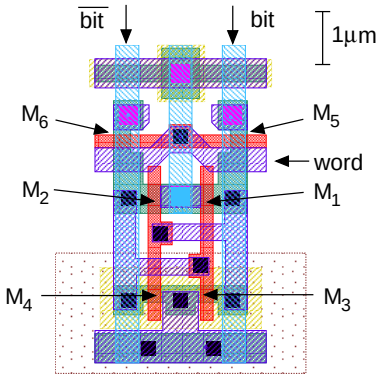


Figure 4.20: Layout of a SRAM cell. The occupied area is 15.5 μm^2.

Suppose the cell has stored a value so that node n1 is high and consequently node n2 is low. In a read access the WORD line is pulled high to connect the cell to the bit-lines by opening the pass transistors M_5 and M_6. The bit-lines were previously precharged to V_{dd} (cf. section 4.3.7). The SRAM cell will now start to discharge the bit-line $\overline{\text{BIT}}$ through the serially connected devices M_6 and M_2, whereas the line BIT will stay high on account of the activated device M_3. After a sufficient difference voltage has been developed between the bit-lines, the sense amplifier will be activated to amplify this difference. A first design constraint arises from the fact that M_6 and M_2 form a resistive divider between the input node $\overline{\text{BIT}}$ and the storage node n2. As a consequence of this, node n2 will rise somewhat after the cell is connected to the bit-lines. If it rises too high, it will cause M_1 to turn on and discharge node n1. The result of this could be a flip of the cell, i.e. the contents of the cell are changed within a read access ("read-upset"). In order to prevent a read-upset the so called cell ratio CR [CHA01], i.e. the ratio of the pull-down (M_2) to the pass gate (M_6) size , has to be chosen correctly.

$$\text{CR} = \frac{W_{M2}/L_{M2}}{W_{M6}/L_{M6}} \tag{4.4}$$

An increasing value of CR results in a rising conductance of the pull-down device relative to the pass gate and therefore decreases the risk of a read-upset. The exact value of the minimum CR necessary to prevent a read-upset depends on the supply voltage, the threshold voltage and parameter mismatch. It has to be found by simulation. Typical values for current fabrication technologies are CR_{min}=1.25-2.0 [CHA01]. For the implemented SRAM cell, a value of CR=1.75 has been used.

A similar constraint arises from the need to keep the memory cell writable: During a write access, one of the bit-lines, for example BIT, is pulled down to ground prior to a connection of the cell. Suppose the cell has stored the same value as in the read access described above (n1 high, n2 low). Now, M_5 and M_3 form a resistive divider between the input node BIT and the storage node n1. In order to write the cell, the conductance of M_5 has to be considerably lower than that of M_3 since

node n1 must be pulled down to a value low enough to flip the cell. This can be expressed by the pullup ratio PR:

$$\mathrm{PR} = \frac{W_{M3}/L_{M3}}{W_{M5}/L_{M5}} \tag{4.5}$$

The maximum pullup ratio for a still writable cell is reached, when the voltage at node n1 is small enough to flip the cell in a write access (approximately at V_T). In the realized cell it is set to PR=1.

Different design approaches are possible to minimize the cell area and simultaneously reach stability. For the presented cell design minimum size pass gates M_5, M_6 were chosen. The widths of the pull-down transistors M_1, M_2 have been increased until the cell ratio was sufficient. Finally, for the pull-up devices M_3, M_4 could also be chosen a minimum size, resulting in a reliably writable cell. The stability of the cell has been verified by simulations under worst case conditions of V_{dd}, V_T, and charge carrier mobility.

Figure 4.20 contains the layout of a single SRAM cell. It requires an area of $2.92\,\mu\mathrm{m} \times 5.32\,\mu\mathrm{m} \approx 15.5\,\mu\mathrm{m}^2$. The assignment of the individual layers is explained at the bottom of figure 4.5 for instance. The top metal layer is not shown here, it covers the whole cell and is used for the routing of the power supply. A single row of the SRAM consists of 512 horizontally aligned cells. 128 of these rows make up the whole array of 65536 cells. In order to minimize the used area, every second row is mirrored so that two consecutive rows share their nwells or substrate contacts respectively. There are 4 bit reserved for every time stamp and a single SRAM row contains the stamps for the 128 pixels in a single row of the HDR window. No redundancy has been implemented in the current SRAM design. Implementation of redundancy to improve silicon productivity is recommended by IBM[7] only for chips with a SRAM content above 1 Mbit.

The horizontally running WORD lines are selected by the final decoders sitting on one side of the cell array. They consist of NOR gates similar to that of the decoders for the pixel array presented in section 4.2.2. The WORD lines are driven by two inverters following the NOR gates with the same driving strength as for the $\overline{\mathrm{RES}}$ line in figure 4.8. Their position is indicated on the chip photograph at the end of the chapter (figure 4.38). The vertically running bit-lines (parasitic capacitance: 240 fF) are connected to the sense amplifiers sitting between the SRAM and the rest of the ITC entity.

4.3.7 Stamp Write

The *stamp write* component has two different tasks: On the one hand, it has to perform the write accesses to the SRAM. On the other hand, it has to provide the possibility to precharge the bit-lines of the memory. The precharging is necessary to guarantee an identical initial state for all bit-lines at the time before the corresponding memory cells are connected to them for read-out. Hereby, a fast read-out of the cell content is possible since the developing voltage difference between the bit-lines of each cell shows the correct sign from the beginning.

Figure 4.21 contains the circuit diagram of the component. Every circuit that is shown in a single grey box is connected to the BIT / $\overline{\mathrm{BIT}}$ lines of a SRAM column. Therefore, if a particular row of the SRAM is selected, every circuit is connected to a single RAM[8] cell. As every stamp consists of four bits this circuit is repeated four times with the inputs NEWSTP [0:3] and the write enable STPWR which is the same for these four bits. The four circuits in turn are repeated 128 times to cover the whole input of 128 stamp write enable signals. The stamp that would be written in case of a positive stamp write enable is always the same (NEWSTP [0:3]). It was shown in section 4.3.1 that the signals STPWR [0:127] simply indicate, which pixels will be reset within the current reset curtain. A high

[7] International Business Machines Corporation

[8] Random Access Memory

STPWR opens the pass gates M_1/ M_2 and therefore enables the update of RAM cells containing this stamp. In case of a low STPWR the corresponding RAM cells would just keep their contents.

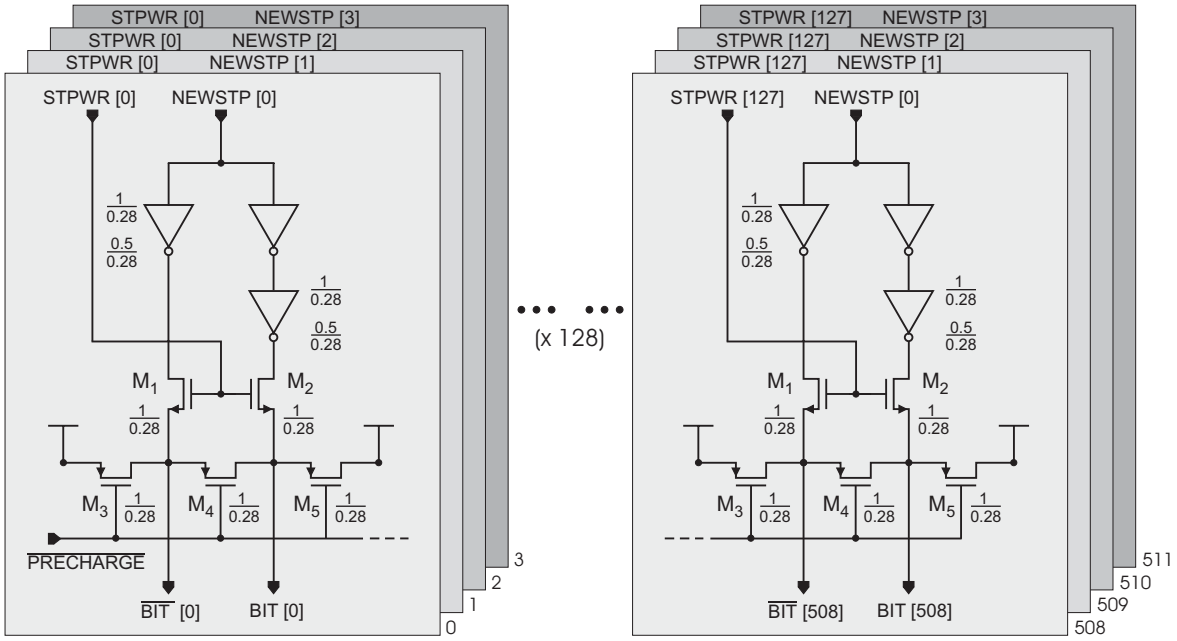


Figure 4.21: Circuit diagram of the SRAM write datapath and the logic to precharge the bit-lines.

During a write cycle the signal $\overline{\text{PRECHARGE}}$ is always high and therefore the pass gates M_3-M_5 are in a non-conductive state. In order to precharge the bit-lines, $\overline{\text{PRECHARGE}}$ is pulled low and hereby connects the bit-lines to the supply voltage. M_4 just accelerates the process of equalizing two bit-lines.

4.3.8 Sense Amplifier Design

Sense amplifiers for memories are circuits which are used for the detection of a small differential input signal to determine the contents of a connected memory cell. In principle they can be regarded as a special kind of comparators. Owing to the fact that the chosen architecture is based on a CMOS latch, it is similar to that of the comparator in section 4.3.2. Nevertheless, the requirements are quite different.

In a single SRAM read access 512 bit are selected and have to be amplified in parallel, which requires 512 sense amplifiers. The sensitivity needed for the comparison depends on the voltage difference ΔV that the memory cell is able to produce between the bit-lines. ΔV on the other hand is dependent upon the available time for a read access, the bit-line capacity and the source/sink capability of the memory cell. For the given memory implementation and requirements for speed the sensitivity is uncritical. Nevertheless, the design should be easily scalable to higher resolutions of the pixel array or/and the HDR window. This would result in a larger memory size and possibly require multiple read/write accesses per ITC cycle or higher ITCCLK frequencies. However, the obtainable sensitivity is limited by the offset distribution of the sense amplifier. A sufficient margin for future implementations can be reached easily by the choice of non-minimum size devices and a proper layout as has been shown for the comparator in section 4.3.2.

Figure 4.22 contains the circuit diagram of the realized sense amplifier. The architecture has been used in a number of Alpha microprocessors [CHA01]. M_5 and M_6 are connected to the bit-lines coming from the selected memory cell. They are used to sense the input voltages. The latch, made up of M_1-M_4, amplifies the detected voltage differential. M_7 is used to activate the latch by providing a current path to ground, whereas M_8 and M_9 precharge the latch. Two inverters are used to buffer the output.

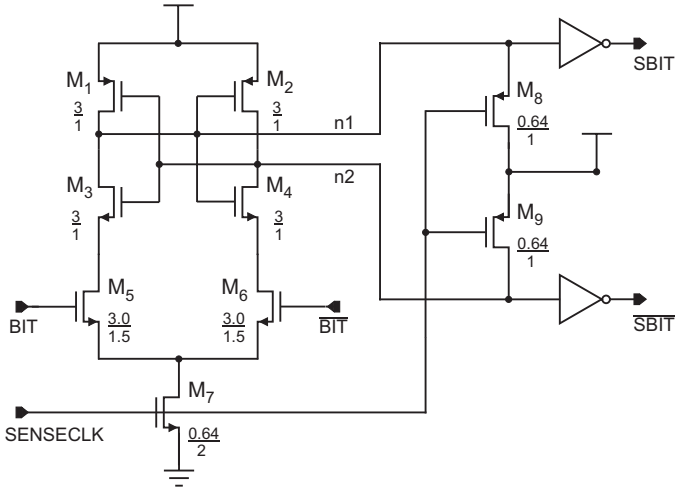


Figure 4.22: Circuit diagram of the sense amplifier.

The sense amplifier operate as follows. There are two phases of operation. In the first phase (SENSECLK=0) nodes n1 and n2 are precharged to V_{dd}. Prior to a connection of the memory cell to the bit-lines, the bit-lines themselves were also precharged to V_{dd} as described in section 4.3.7. When the read access begins, the memory cell starts to discharge one of the bit-lines. This will lead to a slightly lower impedance on the corresponding side of the sense amplifier. After a sufficient voltage difference ($\Delta V > 50\,\mathrm{mV}$) has been build up at the inputs, the sense amplifier is activated by pulling SENSECLK high (phase 2). The lower impedance now leads to a faster discharge of the corresponding node n1 or n2. The asymmetry hereby introduced will be amplified by the positive feedback of the latch up to a voltage difference of $\mid V_{dd} \mid$ between n1 and n2 unless it is stopped by a pull-down of SENSECLK before.

The layout of the sense amplifier is shown in figure 4.23. Compared to figure 4.10 the drawing is rotated by 90° for a compact presentation. *Metal3*, the top metal layer, is not shown here to enhance the recognizability of the other layers. The most sensitive components are completely covered by it to protect them against incident light. The presented layout block is repeated 256 times in y-direction and two times in x-direction summing up to the required 512 sense amplifiers. Communication lines running from left to right connect the inputs of the amplifiers to the SRAM array and their outputs to the stamp registers. Pairs of transistors which are critical with respect to mismatch are designed in a common-centroid geometry AB / BA. Hereby, most of the mismatch caused by process parameter gradients is eliminated by averaging. At the top and at the bottom of figure 4.23 parts of the adjacent sense amplifiers can be observed. The vertical distance of two adjacent transistors belonging to different pairs of cross-coupled transistors is identical with the distance of two transistors within a single pair. Therefore they serve as dummy structures for each other.

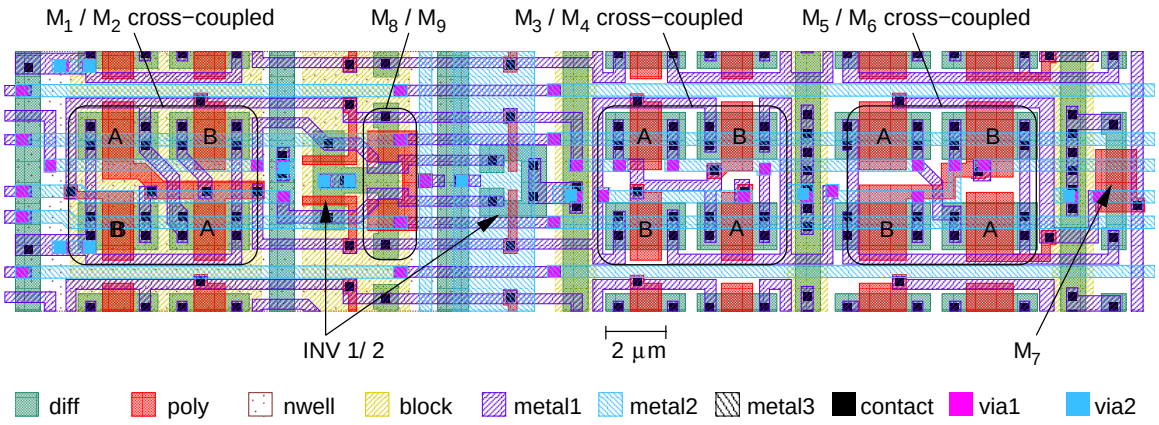


Figure 4.23: Layout of a single sense amplifier. The size of the shown region is $8.5\,\mu\text{m} \times 37\,\mu\text{m}$. The layer *metal3* is not shown here.

4.3.9 Stamp Registers

The *stamp registers* make the time stamps accessible to the outer world. This is necessary since they indicate the integration time of each pixel when they are read out after the application of the last reset curtain to the selected row. A single time stamp has a length of 4 bit. The stamps from one row of pixels comprise 512 bit, which are read from the SRAM in a single read access. In order to be able to read them out serially without blocking up the SRAM in the meantime, the values are stored in registers. The circuit structure of the component can be explained by considering figure 4.24.

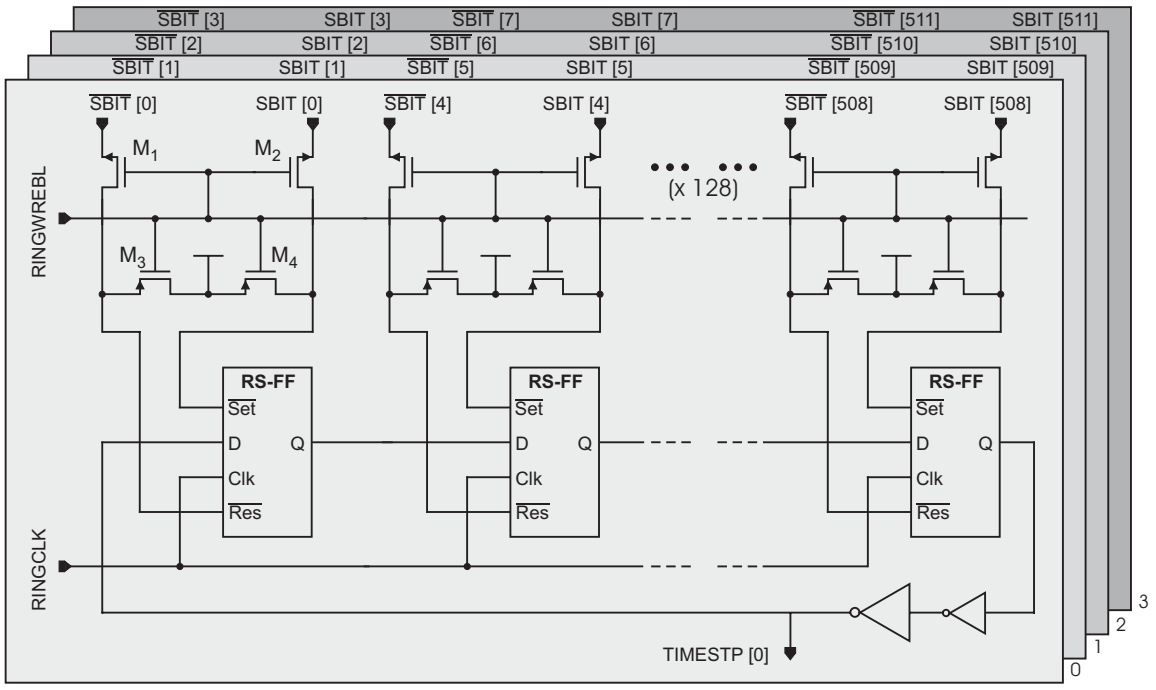


Figure 4.24: Circuit diagram to explain the architecture of the component *stamp registers*. In the upper part the write enable logic can be seen, whereas the lower part shows the shift registers to store the corresponding time stamps.

Each of the grey boxes corresponds to the circuitry that is used to store one of the four stamp-bits from all 128 stamps. Therefore, for every stamp-bit there are 128 flip-flops that are connected with each other to form a shift register, which is controlled by the clock RINGCLK. This way it is possible to read out all stamps serially while the four bits of a single stamp appear in parallel on the bus TIMESTP[0:3].

Owing to the fact that the update of the flip-flops has to be controllable, each of them is connected to a write-enable-circuit consisting of four transistors (M_1-M_4). When RINGWREBL is pulled high, the pass gates M_1 and M_2 are opened and the flip-flop is updated. On the other hand, a low RINGWREBL connects the inputs to the power supply, which leaves the contents unchanged.

The registers are connected with each other in such a way that form four rings, i.e. their outputs are connected with the first inputs. This allows the read-out of a time stamp synchronously to the corresponding analog value of the pixel.

4.4 Double-Sampling Entity

The double-sampling entity is responsible for the analog data acquisition and the serial read-out of the results from integration and from reset of the pixels. The way this has to be done is influenced by the integration time control. A consequence of the time control described in chapter 3 is that all pixels of a certain row reach the end of the integration at the same time. Therefore the entire row is ready to be read out. Subsequently, this row is connected to the column read-out lines, which lead to the comparators on the one side and to the DS entity on the other side. The read-out of the pixel voltages V_{pix} now has to be done quickly since the column lines are blocked for other operations in the meantime. Especially no ITC cycles can be executed to regulate the integration time in other rows. Additionally, due to the fact that the pixels do not stop integrating the pixel voltages will change with time. Therefore the pixel values of a row are copied in parallel to a bank of capacitors. Afterwards, these values are read out serially for the off-chip data acquisition. In order to reach the necessary precision with a reasonable effort, the time that is available for the serial read-out of a single row should be as large as possible. Here, the available time is identical to T_{row}, the time between the selection of two adjacent rows for read-out (cf. chapter 3).

4.4.1 Architecture of the DS Entity

The architecture of the DS entity can be explained by considering figure 4.25. The circuitry for the read-out of a single pixel is enclosed by a grey box. It is repeated 170 times to be able to read out the pixels of a whole row in parallel. For each pixel there are two values that have to be stored, the result from integration and the reset voltage. Hence, the output of each pixel is connected to two sample-and-hold stages. Each of the stages consists of a capacitor to store the analog value and a switch to sample the pixel voltage. The capacitance $C_h = 1\,\mathrm{pF}$ is made up of a PMOS transistor, whose source/drain diffusions are connected to V_{dd}. The switches S_1 and S_2 are realized as single MOS switches with compensation of charge injection. Details of their implementation can be found in the next subsection.

Each of the capacitors is connected to the input of a separate PMOS source follower (M_1-M_3 or M_4-M_6 respectively), which is used to drive one of the two output lines (gain g=0.74).

In view of the fact that the source followers are always used to read out a particular column of pixels, they will be referred to as column amplifiers in the following. They are designed for maximum linearity and can be adjusted by the bias voltage V_{bamp}, which is generated by an external DAC (cf. section 5.1.1). If an amplifier is not in use, it is separated by a CMOS transmission gate

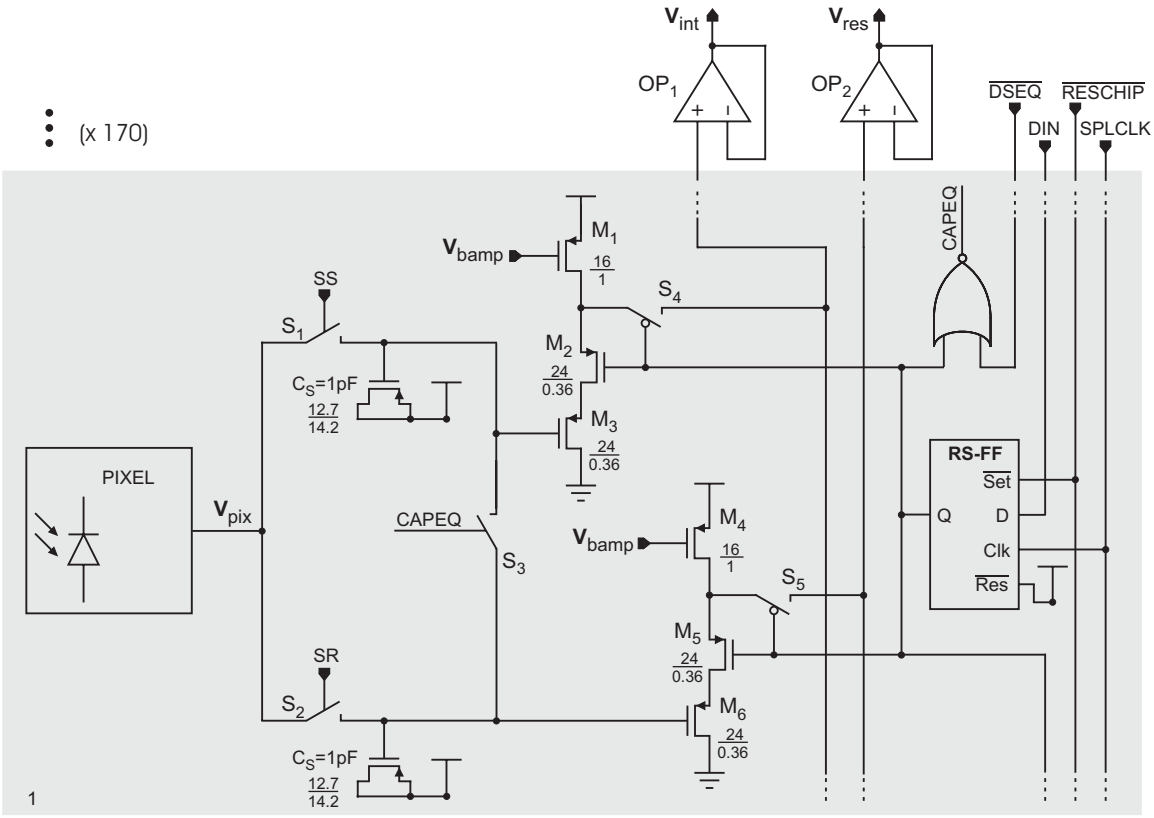


Figure 4.25: Circuit diagram of the DS entity. The circuit enclosed by the grey box is repeated for each column.

(two complementary transistors) from the corresponding output line (switches S_4 and S_5). That way, there is always only one column amplifier connected to each output line and the total capacitance of that line is not unnecessarily increased by the output capacitances of the other amplifiers. The signal which operates the switch is also used to enable/disable the amplifier, so that no static current is flowing if it is not in use. This signal is generated by a shift register, which can be found on the right side of figure 4.25. By the use of SPLCLK the sampled values can be read out one after another, while the integration value V_{int} and the reset value V_{res} of a particular pixel always appear synchronously on the two output lines. The output lines in turn lead to two operational amplifiers, which work as off-chip drivers. Their architecture is presented in section 4.4.3.

As a consequence of device-to-device mismatch, the column amplifiers will exhibit an offset distribution, which deteriorates the signals that are read out. Since V_{res} and V_{int} will finally be subtracted from each other to get the net signal for each pixel, only the offset difference of each pair of column amplifiers has to be known. If this difference was not subtracted from the net signal, it would result in a column-to-column FPN in the final image. In order to measure the offset difference, switch S_3 can be used to connect the two storage nodes with each other, which guarantees that the two column amplifiers have the same input voltage. An additional read-out of the two output lines then allows to correct the previously taken value of $V_{res} - V_{int}$ to get the final net value V_{sig}[9]:

$$V_{sig} = (V_{res} - V_{int}) - (V_{mres} - V_{mint}) \tag{4.6}$$

[9]This method is often referred to as difference double-sampling (DDS) [YAD04-2].

S_3 is realized by a single NMOS pass transistor. Owing to the fact that the stored voltages are limited to a range which can be covered by a single transistor, a CMOS transmission gate is not required here. S_3 is activated by the signal CAPEQ, which is generated in dependence of the shift register output and the signal DSEQ.

Off-chip, the signals V_{res} and V_{int} are connected to the inputs of an analog switch (cf. figure 5.1). Depending on the state of its activation signal SW the switch connects one of its inputs to the input of a 12-bit ADC.

By use of an additional CMOS transmission gate, V_{pix} can also be connected to special input lines (VTEST1 and VTEST2), which can be used to apply test voltages to the column lines. This is important for the test of the comparators and the DS entity. For the sake of simplification, these components, which are used only for test purposes, are omitted in figure 4.25.

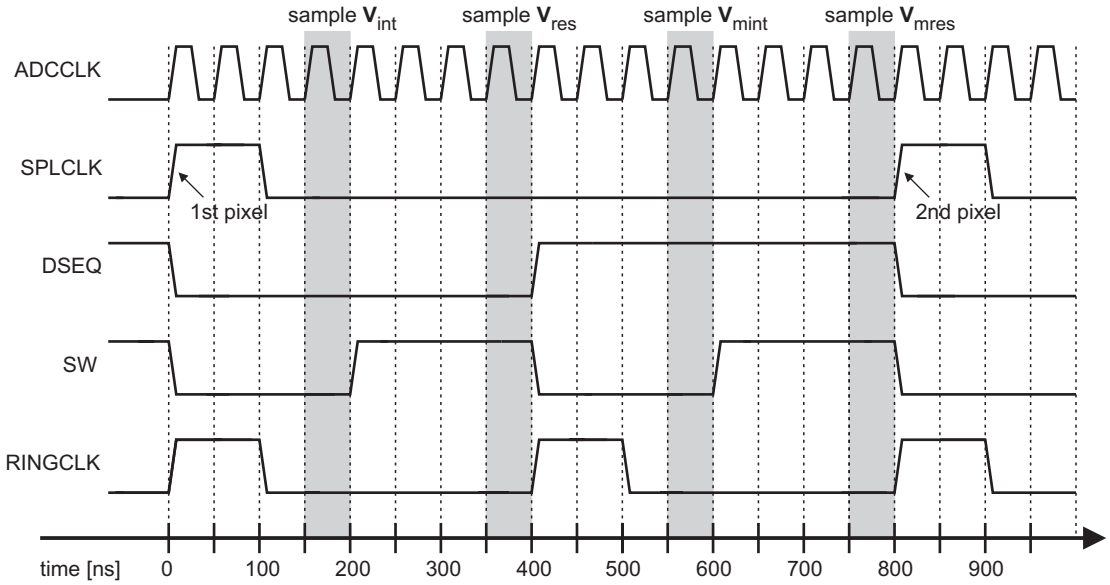


Figure 4.26: Timing diagram for the serial read-out of the analog values. The diagram applies to an integration time of 33 ms.

Prior to the serial read-out, the pixel voltages are sampled on the storage capacitors by the use of SS and SR as described in section 4.5.3. The first shift register is loaded with DIN=0. In the course of the serial read-out, this bit is shifted through the registers for connecting the column amplifiers to the read-out lines (active low). Figure 4.26 represents a detail of the timing diagram during the serial read-out. All signals are generated by the FPGA of the attached test system. The external ADC is operated with a clock frequency of 20 MHz. For each pixel four conversions have to be accomplished by the ADC, which are indicated in grey. The signal RINGCLK is used to read out the corresponding time stamps in parallel. It is connected to the *stamp registers* (cf. section 4.3.9). The presented timing diagram applies to an integration time of 33 ms. Longer integration times lead to higher values of T_{row} and therefore permit a smaller read-out speed. This is used for a 16-times oversampling in the implemented control schemes with T_{int} =157 ms or 524 ms respectively.

4.4.2 Analog MOS Switch

In the previous subsection it was shown that an elementary sample-and-hold circuit is used for the acquisition of the pixel output voltages. Such a circuit combines a sampling switch, which is realized

here as a single transistor MOS switch with a holding capacitor C_h. However, a single transistor is not an ideal switch. When it is turned off by a change of the gate voltage, i.e. the switch is opened, a certain amount of charge will be injected into the data holding node. This charge will lead to a deterioration of the sampled voltage V_{in} and therefore limits the accuracy of the circuit. In order to show how this effect can be minimized and to justify the chosen implementation the origin of the error charge and the influencing parameters will be explained.

The total charge Q_{tot} that is released at switch-off stems from the discharge of the gate-channel capacitance C_{gc} and the gate-diffusion overlap capacitances C_{ov} at the source (C_{gs}) and the drain (C_{gd}) of the switch ($C_{ov} = C_{gs} + C_{gd}$, see figure 4.27). The discharge takes place in two distinct phases. During the first phase, the gate-source voltage $V_G - V_{in}$ is higher than the threshold voltage V_T of the transistor. The channel of the switch conducts and C_{gd} as well as C_{ov} are discharged as consequence of the decreasing gate voltage V_G. When V_G reaches the value $V_T + V_{in}$ the channel is pinched off and the second phase starts. Owing to the fact that C_{gc} nearly disappears at this voltage, only C_{ov} contributes to the released charge in the course of a further decrease of V_G. The contributions of the two phases (Q_{ph1} and Q_{ph2}) sum up to a total released charge of

$$\begin{aligned} Q_{tot} &= Q_{ph1} + Q_{ph2} &&= (C_{gc} + C_{ov})(V_{GON} - V_T - V_{in}) + C_{ov}(V_{in} + V_T) \\ &&&= C_{gc}(V_{GON} - V_T - V_{in}) + C_{ov} V_{GON} \end{aligned} \tag{4.7}$$

where V_{GON} is the maximum gate voltage to close the switch ("on" state, here 2.5 V). Only a fraction of Q_{tot} is injected into the data holding node. This fraction Q_{inj} will cause a shift of the sampled voltage by ΔV:

$$\Delta V = \frac{Q_{inj}}{C_h} \tag{4.8}$$

In principle, the shift ΔV of the sampled voltage would not be a problem in case it would be the same shift for all sample-and-hold stages implemented in the DS entity. An existing voltage shift would be eliminated due to the double-sampling and the subsequent subtraction of the values from each other (see equation 4.6). Unfortunately, ΔV depends on the sampled voltage V_{in} (see equation 4.7), which is different for each stage in general. A further difference arises from the fact that the switch transistor will not have the same properties from stage to stage. As a consequence of process parameter variations the overlap capacitances will vary from switch to switch.

Obviously, there are two possibilities to reduce ΔV: Either by increasing the holding capacitance C_h or by decreasing the turn on voltage V_{GON} of the switch. Both can be realized only to a certain degree. An increase of C_h is limited by the available space and a decrease of V_{GON} would in turn limit the usable range of input voltages as it must be $V_{GON} > V_{in} + V_T$ to keep the switch conductive.

There are further possibilities to reduce ΔV by making use of the effects which determine the fraction of Q_{tot} that is injected into the holding data node. In principle the channel charge can exit the channel through the source, the drain, and the substrate electrodes. The injection into the substrate is only possible if the channel transit time t_{tr}, i.e. the longest time needed by mobile charges to reach one end of the channel, is longer than the switch-off time of V_G. For short-channel transistors, as used for MOS switches (here: L=0.36 μm), t_{tr} is so small that this effect can be neglected [WEG87]. Therefore, Q_{ph1} is shared between the source and the drain node of the switch. The exact partitioning of Q_{ph1} depends on the impedances at these nodes and the switch-off time t_{sw}.

Figure 4.27 shows a simplified circuit for charge injection analysis. The input of the sample-and-hold stage is connected here to a signal source with output resistance R_s. The lumped capacitance at the input node is C_s. If the switch-off time is longer than the input source time constant ($t_{sw} \gg R_s C_s$), the system is always in equilibrium. In this case, the charge injected by the switch will be

compensated by the source as long as the channel is conductive, i.e. as long as $V_G > V_{in} + V_T$. This way, most of the charge injection can be compensated. The disadvantage of this method is the need for long switching transients if $R_s C_s$ is large. This is the case for large pixel arrays (high C_s) with weak in-pixel source followers (high R_s).

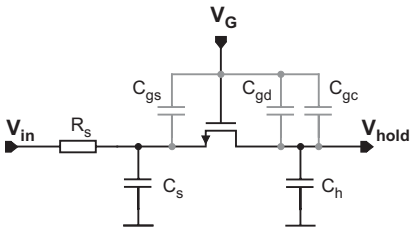


Figure 4.27: Circuit diagram for the analysis of charge injection.

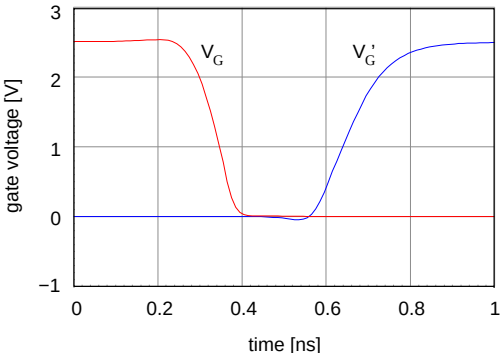


Figure 4.28: Simulation of the control signal slopes for the turn-off of the main switch (V_G) and the turn-on of the dummy switch (V'_G).

For switch-off times $t_{sw} \ll R_s C_s$, the charge partitioning can be described by a model, which finally permits the calculation of that fraction of Q_{ph1} that is injected into C_h. The exact results from the numerical solution of the corresponding differential equation can be found in [WEG87]. In essence it is shown that the partition is determined by the switching parameter B [SHI87], which is defined as

$$B = (V_{GON} - V_{in} - V_T) \cdot \sqrt{\left(\frac{\beta}{a \cdot C_h}\right)} \tag{4.9}$$

where β is the transconductance of the switch and $a = (V_{GON} - V_{in} - V_T)/t_{sw}$ is the slope of the gate voltage. The results show that for small values of B, i.e. for a small switch-off time t_{sw}, half of Q_{ph1} is injected into each node. Large values of B result in a charge partitioning proportional to C_h/C_s, whereas for intermediate values the injected fraction depends strongly on the exact value of B.

A considerable reduction of Q_{inj} could therefore be reached by choosing a large B value and decreasing the ratio C_h/C_s. However, this would require a long switching time and a costly adaptation of C_h/C_s (Owing to equation 4.8 C_s has to be kept large.).

Another way to reduce ΔV is the addition of a dummy switch. This method was finally chosen here for the realization of the required switch. In figure 4.29 the corresponding circuit diagram is shown. The dummy transistor M_2 that is connected to the data holding node is switched on when the main switch M_1 is switched off. The charge that is emitted by this additional transistor should compensate the charge injected by the main switch. In order to reach this, the size of the dummy transistor has to be adapted to the amount of charge it has to compensate. As mentioned above, this amount is dependent on the switch-off slope. In case of moderate switch-off slopes, choosing the size of the dummy switch is non-trivial and a compensation cannot be achieved over a range of V_{in} values [EIC89]. However, for fast switch-off slopes, where half of the total charge is injected, a dummy switch with half of the size of the main switch (W_{M2}=W_{M1}/2) would result in complete compensation. The injected dummy charge will exhibit the same dependencies described by equation 4.7 and therefore reach a compensation independently of V_{in}.

For the control of the implemented switch fast switch-off slopes were used and consequently a dummy with half the width of the main switch was inserted. In order to prevent part of the dummy charge from escaping through the still closed main switch, so that it would not contribute to the compensation, the dummy should be switched on with a delayed signal. The control signals for the switches are generated here with a chain of inverters. Figure 4.28 represents a simulation of the control signal slopes for a main switch that is turned off (dummy switch turned on). The resulting value of the switching parameter is $B \approx 0.2$ which guarantees a 1:1 partitioning of the charge injected by the main switch ($C_h/C_s \approx 3$).

The precision of the compensation depends on the matching of the main switch and the dummy switch. Figure 4.30 contains the layout of the implemented switch. The assignment of the individual layers is explained at the bottom of figure 4.5 for instance. The dummy transistor M_2 is surrounded by two halves of the main switch M_1 to reduce the influence of process parameter gradients. The additional dummy structures at each side have their gates and their source nodes connected to ground.

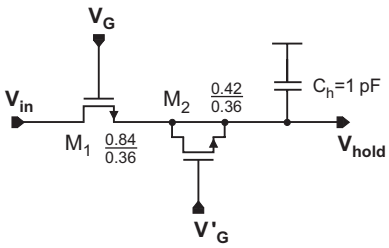


Figure 4.29: Circuit diagram of the implemented sample-and-hold stage.

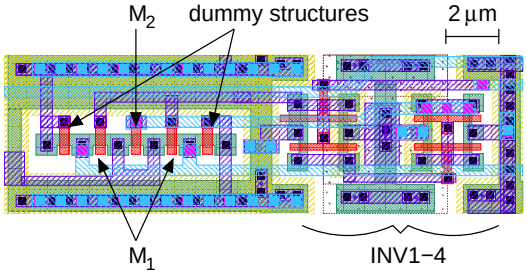


Figure 4.30: Layout of the sample switch including the inverter chain used to generate the necessary control signals.

According to literature [WEG87], with the dummy switch a reduction of the charge injection is possible by one or two orders of magnitude, limited by the process parameter variation. For the given implementation, the uncompensated injected charge amounts to $Q_{inj}=Q_{tot}/2=2.4\,\text{fC}$. Assuming a reduction by a factor of 10 this would lead to a voltage shift of $\Delta V=270\,\mu\text{V}$ on the holding capacitor C_h. Regarding the input voltage V_{in}, which ranges from about 400 mV to 1200 mV (cf. section 4.2.1), the introduced error would be $< 0.1\%$. Besides this error will be further reduced due to the double-sampling and double subtraction to calculate the final net signal voltage (cf. equation 4.6), which eliminates constant offsets.

4.4.3 Output Amplifier

The sampled voltages V_{res} and V_{int} have to be led out of the chip to be connected to the inputs of the external ADC or rather its adaptation circuitry (cf. section 5.1.1). The simple column amplifiers are not able to drive the load capacity consisting of the pad, a bond wire and the input capacitance of the following device. Therefore two additional output amplifiers are required for the two output lines.

The load that has to be driven is estimated to be a max. of 15 pF at a high resistance node ($>$ 1 MΩ). In order to reach a sufficient read-out speed and to take advantage of the full ADC resolution, it is required that the output of the amplifier settles within 150 ns to an accuracy of 12 bit. Assuming an unity-gain buffer, the necessary unity-gain bandwidth GB (in Hz) to settle within ± 0.5 LSB of a

N-bit ADC can be calculated by the equation [ALL02-5]

$$\mathrm{GB} = \frac{1}{\pi \cdot t_s} \ln\left(\frac{4}{\sqrt{3}}\ 2^N\right). \tag{4.10}$$

A settling time of $t_s = 150\,\mathrm{ns}$ therefore requires a GB of $\geq 19.5\,\mathrm{MHz}$ (with N=12).

The circuit diagram of the implemented output amplifier is depicted in figure 4.32. It is made up of a classical two-stage operational-transconductance amplifier with frequency compensation. The input stage consists of the differential pair M_3, M_4 and a current mirror made up of M_1 and M_2. M_3 and M_4 convert the differential input voltage into differential currents. These are amplified by M_1 and M_2 and converted back into a differential voltage. The current of the input stage is provided by transistor M_7. Its gate voltage is generated by a bias circuitry.

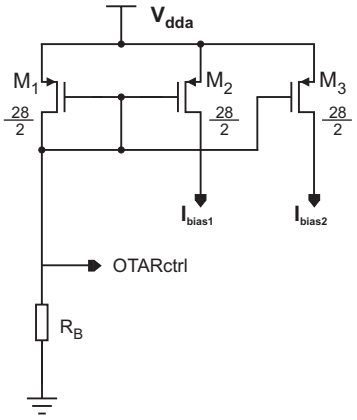


Figure 4.31: Circuitry for the generation of the bias currents.

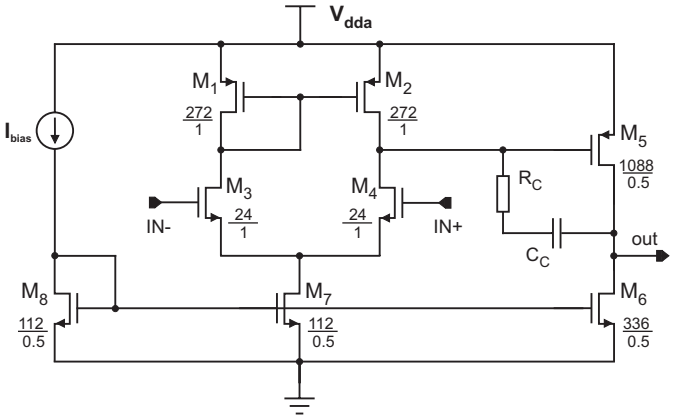


Figure 4.32: Circuit diagram of the frequency compensated two-stage output amplifier.

The output stage consists of a current-sink inverter formed by M_5 and M_6. M_5 converts the second stage input voltage into current. It is loaded by the current-sink load M_6 which converts the current into voltage at the output. In order to increase the stability of the amplifier, i.e. to increase the phase margin, the Miller compensation technique is applied: A capacitor C_c is connected from the output to the input of the second stage. It is realized as a metal-insulator-metal capacitor with a value of 3.3 pF. Compared to MOS capacitors, this type has the advantage of a linear charge-voltage relation over the whole voltage range. Additionally, there is no need for a minimum voltage between the capacitor plates to reach the full capacity. A disadvantage is the lower sheet capacitance, but since only two capacitors are needed, this is accepted here.

The resulting feedforward path through the compensation capacitor can lead to a limitation of the unity-gain bandwidth. One approach to eliminate this effect is to insert a resistor R_c ("nulling resistor") in series with C_c. R_c is implemented as a n^+-diffusion resistor and has a value of $620 \pm 70\Omega$. Due to the high sheet resistance of the diffusions (blocked silicidation, cf. section 4.6) they are especially suitable to build relatively small sized, medium to high ohmic resistors.

As in the previous circuit diagrams, the chosen channel dimensions for every transistor are written close to its identifier M. The lengths of transistors M_1-M_4 are chosen at $1\,\mu$m to make them less sensitive to mismatch. The individual pairs M_1,M_2 and M_3,M_4 are each drawn in a common-centroid geometry. A detailed description of the design procedure, which allows an estimation of the necessary W/L ratios, is given in [ALL02-3].

Figure 4.31 shows the circuitry for the generation of the bias currents for the two output amplifiers. It consists of two current mirrors and a p^+-diffusion resistor of $3\,\mathrm{k}\Omega$. A connection to the pad

OTARctrl provides the possibility to regulate the bias currents from outside the chip in order to adjust the speed and power consumption of the amplifier. This can be done by applying a bias voltage to OTARctrl. Without external regulation the bias current is adjusted to 300 μA, which results in a quiescent current of 900 μA in the output stage of the amplifier.

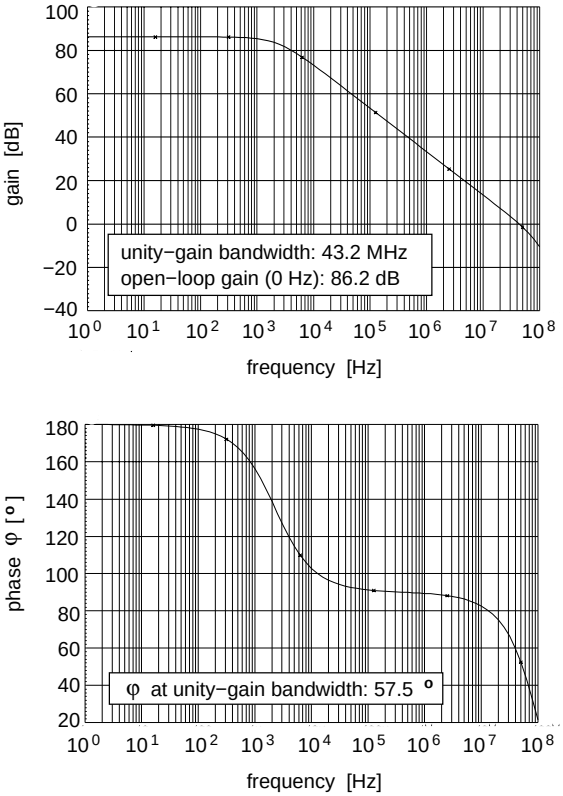


Figure 4.33: Frequency response and phase behaviour of the output amplifier, simulated with standard process parameters.

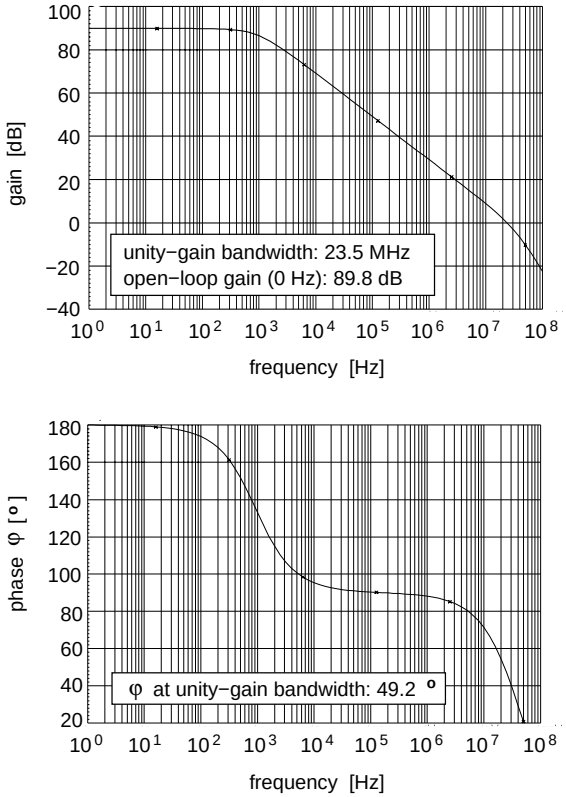


Figure 4.34: Frequency response and phase behaviour of the output amplifier, simulated for a worst case speed scenario ($\Delta\sigma = 1.5$).

The open-loop frequency characteristics of the amplifier have been verified by simulation. The results are shown in figures 4.33 and 4.34. No external regulation has been used here. All simulations have been accomplished with an output load of C_L=15 pF and R_L=1 MΩ. In the upper diagram, the frequency response (Bode-plot) can be seen. The dominant pole lying at 2.3 kHz (-3 dB point) causes a slope of -20 dB/decade. The point where this slope intersects the 0 dB axis is defined as the unity-gain bandwidth GB, which is here 43 MHz. In the lower diagram, the phase response can be seen. A measure of stability is given by the value of the phase at a frequency of GB, referred to as the phase margin. Phase margins of at least 45° are desirable to provide a minimum stability, with 60° preferable in most situations [ALL02-4]. As can be seen, the simulated phase margin amounts to 57°.

The simulations have been repeated for a worst case speed scenario. Process parameter that decisively influence the speed were deteriorated by 1.5 times the standard deviation of their expected distribution, each in the direction of a lower speed. Figure 4.34 contains the corresponding results. In the frequency response diagram the next higher order pole can be observed. The unity-gain bandwidth decreases down to 23 MHz, whereas the phase margin reaches a value of 49°. This shows that even under these conditions the amplifier would be sufficient for the intended application.

4.5 Cycle Control

The cycle control is responsible for the generation of the control signals needed for the adaptive integration time control and the data acquisition. As described before, two different sequences have to be generated, the ITC cycle and the DS cycle. The decision which cycle has to be executed next, the time when this has to be done and the row where it has to be applied to is determined by the external FPGA. The purpose of the cycle control is to guarantee a precise timing of the control signals, to simplify the control of the chip and at the same time to reduce the number of necessary I/O pads. After a short description of the architecture, the timings of the two cycles will be discussed. The detailed connection of the control signals can be taken from the preceding sections.

4.5.1 Architecture of the Cycle Control

The most important inputs of the cycle control are the signals DSCLK and ITCCLK, which are used for to the execution of a DS or an ITC cycle respectively. They are both operated at a speed of 10 MHz. A complete DS cycle consists of 16 successive clock cycles, whereas for an ITC cycle 8 clock cycles are necessary.

The control signals are generated from DSCLK and ITCCLK by a simple concatenation of RS-flip-flops. The circuit diagram for the generation of the ITC control signals is shown in figure 4.35. In the first part, a shift register (FF 1-8) generates a bit pattern 1111110 at its outputs, which is shifted around by use of ITCCLK. In the next, part these outputs are connected to further flip-flops. The way this is done determines the sequence of the control signals. Every control signal is generated by a single flip-flop, whose output is connected to a line driver. A similar circuit is used to generate the DS control signals. In total 36 flip-flops were used for the implementation of the complete cycle control.

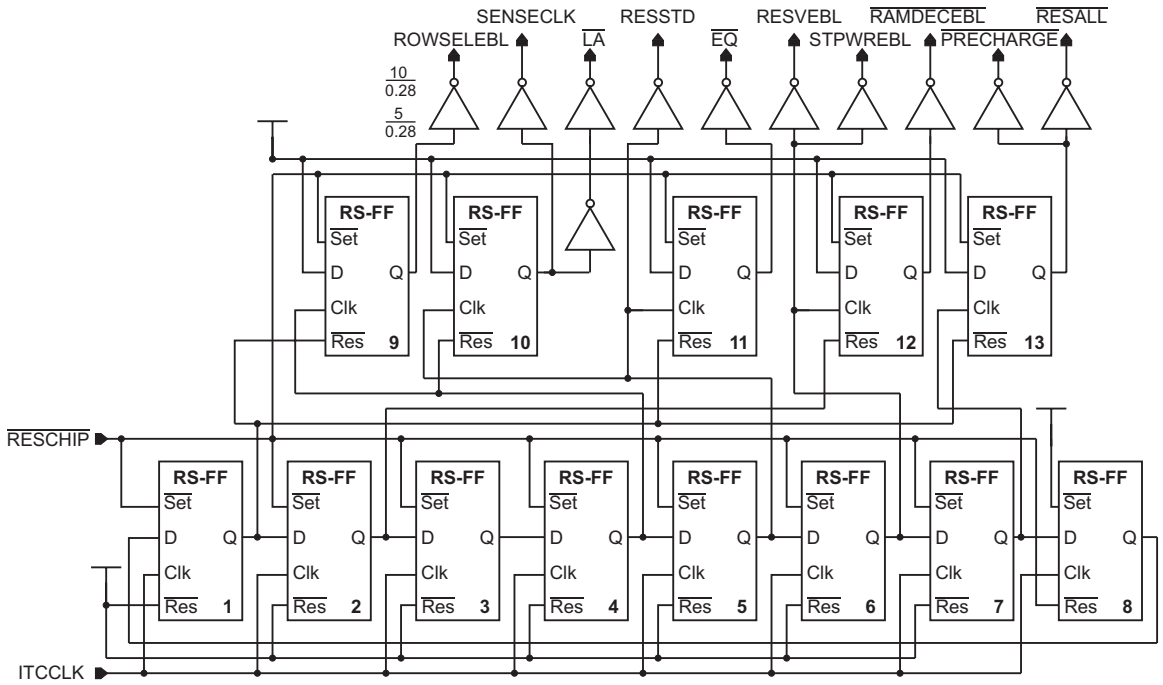


Figure 4.35: Circuit diagram of the logic that generates the ITC control signals. The sequence of the signals is determined by the wiring of the shift register in the lower part with the following flip-flops above.

Since the signal sequence is determined by the connection of the flip-flops, it cannot be changed from outside. However, the durations of the individual control signals can be changed by a variation of the corresponding clock cycle length. This provides a certain flexibility for the test of the imager.

Beside the presented circuits, the cycle control includes a small additional logic whose task is to disable selected control signals or the address decoders of the pixel array and the SRAM. It mainly consists of a few NOR gates, which are not shown here, and is controlled by the signals RAMDECEBL, ARRDECEBL, ARROFF and HDROFF as will be described below.

4.5.2 Timing of an ITC Cycle

The purpose of an ITC cycle is to test whether individual pixels of a particular row have to be reset, to execute the necessary resets and to store the new integration time of these pixels. As mentioned above, the concerning row is previously selected by the FPGA via the bus RADR[7:0]. Owing to the fact that there is an unequivocal assignment of each pixel row to a particular row of SRAM cells, the read address for the memory is hereby also determined. Nevertheless, neither the pixels are connected to the column read-out lines nor the RAM cells are connected to the bit-lines before the ITC cycle gets started. With regard to the components discussed in the preceding sections the tasks which have to be worked off in the course of a ITC cycle are as follows:

1. Read-out of the time stamps from the selected row of the local memory.
2. Comparison of these stamps with those of the previous reset curtain that has been applied to the selected row (2. reset condition).
3. Activation of the comparators to test if $V_{pix} < V_{comp}$ (1. reset condition).
4. Generation of a reset decision from the results of 2. & 3. and execution of a reset where it is necessary.
5. Storage of the new time stamps in local memory.

The complete timing diagram of an ITC cycle is shown in figure 4.36. The signals originating from the cycle control are connected to the corresponding components as indicated in the overview diagrams 4.1 and 4.10. $\overline{\text{EQ}}$ and $\overline{\text{LA}}$ are used to control the comparators, their meaning is explained in section 4.3.2. The point in time were the comparison is done is indicated.

Initially ($t = 0$ ns), the pixel outputs get connected to the column lines by ROWSELEBL. At the same time, the reset transistors of the whole array are separated from the reset control $\overline{\text{RES}}$ (cf. figure 4.2) by pulling up $\overline{\text{RESALL}}$ and the precharging of the bit-lines is switched off ($\overline{\text{PRECHARGE}}$=1). Afterwards, the RAM cells get connected to the bit-lines ($\overline{\text{RAMDECEBL}}$=0). The sense amplifiers are activated at the same time as the comparators ($t = 300$ ns). The results of both are used to generate the decision which pixels of the selected row have to be reset. As described in the preceding sections, this is done by combinatorial logic and does not require any controlling signals. The indication of the pixels which have to be reset is stored in *reset selection* with a rising edge of RESSTD (reset selected) at $t = 400$ ns. At the same time, the outputs of the pixels are separated from the column lines to reduce the power consumption. The information about the pixels which have to be reset, is now available at the sensor array by the bus $\overline{\text{RESSEL[0:169]}}$ and at the component *stamp write* by the bus STPWR[0:127]. At $t = 500$ ns the corresponding time stamps are stored by pulling high STPWREBL (stamp write enable) and the pixels are reset by setting RESVEBL=1. Some of the other control signals are switched back to their initial values preparing the next cycle. The next rising edge of ITCCLK finishes the reset process by pulling high RESVEBL, this is the start of the integration.

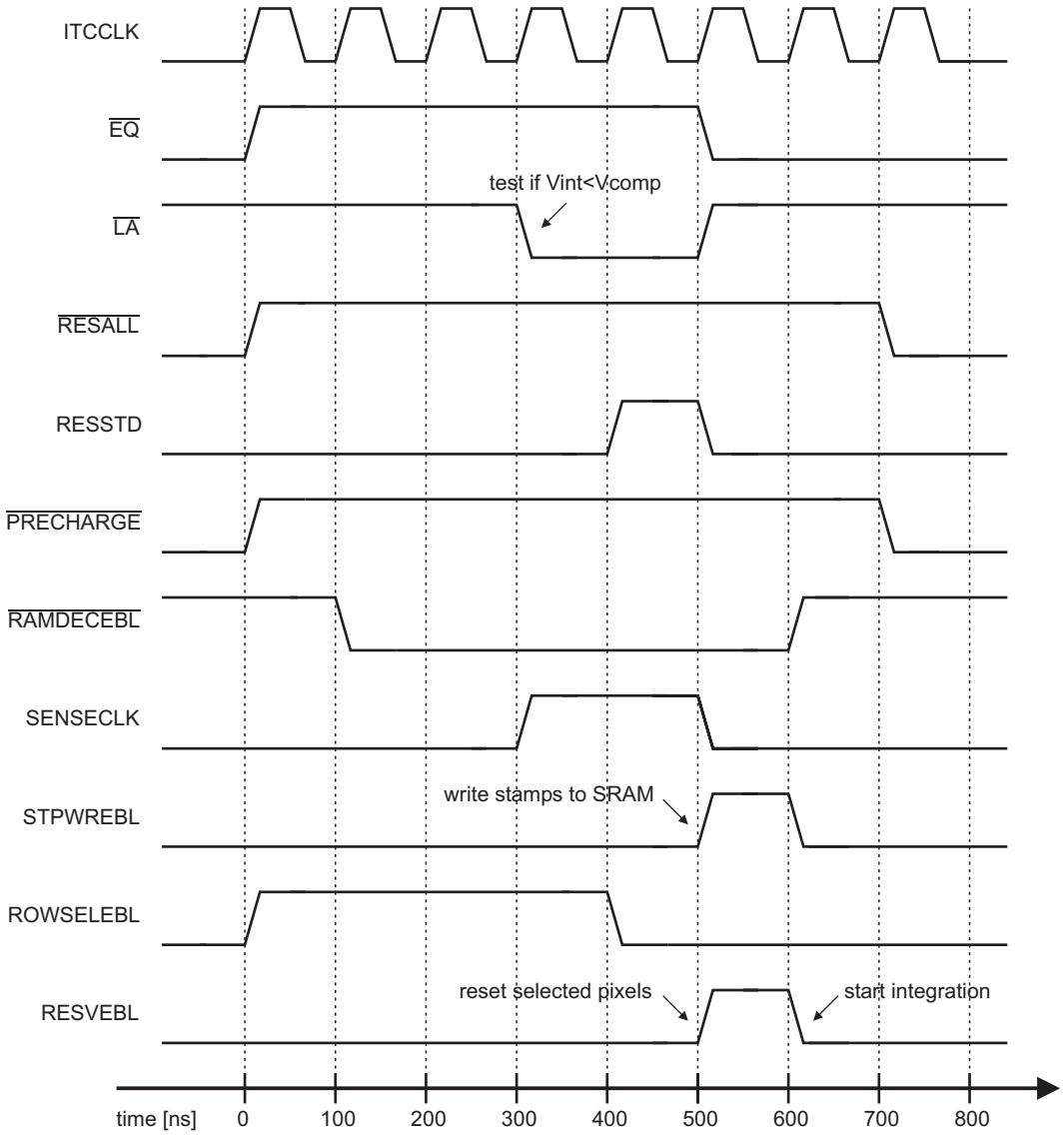


Figure 4.36: Timing diagram of a single ITC cycle.

The RAM cells get disconnected from the bit-lines. Within the last clock cycle, the precharging of the bit-lines is restarted and all pixels of the row are reconnected to the reset control line $\overline{\text{RES}}$. After 800 ns the ITC cycle is finished and all signals show the initial values.

4.5.3 Timing of a DS Cycle

A double-sampling cycle is used to store the analog values from the pixels after integration, to read out the final integration times from the RAM and to prepare the next acquisition cycle. As described in chapter 3 there is only one DS cycle followed by up to 15 ITC cycles, the exact number depending on the implemented control scheme. The conditions at the start of a DS cycle are the same as for an ITC cycle (selected address for the pixel row and the RAM row). In short, the following items have to be worked off:

1. Sampling of the reset and integration values on the capacitor banks of the DS entity.
2. Read-out of the final time stamps which belong to the selected row from the local memory.
3. Storage of the time stamps in a register for the sequential read-out.
4. Reset of all pixels of the selected row (start of integration).
5. Initialization of the RAM cells that correspond to the selected row by writing the first time stamp.

Figure 4.37 contains the complete timing diagram of a DS cycle. The whole cycle has a length of 1.6 μs, which is a consequence of the necessary settling times for the sampling of the reset and the integration values (each 600 ns). Most of the signals are also used in an ITC cycle, for that reason these signals have already been explained in the previous section. The control signals for the comparator $\overline{\text{EQ}}$ and $\overline{\text{LA}}$ as well as the signals $\overline{\text{RESALL}}$ and RESSTD are not used here. Since $\overline{\text{RESALL}}$ stays low during the cycle and a row has always to be reset as a whole in a DS cycle, the signal RESVEBL is sufficient here. The three new signals are SS, SR and RINGWREBL. SS (sample signal) and SR (sample reset) control the switches for the sampling of the corresponding values on the capacitors in the sample-and-hold cells (cf. section 4.4). As can be seen in the timing diagram, the integration value is sampled first. Afterwards, the whole row is reset (RESVEBL=1) and the reset value is sampled. The signal RINGWREBL (ring write enable) stores the final time stamps, which were previously read from the RAM, in the *stamp registers*. The sequential read-out itself is controlled by the FPGA (signals SPLCLK and RINGCLK in sections 4.5 and 4.3.9). It is accomplished between two successive DS cycles.

As indicated in figure 4.37 the current integration period ends 600 ns after the DS cycle has started (falling edge of SS). In the previous section it was shown that the integration started 200 ns before the ITC cycle has finished. Therefore, if a DS cycle is executed directly after an ITC cycle, the integration time of pixels that have been reset in the ITC cycle amounts to 800 ns, equal to the length of one complete ITC cycle. This is the shortest possible integration time. It is important that it amounts to exactly one ITC cycle since the integration time is regulated in multiples of these cycles.

Another remarkable issue is the sampling of the reset value. When the in-pixel reset transistor (M_2 in figure 4.2) is switched off, it injects a certain amount of charge into the capacitance of the photodiode. This amount will vary from pixel to pixel. To permit the acquisition of the hereby changed reset value, the sampling time has to be longer than the reset time (they start at the same time). As can be seen in the timing diagram, the reset of the pixels ends at $t = 1.0\,\mu$s (RESVEBL=0), whereas the sampling of the reset values continues until $t = 1.4\,\mu$s. On the other hand, in case of a very high illumination this can lead to a reduction of the reset values due to their short (but relevant) integration time of 400 ns. The first consequence of this is a decrease of the linearity at high illumination (V_{int} is not yet saturated), which is the same for all pixels. The second is that the net signal $V_{res} - V_{int}$ (difference between integration value and reset value) will start to decrease at very high illumination (V_{int} is already saturated). The measurement of this behavior is presented in section 5.2.2. If these effects are not acceptable for a given application, they could be reduced by a reduction of the 400 ns at the price of a (small) increase of noise.

The timing as shown in figure 4.37 is used for rows that lie within the HDR window. Rows which lie outside always integrate over a constant time T_{int} (here: $T_{int} = T_{max}$). An update of the RAM values is not necessary here, moreover it would overwrite the time stamps of the window. Nevertheless the analog values have to be read out. This is reached by running a DS cycle and setting HDROFF=1. By this, the signals SENSECLK, RAMDECEBL, RINGWREBL and STPWREBL

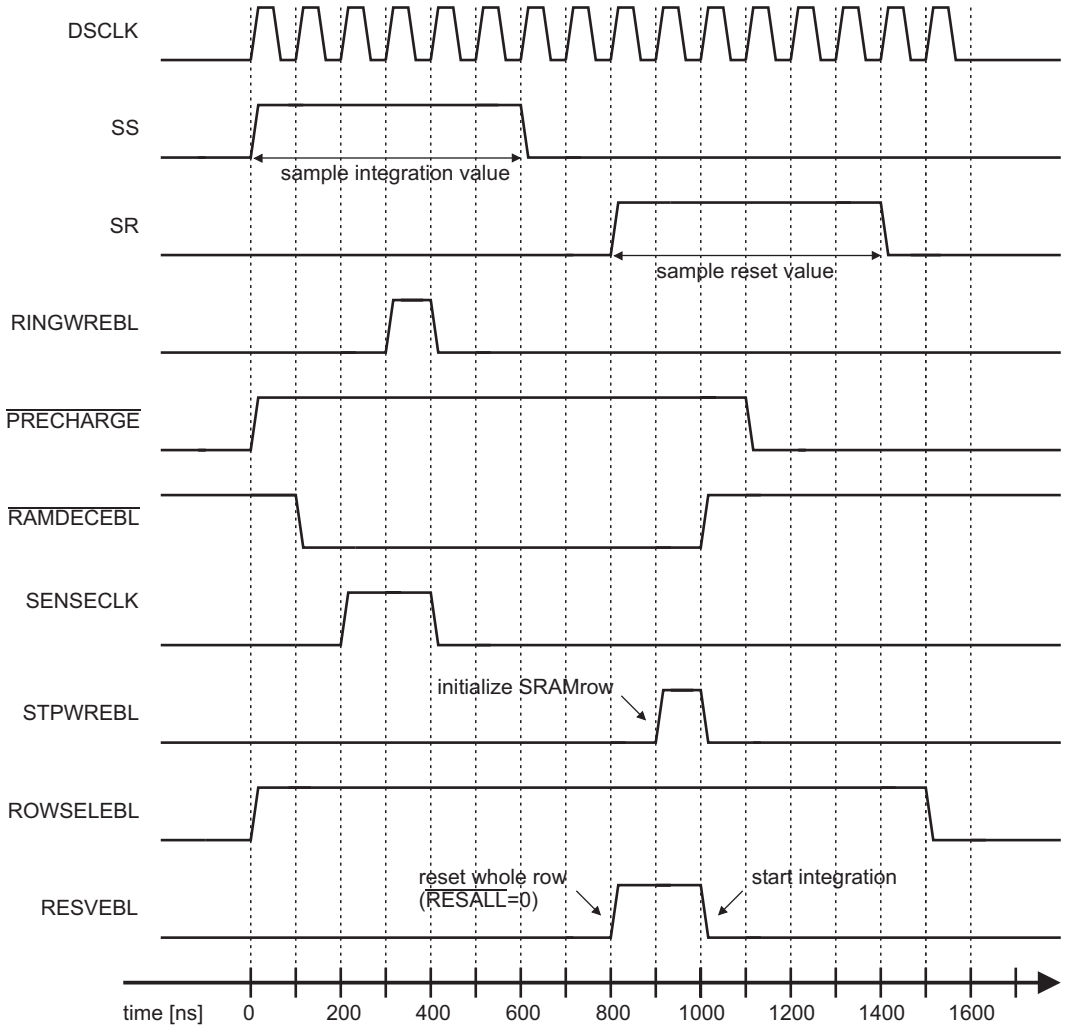


Figure 4.37: Timing diagram of a single DS cycle.

are disabled for the duration of this cycle, i.e. they keep their initial values and neither the SRAM nor the *stamp registers* are used.

4.6 Global Chip Layout

The image sensor that was described in the previous sections has been realized in a IBM CMOS 0.25 μm process (p^+-substrate). Three metal layers and one layer of polysilicon were used in the implementation. The process provides the possibility of MIM[10] capacitors. These were used for the compensation capacitors in the output amplifiers. All other capacitors were realized as MOS transistors using the gate-channel capacitance. Diffusions and polysilicon are silicided automatically in this process to reach low-resistance interconnects. The availability of a silicide block mask made it possible to omit the formation of silicide for the photodiodes, which is essential for the penetration

[10]Metal-insulator-metal capacitor. In the used process these capacitors are formed by adding an additional thin layer of metal between *metal2* and *metal3*.

of the incident light. Shallow trench isolation is used to eliminate the problem of oxide encroachment into the width of a transistor (known as ”bird’s beak”).

Figure 4.38 contains a photomicrograph of the chip. Below the photograph, a block diagram can be seen, which permits the identification of the individual components described so far. The largest part of the chip is occupied by the 178 ×178 pixels (including dummy structures). The pixels are distributed with a pitch of 7.5 μm. On the left side of the chip a second large block can be recognized. It consists of 512×128 SRAM cells, each with a size of $2.92 \times 5.32\,\mu\text{m}^2$. They add up to a memory of 64 kbit.

Since the image sensor is a mixed-signal chip, provisions have to be taken that the analog components are not affected from variations in the power supply or crosstalk caused by the digital parts. The sensitive components which are responsible for the analog data acquisition are therefore positioned as far as possible from the (mainly) digital ITC entity. Most of the digital components are positioned left from the sensor array, whereas the DS entity can be found on the right side. The sensor array itself is surrounded by a guard ring which protects it from the influence of free charge carriers. It has its own power supply (V_{dda}) which is shared only with the output amplifiers. The analog parts of the DS entity have an additional supply (V_{eea}), which has a pad for its own too. On the other hand, all digital components are supplied from the V_{dd} power ring, which surrounds the whole chip. There are tree pads to connect this ring to the external power supply, two on the left side of the chip and one on the right side. All power supplies V_{dd}, V_{dda} and V_{eea} are connected to their own blocking capacitors, which were realized as NMOS transistors. These capacitors serve as fast local charge reservoirs for high transient currents to reduce the switching noise on the power supply lines. They are located at the sensor array (50 pF), at the DS entity (50 pF) and at the SRAM array (100 pF). For the wider power and ground lines, which supply the different components, the top metal layer is used (*metal3*). These lines are also used for shielding devices against incident light.

As can be seen on the photograph, most of the pads are distributed over three sides of the chip. The clock inputs can be found in the vicinity of the small cycle control, which is located in the middle of the lower part of the chip. The analog pads are mostly located on the right side. Each pad has a size of approximately $100 \times 100\,\mu\text{m}^2$ for the connection with a single bond wire. Owing to the integrated ESD[11] structures (in case of digital I/O pads) and the routing channels for the power supply and ground lines the complete pad structure measures approximately $100 \times 350\,\mu\text{m}^2$. The exact distribution of the pads and the corresponding signals can be found in appendix A.

[11] Electro-Static Discharge

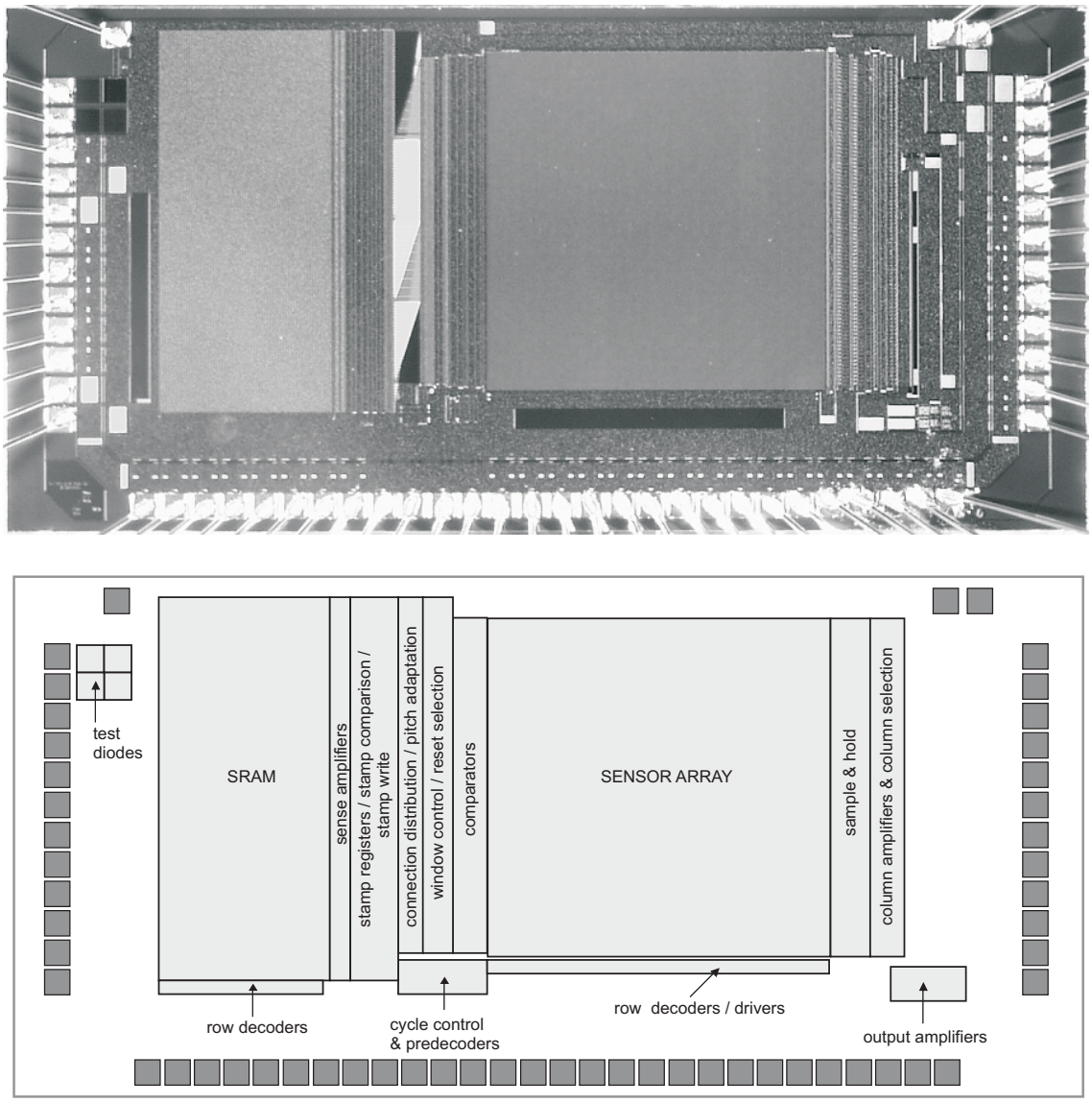


Figure 4.38: Chip photomicrograph. The chip size is $2\,\text{mm} \times 4\,\text{mm}$. In the lower diagram the individual functional blocks are identified.

Chapter 5

Measurements

The following chapter describes the measurements that have been carried out to characterize the image sensor built within this work. First, a description of the test setup which includes the data acquisition system as well as the optical setup is given. The subsequently presented measurement results characterize the image sensor in terms of quantities such as dynamic range, signal-to-noise ratio, quantum efficiency and crosstalk, which permits a comparison with other sensors. Additional measurements refer to specific properties of the sensor implementation. In the last section, sample images of several scenes recorded under varying illumination conditions and different modes of operation are shown. These illustrate the performance of the realized imager.

5.1 Test Setup

The description of the setup for the measurements is divided into three parts: In the first subsection the electronic test setup is presented. After a description of the required signals for operating the chip, the data acquisition system is discussed. On the whole the electronic test setup consists of a PC hosting the acquisition system which is connected to a separate PCB[1] which carries the image sensor. In the second subsection the implemented software used to control the chip and the acquisition system is introduced. The last part describes the optical setup for the measurements. Different sources of light were used in combination with various optical components to stimulate the sensor in a well defined manner.

5.1.1 Electronic Setup

A proper operation of the chip makes certain demands on the mixed signal test environment. On the one hand, several adjustable bias voltages and an analog data acquisition system fast enough to reach video frame rate has to be provided. On the other hand a digital control of the chip, a management of the acquired data and the possibility of user control must be realized. The requirements in detail for the operation of the imager are as follows:

[1] Printed Circuit Board

- *Analog input voltages*
  Four adjustable input voltages are required: V_{comp} defines the voltage range used for photocurrent integration in the pixels. It is connected to the comparators of the DS entity. If the output voltage of a pixel is below V_{comp}, one of the two conditions for reset is fulfilled (cf. chapter 3). V_{bpix} is the bias voltage for pixel source followers, whereas V_{bamp} is the equivalent bias voltage for the column preamplifiers. V_{sel} is used to drive the in-pixel selection transistors of a particular row for the read-out. All analog input voltages are held constant during ordinary operation of the chip. Hence, speed is not an issue here.

- *Digital I/O*
  The only digital output of the sensor is the time stamp (4 bit) which is necessary to be transferred for each sampled pixel. To control the chip 34 digital input lines are required. A detailed description of these signals is given in chapter 4. The used I/O standard is LVCMOS2 with an output source voltage of 2.5 V. A maximum speed of 10 MHz is needed for the signals DSCLK and ITCCLK, which control the double-sampling cycle and the ITC cycle defined in section 4.5.

- *Data acquisition*
  There are two channels for analog read-out of the sampled pixel voltages V_{int} and V_{res}. Both values can be read out in parallel for a particular pixel, but two read-outs on both channels are necessary for each pixel to be able to eliminate the column amplifier offsets. Taking into account the internal durations for the acquisition of a row, a frame rate of 30 Hz requires a sampling rate of 5 MHz per channel (cf. section 4.5.1).

Figure 5.1 gives an overview of the mixed signal test environment. It consists of the chip carrier board connected to a PCI[2] compatible FPGA board, which is hosted by a standard PC.

FPGA Board

The FPGA board (project name *darkwing*) has been developed in our research group as a multipurpose test system for mixed signal ASICs. A detailed description of the board can be found in [BEC01]. Figure 5.1 is confined to the functional blocks relevant for the test system. The FPGA Xilinx Virtex XCV400E [XIL02] is the main control component of the board. Its reconfigurability makes it an ideal tool for systems which demand flexible and interchangeable circuit realizations to cover a wide range of applications. For the test of the sensor it is operated at a clock frequency of 40 MHz. In section 3.5 an overview of the FPGA programming is given. The following will deal with the properties of the used devices on the FPGA board and their application in the test environment. The main tasks that are handled by the implemented circuit are:

- *PCI bus connection*
  The FPGA is connected via a PCI-bridge to the PCI bus which in turn connects the board to the hosting PC. This connection is used to provide user control and to transfer the recorded image data from the on-board SRAM to the PC memory for subsequent visualization and analysis. A transfer protocol implemented on the FPGA for the communication with the PCI-bridge provides easy access to the PCI bus. For a transmission of the uncompressed and unprocessed image data at a frame rate of 30 Hz a bandwidth of 5.5 Mbyte/s is necessary.

[2] Peripheral Component Interconnect

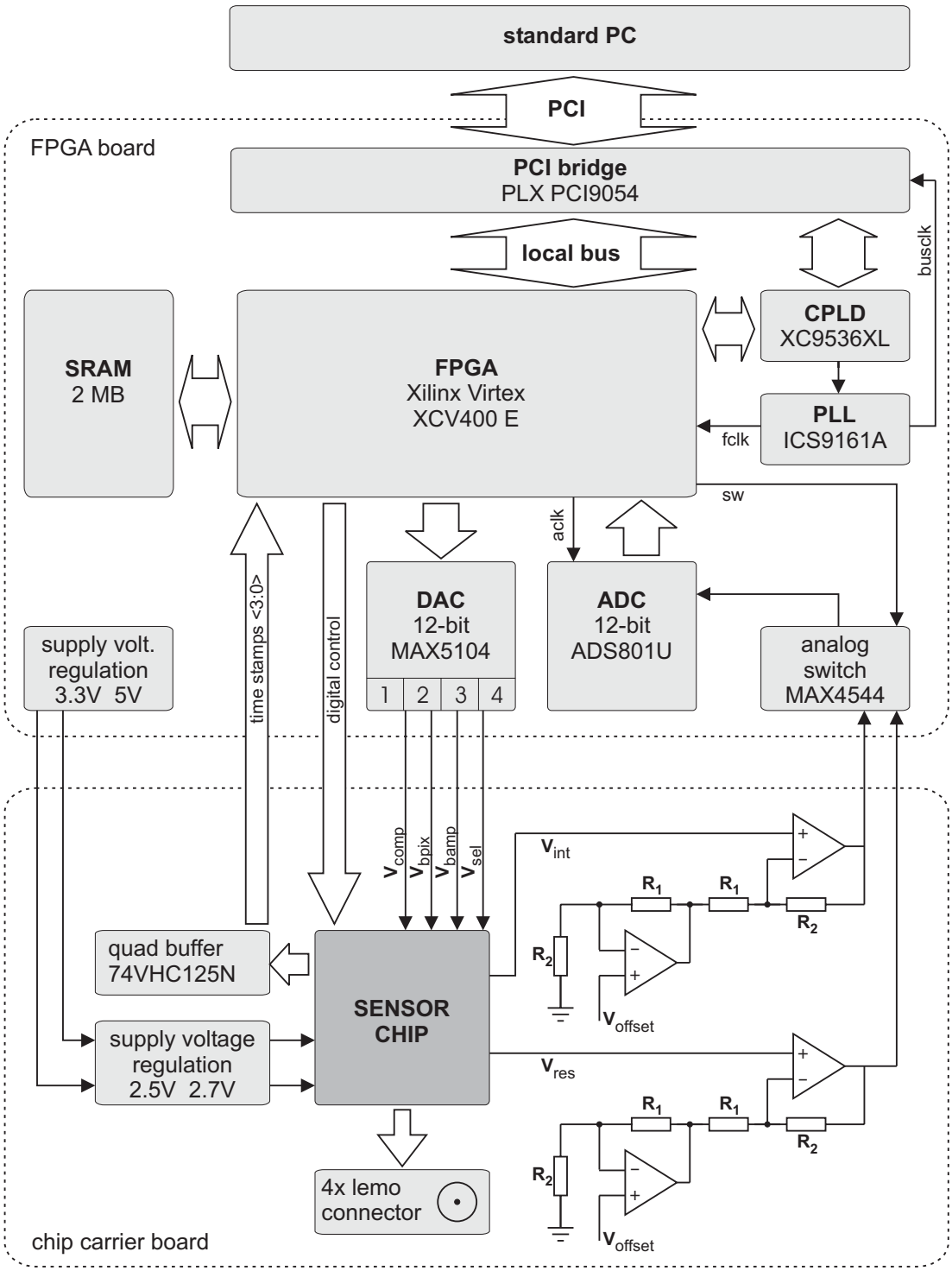


Figure 5.1: Block diagram of the electronic test setup. In the upper part the FPGA based test system for mixed-signal ASICs is presented. The carrier board for the image sensor is shown below.

At a clock frequency of 33 MHz and a width of 32 bit the maximum bandwidth of the PCI bus [PCI95] is 132 Mbyte/s, which is fast enough to transfer the image data to the host PC for imagers with much higher resolution. For the transfer of the image data to the PC RAM burst transfers[3] are used.

- *ADC control*
  Analog data acquisition is done with a 12-bit pipelined ADC, which provides a maximum sampling rate of 25 MHz [BUR96]. The FPGA provides a conversion clock of 20 MHz and receives the digitized data with a delay of 8 clock cycles on account of the pipeline architecture. The high conversion rate makes it possible to use oversampling techniques for the analog-to-digital conversion of the data coming from the image sensor. 16-times oversampling is implemented for integration times of 150 ms and above. By using an analog switch it is possible to sample both analog outputs of the sensor (V_{int} and V_{res}) with a single ADC. The acquisition of the digitized data has to be synchronous to the control of the DS entity, which is also task of the FPGA.

- *DAC control*
  Four 12-bit DACs [MAX99] are used to generate the necessary input voltages for the imager. The DACs can be programmed via daisy chain[4]. The necessary protocol has been implemented in VHDL. At every start of a new frame the user can initiate a reprogramming to the new desired voltages.

- *SRAM control*
  Two SRAM chips are located on the FPGA board. Each is organized in a 36 bit $\times$ 256 k architecture. The width of 36 bit is divided into 4 words of 9 bit with 1 byte for data and one parity bit each. The parity bits are not used here. Connected to the FPGA via a 64 bit wide data bus the two chips can be used as a single 2 Mbyte SRAM. This memory is used to store the image data acquired by the ADC in conjunction with the corresponding time stamp delivered by the imager. The VHDL RAM interface supervises the communication with the SRAM and manages the access depending on the priority of the respective request.

- *Imager control*
  The FPGA is responsible for the digital control of the imager. Besides the provision of necessary control signals (cf. chapter 4) for data acquisition and the adjustment of test modes this also includes the calculation of the sequence for the row accesses. The sequence calculation is described in detail in chapter 3. Reusability of the corresponding VHDL implementation was of particular importance with respect to the possibility of a later on-chip implementation.

The FPGA board is used for a number of different applications within our research group and makes it possible to reuse VHDL modules for particular tasks. The communication interfaces for the PCI-bridge and for the SRAM were taken from existing implementations.

Chip Carrier Board

The carrier board is dominated by the sensor chip itself. The die is packaged in a 108 pin ceramic PGA[5] which was chosen for short bond wires and a practicable bond procedure. Four cable connectors are available for direct access to the on-chip test diodes with a parameter tester. The sensor

[3] In a PCI burst transfer multiple data cycles are performed after a single address cycle.

[4] Daisy chain denotes a hardware configuration in which devices are connected one to another in a series.

[5] Pin Grid Array

chip provides two analog outputs for the data coming from the sensor array: V_{int} and V_{res}. Before these outputs are connected to the ADC they have to be adapted to its input range to make use of the full resolution of 12 bit. This is done by the circuitry shown in figure 5.1: The simple configuration of two high speed operational amplifiers (-3 dB bandwidth (g=1): 300 MHz) [ANA01], allows the subtraction of an offset V_{offset} from the voltage which is applied to the second amplifier. Selection of the resistances permits the adjustment of the gain to the desired value. It can be shown [BUR71] that:

$$\frac{R_1}{R_2} = \frac{R_4}{R_3} \quad \Rightarrow \quad U_{out} = \left(1 + \frac{R_1}{R_2}\right)(V_{int} - V_{offset}) \tag{5.1}$$

For the test setup the resistors were chosen as:

$$R_1 = R_4 = 1.1\,\mathrm{k\Omega} \quad R_1 = R_4 = 499\,\Omega$$

which results in a gain of 3.2. Tolerances of the used resistors were below 0.01 %, which is necessary for this architecture to reach a high common mode rejection ratio.

The chip carrier board provides a 48 pin connector for the communication with the FPGA board. Via an adapter board it can be connected to the CMC[6] connectors of *darkwing*. A cable can be

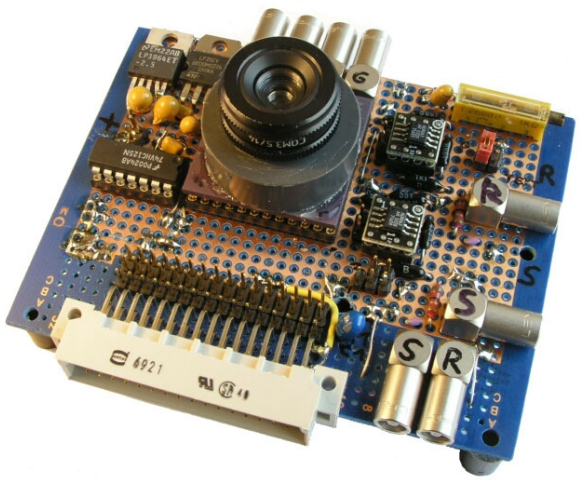


Figure 5.2: Chip carrier board with mounted lens.

inserted for those measurements in which the carrier board has to be mounted on a XY-translation stage. The digital outputs of the sensor are isolated from the cable capacitance by the use of a quad buffer. Figure 5.2 shows a photograph of the chip carrier board with the mounted lens (F/3.5, 14 mm).

5.1.2 Software Control

The standard PC hosting the FPGA board in figure 5.1 is also used as a platform for the software control of the test setup. Running under the operating system *Windows 2000* the software was used to control the settings of the imager and of the FPGA, to visualize the image data in various ways, for first measurements and to store the data for a detailed analysis later on. Programmed completely in C++ the software provides a graphical user interface, which was developed in QT[7], a cross-platform

[6] Common Mezzazine Card

[7] QT is a registered trademark of Trolltech AS.

C++ GUI[8] development framework. Low-level hardware description classes for access to the FPGA card could be reused from previous projects in our group. The GUI consists of the four sheets *Settings, Voltages, Views* and *Tone mapping* shown in figures 5.3 and 5.4, which are explained in the following. The image itself is presented in a separate window, which is not shown here (for sample images see section 5.3).

Settings: With the radio buttons *fast*, *medium* and *slow* the maximum integration time T_{max} is chosen. The available row sequences implemented in the FPGA are $T_{max} = 33\,\text{ms}$, 157 ms and 524 ms. The check box *HDR* activates the integration time control, i.e. enables the HDR window. *Oversampling* is available for *medium* and *slow*, the FPGA carries out 16-times oversampling for analog data acquisition. The averaging of neighboring pixels is adjustable: none, 2×2, 4×4 or 8×8. Sliders can be used for the free adjustment of the position of the HDR window within the sensor array. The test options are intended for separate tests and measurements of the comparators and the DS entity.

Voltages: Input voltages that were described in the previous section can be adjusted here. The *statistics* section permits a measurement of the mean signal value and standard deviation averaged over the whole sensor array or (if activated) over the HDR window only. Values displayed here represent an average of 10 frames.

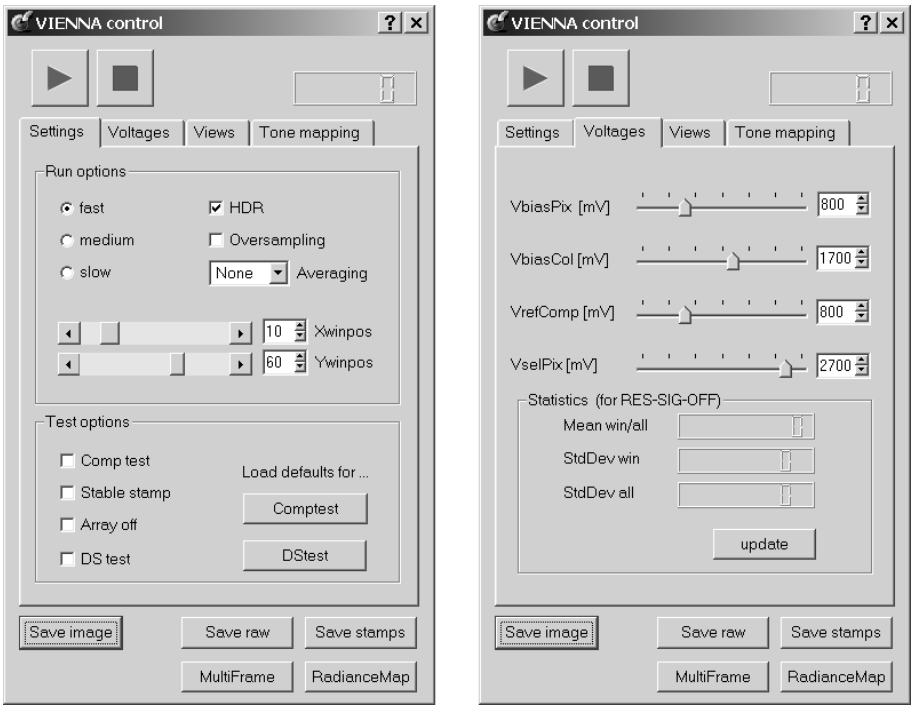


Figure 5.3: General settings and voltage adjustments in the graphical user interface.

[8] Graphical User Interface

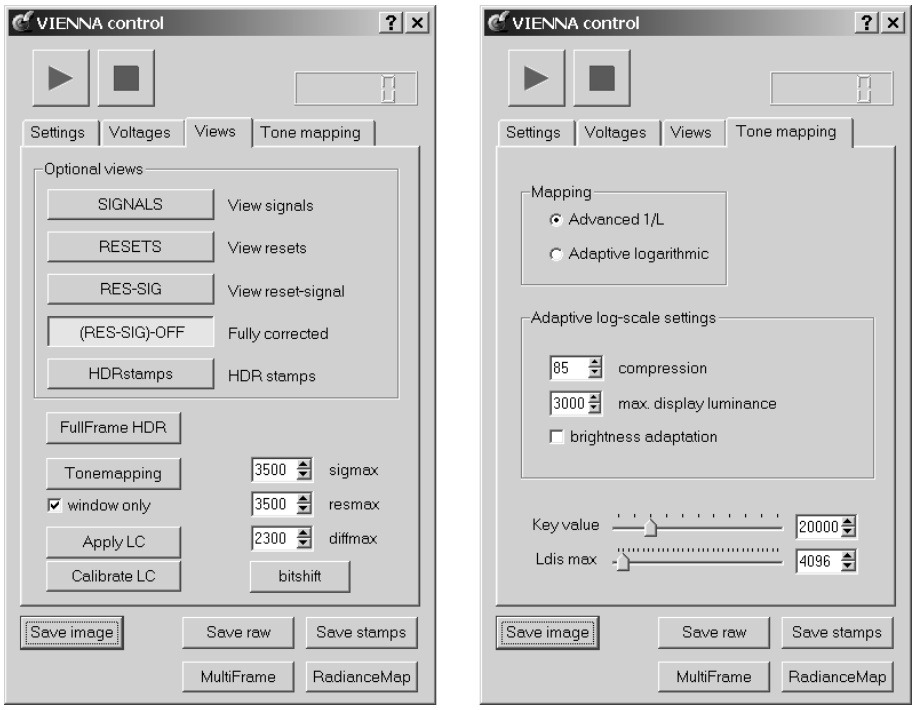


Figure 5.4: Viewing modes and settings for tone mapping in the graphical user interface.

Views: Different possibilities exist to view the recorded image data. The user has the choice between viewing uncorrected integration results V_{int}, reset values V_{res}, the difference of them (corrected for pixel FPN) or the fully corrected image data with subtracted column amplifier offsets. The option *HDRstamps* can be used to visualize the spatial distribution of the different integration times. In this case, every integration time is assigned to a particular color. Activation of *FullFrameHDR* leads to an automatic shift of the HDR window over the sensor array. The presented image is composed of the HDR windows of four frames resulting in an image which provides HDR over the whole sensor array. *Tone mapping* instructs the program to apply elaborated tone mapping algorithms for the presentation of the HDR scenes with the options made in the following settings page. The range of data which is displayed can be adjusted for areas were no DR expansion is used.

Tone mapping: This page allows the user to choose between the two available tone mapping algorithms *Advanced 1/L* and *Adaptive logarithmic*. The algorithms and the available parameters are described in 5.3.1.

Different possibilities for the storage of the recorded data are provided on the main menu page. With *Save Image* the viewed image is saved in PNG[9] format whereas *Save Raw* results in the storage of all sampled values in ASCII[10] format for a later analysis. *Multi frame* offers the opportunity to record 100 successive frames in a compact binary format for noise measurements. *Save Stamps* and *RadianceMap* save only a subset of the raw data.

[9] Portable Network Graphics is a file format for lossless and well-compressed storage of raster images.
[10] American Standard Code for Information Interchange

5.1.3 Optical Setup

Most measurements of the properties of an image sensor require a well defined optical stimulation of the chip. Exact control of the parameters concerning the stimulating light, i.e. the intensity, wavelength, spatial and temporal homogeneity is crucial here. The measurements described in the following were carried out in the optical laboratory of the Kirchhoff-Institute, which provides the necessary sources of light, opto-mechanical devices and instruments. In order to provide the possibility of a precise spatial adjustment, the chip carrier board was mounted on a precision XY-translation stage. The stage itself is set up on a vibration absorbing table. For all measurements the chip was used without a mounted lens. The lens shown in figure 5.2 was exclusively used for taking the sample images presented in section 5.3. In the following the respective optical setups, the used sources of light and calibrations carried out for the different measurements are discussed.

Quantum Efficiency Measurements

In order to calculate the quantum efficiency η (cf. equation 1.6) at a predefined wavelength λ it is necessary to measure the generated photocurrent I_{ph} at the incident light power P_i. The experimental setup for this measurement is shown in figure 5.5. A Xe arc lamp (75 W) was used as a light source. It emits a nearly white spectrum within the visible range of wavelengths (cf. appendix B). The respective wavelength for the stimulation of the chip is selected by an attached monochromator [11]. It was calibrated before by the use of a Hg(Ar) spectral calibration lamp with a characteristic line at 615 nm. To be able to measure η for the different kinds of photodiodes realizable in the CMOS process, test diodes have been implemented on the chip lying outside of the pixel array (see chip photo in figure 4.38). Each of these diodes is designed as a square of $100\,\mu\text{m} \times 100\,\mu\text{m}$. In order to stimulate solely the intended photodiode with a known beam power, the light from the monochromator has to be focussed onto a single spot with a diameter $<100\,\mu$m. Prior to this, a spatial filter system is used to eliminate disturbing border effects from the beam. Within the spatial filter system the beam is passed through an aperture with a diameter of $5\,\mu$m and is subsequently transformed back into a parallel beam by lens. The generated photocurrents were recorded by a semiconductor parameter tester, which is able to measure currents down to 10 fA [HPA94]. The parameter tester

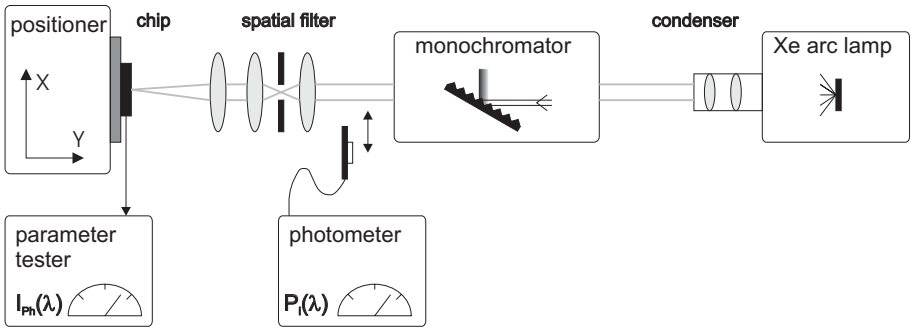


Figure 5.5: Optical setup for the measurement of the wavelength dependent quantum efficiency.

was also used for the generation of the reverse voltage for the operation of the photodiodes (2.5 V). Due to the wavelength dependent efficiency of the monochromator and deviations from the spectral flatness of the arc lamp the optical system had to be calibrated. This was accomplished by the use

[11] Oriel grating monochromator, model 77250 (1200 l/mm grating).

a photometer[12], which recorded the beam power in dependence of the adjusted wavelength. The photometer itself was calibrated by its spectral response curve delivered by the manufacturer. A calibration of the recorded η curve to the absolute value of incident light intensity was reached by a final measurement of the beam power at the position of the chip (laser spot measurement, see below) and the hereby generated photocurrent on the chip.

Response Curve / DR Measurements

As source of light for the determination of the response curve and the DR of the sensor the already mentioned Xe arc lamp was used. Under the condition of an activated integration time control it is necessary to vary the light intensity over 7 decades in magnitude to be able to measure the complete response curve. This was accomplished by the use of several neutral density filters, which were inserted in the optical path. The insertion of more than one filter should lead to an overall attenuation that could be calculated by the multiplication of the individual attenuation factors of the inserted filters. In reality, this is not the case. Multiple reflections occurring between successive filters are responsible for a decrease of the overall attenuation. For the case of two filters for example, a part of the light that passed the first filter is reflected by the second one. This part is reflected again from the first one to finally penetrate through the second. A small tilt of the filter relative to the optical axis reduces this effect, but nevertheless it has to be considered in the measurements. This leads to the necessity of a calibration of the chosen filter adjustment for a given experimental setup. The used filter system consists of 8 neutral density filters, each having an attenuation factor of approximately 0.1. The complete system is described in detail in [TEO97].

As mentioned above, the emitted spectrum of the arc lamp is nearly white but certainly not restricted to the visible range of wavelengths. On the other hand, the sensor is build for stimulation within the visible spectrum. To restrict the spectral bandwidth of the stimulating light and hereby enhance the comparability with other sensors, a bandpass filter has been inserted in front of the sensor. The pass band of this filter is 400-730 nm and is shown in appendix B. Certainly the filter has no influence on the shape of the response curve (sensor signal to incident light intensity), but owing to the eliminated UV and IR part of the lamp spectrum and the finite quantum efficiency of the sensor in this range it leads to a shift of the measured response curve.

Temporal Noise and Homogeneity Measurements

For a determination of the temporal noise of the measured sensor signal many successive frames have to be recorded. To prevent the distortion of the measurement, a light source with high stability is needed. AC[13] driven sources of light always have a portion of intensity varying with the supply frequency and are therefore unsuitable. The Xe arc lamp is driven by a DC[14] supply, but nevertheless turbulence in the Xe gas can lead to an instability in the location of the arc causing a varying intensity. This effect depends on the total lamp current, the current pulsation, cooling, and the remaining life time of the lamp [ORI95]. To avoid these problems a commonly available 50 W halogen bulb was used, driven by a DC power supply with a high PSRR[15]. For the measurement of the homogeneity within the pixel array, i.e. the FPN under illumination, a diffuse reflector was used to reach a spatial homogeneous illumination.

[12] Tektronix J1800 Series LumaColor photometer and irradiance head J1812.
[13] Alternating Current
[14] Direct Current
[15] Power Supply Rejection Ratio

Determination of the Crosstalk

The crosstalk was measured by the stimulation of a single pixel and the recording of the signals from the surrounding pixels. To be able to stimulate only one pixel (pixel pitch = 7.5 μm), a light spot with a small diameter is necessary. It was generated by the use of a red diode laser (wavelength λ=675 nm, max. beam power 10 mW) in combination with the spatial filter mentioned above. Herby a spot diameter of 5 μm was reached. An alignment of this spot by the XY-translation stage was possible with a resolution of 1 μm. The intensity was regulated by the same neutral density filter system that was used for the DR measurements.

5.2 Measurement Results

Within the following subsections the measured characteristics of the image sensor and of the process it has been realized in, are presented. To begin with, the quantum efficiency of the different types of photodiodes that can be used are shown. Next, the average response curve of the sensor is discussed. The measurement of the offset distribution in complete darkness allows a quantitative analysis of the individual contributions to the total FPN. An analysis of the remaining noise under illumination leads to the calculation of the signal-to-noise ratio. Additional measurements of the pixel-to-pixel slope variations, the accuracy of the integration time control, the leakage of the analog memory cells and the crosstalk between the individual pixels complete the characterization.

It should be kept in mind here, that absolute values of the output voltages V_{int} and V_{res} have no meaning since they are dependent on the setting of the bias voltages.

5.2.1 Quantum Efficiency

In a CMOS process different kinds of *pn*-junctions can be used as photodiodes in an image sensor (cf. section 1.2). In order to be able to select the optimal junction for a given application, it is necessary to investigate their spectral efficiency for the conversion of incident photons to electron-hole pairs. Since this efficiency is determined by the specific doping profiles of the junctions it has measured for every new process that is taken into account for usage. The quantity that describes the spectral sensitivity of a photoreceptor is the quantum efficiency η defined in section 1.1. It can be calculated by the measurement of the incident light power and the generated photocurrent at a specific wavelength λ (cf. equation 1.6). Using the optical setup shown in figure 5.5 this was done for the three *pn*-junctions available in the 0.25 μm CMOS process the imager has been realized in. The results of these measurements can be seen in figure 5.6. The first three graphs show the quantum efficiencies for the junctions n^+/substrate, nwell/substrate and p^+/nwell. Due to the limited capabilities of the monochromator grating the measured range of wavelengths is about 450-800 nm. The individual points of the graphs were taken with a step size of 10 nm and connected by straight lines to guide the eye.

The curves can be used for a comparison of the different photodiodes, since they were measured under the same conditions and with the same incident light intensity. The relative values of their quantum efficiencies are therefore well defined. However, the absolute quantum efficiency was calculated from a subsequent calibration at a single wavelength (laser spot), which leads to an accuracy of ± 0.15 of the absolute quantum efficiency values.

In all graphs there is a strong oscillation visible which stems from interference effects in layers made of SiO_2 on top of the silicon wafer (passivation / interlevel dielectrics[16]). The fourth graph

[16] The interlevel dielectrics isolate the conductive layers from each other. They consist of planarized layers of SiO_2.

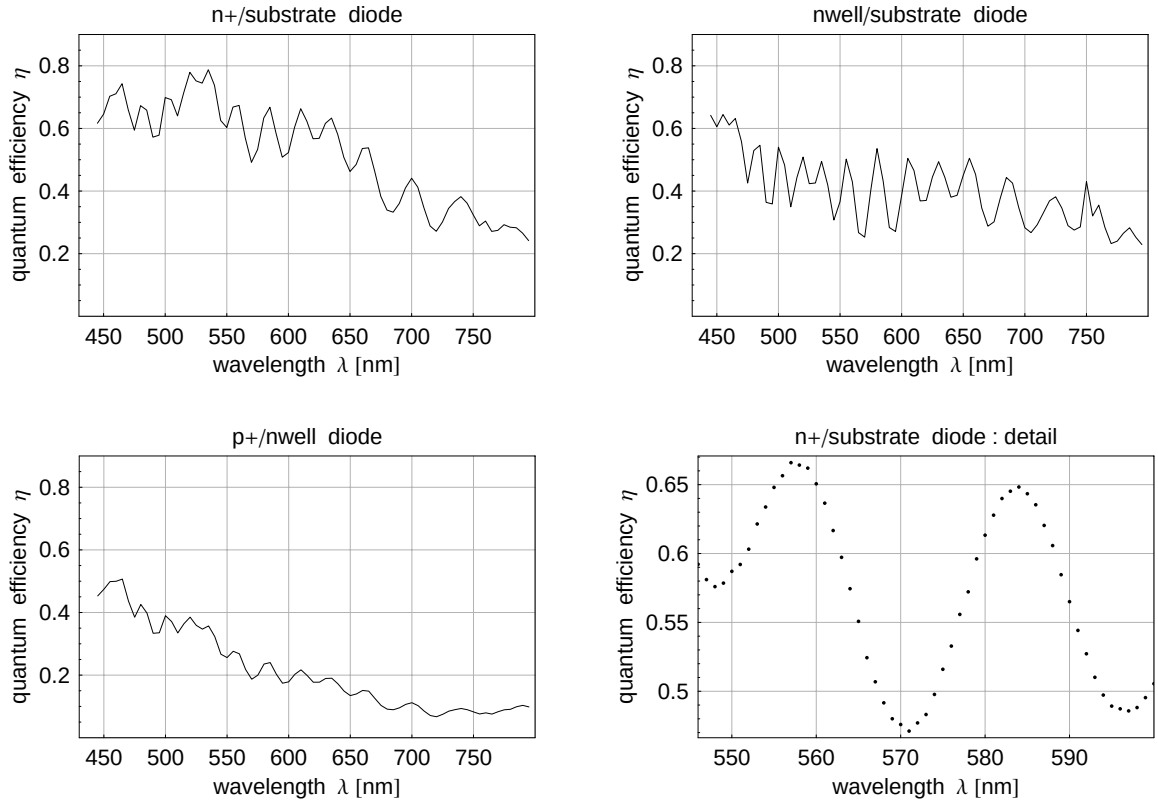


Figure 5.6: Quantum efficiency η of the available photodiodes in the used $0.25\,\mu$m CMOS process. The curve in the lower right is a measurement of a detail in the η-graph of the n^+-substrate-diode taken at a higher resolution.

in figure 5.6 represents a measurement of a detail of this oscillation occurring in the η graph of the n^+/substrate junction. It was taken with a higher resolution of 1 nm. As the maxima result from constructional interference, the wavelengths of two neighboring maxima λ_1 and λ_2 ($\lambda_1 < \lambda_2$) can be used to calculate the thickness d of the interference layer. Under the condition of vertical illumination for every maximum of the interference pattern with index k the following relation is valid:

$$2d = \frac{m_k \lambda_k}{n_{SiO_2}} \qquad k \in \mathbb{N} \tag{5.2}$$

n_{SiO_2}=1.46 is the refractive index of SiO_2 taken from literature and m_k is the number of wave cycles that fit into the thickness d of the interference layer ($m \in \mathbb{N}$). Considering the two successive maxima one additional wave cycle will fit into d at the maximum in λ_2 since $\lambda_1 < \lambda_2$. Substituting $m_1 = m_2 + 1$ in equation 5.2 ($k = 1, 2$) yields

$$m_2 = \frac{\lambda_1}{\lambda_2 - \lambda_1} \tag{5.3}$$

The wavelengths of the two successive maxima can be taken from the fourth graph in figure 5.6: With $\lambda_1 = 558\,$nm and $\lambda_2 = 584\,\text{nm} \Rightarrow m_2 = 21$. Substituting m_2 in equation 5.2 ($k = 2$) and solving for d yields $d = 4.2 \pm 0.2\,\mu$m.

The thickness of the final passivation layer (oxide, nitride and polyimide) given in the IBM process parameters is $5.0 - 6.9\,\mu$m [IBM01]. Subtracting the thickness of the polyimide layer that

should be removed completely in the subsequent etching process (cf. appendix C) yields 1.15 μm. Adding the planarized interlevel dielectrics, which lie above the photodiode, results in a total thickness of 4.7 ± 0.4 μm, which is in agreement with the measurements.

In the following, the quantum efficiencies of the different diodes will be compared with each other. The differences can be explained by the specific widths of the depletion layers and their distance from the surface of the wafer. Unfortunately, the doping profiles of the individual diodes are not available but a first understanding can be reached by the assumption of relative doping depths usually found in CMOS processes. The n^+-substrate-diode shows a maximum in η at smaller wavelengths. This can be explained by the location of the junction close to the surface of the silicon wafer: Most of the light with short wavelengths is absorbed near the surface whereas longer wavelengths are absorbed deeper in the substrate (cf. section 1.1). A similar behavior can be observed for the p^+-nwell-diode. Here, a smaller diffusion depth leads to an even stronger shift of the maximum to smaller wavelengths. Nevertheless the quantum efficiency of the n^+-substrate-diode is generally higher than that of the p^+-nwell-diode. This is due to the fact that the n^+-substrate-diode can also collect parts of the charge carriers that were generated in the deeper regions of the substrate. The p^+-nwell-diode on the other hand is shielded by the nwell-substrate junction against charge carriers that were generated in greater depths, it is therefore relatively inefficient. In the graph of the nwell-substrate-diode, the dependency in the wavelength is not as prominent as for the other diodes. Here, the junction is located deeper under the surface than for the others, it can collect charge carriers from above the junction as well as from deeper in the substrate. To find out why this junction is less efficient than the n^+-substrate-diode a device simulation based on the detailed doping profile would be necessary.

For the implementation of the photoreceptor the n^+-substrate-diode were used. As mentioned in section 4.2.1 it was realized as a n^+-diffusion resistor to get rid of the silicidation layer. To verify the difference an additional n^+-substrate-diode defined as an ordinary diffusion, i.e. with silicidation layer, was implemented. This diode has a quantum efficiency that is decreased by a factor of 5.8 compared to the non-silicided version (measured with a laser spot).

5.2.2 Response Curves and Dynamic Range

The response curve of a photoreceptor describes its output signal as a function of the incident light intensity. The limits of this curve, i.e. the maximum input signal without saturation of the sensor and the minimum input signal that is equivalent to the standard deviation of the read noise, can be used to calculate the DR of the sensor (cf. equation 1.17). In the following subsection the averaged response curve with respect to all pixels lying within the HDR window is presented. A Xe arc lamp in conjunction with the bandpass filter (400 nm-730 nm) described in appendix B was used as light source. The adaptive integration time control was programmed here to regulate the integration time from $T_{max} = 33$ ms to $T_{min} = 4\,\mu$s depending on the incident light intensity. A frame rate of 30 Hz was used.

Prior to the discussion of the response curve in the range of the detectable light intensities, a comparison of the response curves with and without integration time control measured at a higher resolution of data points is drawn. For the variation of the incident light intensity in small steps two polarization filters were rotated against each other. Figure 5.7 contains the results of these measurements. On the y-axis of the diagram the mean signal $\overline{S}$ is given in arbitrary units (a.u.). The reason for this is lies in the covered range of signals: The measured quantity is the offset corrected voltage V_{sig} (see equation 4.6), which has to be multiplied by the integration time dependent expansion factor 2^z (z = time stamp). For higher light intensities this would result in unrealistic high voltage values.

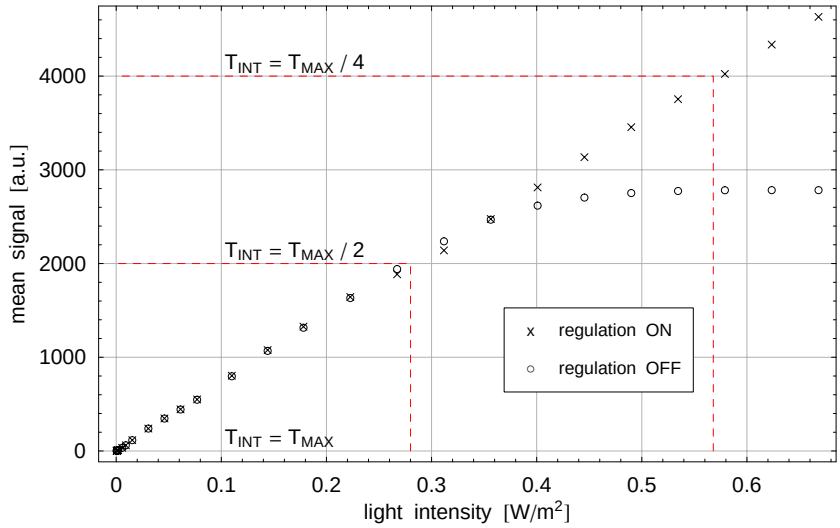


Figure 5.7: Comparison of the measured average response of all pixels lying within the HDR window with (regulation ON) and without (regulation OFF) integration time control.

Therefore, it was decided to use the value of V_{sig} in units of LSB times 2^z for the signal S shown in the response curve. For low light intensities (as long as z=0) S is simply the value of the output voltage V_{sig} in units of measured LSB. Thus, using the known input range of the 12-bit ADC of 4 V and the on-board preamplifiers gain of 3.2 (cf. subsection 5.1.1), V_{sig} can be calculated from the values given in the response curve by

$$V_{sig} = \frac{4\,\mathrm{V}}{2^{12}\,\mathrm{bit} \cdot 3.2} \cdot S = 3.05 \cdot 10^{-4} \frac{\mathrm{V}}{\mathrm{bit}} \cdot S \tag{5.4}$$

This permits a comparison with the results of the noise and slope measurements that will be discussed later on.

The first graph in figure 5.7 (regulation OFF) shows the mean signal $\overline{S}$ of the photoreceptors without integration time control. It represents the response of a standard CMOS APS sensor that is operated with a constant integration time of $T_{int} = 33\,\mathrm{ms}$. The accuracy of the data points is determined by the precision of the photometer used to measure the light intensity. It amounts to $\pm 2\,\%$. The slope of the linear part of the response calculated with equation 5.4 is 2.26 V/(W/m^2). As can be seen, the sensor gets saturated at higher light intensities. For intensities higher than about 0.5 W/m^2 the integration capacitance is completely emptied within T_{int}. During the measurement of the second graph (regulation ON) the adaptive integration time control was active. The measured signal stays proportional to the incident light intensity. With increasing intensity the integration time was shortened by the sensor to prevent a saturation. The individual steps of the regulation can be observed in the change of the time stamps. Each point on the response curve were a new integration time is adjusted by the sensor is indicated by red lines. As the new integration time starts with a reset of the affected pixels, the pixels that are reset here, show the same response curve as after the initial reset. The point were a new integration time is adjusted is defined by the choice of V_{comp}, the reference voltage of the comparators located in the ITC entity (see sections 3.2 and 4.3.2). As can be seen it is chosen in a way, so that only that part of the unregulated response curve is used that is approximately linear. By increasing V_{comp} the intensity were a shorter integration time is adjusted is lowered. This results in an improved linearity at the price of a lower overall DR. The decrease of V_{comp} has the opposite effect. For all measurements within this chapter V_{comp} was set to 800 mV.

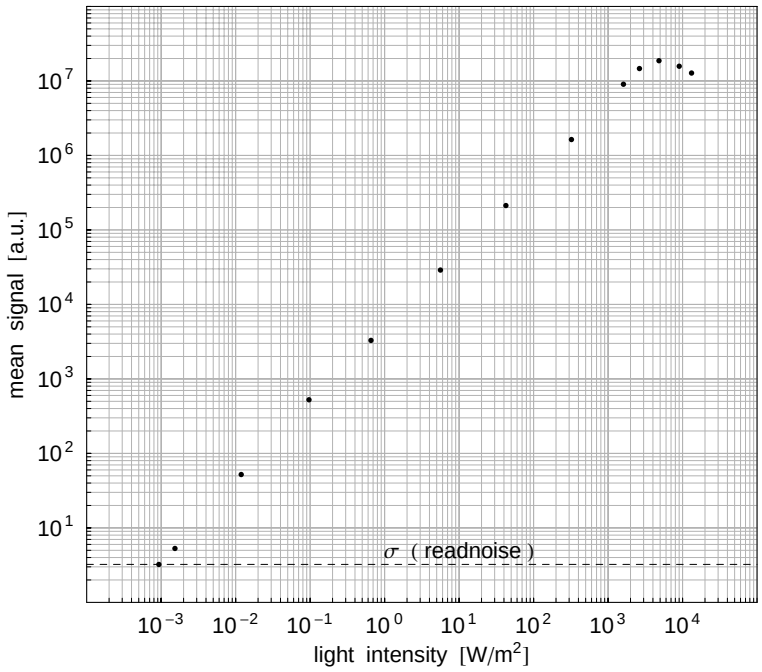


Figure 5.8: Average photoreceptor response with activated integration time regulation measured in a range of 7 decades of light intensity.

In figure 5.8 the response curve of the sensor with activated integration time control was measured within in a range of 7 decades of light intensity. Neutral density filters were used to vary the intensity. At the upper and the lower limit of the curve the supply current of the arc lamp was varied to increase the number of measurement points. Each of these points was calibrated separately by the use of a photometer. As can be seen, with the given settings the sensor is able to cope with intensities in the range from 1 mW/m^2 to 5 kW/m^2, resulting in a dynamic range of

$$\mathrm{DR} = 20 \cdot \log\left(\frac{5\,\mathrm{kW/m^2}}{1\,\mathrm{mW/m^2}}\right) = 134\,\mathrm{dB} \tag{5.5}$$

The lower limit is given by the standard deviation of the noise remaining without illumination of the chip (cf. section 5.2.3). At high intensities a decline of the mean signal is observable. The reason for this is the reduction of the reset values by the incident light before this value can be sampled on the corresponding capacitor. Even though this effect could be further reduced by a specific reduction of sampling durations for the reset values (see section 4.5.3), it is not completely avoidable. On the other hand, an inspection of the reset values is always possible and thus permits discriminating from deteriorated values.

There are several ways to further enhance the DR at the expense of a deterioration of other properties. Depending on the application this may be useful. As mentioned above a decrease of V_{comp} would enhance the DR at the price of a deteriorated linearity. The reduction of reset sampling durations by a reprogramming of the VHDL module would also enhance the maximum detectable intensity to a certain degree. Another possibility is to increase the maximum integration time T_{max} as discussed in chapter 3. This results in a decrease of the minimum detectable intensity at the price of a reduced frame rate. The improvement of the DR is limited here first by the increasing influence of pixel-to-pixel dark current variations and second by the available amount of regulation steps (here

max. 15). For T_{max}=157 ms a maximum DR of 137 dB was reached.

5.2.3 Pixel FPN and Column Amplifier Offsets

In principle, the measurement of the resulting voltage V_{int} of a particular pixel after the integration of the photogenerated charge permits the calculation of the predominant light intensity at the pixels location. Nevertheless the measured value is deteriorated by several sources of noise. The unavoidable device-to-device mismatch leads to a fixed pattern noise (FPN), i.e. a spatial offset distribution in the image data which is constant in time. There are two major contributions to this pattern: Firstly, the mismatch in the pixel circuits itself and secondly the mismatch in the column amplifier circuits.

For integration based sensor concepts, a removal of the pixel related FPN is possible by a subtraction of the sampled reset voltage V_{res}, which exhibits the same offset distribution as the integration result V_{int} (see section 2.1). It should be kept in mind here, that this is strictly true only in the case of no illumination since V_{res} represents the result of an exposure with an integration time close to zero. In reality, there are variations in the slope of the response curve which prevent a complete elimination of the FPN under illumination (see subsection 5.2.5). For the removal of the FPN part which derives from the column amplifiers the known technique of difference double-sampling is used (cf. section 4.4.1): The amplifiers for V_{int} and V_{res} in a particular column are fed with the same input voltage. A subsequent read-out of the outputs V_{mint} and V_{mres} makes it possible to calculate the offset differences of these amplifiers and therefore to calculate the final signal voltage V_{sig} by equation 4.6, which is repeated here:

$$V_{sig} = (V_{res} - V_{int}) - (V_{mres} - V_{mint})$$

In the following, the measurement results for the two contributions to the FPN are presented. They were measured without illuminating the sensor. Figure 5.9 shows the distribution of the column amplifier offset differences $(V_{mres} - V_{mint})$. The left plot presents the spatial distribution over the columns whereas in the right one contains the statistical distribution. As can be seen, there is no systematical drift or regularity. The offset differences scatter with a standard deviation of 1.48 mV (chip output voltage differences). The mean value of -4.39 mV derives from the offset difference of the two off-chip drivers for the analog read-out channels. Thanks to the subsequent subtraction of $(V_{mres} - V_{mint})$ from the term $(V_{res} - V_{int})$, which exhibits the same offset, it has no relevance.

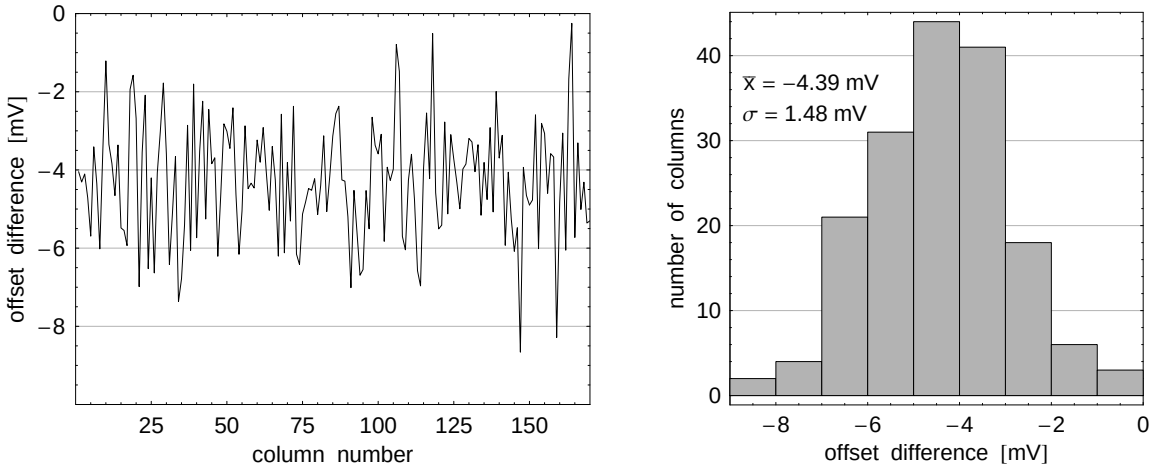


Figure 5.9: Distribution of the offset differences which exist between the column amplifiers for the reset value V_{res} and those for the integration value V_{int}.

A separate measurement of the pixel related FPN is not trivial since V_{int} always contains the unknown column amplifier offsets (only the offset differences are important for the correction). Nevertheless, within a single column there is only pixel related FPN: Figure 5.10 shows the offset distribution within a single column. The average over the standard deviations of all columns yields 6.67 mV, which is much higher than the portion deriving from the amplifier offset differences. The assumption that this value can be taken as the total pixel FPN is supported by the fact that no dominating patterns can be observed in V_{int} (column amplifier offsets become visible after subtraction from V_{res}). This is consistent to the experience that the mismatch of the in-pixel source-followers should lead to a statistical offset distribution for identically designed pixels.

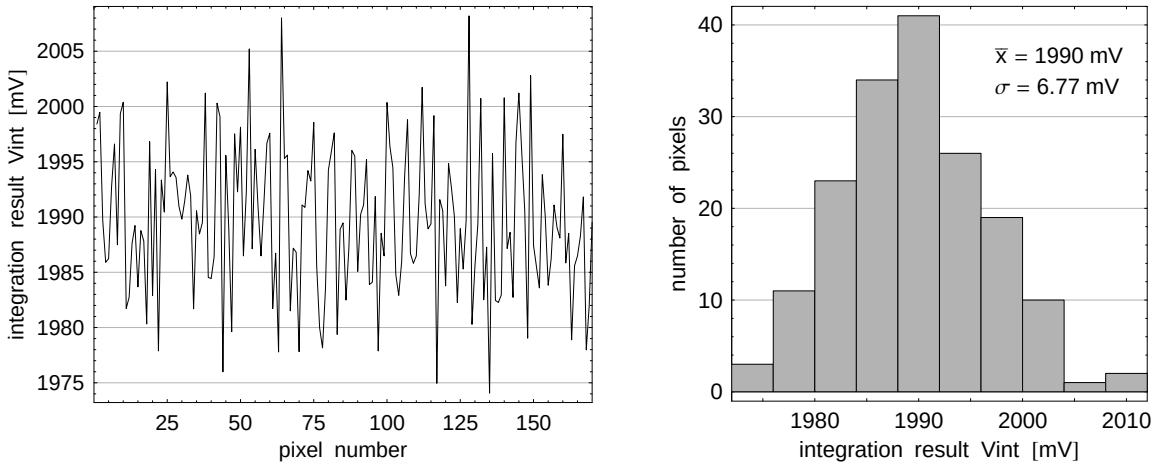


Figure 5.10: Offset distribution of the integration value V_{int} within a sample column. The distribution was measured without illuminating the chip.

Within a single frame, the standard deviation of the total noise in V_{int} amounts to 6.89 mV. The success of the offset reduction described by equation 4.6 can be tested by measuring the remaining noise in V_{sig}. Within a single frame it amounts to only $\sigma = 1.07$ mV. However, it has to be kept in mind here, that these measurements were carried out without illuminating the chip. While the column amplifier offsets are independent from the light intensity the remaining FPN in V_{sig} is influenced by the pixel-to-pixel slope variations of the response curve and eventually further by light induced effects. Therefore, the measured remaining noise of $\sigma = 1.07$ mV represents only the minimum noise without (or with small) illumination. This is the value which is denoted in figure 5.8 as read noise (in a.u.).

The slope variations of the response curves as well as the height and the composition of the remaining noise under illumination are discussed in the next subsections.

5.2.4 Composition of the Remaining Noise

Figure 5.11 shows the distribution of the total remaining noise, i.e. the noise in the signal voltage V_{sig}, present within the HDR window. The measurement was carried out at a light intensity of 0.4 W/m^2 and a frame rate of 30 Hz. No regular patterns can be observed in the spatial noise distribution. The measured standard deviation of the total noise is σ_{tot}=3.09 mV.

To investigate the composition of the remaining noise, a sequence of 100 frames was recorded. This sequence permits to determine the temporal noise in the signal of every individual pixel. The corresponding distribution of the standard deviations is shown in figure 5.13. On the average, the

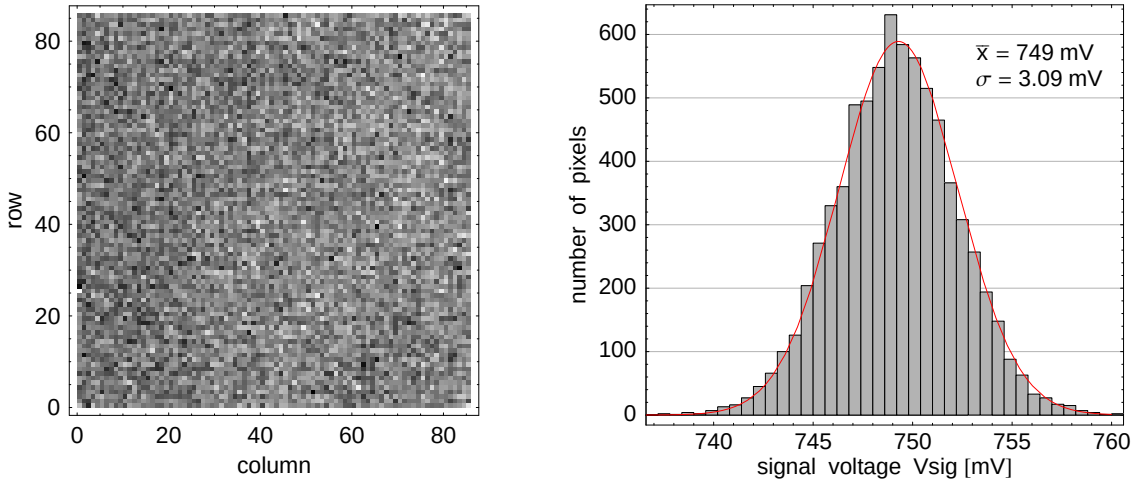


Figure 5.11: Distribution of the total remaining noise within a single frame measured at a light intensity of 0.4 W/m^2. Left: Spatial distribution within the HDR window. The occurring values of V_{sig} were mapped to a scale of grey values from black (737 mV) to white (761 mV). Right: Corresponding statistical distribution.

measured standard deviation σ_{temp} amounts to 1.74 mV. By calculating the mean over the frame sequence it is possible to suppress the temporal noise component by a factor of 10. The variation of the signals within the resulting frame represent the fixed pattern component of the remaining noise. In figure 5.12 the statistical distribution of this frame is presented. It shows a standard deviation of $\sigma_{FPN} = 2.62$ mV, which is 0.28 % of the signal range at the used settings. To verify the normal distribution of the measured values, in both figures a Gaussian function has been fitted to the histogram values using a least-squares fit. The standard deviations gained from the fit match with those resulting from a direct calculation by the application of the correspondent statistical formula to the measured values within ± 0.01 mV.

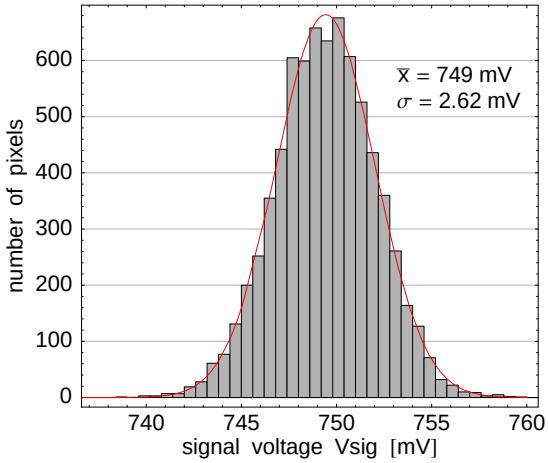


Figure 5.12: Offset distribution within the HDR window at a light intensity of 0.4 W/m^2.

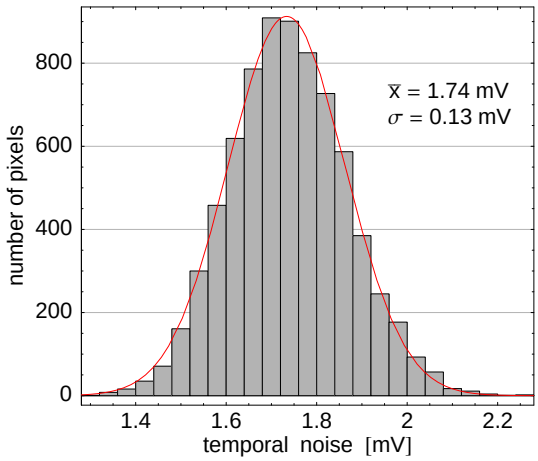


Figure 5.13: Temporal noise distribution measured at 0.4 W/m^2.

Since σ_{temp} and σ_{FPN} should be statistically independent, they should sum up quadratically to the total noise:

$$\sigma_{tot} = \sqrt{\sigma_{temp}^2 + \sigma_{FPN}^2} = \sqrt{1.74^2 + 2.62^2} = 3.15\,\text{mV} \tag{5.6}$$

This is 0.06 mV higher than the value from the distribution in figure 5.11, which can be understood by the fact that the temporal noise is not completely removed in the determination of the FPN.

The knowledge of the noise distribution in figure 5.11 permits to calculate the signal-to-noise ratio that is achieved at a light intensity of $0.4\,\mathrm{W/m^2}$ including the temporal noise as well as the remaining FPN:

$$\mathrm{SNR} = 20 \cdot \log\left(\frac{749\,\mathrm{mV}}{3.09\,\mathrm{mV}}\right) = 47.7\ \mathrm{dB} \tag{5.7}$$

According to the measured response curve (see figure 5.8) and the noise level present even at low light intensities, the SNR decreases until it reaches 0 dB at $1\,\mathrm{mW/m^2}$. For higher intensities the modulation of the SNR that derives from the integration time regulation becomes observable. The step-by-step reduction of the integration time leads to small dips in the SNR at those intensities, where a new integration time is chosen by the chip. Owing to the fact that the integration time is halved at these intensities, the signal is also halved, i.e. the SNR is reduced by 6 dB. With increasing intensity, the SNR rises again until the integration time is reduced once more. This behaviour could be verified up to an intensity of $20\,\mathrm{W/m^2}$ (5 changes of the integration time). The maximum intensity was limited by the measurement setup, which was optimized to guarantee the homogeneity of the stimulating light coming from a diffuse reflector.

5.2.5 Slope Variations

The integration based concept of light detection relies on an approximately linear correlation between the output voltage of the photoreceptor and the incident light intensity. Ideally, the slope that can be assigned to this response curve should be the same for each pixel. As a matter of fact, the deviations in the gain of the in-pixel source follower lead to pixel-to-pixel slope variations. These variations result in a FPN which cannot be eliminated by a subtraction of the reset values.

In the discussion of slope variations for the presented image sensor, it has to be taken into account that the overall response curve shown in figure 5.8 is a result of the adaptive integration time control. In principle, the same part of the pixel response curve (a ”segment”) is run through between two successive regulation steps. The only differences between the successive segments are the different integration time and the different incident light intensity. Therefore, the slope variations within a single segment, i.e. a particular integration time, can be understood as a consequence of the intrinsic gain variations mentioned above. However, if there are any dependencies on the light intensity or on the regulation procedure itself, a measurement of the slope variations across several regulation steps covering a substantial range of intensities should lead to a significant increase of the variations.

The results of these measurements are shown in figures 5.14 and 5.15. They contain the distributions of the slope variations within the HDR window standardized to the mean slope. To suppress deviations due to temporal noise, the slopes were always determined from an average over 100 frames. The Gaussian functions fitted to the histogram values are drawn in red. As can be seen the slope values follow approximately a normal distribution. Within a constant integration time they scatter with a standard deviation of 0.39 %, measured at a medium intensity of $0.6\,\mathrm{W/m^2}$. The right distribution represents the slopes calculated from a first point taken at $0.1\,\mathrm{W/m^2}$ and a second one at $20\,\mathrm{W/m^2}$. Five resets were executed by the integration time control in case of the higher intensity. The resulting standard deviation of the slopes is 0.34 %. The fact of an apparently somewhat smaller deviation can be understood by considering another effect: The assumption of a linear correlation between output voltage of the photoreceptor and the incident light intensity is only valid to a certain degree. The junction capacity of the photodiode depends on the voltage across the junction (see equation 1.9) which results in a small non-linearity. This could lead to a small dependency of the measured slope

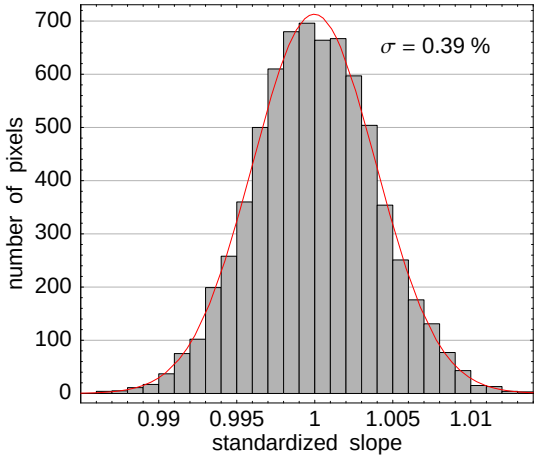


Figure 5.14: Slope distribution within a single step of integration time regulation measured at a medium intensity of 0.6 W/m^2.

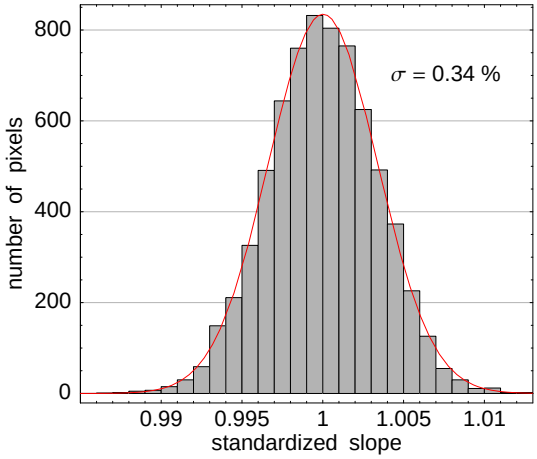


Figure 5.15: Slope distribution measured in an intensity range of 0.1 W/m^2 to 20 W/m^2.

on the position in the response curve where the slope, or one of the points from which it is calculated, is measured.

The standard deviations measured at different points on the response curve were in all cases within a range of $(0.39 \pm 0.05)\,\%$. Therefore, it can be concluded that no significant dependency on the light intensity or the regulation procedure could be observed. The measured slope variations lead to a FPN which is present in the remaining noise discussed in the previous subsection.

5.2.6 Regulation Accuracy

The integration time of a particular pixel is defined by the period between the last reset and the read-out. As discussed in chapter 3 the decision if a particular pixel is reset, relies upon the fulfillment of two conditions: First, the pixel has had to be reset in the previous reset curtain. Second the voltage at the read-out amplifier of the pixel V_{pix} must have decreased below a predefined level V_{comp}, here nearly one half of the voltage range the pixel is able to deliver. Since the second condition is based on a measurement, its limited precision could lead to a false decision. The result of this would be a pixel which is either reset too early or too late. As a consequence of the first case the SNR would be unnecessarily reduced due to the shorter integration time. However, the reduction would only be by a factor of 2 here. The second case would result in a higher deviation from linearity corresponding to the increased non-linearity of the pixel response curve for higher light intensities (cf. section 5.2.2). In the worst case this would lead to an overexposure of the pixel.

Two factors influence the accuracy of the determination of the correct integration time: The offset distribution of the comparators which test if $V_{pix} < V_{comp}$ and the FPN deriving from the unavoidable device-to-device mismatch of the in-pixel source followers. To measure the comparator offset distribution, it is possible to activate a test mode were all signal lines within the pixel array can be fed by a single external voltage. The pixel array itself is switched off in this mode to avoid a deterioration of the applied test voltage V_{test}. The test mode permits to measure only a subset of 40 comparators (total amount: 170), but this is sufficient for a statistical evaluation. V_{test} was changed in steps of 1 mV and the switching point of the individual comparators was observed. The reference voltage was $V_{comp} = 800$ mV, which is the standard setting for all measurements within this chapter. Figure 5.16 shows the measured offset distribution. Owing to the fact that V_{comp} can be

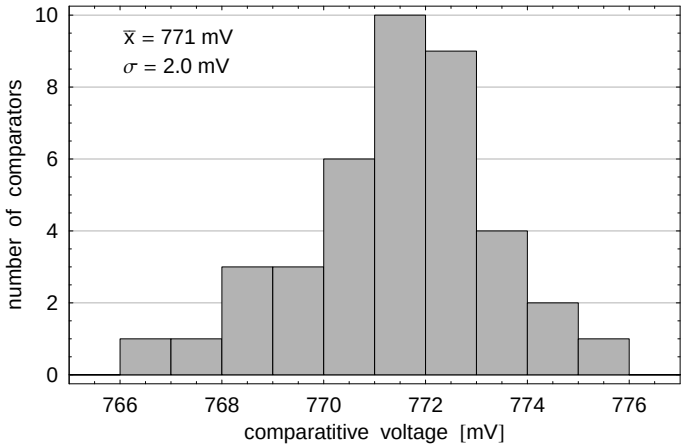


Figure 5.16: Offset distribution of the comparators which are responsible for the verification of the second reset condition $V_{pix} < V_{comp}$.

chosen freely, the mean value is of no relevance here. Its deviation from 800 mV can be explained by the finite resistance of the narrow metal trace for V_{test}, which runs across the chip. The standard deviation of the distribution is $\sigma = 2\,\text{mV}$.

The pixel related FPN amounts to 6.67 mV (cf. section 5.2.3). However, this value is measured at the output of the chip. Since the comparators are connected to the pixel output lines, the variation in V_{pix} has to be calculated. V_{pix} is amplified by the column amplifiers and by the output buffer described in section 4.4.3. The output buffer is operated in a unity gain configuration but the column amplifier gain is 0.74. The fixed pattern noise in V_{pix} is therefore $6.67\,\text{mV}/0.74 \approx 9.0\,\text{mV}$, which is much higher than the offset distribution of the comparators.

Since the FPN and the distribution of the comparator offsets are uncorrelated, the resulting uncertainty for the regulation is clearly dominated by the FPN. The quadratic addition of the standard deviations results in 9.2 mV.

Taking all pixels of the sensor into account ($170 \times 170 = 28900$) a total variation of nearly $\pm 46\,\text{mV}$ (5σ) can be expected. Therefore, choosing V_{comp} about 50 mV above the aimed switching point will guarantee that only the intended range of the response curve is used and no overexposure will occur.

5.2.7 Leakage of Analog Memory Cells

To read out a row of pixels which have finished integration, the integration values as well as the reset values are copied to a bank of analog memory cells (sample-and-hold stages in DS entity). Hereafter, the stored values are sequentially read out to be digitized by the off-chip ADC. A memory cell consists of the storing capacitor and the MOS transistor switches which isolate the capacitor from the signal line and the column amplifier (cf. section 4.4). Since the MOS transistors as well as the capacitor (consisting of a PMOS FET) exhibit leakage currents, a deterioration of the stored voltage during the storage phase occurs. To quantify this effect, a single row was sampled and the stored values were observed over a period of several hundred ms without refreshing these values. The measurement procedure was realized by a change of the VHDL module which controls the sensor: Only a single DS cycle (cf. chapter 3) is executed for the selected row. Next, successive read-outs of the memory cells are carried out. The signal DSEQ is disabled during the read-out to omit charge

equalization between the storing capacitors of V_{res} and V_{int}. Figure 5.17 represents the measured temporal development of the stored voltages. Shown are those two memory cells which exhibit the maximum and the minimum leakage as well as the curve which results from averaging of the read-out results of all memory cells for every single measuring point. 170 data points were taken for each memory cell appearing as a solid line in the graph (points are not connected). The measured voltages were corrected to represent the true voltage stored on the capacitors taking into account the internal gains and offsets known from simulation. As can be seen, the voltage decreases linearly in time, which is a consequence of the approximately constant leakage current I_{leak}. This current can be calculated by

$$I_{leak} = C_h \frac{dU}{dt} \tag{5.8}$$

with the hold capacitance $C_h = 1\,\text{pF}$. The distribution of the leakage currents can be seen in figure 5.18. The decisive measure is the variation of the leakage currents since it can lead to a column FPN if it is too high. The period of time needed for a serial read-out of a row is 157 μs for the frame rate of 30 Hz. Assuming the corresponding storage capacitors exhibit the leakage currents of 0.3 pA and 1.3 pA (limits in figure 5.18), this would result in a difference of 0.16 mV. To measure this difference to a reset voltage of 1.12 V on the capacitor a SNR of 77 dB would be necessary. Since this SNR is not reached by the imager (cf. section 5.2.4) leakage of the memory cells is not an issue in the current implementation. A dependency on illumination has not been investigated explicitly but could not be observed in any measurement under high light intensities: The memory cells are strongly shielded by closed metal coverage. The position of the cells is outside of the pixel array which would also allow additional (and external) shielding.

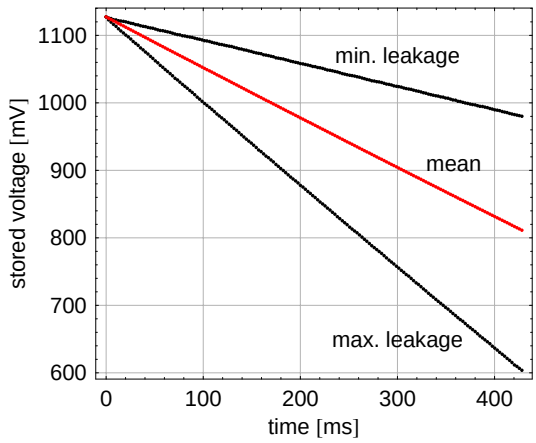


Figure 5.17: Decrease of the stored voltage. The red curve is a mean over all memory cells.

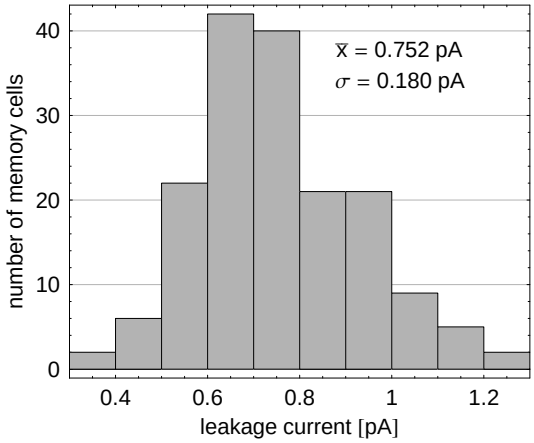


Figure 5.18: Leakage current distribution of the memory cells.

5.2.8 Crosstalk

An important effect influencing the resolution of a sensor is the crosstalk between neighboring pixels. Especially for an HDR imager, which is intended for the use under extreme illumination conditions, this quantity can be of particular relevance. Crosstalk is dependent upon the pixel pitch (here: 7.5 μm), the layout and the process the sensor has been realized in. Hence, the measurement has to be repeated for every new implementation. As described in section 5.1.3 a single pixel was stimulated by a focussed laser spot and the values delivered by the surrounding pixels were recorded. Figure

5.19 shows a bar chart of the spot recorded by the sensor. The chart is restricted to a part of 20×20 pixels to show the details of the measured beam profile. To the right of the bar chart, the corresponding map of time stamps can be seen which the chip used for the regulation of the integration time. In the center of the spot an integration time of $T_{int} = T_{max}/2^z = 33\,\text{ms}/2^{11} = 16\mu\text{s}$ was used. Ideally, the spot should show a symmetrical Gaussian shape but as can be seen in the chart, it is slightly deformed. This is due to its sensitive dependence upon the alignment of the optics and the positioning of the spot on the chip (x,y and tilt).

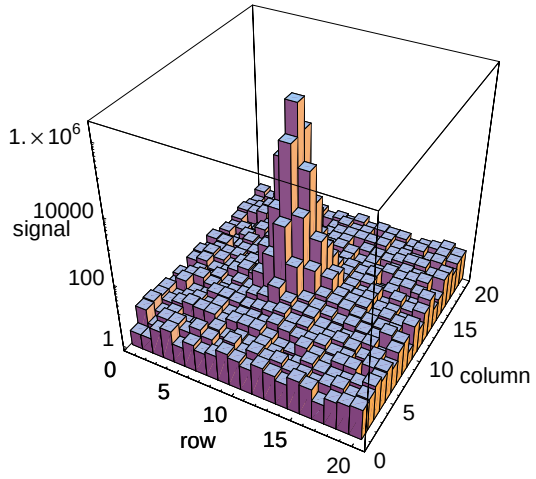


Figure 5.19: Stimulation with a laser spot. The measured signal in a.u. is represented on a logarithmic scale.

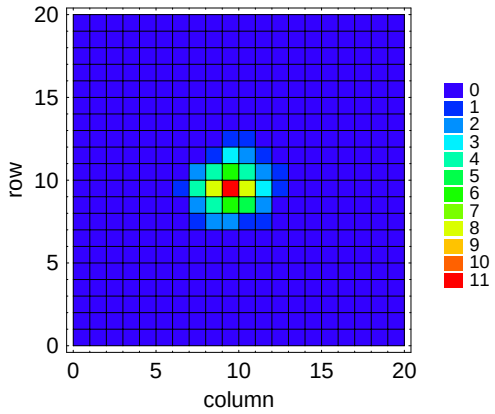


Figure 5.20: Map of time stamps indicating the respective regulation factor $1/2^z$ the chip used in the left image.

In the case of no crosstalk, the adjacent pixels should only show the remaining shape of the spot. Nevertheless the differentiation between parts of the laser spot and the crosstalk can be difficult. Figures 5.21 contains sectional views of two spots with light intensities that differ by a factor of approximately 50. The next neighbors of the central pixel were used for a fit of a Gaussian profile, which is drawn as a red line. Therefore, the fit describes the Gaussian spot with the maximum possible width. The deviations from this profile that can be seen at the other pixels would be the minimum crosstalk in this representation. The diameter of the spot resulting from this fit is $5 \pm 1\ \mu$m (2σ) which is close to the width that is reachable with the spatial filter system. Comparing the brighter spot with the darker one a similar behavior can be observed, whereas the next neighbors of the central pixel in the darker spot show a steeper decrease than in the brighter one.

Nevertheless, when interpreting the results of these measurements it has to be kept in mind, that the visible deviations could also result from a beam profile which is not exactly Gaussian. The results should be rather treated as an estimation of the possible crosstalk. In the HDR images taken with the sensor no disturbances deriving from a strong crosstalk are visible within the HDR window.

5.3 Sample Images

In the following section sample images are presented that were recorded with the sensor developed within this thesis. The recorded images were taken under varying illumination conditions, presenting typical environmental scenes of low to medium DR as well as HDR situations. Different modes of operation were used to illustrate the capabilities of the sensor. In all images the implemented features for an on-the-fly reduction of the FPN (cf. section 4.4) deriving from pixel-to-pixel mismatch and

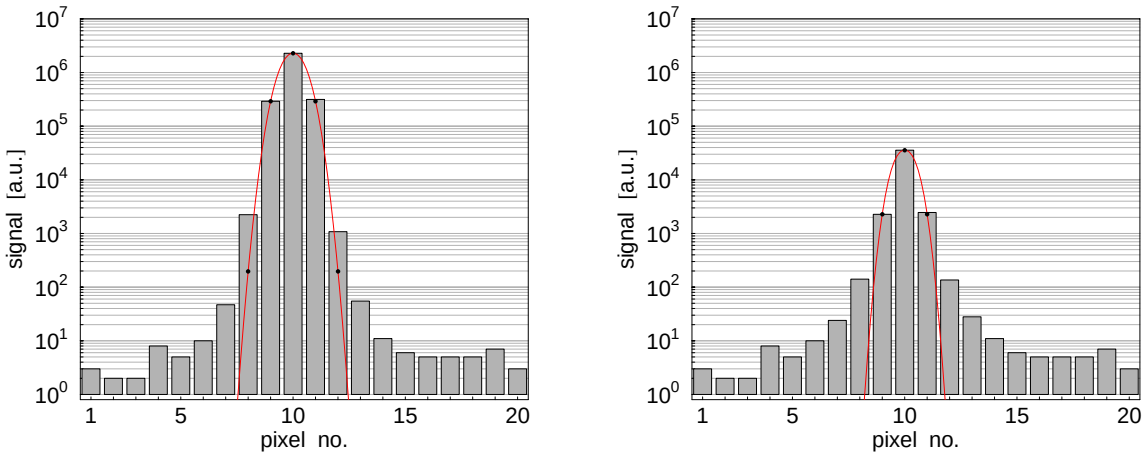


Figure 5.21: Sectional views of the resulting profile after stimulation with a laser spot at different intensities. Left: Laser spot 5.5 decades brighter than the background. Right: Laser spot 4 decades brighter than the background. The red line is a Gaussian fit of the three central pixels.

column amplifier offsets were used. In case of HDR scenes the images had to be mapped to the displayable range of luminances. No other image processing algorithms were used. To illustrate the way HDR images were displayed the first subsection refers to the task of tone reproduction.

5.3.1 Tone Mapping of HDR Images

In the previous section it was shown that images taken with the presented image sensor can exhibit a DR of 134 dB. On the other hand, standard display devices such as CRT/LCD monitors or printers have a DR of luminance of about 40 dB. Therefore the problem comes up of finding a method to represent the HDR images on a display device of low DR, i.e. to compress the occurring luminances to the displayable range of the device. Techniques that accomplish this, are referred to as ”tone mapping” or ”tone reproduction” techniques.

Tone mapping of HDR images is not a new problem. In the history of photography first techniques were developed to bring a real scene to the limited DR range of the photographic paper (e.g. [ADA83]).

In the past decade, considerable work has been done in this field not least on account of recent developments in imaging technology and the increasing demand for displaying HDR scenes in computer graphics [JYR03]. Various approaches have made so far, starting from the attempt to mimic the workings of the human visual system [FER96] or to transfer the methods from photography to computer graphics [DEB97] up to methods that simply rely on visual plausibility. Without going into detail the operators used for tone mapping can be divided into two categories [DEV02]: spatially uniform (”global”) and spatially varying (”local”) operators. While global operators apply the same transformation to each pixel regardless of its position, local operators apply different transformations to different parts of the image. Examples of state-of-the-art local operators can be found in literature [FAT02] [DUR02]. Local operators can provide a strong compression of dynamic range, preserve local contrast and avoid common artifacts. The disadvantage compared to simple global operators is the mathematical complexity that requires an elaborate implementation and high computational power. Therefore, to reach the same frame rate in displaying as the image sensor delivers, global operators were used for the mapping of the recorded data. Two different global operators were implemented in the test software. They are introduced in the following.

Advanced 1/L Scaling

One of the simplest non-linear tone mapping operators used so far is described by the following equation:

$$L_d(x,y) = \frac{1}{1+L(x,y)} \cdot L(x,y) \tag{5.9}$$

with L_d as the luminance on the display at position (x,y) and $L(x,y)$ the luminance in the real world, i.e. the value the image sensor delivers. Despite of its simplicity this operator takes into consideration that the human visual system is much more sensitive to small differences between low luminances than between higher ones. High luminances are scaled by a factor of 1/L while lower ones are scaled by 1. To optimize the use of the displayable range of luminances it would be desirable that the maximum luminance of the display L_{dmax} is adapted to a finite luminance L_{white}, e.g. the maximum luminance occurring in the scene, instead of an infinitely high value. This can be reached by the following extension of 5.9 [REI02]:

$$L_d(x,y) = \frac{\left(1+\frac{L(x,y)}{L_{white}^2}\right)}{1+L(x,y)} \cdot L(x,y) \tag{5.10}$$

The resulting operator allows high luminances to burn out in a controllable fashion. It has been implemented in the test software. Figure 5.22 shows the resulting tone mapping curve for some values of L_{white}. For $L_{white} \to \infty$ it reverts to equation 5.9. The operator allows a very efficient implementation and is sufficient for many HDR scenes. However, it has to be kept in mind, that for very high dynamic range images some details still may get lost. In literature, equation 5.10 is used as a starting point for the development of a more complex local operator [REI02].

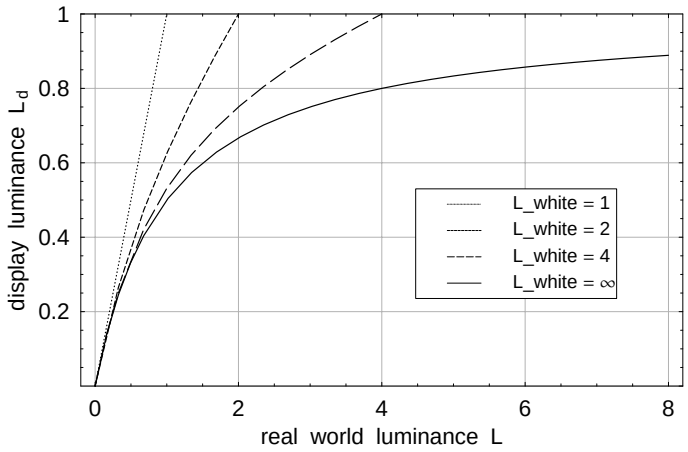


Figure 5.22: Tone mapping curve (equation 5.10) for different values of L_{white}.

Adaptive Logarithmic Mapping

According to the Weber-Fechner law[17] the response of the human visual system to a stimulus can be approximated by a logarithmic function. Drago et al. developed a global operator that is based on

[17] Weber-Fechner law: $B = k \cdot \ln(L/L_0)$ with B describing the response of the human visual system, L the stimulus luminance and L_0 the luminance of the background. k is a constant factor.

logarithmic compression in order to mimic this response and to reach a perceptually accurate tone reproduction [DRA03]. For low luminances a scaling with $\log_2(L(x,y))$ is done, whereas for high luminances $\log_{10}(L(x,y))$ is used. To combine the characteristics of both functions a tone mapping function is used that reaches a smooth interpolation among the logarithmic bases:

$$L_d(x,y) = \frac{L_{dmax} \cdot 0.01}{\log(L_{white}+1)} \cdot \frac{\log(L(x,y)+1)}{\log\left(2+\left(\left(\frac{L(x,y)}{L_{white}}\right)^{\frac{\log(b)}{\log(0.5)}}\right)\cdot 8\right)} \tag{5.11}$$

Designations are the same as in equation 5.10. The parameter b can be used to adjust the compression of high values and visibility of details in dark areas. Even though this operator is more costly than 5.10 it is still fast enough even for interactive applications if it is implemented on a current standard PC. Most of the HDR images of the next section were mapped using 5.11.

5.3.2 High Dynamic Range Scenes

Within the following subsection scenes of medium and high dynamic range are shown. The presented images were taken with activated integration time control following the control scheme discussed in chapter 3. To illustrate the different stages of image acquisition and the working of the integration time control, different representations of a sample scene are discussed next.

Visualization of Integration Time Control

For a scene recorded with activated integration time control the image data is delivered by the sensor via two separate channels of information: First, the analog integration values from the pixels and second the 4 bit digital time stamps z which indicate the corresponding integration times. These values can be used to visualize the working of the adaptive integration time control.

To avoid a restriction to small HDR window for the illustration, the full frame HDR mode has been used. In this mode, the HDR window is shifted over the entire sensor array in four steps. Afterwards, the resulting four HDR parts of the image are put together to a single image at the full resolution in software. The sample scene used in the following shows a typical environmental situation of medium DR: In the front part of the scene a road sign stands in the bright sun. On a parking lot behind the sign a car stands in the shadow of a tree, others park in the sun. In Figure 5.23 different views of this scene are presented.

The first picture shows the image recorded without adaptive integration time control. All pixels integrated for the same period of 33 ms here. This mode corresponds to the ordinary operation of a standard CMOS APS sensor. As can be seen, the car standing in the shadow is correctly exposed. On the other hand, details of the bushes in front of the road sign, of the sign itself and of the cars standing in the sun are lost due to overexposure.

Pictures 2-4 were taken with activated integration time control. Picture 2 shows a map of the different time stamps that were used for the regulation of the integration time. The higher the intensity at a particular pixel location, the shorter is the used integration time ($T_{int} = T_{max}/2^z$). For the whole scene only four stamps were needed. White areas indicate a time stamp of $z = 0$, which means that the maximum integration time of $T_{int} = T_{max} = 33$ ms could be used here without overexposure of pixels. In some regions, a time stamp $z = 1$ is indicated, but in picture 1 the pixel at this location is just not overexposed. This can be explained by the chosen threshold which determines the use of a shorter integration time. The threshold is defined by the reference voltage of the comparators V_{comp}: To use only the linear part of the response curve of a single pixel, V_{comp} is chosen slightly above

one half of the available voltage range (cf. section 4.2.1). Consequently, if the output voltage of the pixel is just slightly smaller than V_{comp} at the time of comparison, it would in principle be possible to continue the integration without the risk of overexposure but at the price of a higher deviation from linearity.

In picture 3 the raw image data neglecting the time stamps is shown. The integration time for each pixel was adjusted by the sensor so that the resulting output voltage lies within the valid voltage range. Regions with the same integration time show the fine shades of brightness of the scene in the real world. However, as a consequence of the neglected time stamps the absolute luminance values are not correct which results in sharp edges separating the different regions from each other. Taking the time stamps into account means to multiply the values of picture 3 with 2^z were the corresponding stamp z is taken from the map in picture 2. This is equivalent to a simple bit shift of the digitized integration value by the stamp value. The result is a linear representation of the scenes luminance values, containing the whole image information delivered by the sensor and spanning the DR of the scene in the real world (within the limits of the sensor).

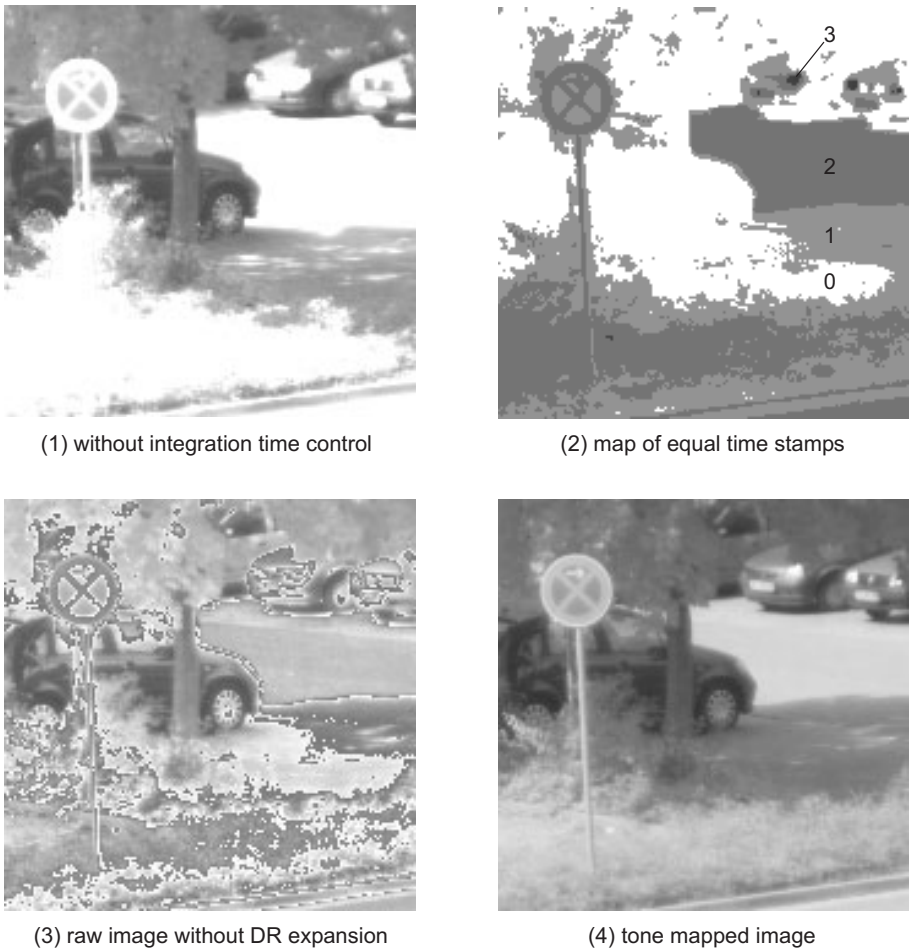


Figure 5.23: Different representations of a recorded scene of medium dynamic range. Picture 1 shows the scene recorded without adaptive integration time control, whereas pictures 2-4 were taken with activated control. Picture 2 illustrates the use of different integration times and picture 3 shows the pixel values neglecting the individual integration times. In the last picture, the final image mapped to the displayable range of luminance values is shown.

Even though this is the desired representation of the information for many image processing applications, it is not directly displayable as discussed in the previous section. It has to be compressed to the DR of the display device by the usage of a tone mapping operator. This has been done in picture 4 (logarithmic adaptive mapping). The result is an image coming close to the natural human perception of the scene.

It should be kept in mind here, that the representation in picture 3 has a significant advantage over the tone mapped image: Details that inevitably get lost due to the insufficiency of the tone mapping operator in preserving local contrast, are clearly visible in regions where the same integration time was used. The higher the total DR of the scene, the stronger is this effect. This illustrates that the linear image data contain much more contrast information than observable in the tone mapped image.

Sample Images

The key feature of the realized image sensor is its ability to cope with situations exhibiting a high optical dynamic range. As illustrated previously, this is reached by a pixel level adaptation of the integration time. Without multiple read-outs of the array this can be done in a window of 85×85 pixels, which can be moved across the sensor array from frame to frame. In Figure 5.24 sample images of high dynamic range scenes are shown. Pictures 2-4 are examples for the active HDR window (taken at a frame rate of 30 Hz), whereas pictures 5 and 6 where taken by a shift of the HDR window over the entire sensor array (4 quadrants) and a subsequent assembly of the 4 windows in a single image.

Picture 1 is the only one that was taken without adaptive integration time control. It has been recorded under indirect sunlight coming through the window and is shown to illustrate the arrangement that is used for the HDR situation in picture 2. In the top left corner a lamp with a light bulb can be seen. The lampshade is adjusted in direction of the sensor chip. Beside the lamp is a white sheet with the KIP[18] logo printed on. It is behind the lampshade so that no direct light coming from the bulb reaches the sheet. For picture 2 the room was darkened and the lamp was switched on. Only the stray light reflected in the vicinity illuminates the sheet. In the recorded image, the brightness outside of the HDR window was increased to make the logo visible: As a consequence of the low light intensity the remaining noise gets noticeable. On the other hand, inside the HDR window, which is positioned on the lamp, the bulb and its filament are clearly visible. The DR of the scene is nearly 95 dB and the used integration times of the individual pixels range from 33 ms down to 4 μs. The contents of the HDR window were always displayed by the use of the tone mapping operators discussed in section 5.3.1 (here adaptive logarithmic). This also explains the sharp edge between the window and the rest of the image: Black is assigned to the darkest regions within the window. A certain amount of local contrast that is on principle available in the data is lost in the displayed window due to the large DR that has to be covered within this region. Certainly the tone mapping could have been applied to the whole image, which would result in a smooth representation except of the overexposed parts outside of the window. On the other hand, the chosen way to carry out a simple linear contrast adaptation outside the window offers the opportunity to visualize also the details in the darker parts of the scene.

Pictures 3 and 4 both contain the same scene, but each with the HDR window at a different position. The scene consists of a bright incandescent lamp (carbon filament) with a lampshade which illuminates a printed page that is turned over. In the shadow of the page the writing is still visible. In picture 4 the window covers a part of the filament as well as a part of the writing. Both regions are visible, even though the local contrast in the displayed window is reduced as a consequence of the tone mapping. The dynamic range of this scene is nearly 80 dB.

[18] Kichhoff-Institute for Physics, University of Heidelberg

Pictures 5 and 6 were taken by four subsequently acquired HDR windows as described above. In picture 5 a close-up of the bright carbon filament is presented, whereas picture 6 deals with an outdoor scene of medium dynamic range on a parking lot. Even the reflections of the sun in the cars headlight and in the street lamp were correctly recorded (no overexposure).

5.3.3 Low Dynamic Range Scenes

In most of the applications for HDR imagers not only situations with extreme illumination conditions have to be handled. The ability of the sensor to cope with scenes of lower contrast and low dynamic range can also be of high relevance. Therefore, in this subsection sample images of such scenes are discussed to illustrate the performance of the imager under these conditions.

In figure 5.25 the recorded images are shown. They were taken at a read-out speed of 30 frames per second. No integration time regulation has been used here, which leads to the same integration time of 33 ms for every pixel. This operational mode is equivalent to the ordinary operation of a standard CMOS APS sensor and results in a DR of 48 dB for the presented image sensor (cf. section 5.2.2). Because of the low DR none of the tone mapping operators discussed in section 5.3.1 had to be used. Instead of this, a simple linear mapping was possible. Just like in all the presented images the FPN could be reduced on-the-fly by the use of the implemented double-sampling (cf. section 4.4). The resulting images are free of visible artifacts. In picture 1 (a bunch of screws) and picture 2 (hand holding a roll of solder wire) small regions of overexposure can be observed. The light reflected from the metal surface in picture 2 exceeds the available DR of the sensor making it impossible to see parts of the winding. In the third picture a couple of electronic devices can be seen, whereas picture 4 shows a wooden ruler, whose scaling is still resolved. Picture 5 represents the lens (F/3.5, 14mm) that was used to record all other images. The diameter of the lens is 20 mm. Picture 5 is the only picture for which a different lens (F/2.8, 10 mm) was mounted. The last picture (no. 6) shows a person standing in the optical laboratory of the institute illuminated by the light coming through the window on a sunny day (no direct sunlight on the persons face).

Regarding the resolution of the presented sample images in this chapter it should be borne in mind, that the resolution of the prototype imager is only 170×170 pixels. Taking the size of the printed images into account, a resolution of 80 dpi is reached[19]. This is comparatively low with respect to the resolution of 150-300 dpi, which is provided by commercial printing studios for the reproduction of digital images in consumer quality.

5.3.4 Averaging

Averaging of neighboring pixels provides the opportunity to take an image at a reduced resolution maintaining the initially chosen region. This way, it is possible to analyze the coarse image structure and to select regions of interest that should be read out at a higher resolution with the following acquisition. The imager offers different averaging modes. Basically, the averaging is accomplished on chip by connecting the photodiodes of the corresponding pixels with each other (cf. section 4.2.3). The user can choose between averaging of 2×2, 4×4 or 8×8 pixels. In figure 5.26 the result of the different modes applied to a scene of low DR can be seen. For increasing levels of averaging a certain loss of contrast is observable for small structures. For instance, the needle of the lamp current indicator fades from step to step. This can be explained by the fact that an increasing number of brighter pixels is included in the averaging process. A similar effect can be observed at the top edge

[19] The resolution of the used printer is 600 dpi.

of the instruments casing: The sloping edge, which is clearly resolved in the first picture, converts to a line of shading gray.

Averaging can also be used in conjunction with HDR expansion. In this case, the HDR window can be expanded over the entire sensor array within a single frame. For 2×2 averaging for example the number of time stamps that have to be stored is only one quarter of the total number of pixels realized in the image sensor. Therefore, the SRAM size is sufficient to hold the stamps for the whole array. This results in a kind of ”overview mode”: At the price of a reduced resolution, the whole scene can be inspected in a single view with HDR. Certainly this can also be used for the 4×4 and 8×8 modes. Figure 5.27 gives an example for a HDR scene recorded without averaging in comparison to the picture of the same scene with 2×2 averaging and activated HDR expansion. In the left picture details of the carbon filament can only be seen inside the HDR window, other parts of it are overexposed, whereas most of the glass of the bulb has a brightness that can also be reproduced outside of the window. In the right picture the whole scene is reproduced correctly but with a reduced resolution. The tone mapping operator had to be used here for the whole sensor array, which resulted in a decrease of contrast in the darker parts of the picture in comparison to the left one. Nevertheless, on principle the contrast information is available in the data as discussed in section 5.3.2.

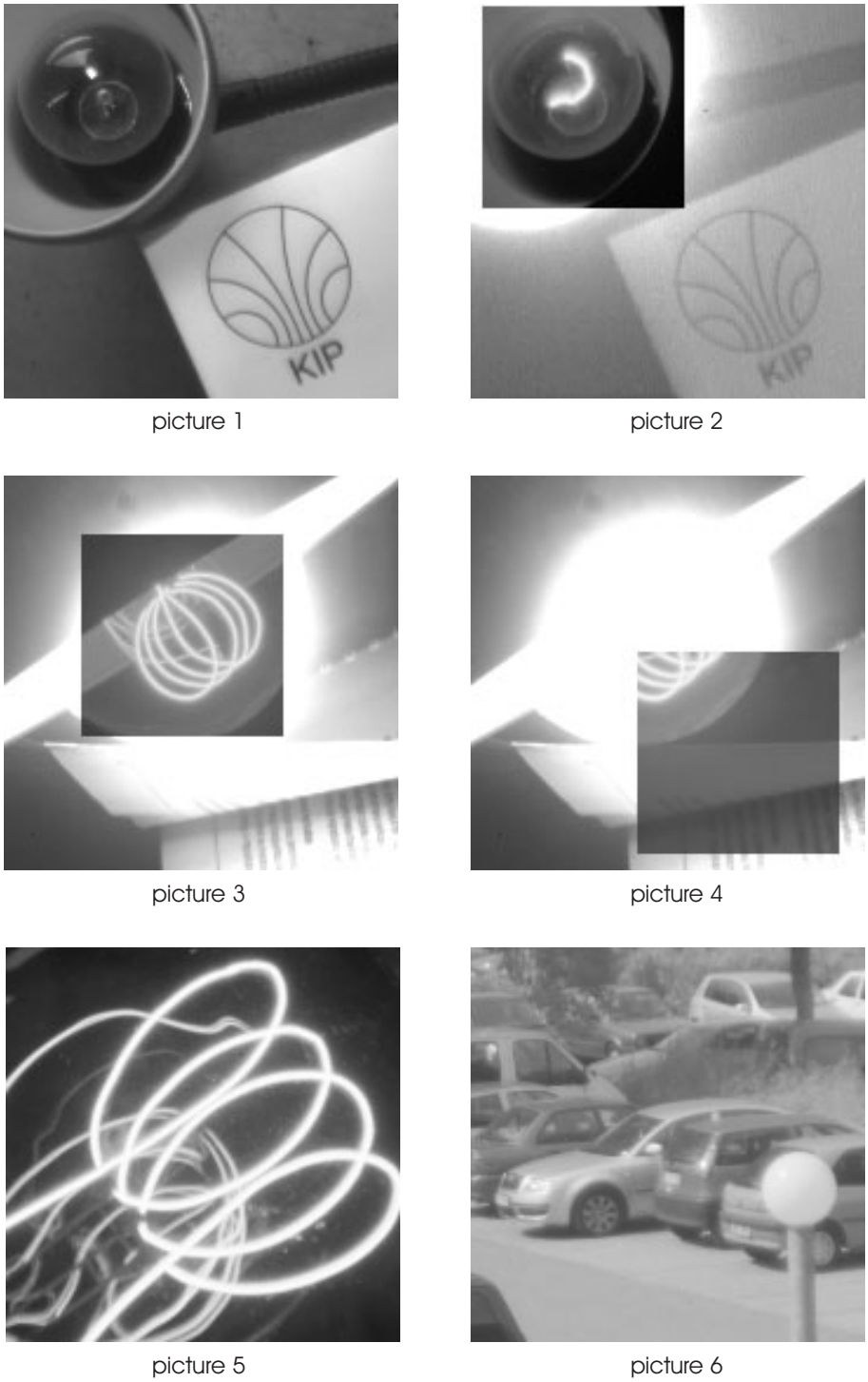
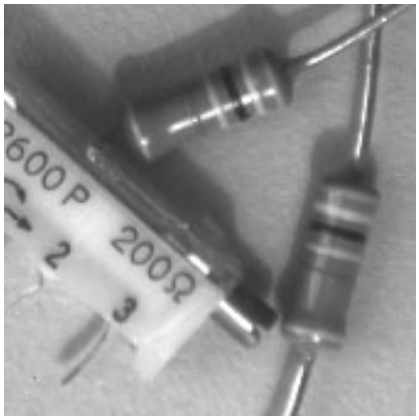


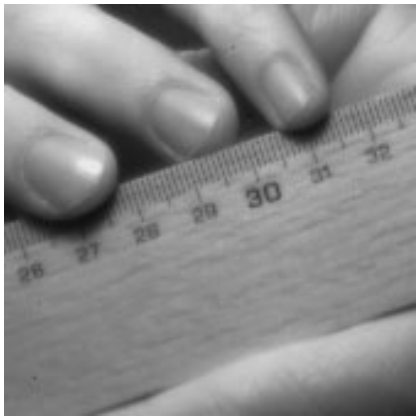
Figure 5.24: Images of high dynamic range scenes taken with the realized sensor chip. All pictures, except of picture 1, were taken with adaptive integration time control. Picture 1 is a scene of low DR (lamp switched off). In Picture 2 the same scene is recorded under extreme illumination conditions (lamp switched on, room darkened). Pictures 3 and 4 both show the same high dynamic range scene, with the HDR window at different positions. Pictures 5 and 6 were recorded by a successive shift of the HDR window over the sensor area.

picture 1

picture 2



picture 3



picture 4

picture 5

picture 6

Figure 5.25: Images of indoor scenes exhibiting a low DR taken with the realized image sensor. All images were recorded without adaptive integration time control.

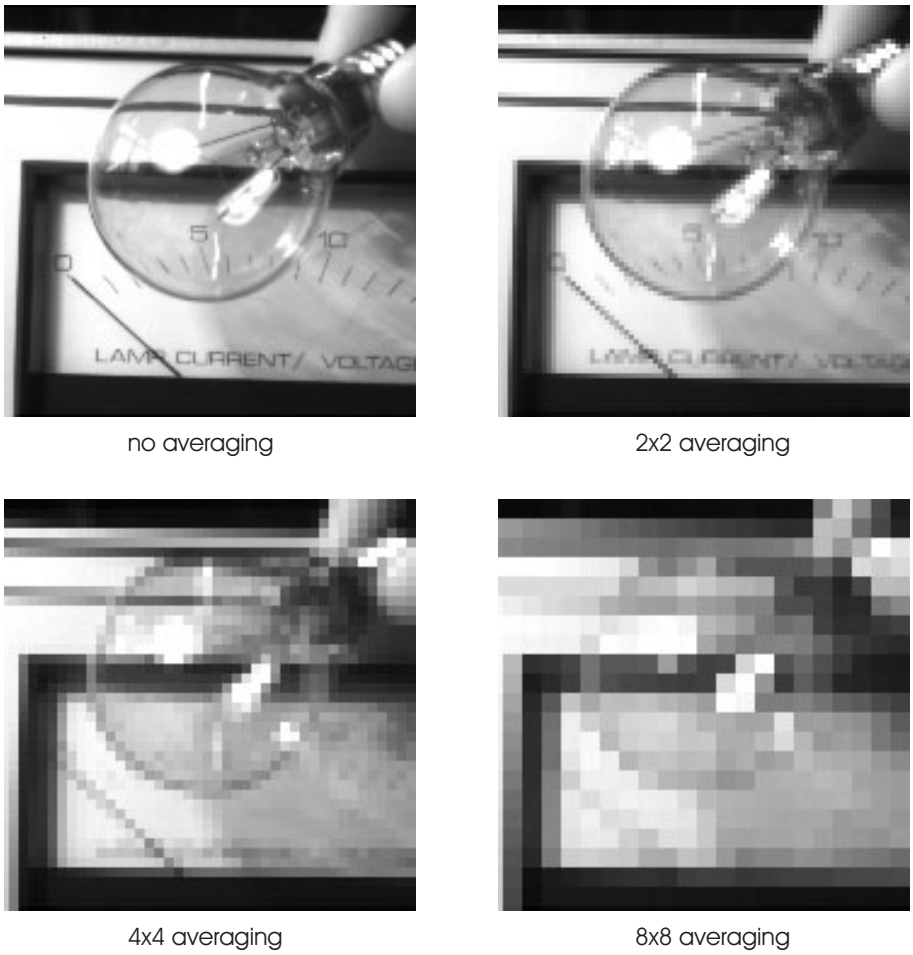


Figure 5.26: Low dynamic range scene recorded with different averaging levels.

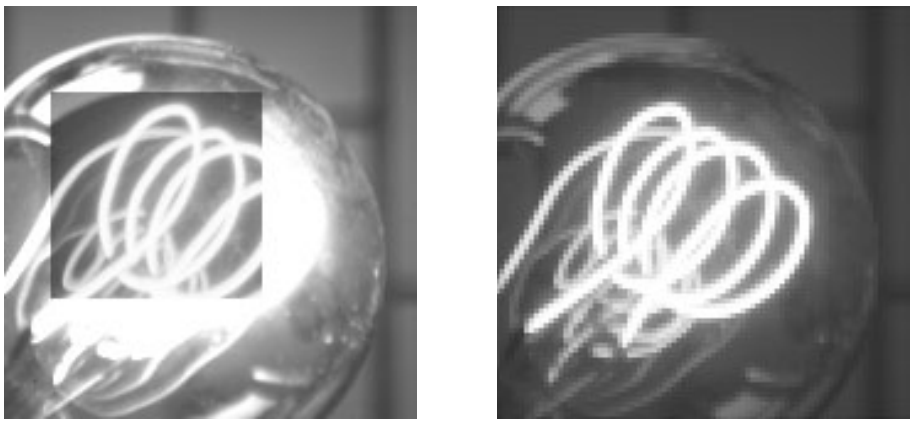


Figure 5.27: Bright incadescent lamp (carbon filament) recorded with different modes of operation: Left: Activated HDR window, no averaging. Right: 2×2 averaging with HDR expansion.

Conclusion

Within the scope of this thesis, a high dynamic range CMOS image sensor with local adaptive integration time control has been developed, implemented and tested. The image sensor has a resolution of 170 ×170 pixels. High dynamic range is reached within a region of 85×85 pixels in size, which can be freely moved across the entire sensor array. The possibility of on-chip averaging of neighboring pixels offers the opportunity to expand the high dynamic range region over the entire sensor array.

A total dynamic range of 134 dB is achieved at a maximum integration time of 33 ms and a frame rate of 30 Hz. With these settings the light intensity covered ranges from 1 mW/m^2 to 5 kW/m^2. The chip offers the possibility to reduce the fixed pattern noise by double-sampling. As a result, a signal-to-noise ratio of 48 dB is achieved, which is limited by the slope variations of the pixel response curves.

The realized image sensor has significant advantages over conventional CCD's found in most commercially available cameras. The high dynamic range of the sensor enables the acquisition of scenes under different illumination conditions without the need of a mechanically adjusted aperture. Furthermore, owing to the fact that the integration time is adjusted individually for each pixel, it is possible to acquire scenes of much higher contrast than possible with a standard CCD. Thanks to the use of CMOS technology, the imager has further advantages over CCD's with respect to power consumption and system integration possibilities. The power consumption of the realized imager amounts to 2.5 mW at a frame rate of 30 Hz.

In the past decade several concepts for widening the dynamic range were suggested (see chapter 2). The concept realized in this thesis offers significant advantages in either pixel size, image quality and/or required effort in post-processing depending on the concept it is compared with.

One branch of concepts consists of sensors using logarithmic compression. There are two serious drawbacks of these sensors: First, the logarithmic compression leads to the fact that the whole dynamic range of the sensor (mostly about 6 decades) is mapped to the small output signal range (mostly 2-3 decades). The result is a reduced sensitivity for local contrasts in comparison to linear integration based concepts. It should be noted here, that the signal compression of logarithmic sensors is advantageous for some applications. An example for this is the representation of the recorded image on a display device: The output signal does not has to be compressed further to lie within the low dynamic range of common display devices[20]. However, the logarithmic compression turns out to be disadvantageous for other applications requiring a linear input (e.g. color calculation). The second drawback results from the strong fixed pattern noise that comes with the use of transistors operating in subthreshold region. The solutions suggested to overcome this problem either rely on off-chip calibration by the use of external facilities, which increases the overall system size and power consumption, or they use on-chip calibration methods, which lead to a significant increase of pixel size (see chapter 2).

In comparison, the dynamic range expansion concept presented in this thesis requires only four

[20] Common display devices such as CRTs and LCDs achieve a dynamic range of 2-3 decades.

transistors per pixel (neglecting extra devices for averaging), which is only one transistor more than in the standard active pixel sensor. Thanks to the possibility of double sampling, an effective reduction of the FPN is feasible without an increase of pixel size or the necessity of off-chip post-processing.

Several concepts for dynamic range expansion were suggested which also rely on the use of different integration times. Some of these are based on the local control of integration time by an in-pixel circuitry, which inevitably results in large pixel sizes. Others perform a global control of the integration time in conjunction with a multiple read-out of the entire sensor array after expiration of the respective integration times. While a multiple read-out permits the realization of small pixels and a high SNR, it requires a costly post-processing of the image data and reduces the achievable frame rate (see section 3.1.2). As a matter of course, the maximum frame rate that is achievable with a sensor that adjusts the integration time is $1/T_{max}$, with T_{max} the given maximum integration time. This is achieved for the realized imager. The entire sensor array has to be read out off-chip only once within the period T_{max}, which results in a lower required read-out speed. Additionally, in contrast to the multiple read-out technique there is, in principle, no need for post-processing to reconstruct the high dynamic range image: Synchronous to the read-out of each analog pixel value the corresponding time stamp z is delivered. The pixel value simply must be multiplied by 2^z to obtain the final result. This is equivalent to a bit-shift of the digitized value by the amount of digits given by the time stamp, which can easily be done on-chip.

Further integration based concepts rely on capacity adjustment or dual-sampling. Both of these suffer from a strong dip in the SNR close to the light intensities where the respective quantity (capacity or integration time) is readjusted. In comparison, the adaptive integration time control with its multiple readjustments provides a consistently high SNR with only small dips (see sections 3.1.1 and 5.2.4).

The major drawback of the implemented high dynamic range enhancement scheme is the necessity of an on-chip memory to store the time stamps, which indicate the individual integration times of the respective pixels. Depending on the size of this memory, it will lead to higher production costs due to an increased total chip size and an eventually lower yield. There are two arguments that take the edge off this problem: First, technology scaling leads to an ever decreasing minimum lithographic feature size. Since the pixel size does not shrink to the same extend as the size of a memory cell, the relative area consumed by the memory will further decrease with progress of technology [WON96]. Second, the amount of memory cells needed for the regulation can be reduced by the restriction to a moveable region of high dynamic range, which was demonstrated here. The memory size can be further reduced by the implementation of an elaborated access control as was shown in section 3.3. In principle, the required memory for the execution of integration time control for the entire sensor array can be reduced to about 1 bit per pixel. A realization of this access control as well as the on-chip integration of the row sequence calculation would be desirable for a future version of the imager.

A restriction for the use of the presented regulation scheme arises from the course of the regulation itself: Since it is based on rolling shutters, distortions of the recorded image result if motion exists that is faster than the frame rate. However, owing to the fact that an array is read out line by line anyway (in any order) this is the case for every sensor concept unless it employs a global shutter.

The image sensor realized in this thesis demonstrates the feasibility and performance of a high dynamic range CMOS image sensor that is based on a local adaptive integration time control. The sensor has several possible applications ranging from consumer electronics to industrial and scientific imaging were high dynamic range is required. Low power consumption and possible system integration make it an appropriate solution especially for mobile systems. Currently, the use of the imager as an optical input channel for an artificial neural network [SCH04] is in preparation.

Appendix A

Pad Diagram and Signal Description

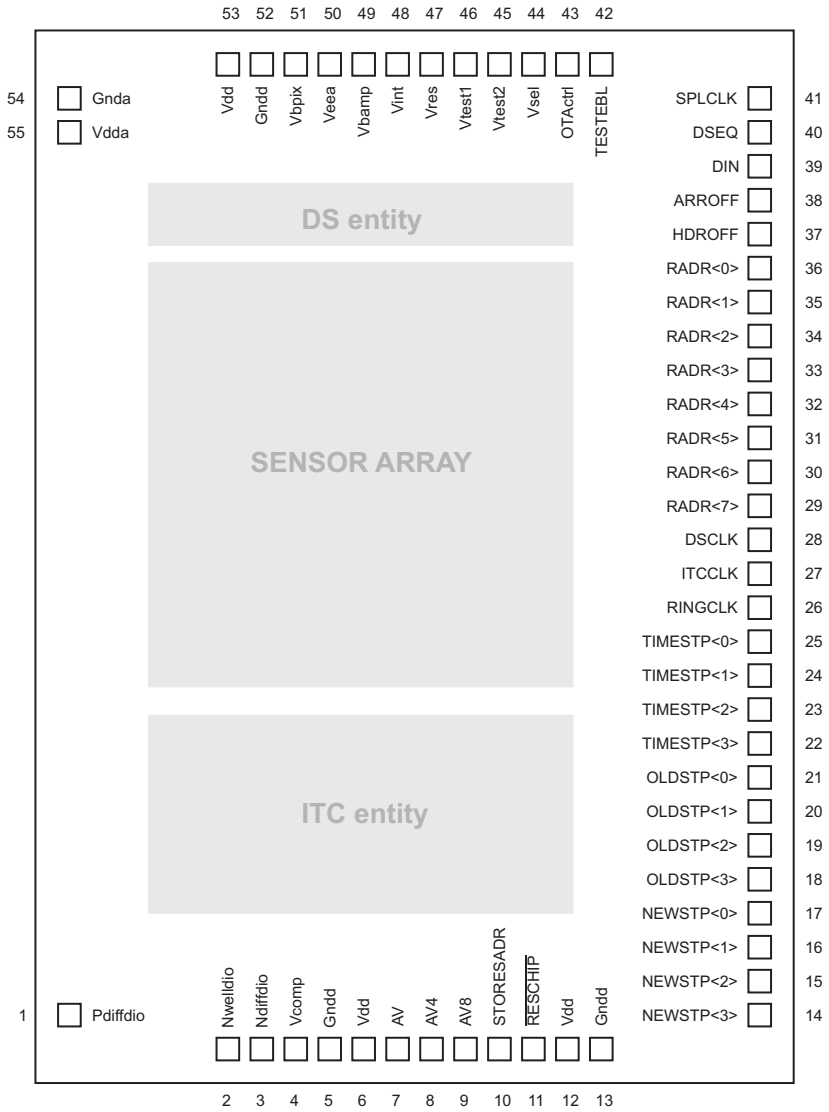


Figure A.1: Pad diagram of the image sensor.

pad no.	pin name	type	description
1	Pdiffdio	control voltage	connected to the p^+-diffusion of the p^+-nwell photodiode teststructure
2	Nwelldio	control voltage	connected to the nwell of the p^+-nwell and nwell-substrate photodiode teststructure
3	Ndiffdio	control voltage	connected to the n^+-diffusion of the n^+-substrate photodiode teststructure (with and without silicidation)
4	Vcomp	control voltage	comparison voltage for the comparators in the ITC entity (typical value 800 mV). Determines the threshold for a reset.
5,13,52	Gndd	ground	ground potential for the digital components of the chip (e.g. SRAM, decoders, shift registers, digital logic)
6,12,53	Vdd	supply	supply voltage (2.5 V) for the digital components of the chip (e.g. SRAM, decoders, shift registers, digital logic). Low FPN requires Vdd=2.7 V, see section 4.2.1.
7	AV	input	control line for averaging of 2×2 pixels
8	AV4	input	control line for averaging of 4×4 pixels (if activated simultaneously with AV)
9	AV8	input	control line for averaging of 8×8 pixels (if activated simultaneously with AV and AV4)
10	STORESADR	input	stores the values RADR[6:0] to a register as new position for the HDR window (rising edge)
11	$\overline{\text{RESCHIP}}$	input	reset of the cycle control and all registers on the chip (active low)
14-17	NEWSTP[3:0]	input	time stamp of the new ITC cycle that has to be executed
18-21	OLDSTP[3:0]	input	time stamp of the ITC cycle before the current one (ordinary operation: OLDSTP=NEWSTP-1)
22-25	TIMESTP[3:0]	output	time stamp that indicates the integration time of an individual pixel
26	RINGCLK	input	clock input for the serial read-out of the time stamps
27	ITCCLK	input	clock input for the execution of an ITC cycle
28	DSCLK	input	clock input for the execution of a DS cycle
29-36	RADR[7:0]	input	address bits for the row selection in the sensor array and the SRAM as well as for the position of the HDR window

Table A.1: Description of the I/O pads (first part).

pad no.	pin name	type	description
37	HDROFF	input	disables the internal signals STPWREBL, SENSECLK, RINGWREBL and $\overline{\text{RAMDECEBL}}$ (cf. section 4.5, required for the read-out of rows lying outside of the HDR window)
38	ARROFF	input	disables the internal signals ROWSELEBL and RESVEBL (cf. section 4.5, required for comparator offset measurements)
39	DIN	input	input bit of the column selection shift register in the DS entity
40	DSEQ	input	connects the capacitors with each other that store the sampled integration and the reset value of the selected column (necessary to measure the column amplifier offset differences)
41	SPLCLK	input	clock input for the serial read-out of the sampled integration values and reset values
42	TESTEBL	input	connects the test voltages Vtest1 and Vtest1 to the corresponding pixel output lines (test option)
43	OTActrl	control voltage	bypass for the bias circuitry of the output operational amplifiers to regulate the bias current (test option, cf. figure 4.31)
44	Vsel	control voltage	voltage used to select a specific row in the sensor array (typical value 2.7 V)
45	Vtest2	test voltage	voltage input to test the performance of the comparators and the DS entity (test option)
46	Vtest1	test voltage	voltage input to test the performance of the comparators and the DS entity (test option)
47	Vres	output	output for the sampled analog reset values
48	Vint	output	output for the sampled analog integration values
49	Vbamp	bias voltage	column amplifier bias (typical value 1.7 V)
50	Veea	supply	supply voltage (2.5 V) for the column amplifiers and reference voltage for the capacitors in the analog memory cells
51	Vbpix	bias voltage	bias for the pixel read-out source follower (typical value 800 mV)
54	Gnda	ground	ground potential for the analog components of the chip (e.g. sensor array, output amplifier)
55	Vdda	supply	supply voltage (2.5 V) for the sensor array and the output amplifiers

Table A.2: Description of the I/O pads (second part).

Appendix B

Spectral Curves

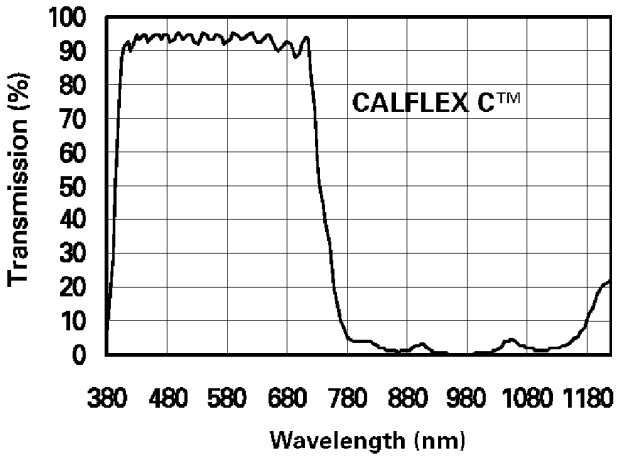


Figure B.1: Spectral curve of the bandpass filter [LIN04] that is used in conjunction with the Xe arc lamp.

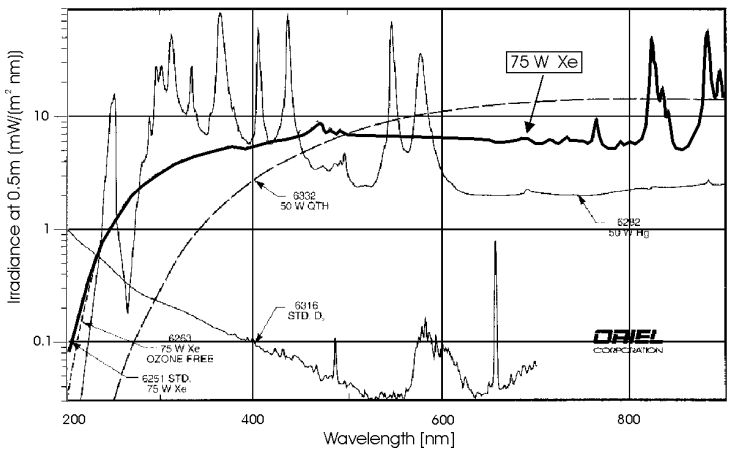


Figure B.2: Spectral curve of the 75 W xenon arc lamp used for the measurement of the response curve [ORI94].

Appendix C

Removal of the Polyimide Layer

For protection against mechanical damage, the chip manufacturer covers the dies with a layer of polyimide. As the chip was produced within a multiproject[1] wafer run, it was not possible to omit this production step. Polyimide is a synthetic polymer resin, which is resistant to high temperature, wear and corrosion. In electronics industry it is also extensively used for flex circuits and printed circuit boards. In the CMOS process the image sensor has been realized in, the polyimide layer has a thickness of approximately 6 μm. It is colored brown but transparent enough so that structures of the chip lying underneath can be recognized under the microscope. In order to reach a maximum quantum efficiency, the polyimide has to be removed. The following procedure was used for the removal. First, the dies are taken into a bath of ethylenediamine ($C_2H_8N_2$) for 30 minutes. The temperature of the bath has to be kept at 100°C for the whole period of time. Ethylenediamine is a medium strong lye (pH=13 at T=20°C). The vaporization at 100°C is very strong. Owing to the fact that the steam is explosive and irritates the respiratory system, the procedure has to take place in a chemical laboratory with the necessary equipment. The liquid is very sticky, therefore every chip should be kept in a separate test tube to avoid contact with other dies. The progress of the removal can be observed as a change in color. The etching stops when the underneath SiO_2 layer is reached but an unnecessarily prolonged etching increases the risk of a damage of the chip (e.g. of the pads). After successful removal of the polyimide layer, the dies must be washed with 2-propanol to stop further etching and to remove the remaining lye. A following ultrasonic bath in distilled water cleans the surface of the sensor to omit disturbing structures from the drying 2-propanol. The success of the removal can be seen on the photomicrographs shown in figure C.1 and C.2. The shown region is located in the lower right corner of the chip (compare with chip photo in figure 4.38). After the removal, the sensor array no longer shows a color of dark brown. It now appears to be white, maybe as a consequence of the reflection from the connection lines (M1, M2 and silicided poly). Most of the chip does not show such a clear change in color. Apart from the sensor array it is nearly completely covered with M3 (last metal layer) for light protection, which seems to be responsible for the observable grainy surface structure.

[1] In a multiproject wafer run designs of different chips are placed on the same wafer for economical reasons (e.g. shared costs for photomask production).

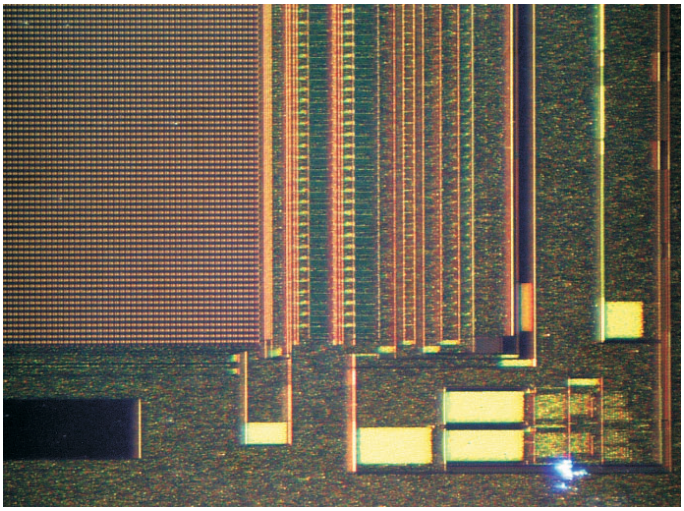


Figure C.1: Photomicrograph of a region located in the lower right corner of the chip. The polyimide is observable as a brown transparent layer, which covers the entire chip.

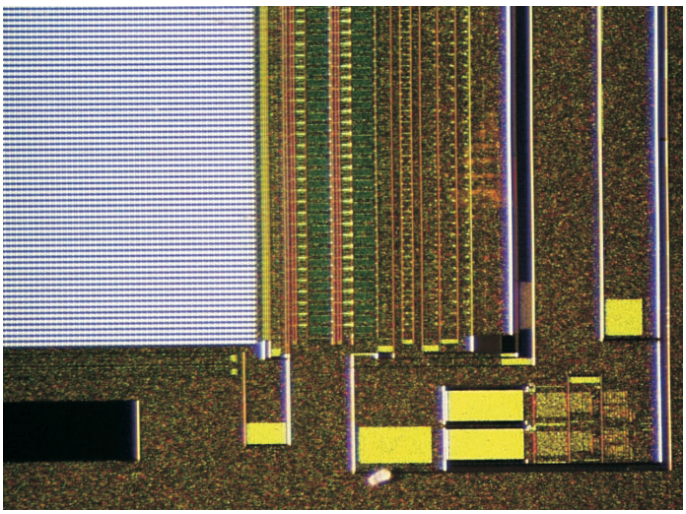


Figure C.2: The same region as shown in figure C.1 after removal of the polyimide.

Appendix D

Parameters for the Integration Time Control

Exemplary, three different ranges (”fast”, ”medium” and ”slow”) for the automatic regulation of the integration time have been implemented. The parameters of the individual ranges are given in table D.1. The meaning of these parameters is explained in section 3.4.

	T_{max}	T_{min}	N_{virt}
fast	32.8 ms	4 μs	39
medium	157.3 ms	4.8 μs	27
slow	524.3 ms	16 μs	37

Table D.1: Parameters of the implemented exemplary timings.

Each of the possible ranges requires a set of parameters for the calculation of the row sequence by the FPGA. These parameters are given in table D.2. Their meaning is explained in section 3.5.

	fast			medium			slow		
index i	cflag	num	stamp	cflag	num	stamp	cflag	num	stamp
1	-	1	0	-	1	0	-	1	0
2	0	34	-	0	228	-	0	196	-
3	1	2	8	1	2	8	1	14	4
4	0	35	-	0	143	-	0	197	-
5	1	3	7	1	26	3	1	27	3
6	0	19	-	0	85	-	0	209	-
7	1	8	5	1	3	7	1	2	8
8	0	7	-	0	37	-	0	185	-
9	1	106	1	1	100	1	1	53	2
10	0	15	-	0	115	-	0	419	-
11	1	2	9	1	2	9	1	3	7
12	0	27	-	0	71	-	0	371	-
13	1	5	6	1	14	4	1	105	1
14	0	3	-	0	61	-	0	97	-
15	1	54	2	1	51	2	1	8	5
16	0	7	-	0	57	-	0	203	-
17	1	2	10	1	2	10	1	2	9
18	0	15	-	0	35	-	0	537	-
19	1	28	3	1	8	5	1	5	6
20	0	3	-	0	59	-	0	101	-
21	1	2	11	1	2	11	1	2	10
22	0	7	-	0	17	-	0	319	-
23	1	15	4	1	5	6	1	2	11
24	0	1	-	0	29	-	0	159	-
25	1	2	12	1	2	12	1	2	12
26	0	4	-	0	23	-	0	79	-
27	1	2	13	1	2	13	1	2	13
28	0	4	-	0	11	-	0	39	-
29	-	-	-	0	2	14	0	2	14
30	-	-	-	0	5	-	0	19	-
31	-	-	-	1	2	15	1	2	15
32	-	-	-	0	5	-	0	19	-

Table D.2: Parameters for the calculation of the row sequence for three exemplary maximum integration times.

Appendix E

List of Acronyms

AC Alternating Current

ADC Analog-to-Digital Converter

APS Active Pixel Sensor

ASCII American Standard Code for Information Interchange

ASIC Application Specific Integrated Circuit

CCD Charge Coupled Device

CDS Correlated Double Sampling

CMC Common Mezzazine Card

CMOS Complementary Metal Oxide Semiconductor

CPLD Complex Programmable Logic Device

CR Cell Ratio

CRT Cathode Ray Tube

DAC Digital-to-Analog Converter

DC Direct Current

DR Dynamic Range

DS Double-Sampling

ESD Electro-Static Discharge

FET Field Effect Transistor

FPGA Field Programmable Gate Array

FPN Fixed Pattern Noise

GB Unity-Gain Bandwidth

GUI Graphical User Interface

HDR High Dynamic Range

IBM International Business Machines Corporation

ITC Integration Time Control

KIP Kichhoff-Institute for Physics, University of Heidelberg

LCD Liquid Crystal Display

LSB Least Significant Bit

MIM Metal-Isolator-Metal (Capacitor)

MOS Metal Oxide Semiconductor

MOSFET Metal Oxide Semiconductor Field Effect Transistor

NMOS N-channel MOS

PCB Printed Circuit Board

PCI Peripheral Component Interconnect

PGA Pin Grid Array

PLL Phase Locked Loop

PMOS P-channel MOS

PNG Portable Network Graphic

PSRR Power Supply Rejection Ratio

RAM Random Access Memory

RTL Register Transfer Logic

SNR Signal-to-Noise Ratio

SRAM Static Random Access Memory

TFA Thin Film on ASIC

VHDL VHDL=VHSIC HDL = Very High Speed Integrated Circuit Hardware Description Language

VIENNA Vision Enhancement for a Neural Network Architecture

VLSI Very large Scale Integration

Bibliography

[ADA83] A. Adams, *"The Print"*, The Ansel Adams Photograhy series. Little, Brown and Company, 1983

[ALL02-1] P. E. Allen, D. R. Holberg, *"CMOS Analog Circuit Design"*, Oxford University Press, 2002, p. 29

[ALL02-2] P. E. Allen, D. R. Holberg, *"CMOS Analog Circuit Design"*, Oxford University Press, 2002, pp. 439-488

[ALL02-3] P. E. Allen, D. R. Holberg, *"CMOS Analog Circuit Design"*, Oxford University Press, 2002, pp. 269-293

[ALL02-4] P. E. Allen, D. R. Holberg, *"CMOS Analog Circuit Design"*, Oxford University Press, 2002, pp. 253-256

[ALL02-5] P. E. Allen, D. R. Holberg, *"CMOS Analog Circuit Design"*, Oxford University Press, 2002, pp. 657-662

[ANA01] Analog Devices Inc., *"AD8062, 300 MHz Rail-to-Rail Amplifiers"*, datasheet, 2001

[BEC01] J. P. Becker, *"Ein FPGA-basiertes Testsystem für gemischt analog/digitale ASICs"*, diploma thesis, Kirchhoff-Institute for Physics, Heidelberg University, 2001

[BER93-1] Bergmann, Schäfer, *"Optik"*, de Gruyter, 1993, pp. 243-244

[BLA01] N. Blanc, *"CCD versus CMOS - has CCD imaging come to an end?"*, Photogrammetric Week '01, D. Fritsch, R. Spiller, eds., Wichmann Verlag, Heidelberg, 2001

[BRE04] A. Breidenassel, K. Meier, J. Schemmel, "A Flexible Scheme for Adaptive Integration Time Control", Proc. of IEEE Sensors 2004, ISBN 0-7803-8693-0, Vienna, pp. 280-283

[BUR71] J. G. Graeme, G. E. Tobey, ed.,*"Operational Amplifiers - Design and Applications"*, McGraw-Hill, 1971

[BUR96] Burr-Brown Corp., *"ADS801, 12-Bit, 25 MHz Analog-to-Digital Converter"*, datasheet, USA, 1996

[CHA01] A. Chandrakasan, W. J. Bowhill, F. Fox, *"Design of High-Performance Microprocessor Circuits"*, IEEE Press, New York, 2001, pp. 290-296

[CUL01] E. Culurciello, R. Ethienne-Cummings, K. A. Boahen, *"Arbitrated address event representation digital image sensor"*, IEEE Int. Solid State Circuits Conf., Dig. Tech. papers, USA, 2001, pp. 92-93

[DEB97] P. E. Debevec, J. Malik, *"Recovering High Dynamic Range Radiance Maps from Photographs"*, Proceedings of SIGGRAPH 97, ACM press / Addison-Wesley Publishing Co., 1997, pp. 369-378

[DEC98] S. Decker, R. D. McGrath, K. Brehmer, C. G. Sodini, *"A* 256×256 *CMOS Imaging Array with Wide Dynamic Range Pixels and Column-Parallel Digital Output"*, IEEE J. Solid State Circuits, vol. 33, no. 12, Dec. 1998, pp. 2081-2091

[DEV02] K. Devlin, *"A review of tone reproduction techniques"*, Technical Report CSTR-02-005, Department of Computer Science, University of Bristol, Nov. 2002

[DRA03] F. Drago, K. Myszkowski, T. Annen, N. Chiba, *"Adaptive Logarithmic Mapping for Displaying High Contrast Scenes"*, EUROGRAPHICS 2003, vol. 22, no. 3, 2003

[DRO99] D. Droste, *"Realisierung eines Wellenfrontsensors mit einem ASIC"*, Ph. D. thesis, Institute of Computer Engineering, University of Mannheim, 1999

[DUR02] F. Durand, J. Dorsey *"Fast Bilateral Filtering for the Display of High-Dynamic-Range Images"*, Proceedings of SIGGRAPH 2002, ACM press, 2002, pp. 257-265

[EIC89] C. Eichenberger, W. Guggenbühl, *"Dummy Transistor Compensation of Analog MOS Switches"*, IEEE J. Solid State Circuits, vol. 24, Aug. 1989, pp. 1143-1146

[FAT02] R. Fattal, D. Lischinski, M. Werman, *"Gradient Domain High Dynamic Range Compression"*, Proceedings of SIGGRAPH 2002, ACM press, 2002, pp. 249-256

[FER96] J. A. Ferwerda, S. Pattanaik, P. S. Shirley, D. P. Greenberg, *"A model of visual adaptation for realistic image synthesis"*, Proceedings of SIGGRAPH 96, ACM press / Addison-Wesley Publishing Co., 1996, pp. 249-258

[FOS93] E. R. Fossum, *"Active Pixel Sensors: Are CCD's Dinosaurs?"*, Proc. SPIE, vol. 1900, 1993, pp. 1-14

[FOS97] E. R. Fossum, *"CMOS image sensors: Electronic Camera on a Chip"*, IEEE Trans. Electron Devices, vol. 44, Oct. 1997, pp. 1689-1698

[FRE00] J. Frey, *"Physikalische Modelle in der Halbleiterelektronik 1999/2000: Rauschen in Halbleitern"*, Universität-Gesamthochschule Siegen, Institut für Halbleiterelektronik, 2000

[GEI90-1] R. L. Geiger, P. E. Allen, N. R. Strader, *"VLSI design techniques for analog and digital circuits"*, McGraw-Hill, Inc., 1990, pp. 174-177

[GEI90-2] R. L. Geiger, P. E. Allen, N. R. Strader, *"VLSI design techniques for analog and digital circuits"*, McGraw-Hill, Inc., 1990, pp. 597-599

[HAM01] T. Hamamoto, K. Aizawa, *"A Computational Image Sensor with Adaptive Pixel-Based Integration Time"*, IEEE J. Solid State Circuits, vol. 36, no. 4, Apr. 2001, pp. 580-585

[HAS01] A. Hastings, *"The Art of Analog Layout"*, Prentice-Hall, Inc., 2001, pp. 426-444

[HAU99] R. Hauschild, *"Integrierte CMOS-Kamerasysteme für die zweidimensionale Bildsensorik"*, Ph. D. thesis, University of Duisburg, 1999

[HPA94] Hewlett Packard, *"HP4155A Semiconductor Parameter Analyser"*, User's Dictionary Reference, Japan, 1994

[IBA95] H. Ibach, H. Lüth, *"Festkörperphysik"*, 4th ed., Springer-Verlag, Heidelberg, 1995, pp. 334-338

[IBM01] IBM Corp., *"CMOS 6SF & CMS 6SFS Design Manual"*, EC number H05619, 2001

[IED78] N. Ieda, Y. Ohmori, K. Takeya, T. Yano, *"Single Transistor MOS RAM Using a Short-Channel MOS Transistor"*, IEEE J. Solid State Circuits, vol. 13, Apr. 1978, pp. 218-224

[ILL01] L. G. McIlrath, *"A Low-Power Low-Noise Ultrawide-Dynamic-Range CMOS Imager with Pixel-Parallel A/D Conversion"*, IEEE J. Solid State Circuits, vol. 36, no. 5, May 2001, pp. 846-853

[IMS00] IMS Stuttgart, *"HDRC VGA Imager and Camera, Data and Features"*, Publicity material, Institut für Mikroelektronik Stuttgart, Germany, 2000

[ISH94] K. Ishibashi et al., *"A 12.5-ns 16-Mb CMOS SRAM with Common-Centroid-Geometry-Layout Sense Amplifier"*, IEEE J. Solid State Circuits, vol. 29, no. 4, Apr. 1994, pp. 411-416

[JYR03] T. Jyrinki, *"High Dynamic Range Images"*, Seminar on computer graphics, T-111.500, Helsinki University of Technology, 2003

[KAV00] S. Kavadias et al., *"A Logarithmic Response CMOS Image Sensor with On-Chip Calibration"*, IEEE J. Solid State Circuits, vol. 35, no. 8, Aug. 2000

[KNO92] K. Knopp, *"Image Sensors"*, in *"Sensors - A Comprehensive Survey"*, W. Göpel, J. Hesse, J. N. Zemel, ed., vol. 6, 1992, p. 244

[LAK86] K. R. Lakshmikumar, R. A. Hadaway, M. A. Copeland, *"Characterization and modeling of mismatch in MOS transistors for precision analog design"*, IEEE J. Solid State Circuits, vol. SC-21, 1986, pp. 1057-1066

[LIN04] LINOS Photonics GmbH & Co. KG, *"Wärmeschutzfilter Calflex C"*, datasheet, 2004

[LOO99-1] M. Loose, *"A Self-Calibrating CMOS Image Sensor with Logarithmic Response"*, Ph. D. thesis, Institute of High Energy Physics, Heidelberg University, 1999, pp. 184-186

[LOO99-2] M. Loose, *"A Self-Calibrating CMOS Image Sensor with Logarithmic Response"*, Ph. D. thesis, Institute of High Energy Physics, Heidelberg University, 1999, pp. 20-30

[LOO01] M. Loose, K. Meier, J. Schemmel, *"A Self-Calibrating Single-Chip CMOS Camera with Logarithmic Response"*, IEEE J. Solid State Circuits, vol. 36, no. 4, Apr. 2001

[LOV98] S. J. Lovett, M. Welten, A. Mathewson, B. Mason, *"Optimizing MOS Transistor Mismatch"*, IEEE J. Solid State Circuits, vol. 33, Jan. 1998, pp. 147-150

[LUL00] T. Lulé, M. Wagner, M. Verhoeven, H. Keller, M. Böhm, *"100000-Pixel, 120-dB Imager in TFA Technology"*, IEEE J. Solid State Circuits, vol. 35, no. 5, May 2000, pp. 732-739

[MAR98] G. F. Marshall, S. Collins, *"A high dynamic range front end for automatic image processing applications"*, Advanced Focal Plane Arrays and Electonic Cameras, Thierry M. Bernard, ed., Proc. SPIE 3410, 1998, pp. 176-185

[MAX99] Maxim Integrated Products, *"MAX5104, Low-Power, Dual, Voltage-Output, 12-Bit DAC with Serial Interface"*, datasheet, 1999

[MIY95] R. Miyagawa, T. Kanade, *"Integration-Time Based Computational Image Sensors"*, Carnegie Mellon University, Robotics Institute, ITE technical report, vol. 19, no. 65, Nov. 1995, pp. 37-41

[MOI00-1] A. Moini, *"Vision Chips"*, Kluwer Academic Publishers, USA, 2000, pp. 261-264

[MOI00-2] A. Moini, *"Vision Chips"*, Kluwer Academic Publishers, USA, 2000, pp. 65-71

[MOI00-2] A. Moini, *"Vision Chips"*, Kluwer Academic Publishers, USA, 2000, p. 75

[NOB68] P. Noble, *"Self-Scanned Image Detector Arrays"*, IEEE Trans. Electron Devices, vol. ED-15, 1968, p. 202

[OHT01] J. Ohta, H. Sakata, T. Tokuda, M. Nunoshita, *"Low-voltage operation of a CMOS image sensor based on pulse frequency modulation"*, Proc. of SPIE, vol. 4306, San Jose, CA, 2001, pp. 319-326

[ORI95] Oriel Corp., *"Sicherer Betrieb von Bogenlampen"*, datasheet, 1995

[ORI94] Oriel Corp., *"Light Sources, Monochromators & Spectrographs, Detectors & Detection Systems, Fiber Optics"*, Oriel catalog, 1994, pp. 1-40

[PCI95] *" PCI Local Bus Specification"*, revision 2.1, PCI Special Interest Group, P.O.Box 14070, Portland, OR97214, USA, 1995

[PEL89] M. J. M. Pelgrom, C. J. Duinmaijer, A. P. G. Welbers, *"Matching Properties of MOS Transistors"*, IEEE J. Solid State Circuits, vol. 24, no. 5, 1989, pp. 1433-1440

[RAZ92] B. Razavi, B. A. Wooley, *"Design Techniques for High-Speed, High-Resolution Comparators"*, IEEE J. Solid State Circuits, vol. 27, Dec. 1992, pp. 1916-1926

[REI02] E. Reinhard et al., *"Photographic Tone Reproduction for Digital Images"*, Proceedings of SIGGRAPH 2002, ACM press, 2002, pp. 267-276

[RIC95] N. Ricquier, B. Dierickx, *"Active Pixel CMOS Image Sensor with On-Chip Non-Uniformity Correction*, IEEE workshop on CCDs and advanced image sensors, California, April 20-22, 1995

[SAR91] R. Sarpeshkar, J. L. Wyatt, N. C. Lu, P. D. Gerber, *"Mismatch Sensitivity of a Simultaneously Latched CMOS Amplifier"*, IEEE J. Solid State Circuits, vol. 26, Oct. 1991, pp. 1413-1422

[SAS04] M. Sasaki, M. Mase, S. Kawahito, Y. Tadokoro, "A Wide Dynamic Range CMOS Image Sensor with Multiple Short-Time Exposures", Proc. of IEEE Sensors 2004, ISBN 0-7803-8693-0, Vienna

[SAY90] M. Sayag, *"Non-linear Photosite Response in CCD Imagers"*, U.S. Patent no. 5055667, 1991 (Filed 1990)

[SEI99] P. Seitz, *"Solid-State Image Sensing"*, published in *"Handbook of Computer Vision and Applications"*, Academic Press, vol. 1, Boston, 1999, pp. 165-222

[SCH99] B. Schneider, P. Rieve, M. Böhm, *"Image Sensors in TFA Thin Film on ASIC Technology"*, published in *"Handbook of Computer Vision and Applications"*, Academic Press, vol. 1, Boston, 1999, pp. 237-270

[SCH00] M. Schanz, C. Nitta, A. Bußmann, B. J. Hosticka, R. K. Wertheimer, *"A High-Dynamic-Range CMOS Image Sensor for Automotive Applications"*, IEEE J. Solid State Circuits, vol. 35, no. 7, Jul. 2000, pp. 932-938

[SCH02] O. Schrey, J. Huppertz, G. Filimonovic, A. Bußmann, W. Brockherde, B. J. Hosticka, *"A 1K×1K High Dynamic Range CMOS Image Sensor With On-Chip Programmable Region-of-Interest Readout"*, IEEE J. Solid State Circuits, vol. 37, no. 7, Jul. 2002, pp. 911-915

[SCH04] J. Schemmel, S. Hohmann, K. Meier, F. Schürmann, *"A Mixed-Mode Analog Neural Network Using Current-Steering Synapses"*, Analog Integrated Circuits and Signal Processing, vol. 38, 2004, pp. 233-244

[SHI87] J.-H. Shie, M. Patil, B. J. Sheu, *"Measurement and Analysis of Charge Injection in MOS Analog Switches"*, IEEE J. Solid State Circuits, vol. sc-22, no. 2, Apr. 1987, pp. 277-281

[SON04] Sony Corp., *"ICX488DQ/DQF - Color CCD for Consumer Digital Still Cameras"*, datasheet, 2004

[SZE81-1] S. M. Sze, *"Physics of Semiconductor Devices"*, John Wiley & Sons, 1981, pp. 74-81

[SZE81-2] S. M. Sze, *"Physics of Semiconductor Devices"*, John Wiley & Sons, 1981, pp. 749-754

[SZE81-3] S. M. Sze, *"Physics of Semiconductor Devices"*, John Wiley & Sons, 1981, pp. 783-789

[TEO97] M. Teodori, *"Design und Messung optoelektronischer Strukturen in CMOS Technologie"*, diploma thesis, Institute of High Energy Physics, Heidelberg University, 1997

[VHD97] *"VHDL Language Reference Manual"*, IEEE Std. 1076.1, Design Automation Standards Committee of the IEEE Computer Society, Institute of Electrical and Electronics Engineers, Inc, New York, USA, 1997

[WEC67] G. P. Weckler, *"Operation of* p-n *Junction Photodetectors in a Photon Flux Integrating Mode"*, IEEE J. Solid State Circuits, vol. sc-2, no. 3, Sept. 1967, pp. 65-73

[WEG87] G. Wegmann, E. A. Vittoz, F. Rahali, *"Charge Injection in Analog MOS Switches"*, IEEE J. Solid State Circuits, vol. sc-22, no. 6, Dec. 1987, pp. 1091-1097

[WON96] H.-S. Wong, *"Technology and Device Scaling Considerations for CMOS Imagers"*, IEEE Trans. on Electron Devices, vol. 43, no. 12, Dec. 1996, pp. 2131-2142

[XIL02] Xilinx Inc., *"Virtex-E 1.8 V Field Programmable Gate Arrays"*, Production Product Specification, 2002

[YAD97] O. Yadid-Pecht, E. R. Fossum, *"Wide Intrascene Dynamic Range CMOS APS Using Dual Sampling"*, IEEE Trans. Electron Devices, vol. 44, no. 10, Oct. 1997, pp. 1721-1723

[YAD99] O. Yadid-Pecht, *"Wide-dynamic-range sensors"*, Optical Engineering, vol. 38, no. 10, Oct. 1999, pp. 1650-1660

[YAD03] O. Yadid-Pecht, A. Belenky, *"In-Pixel Autoexposure CMOS APS"*, IEEE J. Solid State Circuits, vol. 38, no. 8, Aug. 2003, pp. 1425-1428

[YAD04-1] O. Yadid-Pecht, R. Etienne-Cummings, eds., *"CMOS Imagers: From Phototransduction to Image Processing"*, Kluwer Academic Publishers, Boston, 2004, p. 99-115

[YAD04-2] O. Yadid-Pecht, R. Etienne-Cummings, eds., *"CMOS Imagers: From Phototransduction to Image Processing"*, Kluwer Academic Publishers, Boston, 2004, pp. 26-31

[YAD04-2] O. Yadid-Pecht, R. Etienne-Cummings, eds., *"CMOS Imagers: From Phototransduction to Image Processing"*, Kluwer Academic Publishers, Boston, 2004, pp. 161-163

[YAN99-1] D. X. D. Yang, A. El Gamal, *"Comparative Analysis of SNR for Image Sensors with Enhanced Dynamic Range"*, Proc. SPIE, vol. 3649, San Jose, CA, 1999, pp. 197-211

[YAN99-2] D. X. D. Yang, A. El Gamal, B. Fowler, H. Tian, *"A 640×512 CMOS Image Sensor with Ultrawide Dynamic Range Floating-Point Pixel-Level ADC"*, IEEE J. Solid State Circuits, vol. 34, no. 12, Dec. 1999, pp. 1821-1834

[YAN94] W. Yang, *"A wide-dynamic range, low-power photosensor array"*, IEEE Int. Solid State Circuits Conf., San Francisco, CA, Feb. 1994, pp. 230-231

Danksagung · Acknowledgements

An dieser Stelle möchte ich mich bei allen bedanken, die mich während meiner Zeit am Kirchhoff-Institut mit Rat und Tat unterstützt haben und so zum Gelingen dieser Arbeit beigetragen haben. Insbesondere möchte ich mich bedanken bei:

Herrn Prof. Dr. Karlheinz Meier für die Möglichkeit diese Arbeit durchzuführen,

Herrn Prof. Dr. Volker Lindenstruth für die Erstellung des Zweitgutachtens,

Herrn R. Achenbach für die Hilfe bei der Organisation von Messgeräten und Bauteilen für diverse Versuchsaufbauten,

Herrn M. Dorn sowie Herrn M. Keller für die Unterstützung bei der Arbeit mit der Entwurfs- und Simulationssoftware,

Herrn Dr. T. Maucher für ein angenehmes Arbeitsklima und ständige Hilfsbereitschaft bei der Lösung diverser Probleme,

Herrn Dr. J. Schemmel und Herrn J. Langeheine für die hilfreiche und stets kompetente Unterstützung bei der Lösung elektro-technischer Probleme, insbesondere jedoch auch für anregende Diskussionen zu den unterschiedlichsten Themen,

allen Mitgliedern der Arbeitsgruppe *Electronic Vision(s)* für das gemeinsame Miteinander, welches den Alltag mit Leben gefüllt hat,

allen Mitarbeitern des Kirchhoff-Instituts, die sowohl für die notwendige Infrastruktur als auch für ein motivierendes Arbeitsklima gesorgt haben.

Herrn Dr. O. Vossen für die Durchsicht des Manuskripts dieser Arbeit und viele hilfreiche Tipps zur englischen Sprache.

meiner lieben Frau Nicole, die mich nicht nur in meiner Arbeit unterstützt hat sondern auch sonst immer für mich da war,

meinen Eltern dafür, dass sie mir das Studium der Physik ermöglicht haben und mich stets in meinen Entscheidungen unterstützt haben.