MODULAR NEUROMORPHIC VLSI ARCHITECTURES FOR VISUAL MOTION AND TARGET TRACKING

by

Vivek Pant

A Thesis Submitted to the Faculty of the

ELECTRICAL AND COMPUTER ENGINEERING DEPARTMENT

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2003

Acknowledgments

I am greatly thankful to my parents and sisters for their love and continued support throughout my life.

I am thankful to my advisor Chuck Higgins for his guidance, encouragement, and inspiration throughout this work.

I am thankful to Prof. Rozenblit and Prof. Palusinski for serving on my thesis defense committee.

I thank Ania Mitros (Caltech) for her help during the design and implementation of the spiking neuron chip.

I would also like to thank the Doings group of neurobiology for the involving discussions that were highly interesting and informative.

I am thankful to the Higgins Lab for the late-night parleys, and a pleasant and friendly atmosphere courtesy Erhan, Robert, Jad, and Anusha.

And my special thanks to Mansi for being there and keeping me going.

TABLE OF CONTENTS

LIST OF FIGURES

Abstract

Modern age intelligent systems will require extensive computational power, complex parallel processing units, and low-power design. Biologically inspired neuromorphic VLSI systems present a viable solution to the demands of both highly parallel, and low-power consuming processors. Among biological sensory systems, vision is the most important one with the largest portion of the brain devoted to visual computations. Biological models for tasks like visual motion computation, target pursuit, and crash-avoidance have been extensively studied. In this thesis, a biologically inspired target-fixation model has been analyzed and implemented in VLSI. A modular approach for designing a sender-receiver based tracking system has also been discussed. A spiking-neuron sender chip has been implemented using a frequency-encoded event driven communication protocol. This sender chip is used to relay information about changes in image intensity to a computational unit in a modular visual system. The use of these neuromorphic chips has been suggested for developing monolithic and modular target-tracking systems for yaw-torque control in a robot. The mixed-signal chips work in the subthreshold region of the MOSFET and consume very little electrical power. Subthreshold implementations are very well suited for the low-frequency behavior of real-world tracking systems. System level architecture and simulations of such a tracking system have also been presented.

Chapter 1

Introduction

Design of an intelligent machine that can perform human-like tasks is a great engineering challenge. Even the simplest task performed by man being extended to machines remains a distant goal. This is partly due to our incomplete understanding of these mechanisms and partly due to the physical restrictions on the density of wires and the cost of communications imposed by the spatial layout of electronic circuits. At the same time, these tasks can be performed by the smallest of biological organisms with relative ease indicating their feasibility.

The complexity and parallelism of biomimetic algorithms and their real-time applications rule out the use of slow serial processors and software implementations. High-speed digital signal processors are an alternative for doing these computations in real-time. These architectures, though fast, have to be connected to analog sensors. Continuous-time signals from these analog sensors have to be sampled at a very high frequency to reduce aliasing. Moreover, this communication of data at high speeds over finite length buses from the sensor to the processor has become a bottleneck for the overall system performance (Cong *et al.*, 1996). Besides, digital computers are extremely effective when precise numerical solutions are sought. The nervous system and neuromorphic algorithms accept fuzzy, noisy input to produce an approximate result. The design of such systems is therefore fundamentally different from designing digital processors. The analog nature of the neural system can therefore be best implemented by using high-speed, high-density, parallel analog VLSI designs.

1.1 Neuromorphic Engineering

In the field of Neuromorphic Engineering, inspiration is taken from neurobiological circuits and models, which are implemented through analog building blocks and novel circuits. Pioneering work on the implementation of neural circuits on silicon hardware started in the late 1980s (Mead, 1989). Rapid advances in this field have been made since then. Visual, auditory, and other bio-sensory models have been successfully implemented (Koch *et al.*, 1991; Horiuchi *et al.*, 1996; Indiveri *et al.*, 1996b; Lazzaro and Wawrzynek, 1997; Higgins and Shams, 2002). Of these, vision is particularly prominent and has been widely studied. It is no surprise, because out of the 10^{11} neurons (Koch, 1999) that the human brain has, more are devoted to vision than to any other sensory function (Zigmond *et al.*, 1999). Vision is a complex task and neuromorphic systems have been developed to implement various visuo-motor mechanisms. These include self-motion estimation, navigation, target pursuit, and collision avoidance. The visual system of insects is easier to study because of the abundance of the subjects, particularly the fly, and the relatively lower complexity of the insect brain (see Figure 1.1). The fly's eyes are the so-called compound eyes, which consist of about 700 ommatidia (in *Drosophila*) that cover approximately 86% (310°) of the visual field (Huber and Bülthoff, 1998). The light captured by each ommatidium is projected onto 8 photoreceptors (see Figure 1.2). The visual signals from the eyes are processed in three neural layers illustrated in Figure 1.1: (i) the lamina, (ii) the medulla, and (iii) the lobula complex (comprised of the lobula and the lobula plate).

Motion information plays a vital role whenever an object moves in the visual field, or an insect moves in a stationary environment. Let us consider three very common situations:

1. When an insect deviates from a straight course in flight, the visual environment is coherently displaced.

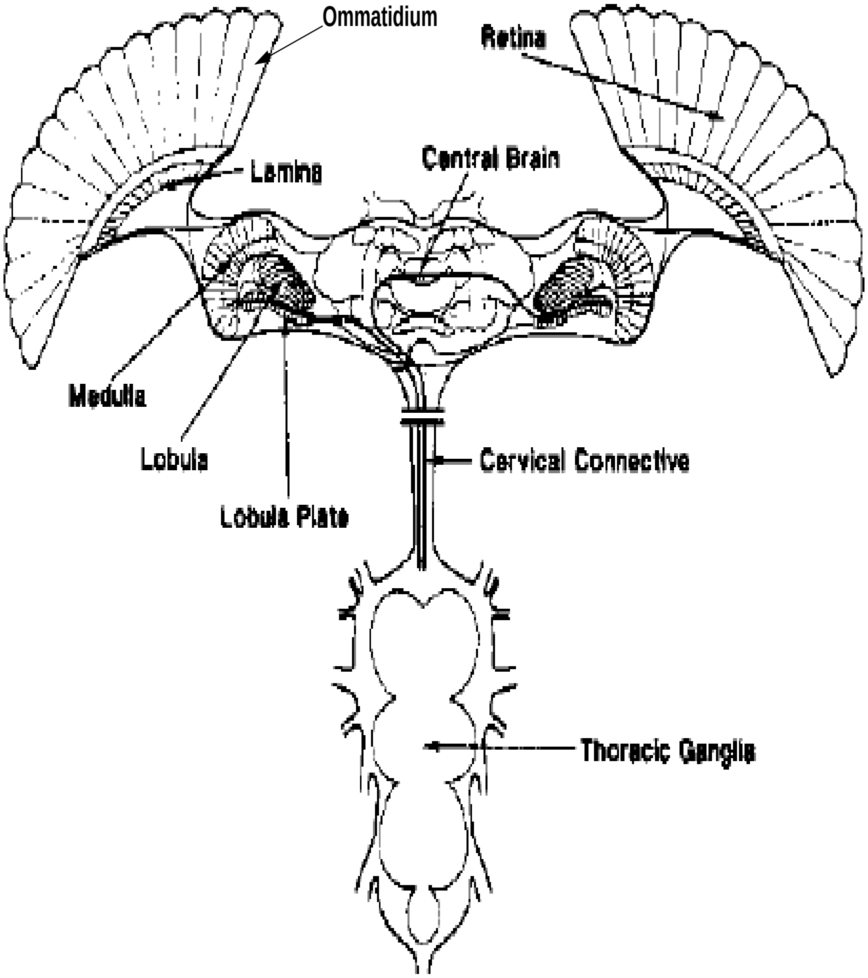


Figure 1.1. Schematic horizontal cross-section of the fly's brain. The retina, three visual ganglia: the lamina, medulla and lobula complex (which is subdivided into the anterior lobula and the posterior lobula plate), the central brain and the thoracic ganglia with the motor control centers constitute the nervous system. The fly's nervous system has approximately 10^5 neurons. Modified from Egelhaaf and Borst (1993a).

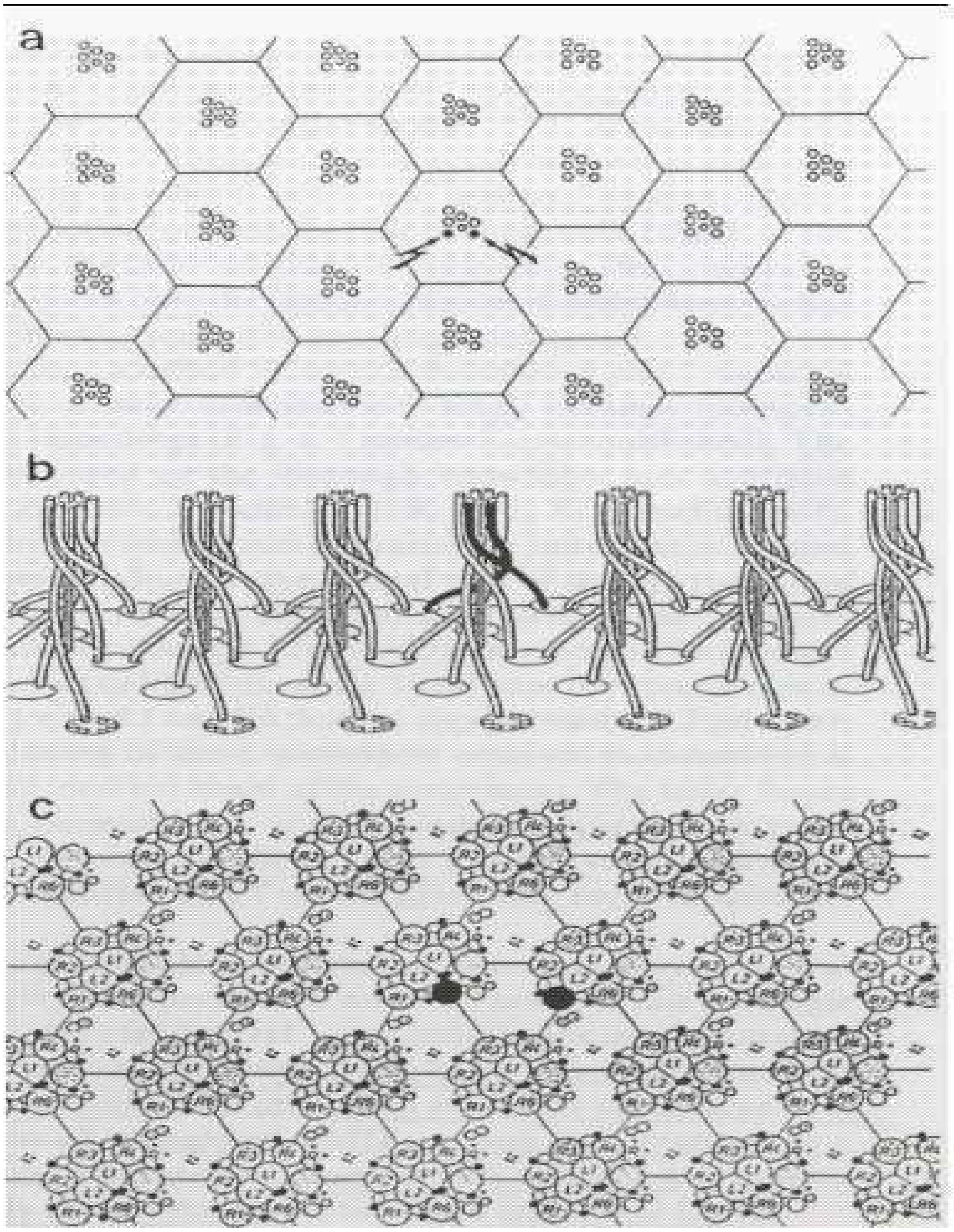


Figure 1.2. Axonal projection of photoreceptor cells from the retina onto the lamina in the eye of the fly. (a) Optical stimulation of two adjacent receptor cells R1 and R6. (b) The axonal projection on the lamina is such that R1 and R6 are immediately adjacent (twisted black) photoreceptor cells. (c) The clustered R1-R6 cells, suggested to increase the signal to noise ratio of the input as received by a movement detector. Reproduced without permission from Franceschini *et al.* (1989).

2. When it approaches an obstacle, there is an expansion of the retinal image.
3. When the insect passes a nearby object in front of a distant background, there is a discontinuity in the visual motion field.

These retinal motion patterns create a different neuronal response for each scenario. This information is used to guide the visual orientation of the insect such that:

1. Rotatory large field motion signals unintended deviations from the course. This response is known as the optomotor response. An auto-steering system may be formed by extracting this information.
2. Expansion of the image signals an approaching obstacle. It may control behavioral responses like deceleration during landing and crash-avoidance.
3. Discontinuities in the visual flow field indicate moving objects in front of a distant background. This information can be extracted by a fixation system and may be used for target pursuit.

All these behaviors have been studied extensively in flies and neuronal circuits for the above have been derived by physiological studies and electrical recordings (Egelhaaf and Borst, 1993b). The robustness of a biological system is inherited from a profusely parallel architecture and distributed information processing units. Low-power analog VLSI circuits operating in the sub-threshold region of field-effect transistors can be densely packed in a small area to obtain a viable processing element. This can be replicated to design a parallel processor performing computations similar to living neurons. Advances in this field provide a new paradigm for solving real-world problems in real-time. The presented thesis is a contribution to this effort.

1.2 Related Work

Smart vision systems performing tasks like motion computation, auto-navigation, and target tracking are an inevitable part of an autonomous mobile robot. Conventional vision systems based on the system level integration (or even chip level integration) of an imager (usually a CCD) camera and a digital processor have limitations in terms of cost, size and complexity of these systems. Due to these factors, conventional vision systems have limited use in the design of small intelligent agents. Vision chips including photosensors and parallel processing elements (analog or digital) have been under research for more than a decade and illustrate promising capabilities.

Highly parallel implementation of biological algorithms is achieved by doing focal-plane computations in a single chip. This monolithic architecture generally consists of an array of computational units processing information transduced into electrical signals by a sensor on the same silicon chip. Mahowald's (1994) silicon retina chip was amongst the first vision chips which implemented a biological facet of vision on silicon. The computation performed by Mahowald's silicon retina was based on models of computation in distal layers of the vertebrate retina, which include the cones, the horizontal cells, and the bipolar cells. Stereo matching chips which use static and dynamic features of images and compute disparity in nine image planes were implemented by Mahowald and Delbrück (1989). Boahen and Andreou (1992) implemented a silicon retina in a compact circuit with a diffusive smoothing circuit. DeWeerth (1992) implemented a centroid detection chip based on an aggregation network. A two-dimensional (2-D) motion detection chip designed for computing spatio-temporal gradients of the input image was among the first few implementations of a motion detecting chip (Moore and Koch, 1991). Harris *et al.* (1990) developed an analog hardware for detecting discontinuities in early vision. The correlation-based motion detector of Delbrück (1993b), unlike some other motion detection systems which depended on spatial and temporal differentiation, used correlation to extract motion information and hence claimed to be more robust than other motion detectors. "Facilitate and sample" algorithms were implemented to calculate the time to travel by Kramer *et al.* (1997). Higgins and Korrapati (2000) implemented the Adelson-Bergen algorithm to compute motion energy. The time-to-crash sensor of Indiveri *et al.* (1996a) was a demonstration of how a neuromorphic vision chip could be used for real life applications.

Focal-plane level processing is very well suited for medium sized algorithms. As the complexity of the algorithm increases, so does the size of the processing unit, until the processing unit becomes so large that the percentage of space devoted to the phototransducer area and therefore the resolution of the system is reduced significantly. Overloading of computations in the focal plane and less imaging area in a monolithic system are the reasons that motivate the splitting of computation into smaller modules. Modular architectures that divide the computations into two or three processing stages have evolved over the years (for review, see: Higgins and Koch, 1999). Various schemes have been developed to relay information from one chip to the other in a modular architecture. Significant work related to developing a frequency encoded address-event representation scheme for communication was done by Boahen (1999). Landolt, Mitros, and Koch (2001) have designed a communication scheme that discards colliding events and allows only error-free events to be transmitted off-chip. Higgins and Koch (1999) implemented multi-chip motion-processing architecture that was based on a sender-receiver system. A biologically-inspired computational architecture for small-field detection and wide-field spatial integration of visual motion was developed by Higgins and Shams (2002) for computing self motion.

Rapid advances have been made in the field of motion processing architectures and their extension into behavioral applications is becoming prevalent. A detailed summary of the works in the field of visual sensors was presented by Moini (1997).

1.3 Presented Work

In this thesis, visual motion computation algorithms and their implementation are studied and a target-tracking sensor based on a neuronal model from the fly is presented.

Chapter 2 discusses a brief summary of the biological algorithms and neuronal circuits for motion detection and target pursuit believed to be present in the brain of the fly and invertebrates in general.

Chapter 3 describes the simulations and VLSI implementation of a biological model used by the fly for target-tracking. The preliminary characterization results of this monolithic target tracking sensor are also presented.

In Chapter 4, the importance of modular architectures, a description of sender-receiver chip based systems, and VLSI implementation of a spiking neuron sender chip are presented. This chip sends the contrast information in its visual field over a digital asynchronous bus. An implementation of this asynchronous communication protocol is also discussed.

In Chapter 5, a target tracking agent based on a biological model is presented. This agent receives the small-field response from a sensor and tracks a moving target in a cluttered environment. The simulation results of this system are shown and discussed.

In Chapter 6, the system limitations and improvements are discussed to conclude this thesis. The modular implementation of the entire system is described. Finally, the future extension of this work is discussed.

CHAPTER 2

BIOLOGICAL ALGORITHMS FOR MOTION DETECTION AND TRACKING

This chapter discusses various biological algorithms that have been proposed to explain the complex visual processing in insects to detect and track objects in the real world. The evolution of life was controlled by sunlight in its initial phase; as a consequence of this solar selection, eyes developed as the primary outpost of the brain (Land and Fernald, 1992). The most primitive visual systems had protein components called opsins, which captured photons from the surrounding environment. In vertebrates, the cone and rod opsins phylogeny maps onto the phylogeny of the parent species, thus showing a long and complex evolutionary lineage (Land and Fernald, 1992). The easy-to-find nature of insects and the relatively lower complexity of their brain make them the ideal candidate for study. In the next section, the anatomy and physiology of the insect eye are discussed. The next few sections discuss two important correlation and spatio-temporal motion algorithms. The neural circuitry behind target fixation and pursuit are then discussed.

2.1 The Insect Eye

Most insects have compound eyes, the most common type being apposition eyes. Each unit, or ommatidium, consists of a lens that forms an image onto the tip of the rhabdom, a light guiding structure of photo pigment (see Figure 2.1). An inverted image is formed behind each facet to delineate each rhabdom's field of view and to increase its brightness. However, the image is not resolved within the rhabdom. The insect sees the overall erect image that is formed by the apposed pixels contributed from the ommatidia (Land and Fernald, 1992). Detection of motion is one of the most important functions of the eyes. It is used for navigation, obstacle avoidance, prey detection, and course stabilization. Exner (1891) proposed that arthropods use motion parallax to estimate distances. Movement of an object in front of a motionless animal, or the self-motion of the animal, produces an orderly sequence of excitations in the photoreceptor cells. This process is same for the compound eyes of insects and the camera eyes of the vertebrates (see Figure 2.1a). The resulting electrical signals produced by the photoreceptor array are processed by specialized neural microcircuits that inform the animal about its motion relative to its ambience. Exner (1894) showed that sequential flashing of two neighboring stationary light spots creates a vivid impression of apparent motion in humans. He postulated the presence of a motion perception center that has ganglion cells responsible for directional motion detection. These directional selective neurons were later discovered in arthropod visual systems. Electrophysiological recordings from these cells have enhanced our understanding of visual motion.

In flies, the lobula complex (third optic ganglion) is divided into two parts. The posterior lobula plate houses 50 identifiable direction selective neurons participating in optomotor control related to head and body movement (Hausen and Egelhaaf, 1989). Most of these tangential cells are wide-field direction selective (DS) cells. The H1 cell found in the lobula plate of the fly is the most experimented-with specimen of DS neurons. This giant neuron has extensive dendritic arbor in the lobula plate and is separated from the photoreceptor cells by at most three or four synapses. This neuron is broad and covers the entire panoramic field of view of the ipsilateral eye. It responds with an increased firing rate to horizontal forward movements and inhibits its response to motion in the backward direction. Small-field motion detection cells, called figure detection (FD) cells, were discovered by Egelhaaf (1985a). FD cells have smaller excitatory fields (see Figure 2.2) and are

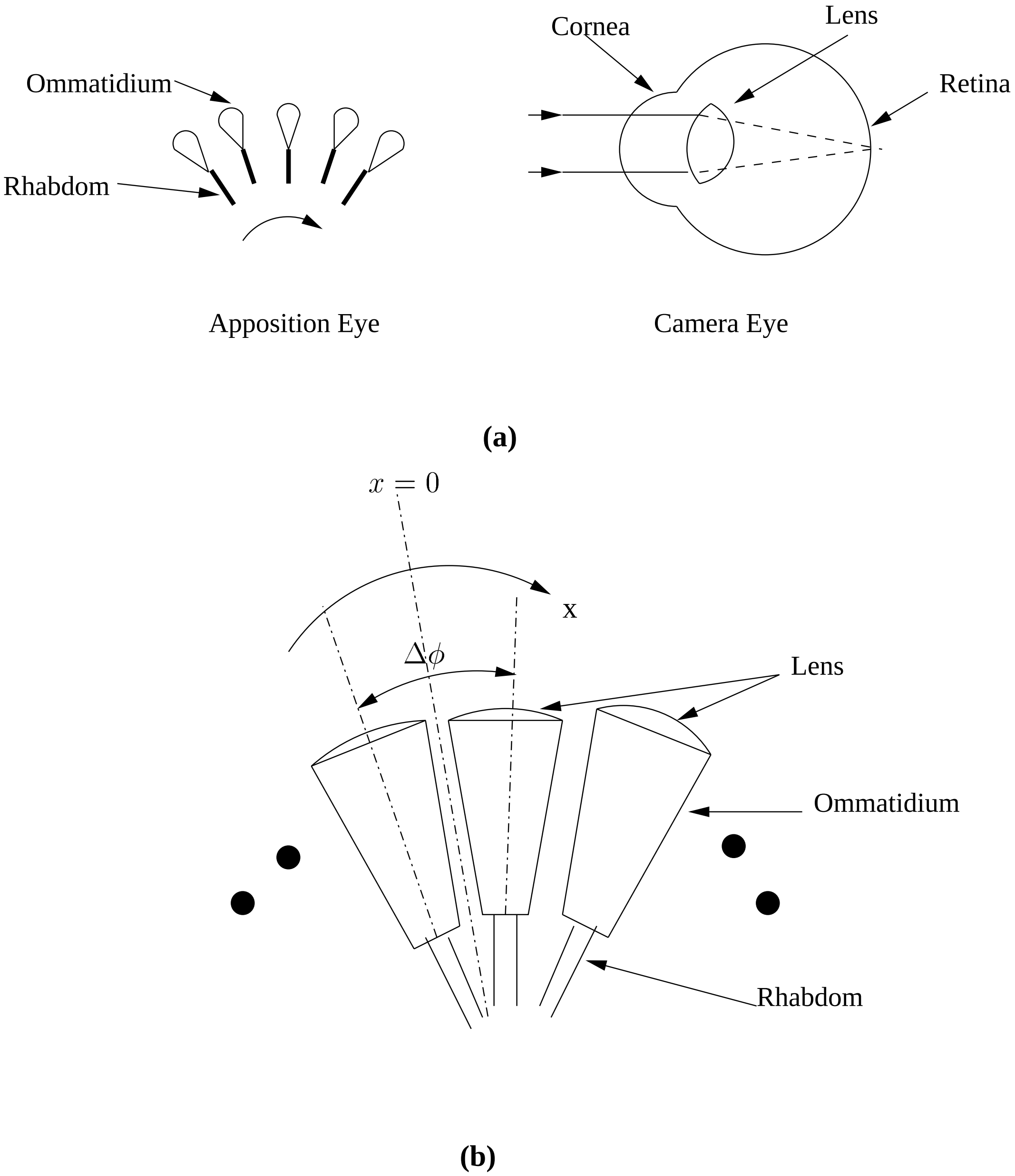


Figure 2.1. Single chambered and compound eyes. (a) Apposition compound eye showing several ommatidia next to each other. Each ommatidium consists of a lens that forms image on the tip of a light-guiding structure, the rhabdom. The camera eye, common in vertebrates including man, has a single chamber consisting of an outer cornea and an inner lens forming an image on the retina. (b) Sketch of ommatidium showing lens and rhabdom. The spatial distance between two adjacent ommatidia is given by an interommatidial angle $\Delta\phi$ along the x-axis.

believed to be a part of a fixation system which induces turning towards objects. These small-field motion detection cells, found in the lobula plate, respond to motion of small patterns in large parts of the fronto-lateral visual-field. The guidance of a fly is supposed to be through the interactive coordination of both horizontal cells and FD cells.

The H1 sensitivity to directional motion is due to pooling of the signals from small field units called the elementary motion detectors (EMDs), which project perpendicularly to the dendritic fan in a retinotopic manner. Since H1 responds readily to local motion between two points in its receptive field that are one interommatidial angle ($\Delta\phi$ in Figure 2.1b) apart (McCann, 1972), the array of EMDs feeding it must be wired very finely to it. It has been observed that small field cells are distally located to H1 (in the medulla: McCann and Dill, 1969) and they do not necessarily drive the DS lobula plate because there exists other direct output links from the medulla to the brain. R1-6 type photoreceptor cells feed their input to the motion detecting circuitry of H1 (McCann and Arnett, 1972). The various experiments by Heisenberg and Buchner (1977) have shown that the optomotor response is mediated predominantly by receptors of the R1-6 type (see Figure 1.2). R1-6 project to the lamina that is believed to be involved in motion detection circuitry driving H1 (Franceschini *et al.*, 1989). Direction selective motion suggests correlation among the different photoreceptor units. In a breakthrough work, Hassenstein and Reichardt (1956) concluded that motion detection by the nervous system requires an interaction of signals from two adjacent ommatidia. In the next section, this model of EMD as proposed by Hassenstein and Reichardt is discussed in detail.

2.2 The Hassenstein-Reichardt (HR) Model

The idea of a correlation-based motion detector was first proposed by Exner (1894). Analysis of the optomotor response of the beetle *Chlorophanus* provided electrophysiological support to the model. Hassenstein succeeded in eliciting that motion detection by the nervous system requires an interaction of signals from two directly adjacent ommatidia. In 1956, Hassenstein and Reichardt developed the correlation model (HR model) to explain their experimental observations. The input/output analysis of the experimental results inferred that the basic neural computation consists of three steps:

- Asymmetric linear filtering of signals from adjacent photoreceptor cells.
- Multiplication of the resulting signals.
- Time averaging of the product obtained.

On the basis of these observations, a correlation-based motion detector model was proposed (see Figure 2.3). According to this model, the two dimensional visual field is covered by local motion detectors, which evaluate a spatio-temporal cross-correlation between the light intensity fluctuations between the two neighboring detectors. Each motion detector consists of two photodetection units followed by two temporal filters and a multiplication sub-unit arranged in a mirror-symmetric fashion. The first of the two temporal filters is a high-pass filter that enhances the sharp temporal features of the image. It is followed by a direct and a delayed path leading to two separate multiplication units. The direct response from one photodetector is multiplied with the delayed input from the adjacent detector. The difference between two such processed signals gives a spatio-temporally tuned response that is directionally selective along the orientation of photodetector axis.

For a one dimensional (1-D) sinusoidal stimulus moving at a velocity v, the image intensity is given by:

$$I(x,t) = I + \Delta I \sin(2\pi f_s x + 2\pi f_t t) \tag{2.1}$$

where I is the mean intensity of the stimulus, f_t is the temporal frequency and f_s is the spatial frequency. The contrast of the grating is $\Delta I/I$. The signal velocity is a ratio of temporal and spatial

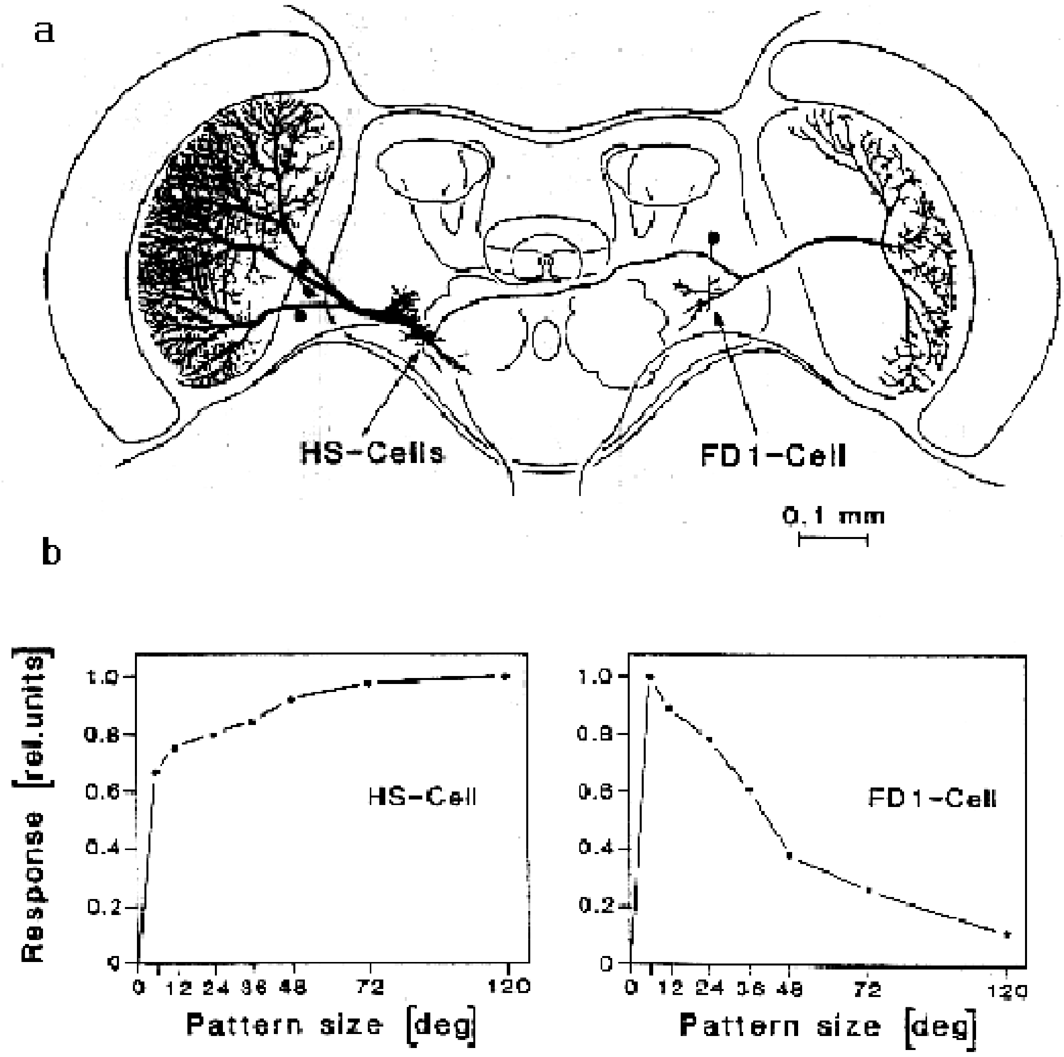


FIGURE 2.2. Spatial integration properties of neural elements. (a) The position of the HS cell and the FD cell in the fly lobula plate. (b) Dependence of the mean response amplitude of a HS cell and a FD cell on the size of the stimulus pattern. The response of HS cell reaches its maximum for motion of large patterns, whereas the FD cell responds strongest when a small pattern is moved in its receptive field. Reproduced without permission from Egelhaaf and Borst (1993a).

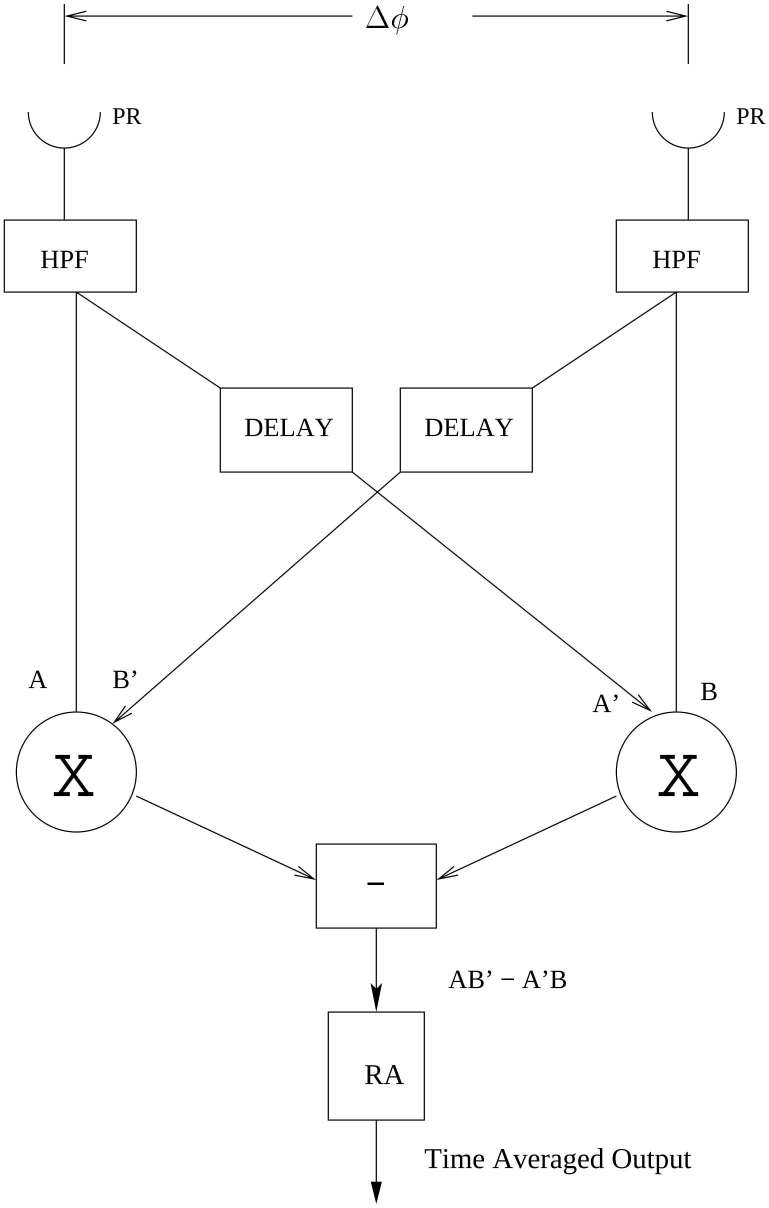


FIGURE 2.3. Modified Hassenstein-Reichardt (HR) model. The photoreceptor (PR) collects visual information and a high-pass filtering stage (HPF) enhances the features of the scene. This HPF stage was a modification introduced by Van Santen and Sperling (1985). This filtered photoreceptor response and a delayed output from an adjacent photoreceptor are correlated by a multiplication (×) stage. The difference of two such adjacent correlation stages gives a direction selective output. A running average is computed by the block RA, which gives the time averaged output from the detector.

frequency, $v = f_t/f_s$. We can re-write the above as:

$$I(x,t) = I + \Delta I \sin(\omega_t t + \omega_s x) \tag{2.2}$$

where $\omega_s = 2\pi f_s$ and $\omega_t = 2\pi f_t$.

As seen in Figure 2.1b, the two sampling points in space are separated by an angle $\Delta\phi$ from each other, so their response can be written as:

$$A = K(\omega_t)\Delta I \sin\left(\omega_t t - \omega_s \frac{\Delta\phi}{2} + \zeta(\omega_t)\right) \tag{2.3}$$

$$B = K(\omega_t)\Delta I \sin\left(\omega_t t + \omega_s \frac{\Delta\phi}{2} + \zeta(\omega_t)\right) \tag{2.4}$$

Here, $K(\omega_t)$ is the magnitude and $\zeta(\omega_t)$ is the phase response of the high-pass filter (Figure 2.3). The delay required by the HR model can easily be approximated as a low-pass filter, so the response after the delay elements is given by:

$$A' = G(\omega_t)K(\omega_t)\Delta I \sin\left(\omega_t t - \omega_s \frac{\Delta\phi}{2} + \zeta(\omega_t) + \psi(\omega_t)\right) \tag{2.5}$$

$$B' = G(\omega_t)K(\omega_t)\Delta I \sin\left(\omega_t t + \omega_s \frac{\Delta\phi}{2} + \zeta(\omega_t) + \psi(\omega_t)\right) \tag{2.6}$$

where, $G(\omega_t)$ is the magnitude and $\psi(\omega_t)$ is the phase response of the temporal low-pass filter, respectively. The correlation of these delayed and non-delayed responses is fed to a subtractor to obtain an opponent output, which is positive for motion in one direction and negative for motion in the opposite direction. This opponent output is given as:

$$O(t) = AB' - A'B = (\Delta I)^2 G(\omega_t) K^2(\omega_t) \sin(\psi(\omega_t)) \sin(\omega_s \Delta\phi) \tag{2.7}$$

This is directionally selective based on the value of $\omega_t = 2\pi v f_s$, where v can have both positive and negative values, hence the opponent output. Note that this output is independent of both x and t, so the position of the motion detector and the time at which the motion occurred does not affect the output of this detector.

This correlation-based technique also takes care of the aperture problem seen in many computer-based velocity determining algorithms. It is known that due to the limited scope of the aperture, the only component of motion that can be computed is oriented perpendicular to the orientation of the segment (Hildreth and Koch, 1987). With correlation-based detectors, apart from certain locations, the two dimensional (2-D) velocity field can be correctly estimated (Reichardt and Egelhaaf, 1988). The correlation model of motion detection has been extensively studied and reviewed in flies. The plausible neural circuit for multiplication has been a questioned stage in the HR model based detector. If the inputs to the multiplicative stage are supposed to be x and time delayed y, the required operation should result in xy. An equivalent form of HR detector multiplies x and $1 - y$, instead of y. The equivalent to this is derived from shunting inhibition (Furman, 1965) in which x is divided by $1 + y$. The first term Taylor series expansion of this gives a reasonable approximation to the multiplication stage and contains the term xy (Torre and Poggio, 1978).

2.3 The Adelson-Bergen Algorithm

The computations in the HR detector are inherently opponent and are composed of a leftward and a rightward motion response. However, the terms AB' and BA' in Figure 2.3 are not leftward and rightward motion energies. Therefore, independent detection of leftward and rightward motion at threshold is not readily accommodated in the HR model (Adelson and Bergen, 1985). Motion is the orientation in space and time. Hence, a measure of this spatio-temporal energy can be used to measure the motion in any orientation (direction). This is the premise of the Adelson-Bergen model, in which the spatial and the temporal impulse response units act as the weighting functions and their combined response gives the spatio-temporal impulse response unit (Adelson and Bergen, 1985). In a sub-unit, there are two spatial filters (SF) followed by two kinds of temporal filters, with two different delays associated with them (Figure 2.4). These low-pass filtered responses (A, A', B, B') are combined to get four different oriented responses. Each of these is squared and added to give the oriented energy. For a 1-D case similar to that discussed in the HR model, the oriented energies are leftward and rightward. The two sets of opposite-oriented energies are subtracted to give the opponent motion output. The spatial filters are modeled as Gabor filters in quadrature and are mathematically expressed as:

$$f_{s1}(x) = e^{-\frac{x^2}{2\cdot\sigma^2}} \cdot \cos(\omega_x \cdot x) \tag{2.8}$$

$$f_{s2}(x) = e^{-\frac{x^2}{2\cdot\sigma^2}} \cdot \sin(\omega_x \cdot x) \tag{2.9}$$

The temporal filters are second and third order Gaussian functions, the temporal impulse response of which is written as:

$$f_{t1}(t) = (kt)^3 \cdot e^{-kt^2} \cdot \left[\frac{1}{3!} - \frac{(kt)^2}{(3+2)!}\right] \tag{2.10}$$

$$f_{t2}(t) = (kt)^5 \cdot e^{-kt^2} \cdot \left[\frac{1}{5!} - \frac{(kt)^2}{(5+2)!}\right] \tag{2.11}$$

A detailed derivation of this model has been done by Korrapati (Korrapati, 2001). The final result is presented here as the result of subtraction of opponent energies for a sinusoidal grating with spatial filters tuned to a spatial frequency $\omega_x = 2\pi f_x$ and temporal filters tuned to temporal frequency $\omega_t = 2\pi f_t$.

$$O(t) = 4 \cdot I^2 \cdot \underbrace{|f_{s1}| \cdot |f_{s2}| \sin(\phi_{s1} - \phi_{s2})}_{spatial} \cdot \underbrace{|f_{t1}| \cdot |f_{t2}| \sin(\phi_{t1} - \phi_{t2})}_{temporal} \tag{2.12}$$

where $|f_{s1}|$ and $|f_{s2}|$ are magnitudes of the two spatial filters, $|f_{t1}|$ and $|f_{t2}|$ are the magnitudes of the two temporal filters, ϕ_{s1} and ϕ_{s2} are the phases of the spatial filters, and ϕ_{t1} and ϕ_{t2} are the phases of the temporal filters. As seen, the Adelson-Bergen algorithm can also be expressed as the product of separable spatial and temporal frequency terms. The output in itself is independent of the position x and time t and has been shown to be an equivalent of the HR model (Van Santen and Sperling, 1985; Adelson and Bergen, 1985). This algorithm is used as a model for the motion response of a primate cortical complex cell. The next two sections describe the two principal mechanisms believed to underlie target tracking in insects.

2.4 FD Cell Based Target Fixation

A fly can distinguish a small moving object from its background on the basis of their relative motion alone (Reichardt *et al.*, 1983). Figure-ground discrimination has been extensively studied in flies, and neural circuitry for the same has been proposed by Egelhaaf (1985b). It is believed that two functional classes of output elements of the visual ganglia are involved in figure-ground discrimination by relative motion in the fly: horizontal cells that respond to large textured patterns, and FD cells which are most sensitive to small moving objects (Egelhaaf, 1985a). Input to the FD cells is provided by two retinotopic arrays of small-field (SF) elementary motion detectors, responding either to front-to-back or back-to-front motion (see Figure 2.2). Four types of FD cells (from front-to-back FD1 and

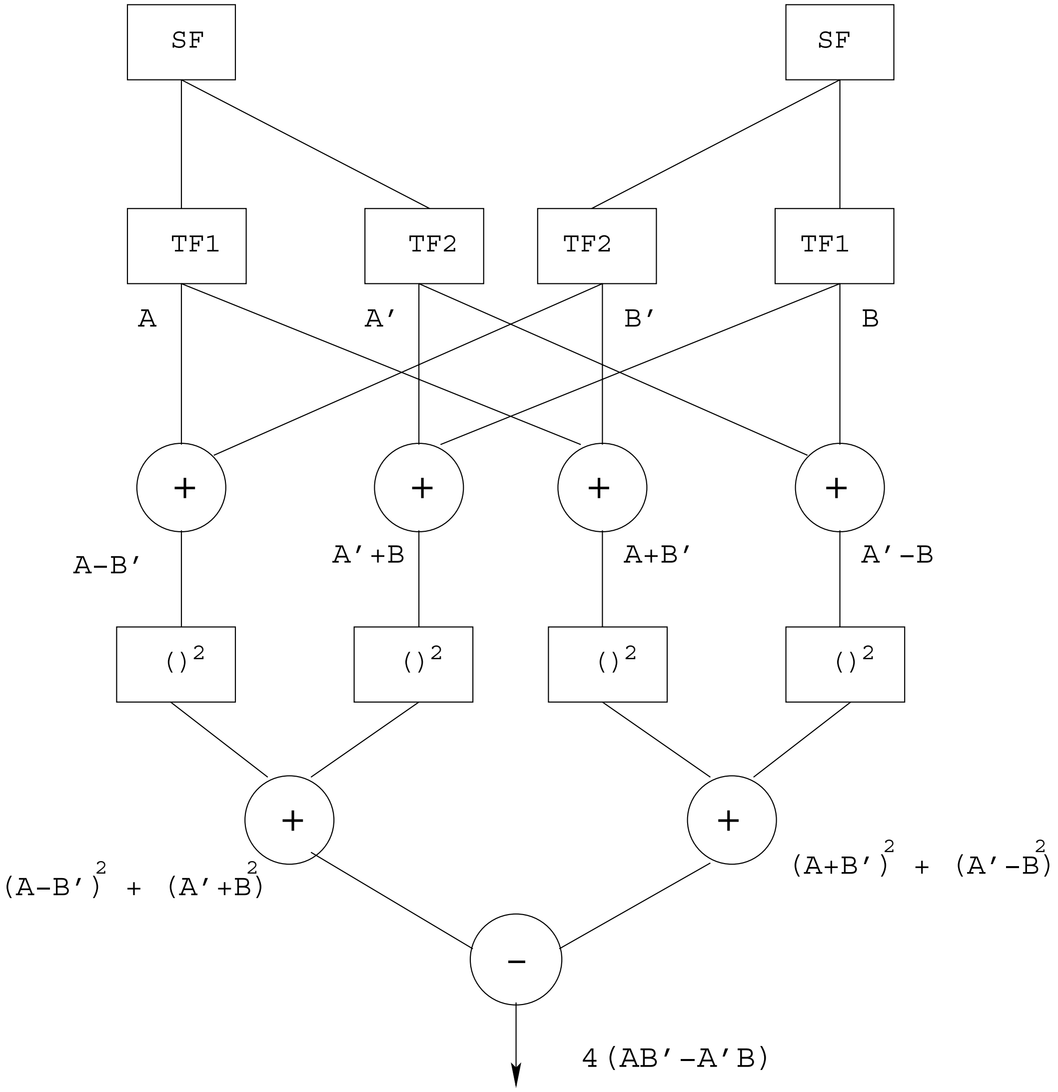


Figure 2.4. Adelson-Bergen motion detector. The input stage consists of spatial filters (SF). A pair of temporal filters (TF1 and TF2) with different time constants treats the output from the SF stage. The summation of two such responses from adjacent receptors gives oriented linear responses. Squaring and adding these gives oriented motion energies. The difference between two such oppositely oriented energies gives a direction selective opponent output.

FD4, and back-to-front FD2 and FD3) receive these motion responses. According to their preferred direction of motion, FD cells are excited by one of these types of motion detectors and inhibited by the other. The inhibition is brought about by synaptic transmission from pool cells (see Figure 2.5) that aggregate the response of the motion detectors from the entire visual field. FD cells are believed to receive at least four different types of motion sensitive input (Egelhaaf, 1985a):

1. Excitatory input by small field motion from the ipsilateral visual field.
2. Inhibitory input by the same part of the visual field, elicited by motion of opposite polarity.
3. Inhibitory input by ipsilateral large-field pool cells sensitive to motion in the preferred direction.
4. Inhibitory input by contralateral large field motion in the null direction.

A postulated neural circuit model based on the FD cells has been described by Egelhaaf (1985b). The FD cell based small-field motion detector unit evaluates the discontinuities in the retinal motion field induced due to the relative motion of a small object to a distant background. Thus, it can indicate a nearby moving object in the presence of a moving background. It can therefore be a part of the fixation system inducing turning response towards targets. The model was further modified by Reichardt *et al.* (1989) and this model circuit is presented in Figure 2.5. The cup-shaped elements are the photoreceptors (PR) and two such adjacent PR units are input to elementary motion detectors (EMDs) based on the HR model. The detector output is split into positive (v_i^+) and negative (v_i^-) responses ($v_i^+ > 0$ and $v_i^- < 0$). As discussed in the previous section, the HR detector sub-units do not have leftward and rightward motion energy. So, this oriented output (v_i^+ or v_i^-) is computed externally from the opponent output of the detector, such that:

$$v_i^+(t) = pos(R_{HR}(t)) \tag{2.13}$$

$$v_i^-(t) = g * neg(R_{HR}(t)) \tag{2.14}$$

where R_{HR} is the HR detector response, and $g < 1$ is a factor indicating that the progressive response is stronger than the regressive response. Two sets of pool cells spatially sum the EMD responses and are thus weakly direction selective. The P^+ pool cells are activated by front-to-back motion and inhibited by back-to-front motion, and P^- pool cells by opposite polarity. The inhibitory response is weighted by a factor $T < 1$ and the pool cell response is given by the following expressions:

$$P^+(t) = \sum [v_i^+(t) + T \cdot v_i^-(t)] \tag{2.15}$$

$$P^-(t) = -\sum [v_i^-(t) + T \cdot v_i^+(t)] \tag{2.16}$$

The pool cells from both sides of the brain are assumed to interact with individual EMDs. The right and left side responses can be combined to form clockwise (cw) and counter clockwise (ccw) turning response with a relative contribution of ipsilateral and contralateral input of $0 < k < 1$.

$$P_{right}^{cw}(t) = P_{right}^+(t) + k \cdot P_{left}^-(t) \tag{2.17}$$

$$P_{right}^{ccw}(t) = P_{right}^-(t) + k \cdot P_{left}^+(t) \tag{2.18}$$

It must be noted that this model is tuned only to the rotatory motion of a fly. This leads to response of opposite sign from the two eyes, since small-field units in both the eyes are modeled to have a positive response to a motion away from the heading direction of the fly. Therefore, this does not handle the translatory motion while tracking a target. Modifications to this model to combine both translatory and rotatory motion are discussed in Chapter 5. The P_{right}^{cw} and P_{right}^{ccw} responses interact with the individual EMD responses. The relative contribution of both the pool cells to individual

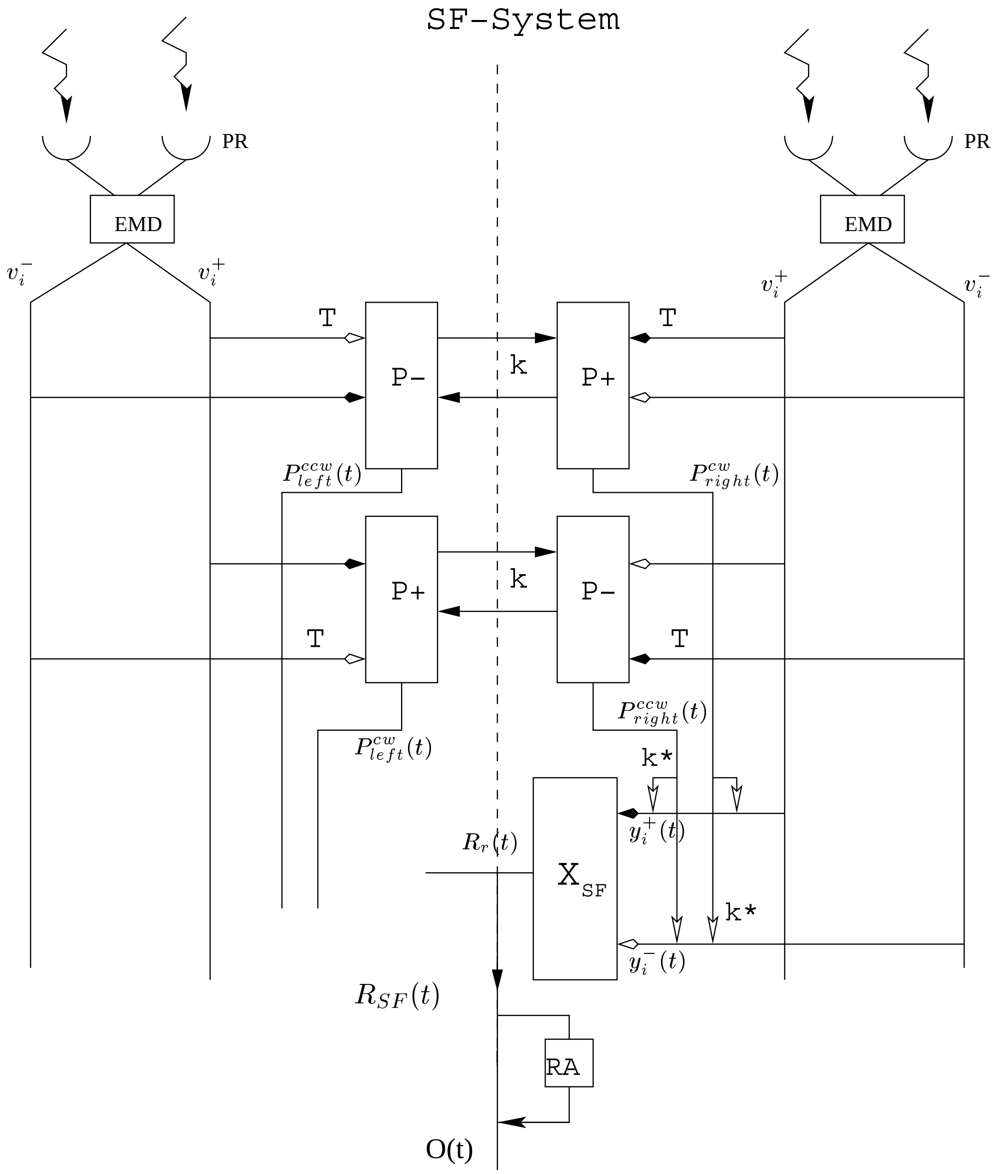


FIGURE 2.5. Model of small-field motion system based on the FD cell. Responses from two adjacent cup-shaped photoreceptors (PR) are input to the motion detectors (EMDs). These EMDs pool their positive and negative input to two types of directionally selective binocular pool cells (P^+, P^-). Excitatory and inhibitory synapses are indicated by black and white tipped arrows, respectively. The contralateral side response is weighted by the factor $0 < k < 1$. Inhibitory input channels synapsing on the pool cells are weighted by the factor $0 < T < 1$. These directionally selective pool cells interact with the individual motion detector output channels prior to their summation by the output cell (X_{SF}). The influence of P^+ and P^- response may differ from one another by a factor $0 < k^* < 1$. The output is a combination of the response from X_{SF} and a running average of this output calculated by RA. Modified from Reichardt, Egelhaaf, and Guo (1989).

EMDs is weighted by a factor $0 < k^* < 1$ to maintain the direction selectivity. This is so because when $k*$ becomes one, each motion component gets divided by the sum of all the motion components. This interaction is usually modeled as a weak response. The response of the detector channels, after shunting inhibition by the pool cells and assuming the saturation of the output approximated by an exponent $q < 1$, is given by:

$$y_i^+(t) = \frac{v_i^+(t)}{\beta + [P_{right}^{cw}(t) + k^* \cdot P_{right}^{ccw}(t)]^q} \tag{2.19}$$

$$y_i^-(t) = \frac{v_i^-(t)}{\beta + [P_{right}^{ccw}(t) + k^* \cdot P_{right}^{cw}(t)]^q} \tag{2.20}$$

The responses are non-linearly transmitted and the excitatory and inhibitory responses are accumulated. The response for one of the eyes is given as:

$$R(t) = \sum_{i=1}^{N} [y_i^+(t)]^n - [|y_i^-(t)|]^n \tag{2.21}$$

The responses from these units on either side of the brain are further subtracted from each other.

$$R_{SF}(t) = R_r(t) - R_l(t) \tag{2.22}$$

The response of the output cell is a combination of the real time response and a running average of the response.

$$O(t) = RA + R_{SF}(t) \tag{2.23}$$

This model response matches very well with the electrophysiological recordings of the torque response of a fly to a moving target in front of a background.

2.5 Male Specific target pursuit system

It has been argued by another leading group of researchers that the FD cell based model for target fixation is not the primary mechanism behind target pursuit (Gronenberg and Strausfeld, 1991; Douglass and Strausfeld, 1995; Buschbeck and Strausfeld, 1996). During sexual pursuit, the target is stabilized in the dorso-frontal acute zone of the compound eyes in male flies (see Figure 2.6). Visual control of insect flight has been observed by compensatory optomotor stabilization of the visual panorama in response to imposed perturbations on the head and body, by tracking and retinal stabilization of small targets, by recognition of shape and color, and by distance perception in navigation (Gronenberg and Strausfeld, 1991). Sex-specific behavior including fast saccades in response to small rapidly moving targets and the pursuit and interception of females by males amongst some species of Diptera was shown by Land and Collet in 1974. The participation of a system of position detectors beneath most of the retina that specifies the geometrical relationship between the pursuer and the pursued is common to both sexes. However, it is argued that an additional role is played by the retinal zone of high acuity in male flies. The responses of the motor channels are weighted according to their location beneath the retina. Hence, those located laterally to the fly body provide more excitation to pre-motor channels than those located medially. There is a transfer of control when the target falls within 0° to 35° from the mid-line and the relative angular velocity controls the angular velocity of the pursuer, not the error angle as in case of larger angular separation from the mid-line. Gilbert and Strausfeld in 1991 argued that 50 pairs of descending neurons (DNs) found in discrete clusters receive afferents from male-specific tangential (MLGs) and columnar (MCol) cell that subtend the retinotopic mosaic.

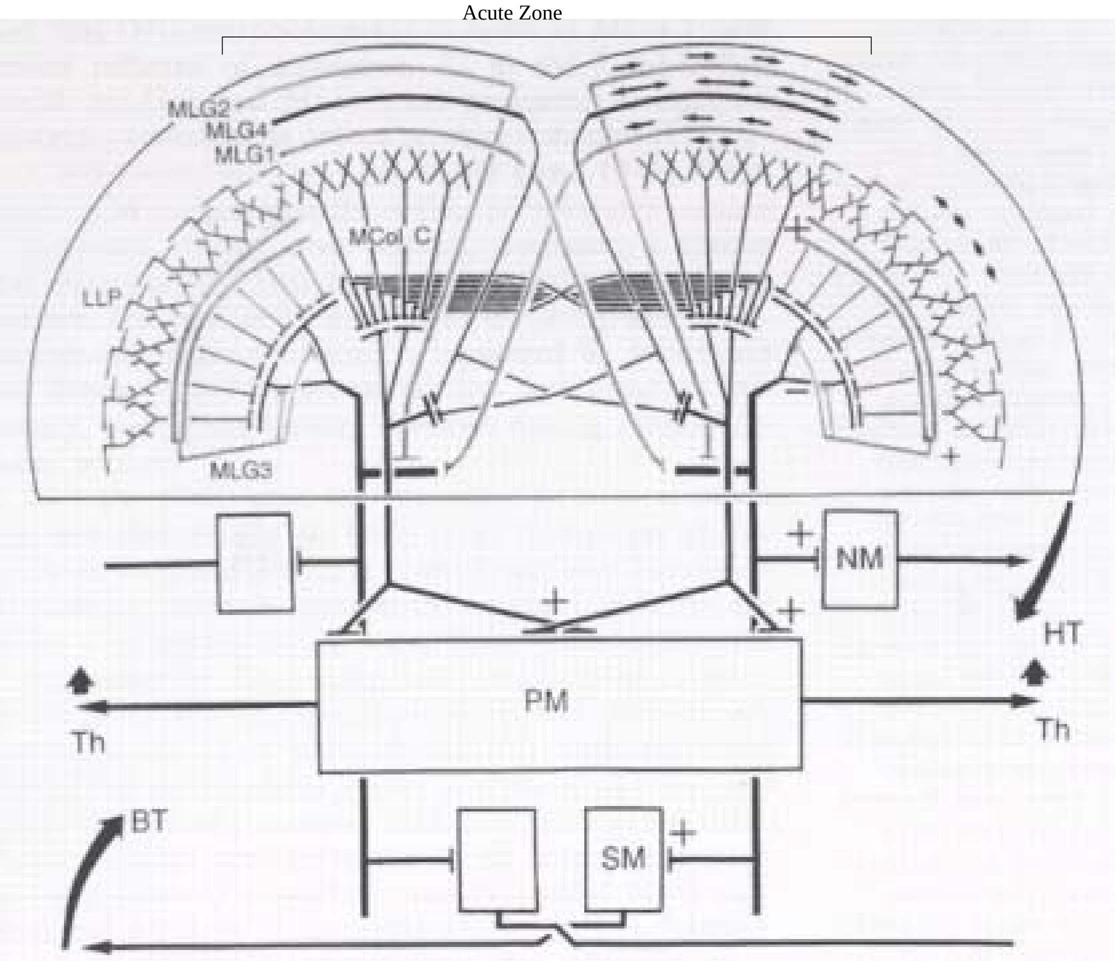


FIGURE 2.6. Gronenberg and Strausfeld model. The velocity system is formed by MLG1 and 2 in the acute zone which serves to control the neck muscle and steering flight muscle motor neurons. The position system is formed by the LLP layer and its interaction with MLG3 is used to bring the target into the acute zone. Together, these drive the yaw response in the head and body (HT, BT). Reproduced without permission from Gronenberg and Strausfeld (1991).

Electrophysiological recordings from MLG1 in flies show phasic ON responses to stationary flashes and a panoramic flow-field, but not to ramp-like increases of light intensity in the acute zone. This indicates:

1. These cells respond to the onset of motion without being directionally selective.
2. These cells respond to motions within the upper frontal field of view.

However, this stimulus does not elicit any response when presented to lateral ommatidia. Together, this suggests that MLG1 is sensitive to “novel initiation” of motion by small contrasting objects in the acute zone.

The modified Land and Collet model circuit is shown in Figure 2.6. This model predicts the yaw torque compensations based on target error angle. Neurons that bridge the lobula and lobula plate (LLP and LPL neurons) form the small-field position sensitive response layer. MLG1 and 2 are velocity sensitive target detectors responding to counter-clockwise and clockwise directions respectively. The male-specific neurons discussed here are purported to play a decisive role in male pursuit of female flies.

The remaining portion of the thesis is dedicated to the implementation of motion and target-tracking algorithms. The next chapter discusses a monolithic implementation of a target tracking sensor in VLSI.

CHAPTER 3

MONOLITHIC TARGET TRACKING SENSOR

Tracking a moving object from a moving platform in a cluttered environment is a computationally demanding task for even the most sophisticated digital computers. Conventional systems that record the object positions in a frame by frame manner and use pattern matching for tracking a target are widely used. However, such systems require high-speed communication buses and hence increase the cost and power dissipation of the system. Biologically inspired implementations of algorithms for target fixation and pursuit provide an alternative solution to this computationally extensive task. The low power, high speed, compact parallel processing units in an analog processor present significant benefits.

In this chapter, the simulations of an FD cell based monolithic tracking sensor and its VLSI implementation are presented. The FD cell based model is a model of the fly's small target fixation behavior. Previous research has revealed that the optomotor response and object fixation behavior may emerge from the joint action of two parallel neural pathways. The HS and FD cells are thought to provide these parallel pathways (Egelhaaf and Borst, 1993a). These pathways are differentially tuned to extract the motion information necessary to perform both behaviors. The FD cell based sensor has been chosen over the Gronenberg and Strausfeld type of circuit because of its simplicity and ease of implementation. It must be noted that the Land and Collet circuit has complex wiring arrangements and requires a larger variety of motion-computation elements.

In recent years, many neuromorphic analog VLSI parallel processors implementing related tracking algorithms have been developed. An analog implementation of centroid localization using focal plane current mode circuits to find a global centroid was done in the early 1990s (DeWeerth, 1992). Neuromorphic hardware emulation of visual tracking based on saccadic eye movement systems triggered by temporal change (Horiuchi *et al.*, 1994) and a smooth pursuit system based on visual motion detectors (Etienne-Cummings *et al.*, 1996) have also been accomplished. In primates selective visual attention is involved in the generation of saccadic (Kowler *et al.*, 1995) and smooth pursuit eye movements (Khurana and Kowler, 1987). This selective attention is believed to be the mechanism for extracting the activity of neurons associated with the target at appropriate time. Koch and Ullman's (1985) saliency map based attentional shifts were modeled by DeWeerth and Morris (1994). A one-dimensional visual tracking chip was implemented by Horiuchi *et al.* (1996) using analog VLSI techniques to model selective visual attention in the control of saccadic and smooth pursuit eye movements in primates. A high-contrast detection and tracking system using a current-mode hysteretic winner-take-all network has been developed by Indiveri *et al.* (2002).

Here the FD cell based small-field motion detector and its use as a target tracking sensor is discussed. This sensor differs from all the above mentioned implementations of tracking systems in two ways. First, it is based on a well understood model of the fixation behavior of the fly and therefore has attributes that can be understood and characterized fairly easily. Second, the uses of this sensor may extend beyond target tracking, since it is generally sensitive to motion discontinuities with the background. It can be used as a small-field motion sensor for detecting small objects and fixating on them.

The FD cell model, as presented in Chapter 2, has been modified to simplify the computations and thus minimize the silicon area. The simulation results of the full FD cell model are presented in the following section. The simplified model with its analysis and an extensive VLSI architecture implementing this model are presented in the subsequent sections.

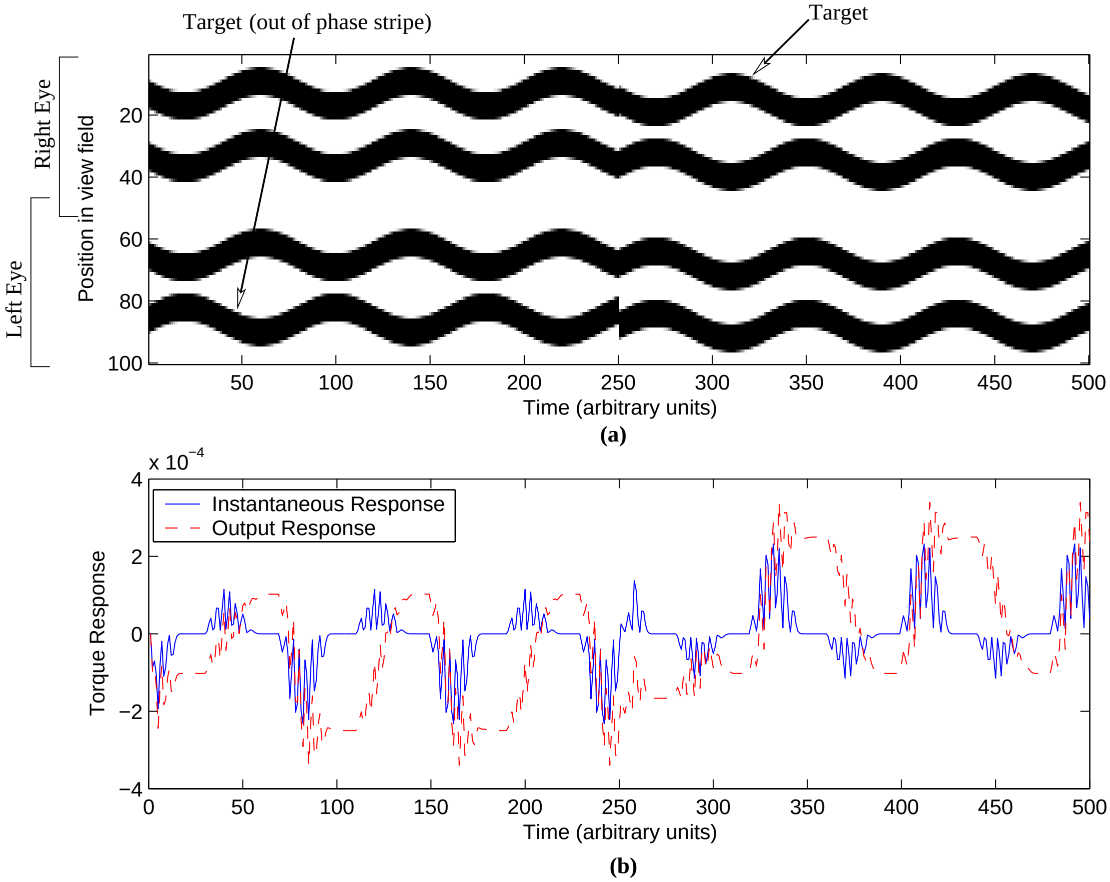


FIGURE 3.1. Simulation results of the FD cell model. (a) The stimulus used for torque response generation. (b) Torque response. The output is negative (counterclockwise yaw movement) if the activity is seen in the left eye and positive (clockwise) for discontinuity in the flow-field in the right eye.

3.1 Simulation of the full FD cell model

In this simulation experiment, the torque-generating behavior of a simulated fly equipped with an FD sensor, as modeled in Chapter 2, is tested. The fly is fixed in space and in angle such that it can only generate torque responses. The visual field of the simulated fly is comprised of 100 photodetectors and is divided into right (1 to 55) and left (45 to 100) halves (Figure 3.1a). Overlapping of the two halves of the view field is used to model the acute zone, where sensors from one eye are activated by activity on the other side of the head. The visual panorama includes 4 stripes located around the midline of the fly. The two stripes closer to the mid-line are separated by 20 units, while the lateral ones are separated by 12 units. This arrangement is similar to that of Reichardt *et al.* (1989), in which the turning response of a fly was recorded. Three stripes out of four oscillate in phase with an amplitude of 4 units, while the fourth stripe moves at a phase lag of 180°. After half completion of the experiment, the position of the lagging stripe is switched from one side of the visual-field to the other (see Figure 3.1a).

The torque response of the robot is as shown in Figure 3.1b. When a motion discrepancy is presented on the left side of the view field, the instantaneous turning response is largely negative and is shown by a solid line in the figure. According to the FD cell model, this instantaneous response is coupled with a running time average to give the final output shown by a dotted line in the figure. This final output response in the simulation is the sum of the instantaneous response and the running average response weighted by a factor $k = 5$. The average output response is seen to be negative for the first half of the experiment, indicating a counter-clockwise turning response. At $t = 250$ units, the motion discrepancy moves to the right side of the view field, causing a discontinuity in the visual motion field. This produces a positive response by the sensor indicating a clockwise turning torque. These results match the simulation results by Reichardt *et al.* (1989) closely.

3.2 Simplified FD cell model

The biological basis of the FD cell based target fixation model has been discussed in Section 2.4, and the simulation result for this model was presented in the previous section. A simplified FD cell based model of the target tracking sensor for VLSI implementation is discussed in this section. Hassenstein-Reichardt (HR) model based elementary motion detectors (EMDs) have been chosen in this model. These detectors fulfill the requirements set by the FD cell based system. Instead of being a pure velocity sensor, the HR model's response is influenced by textural properties of a moving pattern such as its spatial frequency and contrast (Reichardt *et al.*, 1989). The output from this EMD is split into positive and negative parts and fed as input to the target fixation circuit.

The processing elements of the simplified FD cell model are shown in Figure 3.2. The HR model based EMD is the first processing stage for this sensor. This detector has been analyzed in detail in Chapter 2. The output of this detector is split into its positive and negative components, v_i^+ and v_i^- respectively ($v_i^+ > 0$ and $v_i^- > 0$). In the biological model, the positive pool cell accumulates the excitatory v_i^+ and a weighted fraction of the inhibitory v_i^- (refer to Equation 2.17). Since the inhibitory response of the pool cell approximates the common property of neurons that the reversal potential of inhibitory synapses is usually closer to the resting potential than of excitatory synapses (Reichardt *et al.*, 1989), it can be neglected. This can be done without sacrificing the yaw-movement control too much as was seen by the simulation results. These simplified pooling units (positive and negative, and on the left and right side) thus perform the following function:

$$P_r^+ = \sum_i^N v_r^+(i) \tag{3.1}$$

$$P_r^- = \sum_i^N v_r^-(i) \tag{3.2}$$

$$P_l^+ = \sum_i^N v_l^+(i) \tag{3.3}$$

$$P_l^- = \sum_i^N v_l^-(i) \tag{3.4}$$

The cross connection of pool cells from the left side of the visual field to the right and vice-versa forms the clockwise and the counter-clockwise response. It must be noted that the biological FD model assumes only rotational motion of the fly. The motion detector response interacts with this clockwise/counter-clockwise response of the pool cells just before the output cell. This response for a sensor on the right side of the head is given in Equation 2.20. This equation has fitting constants β, k^*, and q that are used to match the model's results to electrophysiological recordings. These fitting constants have not been explicitly included in this simplified model. The factor β, for instance, takes care of the condition in which the summation of the response both from the left and right side of the

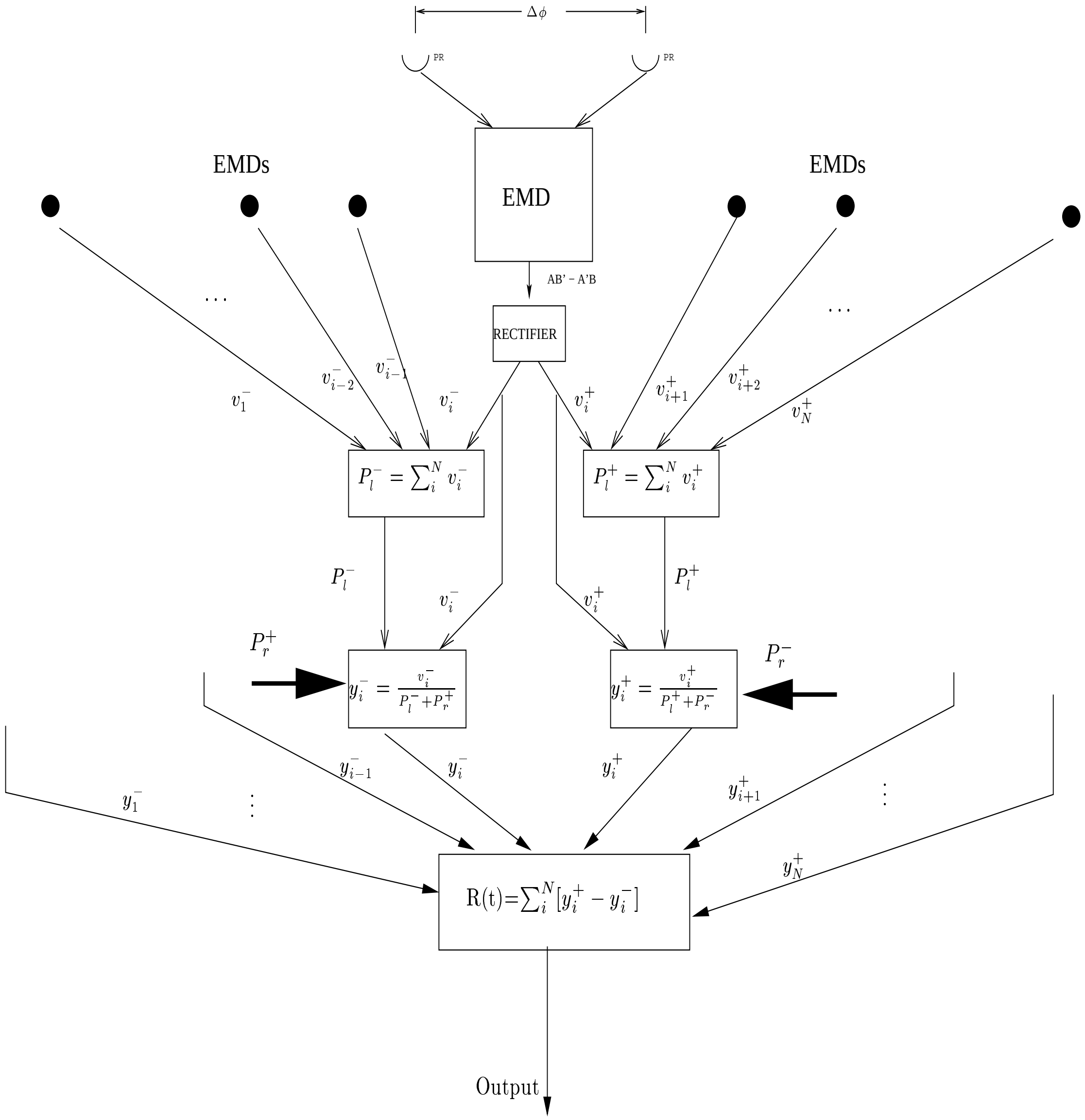


FIGURE 3.2. Simplified FD Sensor (left eye). An array of HR model based detectors pool in their outputs into two pooling units that sum the positive and the negative halves. These individual positive and negative responses are normalized by the sum of corresponding pool cell and opposite pooled response from the right eye. The normalized response from each EMD is then summed to generate the instantaneous output response.

head goes to zero. In VLSI, the normalization circuit handles that condition without any explicit representation.

The FD chip described below implements only one half of this model (either left or right). However, two chips may be used to model the left and the right side sensor. The response of the normalized small-field outputs in the simplified algorithm is given by:

$$y_i^- = \frac{v_i^-}{P_l^- + P_r^+} \tag{3.5}$$

$$y_i^+ = \frac{v_i^+}{P_l^+ + P_r^-} \tag{3.6}$$

In the biological model, these opposite responses pass through a non-linear channel and add up to generate a final response $R(t)$. This non-linearity suppresses the small responses, and expands larger ones, thereby expanding the mean difference of these two signals. In VLSI, however, the non-linear exponentiation behavior of the FD model has not been included for simplicity. Therefore, the small-field output response of the simplified FD model is weaker as compared to its biological counterpart. The response $R(t)$ shown in Figure 3.2 must be computed for both left and right "eyes". The sum of the response of this output cell from one (left or right) eye is given as:

$$R(t) = \sum_{i=1}^{N} [y_i^+ - y_i^-] \tag{3.7}$$

The final output of the small-field system is a difference between the left and right side FD chip responses.

3.3 VLSI Design of Tracking sensor

This section discusses a VLSI architecture (see Figure 3.3) implementing the small-field motion computation presented in the previous section. The EMDs have been modeled as a modified version of Harrison's HR model based chip (Harrison and Koch, 2000). The detectors are followed by a full-wave current splitting circuit to divide the output into positive and negative parts. This output is then fed to a current normalization circuit that mimics the synaptic interaction between the motion detector outputs and the pool cells. An external current from the contralateral side sensor (Figure 3.3b) is also an input to the current normalization circuit. The VLSI implementation that we will present accounts for this by the use of two chips, each implementation having a preferred direction of motion for which the output of the chip is positive. The simplified FD algorithm discussed above requires interaction between preferred and null direction of these two chips. Therefore, the two FD chips are arranged in opposite orientation as illustrated in the figure. These two sets of outputs from the positive and negative normalization circuit are subtracted, and then taken off-chip. The sub-circuit components of this design are described below.

3.3.1 Adaptive Photoreceptor

Modified adaptive photoreceptors (Delbrück, 1993b) are used in this implementation for capturing contrast information from the visual field (Liu, 1999). The adaptive photoreceptor has a continuous-time output that attains low values for static signals and exhibits high gain for transient light signals that are centered around a background adaptation point. Logarithmic encoding of input illumination makes the response largely invariant to absolute light intensity. In Delbrück's receptor, the adaptation time constant is predefined at the design phase. The adaptation time constant of this modified receptor is controllable via a bias voltage *Vb*. The modified adaptive photoreceptor is shown in Figure 3.4.

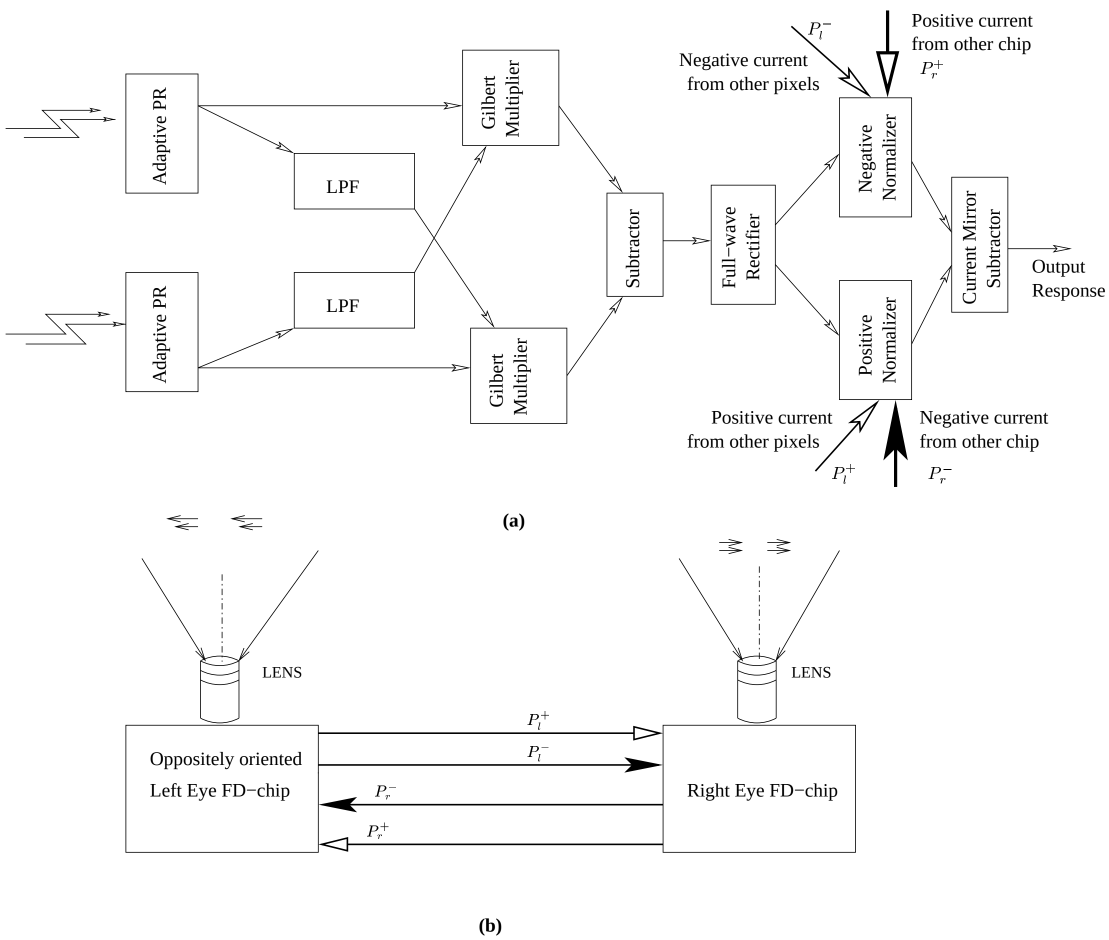


Figure 3.3. Block diagram representation of the FD chip. (a) Phototransduction takes place through modified adaptive photoreceptors. The delay stage is modeled as a low-pass filter (LPF). Correlation between delayed and non-delayed signals takes place in a current-mode Gilbert multiplier circuit. A subtractor circuit computes the difference of the two adjacent Gilbert multipliers and inputs it to the full wave rectifier. This circuit splits the output into a pure negative and a pure positive component. A normalization circuit takes this single polarity current and normalizes it with respect to the sum of the same polarity response from other pixels and an external input corresponding to the contralateral response from another FD chip. A current subtraction circuit takes the normalized positive and negative outputs from the previous stage and performs a subtraction to give the final output. (b) Two FD sensor chips acting as the left- and right-side sensors are shown. As seen, when there is progressive motion of the imager the optical flow-fields for the two sensors are in opposite directions (shown by arrows). The left FD chip is oriented in the opposite direction from the right FD chip so that they both respond with positive motion outputs to the flow field vectors shown. The black arrows indicate currents corresponding to the positive half, while the white ones show negative half.

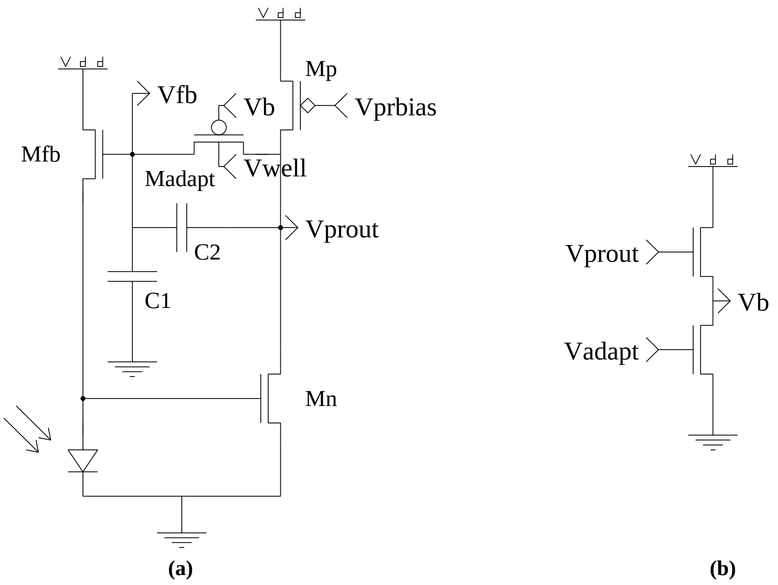


FIGURE 3.4. Adaptive photoreceptor. (a) Circuit which is sensitive to contrast changes over a wide range of illumination. (b) Adaptation-time controlling bias generation circuit.

The photodiode traps the incident photons and converts them into a current signal. A source-follower feedback transistor *Mfb* sits on top of this photodiode and operates in the subthreshold region, such that the source voltage of this transistor is logarithmic in the photo-current. This feedback signal *Vfb* is nothing but a long-term mean response of the photoreceptor. This source terminal is further connected to the gate of transistor *Mn* controlling the amount of current being sunk by it. The current source for this transistor is provided by *Mp*. The transistor *Mp* has an external bias *Vprbias* and together *Mn* and *Mp* act as an amplifying unit with the gain determined by the ratio of the sum of transconductances of *Mn* and *Mp* to the the output conductance of the amplifier. The feedback loop is comprised of a capacitive divider formed by $C1$ and $C2$. In parallel to this capacitive feedback, transistor *Madapt* provides a high resistance feedback path to adapt out slowly varying contrast signals. The adaptation time constant of this photoreceptor is controllable through a bias voltage *Vb*, generated via a source follower configuration from the photoreceptor output *Vprout* and an external bias *Vadapt*. This photoreceptor gives a wide range of intensity indifferent responses.

The HR detector implemented by Harrison and Koch (2000) has a high pass filter circuit cascaded to the adaptive photoreceptor. This is not necessary since the adaptive photoreceptor itself has a band-pass characteristic, which is sufficient for the present application. Thus, the present implementation has an advantage over its predecessor in terms of smaller silicon area and fewer external biases.

3.3.2 Low Pass Filter (LPF)

The HR detector has two delay stages that can be modeled as low-pass temporal filters. The common g_mC filter is utilized for this purpose. A P-type filter is used ensure that critical information is not lost, which might happen due to a low common-mode input voltage in an N-type filter.

Figure 3.5a shows the transistor level implementation of such a filter. The bias current through

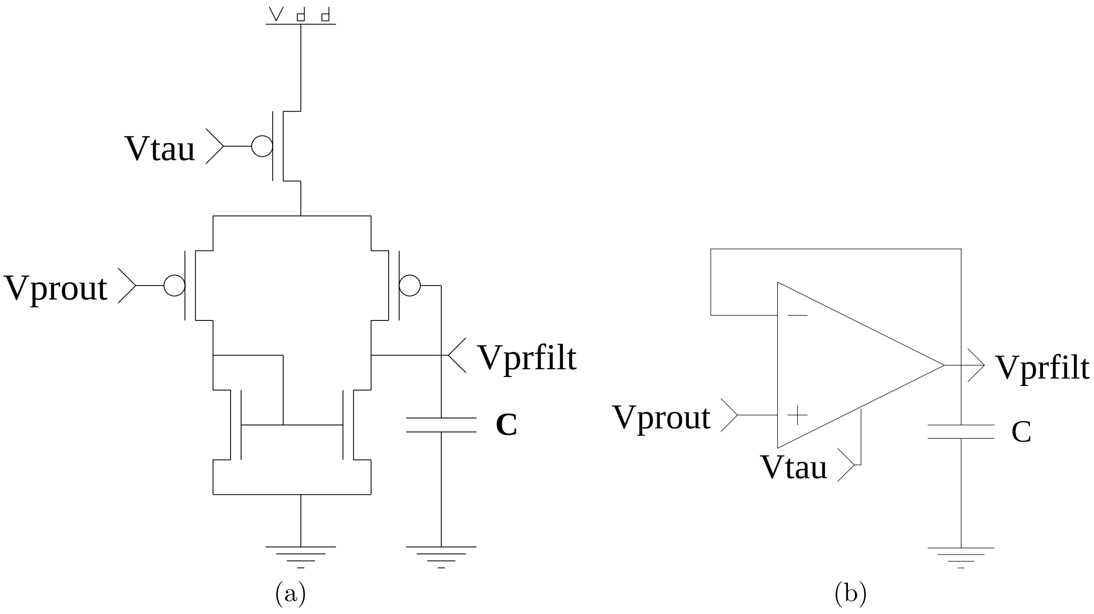


FIGURE 3.5. Low pass filter (LPF) circuit. (a) Circuit performs the low pass filtering on the photoreceptor output. (b) Symbolic representation of the circuit.

this g_mC filter is controlled by a bias voltage *Vtau*, which thereby controls the time constant of the filter. The op-amp equivalent circuit of the same is shown in Figure 3.5b. The input signal to this stage is the phototransduction stage output, *Vprout*. The output current of this stage is given as:

$$C \cdot \frac{dVprfilt}{dt} = I_b \cdot \tanh\left(\frac{\kappa(Vprfilt - Vprout)}{2 \cdot V_T}\right) \tag{3.8}$$

where C is the capacitance, I_b is the bias current controlled by *Vtau*, V_T is the thermal voltage, and κ is the back gate coefficient. For small differences, the hyperbolic tangent function can be approximated as a linear function. So the above equation takes the form:

$$C \cdot \frac{dVprfilt}{dt} = \frac{I_b \cdot \kappa}{2 \cdot V_T} \cdot (Vprfilt - Vprout) \tag{3.9}$$

Taking the Laplace transform of the above and manipulating the coefficients we get

$$Vprfilt = Vprout \cdot \frac{1}{\tau \cdot s + 1} \tag{3.10}$$

where τ is the time constant of the low-pass filter, given as:

$$\tau = \frac{2 \cdot V_T \cdot C}{I_b \cdot \kappa} = \frac{C}{g_m} \tag{3.11}$$

and the transconductance is thus

$$g_m = \frac{\kappa \cdot (I_b/2)}{V_T} \tag{3.12}$$

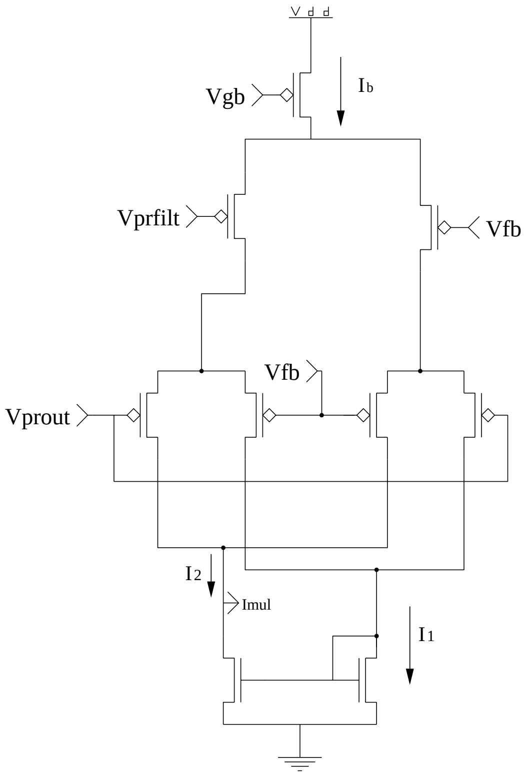


Figure 3.6. Gilbert multiplier circuit. This circuit multiplies the photoreceptor response with the delayed LPF output. The output is a represented by a current I_{mul} (for details see text).

3.3.3 Gilbert Multiplier

The delayed signal derived from the filtering stage is to be correlated with the non-delayed photoreceptor response of the adjacent detector. The simplest mechanism of correlation is employed here by multiplying the two signals. The Gilbert multiplier circuit (Gilbert, 1974) has been implemented as seen in Figure 3.6. This implementation differs from Harrison's implementation of the HR detector in two ways. Firstly, a P-type circuit has been used in the Gilbert multiplier stages. Since the photoreceptor and low pass filter outputs have a DC offset of one or two volts, the P-type circuit ensures that all the transistors will operate in the saturation region. Secondly, Harrison has used a high-pass filtering stage after the phototransduction to get rid of the DC bias (Harrison and Koch, 2000). This is achieved in this implementation by subtracting the long-term mean response of the photoreceptor, *Vfb*, from both the photoreceptor response and the filtering stage output. This makes our implementation more compact. It also has fewer external biases and is therefore more robust to off-chip fluctuations. The Gilbert multiplier circuit operates for a small range of inputs with all the transistors working in the subthreshold region. In this region, the MOS operation is similar to BJTs, both having an exponential voltage-to-current relationship.

The Translinear Network (TN) theorem (Gilbert, 1975) is used to analyze this circuit. Considering the translinear loops between *Vprfilt*-*Vfb* and *Vprout*-*Vfb*, after solving the loop equations using

Kirchoff's laws and substituting these values we arrive at the following current equations:

$$I_1 = I_b \cdot \frac{e^{V_2/V_T}}{e^{V_2/V_T}+1} \cdot \left[\frac{1}{e^{V_1/V_T}+1}\right] + I_b \cdot \frac{1}{e^{V_2/V_T}+1} \cdot \left[\frac{e^{V_1/V_T}}{e^{V_1/V_T}+1}\right] \tag{3.13}$$

$$I_2 = I_b \cdot \frac{e^{V_2/V_T}}{e^{V_2/V_T}+1} \cdot \left[\frac{e^{V_1/V_T}}{e^{V_1/V_T}+1}\right] + I_b \cdot \frac{1}{e^{V_2/V_T}+1} \cdot \left[\frac{1}{e^{V_1/V_T}+1}\right] \tag{3.14}$$

where I_1 and I_2 are the currents in the lower two branches of the circuit ending in a current mirror. The current I_b is the bias current regulated by a bias voltage Vgb, V_T is the thermal voltage, and voltages V_1 and V_2 are given as:

$$V_1 = Vprfilt - Vfb \qquad and \qquad V_2 = Vprout - Vfb \tag{3.15}$$

The output current I_{mul} is the difference between these two currents performed by the current mirror stage at the bottom of the Gilbert multiplier. This current output is equivalent to the HR detector subunit output. The difference of two such multiplying stages corresponding to two adjacent pixels, gives the response equivalent to the HR detector.

$$\begin{aligned} I_{mul} &= I_2 - I_1 \\ &= I_b \cdot \left(\frac{e^{V_2/V_T}-1}{e^{V_2/V_T}+1}\right) \cdot \left(\frac{e^{V_1/V_T}-1}{e^{V_1/V_T}+1}\right) \\ &= I_b \cdot \tanh\left(\frac{V_1}{2 \cdot V_T}\right) \cdot \tanh\left(\frac{V_2}{2 \cdot V_T}\right) \end{aligned}$$

For a small range of values of V_1 and V_2, this can be approximated as a multiplication circuit.

$$\begin{aligned} I_{mul} &= I_b \cdot \left(\frac{V_1}{2 \cdot V_T}\right) \cdot \left(\frac{V_2}{2 \cdot V_T}\right) \\ &= I_b \cdot \left(\frac{Vprout - Vfb}{2 \cdot V_T}\right) \cdot \left(\frac{Vprfilt - Vfb}{2 \cdot V_T}\right) \end{aligned}$$

Here, Vfb is the long-term mean of the photoreceptor and is subtracted to remove the DC offset from the photoreceptor response. As seen by the equation above, the values of the delayed and non-delayed signal can be twice the range of values for which the circuit behaves as a multiplier, which is roughly a few hundred millivolts.

3.3.4 Full Wave Rectifier

A full-wave rectifier circuit is used to split the HR detector response into positive and negative parts. This is a very common circuit comprised of two current mirror stages, one P-type and another N-type as shown in Figure 3.7. A buffer with an NFET and a PFET connected to a common external bias *Vpol* provides the path to the current in both directions. The P-type current mirror provides the negative part (I_{neg}), while the N-type mirror provides the positive part (I_{pos}) of the HR detector response.

3.3.5 Normalization

The full wave rectified responses from the HR detector are the inputs for the FD cell based model. The pooling of the positive and the negative parts is done separately by the Gilbert normalizer

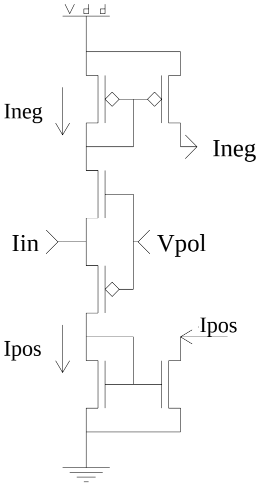


Figure 3.7. Full wave rectifier circuit. The external bias *Vpol* is adjusted such that both positive and negative currents are allowed to flow.

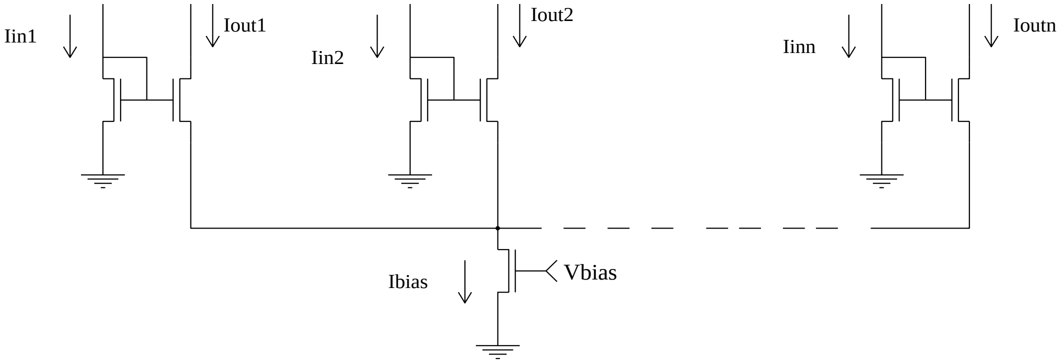


Figure 3.8. Circuit used for normalizing each input current by the sum of all inputs.

circuit (Gilbert, 1984). An n-channel normalizer is shown in Figure 3.8. This circuit is made up of several translinear loops when the transistors are operating in the subthreshold domain, such that:

$$I_{in}(i)I_{out}(j) = I_{in}(j)I_{out}(i) \qquad \forall i, j \tag{3.16}$$

The bias current I_{bias} controls the total output current flowing through this circuit with

$$I_{bias} = \sum_{i=1}^{n} I_{out}(i) \tag{3.17}$$

Solving for all the input/output currents in the translinear loops, we get:

$$\sum_{i=1}^{n} I_{out}(i) = \frac{I_{out}(j)}{I_{in}(j)} \cdot \sum_{i=1}^{n} I_{in}(i) \tag{3.18}$$

which is nothing but,

$$I_{bias} = \frac{I_{out}(j)}{I_{in}(j)} \cdot \sum_{i=1}^{n} I_{in}(i) \tag{3.19}$$

Thus, each of the output currents is a normalized version of the input current and scaled according to the available bias current.

$$I_{out}(j) = I_{bias} \cdot \frac{I_{in}(j)}{\sum_{i=1}^{n} I_{in}(i)} \tag{3.20}$$

As discussed earlier in this section, in the case of a tracking sensor, response from all the HR detecting units as well as external current from the other side of the visual field are fed to two separate positive and negative normalization circuits. Assuming the positive and negative response on the right side of the visual-field to be $I_r^+(i)$ and $I_r^-(i)$, and external currents as I_{left}^+ and I_{left}^-, we have the following positive and negative normalization current outputs:

$$I_{norm}^+(i) = I_{bias} \cdot \frac{I_r^+(i)}{\sum_{j=1}^{n} I_r^+(j) + I_{left}^-} \tag{3.21}$$

$$I_{norm}^-(i) = I_{bias} \cdot \frac{I_r^-(i)}{\sum_{j=1}^{n} I_r^-(j) + I_{left}^+} \tag{3.22}$$

The orientation of the two FD sensor chips in space are opposite of each other. This takes care that the preferred direction response of both left and right side sensors is positive. In the above equation, let us consider the scenario when all I_{in} currents are zero. This implies that the gate voltage of the NFET mirror circuits are all tied to zero (see Figure 3.8). This forces the bias transistor deep into the triode region such that the current flowing through it is only leakage current contributed by all the n-channels of the normalizer. The normalized current outputs of all channels under this condition are close to zero.

This normalized output from each processing unit implements the simplified FD cell model. A circuit representation of all the stages present in the pixel is given in Figure 3.9.

3.3.6 Serial Scanners

The response from the normalizer circuit network is read out by serial scanner circuitry. Each pixel in the two-dimensional array gives out a small field current response. This output is scanned by horizontal and vertical scanning circuits. The scanner circuits in the chip are based on the scanners proposed by Mead and Delbrück (1991). The scanners operate on a single-phase clock. The design of vertical and horizontal scanners is similar. Each scanner has a shift register to store a binary

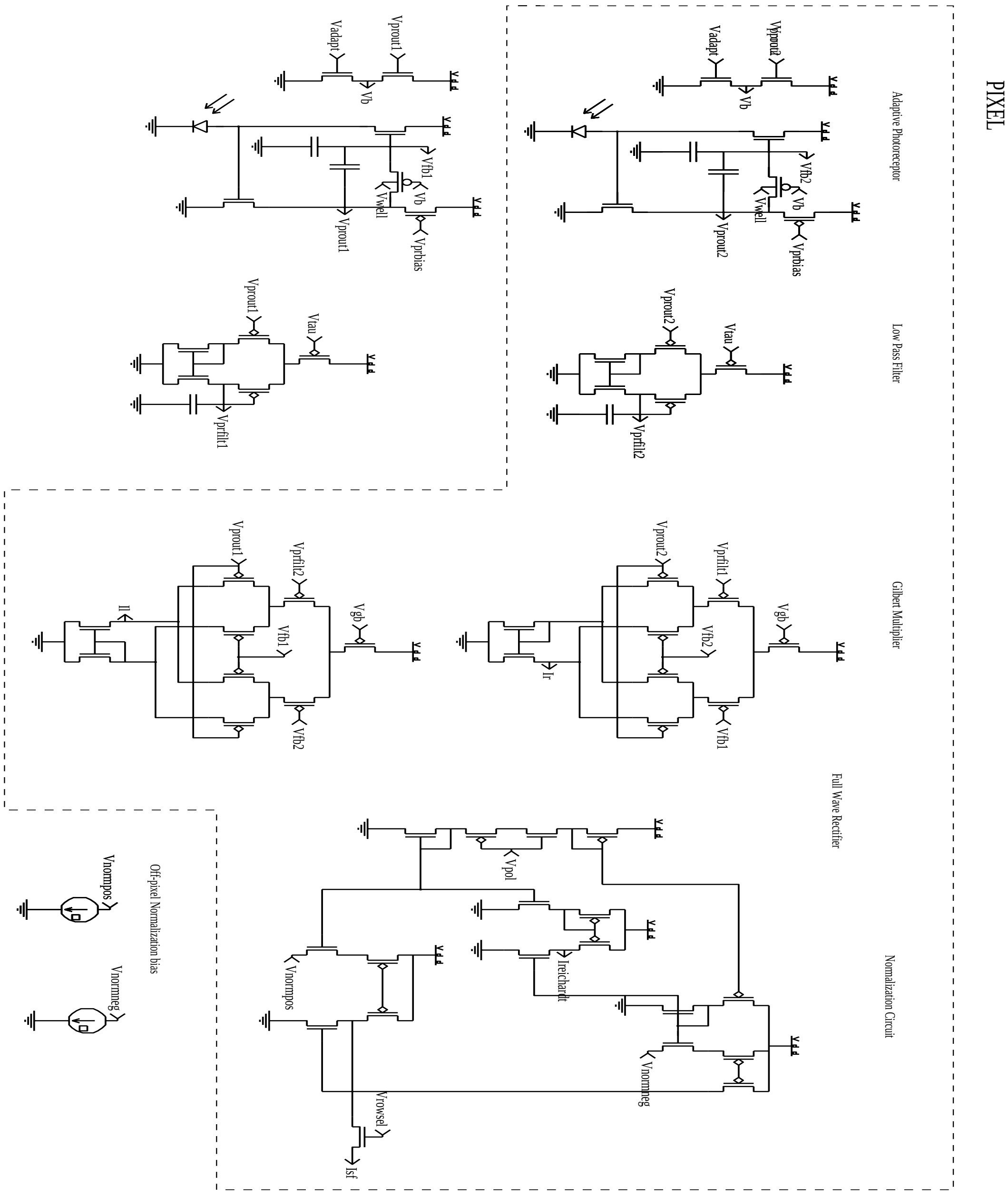


FIGURE 3.9. Circuit diagram showing all the stages in a FD pixel (within the dashed box). The photoreceptor transduces light activity in front of the sensor into electrical signals. A low-pass filter acts as a delay-stage and gives a delayed response corresponding to the photoreceptor signal. This delayed response is multiplied with the instantaneous response of an adjacent photoreceptor by the Gilbert multiplier circuit. Another Gilbert multiplier correlates the direct photoreceptor response of the pixel with a delayed response from the adjacent pixel (photoreceptor and low-pass filtering circuits for the other pixel are shown). The difference between the outputs of these two multipliers is equivalent to the response of a HR detector. A full-wave rectifier stage subtracts and then separates this output. The positive and negative currents are normalized by two normalization circuits. An off-pixel normalization bias circuit regulates the total current flowing through the normalizer. The difference of these two normalization circuit outputs is the small-field response of an individual FD pixel. This output is read out using serial scanners.

state. According to the input bits, this register selects a particular row or column. A logic high in a flip flop of the horizontal scanner selects that particular row. Similarly, a logic high in a flip flop of a vertical scanner selects that particular column. The entire chip can be scanned by continuously shifting bits from one flip flop to the other. To get a global response from the chip, all the flip flops can be made to go high. This output corresponds to the small-field response from one side of the small-field (SF) detector given in Equation 3.7. A particular row and column can be selected by sending the appropriate number of clock pulses. By selecting a particular row and column data can be read from the same pixel continuously. More details about the circuitry involved can be found in Mead and Delbrück (1991).

3.4 Simulation Results

Simulations of the FD chip have been done using the Analog simulator (Gillespie and Lazzarro, 2003).

3.4.1 HR Detector Implementation

The first simulation has been done for the HR type EMD. The detector needs input from at least two adjacent photodetectors to give a motion output. As seen in Figure 3.10, the input to the two adjacent photoreceptors is taken to be sinusoidal. The solid (darker) wave leads the dashed (lighter) one by 90° in the initial part of the simulation. The solid wave is an input to the left photodetector and the dashed one to the right as in Figure 2.3. The outputs of the photodetectors are passed through a temporal filtering (delay) stage and then correlated with the direct output of the adjacent photodetector, using the Gilbert multiplier circuit. The response from two such multiplier circuits when subtracted gives a direction selective output as seen in the lower waveform in Figure 3.10.

The input waveform in the figure has three phases. First, when the darker (solid) one leads the lighter (dashed) by 90°; second, when both are in phase; and third, when the lighter one leads the darker one by 90°. The corresponding response from the detector circuit is shown in the waveform beneath it. The detector circuit is tuned such that its preferred direction (output being positive) is left to right. Thus, when the darker wave leads the lighter one, the output is seen to be positive. When both the lighter and darker waves are in phase, the output settles to zero as predicted by the HR model. This indicates no apparent motion in the field of view. When the lighter wave leads the darker one, the output goes negative indicating that motion is taking place in the null direction.

3.4.2 FD Pixel Implementation

The output of the EMD stage is an input to the FD sensor after its output is split into positive and negative responses. In a second simulation four pixels were used to compute the small-field response. The input to the FD sensor was taken from the HR detector stage. It was then processed by the normalization circuit to generate an individual small-field response for each pixel. A target moving opposite to the background in the visual field of the sensor generates an opposite response in the HR detector. This input was simulated by presenting an out of phase current to the normalization circuit. The out of phase pixel was therefore considered as the one looking at a small moving target, while the other three pixels are looking at the background. Figure 3.11a shows the individual response of two different pixels in the sensor under three different conditions. For the condition where there was no target in the view field, all the pixels were presented with in-phase inputs. The response of the first pixel was enhanced when a target was in its view-field (different motion pattern compared to the background). The responses of all other pixels, including the second pixel under study, were suppressed in this condition. If the case of no target, all the pixels show similar response. This is shown in the middle portion of the graph. When the target moved out of the view-field of the first

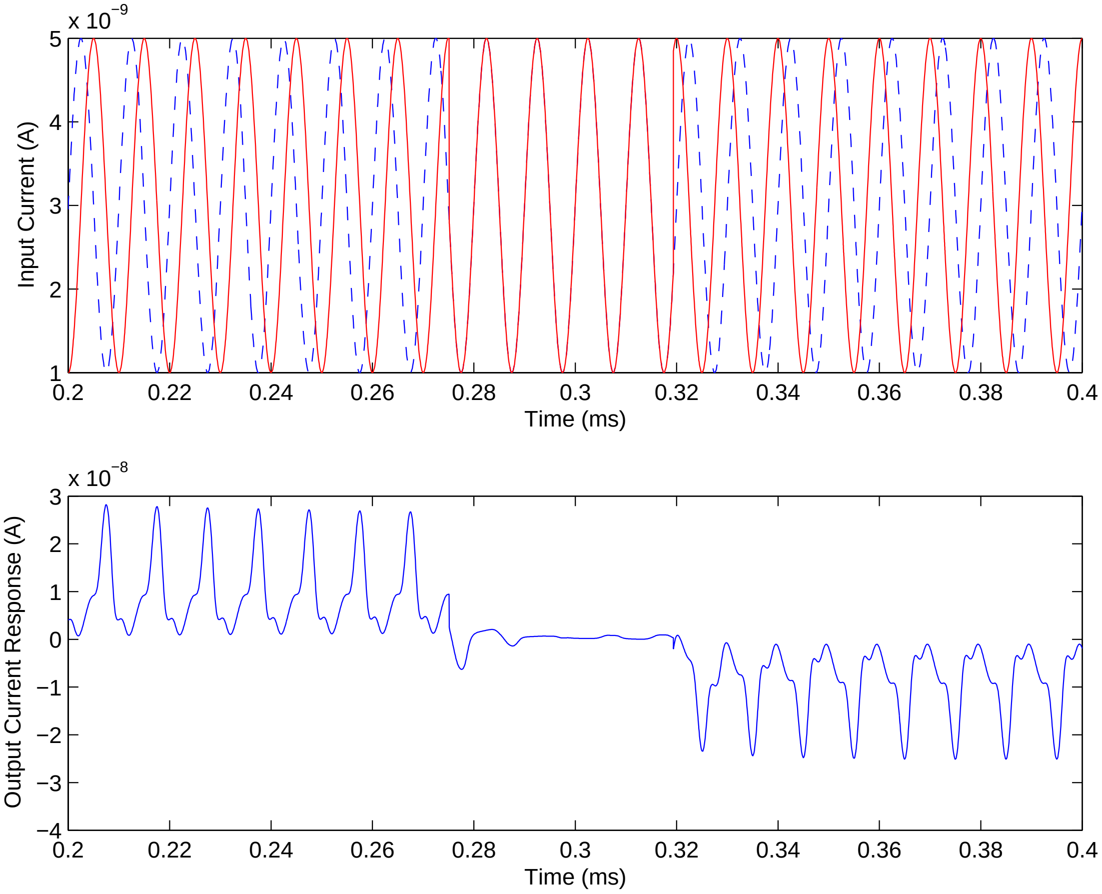
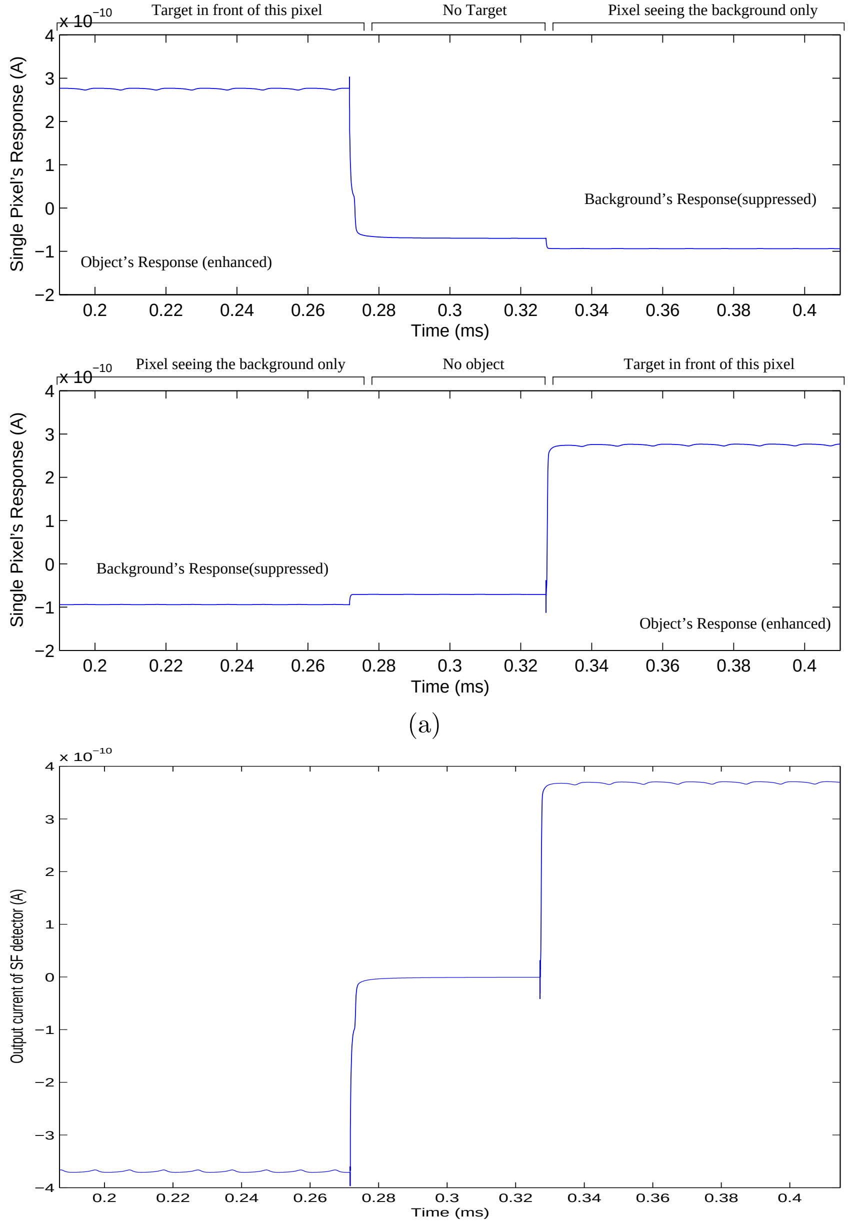


Figure 3.10. Simulation results from the HR detector. The top two traces are a sinusoidal-grating stimulus presented to two adjacent photoreceptors. The phase shift is 90° (for rightward motion) from $0.2 - 0.27$ ms, 0° (no motion) from $0.27 - 0.32$ ms, and -90° (leftward motion) from $0.32 - 0.4$ ms, and the output is correspondingly positive, zero and negative based on the detector orientation.



(a)

(b)

FIGURE 3.11. Simulation results for the small-field motion detector. (a) Output from a single pixel is shown for three conditions: (i) when there is a small target moving in front of the pixel (large response), (ii) when there is no target in the entire view-field, and (iii) when object is in the vicinity but not in front of pixel under study (suppressed response). (b) Summed output over the entire SF detector is shown. Similar stimuli are presented to all except one pixel that is either to the left side or to the right side of the view field. The SF detector response is negative for the target on the left side, zero if no pixel sees the target, and positive if the target is on the right side.

pixel and was in front of the second pixel, its response became similar to other background pixels, while the response of the second pixel, now looking at the target, was enhanced.

3.4.3 SF Detector Implementation

Two FD chip sensors acting as the left and right eye form the SF detector unit. The output corresponds to the torque response as given by Equation 2.22. Figure 3.11b shows the entire small-field response of the SF detector. This result was obtained by subtracting the simulated response of two individual FD sensors. The target is either presented to the left side sensor or the right side sensor. In a third condition, the target has the same motion direction as the background. The small-field response is summed up across the pixels from two separate sensors representing the left and the right side to get a global target tracking response. This response can be seen to produce an instantaneous yaw-turning torque response. It is negative and tends to rotate the body in the counter-clockwise direction when the pixel sees the target moving on the left side. When the background and the target are in phase (no relative motion), the target is undetectable and the response goes to zero. This is the middle portion of the response in the figure. When the target is presented on the right side, the output of the SF detector is a positive yaw-torque indicating motion in the clockwise direction.

The FD chip has three types of outputs: (i) response from the HR detector summed over the entire chip, (ii) small-field response from individual pixels, and (iii) sum of small-field responses of the entire chip. The individual small-field response from each pixel can be scanned out using the peripheral scanner circuit. By the shift operation of the register, a logic '1' is loaded in the flip-flops corresponding to the row and the column of the pixel. Thus, the small-field response of this pixel is read out. For a global small-field response from the chip, all flip-flops in the horizontal and vertical scanner circuits are loaded with logic '1'. This allows the response from all the pixels to add on a common signal line according to Kirchoff's current law. This global small-field response could be used to implement a target tracking agent.

3.5 Characterization

The FD sensor chip has been fabricated through MOSIS in the standard 1.6 μm process. Figure 3.12 shows the layout of a small-field pixel, and Figure 3.13 the complete chip layout. The resolution of this sensor is 13×6 pixels. This chip has been characterized and preliminary data has been collected at two levels: first for the HR detector, and second for the small-field normalized detector output. The characterization was done using computer simulated visual patterns to mimic background and object motion.

3.5.1 HR Detector

The HR detector on this chip produces a global current output which is the sum over all the pixels. We presented a sinusoidal grating of spatial frequency 0.12 cycles/chip-pixel at a temporal frequency of 2 Hz. The response from the FD chip is shown in Figure 3.14. The response is shown for three conditions: (i) for a rightward moving grating; (ii) for a grating moving orthogonal to the sensor orientation; and (iii) for a leftward moving grating. The mean responses for all three stimulus conditions are indicated on the graph. The mean separation between the left and right moving grating is around 500 mV and clearly indicates the direction of motion. By increasing the current through the Gilbert multiplier, this output can be further expanded.

3.5.2 Small-field Normalization Detector

The second set of characterizations were done for the small-field (SF) detector units. The visual stimulus was generated by a computer software and displayed on an LCD screen (see Figure 3.15).

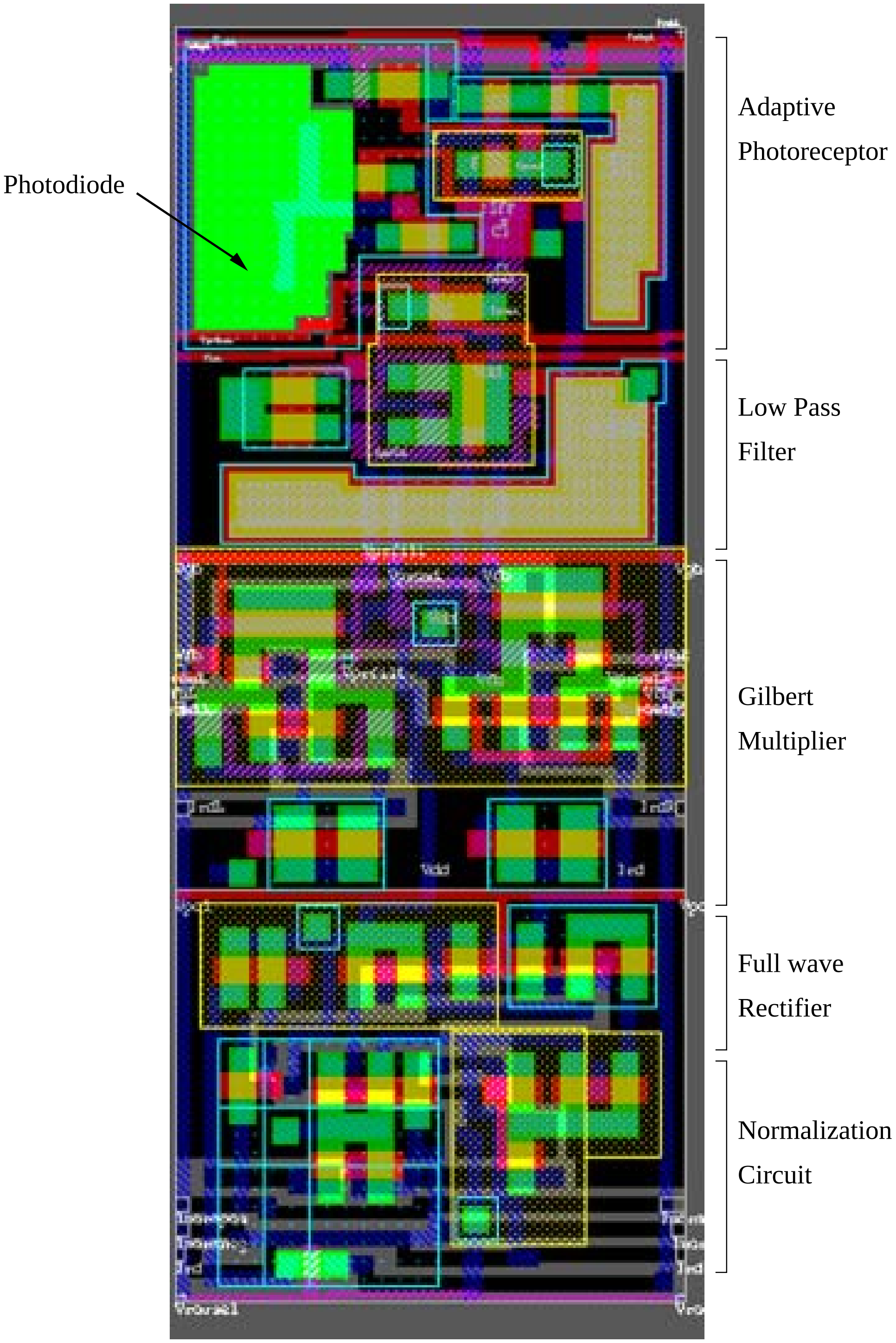
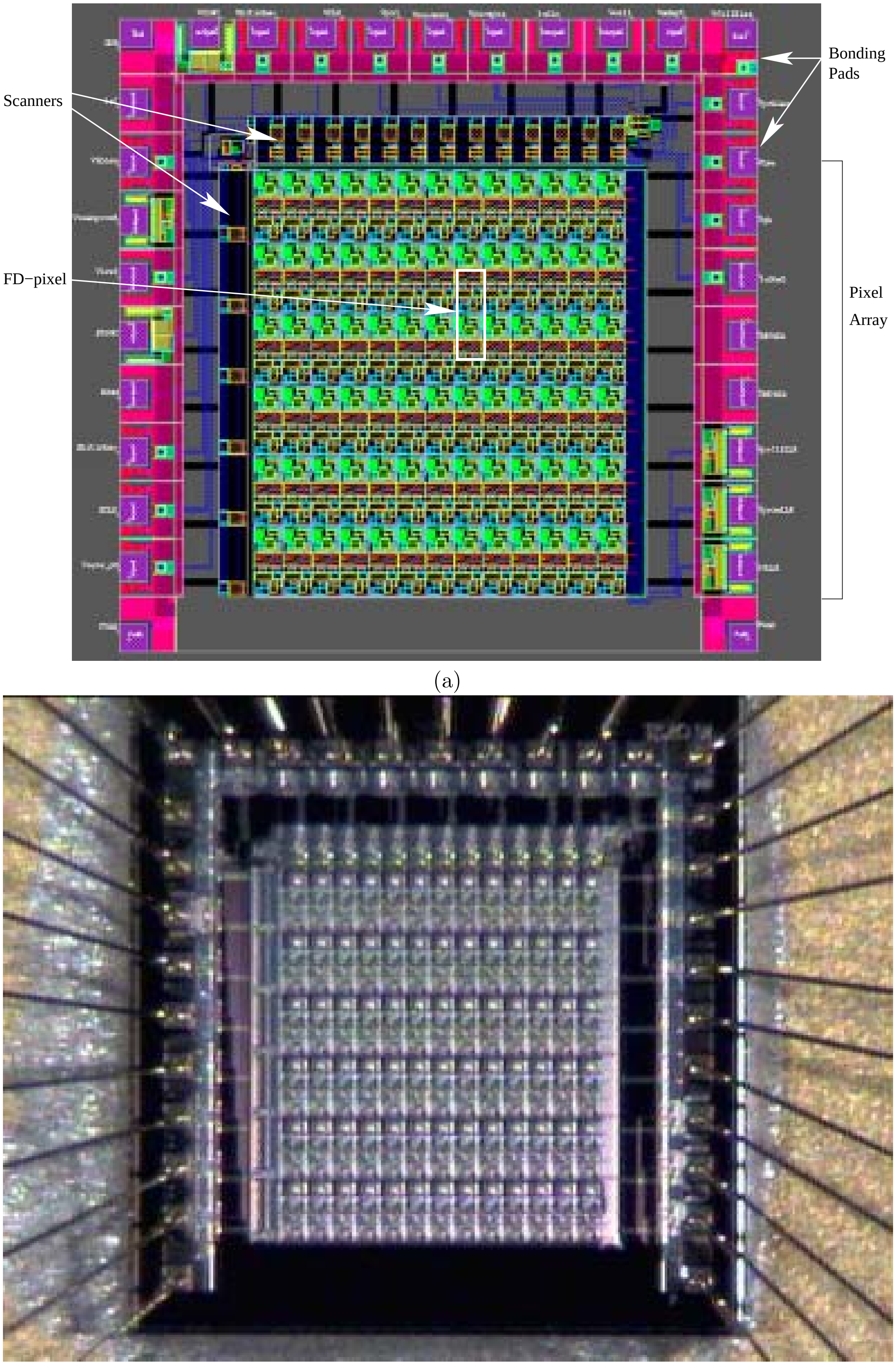


Figure 3.12. Layout of a pixel of the FD chip showing details of the various processing stages.



(a)

(b)

Figure 3.13. FD sensor chip. (a) Layout as drawn in the L-Edit tool with an array size of 13 × 6. The chip was fabricated in a 1.6 μm process with a die size of 2.2 mm × 2.2 mm. (b) Photomicrograph of the fabricated chip.

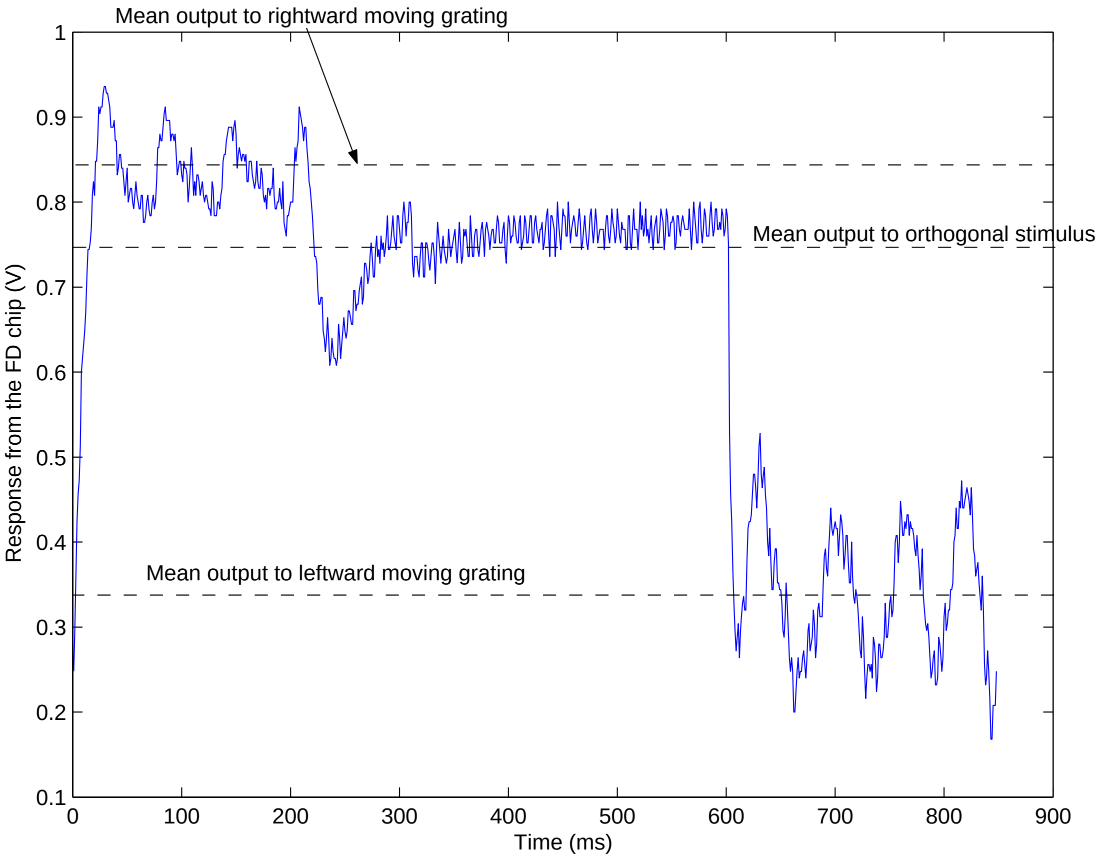


Figure 3.14. Response from the HR detector on the FD chip. A sinusoidal moving grating was presented as the visual input with a spatial frequency of 0.12 cycles/chip-pixel and a temporal frequency of 2 Hz. The response is a sum over 72 processing units. The mean response to the rightward moving stimulus is above the mean response from an orthogonal stimulus. The mean response from a leftward moving grating is below the orthogonal stimulus response. Thus the sensor chip indicates the direction of motion in the mean response.

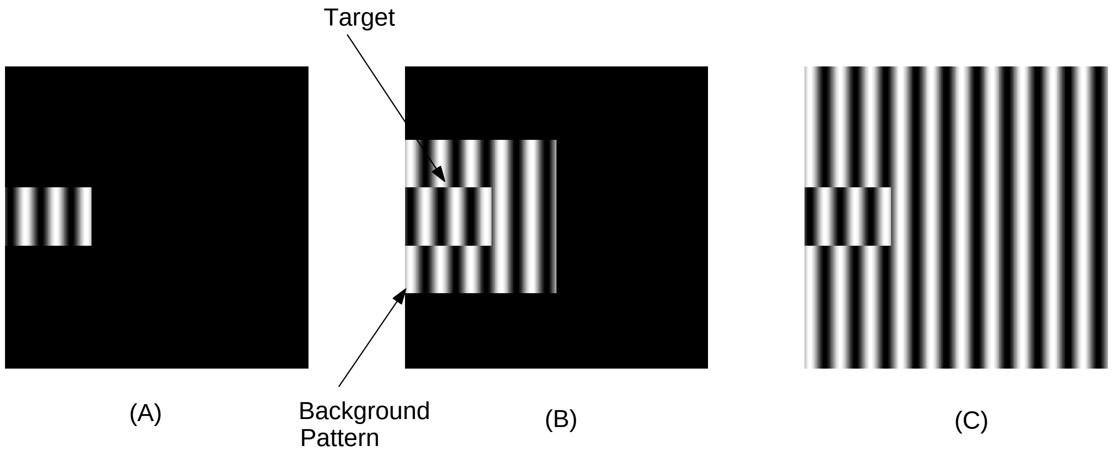


Figure 3.15. Stimuli for the small-field unit with three different background extents. The visual field is shown for three conditions. The target occupies only a two chip-pixel square area of the visual field while the background pattern's size varies. (A) The target and the background occupy only a two chip-pixel square area of the visual field, and they exactly overlap in this condition. Therefore, the background is not visible. (B) The background pattern nearly occupies half of the visual field with the target on the left edge. (C) The background pattern covers the entire visual field in this condition. The extent of the target remains the same as in the previous two conditions. The phase of the target with respect to the background is varied for different experiments.

The visual stimulus for this experiment consisted of a sinusoidal grating as the background and a smaller grating on top of it serving as the target. The discontinuity in the optical flow-field causes the small-field output in the region of the target to be different from the overall response from the chip. The normalization circuit then enhances this aberration such that the response from the pixels activated by the object dominates the output. The response from a single pixel seeing the target was read out by using the serial scanner circuitry. Figure 3.16 shows three traces recorded from the FD chip. It shows the response of an individual small-field unit under three different stimulus conditions as shown in Figure 3.15. The stimulus is such that the background either covers only the pixel under study, or half the extent of the chip, or the entire chip, while the target is always in the same place. First the target and the background move together in the rightward direction (stimulus condition 1 in Figure 3.16). Then, the background keeps moving to the right while the target moves towards the left (stimulus condition 2 in Figure 3.16). The target now keeps moving towards the left while the background switches direction such that both move leftwards (stimulus condition 3 in Figure 3.16). It must be noted that the visual stimulus for the SF unit being stimulated by the target has not changed between stimulus condition 2 and 3, still the response shown in graph C2 is larger as compared to C3 in Figure 3.16. The background extent corresponds to three cases discussed in Figure 3.15A-C. The graphs A1, B1 and C1 show the response of the SF unit for the rightward moving grating and the target. The graphs A2, B2, and C2 show the response when the target is out of phase by 180° with respect to the background still moving rightwards. We see that the normalization due to the background does not affect the SF unit output in this stimulus condition. The graphs A3, B3, and C3 show the in-phase movement of the target and the background towards the left. In C3, due to the normalization of the response because of the large background pattern, the output is smaller as compared to A3. For B3, the background pattern's extent is about half of the visual field, so the normalized output is larger than C3 but still smaller than A3.

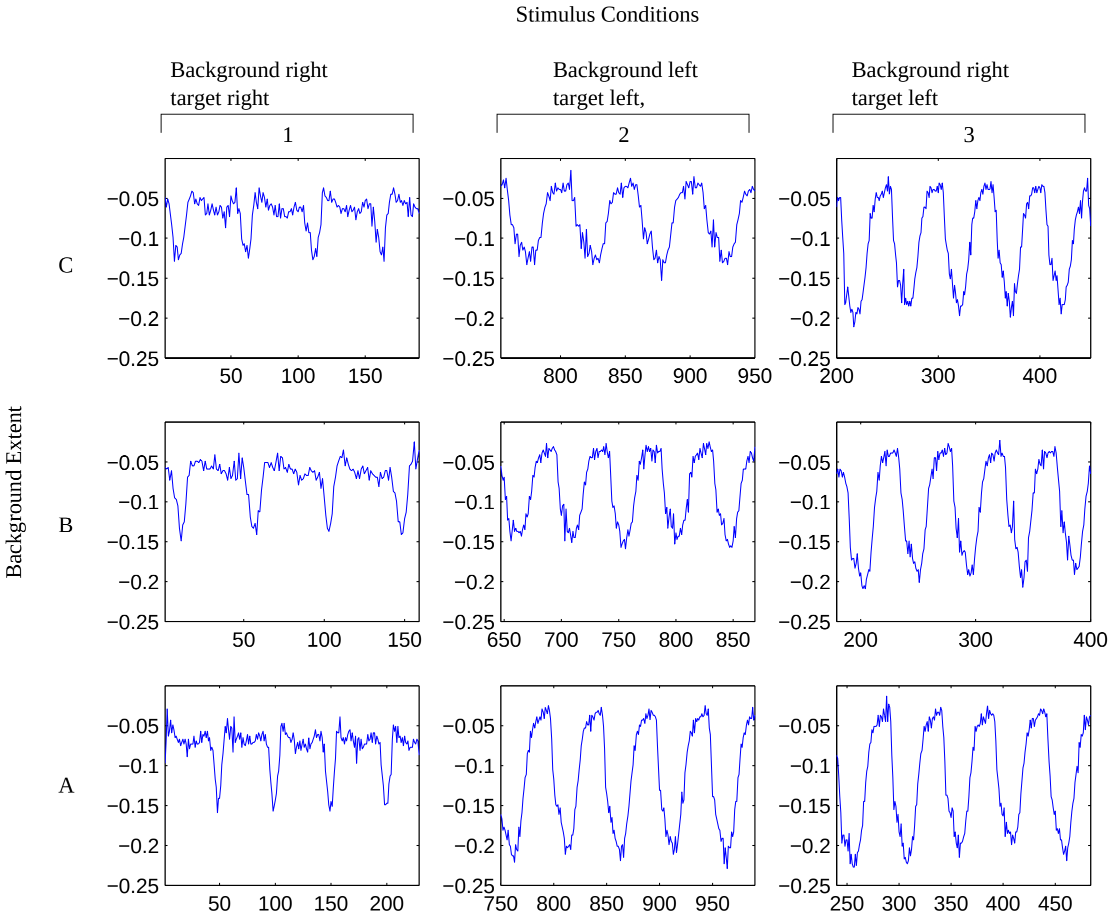


FIGURE 3.16. Response from an SF unit on the FD chip. The response from the FD chip has been recorded for three stimulus conditions and for three background extents. The background and the target move together towards the right first, then the target switches direction and moves towards the left while the background is still moving towards the right, and finally both the background and the target move together towards the left (stimulus conditions 1, 2 and 3 respectively). The extent of the background pattern was varied as shown in Figure 3.15 and corresponds to the cases A, B and C respectively. The spatial frequency of the stimulus was set at 0.12 cycles/chip-pixel at a temporal frequency of 2 Hz. The outputs for the rightward moving target and background are shown in graphs A1, B1, and C1. The response in the cases A2, B2 and C2 remain almost the same when target is present immediately in front of the pixel being recorded, and all the remaining pixels are looking at the background. However, the response to the leftward moving grating was observed to be suppressed due to the normalization over the background extent. It is clearly seen that the normalization over a lower number of pixels (smaller background extent) has a larger response as compared to the case where the background covers the entire visual field (C3). When the background and the target exactly overlap, we see that due to no normalization, the small-field output is equal to the leftward moving background.

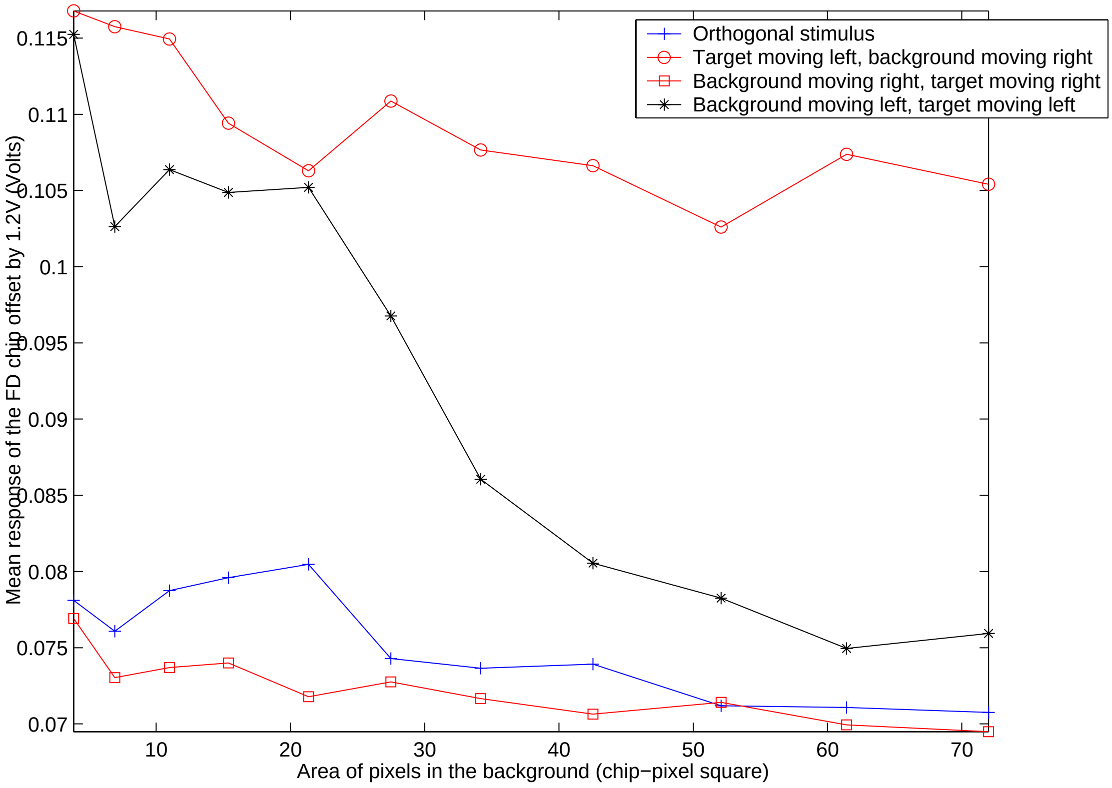


Figure 3.17. Mean response of the small-field unit on the FD chip versus area of pixels stimulated by background. The normalization of the leftward moving background decreases the response as the number of pixels being stimulated by the background increases. The small-field output for the target remains almost the same.

In another experiment, it was also seen that the output where the target and the background move in opposite directions does not change much in all three conditions. The changes in the mean response of the small-field output under various intermediate stimulus conditions were plotted to show the affect of normalization on the background motion. As seen in Figure 3.17, the mean response for the leftward moving background reduces as the number of pixels being stimulated by the background increases. The SF output for the target moving opposite to the background remains nearly a constant. This preliminary data clearly suggests that the response of the small-field detector is stronger for small targets as compared to larger background motion. Thus, it can indicate the presence of a target in the view field and therefore may be used as a target tracking sensor unit.

3.6 Summary

These preliminary characterization results are presently being studied for doing a complete characterization of this chip. The experiments have revealed that the output is very sensitive to the full wave rectifier bias and we are presently working on suggesting an alternate robust circuit implementation.

The compact nature of this design makes a very attractive monolithic tracking unit with some external control circuitry. However, the resolution of this circuit is insufficient for practical real-time applications. A need for a high-density front-end chip with the processing distributed to sender and receiver chips is therefore felt. The next chapter explains the necessity of modular design and describes a sender chip designed and fabricated to improve upon the monolithic tracking sensor.

CHAPTER 4

MODULAR ARCHITECTURES AND SPIKING-NEURON SENSOR

In the previous chapter we presented a monolithic implementation of the FD chip that may be used as a target tracking sensor. In any autonomous mobile robot, there might be several visuo-motor processing units subserving the agent's movement in the arena such as a target tracking system, an obstacle avoidance system, and a course control system. Having separate visual information gathering units for all the subsystems is not an efficient way to design a complex system like this. A solution to this problem is a common contrast sensitive transducer chip as the front-end. It shares visual information among various subsystems making the entire system more compact. This chapter describes such modular architectures for biomimetic systems, and discusses the implementation of a spiking-neuron sender chip.

4.1 Modular VLSI Neuromorphic Architectures

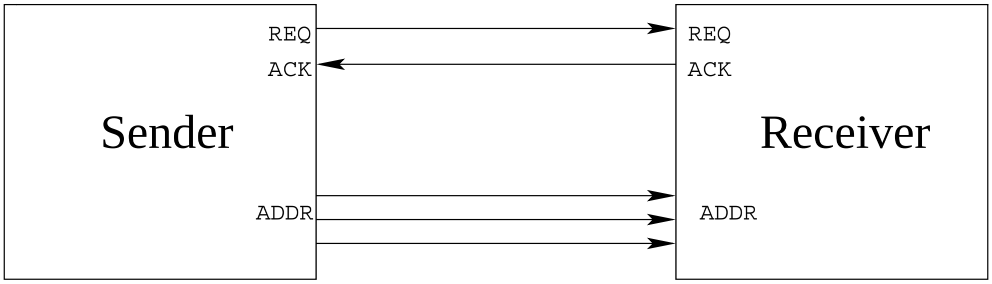
Monolithic neuromorphic architectures perform computations in the focal plane and therefore are well suited for parallel computations in a neuromorphic system. The only problem, however, is that when the computations begin to increase so does the size of each pixel. This leads to lesser resolution and lower fill factor (the percentage of pixel area dedicated to phototransduction). Thus, the amount of computation that can be done at the focal plane level is limited. The solution to this can be taken from the visual system of insects, where the computation is neither done totally in the focal plane nor in an end processor. The answer is to distribute the processing along the pathways through which information is passing. This leads to modular image processing architectures. The modular solution also means that some information, if not all the raw data, is required to be transferred to another stage where the remaining computation is performed. Hence, a time efficient, low power and high-speed communication protocol is required between these intermediary processing units in order not to lose the advantage over a monolithic architecture. In the nervous system, information is encoded as action potentials that are transient changes in the voltage across the membranes of nerve cells. A chain of action potentials emitted by a single neuron is called a spike train - a sequence of stereotyped events which occur at regular or irregular intervals. The Address-Event Representation (AER) protocol mimics this behavior, but in contrast to parallel axons in the nervous system, it has a serial bus relaying the information. The next subsection discusses this asynchronous digital communication protocol for interchip communication.

4.1.1 The Address-Event Representation (AER)

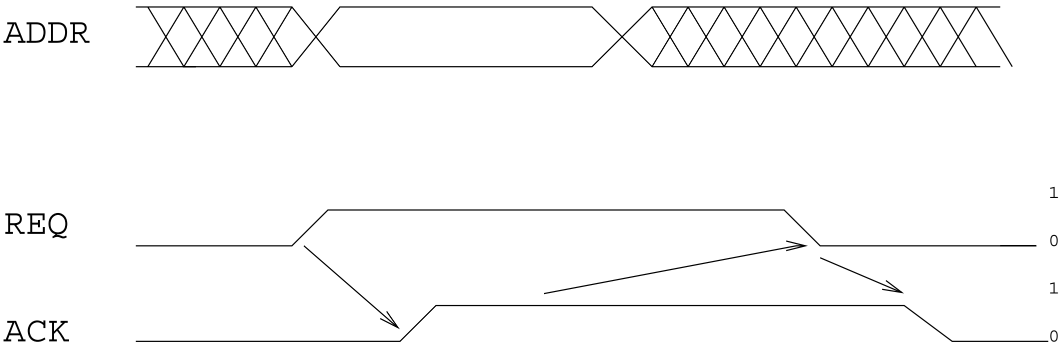
Frequency encoding of the position and time of an event over an asynchronous serial bus for neuromorphic applications was first proposed by Mahowald (1992). The most basic form (see Figure 4.1) has two digital control lines in the form of request (REQ) and acknowledge (ACK), and several address lines to associate this request to its point of origin. There are two characteristic parameters defining the performance of this protocol:

1. **Transmission latency** is the time interval between the onset of the stimulus and event generation at the REQ line.

2. **Temporal dispersion** takes place because of the serialization of the bus due to sharing of the interchip REQ signal, and mismatch between the processing-units (pixels) transmitting this event train.



a) Communication Model



b) Handshaking Protocol

FIGURE 4.1. AER Protocol. (a) The sender and receiver communicate over request (REQ) and acknowledge (ACK) lines. The position of the activity is communicated by the address (ADDR) lines. (b) A four-phase asynchronous handshake initiated by a request with valid address is shown. The receiver latches the address and raises the ACK, which leads to the falling of REQ and subsequently ACK lines.

For an ideal channel, the transmission latency and the temporal dispersion should be zero. In a real system, these parameters have finite values that define the system performance. In a biological system, the retina sends spatiotemporal patterns of light over the optic nerve as spike trains and dendritic integration converts them back into a continuous signal (Boahen, 2000). The activity of a neural array can be represented as an ordered list:

$$\begin{aligned}\xi &= \{(x_0;t_0),(x_1;t_1),\ldots(x_i;t_i),\ldots\};\\ &\quad t_0 < t_1 < \cdots t_i < \cdots\end{aligned}$$

where x_i represents the location and t_i the time of occurrence. Since the channel is real time, the time information is redundant. Hence, the above can be re-written as:

$$\begin{aligned}\xi &= \{(x_0,x_1,\ldots x_i,\ldots\};\\ &\quad t_0 < t_1 < \cdots t_i < \cdots\end{aligned}$$

This real time code is called the address-event representation. There are a number of other issues related to the AER protocol. Since there is a single interchip request line, the channel can either be pre-allocated or dynamically appropriated to the active pixel. The occurrence of simultaneous activities on two pixels can lead to a contention for the request bus. This leads to collision of events and a corrupt signal on the request line, if it is not dealt with in the design. The use of an arbiter to queue up the events or disregard them in the case of collision is a solution. However, queuing of events leads to temporal dispersion, reduction in the channel capacity, and lengthening of the cycle time. Boahen (2000) has done a detailed analysis of this stochastic process. The AER protocol has been used over the years for asynchronous communication between neuromorphic analog chips. The next section discusses various implementation of this protocol in the design of modular biomimetic systems.

4.1.2 Asynchronous Digital VLSI for Neuromorphic Applications

The AER inter-chip communications strategy was first used to transmit visual signals out of a silicon retina by Mahowald and Mead (1991). However, these attempts were dogged by cross-talk and timing problems. The progress made in the field of asynchronous digital VLSI systems in recent years has strengthened and formalized the AER protocol (Martin *et al.*, 1989). Several variants and specializations of the scheme have emerged in the last few years. Boahen (1996a) interfaced two silicon retinas to three receiver chips to implement binocular disparity-selective elements. Venier *et al.* (1997) used an asynchronous interface to a silicon retina to implement orientation-selective receptive fields. Deiss *et al.* (1998) implemented a silicon model of primate visual cortex using interchip communication, and DeWeerth *et al.* (1997) implemented a model of leech intersegmental coordination. Grossberg *et al.* (1997) demonstrated the use of EPROMs for linear or nonlinear address re-mapping in interchip communication. Kumar *et al.* (1998) provided an auditory front-end chip with an asynchronous interface for further off chip processing. Boahen (1996b) published a multi-chip vision processor that computes motion by emulating a model of primate motion computation. Kalayjian and Andreou (1997) created a photosensitive sender chip, and Higgins and Koch (2000) designed a sender-receiver system that computes 2-D optical flow vector field from edge information. The arbitration scheme used by Boahen (1999) takes care of collisions effectively. However, it increases the latency due to queuing, and the temporal dispersion increases. Landolt, Mitros, and Koch (2001) have designed a scheme that discards colliding events and allows only collision-free events to be transmitted off chip. It is this scheme that we will employ in our sender chip communication protocol.

The Landolt, Mitros, and Koch AER scheme: This AER protocol uses a four-phase asynchronous handshake between the sender and receiver guaranteeing reliable communication over the two control lines, REQ and ACK. The $\log_2 N$ number of address lines, for an N-pixel array, is used to communicate the spatial position of the event to the receiver. The average latency of this scheme as published is about $45ns$ (Landolt *et al.*, 2001). A diagram of the information flow of this scheme is presented in Figure 4.2.

Let us consider the case in which only a single pixel is activated by the stimulus (no collision). This launches digital spikes on the two asynchronous lines, EVy and EVx, one shared along that column of pixels and the other along that row. Every row and column line has an SR-latch controlled by the input enable signals IEx and IEy, respectively. When the system is ready to accept a new event, the IE signals are active and allow the event to be latched. When the event has been latched, the CATCH signal goes high. This acts as an acknowledge signal for the event transmitting pixel, and this pixel resets. The ANYx and ANYy signals indicate that an event has been caught, which in turn causes the input enables (IEx, IEy) to become inactive, thereby preventing any new events from being latched. A $\log_2 N$ encoder then encodes the latched address. The outputs of the encoders directly connect to the external ADDR bus and are also input to the validity checking circuitry.

The addresses are encoded in a way that allows verification that no collision has occurred. A dual rail bit representation provides a means to verify the on chip collision occurrence. In this representation, one line is raised high to represent logic '1' and another line is raised high for logic '0' (see Figure 4.3). The lines in their idle state are passively pulled low by pull-down transistors, and are actively pulled high when an event on the corresponding row or column occurs. Thus, both lines low indicates an idle state, both lines high indicates an invalid state, and one active line at a time indicates valid address. The invalid state results from one pixel attempting to set an address bit to zero while a different pixel tries to set it to one. From these ADDR and ZEROS lines, only one of each pair of lines is taken off chip (the ADDR line), while the other is used to perform error checking on the chip (the ZEROS line).

In case of no collision, a valid address is detected and VALx and VALy go high. This in turn results in REQ being raised to signal the off chip circuitry that an address awaits reading. When the address has been read, ACK acknowledges and the reset procedure begins. First, CLR is raised signaling that the event latches should be cleared. When the latches are cleared, the addresses in the address encoder are reset, VALx and VALy fall, and finally RESx and RESy go high to indicate that the address lines have reset. Then, CLR falls and in turn allows IEx and IEy to get activated, ready to latch the next event.

It is also possible that two events might occur very closely in time. The ideal condition would be that input enables inactivate immediately when a single event is caught in order to prohibit simultaneous latching of two events. However, because disabling the latches takes about $1.5ns$ (Landolt *et al.*, 2001), two events may be simultaneously latched. As described above, both the lines representing the bits in which the two requesting pixels differ are raised. Thus an invalid address gets detected. CLR is raised in this case and the reset procedure proceeds as above, but REQ never goes high. This avoids wastage of time and error-checking resources necessitated by sending invalid addresses off chip.

Several events may occur while an event is being processed and IEx and IEy are inactive. These active pixels continue to attempt to broadcast their event to the latches until the latches acknowledge their data. On the successful completion of the request from a pixel, the latches are cleared out and the input enables (IEx and IEy) are reactivated. Upon activation of the input enables, latches immediately receive and latch all events which occurred while the first event was being processed. This cycle continues as described above. In case of a single event, a valid request and address are transmitted. If multiple events occur, collision results in an invalid address on the lines and the system resets without off-chip communication.

This version of the AER protocol has been chosen for implementation because of its low latency and temporal dispersion. The following section presents a spiking-neuron sender chip employing the Landolt-Mitros-Koch AER scheme.

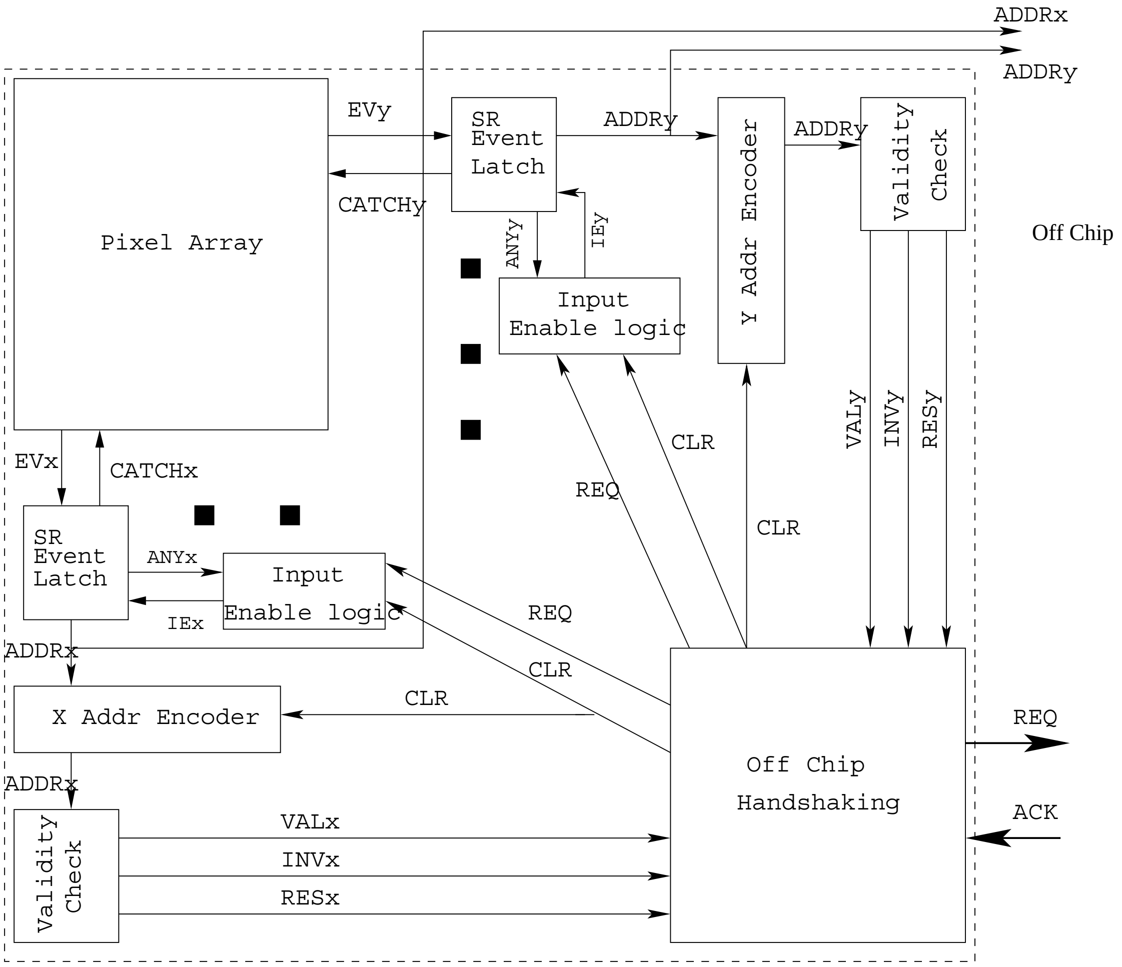


FIGURE 4.2. Schematic showing the Landolt-Mitros-Koch AER Signal-flow. EV is the event generated by a pixel; CATCH is the acknowledge signal from the peripheral circuitry to the pixel; ADDR is the address bus; ANY indicates that the address is latched; IE is the enable signal for the latch; VAL indicates a valid address; INV an invalid address; RES indicates that the address line has been reset; CLR initiates the clearing of the encoders; and REQ and ACK are used for handshaking between the sender and the receiver in the event of valid address. The dashed box represents the on chip circuits. The SR event latch circuit is repeated according to the number of rows and columns in the chip. Off chip REQ, ACK and ADDR buses are also shown. Reproduced without permission from Landolt, Mitros, and Koch 2001.

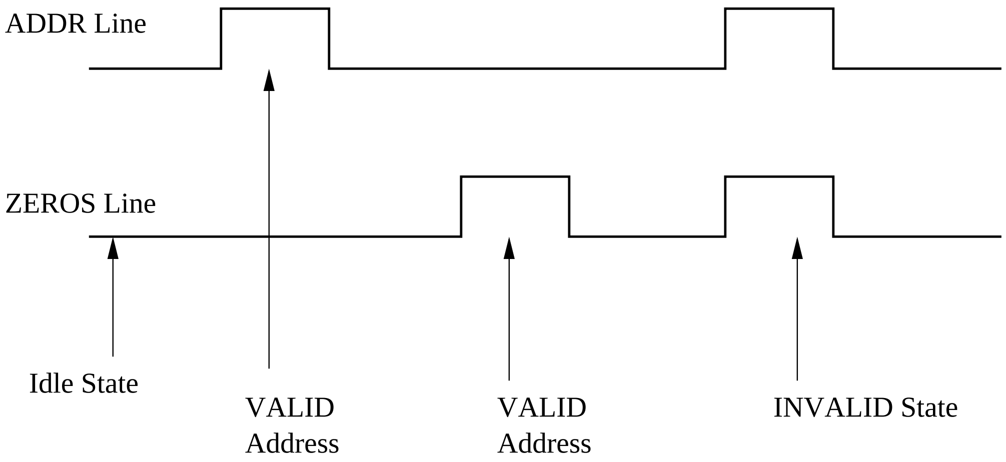


Figure 4.3. Address (ADDR) and Zeros (ZEROS) line based arbitration. Zero on both lines indicates an idle state and the pixels can transmit requests. A high on the ADDR line indicates a logic '1', and a high on the ZEROS line indicates a logic '0' code for the address. Simultaneous activity on both the lines is an invalid state, and such a request is terminated.

4.2 Circuit Design of the Spiking-Neuron Sensor

The VLSI implementation of a sender chip acting as the visual front end of a target tracking system or other similar visual system is presented here. This chip consists of pixels that convert contrast information into electrical signals and relay this information as frequency encoded digital events over an asynchronous serial bus. The design of a phototransducing pixel and surrounding AER arbitration circuit is presented in this section. A block diagram representing the various sub-circuit stages for this implementation is shown in Figure 4.4. The pixel resembles the functioning of a spiking neuron and sends a train of spikes off the chip to communicate the edge locations. There are several types of spiking-neuron models ranging from detailed biophysical ones to the integrate-and-fire type (for an excellent review, see Gerstner, 1998), which form the basis of spiking-neuron networks (Maass, 2002). A class of more detailed models, known as conductance-based ones, have their origin in the classic work by Hodgkin and Huxley in 1952. Hodgkin and Huxley summarized their experimental studies of the giant axon of a squid in four differential equations. The main equation describes the conservation of electric charge on a piece of membrane capacitance of a neuron under the influence of a current and a voltage (for a detailed mathematical analysis see Cronin, 1987). The Hodgkin and Huxley equations can be regarded as describing the variations of membrane potential and ion conductances that occur naturally at a fixed point on the axon. A different approach takes the neuron as a leaky integrator which is reset if firing occurs (Stein, 1967). Our sensor adopts this scheme of integrate-and-fire, and generates spikes corresponding to the activity of the photoreceptor. The generated spikes (called events) are then trapped by latches and using the Landolt-Mitros-Koch AER scheme; these events can then be transferred to a receiver unit for further processing. The circuit details of all the stages are presented in the next few subsections.

4.2.1 Adaptive Photoreceptor

The local light intensity is transduced using the adaptive photoreceptor (Liu, 1999). This circuit is the same as has been described in Section 3.3.1. This phototransduction results in two output signals: $Vprout$, the instantaneous response of the detector, and Vfb, which represents the long-term

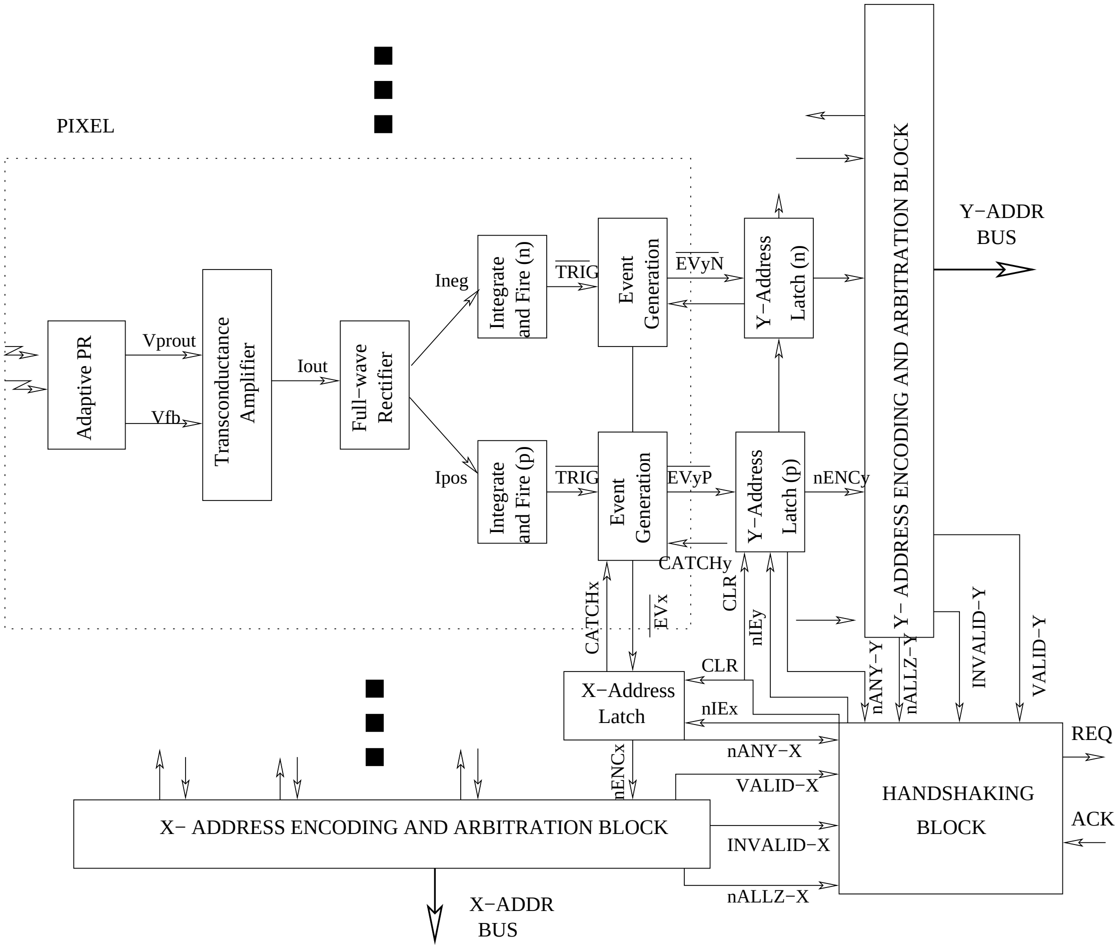


FIGURE 4.4. Schematic representation of circuit stages in the spiking-neuron chip. The block marked pixel is a single processing unit, and this chip is formed by a 5×4 array of these pixels (indicated by dots). Each pixel has an adaptive photoreceptor that transduces visual information into electrical signals. These signals are conditioned and amplified by a transconductance amplifier stage. This contrast information is rectified by a full-wave rectifier circuit block. The rectified responses (I_{neg} and I_{pos}) are used to generate trigger pulses ($\overline{TRIG}$) in two integrate-and-fire blocks of differing polarity. Two event generation circuit blocks then communicate with peripheral AER circuits to communicate this activity. The event is latched if the address line is idle. The AER circuitry is similar along X- and Y-axes. These blocks communicate with a handshaking block that generates the request signal.

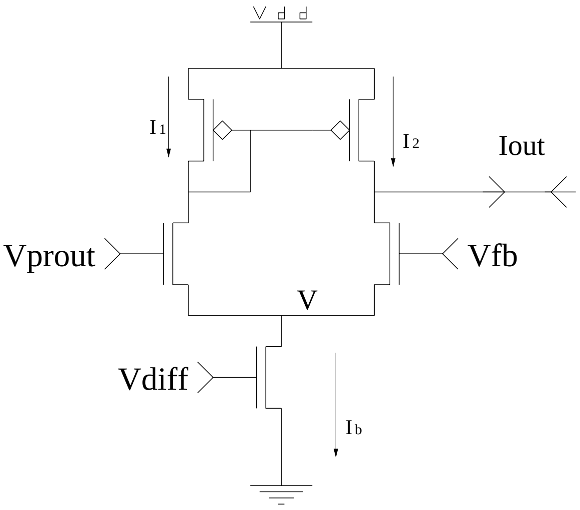


FIGURE 4.5. Operational Transconductance Amplifier (OTA). The amplified current output is bidirectional based on the difference of the two inputs.

mean response of the detector (Delbrück, 1993a). The output of this detector stage has a band-pass characteristic as it enhances the sharp contrast variation information. This is highly desired as this chip is to be used for visuo-motor applications to transfer contrast information across to a receiver chip.

4.2.2 Transconductance Amplifier

The response from the photoreceptor *Vprout* has contrast information encoded in it, which rides over the average background illumination response *Vfb*. To enhance the contrast information, an operational transconductance amplifier (OTA) stage is used. This amplifier generates a current, which is a function of the difference between the two input voltages *Vprout* and *Vfb* (see Figure 4.5). The bottom transistor is used as a current source. The bias current I_b is controlled by the bias voltage *Vdiff*.

The current through the two differential arms, I_1 and I_2, bears an exponential relationship to the differential voltage inputs. All the transistors in the OTA work in the subthreshold region and the relation is specified as:

$$\begin{aligned} I_1 &= I_0 e^{\kappa Vprout - V} \\ I_2 &= I_0 e^{\kappa Vfb - V} \end{aligned}$$

and these two currents are made up of the bias current I_b.

$$I_b = I_1 + I_2 = I_0 e^{-V} \left(e^{\kappa Vprout} + e^{\kappa Vfb} \right)$$

The current mirror at the top of the OTA subtracts the current in the positive arm of the amplifier from the current in the negative arm to give an output current I_{out} as a function of the difference of the two inputs.

$$\begin{aligned} I_{out} &= I_1 - I_2 \\ &= I_b \tanh \frac{\kappa(Vprout - Vfb)}{2} \end{aligned}$$

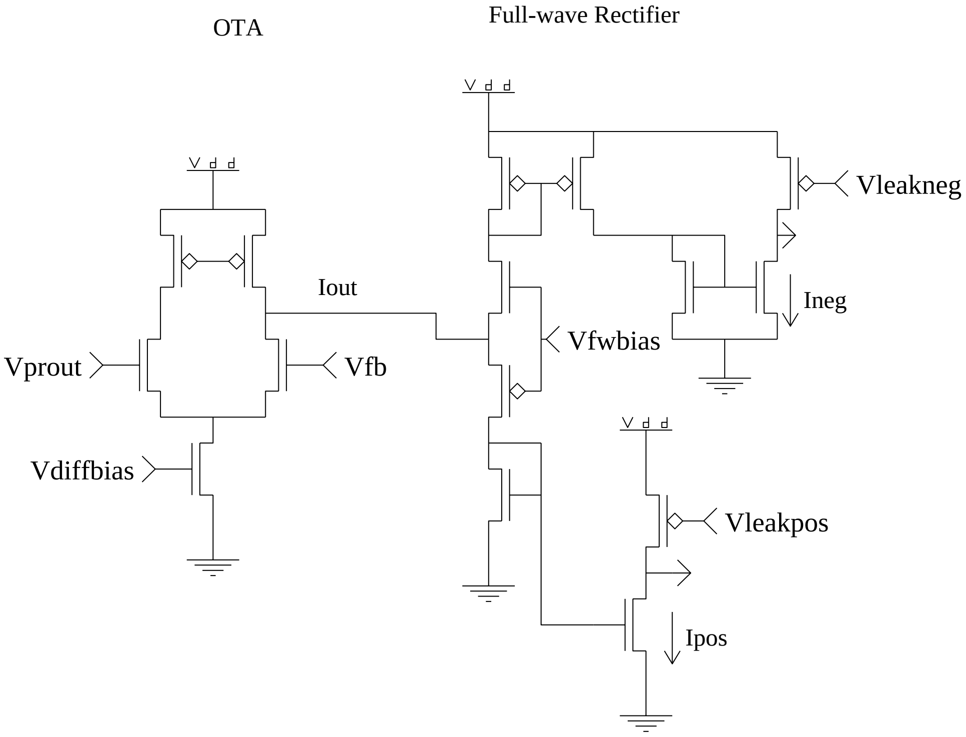


FIGURE 4.6. Analog part of the neuron pixel showing OTA stage and the separation of negative and positive currents by using a full-wave rectifier circuit.

This current is bidirectional depending on the relative values of the differential inputs. When the long-term mean response *Vfb* is greater than the instantaneous output *Vprout*, the current is negative, and vice-versa. This bidirectional current thus specifies changes in contrast. For details about the analysis of this circuit, see Mead (1989).

4.2.3 Full Wave Rectifier

A full wave rectifier circuit is used to split this bidirectional current into two unidirectional currents. This is a very common circuit comprised of two current mirror stages, one P-type and another N-type as shown in Figure 4.6. A buffer with an NFET and a PFET connected to a common external bias $Vfwbias$ provides a path to the current in both directions. The P-type current mirror provides the negative part (I_{neg}), while the N-type mirror provides the positive part (I_{pos}) of the response. The rectified current may be nonzero even when there is no activity at the photoreceptor due to leakage. This leakage current is subtracted from the rectified currents by an external bias (*Vleakneg* and *Vleakpos*). This response is fed to the digital stage that implements the integrate-and-fire circuit. This circuit has been discussed earlier in Section 3.3.4.

4.2.4 Integrate-and-Fire (IF) Circuit

The simplified description of a neuron's function may be summarized in two steps: the integration of synaptic inputs leading to membrane depolarization, and the initiation of an action potential that may propagate to the output terminals of the neuron. Such neuron models are called "integrate-

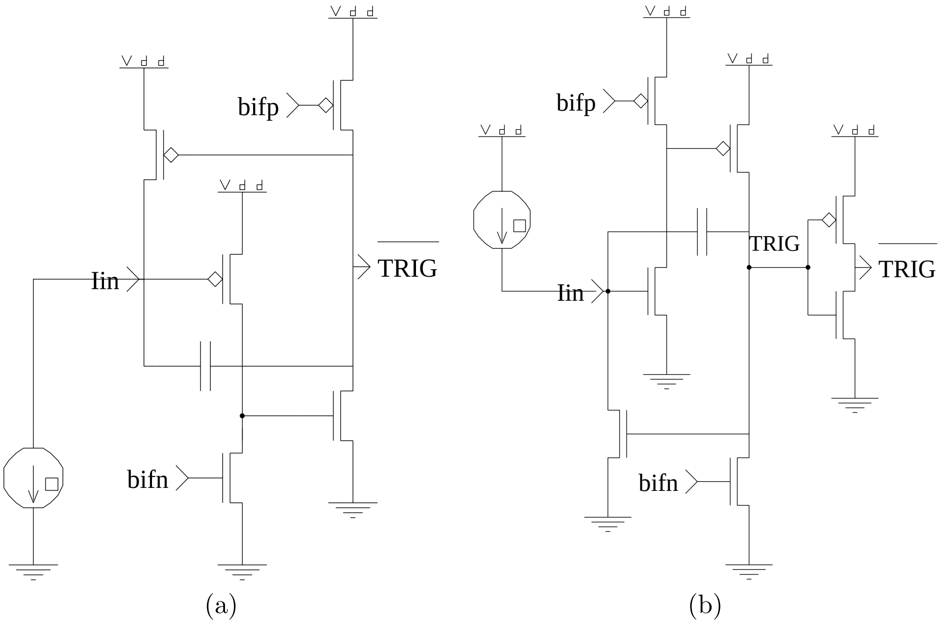


FIGURE 4.7. Integrate-and-Fire circuits (from Landolt *et al.*, 2001). (a) P-type circuit, the input of which is a current sink at node I_{in}. (b) N-type circuit, the input of which is a current source at node I_{in}.

and-fire" models. The integrate-and-fire (IF) circuit implemented here mimics the leaky-IF model of neurons (Landolt *et al.*, 2001). Since input currents may go in either direction in general, there are two different IF circuits; one for current being sunk by the circuit (N-type) and the other which requires a sink to dump the current (P-type) as illustrated in Figure 4.7. Based on the polarity of incoming current, a capacitor in the IF circuit provides a feedback path from the output and toggles it after a certain amount of charge accumulation. A bias voltage *bifn* controls the threshold voltage to make the capacitor spike. Another bias *bifp* controls the reset time of the output, and thereby determines the maximum spike generation frequency of the pixel. The N- and P-type IF circuits produce TRIG and $\overline{TRIG}$ signals respectively, which are captured by the latches in the arbitration block. The TRIG signal is passed through an inverting stage to utilize a common latch for both the positive and negative pulse trains. This integrate-and-fire scheme thus produces a train of spikes with a frequency proportional to the current input to the IF circuit (Mitros, 2002).

4.2.5 Event Generation

The pulse train generated by the IF circuit is fed to an event generating circuit (Landolt *et al.*, 2001). There are two such blocks in each pixel, one to capture the event-train of the positive current, and the other for negative current. Figure 4.8 shows the transistor level implementation of this stage. Events are generated based on the availability of external address lines. Two signal lines, CATCHx and CATCHy, convey this information from an external latch to the event generation circuit. These lines are low when the latch is ready to accept an event, and is raised high when the address lines are not available for the pixel.

This asynchronous circuit transmits an event based on the same threshold as the IF circuit. This is controlled by the bias voltage $bifn$. When the external latch is busy, the CATCHx and CATCHy signals are high. This makes the NFETs connected to these signals to be on. This keeps the circuit

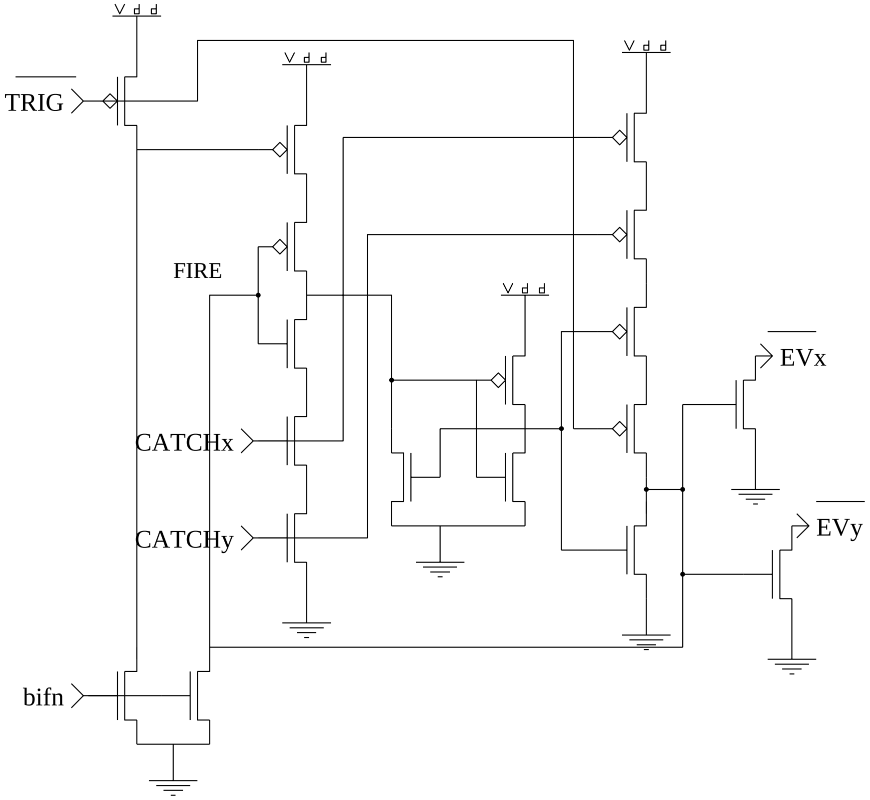


Figure 4.8. Event generation circuit (from Landolt *et al.*, 2001). The spikes from the IF circuit are processed to generate an event that is latched on by the external AER circuit, if there is no collision. A feedback signal (CATCH) is provided by the external AER circuitry along both axes, to indicate that the event has been latched.

in a latent state. When the external latch becomes available, the CATCH signals go low switching off the connected NFETs. The current from this part of the circuit stage is diverted to a latch that relays the events to the external circuitry. For the events to be propagated, the current has to exceed the pull-down force being applied by the $bifn$ bias. This node also has a feedback connection to the FIRE node. Thus, the circuit resets immediately after sending the event.

These events are fed to an external latching stage. A transistor level circuit diagram of all the stages within a pixel is shown in Figure 4.9.

4.2.6 Event Latches and Addressing

In this and subsequent subsections, we present the peripheral AER circuitry as designed by Landolt, Mitros, and Koch (2001). The layout for this arbitration circuit was generated by computer software written in the silicon compiler of the L-Edit tool (Tanner-EDA, 2003). The events generated from each pixel are available to external latches, each event line corresponding to a latch. In effect one row corresponds to two latches and a column to one latch. This is because the pixels are arranged laterally, due to which they have a positive and negative response along the rows, but not along the columns. The latches have an input enable logic based on the AER scheme. As seen in Figure 4.10, the nIE signal is generated when no event line is active, no pixel is sending a request, and

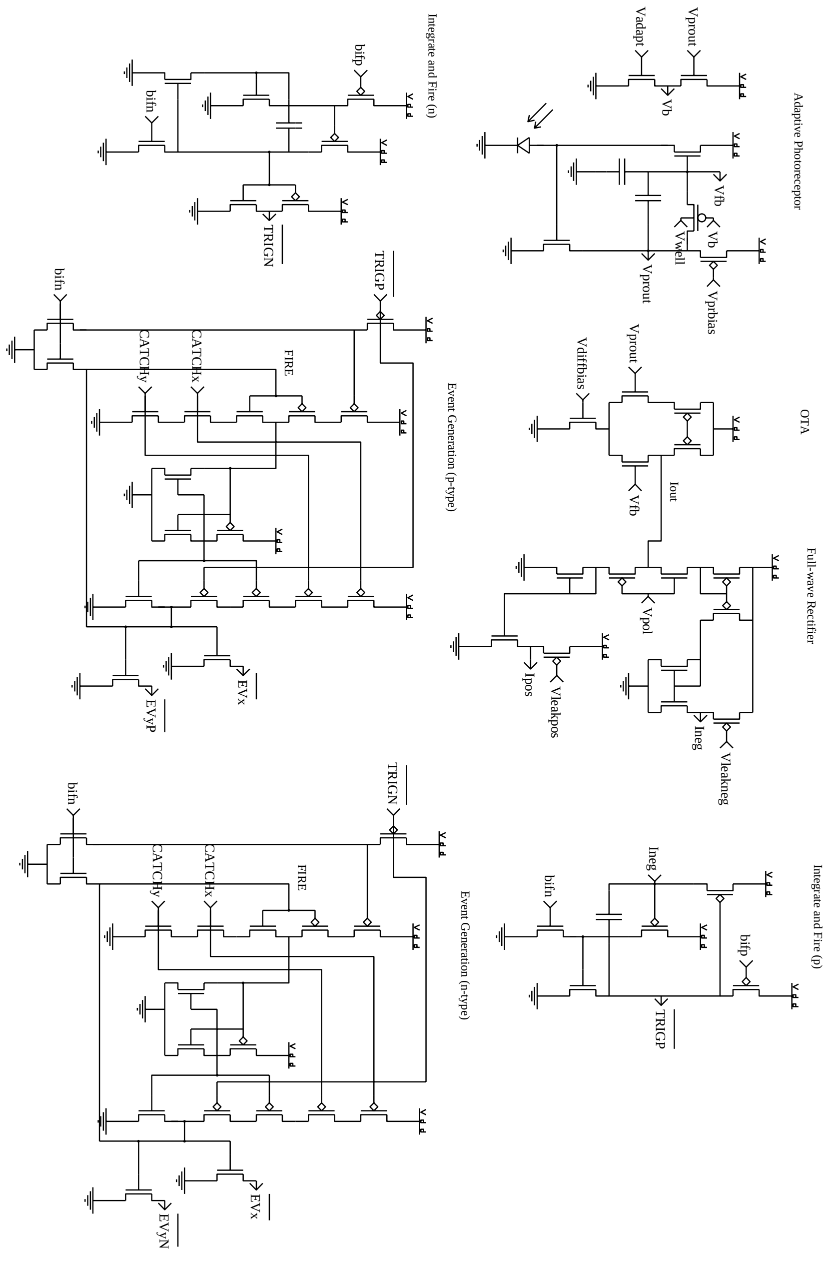


FIGURE 4.9. Circuit diagram showing transistor level implementation of a pixel in the spiking-neuron chip. The adaptive photoreceptor transduces visual information, and a transconductance amplifier (OTA) stage amplifies the signal content after subtracting the long-term mean. This output current is rectified by a full-wave rectifier. Positive and negative currents are separated and a leak-current is subtracted to adjust for zero-error (current in the absence of visual activity). These currents drive P- and N-type integrate-and-fire circuit. The N-type circuit is followed by an inverter. The outputs of the integrate-and-fire circuits are input to two event generation circuit blocks. These blocks communicate with the peripheral AER circuitry to generate requests.

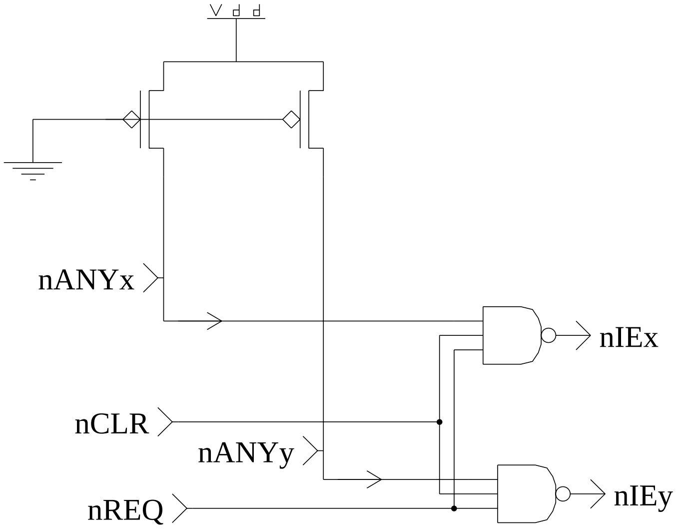


Figure 4.10. IE generating circuit (from Landolt *et al.*, 2001). This serves as the input enable for the external AER latches and is active low based on the inputs indicating idle address lines and no collision.

clear signal is not asserted. This nIE signal enables a latch (see Figure 4.11) to store an event being generated by a spiking pixel. This arrangement of trapping the events by the latches provides the coordinate location of the event on the chip grid. Thus, we can localize the event and associate an address by encoding this information in a $\log_2 n$ manner, where n is the number of rows or columns. In this chip, we have a pixel array of 5×4 which translates into ten rows and four columns. The number of rows is twice that of the pixel value because one part of the response corresponds to the positive half, and the other to the negative half.

The way these address-encoders work is that there are pull-down transistors which pull the address lines to zero - the default value. Moreover, there are some wider transistors, which can overpower the pull-down FETs to pull some of the lines high, generating an address. As discussed earlier, this AER scheme utilizes two lines of address for every bit, one for logic ‘1’ (ADDR line), and the other for logic ‘0’ (ZEROS line). The size of the pull-down FETs was designed considering that they had to be wide enough to pull the ADDR and ZEROS lines to ground within a reasonably short length of time. At the same time they had to be weak enough to be easily overpowered so the lines could be pulled high reasonably quickly. These are inverted to avoid adding more load, that is, to provide enough driving power. The inverters that drive the nADDR lines are four times wider than the ones driving nZEROS. This is because the nADDR address bits are taken to the pads and it is highly advantageous to have high current drive in a signal that goes off-chip. These nADDR and nZEROS lines are used by the arbitration circuitry.

4.2.7 Arbitration and REQ generation

In this subsection, the arbitration logic circuit generated by the silicon compiler is described. This is where the event train is used to generate the REQ signal, the collision between simultaneous events is detected, and the CLR signal is generated to unlock the invalid (collision) state or simply allow the latches to capture a new event. The collision is detected using the dual inverted address lines: nADDR and nZEROS. Figure 4.12 shows an active-low circuit implementation. The collision state is detected by AND-ing each of the ADDR and ZEROS lines, and then OR-ing these outputs to

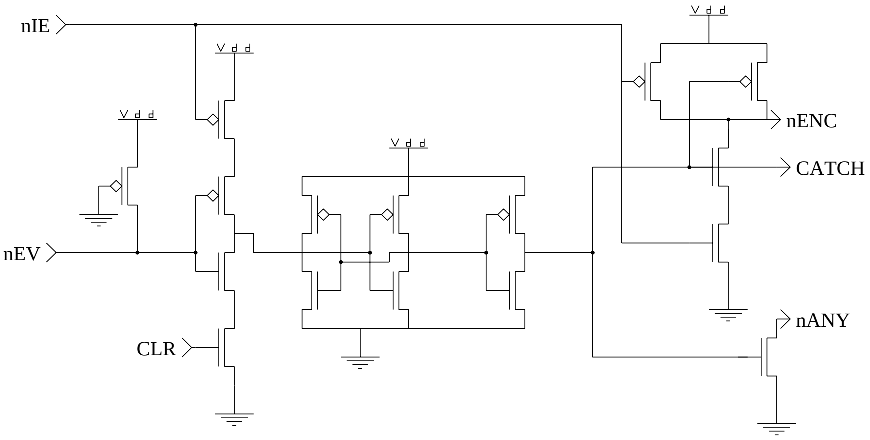


FIGURE 4.11. AER latch circuit (from Landolt *et al.*, 2001). This digital latch with nIE as enable and CLR as the disable signal latches the request from a pixel. The successful encoding of the request leads to generation of an acknowledge signal back to the pixel (CATCH) and a control signal (nANY) to indicate that a request has been made.

check for all the address bits. The INVALID signal asserted by this logic circuit is used to identify a collision, if any.

Figure 4.13 shows an active low implementation of the logic circuit. The VALID signal pertaining to no collision is generated by exclusive-OR-ing the ADDR and ZEROS lines, and then AND-ing the individual address bit outputs. This is used by another logic-block to assert the REQ line. The idle state of these address lines is identified by checking if all the address lines are in a low state. The ADDR and ZEROS lines are individually OR-ed and this output for each address bit is AND-ed to check if they are all low. If so, the active-low nALLZ signal line is asserted. There are two such signal lines for the X- and Y-axes.

The digital logic in generating the REQ and CLR signals is shown in Figure 4.14. This circuitry generates the REQ signal to let off chip circuitry know that the addresses have settled and are ready to be read. The REQ line should be asserted once the address lines have settled to a steady state, so it is better to have a buffer precede the VALID and INVALID circuitry to provide the requisite delay before the generation of request, even in the most unlikely state (Mitros, 2002). This logic-block has been implemented keeping this criterion in mind.

The VALID, INVALID, and nALLZ lines from both the axes are used to generate the request or clear the latches in case of a collision. In the collision scenario, first, one spike causes a valid address to be generated and the VALID flag is raised. A moment later, a second spike collides and the address gets invalidated. To avoid a glitch on the REQ line resulting from the brief instant when the VALID line is active, it has to be ensured that the necessary gate is just a little slower. This gate is shown as the “delayed” NAND gate in Figure 4.14. This gate has to be slow enough to allow latching of the addresses on the bus and to enable the signals IEx and IEy to disallow latching of any more addresses.

4.2.8 Scanner Circuitry

A serial scanner circuit is used to read out the signals from a specific pixel in the chip. Scanner circuits for neuromorphic applications were studied by Mead and Delbrück (1991). This chip employs a vertical scanner (for rows) and a horizontal scanner (for columns). The scanner consists of a shift

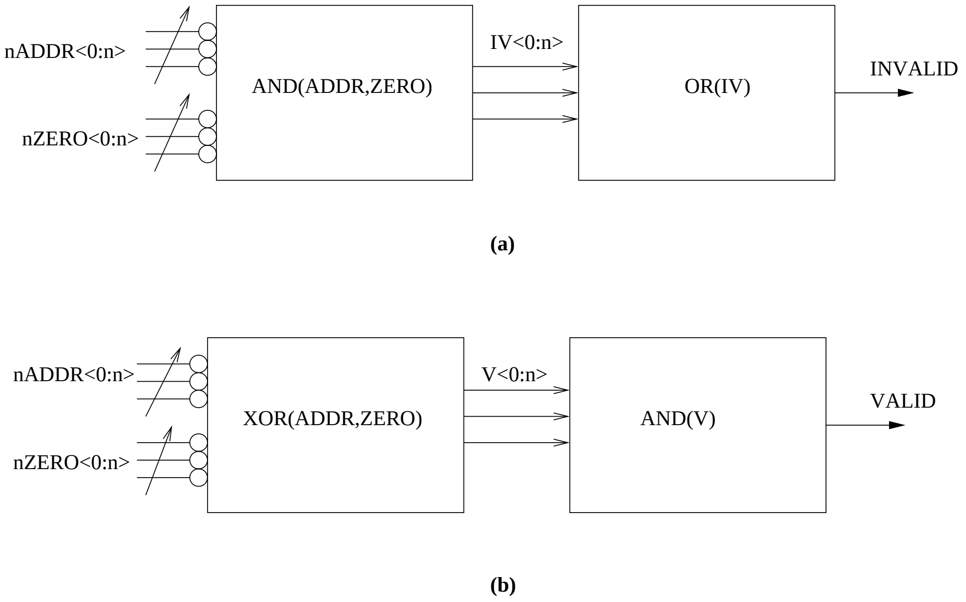


Figure 4.12. Valid-Invalid circuit (from Landolt *et al.*, 2001). (a) The INVALID signal is generated if there is a simultaneous assertion on both ADDR and ZERO lines. Technically, it is the AND of the ADDR and ZERO lines individually, and then OR-ing of the output from this stage. (b) The VALID signal is generated when only one of each pair of ZERO and ADDR lines is high, representing a valid address state. Technically, it is the XOR of the ADDR and ZERO lines individually, and then AND-ing of the output from this stage.

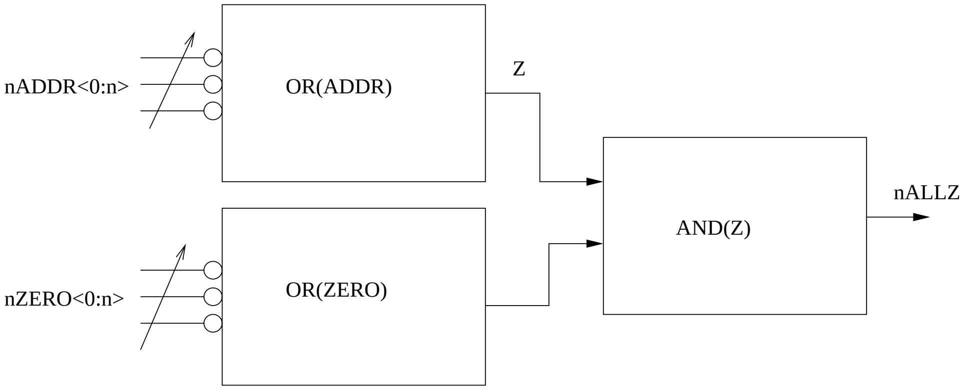


Figure 4.13. Idle state signal generation (from Landolt *et al.*, 2001). The ADDR and ZERO lines produce an active low idle-state output (nALLZ) when none of them is carrying any address. Technically, it is the OR of all the ADDR and ZERO lines, and AND of the output of that stage.

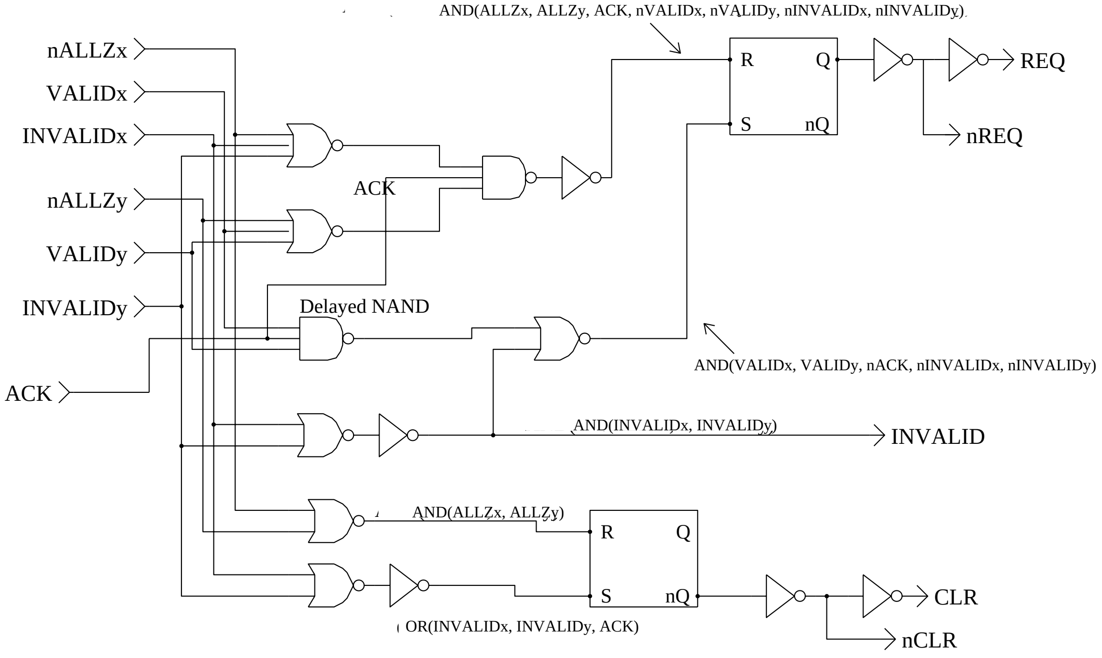


Figure 4.14. REQ and CLR signal generation circuit (from Landolt *et al.*, 2001). Based on the VALID, INVALID and nALLZ signal lines, the REQ line is asserted in case of no collision. The “delayed” NAND gate ensures that the REQ line is asserted only when all the signals have settled. A collision results in raising of the INVALID and CLR signals. Acknowledge of a request also asserts the CLR signal.

register. A logic '1' when stored in a flip flop activates that row or column. A logic '0' disables it. This circuit is mainly used in this chip to access outputs from a specific pixel, and for tuning of the photoreceptor response. This circuit is same as that discussed in Section 3.3.6.

4.2.9 Summary

This chip has been fabricated through MOSIS in a standard 1.6 μm process. The layout was done using the L-Edit tool (Tanner-EDA, 2003). Layout of the spiking-neuron pixel showing the sub-circuit stages is shown in Figure 4.15. The entire chip was compiled using a silicon compiler program (User-Programmable Interface: Tanner-EDA, 2003). Computer software was written in this compiler such that for any given size of the pixel array, peripheral arbitration circuitry layout was generated automatically. The layout as generated by the computer software for a size of 5 pixels is shown in Figure 4.16. These digital peripherals circuits were then arranged using the same code to generate the full layout of the entire chip. The layout generated by the computer code, and a photomicrograph of the fabricated chip, are shown in Figure 4.17.

4.3 Mixed Signal Design Considerations in the Spiking-Neuron Sensor

Mixed signal designs in CMOS require multiple techniques to achieve success. The spiking-neuron chip has a complicated design with analog and digital sub-circuits in every processing unit (pixel). In this chip, the noise sensitive analog circuit is embedded with a digital spike generating circuit in a single pixel, and has been carefully designed. The following issues that may hamper performance in a mixed signal architecture have been addressed.

4.3.1 Digital CMOS Process

This chip has been fabricated in a CMOS process for digital designs. A digital CMOS process optimizes parameters for logic functionality: switching speed, low voltage power supplies, submicron geometry, and high component density. The process does not address analog design issues related to process variations, parasitic resistance and capacitance, and parametric variance. Parametric variance is generally higher for smaller devices. In our chip, we have addressed this by making the analog transistors bigger than the minimum geometry.

4.3.2 Power Isolation

High frequency noise is coupled to the power lines due to high speed switching of the digital circuits. Analog circuits are very sensitive to such fluctuations in their biases. An isolated power supply for analog circuits therefore becomes mandatory. Our chip has a separate analog and digital power supply line. External low frequency power filtering (a capacitor between power and ground) was used to eradicate noise coupling effects. This technique, though, becomes ineffective at very high frequencies due to dominance of inductance and capacitance of the bonding wire and integrated circuit package.

4.3.3 Ground Noise

The ground reference has a dynamic variance due to substrate currents, and this may induce coupling of noise from switching transients between grounds across the integrated circuit. Small induced voltages due to this effect may lead to large amount of current modulation in a circuit like the current mirror. In our chip, the ground line has been connected to the bonding pads at regular intervals to reduce the effect of substrate leakage currents (see Figure 4.17).

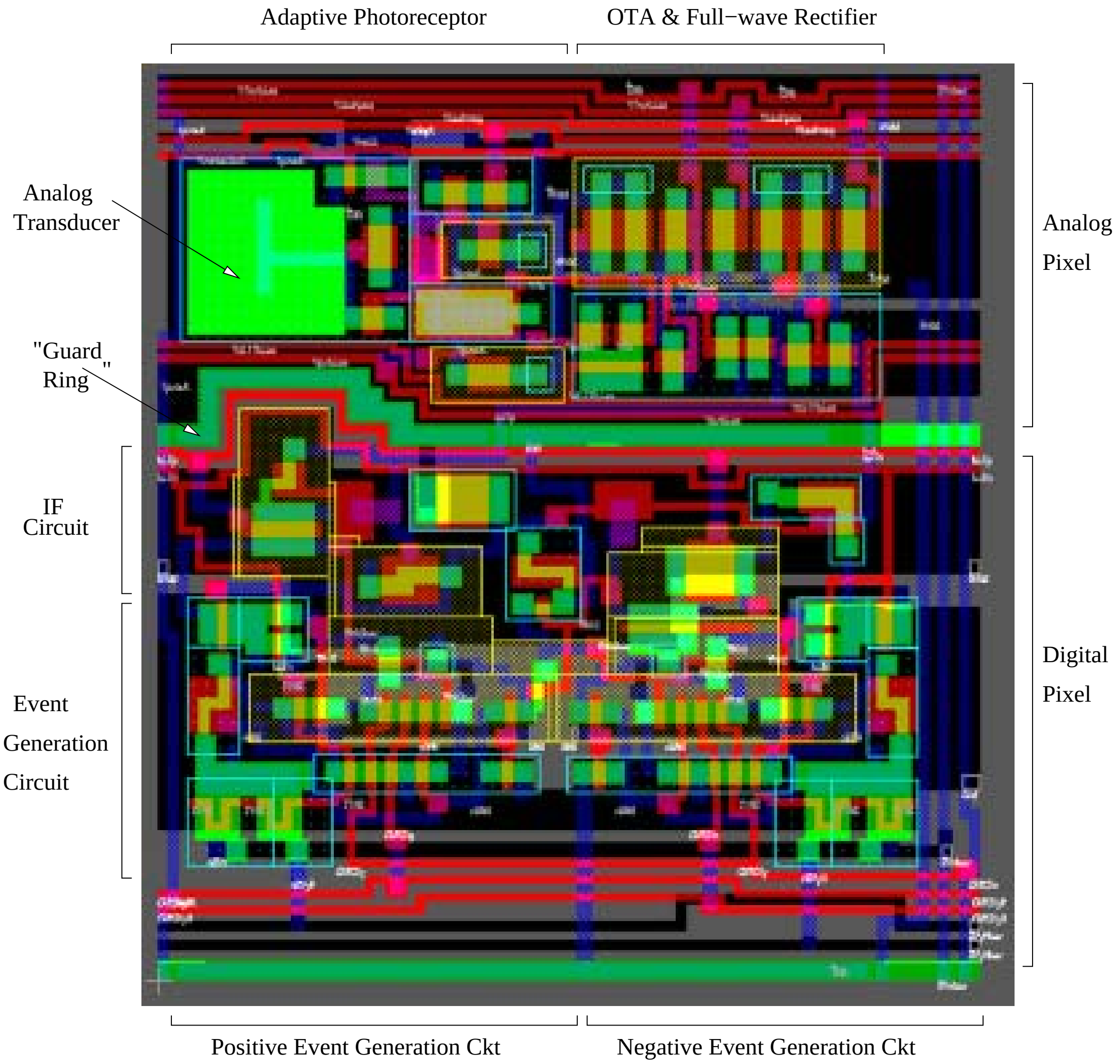


FIGURE 4.15. Layout of the spiking-neuron pixel showing the analog and digital sub-pixels. The internal AER sub-circuits and analog transducer have been shown.

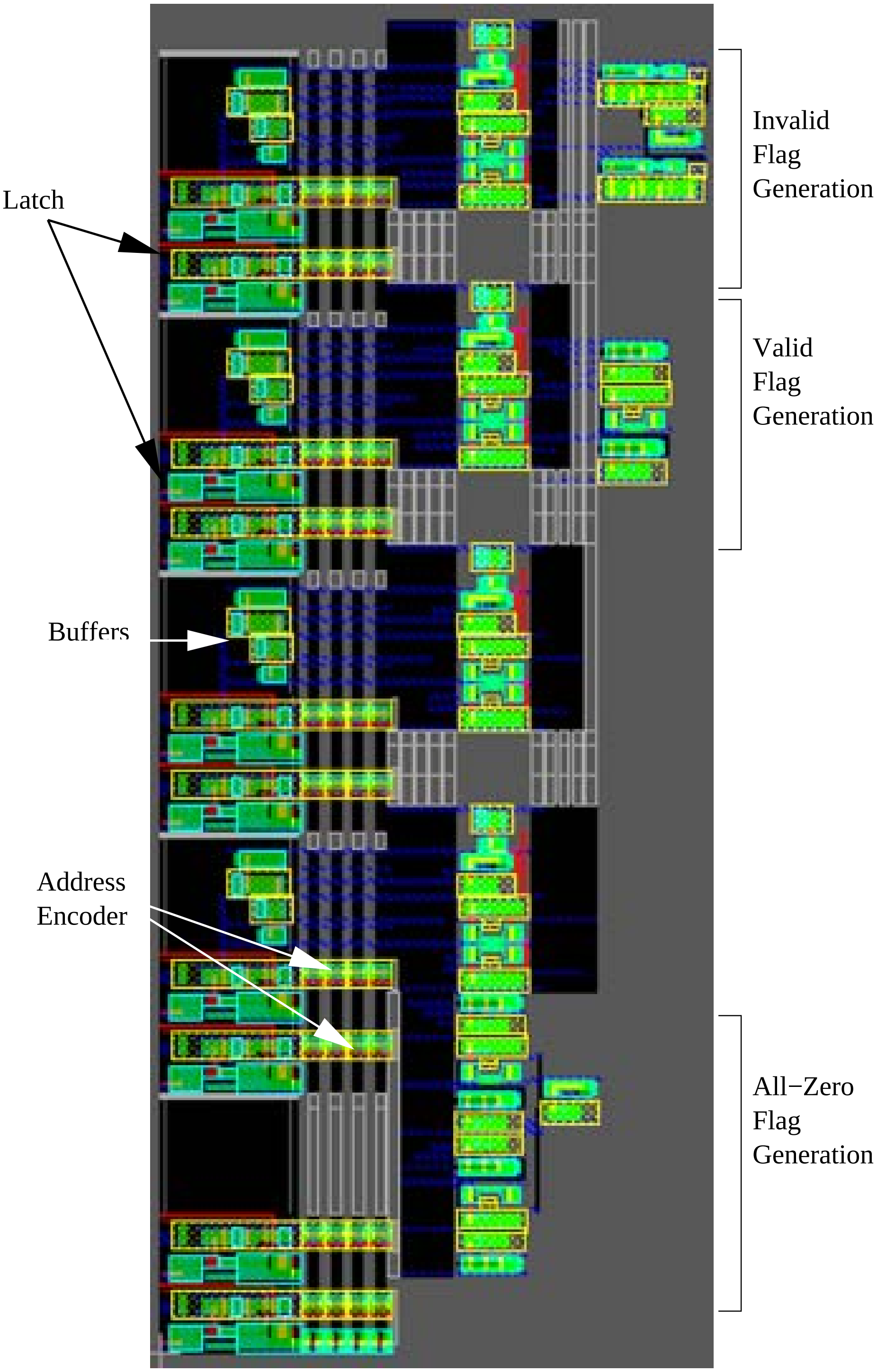
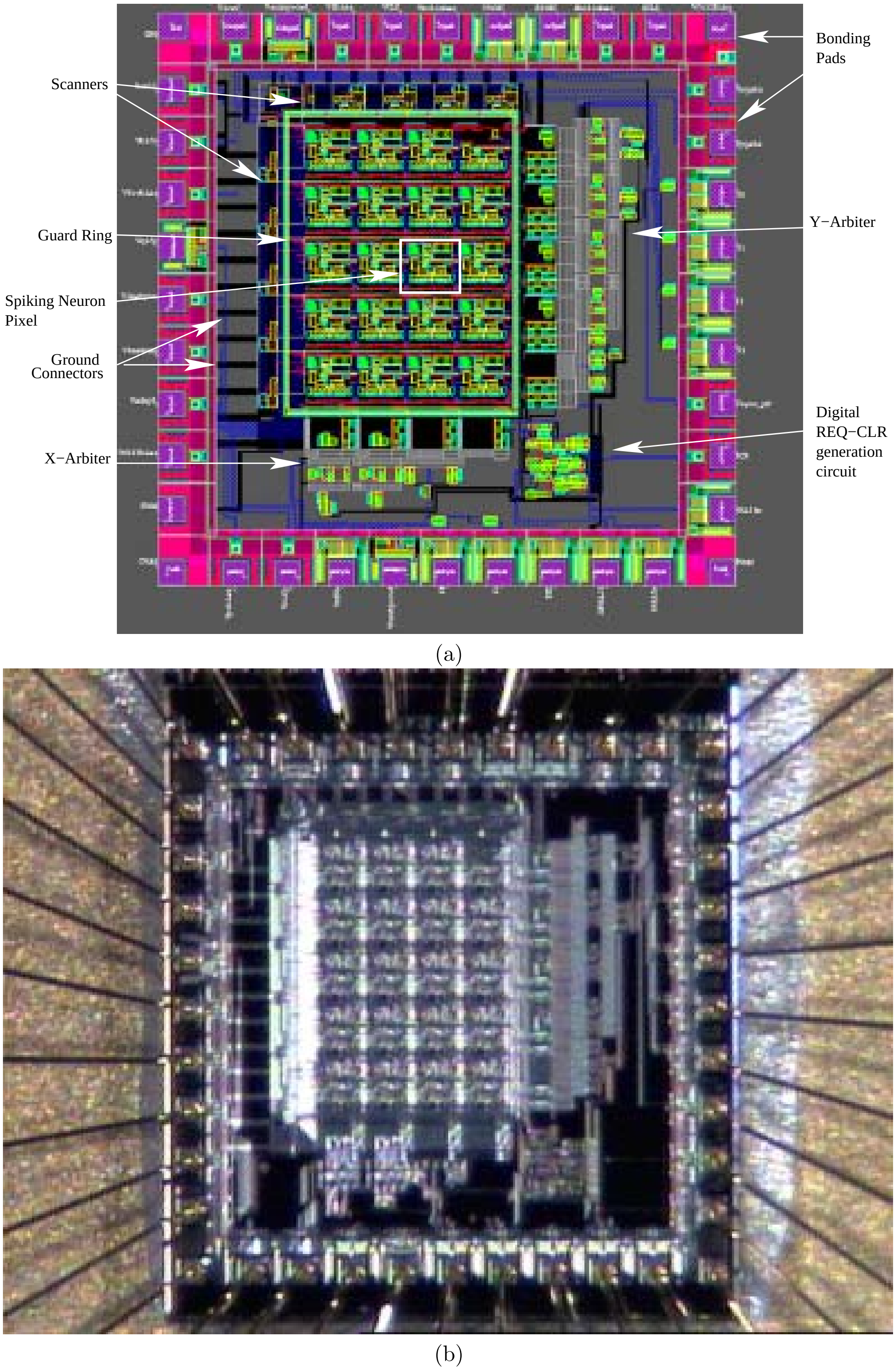


FIGURE 4.16. Layout of the AER arbitration circuit. The address latches, encoders, and arbitration blocks have been shown for a size of 5 pixels.



(a)

(b)

FIGURE 4.17. Layout of the spiking-neuron sender chip. (a) Layout in the L-Edit tool showing an array size of 4 × 5 pixels. The various processing sub-circuits including the scanners have been indicated. (b) Photomicrograph of the fabricated chip in a 1.6 μm process.

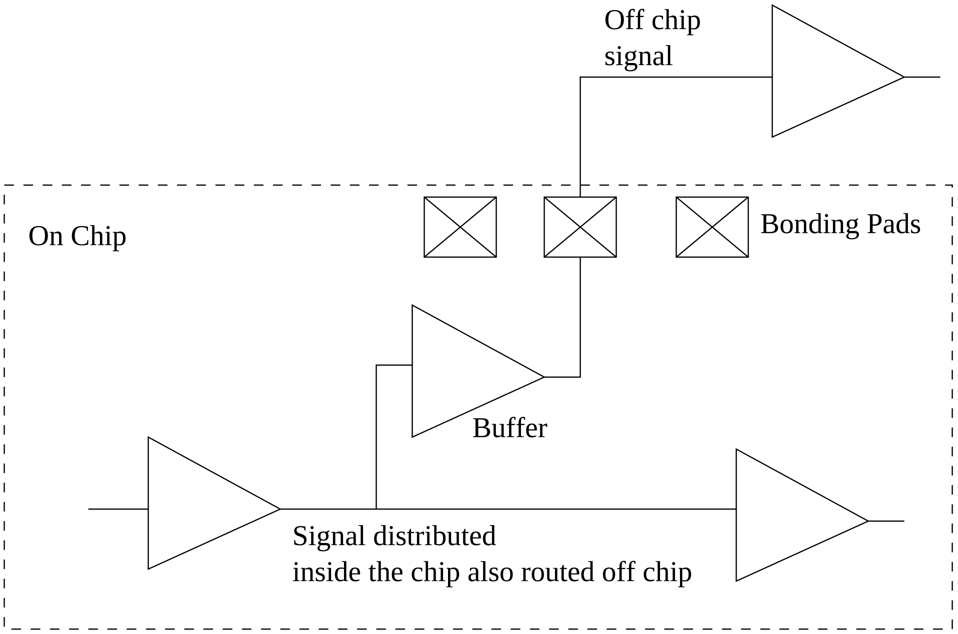


Figure 4.18. Passing signals out of the chip. A buffer amplifier prevents noise off-chip from coupling with the internal signal line.

4.3.4 Differential-mode Output for Noise Immunity

Our chip operates in a noisy environment and the effects of noise coupling cannot be eliminated entirely. However, the use of differential circuits greatly minimizes noise coupling by common-mode signal rejection. In our chip, we have implemented a differential transconductance amplifier (OTA) that removes the long-time mean signal from the instantaneous response of the photoreceptor, and converts this differential response into a current (see Section 4.2.2). This reduces the common-mode noise induced by the digital circuitry on the photoreceptor response and its long term mean.

4.3.5 Off-Chip Signals

The signals going outside a chip have more noise coupling than signals inside it. This couples external noise on that pin to the circuits connected with that signal line inside a chip. Figure 4.18 shows that a buffer amplifier can be used to prevent coupling of external noise to the internal signal. In our chip, all the address lines are distributed to several processing blocks for arbitration and request generation. The addresses lines are also off-chip signals of the AER protocol. We have used a two-stage inverter to buffer these off-chip going signals (see Figure 4.18).

4.3.6 Layout Considerations

As digital circuits switch, a fast transition can inject a lot of carriers into the substrate. This induces unwanted currents in analog circuits. For less coupling, analog and digital circuits have been isolated by a guard ring (see layout in Figure 4.15). Since peripheral transistors are made wider and faster to make the AER bus run sufficiently fast, they introduce more switching noise than the AER-related transistors within the pixel. Hence, a guard ring between the peripheral AER circuitry and the core array of pixels was laid out (see Figure 4.17). The guard ring consists of native and well-type

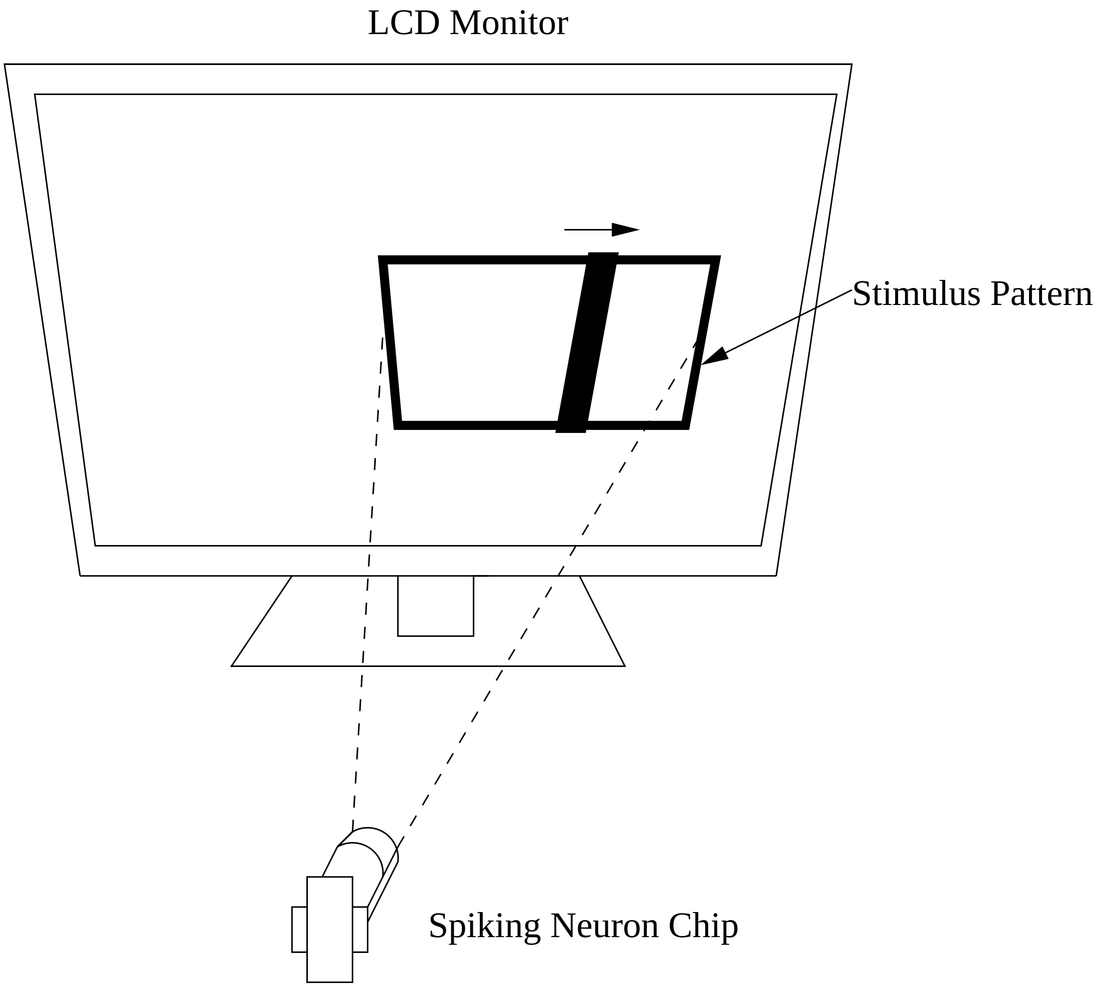


FIGURE 4.19. Setup for characterization of the spiking-neuron chip. The stimulus pattern is generated using a C program and presented on a LCD screen kept at a distance of roughly 17 inches from the chip.

regions. The native region is tied to ground, while the well-type region is connected to *Vdd*. Thus, the native region absorbs the holes, while the well-region absorbs the electrons released into the substrate by the switching in digital circuit.

A pixel (or array element, more generally) behaves more like another pixel if both are surrounded by the same type of structures. It is therefore a good design strategy to make dummy artifacts as the corner rows and columns. In this design, however, no such pseudo-structure has been made on-chip because of area constraints. Therefore, the pixels bordering the peripheral AER circuitry show edge effects. We have used a current mirror with leakage current subtraction mechanism (see Section 4.2.3) to take care of this added noise from the AER circuit, and a threshold can be set for noise immunity.

Our chip has been fabricated keeping in mind the above mentioned design issues. This does eliminate noise to some extent but control signal biases are still prone to noise coupling effects. In the next section, we discuss the characterization results of our spiking-neuron chip.

4.4 Spiking-Neuron Sensor Characterization

The spiking-neuron chip has been fabricated through MOSIS in a 1.6 μm process on a 2.2 mm × 2.2 mm die. In this section, an initial characterization of the spiking-neuron sender chip is presented. The experimental set up is as shown in Figure 4.19. Computer software was used to generate a visual input pattern for the chip, and presented on an LCD screen. The stimulus consisted of a square wave grating with a spatial frequency of 0.01 cycles/chip-pixel, and a temporal frequency of 1 Hz. This stimulus was used to present a single sharply rising contrast edge to generate a spike train on the AER bus. The sharply rising edge causes the contrast to change suddenly and then remain fixed at a value. This produces a burst of spikes. Then the pixel adapts to the changed illumination value and the spiking subsides. If the output of the OTA stage after rectification is sufficiently large to overcome the leakage current, multiple requests are sent on the AER bus (see Figure 4.20). Each request for the AER bus is arbitrated independently. Hence in the case of multiple spikes (a burst of spikes) from a particular pixel, the requests are interleaved with requests from other pixels. The burst length of this sensor is measured by stimulating it with the computer generated stimulus and by connecting the REQ line directly to the ACK line. This is the fastest operation mode for the AER bus and is called self-acknowledge mode. It yields the fastest possible event cycle, taking approximately 62 ns per request-acknowledge cycle. The maximum spike rate over the REQ line was measured to be 16.34 MHz. The IF circuitry imposes a threshold on the output of the rectified OTA-circuit output. Thus, each sender pixel fires only in response to a stimulus above a fixed phototransduced current. Due to inevitable random noise in the analog part of this system, a sender pixel will fire probabilistically when visually stimulated near a threshold current.

Figure 5.4 shows the variation in the spiking response of the pixel with variation in the contrast of the stimulus. As stimulus contrast is varied, the burst width increases and then saturates. No response is seen at very low contrast due to the threshold set by the spike-generating circuitry. The reason for shorter burst width at lower contrast is due to the smaller photoreceptor output. The pixel therefore adapts to this small photoreceptor output quickly, so causing activity on the request line for a shorter time. In case of higher contrasts, the photoreceptor response is large, so the pixel takes a longer time to adapt, due to which the activity on the request line goes on for a longer time period. An external circuit consisting of an address comparator and 555 timer was used to record the length of the burst.

4.5 Summary

The advantages of modular architectures over monolithic implementations have been discussed in this chapter. The spiking-neuron sender chip described in this chapter can be used to develop a full-scale sender-receiver system implementing the FD-cell based target tracking scheme. The next chapter discusses the simulation of such an agent, and possible extension of the monolithic target tracking system into a modular target tracking agent.

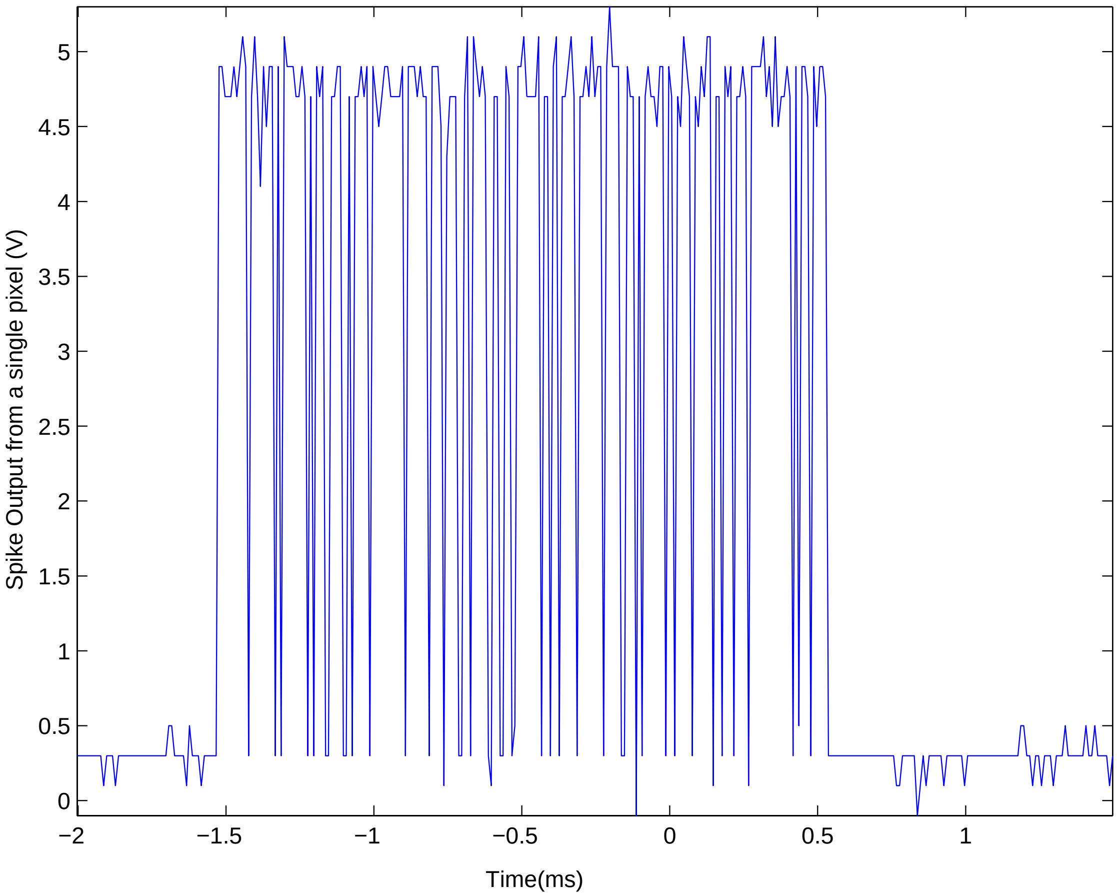


FIGURE 4.20. Burst of spikes from a pixel in the spiking-neuron chip. The actual request acknowledge cycle of the AER bus is of the order of nanoseconds. In this plot, the spike duration has been lengthened with an external 555-timer circuit to show the burst of spikes. The stimulus was a square wave grating generated by a computer program, displayed on an LCD screen.

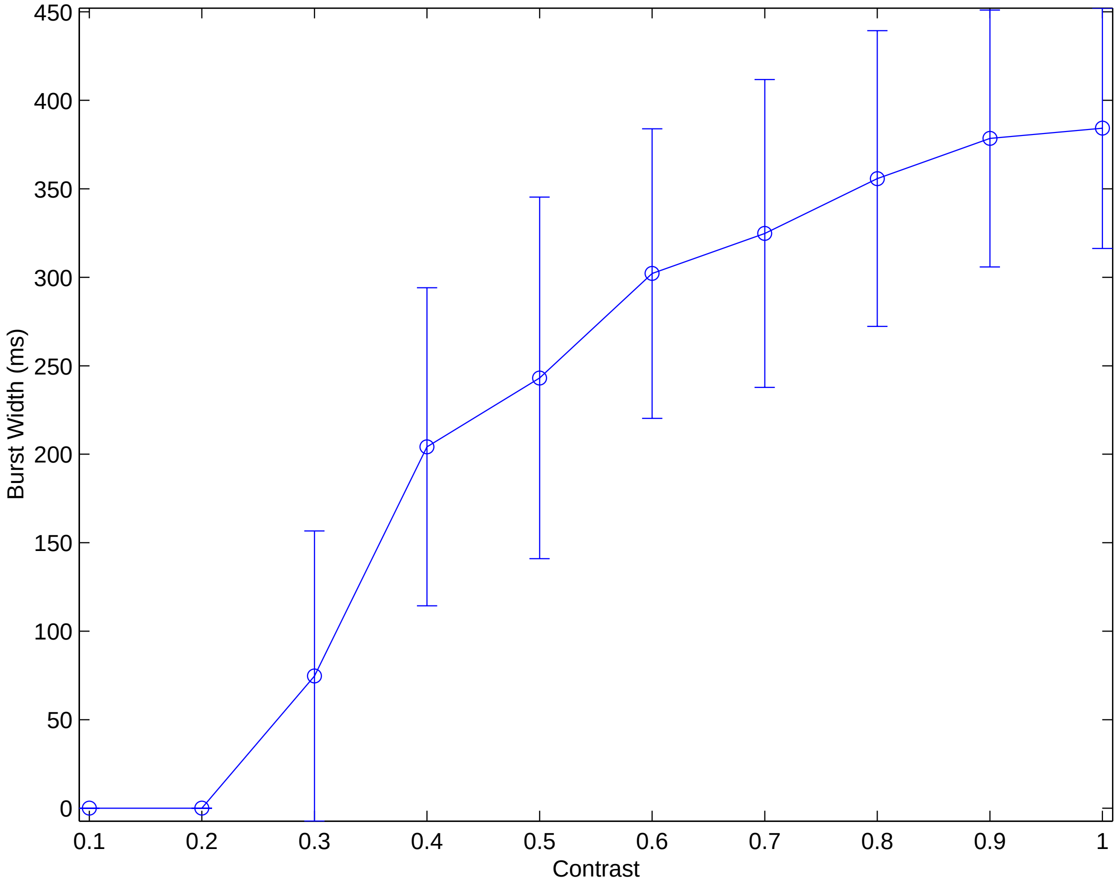


Figure 4.21. The burst width of the spiking-neuron pixel output versus contrast. The burst width increases with contrast and tends to saturate. Circles represent the mean of 10 readings each taken from 2 different pixels. The error bars represent the standard deviation.

Chapter 5

Target Tracking Agent

This chapter presents simulations which predict the target tracking performance that would be possible for an autonomous mobile robot using a modified FD algorithm. Tracking is an involved procedure that requires motor activity along with visual computations. This is necessary to ensure that the tracking system does not lose sight of the object being pursued. Thus, the implementation of such a tracking system involves capturing visual data and computing a turning response for the robot. The FD sensor can be used for generating this turning response. This response is used by a control system to steer the platform movement of the robot. In this chapter we describe the performance of the simulated robot without regard to implementation. The implementation of this system has been suggested in Chapter 6 as the future extension of this work.

The FD-cell based sensor, as has been described in Chapter 3, provides a small-field motion output that can be scanned out using the scanner circuitry. This individual pixel response can be summed up over the whole chip to give an approximate small-field response. This output is proportional to the turning torque response in a fly (Reichardt *et al.*, 1983). Hence, this response can be used to generate a signal that can control the yaw movement of a motor unit. The fixation behavior of the fly has been studied and modeled extensively. A simulation and robot implementation of fly visual orientation behavior has been done by Huber and Bülthoff (1998). Huber and Bülthoff showed in their study that a large-field motion response can be used to design autonomous agents that have a fixation behavior in uncluttered scenarios. In our simulations we have considered cluttered environments and presented an algorithm that can be used by a target tracking agent in such arenas.

5.1 Modifications to the FD algorithm

The FD algorithm as suggested by Reichardt *et al.* (1989) models target fixation behavior of a fly from a fixed platform. Therefore, it does not take into consideration the complexity of the optical visual-field while following a target. During pursuit, the translatory motion of the fly (robot) is coupled with the rotatory motion required to keep the target in the center of the view field. The translatory motion of the robot produces an expanding view field with the focus of expansion (FOE) shifted to one side of the view field due to rotatory motion. If the rotatory motion is small, this shift is small and can be approximated as a pure translatory motion. However, if the rotatory motion is large, the FOE shifts completely out of the view field. The robot perceives a pure rotatory motion in that case. The original FD algorithm considers only the latter case where the motion is purely rotational. In our algorithm, we have accounted for both these conditions. When the angular velocity of the robot is smaller than a threshold value, the motion is approximated by a pure translatory motion. Under this condition, unlike the clockwise and counter-clockwise pooling as shown in Equations 3.5 and 3.6, pooling of the preferred or null direction motion components from both "eyes" is computed. So the normalization stage is now represented by:

$$y_i^+ = \frac{v_i^+}{P_l^+ + P_r^+} \tag{5.1}$$

$$y_i^- = \frac{v_i^-}{P_l^- + P_r^-} \tag{5.2}$$

where the notation is the same as that used in Chapter 3. When the angular velocity of the robot exceeds the set threshold, the output of the normalization stage is represented by Equations 3.5

and 3.6. This can be implemented by using a binary signal to switch the positive and negative components between the two FD chips being used as the sensors.

Also, the original FD algorithm assumes direction selective cells responsible for small-field motion detection. So the output of Equation 2.21 may be positive or negative depending upon the direction of motion in front of that eye. The final output of the small-field system, if positive, triggers a clockwise turning response and vice-versa. Thus, motion in the leftward direction in front of the right eye will generate a torque such that the robot moves towards the left. This would make the robot swerve away from the target. In our algorithm, we take only the amplitude of the response from both left and right side sensors. Therefore, any motion discrepancy on the right side will always result in a positive response, and similarly for negative on the left side. The response from both the "eyes" are symmetric. The response of the right side sensor is given by:

$$R_r(t) = \sum_{i=1}^{N} abs\left([y_i^+]^n - [|y_i^-|]^n\right) \tag{5.3}$$

where the notation is same as that in Chapter 2. The exponent n is used to enhance the target response, and in our algorithm n is a cubing factor ($n = 3$). The present iteration of the FD chip does not include this functionality and modifications to the hardware for the same have been suggested in Chapter 6.

The final response of the sensor guiding the robot is given as:

$$R(t) = R_r(t) - R_l(t) \tag{5.4}$$

Finally, the response from the FD cell based model is a combination of a running average and its instantaneous response to produce a torque output. It can be shown from basic physics that the angular displacement of an object on application of a torque is given by a combination of the average and instantaneous value of the applied torque. The small-field motion output from the FD-sensor is modeled as:

$$R(t) = K_{fd} \cdot R_{avg} + R_{ins}(t) \tag{5.5}$$

where $R_{ins}(t)$ is the instantaneous small-field response, R_{avg} is the average response computed over N time steps, and K_{fd} is a constant. Since the average response does not change much, it does not have much control over the instantaneous angular velocity. In our algorithm, we have used only the instantaneous response of the FD sensor given by Equation 5.4.

However, it must be noted that as the number of objects in the background increase, so does the compensation of the output signal. Thus, the strength of the instantaneous small-field response of the FD sensor goes down. So, a variable gain is required such that it can compensate for the normalization induced by the background motion components. In our simulations we have handled this by turning up the gain factor such that the response in a cluttered scenario is also able to induce a turning response. To compensate for the large response in case of fewer background objects, we have set a limit for the maximum angular turn the robot can take in one time step. In the next section we discuss the dynamics of the robot used in our simulations.

5.2 Dynamics of the Simulated Robot

The control signal generated by the FD sensor is used to steer the platform motion of a simulated robot. As seen in Figure 5.1, we have simulated a system where the robot has a fixed translatory velocity v_r, and the response produced by the FD sensor is proportional to its angular velocity. This is modeled as:

$$\dot{\theta}_r = G \cdot R(t) \tag{5.6}$$

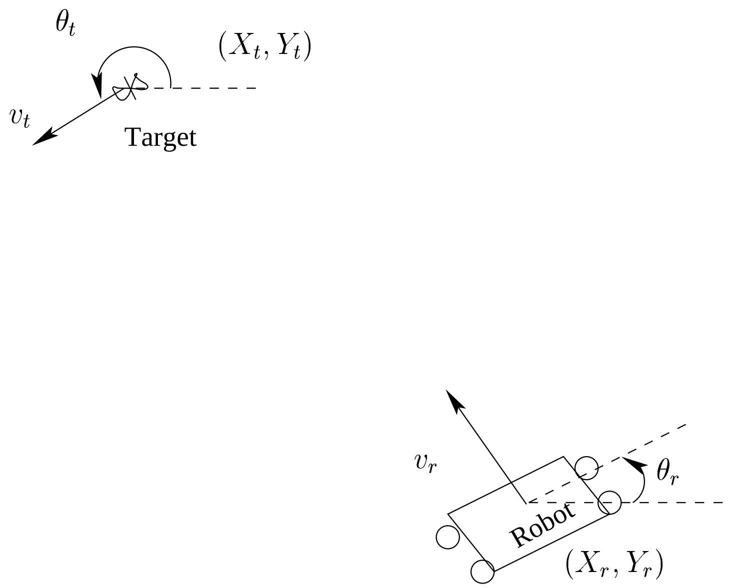


FIGURE 5.1. Sketch of the tracking system used in our simulation. The robot has a translatory velocity v_r modulated by its angular orientation in the world θ_r. The target moves with a velocity v_t in the arena.

where G is the gain of the control system, $R(t)$ is the instantaneous output from the FD sensor, and $\dot{\theta}_r$ is the angular velocity of the robot.

This angular velocity generates an angular movement of the robot given as:

$$\theta_r = \theta_{r0} + \int_0^t \dot{\theta}_r dt \tag{5.7}$$

where θ_{r0} is the initial orientation of the robot and θ_r is the angular orientation of the robot at time t. This angular velocity together with the translatory speed v_r changes the position of the robot, given as:

$$X_r = X_{r0} + \int_0^t v_r \cos\theta_r dt \tag{5.8}$$

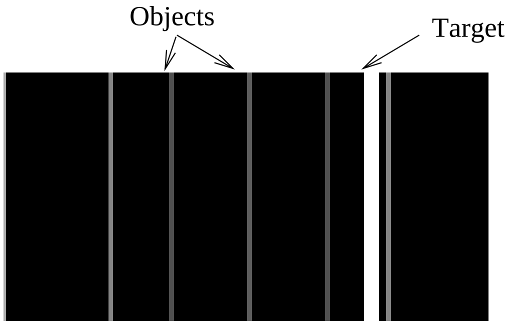
$$Y_r = Y_{r0} + \int_0^t v_r \sin\theta_r dt \tag{5.9}$$

where X_{r0}, Y_{r0} are the coordinates of the starting point of the robot, and X_r, Y_r represent the current position of the robot.

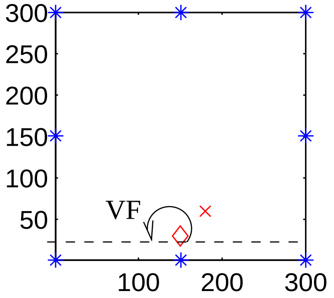
The target moves either linearly or in a sinusoidal manner with a speed v_t and oriented as shown in the figure.

5.3 Simulation Setup

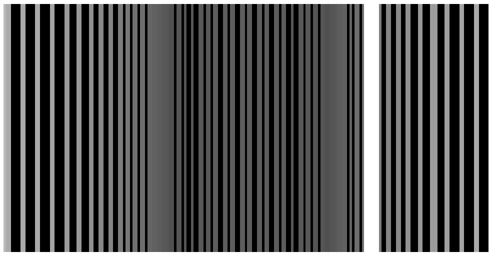
The above model of a robot tracking system has been simulated in MATLAB. The robot was situated in an arena which was 300 × 300 space-units in size. The simulated world is two dimensional and the robot can move along both axes. However, the visual field of the robot is only 1-D, that is the objects and the target have no height. This arena is surrounded by four walls. Each wall is



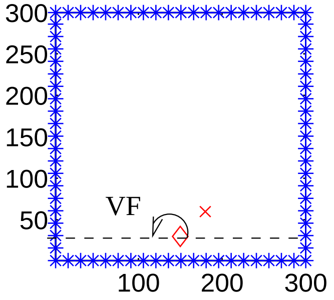
(a) Visual field with 2 objects/wall



(b) Top View with 2 objects/wall

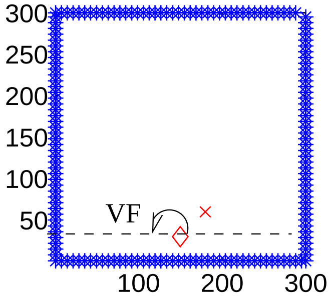


(c) Visual field with 20 objects/wall



(d) Top view with 20 objects/wall

(e) Visual field with 40 objects/wall



(f) Top view with 40 objects/wall

FIGURE 5.2. Visual field of the robot and top view of the arena with robot and objects. The arena is a 2-D plane 300 × 300 space-units in size. The robot is shown by a diamond ($\diamond$), the target by a cross ($\times$), and fixed objects on the walls by asterisks ($*$). The robot's visual field (VF) is 180° and has been indicated in the top view graphs. The contrast has been scaled for distant objects with $K_D = 100$. (a) and (b) Visual field of the robot and the top view of the arena with 2 objects on each wall. The target and the fixed objects are well separated in the visual field. (c) and (d) Cluttered visual field with 20 objects on each wall. The target can be distinguished by its relative motion to the background. (e) and (f) Blinding effect due to 40 objects on each wall on the arena. The target's edges are non-separable from the cluttered background.

composed of variable number of objects controlled by a parameter n_{obj}. The contrast of the objects varies with distance as:

$$C_i = min\left(1.0, \frac{K_D}{D_i}\right) \tag{5.10}$$

where D_i is the distance of an object from the robot, and K_D is the distance within which the contrast is true (not scaled). The contrast C_i makes the further objects less visible in the robot's visual field. The contrast and the number of fixed objects on each wall, as seen by the robot, can be set as per experimental requirements. The object size varies inversely with its distance from the robot. Thus, nearer objects cover a larger part of the visual field as compared to further ones. The visual field of the robot has an angular extent of 180°. The starting position of the robot (X_{r0}, Y_{r0}) and the target (X_{t0}, Y_{t0}) may also be set as per the experimental requirements. Figure 5.2 shows the visual field of the robot and the top view of the arena for two, twenty and forty fixed objects on each wall of the arena. The contrast was scaled for distant objects by setting $K_D = 100$.

In the case of only two objects on each wall of the arena, the target was easily detectable. A collision was detected when the distance between the target and the robot was within 6 space-units. This arbitrary threshold was selected assuming that the spatial extent on both the target and the robot is 3 space-units each. As the number of fixed objects increases to twenty, the robot's visual field becomes cluttered. In Figure 5.2e there are forty objects on each wall of the arena. The presence of this many objects makes the entire visual field extremely cluttered such that no distinction can be made between the edges of the objects with each other and the target. The top view of the arena under the discussed three scenarios are also shown.

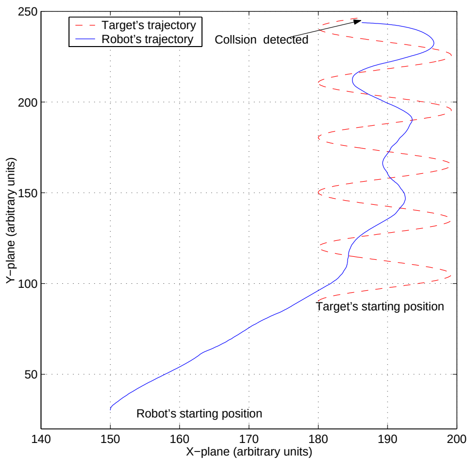
5.4 Results

In this section, the various results from simulation experiments are presented. The parameters and background settings have been varied and the results have been plotted.

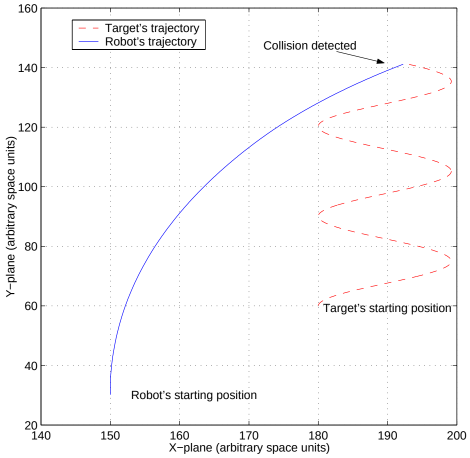
5.4.1 Experiment 1

In this experiment, the approach behavior of the robot towards a moving target was studied. The robot speed was fixed to be a constant, $v_r = 18$ space-units/sec. The target moves sinusoidally in the arena. Its speed along the Y-axis was $v_t = 12$ space-units/sec and it moves sinusoidally along X-axis with an amplitude of 90 space-units. The initial position of the robot was (150,30) and that of the target was (180,90), such that the target was on the right side of the visual field (see Figure 5.3). The gain of the control loop was set at $G = 200$. The contrast scaling factor was set at $C = 1$ for all the objects (K_D infinite). Three sets of stimulus conditions are presented to the robot as shown in Figure 5.2, however, there was no contrast scaling. In the first case, only the target forms the 1-D visual field and there are no other fixed objects ($n_{obj} = 0$). In the second case, twenty fixed objects are present on each wall of the arena ($n_{obj} = 20$). In the last case, the background was made up of forty-five objects on each wall ($n_{obj} = 45$).

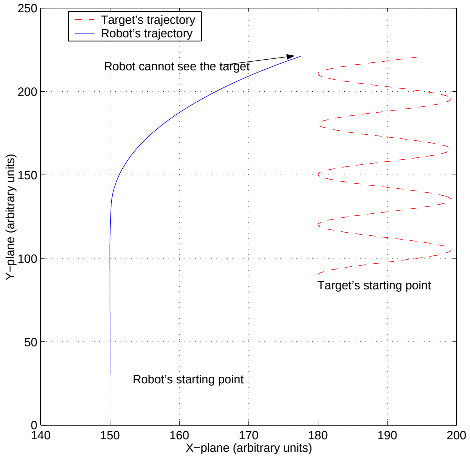
Figure 5.3 shows the robot and the target trajectory for the three mentioned conditions. The robot was able to track the target successfully in the first two scenarios. It can be seen from the robot's trajectory that the response of the FD sensor was stronger when there are no background objects. This makes the robot follow the target very closely until it collides with the target. In the second scenario, the presence of twenty objects on each wall of the arena produced motion components that normalized the response from the FD sensor. So, the response was weaker but sufficient to track and collide with the object. In the third scenario, the presence of forty-five objects on each wall makes it impossible to distinguish between the edges of the fixed objects and the target. So, initially the course of the robot was straight ahead, just the same as it would have had been if it was moving with its "eyes" closed (no visual input). When the robot becomes sufficiently close to the target, the strength of the target increases, and thereby causes the robot to swerve



(a) No fixed object on the wall



(b) 20 fixed objects on each wall



(c) 45 fixed objects on each wall

FIGURE 5.3. Simulation of a moving target with variable number of fixed background objects. (a) Robot sees only the moving target and no background objects. It can be seen that the robot's trajectory is strongly influenced by the target's trajectory. (b) Robot sees the target in the presence of 20 fixed objects on each wall of the arena. The robot's trajectory is affected by the target in a weaker manner. (c) Robot sees the target in the presence of 45 fixed objects on each wall of the arena. As seen, the robot was not able to distinguish between the target and the object till it has come quite close to the target. This late response by the robot was not sufficient to track the target.

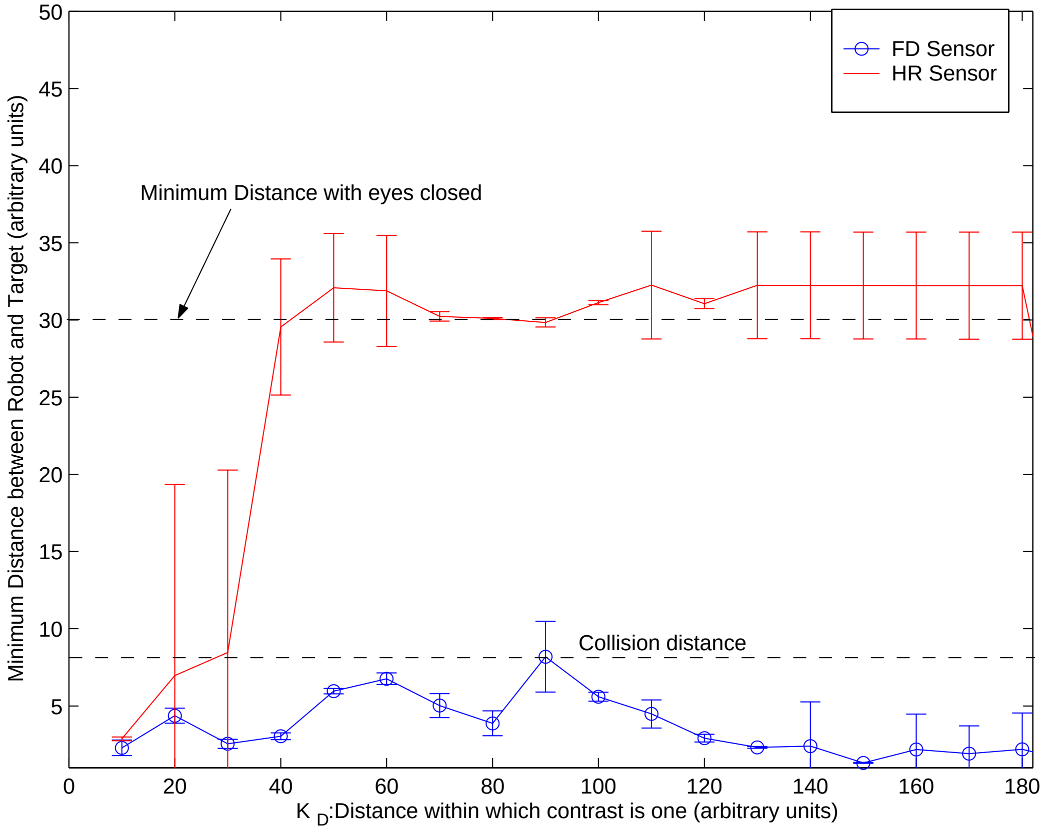


FIGURE 5.4. Simulation results for the HR and FD sensors. For small values of K_D, the HR sensor is able to track the target as well as the FD sensor. However, the minimum distance between the target and the robot increases with reduced contrast scaling in the case of the HR sensor. The FD sensor is more or less unaffected by the contrast scaling. The collision distance has been set at 6 space-units. The minimum distance between the robot and the target when the robot is moving without any visual inputs is also shown. The vertical bars indicate standard deviation over ten simulations.

towards the target. This late movement was not sufficient to track the target and the robot soon crosses past it. The trajectory of the target is shown by a dashed line, and the robot's course by a solid line. The robot implemented by Huber and Bülthoff could fixate only on a single black stripe on a white background. These results indicate that our sensor could be used for tracking in cluttered environments as effectively as for tracking a single moving target with no objects.

5.4.2 Experiment 2

In this experiment, we present a comparison of the performance between the HR sensor and the FD sensor for target tracking. The HR sensor output was obtained by taking only the large-field response of the FD sensor before the normalization, given as:

$$R_{HR}(t) = [P_r^+ - P_r^-] - [P_l^+ - P_l^-] \tag{5.11}$$

where the notation is same as that given by Equations 3.1-refeqHRpool2. This was used to vary the angular velocity of the robot with the control loop gain set at $G = 50$. The arena was generated such that the contrast of the objects was scaled according to their distance from the robot for each simulation step. The contrast scaling distance K_D was varied from 10 to 180 ain steps of 10 to record data for varying contrast scaling. For this experiment, twenty objects ($n_{obj} = 20$) were set on each wall of the arena. The initial position and speed of the robot and the target, and the target's trajectory, were not changed from the previous experiment values. The phase of the target's sinusoidal motion along X-axis was alternately set to be either along or against the axis orientation. Ten sets of readings have been taken for each scenario.

The minimum distance between the robot and the target was recorded for all the scenarios for both HR and FD sensors. The minimum distance is defined as the smallest distance between the target and the robot during each simulation, which ends with a collision or when the robot goes ahead of the target (and thus can never catch it). Figure 5.4 shows a plot of the minimum distance between the robot and the target during the pursuit against the distance after which the contrast of an object was scaled. The vertical bars show the standard deviation over ten recordings. In the case of a nearly dark background (K_D small), both the HR detector and the FD sensor show a similar kind of behavior. But as soon as the contrast scaling distance was big enough to make the walls visible, the HR sensor was not able to track the target as is seen in the figure. The minimum distance reached by the HR sensor in these cases was nearly equal to the robot's minimum distance from the target when it moved with its "eyes" closed (no visual input). On the other hand, the response of the FD sensor was largely unaffected by contrast scaling and so it was able to track the target in all experimented scenarios. This shows that the FD sensor can outperform HR sensor based sensors for target tracking applications.

5.4.3 Experiment 3

In this experiment, the minimum distance between the robot and the target was computed for an increasing number of objects on each wall of the arena. The number of wall objects n_{obj} were varied from none to sixty in steps of two. The starting parameters for this experiment were similar to the first experiment. The contrast scaling factor was set to $C = 1$ for all the objects.

The simulation results are shown in Figure 5.5. The robot is able to easily track the target in the case of no background objects. However, when two objects are placed on each wall the minimum distance recorded by the robot increases. This is due to the fact that the robot's self motion produces an apparent motion for the fixed wall objects. Since there are very few background objects, the normalization by the FD sensor is weak and so the fixed objects also generate a weak turning response which reduces the overall sensor response leading to an increase in the minimum distance between the robot and the target. As the number of objects on the walls increases, the background becomes more balanced on both sides and the normalization suppresses the response due to fixed objects. This remains true till the objects on each wall reach forty-two. When the number of objects on each wall are between forty-two and fifty, the minimum distance the target reaches is greater than the collision threshold but still less than the minimum "eyes" closed distance. These cases are similar to the last condition observed in Experiment 1. The robot has no clue about the target for a very long time and when it is sufficiently close to the target, it starts to move towards it but fails to "collide". Once $n_{obj} > 50$, the target's edge information becomes undetectable by the sensor. As seen in the plot, this minimum distance is even more than the "eyes" closed minimum distance.

5.5 Summary

The simulation results show that the FD sensor based tracking algorithm is more robust than the HR sensor based algorithm. This algorithm might be used for visual tracking applications in cluttered environments. The modifications suggested in the FD algorithm make our modified algorithm robust for a larger number of scenarios. In the next chapter we will discuss the limitations in the design of

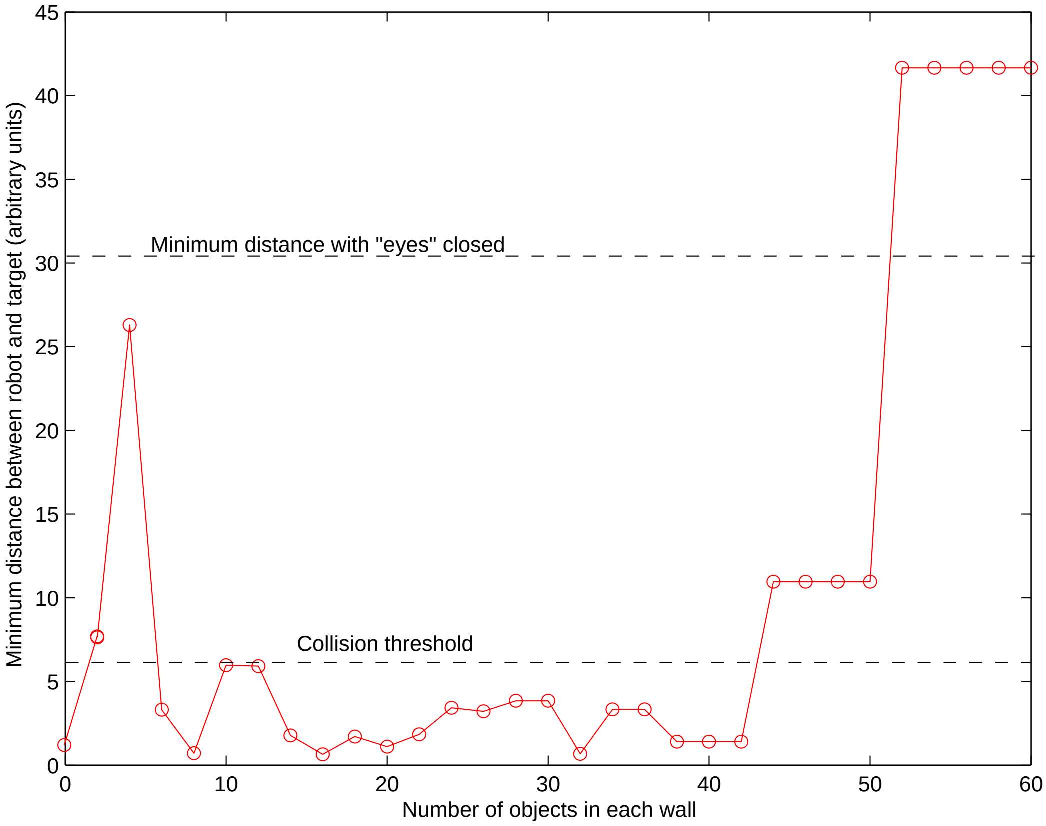


FIGURE 5.5. Simulation results of the minimum distance between robot and target with an increasing number of objects on each wall of the arena. The minimum distance increases for very few background objects, and reduces as a larger number of objects appear in the background. The minimum distance increases again when the number of objects on each wall becomes so high that the edge information of the target becomes undetectable. The collision distance has been set at 6 space-units. The minimum distance between the robot and the target when the robot is moving without any visual inputs is also shown.

the FD chip and circuit level improvements. A suggestion for the modular implementation of this system will also be discussed.

Chapter 6

Concluding Remarks

An FD-cell based small-field motion detection chip and a spiking-neuron sensor chip have been implemented and fabricated during the course of this study. These low-power, compact, real-time mixed-signal processing chips may be used to implement a modular motion sensing and target tracking system. A new algorithm has been proposed for target tracking agents. It accounts for the robot's own translatory motion as well as the rotatory motion required for tracking. There are a few modification required in the design of the FD chip to incorporate all the features of the new tracking algorithm. The FD chip and the spiking-neuron chip have certain limitations, a few due to design issues, and others simply because of the process variations during fabrication. The following sections discuss the limitations of this system, the circuit level improvements and the future extension of this work.

6.1 Discussion

In this section the system level limitations and modifications required to improve upon the design are presented.

6.1.1 HR Detector Design

The HR detector implementation (discussed in Chapter 3) is more compact and has fewer biases as compared to the implementation by Harrison and Koch (2000). The high pass filtering of the signal is left to the band-pass filter characteristics of the adaptive photoreceptor. The *Vadapt* bias of the photoreceptor can be used to control the adaptation time (see Figure 3.4). The response of the HR detector in the FD chip is a current output from a full wave rectifier circuit. This output is sensitive to the bias condition and the DC offset voltage levels of the photoreceptor and the low-pass filtered responses. The photoreceptor was biased such that it had a larger DC offset, and this provided sufficient voltage range for the bidirectional operation of the full wave rectifier circuit.

6.1.2 FD Model Implementation

The FD model implementation includes a normalization circuit that is an over-simplification of a rather complex shunting inhibition computation. In our VLSI implementation we have assumed the following:

$$\frac{v_i^+(t)}{P_l^+(t) + P_r^-(t)} \approx \frac{v_i^+(t)}{\beta + [P_r^{ccw}(t) + k^* \cdot P_r^{cw}(t)]^q}$$

This simplification assumes that the response from the background saturates and so there is no need to use the exponentiation done by the q ($q \approx 0.5$) operator. This factor was used in the original algorithm for matching electrophysiological data. It therefore has no critical influence on the characteristic response of the sensor. Its omission from the circuit simplifies our design. Moreover, the denominator has to be non-zero even in the absence of any motion in the visual field (the factor β ensures this in the right side equation). In the circuit implementation, in this condition, the bias transistor of the normalizer circuit goes into the triode region of operation and the response from individual small-field pixels becomes zero.

It seems from the initial characterization results that the range of voltage required by the full wave rectifier circuit for its normal operation is not available to it. This is because the rectifier circuit is connected to the NFET mirror circuit from the Gilbert multiplier, which has a small voltage range in which all the transistors remain in saturation. Thus, biasing of the full wave rectifier circuit heavily influences the HR detector and the small-field response from the chip. A solution to this is biasing the photoreceptor at a higher DC offset value. This, along with biasing of the Gilbert multiplier circuit such that small current outputs are generated allow this chip to function well within this range of bias conditions.

It has also been found that the vertical scanner circuitry in the FD chip has been laid out in flipped manner. This makes the vertical scanners not work and so we cannot get the output from the entire chip. In the next iteration of this chip, the layout will be corrected to make the vertical scanners work.

Also, the new tracking algorithm used for the target tracking simulations requires the output from the FD chips to have an amplitude response only, and not the direction. This modification can be done by using an absolute value circuit after the small field output has been generated. This modification will also be incorporated in the next iteration of this chip.

6.1.3 Spiking-neuron Sensor

In the design of the spiking-neuron sender chip, two types of integrate-and-fire (IF) circuits have been implemented in each pixel, P- and N-type. The P-type IF circuit takes negative current as input (or gives out positive current at its input node), while the N-type IF circuit draws in current. However, the current from the full-wave rectifier stage is negative for both the IF circuits. The leakage current has to be turned up high in order to make the N-type IF circuit work. Therefore, the $V_{leakneg}$ bias has to be continuously varied to get a proper response from the spiking pixel. This error in design makes the sender chip difficult to bias. The solution is as simple as using the P-type IF circuit for both the positive and negative currents. A more sophisticated way is to use another P-type mirror on top of the negative current and change its polarity to positive.

6.2 Future Work

The modifications required in the current FD chip will be incorporated and a revised design will be sent for fabrication soon. Also, it is advantageous to have a common front-end chip for various motion processing subunits that might be a part of an autonomous mobile agent. So, the future extension of this work will be the design of modular target tracking architectures as presented in the following subsections.

6.2.1 Sender-Receiver Tracking Architecture

A suggested implementation of a multi-chip FD-sensor based tracking system is discussed in this section. A sketch of such a sensor is shown in Figure 6.1. The spiking-neuron sender chip, described in Chapter 4, is appropriate as the front-end imaging sensor required for this modular system.

The motion computations will be carried out by the receiver chip. The receiver will employ a filter that integrates the spikes from the sender at each respective pixel position. This re-constructed signal will then be processed through all the respective stages of the FD-sensor as discussed in Chapter 3. The delayed signal in each pixel is created by a low-pass circuit. The Gilbert Multiplier circuit correlates the delayed and non-delayed signals, and two such sub-units correspond to a HR type movement detector. This response is further split by a full-wave rectifier circuit and is normalized using the normalization circuit. The individual and global small-field response can be computed by extracting the single pixel small-field output and the sum of all the pixel outputs, respectively. These compact processing blocks produce a better small-field response because of the potentially higher spatial resolution of the sender. Thus, this response from the FD-receiver chip will lead to a robust real-time tracking system. However, during the course of this study, such a system has

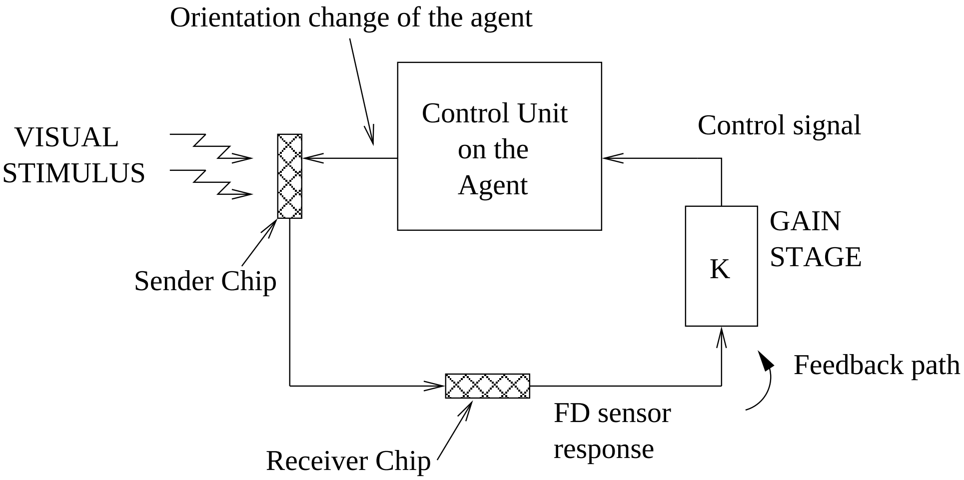


FIGURE 6.1. Modular Tracking agent with sender and receiver chips. The visual scene activates the sender chip that transmits the feature information over the high-speed AER bus. The FD receiver chip processes this information and produces a small-field response. This response is treated by a gain stage and is fed into a control system that steers the agent in the environment. This movement of the agent induces a change in the visual scene in front of the sender chip. Thus, it forms a feedback loop and tracks the moving target such that the response from the sender and receiver system is minimum.

not been implemented. This extension of the present design into a modular implementation of the tracking system is the future course for this research.

6.2.2 Gronenberg and Strausfeld Tracking System

This work could be extended to implement a more involved target-pursuit system based on the circuit by Gronenberg and Strausfeld (1991). This is a neural circuit explaining the sex-specific pursuit of the fly, and has been discussed in Section 2.5. This model explains the smooth pursuit behavior of insects based on the neural circuitry thought to be present in the acute zone of a male fly's eye.

Target-tracking is difficult to do in real-time by even the most sophisticated digital computers. Analog VLSI systems that process real world data in real-time provide a robust solution. A neuromorphic VLSI architecture based on Gronenberg and Strausfeld's model would be a reliable and high-speed tracking system. This model could be used to design a modular tracking sensor comprised of a high-density sender chip and a receiver to do the bulk of the computations. The sender chip will functionally resemble a spiking-neuron circuit as described in Chapter 3. As discussed in the previous section, the receiver chip of this modular tracking system is where the motion computations would take place. Contrast sensitive image information would be transferred to the receiver from the sender in the form of spike trains. The receiver would then do what is essentially a reconstruction of the information by utilizing the specific information about the position and contrast of the image, relayed by the AER bus. This conversion is necessary so that rest of the computations could be realized by utilizing fast custom analog circuits. The position sensitive response is allowed to generate what would be a yaw-torque response for the robot housing the sensor. The entire agent would move and turn as a single unit (here we differ from the Gronenberg and Strausfeld's model, which has a body yaw and a head yaw movement). The error-angle computation corresponds to the measurement of the position from the mid-line to where there is a discontinuity in the motion-field, suggesting a small object in front of the background. The basic design of the system would be similar to the work presented in this thesis. Only the position-detection circuit and weight multiplication according to the lateral position of the object circuit would be added. The implementation of such a position-detection circuit has been studied and published (Korrapati, 2001). The output from this sender-receiver system would control the yaw-turning signal of the robot. This robust system would present a more involved real-time multi-chip tracking system.

6.3 Summary

The study of biological algorithms and their VLSI implementation provides a road map to develop highly intelligent agents in the future. The FD cell based small-field circuit and the spiking neuron model have been implemented in this thesis. The use of the FD chip as a monolithic target-tracking sensor with simulations of a target tracking agent has been discussed. The need and importance of modular architectures with growing computations in the focal plane motivates the design of the multi-chip implementation of this sensor. The implementation of a spiking-neuron chip as the sender chip in a modular architecture has been described. The use of this contrast transducing sender chip transcends its role as a sender chip for a target tracking agent alone. It may be used as the front end for more than one visual motion processing unit on a robot. Further in this thesis, target tracking in a cluttered environment using the revised FD algorithm has been studied. The simulations show that our sensor detects and tracks the target even in cluttered environment, as compared to the fixation of a single stripe by Huber and Bülthoff's (1998) implementation. This FD cell based system provides a compact, efficient, and low-power solution for a complicated real-time tracking problem. The future of the tracking system design will be based on more detailed understanding of the bio-sensory system. It will provide useful insight for the development of more robust agents.

REFERENCES

Adelson, E. H. and J. R. Bergen (1985). Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* 2: 284–299.

Boahen, K. (1996a). *NSF Neuromorphic Engineering Workshop Report.* Telluride, CO.

Boahen, K. (1996b). Retinomorphic vision systems. In *Proceedings of the International Conference on Microelectronics for Neural Networks and Fuzzy Systems.* IEEE.

Boahen, K. (1999). A throughput-on-demand 2-D address-event transmitter for neuromorphic chips. In *Proc. of the 20th Conference on Advanced Research in VLSI*, Atlanta, GA.

Boahen, K. A. (2000). Point-to-point connectivity between neuromorphic chips using address events. *IEEE Trans. Circuits Syst. II-Analog Digit. Signal Process.* 47: 416–434.

Boahen, K.A. and A.G. Andreou (1992). A contrast sensitive silicon retina with reciprocal synapses. In Touretzky, D.S., M.C. Mozer, and M.E. Hasselmo, editors, *Advances in neural information processing systems*, Vol. 4. IEEE, MIT Press.

Buschbeck, EK and NJ Strausfeld (1996). Visual motion-detection circuits in flies: small field retinotopic elements responding to motion are evolutionarily conserved across taxa. *J. Neurosci.* 16: 4563–4578.

Cong, J., L. He, C. Koh, and P.H. Madden (1996). Performance Optimization of VLSI Interconnect Layout. *Integration, the VLSI Journal* 21: 1–94.

Cronin, J. (1987). *Mathematical aspects of Hodgkin-Huxley neural theory.* Cambridge University Press.

Deiss, S. R., R. J. Douglas, and A. M. Whatley (1998). A pulse-coded communications infrastructure for neuromorphic systems. In Maass, W. and C. M. Bishop, editors, *Pulsed Neural Networks*, chapter 6, pp. 157–178. MIT Press.

Delbrück, T. (1993a). Investigations of analog VLSI visual transduction and motion processing. Ph.D. dissertation, Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA.

Delbrück, T. (1993b). Silicon retina with correlation-based, velocity-tuned pixels. *IEEE Trans. Neural Networks* 4: 529–541.

DeWeerth, S., G. Patel, M. Simoni, D. Schimmel, and R. Calabrese (1997). A VLSI architecture for modeling intersegmental coordination. In *Proc. of the 17th conference on Advanced Research in VLSI*, Ann Arbor, MI.

DeWeerth, S.P. (1992). Analog VLSI circuits for stimulus localization and centroid computation. *International J. Computer Vision* 8: 191–202.

DeWeerth, S.P. and T.G Morris (1994). Analog VLSI circuits for primitive sensory attention. In *Proc. IEEE Int. Symp. Circuits and Systems*, Vol. 6, pp. 507–510. IEEE.

Douglass, JK and NJ Strausfeld (1995). Visual motion detection circuits in flies: Peripheral motion computation by identified small field retinotopic neurons. *J. Neurosci.* 15: 5596–5611.

Egelhaaf, M. (1985a). On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. II. Figure-detection cells, a new class of visual interneurons. *Biol. Cybern.* 52: 195–209.

Egelhaaf, M. (1985b). On the neuronal basis of figure-ground discrimination by relative motion in the visual system of the fly. III. Possible input circuitries and behavioral significance of the FD-cells. *Biol. Cybern.* 52: 267–280.

Egelhaaf, M. and A. Borst (1993a). A look into the cockpit of the fly - visual orientation, algorithms, and identified neurons. *J. Neurosci.* 13: 4563–4574.

Egelhaaf, M. and A. Borst (1993b). Motion computation and visual orientation in flies. *Comp. Biochem. Physiol.* 104A: 659–673.

Etienne-Cummings, R., J. Spiegel, and P. Mueller (1996). An visual smooth pursuit tracking chip. *Advances in Neural Information Processing Systems* 8: 706–712.

Exner, S. (1891). Die Physiologie der facettierten Augen von Krebsen und Insekten. *Deuticke* .

Exner, S. (1894). Entwurf zu einer physiologischen Erklarung der psychischen Frscheinungen. *I. Teil* pp. 37–140.

Franceschini, N., A. Riehle, and A. Le Nestour (1989). Directionally selective motion detection by insect neurons. In Stavenga, D. G. and R. C. Hardie, editors, *Facets of Vision*, chapter 17, pp. 360–390.

Furman, G.G. (1965). Comparison of models for subtractive and shunting lateral inhibition in receptor-neuron fields. *Kybernetik* 2: 257–274.

Gerstner, W. (1998). Spiking neurons. *Pulsed Neural Networks* pp. 261–295.

Gilbert, B. (1974). A high-performance monolithic multiplier using active feedback. *IEEE J. Solid-State Circuits* SC-9: 267–276.

Gilbert, B. (1975). Translinear circuits: A proposed classification. *Electronics Letters* 19: 14–16.

Gilbert, B. (1984). A monolithic 16-channel analog array normalizer. *IEEE J. Solid-State Circuits* 19: 956–963.

Gillespie, D. and J. Lazzarro (2003). Analog circuit schematic capture tool and simulation environment. `http://www.cs.berkeley.edu/~lazzaro/chipmunk/document/log`.

Gronenberg, W. and N. J. Strausfeld (1991). Descending pathways connecting the male-specific visual-system of flies to the neck and flight motor. *J. Comp. Physiol. A-Sens. Neural Behav. Physiol.* 169: 413–426.

Grossberg, S., G. Carpenter, E. Schwartz, E. Mingolla, D. Bullock, P. Gaudiano, A. Andreou, G. Cauwenberghs, and A. Hubbard (1997). Automated vision and sensing systems at Boston University. In *Proc. of the DARPA Image Understanding Workshop*, New Orleans, LA.

Harris, J.G., C. Koch, E. Staats, and J. Luo (1990). Analog hardware for detecting discontinuities in early vision. *Int. J. Comput. Vis.* 4: 211–223.

Harrison, R. and C. Koch (2000). A robust analog VLSI Reichardt motion sensor. *Analog Integrated Circuits and Signal Processing* 24: 213–229.

Hassenstein, B. and W. Reichardt (1956). Systemtheorische analyse der Zeit-, Reihenfolgen- und Vorzeichenauswertung bei der Bewegungsperzeption des Rüsselkäfers Chlorophanus. *Zeitschrift für Naturforschung* 11b: 513–524.

Hausen, K. and M. Egelhaaf (1989). Neural mechanisms of visual course control in insects. In Stavenga, D. G. and R. C. Hardie, editors, *Facets of Vision*, chapter 18, pp. 391–424.

Heisenberg, M. and E. Buchner (1977). The role of retinal cell types in visual behavior of dorsophila malanogaster. *J. Comp. Phsiol. A* 117: 127–162.

Higgins, C. M. and C. Koch (1999). Multi-chip neuromorphic motion processing. In *Proc. of the 20th Conference on Advanced Research in VLSI*, Atlanta, GA.

Higgins, C. M. and C. Koch (2000). A modular multi-chip neuromorphic architecture for real-time visual motion processing. *Analog Integrated Circuits and Signal Processing* 24(3).

Higgins, C.M. and S.K. Korrapati (2000). An analog VLSI motion energy sensor based on the Adelson-Bergen algorithm. In *Proceedings of the International Symposium on Biologically-Inspired Systems.*

Higgins, C.M. and S.A. Shams (2002). A biologically-inspired modular VLSI system for visual measurement of self-motion. *IEEE Sensors Journal special issue on Integrated Multi-Sensor Systems and Signal Processing* 2(6): 508–528.

Hildreth, E. C. and C. Koch (1987). The analysis of visual-motion - from computational theory to neuronal mechanisms. *Annu. Rev. Neurosci.* 10: 477–533.

Hodgkin, A. L. and A F. Huxley (1952). A quantitative description of membrane current and its application to conduction and excitation in a nerve. *J. of Physiol.* 117: 500–544.

Horiuchi, T., B. Bishofberger, and C. Koch (1994). An analog VLSI saccadic system. *Advances in Neural Information Processing Systems* 6: 582–589.

Horiuchi, T, T Morris, C Koch, and SP DeWeerth (1996). Analog VLSI circuits for attention-based, visual tracking. In *Advances in Neural Information Processing Systems.*

Huber, S.A. and H. H. Bülthoff (1998). Simulation and Robot Implementation of Visual Orientation behaviors of flies. *From animals to animats, proceedings of the fifth conference on the simulation of adaptive behaviour* 5: 77 85.

Indiveri, G., J. Kramer, and C. Koch (1996a). Parallel analog VLSI architectures for computation of heading direction and time-to-contact. In Touretzky, D.S., M.C. Mozer, and M.E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, Vol. 8, pp. 720–726, Cambridge, MA. MIT.

Indiveri, G., J. Kramer, and C. Koch (1996b). System implementations of analog VLSI velocity sensors. *IEEE Micro* 16: 40–49.

Indiveri, G., P. Oswald, and J. Kramer (2002). An adaptive visual tracking sensor with a hysteretic Winner-Take-All network. *IEEE* pp. 324–327.

Kalayjian, Z. and A. G. Andreou (1997). Asynchronous communication of 2D motion information using Winner-Take-All arbitration. *Analog Integr. Circuits Process.* 13: 103–109.

Khurana, B. and E. Kowler (1987). Shared attentional control of smooth eye movement and perception. *Vision Research* 27(9): 1603–1618.

Koch, C. and S. Ullman (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology* 4: 219–227.

Koch, C., H.T. Wang, R. Battiti, B. Mathur, and C. Ziomkowski (1991). An adaptive multi-scale approach for estimating optical flow: Computational theory and physiological implementation. *IEEE* –: 111–122.

Koch, Christof (1999). *Biophysics of Computation: Information Processing in Single Neurons.* Oxford University Press: New York.

Korrapati, S. (2001). An analog VLSI motion energy sensor and its applications in system level robotic design. Master's thesis, University of Arizona, Electrical and Computer Engineering Department.

Kowler, E., E. Anderson, B. Dosher, and E. Blaser (1995). The role of attention in the programming of saccades. *Vision Research* 35(13): 1897–1916.

Kramer, J., R. Sarpeshkar, and C. Koch (1997). Pulse-based analog VLSI velocity sensors. *IEEE Trans. Circuits Syst. II-Analog Digit. Signal Process.* 44: 86–101.

Kumar, N., W. Himmelbauer, G. Cauwenberghs, and A. G. Andreou (1998). An analog VLSI chip with asynchronous interface for auditory feature extraction. *IEEE Trans. on Circuit and Systems II* 45(5): 600–606.

Land, M. F. and R. D. Fernald (1992). The evolution of eyes. *Annu. Rev. Neurosci.* 15: 1–29.

Landolt, O., A. Mitros, and C. Koch (2001). Visual sensor with resolution enhancement by mechanical vibrations. *Proceedings 2001 Conference on Advanced Research in VLSI* pp. 249–264.

Lazzaro, J. and J. Wawrzynek (1997). Speech recognition experiments with silicon auditory models. *Analog Integr. Circuits Process.* 13: 37–51.

Liu, S.C. (1999). Silicon retina with adaptive filtering properties. *Analog Integrated Circuits and Signal Processing* 18: 1–12.

Maass, W. (2002). *Computation with spiking neurons.* The MIT Press (Cambridge), 2nd edition edition.

Mahowald, M. (1994). Analog VLSI chip for stereocorrespondence. In *Proc. IEEE Int. Symposium on Circuits and Systems*, Vol. 6, pp. 347–350.

Mahowald, M. and T. Delbrück (1989). Cooperative stereo matching using static and dynamic image features. In Mead, C. and M. Ismail, editors, *Analog VLSI Implementation of Neural Systems*, pp. 213–238. Kluwer Academic Publishers.

Mahowald, M and C. Mead (1991). The silicon retina. *Scientific American* 264(5): 76–82.

Mahowald, M.A. (1992). VLSI analogs of neuronal visual processing: a synthesis of form and function. Ph.D. dissertation, Department of Computation and Neural Systems, California Institute of Technology, Pasadena, CA.

Martin, A.J., S. M. Burns, T. K. Lee, D. Borkovic, and P.J. Hazewindus (1989). The design of an asynchronous microprocessor. *Advanced Reserch in VLSI: Proceedings of the Decennial Caltech Conference* pp. 351–373.

McCann, G.D. (1972). The fundamental mechanism of motion detection in the insect visual system. *Kybernetik* 12: 64–73.

McCann, G.D. and D.W. Arnett (1972). Spectral and polarization sensitivity of the Dipteran visual system. *J. Gen. Physiol.* 59: 534–558.

McCann, G.D. and J.C. Dill (1969). Fundamental properties of intensity, form and motion perception in the visual nervous system of calliphora phaenicia and musca domestic. *J. Gen. Physiol.* 53: 385–413.

Mead, C. A. and T. Delbrück (1991). Scanners for visualizing activity of analog VLSI circuitry. Technical report 11, California Institute of Technology, Department of Computation and Neural Systems.

Mead, C.A. (1989). *Analog VLSI and Neural Systems.* Addison-Wesley, Reading, MA.

Mitros, A. (2002). AER scheme discussions. Personal communication.

Moini, A. (1997). Vision chips or seeing silicon. Technical report 3rd ed., The Univ. of Adelaide, The Centre for High Performance Integrated Technologies and Systems.

Moore, A. and C. Koch (1991). A multiplication based analog motion detection chip. *Visual Information processing* 1473: 66–75.

Reichardt, W. and M. Egelhaaf (1988). Movement detectors provide sufficient information for local computation of 2-D velocity field. *Naturwissenschaften* 75: 313–315.

Reichardt, W., M. Egelhaaf, and A.K. Guo (1989). Processing of figure and background motion in the visual-system of the fly. *Biol. Cybern.* 61: 327–345.

Reichardt, W., T. Poggio, and K. Hausen (1983). Figure-ground discrimination by relative movement in the visual system of the fly. II. towards the neural circuitry. *Biol. Cybern.* 46: 1–30.

Stein, R. B. (1967). The frequency of nerve action potential generated by applied currents. *Proc. R. Soc. London, B* 167: 64–86.

Tanner-EDA (2003). L-Edit: The Layout Editor. A division of Tanner Research, Inc.

Torre, V. and T. Poggio (1978). Synaptic mechanism possibly underlying directional selectivity to motion. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 202: 409–416.

Van Santen, J. P. H. and G. Sperling (1985). Elaborated Reichardt detectors. *J. Opt. Soc. Am. A-Opt. Image Sci. Vis.* 2: 300–321.

Venier, P., A. Mortara, X. Arreguit, and E. Vittoz (1997). An integrated cortical layer for orientation enhancement. *IEEE Journal of Solid State Circuits* 32(2): 177–186.

Zigmond, J. Michael, E.F. Bloom, C.L. Story, J.L. Roberts, and L.R. Squire (1999). *Fundamental Neuroscience* .