UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title

Modeling the biological visual system: from static and computational to active and data-driven

Permalink

https://escholarship.org/uc/item/5705m08z

Author

Xu, Aiwen

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

University of California
Santa Barbara

Modeling the biological visual system: from static and computational to active and data-driven

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Computer Science

by

Aiwen Xu

Committee in charge:

Professor Michael Beyeler, Chair
Professor Ambuj Singh
Professor Linda Petzold

June 2024

The Dissertation of Aiwen Xu is approved.

Professor Ambuj Singh

Professor Linda Petzold

Professor Michael Beyeler, Committee Chair

June 2024

Modeling the biological visual system: from static and computational to active and
data-driven

Copyright © 2024

by

Aiwen Xu

To my family.

Acknowledgements

I would like to express my deepest gratitude to my advisor, Michael Beyeler, for his unwavering support, guidance, and encouragement throughout my doctoral journey. His expertise, patience, and insightful feedback have been invaluable in shaping this dissertation.

I am immensely grateful to my dedicated colleagues at the Bionic Vision Lab for their collaboration and shared passion for research. Their friendship, intellectual exchange, and willingness to lend a helping hand have enriched my academic experience immeasurably.

Special thanks are due to my committee members, Ambuj Singh and Linda Petzold, for their valuable insights and support. Their expertise and commitment to academic excellence have significantly enriched the quality of this work.

I also want to extend my appreciation to my former advisor, Tobias Höllerer, for their mentorship, wisdom, and contributions to my academic growth, and my previous colleagues at the Four Eyes Lab, for their encouragement during the early years.

To my parents, Rong Xu and Yaoping Cai, I owe a debt of gratitude that words cannot adequately convey. Your unwavering love, encouragement, and sacrifices have been the foundation upon which I've built my academic pursuits.

Last but certainly not least, I want to thank my husband, Siqing Zhang, for his unwavering support, understanding, and patience throughout this journey.

To all those who have supported me along the way, whether through words of encouragement, constructive criticism, or a comforting presence, I am truly grateful. This dissertation would not have been possible without your support and belief in me.

Thank you from the bottom of my heart.

Curriculum Vitæ
Aiwen Xu

Education

2024	Ph.D. in Computer Science University of California, Santa Barbara
2023	M.S. in Computer Science University of California, Santa Barbara
2018	B.S. in Computer Science, B.S. in Mathematics New York University Shanghai

Publications

Aiwen Xu, Yuchen Hou, Cristopher M. Niell, Michael Beyeler. Multimodal Deep Learning Model Unveils Behavioral Dynamics of V1 Activity in Freely Moving Mice. *NeurIPS 2023: Thirty-seventh Conference on Neural Information Processing Systems.*

Aiwen Xu, Michael Beyeler. Retinal ganglion cells undergo cell type—specific functional changes in a computational model of cone-mediated retinal degeneration. *Frontiers in Neuroscience*, 2023.

Nicole Han, Sudhanshu Srivastava, **Aiwen Xu**, Devi Klein, Michael Beyeler. Deep learning–based scene simplification for bionic vision. *AHs '21: Augmented Humans International Conference*, February 22-24, 2021, online. ACM, New York, NY, USA, 10 pages. **Honorable mention**.

Yi Ding, Brandon Huynh, **Aiwen Xu**, Tom Bullock, Hubert Cecotti, Matthew Turk, Barry Giesbrecht, Tobias Höllerer. Multimodal Classification of EEG During Physical Activity. *ICMI '19: 2019 International Conference on Multimodal Interaction.*

Vrishab Krishna, Yi Ding, **Aiwen Xu**, Tobias Höllerer. Multimodal Biometric Authentication for VR/AR using EEG and Eye Tracking. *ICMI '19: Adjunct of the 2019 International Conference on Multimodal Interaction.*

Ke Xu, Shunan Guo, Nan Cao, David Gotz, **Aiwen Xu**, Huamin Qu, Zhenjie Yao, Yixin Chen. ECGLens: Interactive Visual Exploration of Large Scale ECG Data for Arrhythmia Detection. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2018.* **Honorable mention**.

Conference Abstracts

Aiwen Xu, Michael Beyeler. A biophysically detailed model of retinal degeneration. *COSYNE 2023: Computational and Systems Neuroscience.*

Yuchen Hou, **Aiwen Xu**, Dylan Martins, Amirali Vahid, Elliott Abe, Cristopher Niell, Michael Beyeler. Retinal scene statistics for freely moving mice. *COSYNE 2023: Computational and Systems Neuroscience*

Aiwen Xu, Nicole Han, Sudhanshu Srivastava, Devi Klein, Michael Beyeler. Enhancing simulated prosthetic vision with deep learning–based scene simplification strategies. *Vision Sciences Society 2021.*

Professional Experience

Software Engineering Intern, Google 06/2023 - 09/2023

- Implemented recitation-based dynamic few-shot prompting in Bard, an AI chat tool based on large language models (PaLM 2)
- Developed a dynamic few-shot prompting recipe in Bard via retrieving examples similar to the current user query from the supervised training mixture, based on cosine similarity of T5 sentence embeddings
- Evaluated Bard performance in different pillars (creativity, factuality, coding, reasoning) and settings (supervised finetuning, retrieval-based few-shot prompting, and a mixture of both)
- Demonstrated that the performance enhancement due to retrieval-based few-shot prompting was strongly correlated with that due to supervised finetuning via automatic side-by-side evaluation

Software Engineering Intern, Google 06/2022 - 09/2022

- Increased the top-1 accuracy of the Universal Vision Transformer (UViT) on ImageNet from 80% to 82% via hyperparameter tuning and distillation
- Reduced the latency of UViT by 15% by identifying a performance bottleneck and reimplementing the multi-head attention layer to remove unnecessary transposes in Keras and Tensorflow
- Performed neural architecture search and quantization to further reduce the latency of the model by 54% compared to the baseline while maintaining similar accuracy

Abstract

Modeling the biological visual system: from static and computational to active and data-driven

by

Aiwen Xu

A more complete understanding of the biological visual system can inspire the design of computer vision algorithms, and building accurate models constitutes an important step to such an understanding. We utilize computational and deep learning approaches to close the gaps in the literature on modeling the retina and the primary visual cortex (V1), the two important components of the early visual processing pathway.

Firstly, to address the lack of a comprehensive computational model of retinal degeneration, we present a biophysically detailed model of the cone pathway in the retina that simulates responses to light and electrical stimulation. Anatomical and neurophysiological changes due to retinal degenerative diseases were systematically introduced. The model was not only able to reproduce common findings about retinal ganglion cell (RGC) activity in the degenerated retina, but also offered testable predictions about the underlying neuroanatomical mechanisms. These insights may further our understanding of retinal processing and inform the design of retinal prostheses.

Secondly, to argue for more emphasis on freely moving experimental design, we propose an analysis of the retinal input during free exploration in mice. Mice were able to employ compensatory and gaze-shifting eye-head movements to sample the visual environment during natural locomotion. We found that eye movements preferred features such as edges and textures. A deep learning predictive model of gaze shifts indicated that the upper peripheral visual field contributed most to the prediction, consistent with animal behavior such as predator detection. These results may provide implications for visual processing beyond head-fixed preparations.

Lastly, to bridge the gap in predictive modeling tailored for neural data gathered from freely moving experimental paradigms, we introduce a multimodal recurrent neural network that integrates gaze-contingent visual input with behavioral and temporal dynamics to explain V1 activity in freely moving mice. The model achieves state-of-the-art predictions of V1 activity during free exploration. Analyzing our model using maximally activating stimuli and saliency maps, we reveal new insights into cortical function, including the prevalence of mixed selectivity for behavioral variables in mouse V1. Our model offers a comprehensive deep-learning framework for exploring the computational principles underlying V1 neurons in freely-moving animals engaged in natural behavior.

Contents

Curriculum Vitae **vi**

Abstract **viii**

1 Introduction **1**
1.1 Motivation 1
1.2 Aims and objectives 3
1.3 Organization 5

2 Background **7**
2.1 The early visual system 7
2.2 Computational models of the early visual system 10

3 A biophysically detailed computational model of the retina **16**
3.1 Introduction 16
3.2 Methods 18
3.3 Results 27
3.4 Discussion 31

4 A computational model of retinal degeneration **32**
4.1 Introduction 32
4.2 Methods 34
4.3 Results 38
4.4 Discussion 48

5 Retinal scene statistics for freely moving mice **51**
5.1 Introduction 51
5.2 Methods 53
5.3 Results 59
5.4 Discussion 67

6 A multimodal deep learning model of V1 in freely moving mice 70
6.1 Introduction 70
6.2 Related Work 72
6.3 Methods 74
6.4 Results 79
6.5 Discussion 85

7 Conclusion and future work 89

A Appendix 92
A.1 Appendix to Chapter 4 92
A.2 Appendix to Chapter 5 96
A.3 Appendix to Chapter 6 96

Bibliography 102

Chapter 1

Introduction

1.1 Motivation

Artificial intelligence (AI) and neuroscience have been closely connected and their interplay has contributed to the progress of both disciplines since the start of the computer age [1, 2, 3, 4].

Neuroscience can benefit the development of AI through both inspiration and validation [1]. On one hand, the brain's computational principles uncovered by neuroscience research offer a wealth of inspiration for novel algorithms and architectures in AI [1]. For example, the perceptron, the precursor to modern artificial neural networks, is modeled after the organizational principles of neurons in the brain [5]. In the same vein, modern artificial neural networks, which are successful at vision tasks, language tasks, and beyond, closely replicate brain-like computations [6]. They are structured with layers of neuron-like units linked by adaptable weights with a striking similarity to the synapses in the biological brain [2]. On the other hand, if an algorithm is found to be also implemented in the brain, it is more plausible to conclude that it can be an integral part of general intelligence [1]. For instance, although the transformer architecture does not seem to be directly inspired by the brain [7], neuroscientists have discovered similar mechanisms in the brain [8, 9, 10], which can potentially lead to a better understanding of why transformers work so well and how they might be utilized in a general intelligence system.

In the other direction, AI has also been revolutionizing neuroscience research, leading to a better understanding of the brain. Modern deep learning methods resulting from active research in AI have provided a more effective toolset for conducting neuroscience research, including analyzing data on a large scale [11, 12, 13], generating experimental stimuli to elicit precise response from neurons [14, 15], and decoding brain activity [16, 17]. Perhaps more importantly and more fundamentally, a comparison between trained artificial neural networks and real neural data can often reveal new insights into visual processing in higher visual areas of the biological brain [18, 19, 20].

Building accurate models so that the neural responses to arbitrary stimuli can be reliably predicted constitutes an important step to a more complete understanding of the visual system [21]. Given the close interaction between AI and neuroscience, deep learning models have been built to understand the different components of the visual system, including the retina [22, 23, 24], the primary visual cortex (V1) [25, 26, 27, 28, 29], and higher visual areas [30, 31, 18].

However, these models are limited in two ways.

- First, they are somewhat reductionist compared to the brain, which includes multiple cell types that are modulated by various neurotransmitters and linked via feedforward, lateral and feedback connections. It is important to increase the structural and mechanistic correspondence of AI models to the brain to advance the understanding of the brain and in turn to encourage the development of new AI algorithms [4].

- Second, these models are usually trained with a curated and static dataset, and do not consider the action of underlying agent. But vision is developed with an active interaction with the environment [32], and other sensory modalities during movement might affect visual processing as well [33, 34]. A comprehensive model of visual processing integrating movement and multimodal inputs may potentially inspire algorithms relevant to embodied AI [35, 36].

1.2 Aims and objectives

The goal of this dissertation is to use tools from computer science to address the gaps in the literature on modeling the early visual system, including the retina and the V1. Specifically, this dissertation introduces a large-scale, biophysically detailed model of the retina, analyzes the retina-centric visual inputs during natural locomotion, and proposes a multimodal recurrent neural network that integrates visual, behavioral and temporal dynamics designed for neuroscience data collected from freely-moving animals.

Thus, the contributions of this thesis are as follows:

- To fill the gap that a realistic model of the retina is missing, we introduce a network-level model of the retina including feed-forward and lateral connections in 3D space with accurate biophysical modeling of the retinal ganglion cell (RGC), which is one of the first computational models in the literature that allows simultaneous simulation of thousands of neurons with biophysical detail. The model includes various cell types and is able to simulate the network-level response to both light and electrical stimulation. Then we systematically introduce anatomical and neurophysiological changes caused by retinal degenerative diseases (e.g., reduced light sensitivity of photoreceptor, cell death, cell migration) to the network and study their effect on network activity. We find that ON and OFF pathways are sensitive to different network changes and thus affected differently, which may explain why the biological visual system is more robust than the artificial counterpart and argue for a similar two-pathway design in artificial neural networks. In addition, simulating the degenerated retinal model with electric stimulation may inform the design of retinal prostheses.

- To understand the transformation of visual input during natural locomotion, we tap into a large experimental dataset collected by our collaborator Prof. Cris Niell at the University of Oregon, to analyze the retinal input encountered during free exploration by measuring the mouse's visual scene and eye and head position using head-mounted cameras and an inertial measurement unit (IMU). We train a neural network to estimate the retinal input by gaze-correcting the worldcam video with an affine transformation. We find that eye movements kept overall luminance and contrast the same and increased edge density and Difference of Gaussian (DoG) entropy, which indicates that higher-level features such as edges and textures are more important for visual processing during natural locomotion. To discover which visual features drove gaze-shifting eye movements, we fine-tune EfficientNet-B0 to predict the horizontal angle of the saccadic endpoint from video frames immediately preceding a gaze shift. Our model could account for 54.4% of the variance in saccade targeting. A saliency map analysis of the fine-tuned model revealed that the pixels that were most predictive of saccadic endpoints were often localized to the upper visual field, consistent with animal behavior such as scanning for predators. These results may provide insight into how mice use their eyes to sample the visual scene, with implications for visual processing beyond head-fixed preparations.
- To address the problem of modeling the dynamic, continuous visual stimuli that arise during movement in the real world, and the integration of natural visual input and other sensory modalities over time, we introduce a multimodal recurrent neural network that integrates gaze-contingent visual input with behavioral and temporal dynamics

to explain V1 activity in freely moving mice. We show that the model achieves state-of-the-art predictions of V1 activity during free exploration and demonstrate the importance of each component in an extensive ablation study. Analyzing our model using maximally activating stimuli and saliency maps, we reveal new insights into cortical function, including the prevalence of mixed selectivity for behavioral variables in mouse V1. Our model offers a comprehensive deep-learning framework for exploring the computational principles underlying V1 neurons in freely-moving animals engaged in natural behavior.

1.3 Organization

The remainder of this thesis is organized as follows.

Chapter 2 outlines the necessary background knowledge. It introduces the early visual system, including the retina and V1, and reviews existing modeling approaches for these components in the literature.

Chapter 3 summarizes our work on implementing a biophysically detailed model of the retina. Our model simulates a total of $11,138$ cells belonging to nine different cell types (cone photoreceptors, horizontal cells, ON/OFF bipolar cells, ON/OFF amacrine cells, and ON/OFF ganglion cells) confined to a $300 \times 300 \times 210\,\mu\text{m}$ patch of the parafoveal retina. Each neuron is governed by a dynamical system with realistic modeling of various ion channels. Our model is one of the first in the literature that allows simultaneous simulation of more than ten thousand cells with biophysical detail, which is made possible with GPU enhanced Neuronal Networks (GeNN) backend.

Chapter 4 summarizes our work on systematically introducing anatomical and neurophysiological changes caused by retinal degenerative diseases to the retinal model introduced in Chapter 3. By studying the changes in population and single-neuron responses to both light and electrical stimulation, we uncover insights that may further our understanding of retinal processing, inspire new architectures in computer vision, and inform the design of retinal prostheses.

Chapter 5 summarizes our efforts in quantifying visual coding in freely moving animals by using deep learning methods to correct raw video feed recorded from the head-mounted camera for eye and head movements and to discover which visual features drive gaze-shifting eye movements. Our analysis reveals the difference in visual input during free locomotion due to the effects of eye and head movement, which may inform the development of embodied AI algorithms.

Chapter 6 summarizes our work on designing a multimodal deep-learning model that combines visual, behavioral and temporal dynamics to explain V1 neural activity collected from freely-moving animals. Our study opens a new perspective on modeling neural activity during natural conditions, and offers a comprehensive deep-learning framework that is able to reveal new computational insights into cortical function in freely-moving animals.

Chapter 7 concludes this dissertation and proposes future research directions, including training large foundation models and moving beyond supervised learning.

Chapter 2

Background

Artificial intelligence (AI) and neuroscience have been influencing each other for decades. Neuroscience has inspired recent deep learning models; in turn, these models could be used to further our understanding of visual processing in the biological brain. In this section, we briefly introduce the readers to the retina and the primary visual cortex (V1), the two initial stages of the early visual processing pathway. The visual processing pathway from the retina to the cortex is common to all mammals: It consists of the retina which converts light into action potentials, and through lateral geniculate nucleus (LGN), the information is passed to the V1 for further processing [37].

2.1 The early visual system

2.1.1 Retina

The retina is a sheet of neural tissue that lines the back of the eye. Its main function is to transduce light into electrical signals, perform initial processing and send the information to the cortex for further processing. Despite being only around 200 µm thick, the retina is a very complicated neural network, both structurally and functionally [38, 39].

In a simplified view, the retina contains five main types of neurons, which are photoreceptors, horizontal cells, bipolar cells, amacrine cells and retinal ganglion cell (RGC). The main feed-forward pathway consists of photoreceptors, bipolar cells and RGC, while horizontal cells and amacrine cells provide feedback and lateral processing [38]. Note that feedback and lateral processing is an indispensable part of retinal processing. For example, the amacrine cells receive signals from bipolar cells and other amacrine cells, and output signals to bipolar cells, RGCs and other amacrine cells, thus playing roles in feedback inhibition, feed-forward inhibition and lateral inhibition. A single amacrine cell may have a subset of those functions or even all of them [38].

A more detailed view reveals an interconnected network of more than 60 distinct type of neurons [38]. There are at least 10 distinct types of bipolar cells in the mammalian retina. At least 30 types of amacrine cells have been identified in the mammalian retina, and the function of most of them are still poorly understood. RGCs compose the last layer of the retina and project to the brain, and there are 17 distinct types, each of which covers the visual field with a regular structure [39]. Recent research suggests that visual signals from different RGC types may remain separate in cortical processing [39].

Due to the variety of cell types and connection types, the processing of the retina is complex and highly nonlinear, resulting in functions such as omitted stimulus response, latency encoding, motion reversal and motion anticipation [23, 24]. Those retinal signals provide an important foundation for downstream processing in V1 and higher visual areas.

Understanding how the neural signals in the retina change is essential for facilitating research in neuroengineering applications such as retinal implants, which have the potential to restore vision in individuals with visual impairment. The retinal network undergoes complex and multistage alterations during degenerative diseases, resulting in a drastic change in the neural signaling [40, 41]. The process of retinal degeneration starts with photoreceptor stress that leads to outer segment truncation (Phase I) and progressive photoreceptor cell death (Phase II), followed by a protracted period of global cell migration and cell death (Phase III). Those main changes are accompanied with the hypertrophy of Müller cells, the growth of abnormal neurites, and the growth and regression of microneuromas.

2.1.2 Primary visual cortex (V1)

V1 is the first step of the cortical processing of visual information. V1 neurons can be classified into two types depending on the structure of their receptive fields: simple cells and complex cells [42, 43]. In the receptive field of a visual neuron, there are ON subregions which are excited by light and inhibited by dark, and OFF subregions whose response property is the exact opposite [44]. Simple cells have separate ON and OFF subregions, while complex cells have superimposed ON and OFF subregions, meaning that the same location in the receptive field could respond to light and dark.

V1 neurons exhibit various types of stimulus selectivity, including orientation selectivity [42], spatial frequency selectivity [45], direction selectivity [46, 47], temporal frequency selectivity [48], and color selectivity [49, 50].

V1 neuron properties may differ in different species. For example, in the mouse, which has become an excellent model organism and is perfect for studying vision during natural locomotion, V1 is distinctly different in the following ways [51]. Firstly, although V1 neurons in the mouse still encode low-level visual features [52], other information has widespread influence on V1 activity, including eye and head positions [53, 54, 55], locomotion [34, 56, 57], arousal [58, 59], the recent history of the animal [60], and reward [61]. Secondly, some neurons in mouse visual cortex exhibit wider tunings: they are less selective in terms of orientation and spatial frequency [62].

2.2 Computational models of the early visual system

2.2.1 Computational models of the retina

Modeling via direct data fitting

In the literature of modeling the retina, one approach is to directly fit a model with the stimulus shown to the retina and the recorded activity from RGCs. The model might loosely correspond to the general biology of the retina, but it does not include detailed modeling of the underlying biophysical mechanisms. Researchers have traditionally used the linear-nonlinear-Poisson (LNP) model to study the retina, and more recently, deep learning models have become the state-of-the-art for modeling the retina.

The goal of the LNP model is to estimate the firing rate $r(t)$ of a neuron through a combination of linear and nonlinear operations given the stimulus $s(t)$, and generate the spikes according to the estimated firing rate via an inhomogeneous Poisson process [63]. Mathematically, the linear and nonlinear operations can be summarized as

$$r(t) = N(k_1 \cdot s(t), k_2 \cdot s(t), ..., k_m \cdot s(t)),$$

where $\{k_1, ..., k_m\}$ is a set of linear filters and N is a nonlinear function that maps the output of the linear filters to a firing rate [64]. The stimulus vector $s(t)$ is spatiotemporal, and its duration in time is typically empirically determined to be the period over which the stimulus can affect the response [65]. It is typically assumed that the generation of spikes does not depend on the history of previous spikes, so a Poisson process is used to generate the spikes according to the estimated firing rate $r(t)$.

Despite being relatively simple, the LNP model has uncovered many mechanisms underlying the responses of the RGCs to the visual stimulus. For example, the LNP model can describe how blue-ON/yellow-OFF RGCs process the inputs from short wavelength sensitive cones [66]. The LNP model can also describe the contrast adaptation of the RGCs, and the linear filters reveal the differences in adaptation in different cell types [67]. The LNP model can also be extended to incorporate more complex dynamics. With the addition of nonlinear subunits, the LNP model can describe the light responses of RGCs and the subunits are consistent with the structure of the retina [68, 69]. With extensions such as post spike filters (to model the dependency on the spike history) and coupling filters (to model the synaptic connections between neurons), the LNP is able to reveal the properties of the neural encoding at the population level [70]. However, the LNP model cannot accurately describe the responses of the RGCs to naturalistic stimuli [71]. This result has led to the development of deep learning models, which are more expressive and align with the retinal computations more closely.

The most successful deep learning model of the retina is a two-hidden layer convolutional neural network (CNN), closely mimicking the feed-forward pathway of the retina which also contains two layers of neurons before the RGCs [22] (Figure 2.1). The deep learning model has been shown to be superior in describing the responses of the RGCs to simple visual stimuli compared to the LNP model [72, 73]. More importantly, it is able to describe the responses of the RGCs to naturalistic visual stimuli, the task which the LNP model has failed at: Both the CNN [22] and the recurrent neural network (RNN) [74] have achieved good performance. The CNN model which is fit to only the RGC data in response to naturalistic visual stimuli exhibits known phenomena in response to artificial stimuli, and for most of those phenomena the CNN model employs the same computational principles established in previous literature [24, 23]. However, it predicts the response to images, but the visual processing system does not deal with static images. In addition, it is not clear how to adapt the CNN model to predict the retinal response to electrical stimuli, which limits its usefulness in neuro-engineering applications.

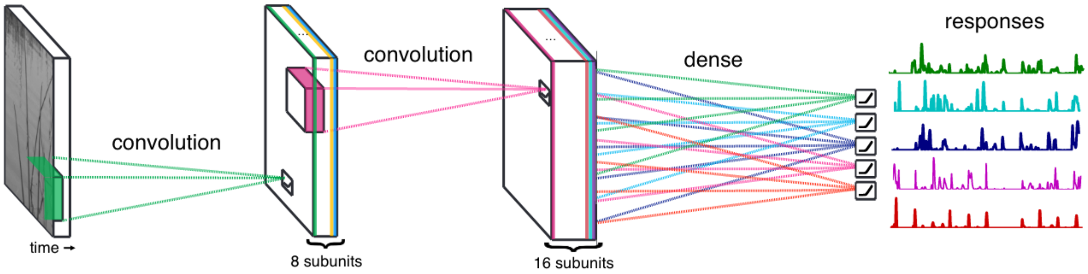

Figure 2.1: Diagram of the best CNN model of the retina (from [22]). It consists of two convolutional layers each followed by a non-linearity, and a fully connected layer for the final prediction of the RGC firing rate.

Modeling via biophysical mechanisms

The other popular method of modeling the retina is to more closely follow the biophysical mechanisms of the retinal cells. Those models range from Hodgkin-Huxley-type models to multi-compartmental models, with an increasing amount of detail. It is very computationally expensive to simulate biophysical details, especially if we are modeling a large population of retinal neurons each modeled by a Hodgkin-Huxley-type model, or neurons with a lot of detailed compartments. Recently, this has become feasible via the development of Python packages that allows the model to be run on GPUs [75, 76].

The membrane potential V of a neuron can be modeled according to its electrical properties as

$$c_{\mathrm{m}} \frac{dV}{dt} = -i_{\mathrm{m}},$$

where c_{m} is the specific membrane capacitance, i_{m} is the total current per unit area flowing across the membrane.

In a Hodgkin-Huxley-type model, i_{m} is modeled more realistically by considering voltage-dependent conductances, synaptic conductances and/or Ca^{2+}-dependent conductances. The prominent example is of course the Hodgkin-Huxley model itself [77], which models the membrane current as

$$i_{\mathrm{m}} = -\bar{g}_{\mathrm{L}}(V - E_{\mathrm{L}}) + \bar{g}_{\mathrm{K}} n^4 (V - E_{\mathrm{K}}) + \bar{g}_{\mathrm{Na}} m^3 h (V - E_{\mathrm{Na}}),$$

which is the sum of a leakage current, a delayed-rectified K^+ current, and a transient Na^+ current. Both of the ion channel conductances are voltage-dependent. Here, $\bar{g}_{\mathrm{K}}$ and $\bar{g}_{\mathrm{Na}}$ are the maximal conductance, i.e., the conductance per unit area if all the channels of a certain type are open. m, n and h are the gating variables that determine the fraction of channels in the open state, and they are governed by separate differential equations to accurately model their voltage dependency. E_{K} and E_{Na} are the equilibrium potentials of potassium and sodium respectively. In summary, the ion channels modeled here tend to move the membrane potential of the neuron toward the equilibrium potential via a voltage-dependent conductance. The Hodgkin-Huxley model can successfully explain the mechanism behind the generation of action potentials.

Synaptic conductances and Ca^{2+}-dependent conductances can be modeled in a similar way. For synaptic conductances, the fraction of channels in the open state is made to be dependent on the concentration of transmitter while for Ca^{2+}-dependent conductances, the fraction of channels in the open state is made to be dependent on the concentration of Ca^{2+}.

In the context of accurately modeling the spiking dynamics of RGC, researchers have determined the specific ion channels to include and the parameters of those ion channels according to experimental data. Traditionally, the RGC model contains five ion channels [78, 79], and recent research has suggested two more ion channels to model the RGC in more detail [80].

The Hodgkin-Huxley-type model can be used to determine how different types of electrical stimulation cause different responses in the retina. For example, introducing an interphase gap in traditional biphasic electrical stimulation can reduce the threshold of the RGC [81]. Non-rectangular capacitive-like voltage pulses caused by photovoltaic material result in an increase of the network-mediated activity, which implies more focal stimulation [82]. [83] and [84] explore a novel hexagonal configuration of the electrodes by modeling the whole RGC layer altogether, inspired by the model in [85].

A hidden assumption in the Hodgkin-Huxley-type model is that the membrane potential of a neuron is constant across its surface. In multi-compartmental models, a neuron can have different values of membrane potentials at different locations: The neuron is split into an appropriate number of compartments, each compartment comes with its own membrane current and membrane potential, and the compartments are electrically coupled with connected compartments. Very often, each of those compartments is modeled by a separate Hodgkin-Huxley-type model to reflect the complex conductances.

Multi-compartmental models can reproduce a lot of the experimental data. Multi-compartmental models are able to reconstruct and predict response properties of ON and OFF RGCs recorded experimentally [80], even when the stimulation is high-frequency [86]. Multi-compartmental models show that the increase of electrode diameter and the electrode-retina distance results in higher values of threshold current [87], and this applies to different areas of the macula where the densities of the RGCs are different [88].

However, multi-compartmental models are computationally intensive and can be hard to scale. A workaround is demonstrated in [89, 90], where multi-compartmental detailed cellular networks are duplicated and embedded into a large-scale retina space where electrical stimulation is modeled.

2.2.2 Computational models of the V1

Traditionally, researchers have utilized the LNP model to explain V1 simple cell computations [91]. Similar to the LNP model that is used to model retinal neurons, the LNP model for V1 simple cells first filter the stimulus with a single spatial filter, then pass the result to a nonlinearity, and lastly spike trains are generated with a Poisson process [92, 21, 93].

For V1 complex cells, however, multiple linear filters in space are needed to capture the computation, because they have superimposed ON and OFF subregions. The energy model are traditionally used by researchers to explain V1 complex cell computations, which can be thought of a combination of several simple cell models [21, 94]. Different ways to combine the simple cell models have been proposed, including hierarchical, parallel, and recurrent [95].

More recently, researchers have started to train CNNs to predict the activity of V1 neurons, and have achieved state-of-the-art performance in both macaque V1 [26, 29, 96] and mouse V1 [28, 25, 97, 98]. This success indicates that visual features learned by the CNN for object recognition can well explain V1 activity. However, CNNs demonstrate greater accuracy in predicting macaque V1 than mouse V1, suggesting that object recognition may not be suitable for understanding mouse vision, and further modeling efforts are required to gain a thorough understanding of mouse V1.

Chapter 3

A biophysically detailed computational model of the retina

3.1 Introduction

Understanding how the retina responds to light can inspire the design of computer vision algorithms while understanding how the retina responds to electrical stimulation is key to advancing neuroscience and neuroengineering applications such as retinal prostheses. Computational models have been built either at the single-cell level or network level to understand the response properties of the healthy retina (for a recent review, see Guo et al. [99]). These include single-compartment models ("point models") to simulate neuronal response as a function of ionic currents flowing across the neuronal membrane (e.g., Fohlmeister et al., [79]; Fohlmeister and Miller, [78], Wohrer and Kornprobst, [100]), morphologically realistic models based on detailed anatomical representations of the physical components of biological neurons (e.g., Smith, [101]; Greenberg et al., [102]), and convolutional neural networks (e.g., McIntosh et al., [22]). Several studies did not just focus on the retina's light response but also on the response to electrical stimulation

[103, 80, 104, 105, 106], which may inform treatment options for people blinded by retinal degenerative diseases.

However, a network-level model of the retina that can simulate thousands of cells simultaneously while capturing the various biophysical details of the retinal neurons is not readily accessible for utilization. The most complete network-level model of the retina so far lays out the neurons in a three-dimensional space, models synapses between retinal neurons realistically and provides mechanisms for simulating both light and electrical stimulation to the retina [103]. Unfortunately, the retinal ganglion cells (RGCs), which compose the output layer of the retina, are modeled inadequately: They simply output a firing rate according to the membrane potential, and the detailed spiking dynamics of the RGCs are not included. The RGCs are known to exhibit several detailed spiking dynamics, including spike frequency adaptation, burst firing, and rebound excitation, and they can be modeled only with a Hodgkin-Huxley type model with specific ion channels [79, 78, 107, 80]. Accurately modeling those spiking dynamics can provide a detailed account of the retina under varied electrical stimuli (with different pulse shapes, amplitudes, and frequencies) and facilitate the study of temporal changes and the calculation of the electrical threshold [81], crucial aspects for leveraging the model in enhancing retinal prosthesis design.

To bridge this gap, we present a biophysically inspired *in silico* computational model of the cone pathway in the retina. We choose to implement this model in Python to facilitate potential interaction with popular deep learning frameworks, and open-source the code for the wider research community to use. Following previous research [103], the three-dimensional model contains more than ten thousand cells from nine different types, interconnected with excitatory and inhibitory synapses to provide both feed-forward and lateral processing. Each neuron in the RGC population is modeled with the best biophysical model of the single RGC in the literature, which includes seven ion channels with accurate parameters to reflect the different spiking dynamics in ON and OFF RGCs [80]. To make the simultaneous simulation of thousands of neurons with biophysical details feasible, we utilize the GPU enhanced Neuronal Networks (GeNN) backend to run the model. [75, 76]. We simulate the network-level response to light stimulation and verify the model reproduces seminal findings about the light response of RGCs in ON and OFF pathways. This model lays a solid foundation for studying the retina's response to electrical stimulation and introducing degenerative changes to the retina.

3.2 Methods

Inspired by [103], we started by simulating a three-dimensional patch ($300\,\mu\text{m} \times 300\,\mu\text{m} \times 210\,\mu\text{m}$) of the cone pathway in the parafoveal retina. The network consisted of $11,138$ cells belonging to nine different cell types ($4,149$ cones, 537 horizontal cells, $3,508$ ON/OFF bipolar cells, 779 ON/OFF wide-field amacrine cells, 723 narrow-field amacrine cells, and $1,442$ ganglion cells), connected via generally accepted [108, 109] synaptic connections (see Fig. 3.1).

Briefly, upon photoactivation, cone photoreceptors (labeled "PR" in Fig. 3.1A) produced a photocurrent that led to hyperpolarization in OFF bipolar cells and depolarization in ON bipolar cells ("BP"). In addition, cones excited horizontal cells ("HRZ"), which in turn inhibited cone terminals, thus generating an inhibitory surround in the bipolar cell response. ON and OFF bipolar cells then excited ON and OFF amacrine cells ("AMA") as well as ON and OFF ganglion cells ("RGC"), respectively, to generate an inhibitory surround in the ganglion cell response. ON and OFF amacrine cells also provided lateral inhibition to ON and OFF ganglion cells, respectively. Lastly, we included a unilateral inhibitory connection from a special type of narrow-field ON amacrine cell to OFF ganglion cells [110]. Rod circuitry was not implemented.

The retina model was implemented using Brian 2 [75] and Brian2GeNN [76] in Python. All simulations were run on a single NVIDIA RTX 3090 (24 GB of GPU memory), and all our code is available at `https://github.com/bionicvisionlab/2023-Xu-Retinal-Degeneration`.

3.2.1 Modeling individual neurons

To implement the biophysical properties of retinal neurons, we modified a leaky integrator model [103] by adding membrane and synaptic conductances (Fig. 3.1B). We assumed that neurons are electronically compact, so their activation levels could be described by a single membrane potential. All $11,138$ neurons had a spatially nonzero soma with non-gated ion channels (leakage channels) modeled by a constant linear conductance (G_m) in series with a constant single-cell battery (i.e., the cell's resting voltage, E_rest) and

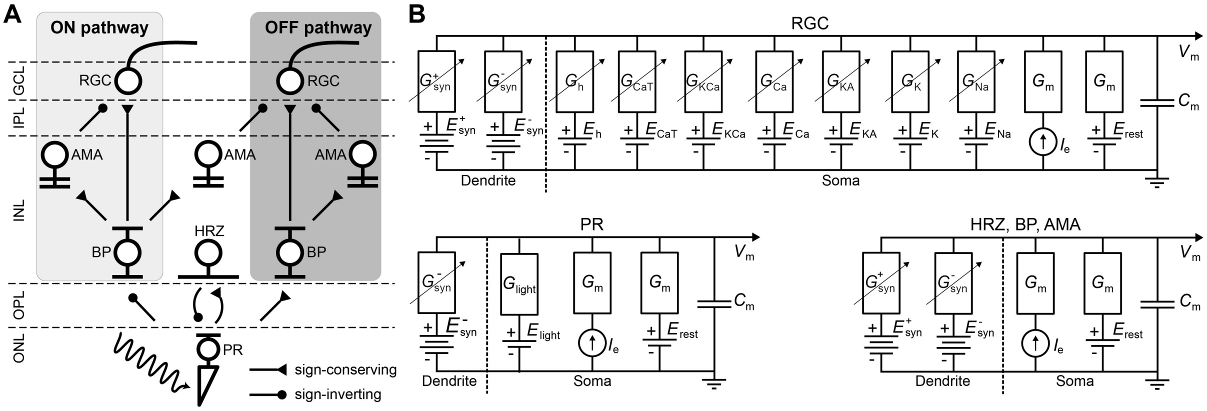

Figure 3.1: **A)** Diagram of the connections between the retinal neurons. PR: photoreceptor, HRZ: horizontal cell, BP: bipolar cell, AMA: amacrine cell, RGC: retinal ganglion cell. ONL: outer nuclear layer, OPL: outer plexiform layer, INL: inner nuclear layer, IPL: inner plexiform layer, GCL: ganglion cell layer. **B)** RC circuit model of a neuron's membrane potential. All neurons included a membrane capacitance (C_m), a leakage current (I_leak), an external current driven by the extracellular potential gradient (I_e), and synaptically gated ionic currents (I^+_syn and I^-_syn). RGCs had additional voltage-gated and ligand-gated ionic currents [80] and photoreceptors had a photo-sensitive current (I_light) as described in [103]. Dendritic trees and ganglion cell axons were not modeled.

an extracellular current that modeled extracellular electrical stimulation.

With the exception of RGCs, all other cell types (labeled "HRZ", "BP", "AMA" in Fig. 3.1B) were modeled as leaky integrators [103], whose membrane potential (v_m) followed the following differential equation:

$$C_\mathrm{m}\frac{dv_\mathrm{m}}{dt} = \sum_\mathrm{syn} i_\mathrm{syn} + i_\mathrm{ext} + i_\mathrm{leak}, \tag{3.1}$$

where C_m was the cell type–specific membrane conductance, the sum was over all presynaptic currents i_syn (see Section 3.2.2), i_ext was the external current resulting from extracellular electrical stimulation (see Section 4.2.2), and i_leak was a leakage current modeled by a constant linear conductance (G_m) in series with a constant single-cell battery (i.e., the cell's resting voltage, E_rest):

$$i_\mathrm{leak} = -G_\mathrm{m}\big(v_\mathrm{m} - E_\mathrm{rest}\big). \tag{3.2}$$

It is worth noting that in reality these cell types contain a variety of voltage-gated and ligand-gated ion channels. However, modeling the behavior of these neurons with multiple ion channels would have further increased the complexity and computational cost of the model. For the sake of practical feasibility, we therefore had to limit ourselves to a single ion channel. [103] demonstrated that this can still lead to realistic light responses.

The light-cone interaction of photoreceptors (labeled "PR" in Fig. 3.1B) was modeled with an additional current as a synapse, described in detail in [103] and given as:

$$C_\mathrm{m}\frac{dv_\mathrm{m}}{dt} = \sum_\mathrm{syn} i_\mathrm{syn} + i_\mathrm{ext} + i_\mathrm{leak} + i_\mathrm{light}, \tag{3.3}$$

$$i_\mathrm{light} = -g_\mathrm{light}\big(v_\mathrm{m} - E_\mathrm{light}\big) \tag{3.4}$$

where $E_{\text{light}} = -8\,\text{mV}$ was the reversal potential and g_{light} was the synaptic conductance that depended on the time-dependent light intensity $l(t) \in [0, 1]$:

$$g_{\text{light}} = G_{\text{light}}\big(1 - l(t)\big). \tag{3.5}$$

As one of our main goals was to study the RGC response to electrical stimulation, which is often delivered by short biphasic pulses, we considered it important that our simulated RGC population exhibited detailed temporal responses. Thus, our implementation of RGCs (labeled "RGC" in Fig. 3.1B) deviated from [103], as they were modeled as Hodgkin-Huxley neurons with seven ion channels (Eq. 3.6) that were previously shown to capture the firing dynamics of RGCs in the rabbit retina [78, 80]:

$$C_m \frac{dV_{\text{m}}}{dt} = \sum_{\text{syn}} i_{\text{syn}} + i_{\text{ext}} + i_{\text{ion}}, \tag{3.6}$$

where the ionic current i_{ion} was given as the product of the neuron's surface area A and the sum of several ionic current densities:

$$\begin{aligned}
i_{\mathrm{ion}} &= A(j_{\mathrm{Na}} + j_{\mathrm{Ca}} + j_{\mathrm{K}} + j_{\mathrm{KA}} + j_{\mathrm{KCa}} + j_{\mathrm{h}} + j_{\mathrm{CaT}} + j_{\mathrm{leak}}), \qquad (3.7)\\
j_{\mathrm{Na}} &= -G_{\mathrm{Na}} m^3 h (v_{\mathrm{m}} - E_{\mathrm{Na}}),\\
j_{\mathrm{Ca}} &= -G_{\mathrm{Ca}} c^3 (v_{\mathrm{m}} - E_{\mathrm{Ca}}),\\
j_{\mathrm{K}} &= -G_{\mathrm{K}} n^4 (v_{\mathrm{m}} - E_{\mathrm{K}}),\\
j_{\mathrm{KA}} &= -G_{\mathrm{KA}} m_{\mathrm{A}}^3 h_{\mathrm{A}} (v_{\mathrm{m}} - E_{\mathrm{K}}),\\
j_{\mathrm{KCa}} &= -G_{\mathrm{KCa}} m_{\mathrm{KCa}} (v_{\mathrm{m}} - E_{\mathrm{K}}),\\
j_{\mathrm{h}} &= -G_{\mathrm{h}} n_{\mathrm{h}} (v_{\mathrm{m}} - E_{\mathrm{h}}),\\
j_{\mathrm{CaT}} &= -G_{\mathrm{CaT}} m_{\mathrm{T}}^3 h_{\mathrm{T}} (v_{\mathrm{m}} - E_{\mathrm{Ca}}),\\
j_{\mathrm{leak}} &= -G_{\mathrm{m}} (v_{\mathrm{m}} - E_{\mathrm{rest}}).
\end{aligned}$$

Here, j_{Na} was a voltage-gated sodium channel with gating variables m and h; j_{Ca} was a voltage-gated calcium channel with gating variable c and E_{Ca} modeled with the Nernst equation (see Guo et al. [80] for details); j_{K} was a non-inactivating potassium channel with gating variable n; j_{KA} was an inactivating potassium channel with gating variables m_{A} (called A in Guo et al. [80]) and h_{A}; j_{KCa} was a Ca^{2+}-activated potassium channel gated by m_{KCa} whose value was dependent on the internal calcium concentration (see Guo et al. [80] for details); j_{h} was a hyperpolarization-activated non-selective cationic channel with gating variable n_{h}; and j_{CaT} was a low-threshold voltage-activated calcium channel with gating variables m_{T} and h_{T} [78, 80].

The equations for the gating variables were identical to [80] (see their Tables 2–3). In short, all gating variables, except the inactivating gating variable h_T of j_{CaT}, followed first-order kinetics:

$$\frac{dx}{dt} = \alpha_x(1 - x) - \beta_x x, \qquad (3.8)$$

where x was the gating variable, α_x was the opening rate and β_x was the closing rate of the channel. The inactivating gating variable h_T of j_CaT (but not m_T; note the typo in Guo et al. [80]) followed second-order dynamics:

$$\frac{dh_\mathrm{T}}{dt} = \alpha_{h_\mathrm{T}}(1 - h_\mathrm{T} - d_\mathrm{T}) - \beta_{h_\mathrm{T}} h_\mathrm{T}, \tag{3.9}$$

$$\frac{d(d_\mathrm{T})}{dt} = \alpha_{d_\mathrm{T}}(1 - h_\mathrm{T} - d_\mathrm{T}) - \beta_{d_\mathrm{T}} d_\mathrm{T}. \tag{3.10}$$

The initial values of the gating variables and the internal calcium concentrations of the RGCs can be found in Appendix A.1.4.

Neurons were assumed to contain a spherical soma with either 26 µm diameter in the case of RGCs [111, 112] or 7 µm otherwise [103]. The initial values of the membrane voltages were set according to normal distributions, whose parameters can be found in Appendix A.1.4.

Ionic current densities were multiplied by the surface area A of the RGC to convert to a current. G_leak was set to a value so that the spontaneous firing rate under 0.5 light was around 2 Hz [113]. A spike was recorded whenever the membrane potential exceeded $-10\,\mathrm{mV}$.

3.2.2 Modeling the retinal circuitry

Neurons were connected as shown in Fig. 3.1 using parameters given in Table 3.3. All neurons belonging to a particular cell type were arranged in a hexagonal mosaic, where the x and y coordinates of a neuron were given as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} i & j \end{bmatrix} \times \begin{bmatrix} 1 & \sqrt{3} \\ 1 & -\sqrt{3} \end{bmatrix} \lambda, \tag{3.11}$$

	PR	HRZ	$\mathrm{BP}_{\mathrm{ON/OFF}}$	$\mathrm{AMA}^{\mathrm{WF/NF}}_{\mathrm{ON/OFF}}$	$\mathrm{RGC}_{\mathrm{ON}}$	$\mathrm{RGC}_{\mathrm{OFF}}$
C_m (pF)	80.0	210.0	50.0	50.0	50.0	50.0
G_m (nS)	4.0	2.5	2.0	2.0	—	—
G_m ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	0.3	0.274
G_light (nS)	0.9	—	—	—	—	—
E_rest(mV)	-50.0	-65.0	-45.0	-50.0	−66.5	−70.5
E_Na(mV)	—	—	—	—	35.0	35.0
E_K(mV)	—	—	—	—	−72.0	−68.0
E_h(mV)	—	—	—	—	−45.8	−26.8
G_Na ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	1072.0	249.0
G_K ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	40.5	68.85
G_KA ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	94.5	18.9
G_Ca ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	2.1	1.6
G_KCa ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	0.04	0.0474
G_h ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	0.4287	0.1429
G_CaT ($\mathrm{mS\,cm^{-2}}$)	—	—	—	—	0.008	0.1983

Table 3.1: Neuronal membrane parameters. G_m for RGCs was set such that the spontaneous firing rate was around 2 Hz. All other values for G_m and C_m were adopted from [103]. All others were adopted from [80].

where $i, j = 0, \pm 1, \pm 2...$ and λ was cell-type specific (Table 3.2). The x and y coordinates were further jittered by Gaussian noise: $\mathcal{N}(0, 1\,\mu\mathrm{m})$. Neurons were confined to different z locations depending on their cell type (Table 3.2). Within each band, the z coordinate was assigned by sampling from a random uniform distribution.

Synapses were assumed to lie at the center of a neuron's dendritic field (excitatory if $E_\mathrm{syn} > E_\mathrm{rest}$ and inhibitory if $E_\mathrm{syn} < E_\mathrm{rest}$). The synaptic connection from presynaptic neurons of the same type to a postsynaptic neuron was modeled with a current i_syn (see Equations 3.1, 3.3, and 3.6) in the postsynaptic neuron via:

$$i_\mathrm{syn} = -g_\mathrm{syn}(v_\mathrm{m} - E_\mathrm{syn}), \tag{3.12}$$

where E_syn was different for each synaptic type (see Table 3.3). g_syn was computed as a spatially weighted sum over the conductances of all the channels from the presynaptic neurons:

$$\begin{aligned} g_{\text{syn}} &= \frac{1}{W} \sum_{\mathbf{p}} g_{\text{syn},\mathbf{p}} \exp\Big(- \frac{D(\mathbf{p_0}, \mathbf{p})}{\sigma} \Big), \\ W &= \sum_{\mathbf{p}} \exp\Big(- \frac{D(\mathbf{p_0}, \mathbf{p})}{\sigma} \Big), \end{aligned} \tag{3.13}$$

where $D(\mathbf{p_0}, \mathbf{p})$ was the Euclidean distance between the center of the postsynaptic neuron $\mathbf{p_0}$ and the center of a presynaptic neuron $\mathbf{p}$, and σ was a decay constant that determined how the presynpatic currents were weighted. σ was different for each synaptic type (see Table 3.3). Note that the choice of σ led to a configuration where each bipolar cell received input from a single cone [114, 103].

Following [103], we modeled the individual channel conductance $g_{\text{syn},\mathbf{p}}$ as either a monotonically increasing (type "I") or monotonically decreasing (type "D") function of the membrane potential of the presynaptic neuron with the following equation:

$$g_{\text{syn}}(t) = \begin{cases} G_{\min} + (G_{\max} - G_{\min}) \times \bigg(1 - \Big(1 + \exp\big(\frac{v_{\text{pre}}(t-\tau) - V_{50}}{\beta}\big)\Big)^{-1}\bigg) & \text{if type is "I"}, \\ G_{\min} + (G_{\max} - G_{\min}) \times \Big(1 + \exp\big(\frac{v_{\text{pre}}(t-\tau) - V_{50}}{\beta}\big)\Big)^{-1} & \text{if type is "D"}, \end{cases} \tag{3.14}$$

where $G_{\min}$ and $G_{\max}$ were the lower and upper bound of values for the synaptic conductance, v_{pre} was the membrane potential of the presynaptic neuron, τ was the synaptic delay, V_{50} determined the function's center operating point and β determined the function's steepness (see Table 3.3).

Synaptic delays (τ in Table 3.3) were set such that ON RGCs fired their first spikes roughly 20 ms after stimulus onset and OFF RGCs fired roughly 50 ms after stimulus onset [115]. To achieve these response times, we assumed a constant transmission speed and calculated the average distance between each connected pair of cells to set τ accordingly.

	PR	HRZ	$\mathrm{BP}_{\mathrm{ON/OFF}}$	$\mathrm{AMA}^{\mathrm{WF}}_{\mathrm{ON/OFF}}$	$\mathrm{AMA}^{\mathrm{NF}}_{\mathrm{ON}}$	$\mathrm{RGC}_{\mathrm{ON/OFF}}$
λ (µm)	2.5	7.0	3.85	8.0	6.0	6.0
z_{min} (µm)	170	100	100	80	80	25
z_{max} (µm)	205	128	128	101	101	39
G_{ext} (nS)	4.0	2.5	2.0	2.0	2.0	2.0

Table 3.2: Spatial layout parameters for the different cell types [103]. PR: photoreceptor, HRZ: horizontal cell, BP: bipolar cell, AMA: amacrine cell, RGC: retinal ganglion cell, WF: widefield, NF: nonwide-field.

	τ (ms)	E_{syn} (mV)	G_{min} (nS)	G_{max} (nS)	V_{50} (mV)	β (mV)	type	σ (µm)
PR→HRZ	7	0.0	0.0	7.0	-43.0	2.0	I	10.5
HRZ→PR	7	-67.0	0.0	3.0	-29.5	7.4	I	2.5
$\mathrm{PR}\rightarrow\mathrm{BP}_{\mathrm{ON}}$	5	0.0	0.1	1.1	-47.0	1.7	D	3.85
$\mathrm{PR}\rightarrow\mathrm{BP}_{\mathrm{OFF}}$	13	0.0	0.0	3.75	-41.5	1.2	I	3.85
$\mathrm{BP}_{\mathrm{ON}}\rightarrow\mathrm{AMA}^{\mathrm{WF}}_{\mathrm{ON}}$	5	0.0	0.0	1.0	-33.5	3.0	I	24.0
$\mathrm{BP}_{\mathrm{ON}}\rightarrow\mathrm{AMA}^{\mathrm{NF}}_{\mathrm{ON}}$	5	0.0	0.0	0.2	-35.0	3.0	I	6.0
$\mathrm{BP}_{\mathrm{OFF}}\rightarrow\mathrm{AMA}^{\mathrm{WF}}_{\mathrm{OFF}}$	11	0.0	0.0	1.8	-44.0	3.0	I	24.0
$\mathrm{BP}_{\mathrm{ON}}\rightarrow\mathrm{RGC}_{\mathrm{ON}}$	5	0.0	0.0	2.5	-33.5	3.0	I	6.0
$\mathrm{AMA}^{\mathrm{WF}}_{\mathrm{ON}}\rightarrow\mathrm{RGC}_{\mathrm{ON}}$	5	-70.0	0.0	2.0	-42.5	2.5	I	6.0
$\mathrm{BP}_{\mathrm{OFF}}\rightarrow\mathrm{RGC}_{\mathrm{OFF}}$	13	0.0	0.0	2.5	-44.0	3.0	I	6.0
$\mathrm{AMA}^{\mathrm{WF}}_{\mathrm{OFF}}\rightarrow\mathrm{RGC}_{\mathrm{OFF}}$	12	-70.0	0.0	2.5	-34.4	2.5	I	6.0
$\mathrm{AMA}^{\mathrm{NF}}_{\mathrm{ON}}\rightarrow\mathrm{RGC}_{\mathrm{OFF}}$	12	-80.0	0.0	2.0	-47.5	2.0	I	6.0

Table 3.3: Synaptic parameters. The synaptic delays were set such that the latency of ON RGCs were around 20 ms and the latency of OFF RGCs were around 50 ms. All other values are from [103].

3.2.3 Estimating spatiotemporal receptive fields

To measure the spatiotemporal receptive field of an RGC, we fit a generalized linear model to its spiking response to a spatially correlated “cloud” stimulus [116]. The stimulus consisted of spatiotemporal Gaussian white noise (pixel size = 6.25 µm, mean brightness

= 0.5, standard deviation of brightness = 0.175, refresh rate = 20 Hz) filtered with a two-dimensional spatial Gaussian filter (standard deviation = 12.5 µm). The resulting spikes were binned at the refresh rate of the stimulus.

The generalized linear model predicted the firing rate of a RGC as:

$$r(t) = f\Big(\sum_i k_i s_i(t)\Big), \tag{3.15}$$

where $s_i(t)$ denoted the relevant stimulus frames before and at time t, k_i denoted the spatial filter for the stimulus frame $s_i(t)$, and f was a nonlinear function. We trained a generalized linear model in PyTorch with a spatiotemporal filter spanning five time steps and a rectified linear unit (ReLU) as the nonlinear function. The generalized linear model was regularized with a Laplacian square penalty on the spatial filters [116]. We used the mean squared error as loss function. The generalized linear model was trained for 3000 epochs with the Adam optimizer and a learning rate of 0.000001 decayed by 10 % every 2000 epochs.

3.3 Results

3.3.1 Light response of RGC population

Fig. 3.2 shows the response of the retina to a bright disk stimulus (40 µm in radius, with light intensity 1.0) presented against a gray background (300 µm by 300 µm, with light intensity 0.5). Fig. 3.2B breaks down the retinal response to the disk stimulus layer by layer. Upon stimulus onset, photoreceptors in the center of the mosaic became hyperpolarized and were surrounded by a thin ring of depolarized cells due to reduced lateral inhibition provided by the horizontal cells. This activity spread through both ON and OFF pathways, leading to depolarized ON bipolar cells and spiking ON RGCs,

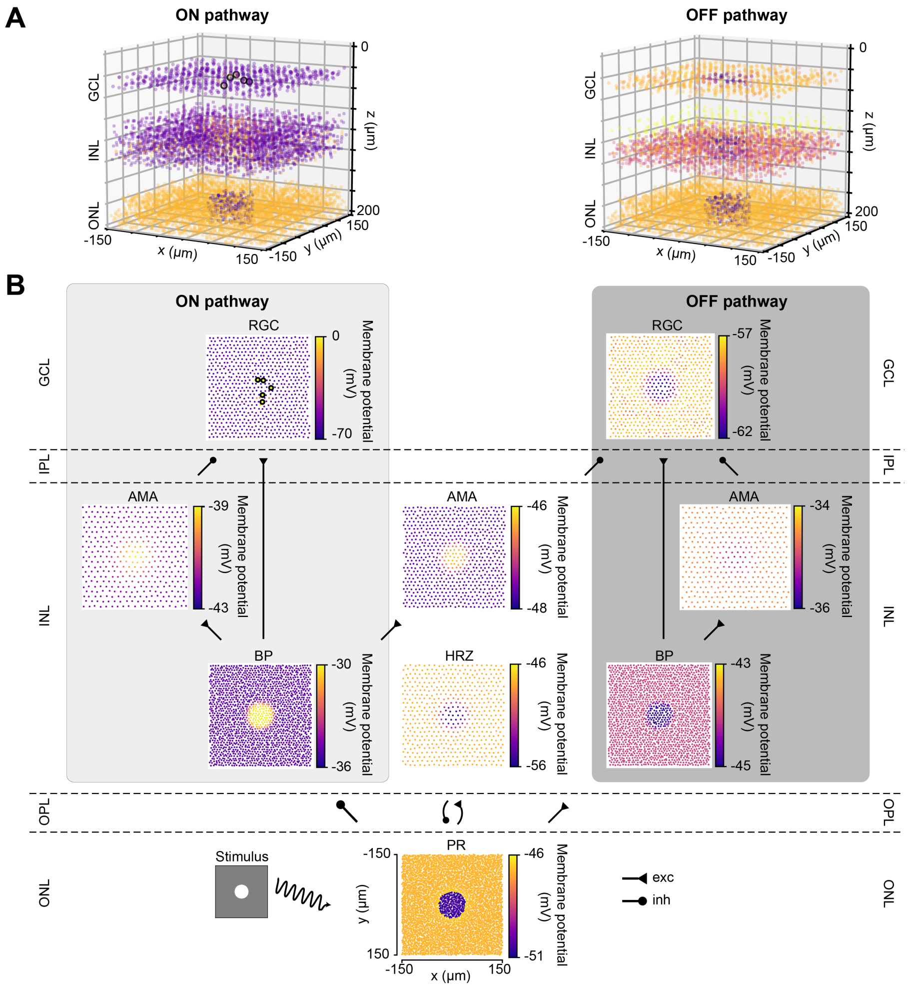

Figure 3.2: The light response of the retinal network, presented both in 3D and by cell type. Abbreviations same as in Fig. 3.1. **A)** The light response presented separately for the ON pathway (left) and the OFF pathway (right) in 3D. The light stimulus that was used to elicit the response was a bright disk (40 μm in radius with light intensity 1) placed at the center of a gray background (300 μm by 300 μm with light intensity 0.5) (illustrated by the bottom left inset of *B*). The light response shown occured 110 ms after stimulus onset. Each circle represents the (x, y, z) location of a neuron, and the color of each circle indicates the membrane potential of each neuron, with the color bars in *B*. Enlarged circles with black border indicate neuronal spikes. The plot of the ON pathway includes cone photoreceptors, horizontal cells and all the cells belonging to the ON pathway, and the plot of the OFF pathway includes cone photoreceptors, horizontal cells and all the cells belonging to the OFF pathway. **B)** The light response presented by cell type, corresponding to each layer in *A*. Each circle represents the x-y location of a neuron, and the color of each circle indicates the membrane potential of each neuron. Enlarged circles with black border indicate neuronal spikes. For an animated version of this figure, see Supplemental Video 1.

whereas corresponding cells in the OFF pathway were hyperpolarized.

The spatiotemporal evolution of RGC activity in response to the above mentioned stimulus is shown in Fig. 3.3, which was modeled after Fig. 5 in [103]. The stimulus mentioned above was modulated in time by a square wave signal (200 ms phase duration) at four contrast levels: 100 %, 50 %, −50 %, −100 %. Consistent with conduction delays in the rabbit retina [115], ON RGCs first fired roughly 20 ms after stimulus onset, whereas OFF RGCs took 50 ms to respond (Fig. 3.3A–B). RGC unaffected by the stimulus exhibited at a 2 Hz spontaneous firing rate, which was achieved by setting the conductance of the leakage current (see Section 3.2.1). Synchronization of firing increased with stimulus strength for both ON and OFF populations. The center-surround structure of RGC receptive fields is evident in Fig. 3.3C–D.

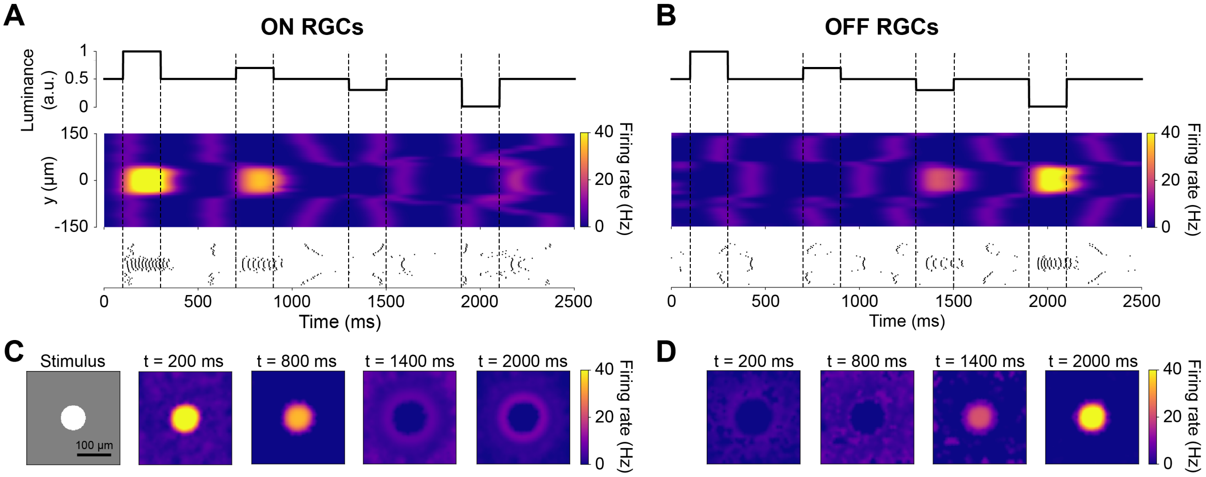

Figure 3.3: The spatiotemporal response of RGCs to a temporally varying light stimulus (inspired by Cottaris and Elfar, [103]). The light stimulus (illustrated in the bottom-left inset) was a disk (40 μm in radius) placed at the center of a gray background (300 μm × 300 μm with light intensity 0.5), varying in intensity over time. **A–B)** Spatiotemporal profile of RGC firing rate for neurons located at $x = 0\,\mu\text{m}$, visualized both as a heatmap (smoothed with a 50 ms Gaussian sliding window) and raster plot. The vertical dotted lines indicate the time of change in light intensity of the disk stimulus. **C–D)** Spatial activity profiles of RGC firing rate taken at different points in time.

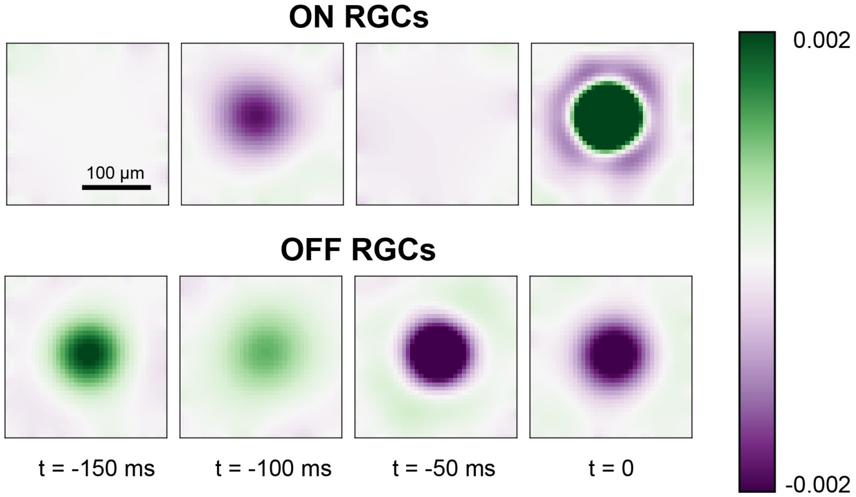

Figure 3.4: Generalized linear model fit for ON cells (*top*) and OFF cells (*bottom*). The colorbar represents the range of values of the linear filters k_i (see 3.2.3) at each spatial and temporal location, where green indicates excitatory values and purple indicates inhibitory values.

3.3.2 Receptive field of single RGCs

To ensure RGCs in the model have expected spatiotemporal receptive field profiles, we fit a generalized linear model to the spiking response of RGCs to a “cloud” stimulus consisting of spatiotemporal Gaussian white noise filtered with a two-dimensional spatial Gaussian filter (Shi et al., [116]; for details see Section 3.2.3).

The fitted spatiotemporal receptive field of two example cells located at the center of the simulated retinal patch is shown in Fig. 4.3. As expected, the receptive field profile of the healthy ON cell showed a clear excitatory center and an inhibitory surround at the time of a spike ($t = 0\,\mathrm{ms}$), with reversed polarity at $t = -100\,\mathrm{ms}$.

The receptive field profile of the OFF cell also exhibited a center-surround structure, though the excitatory surround was most pronounced at $t = -50\,\mathrm{ms}$ due to the longer synaptic delay of the OFF pathway (see Section 3.2.2). This also led to a prolonged response at the center of the receptive field profile.

3.4 Discussion

In this work, we present a biophysically-detailed computational model of the cone pathway in the retina. Our model represents retinal neurons across distinct layers in a three-dimensional space, faithfully replicates the feed-forward and lateral circuitry in the retina, and incorporates essential ion channels in each RGC neuron to accurately reflect spiking dynamics. We make the simultaneous simulation of thousands of neurons feasible by using GeNN. We show that our model can reproduce seminal findings about the light response of RGCs in ON and OFF pathways, at the single-neuron level and at the population level.

Compared to descriptive models produced by direct data fitting such as linear-nonlinear-Poisson (LNP) and convolutional neural network (CNN), our model has several advantages. First, it contains intermediate model neurons that mirror real-life counterparts, thus enabling future investigations into network responses under local or global changes in network connections. Second, the spatially accurate arrangement of neurons in the model permits the simulation of electrical stimuli delivered by electrodes, with flexibility in location, size and shape, which are crucial free parameters in neuro-engineering applications. Last but not least, our model outputs spikes with realistic temporal dynamics, facilitating studies in both rate coding and temporal coding.

In conclusion, our model provides a solid framework for studying the retina's response to electrical stimulation and introducing degenerative changes to the retina, critical elements in modeling for developing neuro-engineering devices. Those aspects will be examined in depth in Chapter 4.

Chapter 4

A computational model of retinal degeneration

4.1 Introduction

Understanding how the retina responds to light and electrical stimulation is a key issue for neuroscience and neuroengineering applications such as retinal prostheses. The retinal network undergoes drastic neuroanatomical alterations during retinal degeneration [40, 41] such as retinitis pigmentosa, which are of clinical importance to rehabilitative strategies such as retinal prostheses [117, 118, 119]. These alterations are complex and multi-stage [40], starting with photoreceptor stress that leads to outer segment truncation (Phase I) and progressive photoreceptor cell death (Phase II), followed by a protracted period of global cell migration and cell death (Phase III). The consequences of these alterations on retinal ganglion cell (RGC) firing are manifold, which include hyperactivity [120, 121, 122, 123], emergence of oscillations [124, 125], and increased electrical stimulation thresholds [126, 127, 128, 129, 130]. Oscillations are thought to arise from the network of electrically-coupled AII amacrine cells and ON cone bipolar cell [131, 132, 125]. Previous research has also identified retinoic acid as a trigger for hyperactivity [123].

Previous computational work modeled retinal degeneration, but often stopped short of simulating the global retinal remodeling typified by the progressive nature of these diseases. For instance, [103] built a model of the healthy retina and removed the cone population without addressing biophysical changes to the inner retina. [133] simulated degeneration by removing a fraction of simulated neurons, increasing connectivity among the surviving neurons, and increasing the noise level, but did not address the progressive nature of these diseases. Other models stopped at reducing the thickness of different retinal layers [106, 134] or hard-coded known physiological changes, such as increased spontaneous activity, into their model [135]. To the best of our knowledge, a comprehensive computational model of retinal degeneration is still lacking.

To address this, we systematically introduced anatomical and neurophysiological changes to a biophysically detailed network-level model of the retina (detailed in Chapter 3) and studied their effect on network activity. In early phases of this simulated cone-mediated retinal degeneration, we found that reduced light sensitivity and subsequent death of cones differentially affected ON and OFF RGC firing: whereas the light response of ON RGCs diminished more quickly than that of OFF RGCs, the spontaneous firing rate of OFF RGCs steadily increased. In late phases of degeneration, we found that migration and progressive death of inner retinal neurons led to a steady increase in electrical activation thresholds of both ON and OFF RGCs, especially for epiretinal stimulation.

Our findings demonstrate how biophysical changes associated with cone-mediated retinal degeneration affect retinal responses to both light and electrical stimulation. A detailed model of the retina in health and disease has the potential to further our understanding of visual processing in the retina. It may also inform the design of retinal prostheses, for the effective treatment of inherited retinal degenerative diseases such as retinitis pigmentosa and age-related macular degeneration.

4.2 Methods

4.2.1 Modeling retinal degeneration

Inherited retinal degenerative diseases, specifically photoreceptor-initiated ones, are commonly described in the literature as progressing in three phases [40, 41]:

- Phase I starts with either cone or rod stress, which leads to the truncation of the outer segments. The population of the affected photoreceptors starts to decrease and their neurites start to extend.
- In Phase II, the other class of photoreceptors also start to die and extend their neurite. Cones continue to truncate. Muller cells move to the outer nuclear layer and start to seal off the retina from the choroid. This process is called subretinal fibrosis, which will later evolve to a glial seal. Horizontal cells begin to hypertrophy and extend their neurites, while rod and cone bipolar cells retract their dendrites.
- Phase III is a protracted period of cell death and leads to global retinal remodeling. In early Phase III, Muller cells hypertrophy and form the glial seal. Neurons start to die, while microneuromas start to form. They often contain active synapses despite lacking normal signaling abilities. In middle Phase III, progressive neuronal death and microneuroma formation continue, while the remaining neurons start to migrate. Specifically, amacrine and bipolar cells move to the inner plexiform and the ganglion cell layer, and amacrine cells and RGCs move to the glial seal. In late Phase III, the microneuromas regress with continued cell death, accompanied with

the hypertrophy of Muller cells and vessels.

To make the modeling of such a complex process feasible, we limited ourselves to the major neuroanatomical changes that may have an impact on RGC signaling, and introduced them in a systematic step-wise manner. These changes are summarized in Table 4.1.

	Changes in the retina	Changes in the model
Phase I/II	cone truncation	gradual decrease of G_{light} (Eq. 3.5)
	cone cell death	gradual decrease of cone population
Phase III	neuronal cell death	retain $0\,\%$ cones and horizontal cells
		gradual decrease of the population of bipolar and amacrine cells
	retinal remodeling	migration of bipolar, amacrine, ganglion cells

Table 4.1: Phases of retinal degeneration [40], with corresponding changes in the retina and in our model.

Phase I/II

Because our model retina was intended to simulate the parafoveal retinal region and thus did not include rods, we combined Phases I and II into Phase I/II, where we gradually reduced the cone population and shortened the outer segments of the surviving cones, the latter of which was modeled by a gradual reduction in the ceiling of a cone's light response, G_{light} (Eq. 3.5). To model disease progression over time (Fig. 4.1E), we assumed a linear reduction in cone segment length and cone population over time.

In inherited retinal degeneration, most cones die by the end of Phase I/II [41], and the few that remain lose the ability to communicate with the inner retina. Muller cells play an important role in this phase of retinal degeneration, as they tend to move around and start to seal off the retina from the choroid. Although we did not explicitly model the movement and hyperproliferation of Muller cells, we modeled their functional consequence on the cones, which is to isolate them from the other retinal cells. To model this combined effect, we removed all photoreceptors and horizontal cells, which marked the end of Phase I/II.

Phase III

To simulate global retinal remodeling in Phase III, we restricted ourselves to simulating cell death and migration.

First, we gradually reduced the population of bipolar and amacrine cells (but not RGCs). Second, according to the literature, a fraction of the surviving amacrine and bipolar cells tend to migrate to the inner plexiform layer and the ganglion cell layer, whereas amacrine cells and RGCs tend to migrate to the glial seal [40]. We simulated this by migrating a randomly chosen subset of cells to different layers. A subset of amacrine cells was moved to the horizontal cell layer ($z \in [100\,\mu\mathrm{m}, 128\,\mu\mathrm{m}]$, close to the hypothetical glial seal), the inner plexiform layer ($z \in [40\,\mu\mathrm{m}, 80\,\mu\mathrm{m}]$), and the ganglion cell layer ($z \in [25\,\mu\mathrm{m}, 39\,\mu\mathrm{m}]$) in equal proportions. Half of the migrating bipolar cells were moved to the inner plexiform layer and the other half to the ganglion cell layer. The migrating RGCs were moved to the horizontal cell layer. The z coordinates of the migrating cells were sampled from a random uniform distribution in the respective range of z values that make up the different retinal layers (Table 4.2), whereas x, y coordinates remained unchanged. Synaptic weights and delays were unaffected by these coordinate changes.

To model disease progression over time (Fig. 4.1E), we assumed a linear reduction in cell survival rate (from 100 % at the beginning of Phase III to 0 at the end of Phase III) and a linear increase in cell migration rate (from 0 to 50 %).

Our model did not include Muller cells or microneuromas. However, some of the modeled changes may be indirectly due to Muller cell activity, such as the progressive death of inner retinal neurons.

4.2.2 Modeling extracellular electrical stimulation

Extracellular electrical stimulation was assumed to be generated by a disk electrode with diameter α µm placed at (x_e, y_e, z_e). The extracellular electrical potential v_{ext} at location (x, y, z) was given by:

$$v_{\text{e}}(x,y,z) = \frac{2V_0}{\pi} \arcsin\left(\frac{2\alpha}{\sqrt{(r+\alpha)^2+d^2}+\sqrt{(r-\alpha)^2+d^2}}\right), \tag{4.1}$$

where V_0 was the electrical potential of the disk, $r = \sqrt{(x-x_e)^2+(y-y_e)^2}$, and $d = z - z_e$ [136]. v_{e} was converted to a current, i_{e} (see Equations 3.1, 3.3, and 3.6), as follows:

$$i_{\text{e}} = \frac{1}{2} G_{\text{ext}} \langle v_{\text{e}} \rangle, \tag{4.2}$$

where G_{ext} was a conductance (see Table 3.2) and $\langle v_e \rangle$ was the average of the absolute voltage differences between 500 uniformly sampled, diametrically opposing points on the neuron's spherical soma [137].

Although epiretinal electrodes are known to activate passing axon fibers [138, 139, 140], we did not include RGC axons in our simulations.

The epiretinal electrode was centered at $(x, y, z) = (0\,\mu\text{m}, 0\,\mu\text{m}, -2\,\mu\text{m})$. The subretinal electrode was centered at $(x, y, z) = (0\,\mu\text{m}, 0\,\mu\text{m}, 135\,\mu\text{m})$. Both electrodes had a diameter of $\alpha = 80\,\mu\text{m}$.

	$\mathrm{BP}_{\mathrm{ON/OFF}}$	$\mathrm{AMA}^{\mathrm{WF/NF}}_{\mathrm{ON/OFF}}$	$\mathrm{RGC}_{\mathrm{ON/OFF}}$
healthy $(z_{\min}, z_{\max})$	$(100, 128)$	$(80, 101)$	$(25, 39)$
degenerated $(z_{\min}, z_{\max})$	$(25, 39)$	$(25, 39)$	$(100, 128)$
	$(40, 80)$	$(40, 80)$	-
	-	$(100, 128)$	-

Table 4.2: The range of z coordinates $(z_{\min}, z_{\max})$ where bipolars, amacrines and RGCs could be found during degeneration, given in µm. The bipolar cells migrated to the inner plexiform layer and the ganglion cell layer. The amacrine cells migrated to the horizontal cell layer, the inner plexiform layer and the ganglion cell layer. The RGCs migrated to the horizontal cell layer.

4.3 Results

4.3.1 Retinal degeneration differentially affects the spontaneous firing of ON and OFF cells

After verifying the light response of the healthy retina model, we gradually introduced neuroanatomical and neurophysiological changes to the network in order to model retinal degeneration (Fig. 4.1). Retinal degenerative diseases are commonly described in the literature as progressing in three phases [40, 41], starting with photoreceptor stress that leads to outer segment truncation (Phase I) and progressive photoreceptor cell death (Phase II), followed by a protracted period of global cell migration and cell death (Phase III). To make the modeling of such a complex process feasible, we limited ourselves to the major changes that may have an impact on RGC signaling as outlined below (see Section 4.2.1 for details). Because our model did not include rods, we combined Phases I and II into Phase I/II, where we gradually reduced the cone population and shortened the outer segments of the surviving cones (Fig. 4.1A). The complete loss of photoreceptors and horizontal cells marked the end of Phase I/II (Fig. 4.1B). To simulate global retinal remodeling in Phase III, we gradually reduced the population of bipolar and amacrine cells while migrating a randomly chosen fraction of inner retinal neurons to different retinal layers (Fig. 4.1C). To model disease progression over time (Fig. 4.1E), we assumed a linear reduction in cone segment length and cone population during Phase I/II, and a linear reduction in cell survival rate as well as a linear increase in cell migration rate during Phase III.

Although we did not alter the inherent excitability of RGCs, the network-level changes described above had a profound impact on RGC activity. Fig. 4.1D shows the spontaneous firing rate of RGCs as a function of disease progression, measured over 2500 milliseconds and averaged across all RGCs. Whereas most ON RGCs were silenced as soon as the cone population and outer segment length dropped below 80 % of their initial values, OFF RGCs experienced a gradual increase in spontaneous firing rate throughout Phase I/II, which peaked at roughly 300 % of its initial value at the beginning of Phase III. After that, the average spontaneous firing rate of the OFF RGC population gradually dropped to zero, although individual cells differed greatly in their activation profiles. Despite the increased variability in mean activity, no bursting or oscillatory activity emerged, as evidenced by Poissonian inter-spike interval distributions (see Fig. A.1 in the Appendix). This could be attributed to the fact that our model did not incorporate direct electrical coupling between cells, which is believed to be responsible for the development of oscillatory behavior in retinal degeneration [131, 132, 125].

We identified the underlying mechanistic causes of the change in spontaneous firing rate and found that they differed for ON and OFF RGCs. Whereas ON cells increased their spontaneous firing mainly as a function of cone outer segment truncation (Fig. 4.1F), OFF cells were mainly affected by the size of the surviving cone population (Fig. 4.1G). The full range of light responses as a function of outer segment truncation and cone population size is given in Fig. A.2.

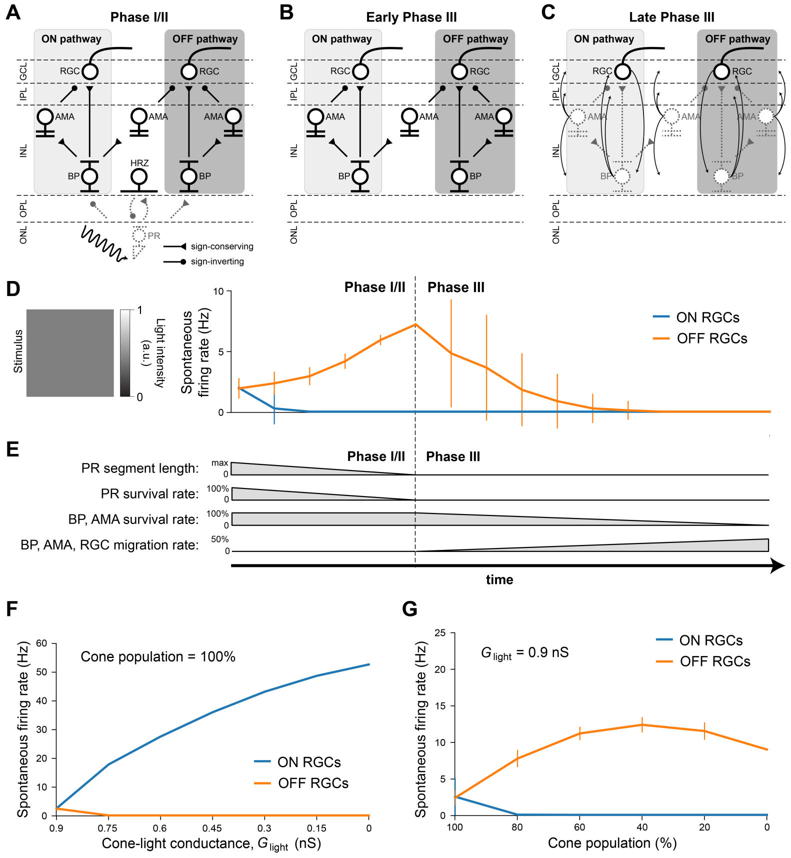

Figure 4.1: Simulating retinal degeneration. Abbreviations same as in Fig. 3.1. **A)** To simulate Phase I/II of retinal degeneration, we gradually shortened the cone outer segment length while simultaneously reducing the cone population. **B)** The complete loss of the cone population marks the beginning of Phase III. **C)** During Phase III, we gradually reduced the population of bipolar and amacrine cells. In addition, a fraction of surviving cells migrated to different layers: amacrine cells started to migrate to the horizontal cell layer, the inner plexiform layer, and the ganglion cell layer; bipolar cells started to migrate to the inner plexiform layer and the ganglion cell layer; RGCs started to migrate to the horizontal cell layer. At the end of Phase III, all inner retinal neurons have degenerated. **D)** Spontaneous firing rate of ON and OFF RGCs as a function of disease progression. Input was a full-field stimulus of $L(t) = 0.5$ light intensity (Eq. 3.5). Values indicate the spontaneous firing rate measured over 2500 milliseconds and averaged across all RGCs, with vertical bars indicating the standard deviation. **E)** To simulate disease progression over time, we assumed a constant rate of change for PR segment length and neuron survival. **F)** RGC spontaneous firing rate as a function of cone outer segment truncation (simulated as a reduction in the cone-light conductance G_{light}, see Eq. 3.5), averaged across RGCs, while the cone population size was held constant. **G)** RGC spontaneous firing rate as a function of the size of the cone population, averaged across the population of surviving RGCs, while G_{light} was held constant.

4.3.2 Light response of ON cells decreases more quickly than that of OFF cells during degeneration

To investigate how the light response of RGCs changed as a function of disease progression, we presented a constant stimulus in all stages of the disease (Fig. 4.2). The stimulus was a bright disk (40 µm in radius with maximal light intensity, $l(t) = 1$, Eq. 3.5) surrounded by a dark ring (40 µm in inner radius and 80 µm in outer radius with light intensity 0) placed on a gray background simulated with light intensity 0.5 (Fig. 4.2B, *left*). The stimulus was presented for 1000 milliseconds, during which the mean firing rate of each RGC was calculated.

Whereas both ON and OFF cells initially responded with similar firing rates, ON cells saw a much quicker reduction in firing rate during Phase I/II than OFF cells, remaining silent for the second half of Phase I/II and all throughout Phase III (Fig. 4.2A).

The spatial response profile at different time steps (where the center of the image is aligned with the corresponding time of disease progression on the x axis) is shown Fig. 4.2B. Whereas the center-surround structure of the retinal response is preserved during early stages of Phase I/II, spatial specificity is quickly lost during later stages of Phase I/II. During Phase III, it is not uncommon for the most active OFF cells to be found far away from the site of stimulation.

4.3.3 Ganglion cells quickly lose spatial selectivity during Phase I/II

To further illustrate the change in spatiotemporal receptive field profiles, we fit a generalized linear model to the spiking response of RGCs to a "cloud" stimulus consisting of spatiotemporal Gaussian white noise filtered with a two-dimensional spatial Gaussian filter (Shi et al., [116]; for details see Section 3.2.3).

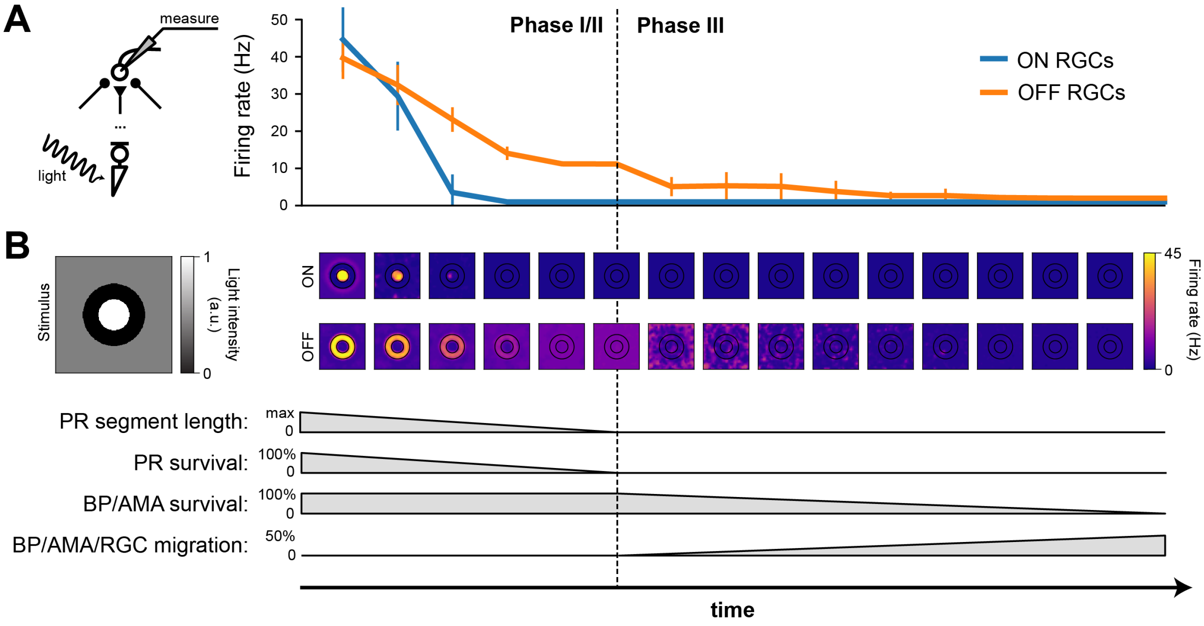

Figure 4.2: RGC firing rate (in response to light stimulation) decreases with the progression of degeneration. The stimulus was a bright disk (40 µm in radius with light intensity 1) surrounded by a dark ring (40 µm in inner radius and 80 µm in outer radius with light intensity 0) placed on a gray background (300 µm by 300 µm with light intensity 0.5). **A)** Firing rate of ON and OFF RGCs, measured over the 1000 ms stimulus presentation, plotted against time. The mean and standard deviation were calculated from the ON RGCs under the bright disk and the OFF RGCs under the dark ring. **B)** The spatial activity of ON RGCs (top row) and OFF RGCs (bottom row) plotted at different time points of degeneration. The borders of the bright center and dark surround are outlined in black.

The fitted spatiotemporal receptive field of two example cells located at the center of the simulated retinal patch is shown in Fig. 4.3. As expected, the receptive field profile of the healthy ON cell showed a clear excitatory center and an inhibitory surround at the time of a spike ($t = 0\,\mathrm{ms}$), with reversed polarity at $t = -100\,\mathrm{ms}$. In early Phase I/II (where cone population and outer segment length were at 80 % of their healthy values), inhibitory subregions at times $t < 0$ were much broader and much more pronounced, and were followed by an enlarged excitatory center that had lost its circular shape. In later stages of Phase I/II (where cone population and outer segment length were at 40 % of their healthy values), the ON cell was no longer sufficiently responsive to light stimulation

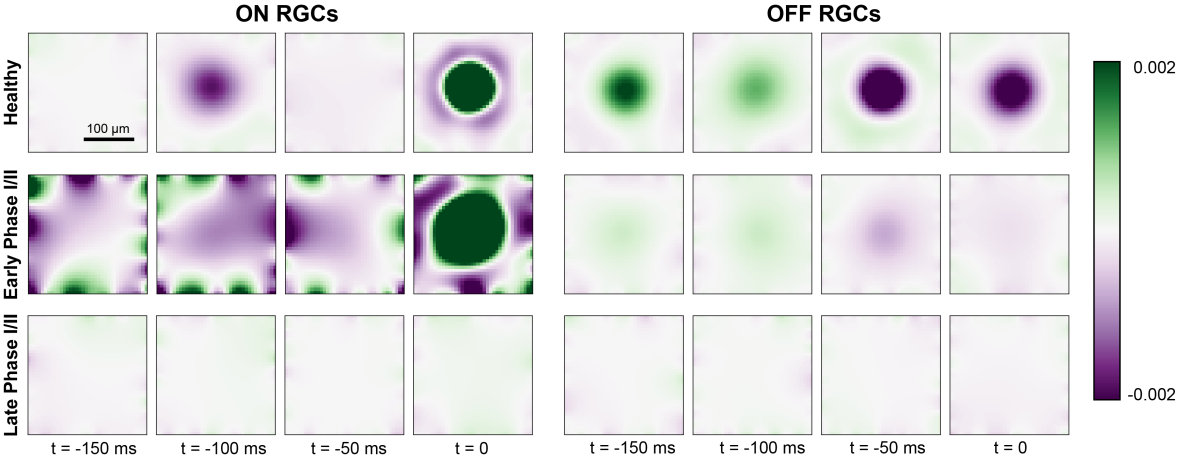

Figure 4.3: Generalized linear model fit for ON cells (*left*) and OFF cells (*right*) for a healthy retina (*top row*), early Phase I/II (*middle row*; 100 % cones surviving and $G_{\text{light}} = 0.75$ for ON, 60 % cones surviving and $G_{\text{light}} = 0.6$ for OFF), and late Phase I/II (*bottom row*; 80 % cones surviving and $G_{\text{light}} = 0.75$ for ON, 40 % cones surviving and $G_{\text{light}} = 0.45$). The colorbar represents the range of values of the linear filters k_i (see 3.2.3) at each spatial and temporal location, where green indicates excitatory values and purple indicates inhibitory values.

and its spatial response profile was lost.

The receptive field profile of the OFF cell also exhibited a center-surround structure, though the excitatory surround was most pronounced at $t = -50\,\text{ms}$ due to the longer synaptic delay of the OFF pathway (see Section 3.2.2). This also led to a prolonged response at the center of the receptive field profile. In early Phase I/II, the spatial response profile of the OFF cell seemed to be largely preserved, although the overall response as weakened. In late Phase I/II, the OFF cell was no longer sufficiently responsive to light stimulation and its spatial response profile was lost.

4.3.4 Electrical thresholds increase throughout retinal degeneration

Understanding how the degenerated retina responds to electrical stimulation is crucial for treatment options such as retinal prostheses. We thus placed a simulated disk electrode

(80 µm) either epiretinally (i.e., 2 µm above the ganglion cell layer; Cottaris and Elfar [103]) or subretinally (i.e., at $z = 135\,\mu\text{m}$, close to horizontal and bipolar cells) and measured the RGC response to a 20 Hz cathodic-first biphasic pulse train of 1 s duration with 0.45 ms phase duration (Fig. 4.4).

Consistent with the literature [126, 127, 128, 129, 130], electrical thresholds in later stages of degeneration rose to 200 %–400 % of those in the healthy retina (Fig. 4.4). Here, threshold was defined as the smallest stimulus amplitude that elicited a spike on at least half of 20 trials [141, 81], and the resulting thresholds were averaged across the population of surviving RGCs.

Interestingly, our model predicted that degeneration should affect the electrical thresholds of ON and OFF RGCs differently: whereas thresholds for ON cells tended to rise rapidly in Phase I/II, OFF cell thresholds decreased throughout Phase I/II to a point where the threshold was effectively zero, due to increased spontaneous firing. As cells started to migrate in Phase III, thresholds rose again. Notably, epiretinal stimulation thresholds kept rising (Fig. 4.4A, reaching thresholds up to 400 % of those found in a healthy retina, whereas subretinal thresholds were more stable during Phase III and plateaued at around 200 % of the healthy thresholds (Fig. 4.4). The standard deviations in Phase III were significantly larger than those in Phase I/II because of the cell migration in Phase III.

In addition, the spatial response profiles were strongly affected by degeneration (heat maps in Fig. 4.4, here shown for a 20 Hz biphasic pulse train with 60 mA amplitude). A ring-like structure is noticeable in both ON and OFF cell responses due to the influence of the extracellular potential being highest along the edge of the electrode, allowing activity to spread far beyond the size of the electrode. Epiretinal stimulation was able to elicit solid OFF cell responses throughout Phase I/II, whereas ON cell responses lasted only halfway through. After that, cell migration started to disrupt spatial response profiles in

Phase III. The story was similar for subretinal stimulation, though a few differences were noticeable. First, ON cell responses vanished almost immediately in Phase I/II, only to come back in late stages of Phase III as more and more ganglion cells started to migrate closer to the subretinal electrode. Second, the spatial activation profile of OFF cells was more confined early in Phase I/II, but began to widen due to increased spontaneous RGC later in Phase I/II. Third, cell migration disrupted the spatial response profiles more quickly and more thoroughly, but continued to elicit responses all the way to the end of Phase III.

4.3.5 Cell death and migration affect ON and OFF cells differently

To isolate the network changes responsible for the altered electrical response properties of RGCs, we simulated frequency-current (F-I) curves at different stages of degeneration for three different stimulation modes: current injection, epiretinal electrical stimulation, and subretinal electrical stimulation (Fig. 4.5). During Phase I/II, cone death reduced the response of ON cells for all three stimulation modes (Fig. 4.5, *top row*), but left OFF cells unaffected. This is consistent with the retina's light response in Fig. 4.1F. During Phase III, bipolar and amacrine cell death reduced the response of OFF cells for all three stimulation modes, but left the ON cells mostly unaffected (Fig. 4.5, *top row*); the one exception being subretinally stimulated ON cells, which saw the greatest reduction in activity. Finally, cell migration affected the response of both ON and OFF cells for both epiretinal and subretinal stimulation, increasing response variability across the RGC population.

Overall, these results demonstrate how network-level changes may affect RGC firing at different stages of retinal degeneration.

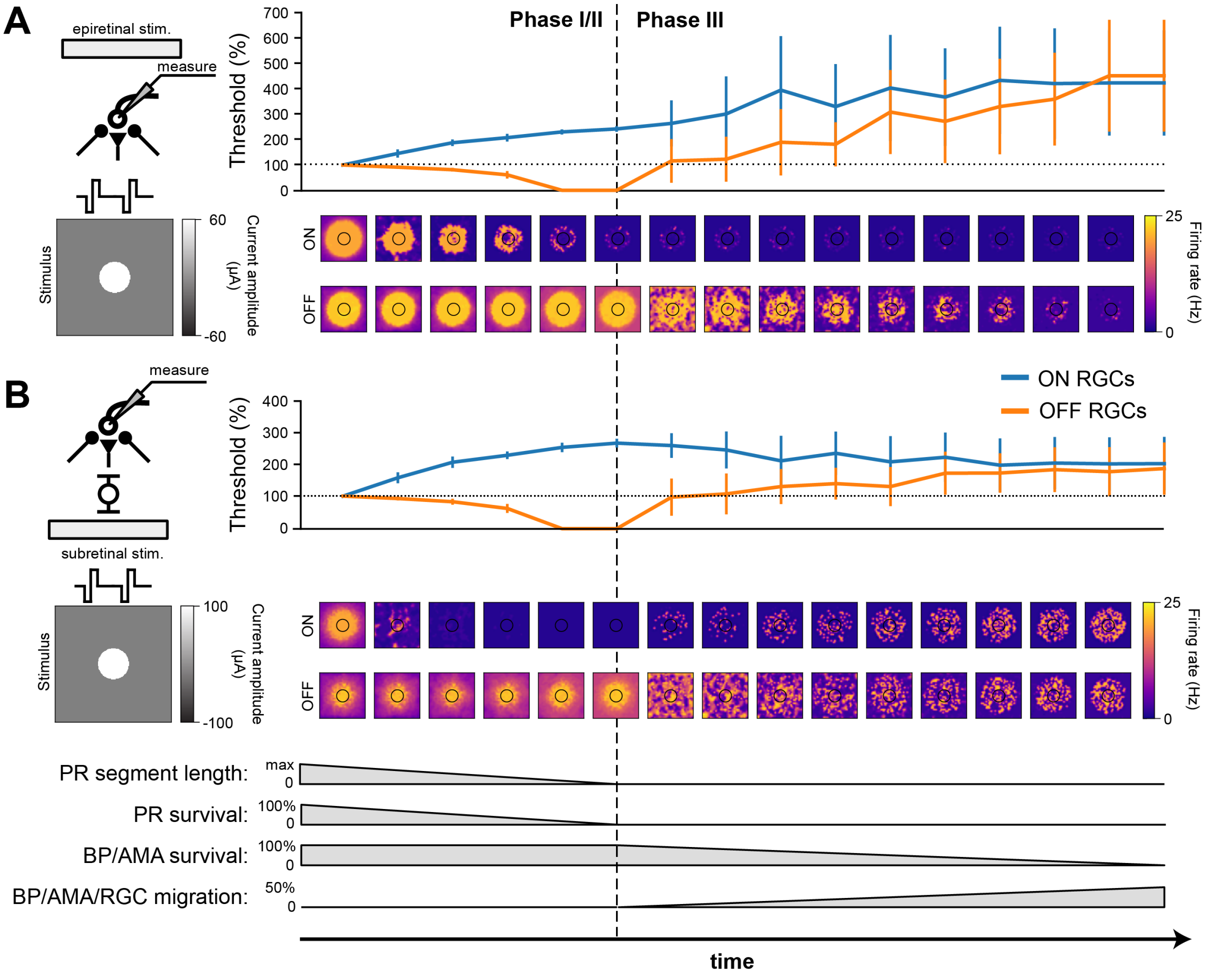

Figure 4.4: RGC response to electrical stimulation with a 20 Hz biphasic cathodic-pulse train (0.45 ms phase duration, 1 s stimulus duration). Mean values were calculated over 1000 ms (stimulus presentation) and averaged across neurons located directly below the electrode; vertical bars the SD. **A)** Epiretinal stimulation. Electrical stimulation thresholds during retinal degeneration reported relative to healthy thresholds (100 %, horizontal dotted line). The spatial response profile of ON and OFF RGCs is given below for a pulse train of 60 µA amplitude. The borders of the electrode are outlined in black. **B)** Subretinal stimulation. Electrical stimulation thresholds during retinal degeneration reported relative to healthy thresholds (100 %, horizontal dotted line). The spatial response profile of ON and OFF RGCs is given below for a pulse train of 60 µA amplitude. The borders of the electrode are outlined in black.

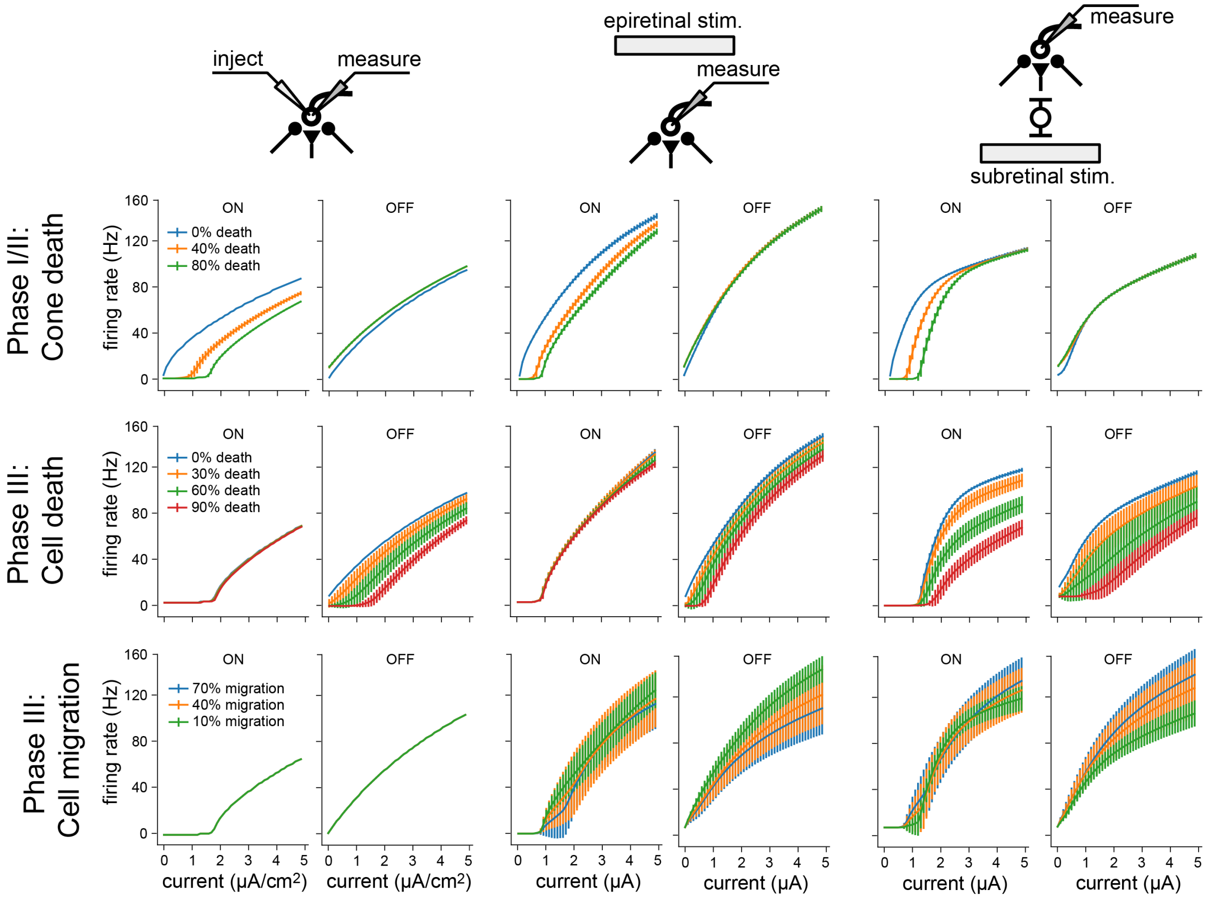

Figure 4.5: Frequency-current (F-I) curves for three modes of neuronal stimulation: current injection (*two leftmost columns*), epiretinal electrical stimulation (*two center columns*) and subretinal electrical stimulation (*two rightmost columns*). F-I curves are shown for RGCs at different stages of retinal degeneration: as a function of cone death during Phase I/II (*top row*), as a function of bipolar and amacrine cell death during Phase III (*middle row*), and as a function of bipolar, amacrine, and ganglion cell migration during Phase III (*bottom row*). Values averaged across RGCs; vertical bars are the standard deviation.

4.4 Discussion

We have developed a biophysically inspired *in silico* model of the cone pathway in the retina that simulates the network-level response to both light and electrical stimulation, and found that simulated cone-mediated retinal degeneration differentially affects ON and OFF RGCs. Existing computational models of retinal degeneration largely focus on RGC activity in the absence of photoreceptor input (e.g., Cottaris and Elfar, [103]; Golden et al., [133]), but do not consider the global retinal remodeling that may impact the responsiveness of RGCs. To this end, our simulations do not just reproduce commonly reported findings about the changes in RGC activity encountered during retinal degeneration (e.g., hyperactivity, increased electrical thresholds) but also offer testable predictions about the neuroanatomical mechanisms that may underlie altered RGC activity as a function of disease progression.

Consistent with the literature [120, 121, 142, 122, 123], we found that RGCs exhibited elevated spontaneous firing rates during retinal degeneration (Fig. 4.1). In our model this was mainly restricted to the OFF RGC population, which became more active over time (Fig. 4.1D, G). Similar observations have been made in degenerated retinas of mouse models [120, 121, 142], which are dominated by OFF cell activity. However, we identified photoreceptor cell death in Phase I/II as the main driving force behind this hyperactivity, as opposed to an intrinsic change to RGC excitability [123]. Moreover, the complete loss of cones did not drive RGCs into an oscillatory state (Fig. A.2). However, this may be due to the fact that our simulations did not include gap junctions (see below). On the other hand, ON cells showed increased activity in response to cone outer segment truncation (Fig. 4.1F), which may occur only early in degeneration before most photoreceptors are lost, after which their spontaneous activity is expected to quickly vanish.

As the light response of the RGC population slowly subsided (Fig. 4.2), ON cells saw a much quicker reduction in firing rate than OFF cells, remaining silent for the second half of Phase I/II and all throughout Phase III (Fig. 4.2). A generalized linear model revealed a brief broadening of the spatiotemporal receptive field for ON RGCs early in Phase I/II while these cells were losing their inhibitory surround, before the spatial properties of both ON and OFF receptive field were lost (Fig. 4.3). This is consistent with studies that have documented cell type–specific functional changes in RGCs across animal models [120, 142, 143], where spatial receptive fields often lose their circular shape and appear spotty before they vanish [143].

As degeneration progressed, electrical thresholds tended to increase for both subretinal and epiretinal stimulation (Fig. 4.5), which is consistent with most literature on the subject [126, 127, 128, 129, 130]—though see [141]. Mirroring the changes in the light response, ON cells also displayed diminished responses to electrical stimulation (Fig. 4.5). This resulted in higher activation thresholds for ON cells as compared to OFF cells (Fig. 4.4), which is a phenomenon previously documented in the degenerated mouse retina [144]. Interestingly, our model also predicts a brief period of degeneration during which OFF cells are so active that their electrical threshold is effectively zero. Furthermore, our model offers testable predictions of how cone death (Phase I/II), bipolar and amacrine cell death (Phase III), and cell migration (Phase III) affect the responsiveness of RGCs (Fig. 4.5).

Overall, our findings demonstrate how biophysical changes associated with retinal degeneration affect retinal responses to both light and electrical stimulation, which may have important implications for the design and application of retinal prostheses. In specific, our results suggest that spatially confined responses might be more easily elicited with subretinal stimulation in Phase I/II and with epiretinal stimulation in Phase III (Fig. 4.4). This implies a role for subretinal prostheses in early stages of the disease, whereas epiretinal prostheses may be more effective in later stages. However, cell migration might strongly affect both stimulation modes (Fig. 4.5), leading to spotty activation of the RGC population that may obscure the perceptual interpretation of these electrical stimuli.

Moreover, both subretinal and epiretinal stimulation are expected to more easily activate OFF cells, as they remain active through most of Phase III. This suggests that OFF cell activity may play a greater role in prosthetic vision than previously assumed [145], which could have important implications for the differential activation of RGC subtypes [146, 144, 147, 148].

Although our model captures a range of biophysical changes common to cone-mediated retinal degeneration, we were forced to make some simplifying assumptions due to the complex nature of the degeneration process. Most notably, our model did not include rod circuitry and gap junctions. This may explain why we did not observe oscillations in our model, since rod bipolar cells are thought to participate in an oscillatory network in the outer retina [131, 149] that may arise from electrically coupled networks [150, 131, 149, 132, 151, 125]. In addition, we did not consider it feasible to model the hyperproliferation of Muller cells, the formation of microneuromas, or other morphological changes commonly observed during inherited retinal degeneration [40].

Nevertheless, as we did not modify the intrinsic properties of the RGC population, our results suggest that commonly documented physiological changes such as RGC hyperactivity and increased electrical thresholds [152, 153, 123] may have additional network-mediate causes that are presynaptic to RGCs. This work thus offers testable predictions to further our understanding of retinal processing in health and disease. Future work could focus on adding additional morphological and topological detail to the simulation in order to obtain a more complete picture of the changes in RGC response properties associated with inherited retinal degeneration.

Chapter 5

Retinal scene statistics for freely moving mice

5.1 Introduction

Most computational models of early vision are trained on static stimuli (i.e., images) from a fixed dataset. However, the neural circuits of an organism are thought to adapt to the natural environment it inhabits due to evolutionary pressure. According to the efficient coding hypothesis, information processing in the visual system should be optimized for the statistics of the visual scenes from the natural environment [154]. Therefore, an accurate characterization of the visual stimuli encountered in the natural environment is imperative for a complete understanding of the visual system [155, 156, 157].

The mouse has become a prominent model organism in neuroscience due to its unparalleled experimental access to the mammalian cerebral cortex [158, 51], but most studies focus on using sparse artificial stimuli to probe the visual processing of head-fixed mice. Studying visual processing in mice in a more ecologically relevant setting is still in its nascent stages [159, 160, 161]. Even so, several studies have already highlighted the importance of using more naturalistic stimuli and revealed findings such as enhanced spatial frequency tuning in primary visual cortex (V1) neurons and respective sensitivity to green and UV in the dorsal and ventral retina [162, 163, 164]. However, few previous studies have considered the effect of head and eye movement on natural scene statistics; the mouse was either head-fixed or the footage analyzed was collected with a camera mimicking the animal's location in the natural environment.

To understand how the mouse uses head and eye movement to sample the natural scene, it is crucial to record the visual input from freely moving mice that exhibit naturalistic head-eye coordination. Mice explore their visual surroundings through a combination of head and eye movements, which can be classified into two types: gaze-shifting eye movements, where the eyes move in conjunction with the head, and compensatory eye movements, where the eyes move counter to the head's motion [165, 166, 167]. Through an analysis of the visual footage recorded from freely moving mice, we might be able to uncover visual scene characteristics resulting from eye, body, and head movements during natural locomotion, similar to studies conducted in humans [168].

We combined high-density silicon probe recordings with miniature head-mounted cameras, including one camera aimed outwards to capture the visual scene from the mouse's perspective and the other one aimed at the eye to measure the pupil position, as well as an inertial measurement unit (IMU) to quantify head orientation. Three mice equipped with this state-of-the-art recording system freely explored an arena for around 1 hour. To estimate the retinal input during free exploration, we followed previous work [15, 169, 55] to train a shifter network that learns an affine transformation to correct the footage for eye movements. The shifter network was trained end-to-end with a neural activity prediction model such that the shifts were optimized to produce the best fit in predicting the firing rates of the recorded V1 neurons. This allowed us to transform each frame of the recorded visual input into an eye-centered coordinate frame. Following [170], we characterized each frame of the head-centered and eye-centered video based on various spatiotemporal features (e.g., luminance, contrast, edge density). To determine the visual cues influencing gaze-shifting eye movements, we fine-tuned EfficientNet-B0 to predict the horizontal angle of the endpoint of each gaze-shifting eye movement. We then used saliency map analysis to demonstrate which pixels were most predictive of gaze-shifting eye movements.

5.2 Methods

5.2.1 Data acquisition and preprocessing

In this subsection, we provide a brief overview of the data collection and preprocessing. For a more complete and detailed description of the procedure, see [55].

The data from three adult mice was used in this study. Equipped with the state-of-the-art recording system designed for the freely-moving condition, each mouse was free to explore an approximately 48 cm long × 37 cm wide × 30 cm high arena for about 1 hour. The head-centric visual input, neural activity in V1, and various behavioral variables were available as simultaneous data streams for further analysis (Figure 5.1.A).

The recording system was composed of two head-mounted miniature cameras, a head-mounted IMU, and high-density silicon probes implanted into V1. One of the cameras (about 120° wide) was forward-facing and pointed towards the direction of the right eye, and it recorded the head-centric visual input at 16 ms per frame. The other miniature camera was pointed at the right eye, and it recorded a video of the right eye at 30 Hz. The IMU collected the gyroscope and accelerometer information at 30 kHz. The high-density silicon probe (11 µm × 15 µm, 128 channels), implanted in the left V1 center, acquired electrophysiology data at 30 kHz. An additional top-down camera recorded a video of the mouse moving around in the arena at 60 Hz.

The arena was composed of four walls and one floor. The floor of the arena was gray and scattered with black and white Lego blocks. One wall of the arena was a monitor, displaying a moving black and white dot stimulus. The other three walls presented static stimuli, including white noise and black and white gratings with high and low spatial frequency.

After data collection, the following data preprocessing procedure was performed. The head-centric visual input was downsampled to 60×80 pixels. DeepLabCut [171] was used to extract the horizontal eye position θ, vertical eye position ϕ, and pupil radius σ from the video of the right eye, and estimate the locomotion speed s from the top-down video. The pitch (ρ) and roll (ω) of the mouse's head were gathered from the IMU. Raw data traces from the electrophysiological recording were bandpass-filtered between 0.01 Hz and 7.5 kHz, and spike-sorted with Kilosort 2.5 [172]. Phy2 was used to select single units ([167]), and further processing was performed to remove inactive units (mean firing rate < 3 Hz). In the end, there were respectively 68, 32, and 49 active units for the three mice. All data streams were resampled at 20.83 Hz (48 ms per frame), and then were used to train a shifter network and a multimodal deep learning network end-to-end to predict the neural activity for each mouse.

5.2.2 Shifter network

To accurately analyze the scene statistics of the retinocentric visual input for freely moving mice, transformations need to be applied to the camera feed from the head-mounted camera so that the visual input is in retinal coordinates before analysis.

To solve this problem, we followed previous work in the field [15, 169, 55] and trained a shifter network to predict the correct transformation parameters based on recorded

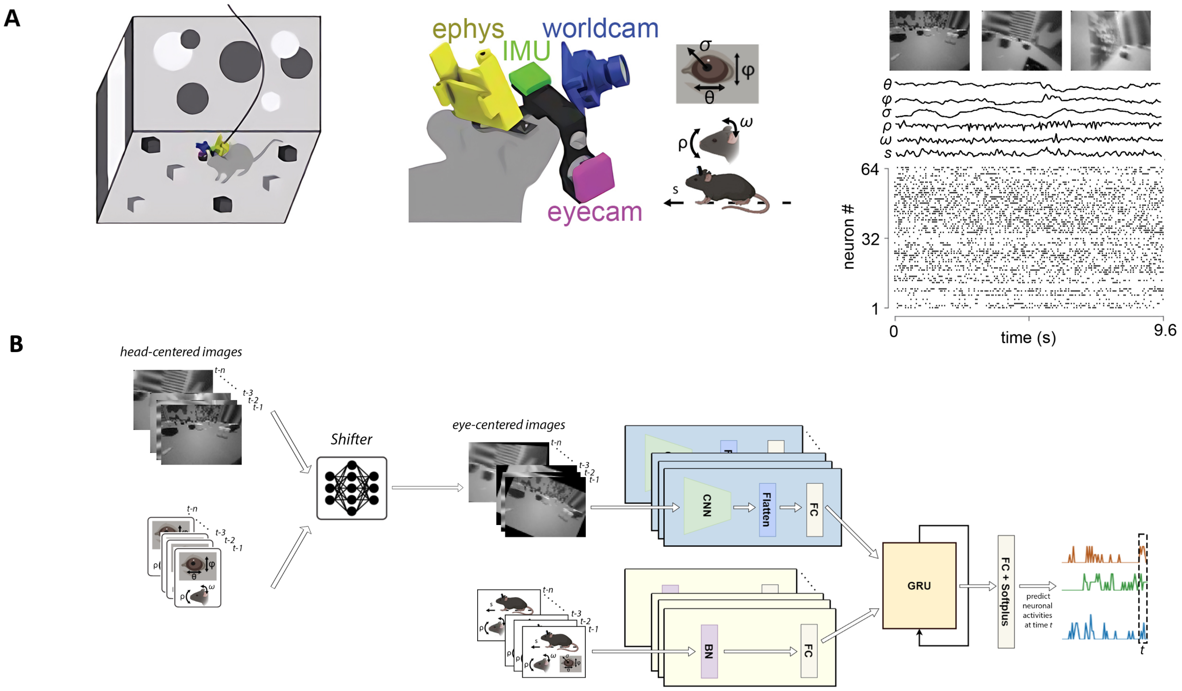

Figure 5.1: A) Schematic of the head-mounted recording system for freely moving mice (adapted from [55]). Left: Three mice freely explored a 48 cm long × 37 cm wide × 30 cm high arena. Middle: Preparation included a silicon probe for electrophysiological recording in V1 (yellow), miniature cameras for recording the mouse's eye position and pupil size (θ, ϕ, and σ; magenta), and visual scene (blue), and inertial measurement unit for measuring head orientation (ρ and ω; green). Right: Sample data from a 9.6 s period during free movement showing (from top) visual scene, horizontal and vertical eye position, pupil size, head pitch and roll, locomotor speed, and a raster plot of 64 units. B) The worldcam video was gaze-corrected based on measured eye movements and head movements with an affine transformation learned by a shifter network. The shifter network was trained end-to-end with a neural activity prediction model to maximize the accuracy of neural activity prediction. The neural activity prediction model will be explained in more detail in Chapter 6.

behavioral variables. Specifically, the shifter network consisted of three fully-connected layers which each were followed by a Tanh layer and a BatchNorm layer. The input to the shifter network included the horizontal eye position θ, the vertical eye position ϕ, the pitch of the head ρ, and the roll of the head ω. The shifter network predicted a rotation and a shift for each frame to convert it from head-centric to retinocentric coordinates. The rotation was bounded by ±36°, and the shift was bounded by ±16 pixels horizontally and ±12 pixels vertically. The shifter network was trained end-to-end with a state-of-the-art multimodal deep learning network to determine the transformation parameters that enabled the best prediction of activity in V1 for each mice [173] (Figure 5.1.B).

The multimodal deep learning network predicts V1 activity jointly for all recorded neurons per frame in a continuous fashion by extracting visual features with a 3-layer convolutional neural network (CNN) from the transformed stimulus frame and behavioral features from a novel behavioral encoder, fusing those two features, and then feeding the fused features into a gated recurrent unit (GRU) unit. Chapter 6 explains this model in more detail. This architectural design allowed for flexible inclusion of visual, behavioral and temporal dynamics for accurate neural activity prediction. Different lengths of history for the GRU were experimented with, and the shifter trained with the best-performing network was extracted for each session and used to obtain the eye-shifted videos, which were used for subsequent analysis.

5.2.3 Types of eye movements

Mice sample the visual scene by moving their head and eyes. Their eye movements can be categorized into gaze-shifting eye movements where the eyes move with the head or compensatory eye movements where the eyes move against the head [165, 166, 167]. Since different eye movements may result in different scene statistics of the visual scenes in retinal coordinates, we followed previous work to classify eye movements based on gaze velocity, which is defined as the sum of horizontal eye movement and head movement velocities [167]. Eye movements accompanying high head velocities (greater than $60\,^{\circ}/\mathrm{s}$) and high gaze velocities (greater than $120\,^{\circ}/\mathrm{s}$) were considered gaze-shifting, while those accompanying high head velocities (greater than $60\,^{\circ}/\mathrm{s}$) and low gaze velocities (less than

120 °/s) were considered compensatory.

5.2.4 Computation of scene statistics

To analyze the mouse's strategy of sampling the visual scene during free movement, we extracted the spatiotemporal features from each frame in the head-centered and eye-centered videos and compared those scene statistics from the corresponding head-centered frame and eye-centered frame. The spatiotemporal features included luminance, contrast, edge density, and Difference of Gaussian (DoG) entropy [170]. Prior to the computation of spatiotemporal features, we normalized the pixel values in both the head-centered video and the eye-centered video to the range $[0, 1]$. For the eye-centered video, we masked the pixels that became out of the camera's field of view after the transformation predicted by the shifter network was applied to exclude those invalid pixels from further analysis. In addition, we computed those scene statistics separately for the set of pixels belonging to the whole frame, the top half frame, and the bottom half frame. The respective analysis of the top and bottom visual field was inspired by the fact that the mouse retina might have specialized circuitries for the top and bottom half, speculated to have evolved in response to the visual features of its natural habitat due to survival pressure [163, 164].

Luminance For each set of pixels in a given frame, we calculated the average luminance, which was simply the mean of the pixel grayscale values.

Contrast For each set of pixels in a given frame, we calculated the average contrast, which was defined as the standard deviation of the pixel grayscale values divided by the mean of the pixel values.

Edge density For each set of pixels in a given frame, we extracted the edges from the frame with a Canny edge detector with the `cv2` python package and then computed the edge density, which was defined as the number of nonzero pixels in the output of the edge detection divided by the total number of pixels.

DoG entropy For each set of pixels in a given frame, we computed its DoG entropy [170]. Firstly, DoG was computed by filtering the image with two different Gaussian kernels with the size of 11 pixels and standard deviations of 1 and 1.61 pixels respectively using the cv2 python package, and taking the difference between the two filtered images. Then, the Shannon entropy was computed with the difference image using the scikit-image python package. DoG can indicate the presence of texture [170].

5.2.5 Eye position prediction

In order to understand what visual features drove gaze-shifting eye movements, we trained a deep convolutional network to predict the horizontal endpoint of each gaze shift based on the video frames preceding the gaze shift and then used different attribution methods to visualize the visual features that contributed to the prediction.

We replaced the last linear layer of the pretrained EfficientNet-B0 and finetuned the model with a mean-squared error loss so that the model could predict the horizontal eye position according to the three video frames preceding the gaze shift. The three video frames were used as separate channels in the input image to the network. The dataset was randomly split so that 80%, 10%, and 10% of the data were used as the train, validation, and test set respectively. The model was optimized with Adam (batch size: 64, learning rate: .0002) for a maximum of 100 epochs with early stopping on the validation set (patience: 10 epochs), and the model that achieved the best validation loss was saved for attribution analysis.

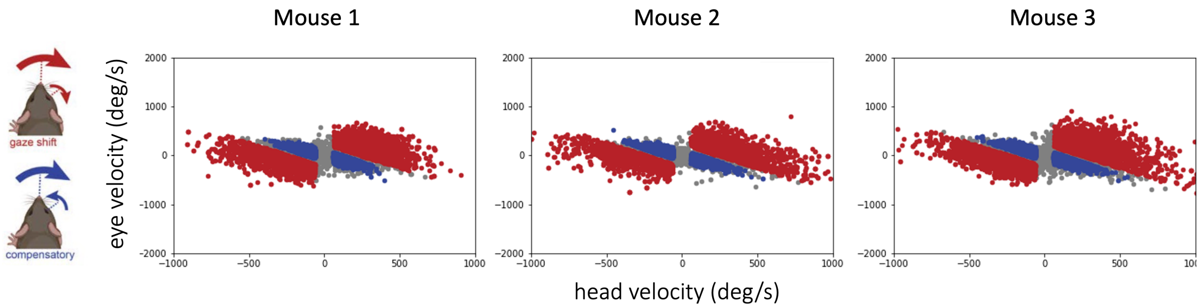

Figure 5.2: Most of the eye movements of all three mice can be classified into either compensatory or gaze-shifting eye movements. Each dot represents an eye movement during the recording. Red dots represent gaze-shifting eye movements, and blue dots represent compensatory eye movements.

We used a variety of attribution algorithms including the saliency map [174], Guided Grad-CAM [175], deconvolution [176], and guided backpropagation [177] to visualize which pixels contributed to the prediction of the horizontal landing position of gaze shifts. Those visualizations were computed with the whole dataset using the Captum package [178].

5.3 Results

5.3.1 Eye movements in freely moving mice can be categorized into compensatory and gaze-shifting types

Following previous work, we categorized the mouse's eye-head movements into compensatory eye movements and gaze-shifting eye movements (see Section 5.2.3). Figure 5.2 shows that for each mouse, most of the eye movements can be classified as either compensatory eye movements or gaze-shifting eye movements. This suggests that although the mice were equipped with head-mounted recording devices, our experimental setup was natural enough for the mice to make head and eye movements to sample the visual scene effectively.

5.3.2 The shifter network enables accurate prediction of neural activity

Whether the shifter network improves the prediction of neural activity is an important indicator of its correctness in transforming the video frames from head-centered coordinates to retina-centered coordinates [55]. Therefore, we compared the predictive performance of the model with the shifter and that of the model without the shifter. To determine the best model for each mouse, we tried different lengths of history (from 48 ms to 384 ms) for the GRU unit inside the model.

	Mouse 1		Mouse 2		Mouse 3	
History	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓
48 ms	.626 ± .145	.0574	.487 ± .180	.0931	.580 ± .133	.0849
96 ms	.631 ± .149	.0560	.474 ± .184	.0981	.592 ± .132	.0840
144 ms	.625 ± .154	.0597	.505 ± .174	.0897	.591 ± .133	.0839
192 ms	.625 ± .153	.0581	.482 ± .174	.0965	.588 ± .129	.0853
240 ms	.630 ± .152	.0556	.484 ± .179	.0990	.576 ± .136	.0851
288 ms	.625 ± .156	.0564	.508 ± .179	.0908	.575 ± .137	.0865
336 ms	.622 ± .155	.0586	.469 ± .192	.0981	.579 ± .137	.0860
384 ms	.628 ± .151	.0569	.517 ± .182	.0886	.580 ± .143	.0866

Table 5.1: The model trained with different lengths of history and without the shifter network. *cc*: cross-correlation, mean ± standard deviation (↑: the higher the better), MSE: mean-squared error (↓: the lower the better).

The predictive performance of the model without the shifter and with the shifter is shown in Table 5.1 and Table 5.2, respectively. The shifter network improved the prediction of the neural activity in all three mice under different lengths of temporal history in terms of both cross-correlation and mean-squared error, suggesting that the shifter network successfully predicted the transformation parameters to convert head-centered

	Mouse 1		Mouse 2		Mouse 3	
History	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓
48 ms	.637 ± .143	.0536	.494 ± .161	.0901	.603 ± .131	.0804
96 ms	.647 ± .139	.0551	**.553 ± .168**	.0850	.605 ± .134	.0824
144 ms	.647 ± .137	.0555	.523 ± .174	.0875	.606 ± .136	.0791
192 ms	.648 ± .138	.0553	.541 ± .173	**.0830**	.599 ± .133	.0796
240 ms	.653 ± .142	.0525	.534 ± .172	.0881	.606 ± .137	.0787
288 ms	.651 ± .139	.0530	.549 ± .171	.0857	.597 ± .141	.0800
336 ms	.650 ± .139	.0538	.543 ± .167	.0857	**.613 ± .136**	**.0781**
384 ms	**.657 ± .140**	**.0513**	.531 ± .174	.0868	.603 ± .134	.0799

Table 5.2: The model trained with different lengths of history and with the shifter network. Best performing networks are indicated in bold. *cc*: cross-correlation, mean ± standard deviation (↑: the higher the better), MSE: mean-squared error (↓: the lower the better).

frames to retina-centered frames based on the eye positions and head positions.

The amount of temporal information needed for the model to reach the best performance in terms of cross-correlation was 384 ms, 96 ms, 336 ms for Mouse 1, Mouse 2 and Mouse 3, respectively. We extracted the shifter network from the best model for each mice, and then used the shifter network to transform each frame collected from the head-mounted camera (referred to as "head-centered" in subsequent sections) to an retina-centered coordinate frame (referred to as "eye-centered" in subsequent sections).

5.3.3 Average luminance and contrast are unaffected by eye movements

To understand the difference between the head-centered frames and the eye-centered frames, we started out by analyzing the low-level features including the average luminance and the average contrast of each frame. We calculated and compared the average

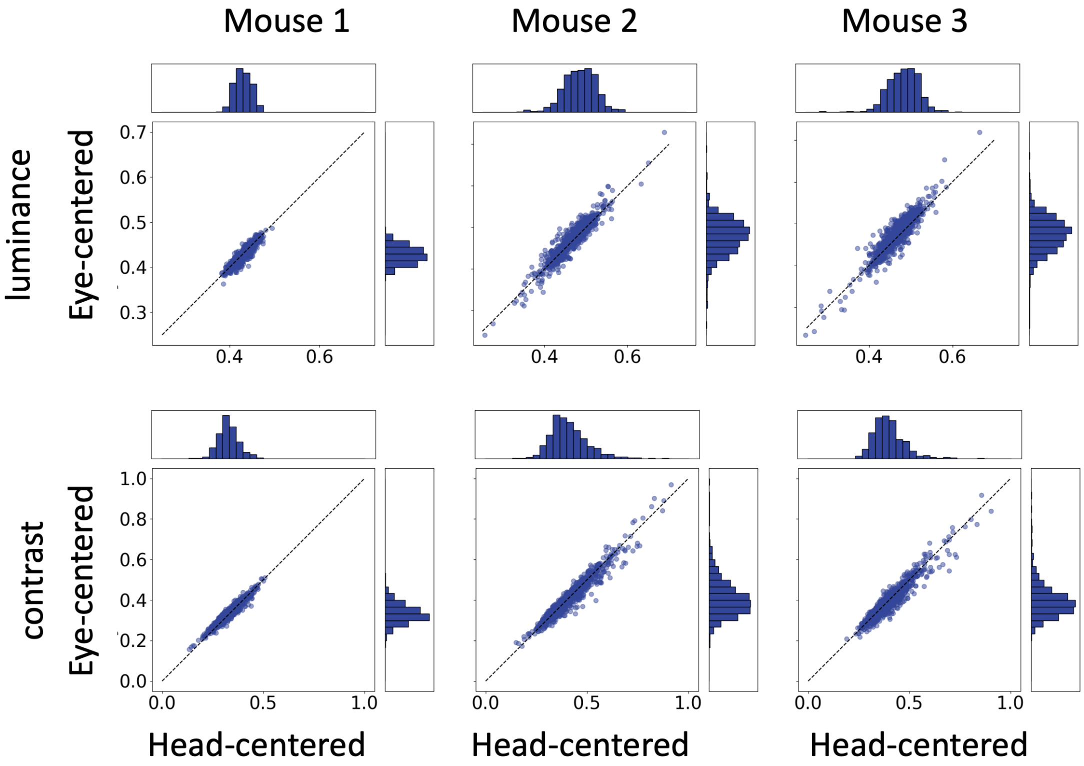

Figure 5.3: A comparison of luminance (top) and contrast (bottom) between head–centered frames and eye-centered frames. Each scatterplot shows a random subset of 1000 pairs of frames. Each dot in the scatterplot represents a pair of head-centered and eye-centered frames. The distribution of the values in the scatterplot is visualized in the accompanying histograms.

luminance value and the average contrast value for each pair of the head-centered frame and the eye-centered frame (Figure 5.3). There was no noticeable difference between the head-centered frames and the eye-centered frames in terms of either luminance or contrast across the three mice. This suggests that the mouse does not make preferential choices regarding luminance and contrast when sampling the visual scene.

5.3.4 Edge density and DoG entropy are increased by eye movements

Next, we wondered whether the mouse would make eye movements that are preferential for higher-order features including edge density and DoG entropy, and whether those higher-order features would differ in the top and the bottom half of the visual field. To do this, we calculated and compared edge density and DoG entropy for each pair of the head-centered frame and the eye-centered frame. To delineate the difference between the top and the bottom half of the visual field, we not only compared edge density and DoG entropy between the whole head-centered frame and the whole eye-centered frame, but also respectively contrasted the top half of the head-centered frame and the top half of the eye-centered frame, and the bottom half of the head-centered frame and the bottom half of the eye-centered frame.

The result of edge density comparison is shown in Figure 5.4. The edge density values in the eye-centered frames are higher than those in the head-centered frames across the three mice, indicating that the mouse makes eye movements that are preferential for the presence of edges. In addition, in the lower half, the difference in edge density between head-centered frames and eye-centered frames is more pronounced than that in the upper half across the three mice. This result might indicate that the mouse's lower visual field is more devoted to processing edges compared to its upper visual field.

The result of DoG entropy comparison is shown in Figure 5.5. The DoG entropy values in the eye-centered frames are higher than those in the head-centered frames. This result suggests that the mouse makes eye movements are preferential for the presence of texture. However, there is no prominent difference between the upper visual field and the lower visual field in terms of DoG entropy.

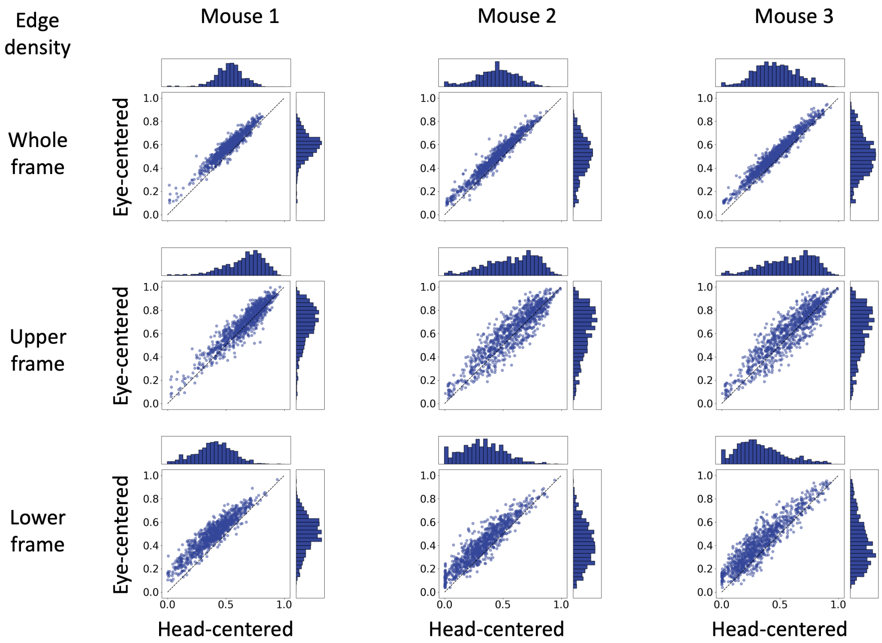

Figure 5.4: A comparison of edge density between head-centered frames and eye-centered frames in terms of the whole frame (top), the upper half frame only (middle) and the bottom half frame only (bottom). Each scatterplot shows a random subset of 1000 pairs of frames. Each dot in the scatterplot represents a pair of head-centered and eye-centered frames. The distribution of the values in the scatterplot is visualized in the accompanying histograms.

5.3.5 Upper peripheral visual field drives gaze-shifting eye movements

We fine-tuned a pretrained EfficientNet-B0 model to predict the horizontal eye position after each gaze-shifting eye movement at time t based on the video frames at time $t-3$ (144 ms before the gaze shift), $t-2$ (96 ms before the gaze shift) and $t-1$ (48 ms before the gaze shift). We trained the model for the eye position prediction task with the data from each mouse separately and all the data from all three mice together, and the R^2 score is summarized in Table 5.3. When the model was trained separately for each mouse, it severely overfit the training data. When the model was trained with all the data, the R^2 score was similar across the training, validation, and testing data, and the R^2 score of

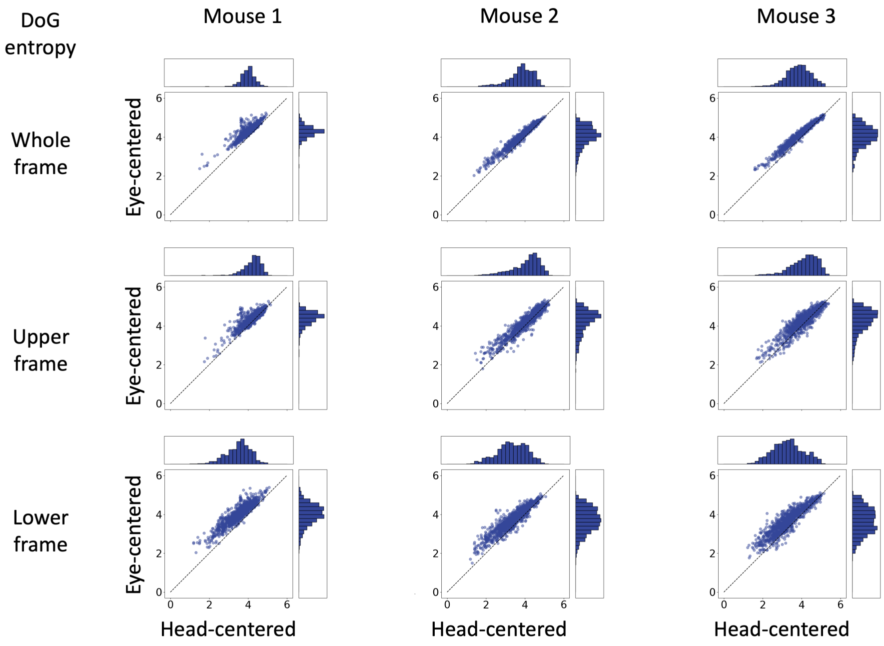

Figure 5.5: A comparison of DoG entropy between head-centered frames and eye-centered frames in terms of the whole frame (top), the upper half frame only (middle) and the bottom half frame only (bottom). Each scatterplot shows a random subset of 1000 pairs of frames. Each dot in the scatterplot represents a pair of head-centered and eye-centered frames. The distribution of the values in the scatterplot is visualized in the accompanying histograms.

the test set was the highest. Therefore, we used the model trained with all the data to perform the subsequent attribution analysis.

We computed the saliency map of the three video frames preceding each gaze-shifting eye movement, and computed the mean of all the saliency maps (Figure 5.6). The

	Training R^2	Validation R^2	Testing R^2
Mouse 1	.990	.305	.397
Mouse 2	.990	.518	.541
Mouse 3	.985	.512	.511
All	.681	.508	.544

Table 5.3: R^2 score of eye position prediction from training separately for each mouse and training together for the data from all mice.

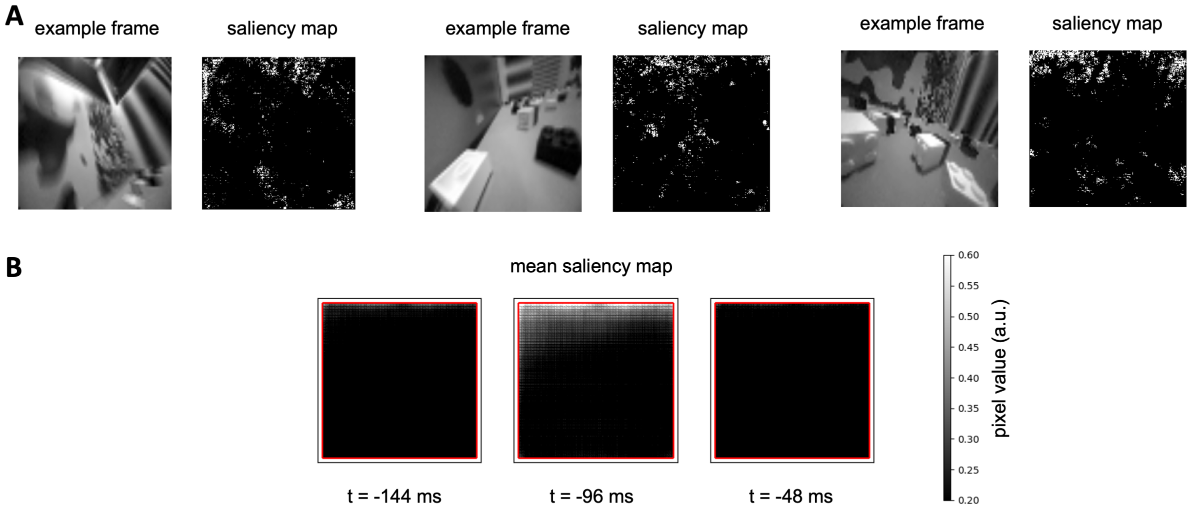

Figure 5.6: Saliency maps generated from the eye position prediction model trained from all data. A) Three example pairs of the original video frame and its corresponding saliency map. B) The mean of all the saliency maps, computed separately for the frames 144 ms, 96 ms and 48 ms before gaze shifts. The borders of the saliency maps were cropped by 6 pixels (indicated by the red box) to avoid border artifacts.

brightness value of a pixel in the saliency map indicates how strong the pixel contributed to the prediction of the landing position of the gaze-shifting eye movements. In Figure 5.6.A, three example pairs of the original video frame and its corresponding saliency map are shown, and in those examples the top parts of the saliency maps were consistently highlighted. Indeed, an inspection of the mean of all the saliency maps (Figure 5.6.B) reveals that the video frame 96 ms before the gaze shift contributed the most to the eye position prediction, and in particular the upper peripheral visual field was the most important region. The highlighted region appeared slightly slanted, which may be as a result of the data collection from one eye exclusively. Other attribution methods (Guided Grad-CAM [175], deconvolution [176], and guided backpropagation [177]) produced similar visualization (Appendix A.2.1). This result suggests that a major part of the gaze-shifting eye movements in mice might be devoted to detecting salient regions in the upper peripheral visual field.

5.4 Discussion

In this work, we have proposed to analyze the statistics of the visual scenes during the natural locomotion of the mouse by utilizing a state-of-the-art recording system allowing for simultaneous recording of visual, behavioral, and neural data and training a shifter network to estimate the retinal input by predicting a correcting transformation based on the head and eye positions. Consistent with previous studies, we found that the mouse made two distinct types of eye-head movements and that the shifter network enabled more accurate prediction of neural activity. A frame-based analysis of the head-centered video and the eye-centered video revealed that eye movements did not affect luminance or contrast, but interestingly increased edge density and DoG entropy. Via an attribution analysis of the deep learning network that predicted the horizontal landing endpoint of gaze-shifting eye movements, we discovered that the upper peripheral visual field contributed the most to gaze-shifting eye movements. In conclusion, our work not only confirms previous results on the mouse's head-eye movement coordination but also offers new insights into the visual processing of freely moving mice.

Similar to prior research, we observed that mice made two different types of eye-head movements (compensatory and gaze-shifting) while they explored freely in the arena while wearing miniature recording devices. This indicates that our experimental setup facilitated natural eye and head movements in mice, enabling them to effectively sample the visual environment. It underscores the significance of employing a freely moving experimental paradigm.

Like previous research, we employed a shifter network to correct the recording captured by the head-mounted camera and transform the frames to retina-centric and found that the shifter network improved the accuracy of the prediction of neural activity. This result validates the correctness of our approach and furthermore demonstrates that the shifter network is a versatile tool that works with a variety of predictive models including the generalized linear model (GLM) [55], CNN [15, 97], and, in our case, a multimodal temporally-recurrent artificial neural network. The shifter network proves to be an important tool for analyzing the retinal input of mice since mice can't be trained to fixate.

A comparison between the head-centered frames and the eye-centered frames showed that eye movements did not change the luminance or contrast, but preferred more edges and more texture. This indicates that higher-level features such as edges and textures might be more important for visual processing in mice during natural locomotion compared to low-level ones. In addition, we found that eye movements increased the edge density more noticeably in the lower half of the visual field compared to the upper half of the visual field. This aligns with the efficient coding hypothesis: In the natural habitat of mice, the bottom half of the visual field mostly consists of grass, and during evolution, the bottom half of the visual field becomes more efficient for processing edge-related features. This could explain why mice like to focus on edges with the bottom half of their visual field during the exploration of the arena. However, we did not find a similar trend in terms of DoG. A possible explanation is that DoG indicates the presence of texture, but does not differentiate different types of textures. In the future, a more detailed texture analysis and a more realistic arena with various natural elements such as grass and tree bark could be beneficial in disentangling how eye movements affect textures with greater precision.

We found that the upper peripheral visual field 96 ms before the gaze-shifting eye movement was most predictive of the horizontal endpoint of the eye movement. This indicates that a large part of visual processing in mice during natural locomotion might be devoted to predator detection on a fast time scale. We did not observe that the pixels belonging to the lower part of the visual field contributed much to the prediction of the gaze-shifting position. Future work on eye movement analysis during tasks with ecological relevance, such as prey capture, could reveal different results [166].

Chapter 6

A multimodal deep learning model of V1 in freely moving mice

6.1 Introduction

Computational models have been crucial in providing insight into the underlying mechanisms by which neurons in the visual cortex respond to external stimuli. Deep convolutional neural networks (CNNs) have had immense success as predictive models of the primate ventral stream, in cases where the animal was passively viewing stimuli or simply maintaining fixation [31, 19, 179, 26, 27]. Despite their success, these CNNs are poor predictors of neural responses in mouse visual cortex [180], which is thought to be shallower and more parallel than that of primates [181, 182]. According to the best models in the literature [28, 25, 97, 98, 183, 184], the mouse visual system is more broadly tuned and operates on relatively low-resolution inputs to support a variety of behaviors [158]. However, these models were limited to predicting neural responses to controlled (and potentially ethologically irrelevant) stimuli that were passively viewed by head-fixed animals.

Movement is a critical element of natural behavior. In the visual system, eye and head movements during locomotion and orienting transform the visual scene in potentially both beneficial (e.g., by providing additional visual cues) and detrimental ways (e.g., by introducing confounds due to self-movement) [32, 185, 57]. Movement-related activity is widespread in mouse cortex [186, 33] and prevalent in primary visual cortex (V1) [34, 56]. For instance, V1 neurons of freely moving mice show robust responses to head and eye position [53, 54], which may contribute a multiplicative gain to the visual response [55] that cannot be replicated under head fixation. V1 activity may be further modulated by variables that depend on the state of the animal and its behavioral goals [60, 56, 33, 58]. However, how these behavioral variables may integrate to modulate visual responses in V1 is unknown. Furthermore, a comprehensive predictive model of V1 activity in freely moving animals is still lacking.

To address these challenges, we make the following contributions:

- We introduce a multimodal recurrent neural network that integrates gaze-contingent visual input with behavioral and temporal dynamics to explain V1 activity during natural vision in freely moving mice.

- We show that the model achieves state-of-the-art predictions of V1 activity during free exploration based on visual input and behavior, demonstrating the ability to accurately model neural responses in the dynamic regime of movement through the visual scene.

- We uncover new insights into cortical neural coding by analyzing our model with maximally activating stimuli and saliency maps, and demonstrate that mixed selectivity of visual and behavioral variables is prevalent in mouse V1.

6.2 Related Work

The mouse, as a model organism, offers unparalleled experimental access to the mammalian cerebral cortex [51]. Computational models of mouse V1, including generalized linear models (GLMs) [55, 187] and customized models mimicking the mouse visual hierarchy [188], have been crucial in providing deeper insights into the range of computations performed by visual cortex. More recently, deep CNNs have also been used to model mouse V1 [28, 25, 97, 98, 183, 189, 20, 190, 191].

Despite their success in predicting neural activity in the macaque visual cortex [96], deep CNNs trained on ImageNet have had limited success in predicting mouse visual cortical activity [180]. This is perhaps not surprising, as most ImageNet stimuli belong to static images of human-relevant semantic categories and may thus be of low ethological relevance for rodents. More importantly, these deep CNNs may not be the ideal architecture to model mouse visual cortex, which is known to be shallower and more parallel than primate visual cortex [192, 193]. In addition, mice are known to have lower visual acuity than that of primates [181, 182], and much of their visual processing may be devoted to active, movement-based behavior rather than passive analysis of the visual scene [166, 194, 56]. Although the majority of V1 neurons is believed to encode low-level visual features [52], their activity is often strongly modulated by behavioral variables related to eye and head position [53, 54, 55], locomotion [34, 56, 57], arousal [58, 59], and the recent history of the animal [60]. Furthermore, mouse V1 is highly interconnected with both cortical and subcortical brain areas, which contrasts with feedforward, hierarchical models of visual processing [56].

A common architectural approach that has proved quite successful is to split the network into different components (first introduced by [97]):

- a “core” network, which typically consist of a CNN used to extract convolutional features from the visual stimulus [97, 15, 98, 55], sometimes in combination with a recurrent network [97];

- a “shifter” network, which mimics gaze shifts by learning a (typically affine) transformation from head- to eye-centered coordinates, either applied to the pixel input [97, 55] or a CNN layer [98];
- a “readout” network, which learns a mapping from artificial to biological neurons [97, 15, 98].

Owing to the difficulty of developing a predictive model of mouse cortex, Willeke *et al.* [184] recently invited submissions to the Sensorium competition held at NeurIPS ’22. The competition introduced a benchmark dataset of V1 neural activity recorded from head-fixed mice on a treadmill viewing static images, with simultaneous measurements of running speed, pupil size, and eye position. A baseline model was provided as well, which consisted of a 4-layer CNN core in combination with a shifter and readout network [98]. Even though 26 teams submitted 194 different models, the overall improvement to the baseline performance was modest, raising the single trial correlation from .287 to .325 in the Sensorium and from .384 to .453 in the Sensorium+ competition. Architectural innovations (e.g., Transformers, Normalizing Flows, YOLO, and knowledge distillation), were unable to make an impact, as most improvements were gained from ensemble methods. A promising direction was taken by the winning model, which attempted to learn a latent representation of the “brain state” from the various behavioral variables, inspired by [33]. However, the model utilized the timestamps of the test set to estimate recent neuronal activities, which the other competitors did not have access to.

Taken together, we identified three main limitations of previous work that this study aims to address:

- **Head-fixed preparations.** Most previous models operated on data from animals in head-fixed conditions with static stimuli, which do not mirror natural behavior and thus provide limited insight into visual processing in real-world environments. In contrast, the present work is applied to state-of-the-art neurophysiological recordings of V1 activity in freely moving mice. This represents a dramatic shift in the "parameter space" of visual input, from static images to dynamic, real-world visual input. One could imagine that this will make the modeling process more difficult, because the stimulus set is more complex, or easier, because it is more matched to the computational challenge the brain evolved for.

- **Limited influence of behavioral state.** Previous models often limited the influence of behavioral state to eye measurements and treadmill running speed, which were either concatenated with the visual features [59, 184], utilized in the shifter network to determine the gaze-contingent retinal input [97, 184], or used to predict a multiplicative gain factor [97].

- **Missing temporal dynamics.** Most previous modeling works ignored the temporal factors that might influence V1 activity and overlooked the dynamic nature of visual processing (but see [97]). We overcome this limitation by utilizing approximately 1-hour-long recordings of three mice freely exploring an arena, and our model is capable of handling continuous data streams of any length.

6.3 Methods

Head-mounted recording system We had access to data from three adult mice who were freely exploring 48 cm long × 37 cm wide × 30 cm high arena (Fig. 6.1A), collected with a state-of-the-art recording system [55] that combined high-density silicon probes with miniature head-mounted cameras (Fig. 6.1B). One camera was aimed outwards to capture the visual scene from the mouse's perspective ("worldcam") at 16 ms per frame (downsampled to 60 × 80 pixels). A second camera, aimed at the eye, was used to extract eye position (θ, ϕ) and pupil radius (σ) at 30 Hz using DeepLabCut [171]. Pitch (ρ) and roll (ω) of the mouse's head position were extracted at 30 kHz from the inertial measurement unit (IMU). θ, ϕ, ρ, and ω allowed for the worldcam video to be corrected for eye movements: A 3-layer fully-connected shifter network (where each linear layer was accompanied by Tanh and BatchNorm) was trained to predict a rotation (bounded by ±36°) and a shift (bounded by ±16 pixels horizontally and ±12 pixels vertically) based on θ, ϕ, ρ, and ω to convert each frame to head- to eye-centered coordinates. Locomotion speed (s) was estimated from the top-down camera feed using DeepLabCut [171]. Electrophysiology data was acquired at 30 kHz using a 11 µm × 15 µm multi-shank linear silicon probe (128 channels) implanted in the center of the left monocular V1, then bandpass-filtered between 0.01 Hz and 7.5 kHz, and spike-sorted with Kilosort 2.5 [172]. Single units were selected using Phy2 (see [167]) and inactive units (mean firing rate $<$ 3 Hz) were removed. This yielded 68, 32, and 49 active units for Mouse 1–3, respectively. To prepare the data for machine learning, all data streams were deinterlaced and resampled at 20.83 Hz (48 ms per frame; Fig. 6.1C). For a more detailed description of the dataset, see Appendix A.3.1 and Ref. [55].

Model architecture We used a 3-layer CNN (kernel size 7, 128 × 64 × 32 channels) to encode the visual stimulus. Each convolutional layer was followed by a BatchNorm layer, a ReLU, and a Dropout layer (0.5 rate). A fully-connected layer transformed the learned visual features into a visual feature vector, $\boldsymbol{v}$ (Fig. 6.2, *top-right*). In a purely visual version of the model, $\boldsymbol{v}$ was fed into a fully-connected layer, followed by a softplus layer, to yield a neuronal response prediction.

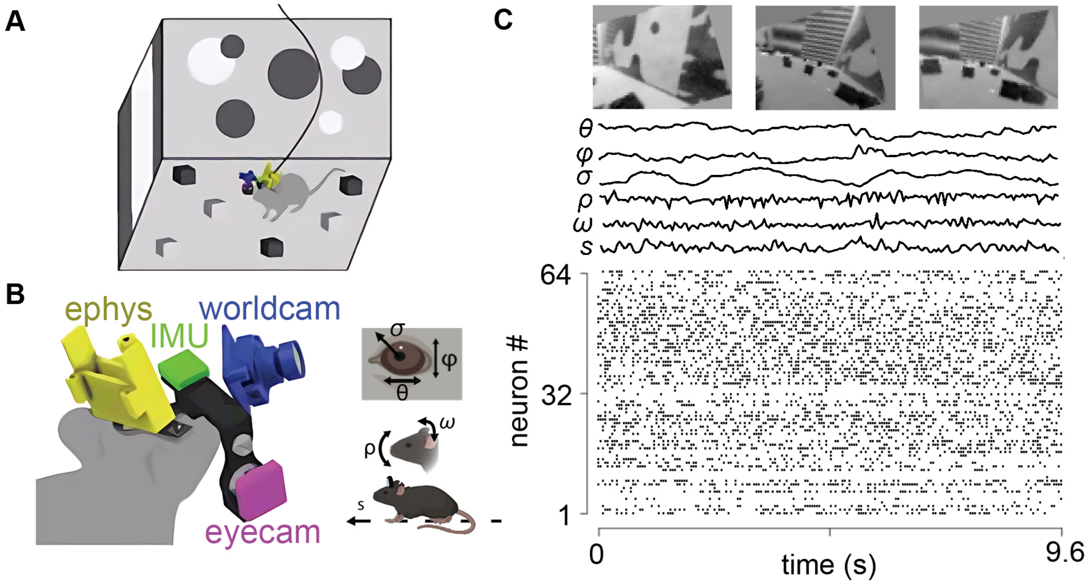

Figure 6.1: Schematic of the head-mounted recording system for freely moving mice (adapted from [55]). A) Three mice freely explored a 48 cm long × 37 cm wide × 30 cm high arena. B) Preparation included a silicon probe for electrophysiological recording in V1 (yellow), miniature cameras for recording the mouse's eye position and pupil size (θ, ϕ, and σ; magenta), and visual scene (blue), and inertial measurement unit for measuring head orientation (ρ and ω; green). C) Sample data from a 9.6 s period during free movement showing (from top) visual scene, horizontal and vertical eye position, pupil size, head pitch and roll, locomotor speed, and a raster plot of 64 units.

To encode behavioral state, we constructed an input vector from different sets of behavioral variables:

- $\mathcal{S}$: all behavioral variables used in the Sensorium+ competition [184], consisting of running speed (s), pupil size (σ), and its temporal derivative ($\dot{\sigma}$);
- $\mathcal{B}$: all behavioral variables used in [55], consisting of eye position (θ, ϕ), head position (ρ, ω), pupil size (σ), and running speed (s);
- $\mathcal{D}$: the first-order derivatives of the variables in $\mathcal{B}$, namely $\dot{\theta}$, $\dot{\phi}$, $\dot{\omega}$, $\dot{\rho}$, $\dot{\sigma}$, and s.

To test for interactions between behavioral variables, these sets could also include the pairwise multiplication of their elements; e.g., $\mathcal{B}_\times = \{b_i b_j \, \forall \, (b_i, b_j) \in \mathcal{B}\}$. The input

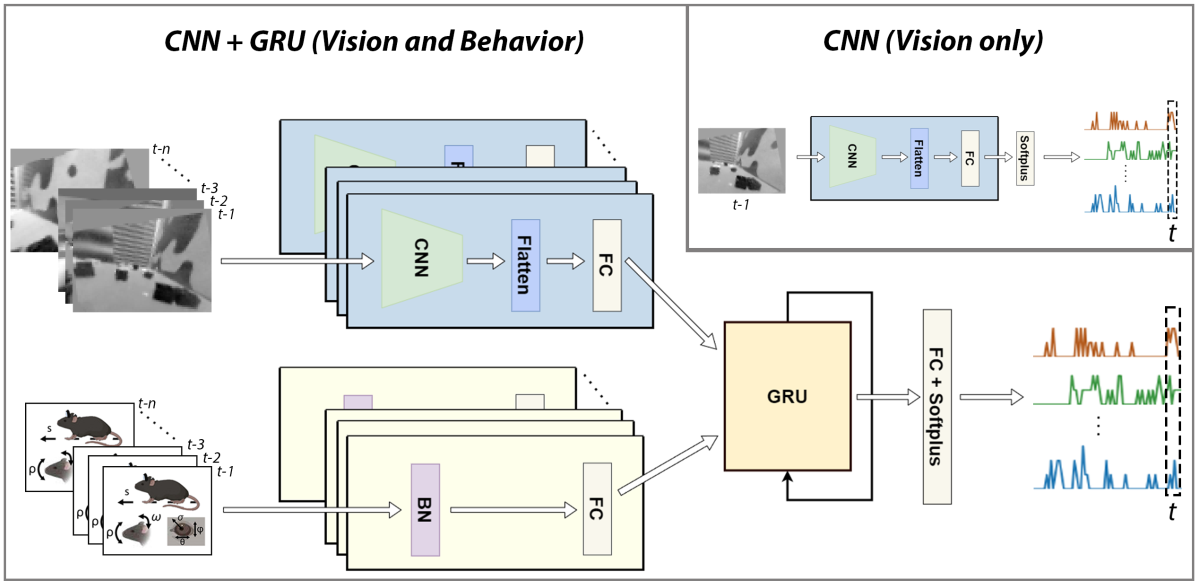

Figure 6.2: Model architecture diagram. The vision-only network (top-right) was a CNN network, predicting the neural activity at time t given the visual input at time $t-1$ (48 ms bins). The full model combined the CNN with a behavioral encoder and a gated recurrent unit (GRU), predicting the neural activity at time t given the visual and behavioral inputs from time $t-1$ to $t-n$.

vector was then passed through a batch normalization layer and a fully connected layer (subjected to a strong L1 norm for feature selection) to produce a behavioral vector, $\boldsymbol{b}$.

We then concatenated the vectors $\boldsymbol{v}$, $\boldsymbol{b}$, and their element-wise product $\boldsymbol{v} \odot \boldsymbol{b}$ (all calculated for each individual input frame), fed them through a batch normalization layer, and input them to a 1-layer gated recurrent unit (GRU) (hidden size of 512). To incorporate temporal dynamics, we constructed different versions ($\mathrm{GRU_k}$) of the model that had access to k previous frames. A fully-connected layer and a softplus activation function were applied to yield the neuronal response prediction.

Training and model evaluation Since the visual input depended on the movement of the mouse and the mouse could be in very different behavioral states over the length of the recording, the data was highly inhomogeneous across time. To deal with the continuous and dynamic nature of the data, we therefore split the ~ 1 h-long recording into 10 consecutive segments. The first 70 % of each segment were then reserved for training (including an 80-20 validation split) and the remaining 30 % for testing.

Note that it is unlikely for data to “leak” from the train segment into the test segment. While it is possible that the mouse could have been exploring the same part of the arena at different segments of the recording, it was free to move its head, eyes, and body as it saw fit. Thus two duplicate data points could only be produced by the animal exactly duplicating the time courses of its eye, head, and body movement in the exact same location of the arena.

Models were separately trained on the data from each mouse. Model parameters were optimized with Adam (batch size: 256, CNN learning rate: .0001, full model: .0002) to minimize the Poisson loss between predicted neuronal response ($\hat{r}$) and ground truth (r) : $\frac{1}{N}\sum_{i=1}^{N}(\hat{r}_i - r_i \log \hat{r}_i)$, where N denotes the number of recorded neurons for each mouse. We used early stopping on the validation set (patience: 5 epochs), which led all models to converge in less than 50 epochs. Due to the large number of hyper-parameters, the specific network and training settings were determined using a combination of grid search and manual exploration on a validation set (see Appendix A.3.2).

To evaluate model performance, we calculated the cross-correlation (cc) between a smoothed version (2 s boxcar filter) of the predicted and ground-truth response for each recorded neuron [55].

All models were implemented in PyTorch and trained on an NVIDIA RTX 3090 with 24GB memory. All code, data used to train the models, and weights of the trained model can be found at `https://github.com/bionicvisionlab/2023-Xu-Multimodal-Mouse-V1`.

Maximally activating stimuli We used gradient ascent [15] to discover the visual stimuli that most strongly activate a particular model neuron in our network. The visual input was initialized with noise sampled in $\mathcal{N}(.5, 2)$. The behavioral variables were initialized to a vector of all ones, and updated in the loop with the visual stimuli. We used the Adam optimizer to repeatedly add the gradient of the target neuron's activity with respect to its inputs. We also applied L2 regularization (weight of .02) and Laplacian regularization (weight of 0.01) [14] on the image. This procedure was repeated 6400 times. The resulting, maximally activating visual stimuli were smoothed with a Butterworth filter (low-pass, .05 cutoff frequency ratio) to reduce the impact of high-frequency noise.

Saliency map We computed a saliency map [174] of the behavioral vector for each neuron to discover which behavioral variables contributed most strongly to each model neuron's activity. We iterated through the test dataset, recorded the gradient of each behavioral input with respect to each neuron's prediction, and then averaged the gradients per neuron to obtain the saliency map.

6.4 Results

Mouse V1 activity is best predicted with a 3-layer CNN To determine the purely visual contribution to V1 responses, we experimented with a large number of vision architectures (see Appendix A.3.2). In the end, a vanilla 3-layer CNN (kernel size 7, $128 \times 64 \times 32$ channels) yielded the best cross-correlation between predicted and ground-truth responses (Table 6.1), outperforming the best autoencoder architecture (kernel size: 7, encoder: $64 \times 128 \times 256$ channels, decoder: $256 \times 128 \times 64$ channels), ResNet-18 [195] (a 20-layer CNN with the first input channel being replaced by 1), EfficientNet-B0 [196] (a 65-layer CNN with the first input channel being replaced by 1), and the Sensorium baseline [98] (a 4-layer CNN with a readout network). The greatest improvement in cross-correlation was achieved for Mouse 1.

Behavioral variables improve most neuronal predictions Once we identified the 3-layer CNN as the best visual encoder, we added the different sets of behavioral variables to the network. To allow for a fair comparison with the Sensorium+ baseline [184], we first limited ourselves to $\mathcal{S} = \{\sigma, \dot{\sigma}, s\}$, but then gradually added more behavioral variables ($\mathcal{B}$) [55] as well the derivatives of these variables ($\mathcal{D}$) and multiplicative pairs ($\mathcal{B}_\times$ and

	Mouse 1		Mouse 2		Mouse 3	
Model	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓
CNN	**.596 ± .134**	**.0626**	**.424 ± .141**	**.100**	**.552 ± .138**	**.0908**
Autoencoder	.555 ± .140	.0710	.370 ± .145	.112	.521 ± .144	.0974
ResNet-18 [195]	.517 ± .159	.0782	.366 ± .171	.107	.511 ± .138	.0944
EfficientNet-B0 [196]	.542 ± .153	.0694	.393 ± .165	.103	.510 ± .127	.0965
Sensorium [98]	.519 ± .149	.0754	.381 ± .128	.119	.497 ± .136	.100

Table 6.1: Best-performing vision models, compared to the Sensorium baseline [98] (see Appendix A.3.2 for more). Best-performing networks are indicated in bold. *cc*: cross-correlation, mean ± standard deviation across neurons (↑: the higher the better), MSE: mean-squared error (↓: the lower the better).

		Mouse 1		Mouse 2		Mouse 3	
	Feature Set	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓
CNN+GRU$_1$	$\{\mathcal{B} \cup \mathcal{D}\}_\times$	**.646 ± .136**	**.0543**	**.508 ± .166**	**.0917**	**.607 ± .132**	.0801
	$\mathcal{B} \cup \mathcal{D}$	.644 ± .141	**.0543**	.467 ± .180	.0966	.599 ± .134	**.0797**
	$\mathcal{B}_\times$	.639 ± .141	.0555	.484 ± .165	.0925	.593 ± .132	.0812
	$\mathcal{B}$	.641 ± .136	.0555	.450 ± .158	.0962	.583 ± .136	.0825
	$\mathcal{S}$	.623 ± .142	.0551	.498 ± .154	.0911	.579 ± .135	.0828
	Sensorium+	.540 ± .138	.0696	.441 ± .181	.101	.487 ± .146	.0975

Table 6.2: CNN+GRU$_1$ model, compared to the Sensorium+ baseline [98], trained on different behavioral feature sets. $\mathcal{S} = \{\sigma, \dot{\sigma}, s\}$: the set of variables used in the Sensorium+ competition [184]. $\mathcal{B} = \{\theta, \phi, \omega, \rho, \sigma, s\}$: the set of variables from [55]. $\mathcal{D} = \{\dot{\theta}, \dot{\phi}, \dot{\omega}, \dot{\rho}, \dot{\sigma}, s\}$: the derivatives of $\mathcal{B}$. $\mathcal{A}_\times = \{a_i a_j \; \forall \; (a_i, a_j) \in \mathcal{A}\}$ denotes the set of all multiplicative pairs. ∪ denotes the union operator. Best performing networks are indicated in bold. *cc*: cross-correlation, mean ± standard deviation across neurons (↑: the higher the better), MSE: mean-squared error (↓: the lower the better).

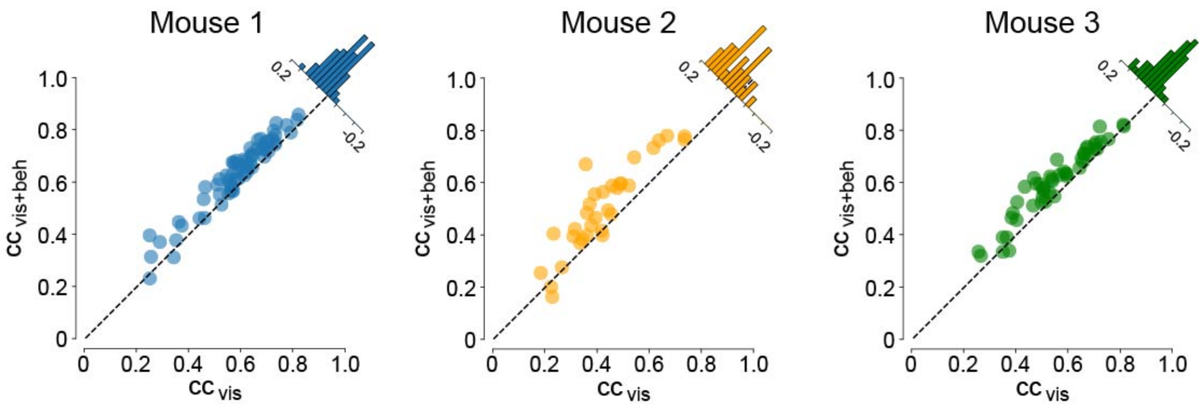

Figure 6.3: The integration of behavioral variables improved the cross-correlation (cc) for the majority of neurons. Each dot represents a neuron. A dot above the dashed diagonal indicates a higher cc with the inclusion of behavioral variables. Histograms (small insets) illustrate the distribution of the improvement in cc across the neuronal population.

$\{\mathcal{B} \cup \mathcal{D}\}_\times$).

The results are shown in Table 6.2. All models were able to outperform the Sensorium+ baseline, and the addition of behavioral variables and their interactions further improved model performance. Note that although the full model used a GRU to combine visual and behavioral features, the input sequence length was always 1 (i.e., GRU_1). That being said, it is possible that the GRU learned long-term correlations that the Sensorium+ baseline model did not have access to. Nevertheless, the biggest performance improvements were gained through the addition of behavioral variables related to head and eye position (which are present in $\mathcal{B}$ but not in $\mathcal{S}$), their derivatives ($\mathcal{D}$), and multiplicative interactions between these variables ($\{\mathcal{B} \cup \mathcal{D}\}_\times$).

We also wondered whether the prediction of only some V1 neurons would benefit from the addition of these behavioral variables. To our surprise, the cross-correlation between predicted and ground-truth responses improved for almost all recorded V1 neurons (Fig. 6.3).

Access to longer series of data in time further improves predictive performance After we identified the full behavioral feature set ($\{\mathcal{B} \cup \mathcal{D}\}_\times$) as the one yielding the best model performance, we extended the GRU's temporal dependence by allowing the input to vary from one frame (48 ms) to a total of eight frames (384 ms), and assessed the model's performance.

The results are shown in Table 6.3. The amount of temporal information needed by the model to reach peak predictive performance was similar across mice (288 ms, 192 ms, and 192 ms in terms of cross-correlation, 192 ms, 192 ms, and 192 ms in terms of mean-squared error, respectively). This indicates that temporal information is important for predicting dynamic neural activity. However, the dependence on temporal information has a limit, and different neurons in V1 might possess different temporal capacities.

Well-defined visual receptive fields emerge To assess whether the CNN+GRU$_1$ model learned meaningful visual receptive fields, we used gradient ascent (see Methods) to find the maximally activating stimulus for each neuron. Receptive fields for the 32

		Mouse 1		Mouse 2		Mouse 3	
Model	History	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓
CNN+GRU$_1$	48 ms	.646 ± .136	.0543	.508 ± .166	.0917	.607 ± .132	.0801
CNN+GRU$_2$	96 ms	.649 ± .139	.0528	.506 ± .174	.0898	.607 ± .133	.0811
CNN+GRU$_3$	144 ms	.653 ± .139	.0528	.528 ± .160	.0843	.604 ± .134	.0790
CNN+GRU$_4$	192 ms	.650 ± .142	**.0525**	**.566 ± .169**	**.0799**	**.614 ± .136**	**.0773**
CNN+GRU$_5$	240 ms	.645 ± .144	.0556	.519 ± .177	.0933	.598 ± .134	.0807
CNN+GRU$_6$	288 ms	**.654 ± .142**	.0526	.549 ± .175	.0823	.598 ± .138	.0798
CNN+GRU$_7$	336 ms	.644 ± .148	.0534	.533 ± .169	.0931	.596 ± .133	.0806
CNN+GRU$_8$	384 ms	.646 ± .146	.0547	.546 ± .179	.0840	.594 ± .141	.0825

Table 6.3: CNN+GRU$_k$ model trained with input from k timesteps on the full feature set ($\{\mathcal{B} \cup \mathcal{D}\}_\times$). Best performing networks are indicated in bold. *cc*: cross-correlation, mean ± standard deviation (↑: the higher the better), MSE: mean-squared error (↓: the lower the better).

best-predicted neurons are shown in Fig. 6.4. Interestingly, most of them had well-defined excitatory and inhibitory subregions, often resembling receptive fields of orientation-selective neurons. Most excitatory and inhibitory subregions spanned approximately 30° of visual angle (the full width of the frame, 80 pixels, roughly corresponding to 120° of visual angle), which is roughly on the same order of magnitude compared to receptive field sizes typically observed in mouse V1, varying from 10° to 30° [182, 197, 55].

Receptive fields were noticeably different across mice. Whereas Mouse 1 and 3 had visual receptive fields with strongly excitatory subregions, most model neurons for Mouse 2 appeared to be inhibited by visual signals (same colorbar across panels). In addition, several model neurons lacked pronounced visual receptive fields, indicating that they were more strongly driven by behavioral variables. Even though model fits were repeated with different initial values for the behavioral variables, the resulting visual receptive fields looked qualitatively the same (see Appendix A.3.3), thus demonstrating the validity of the generated receptive fields. In addition, even some of the best-predicted neurons lack a pronounced or spatially structured receptive field, implying that these neurons could be primarily driven by behavioral variables.

Analysis of behavioral saliency maps reveals different types of neurons Intrigued by the fact that some neurons lacked pronounced visual receptive fields, we aimed to analyze the influence of behavioral state on the predicted neuronal response by performing a saliency map analysis on the behavioral inputs (see Methods). Since different behavioral variables operate on different input ranges, we first standardized the saliency map activities for each behavioral variable across the model neuron population. Saliency map activities further than 1 standard deviation from the mean were then interpreted as "driving" the neuron, allowing us to categorize each neuron as being driven by one or multiple behavioral variables (Fig. 6.5).

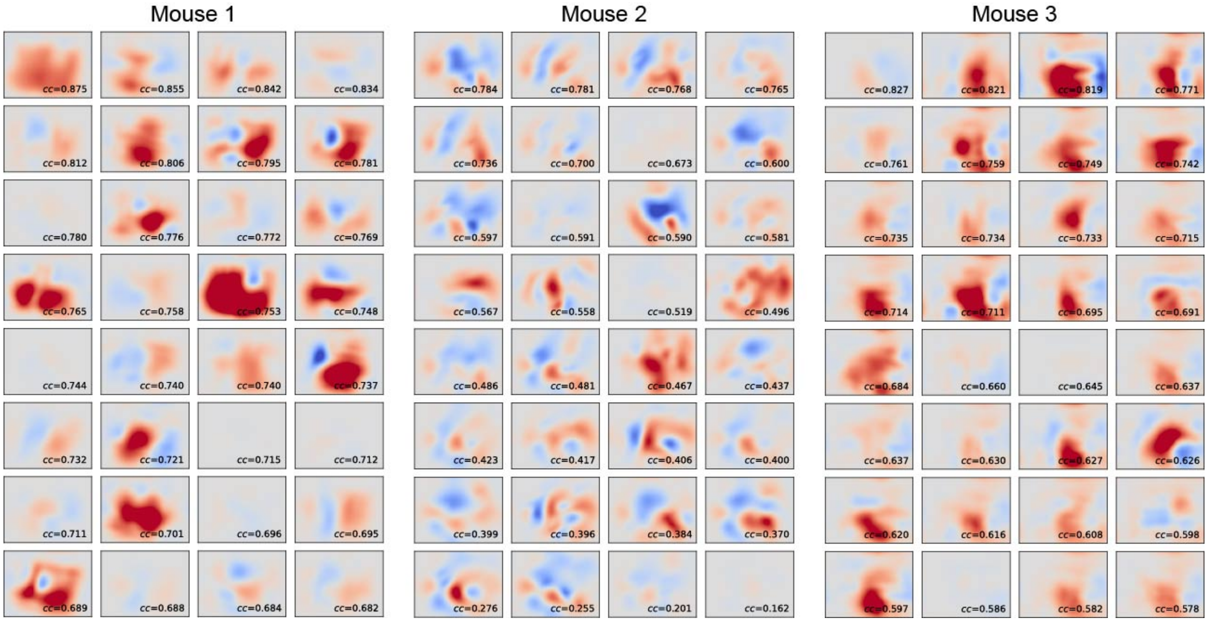

Figure 6.4: The maximally activating stimuli learned in CNN+GRU$_1$, generated via gradient ascent. The 32 neurons with the highest cross-correlation (*cc*) from each mouse are shown, sorted by *cc*.

We first asked which neurons in our model were driven by which behavioral variables (Fig. 6.5, *top*). Consistent with [55], we found a large fraction of model neurons driven by eye and head position, and smaller fractions driven by locomotion speed and pupil size. Approximately 20-30% of neurons were not driven by any of these behavioral variables, rendering their responses purely visual.

However, a particular neuron could be driven by multiple behavioral variables. Repeating the above analysis, we found that most model neurons showed mixed selectivity (i.e., responding to different categories of information, such as visual and motor signals, or stimulus and reward signals), with only a minority of cells responding exclusively to a single behavioral variable, (Fig. 6.5, *middle*). Adding the interaction terms between behavioral variables (Fig. 6.5, *bottom*) did not change the fact that most model V1 neurons encoded combinations of multiple behavioral variables, often relating information about the animal's eye position to head position and locomotor speed.

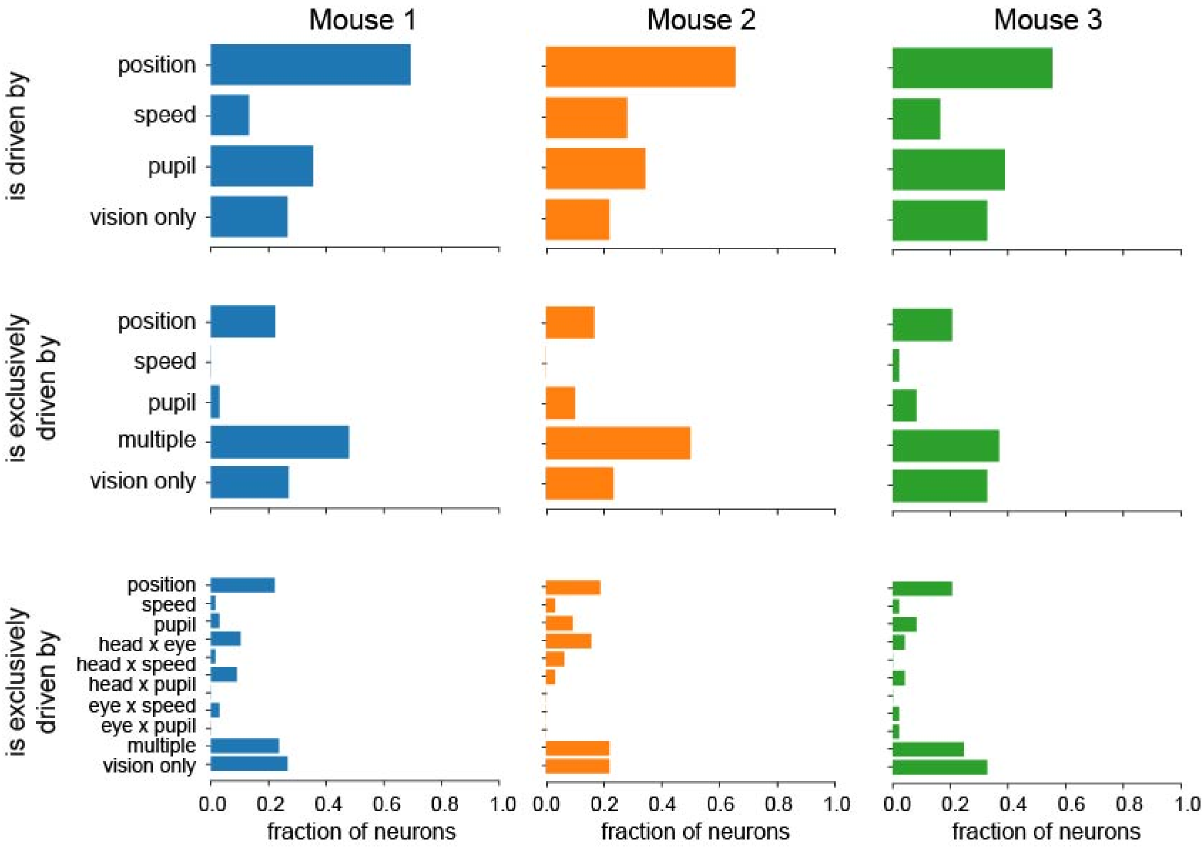

Figure 6.5: Effect of behavioral variables on model neuron activity, inferred by the saliency analysis. A) Fraction of neurons that are "driven by" (i.e., their saliency map activation is further than 1 standard deviation from the mean) different behavioral variables (similar to [55]). A neuron that responds to (e.g.) both position and speed may be counted twice. Neurons without a strong behavioral drive are categorized as "vision only". B) Fraction of neurons that are uniquely driven by a specific behavioral variable. Still, a large fraction of neurons are driven by multiple behavioral variables. C) Same as B), but split with interaction terms.

6.5 Discussion

In this paper, we propose a deep recurrent neural network that achieves state-of-the-art predictions of V1 activity in freely moving mice. We discovered that our model outperforms previous models under these more naturalistic conditions, which could be attributed to the better alignment of this data with the computations performed by the mouse visual system, based on its natural visual environment and behavioral characteristics. Similar to previous models, we found that a simple CNN architecture is sufficient to predict the visual response properties of cells in mouse V1.

In addition, mouse V1 is known to be strongly modulated by signals related to the movement of the animal's eyes, head, and body [34, 56, 55], which are severely restricted in head-fixed preparations. Models trained on head-fixed preparations may thus be limited in their predictive power. In contrast, our model was able to predict V1 activity on a 1-hour continuous data stream, during which the animal freely explored a real-world arena. Our analyses demonstrate the impact of the animal's behavioral state on V1 activity and reveal that most model V1 neurons exhibit mixed selectivity to multiple behavioral variables.

Accurate predictions of mouse V1 activity under natural conditions Our brains did not evolve to view stationary stimuli on a computer screen. However, most research on neural coding in vision has been conducted under head-fixed conditions, which do not mirror natural behavior and thus provide limited insight into visual processing in real-world environments. Some visual functions mediated by the ventral stream, such as identifying faces and objects, resemble this condition, but the real visual environment is constantly shifting due to self-motion, leading to dynamic activities such as navigation or object reaching, typically mediated by the dorsal stream. To truly understand visual perception in natural environments, we need to capture the computational principles when the subjects are in motion.

In this research, we take the initial steps towards this by modeling a novel data type encompassing neural activity coupled with a visual scene captured from a freely moving animal's perspective. This represents a dramatic (but, in our opinion, crucial) shift in the "parameter space" of visual input, from static images projected on a screen to dynamic, real-world visual input.

Surprisingly, visual responses were best predicted with a standard three-layer ("vanilla") CNN (Table 6.1), as compared to a multitude of more sophisticated models that included autoencoders, variational autoencoders, filter bank models, and pre-trained ResNet and EfficientNet architectures. One possible explanation might be that the neurons in our dataset were selective for other behavioral inputs that we did not have access to, and that the vanilla CNN architecture imposed the fewest assumptions about how visual input contributed to the neural activity. In addition, visual receptive fields for Mouse 2 were noticeably different from the other two mice, exhibiting pronounced inhibitory subregions (Fig. 6.4, *center*). This is consistent with the fact that the cortical probes of Mouse 2 were more superficial compared to the other two mice [55], so the recorded neurons may have both different anatomical inputs and different visual responses.

Mixed selectivity of behavioral variables Our experiments demonstrated that the models incorporating behavioral variables and their interactions performed substantially better than the models relying exclusively on visual inputs. Moreover, our saliency map analysis showed that only around 25% of model neurons could be considered purely visual, with the majority of model neurons driven by multiple behavioral variables.

This widespread mixed selectivity is consistent with previous literature suggesting that V1 neurons may be modulated by a high-dimensional latent representation of several behavioral variables related to the animal's movement, recent experiences, and behavioral goals [33]. It is also consistent with the idea of a basis function representation [198, 199], which allows a population of neurons to conjunctively represent multiple behaviorally relevant variables. Such representations are often employed by higher-order visual areas in primate cortex to implement sensorimotor transformations [200, 185, 198]. It is intriguing to find computational evidence for such a representation as early as V1 in the mouse. Future computational studies should therefore aim to study the mechanisms by which V1 neurons might construct a nonlinear combination of behavioral signals.

Limitations and future work. While our study opens a new perspective on modeling neural activity during natural conditions, there are a few limitations that need to be acknowledged. First, our data was relatively limited (around 50 neurons per animal, for 3 animals). The development of a Sensorium-style standardized dataset [184] for freely-moving mice would significantly benefit future research in this area, enabling more robust comparisons between different modeling approaches. Second, it would be beneficial to integrate other modalities that are known to be encoded in mouse V1 into the model. One such example is reward signals [61], which could provide additional information about the animal's decision-making processes and motivations during exploration.

Chapter 7

Conclusion and future work

This dissertation outlines our efforts in closing gaps and proposing new directions for predictive modeling of the biological visual system, which can have implications for both artificial intelligence (AI) and neuroscience research. First, noticing that comprehensive modeling for the retina and its degeneration was missing, we proposed a computational model of the healthy retina and retinal degeneration that was able to reproduce experimental findings and generate new testable hypotheses to advance the understanding of the degenerated retina. Second, we analyzed the retinal input in freely-moving mice, demonstrated the difference in visual input during active exploration, and thus called for more emphasis on freely moving experimental paradigms. Third, we designed a novel data-driven deep-learning-based model designed for neural and behavioral data collected continuously during free movement which achieved state-of-the-art performance by combining visual, behavioral, and temporal dynamics, and showed that the model revealed novel computational principles of primary visual cortex (V1). To this end, we have moved from a computational and static modeling approach to a data-driven and active one.

Despite these advancements, there are still important directions awaiting exploration in future research.

Training large foundation models Nowadays, large models trained with a huge amount of data have proved to become game changers in language modeling and vision-language modeling. Exciting development intersecting large model training and neuroscience has emerged more recently, employing techniques including pre-training with a large amount of data, fine-tuning for downstream tasks, and aligning multiple data modalities [201, 202, 203]. It will be particularly interesting to integrate those techniques into solving the modeling problem for neural data collected from freely moving animals. The neural data collected from freely moving animals usually contain fewer active units and can be subject to more noise due to the nature of single trial, so casting this problem as a downstream task for a certain large foundation model has the potential of drastically improving the accuracy in neural activity prediction, and thus revealing more computational principles underlying visual processing in animals during free movement. However, some open questions remain. In each trial, data from different neurons with different reliability from different animals can be collected, and it is not yet clear what is the best and most consistent method to handle this heterogeneous characteristic of neural data.

Moving beyond supervised learning The deep-learning modeling approach in this dissertation is limited to supervised training. However, there is evidence that mouse visual system could be better modeled with approaches beyond supervised learning, including unsupervised learning and reinforcement learning [204, 20]. Utilizing unsupervised learning and reinforcement learning might prove valuable in comparing representational similarity of the artificial neural network and the mouse brain [190], modeling the reward signal effectively [61], and aligning models more closely with naturalistic tasks such as prey capture and navigation [166].

Taken together, this dissertation addresses some key gaps in the modeling of the biological visual system and lays the groundwork for deep-learning modeling using neural data obtained from animals in free motion.

Appendix A

Appendix

A.1 Appendix to Chapter 4

A.1.1 Inter-spike intervals

Fig. A.1A shows the inter-spike interval distribution for ON and OFF retinal ganglion cells (RGCs) as a function of disease progression. Fig. A.1B highlights the histograms of the inter-spike interval distributions presented in Panel A during Phase I/II. Note that both the mean firing rate (see Fig. 4.1) and the mean inter-spike interval of ON cells (Fig. A.1B, *left*) decreased as degeneration progressed. This is because most ON cells did not spike at all and therefore did not contribute any finite inter-spike interval values. However, a select few did spike a lot, therefore contributing many short inter-spike interval values. As a result, mean inter-spike interval decreased while mean firing rate decreased as well. Furthermore, we did not observe any bursting behavior in the spontaneous firing of retinal ganglion cells (RGCs).

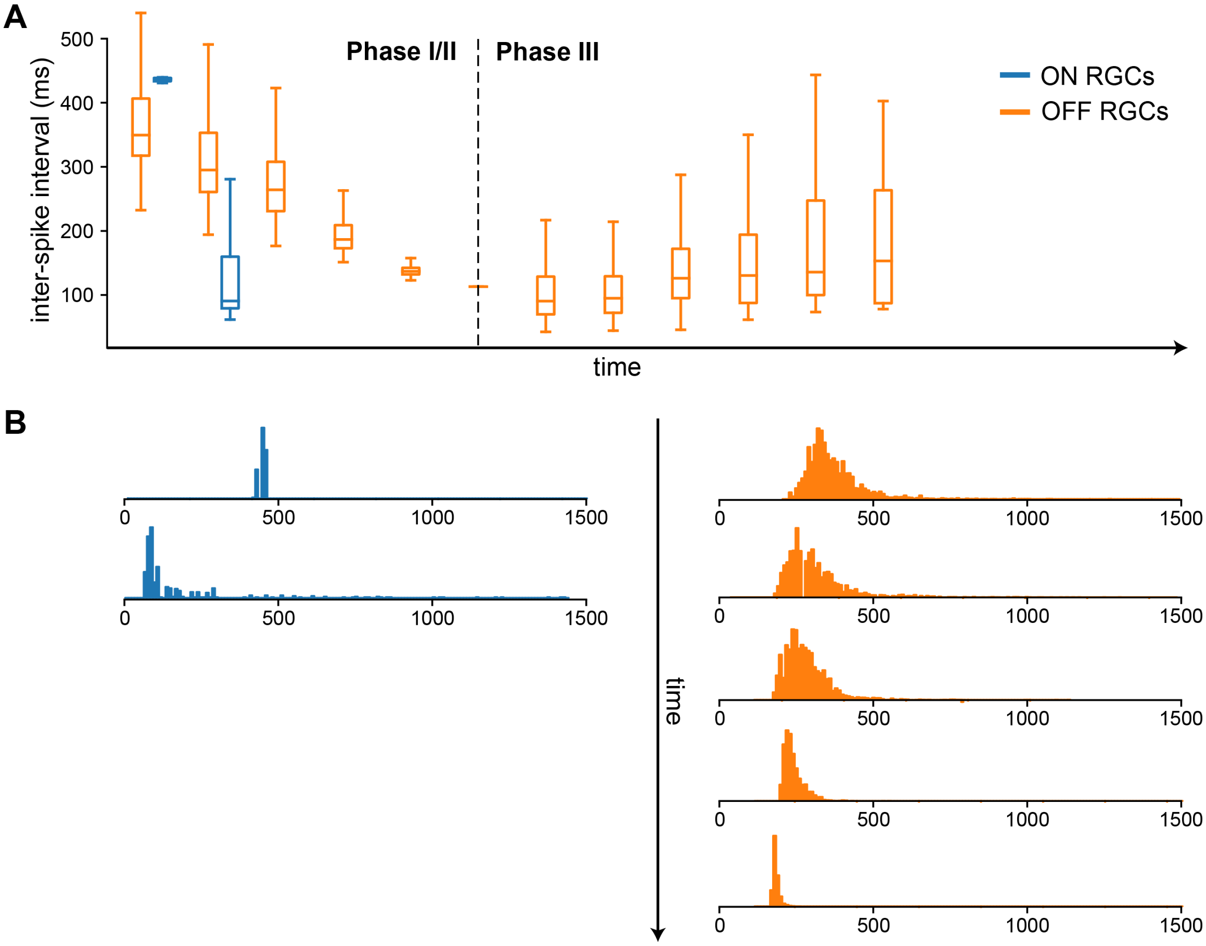

Figure A.1: Inter-spike interval distributions as a function of disease progression. **A)** Box plots of inter-spike interval during Phase I/II and Phase III. **B)** Histograms for the data from Phase I/II that is presented in Panel A.

A.1.2 Phase I/II: Light response

Fig. A.2 shows the firing rate of ON and OFF retinal ganglion cells (RGCs) in response to full-field stimuli of a given light intensity $l(t) = \text{const.}$ (Eq. 3.5), averaged across the population of surviving cells, as a function of both cone outer segment truncation (simulated by a reduction in G_{light} and cone survival rate. As light intensity varied between 0 (black) and 1 (white), the RGC response to 0.5 intensity was interpreted as the spontaneous firing rate.

At high light intensities ($l(t) > 0.5$), both ON and OFF RGC activity was mainly a

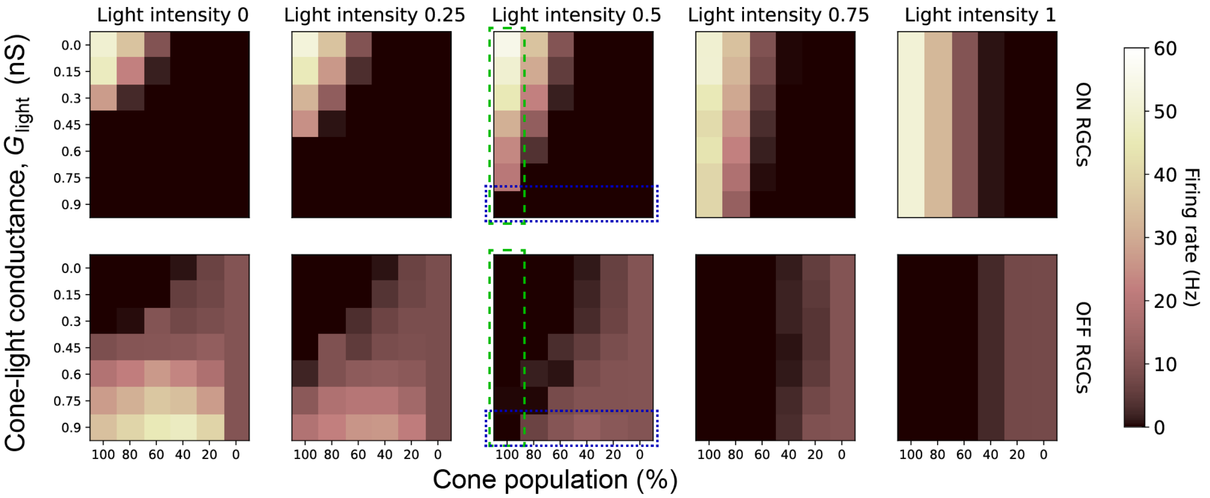

Figure A.2: Light response of RGCs as a function of cone outer segment truncation and cone survival rate. The values encapsulated by the green dashed box are the same as the mean values presented in Fig. 4.1F. The values encapsulated by the blue dotted box are the same as the mean values presented in Fig. 4.1G.

function of the size of the surviving cone population. At low light intensities ($l(t) \leq 0.5$), both cone outer segment truncation and cone population size had an effect on RGC firing, with ON firing dominated by outer segment truncation and OFF firing dominated by cone population. The linear time axis of disease progression corresponded to the diagonal in each heatmap (top-left corner: healthy, bottom-right corner: end of Phase I/II).

A.1.3 Ganglion cell response to electrical stimulation

Fig. A.3 shows the RGC response to a constant electrical stimulus delivered either epiretinally (Fig. A.3A) or subretinally (Fig. A.3). The stimulus was the same as for the spatial profiles in Fig. 4.4; that is, a 20 Hz cathodic-first biphasic pulse train of 1 s duration, with pulse duration 0.45 ms and 60 µA amplitude. Electrode dimesnions are given in Section 4.2.2.

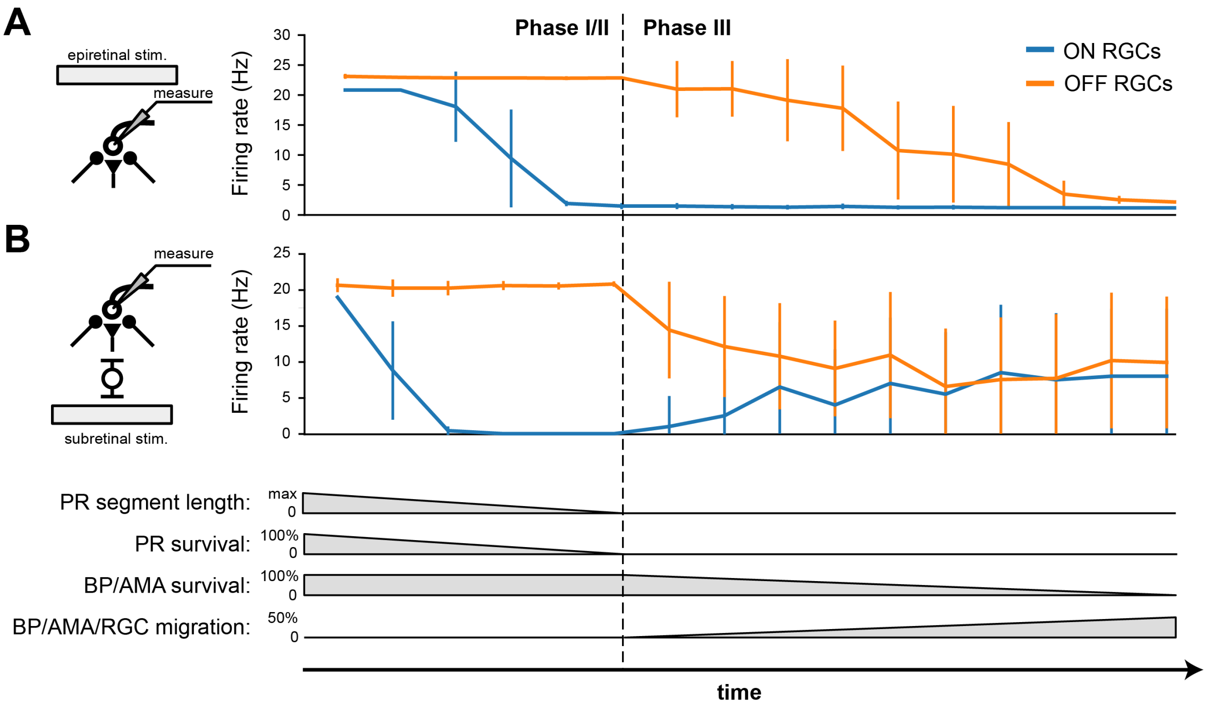

Figure A.3: RGC firing rate in response to epiretinal (A) and subretinal (B) stimulation. The mean and standard deviation were calculated from the ON and OFF RGCs that were located directly under the electrode.

A.1.4 Initial values

Table A.1 shows the initial values for the gating variables in the Hodgkin-Huxley model of the RGCs (Eq. 3.7). The gating variables are in units of Hz. The internal calcium concentration are in units of nM (nanomolar).

m	h	c	n	A	h_A	y	m_T	h_T	d_T	$[\mathrm{Ca}^{2+}]_i$
0.5	0.5	0.5	0.5	0	1	0.5	0	1	1	0.0001

Table A.1: Initial values for the Hodgkin-Huxley model.

Table A.2 shows the initial values of the membrane potential for the different cell types. Initial values for each cell type followed a normal distribution with a mean μ and a standard deviation σ.

Cell type	μ (mV)	σ (mV)
cone photoreceptors	-46.8	3
horizontal cells	-47.7	3
ON bipolar cells	-35	3
OFF bipolar cells	-44	3
ON wide-field amacrine cells	-42	3
OFF wide-field amacrine cells	-34.5	3
ON narrow-field amacrine cells	-47.6	3
ON RGCs	-66.5	3
OFF RGCs	-70.5	3

Table A.2: Initial values for the membrane potential of the different cell types.

A.2 Appendix to Chapter 5

A.2.1 Other attribution methods

We applied other attribution methods (Guided Grad-CAM [175], deconvolution [176], and guided backpropagation [177]) to the eye position prediction model to uncover which pixels most strongly contributed to the prediction. The result (Figure A.4) appeared similar to the mean saliency maps (Figure 5.6).

A.3 Appendix to Chapter 6

A.3.1 Data Acquisition

The full procedure of data collection and data preprocessing is described in detail in Ref. [55], but are briefly described below for the reader's convenience.

Visual scenes from the mouse's perspective, neural activity in V1, and various behavioral variables were simultaneously recorded from three adult mice who were freely exploring an arena with a state-of-the-art head-mounted recording system [55]. The recording system consisted of high-density silicon probes, two miniature cameras and

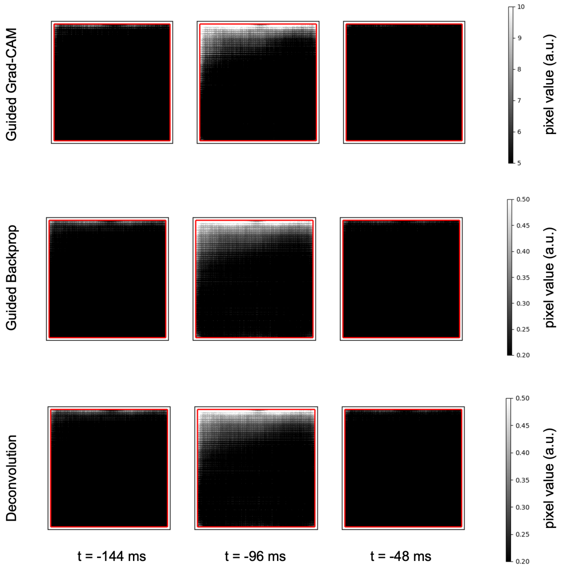

Figure A.4: The visualization results from Guided Grad-CAM (top), guided backpropagation (middle) and deconvolution (bottom).

an inertial measurement unit (IMU). Electrophysiology data was acquired at 30 kHz using a 11 µm × 15 µm multi-shank linear silicon probe (128 channels) implanted in the center of the left monocular V1. One wide-angled camera (around 120°) was aimed outwards to capture the visual scene available to the right eye of the mouse at 16 ms per frame ("worldcam"). A second camera was aimed at the right eye (illuminated by an infrared-LED) to record a video feed of the right eye at 30 Hz. The inertial measurement unit (IMU) acquired three-axis gyroscope and accelerometer data at 30 kHz. In addition, a top-down camera recorded the mouse in the arena at 60 Hz.

During experiments, mice were placed in an arena where they could move around freely for about 1 hour. The arena was approximately 48 cm long × 37 cm wide × 30 cm high. The gray floor was covered with black-and-white Legos to provide visual contrast. One wall of the arena was a monitor displaying a moving black-and-white spots stimulus, and the other three walls were covered with wallpaper with static stimuli including white noise, black-and-white high-spatial-frequency gratings, and black-and-white low-spatial-frequency gratings. In order to encourage foraging behavior during the recording sessions, small fragments of tortilla chips were sparsely distributed across the arena.

The worldcam video was downsampled to 60×80 pixels. DeepLabCut [171] was used to extract extract eye position (θ, ϕ) and pupil radius (σ). Pitch (ρ) and roll (ω) of the mouse's head position were extracted from the IMU. Locomotion speed (s) was estimated from the top-down camera feed using DeepLabCut [171]. Electrophysiology data bandpass-filtered between 0.01 Hz and 7.5 kHz, and spike-sorted with Kilosort 2.5 [172]. Single units were selected using Phy2 (see [167]) and inactive units (mean firing rate < 3 Hz) were removed. This yielded 68, 32, and 49 active units for Mouse 1–3, respectively. To prepare the data for machine learning, all data streams were deinterlaced and resampled at 20.83 Hz (48 ms per frame; Fig. 6.1C).

A.3.2 Vision-Only Models

Hyperparameter Tuning

We performed a grid search to find the optimal CNN kernel size (3, 5, 7, 9), number of channels (32, 64, 128, 256, 512; in various combinations), and dropout rate (0, 0.25, 0.5). While other models often rely on kernel size 3 for their convolutional neural network (CNN), we found these small kernels to lead to worse performance, perhaps due to the mouse's low-resolution vision. Kernel size 7 performed best.

We repeated the grid search for CNNs with different number of convolutional layers. The resulting 3-layer CNN with 0.5 dropout rate outperformed many differently sized networks, such as a 1-layer CNN with 1024 channels (i.e., a shallow but wide network), a 2-layer CNN, or a 4-layer CNN. Choice of learning rates and optimizers had no notable effect on the final performance of the networks.

Autoencoder

We initially hypothesized that an autoencoder could provide regularization benefits over a “vanilla” CNN, because the reconstruction loss might encourage the model to learn visual features that are useful for decoding. We used an encoder ϕ to map the original frame $\mathcal{F}$ to a vector $\mathcal{V}$ in the latent space, which was present at the bottleneck, while the decoder ψ then mapped the vector $\mathcal{V}$ from the latent space to the output.

$$\phi: \ \mathcal{F} \rightarrow \mathcal{V}, \tag{A.1}$$

$$\psi: \ \mathcal{V} \rightarrow \mathcal{F}, \tag{A.2}$$

$$\phi, \psi = \text{argmin}_{\phi,\psi} ||\mathcal{F} - (\psi \cdot \phi)\mathcal{F}||^2. \tag{A.3}$$

After hyperparameter search, we settled on size 256 for the latent space vector, and the weight of the reconstruction loss relative to the Poisson loss was fixed at 0.5. Both the encoder and the decoder were 3-layer CNNs, and their numbers of channels were symmetric. However, after testing a number of autoencoders with different configurations (Table A.3), we found that a simple 3-layer CNN outperformed any and all of them.

A.3.3 Visual Receptive Fields

To test whether the recovered visual receptive fields are sensitive to the choice of initial values for the behavioral variables, we repeated our experiments by initializing the behavioral variables with noise sampled from the uniform distribution $[-1, 1)$. The values remained the same throughout the process of gradient ascent.

We found that different values of behavioral variables resulted in similar visual receptive fields. A representative example is shown in Fig. A.5. Although some of the absolute values changed, receptive field structure stayed qualitatively the same, with identical

	Mouse 1		Mouse 2		Mouse 3	
Kernel size, encoder #channels	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓	*cc* ↑	MSE ↓
3, 16 × 32 × 64	.539 ± .149	.0728	.389 ± .128	**.107**	.502 ± .129	.0996
5, 16 × 32 × 64	.550 ± .147	.0728	.363 ± .116	.109	.508 ± .135	.0983
7, 16 × 32 × 64	.525 ± .152	.0732	.353 ± .121	.117	.509 ± .131	**.0980**
9, 16 × 32 × 64	.518 ± .147	.0752	.315 ± .101	.119	.492 ± .135	.0997
3, 32 × 64 × 128	.543 ± .144	.0737	.367 ± .128	.109	.503 ± .131	.100
5, 32 × 64 × 128	.551 ± .149	.0723	.361 ± .109	.119	.514 ± .132	.0984
7, 32 × 64 × 128	.539 ± .145	.0739	**.390 ± .118**	.109	.492 ± .129	.100
9, 32 × 64 × 128	.510 ± .155	.0758	.331 ± .119	.112	.500 ± .134	.101
3, 64 × 128 × 256	.541 ± .146	.0758	.374 ± .123	.110	.514 ± .127	.0990
5, 64 × 128 × 256	.552 ± .145	.0777	.362 ± .119	.110	.508 ± .134	.104
7, 64 × 128 × 256	**.553 ± .134**	**.0688**	.369 ± .104	.111	**.530 ± .136**	.0992
9, 64 × 128 × 256	.537 ± .146	.0811	.355 ± .109	.119	.500 ± .128	.105

Table A.3: Performance of different autoencoders. The numbers of channels in the decoder were symmetric with those of the encoder. Best performing networks are indicated in bold. *cc*: cross-correlation, mean ± standard deviation across neurons (↑: the higher the better), MSE: mean-squared error (↓: the lower the better).

excitatory and inhibitory subregions to the ones reported in Fig. 6.4.

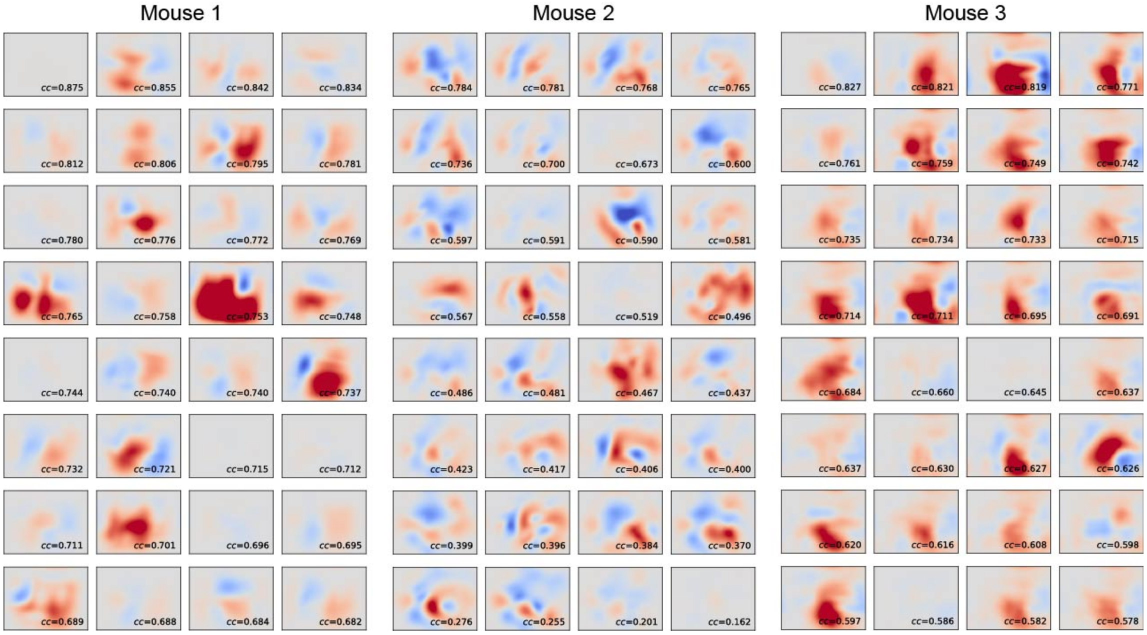

Figure A.5: The maximally activating stimuli learned in CNN+GRU_1, generated via gradient ascent with a fixed and randomly initialized behavioral variable. The 32 neurons with the highest cross-correlation (cc) from each mouse are shown, sorted by cc.

Bibliography

[1] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, *Neuroscience-Inspired Artificial Intelligence*, *Neuron* **95** (July, 2017) 245–258. Publisher: Elsevier.

[2] S. Ullman, *Using neuroscience to develop artificial intelligence*, *Science* **363** (Feb., 2019) 692–693. Publisher: American Association for the Advancement of Science.

[3] B. A. Richards, T. P. Lillicrap, P. Beaudoin, Y. Bengio, R. Bogacz, A. Christensen, C. Clopath, R. P. Costa, A. de Berker, S. Ganguli, C. J. Gillon, D. Hafner, A. Kepecs, N. Kriegeskorte, P. Latham, G. W. Lindsay, K. D. Miller, R. Naud, C. C. Pack, P. Poirazi, P. Roelfsema, J. Sacramento, A. Saxe, B. Scellier, A. C. Schapiro, W. Senn, G. Wayne, D. Yamins, F. Zenke, J. Zylberberg, D. Therien, and K. P. Kording, *A deep learning framework for neuroscience*, *Nature Neuroscience* **22** (Nov., 2019) 1761–1770. Publisher: Nature Publishing Group.

[4] T. Macpherson, A. Churchland, T. Sejnowski, J. DiCarlo, Y. Kamitani, H. Takahashi, and T. Hikida, *Natural and Artificial Intelligence: A brief introduction to the interplay between AI and neuroscience research*, *Neural Networks* **144** (Dec., 2021) 603–613.

[5] F. Rosenblatt, *The perceptron: A probabilistic model for information storage and organization in the brain*, *Psychological Review* **65** (1958), no. 6 386–408. Place: US Publisher: American Psychological Association.

[6] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, *Nature* **521** (May, 2015) 436–444. Publisher: Nature Publishing Group.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is All you Need*, in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[8] J. C. R. Whittington, J. Warren, and T. E. J. Behrens, *Relating transformers to models and neural representations of the hippocampal formation*, Oct., 2021.

[9] D. K. Kim, J. Kwon, M. Cha, and C. Lee, *Transformer as a hippocampal memory consolidation model based on NMDAR-inspired nonlinearity*, *Advances in Neural Information Processing Systems* **36** (Dec., 2023) 14637–14664.

[10] L. Kozachkov, K. V. Kastanenka, and D. Krotov, *Building transformers from neurons and astrocytes*, *Proceedings of the National Academy of Sciences* **120** (Aug., 2023) e2219150120. Publisher: Proceedings of the National Academy of Sciences.

[11] R. M. Cichy, D. Pantazis, and A. Oliva, *Resolving human object recognition in space and time*, *Nature Neuroscience* **17** (Mar., 2014) 455–462. Publisher: Nature Publishing Group.

[12] T. Çukur, S. Nishimoto, A. G. Huth, and J. L. Gallant, *Attention during natural vision warps semantic representation across the human brain*, *Nature Neuroscience* **16** (June, 2013) 763–770. Publisher: Nature Publishing Group.

[13] N. Kriegeskorte and R. A. Kievit, *Representational geometry: integrating cognition, computation, and the brain*, *Trends in Cognitive Sciences* **17** (Aug., 2013) 401–412. Publisher: Elsevier.

[14] P. Bashivan, K. Kar, and J. J. DiCarlo, *Neural population control via deep image synthesis*, *Science* **364** (May, 2019) eaav9436. Publisher: American Association for the Advancement of Science.

[15] E. Y. Walker, F. H. Sinz, E. Cobos, T. Muhammad, E. Froudarakis, P. G. Fahey, A. S. Ecker, J. Reimer, X. Pitkow, and A. S. Tolias, *Inception loops discover what excites neurons most using deep predictive models*, *Nature Neuroscience* **22** (Dec., 2019) 2060–2065. Number: 12 Publisher: Nature Publishing Group.

[16] Y. Miyawaki, H. Uchida, O. Yamashita, M.-a. Sato, Y. Morito, H. C. Tanabe, N. Sadato, and Y. Kamitani, *Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders*, *Neuron* **60** (Dec., 2008) 915–929. Publisher: Elsevier.

[17] Y. Kamitani and F. Tong, *Decoding the visual and subjective contents of the human brain*, *Nature Neuroscience* **8** (May, 2005) 679–685. Publisher: Nature Publishing Group.

[18] D. L. K. Yamins and J. J. DiCarlo, *Using goal-driven deep learning models to understand sensory cortex*, *Nature Neuroscience* **19** (Mar., 2016) 356–365. Publisher: Nature Publishing Group.

[19] S.-M. Khaligh-Razavi and N. Kriegeskorte, *Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation*, *PLOS Computational Biology* **10** (Nov., 2014) e1003915. Publisher: Public Library of Science.

[20] A. Nayebi, N. C. L. Kong, C. Zhuang, J. L. Gardner, A. M. Norcia, and D. L. K. Yamins, *Mouse visual cortex as a limited resource system that self-learns an ecologically-general representation*, *PLOS Computational Biology* **19** (Oct., 2023) e1011506. Publisher: Public Library of Science.

[21] M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust, *Do We Know What the Early Visual System Does?*, *Journal of Neuroscience* **25** (Nov., 2005) 10577–10597. Publisher: Society for Neuroscience Section: Symposia and Mini-Symposia.

[22] L. McIntosh, N. Maheswaranathan, A. Nayebi, S. Ganguli, and S. Baccus, *Deep Learning Models of the Retinal Response to Natural Scenes*, in *Advances in Neural Information Processing Systems 29* (D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, eds.), pp. 1369–1377. Curran Associates, Inc., 2016.

[23] H. Tanaka, A. Nayebi, N. Maheswaranathan, L. McIntosh, S. A. Baccus, and S. Ganguli, *From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction*, .

[24] N. Maheswaranathan, L. T. McIntosh, H. Tanaka, S. Grant, D. B. Kastner, J. B. Melander, A. Nayebi, L. E. Brezovec, J. H. Wang, S. Ganguli, and S. A. Baccus, *Interpreting the retinal neural code for natural scenes: From computations to neurons*, *Neuron* **111** (Sept., 2023) 2742–2755.e4. Publisher: Elsevier.

[25] A. S. Ecker, F. H. Sinz, E. Froudarakis, P. G. Fahey, S. A. Cadena, E. Y. Walker, E. Cobos, J. Reimer, A. S. Tolias, and M. Bethge, *A rotation-equivariant convolutional neural network model of primary visual cortex*, Dec., 2018.

[26] S. A. Cadena, G. H. Denfield, E. Y. Walker, L. A. Gatys, A. S. Tolias, M. Bethge, and A. S. Ecker, *Deep convolutional models improve predictions of macaque V1 responses to natural images*, *PLOS Computational Biology* **15** (Apr., 2019) e1006897. Publisher: Public Library of Science.

[27] W. F. Kindel, E. D. Christensen, and J. Zylberberg, *Using deep learning to probe the neural code for images in primary visual cortex*, *Journal of Vision* **19** (Apr., 2019) 29.

[28] D. Klindt, A. S. Ecker, T. Euler, and M. Bethge, *Neural system identification for large populations separating " what" and " where"*, in *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017.

[29] Y. Zhang, T. S. Lee, M. Li, F. Liu, and S. Tang, *Convolutional neural network models of V1 responses to complex patterns*, *Journal of Computational Neuroscience* **46** (Feb., 2019) 33–54.

[30] D. L. Yamins, H. Hong, C. Cadieu, and J. J. DiCarlo, *Hierarchical Modular Optimization of Convolutional Networks Achieves Representations Similar to Macaque IT and Human Ventral Stream*, in *Advances in Neural Information Processing Systems*, vol. 26, Curran Associates, Inc., 2013.

[31] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, *Performance-optimized hierarchical models predict neural responses in higher visual cortex*, *Proceedings of the National Academy of Sciences* **111** (June, 2014) 8619–8624. Publisher: Proceedings of the National Academy of Sciences.

[32] J. J. Gibson, *The Ecological Approach to Visual Perception: Classic Edition.* Psychology Press, Nov., 2014. Google-Books-ID: 8BSLBQAAQBAJ.

[33] C. Stringer, M. Pachitariu, N. Steinmetz, C. B. Reddy, M. Carandini, and K. D. Harris, *Spontaneous behaviors drive multidimensional, brainwide activity*, *Science* **364** (Apr., 2019) eaav7893. Publisher: American Association for the Advancement of Science.

[34] C. M. Niell and M. P. Stryker, *Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex*, *Neuron* **65** (Feb., 2010) 472–479.

[35] J. Duan, S. Yu, H. L. Tan, H. Zhu, and C. Tan, *A Survey of Embodied AI: From Simulators to Research Tasks*, *IEEE Transactions on Emerging Topics in Computational Intelligence* **6** (Apr., 2022) 230–244.

[36] M. Deitke, D. Batra, Y. Bisk, T. Campari, A. X. Chang, D. S. Chaplot, C. Chen, C. P. D'Arpino, K. Ehsani, A. Farhadi, L. Fei-Fei, A. Francis, C. Gan, K. Grauman, D. Hall, W. Han, U. Jain, A. Kembhavi, J. Krantz, S. Lee, C. Li, S. Majumder, O. Maksymets, R. Martín-Martín, R. Mottaghi, S. Raychaudhuri, M. Roberts, S. Savarese, M. Savva, M. Shridhar, N. Sünderhauf, A. Szot, B. Talbot, J. B. Tenenbaum, J. Thomason, A. Toshev, J. Truong, L. Weihs, and J. Wu, *Retrospectives on the Embodied AI Workshop*, Dec., 2022. arXiv:2210.06849 [cs].

[37] J. D. Boyd, I. Khaytin, and V. A. Casagrande, *Evolution of the Visual System in Mammals – Comparative Evolutionary Aspects across Orders*, in *Encyclopedia of Neuroscience* (M. D. Binder, N. Hirokawa, and U. Windhorst, eds.), pp. 1448–1455. Springer, Berlin, Heidelberg, 2009.

[38] R. H. Masland, *The Neuronal Organization of the Retina*, *Neuron* **76** (Oct., 2012) 266–280. Publisher: Elsevier.

[39] G. Field and E. Chichilnisky, *Information Processing in the Primate Retina: Circuitry and Coding*, *Annual Review of Neuroscience* **30** (July, 2007) 1–30.

[40] R. E. Marc, B. W. Jones, C. B. Watt, and E. Strettoi, *Neural remodeling in retinal degeneration*, *Progress in Retinal and Eye Research* **22** (Sept., 2003) 607–655.

[41] B. W. Jones, R. L. Pfeiffer, W. D. Ferrell, C. B. Watt, M. Marmor, and R. E. Marc, *Retinal remodeling in human retinitis pigmentosa*, *Exp Eye Res* (Mar., 2016).

[42] D. H. Hubel and T. N. Wiesel, *Receptive fields of single neurones in the cat's striate cortex*, *The Journal of Physiology* **148** (Oct., 1959) 574–591.

[43] D. H. Hubel and T. N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*, *The Journal of Physiology* **160** (Jan., 1962) 106–154.2.

[44] T. Z. Lauritzen and K. D. Miller, *Different Roles for Simple-Cell and Complex-Cell Inhibition in V1*, *The Journal of Neuroscience* **23** (Nov., 2003) 10201–10213.

[45] R. L. De Valois and K. K. De Valois, *Spatial vision*, *Annual Review of Psychology* **31** (1980) 309–341.

[46] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman, *Receptive-field dynamics in the central visual pathways*, *Trends in Neurosciences* **18** (Oct., 1995) 451–458.

[47] M. Carandini, D. J. Heeger, and J. Anthony Movshon, *Linearity and Gain Control in V1 Simple Cells*, in *Cerebral Cortex* (E. G. Jones, A. Peters, P. S. Ulinski, E. G. Jones, and A. Peters, eds.), vol. 13, pp. 401–443. Springer US, Boston, MA, 1999. Series Title: Cerebral Cortex.

[48] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, *Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex*, *The Journal of Physiology* **283** (Oct., 1978) 101–120.

[49] K. R. Gegenfurtner and D. C. Kiper, *Color vision*, *Annual Review of Neuroscience* **26** (2003) 181–206. Place: US Publisher: Annual Reviews.

[50] G. D. Horwitz and C. A. Hass, *Nonlinear analysis of macaque V1 color tuning reveals cardinal directions for cortical color processing*, *Nature Neuroscience* **15** (June, 2012) 913–919. Publisher: Nature Publishing Group.

[51] C. M. Niell and M. Scanziani, *How Cortical Circuits Implement Cortical Computations: Mouse Visual Cortex as a Model*, *Annual Review of Neuroscience* **44** (2021), no. 1 517–546. _eprint: https://doi.org/10.1146/annurev-neuro-102320-085825.

[52] C. M. Niell, *Cell Types, Circuits, and Receptive Fields in the Mouse Visual Cortex*, *Annual Review of Neuroscience* **38** (2015), no. 1 413–431. _eprint: https://doi.org/10.1146/annurev-neuro-071714-033807.

[53] A. F. Meyer, J. Poort, J. O'Keefe, M. Sahani, and J. F. Linden, *A Head-Mounted Camera System Integrates Detailed Behavioral Monitoring with Multichannel Electrophysiology in Freely Moving Mice*, *Neuron* **100** (Oct., 2018) 46–60.e7.

[54] G. Guitchounts, J. Masís, S. B. E. Wolff, and D. Cox, *Encoding of 3D Head Orienting Movements in the Primary Visual Cortex*, *Neuron* **108** (Nov., 2020) 512–525.e4.

[55] P. R. L. Parker, E. T. T. Abe, E. S. P. Leonard, D. M. Martins, and C. M. Niell, *Joint coding of visual input and eye/head position in V1 of freely moving mice*, *Neuron* (Sept., 2022).

[56] E. Froudarakis, P. G. Fahey, J. Reimer, S. M. Smirnakis, E. J. Tehovnik, and A. S. Tolias, *The Visual Cortex in Context*, *Annual Review of Vision Science* **5** (2019), no. 1 317–339. _eprint: https://doi.org/10.1146/annurev-vision-091517-034407.

[57] P. R. L. Parker, M. A. Brown, M. C. Smear, and C. M. Niell, *Movement-Related Signals in Sensory Areas: Roles in Natural Behavior*, *Trends in Neurosciences* **43** (Aug., 2020) 581–595.

[58] S. Schröder, N. A. Steinmetz, M. Krumin, M. Pachitariu, M. Rizzi, L. Lagnado, K. D. Harris, and M. Carandini, *Arousal Modulates Retinal Output*, *Neuron* **107** (Aug., 2020) 487–495.e9.

[59] K. Franke, K. F. Willeke, K. Ponder, M. Galdamez, N. Zhou, T. Muhammad, S. Patel, E. Froudarakis, J. Reimer, F. H. Sinz, and A. S. Tolias, *State-dependent pupil dilation rapidly shifts visual feature selectivity*, *Nature* **610** (Oct., 2022) 128–134. Number: 7930 Publisher: Nature Publishing Group.

[60] L. Busse, A. Ayaz, N. T. Dhruv, S. Katzner, A. B. Saleem, M. L. Schölvinck, A. D. Zaharia, and M. Carandini, *The Detection of Visual Contrast in the Behaving Mouse*, *Journal of Neuroscience* **31** (Aug., 2011) 11351–11361. Publisher: Society for Neuroscience Section: Articles.

[61] M. G. Shuler and M. F. Bear, *Reward Timing in the Primary Visual Cortex*, *Science* **311** (Mar., 2006) 1606–1609. Publisher: American Association for the Advancement of Science.

[62] A. M. Kerlin, M. L. Andermann, V. K. Berezovskii, and R. C. Reid, *Broadly Tuned Response Properties of Diverse Inhibitory Neuron Subtypes in Mouse Visual Cortex*, *Neuron* **67** (Sept., 2010) 858–871.

[63] P. Dayan and L. Abbott, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* MIT Press, 2014.

[64] O. Schwartz, E. J. Chichilnisky, and E. P. Simoncelli, *Characterizing neural gain control using spike-triggered covariance*, in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, NIPS'01, (Cambridge, MA, USA), pp. 269–276, MIT Press, Jan., 2001.

[65] E. J. Chichilnisky, *A simple white noise analysis of neuronal light responses*, *Network: Computation in Neural Systems* **12** (Jan., 2001) 199–213. Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/net.12.2.199.213.

[66] E. J. Chichilnisky and D. A. Baylor, *Receptive-field microstructure of blue-yellow ganglion cells in primate retina*, *Nature Neuroscience* **2** (Oct., 1999) 889–893. Number: 10 Publisher: Nature Publishing Group.

[67] D. Chander and E. J. Chichilnisky, *Adaptation to Temporal Contrast in Primate and Salamander Retina*, *Journal of Neuroscience* **21** (Dec., 2001) 9904–9916. Publisher: Society for Neuroscience Section: ARTICLE.

[68] J. Freeman, G. D. Field, P. H. Li, M. Greschner, D. E. Gunning, K. Mathieson, A. Sher, A. M. Litke, L. Paninski, E. P. Simoncelli, and E. Chichilnisky, *Mapping nonlinear receptive field structure in primate retina at single cone resolution*, *eLife* **4** (Oct., 2015) e05241. Publisher: eLife Sciences Publications, Ltd.

[69] J. K. Liu, H. M. Schreyer, A. Onken, F. Rozenblit, M. H. Khani, V. Krishnamoorthy, S. Panzeri, and T. Gollisch, *Inference of neuronal functional circuitry with spike-triggered non-negative matrix factorization*, *Nature Communications* **8** (July, 2017) 149. Number: 1 Publisher: Nature Publishing Group.

[70] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli, *Spatio-temporal correlations and visual signalling in a complete neuronal population*, *Nature* **454** (Aug., 2008) 995–999.

[71] A. Heitman, N. Brackbill, M. Greschner, A. Sher, A. M. Litke, and E. Chichilnisky, *Testing pseudo-linear models of responses to natural scenes in primate retina*, preprint, Neuroscience, Mar., 2016.

[72] T. H. Moskovitz, N. A. Roy, and J. W. Pillow, *A comparison of deep learning and linear-nonlinear cascade approaches to neural encoding*, *bioRxiv* (Nov., 2018) 463422. Publisher: Cold Spring Harbor Laboratory Section: New Results.

[73] Q. Yan, Y. Zheng, S. Jia, Y. Zhang, Z. Yu, F. Chen, Y. Tian, T. Huang, and J. K. Liu, *Revealing Fine Structures of the Retinal Receptive Field by Deep-Learning Networks*, *IEEE Transactions on Cybernetics* (2020) 1–12. Conference Name: IEEE Transactions on Cybernetics.

[74] E. Batty, J. Merel, N. Brackbill, A. Heitman, A. Sher, A. Litke, E. J. Chichilnisky, and L. Paninski, *Multilayer Recurrent Network Models of Primate Retinal Ganglion Cell Responses*, in *ICLR*, 2017.

[75] M. Stimberg, R. Brette, and D. F. Goodman, *Brian 2, an intuitive and efficient neural simulator*, *eLife* **8** (Aug., 2019) e47314. Publisher: eLife Sciences Publications, Ltd.

[76] M. Stimberg, D. F. M. Goodman, and T. Nowotny, *Brian2GeNN: accelerating spiking neural network simulations with graphics hardware*, *Scientific Reports* **10** (Jan., 2020) 410. Bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Number: 1 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Computational neuroscience;Software Subject_term_id: computational-neuroscience;software.

[77] A. L. Hodgkin and A. F. Huxley, *A quantitative description of membrane current and its application to conduction and excitation in nerve*, *The Journal of Physiology* **117** (Aug., 1952) 500–544.

[78] J. F. Fohlmeister and R. F. Miller, *Impulse Encoding Mechanisms of Ganglion Cells in the Tiger Salamander Retina*, *Journal of Neurophysiology* **78** (Oct., 1997) 1935–1947.

[79] J. Fohlmeister, P. Coleman, and R. Miller, *Modeling the repetitive firing of retinal ganglion cells*, *Brain Research* **510** (Mar., 1990) 343–345.

[80] T. Guo, D. Tsai, J. W. Morley, G. J. Suaning, T. Kameneva, N. H. Lovell, and S. Dokos, *Electrical activity of ON and OFF retinal ganglion cells: a modelling study*, *Journal of Neural Engineering* **13** (Feb., 2016) 025005. Publisher: IOP Publishing.

[81] A. C. Weitz, M. R. Behrend, A. K. Ahuja, P. Christopher, J. Wei, V. Wuyyuru, U. Patel, R. J. Greenberg, M. S. Humayun, R. H. Chow, and J. D. Weiland, *Interphase gap as a means to reduce electrical stimulation thresholds for epiretinal prostheses*, *Journal of Neural Engineering* **11** (Feb., 2014) 016007.

[82] N. A. L. Chenais, M. J. I. A. Leccardi, and D. Ghezzi, *Capacitive-like photovoltaic epiretinal stimulation enhances and narrows the network-mediated activity of retinal ganglion cells by recruiting the lateral inhibitory network*, *Journal of Neural Engineering* **16** (Oct., 2019) 066009. Publisher: IOP Publishing.

[83] S. A. Joarder, M. Abramian, G. J. Suaning, N. H. Lovell, and S. Dokos, *A continuum model of retinal electrical stimulation*, *Journal of Neural Engineering* **8** (Oct., 2011) 066006. Publisher: IOP Publishing.

[84] M. Abramian, N. H. Lovell, A. Habib, J. W. Morley, G. J. Suaning, and S. Dokos, *Quasi-monopolar electrical stimulation of the retina: a computational modelling study*, *Journal of Neural Engineering* **11** (Feb., 2014) 025002. Publisher: IOP Publishing.

[85] S. Dokos, G. J. Suaning, and N. H. Lovell, *A bidomain model of epiretinal stimulation*, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **13** (June, 2005) 137–146. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[86] T. Guo, N. H. Lovell, D. Tsai, P. Twyford, S. Fried, J. W. Morley, G. J. Suaning, and S. Dokos, *Selective activation of ON and OFF retinal ganglion cells to high-frequency electrical stimulation: A computational modeling study*, in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6108–6111, Aug., 2014. ISSN: 1558-4615.

[87] X. Li and L. Li, *An Optimized Computational Model of Retinal Ganglion Cells and Electrical Stimulation with Varied Epiretinal Electrode*, in *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation*, pp. 133–136, June, 2015. ISSN: 2157-1481.

[88] S. Qiu, X. Li, X. Li, and L. Li, *A comparative study on electrically evoked responses of retinal ganglion cells in distinct retinal areas by computational model*, in *2017 8th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 308–311, May, 2017. ISSN: 1948-3554.

[89] K. Loizos, G. Lazzi, J. S. Lauritzen, J. Anderson, B. W. Jones, and R. Marc, *A Multi-Scale Computational Model for the Study of Retinal Prosthetic Stimulation*, *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference* **2014** (2014) 6100–6103.

[90] K. Loizos, A. K. RamRakhyani, J. Anderson, R. Marc, and G. Lazzi, *On the computation of a retina resistivity profile for applications in multi-scale modeling of electrical stimulation and absorption*, *Physics in Medicine and Biology* **61** (May, 2016) 4491–4505. Publisher: IOP Publishing.

[91] Y. Lian, A. Almasi, D. B. Grayden, T. Kameneva, A. N. Burkitt, and H. Meffin, *Learning receptive field properties of complex cells in V1*, *PLoS Computational Biology* **17** (Mar., 2021) e1007957.

[92] J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, *Spatial summation in the receptive fields of simple cells in the cat's striate cortex.*, *The Journal of Physiology* **283** (Oct., 1978) 53–77.

[93] M. Carandini, *What simple and complex cells compute*, *The Journal of Physiology* **577** (Dec., 2006) 463–466.

[94] A. Eh and B. Jr, *Spatiotemporal energy models for the perception of motion*, *Journal of the Optical Society of America. A, Optics and image science* **2** (Feb., 1985). Publisher: J Opt Soc Am A.

[95] L. M. Martinez and J.-M. Alonso, *Complex receptive fields in primary visual cortex*, *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry* **9** (Oct., 2003) 317–331.

[96] G. Pogoncheff, J. Granley, and M. Beyeler, *Explaining V1 Properties with a Biologically Constrained Deep Learning Architecture*, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., Nov., 2023.

[97] F. Sinz, A. S. Ecker, P. Fahey, E. Walker, E. Cobos, E. Froudarakis, D. Yatsenko, Z. Pitkow, J. Reimer, and A. Tolias, *Stimulus domain transfer in recurrent models for large scale cortical population prediction on video*, in *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018.

[98] K.-K. Lurz, M. Bashiri, K. Willeke, A. Jagadish, E. Wang, E. Y. Walker, S. A. Cadena, T. Muhammad, E. Cobos, A. S. Tolias, A. S. Ecker, and F. H. Sinz, *Generalization in data-driven models of primary visual cortex*, Jan., 2021.

[99] T. Guo, D. Tsai, S. Bai, J. W. Morley, G. J. Suaning, N. H. Lovell, and S. Dokos, *Understanding the Retina: A Review of Computational Models of the Retina from the Single Cell to the Network Level*, *Critical Reviews™ in Biomedical Engineering* **42** (2014), no. 5. Publisher: Begel House Inc.

[100] A. Wohrer and P. Kornprobst, *Virtual Retina: A biological retina model and simulator, with contrast gain control*, *Journal of Computational Neuroscience* **26** (Apr., 2009) 219–249.

[101] R. G. Smith, *Simulation of an anatomically defined local circuit: The cone-horizontal cell network in cat retina*, *Visual Neuroscience* **12** (May, 1995) 545–561. Publisher: Cambridge University Press.

[102] R. J. Greenberg, T. J. Velte, M. S. Humayun, G. N. Scarlatis, and E. de Juan, *A computational model of electrical stimulation of the retinal ganglion cell*, *IEEE Trans Biomed Eng* **46** (May, 1999) 505–14.

[103] N. P. Cottaris and S. D. Elfar, *How the retinal network reacts to epiretinal stimulation to form the prosthetic visual input to the cortex*, *Journal of Neural Engineering* **2** (Feb., 2005) S74–S90. Publisher: IOP Publishing.

[104] P. Werginz, M. Im, A. E. Hadjinicolaou, and S. I. Fried, *Visual and electric spiking responses of seven types of rabbit retinal ganglion cells*, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2434–2437, July, 2018. ISSN: 1558-4615.

[105] M. Beyeler, *Biophysical model of axonal stimulation in epiretinal visual prostheses*, in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 348–351, Mar., 2019. ISSN: 1948-3554.

[106] J. Paknahad, K. Loizos, M. Humayun, and G. Lazzi, *Targeted Stimulation of Retinal Ganglion Cells in Epiretinal Prostheses: A Multiscale Computational Study*, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **28** (Nov., 2020) 2548–2556. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[107] T. Kameneva, H. Meffin, and A. N. Burkitt, *Modelling intrinsic electrophysiological properties of ON and OFF retinal ganglion cells*, *Journal of Computational Neuroscience* **31** (Nov., 2011) 547–561.

[108] R. W. Rodieck, *The first steps in seeing.* The first steps in seeing. Sinauer Associates, Sunderland, MA, US, 1998. Pages: xi, 562.

[109] H. Wassle and B. B. Boycott, *Functional architecture of the mammalian retina*, *Physiological Reviews* **71** (Apr., 1991) 447–480. Publisher: American Physiological Society.

[110] J. Wyatt and J. Rizzo, *Ocular implants for the blind*, *IEEE Spectrum* **33** (May, 1996) 47–53. Conference Name: IEEE Spectrum.

[111] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, *BioNumbers—the database of key numbers in molecular and cell biology*, *Nucleic Acids Research* **38** (Jan., 2010) D750–D753.

[112] J. Crooks and H. Kolb, *Localization of GABA, glycine, glutamate and tyrosine hydroxylase in the human retina*, *The Journal of Comparative Neurology* **315** (Jan., 1992) 287–302.

[113] X. Tao, J. Sabharwal, S. M. Wu, and B. J. Frankfort, *Intraocular Pressure Elevation Compromises Retinal Ganglion Cell Light Adaptation*, *Investigative Opthalmology & Visual Science* **61** (Oct., 2020) 15.

[114] H. Kolb, R. F. Nelson, P. K. Ahnelt, I. Ortuño-Lizarán, and N. Cuenca, *The Architecture of the Human Fovea*, in *Webvision: The Organization of the Retina and Visual System* (H. Kolb, E. Fernandez, and R. Nelson, eds.). University of Utah Health Sciences Center, Salt Lake City (UT), 1995.

[115] G. Zeck, A. Lambacher, and P. Fromherz, *Axonal Transmission in the Retina Introduces a Small Dispersion of Relative Timing in the Ganglion Cell Population Response*, *PLOS ONE* **6** (June, 2011) e20810. Publisher: Public Library of Science.

[116] Q. Shi, P. Gupta, A. K. Boukhvalova, J. H. Singer, and D. A. Butts, *Functional characterization of retinal ganglion cells using tailored nonlinear modeling*, *Scientific Reports* **9** (Dec., 2019) 8713.

[117] K. Stingl, K. U. Bartz-Schmidt, D. Besch, A. Braun, A. Bruckmann, F. Gekeler, U. Greppmaier, S. Hipp, G. Hortdorfer, C. Kernstock, A. Koitschev, A. Kusnyerik, H. Sachs, A. Schatz, K. T. Stingl, T. Peters, B. Wilhelm, and E. Zrenner, *Artificial vision with wirelessly powered subretinal electronic implant alpha-IMS*, *Proc Biol Sci* **280** (Apr., 2013) 20130077.

[118] Y. H.-L. Luo, J. J. Zhong, M. Clemo, and L. d. Cruz, *Long-term Repeatability and Reproducibility of Phosphene Characteristics in Chronically Implanted Argus II Retinal Prosthesis Subjects*, *American Journal of Ophthalmology* **170** (Oct., 2016) 100–109. Publisher: Elsevier.

[119] D. Palanker, Y. Le Mer, S. Mohand-Said, M. Muqit, and J. A. Sahel, *Photovoltaic Restoration of Central Vision in Atrophic Age-Related Macular Degeneration*, *Ophthalmology* (Feb., 2020).

[120] M. Pu, L. Xu, and H. Zhang, *Visual Response Properties of Retinal Ganglion Cells in the Royal College of Surgeons Dystrophic Rat*, *Investigative Ophthalmology & Visual Science* **47** (Aug., 2006) 3579–3585.

[121] S. F. Stasheff, *Emergence of Sustained Spontaneous Hyperactivity and Temporary Preservation of off Responses in Ganglion Cells of the Retinal Degeneration (rd1) Mouse*, *Journal of Neurophysiology* **99** (Mar., 2008) 1408–1421. Publisher: American Physiological Society.

[122] S. F. Stasheff, M. Shankar, and M. P. Andrews, *Developmental time course distinguishes changes in spontaneous and light-evoked retinal ganglion cell activity in rd1 and rd10 mice*, *Journal of Neurophysiology* **105** (June, 2011) 3002–3009. Publisher: American Physiological Society.

[123] M. Telias, B. Denlinger, Z. Helft, C. Thornton, B. Beckwith-Cohen, and R. H. Kramer, *Retinoic Acid Induces Hyperactivity, and Blocking Its Receptor Unmasks Light Responses and Augments Vision in Retinal Degeneration*, *Neuron* **102** (May, 2019) 574–586.e5.

[124] D. J. Margolis, A. J. Gartland, J. H. Singer, and P. B. Detwiler, *Network Oscillations Drive Correlated Spiking of ON and OFF Ganglion Cells in the rd1 Mouse Model of Retinal Degeneration*, *PLOS ONE* **9** (Jan., 2014) e86253. Publisher: Public Library of Science.

[125] J. Ahn, S. Cha, K.-E. Choi, S.-W. Kim, Y. Yoo, and Y. S. Goo, *Correlated Activity in the Degenerate Retina Inhibits Focal Response to Electrical Stimulation*, *Frontiers in Cellular Neuroscience* **16** (2022).

[126] J. F. Rizzo, III, J. Wyatt, J. Loewenstein, S. Kelly, and D. Shire, *Methods and Perceptual Thresholds for Short-Term Electrical Stimulation of Human Retina with*

Microelectrode Arrays, *Investigative Ophthalmology & Visual Science* **44** (Dec., 2003) 5355–5361.

[127] T. M. O'Hearn, S. R. Sadda, J. D. Weiland, M. Maia, E. Margalit, and M. S. Humayun, *Electrical stimulation in normal and retinal degeneration (rd1) isolated mouse retina*, *Vision Research* **46** (Oct., 2006) 3198–3204.

[128] R. J. Jensen and J. F. Rizzo, *Activation of retinal ganglion cells in wild-type and rd1 mice through electrical stimulation of the retinal neural network*, *Vision Research* **48** (June, 2008) 1562–1568.

[129] Y. S. Goo, J. H. Ye, S. Lee, Y. Nam, S. B. Ryu, and K. H. Kim, *Retinal ganglion cell responses to voltage and current stimulation in wild-type and rd1 mouse retinas*, *Journal of Neural Engineering* **8** (May, 2011) 035003. Publisher: IOP Publishing.

[130] A. Cho, C. Ratliff, A. Sampath, and J. Weiland, *Changes in ganglion cell physiology during retinal degeneration influence excitability by prosthetic electrodes*, *Journal of Neural Engineering* **13** (Feb., 2016) 025001. Publisher: IOP Publishing.

[131] W. Haq, B. Arango-Gonzalez, E. Zrenner, T. Euler, and T. Schubert, *Synaptic remodeling generates synchronous oscillations in the degenerated outer mouse retina*, *Frontiers in Neural Circuits* **8** (2014).

[132] S. Trenholm and G. B. Awatramani, *Origins of spontaneous activity in the degenerating retina*, *Frontiers in Cellular Neuroscience* **9** (2015).

[133] J. R. Golden, C. Erickson-Davis, N. P. Cottaris, N. Parthasarathy, F. Rieke, D. Brainard, B. Wandell, and E. J. Chichilnisky, *Simulation of visual perception and learning with a retinal prosthesis*, *Journal of Neural Engineering* (2018).

[134] J. Paknahad, K. Loizos, L. Yue, M. S. Humayun, and G. Lazzi, *Color and cellular selectivity of retinal ganglion cell subtypes through frequency modulation of electrical stimulation*, *Scientific Reports* **11** (Mar., 2021) 1–13. Number: 1 Publisher: Nature Publishing Group.

[135] K. Loizos, R. Marc, M. Humayun, J. R. Anderson, B. W. Jones, and G. Lazzi, *Increasing Electrical Stimulation Efficacy in Degenerated Retina: Stimulus Waveform Design in a Multiscale Computational Model*, *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **26** (June, 2018) 1111–1120. Conference Name: IEEE Transactions on Neural Systems and Rehabilitation Engineering.

[136] J. D. Wiley and J. G. Webster, *Analysis and Control of the Current Distribution under Circular Dispersive Electrodes*, *IEEE Transactions on Biomedical Engineering* **BME-29** (May, 1982) 381–385. Conference Name: IEEE Transactions on Biomedical Engineering.

[137] D. Knuth, *The Art of Computer Programming : Volume 2 : Seminumerical Algorithms*. Addison-Wesley Ser. in Computer Science & Information Processing. Reading [etc.] : Addison-Wesley Publishing Company, 1969.

[138] J. F. Rizzo, J. Wyatt, J. Loewenstein, S. Kelly, and D. Shire, *Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials*, *Invest Ophthalmol Vis Sci* **44** (Dec., 2003) 5362–9.

[139] S. I. Fried, A. C. W. Lasker, N. J. Desai, D. K. Eddington, and J. F. Rizzo, *Axonal Sodium-Channel Bands Shape the Response to Electric Stimulation in Retinal Ganglion Cells*, *Journal of Neurophysiology* **101** (Apr., 2009) 1972–1987.

[140] M. Beyeler, D. Nanduri, J. D. Weiland, A. Rokem, G. M. Boynton, and I. Fine, *A model of ganglion axon pathways accounts for percepts elicited by retinal implants*, *Scientific Reports* **9** (June, 2019) 1–16.

[141] C. Sekirnjak, C. Hulse, L. H. Jepson, P. Hottowy, A. Sher, W. Dabrowski, A. M. Litke, and E. J. Chichilnisky, *Loss of Responses to Visual But Not Electrical Stimulation in Ganglion Cells of Rats With Severe Photoreceptor Degeneration*, *Journal of Neurophysiology* **102** (Dec., 2009) 3260–3269. Publisher: American Physiological Society.

[142] C. Sekirnjak, L. H. Jepson, P. Hottowy, A. Sher, W. Dabrowski, A. M. Litke, and E. J. Chichilnisky, *Changes in physiological properties of rat ganglion cells during retinal degeneration*, *Journal of Neurophysiology* **105** (May, 2011) 2560–2571.

[143] W.-Q. Yu, N. M. Grzywacz, E.-J. Lee, and G. D. Field, *Cell type-specific changes in retinal ganglion cell function induced by rod death and cone reorganization in rats*, *Journal of Neurophysiology* **118** (July, 2017) 434–454.

[144] C. Y. Yang, D. Tsai, T. Guo, S. Dokos, G. J. Suaning, J. W. Morley, and N. H. Lovell, *Differential electrical responses in retinal ganglion cell subtypes: effects of synaptic blockade and stimulating electrode location*, *Journal of Neural Engineering* **15** (Aug., 2018) 046020.

[145] M. Im and S. I. Fried, *Indirect activation elicits strong correlations between light and electrical responses in ON but not OFF retinal ganglion cells*, *J Physiol* **593** (Aug., 2015) 3577–96.

[146] P. Twyford, C. Cai, and S. Fried, *Differential responses to high-frequency electrical stimulation in ON and OFF retinal ganglion cells*, *J Neural Eng* **11** (Apr., 2014) 025001.

[147] Y.-C. Chang, D. H. Ghaffari, R. H. Chow, and J. D. Weiland, *Stimulation strategies for selective activation of retinal ganglion cell soma and threshold reduction*, *Journal of Neural Engineering* **16** (Feb., 2019) 026017.

[148] W. Tong, H. Meffin, D. J. Garrett, and M. R. Ibbotson, *Stimulation Strategies for Improving the Resolution of Retinal Prostheses*, *Frontiers in Neuroscience* **14** (2020). Publisher: Frontiers.

[149] T. Euler and T. Schubert, *Multiple Independent Oscillatory Networks in the Degenerating Retina*, *Frontiers in Cellular Neuroscience* **9** (Nov., 2015).

[150] C. W. Yee, A. H. Toychiev, and B. T. Sagdullaev, *Network Deficiency Exacerbates Impairment in a Mouse Model of Retinal Degeneration*, *Frontiers in Systems Neuroscience* **6** (2012).

[151] Y. S. Goo, D. J. Park, J. R. Ahn, and S. S. Senok, *Spontaneous Oscillatory Rhythms in the Degenerating Mouse Retina Modulate Retinal Ganglion Cell Responses to Electrical Stimulation*, *Frontiers in Cellular Neuroscience* **9** (Jan., 2016).

[152] Z. Chen, Y. Song, J. Yao, C. Weng, and Z. Q. Yin, *Alterations of Sodium and Potassium Channels of RGCs in RCS Rat with the Development of Retinal Degeneration*, *Journal of Molecular Neuroscience* **51** (Nov., 2013) 976–985.

[153] S. Saha, U. Greferath, K. A. Vessey, D. B. Grayden, A. N. Burkitt, and E. L. Fletcher, *Changes in ganglion cells during retinal degeneration*, *Neuroscience* **329** (Aug., 2016) 1–11.

[154] H. B. Barlow, *Possible Principles Underlying the Transformations of Sensory Messages*, in *Sensory Communication* (W. A. Rosenblith, ed.), pp. 216–234. The MIT Press, Sept., 2012.

[155] E. P. Simoncelli and B. A. Olshausen, *Natural Image Statistics and Neural Representation*, *Annual Review of Neuroscience* **24** (Mar., 2001) 1193–1216. Publisher: Annual Reviews.

[156] J. Golz and D. I. A. MacLeod, *Influence of scene statistics on colour constancy*, *Nature* **415** (Feb., 2002) 637–640. Publisher: Nature Publishing Group.

[157] D. Stansbury, T. Naselaris, and J. Gallant, *Natural Scene Statistics Account for the Representation of Scene Categories in Human Visual Cortex*, *Neuron* **79** (Sept., 2013) 1025–1034.

[158] A. D. Huberman and C. M. Niell, *What can mice tell us about how vision works?*, *Trends in Neurosciences* **34** (Sept., 2011) 464–473.

[159] J. W. Krakauer, A. A. Ghazanfar, A. Gomez-Marin, M. A. MacIver, and D. Poeppel, *Neuroscience Needs Behavior: Correcting a Reductionist Bias*, *Neuron* **93** (Feb., 2017) 480–490. Publisher: Elsevier.

[160] S. R. Datta, D. J. Anderson, K. Branson, P. Perona, and A. Leifer, *Computational Neuroethology: A Call to Action*, *Neuron* **104** (Oct., 2019) 11–24.

[161] U. Hasson, S. A. Nastase, and A. Goldstein, *Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks*, *Neuron* **105** (Feb., 2020) 416–434.

[162] L. Dyballa, M. S. Hoseini, M. C. Dadarlat, S. W. Zucker, and M. P. Stryker, *Flow stimuli reveal ecologically appropriate responses in mouse visual cortex*, *Proceedings of the National Academy of Sciences* **115** (Oct., 2018) 11304–11309. Publisher: Proceedings of the National Academy of Sciences.

[163] Y. Qiu, Z. Zhao, D. Klindt, M. Kautzky, K. P. Szatko, F. Schaeffel, K. Rifai, K. Franke, L. Busse, and T. Euler, *Natural environment statistics in the upper and lower visual field are reflected in mouse retinal specializations*, *Current Biology* **31** (Aug., 2021) 3233–3247.e6.

[164] L. Abballe and H. Asari, *Natural image statistics for mouse vision*, *PLOS ONE* **17** (Jan., 2022) e0262763. Publisher: Public Library of Science.

[165] A. F. Meyer, J. O'Keefe, and J. Poort, *Two Distinct Types of Eye-Head Coupling in Freely Moving Mice*, *Current Biology* **30** (June, 2020) 2116–2130.e6.

[166] A. M. Michaiel, E. T. Abe, and C. M. Niell, *Dynamics of gaze control during prey capture in freely moving mice*, *eLife* **9** (July, 2020) e57458. Publisher: eLife Sciences Publications, Ltd.

[167] P. R. L. Parker, D. M. Martins, E. S. P. Leonard, N. M. Casey, S. L. Sharp, E. T. T. Abe, M. C. Smear, J. L. Yates, J. F. Mitchell, and C. M. Niell, *A dynamic sequence of visual processing initiated by gaze shifts*, Aug., 2022. Pages: 2022.08.23.504847 Section: New Results.

[168] J. S. Matthis, K. S. Muller, K. L. Bonnen, and M. M. Hayhoe, *Retinal optic flow during natural locomotion*, *PLOS Computational Biology* **18** (Feb., 2022) e1009575. Publisher: Public Library of Science.

[169] J. L. Yates, S. H. Coop, G. H. Sarch, R.-J. Wu, D. A. Butts, M. Rucci, and J. F. Mitchell, *Beyond Fixation: detailed characterization of neural selectivity in free-viewing primates*, Nov., 2021. Pages: 2021.11.06.467566 Section: New Results.

[170] Y. Yu, J. N. Stirman, C. R. Dorsett, and S. L. Smith, *Selective representations of texture and motion in mouse higher visual areas*, *Current Biology* **32** (July, 2022) 2810–2820.e5.

[171] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, *DeepLabCut: markerless pose estimation of user-defined body parts with deep learning*, *Nature Neuroscience* **21** (Sept., 2018) 1281–1289. Number: 9 Publisher: Nature Publishing Group.

[172] N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Böhm, M. Broux, S. Chen, J. Colonell, R. J. Gardner, B. Karsh, F. Kloosterman, D. Kostadinov, C. Mora-Lopez, J. O'Callaghan, J. Park, J. Putzeys, B. Sauerbrei, R. J. J. van Daal, A. Z. Vollan, S. Wang, M. Welkenhuysen, Z. Ye, J. T. Dudman, B. Dutta, A. W. Hantman, K. D. Harris, A. K. Lee, E. I. Moser, J. O'Keefe, A. Renart, K. Svoboda, M. Häusser, S. Haesler, M. Carandini, and T. D. Harris, *Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings*, *Science* **372** (Apr., 2021) eabf4588. Publisher: American Association for the Advancement of Science.

[173] A. Xu, Y. Hou, C. Niell, and M. Beyeler, *Multimodal Deep Learning Model Unveils Behavioral Dynamics of V1 Activity in Freely Moving Mice*, *Advances in Neural Information Processing Systems* **36** (Dec., 2023) 15341–15357.

[174] K. Simonyan, A. Vedaldi, and A. Zisserman, *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*, Apr., 2014. arXiv:1312.6034 [cs].

[175] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, *Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization*, *International Journal of Computer Vision* **128** (Feb., 2020) 336–359. arXiv:1610.02391 [cs].

[176] M. D. Zeiler and R. Fergus, *Visualizing and Understanding Convolutional Networks*, Nov., 2013. arXiv:1311.2901 [cs].

[177] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, *Striving for Simplicity: The All Convolutional Net*, Apr., 2015. arXiv:1412.6806 [cs].

[178] N. Kokhlikyan, V. Miglani, M. Martin, E. Wang, B. Alsallakh, J. Reynolds, A. Melnikov, N. Kliushkina, C. Araya, S. Yan, and O. Reblitz-Richardson, *Captum: A unified and generic model interpretability library for PyTorch*, Sept., 2020. arXiv:2009.07896 [cs, stat].

[179] U. Güçlü and M. A. J. v. Gerven, *Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream*, *Journal of Neuroscience* **35** (July, 2015) 10005–10014. Publisher: Society for Neuroscience Section: Articles.

[180] S. A. Cadena, F. H. Sinz, T. Muhammad, E. Froudarakis, E. Cobos, E. Y. Walker, J. Reimer, M. Bethge, A. Tolias, and A. S. Ecker, *How well do deep neural networks trained on object recognition characterize the mouse visual system?*, Oct., 2019.

[181] G. T. Prusky, P. W. R. West, and R. M. Douglas, *Behavioral assessment of visual acuity in mice and rats*, *Vision Research* **40** (July, 2000) 2201–2209.

[182] C. M. Niell and M. P. Stryker, *Highly Selective Receptive Fields in Mouse Visual Cortex*, *The Journal of Neuroscience* **28** (July, 2008) 7520–7536.

[183] S. Bakhtiari, P. Mineault, T. Lillicrap, C. Pack, and B. Richards, *The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning*, in *Advances in Neural Information Processing Systems*, vol. 34, pp. 25164–25178, Curran Associates, Inc., 2021.

[184] K. F. Willeke, P. G. Fahey, M. Bashiri, L. Pede, M. F. Burg, C. Blessing, S. A. Cadena, Z. Ding, K.-K. Lurz, K. Ponder, T. Muhammad, S. S. Patel, A. S. Ecker, A. S. Tolias, and F. H. Sinz, *The Sensorium competition on predicting large-scale mouse primary visual cortex activity*, June, 2022. arXiv:2206.08666 [cs, q-bio].

[185] M. Beyeler, N. Dutt, and J. L. Krichmar, *3D Visual Response Properties of MSTd Emerge from an Efficient, Sparse Population Code*, *Journal of Neuroscience* **36** (Aug., 2016) 8399–415.

[186] L. Busse, J. A. Cardin, M. E. Chiappe, M. M. Halassa, M. J. McGinley, T. Yamashita, and A. B. Saleem, *Sensation during Active Behaviors*, *Journal of Neuroscience* **37** (Nov., 2017) 10826–10834. Publisher: Society for Neuroscience Section: Symposium and Mini-Symposium.

[187] P. M. Goltstein, S. Reinert, T. Bonhoeffer, and M. Hübener, *Mouse visual cortex areas represent perceptual and semantic features of learned visual categories*, *Nature Neuroscience* **24** (Oct., 2021) 1441–1451. Number: 10 Publisher: Nature Publishing Group.

[188] J. Antolík, S. B. Hofer, J. A. Bednar, and T. D. Mrsic-Flogel, *Model Constrained by Visual Hierarchy Improves Prediction of Neural Responses to Natural Scenes*, *PLOS Computational Biology* **12** (June, 2016) e1004927. Publisher: Public Library of Science.

[189] M. Bashiri, E. Walker, K.-K. Lurz, A. Jagadish, T. Muhammad, Z. Ding, Z. Ding, A. Tolias, and F. Sinz, *A flow-based latent state generative model of neural population responses to natural images*, in *Advances in Neural Information Processing Systems*, vol. 34, pp. 15801–15815, Curran Associates, Inc., 2021.

[190] J. Shi, E. Shea-Brown, and M. Buice, *Comparison Against Task Driven Artificial Neural Networks Reveals Functional Properties in Mouse Visual Cortex*, in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.

[191] J. Shi, B. Tripp, E. Shea-Brown, S. Mihalas, and M. A. Buice, *MouseNet: A biologically constrained convolutional neural network model for the mouse visual cortex*, *PLOS Computational Biology* **18** (Sept., 2022) e1010427. Publisher: Public Library of Science.

[192] J. A. Harris, S. Mihalas, K. E. Hirokawa, J. D. Whitesell, H. Choi, A. Bernard, P. Bohn, S. Caldejon, L. Casal, A. Cho, A. Feiner, D. Feng, N. Gaudreault, C. R. Gerfen, N. Graddis, P. A. Groblewski, A. M. Henry, A. Ho, R. Howard, J. E. Knox, L. Kuan, X. Kuang, J. Lecoq, P. Lesnar, Y. Li, J. Luviano, S. McConoughey, M. T. Mortrud, M. Naeemi, L. Ng, S. W. Oh, B. Ouellette, E. Shen, S. A. Sorensen, W. Wakeman, Q. Wang, Y. Wang, A. Williford, J. W. Phillips, A. R. Jones, C. Koch, and H. Zeng, *Hierarchical organization of cortical and thalamic connectivity*, *Nature* **575** (Nov., 2019) 195–202. Number: 7781 Publisher: Nature Publishing Group.

[193] J. H. Siegle, X. Jia, S. Durand, S. Gale, C. Bennett, N. Graddis, G. Heller, T. K. Ramirez, H. Choi, J. A. Luviano, P. A. Groblewski, R. Ahmed, A. Arkhipov, A. Bernard, Y. N. Billeh, D. Brown, M. A. Buice, N. Cain, S. Caldejon, L. Casal, A. Cho, M. Chvilicek, T. C. Cox, K. Dai, D. J. Denman, S. E. J. de Vries, R. Dietzman, L. Esposito, C. Farrell, D. Feng, J. Galbraith, M. Garrett, E. C. Gelfand, N. Hancock, J. A. Harris, R. Howard, B. Hu, R. Hytnen, R. Iyer, E. Jessett, K. Johnson, I. Kato, J. Kiggins, S. Lambert, J. Lecoq, P. Ledochowitsch, J. H. Lee, A. Leon, Y. Li, E. Liang, F. Long, K. Mace, J. Melchior, D. Millman, T. Mollenkopf, C. Nayan, L. Ng, K. Ngo, T. Nguyen, P. R. Nicovich, K. North, G. K. Ocker, D. Ollerenshaw, M. Oliver, M. Pachitariu, J. Perkins, M. Reding, D. Reid, M. Robertson, K. Ronellenfitch, S. Seid, C. Slaughterbeck, M. Stoecklin, D. Sullivan, B. Sutton, J. Swapp, C. Thompson, K. Turner, W. Wakeman, J. D. Whitesell, D. Williams, A. Williford, R. Young, H. Zeng, S. Naylor, J. W. Phillips, R. C. Reid, S. Mihalas, S. R. Olsen, and C. Koch, *Survey of spiking in the mouse visual system reveals functional hierarchy*, *Nature* **592** (Apr., 2021) 86–92. Number: 7852 Publisher: Nature Publishing Group.

[194] A. B. Saleem, *Two stream hypothesis of visual processing for navigation in mouse*, *Current Opinion in Neurobiology* **64** (Oct., 2020) 70–78.

[195] K. He, X. Zhang, S. Ren, and J. Sun, *Deep Residual Learning for Image Recognition*, Dec., 2015. arXiv:1512.03385 [cs].

[196] M. Tan and Q. V. Le, *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*, Sept., 2020. arXiv:1905.11946 [cs, stat] version: 5.

[197] G. Van den Bergh, B. Zhang, L. Arckens, and Y. M. Chino, *Receptive-field properties of V1 and V2 neurons in mice and macaque monkeys*, *Journal of Comparative Neurology* **518** (2010), no. 11 2051–2070. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.22321.

[198] M. Beyeler, E. L. Rounds, K. D. Carlson, N. Dutt, and J. L. Krichmar, *Neural correlates of sparse coding and dimensionality reduction*, *PLOS Computational Biology* **15** (June, 2019) e1006908.

[199] S. Fusi, E. K. Miller, and M. Rigotti, *Why neurons mix: high dimensionality for higher cognition*, *Curr Opin Neurobiol* **37** (Apr., 2016) 66–74.

[200] A. Pouget and L. H. Snyder, *Computational approaches to sensorimotor transformations*, *Nat Neurosci* **3 Suppl** (Nov., 2000) 1192–8.

[201] E. Y. Wang, P. G. Fahey, K. Ponder, Z. Ding, A. Chang, T. Muhammad, S. Patel, Z. Ding, D. Tran, J. Fu, S. Papadopoulos, K. Franke, A. S. Ecker, J. Reimer, X. Pitkow, F. H. Sinz, and A. S. Tolias, *Towards a Foundation Model of the Mouse Visual Cortex*, *bioRxiv* (Apr., 2023) 2023.03.21.533548.

[202] A. Antoniades, Y. Yu, J. Canzano, W. Wang, and S. L. Smith, *Neuroformer: Multimodal and Multitask Generative Pretraining for Brain Data*, Mar., 2024. arXiv:2311.00136 [cs, q-bio].

[203] J. O. Caro, A. H. d. O. Fonseca, C. Averill, S. A. Rizvi, M. Rosati, J. L. Cross, P. Mittal, E. Zappala, D. Levine, R. M. Dhodapkar, I. Han, A. Karbasi, C. G. Abdallah, and D. v. Dijk, *BrainLM: A foundation model for brain activity recordings*, Jan., 2024. Pages: 2023.09.12.557460 Section: New Results.

[204] J. Merel, D. Aldarondo, J. Marshall, Y. Tassa, G. Wayne, and B. Olveczky, *Deep neuroethology of a virtual rodent*, Feb., 2022.