

Simulating and optimizing electrical current flow and neuronal activation in retinal implants

Timothy Esler

May, 2018

Submitted in total fulfillment of
the requirements of the degree of
Doctor of Philosophy

Department of Electrical & Electronic Engineering
The University of Melbourne

Abstract

A number of challenges facing electrical stimulation of the retina necessitate more sophisticated methods of stimulation in order to achieve targeted and meaningful perception. Computational models of current flow and neural activation in complex retinal tissue provide a means for the exploration of novel stimulation strategies and the development of stimulus optimization techniques.

The first component of this project develops multi-layered biophysical models of the retina that capture the characteristic anatomical properties of each retinal layer. Models used in this research build on single-layered models of electrical stimulation that include the microscopic structure and physiology of neural tissue. A novel four-layer model is presented that models current flow and passive membrane activation due to simultaneous multi-electrode stimulation, taking into account the organized structure of parallel fibers in the nerve fiber layer. Using this model, simulations are used to elucidate strategies for the preferential activation of the target ganglion cell layer over the nerve fiber layer in epiretinal stimulation. These strategies indicate that preferential activation of retinal ganglion cells over passing axons is achievable given carefully chosen electrode locations and stimulation waveforms.

The project then focuses on the validation of simpler linear-nonlinear models of retinal ganglion cell activity for the purpose of optimizing multi-electrode array stimulation parameters. Models of spiking probability are compared and validated against the developed multi-layered biophysical model of retinal stimulation and current flow in order to establish a biophysical basis for the simplified model's structure. Results demonstrate that the linear electrical receptive field of the linear-nonlinear model matches the transmembrane currents induced by electrodes, known as the activating function, at the site of action potential initiation. Importantly, this result allows for the typically experimentally-determined electrical receptive field of the retinal ganglion cell to be estimated using a biophysical model, yielding a biophysical version of linear-nonlinear model for biphasic stimulation.

Finally, a method is demonstrated for the calculation of optimal stimulus currents given desired patterns of retinal ganglion cell activation. The four-layer biophysical model is used to simulate the electrical receptive field of retinal ganglion cells across the retina. Combined with the linear-nonlinear model, this allows for a high-resolution description of activation in response to arbitrary multi-electrode stimulation. This model is shown to be equivalent to a specific type of feed-forward artificial neural network. The fitting of this model to target stimulation patterns (images) using state-of-the-art optimization algorithms is shown to produce stimulus currents that achieve an optimal pattern of activation across the retina.

Declaration of authorship

I declare that this thesis is comprised of original work completed by me, the author. All results presented in this thesis are the output of research conducted by the author. Unless acknowledged below, all data collection, analysis, investigation, software programming, creation of reports (such as this), visualization, and figure generation were completed in total by the author.

The following researchers are acknowledged for their contributions:

- Anthony Burkitt, David Grayden, Robert Kerr, and Hamish Meffin aided in the conceptualization, review and discussion of each study.
- Bahman Tahayori aided in technical troubleshooting and review of mathematical models used in Chapter 2.
- Matias Maturana aided in the review of Chapter 3, and supplied experimental data used in comparisons with simulation results.

This thesis is fewer than the maximum word limit of 100,000 words.

Signed: Timothy Esler

6th May, 2018

Acknowledgements

I would like to thank my supervisors Anthony Burkitt, David Grayden, Robert Kerr, and Hamish Meffin for their guidance, support and encouragement, and for all the time they have dedicated to this project. Thank you for continuing to test my thinking, pushing me to improve, and letting me work independently.

I would like to thank Angus Johnston, Georgina Such, and David Grayden for giving me the opportunity to work in their labs during my Masters degree, giving me an insight into research. To the lecturers in the Department of Biomedical Engineering, particularly David Grayden and Leigh Johnston, thank you for inspiring my interest in neural engineering, and for just being excellent teachers.

Thank you to Matias Maturana and Bahman Tahayori for their help in conceptualization, debugging code and maths, and reviewing the work that went into Chapters 2 and 3.

My research and the travel it included was made possible by the financial support I received from an Australian Government Research Training Program Scholarship, the Gowrie Scholarship Fund, and an Australian Research Council Discovery Project. I would also like to thank the members of my advisory panel Leigh Johnston and Marcus Brazil for their time.

I would like to thank my colleagues and friends in the NeuroEngineering Laboratory for discussion and camaraderie, including the esteemed members of the Dream Team, Ewan Nurse and Pip Karoly, and those who made Friday afternoon drinks something to look forward to: Kerry, Shannon, Rob, and Matias.

I would like to thank IBM Research, Australia, and John Wagner and Rob Kerr, for giving me the opportunity to complete an internship in the first year of my PhD. To Alex Feigin and Andrew Coulthurst, thank you for giving me the flexibility to continue working throughout my PhD.

Thank you to my non-PhD friends for providing necessary distractions and good times: Jessica, Bob, Claire, Rob, Archie, Tom, and Luke.

Finally, I would like to thank my family, Mum, Dad, Isaac, Ben, and Matt, for their continuing support and encouragement in all things I do.

Publications

The following chapters of this thesis have been published as peer-reviewed journal papers:

- **Chapter 2:** Timothy B. Esler, Robert R. Kerr, Bahman Tahayori, David B. Grayden, Hamish Meffin, Anthony N. Burkitt. (2018) “Minimizing activation of overlying axons with epiretinal stimulation: The role of fiber orientation and electrode configuration.” PLOS ONE 13(3): e0193598.
- **Chapter 3:** Timothy B. Esler, Matias I. Maturana, Robert R. Kerr, David B. Grayden, Anthony N. Burkitt, Hamish Meffin. “Biophysical basis of the linear electrical receptive fields of retinal ganglion cells.” J. Neural Eng. 15(5): 055001.

Parts of the following chapters of this thesis have been presented at peer-reviewed conferences as abstracts or conference papers:

- **Chapter 2:** Timothy B. Esler, Anthony N. Burkitt, David B. Grayden, Robert R. Kerr, Bahman Tahayori, Hamish Meffin. “Modeling orientation-selective electrical stimulation with retinal prostheses”. CNS*2016 Computational Neuroscience Meeting, BMC Neuroscience 17(Suppl 1), P97, Jeju, South Korea, July 2016.
- **Chapter 2:** Timothy B. Esler, Anthony N. Burkitt, David B. Grayden, Robert R. Kerr, Bahman Tahayori, Hamish Meffin. “A computational model of orientation-dependent activation of retinal ganglion cells”. EMBC 2016 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Orlando, USA, August 2016.
- **Chapter 2:** Timothy B. Esler, Anthony N. Burkitt, David B. Grayden, Robert R. Kerr, Bahman Tahayori, Hamish Meffin. “The effect of inner retina fiber orientation on preferential activation of retinal ganglion cells”. 4th International Conference on Medical Bionics, Brisbane, Australia, November 2016.
- **Chapter 3:** Timothy B. Esler, Matias I. Maturana, Robert R. Kerr, David B. Grayden, Anthony N. Burkitt, Hamish Meffin. “The biophysical basis for the linear electrical receptive field of the retinal ganglion cell”. NeuroEng 2016: 9th Australasian Workshop on Neuro-Engineering and Computational Neuroscience, Brisbane, Australia, November 2016.
- **Chapter 3:** Timothy B. Esler, Matias I. Maturana, Robert R. Kerr, David B. Grayden, Anthony N. Burkitt, Hamish Meffin. “The biophysical basis for the linear electrical receptive field of the retinal ganglion cell”. 10th Eye and The Chip World Congress, Detroit, USA, September 2017.

Contents

Abstract	iii
Declaration of authorship	iv
Acknowledgements	v
Publications	vi
Contents	x
List of Figures	xii
List of Tables	xiii
List of Abbreviations	xiv
1 Introduction	1
1.1 Retinal prostheses	1
1.1.1 Retinal anatomy	1
1.1.2 Degenerative outer retinal diseases	3
1.1.3 Prosthesis types	3
1.1.4 Clinical outcomes to-date	5
1.1.5 Existing challenges	6
1.2 Biophysical modeling of retinal stimulation	8
1.2.1 Volume conductor models	9
1.2.2 Limitations of standard volume conductor models	10
1.2.3 Simulating morphological reconstructions	11
1.2.4 Existing models	12
1.2.5 The cellular composite model	13
1.3 Phenomenological modeling of retinal stimulation	18
1.3.1 Block-structured models/cascaded filter models	18
1.3.2 The linear-nonlinear model of extracellular stimulation	19
1.3.3 Underlying assumptions	23
1.4 Optimization of multi-electrode stimulus currents	23
1.4.1 Linear programming approaches to spatial shaping	24
1.4.2 Spatial shaping using the linear-nonlinear model	26

1.5	Conclusion	27
1.6	Thesis overview	27
2	Minimizing activation of overlying axons	30
2.1	Abstract	30
2.2	Introduction	31
2.3	Methods	33
2.3.1	Distribution of orientations in the ganglion cell layer	33
2.3.2	Tissue geometry and governing equations	34
2.3.3	Neurite equations	38
2.3.4	Calculating membrane potential thresholds	39
2.3.5	Simulation Methods	40
2.4	Results	41
2.4.1	Analysis of tissue anisotropy	41
2.4.2	Comparison of one-, two-, and four-electrode configurations	44
2.4.3	Effect of pulse duration and electrode-retina separation	44
2.4.4	Sensitivity of results to threshold membrane potential	47
2.4.5	Performance of simultaneous four-electrode stimulation	47
2.4.6	Non-ideal electrode array placement	48
2.5	Discussion	51
2.5.1	Key factors influencing preferential retinal activation	51
2.5.2	Choosing a stimulation strategy	52
2.5.3	Determining membrane thresholds	53
2.5.4	Experimental validation	54
2.5.5	Optimizing electrode currents	55
2.6	Conclusion	55
3	The biophysical basis for the linear electrical receptive field	56
3.1	Abstract	56
3.2	Introduction	57
3.2.1	Multi-electrode stimulation	57
3.2.2	Linear-nonlinear models	57
3.2.3	The biophysical basis for the RGC linear-nonlinear model	58
3.3	Methods	60
3.3.1	Electrode geometry and cell locations	60
3.3.2	The linear-nonlinear model	62
3.3.3	Cell morphology	63
3.3.4	Biophysical model used for validation	64
3.3.5	Comparison of the electrical receptive field and the activating function	66
3.4	Results	68
3.4.1	Recovery of the electrical receptive field	68
3.4.2	Linearity of the electric field and the activating function	69
3.4.3	Comparison of the electrical receptive field and the activating function	71
3.4.4	Comparison of active and leaky integrate-and-fire neuron models of RGCs	73
3.4.5	Biophysical determination of activating function contributions to the ERF	75
3.4.6	Multiple electrical receptive fields for multiple initiation sites	77

3.5	Discussion	80
3.5.1	Comparison to previous studies	83
3.5.2	Biophysical reproduction of the ERF	83
3.5.3	Limitations	84
3.5.4	Improved stimulation strategies	88
3.6	Conclusion	88
4	Spatial shaping of retinal ganglion cell activation	89
4.1	Introduction	89
4.2	Methods	92
4.2.1	The linear-nonlinear model	92
4.2.2	The linear-nonlinear model for spatial shaping	94
4.2.3	The linear-nonlinear neural network	95
4.2.4	Neural network training algorithm	98
4.2.5	Existence of globally optimal solutions under different conditions	100
4.2.6	Network weight initialization	101
4.2.7	Stimulus current constraints	102
4.2.8	Inversion using genetic algorithms	102
4.2.9	Biophysical model of the ERF	103
4.2.10	Analytic parameters	106
4.3	Results	108
4.3.1	Spatial distribution of ERFs	108
4.3.2	Visualization of the objective function	108
4.3.3	Optimality of gradient descent solutions	110
4.3.4	Current steering and virtual electrodes	110
4.3.5	Minimizing activation of the NFL	114
4.3.6	Optimization of complex stimuli	115
4.3.7	Epiretinal versus subretinal stimulation	115
4.3.8	Comparison of animal model anatomies	119
4.4	Discussion	119
4.4.1	Optimality of solutions	121
4.4.2	Constrained optimization	121
4.4.3	Real-world application	122
5	Conclusion	124
5.1	Summary	124
5.1.1	Minimizing activation of overlying axons	124
5.1.2	The biophysical basis for the linear electrical receptive field	124
5.1.3	Spatial shaping of retinal ganglion cell activation	125
5.2	Future areas of investigation	125
5.2.1	Minimizing activation of overlying axons	125
5.2.2	Spatial shaping of retinal ganglion cell activation	126
5.3	Final remarks	127
A	Supporting Material for Chapter 2	128
A.1	Solution of volume equations	128

A.2	Admittivity of the nerve fiber layer	130
A.3	Generalisation of neurite equations	131
A.4	Alternative non-ideal electrode array placement	133
B	Supporting Material for Chapter 3	135
B.1	Analytic derivation of weights for a cylinder	135
C	Supporting Material for Chapter 4	137
C.1	Number of local optima for biphasic stimulation	137
	References	139

List of Figures

1.1	Anatomy of the retina showing the different retinal layers	2
1.2	The effects of retinitis pigmentosa and age-related macular degeneration	3
1.3	Diagram of epiretinal and subretinal prostheses	4
1.4	Retinal ganglion cell calcium imaging showing neuron activation	7
1.5	The cable equation for an infinite cylinder	10
1.6	Diagram of a neurite under longitudinal and transverse modes of stimulation	11
1.7	Diagram of the neurite plus thin extracellular sheath (N+TES) cylindrical neurite geometry	13
1.8	Calculated membrane potential from the cellular composite model and simplifications .	17
1.9	Linear-nonlinear model for visual stimulation	19
1.10	Example of fixed-width biphasic pulses with randomly generated amplitudes delivered to a 20-electrode multi-electrode array	21
1.11	Recovery of positively- and negatively projected spike-triggered averages	22
2.1	Unwanted stimulation of retinal ganglion cell axons of passage	31
2.2	Fiber orientation distributions along the length of the axon	35
2.3	Geometry of the four-layer model of the retina	36
2.4	Simulation of experiments from Fried et al. (2009)	40
2.5	Normalized spread of extracellular potential with distance from a stimulating electrode .	43
2.6	Geometry and simulated membrane potentials for axons of passage and axon initial segments at a variety of x - y orientations	45
2.7	Proportion of axon initial segment orientations preferentially activated for different electrode-retina separations (d_{ER}) and pulse durations	46
2.8	Performance of different electrode configurations with respect to GCL activation level, required stimulus charge density per phase, and radius of activation	49
2.9	Preferential stimulation for two non-ideal electrode array placements	50
3.1	Alternative biophysical approaches to modeling activation	59
3.2	Electrode and tissue simulation geometry	61
3.3	The linear-nonlinear model for electrical stimulation	67
3.4	Illustration of the linear component of the biophysical model	70
3.5	Comparison of the electrical receptive field and the activating vector	72
3.6	Spike-triggered analyses conducted on cells with active versus leaky integrate-and-fire (LIF) membrane dynamics	74

3.7	Linear integration of activating function inputs	76
3.8	The biophysical representation of the linear-nonlinear model	78
3.9	Multiple electrical receptive fields for multiple sites of action potential initiation	79
3.10	Demonstration of sorting SOCB and soma generated spikes	81
3.11	Biophysical representation of multiple ERFs in a single cell	82
4.1	Linear-nonlinear neural network architecture	97
4.2	Schematic of transformation of two-dimensional ERF data	105
4.3	Linear-nonlinear network with GCL and NFL outputs	107
4.4	Distribution of the ERF from a single electrode in space	109
4.5	Map of the objective surface for the linear-nonlinear network	111
4.6	Optimality of fitting procedure for the linear-nonlinear network	112
4.7	Optimization of currents for focal stimulation	113
4.8	Optimization of currents for creating virtual electrodes	114
4.9	Minimizing activation of passing axons	116
4.10	Optimization of complex stimuli for epiretinal stimulation	117
4.11	Interactions of NFL constraints and current constraints	118
4.12	Optimization of subretinal stimulation	119
4.13	Comparison of stimulation of rat and human retinas	120
A.1	Alternative non-ideal electrode array placement	134

List of Tables

1.1	Cellular composite model parameters from Meffin et al. (2012)	15
2.1	Conductivity and thickness of layers for epiretinal model	36
2.2	Multi-layer retinal model parameters	42
3.1	Conductivity and thickness of layers for subretinal model	61
A.1	Cellular composite model parameters for Chapter 2	132

List of Abbreviations

AF	Activating function
AFF	Anisotropic far-field
AIS	Axon initial segment
AMD	Age-related Macular Degeneration
ANF	Anisotropic near-field
AOP	Axon of passage
AV	Activating vector
CCM	Cellular composite model
ERF	Electrical receptive field
GCL	Ganglion cell layer
INL	Inner nuclear layer
IPL	Inner plexiform layer
LIF	Leaky integrate-and-fire
NFL	Nerve fiber layer
ONL	Outer nuclear layer
OPL	Outer plexiform layer
PCA	Principal component analysis
RGC	Retinal ganglion cell
RP	Retinitis Pigmentosa
SOCB	(High-density) sodium channel band
STA	Spike-triggered average
STC	Spike-triggered covariance

Chapter 1

Introduction

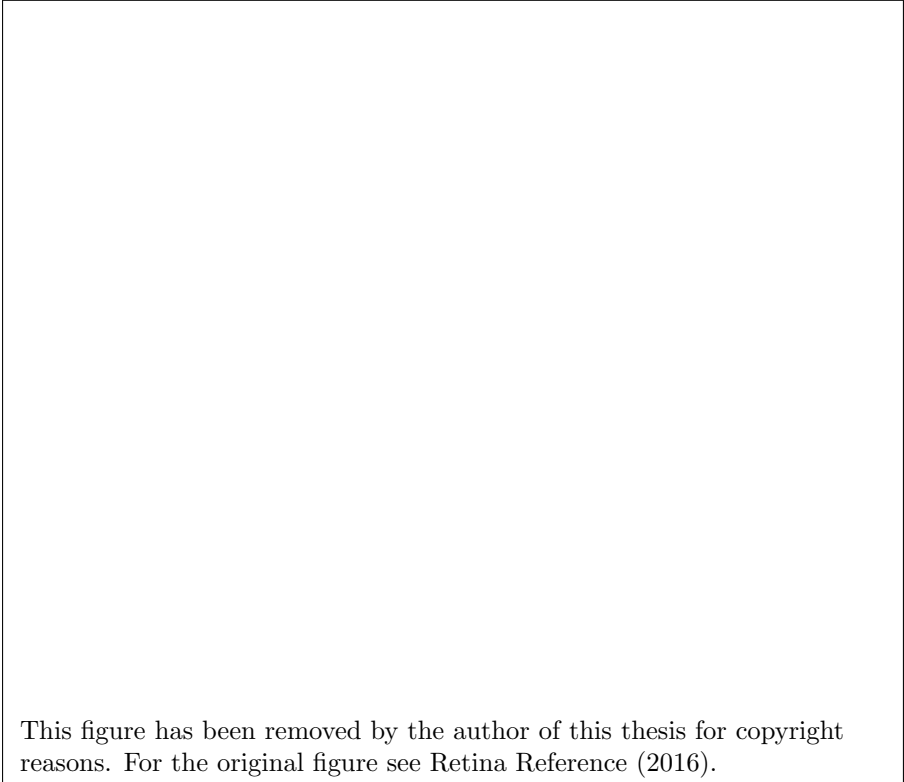
This project deals with the challenge of improving strategies for the electrical stimulation of the retina, with the eventual goal of helping improve therapeutic outcomes for people with retinal prostheses. The objectives of the project investigate two main approaches to developing improved retinal stimulation. The first objective uses detailed, biophysically-accurate models of the retina to gain a greater understanding of the dynamics of current flow and neural activation in complex retinal tissue, which can then be translated into experimental predictions and suggest better stimulation methods. The second and third objectives are directed toward validating and then applying models of retinal activation to calculate optimal stimulation parameters directly via model inversion. This approach requires the use of methods that are able to efficiently invert potentially complex models of neural activation under multi-electrode electrical stimulation.

This literature review will first introduce relevant retinal anatomy and then describe the various existing retinal prostheses and the conditions they have been developed to treat. Existing literature on computational modeling of electrical stimulation of the retina will be presented in two sections. The first section will investigate detailed biophysical modeling of retinal current flow and activation; the second section will investigate heuristic or phenomenological models of spiking probability. Each of these modeling approaches offers different but complementary avenues for the improvement of retinal stimulation outcomes.

1.1 Retinal prostheses

1.1.1 Retinal anatomy

The retina is a light-sensitive layer of neural tissue that lines the back of the eye. The cells of the retina can be classified into five main types: photoreceptors, bipolar cells, amacrine cells, horizontal cells, and retinal ganglion cells (RGCs). Each of these cell types play a role in detecting light or processing visual signals before they are carried to the brain. The location of each major cell type in the retina is shown in Figure 1.1. A direct three-cell chain involving photoreceptors, bipolar cells, and ganglion cells is the major route of information flow in the retina. Photoreceptor cells, divided into rods and



This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Retina Reference (2016).

Figure 1.1: **Anatomy of the retina showing the different retinal layers.** Starting from the epiretinal surface, the retina consists of the nerve fiber layer (NFL), ganglion cell layer (GCL), inner plexiform layer (IPL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), and photoreceptor layer (IS/OS). The 5 main retinal neural cell types and their respective locations are also shown: photoreceptors (rods and cones), horizontal cells, bipolar cells, amacrine cells and retinal ganglion cells (ganglionic cells). Figure taken from Retina Reference (2016) (<http://www.retinareference.com/anatomy/>).

cones, convert incident light energy into neural signals that are then transferred to RGCs via bipolar cells. The RGC population transmits a complete representation of the visual signal from the retina to the brain. The two remaining cell types, amacrine and horizontal cells, are interneurons that carry signals laterally across the retina (Remington, 2005).

The retina is a highly layered structure. The neural component of the retina (excluding bounding structures and non-neural elements) can be broken down into the following layers, proceeding from the outside of the retina to the inside: photoreceptor layer, outer nuclear layer, outer plexiform layer, inner nuclear layer, inner plexiform layer, ganglion cell layer (GCL), and nerve fiber layer (NFL), as shown in Figure 1.1) (Remington, 2005; Retina Reference, 2016). The organization of cells in the retina with respect to incident light, known as the retinotopic organization, is established by the physical location of photoreceptors in the retinal plane, and is preserved throughout each layer of the retina. For this reason, the information carried by a particular RGC is primarily determined by incident light in some region around its cell body. This information is encoded as action potentials passed out of the retina through RGC axons that stream across the inner surface of the retina to their exit point at the optic disc.

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Vision Australia (2012).

Figure 1.2: **The effects of retinitis pigmentosa and age-related macular degeneration.** Images showing normal vision (A) and the visual impact of retinitis pigmentosa and age-related macular degeneration. Retinitis pigmentosa (B) results in ‘tunnel vision’ due to degeneration of peripheral photoreceptor cells. Age-related macular degeneration (C) results in loss of sight in the central visual region due to degeneration of photoreceptors in the macula. Taken from Vision Australia (2012).

Of particular interest in this project are the layers that include retinal ganglion cells: the nerve fiber layer and the ganglion cell layer. Notably, the ganglion cell layer contains only the RGC somas and initial segments of their axons, whereas the nerve fiber layer comprises distal RGC axons, known as axons of passage (AOPs). The NFL has a very different anatomical structure to the other layers of the retina due to the approximately parallel alignment of axons as they pass across the surface of the retina.

1.1.2 Degenerative outer retinal diseases

A key requirement for the success of a retinal prosthesis system is that responses elicited in the retina by the device can be carried to the brain, enabling perception by the user (Weiland and Humayun, 2014; Luo and da Cruz, 2016). As a result, retinal prostheses can only be used to treat retinal conditions that leave at least a partially intact population of RGCs. Two primary degenerative retinal diseases have been identified as having this property: age-related macular degeneration (AMD) and retinitis pigmentosa (RP), each of which result in vision loss due to degeneration of photoreceptors (Hollyfield et al., 2006). Typically, AMD initially results in deterioration of the macular, or central, region of the retina, while RP results in deterioration of the peripheral retina; the whole retina can be affected in later stages of either disease. Figure 1.2 presents the aspects of visual deterioration typical to each condition.

1.1.3 Prosthesis types

Retinal prostheses aim to restore vision to the vision impaired via electrical stimulation of retinal tissue. This requires that these devices can detect light, convert the detected signal to electrical energy, and then deliver appropriate electrical stimulation to the retina (Weiland et al., 2011; Luo and da Cruz, 2016). Ideally, the resulting retinal response from such stimulation would closely mimic the response observed in a healthy retina given the same stimulus light. For the two primary targets for treatment with retinal prostheses, AMD and RP, stimulation aims to excite non-photoreceptor cells in the retina,

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Figure 1 in Weiland et al. (2011).

Figure 1.3: Epiretinal and subretinal prostheses. The main acquisition, signal transmission and stimulation components are shown. External light stimuli are recorded using a video camera and signal processing system, before being transmitted by laser or RF signals to an implanted stimulation device in the eye. Epiretinal prostheses lie in the vitreous body of the eye and are fastened to the inner surface of the retina. Subretinal implants are inserted at the back of the eye, either adjacent to the degenerated photoreceptor cells, or in place of them. Taken from Weiland et al. (2011).

where the primary cell being targeted is determined by both the placement of the electrode array and the stimulation waveform. The two most common types of retinal prostheses, and those dealt with in this project, are epiretinal and subretinal prostheses. The approximate placement of each type within the eye is shown in Figure 1.3. Epiretinal prostheses sit against the inner surface of the retina in the vitreous body whereas subretinal prostheses sit on the outer surface of the retina between the photoreceptor layer and the retinal pigment epithelium. Alternative retinal prostheses placements include suprachoroidal, in which stimulating devices are placed between the sclera and the choroid or within the sclera, and episcleral, in which devices are located on the outer surface of the sclera.

Due to the location of epiretinal prostheses, they tend to stimulate RGCs directly, with less activation of other retinal cells, such as bipolar cells, and less reliance up network-mediated (indirect) activation. This results in epiretinal stimulation having the potential to treat a larger number of conditions and severity levels due to their reliance only on stimulating ganglion cells (Boinagrov et al., 2014). Furthermore, the vitreous body, which surrounds epiretinal prostheses, acts a heat sink, mitigating complications associated with electrical heating of the stimulating device (Piyathaisere et al., 2003).

An associated challenge is that stimulation methods employed in epiretinal devices must attempt to reproduce the full effect of retinal processing that is performed by non-RGCs in the retina (Weiland et al., 2005). Most commonly, epiretinal prostheses utilize an external camera for image acquisition and processing before the processed signal is transferred wirelessly to the implanted component. The orientation of the acquisition camera, normally mounted on glasses, determines the field of view of the wearer, which is usually a subset of the captured image (Figure 1.3) (Weiland et al., 2011; Ghezzi, 2015).

The location of subretinal prostheses is best suited to eliciting network-mediated activation of the retina through stimulation of retinal cells directly adjacent to degenerated photoreceptors. This approach preserves the processing functions of the remaining retinal layers, provided they exist (Weiland et al., 2005). Hence, the representation of desired percepts can be transferred into a pattern of electrical stimulation with only minimal processing, in the hope that the networks in the intact retina will operate as normal. As a result of this property, implementations have been developed for subretinal stimulation that use light-sensitive microphotodiodes, which acts as the system's image acquisition and stimulation system, and require no external device component (Zrenner, 2002). This represents a more direct replacement of the function of the damaged photoreceptor cells. However, several existing challenges associated with this method are the difficulty in generating sufficient stimulation current in the absence of signal amplification and the inability of these systems to perform image processing such as luminance gain control and local contrast enhancement (Weiland and Humayun, 2014). The focus of this thesis is on classical stimulating multi-electrode array designs, and so the use of microphotodiodes will not be addressed in any more detail.

1.1.4 Clinical outcomes to-date

There has been significant progress over the past decade in the development of retinal prostheses for those with retinal pathologies such as RP and AMD. Preliminary acute studies into retinal prostheses successfully established their feasibility, specifically showing that stimulation of the degenerated retina produces perceptible spots of light (phosphenes), that multi-electrode stimulation can produce complex percept shapes, and that elicited perception can be achieved without discomfort to the wearer of the implant (Weiland et al., 2011; Rizzo et al., 2003a,b; Humayun et al., 1996, 1999). This led the way to more advanced clinical assessment of various retinal prostheses designs. Clinical trials of a range of modern retinal prosthetic devices have found that patients can reliably report visual percepts arising from stimulation and perform simple identification tasks (Zrenner et al., 2011; Benav et al., 2010; da Cruz et al., 2013; Humayun et al., 2012; Ahuja et al., 2011).

Several groups have developed epiretinal prostheses, including Second Sight with Argus II (Humayun et al., 2012), the Epi-Ret Consortium with Epi-Ret (Roessler et al., 2009), and a device by Intelligent Medical Implants (IMI) (Hornig et al., 2007). In addition, a device incorporating the technology developed by IMI, the Iris Bionic Vision System, has been developed by Pixium Vision (Hornig et al., 2017). Although the Epi-Ret and IMI devices were designed only for semi-chronic implantation, these devices could elicit visible spots and patients were able to recognize simple patterns and gratings. The Argus II was developed to have an estimated longevity of greater than 10 years and was successful in allowing users of the implant to perceive light, localize objects, and detect object motion. The Argus II was the first retinal prosthetic to receive FDA approval in the US and CE marking for sale in Europe (Luo and da Cruz, 2016), and Pixium's Iris has recently received the CE mark also.

There has also been success in clinical testing of subretinal prostheses. Although, so far, no purely passive and fully implantable subretinal device (requiring no external power source) has yielded significant success, a device that utilizes a microphotodiode array for image acquisition and externally amplified stimulation has shown substantial promise (Zrenner et al., 2011). The device, developed by Retina Implant, GmbH, has been shown to allow patients to detect patterns and simple lines as well as magnified letters. Both the first- and second-generation commercial versions of this device, the Alpha-IMS and Alpha-AMS, respectively, have received CE marking in Europe (Weiland and Humayun, 2014; Stingl et al., 2013). Notably, the Alpha-AMS has a larger number of stimulating units compared to the its epiretinal counterparts (i.e., 1600 compared with less than 200).

Although progress to date is highly encouraging, many aspects of the performance of retinal prostheses remain limited and hinge on the ability of these devices to target either specific retinal cell types or more precise retinal volumes.

1.1.5 Existing challenges

A range of challenges exist relating to the physical design and implantation of retinal prostheses. A key consideration is the ability of devices to have sufficient longevity once implanted. The development of robust hermetic packaging techniques has allowed devices such as the Argus II to achieve expected lifetimes of 10 years or more; however, with current packaging technology, this has required greater implant size (Weiland and Humayun, 2014). Analysis of system lifetime and electrode density for existing devices has highlighted a trade-off between the two variables, indicating the need for enhanced spatial economies for both hermetic sealing and electrode arrays. In addition, long-term implantation of retinal devices requires reliable attachment to the retina and stability over time (Walter et al., 1999). This is particularly challenging for epiretinal prostheses due to the lack of tissue to hold the electrode array in place.

Retinal prostheses must be able to deliver electrical stimulation that has a desired effect on complex, layered retinal tissue. This raises several challenges relating to the geometry of electrode arrays and the form of stimulation waveforms delivered to electrodes. In the case of epiretinal stimulation, a factor that limits performance is the inability of electrical stimulation to preferentially activate target neuronal structures in the ganglion cell layer while avoiding activation of axons in the overlying nerve fiber layer (Grosberg et al., 2017; Jensen et al., 2003; Werginz and Rattay, 2015; Rizzo et al., 2003b; Jepson et al., 2013; Rattay and Resatz, 2004; Werginz et al., 2014; Cao et al., 2015; Weitz et al., 2013; Abramian et al., 2015; Jeng et al., 2011). As mentioned, the retinotopic organization of cells in the retina with respect to incident light is preserved throughout the outer retina and in retinal ganglion cells in the ganglion cell layer. However, this organization is lost in the nerve fiber layer as RGC axons traverse the surface of the retina toward the optic disk (Figure 1.1). Because of this structure, epiretinal electrical stimulation faces the challenge of stimulating the deeper, retinotopically-organized GCL while minimizing activation of axon in the NFL. Often, the shapes of visual percepts described by recipients of epiretinal implants are irregular. Although there may be other causes, this is often attributed to stimulation of axons of passage, which significantly limits perceptual efficacy (Werginz et al., 2014; Rizzo et al., 2003b; Nanduri et al., 2008; Chen et al., 2009). This effect has been confirmed both experimentally and in simulations, and results in a reduction in the spatial selectivity of epiretinal stimulation (Grosberg et al., 2017; Weitz et al., 2013; Rattay and Resatz, 2004; Greenberg et al., 1999; Werginz et al., 2014; Cao et al., 2015; Nanduri et al., 2008; Chen et al., 2009).

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Figure 7 in Weitz et al. (2013).

Figure 1.4: **Retinal ganglion cell calcium imaging showing neuron activation.** (A) Quantized activation levels in responses to biphasic electrical stimulation are obtained from fluorescence microscopy images (B). The level of fluorescence intensity indicates calcium processing due to neural activity. The electrode perimeter is outlined in grey. These results show that retinal ganglion cells 100's of microns away from the stimulating electrode can be activated due to activation of overlying axons of passage. Taken from Weitz et al. (2013).

Figure 1.4 shows the results of a study that used calcium imaging to track neural activity in RGC somas across the retina (Weitz et al., 2013). Using typical biphasic stimulation, it was shown that a stream of spiking responses occurred along the trajectory of passing RGC axons, resulting from the stimulation of axons of passage. Previous clinical studies have reported undesirable perceptual shapes in agreement with the pattern of activation suggested by Figure 1.4 (Rizzo et al., 2003a). These findings strongly suggest that stimulation of overlying axons in the nerve fiber layer results in percepts that extend retinotopically to the location of their somas. Recently, detailed experimental findings using high-density epiretinal electrode arrays has both clearly reaffirmed the existence of this problem while suggesting that it may be overcome using electrode stimuli that are carefully tuned using stimulation and recording schemes (Grosberg et al., 2017).

Another key question surrounding stimulation with multi-electrode arrays is how best to achieve an overall pattern of activation that best matches a desired image or a desired pattern of activation. The ability of electrical stimulation of the retina to evoke visible spots at the approximate location of electrodes has long been established. A logical next step is the development of stimulation strategies that account for the overall influence of a set of simultaneously driven electrodes. However, simultaneous multi-electrode stimulation has generally been avoided in the past due to the belief that interactions between the electric fields induced by adjacent electrodes resulted in what were considered to be unpredictable and undesirable percepts (Horsager et al., 2010; Wilke et al., 2011; Rizzo et al., 2003b). Specifically, a negative effect on both spatial resolution (Wilke et al., 2011) and on the wearer's ability to discriminate objects (Horsager et al., 2010) has been observed when using synchronous stimulation. However, in each of these studies it was suggested that benefit could still be gained from simultaneous stimulation provided electric field interactions are well understood. Overcoming these issues requires the development of computational or phenomenological models that accurately predict neural activation in response to arbitrary multi-electrode stimulation, allowing for strategies that harness the effects of electric field interactions instead of avoiding them.

A primitive method to determine simultaneous multi-electrode stimulation is to simply map image intensity across a field of view to proportional electrical currents at each electrode. However, due to the complex interaction of different retinal neurons and the interaction between driven electrical currents from adjacent electrodes, this presents itself as a challenging problem. A more sophisticated approach to stimulation utilizes methods of *spatial shaping* of retinal activation patterns: using computational models of that activation that attempt to adequately account for the intricacies of retinal tissue and spiking nonlinearities. Through the inversion of models such as these, sets of multi-electrode stimulus currents may be selected to optimally achieve desired patterns of activation.

1.2 Biophysical modeling of retinal stimulation

A potential way to minimize activation of axons of passage in the retina is to take advantage of differences in neurite orientation in the NFL and GCL. The direction of overlying passing axon tracts represents the dominant fiber orientation at each location in the NFL as the axons are packed together as largely parallel fibers (Grosberg et al., 2017; Rattay and Resatz, 2004; Rizzo et al., 2003b). As a result, current flow under extracellular electrical stimulation spreads through retinal tissue in a highly anisotropic way. In contrast to the distal RGC axons in the NFL, proximal axon regions such as the axon initial segment (AIS), located in the GCL, have a wider distribution of orientations as they pass out from the soma. It is expected that, based on these anisotropic tissue characteristics, the orientation of a neurite in retinal tissue can have a significant effect on its activation. Knowledge and understanding of this effect may offer a way to develop stimulation methods that reduce activation of the NFL. To assess the effect of neurite orientation on activation under electrical stimulation, detailed biophysical computational models of current flow and axonal activation should be developed that can accurately capture the anisotropic characteristics of different retinal tissue layers using detailed descriptions of retinal physiology and structure.

A key obstacle in creating simulations of electrical stimulation of the retina is the difficulty in developing models that capture dynamics over the range of relevant spatial scales. These scales can range from 100's of micrometers, for electrode-retina separation distances, dendritic arbors and extent of tissue volumes, to a few nanometers for synaptic clefts (Meffin et al., 2012). As a result, to the author's knowledge, there exist no full-scale, three-dimensional computational models of retinal volume stimulation that include all relevant elements (i.e., electrodes, tissue, neurons, and synapses). Such a model would likely need to be simulated using finite element methods, but this would require a prohibitive amount of computational power.

Several elegant biophysical approximations of electrical stimulation of retinal volumes exist. Two popular approaches include (1) bidomain or continuum models and (2) volume conductor models. Bidomain models use a mean description of intracellular and extracellular conductivity to approximate tissue as a homogeneous medium. Intracellular and extracellular volumes are then modeled as continuous domains existing everywhere in the tissue. This model form allows the simultaneous calculation of current flow in the tissue and activation of neural membranes (Guo et al., 2014; Dokos et al., 2005).

Unlike bidomain models, volume conductor models simulate current flow and neural activation in two distinct steps. Like bidomain models, volume conductor models use a mean approximation of the bulk tissue conductivity to calculate current flow in the volume being simulated. This is then used as

input into a distinct stage 2 neuron model. While in many cases the volume conductor approach is more computationally intensive, it has added flexibility since the stage 2 model may be of arbitrary form, allowing, for example, the use of detailed morphologically-realistic neuron models. It has recently been demonstrated that a class of volume conductor models that uses a linear (sub-threshold) representation of neuron activity has mathematically equivalent bidomain model representations (Meffin et al., 2014).

1.2.1 Volume conductor models

Stage 1 of a volume conductor model simulates the electric field or current density due to stimulating electrodes using some linear, continuous, macroscopic approximation of the tissue. In the general case, Laplace- or Poisson-type equations govern current density in the volume considered:

$$\nabla \cdot \sigma(\mathbf{x}, t) \nabla \phi_e(\mathbf{x}, t) = -\mathbf{I}_0(\mathbf{x}, t) \quad (1.1)$$

where ∇ is the three-dimensional differential operator, σ is the tissue conductivity tensor (a 3x3 matrix), ϕ_e is the extracellular potential, \mathbf{I}_0 is the current source per unit volume, and $\mathbf{x} = (x, y, z)$.

This description is either solved analytically given suitable assumptions and constraining boundary conditions or evaluated numerically using finite element modeling software. A very common simplification of the general description is to use a point source electrode in an infinite isotropic and purely resistive (i.e., no capacitive component) tissue, which yields the solution

$$\phi_e(\mathbf{x}, t) = \frac{\mathbf{I}_0(t)}{4\pi\sigma r(\mathbf{x})}, \quad (1.2)$$

where $r(\mathbf{x})$ is the distance from the position of analysis, \mathbf{x} , to the point source (Malmivuo and Plonsey, 2012). For less trivial and more accurate tissue descriptions, such as retinal models that include multiple distinct anisotropic layers, analytic solutions can sometimes be obtained using Fourier domain approaches (Meffin et al., 2012, 2014), an example of which is given in Section 1.2.5.

The second stage of a volume conductor model takes the calculated extracellular potential, ϕ_e , or current density, \mathbf{J}_e , from stage 1 as input into a more detailed model of a neuron to calculate membrane potential. Stage 2 neuron models are typically either nonlinear, morphologically-realistic, Hodgkin-Huxley-type models or linear, sub-threshold models of membrane potential. Shown in Figure 1.5, a simplistic volume conductor implementation models neurons as uniform cylindrical conductors using the well-known cable equation with a driving extracellular potential (Rattay, 1989; Hodgkin and Huxley, 1990),

$$c_m \frac{\partial V_m}{\partial t} = -\frac{V_m}{r_m} + \frac{1}{r_a} \left(\frac{\partial^2 V_m}{\partial x^2} + \frac{\partial^2 \phi_e}{\partial x^2} \right) \quad (1.3)$$

where V_m is the membrane potential, c_m is the membrane capacitance per unit length ($\text{F} \cdot \text{m}^{-1}$), r_m is the membrane unit length resistance ($\Omega \cdot \text{m}$), r_a is the intracellular axial resistance per unit length ($\Omega \cdot \text{m}^{-1}$), x is displacement along the neuronal cylinder, and a is the neurite radius. This expression is used to model subthreshold behavior in neurites. To model nonlinear spiking behavior, the term V_m/r_m is replaced with nonlinear ionic currents, i_{ionic} , as in the Hodgkin-Huxley equations and conductance-based modeling software such as NEURON (Hines and Carnevale, 1997; Rattay, 1986; Coburn, 1989; Rattay, 1999; Greenberg et al., 1999).

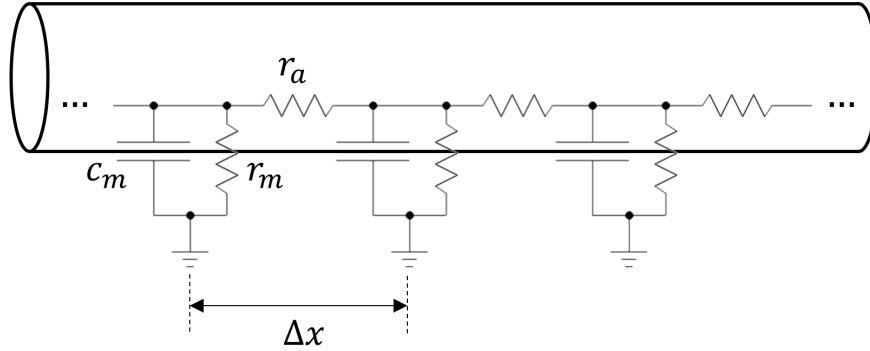


Figure 1.5: **The cable equation for an infinite cylinder.** Using an infinite cylinder approximation for neurites, sub-threshold membrane potential is often described using cable equation. A discrete form of the cable equations models the membrane as a set of linked resistor-capacitor circuits separated by Δx . The continuous form of the cable equation, Equation 1.3, is recovered by setting $\Delta x \rightarrow 0$.

The cable model framework models only the activation of the neural membrane caused by the gradual transfer of current from the extracellular to the intracellular space as it flows longitudinally along the length of the neurite, known as the longitudinal mode of activation. However, more complex modes of stimulation also contribute to overall activation, as shown in Figure 1.6. The overriding assumption of the cable equation is that the voltage gradient is negligible in the cross-sectional plane of the neurite, resulting in predominantly longitudinal current flow in and around the neurite and rotationally symmetric activation, as shown in Figure 1.6(A) and (B).

In the real system, there are also more complex modes of stimulation, known as transverse modes, that include rotational dependence. Examples of these modes are depicted in Figure 1.6(C) and (D). Transverse modes are induced by the cross-sectional flow of current across a neurite. Although the transverse mode is typically neglected due to the assumption that its magnitude is small compared with that of the longitudinal mode, it has been found to be significant, and even dominant, under conditions relevant to retinal stimulation (Tahayori et al., 2012, 2014). Specifically, the transverse mode becomes more relevant for higher frequency stimulation and larger electrode-neurite separation distances. Due to the focus in this project on orientation-dependence in neural activation, models of the transverse mode will be dealt with in some detail later.

1.2.2 Limitations of standard volume conductor models

The major advantage of the volume conductor approach is that it overcomes issues related to the range of relevant spatial scales in neural tissue by describing the tissue's macroscopic properties using bulk conductivity. However, this approximation requires a series of assumptions to be made about the underlying tissue. Typically, this method assumes the tissue to be smooth and homogeneous on a microscopic scale. As a result, important geometrical and dynamical properties of the microscopic units that comprise the tissue (i.e., neurons) are neglected when calculating macroscopic current spread in stage 1. In reality, neural tissue has a significant mean capacitance, which can result in markedly more complex responses to stimulation than would be expected of purely resistive tissue.

A recently highlighted side-effect of this approximation is internal model inconsistency between

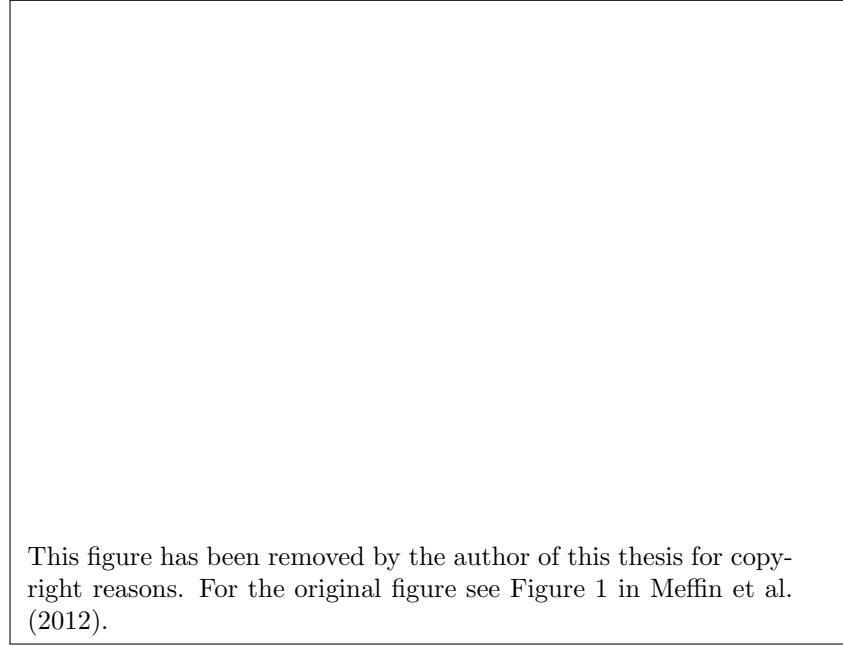


Figure 1.6: A neurite under different longitudinal and transverse modes of stimulation. Different modes are characterized by the the orientation and rotational dependence of current flow. Arrows indicate the direction of current flow in each mode. (A-B) For the longitudinal mode, current enters the neurite progressively and flows axially within the cylinder. (C-D) In each transverse mode, the current flows in the neurite’s cross-sectional plane. Adapted from Meffin et al. (2012).

the first and second stages of the volume conductor model. The calculated membrane potential can be significantly different depending on whether the model’s boundary conditions, and the extracellular input into stage 2 of the model, are specified in terms of current density or electric potential (Meffin et al., 2012; Tahayori et al., 2012). In essence, this reflects the fact that the inherent ohmic transfer function is different in stages 1 and 2 of the model.

1.2.3 Simulating morphological reconstructions

In the more general case of arbitrary neuron morphology, multi-compartment models are generally used. In this case, a spatially-discretized version of Equation (1.3) with ionic currents may be used for stage 2 of the volume conductor:

$$C_{m,n} \frac{dV_{m,n}}{dt} = -I_{\text{ionic},n} + \frac{V_{m,n-1} - V_{m,n}}{\frac{1}{2}(R_{a,n-1} + R_{a,n})} + \frac{V_{m,n+1} - V_{m,n}}{\frac{1}{2}(R_{a,n+1} + R_{a,n})} + \dots + f_n, \quad (1.4)$$

where $V_{m,n}$, $R_{a,n}$, and $C_{m,n}$ are the membrane potential, axial resistance, and capacitance, respectively, at compartment n . The above equation can accommodate neural compartments with more than two neighboring components (i.e., branch points) via the (\dots) term. The value of f_n is given by

$$f_n = \frac{V_{e,n-1} - V_{e,n}}{\frac{1}{2}(R_{a,n-1} + R_{a,n})} + \frac{V_{e,n+1} - V_{e,n}}{\frac{1}{2}(R_{a,n+1} + R_{a,n})} + \dots, \quad (1.5)$$

where $V_{e,n}$ is the extracellular voltage at compartment n as determined by the output of stage 1 of the volume conductor.

The function f_n is known as the activating function, a concept pioneered by Rattay (1986, 1989, 1999), represented above in discrete form. The continuous form of the activating function can be inferred from Equation (1.5) as

$$f = \frac{1}{r_a} \frac{\partial^2 \phi_e}{\partial x^2}.$$

The activating function provides a general framework for the combination of arbitrary stage 1 current flow models with morphologically-accurate, conductance-based, multi-compartmental neural models. Commonly, models such as these are implemented using a combination of linear analysis or finite element software (e.g., MATLAB or COMSOL) and conductance-based neural modeling software (e.g., NEURON (Hines and Carnevale, 1997)).

In either the continuous or discrete cases, the activating function represents the driving force acting upon a neuron due to electrical stimulation as an equivalent transmembrane current. As can be seen, the activating function is proportional to the second derivative of the extracellular potential along the neurite or, in the discrete case, the second difference quotient of the extracellular potential along the neurite. This dependence on the second derivative of the extracellular potential is a key principle of extracellular electrical stimulation.

1.2.4 Existing models

A modeling study by Rattay and Resatz (2004) assessed the influence of electric field orientation with respect to neurites in the NFL. This model showed that, by orientating long, rectangular electrodes parallel to axons in the NFL, the activation of those axons could be reduced. In addition, Cao et al. (2015) developed a model with a 9-layer description of retinal tissue that reproduced experimentally-reported irregular phosphene shapes resulting from preferential activation of axons. Although these models successfully utilize volume conductor models to overcome issues relating to spatial scale in neural tissue simulations, each model relies on key simplifications that have implications for the current research. Each of the above models was constructed using the standard volume conductor framework using purely isotropic and resistive descriptions of current flow in each retinal layer and conductance-based multi-compartment models. This means that the temporal processing characteristics of retinal tissue are neglected, as well as the potentially significant interplay between fiber orientation and tissue anisotropy.

Dokos et al. (2005) reported a bidomain model that coupled a standard extracellular domain with an active, conductance based intracellular domain. This work extended on previous attempts by capturing the capacitive effect of the neural tissue on extracellular potential. The bidomain model is described by the coupled equations

$$\sigma_e \nabla^2 \phi_e = -\mathbf{J}_m \tag{1.6a}$$

$$\sigma_{i,x} \frac{\partial^2 \phi_i}{\partial x^2} = \mathbf{J}_m \tag{1.6b}$$

$$\beta \left(c_m \frac{\partial V_m}{\partial t} + i_{\text{ionic}} \right) = \mathbf{J}_m \tag{1.6c}$$

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Figure 3 in Meffin et al. (2012).

Figure 1.7: The neurite plus thin extracellular sheath (N+TES) cylindrical neurite geometry. Neural tissue is comprised of densely packed neural fibers with confined extracellular space, represented as a thin surrounding sheath. Neurite membrane potential is governed by the intra- and extracellular resistivity (ρ_i and ρ_e) and by the electrical properties of the neural membrane (r_M and c_M), which are modeled by a distributed RC electrical circuit. Taken from Meffin et al. (2012).

$$V_m = \phi_i - \phi_e \quad (1.6d)$$

where β is the surface to volume ratio for the intracellular domain, σ_e is the 3x3 conductivity tensor of the extracellular space, and $\sigma_{i,x}$ is the intracellular conductivity in the axial direction, x , which is assumed to be zero in the y - and z -directions. Due to the nonlinearity introduced by the ionic currents, i_{ionic} , an analytic solution of this system does not exist. Consequently, the model was implemented numerically using custom finite element software. Although this model neglected the impact of the highly anisotropic NFL, it was shown to improve on previous models in demonstrating the previously ignored time-varying behavior of the induced extracellular potential caused by the tissue capacitance. Subsequently, an extension of this model type that incorporated multiple distinct retinal layers and anisotropy was reported by Shalhaf et al. (2015).

Recently, a volume conductor model was presented by Meffin et al. (2012, 2014) and Tahayori et al. (2012, 2014) that connected the microscopic properties of the tissue's constituent fibers to the macroscopic electrical properties of the simulated volume. This formulation, known as the cellular composite model, overcame several existing limitations of biophysical models of electrical stimulation by generalizing the usual effective conductivity term with a time and spatially varying complex admittivity kernel. This model forms the basis for a significant portion of the current work and is discussed in detail in the following section.

1.2.5 The cellular composite model

In the cellular composite model, the spatially- and temporally-static volume conductor conductance is replaced by an admittivity kernel derived using the confined geometry and electrical properties of the constitutive neural tissue, ensuring model self-consistency. Admittivity has both spatial and temporal (or frequency) dependence, unlike conductivity, which varies only in space. This generalization of conductivity allows for spatially- and temporally-dependent relationships between the extracellular potential and current.

The constitutive and continuity equations for stage 1 of the cellular composite model are given by the following, using a generalized form of Ohm's law for spatially and temporally dependent admittivity:

$$\text{constitutive equation: } \mathbf{J}_e = -\frac{1}{2\pi} \xi_e(x, t) *_{x,t} \nabla \phi_e \quad (1.7a)$$

$$\text{continuity equation: } \nabla \cdot \mathbf{J}_e = 0, \quad (1.7b)$$

where ξ_e is the neural tissue admittivity kernel. In the Fourier domain with respect to axial location, x , and time, t , the model simplifies to

$$\text{constitutive equation: } \check{\mathbf{J}}_e = -\check{\xi}_e(k_x, \omega) \check{\nabla} \check{\phi}_e \quad (1.8a)$$

$$\text{continuity equation: } \check{\nabla} \cdot \check{\mathbf{J}}_e = 0 \quad (1.8b)$$

where $\check{\cdot}$ indicates the Fourier transform of the specified quantity with respect to x and t , with Fourier transform pairs k_x and ω , respectively.

The admittivity kernel, $\xi_e(x, t)$, is derived from an expression of the transimpedance of neural tissue with a *Neurite plus Thin Extracellular Sheath* (N+TES) geometry and membrane capacitance, as shown in Figure 1.7. Transimpedance was determined by relating analogous stage 2 expressions for membrane potential using both voltage and current density boundary conditions. The transimpedance and corresponding admittivity found using this approach guarantees model self-consistency by ensuring that the current-voltage relationship is equivalent for both the stage 1 volume model and the stage 2 neurite model. As shown in Figure 1.8, unlike the cellular composite model, common simplifying approximations such as near- and far-field approximations lead to large discrepancies between the results obtained with each set of boundary conditions. For each approximate model, the calculated membrane potential varies by as much as an order of magnitude between sets of boundary conditions. In addition, the cellular composite model accurately calculates membrane potential for the full range of tested electrode-neurite separation distances.

The following three-dimensional Laplace equations govern the spatial distribution of intracellular and extracellular potential:

$$\nabla^2 \phi_i(\mathbf{x}, t) = 0 \quad (1.9a)$$

$$\nabla^2 \phi_e(\mathbf{x}, t) = 0, \quad (1.9b)$$

with the normal RC circuit representation of the neurite membrane,

$$C_m \frac{dV_m}{dt} + \frac{V_m}{R_m} = J_m \quad (1.10a)$$

$$V_m = \phi_i - \phi_e. \quad (1.10b)$$

Neurite equations are obtained by considering a cylinder of radius a of infinite extent. A key point of difference in the cellular composite model is that the modeled N+TES geometry correctly captures the extremely confined extracellular space evident in dense neural tissue such as the retina, particularly in the NFL. Given appropriate voltage or current boundary conditions, the system defined

Table 1.1: Cellular composite model parameters from Meffin et al. (2012)

Parameter	Description	Unit
a	Neurite radius	m
b	Outer cylinder radius	m
d	Width of extracellular sheath, $d = b - a$	m
ρ_i	Intracellular resistivity	$\Omega \text{ m}$
ρ_e	Extracellular resistivity	$\Omega \text{ m}$
r_i	Intracellular resistance per unit length, $r_i = \rho_i/(\pi a^2)$	Ω/m
r_e	Extracellular resistance per unit length, $r_e = \rho_e/(\pi(b^2 - a^2))$	Ω/m
R_m	Membrane unit area resistance	$\Omega \text{ m}^2$
r_m	Membrane unit length resistance, $r_m = R_m/(2\pi a)$	$\Omega \text{ m}$
$R_{e,J}(1)$	Transverse extracellular specific resistance for current density boundary conditions, $R_{e,J}(1) = \rho_e b^2/d$	Ω/m^2
C_m	Membrane capacitance per unit area	F/m^2
c_m	Membrane capacitance per unit length	F/m
τ_m	Membrane time constant, $\tau_m = R_m C_m = r_m c_m$	s
Z_m	Membrane specific impedance, $Z_m = (1/R_m + j\omega C_m)^{-1}$	$\Omega \text{ m}^2$
λ_{0J}	Static electrotonic length constant for current density boundary conditions, $\lambda_{0J} = \sqrt{r_m/(r_e + r_i)}$	m
$\lambda_J(\omega)$	Frequency-dependent electrotonic length constant for current density boundary conditions, $\lambda_J(\omega) = \lambda_{0J}/\sqrt{1 + j\omega\tau_m}$	m
λ_{0V}	Static electrotonic length constant for voltage boundary conditions, $\lambda_{0V} = \sqrt{r_m/r_i}$	m
$\lambda_V(\omega)$	Frequency-dependent electrotonic length constant for voltage boundary conditions, $\lambda_V(\omega) = \lambda_{0V}/\sqrt{1 + j\omega\tau_m}$	m

by Equations (1.9) and (1.10) can be solved for the N+TES by cylindrical harmonic expansion of the three-dimensional Laplace equations (1.9) in the intracellular and extracellular space. This yields expressions for both the longitudinal and transverse modes of stimulation, given in the Fourier domain as

$$\text{Voltage B.C.s: } \check{V}_{m,L} = \frac{-k_x^2 \lambda_V^2(\omega)}{1 + k_x^2 \lambda_V^2(\omega)} \check{V}_{e,L} \quad (1.11a)$$

$$\check{V}_{m,T} = -2\check{V}_{e,T} \quad (1.11b)$$

$$\text{Current B.C.s: } \check{V}_{m,L} = \frac{2\pi b r_e \lambda_J^2(\omega)}{1 + k_x^2 \lambda_J^2(\omega)} \check{\mathbf{J}}_{e,L} \quad (1.11c)$$

$$\check{V}_{m,T} = -\frac{2\rho_e b^2}{d} \check{\mathbf{J}}_{e,T}. \quad (1.11d)$$

All relevant quantities in Equations (1.11) are defined in Table 1.1. These expressions define the underlying Ohmic relationship between current and potential in neural tissue. To determine the corresponding Ohmic transfer function, or the transimpedance, expressions for membrane potential for each set of boundary conditions are equated (i.e., equate (1.11a) with (1.11c) and equate (1.11b) with (1.11d)),

$$\frac{\check{V}_{e,L}}{\check{\mathbf{J}}_{e,L}} = \check{Z}_{e,L} = -\frac{2\rho_i}{k_x^2 b} \left(\frac{1 + k_x^2 \lambda_V^2(\omega)}{1 + k_x^2 \lambda_J^2(\omega)} \right) \quad (1.12a)$$

$$\frac{\check{V}_{e,T}}{\check{\mathbf{J}}_{e,T}} = \check{Z}_{e,T} = \frac{\rho_e b^2}{d}. \quad (1.12b)$$

By using this transimpedance to derive the following tissue admittivity, the cellular composite model captures the microscale geometry and physiology of the N+TES geometry in stage 1 of the model while simultaneously guaranteeing internal model consistency. The final cellular composite model admittivity is obtained by combining Equation (1.12) with the constitutive Equation (1.7a), giving

$$\begin{aligned} \check{\xi}_e(k_x, \omega) &= \begin{bmatrix} \check{\xi}_{e,x} & 0 & 0 \\ 0 & \check{\xi}_{e,y} & 0 \\ 0 & 0 & \check{\xi}_{e,z} \end{bmatrix} \\ &= \begin{bmatrix} \check{\xi}_{e,L} & 0 & 0 \\ 0 & \check{\xi}_{e,T} & 0 \\ 0 & 0 & \check{\xi}_{e,T} \end{bmatrix} \end{aligned} \quad (1.13)$$

where

$$\check{\xi}_{e,T} = \frac{d}{b\rho_e} \quad (1.14)$$

$$\check{\xi}_{e,L}(k_x, \omega) = \frac{1}{\rho_i} \frac{1 + j\omega\tau_m + k_x^2 \lambda_{0J}^2}{1 + j\omega\tau_m + k_x^2 \lambda_{0V}^2}. \quad (1.15)$$

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Figure 7 in Tahayori et al. (2014).

Figure 1.8: Comparison of the calculated membrane potential from the cellular composite model (CCM), an anisotropic near-field approximation (ANF) and an anisotropic far-field approximation (AFF) for both voltage (A) and current density (B) boundary conditions. The maximum membrane potential has been calculated for a range of electrode separation distances to illustrate the limitations of the near- and far-field approximations at large and small separation distances, respectively. Taken from Tahayori et al. (2014).

Importantly, the cellular composite model provides a common framework for calculating the relative contributions of the longitudinal and transverse modes of stimulation. Analysis presented in previous work reveals that both modes are significant for a range of temporal (pulse duration) and spatial (electrode-neurite separation distance) stimulation characteristics, supporting the combined consideration of both modes (Meffin et al., 2012; Tahayori et al., 2012). Furthermore, it has been shown that simpler volume conductor models are only valid for a restricted set of situations (e.g., for large electrode-neurite separations or low stimulation frequencies) due to neglecting important spatiotemporal tissue characteristics and model self-inconsistency. Outside of these situations, these models diverge from true values by up to an order of magnitude.

1.3 Phenomenological modeling of retinal stimulation

A distinct class of computation models aims to simulate neural responses in the retina to electrical stimulation using simplified functional models of neuron spiking. This facilitates the development of more parsimonious and computationally efficient descriptions of retinal tissue at the expense of accuracy and potentially interpretability. Simplified models may be obtained by either the progressive application of simplifying assumptions to detailed biophysical models or by the empirical construction of phenomenological models given experimental observations. A combination of both approaches may also be used whereby the model structure is informed by some knowledge of the neural system being model, with model parameters then being fit to experimental data.

1.3.1 Block-structured models/cascaded filter models

Block-structured models, also known as cascaded filter models, aim to represent single-cell, network, or mean retinal responses using black- or grey-box approaches. Primarily, these models are structured to most efficiently capture the statistical relationships between multi-dimensional stimuli and RGC spiking rates without explicitly modeling the anatomy or physiological dynamics of the system.

Block-structured models are defined by a series of cascaded linear and nonlinear filters. In this way, different retinal cell types or different signal transformations that occur can be represented by a distinct series of filters or transfer functions (Guo et al., 2014; Pillow, 2005; Pillow et al., 2008). Although these cascades do not attempt to recreate biophysical properties, they can be effective in reproducing functional responses for both single cells and populations (Curlander and Marmarelis, 1987; Teeters et al., 1997; Wohrer and Kornprobst, 2009). The main advantage of such models is their simplicity, generally requiring only a few model parameters to be fitted to reproduce functional behavior. This property also facilitates the translation of models between species more easily.

Historically, block-structured models have been used predominantly to model retinal responses to light stimuli, with the primary focus being on the behavior of outer retinal cells such as photoreceptor and bipolar cells (van Hateren and Snippe, 2007; Hamer et al., 2005). In this context, models have been used successfully to investigate nonlinear signal processing mechanisms in both the healthy and degenerate retina in response to natural light stimulation (Hood et al., 1993). A series of block-structured models have also been shown to characterize spatial summation properties accurately in several types of RGCs. Block-structured models have also been effective when incorporated into spike train simulators, which use the output from a series of filter blocks to drive a probabilistic (typically

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Figure 4 in Keat et al. (2001).

Figure 1.9: Spiking model of Keat et al. (2001). The light stimulus, $s(t)$, undergoes a linear transformation, $F(t)$, producing a generator potential signal, $g(t)$, that is used to drive a nonlinear spiking process governed by threshold, θ , producing a spike train, $r(t)$. A negative feedback loop models the return of the generator potential to below threshold after each spike, and two Gaussian noise signals, $a(t)$ and $b(t)$, are used to generate observed trial-to-trial variability. Taken from Keat et al. (2001).

Poisson) spiking process. A model developed by Keat et al. (2001) used this technique to reproduce the unique spiking behaviors of a broad class of RGC types in various species. A diagrammatic description of the model is shown in Figure 1.9. A single linear filter, $F(t)$, is combined with a spiking nonlinearity based on a constant spiking threshold, θ , with a negative feedback interconnection, $P(t)$. These filters were connected as shown to map a set of light stimulus inputs, $s(t)$, to spiking responses, $r(t)$. Filter-based models like these have also proven the most effective in investigating more complex retinal functions, such as motion detection and motion prediction in the retina, capturing the relative influence of different cell types, cell-cell connections, and network architecture in performing these functions (Berry et al., 1999; Baccus et al., 2008).

The main limitation of block-structured models is that, although they may be able to accurately simulation functional behavior, they do not generally elucidate the underlying physiological or morphological mechanisms that yield the observed behavior. In reality, the physiological processes governing signal processing in the retina and the influence of cell and tissue anatomy are far more complex than a series of cascaded spatial and temporal filters. Partially due to this, until recently, there has been limited application of these types of models to the simulation of RGC responses to extracellular electrical stimulation (Guo et al., 2014).

1.3.2 The linear-nonlinear model of extracellular stimulation

Recently, a block-structured model used by Maturana et al. (2016), known as the linear-nonlinear model, has demonstrated utility in accurately simulating RGC responses to subretinal multi-electrode array stimulation. The linear-nonlinear model uses a linear dimensionality-reducing filter to characterize the stimulus space as an N-dimensional subspace. The output of this transformation is then passed through a nonlinearity to generate spiking probability or spiking rate. When applied to individual RGCs, the output of the linear transformation is thought to be a scaled approximation of the resultant

transmembrane current in the cell; the output of the nonlinearity is the resultant probability of a spike occurring given the stimulus. Hence, the linear and nonlinear components aim to represent the cell's integration of many stimulus currents and its intrinsic nonlinear spiking behavior, respectively.

The original stimulus space used as input into the linear-nonlinear model is composed of the amplitudes of fixed-period biphasic pulses delivered to each of the electrodes of a subretinal multi-electrode array. Using this description of the stimulus space, the model simulates the combined effect of arbitrary stimulus amplitudes across the electrodes of an array. The most general form of the model may be described by the following: given a stimulus vector, \vec{S}_t , of stimulus amplitudes, the probability of a spike being generated is given by

$$P(R_t = \text{spike} | \vec{S}_t) = \mathcal{N}_N(\vec{v}_1 \cdot \vec{S}_t, \vec{v}_2 \cdot \vec{S}_t, \dots, \vec{v}_N \cdot \vec{S}_t), \quad (1.16)$$

where $R_t \in \{\text{spike}, \text{not spike}\}$ is the RGC response at time t , \mathcal{N}_N is the nonlinearity operating on the N -dimensional stimulus subspace, and \vec{v}_i for $(i = 1, 2, \dots, N)$ are a set of orthogonal vectors defining a projection from the full stimulus space to an N -dimensional subspace.

The set of vectors, \vec{v}_i , are the N significant principal components calculated on the subset of stimulus data that elicited direct (short-latency) spiking responses, which can be obtained using an experimental stimulation and recording scheme. In the analysis presented by Maturana et al. (2016), the experimental stimuli consisted of biphasic pulses of 500 μs phase duration and a 50 μs interphase gap. A sample set of electrode stimuli are shown in Figure 1.10, where current amplitudes for each electrode in each stimulus application were sampled randomly from a Gaussian distribution with an appropriate variance, σ_{exp} (chosen to avoid over-stimulation), and a mean of 0 μA . Since all stimuli are biphasic and charge-balanced, a positive current amplitude is an anodic-first pulse and a negative current is a cathodic-first pulse.

To calculate the principal components, the complete stimulus data, \mathbf{S}_T , was filtered to include only the stimuli that resulted in direct RGC spiking responses, \mathbf{S}_D . For an array with N_E electrodes, principal component analysis (PCA) is then performed on \mathbf{S}_D , yielding the set of N_E eigenvectors of its $N_E \times N_E$ covariance matrix, known as the principal components. The corresponding eigenvalues are the variance in the direction of each principal component. To determine which principal components are significant, if any, the value of these eigenvalues for \mathbf{S}_D are compared to those for \mathbf{S}_T , which are approximately equal to σ_{exp} by virtue of the Gaussian distribution from which stimuli were drawn. Those eigenvalues that are significantly different correspond to the set of orthogonal stimulus space projections that carry information about RGC spiking probability. For most cells analyzed, Maturana et al. (2016) found that only one principal component was significant, meaning that a one-dimensional reduction accurately captured the direct spiking activity. Equation (1.16) then becomes

$$P(R_t = \text{spike} | \vec{S}_t) = \mathcal{N}_1(\vec{v}_1 \cdot \vec{S}_t). \quad (1.17)$$

To simulate the asymmetrical responses to cathodic- and anodic-first stimuli, which has been presented in the literature (Jensen and Rizzo, 2006), the nonlinear function \mathcal{N}_1 is generalized to be of a double-sigmoid form,

$$\begin{aligned} P(R_t = \text{spike} | \vec{S}_t) &= \mathcal{N}_1(\vec{v}_1 \cdot \vec{S}_t) \\ &= \mathcal{N}_+(\vec{w}_+ \cdot \vec{S}_t) + \mathcal{N}_-(\vec{w}_- \cdot \vec{S}_t), \end{aligned} \quad (1.18)$$

This figure has been removed by the author of this thesis for copyright reasons. For the original figure see Figure 2 in Maturana et al. (2016).

Figure 1.10: Example of fixed-width biphasic pulses with randomly generated amplitudes delivered to a 20-electrode multi-electrode array. (A) A sample set of biphasic pulses with color indicating current amplitude in μA . (B) The set of stimulation timecourses corresponding to the amplitudes in (A). Stimulus current amplitudes are drawn from a zero mean Gaussian distribution with variance σ_{exp}^2 . Taken from Maturana et al. (2016).

where \vec{w}_+ and \vec{w}_- are the spike triggered averages (STAs) for stimuli with positive and negative values, respectively, following projection by the first principal component, \vec{v}_1 , called the *electrical receptive field*. An electrical receptive field is defined as a given cell's sensitivity to stimulation from each electrode in a multi-electrode array, defined by a vector of weights equal in length to the number of electrodes. To obtain these ERF vectors, \mathbf{S}_D is split according to the sign of $\vec{v}_1 \cdot \mathbf{S}_D$, as shown in Figure 1.11B, giving \mathbf{S}_D^+ and \mathbf{S}_D^- . The means of the amplitudes for each electrode in each of these subsets are vectors corresponding to \vec{w}_+ and \vec{w}_- , representing the *average* net anodic-first and net cathodic-first spike-triggered averages (STA), respectively.

Recovery of \vec{w}_+ and \vec{w}_- is depicted visually in Figure 1.11. Figure 1.11(A) shows an example of an applied stimulus train, which is sub-sampled to yield only the stimuli that elicited spikes, \mathbf{S}_D . Following the calculation of the principal components of \mathbf{S}_D , spike-triggered stimuli are projected onto the first principal components, as shown in Figure 1.11(B). The spike-triggered averages \vec{w}_+ and \vec{w}_- are the average of the stimuli that have a positive or negative value, respectively, when projected onto the first principal component (i.e., the horizontal axis in Figure 1.11(B)).

The sigmoidal nonlinearities of the linear-nonlinear model are implemented as logistic functions,

$$\mathcal{N}_+(\vec{w}_+ \cdot \vec{S}_t) = \frac{a_+}{1 + \exp(-b_+(\vec{w}_+ \cdot \vec{S}_t - c_+))} \quad (1.19a)$$

$$\mathcal{N}_-(\vec{w}_- \cdot \vec{S}_t) = a_- - \frac{a_-}{1 + \exp(-b_-(\vec{w}_- \cdot \vec{S}_t - c_-))}, \quad (1.19b)$$

where the coefficients a_+ , b_+ , and c_+ represent the maximum spiking probability, slope, and threshold, respectively, and similarly for a_- , b_- , and c_- .

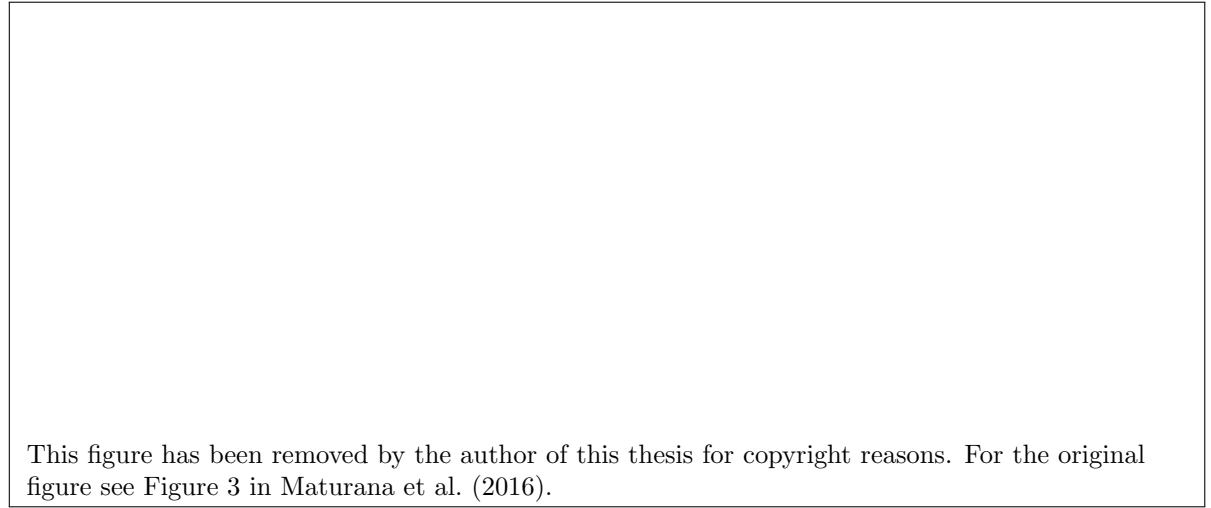


Figure 1.11: Recovery of positively- and negatively projected spike-triggered averages. (A) An example set of binary cell responses (where a spike indicates a direct spiking response) and the associated set of stimuli, \mathbf{S}_T . Each stimulus in \mathbf{S}_T (represented as a line) consists of fixed-width biphasic pulses applied to each electrode with randomly generated amplitudes. The spike-triggered stimuli, \mathbf{S}_D , are those stimuli which resulted in a spike. (B) Scatter plot of the stimuli projected onto the first 2 principal components of \mathbf{S}_D . Black markers show all stimuli and white markers show only spike-triggered stimuli. The centers of the positive and negative spike-triggered stimulus clouds, a diamond and circle, respectively, represent the projections of the electrical receptive fields. This shows the typical scenario where only the first principal component calculated on the spike-triggered data contains significant information about spiking probability. Taken from Maturana et al. (2016).

1.3.3 Underlying assumptions

The one-dimensional form of the linear-nonlinear model described in Section 1.3.2 is structured according to the following series of simplifying assumptions:

1. Retinal ganglion cell responses are determined primarily by a one-dimensional linear projection of stimulus current amplitudes.
2. The nonlinear spiking behavior of RGCs can be accurately represented by a double-sided sigmoidal function.
3. More complex neural dynamics such as stimulation time-courses and temporal interactions between electrode currents, which are neglected by using a rate approximation, have little effect.

The first assumption is the most significant and has important implications for our understanding of cellular processing by RGCs. A key finding by Fried et al. (2009) revealed a high-density sodium channel band (SOCB) in the RGC axon initial segment. This short axonal segment was found to have a significantly lower stimulation threshold than the soma or other parts of the axon, and was shown to be the primary location of spike initiation and overall response shaping. Significantly, this finding is supported by that of Maturana et al. (2016). The existence of a single cellular location responsible for driving cell activity during electrical stimulation implies that cell input may be approximated using a one-dimensional projection of the stimulus space. In this way, the potentially complex nature of spatial summation of inputs from across the cell's entire surface may be replaced by a single principal component. With this in mind, it can be hypothesized that the spike-triggered average projections of the linear-nonlinear model are proportional to the transmembrane currents, or the activating function, at the SOCB. If this hypothesis holds, the linear-nonlinear model can thus be thought of as being an accurate reduction to a single-point neuron model.

All three of the above assumptions correspond to a departure from the more common volume conductor approach to modeling extracellular stimulation of the retina. As such, comparison of the output of a volume conductor model that captures (1) whole cell input integration, (2) complex spatiotemporal tissue properties, and (3) conductance-based spiking nonlinearities may be used to test the validity of the linear-nonlinear model assumptions, and is the focus of Chapter 3 of this thesis.

1.4 Optimization of multi-electrode stimulus currents

The electric fields generated by stimulation from multiple adjacent electrodes can interact in complex ways, and have a nonlinear relationship to resulting neural activation. Consequently, unexpected results have been observed when stimulating the retina with simultaneous multi-electrode stimulation (Horsager et al., 2010; Wilke et al., 2011). To avoid complications arising from electrode interactions, stimulation of the retina has typically been performed using sequential single-electrode stimulation, despite the larger variety of retinal responses that may be elicited from multi-electrode stimulation.

However, recent theoretical and experimental evidence suggests substantial benefit from more complex, simultaneous stimulation strategies (Opie et al., 2013; Matteucci et al., 2013; Spencer et al., 2016; Jepson et al., 2013). Current steering involves the recruitment of multiple stimulating electrodes

to activate regions of neural tissue that could not normally be activated under single-electrode stimulation. Different current steering approaches have been developed for activating smaller regions of tissue (known as *focal activation* or *current focusing*) (Opie et al., 2013; Matteucci et al., 2013; Spencer et al., 2016; Jepson et al., 2013) and for activating regions in between electrodes (known as *virtual electrodes*) (Spencer, 2018; Dumm et al., 2014; Loizos et al., 2016).

Although approaches such as these show the potential of simultaneous multi-electrode stimulation for improving clinical outcomes, they do not provide a strategy for selecting electrode stimuli to achieve complex/real-world spatial activation patterns. Moreover, using heuristic approaches is likely to become intractable when attempting to develop effective stimulation strategies for large numbers of electrodes due to the large number of potential stimuli that must be considered.

An alternative approach for multi-electrode stimulation uses computational models of neural activation under multi-electrode stimulation that allow for model inversion, thereby allowing for the selection of stimuli that achieve desired neural activation patterns. The optimization of multi-electrode stimulation to best achieve desired spatial patterns of activation in the retina is here termed *spatial shaping* (Meffin et al., 2013b). Spatial shaping utilizes models of spiking probability across the retina in response to simultaneous multi-electrode stimulation. Through efficient inversion of such models of neural responses, spatial shaping aims to find optimal sets of stimulus currents given target retinal activation patterns. This approach has the potential to select optimal stimulus currents for large numbers of electrodes given complex target activation patterns. Furthermore, an appropriate solution to spatial shaping should be able to recover previously validated current steering results via the presentation of appropriate stimuli.

As evidenced by the structure of both the biophysical volume conductor and phenomenological linear-nonlinear models introduced above, electrical stimulation of the neural tissue is best represented by a combination of linear and nonlinear components. In the case of volume conductor models, the linear component of the model represents the extracellular electric field generated by multiple electrodes and the nonlinear component represents the cellular integration of inputs and spiking behavior of an active neuron. Approaches to the optimization of stimulus currents via model inversion aim to optimize either (1) the output of the linear part of the system or (2) the output of the nonlinear part of the system (i.e., the complete model).

1.4.1 Linear programming approaches to spatial shaping

The advantage of inverting only the linear component of the system is that linear programming techniques can be applied, significantly simplifying model inversion. A simple approach is to invert a linear model of the electric field in a volume of the retina. Since the model is linear, it can be represented as a pre-calculated matrix, \mathbf{A} , where \mathbf{A} is a four-dimensional matrix with three spatial dimensions and one dimension for the set of electrodes. Commonly, the matrix \mathbf{A} is flattened into a two-dimensional representation to ease manipulation. Each element in \mathbf{A} is the electric field induced at a particular three-dimensional location by nominal stimulation (e.g., $1 \mu\text{A}$) from a particular electrode. Linearity dictates that the electric fields generated by different electrodes superimpose linearly for a given stimulus. Hence, given a set of electrode stimulus currents, \vec{S} , the resulting extracellular electric field at

every point in the volume of interest can be calculated using an equation of the form,

$$\mathbf{V}_e = \mathbf{A}\vec{S}. \quad (1.20)$$

To find an optimal set of stimulus currents, \vec{S} , given a desired electric field, \mathbf{V}_e , the model is inverted by finding the solution to the optimization problem,

$$\underset{\vec{S}}{\text{minimize}} \quad \|\mathbf{A}\vec{S} - \mathbf{V}_e\|^2. \quad (1.21)$$

This problem can be solved in closed form using ordinary least squares (Gui-Jing et al., 1981).

In practice, optimization of electrode currents should be constrained such that selected currents remain with clinical safety bounds. As such, the optimization problem becomes

$$\begin{aligned} &\underset{\vec{S}}{\text{minimize}} \quad \|\mathbf{A}\vec{S} - \mathbf{V}_e\|^2 \\ &\text{subject to} \quad \vec{C}_1 < \vec{S} < \vec{C}_2, \end{aligned} \quad (1.22)$$

where \vec{C}_1 and \vec{C}_2 are sets of lower and upper bounds, respectively, on electrode current amplitudes. This optimization problem can be solved by several efficient iterative linear programming techniques such as the interior-point algorithm (Potra and Wright, 2000) or the trust-region-reflective algorithm (Coleman and Li, 1996). A similar approach has been applied with some success in the application of transcranial direct current stimulation (Dmochowski et al., 2011).

The main drawback of this approach is that, since it only inverts the linear portion of the system, it does not directly optimize nonlinear neural activation. This is problematic since the assumption that a linear approximation correlates well with activation is generally not well-founded. In an attempt to account for this shortcoming, a method has been proposed in which a mixed integer linear programming approach is used to optimize the set of activated locations in the tissue (Abouelseoud et al., 2018). This method uses a binary approximation of neural activation, by comparing the achieved linear electric field/current density to an activation threshold. Based on this approximation, an iterative optimization algorithm is used to find electrode currents that result in the least number of locations that are incorrectly activated. Although this approach moves closer to the direct optimization of stimulus currents to desired activation patterns, several limitations remain.

The simplified binary representation of neural activation fails to capture the probabilistic relationship between generating signals (such as the electric field) and activation (Jepson et al., 2014; Escola et al., 2011), which are observed even in single cell recordings (Maturana et al., 2016). Furthermore, in the application of inversion strategies to retinal prostheses, it will be necessary to map recorded images (e.g., grayscale images) to desired patterns of activation. This is precluded by the use of a binary model for all but black-and-white (i.e., activated or not) images.

In addition, methods such as mixed integer linear programming are only suitable for monophasic stimulation. As described by (Abouelseoud et al., 2018), the proposed optimization algorithm is guaranteed to find the globally optimal solution. However, this is the case because only a single phase of the stimuli were considered, even for biphasic stimuli. For biphasic stimuli, as is modeled by the linear-nonlinear model described above, the potential number of solutions grows dramatically and significantly complicates model inversion due to the possibility of excitation in either phase.

Finally, the mixed integer linear programming approach was shown to supply updated optimized stimuli on time frames in the order of 10 minutes for relatively simple target activation patterns (Abouelseoud et al., 2018). For an approach to spatial shaping to be clinically feasible for retinal prostheses, it will need to supply updated optimized stimulus currents on sub-second intervals. To our knowledge, there are no existing methods for the inversion of nonlinear models of neural activity on this time scale and no methods that shape responses to complex target activation patterns, such as real-world images.

1.4.2 Spatial shaping using the linear-nonlinear model

The utility of models such as that of Maturana et al. (2016) for optimizing electrical stimulation offers exciting opportunities. Although biophysical models have existed for some time that are able to accurately characterize RGC responses to electrical stimulation (Rattay and Resatz, 2004; Werginz et al., 2014; Cao et al., 2015; Greenberg et al., 1999), they have been too computationally demanding or complex to be inverted in real-time solutions. In addition, these models are extremely difficult to constrain using experimental data owing to the large number of free parameters that govern dynamics (Meffin et al., 2012, 2014). The simplicity of the linear-nonlinear model and the low number of parameters required to be fit makes it suitable for implementation in an inversion system aimed at optimizing stimulus parameters. Furthermore, it has recently been demonstrated that multi-electrode retinal stimuli informed by the linear-nonlinear model produce predictable patterns of cortical activation in response (Halupka, 2017).

Since the linear-nonlinear model can predict spiking activity patterns in response to arbitrary combinations of electrode stimulus amplitudes, it is also well suited to the selection of stimulation amplitudes through model inversion, given desired activity patterns. However, inversion of the linear-nonlinear model is not straightforward due to the nonlinearity of the double-sided sigmoidal function, which is required to model responses to biphasic stimulation. Due to its non-monotonicity, the inversion of the linear-nonlinear model is an ill-posed problem, with many potential solutions to a given inversion. Chapter 4 of this thesis deals with methods of solving this problem.

To allow for the true optimization of stimulus currents for retinal prostheses, it is necessary to develop a process for the optimal, or near-optimal, inversion of the double-sided linear-nonlinear model. In addition, this inversion strategy must meet the following requirements:

1. Inversion should be efficient enough to be performed in real time with a portable device.
2. Responses should be shaped to represent complex target images or moderate simplifications of those images.
3. Stimulus currents should be constrained to ensure clinical safety and limit power consumption.
4. The condition of *optimality* should include as much available information as possible, including the resulting patterns of activation in multiple retinal layers (such as the NFL and the GCL) and multiple neuron populations (such as ON and OFF RGCs).

1.5 Conclusion

This chapter presented an overview of two main approaches to the simulation of electrical stimulation of the retina: biophysical models and block-structured models. Two models were explored in detail based on their relevance to the current research outcomes: the cellular composite model (Meffin et al., 2012; Tahayori et al., 2012; Meffin et al., 2014; Tahayori et al., 2014) and the linear-nonlinear model (Maturana et al., 2016). Each model may be applied in different ways, with the biophysically-accurate cellular composite model providing a means for the investigation of the implications of retinal structure and the linear-nonlinear model providing a means for the optimization of complex multi-electrode stimulation parameters.

As described in Section 1.1.5, an ongoing issue with epiretinal stimulation of the retina is the stimulation of overlying axons of passage. This phenomenon limits the efficacy of epiretinal prostheses due to the creation of irregular and confusing percepts for patients. The cellular composite model can capture both the effect of the microscale anatomy of neural tissue and the mean effect of the electrical properties (resistive and capacitive) of its cellular constituents. By first constraining the cellular composite model to the specific tissue geometry of relevant retinal tissue layers before integrating it into a multi-layered model of the retina, a computational modeling approach can be used to understand the problem of stimulation of unintended targets. Furthermore, through an exploration of the stimulus space, this method can potentially elucidate stimulation parameter combinations that minimize or eliminate the problem, improving clinical outcomes for epiretinal prostheses. This forms the first aim of this project.

For retinal prostheses in general, following the establishment of electrical stimulation as a promising treatment for diseases such as age-related macular degeneration and retinitis pigmentosa, a crucial milestone will be the ability of devices to effectively construct stimulation schemes that reproduce complex inputs in the pattern of neural activation across the retina. This is needed to ensure that the fidelity of perception for prostheses recipients can improve to more desirable levels. One approach to constructing these stimulation patterns is through the inversion of models of neural response. The linear-nonlinear model of Maturana et al. (2016) accurately predicts spiking responses from extracellular electrical stimulation of the retina. The simplicity afforded by the one-dimensional stimulus projection ensures this model is well-suited to this problem and, as such, it is important that the model's assumptions are sound. The second phase of this project aims to apply a volume conductor model that combines a multi-layered implementation of the cellular composite model with morphologically-accurate compartmental neuron models to explore the validity of the linear-nonlinear model assumptions. Using the method described in Section 1.2.3, the simulated activating function (stimulation-induced transmembrane current) will be compared to the simulated ERF of the simpler linear-nonlinear model. Following model validation, numerical inversion techniques will be applied to the linear-nonlinear model to optimize multi-electrode stimulation parameters, forming the final objective of this work.

1.6 Thesis overview

This research aims to address several outstanding challenges in electrical stimulation of the retina that were highlighted in the previous section. The problem of stimulation of RGC axons of passage dur-

ing epiretinal stimulation suggests the need for a greater understanding of the way electrical current passes through the first several layers of the retina, including the effect of the specific anatomical structures and physiological properties of each layer. This prompts the development of more sophisticated computational models of this tissue that can explore the effects of relevant temporal, orientation and geometrical differences in the retina.

The challenge of developing methods for the overall shaping of retinal activity in response to multi-electrode array stimulation requires the development of parsimonious models of neural activation that lend themselves to model inversion and optimization techniques. Model inversion offers a method for calculating optimal stimulation waveforms for achieving a given neural activation pattern, as opposed to inferring approximate stimulation patterns from the output of simulations of neural activation or from experimental recordings. A preliminary step in the development of such methods is the validation of these simpler models, intended for spatial shaping, against more detailed biophysical computational models and against experimental data. This research focuses primarily on direct stimulation of retinal ganglion cells, however, generalization of findings to indirect or network mediated stimulation is discussed where relevant.

In Chapter 2 of this thesis, we demonstrate that preferential activation of axon initial segments can be achieved by simultaneous stimulation with multiple electrodes along the length of passing axons. A present challenge in electrical stimulation with epiretinal prostheses is the avoidance of stimulation of axons of passage in the nerve fiber layer (NFL) that originate from distant regions of the ganglion cell layer (GCL) (Jensen et al., 2003; Werginz and Rattay, 2015; Rizzo et al., 2003b; Jepson et al., 2013; Rattay and Resatz, 2004; Werginz et al., 2014; Cao et al., 2015; Weitz et al., 2013). When attempting to stimulate RGCs in the ganglion cell layer, collateral stimulation of overlying axons results in irregular visual percepts, which can significantly limit perceptual efficacy. Multi-layered, biophysical models of the retina are employed that capture the characteristic distributions of fiber orientation in each retinal layer. Taking advantage of these distributions, stimulation methods are demonstrated that preferentially activate target RGC structures and volumes. Models used in this research build on novel single-layered models of electrical stimulation developed by Meffin et al. (2012, 2014) and Tahayori et al. (2012, 2014) that have been shown to capture the microscopic structure and physiology of neural tissue more accurately.

Chapter 3 presents a validation of linear-nonlinear models of neural response against a biophysically-detailed model of electrical stimulation of the retina. Recent work has shown the effectiveness of linear-nonlinear models in predicting neural responses to electrical stimulation with subretinal multi-electrode arrays. These models, such as that of Maturana et al. (2016), utilize linear projections of stimulating electrode amplitudes onto a low-dimensional subspace via principal component analysis, yielding the linear electrical receptive field of the RGC. The projection then undergoes a nonlinear transformation to yield spiking probability. Experimental procedures used to generate cell responses are simulated using a combination of a macroscopic current flow model developed in Chapter 2 and conductance-based, compartmental neuron models. Multi-layered biophysical models of the retina are used to determine the stimulus-induced transmembrane current flow (activating function) at each neural compartment. Through a comparison of the activating function at various neural structures and the ERF, the equivalence of the two quantities is assessed. This method allows for validation of the hypothesis that the linear ERF of the RGC is determined predominantly by a linear function of input currents at and near the site of action potential initiation.

In chapter 4, we develop and validate methods for the spatial shaping of neural activity in the retina through numerical inversion of the linear-nonlinear model. A proposed strategy for improving electrical stimulation of the retina involves the optimization of stimulus parameters to achieve a desired activity profile across the retinal plane. This chapter of the thesis focuses on the inversion of accurate models of spiking activity in response to biphasic multi-electrode array stimulation to determine optimal sets of stimulus parameters for desired spatial responses. Through the translation of the linear-nonlinear model into an equivalent artificial neural network, procedures are demonstrated for the calculation of stimulus amplitudes across multi-electrode arrays that result in an optimal match between the evoked pattern of spatial activity and a target pattern. An efficient and optimal strategy for the inversion of the linear-nonlinear is demonstrated for both experimentally-estimated ERFs, as in Maturana et al. (2016), and for biophysically-estimated ERFs, based on the outcome of Chapter 3.

Chapter 2

Minimizing activation of overlying axons

This content of this chapter is a slightly modified version of the following published journal article:

*Timothy B. Esler, Robert R. Kerr, Bahman Tahayori, David B. Grayden, Hamish Mef-
fin, Anthony N. Burkitt. (2018) "Minimizing activation of overlying axons with epiretinal
stimulation: The role of fiber orientation and electrode configuration." PLOS ONE 13(3):
e0193598.*

2.1 Abstract

Currently, a challenge in electrical stimulation of the retina with a visual prosthesis (bionic eye) is to excite only the cells lying directly under the electrode in the ganglion cell layer, while avoiding excitation of axon bundles that pass over the surface of the retina in the nerve fiber layer. Stimulation of overlying axons results in irregular visual percepts, limiting perceptual efficacy. This research explores how differences in fiber orientation between the nerve fiber layer and ganglion cell layer leads to differences in the electrical activation of the axon initial segment and axons of passage. *Approach.* Axons of passage of retinal ganglion cells in the nerve fiber layer are characterized by a narrow distribution of fiber orientations, causing highly anisotropic spread of applied current. In contrast, proximal axons in the ganglion cell layer have a wider distribution of orientations. A four-layer computational model of epiretinal extracellular stimulation that captures the effect of neurite orientation in anisotropic tissue has been developed using a volume conductor model known as the cellular composite model. Simulations are conducted to investigate the interaction of neural tissue orientation, stimulating electrode configuration, and stimulation pulse duration and amplitude. *Main results.* Our model shows that simultaneous stimulation with multiple electrodes aligned with the nerve fiber layer can be used to achieve selective activation of axon initial segments rather than passing fibers. This result can be achieved while reducing required stimulus charge density and with only modest increases in the spread of activation in the ganglion cell layer, and is shown to extend to the general case of arbitrary electrode array positioning and arbitrary target volume. *Significance.* These results elucidate a strategy for more

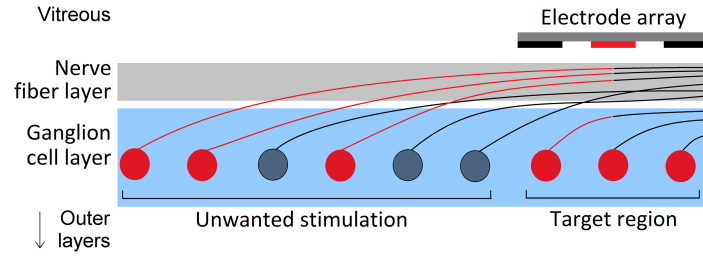


Figure 2.1: **Unwanted stimulation of retinal ganglion cell axons of passage.** Retinal ganglion cell somas and axon initial segments represent the target regions for epiretinal stimulation (region shaded blue). Activation of passing axons in the nerve fiber layer (gray shaded region) results in long, arc-shaped visual percepts and degradation of the quality of artificial vision. Retinal ganglion cell axon bundles in the nerve fiber layer that pass close to stimulating electrodes may be stimulated preferentially to target locations in the ganglion cell layer. Activated retinal ganglion cells are colored red from the soma to the location of axon potential initiation. Note that the orientation of initial axonal segments is much more varied in reality than shown in this schematic.

targeted stimulation of retinal ganglion cells with experimentally-relevant multi-electrode geometries and achievable stimulation requirements.

2.2 Introduction

There has been significant progress over the past decade in the development of retinal prostheses for those with retinal pathologies such as Retinitis Pigmentosa. Clinical trials of retinal prostheses have found that patients can reliably report visual percepts arising from stimulation and can perform simple identification tasks (Ayton et al., 2014; Werginz and Rattay, 2015; Weiland et al., 2011; Rizzo et al., 2003a; Zrenner et al., 2011; da Cruz et al., 2013; Fujikado et al., 2011). Although progress to date is highly encouraging, many aspects of the performance of retinal prostheses remain limited, hinging on the ability of these devices to target either specific retinal cell types (Twyford et al., 2014; Jensen and Rizzo, 2006) or more precise retinal volumes (Grosberg et al., 2017; Werginz and Rattay, 2015; Rizzo et al., 2003a; Rattay and Resatz, 2004). In the case of epiretinal stimulation, a factor limiting performance is the inability of electrical stimulation to preferentially activate target neuronal structures in the ganglion cell layer (GCL), such as the axon initial segment (AIS), while avoiding activation of overlying axons in the nerve fiber layer (NFL) (Grosberg et al., 2017; Jensen et al., 2003; Werginz and Rattay, 2015; Rizzo et al., 2003b; Jepson et al., 2013; Rattay and Resatz, 2004; Werginz et al., 2014; Cao et al., 2015; Weitz et al., 2013; Abramian et al., 2015; Jeng et al., 2011), illustrated in Figure 2.1.

Since the axons of retinal ganglion cells (RGCs) traverse the inner surface of the retina in the NFL, epiretinal electrical stimulation faces the challenge of stimulating the deeper, favorably-organized GCL while minimizing activation of axons of passage (AOPs) in the NFL. Recipients of epiretinal implants commonly describe irregular visual percept shapes due to stimulation of axons of passage (Werginz et al., 2014; Rizzo et al., 2003b; Nanduri et al., 2008; Chen et al., 2009). This effect has been confirmed experimentally and in simulations, and results in a reduction in the spatial selectivity of epiretinal stimulation (Grosberg et al., 2017; Weitz et al., 2013; Rattay and Resatz, 2004; Greenberg et al., 1999; Werginz et al., 2014; Cao et al., 2015; Nanduri et al., 2008; Chen et al., 2009). Recent experimental

findings by Grosberg et al. (2017) have both reaffirmed the existence of this problem while suggesting that it may be overcome using electrode stimulus amplitudes carefully tuned via detection of RGC activation in response to stimulation.

The AIS is typically defined as the section of axon beginning from the soma or axon hillock and extending for up to 100 μm . Owing to the presence of an experimentally-observed, high-density sodium channel band (SOCB) within or near the AIS, it has been shown to be the most excitable part of a RGC (Fried et al., 2009). Hence, it is important to compare the activation of AOPs with the AIS, as opposed to the soma. A potential way to minimize activation of AOPs is to take advantage of differences in neurite orientation in the NFL and GCL. The direction of overlying axon tracts represents the dominant fiber orientation in a given location in the NFL. These axons are packed together as mostly parallel fibers (Grosberg et al., 2017; Rattay and Resatz, 2004; Rizzo et al., 2003b). As a result, current flow from epiretinal electrical stimulation spreads through retinal tissue in a highly anisotropic way. In contrast to the distal RGC AOPs in the NFL, proximal axon regions, such as the AIS located in the GCL, have a much wider distribution of orientations as they pass out from the soma. Based on these anisotropic tissue characteristics, it is expected that the orientation of a neurite in retinal tissue can have a significant effect on its activation. However, a common approximation employed by existing computational models of epiretinal stimulation is that the retinal layers are isotropic (Abramian et al., 2015; Rattay and Resatz, 2004; Cao et al., 2015; Eiber et al., 2016). In order to assess the effect of neurite orientation and its interaction with different multi-electrode configurations, computational models of current flow and axonal activation must be developed that can describe the anisotropic characteristics of key retinal layers.

In the absence of sufficient data to model the anisotropy of the NFL, an alternative approach is to derive layer anisotropy from first principles using a geometric description of the axonal units that comprise the tissue. The cellular composite model, introduced by Meffin et al. (Meffin et al., 2012; Tahayori et al., 2012; Meffin et al., 2014; Tahayori et al., 2014), provides a modeling framework that accomplishes this while addressing a number of limitations of conventional volume conductor models. Common two-stage approaches, such as those that combine a linear, numerically-integrated volume model for extracellular voltage (typically simulated in COMSOL Multiphysics) with a conductance-based neuron model (typically simulated in NEURON), are confounded by inconsistencies between the two simulation stages (Meffin et al., 2013a; Tahayori et al., 2014). This effect was exemplified by Tahayori et al. (2014), who showed that, in the presence of model inconsistency, the choice of simulating either extracellular current or extracellular voltage could result in changes in simulated membrane potential by up to an order of magnitude. This is caused by a discrepancy between the tissue impedance underlying the two simulation stages. To more accurately capture the structural and temporal properties of neural tissue and to guarantee model self-consistency, the cellular composite model maps extracellular current to voltage using an expression for impedance derived directly from the geometry and physiology of the tissue’s microscopic constituent axons, providing consistent descriptions of both extracellular voltage/current and neural activation. Here, we present a multi-layered generalization of the cellular composite for which a closed-form solution exists in the Fourier domain. This solution yields modeling results that are more readily interpretable than three-dimensional finite-element model simulations. Furthermore, the relative computational efficiency of this approach allows for large-scale parameter sweeps. Although not explored in this research, the simulation approach presented here also lends itself to (closed-form) model inversion, which can be applied to studying the inverse stimulation problem: determining optimal stimulus currents given desired tissue activation

patterns.

In addition to intrinsic tissue anisotropy, RGC activation will also depend on the orientation of the applied electric field. One existing modeling study by Rattay and Resatz (2004) assessed the influence of electric field orientation with respect to neurites in the NFL. This study showed that, by orientating long, rectangular electrodes parallel to axons in the NFL, the activation of those axons could be reduced. The basis for this result is that the membrane potential response of an axon to extracellular stimulation is largely determined by the activating function: the second spatial derivative of the extracellular potential along the axon’s length (Rattay, 1999). By ‘flattening’ the extracellular potential along the length of the axon using long parallel electrodes, the activation of the axon is minimized.

The aim of this study is to demonstrate a multi-electrode stimulation strategy for the avoidance of activation of axon bundles, while achieving focal activation of axon initial segments in the GCL. We present a model that captures both the effect of electric field orientation imposed by multi-electrode stimulation and the effect of the highly anisotropic geometry of the NFL. Increased sensitivity to stimulation at the AIS attributed to the high-density sodium channel band (Fried et al., 2009) has been incorporated into the model via an adjustment of threshold membrane potentials. Simulation results are presented that illustrate the achievable levels of preferential activation for one-, two-, and four-electrode configurations. An exploration of the effect of electrode-retina separation distance and pulse duration are presented, as well as the effect of different strategies on key performance metrics: required stimulus charge, GCL activation, and activation radius. The proposed multi-electrode array strategy is then validated against a more general set of electrode geometries and target volumes.

2.3 Methods

2.3.1 Distribution of orientations in the ganglion cell layer

To quantify the distribution of proximal axon orientations in the GCL, we analyzed mammalian RGC reconstructions obtained from the NeuroMorpho.org database (Ascoli et al., 2007; Badea and Nathans, 2011; Coombs et al., 2006; Fletcher et al., 2014; Kong et al., 2005; Toris et al., 1995; Poria and Dhingra, 2015; Rodger et al., 2012; Sümbül et al., 2014; Lee and Stevens, 2007; Snider et al., 2010; Mazzoni et al., 2008). At the time of analysis, 749 of the available cell reconstructions included at least 100 μm of the cell’s axon and so were included in our calculations. It is challenging to determine the distribution of RGC subtypes in the sample due to inconsistent classification schemes across studies that span many years and different methodologies. However, we expect that the sample used is representative of the overall RGC population since the majority of cells come from morphological classification studies of the whole population. This is appropriate for the current study as simulations only consider axonal orientation and not cell morphology, so general RGC properties are adequate. In addition, the modeling approach used in this study accounted only for axonal orientation and not direction. Hence, differences in the propagation of activity in different directions was neglected. Cell morphologies were imported and processed in MATLAB (The Mathworks, Release 2016a) with the third-party TREES toolbox (Cuntz et al., 2010). The change in orientation between the high-density sodium channel band of the AIS (defined as the segment from 40 μm to 80 μm from the soma (Fried et al., 2009)) and various locations along the axon was calculated. Figures 2.2(a) and 2.2(b)

show the proportion of cells with orientation in different ranges. For each cell, the orientation was calculated as the angle between the AIS and the axon, and was measured at axonal locations 100, 300, and 500 μm from the soma. Since each cell has a different length of axon included in the morphological reconstruction (due to differing imaging and sample preparation limitations) different cell sub-populations were available for the distribution at each axonal location. All cells (749) had at least 100 μm of reconstructed axon, 158 cells had at least 300 μm , and 44 had at least 500 μm .

As shown in Figure 2.2(a), the orientation in the x - y plane approaches a uniform distribution for locations 500 μm (or more) distal from the soma. This was validated statistically by comparing the goodness-of-fit of uniform and exponential distributions for each location using likelihood-ratio tests. Exponential fits to the distribution in Figure 2.2(a) are shown in Figure 2.2(b) with 95% confidence bounds. To account for the fact that fewer cells were available for orientation measurements at 500 μm from the soma, likelihood was calculated for 100 μm ($n = 749$) and 300 μm ($n = 158$) by averaging fits across 1000 random samples of size $n = 44$. This test showed that an exponential fit was more appropriate for the orientation at 100 μm ($p < 0.001$) and 300 μm ($p = 0.0251$), whereas the distribution of orientations at 500 μm was consistent with a uniform fit ($p = 0.1078$). Similar results were found using alternative metrics such as the Akaike information criterion. In contrast, Figure 2.2(c) shows that there is little change in orientation between the AIS and the distal axon in terms of altitudinal orientation indicating that, beyond the AIS, axons remain predominantly parallel to the surface of the retina. Statistically, an exponential fit was more appropriate than uniform for all distributions in Figure 2.2(c). Based on the knowledge that fibers in the NFL are approximately parallel to each other at a given location (Grosberg et al., 2017; Rattay and Resatz, 2004; Rizzo et al., 2003b), this analysis suggests an approximately circular (but not spherical) uniform distribution is an appropriate approximation for the orientation of AISs in the GCL. A circular distribution will be used in the remainder of this chapter.

2.3.2 Tissue geometry and governing equations

The model employed here uses a two-stage volume conductor framework. The first stage models the electric field induced by the stimulating electrodes. The second stage uses the calculated extracellular potential from the first stage as input into a passive neurite model to calculate membrane potential. The present modeling approach uses a four-layer description of retinal geometry for stage 1 (Figure 2.3). The modeled layers are the insulating substrate of the electrode array, the vitreous, the nerve fiber layer, and an approximation of the remaining retinal layers, including the ganglion cell layer. The conductivity/admittivity and directional dependence properties of each layer are presented in Table 2.1. Admittivity is a spatially- and temporally-dependent generalization of conductivity and is the inverse of impedivity, containing both resistive (real) and reactive (imaginary) parts. The anisotropic admittivity of the NFL is incorporated into the complex admittivity kernel provided by the cellular composite model of Meffin et al. (2014).

The description of anisotropy/isotropy of each layer is based on a mean-field approximation of the cells that comprise the tissue in that layer. The nerve fiber layer has a markedly anisotropic geometry as it is composed of largely parallel axon bundles. Due to the mix of different cell types and the presence of cell bodies, the GCL and remaining outer retinal layers can be approximated as isotropic on a macro-scale (i.e., isotropic when averaged over some distance). Additionally, it has been shown previously that a combined model of the GCL and outer retinal layers yields approximately

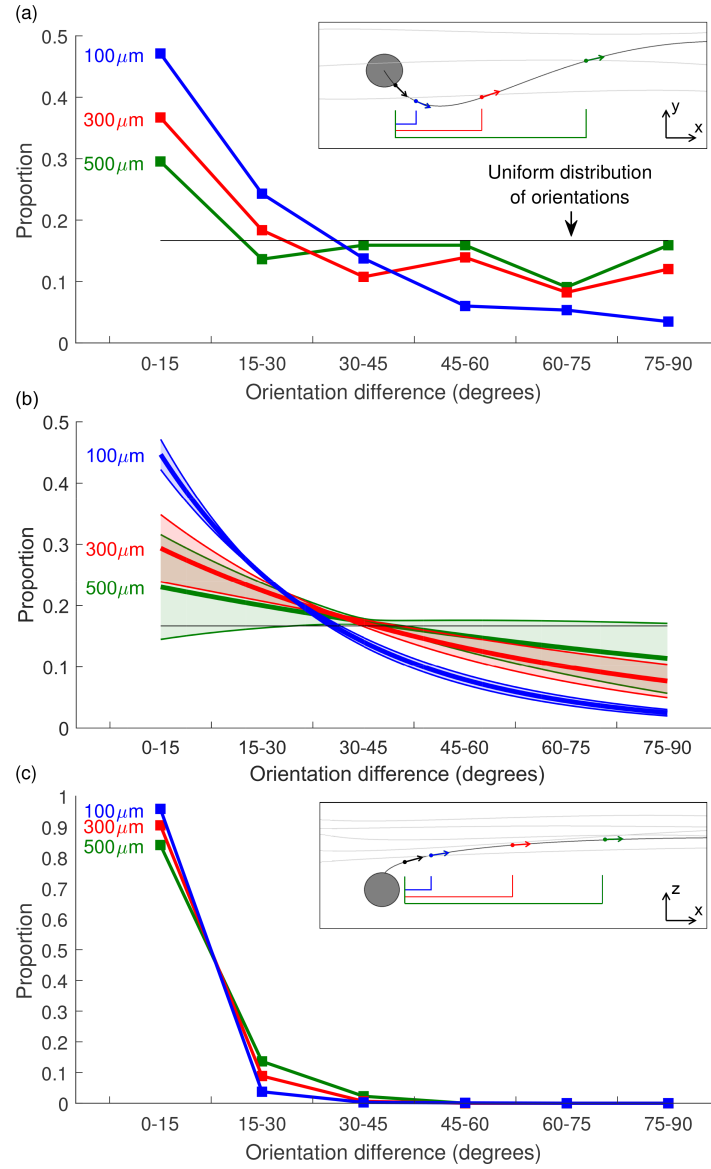


Figure 2.2: **Fiber orientation distributions along the length of the axon.** Fiber orientation is calculated relative to the axon initial segment, and determined from all available mammalian retinal ganglion cell reconstructions on NeuroMorpho.org. The distribution of orientations at axonal locations of 100 μm (blue), 300 μm (red), and 500 μm (green) from the soma are shown. (a) Azimuthal (i.e., x - y) change in orientation between the axon initial segment and more distal axonal locations. (b) Exponential fits to the distributions in (a) with 95% confidence intervals. Statistical analysis of these fits is discussed in the main text. (c) Altitudinal (i.e., z) change in orientation between the axon initial segment and more distal axonal locations, with all following an exponential distribution. Insets illustrate planes in which orientations are compared. All orientations are calculated relative to the orientation of the axon initial segment. Due to variation in the length of axon reconstructions, each trace is calculated using a different subset of cells (100 μm - all 749 cells, 300 μm - 158 cells, 500 μm - 44 cells).

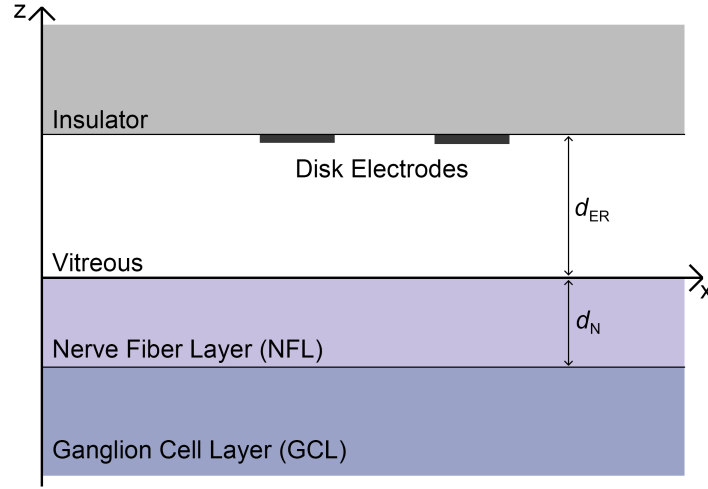


Figure 2.3: **Geometry of the four-layer model of the retina.** Modeled layers are the insulator, vitreous, nerve fiber layer, and ganglion cell layer. The insulator is assumed to have zero conductivity and is modeled using a zero flux boundary condition. The GCL is assumed to have infinite extent in the z -direction. The distance from electrodes to the retinal surface and the thickness of the NFL are denoted by d_{ER} and d_N , respectively.

Table 2.1: Conductivity and thickness of modeled layers

Layer	Directional dependence	Conductivity (S/m)	Thickness (μm)
Insulator	Isotropic	0	Infinite extent
Vitreous	Isotropic	1.78*	10-500 (based on electrode placement)
NFL	Anisotropic	see A.2	100*
GCL	Isotropic	0.1 [†]	Infinite extent

*Mazzoleni et al. (1986)

^Oishi et al. (2009); Frenkel et al. (2005); Wang et al. (2013); Thomson et al. (2015)

[†]Tahayori et al. (2014); Gabriel et al. (2009); Latikka et al. (2001); Ranck (1963); Kasi et al. (2011); Wang and Weiland (2015); Karwoski et al. (1985)

equivalent results to a more sophisticated five-layer model of these layers (Werginz et al., 2014).

Due to its (approximately) zero conductivity, the insulator layer is included via a zero current boundary condition applied at the insulator-vitreous interface. Electrodes are modeled as two-dimensional, circular disks lying on this boundary. In this study, we neglect the formation of an electrical double layer at the electrode-vitreous boundary and any associated voltage drop in this region. Experimental and simulation studies have shown that, given the distances between the electrode array and tissue that we consider here, the impact of an electrical double layer is negligible in magnitude as well as being both highly transient and highly spatially localized (Behrend et al., 2008; Abramian et al., 2015, 2011).

The flow of current in the extracellular space in each layer is described by a separate Poisson-type equation, allowing for differing tissue admittivities in each layer, with the driving current delivered by disk electrodes entering as an explicit term on the right-hand-side of the vitreous layer continuity

equation:

$$\nabla \cdot \mathbf{J}_V(x, y, z, t) = \sum_i^M \frac{I_i(t)}{\pi q^2} g_q(x - x_i, y - y_i) \delta(z - z_i), \quad (2.1a)$$

$$\nabla \cdot \mathbf{J}_N(x, y, z, t) = 0, \quad (2.1b)$$

$$\nabla \cdot \mathbf{J}_G(x, y, z, t) = 0, \quad (2.1c)$$

where subscripts V, N, and G associate quantities with the vitreous, NFL, and GCL, respectively. \mathbf{J}_α is the extracellular current density in layer α and each set of (x_i, y_i, z_i) represents the three-dimensional location of one of the M electrodes. Each electrode has radius q and stimulus current waveform $I_i(t)$. The function $g_q(x, y)$ is the unit circular step function of radius q in the x - y plane and $\delta(z)$ represents the Dirac delta function. $\nabla = [\partial/\partial x; \partial/\partial y; \partial/\partial z]$ is the differential operator. For this model, each layer boundary is approximated by an infinite flat plane parallel to the x - y plane, so that the set of electrode heights, z_i , are equal. Furthermore, if the origin is fixed on the vitreous-NFL boundary, then z_i is equivalent to the electrode-retina separation distance, d_{ER} .

A generalized form of Ohm's Law is used to describe extracellular current density and potential, which is governed by each layer's admittivity kernel. This admittivity kernel incorporates the dependence of the extracellular current density on the electric field at previous times and at remote locations in the extracellular space. These atypical dependencies arise due to the passage of current across the cellular membrane and through the intracellular space. The relationship between extracellular potential and current density is described by

$$\begin{aligned} \mathbf{J}_\alpha &= -\frac{1}{4\pi^2} \xi_\alpha(x, y, z, t) * \nabla \phi_\alpha(x, y, z, t) \\ &= -\frac{1}{4\pi^2} \iint_{\mathbf{r}' t'} \xi_\alpha(\mathbf{r}', t') \nabla \phi_\alpha(\mathbf{r} - \mathbf{r}', t - t') d\mathbf{r}' dt', \end{aligned} \quad (2.2)$$

where ξ_α is the 3x3 admittivity kernel and ϕ_α is the extracellular potential of layer $\alpha \in \{V, N, G\}$. In the most general case, where ξ_α varies in three spatial dimensions and time, $*$ represents a convolution over three spatial dimensions and time. For brevity, the spatial coordinates (x, y, z) have been represented by the vector \mathbf{r} in the integral expression for the convolution.

For layers with infinite extent in the x - and y -directions, as in the present model, boundary conditions are specified at the layer boundaries:

$$\phi_G|_{z=-\infty} = 0, \quad (2.3a)$$

$$\phi_N|_{z=-d_N} = \phi_G|_{z=-d_N}, \quad (2.3b)$$

$$\phi_V|_{z=0} = \phi_N|_{z=0}, \quad (2.3c)$$

$$\mathbf{J}_{Nz}|_{z=-d_N} = \mathbf{J}_{Gz}|_{z=-d_N}, \quad (2.3d)$$

$$\mathbf{J}_{Vz}|_{z=0} = \mathbf{J}_{Nz}|_{z=0}, \quad (2.3e)$$

$$\mathbf{J}_{Vz}|_{z=d_{ER}+d_{EI}} = 0. \quad (2.3f)$$

These boundary conditions ensure that the described system has finite energy (Equation (2.3a)), that current density and voltage vary continuously across layer boundaries (Equations (2.3b-e)), and that no current can flow into the insulating substrate (Equation (2.3f)).

Since the current sources are at the same z -location as the insulator's zero current condition, we initially define the geometry such that the insulator is separated from the electrodes by some distance, d_{EI} . This is eliminated subsequently by computing the limit from above as d_{EI} goes to zero. As a result, we maintain a zero current boundary condition at the insulator, except for current coming out of the electrode sources, which are modeled as explicit current sources in Equation (2.1a).

Solution of the system of elliptic partial differential equations defined by Equations (2.1) and (2.2) using layer boundary conditions (2.3) yields expressions for the extracellular potential in each layer. To find a closed-form solution to this system, we first assume that within each layer tissue admittivity is independent of z , reducing the above four-dimensional convolutions to three dimensions. Fourier domain approaches are then applied to reduce the convolutions shown in Equation (2.2) to multiplications. Equation (2.1) and Equation (2.2) are transformed into the Fourier domain with respect to x , y , and t . The system is then represented by a system of partial differential equations for which an analytic solution exists of the form

$$\hat{\phi}_V = A_1 e^{-\eta_V z} + A_2 e^{\eta_V z} + \sum_i^M \frac{m_i}{2\eta_V} e^{-\eta_V |z - z_i|}, \quad (2.4a)$$

$$\hat{\phi}_N = B_1 e^{-\eta_N z} + B_2 e^{\eta_N z}, \quad (2.4b)$$

$$\hat{\phi}_G = C_1 e^{-\eta_G z} + C_2 e^{\eta_G z}, \quad (2.4c)$$

where A_1 , A_2 , B_1 , B_2 , C_1 , and C_2 are constants of integration and remaining parameters are defined in A.1, along with a detailed derivation of this solution. The hat symbol ($\hat{\cdot}$) indicates the Fourier transform of the specified quantity with respect to x , y , and t . Quantities m_i and η_α are defined in terms of the Fourier domain pairs of x , y , and t . The form of m_i , shown in A.1, defines both the geometry of the electrodes (the spatial Fourier transform of a disk), and the current stimuli (the temporal Fourier transform of a biphasic square pulse). Stage 1 of the volume conductor model is completed by determining appropriate values for the admittivity or conductivity of each of the modeled layers and is presented in A.2.

2.3.3 Neurite equations

Stage 2 of the cellular composite model involves the calculation of the passive membrane potential in the neurite of interest in either the NFL or the GCL. This is achieved using the neurite equations of Meffin et al. (2012), which provide expressions for membrane activation due to modes of current flow that are both longitudinal ($V_{m,L}$) and transverse ($V_{m,T}$) with respect to the fibers. Expressions for the membrane potential along a single fiber in a fiber-bundle with orientation parallel to the y -axis (as in the NFL) are supplied in the x, y, t -Fourier domain by the cellular composite model,

$$\hat{V}_{m,L}(k_y, \omega; k_x, z) = -\frac{k_y^2 \lambda_V^2(\omega)}{1 + k_y^2 \lambda_V^2(\omega)} \hat{\phi}_\alpha(k_y, \omega; k_x, z), \quad (2.5a)$$

$$\hat{V}_{m,T}(k_y, \omega; k_x, z) = -b \sqrt{\left[-k_x^2 \hat{\phi}_\alpha(k_y, \omega; k_x, z)^2 + \left(\frac{\partial \hat{\phi}_\alpha(k_y, \omega; k_x, z)}{\partial z} \right)^2 \right]}, \quad (2.5b)$$

where k_x , k_y , and ω are the Fourier transform pairs of x , y , and t , respectively. $\hat{\phi}_\alpha(k_y, \omega; k_x, z)$ is the Fourier domain representation of extracellular potential along the neurite axis for a straight neurite oriented parallel to the y -axis at a point (k_x, z) . $\lambda_V(\omega)$ is the frequency-dependent electrotonic length constant for voltage boundary conditions and is defined in A.2. Equation (2.5a) is a Fourier domain representation of the cable equation for extracellular stimulation and indicates the dependence of $V_{m,L}$ on the second spatial derivative of the extracellular potential in the direction of the neurite, known as the activating function (Rattay, 1999). Here, the activating function is represented in the Fourier domain as $-k_y^2 \phi_\alpha$.

A.3 provides an extension of expressions for the longitudinal and transverse components of the membrane potential to straight neurites of arbitrary x - y orientation, allowing for analysis of fibers in both the NFL (fibers with parallel orientation) and GCL (fibers with arbitrary x - y orientation).

2.3.4 Calculating membrane potential thresholds

Several studies have examined the difference in excitability of the AIS and the AOP or axon bundles (Fried et al., 2009; Grosberg et al., 2017). Electrical stimulation experiments conducted by Fried et al. (2009) found that a high-density sodium channel band exists in the RGC AIS. They confirmed the existence of the high-density sodium channel band using both electrical recording of cell responses and immunochemical cell staining with an antibody stain of ankyrin-G that colocalizes with sodium channels. When compared to the soma and the distal axon or AOP, the high-density sodium channel band at the AIS had a greater sensitivity to electrical stimulation. Heightened excitability at the AIS has also been demonstrated in the central nervous system more generally. In the cortex, this has been shown experimentally for pyramidal cells via investigations of the site of action potential initiation (Stuart and Sakmann, 1994; Colbert and Johnston, 1996), action potential initiation thresholds (Coombs et al., 1957; Colbert and Johnston, 1996), and sodium channel density (Colbert and Johnston, 1996).

As ion channels are not modeled in the present model, the difference in excitability of the AIS and AOP is instead captured by adjusting the threshold membrane potential at each location. Threshold potential values for the AIS and AOP have been determined from simulations that replicate the experimental procedures of Fried et al. (2009). By matching the experimental electrode geometry, electrode location, neurite orientation, nerve fiber layer thickness, and stimulation frequency, Fried's experimentally-determined threshold stimulus currents were mapped to corresponding threshold membrane potentials in the computational model presented here. A recent study by Chichilnisky and colleagues (Grosberg et al., 2017) has also demonstrated conditions under which activation of passing axon bundles and the AIS occurs, however we were not able to reproduce these experiments in simulations without cell morphologies and the precise three-dimensional location of the stimulating electrodes relative to the different axon locations.

In order to design simulations that most closely match the experimental methodology, nerve fiber layer thickness, d_N , was set to 25 μm , appropriate for a rabbit retina. A single electrode with a radius, q , of 15 μm was used to deliver a single cathodic-first, biphasic pulse from a location 25 μm from the surface of the retina ($d_{ER} = z_i = 25 \mu\text{m}$). Stimulus pulse amplitudes were chosen to approximately match the stimulus threshold levels reported by Fried et al. (2009). Experimentally reported stimulus current thresholds for the AIS and the distal axon were then used as pulse amplitudes in simulations,

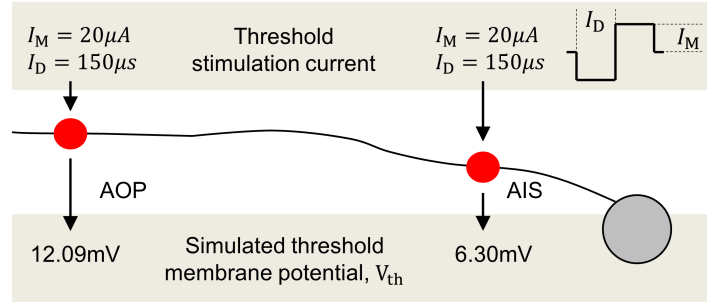


Figure 2.4: **Simulation of experiments from Fried et al. (2009).** The current waveforms used for stimulation and the maximum simulated membrane potential responses are shown. The maximum simulated membrane potential for each simulation corresponds to the membrane threshold, V_{th} , for that location in the axon.

which are illustrated in Figure 2.4. These stimulus current thresholds happened to be approximately $20\ \mu A$ for both the initial and distal axon. This is due to the fact that, although the AIS has a markedly lower threshold membrane potential (not a lower *stimulus* current threshold), the AOP is closer to the stimulating electrode. In this analysis, the AIS was assumed to be $5\ \mu m$ below the surface of the GCL and the AOP was assumed to be centered in the NFL, $12.5\ \mu m$ below the retinal surface. The experimental procedure of Fried et al. used narrow conical electrodes with no backing insulator and so the insulator layer was removed in these simulations. Using these parameters, the maximum simulated membrane depolarization achieved in an axon below the electrode corresponded to the relevant membrane threshold. Membrane thresholds were 12.09 mV and 6.30 mV above resting membrane potential for the AOP and the AIS, respectively. Note that this analysis used a NFL thickness of $25\ \mu m$ appropriate for the rabbit retina whereas human NFL thickness is in the order of $100\ \mu m$, and is considered for the remainder of the chapter. The assumption is that the ratio of membrane thresholds between the AIS and AOP is suitably consistent across vertebrate species (Sanes and Zipursky, 2010). To ensure the present result is robust to this assumption and to the calculated threshold membrane potential values, a sensitivity analysis was conducted to assess the dependence of the final simulation results, presented in the Results section.

2.3.5 Simulation Methods

Simulations of a wide range of electrode and current waveform variations were conducted in MATLAB. All computations of induced extracellular potential and membrane potential were first calculated in the spatial and temporal frequency domains using Fourier domain solutions of the modeled system. The frequency representation of the longitudinal and first transverse components of a neurite's or a volume's membrane potential were summed together prior to calculating the inverse Fourier transform, yielding the final membrane potential. The solution to the system described above is found in the Fourier domain with respect to the x and y spatial dimensions and the temporal dimension. Due to this, each simulation required the calculation of extracellular and membrane potential in an entire spatial plane and for the full temporal extent of the simulation before the inverse Fourier transform was calculated.

This analysis considered only direct cell responses and neglected the effect of retinal networks. As such, the output of the passive membrane potential model was compared to pre-calculated membrane

thresholds for the AIS and AOP to determine corresponding levels of activation.

To determine the proportion of fibers activated at a given location within the retina, the activity of fibers with an appropriate range of orientations in the x - y plane was first calculated and then combined in a weighted sum, where the weights were sampled from an assigned distribution of orientations. For locations in the NFL, a single parallel orientation was assumed, whereas, for the GCL, a uniform distribution of orientations was applied in the x - y plane, as validated in Figure 2.2.

We describe and analyze the results of simulations of straight cylindrical neurites embedded in the modeled four-layer retinal structure. For all simulations, 100 μm diameter disk electrodes were used unless stated otherwise. For simulations of multi-electrode stimulation, electrodes were arranged in a regular grid with 200 μm center-to-center spacing between electrodes. Unless otherwise stated, stimuli used were cathodic-first, biphasic pulses with a pulse width of 200 μs . In addition, for all multi-electrode simulations, equal currents were applied to each electrode.

Due to the importance of anisotropy in the NFL, which depends on the thickness of the NFL, it is required that human retinal geometries be used in simulations. In general, the required model parameters are not expected to be specific to the animal being simulated. The RGC axon radius, a , has been shown experimentally to be consistent across several animal and human studies (Ascoli et al., 2007; Badea and Nathans, 2011; Coombs et al., 2006; Fletcher et al., 2014; Kong et al., 2005; Toris et al., 1995; Poria and Dhingra, 2015; Rodger et al., 2012; Sümbül et al., 2014; Lee and Stevens, 2007; Snider et al., 2010; Mazzoni et al., 2008; FitzGibbon and Taylor, 2012; Jonas et al., 1992). All relevant model parameter values used are presented in Table 2.2. Note that since the model includes a description of both the intracellular and extracellular components of the relevant retinal layers and the effect of the constrained extracellular space, we use the extracellular resistivity associated with the extracellular medium, and not the *effective* resistivity of the tissue, which is influenced by cells and the extracellular medium. To compare the activation of AOPs in the NFL and AISs in the GCL, we considered characteristic axons located just ($z = 10 \mu\text{m}$) below the surface of their respective retinal layer, as structures at these locations are most sensitive to epiretinal stimulation.

An implementation of the computational model used in this study is hosted publicly on GitHub (<https://github.com/timesler/FourLayerRetinalModel-Esler2017>).

2.4 Results

2.4.1 Analysis of tissue anisotropy

A prerequisite for tissue orientation-dependent activation of RGCs is that tissue anisotropy translates into anisotropic spread of extracellular potential. Figure 2.5 shows the normalized spread of current versus depth in the retina for planes parallel and perpendicular to the orientation of fibers in the NFL. Anisotropy of current spread is demonstrated, with a 760 μm increase in the half-width at full-maximum extracellular potential across the NFL in the y - z plane (parallel to the orientation of fibers in the NFL) compared to 500 μm in the x - z plane. This represents a 1.52x greater spread of extracellular potential in the direction of passing fibers.

Table 2.2: Model parameter values, unless stated otherwise

Parameter	Description	Value	Reference
a	Neurite radius	$0.47 \mu\text{m}$	Ascoli et al. (2007) Badea and Nathans (2011) Coombs et al. (2006) Fletcher et al. (2014) Kong et al. (2005) Toris et al. (1995) Poria and Dhingra (2015) Rodger et al. (2012) Sümbül et al. (2014) Lee and Stevens (2007) Snider et al. (2010) Mazzoni et al. (2008) FitzGibbon and Taylor (2012) Jonas et al. (1992)
d	Extracellular sheath width	30 nm	Syková and Nicholson (2008)
ρ_i	Intracellular resistivity	$0.7 \Omega\text{m}$	Gleixner and Fromherz (2006)
ρ_e	Extraellular resistivity	$0.7 \Omega\text{m}$	Stuart and Spruston (1998) Gleixner and Fromherz (2006) Stuart and Spruston (1998)
R_m	Membrane resistance	$1 \Omega \text{ m}^2$	
C_m	Membrane capacitance	0.01 F/m^2	
d_N	Nerve fiber layer thickness	$100 \mu\text{m}$	Oishi et al. (2009) Frenkel et al. (2005) Wang et al. (2013) Thomson et al. (2015)
d_{ER}	Electrode-retina separation	$10\text{-}500 \mu\text{m}$	
q	Radius of disc electrodes	$50 \mu\text{m}$	

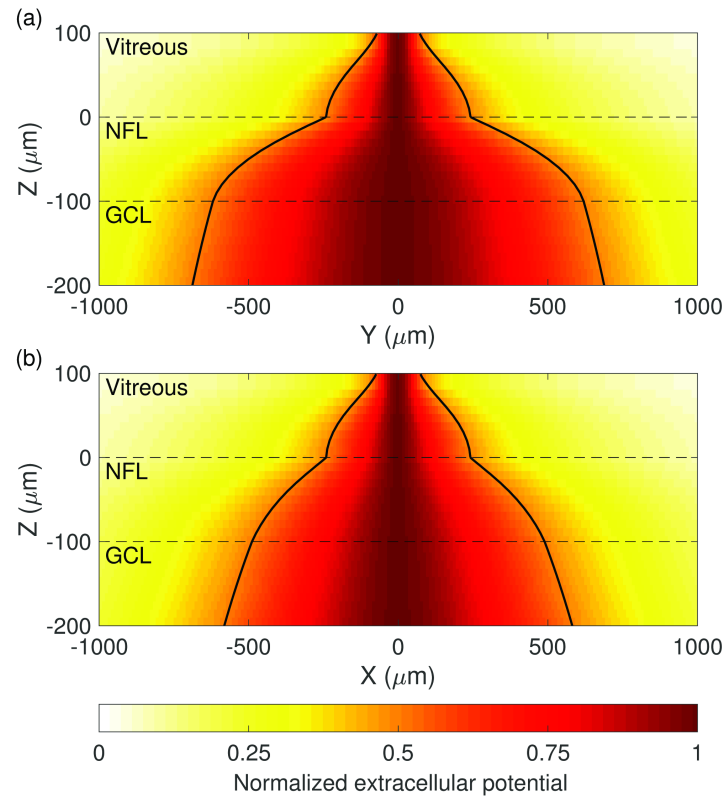


Figure 2.5: **Normalized spread of extracellular potential with distance from a stimulating electrode.** Spread is shown in (a) the $y-z$ plane, parallel to the orientation of AOPs, and in (b) the $x-z$ plane, perpendicular to the orientation of AOPs. The simulated extracellular potential at each z -slice is normalized to the range $[0, 1]$ by subtracting the minimum and scaling the maximum per slice to 1. This is done for illustrative purposes due to the rapid fall-off of extracellular potential with increasing distance from the electrode. Contour lines indicate the full-width at half-maximum potential. Stimulation is with a single electrode located $100 \mu\text{m}$ above the retinal surface at the origin in the $x-y$ plane. Dashed lines indicate layer boundaries.

2.4.2 Comparison of one-, two-, and four-electrode configurations

To establish a basis for fiber orientation-dependent activation in the retina, simulations were run to determine the activation for 1) a single characteristic AOP and 2) AISs with a range of sampled azimuthal orientations. The geometry of these simulations is represented in Figure 2.6(a). Figures 2.6(b), (c), and (d) show the membrane potential resulting from stimulation with one, two, and four electrodes, respectively, for fibers with orientations illustrated in Figure 2.6(a). Figure 2.6(b) highlights the influence of the NFL anisotropy on the activation of GCL fibers of different orientations, with fibers orientated perpendicularly to the AOP experiencing 1.9 times the depolarization of parallel fibers. Figure 2.6(b) also highlights the problem being addressed by this research: although the target AISs in the GCL are more excitable due to the presence of the high-density sodium channel band (represented here by a lower threshold), the proximity of the NFL to stimulating electrodes results in the preferential activation of AOPs.

Figures 2.6(c) and 2.6(d) provide an initial assessment of the combined influence of tissue anisotropy and electric field orientation on activation of AOPs and AISs. The work of Rattay and Resatz (2004) indicated that the activation of a passing fiber may be limited by controlling the way in which the induced electric field changes along the length of that fiber. Hence, simulations have been designed that recruit a number of electrodes aligned with the direction of the considered AOP. As can be seen from Figure 2.6, the level of AIS versus AOP activation increases markedly as the number of electrodes increases. With four electrodes, it is possible to activate 78% of AIS fibers before activation of the overlying layer. When compared to Figure 2.6(b), there is a consistent increase in the relative activation of perpendicular (green) and parallel (brown) AISs in the GCL for two- and four-electrode configurations. For comparison, the ratios of perpendicular to parallel AIS activation are 1.9, 2.1, and 2.5 for one, two, and four electrodes, respectively.

2.4.3 Effect of pulse duration and electrode-retina separation

A parameter sweep was conducted over pulse duration and electrode-retina separations. For each set of parameters, simulations were run to compare the membrane activation of parallel neurites in the NFL and neurites with a range of rotated orientations in the GCL. For both the single orientation in the NFL and the range of simulated fiber orientations in the GCL, membrane potential was calculated for fibers across the full plane at the appropriate retinal depth. Under the assumption that the xy -orientation of AISs is described by a uniform distribution, the proportion of preferentially activated AIS fiber orientations was determined. This is illustrated in Figure 2.7, which shows the level of preferential activation achieved for a variety of stimulation parameter combinations. In this analysis, preferential activation is defined as when the membrane potential of an AIS is driven to its threshold potential at a lower stimulus current than is required to drive *any* AOP to threshold.

Figure 2.7(a) highlights the challenge of achieving preferential activation of the GCL using single-electrode stimulation. Only very small levels of selectivity are obtained even with the most favorable stimulation parameters (large electrode height and pulse duration). A dramatic increase can be seen in the range of stimulation parameters at which preferential activation is achieved when moving from the one- or two-electrode configurations to four electrodes.

A comparison of the membrane potential induced by four-electrode stimulation with small and

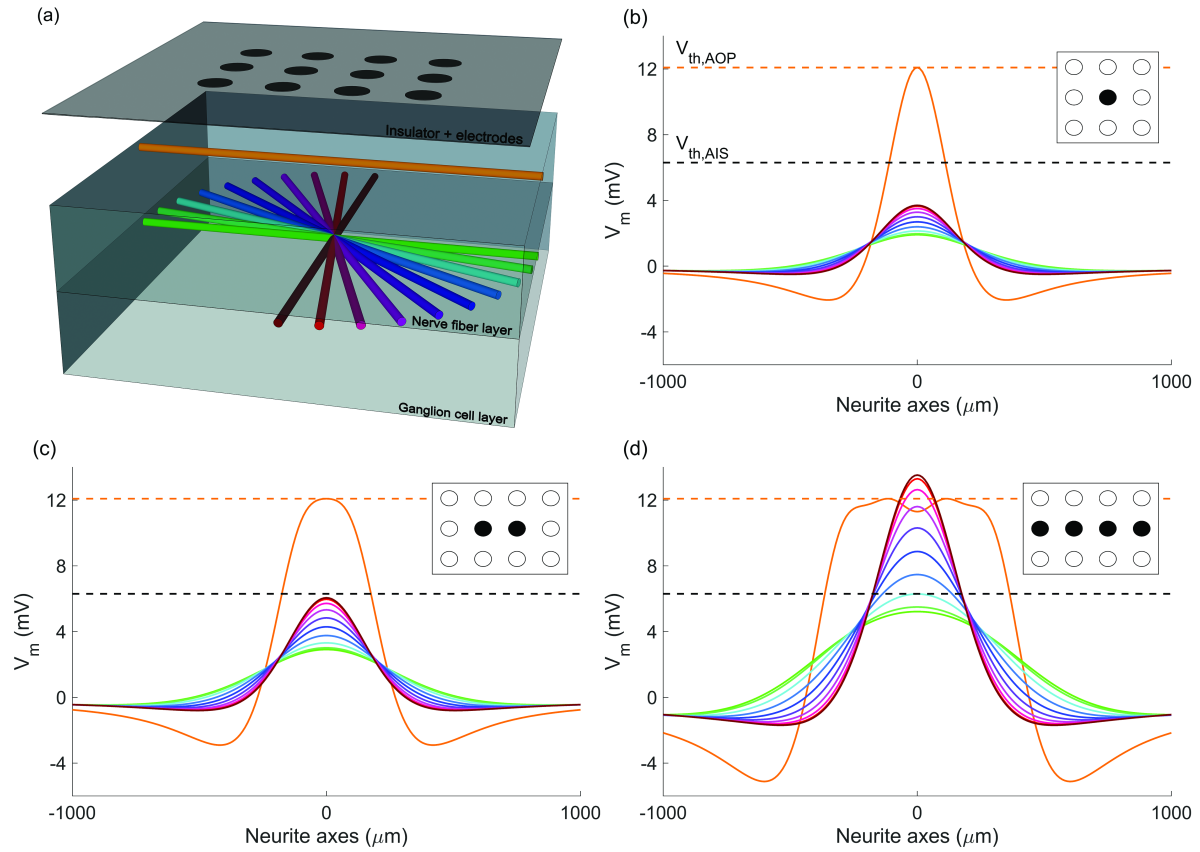


Figure 2.6: **Geometry and simulated membrane potentials for axons of passage and axon initial segments at a variety of x - y orientations.** (a) Four-layer model geometry showing the electrode array, an example of a parallel axon of passage (orange), and the neurite orientations considered in the ganglion cell layer (green-brown). Membrane potential at the end of the cathodic phase is shown along the axes of the neurites being simulated for configurations of (b) one, (c) two, and (d) four electrodes aligned with the axon of passage. Dotted lines represent membrane thresholds for axons of passage (orange) and axon initial segments (black). Stimulus currents have been chosen such that they drive the axon of passage precisely to its threshold level. Colors in (b)-(d) indicate corresponding neurites in (a).

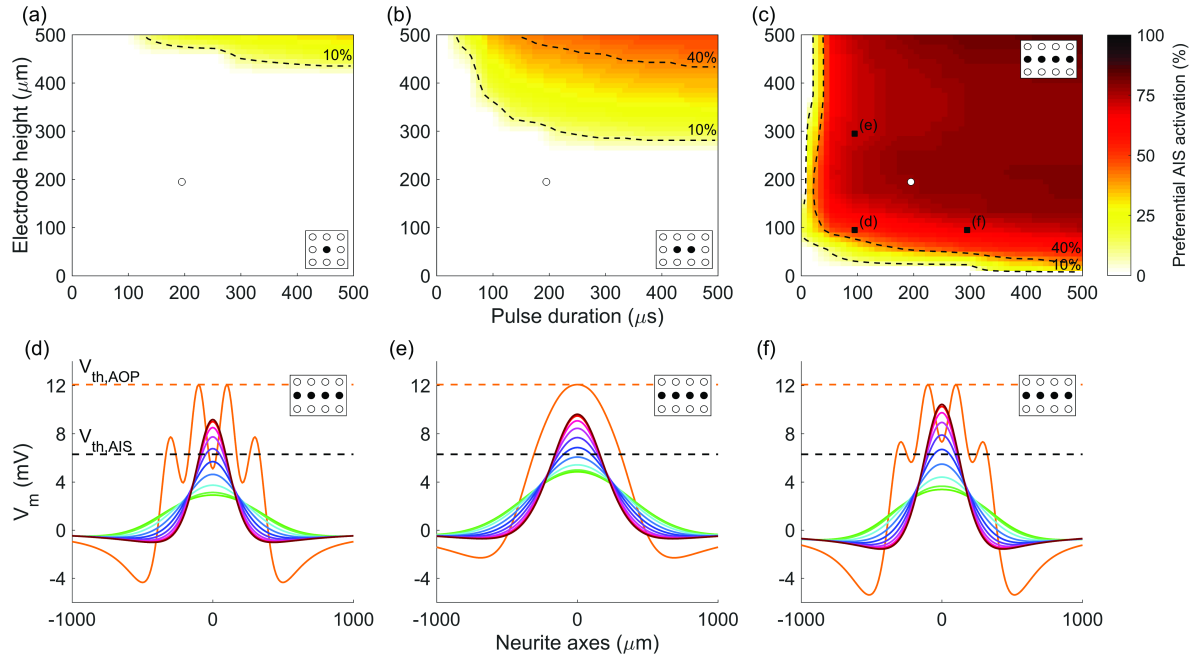


Figure 2.7: **Proportion of axon initial segment orientations preferentially activated for different electrode-retina separations (d_{ER}) and pulse durations.** Heat maps indicate the proportion of axon initial segments activated at a lower stimulus current than any fibers in the nerve fiber layer for (a) one-, (b) two- and (c) four-electrode configurations (aligned with the axon of passage). Regions of low (<10%), medium (10-40%), and high (>40%) stimulation selectivity are separated by dotted contours. White markers indicate the parameters used in Figure 2.6, and black markers indicate the parameters used for subplots (d), (e) and (f), which show examples of simulated membrane potentials for axons of passage and axon initial segments. Colors in (d)-(f) correspond to those in Figure 2.6(a).

large electrode-retina separation can be seen by comparing Figure 2.7(d) and Figure 2.7(e). A clear effect is that, due to the smoothing effect of increased current spread with greater electrode-retina separation, the AOP membrane potential along the axon has a much smoother shape for electrodes positioned further from the retina. Less intuitively, larger separation distances result in increases in preferential activation of the GCL. This is due to the increased opportunity for summation of currents originating from adjacent electrodes. Similarly, as can be seen in Figures 2.7(d) and 2.7(f), increases in pulse duration also result in increases in preferential activation of the GCL. Importantly, however, the overwhelming majority of the change in preferential activation occurs for pulse durations of less than $50 \mu\text{s}$ and separation distances of less than $100 \mu\text{m}$. Although preferential activation increases with pulse duration, above $50 \mu\text{s}$ the increases become negligible, so a duration of $50 \mu\text{s}$ may be preferred due to the lower power required.

2.4.4 Sensitivity of results to threshold membrane potential

Due to the dependence of these findings on the chosen threshold values at the AIS and AOP, a sensitivity analysis was performed. Figure 2.7(c) presents four-electrode stimulation results obtained using an AOP to AIS membrane potential threshold ratio of approximately 2 (12.09 mV versus 6.30 mV), as determined from simulations of experiments conducted by Fried et al. (2009). Identical four-electrode analyses were conducted using ratios of 1.5 and 1. Each case was compared by calculating the proportion of the plot area for which greater than 10% preferential AIS activation was achieved. This yields 93% for a ratio of 2 as shown in Figure 2.7(c), 80% for a ratio of 1.5, and 24% for a ratio of 1. As expected, increasing the relative membrane threshold of the AIS decreases its propensity for preferential activation. This decrease is modest for a more conservative ratio of 1.5. Even with a ratio of 1, corresponding to equal membrane potential thresholds at the AIS and AOP, a level of preferential activation is achievable with four-electrode stimulation. This is due to the beneficial influence of electrode alignment and tissue anisotropy on the shape of the induced electric field. The resultant shape tends to more readily stimulate fibers with an orientation perpendicular to those in the NFL, causing preferential activation of some fibers in GCL target region.

2.4.5 Performance of simultaneous four-electrode stimulation

An important assessment of these results is how the increase in preferential activation of the GCL affects key clinical performance metrics, such as the required stimulus charge density and the spatial selectivity of activation, which is measured here using activation radius. In the following analysis, GCL activation level is defined as the percentage of AIS orientations that are activated (depolarized to above membrane threshold) given a specific stimulus. This percentage is taken at the point in the plane of analysis that is maximally activated, which in all simulated examples is centered with respect to the activated electrodes. Activation radius is used to show the width of the region that is activated by a given stimulus, which will directly affect the resolution achievable with an implanted device and is defined as the radius of the smallest circle that encloses all areas with non-zero activation.

Figure 2.8(a) shows the relationship between stimulus charge density and GCL activation level and how this relationship changes with electrode configuration and electrode-retina separation distance, d_{ER} . As expected, to achieve an equal level of activation for more distant electrodes, greater stimulus charge is required. Figure 2.8(b) shows the variation in activation radius with stimulus charge density

and Figure 2.8(c) shows the correspondence between activation level and activation radius in the GCL. In each of Figures 2.8(a)-(c), dashed curve regions indicate undesirable stimulation configurations, in which AOPs are activated preferentially or in addition to AISs. In terms of isolating the optimal stimulus level, it is important to consider whether this will result in co-activation of passing axons (as indicated by dashed regions), the level of activation achieved in the GCL, and the resulting radius of activation in the GCL. To facilitate comparison of the spread of activation in the GCL induced by one-, two-, and four-electrode configurations, two-dimensional maps of activation in the x - y plane are shown in Figures 2.8(d)-(f), along with the locations of the stimulating electrodes. Importantly, despite utilizing four times the number of electrodes, the activation radius at a given activation level for the four-electrode configuration is typically less than 200% of the activation radius for one electrode, as indicated by dotted lines in Figure 2.8(c). Furthermore, for four electrodes, the required stimulus charge density is reduced to 0.34x of that required for one electrode.

2.4.6 Non-ideal electrode array placement

In practice, electrodes are unlikely to be ideally aligned with passing axons as in Figure 2.8(f). This is due to both the placement of the implanted device and the curvature of passing axons as they pass under the electrode array. To test the application of the multi-electrode stimulation strategy for non-ideal electrode placement, several more challenging geometries were simulated. In each case, the electrodes recruited for stimulation were chosen to represent the most logical extension of the ideal four-electrode configuration presented above and the electrodes were stimulated with equal current.

Figure 2.9 shows an assessment of two such geometries: one where the target for stimulation is centered between four electrodes and another where the target for stimulation is centered between two electrodes and with a non-parallel axon of passage orientation of 22.5 degrees relative to the electrode array orientation, as shown in the insets in Figures 2.9(a)-(b). For the former case, stimulation current was delivered by eight electrodes in total. Another obvious choice of AOP orientation to analyze is 45 degrees. However, because this orientation aligns with diagonal rows of electrodes, the outcome was very similar to the ideal, 0 degree case and so has been omitted here.

The resulting membrane potentials along the axis of an AOP and AISs with varied orientations are presented in Figures 2.9(a)-(b). For each configuration, preferential activation of AISs was achieved, with 70% and 44% of AIS orientations being activated at lower stimulus currents than any AOPs for the eight- and four-electrode configurations shown, respectively (compared to 61% for the ideal four-electrode configuration). As shown in Figures 2.9(c)-(d), the relationship between GCL activation and activation radius is comparable with that of the ideal configuration. Finally, the x - y activation maps in Figures 2.9(e)-(f) indicate only modest increases in the spread of activation when compared to the ideal case. Simulations of several additional edge cases such as off-center target region, are presented in A.4

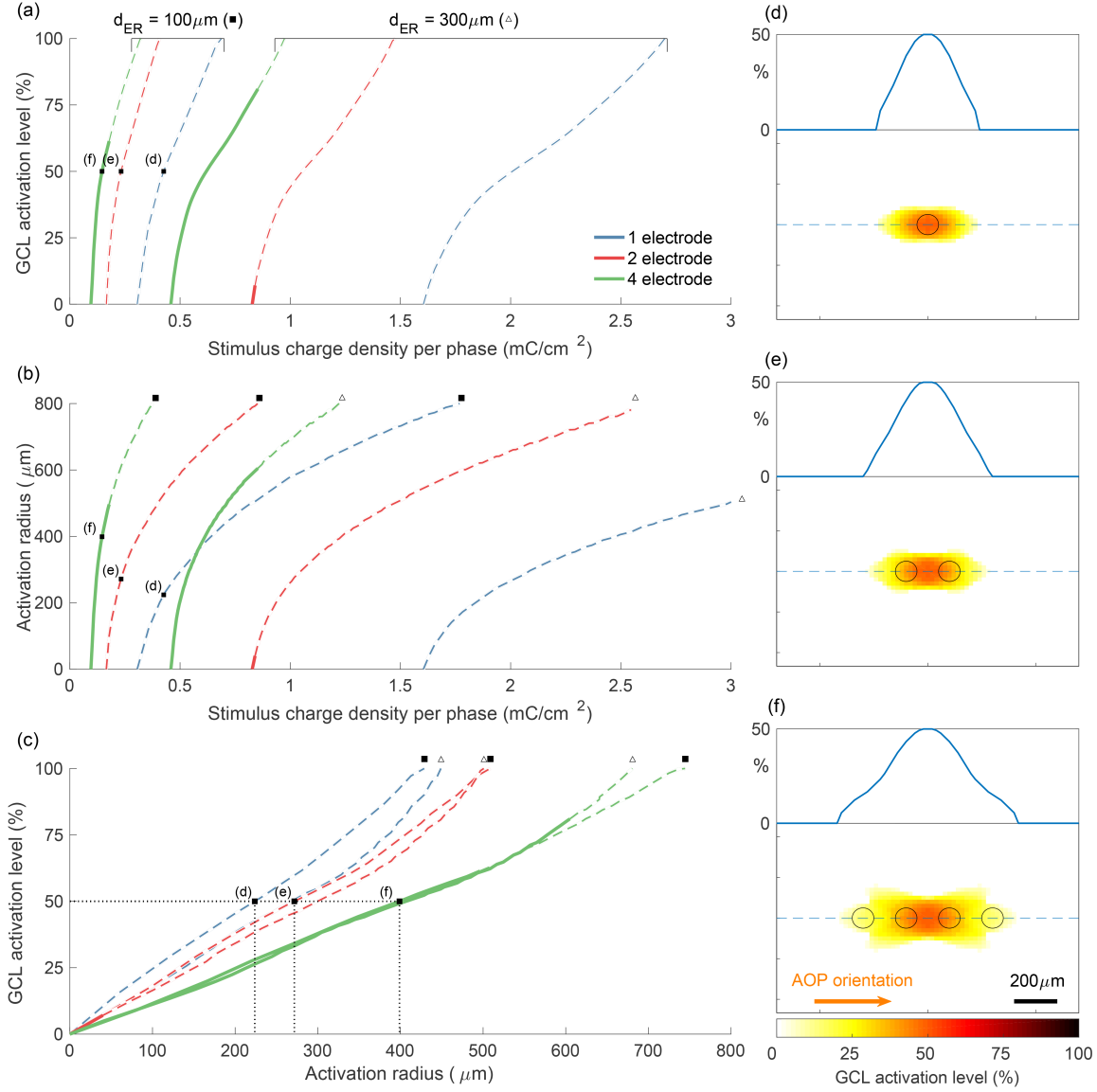


Figure 2.8: Performance of different electrode configurations with respect to GCL activation level, required stimulus charge density per phase, and radius of activation. (a) The proportion of AIS orientations activated vs. stimulus charge density for various electrode configurations and electrode-retina separation distances, d_{ER} . (b) The radius of the activated region vs. stimulus charge density per phase. (c) The relationship between activation radius and activation level. Stimulation strategies analyzed in (a)-(c) include one-, two-, and four-electrode configurations, as well as separation distances of $100 \mu\text{m}$ (filled square, \blacksquare) and $300 \mu\text{m}$ (unfilled triangle, \triangle). Solid and dashed regions in (a)-(c) represent configurations that result in preferential activation of AISs and preferential activation of AOPs, respectively. Labeled points in (a)-(c) correspond to the examples plotted in (d)-(f), which show the spread of GCL activation in the x - y plane when stimulus charge is set to achieve maximum GCL activation of 50%. Dashed blue lines in (d)-(f) correspond to one-dimensional insets. All simulations used a pulse phase duration of $200 \mu\text{s}$, with amplitudes indicated in terms charge density per phase.

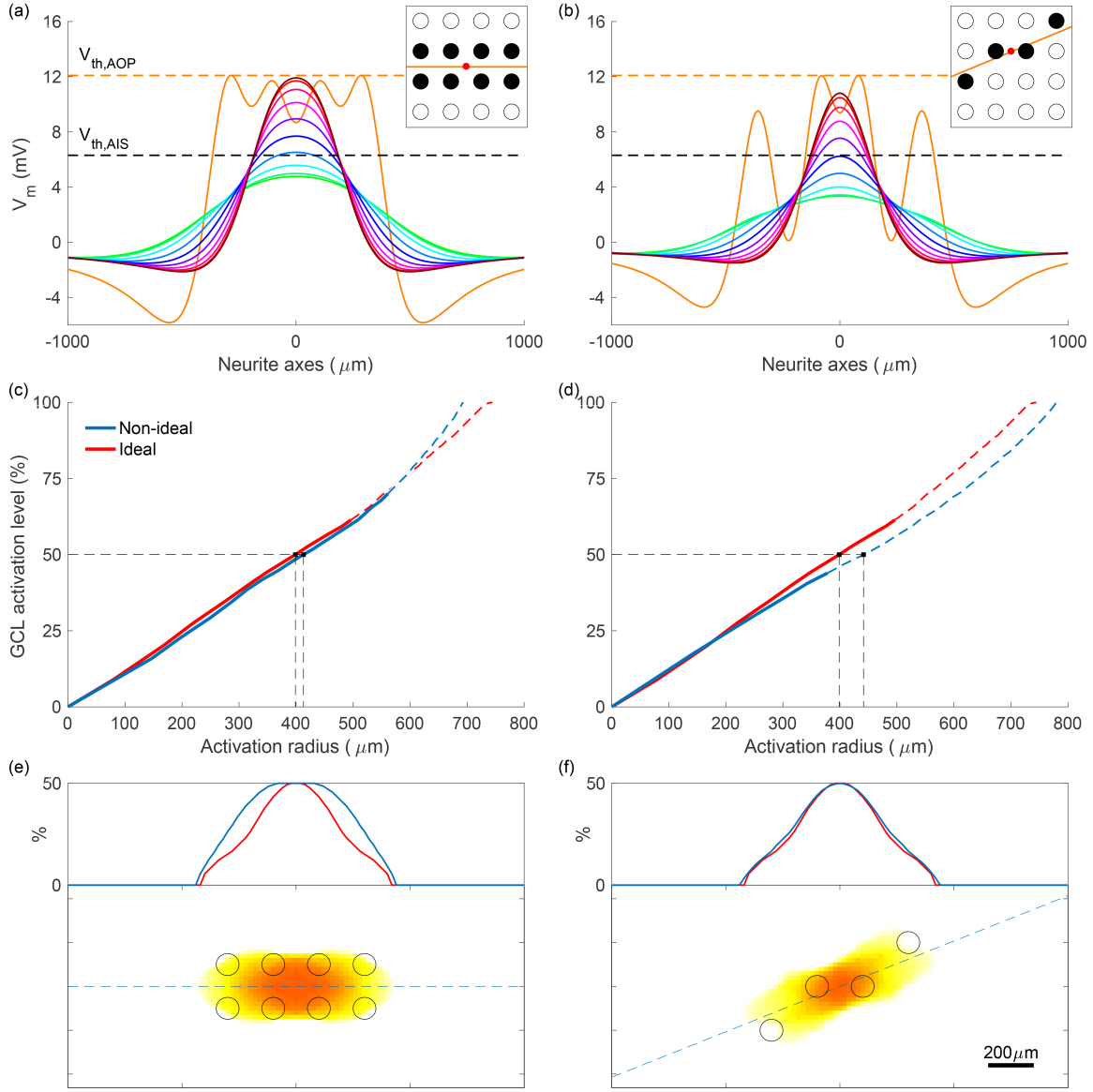


Figure 2.9: Preferential stimulation for two non-ideal electrode array placements. (a)-(b) Membrane potential along neurite axes for axons of passage and axon initial segments, with stimulus current chosen to maximally activate initial segments without activating any passing axons. Colors correspond to those in Figure 2.6(a), with green parallel to axons of passage and brown perpendicular. Insets describe the geometry of each simulation, indicating target region (red), electrodes used (black), and the orientation of axons of passage (orange). (c)-(d) Ganglion cell layer activation level vs. activation radius for non-ideal and ideal (as in Figure 2.8(f)) geometries. Transitions from solid to dashed lines represent the transitions from axon initial segment to axon of passage preferential activation. (e)-(f) The spread of ganglion cell layer activation in the x - y plane. The dashed blue line corresponds to the one-dimensional inset. Colors are mapped according to the color bar in Figure 2.8. All simulations used a pulse phase duration of $200 \mu\text{s}$ and electrode-retina separation of $100 \mu\text{m}$.

2.5 Discussion

2.5.1 Key factors influencing preferential retinal activation

The observed dependence of activation on neurite orientation is a result of several competing factors. The dominant orientation of axons in the NFL results in highly anisotropic spread of extracellular potential under stimulation. As a result, the orientation of fibers in the GCL with respect to this anisotropy influences membrane potential. The overall probability of eliciting a response selectively in the GCL and not in the NFL then depends on the relative influence of fiber rotation, membrane threshold, and fiber depth.

As shown by Figure 2.5, current spreads through the NFL much more readily in the direction of the overlying fiber tracts than it does perpendicularly to them. This leads to a more rapid change in extracellular potential when moving away from stimulating electrodes in the direction perpendicular to the AOPs. This, in turn, results in the directional spatial derivatives of extracellular potential being greater in this perpendicular direction, leading to maximal activation of AISs with perpendicular orientation in the GCL, as seen in Figure 2.6(b). Specifically, the activation due to orientation is influenced via differences in the second spatial derivative of the extracellular potential, which manifests in the frequency domain in Equation (2.5a) as $-k_y^2 \hat{\phi}_\alpha$. This analysis also shows that modulating the spread and orientation of the electric field by driving multiple, aligned stimulating electrodes can be used to minimize activation of fibers with specific orientations, such as passing axons.

A secondary effect of using multiple electrodes aligned with passing axons is that the ratio of depolarization of perpendicular AISs to parallel AISs increases with the number of electrodes. In the results summarized in Figure 2.6, the ratio of the maximal depolarization for perpendicular AISs to parallel AISs is 1.9, 2.1, and 2.5 for one-, two-, and four-electrode configurations, respectively. The cause of the increase from one to four electrodes is likely that, by aligning electrodes with passing axons, the activity of similarly oriented AISs in the GCL is also reduced, while having little effect on the depolarization of perpendicular AISs. This effect is far less pronounced for parallel fibers in the GCL when compared to the NFL due to the natural spread of current at greater retinal depths; the artificial spread of current introduced by using multiple electrodes is less pronounced when compared to the natural longitudinal spread caused by the geometry of the NFL. In contrast, the anisotropic spread introduced by the NFL, shown in Figure 2.5, has little effect on superficial AOPs as they are close to the retinal surface where current spread is still predominantly isotropic and so we must rely on electrode configuration to control the profile of extracellular potential.

A related phenomenon is highlighted in Figure 2.8(c), which shows that, given a certain level of GCL activation, there is an increase in the spread of activation as electrodes are moved further from the retinal surface; however, this increase is not seen for four electrodes. The reason for this is that, due to the wider distribution of current at the electrode array when recruiting four electrodes, the increase in spread due to greater electrode-retina distance is marginal. Another key feature of the system is that the region in which the largest spread occurs is in the NFL, in the direction of passing axons. Therefore, an increase in the distance of vitreous fluid through which current flows has a less pronounced effect on total spread in that direction. An increase in the spread of activation in the direction perpendicular to passing axons can be observed as electrodes are moved away from the retina; however, spread is always more pronounced in the direction of passing fibers.

Although these results are based on simulations of cylindrical neurites, the developed method for the analysis of arbitrarily rotated fibers can be applied directly to the simulation of unbranched axons with arbitrary morphology, as discussed in A.3. A preliminary next step will be to validate the current results using ganglion cell axon reconstructions. A key point of interest will be whether the effect is maintained when axonal orientation changes along the length of the simulated fiber. This will depend on the length constants associated with both axon curvature and membrane activation. If the latter is relatively smaller, axon curvature will have little effect and localized fiber orientation will determine the level of activation along the axon.

2.5.2 Choosing a stimulation strategy

As can be appreciated from Figures 2.7(a)-(c), of the electrode configurations that were simulated, preferential stimulation with clinically desirable parameters can only be achieved with four electrodes. Ideally, electrodes should be placed as close as possible to the surface of the retina without causing damage. This reduces the required stimulus current and limits current spread, thereby increasing the achievable device resolution. From Figure 2.7(c), most of the change in AIS activation with varying electrode height is seen to occur in the first 100 μm , suggesting that the optimal electrode height considering both preferential AIS activation and activation radius is around 100 μm . Beyond this height, little is gained in terms of preferential activation, with reductions in resolution and larger required stimulus currents.

Given the electrode-electrode separation used in this study, for separation distances of less than 50 μm , preferential stimulation is limited due to a lack of lateral summation of currents from adjacent electrodes; activation under each electrode will occur in a similar way to one electrode. This highlights the fact that these results rely on current spread from adjacent electrodes overlapping and summing together. The level of this summation depends on both the distance between electrodes (the x - y distance that current must spread) and the distance from the electrodes to the retinal surface (the z distance over which current can spread). In theory it is expected that, in the limit of infinitesimally small electrodes that are infinitesimally close together, preferential activation could be achieved with electrodes arbitrarily close to the retinal surface. In reality, the optimal electrode-retina separation distance will depend on the geometry of the electrode array and may differ from the results presented here.

The combination of results shown in Figures 2.7 and 2.8 provide a starting point for choosing a clinically relevant stimulation strategy. If the height of the electrode array above the retina is fixed at 100 μm and pulse duration is greater than 50 μs , the chosen pulse duration has little influence on activation provided appropriate current magnitudes are delivered. Key remaining considerations are the required current or charge density, level of activation in the GCL, and size of the activated region, which can be determined from Figure 2.8. It is unclear exactly how either GCL activation level or activation radius in the current model will map to perception by patients with an implanted device. As such, a suitable stimulus charge may need to be determined either experimentally or based on direct feedback from device users. A suitable charge density will depend on the trade-off between GCL activation level and activation radius (Figure 2.8(c)), and should always be kept below the level required for AOP-related perception and within clinically determined safety limits. A conservative and commonly cited safety limit based on the theoretical non-gassing limit is 0.35 mC/cm^2 per phase for platinum electrodes (Brummer and Turner, 1977). For four-electrode stimulation with an array

positioned $100\ \mu\text{m}$ above the retina (green line indicated by filled square), the maximum charge density per phase before unwanted AOP activation is $0.18\ \text{mC}/\text{cm}^2$ per phase, as indicated by the transition from solid to dotted lines in Figures 2.8(a) or (b). A valuable implication of using four-electrode stimulation is that it results in an approximately 3x times decrease in the required charge density and hence is consistently below safe charge density limits. Although the total current and total power required is slightly more for four-electrode stimulation when compared to one-electrode stimulation, this increase is well below proportional.

Although Figure 2.8 shows that by recruiting more stimulating electrodes the induced area activated becomes greater, it should be noted that this will not necessarily reduce perceived resolution. Previously, recipients of epiretinal implants have reported elongated and line-like phosphenes, thought to be caused by stimulation of passing axons in the NFL that originate from distant regions of the GCL (Jensen et al., 2003; Werginz and Rattay, 2015; Rizzo et al., 2003b; Jepson et al., 2013; Rattay and Resatz, 2004; Werginz et al., 2014; Cao et al., 2015; Weitz et al., 2013; Abramian et al., 2015; Jeng et al., 2011). Hence, despite an increase in the region of activation in the GCL when using a four-electrode stimulation strategy, the overall resolution is expected to increase due to the elimination of activation of the NFL. Furthermore, phosphene regularity is expected to be greater under the proposed strategy, more readily facilitating the development of more complex stimulus patterns built up from this perceptual subunit.

2.5.3 Determining membrane thresholds

To the best of our knowledge, although threshold stimulus currents have been reported for the AIS and distal axon of RGCs (as by Fried et al. (2009) and Grosberg et al. (2017)), there exists no experimental data on the membrane thresholds of RGCs at these locations. In previous modeling studies, the low threshold of the AIS has only been incorporated into active, conductance-based models of RGCs. In these models, the threshold is reduced at the AIS by increasing the sodium channel density by a factor ranging from 2 to 40, generally chosen to reproduce desired physiological responses (Maturana et al., 2014; Jeng et al., 2011; Werginz et al., 2014).

As ion channels are not modeled here, the sensitivity of the AIS was adjusted in our passive model by using different membrane potential thresholds. To avoid arbitrarily choosing a reduced threshold for the AIS, membrane threshold levels were determined using an approximate reproduction of the experimental procedure of Fried et al. (2009). This was made possible as the stimulus thresholds reported by Fried et al. (2009) were accompanied by measurements of axonal morphology and three-dimensional electrode location. Simulating the experiment using the same modeling framework in which the thresholds were later applied ensured that the chosen values were most representative of the reported experimental data and were relevant to the current model. The ratio of the calculated membrane threshold for the AOP and AIS was approximately 2. Due to the scale of sodium channel densities utilized in previous models, this suggests that our estimation of the range of parameters for which preferential activation is achievable is conservative. This value is dependent on model parameters for extracellular tissue impedance and the assumed location of the RGC in simulations of the Fried et al. experiments. For instance, if the AIS was assumed to be further from the electrode, an lower membrane potential threshold would have been estimated at the AIS. For this reason, the AIS was assumed to be located just below ($5\ \mu\text{m}$) the surface of GCL, yielding the highest, and therefore most conservative, estimate of threshold membrane potential. Furthermore, as demonstrated by the

sensitivity analysis presented in the Results, an even more conservative membrane threshold ratio of 1.5 may be used without much loss of effect.

2.5.4 Experimental validation

Controlled experimental validation of these results requires techniques for the measurement of RGC activation at multiple locations in the retina simultaneously and a method for imaging axon trajectories relative to implanted devices. *In vitro* studies in which the average trajectory of passing axons in the NFL is known will allow for measurements of activation to be taken in the GCL at both the region being targeted by stimulation and at more distant locations that lie under the trajectory of passing axons. Potentially useful methods have also been developed for monitoring GCL activity across the whole retina using calcium imaging (Weitz et al., 2013) or micron-scale electrical imaging of axonal action potential transmission (Zeck et al., 2011; Bakkum et al., 2013). Recent developments in high-density stimulation and recording, which enable initiation of action potentials in precise locations of a neuron and imaging of their propagation, may also serve as valuable, non-clinical platforms for the validation of a strategy such as this (Radivojevic et al., 2016; Eickenscheidt and Zeck, 2014). A challenge with quantitatively validating the result in this chapter is that the small distance between the electrode array and the surface of the retina must be very tightly controlled.

Due to the dependence of these results on the anisotropy of the NFL, it is expected that varying the thickness of the NFL will have a marked effect. The chosen NFL layer thickness is based on an approximation of the human retina, and so these results are relevant only to human retinal stimulation. Rodent models used for research and testing of epiretinal implants have thinner NFLs and so the influence of retinal layer orientation will be less pronounced. Although this in no way confounds the current findings, it suggests that experimental validation would be best carried out in the primate retina. A potential solution for other animal models may be to modify the present model to represent the appropriate animal model so that any observed evidence can be extrapolated to human-like retinal geometries. It is important to note that a large part of the present result derives from electrode configuration, which can be kept consistent across different animal models.

In a real-world implanted system, the assumption that the distance between each electrode and the NFL is the same is unlikely to hold. In addition, the electrode-retina separation distance is likely to change over time due to device settling, immune responses and changes to the implant's environment caused by vitrectomy. Challenges such as these will require careful modification of the idealized solution presented here, and are likely to require either measurement for each patient of the precise three-dimensional location of the electrode array over time using methods such as OCT or measurement of the retina's electrical response to stimulation. Procedures such as those used by Grosberg et al. (2017) for electrical recording of RGCs following stimulation provide a potential solution. This method enables the classification of responses as either AOP activation or AIS/soma activation. This information could then be used to tune the multi-electrode stimulation strategy over time.

2.5.5 Optimizing electrode currents

For stimulation strategies that utilize more than two electrodes, it is likely that delivering equal currents to all electrodes does not represent the optimal stimulus for achieving preferential activation with minimal activation radius. It may be possible to more optimally distribute currents across the recruited electrodes in a way that minimizes the activating function along AOPs. As can be seen from Figures 2.8(d)-(f), for one- and two-electrode configurations, the profile of activation about the center of the electrode array follows a simple curve with monotonic first derivative. In contrast, with four electrodes, the profile has a more complex shape due to the added degree of freedom. In this case, this extra degree of freedom can be represented by the ratio of the current delivered to the two internal and two external electrodes.

As highlighted by Figure 2.9, there is a range of electrode/AOP orientations that must be dealt with by a proposed stimulation strategy. We have demonstrated that the approach proposed in this chapter is robust to changes in relative electrode array to AOP orientation, and can target off-centered tissue volumes. However, it is again likely that delivering equal currents to each electrode is sub-optimal. In this more general case, optimal electrode currents will also depend on the particular pattern of electrodes that is being used. For instance, the optimal ratio of internal electrode currents to external electrode currents for the case presented in Figure 2.9(b) will be different than for a set of four electrodes perfectly aligned with the AOP.

With four-electrode stimulation, which can be tuned by a single parameter, optimization could be achieved using a simple brute force search through possible current ratios. However, the model presented here is linear and has an analytic solution in the Fourier domain. This means that a closed-form solution to this optimization problem can be found using least squares or some other linear optimization algorithm. This approach could be applied to the optimization of currents delivered to an arbitrary number of electrodes to minimize activation of the NFL. Optimization of multiple electrode currents to achieve both focal activation of the GCL and minimal activation of the NFL will be the subject of a subsequent study.

2.6 Conclusion

This paper demonstrates that activation of RGCs in the inner retina under epiretinal stimulation depends on both axonal orientation and orientation of the stimulating electric field relative to the orientation of AOPs in the NFL. The developed model allows for an analysis of this dependence by capturing the distinct distributions of fiber orientation of the nerve fiber layer and the ganglion cell layer. A four-electrode stimulation strategy has been proposed that accomplishes preferential activation of the retinal ganglion cell AIS over passing axons in the NFL using clinically suitable stimulus charge densities and electrode configurations. Although concessions must be made with regard to activation radius in the GCL, these are relatively minor, and the proposed strategy is expected to enable higher resolutions and more clearly interpretable percepts by users of epiretinal prostheses.

Chapter 3

The biophysical basis for the linear electrical receptive field

This content of this chapter is a slightly modified version of the following published journal article:

Timothy B. Esler, Matias I. Maturana, Robert R. Kerr, David B. Grayden, Anthony N. Burkitt, Hamish Meffin. "Biophysical basis of the linear electrical receptive fields of retinal ganglion cells." J. Neural Eng. 15(5): 055001

3.1 Abstract

Responses of retinal ganglion cells to direct electrical stimulation have been shown experimentally to be well described by linear-nonlinear models. These models rely on the simplifying assumption that retinal ganglion cell responses to stimulation with an array of electrodes are driven by a simple linear weighted sum of stimulus current amplitudes from each electrode, known as the “electrical receptive field”. *Objective.* This paper aims to demonstrate the biophysical basis of the linear-nonlinear model and the electrical receptive field to facilitate the development of improved stimulation strategies for retinal implants. *Approach.* We compare the linear-nonlinear model of subretinal electrical stimulation with a multi-layered, biophysical, volume conductor model of retinal stimulation. *Main results.* Our results show that the linear electrical receptive field of the linear-nonlinear model matches the transmembrane currents induced by electrodes (the activating function) at the site of the high-density sodium channel band with only minor discrepancies. The discrepancies are mostly eliminated by including axial current flow originating from adjacent cell compartments. Furthermore, for cells where a single linear electrical receptive field is insufficient, we show that cell responses are likely driven by multiple sites of action potential initiation with multiple distinct receptive fields, each of which can be accurately described by the activating function. *Significance.* This result establishes that the biophysical basis of the electrical receptive field of the linear-nonlinear model is the superposition of transmembrane currents induced by different electrodes at and near the site of action potential initiation. Together with existing experimental support for linear-nonlinear models of electrical stimulation, this provides a firm basis for using this much simplified model to generate more optimal stimulation patterns for retinal implants.

3.2 Introduction

Progress in retinal prostheses (bionic eyes) has led to promising developments for people with retinal pathologies such as Retinitis Pigmentosa. Present-day retinal prostheses are able to produce visual percepts through electrical stimulation of the retina, and can allow users to perform simple navigation and identification tasks (Ayton et al., 2014; Werginz and Rattay, 2015; Weiland et al., 2011; Rizzo et al., 2003a; Zrenner et al., 2011; da Cruz et al., 2013; Fujikado et al., 2011; Hadjinicolaou et al., 2015). Despite this progress, percepts produced by current devices vary in shape, intensity, and size due to difficulty in using electrical stimulation to reproduce patterns of neural activity with that would normally occur in a healthy retina with visual stimulation. To elicit more natural activation of retinal tissue, methods need to be developed that provide more targeted and natural activation of specific cells, cell types (Twyford et al., 2014; Jensen and Rizzo, 2006), and retinal volumes (Werginz and Rattay, 2015; Rizzo et al., 2003a; Rattay and Resatz, 2004).

3.2.1 Multi-electrode stimulation

A key question regarding stimulation with multi-electrode arrays is how best to achieve an overall pattern of activation that matches a desired image or a desired pattern of activation. A promising approach is the development of stimulation strategies that account for the overall influence of simultaneously driven electrodes on each cell. Progress from sequential single-electrode stimulation to simultaneous multi-electrode stimulation could enable greater targeting of specific locations and cell types. This is a challenging problem due to the complex responses of different retinal neurons and the interactions of electrical currents from adjacent electrodes (Wilke et al., 2011; Shivdasani et al., 2012).

A potential solution is to utilize more sophisticated methods of stimulation to achieve *spatial shaping* of retinal activation that is informed by computational models of responses of cells to multi-electrode stimulation (Matteucci et al., 2013; Habib et al., 2013). Ideally, these models may be fit to individual patients, such as by using implants that can record the retinal activity elicited by stimulation. Methods for shaping retinal activity in response to multi-electrode stimulation require the development of parsimonious models of neural activation that lend themselves to model inversion and optimization techniques. Model inversion will offer a method to directly calculate optimal stimulation patterns for achieving a desired neural activation pattern. One form of candidate model is the linear-nonlinear model of electrical stimulation (Maturana et al., 2016; Halupka et al., 2017).

3.2.2 Linear-nonlinear models

Linear-nonlinear models of retinal activation have been demonstrated to predict retinal ganglion cell activity accurately in response to both light and multi-electrode electrical stimuli (Maturana et al., 2016; Jepson et al., 2014; Halupka et al., 2016, 2017; Sekhar et al., 2016). These models are structured to efficiently capture the statistical relationships between multi-dimensional stimuli and retinal ganglion cell (RGC) spiking rates without explicitly modeling the anatomy or physiology of the system. Although these models have historically been applied to light stimuli (Chichilnisky, 2001), their utility in predicting electrically-evoked RGC responses has recently been demonstrated (Maturana et al., 2016; Jepson et al., 2014; Halupka et al., 2016, 2017; Sekhar et al., 2016).

Linear-nonlinear models are simple, cascaded-filter models that represent single-cell responses using a linear filter to characterize the high-dimensional space of stimulus currents (equal to the number of electrodes) in terms of a low-dimensional signal that is responsible for driving the neuron. The linear filter is a simple weighted sum of the stimulus currents from different electrodes, and represents the observed sensitivity of the cell to stimulation with each electrode. The output of this filter is passed through a static nonlinearity to generate spiking probability. Furthermore, it has been shown that, for the majority of RGCs, a single weighted sum (i.e., a single filter) of the stimulus current amplitudes is sufficient, when combined with a spiking nonlinearity, to accurately predict responses to direct electrical activation. This single filter is known as the *electrical receptive field* (ERF) (Maturana et al., 2016; Jepson et al., 2014). The ERF describes which electrodes influence the cell and by how much, and should not be confused with the well-studied visual receptive field, which describes the region of the visual field that influences a cell. The ERF assumes that the electrodes interact linearly in driving direct activation of the cell. For the minority of cells for which a single linear filter does not accurately predict direct electrically-evoked activation, a piece-wise linear description has been shown to be sufficient (Jepson et al., 2014).

A main limitation of linear-nonlinear models is that, although they are able to accurately predict functional behavior, they do not elucidate the underlying physiological or morphological mechanisms. The performance of linear-nonlinear models, however, suggests that the complex anatomy and physiological processes that govern RGC activity can be largely described by a far simpler representation. In this paper, we aim to recover the linear electrical receptive field of the linear-nonlinear model from a detailed biophysical description of the retina, thereby highlighting the biophysical underpinning of the model.

3.2.3 The biophysical basis for the RGC linear-nonlinear model

As shown in Figure 3.1, a biophysical description of electrical stimulation of a RGC consists of (1) the flow of current through the extracellular space from each electrode to the cell, (2) currents entering each location in the cell, (3) the temporal and spatial integration/summation of inputs within the cell, and (4) the membrane potential dynamics of action potential initiation in response to that intracellular current flow. The accuracy of the linear-nonlinear model indicates that these four stages can be effectively reduced to a single linear filter/projection and a single sigmoidal nonlinearity.

Extracellular electrical stimulation of neural tissue is typically modeled using Scheme 1 shown in Figure 3.1, where extracellular current flow and the induced transmembrane current are modeled as linear, or passive, and cellular integration and spiking is modeled as nonlinear, or active. However, this description suggests a higher-dimensional and nonlinear electrical receptive field, in disagreement with the linear-nonlinear model. Schemes 2 and 3 in Figure 3.1 represent potential simplifications of the detailed biophysical approach, and may be used to explain the linear-nonlinear model.

Experimental evidence indicates that RGC electrical responses are strongly influenced by the high-density sodium channel band (SOCB) and that this is commonly the site of action potential initiation (Fried et al., 2009). This suggests a potential explanation for the one-dimensional (i.e., a single linear filter) description of the stimulus space used by the linear-nonlinear model when fitted to electrical stimuli. If RGC spiking initiates at the SOCB, it may be that the combination of stages 1 and 2 can be well approximated by a linear weighting of electrode currents, corresponding to the

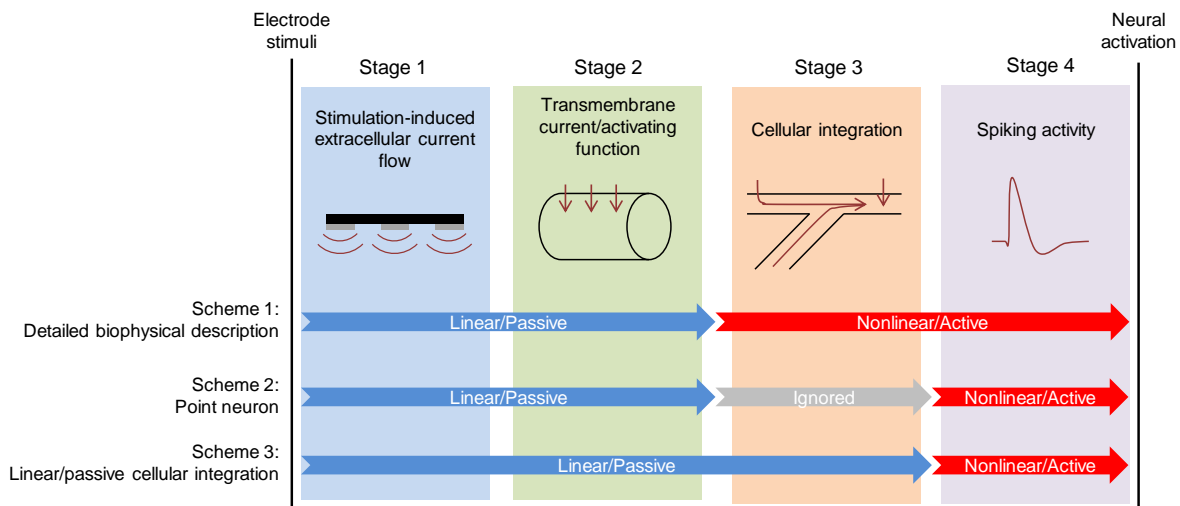


Figure 3.1: Alternative biophysical approaches to modeling retinal ganglion cell activation. Models of neural activation in response to electrical stimulation map electrode currents to corresponding neural activation, using a combination of linear and nonlinear components, indicated by blue and red arrows, respectively. Scheme 1: A typical biophysically-detailed approach is to describe the extracellular flow of current and the resulting transmembrane currents with a function that is linear with respect to the stimulus currents, followed by a nonlinear description of the spatial and temporal integration of cell inputs and spiking behavior. Scheme 2: In cases where neural activity is dominated by cell inputs at a single location, the more detailed description can be simplified by neglecting the impact of cellular integration. Scheme 3: In cases where spiking predominantly initiates at a single location in the cell, but is influenced by cell inputs at a range of locations, cellular integration may be treated as approximately linear/passive, with a nonlinearity applied only at the site of spike initiation.

cell’s electrical receptive field (ERF), that defines the current entering the SOCB (including both transmembrane currents and axial current flow). If the current entering the cell at the SOCB dominates its activity, it may be that the effects of cellular integration can be mostly ignored, as in Scheme 2 in Figure 3.1. In this way, for either of Scheme 2 or Scheme 3 in Figure 3.1, the potentially complex nature of spatial summation of current inputs from across the cell’s entire surface is replaced by a single weighted sum of the stimulus amplitudes. The nonlinear component of the linear-nonlinear model then describes the active spiking dynamics at the SOCB in response to that linear combination of stimuli.

Alternatively, it may be that cellular integration within the cell cannot be neglected, but can be approximated as linear, as in Scheme 3 in Figure 3.1. Provided that spiking predominantly initiates at a single location in the cell, presumably the SOCB, it may be that cellular integration of inputs can be largely approximated as linear/passive, with a spiking nonlinearity applied only at the site of action potential initiation.

For RGCs for which a single weighted sum of the stimulus space does not, on its own, accurately predict direct responses, it has been proposed that this is due to the existence of multiple, distinct sites of action potential initiation (Jepson et al., 2014). This description implies that multiple ERFs may exist for a single cell, one for each site of spike initiation. Each ERF then represents the weighted sum applied to stimulus electrode currents that defines extracellular current flow and summation of intracellular currents to that site of initiation. This would also suggest a piece-wise linear description of the cell’s ERF, where each ‘piece’ relates to the ERF associated with a single site of action potential initiation (Jepson et al., 2014).

This paper aims to illustrate the biophysical basis of the linear ERF of the linear-nonlinear model by comparing it with a detailed biophysical model of retinal tissue and morphologically-accurate, active neuron models. A biophysically-accurate, multi-layered description of retinal tissue is used to determine the activating function (Tahayori et al., 2012; Rattay, 1986, 1989, 1999), which is an approximation of the induced transmembrane current along the surface of a neuron. Simulated activating function data are then used as current inputs in the compartments of a multi-compartment neuron model with active RGC membrane dynamics. This allows the ERF to be estimated from the simulated spiking response to an ensemble of stimuli.

By drawing comparisons between the activating function (transmembrane current) at different locations in the cells and the electrical receptive field the linear-nonlinear model will be biophysically reconstructed. The concept that the linear component, or ERF, of the linear-nonlinear model represents the currents entering the cell near the SOCB will first be examined (Scheme 2). The ERF will then be recovered from a reduced biophysical model in which cellular integration is treated as linear (Scheme 3), demonstrating that the ERF is equivalent to the passively integrated membrane potential at the site of action potential initiation.

3.3 Methods

3.3.1 Electrode geometry and cell locations

Simulations described in this study were designed to replicate the experimental setup of Maturana et al. (2016), who reported the use of the linear-nonlinear model for modeling direct electrically-evoked

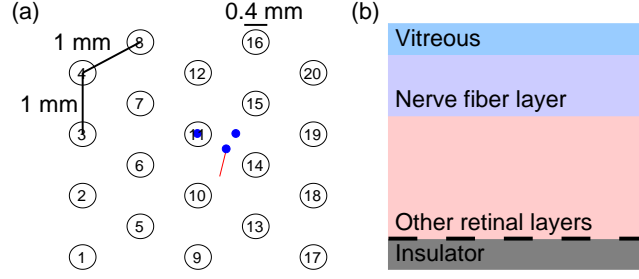


Figure 3.2: **Electrode and tissue simulation geometry.** (a) The electrode array consisted of 20 electrode disks of $400\ \mu\text{m}$ diameter. Electrodes were arranged in 5-by-4 configuration, with center-to-center pitch of 1 mm. For simulation, neurons were oriented such that the end of the axon (example in red) was directed toward the optic disk, which was located under electrode 9 as was the case in Maturana et al. (2016). Simulations were run with six reconstructed cells at each of the three locations (18 combinations in total) indicated by blue markers. (b) The retina was modeled by four distinct layers: the vitreous, the nerve fiber layer, remaining retinal layers, and an insulating substrate. The electrode array was planar and located at the insulator-retina boundary (black dashes). Neurons were simulated with their somas located $10\ \mu\text{m}$ below the nerve fiber layer.

Table 3.1: Conductivity and thickness of modeled layers

Layer	Directional dependence	Conductivity (S/m)	Thickness (μm)
Vitreous	Isotropic	1.78^*	Infinite extent
NFL	Anisotropic	varied^\wedge	40^\dagger
Remaining neural layers	Isotropic	0.1^\ddagger	360^{**}
Insulator	Isotropic	0	Infinite extent

*Mazzoleni et al. (1986)

$^\wedge$ Meffin et al. (2012, 2014)

† Abbott et al. (2014)

‡ Tahayori et al. (2014); Gabriel et al. (2009); Latikka et al. (2001); Ranck (1963); Kasi et al. (2011); Wang and Weiland (2015); Karwoski et al. (1985)

**Zhao et al. (2015)

RGC activation. The modeled electrode array was located subretinally and stimulated retinal tissue with geometry consistent with a rat retina, as that was the animal model used in the experiments of Maturana et al. (2016).

The modeled electrode array consisted of 20 planar disk electrodes of $400\ \mu\text{m}$ diameter with a center-to-center pitch of 1 mm, as shown in Figure 3.2(a). For simulations, neurons were oriented such that the end of the axon was directed toward the optic disk, which was located under electrode 9, as was the case for Maturana et al. (2016). Simulations were run with each of six reconstructed cells at each of the three locations indicated by blue markers in Figure 3.2(a). The retina was modeled by four distinct layers: the vitreous, the nerve fiber layer, remaining retinal layers, and an insulating substrate, as shown in Figure 3.2(b). Further details of this model are given in Section 3.3.4 of the Methods.

3.3.2 The linear-nonlinear model

The input into the linear-nonlinear model was composed of the amplitudes of fixed-period, biphasic current pulses delivered to each of the electrodes of a multi-electrode array; $\vec{S} = [s_1, \dots, s_N]$, where the current on each electrode, s_i , was positive if the biphasic pulse was cathodic-first and negative if it was anodic-first. Using this description of the stimulus space, the model was able to simulate the combined effect of arbitrary stimulus amplitudes across the electrodes of an array. The most general form of the model can be described by the following: given a stimulus vector, \vec{S}_t , applied at time t , the probability of a spike being generated is given by

$$P(R_t = \text{spike} | \vec{S}_t) = \mathcal{N}_N(\vec{v}_1 \cdot \vec{S}_t, \vec{v}_2 \cdot \vec{S}_t, \dots, \vec{v}_N \cdot \vec{S}_t), \quad (3.1)$$

where $R_t \in \{\text{spike}, \text{not spike}\}$ is the binary RGC response, \mathcal{N}_N is the nonlinearity operating on the N -electrode stimulus subspace, and \vec{v}_i for $i \in \{1, 2, \dots, N\}$ are a set of vectors defining a projection from the original stimulus space to an N -dimensional subspace (which can be chosen to be orthogonal without loss of generality). Each vector, \vec{v}_i defines a new direction in the N -electrode stimulus space using a set of N weights. The magnitude of a projection of a particular stimulus onto a vector \vec{v}_i , indicates the alignment of that stimulus with that vector. In the experimental analysis of Maturana et al. (2016) and in the simulations reported here, the stimuli consisted of charge-balanced biphasic pulses of 500 μs phase duration and a 50 μs interphase gap. In this study, we ran only single-neuron simulations and, hence, all measured spikes were directly activated spikes (i.e., without input from the presynaptic network).

The set of vectors, \vec{v}_i , were estimated using a stimulation and recording scheme in which each stimulus current was chosen randomly and independently from a Gaussian distribution. The stimulus data, S_T , were filtered to include only the stimuli that resulted in RGC spiking responses, S_D . Using this spike-triggered data, the *spike-triggered covariance* matrix was calculated (Schwartz et al., 2006),

$$\mathbf{C}_S = \text{cov}(S_D), \quad (3.2)$$

where for N electrodes, \mathbf{C}_S is an $N \times N$ matrix. This covariance matrix was used to perform principal component analysis (PCA) to generate the set of vectors, \vec{v}_i : the principal components calculated on the spike-triggered data. These principal components are equivalent to the eigenvectors of \mathbf{C}_S . The corresponding eigenvalues indicate the amount of variation in spiking behavior that was explained by each principal component. To determine which principal components were significant, the eigenvalues were tested to determine if they were significantly different from chance (Schwartz et al., 2006). This was done by randomly shuffling the sequence of applied stimuli relative to spiking responses to create a set of randomized spike-triggered stimuli and recovering a null set of eigenvalues from this set using PCA. This process was completed 1000 times to construct a null distribution of eigenvalues and a 95% confidence interval. If the largest or smallest eigenvalue calculated on the raw data was outside the confidence intervals, it was considered significant. The principal component corresponding to the significant eigenvalue was then projected out of the raw data and the randomization test was repeated to test for the significance of the second largest/smallest eigenvalues until all remaining eigenvalues fell within the confidence bounds.

Maturana et al. (2016) demonstrated that the majority of information about direct spiking is captured by the first principal component, which reduces the N -dimensional projection shown in

Equation 3.1 to a one-dimensional projection,

$$P(R_t = \text{spike} | \vec{S}_t) = \mathcal{N}_1(\vec{v}_1 \cdot \vec{S}_t). \quad (3.3)$$

In the model above, \vec{v}_1 is the eigenvector for the spike-triggered covariance matrix which had the largest corresponding eigenvalue and indicated the cell's sensitivity to currents from each electrode; it will be referred to as the cell's electrical receptive field (ERF) in subsequent sections of this paper. The one-dimensional projection defined by \vec{v}_1 defines a simple weighted average applied to the vector of stimulus currents. This can also be referred to as the linear filter of the linear-nonlinear model. The terms 'linear filter', 'one-dimensional projection', and 'weighted average' are synonymous and are all used in the literature to describe the ERF or similar concepts.

To account for the asymmetrical responses to cathodic-first and anodic-first biphasic current pulses that has been presented in the literature (Jensen and Rizzo, 2006), the nonlinear function, \mathcal{N}_1 , has a double-sigmoid form,

$$\begin{aligned} P(R_t = \text{spike} | \vec{S}_t) &= \mathcal{N}_1(\vec{v}_1 \cdot \vec{S}_t) \\ &= \mathcal{N}_+(\vec{v}_1 \cdot \vec{S}_t) + \mathcal{N}_-(\vec{v}_1 \cdot \vec{S}_t). \end{aligned} \quad (3.4)$$

Here, a cathodic-first pulse is defined to have a positive amplitude and an anodic-first pulse has a negative amplitude. Equation 3.4 gives the spike probability in terms of a sum of the spiking probability in response to cathodic-first and anodic-first pulses. The net effect on the cell is cathodic-first or anodic-first depending on the sign of $\vec{v}_1 \cdot \vec{S}_t$. The sigmoidal nonlinearities are implemented as logistic functions,

$$\mathcal{N}_+(\vec{v}_1 \cdot \vec{S}_t) = \frac{a_+}{1 + \exp(-b_+(\vec{v}_1 \cdot \vec{S}_t - c_+))} \quad (3.5a)$$

$$\mathcal{N}_-(\vec{v}_1 \cdot \vec{S}_t) = a_- - \frac{a_-}{1 + \exp(-b_-(\vec{v}_1 \cdot \vec{S}_t - c_-))}, \quad (3.5b)$$

where the coefficients a_+ , b_+ , and c_+ represent the maximum spiking probability, slope, and threshold, respectively, and similarly for a_- , b_- , and c_- .

3.3.3 Cell morphology

Six cell morphologies were traced using neuron reconstruction software, NeuronStudio (v0.9.92) (Rodriguez et al., 2008; Wearne et al., 2005), from three-dimensional fluorescent confocal microscope images of the cells analyzed by Maturana et al. (2016). The SOCB was defined as the region 30 μm from the soma with a length of 40 μm . To assess the influence of cell location on its ERF, stimulation experiments were simulated for each cell at three characteristic locations. These locations were defined by the location of the soma: directly under electrode 11; midway between electrodes 11 and 14; and midway between electrodes 11, 14, and 15, as shown by the blue markers in Figure 3.2(a).

3.3.4 Biophysical model used for validation

To validate the linear-nonlinear model against more complex biophysical representations of retinal stimulation, a modified version of the four-layer volume conductor model described previously by Esler et al. (2018a) (Chapter 2) was used. This modified volume conductor model describes extracellular potential throughout the simulated retina, and uses a morphologically-accurate, Hodgkin-Huxley-type model implemented in the NEURON environment to describe neuron physiology.

The extracellular potential is described by a linear, four-layer model of the retina, including layer descriptions for the insulating substrate of the electrode array, the vitreous, the nerve fiber layer (NFL), and an combined approximation of the remaining retinal layers. Since the linear-nonlinear model to be validated in this paper aims to predict responses to subretinal stimulation, the order of model layers was modified from the epiretinal configuration used in previous work (Esler et al., 2018a) (Chapter 2) to be (1) insulator substrate, (2) isotropic neural layers, (3) NFL, and (4) vitreous, as shown in Figure 3.2(b). The continuity of current flow is conserved at all points in each layer and is consequently governed by a continuity equation:

$$\nabla \cdot \mathbf{J}_{e,\ell} = 0 \quad (3.6)$$

where $\mathbf{J}_{e,\ell}$ is the extracellular current density in layer ℓ , and $\nabla = \left[\partial/\partial x; \partial/\partial y; \partial/\partial z \right]$ is the differential operator. A generalized form of Ohm's Law is used to relate extracellular current density and potential:

$$\mathbf{J}_{e,\ell} = -\frac{1}{4\pi^2} \xi_\ell * \nabla V_{e,\ell} \quad (3.7)$$

where $V_{e,\ell}$ is the extracellular potential in layer ℓ and $*$ represents a convolution over space and time and captures the flow of current between points in the extracellular space. We refer the reader to Meffin et al. (2012); Tahayori et al. (2012); Meffin et al. (2014); Tahayori et al. (2014) for further details and explanation. The conductivity or admittivity kernels, ξ_ℓ , of each layer, ℓ , are given in Table 3.1. Admittivity is a spatially- and temporally-dependent generalization of conductivity and is the inverse of impedivity. The anisotropic admittivity of the NFL is incorporated into the admittivity kernel provided by the cellular composite model of Meffin et al. (2012); Tahayori et al. (2012); Meffin et al. (2014); Tahayori et al. (2014). Note that the conclusions presented in this paper rely only on the linearity of Equation 3.7, and so are not altered if this is substituted with a different linear model of extracellular potential or current flow.

Following substitution of Equation 3.7 into Equation 3.6, a solution to the system is determined by applying boundary conditions at the layer boundaries that ensure finite energy and continuity of current and potential across retinal layers. The system was formally solved in the spatio-temporal Fourier domain, where the expression for the admittivity of a tissue composed of parallel fibers, such as the NFL, can be simplified. The solution was then numerically inverted back into the space-time domain via the discrete inverse Fourier transform to give the extracellular potential, ϕ_ℓ on a uniform 3D grid of points that sampled the layered structure, as well as over time. Full details appear in Esler et al. (2018a) and Chapter 2 of this thesis.

Following simulation of the extracellular potential using the above volume conductor model, the resulting input into a neuron was determined using the activating function (Tahayori et al., 2012; Rattay, 1986, 1989, 1999). In the general case of arbitrary neuron morphology, multi-compartmental

neuron models can be used. In this work, a spatially-discretized description of the neural membrane with ionic currents was used for determining membrane potential dynamics,

$$C_{m,n} \frac{dV_{m,n}}{dt} = -I_{\text{ionic},n} + \frac{V_{m,n-1} - V_{m,n}}{\frac{1}{2}(R_{a,n-1} + R_{a,n})} + \frac{V_{m,n+1} - V_{m,n}}{\frac{1}{2}(R_{a,n+1} + R_{a,n})} + \dots + f_n, \quad (3.8)$$

where $V_{m,n}$, $R_{a,n}$, and $C_{m,n}$ are the membrane potential, axial resistance, and capacitance, respectively, at compartment n . $I_{\text{ionic},n}$ represents the conductance-based inputs into the cell resulting from active membrane dynamics. The above equation can accommodate neural compartments with more than two neighboring components (i.e., branch points) via the (\dots) term. The activating function, f_n , approximates the stimulation-induced transmembrane current and is represented here in discrete form as

$$f_n = \frac{V_{e,n-1} - V_{e,n}}{\frac{1}{2}(R_{a,n-1} + R_{a,n})} + \frac{V_{e,n+1} - V_{e,n}}{\frac{1}{2}(R_{a,n+1} + R_{a,n})} + \dots, \quad (3.9)$$

where $V_{e,n}$ is the extracellular potential at compartment n as determined by the linear extracellular component of the volume conductor model. Note that, in this work, the activating function is represented as a current, as by Rattay et al. (2017), as opposed to its alternative representation, which is normalized by capacitance (Rattay, 1999).

The activating function provides a framework for the combination of arbitrary models of extracellular potential with morphologically-accurate, conductance-based, multi-compartmental neural models. Here, models of extracellular potential and neural response were implemented using a combination of customized linear analysis software (MATLAB, Mathworks, Release 2016a) and conductance-based neural modeling software (NEURON, Release 7.4 (Hines and Carnevale, 1997)), respectively.

The model of the RGC neural membrane was the same as that used by Jeng et al. (2011), which was primarily adapted from the model presented by Sheasby and Fohlmeister (1999). Similar models have been validated in multiple previous studies (Schiefer and Grill, 2006; Greenberg et al., 1999; Fohlmeister and Miller, 1997). The model consisted of a leak current and five active membrane channels: voltage-gated sodium, calcium, delayed rectifier non-inactivating potassium, inactivating A-type potassium, and calcium-activated potassium. The sodium channel conductance at the SOCB was set to be 20 times that of the soma and the remainder of the axon. In line with Jeng et al. (2011), the voltage-gated potassium conductance was increased at the SOCB to a level that prevented spontaneous spiking.

RGC morphologies were embedded in the simulated retina by placing the soma at one of the locations indicated in Figure 3.2 by blue markers, 10 μm below the NFL with the axon extending into the NFL. Using trilinear interpolation, the time course of the extracellular potential was calculated at each three-dimensional compartment location from the grid of calculated extracellular potentials. These extracellular potentials were then used to calculate the resultant transmembrane current (or activating function) in each compartment using Equation (3.9). Activating functions values were then used to drive the simulation of cell spiking dynamics in NEURON. This biophysical modeling framework is equivalent to Scheme 1 in Figure 3.1. Simplifications of this model were subsequently applied to obtain models in the form of Scheme 2 and 3 in order to biophysically-recover the linear-nonlinear model.

3.3.5 Comparison of the electrical receptive field and the activating function

To validate the system for a wide range of stimulus inputs, the 20-electrode stimulus current space used by Maturana et al. (2016) was sampled from a zero-mean Gaussian distribution. Application of each generated combination of stimulus currents was simulated using the biophysical model to yield simulated spiking responses. The ERF of the simulated cell was then calculated from the model output. To establish a biophysical basis for the simplifying assumptions employed by the linear-nonlinear model, the ERF was compared to transmembrane currents (the activating function) at key locations in the simulated cell’s morphology.

Since the biophysical model describes extracellular potential and transmembrane current as linear functions of the stimulus currents, the value of the activating function at a particular cell compartment was the sum of contributions from each electrode at that location. Hence, for a particular cell compartment, the activating function (i.e., the stimulus-induced transmembrane current) was the sum of the transmembrane currents induced by each electrode. To compare the activating function at a particular compartment to the ERF of the cell (which is a vector of weights or sensitivities assigned to each electrode), the activating function was decomposed into the contributions made by each electrode and its amplitude was averaged over the course of the first phase of stimulation. The contribution of each electrode was calculated by stimulating each electrode with a nominal stimulus current of $1\ \mu\text{A}$ and calculating the resulting activating function. For each neuron compartment, this yielded a vector equal in length to the number of electrodes, where each element corresponded to the contribution of an electrode to the activating function at that compartment in μA . Henceforth, this vector will be referred to as the *activating vector* (AV). For a particular compartment, n , the activating vector will be denoted by \vec{f}_n in this paper, as distinct from the activating function itself (the combined influence of all electrodes), which will be denoted by f_n .

The ERF and activating vector were converted to unit vectors to eliminate differences in scale. To test for equivalence between the two vectors, the Pearson product-moment correlation (hereafter referred to as correlation) of the two was calculated and compared to a null distribution to determine statistical significance. For each calculated correlation, an appropriate null distribution was generated, defined as the population of correlations between activating vectors at different locations in the cell and the cell’s ERF. An activating vector was considered to be significantly similar to the ERF if the correlation between them was in the top 5% of correlations in the null distribution. For the activating vector at the soma, which is a single compartment, this null distribution consisted of the correlations of the activating vectors at each compartment with the ERF. For a SOCB comprised of n adjacent compartments, the null distribution consisted of the correlation of each set of n adjacent compartments in the cell with the ERF. To calculate significance of correlations across all cell reconstructions, one-sided Student’s t-tests were performed to compare the set of correlations (one per cell and location) with the pooled null distributions from all cells.

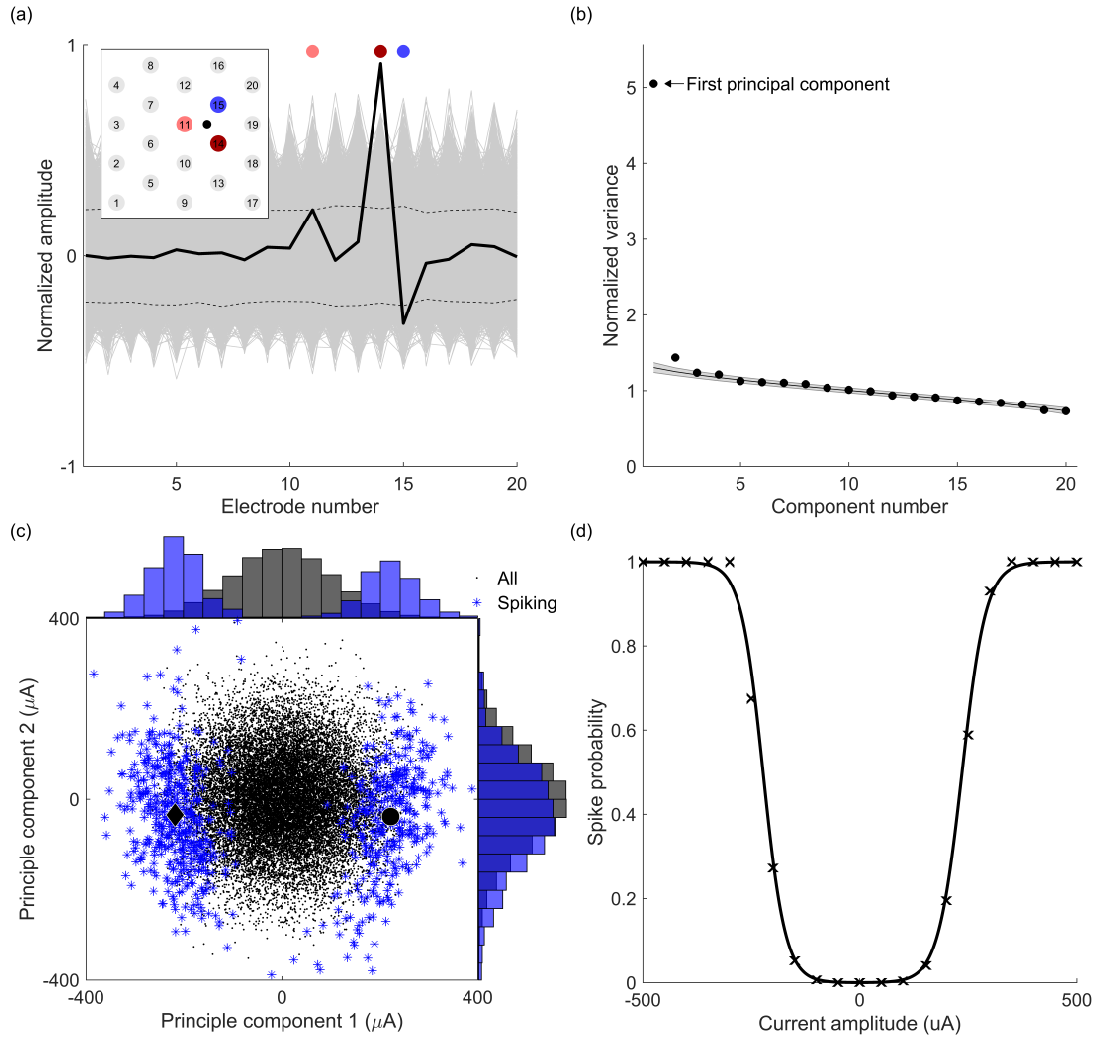


Figure 3.3: **The linear-nonlinear model for electrical stimulation.** Following simulation of 20,000 random amplitude multi-electrode stimuli, spike-triggered covariance (STC) (de Ruyter van Steveninck and Bialek, 1988; Schwartz et al., 2002, 2006) analysis was conducted on the stimuli which elicited a spiking response. (a) The ERF recovered from PCA on the spike-triggered data (solid black) was compared to the root mean square (dashed black) of a null distribution of ERFs. Electrodes found to be significant are indicated by colored dots. The inset shows the location of the cell being stimulated (black) relative to the array. Colors shown on electrodes correspond to the magnitudes in the line plot with dark red corresponding to positive 1 and dark blue corresponding to negative 1. (b) Eigenvalues of the spike-triggered stimuli recovered by PCA are plotted, normalized by the variance of the input stimuli. The shaded region represents the 95% confidence intervals around the null model, revealing one dominant and one marginally significant component. (c) All stimuli were projected onto the first two principal components, with spike-triggered data represented by blue stars. Positive and negative stimuli were defined using the sign of the projection onto the first component, yielding positive (circle) and negative (diamond) spike-triggered averages. (d) The nonlinear function recovered by fitting a double sigmoid (solid black) to the spike probability data projected by the ERF (crosses).

3.4 Results

3.4.1 Recovery of the electrical receptive field

To demonstrate the relevance of the approach used in this paper, the same statistical analysis conducted by Maturana et al. (2016) on their experimental data was performed here on the simulated data of the full biophysical model (Scheme 1 in Figure 3.1). By mimicking the experimental setup of Maturana et al. (2016) in the simulations presented here, experimental and simulation results can be compared. Simulations of neural activity in response to 20,000 random stimuli were conducted for each of the six cells at each of the three locations, indicated by the blue markers in Figure 3.2. Each simulation consisted of a single neuron stimulated simultaneously by biphasic pulses on all electrodes, where the amplitude of each pulse was randomly selected from a zero-mean Gaussian distribution with standard deviation between 100 and 150 μA , chosen to ensure a suitable number of stimulus applications ($>2\%$) elicited spikes.

To estimate the ERF of the cell, principal component analysis was conducted on the applied stimuli that elicited a spiking response from the cell (see Methods). Principal component analysis provides a way to determine which combinations of electrode currents effect spike probability. In principle there may be several distinct combinations of electrodes that effect spike probability differently. These combinations correspond to the different principal components.

Figure 3.3 shows an example of the outcome of this analysis. In agreement with Maturana et al. (2016), the neural responses of the majority of cells were well predicted using a single principal component, defined as the ERF. The first principal component is shown in Figure 3.3(a), along with the location of the electrodes and the cell (black dot) in the inset. The ERF corresponds to a projection of the 20-electrode stimulus space onto the neuron’s preferred stimulus, representing a one-dimensional linear projection. For the neuron shown in Figure 3.3, the spiking response could be well predicted by a filter corresponding to a weighted sum of the currents from the three surrounding electrodes. In the majority of cases, the variance in spiking probability described by the first component was substantially larger than the second and higher components. This is exemplified in Figure 3.3(c), which shows the set of all 20,000 stimuli (black dots) projected onto the first two principal components. Histograms along each axis show the distributions of spike-triggered stimuli along each component (blue) relative to the original distribution of stimuli (grey), revealing far greater divergence from the original distribution along the first component. This indicates that the magnitude of stimuli along this dimension (i.e., their projection) determined the probability of spiking.

This is quantified further in Figure 3.3(b), which shows that the first principal component dominated. By conducting the statistical test described in Section 3.3.2, it was revealed that only the first principal component was significant for the sample cell shown. The distance between the null distribution mean (black line) and the eigenvalues was approximately 20 times greater for the first principal component than the second. Finally, to complete the estimation of the linear-nonlinear model, the one-dimensionally projected stimuli were passed through a double sigmoid nonlinearity, such as that shown in Figure 3.3(d). This function was fit to the cell’s output spikes to yield a prediction of spiking probability. Each of these results is in agreement with the results of similar analysis conducted on experimental data by Maturana et al. (2016).

As found in several previous experimental studies, some cells (one of six in this study) appeared

to have an ERF that required more than one dimension or filter (Maturana et al., 2016; Jepson et al., 2014), meaning that more than one important principal component was recovered during PCA. This example is explored in more depth in section 3.4.6.

3.4.2 Linearity of the electric field and the activating function

The stimulation of neurons by one or more electrodes is described biophysically using the combination of linear and non-linear transformations. This biophysical description is mirrored in the structure of the phenomenological linear-nonlinear model. However, it is not immediately clear which aspects of the biophysical system can be approximated as linear before applying a nonlinearity, as illustrated in Figure 3.1. The chain of transformations that map multi-electrode stimuli to neuron spiking can be broken into (1) the generation of an electrical potential around the cell by the electrodes, (2) generation of the activating function (transmembrane currents) at each location in the cell due to the extracellular potential, (3) cellular integration of transmembrane currents, and (4) neuronal spiking based on integrated current inputs. As in the full biophysical model used in this work, steps 1 and 2 are commonly treated as linear with respect to stimulus currents, and steps 3 and 4 are commonly treated as nonlinear (e.g., multi-compartment conductance-based models). In this section, we will investigate whether the linear activating function, representing the first two transformations above, can be used to reproduce the ERF. This approach is equivalent to Scheme 2 in Figure 3.1, where the activating function at one or few locations drive the cell and cellular integration can be neglected. Following this, in Sections 3.4.4 and 3.4.5, we will investigate whether step 3 - the integration of activating function inputs - can also be approximated as linear, with only spiking behavior being treated as nonlinear.

The input into the cell from electrical stimulation (the activating function) is determined by the extracellular potential via a second-order difference equation, as shown in Equation 3.9. Hence, the activating function is a linear function of the electrode stimuli. Figure 3.4 shows an example of extracellular potential and the activating function for the simulated system. As illustrated in Figure 3.4(a-f), due to linearity, both the extracellular potential and the activating function induced by multiple electrode stimulation can be determined by superposition of the induced behavior from stimulation with each electrode individually. This quality allows us to separate out the activating function for each compartment into an activating vector with components arising from each electrode. Subsequently, to determine the transmembrane current induced by arbitrary simultaneous stimulation from all electrodes, we need only to scale the elements of this vector by the set of stimulus current amplitudes and sum together to obtain the superposition.

As can be seen in Figure 3.4(d-f), the activating function can vary significantly from one compartment to another, even over a small distance. The reason for this is illustrated in Figure 3.4(g) and (h). The first derivative of the extracellular potential along a neurite (i.e., an axon or dendrite) is proportional to the current flowing along that neurite. The second derivative of the extracellular potential then indicates the change in that extracellular current flow along the neurite, corresponding to the amount of current flowing into the neurite: the activating function. As a second derivative in space, the sign and magnitude of the activating function at a particular location is highly sensitive to the three-dimensional orientation and curvature of the neurite at that location. For a convoluted structure such as a neuron, this results in the rapidly changing activating function seen in Figure 3.4(d-f) and exemplified in Figure 3.4(h).

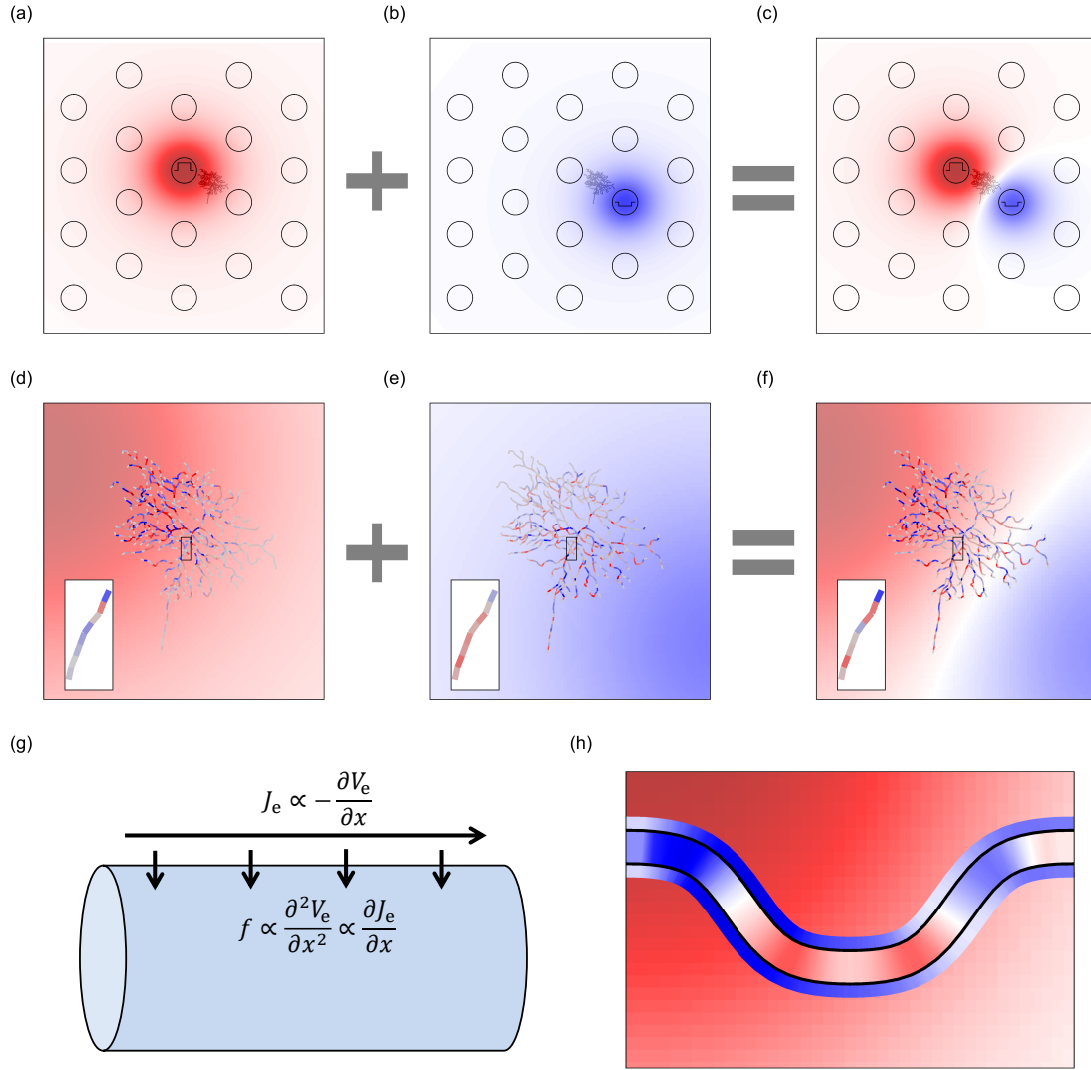


Figure 3.4: **Illustration of the linear component of the biophysical model.** (a-c) Due to linearity, the addition of the induced extracellular potential from two single-electrode stimuli is equal to the induced extracellular potential from simultaneous stimulation with both electrodes. Stimuli waveforms are shown as small traces over electrodes, and the outline of the cell being stimulated is shown to right of the center electrode. (d-f) Similarly, the induced activating function at a cell compartment in response to two single-electrode stimuli (two components of the activating vector) can be summed together to determine the induced activating function in response to simultaneous stimulation. Insets show the activating function along the SOCB. (g) Illustration of the first and second derivative of the extracellular potential along a neurite (i.e., an axon or dendrite). The first derivative is proportional to the extracellular axial current flow (J_e) and the second derivative is proportional the transmembrane current (i.e., the activating function, f). (h) Illustration of the sensitivity of the activating function to both extracellular potential (background color) and neurite orientation for a short segment of curved neurite. The color of the outer layer of the neurite represents the first derivative of the extracellular potential in space along its axis. The inner layer of the neurite represents the second derivative of the extracellular potential along its axis. In all subplots and for all measures shown (activating function, extracellular potential, and its derivatives), red represents a positive value, white represents zero, and blue represents a negative value.

3.4.3 Comparison of the electrical receptive field and the activating function

To assess the validity of Scheme 2 in explaining the linear-nonlinear model, each RGC's ERF was compared to the activating vector at the soma and the SOCB. For these and subsequent comparisons, the activating vector refers to the vector of average current flow into a compartment during the first phase of a 1 μ A stimulus from each electrode. For the SOCB, since it consisted of a series of adjacent cell compartments, the activating vector across all those compartments was first summed together to determine the total current flowing into or out of the SOCB. To compare ERFs with activating vectors, both were first converted to unit vectors to eliminate differences in scale.

In agreement with the assumption that the SOCB is the site of action potential initiation (Fried et al., 2009), the activating vector at the SOCB generally matched well with the ERF of the cell (Figure 3.5(b)), whereas the soma had less agreement (Figure 3.5(c)). The discrepancies seen between the SOCB activating vector and the ERF were likely due to transmembrane currents (activating function inputs) at other locations in the cell having an impact on the cell response. Due to the large size and proximity of the soma with the SOCB, axial current flow into the SOCB that originated from the soma was incorporated. A weighted sum, \vec{f}_{ws} , of the activating vector at the SOCB and soma was calculated as

$$\vec{f}_{ws} = (1 - w_{\text{soma}})\vec{f}_{\text{SOCB}} + w_{\text{soma}}\vec{f}_{\text{soma}}. \quad (3.10)$$

This represents a model of the SOCB in which membrane depolarization and spiking is driven by transmembrane currents entering at the SOCB and axial currents that originate as transmembrane currents in the soma.

The weight attributed to the soma activating vector, w_{soma} , was varied from 0 to 1 in increments of 0.0001 and the optimal value selected based on correlation with the ERF. The optimized weighted sum of activating vectors generally achieved a markedly higher correlation with the cell's ERF when compared to the SOCB or soma activating vectors alone. Figure 3.5(a) shows this optimization for the same cell analyzed in Figure 3.3, with the correlation of the SOCB, soma, and optimized weighted combination of both indicated by square black markers. For the example shown, the optimal soma weight was approximately 0.36, giving an optimized correlation with the ERF of 0.98, compared with 0.82 for the SOCB activating vector and 0.37 for the soma activating vector. Figures 3.5(a), (b), and (c) compare the ERF with the SOCB, soma, and weighted sum activating vectors, respectively.

To determine the overall equivalence of the activating vector and the ERF across all six cells and at all three locations (see Figure 3.2), the activating vector and ERF vectors from each combination (18 in total) were concatenated before calculating correlation between the two concatenated vectors. This was completed for the activating vector at the SOCB, the soma, and the optimized weighted sum of both. This demonstrated that the activating vector at the SOCB, with a correlation of 0.841 ($p = 0.009$), performed far better at predicting the ERF than the activating vector at the soma, with a correlation of 0.730 ($p = 0.281$). Furthermore, the optimized weighted sum of the SOCB and the soma activating vector did better still, with an overall correlation of 0.932 ($p = 0.000120$). Across all cells and locations tested, optimized soma weights ranged from 0 to 0.36, with an average of 0.088. The p-values reported above were calculated as described in Section 3.3.5.

This result suggests that the ERF captures aspects of the activating function, or the currents entering the cell, as well as the intracellular integration of an important subset of those inputs; in this

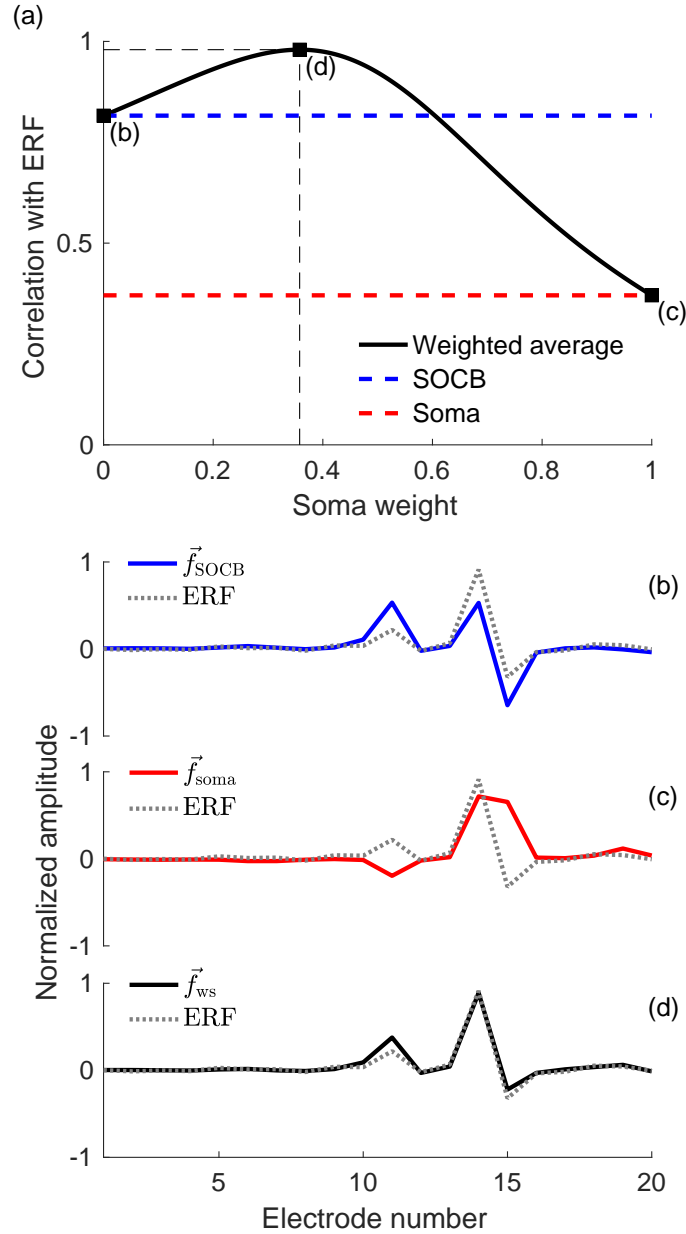


Figure 3.5: **Comparison of the electrical receptive field and the activating vector.** (a) The weighted sum of the activating vector at the SOCB and soma, calculated as $\vec{f}_{\text{ws}} = (1 - w_{\text{soma}})\vec{f}_{\text{SOCB}} + w_{\text{soma}}\vec{f}_{\text{soma}}$, is compared to the ERF. The correlation of the weighted sum of activating vectors and the ERF is maximized by a soma weighting of approximately 0.36 (dashed black lines), representing the influence on spiking activity of axial currents flowing into the SOCB from the soma. The correlation of the activating vector at the SOCB and the soma with the ERF is indicated by blue and red dashed lines, respectively. (b), (c), and (d) show comparisons of the ERF (dotted grey) with the SOCB (blue), soma (red) and optimized weighted sum activating vectors (black), respectively. Square black markers in (a) correspond to the activating vector combinations in (b), (c), and (d). The analysis presented in this Figure was conducted on the same cell as for Figure 3.3.

case, the inputs into the SOCB and the soma. Importantly, since the ERF has been reconstructed using a linear combination of the activating vectors at the SOCB and the soma, it suggests that intracellular integration of inputs from these compartments can be approximated by a linear function.

3.4.4 Comparison of active and leaky integrate-and-fire neuron models of RGCs

The true biophysical system describing neural activation of RGCs is complex, involving nonlinear spatio-temporal dynamics across the compartments of the neuron. To clarify the essential biophysical processes that allow the behavior of this model to be captured by the linear-nonlinear model and the ERF, we considered whether a simpler and largely linear biophysical model (Scheme 3) would yield a similar ERF to the biophysical model with active cell dynamics (Scheme 1).

For cells in which spiking is thought to predominantly initiate at one or a small number of locations (e.g., the SOCB), it is possible that the spiking response of the cell can be accurately represented by a biophysical model that largely neglects active membrane dynamics, with only fixed spiking thresholds applied at the sites of spike initiation and passive membrane properties elsewhere (i.e., leak conductance and capacitance). In contrast to the active biophysical model of Scheme 1, this approach treats the summation/integration of spatially distributed inputs into the cell as linear, in addition to using a linear description of extracellular potential and the activating function, as in Scheme 3 (see Figure 3.1). In this model, nonlinear behavior is modeled only by the application of a spiking threshold at one or several fixed locations, operating on the passive/linear membrane potential. To investigate this, the results of an active model, such as that analyzed in Figures 3.3 and 3.5, were compared to the response from a neuron model with passive (linear) membrane dynamics and a firing threshold applied at the SOCB. The passive model differs in that all active membrane dynamics are replaced by a single firing threshold at the SOCB, but the models are otherwise identical. For consistency, the threshold applied at the SOCB was chosen to ensure a similar proportion of stimuli resulted in spikes as was seen for the active model. These integrate-and-fire thresholds varied from 8 mV to 15 mV above resting membrane potential, with a mean of 11.8 mV.

The result of this comparison is shown in Figure 3.6 for the same cell analyzed in Figures 3.3 and 3.5. Figure 3.6(a) shows that a single principal component dominated for both the active and passive models. Although in both active and passive cases only one component was significant, the second component in the active model contained a small amount of information. This result suggests that the second component for the active model was influenced by spikes initiated at locations other than the SOCB. Despite this small discrepancy, Figure 3.6(c) shows that the ERFs recovered from both models were similar. The observed discrepancies can be related to the differences between subthreshold dynamics in active and passive models, and are discussed further in section 3.5.2. In addition, by conducting the same analysis of activating vectors at the SOCB and soma, almost identical weights for each location were suggested by both active and passive models (as demonstrated in Figure 3.6(b)). The soma weight was 0.358 for the active model and 0.366 for the passive model, resulting in the almost identical weighted sum activating vectors shown in Figure 3.6(d).

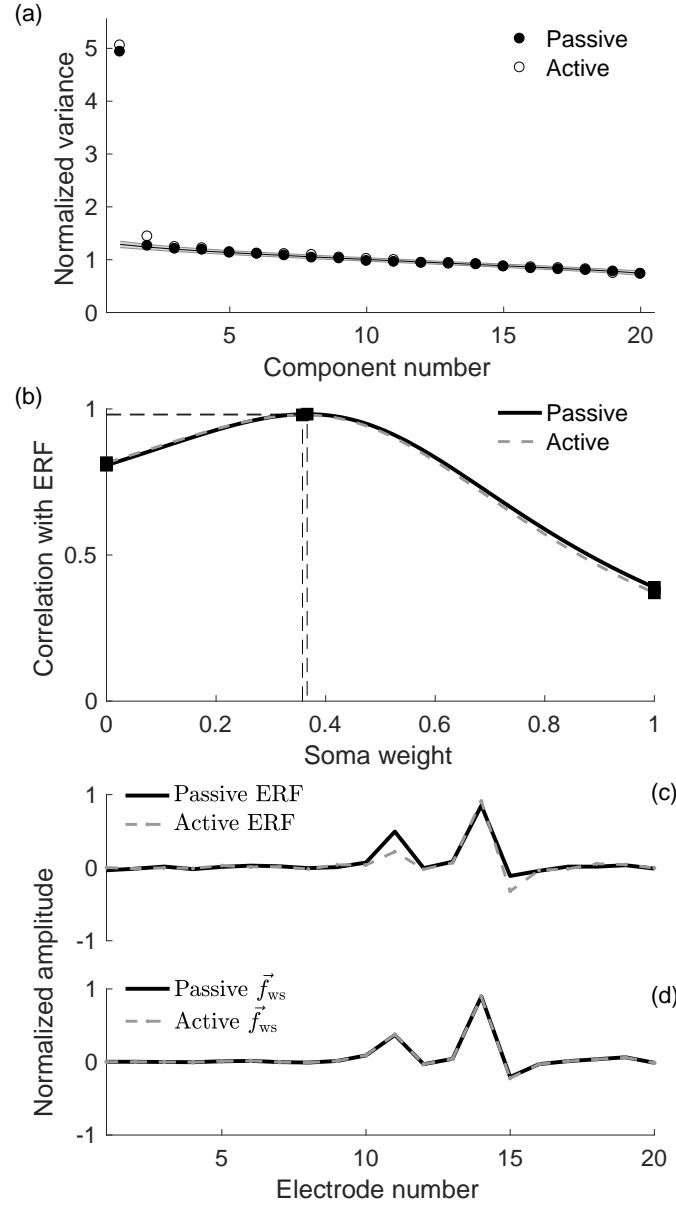


Figure 3.6: Spike-triggered analyses conducted on cells with active versus leaky integrate-and-fire (LIF) membrane dynamics. (a) The normalized eigenvalues recovered from PCA on spike-triggered data simulated using active and LIF (passive) models of neural activity. Characteristics of the active model are mostly captured using a passive model of membrane potential with a single spiking threshold applied at the SOCB. (b) The correlation between 1) the weighted sum of SOCB and soma activating vectors and 2) the ERF calculated from active and LIF neuron models. The optimal weighted sum of activating vectors at the SOCB and soma are very similar for both models, implying similar contributions from each location. (c) Comparisons of the ERFs recovered from PCA on data simulated using active (dashed) and passive (solid) models. (d) Comparison of optimized activating vector weighted sums between active (dashed) and passive (solid) neuron models. The analysis presented in this Figure was conducted on the same cell as for Figures 3.3 and 3.5.

3.4.5 Biophysical determination of activating function contributions to the ERF

The forgoing result indicates that the ERF of the linear-nonlinear model is well approximated by an optimal weighting of the activating vectors from a small number of compartments: the SOCB and the soma. Furthermore, we have shown that the ERF is predominantly determined by passive integration of inputs from these locations to give the membrane potential at the SOCB. More generally, this indicates that the ERF is the result of passive integration of inputs at locations near the site or sites of action potential initiation. To test this further, we examined whether the weighting of inputs at different compartments could be determined biophysically from the linear properties of the cell as opposed to through a brute-force optimization, and whether or not this weighting is effective in predicting the ERF.

Intuitively, the size of the contribution of each compartment to membrane depolarization at the site of action potential initiation should depend on the morphology and membrane dynamics of the cell only. If the cell can be accurately described using a linear membrane model with spiking determined by a nonlinear threshold at a single location, then linear cable theory and its extension to multi-compartmental models implies that the membrane depolarization at that location at a given point in time can be approximated by

$$V_{m,t} \approx \vec{S}_t \cdot \vec{f}_{ws} = \vec{S}_t \cdot \sum_n w_n \vec{f}_n, \quad (3.11)$$

where \vec{f}_n is the activating vector for the n th compartment in the cell and S_t is the vector of current amplitudes delivered to each electrode in a stimulus at time, t . The set of w_n 's are the weights assigned to each compartment, describing their influence on the membrane potential at the SOCB, and \vec{f}_{ws} is the weighted sum of activating vectors. Hence, at each point in time, the linear intracellular integration of inputs into the cell can be represented by a linear combination of the current flowing into each compartment. In the case of a simple cylinder, the weight assigned to each compartment, w_n , in Equation 3.11 can be determined analytically from the solution of the cable equation. This is demonstrated in detail in Appendix A.

Figure 3.7 demonstrates the linear integration at the SOCB of several spatially-distributed cellular inputs. Figure 3.7(a) shows the locations of four compartments receiving injections: one in a dendrite, one in the soma, one in the SOCB, and one in the axon. The calculated total SOCB depolarization, shown in Figure 3.7(f), is equal to the sum of the depolarization observed due to each input in isolation, shown in Figure 3.7(b-e). Since the location of one of these injections corresponds to the location at which we are measuring depolarization (the SOCB), this injection has the largest contribution while compartments further away have progressively attenuated contributions. The depolarizations are also more delayed and more temporally spread out the further the injection site is from the SOCB. The relative size of the contribution of each compartment, w_n , can thus be determined by injecting current into each compartment in a passive model of the cell and measuring the resulting membrane depolarization at the site of spike initiation.

Figure 3.8 shows the result of this analysis graphically for the same cell analyzed in Figures 3.3, 3.5, and 3.6. Figure 3.8(a) shows the weights calculated for each compartment, as determined by the size of the membrane depolarization at the SOCB in response to a current injection, measured shortly after stimulus onset (approximately 0.1 ms after onset). This time is analogous to t_{meas} in

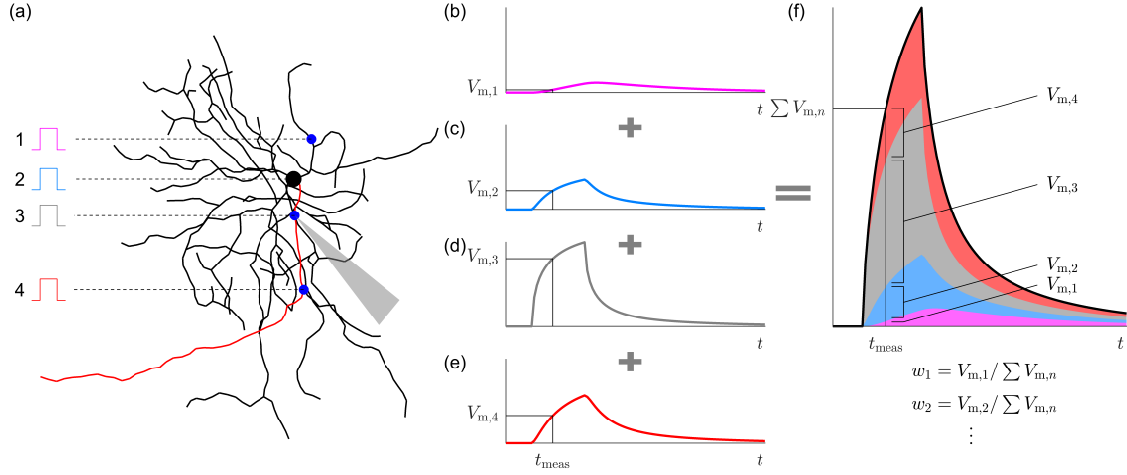


Figure 3.7: **Linear integration of activating function inputs.** When modeling a multi-compartment neuron with passive membrane dynamics, integration of spatially-distributed current inputs is linear. (a) Equal magnitude currents were injected one by one at several locations (indicated by dotted lines) in a simulated multi-compartmental neuron. The locations correspond to a dendrite (1), the soma (2), the SOCB (3), and the axon (4). The axon is colored red and the soma is shown as a large black circle. The membrane depolarization in response to each of these injections was recorded at location 3 in the SOCB, indicated by the grey triangle. (b-e) Simulated membrane depolarization, d_i , at a single location in the SOCB is shown in response to each current injection. The color of the trace corresponds to the color of the injection in (a). (f) Stacked area plot showing depolarization at the SOCB in response to simultaneous current injection at all four locations. Colored sections of the area plot indicate the contribution of the individual injections/depolarizations. Due to the linearity of integration, the depolarization observed when injecting at all locations simultaneously is equal to the sum of the depolarizations observed from each injection in isolation. For a particular point in time, t_{meas} , the linear integration of inputs at the SOCB (or any location) can be represented as a weighted sum of injected inputs. Following injections of equal magnitude into every compartment, the weight of a particular compartment can be calculated as the ratio of its contribution to the sum of all depolarizations at that point in time.

Figure 3.7. To convert voltage depolarizations at the SOCB to weight percentages, depolarizations for all compartments were normalized by their sum and multiplied by 100. To determine which compartments in the cell would be most influential on spiking, the product of the passive compartment weights and the norm of their activating vectors was calculated and is displayed in Figure 3.8(b) (the norm of the activating vector at a compartment is a measure of the influence of electrical stimulation on that compartment). The size of the bubbles represents the relative expected influence on the cell's response. Note that while the soma has the greatest weight of any single compartment, the overall influence of the SOCB is greater because it comprises several compartments with a greater sum of weights.

To reconstruct the ERF using a fully biophysically-determined linear combination of activating vectors, the neuron model compartments were progressively weighted and summed together according to the importance order indicated in Figure 3.8(b), each weighed by the weights indicated in Figure 3.8(a). The result of this process is shown in Figure 3.8(c), with almost perfect agreement between the ERF and a weighted combination of the activating vectors of 10 compartments (correlation of 0.997, Figure 3.8(d) top). Namely, these compartments were the somatic compartment, eight of fourteen SOCB compartments and one additional axonal compartment. The correlation was reduced by the addition of more components to the linear combination, with the correlation of the ERF using all compartments being 0.904. The introduction of this error was likely to be due to the differences between passive and active sub-threshold membrane dynamics and is explored in the Discussion. These results indicate that RGC responses to electrical stimuli are primarily driven by an approximately linear integration of inputs (as in Scheme 3 of Figure 3.1), and that this linear integration corresponds to the ERF.

3.4.6 Multiple electrical receptive fields for multiple initiation sites

Unlike the example cell used in Figure 3.3, principal component analysis performed on one cell analyzed did not yield a single dominant component. The requirement for multiple principal components means that the stimulus-space boundaries between spike-triggered and non-spike-triggered data are non-linear, as shown in Figure 3.9(a). For this cell, it was observed that although some spikes were initiated at the SOCB a subset of spikes were initiated elsewhere in the cell, in the soma in this case. Figure 3.9(a) shows the stimuli projected onto the first two principal components. In both Figures 3.3(c) and 3.9(a), the boundaries between spiking and non-spiking data represent a map of the cell's stimulus threshold with respect to the first two principal components. Unlike in Figure 3.3(c), in which the stimulus threshold is approximately linear, the boundaries in Figure 3.9(a) between spiking and non-spiking data are better described using a piece-wise linear function. Similar effects are expected in situations where a proportion of spikes are initiated in the distal axon and backpropagate to the SOCB.

Following simulation of 20,000 randomly generated stimuli, spikes were sorted by initiation location, either the SOCB (blue) or the soma (red). For the cell used to generate Figure 3.9, the activating function at the soma had opposite polarity to the activating function at the SOCB for the vast majority of stimuli, meaning that a depolarizing input at the SOCB almost always coincided with a hyperpolarizing input at the soma. Due to this, as illustrated in Figure 3.10, spikes were able to be classified as soma-initiated spikes if the membrane potential at the SOCB was depolarized (hyperpolarized) at stimulus onset, before the arrival of axial current originating from the soma, but hyperpolarized (depolarized) at the end of the first stimulus phase, due to the dominating influence of axial current from

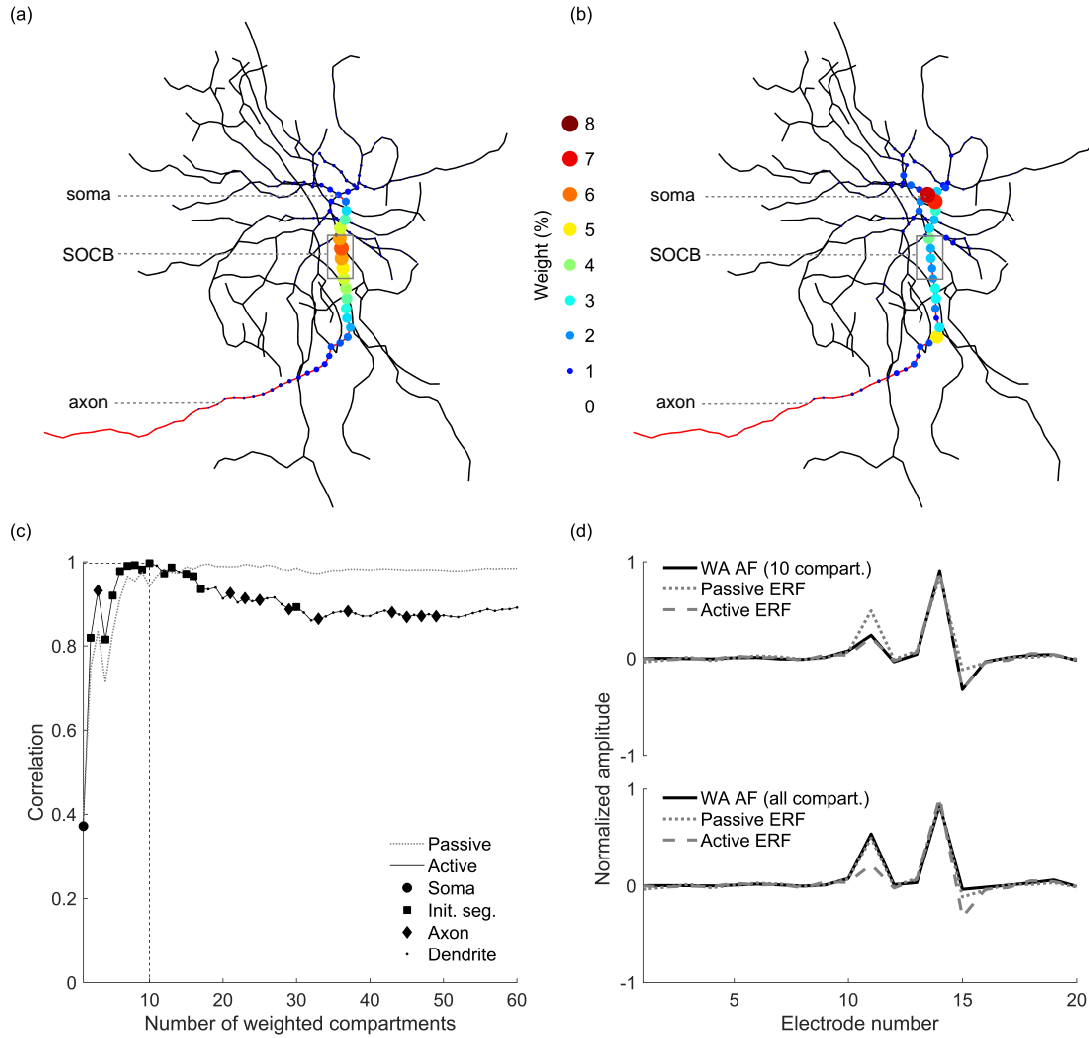


Figure 3.8: The biophysical representation of the linear-nonlinear model. (a) Biophysical weights are determined from the passive membrane depolarization resulting from the injection of equal current into each compartment. These weights are represented by bubble size and color. (b) To determine the order of importance of each compartment on spiking at the SOCB, the biophysical weight is multiplied by the norm of the activating vector at each compartment. Larger bubbles represent compartments expected to have a greater influence on cell response. (c) Using the importance ordering depicted in (b), the activating vectors at each compartment are progressively weighted together using the weights depicted in (a), and correlation with the ERF is calculated. The type of compartment (soma/initial segment/axon/dendrite) added to the weighted sum activating vector at each step is indicated by marker shape. Note that the initial segment consists of the first 130 μm of the axon originating from the soma and includes the SOCB. The best reproduction of the ERF is achieved with 10 compartments, indicated by dashed black lines. (d) The 10 compartment (top) and all compartment (bottom) biophysically-weighted activating vector is compared to the ERFs recovered from LIF (passive) and active neuron models. The analysis presented in this Figure was conducted on the same cell as for Figures 3.3, 3.5, and 3.6.

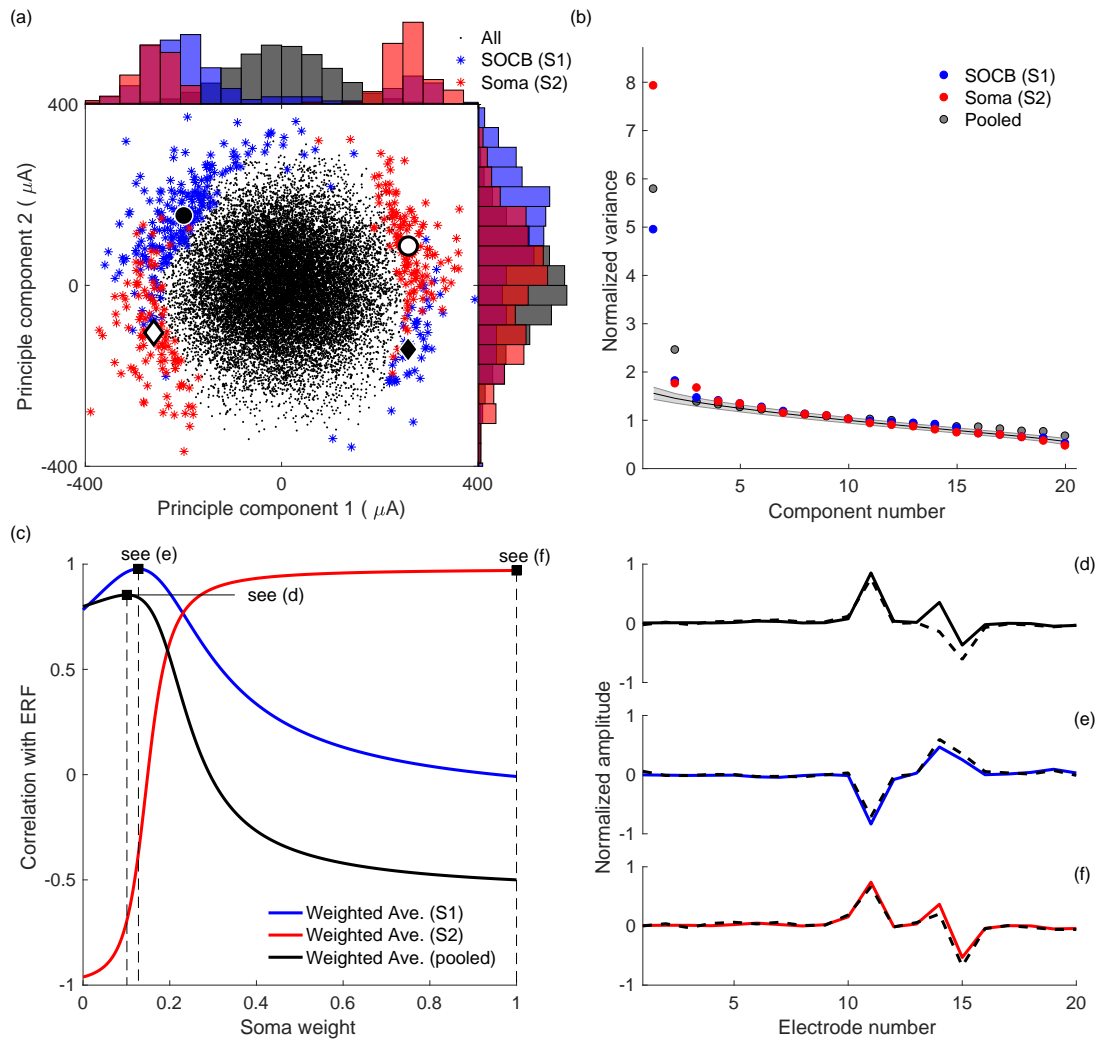


Figure 3.9: Multiple electrical receptive fields for multiple sites of action potential initiation. (a) Stimuli are projected onto the first two principal components recovered from PCA on all spike-triggered data. Spikes found to be initiated at the SOCB (S1) and the soma (S2) are indicated by blue and red stars, respectively. PCA was repeated on the SOCB- and soma-initiated spike-triggered data separately, yielding independent positive (circle) and negative (diamond) spike-triggered averages for each spiking location. (b) Eigenvalues of the spike-triggered stimuli recovered by PCA are plotted. Eigenvalues recovered from PCA of all spiking data, SOCB-spiking data, and soma-spiking data are shown by gray, blue, and red circles, respectively. The shaded region represents the 95% confidence intervals around the null model for the SOCB-spiking analysis. The ratio of the first to second eigenvalue is decreased in each of the SOCB- and soma-only analyses when compared to the pooled spikes. (c) The optimal activating vector weight for the soma is calculated for each spiking initiation location, with the optimization curves shown in blue and red. The solid black curve shows the same analysis for the ERF calculated on the pooled spiking data. (d-f) The pooled (d), SOCB-specific (e), and soma-specific (f) ERFs are compared to the optimized weighted sum activating vector reproductions at each location. In each sub-figure, the dashed line shows the appropriate ERF and the solid line shows the weighted activating vector. Note that for the soma optimization (red), the sum of the activating vector at the soma and the set of adjacent compartments was used to represent the soma. Square black markers in (c) correspond to the activating vector combinations in (d), (e), and (f).

the soma. When this behavior was observed, it indicated that the membrane potential at the SOCB was being primarily driven by the soma, and hence these spikes were classified as soma-initiated spikes. Spikes were classified as SOCB-initiated spikes if the membrane potential at the SOCB was depolarized (hyperpolarized) at stimulus onset and remained depolarized (hyperpolarized) for the remainder of the first stimulus phase.

In contrast to the pooled spike-triggered data, each of the sub-populations of spikes was more accurately described by a single linear filter (a weighted sum of stimulus currents). This can be seen in the eigenvalues shown in Figure 3.9(b), which show that the relative importance of the second eigenvalue was much larger for the pooled spike-triggered data when compared to the SOCB- and soma-initiated sub-populations. Hence, although the ERF of a cell may not be fully captured by a single linear filter, the ERF associated with each initiation site can be.

For both spike-initiation sites, a weighted sum of the activating vector at the SOCB and soma was optimized to maximize correlation with the relevant ERF, as shown in Figure 3.9(c). This shows that for spikes initiated at the SOCB (blue), a weighted sum of the activating vector at the SOCB and soma best reproduced the ERF ($r = 0.978$, $w_{\text{soma}} = 0.128$) and for spikes initiated at the soma (red), the weighted sum of the activating vector at the soma and adjacent compartments best reproduced the ERF ($r = 0.970$, $w_{\text{soma}} = 1$). In both cases, the appropriate ERF was more accurately reconstructed using activating vectors than could be achieved for the ERF determined from the pooled spiking data (black) ($r = 0.854$, $w_{\text{soma}} = 0.101$). The difference in the ERF between the SOCB and the soma, and the high accuracy of these optimized fits is evident in Figures 3.9(e) and (f).

Finally, the ERF associated with each site of action potential initiation was reconstructed biophysically using a linear combination of activating vectors at nearby compartments. The weights and order of importance of each compartment in the cell were determined as in the previous section: by measuring membrane depolarization at the site of action potential initiation in response to current injection in each compartment. This yielded weights for each compartment for each of the SOCB and the soma, as shown in Figure 3.11(a) and (d), respectively. By progressively combining the activating vectors from each compartment using these weights, it was found that the ERF of each site could be accurately reproduced by the linear combination of the activating vectors from only 8 nearby compartments, as shown by Figure 3.11(b) and (e).

3.5 Discussion

Using a detailed modeling approach, the biophysical interpretation of the linear electrical receptive field of the retinal ganglion cell has been elucidated using a vector representation of the activating function. The activating vector of a compartment gives the transmembrane currents induced by stimulation with each electrode. In all six cells examined, the ERF could be accurately reconstructed using weighted sums of activating vectors, which are, in turn, linear functions of the stimulus currents. For cells for which multiple sites of spike initiation exist, the analysis presented here suggests that, if an ERF is defined per site, each can be accurately reconstructed using weighted sums of activating vectors. In each case, the weights comprising these linear combinations are determined by the passive properties of the cell: its morphology and passive membrane dynamics.

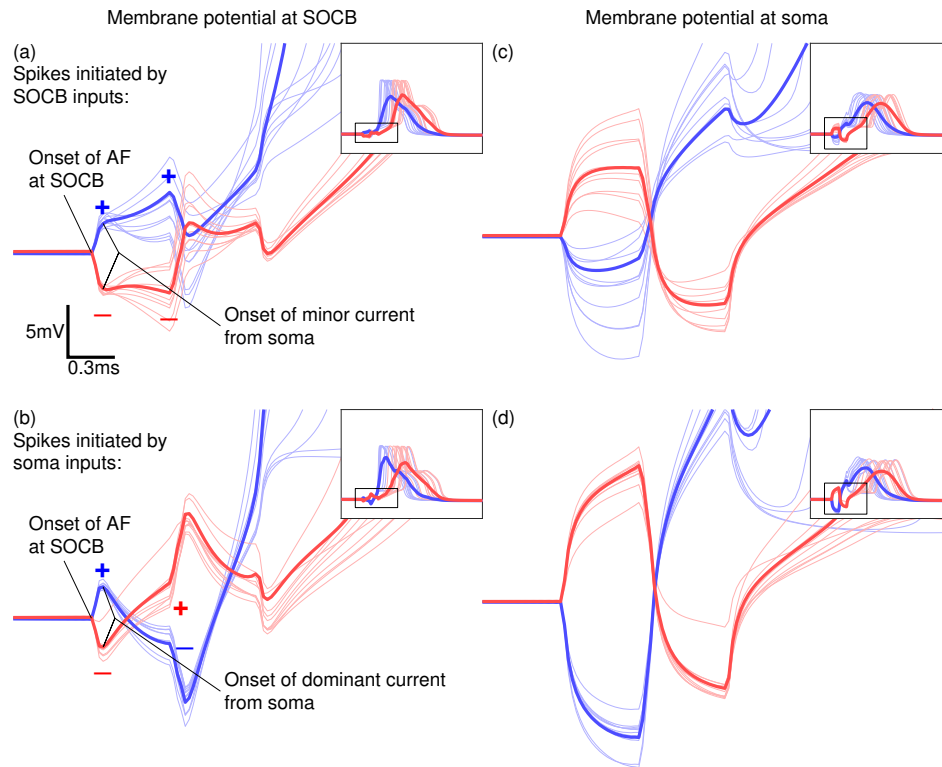


Figure 3.10: **Demonstration of sorting SOCB and soma generated spikes.** (a) Membrane potential traces from the SOCB for spikes initiated by the activating function (AF) entering the SOCB. For both net cathodic-first (blue) and net anodic-first (red) current inputs, SOCB-initiated spikes are characterized by the polarity of the depolarization seen at stimulus onset persisting through the first stimulus phase. When the axial current originating from the soma propagates to the SOCB (indicated), it results in a minor, but not dominant suppression of polarization. Thick blue and red lines show the average membrane potential for SOCB-initiated spikes. (b) Membrane potential traces from the SOCB for spikes initiated by the activating function entering the soma. For both net cathodic-first (blue) and net anodic-first (red) current inputs, soma-initiated spikes are characterized by the polarity of the depolarization seen at stimulus onset switching during the first stimulus phase. When the axial current originating from the soma propagates to the SOCB (indicated), it results in causes a reversal of polarization at the SOCB. Thick blue and red lines show the average membrane potential for soma-initiated spikes. The full membrane potential traces are shown in insets. (c) and (d) show membrane potential traces at the soma in response to the same stimuli as in (a) and (b), respectively. The analysis presented in this Figure was conducted on the same cell as for Figure 3.9.

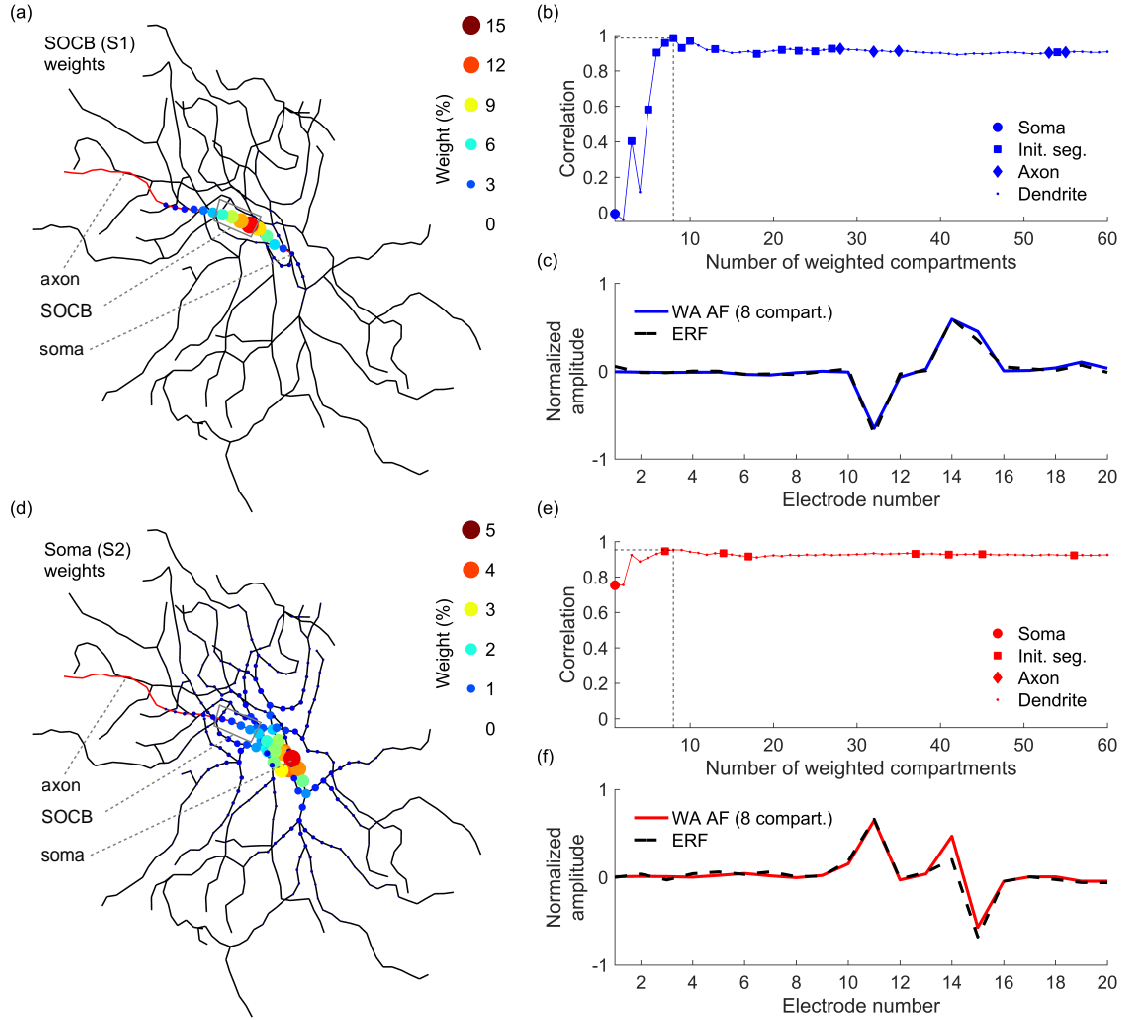


Figure 3.11: Biophysical representation of multiple ERFs in a single cell. Biophysically-determined compartment weights are determined for each of the SOCB (a) and the soma (d) from passive membrane depolarization at each location in response to current injection in each compartment. (b, e) The activating vectors at each compartment are progressively weighted together using the weights depicted in (a) and (d), and correlation with the ERF at the appropriate location (SOCB or soma) is calculated. The type of compartment (soma/initial segment/axon/dendrite) added to the weighted sum activating vector at each step is indicated by marker shape. Note that the initial segment consists of the first $130\ \mu\text{m}$ of the axon originating from the soma and includes the SOCB. (c, f) The best weighted sum activating vector (solid line) for each spike initiation location (SOCB - blue, soma - red) is compared to the ERF at that location (dashed line).

3.5.1 Comparison to previous studies

Jepson et al. (2014) showed empirically that the neural responses recorded from RGCs could be well described by using linear functions of electrode currents, with a subset requiring piecewise linear descriptions. Similarly, Maturana et al. (2016) found, using the same linear-nonlinear model used here, that most cells were well approximated using a one-dimensional linear projection of stimulus current amplitudes, with a subset requiring a two-dimensional projection.

Our results demonstrate that the linear relationship between RGC activation and the currents of multiple electrodes is due to the ERF being determined largely by passive integration of stimulus-induced transmembrane currents at a single site of action potential initiation, typically the SOCB, as shown in Figures 3.5 and 3.8. In addition, we showed that the existence of additional sites of initiation can result in the cell response being driven by multiple distinct linear projections of electrode currents. These linear projections were shown to correspond to the ERFs associated with each initiation site, as shown in Figures 3.9 and 3.11. In Figure 3.9(a), the boundaries between spiking and non-spiking data indicate a map of the cell's stimulus threshold with respect to the first two principal components. In agreement with Jepson et al. (2014), this threshold is best described by a piecewise linear function, shown here to be caused by the interaction of multiple initiation sites.

3.5.2 Biophysical reproduction of the ERF

Figures 3.5(a) and 3.9(c) show the results of optimizing a weighted sum of the activating vectors at the SOCB and the soma to maximize correlation with the ERF. This reveals that there exists a weighted combination of these activating vectors that more closely resembles the cell's ERF than either component in isolation. Although this may seem compelling based on the size of the correlation between the two vectors, there is no guarantee of causality between the activating function and the ERF. This ambiguity is largely overcome by removing any reliance on brute force optimization of the weight, and instead inferring weights by probing the passive properties of the cell, as shown in Figures 3.8 and 3.11. These biophysical weights are what is expected to have a causal influence on spiking. By determining the weights used to combine activating vectors biophysically, the possibility of finding a high correlation by chance is reduced.

For both optimized weights and biophysically determined weights, it is important to make clear the distinction between the weight of a particular compartment and its influence on the ERF. In Figure 3.8, we show that the weight of each compartment can be determined purely from the morphology and passive properties of the cell, and hence is independent of the location of the cell relative to the electrodes. The weight is a factor that is used to scale the transmembrane currents flowing into the compartment to determine its influence on the cell. It is the transmembrane currents that are dependent on the location and orientation of the cell. In theory, even if the weight associated with a compartment is large, its *influence* may be small if the activating function at the compartment is small enough.

Figure 3.8 most clearly demonstrates the biophysical basis for the ERF of the linear-nonlinear model. As can be seen in Figure 3.8(d), the ERF of a RGC with active membrane dynamics can be well reproduced by a fully linear description of the cell and the retinal tissue, except for a spiking threshold applied at the site of action potential initiation. In this simplified model, the only signal

that determines spiking activity is the membrane potential at the site of action potential initiation. Furthermore, as exemplified by Figures 3.4 and 3.7, this membrane potential is fully determined by a linear combination of transmembrane currents (activating function inputs) at compartments near that site, which are in turn a linear combination of stimulus currents. This represents a reproduction of the linear-nonlinear model in biophysical terms, where the linear filter (the ERF) is a linear combination of activating vectors and the nonlinearity is a spiking threshold applied at the spike initiation site.

The observed discrepancies between weighted combinations of the activating vector and the ERF, as shown in Figures 3.5(d) and 3.8, are likely due to differences between passive and active descriptions of the neural membrane. Since the activating function at each location in the cell is a linear function of the currents originating from each electrode, a linear combination of these activating function terms is also a linear function. Although only small errors exist in the linear (passive) approximation of the ERF by the activating function, in the complete active system, the ERF will be influenced to some degree by active membrane dynamics at locations other than the sites of action potential initiation. For example, due to the increasing influence of active dynamics as membrane potential approaches threshold, locations that remain below threshold but are driven close to threshold will likely have a greater influence on spiking in an active model than a passive model.

This effect is also expected to be responsible for the differences in the ERF seen when the cell is modeled with a leaky integrate-and-fire (LIF) model compared to an active model, as shown in Figure 3.6(c). Small differences in the relative size of currents flowing into the SOCB from each compartment means that the ERF is comprised of a slightly different weighted combination of activating vectors. Another likely reason for the discrepancies between actively- and passively-determined ERFs is that the LIF model only captures spikes originating at the set of assumed sites of action potential initiation. The loss of the second significant principal component in the passive model explored in Figure 3.6(a) suggests that the second significant principal component in the active model is introduced by a small subset of spikes initiated at a second location.

As can be seen from the curves of correlation versus number of weighted compartments shown in Figure 3.8(c) and Figures 3.11(a) and (b), the ERF is best reproduced by a weighted combination of a small number of compartments (< 20). Beyond this optimal number of compartments, the correlation with the ERF decreases as more compartments are added to the weighted sum. The quality of the reproduction ($r > 0.95$) suggests that the relative size of the weights for the handful of most important compartments is accurate. However, the fact that the correlation decreases with more compartments suggests that the weights associated with the remaining, less important compartments are too large. The overweighting of these compartments, which have little effect on cell response in reality, is again likely to have been caused by using a passive neuron description to approximate an active system. As discussed above, active dynamics will tend to accentuate larger depolarizations/weights, reducing the number of important compartments when compared to passive models. This is evidenced by the fact that the ERF calculated from a cell with leaky integrate-and-fire dynamics is almost perfectly reconstructed by the full combination of compartments (correlation of 0.993, Figure 3.8(d) bottom).

3.5.3 Limitations

Network-mediated activation. The focus of this study was direct, single-cell activation. As a result, no consideration was given to network effects in the retina and synaptic inputs. Although

direct activation of RGCs is important, it is expected that excitation of the full retinal network will complicate the results presented here. For example, activation of upstream cells in the retinal network (e.g., bipolar cells) may result in more complicated, nonlinear responses from RGCs, particularly for subretinal prostheses. Identifying sites of activation across the network, in addition to within a single cell, could offer an extension to the approach used here, associating distinct ERFs with bipolar-initiated activation and direct RGC activation.

Subretinal versus epiretinal stimulation. The analysis presented here and the experiments of Maturana et al. (2016) deal only with subretinal stimulation. Epiretinal stimulation could result more frequently in multiple sites of action potential initiation and hence multiple ERFs. Specifically, due to the proximity to epiretinal electrode arrays of overlying RGC nerve fiber bundles in the nerve fiber layer, the probability of activation of the axon of a distal neuron is higher under epiretinal stimulation (Grosberg et al., 2017; Rattay and Resatz, 2004; Werginz et al., 2014; Esler et al., 2018a). Since membrane potential is a function of distance along the axon, a spike will result if the threshold is exceeded at any point. As a second site of action potential initiation, this would manifest as a secondary ERF for an affected cell.

This phenomenon was not considered in this work due to the focus on subretinal stimulation, under which activation of passing distal axons is less likely to occur. In future work, the accuracy of this model for epiretinal stimulation should also be considered. Since the length of reconstructed axons rarely extends beyond 500 μm , this analysis will require either new morphological reconstructions or an approximation of RGC axon trajectory.

Temporal dependencies. A main focus of this work was to explain the experimental results of Maturana et al. (2016) and, therefore, our analysis of the simulated data followed the same protocols as previously presented. For this reason and for computational simplicity, only spatial relationships in neural activation (spatial ERFs) were considered, with no consideration of spatiotemporal interactions. In reality, membrane potential is a function of time and depends on the temporal filtering properties of the tissue and the neuron being simulated. This can potentially result in complicated linear responses in membrane potential. As described in Section 3.3.5, the system was simplified by assuming that spiking behavior could be inferred using the membrane potential 0.1 ms after stimulus onset (during the first 0.5 ms phase). Although this is a strong assumption, it is validated by the high quality of reproduction of the ERF, as shown in Figure 3.8. Therefore, extension of the method to spatiotemporal interactions is unlikely to result in large improvements. Nevertheless, due to the demonstrated applicability of linear-nonlinear-type models in capturing spatiotemporal responses to light stimuli, it is expected that the method generalizes well to the spatiotemporal domain.

There may also be some level of influence on spiking from previous stimulus applications. In this research, simulation of all stimulus applications were performed on cells at rest, neglecting any influence from previous applications. In the experimental work of Maturana et al. (2016), white noise stimuli were used to minimize the correlation between inputs. This approach is thought to reduce cross-stimuli dependencies (Freeman et al., 2010; Touryan et al., 2002). In the application of models such as the linear-nonlinear model in a clinical context, it may be necessary to take into account the influence of the non-random stimuli being applied over time.

Choice of nonlinearity. Experimentally, the spiking response of a neuron to electrical stimulation can appear to be stochastic and not deterministic. Empirically, the probability of a spike occurring is related to stimulus current by a sigmoidal nonlinearity. Consequently, the linear-nonlinear

model uses the sigmoidal logistic function as its nonlinearity. The full biophysical model has complex dynamic non-linearities arising from its active channel dynamics, but results in the same double-sided sigmoidal non-linearity seen experimentally in the recovered linear-nonlinear model. In contrast, the integrate-and-fire version of the biophysical model used to recover the linear-nonlinear model used a single-valued threshold to represent the spiking nonlinearity, which is equivalent to a logistic function with infinite slope (i.e., a step function). Both biophysical models are deterministic and so do not capture the stochastic aspects of experimentally observed neuron responses. The difference in non-linearity between these two biophysical models arises because integrate-and-fire membrane dynamics cannot capture all possible factors influencing neuronal spiking. These factors, which remain undetermined here, account for stimuli that are suprathreshold for the integrate-and-fire biophysical model, but nonetheless deterministically fail to produce a spike in the full biophysical model, and vice versa. Furthermore, the observation of stochastic behavior in neurons represents a baseline level of error in reproducing biological responses.

For each site of spike initiation, we observed a proportion of spikes for both cathodic-first and anodic-first stimulation, as can be seen in Figure 3.10. For this reason, a double-sided sigmoid is needed to represent the probability of spiking. It is likely that it is the transmembrane current during the net-cathodic phase of the input that determines spiking (as this depolarizes the membrane), whether a stimulus is cathodic-first or anodic-first. This was confirmed from examination of histograms of spiking latency. It may be that for some cells, the likelihood of spiking is equal whether the stimulus is cathodic-first or anodic-first. In these cases, a single-sided sigmoid could be used and instead applied to the absolute value of the stimulus current instead of the amplitude. However, the assumption that RGCs are equally sensitive to cathodic-first and anodic-first stimulation is difficult to justify, and not necessary for this analysis. Using the double-sided form of the nonlinearity preserves its generality, which has been shown to be necessary for some applications of the model, such as the modeling of network-mediated RGC responses (Maturana et al., 2016). Furthermore, it was found by Maturana et al. (2016) that the flexibility of a double-sided sigmoid was required to accurately model direct responses in some cells.

Electrode geometry. In line with the work of Maturana et al. (2016), simulations presented in this paper used large (400 μm diameter) electrodes, with large inter-electrode spacing (1 mm). Although large electrodes are relevant to some existing clinical visual prosthesis electrode arrays, small-electrode, high-density arrays are becoming increasingly common (Grosberg et al., 2017; Sekirnjak et al., 2008; Hottowy et al., 2012; Jepson et al., 2014). However, due to the perpendicular distance between the ganglion cell layer and electrodes in subretinal stimulation (i.e., the majority of the thickness of the retina), there is a significant amount of current spread from each electrode and interactions of the electric fields of adjacent electrodes before current reaches RGCs. Hence, it is possible that, for high-density subretinal arrays, contributions from adjacent electrodes to the ERF (and the activating function) are similar and do not significantly alter the result presented here. More benefit may be gained from high-density arrays, however, if more sophisticated stimulation procedures are used that account for interactions in the electric fields of adjacent electrodes.

In the case of epiretinal stimulation, however, smaller electrodes and higher density arrays may have a clearer impact. Although dense electrode arrays may afford a greater specificity in stimulating specific cells or even locations within cells, the greater proximity to cells and concentration of stimulation current may result in more frequent cases of multiple spike initiation sites. This effect has been recently demonstrated experimentally using a high-density array on cortical cells (Radivojevic et al.,

2016). Despite these differences, similar descriptions of the ERF should apply due to the similarity between the results presented here and those for high-density epiretinal stimulation studies such as Jepson et al. (2014).

Neural dynamics. The neuron membrane model used was based on that used by Jeng et al. (2011), which was adapted from the model originally presented by Sheasby and Fohlmeister (1999). As mentioned in the Methods, the sodium channel conductance at the SOCB was set to be 20 times higher than the rest of the cell. This value was chosen as it is approximately in the middle of the range of values suggested by Jeng et al. (2011). However, it is likely that this value would vary for each cell and represents a limitation of the membrane model used here. Although this and similar models have been validated in several previous studies (Schiefer and Grill, 2006; Greenberg et al., 1999; Fohlmeister and Miller, 1997), there are likely to be additional parameters that influence RGC activation, such as additional ion channels (Kameneva et al., 2011) or temperature (Fohlmeister et al., 2010). Despite this, the ability of the chosen model to reproduce the experimentally-observed behavior of the linear-nonlinear model, as shown in Figure 3.3, demonstrates that the chosen membrane model is likely to be sufficient. Furthermore, due to the high level of agreement between the active membrane model and the leaky integrate-and-fire model, as shown in Figure 3.6, it is unlikely that a model with a more complex set of active membrane channels would yield substantially better results. However, one observation made during the development of the simulations for this paper was that the shape of a cell's nonlinearity could be altered by varying the leak conductance of the membrane. Specifically, it was seen that for small leak conductances, only a one-sided nonlinearity was observed, meaning that the cell was only spiking due to biphasic pulses that were of a specific polarity (either anodic-first or cathodic-first) and not both. This is likely because, for low leak conductances, the resistive properties of the cell membrane are more dominant, causing the membrane voltage to switch polarity more slowly between stimulus phases. For this reason, maximum depolarization/hyperpolarization is generally larger during the first stimulus phase compared to the second. Hence, with lower leak conductances, spiking is only seen for stimuli that result in a net-cathodic-first input into the cell's site of activation.

In all simulations presented here, inputs into neurons were calculated using the activating function, which is related to the second derivative of the extracellular potential longitudinally along each neurite. Hence, this approach captures the longitudinal mode of neural stimulation: stimulation caused by the flow of current axially along the neurite. This description of neural stimulation is in line with the cable equation for active neurons, as shown in Equation 3.8. A transverse mode of activation also exists, caused by the flow of current orthogonally across the neurite, depolarizing one side and hyperpolarizing the other. The transverse mode is normally neglected, as was the case in the present model, being generally considered to have a much smaller influence than the longitudinal mode. However, in Tahayori et al. (2014) an analysis was completed showing the conditions under which each mode dominates, showing that, for the conditions considered here (stimulus pulse duration of 500 μ s and electrode-neuron separation of 360 μ m), the transverse mode may have a non-negligible influence. Future work should be completed to incorporate the transverse mode of activation and assess its influence on spiking behavior.

3.5.4 Improved stimulation strategies

The equivalence of the activating function and the electrical receptive field may be used to develop methods for improved stimulation of retinal tissue. The activating function is determined completely by the electric field generated in the volume in which a cell resides and the trajectory and diameter of the cell's neurites. By developing a fully biophysical description of the linear-nonlinear model that depends only on known anatomical and physiological features of retinal tissue and RGCs, it will be possible to estimate the distribution of ERFs over populations of RGCs and areas of the retina.

If the distribution of neurite orientations and sizes can be determined (or assumed), a model may be developed for the distribution of activating vectors in a particular region of the retina. This may then be used to obtain a mean-field linear-nonlinear approximation of neural response to multi-electrode stimulation across the retina and for different cell sub-types. Due to the simplicity of this type of model, it may be feasible to perform numerical inversion, giving a set of electrode currents that optimally achieve a desired RGC activation pattern. This approach may allow for increases in resolution and perceptual outcomes given the same stimulation hardware.

3.6 Conclusion

The simple linear-nonlinear model of electrically-elicited RGC responses presented by Maturana et al. (2016) has been recovered from a multi-layered biophysical description of the retinal tissue and a morphologically-accurate description of RGCs. This process validates the assumptions made by the linear-nonlinear model and offers an explanation for those cases which deviate from the model. The linear electrical receptive field of the linear-nonlinear model is determined by the stimulus induced transmembrane current, or activating function, at key locations in the cell: the high-density sodium channel band and the soma. Here, we show that a RGC can be accurately represented by (1) a purely linear approximation of multi-electrode stimulation and the integration of inputs in the cell followed by (2) a fixed spiking threshold at one or a small number of locations. Furthermore, these two phases correspond directly to the linear projection and spiking nonlinearity of the linear-nonlinear model, respectively. The results demonstrate that a linear ERF can be accurately recovered for each identified site of action potential initiation, even for cells for which a single linear description is insufficient. Furthermore, for each site of action potential initiation, the ERF can be almost perfectly reconstructed using a weighted sum of activating vectors from a small set (< 20) of nearby compartments.

Chapter 4

Spatial shaping of retinal ganglion cell activation

4.1 Introduction

Established approaches to electrical stimulation of the retina with multi-electrode arrays use sequential, single-electrode stimulation or disregard electrical interactions to provide a coarse level of perception for the user of a retinal prosthesis. This approach avoids interactions between the electric fields generated by different electrodes, which are often considered to result in unpredictable results (Horsager et al., 2010; Wilke et al., 2011). However, approaches such as these also limit the number and complexity of achievable percepts to those that result from single-electrode stimulation. With the development of phenomenological and biophysical models of patterns of neural activity generated by simultaneous multi-electrode stimulation (Maturana et al., 2016; Halupka et al., 2017; Jepson et al., 2014), it is possible to take advantage of electric field interactions between different electrodes to gain access to a greater ensemble of potential activation patterns. Approaches that effectively utilize simultaneous multi-electrode stimuli are expected to enable greater perceptual resolution and more natural patterns of neural activation using hardware that is available today.

As described in previous chapters, there is substantial benefit in developing retinal stimulation strategies that elicit more natural and accurate activation of retinal tissue via simultaneous multi-electrode stimulation. One approach to achieving more efficacious electrical stimulation is to use model inversion (Dmochowski et al., 2011; Abouelseoud et al., 2018; Maturana et al., 2016; Halupka, 2017), where a model is developed that predicts retinal activity in response to multi-electrode stimulation and that facilitates either closed-form inversion or rapid numerical inversion. The inverted model will yield the set of electrical stimuli that best achieve a desired pattern of activation in the retina. For clinical applicability, this model inversion should be performed in real-time and should be subject to a set of clinically-relevant constraints such as maximum allowed power and current density to ensure the optimized stimulation is both feasible and safe.

Several examples exist in the literature of using forward models of neural activity to inform improved multi-electrode stimuli. Guo et al. (2018) recently demonstrated that increased levels of

preferential activation of OFF/ON retinal ganglion cells (RGCs) could be achieved using a model-based closed-loop feedback system during *in vitro* stimulation. Retinal responses were simulated and compared to desired activity. Comparisons were passed through a feedback loop and used to update stimulus parameters until the stimuli was deemed optimal.

Current steering approaches are also generally informed by either detailed forward models of neural stimulation or empirical observations (Opie et al., 2013; Matteucci et al., 2013; Spencer, 2018; Dumm et al., 2014; Spencer et al., 2016; Jepson et al., 2013). Although these approaches are informed by models of neural activation in response to simultaneous multi-electrode stimulation, they have only been demonstrated with simple activation targets and small numbers of simultaneously driven electrode. There exists no method to date for the direct optimization of stimulus currents for neural activation through model inversion. Approaches such as this are necessary for the reproduction of complex target activation patterns with stimulation from many electrodes.

As explored in detail in Chapter 3 (Esler et al., 2018b), the electrical stimulation of neural tissue can be decomposed into an approximately linear component describing the induced electric field or current density and a nonlinear component describing the resultant neural activation. Due to the linearity of the first component, the electric field or current density resulting from multiple electrode stimulation can be determined using linear superposition of the fields generated by each electrode. Due to this property, previous attempts at optimization of multi-electrode stimuli have endeavored to control the induced electric field as opposed to the nonlinear neural activity. This enables well-established linear algebra techniques (e.g., least squares) to be used to select stimulus currents (Dmochowski et al., 2011). Furthermore, these techniques can be constrained by clinical safety requirements on electrode currents (e.g., constrained least squares) (Dmochowski et al., 2011). However, there exists a highly nonlinear relationship between the linear electric field and nonlinear neural activation under electrical stimulation (for both single and multiple electrodes). Consequently, methods for controlling only the electric field fail to accurately reproduce desired patterns of neural activation.

An attempt has been made to improve upon simple linear model inversions and optimizations by incorporating a penalty or constraint that is informed by neural activation. This approach uses mixed integer linear programming (Abouelseoud et al., 2018) to control the extracellular electric field or current density while constraining solutions based on clinical safety requirements and binary on/off estimates of neural activity, although it does not optimize neural activity directly. By focusing on electric fields or current density, neural activation is not strictly optimized, and the use of a binary simplification of nonlinear neural dynamics fails to capture the probabilistic relationship between neural inputs and spiking behavior (Jepson et al., 2014; Escola et al., 2011). It is also challenging to use binary models of spiking to map desired neural activity patterns to continuously varying target activity patterns (i.e., images) and perception. In addition, the mixed integer linear programming approach was shown to be substantially slower (minutes) than necessary for online image processing (sub-second), with the authors suggesting its suitability for session-based therapies as opposed to visual prostheses (Abouelseoud et al., 2018). Finally, no electrode stimulus optimization regime has yet been successfully applied to complex, high-resolution activity patterns and biphasic stimulation waveforms.

The linear-nonlinear model explored in detail in Chapter 3 is a candidate for spatial shaping through model inversion because of its parsimony and accuracy. In this chapter, the linear-nonlinear model for biphasic stimulation is inverted to generate optimal sets of stimulus currents given desired

patterns of activation. Maturana et al. (2016) demonstrated the experimental recovery of the electrical receptive field (ERF) from white noise stimulation of the retina and electrical recording of RGC responses. This yielded the statistical model of the ERF and the linear-nonlinear model, where the ERF is an estimate of the contribution of each electrode on the signal (e.g., membrane potential) that determines cell spiking. In the application of this model to spatial shaping, RGC activity would need to be recorded at several locations in the retina, enabling the estimation of the ERF at each location. The linear-nonlinear model can then provide an estimate of the spiking activity at each location. Given a desired pattern of spiking activity at the retinal locations, inversion of the linear-nonlinear model will yield optimal multi-electrode stimulus currents. In this way, spatial shaping may be developed using experimentally estimated ERFs at a range of locations in the retina.

A similar approach may also be applied for ERFs estimated using biophysical models. The outcome of Chapter 3 was that the ERF could also be recovered using a linear biophysical model of subthreshold (passive) membrane potential. The ERF of a cell can be represented as the vector of simulated passive membrane potentials induced by each electrode at the location of action potential initiation at a particular time. Hence, an alternative to using the experimental, statistical approach is to use a computational model of passive membrane potential in response to electrical stimulation. Such a computational model was introduced in Chapter 2 of this thesis; however, any suitable linear model of membrane potential may be used. This approach for estimating ERFs is used in this chapter and has the following advantages when compared to experimentally-determined ERFs:

- much higher resolution of ERF estimation can be achieved,
- the need to record neural responses, potentially at high spatial resolution, places challenging constraints on device manufacture,
- results are biophysical and deterministic and so do not require the application of large numbers of random stimuli,
- simulations can be completed offline and, potentially, very rapidly, requiring less patient and practitioner time,
- ERFs, and hence activation, can be determined in multiple relevant layers in the retina, such as the ganglion cell layer (GCL) and the nerve fiber layer (NFL).

In addition, using a biophysical version of the linear-nonlinear model enables the rapid testing of various electrode configurations, retinal anatomies, and model assumptions. The shortcomings of a biophysical approach lie in the assumptions required to develop the computational model of the ERF and the difficulty in tailoring models to specific patients and device placements. For example, a bidomain or mean-field model of passive membrane potential must assume either constant or fixed distributions for morphological parameters, such as the diameters, curvatures, and orientations of axons. In the retina, morphological and physiological parameters may vary significantly from neuron to neuron and may cause substantial deviation from the biophysically-estimated ERF. Consequently, biophysically-estimated ERFs are better interpreted as the *average* ERF in a region of the retina. Due to the relative strengths of experimental and modeling approaches, the methods used in this study are developed to be applicable to both.

In this chapter, we first describe a form of the linear-nonlinear model for modeling the activation of a population of retinal ganglion cells, using either experimentally or computationally estimated ERFs for each cell in the population. A complication in applying models of spiking activity to the

problem of spatial shaping is that desired spatial patterns of activity are likely to be specified in two dimensions (such as images recorded from a camera), but a higher-dimensional representation of the retina may be desired (such as multiple two-dimensional layers or multiple distinct cell types within the same layer). In particular, in this chapter, we model neurites in the ganglion cell layer using a series of overlaid representations with different characteristic orientations (as in Chapter 2), while mapping activity to a single two-dimensional target image.

To overcome this complication, the linear-nonlinear model is generalized to enable the specification of one or more two-dimensional target activity patterns while modeling a higher-dimensional representation of the retina and its ERFs. It is shown that the linear-nonlinear model can be represented as a particular type of convolutional neural network (CNN) (LeCun and Bengio, 1995; LeCun et al., 2015), where the network weights are equal to the optimal set of stimulus currents. Methods for efficient spatial shaping of RGC responses are demonstrated via the constrained iterative inversion of the neural network model to yield optimized stimulus currents for arbitrary target activation patterns. The inversion problem is shown to be equivalent to training the weights of the neural network. The conditions under which the inversion is globally optimal are explored.

The quality of the achieved activation using the inversion strategy is tested with a wide variety of target activity patterns ranging from single phosphenes to complex real-world images. A range of stimulation conditions are also varied to test their importance. These include electrode array location (epiretinal or subretinal), electrode arrangement (hexagonal or grid), stimulation form (monophasic or biphasic), and animal model (human or rodent).

4.2 Methods

4.2.1 The linear-nonlinear model

The linear-nonlinear model of direct RGC activation under biphasic stimulation for a single cell (Maturana et al., 2016) is given by

$$\mathcal{N}(\vec{v}_1 \cdot \vec{S}_t) = \frac{1}{1 + \exp(-(\vec{v}_1 \cdot \vec{S}_t - c^+))} + \frac{1}{1 + \exp(-(-\vec{v}_1 \cdot \vec{S}_t - c^-))}, \quad (4.1)$$

where \vec{v}_1 is the electrical receptive field (ERF), \vec{S}_t is the vector of first-phase electrode amplitudes of the stimulus delivered at time t , and c^+ and c^- are the positive and negative thresholds for the spiking nonlinearity, respectively. Note that the sub-scripted 1 in \vec{v}_1 is for consistency with the linear-nonlinear model presented in Chapters 1 and 3, where it represented that the ERF was recovered as the first eigenvector of a principal component analysis.

This model captures spiking behavior resulting from both net cathodic-first and net anodic-first stimulation, where the *net* polarity of stimulation at a particular location in the tissue is determined by the superposition of the induced electric fields from each electrode. To simulate the response to monophasic stimulation, it is likely that the second term on the right-hand-side of Equation 4.1 may be safely removed without loss of accuracy. The ERF, \vec{v}_1 , approximates the cell's response to stimulation with each electrode. For N electrodes, the ERF for a single cell is a $1 \times N$ vector.

Using the experimental approach described by Maturana et al. (2016) to estimate ERFs at multiple locations in the retina (using multiple recording electrodes) yields a series of ERF vectors. These vectors can be combined into a matrix representation, where each row is a recovered ERF,

$$\mathbf{V} = \begin{bmatrix} \vec{v}_{1,1} \\ \vec{v}_{1,2} \\ \vdots \\ \vec{v}_{1,M} \end{bmatrix}. \quad (4.2)$$

Similar to measuring ERFs using the experimental approach, a biophysical model can be used to simulate the ERF as passive membrane potential at arbitrarily many locations in the retina, as demonstrated in Chapter 3, yielding a matrix structure,

$$\mathbf{V} = \begin{bmatrix} \vec{u}_{m,1} \\ \vec{u}_{m,2} \\ \vdots \\ \vec{u}_{m,M} \end{bmatrix}, \quad (4.3)$$

where each simulated ERF, $\vec{u}_{m,i}$, is a $1 \times N$ vector of membrane potential depolarizations at location i in response to stimulation from each of N electrodes with a nominal stimulation amplitude. In this case, the number of locations at which the ERF is determined, M , may be very large. For example, the model may calculate ERFs over a $10 \mu\text{m} \times 10 \mu\text{m}$ sampling of a plane in the retina. Furthermore, there may be a distinction between different cell types or axonal orientations within that plane, yielding multiple overlaid two-dimensional sets of ERFs.

For parsimonious models of spiking response to be of most use, we need a way to invert them; i.e., to find sets of optimal stimulus currents given desired patterns of RGC activation. Using either of the matrix representations of the ERF described above, the multi-cell/population form of the linear-nonlinear model is

Monophasic:

$$Y' = \mathcal{N}(\mathbf{V}\vec{S}_t) = \frac{1}{1 + \exp(-(\mathbf{V}\vec{S}_t - c))} \quad (4.4a)$$

Biphasic:

$$Y' = \mathcal{N}(\mathbf{V}\vec{S}_t) = \frac{1}{1 + \exp(-(\mathbf{V}\vec{S}_t - c^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}\vec{S}_t - c^-))}, \quad (4.4b)$$

The challenge of spatial shaping is to find the set of electrode stimuli, \vec{S}_t , that minimize the error between a desired pattern of activation, Y , and the activation achieved, $Y' = \mathcal{N}(\mathbf{V}\vec{S}_t)$.

4.2.2 The linear-nonlinear model for spatial shaping

A feature of the model defined by Equation 4.4 is that the matrices defining the target activity, the achieved activity, and the ERFs all have the same number of observations or sampled locations, M . The problem of spatial shaping may be complicated further when attempting to map multiple distinct representations (e.g., cell types, axonal orientations, or different retinal depths) to a single target activity (e.g., a grayscale image). It is likely that we will desire the achieved neural activity to resemble a two-dimensional target image, such as that recorded from a camera; however, the model of the ERF and spiking activity may include a three-dimensional volume in the retina or distinct descriptions for multiple cell types.

When attempting to map multiple, distinct, two-dimensional sets of ERFs, such as multiple axonal orientations or multiple retinal depths, to a single two-dimensional achieved activity, Equation 4.4 no longer holds (Y' and \mathbf{V} have incompatible dimensionality). As a concrete example, if ERFs can be estimated in a three-dimensional volume of the retina for axons with several different characteristic orientations, our input matrix, \mathbf{V} , is four-dimensional: three spatial dimensions and one dimension for overlaid cell representations. In contrast, the desired target activity, Y , is a two-dimensional grayscale image of a visual scene. In this case, the desired activity at each two-dimensional location can be approximated by the average of the multiple representations of the ERF at that location; i.e., the average two-dimensional activation across cell populations, orientations, and retinal depths. The model to be inverted then becomes

Monophasic:

$$Y' = \frac{1}{R} \sum_r \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r))}, \quad (4.5a)$$

Biphasic:

$$Y' = \frac{1}{R} \sum_r \left(\frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - c_r^-))} \right), \quad (4.5b)$$

where Y' is the two-dimensional achieved activity, R is the number of distinct two-dimensional representations, and $\mathcal{N}_r(x)$ is the nonlinearity applied to the r th two-dimensional representation, with spiking threshold, c_r . This enables, for example, different cells in different retinal layers to have different membrane thresholds (as was the case for axon initial segments and overlying axons in Chapter 2). \mathbf{V}_r is the set of M ERFs in the r th two-dimensional representation. Hence, each of the summands in the right-hand-side of Equation 4.5 has the same dimension as the output, Y' .

Furthermore, it will be advantageous to map multiple distinct representations of the retina to several, but fewer, target activity patterns. An example of this, which is explored in this chapter, is the specification of separate target activity patterns in the NFL and the GCL. As introduced in Chapter 1 and explored in detail in Chapter 2, epiretinal stimulation is challenging due to the overlying layer of passing axons in the NFL that may be preferentially stimulated. The results of Chapter 2 indicated that activation of these passing fibers could be avoided using carefully chosen simultaneous multi-electrode stimulation. To address challenges such as this, when different patterns of activation are

desired in different retinal layers or cell types, a spatial shaping solution should allow the specification of multiple two-dimensional target activities.

For multiple two-dimensional target activities, such as activation in the GCL and NFL, Equation 4.5 may be further extended to

Monophasic:

$$Y'_1 = \frac{1}{R_1} \sum_{r \in \mathcal{S}_1} \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r))}, \quad (4.6a)$$

$$Y'_2 = \frac{1}{R_2} \sum_{r \in \mathcal{S}_2} \frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r))}, \quad (4.6b)$$

\vdots

Biphasic:

$$Y'_1 = \frac{1}{R_1} \sum_{r \in \mathcal{S}_1} \left(\frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - c_r^-))} \right), \quad (4.6c)$$

$$Y'_2 = \frac{1}{R_2} \sum_{r \in \mathcal{S}_2} \left(\frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - c_r^-))} \right), \quad (4.6d)$$

\vdots

where Y'_i is the i th achieved activation pattern, R_i is the number of ERF representations that are mapped to the i th activation pattern, and \mathcal{S}_i is the set of representations that are mapped to the i th activation pattern. For example, if each of the multiple retinal representations, \mathbf{V}_r , are depths in the retina, then \mathcal{S}_1 may correspond to the simulated depths occupying the GCL and \mathcal{S}_2 may correspond to the simulated depths occupying the NFL. The achieved two-dimensional GCL and NFL activities, Y'_1 and Y'_2 , respectively, are the averages of the activities achieved in those three-dimensional retinal layers. Similar approaches may be used to target activation to certain cell subtypes while limiting activation of others. During model inversion, each specified target pattern, Y_i , may be weighted according to its importance. For simplicity, the following discussion of model inversion will focus on a single output activity pattern, Y' , and target activity, Y , as in Equation 4.5, but will later be generalized to the case of multiple output activity patterns.

4.2.3 The linear-nonlinear neural network

Inspection of Equation 4.4a (i.e., for a single two-dimensional retinal representation and a single output image) reveals that the equation is equivalent to a logistic regression model (McCullagh and Nelder,

1989; Semenovich et al., 2010). The general form of a logistic regression model is

$$y' = P(y|\theta) = \frac{1}{1 + e^{-\theta^T x}}, \quad yK \sim \mathcal{B}(K, p), \quad (4.7)$$

where x is a vector of observed regressors, θ is a vector of regression coefficients, y' is the predicted probability/proportion, y is the true value, and \mathcal{B} is the binomial distribution. In this form of logistic regression, the distribution of y is free to follow either a Bernoulli distribution (i.e., $K = 1$ and $y \in \{0, 1\}$) or a Binomial proportion. To translate this into terms consistent with the linear-nonlinear model, y (and y') are activation probabilities or proportions, θ is the electrode current vector, \vec{S}_t , and x is an ERF.

Thus, given a set of target activations, Y , solving this problem for monophasic stimulation is equivalent to fitting a logistic regression model, with Y and \mathbf{V} the dependent and independent variables, respectively. The fitting procedure minimizes some measure of distance between the desired activation, Y , and the achieved activation, Y' . This distance is generally quantified using logarithmic loss, which will be discussed in detail later. Once the model is fit, the model coefficients are the optimized electrode currents, \vec{S}_t .

It is important to note the difference between this application of logistic regression and its use in predictive modeling and inference. In the case of prediction or inference, models are typically trained on a single large set of data using pre-coded binary labels or probabilities. Trained models are then fixed and used to either infer relationships in the training data (inference) or used to predict labels on future data (predictive modeling). In contrast, for its application in optimization of stimulus currents (i.e., model coefficients) for retinal stimulation, the labelled probabilities in the training data change as the target image changes with time. This means that the trained model is not fixed, but continuously retrained using new images. As such, for stimulus optimization, it is the algorithm used for model training (and its accuracy and efficiency) that is of interest rather than the predictive power of the model itself.

Similarly, Equations 4.5 or 4.6, which represent a mapping from a higher-dimensional retinal representation to one or several output images, are equivalent to a particular type of feed-forward convolutional neural network (LeCun and Bengio, 1995; LeCun et al., 2015). This equivalent network is shown in Figure 4.1 and consists of an input layer, a *non-overlapping* convolution layer, an offset layer, a sigmoidal activation layer and an average-pooling layer. Henceforth, this neural network architecture will be referred to as the *linear-nonlinear network*. The convolutional layer corresponds to the scaling of the different ERF representations by a common set of electrode currents (i.e., the same stimuli are experienced by the whole retina) to yield the approximate passive membrane potential. The offset and sigmoidal activation layers correspond to the application of the spiking threshold to the membrane potential and the spiking nonlinearity, respectively. Finally, the average pooling layer combines the multiple representations to determine the average two-dimensional spiking probability. Each of the layers in the neural network can be equated to mathematical operations in Equation 4.5, as shown in the right-hand panel of Figure 4.1. The only network weights that need to be trained are those in the convolutional layer, \vec{S}_t , and, once trained, these are the optimized electrode currents. Note that the convolutional layer does not strictly perform a convolution over the input (as the filter is applied in a non-overlapping manner); however, the layer weights, \vec{S}_t , are fixed across all R representations in the manner of a typical convolutional layer. Furthermore, since Equation 4.4 is a special case of Equation 4.5, in which there is only a single representation, r , the neural network approach can be

applied in all cases considered here.

This application of neural networks is unconventional as training is not performed once using stationary data for prediction purposes, but rather the network is retrained for each target image. Therefore, the efficiency of the chosen training or optimization algorithm is of key importance. The translation of Equation 4.5a into an equivalent neural network is not strictly required for model inversion, but doing so enables ready application of the wealth of neural network optimization algorithms that have become available in recent years. In addition, while the linear-nonlinear model was initially developed to map stimulus currents to spiking probability via the ERF, here we instead structure the model to map a set of ERFs to a pattern of spiking activity via the electrode currents. This is because, for the purposes of stimulus optimization as opposed to model-fitting, it is necessary that the model be parameterized by the stimulus currents (\vec{S}_t) rather than the ERFs (\mathbf{V}), so that it is the stimulus currents that undergo optimization.

In the case of biphasic stimulation, in which stimuli of opposite polarity may result in approximately equal activation, the spiking nonlinearity of the linear-nonlinear model is double sided, as shown in Equations 4.4b and 4.5b. As a result, no unique inverse exists and the model cannot be represented as an equivalent logistic regression model. However, the linear-nonlinear network described above may be modified, by applying a double-sided nonlinearity, to model the response to biphasic stimuli, as shown in Figure 4.1. An additional neural network layer (layer 5 in Figure 4.1) is added to sum the output of each side of the nonlinearity, as in Equation 4.4b. Although optimization of the set of weights (i.e., currents) does not have a unique solution due to the double-sided nonlinearity of Equation 4.5b (see Appendix C.1), backpropagation combined with an iterative optimization algorithm may be used to find locally optimal weights (LeCun et al., 2015). To gauge the effectiveness of this approach, locally optimal solutions found using iterative approaches were compared to solutions obtained by search methods and solutions obtained using evolutionary algorithms, as described below.

4.2.4 Neural network training algorithm

For all models used in this work, the model output, Y' , and the target activation, Y , represent a two-dimensional map of spiking probabilities. Alternatively, these values may represent a proportion of cells that are spiking. As a result, error in the fit of a model to Y is determined using binomial logarithmic loss (McCullagh and Nelder, 1989; Semenovich et al., 2010):

$$\mathcal{L}(\vec{S}_t) = -\frac{1}{M} \sum_m^M (1 - y_m) \log(1 - y'_m) + y_m \log y'_m, \quad (4.8)$$

where M is the number of locations at which the ERF is estimated, y_m is the target activation at location m , and y'_m is the activation achieved with stimulus currents, \vec{S}_t , at location m . Depending on context, this measure is sometimes referred to as negative binomial log-likelihood (Semenovich et al., 2010; McCullagh and Nelder, 1989) or binary cross-entropy (Goodfellow et al., 2016; Chollet, 2015). Equation 4.8 is the objective function used for training models and is minimized to determine an optimal set of electrode currents. The optimization/inversion problem may be expressed as

$$\begin{aligned}
& \underset{\vec{S}_t}{\text{minimize}} && \mathcal{L}(\vec{S}_t; Y) \\
& \text{subject to} && \vec{S}_t \in C,
\end{aligned} \tag{4.9}$$

where C is the set of allowed values for the stimulus currents \vec{S}_t . If the stimulus currents are allowed to take any value, $C = \mathbb{R}$, equivalent to unconstrained optimization. The possibility of constraining the optimization using C is addressed in Section 4.2.7 of the Methods. Hereafter, the terms training, fitting, and inverting will be used interchangeably.

Due to the greater generality of the linear-nonlinear network when compared to the logistic regression form of the linear-nonlinear model, this work focuses on the inversion of the neural network. The linear-nonlinear network was trained using a modification of the mini-batch gradient descent algorithm known as Adam: Adaptive Moment Estimation (Kingma and Ba, 2015). Mini-batch gradient descent is an algorithm that uses a randomly sampled subset of the training data (in our the pixel intensities of an image of a visual scene) for each update of the network's weights, \vec{S}_t . At each step, s , mini-batch gradient descent updates the weights using the general update rule,

$$\vec{S}_{t,s+1} = \vec{S}_{t,s} - \eta \cdot \gamma(\vec{S}_{t,s}; \mathbf{V}_{\text{batch}}; Y_{\text{batch}}), \tag{4.10}$$

where η is the learning rate and the function γ is related to the gradient of the objective function, \mathcal{L} , with respect to \vec{S}_t . $\mathbf{V}_{\text{batch}}$ and Y_{batch} are random subsamples of fixed size (batches) of the independent and dependent variables, respectively. The use of mini-batch updates means that, even for a very high-resolution target image/activity, stimulus currents can converge to near ideal values very rapidly through a sampled representation of the image, which increases in resolution with each mini-batch.

The Adam algorithm supplies definitions for η and γ in Equation 4.10 and utilizes the gradient descent concepts of momentum (Qian, 1999) and adaptive learning (Duchi et al., 2011; Zeiler, 2012; Kingma and Ba, 2015). Adam is among the most robust and popular neural network learning algorithms currently available (Kingma and Ba, 2015; Ruder, 2016; Bengio et al., 2015). Unlike early gradient descent algorithms used for training neural networks, Adam updates the learning rate for each network weight individually during training. In addition, Adam incorporates a momentum-like mechanism in the form of an exponentially-decaying mean of past gradients of the objective function and an exponentially-decaying mean of the square of those gradients. These two means are estimates of the first moment (mean) and second moment (uncentered variance) of the objective function gradient. Momentum mechanisms such as this enable those weights in \vec{S}_t for which the gradient is consistently positive (negative) to gain momentum in the positive (negative) direction. Thus, gradient descent proceeds more rapidly in directions that offer the most benefit (Ruder, 2016; Bengio et al., 2015).

Adam requires the following user-defined parameters:

- α : the step size,
- $\beta_1, \beta_2 \in [0, 1)$: the exponential decay rates for the moment estimates,
- $\mathcal{L}(\vec{S}_t)$: the objective function,
- \vec{S}_t^0 : the initial weight vector.

Given the above parameters, the Adam algorithm is implemented as described in Algorithm 1. The solution is deemed to have converged when the step-on-step improvement in the objective function, \mathcal{L} ,

drops below a specified value. In the above algorithm, the gradient of the optimization function, g , with respect to the network weights is determined using backpropagation.

Algorithm 1: *Adam* gradient descent optimization algorithm (Kingma and Ba, 2015).

```

 $\mu_0 \leftarrow 0$  (Initialize first moment);
 $\nu_0 \leftarrow 0$  (Initialize second moment);
while  $\vec{S}_t$  has not converged do
     $s \leftarrow s + 1$ ;
     $g_s \leftarrow \nabla \mathcal{L}(\vec{S}_t)$  (Get gradients of objective function w.r.t. weights);
     $\mu_s \leftarrow \beta_1 \mu_{s-1} + (1 - \beta_1) g_s$ ;
     $\nu_s \leftarrow \beta_2 \nu_{s-1} + (1 - \beta_2) g_s^2$ ;
     $\hat{\mu}_s \leftarrow \mu_s / (1 - \beta_1^s)$ ;
     $\hat{\nu}_s \leftarrow \nu_s / (1 - \beta_2^s)$ ;
     $\vec{S}_t \leftarrow \vec{S}_t - \alpha \hat{\mu}_s / (\sqrt{\hat{\nu}_s} + \epsilon)$ ;
end

```

The linear-nonlinear neural network was specified and trained in statistical computing software R 3.4.2 (R Core Team, 2017) using third-party R packages keras (Allaire and Chollet, 2017) and tensorflow (Allaire and Tang, 2017), which offer R interfaces to neural network software packages Keras (Chollet, 2015) and Tensorflow (Abadi et al., 2016), respectively. Efficient manipulation of data within R was achieved using the third-party R package, data.table (Dowle and Srinivasan, 2017).

4.2.5 Existence of globally optimal solutions under different conditions

To demonstrate that a fitting procedure or optimization algorithm can find global optima, it is necessary to examine the topology of the objective function, in our case \mathcal{L} , with respect to the model coefficients being fit, in our case \vec{S}_t . Since the Adam training procedure is a gradient descent method, it is necessary to consider the possibility that gradient descent will be caught in local optima and terminate with a sub-optimal solution.

As described previously, when the target activation, Y , is specified using the same number of two-dimensional representations as the input, V , spatial shaping can be achieved via the training of Equation 4.4, represented as a logistic regression model. Both logistic regression and the linear-nonlinear model are fully characterized by a sigmoidal transformation of a linear, multi-variate function. It is an established result that the unconstrained fitting of the exponential family of generalized linear models (such as logistic regression) by maximizing log-likelihood is a convex problem (Wedderburn, 1976; Semenovich et al., 2010). This means that for unconstrained optimization, every local optimum is also a global optimum and the gradient of the objective function is monotonically non-decreasing with respect to each model coefficient. Hence, for the case of monophasic stimulation and compatible input and output representations, any gradient descent algorithm (such as Adam) will obtain an estimate arbitrarily close to the global optima given sufficient descent iterations. This result holds for both logistic regression and neural network implementations of Equation 4.4a as they are equivalent mathematically and minimize the same logarithmic loss function. However, convexity is not generally the case when constraints are applied to the values of the model parameters (i.e., electrode currents) (Semenovich et al., 2010).

In the case of biphasic stimulation, gradient descent is not guaranteed to find the globally optimal set of stimulus currents. This can be seen from inspection of Equation 4.4b. If we assume that $c_r^+ = c_r^-$, the same Y' is achieved with either of \vec{S}_t or $-\vec{S}_t$. Hence, if one local optimum exists at \vec{S}_t , another will exist at $-\vec{S}_t$. Furthermore, since $\mathbf{V}_r \vec{S}_t$ is a vector of length M , each element of which is passed through a double-sided nonlinearity, there are potentially 2^M locally optimal solutions to the inversion problem for a given Y . A mathematical demonstration of this is provided in Appendix C.1.

In the case of a single two-dimensional representation for Y and multiple two-dimensional representations included in V , we have Equation 4.5. As for above, in the case of biphasic stimulation, many locally optimal solutions of Equation 4.9 exist. In the case of monophasic stimulation and the neural network shown in Figure 4.1 (described by Equation 4.5a), Y' is represented by a sum of sigmoid functions. There is no straightforward method for extending established convexity proofs for single sigmoids to this sum. Furthermore, as demonstrated in Section 4.3.2, the log-likelihood function, Equation 4.8, applied to Equation 4.5a can be shown empirically to be non-convex. Despite this, it is still possible that a unique solution exists to the optimization problem described by Equation 4.9.

To demonstrate the existence of a single (multiple) optima for monophasic (biphasic) stimulation, network weights were varied over a grid to construct a map of the objective function with respect to the weights. This analysis is presented in Section 4.3.2.

4.2.6 Network weight initialization

To ensure reproducibility of results and decrease training time, the weights of the linear-nonlinear network were initialized using multiple linear regression. In the case of monophasic stimulation with an equal number of input and output representations, Equation 4.4a was rearranged to give,

$$\mathbf{V} \vec{S}_t = -\log(1/Y' - 1) + c. \quad (4.11)$$

An initialized set of weights, \vec{S}_t , was found by substituting the target, Y , for Y' and solving the multiple linear regression problem,

$$\underset{\vec{S}_t}{\text{minimize}} \quad \|\mathbf{V} \vec{S}_t - (-\log(1/Y - 1) + c)\|. \quad (4.12)$$

In practice, to ensure that $\log(1/Y - 1)$ was always defined, Y was replaced by $Y(1 - 2\epsilon) + \epsilon$, where ϵ is a small positive value (10^{-10} in this work).

In the case of a single output representation and multiple input representations, a direct inversion as in Equation 4.11 does not exist. Instead, an approximate inversion was obtained by averaging \mathbf{V} across representations, before using the same approach as in Equations 4.11 and 4.12. This is equivalent to moving the average-pooling layer shown in Figure 4.1 to be before the convolutional layer. Although this process does not find the optimal solution, it seeds the network weights at a more sensible location than random initialization and significantly reduces the required number of gradient descent training iterations.

As discussed above, for biphasic stimulation there are potentially many locally optimal solutions to the inversion problem. In the absence of a method for choosing one solution over another, the aforementioned initialization approach for monophasic stimulation was also used for biphasic stimulation.

4.2.7 Stimulus current constraints

Inequality constraints

As introduced in Equation 4.9, the optimization of electrode currents may be constrained. This enables the training algorithm to conform to specific power and safety constraints on the amount of current delivered by an electrode array to tissue. Inequality constraints were applied in this work to limit the L2-norm of the stimulus current vector, $\|\vec{S}_t\|_2$, to remain under a fixed value. Although placing other constraints, such as limiting the infinity-norm (i.e., limiting the maximum electrode current or charge magnitude) and the L1-norm (i.e., limiting the sum of electrode current magnitudes) may be more directly applicable to the problem of multi-electrode array stimulation, they are not commonly used as constraints in neural networks. It is left to future work to implement these types of constraints.

Equality constraints

It is also advantageous to be able to apply equality constraints on a subset of network weights (i.e., electrode currents) during model training. An example is the development of methods for focal stimulation (Opie et al., 2013; Matteucci et al., 2013; Spencer, 2018; Dumm et al., 2014; Spencer et al., 2016; Jepson et al., 2014). Current delivered to a single electrode can be fixed at a pre-specified value and the current delivered to surrounding electrodes may be optimized to narrow the achieved region of activation.

To impose equality constraints on a subset of electrodes, we can split the electrode amplitudes \vec{S}_t , and the components of the ERF sets, \mathbf{V}_r , into fixed and trainable components. Equation 4.5b can be modified to

$$Y' = \frac{1}{R} \sum_r \left(\frac{1}{1 + \exp(-(\mathbf{V}_r \vec{S}_t + \mathbf{V}_r^f \vec{S}_t^f - c_r^+))} + \frac{1}{1 + \exp(-(-\mathbf{V}_r \vec{S}_t - \mathbf{V}_r^f \vec{S}_t^f - c_r^-))} \right), \quad (4.13)$$

where \vec{S}_t^f is the set of fixed electrode currents and \mathbf{V}_r^f is the ERF elements corresponding to those electrodes. In order to include this type of constraint in the neural network shown in Figure 4.1, an additional offset layer is added between the convolutional and replication layers which adds in the fixed terms, $\mathbf{V}_r^f \vec{S}_t^f$.

4.2.8 Inversion using genetic algorithms

As mentioned above, there are many locally optimal solutions to the inversion of the linear-nonlinear network for biphasic stimulation. Hence, iterative gradient descent optimization algorithms are not guaranteed to find global optima. To assess the optimality of the stimulus currents obtained using gradient descent, they were compared to the stimulus currents obtained by inverting Equation 4.5b using a genetic algorithm (Addis et al., 2011; Sivanandam and Deepa, 2008).

Genetic algorithms are a universal search algorithm, meaning that they search across the entire weight space for the best solution. They are commonly used to generate high-quality solutions by relying on evolution-inspired processes including mutation, crossover, and selection. Although a genetic algorithm is not guaranteed to find the global optima, it searches through a much wider class of

potential solutions, generally at the expense of efficiency.

Starting with a randomly generated population of weight sets, $\{S_{t,1}, S_{t,1}, \dots, S_{t,n}\}$, a genetic algorithm was employed to *evolve* the population toward an optimal solution. This was achieved by:

1. Calculating the value of the loss function, \mathcal{L} , for each set of weights in the population.
2. Selecting a proportion, p_s , of the population based on the value of the loss function. This subset is *bred* with each other using crossover: contiguous sections of weights (e.g., the first 10 weights) are exchanged between randomly selected pairs of candidate solutions, where this crossover occurs with probability p_c . This process is repeated as necessary to maintain the same population size. A small proportion, p_e , of the very best weight sets from the previous generation are carried forward unchanged.
3. Mutating the new generation by randomly altering a proportion, p_m , of weights in each weight set.
4. Repeating steps 1-3 until a satisfactory solution is found.

The results presented here used the following parameters to control the algorithm: $p_s = 0.5$, $p_c = 0.8$, $p_e = 0.05$, $p_m = 0.1$. These values are the default values for the software implementation used and were not altered as the results were satisfactory and further optimization seemed beyond the scope of this research. All genetic algorithms were run using the third-party R package GA (Scrucca, 2013).

4.2.9 Biophysical model of the ERF

To this point, all methods relating to the linear-nonlinear model, the linear-nonlinear network, and their inversion are not specific to the method of estimation of the ERFs across the retina. All above descriptions may be used to develop an approach to spatial shaping for ERFs determined either experimentally or derived from biophysical simulation. The four-layer biophysical model of the retina presented in Chapter 2 is now used to derive a biophysical approximation of the set of required ERFs, \mathbf{V}_r , for all representations, r .

In Chapter 3, the biophysical basis for the linear-nonlinear model of direct RGC responses to electrical stimulation was determined. The linear component was shown to correspond to a linear approximation of the integration of stimulus-induced transmembrane currents in a cell. That is, the linear component is simply the linear subthreshold membrane dynamics and the nonlinear component is a fixed spiking nonlinearity. In Chapter 2, a model was developed, based on several underlying assumptions, that supplies a linear, closed-form description of subthreshold membrane depolarization resulting from direct electrical stimulation. In this chapter, this model was used to approximate direct RGC activation from an arbitrary set of stimuli across the entire retina at a desired resolution. Chapter 2 also provides and justifies a choice of membrane threshold that was used to define the nonlinearity of the linear-nonlinear model and was required to predict activation from subthreshold dynamics.

From Chapter 2, the passive membrane potential in response to direct electrical stimulation is given by

$$\hat{V}_m(k_x, k_y, z, \omega) = -\frac{k_y^2 \lambda^2(\omega)}{1 + k_y^2 \lambda^2(\omega)} \hat{\phi}_e(k_x, k_y, z, \omega), \quad (4.14)$$

where \hat{V}_m is the membrane potential in the x, y, t -Fourier domain and $\hat{\phi}_e$ is the extracellular voltage in

the x, y, t -Fourier domain resulting from stimulation with N electrodes, where the x and y -directions are parallel to the surface of the retina and z represents depth. k_x , k_y , and ω are the Fourier domain pairs of x , y , and t , respectively. The frequency-dependent length constant, $\lambda(\omega)$, is given by

$$\lambda(\omega) = \frac{\sqrt{r_m}}{\sqrt{r_i(1 + 2\pi a j \omega r_m C_m)}},$$

where r_m and r_i are the membrane and intracellular resistance per unit length, respectively, C_m is the membrane capacitance per unit area, and a is the radius of the simulated neurites. The calculation of $\phi_e(k_x, k_y, z, \omega)$ in Equation 4.14 is described in detail in Chapter 2 and depends on the stimuli delivered to each electrode and the type of retinal stimulation (epiretinal or subretinal) being simulated.

Equation 4.14 was used to determine the Fourier-domain representation of the passive membrane depolarization at a particular depth, z , in the retina in response to stimulation with electrodes. The inverse Fourier transform with respect to k_x , k_y , and ω was then calculated. This yielded the time course of the passive membrane depolarization in an x - y plane. As in Chapter 3, the time course was sampled at $t_{\text{meas}} = 0.1$ ms after stimulus onset to obtain a single depolarization value per location.

Figure 4.2 demonstrates how a set of ERFs, \mathbf{V}_r , were generated using the four-layer biophysical model. Since each element of the ERF is the approximate passive membrane depolarization in response to stimulation with a single electrode, each electrode was used to deliver $1 \mu\text{A}$ of current and the resulting membrane depolarization at all desired locations was calculated using Equation 4.14. In this way, the set of ERFs for a particular representation, r , is:

$$\begin{aligned} \mathbf{V}_r &= \begin{bmatrix} \vec{u}_{m,1} \\ \vec{u}_{m,2} \\ \vdots \\ \vec{u}_{m,M} \end{bmatrix} \\ &= \begin{bmatrix} V_{m,1}(x_1, y_1, z_r, t_{\text{meas}}) & V_{m,2}(x_1, y_1, z_r, t_{\text{meas}}) & \dots \\ V_{m,1}(x_2, y_2, z_r, t_{\text{meas}}) & V_{m,2}(x_2, y_2, z_r, t_{\text{meas}}) & \dots \\ \vdots & \vdots & \vdots \\ V_{m,1}(x_M, y_M, z_r, t_{\text{meas}}) & V_{m,2}(x_M, y_M, z_r, t_{\text{meas}}) & \dots \end{bmatrix}, \end{aligned} \quad (4.15)$$

where $V_{m,e}(x_i, y_i, z_r, t_{\text{meas}})$ is the membrane depolarization at x - y location (x_i, y_i) induced by a $1 \mu\text{A}$ stimulation current delivered by electrode, e . Note that a representation here corresponds to a particular fiber orientation and retinal depth (i.e., the GCL or the NFL).

To assess the degree to which desired stimulation was occurring in the GCL and unwanted stimulation was occurring in the NFL, ERFs were calculated for a two-dimensional plane in each retinal layer. As discussed in Chapter 2, the GCL is best represented by considering fibers with a uniform distribution of orientations, whereas the NFL may be approximated as having a single parallel fiber orientation. To account for this, multiple ERF representations were used to describe the GCL, with each representation describing the set of ERFs for fibers with a specific orientation. One representation was used for the NFL, resulting in a total of $R = R_{\text{GCL}} + 1$ representations.

The neural network architecture used in this research is shown in Figure 4.3. As can be seen, the model consisted to two outputs: one for the GCL activation and one for the NFL activation. To

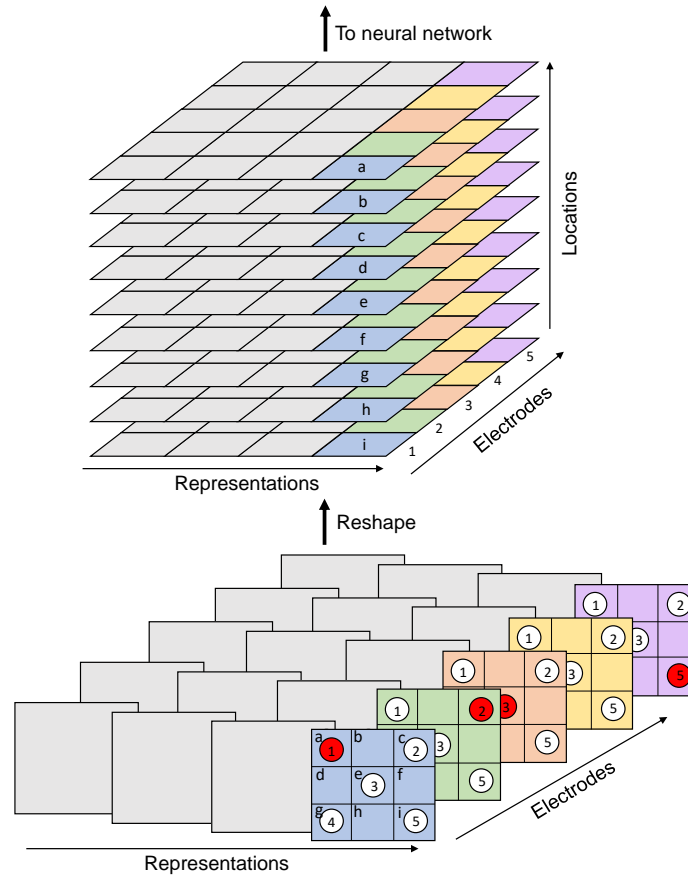


Figure 4.2: **Schematic of transformation of two-dimensional ERF data to a format suitable for input into the neural network of Figure 4.1.** For each orientation in the GCL and for the NFL (i.e., for all representations), simulations were run of stimulation by each electrode, one by one, with a stimulus amplitude of $1 \mu\text{A}$. Each two-dimensional simulation generates a single element in the ERF for each of the M locations. Each two-dimensional grid of passive membrane depolarizations provided by simulations is reshaped into a one-dimensional vector to prepare data for the neural network. Colors indicate the mapping of calculated values. The mapping of individual values is indicated by lettered elements, a-i, for the blue plane. The mapping of data from stimulation with each electrode is indicated by numbers 1-5.

fit this model, two target activations, Y_{GCL} and Y_{NFL} were specified. For all optimizations presented here in which activation in the NFL is considered, the target for the NFL was the zero matrix (i.e., no activation) since it is ideal to minimize activation of the NFL. A range of different targets were used for the GCL target activity.

Finally, since the linear-nonlinear network was being fit to multiple target patterns of activation, a weighting factor, λ , was applied that varied the relative importance of each target in the optimization problem. The objective function in Equation 4.8 is then

$$\begin{aligned} \mathcal{L}(\vec{S}_t) = & -\frac{\lambda}{M} \sum_m^M (1 - y_{\text{GCL},m}) \log(1 - y'_{\text{GCL},m}) + y_{\text{GCL},m} \log y'_{\text{GCL},m} \\ & - \frac{(1 - \lambda)}{M} \sum_m^M (1 - y_{\text{NFL},m}) \log(1 - y'_{\text{NFL},m}) + y_{\text{NFL},m} \log y'_{\text{NFL},m}. \end{aligned} \quad (4.16)$$

The importance weighting, λ , was varied from 0 to 1, where 1 gives importance only to the GCL target and ignores activation of the NFL, whereas 0 gives importance only to the NFL target. For the network weight initialization described above, the same λ weighting was used to solve the norm minimization problem in Equation 4.12.

Along with assessing the quality of inversion, analyses were also conducted to assess the differences in achievable stimulation between epiretinal and subretinal stimulation. This was achieved by moving the electrode layer from the epiretinal to the subretinal space. All biophysical simulations of the ERF were conducted using MATLAB (The Mathworks, Release 2016a).

4.2.10 Analytic parameters

Unless otherwise stated, all simulations and optimizations employed the following parameters:

1. **Electrode location.** Except for comparisons between epiretinal and subretinal stimulation, epiretinal electrode array placement was used. For epiretinal stimulation, the electrode array was located 100 μm away from the inner surface of the retina in the vitreous body. For subretinal stimulation, the electrode array was located on the outer surface of the modeled retinal layers. In both cases, the array is assumed to be flat and parallel to the retinal surface.
2. **Electrode array geometry.** Both rectangularly- and hexagonally-arranged electrode arrays were tested. In both cases, arrays had electrodes with diameters of 100 μm and a center-to-center electrode pitch of 200 μm . For all simulations with rectangularly arranged electrodes, a 9×9 grid of regularly spaced electrodes was used.
3. **Retinal anatomy.** This research focused on human or primate-like retinal geometry, except for simulations comparing the influence of human and rodent anatomies. As such, the thickness of the NFL was generally set to 100 μm (Oishi et al., 2009; Frenkel et al., 2005; Wang et al., 2013; Thomson et al., 2015).
4. **RGC spiking thresholds.** In all results presented here, the sensitivity of the cells was assumed to be equal for net cathodic-first and net anodic-first stimuli: $c_r^+ = c_r^-$. This was shown to be an appropriate assumption by Maturana et al. (2016).

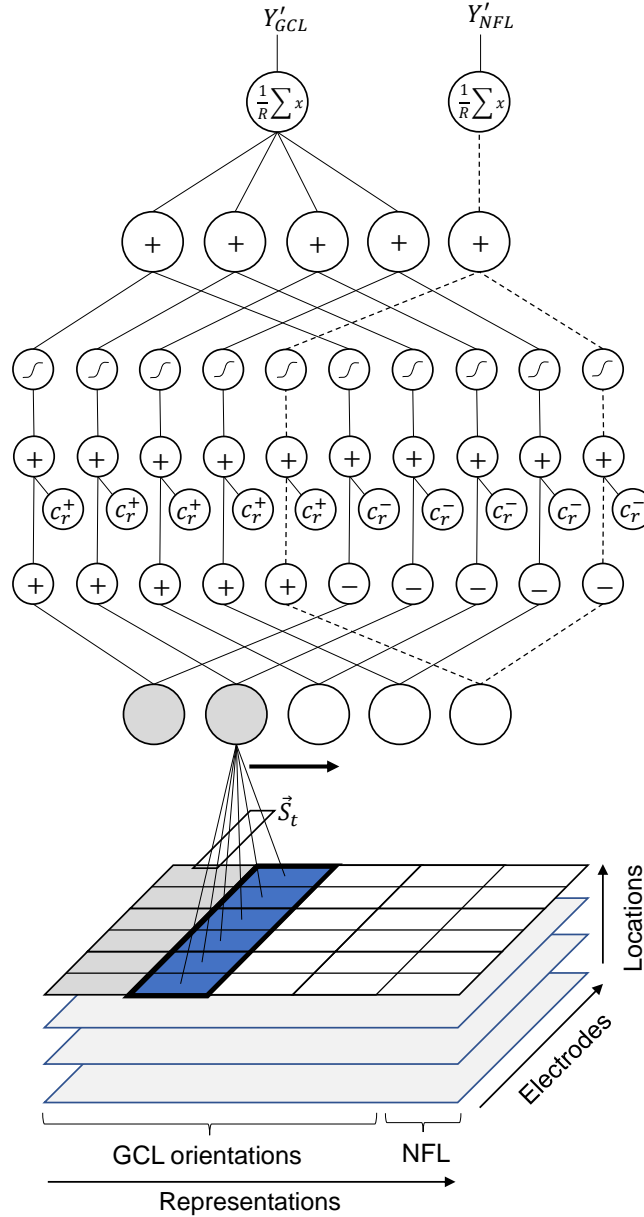


Figure 4.3: **Linear-nonlinear network with GCL and NFL outputs.** As in Figure 4.1, the input dataset has dimensions $R \times M \times N$, where R is the number of two-dimensional representations of the retina, M is the number of locations sampled in each two-dimensional representation, and N is the number of electrodes. Each layer of the network contributes the same mathematical operation as described in Figure 4.1. From Chapter 2, the GCL is best represented by a range of uniformly sampled orientations. Using the terminology introduced in this chapter, each of these orientations is a two-dimensional representation of the GCL. The output of the linear-nonlinear network for the GCL is the average activation across all modeled orientations. Since the NFL is better modeled as having a single orientation of fibers, only one representation is used. The activation achieved for this one representation is the output of the model for the NFL. The passage of the single NFL representation through the network, from input ERFs to output neural activity (Y'_{NFL}), is indicated by dashed lines.

5. **Target locations.** Activation in the GCL was targeted to a plane a short distance into the GCL on the side of the layer closest to the electrode. For epiretinal stimulation, the target plane was $10\ \mu\text{m}$ into the GCL from the NFL. Similarly, the resulting activation in the NFL was calculated at $10\ \mu\text{m}$ from the retinal surface.
6. **Stimulus waveform.** Biphasic stimulation was used for all presented results, except where comparisons between monophasic and biphasic stimulation are presented.

4.3 Results

The results of an analysis of properties of the ERF and the linear-nonlinear network objective surface will first be presented. This will be followed by a series of analyses aimed at demonstrating the ability of the proposed optimization algorithm to reproduce established results in multi-electrode stimulation (e.g., current steering) and in selecting appropriate stimuli given complex, real-world images. Due to the focus of this Chapter (and thesis) on avoiding unwanted activation of overlying axons in the NFL, most of the following results are generated using epiretinal stimulation.

4.3.1 Spatial distribution of ERFs

A primary reason that spatial shaping methods are necessary for effective multi-electrode stimulation is that there is significant interaction of the electric fields and stimulating currents generated by adjacent electrode by the time they reach the target neural tissue. If the distance were such that this interaction could be neglected, spatial shaping would not be necessary. However, due to the push in recent years for higher-density electrode arrays, this is unlikely to be a feasible approach. Furthermore, large separations between electrodes will result in discrete spotted percepts, precluding any recovery of complex, continuously spatially varying visual scenes.

Figure 4.4 demonstrates the resulting passive membrane potential from stimulation with a single electrode. This is equivalent to the elements of the ERF for this electrode (a single column) across all positions (rows) in the matrix \mathbf{V} . In this simulation, the center-to-center electrode pitch was $200\ \mu\text{m}$, and the electrode array was located epiretinally, $100\ \mu\text{m}$ from the surface of the retina. This geometry is used for most of the results presented in the remainder of this Chapter. The spread shown in Figure 4.4 is that observed in the GCL, the region being targeted by stimulation. As can be seen, due to the spread of the electric field from one electrode and the corresponding distribution of the ERF, interactions between adjacent electrodes are significant for the electrode geometry and placement modeled here, necessitating methods for spatial shaping in selecting simultaneous multi-electrode stimulation parameters.

4.3.2 Visualization of the objective function

To validate the analysis presented in the Methods regarding the existence of local and global optima for biphasic and monophasic stimulation, respectively, the surface of the objective function, \mathcal{L} , was mapped, as shown in Figure 4.5. The neural network architecture shown in Figure 4.1 was defined along with a simple target activity pattern which is shown in the inset. For simplicity, this experiment

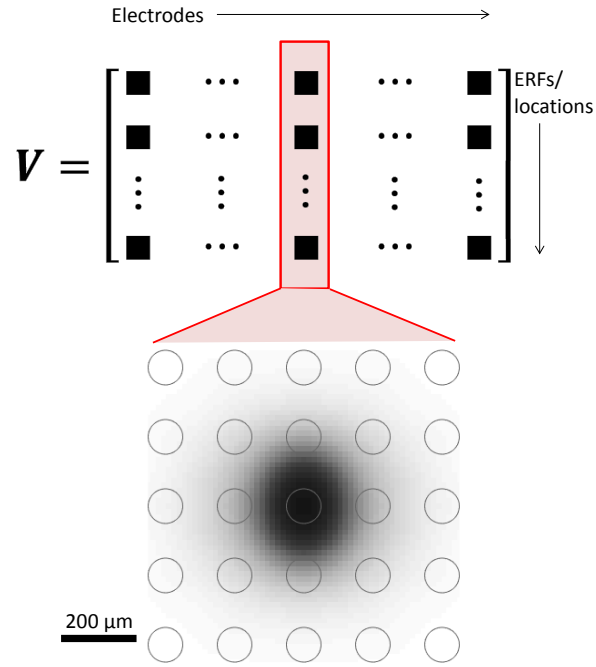


Figure 4.4: **The distribution of the ERF from a single electrode in space.** The matrix V contains the ERF vectors for locations in the retina, where each ERF is represented in a row. Hence, each column represents the element of each ERF contributed by a single electrode. Shown here is the value of one column in V across a plane in the GCL. As described in Chapter 3, this is equivalent to the passive membrane potential induced by stimulation with one electrode. The center-to-center electrode pitch is 200 μm and the electrode array is located epiretinally, 100 μm from the surface of the retina.

was conducted using only two electrodes (so that it might be visualized) and with a single centered phosphene as the activation target. The electrode currents (network weights) assigned to each electrode were varied in a regular, two-dimensional grid with a resolution of $40\ \mu\text{A}$ and the value of the objective function was calculated at each grid point. For this simulation, an epiretinal array with a center-to-center electrode pitch of $200\ \mu\text{m}$ was used. Only the achieved activation pattern in the GCL was considered by setting $\lambda = 1$ in Equation 4.16.

As can be seen in Figure 4.5, for monophasic stimulation, there is a single globally optimal solution to the inversion problem (at $[80\ \mu\text{A}, 80\ \mu\text{A}]$). Hence, a gradient descent algorithm is guaranteed to find the optimal set of stimulus currents given enough optimization iterations. In contrast, there are four distinct local optima for biphasic stimulation, one in each quadrant of the stimulus space. In addition, since the positive and negative thresholds are equal, the loss function is reflected about the origin (i.e., each local optimum has a pair in the opposite quadrant). For biphasic stimulation, different optima may be obtained by gradient descent depending on the starting position. Also, even for monophasic stimulation, the objective function is not convex; the derivative of the objective function is not monotonically non-decreasing with respect to each electrode current.

4.3.3 Optimality of gradient descent solutions

The potentially large number of local optima for the biphasic optimization necessitates an assessment of the level of optimality of solutions obtained using the Adam gradient descent algorithm. This was achieved by comparing the performance of solutions obtained using Adam with those obtained using (1) a genetic algorithm and (2) a wide range of weight initializations followed by gradient descent. For (2), to ensure that a suitable amount of the potential solution space is explored, a gradient descent optimization was performed in each orthant. This was achieved by initializing the electrode currents in each orthant and performing gradient descent from this point while constraining the optimization to remain in the same orthant. To make this approach feasible, optimizations were performed with only 9 electrodes, yielding $2^9 = 512$ orthants (and 512 gradient descent solutions for comparison).

Figure 4.6 shows the results of this analysis for three sample target images. Figure 4.6(a), (c), and (e) are histograms, which demonstrate that, in each case, the loss achieved via Adam gradient descent (thick black lines) is as low or lower than that achieved using either the genetic algorithm (red dashed line) or the orthant search (grey lines) approaches. Figure 4.6(b), (d), and (f) show the trajectory of the loss during the gradient descent procedure.

4.3.4 Current steering and virtual electrodes

Two promising enhancements to retinal stimulation currently being investigated are focal stimulation and virtual electrodes. Virtual electrodes refer to stimuli aimed at eliciting activation at locations between electrodes by recruiting multiple neighboring electrodes. Focal stimulation (or focused multipolar stimulation) aims to recruit multiple neighboring electrodes to narrow the region of activation, thereby increasing the achievable resolution from a retinal prosthesis. Both techniques aim to control the local spread of retinal neural activation under electrical stimulation through simultaneous stimulation with multiple electrodes. As such, the optimization framework described here is ideal for achieving the outcomes intended by both techniques.

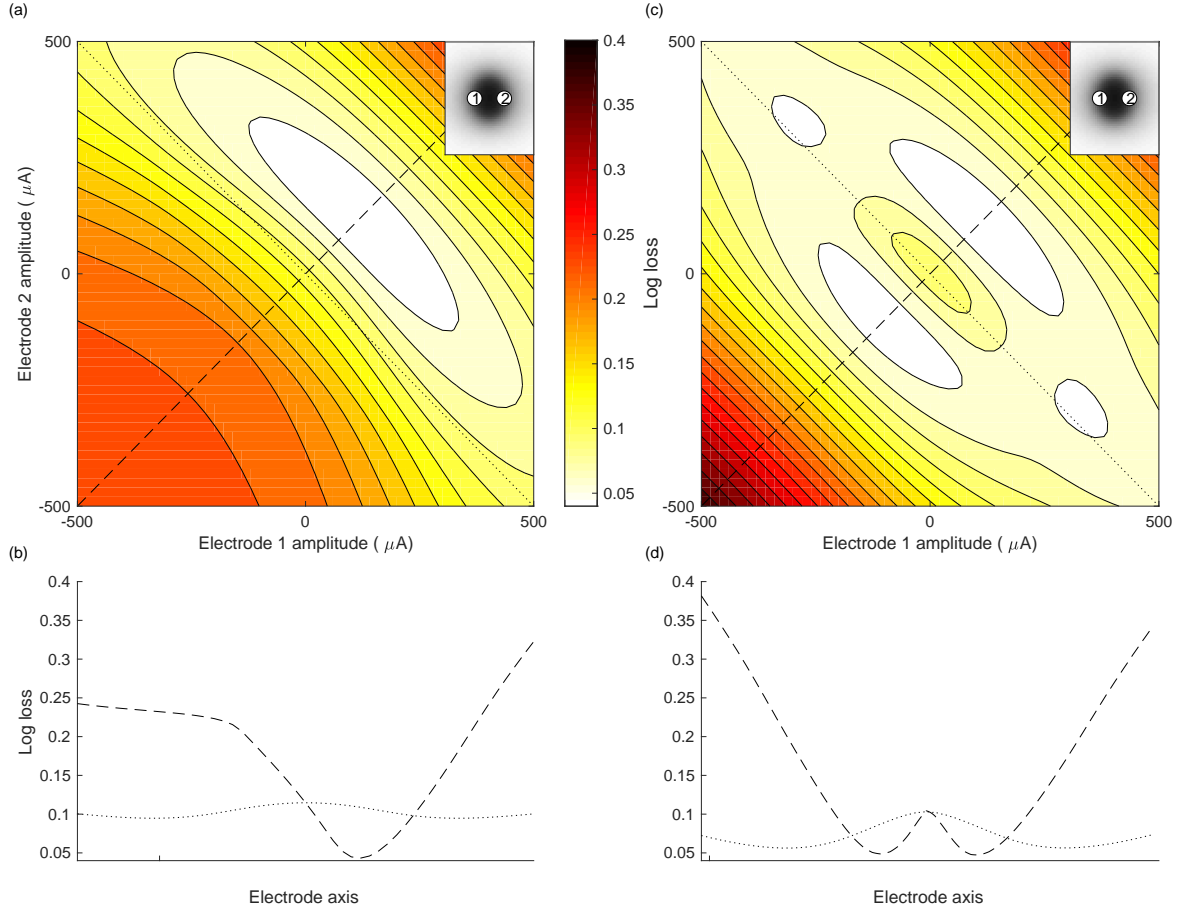


Figure 4.5: **Map of the objective surface.** The objective function, Equation 4.16, is shown with respect to two electrode current amplitudes. As shown in the insets of (a) and (c), this mapping was performed with only two electrodes for the purposes of visualization and the target activation pattern consisted of a single phosphene in the GCL. The properties of the shown optimization surfaces are consistent for more complex targets. (a) For monophasic stimulation, a single global stationary point was found, in agreement with the mathematical result presented in the Methods section 4.2.5. (b) The value of the objective function along the dashed and dotted lines of (a). (c) For biphasic stimulation, a potential 2^M locally optimal solutions exist, of which four can be clearly seen here. (d) The value of the objective function along the dashed and dotted lines of (c). The center-to-center electrode pitch was $200\ \mu\text{m}$ and the electrode array was located epiretinally, $100\ \mu\text{m}$ from the inner retinal surface. In this analysis, $\lambda = 1$ (Equation 4.16).

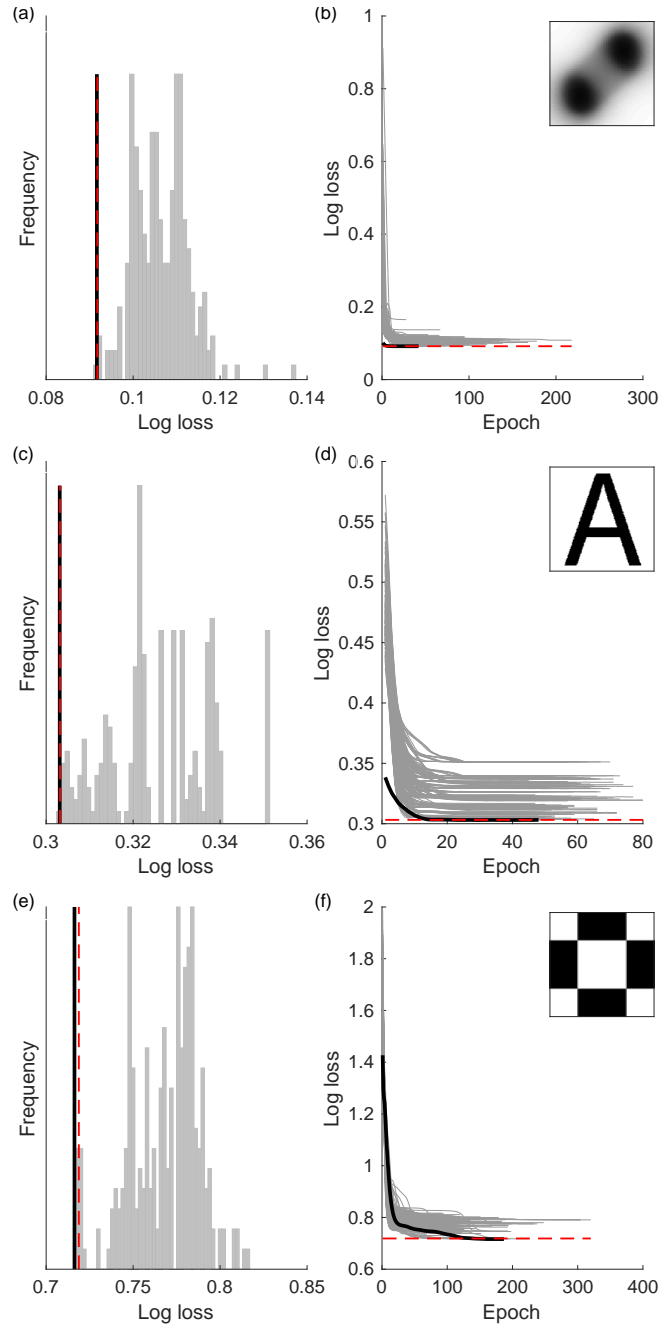


Figure 4.6: **Optimality of fitting procedure.** The optimality of solutions obtained using least-squares initialization followed by the Adam gradient descent algorithm (thick black lines) is compared to solutions obtained from genetic algorithms (dashed red line) and from gradient descent within each orthant in the weight space (gray lines). The gray lines represent the loss achieved by initializing the weights in each orthant, followed by Adam gradient descent constrained to that orthant. For three sample images, (a), (c), and (e) show histograms of the log loss achieved from the orthant search solutions, with the loss obtained by genetic algorithm optimization and unconstrained Adam gradient descent shown. (b), (d), and (f) show the trajectory of the loss during training of the network for each gradient descent.

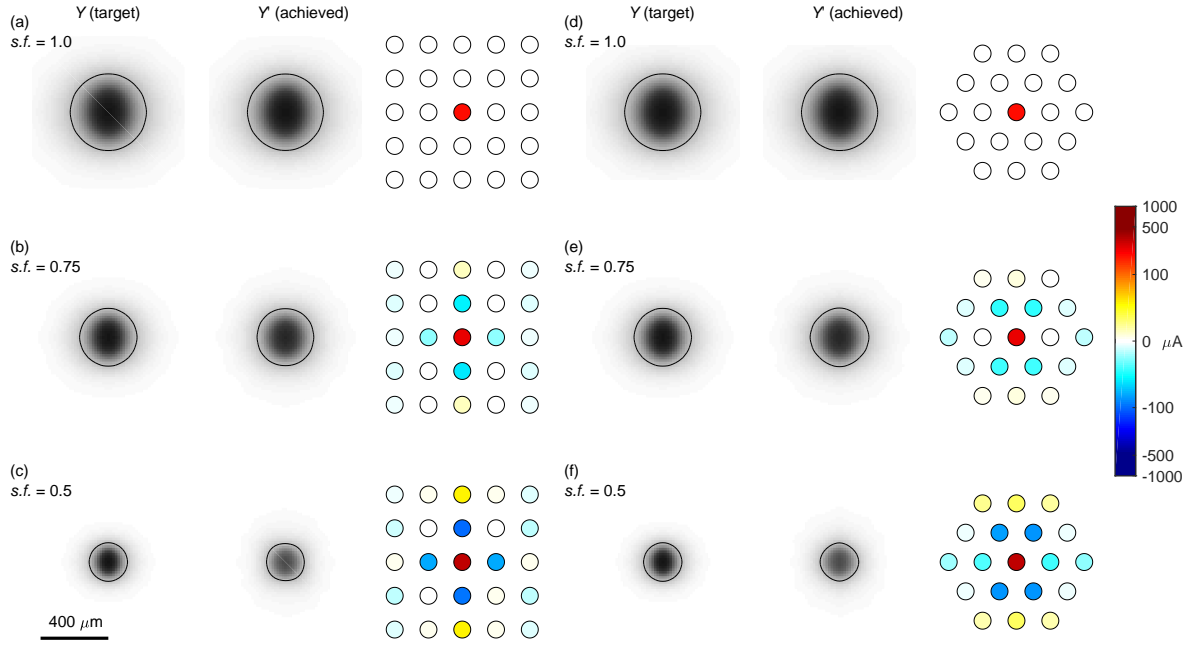


Figure 4.7: **Optimization of currents for focal stimulation.** Using biphasic stimulation, the linear-nonlinear network was trained to stimulate progressively smaller regions of the GCL. The target activation was based on the simulated activation resulting from stimulation with a single electrode. This target was then shrunk by a proportional spread factor ($s.f.$) to 75% and 50% of its size. (a-c) The target activation pattern, achieved activation, and optimized stimulus currents (trained network weights), respectively, for a rectangular electrode arrangement. (d-f) The target activation pattern, achieved activation, and optimal stimulus currents for a hexagonal electrode arrangement. Solid black lines represent the 0.25 activation (i.e., 25%) contour lines (1 represents maximum activation). For all simulations, the center-to-center electrode pitch was $200\ \mu\text{m}$ and the electrode array was located epiretinally, $100\ \mu\text{m}$ from the inner retinal surface. Only activation in the GCL was assessed ($\lambda = 1$).

Focal activation

Figure 4.7 demonstrates the optimization of electrode currents for focal stimulation for a range of decreasing spread factors. Here, a spread factor of 1 corresponds to the radius of activation achieved using stimulation with a single electrode (with a stimulus of $200\ \mu\text{A}$). A lower spread factor corresponds to a smaller radius of activation. As before, the center-to-center electrode pitch was $200\ \mu\text{m}$ and the electrode array was located epiretinally, $100\ \mu\text{m}$ from the inner retinal surface. Only a single target activation pattern (the GCL) was considered in this simulation by setting $\lambda = 1$ (Equation 4.16).

In agreement with experimental research on current steering and focal activation (Opie et al., 2013; Matteucci et al., 2013; Spencer, 2018), optimization of stimulus currents achieves focal activation by stimulating with surrounding electrodes at an opposite polarity to the center electrode. No significant difference could be seen between rectangular or hexagonally-arranged electrodes. Due to the tissue anisotropy introduced by the NFL, the optimal current delivered to the surrounding electrodes is not quite radially symmetric about the center electrode.

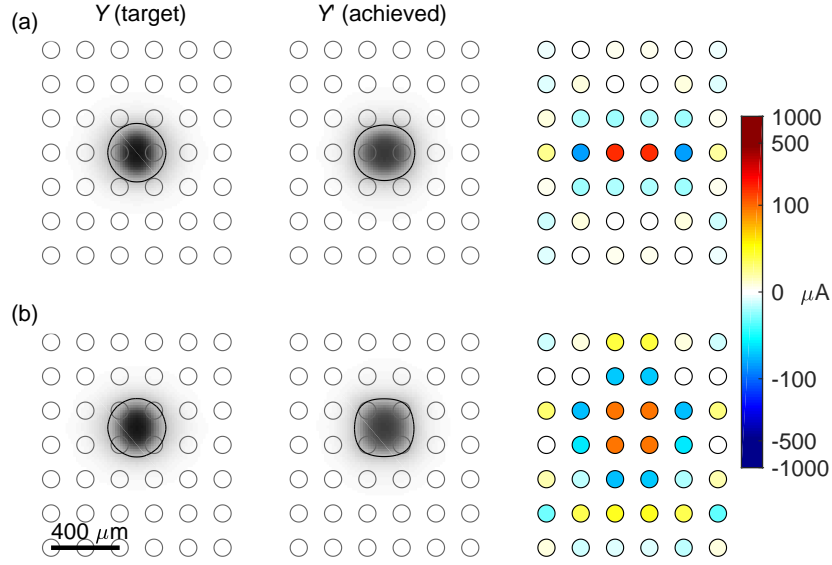


Figure 4.8: **Optimization of currents for creating virtual electrodes.** For several target locations, virtual electrodes with a focal spread factor of 0.75 were achieved using biphasic stimulation via inversion of the linear-nonlinear network model. Virtual electrodes were centered (a) in-line between two electrodes and (b) between four electrodes. Solid black lines represent the 0.25 activation (i.e., 25%) contour lines. The center-to-center electrode pitch was $200\ \mu\text{m}$ and the electrode array was located epiretinally, $100\ \mu\text{m}$ from the inner retinal surface. Only activation in the GCL was assessed ($\lambda = 1$).

Virtual electrodes

Using the same electrode array geometry, Figure 4.8 demonstrates the optimization of electrode currents for creating virtual electrodes. This is achieved while simultaneously focusing the field of activation (achieving both focal activation and manifesting virtual electrodes simultaneously). Virtual electrodes are targeted at multiple locations: mid-way between two electrodes and mid-way between four electrodes. In each case, inversion of the linear-nonlinear network recovered electrode currents that accurately reproduced the target activity.

4.3.5 Minimizing activation of the NFL

The above analyses aimed to optimize patterns of activation in the GCL, but neglected any collateral activation that may have occurred in the NFL. To simultaneously optimize GCL target activation and minimize activation of the NFL, target patterns were specified for both the GCL and NFL, corresponding to multiple outputs from the neural network, as shown in Figure 4.3. In this case, the NFL target was a vector of zeros, since no activation of the NFL was desired. A center-to-center electrode pitch of $200\ \mu\text{m}$ and epiretinal electrodes were used, located $100\ \mu\text{m}$ from the retinal surface.

Figure 4.9 shows the result of stimulus current optimization for a range of values of the weighting factor, λ , between the GCL and NFL targets. As described in Section 4.2.9 of the Methods, $\lambda = 1$ corresponds to ignoring activation of the NFL and $\lambda = 0$ corresponds to ignoring the GCL. Figure 4.9(a-

d) shows the results for λ values of 1, 0.75, 0.5, and 0.25, respectively. In agreement with the results presented in Chapter 1 of this thesis, activation of the NFL was minimized by stimulating with multiple electrodes in-line with the orientation of passing fibers in the NFL. A consequence of this stimulation pattern is the elongation of activation in the GCL. Figure 4.9(e-h) presents similar results, but where the current delivered to the central electrode was fixed at 200 μ A and the surrounding electrodes were optimized. This approach appears to perform less favorably, resulting in greater elongation of GCL activation and greater NFL activation.

4.3.6 Optimization of complex stimuli

An important test of the optimization regime presented here is its performance with more complicated target patterns, such as text and real-world visual scenes. Figure 4.10 shows the performance of the algorithm for each of these cases, where activity in both the GCL and NFL was optimized. For comparison, the pattern of activation achieved using a naive stimulation strategy is shown in Figure 4.10(c) and (d). Figure 4.10(e)-(l) demonstrate the achievable patterns of stimulation in the GCL and NFL as the importance of each target is changed by varying λ . For all simulations presented in this section, an 9×9 epiretinal electrode array was used with electrode pitch of 200 μ m.

From inspection of Figure 4.10, it is clear that a side-effect of increasing the importance of unwanted activation in the NFL is reduction of overall stimulus current amplitudes. In this sense, dual optimization of activation in the GCL and the NFL acts in a similar way to constrained optimization, with constraints applied to stimulus magnitudes. To investigate this, Figure 4.11 shows the interaction of the GCL/NFL weighting term, λ , and an inequality constraint applied to the the norm of the stimulus current vector, $|\vec{S}_t|$. In Figure 4.11, the norm constraint varies along the x-axis and λ varies along the y-axis. Each set of two rows shows the optimized activation in the GCL (top) and the NFL (bottom). This demonstrates that constraining optimization of stimulus currents with an NFL penalty is more effective at reducing NFL activation than applying a norm constraint on the vector of currents. Only for $\lambda = 1$ did the norm constraint have a clear impact on NFL activation. In contrast, a substantial reduction in NFL activation was observed using $\lambda = 0.75$ relative to $\lambda = 1$, with diminishing returns for progressively smaller values of λ . The downside in constraining stimuli using activation in the NFL is seen in the reduction in resolution in activation in the GCL.

4.3.7 Epiretinal versus subretinal stimulation

The challenge of avoiding stimulation of the NFL when conducting epiretinal stimulation, as illustrated in Figures 4.9, 4.10, and 4.11, necessitates a comparison of epiretinal and subretinal stimulation using the methodology developed here. Due to the greater distance from the NFL, subretinal electrode arrays are less likely to activate passing fibers in the NFL. This should have the effect of reducing the patterns of unwanted stimulation seen above.

Figure 4.12 shows the achieved activation patterns in the GCL and NFL under subretinal stimulation using the two target images shown in Figure 4.10(a) and (b). For each image, very little activation is observed in the NFL, even when NFL activity is ignored by the optimization (i.e., $\lambda = 1$). Furthermore, from inspection of the optimized patterns of activation in the GCL, a similar quality of image reproduction is achieved for subretinal stimulation as was observed for epiretinal stimulation.

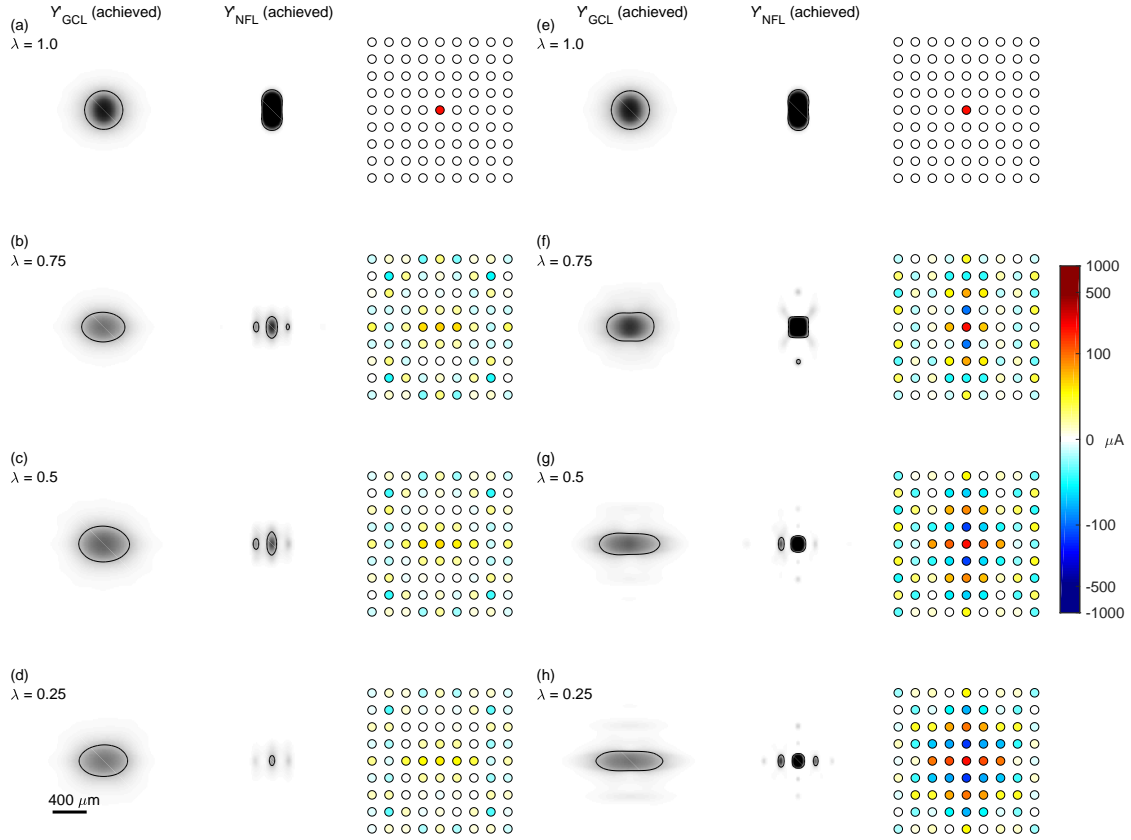


Figure 4.9: **Minimizing activation of passing axons.** Using separate target outputs from the GCL and NFL, biphasic electrode currents were optimized for a range of weightings between the two targets. In each case presented, the NFL target is a vector of zeros and the GCL target is a single phosphene (identical to that shown in (a)). (a-d) The activations achieved in the NFL and GCL for $\lambda \in \{1.0, 0.75, 0.5, 0.25\}$, respectively, where λ represents the weight assigned to the optimization of activation in the GCL. (e-h) The same analyses as in (a-d), respectively, with the current to the center electrode constrained to 200 μA . Solid black lines represent the 0.25 activation (i.e., 25%) contour lines. For all simulations shown, the center-to-center electrode pitch was 200 μm and the electrode array was located epiretinally, 100 μm from the inner retinal surface.

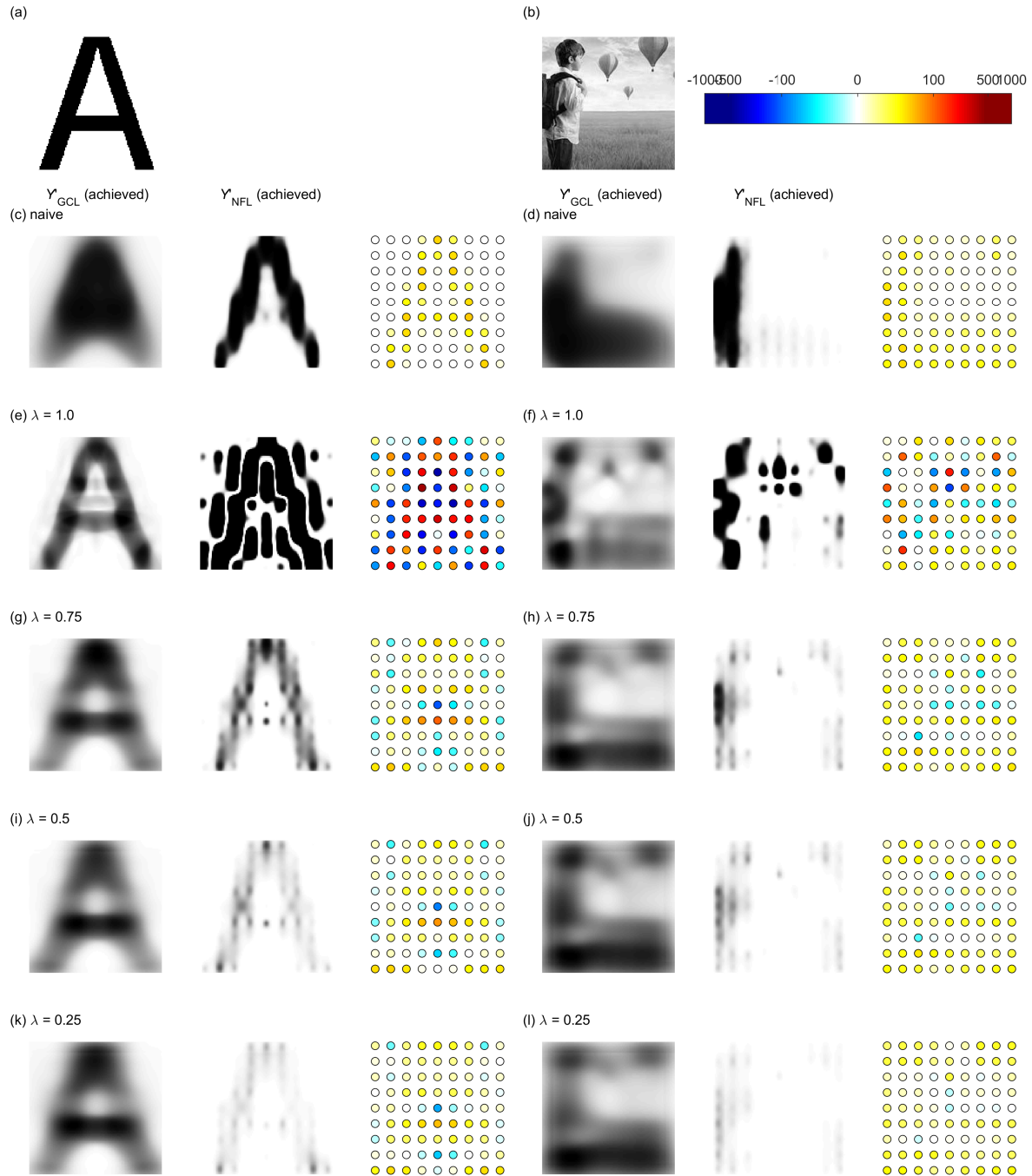


Figure 4.10: **Optimization of complex stimuli for epiretinal stimulation.** Image of (a) a letter and (b) a real-world scene were used to test the performance of the biphasic linear-nonlinear network inversion scheme. (c, d) The pattern of activation achieved for each image using a naive stimulation strategy, in which the stimulus current delivered to each electrode is proportional to the intensity of the image in the surrounding region. (e,f), (g,h), (i,j), and (k,l) Optimized GCL activation, NFL activation, and stimulus currents for λ -values of 1.0, 0.75, 0.5, and 0.25, respectively, where λ represents the fractional weight applied to the GCL target and $1 - \lambda$ is the NFL weight. All simulations presented used a 9×9 epiretinal electrode array with electrode pitch of $200 \mu\text{m}$.

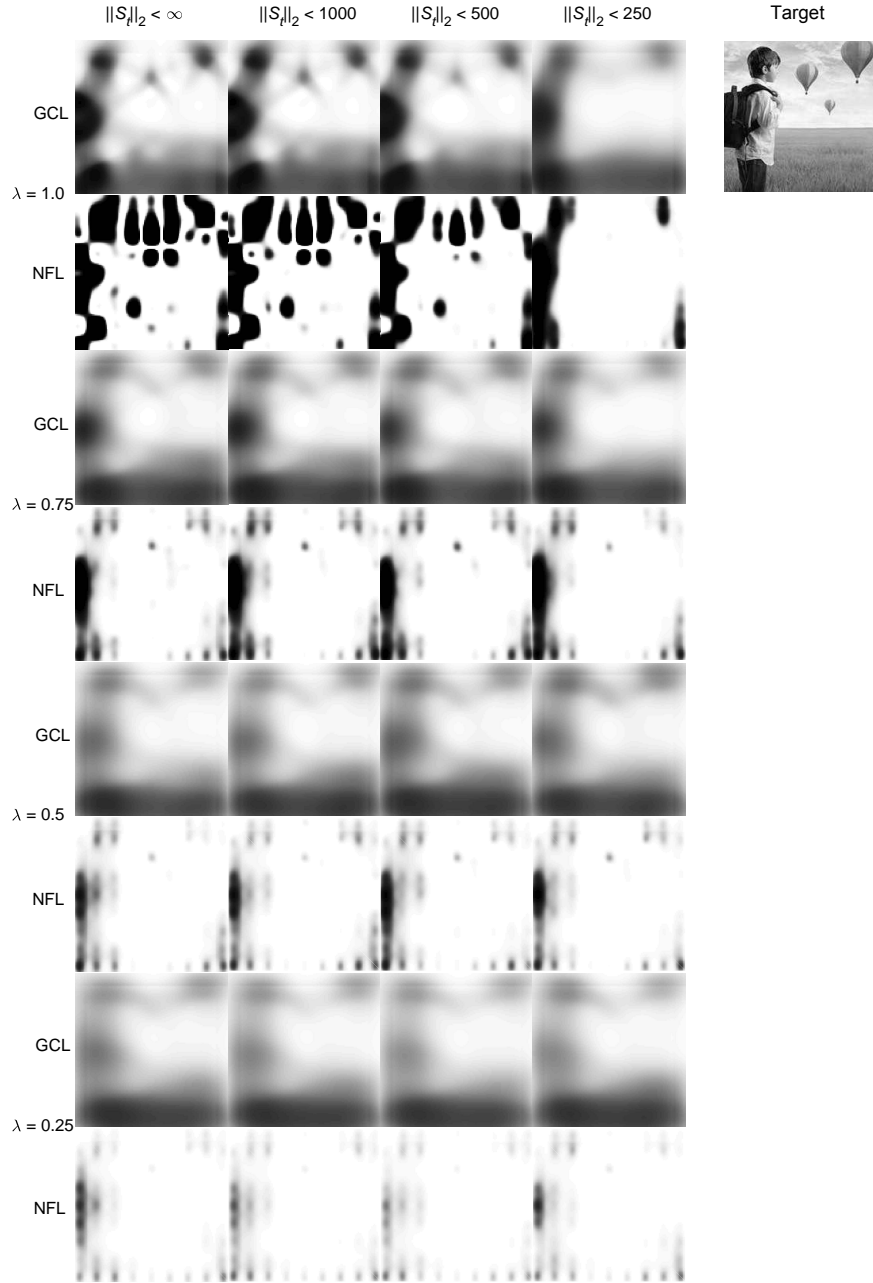


Figure 4.11: **Interactions of NFL constraints and current constraints.** For the real-world visual scene shown in the right-hand side of the figure, the patterns of GCL and NFL activation were optimized for various values of the GCL/NFL weight factor, λ , and an inequality constraint applied to the norm of the stimulus current vector, \vec{S}_t . For each set of two rows, the top row shows activation in the GCL and the bottom row shows activation in the NFL. All simulations presented used a nine-by-nine epiretinal electrode array with electrode pitch of $200 \mu\text{m}$.

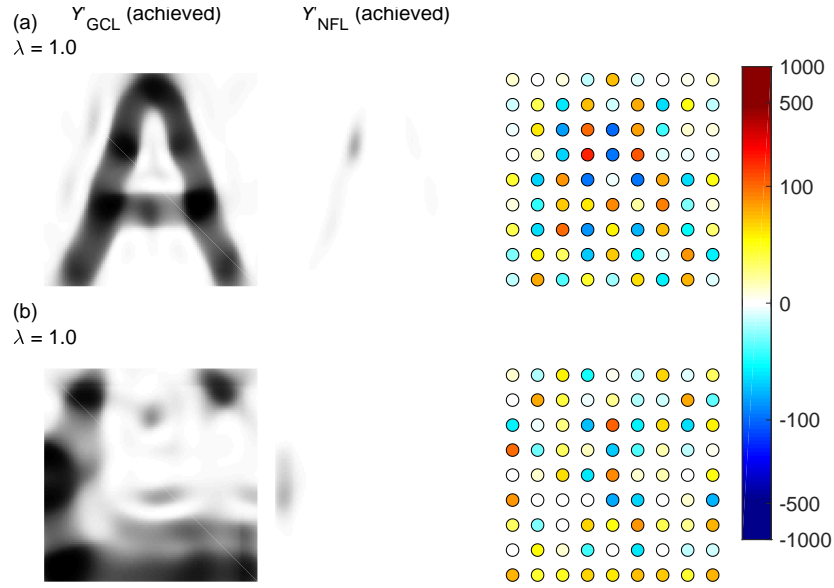


Figure 4.12: **Optimization of subretinal stimulation.** For the (a) letter A and (b) real-world scene target images shown in Figure 4.10, the biphasic linear-nonlinear network was inverted to optimize subretinal stimuli. In these optimizations, only the target pattern for the GCL (i.e., the target image) was considered, with activation in the NFL being ignored by the optimizer. A 9×9 electrode array with electrode pitch of $200 \mu\text{m}$ was used.

4.3.8 Comparison of animal model anatomies

Due to the variation in thickness of the NFL between different species (human: $100 \mu\text{m}$ (Oishi et al., 2009; Frenkel et al., 2005; Wang et al., 2013; Thomson et al., 2015), rat: $45 \mu\text{m}$ (Abbott et al., 2014)), analyses were also conducted using the geometry of the rat for the NFL by setting the thickness to $45 \mu\text{m}$. Figure 4.13 shows the results of this analysis for the two target images from Figure 4.10. Despite the lower NFL thickness of the rat and lower distance between electrodes and GCL, similar results were observed for both animal models.

4.4 Discussion

In this chapter, we have demonstrated a method for the spatial shaping of neural activity in retinal ganglion cells via inversion of the linear-nonlinear network. The approach is capable of rapidly optimizing stimulus currents to desired complex images or patterns using a biophysical model of the retina. Furthermore, the strategy can be used to fit to multiple target activation patterns for multiple retinal layers, and can be constrained to ensure clinically-acceptable stimulus parameters. The artificial neural network form of the linear-nonlinear model used here is flexible and it would be straightforward to generalize it to account for complexities such as asymmetric stimulation thresholds and temporally-dynamic ERFs. Furthermore, the approach developed in this study is applicable to any application in which linear-nonlinear models are relevant, including cortical stimulation.

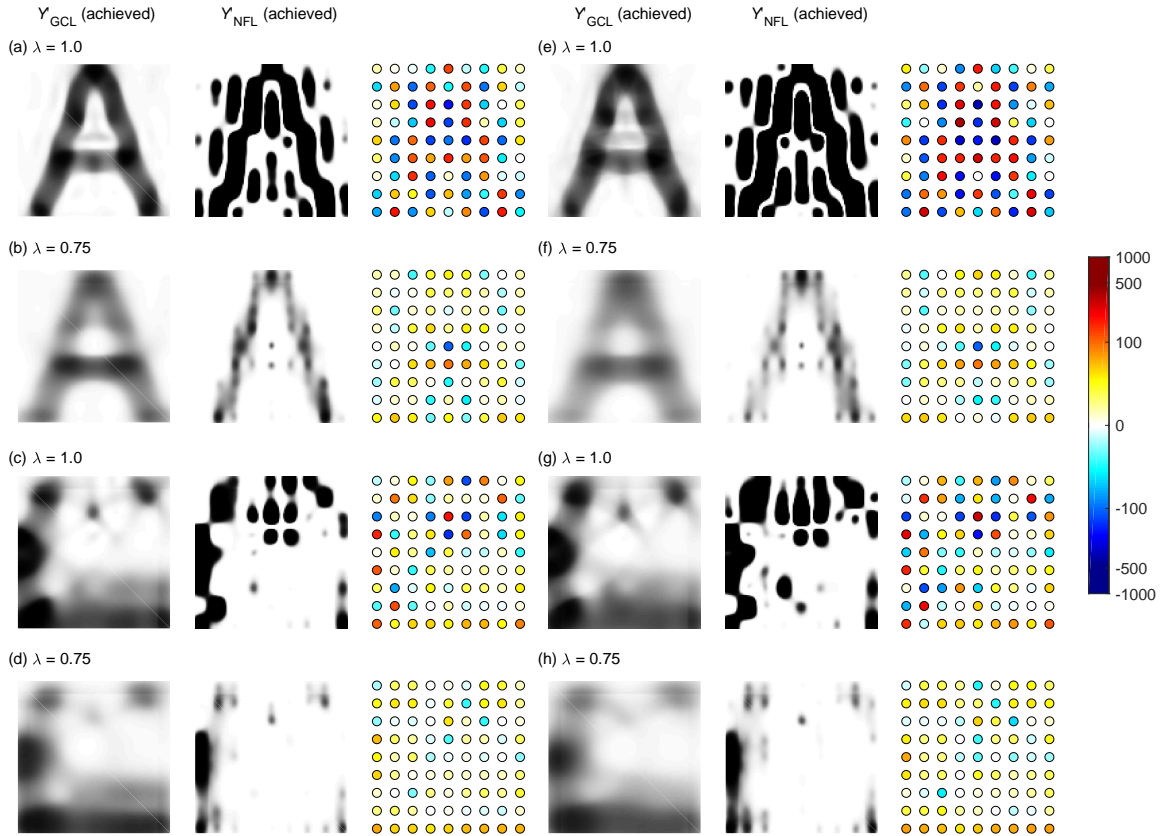


Figure 4.13: **Comparison of stimulation of rat and human retinas.** The biphasic linear-nonlinear network was inverted to optimize epiretinal stimulus currents for human and rat NFL thicknesses of $100 \mu\text{m}$ and $45 \mu\text{m}$, respectively. For each retinal geometry, optimizations were performed by each of the target images shown in Figure 4.10. (a-d) Achieved GCL and NFL activation pattern for rat-like retinal geometry for λ -values of 1.0 and 0.75. (e-h) Achieved GCL and NFL activation pattern for human-like retinal geometry for λ -values of 1.0 and 0.75. The simulated electrode array and retinal geometry used was identical to that of Figure 4.10 apart from the thickness of the NFL.

4.4.1 Optimality of solutions

As demonstrated in Figure 4.5 and Appendix C.1, inversion of the linear-nonlinear model with biphasic stimulation is challenging due to the potentially large number of locally-optimal solutions. As a result, there is no guarantee that a gradient-descent algorithm, such as Adam, will find the globally optimal inversion. As shown in Figure 4.6, the inversion strategy achieves solutions with a range of loss values when constrained to localized regions of the solution space. In contrast, when allowed to traverse the entire solution space, the inversion strategy finds an optimal or near-optimal solution for each of the target images tested.

This result can be explained in two possible ways: 1) the Adam gradient-descent algorithm is effective when searching across the solution space to find the best solution, avoiding becoming caught in local optima or 2) the least-squares initialization used in this work is effective at initializing the algorithm at a location in the vicinity of the global optimum. Similar results to those presented in Figure 4.6 were achieved when the analysis was repeated with random initialization of network weights (despite requiring longer to converge for some targets), indicating that it is the robustness of the Adam algorithm that is responsible for the optimality of the results. As described in the Methods, the Adam algorithm contains a mechanism for *momentum*, which acts to prevent the algorithm from converging prematurely into local minima (Kingma and Ba, 2015). Due to the large number of local optima for biphasic stimulation, the objective/optimization surface will have an inherent *bumpiness*. The second order momentum-like mechanism in the Adam algorithm is able to self-tune in a way that enables it to traverse that surface without getting caught in localized troughs (local optima) (Ruder, 2016).

It is important to note that the analysis of optimality shown in Figure 4.6 was conducted for only a small number of electrodes (9). It may be that for more electrodes, or different electrode array geometries, results will vary. However, due to the computational difficulty in determining the true optimal solution, it was not feasible to conduct the analysis for more complex electrode arrays given time and resource constraints. If, in the application of a strategy such as this, low-quality inversions are observed, alternative, non-biphasic stimulation strategies may be considered. For example, the use of charge-balanced tri-phasic stimuli may limit activation to a single stimulus phase (Jepson et al., 2013), thereby reducing or eliminating the need for a double-sided nonlinearity.

4.4.2 Constrained optimization

As demonstrated in Figure 4.11, inversion of the linear-nonlinear neural network can be constrained to ensure the clinical safety of optimized stimuli. However, it is also evident from the interaction of λ (the weighting of GCL activation over NFL activation) and norm constraints that constraining solutions by penalizing unwanted NFL activation has a similar, but more marked, effect. The dominance of the effect of λ shown in Figure 4.11 demonstrates that co-targeting both the GCL and the NFL reduces the need to constrain the currents. Furthermore, unlike a hard constraint on stimulus amplitudes, constraining based on NFL activation is informed by desired patterns of activation, and so is a preferable method of constraining currents. However, since including a term for the NFL does not *guarantee* clinical constraints on current amplitude and power are met, the best solution might include both approaches, with a fixed current constraint serving predominantly as a backup measure.

A key result from Chapter 2 of this thesis was that activation of the NFL could be reduced by

simultaneously stimulating with multiple electrodes aligned with the axons of passage in the NFL. In agreement with this result, Figure 4.9 shows that, as activation in the NFL is given more weight in the optimization scheme, multiple electrodes aligned with fibers in the NFL are recruited. This can be seen by comparing Figure 4.9(a), in which a single electrode was used, with Figure 4.9(b-d), in which sets of horizontally-adjacent electrodes with equal polarity were used. Similarly, the same result can be seen by comparing Figure 4.9(e) with Figure 4.9(g-h). As seen most clearly in Figure 4.9(h), stimulus currents were optimized such that horizontally adjacent electrodes (aligned with the NFL) received stimuli of the same polarity, whereas vertically adjacent electrode received stimuli with alternating polarity.

Constraining terms may also be used to invert an undetermined form of the linear-nonlinear model. When using experimentally-determined ERFs as opposed to ERFs estimated with a computational model, it may be that fewer recording electrodes are used than stimulating electrodes, resulting in fewer sampled locations, M , than electrodes, N . In this situation, the problem of selecting an optimal set of N electrode currents given a desired pattern of activation across M locations in the retina is an undetermined problem, yielding infinitely many solutions. A common method for solving this problem in statistics is to use ridge regression (also known as L2-regularization, Tikhonov regularization, or weight decay) (Hoerl and Kennard, 1970; Tikhonov et al., 1995). Ridge regression uses an altered objective function in which a term is included that penalizes potential solutions based on the size of the weights (current amplitudes in our case) (Ng, 2004). This biases an inversion algorithm toward a smaller subset of potential solutions; those that require less current. Following appropriate tuning of the regularization parameters, this allows for the approximate inversion of the undetermined system (Gerwinn, 2010).

4.4.3 Real-world application

Although not examined in this work, a crucial next step towards the *in vivo* and clinical application of a spatial shaping method is determining its ability to select appropriate stimuli in real time. A key difference between a proof of concept such as this and real-time implementation is that the target images are likely to be constantly changing as the camera surveys the visual scene. In this respect, a neural network and an iterative optimization algorithm, such as Adam, are likely to be well suited to real-time implementation. As described in Section 4.2.3, the application of neural networks used in this chapter is unconventional as it is the training procedure, and not its predictive application, that is of interest. Despite this, neural networks are well suited for applications that required both online training (relevant to spatial shaping) and online prediction (not relevant to spatial shaping) (LeCun et al., 2015).

The use of mini-batch stochastic gradient descent, as in Adam, provides a sensible approach to the online application of such as model. At initialization, when stimulus currents are likely to be far from optimal, mini-batch stochastic gradient descent allows the visual scene to be coarsely sampled to obtain an approximately optimal set of stimulus currents very rapidly. From this point onward, for static or continuously changing visual scenes, the quality of the solution would be expected to improve as more samples are used for optimization. In the case of a dramatic and sudden change in the visual scene, the optimization algorithm is able to rapidly adapt to a near-optimal solution using sampled representations. In this way, stochastic gradient descent will inherently tune the currents to a level of optimality appropriate to the level of change in the visual scene. Provided a single iteration of

the optimization algorithm is fast enough, this will maintain an effective set of stimuli in response to changing target images. All optimization conducted in this study were run using a commodity CPU (quad-core, 3.6 GHz), and so are not representative of the best-case performance of the algorithm. Online training of neural networks can be implemented far more efficiently on a field-programmable gate array, allowing for extremely efficient, hardware-based, stimulus current optimization (Bahoura and Park, 2011; Gironés et al., 2005; Savich et al., 2012). Despite this, based on the optimizations run during this analysis, iterative optimization updates can be provided at a rate of at least 100 Hz, and generally requires less than 10 iterations to obtain a solution within 5% of the optimal value of the objective function.

In contrast to this, approaches that involve repeated matrix inversion (e.g., constrained least squares approaches) may not be well-suited to online optimization. For constrained optimization, these approaches will require an approximate matrix inversion for each stimulus current update. Although this may also be performed efficiently using hardware such as field-programmable gate arrays, the requirement for all data (i.e., all pixels in an image) to be processed before providing an updated set of stimuli reduces the ability to track constantly changing visual scenes. In addition, an online approach that initializes the update step with the previously learned stimuli allows for the continuous changing of stimulus currents, whereas matrix-inversion approaches may result in sharp and large changes in stimulus patterns. This is due to the fact that there is no connection between one solution and the next, potentially resulting in confusing percept transitions.

There are undoubtedly errors in the biophysical linear-nonlinear model presented here in predicting true responses from retinal ganglion cells, as would be expected with any computational model. Complications include macroscopic heterogeneity in the density and type of retinal ganglion cells across the retina, differences in the conductivity of different retinal layers, and uneven placement of electrode arrays with respect to the retinal surface; these will all cause model prediction errors and lead to suboptimal estimated stimulus currents. To account for issues such as these, mechanisms can be put in place in clinical application to tailor solutions to specific wearers of retinal prostheses. For instance, some wearers may experience a greater sensitivity to certain electrodes due to their individual retinal anatomy or irregular device placement. In this situation, an approximate sensitivity matrix can be determined from the individual by stimulating with one electrode at a time. This matrix can then be implemented in the linear-nonlinear network as an additional network layer so that model inversion automatically accounts for varying sensitivities across the electrode array. Alternatively, or in addition, separate sets of stimulating and recording electrodes may be used to aid in determining ERFs.

Finally, it is important to note that, while a particular biophysical model was used here to estimate ERFs, this could be replaced with either a different linear model or a set of experimentally-estimated ERFs. In fact, provided passive (i.e., linear) membrane dynamics are used, a full morphologically-accurate representation of a population of neurons could be used to estimate ERFs, without any loss to the efficiency of the inversion strategy. This extension represents a significant body of work and requires the building of many morphologically-diverse neural models, but may offer a method for more accurately estimating patterns of activation among neural populations. Consequently, it should form a key component of future work in this area.

Chapter 5

Conclusion

5.1 Summary

5.1.1 Minimizing activation of overlying axons

In the first component of this project, a novel multi-layered model of current flow and passive neural activation in the retina was developed. Based on the cellular composite model (Meffin et al., 2014), this model addresses a number of limitations in previously reported retina models by eliminating internal model inconsistency, capturing the spatial and temporal dependencies associated with electrical stimulation of neural tissue, and explicitly modeling the effect of fiber orientation. As described in Chapter 2, this model was used to demonstrate that activation of retinal ganglion cells in the inner retina under epiretinal stimulation depends on both axonal orientation and orientation of the stimulating electric field relative to the orientation of axons of passage in the nerve fiber layer. The developed model allowed for an analysis of this dependence by capturing the distinct distributions of fiber orientation of the nerve fiber layer and the ganglion cell layer. A four-electrode stimulation strategy has been proposed that accomplishes preferential activation of the retinal ganglion cell axon initial segment over passing axons in the nerve fiber layer using clinically suitable stimulus charge densities and electrode configurations. Although concessions must be made with regard to the radius of activation in the ganglion cell layer, these are relatively minor, and the proposed strategy is expected to enable higher resolutions and more clearly interpretable percepts by users of epiretinal prostheses.

5.1.2 The biophysical basis for the linear electrical receptive field

Based on the multi-layered biophysical model developed in the previous study (Chapter 2) and an active morphologically-accurate neuron model, the linear-nonlinear model of electrically-elicited retinal ganglion cell responses was recovered. This process validates the assumptions made by the linear-nonlinear model and offers an explanation for those cases that deviate from the model. The linear electrical receptive field of the linear-nonlinear model was shown to be determined by a linear combination of the stimulus-induced transmembrane current, or activating function, at key locations in the

cell: at and near the site of action potential initiation. We have shown that a retinal ganglion cell can be accurately represented by (1) a purely linear approximation of multi-electrode stimulation and the integration of inputs in the cell followed by (2) a fixed spiking threshold at one or a small number of locations. Furthermore, these two phases correspond directly to the linear projection and spiking nonlinearity of the linear-nonlinear model, respectively. The results from this study demonstrate that a linear electrical receptive field can be accurately recovered for each identified site of action potential initiation, even for cells for which a single linear description is insufficient. For each site of action potential initiation, the electrical receptive field was almost perfectly reconstructed using a weighted sum of activating function terms from a small set (< 20) of nearby compartments.

5.1.3 Spatial shaping of retinal ganglion cell activation

In Chapter 4, a method was demonstrated for the spatial shaping of neural activity in retinal ganglion cells via inversion of a generalized form of the linear-nonlinear model. The linear-nonlinear model was first extended to allow for a high-dimensional representation of the retinal tissue (e.g., multiple volumes of multiple cell types/orientations). This model was then translated into an equivalent artificial neural network: the linear-nonlinear network, which is applicable to both experimentally- and computationally-estimated electrical receptive fields. Based on the biophysical validation presented in Chapter 3, electrical receptive fields were computed for a population of retinal ganglion cells using the biophysical model of passive activation developed in Chapter 2. This allowed for a high-resolution representation of the retina and target patterns of activation. Using the Adam optimization algorithm, developed for training neural networks, an efficient and optimal method for the inversion of the linear-nonlinear network was demonstrated. The presented approach is capable of rapidly optimizing stimulus currents to desired complex images or patterns using a complex biophysical model of the retina. Furthermore, the strategy can be used to fit to multiple target activation patterns for multiple retinal layers, and can be constrained to ensure clinically-acceptable stimulus parameters. The artificial neural network form of the linear-nonlinear model used here is flexible, and is straightforward to generalize to account for complexities such as asymmetric cathodic/anodic stimulation thresholds and temporally-dynamic ERFs.

5.2 Future areas of investigation

Due to the computational and theoretical nature of the research conducted throughout this doctorate, it is imperative that the outcomes of each study be translated into experimental predictions and to suggest methods for experimental validation.

5.2.1 Minimizing activation of overlying axons

As this study deals with activation in two distinct regions of the retina, the GCL and the NFL, controlled experimental validation of these results requires techniques for the measurement of RGC activation at multiple locations in the retina simultaneously. *In vitro* studies in which the average trajectory of passing axons in the NFL is known will allow for measurements of activation to be taken in the GCL at both the region being targeted by stimulation and at more distant locations that

lie under the trajectory of passing axons. Micron-scale electrical imaging of axonal action potential propagation (Zeck et al., 2011; Bakkum et al., 2013) and methods for monitoring GCL activity across the whole retina using calcium imaging (Weitz et al., 2013) may facilitate accurate validation. Recent developments in high-density stimulation and recording, which enable initiation of action potentials in precise locations of a neuron and imaging of their propagation, may also serve as valuable, non-clinical platforms for the validation of a strategy such as this (Radivojevic et al., 2016; Eickenscheidt and Zeck, 2014). A challenge with quantitatively validating the result in this chapter is that the small distance between the electrode array and the surface of the retina must be very tightly controlled.

Due to the dependence of these results on the anisotropy of the NFL, it is expected that varying the thickness of the NFL will have a marked effect. The chosen NFL layer thickness is based on an approximation of the human retina, and so these results are relevant only to human retinal stimulation. Rodent models used for research and testing of epiretinal implants have thinner NFLs and so the influence of retinal layer orientation will be less pronounced. Although this in no way confounds the current findings, it suggests that experimental validation would be best carried out in the primate retina. A potential solution for other animal models may be to modify the present model to represent the relevant animal so that any observed evidence can be extrapolated to human-like retinal geometries. It is important to note that a large part of the present result derives from electrode configuration, which can be kept consistent across different animal models.

In a real-world implanted system, the assumption that the distance between each electrode and the NFL is the same is unlikely to hold. In addition, the electrode-retina separation distance is likely to change over time due to device settling, immune responses and changes to the implant's environment caused by vitrectomy. Challenges such as these will require careful modification of the idealized solution presented here, and are likely to require either measurement for each patient of the precise three-dimensional location of the electrode array over time using methods such as OCT or measurement of the retina's electrical response to stimulation. Procedures such as those used by Grosberg et al. (2017) for electrical recording of RGCs following stimulation provide a potential solution. This method enables the classification of responses as either AOP activation or AIS/soma activation. This information could then be used to tune the multi-electrode stimulation strategy over time.

5.2.2 Spatial shaping of retinal ganglion cell activation

The results presented here demonstrate the potential of an approach to spatial shaping of neural activity in the retina. As discussed Chapter 4, in order to implement the algorithm in an animal model or clinical setting, it needs to be implemented in a flexible and efficient manner. We believe that the neural network inversion strategy presented here is both sufficiently flexible and sufficiently efficient (i.e., sub-second updates) to be implemented in an online manner and to account for a number of potential complications that may be encountered in a real-world scenario. A next step towards this goal is to demonstrate the online application of the stimulus optimization approach in a simulated environment, consisting of constantly changing target images taken from video recordings. Following this, an experimental proof-of-concept can hopefully be conducted using simultaneous multi-electrode stimulation in an *in vitro* animal model.

Another focus of the work presented in this study was to examine the potential of spatial shap-

ing to not only shape the pattern of activity in a target retinal layer (the GCL) but also minimize undesirable collateral activation (in the NFL), shown in Figure 4.10. This was achieved by using a weighted objective function that included terms for both regions of interest. Although this approach is warranted, it fails to capture the effect of unwanted activation in the NFL on activation in the the GCL. That is, when an overlying axon is activated in the NFL, it corresponds to activation of a retinal ganglion cell in the GCL at a potentially distant location. It is the final pattern of activation in the GCL (resulting from both direct activation and activation of the NFL) that corresponds most closely with perception. For this reason, it may not be ideal to blindly minimize activation of the NFL. Given the framework presented in this study, it is possible to combine *direct* activation in the GCL and NFL in order to optimize the *final*, combined activation in the GCL. This can be achieved by including an additional neural network layer that combines the impact of activation in the NFL on the GCL. In doing so, stimulus currents can be optimized to take advantage of activation in the NFL where it results in advantageous activation in the GCL, a concept that has not yet been explored in either experimental or theoretical research.

Finally, the approach presented in this Chapter 4 is capable of accounting for 1) temporal dependencies in the ERF and 2) differing sensitivities to cathodic-first and anodic-first stimuli. Both of these complexities have been shown to be important under certain circumstances (Maturana et al., 2016, 2018). As a result, care was taken in this study to preserve the generality of the model with respect to these extensions. The form of the linear non-linear model used here, described by Equation 4.5(b) and Equation 4.6(b), allows for differing thresholds between cathodic-first and anodic-first stimuli via the terms c_r^+ and c_r^- , respectively. Similarly, due to the flexibility of the neural network framework, input ERFs, presented as vectors in this work (see Figure 4.3), can be presented as matrices by increasing the dimensionality of the model's input. This will allow for temporally-varying ERFs to be used for stimulus optimization.

Finally, the linear-nonlinear model explored in Chapter 3 and generalized in Chapter 4 is not tied to a particular model of passive membrane activation, or the ERF. As such, more complex linear models of passive activation could be substituted to more accurately capture the diversity of neuron morphologies and membrane dynamics. Using a large set of morphological neuron reconstructions, a more morphologically-informed version of the ERF matrix, \mathbf{V} could be developed using passive membrane model in software such as NEURON (Hines and Carnevale, 1997).

5.3 Final remarks

In this thesis, the use of simultaneous multi-electrode stimulation in developing improved stimulation strategies for the retina has been explored. Various strategies for controlling electrically-induced retinal activation have been elaborated using detailed biophysical models, simple phenomenological models, and a combination of each. Although further experimental validation of the findings presented here is required prior to clinical implementation, as with any computational work, it is the expectation of the author that strategies better-informed by accurate modeling and model inversion are key to the development of an effective visual prosthesis system.

Appendix A

Supporting Material for Chapter 2

A.1 Solution of volume equations

Expanding Equation (2.2) for each layer gives

$$\mathbf{J}_V = \frac{-1}{4\pi^2} \left(\xi_{V_x} * \frac{\partial \phi_V}{\partial x} \mathbf{e}_x + \xi_{V_y} * \frac{\partial \phi_V}{\partial y} \mathbf{e}_y + \xi_{V_z} * \frac{\partial \phi_V}{\partial z} \mathbf{e}_z \right), \quad (\text{A.1a})$$

$$\mathbf{J}_N = \frac{-1}{4\pi^2} \left(\xi_{N_x} * \frac{\partial \phi_N}{\partial x} \mathbf{e}_x + \xi_{N_y} * \frac{\partial \phi_N}{\partial y} \mathbf{e}_y + \xi_{N_z} * \frac{\partial \phi_N}{\partial z} \mathbf{e}_z \right), \quad (\text{A.1b})$$

$$\mathbf{J}_G = \frac{-1}{4\pi^2} \left(\xi_{G_x} * \frac{\partial \phi_G}{\partial x} \mathbf{e}_x + \xi_{G_y} * \frac{\partial \phi_G}{\partial y} \mathbf{e}_y + \xi_{G_z} * \frac{\partial \phi_G}{\partial z} \mathbf{e}_z \right), \quad (\text{A.1c})$$

where \mathbf{e}_x , \mathbf{e}_y , and \mathbf{e}_z are unit vectors in x , y , and z directions, respectively. Note that each of the dimension-specific admittivity terms in the above expression may have (x, y, z, t) dependence.

By assuming that within each layer tissue admittivity is independent of z , we can reduce the above four dimensional convolutions to three dimensions. Using the x -component of the admittivity in the vitreous layer to illustrate, we have

$$\begin{aligned} \xi_{V_x} * \frac{\partial \phi_V}{\partial x} &= \sqrt{2\pi} \iiint_{x' y' z' t'} \bar{\xi}_{V_x}(x', y', t') \delta(z') \times \frac{\partial \phi_V(\mathbf{r} - \mathbf{r}', t - t')}{\partial x} dx' dy' dz' dt' \\ &= \sqrt{2\pi} \iiint_{x' y' t'} \bar{\xi}_{V_x}(x', y', t') \frac{\partial \phi_V(\mathbf{r} - \mathbf{r}', t - t')}{\partial x} dx' dy' dt', \end{aligned} \quad (\text{A.2})$$

where $\bar{\xi}_{V_x}$ is the x -component of the admittivity in the vitreous layer at a given point in (x, y, t) space, which does not vary with z within a layer. Identical simplifications can be shown for each of the nine convolutions in Equations (A.1). Subsequently, this will allow for the removal of these convolutions using a Fourier transform in three dimensions instead of four.

All Fourier domain transformations performed in these analyses are of the following form:

$$F(k_x) = \frac{1}{\sqrt{2\pi}} \int_x f(x) e^{-jk_x x} dx, \quad (\text{A.3a})$$

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{k_x} F(k_x) e^{jk_x x} dk_x, \quad (\text{A.3b})$$

where k_x and $F(k_x)$ are the Fourier transform pairs of x and $f(x)$, respectively.

Taking the Fourier transform of each of Equations (2.1) with respect to x , y , and t gives the following set of equations:

$$\hat{\nabla} \cdot \hat{\mathbf{J}}_V = \sum_i^M \frac{\hat{I}_i(\omega)}{\pi q} \frac{\mathbb{J}_1(q\sqrt{k_x^2 + k_y^2})}{\sqrt{k_x^2 + k_y^2}} \times \delta(z - z_i) e^{-j(k_x x_i + k_y y_i)}, \quad (\text{A.4a})$$

$$\hat{\nabla} \cdot \hat{\mathbf{J}}_N = 0, \quad (\text{A.4b})$$

$$\hat{\nabla} \cdot \hat{\mathbf{J}}_G = 0, \quad (\text{A.4c})$$

where the hat symbol ($\hat{\cdot}$) indicates the Fourier transform of the specified quantity with respect to x , y , and t , with Fourier transform pairs k_x , k_y , and ω , respectively. $\hat{\nabla} = [jk_x; jk_y; \partial/\partial z]$ is the Fourier transform of the differential operator and $\mathbb{J}_1(\cdot)$ represents the Bessel function of the first kind of order 1. Similarly, taking the Fourier transform of Equations (A.1) yields

$$\hat{\mathbf{J}}_V = -jk_x \hat{\xi}_{V_x} \hat{\phi}_V \mathbf{e}_x - jk_y \hat{\xi}_{V_y} \hat{\phi}_V \mathbf{e}_y - \hat{\xi}_{V_z} \frac{\partial \hat{\phi}_V}{\partial z} \mathbf{e}_z, \quad (\text{A.5a})$$

$$\hat{\mathbf{J}}_N = -jk_x \hat{\xi}_{N_x} \hat{\phi}_N \mathbf{e}_x - jk_y \hat{\xi}_{N_y} \hat{\phi}_N \mathbf{e}_y - \hat{\xi}_{N_z} \frac{\partial \hat{\phi}_N}{\partial z} \mathbf{e}_z, \quad (\text{A.5b})$$

$$\hat{\mathbf{J}}_G = -jk_x \hat{\xi}_{G_x} \hat{\phi}_G \mathbf{e}_x - jk_y \hat{\xi}_{G_y} \hat{\phi}_G \mathbf{e}_y - \hat{\xi}_{G_z} \frac{\partial \hat{\phi}_G}{\partial z} \mathbf{e}_z. \quad (\text{A.5c})$$

Substituting Equations (A.5) into (A.4), the system may be written as

$$\frac{\partial^2 \hat{\phi}_V}{\partial z^2} - \hat{\phi}_V \eta_V^2 = - \sum_i^M m_i \delta(z - z_i), \quad (\text{A.6a})$$

$$\frac{\partial^2 \hat{\phi}_N}{\partial z^2} - \hat{\phi}_N \eta_N^2 = 0, \quad (\text{A.6b})$$

$$\frac{\partial^2 \hat{\phi}_G}{\partial z^2} - \hat{\phi}_G \eta_G^2 = 0, \quad (\text{A.6c})$$

where

$$m_i = \frac{\hat{I}_i(\omega)}{\hat{\xi}_{V_z} \pi q} \frac{\mathbb{J}_1(q\sqrt{k_x^2 + k_y^2})}{\sqrt{k_x^2 + k_y^2}} e^{-j(k_x x_i + k_y y_i)},$$

and

$$\eta_\alpha^2 = \frac{k_x^2 \hat{\xi}_{\alpha_x} + k_y^2 \hat{\xi}_{\alpha_y}}{\hat{\xi}_{\alpha_z}}, \quad \alpha \in \{V, N, G\}.$$

Solutions to Equations (A.6) are of the form shown in Equation (2.4). By substituting the boundary conditions in Equations (2.3) into Equations (A.5) and (2.4), the following simultaneous equations define the system's constants of integration

$$C_1 = 0, \quad (\text{A.7a})$$

$$A_1 + A_2 - B_1 - B_2 = -\frac{m_i}{2\eta_V} e^{-\eta_V d_{\text{ER}}}, \quad (\text{A.7b})$$

$$B_1 e^{\eta_N d_N} + B_2 e^{-\eta_N d_N} - C_1 e^{-\eta_G d_N} = 0, \quad (\text{A.7c})$$

$$A_1 \hat{\xi}_{V_z} \eta_V - A_2 \hat{\xi}_{V_z} \eta_V - B_1 \hat{\xi}_{N_z} \eta_N + B_2 \hat{\xi}_{F_z} \eta_F = \frac{m_i \hat{\xi}_{V_z}}{2} e^{-\eta_V d_{\text{ER}}}, \quad (\text{A.7d})$$

$$B_1 \hat{\xi}_{N_z} \eta_N e^{\eta_N d_N} - B_2 \hat{\xi}_{N_z} \eta_N e^{-\eta_N d_N} + C_2 \hat{\xi}_{G_z} \eta_G e^{-\eta_G d_N} = 0, \quad (\text{A.7e})$$

$$A_1 \hat{\xi}_{V_z} \eta_V e^{-\eta_V (d_{\text{ER}} + d_{\text{EI}})} - A_2 \hat{\xi}_{V_z} \eta_V e^{\eta_V (d_{\text{ER}} + d_{\text{EI}})} = \frac{m_i \hat{\xi}_{V_z}}{2} e^{-\eta_V d_{\text{EI}}}. \quad (\text{A.7f})$$

Given the complexity of the resultant expressions, this set of equations was solved with the symbolic mathematics engine, Mathematica (Wolfram Research, Version 10). Note that the integration constants are obtained separately for each electrode. Owing to the model's linearity, multi-electrode simulations are implemented via the superposition of the electric field generated from several single-electrode simulations. In order to eliminate d_{EI} from the resulting solution, the right limit as d_{EI} goes to zero was also computed with Mathematica.

A.2 Admittivity of the nerve fiber layer

As shown in Table 2.1, the insulator, vitreous, and ganglion cell layers are described using scalar conductivities. This is equivalent to these layers having a spatially- and temporally-independent admittivity term that is also isotropic. Like the NFL, the GCL consists of active neural tissue, and hence would be expected to have some capacitive response to stimulation, which would lead to a nonzero permittivity. However, since the GCL is predominantly composed of cell somas, the volume proportion taken up by neural membrane is much lower than that of the NFL, which is predominantly composed of densely packed and thinner fibers. For this reason, the GCL, as for the insulator and vitreous layers, is assumed to have a constant conductance, and is considered to be isotropic:

$$\xi_{I_x} = \xi_{I_y} = \xi_{I_z} = \sigma_I, \quad (\text{A.8a})$$

$$\xi_{V_x} = \xi_{V_y} = \xi_{V_z} = \sigma_V, \quad (\text{A.8b})$$

$$\xi_{G_x} = \xi_{G_y} = \xi_{G_z} = \sigma_G. \quad (\text{A.8c})$$

The cellular composite model provides expressions for both the tissue admittivity kernel for the NFL and for the membrane potential of a neurite given the extracellular potential along its axis (irrespective of which layer the neurite is in) (Meffin et al., 2014).

We assume that in the NFL fibers are oriented in the y -direction. Then, in the time and space domains, the NFL admittivity is given by

$$\begin{aligned} \xi_N(y, t) &= \begin{bmatrix} \xi_{N_x} & 0 & 0 \\ 0 & \xi_{N_y} & 0 \\ 0 & 0 & \xi_{N_z} \end{bmatrix} \\ &= \begin{bmatrix} \xi_{N_T} & 0 & 0 \\ 0 & \xi_{N_L} & 0 \\ 0 & 0 & \xi_{N_T} \end{bmatrix}, \end{aligned} \quad (\text{A.9})$$

where

$$\xi_{\text{N}_T}(y, t) = \frac{2\pi d}{b\rho_e} \delta(y) \delta(t), \quad (\text{A.10a})$$

$$\xi_{\text{N}_L}(y, t) = \frac{2\pi \delta(y) \delta(t)}{\rho_i} - \frac{\sqrt{\pi} H(t) \tau_m^{\frac{3}{2}}}{4\rho_i \lambda_{0V} t^{\frac{5}{2}}} \left(\frac{2t}{\tau_m} - \frac{y^2}{\lambda_{0V}^2} \right) e^{-\frac{t}{\tau_m} - \frac{y^2 \tau_m}{4\lambda_{0V}^2 t}}, \quad (\text{A.10b})$$

in which ξ_{N_T} and ξ_{N_L} are the transverse and longitudinal components of the admittivity kernel, respectively, and H is the Heaviside step function. The remaining terms, d , b , ρ_e , ρ_i , τ_m , and λ_{0V} each represent different physical or electrical properties of the tissue and are defined in Table A.1. The above expressions reduce to the following by taking the Fourier transform with respect to y and t :

$$\hat{\xi}_{\text{N}_T}(k_y, \omega) = \frac{d}{b\rho_e}, \quad (\text{A.11a})$$

$$\hat{\xi}_{\text{N}_L}(k_y, \omega) = \frac{1}{\rho_i} \frac{1 + j\omega\tau_m + k_y^2 \lambda_{0J}^2}{1 + j\omega\tau_m + k_y^2 \lambda_{0V}^2}. \quad (\text{A.11b})$$

Specific layer admittivities and conductivities define the level of anisotropy for each layer as well as unique spatiotemporal dependencies. For the NFL and GCL, the form of the admittivity represents the assumed distribution of fiber orientations in each layer. The NFL is modeled as a parallel fiber bundle (anisotropic) and fibers in the GCL are modeled as having a uniform distribution of orientations (isotropic). Equations (A.10) describe a non-local, non-instantaneous admittivity for the NFL, which is derived from an accurate characterization of the spatiotemporal electrical properties of individual neurites that comprise the tissue (Meffin et al., 2014).

A.3 Generalisation of neurite equations

Owing to the high anisotropy of the NFL, fibers with different orientations, whether they are in the NFL or lower layers, will achieve different levels of activation given an applied stimulus. The volume conductor model derived above assumes that membrane potential is calculated for fibers with a single orientation: parallel to the NFL fiber bundle.

The model includes distinct stages for the calculation of extracellular potential (which captures the anisotropy of the NFL) and the calculation of membrane potential. As a result, it is possible to calculate the membrane potential for an unbranched axon with arbitrary morphology if we assume that the orientation of the axon of interest has a negligible effect on the distribution of extracellular potential as a whole. We can first calculate extracellular potential in the region of interest using Equations (2.4) and then sample it along the desired axon trajectory. This sampled extracellular potential can then be substituted into Equations (2.5) to yield the membrane potential along the axon.

Although this approach achieves a high level of generality, it requires an extra two computationally expensive Fourier transform calculations: one to convert the extracellular potential into the spatial domain from the spatial frequency domain to allow for sampling and another to convert the sampled data back into the frequency domain.

In the case of simple x - y rotations of straight axons, as assumed in the GCL, the required rotation can be performed directly in the Fourier domain, since rotation is preserved under unitary

Table A.1: Cellular composite model parameters for Chapter 2

Parameter	Description	Unit
a	Neurite radius	m
b	Outer cylinder radius	m
d	Width of extracellular sheath, $d = b - a$	m
ρ_i	Intracellular resistivity	$\Omega \text{ m}$
ρ_e	Extracellular resistivity	$\Omega \text{ m}$
r_i	Intracellular resistance per unit length, $r_i = \rho_i/(\pi a^2)$	Ω/m
r_e	Extracellular resistance per unit length, $r_e = \rho_e/(\pi(b^2 - a^2))$	Ω/m
R_m	Membrane unit area resistance	$\Omega \text{ m}^2$
r_m	Membrane unit length resistance, $r_m = R_m/(2\pi a)$	$\Omega \text{ m}$
C_m	Membrane capacitance per unit area	F/m^2
τ_m	Membrane time constant, $\tau_m = R_m C_m$	s
λ_{0J}	Static electrotonic length constant for current density boundary conditions, $\lambda_{0J} = \sqrt{r_m/(r_e + r_i)}$	m
$\lambda_J(\omega)$	Frequency-dependent electrotonic length constant for current density boundary conditions, $\lambda_J(\omega) = \lambda_{0J}/\sqrt{1 + j\omega\tau_m}$	m
λ_{0V}	Static electrotonic length constant for voltage boundary conditions, $\lambda_{0V} = \sqrt{r_m/r_i}$	m
$\lambda_V(\omega)$	Frequency-dependent electrotonic length constant for voltage boundary conditions, $\lambda_V(\omega) = \lambda_{0V}/\sqrt{1 + j\omega\tau_m}$	m
d_N	Nerve fiber layer thickness	m
d_{ER}	Electrode-retina separation distance	m
d_{EI}	Electrode-insulator separation, set to zero in final solution	m
q	Radius of disc electrodes	m
x_i, y_i, z_i	Location of center of electrode i	m

transformations such as the Fourier transform. Extracellular potential is thus calculated using a straightforward modification of Equations (2.4) in order to rotate the k_y -axis (and equivalently the y -axis to align with the desired fiber orientation:

$$\begin{aligned} \hat{\phi}_V(k'_x, k'_y, \omega; z) &= A_1 e^{-\eta_V(k'_x, k'_y)z} + A_2 e^{\eta_V(k'_x, k'_y)z} \\ &+ \sum_i^N \frac{m_i(k'_x, k'_y, \omega)}{2\eta_V(k'_x, k'_y)} e^{-\eta_V(k'_x, k'_y)|z-z_i|}, \end{aligned} \quad (\text{A.12a})$$

$$\hat{\phi}_N(k'_x, k'_y, \omega; z) = B_1 e^{-\eta_N(k'_x, k'_y)z} + B_2 e^{\eta_N(k'_x, k'_y)z}, \quad (\text{A.12b})$$

$$\hat{\phi}_G(k'_x, k'_y, \omega; z) = C_1 e^{-\eta_G(k'_x, k'_y)z} + C_2 e^{\eta_G(k'_x, k'_y)z}, \quad (\text{A.12c})$$

where, for a rotation in the x - y plane of θ , we use

$$\begin{bmatrix} k'_x \\ k'_y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} k_x \\ k_y \end{bmatrix}.$$

The corresponding rotated extracellular potential can then be applied directly to Equations (2.5).

A.4 Alternative non-ideal electrode array placement

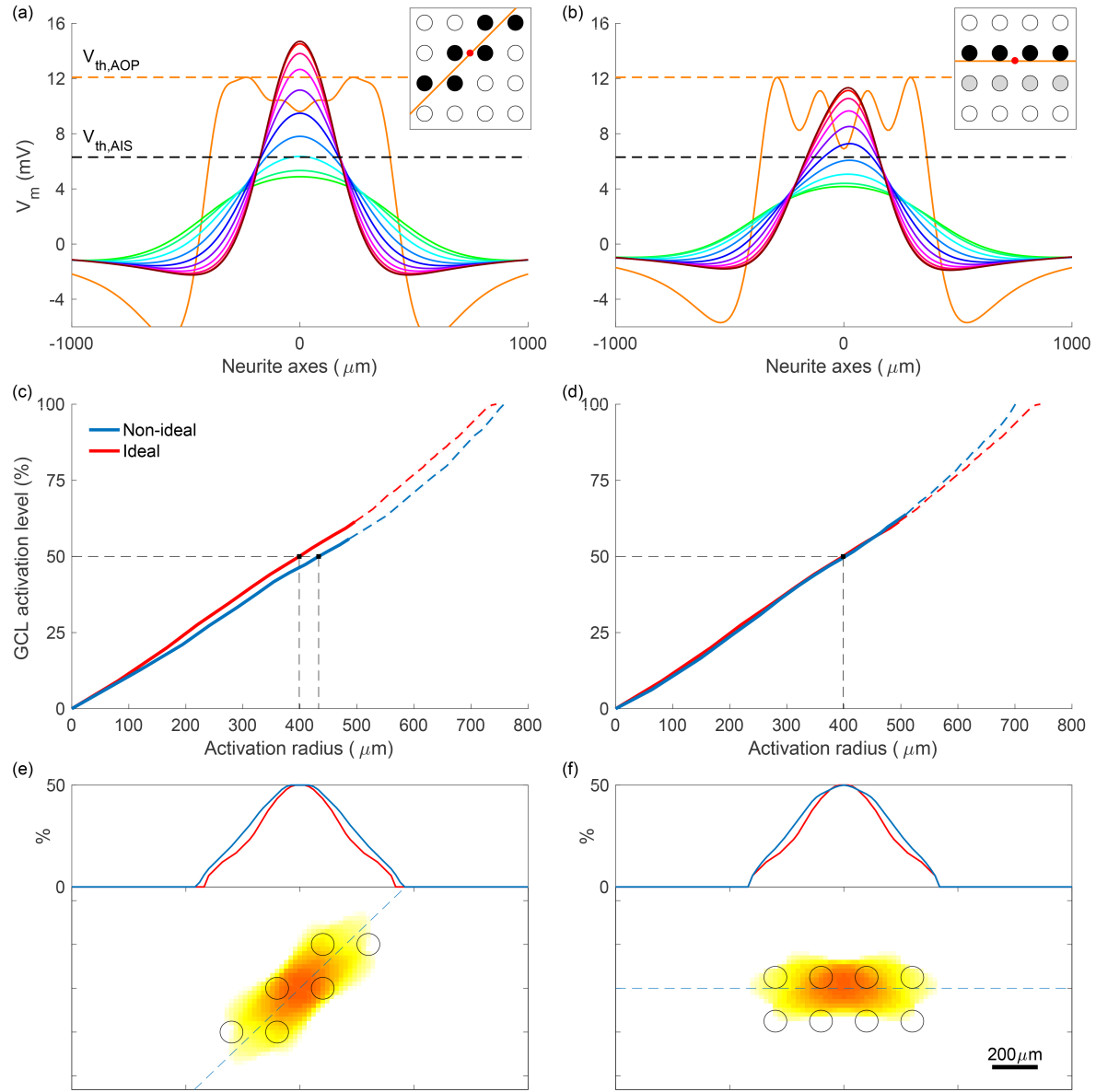


Figure A.1: **Alternative non-ideal electrode array placement.** (a)-(b) Membrane potential along neurite axes for axons of passage and axon initial segments, with stimulus current chosen to maximally activate initial segments without activating any passing axons. Colors correspond to those in Figure 2.6(a), with green parallel to axons of passage and brown perpendicular. Insets describe the geometry of each simulation, indicating target region (red), electrodes used (black), and the orientation of axons of passage (orange). (c)-(d) Ganglion cell layer activation level vs. activation radius for non-ideal and ideal (as in Figure 2.8(f)) geometries. Transitions from solid to dashed lines represent the transitions from axon initial segment to axon of passage preferential activation. (e)-(f) The spread of ganglion cell layer activation in the x - y plane. The dashed blue line corresponds to the one-dimensional inset. Colors are mapped according to the color bar in Figure 2.8. The left-hand panel (a, c, and e) shows stimulation with 6 electrodes, each with equal current. The right-hand panel (b, d, and f) shows stimulation of an off-center region of the GCL by halving the current delivered from the bottom row of electrodes. All simulations used a pulse phase duration of $200 \mu\text{s}$ and electrode-retina separation of $100 \mu\text{m}$.

Appendix B

Supporting Material for Chapter 3

B.1 Analytic derivation of weights for a cylinder

In the special case of a straight cylinder of infinite length, the passive (linear) membrane depolarization in response to one or many injected inputs can be determined analytically via solution of the linear cable equation. Here, we demonstrate the calculation of this depolarization for a single injected current and then for two injected currents.

The cable equation with a single current injection I_1 at location $x = x_1$ is given by

$$C_m \frac{\partial V_m(x, t)}{\partial t} - \frac{1}{R_a} \frac{\partial^2 V_m(x, t)}{\partial x^2} + \frac{V_m(x, t)}{R_m} = I_1(t) \delta(x - x_1), \quad (\text{B.1})$$

where C_m is the membrane capacitance, R_a is the axial resistance, R_m is the membrane resistance, V_m is the membrane voltage (relative to resting voltage), and $\delta(x - x_1)$ is the Dirac delta function.

By converting Equation B.1 into the temporal and spatial frequency domains using the Fourier transform, the solution can be given as

$$\hat{V}_m(k, \omega) = \frac{R_a e^{-j k x_1}}{k^2 + \alpha(\omega)^2} \hat{I}_1(\omega), \quad (\text{B.2})$$

where

$$\alpha(\omega)^2 = R_a C_m j \omega + \frac{R_a}{R_m}, \quad (\text{B.3})$$

and \hat{V}_m and \hat{I}_1 are the Fourier transforms of V_m and I_1 , respectively; and k and ω are the Fourier transform pairs of x and t , respectively.

For many current injections, at many different locations, the cable equation becomes

$$C_m \frac{\partial V_m(x, t)}{\partial t} - \frac{1}{R_a} \frac{\partial^2 V_m(x, t)}{\partial x^2} + \frac{V_m(x, t)}{R_m} = \sum_n I_n(t) \delta(x - x_n). \quad (\text{B.4})$$

Similarly, the Fourier domain solution for multiple injections is given by

$$\hat{V}_m(k, \omega) = \sum_n \frac{R_a e^{-jkx_n}}{k^2 + \alpha(\omega)^2} \hat{I}_n(\omega). \quad (\text{B.5})$$

Hence, the response to many simultaneous injections is equal to the sum of the responses to each injection, as can be seen by comparing Equations B.2 and B.5.

Equation B.5 gives as a similar expression to that of Equation 3.11 in Section 3.4.5, where here the weights are represented in the Fourier domain by

$$\hat{w}_n(k, \omega) = \frac{R_a e^{-jkx_n}}{k^2 + \alpha(\omega)^2}. \quad (\text{B.6})$$

The above expression for $w_n(k, \omega)$ depends on both location along the neurite/cell, x and time, t , via their Fourier domain representations. Note that in Equation 3.11, we consider the weights at a specific location (the site of action potential initiation) and at a specific point in time, reducing the set of weights, w_n , to a set of constants.

Appendix C

Supporting Material for Chapter 4

C.1 Number of local optima for biphasic stimulation

This appendix will demonstrate that potentially 2^M locally optimal solutions exists to the optimization problem defined in Equation 4.9 when applied to the inversion of the biphasic version of the linear-nonlinear model. For simplicity, this derivation will use a simplified version of the model given by Equation 4.4(b). Assuming that $c^+ = c^-$, the double sided sigmoid is approximately equivalent to:

$$Y' = \frac{1}{1 + \exp(-(|\vec{v}_1 \cdot \vec{S}_t| - c))}. \quad (\text{C.1})$$

The optimization/inversion problem can then be represented by the minimization of:

$$\mathcal{L}(\vec{S}_t) = -\frac{1}{M} \sum_m^M (1 - y_m) \log \left(1 - \frac{1}{1 + e^{-(|\vec{v}_{1,m} \cdot \vec{S}_t| - c)}} \right) + y_m \log \left(\frac{1}{1 + e^{-(|\vec{v}_{1,m} \cdot \vec{S}_t| - c)}} \right), \quad (\text{C.2})$$

where $\vec{v}_{1,m} = [w_{m,1}, w_{m,2}, \dots, w_{m,N}]$ is the estimated ERF at location m .

Taking the derivative with respect to \vec{S}_t gives the following:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \vec{S}_t} &= \frac{1}{M} \sum_m^M \begin{bmatrix} w_{m,1} \left(\frac{\vec{v}_{1,m} \cdot \vec{S}_t ((1-y_m)e^{|\vec{v}_{1,m} \cdot \vec{S}_t|} + y_m e^c)}{|\vec{v}_{1,m} \cdot \vec{S}_t| (e^{|\vec{v}_{1,m} \cdot \vec{S}_t|} + e^c)} \right) \\ \vdots \\ w_{m,N} \left(\frac{\vec{v}_{1,m} \cdot \vec{S}_t ((1-y_m)e^{|\vec{v}_{1,m} \cdot \vec{S}_t|} + y_m e^c)}{|\vec{v}_{1,m} \cdot \vec{S}_t| (e^{|\vec{v}_{1,m} \cdot \vec{S}_t|} + e^c)} \right) \end{bmatrix} \\ &= \frac{1}{M} \sum_m^M \left(\frac{\vec{v}_{1,m} \cdot \vec{S}_t ((1-y_m)e^{|\vec{v}_{1,m} \cdot \vec{S}_t|} + y_m e^c)}{|\vec{v}_{1,m} \cdot \vec{S}_t| (e^{|\vec{v}_{1,m} \cdot \vec{S}_t|} + e^c)} \right) \vec{v}_{1,m}^\top \end{aligned} \quad (\text{C.3})$$

To find the stationary points (local minima) from the above expression, we set $\frac{\partial \mathcal{L}}{\partial \vec{S}_t} = \vec{0}$, where $\vec{0}$ is a vector of zeros. Due to the existence of the absolute value operator, there are 2^M cases that must be considered in order to find all possible solutions. For each of these 2^M cases, a potential zero for

equation Equation C.3, and a potential local minima of Equation C.2, exists. For example, consider the case where

$$\vec{v}_{1,m} \cdot \vec{S}_t > 0 \quad \forall m \in \{1, 2, \dots, M\}. \quad (\text{C.4})$$

In this case, Equation C.3 becomes,

$$\frac{\partial \mathcal{L}}{\partial \vec{S}_t} = \frac{1}{M} \sum_m^M \left(\frac{(1 - y_m) e^{\vec{v}_{1,m} \cdot \vec{S}_t} + y_m e^c}{e^{\vec{v}_{1,m} \cdot \vec{S}_t} + e^c} \right) \vec{v}_{1,m}^\top \quad (\text{C.5})$$

It is beyond the scope of this work, and likely very involved, to find an expression for the 2^M potential local minima. In practice, although it is possible that, given a particular target image and set of ERFs, 2^M distinct minima exist, some of these may be degenerate, resulting in a fewer number of effective minima. Furthermore, from what has been observed throughout this work, the topology of the objective surface described by Equation C.2 is dominated by the number of stimulating electrodes, N , and not by the number of sampled locations at which the ERF is estimated, M . This can be seen in clearly in Figure 4.5(c), which shows four (2^N) clear minima in the objective surface instead of 2^M . It may be that less obvious local minima may be apparent if the surface is sampled with a much higher resolution. However, due to the smoothness of the surface, these are unlikely to have a lower level of loss than the dominant optima visible.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv*, 1603.04467.
- Abbott, C. J., Choe, T. E., Burgoyne, C. F., Cull, G., Wang, L., and Fortune, B. (2014). Comparison of retinal nerve fiber layer thickness in vivo and axonal transport after chronic intraocular pressure elevation in young versus older rats. *PLOS One*, 9(12):e114546.
- Abouelseoud, G., Abouelseoud, Y., Shoukry, A., Ismail, N., and Mekky, J. (2018). A mixed integer linear programming approach to electrical stimulation optimization problems. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 26(2):527–537.
- Abramian, M., Lovell, N. H., Morley, J. W., Suaning, G. J., and Dokos, S. (2011). Activation of retinal ganglion cells following epiretinal electrical stimulation with hexagonally arranged bipolar electrodes. *J. Neural Eng.*, 8(3):35004.
- Abramian, M., Lovell, N. H., Morley, J. W., Suaning, G. J., and Dokos, S. (2015). Activation and inhibition of retinal ganglion cells in response to epiretinal electrical stimulation: a computational modelling study. *J. Neural Eng.*, 12(1):016002.
- Addis, B., Cassioli, A., Locatelli, M., and Schoen, F. (2011). A global optimization method for the design of space trajectories. *Comput. Optim. Appl.*, 48(3):635–652.
- Ahuja, A. K., Dorn, J. D., Caspi, A., McMahon, M. J., Dagnelie, G., DaCruz, L., Stanga, P., Humayun, M. S., and Greenberg, R. J. (2011). Blind subjects implanted with the Argus II retinal prosthesis are able to improve performance in a spatial-motor task. *Br. J. Ophthalmol.*, 95(4):539–543.
- Allaire, J. and Chollet, F. (2017). keras: R Interface to 'Keras'. *The Comprehensive R Archive Network*.
- Allaire, J. and Tang, Y. (2017). tensorflow: R Interface to 'TensorFlow'. *The Comprehensive R Archive Network*.
- Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). NeuroMorpho.Org: A central resource for neuronal morphologies. *J. Neurosci.*, 27(35):9247–9251.
- Ayton, L. N., Blamey, P. J., Guymer, R. H., Luu, C. D., Nayagam, D. A. X., Sinclair, N. C., Shivdasani, M. N., Yeoh, J., McCombe, M. F., Briggs, R. J., Opie, N. L., Villalobos, J., Dimitrov, P. N., Varsamidis, M., Petoe, M. A., McCarthy, C. D., Walker, J. G., Barnes, N., Burkitt, A. N., Williams,

- C. E., Shepherd, R. K., and Allen, P. J. (2014). First-in-human trial of a novel suprachoroidal retinal prosthesis. *PLOS One*, 9(12):e115239.
- Baccus, S. A., Olveczky, B. P., Manu, M., and Meister, M. (2008). A retinal circuit that computes object motion. *J. Neurosci.*, 28(27):6807–6817.
- Badea, T. C. and Nathans, J. (2011). Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: Analysis of wild type and mutant cells using genetically-directed sparse labeling. *Vision Res.*, 51(2):269–279.
- Bahoura, M. and Park, C. W. (2011). FPGA-implementation of high-speed MLP neural network. In *2011 18th IEEE Int. Conf. Electron. Circuits, Syst.*, pages 426–429.
- Bakkum, D. J., Frey, U., Radivojevic, M., Russell, T. L., Müller, J., Fiscella, M., Takahashi, H., and Hierlemann, A. (2013). Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nat. Commun.*, 4:2181.
- Behrend, M. R., Ahuja, A. K., and Weiland, J. D. (2008). Dynamic current density of the disk electrode double-layer. *IEEE Trans. Biomed. Eng.*, 55(3):1056–1062.
- Benav, H., Bartz-Schmidt, K. U., Besch, D., Bruckmann, A., Gekeler, F., Greppmaier, U., Harscher, A., Kibbel, S., Kusnyerik, A., Peters, T., Sachs, H., Stett, A., Stingl, K., Wilhelm, B., Wilke, R., Wrobel, W., and Zrenner, E. (2010). Restoration of useful vision up to letter recognition capabilities using subretinal microphotodiodes. In *32nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 5919–5922.
- Bengio, Y., Goodfellow, I. J., and Courville, A. (2015). Optimization for training deep models. *Deep Learn.*, pages 238–290.
- Berry, M. J., Brivanlou, I. H., Jordan, T. A., and Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398(6725):334–338.
- Boinagrov, D., Pangratz-Fuehrer, S., Goetz, G., and Palanker, D. (2014). Selectivity of direct and network-mediated stimulation of the retinal ganglion cells with epi-, sub- and intraretinal electrodes. *J. Neural Eng.*, 11(2):026008.
- Brummer, S. B. and Turner, M. J. (1977). Electrical stimulation with Pt electrodes: II-estimation of maximum surface redox (theoretical non-gassing) limits. *IEEE Trans. Biomed. Eng.*, 24(5):440–443.
- Cao, X., Sui, X., Lyu, Q., Li, L., and Chai, X. (2015). Effects of different three-dimensional electrodes on epiretinal electrical stimulation by modeling analysis. *J. Neuroeng. Rehabil.*, 12(1):73.
- Chen, S. C., Suaning, G. J., Morley, J. W., and Lovell, N. H. (2009). Simulating prosthetic vision: I. Visual models of phosphenes. *Vision Res.*, 49(12):1493–1506.
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, 12(2):199–213.
- Chollet, F. (2015). Keras. *Github*.
- Coburn, B. (1989). Neural modeling in electrical stimulation. *Crit. Rev. Biomed. Eng.*, 17(2):133–78.
- Colbert, C. M. and Johnston, D. (1996). Axonal action-potential initiation and Na⁺ channel densities in the soma and axon initial segment of subicular pyramidal neurons. *J. Neurosci.*, 16(21):6676–86.
- Coleman, T. F. and Li, Y. (1996). A reflective newton method for minimizing a quadratic function subject to bounds on some of the variables. *SIAM J. Optim.*, 6(4):1040–1058.
- Coombs, J., van der List, D., Wang, G.-Y. Y., and Chalupa, L. M. (2006). Morphological properties of mouse retinal ganglion cells. *Neuroscience*, 140(1):123–136.

- Coombs, J. S., Curtis, D. R., and Eccles, J. C. (1957). The generation of impulses in motoneurones. *J. Physiol.*, 139(2):232–249.
- Cuntz, H., Forstner, F., Borst, A., Häusser, M., Kuhlman, S., and Saggau, P. (2010). One rule to grow them all: A general theory of neuronal branching and its practical application. *PLOS Comput. Biol.*, 6(8):e1000877.
- Curlander, J. C. and Marmarelis, V. Z. (1987). A linear spatio-temporal model of the light-to-bipolar cell system and its response characteristics to moving bars. *Biol. Cybern.*, 57(6):357–363.
- da Cruz, L., Coley, B. F., Dorn, J., Merlini, F., Filley, E., Christopher, P., Chen, F. K., Wuyyuru, V., Sahel, J., Stanga, P., Humayun, M. S., Greenberg, R. J., and Dagnelie, G. (2013). The Argus II epiretinal prosthesis system allows letter and word reading and long-term function in patients with profound vision loss. *Br. J. Ophthalmol.*, 97(5):632–636.
- de Ruyter van Steveninck, R. and Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. *Proc. R. Soc. B Biol. Sci.*, 234(1277):379–414.
- Dmochowski, J. P., Datta, A., Bikson, M., Su, Y., and Parra, L. C. (2011). Optimized multi-electrode stimulation increases focality and intensity at target. *J. Neural Eng.*, 8(4):046011.
- Dokos, S., Suanning, G. J., and Lovell, N. H. (2005). A bidomain model of epiretinal stimulation. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 13(2):137–146.
- Dowle, M. and Srinivasan, A. (2017). *data.table: Extension of ‘data.frame’*.
- Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.*, 12:2121–2159.
- Dumm, G., Fallon, J. B., Williams, C. E., and Shivdasani, M. N. (2014). Virtual electrodes by current steering in retinal prostheses. *Invest. Ophthalmol. Vis. Sci.*, 55(12):8077–8085.
- Eiber, C. D., Dokos, S., Lovell, N. H., and Suanning, G. J. (2016). A spectral element method with adaptive segmentation for accurately simulating extracellular electrical stimulation of neurons. *Med. Biol. Eng. Comput.*, 55(5):823–831.
- Eickenscheidt, M. and Zeck, G. (2014). Action potentials in retinal ganglion cells are initiated at the site of maximal curvature of the extracellular potential. *J. Neural Eng.*, 11(3):036006.
- Escola, S., Fontanini, A., Katz, D., and Paninski, L. (2011). Hidden markov models for the stimulus-response relationships of multistate neural systems. *Neural Comput.*, 23(5):1071–1132.
- Esler, T. B., Kerr, R. R., Tahayori, B., Grayden, D. B., Meffin, H., and Burkitt, A. N. (2018a). Minimizing activation of overlying axons with epiretinal stimulation: The role of fiber orientation and electrode configuration. *PLOS One*, 13(3):e0193598.
- Esler, T. B., Maturana, M. I., Kerr, R. R., Grayden, D. B., Burkitt, A. N., and Meffin, H. (2018b). Biophysical basis of the linear electrical receptive fields of retinal ganglion cells. *J. Neural Eng.*, 15(5):055001.
- FitzGibbon, T. and Taylor, S. F. (2012). Mean retinal ganglion cell axon diameter varies with location in the human retina. *Jpn. J. Ophthalmol.*, 56(6):631–637.
- Fletcher, L. N., Coimbra, J. P., Rodger, J., Potter, I. C., Gill, H. S., Dunlop, S. A., and Collin, S. P. (2014). Classification of retinal ganglion cells in the southern hemisphere lamprey *Geotria australis* (Cyclostomata). *J. Comp. Neurol.*, 522(4):750–771.

- Fohlmeister, J. F., Cohen, E. D., and Newman, E. A. (2010). Mechanisms and distribution of ion channels in retinal ganglion cells: Using temperature as an independent variable. *J. Neurophysiol.*, 103(3):1357–1374.
- Fohlmeister, J. F. and Miller, R. F. (1997). Impulse encoding mechanisms of ganglion cells in the tiger salamander retina. *J. Neurophysiol.*, 78(4):1935–1947.
- Freeman, D. K., Eddington, D. K., Rizzo, J. F., and Fried, S. I. (2010). Selective activation of neuronal targets with sinusoidal electric stimulation. *J. Neurophysiol.*, 104(5):2778–2791.
- Frenkel, S., Morgan, J. E., and Blumenthal, E. Z. (2005). Histological measurement of retinal nerve fibre layer thickness. *Eye*, 19(5):491–498.
- Fried, S. I., Lasker, A. C. W., Desai, N. J., Eddington, D. K., and Rizzo, J. F. (2009). Axonal sodium-channel bands shape the response to electric stimulation in retinal ganglion cells. *J. Neurophysiol.*, 101(4):1972–1987.
- Fujikado, T., Kamei, M., Sakaguchi, H., Kanda, H., Morimoto, T., Ikuno, Y., Nishida, K., Kishima, H., Maruo, T., Konomi, K., Ozawa, M., and Nishida, K. (2011). Testing of semichronically implanted retinal prosthesis by suprachoroidal-transretinal stimulation in patients with retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.*, 52(7):4726–4733.
- Gabriel, C., Peyman, A., and Grant, E. H. (2009). Electrical conductivity of tissue at frequencies below 1MHz. *Phys. Med. Biol.*, 54(16):4863–4878.
- Gerwinn, S. (2010). Bayesian inference for generalized linear models for spiking neurons. *Front. Comput. Neurosci.*, 4:12.
- Ghezzi, D. (2015). Retinal prostheses: progress toward the next generation implants. *Front. Neurosci.*, 9:290.
- Gironés, R. G., Palero, R. C., Boluda, J. C., and Cortés, A. S. (2005). FPGA implementation of a pipelined on-line backpropagation. *J. VLSI Signal Process. Syst. Signal Image. Video Technol.*, 40(2):189–213.
- Gleixner, R. and Fromherz, P. (2006). The extracellular electrical resistivity in cell adhesion. *Biophys. J.*, 90(7):2600–2611.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- Greenberg, R., Velte, T., Humayun, M., Scarlatis, G., and De Juan, E. (1999). A computational model of electrical stimulation of the retinal ganglion cell. *IEEE Trans. Biomed. Eng.*, 46(5):505–514.
- Grosberg, L. E., Ganesan, K., Goetz, G. A., Madugula, S. S., Bhaskhar, N., Fan, V., Li, P., Hottowy, P., Dabrowski, W., Sher, A., Litke, A. M., Mitra, S., and Chichilnisky, E. (2017). Activation of ganglion cells and axon bundles using epiretinal electrical stimulation. *J. Neurophysiol.*, 118(3):1457–1471.
- Gui-Jing, C., Lai, T. L., and Wei, C. Z. (1981). Convergence systems and strong consistency of least squares estimates in regression models. *J. Multivar. Anal.*, 11(3):319–333.
- Guo, T., Tsai, D., Bai, S., Morley, J. W., Suaning, G. J., Lovell, N. H., and Dokos, S. (2014). Understanding the retina: a review of computational models of the retina from the single cell to the network level. *Crit. Rev. Biomed. Eng.*, 42(5):419–436.
- Guo, T., Yang, C. Y., Tsai, D., Muralidharan, M., Suaning, G. J., Morley, J. W., Dokos, S., and Lovell, N. H. (2018). Closed-loop efficient searching of optimal electrical stimulation parameters for preferential excitation of retinal ganglion cells. *Front. Neurosci.*, 12:168.

- Habib, A. G., Cameron, M. A., Suaning, G. J., Lovell, N. H., and Morley, J. W. (2013). Spatially restricted electrical activation of retinal ganglion cells in the rabbit retina by hexapolar electrode return configuration. *J. Neural Eng.*, 10(3):036013.
- Hadjinicolaou, A. E., Meffin, H., Maturana, M. I., Cloherty, S. L., and Ibbotson, M. R. (2015). Prosthetic vision: Devices, patient outcomes and retinal research. *Clin. Exp. Optom.*, 98(5):395–410.
- Halupka, K. (2017). *Prediction and shaping of visual cortex activity for retinal prostheses*. PhD thesis, The University of Melbourne.
- Halupka, K. J., Abbott, C. J., Wong, Y. T., Cloherty, S. L., Grayden, D. B., Burkitt, A. N., Sergeev, E. N., Luu, C. D., Brandli, A., Allen, P. J., Meffin, H., and Shivdasani, M. N. (2017). Neural responses to multielectrode stimulation of healthy and degenerate retina. *Invest. Ophthalmol. Vis. Sci.*, 58(9):3770–3784.
- Halupka, K. J., Shivdasani, M. N., Cloherty, S. L., Grayden, D. B., Wong, Y. T., Burkitt, A. N., and Meffin, H. (2016). Prediction of cortical responses to simultaneous electrical stimulation of the retina. *J. Neural Eng.*, 14(1).
- Hamer, R. D., Nicholas, S. C., Tranchina, D., Lamb, T. D., and Jarvinen, J. L. P. (2005). Toward a unified model of vertebrate rod phototransduction. *Vis. Neurosci.*, 22(4):417–36.
- Hines, M. L. and Carnevale, N. T. (1997). The NEURON simulation environment. *Neural Comput.*, 9(6):1179–1209.
- Hodgkin, A. and Huxley, A. (1990). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Bull. Math. Biol.*, 52(1-2):25–71.
- Hoerl, A. E. and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67.
- Hollyfield, J. G., Anderson, R. E., and LaVail, M. M., editors (2006). *Retinal Degenerative Diseases*, volume 572 of *Advances in Experimental Medicine and Biology*. Springer US, Boston.
- Hood, D. C., Shady, S., and Birch, D. G. (1993). Heterogeneity in retinal disease and the computational model of the human-rod response. *J. Opt. Soc. Am.*, 10(7):1624–1630.
- Hornig, R., Dapper, M., Le Joliff, E., Hill, R., Ishaque, K., Posch, C., Benosman, R., LeMer, Y., Sahel, J.-A., and Picaud, S. (2017). Pixium Vision: First Clinical Results and Innovative Developments. In *Artificial Vision*, pages 99–113. Springer International Publishing, Cham.
- Hornig, R., Zehnder, T., Velikay-Parel, M., Laube, T., Feucht, M., and Richard, G. (2007). The IMI retinal implant system. In *Artif. sight basic Res. Biomed. Eng. Clin. Adv.*, pages 111–128. Springer.
- Horsager, A., Greenberg, R. J., and Fine, I. (2010). Spatiotemporal interactions in retinal prosthesis subjects. *Invest. Ophthalmol. Vis. Sci.*, 51(2):1223–1233.
- Hottowy, P., Skoczeń, A., Gunning, D. E., Kachiguine, S., Mathieson, K., Sher, A., Wiacek, P., Litke, A. M., and Dabrowski, W. (2012). Properties and application of a multichannel integrated circuit for low-artifact, patterned electrical stimulation of neural tissue. *J. Neural Eng.*, 9(6):066005.
- Humayun, M. S., de Juan, E., Dagnelie, G., Greenberg, R. J., Propst, R. H., and Phillips, D. H. (1996). Visual perception elicited by electrical stimulation of retina in blind humans. *Arch. Ophthalmol.*, 114(1):40–46.
- Humayun, M. S., De Juan, E., Weiland, J. D., Dagnelie, G., Katona, S., Greenberg, R. J., and Suzuki, S. (1999). Pattern electrical stimulation of the human retina. *Vision Res.*, 39(15):2569–2576.

- Humayun, M. S., Dorn, J. D., da Cruz, L., Dagnelie, G., Sahel, J.-A., Stanga, P. E., Cideciyan, A. V., Duncan, J. L., Elliott, D., Filley, E., Ho, A. C., Santos, A., Safran, A. B., Ardit, A., Del Priore, L. V., and Greenberg, R. J. (2012). Interim results from the international trial of second sight's visual prosthesis. *Ophthalmol.*, 119(4):779–788.
- Jeng, J., Tang, S., Molnar, A., Desai, N. J., and Fried, S. I. (2011). The sodium channel band shapes the response to electric stimulation in retinal ganglion cells. *J. Neural Eng.*, 8(3):036022.
- Jensen, R. J. and Rizzo, J. F. (2006). Thresholds for activation of rabbit retinal ganglion cells with a subretinal electrode. *Exp. Eye Res.*, 83(2):367–373.
- Jensen, R. J., Rizzo, J. F., Ziv, O. R., Grumet, A. E., and Wyatt, J. (2003). Thresholds for activation of rabbit retinal ganglion cells with an ultrafine, extracellular microelectrode. *Invest. Ophthalmol. Vis. Sci.*, 44(8):3533–3543.
- Jepson, L. H., Hottowy, P., Mathieson, K., Gunning, D. E., Dabrowski, W., Litke, A. M., and Chichilnisky, E. J. (2013). Focal electrical stimulation of major ganglion cell types in the primate retina for the design of visual prostheses. *J. Neurosci.*, 33(17):7194–7205.
- Jepson, L. H., Hottowy, P., Mathieson, K., Gunning, D. E., Dabrowski, W., Litke, A. M., and Chichilnisky, E. J. (2014). Spatially patterned electrical stimulation to enhance resolution of retinal prostheses. *J. Neurosci.*, 34(14):4871–4881.
- Jonas, J. B., Schmidt, a. M., Mullerbergh, J. a., Schlotzschrehardt, U. M., and Naumann, G. O. H. (1992). Human optic-nerve fiber count and optic disk size. *Invest. Ophthalmol. Vis. Sci.*, 33(6):2012–2018.
- Kameneva, T., Meffin, H., and Burkitt, A. N. (2011). Modelling intrinsic electrophysiological properties of ON and OFF retinal ganglion cells. *J. Comput. Neurosci.*, 31(3):547–561.
- Karwoski, C. J., Frambach, D. A., and Proenza, L. M. (1985). Laminar profile of resistivity in frog retina. *J. Neurophysiol.*, 54(6):1607–1619.
- Kasi, H., Meissner, R., Babalian, A., Lintel, H. V., Bertsch, A., and Renaud, P. (2011). Direct localised measurement of electrical resistivity profile in rat and embryonic chick retinas using a microprobe. *J. Electr. Bioimpedance*, 1(1):84–92.
- Keat, J., Reinagel, P., Reid, R., and Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron*, 30(3):803–817.
- Kingma, D. P. and Ba, J. L. (2015). Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent. 2015*, pages 1–15.
- Kong, J.-H. H., Fish, D. R., Rockhill, R. L., and Masland, R. H. (2005). Diversity of ganglion cells in the mouse retina: Unsupervised morphological classification and its limits. *J. Comp. Neurol.*, 489(3):293–310.
- Latikka, J., Kuurne, T., and Eskola, H. (2001). Conductivity of living intracranial tissues. *Phys. Med. Biol.*, 46(6):1611–1616.
- LeCun, Y. and Bengio, Y. (1995). Convolutional networks for images, speech, and time-Series. In *Handb. brain theory neural networks*, pages 255–258. MIT Press, Cambridge.
- LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- Lee, S. and Stevens, C. F. (2007). General design principle for scalable neural circuits in a vertebrate retina. *Proc. Natl. Acad. Sci. U.S.A.*, 104(31):12931–12935.
- Loizos, K., Lazzi, G., Marc, R., and Cela, C. (2016). Virtual electrode design for increasing spatial resolution in retinal prosthesis. *Healthc. Technol. Lett.*, 3(2):93–97.

- Luo, Y. H. L. and da Cruz, L. (2016). The Argus® II retinal prosthesis system. *Prog. Retin. Eye Res.*, 50:89–107.
- Malmivuo, J. and Plonsey, R. (2012). *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*, volume 20. Oxford University Press.
- Matteucci, P. B., Chen, S. C., Tsai, D., Dodds, C. W. D., Dokos, S., Morley, J. W., Lovell, N. H., and Suaning, G. J. (2013). Current steering in retinal stimulation via a quasimonopolar stimulation paradigm. *Invest. Ophthalmol. Vis. Sci.*, 54(6):4307–4320.
- Maturana, M. I., Apollo, N. V., Garrett, D. J., Kameneva, T., Cloherty, S. L., Grayden, D. B., Burkitt, A. N., Ibbotson, M. R., and Meffin, H. (2018). Electrical receptive fields of retinal ganglion cells: Influence of presynaptic neurons. *PLOS Comput. Biol.*, 14(2):e1005997.
- Maturana, M. I., Apollo, N. V., Hadjinicolaou, A. E., Garrett, D. J., Cloherty, S. L., Kameneva, T., Grayden, D. B., Ibbotson, M. R., and Meffin, H. (2016). A simple and accurate model to predict responses to multi-electrode stimulation in the retina. *PLOS Comput. Biol.*, 12(4):e1004849.
- Maturana, M. I., Kameneva, T., Burkitt, A. N., Meffin, H., and Grayden, D. B. (2014). The effect of morphology upon electrophysiological responses of retinal ganglion cells: simulation results. *J. Comput. Neurosci.*, 36(2):157–175.
- Mazzoleni, A. P., Siskin, B. F., and Kahler, R. L. (1986). Conductivity values of tissue culture medium from 20C to 40C. *Bioelectromagnetics*, 7(1):95–99.
- Mazzoni, F., Novelli, E., and Strettoi, E. (2008). Retinal ganglion cells survive and maintain normal dendritic morphology in a mouse model of inherited photoreceptor degeneration. *J. Neurosci.*, 28(52):14282–14292.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, second edition.
- Meffin, H., Tahayori, B., Grayden, D. B., and Burkitt, A. N. (2012). Modeling extracellular electrical stimulation: I. Derivation and interpretation of neurite equations. *J. Neural Eng.*, 9(6):065005.
- Meffin, H., Tahayori, B., Grayden, D. B., and Burkitt, A. N. (2013a). Internal inconsistencies in models of electrical stimulation in neural tissue. In *35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 5946–5949.
- Meffin, H., Tahayori, B., Greene, E., Grayden, D. B., and Burkitt, A. N. (2013b). Spatial shaping of neural activity using electrical stimulation. *BMC Neurosci.*, 14(Suppl 1):P49.
- Meffin, H., Tahayori, B., Sergeev, E. N., Mareels, I. M. Y., Grayden, D. B., and Burkitt, A. N. (2014). Modelling extracellular electrical stimulation: III. Derivation and interpretation of neural tissue equations. *J. Neural Eng.*, 11(6):065004.
- Nanduri, D., Humayun, M. S., Greenberg, R. J., McMahon, M. J., and Weiland, J. D. (2008). Retinal prosthesis phosphene shape analysis. In *30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 1785–1788.
- Ng, A. Y. (2004). Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *21st Int. Conf. Mach. Learn.*, volume 2004, page 78.
- Oishi, A., Otani, A., Sasahara, M., Kurimoto, M., Nakamura, H., Kojima, H., and Yoshimura, N. (2009). Retinal nerve fiber layer thickness in patients with retinitis pigmentosa. *Eye*, 23(3):561–566.
- Opie, N. L., Lovell, N. H., Suaning, G. J., Preston, P., and Dokos, S. (2013). Current steering for high resolution retinal implants. In *37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 2760–2763.

- Pillow, J. W. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J. Neurosci.*, 25(47):11003–11013.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.
- Piyathaisere, D. V., Margalit, E., Chen, S.-J., Shyu, J.-S., D’Anna, S. A., Weiland, J. D., Grebe, R. R., Grebe, L., Fujii, G., Kim, S. Y., Greenberg, R. J., De Juan, E. J., and Humayun, M. S. (2003). Heat effects on the retina. *Ophthalm. Surg. Lasers Imaging*, 34(2):114–120.
- Poria, D. and Dhingra, N. K. (2015). Spontaneous oscillatory activity in rd1 mouse retina is transferred from ON pathway to OFF pathway via glycinergic synapse. *J. Neurophysiol.*, 113(2):420–425.
- Potra, F. A. and Wright, S. J. (2000). Interior-point methods. *J. Comput. Appl. Math.*, 124(1-2):281–302.
- Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151.
- R Core Team (2017). R: A Language and Environment for Statistical Computing.
- Radivojevic, M., Jäckel, D., Altermatt, M., Müller, J., Viswam, V., Hierlemann, A., and Bakkum, D. J. (2016). Electrical identification and selective microstimulation of neuronal compartments based on features of extracellular action potentials. *Sci. Rep.*, 6(1):31332.
- Ranck, J. (1963). Specific impedance of rabbit cerebral cortex. *Exp Neurol*, 7(2):144–152.
- Rattay, F. (1986). Analysis of models for external stimulation of axons. *IEEE Trans. Biomed. Eng.*, 33(10):974–977.
- Rattay, F. (1989). Analysis of models for extracellular fiber stimulation. *IEEE Trans. Biomed. Eng.*, 36(7):676–682.
- Rattay, F. (1999). The basic mechanism for the electrical stimulation of the nervous system. *Neuroscience*, 89(2):335–346.
- Rattay, F., Bassereh, H., and Fellner, A. (2017). Impact of electrode position on the elicitation of sodium spikes in retinal bipolar cells. *Sci. Rep.*, 7(1):17590.
- Rattay, F. and Resatz, S. (2004). Effective electrode configuration for selective stimulation with inner eye prostheses. *IEEE Trans. Biomed. Eng.*, 51(9):1659–1664.
- Remington, L. A. (2005). Clinical Anatomy of the Visual System. In *Clin. Anat. Vis. Syst.*, chapter Retina, pages 55–86. Elsevier, Philadelphia.
- Retina Reference (2016). Normal Retinal Anatomy, url: <http://retinareference.com/anatomy/>.
- Rizzo, J. F., Wyatt, J., Loewenstein, J., Kelly, S., and Shire, D. (2003a). Methods and perceptual thresholds for short-term electrical stimulation of human retina with microelectrode arrays. *Invest. Ophthalmol. Vis. Sci.*, 44(12):5355–5361.
- Rizzo, J. F., Wyatt, J., Loewenstein, J., Kelly, S., and Shire, D. (2003b). Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials. *Invest. Ophthalmol. Vis. Sci.*, 44(12):5362–5369.
- Rodger, J., Drummond, E. S., Hellström, M., Robertson, D., and Harvey, A. R. (2012). Long-term gene therapy causes transgene-specific changes in the morphology of regenerating retinal ganglion cells. *PLOS One*, 7(2):e31061.

- Rodriguez, A., Ehlenberger, D. B., Dickstein, D. L., Hof, P. R., and Wearne, S. L. (2008). Automated three-dimensional detection and shape classification of dendritic spines from fluorescence microscopy images. *PLOS One*, 3(4):e1997.
- Roessler, G., Laube, T., Brockmann, C., Kirschkamp, T., Mazinani, B., Goertz, M., Koch, C., Krisch, I., Sellhaus, B., Trieu, H. K., Weis, J., Bornfeld, N., R  thgen, H., Messner, A., Mokwa, W., and Walter, P. (2009). Implantation and explantation of a wireless epiretinal retina implant device: Observations during the EPIRET3 prospective clinical trial. *Invest. Ophthalmol. Vis. Sci.*, 50(6):3003–3008.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv*, 1609.04747.
- Sanes, J. R. and Zipursky, S. L. (2010). Design principles of insect and vertebrate visual systems. *Neuron*, 66(1):15–36.
- Savich, A., Moussa, M., and Areibi, S. (2012). A scalable pipelined architecture for real-time computation of MLP-BP neural networks. *Microprocess. Microsyst.*, 36(2):138–150.
- Schiefer, M. A. and Grill, W. M. (2006). Sites of neuronal excitation by epiretinal electrical stimulation. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 14(1):5–13.
- Schwartz, O., Chichilnisky, E. J., and Simoncelli, E. P. (2002). Characterizing neural gain control using spike-triggered covariance. *Adv. Neural Inf. Process. Syst.* 14, 14:269–276.
- Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *J. Vis.*, 6(4):13.
- Scrucca, L. (2013). GA: A Package for Genetic Algorithms in R. *J. Stat. Softw.*, 53(4):1–37.
- Sekhar, S., Jalligampala, A., Zrenner, E., and Rathbun, D. L. (2016). Tickling the retina: Integration of subthreshold electrical pulses can activate retinal neurons. *J. Neural Eng.*, 13(4):046004.
- Sekirnjak, C., Hottowy, P., Sher, A., Dabrowski, W., Litke, A. M., and Chichilnisky, E. J. (2008). High-resolution electrical stimulation of primate retina for epiretinal implant design. *J. Neurosci.*, 28(17):4446–4456.
- Semenovich, D., Cai, Y., and Heppell, I. (2010). Convex optimization: a new approach to common challenges in premium rating. Technical report, Institute of Actuaries of Australia.
- Shalhaf, F., Du, P., Lovell, N. H., Dokos, S., and Vaghefi, E. (2015). A 3D-continuum bidomain model of retinal electrical stimulation using an anatomically detailed mesh. In *37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, volume 2015-Novem, pages 2291–2294. IEEE.
- Sheasby, B. W. and Fohlmeister, J. F. (1999). Impulse encoding across the dendritic morphologies of retinal ganglion cells. *J. Neurophysiol.*, 81(4):1685–1698.
- Shivdasani, M. N., Fallon, J. B., Luu, C. D., Cicione, R., Allen, P. J., Morley, J. W., and Williams, C. E. (2012). Visual cortex responses to single- and simultaneous multiple-electrode stimulation of the retina: Implications for retinal prostheses. *Invest. Ophthalmol. Vis. Sci.*, 53(10):6291–6300.
- Sivanandam, S. N. and Deepa, S. N. (2008). *Introduction to genetic algorithms*. Springer-Verlag, Berlin, Heidelberg.
- Snider, J., Pillai, A., and Stevens, C. F. (2010). A universal property of axonal and dendritic arbors. *Neuron*, 66(1):45–56.
- Spencer, T. C. (2018). Creating virtual electrodes with two-dimensional current steering. *J. Neural Eng.*, 15(3):1–35.

- Spencer, T. C., Fallon, J. B., Thien, P. C., and Shivdasani, M. N. (2016). Spatial restriction of neural activation using focused multipolar stimulation with a retinal prosthesis. *Invest. Ophthalmol. Vis. Sci.*, 57(7):3181–3191.
- Stingl, K., Bartz-Schmidt, K. U., Besch, D., Braun, A., Bruckmann, A., Gekeler, F., Greppmaier, U., Hipp, S., Hörtdörfer, G., Kernstock, C., Koitschev, A., Kusnyerik, A., Sachs, H., Schatz, A., Stingl, K. T., Peters, T., Wilhelm, B., and Zrenner, E. (2013). Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS. *Proc. R. Soc. B Biol. Sci.*, 280(1757):20130077.
- Stuart, G. and Spruston, N. (1998). Determinants of voltage attenuation in neocortical pyramidal neuron dendrites. *J. Neurosci.*, 18(10):3501–3510.
- Stuart, G. J. and Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, 367:69–72.
- Sümbül, U., Song, S., McCulloch, K., Becker, M., Lin, B., Sanes, J. R., Masland, R. H., and Seung, H. S. (2014). A genetic and computational approach to structurally classify neuronal types. *Nat. Commun.*, 5:3512.
- Syková, E. and Nicholson, C. (2008). Diffusion in brain extracellular space. *Physiol. Rev.*, 88(4):1277–1340.
- Tahayori, B., Meffin, H., Dokos, S., Burkitt, A. N., and Grayden, D. B. (2012). Modeling extracellular electrical stimulation: II. Computational validation and numerical results. *J. Neural Eng.*, 9(6):065006.
- Tahayori, B., Meffin, H., Sergeev, E. N., Mareels, I. M. Y., Burkitt, A. N., and Grayden, D. B. (2014). Modelling extracellular electrical stimulation: IV. Effect of the cellular composition of neural tissue on its spatio-temporal filtering properties. *J. Neural Eng.*, 11(6):065005.
- Teeters, J., Jacobs, A., and Werblin, F. (1997). How neural interactions form neural responses in the salamander retina. *J. Comput. Neurosci.*, 4(1):5–27.
- Thomson, K. L., Yeo, J. M., Waddell, B., Cameron, J. R., and Pal, S. (2015). A systematic review and meta-analysis of retinal nerve fiber layer change in dementia, using optical coherence tomography. *Alzheimer’s Dement. Diagnosis, Assess. Dis. Monit.*, 1(2):136–143.
- Tikhonov, A. N., Goncharsky, A. V., Stepanov, V. V., and Yagola, A. G. (1995). Numerical methods for the approximate solution of ill-posed problems on compact sets. *Numer. Methods Solut. Ill-Posed Probl.*, 32(144):65–79.
- Toris, C. B., Eiesland, J. L., and Miller, R. F. (1995). Morphology of ganglion cells in the neonatal tiger salamander retina. *J. Comp. Neurol.*, 352(4):535–559.
- Touryan, J., Lau, B., and Dan, Y. (2002). Isolation of relevant visual features from random stimuli for cortical complex cells. *J. Neurosci.*, 22(24):10811–10818.
- Twyford, P., Cai, C., and Fried, S. I. (2014). Differential responses to high-frequency electrical stimulation in on and off retinal ganglion cells. *J. Neural Eng.*, 11(2):025001.
- van Hateren, J. H. and Snippe, H. P. (2007). Simulating human cones from mid-mesopic up to high-photopic luminances. *J. Vis.*, 7(4):1.
- Vision Australia (2012). Eye Conditions, url: <https://www.visionaustralia.org/eye-health/eye-con>.
- Walter, P., Szurman, P., Vobig, M., Berk, H., Lüdtke-Handjery, H. C., Richter, H., Mittermayer, C., Heimann, K., and Sellhaus, B. (1999). Successful long-term implantation of electrically inactive epiretinal microelectrode arrays in rabbits. *Retina*, 19(6):546–552.

- Wang, B. and Weiland, J. D. (2015). Resistivity profiles of wild-type, rd1, and rd10 mouse retina. In *37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, pages 1650–1653. IEEE.
- Wang, Y. X., Pan, Z., Zhao, L., You, Q. S., Xu, L., and Jonas, J. B. (2013). Retinal nerve fiber layer thickness. The Beijing Eye Study 2011. *PLOS One*, 8(6):e66763.
- Wearne, S. L., Rodriguez, A., Ehlenberger, D. B., Rocher, A. B., Henderson, S. C., and Hof, P. R. (2005). New techniques for imaging, digitization and analysis of three-dimensional neural morphology on multiple scales. *Neuroscience*, 136(3):661–680.
- Wedderburn, R. (1976). On the existence and uniqueness of the maximum likelihood estimator for certain generalized linear models. *Biometrika*, 63(June):27–32.
- Weiland, J. D., Cho, A. K., and Humayun, M. S. (2011). Retinal prostheses: Current clinical results and future needs. *Ophthalmol.*, 118(11):2227–2237.
- Weiland, J. D. and Humayun, M. S. (2014). Retinal prosthesis. *IEEE Trans. Biomed. Eng.*, 61(5):1412–1424.
- Weiland, J. D., Liu, W., and Humayun, M. S. (2005). Retinal prosthesis. *Annu. Rev. Biomed. Eng.*, 7:361–401.
- Weitz, A. C., Behrend, M. R., Lee, N. S., Klein, R. L., Chiodo, V. a., Hauswirth, W. W., Humayun, M. S., Weiland, J. D., and Chow, R. H. (2013). Imaging the response of the retina to electrical stimulation with genetically encoded calcium indicators. *J. Neurophysiol.*, 109(7):1979–1988.
- Werginz, P., Fried, S. I., and Rattay, F. (2014). Influence of the sodium channel band on retinal ganglion cell excitation during electric stimulation - a modeling study. *Neuroscience*, 266:162–177.
- Werginz, P. and Rattay, F. (2015). Past, present, future: A review on visual prostheses. *Minerva Med.*, 105(1):65–77.
- Wilke, R. G., Moghadam, G. K., Lovell, N. H., Suaning, G. J., and Dokos, S. (2011). Electric crosstalk impairs spatial resolution of multi-electrode arrays in retinal implants. *J. Neural Eng.*, 8(4):046016.
- Wohrer, A. and Kornprobst, P. (2009). Virtual Retina: A biological retina model and simulator, with contrast gain control. *J. Comput. Neurosci.*, 26(2):219–249.
- Zeck, G., Lambacher, A., and Fromherz, P. (2011). Axonal transmission in the retina introduces a small dispersion of relative timing in the ganglion cell population response. *PLOS One*, 6(6):e20810.
- Zeiler, M. D. (2012). ADADELTA: An adaptive learning rate method. *CoRR*, abs/1212.5.
- Zhao, D., He, Z., Vingrys, A. J., Bui, B. V., and Nguyen, C. T. O. (2015). The effect of intraocular and intracranial pressure on retinal structure and function in rats. *Physiol. Rep.*, 3(8):e12507.
- Zrenner, E. (2002). Will retinal implants restore vision? *Science*, 295(5557):1022–1025.
- Zrenner, E., Bartz-Schmidt, K. U., Benav, H., Besch, D., Bruckmann, A., Gabel, V.-P., Gekeler, F., Greppmaier, U., Harscher, A., Kibbel, S., Koch, J., Kusnyerik, A., Peters, T., Stingl, K., Sachs, H., Stett, A., Szurman, P., Wilhelm, B., and Wilke, R. (2011). Subretinal electronic chips allow blind patients to read letters and combine them to words. *Proc. R. Soc. B Biol. Sci.*, 278(1711):1489–1497.