CLOSED-LOOP HIGHLY-SCALABLE RETINAL IMPLANT WITH FULLY-ANALOG ED-BASED ADAPTIVE-THRESHOLD SPIKE DETECTION AND POISSON-CODED TEMPORALLY DISTRIBUTED OPTOGENETIC STIMULATION

TAYEBEH YOUSEFI

A DISSERTATION SUBMITTED TO

THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

YORK UNIVERSITY

TORONTO, ONTARIO

Winter 2024

© Tayebeh Yousefi, 2024

Abstract

Intraocular stimulators show promise for treating retinal degeneration by restoring visual input to the damaged retina. This is achieved by capturing images with a wearable camera and accordingly stimulating remaining retinal cells (bipolar, ganglion), effectively bypassing dysfunctional photoreceptors. State-of-the-art retinal stimulators face a major challenge due to the lack of cell-type specificity of electrical stimulation and activating both ON and OFF pathways in the retina, leading to limited visual perception due to sending contradictory messages to the brain. This fundamental limit motivated us to investigate the development of an optogenetic-based retinal prosthesis, that uses specific promoters for selective modification and activation of ON bipolar cells, offering a more natural vision restoration.

In developing such a device, the first challenge we faced was optimizing stimulation strategy for optimal therapeutic efficacy. Responding to this challenge, we first present a retina-inspired computational framework to evaluate and optimize an optogenetic epi-retinal neurostimulator. This framework reveals that optical stimulation, compared to electrical stimulation, provides superior visual perception, which improves with increased μLED array resolution. The framework also explores optical stimulation factors and μLED specifications like light intensity and wavelength spatial resolution and light divergence.

A critical issue in optogenetics is controlling opsin distribution, as uneven distribution affects light sensitivity across the retina. Variations in tissue properties and fluid dynamics introduce unpredictability in stimulation effectiveness. Our solution to this issue includes a scalable optogenetic stimulator IC, which features channel-specific closed-loop calibration for defining the optimal stimulation intensity using a temporally-adaptive-threshold spike detection circuit.

The second challenge we addressed was scalability, and by association, energy efficiency of the device. Scaling implantable stimulators is limited by instantaneous power demands during multi-channel stimulation. We address this by exploiting opsins' sensitivity to integrated optical energy (validated in vitro), using Poisson coding, temporally-distributed stimulation to evenly distribute the stimulation power consumption, enabled by our raster scanning technique for efficient μLED addressing. This reduces wireless data communication

requirements, and significantly reduces IC-to-optrode interconnections, making large-scale implementation feasible.

Our wireless and battery-less stimulator implant comprises blocks for optical stimulation, fully-adaptive spike detection, and closed-loop calibration. It calibrates light intensity for each μLED row based on recorded spiking activity. The IC is connected to a custom-designed optrode with a 10×10 μLED array, featuring a shared electrical recording pad for every four LEDs.

To the memory of my grandmother, whose heart was a wellspring of kindness

To my parents and sister, for their unwavering support and love

To my husband, the source of love and joy in my life.

Acknowledgements

I extend my deepest gratitude to my supervisor, Dr. Hossein Kassiri, for his invaluable guidance and support throughout my research journey. His expertise and insightful feedback have been crucial in shaping my work. More importantly, his remarkable support during times of challenge, particularly when we encountered mistakes, has been a source of great encouragement.

My sincere thanks also go to my defense committee members, Dr. Georg Zoidl, Dr. Amir-Soleimani, Dr. Peter Lian, Dr. Azadeh Yazdan-Shahmorad and Dr. Nima Tabatabaei for their insightful feedback and contributions to my work. A special note of thanks to Dr. Georg Zoidl, whose collaboration was instrumental in our success. His assistance in providing access to lab facilities and equipment, and his insightful discussions on the biological aspects of our work, were invaluable.

I extend my heartfelt thanks to Cadence Design System, Dr. Michael Venditti, and Dr. Louis Francois Tanguay for providing me with an invaluable internship opportunity during my PhD studies. The insights and skills I acquired under their guidance have been crucial to my professional development.

My journey would not have been the same without the friendship and support of my col-

leagues. I would like to express my appreciation to my labmates, Mansour Taghadosi, Mina Sayedi, Rojin Salahi, Abdul Muneeb, Haydar Isa Senturk, and my colleagues Mahdi Nekoui Shahraki, Milad Akbari, Milad Ghalamboran, and Iman Babaiyazdi. Their presence made this experience both enjoyable and memorable.

I am sincerely grateful to my parents and sister, whose unwavering affection and encouragement have been my constant source of strength and inspiration. Last but certainly not least, I am eternally grateful to my husband, whose love, support, and belief in me were the bedrock upon which I leaned at every step of this journey. His presence has been a guiding light in both my personal and professional life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADC - Analog-to-Digital Converter

AMD - Age-Related Macular Degeneration

ASIC - Application-Specific Integrated Circuit

ASK - Amplitude Shift Keying

BCIs - Brain-Computer Interfaces

BMI - Brain-Machine Interfaces

CMOS - Complementary Metal-Oxide-Semiconductor

COTS - Commercial Off-The-Shelf

CSF - Cerebrospinal Fluid

DAC - Digital-to-Analog Converter

DNL - Differential Non-Linearity

DSP - Digital Signal Processing

ED - Energy of Derivative

IIRN - Integrated Input-Referred Noise

IC - Integrated Circuit

INL - Integral Non-Linearity

LFPs - Local Field Potentials

LFSR - Linear Feedback Shift Register

LSK - Load Shift Keying

NEF - Noise Efficiency Factor

NEO - Nonlinear Energy Operator

OBG - Oregon Green^{TM} 488 BAPTA-1, AM

OMS - Object Motion-Sensitive

OTA - Operational Transconductance Amplifier

PCB - Printed Circuit Board

PBS - Phosphate Buffered Saline

PRNG - Pseudo-Random Number Generator

RGCs - Retinal Ganglion Cells

SAR - Specific Absorption Rate

SEM - Scanning Electron Microscope

SNDR - Signal-to-Noise and Distortion Ratio

SNNs - Spiking Neural Networks

SNR - Signal-to-Noise Ratio

SoC - System on Chip

SU-8 - A commonly used epoxy-based negative photoresist

WSCM - Wide-Swing Current Mirror

μ**LED** - Micro Light Emitting Diode

Chapter 1

Introduction

1.1 Biological Background

As the leading cause of vision loss, age-related macular degeneration (AMD) affects 196 million people worldwide— a number that is projected to double by 2050 [1, 2]. AMD patients gradually lose their vision due to the degeneration of photo-receptor layer cells in their retina. Regardless of the underlying reason being environmental, pathologic damage, or genetic mutation, currently this progressive neurological disorder has no established cure, and various pharmacological therapies can only slow down its progression [3].

Understanding the retina's complex neural network and its role in visual processing provides a critical context for exploring treatment options for retinal degeneration. This understanding is essential for comprehending the functional basis upon which the treatment methodologies are designed.

As shown in Fig. 1.1, the mammalian retina is located at the back of the eyeball and

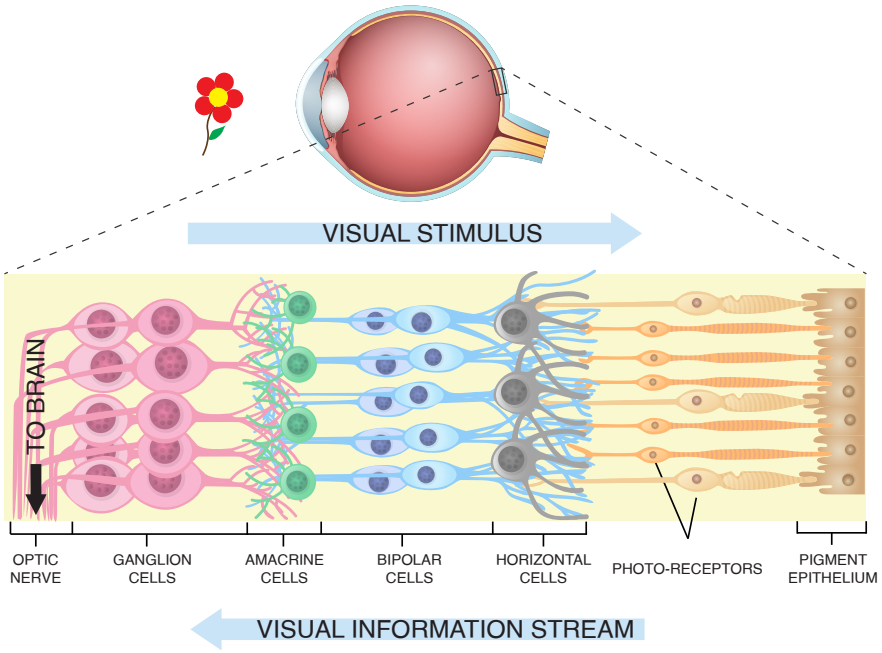


Figure 1.1: The structure of different layers in a mammalian retina, illustrating the visual input light path and the visual data processing network.

consists of rod and cone photo-receptors, horizontal cells, bipolar cells, amacrine cells, and retinal ganglion cells (RGCs). It is the axons of RGCs that are bundled together to form the optic nerve, which is responsible for transmitting visual information from the retina to the brain. As illustrated by the "visual stimulus" direction in the picture, the light passes through layers of neurons before activating the rod and cone photo-receptors.

In a healthy retina, upon receiving an incident light and proportional to its intensity, photo-receptors generate photo-currents that trigger the downstream retinal network indicated by the "visual information stream" direction. For a degenerated retina, the loss of functionality in photo-receptors prevents the first stage of this chain from generating visually-evoked signals, which results in loss of vision despite the fact that the downstream network of neurons (i.e., horizontal, bipolar, amacrine, and RGCs) is still functional [3].

The downstream network of the retina operates as a series of parallel information pro-

cessing paths, each involving various cell types within distinct retinal layers. These cells are specialized in extracting different features from the stimulus image. The process initiates with photoreceptors, which are responsible for transducing light into neural signals. These signals are then relayed to bipolar cells, serving as intermediary processors that modify and forward the information to the retinal ganglion cells (RGCs).

In the retinal network, horizontal and amacrine cells, both of which being inhibitory cell types, play a significant role in the modulation of signals. These cells are integral in enhancing specific visual aspects, such as contrast and motion detection, through lateral interactions. The inhibitory nature of these cells allows for effective adjustment control and regulation of the signal processing within the retina. This modulation is critical for the accurate interpretation of visual stimuli, enabling the identification of edges, movement, and changes in light intensity. The presence of these inhibitory cells ensures that the signal processing is finely tuned, enhancing the retina's ability to respond to a diverse range of visual environments.

The RGCs, situated at the final stage of this processing pathway, integrate and transmit the processed visual information to the brain. Each RGC subtype is responsive to particular features extracted from the visual stimulus, such as movement, color contrast, or brightness changes. This specialization facilitates a detailed representation of the visual scene, with different RGCs activating in response to distinct stimulus characteristics. The RGCs, as the concluding element of this network, encode and relay this information to the brain, enabling complex visual perception.

Fig. 1.2 illustrates the retina's structure, which includes two distinct pathways known as

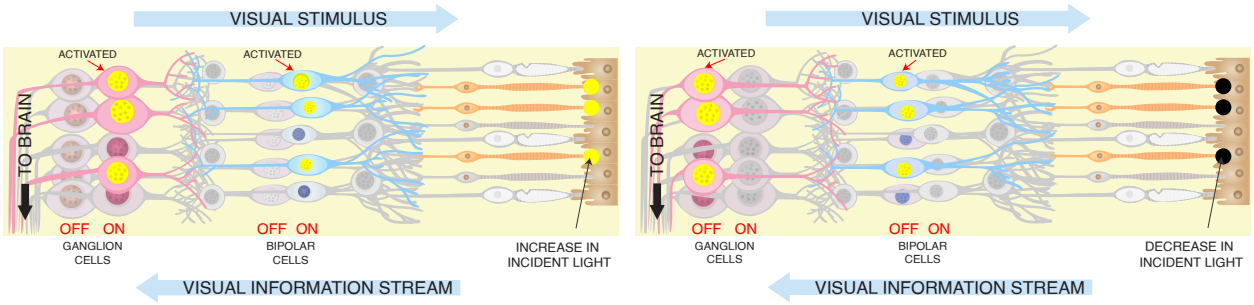


Figure 1.2: The structure and the differences of the retina's ON and OFF pathways.

the ON and OFF pathways. The ON pathway is activated by an increase in incident light, while the OFF pathway responds to a decrease in light, which results in independent and complementary activities in these pathways. These pathways are intricately woven throughout the entire retinal network, implying the presence of both ON and OFF cells across all layers of the retina, including the photoreceptors, bipolar cells, and RGCs.

1.2 Vision Restoration Techniques

Over the past few decades, a variety of multi-disciplinary approaches, including electrical retinal prostheses, optogenetics, chemical neural interfaces, gene therapy, and cell transplantation have been used in an attempt to restore visual function through bypassing the damaged photo-receptor layer [4–7]. Most of these methods rely on the fact that despite the photo-receptors deterioration and reorganization, the inner retinal neurons largely retain their capacity for signal transmission [3].

The subsequent section shifts the focus to these prostheses and approaches. It explores their potential as treatment options for retinal degeneration, detailing how they artificially stimulate neuronal layers to convey visual information to the brain.

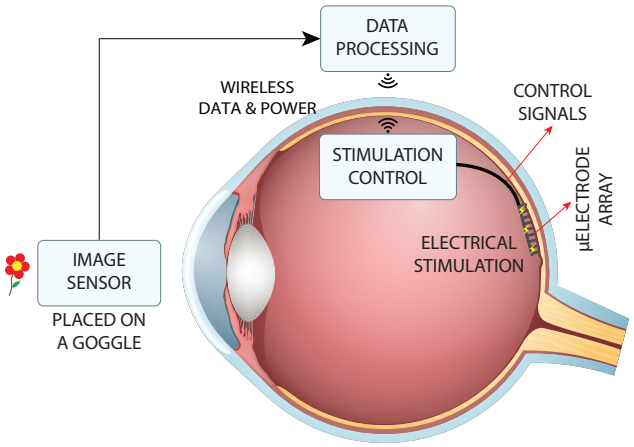


Figure 1.3: Simplified conceptual structure of electrical retinal prosthesis.

1.2.1 Electrical Retinal Prostheses

Implantable retinal prostheses using electrical stimulation have shown potential to offer a promising treatment option for retinal degeneration by manipulating its response to light through artificial electrical stimulation of the retina layers that carry visual information to the brain [8–11, 13, 14]. This section examines the technological developments in this field, categorizing the prostheses based on their placement relative to the retina and discussing their operational mechanisms.

As shown in Fig. 1.3 , in a retinal prosthesis, the visual stimulus is captured using an image sensor (typically, a camera placed on a goggle or an eye glass) and is sent to an external processing unit to generate corresponding stimulation patterns for artificial electronic stimulation of the retina layers (i.e., bipolar cells or RGCs). As shown, the external unit wirelessly communicates these commands to an implantable chip, where it is decoded to generate the control and timing signals required for conducting the stimulation. Based on the placement of the implant with respect to the retina, the electrical retinal prostheses are divided into two main categories of epi-retinal and sub-retinal.

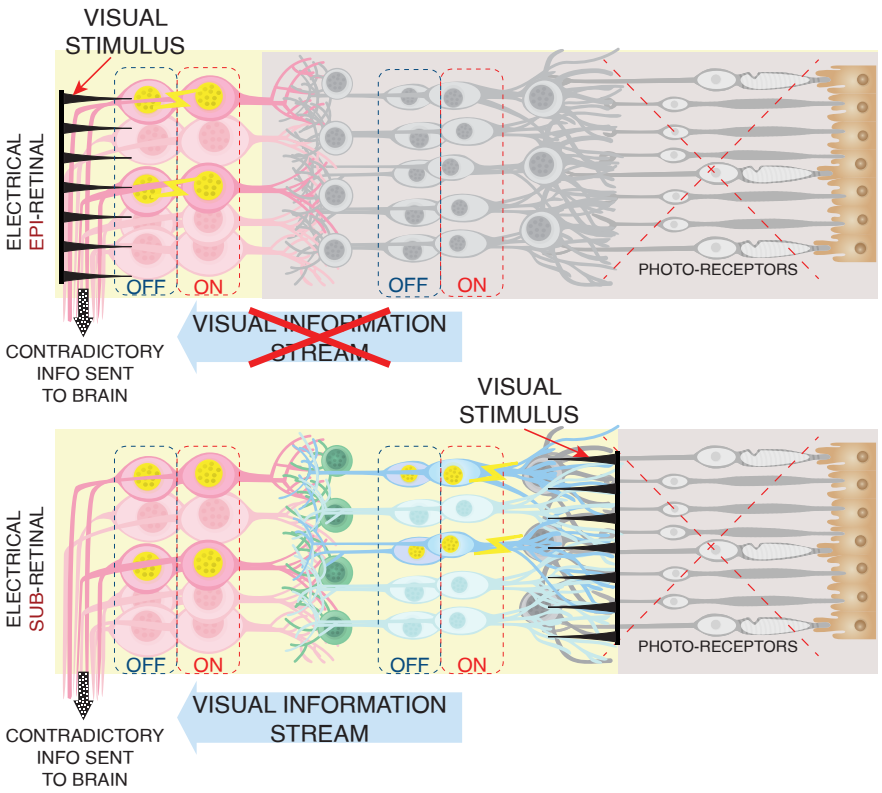


Figure 1.4: (a) an epi-retinal electrical stimulator and (b) a sub-retinal electrical stimulator.

The epi-retinal prosthesis (shown in Fig. 1.4(a)) is attached to the innermost layer of retina (i.e., the RGCs), which is the most common placement option, mainly thanks to its simpler surgical implantation [8–12]. However, since they are in direct contact with RGCs (i.e., the last layer in the retinal downstream network of neurons), all other healthy layers (i.e., bipolar, amacrine, etc.) are also bypassed. The job of these bypassed layers is to extract and accentuate features (i.e., visual information) from the photo-receptors' output and to pass along visual information and generate corresponding spiking patterns at RGC layer to relay the message to the brain. In the absence of these layers, this job needs to be done by the electronic circuits, using image processing techniques or artificial neural networks, implemented on-chip to mimic the functions of the bypassed downstream neuronal network

[15–17].

Due to the complexity and time-variability of the bypassed network, the cross-patient variations, and the limited computational resources, the implementation of such an image processing unit is neither efficient nor practical. In fact, the full extent of image processing that is taken place in the retina downstream network has not been fully recognized and there are more than 20 types of RGCs, each assumed to convey a different aspect of visual information [5, 18].

Motivated by this, sub-retinal stimulators are placed on the other end of retinal structure, where the degenerated photo-receptor layer is located (shown in Fig. 1.4(b)) [19, 20]. This type of prostheses can target the middle layer of retina (i.e., bipolar layer) and rely on the downstream retina network to perform the signal processing in the natural way [3, 4]. As such, stimulation pulses only need to mimic the photo-receptors' outputs, which are proportional to the incident light sensed by the image sensor, a significantly more feasible task to do compared to epi-retinal stimulators. However, the implantation surgery for sub-retinal implants are more challenging and invasive and it can cause severe inflammation and implant rejection. Additionally, the limited available space for the implant and being away from vitreous gel which has temperature regulatory feature, result in tight restrictions on implant's size and heat dissipation, respectively [13, 14, 21].

Regardless of the implantation being epi- or sub-retinal, a fundamental barrier that severely limits electrical (voltage or current mode) stimulators' efficacy in restoring vision is their indiscriminate activation of all types of retina cells that are in the proximity of stimula-

tion electrodes. As will be described in detail in chapter 2, this could result in contradicting messages being communicated to the brain through ON and OFF pathways, resulting in a poor or noise-like image reconstruction [22, 23].

In an effort to address this, and to improve visual perception's quality and consistency, neuro-chemical retinal prostheses are proposed (e.g., [24]) that employ certain chemical neurotransmitters to discriminately target different types of retinal cells to achieve a target stimulation pattern. However, delivery of these chemical substances in a spatially- and temporally-controlled fashion poses significant challenges that do not bode well with stimulation channel-count scaling, thus preventing the development of a multi-channel implantable device capable of performing this task to date.

1.2.2 Optogenetic Stimulator

Alternative techniques for vision restoration, such as optogenetic and gene therapy methods, have been explored in recent research. These approaches involve artificially sensitizing other retinal layers, like the Retinal Ganglion Cells (RGC) or Bipolar cells, to light. This is achieved through optogenetics, a process where optical stimulation is used to modulate the activity of neurons that have been genetically altered to respond to light. This genetic alteration involves the incorporation of specific proteins known as opsins into the neuron's membrane, making the cell responsive to certain light wavelengths depending on the opsin type used.

As illustrated in Fig. 1.5, a significant challenge with this approach is that the genetically modified cells do not possess the same level of light sensitivity as the photoreceptors in a healthy retina. Consequently, these cells are unable to respond to ambient light. In studies

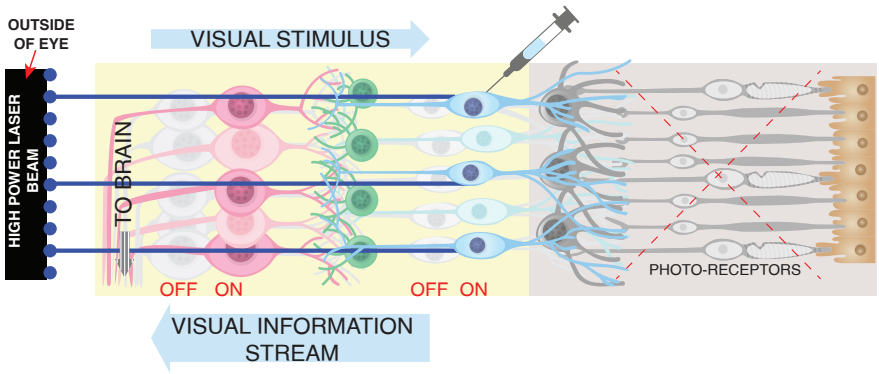


Figure 1.5: Using optogenetic as a vision restoration technique.

utilizing optogenetic stimulation for vision restoration, a powerful external laser light source, often placed outside of the eye, is required to achieve the necessary level of optical stimulation. This necessity highlights a key limitation in the current optogenetic methods and underscores the need for further advancements to make these techniques more viable for practical vision restoration applications.

1.3 Thesis Objectives and Design Challenges

The core objective of this thesis is to design and develop an innovative implantable retinal prosthesis that harnesses the power of optogenetics for cell-type specific stimulation, with the ultimate goal of restoring vision in patients affected by retinal degeneration. This goal necessitates addressing several key challenges. Below, we outline the primary design challenges and detail the solutions implemented to overcome these obstacles, focusing on scalability, power efficiency, adaptivity, and closed-loop operation as central themes of our research.

1) Power Efficiency and Scalability

Scalability and power efficiency are intertwined challenges that are crucial for the successful implementation of an implantable retinal prosthesis. Scalability refers to the

system's capacity to accommodate a large number of stimulation channels, which is essential for restoring high-resolution vision. Power efficiency, on the other hand, is crucial for ensuring that the system can operate within the constraints of a limited power supply, typically provided in a wireless fashion to the implanted device.

To address these challenges, we employ strategies to promote both scalability and power efficiency. Through use of photon integration properties of opsins, coupled with a raster scanning scheme and Poisson temporal coding, we show that the system's instantaneous power consumption profile could be significantly flattened, removing large instantaneous power peaks. This not only reduces the total instantaneous power requirement but also minimizes the need for extensive wiring, thus supporting a higher density of μLEDs within the array. The result is a system that can support a greater number of stimulation channels without compromising on safety or functionality, directly translating to improved potential resolution for prosthetic vision. This strategic approach demonstrates our system's enhanced scalability and power efficiency, which are vital for the successful deployment of high-resolution retinal prostheses. In Chapter 4, we focus on this challenge and our proposed solution.

2) Spatio-Temporal Adaptively and Closed-Loop Operation

The adaptivity and closed-loop operation of our system are critical for adjusting the stimulation to the opsin distribution variations across the retina and over time. This challenge necessitates a solution that can dynamically adjust stimulation parameters in real-time to optimize visual restoration. In Chapter 5, we detail our approach to

achieving this through an innovative ADC-less spike detection mechanism that serves as the foundation of our closed-loop system.

This mechanism enables the system to adaptively modify the light intensity emitted by each μLED based on real-time feedback from the spiking cells in the RGC layer. By incorporating a fully-analog, temporally-adaptive threshold spike detection circuit, our system can continuously adjust to changes in retinal conditions, ensuring that each cell receives the most effective stimulation intensity. This adaptivity, facilitated by our closed-loop operation, not only optimizes power consumption but also significantly enhances the quality and consistency of the visual perception restored by the prosthesis ensuring that the stimulation neither fail to stimulate or over-saturate the opsins.

1.4 System-Level Architecture

This section discusses the system architecture of our implantable retinal prosthesis. As shown in Fig. 1.6, our design strategically positions the implant epi-retinally, targeting the optogenetically modified ON bipolar cells through optical stimulation, while recording the spiking RGCs. This approach effectively combines the advantages of both epi-retinal and sub-retinal electrical prostheses.

The top-level block diagram of Fig. 1.7 illustrates the interaction of components within the implantable retinal prosthesis system. The system includes an external Digital Signal Processing (DSP) unit, which is connected to an image sensor mounted on goggles. This sensor is responsible for capturing visual stimuli from the environment. The captured stimuli are processed by the DSP unit to generate optimal stimulation patterns for optical stimulation.

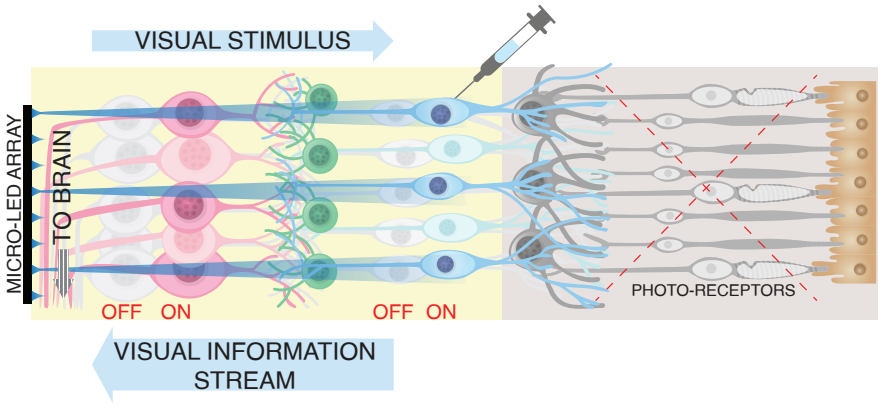


Figure 1.6: Envisioned placement of the proposed optogenetic retina prosthesis in contact with the RGC layer while stimulating the ON bipolar cells.

The processed data is then wirelessly transmitted to the implantable part of the prosthesis. This part consists of a flexible PCB equipped with power and data telemetry, facilitated by two pairs of coils for efficient energy transfer and data communication. Additionally, the PCB incorporates clock recovery systems, an Amplitude Shift Keying (ASK) data decoder, and voltage regulators. These regulators are tasked with providing stable 1.8V and 3.3V outputs, essential for the operation of the stimulator IC.

The signal and power regulation from this PCB are directed towards another PCB, which houses the optogenetic stimulator Integrated Circuit (IC). This IC is directly connected to a custom-designed 100-channel μLED array, which interfaces with the retinal cells. The stimulator IC is responsible for delivering precise optical stimulation, calibrated through a closed-loop control mechanism to ensure optimal intensity and efficacy.

Incorporated within the system are several novel innovations at both the system-level and circuit-level. Notably, the development of an ADC-less spike detection mechanism with adaptive thresholding stands out for its ability to fine-tune stimulation parameters, based on the immediate feedback from the retina.

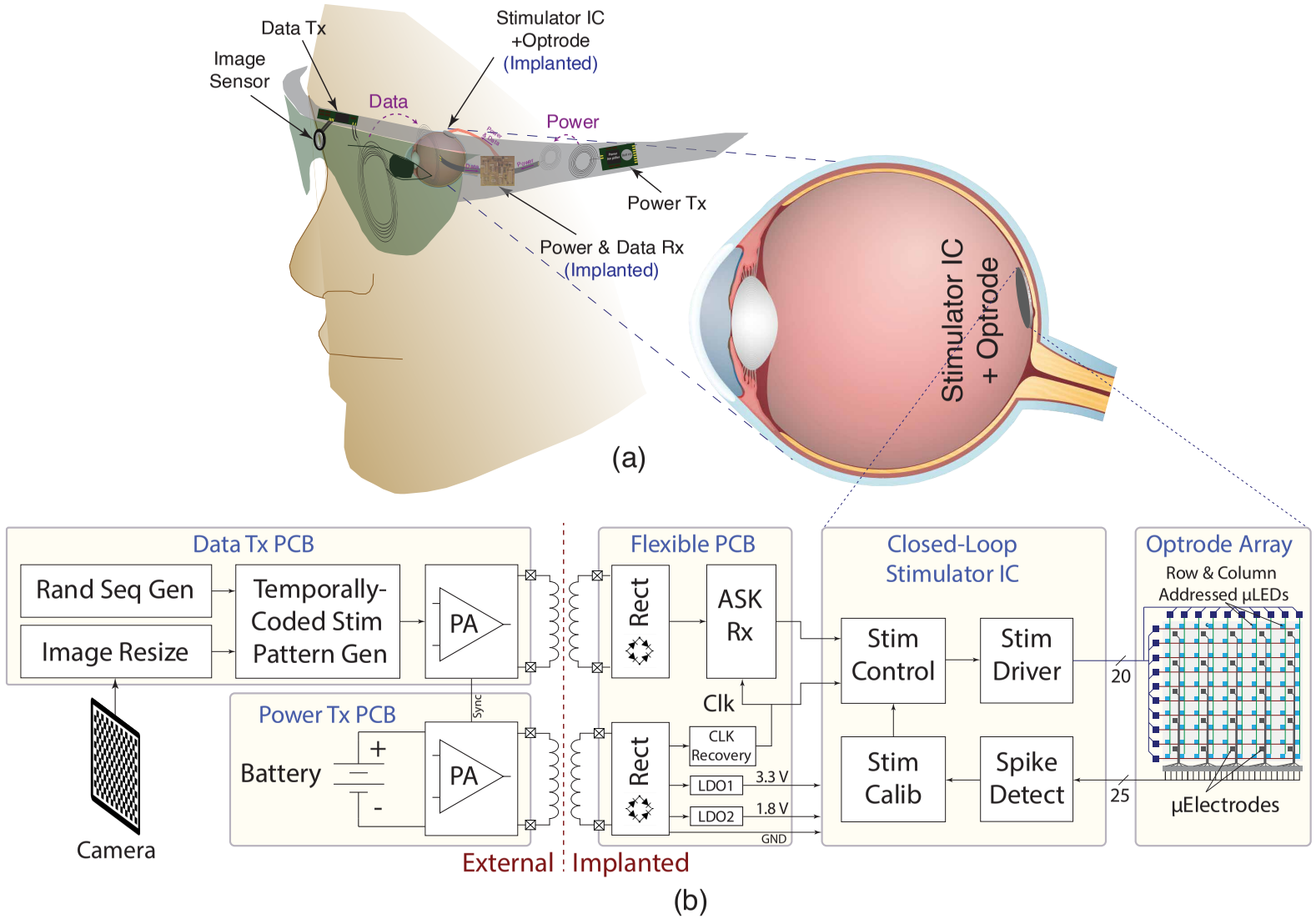


Figure 1.7: (a) Envisioned implantation of the retinal prosthesis and (b) Top-level block diagram of the proposed retinal prosthesis.

Additionally, the key to the system's power efficiency and scalability is the strategic utilization of photon integration feature and Poisson-coded, temporally-distributed stimulation. These methods help in managing the instantaneous power consumption and reducing the load on wireless data transmission. Additionally, the use of raster scanning technology significantly cuts down the number of required interconnections between the IC and the optrode array, enhancing the system's overall efficiency and scalability.

1.5 Dissertation Organization

The organization of the dissertation is as follows. The second chapter serves as a stepping stone for this project. It focuses on the development of an optogenetic neuro-stimulator, a critical component in designing our target retinal prosthesis. The chapter focus on the design, development, and experimental validation of a general-purpose optogenetic stimulator, applicable across a broad spectrum of optogenetic applications. A key concept introduced and validated here is duty-cycled stimulation, which is integral not only for its technical significance but also for its crucial role in the broader context of our proposed retinal prosthesis. This foundational concept sets the stage for future techniques and ideas explored in subsequent chapters.

The third chapter explores the feasibility of employing an implantable wireless optical neuro-stimulator for treating retinal degeneration. We compare its performance and potential with that of an electrical stimulator from various perspectives, motivated by the inherent cell-type specificity of optogenetics that enables pathway-specific neuron activation. The chapter presents a customized computational model developed to translate visual stimuli into optical

stimulation patterns, estimate the brain's visual perception for given stimulation patterns, and evaluate the impact of different optical stimulation parameters on inferred perception. This model facilitates a quantitative comparison of optogenetic and electrical stimulation, focusing on the quality of visual perception, scalability of the μLED/μelectrode array, and achievable spatial resolution.

The focus of the fourth chapter is on addressing the circuit- and system-level implementation challenges in designing an implantable retinal prosthesis specifically the generation of the target stimulation pattern and the optical stimulator circuit. We explore solutions to meet essential design requirements such as low power consumption, small size, reliable power and data links, electrode selection, stimulation pattern flexibility, and patient safety, all while ensuring scalability for high-density retinal prostheses.

The fifth chapter addresses the closed-loop calibration in our retinal prosthesis design, essential due to the variable distribution of opsins across the retina. This variability can lead to inconsistent light sensitivity, affecting visual perception quality. To combat this, we utilize a closed-loop system optimized for individual cell stimulation, minimizing power consumption. Enabling this closed-loop calibration requires a fully adaptable spike detection system, which is a key component of our design. This chapter also addresses the broader interest in recording brain electrophysiological activity with high channel count, an essential feature for applications like high visual acuity retinal prostheses and Brain-Machine Interfaces (BMI). We discuss the necessity of adaptively finding spike detection thresholds in systems with a high channel count and discuss methods to enhance the signal-to-noise ratio in environments

with significant background noise. The chapter outlines the proposed method for spike detection in the analog domain and explores strategies to accentuate spiking activities. Finally, we will demonstrate the operation of the closed-loop calibration, where the stimulation current is determined based on the sensitivity of the target cells which are assumed to have varying predetermined opsins concentrations.

The last chapter concludes the dissertation, summarizing the key contributions and publications resulting from this research. It also outlines future work trends and suggestions for further advancement in the field of retinal prostheses and related technologies.

Chapter 2

Implantable Optogenetic Stimulator SoC with Extended Optical Pulse Width Enabled by Duty-Cycled Stimulation

2.1 Introduction

The development of an implantable optogenetic neuro-stimulator SoC is a critical component in the broader project of designing our target pathway-specific retinal prosthesis for vision restoration in retinal degeneration cases. This chapter details the design, development, and experimental validation of a general-purpose optogenetic stimulator, suitable for a wide range of optogenetic applications. In this chapter, we introduce and experimentally validate the

concept of duty-cycled stimulation within the framework of developing an implantable optogenetic neuro-stimulator SoC.

The emphasis on duty cycled stimulation in this chapter is not merely a technical detail; it is a fundamental concept that enhances the functionality of the general-purpose optogenetic neuro-stimulator SoC in a broader spectrum of applications. More importantly, it plays a crucial role in the specific context of our proposed retinal prosthesis, laying the groundwork for the future techniques and ideas reported in the later chapters.

2.2 Fully-Implantable General-Purpose Optogenetic Stimulator

Optogenetics is widely considered as a promising solution for understanding the underlying mechanisms of complex brain functions (e.g., from sensory-motor integration to memory) and dysfunctions (e.g., various neurological disorders [29,30]). Its key advantage points compared to other neural stimulation methods are cell-type specificity, high spatial resolution, millisecond temporal precision, and rapid reversibility. Recently, it has also shown potential as an alternative treatment option for various brain disorders such as Parkinson's disease and epilepsy [31–33].

The main goals in the design of a fully-implantable optogenetic stimulator are (a) delivering the required irradiance (typically 1-10mW/mm^2) to the target tissues, (b) precise control of the optical pulse shape (i.e., intensity, frequency, duty cycle, etc.), and (c) maintaining the surrounding temperature within a certain safe limit, all while maximizing the energy efficiency. A typical optogenetic experiment requires driving the light source (i.e., a

μLED for implantable stimulators) with a 1-10mA, 1-100Hz current pulse, which translates into a power consumption in the order of 10s of milliwatts during stimulation [35, 36, 89]. This is too high to be continuously supplied through a weakly-coupled mm-scale wireless powering link, hence necessitates the use of a storage capacitor to accumulate energy during no-stimulation periods, and supply it during stimulation. However, using a capacitor as the power supply introduces challenges of its own. The μLED driver circuit should generate a constant current (I_{LED}) for the duration of the stimulation pulse, while its supply voltage is decreasing as the storage capacitor is being discharged by the same I_{LED}. One solution is to increase the storage capacitor's size to make the V_{DD} drop negligible even during the worst-case scenario. However, our calculations (Section III.B) show that to maintain V_{DD} variations <10%, a storage capacitor >10mF is required, which is too large (i.e., ∼1cm diameter with a thickness of ∼2.5cm) to be a part of an implantable device, even as a discrete component.

The above is the main reason that all reported implantable optogenetic stimulators, to the best of authors' knowledge, can only support a very limited range for frequency and light intensity. Optogenetic stimulators reported over the past decade can be grouped into two categories of head-mounted and implantable. Being external to the body, the head-mounted devices' design requirements are more relaxed in terms of size, power consumption, and heat generation [37–39]. However, the fully-implantable devices offer a significantly smaller form factor, allowing for a wider range of freely-moving behavioral animal studies, though typically at the expense of losing precise timing/magnitude control [40–42], or only being able to

provide limited current amplitude (e.g. 0.1-1mA [43]) and/or a very short stimulation pulse width (e.g. 0.5-2ms [42]).

In this chapter, as summarized in the next few paragraphs, we present circuit- and system-level techniques to meet the above-mentioned versatility, energy efficiency, and safety design requirements.

To widen the range of stimulation intensity and speed, we employ and experimentally validate an alternative stimulation approach that leverages the photon-integrative feature in opsins to achieve a significant size reduction in the required energy storage capacitor, thus facilitating a wider range of pulse widths and magnitudes for optical stimulation. The approach works based on the fact that opsins with slower kinetics (e.g. step-function opsins [89]) have a higher light sensitivity as they integrate incident photons, thus their responsivity is determined by the total photon exposure (i.e., total integrated power delivered during responsivity window) [90–93]. Inspired by this, we toggle the μLED light on and off during the stimulation phase with a frequency (f_{toggle}) that is beyond the opsin's responsivity bandwidth (e.g., >1kHz) to imitate a constant-light condition. As shown in Fig. 2.1, to keep the integrated power delivered to the target tissue during the stimulation phase (Δt_{stim}) at the same level as the conventional scenario (Fig. 2.1(a)), the high-frequency pulse (Fig. 2.1(b)), should have a current magnitude of $\frac{1}{D_{toggle}} \times I_{stim}$, where D_{toggle} is the toggling duty cycle. D_{toggle} will be set small enough to allow sufficient time for the inductive link to recharge the $C_{Storage}$. Intuitively speaking, the opsins see the high-frequency optical pulse trains as a constant light (see opsin's perception in Fig. 2.1(b)), similar to the way that human eyes

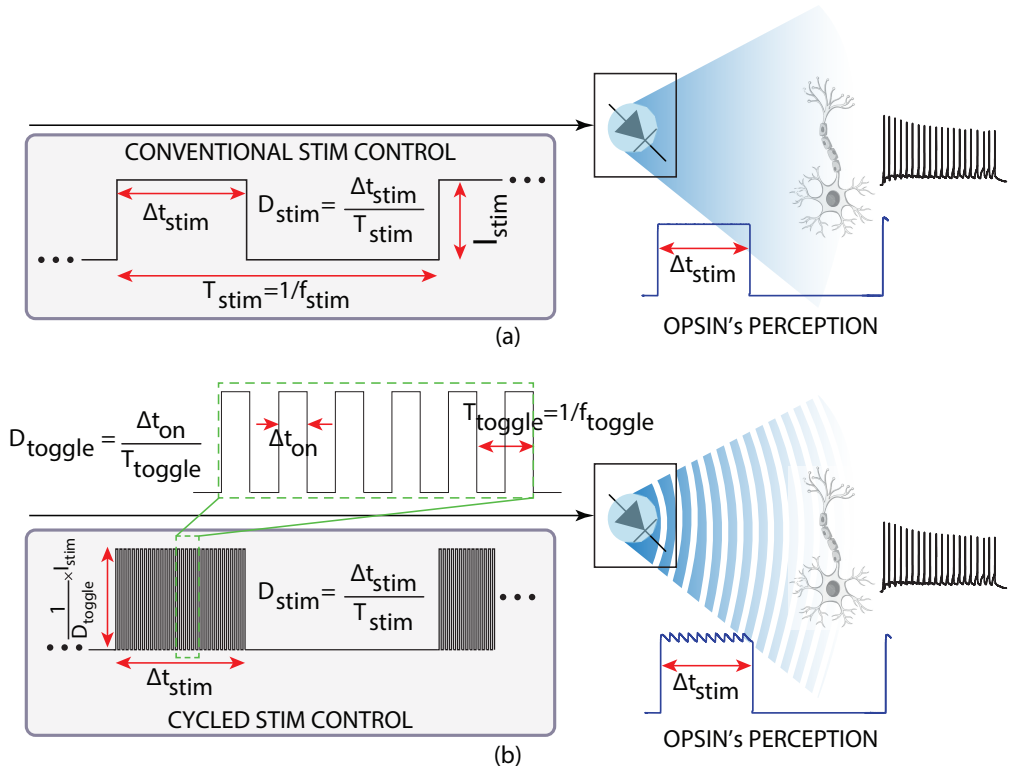


Figure 2.1: Stimulation control signals and opsin's perception for (a) conventional constant-light stimulation approach, and (b) the proposed high-frequency light-toggling stimulation approach.

see high-frame-rate picture display. While this approach will not reduce the overall required power for stimulation, we will describe in detail (in Section III.B) how we have used this approach to shrink the required storage capacitor's size. In Section VI, we will present in-vitro experimental results to validate the efficacy of the proposed light-toggling optogenetic stimulation approach and to compare the results to the conventional stimulation.

To enhance optical stimulation's energy efficiency, we have taken a two-pronged approach. First is employing a novel current-mode driver circuit in each stimulation channel that supplies up to 10mA to the μLED with excellent linearity (i.e., DAC+Amplification's $|INL|$ <0.5LSB at 3-bit level), while being capable of tolerating ~12% V_{DD} drop and requiring the smallest-in-literature voltage headroom. The driver can operate with a supply volt-

age as low as 2.9V, considering the V_{DD} drop (smallest reported for implantable optogenetic stimulators), which is a significant advantage in terms of stimulation power consumption and heat generation. Quantitatively speaking, knowing that the I_{LED} can be up to 10mA and that the $I_{LED} \times V_{headroom}$ is wasted, this reduction of the required V_{DD} could save tens of milliwatts. Additionally, the driver's high tolerance to V_{DD} drops makes it ideal for a battery-less implantable device that uses a capacitor as the power supply. Second is bonding a custom-designed inkjet-printed μlens onto the μLED to converge its inherently-divergent light beams, leading to a significantly larger portion of the emitted light reaching the target cells (>30× improvement). This light directivity enhancement proportionally reduces the required μLED current, thus the overall power consumption needed to deliver a targeted irradiance.

To prevent tissue overheating caused by optical stimulation, which can suppress neural firings and impact the animal behavior (e.g., affecting locomotor activity [95]), or damage the tissue, a temperature sensor as well as circuit blocks for continuous temperature monitoring and neuro-electrical recording, both simultaneous with the optical stimulation, are also integrated on the chip. This is in addition to our previously-mentioned efforts on energy efficiency improvement, which also lead to a reduction in heat generation.

All the above features have been tested through electrical and in-vitro biological experiments to validate the proposed stimulation approach and the functionality of the implemented circuitry. Additionally, this chapter introduces the light-toggling approach, analytically discusses the details of its implementation, and presents experimental results demon-

strating its efficacy.

The rest of the chapter is organized as follows. Section II describes the system architecture of the proposed fully-implantable all-wireless temperature-aware bidirectional optogenetic neuro-stimulator SoC. Section III explains the circuit implementation of the major blocks including the optical stimulator, temperature readout circuitry, electrophysiological recording and digitization, power management, and data telemetry. Section IV describes the modeling, design, optimization, and testing of the inkjet-printed μlenses. Section V describes the system-level specification of the presented SoC. Performance summary and comparison with the state-of-the-art are also included in this section. Section VI covers the in-vitro validation experiments and results, and Section VII concludes the chapter.

2.3 System Architecture

Fig. 2.2(a) shows the top-level block diagram of the presented implantable wireless and battery-less bidirectional optogenetic neuro-stimulator. As shown, in addition to the main blocks designed for optical stimulation, electrical recording, temperature sensing, power management, and wireless data communication, the microsystem integrates DSP unit for data processing and organization (i.e., serializing/deserializing, modulation/demodulation, etc.), clock generation, and timing control of all blocks. Fig. 2.2(b) shows how the inductive coil, the energy storage capacitor, and the temperature-sensing thermistor are interfaced with the SoC. The figure also shows the envisaged implantation of the presented freely-distributed inductively-powered microchip along with a 3D illustration of how the silicon SoC is interfaced with the four μLEDs and their encompassing dome-shaped μlenses used for light

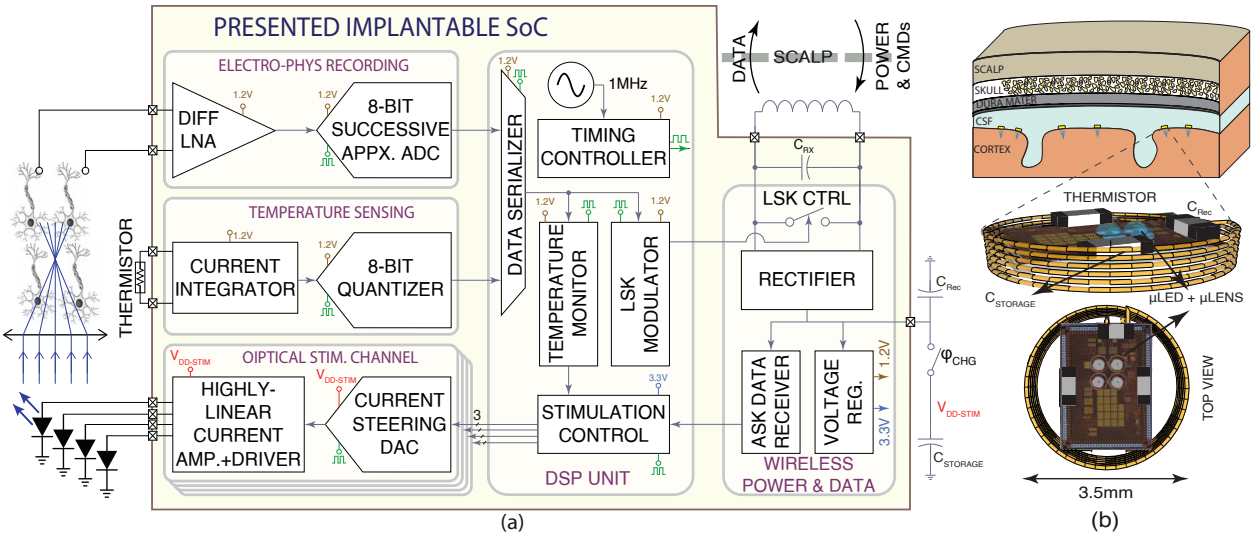


Figure 2.2: (a) Top-level block diagram of the presented SoC, and (b) the implantation and top-view of the envisaged free-floating mm-scale neurostimulator device.
directivity enhancement.

The electrophysiological signal recording and digitization circuits allow for monitoring the neural activity before, during, and after the optical stimulation. The recorded data is used to evaluate the efficacy of stimulation in real-time, and to adjust the stimulation parameters to excite/inhibit a certain neural activity. Considering the variations in opsins' activation threshold and their dependence on their proximity to the light source, this allows for adaptive adjustment of the stimulation intensity to the minimum required level that can activate neural activity.

Continuous temperature sensing and digitization circuitry are also integrated on the SoC to ensure that the temperature increase during stimulation does not exceed the safe limit [49]. Since the main sources of heat generation are the μLEDs (encompassed by transparent optical μlenses), an off-chip thermistor-based temperature sensor is chosen over an on-chip alternative to ensure the closest proximity of the sensing element to the μLEDs and the tar-

get tissue. More importantly, thermistors offer higher sensitivity to temperature variations compared to on-chip options (BJTs, resistors, etc.)[96], which significantly relaxes the read-out circuit's design requirement (described in details in section III.B), thus reducing their power consumption.

The recorded and quantized electrophysiological signals and temperature values are fed to a backend digital signal processing (DSP) block that (i) conducts data serialization and modulation for wireless transmission, (ii) monitors temperature variations and shuts off stimulation in case of detecting an unsafe level, and (iii) generates (using an on-chip 1MHz ring oscillator) and sends timing control signals to all blocks across the chip, including the stimulation channels, recording circuits, and wireless data and power telemetry circuits.

The system is wirelessly powered using an inductive link with an off-chip receiver coil. A rectifier and two regulators are used to generate 3.3V and 1.2V supplies for different circuits on the chip. Due to the aforementioned high instantaneous power required during optical stimulation, an off-chip capacitor (KYOCERA AVX, 02016D475MAT2A, 4.7μF) is used along with the inductive receiver to store the excess received energy during the stimulator's off cycles and to supply the μLEDs during the on cycles [50, 51]. While the LED branch is supplied from the storage capacitor, all other blocks are powered through the 1.2V and 3.3V regulators outputs. This prevents the stimulation from negatively impacting the performance of the recording and temperature sensing blocks.

Beside wireless power delivery, a truly implantable system requires a two-way wireless data communication so that the SoC can receive control commands and transmit the recorded

data from/to an external unit. In the presented systems as will be discussed in Section III.E, both communications are performed by reusing the inductive powering link. We have used load shift keying (LSK) to back-propagate the recorded data outside of the body (i.e., the uplink). Communicating control commands (e.g., for stimulation pulse magnitude control) from outside the body to the chip (i.e., the downlink) is carried out through amplitude shift keying (ASK) of the inductive powering waveforms.

2.4 Circuit Implementation

2.4.1 A Highly-Linear Small-Headroom Current Amplifier

The optical stimulator circuit must drive the μLED (typical I_{LED}:1-10mA, 1-100Hz) with high precision and linearity from a capacitive voltage supply ($C_{Storage}$). A major challenge is the high variability of the driver's V_{DD} (i.e., $C_{Storage}$'s voltage) during stimulation, because $C_{Storage}$ gets discharged by the large I_{LED}. At the extreme, the $C_{Storage}$ discharge could turn off the μLED if the V_{DD} drops down to $V_{D,ON}$ (i.e., the μLED's forward bias voltage). While this motivates for working with a V_{DD} that is much higher than $V_{D,ON}$, it should be noted that only $I_{LED}\times V_{D,ON}$ is converted to optical power, and (V_{DD}-$V_{D,ON}$)$\times I_{LED}$ is wasted, which is significant as I_{LED} could be up to 10mA. Therefore, the closer the V_{DD} is to $V_{D,ON}$ the better energy efficiency could be achieved. This requires the driver circuit to work not only with a variable V_{DD}, but also with the smallest possible required headroom voltage (i.e., minimum V_{DD}-$V_{D,ON}$=min($V_{DS_{M_D}}$)).

Fig. 2.3(a) shows the conventional wide-swing current mirror (WSCM) used for driving μLEDs [37]. The digital commands specifying the pulse magnitude, width, and frequency are

first converted to a current (I_{DAC}), and then amplified by the WSCM to drive the μLED. As shown, $C_{Storage}$ is used as the supply voltage of the μLED driver (I_{LED}) and its mirrored branch (I_{MID}). To minimize the circuit's power consumption, the current amplification gain (i.e., I_{LED}/I_{MID}) should be maximized, and to ensure linear performance, it should be independent of V_{DD} variations. The latter becomes very challenging for higher I_{LED} values (e.g., 10mA), for which the μLEDs $V_{D,ON}$ rises to almost 2.8V (even for the most efficient InGaN LEDs [52, 53]), while V_{DD} is rapidly decreasing due to $C_{Storage}$ being discharged by the large I_{LED}. The combination of the decreasing V_{DD} and large $V_{D,ON}$ shrinks the voltage headroom (V_{DD}-$V_{D,ON}$) and forces M_D and M_{CASC} to triode, deteriorating current mirror's gain and linearity.

Fig. 2.3(b) shows the proposed current driver, in which, two OTAs are employed that ensure a constant current amplification gain regardless of the operating point of the current mirror transistors (i.e., M_3 and M_D) or their available headroom. OTA1, similar to the WSCM, ensures that V_{DS} of M_D and M_3 are equal. OTA2, on the other hand, plays a feedback role between the two current mirrors in the circuit. It allows the gate voltage of M_D and M_3 to be automatically set to any voltage between 0 to V_{DD} that results in I_{MID}=I_{DAC}. This double-feedback configuration offers adaptive control of both V_{DS} and V_{GS} of both current mirrors and automatically adjusts them to compensate for any changes in the operating region of transistors, yielding a constant current amplification gain that is independent of I_{LED} and V_{DD} variations. Both OTAs are connected to the gate terminal of a MOSFET, hence, do not drive any significant current. Considering the high output swing

requirements, we used low-power folded-cascode architecture for both OTA1 and OTA2.

Fig. 2.3(c) compares the simulation results of the WSCM with the proposed circuit while a $C_{Storage}$ is used as the supply voltage. As shown, for the WSCM, the V_{DD} drop results in forcing M_D out of saturation and into the triode region (pink shade), which effectively disables the current mirror and leads to a significant I_{LED} drop. However, in the proposed circuit, the gate voltage of the mirroring transistors (M_D and M_3) is automatically adjusted such that the I_{LED} remains constant regardless of the operating point of these transistors and the varying supply voltage. The proposed architecture also stays longer in saturation as it has one less transistor in the LED branch.

2.4.2 Cycled Light-Toggling Stimulation

Even with the proposed architecture, the stimulator stops working if the V_{DD} drops below V_{Req}=$V_{D,ON}$ +min($V_{DS_{M_D}}$). For a ΔV=V_{DD}-V_{Req}, the smallest $C_{Storage}$ to ensure proper operation is,

$$C_{storage-min} = \frac{I_{stim} \times D_{stim}}{\Delta V \times f_{stim}}. \tag{2.1}$$

where I_{stim}, f_{stim}, and D_{stim} are stimulation current magnitude, frequency, and duty cycle, respectively. This suggests that for the challenging, but not uncommon, stimulation scenario of I_{stim}=10mA, f_{stim}=1Hz, D_{stim}=50%, a $C_{Storage}$ greater than 10mF is required (assuming ΔV=0.4V, i.e., 12% VDD drop), which is too large to be implanted even as a discrete component. To address this, as mentioned earlier, we have taken advantage of the opsins'

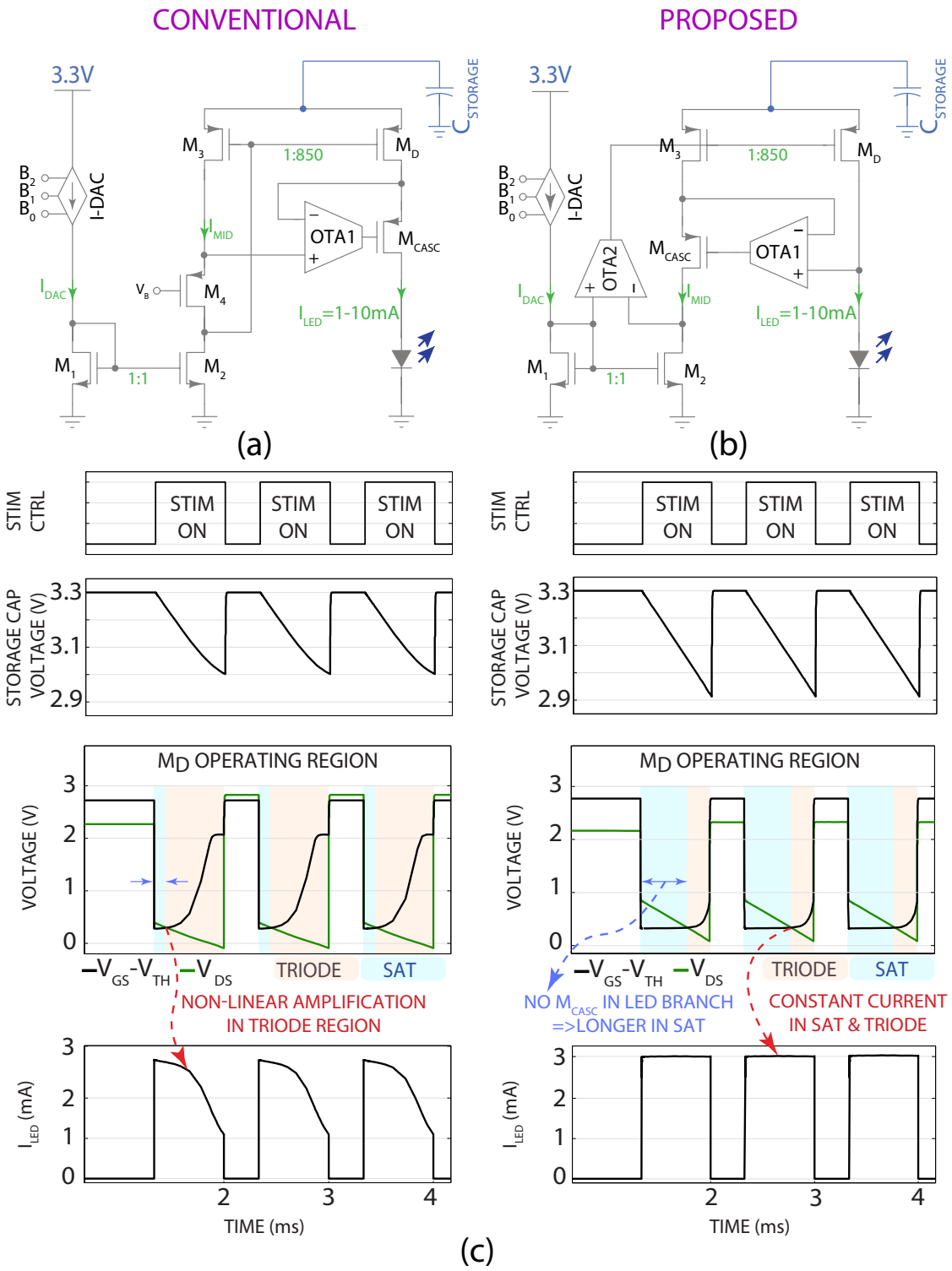


Figure 2.3: The circuit schematic of (a) the conventional wide-swing current mirror, and (b) the proposed current amplifier, when both employ a storage capacitor as the voltage supply. (c) Simulation results of the two circuits stimulating with 3mA and 1kHz while supplied by a capacitor.

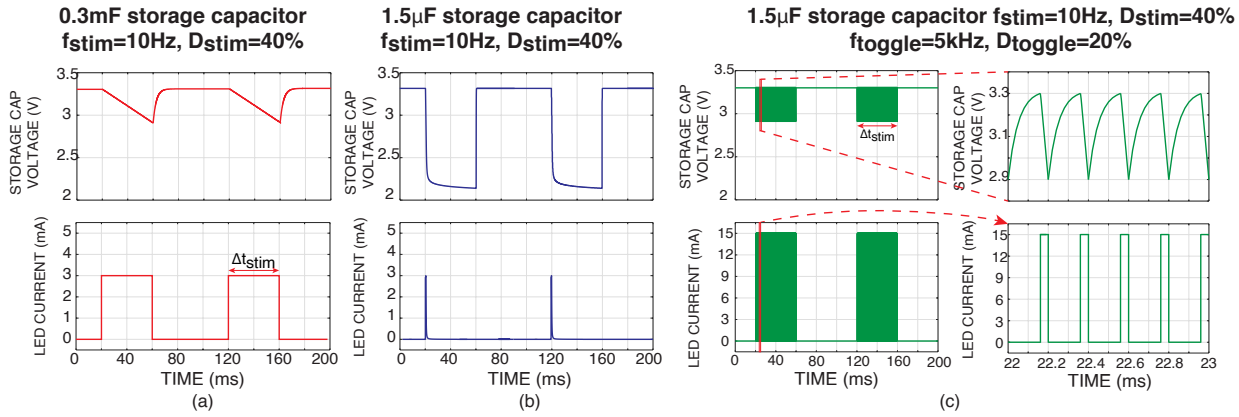


Figure 2.4: Simulation results showing the storage capacitor voltage and the μLED current of the proposed current driver under a capacitive supply and for stimulation scenarios of (a) constant light with 0.3mF storage capacitor, (b) constant light with 1.5μF storage capacitor, and (c) cycled light-toggling with 1.5μF storage capacitor.

photon integration property and toggled the LED on and off with a frequency higher than opsin's responsivity, so that it is perceived as a constant-light stimulation by the opsin [90–93]. This way, an optical stimulation episode with a constant current of I_{stim} for the duration of ΔT_{stim} is substituted by an episode of the same duration (ΔT_{stim}), in which, the LED turns on and off with a frequency of f_{toggle}, a duty cycle of D_{toggle}, and a current magnitude of $\frac{1}{D_{toggle}} \times I_{stim}$, as both scenarios deliver the same amount of integrated power to the opsin.

Using this approach, $C_{Storage}$'s size is defined based on I_{stim} and f_{toggle},

$$C_{storage} = \frac{I_{stim} \times \frac{1}{D_{toggle}} \times D_{toggle}}{\Delta V \times f_{toggle}} = \frac{I_{stim}}{\Delta V \times f_{toggle}}. \tag{2.2}$$

Recalculating the minimum required $C_{Storage}$ for the previously-discussed scenario (10mA, 1Hz, 50%) results in only 5μF when f_{toggle}=5kHz is used (same ΔV for both cases).

It should be noted that while the f_{toggle} is chosen to be much larger than the opsin's responsivity speed, it must also be much smaller than the frequency used for inductive reso-

nance power transfer (typically 1-20MHz [54]). In this chapter, we have used the 1-to-10 kHz range to satisfy both these conditions. The D_{toggle} is desired to be small to allow sufficient off time so that the inductive link can recharge the $C_{Storage}$. At the same time, it cannot be too small as the I_{stim} is scaled up by a factor of $\frac{1}{D_{toggle}}$ for the proposed cycled-stimulation approach.

Circuit simulation results illustrating the efficacy of the proposed approach in reducing the required $C_{Storage}$ are shown in Fig. 2.4. As shown, without the proposed cycled-stimulation approach, reducing the $C_{Storage}$ from a non-implantable size of 300μF (Fig. 2.4(a)) to a more reasonable size of 1.5μF (Fig. 2.4(b)) results in a complete failure of stimulation. As expected from theory, once the stimulation pulse starts, the 1.5μF $C_{Storage}$ quickly loses its voltage, which results in the μLED current dropping to zero. On the other hand, for the case where the proposed approach is implemented (Fig. 2.4(c)), the light-toggling provides sufficient time for the $C_{Storage}$ to recharge and keep the circuit functional throughout the stimulation phase (Δt_{stim}). In this example, an I_{LED}=15mA, is used which is toggled with f_{toggle} of 5kHz and D_{toggle} of 20% (resulting in an effective I_{LED}=3mA, which is equivalent to the cases shown in Fig. 2.4(a)).

The rate at which the light toggling method can reduce the capacitor size depends on a number of parameters including the stimulation current magnitude, maximum tolerated supply voltage drop (ΔV), $C_{Storage}$ size, available time for recharging $C_{Storage}$, and the maximum deliverable power through the wireless link. Generally, ΔV is set by how much variation the LED current driver circuit can tolerate (i.e., exhibits negligible nonlinearity),

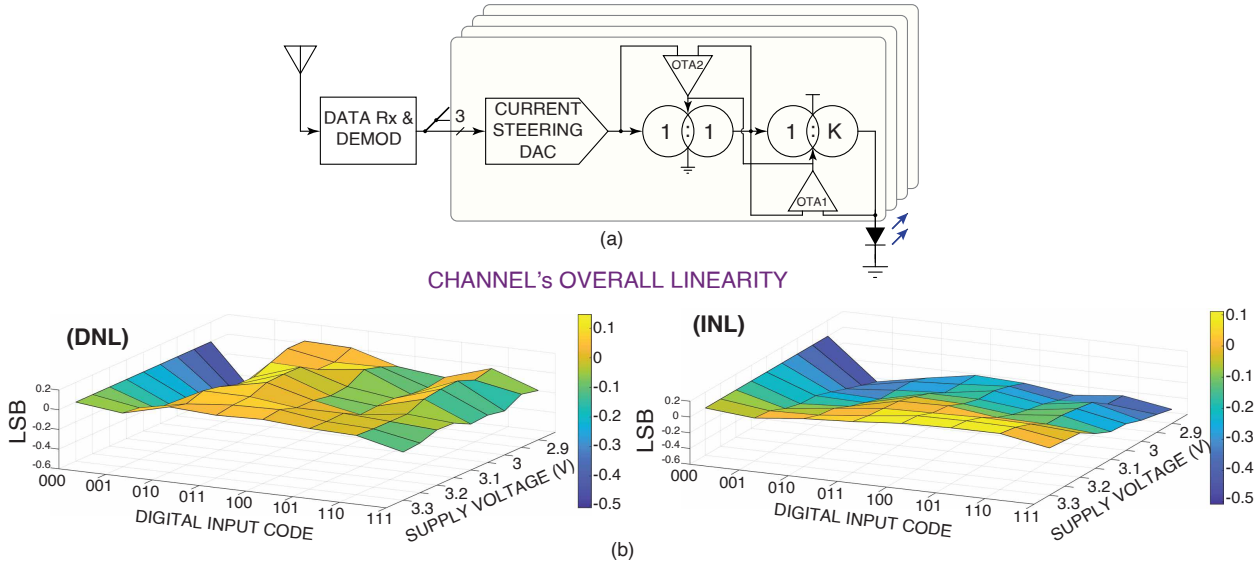


Figure 2.5: (a) Simplified block diagram, and (b) experimental measurement results showing the DNL and INL of the entire stimulation circuit (DAC + current amplifier/driver), with a varying supply voltage.

which is significantly increased in this chapter, thanks to the proposed circuit architecture.

The stimulation frequency and duty cycle are dictated by the application, therefore, if the proposed toggling approach is not used, the available time to recharge the storage capacitor is fixed. Employing the proposed light-toggling allows for breaking this application-enforced barrier by opening charging windows during stimulation. Of course, reducing the D_{toggle} allows for more time available for recharging, but comes at the cost of increasing the LED current proportionally. The required available time, among other things, also depends on the difference between the maximum power delivered through the wireless link and the required instantaneous power for stimulation. D_{toggle} is chosen as a trade-off between the above two criteria, so that sufficient recharging can be done within the available time and using the wirelessly delivered power, all while the LED current does not increase beyond a safe level. It is important to note that since this current passes through the LED and is not injected into the tissue (unlike electrical stimulation), the limit is mainly decided by the amount of heat

generation, which is comparable to a normal (i.e., no toggling) optogenetic stimulation due to proportional duty-cycle scale down. It is also very important to note that employing light toggling is possible not only because of the described biological phenomenon (i.e., photon integrative feature), but also because of the exponential I-V characteristics of LEDs, which allows for significant current increase with a slight $V_{D,ON}$ change. Therefore, the average power remains more-or-less the same. In contrast, for electrical stimulation, increasing stimulation current results in almost linear increase in voltage compliance, leading to a quadratic increase in the required instantaneous power, and a linear increase in the average power.

Fig. 2.5 shows a simplified block diagram of the stimulation channels as well as the experimental measurement results for the differential non-linearity (DNL) and integral non-linearity (INL) of the entire stimulator circuit (i.e., the 3-bit current steering DAC plus the current amplifier) under a varying supply voltage that changes from 2.9V to 3.3V. The results show that for supply voltage of 3.3V the DAC plus the current amplifier have $|INL|$ <0.1LSB and this number changes to $|INL|$ <0.5LSB when the supply voltage reduces to 2.9V. These measurements are conducted with the CREE TR 2227 LED, which is available in both blue and green colors.

2.4.3 Temperature Sensor and Readout Circuitry

While a wide input range is not typically required for bio-implantable temperature sensors, a high sensitivity (e.g., 0.1°C resolution) and small conversion time (e.g., <1ms) are critical for these applications [96]. There are several components in CMOS technology that can be used for temperature sensing such as BJT transistors and poly-silicon resistors. However,

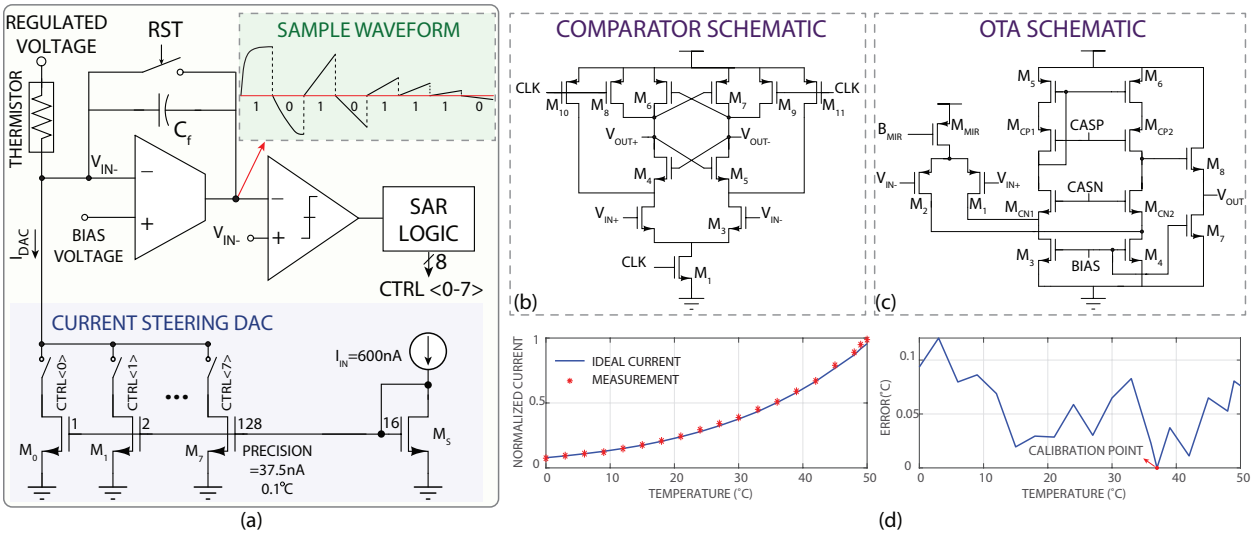


Figure 2.6: (a) Block diagram of the presented temperature readout circuitry, and the circuit schematic of (b) the comparator, and (c) the OTA used in it. (d) Experimental measurement results of the temperature sensor's readout output current (normalized) and error for 0-50°C.

considering the layers of encapsulation around the implant, the temperature sensed by an on-chip transducer is not an accurate representative of the surrounding tissue temperature. As such, we used an off-chip thermistor (ERTJ1VS104HA, Panasonic EC) to achieve a more realistic temperature reading. It also has a much higher sensitivity ($\Delta R/R$) when compared with on-chip temperature-dependent alternatives. The higher sensitivity relaxes the design requirements for the readout circuitry, hence, reduces the power consumption. Quantitatively speaking, For a temperature resolution of 0.1°C at 37°C, the readout circuit must resolve a ΔR of 300Ω in 40kΩ, achievable by a 7-bit quantizer. However, achieving the same sensing precision using an on-chip resistor or a BJT-based transducer requires at least a 12-bit quantizer [97].

Fig. 2.6(a) shows the proposed readout circuit for temperature sensing. The transducer (i.e., thermistor) is connected between a voltage regulator's output and the inverting input of an OTA, thus a fixed controlled voltage is applied across it. The current through the

thermistor, which is the transduced temperature, is subtracted from the output of a current steering DAC (I_{DAC}), and the result is integrated on a capacitor (C_f). The current integrator's output is fed to a comparator. To mitigate the effect of the voltage difference between the inverting and non-inverting inputs of the OTA, the positive input of the comparator is connected to the inverting input of the OTA. The comparator's output is fed to a SAR logic that controls the above-mentioned 8-bit current steering DAC. The updated I_{DAC} is subtracted from the $I_{Thermistor}$ and is fed to the integrator that has been reset (V_{out}=V_{in-}). To boost the quantization speed we used a current integrator with a small C_f (very high impedance at DC) instead of a current amplifier to maximize the DC gain without the need for a bulky passive resistor.

As shown in the sample waveform of the OTA's output in Fig. 2.6(a), this could result in a highly non-linear current amplification or even output saturation, which has no importance as the amplified current is fed to the most nonlinear circuit (i.e., the comparator) that only cares about the polarity of its input signal. After 8 clock cycles, I_{DAC} is set equal to $I_{Thermistor}$ with 8 bits precision and the output of the SAR logic is used as the digital equivalent of the temperature. Fig. 2.6(b) and (c) show the circuit-level schematics of the strong-arm comparator and the folded cascode OTA used in this design, respectively.

Fig. 2.6(d) shows the measurement results of the readout circuitry for 0-50°C, confirming the accuracy of the circuit in temperature monitoring (error of <0.12°C for conversion time of 1ms). This measurement is performed by changing the thermistor's surrounding temperature by steps of 3°C and reading the circuit's output. Considering the non-linear relationship

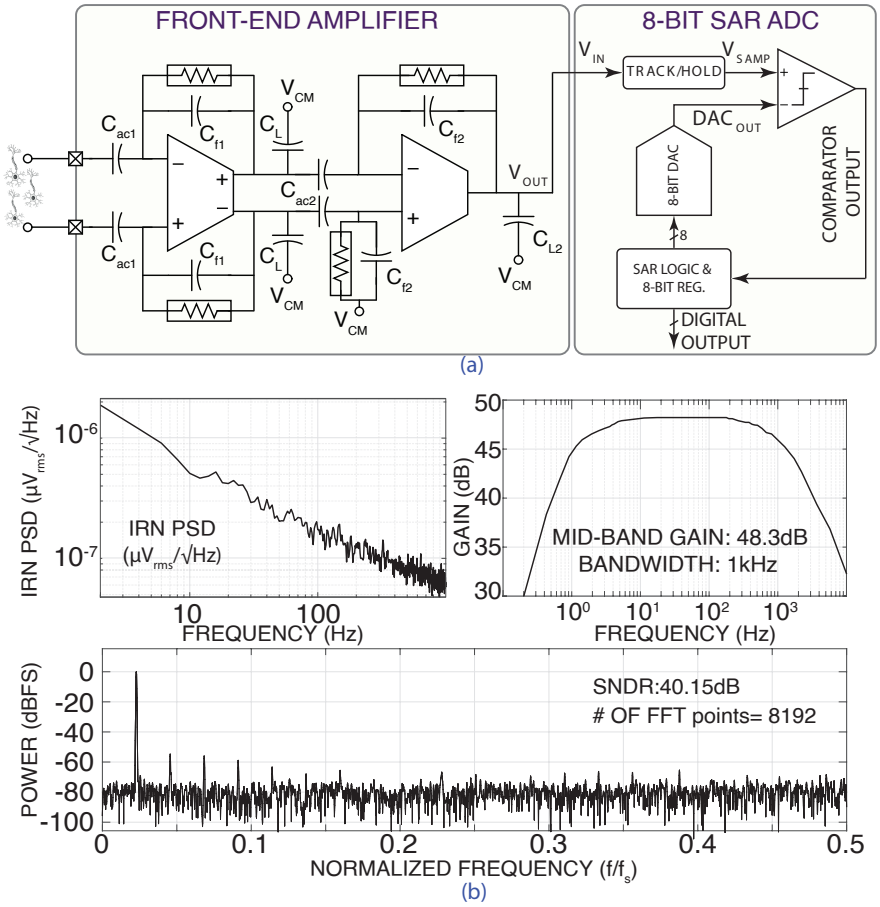


Figure 2.7: (a) Circuit schematic of the recording channel comprised of a two-stage amplifier and an 8-bit SAR ADC. (b) Experimental measurement results of the recording channel.

between the temperature and the thermistor current, a one-time calibration was conducted, during which the digital code for maximum safe temperature was determined and fed to the close-loop temperature monitoring system.

2.4.4 Electrophysiological Recording and Digitization

The presented SoC also includes an electro-physiological recording channel with a differential two-stage capacitively-coupled amplifier and an 8-bit SAR ADC as shown in Fig. 2.7. The experimental measurement results show that the neural front-end yields a mid-band gain of 260V/V, and an integrated input-referred noise of 7.79μVrms over 1kHz bandwidth. In addi-

tion, the SAR ADC achieves an SNDR of 40.15dB, meeting the gain/BW/noise requirements for high-quality neural recording [57].

2.4.5 Wireless Power and Data Telemetry

Inductive power reception is realized by a voltage-mode receiver followed by an on-chip rectifier and two regulators. The regulators generate a 3.3V for the stimulator branch (that uses thick-oxide MOSFETs) and a 1.2V for the remaining blocks on the chip. Logic level shifters are used for control signals traveling between high-voltage and low-voltage blocks (e.g., the signal that terminates optical stimulation in case of an unsafe temperature increase).

The link that we used for testing the data telemetry performance operates at 15.7MHz and yields a power transfer efficiency of 2.24%, with a coil distance of 0.4cm and C_{RES} of 30pF and Rx coil inductance of 3.4μH. It must be mentioned that we used this link to conduct benchtop characterization showing the efficacy of our high-frequency light-toggling idea and the data communication blocks. However, the presented idea for relaxing the wireless power transfer requirement is applicable to any inductive link regardless of their coil size, separation, and efficiency. To ensure safety and avoid tissue damage, the link's transmitter power will be set to keep the specific absorption rate (SAR) below the safe limit [58, 59].

The stimulation commands are communicated wirelessly to the chip (i.e., downlink) by amplitude modulation (ASK) of energy waveforms of the inductive link. As shown in Fig. 2.8, to decode the commands, the rectified version of the waveform at the inductive link's Rx coil is fed to a Schmitt trigger to extract the high and low levels. The transition point of the Schmitt trigger's voltage transfer characteristics is shifted up (without increasing the

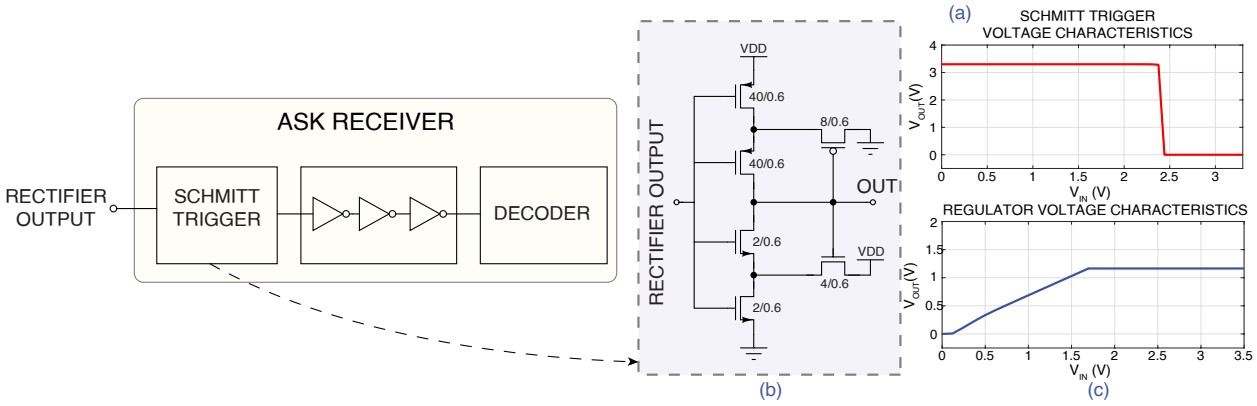


Figure 2.8: (a) a simplified block diagram of the ASK receiver, (b) the circuit schematic of the Schmitt trigger, and (c) voltage characteristics of the Schmitt trigger and the 1.2V voltage regulator.

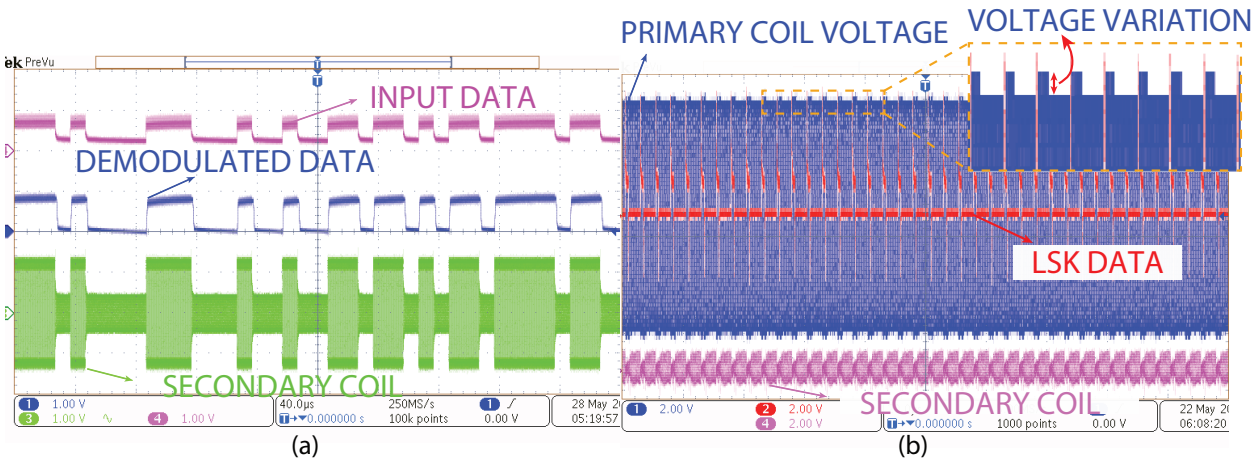


Figure 2.9: Experimental measurement results of (a) the downlink ASK receiver, and (b) uplink LSK transmitter.

bit error rate) to the “high" level of the ASK. This reduces the noise margin but ensures that the rectifier’s output magnitude is sufficiently high at all times (i.e., both high and low phases of ASK) for the 1.2V voltage regulator to remain functional. Fig. 2.9(a) shows the measurement results of the input data, ASK recovered data, and the voltage across the secondary coil, from top to bottom, respectively.

Load Shift Keying (LSK) modulation is employed to transmit the digitized neural signals and temperatures to outside of the body (i.e., the uplink). The simplicity of the LSK

transmitter allows for minimizing its required on-chip power and transferring the complexity to the external receiver where the power and area constraints are not as strict. In our implementation, the data is transmitted by changing the loading of the secondary coil through a switch connected across its terminals, which induces a subsequent amplitude change on the primary coil [60]. To minimize the impact of the LSK switchings on the wireless power transfer efficiency, an edge detection data modulation is adopted to reduce the time duration that the secondary coil is shorted (i.e., out of resonance). In this method, the LSK modulator sends a short pulse only when the LSK data changes [61] and the switch is open otherwise. At the external unit, the variations in the envelope of the signal at the primary coil are detected and used to demodulate the LSK data. Fig. 2.9(b) shows the measurement results of the voltage across the primary coil, input data, and the voltage across the secondary coil. Considering that both the electrical recording and temperature sensing are operating at a maximum 10kS/s, and including 10% overhead for data wrapping and packaging, the link was designed and tested to support a minimum data rate of 22kSps.

2.5 μlens Design, Optimization, and Fabrication

Fig. 2.10(a) shows a conceptual view of the effect of adding a dome-shaped μlens on top of a μLED in order to enhance directivity of its radiation pattern. We designed and optimized the μlenses based on the target stimulation area and its distance from the μLED. In the developed model, the refractive index, absorption coefficient, and thermal conductivity of the lens material, cerebrospinal fluid (CSF), and brain tissue were incorporated. Inkjet printing was chosen to fabricate the optimized μlenses. This fabrication method is particularly

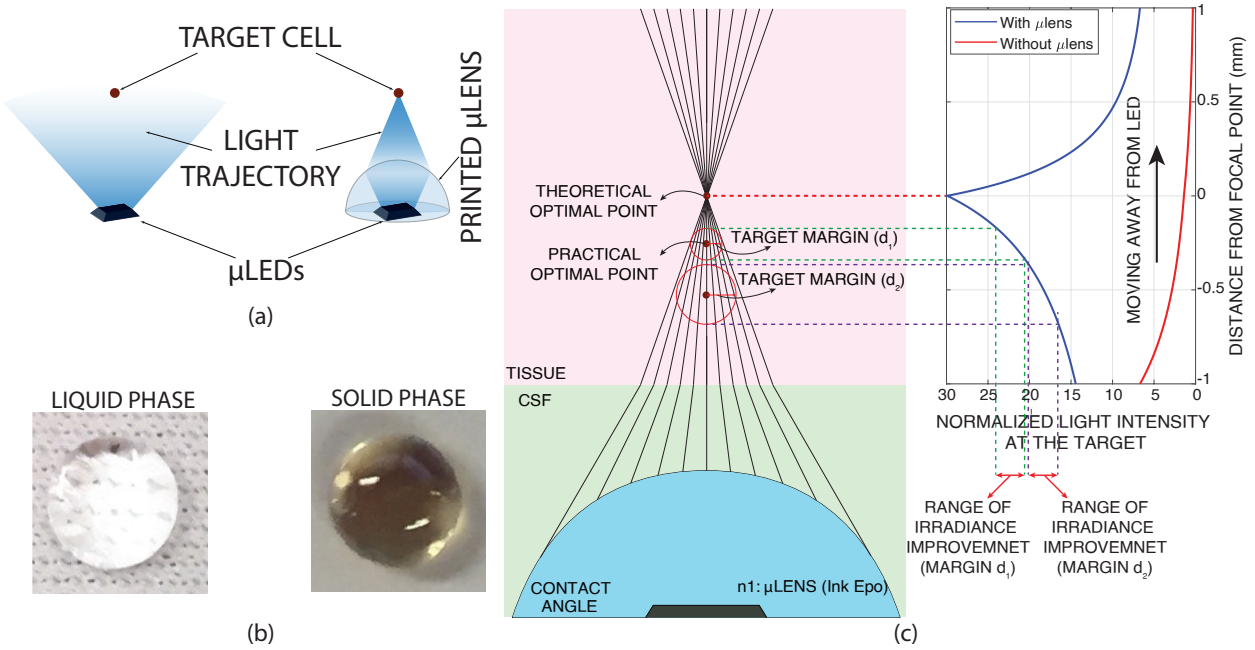


Figure 2.10: (a) Effect of using a μlens for converging the LED light. (b) Inkjet-printed μlens before and after solidification. (c) Irradiance improvement at the theoretical optimal point and more practical optimal points when there is a margin for displacement error.

advantageous due to being easily scalable and highly customizable. A SU-8-based polymer (light transparent and biocompatible) called InkEpo (Microresist technology, Germany) was selected as the ink material for fabricating the μlenses. The μlenses are inkjet printed using the process described in [62].

The improvement in stimulation energy efficiency caused by the μlens depends on how steady the focal point is, which could deviate from the actual targeted point due to the variations in the electrical and physical properties of the layers between the implant and the target cells. As shown in Fig. 2.10(c), even with such variations, the use of a light-converging μlens leads to an improved irradiance at the target compared to a lens-less situation, especially considering that it is a passive component and comes at no extra power cost. Additionally, thanks to the quick fabrication and assembly process of inkjet-printed μlenses, any shape adjustments/reoptimizations could be performed during a calibration phase prior to long-term

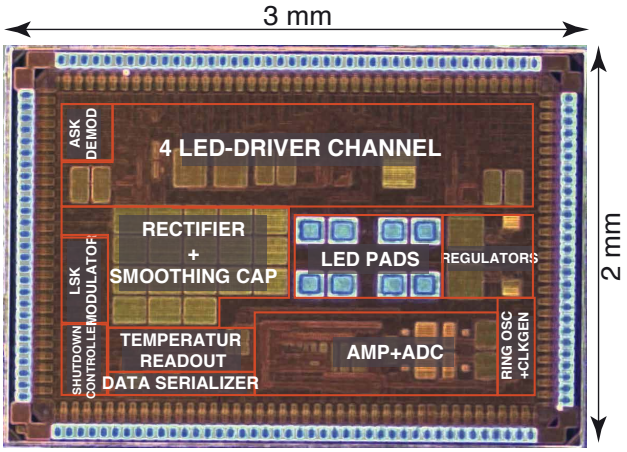


Figure 2.11: Annonated chip micrograph

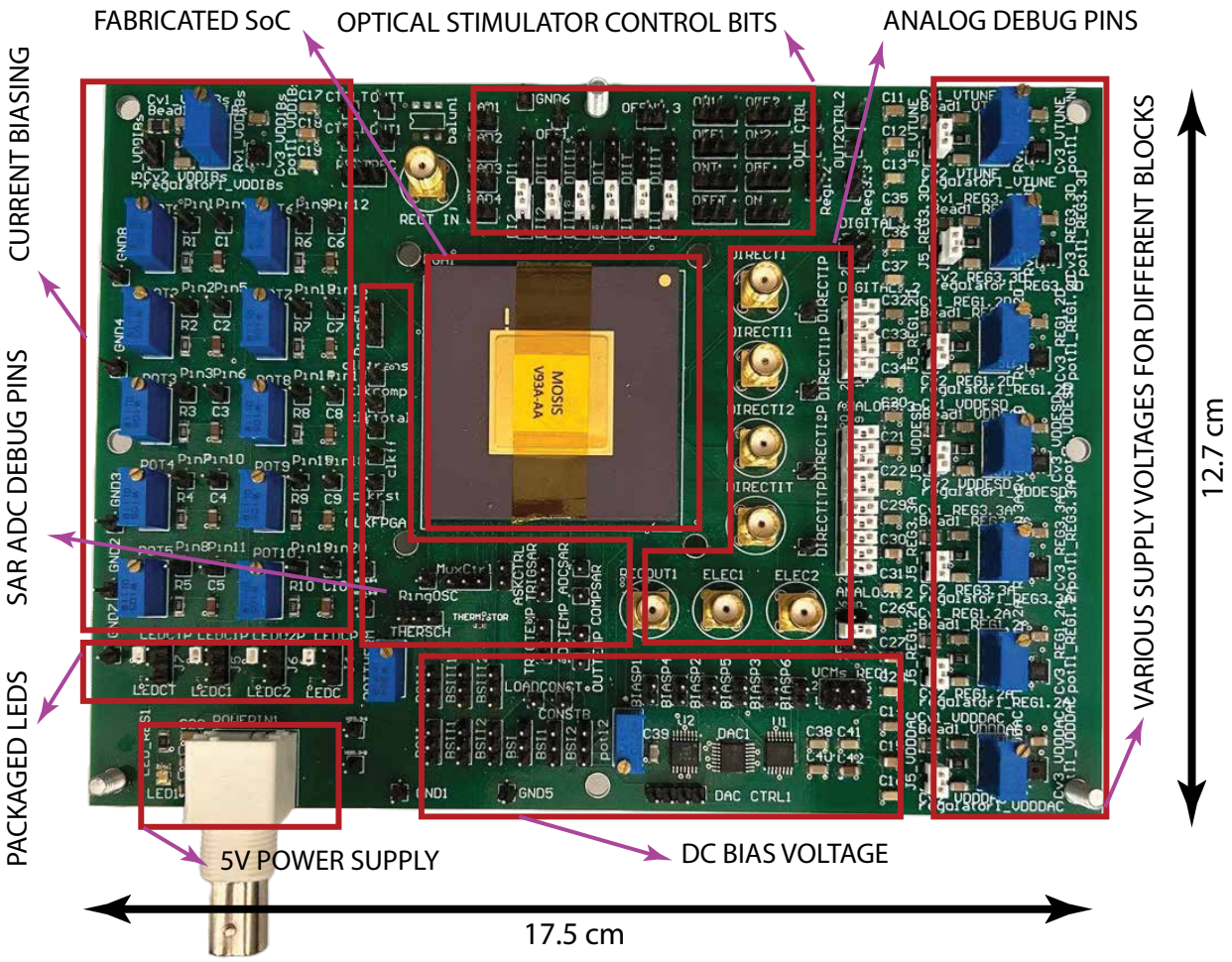


Figure 2.12: Designed PCB board for electrical verification with the main blocks annotated.

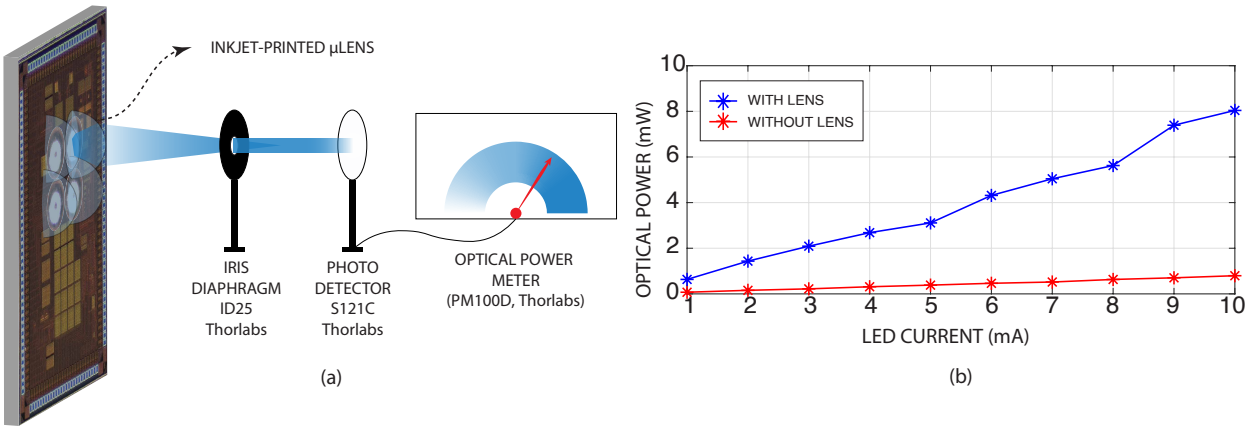


Figure 2.13: (a) Optical power measurement setup. (b) Optical experimental measurement results showing irradiance improvement after adding the μlens, for μLED current range of 1-10mA and an aperture diameter of 2mm.

Table 2.1: COMSOL simulation results with and without the μlens to reach the same irradiance at the target point.

	Without μlens	With μlens
Irradiance at target	10mW/mm^2	10mW/mm^2
Source optical power	6.37mW	0.215mW
Required LED current	15.72mA	0.52mA
Max temperature increase	1.06°C	0.03°C

implantation. Nevertheless, to ensure practicality, we decided to leave a margin, illustrated in Fig. 2.10(c), for any displacement between the target tissue and the μlens focal point. This target margin will define a spherical area as the practical optimal target which is different from the theoretical target point. As shown in Fig. 2.10(c), the smaller this margin is chosen, the higher irradiance improvement is achievable.

Table 2.1 reports the results of optical COMSOL simulations showing the effect of adding the μlens on the stimulation's energy efficiency. It confirms that the μlens converges the μLED light on the target tissue without exceeding the safe temperature limits. It compares the two scenarios with and without a μlens in terms of the required power to deliver the same irradiance (10mW/mm^2) to the target tissue. The results show a remarkable power

Table 2.2: Presented SoC's Performance Summary

SYSTEM LEVEL		**ELECTRICAL RECORDING**	
SoC area	6mm^2	Supply voltage	1.2V
Technology	TSMC 130nm	Supply current	48μA
# of stim channels	4	Voltage gain	48.3dB
# of rec channels	1	Bandwidth	1kHz
Temperature sensing	Yes	IIRN	7.79μV_{RMS}
OPTICAL STIMULATION		ADC resolution	8-bits
Supply voltage	3.3V	ADC SNDR	40.15dB
Supply current	0.1-10mA	**LIGHT CONVERGING μ-LENS**	
DAC resolution	3-bits	Fabrication method	Inkjet printing
Stimulation frequency	DC-100Hz	Material	InkEpo
Current gain	850	Spatial resolution	100μm^2
Stim pattern	Arbitrary (3-bit)	Irradiance improvement	30.46×
TEMPERATURE SENSING		**POWER MGMT AND DATA**	
Sensing element	Thermistor	Regulators	1.2V and 3.3V
Supply voltage	1.2V	Quiescent current (combined)	25μA
Supply current (@37°C)	30μA		
Logic	8-bit	Uplink data telemetry	LSK (22kbps)
Resolution	0.12°C	LSK data modulation	Edge detection
Conversion time	1ms	Downlink data telemetry	ASK

saving of >30× when a μlens is used. The other significant advantage shown from COMSOL temperature distribution simulation of the case using the μlens is that achieving the target irradiance is possible without exceeding the safe temperature limits, which is not true for the case without the μlens that increases the temperature by more than 1°C. The μLED current is calculated assuming wall plug efficiency (i.e. the μLED's electrical-to-optical energy conversion efficiency) of 15%, taken from the datasheet of the LED used in this system (CREE TR2227).

2.6 System-Level Specifications

The presented SoC was fabricated in a 130nm standard CMOS technology with a total area of 3×2mm^2. The optical μlenses were inkjet-printed and placed on top of the μLEDs (CREE

TR2227), which were assembled onto the aerial pads that are annotated on the chip micrograph in Fig. 2.11. The picture of the PCB board that has been designed for electrical verification of the fabricated SOC is shown in Fig. 2.12. In order to evaluate the effect of μlenses on the delivered irradiance to the target area, the μlens is directly printed on the package-less μLED. This configuration is used in the optical power measurement setup shown in Fig. 2.13, which achieves an average of 9.2$\times$ improvement in the amount of optical power that passes the IRIS diaphragm with a diameter of 2mm. Based on the COMSOL simulation results, we expect the improvement to be stronger for smaller aperture diameters, which we didn't have access to for our experiments.

Table 2.2 summarizes the presented SoC's electrical, optical, and system-level performance measurement results. Table 2.3 summarizes the SoC's comparison with the state-of-the-art, both in terms of the circuit- and system-level performance parameters. The device is capable of driving μLEDs (up to 10mA) at the lowest reported supply voltage and enhances the light directivity by more than 30$\times$, which together result in an excellent overall system's energy efficiency. Compared to the state of the art, it integrates the most complete set of features for an optogenetic stimulation, which includes electrophysiological recording, temperature sensing, wireless power, and bidirectional data telemetry. Additionally, thanks to the proposed light toggling approach, the required size for energy storage capacitor is significantly reduced, yielding an estimated 3-5mm thickness for the presented implantable device.

Table 2.3: Comparison with the state-of-the-art

		[40]	[41]	[37]	[43]	[42]	[62]	This work
Technology		COTS	Custom Fab	0.13μm	0.35μm	0.35μm	0.13μm	0.13μm
Device Type		Implant	Implant	Head-mounted	Implant	Implant	Implant	Implant
# of Stim Channels		1	4	4	6	16	2	4
LED Branch Supply		NA	NA	3.7	5	5	3.3	3.3
Type		NA	NA	Battery	Regulator	Storage Cap	Regulator	Storage Cap
Light Directivity Enhancement		No	No	No	No	Yes	Yes	Yes
Optical Stim	Max I	NA	NA	35mA	1.1mA	10mA	15mA	15mA
	Current Ctrl	No	No	Yes(1-bit)	Yes(8-bit)	Yes(2-bit)	Yes(3-bit)	Yes(3-bit)
	Pattern Ctrl	No	No	No	Biphasic	No	Arbitrary	Arbitrary
Extended Stim Capacity		No	No	Yes	No	No	No	Yes
(I_{LED}:1-10mA, 1-100Hz)		-	-	(Battery-pwr)	-	-	-	(Cycled stim)
Electrical Recording		No	No	Yes	Yes	No	Yes	Yes
(# of Channels)		-	-	10	4	-	2	1
ADC		No	No	$\Delta\Sigma$(14-bits)	SAR(10-bits)	No	SAR(8-bits)	SAR(8-bits)
Temperature Sensing		No	Yes	No	Yes	No	No	Yes
(Sensing Element)		-	Resistor	-	Resistor	-	-	Thermistor
Temperature Readout Res		-	NR	-	6-bits	-	-	8-bits (0.12°C)
Temperature Control		No	No	No	Yes	No	No	Yes
Wireless Data Telemetry		No	No	COTS	COTS	on-chip	No	on-chip

2.7 In-Vitro Experiments

Three in-vitro experiments were conducted to further validate the presented device, and particularly, the proposed high-frequency light-toggling stimulation approach. The objective of the first experiment was to compare the biological effect of a constant-light stimulation with the proposed light-toggling approach. In this experiment, calcium imaging was chosen to allow precise quantification of the activation level with single-cell resolution.

Mouse Neuro2a cells were transfected with pAAV-CaMKIIa-hChR2(H134R)-mCherry on 35mm glass-bottom dishes (MatTek Corporation, Ashland, MA, USA) for 48 hours. Cells were washed with Phosphate buffered saline (PBS) and incubated in DMEM$^{(-phenolred)}$ containing 5μM Oregon GreenTM 488 BAPTA-1, AM Cell Permeant (OBG) calcium indicator (Thermo Fisher Inc., Mississauga, ON, Canada) for 20 minutes. A washing step with PBS was performed prior to experimentation. Cell images were captured 48 hours post-transfection

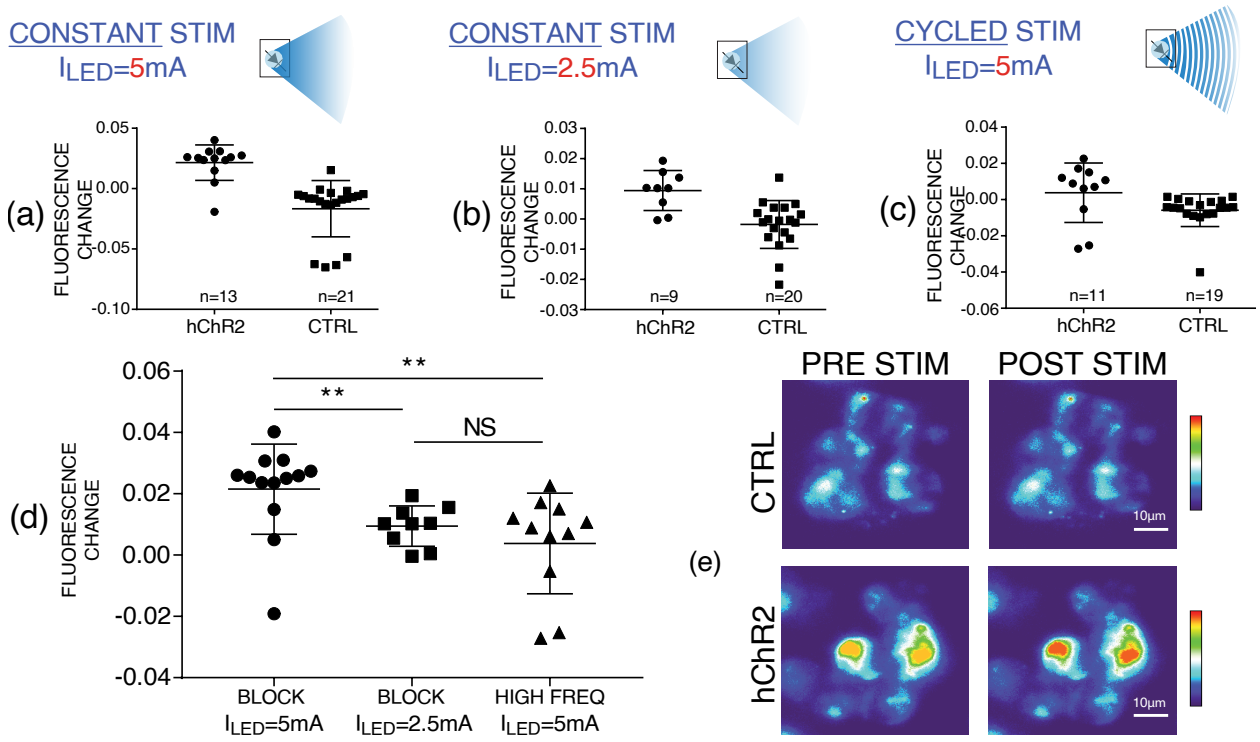


Figure 2.14: In-vitro measurement results for optogenetic modulation of calcium concentration (average and distribution shown) in hChR2-transfected and CTRL cells using the presented device: (a) 5mA constant stimulation, (b) 2.5mA constant stimulation, (c) 5mA cycled stimulation, (d) comparison of the three above cases. (e) Measurement results from the second in-vitro experiment comparing expression of hChR2(H134R)-mCherry in hChR2-transfected cells versus CTRL cells before and after stimulation.

using the Zeiss Observer Z1 spinning disk confocal with Evolve TIRF camera under 40× magnification.

The quantifications of changes in the intracellular calcium levels from cells expressing hChR2(H134R)-mCherry and the untransfected control cells (CTRL) are shown in Fig 2.14(a-c), where all cells were stimulated for a duration of 5 seconds (Δt_{stim}=5s) and the images right before the onset of stimulation and right after the stimulation cessation are taken into account to calculate the level of activation. Fluorescence change was calculated by subtracting ΔF/F of background from ΔF/F of cells, where ΔF is fluorescence$_{poststim}$-fluorescence$_{prestim}$ and F is fluorescence$_{prestim}$. Cells expressing hChR2 showed a significantly

higher calcium fluorescence change compared to the CTRL group when stimulated by LED current of 5mA for Δt_{stim} of 5s (hChR2: 0.022±0.015, n=13; CTRL: -0.017±0.023, n=21; p<0.0001)(Fig 2.14(a)). A constant light stimulation with current of 2.5mA for Δt_{stim} of 5s also showed an increase in calcium fluorescence change compared to the CTRL group (hChR2: 0.0094±0.0066, n=9; CTRL: -0.0018±0.0079, n=19; p=0.0008) (Fig 2.14(b)). Light-toggling with f_{toggle} of 5kHz and D_{toggle} of 50% with current of 5mA resulted in an increase of intracellular fluorescence compared to the CTRL cells (hChR2: 0.0038±0.016, n=11; CTRL: -0.0059±0.0090, n=19; p=0.016) (Fig 2.14(c)). Therefore, a rise in the intracellular calcium levels was detected in response to all three of the light stimulation conditions. As shown in Fig 2.14(d), comparing the level of fluorescence changes in these three scenarios illustrates that the constant light stimulation with LED current of 5mA resulted in the greatest fluorescence change (p=0.0056) and no significant difference in fluorescence changes was observed between the constant light stimulation with LED current of 2.5mA and light-toggling with f_{toggle} of 5kHz and D_{toggle} of 50% with current of 5mA (p=0.77) which confirms the proposed substitute stimulation approach. Values reported are the mean±standard deviation. Images were analyzed using ImageJ (Image Processing and Analysis in Java) and the graph was generated using Prism (GraphPad). The designed experiment proved that constant light stimulation with I_{stim} for duration of Δt_{stim} has the same effect as stimulation scenario that the LED light toggles with f_{toggle} and D_{toggle} and current of $\frac{1}{D_{toggle}} \times I_{stim}$ for the same duration.

Fig 2.14(e) presents the results of the second experiment, in which the expression of

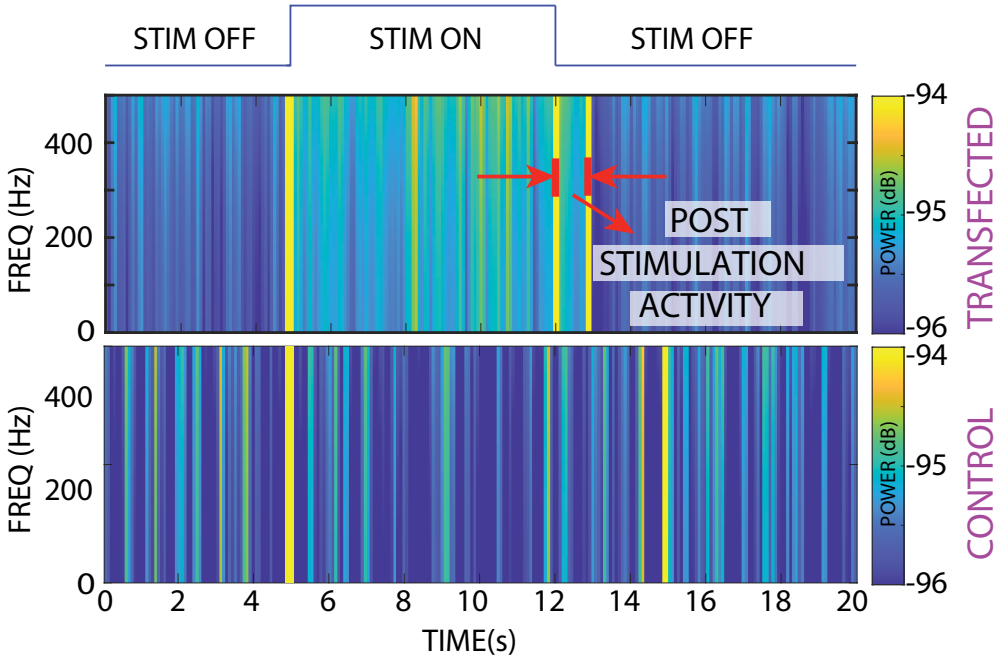


Figure 2.15: In-vitro measurement results of simultaneous optical stimulation and electrical recording showing the power spectral density of the recorded neural activity for the hChR2-transfected and CTRL cells.

hChR2(H134R)-mCherry in transfected Neuro2a cells versus the CTRL cells before and after stimulation are shown. The cell images are a combination of both channels (green OGB and red rhodopsin), imaged before and after stimulation (450nm LED) and using the Plan Aprochromat 63×/1.4 Oil Objective (Scale bar: 10μm).

In the third in-vitro experiment, we conducted simultaneous electrical recording and optical stimulation. The cell preparation process is similar to the previous experiment with the single difference that there is no need for adding the calcium indicator (OGB). The experiment was performed by connecting the differential inputs of the electrophysiological recording block to two unipolar needle electrodes (EL452, BIOPAC Systems, Inc.) and placing both inside the dish. Fig. 2.15 shows the power spectral density of the recorded neuro-electrical signal before, during, and after optical stimulation, for the hChR2 transfected cells, which

illustrates increased activity during and after optical stimulation, confirming effective activation of neural activity. While some photoelectric effect is observed at the positive and negative edges of stimulation, they are distinct from the actual neural activity that occurs post-stimulation (annotated on the figure). Additionally, Fig. 2.15 shows the power spectral density of the recorded signals for a similar optical stimulation scenario but with the CTRL cells, in which no increased activity is observable during or after the stimulation.

Chapter 3

A Retina-Inspired Computational Model for Stimulation Pattern Optimization of Implantable Optogenetic Neuro-Stimulators

In this chapter, through development of a customized computational model, we will investigate the feasibility of employing an implantable wireless optical neuro-stimulator (e.g., our recently-reported device in [62]) as a treatment option for patients with retinal degeneration, and compare its performance and potential to an electrical stimulator from various aspects. This is mainly motivated by the optogenetics' inherent cell-type specificity that enables pathway-specific activation of neurons, allowing for a visual perception quality that

is practically unachievable with electrical stimulation. Optogenetic stimulation requires genetic modification of the degenerated retinal cells using microbial opsins to restore their light sensitivity. While this has been done in the past, it has been shown (e.g., in [7]) that the opsins' sensitivity to ambient light is far less than a healthy eye's photo-receptors. As such, we propose to use an implantable optical stimulator to activate genetically-modified opsin-transfected cells of a retina layer (i.e., bipolar cells or RGCs).

The presented customized computational model is developed based on a widely-used framework for modeling human vision and is designed to be capable of (a) translating a given visual stimulus input to optical stimulation patterns, (b) estimating the brain's visual perception for a given optical stimulation pattern, and (c) evaluating the impact of varying optical stimulation parameters (e.g., μLED array size, light intensity, light wavelength, illumination divergence angle, etc.) on the inferred perception. Using this model, we have compared optogenetic and electrical stimulation performance and potential for an implantable medical device aimed at restoring vision in patients with retinal degeneration. The quantitative comparison is made in terms of the quality of the inferred visual perception, the scalability of the μLED/μelectrode array size and density, and the corresponding achievable spatial resolution. The model is also used to investigate the sensitivity of the visual perception to various optical stimulation parameters (e.g., μLEDs' type/wavelength, array pitch and density, etc.).

The rest of the chapter is organized as follows. Section II explains the significance of cell-type-selective activation in a retinal prosthesis and highlights the advantageous of opto-

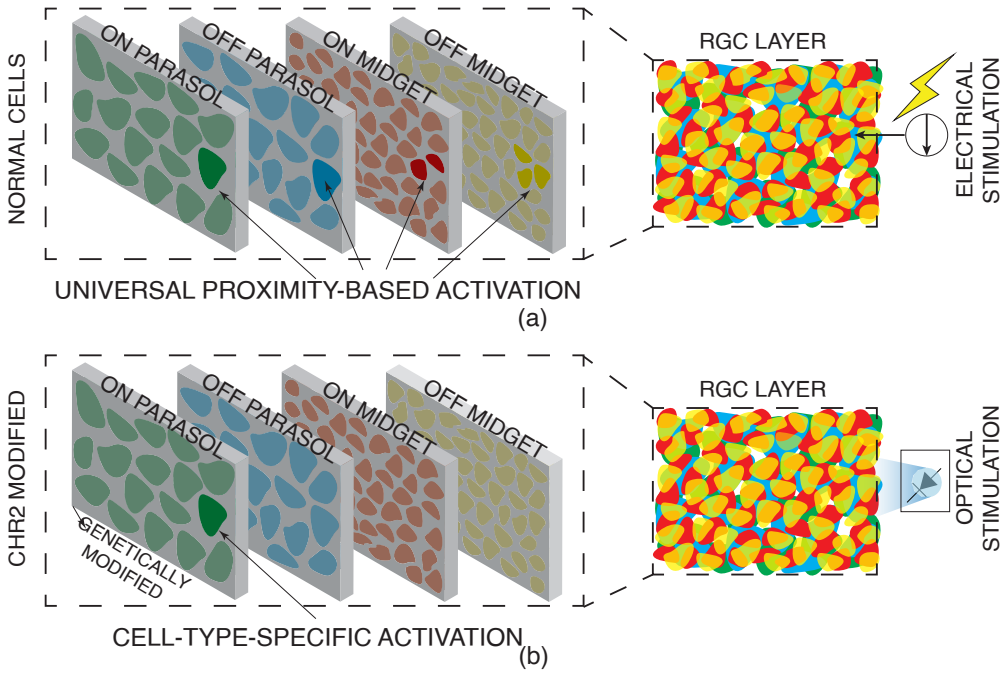


Figure 3.1: (a) Electrical stimulation evokes excitation in all RGC types in the proximity of the electrode. (b) Optical stimulation only evokes excitation in the genetically-modified cells and is capable of cell-type-specific excitation.

genetic stimulation in this regard, compared to electrical stimulation. Section III describes the development details of the presented computational model. It also includes a comparative discussion on the efficacy of the electrical and optogenetic stimulations based on the model's simulation results. Section IV investigates the effect of different optical stimulation parameters and the μLED array's physical specifications through conducting various simulations and discusses their implications in optimizing the design of a retinal prosthesis in terms of biological efficacy, spatial resolution, and energy efficiency. Section V concludes the chapter.

3.1 Cell-type Specificity: A Requirement in Retinal Prostheses

As mentioned, the downstream retinal chain (from photo-receptors to RGCs) acts as a series of parallel signal processing network with more than 20 types of RGCs whose receptive fields tile the retinal surface and convey distinct visual information about the input visual stimulus [66]. In a healthy retina, the changes in the power of incident light and its associated photocurrent trigger two morphologically distinct pathways, called ON and OFF pathways, which are activated by increments or decrements of light, respectively. Fig. 3.1 shows how the parasol and midget cells (two types of RGCs) tile the surface of the RGC layer. These numerically dominant cell types are further divided into ON and OFF sub-types, which respond antagonistically to the changes in light intensity. Under electrical stimulation (Fig. 3.1(a)), all four RGC types are activated as long as they are close enough to the stimulation electrode. This results in an unnatural and even contradictory message communicated to the brain. In other words, due to the opposite situations that excite ON and OFF pathways (i.e., increment and decrement of light intensity, respectively) the signals sent to the brain following an electrical stimulation have a destructive effect on each other, resulting in creation of a vague or unstable perception in the brain [23, 24].

Selective activation of different RGCs using electrical stimulation is an open research topic. There are different approaches reported to bias the stimulation toward either ON or OFF pathways. Using smaller electrodes and pitch, manipulating the stimulus polarity, pulse duration, and frequency, and adding dedicated local return electrodes are some of the

methods investigated for enhancing the selectivity of electrical stimulation [67–69]. Despite some level of success and perception improvements, these methods require calibration over time, making them less practical when considering the post-implantation variation [68] and usually achieved at expense of increased power consumption. Additionally, with electrical stimulation, charge balancing needs to be performed for each channel with a high precision considering the large number of electrodes, resulting in extra power consumption. On the other hand, through genetic modification, it is possible to make different retina cell types sensitive to distinct wavelengths. Therefore, as shown in Fig. 3.1(b), by generating different light wavelengths (e.g., different μLED types embedded in an implantable array), selective activation of different RGC types can be achieved. It should be mentioned that currently, the specific promoter to genetically modify the different bipolar cell types has been reported [7, 70–72], which is in line with our idea of placing the implant epi-retinal and stimulate the bipolar cells to take advantage of the remaining downstream retina network. Unlike methods used for electrical stimulation, this cell-type specificity comes as an inherent advantage of optogenetic stimulation and does not require precise control of stimulation pulses or using microfabrication techniques for manipulating the electrode array's physical properties.

3.2 A Computational Model for Retinal Stimulation

To validate our hypothesis and to investigate the efficacy of optical stimulation as a potential treatment option for patients with retinal degeneration, we developed our model based on the ISETBio toolbox (Image System Engineering Toolbox-Biology) [73], which is a computa-

tional framework designed for calculating the properties of the front-end of biological visual systems and used in modeling the activation of retina layers for image formation. In doing so, the same biological properties of the retina from large-scale activity recording that has been used in ISETBio toolbox are considered with the exception that ON and OFF bipolar cells are assumed sensitive to different light wavelengths [73].

In [23], the authors have developed an ISETBio-based model to generate the visual perception resulted from activation of bipolar cell layer in response to electrical stimulation (i.e., bypassing the photo-receptor layer). In their model, the electrical stimulation parameters (i.e., spatial and temporal intensity) are defined based on the visual stimulus captured by a camera. The patient's perception estimation through linear reconstruction of retinal ganglion cells activation is also included in their model which can be used to study the adjustment of different stimulation parameters.

In this chapter, we have developed a an ISETBio-based framework that translates the captured visual stimulus into optical stimulation pattern that are used for activation of genetically-modified bipolar cells, with each pathway being transfected with a distinct opsin. By doing so, the ON and OFF pathways were made sensitive to visual stimulus irradiance increment and decrement, respectively. The patient's perception estimation is reconstructed using linear model based on the induced activities of RGCs. Beside the pathway-specific stimulation, this model also allows for defining the optical stimulation parameters (e.g., intensity, frequency, and duty cycle) and the μLED array physical specification (e.g., μLED array spatial resolution, spatial coverage, placement, and light divergence). This is used to study the

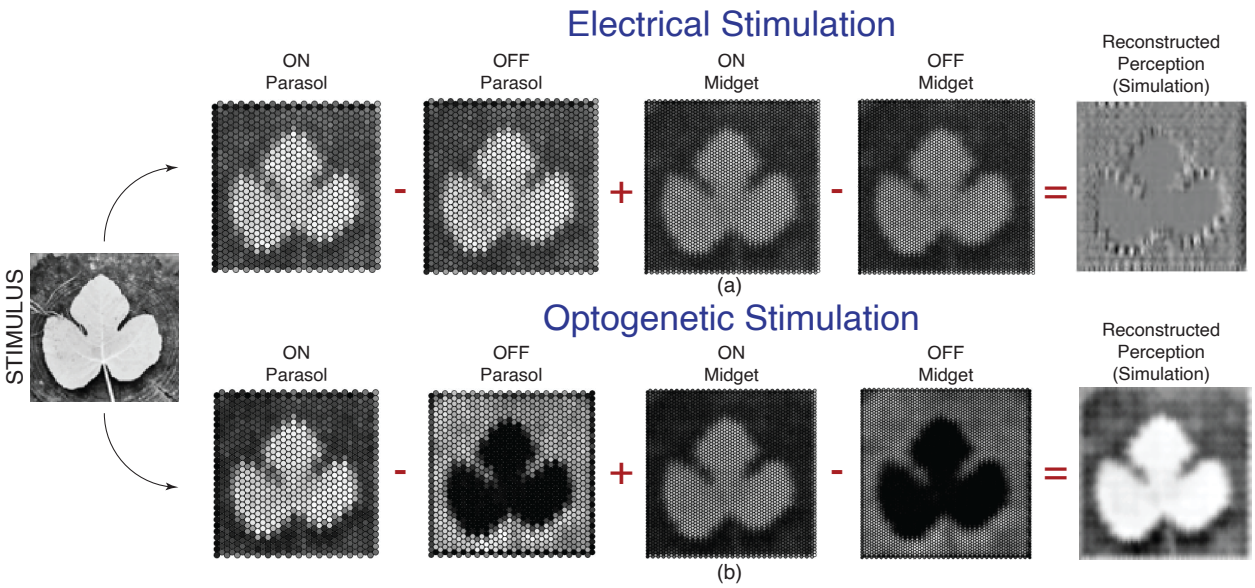


Figure 3.2: Simulation results showing the input stimulus (left), the four types of RGCs spiking activity density and their linear reconstruction for (a) electrical stimulation, and (b) optogenetic stimulation.

effect of these adjustments on the reconstructed perception and to optimize the parameters while considering the area and power restrictions of an implantable neuro-stimulator device.

3.2.1 Stimulus to Inference with Electrical Stimulation

To generate a baseline for comparison, we first regenerated the response (i.e., visual perception) to electrical stimulation based on what is developed in [23]. Fig. 3.2(a) shows the spiking activity density of different RGC types in response to electrical stimulation of bipolar cells assuming natural function for the rest of retina downstream network. As expected, due to the indiscriminate activation of electrical stimulation, the ON and OFF cells have similar spiking density, which is higher in the brighter areas of the visual stimulus and lower in the darker regions. This is contrary to what happens in a healthy retina, where these cells spiking activities peak at opposite light intensities of the input visual stimulus. Therefore, two contradicting messages are communicated by the ON and OFF pathways in response to

the electrical stimulation, which combine destructively during linear reconstruction of the image. Consequently, the estimated perception is only marginally more correlated to the visual stimulus than noise, which is the same outcome also reported in [23] and [22].

3.2.2 Stimulus to Inference with Optical Stimulation

Fig. 3.2(b) shows the spiking activity density in RGCs in response to the same visual stimulus that is translated into optical stimulation pattern for bipolar cell layer and the natural downstrean network is assumed for defining the RGCs activities. The optical stimulation employs two distinct optical wavelengths for ON and OFF pathways, which respond to increment and decrement of light intensity. In practice, this can be done by employing a dual-color μLED array, where the ON pathway cells are genetically modified by a blue-light-sensitive opsin (e.g. Chronos [75]) and the OFF pathway cells are genetically modified by a red-shifted opsin (e.g. Chrimson [75]). It should be noted that these two opsin light sensitivity wavelength lies within the visible range. However, considering that our current target is a fully degenerated retina, there is no constraint about the remaining healthy photoreceptors. Under these circumstances, the ON and OFF pathways cells exhibit distinctly different activations, same as what happens in a healthy retina. As a result, as shown in Fig. 3.2(b), the estimated perception calculated using linear reconstruction is highly resemblant to the visual stimulus.

3.3 Analysis and Optimization of Stimulation Parameters and μLED Array Specifications

In this section, the effect of different optical stimulation parameters and the μLED array's physical specifications are investigated. We have also included a discussion on each simulation result and how they could be used for parameter optimization.

3.3.1 μLED Array Spatial Resolution

The developed model is used to investigate the impact of increasing the optical stimulation's μLED array density in improving the visual perception, and also to compare that to the same μelectrode array density in electrical stimulation. This will compare the two modalities in terms of their potential for channel-count scalability as well. To do this, the same visual stimulus as in Fig. 3.2 is fed to the models for optical and electrical stimulation. Then the structural similarity of the reconstructed visual perception to the input visual stimulus is calculated and the results are normalized with respect to white noise's (i.e., when random numbers are used for each pixel's brightness with uniform distribution) structural similarity to the visual stimulus. The three terms that make up the structural similarity quality assessment index are brightness, contrast, and structure. The three terms are multiplied to create the overall index [76]. This experiment is repeated for 12 different μ-electrode/μLED array densities. As shown in Fig. 3.3 (a), our simulation results made it evident that the perception degradation due to the cell-type-non-specific activation in electrical stimulation is so severe that negatively affect factors such as reducing electrode size/pitch. In contrast, the results indicate that increasing the μLED array's density yields a considerable improvement in the

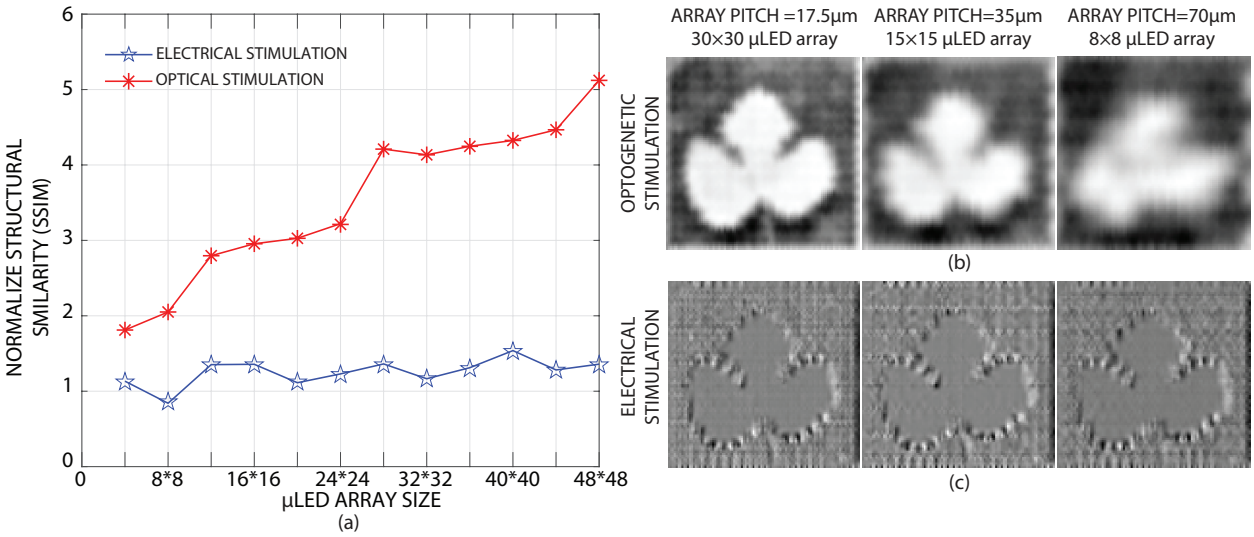


Figure 3.3: Simulation results showing (a) comparison between optical and electrical stimulation in terms of normalized structural similarity of the visual stimulus to the estimated perception as a function of the μLED/μelectrode array density and the reconstructed estimated perception as a result of (b) optogenetic stimulation and (c) electrical stimulation for 17.5μm, 35μm, and 70μm μLED/μelectrode array pitch.

visual perception. As a more visual example, Fig. 3.3 shows the linear reconstruction results for optogenetic and electrical stimulations with μLED/μelectrode array of 17.5μm, 35μm, and 70μm pitch, respectively. These results confirm the efficacy of increasing optical stimulation array density in improving the resulted visual perception in optogenetic stimulation, while no notable improvement is observed for electrical stimulation.

3.3.2 μLED Array Spatial Placement

It is ideal for a stimulator designed to restore the functionality of a degenerated retina to be implanted where the density of bipolar cells (or RGCs, depending on the stimulation target) is at the highest. This is mainly to impact more neuronal pathways to the brain (i.e., higher spatial coverage), to maximize the stimulation efficacy. In the human eye's retina, placing the

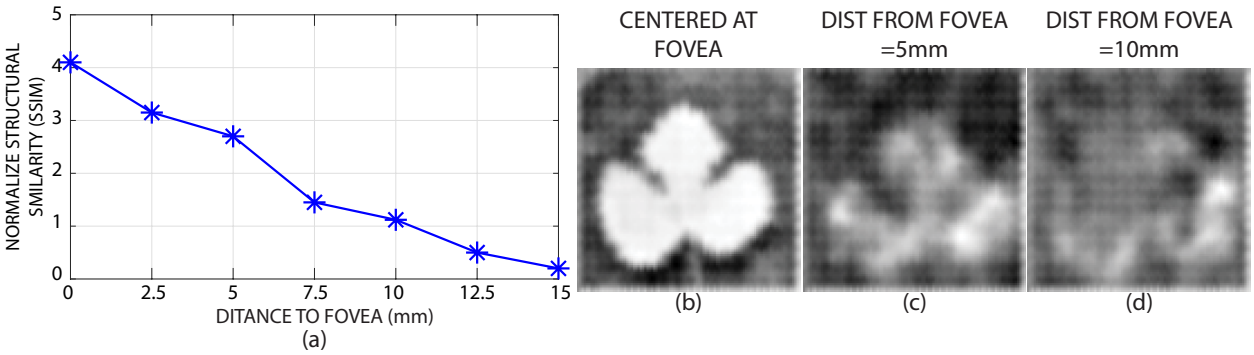


Figure 3.4: Simulation results showing (a) the normalized structural similarity of the visual stimulus to the estimated perception as a function of distance to fovea and the reconstructed estimated perception as a result of optical stimulation when the implant is (b) centered at fovea (c) at a 5mm distance from fovea, and (d) at a 10mm distance from fovea.

stimulator's μelectrode/μLED array in a zone called fovea is ideal for this exact reason [77]. However, due to their indiscriminating activation, electrical stimulation electrodes are often placed peripheral to the fovea (e.g. in low-density areas such as raphe) to avoid unwanted cell stimulation [77]. This results in the stimulation to only affect parts of the retina that have a sparse cell population, further reducing the chance of success in effective vision restoration using this approach. Thanks to their cell-type specificity, this is not the case for optical stimulators and they can directly target areas with high cell-density.

We used the developed model to quantitatively illustrate the impact of placing an optical neuro-stimulator closer to the fovea, in achieving a better visual perception. Fig. 3.4(a) shows the normalized structural similarity of the reconstructed visual perception for a visual stimulus input with respect to white noise input at different distances from fovea. The figure clearly shows the importance of implant placement in achieving a better visual perception (same array size and light intensity used for all cases). This is more visually evident in the three example pictures shown in Fig. 3.4.

3.3.3 μLED Array Light Divergence

A practical concern often cited about implantable μLEDs used for optogenetic stimulation is their inherently-divergent light beams [78]. The main problem is that the poor light directivity leads to a poor spatial resolution for the optical stimulation, hence the LED light activates the nearby cells. While optogenetics ensures that only a specific type of cell is activated by the optical stimulation, it cannot stop the neighboring cells from being activated if they are of the same type as the target cells (i.e., similarly-transfected). Fortunately, in human eye's retina, it has been shown that while the neighbor cells from different types have little to no correlation, the neighbor cells of the same type are highly correlated [74, 79, 80]. This is mainly due to the fact that these cells are usually affected by similar inputs from the previous retina layers. Therefore, the light divergence, at the microscopic level (i.e., multi-cell activation instead of a single-cell activation) can be even beneficial as it engages more relevant cells in the information transfer path.

The above discussion is not valid at the macroscopic level, meaning that cells of the same type that are far from each other are likely to be uncorrelated, hence, should not accidentally get activated together. This suggests that there is an optimal radiation field for the μLEDs that results in the highest visual perception quality. Since the radiation field depends on both the μLEDs' spacing (i.e., the pitch) and their divergence angle, to investigate this optimal point, we first introduce a parameter called the divergence factor as,

$$\text{Divergence Factor} = \frac{\text{Radiation Field Diameter}}{\mu\text{LED Array Pitch}}, \tag{3.1}$$

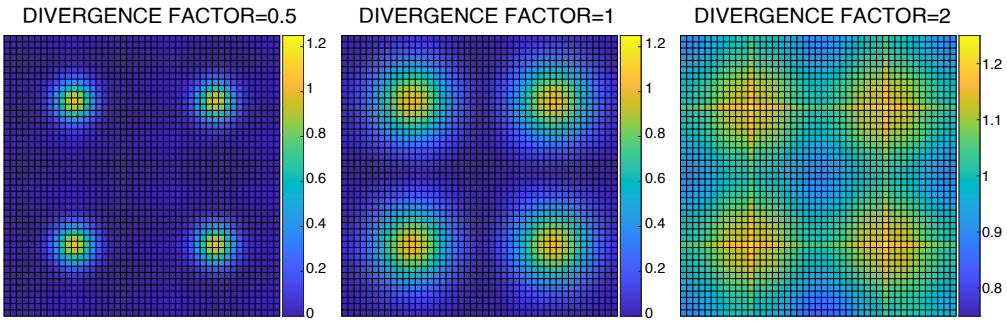


Figure 3.5: The spatial filters applied to the radiation of a 2×2 array of μLEDs to implement divergence factors of 0.5, 1, and 2, assuming a 2-D Gaussian-distributed illumination, centered at each μLED.

where the μLED's radiation field diameter is defined as the diameter of the surface area at the targeted distance from the optical source that all the transfected neurons inside it are stimulated by the optical illumination of that μLED; and the μLED array pitch is defined as the physical distance between the centers of the two neighboring μLEDs in an array. In order to model the effect of light divergence, a spatial filter is applied to the optical stimulation pattern. Fig. 3.5 shows the spatial filters used to model different divergence factors, in which a Gaussian distribution is assumed for the μLEDs radiation. For a divergence factor <1 (e.g., 0.5, as shown in Fig. 3.5), the radiation fields do not extend sufficiently to cover the entire distance between neighboring μLEDs, which results in blind spots (i.e., not stimulated areas). On the other hand, for a divergence factor >1 (e.g., 2, as shown in Fig. 3.5), the radiation patterns from neighboring μLEDs interfere, resulting in some of the cells being affected by more than one μLED.

Fig. 3.6(a) shows the reconstructed visual perception for three different divergence factors, at two different μLED array pitches. The figure confirms that the best result is for divergence factor of 1, while higher divergence factors result in blurring the perception (due

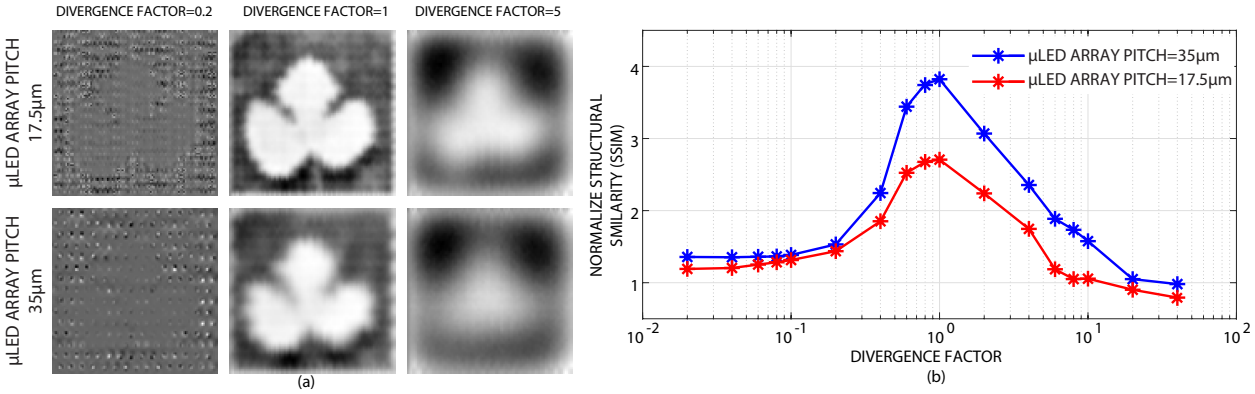


Figure 3.6: (a) Simulation results showing the reconstructed estimated perception for divergence factors of 0.2, 1, and 5 for μLED array pitches of 17.5μm and 35μm. (b) Simulation results showing the normalized structural similarity of the visual stimulus to the estimated perception as a function of divergence factor for μLED array pitches of 17.5μm and 35μm.

to sending wrong messages to the brain by accidentally co-activating opsins that are far from each other), and lower values result in a fewer number of cells being activated, hence a proportional drop in the reconstructed perception's quality. Fig. 3.6(b) shows the normalized structural similarity of a particular visual stimulus (same as the one in Fig. 5) with respect to a white-noise visual stimulus, at different divergence factors and for two different μLED array pitches. The results confirm that regardless of the μLED array pitch, the highest structural similarity is achieved for a unity divergence factor. These simulation results imply that for a specific μLED that has a known radiation field, the optimum pitch for the μLED array would be equal to the diameter of the radiation field.

3.3.4 Optical Stimulation Light Intensity

The instantaneous power required for driving the stimulating μLEDs is orders of magnitude larger than all other blocks on an inductively-powered implantable retinal prosthesis (e.g., data receiver, power management, signal processing, etc.). Therefore, it is of critical

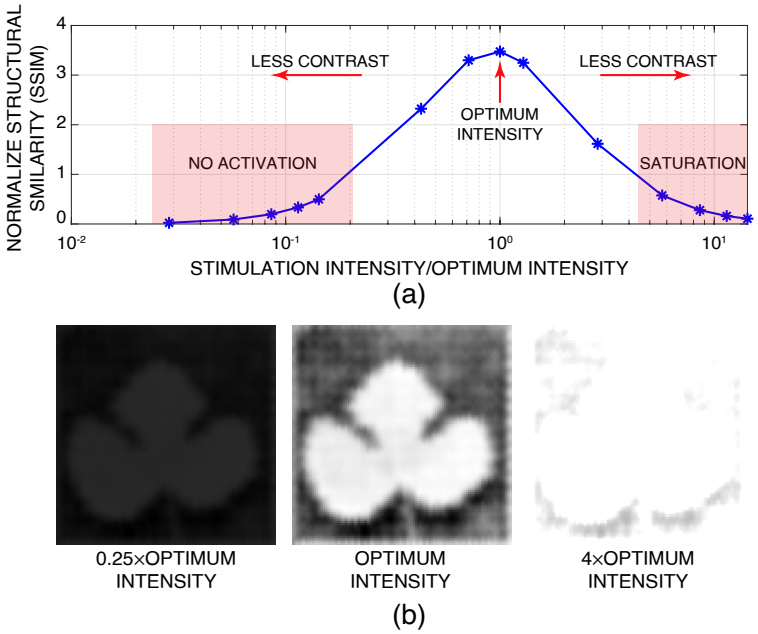


Figure 3.7: (a) Simulation results showing the normalized structural similarity of the visual stimulus to the estimated perception as a function of stimulation intensity divided by the optimum intensity. (b) Simulation results showing the reconstructed estimated perception for three different light intensities.

importance for the stimulation light intensity to be optimized to avoid a power dissipation more than the minimum required (to maintain a high energy efficiency for the device) while ensuring that all targeted opsins receive sufficient power needed for activation.

We varied optical stimulation intensity over a range of more than two orders of magnitude and compared the resulted visual perception with the stimulus input. As shown in Fig. 3.7(a), the perception's quality clearly peaks at an optimal intensity and diminishes for higher and lower values. While the quality degradation was expected for lower light intensities (due to inactivation of some of the opsins), we observed that a higher-than-optimal intensity could also result in diminishing the visual perception's quality, in addition to making the device less energy efficient. Fig. 3.7(b) illustrate the effect of non-optimal and optimal intensities with three visual examples. Our interpretation is that for the lower intensities, the visual

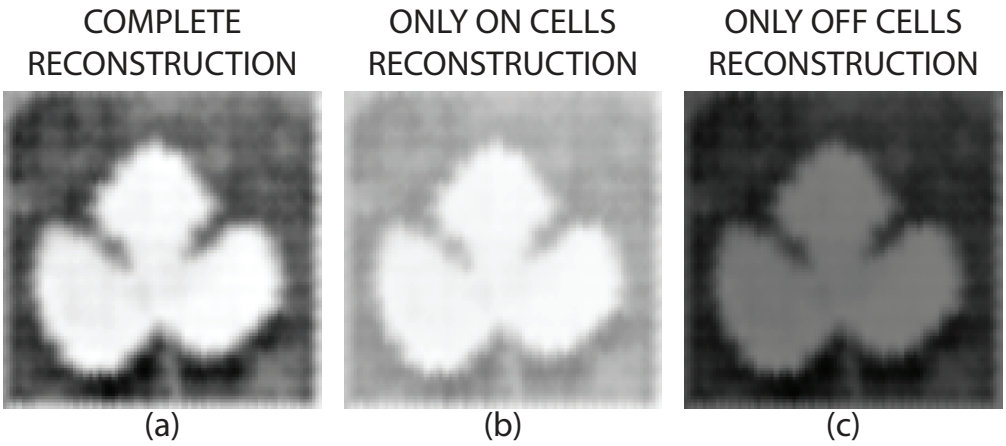


Figure 3.8: Simulation results showing the reconstructed estimated perception as a result of optical stimulation when (a) both ON and OFF pathways are activated, (b) only ON pathway is activated, and (c) only OFF pathway is activated.

perception has a lower contrast due to some opsins not receiving sufficient optical power to get activated. For the higher intensities, we assume that the quality degradation is due to saturation of the genetically-modified cells because of receiving too much power [81].

The above results indicate the importance of real-time neural activity recording for implantable optogenetic stimulators as also mentioned in [62]. Besides enabling closed-loop operation that could be used for diagnostic applications, real-time recording of neuronal activities (typically, electro-physiological recording) allows for calibrating the stimulation light intensity to avoid the above-mentioned problems at both sides of the spectrum.

3.3.5 Single-Wavelength Optical Stimulation

While implantable multi-wavelength (i.e., multi-color) μLED arrays have been reported in literature [82, 83], custom micro-fabrication of a single-wavelength array poses far less complexity [82, 83]. Additionally, from an energy efficiency perspective, if we can conduct optical stimulation by turning on only one μLED color, it could significantly (almost by 50%) reduce the required power consumption for stimulation.

To investigate this, we adjusted our developed model to only consider genetic modification for one of the pathways (i.e., ON or OFF). Fig. 3.8 shows the reconstructed perception due to optical stimulation of only ON or only OFF pathways. The results show that the above-mentioned fabrication and energy efficiency benefits come at the cost of quality degradation compared to a two-wavelength optical stimulation, but still a significantly better quality than the electrical stimulation. While in electrical stimulation the messages from two pathways had a destructive effect on each other, here only one pathway gets activated, hence, the message sent to the brain is still correlated with the stimulus input. Intuitively speaking, by using only one wavelength for optical stimulation, the optogenetic-based activation gain is reduced approximately by half, while the destructive effect of non-discriminate electrical stimulation is still avoided.

Chapter 4

Poisson-coded Temporally-Distributed Spiking Pattern Generation for Optimized Optical Stimulation

In this chapter we focus on circuit and system-level implementation challenges of our proposed technique of poison-coded temporally-distributed spiking pattern generation for our target optimal optogenetic stimulation. Using this technique, we achieve the main design criteria for an implantable retinal prosthesis which are low power consumption, compact size, precise electrode selection, versatile stimulation pattern capabilities, patient safety, robust power and data transmission [85, 86]. Additionally, to achieve a high-density retinal prosthesis to accommodate high visual acuity and high spatial resolution, the aforementioned requirement should be achieved while ensuring scalability.

In the domain of retinal prostheses, a fundamental differentiation is found in the approach to visual information acquisition, divided into external image sensor-based systems and those integrating internal photo-diode-based sensors. Each method brings distinct attributes and challenges, shaping their applicability and impact on patient's perception of the applied stimulation.

Systems utilizing external image sensors involve cameras mounted on external devices like glasses, capturing the visual scene. These images are then processed by an external unit, enabling the application of advanced image processing techniques. This capability allows for significant enhancement in the generated stimulation pattern using the suitable processing technique that can best mimics bypassed layers of the retina. The following elucidates key features inherent in a healthy retina [87], which can be integrated into the image processing algorithms within the external unit:

- Luminance Contrast: Some RGCs signal local differences in brightness, known as luminance contrast. Enhancing image quality by emphasizing these contrast differences could stimulate these RGCs more effectively.

- Motion Processing: RGCs in the retina process motion, distinguishing between movements of the observer (self-motion) and movements of objects in the observed world (object motion). Enhancing the clarity and distinction of motion in images can help stimulate these motion-processing circuits in the retina more distinctly.

- Object Motion Sensitivity: Some RGCs, known as object motion-sensitive (OMS) RGCs, are specifically attuned to detecting moving objects. They differentiate local

motion in a scene from global image motion caused by head and eye movements. Improving the distinctiveness of moving objects against static backgrounds in images can stimulate these RGCs more effectively.

- Looming Detection: This involves the detection of objects on a collision course with the observer, a critical feature for evoking defensive responses. Enhancing the representation of rapidly expanding objects (looming) in images can effectively stimulate these specific RGCs.

- Orientation Selectivity: RGCs also show preferences for the orientation of static or moving stimuli. Enhancing the orientation features in images, such as emphasizing horizontal or vertical lines, can stimulate these orientation-selective RGCs.

Additionally, the external setup also offers the benefit of substitution and upgrade whenever necessary without requiring to perform an invasive surgery. However, this method requires to wirelessly transmit the generated stimulation pattern to the implantable unit which requires to budget the available power consumption of the implant accordingly.

On the other hand, internal photo-diode-based retinal prostheses integrate photo-diodes directly into the implant. These diodes capture light entering the eye and directly convert it into electrical signals for retinal stimulation, through simple current amplification. The primary advantages of this system lie in its simplicity and autonomy, removing the necessity for external devices. The direct conversion from light to electrical stimulation offers a more immediate and responsive visual experience. Despite these benefits, the reliance on ambient light levels negatively affects the reliability of the visual experience. Additionally, the lack

of external processing restricts the ability to enhance or modify visual inputs, potentially resulting in a less refined visual perception. It should also be noted that the advantages associated with this method are achieved if the photo-diodes and the electrical circuitry required for stimulation are fabricated on the same substrate which means that it can not be done using standard CMOS process as the photo-diodes require special material which are not present in standard CMOS process.

In conclusion, while both external image sensor-based and internal photo-diode-based retinal prostheses present viable pathways for vision restoration, for our retinal prosthesis design, the choice of an external image sensor-based system is preferable.

As mentioned earlier, our target is to place the μLED array epiretinally (i.e., in contact with e RGC layer), and optically stimulate the Bipolar cells. Using this structure, we can employ the remaining network between the bipolar layer and RGC layer and achieve a more natural signal transmission to the brain, while having a less invasive surgery procedure in comparison with subretinal placement of the μLED array.

In our proposed system, we have a two-part implant where one part is fabricated as a flexible PCB which is placed on the sclera in a biocompatible casing and host the circuitry for power management (i.e., power RX coil, 3.3V and 1.8V regulators) and data telemetry (i.e., data RX coil, clock recovery and ASK data decoder). Therefore, 5 wires (i.e., GND, 1.2V supply, 3.3V supply, recovered clock and received data) are inserted inside the sclera travelling to the other part of the system that is placed in contact with the RGC layer which hosts the μLED array and our fabricated IC. Having a fixed number of wires that

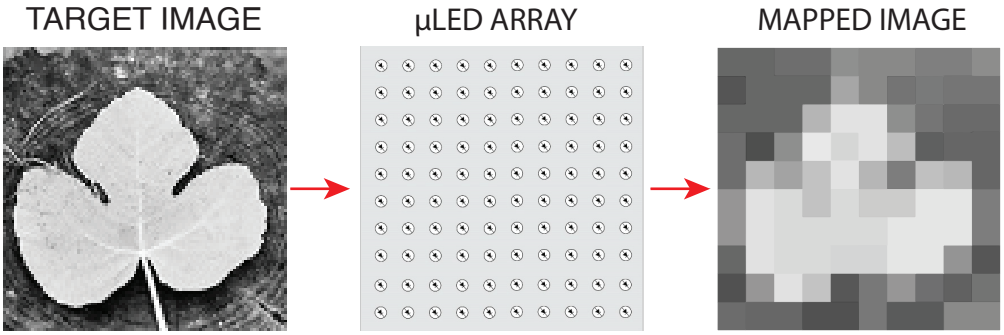


Figure 4.1: Mapping the target visual stimulus on the implanted μLED array.

are inserted inside the sclera significantly improves the scalability of the device. Although, this technique can be applied to systems that use electrical stimulation as well, our special advantage lies in the number of wires that are required to control the μLED array to evoke the target stimulation pattern.

As illustrated in Figure 4.1, the objective is to map the target visual stimuli onto the μLED array. This projection can be performed after applying necessary image processing techniques, including edge enhancement and contrast optimization. Consequently, to facilitate this process, it is required that each μLED within the array is capable of being individually addressed and controlled.

For a passive μelectrode array to be individually addressable, each electrode should be connected to a peripheral wire. Therefore, the number of required wires to control a μelectrode array scales with the size of the array (i.e., M×N), which is a scalability constraint preventing from increasing the electrode count to hundreds/thousands. For a μLED arrays, the individual addressing can occur with two different approaches. Performing simultaneous stimulation requires M×N wires which has the same scalability issue [88]. However, by leveraging the concept of photon integration (which has been comprehensively detailed from biological aspect in Chapter 2) and the low update rate for human vision (24 frame/s)

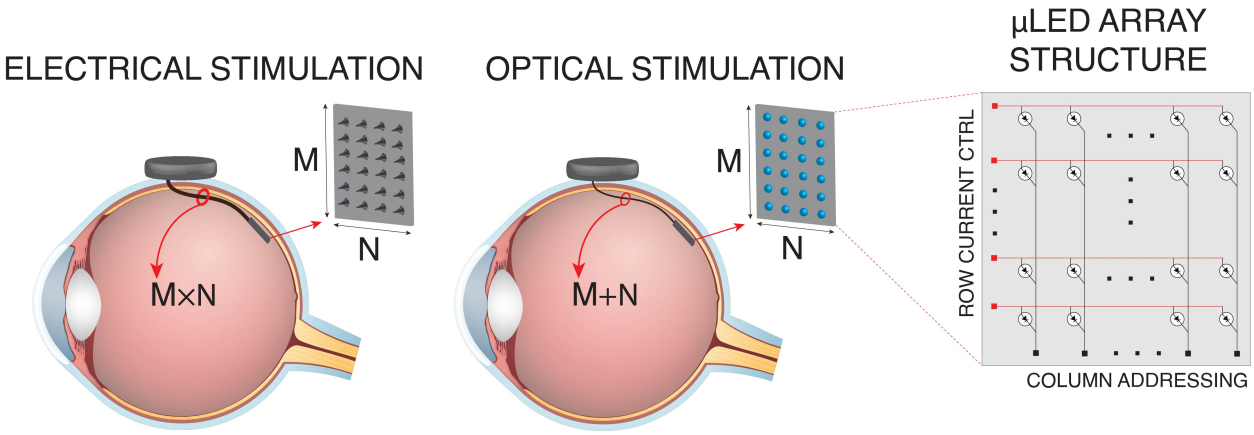


Figure 4.2: Required number of wires for a retinal prosthesis with electrical stimulation or optical stimulation using row and column addressing structure.

we can perform sequential addressing to control the μLEDs.

Considering that unlike electrodes, which only have one terminal, each LED has two terminals prompts the possibility of an alternative routing methods which connects the anodes of all the electrodes in each row and the cathodes of all the electrodes in each column. As shown in Fig. 4.2, using this structure the number of required wires to control the μLED arrays is the number of rows plus the number of columns (i.e., M+N), which is a significant reduction in comparison with the number of wires required in electrical stimulation (i.e., M×N) which facilitates scalability.

Performing the sequential μLED control can occur by stimulation each single LED and scanning the whole image. However, in this situation, there would be very limited time to deliver the whole power for each LED, which not only result in over designing the power transfer link, but also may harm the tissue. To address this, the generation of stimulation patterns is performed in a raster scanning (row-by-row) methodology (intuitively speaking, similar to the mechanism of a television display). In this context, rows are activated in a successive manner, with the selection of specific LEDs within each row achieved by connect-

ing their cathodes to GND. Furthermore, the wires interfacing with the rows and columns serve as path for the stimulation current, which originates from the active circuitry of the implant.

While the raster scanning technique resolves the problem that the LEDs cannot be controlled simultaneously, remains the issue that to generate a visual scene, it is required to drive each LED with a different optical intensity. However, in the raster scanning technique all the LEDs that are turned on in one row would have the same stimulation current and therefore have the same intensity. Thus, the challenge here is how to distribute the total row current through different LEDs of the same row in an uneven fashion, such that each of them gets the exact current that is intended for that LED. To address this challenge, we propose to temporally modulate the stimulation pattern instead of controlling the stimulation pulse amplitude. This temporal coding approach also leverages the photon integration feature of opsins.

4.1 Optimal Stimulation Pattern Generation

This section is dedicated to the generation of an optimal stimulation pattern for the control of the μLED array, tailored specifically to accommodate our proposed raster scanning scheme while addressing the challenge of delivering uneven stimulation current to the different LEDs in the same row to replicate the target visual stimulus.

Conventionally, the input visual stimuli are coded by a specific stimulation intensity per pixel which translates to the amplitude of the stimulation pulse (e.g., the sample simple image in Fig. 4.3). By leveraging the inherent opsins' photon-integration feature, we propose

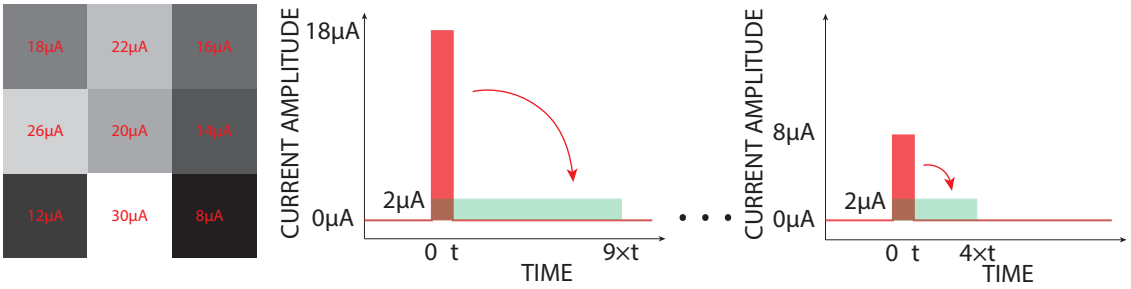


Figure 4.3: An example of target intensity in a stimulation pattern for a specific visual stimuli and temporal coding versus amplitude coding for two different pixels with various stimulation intensity.

to differentiate the stimulation intensity among different pixels by the duration of stimulation while the magnitude of stimulation current is the same for all the pixels. Using this scheme, at any moment in time, all the LEDs in one row would be turned on with the same current. Therefore, it would resolve the challenge of even distribution of the injected current to the ON LEDs in each row.

Fig. 4.3 shows an illustration of temporal coding versus amplitude coding for two different pixels with various stimulation intensity. It should be noted that this temporal coding scheme and the row-by-row addressing scheme are both possible since the retina networks functions at a low frequency (i.e., 24 frame per second). It should be noted that this technique not only significantly reduces the number of necessary interconnections (from N^2 to 2N for an N×N array), but also greatly eases the requirement for wireless data communication throughput. This point is further described when we are talking about the circuit-level implementation of the stimulator circuit.

4.1.1 Power Consumption Distribution

A significant barrier to scale implantable electrical/optical stimulators' channel count, is the instantaneous power demand during multi-channel stimulation, which typically far exceeds

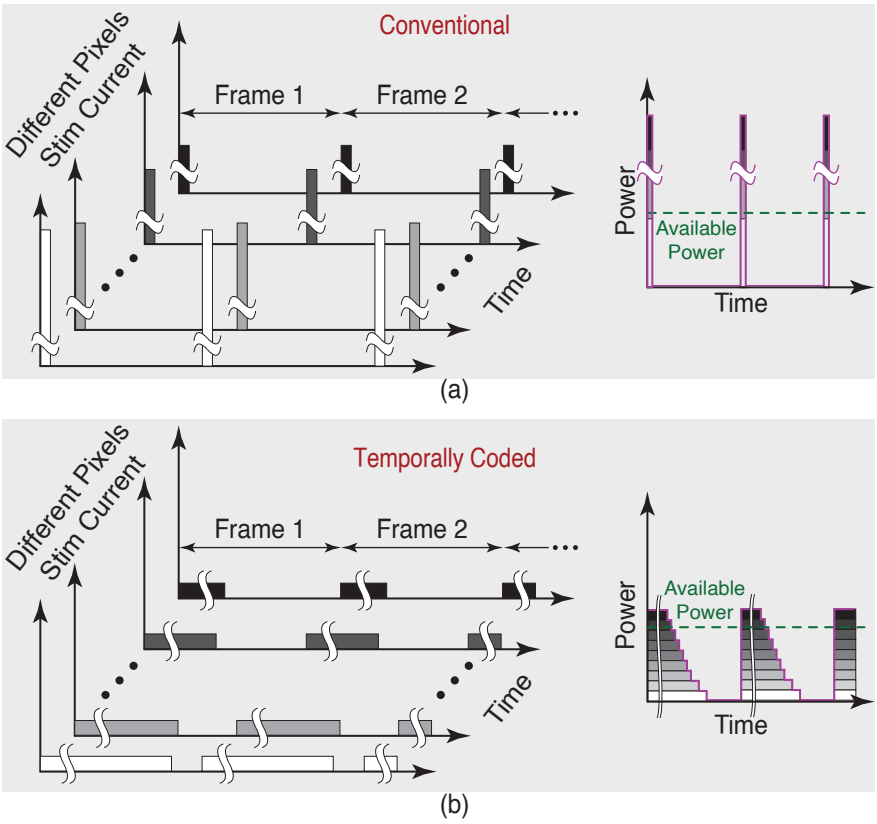


Figure 4.4: Power consumption distribution in (a) amplitude coding vs (b) temporal coding.

the safe power transmission capacity of the inductive link, thereby substantially restricting scaling. However, the introduction of our proposed technique, which employs temporally-coded stimulation pattern generation, addresses this issue. It achieves a reduction in instantaneous power consumption and effectively distributes the power load across the time span of each frame, thereby substantially improving the device's scalability.

As illustrated in Fig. 4.4(a), in the conventional amplitude coding technique where all the pixels (μelectrodes) are activated simultaneously, there would be a large instantaneous power consumption associated with each frame. However, as shown in Fig. 4.4(b) using the temporally-coded scheme will distributes the power load over the span of each frame. The instantaneous power distribution in this situation have an initial surge at the commencement

of each frame, which subsequently diminishes over the frame's duration. To clarify, this imbalance in the instantaneous power consumption because the LEDs that are associated with the pixels with lower intensity would be turned on for a shorter duration compared to those needing higher intensity. It should be noted that while the cumulative power consumption remains consistent across both coding methodologies, temporal coding uniquely offers the advantage of dispersing power usage and mitigating the high instantaneous power demands inherent in amplitude coding.

4.1.2 Random Distribution of Optical Stimulation Pattern

In the implementation of temporally-coded stimulation patterns, the simultaneous activation of all pixels leads to an imbalanced power consumption distribution. Although this approach does reduce the instantaneous power consumption relative to the traditional amplitude coding scheme, the distribution of power remains uneven. Consequently, it necessitates that the power delivery system be designed with the capacity to accommodate the peak power requirements encountered during each frame's duration. Therefore, this would result in an overdesign for the remaining duration of the frame, as the power requirements are significantly lower outside of these peak periods.

To resolve this unevenness and ensure a more uniform distribution of power consumption, our methodology entails the random distribution of the temporally coded pulses, with a current denoted as $I_{UNIT-STIM}$ and a duration of $\Delta T_{UNIT-STIM}$. This duration is calculated based on the number of rows in the μLED array and the desired resolution. This, in turn, correlates to the quantity of frames into which each visual stimulus (or input im-

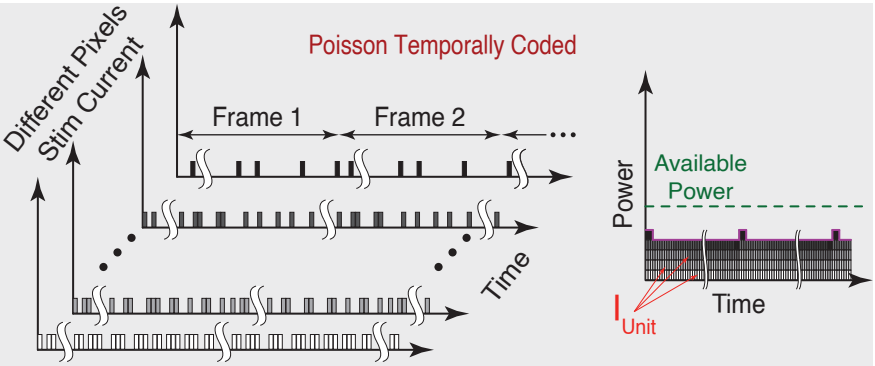


Figure 4.5: Power consumption distribution in random distribution of stimulation pattern.

age) is translated to. Consequently, it is essential to find the precise $I_{UNIT-STIM}$ tailored to the targeted tissue. This determination involves monitoring the electrophysiological activity of RGCs while optically stimulating the bipolar cells, with the objective of identifying the minimal level of optical stimulation necessary to evoke spiking activity in the RGC layer. The required current value for stimulation is highly dependent on the distribution of opsins within the target tissue, which in turn influences the tissue's sensitivity to optical stimulation. Thus, the identification of this $I_{UNIT-STIM}$ is a critical component of this project, and is the focal point of the next chapter.

As depicted in Fig. 4.5, this coding strategy ensures an even distribution of power consumption across the entire frame. Such an approach not only alleviates the burden on power delivery and storage mechanisms but also plays a crucial role in mitigating tissue overheating. By distributing the generated heat more evenly over time, this method allows the intraocular fluids more opportunity to regulate temperature, thereby enhancing the overall safety and efficiency of the system. This balanced distribution is necessary in maintaining the integrity and longevity of the implant, ensuring consistent performance and reducing the risk of damage to sensitive ocular tissues.

Furthermore, the adoption of this technique would enhance patients' perception of the visual stimulus. This method, not only disperses the stimulation temporally but also ensures its random distribution across the area of the μLED array. To clarify, in the absence of random distribution, the commencement of each frame would typically see a simultaneous activation of adjacent LEDs, thereby escalating interference. In contrast, the proposed randomization effectively spreads the stimulation across the expanse of the μLED array. This strategic dispersion results in a significant reduction of interference, thereby yielding improved spatial resolution and a more accurate visual representation.

4.2 Circuit-Level Implementation of the Optical Stimulator

In this section, we discuss the circuit-level implementation of the stimulator circuit, specifically designed to facilitate the application of the proposed Poisson-coded temporally-distributed stimulation pattern. As shown in Fig. 4.6 central to this circuit is the calibration phase, where the current output of a 5-bit DAC (I_{CALIB}) is adjusted to the lowest value sufficient to induce spiking activities in the associated RGC layers. The calibration process commences with the first LED in the first row, incrementally raising the I_{CALIB} DAC current until it evokes spiking activity in the RGC layer. Subsequently, the procedure advances to the next LED, initiating with the current level determined for the preceding LED. This iterative process extends to the last LED in the row, resulting in the storage of a 5-bit current code specific to that row, to be utilized during its stimulation.

However, this scheme encounters a challenge: if a patch of bipolar cells is either un-

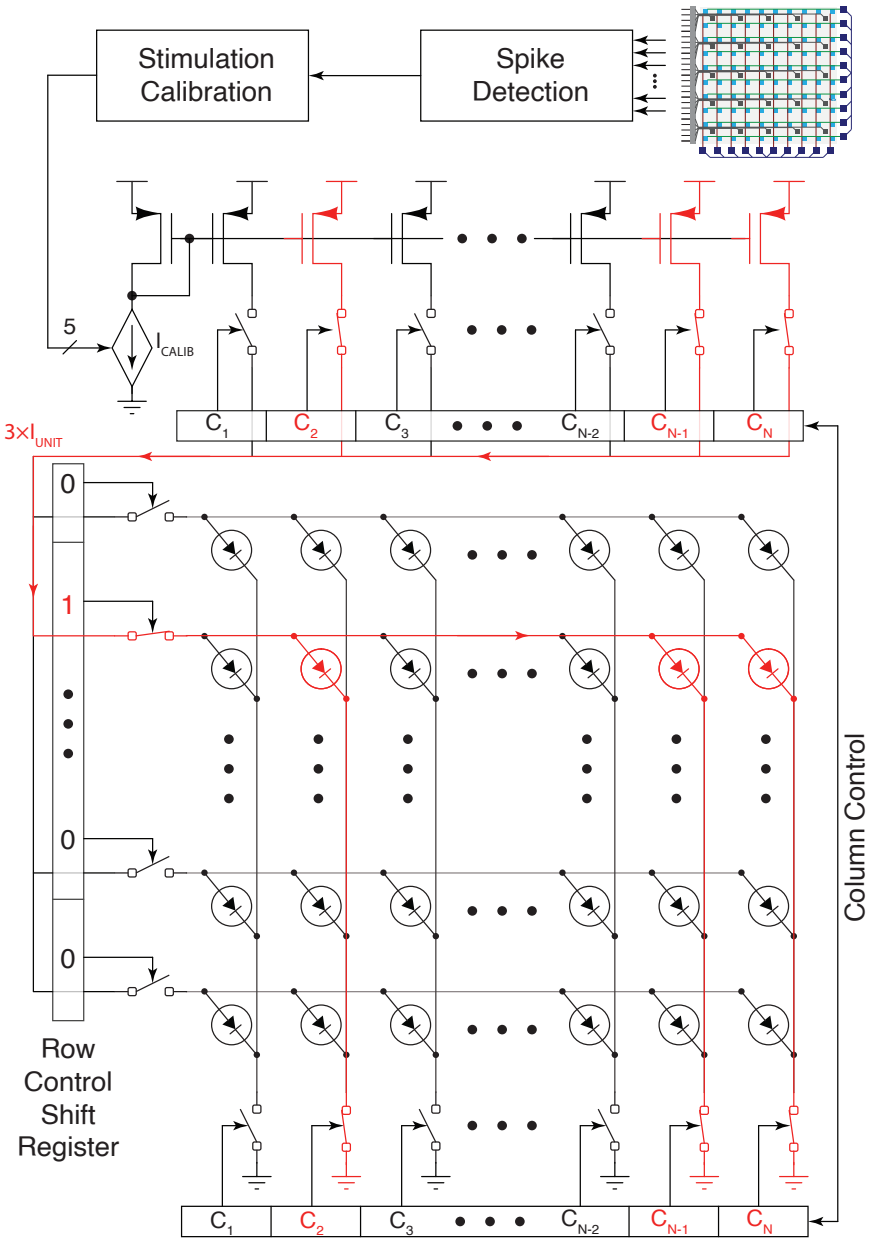


Figure 4.6: Detailed implementation of the raster-scan-based stimulation circuit controlled by the calibration loop.

modified or exhibits low opsin concentration, it could unnecessarily increase the stimulation current for that row. To mitigate this issue, the newly computed code is compared with the last stored value. If the difference falls within an acceptable range, the stored code will be updated; otherwise, the affected cell patch is disregarded. This approach is then replicated in subsequent rows. Given that all LEDs in a row are stimulated with uniform current in optical stimulation, it is only required to store a unique 5-bit code for each row, which is subsequently supplied to the I_{CALIB} DAC during its respective calibration phase.

During the stimulation phase, row selection occurs sequentially, managed by a circular shift register controlling the row selection switches. The column selection control code constitutes the only required data requisite for transmission to the stimulation unit, significantly simplifying the demands on wireless data communication throughput. For example, in a 10×10 array scenario, each image frame with 8-bit resolution requires the wireless transmission of only 256×100 bits to control the columns, amounting to a data rate of 614.4kbps for a 24-frame per second vision system.

Additionally, the current injected into each row must be equivalent to the number of activated LEDs in that row times the respective $I_{Stim-Unit}$. To generate this requisite current, a current mirror array, with the same number of branches as the column count, is employed. These mirrors are controlled by the same digital code that controls the column addressing shift register, ensuring the correct amount of current flowing through the each row.

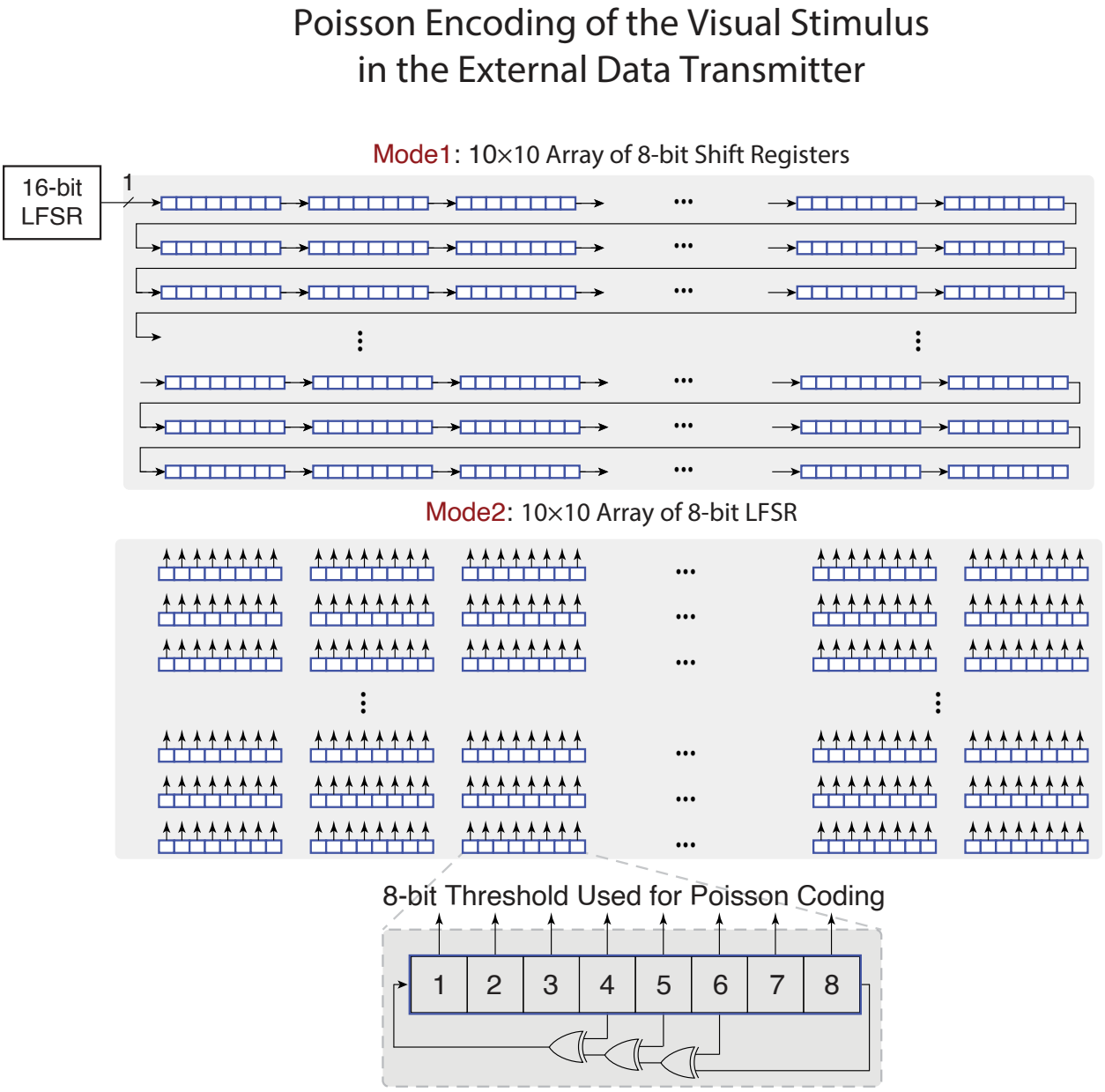


Figure 4.7: The proposed architecture for implementing the array of 10×10 pixels that randomly shuffles through all possible numbers between 1 and the number of frames and the structure of an 8-bit maximum-length LFSR.

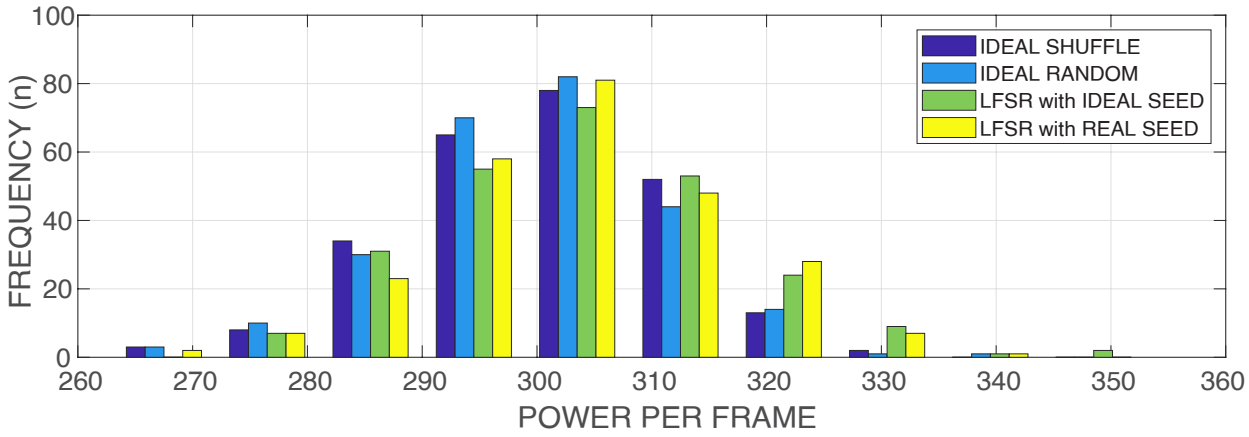


Figure 4.8: The histogram plot for power per frame using different implementations.

4.3 Hardware Implementation of Poisson-Coded Temporal Distribution

The challenge here is the hardware implementation of the random distribution. Assuming that the intensity of each pixel is coded with 8 bits, for each pixel with intensity of k, we need to randomly choose k numbers between 1 and 255. These numbers define the specific frames in the total sequence of 255 frames that should turn on that specific pixel.

The algorithm to implement this random distribution for each pixel is to shuffle the numbers between 1 and 255 and then compare the value of the pixel intensity with each number. In the frame sequence for that specific pixel we have 1 if the number of intensity code is higher that the shuffled value and 0 if it is lower. This structure should be repeated for each pixel and the numbers should be shuffled individually for each pixel. Implementing this exact algorithm in hardware is very complicated and not practical as it is required to communicate the stimulation pattern with minimum delay.

The hardware algorithm to implement this system uses pseudo-random number generator (PRNG) in specific linear-feedback-shift registers (LFSR). As shown in Fig. 4.7, an LFSR

is a shift register that, when clocked, advances the signal through the register from one bit to the next most-significant bit. Some of the outputs are combined in exclusive-OR configuration to form a feedback mechanism. A linear feedback shift register can be formed by performing exclusive-OR on the outputs of two or more of the flip-flops together and feeding those outputs back into the input of one of the flip-flops, to generate a highly randomized pattern. When the outputs of the flip-flops are loaded with a seed value (anything except all 0s, which would cause the LFSR to produce all 0 patterns) and when the LFSR is clocked, it will generate a pseudo-random pattern of 1s and 0s. It should be noted that, the only signal necessary to generate the test patterns is the clock and at each clock only 1 random bit is generated. Therefore, to generate an N bit random number, we need N separate LFSR. As shown in Fig. 4.7, our proposed solution is to assign an 8-bit LFSR to each pixel and implement the feedback to operate as a maximum length LFSR. This would act as shuffling the numbers if the initial seeds of each LFSR is chosen randomly.

Additionally, we need a 16-bit LFSR to generate a 1-bit random number, which serves the purpose of filling all the LFSRs, which are initially connected in a series configuration. Following this, the LFSRs are randomly assigned an initial seed value, and the series connection is then removed. The feedback loop in each LFSR is subsequently closed, thereby generating the desired pseudo-random shuffling. To assess the efficacy of our proposed hardware implementation method, we compared the histogram plot of power distribution with other scenarios.

To evaluate the proposed hardware implementation, we compared the histogram plot of

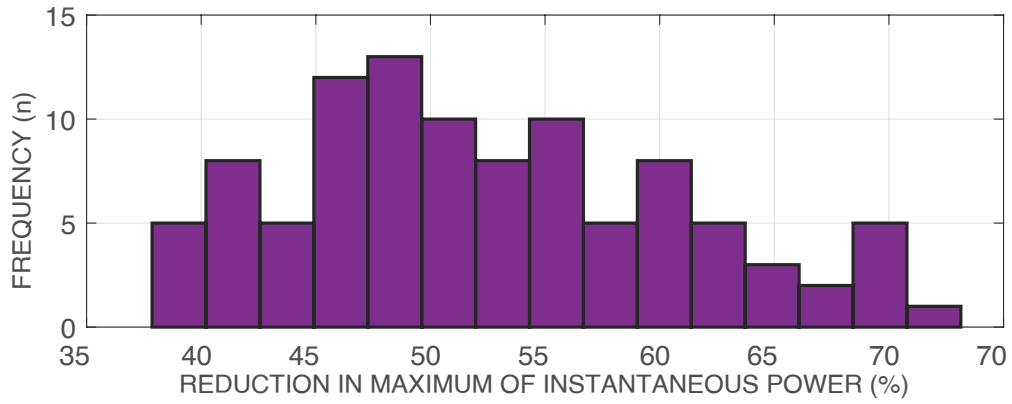


Figure 4.9: The Histogram plot for percentage of maximum instantaneous power reduction in 100 natural images.

power distribution with other scenarios, including ideal shuffling (software-based), ideal random number generation (one random number for the threshold of each pixel per frame), using the LFSR for random shuffling with random initial seed value, and our proposed method, which uses LFSR for random shuffling and setting the initial value using the output of 16-bit LFSR. As shown in Fig. 4.8, our proposed method showed a similar power distribution as the other methods, indicating that it can generate thresholds with good levels of randomness and uniform distribution.

To evaluate the impact of temporal coding with random distribution, we conducted experiments using natural images from the CG-1050 dataset [98]. As shown in Fig. 4.9, our results demonstrate an average reduction of 52% in the maximum power consumption. However, it should be noted that this reduction varies depending on the specific image, with values ranging from 38% to 73% for the images in our dataset.

To better illustrate the impact of the proposed temporal coding techniques on instantaneous power reduction for various images, in Fig. 4.10 we are presenting plots of power per frame for two scenarios: (1) using only temporal coding and (2) using temporal coding with

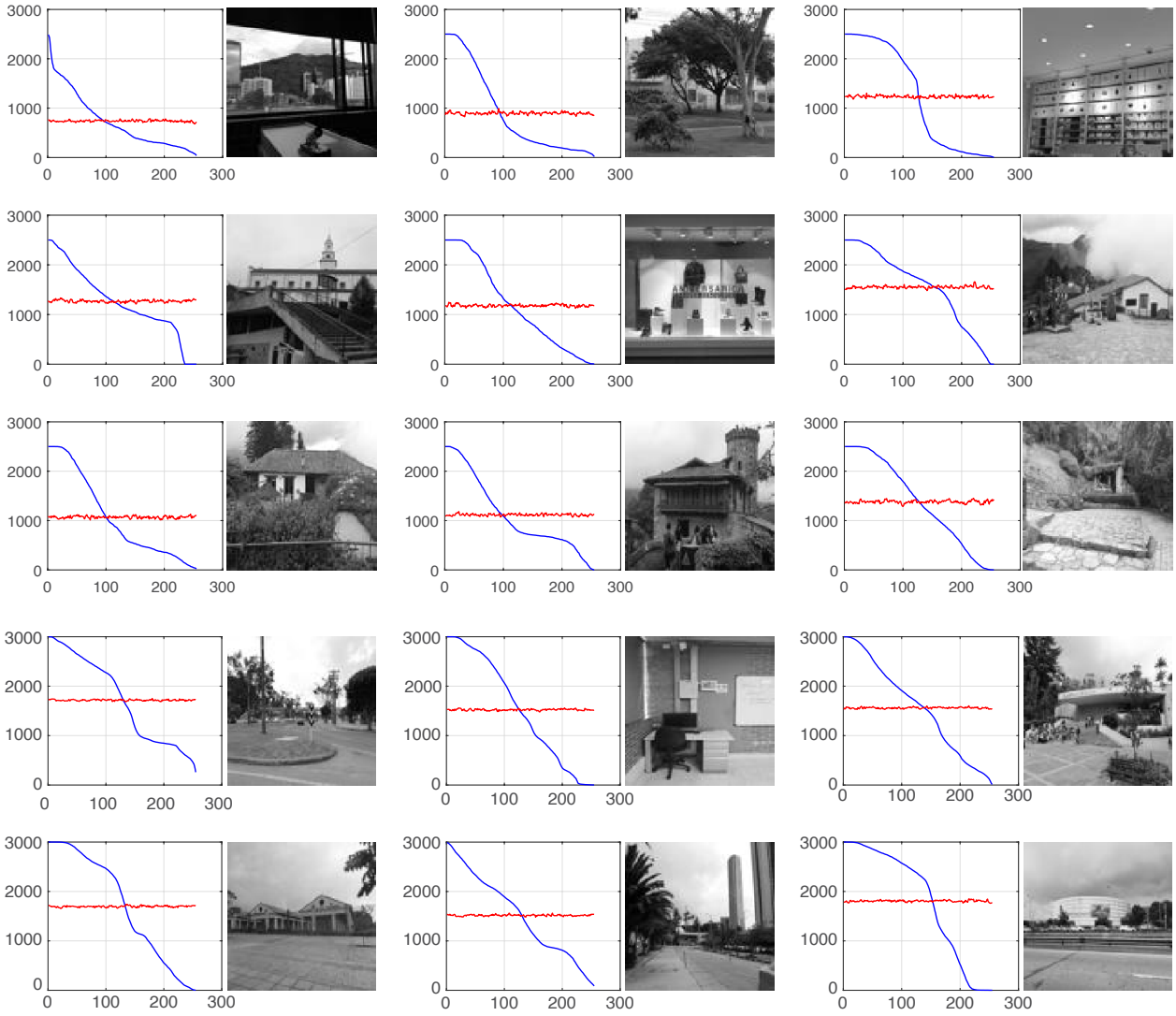


Figure 4.10: The instantaneous power consumption per frame for 15 different visual stimuli.

random distribution. The input visual stimulus for each plot is also shown, demonstrating how the power plots vary for different input images.

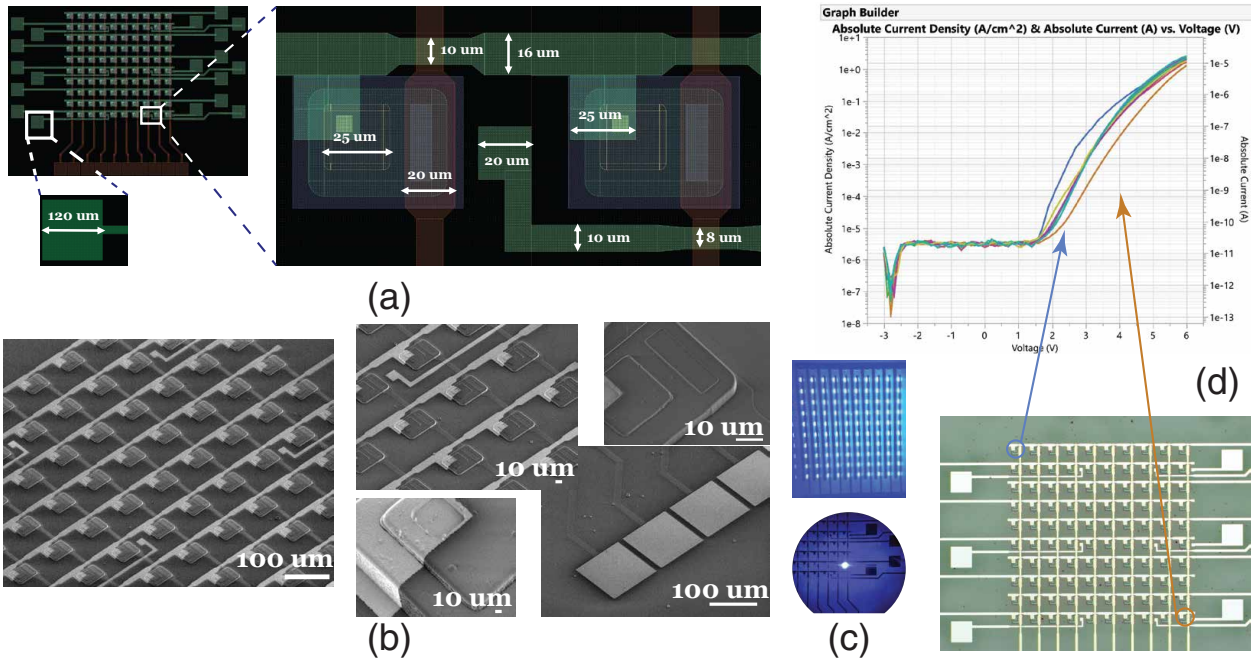


Figure 4.11: (a) Annotated layout and (b) SEM images of different zoom levels for the in-house custom-designed and fabricated μLED-μelectrode (Optrode) array. (c) example of turned ON (all and one) LEDs. (d) Measured I-V plots for different LEDs across the array.

4.4 A Row and Column Addressable μLED Array

In this section we focus on the custom-design μLED array developed in collaboration with the University of Waterloo. This array is specifically designed to be compatible with our proposed row and column addressing scheme, ensuring precise control on activation of individual LEDs. This is a critical part of the design of the μLED as it required to prevent any leakage of the current to the neighboring LEDs. As shown in Fig. 4.11, in this prototype, we have a 10×10 array of μLEDs (LED area: $40\times50\mu m^2$, pitch: 105μm), with shared $20\times20\mu m^2$ electrical recording pads which is used for recording the spiking activity of the RGC layer that is used for calibration purposes. Placement of the μelectrode array within the μLED array should be such that we have a close μelectrode to each μLED so it is possible to record the activation of

the associated RGC cells evoked from the optogenetic activation of specific patch of bipolar cells. For the current prototype, the μLED array and the fabricated IC are envisioned to be placed on opposite sides of a PCB substrate. Connections between the μLEDs and the IC will be established through wire bonding, linking the LED array's row and column control signals and the μelectrodes to the PCB. The same signals should be wirebonded to the associated pads on the fabricated IC to control the μLED array.

An alternative integration method involves directly placing the LED array on the IC surface and establish the connections by perform diebonding by placing the required bonding pads on the surface of the IC (Intuitiveley speaking the ponding would be like a BGA package). This direct bonding method, which would require performing a laser lift-off to separate the fabricated μLED array from its sapphire substrate, simplifies the integration process significantly and has no detrimental impact on the μLED fabrication process.

Chapter 5

ADC-less Adaptive σ-Based Threshold for Spike Detection in SNR-Enhanced Neural Recording

5.1 Introduction

Neural spikes, as captured in electrophysiological recordings, particularly through extracellular methods, are crucial for understanding the complex dynamics of the brain. Unlike action potentials that occur within individual neurons, spikes in these recordings often represent the combined activity of multiple neurons firing near-simultaneously. This collective neural activity is vital for interpreting the functional aspects of brain networks. It should be noted that the extracellular recording includes both spikes and Local Field Potentials (LFPs) which represent the summed electrical activity of a larger neuron population within a certain vol-

ume of brain tissue. LFPs typically exhibit lower frequencies compared to spikes, generally in the range of 0.5 Hz to 300 Hz. As a result, using a highpass filter, it is fairly simple to remove the LFP content of the recording signal even though the LFP signal may have larger amplitude in comparison with the spiking activities. Spikes provide a window into the cooperative and synchronized activities of neuronal populations, revealing how groups of neurons communicate and process information in concert. The ability to accurately detect and analyze these spikes is therefore essential for decoding the complex patterns of neural communication and understanding the emergent properties of neural networks. This focus on spikes reflects a broader perspective in neuroscience, emphasizing the importance of collective neural behavior over individual neuronal events in understanding brain function and disorders.

The detection and analysis of neural spikes are particularly crucial in the development and refinement of Brain-Computer Interfaces (BCIs) and other advanced medical technologies. As shown in Fig. 5.1, in a general BCI application, accurate spike detection is essential for translating neural activity into commands that control external devices, offering life-changing possibilities for individuals with mobility or communication impairments. This technology relies heavily on the precise detection of spike patterns, as these represent the collective neural activity that can be harnessed to operate prosthetics, computer interfaces, and other assistive devices. Beyond BCIs, the insights gained from spike analysis are instrumental in the development of neurotherapeutic strategies and rehabilitation techniques.

A key factor in enhancing the effectiveness of applications like Brain-Computer Inter-

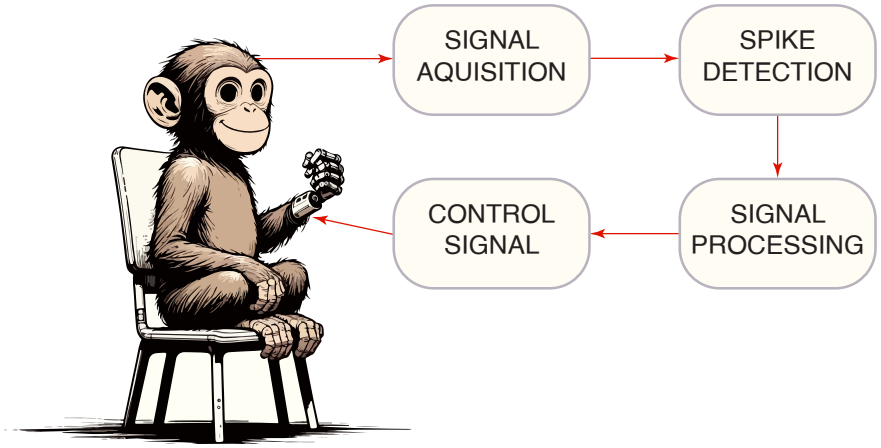


Figure 5.1: Conceptual block diagram of a BCI application.

faces (BCIs) and advanced medical technologies is the ability to utilize a large number of recording electrodes [25,99,100]. This multiplicity of electrodes allows for a more comprehensive capture of neural activity, providing a more detailed representation of brain function. In BCIs, for instance, having a greater number of electrodes enhances the system's ability to interpret neural signals accurately, leading to more precise control of external devices and a better user experience. Similarly, in medical diagnostics and treatment, multiple electrodes enable more complete monitoring and analysis of neural patterns.

In high-channel recording systems, particularly those used in implantable applications, the process of spike detection presents unique challenges. Generally, spike detection relies on setting a threshold level, with neural activity above this threshold identified as a spike. However, in many research works that perform spike detection and spike sorting, this threshold level has to be set manually. Thus, in systems with a high number of electrodes, this approach becomes increasingly complex.

Firstly, the threshold for spike detection needs to be individually defined for each electrode, accommodating variations that can occur between different individuals, over time,

and across the areas of the high channel count μelectrode array. Secondly, for manually setting the threshold level in an implantable device, it is required to record the data with high sampling rate (2$\times$ spike activity highest frequency content$\approx$15kHz), digitize the recordings, transmit the data to external unit, find the threshold value and transmit the threshold information back to the IC to set the threshold value. All of these steps need to be repeated once in a while to account of the changes in the threshold value over time. Therefore, even if after setting the threshold value there is no need for constant recording with high bandwidth or even digitization, this features should be available on the implanted IC, which restricts the design process.

Lastly, having a high channel count system in an implantable application where power consumption is a critical constraint, results in very limited power budget per channel. Due to the power-noise trade-off, this restrict on power consumption, results in the recording front-end's integrated input referred noise to be large. This along with the environment background noise, results in operating in conditions of low Signal-to-Noise Ratio (SNR) recording, making the detection of genuine spikes using raw recording signal more difficult.

5.2 Adaptive Spike Detection

The primary objective of adaptive spike-detection algorithms is to finely tune the detection threshold to effectively capture spiking activities while excluding incidental peaks attributable to background noise. A fundamental assumption in these techniques is that the background noise follows a Gaussian distribution with a mean of zero. This assumption is largely supported by empirical observations from actual neural recordings, which typically

exhibit a Gaussian noise distribution, albeit with slightly broader tails as indicated in literature [101]. Consequently, the noise in these recordings can be adequately characterized by its root mean square (rms) value, which, in the case of a Gaussian distribution, corresponds to its standard deviation (σ).

The main target of these algorithms is to accurately determine the standard deviation of the background noise. Once this parameter is established, the algorithm sets a threshold at a multiple of σ value, effectively filtering out the majority of the background noise. For instance, implementing a threshold of 5σ significantly minimizes the likelihood of false detections due to noise. In statistical terms, with a 5σ threshold, the probability of Gaussian noise erroneously triggering the spike detector is approximately 3×10^{-7}. However, as in biomedical application we do not usually require such high precision, the threshold level is usually set between 2-4 times of the standard deviation balancing the sensitivity to detect all spiking activity and robustness against noises.

5.2.1 Analog vs Digital Implementation

The process of calculating the standard deviation of the background noise can be executed in both analog and digital domains. However, these two approaches offer distinct advantages and challenges, especially in the context of high channel count recording applications. In digital implementation, it is required to digitize all the data using Nyquist rate sampling rate (15kHz) and store a large enough amount of data to perform the digital computation. This method is acceptable if we need the this large amount of data to completely reconstruct the data in certain research applications. However, in real-life applications especially when

having a large number of channels, it is not practical to handle all this data. Therefore, the goal is to perform as much computing as possible in the analog domain and ideally extract the spiking activity which can trigger real-time application or trigger specific algorithms that have been implemented using spiking neural networks which can best imitate the biological activities [102–104]. Therefore, while both analog and digital domains have their merits in adaptive spike detection, analog implementation emerges as a more suitable option for high-channel count applications considering the advantages listed below:

- **No Need for Data Storage**: Analog systems inherently process signals in real-time, eliminating the need for large-scale data storage, which is a significant advantage in high-channel count applications.
- **Energy Efficiency**: Analog processing can be more energy-efficient, a critical factor in implantable high-channel-count neural recording systems.
- **Real-Time Processing**: The ability to process signals continuously and in real-time allows for lower latency in spike detection, which is crucial for applications requiring immediate response, such as in some BCI implementations.
- **Compactness**: Analog circuits can be more compact than digital ones for simple calculations, beneficial in applications where channel area is a constraint.

5.2.2 Spike Detection Using Raw Data vs SNR-Enhanced Data

As shown in Fig. 5.2, spike detection can be approached in two primary ways: directly using raw data (post-amplification and high-pass filtering to isolate spiking activity) or after ap-

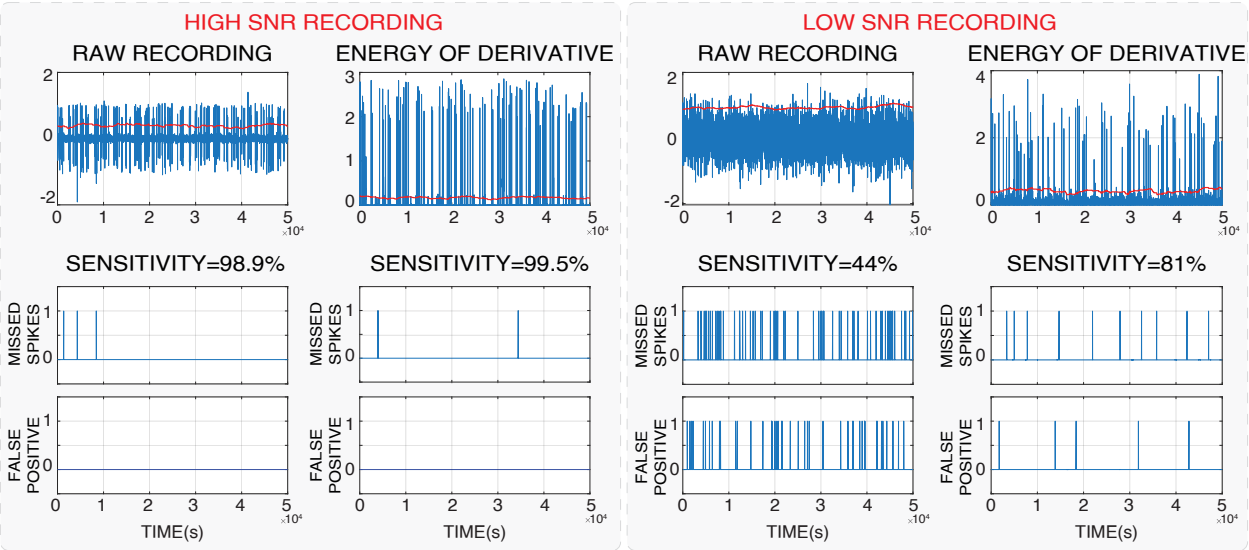


Figure 5.2: Performance of spike detection for high and low SNR recordings using raw and SNR enhanced signals which shows sensitivity improvements especially in low SNR conditions

plying Signal-to-Noise Ratio (SNR) enhancement techniques, such as the Nonlinear Energy Operator (NEO) or Energy of Derivative (ED). The SNR enhancement techniques function by attenuating the low-frequency and low amplitude content of the signal (often where most of the noise resides) and accentuating the high-frequency and high amplitude content, which is indicative of spiking activity [105–108]. By doing so, they effectively enhance the energy content of spikes relative to the background noise. This results in a clearer distinction between neural spikes and noise, facilitating the task of defining the spike detection threshold.

Therefore, as mentioned before, in high channel count applications, where power constraints are a primary concern and each channel typically operates with a limited power budget [109]. This restriction often leads to a higher level of integrated input-referred noise in the recording front-end, resulting in a lower SNR, which can significantly hinder the effectiveness of spike detection when relying solely on raw data. The amplified noise level can obscure the spikes, leading to increased false negatives (missed spikes) or false posi-

tives (noise mistaken for spikes). By contrast, SNR enhancement techniques can mitigate these issues by boosting the signal component relative to the noise, making the spikes more prominent and distinguishable even in a low SNR setting. Fig. 5.2 shows examples of spike detection with and without ED operator and how the ED operator is more necessary in case of high background noise. In the following we provide a brief explanation about the main SNR enhancement techniques in literature:

- **Nonlinear Energy Operator (NEO)**: The Nonlinear Energy Operator (NEO) is a signal processing technique designed to enhance the detection of energy transients in signals, such as neural spikes [105–108]. NEO works by nonlinearly combining a signal with its first and second derivatives to accentuate high-energy features. For a continuous signal $x(t)$, the NEO is defined as:

$$\Psi_{NEO}(x(t)) = x'(t)^2 - x(t)x''(t) \tag{5.1}$$

  This equation amplifies points in the signal where there is a significant energy difference, thereby enhancing sharp transients.Therefore, NEO is especially effective in suppressing low-frequency components prevalent in background noise, while enhancing high-frequency components associated with spikes.

- **Energy of Derivative (ED)**: The Energy of Derivative (ED) technique emphasizes the energy content of a signal's temporal derivatives to enhance spike detectability. ED is based on the premise that neural spikes, due to their rapid onset and offset, exhibit

considerable energy in their first derivatives. The energy of the derivative is calculated as follows:

$$\Psi_{ED}(x(t)) = x'(t)^2 \tag{5.2}$$

ED amplifies parts of the signal with rapid changes, such as the edges of a spike, aiding in distinguishing spikes from slower signal components.

In selecting the most suitable technique for enhancing the detectability of spikes in our work, both the NEO and ED operators have been considered. However, we have opted to employ the ED operator due to the following reasons:

1) Simplification of Thresholding: The ED technique, which focuses on the square of the signal's derivative, simplifies the thresholding process by eliminating the need for dual-threshold levels required when using the NEO output. This is also another challenge when dealing with raw data, where having two thresholds for spike detection is crucial. By working with the square of the derivative, the ED operator ensures that all data points are positive, allowing for a single threshold level that can be determined in the analog domain.

2) Analog Domain Compatibility: The implementation of our adaptive threshold generation circuit necessitates the entire process to remain within the analog domain to avoid the use of high resolution ADCs. Therefore, the ED operator is chosen as it provide a straightforward and resource-optimized analog implementation.

Beside the qualitative explanation, the superiority of analog implementation and SNR enhancement techniqes are validated using quantitative simulation results. Fig. 5.3 shows the receiver operating characteristic (ROC) curve for spike detection sensitivity versus the false

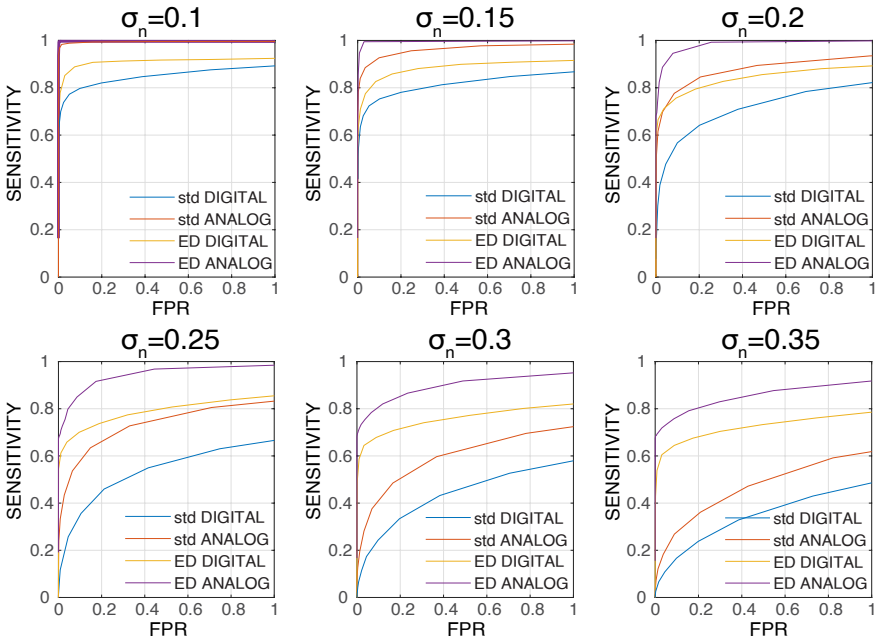


Figure 5.3: Simulation result for sensitivity versus false positive rate ROC for R_{STD} and ED_{STD} methods with analog and digital implementation.

positive rate which is generated by changing the coefficient that is used to set the threshold voltage based on the value of the raw signal standard deviation *(R_{STD} method)* or standard deviation of the energy of derivative (ED) signal *(ED_{STD} method)*. The comparison shows how the spike detection of the raw signal can be effective for low-noise recording but it will fail for recording with higher noise and the SNR enhancement performed by the ED operator makes it possible to detect the spike even in a noisy background.

In the digital implementation for both methods, a constant number of previous samples are used to define the sigma of the raw data or sigma of the ED signal. For analog implementation all the data points from the start of simulation are considered to calculate the sigma at each point. As a result, with analog implementation, the spiking activities have minimal effect on the calculated standard deviation due to the sparse spiking activity. With digital implementation, the spiking activities will effect the calculation of the background noise and

that is the reason that analog implementation has better performance than their digital counterparts. This issue is a well-studied challenge and the proposed solution is to somehow remove the spiking activity from the background to avoid its effect on the background noise [110, 111]. This cause extra complexity in the process and require a large buffer to keep the previous data points or interacting with memory blocks on the chip.

The other point that is visible from the simulation result is that for data from low-noise recording, the analog implementation of RSTD method has better performance of the digital implementation of EDSTD method. However, for high-noise recording, the performance of the RSTD method degrades and thanks to the SNR improvement feature of the ED operator, even the digital implementation of the EDSTD method has better performance.

Based on the advantages and limitations outlined in the preceding sections, our goal is to achieve *adaptive spike detection* that not only operates robustly within the *analog domain* but also incorporates *SNR enhancement* techniques to ensure the precision and robustness of spike detection.

5.3 System-Level Architecture

In this section, we focus on the system-level implementation of our novel adaptive threshold, ED-based spike detection methodology. Fig. 5.4 shows the concept of analog adaptive threshold for raw neurological recording. Considering that the spikes are sparse activities the target is to find the standard deviation of the background noise. It has been shown that measured background noise from actual recording have almost a Gaussian distribution. Consequently, when a data stream characterized by a Gaussian noise distribution is compared against its

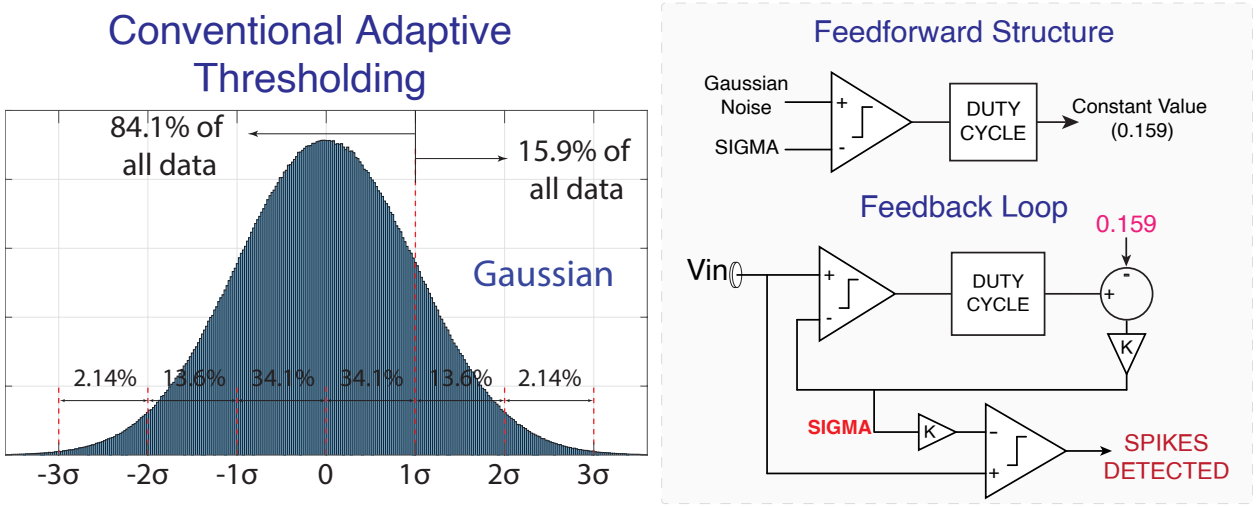


Figure 5.4: System-level implementation of adaptive threshold generation for raw recording data.

own standard deviation, the duty cycle (defined as the proportion of time the output signal is 'high' or at a logic level of 1, over the total observation time) would stabilize at a value of $0.159 \times \mathrm{VDD}_{Comp}$. This value represents the percentage of time that the signal exceeds its own standard deviation within the Gaussian distribution. To use this concept in calculating the value of the the standard deviation of the background noise, the feedback loop of Fig. 5.4 is used when the high gain loop (Due to the presence of the Comparator) will set the voltage at the negative input of the comparator equal to the standard deviation of the Gaussian noise. It should be noted that the comparator in this structure is followed by a latch and therefore, there is not reset phase in the output of the comparator.

To extend the principle of adaptive thresholding to the output of our ED operation, a prerequisite condition must be satisfied: the ED output should exhibit a constant distribution when the input has Gaussian distribution. This consistency is crucial because our goal is to determine the adaptive threshold by identifying the standard deviation of the ED output

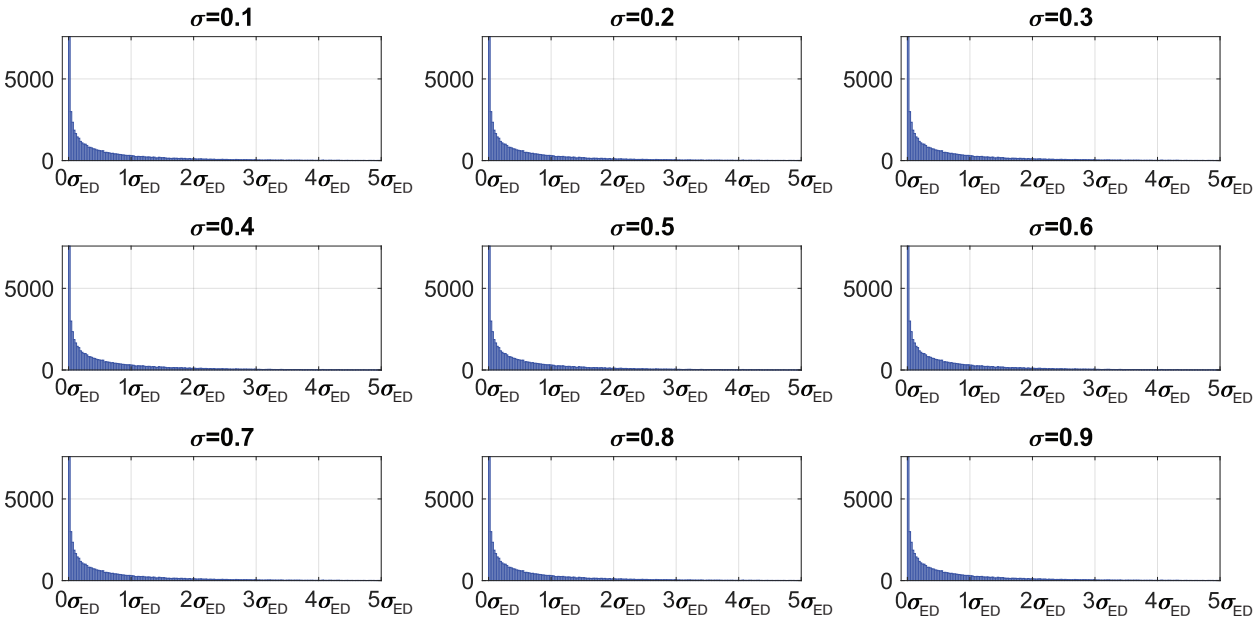


Figure 5.5: Simulation results of ED processed Gaussian noise with different standard deviation values with the x-axis of the distribution plots normalized against σ_{ED}.

(σ_{ED}).

To validate this requirement, we have fed various data stream with Gaussian distributions ($\mu = 0$ and varying σ) to the ED operation. As shown in Fig. 5.5, the transformed distributions retain a similar shape when the x-axis of the distribution plots is normalized against σ_{ED}.

It is evident from this distribution that across different noise levels, 23.4% of the data lies beyond σ_{ED}. In other words, if we feed the output of the ED operation and its computed σ_{ED} into a comparator and measure the duty cycle of the comparator's output, it will reach a steady-state value of $0.234\times\text{VDD}_{Comp}$, following the dissipation of initial transients.

Leveraging this property, the feedback structure illustrated in Fig. 5.6 is employed. Within this setup, the closed-loop system (owing to the high gain provided by the comparator) ad-

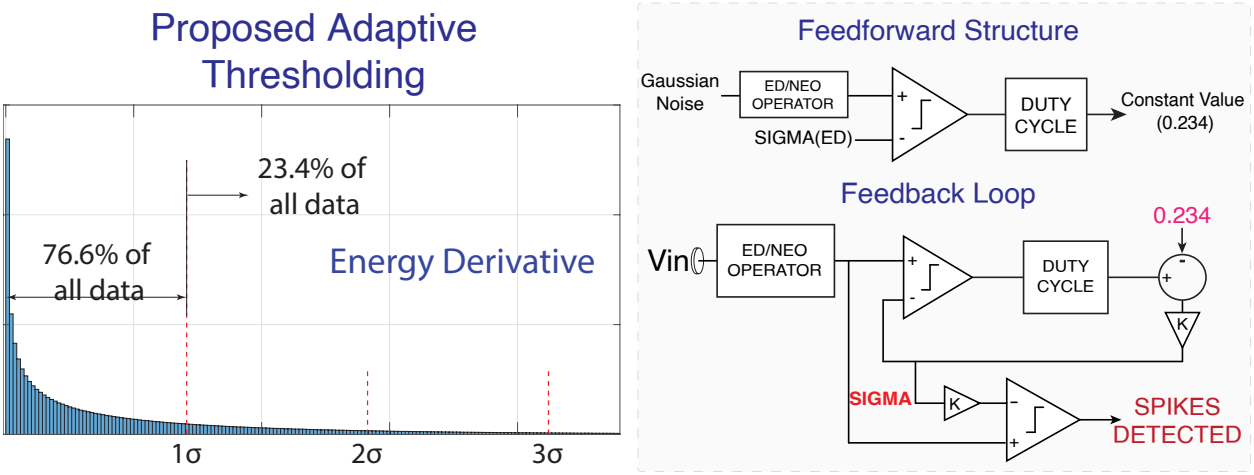


Figure 5.6: System-level implementation of the proposed adaptive threshold generation for ED-processed data.

justs the voltage at the comparator's negative input to σ_{ED} value, such that the duty cycle of the comparator's output stabilizes at 0.234.

5.4 Circuit-Level Implementation

In this section we explain the circuit-level implementation of our proposed technique for adaptive threshold generation for the output of ED operator. Fig. 5.7 shows the to-level implementation of this structure where the first stage is the low-noise and low-power bio-amplifier which is implemented by a two-stage capacitively coupled structure which also implements a 300-8kHz bandpass filter. The next stage is to find the signal derivative which is implemented in a discrete fashion using passive switched capacitor structure. This eliminated the need for an opamp to implement the derivative which can cause instability. The signal derivative is then fed to a squarer circuit which is implemented by Gilbert cell structure. The Gilbert cell has a single-ended structure and the output is the result of the ED operator. This signal is then fed to the loop which defines its standard deviation (σ_{ED}). In this structure,

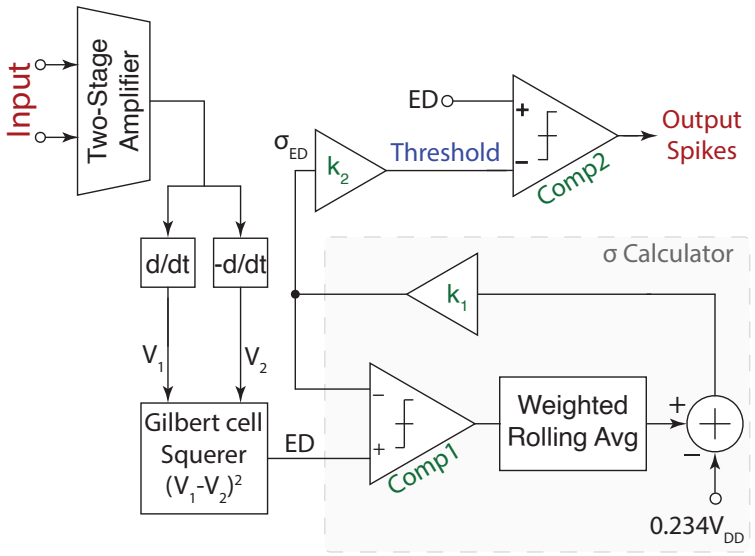


Figure 5.7: Top-level implementation of the adaptive threshold generation using a feedback loop structure based on the standard deviation of the background noise.

the duty cycle of output of the comparator is calculated using a passive switch capacitor structure. The output of the switched capacitor is then followed by a low-pass RC filter to calculate the duty cycle value. The output of the duty cycle is then compared with the constant value of 0.234×VDD and the subtracted voltage is fed to the negative input of the comparator which would be equal to σ_{ED}. The threshold of spike detection is then set to 2-4 times of the σ_{ED}, which is connected to the negative input of the spike detection comparator where the positive input is the ED output and the output of this comparator will show the detected spiking activities. In the following the main blocks and sections of this top-level structure is explained in detail.

5.4.1 Two-Stage Amplifier

As shown in Fig. 5.8(a), the first stage of the system focuses on signal amplification and filtering, employing a two-stage capacitively-coupled amplifier. Although it is possible to design the recording front-end to record both LFP and spikes, considering their different

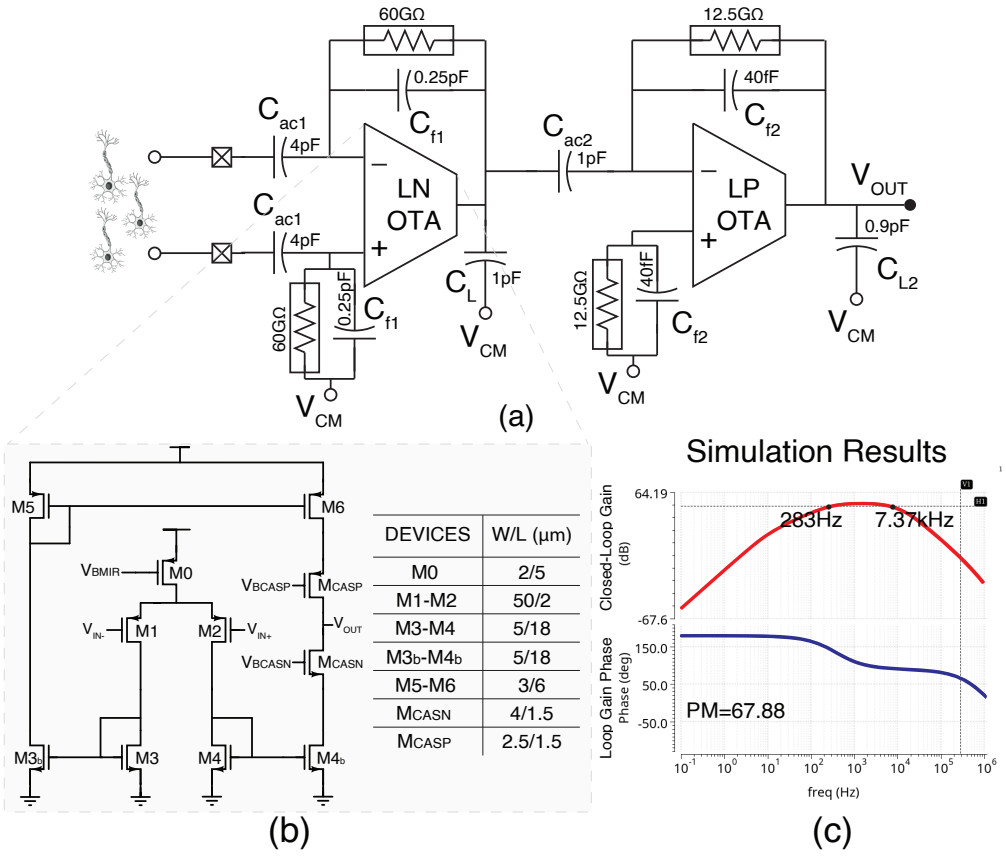


Figure 5.8: The two-stage capacitively-coupled OTA and the schematic of the low-noise OTA with the transistor sizing and the simulation results showing the closed-loop gain and phase margin.

signal characteristics and frequency bands, it would be a suboptimal design to record both signal with the same front-ens and it is more efficient to have dedicated circuit with different bandwidth for each application (i.e., 10-300Hz for LFP and 300-8kHz for spikes). As the target application of this structure is to detect the spiking activity, this two-stage amplifier tailored for spike activity recording and Low-Frequency Potential (LFP) signal rejection. In this design, the constant gain is secured through a capacitive gain structure, the high pass pole is determined by the feedback RC network, and the low pass pole is derived from $G_m/A_M C_L$ with $A_M = C_{in}/C_f$, which are designed to provide a gain of 52dB and bandwidth of 300Hz-8kHz.

As the first stage is more prone to noise, a low-noise OTA is used in this stage. Following this initial amplification, the second stage employs a low-power OTA, as the signal is now more robust against noise factors. Fig. 5.8(b) shows the schematic and transistor sizing of the low-noise OTA used in the first stage. The low-power OTA of the second stage is designed with the exact same architecture and transistor sizing minus the size of the M_0 transistor which has W/L of 1μm/5μm, making the DC bias current of the low-power OTA (0.25μA) half of the low-noise OTA (0.5μA). In this design, the current mirror OTA is selected for its robustness and potential to achieve a low Noise Efficiency Factor (NEF). Adhering to the g_m/I_D methodology outlined in [112], the integrated input referred noise of the OTA is shown to be inversely proportional to the g_m of transistors M_1 and M_2, and directly proportional to the g_m of transistors M_3, M_4, M_{3b}, M_{4b}, M_5, and M_6. This necessitates maximizing the g_m/I_D ratio for the input transistors, which are set to operate in the sub-threshold region,

thus demanding a large W/L ratio. However, for M_{3-6}, M_{3b}, and M_{4b}, the g_m/I_D ratio should be minimized, which is achieved by having a W/L ratio less than unity. Given that flicker noise is inversely proportional to the transistor dimensions ($W \times L$), enlarging both W and L is a strategic choice to reduce flicker noise.

After the initial design, the noise summary tool is employed to pinpoint and attenuate the primary noise contributors. If thermal noise from the feedback resistor predominates, lowering the low cut-off frequency below the target can be advantageous. Should thermal noise from the input transistors be significant, increasing the tail bias current or enlarging the W/L ratio of the transistors is beneficial. It should be noted that the g_m/I_D plot plateaus after some point and therefore, increasing the W/L of the transistor will not increase the gm proportionally. In cases where flicker noise is predominant, increasing transistor size or employing a chopper structure is recommended, even though the latter introduces complexities that necessitate additional circuitry and specialized filters.

5.4.2 Energy of Derivative Calculation

The subsequent stage in our system is responsible for computing the Energy of Derivative (ED), which is derived from squaring the signal's derivative. As previously discussed, the purpose of this stage is to amplify the signal-to-noise ratio (SNR) of spike recordings. This is achieved by emphasizing the high-frequency components, representative of spiking activities, while suppressing the low-frequency components, associated with noise. In this block, differentiation is executed through a discrete-time, switched-capacitor-based differentiator, and

the squaring function is realized using a Gilbert-cell multiplier. Both of these components are elaborated in the following.

5.4.2.1 Switched-Capacitor-Based Diffrentiator

Derivative calculation can be performed using an OTA with resistive feedback and capacitive input which would generate the transfer function of SR_fC_{in}. However, using this structure is challenging due to: 1) Its sensitivity to high frequency noise as this circuit inherently amplifies high-frequency signal. 2) The stability issues due to the phase shift introduced by the capacitive input. 3) Its limited dynamic range that is constrained by the OTA's output range which can results in saturation in the presence of high frequency noise. 4) The OTA in this structure is an active component and consumes power. To resolve these issues, in this design we have implemented the derivative using a discrete-time switched-capacitor-based Diffrentiator. In this structure, the differentiation is implemented using passive components and therefore is more power efficient in comparison with the OTA-based implementation. The design is implemented in discrete fashion and therefore, is inherently stable. Also, the clock frequency is selected based on the signal Nyquist rate and therefore, it is not prone to high frequency noise saturation.

Fig. 5.9 shows the circuit-level implementation of the switched-capacitor-based Diffrentiator. In this structure, during φ_1 phase the signal is sampled on C_{S1} and during the φ_2 phase the signal is sampled on C_{S2}. Next, during the φ_O phase, the C_{S1} and C_{S2} capacitors are connected in series with opposite polarities which means that their voltages are subtracted from each other and is stored on the C_O capacitor. The M_O transistor is turned on with a

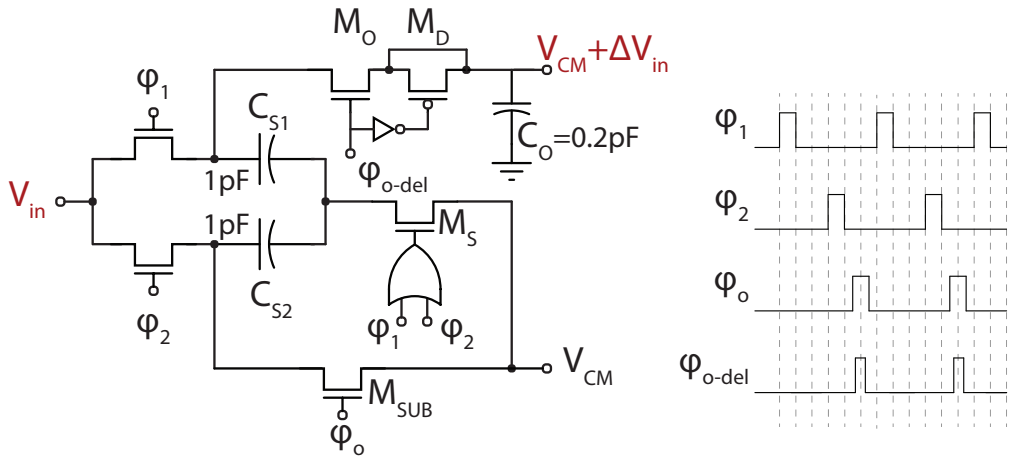


Figure 5.9: The implementation of discrete-time diffrentiator.

delay to ensure that the C_O capacitor does not affect the subtraction and is turned off before the M_{SUB} transistor to avoid the effect of its charge injection. Also, the M_D transistor is used in a dummy structure to absorb the charge injection of the M_O transistor. It should be noted that with this structure the differentiation is performed in a non-overlapping fashion.

Using this structure also simplifies the biasing of the proceeding stage which is the input of the Gilbert cell, as the sampled voltage on the C_O capacitor is equal to:

$$V_{out} = V_{CM} + (V_{in}[\varphi_1] - V_{in}[\varphi_2])\frac{0.5p}{0.5p + 0.2p} \tag{5.3}$$

where the V_{CM} voltage is the DC bias of the output and can be set to whatever is suitable for the next stage without affecting the performance of this stage. Additionally, we can see from this formula that the size of the C_O should be as small as possible to avoid attenuating the output voltage. However, reducing the size of the storage capacitor, makes it prone to fast discharging and clock feedthrough. In order to resolve this issues, a duplicate of this switched-capacitor-based Diffrentiator is used where the φ_1 and φ_2 phases are opposite of

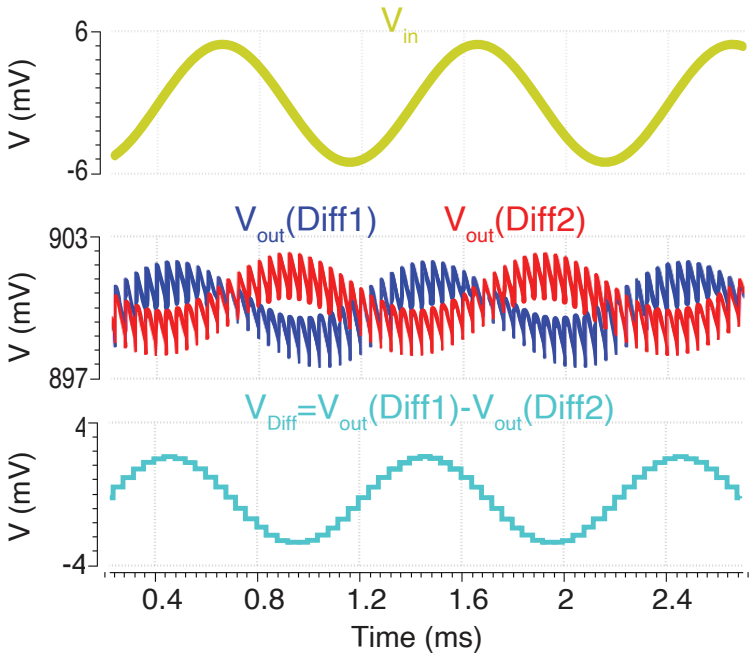
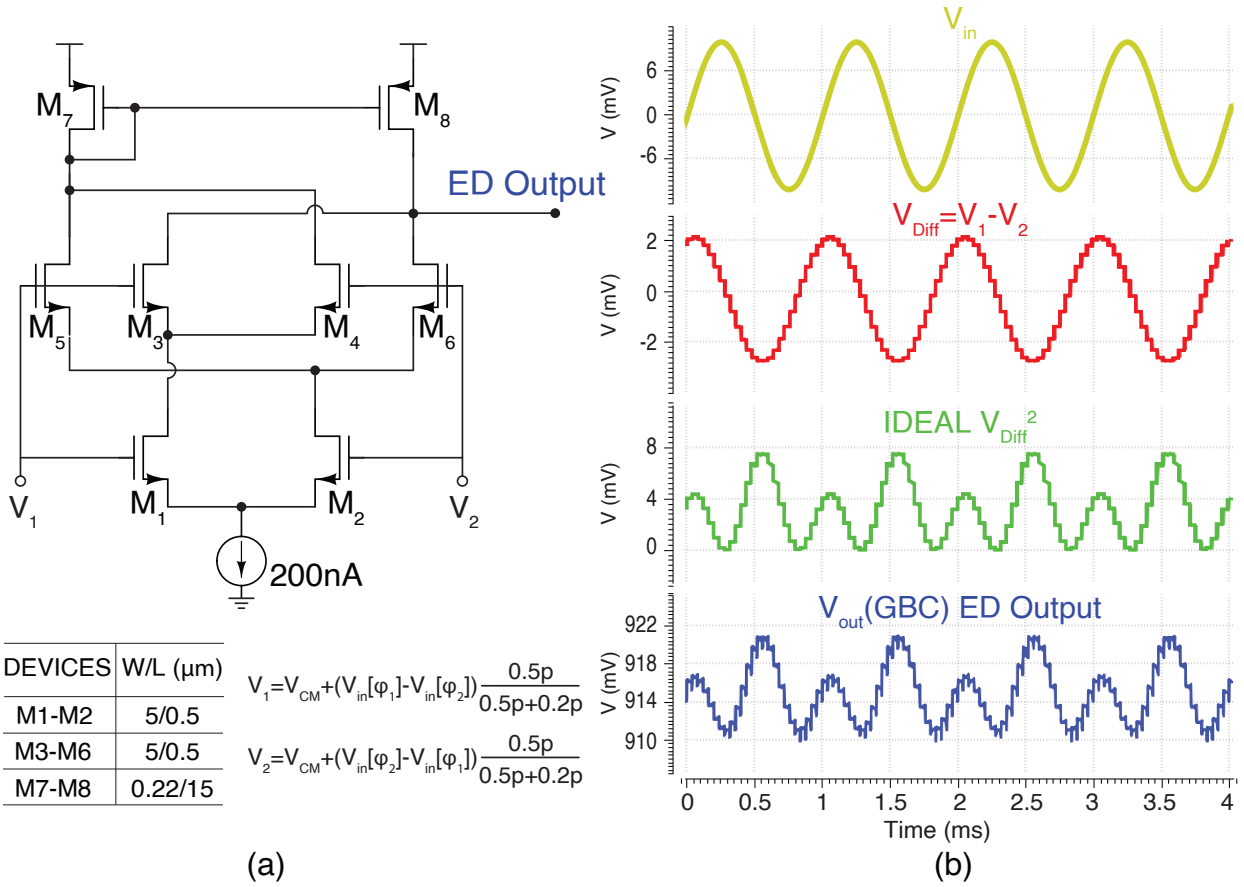


Figure 5.10: Simulation results showing the functionality of the pair of opposite phase Diffrentiators.

Fig. 5.9, but the φ_O is the same. As shown in Fig. 5.10, using this techniques, the output of the two Diffrentiator have the same non-idealities which is canceled out on the difference of the two waveforms which shows a clean discrete-form differentiation of the input voltage.

5.4.2.2 Gilbert Cell Squarer

Fig. 5.11(a) shown the MOSFET-based Gilbert cell structure that is used to implement the the multiplication function. In this structure by connecting the bottom differential pair (M_1 and M_2) and the top differential pairs (M_3 and M_6) to the same voltages. In general application of Gilbert cell the bottom pair is connected to the radio frequency (RF) signal and the top pairs are connected to the local oscillator (LO) signal and the output is proportional to RF$\times$ LO signal. Using the same concept, the output of the Gilbert cell with the structure shown in Fig. 5.11(a) is $(V_1 - V_2)^2$. Considering that the V_1 and V_2 voltages are proportional



DEVICES	W/L (μm)
M1-M2	5/0.5
M3-M6	5/0.5
M7-M8	0.22/15

$$V_1 = V_{CM} + (V_{in}[\varphi_1] - V_{in}[\varphi_2]) \frac{0.5p}{0.5p+0.2p}$$

$$V_2 = V_{CM} + (V_{in}[\varphi_2] - V_{in}[\varphi_1]) \frac{0.5p}{0.5p+0.2p}$$

Figure 5.11: (a) The schematic of the gilbert cells along with the transistor sizings. (b) Simulation results showing the functionality of the Gilbert cell.

of dV_{in} and $-dV_{in}$, this subtraction also provides a gain of 4.

Fig. 5.11(b) shows the simulation result of the Gilbert cell which shows that the circuit can generate the $(V_1 - V_2)^2$ waveform with good matching with the ideal square of the voltage difference. It should be noted that the reason for asymmetry in this results is that the difference waveform is calculated based on non-overlapping sampling and as a result, the V_{Diff} waveform is not symmetric.

5.4.3 σ_{ED} and Threshold Calculation

Following the calculation of the Energy of the Derivative (ED) value, the next analytical step is to ascertain its standard deviation, represented as σ_{ED}. This parameter serves a critical role in the establishment of the adaptive threshold for spike detection, typically set to a range of 2 to 4 times σ_{ED}. The rationale for selecting σ_{ED} as the representative measure of background noise variance is the observation that spiking events are sparse relative to the background noise within the signal. Consequently, σ_{ED} predominantly reflects the fluctuations of the background activity, thereby providing a robust baseline against which spikes, characterized by their relatively higher amplitude, can be detected.

Fig. 5.7 illustrates the configuration of the feedback loop designed to generate the voltage of σ_{ED}. The output from the Gilbert cell, representing the computed ED, is fed into the positive terminal of a comparator. This stage is followed by a latch which means there is no reset phase. Subsequently, the duty cycle of the comparator's output is determined. This output is then subtracted from a preset voltage level of $0.234\times$VDD, and the result is directed to the comparator's negative input. The architecture of the feedback loop inherently adjusts the voltage at the comparator's negative input to σ_{ED} voltage. While this loop is similar to the configuration presented in [102], which was employed to establish the threshold for raw recordings, there are some main differences that are elaborated in the following.

5.4.3.1 Subtraction and Threshold Calculation

The primary distinction of the employed methodology resides in the unique output characteristics of the Gilbert cell, which yields an expression of $(V_1 - V_2)^2$ added to a constant

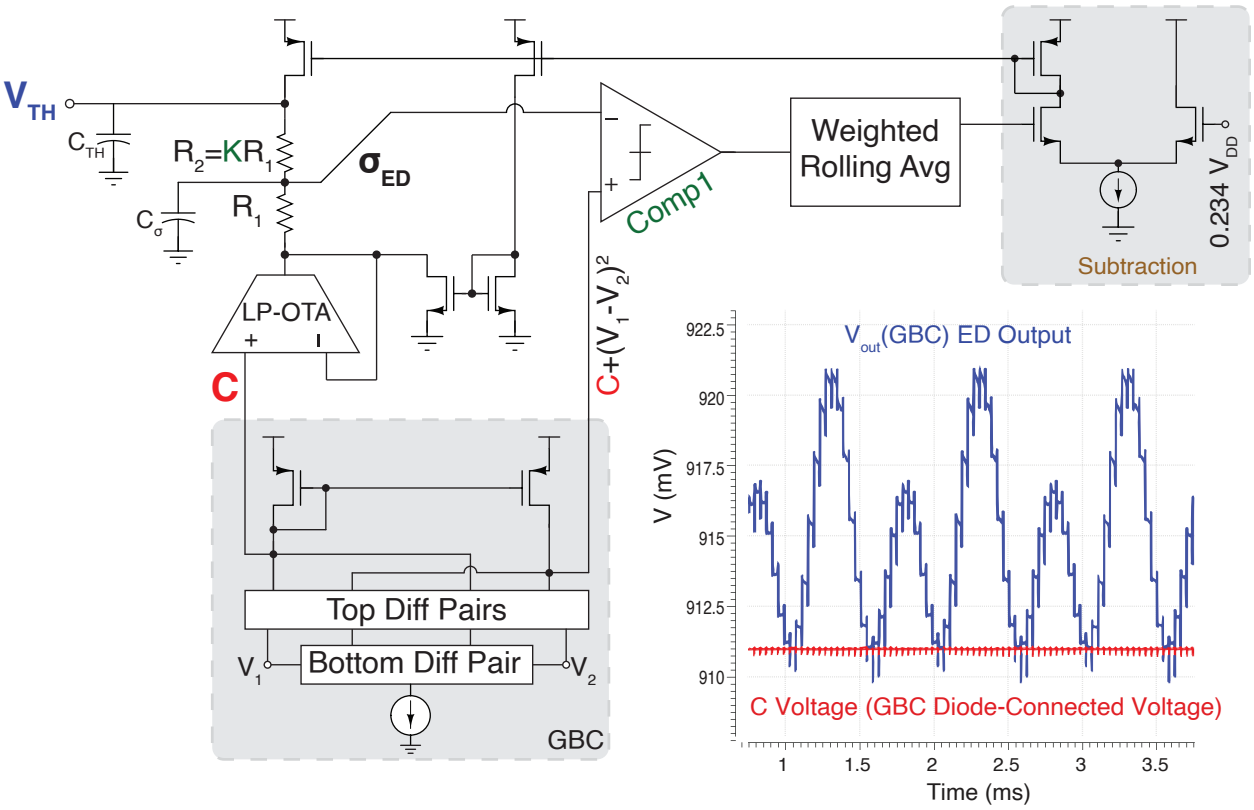


Figure 5.12: Detailed implementation of subtraction and hreshold voltage calculation in the proposed feedback loop and the simulation result showing the variations at Gilbert cell's low and high impedance outputs.

voltage term (C). From a statistical standpoint, the addition of a constant to every element within a dataset does not alter its standard deviation; hence, this constant voltage component does not affect the standard deviation of the ED signal. Nevertheless, the feedback loop in discussion is designed to set the comparator's negative input to $\sigma_{ED} + C$. Therefore, if the resistor R_1 were connected to ground as in a conventional structure, the resulting threshold voltage would be $K \times (\sigma_{ED} + C)$, deviating from the intended value of $C + K \times \sigma_{ED}$.

To rectify this and achieve the desired threshold, R_1 is not connected to ground but rather to the constant C voltage. It is important to clarify that this voltage does not represent the mean of the ED signal, but rather indicates the DC shift, the minimum value or the operating point of this signal. As depicted in Fig. 5.12, the voltage across the diode-connected transistor within the Gilbert cell architecture reflects the extent of this upward shift, considering it is a node of low impedance. Simulation results corroborate that this voltage corresponds to the minimum of the ED voltage, not its average.

In pursuit of this design objective, connecting R_1 directly to this node is avoided to prevent any interference with the Gilbert cell's operation by the injection of R_1 current. For optimal functioning of this buffer, the current passing through R_1, which is directed into the buffer, must be substantially lower than the buffer's bias current, adhering to the criterion $I_{Bias} > 10 \times I_{R_1}$.

In order to mitigate the increase in the buffer's power consumption and utilizing a low-power OTA, a specific technique is adopted. This method involves drawing the current of R_1 from the OTA output via a current mirror structure. This approach is crucial as it effectively

isolating the OTA's current flow from the current flow through R_1. Consequently, the OTA is able to maintain its functionality with a minimal bias current. This inventive configuration not only ensures efficient voltage buffering but also preserves the OTA's performance without necessitating a significant increase in power consumption.

5.4.3.2 Switched-Capacitor-Based Weighted Rolling Average

Another distinction in our feedback loop implementation, compared to conventional designs, lies in the duty cycle calculation mechanism. This circuit is tasked with quantifying the proportion of time during which the Comparator's output remains high, starting from the circuit's initial operation. This implies that rather than evaluating a fixed time window, the circuit retains all prior data, assigning greater significance to more recent inputs. This computation is fundamentally different from that of an integrator.

For a basic implementation, an RC structure would suffice to achieve this operation. However, the output from this block, which feeds into a subtraction circuit, requires a relatively stable value. Given that the comparator's output alternates between 0 and VDD, to minimize fluctuations at the RC output, a small pole, necessitating a large resistor and capacitor (time constant $>$ 100ms), is needed. In the conventional structure, this has been accomplished using a capacitor coupled to a g_m cell to generate a high resistive value of $1/g_m$. Yet, to attain the necessary time constant, such a configuration would demand a large capacitor (>10-100pF), which is impractical for on-chip implementation, given the area limitations in high-channel-count recording systems.

To overcome this challenge, we implemented a switched-capacitor-based weighted rolling

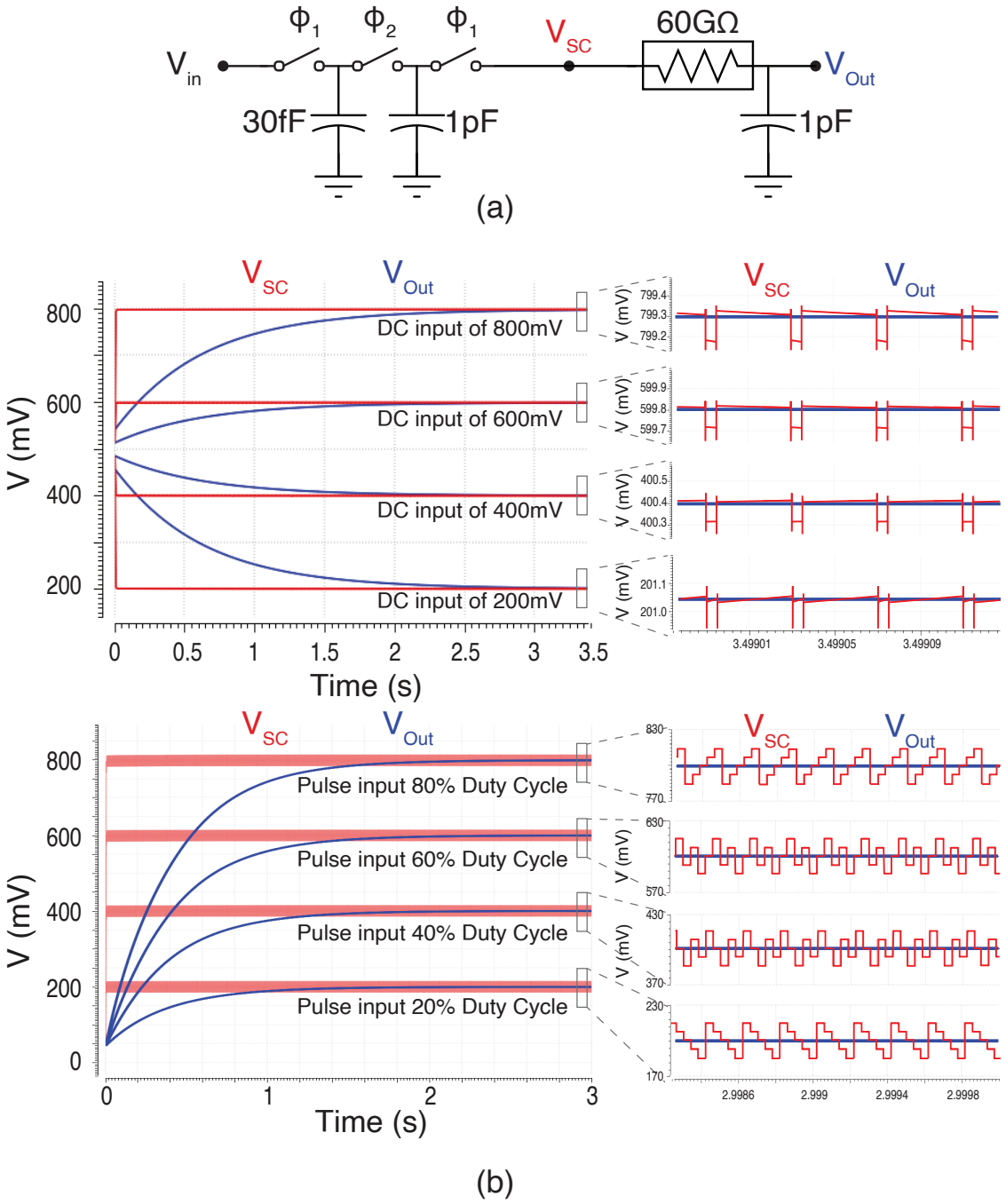


Figure 5.13: (a) Circuit-level implementation of the switched-capacitor based rolling average and (b) the simulation results for varying input DC levels and input pulse signal with varying duty cycles.

average structure, depicted in Fig. 5.13(a). To understand this circuit's operation, consider initially just the first two capacitors, C_1 and C_2. During the ϕ_1 phase, the input is sampled onto a 30fF capacitor. In the ϕ_2 phase, C_1 and C_2 are connected to equalize their voltages. The discrete-time operation of this circuit is expressed as follows:

$$Y_{[n]} = \frac{C_1 \times X_{[n]} + C_2 \times Y_{[n-1]}}{C_1 + C_2} \tag{5.4}$$

This equation represents the weighted rolling average of previous samples, allowing the circuit output to fluctuate around the target average. An additional output stage in this design aims to smooth the output of the C_1 and C_2 capacitors. For further clarification, Fig. 5.13(b) presents simulation results under two conditions: (a) when the input is a constant DC level and (b) when the input is a pulse waveform with varying duty cycles. The simulations confirm that the circuit accurately computes the average, with the switched-capacitor arrangement results in minor deviations around the target value and the subsequent stage applying a low-pass filter to produce a near-DC output voltage.

5.5 Measurement Results

In this section we focus on the experimental measurement results to validate the performance of the adaptive threshold spike detection circuit. Fig. 5.14 shows (a) the measured integrated input referred noise and (b) the voltage gain of the input two-stage amplifier, confirming an IIRN of 8.82μV in the band of interest, mid-band gain of 51.45 dB and a bandwidth of 237.5Hz-8.27kHz.

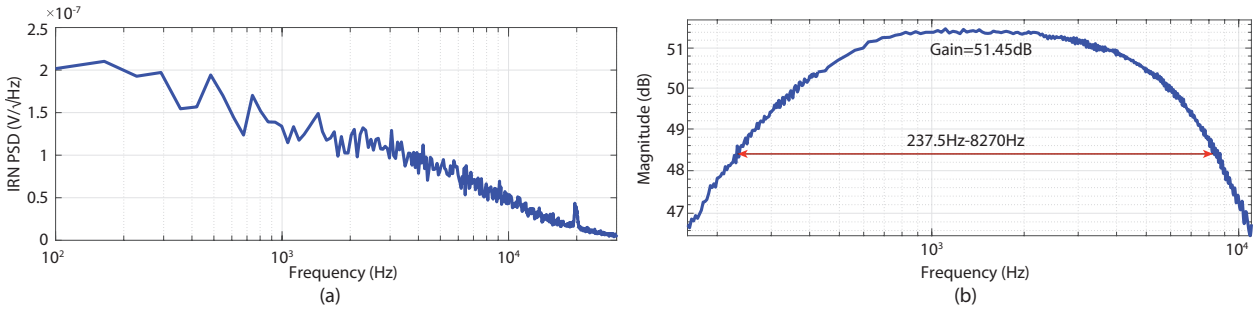


Figure 5.14: Experimental measurement results of (a) input referred noise, and (b) gain and bandwidth of the two-stage amplifier.

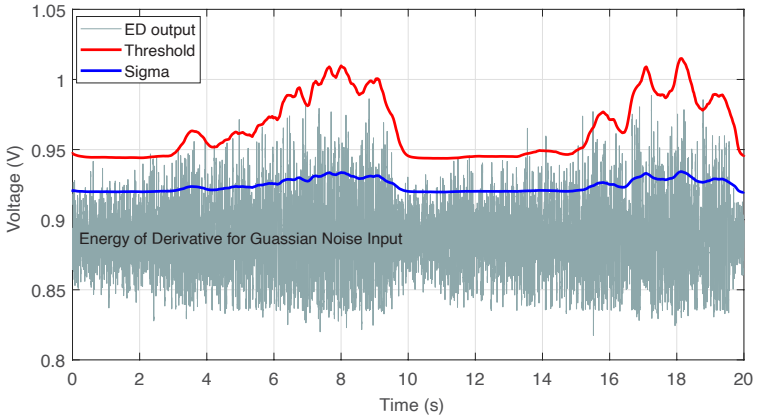


Figure 5.15: Dynamic response of the spike detection system to Gaussian noise input, showing the automatically calculated standard deviation and the adaptive threshold over a 20-second interval.

In the next experiment to validate the performance of the σ detection loop, a Gaussian noise with temporally varying standard deviation is fed to the recording front-end. In Fig. 5.15 we show the automatically calculated σ_{ED} and threshold voltages for a Gaussian noise with temporally varying standard deviation input signal. We can see that the threshold is set at a voltage that is effectively higher that the noise activity except in very few points (6 points in 20s of data).

In a subsequent experiment designed to further clarify the functionality of the (σ) detection loop, Gaussian noise characterized by a temporally varying standard deviation is fed to the recording front-end. Fig. 5.16 illustrates the feedback loop's capability to ascertain

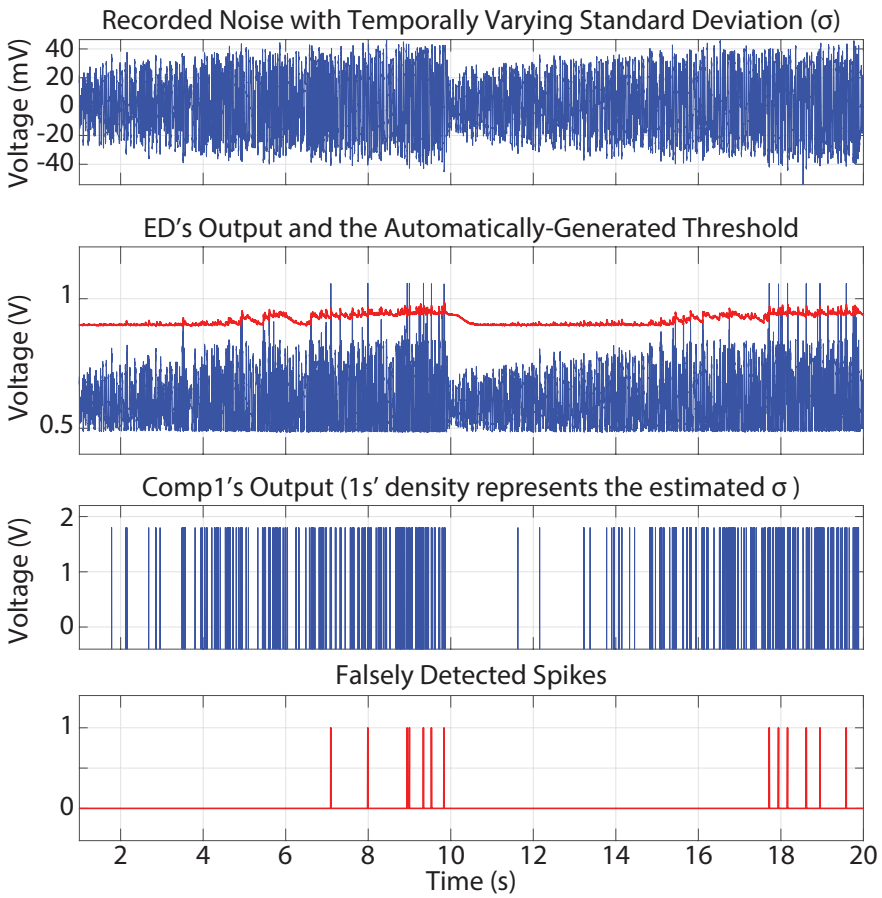


Figure 5.16: Experimental measurement results of (a) Input noise with varying standard deviation, (b) ED output with the automatically generated threshold, (c) in-loop comparator output indicating 1s density, and (d) identification of falsely detected spikes.

σ_{ED}, which fluctuates in accordance with the amplitude of the background noise. Furthermore, the in-loop comparator's output was analyzed, showing the proportionality between the incidence of 'ones' in the comparator's output and the standard deviation of the ED signal. The final graph depicts instances of falsely detected spikes, which are dominantly occur under conditions of exceedingly high noise levels.

Fig. 5.17 presents the measurement outcomes when pre-recorded neural signals, encompassing both local field potentials (LFP) and action potentials (spikes), were input into the system. The top plot displays the combined LFP and spike signal. The subsequent plot demonstrates the efficacy of the bandpass amplifier in isolating the spikes by attenuating the LFP components. The third plot illustrates the spikes being accurately identified using the Energy of the Derivative (ED) method, with the automatically generated threshold indicating the detection cutoff. The final plot showcases that spikes with extended duration are not falsely counted multiple times, owing to the refractory period mechanism integrated within the spike detection algorithm. This refractory period ensures that once a spike is detected, a certain time must elapse before another spike can be identified, thereby preventing overcounting and improving the fidelity of spike detection.

Fig. 5.18 presents the experimental measurement result of a comparative evaluation of the presented spike detection circuit in environments with different levels of background noise, using a dataset of prerecorded neural signals. The top half of the figure represents conditions with low background noise. It shows the original input signal as the ground truth provided by the QQ Standard Spike Dataset from the University of Leicester [113]. The fol-

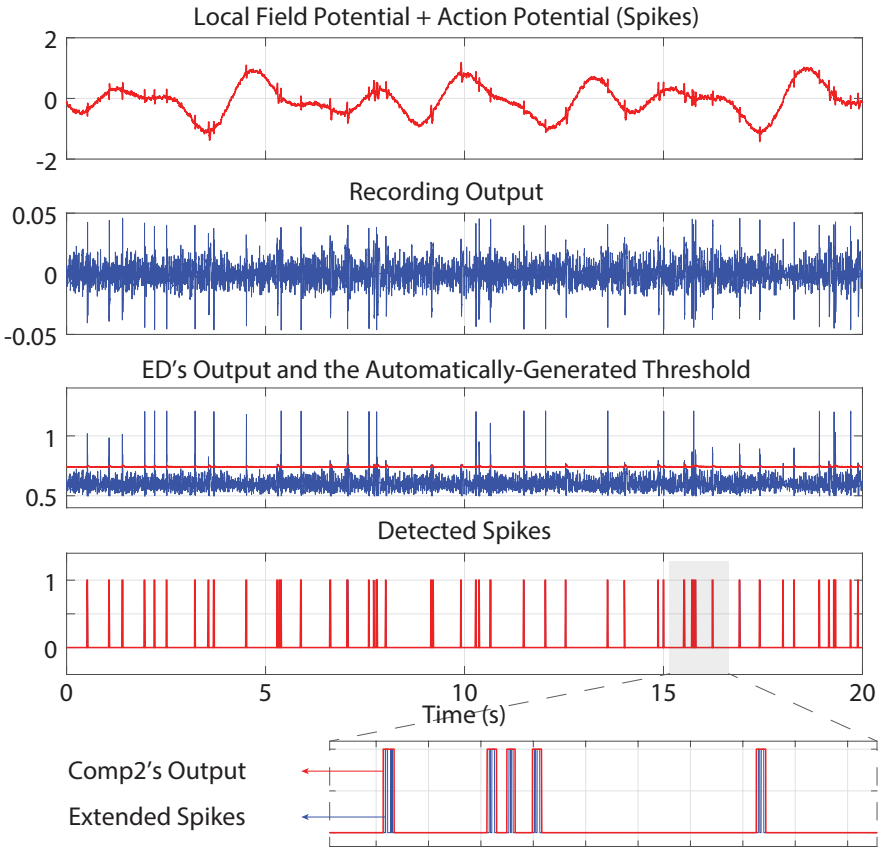


Figure 5.17: Experimental measurement results showing the functionality of the spike detection in the presence of LFP signal ED-based spike detection, and prevention of over-counting extended spikes due to an implemented refractory period.

lowing plot reveals the recording output, where the low noise level is evident. Subsequently, the ED output is displayed along with the automatically generated threshold, demonstrating the system's capability to detect spikes against the background noise.in the final plot, the detected spikes are illustrated, indicating a successful identification process.

Conversely, the bottom half of the figure depicts the scenario when the background noise is significantly higher, posing a challenge for spike detection. The initial plot once again shows the input signal which is the ground through to validate the effectiveness of the ED and adaptive threshold loop in identifying the spiking activity. The recording output, with increased noise level, underscores the difficulty in setting an appropriate threshold for the raw data. Nevertheless, the application of the ED operator transforms the signal in such a manner that threshold setting becomes feasible, as shown in the third plot. The ED operator effectively enhances the spikes, making them distinguishable from the noisy background. The final plot in the sequence illustrates the detected spikes, confirming that, despite the presence of substantial noise, the system is able to accurately identify spikes, thereby validating the effectiveness of the ED operator in high-noise conditions. The post processing of 5 data epochs (each comprising 100-ms windows of prerecorded neural signals) shows the spike detection circuit achieves a 97% sensitivity and 0.4% false alarm rate in low noise conditions and 92% sensitivity and 3.8% false alarm rate in high noise condition.

Table 5.1 presents a comparative analysis between our proposed adaptive threshold spike detection system and the state-of-the-art spike detection system that is targeting high

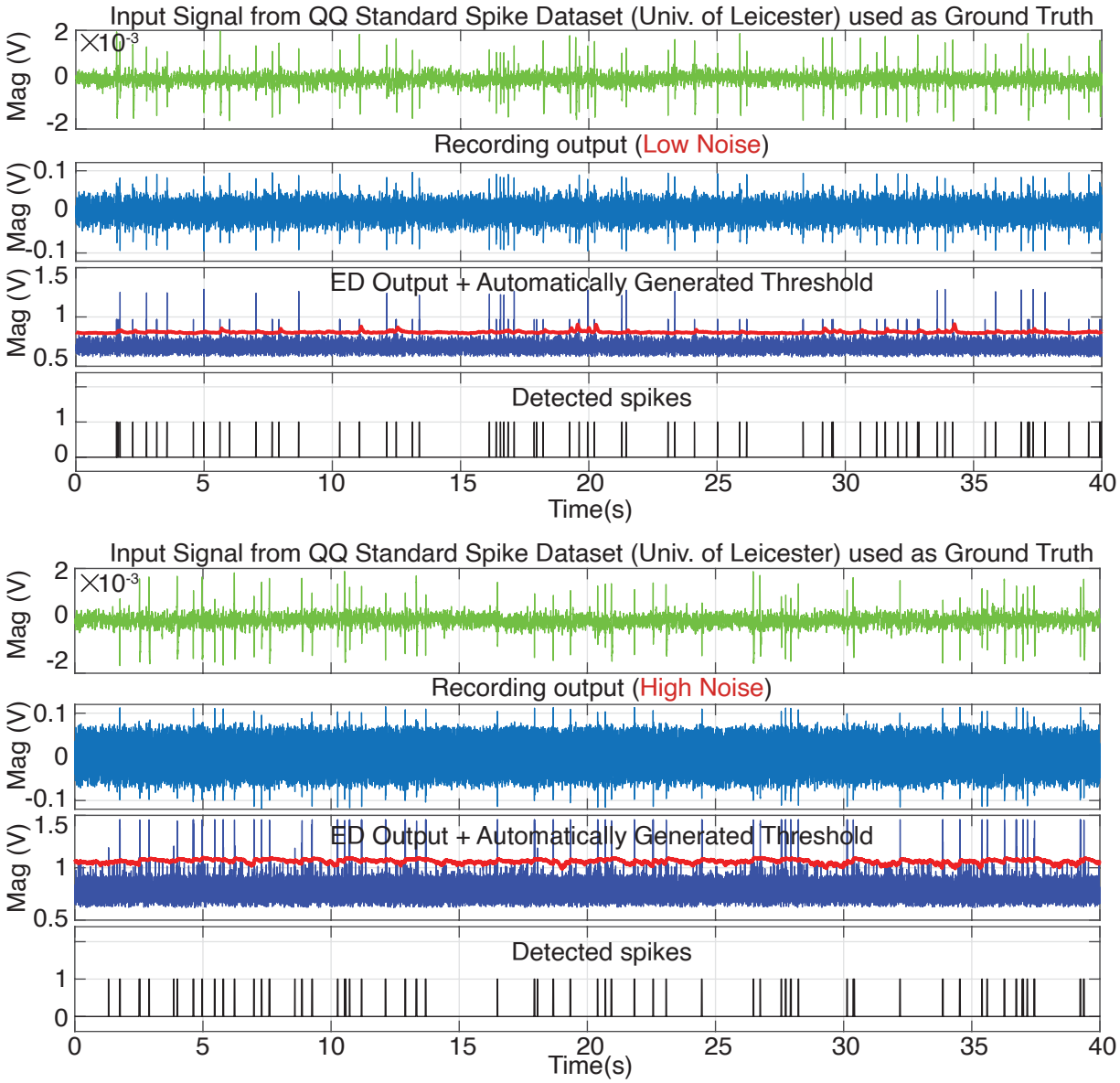


Figure 5.18: Experimental measurement results of performance evaluation of spike detection in varied noise environments using prerecorded neural data. Top panels: Low-noise scenario and Bottom panels: High-noise scenario, showcasing the efficacy of the ED operator in SNR enhancement and spike identification.

Table 5.1: Comparison between the state-of-the-art in high channel-count spike detection and our proposed adaptive threshold spike detection

Parameter	**VLSI 2021 [114]**	**This Work**
Application	Neural Implant	Retinal Prosthesis
Stimulation	Yes (Electrical)	Yes (Optical)
Spike Detection	Yes Local Digital	Yes Fully Analog
Adaptive Threshold	No	Yes
Process	65nm	180nm
Supply Voltage (V)	2.5	1.8 Recording 3.3 Stimulation
Gain (V/V)	411-2661	375
Bandwidth	300Hz-10kHz	237.5Hz-8.27kHz
In-Band IIRN	8.98 μV	8.82 μV
Recording Area/Channel (mm^2)	0.062	0.018 Recording 0.026 Spike Det
Recording Power/Channel	2.72 μW	2.7 μW
Adaptive Spike Detection Power	NR	1.02 μW

channel-count applications.

5.6 Top-Level System Integration

Up to this section in this chapter, the target was to introduce the proposed ADC-less adaptive σ-based threshold for spike detection in SNR-enhanced neural recordings. This methodology has broad range of applications and can be deployed across a spectrum of Brain-Computer Interface (BCI) applications, such as in the control of prosthetic limbs. The flexibility and general applicability of this technique make it a valuable tool for various BCI systems, where accurate and adaptive spike detection is crucial. However in this section we are analyzing the application of this circuit in the context of our retinal prosthesis and its overall system architecture.

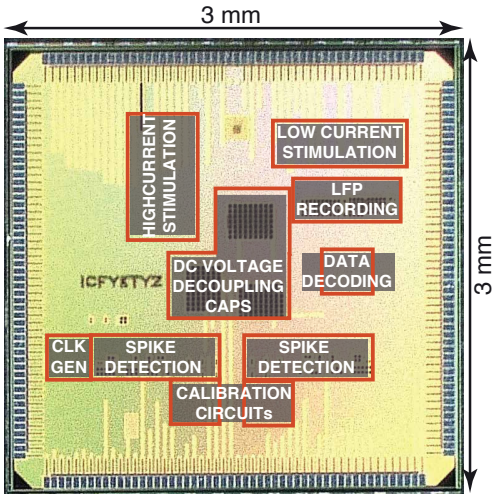


Figure 5.19: Annotated chip micrograph of the fabricated IC.

Fig. 5.19 shows the chip micrograph of the presented IC, designed and fabricated in a 180nm standard CMOS technology. This IC includes analog blocks for (1) Adaptive threshold spike detection, (2) Controlled stimulation block for a 10×10 custom-designed LED array and a 10×10 LED array designed using commercially available LEDs, and (3) LFP recording and digital blocks for (1) Input data decoding and distribution, (2) Clock generation and synchronization between input data, stimulation block and spike detection, and (3) Closed-loop calibration for monitoring the RGC activity signal to set the minimum required current for stimulation that can evoke spiking activity. The digital block are designed using Verilog hardware language, synthesized using Design Compiler, transferred to layout using Innovus, and integrated with analog block in the top level IC. Fig. 5.20 shows the annotated PCB board designed for the electrical verification and its interface with the PCB-based μLED array and a DE10 Altera FPGA board.

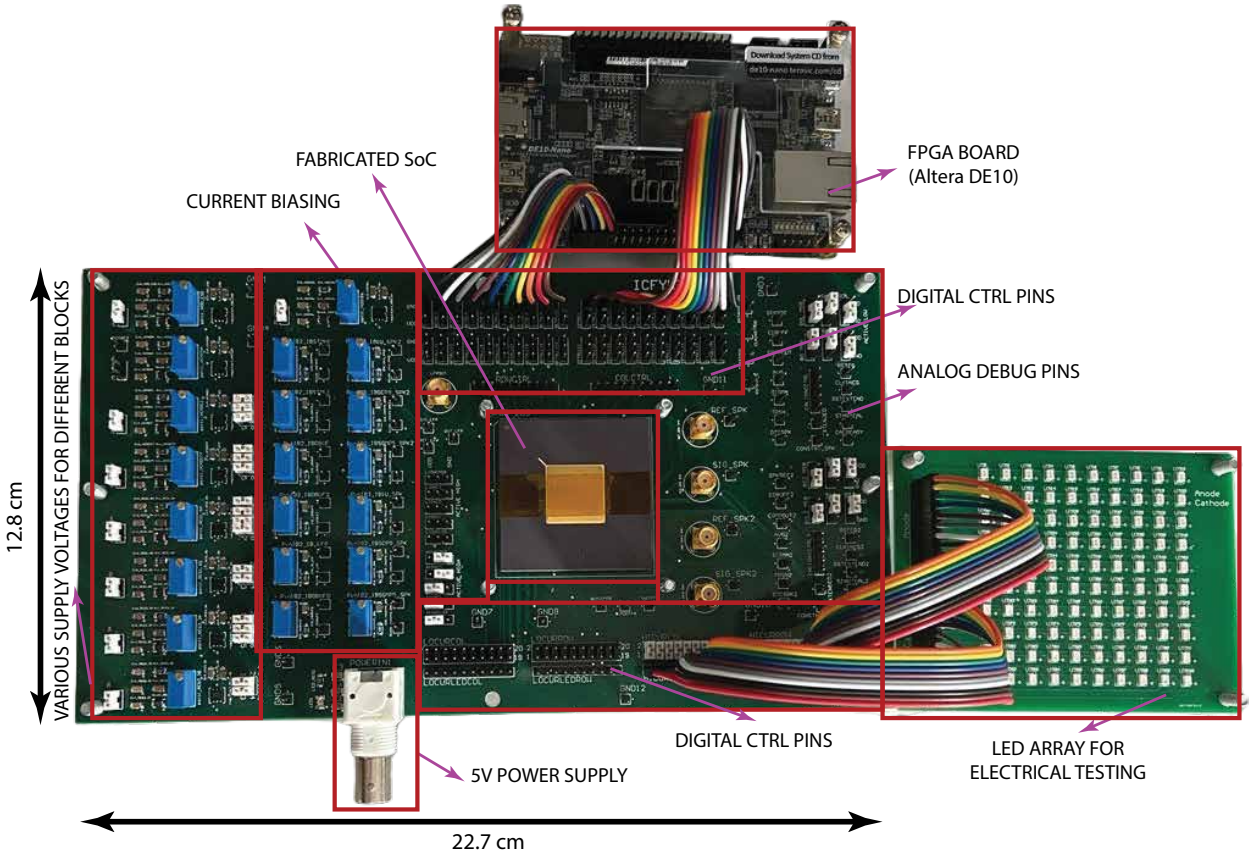


Figure 5.20: Annotated PCB board designed for the electrical verification and its interface with the PCB-based μLED array and a DE10 Altera FPGA board.

5.6.1 Closed-Loop Calibration

In this section, we turn our focus to explain the specific function of this circuit within the closed-loop framework of our retinal prosthesis. Within this architecture, the adaptive threshold spike detection circuit plays a vital role in the calibration process of our implant. In this process, through dynamic adjustment of a 5-bit current steering DAC, our system determines the minimum current that is adequate for effective opsin activation.

This calibration process is not merely a technical requirement but a response to a fundamental challenge in optogenetics(i.e., the unpredictable and often non-uniform opsin distribution within the target layer). Such variability can lead to inconsistent light sensitivities, causing an uneven quality of visual perception. Distance variations and the optical characteristics of the media between the μLED array and the opsins lead to this inconsistency, introducing a degree of unpredictability that must be mitigated for the prosthetic system to function effectively. Tissue property fluctuations and the dynamics of intraocular fluids further accentuate the need for a robust and responsive calibration protocol.

Given these considerations, the calibration of light intensity becomes an essential step in our system. It ensures that each region of the retina receives a tailored stimulation, preventing under- or over-activation of opsins and thus preserving the integrity of the visual signal relayed to the brain. By incorporating this calibration within our system, we address one of optogenetics' most significant hurdles, ensuring that each individual's unique retinal structure and the dynamic changes it may undergo are accounted for, resulting in a consistent and high-quality visual experience.

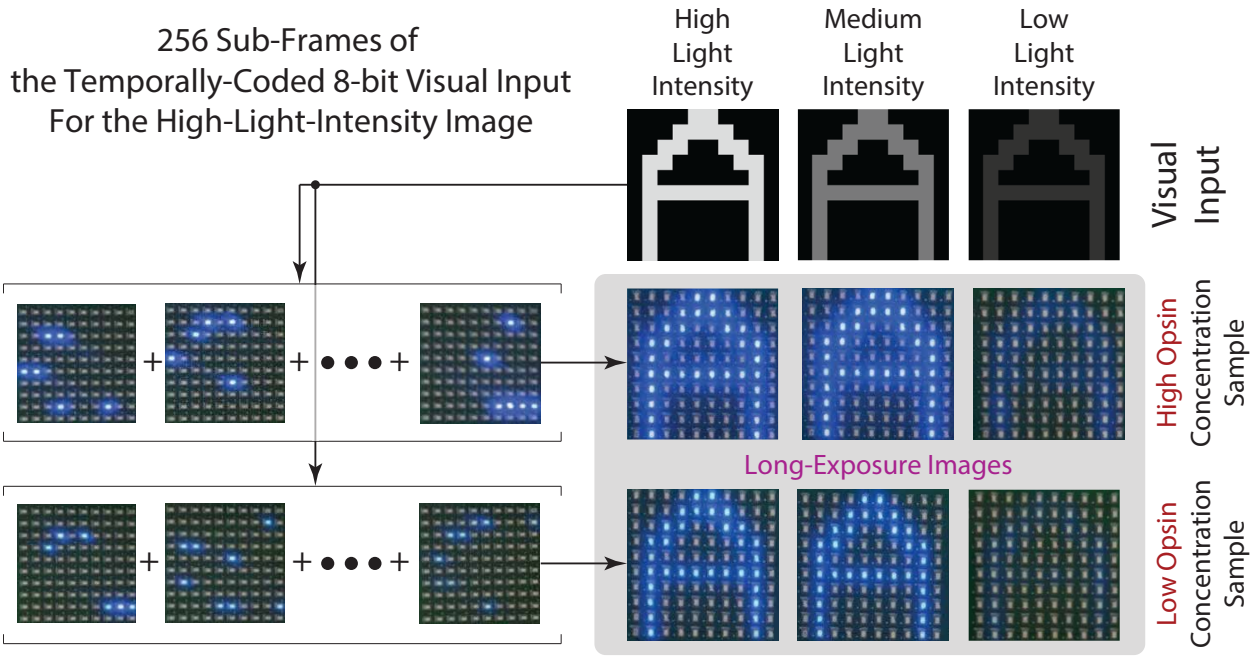


Figure 5.21: Experimental results validating the closed-loop calibration and the temporal coding strategy showing the practicality of the system in adjusting the stimulation level to opsin distribution and the effectiveness of the temporal coding in generation of the target perception.

Fig. 5.21 shows the results from an experimental validation of a closed-loop control system for stimulation intensity, tailored to the concentration of opsins within a retinal prosthesis. To enhance visibility, we developed a 10×10 array using commercial μLEDs and connected it to the IC responsible for controlling the stimulation intensity. In this experiment a synthesized recording data is fed to the calibration unit that emulates the low and high opsin concentration.

The experiment substantiates the IC's capability to discern varying opsin concentration levels based on the spikes recorded from the input. The system dynamically adjusts the light intensity, delivering higher intensity for data associated with a lower concentration of opsins, illustrated in the top row of the figure. Conversely, for data associated with a higher concentration of opsins, as shown in the bottom row, the system correspondingly reduces the

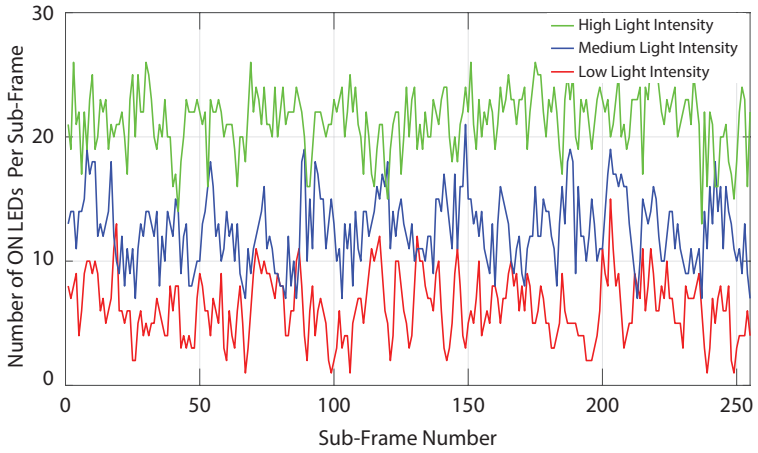


Figure 5.22: Uniform distribution of ON LEDs per sub-frame in total of 255 frames for different levels of stimulus intensities.

light intensity. This adaptability ensures that each region receives the appropriate amount of stimulation, based on the local opsin density.

This setup also validates our temporal coding scheme's effectiveness in creating the desired perception of the visual stimulus. The images on the left represent individual sub-frames within the stimulation pattern, where only a few LEDs are activated. However, when these sub-frames are temporally accumulated, as depicted in the long-exposure images on the right, the full visual stimulus is successfully reconstructed. This phenomenon demonstrates the temporal coding scheme's efficacy, where temporally-coded subframes collectively generate the entire visual input, ensuring that the stimulated perception in the brain is accurate. Additionally, the functionality of the temporal coding is validated across three distinct light intensities, verifying that less number of LEDs are turned on for lower intensity images but the distribution of the ON LEDs per frames is almost uniform across all the frames (As shown in Fig. 5.22).

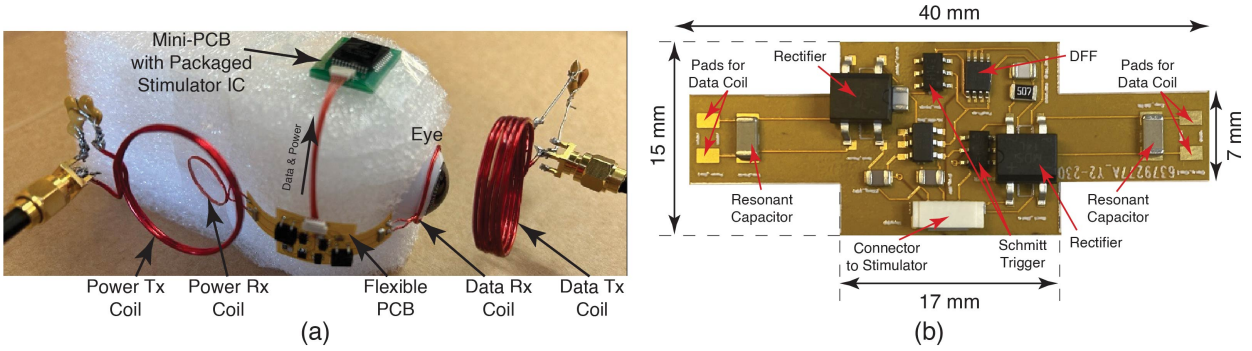


Figure 5.23: (a) Envisioned implantation of the flexible PCB used for power and data reception and its connection to the two inductive receiver coils used for power and data transfer and (b) the flexible power/data PCB with components annotated.

5.6.2 Power and Data Communication

Fig. 5.23 shows the fabricated prototype of our retinal prosthesis highlighting the wireless power and data transmission to the implantable device. The prototype employs a dual-pair coil system, distinguishing between power and data links to maintain the integrity of the power supply while accommodating a high downlink data rate for stimulation pattern control. The envisioned integration showcases two receiver coils: one positioned anteriorly relative to the eye and the other laterally, both connected to a flexible PCB. This flexible PCB is equipped with the essential components for power regulation (delivering 1.8V and 3.3V), rectification, and clock recovery associated with the power coil, alongside amplitude-shift keying (ASK) demodulation circuitry connected to the data receiver coil. Five wires (GND, 1.8V, 3.3V, Clk, and Data) extend from this flexible PCB to a secondary PCB, which houses the custom-fabricated IC and the custom-designed μLEDs array.

The coils designated for power transmission are carefully engineered in size and placement to optimize energy delivery to the implant while adhering to Specific Absorption Rate (SAR) safety standards. The data transmission component employs ASK techniques to re-

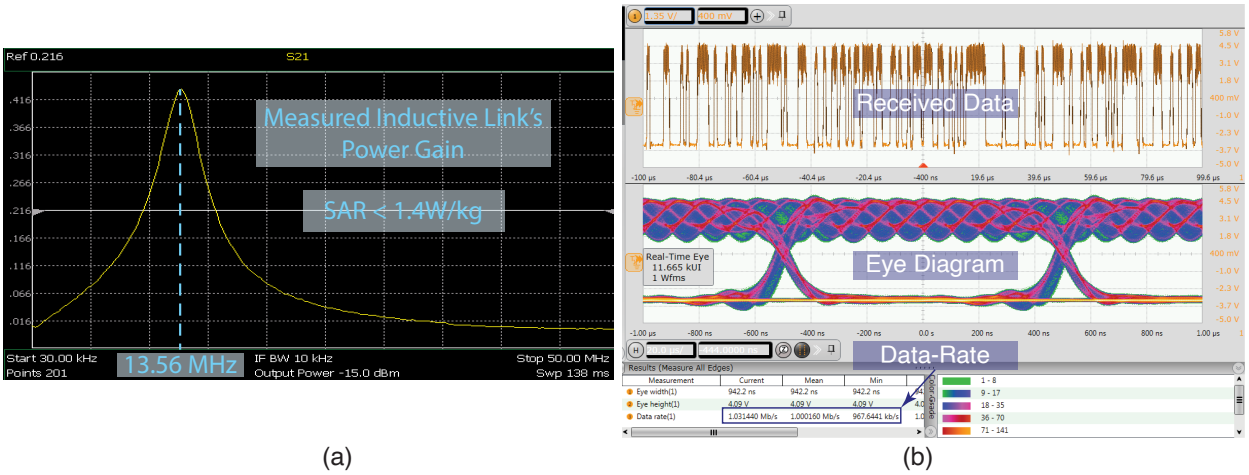


Figure 5.24: (a) The measured S21 of the inductive powering link showing the power gain and resonance at 13.56MHz while operating within SAR limits. (b) Measured Eye diagram showing example of the received data using the inductive link with the >1Mbps data-rate.

duce the power required for decoding, an essential consideration in the design of implantable medical devices.

The functionality of our system is validated by the measurement results presented in the Fig. 5.24. The S-parameter plot illustrates the power gain at the resonance frequency of 13.56MHz, operating within the SAR safety threshold while providing 10mW power to the implant at 15mm distance with 21% power transfer efficiency. The eye diagram, demonstrating data reception at rates exceeding 1Mbps, attests to the efficiency and reliability of the data link and the receiver's circuitry. Collectively, these results not only confirm the system's operational efficacy but also its adherence to safety regulations.

Table 5.2 provides a comparison of state-of-the-art retinal prosthesis systems and the system introduced in this work. Our system distinguishes itself by providing optical stimulation with an implanted source, align with progresses made in the field of optogenetic and gene therapy. Although the current system has 100 channels, scalability is one of the main goals of the system. Furthermore, our approach achieves cell-type specificity, a critical aspect

Table 5.2: Comparative Analysis of Retinal Prosthesis Systems

Feature	**[10]**	**[115]**	**[116]**	**[117]**	**This Work**
Stimulation	Electrical	Optical	Electrical	Electrical	Optical
Stimulation source	Implanted	External	Implanted	Implanted	Implanted
# channel	512	8100	1512	505	100
Placement	Epi-Retina	External	Sub-Retina	Sub-Retina	Epi-Retina
Stimulation target layer	RGC layer	RGC layer	Bipolar layer	RGC layer	Bipolar layer
Closed-loop stimulation control	No	No	No	No	Yes
Cell-type specificity	No	Yes	No	No	Yes
Inst power/Avg power	Variable	Variable	Variable	Variable	~1

for eliciting precise visual responses and a significant enhancement over other systems that lack this capability. Additionally, the proposed system, targets the bipolar layer of the retina, which is pivotal for more naturalistic visual restoration.

The table also highlights the superb power efficiency of our design. On one hand due to the proposed Poisson-coded temporal distribution of the stimulation pattern we can achieve an almost uniform ratio of instantaneous to average power consumption. On this other hand, this level of average power is also optimized due to the integration of closed-loop stimulation calibration, a feature not present in earlier designs. This control facilitates real-time adjustment of stimulation parameters, ensuring optimal activation of the target cells. Additionally, this closed-loop calibration is conducted without the need for any manual tuning thanks to our proposed fully adaptive spike detection.

Chapter 6

Conclusion, Contribution and Future Work

6.1 Contributions and Relevant Publications

As we conclude this thesis, it is imperative to synthesize the contributions made to the field of visual prosthesis and lay the groundwork for future exploration. The advent of intraocular stimulators has paved the way for new treatments of retinal degeneration, such as age-related macular degeneration. By translating the captured information by the image sensor into stimulations of the remaining retinal cells, these devices circumvent malfunctioning photoreceptors, offering a possible solution for visual restoration.

However, conventional electrical retinal stimulators present a critical limitation: a lack of cell-type specificity. This results in the simultaneous activation of both ON and OFF retinal pathways, sending conflicting signals to the brain and thus compromising visual per-

ception. This thesis introduced an innovative optogenetic-based retinal prosthesis, leveraging the specificity afforded by optogenetic techniques to selectively activate ON pathway cells, promising a more naturalistic restoration of vision. Additionally in our proposed system, the optical stimulator's less invasive epi-retinal implantation and exploitation of the retinal neural network's transparency to light further improves the performance of the prosthesis.

The first step toward our final design was development of a general-purpose optogenetic neuro-stimulator. This device was not only pivotal for our targeted retinal prosthesis but also held potential for a myriad of optogenetic applications. The pioneering concept of duty-cycled stimulation, introduced and substantiated within this chapter, has proven indispensable. It carved a path for future advancements and set a precedent for the methodologies employed in the ensuing research. This chapter has resulted in a conference paper and an extended journal paper that are published in 2021 IEEE Biomedical Circuits and Systems Conference (BioCAS) [118] and the IEEE Transactions on Biomedical Circuits and Systems (TBioCAS) [94], respectively.

In the next phase of this endeavor we addressed the challenge of identifying optimal stimulation patterns and a model to estimate the brain's inferred perception. A computational framework inspired by the retina's own processing was developed and validated, optimizing the deployment of an optogenetic epi-retinal neuro-stimulator. This framework allowed for a detailed investigation into the functional benefits of optogenetic stimulation, providing evidence that optical stimulation could indeed surpass electrical stimulation in creating a more accurate visual perception in the brain. This was further enhanced by increasing the

μLED array resolution, illuminating the path for advanced visual prosthesis. This chapter has resulted in a conference paper published in 2021 IEEE Biomedical Circuits and Systems Conference (BioCAS) [65] and its extended version is submitted to IEEE Access Journal.

Advancing into the implementation of our theoretical groundwork, the design of our retinal prosthesis system was approached with two pivotal objectives: scalability and power efficiency. In pursuit of scalability, we employed an innovative μLED array, addressed via row and column, enabling a high-density control for high resolution vision. This array's functionality was augmented by our novel temporal coding technique, rooted in the principles of photon integration and duty-cycled stimulation.

Power efficiency was further enhanced by the incorporation of Poisson coding technique. This approach allowed us to smooth out the power demand, achieving a nearly uniform ratio of instantaneous to average power consumption. Such an achievement is instrumental in relaxing the design constraints on the inductive power transmission link.

Having achieved a flat power consumption profile through our Poisson coding technique, our next objective was to ensure that the power level is not just consistent, but also as low as possible to maximize efficiency. This was accomplished through the addition of a calibration phase in our system, a step that is also crucial from biological point of view.

Primarily, this calibration phase was a strategic response to one of the fundamental challenges in optogenetics: the unpredictable and non-uniform distribution of opsins within the target layer. Such variability can result in inconsistent light sensitivities across the retina, leading to a inconsistent quality of visual perception. This is resulted from factors such as

distance and optical characteristics of the media between the μLED array and the opsins, as well as the dynamic changes in tissue properties and intraocular fluid dynamics. These factors necessitate a robust and responsive calibration protocol to ensure effective functioning of the prosthetic system. Therefore, the calibration phase was designed to determine the minimum current required for effective opsin activation, thereby optimizing power usage. Secondarily, this calibration phase served an additional critical role: it optimizes power efficiency. By determining the minimum effective current for opsin activation, the system not only adapted to the varying light sensitivities of different retinal regions but also minimized the overall power consumption.

At the core of our system's calibration process was the adaptive threshold spike detection circuit for SNR-enhanced recording data. This circuit was a key innovation of our design, enabling closed-loop calibration without the need for manual tuning. Our fully adaptive spike detection approach autonomously adjusted to the varying conditions within the retina, ensuring optimal stimulation without the need for any manual tuning. Additionally, the implementation of energy of derivative techniques equipped the system with the capability to attain high detection sensitivity and a low false alarm rate, even with significant background noise. This scenario often arises in systems with a high number of channels, where each channel operates within a constrained power budget. While this circuit has been specifically designed for our retinal prosthesis, its potential applications extend far beyond. It holds promise for a wide range of Brain-Computer Interface (BCI) applications, particularly in scenarios that require spike detection in systems with a high channel count. This

adaptability and precision position our spike detection circuit as a versatile component suitable for the next generation of neural interfacing applications. The overall implementation of our proposed retinal prosthesis system has resulted in a conference paper published in 2024 International Solid-State Circuits Conference (ISSCC) and its extended version will be submitted to IEEE Transactions on Biomedical Circuits and Systems (TBioCAS).

6.2 Contribution

The following summarizes a list of accomplished contributions in terms of conference and journal publications since the start of my PhD program (Fall 2019):

T. Yousefi, and H. Kassiri. (2024). Adaptive Threshold Calibration for Spike Detection. United States Provisional **Patent Application No. 63/556079**. Filed February 21, 2024.

T. Yousefi, and H. Kassiri, "Closed-Loop 100-Channel Highly-Scalable Retinal Implant with 1.02μW Analog ED-Based Adaptive-Threshold Spike Detection and Poisson-Coded Temporally-Distributed Optogenetic Stimulation," in *International Solid-State Circuits Conference* ***(ISSCC)***, 2024.

T. Yousefi, K. Timonina, G. Zoidl and H. Kassiri, "An Implantable Optogenetic Neuro-Stimulator SoC with Extended Optical Pulse-Width Enabled by Supply-Variation-Immune Cycled Light-Toggling Stimulation," in *IEEE Transactions on Biomedical Circuits and Systems* ***(TBioCAS)***, 2022.

T. Yousefi, K, Timonina, G.Zoidl and H. Kassiri, "A Temperature-Aware Fully-Wireless mm-Scale Optically-Enhanced Optogenetic Neuro-Stimulator," in *IEEE Biomedical Circuits and Systems Conference* ***(BioCAS)***, pp. 1-5, 2021.

T. Yousefi and H. Kassiri, "A Biologically-Informed Computational Framework for Pathway-Specific Spiking Patterns Generation and Efficacy Evaluation in Retinal Neurostimulators," in *IEEE Biomedical Circuits and Systems Conference* ***(BioCAS)***, pp. 1-5, 2021.

T. Yousefi, M.Taghadosi, A. Dabbaghian, R. Siu, G,Grau, G.Zoidl and H. Kassiri, "An Energy-Efficient Optically-Enhanced Highly-Linear Implantable Wirelessly-Powered Bidirectional Optogenetic Neuro-Stimulator," in *IEEE Transactions on Biomedical Circuits and Systems* ***(TBioCAS)***, vol. 14, no. 6, pp. 1274-1286, 2020.

T. Yousefi, A.Dabbaghian and H. Kassiri, "A 3.12 pJ°C^2 Ultra-Low-Power Direct-ADC Multi-Range Temperature Sensor for IoT Nodes," in *IEEE International Symposium on Circuits and Systems* ***(ISCAS)***, pp. 1-5, 2020.

T. Yousefi, M.Taghadosi, A. Dabbaghian, R. Siu, G,Grau, G.Zoidl and H. Kassiri, "A 12.5 mg mm-Scale Inductively-Powered Light-Directivity-Enhanced Highly-Linear Bidirectional Optogenetic Neuro-Stimulator," in *IEEE Custom Integrated Circuits Conference* ***(CICC)***, pp. 1-4, 2020.

6.3 Future Work

As we look toward the future, there are several exciting avenues for expanding upon the work presented in this thesis. In the following, we have divided the possible paths that can be taken to improve upon the current work under two categories of system- and circuit-level improvements.

6.3.1 System-Level Improvements

Advancements in Device Integration: A key aspect of our design is the μLED array, which was integrated onto a mini-PCB along with the fabricated IC. A more advanced system-level integration could involve directly placing the LED array onto the IC surface, connecting them via die bonding. This method simplifies the integration process substantially. A laser lift-off technique could be employed to separate the fabricated μLED array from its sapphire substrate, facilitating this integration without affecting the μLED fabrication process. This approach is particularly advantageous as it allows for the placement of μelectrodes directly on the IC surface, connected to the tissue with stud bumps, thereby eliminating the need for individual wire connections to the μelectrodes. Such a design would significantly streamline the assembly process and could lead to more compact, efficient, and reliable retinal prostheses.

Improved Wireless Communication Protocols: Investigating advanced wireless communication protocols that optimize power efficiency and boost data transmission rates is crucial for the next generation of medical implants. This exploration could

encompass innovative modulation techniques and the integration of wireless technologies specifically designed for biomedical applications. Our current design employs two pairs of coils for power and data transmission, setting a foundation for targeted enhancements in each domain independently. Considering that in this system, the visual information is sent to the implant via the downlink data transmission channel,as we scale the system, the necessity for transmitting at higher data rates becomes significant. To this end, the adoption of high data rate antennas emerges as a promising solution.

Tailored Optogenetic Tools: Refining optogenetic tools and vectors for precise and efficient opsin expression tailored to specific retinal cell types stands as a pivotal advancement. This customization aims to elevate the system's efficacy by ensuring more targeted stimulation of retinal cells. Up to this point, existing systems have facilitated a monochromatic visual experience for patients. However, through these enhancements, there exists the potential to develop the prosthesis system to generate colored visual perception.

6.3.2 Circuit-Level Improvements

Mode Control of the Recording Front-End: Another potential application of the proposed adaptive threshold spike detection is in the development of an event-driven recording system. Such a system could feature different operational modes for power consumption, noise level, and sampling frequency. By using the adaptive threshold to detect the events in real-time, the system could dynamically switch between these

modes, optimizing power usage and data fidelity according to the current activity level. This approach would significantly enhance the efficiency and adaptability of neural recording systems, especially in long-term monitoring applications.

Direct Connection to Spiking Neural Networks (SNNs): In this thesis, we have developed an adaptive spike detection technique for SNR-enhanced recording data. The output of this detection is a pulse waveform that encapsulates crucial information about the spikes, such as inter-spike interval, spike type, and frequency. This output is ideally suited for direct input into an SNN. An SNN can utilize this data for various processing tasks, including spike sorting or other control processes, particularly in real-time applications. This integration could open new frontiers in neural processing and interface technology, enhancing the capabilities of real-time neural monitoring and response systems.

Each of these future directions holds the potential to significantly advance the field of neuroprosthetics and brain-machine interfaces, pushing the boundaries of what is currently possible in neural engineering and providing new solutions for visual restoration and neural monitoring.

Bibliography

[1] Age-related macular degeneration: Facts & Figures [https://www.brightfocus.org/macular/article/age-related-macular-facts-figures]. Accessed 1 March 2022.

[2] W.L. Wong *et al.*, "Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis," in *The Lancet Global Health*, vol. 2, no. 2, pp. e106-e116, 2014.

[3] L. Yue *et al.*, "Retinal stimulation strategies to restore vision: Fundamentals and systems," in *Progress in retinal and eye research*, vol. 53, pp. 21-47, 2016.

[4] L. Theogarajan, "Strategies for restoring vision to the blind: current and emerging technologies," in *Neuroscience letters*, vol. 519, no. 2, pp. 129-133, 2012.

[5] B. Roska and J.A. Sahel "Restoring vision," in *Nature*, vol. 557, no. 7705, pp. 359-367, 2018.

[6] V. Wang and A.E. Kuriyan, "Optoelectronic devices for vision restoration," in *Current Ophthalmology Reports*, vol. 8, no. 2, pp. 69-7, 2020.

[7] M.E. McClements *et al.*, "Optogenetic gene therapy for the degenerate retina: recent advances. frontiers in neuroscience," in *Frontiers in Neuroscience*, vol. 14, pp. 1187-1203, 2020.

[8] M. S. Humayun *et al.*, "Visual perception in a blind subject with a chronic microelectronic retinal prosthesis," in *Vision research*, vol. 43, no. 24, pp. 2573-2581, 2003.

[9] S. Klauke *et al.*, "Stimulation with a wireless intraocular epiretinal implant elicits visual percepts in blind humans," in *Investigative ophthalmology & visual science*, vol. 52, no. 1, pp. 449-455, 2011.

[10] M. Monge *et al.*, "A fully intraocular high-density self-calibrating epiretinal prosthesis," in *IEEE transactions on biomedical circuits and systems*, vol. 7, no. 6, pp. 747-760, 2013.

[11] J. H. Park *et al.*, "1225-channel neuromorphic retinal-prosthesis SoC with localized temperature-regulation," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 6, pp. 1230-1240, 2020.

[12] Y. -K. Lo *et al.*, "A fully-integrated high-compliance voltage SoC for epi-retinal and neural prostheses," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 7, no. 6, pp. 761-772, 2013.

[13] S.K. Kelly *et al.*, "A hermetic wireless subretinal neurostimulator for vision prostheses," in *IEEE transactions on biomedical engineering*, vol. 58, no. 11, pp. 3197-3205, 2011.

[14] K. Stingl *et al.*, "Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS," in *Proceedings of the Royal Society B: Biological Sciences*, vol. 280, no. 1757, pp. 20130077, 2013.

[15] A. Asher *et al.*, "Image processing for a high-resolution optoelectronic retinal prosthesis," in *IEEE transactions on Biomedical Engineering,*, vol. 54, no. 6, pp. 993-1004, 2007.

[16] Z. Yu *et al.*, "Toward the next generation of retinal neuroprosthesis: visual computation with spikes," in *Engineering,*, vol. 6, no. 4, pp. 449-461, 2020.

[17] S. Nirenberg *et al.*, “Retinal prosthetic strategy with the capacity to restore normal vision," in *Proceedings of the National Academy of Sciences*, vol. 109, no. 37, pp. 15012-15017, 2012.

[18] A. M. Litke *et al.*, “ What does the eye tell the brain?: Development of a system for the large-scale recording of retinal output activity," in *IEEE Transactions on Nuclear Science*, vol. 51, no. 4, pp. 1434-1440, 2004.

[19] A. Rothermel *et al.*, “A 1600-pixel subretinal chip with DC-free terminals and ±2V supply optimized for long lifetime and high stimulation efficiency,” in *IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 144-602. Feb. 2008.

[20] D. -H. Choi *et al.*, “A 4.49nW/pixel light-to-stimulus duration converter-based retinal prosthesis chip,” in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 6, pp. 1140-1148, 2021.

[21] N.L. Opie *et al.*,“Heating of the eye by a retinal prosthesis: modeling, cadaver and in vivo study," in *IEEE transactions on biomedical engineering*, vol. 59, no. 2, pp. 339-345, 2011.

[22] B. Abbasi and J.F. Rizzo, “Advances in neuroscience, not devices, will determine the effectiveness of visual prostheses,” in *Seminars in Ophthalmology*, vol. 12, no. 3, pp. 1-8, 2021.

[23] J.R. Golden *et al.*, “Simulation of visual perception and learning with a retinal prosthesis,” in *journal of Neural Engineering*, vol. 16, no. 2, pp. 025003, 2019.

[24] R. Iezzi *et al.*, “Neurotransmitter-based retinal prosthesis modulation of retinal ganglion cell responses in-vivo,” in *Investigative Ophthalmology and Visual Science*, vol. 44, no. 13, pp. 5083-5083, 2003.

[25] D.G. Muratore *et al.*, "A Data-compressive wired-or readout for massively parallel neural recording," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1128-1140, 2019.

[26] S. Maity *et al.*, "On the safety of human body communication," in *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 12, pp. 3392-3402, 2020.

[27] M. Kiani and M. Ghovanloo, "A 13.56-Mbps pulse delay modulation based transceiver for simultaneous near-field data and power transmission," in *IEEE transactions on Biomedical Engineering,*, vol. 9, no. 1, pp. 1-11, 2014.

[28] H. Cho *et al.*, "A 79 pJ/b 80 Mb/s full-duplex transceiver and a 42.5μW 100 kb/s super-regenerative transceiver for body channel communication," in *IEEE Journal of Solid-State Circuits,*, vol. 51, no. 1, pp. 310-317, 2016.

[29] Y.P. Zhang, and T.G. Oertner. "Optical induction of synaptic plasticity using a light sensitive channel," in *Nature Methods*, vol. 4, no. 2, pp. 139-141, 2007.

[30] P. Schoenenberger *et al.*, "Optimizing the spatial resolution of Channelrhodopsin-2 activation," in *Brain cell biology*, vol. 36, no. 1-4, pp. 119, 2008.

[31] R.C. Wykes *et al.*, "Optogenetic and potassium channel gene therapy in a rodent model of focal neocortical epilepsy," in *Science translational medicine*, vol. 4, no. 161, pp. 152-161, 2012.

[32] E. Krook-Magnuson *et al.*, "On-demand optogenetic control of spontaneous seizures in temporal lobe epilepsy," in *Nature communications*, vol. 4, no. 1, pp. 1-8, 2013.

[33] J.A. Steinbeck *et al.*, "Optogenetics enables functional analysis of human embryonic stem cell-derived grafts in a Parkinson's disease model," in *Nature biotechnology*, vol. 33, no. 2, pp. 204-209, 2015.

[34] A. Guru *et al.*, "Making sense of optogenetics," in *International Journal of Neuropsychopharmacology*, vol. 18, 2015.

[35] K. Paralikar *et al.*, "An implantable optical stimulation delivery system for actuating an excitable biosubstrate," in *IEEE Journal of Solid-State Circuits*, vol. 46, no. 1, pp. 321-332, Jan. 2011.

[36] I. Diester *et al.*, "An optogenetic toolbox designed for primates," in *Nature neuroscience*, vol. 14, no. 3, pp. 387-397, 2011.

[37] G. Gagnon-Turcotte *et al.*, "A 13μm CMOS SoC for simultaneous multichannel optogenetics and electrophysiological brain recording," in *IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 466-468. Feb. 2018.

[38] S. Nag *et al.*, "Multi-function optogenetic stimulator and neural amplifier for wirelessly controlled neural interface," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Atlanta, GA, pp. 1-4, 2015.

[39] H. Kassiri *et al.*, "Arbitrary-waveform electro-optical intracranial neurostimulator with load-adaptive high-voltage compliance," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 582-593, Aug. 2019.

[40] K.L. Montgomery *et al.*, "Wirelessly Powered, Fully Internal Optogenetics for Brain, Spinal and Peripheral Circuits in Mice," in *Nature Methods*, vol. 12, no. 10, pp. 969-974, Aug. 2015.

[41] T. Kim *et al.*, "Injectable, cellular-scale optoelectronics with applications for wireless optogenetics," in *Science*, vol. 340, no. 6129, pp. 211-216, 2013.

[42] Y. Jia *et al.*, "A mm-sized free-floating wirelessly powered implantable optical stimulating system-on-a-chip," in *IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 468-469. Feb. 2018.

[43] R. Ramezani *et al.*, "On-probe neural interface ASIC for combined electrical recording and optogenetic stimulation," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 3, pp. 576-588, June. 2018.

[44] A. Berndt, O. Yizhar, L.A. Gunaydin, P. Hegemann, and K. Deisseroth, "Bi-stable neural state switches," in *Nature neuroscience*, vol. 12, no. 2, pp. 229-234, 2009.

[45] O. Yizhar *et al.*, "Neocortical excitation/inhibition balance in information processing and social dysfunction," in *Nature*, vol. 477, no. 7363, pp. 171-178, 2011.

[46] X. Gong *et al.*, "ultra-sensitive step-function opsin for minimally invasive optogenetic stimulation in mice and macaques," in *Neuron*, vol. 107, no. 1, pp. 38-51, 2020.

[47] J. Mattis *et al.*, "Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins." in *Nature methods*, vol. 9, no. 2, pp. 159-172, 2012.

[48] S. Owen *et al.*, "Thermal constraints on in vivo optogenetic manipulations," in *Nature neuroscience*, vol. 22, no. 7, pp. 1061-1065, 2019.

[49] O. Yizhar *et al.*, "Optogenetics in neural systems," in *Neuron*, vol. 71, no. 1, pp. 9-34, 2011.

[50] M. Taghadosi and H. Kassiri, "A real-time-link-adaptive operation scheme for maximum energy storage efficiency in resonant CM wireless power receivers," in *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 1, pp. 510-523, 2021.

[51] M. Taghadosi and H. Kassiri, "A calibration-free energy-efficient IC for link-adaptive real-time energy storage optimization of CM inductive power receivers," in *IEEE Journal of Solid-State Circuits*, vol. 57, no. 3, pp. 793-802, 2022.

[52] y. Yang, X.A. Cao, C. Yan, "Investigation of the non-thermal mechanism of efficiency rolloff in InGaN light-emitting diodes," in *IEEE transactions on electron devices*, vol. 55, no. 7, pp. 1771-1775, 2008.

[53] T. Egawa, B. Zhang, H. Ishikawa, "High performance of InGaN LEDs on (111) silicon substrates grown by MOCVD," in *IEEE Electron Device Letters*, vol. 26, no. 3, pp. 169-171, 2005.

[54] H. Kassiri *et al.*, "Closed-loop neurostimulators: A survey and a seizure-predicting design example for intractable epilepsy treatment," in *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 5, pp. 1026-1040, 2017.

[55] M.A. Ghanad, M.M. Green, C. Dehollain, "A 30 μW remotely powered local temperature monitoring implantable system," in *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 1, pp. 54-63, 2017.

[56] K. Souri, K.A. Makinwa, "A 0.12 mm^2 7.4μW micropower temperature sensor with an inaccuracy of $\pm$0.2°C (3σ) trom $-$30°C to 125°C," in *IEEE Journal of Solid-State Circuits*, vol. 46, no. 7, pp. 1693-1700, 2011.

[57] H. Kassiri *et al.*, "Battery-less tri-band-radio neuro-monitor and responsive neurostimulator for diagnostics and treatment of neurological disorders," in *IEEE Journal of Solid-State Circuits*, vol. 51, no. 5, pp. 1274-1289, May 2016.

[58] M. Zargham and P. G. Gulak, "Fully integrated on-chip coil in 0.13 μm CMOS for wireless power transfer through biological media," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 2, pp. 259-271, April 2015.

[59] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues," in *Physics in Medicine and Biology*, vol. 41, no. 11, p. 2271, 1996.

[60] Z. Tang *et al.*, "Data transmission from an implantable biotelemeter by load-shift keying using circuit configuration modulator," in *IEEE Transactions on Biomedical Engineering*, vol. 42, no. 5, pp. 524-528, 1995.

[61] C.H. Cheng *et al.*, "A fully integrated 16-channel closed-loop neural-prosthetic CMOS SoC with wireless power and bidirectional data telemetry for real-time efficient human epileptic seizure control," in *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3314-3326, 2018.

[62] T. Yousefi *et al.*,"An energy-efficient optically-enhanced highly-linear implantable wirelessly-powered bidirectional optogenetic neuro-stimulator," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 6, pp. 1274-1286, 2020.

[63] D.D. Bezshlyakh *et al.*, "Directly addressable GaN-based nano-LED arrays: fabrication and electro-optical characterization," in *Microsystems & Nanoengineering*, vol. 6, no. 1, pp. 1-10, 2020.

[64] M. Asad *et al.*, "Thermal and optical properties of high-density GaN micro-LED arrays on flexible substrates," in *Nano Energy*, vol. 73, pp. 104724, 2020.

[65] T. Yousefi and H. Kassiri, "A biologically-informed computational framework for pathway-specific spiking patterns generation and efficacy evaluation in retinal neurostimulators," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1-5, 2021.

[66] E. Richard *et al.*, "Recognizing retinal ganglion cells in the dark," in *Advances in Neural Information Processing Systems*, vol. 28, pp. 2476-2484, 2015.

[67] C. Sekirnjak *et al.*, "High-resolution electrical stimulation of primate retina for epiretinal implant design," in *Journal of Neuroscience*, vol. 28, no. 17, pp. 4446-4456, 2008.

[68] L.H. Jepson *et al.*, "Focal electrical stimulation of major ganglion cell types in the primate retina for the design of visual prostheses," in *Journal of Neuroscience*, vol. 33, no. 17, pp. 7194-7205, 2013.

[69] V.H. Fan *et al.*, "Epiretinal stimulation with local returns enhances selectivity at cellular resolution," in *Journal of Neural Engineering*, vol. 16, no. 2, pp. 025001, 2019.

[70] A. Berndt *et al.*, "High-efficiency channelrhodopsins for fast neuronal stimulation at low light levels," in *Proceedings of the National Academy of Sciences*, vol. 108, no. 18, pp. 7595-7600, 2020.

[71] Z. H. Pan *et al.*, "Optogenetic approaches to restoring vision," in *Annual review of vision science*, vol. 1, pp. 185-210, 2015.

[72] S. D. Klapper *et al.*, "Biophysical properties of optogenetic tools and their application for vision restoration approaches," in *Frontiers in Systems Neuroscience*, vol. 10, pp. 74, 2016.

[73] N.P. Cottaris *et al.*, "A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine," in *Journal of vision*, vol. 19, no. 4, pp. 8-8, 2019.

[74] J.W. Pillow *et al.*, "Spatio-temporal correlations and visual signaling in a complete neuronal population," in *Nature*, vol. 454, no. 7207, pp. 995-999, 2008.

[75] N.C. Klapoetke *et al.*, "Independent optical excitation of distinct neural populations," in *Nature methods*, vol. 11, no. 3, pp. 338-346, 2014.

[76] Z. Wang *et al.*, "Image quality assessment: from error visibility to structural similarity," in *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600-612, 2004.

[77] L. E. Grosberg *et al.*, "Activation of ganglion cells and axon bundles using epiretinal electrical stimulation," in *Journal of neurophysiology*, vol. 118, no. 3, pp. 1457-1471, 2017.

[78] T. Yousefi *et al.*, "A 12.5 mg mm-scale inductively-powered light-directivity-enhanced highly-linear bidirectional optogenetic neuro-stimulator," in *IEEE Custom Integrated Circuits Conference (CICC)*, pp. 1-4, 2020.

[79] S.A. Baccus , "Timing and computation in inner retinal circuitry," in *Annu. Rev. Physiol.*, vol. 69, pp. 271-290, 2007.

[80] M. Vidne *et al.*, "Modeling the impact of common noise inputs on the network activity of retinal ganglion cells," in *Journal of computational neuroscience*, vol. 33, no. 1, pp. 97-121, 2012.

[81] N. Grossman *et al.*, "Modeling study of the light stimulation of a neuron cell with channelrhodopsin-2 mutants," in *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 6, pp. 1742-1751, 2011.

[82] C. M. Kang *et al.*, "Fabrication of a vertically-stacked passive-matrix micro-LED array structure for a dual color display," in *Optics Express*, vol. 25, no. 3, pp. 2489-2495, 2017.

[83] S. Zhao *et al.*, "Molecular beam epitaxy of III-nitride nanowires: Emerging applications from deep-ultraviolet light emitters and micro-LEDs to artificial photosynthesis," in *IEEE Nanotechnology Magazine*, vol. 13, no. 2, pp. 6-16, 2019.

[84] F. Soto *et al.*, "Efficient coding by midget and parasol ganglion cells in the human retina," in *Neuron*, vol. 107, no. 4, pp. 656-666, 2020.

[85] K. Chen *et al.*, "An integrated 256-channel epiretinal prosthesis," in *IEEE Journal of Solid-State Circuits*, vol. 45, no. 9, pp. 1946-1956, 2010.

[86] M. Ortmanns *et al.*, "A 232-channel epiretinal stimulator ASIC," in *IEEE Journal of Solid-State Circuits*, vol. 42, no. 12, pp. 2946-2959, 2007.

[87] D. Kerschensteiner, "Feature detection by retinal ganglion cells," in *Annual review of vision science*, vol. 8, pp. 135-169, 2022.

[88] D.D. Bezshlyakh *et al.*, "Directly addressable GaN-based nano-LED arrays: fabrication and electro-optical characterization," in *Journal of Microsystems & Nanoengineering*, vol. 6, no. 1, pp. 1-10, 2020.

[89] A. Guru *et al.*, "Making Sense of Optogenetics," in *International Journal of Neuropsychopharmacology*, vol. 18, 2015.

[90] A. Berndt, O. Yizhar, L.A. Gunaydin, P. Hegemann, and K. Deisseroth, "Bi-stable neural state switches," in *Nature neuroscience*, vol. 12, no. 2, pp. 229-234, 2009.

[91] O. Yizhar *et al.*, "Neocortical excitation/inhibition balance in information processing and social dysfunction," in *Nature*, vol. 477, no. 7363, pp. 171-178, 2011.

[92] X. Gong *et al.*, "ultra-sensitive step-function opsin for minimally invasive optogenetic stimulation in mice and macaques," in *Neuron*, vol. 107, no. 1, pp. 38-51, 2020.

[93] J. Mattis *et al.*, "Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins." in *Nature methods*, vol. 9, no. 2, pp. 159-172, 2012.

[94] T. Yousefi *et al.*, "An implantable optogenetic neuro-stimulator SoC with extended optical pulse-width enabled by supply-variation-immune cycled light-toggling stimulation," in *IEEE Transactions on Biomedical Circuits and Systems*, 2022.

[95] S. Owen et al., "Thermal constraints on in vivo optogenetic manipulations," in *Nature neuroscience*, vol. 22, no. 7, pp. 1061-1065, 2019.

[96] M.A. Ghanad, M.M. Green, C. Dehollain, "A 30 μW remotely powered local temperature monitoring implantable system," in *IEEE transactions on biomedical circuits and systems*, vol. 11, no. 1, pp. 54-63, 2017.

[97] K. Souri, K.A. Makinwa, "A 0.12 mm^2 7.4μW micropower temperature sensor with an inaccuracy of $\pm$0.2°C (3σ) from $-$30°C to 125°C," in *IEEE Journal of Solid-State Circuits*, vol. 46, no. 7, pp. 1693-1700, 2011.

[98] M. Castro, D.M. Ballesteros, D. Renza "CG-1050: Original and tampered images (Color and grayscale)" in *Mendeley Data*, 2019.

[99] U. Shin et al., "NeuralTree: A 256-Channel 0.227-μJ/Class versatile neural activity classification and closed-loop neuromodulation SoC," in *IEEE Journal of Solid-State Circuits*, vol. 57, no. 11, pp. 3243-3257, 2022.

[100] B. Zhu, U. Shin and M. Shoaran, "Closed-loop neural prostheses with on-chip intelligence: a review and a low-latency machine learning model for brain state detection," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 5, pp. 877-897, 2021.

[101] P.T. Watkins, "Validation of adaptive threshold spike detector for neural recording," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 4079-4082, 2004.

[102] R.R. Harrison, "The design of integrated circuits to observe brain activity," in *Proceedings of the IEEE*, vol. 96, no. 7, pp. 1203-1216, 2008.

[103] U. Shin et al., "Adaptive threshold neural spike detector using stationary wavelet transform in CMOS," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 946-955, 2015.

[104] E. Yao, Y. Chen and A. Basu, "A 0.7V, 40nW compact, current-mode neural spike detector in 65 nm CMOS," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 2, pp. 309-318, 2015.

[105] J.H. Choi et al., "A new action potential detector using the MTEO and its effects on spike sorting systems at low signal-to-noise ratios," in *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 4, pp. 738-746, 2006.

[106] H. Semmaoui et al., "Setting adaptive spike detection threshold for smoothed TEO based on robust statistics theory," in *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 2, pp. 474-482, 2011.

[107] R. Fiorelli et al., "Charge-redistribution based quadratic operators for neural feature extraction," in *IEEE Transactions on Biomedical Engineering*, vol. 14, no. 3, pp. 606-619, 2020.

[108] Y. Yang and A. J. Mason, "Hardware efficient automatic thresholding for NEO-based neural spike detection," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 4, pp. 826-833, 2017.

[109] K.H. Kim, S.J. Kim, "Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier," in *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 10, pp. 1406-1411, 2000.

[110] S. Dwivedi and A. K. Gogoi, "A novel adaptive real-time detection algorithm for an area-efficient CMOS spike detector circuit," in *International Journal of Electronics and Communications*, vol. 88, pp. 87-97, 2018.

[111] E. Koutsos et al., "A 1.5 μW NEO-based spike detector with adaptive-threshold for calibration-free multichannel neural interfaces," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1922-1925, 2013.

[112] R.R. Harrison and C. charles "A wearable 8-channel active-electrode EEG/ETI acquisition system for body area networks," in *IEEE Journal of Solid-State Circuits*, vol. 38, no. 6, pp. 958-965, 2003.

[113] Quiroga, Rodrigo Quian (2020). Simulated dataset. University of Leicester. Dataset. https://doi.org/10.25392/leicester.data.11897595.v1

[114] D. Y. Yoon et al., "A 1024-channel simultaneous recording neural SoC with stimulation and real-time spike detection," in *2021 Symposium on VLSI Circuits, Kyoto, Japan*, pp. 1-2, 2021.

[115] A. Soltan et al., "High density, high radiance μLED matrix for optogenetic retinal prostheses and planar neural stimulation," in *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 2, pp. 347-359, 2017.

[116] A. Akinin et al., "An optically-addressed nanowire-based retinal prosthesis with 73% RF-to-stimulation power efficiency and 20nC-to-3μC wireless charge telemetering," in *IEEE International Solid- State Circuits Conference (ISSCC), San Francisco*, pp. 276-278, 2021.

[117] K. Eom et al., "A stimulus-scattering-free pixel-sharing sub-retinal prosthesis SoC with 35.8dB dynamic range time-based photodiode sensing and per-pixel dynamic voltage scaling," in *IEEE International Solid- State Circuits Conference (ISSCC), San Francisco*, pp. 480-482, 2023.

[118] T. Yousefi *et al.*, "A temperature-aware fully-wireless mm-scale optically-enhanced optogenetic neuro-stimulator," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pp. 1-5, 2021.