

國立交通大學

電子工程學系電子研究所

碩士論文

仿效兔子視網膜開持續中樞神經細胞之
視覺神經晶片



**A NEUROMORPHIC CHIP THAT IMITATES THE
ON SLUGGISH SUSTAIN GANGLION CELL SET IN
THE RETINA OF RABBITS**

研究生：許筱妊

指導教授：吳重雨 教授

中華民國 九十七 年 八 月

仿效兔子視網膜開持續中樞神經細胞之視覺神經晶片

**A NEUROMORPHIC CHIP THAT IMITATES THE ON SLUGGISH
SUSTAIN GANGLION CELL SET IN THE RETINA OF RABBITS**

研究生：許筱妊

Student: Xie-Ren Hsu

指導教授：吳重雨 教授

Advisor: Chung-Yu Wu



Submitted to Institute of Electronics

College of Electrical Engineering and Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Electronic Engineering

August 2008

Hsin-Chu, Taiwan, Republic of China

中華民國九十七年八月

仿效兔子視網膜開持續中樞神經細胞之視覺神經晶片

研究生：許筱妊

指導教授：吳重雨 老師

國立交通大學

電子工程學系 電子研究所碩士班

摘要

哺乳類動物視網膜由五種不同的細胞構成，而每一種細胞又各自有不同的分類與功能。除了將影像世界的訊息轉變為神經訊號外，視網膜還負責 0.3% 的大腦視覺處理功能。在本研究中，我們嘗試以工程的角度去瞭解視網膜的運作功能。

於本論文中，我們提出了一個新的設計方法用以設計金氧半場效電晶體神經型態晶片，此神經型態晶片依據由哺乳類動物視網膜建立之生物模型所設計的仿視網膜矽晶片，模擬了兔子視網膜的開持續中樞神經節細胞組；文中詳論其特性與工作原理。

此晶片包含一個大小為 32x32 的影像感測陣列，可仿效兔子視網膜開持續中樞神經細胞的功能。陣列中包含了許多相同的基本單元，而每個基本單元中包含了光輸入元件、感光細胞、水平細胞、開兩極細胞、關兩極細胞、Amacrine 細胞與中樞神經細胞元件。晶片中也包括了輸入輸出接墊 (I/O pads) 的靜電防護、位址解碼 (address decoder) 及訊號讀出電路。

最後，我們比較原本生物模型的結果與電路模擬及晶片的量測結果之異同，檢驗此晶片的正確性與待改進之處。量測結果顯示，該金氧半場效電晶體神經型態晶片在時域及空間域的特性與生物量測結果相符，驗證了該晶片的仿生功能。經由實驗驗證，這個設計方法所設計的神經型態晶片能協助尚未揭露的視網膜細胞行為與視覺語言，並且也可用以設計所有的視網膜神經節細胞組，將視網膜功能完全由晶片實現。此外，該研究結果也使得極具潛力的視網膜晶片應用，諸如運動偵測、電腦視覺、人工視網膜和生醫元件方面變的更有可行性。

A NEUROMORPHIC CHIP THAT IMITATES THE ON SLUGGISH SUSTAIN GANGLION CELL SET IN THE RETINA OF RABBITS

Student: Xie-Ren Hsu

Advisor: Chung-Yu Wu

Department of Electronics Engineering & Institute of Electronics

National Chiao Tung University



ABSTRACT

The mammalian retina is comprised of five different kinds of cells, for each kind can be divided into more types. Besides transducing the visual world into neural signals, these cells are in charge of 0.3% of the visual function of the brain. In this research, we try to understand the retinal functions from the engineering point of view.

In this thesis, a new design methodology is proposed to implement CMOS neuromorphic chips which imitate the ON sluggish sustain ganglion cell set of rabbits' retinas. A retinal chip based on the biological model derived from the mammalian retina is proposed.

The retinal chip contains a focal plane sensory array of 32x32 similar basic cells that perform the functions. Each basic cell contains photo-input, photoreceptor, horizontal cell, on bipolar cell, off bipolar cell, amacrine cell and ganglion cell. ESD protection circuit for I/O pads, address decoders and readout circuits are also included in this chip.

Finally, the results of HSPICE simulation and chip measurement are compared with that of the original model to examine the consistency and the difference for further improvement. The

measurement results on the fabricated CMOS neuromorphic chip are consistent with the biological measurement results. Thus the biological functions of the chip have been successfully verified. It can be used to understand more biological behaviors and visual language of retina under different input optical images which have not yet been tested in biological experiments. Based on the results, the full ganglion cell sets of retina can be designed. Otherwise, many potential applications of retinal chips on motion sensors, computer vision, retinal prosthesis, and biomedical devices are feasible.



ACKNOWLEDGEMENT

感謝指導教授吳重雨博士，給予我能夠受到這個環境薰陶的機會。一年半來耐心的指導與鼓勵，使我能順利完成碩士學位。在吳教授循序漸進的諄諄教誨下，前瞻的見解與淵博的專業知識，給予我指導與啟發，讓我不僅獲得許多積體電路設計的專業知識，更學習到研究時應有的挑戰困難及解決問題的態度與方法，面對問題時如何著手構思和勇於嘗試，使我得以一點一滴地累積自己的實力，雖然在過程倍感艱辛，但卻獲益良多。此外，我要特別感謝林例如教授和楊文嘉學長，在研究上所給予的教導與協助，並給予我寶貴的建議與指導，真的很感謝你們。祝福你們研究順利、事事順心，也祝福學長早日畢業。

感謝307實驗室的王文傑學長、虞繼堯學長、陳勝豪學長、黃祖德學長、盧台祐學長、黃鈞正學長、蘇烜毅學長、蔡夙勇學長、陳旻琰學長在研究上所給予的指導與協助，給予我指點和研究上的經驗傳承。同時也感謝所有碩班的學長姐與同學。感謝國忠、順哥、萬謙、維德、豪哥、松諭、巧玲、威宇、建名、期聖、國權、政邦、亭州、區文、世範、宗恩、紹岐、柏硯、歐陽、昕華、國維、晟佑等學長姐的諸多幫助。還要謝謝惠雯、佳琪、怡歆、家菱、炯為、宗昀、韋丞、順天、暄泰、彥良、哲倫、世昕、凱悌、彥璋、育祥、詠儒以及所有新進來的學弟妹們一起奮鬥玩樂所建立那起無堅不摧的友誼情感，在生活上所帶來的歡樂與相互幫忙。謝謝你們一直以來的陪伴與關懷。

此外，特別感謝當初帶我進入電路設計領域的大學專題教授：翁若敏博士。感謝老師的啟蒙與教導，平日的叮嚀與教誨，一直以來的提攜與指引、支持與鼓勵，讓我得以在電路設計領域紮下穩固的基礎進而深入研究。如師亦友的相處，也為冷淡的生涯平添華麗且優美樂章！更教導了人生旅途上遇到的挫折該如何面對，使得我在專業知識增長的同時，更學習到為人處世的珍貴態度。在此獻上最誠懇的謝意。

最後感謝我的父親許清莖先生、母親張淑媛女士，感謝你們無怨無悔、永無止境的付出、照顧、支持與鼓勵，給予我最溫暖的關心，你們是我學習過程中最大的後盾，衷心的感謝你們。在此，僅以這份小小的成果與你們分享，並致上深深的感謝之意。

謹以此論文獻給所有關心我的人，並祝福我的師長、家人、朋友及學弟妹們事事順心、身體健康。我在撰寫的過程中，雖力求嚴謹，然誤謬之處，在所難免，尚祈各位讀者賜予寶貴意見，使本論文能更加完善。

許筱妊
誌於 風城交大
九十七年 盛夏

CONTENTS

| | |
|---------------------------------|-------------|
| ABSTRACT (CHINESE) | i |
| ABSTRACT (ENGLISH) | ii |
| ACKNOWLEDGEMENT | iv |
| CONTENTS | v |
| TABLE CAPTIONS | vii |
| FIGURE CAPTIONS | viii |

| | |
|--|----------|
| CHAPTER 1 Introduction | 1 |
| 1.1 BACKGROUND..... | 1 |
| 1.1.1 <i>Morphology and physiology of retina</i> | 1 |
| 1.1.2 <i>The classification of ganglion cells of the rabbits' retina</i> | 4 |
| 1.2 OPERATIONAL PRINCIPLE – FROM THE ENGINEERING POINTS OF VIEW..... | 5 |
| 1.3 MOTIVATIONS..... | 8 |
| 1.4 MAIN RESULTS..... | 9 |
| 1.5 THESIS ORGANIZATION..... | 10 |

| | |
|---|-----------|
| CHAPTER 2 The CNN Model of the ON Sluggish Sustain Ganglion Cell | 15 |
| 2.1 NEUROMORPHIC MODEL OF THE RETINAL CELL SET..... | 15 |
| 2.2 MACROMODEL FOR CHIP IMPLEMENTATION..... | 17 |

| | | |
|-------------------|--|-----------|
| 2.3 | SIMULATION RESULTS..... | 18 |
| CHAPTER 3 | Circuit Design..... | 26 |
| 3.1 | ARCHITECTURE OF A BASIC CELL OF THE RETINAL CHIP..... | 26 |
| 3.2 | CIRCUIT DESIGN OF THE BASIC CELL..... | 27 |
| 3.2.1 | <i>Photoreceptor and Horizontal cell.....</i> | <i>27</i> |
| 3.2.2 | <i>On bipolar cell.....</i> | <i>28</i> |
| 3.2.3 | <i>Off bipolar cell.....</i> | <i>29</i> |
| 3.2.4 | <i>Amacrine and Ganglion cells.....</i> | <i>34</i> |
| 3.2.5 | <i>Impacts of device leakage and mismatch.....</i> | <i>34</i> |
| 3.3 | ARCHITECTURE OF 2-D RETINAL ARRAY..... | 35 |
| 3.4 | SIMULATION RESULTS..... | 35 |
| CHAPTER 4 | Layout Descriptions and Experimental Results..... | 53 |
| 4.1 | LAYOUT DESCRIPTIONS..... | 53 |
| 4.2 | EXPERIMENTAL ENVIRONMENT..... | 54 |
| 4.3 | EXPERIMENTAL RESULTS..... | 55 |
| CHAPTER 5 | Conclusions and Future Work..... | 72 |
| 5.1 | CONCLUSIONS..... | 72 |
| 5.2 | FUTURE WORK..... | 73 |
| REFERENCES | | 74 |



TABLE CAPTIONS

CHAPTER 1

Table I Ganglion cell types, defined by response to spot sizes and level of dendritic stratification [13].....11

CHAPTER 3

Table II The transistor's sizes of circuits in Fig. 3. 2(a)-(d).....39

CHAPTER 4

Table III The bias condition to measure the spatiotemporal patterns of Fig. 4. 7.....58

Table IV Summary of characteristics of the proposed 32x32 retinal chip.....69



FIGURE CAPTIONS

CHAPTER 1

Fig. 1.1 Schematic of the mammalian retina. Reprinted from [1].....12

Fig. 1.2 The major cell types of a typical mammalian retina. From the top row to the bottom, photoreceptors, amacrine cells and ganglion cells. Reprinted from [2].....12

Fig. 1.3 Bipolar cell types of the primate retina. Reprinted from [1].....13

Fig. 1.4 Spiking patterns in each row (measured from four different members of the same cell class) for five different classes of ganglion cells. The numbers at the beginning of each row refer to the classification shown in Table 1. 1. The arrows in row 5 (cell type 7) point to the enhanced activity at the stimulus edge, a feature that is characteristic of this class [13].....13

Fig. 1.5 A sketch of the processing structure of the CNN model of the mammalian rabbit retina [14].....14

CHAPTER 2

Fig. 2.1 The neuromorphic model of the ON sluggish sustain ganglion cell set of rabbits' retinas. The parameter τ denotes the time constant in millisecond of the LPF. The parameter D denotes the space constant, which is defined by the laterally diffusing range in cell number in a 180-cell array.....21

Fig. 2.2 The RC equivalent model of the ON sluggish sustain ganglion cell set of rabbits' retinas.....21

Fig. 2.3 The macromodel of a single pixel of the implemented chip.....22

Fig. 2.4 The HSPICE model-simulated time domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These waveforms are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 15th to the 20th pixel at time point 1001 to 2000.....23

Fig. 2.5 The HSPICE model-simulated space domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These waveforms are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 15th to the 20th pixel at time point 1001 to 2000.....24

Fig. 2.6 The HSPICE model-simulated time domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These waveforms are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 15th to the 20th pixel at time point 1001 to 2000. The waveform at the left of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. The $V_{sm} = 2.2V$25

CHAPTER 3

- Fig. 3.1** The block diagram of a single pixel.....40
- Fig. 3.2** (a) The circuit of the photoreceptor and horizontal cell (PH1, PH2, H); (b) The circuit of ON bipolar cell (ONBIP); (c) OFF bipolar cell (OFFBIP); (d) The circuit of amacrine and ganglion cells (Ama, GC).....43
- Fig. 3.3** (a) A linearized transconductor using floating voltages sources; (b) Flipped Voltage Follower implementation; (d) Transconductor using two identical FVFs to implement the floating batteries.....44
- Fig. 3.4** (a) The whole chip architecture of the implemented chip, and (b) the implementation of smoothing networks.....45
- Fig. 3.5** The schematic of the address decoder used in the chip (a) column address decoder; (b) row address decoder; (c) buffer; (d) AND gate.....48
- Fig. 3.6** The monte carlo analysis of OFF bipolar cell with (a) the original V-I converter structure (b) the modified linear transconductor structure.....49
- Fig. 3.7** The HSPICE circuit-simulated time domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. The light stimulus is applied at 1000th~2000th micro-second and $V_{sm} = 2.2V$50
- Fig. 3.8** The HSPICE circuit-simulated space domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. The light stimulus is applied at 15th~20th pixel and $V_{sm} = 2.2V$51
- Fig. 3.9** The HSPICE circuit-simulated spatiotemporal patterns for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f)

ganglion cell. These waveforms are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 15th to the 20th pixel at time point 1001 to 2000. The waveform at the left of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. $V_{sm} = 2.2V$52

CHAPTER 4

| | | |
|-----------------|--|----|
| Fig. 4.1 | Layout of the whole retinal chip. Main parts of the chip are labeled. Sensory array is in the center, the row and column address decoders are in the upper and left of the array respectively..... | 59 |
| Fig. 4.2 | Layout and floorplan of a basic cell of the sensory array. The six regions of main parts of the basic cell and the phototransistor are labeled..... | 60 |
| Fig. 4.3 | Photograph of the whole retinal chip. Main parts of the chip are labeled. Sensory array is in the center, the row and column address decoder are in the upper and left of the array respectively..... | 61 |
| Fig. 4.4 | Photograph of a basic cell. Only the photo-BJT region is not covered by metal..... | 61 |
| Fig. 4.5 | The read-out circuit to translate the output current into voltage..... | 62 |
| Fig. 4.6 | The measurement setup diagram..... | 62 |
| Fig. 4.7 | Comparison of the chip measured transient response to the model simulation results and circuit simulation results. (a) photoreceptor, (b) horizontal cell. The time units for model simulation and circuit simulation are the same, $17\mu s/frame$; time unit for chip | |

measurement is 20 μ s/frame. The light stimulus at 170th~340th frame and $V_{sm}=2.4V$63

Fig. 4.8 Comparison of the chip measured transient response to the model simulation results and circuit simulation results. (a) on bipolar cell, (b) off bipolar cell. The time units for model simulation and circuit simulation are the same, 17 μ s/frame; time unit for chip measurement is 20 μ s/frame. The light stimulus at 170th~340th frame and $V_{sm}=2.4V$64

Fig. 4.9 Comparison of the chip measured transient response to the model simulation results and circuit simulation results. (a) amacrine cell, (b) ganglion cell. The time units for model simulation and circuit simulation are the same, 17 μ s/frame; time unit for chip measurement is 20 μ s/frame. The light stimulus at 170th~340th frame and $V_{sm}=2.4V$65

Fig. 4.10 Comparison of the chip measured spatial response to the model simulation results and circuit simulation results. (a) photoreceptor, (b) horizontal cell. The light stimulus at 15th~21th pixel and $V_{sm}=2.4V$66

Fig. 4.11 Comparison of the chip measured spatial response to the model simulation results and circuit simulation results. (a) on bipolar cell, (b) off bipolar cell. The light stimulus at 15th~21th pixel and $V_{sm}=2.4V$67

Fig. 4.12 Comparison of the chip measured spatial response to the model simulation results and circuit simulation results. (a) amacrine feed-forward cell, (b) ganglion cell. The light stimulus at 15th~21th pixel and $V_{sm}=2.4V$68

Fig. 4.13 Measured transient response with light stimulus at time point 200msec~400msec. Curve (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. Under bias condition: $V_{sm}=2.4V$69

Fig. 4.14 The measured spatiotemporal patterns for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These patterns are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 19th to the 23rd

pixel at time point 1001 to 2000. The waveform at the left of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. Under bias condition: $V_{sm}=2.4V$70





CHAPTER 1 Introduction

1.1 BACKGROUND

1.1.1 Morphology and physiology of retina

The retina is a thin sheet of brain tissue (100 to 250 μ m thick) that grows out into the eye to provide neural processing for photoreceptor signals. It includes both photoreceptor and the first two to four stages of neural processing. Its output projects centrally over many axons, and analysis of these information channels occupies about half of the cerebral cortex. Moreover, it comprises about 75 discrete neuron types connected in specific, highly stereotyped patterns. It sends different 'images' of the outside world to the brain – an image of contours (line drawing), a color image (watercolor painting) or an image of moving objects (movie). This is commonly referred to as parallel processing and starts as early as the first synapse of the retina, the cone pedicle. The schematic of the mammalian retina is shown in Fig. 1.1 [1]. There are six classes of neuron in the mammalian retina: rod (1), cones (2), horizontal cells (3), bipolar cells (4), amacrine cells (5), and retinal ganglion cells (RGCs) (6). They have a laminar distribution (OS/IS, outer and inner segments of rods and cones; ONL, outer nuclear layer; OPL, outer plexiform layer; INL, inner nuclear layer; IPL, inner plexiform; GCL, ganglion cell layer; NFL, optic nerve fiber layer). Rods and cones together are labeled as photoreceptors.

As the synaptic terminals of rods and cones, the light-evoked signals are transferred onto bipolar and horizontal cells. Horizontal cells, of which there are between one and three types in mammalian retinas, provide lateral interactions in the outer plexiform layer. One type of rod bipolar cell and at least nine types of cone bipolar cell transfer the light signals into the inner plexiform layer (IPL), onto the dendrites of amacrine and ganglion cells. Cone bipolar cells fall into two main groups: ON and OFF bipolar cells. Amacrine cells are inhibitory interneurons, and there are as many as 50 morphological types. Ganglion cell dendrites collect the signals to the visual center of the brain. At least 10-15 morphological types of ganglion cell are found in any mammalian retina. The

major cell types of a typical mammalian retina are shown in Fig. 1.2 [2]. The cells are introduced below:

Photoreceptor:

The photoreceptor mosaic is optimized to cover the full range of environmental light intensity (10¹⁰). This design specification requires two types of detected with different sensitivities, the rod and the cone. The rod serves under starlight where photons are so sparse that over 0.2 second (the rod integration time), they cause only ~1 photoisomerization (R^*)/10,000 rods. Consequently, under starlight and for 3 log units brighter, a rod must give a binary response reporting over each integration time the occurrence of either 0 or 1 R^* . The rod continues to serve at dawn as photons arrive more densely, providing more than one R^* /integration time. The rod sums these linearly up to 20 R^* /integration time and then gradually saturates, with 100 R^* evoking a maximal photocurrent of ~20pA.

The cone serves under full daylight, beginning when photon density reaches ~100 photons/receptor/integration time. The cone actually absorbs and transduces single photons, but because its gain is 50-fold lower than the rod's, it requires 100 R^* for the signal to rise above the continuous dark noise. By 1000 R^* /integration time, when rods are nearly saturated, the cone responds strongly. The cone photocurrent saturates at 30~pA (similar to the rod), but this occurs at much higher intensities up to 10⁶ R^* /cone/integration time. Consequently, whereas the rod signal is at first binary and then graded, but always corrupted by photon noise, the cone signal is always graded and far less noisy. Thus, the so-called photoreceptors in silicon retinas are actually cones.

Horizontal cell:

Horizontal cell dendrites are inserted as lateral elements into the invaginating contacts of cone pedicles, and horizontal cell axon terminals form the lateral elements within rod spherules. It is assumed that horizontal cells release the inhibitory transmitter and provide feedback inhibition at the photoreceptor synaptic terminal. As horizontal cells summate light signals from several cones, such feedback would cause lateral inhibition, through which a cone's light response is reduced by the

illumination of neighboring cones. This mechanism is thought to enhance the response to the edges of visual stimuli and to reduce the response to areas of uniform brightness. There is also evidence that light-dependent release of GABA from horizontal cells provides feed-forward inhibition of bipolar cell dendrites. Irrespective of their precise mode of action, horizontal cells sum light responses across a broad region, and subtract it from the local signal. Because horizontal cells are coupled through gap junctions, their receptive fields can be much wider than their dendritic fields.

Bipolar cells:

Bipolar cell types of the primate retina are shown in Fig. 1.3 [3]. Their axons terminate at different levels in the IPL; those terminating in the outer half are putative OFF cone bipolar cells, and those terminating in the inner half are ON bipolar cells. The axons of OFF and ON cone bipolar cells terminate at different levels (strata) within the IPL: OFF in the outer half, ON in the inner half. However, superimposed on this ON/OFF dichotomy, further bipolar cell types have been described in Fig. 1.3 and every mammalian retina that has been studied contains at least four types of OFF and four types of ON cone bipolar cell [4], [5]. We are just beginning to understand their functional roles [6]. Axons that carry more transient OFF light signals terminate in the middle of the IPL; those that carry sustained OFF light signals are found in a more peripheral position [7], [8].

Amacrine cells:

There are twenty-nine types of amacrine cells. All retinal ganglion cells receive input from cone bipolar cells, but most direct synapses on the ganglion cells are from amacrine cells. The exact fraction varies among different functional types of ganglion cells, ranging from roughly 70% for alpha cells (large, movement-sensitive ganglion cells found in most mammals) to 50% for the midget ganglion cells located in the monkey central fovea. Amacrine cells also make inhibitory synapses on the axon terminals of bipolar cells, thus controlling their output to ganglion cells. In contrast to horizontal cells, which have a single broad role, amacrine cells have dedicated functions – they carry out narrow tasks concerned with shaping and control of ganglion cell responses. The different amacrine cells have distinct pre- and postsynaptic partners, contain a

variety of neurotransmitters. Amacrine cells seem to account for correlated firing among ganglion cells. Shared input from a common amacrine cell will tend to make ganglion cells fire together [2].

Ganglion cells:

The retinal ganglion cells summarized signals of previous retinal cells, and then send neural spiking to the brain. They have been found to diverse in both stratification and physiological properties. There are at least 10-15 different morphological types of ganglion cell in any mammalian retina [3], [9]. The underlying belief is that cells with distinct morphologies have distinct physiological functions. It was thought that they represented feature detectors that react to specific light stimuli. Among them were direction selective ganglion cells, which respond to light spots that move in a certain direction across their receptive field. In the primate retina, two types of concentrically organized receptive field have been found. One type showed no chromatic receptive field organization, whereas in the other type the centre and surround were chromatically selective [10]. The detailed classification of ganglion cells is described in section 1.1.2.

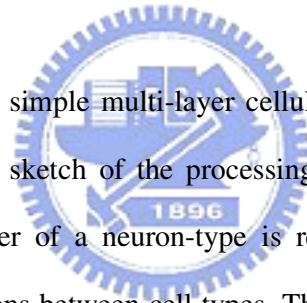
1.1.2 The classification of ganglion cells of the rabbits' retina

The physiological classification of ganglion cells has been the most detailed in rabbit of any mammalian species. So the classification of ganglion cells of the rabbit is introduced here. Many neuroscientists have made classifications of ganglion cells. Although there are areas of disagreement all neuroscientists confirm this diversity. Morphologically, the retinal ganglion cells are divided into two types of concentrically organized receptive fields, one with a small, linearly summing receptive field center (X cell) and another with a large, non-linear responsive area (Y cell). In the cat, the correspondence between X-cells and β , and Y-cells and α was established long ago, as was the analogous match between P and M, midget and parasol cells in the monkey [2]. Physiologically, the retinal ganglion cells are divided into ON center/OFF surround and OFF center/ON surround cells [1].

In the thesis, the classification of [11], [12] is adopted. Here the ganglion cells are divided into

with concentric receptive fields and with complex receptive fields. Concentric ganglion cells were those that had ON or OFF centers with antagonistic surrounds and were classified into different group by extracellular recordings of their ON- or OFF-center response sign, excitatory receptive field center size, linearity of spatial summation, and brisk vs. sluggish and transient vs. sustained response to step changes in light intensity [11]. Ganglion cells that had complex receptive field properties are ON-OFF and ON direction-selective cells, orientation-selective cells, local edge detectors and uniformity detectors. Cells were first classified by their characteristic extracellular response to manually controlled stimuli similar to those which have been used in previous in vivo studies [12]. The recorded cell types are presented in Table 1.1 [13]. In [13], space-time patterns of concentric ganglion cells are reported. Spiking patterns of five different classes of ganglion cells are shown in Fig. 1.4. We can see in Fig. 2.4 that even for the same cell, the space-time patterns vary in every record.

Based on the results of [13], a simple multi-layer cellular neural /non-linear network (CNN) model was built [14]. Fig. 2.5 is a sketch of the processing structure of the CNN model of the mammalian rabbit retina. Each layer of a neuron-type is represented by a horizontal line. The vertical lines represent the connections between cell types. The third row contains the retina output cells, where the names are the neurobiologic names and their positive input is the bipolar layer, and their negative inputs are amacrine FF layers. In chapter 2, the ON Sluggish sustain ganglion cell is chosen to verify the design methodology proposed in this thesis.



1.2 OPERATIONAL PRINCIPLE – FROM THE ENGINEERING POINTS OF VIEW

The instantaneous dynamic range of the photoreceptors spans only about 1000:1, but the mean level around which this instantaneous range operates can move, facilitated by gain adjustments in the photoreceptors themselves, over a range of 10^6 to 1 or more, from twilight to bright sunlight. There is a range of intensities of about 10^3 :1 near twilight where the cone photoreceptors do not

operate. In this range the visual message is conveyed through the rod system that hijacks the cone circuitry within the retina. About this intensity range the cones predominate. The switch is obvious to us: it is the intensity at which we begin to see color.

Photoreceptor activity is passed to the next processing layer, the horizontal cells. These cells are strongly interconnected so they form a broad diffusion layer. Each horizontal cell receives input from a local population of photoreceptors, and because of the interconnections, the horizontal cell layer acts to diffuse the image, so the visual image appears to be quite blurred with a space constant of a few degrees. Horizontal cell activity at any point represents a local spatial (and temporal) average of neural activity.

The horizontal cells “feed back” with a sign-inverting phase to the cone photoreceptors to modify the representation generated by the transduction component of the photoreceptors. The range of neural activity in the feedback-enhanced image is reduced so that small incremental changes around the mean are well-represented, while changes far from the mean level are suppressed. The feedback of a blurred image has an effect similar to that used by early photoreceptors to enhance the edges of an image. They developed a positive transparency on top a slightly blurred negative transparency. The negative feedback from horizontal cells to cones accomplishes a similar effect, edge-enhancing the resulting visual representation.

This interaction is “read out” by the bipolar cell. These representations have been enhanced in a variety of ways. They are gain adjusted to fit the ambient level of the original scene, and then mean adjusted by the horizontal cell network, and the features of the scene are edge-enhanced as a results of the interaction of the sharp and blurred images. Bipolar cells exist in two main varieties: those that represent (by activity moving back into the plane) the brightness of the image, and those that represent (by activity moving back into the plane) the darkness of the image. These two representations are useful because the visual system is best at identifying small changes in intensity above and below the mean level, and having two separate systems, each tuned to one of these representations appears to be an effective implementation of this phenomenon.

The nature of synaptic connections in the nervous system is such that it favors transmission of increases in activity (out of the plane) over decreases in activity (into the plane). Consequently, the signals leaving the two bipolar cell arrays are “rectified” in that the (out of the plane) activity is accentuated and the activity into the plane is lost. These two representations of the visual world are now presented to the next layer of processing for further enhancement of the visual representation. This is the region where the dozen movies of the visual representation. This is the region where the dozen movies of the visual world are generated. Each of the representations is “filtered” in space and time so that numerous representations, each with a different space or time constant, are generated.

Then different space–time components interact with each other to generate a further enhanced representation. These interrelationships between different space-time representations are implemented by a second set of interneurons, the amacrine cells in the inner retina. Unlike the horizontal cells discussed above that form a highly uniform interconnected. Each seems to receive input from a highly uniform interconnected system, there is great diversity in the amacrine cell population, and they are not strongly interconnected. Each seems to receive input from one or more space-time representations, delivered by the bipolar cells, and communicate by inhibiting one or more of the other space-time representations.

If we think of each space-time representation as an extracted “visual feature”, then the amacrine cells form a set of interactions between features. One way to think about this would be to imagine that each feature is broadcasting, via the amacrine cells, to turn down its representation in other feature detectors, and simultaneously, all other feature detectors, via the amacrine cells are turning down their specific feature quality in adjacent features. This interaction represents a form of lateral inhibition, not in physical space but in the abstract space of the features themselves.

1.3 MOTIVATIONS

The previously described model is proposed for the capability of being implemented on CNN-UM. The CNN-UM is a cellular analog stored-program multidimensional array computer with distributed local analog and logic memory, analogic (analog and logic) computing units, as well as local and global communication and control unit [15]. A complex CNN-UM chip is needed to implement the previous mentioned multilayer model. If a complex multilayer CNN-UM is not available, iterative operations are needed to complete the whole processing of the model, and thus real time processing could not be accomplished.

The motivation of this thesis is to realize CMOS neuromorphic chips which imitate the ON sluggish sustain ganglion cell set of rabbits' retina with the modified circuit to obtain better performance. Furthermore, based on the IC realization of ON brisk transient and the ON sluggish sustain ganglion cell set of rabbits' retina, every kind of ganglion cell sets can be implemented and integrated similarly to form the whole silicon retina.

It is important to duplicate successfully the retinal functions, channels, and visual language on silicon chip because of the key advantages this might provide. First, it could help neuroscientists to understand retinal functions and visual language. Since biological experiments can only be performed on a very limited number of cells, it is very difficult to see the global spatiotemporal features of retinal cells using this method alone. Second, it could provide valuable clues concerning neural activities in the visual cortex and thus move a few more steps toward the discovery of the visual processing of the brain. Third and finally, duplication could enable important applications in the areas of intelligent visual sensor systems and retinal prostheses. Since the neuromorphic multilayer CNN model [14], [16]-[17] shed light to understand the role of different circuits and layer parameters. It is feasible to implement the complete retinal circuitry with nowadays VLSI technology.

In this thesis, a retinal chip that realizes the On Sluggish sustain ganglion cell of the CNN model is proposed. The model is slightly simplified for chip design. Image sensing devices are the

interfaces between CNN-UM and imager are no more needed.

1.4 MAIN RESULTS

Based on the previous work on ‘ON brisk transient’ ganglion cell set [20], we developed a novel retinal chip that imitates the ON sluggish sustain ganglion cell set in the retina of rabbits.

An OFF bipolar cell block features low DC current variation is proposed to solve the DC level variation problem in the previous work. The simulation result of the original structure shows that the variation of output current level is 8mA while the variation of the modified circuit is less than 0.04 μ A. The reduction of DC current variation is verified by the measurement results indicating that the maximum inter-pixel variations of DC current output is less than 0.26 μ A.

Besides the difference of the OFF bipolar cell, the ON bipolar cell block is also different in its structure. In the previous work, the ON bipolar cell performs bandpass-filtering on the signals, whereas the ON bipolar cell of this work performs large time delay and performs lateral diffusion in space. The structures are dissimilar.

The retinal chip contains a focal plane sensory array of 32x32. The fill factor is 9.14%, the maximum inter-pixel variation is less than 0.26mA and the total power consumption is 1.15W.

It is very important to duplicate successfully the retinal functions, channels, and visual language on silicon chips because of the key advantages this might provide. First, it could help neuroscientists to understand retinal functions and visual language. Since biological experiments can only be performed on a very limited number of cells, it is very difficult to see the global spatiotemporal features of retinal cells using this method alone. Second, it could provide valuable clues concerning neural activities in the visual cortex and thus move a few more steps toward the discovery of the visual processing of the brain. Third and finally, duplication could enable important applications in the areas of intelligent visual sensor systems and retinal prostheses. To date some CMOS neuromorphic chips have been designed to imitate the retinal channels [16], [21]-[22].

In this paper, the ON sluggish sustain ganglion cell set of a rabbit's retina is adopted and implemented on a CMOS neuromorphic chip. The neuromorphic model of the ON sluggish sustain ganglion cell set, which is directly derived from the biological measurements, is considered. The model-building approach is to incorporate the available knowledge concerning morphology, electro-physiology, and pharmacology and by only using elementary building blocks [17]. Then the model is transformed into a RC equivalent circuit which consists of gain blocks, resistors, and capacitors. Based on the RC equivalent circuit, a suitable macromodel is developed for CMOS circuit implementation. The resultant chip has been fabricated and measured, and its retinal functions have been verified successfully.

1.5 THESIS ORGANIZATION

This thesis is divided into five chapters. The first chapter introduces the background and motivation of this thesis. In chapter two of this paper, the retinal model of the ON sluggish sustain ganglion cell set is presented. The chip architecture and circuit design are elaborated in the following chapter. Simulation results of both the biological model and the circuit are also presented in this chapter. Chapter four shows the layouts of the chip and the measurement results. The various spatial, temporal, and spatiotemporal characteristics of the neuromorphic model and the implemented chip are shown. In the last chapter, we draw conclusions and bring up the direction of future works.

Table I

Ganglion cell types, defined by response to spot sizes and level of dendritic stratification [13]

| Cell type ¹ | Previous classification [11] | Response Polarity | Spiking response to small spots (100-200um) | Spiking response to large spots (600-100um) | Morphology |
|------------------------|-------------------------------|---|---|---|--|
| 1(22) | OFF Brisk Linear 1 | OFF | Sustained ² | Sustained | Monostratified |
| 2 (7) | OFF Brisk Linear 2 | OFF | Sustained | Sustained | Monostratified, dye-coupled to amacrine cells ³ |
| 3(36) | OFF Local Edge Detector | ON OFF for 50-200um OFF for > 200um spots | Sustained | No response | Monostratified |
| 4(25) | OFF Brisk Transient | OFF | Transient | Transient | Monostratified |
| 5(16) | OFF Brisk Transient Nonlinear | ON | Transient | Transient | Monostratified |
| 6(20) | ON Brisk Transient Linear | ON | Transient | Very small response or no response | Monostratified |
| 7(28) | ON Brisk Sustained Linear | ON | Sustained | Transient | Diffusively stratified |
| 8(6) | On Sluggish | ON | Sustained | Sustained | Monostratified |
| 9(10) | ? | ON | Sustained | Sustained | Bi-stratified |
| 10(4) | ON Local Edge Detector | ON | Sustained | No response | ? |

¹ Numbers in brackets refer to the number of measured cells. (16 ON cells and 13 OFF cells were not classified.)

² The response was defined Transient if at the end of a 1-s flash there was no response. Otherwise it was defined as Sustained. ON cells were tested with white spots; OFF cells with black spots.

³ This cell type had a little wider ramification than the other monostratified cell classes.

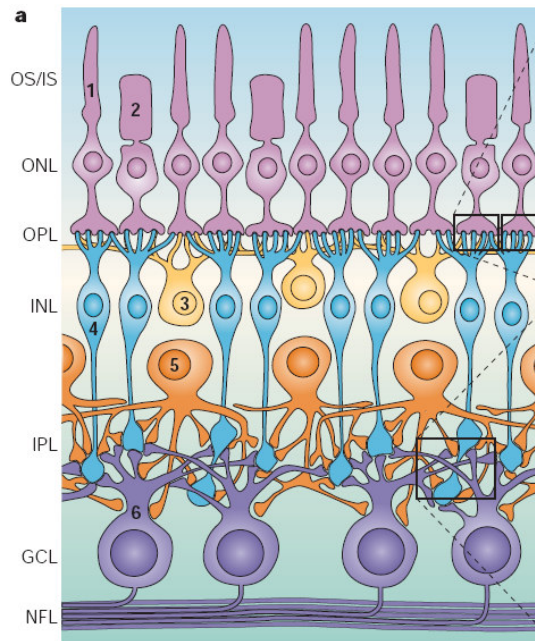


Fig. 1. 1. Schematic of the mammalian retina. Reprinted from [1].

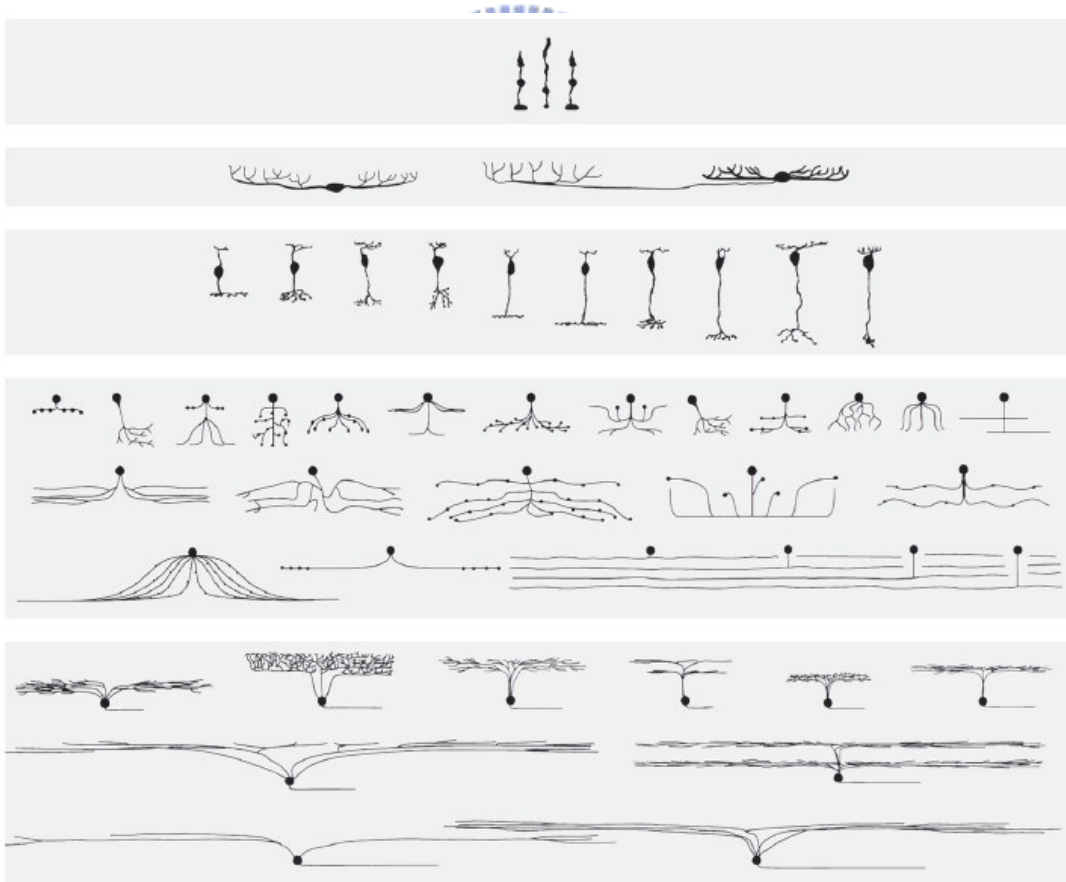


Fig. 1. 2. The major cell types of a typical mammalian retina. From the top row to the bottom, photoreceptors, amacrine cells and ganglion cells. Reprinted from [2].

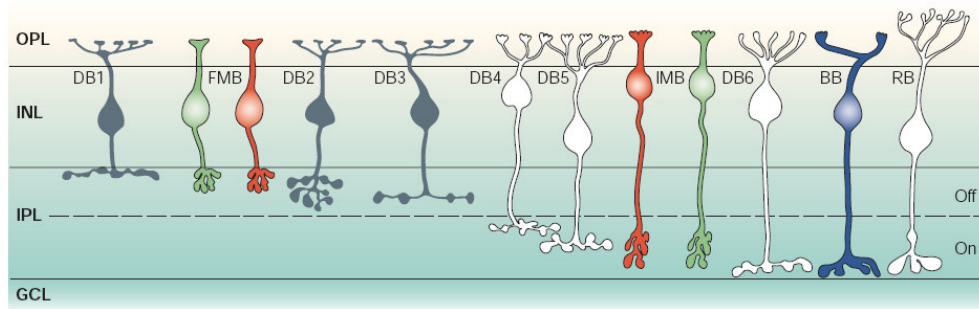


Fig. 1. 3. Bipolar cell types of the primate retina. Reprinted from [1].

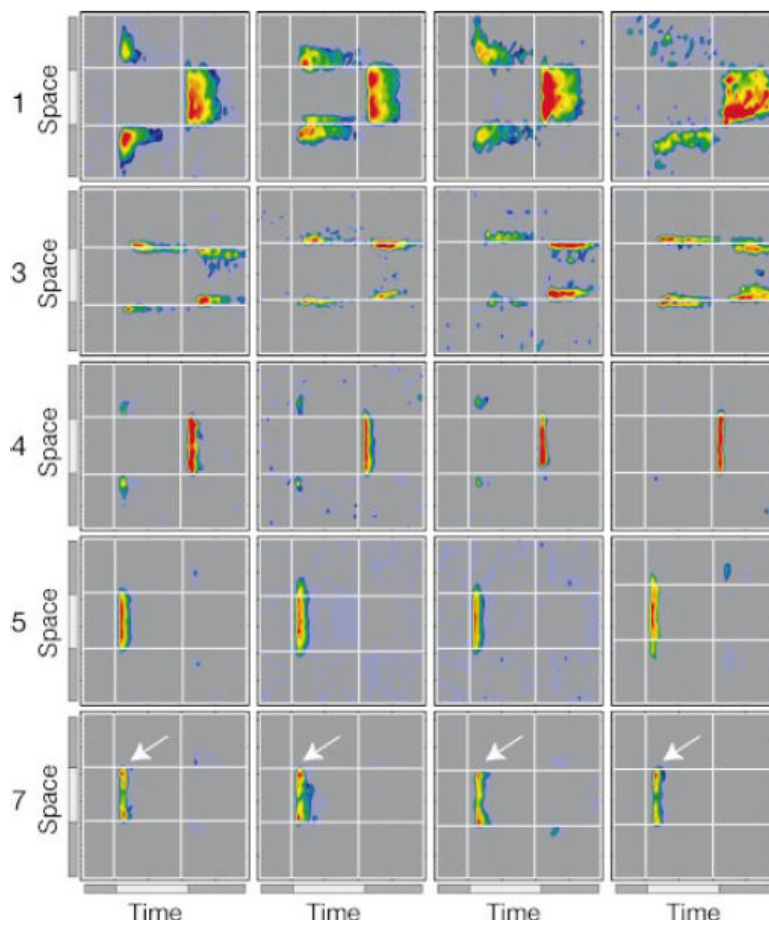


Fig. 1. 4. Spiking patterns in each row (measured from four different members of the same cell class) for five different classes of ganglion cells. The numbers at the beginning of each row refer to the classification shown in Table 1. 1. The arrows in row 5 (cell type 7) point to the enhanced activity at the stimulus edge, a feature that is characteristic of this class [13].

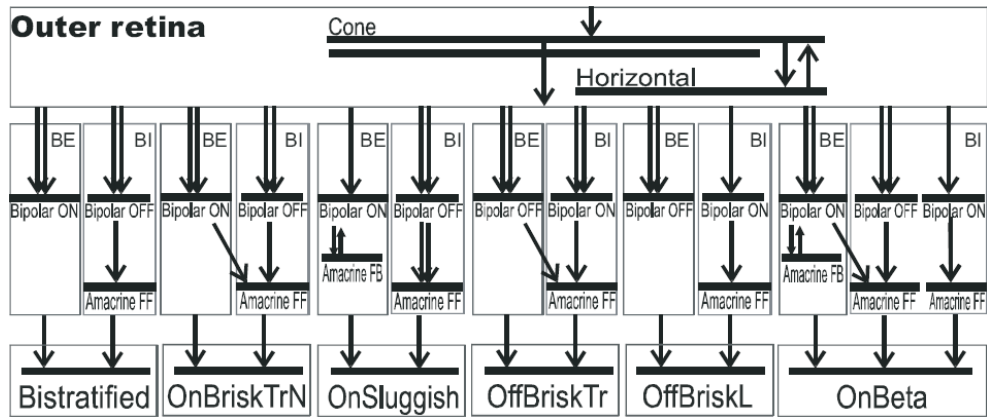


Fig. 1. 5. A sketch of the processing structure of the CNN model of the mammalian rabbit retina [14].



CHAPTER 2 The CNN Model of the ON Sluggish Sustain Ganglion Cell

2.1 NEUROMORPHIC MODEL OF THE RETINAL CELL SET

The block diagram of the neuromorphic model of the ON brisk transient ganglion cell set is shown in Fig. 2.1, where each block has a specific name and is defined as the abstract neuron in [23]. The parameter τ denotes the time constant in ms of the lowpass filter (LPF). The parameter D denotes the space constant, which is defined by the laterally diffusing range in cell number in a 180-cell array. In [23], D is defined as the spatial direct coupling strength between neurons. The physical meaning is defined by the spatial range where the signals decrease to 10% of its highest value. The two parameters τ and D can be realized by a simple LPF and a laterally diffusive network, respectively. The small block with the parameter G is the gain factor of each interconnection between the two blocks. The black block denoted as R means the signals are positively rectified through it. Due to biodiversity, the relationship between both the time and space constants is more important than their absolute values. Therefore, by scaling down these parameters similar spatiotemporal features are obtained after normalization. The space constants are obtained from morphological and electrophysiological data derived from living neural tissue, which may contain information concerning both cell sizes and lateral coupling between cells. However, it has been found that there are only three levels of lateral inhibition mediated by the horizontal cell, the OFF bipolar cell, and the amacrine cell [24]. The lateral inhibitions compress the spatial representation of the stimulus, thus they have some key effects on the spatiotemporal characteristics of the ganglion cell channels. Therefore, only the space constants of the horizontal cell, the OFF bipolar cell, and the amacrine cell contain the information of lateral coupling between cells and, as a consequence, they play more important roles than the other space constants. Corresponding to the actual cells, the blocks PH1 and PH2 together comprise the photoreceptor, where PH1 performs the spatial sensing function of a photoreceptor and PH2 performs the temporal function of the photoreceptor feedback.

The photoreceptor receives photo stimuli and transduces them into electrical signals. The block H provides functions of a horizontal cell, which performs lateral wide-range diffusion operation on signals of the photoreceptor and sends inhibitory feedback to the photoreceptor. The feedback from both PH2 and H are subtracted from the photo-stimuli at the block PH1 and then sent to the bipolar cells. The PH1 signals are amplified to four times their existing level before they enter the bipolar cells. There are two bipolar cells, namely the ON and OFF bipolar cells, in the proposed cell set and both the ON and OFF bipolar cells are of a transient type. The transient-type bipolar cells act like bandpass filters (BPF). Since the neuromorphic model is built by using LPF blocks, two LPF blocks are needed to realize the function of a bipolar cell. As can be seen in Fig. 2.1, the blocks ON Bipolar Cell comprise the ONBIP and Amacrine Feed-back, whereas OFFHELP and OFFBIP together comprise the OFF Bipolar Cell. Since the block OFFBIP has a smaller time constant ($\tau=13\text{ms}$) than that of the block OFFHELP ($\tau=85\text{ms}$), subtracting the output of OFFHELP from that of OFFBIP can generate BPF signals with the up 3dB frequency at 11.77Hz and the down 3dB frequency at 76.92Hz in the OFF Bipolar Cell. Moreover, the OFF bipolar cells perform a very narrow lateral diffusion. Subsequently, the ON and OFF bipolar cell signals are positively rectified before they enter the amacrine and ganglion cells. The block named Amacrine Cell provides the functions of an amacrine feed-forward cell. In the proposed cell set, the amacrine cell receives positively rectified signals from OFF bipolar cells a gain factors of 1. It performs a lateral diffusion in a small range and then sends inhibitory signals to the ganglion cell. The block named Ganglion Cell performs the functions of the ganglion cell. It receives the positively rectified signal from the ON bipolar cell with a gain factor of 2 and the signal from the amacrine cell with a gain factor of -2. Finally, the signals are positively rectified to achieve the complete function of the ganglion cell. This ganglion cell signal can be used to generate the neural spiking necessary to communicate with the brain. The actual RC equivalent circuit of the electrical model is shown in Fig. 2.2, where τ is realized by a simple RC LPF and D is realized by a 2-D resistor array. To understand the behavior and spatiotemporal pattern of every cell, the RC equivalent circuit is simulated and analyzed. The

simulated results are presented in Chapter 2.3. Since the neuromorphic model is based on the available neuroscientific knowledge of the retina, the simulated spatiotemporal patterns of the RC equivalent circuit similar to the biologically measured results.

2.2 MACRO MODEL FOR CHIP IMPLEMENTATION

To facilitate the integrated circuit implementation, the neuromorphic model in Fig. 2.1 is transformed into the macromodel in Fig. 2.3. This macromodel is used to integrate the cell set into a single pixel. A sensory chip containing 32x32 pixels is employed to verify the function of the macromodel. The macromodel consists of a photoreceptor (PH1 and PH2), a horizontal cell (H), an ON and OFF bipolar cells (ONBIP and OFFBIP), an amacrine cell (Ama), and a ganglion cell (GC). Block R ensures that the signals are positively rectified as they pass through it. The + and - signs within a circle indicate the signals which are added and subtracted, respectively, at that node. Due to the variations of biological cells, the constants in space and time for each cell have their varying degrees of tolerance. This allows inevitable process variations in CMOS circuit realization of these space and time constants. There were some modifications in the construction of the macromodel from the neuromorphic model of Fig. 2.1. First, a gain stage with a value -8 was added in front of the pixel circuit. This is used to enlarge the input photocurrent to facilitate the tracking of the circuit operations. Second, space constants of PH1 was removed. The space constants of the horizontal cell, the ON bipolar cell, the OFF bipolar cell, the amacrine cell and the ganglion cell are considered in order to produce the complex interconnection while retaining the functions of the six levels of lateral inhibition. Furthermore, since the pixel array shrinks to 32x32, the space constants of the horizontal cell, the ON bipolar cell, the amacrine feedback cell, the OFF bipolar cell, the amacrine cell and the ganglion cell are proportionally scaled down to 27, 14, 12, 6, 12 and 10, respectively. Third, there are better methods to implement a BPF with CMOS circuit rather than subtracting a LPF from the other. Therefore, at the OFF bipolar cell, the two blocks OFFHELP and OFFBIP were merged

into the one block OFFBIP in the macromodel, as can be seen in Fig. 2.3. The constants of OFFHELP and OFFBIP were also merged into a single time constant, τ_{BPF} . The two values in τ_{BPF} denote the upper and lower 3dB time constants. Fourth, the time constant of the blocks PH1 and H, were modified to τ_L , and that of block PH2 was modified to $4\tau_L$ to facilitate chip implementation. For the same reason, the higher 3dB time constant of block OFFBIP was changed to $4\tau_L$, and the lower 3dB time constant of block OFFBIP was changed to τ_H , respectively. Finally, since the time constants of the amacrine cell and the ganglion cell were smaller than those of the ON and OFF bipolar cells, their neglect had no effect on the signal flow. Therefore, the time constants of the blocks Ama and GC were removed in the macromodel.

2.3 SIMULATION RESULTS

The RC equivalent circuit in Fig. 2. 1 is used to construct a 32x32 array which is simulated by HSPICE. Since the array size is differs from the original definition of the space constants D, all the space constants in the simulation are shrunk with the same factor, 32/180. The simulation results are shown in Fig. 2. 4 ~ Fig. 2. 6. In this simulation, a voltage pulse is applied to the 5x5 cells in the center of the array to imitate the 500Hz flashing light stimulus and the spatiotemporal patterns of the 17th row are observed.

First we show the transient response of the circuit when a pulse-stimulus is supplied. Fig. 2. 4 shows the output currents of the center cell of the 2-D array. The simulation is performed under the condition that a pulse signal with a turned-on duration of around 1ms is incident on the middle six cells of the array. The simulated input to the middle six cells are periodic stimulus with 100pA transient current added to a 100pA background-induced DC current. All other twenty-six cells are supplied with background currents of 100pA. All above-mentioned currents are supplied to the bases of the photo-BJT of all cells. The output of photoreceptor in Fig 2. 4(a) has overshooting and undershooting at the turn-on and turn-off similar to the CNN model simulation results in the

previous section. It is also clear in Fig. 2. 4(b) that the output of the horizontal is analogous to that of the PH1 but has a larger magnitude.

As mentioned in the previous section, the gate bias of MSM11, MSM12, MSM21, MSM22, MSM31, MSM32, MSM41, MSM42, MSM51, MSM52, MSM61 and MSM62, VSM, controls the diffusion range of the horizontal, on bipolar cell, off bipolar cell, amacrine cell and ganglion cell, respectively. Fig 2. 5 shows the steady state output of space domain response with subjected to suitable gate bias voltages of V_{SM} . Fig 2. 5(b) is the results of the horizontal when the middle six cells are incident with stronger light while the other cells are incident with background light. As could be seen in the figure, a higher gate biasing voltage, VSM, causes lower resistance of the smoothing network, and a wider diffusion range is achieved. The results of photoreceptor with the same stimulus are shown in Fig 2. 5(a). As discussed previously, the edge of the incident pattern would have higher contrast in the output of PH1 but the extent of contrast varies with different diffusion range.

In Fig. 2. 6, the x-axis is time and the y-axis is the pixel location which denotes space. The black and white bars denote the spatial and temporal region of the input stimulus. The stimulus is applied to the 15th to the 20th pixel from 1001 μ sec to 2000 μ sec. The waveform at the right of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. Fig. 2. 6(a) and (b) represent the spatiotemporal patterns of the photoreceptor and horizontal cell, respectively. It can be seen from Fig. 2. 6(a) that the photoreceptor's signal level drops when there is a stimulus, and it returns to its original level when the stimulus disappears. Slight undershooting and overshooting in temporal domain can be expected in the periphery as it reacts to the stimulus directly. In the spatial domain, there is strong contrast at the edge of the stimulus. At the edge pixels inside the stimulus, the signal level is higher than the other stimulated pixels. Contrarily, at the edge pixels where the stimulus is just absent, the signal level is higher than the other silent pixels. Therefore, the spatial range of the stimulus can be well

defined. In the temporal domain, the horizontal cell has a similar reaction to the photoreceptor, as can be seen in Fig. 2. 6(b). However, since the horizontal cell performs lateral diffusion in space, its spatial domain waveform spreads wider than that of the photoreceptor. Fig. 2. 6(c) and (d) represent the spatiotemporal patterns of the ON and OFF bipolar cells, respectively. The OFF bipolar cell performs bandpass-filtering on the signals from the photoreceptor. In the amacrine cell, the rectified signals from off bipolar cell with a gain factor, as shown in Fig. 2. 6(e). Moreover, the response to the appearance of the stimulus is weaker than the response to the disappearance of the stimulus. Thus the amacrine cell provides a strong inhibition to the ganglion cell when the stimulus disappears in time. Therefore, the ganglion cell exhibits clear turned-ON reaction, as can be seen in Fig. 2. 6 (f).



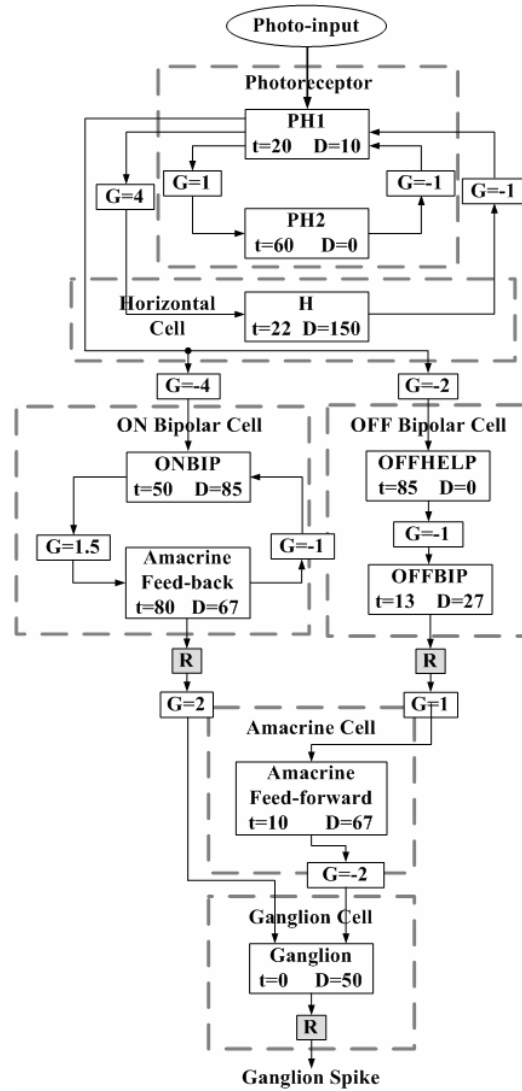


Fig. 2. 1. The neuromorphic model of the ON sluggish sustain ganglion cell set of rabbits' retinas. The parameter τ denotes the time constant in millisecond of the LPF. The parameter D denotes the space constant, which is defined by the laterally diffusing range in cell number in a 180-cell array.

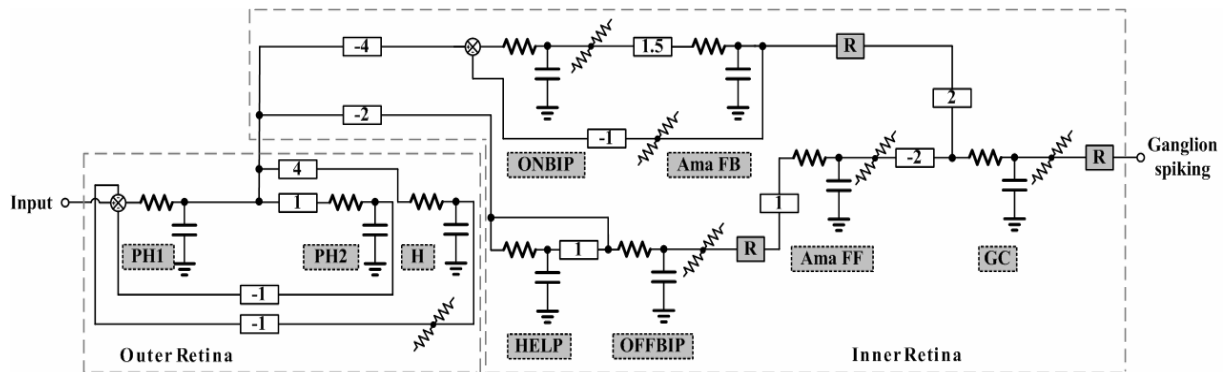


Fig. 2. 2. The RC equivalent model of the ON sluggish sustain ganglion cell set of rabbits' retinas.

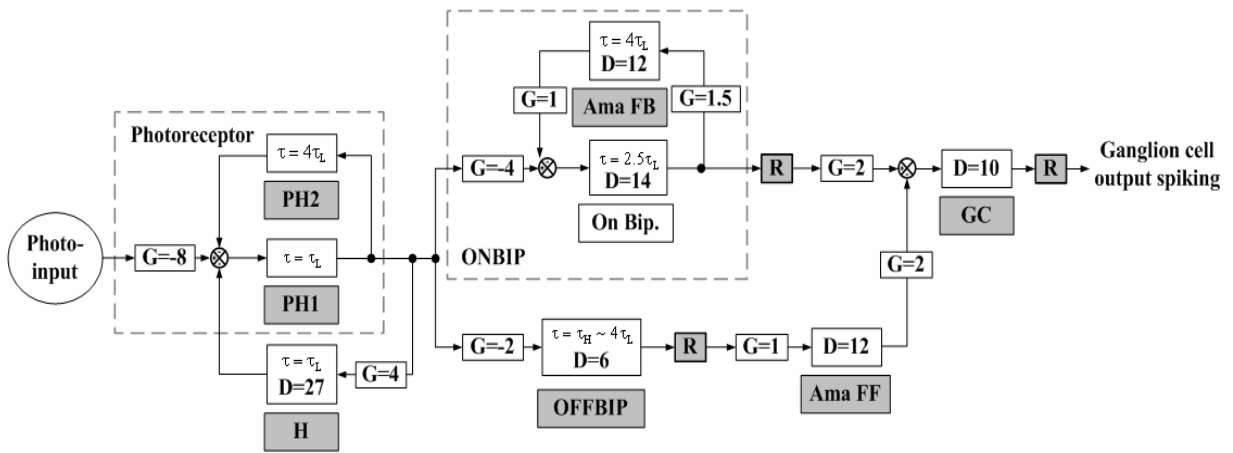


Fig. 2. 3. The macromodel of a single pixel of the implemented chip.



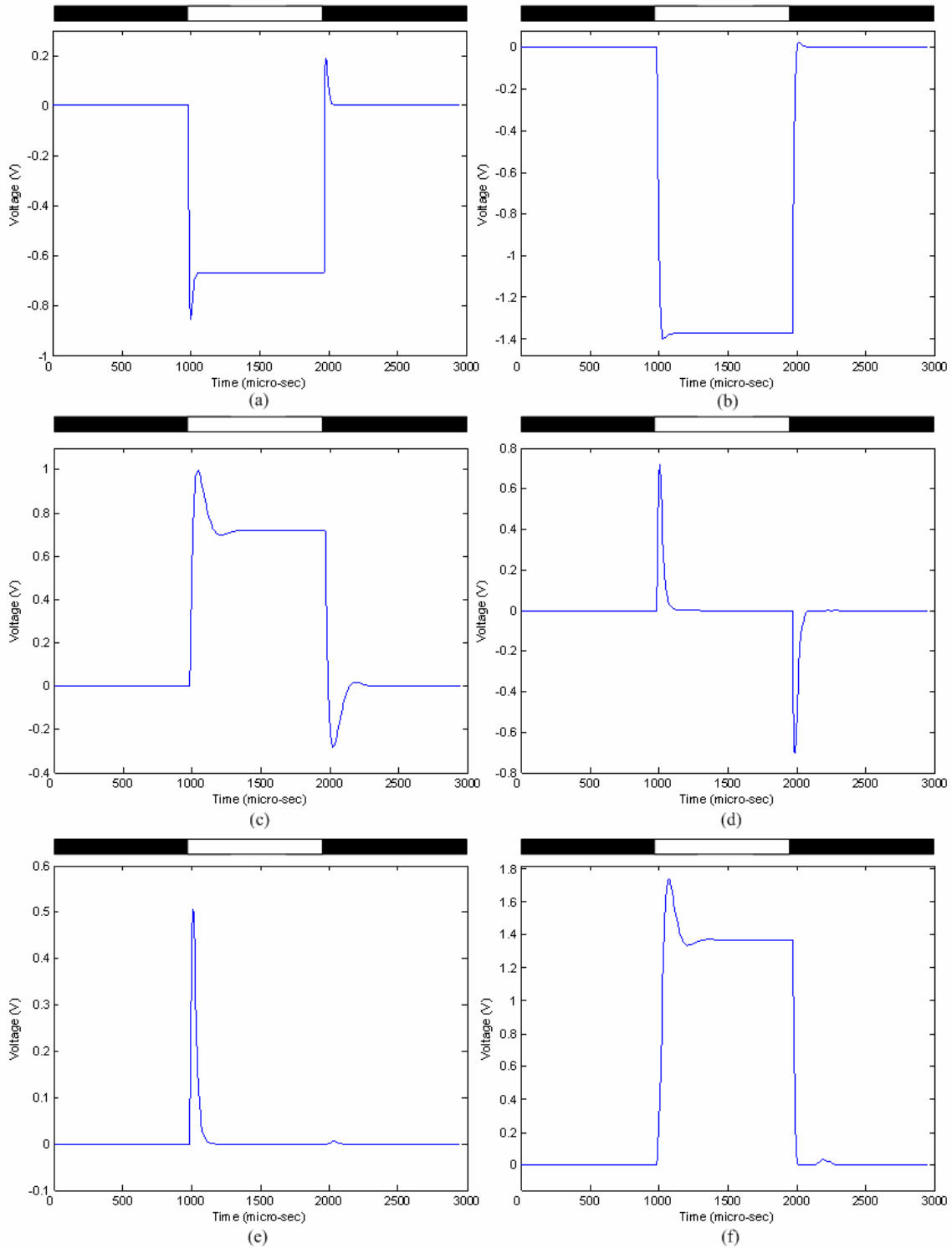


Fig. 2. 4. The HSPICE model-simulated time domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. The light stimulus is applied at 1000th~2000th micro-second and $V_{sm} = 2.2V$.

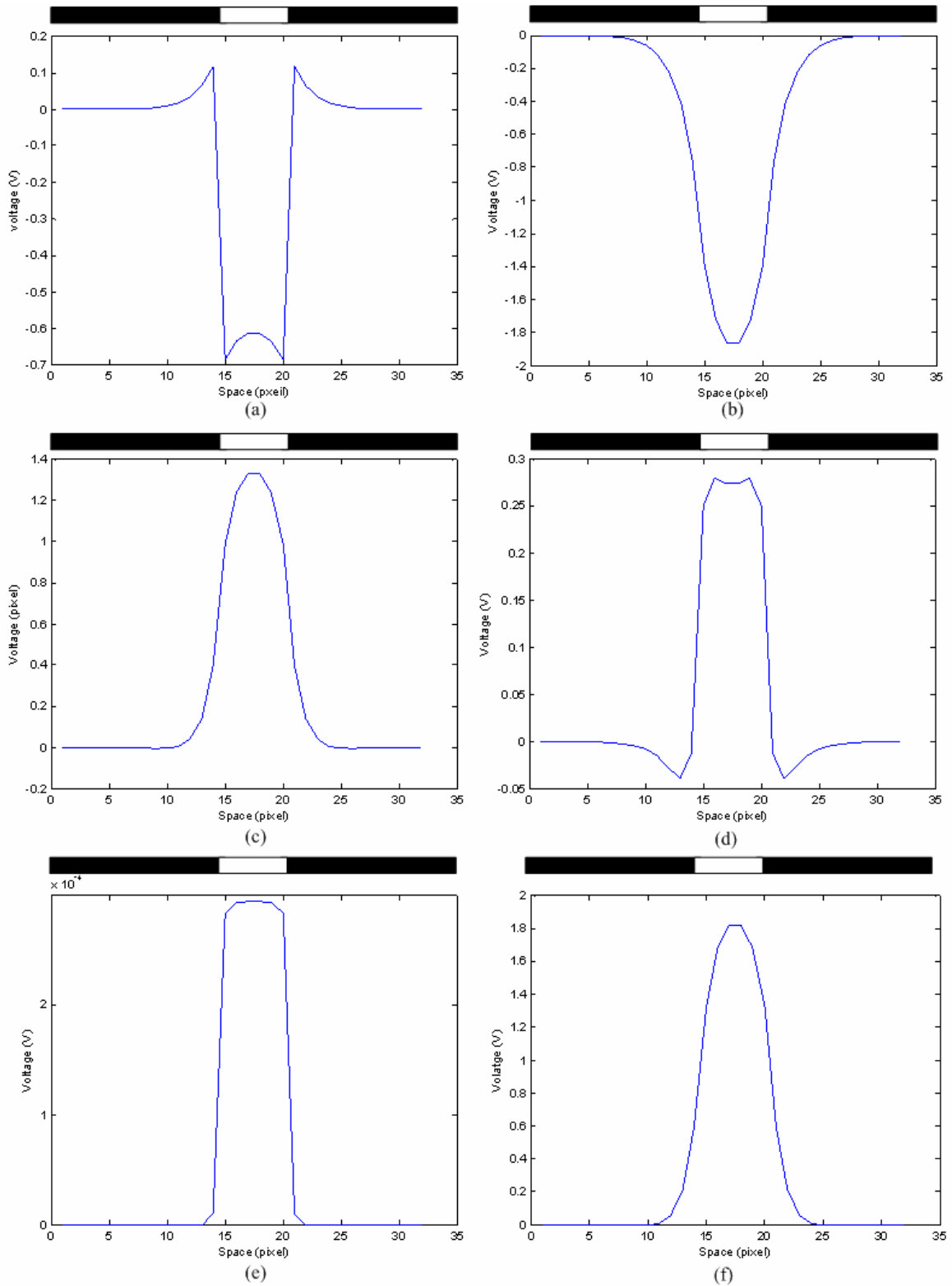


Fig. 2. 5. The HSPICE model-simulated space domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. The light stimulus is applied at 15th~20th pixel and $V_{sm} = 2.2V$.

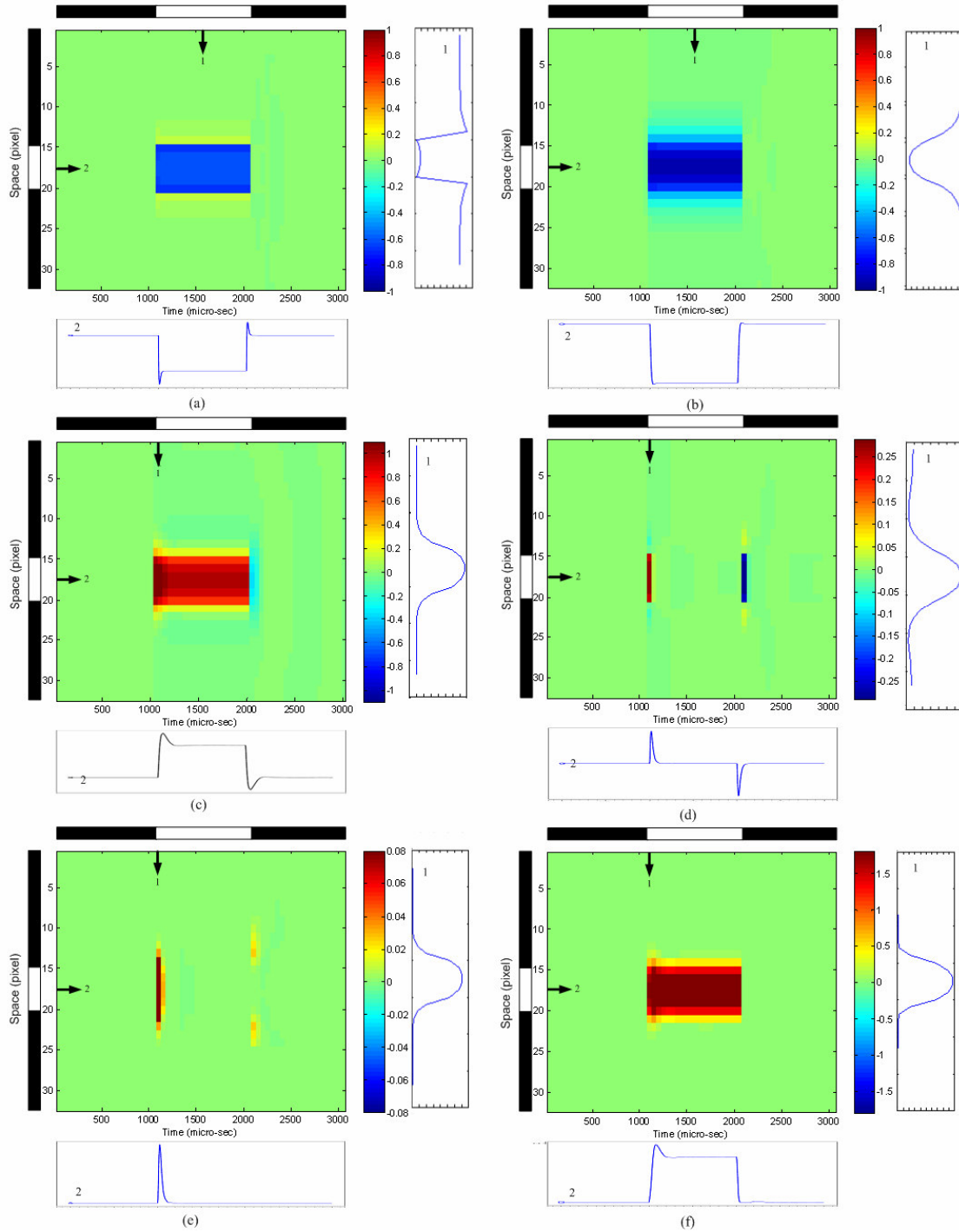


Fig. 2. 6. The HSPICE model-simulated time domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These waveforms are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 15th to the 20th pixel at time point 1001 to 2000. The waveform at the left of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. The $V_{sm} = 2.2V$.

CHAPTER 3 Circuit Design

3.1 ARCHITECTURE OF A BASIC CELL OF THE RETINAL CHIP

Fig. 3.1 is the block diagram corresponding to the macromodel in Fig. 2. 2. A vertically parasitic p+-n-well-p-substrate BJT Q_{PHO} with a floating base is used as a photo-BJT. The photo-BJT traduces light stimuli into electrical currents. Then the photocurrent is read out, inverted, amplified by eight times, and sent to the block PH1. Then the PH1 signal is sent to PH2 and H to perform temporal delay and spatial diffusion, respectively. Then the signals from PH2 and H are fed back into PH1. Since the input of this circuit is a photocurrent, all signals in PH1, PH2, and H are in current-mode. An additional cascode current mirror (denoted by sub-block CM) is used to inverse it and send it to the OFF bipolar cell (OFFBIP) in order to achieve the inverse band-pass-filtering function. The output current I_{ON} of PH1 directly enters the ON bipolar cell (ONBIP). Since the retina operation is very slow, very low frequency BPFs (denoted by sub-block BPF) are needed for the bipolar cells. For OFF bipolar cell, the current-mode signals from PH1 and CM are converted into voltage-mode signals by Rm amplifiers (denoted by the sub-block Rm Amp.) and fed into the BPF sub-blocks. The output signals of the sub-blocks BPF are converted to current-mode signals again using the sub-blocks V-I converter. The absolute-value circuits (denoted by the sub-block ABS) are used to rectify the bipolar cell signals. At the amacrine cell (Ama), the positively rectified signals of the OFF bipolar cells with a gain factor. The ganglion cell (GC) subtracts the amacrine cell's signal from the positively rectified signal of the ON bipolar cell. Another sub-block ABS is used to rectify the ganglion cell's signal in order to realize the complete function of the ON sluggish sustain ganglion cell set. There are six smoothing networks at the horizontal cell, ON bipolar cell, OFF bipolar cell, amacrine cell and ganglion cell, respectively, which are used to realize the space constants of these six cells. The space constants are realized by using tunable NMOSFET smoothing networks [4]. Such a technique can control the space constants by controlling the gate biases of the NMOSFETs, thus allowing greater flexibility in designing the space constants. Moreover, since the

horizontal cell has a relatively large space constant compared with the other cells, the NMOSFETs used in its smoothing network have a larger W/L ratio than the other five cells, as shown in Table II. The detailed transistor-level circuit of a pixel is shown in Fig. 3. 2 (a)-(d), and the transistor sizes are listed in Table II.

3.2 CIRCUIT DESIGN OF THE BASIC CELL

3.2.1 Photoreceptor and Horizontal cell

Fig. 3. 2(a) represents the circuit of the photoreceptor (PH1 and PH2) and horizontal cell (H). Each block is constructed basically using cascode current mirrors to deal with the light-induced photocurrent. Because the photocurrent is around one hundred-pA to several hundred-nA, it can benefit from the good linearity of cascode current mirrors. Moreover, since the photocurrent might flow in or out of the circuit, complementary current mirrors are used to enable both current directions. The transistors M1-M6 are used to bias the emitter of Q_{PHO} at V_{B1} and to direct the photocurrent into the circuit. The transistors M7 - M18 provide a current gain of -8, as described in Fig. 3. 1. The CMOS transmission gates composed of transistors M_{PLP1} - M_{PLP12} and M_{LPON1} - M_{LPON8} in Fig. 3. 2 (a) and (b) are used as resistors. Their resistances are controlled by V_{NLP1} and V_{PLP1} for NMOSFET and PMOSFET, respectively. The resistances and the gate capacitances compose the LPF time constants of the blocks PH1, PH2, H, and ONBIP. The unit LPF time constant τ_L can be controlled by $V_{NLP1,2}$ and $V_{PLP1,2}$ simultaneously. On the other hand, the ratios of time constants can be set by designing the MOSFET's sizes to get the suitable gate capacitances. The current mirror composed of transistors M_{PH11} - M_{PH14} and M_{PH15} - M_{PH18} is designed to have a current gain of four times. Thus the current of block PH1 is enlarged to enter block H. The enlarged current is spread by a diffusion network controlled by V_{SM} , and then it is fed back to the block PH1 by another complementary current mirror, including transistors M_{H1} - M_{H16} . On the other hand, the current of the block PH1 is directly repeated to enter the block PH2 via the current mirror. Complementary current

mirror including transistors M_{PH21} - M_{PH24} is used to perform the temporal delay and to send feedback to the block PH1. Therefore, the channel length and width of transistors M_{PH21} , M_{PH22} , M_{PH23} , and M_{PH24} , are both twice as large as those of M_{PH11} , M_{PH12} , M_{PH13} , and M_{PH14} . Thus, the current gain is kept as unity because the W/L ratio doesn't change. However, the time constant of the block PH2 is four times larger than that of the block PH1 because the gate capacitances of M_{PH21} and M_{PH24} are four times larger than those of M_{PH11} and M_{PH14} .

The output current of block PH1 is enlarged four times by the current mirrors composed of transistors M_{PH11} - M_{PH14} , M_{CM1} - M_{CM4} , M_{CM5} - M_{CM12} and M_{CM13} - M_{CM16} to be sent to the bipolar cells. Transistors M_{CM1} - M_{CM12} generate I_{ON} , which is to be sent to the block ONBIP. The drain current of transistors M_{CM2} and M_{CM3} is inversed using a complementary current mirror (CM). Transistors M_{CM1} - M_{CM8} and M_{CM13} - M_{CM16} generate I_{OFF} , which is to be sent to the block OFFBIP. The drain current of transistors M_{CM2} and M_{CM3} is inversed using a complementary current mirror (CM).

At the output stage, the transistors M_{O1} and M_{O2} , M_{O3} and M_{O4} , and M_{O5} and M_{O6} are used to repeat the output currents of the blocks PH1, PH2, and H, respectively, to enable chip measurement. Since there are external biases applied to measure the chip, simple structure instead of the cascode type slaves are used to save the pixel area.

3.2.2 On bipolar cell

The CMOS transmission gates composed of transistors M_{LPON1} ~ M_{LPON4} and M_{LPON5} ~ M_{LPON8} are used as transistors. Their resistances are controlled by V_{NLP2} and V_{PLP2} for NMOSFET and PMOSFET, respectively. The ratios of time constants can be set by designing the MOSFET's sizes to get the suitable gate capacitances. Transistors M_{ON1} ~ M_{ON8} are used to perform the temporal delay and to send feedforward to block Ama FB.

The function of amacrine feed-back cell is to generate a low-passed signal from on bipolar cell's output with a large time constant coefficient, τ , and then to feed the low-passed signal back to on bipolar cell for subtraction. A couple of current mirrors are used to realized amacrine feed-back

cell for subtraction. This generates undershoot and overshoot waveforms similar to biological measurements. To realize the architecture, we need a low-pass filter with a large time constant. A circuit named “continuous time current delay element” is used to implement the current-mode low-pass filter. The current delay element is a current mirror with a capacitor connected between the gates and sources of the two MOSFET.

The result of current subtraction at the drain of M_{ON26} and M_{ON27} are then passed to a modified absolute value circuit (ABS) composed of the transistors $M_{ON29} \sim M_{ON35}$. Transistors $M_{ON29} \sim M_{ON32}$ form a simple current mirror, and transistors M_{ON33} and M_{ON34} play the roles of adjustable switches that are controlled by V_{absN} and V_{absP} , respectively. The transistor M_{ON33} sends the positively rectified signal of the block ONBIP to the block GC as I_{ON_GC} .

3.2.3 Off bipolar cell

The circuit of the off bipolar cell stage is shown in Fig. 3. 2(c). At the OFFBIP path, an Rm amplifier composed of transistors $M_{RM1} \sim M_{RM4}$ is used to transduce the current I_{OFF} into voltage-mode. Afterward, a poly capacitor C_{OFF} and a CMOS transmission gate, including transistors M_{HP1} and M_{HP2} , are used to form a highpass filter (HPF). This HPF and the LPF composed of M_{PLP13} , M_{PLP14} , and M_{VI1} form the sub-block BPF. The bandwidth of the HPF, which is also the lower 3dB time constant of the BPF τ , can be tuned by the biases V_{NHP} and V_{PHP} at gates of M_{HP1} and M_{HP2} , respectively. The transmission gate is connected to the reference bias V_{ref} , which is set to a half of the supply voltage to balance both upward and downward signal swings. The voltage-mode signal of the sub-block BPF is transduced into the current-mode signal again by using a transconductor composed of transistors $M_{VI1} \sim M_{VI10}$ [25]. The transconductor is a low-voltage low-power class-AB linear transconductor [26]. Its positive input node connects to the output of the sub-block BPF, and its negative input node connects to the reference bias V_{ref} . The circuit performs as a linear transconductor but it works in class-AB. This is due to the FVF (flipped voltage follower) [27] low impedance buffers that only force the current of the circuit under quiescent conditions, and

can source any amount of current demanded by M_{V11} and M_{V12} when a differential input signal is applied. This fact makes this linear OTA very attractive for low-power applications as the maximum current at the output $I_{MV11}-I_{MV12}$ is comparable to the total current consumed by the circuit $I_{MV13}+I_{MV14}$, where I_{MV13} and I_{MV14} are the drain currents of transistors M_{V13} and M_{V14} , respectively.

In order to present this principle, let's consider the differential pair of figure 3.3. The drain currents of transistors M1 and M2 are given by

$$\begin{aligned} I &= \frac{1}{2}k_p \left(\frac{W}{L} \right) (V_{SG} - |V_{tp}|)^2 \\ &= \frac{1}{2} \cdot \beta \cdot (V_{SG} - |V_{tp}|)^2, \quad \beta = k_p \left(\frac{W}{L} \right) \end{aligned} \quad (1)$$

β is the transistor gain factor defined as $\beta=k_p(W/L)$. We assume that M1 and M2 are identical p-MOS transistors working in the saturation region, so that $\beta_1=\beta_2=\beta$. The differential output current of this circuit is defined as:

$$\begin{aligned} I_{out} &= I_2 - I_1 \\ &= \frac{1}{2} \cdot \beta \cdot \left[(V_{SG2} - |V_{tp}|)^2 - (V_{SG1} - |V_{tp}|)^2 \right] \\ &= \frac{1}{2} \cdot \beta \cdot (V_{SG1} + V_{SG2} - 2|V_{tp}|) \cdot (V_{SG2} - V_{SG1}) \end{aligned} \quad (2)$$

$$\begin{aligned} |V_1 - V_{SG1} + (V_X + |V_{tp}|) - V_{SG2} + (V_X + |V_{tp}|) &= V_1 \\ \Rightarrow V_{SG1} - (V_X + |V_{tp}|) + V_{SG2} - (V_X + |V_{tp}|) &= 0 \\ \Rightarrow V_{SG1} + V_{SG2} &= 2 \cdot (V_X + |V_{tp}|) \end{aligned} \quad (3)$$

From equation (2), the output current I_o will be linear if the sum of gate-source voltages (term $V_{SG1}+V_{SG2}-2|V_{tp}|$) is kept constant. A possible approach to maintain a constant sum of gate-source voltages is by using two floating DC voltage sources with values $V_X+|V_{tp}|$, as it is shown in figure 3.3 (a). From the circuit in figure 3.3 (a), if we consider the loop around the input transistors, we can derive

$$\begin{aligned}
V_1 - V_{SG1} + (V_X + |V_{tp}|) &= V_2 \\
\Rightarrow V_1 - V_2 &= V_{SG1} - (V_X + |V_{tp}|) \\
\Rightarrow V_{SG1} &= (V_1 - V_2) + (V_X + |V_{tp}|)
\end{aligned} \tag{4}$$

So it is easy to demonstrate that $V_{SG1} + V_{SG2} = 2(V_X + |V_{tp}|)$. Writing the relationship between V_{SG1} and V_{SG2} as a function of V_1 and V_2 ,

$$\begin{aligned}
V_1 - (V_X + |V_{tp}|) + V_{SG2} &= V_2 \\
\Rightarrow V_2 - V_1 &= V_{SG2} - (V_X + |V_{tp}|) \\
\Rightarrow V_{SG2} &= (V_2 - V_1) + (V_X + |V_{tp}|)
\end{aligned} \tag{5}$$

$$\begin{aligned}
I_{out} &= I_2 - I_1 \\
&= \frac{1}{2} \cdot \beta \cdot \left[(V_{SG2} - |V_{tp}|)^2 - (V_{SG1} - |V_{tp}|)^2 \right] \\
&= \frac{1}{2} \cdot \beta \cdot (V_{SG1} + V_{SG2} - 2|V_{tp}|) \cdot (V_{SG2} - V_{SG1}) \\
&= \frac{1}{2} \cdot \beta \cdot \left[(V_1 - V_2) + (V_X + |V_{tp}|) + (V_2 - V_1) + (V_X + |V_{tp}|) - 2|V_{tp}| \right] \\
&\quad \cdot \left[(V_2 - V_1) + (V_X + |V_{tp}|) - (V_1 - V_2) - (V_X + |V_{tp}|) \right] \\
&= \frac{1}{2} \cdot \beta \cdot (2V_X) \cdot [2(V_2 - V_1)] \\
&= 2 \cdot \beta \cdot V_X \cdot (V_2 - V_1)
\end{aligned} \tag{6}$$

Replacing $V_{SG1} + V_{SG2}$ and $V_{SG1} - V_{SG2}$ in expression (2), the output differential current is given by

$$I_{out} = G_m \cdot (V_2 - V_1), \quad G_m = 2 \cdot \beta \cdot V_X \tag{7}$$

$$\begin{aligned}
I_b &= \frac{1}{2} \cdot \beta_5 \cdot (V_{SG5} - |V_{tp}|)^2 \\
\Rightarrow V_{SG5} - |V_{tp}| &= \sqrt{\frac{2I_b}{\beta_5}} \\
\Rightarrow V_{SG5} &= \sqrt{\frac{2I_b}{\beta_5}} + |V_{tp}| \\
\Rightarrow V_{SG5} = V_X + |V_{tp}|, \quad V_X &= \sqrt{\frac{2I_b}{\beta_5}}
\end{aligned} \tag{8}$$

$$\begin{aligned}
I_b &= \frac{1}{2} \cdot \beta_6 \cdot (V_{SG6} - |V_{tp}|)^2 \\
\Rightarrow V_{SG6} - |V_{tp}| &= \sqrt{\frac{2I_b}{\beta_6}} \\
\Rightarrow V_{SG6} &= \sqrt{\frac{2I_b}{\beta_6}} + |V_{tp}| \\
\Rightarrow V_{SG6} &= V_X + |V_{tp}|, \quad V_X = \sqrt{\frac{2I_b}{\beta_6}}
\end{aligned} \tag{9}$$

$$\begin{aligned}
V_{S1} &= V_2 + (V_X + |V_{tp}|) \\
V_{S2} &= V_1 + (V_X + |V_{tp}|)
\end{aligned} \tag{10}$$

Assuming n-well technology, $V_{SB}=0$ for all transistors, $\beta_1=\beta_2$, $\beta_3=\beta_4$, and $\beta_5=\beta_6$, the equations that describe the circuit are given by:

$$\begin{aligned}
I_1 &= \frac{1}{2} \cdot \beta_1 \cdot (V_{S1} - V_1 - |V_{tp}|)^2 \\
&= \frac{1}{2} \cdot \beta_1 \cdot [V_2 + (V_X + |V_{tp}|) - V_1 - |V_{tp}|]^2 \\
&= \frac{1}{2} \cdot \beta_1 \cdot [V_X + (V_2 - V_1)]^2 \\
&= \frac{1}{2} \cdot \beta_1 \cdot \left[\sqrt{\frac{2I_b}{\beta_6}} + (V_2 - V_1) \right]^2
\end{aligned} \tag{11}$$

$$\begin{aligned}
I_2 &= \frac{1}{2} \cdot \beta_2 \cdot (V_{S2} - V_2 - |V_{tp}|)^2 \\
&= \frac{1}{2} \cdot \beta_2 \cdot [V_1 + (V_X + |V_{tp}|) - V_2 - |V_{tp}|]^2 \\
&= \frac{1}{2} \cdot \beta_2 \cdot [V_X - (V_2 - V_1)]^2 \\
&= \frac{1}{2} \cdot \beta_2 \cdot \left[\sqrt{\frac{2I_b}{\beta_6}} - (V_2 - V_1) \right]^2
\end{aligned} \tag{12}$$

Replacing V_{SG1} and V_{SG2} as a function of the input voltages V_1 and V_2 , the following expressions result,

$$\begin{aligned}
I_{\text{out}} &= I_2 - I_1 \\
&= \frac{1}{2} \cdot \beta \cdot \left[[V_X - (V_2 - V_1)]^2 - [V_X + (V_2 - V_1)]^2 \right] \\
&= \frac{1}{2} \cdot \beta \cdot [V_X - (V_2 - V_1) + V_X + (V_2 - V_1)] \cdot [V_X - (V_2 - V_1) - V_X - (V_2 - V_1)] \quad (13) \\
&= \frac{1}{2} \cdot \beta \cdot (2V_X) \cdot [2(V_1 - V_2)] \\
&= 2 \cdot \beta \cdot V_X \cdot (V_1 - V_2)
\end{aligned}$$

If we subtract (6) and (5) and identify the result with the expression $I_o = I_2 - I_1 = G_m(V_2 - V_1)$, we obtain the transconductance value,

$$I_{\text{out}} = G_m \cdot (V_1 - V_2), \quad G_m = 2 \cdot \beta \cdot V_X = 2 \cdot \beta \cdot \sqrt{\frac{2I_b}{\beta_6}} \quad (14)$$

This OTA has a very high current efficiency [28], as the static power consumption can be much lower than the power consumption when the signal is applied at the input terminals. Its output current is passed to a modified absolute value circuit composed of the transistors $M_{\text{OFF1}}-M_{\text{OFF6}}$.

The transistors $M_{\text{OFF1}}-M_{\text{OFF2}}$ and $M_{\text{OFF5}}-M_{\text{OFF6}}$ form a simple current mirror. The transistors M_{OFF3} and M_{OFF4} play the roles of adjustable switches that are controlled by V_{absn} and V_{absp} , respectively. When M_{OFF3} and M_{OFF4} are biased suitably, the output of sub-block BPF, which is smaller than V_{ref} , will be converted into current-mode and mirrored only by M_{OFF1} and M_{OFF2} ; while that larger than V_{ref} will be converted and mirrored only by M_{OFF5} and M_{OFF6} . The transistors M_{OFF2} and M_{OFF6} repeat the complete signal of the block OFFBIP and send it out of the chip. The transistor M_{OFF7} repeats the signal of the block OFFBIP only when the output of the sub-block BPF is smaller than V_{ref} , and sends it to the block Ama as I_{OFF_A} . COFF is designed to be 10fF in order to achieve the relative ratio of time constants of the blocks OFFBIP. There is a smoothing network at the output node of the transconductor of the OFFBIP path to perform the lateral diffusion of the OFF bipolar cell.

3.2.4 Amacrine and Ganglion cells

The circuit of the amacrine and the ganglion cell is shown in Fig. 3. 2(d). The block Ama is constructed using the transistors M_{A1} - M_{A10} . The transistor M_{A1} receives I_{OFFBIP} , shown in Fig. 3. 2(d). The current is laterally diffused by a smoothing network at the drain node of transistors M_{A1} and the transistors M_{A1} - M_{A4} is a cascode current mirror used to repeat the current. Then this diffused current is inversely repeated by another cascode current mirror composed of transistors M_{A5} - M_{A8} and is sent out of the chip via M_{A11} as I_{Ama} in proper polarization. Finally, the current is inversely repeated again to enter the block GC by the cascode current mirror composed of transistors M_{A9} , M_{A10} , M_{G1} , and M_{G2} . This current mirror is also designed to implement the gain block after the block Ama, as shown in Fig. 3. 2. The block GC receives I_{ONBIP} as shown in Fig. 3. 2(c) and subtracts it from the current mirrored using the transistors M_{A9} , M_{A10} , M_{G1} , and M_{G2} . The subtracted current is sent to the sub-block ABS composed of the transistors M_{G3} - M_{G9} to be positively rectified. Thus, only the signal on the positive side is repeated and sent to the output by M_{G9} as I_{GC} .

3.2.5 Impacts of device leakage and mismatch

Since we use many current mirrors in our circuit, device mismatched may affect the performance of the chip. However, the goal of our chip is to reproduce response of the real retina qualitatively, and therefore the parameters, time constants and space constant, are not necessary to be accurate values. As long as the characteristics of the response of the chip. Overshooting and undershooting or smoothing function, are ensure to be correct, device mismatched could be tolerated.

Since the photocurrent induced by light is rather small compared to the leakage current of the devices, the input signal is applied, the output current might be affected by leakage currents or noise. The output current might not be observable consequently.

3.3 ARCHITECTURE OF 2-D RETINAL ARRAY

The architecture of the fabricated chip is shown in Fig. 3. 4. The chip consists of an array of 32x32 pixels and address decoders.

A number of identical basic cells are arranged to form a rectangular array. Two decoders are used to control the output of the cell. The row decoder decodes the m row address bits and activates one of 2^m row control signals. The column decoder decodes the n column address bits and activates one of 2^n column control signals. In our design, a 32x32 2-D retina array is designed, and therefore two 5-bit addresses are needed for row and column address to select one cell from all of 32x32 cells. The circuit of the address decoder is shown in fig. 3.5. The fundamental blocks of the address decoder circuit is the five-input and gate. The inputs of each and gate are five non-inverting or inverting address bits, and none of arbitrary two input configurations are the same because an address can only activates one row or one column. The selected address line is set to logic high or the most positive voltage supply of the chip, that is, VDD.

The pixel circuit mentioned above is arranged in a 32x32 array. A row decoder and a column decoder are used to select the pixel that is to be observed. Every time there is only one pixel that can be observed, and there are six observable signals in each observed pixel: I_{PH1} , I_H , I_{ONBIP} , I_{OFFBIP} , I_{Ama} , and I_{GC} as shown in Fig. 3. 2(a)-(c). The six signals of each pixel are connected to six output channels separately through row and column switches. Thus, only signals of the selected pixel reach the output channels. Using this method, the behavior of each pixel, as well as that of the whole array, can be observed by setting different controlling strategies using the row and column decoders.

3.4 SIMULATION RESULTS

The circuit in Fig. 3. 2 is used to construct a 32x32 array which is simulated by HSPICE. Since the array size is differs from the original definition of the space constants D, all the space constants in the simulation are shrunk with the same factor, 32/180. The simulation results are shown in Fig. 3.

6 ~ Fig. 3. 9. In this simulation, a voltage pulse is applied to the 5x5 cells in the center of the array to imitate the 500Hz flashing light stimulus and the spatiotemporal patterns of the 17th row are observed.

Fig. 3.6 is the monte carlo analysis of OFF bipolar cell with the original V-I converter structure and the modified linear transconductor structure. Monte carlo analysis target on MOSFETs of channel width and threshold voltage of Rm amplifier, band-pass filter, and transconductor. The simulation result of the original structure shows that the variation of output current level is 8mA while the variation of the modified circuit is less than 0.04mA. An OFF bipolar cell block features low DC current variation is proposed to solve the DC level variation problem in the previous work and to increase the stability of the circuit and the chip.

Then we show the transient response of the circuit when a pulse-stimulus is supplied. Fig. 3. 7 shows the output currents of the center cell of the 2-D array. The simulation is performed under the condition that a pulse signal with a turned-on duration of around 1ms is incident on the middle six cells of the array. The simulated input to the middle six cells are periodic stimulus with 100pA transient current added to a 100pA background-induced DC current. All other twenty-six cells are supplied with background currents of 100pA. All above-mentioned currents are supplied to the bases of the photo-BJT of all cells. The output of photoreceptor in Fig 3. 7(a) has overshooting and undershooting at the turn-on and turn-off similar to the CNN model simulation results in the previous section. It is also clear in Fig. 3. 7(b) that the output of the horizontal is analogous to that of the PH1 but has a larger magnitude.

As mentioned in the previous section, the gate bias of M_{SM11} , M_{SM12} , M_{SM21} , M_{SM22} , M_{SM31} , M_{SM32} , M_{SM41} , M_{SM42} , M_{SM51} , M_{SM52} , M_{SM61} and M_{SM62} , V_{SM} , controls the diffusion range of the horizontal, on bipolar cell, off bipolar cell, amacrine cell and ganglion cell, respectively. Fig 3. 8 shows the steady state output of space domain response with subjected to suitable gate bias voltages of V_{SM} . Fig 3. 8(b) is the results of the horizontal when the middle six cells are incident with stronger light while the other cells are incident with background light. As could be seen in the figure,

a higher gate biasing voltage, VSM, causes lower resistance of the smoothing network, and a wider diffusion range is achieved. The results of photoreceptor with the same stimulus are shown in Fig 3. 8(a). As discussed previously, the edge of the incident pattern would have higher contrast in the output of PH1 but the extent of contrast varies with different diffusion range.

In Fig. 3. 9, the x-axis is time and the y-axis is the pixel location which denotes space. The black and white bars denote the spatial and temporal region of the input stimulus. The stimulus is applied to the 15th to the 20th pixel from 1001 μ sec to 2000 μ sec. The waveform at the right of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. Fig. 3. 9(a) and (b) represent the spatiotemporal patterns of the photoreceptor and horizontal cell, respectively. It can be seen from Fig. 3. 9(a) that the photoreceptor's signal level drops when there is a stimulus, and it returns to its original level when the stimulus disappears. Slight undershooting and overshooting in temporal domain can be expected in the periphery as it reacts to the stimulus directly. In the spatial domain, there is strong contrast at the edge of the stimulus. At the edge pixels inside the stimulus, the signal level is higher than the other stimulated pixels. Contrarily, at the edge pixels where the stimulus is just absent, the signal level is higher than the other silent pixels. Therefore, the spatial range of the stimulus can be well defined. In the temporal domain, the horizontal cell has a similar reaction to the photoreceptor, as can be seen in Fig. 3. 9(b). However, since the horizontal cell performs lateral diffusion in space, its spatial domain waveform spreads wider than that of the photoreceptor. Fig. 3. 9(c) and (d) represent the spatiotemporal patterns of the ON and OFF bipolar cells, respectively. The OFF bipolar cell performs bandpass-filtering on the signals from the photoreceptor. In the amacrine cell, the rectified signals from both bipolar cells are added with different gain factors, as shown in Fig. 3. 9(e). Moreover, the response to the appearance of the stimulus is weaker than the response to the disappearance of the stimulus. Thus the amacrine cell provides a strong inhibition to the ganglion cell when the stimulus disappears in time. Therefore, the ganglion cell exhibits clear turned-ON

reaction, as can be seen in Fig. 3. 9(f).



Table II

The transistor's sizes of circuits in Fig. 3. 2(a)-(d)

| Transistor numbers | W/L ($\mu\text{m}/\mu\text{m}$) | Transistor numbers | W/L ($\mu\text{m}/\mu\text{m}$) |
|--|--------------------------------------|------------------------------------|--------------------------------------|
| M1, M2 | 1.5/0.8 | M3, M4, M5 | 2/0.8 |
| M6 | 1/0.8 | M7, M9 | 2/0.4 |
| M8, M10, M12, M14- M18 | 4/0.4 | M11, M13, MRM2 | 1/0.4 |
| MPH11~MPH14, MH1~MH16, MO1~MO4, MCM1~MCM4 | 4/0.35 | MCM13~MCM16, MON35, MVI7~MVI8 | 2/1 |
| MPH21-MPH24, MO5~MO6 | 8/0.7 | MPH15~MPH18 | 16/0.35 |
| MPLP1, MPLP3, MPLP9, MPLP11 | 0.4/0.35 | MPLP2, MPLP4, MPLP10, MPLP12 | 1.2/0.35 |
| MPLP5, MPLP7, MLPON5, MLPON7 | 0.4/20 | MPLP6, MPLP8, MLPON6, MLPON8 | 1.2/20 |
| MCM5~MCM8, MON1~MON34, MOFF1~MOFF7 | 1/1 | MA1~MA6, MA9~MA11, MG3, MG6~MG9 | 2/2 |
| MCM9~MCM12 | 4/1 | MLPON1, MLPON3 | 0.4/1.5 |
| MLPON2, MLPON4 | 1.2/1.5 | MRM1 | 3/0.4 |
| MRM3 | 2.25/1 | MRM4 | 0.7/1 |
| COFF | 0.01pF | MHP1 | 2.25/4 |
| MHP2 | 0.7/4 | MPLP13 | 0.4/1 |
| MPLP14 | 1.2/1 | MVI1~MVI2 | 1/10 |
| MVI3~MVI4 | 5/1 | MVI5~MVI6 | 0.5/4 |
| MVI9~MVI10 | 1/5 | MA7~MA8, MG1~MG2 | 4.4/2 |
| MG4 | 3/1 | MG5 | 6/1 |

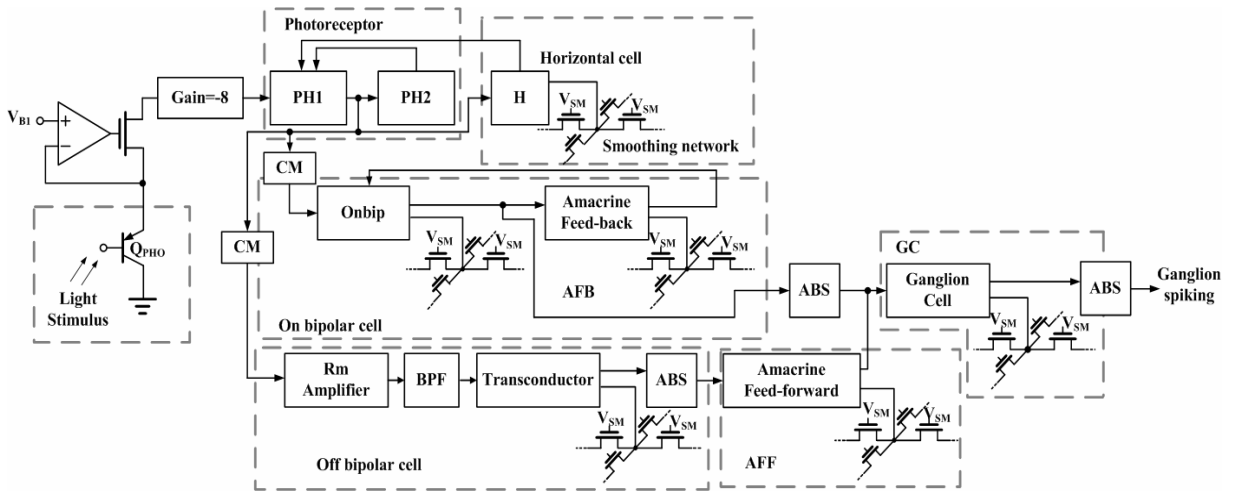
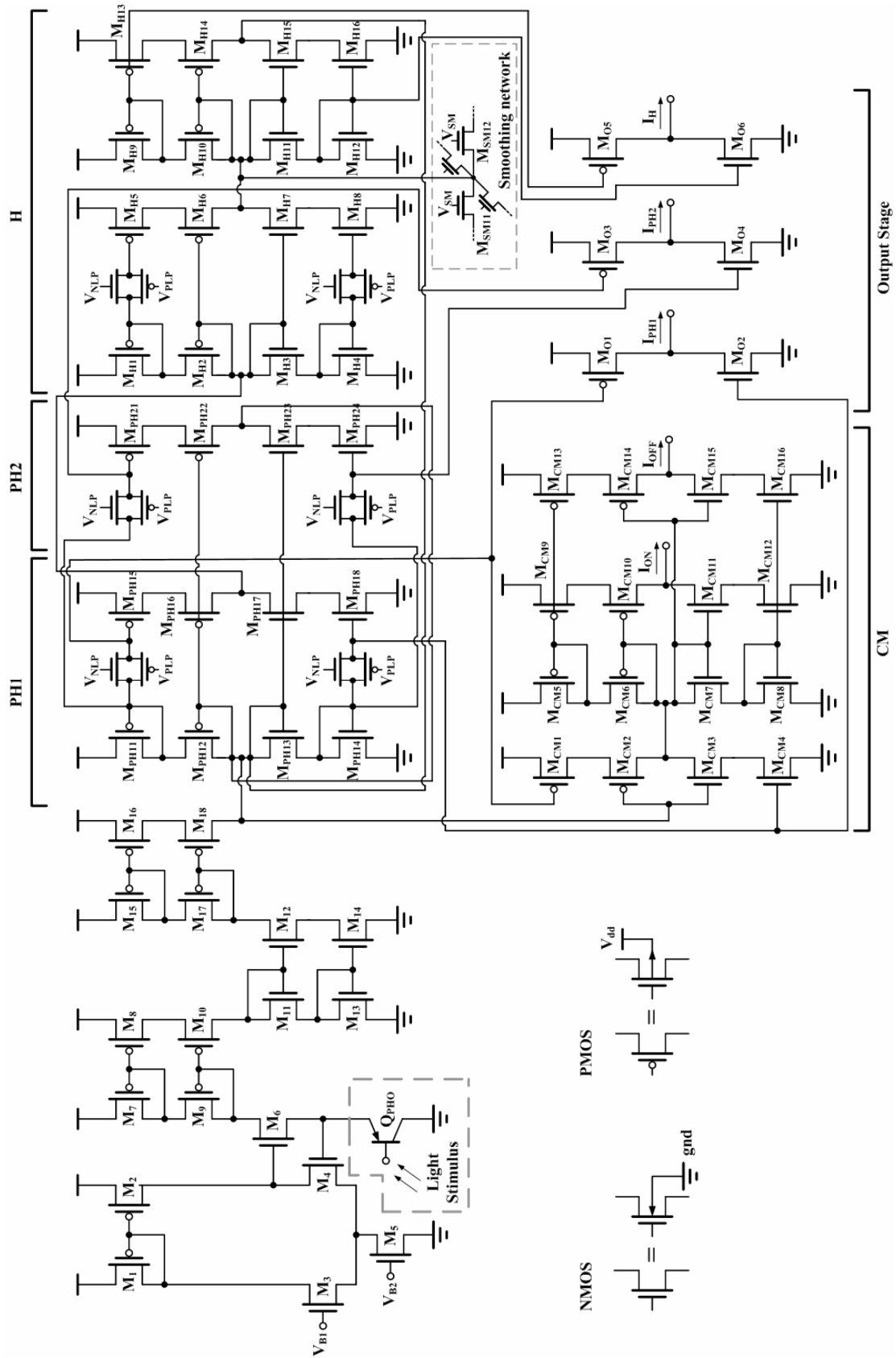


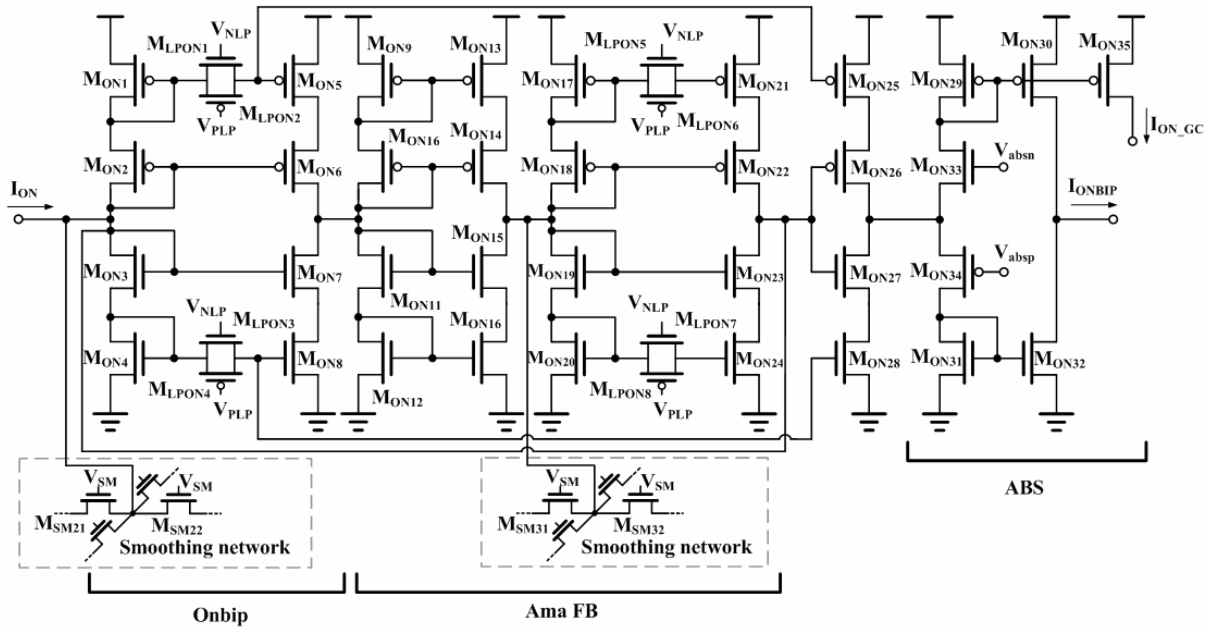
Fig. 3. 1. The block diagram of a single pixel.





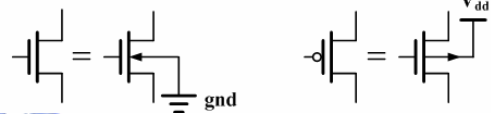
(a)

On bipolar cell



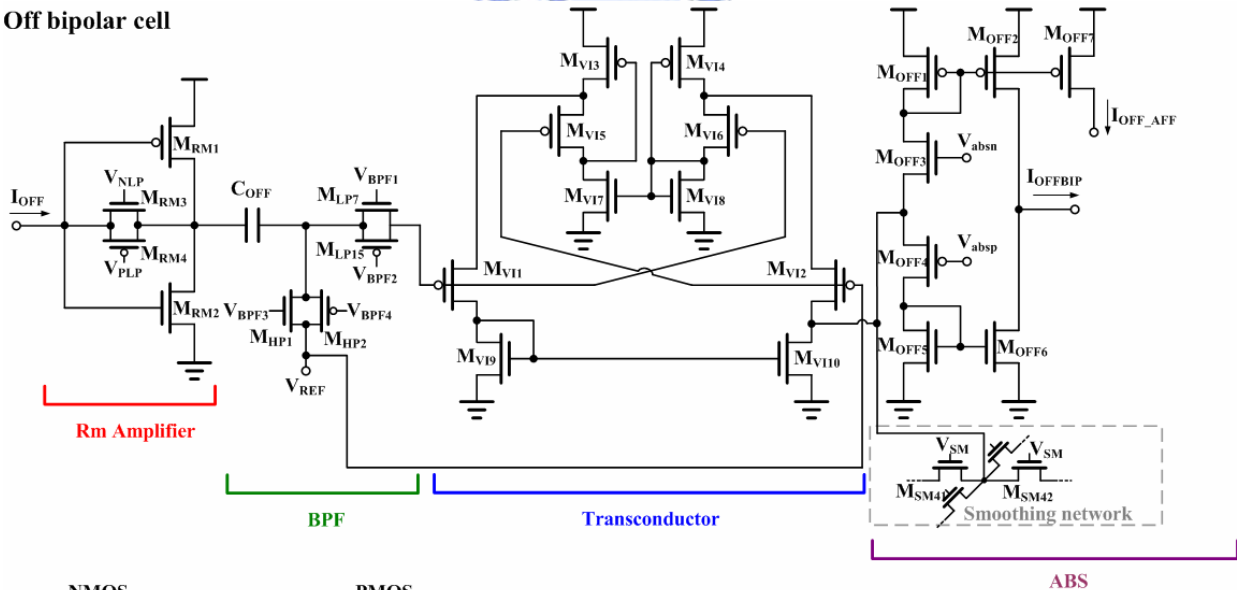
NMOS

PMOS



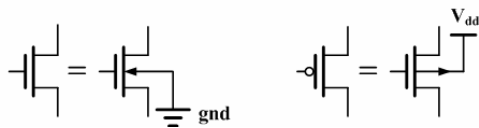
(b)

Off bipolar cell

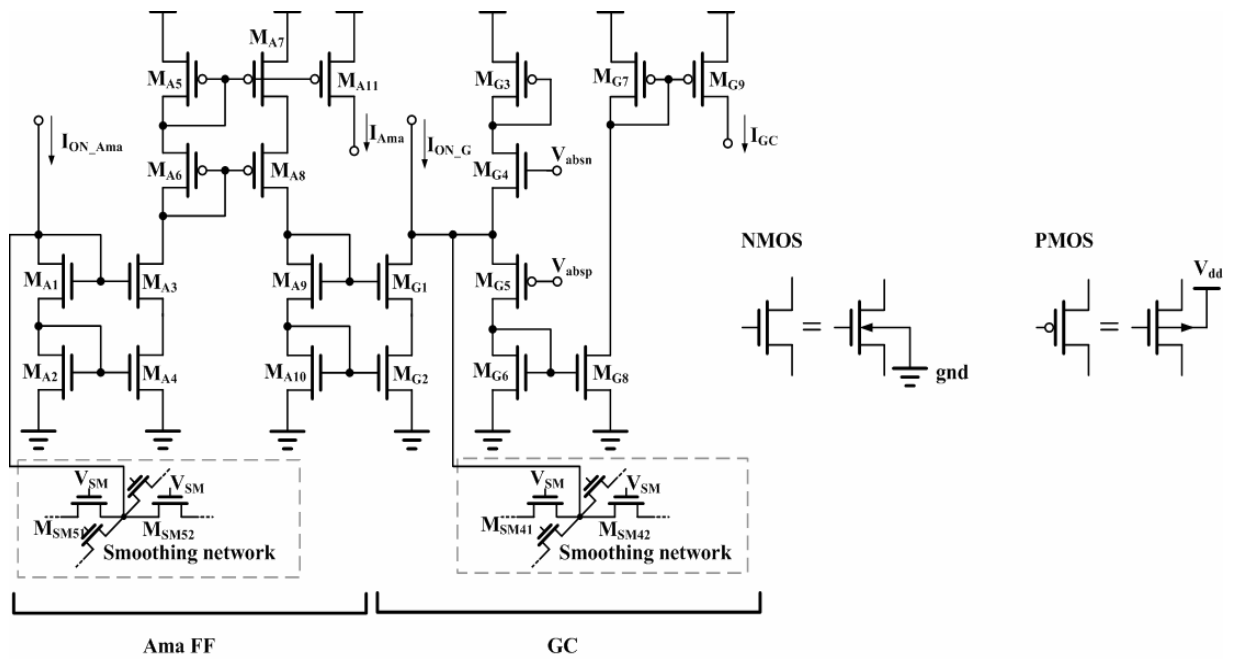


NMOS

PMOS



(c)



(d)

Fig. 3. 2. (a) The circuit of the photoreceptor and horizontal cell (PH1, PH2, H); (b) The circuit of ON bipolar cell (ONBIP); (c) OFF bipolar cell (OFFBIP); (d) The circuit of amacrine and ganglion cells (Ama, GC).



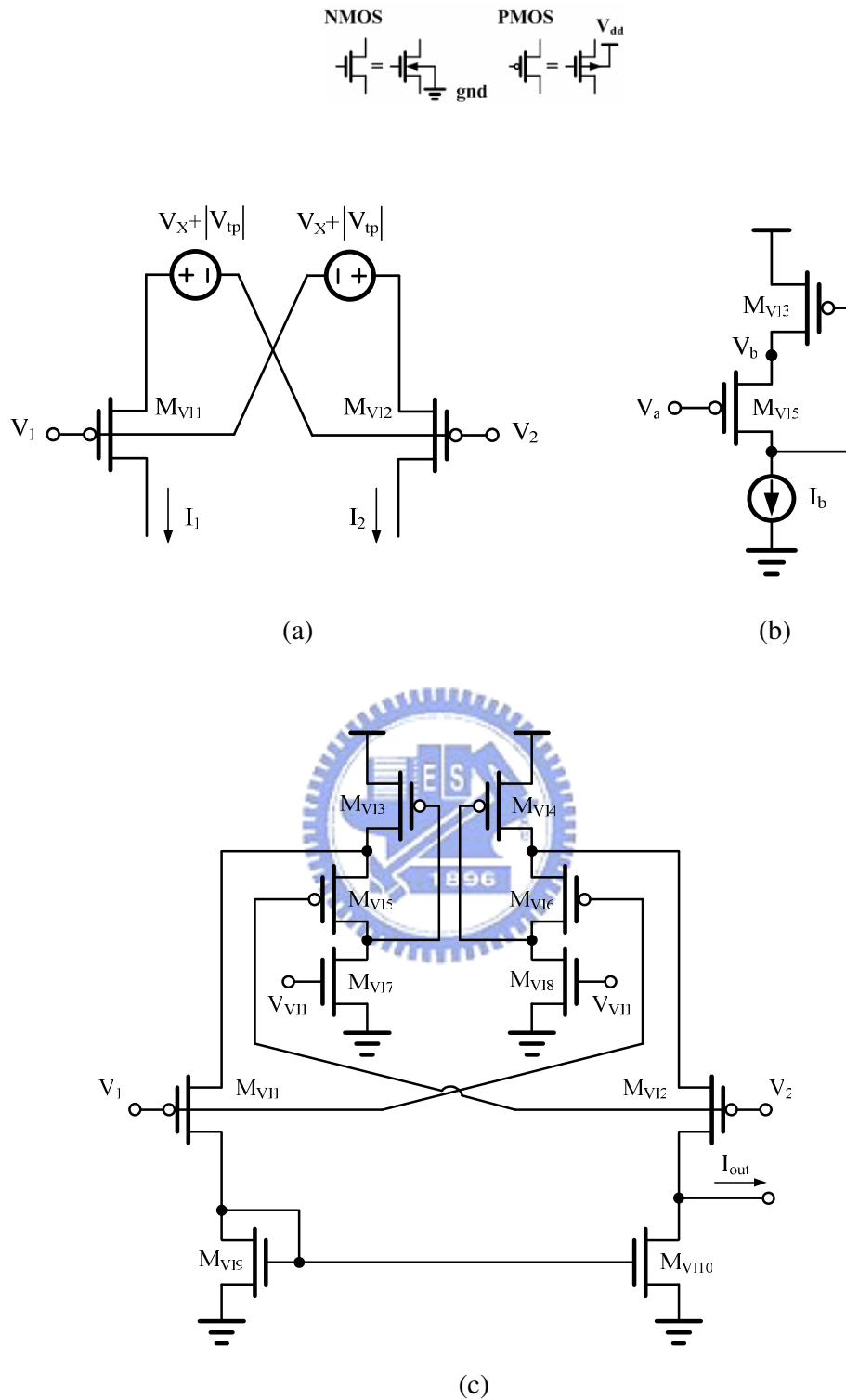


Fig. 3. 3. (a) A linearized transconductor using floating voltages sources; (b) Flipped Voltage Follower implementation; (d) Transconductor using two identical FVFs to implement the floating batteries.

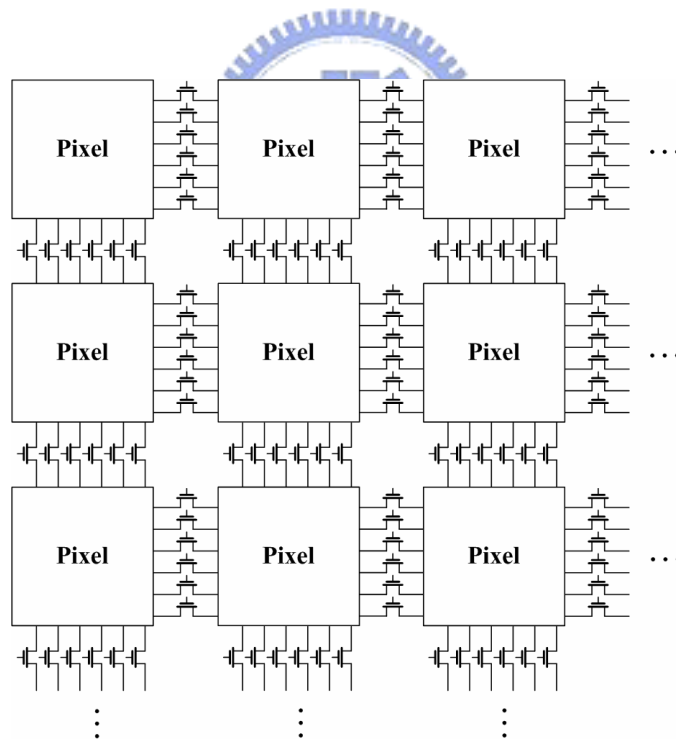
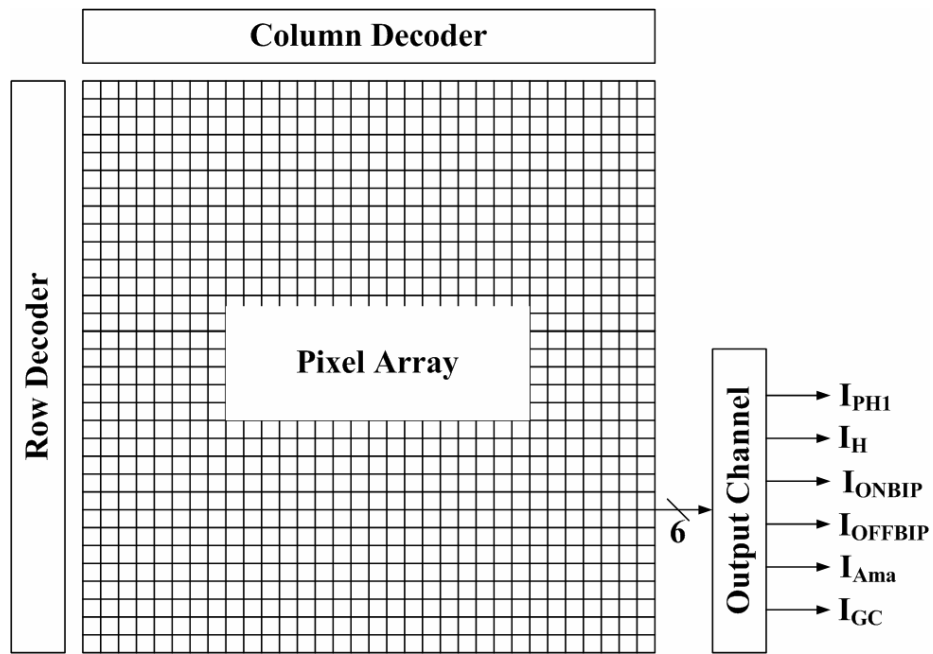
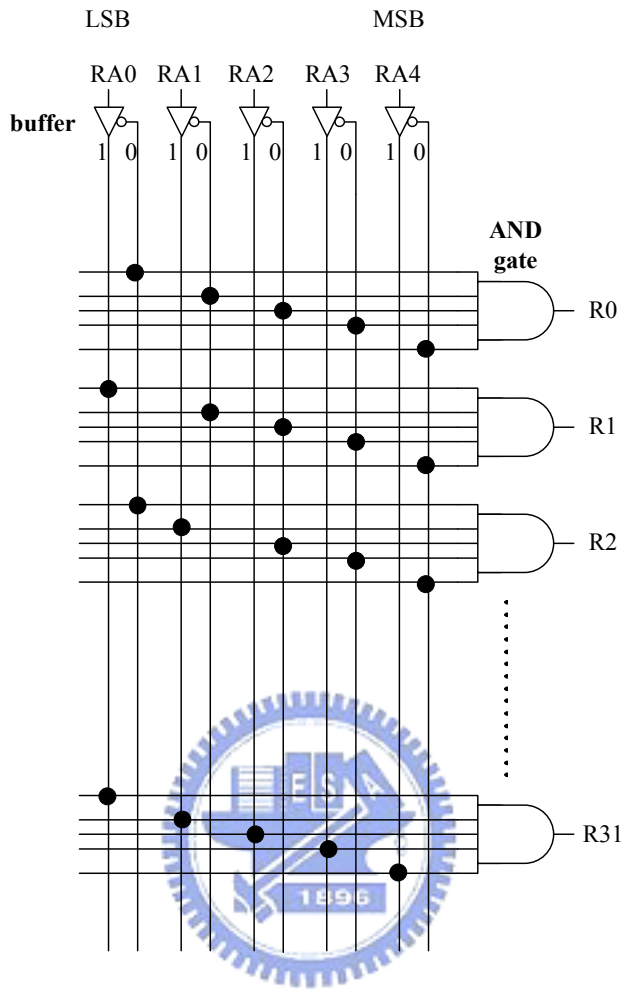
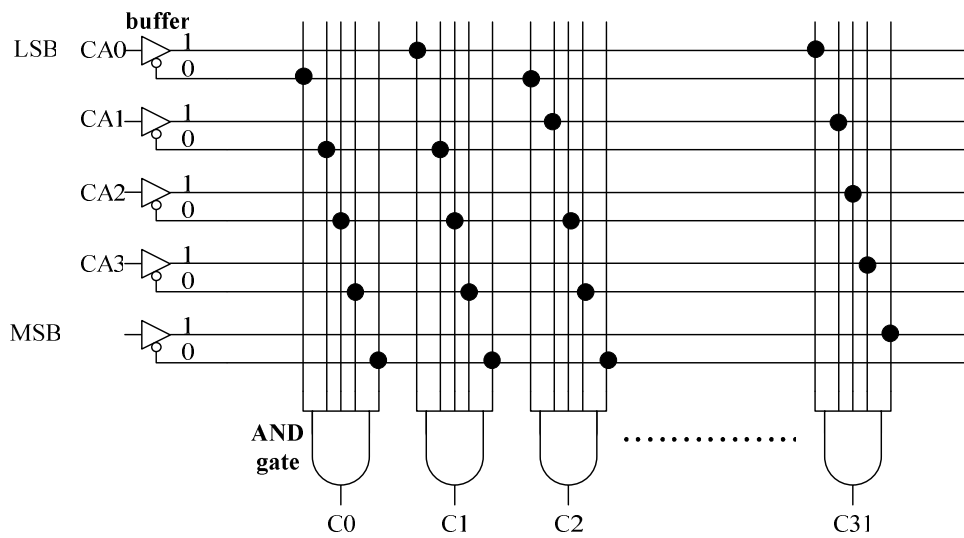


Fig. 3. 4. (a) The whole chip architecture of the implemented chip and (b) the implementation of smoothing networks.

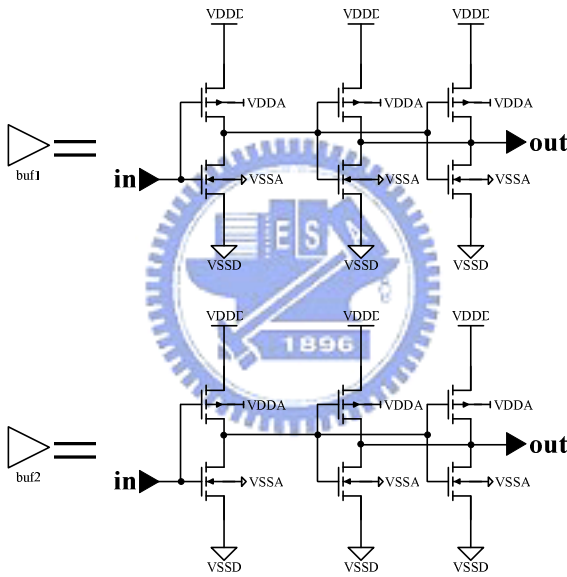
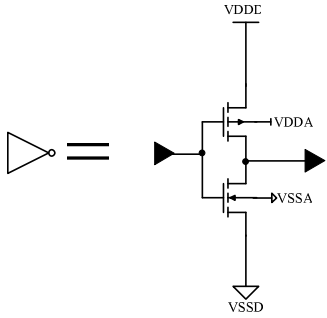
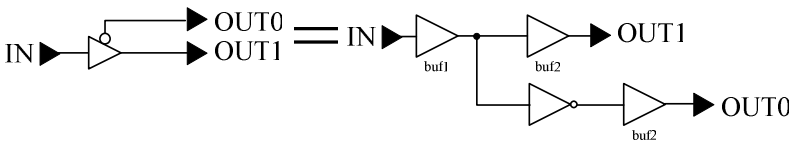


(a)



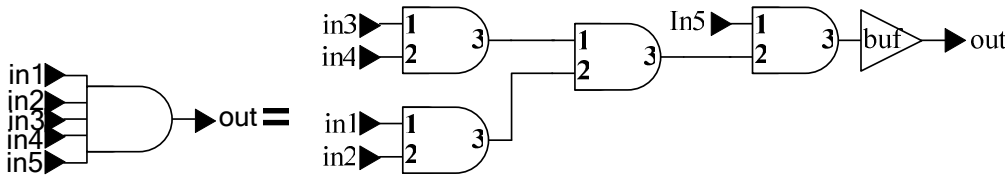
(b)

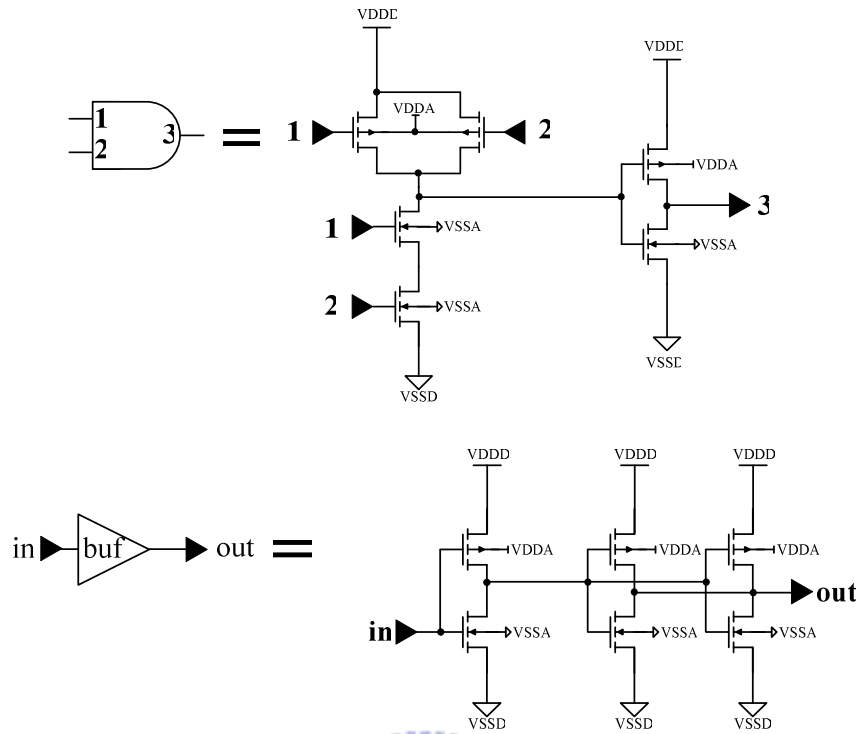
Buffer:



(c)

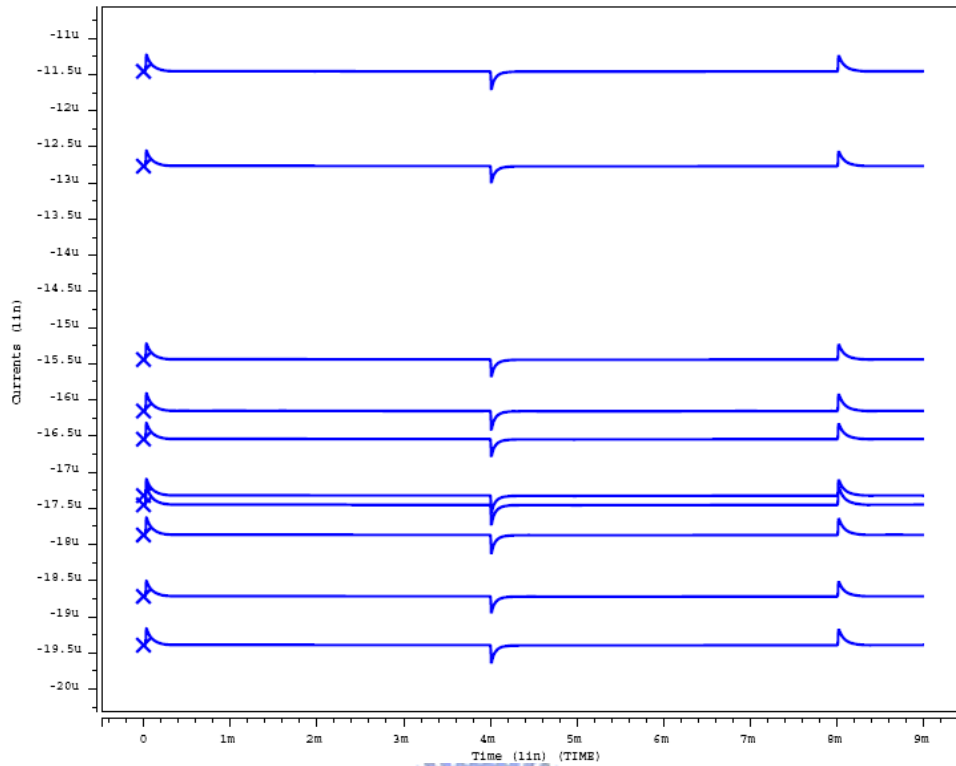
AND gates:



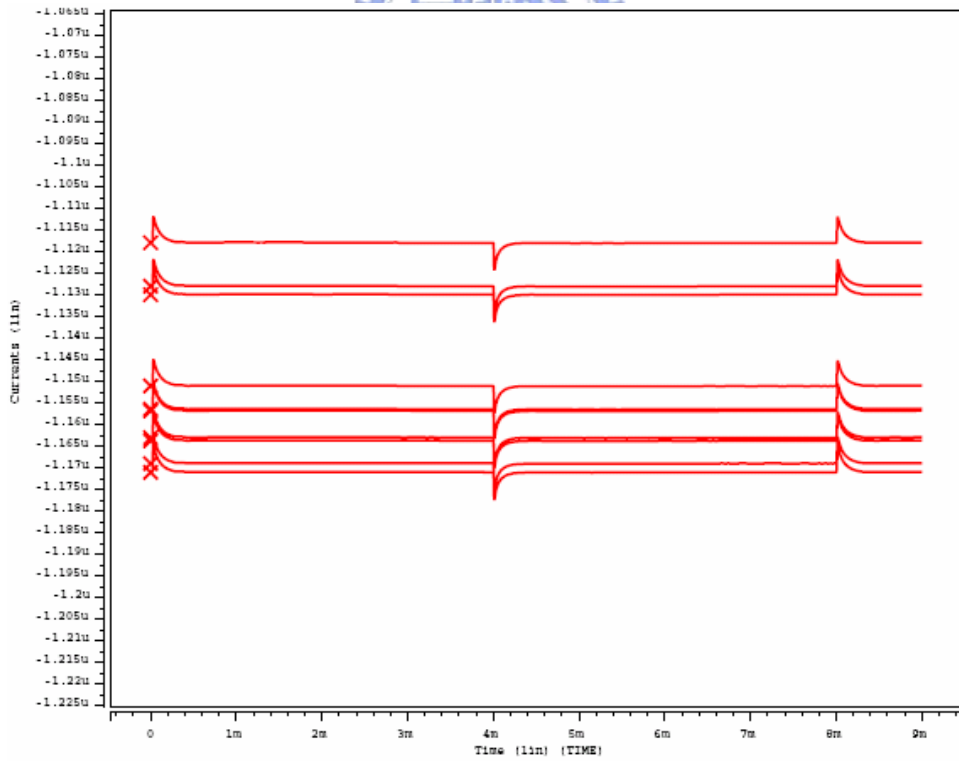


(d)

Fig. 3. 5. The schematic of the address decoder used in the chip (a) column address decoder; (b) row address decoder; (c) buffer; (d)AND gate.



(a)



(b)

Fig. 3. 6. The monte carlo analysis of OFF bipolar cell with (a) the original V-I converter structure (b) the modified linear transconductor structure.

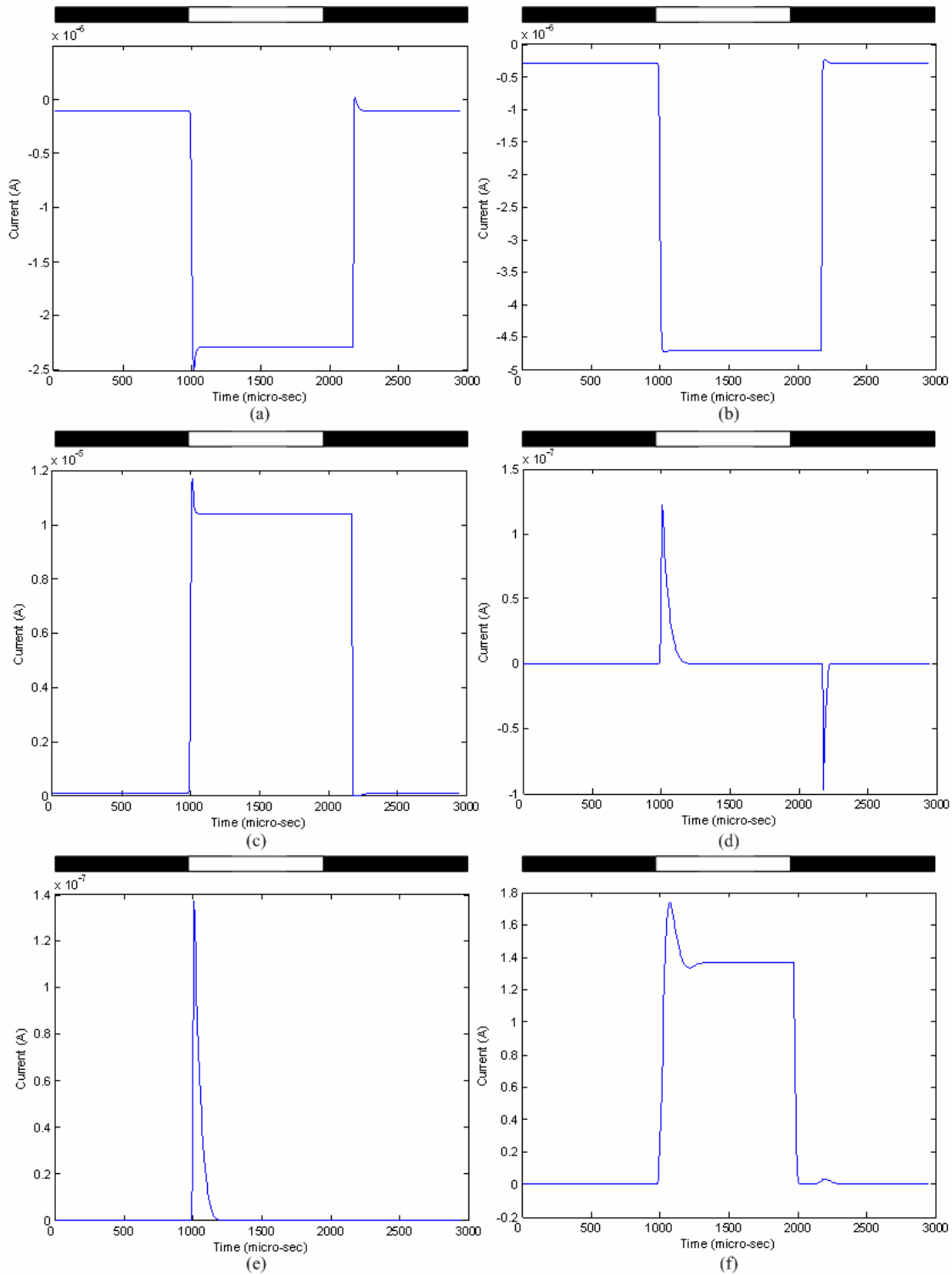


Fig. 3. 7. The HSPICE circuit-simulated time domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. The light stimulus is applied at 1000th~2000th micro-second and $V_{sm} = 2.2V$.

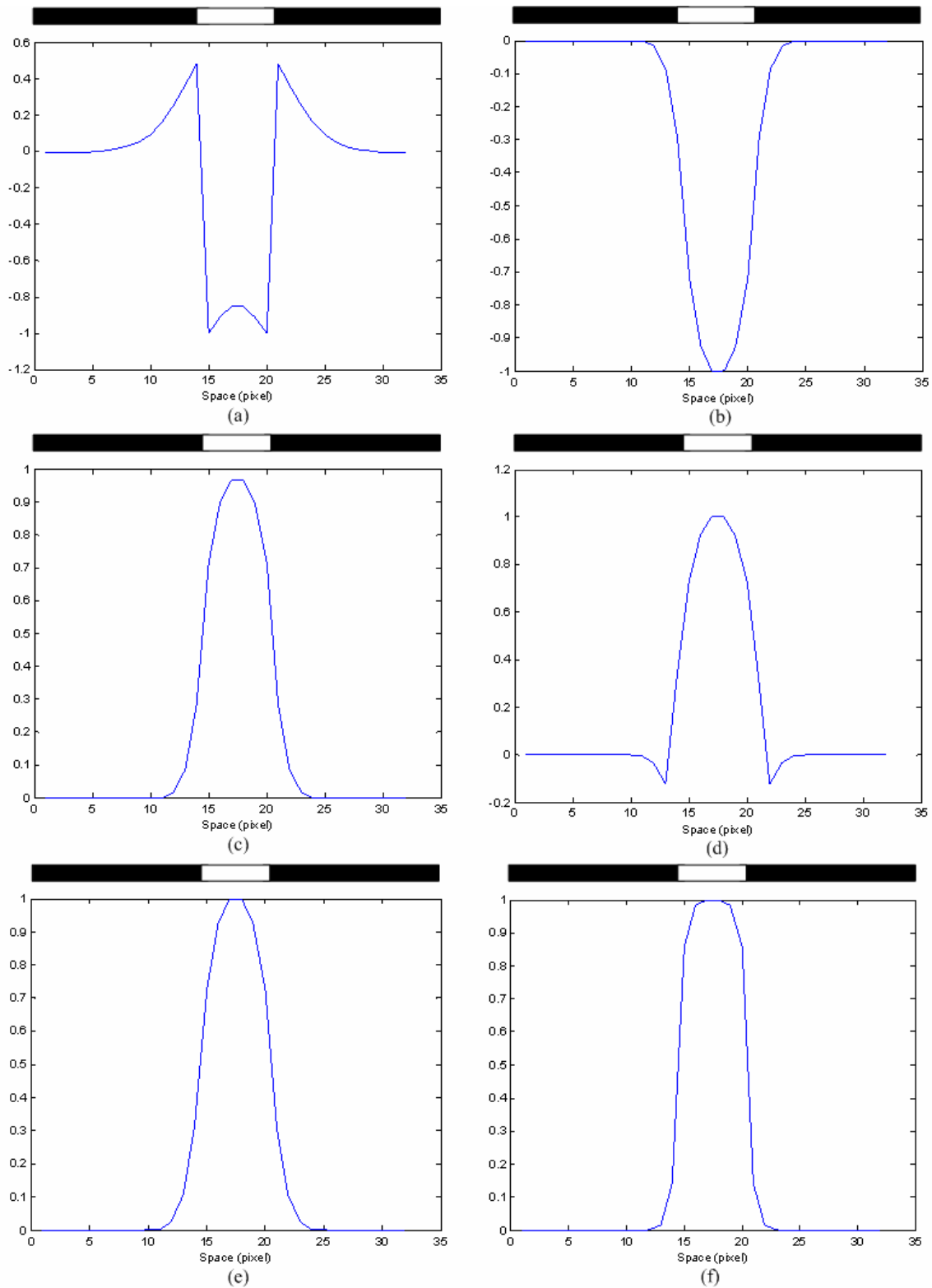


Fig. 3. 8. The HSPICE circuit-simulated space domain waveforms of the neuromorphic model of the ON sluggish sustain cell set in a 32x32 array for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. The light stimulus is applied at 15th~20th pixel and $V_{sm} = 2.2V$.

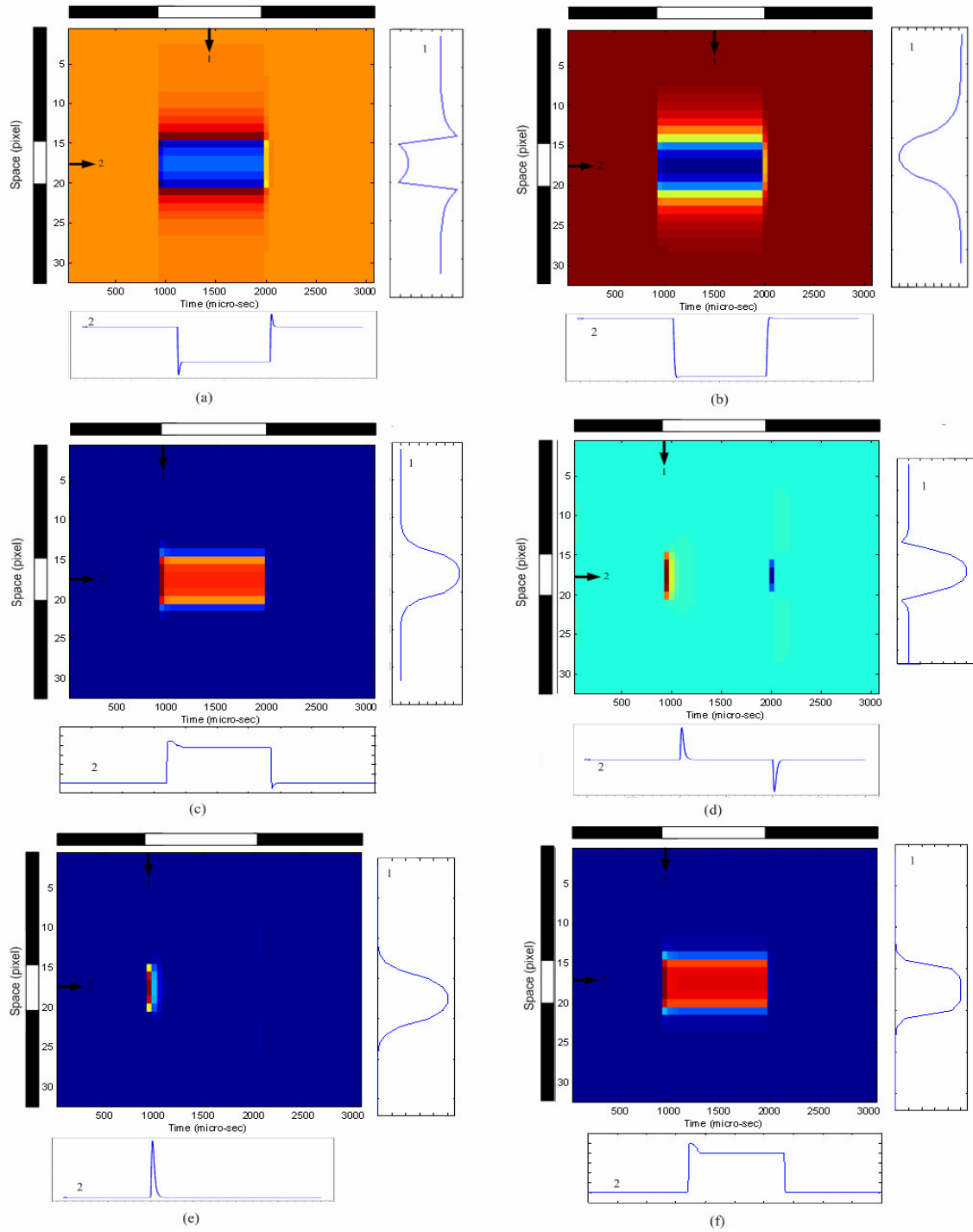


Fig. 3. 9. The HSPICE circuit-simulated spatiotemporal patterns for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These waveforms are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 15th to the 20th pixel at time point 1001 to 2000. The waveform at the left of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. $V_{sm} = 2.2V$.

CHAPTER 4 Layout Descriptions and Experimental Results

In this chapter, the layout and chip photographs of the retinal chip is described in section 4.1. The experimental environment is illustrated in sections 4.2. Finally, the experimental results are shown in the last section to verify the functions of the chip.

4.1 LAYOUT DESCRIPTIONS

The retinal chip is designed and fabricated in 0.35 μm double-poly quadruple-metal CMOS process. The architecture and circuit are described in detail in the previous chapter. The layout of the chip is shown in Fig. 4. 1 and Fig. 4. 2.

The retinal chip contains a sensory array of 32x32, address decoders and other circuits. Fig. 4.1 shows the layout of whole retinal chip. The sensory array is in the center and the input and output pads are arranged in its peripheral. The row address decoder and the column address decoder are on the left side and the upper side of the sensory array respectively as labeled in the figure. Total area of the chip is around 4.3mm x 4.2mm. ESD (Electrostatic Discharge) protection circuit for input and output pads are included in the chip.

The layout of a basic cell in the chip is shown in Fig. 4. 2. In each basic cell, there are ninety-seven NMOS devices, eighty-three PMOS devices, and a parasitic PNP phototransistor, including six NMOS switches. Each basic cell occupies an area of 101 μm x 100 μm with a fill factor of 9.14%. The region of the photo-BJT is labeled in the figure. It is surrounded by dashed line in the lower-left of the figure. Except the phototransistor region, the other region of the basic cell is covered by metal layers to prevent the incident light from affecting the other part of the circuit. The phototransistor region which is not covered by metal layers is transparent and light could pass through easily. The locations of other parts of the cell are shown in the figure as labeled.

A photograph of the whole chip and the basic cell are shown in Fig. 4. 3 and Fig. 4. 4, respectively. In Fig. 4. 4, the region of six main part of a basic cell, photo-input, photoreceptor and horizontal cell, on bipolar cell, off bipolar cell, amacrine and ganglion cells, and photo-BJT are labeled. Only the area of photo-BJT is not covered by metal layers and is transparent.

4.2 EXPERIMENTAL ENVIROMENT

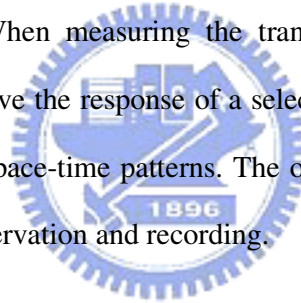
The chip is designed using the TSMC 0.35 μ m double-poly-quadruple-metal standard CMOS technology. The image/photograph of the whole chip is shown in Fig. 4. 3 and that of a single cell is shown in Fig. 4. 4. The whole chip area is 4.3mm x 4.2mm. The fill factor of the photo-BJT is 9.14%. It can be seen from Fig. 4. 3 that the pixel array occupies most of the chip area. The row decoder and column decoder are on the left and top of the chip. The ESD pads are used to protect the chip from electrostatic damages. As can be seen in Fig. 4. 4, there is a photo-BJT in the left to transduce light stimuli into photocurrents. Since the designed chip sends pixel signals in current format, external current-to-voltage converters are needed to facilitate measurement. The read-out circuit of each output is shown in Fig. 4. 5. Since there are six output signals generated by this chip, as shown in Fig. 3. 4, there are six similar read-out circuits. The circuit contains one operational amplifier (OP AMP), a resistor, and a LPF composed of a resistor and a capacitor. The operational amplifier with a negative feedback provides a virtual bias at the negative input node, which is connected to the chip's output node. The value of the virtual bias is determined by V_B . The resistor R_{OUT} converts the currents into voltages. Therefore, the current flowing out from the chip I_{OUT} will be transduced into V_{OUT} through the readout circuit. Assuming the voltage gain of the operational amplifier is A , then the relationship between I_{OUT} and V_{OUT} is:

$$V_{out} = \left(\frac{A}{A+1} \right) (-I_{out}R_{out} + V_B) \quad (15)$$

The voltage-mode signals V_{OUT} are actually observed and recorded. The measured data are presented and analyzed below. Considering the different signal swings of the output currents, the

R_{OUT} for I_{PH1} and I_{ONBIP} are set as $1M\Omega$, I_H and I_{GC} are set as $0.5M\Omega$, while the R_{OUT} for I_{OFFBIP} and I_{Ama} are set as $4M\Omega$ to optimize experimental conditions. The optimization of the experimental condition will yield the largest but unsaturated signal levels. A white light LED is used to generate flashing light to stimulate the test chip. The LED is controlled by a function generator, so that the amplitude and frequency of the incidental light can be monitored. However, the input stimuli are always repetitive because of the inherent nature of the function generator. An optical lens is used to concentrate the light from the LED onto the test chip, thus only a small part (around 5×5 pixels) of the chip is stimulated while the other part remains relatively dark.

The diagram of the measurement setup is demonstrated in Fig. 4. 6. The projector and the LED are used to project. By controlling the function generator that drives the LED, we could control the flash frequency and luminance of the light. Via a convex lens, flashing light of the LED produces the light spot focusing on the chip. When measuring the transient response, the row and column addresses are kept the same to observe the response of a selected cell. Additional counter is needed to measure the spatial response or space-time patterns. The output of the chip are connected to the terminals of the oscilloscope for observation and recording.



4.3 EXPERIMENTAL RESULTS

The bias condition for measuring the spatiotemporal patterns is listed in Table III. The control voltage of smoothing network is slightly modified to $2.4V$ to obtain the obvious temporal waveform and spatial edge of light stimulus. The flashing frequency of white light of LED is $500Hz$ and the light-induced photocurrent is around one hundred-pA to several hundred-nA. To verify that the measured output of the chip is consistent with the HSPICE circuit simulation results and the original CNN model, a normalized output of each part of the cell and simulation curves are plotted in the same drawing for comparison. The measured transient response comparison with model and circuit simulation are shown in Fig. 4.7 ~ Fig. 4.9. The stimulus is applied to the 15^{th} to the 20^{th} column and

15th to the 20th row. The different location on the chip of light stimulus for the measured transient response is the same because of the smoothing network is connected continuous without disconnected. Fig. 4.7 ~ Fig. 4.9 show the results of measurement when a periodic light source is incident on the chip. In the figure, overshooting and undershooting of photoreceptor could be observed clearly and so could that of the horizontal as simulated previously. It could be found from the figure that the measured curves match the simulation curves qualitatively. The measured spatial response comparison with model and circuit simulation are shown in Fig. 4.10 ~ Fig. 4.12. As mentioned in the previous discussion, the smoothing or diffusion range could be tuned by controlling bias V_{sm} . A large V_{sm} makes a wider smoothing range. In Fig. 4.10 (a), the measurement result is not as perfect as that simulated waveform because of the imperfection of the light stimulus. However, the spatial edge of light stimulus still can be recognized. Measured transient response with large light stimulus interval (200msec~400msec) are shown in Fig. 4.13. With large light stimulus interval, the output signals of ON bipolar cell and ganglion cell decrease when light stimulus large than 310msec.

The normalized measured spatiotemporal patterns of the chip with subjected to periodic signal are shown in Fig. 4. 14. The light stimulated region is from the 19th pixel to the 23rd pixel in space, and from the 1001 point to the 2000 point in time. At the bottom of each pattern is the temporal domain waveform recorded at the 21st pixel, and at the right is the spatial domain waveform(s) recorded at the time point(s) marked by the vertical arrow(s). In Fig. 4. 14(a), (b) and (c), the spatial domain waveforms are recorded at the time point 900. In Fig. 4. 14(d), (e) and (f), the spatial domain waveforms are recorded at the time points 1000 and 2000. The effects of the spatial diffusions can be seen in Fig. 4. 14(b)~(f).

Fig. 4. 14(a) illustrates the spatiotemporal pattern of the photoreceptor. The spatiotemporal pattern of the horizontal cell in Fig. 4. 14(b) covers a wide area of space. The phenomenon can be seen clearly from its spatial waveform represented at the right of Fig. 4. 14(b). Because of this phenomenon, the effects of inter-pixel variation and imperfection of the light stimulus are not as

strong as those presented in Fig. 4. 14(a).The spatiotemporal patterns of the ON and OFF bipolar cells are shown in Fig. 4. 14(c) and (d). Both cells are affected by the photoreceptor. They responds to both appearance and disappearance of the light stimulus in time, as can be seen at the bottom of Fig. 4. 14(c) and (d). The pattern of the amacrine cell, as shown in Fig. 4. 14(e). The spatiotemporal pattern of the ganglion cell in Fig. 4. 14(f) only responds to the appearance of the light stimulus in time. The patterns are not as perfect as that in Fig. 3. 8 because of the inter-pixel variation and imperfection of the light stimulus. However, the spatial edge of light stimulus still can be recognized by the contrast between grey and white rows.

Table IV summarizes a brief comparison on the measurement results of proposed neuromorphic chip and the previous work [20] to show the improvements from the proposed structure. The maximum inter-pixel variations of DC current output is recorded from the measurement results of OFF bipolar cell, which is less than $0.26\mu\text{A}$. In this thesis, the ON bipolar cell and OFF bipolar cell block is different from the previous work [20], so the chip total power consumption of this work is decreased even though its power supply is slightly larger than previous work [20].

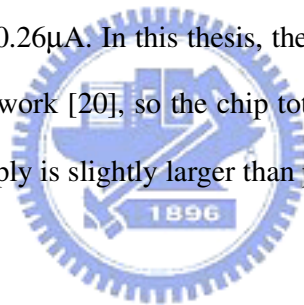


Table III

The bias condition to measure the spatiotemporal patterns of Fig. 9.

| | | | | | | |
|------------|------------|------------|------------|--------------------|------------|------------|
| V_{DD} | V_{B1} | V_{B2} | V_{NLP1} | V_{PLP1} | V_{NLP2} | V_{PLP2} |
| 3.3V | 0.326V | 0.244V | 1.7V | 1.6V | 1.7V | 1.6V |
| V_{absn} | V_{absp} | V_{NLP3} | V_{PLP3} | V_{BPF1} | V_{BPF2} | V_{BPF3} |
| 2.13V | 0.50V | 1.91V | 0.27V | 1.8V | 0.5V | 0.62V |
| V_{BPF4} | V_{REF} | V_{VII} | V_{SM} | Flashing Frequency | | |
| 1.9V | 1.38V | 1V | 2.4V | 500Hz | | |



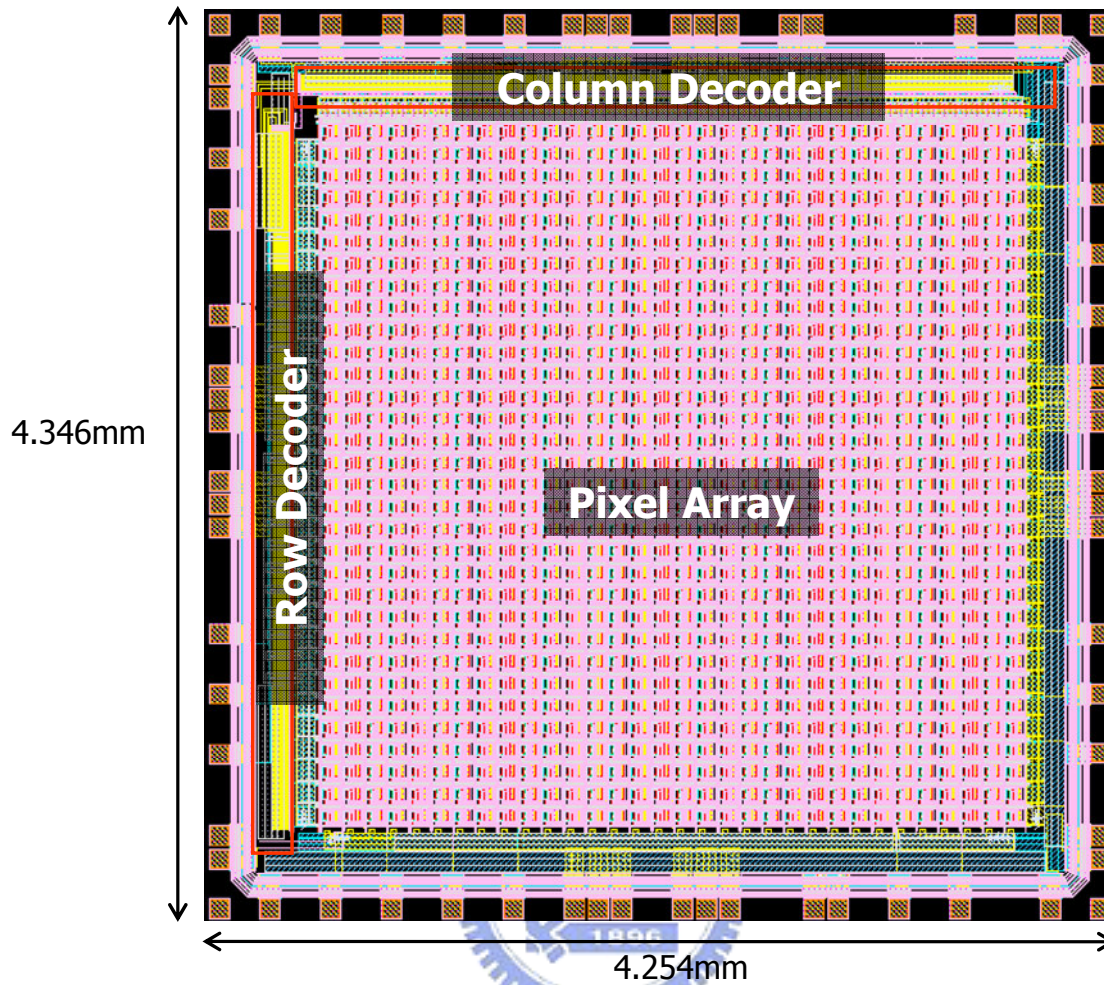


Fig. 4. 1. Layout of the whole retinal chip. Main parts of the chip are labeled. Sensory array is in the center, the row and column address decoders are in the upper and left of the array respectively.

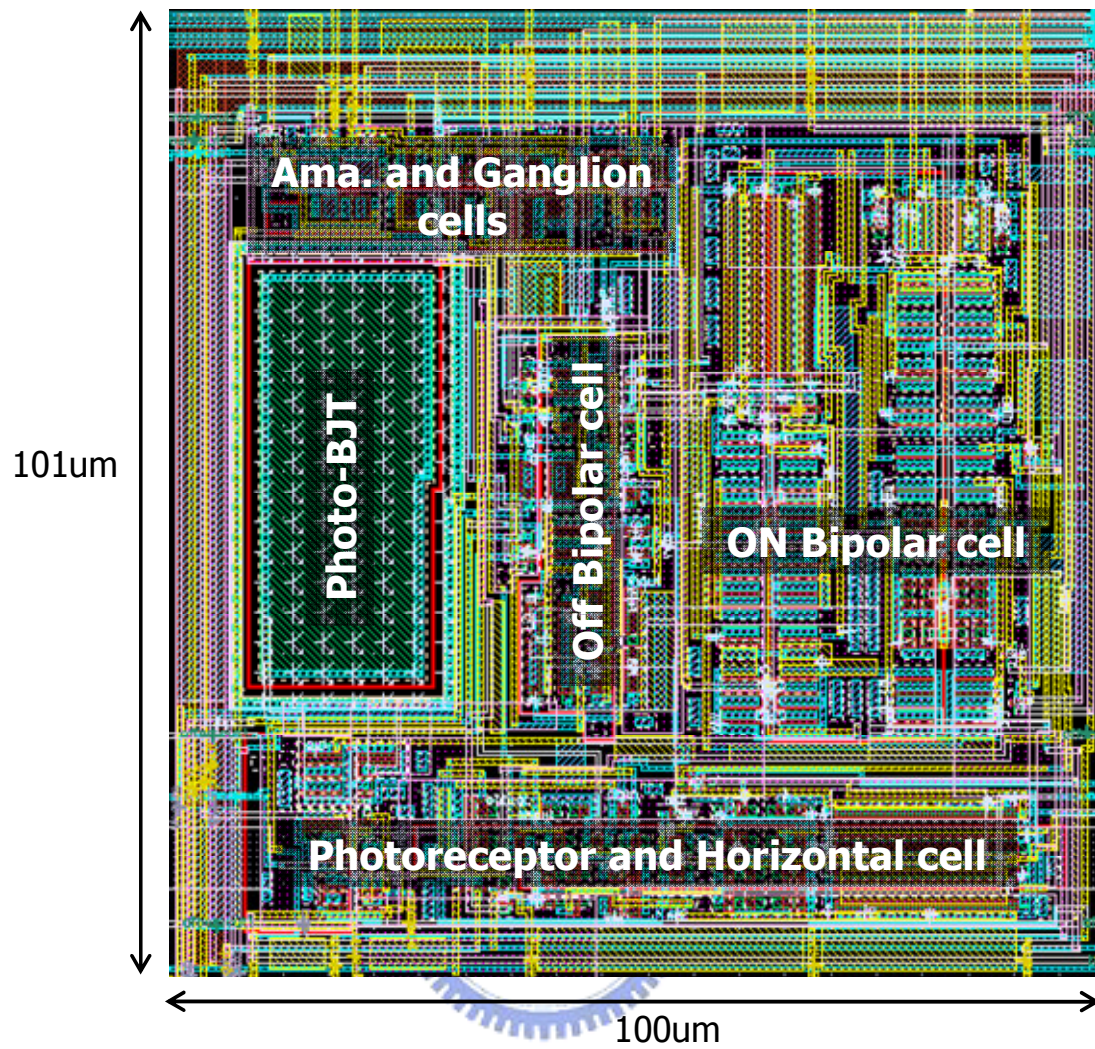


Fig. 4. 2. Layout and floorplan of a basic cell of the sensory array. The six regions of main parts of the basic cell and the phototransistor are labeled.

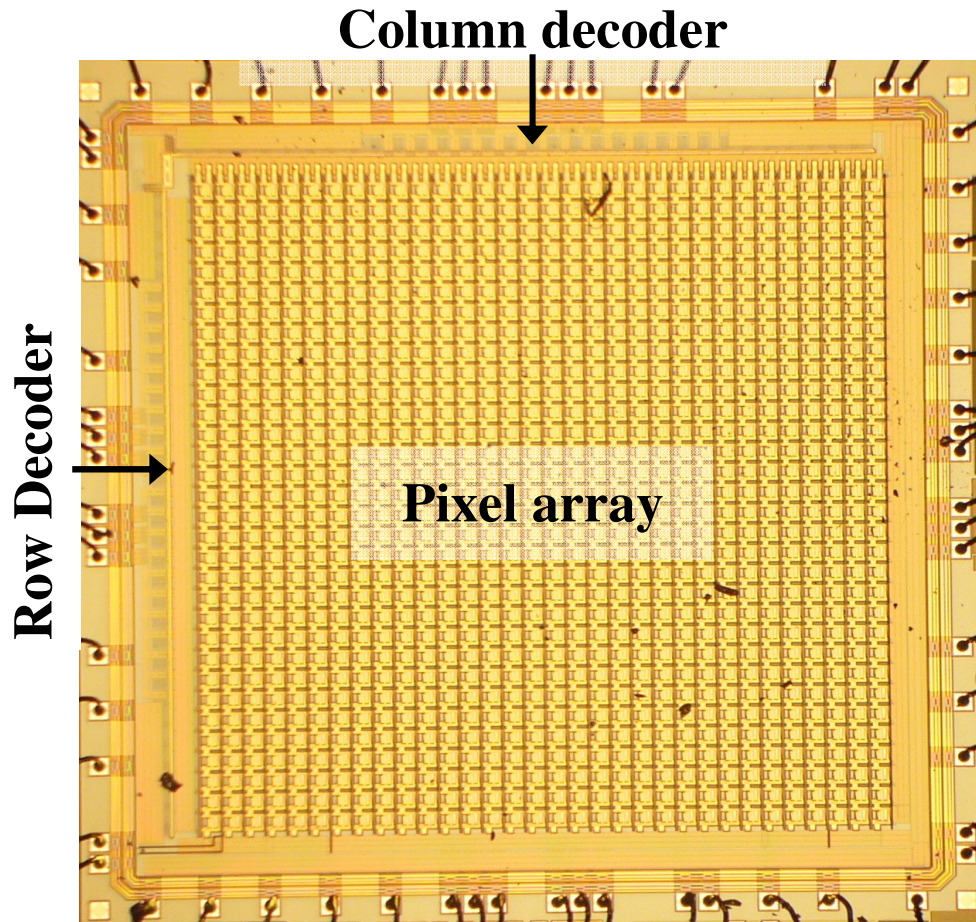


Fig. 4. 3. Photograph of the whole retinal chip. Main parts of the chip are labeled. Sensory array is in the center, the row and column address decoder are in the upper and left of the array respectively.

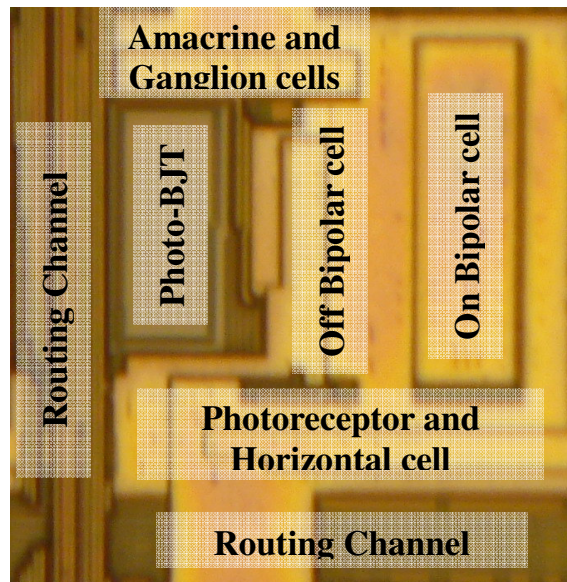


Fig. 4. 4. Photograph of a basic cell. Only the photo-BJT region is not covered by metal.

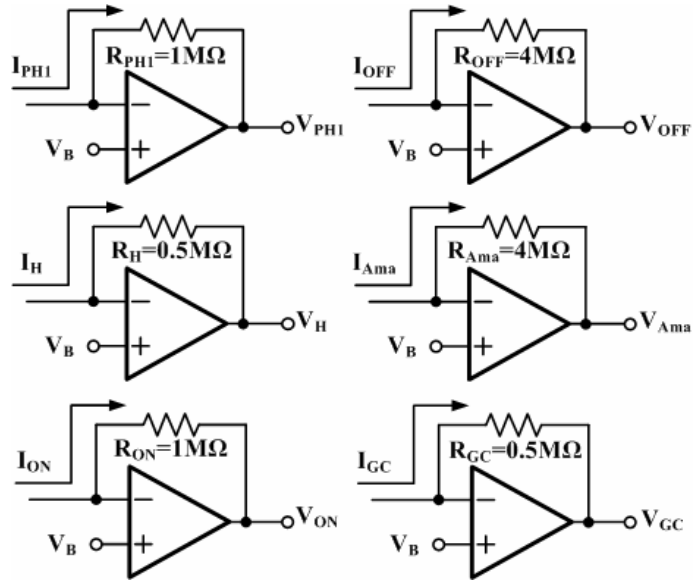


Fig. 4. 5. The read-out circuit to translate the output current into voltage.

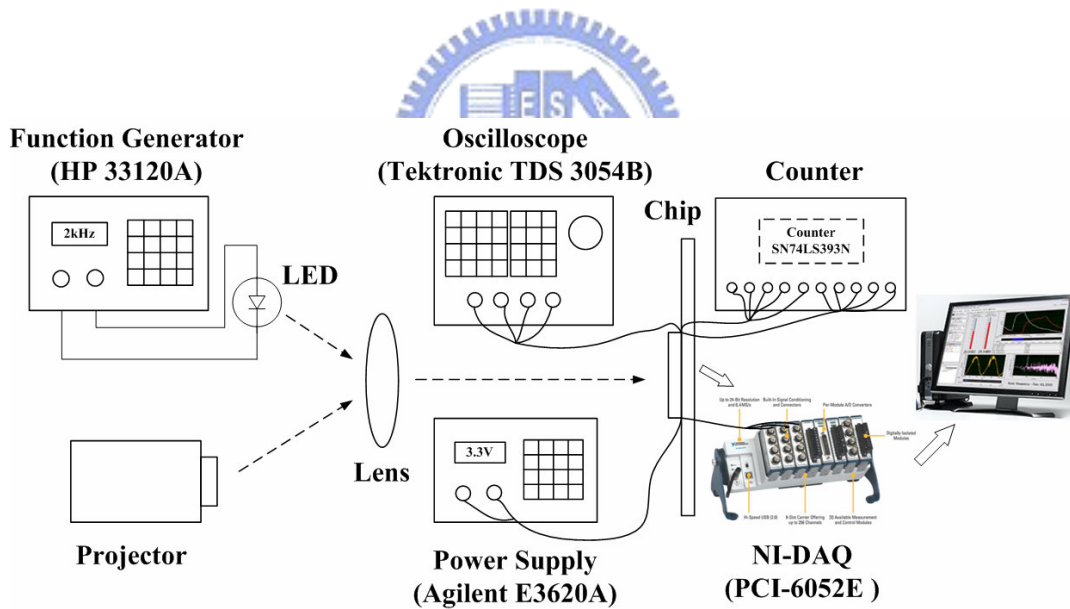
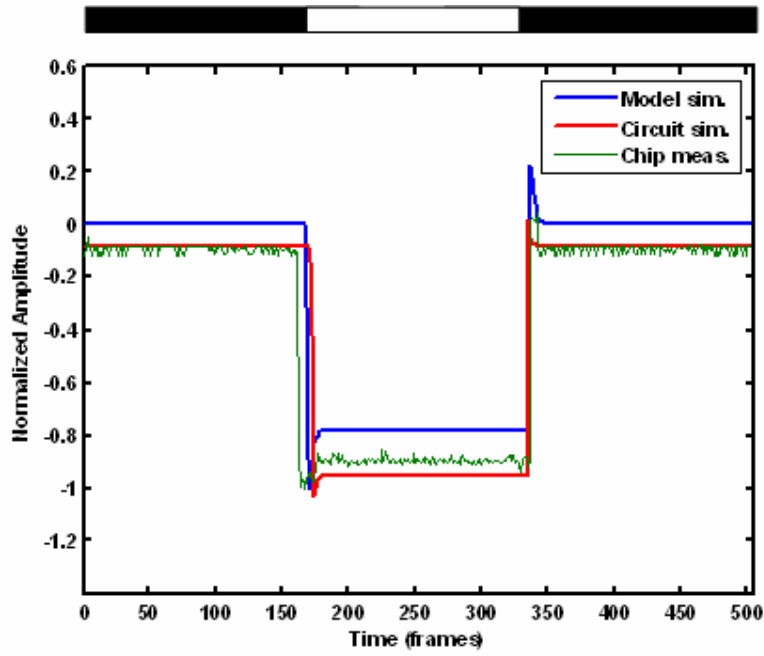
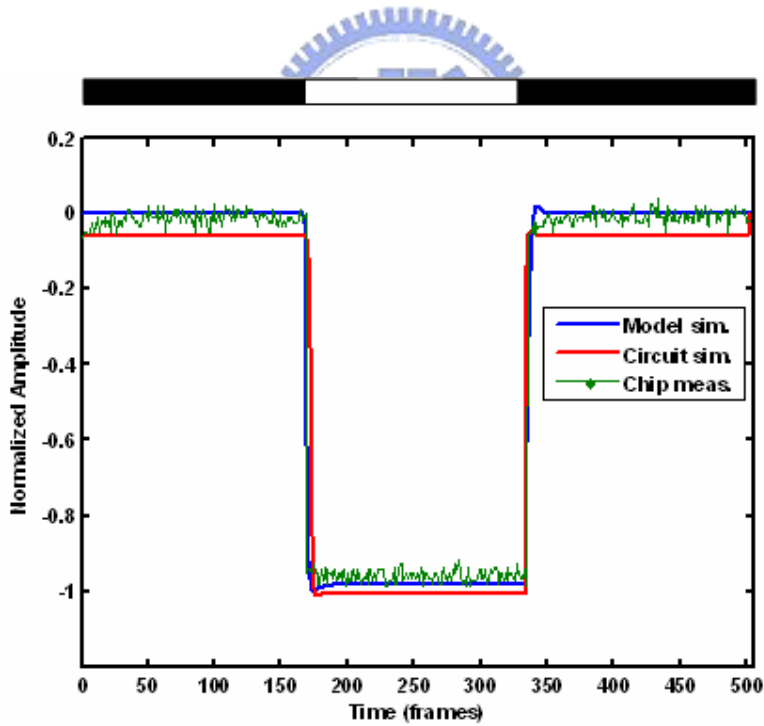


Fig. 4. 6. The measurement setup diagram.

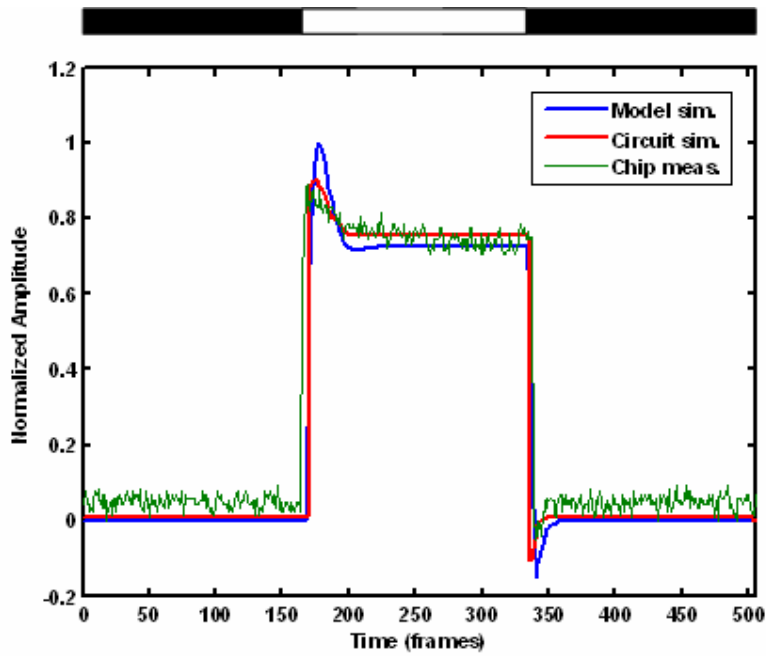


(a)

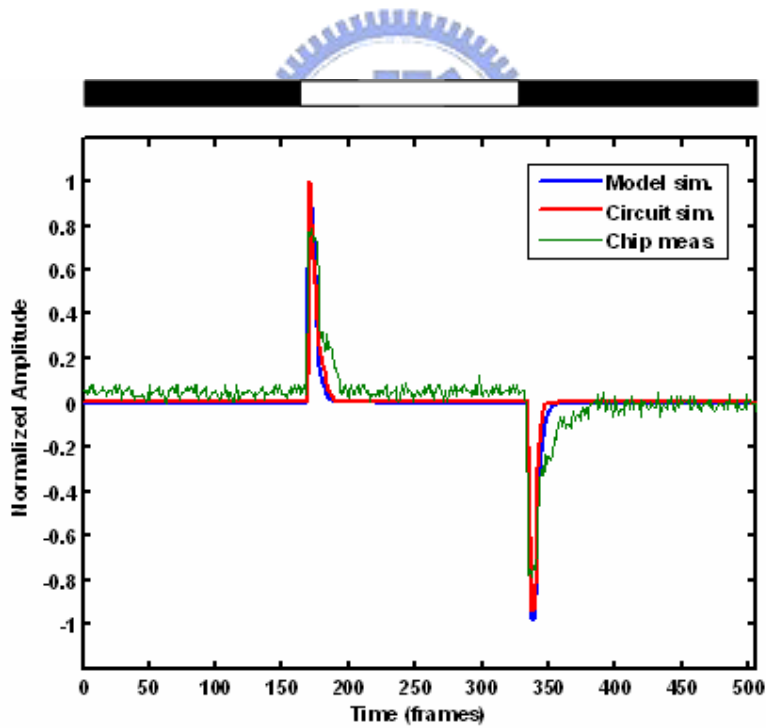


(b)

Fig. 4. 7. Comparison of the chip measured transient response to the model simulation results and circuit simulation results. (a) photoreceptor, (b) horizontal cell. The time units for model simulation and circuit simulation are the same, $17\mu\text{s}/\text{frame}$; time unit for chip measurement is $20\mu\text{s}/\text{frame}$. The light stimulus at $170^{\text{th}}\sim 340^{\text{th}}$ frame and $V_{\text{sm}}=2.4\text{V}$.

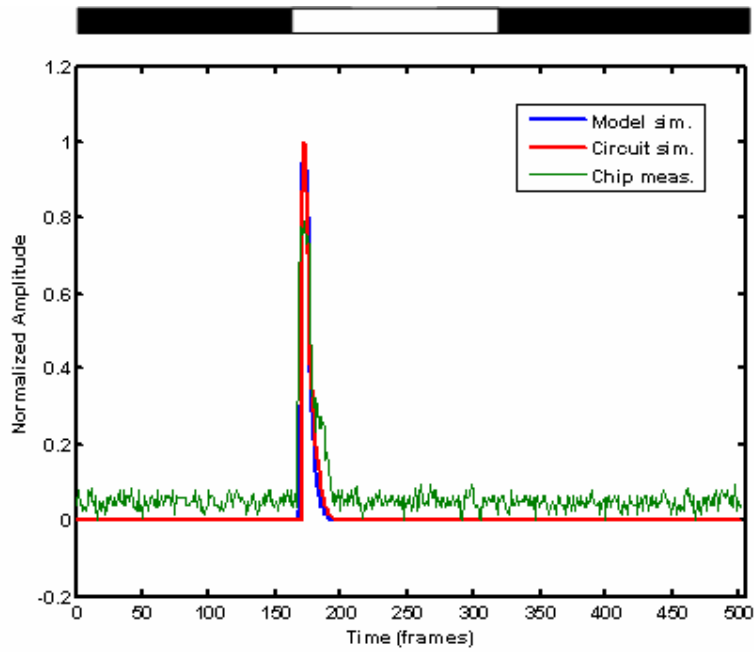


(a)

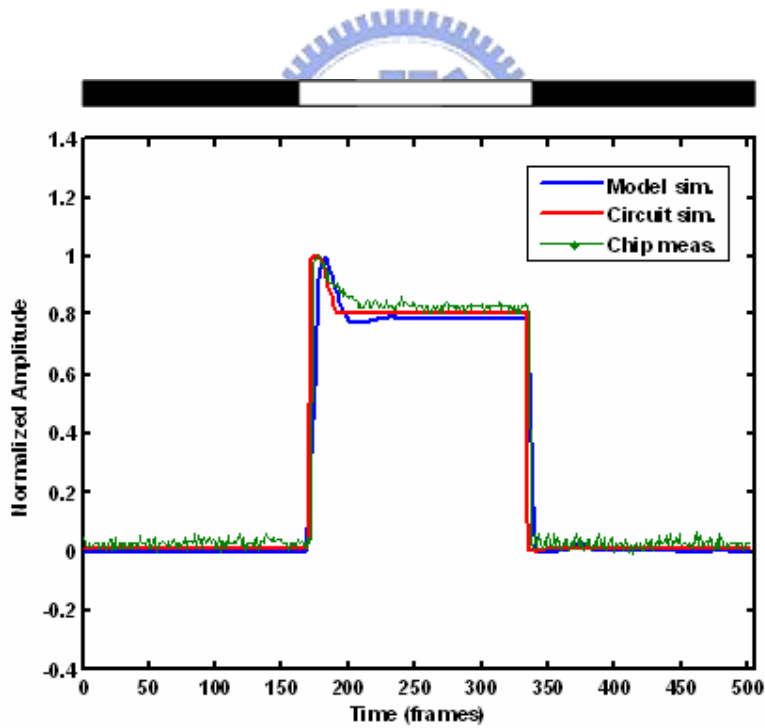


(b)

Fig. 4. 8. Comparison of the chip measured transient response to the model simulation results and circuit simulation results. (a) on bipolar cell, (b) off bipolar cell. The time units for model simulation and circuit simulation are the same, $17\mu\text{s}/\text{frame}$; time unit for chip measurement is $20\mu\text{s}/\text{frame}$. The light stimulus at $170^{\text{th}}\sim 340^{\text{th}}$ frame and $V_{\text{sm}}=2.4\text{V}$.

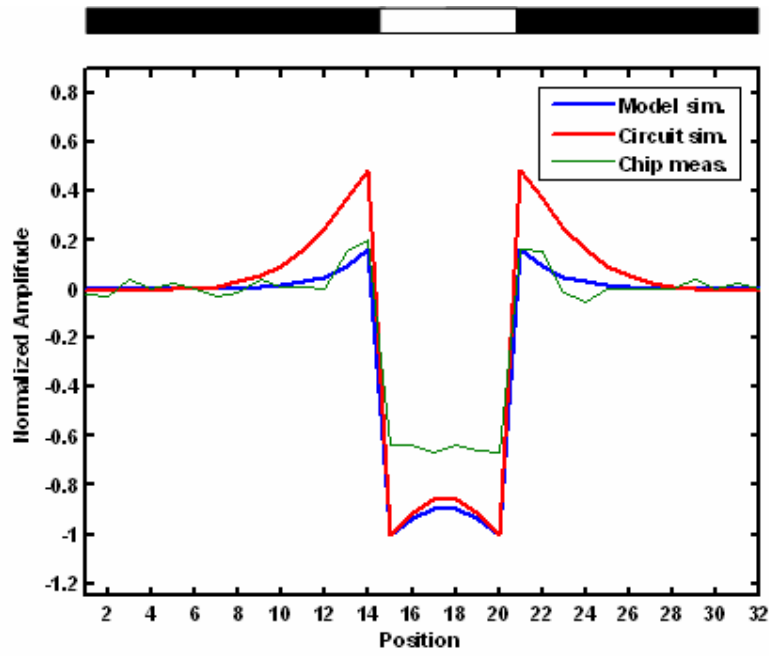


(a)

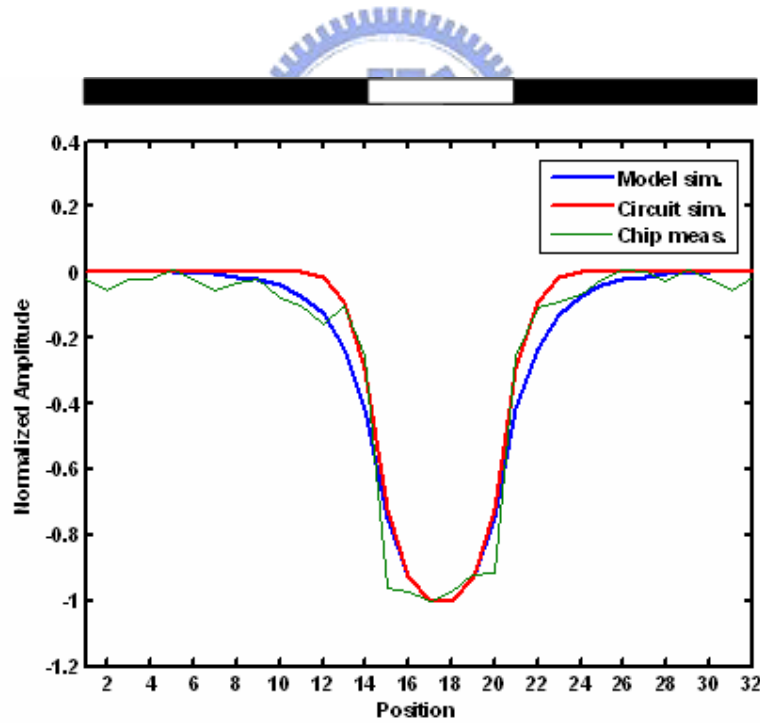


(b)

Fig. 4. 9. Comparison of the chip measured transient response to the model simulation results and circuit simulation results. (a) amacrine cell, (b) ganglion cell. The time units for model simulation and circuit simulation are the same, $17\mu\text{s}/\text{frame}$; time unit for chip measurement is $20\mu\text{s}/\text{frame}$. The light stimulus at $170^{\text{th}}\sim 340^{\text{th}}$ frame and $V_{\text{sm}}=2.4\text{V}$.

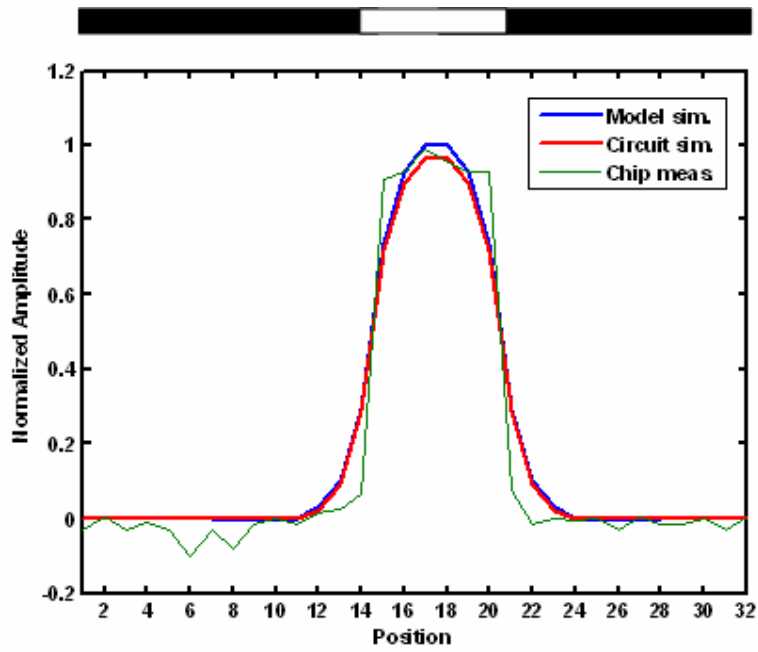


(a)

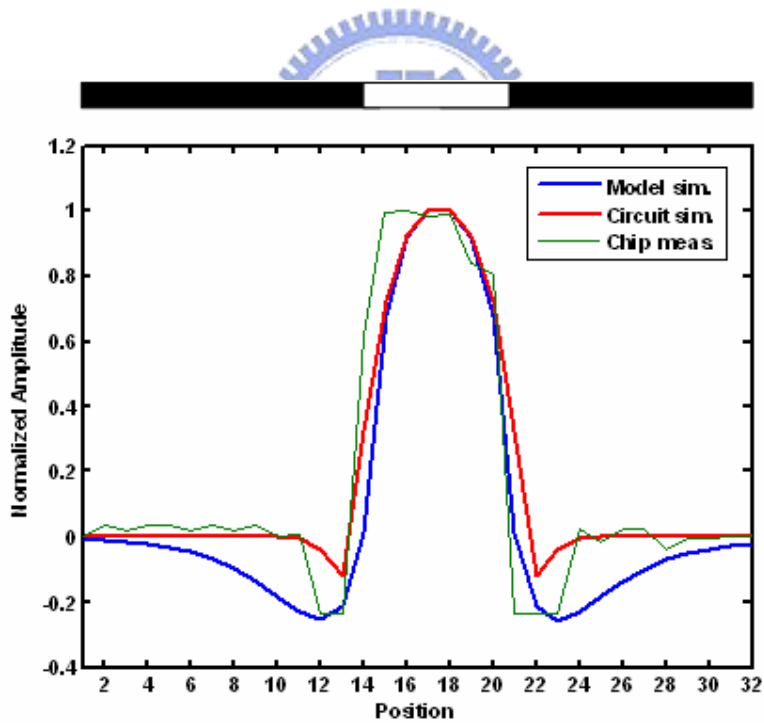


(b)

Fig. 4. 10. Comparison of the chip measured spatial response to the model simulation results and circuit simulation results. (a) photoreceptor, (b) horizontal cell. The light stimulus at 15th~21th pixel and $V_{sm} = 2.4V$.

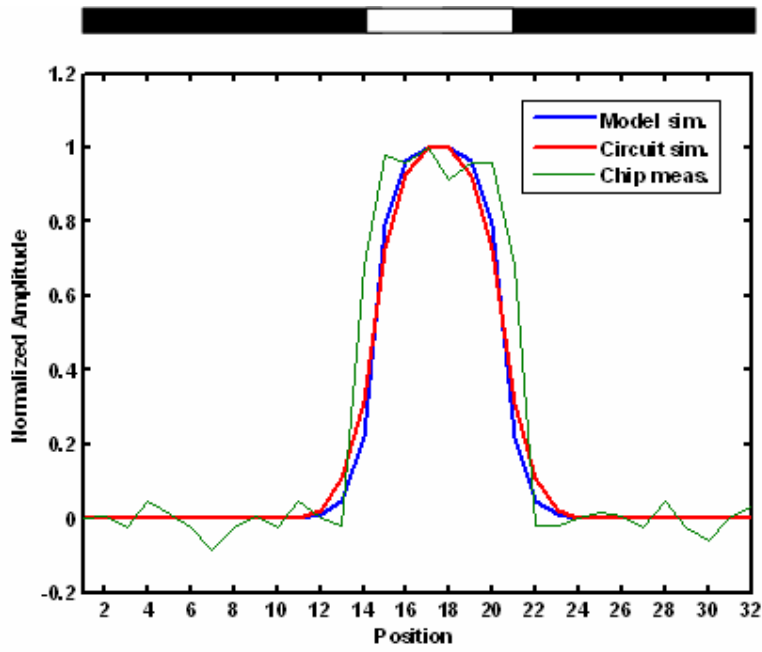


(a)

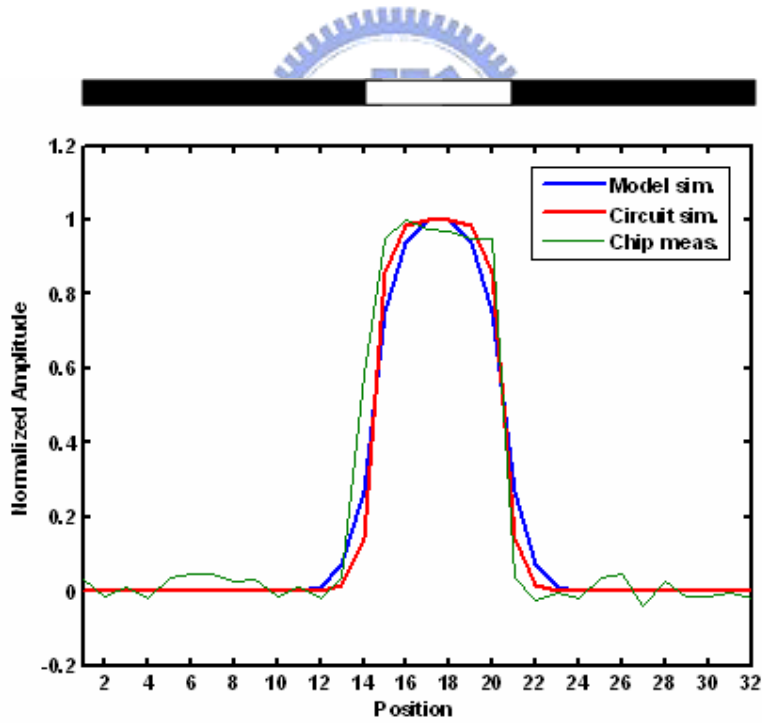


(b)

Fig. 4. 11. Comparison of the chip measured spatial response to the model simulation results and circuit simulation results. (a) on bipolar cell, (b) off bipolar cell. The light stimulus at 15th~21th pixel and $V_{sm} = 2.4V$.



(a)



(b)

Fig. 4. 12. Comparison of the chip measured spatial response to the model simulation results and circuit simulation results. (a) amacrine feed-forward cell, (b) ganglion cell. The light stimulus at 15th~21th pixel and $V_{sm} = 2.4V$.

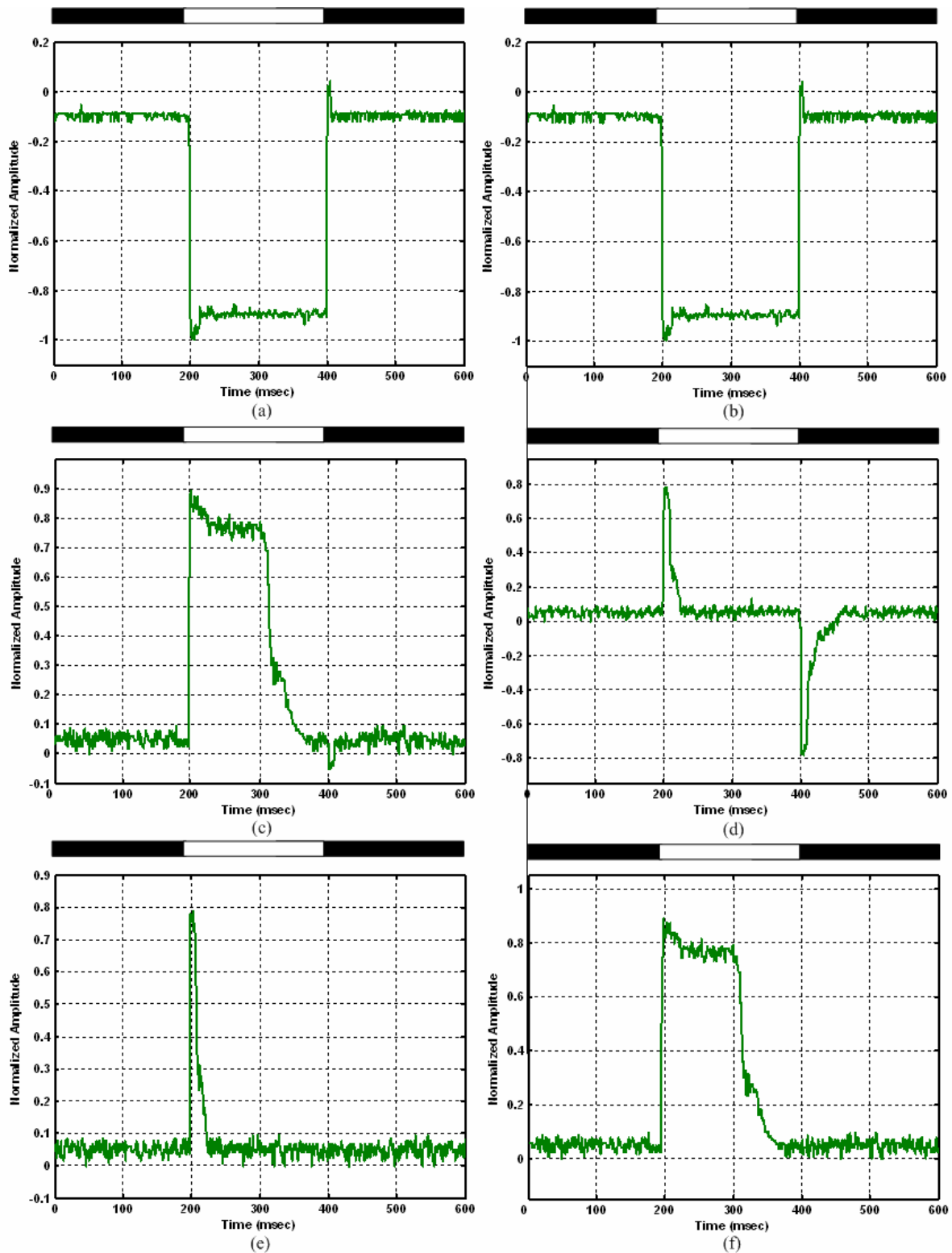


Fig. 4. 13. Measured transient response with large light stimulus interval (200msec~400msec). Curve (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. Under bias condition: $V_{sm}=2.4V$.

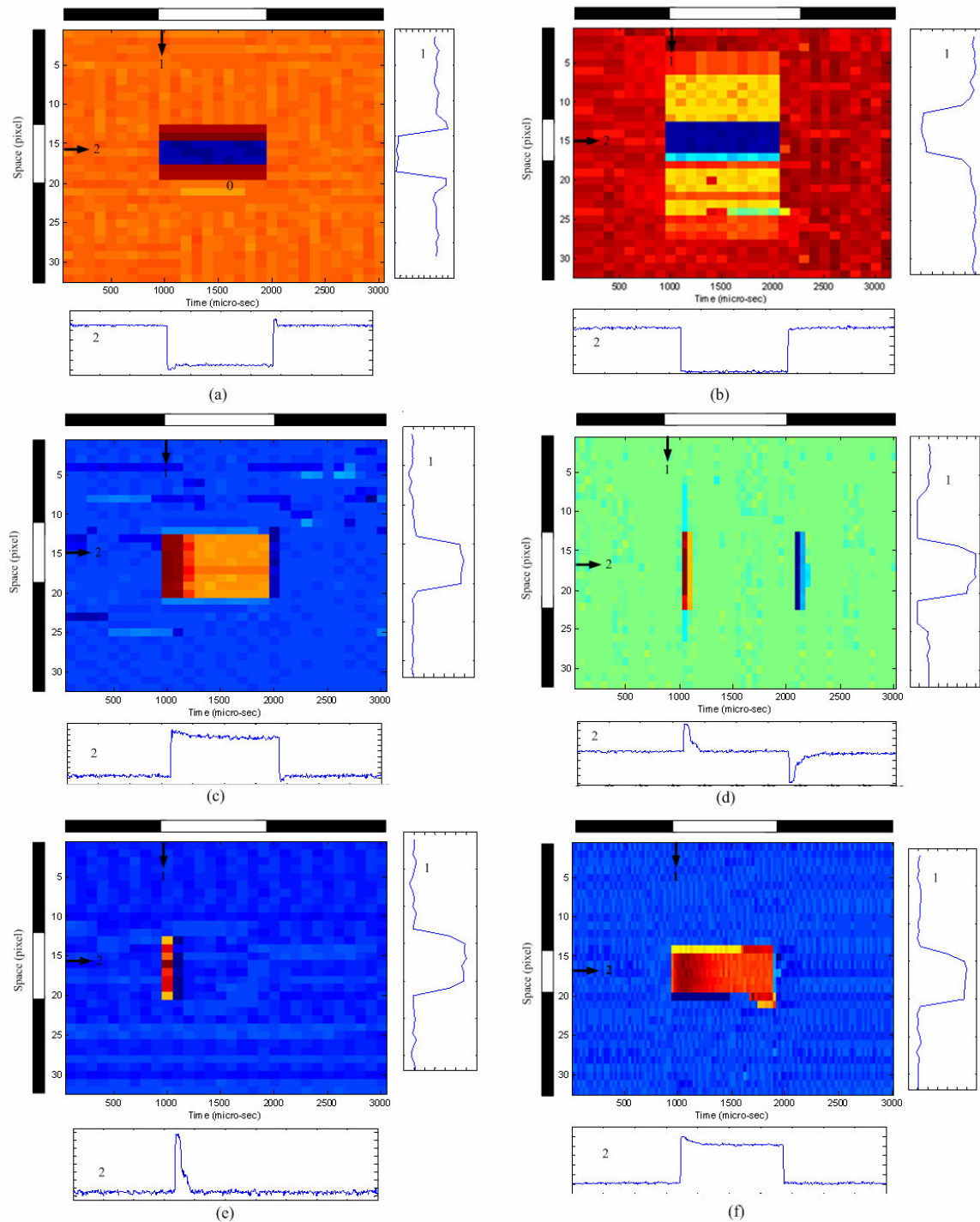


Fig. 4. 14. The measured spatiotemporal patterns for (a) photoreceptor, (b) horizontal cell, (c) ON bipolar cell, (d) OFF bipolar cell, (e) amacrine cell, and (f) ganglion cell. These patterns are recorded from the 17th row of the array. The x-axis is normalized time and the y-axis is the pixel location which denotes space. The stimulus is applied to the 19th to the 23rd pixel at time point 1001 to 2000. The waveform at the left of each pattern is the spatial domain waveform(s) obtained at the time marked by the vertical arrow(s). The waveform at the bottom of each pattern is the temporal domain waveform obtained at the location marked by the horizontal arrow. Under bias condition: $V_{sm}=2.4V$.

Table IV

Summary of characteristics of the proposed 32x32 retinal chip

| | | |
|---|--|---|
| | This work | L. J. Lin, "Implementations and applications of the retinal functions on integrated circuits", National Chiao Tung University, Degree of Doctor of Philosophy in Electronic Engineering, June 2007. |
| Process | TSMC 0.35UM Mixed-Signal 2P4M Polycide 3.3/5V | TSMC 0.35UM Mixed-Signal 2P4M Polycide 3.3/5V |
| Power supply | 3.3V | 2.5V |
| Resolution | 32 x 32 | 32 x 32 |
| *DC current level variation (OFF bipolar cell) | 0.04 μ A | 8 μ A |
| Maximum inter-pixel variations of DC current output | 0.26 μ A | N/A |
| Fill Factor | 9.14% | 7.44% |
| Frame rate | 16kHz | 16kHz |
| Basic Cell Area | 101 x 100 μ m ² | 105 x 105 μ m ² |
| Whole Chip Area | 4.3 x 4.2 mm ² | 4.4 x 4.3 mm ² |
| Chip power dissipation | 1.15W | 1.675W |

*Simulation date

CHAPTER 5 Conclusions and Future work

5.1 CONCLUSIONS

Based on the previous work on ‘ON brisk transient’ ganglion cell set, we developed a novel retinal chip that imitates the ON sluggish sustain ganglion cell set in the retina of rabbits. The model-building approach of the design methodology is to incorporate the available concerning morphology, electro-physiology, and pharmacology and by only using elementary building blocks. An OFF bipolar cell block features low DC current variation is proposed to solve the DC level variation problem in the previous work. The simulation result of the original structure shows that the variation of output current level is $8\mu\text{A}$ while the variation of the modified circuit is less than $0.04\mu\text{A}$. The reduction of DC current variation is verified by the measurement results indicating that the maximum inter-pixel variations of DC current output is less than $0.26\mu\text{A}$.

Besides the difference of the OFF bipolar cell, the ON bipolar cell block is also different in its structure. In the previous work, the ON bipolar cell performs bandpass-filtering on the signals, whereas the ON bipolar cell of this work performs large time delay and performs lateral diffusion in space. The structures are dissimilar.

The chip contains 32×32 pixels where each pixel imitates one cell set of the ON sluggish sustain ganglion cell. The chip is manufactured in TSMC $0.35\mu\text{m}$ double-poly quadruple-metal standard CMOS technology. The area of a basic retina cell is a square of $100\mu\text{m}$, and the total area of the whole chip is approximately $4.3 \text{ mm} \times 4.2 \text{ mm}$. The fill factor is 9.14% and the maximum inter-pixel variation is less than 0.26mA . Most parts of the chip deal with current-mode signals. The total power consumption is 1.15W under light stimuli.

An experimental chip which imitate the ON sluggish sustain ganglion cell set of rabbits’ retinas is fabricated and measured to verify the design methodology. Through HSPICE simulation and chip measurement, the functions of the chip are verified. Such consistency strongly suggests that the chip,

in extracting the features of the visual world, behaves in a way which is similar to that of real retinal cells. Therefore, the verifications of the implemented chip establish the success of the proposed design methodology. The tunable parameter associated with the CNN model is the space constant of the horizontal. Varying bias voltage, the desired space constant could be obtained. The functions of designed chip has been verified by the measured spatiotemporal responses under flashing light stimulus which shows the function consistence between the designed chip and the rabbit retina.

5.2 FUTURE WORK

Though the function of the chip is verified to be correct, there is still room to improve. First, reducing the power consumption of the designed chip and shrinking the size of the cell. The cell area of the proposed retina is still too large with compared to that of a real retina cell. Shrinking the size of the cell may make some potential applications such as implantation of artificial retina in human body more practical.

Secondly, the chip contains 32x32 pixels where each pixel imitates one cell set of the ON sluggish sustain ganglion cell. In the practical application, the output signal of each pixel should transmit the ganglion spiking to an electrode to stimulate and imitate the ON sluggish sustain ganglion cell set of rabbits' retina, respectively.

Third, based on the 'ON brisk transient' and the 'ON sluggish sustain' ganglion cell set of rabbits' retina, every kind of ganglion cell sets can be implemented and integrated in the same way. Based on the results, the full ganglion cell sets of retina can be designed and realized. Thus the research will help the blind to restore vision.

Finally, with the design methodology proposed in this thesis, many potential applications of retinal chips including motion sensors, computer vision, retinal prosthesis, and biomedical devices are feasible.

REFERENCES

- [1] H. Wässle, "Parallel processing in the mammalian retina," *Nature Reviews Neuroscience*, vol. 5, pp. 1-11, October 2004.
- [2] R. H. Masland, "The fundamental plan of the retina," *Nature Neuroscience*, vol. 4, no. 9, September 2001.
- [3] R. L. Rockhill, F. J. Daly, M. A. MacNeil, S. P. Brown, and R. H. Masland, "The Diversity of Ganglion cells in a Mammalian Retina," *J. Neuroscience*, vol. 22, no. 9, pp.3831-3843, May, 2002.
- [4] K. K. Ghosh, S. Bujan, S. Haverkamp, A. Feigenspan, and H. Wässle, "Types of bipolar cells in the mouse retina," *J. Comp. Neurol.*, vol. 469, pp. 70-82, 2004.
- [5] M. A. MacNeil, J. K. Heussy, R. F. Dacheux, E. Raviola, and R. H. Masland, "The population of bipolar cells in the rabbit retina," *J. Comp. Neurol.*, vol. 472, pp. 73-86, 2004.
- [6] K. A. Zaghoul and K. Boahen, "Optic nerve signals in a neuromorphic chip I: Outer and inner retina models," *IEEE. Trans. Biomedical Engineering*, vol. 51, no. 4, pp. 657-666, April 2004.
- [7] G. B. Awatramani and M. M. laughter, "Origin of transient and sustained responses in ganglion cells of the retina," *J. Neurosci.*, vol. 16, pp. 7087-7095, 2000.
- [8] S. M. Wu, F. Gao, and B. R. Maple, "Functional architecture of synapses in the inner retina: segregation of visual signals by stratification of bipolar cell axon terminals," *J. Neurosci.*, vol. 20, pp. 4462-4470, 2000.
- [9] R. H. Masland, "Neuronal diversity in the retina," *Curr. Opin. Neurobiol.*, vol. 11, pp. 431-436, 2001.
- [10] D. M. Dacey, and O. S. Packer, "Colour coding in the primate retina: diverse cell types and cone-specific circuitry ," *Curr. Opin. Neurobiol.*, vol. 13, pp. 421-427, 2003.

- [11] F. R. Amthor, E. S. Takahashi, and C. W. Oyster, "Morphologies of rabbit retinal ganglion cells with concentric receptive fields," *J. Comp. Neurol.*, vol. 280, pp.72-96, 1989.
- [12] E. B. Goldstein, "Sensation and Perception," 6th edition.
- [13] B. Roska and F. Werblin, "Vertical interactions across ten parallel, stacked representations in the mammalian retina," *Nature*, vol. 410, pp. 583-587, March 2001.
- [14] D. Bálya, B. Roska, T. Roska, and F. S. Werblin, "A CNN framework for modeling parallel processing in a mammalian retina," *Int. J. of Circuit Theory and Applications*, vol. 30, pp. 363-393, 2002.
- [15] T. Delbruck and C. A. Mead, "Analog VLSI Phototransduction by Continuous-Time, Adaptive, Logarithmic Photoreceptor Circuits," *Computation and Neural Systems Program*, Memo, no. 30, California Institute of Technology, July 1994.
- [16] F. Werblin, B. Roska, D. Balya, C. Rekeczky, and T. Roska, "Implementing a retinal visual language in CNN: a neuromorphic study," in *Proc. IEEE Int. Symp. Circuits Syst.*, vol. 2, pp. 333 -336, 2001.
- [17] D. Bálya, I. Petrás and T. Roska, "Implementing the multilayer retinal model on the complex-cell CNN-UM chip prototype," *Int. J. Bifurcation and Chaos*, vol. 14, no. 2, pp. 427-451, 2004.
- [18] P. Sterling, "How retinal circuits optimize the transfer of visual information," in *The Visual Neurosciences*, vol. 1, L. M. Chalupa and J. S. Werner, Ed. Cambridge, MA: The MIT Press, 2003, pp. 234-278.
- [19] C.Y. Wu and H.C. Jiang, "An improved BJT-based silicon retina with tunable image smoothing capability," *IEEE Transactions on VLSI Systems*, vol. 7, no. 2, pp. 241-248, June 1999.
- [20] L. J. Lin, C. Y. Wu, B. Roska, F. Werblin, D. Bálya, and T. Roska, "A neuromorphic chip that imitates the ON brisk transient ganglion cell set in the retinas of rabbits", *IEEE Sensors Journal*,

vol. 7, no. 9, Sep 2007.

- [21] C.Y. Wu and C.T. Chiang, "A low-photocurrent CMOS retinal focal-plane sensor with a pseudo-BJT smoothing network and an adaptive current Schmitt trigger for scanner applications," *IEEE Sensors Journal*, vol. 4, no. 4, pp. 410-418, August 2004.
- [22] C.Y. Wu and C.F. Chiu, "A new structure of 2-D silicon retina," *IEEE Journal of Solid-State Circuits*, vol. 30, no. 8, pp. 890-897, August 1995.
- [23] A. G. Andreou, K. A. Boahen, P. O. Pouliquen, A. Pavasovic, R. E. Jenkins and K. Strohhahn, "Current-mode subthreshold MOS circuits for analog VLSI neural systems," *IEEE Transaction on Neural Network*, vol. 2, no. 2, pp. 205-213, 1991.
- [24] B. G. Cleland, M. W. Dubin, and W. R. Levick, "Sustained and transient neurons in the cat's retina and lateral geniculate nucleus," *J. Physiol.* vol. 213, pp. 473-496, 1971.
- [25] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," *J. Neurophysiol.*, vol. 16, pp. 37-68, 1953.
- [26] Laguna, M. De la Cruz-Blas, C. Torralba, A. Carvajal, R.G. Lopez-Martin, A. Carlosena, A., "A novel low-voltage low-power class-AB linear transconductor,". *Proc. of the IEEE International Symposium on Circuits and Systems, ISCAS*, vol. I, pp. 725- 728, 2004.
- [27] J. Ramirez-Angulo, R.G. Carvajal, A. Torralba, J.A.Galan, A.P. Vega-Leal and J. Tombs, "The flipped voltage follower: A useful cell for low-voltage, low-power circuit design". *Proc. of the IEEE International Symposium on Circuits and Systems, ISCAS*, vol. III, pp. 615 – 618, 2002.
- [28] V. Peluso, P. Vancorenland, M. Steyaert and W. Sansen, "900 mV differential class AB OTA for switched opamp applications," *Electronics Letters*, vol. 33 , no. 17, pp. 14455-1456, Aug. 1997.