Neural Circuit Dynamics Estimation and Control

Sang Min Han

May 1, 2024

Acknowledgement

This work was partially supported by the National Science Foundation Graduate Research Fellowship.

Neural Circuit Dynamics Estimation and Control

by

Sang Min Han

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering — Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chunlei Liu, Chair
Associate Professor Hillel Adesnik
Professor Kannan Ramchandran

Fall 2021

Neural Circuit Dynamics Estimation and Control

Abstract

Neural Circuit Dynamics Estimation and Control

by

Sang Min Han

Doctor of Philosophy in Engineering — Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Chunlei Liu, Chair

We present a high-throughput, scalable, and biophysically informed effective connectivity mapping method for a population of neurons via two photon holography optogenetics. Specifically, we derive a simple neural circuit dynamics model from basic biophysical principles and introduce a fast algorithm that best estimates the connectivity among the neurons in a network given the observed neural population activity. The algorithm leverages the state-of-the-art two photon holography optogenetic technique and calcium trace imaging as a proxy for neuron membrane potential to optimally estimate the connectivity of neurons residing inside a three-dimensional volume with dimensions spanning hundreds of microns. Using 3D-SHOT, a two photon holography optogenetics technique capable of stimulating custom ensembles of neurons with cellular resolution and millisecond-order time precision, combined with GCaMP6, a contemporary genetically encoded calcium indicator, imaging, we observe the activity of both stimulated and neighboring neurons inside the neocortex of both awake and anesthetized mice in vivo. With these modern technologies at hand, we use the derived deterministic and linear autoregressive model of an arbitrary order with a feed-forward optogenetic stimuli input to describe the population-level time evolution of neural activities in a network. We utilize the ideas in causal inference, compressed sensing, and parallel computing to efficiently estimate the aforementioned model parameters, which directly translate to the connectivity matrices that characterize the effective interactions among the observed neurons. Furthermore, interference framework in experimental design and network control algorithm based on a graph-theoretic centrality measure are applied to provide a higher fidelity summary statistics of the connections among a subset of selected neurons and to artificially drive the network to specific brain states, respectively. With the estimated biophysical model describing the partial dynamics of neuronal interactions, inferences regarding both the spatial and temporal signatures of a local region of the brain can be made.

To my family

Contents

List of Figures

Acknowledgments

Over the past 6.5 years that I have been at UC Berkeley to complete my PhD, I met a lot of wonderful people that I must thank. First and foremost, I could not have made it this far without the help of everyone in the Chunlei Liu Lab: Miriam Hernández-Morales, Koyam Morales, Hongjiang Wei, Yuyao Zhang, Jingjia Chen, Victor Han, Zoe Cohen, Maruf Ahmed, and Shafeeq Ibraheem. I also owe thanks to the wonderful undergraduates Raymond Wang, Soobin Hong, Amy Wang, Jiwoong Han, and Cynthia Sor for working with me, teaching me, and helping me develop as a mentor. I must also express my gratitude toward the members of the Hillel Adesnik Lab for welcoming me and patiently teaching me neuroscience, optics, and biology. I would like to thank Lamiae Abdeladim, Janine Beyer, Jenny Brown, Nikhil Bhatla, Hayley Bounds, Yun Chiu, Marta Gajowa, Karthika Gopakumar, Will Hendricks, Uday Jagadisan, Alan Mardinly, Dan Mossing, Mora Ogando, Ian Oldenburg, Nicolas Pégard, Daniel Quintana, Masato Sadahiro, Hyeyoung Shin, Savitha Sridharan, Greg Telian, Silvio Temprana, and Yi Xue. I also thank the people in Miki Lustig's group, especially Jon Tamir, Alan Dong, and Ke Wang, for managing the shared computing resources that were vital to my work.

Tom Courtade welcomed me into UC Berkeley and helped me transition from being a college student to an independent graduate student researcher. As my advisor for the first 2.5 years of graduate school, Tom taught me the skills and knowledge that I use even today. Tom's amazing brain power and gifted problem solving skills have always been a source of inspiration to me. Thank you, Tom!

Sam Pimentel served on my qualifying exam committee, and he was instrumental in helping me develop the experimental design and causal inference formulation of the work detailed in this thesis. Sam's statistical perspective and careful approach to research problems helped me keep my work on the right track.

Kannan Ramchandran has consistently supported me throughout the course of my graduate school career. Kannan has served on both of my preliminary exam committees as well as my qualifying exam committee. Kannan continued to provide guidance and help as an essential member of my dissertation committee. Seeing his enthusiasm for solving problems and his broad interest in research has been very encouraging.

My collaboration, interactions, and discussions with Ian Oldenburg really defined my work in graduate school. Ian's extensive knowledge and expertise in science continues to amaze me to this day, and his joyful and lively attitude toward research is something I hope to emulate. Ian has been my teacher, my mentor, and the person who really made our work together come to fruition. It is definitely not an exaggeration to say that my PhD could not have been made possible without Ian's contributions.

Although not officially listed on paper, I have always considered Hillel Adesnik as my co-advisor. Hillel welcomed me into his lab, and provided me with the perfect environment for me to develop as a researcher in computational and systems neuroscience. His intelligence, comprehensive knowledge in a wide variety of disciplines, and passion for science and engineering never cease to impress and inspire me.

Above all, I must thank my advisor, Chunlei Liu. Chunlei saved not only my graduate school career, but also my overall career in research. He inspired and motivated me to continue in my PhD journey under his guidance, and gave me the perfect combination of independence and supervision throughout my time in Berkeley. He really brought out the best in me, and made graduate school a fulfilling and rewarding experience. I could always rely on and count on Chunlei's help in times of need, and I thank Chunlei for all his support and guidance during times of hardship. I am also grateful to Chunlei for giving me the opportunity and the freedom to explore and learn. It was truly a privilege and an honor to have had the opportunity to learn from his brilliance and wisdom. I consider Chunlei as my role model in academia, and I look forward to continuing to work with him in the future. Thank you, Chunlei!

In addition to learning how to conduct research, I learned how to manage courses and to teach effectively from excellent educators. I thank Tom Courtade for having me as the Graduate Student Instructor for EE 120: Signals and Systems, as well as Vladmir Stojanovic and Chunlei Liu for having me as the Graduate Student Instructor for EE 16A: Designing Information Devices and Systems I. I was also very fortunate to take courses taught by Tom Courtade, Venkat Anantharam, Michael Jordan, Laura Waller, Johannes Royset, Laurent El Ghaoui, Martin Wainwright, Chunlei Liu, Miki Lustig, Sam Pimentel, Eric Betzig, Na Ji, and Bruno Olshausen.

I must also thank Sangjae Bae and Scott Moura for guiding me through my first collaboration efforts. I am also indebted to Shirley Salanio for always being available at times of need, and for convincing me to stay and finish my PhD.

I also need to acknowledge the members of the Korean EECS community: Jaeyoung Choi, Jichan Chung, Seobin Jung, Edward Kim, Hokeun Kim, Hyungjin Kim, Jinkyu Kim, Kangwook Lee, Kyoungtae Lee, Suhong Moon, and Dongjin Seo; and the BLISS community: Ashwin Pananjady, Kangwook Lee, Nihar Shah, Rashmi Vinayak, Vasuki Narasimha Swamy, Fanny Yang, Yuting Wei, Reza Abbasi-Asl, Payam Delgosha, Dong Yin, Orhan Ocal, Vidya Muthukumar, Raaz Dwivedi, Soham Phade, Billy Fang, Joseph Borja, Avishek Ghosh, Vipul Gupta, Laura Brink, Aviva Bulow, Koulik Khamaru, Efe Aras, Kuan-Yun Lee, Jichan Chung, Kabir Chandrasekher, Leah Dickstein, Tavor Baharav, Simon Li, Reinhard Heckel, and Ilan Shomorony. I am indebted to Raaz Dwivedi, Soham Phade, and David Fridovich-Keil for helping me survive Berkeley as well as for the great memories. I am also grateful to Kyoungtae Lee, Raejoon Jung, Donggun Lee, Saehong Park, and Sangjae Bae for the much needed laughs and fun. Thanks also to Amy Goh and Wonjung Yun for helping me take breaks from work and for reminding me to enjoy life.

I especially thank Soyeon Kim for cheering me on, encouraging me, and for being the source of my happiness. Most importantly, I am forever thankful to my family for their unconditional love and support. No words can express the amount of appreciation and gratitude I have for my parents Joon Soo Han and Jin Sook Han, my brother Sang Wook Han, my sister-in-law Joo Lee, and my nephew Leo Han.

Chapter 1

Introduction

In this dissertation, we present an explainable artificial intelligence-based method to decode population-level neural activity. Rather than relying on black box architecture-based methods, we ground our method on a biophysically informed model. Using calcium trace imaging data of neural population activity obtained with direct photostimulations using two photon holography optogenetics in both awake and anesthetized mice, we optimize our model parameters. The fitted model is then used to provide biological insights and to yield intuition into how the investigated neural circuit can be modulated to achieve specific network-level characteristics and dynamics.

1.1 Objective

The overarching objective of the work presented in this dissertation is to utilize the newly developed computational method that combines imaging, modeling, and modulating the dynamics of neural circuits with extremely high spatiotemporal precision to understand and to ultimately understand perception and behavior. The long-term goal of the project detailed in this thesis is to achieve fundamental new understandings of neural circuit dynamics that may have significant science and health implications. To this end, we demonstrate that we can compute the aforementioned model parameters in a fast and accurate manner, and propose a scheme that uses the learned biophysical model for closed-loop neural stimulations. Figure 1.1 illustrates our method pipeline as well as our overall objective of achieving closed-loop neural modulations in awake, behaving animals.

This dissertation outlines the derivations that lead to an autoregressive equation that defines the biophysical model, which characterizes and predicts neural circuit dynamics. It also details the use of optogenetic stimulation and the optimal spatiotemporal stimulation patterns for enabling online and closed-loop circuit estimation and modulation. The custom developed estimation algorithm that employs key ideas in compressed sensing is further described. We validate the circuit properties and dynamics of the learned biophysical autoregressive model and propose an optimized strategy to efficiently modulate circuit dynamics

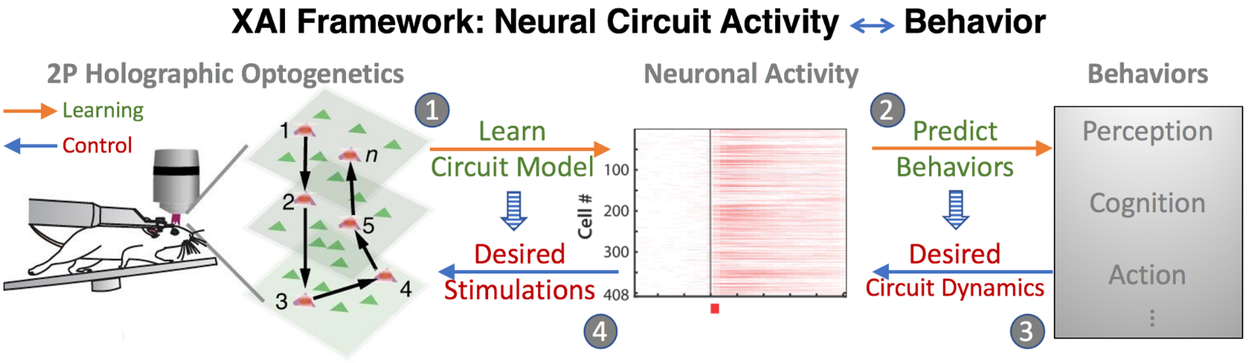

Figure 1.1: Explainable artificial intelligence framework to decode and modulate population-level neural activity in awake, behaving animals.

with optogenetic stimulation.

1.2 Background

Recent technical advances have revolutionized our ability to observe and manipulate neurons at a large scale and to modulate behaviors [14, 32, 3]. Simultaneous two photon (2P) optogenetic stimulation and calcium imaging have enabled rapid, reversible control and recording of genetically and spatially defined neurons [27, 31, 33, 30]. Targeted activation of neuronal ensembles in the neocortex may be sufficient to elicit percepts and possibly influence behavioral outcomes [28, 12]. Although these findings demonstrate the power of precise optogenetic manipulations, a more generalizable and unsupervised approach to design precise optogenetic perturbations that will optimally provide insight into the underlying neural mechanisms of neural coding is necessary. Using the state-of-the-art two photon holography optogenetic technique, we are able to stimulate tens of neurons together at the same time. However, as the number of neurons that can be simultaneously targeted grows, the number of possible perturbations will increase combinatorially. With this technological advance forthcoming, selecting the most informative ensembles of cells, given limited time, is essential. Straightforward applications of machine learning have resulted in "black box" correlations between neural activity and behaviors, producing little mechanistic understanding of the underlying neural circuit [35, 17, 19, 45]. By combining functional network mapping via holographic optogenetics with innovative autoregressive (AR) modeling, we developed an explainable artificial intelligence (XAI) approach that can reveal the governing equations underlying the dynamics of a network comprised of a large number of neurons. With further inspection, we anticipate that the novel approach described in this dissertation will provide deep mechanistic insight into how population activity in the cortex drives perceptual behaviors and cognition that cannot be obtained via other means.

We hypothesized that the dynamics of a neural circuit can be modeled by an augmented autoregressive equation that can inform the precise types of perturbations needed to informatively model and modulate circuit activities linked to cortical function. Our approach to test this hypothesis hinges on two innovations: three-dimensional holographic optogenetic circuit mapping and a physical circuit model powered by machine learning algorithms. Experimentally, we leverage an all-optical circuit interrogation technique named *Three-dimensional scanless holographic optogenetics with temporal focusing* (3D-SHOT) [34]. 3D-SHOT allows precise, simultaneous photo-activation of sets of neurons at near single-neuron spatial resolution and millisecond time precision. By holographically activating user-defined ensembles of neurons in vivo, we obtain the data needed to train the autoregressive biophysical model that can then predict the optimal neural circuit perturbations needed to drive robust effects on behavior, which can provide insight into the key aspects of the population codes that drive perception and behavior. Capitalizing on the more recent innovations in both the optical aspects of 3D-SHOT and the development of substantially more potent microbial opsins, we interrogated close to one thousand neurons within the primary visual cortex (V1) during each experiment session. It is our hope to simultaneously study 100,000 neurons across V1 and higher cortical areas in the future using a newly developed holographic mesoscope. The number of unknowns in a circuit connectivity matrix scales with the square of the number of neurons. To address this computational complexity, we probe the circuit by co-stimulating large, spatially distributed random sets of neurons, and use compressed sensing, an artificial intelligence algorithm, to learn what we term effective connectivity among all the neurons in the field-of-view (FOV) of the microscope. The algorithm is fast, which allows online and closed-loop modeling and modulation of the neural circuits.

1.3 Innovation

Our novel explainable artificial intelligence approach is devised based on known biology and physics. Prior methods predominantly relied on the use of "black box" machine learning techniques that fit neural recordings data from specific brain regions directly to specific behaviors. Compared to these approaches that cannot be generalized to other brain regions or to new data, we tackle the problem of decoding population neural activity in two steps: 1) learn a generalizable circuit model; 2) use the learned circuit properties to explain behaviors. We also employ compressed sensing in the interrogation of neural circuit functional connectivity. Prior works have conducted limited perturbation of only few neurons at a time to estimate circuit dynamics, or relied on observations without targeted stimulation [42, 43]. We take a novel systems approach to interrogate the circuit with much larger scale photo-stimulation and compressed sensing [16], which capitalizes on the sparse nature of the circuit connections. Further novelty comes from the application of three-dimensional holographic photostimulation to train predictive circuit models. The precise targeted ensemble stimulation allows statistical estimation of model parameters rather than inference, which does not scale well in computational complexity and requires probabilistic assumptions and priors. Past works related to functional mapping have stimulated one neuron at a time. We stimulated tens of neurons at a time to increase throughput and obtain more appropriate training data for large-scale circuit perturbations-based mapping. Our approach stimulates, observes, and computes in the most effective and efficient manner to date.

Chapter 2

Approach

Our objective is to develop a data-driven explainable artificial intelligence-based model that both provides insights about the inner working of the circuit and generates useful predictions for perturbing it to probe downstream dynamics and behavior. To achieve this goal, we take an approach that involves three steps. First, we obtain training data sets by mapping effective connections within a large population of cortical neurons with 3D-SHOT, photo-activating ensembles of approximately 30 neurons per time step while monitoring the resulting impacts on all the other neurons through dense, volumetric calcium imaging. Second, this data is fed into an estimation algorithm that computes the model parameters, which is the set of detailed connectivity matrices characterizing the effective functional pairwise influence weights among all probed neurons. Third, the biophysically inspired circuit model determined by these matrices are used to identify further photo-activation patterns that can drive local dynamics that may optimally drive higher cortical areas and modify behavior. Training a full biophysical Hodgkin-Huxley model of the entire network is not feasible due to limitations of current technology, but we still aim for a model that treats each neuron and its connections in a physiological manner. This plan of attack facilitates gaining insight into the neural circuit's underlying mechanisms. The approach we take overcomes the above limitations by modeling the current-voltage propagation relationship with a compound transconductance function and using calcium imaging as a surrogate for action potentials in the training data. Algorithms for both offline and online estimation have been developed, and we demonstrate that compared to conventional correlation analysis, our approach significantly improves the predictive power of the data-driven model, which informs the precise perturbations that yield insight into the computations underlying perception and behavior.

The aforementioned algorithms compute the autoregressive circuit parameters using a sparsity prior. The estimation algorithm is high-throughput, scalable, and learns the neural circuit parameters of an ensemble of neurons in the optogenetics setting using the data collected from the circuit interrogation scheme outlined above. We leverage the ideas in compressed sensing and parallel computing to efficiently estimate the parameters of the autoregressive biophysical model. The model parameters directly translate to the effective connectivity matrices that characterize the effective connections. These connectivity matri-

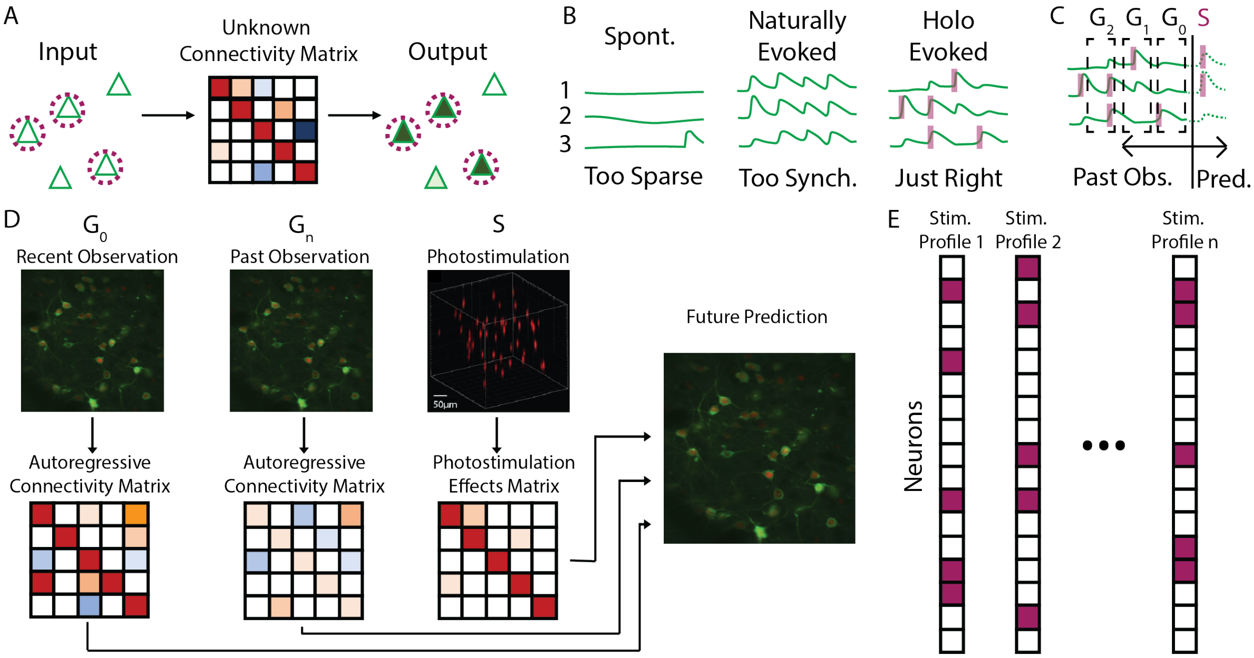

Figure 2.1: Autoregressive biophysical model determined using compressed sensing with multiphoton optogenetics. A) Connectivity parameter estimation: Statistical estimation of connectivity made possible by custom ensemble stimulation of opsin expressing cells and calcium imaging. B) Neural perturbation: Connectivity between cells cannot be efficiently learned via spontaneous or naturally evoked activity, as they are too sparse or too synchronized, respectively. Instead, many custom precise stimulation patterns generated by multiphoton holographic optogenetics are needed. C-D) AR model: We use a sparse autoregressive model that takes into account connectivity dynamics from past observations $G_0, \ldots, G_{p-1}$, as well as the current photostimulation effects over time $S_0, \ldots, S_{q-1}$ to predict future neuronal activities. E) Random stimulations: Random pattern stimulation profile inputs are transformed by the sparse connectivity matrices to explain observed responses.

ces capture effects of both functional and true physical synaptic connectivity that might be mono- or poly-synaptic. We derive the mathematical formulations required to transform the problem of estimating the unknown model parameters to solving a slightly modified version of the least absolute shrinkage and selection operator (LASSO) [44], a convex optimization problem, which is the hallmark of compressed sensing. The LASSO problem is solved using a state-of-the-art variant of the proximal gradient descent algorithm termed the Greedy Fast Iterative Shrinkage-Thresholding Algorithm (Greedy FISTA) [2, 25]. The appropriate sparsity level in the recovered solution is determined from the validation data set using cross validation.

We check the accuracy of our algorithm by conducting in silico computer simulations.

The primary advantage of such simulations is that the circuit connectivity is entirely known a priori. After constructing a biophysically meaningful ground truth connectivity matrices with sparsity levels and connection strengths that match what is known in neuroscience, we demonstrate the accuracy of our effective connectivity mapping approach on simulated photo-stimulation training data.

2.1 3D-SHOT Optimization

We further optimize 3D-SHOT photo-stimulation for acquiring training data for our estimation routine. 3D-SHOT allows the user to photo-stimulate a custom ensemble of neurons optogenetically with near cellular resolution and millisecond precision [27, 34]. It thus serves the need for acquiring the relevant training data for the compressed sensing estimation algorithm. We update and tailor the specific 3D-SHOT photo-stimulation and 2P imaging settings to acquire the best possible training data for our problem formulation. As shown in Figure 2.2 showing the 3D-SHOT setup, a spatial light modulator (SLM) diffracts the excitation laser into a hologram comprised of a set of neuron-sized, temporally focused light spots where each spot targets one of a subset of neurons for optogenetic stimulation. Combined with volumetric calcium imaging, this two photon read and write platform allows one to activate arbitrarily chosen sets of neurons while simultaneously monitoring the impact on all other neurons within the FOV. Exploiting the recent optical advances, the updated 3D-SHOT system both images and photo-stimulates across a brain volume approximately $800\mu\text{m} \times 800\mu\text{m} \times 100\mu\text{m}$. In the mouse primary visual cortex (V1), this volume encompasses a substantial fraction of Layer 2/3 (L2/3). Within this volume, the system records the activity of each of approximately 1,000 neurons using genetically encoded calcium indicator GCaMP6s whose relative fluorescence changes report firing changes approximately linearly [10, 48]. With recent developments of even more potent opsins for 2P optogenetics, named ChroME2.0, we selected about 30 neurons to stimulate at a time [41]. The optimized setup can excite a different group of neurons at each refresh of the spatial light modulator, which has maximum rate of 300 Hz. The calcium dynamics of nearly every GCaMP-expressing neuron in the imaging volume are sampled with fast remote focusing via a second SLM. Three planes are sampled at approximately 6 Hz imaging rate. There is an inherent tradeoff between the number of neurons the system can sample and the sampling rate. We deliberately choose to acquire data from many imaging planes (many neurons) at lower frame rates for two reasons: first, the AR model will perform best when most of the potential inputs to each neuron are probed; second, the slow kinetics of calcium transients limits the amount of added information obtained from higher sampling frequency.

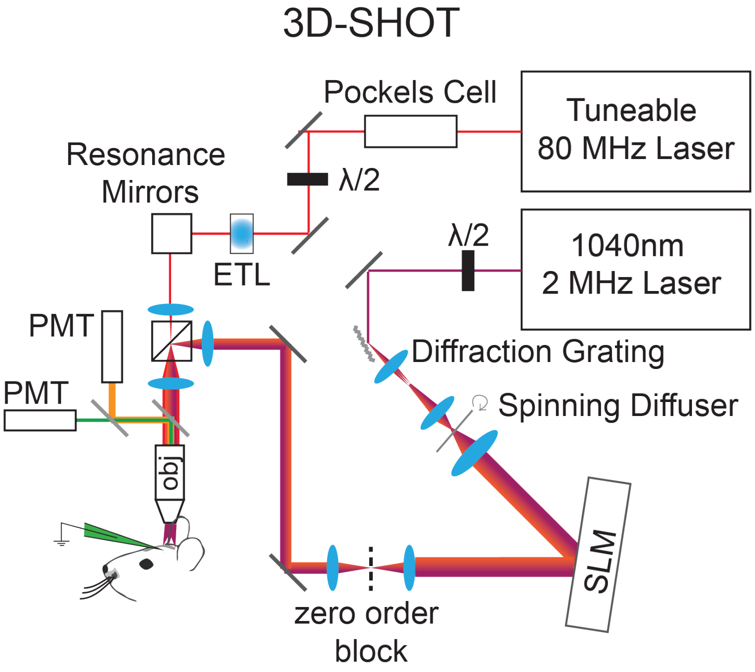

Figure 2.2: Schematic diagram of the 3D-SHOT rig setup.

2.2 Off-Target Effects

An important point to consider is the effective photo-stimulation resolution. Although the resolution is on the order of the size of a single cortical pyramidal neuron, we take several approaches to minimize contamination from “off-target” photo-excited neurons as they may corrupt the model estimation procedure. First, we use recently developed high potency opsins (ChroME2.0) that allow us to drive neurons with light levels that are far from saturating the opsin, maximizing the resolution benefit from 2P excitation [9, 41]. Second, we stimulate each opsin-positive cell with the minimum laser power needed to produce reliable photo-activation, minimizing activation of nearby neurons as exact targeting is never perfect. Third, in cases where photo-excitation of ‘off-target’ cells is still unavoidable (as measured through the volumetric GCaMP6s imaging) will be excluded from analysis. These data demonstrate that our system is capable of activating large user-defined sets of neurons in the imaging volume, a critical technical requirement for training the AR model.

Chapter 3

Mathematical Formulation

Recall the optimized 3D-SHOT setting where a custom ensemble of neurons in the neocortex region of the brain can be stimulated optically using laser light with near cellular spatial resolution and millisecond-order time precision. We consider a region of interest [1] comprised of three cross sections of 800 μm $\times$ 800 μm in size that cover a depth of approximately 100μm in the cortex. Within this field-of-view, each of approximately 1000 neurons' membrane potentials are recorded using GCaMP6s, a calcium trace indicator outputting relative fluorescence changes that translate to a cell's action potential event. The brighter the relative GCaMP fluorescence signature, the higher the membrane potential of the neuron. Using 3D-SHOT, we target at most 50 neurons simultaneously. In 3D-SHOT, each hologram defines one stimulation profile, which is comprised of a subset of neurons selected as targets for optical stimulation. The turnaround time for switching from one stimulation profile to another is only a couple of milliseconds. Nevertheless, up to 10 seconds may be needed to prepare and compile a new hologram if a custom ensemble of neurons needs to be stimulated online based on immediate feedback. In other words, millisecond-order turnaround time for stimulation is possible only with pre-determined and pre-prepared ensembles. With these physical spatiotemporal settings in mind, we formulate our problem mathematically using definitions and techniques often used in experimental design and causal inference.

3.1 Mathematical Formulation of Experimental Design

The experimental design framework fits naturally in the optogenetics setting. The mathematical formulation of the effective connectivity estimation in optogenetics investigated here will follow the notations commonly used in the experimental design literature.

Consider a finite population U of neurons indexed by $n = 1, \dots, N$, where $|U| = N$. Denote by m the discrete time index corresponding to one imaging sample. Let the treatment

[1] the field-of-view is a trapezoid due to the physics of the optogenetics setup

assignment $d_n \in \{0,1\}$ represent the stimulation assignment corresponding to neuron n such that $d_n = 1$ indicates that neuron n is targeted for optogenetic stimulation. We assume that there is perfect compliance. This is a valid assumption as we excite a neuron with a set of approximately 10 pulses at 30 Hz to ensure stimulation with very high probability. Let the potential outcome y_n for neuron n be the high-pass-filtered fluorescence readout of GCaMP–the calcium trace indicator–relative to the baseline level, $\frac{\Delta F}{F_0}$, of neuron n. High-pass filtering removes the undesirable low-frequency baseline shifts in the recorded, raw $\frac{\Delta F}{F_0}$ signals. For brevity and clarity, $\frac{\Delta F}{F_0}$ will refer to the high-pass filtered $\frac{\Delta F}{F_0}$ data. Define

$$y_n \coloneqq \left(\frac{\Delta F}{F_0}\right)_n. \tag{3.1}$$

Then, under randomization-based treatment assignment, the potential outcome for neuron n can be written:

$$y_n = d_n y_n^{(1)} + (1-d_n) y_n^{(0)}, \tag{3.2}$$

where $y_n^{(1)}$ and $y_n^{(0)}$ are the potential outcomes, the observed $\frac{\Delta F}{F_0}$ trace, of unit n under treatment and control, respectively. In the context of optogenetics, $y_n^{(1)}$, the outcome under treatment, and $y_n^{(0)}$, the outcome under control, varies over time for every neuron n due to noise (spontaneous activity), interference, and the discrete nature of sampling. Treatment can only be assigned to neurons expressing opsins, which are light-sensitive proteins that enable optical stimulation. We restrict treatment assignments to the support set of opsin-positive neurons:

$$\Omega \coloneqq \{n : \text{Prob}(d_n) > 0\}, \tag{3.3}$$

i.e., the set of neurons that can be optically stimulated.

We assume the constant additive treatment effect model for optogenetics:

$$y_n^{(1)} = y_n^{(0)} + \tau_n \ \forall \ n = 1, \ldots, N, \tag{3.4}$$

and presume that the optogenetic stimulus increases the $\frac{\Delta F}{F_0}$ of neuron n in the support Ω by a constant amount of τ_n. The constant additive treatment effect model above defines a superpopulation model, where one potential outcome at time steps m for each neuron n is regarded as a sample drawn from one of two distributions: one distribution corresponding to the stimulated state and another corresponding to the resting state of the neuron. Given this superpopulation model, we can quantify the effects of the treatment, optogenetic stimulation, on the neurons' observed $\frac{\Delta F}{F_0}$, or y_n. Due to the nature of biology, e.g. discrepancies in the GCaMP expression levels and nonuniform transfection rate of the opsins in the neurons, different neurons react differently to the optogenetic stimulation. Therefore, treatment effect is not homogeneous across all neurons, and there exists treatment effect heterogeneity. We define the Individual Treatment Effect (ITE) for every neuron n as

$$\text{ITE}_n \coloneqq \tau_n = y_n^{(1)} - y_n^{(0)}. \tag{3.5}$$

Fundamental Problem of Causal Inference

In the optogenetics setting, we can circumvent the fundamental problem of causal inference, which states that it is impossible to observe multiple treatment effects or potential outcomes under both treatment and control for one unit. For every unit neuron n, we can observe both the baseline state of a neuron as well as its potential outcome under treatment with the assumption that reasonable stationarity conditions hold, i.e. plasticity does not develop, because a neuron's membrane potential always returns to its resting state after an action potential event. Therefore, if an experiment is performed within a reasonable amount of time (approximately 2 hours) such that plasticity or other significant change in the brain does not occur to a degree that alters the circuit, we can disregard the fundamental problem of causal inference. Then, estimating the ITE for neuron n, i.e. τ_n, for every neuron n is straightforward.

3.2 Potential Treatment Outcome Framework

We consider a superpopulation model, where potential outcomes are regarded as samples drawn from two distributions: one distribution corresponding to the stimulated state and another corresponding to the resting state. Given this superpopulation model, we estimate the Individual Treatment Effect (ITE) for every neuron in the context of treatment effect heterogeneity, and propose a scheme for calculating point estimates of the interference estimands of interest in the subsequent sections. To estimate this superpopulation model as well as the ITE for every neuron n, we fit a two-Gaussian mixture model on the high-pass filtered $\frac{\Delta F}{F_0}$ histogram data. These estimates are then used to decide whether each neuron's $\frac{\Delta F}{F_0}$ at every time step m corresponds to the stimulated state or the baseline state. In addition, the estimates are used to standardize and normalize the $\frac{\Delta F}{F_0}$ data, which is essential for estimating population-level relationships among all neurons without systematic biases.

Suppose that the potential outcomes $\frac{\Delta F}{F_0}$ for each neuron n are realizations of a parametric model: a Gaussian mixture model with two mixture components: $\mathcal{N}(\mu_{n0}, \sigma_{n0}^2)$ for the baseline and $\mathcal{N}(\mu_{n1}, \sigma_{n1}^2)$ for the excited state. For every neuron n, denote by $f_n(x)$ the approximate probability density function of the high-pass-filtered $\frac{\Delta F}{F_0}$ histogram data. Then,

$$f_n(x) = w_0 f_{\mathcal{N}(\mu_{n0},\sigma_{n0}^2)}(x) + w_1 f_{\mathcal{N}(\mu_{n1},\sigma_{n1}^2)}(x). \tag{3.6}$$

We subsequently have an approximate superpopulation model for each neuron n, and have the following estimate of the ITE:

$$\widehat{\text{ITE}}_n = \mu_{n1} - \mu_{n0}. \tag{3.7}$$

Because every neuron n reacts differently to treatment, the optogenetic stimulation, we quantify the treatment effect heterogeneity using the ITE estimate, and leverage this ITE to normalize and standardize the data for population-level analysis. We also use this ITE to determine whether each observation $\frac{\Delta F}{F_0}$ for every neuron n is sampled from the baseline distribution or the excited distribution. Using the estimated Gaussian mixture model, we have the following hypotheses to test for each neuron n at every time step m:

$$\mathcal{H}_0 : y_n \sim \mathcal{N}(\mu_{n0}, \sigma_{n0}^2), \tag{3.8}$$

$$\mathcal{H}_1 : y_n \sim \mathcal{N}(\mu_{n1}, \sigma_{n1}^2). \tag{3.9}$$

We test for

$$\Lambda\left(y_n\right) = \frac{w_1 f_{Y_n=y_n|\mathcal{N}(\mu_{n1},\sigma_{n1}^2)}\left(y_n\right)}{w_0 f_{Y_n=y_n|\mathcal{N}(\mu_{n0},\sigma_{n0}^2)}\left(y_n\right)} = \frac{\frac{w_1}{\sqrt{2\pi}\sigma_{n1}} \exp\left\{\frac{-(y_n-\mu_{n1})^2}{2\sigma_{n1}^2}\right\}}{\frac{w_0}{\sqrt{2\pi}\sigma_{n0}} \exp\left\{\frac{-(y_n-\mu_{n0})^2}{2\sigma_{n0}^2}\right\}} \tag{3.10}$$

$$= \frac{w_1\sigma_{n0}}{w_0\sigma_{n1}} \exp\left\{\frac{1}{2}\left[\left(\frac{y_n-\mu_{n0}}{\sigma_{n0}}\right)^2 - \left(\frac{y_n-\mu_{n1}}{\sigma_{n1}}\right)^2\right]\right\} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 1. \tag{3.11}$$

Taking the log-likelihood ratio, we can equivalently test for

$$\left(\frac{y_n-\mu_{n0}}{\sigma_{n0}}\right)^2 - \left(\frac{y_n-\mu_{n1}}{\sigma_{n1}}\right)^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\log\left(\frac{w_0\sigma_{n1}}{w_1\sigma_{n0}}\right). \tag{3.12}$$

Simplifying, we obtain

$$\frac{y_n^2\sigma_{n1}^2 - 2y_n\mu_{n0}\sigma_{n1}^2}{\sigma_{n0}^1\sigma_{n1}^2} - \frac{y_n^2\sigma_{n0}^2 - 2y_n\mu_{n1}\sigma_{n0}^2}{\sigma_{n0}^2\sigma_{n1}^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 2\log\left(\frac{w_0\sigma_{n1}}{w_1\sigma_{n0}}\right) + \frac{\mu_{n1}^2}{\sigma_{n1}^2} - \frac{\mu_{n0}^2}{\sigma_{n0}^2}, \tag{3.13}$$

where the optimal threshold for determining if y_n is a sample from the baseline distribution or the stimulated distribution becomes a function of the solutions to the following quadratic inequalities:

$$c_1 y_n^2 + c_2 y_n + c_3 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} 0, \tag{3.14}$$

where

$$c_1 \coloneqq \sigma_{n1}^2 - \sigma_{n0}^2, \tag{3.15}$$

$$c_2 \coloneqq 2\left(\mu_{n1}\sigma_{n0}^2 - \mu_{n0}\sigma_{n1}^2\right), \tag{3.16}$$

$$c_3 \coloneqq \mu_{n0}^2\sigma_{n1}^2 - \mu_{n1}^2\sigma_{n0}^2 - 2\sigma_{n0}^2\sigma_{n1}^2\log\left(\frac{w_0\sigma_{n1}}{w_1\sigma_{n0}}\right). \tag{3.17}$$

The two critical y_n^*'s are then:

$$y_n^* = \frac{\mu_{n1}\sigma_{n0}^2 - \mu_{n0}\sigma_{n1}^2 \pm \sqrt{\sigma_{n0}^2\sigma_{n1}^2\left[\left(\mu_{n0}-\mu_{n1}\right)^2 - 2\left(\sigma_{n0}^2 - \sigma_{n1}^2\right)\log\left(\frac{w_0\sigma_{n1}}{w_1\sigma_{n0}}\right)\right]}}{\sigma_{n0}^2 - \sigma_{n1}^2}. \tag{3.18}$$

Algorithm 1: Threshold γ calculation algorithm

```
Input  : Observation $y$ and the parameters of the fitted Gaussian mixture models:
         $w_0, w_1, \mu_0, \mu_1, \sigma_0, \sigma_1$
Output: Threshold $\gamma$ for optimally deciding between $\mathcal{H}_1$ and $\mathcal{H}_0$.
Initialize: Calculate
    $\gamma^*_{n1} = \frac{\mu_{n1}\sigma^2_{n0} - \mu_{n0}\sigma^2_{n1} - \sqrt{\sigma^2_{n0}\sigma^2_{n1}\left[\widehat{\mathrm{ITE}}^2_n - 2(\sigma^2_{n0} - \sigma^2_{n1})\log\left(\frac{w_{n0}\sigma_{n1}}{w_{n1}\sigma_{n0}}\right)\right]}}{\sigma^2_{n0} - \sigma^2_{n1}}$,
    $\gamma^*_{n2} = \frac{\mu_{n1}\sigma^2_{n0} - \mu_{n0}\sigma^2_{n1} + \sqrt{\sigma^2_{n0}\sigma^2_{n1}\left[\widehat{\mathrm{ITE}}^2_n - 2(\sigma^2_{n0} - \sigma^2_{n1})\log\left(\frac{w_{n0}\sigma_{n1}}{w_{n1}\sigma_{n0}}\right)\right]}}{\sigma^2_{n0} - \sigma^2_{n1}}$.
if $\sigma^2_1 \geq \sigma^2_0$ then
    $\gamma = \gamma^*_2$.
else
    $\gamma = \gamma^*_1$.
end
Return $\gamma$.
```

Denote by γ^*_{n1} and γ^*_{n2} the two critical y^*_n such that $\gamma^*_{n1} \leq \gamma^*_{n2}$. Then, for every y_n, we have Algorithm 1 for calculating the threshold γ_n for optimally deciding between $\mathcal{H}_1$ and $\mathcal{H}_0$ according to the Neyman-Pearson lemma.

This threshold is a function of

$$\gamma^*_{n\pm} = \frac{\mu_{n1}\sigma^2_{n0} - \mu_{n0}\sigma^2_{n1} \pm \sqrt{\sigma^2_{n0}\sigma^2_{n1}\left[\widehat{\mathrm{ITE}}^2_n - 2\left(\sigma^2_{n0} - \sigma^2_{n1}\right)\log\left(\frac{w_0\sigma_{n1}}{w_1\sigma_{n0}}\right)\right]}}{\sigma^2_{n0} - \sigma^2_{n1}}, \tag{3.19}$$

where $\gamma_{n1} = \gamma_{n-}$ and $\gamma_{n2} = \gamma_{n+}$. If $\sigma^2_{n1} \geq \sigma^2_{n0}$, the parabola specified by the quadratic inequality opens up, and γ^*_+ is the optimal threshold for deciding between $\mathcal{H}_0$ and $\mathcal{H}_1$. Otherwise, the parabola opens down, and γ_{n1} is the optimal threshold. In conclusion, the optimal threshold is

$$\begin{aligned}\gamma_n &= \mathbb{I}\left(\sigma^2_{n1} \geq \sigma^2_{n0}\right)\gamma^*_{n+} + \left[1 - \mathbb{I}\left(\sigma^2_{n1} \geq \sigma^2_{n0}\right)\right]\gamma^*_{n-} && (3.20)\\ &= \begin{cases}\gamma_{n+}, & \text{if } \sigma^2_{n1} \geq \sigma_{n0}\\ \gamma_{n-}, & \text{otherwise}\end{cases}, && (3.21)\end{aligned}$$

where $\mathbb{I}(\cdot)$ is the indicator function. Note that the critical y^*_n, and thus γ_n, is a function of the ITE for every neuron n. The optimal threshold γ_n is used to compute the binary activation state of the recorded $\frac{\Delta F}{F_0}$ to standardize the observations across all neurons in the field-of-view.

Chapter 4

Biophysics Model

The interpretability of AI algorithms can be achieved with either model-based machine learning and/or post-hoc analysis. Our approach integrates both methods. We decouple the circuit-behavior relationship into two separate problems: first, understanding the mechanisms underlying circuit dynamics; and second, relating the local circuit dynamics to downstream cortical area activation and behavior. In our model, circuit dynamics are governed by neuronal and synaptic biophysics that are modeled by a set of autoregressive processes, and thus are generalizable to other circuits and brain regions. The relationship between circuit dynamics and behavior is to be learned from the impact of closed-loop model-designed optical perturbations. The model is autoregressive because it relates current-time neuronal activity with past activity and stimulation. The parameters of the autoregressive model define a set of dynamic connectivity matrices. Existing approaches for neural modulation attempt to record neural activity patterns associated with specific behavior and "replay" the activity patterns in the same group of neurons in order to elicit the same behavior. The aforementioned approach completely neglects the circuit dynamics. A network of neurons may not exhibit stationarity when presented with the replayed patterns due to inherent network dynamics. By modeling, predicting, and empirically probing the circuit dynamics, our approach has the potential to provide more potent and predictable modulations of behavior.

4.1 Vector Notation for Neural Population

Denote by $d[m] \in \{0,1\}^N$ and $y[m] \in \mathbb{R}^N$ the N-dimensional vectors with the n^{th} element equivalent to $d_n[m]$ and $y_n[m]$, respectively. We refer the the vector $d[m]$ as the stimulation profile at time m and the vector $y[m]$ as the response of the network at time m. We obtain the input-output pairs $(d_n[m], y_n[m])$ from optical stimulation and calcium imaging of a three-dimensional field-of-view of the brain at time steps indexed by m for every neuron n.

Effective Connectivity Matrix

The overarching objective is to establish population-level signatures associated to certain brain states, which may be directly linked to behavior. We would like to establish how neurons in one region of the brain interact with each other. Let $G \in [-1,1]^{N\times N}$ be the effective connectivity matrix of the neural network under observation, where $[-1,1]^{N\times N}$ denotes the set of N-dimensional square matrices with the element $g_{n\ell}$ (located at the n^{th} row and ℓ^{th} column) that is less than 1 in magnitude for every $n, \ell \in \{1,\ldots,N\}$. Edges with positive weights correspond to neural links that are excitatory, and edges with negative weights correspond to neural links that are inhibitory. We follow the convention where $g_{n\ell}$ specifies how neuron ℓ influences neuron n. Suppose that G does not change with respect to time, i.e., there does not exist plasticity. In order to assure this assumption, we must estimate G within the time frame of approximately 2 hours. This effective connectivity matrix G represents a graph with N nodes, with each node representing a neuron.

Deficiency of the Simple Linear Model

The goal here is to find the effective connectivity matrix G with the constraints that the elements $g_{n\ell}$ satisfy $|g_{n\ell}| \leq 1$ for all $n, \ell \in \{1,\ldots,N\}$. Positive weights correspond to excitatory connections and negative weights correspond to inhibitory connections. We define the effective connectivity to be the matrix G that specifies the linear relationship between the optogenetic stimulation input $d[m]$ and the GCaMP calcium trace level output $y[m]$ at each discrete imaging step m:

$$y[m] = Gd[m], \tag{4.1}$$

and our objective is to solve for the elements of G. We can write the aforementioned problem of estimating G in the linear system of equations as a solution to the following underdetermined system of linear equations at every time step m:

$$y[m] = D[m]g, \tag{4.2}$$

where

$$D[m] \coloneqq \begin{bmatrix} [d_1[m]\ 0_{1\times N-1}] & [d_2[m]\ 0_{1\times N-1}] & \cdots & [d_N[m]\ 0_{1\times N-1}] \\ [0\ d_1[m]\ 0_{1\times N-2}] & [0\ d_2[m]\ 0_{1\times N-2}] & \cdots & [0\ d_N[m]\ 0_{1\times N-2}] \\ \vdots & & & \\ [0_{1\times N-1}\ d_1[m]] & [0_{1\times N-1}\ d_2[m]] & \cdots & [0_{1\times N-1}\ d_N[m]] \end{bmatrix} \tag{4.3}$$

$$= \begin{bmatrix} d_1[m]I_{N\times N} & d_2[m]I_{N\times N} & \cdots & d_N[m]I_{N\times N} \end{bmatrix} \in \mathbb{R}^{N\times N^2}, \tag{4.4}$$

and

$$g \coloneqq \operatorname{vec}(G) \tag{4.5}$$

is the vectorized (in column-stack manner) G. The least-norm solution to the above underdetermined system at time step m is

$$g_{\text{ln}}[m] = D[m]^\top \left(D[m]D[m]^\top\right)^{-1} y[m]. \tag{4.6}$$

One naïve approach to estimating G is to refer back to the definition of matrix multiplication. We have

$$y[m] = Gd[m] = \vec{g}_1 d_1[m] + \vec{g}_2 d_2[m] + \cdots + \vec{g}_N d_N[m], \tag{4.7}$$

where $\vec{g}_n$ is the n^{th} column of G for $n = 1, \ldots, N$. By allowing only one element of $d[m]$ be one and the rest be zeros, i.e. stimulate one neuron at a time, we can figure out G in $\mathcal{O}(N)$ time. However, stimulating one neuron may not provide enough impetus to yield useful observations for estimation. One neuron is often not powerful enough to drive other neighboring neurons. The matrix D, although high-dimensional, is sparse. Sparse matrix multiplication algorithms can be employed to further speed up the computation of the least-norm solution as described, but rank deficiency caused by the severe sparsity presents a problem in the above method. Because D is extremely sparse, especially when we can excite only a few neurons in an ensemble, the inverse problem is extremely ill-posed and ill-conditioned. Therefore, a method that results in a denser D is necessary. Furthermore, in order to meet the model criteria, we must instead solve the following optimization problem, which has the constraints $|g_{n\ell}| \leq 1$ for every element $g_{n\ell}$ of G:

$$\min_{g} \quad \|g\|_2 \tag{4.8}$$

$$\text{s.t.} \quad D[m]g = y[m] \tag{4.9}$$

$$|g_{n\ell}| \leq 1 \; \forall \; n, \ell \in \{1, \ldots, N\}, \tag{4.10}$$

which adds to the complexity. Moreover, the aforementioned model is too simplistic to capture the dynamics of the neural network. The response of the network is not solely a function of the stimulation profile d. Instead, it is a function of both the stimulation to the system as well as the activation states of the neurons that constitute the network. Therefore, a model that is complex enough to capture the main network dynamics while staying simple enough to remain both biologically and physically meaningful, as well as bypass overfitting to the data, is presented in the following sections.

4.2 Autoregressive Model

Neural responses vary from one trial to another even when the same stimulus is applied. Potential sources of variation include differing levels of arousal and attention, and randomness associated with biological and cognitive processes affecting neuronal firing. This lack of regularity prevents accurate deterministic modeling of when the action potential will fire. Instead, probabilistic models are used in the literature to predict spike sequences based on specific stimuli. We do not aim to model the neural spike train of every neuron or the actual membrane potential dynamics of every neuron. The commonly used leaky integrate-and-fire, Hodgkin-Huxley, and Izhikevich neuron models all increase the complexity of our mesoscale dynamics model, and thus are used only to derive the appropriate model for the context of our problem setting. We are more interested in modeling the population membrane potential dynamics (of all neurons residing within the field-of-view), and take advantage of the slow dynamics of GCaMP relative to that of the action potential to justify the following linear model. This population-level decoding approach mitigates the aforementioned issues. We derive a neural dynamic model from the fundamental Hodgkin-Huxley model, which results in an autoregressive (AR) process of network responses that further incorporates feed-forward photostimulation profile inputs.

Model Derivation

We derive the autoregressive model of the V1 L2/3 circuit with Hodgkin-Huxley neurons. Figure 4.1 summarizes the autoregressive model derivation.

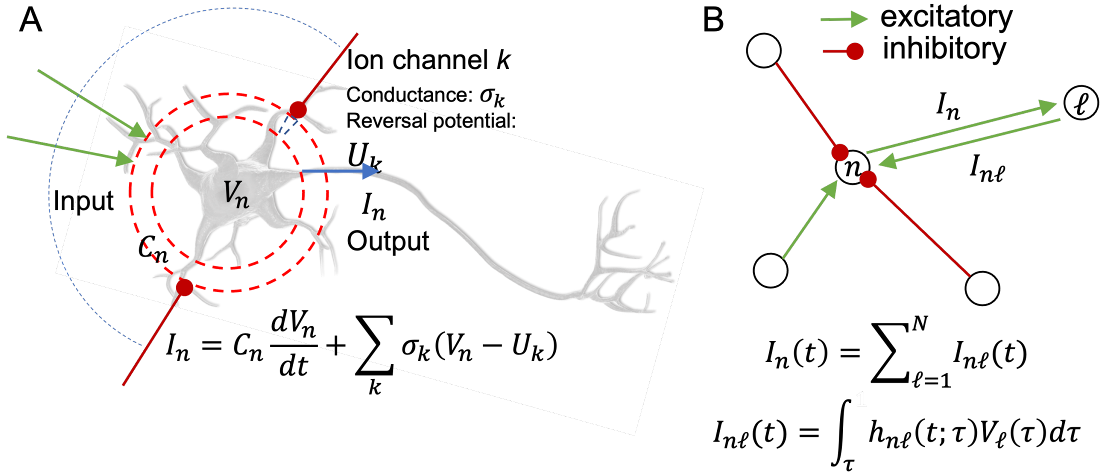
A
Ion channel k
Conductance: σ_k
Reversal potential: U_k
V_n
Input
I_n
Output
C_n
$I_n = C_n \frac{dV_n}{dt} + \sum_k \sigma_k (V_n - U_k)$
B
excitatory
inhibitory
I_n
ℓ
n
$I_{n\ell}$
$I_n(t) = \sum_{\ell=1}^{N} I_{n\ell}(t)$
$I_{n\ell}(t) = \int_\tau h_{n\ell}(t;\tau) V_\ell(\tau) d\tau$

Figure 4.1: Summary of the autoregressive model derivation from Hodgkin-Huxley model neurons.

We model the dynamics of the neural network as an autoregressive process of photostimulation profiles and network responses. To achieve this, we derive a system of equations that models the circuit dynamics using neurons that are connected via axons and synapses that conduct electrical currents. Each neuron receives both excitatory and inhibitory input currents. Individual neurons follow the Hodgkin-Huxley model. The Hodgkin-Huxley model states that for a single neuron, the total membrane current is

$$I = C \frac{\mathrm{d}V}{\mathrm{d}t} + \sum_{k=1}^{K} \sigma_k \left(V - U_k \right), \tag{4.11}$$

where C is the membrane capacitance, V is the membrane potential, σ_k is the electrical conductance of ion or leak channel k, and U_k is the reversal potential of channel k. Conduction of currents in this circuit follows known laws of physics (e.g., Kirchhoff's Current and Voltage laws) with consideration of propagation delay. For a circuit of N neurons, we can regard every neuron as a junction or a node in a network, where I_n is the total current generated by neuron n and I_{nm} is the current flow from neuron n to neuron m. By Kirchhoff's Current Law (KCL),

$$I_n = \sum_{\ell=1}^{N} I_{n\ell}. \tag{4.12}$$

Then, we have

$$C\frac{\mathrm{d}V_n}{\mathrm{d}t} + \sum_{k} \sigma_k \left(V_n - U_k\right) = \sum_{\ell=1}^{N} I_{n\ell} = I_n. \tag{4.13}$$

Assuming that the extracellular fluid acts as the common ground, we have

$$\sum_{n=1}^{N} I_n = 0 \tag{4.14}$$

from Kirchhoff's Current Law and

$$\sum_{n=1}^{N} V_n = 0 \tag{4.15}$$

from Kirchhoff's Voltage Law (KVL). Combining KCL and KVL, we have

$$I_n = \sum_{\ell=1}^{N} a_{n\ell} V_\ell, \tag{4.16}$$

where a_{nm} terms are to be determined. Substituting, we obtain

$$C\frac{\mathrm{d}V_n}{\mathrm{d}t} + \sum_{k=1}^{K} \sigma_k \left(V_n - U_k\right) = \sum_{\ell=1}^{N} a_{n\ell} V_\ell. \tag{4.17}$$

We assume that only DC current exists. Let us now consider propagation delay. With propagation delay, we obtain

$$C\frac{\mathrm{d}V_n}{\mathrm{d}t} + \sum_{k} \sigma_k \left(V_n - U_k\right) = \sum_{i=1}^{p} \sum_{\ell=1}^{N} a_{n\ell}(t_i) V_\ell(t_i). \tag{4.18}$$

Because our experimental measurement of neural activity is acquired at discrete time points, we discretize Equation 4.18. In discrete time and assuming that we have a causal system, we obtain

$$C\frac{V_n[m] - V_n[m-1]}{\Delta t} + \sum_{k} \sigma_k \left(V_n[m] - U_k\right) = \sum_{i=1}^{p} \sum_{\ell=1}^{N} a_{n\ell}[i] V_\ell[m-i] \tag{4.19}$$

$$\frac{C}{\Delta t} V_n[m] - \frac{C}{\Delta t} V_n[m-1] + \sum_{k} \sigma_k V_\ell[m] - \sum_{k} \sigma_k U_k = \sum_{i=1}^{p} \sum_{\ell=1}^{N} a_{n\ell}[i] V_\ell[m-i]. \tag{4.20}$$

Combining the discretized equations and rearranging, we have

$$\left(\frac{C}{\Delta t}+\sum_k \sigma_k\right)V_n[m] = \sum_{i=1}^{p}\sum_{\ell=1}^{N} a_{n\ell}[i]V_\ell[m-i]+\frac{C}{\Delta t}V_n[m-1]+\sum_k \sigma_k U_k \tag{4.21}$$

$$= \sum_{i=1}^{p}\sum_{\ell=1}^{N} a_{n\ell}[i]V_\ell[m-i]+\frac{C}{\Delta t}V_n[m-1]+\sum_k \sigma_k U_k \tag{4.22}$$

$$= \sum_{i=1}^{p}\sum_{\ell=1}^{N} \left(a_{n\ell}[i]+\frac{C}{\Delta t}\delta_{n\ell}\delta k1\right)V_\ell[m-i]+\sum_k \sigma_k U_k. \tag{4.23}$$

where δ_{ij} is the Kronecker delta. Define now the following:

$$g_{n\ell}[i] \coloneqq \frac{a_{n\ell}[i]+\frac{C}{\Delta t}\delta_{n\ell}\delta k1}{\frac{C}{\Delta t}+\sum_k \sigma_k}, \tag{4.24}$$

$$V_0 \coloneqq \frac{\sum_k \sigma_k U_k}{\frac{C}{\Delta t}+\sum_k \sigma_k}. \tag{4.25}$$

Note that V_0 is a constant. Then, we obtain

$$V_n[m] = \sum_{i=1}^{p}\sum_{\ell=1}^{N} g_{n\ell}[i]V_\ell[m-i]+V_0. \tag{4.26}$$

Considering stimulation and noise, which subsumes V_0, we arrive at

$$V_n[m] = \sum_{i=1}^{p}\sum_{\ell=1}^{N} g_{n\ell}[i]V_\ell[m-i]+s_{n\ell}d_\ell[m]+\varepsilon_n[m], \tag{4.27}$$

which describes an autoregressive relationship for a sequence of voltage $V_n[m]$ of neuron n at time step m. Equation 4.27 also includes the external stimulation d_n and noise ε_n associated with neuron n. $g_{n\ell}[k]$ is a dimensionless physical parameter that relates the voltage of neuron ℓ to the present and past voltage of neuron n at time point $m-i$. This term includes contributions from propagation delay $h_{n\ell}$, membrane capacitance C_n and ion channel conductance σ_k. On the other hand, $s_{n\ell}$ is a parameter that describes the direct effect of the photostimulation. $g_{n\ell}[k]$ and $s_{n\ell}$ are the unknowns that need to be estimated by varying the stimulation patterns $d_n[m]$. Given N neurons in the circuit, the total number of unknowns are of order N^2. The parameter $s_{n\ell}$ defines the elements of the photostimulation connectivity matrix S; $g_{n\ell}[i]$ defines the elements of the autoregressive connectivity matrix G_i of order i. These matrices characterize the effective connections among the observed neurons as well as the coefficients that describe the temporal dynamics of the circuit.

Noise Model

The term $\varepsilon[m]$ is the spontaneous activity noise vector based on the Poisson arrival process. We assume that there is no image noise in the observed image data, and use the spontaneous activities of individual neurons as the sole quantified noise source in the autoregressive model. For each neuron, we model its spontaneous activity by a Poisson arrival process with a common rate parameter ν, which we can estimate from observation before stimulation. We choose this stochastic process model because with the Poisson arrival process, the interarrival time between two arrival events follows the exponential distribution. We can exploit the memorylessness property of exponential random variables to simplify our observation model. Consider $T_n \sim \text{Exp}(\nu)$ for $n = 1, \ldots, N$, and let τ be the sampling interval, i.e., the period between observations or the time lag between steps $m+1$ and m is τ. Then, the n^{th} element of $\varepsilon[m]$ is given by

$$\varepsilon_n[m] = \mathbb{I}(T_n \leq \tau), \tag{4.28}$$

where $\mathbb{I}(\cdot)$ denotes the indicator random variable.

4.3 GCaMP Autoregressive Model

The network responses are mapped to normalized potential outcome values to form the following model. For every $m \geq m_k$, where m denotes the time index of imaging m_k is the start time index of the k^{th} stimulation, we have $m_k = Lk$, where L is the interstimulus interval (ISI), the number of imaging time steps taken per one stimulation step. For brevity and clarity in explanation, we simply refer to this sampling rate expander as d. Recall that y is the standardized network response vector, where standardization was performed using the cumulative distribution function (CDF) of the activated state Gaussian in the approximate superpopulation model. The CDF function specific to every neuron profile normalizes the respective potential outcomes for population-level analysis. The term ε models the spontaneous activity noise in the system. The model parameters of interest are G_is, the effective connectivity matrices, and S_js, the photostimulation effects matrices. We also estimate ψ, the vector weight coefficients that serve the role of the convolution kernel capturing the temporal GCaMP dynamics. This model can be written as follows if the interstimulus interval L is large enough to account for the complete decay of the GCaMP dynamics. In this case, one stimulation profile is decoupled from another, and the autoregressive model after ignoring noise for every $m \geq m_s$, where m_s is the start time index of the k^{th} stimulation-response observation pair, becomes the following.

Initialize $m_h = m_k$. For every $m \geq m_k$ and the k^{th} photostimulation profile d_k defined by the corresponding 3D-SHOT hologram, we have

$$y\left[m\right] = \sum_{i=0}^{p-1} \frac{G_i^+[m] \max\left\{Y^{(i)}[m]\right\}}{N_G[m]} \psi\left(m; m_h\right) + S_j^+[m] d_k, \tag{4.29}$$

$$m_h = \begin{cases} m' + 1, & \text{if } y[m'] \leq \theta_0,\ m_h \leq m' \leq m \\ m', & \text{if } y[m'] \geq \theta_1,\ m_h \leq m' \leq m\ , \\ m_h, & \text{otherwise} \end{cases} \tag{4.30}$$

where

$$G_i^+[\cdot] \coloneqq G_i \mathbb{I}(\cdot - (iu+1) \geq m_s), \tag{4.31}$$

$$N_G[\cdot] \coloneqq \max\left\{1, \sum_{j=0}^{p-1} \mathbb{I}(\cdot - (ju+1) \geq m_s)\right\}, \tag{4.32}$$

$$\max\left\{Y_n^{(i)}[\cdot]\right\} \coloneqq \max\left\{y_n[\cdot - (i+1)u] \mathbb{I}(\cdot - (i+1)u \geq m_s), \ldots,\right. \tag{4.33}$$

$$\left. y_n[\cdot - (iu+1)] \mathbb{I}(\cdot - (iu+1) \geq m_s)\right\}\ \forall\ n, \tag{4.34}$$

$$S_j^+[\cdot] \coloneqq S_{\min\{\lfloor \frac{\cdot - m_d - m_k}{v} \rfloor, q-1\}} \mathbb{I}(\cdot - m_d \geq m_k), \tag{4.35}$$

$$\psi\left(\cdot; m_h\right) \coloneqq b + (1-b) e^{-\tau[\cdot - m_h]_+}. \tag{4.36}$$

There are p autoregressive connectivity matrices G_is and q photostimulation effects matrices S_js. The indicator functions are denoted by the superscript + in the above autoregressive model equation in the definitions of the connectivity and photostimulation effects matrices G_i^+s and S_j^+s, respectively. These indicator functions ensure that the network responses from decoupled stimuli do not affect the temporal portion of the dynamics under investigation. In terms of estimation, the indicator functions prevent data from other decoupled stimuli from contaminating the estimation process. The term N_G counts the number of active autoregressive connectivity matrices G_is as higher order G_is may not be active near the beginning of the start of the response observation of every stimulation. This term enforces proper normalization of the right-hand-side regardless of the number of active G matrices.

The model takes the maximum response observed in each time bin window in the past corresponding to the respective autoregressive connectivity matrix G_i to explain the present observation. The parameters u and v specify how many imaging frames each G matrix and S matrix integrates, respectively, to explain the present observation. This approach strengthens the model's robustness, especially against slight unwanted temporal variations in the fluorescence observations due to the slow decay rate of GCaMP with respect to the sampling rate. The slowly decaying GCaMP dynamics is modeled by ψ, which depends on both the current time index m and the time index of the last fully activated or rested state of every neuron m_h. A neuron's state is determined using the fitted two-component Gaussian mixture model profile of that particular neuron. Whenever a fluorescence signature of a neuron is not at its fully excited or rested state, a timer starts, and exponential decay modeling the dynamics of GCaMP is applied to prevent GCaMP decay dynamics from degrading the interpretation of the autoregressive connectivity matrices. Subscript + denotes the Rectified Linear Unit activation function. Due to the slow dynamics of GCaMP, observations are delayed by approximately a few hundred milliseconds. This system-wide delay factor is captured by m_d. The above equations collectively describe the mesoscale dynamics of the population of neurons under observation. θ_0 is set to be the mean of the baseline Gaussian and θ_1 is set to be the mean of the Gaussian corresponding to the activated state in the fitted Gaussian mixture model. The term b is set to be the standard deviation of baseline activity, or the noise standard deviation.

4.4 Spike Autoregressive Model

GCaMP has served as the proxy for action potentials in our system. The above autoregressive model can easily be converted to a spike-based model if neural population activity is measured directly in terms of action potential spikes. In the case where sampling rate is fast enough to capture single action potential spikes, y can be interpreted as a binary vector of action potential events for all neurons in the field-of-view. Then, the spike-based autoregressive model can be written as follows.

Drawing parallels, we initialize $m_h = m_k$. Then, for every $m \geq m_k$ and the k^{th} photostimulation profile d_k, we have

$$y\left[m\right] = \sum_{i=0}^{p-1} \frac{G_i^+[m] \max\left\{Y^{(i)}[m]\right\}}{N_G[m]} + S_j^+[m] d_k, \tag{4.37}$$

where

$$G_i^+[\cdot] \coloneqq G_i \mathbb{I}(\cdot - (iu+1) \geq m_s), \tag{4.38}$$

$$N_G[\cdot] \coloneqq \max\left\{1, \sum_{j=0}^{p-1} \mathbb{I}(\cdot - (ju+1) \geq m_s)\right\}, \tag{4.39}$$

$$\max\left\{Y_n^{(i)}[\cdot]\right\} \coloneqq \max\left\{y_n[\cdot - (i+1)u]\mathbb{I}(\cdot - (i+1)u \geq m_s), \ldots,\right. \tag{4.40}$$

$$\left. y_n[\cdot - (iu+1)]\mathbb{I}(\cdot - (iu+1) \geq m_s)\right\} \;\forall\; n, \tag{4.41}$$

$$S_j^+[\cdot] \coloneqq S_{\min\{\lfloor\frac{\cdot - m_d - m_k}{v}\rfloor, q-1\}} \mathbb{I}(\cdot - m_d \geq m_k), \tag{4.42}$$

$$\tag{4.43}$$

The notations used in the spike-based autoregressive model remain the same as in the aforementioned GCaMP-based autoregressive model. By removing the GCaMP dynamics model and changing the representation of the observation vector y, we arrive at a model that can be used to address questions regarding synaptic–rather than effective–connections. This derivative version of the model may be insightful and useful in cases where action potentials of neurons in the field-of-view can be measured in conjunction with holography optogenetics setup.

Chapter 5

Estimation Algorithm

We interrogate circuit dynamics efficiently with compressed sensing and optogenetics. The AR model will characterize the circuit dynamics with order N^2 parameters. Determining these parameters generally requires at least N^2 independent experimental measurements, which is impractical. To overcome this challenge, we employ sparse 2P holographic optogenetic stimulation. The approach combines optogenetic stimulation via 3D-SHOT and large-scale volumetric calcium imaging to optimally estimate the unknown AR model parameters describing the circuit dynamics of neurons residing inside a volume that is hundreds of microns in dimensions. The problem of choosing which neurons to stimulate at each stimulation step to recover the underlying neural network dynamics with as few observations as possible is equivalent to designing the measurement (sensing) matrix in compressed sensing. By parallelizing and constructing a general sensing matrix, we can design an incoherent sensing matrix that can recover the unknowns of the underdetermined system of linear equations at hand. According to compressed sensing theory, the best method of selection is to randomly choose from a set of available neurons to stimulate [16]. In the context of our problem formulation, we have a sensing matrix with columns that are fixed (elements are time lags of the potential outcome observations) while the last qN columns correspond to the indicator of each neuron's inclusion in the stimulation profile for every corresponding photostimulus matrix. In this part of the sensing matrix where we can design, every row is the stimulation profile vector d, where the elements of the vector are randomly selected to be 1 or 0 ($d_n = 1$ specifies that neuron n is targeted for stimulation) such that the sum of the vector is equal to the maximum number of neurons that can be stimulated at one instance in time. The compressed sensing framework allows us to efficiently interrogate the circuit and learn the AR model.

5.1 Parameter Estimation

In this section, we detail how we estimate the parameters of the autoregressive model of arbitrary order p as well as arbitrary number q of feed-forward dynamics governed by the photostimulations: the effective connectivity matrices G, which we know from neurobiology to be sparse, as well as the photostimulation effect matrices S.

5.2 Compressed Sensing

The problem of choosing which neurons to stimulate at each stimulation step to recover the underlying neural network with as few observations as possible is equivalent to designing the measurement (sensing) matrix in compressed sensing. By parallelizing and constructing a general sensing matrix, we can design an incoherent sensing matrix that can recover the unknowns ($g_{n\ell}$, the elements of G_i for all $n, \ell \in \{1, \ldots, N\}$ and for all $i \in \{0, \ldots, p-1\}$; and $s_{n\ell}$, the elements of S_j for all $n, \ell \in \{1, \ldots, N\}$ and for all $j \in \{0, \ldots, q-1\}$) of the underdetermined system of linear equations at hand. According to compressed sensing theory, the best method of selection is to randomly choose from a set of available neurons to stimulate. In the context of our problem formulation, we have a sensing matrix with columns that are fixed (elements are time lags of the potential outcome observations) while the last N columns correspond to the indicator of each neuron's inclusion in the stimulation profile. In this part of the sensing matrix where we can design, every row is the stimulation profile vector d, where $d_n(\cdot) \in \{0, 1\}$, $n \in \{1, \ldots, N\}$, is randomly selected such that $\sum_{n=1}^{N} d_n(\cdot) = N_{\max}$, the maximum number of neurons that can be stimulated at one instance in time. The technology we have today allows stimulation of up to 50 neurons at one time. To elaborate, $N_{\max}$ randomly selected elements of $d(\cdot)$ are 1, and the rest are 0. The structure of the sensing matrix that recovers the estimands of our problem is detailed below.

In the context of our problem formulation, compressed sensing concerns recovering the sparse solution $g \in \mathbb{R}^{N^2}$ in an underdetermined system $y = Ag$, where $y \in \mathbb{R}^M$ and $A \in \mathbb{R}^{M \times N^2}$ is the sensing matrix. Using compressed sensing techniques, we need only $M = \mathcal{O}\left(k \log N^2\right)$ to uniquely recover g, where k is the sparsity level or the number of nonzero entries of g, i.e. $\|g\|_0$. $M = \mathcal{O}\left(k \log N^2\right)$ is known as the information rate.

Compressed Sensing for Individual Neuron Model

The simple model failed to capture the network dynamics, so the autoregressive model was developed. Even with the autoregressive model, the fixed, restrictive structure of the model renders the associated inverse problem of recovering the unknowns of the model impossible. In the autoregressive model, estimating the effective connectivity matrix and the autoregressive coefficients using compressed sensing at once becomes impractical. We circumvent this issue by recasting the full autoregressive model into what we call "individual neuron observations."

Focusing on one single neuron and designing a compressed sensing scheme for estimating all parameters of the model related to the individual neuron frees the inverse problem from the limiting structure. For every neuron n, we have a total of M observations.

5.3 Compressed Sensing for Individual Neurons

We can write the m^{th} observation for neuron n as

$$
\begin{aligned}
y_n[m] = \sum_{i=0}^{p-1} & \frac{g_{ni}^\top \max\left\{Y^{(i)}[m]\right\} \psi_n\left(m; m_h\right) \mathbb{I}\left(m - (iu+1) \geq m_s\right)}{\max\left\{1, \sum_{j=0}^{p-1} \mathbb{I}(m - (ju+1) \geq m_s)\right\}} && (5.1) \\
& + s^\top_{n \min\{\lfloor \frac{m - m_d - m_k}{v} \rfloor, q-1\}} d_k \mathbb{I}\left(m - m_d \geq m_k\right) && (5.2) \\
= s^\top_{n \min\{\lfloor \frac{m - m_d - m_k}{v} \rfloor, q-1\}} & d_k && (5.3) \\
& + g_{n0}^\top \frac{\max\left\{Y^{(0)}[m]\right\} \psi_n\left(m; m_h\right)}{\max\left\{1, \sum_{j=0}^{p-1} \mathbb{I}(m - (ju+1) \geq m_s)\right\}} && (5.4) \\
& + g_{n1}^\top \frac{\max\left\{Y^{(1)}[m]\right\} \psi_n\left(m; m_h\right)}{\max\left\{1, \sum_{j=0}^{p-1} \mathbb{I}(m - (ju+1) \geq m_s)\right\}} && (5.5) \\
& + \cdots && (5.6) \\
& + g_{n(p-1)}^\top \frac{\max\left\{Y^{(p-1)}[m]\right\} \psi_n\left(m; m_h\right)}{\max\left\{1, \sum_{j=0}^{p-1} \mathbb{I}(m - (ju+1) \geq m_s)\right\}}, && (5.7)
\end{aligned}
$$

where $g_n^\top$ and $s_n^\top$ are the n^{th} rows of G and S, respectively, and indicator functions have been omitted for brevity. For every neuron n, define the following:

$$
\begin{aligned}
\rho_n[m] &:= y_n[m], && (5.8) \\
\chi_n &:= \begin{bmatrix} g_{n0}^\top & g_{n1}^\top & \cdots & g_{n(p-1)}^\top & s_{n0}^\top & \cdots & s_{n(q-1)}^\top \end{bmatrix}^\top . && (5.9)
\end{aligned}
$$

For brevity, we omit the indicator functions in the following equations. Denote by $a_m^\top$ the m^{th} row of the measurement matrix A:

$$
a_m^\top := \begin{bmatrix} \frac{\max\left\{Y^{(0)}[m]\right\}\psi_n(m;m_h)}{\max\left\{1, \sum\limits_{j=0}^{p-1} \mathbb{I}(m-(ju+1) \geq m_s)\right\}} & \cdots & \frac{\max\left\{Y^{(p-1)}[m]\right\}\psi_n(m;m_h)}{\max\left\{1, \sum\limits_{j=0}^{p-1} \mathbb{I}(m-(ju+1) \geq m_s)\right\}} & \overset{d_k^\top \otimes}{e^\top_{\min\{\lfloor \frac{m-m_d-m_k}{v} \rfloor, q-1\}}} \end{bmatrix}, \quad (5.10)
$$

where e_k is the $(q-1)$-length standard basis vector. Then, rewriting to solve for the unknowns, the m^{th} observation for neuron n becomes

$$
\rho_n[m] = a_m^\top \chi_n. \quad (5.11)
$$

Collecting M measurements and letting ρ_n denote the vector with elements $\rho_n[m]$, the autoregressive model parameter estimation procedure–the problem of estimating the effective connectivities and the photostimulus effect weights–materializes into the following constrained

Least Absolute Shrinkage and Selection Operator (LASSO) optimization problem:

$$\min_{g_n, s_n} \; \mathcal{F}(g_n, s_n) \coloneqq \frac{1}{2M} \|\rho_n - A\chi_n(g_n, s_n)\|_2^2 + \lambda_G \|\chi_n(g_n)\|_1 + \lambda_S \|\chi_n(s_n)\|_1 \tag{5.12}$$

$$\text{s.t.} \quad |g_n| \leq 1 \tag{5.13}$$

$$|s_n| \leq 1. \tag{5.14}$$

Let

$$\chi_n^* \coloneqq \begin{bmatrix} g_{n0}^{*\top} & g_{n1}^{*\top} & \cdots & g_{n(p-1)}^{*\top} & s_{n0}^{*\top} & \cdots & s_{n(q-1)}^{*\top} \end{bmatrix}^\top \in \mathbb{R}^{(p+q)N \times 1}. \tag{5.15}$$

denote the unique solution to the above convex optimization problem. We can solve the above optimization problem using the Greedy Fast Iterative Shrinkage-Thresholding Algorithm (Greedy FISTA), which is the novel accelerated version of the proximal gradient algorithm on the LASSO, with projection steps to address the constraints. The implemented iterative method is detailed in Algorithm 2 is detailed below.

Algorithm 2: Greedy FISTA to find ϵ-optimal solution

```
Input  : Sensing matrix $A$, observation $y$, regularization parameters $\lambda$, and
         $L = \frac{1}{M}\lambda_{\max}\left(A^\top A\right)$, the smallest Lipschitz constant of the gradient
         $\nabla\mathcal{F}(\chi_n)$.
Output: $u^{(k)}$.
Initialize: Let $t \in [\frac{1}{L}, \frac{2}{L}]$, $\xi < 1$, $S > 1$, $u^{(0)} \in \mathbb{R}^N$, $u^{(-1)} = u^{(0)}$.
$k_{\max} = \left\lceil \frac{C}{\sqrt{\epsilon}} - 1 \right\rceil$, where $C = \sqrt{2L\|u^{(0)} - u^*\|_2^2}$.
for $k = 1$ to $k_{\max}$ do
    1. $v^{(k)} = u^{(k)} + \left(u^{(k)} - u^{(k-1)}\right)$
    2. $u^{(k+1)} = p_{\frac{1}{t}}\left(v^{(k)}\right)$
    Restarting:
    if $\left(v^{(k)} - u^{(k)}\right)^\top \left(u^{(k+1)} - u^{(k)}\right) \geq 0$ then
        $v^{(k)} = u^{(k)}$
    end
    Safeguard:
    if $\left\|u^{(k+1)} - u^{(k)}\right\|_2 \geq S\left\|u^{(1)} - u^{(0)}\right\|_2$ then
        $t = \max\left\{\xi t, \frac{1}{L}\right\}$
    end
end
```

FISTA has the convergence rate

$$\mathcal{F}(\chi_n^{(k)}) - \mathcal{F}(\chi_n^*) \leq \frac{4\lambda_{\max}\left(A^\top A\right)\|\chi_n^{(0)} - \chi_n^*\|_2^2}{(k+1)^2}, \tag{5.16}$$

where the operator $\lambda_{\max}$ denotes the maximum eigenvalue of the argument. After $M = \mathcal{O}\left(k \log(p+1)N\right)$ observations, we can expect to recover the unique solution χ_n^*.

We expect that the sparsity level k is in the order of 10% of the total number of possible connections N^2, i.e., $k = 0.1N^2$ [40]. The unknowns $g_n^\top$ and $s_n^\top$ in the individual neuron observation model can be mapped directly to the full autoregressive model parameters. For example, $g_n^\top$ is the n^{th} row of G, i.e.,

$$G_i = \begin{bmatrix} \longleftarrow & g_{0i}^{*\top} & \longrightarrow \\ \longleftarrow & g_{1i}^{*\top} & \longrightarrow \\ & \vdots & \\ \longleftarrow & g_{Ni}^{*\top} & \longrightarrow \end{bmatrix} \in [-1,1]^{N\times N}. \tag{5.17}$$

Parallel Computation

By selecting a general incoherent sensing matrix A for every neuron $n \in \{1,\ldots,N\}$, we can significantly decrease the time needed to compute the estimated Gs and Ss from compressed sensing. Because we observe the calcium trace of every neuron in the field-of-view, we can solve the aforementioned optimization problem for all $n \in \{1,\ldots,N\}$, i.e., run Greedy FISTA N times in parallel, once per every neuron n. We can then obtain Gs and Ss using only $M = \mathcal{O}\left(k \log(p+1)N\right)$ observations.

5.4 Measurement Matrix

We sacrifice the maximum incoherence we can obtain for the measurement (sensing) matrix A in order to allow for parallel computation. By having one general A for all neurons $n \in \{1,\ldots,N\}$, we can significantly decrease the computational cost associated with reconstructing Gs and Ss.

Due to the autoregressive nature of the model and how the unknowns of the system are placed in the model, we have, given that the experiment starts at time-step m:

$$A := \left[\begin{array}{ccc|c} \frac{\max\left\{Y^{(0)}[m]\right\}\psi_n(m;m_h)}{\max\left\{1,\sum_{j=0}^{p-1}\mathbb{I}(m-(ju+1)\geq m_s)\right\}} & \cdots & \frac{\max\left\{Y^{(p-1)}[m]\right\}\psi_n(m;m_h)}{\max\left\{1,\sum_{j=0}^{p-1}\mathbb{I}(m-(ju+1)\geq m_s)\right\}} & \begin{array}{c} d_k^\top \otimes \\ e^\top_{\min\{\lfloor \frac{m-m_d-m_k}{v}\rfloor,q-1\}} \end{array} \\ \frac{\max\left\{Y^{(0)}[m+1]\right\}\psi_n(m+1;m_h)}{\max\left\{1,\sum_{j=0}^{p-1}\mathbb{I}(m+1-(ju+1)\geq m_s)\right\}} & \cdots & \frac{\max\left\{Y^{(p-1)}[m+1]\right\}\psi_n(m+1;m_h)}{\max\left\{1,\sum_{j=0}^{p-1}\mathbb{I}(m+1-(ju+1)\geq m_s)\right\}} & \begin{array}{c} d_k^\top \otimes \\ e^\top_{\min\{\lfloor \frac{m+1-m_d-m_k}{v}\rfloor,q-1\}} \end{array} \\ & \vdots & & \vdots \\ \frac{\max\left\{Y^{(0)}[m+M-1]\right\}\psi_n(m+M-1;m_h)}{\max\left\{1,\sum_{j=0}^{p-1}\mathbb{I}(m+M-1-(ju+1)\geq m_s)\right\}} & \cdots & \frac{\max\left\{Y^{(p-1)}[m+M-1]\right\}\psi_n(m+M-1;m_h)}{\max\left\{1,\sum_{j=0}^{p-1}\mathbb{I}(m+M-1-(ju+1)\geq m_s)\right\}} & \begin{array}{c} d_k^\top \otimes \\ e^\top_{\min\{\lfloor \frac{m+M-1-m_d-m_k}{v}\rfloor,q-1\}} \end{array} \end{array}\right], \tag{5.18}$$

$A \in \mathbb{R}^{M\times(p+q)N}$, where $d_n(\cdot) \in \{0,1\}$, $n \in \{1,\ldots,N\}$, is randomly assigned such that $\sum_{n=1}^{N} d_{kn}(\cdot) = N_{\max}$, the maximum number of neurons that can be stimulated ($N_{\max}$ randomly selected elements of $d_k(\cdot)$ are 1, and the rest are 0).

Figure 5.1 below highlights the dichotomous structure of the above compressed sensing measurement matrix, where the left pN columns are constructed from the past activity observations of the neurons in the field-of-view and are used to estimate G_is; and the right qN columns are based on the photostimulation profiles and are used to estimate S_js.

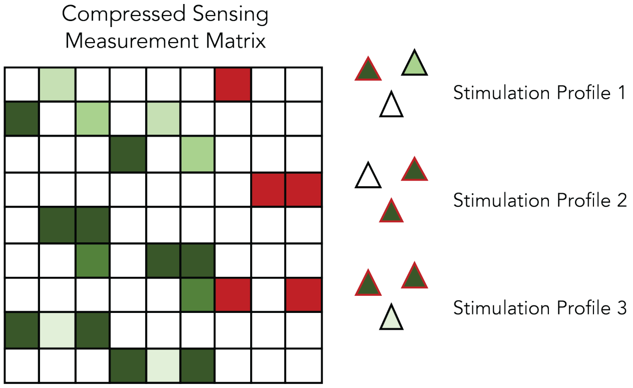

Figure 5.1: Graphical representation of the compressed sensing measurement matrix specific to the autoregressive model setting. The matrix is built from the past observations of neurons in the field-of-view as well as the photostimulation profiles.

The Greedy Fast Iterative Shrinkage-Thresholding Algorithm (Greedy FISTA) applied to the optimization problem 5.12 is detailed in Algorithm 2. Denote by $\mathcal{G} : \mathbb{R}^{(p+1)N} \mapsto \mathbb{R}$ a continuous, convex function which may be nonsmooth, and denote by $\mathcal{H} : \mathbb{R}^{(p+1)N} \mapsto \mathbb{R}$ a smooth, convex function that is continuously differentiable with Lipschitz continuous gradient, i.e. $\|\nabla\mathcal{H}(x) - \nabla\mathcal{H}(y)\| \leq L\|x - y\| \ \forall\ x, y \in \mathbb{R}^{(p+1)N}$, where L is the Lipschitz constant of the gradient of $\mathcal{H}$. Following the convention by Beck and Teboulle [2], denote by $p_L(v)$ the solution to the proximal operator, i.e.,

$$p_L(v) = \arg\min_u \quad \left\{\mathcal{G}(u) + \frac{L}{2}\big\|u - \big(v - \frac{1}{L}\nabla\mathcal{H}(v)\big)\big\|_2^2\right\}. \tag{5.19}$$

For $\mathcal{G}\left(\chi_n\right) = \lambda\left\|\chi_n\right\|_1$ and $\mathcal{H}\left(\chi_n\right) = \frac{1}{2}\left\|\rho_n - A\chi_n\right\|_2^2$,

$$p_L(v) = \mathcal{T}_{\frac{\lambda}{L}}\big(v - \frac{2}{L}A^\top(Av - y)\big), \tag{5.20}$$

where the soft-thresholding operator is defined

$$\mathcal{T}_\alpha(\cdot) \coloneqq (|\cdot| - \alpha)_+ \operatorname{sgn}(\cdot). \tag{5.21}$$

Algorithm 2 finds the ϵ-optimal solution, i.e. $\mathcal{F}\left(\chi_n\right) - \mathcal{F}\left(\chi_n^*\right) \leq \epsilon$.

Chapter 6

In silico Results and Validations

In silico, we built a ground truth network consisting of 500 cortical neurons (a purposely smaller number of neurons was considered for best visibility) whose dynamics follows the AR model with considerable noise. We set both connection probability and strength in the network as functions of distance between pairs of neurons. Connection probabilities were sampled from the Gamma distribution with parameters $\alpha = 3$ and $\beta = 1/40$ to best match what has been observed in vivo. The connection strengths were set such that the strength was inversely proportional to the distance between the two connected neurons. Using the aforementioned algorithm and the estimation procedure, we calculated the model parameters using 400 stimulation profiles, each consisting of 30 neurons being stimulated at a time. With the sampling interval matching that of the in vivo experiment to be described in the next chapter ($\tau = 1/6.3$ s), and reasonable spontaneous activity built into the simulated system (the exponential random variable with rate parameter is set to $\nu = \tau/10$), we demonstrate near-perfect recovery of the ground truth effective connectivity matrices.

6.1 In silico Experiment Setup

A key consideration in actual experiments is that neurons outside the field-of-view (FOV), which can neither be controlled nor imaged, will likely influence the neurons within the FOV through their spontaneous activity and potentially through polysynaptic effects. Although much of cortical connectivity is local, and in real experiments, we attempt to minimize the number of these neurons through large FOVs and dense imaging, some neurons will always evade direct observation (such as those from deeper brain structure or distant brain areas) [40]. Computationally, we treat these inaccessible neurons as unknown sources in our model. To investigate the impact of this uncertainty, we simulated the setting where we stimulate and observe only a partial region (the center field-of-view marked by the green square) of the defined circuit while some neurons are deliberately excluded. This setup is illustrated in Figure 6.1 below, which shows the ground truth connectivity represented in one photostimulus matrix S. We performed in silico 2P imaging and photostimulation in a constrained region containing 129 neurons while the system as whole contained 500 neurons. As described above, these neurons were placed randomly in the field-of-view with connection strengths and sign determined with statistics grounded in what is known in mouse V1. Algorithm 3 detailed below was used to generate the ground truth system to be recovered using our estimation algorithm.

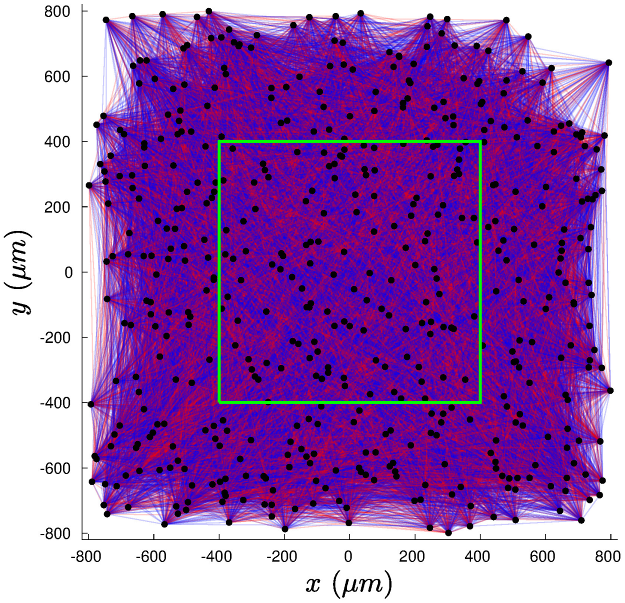

Figure 6.1: In silico experiment setup where observations and photostimulations are limited to the neurons residing in the center field-of-view denoted by the green square. All neurons in the system, including those outside the field-of-view, determine the observed dynamics. The neurons outside this field-of-view are considered latent, and are treated as noise sources.

6.2 Results

Despite the limited field-of-view and the additional unknown sources simulated as coming from neurons outside the FOV, the reconstruction algorithm correctly recovered the ground truth model parameters when compared with the ground truth. The results are shown in Figure 6.2. As technology advances, we will address these "non-observed neurons" by scaling up our imaging and perturbation capabilities, primarily by increasing our field of view via a 2P holographic mesoscope and fast remote focusing to acquire more z-planes of the volume.

Algorithm 3: Ground truth connectivity matrices generation algorithm

```
Input  : the number of neurons $N$ and the sparsity level $k$.
Output: $G_i$s ans $S_j$s.
Initialize: $i = 0$.
Randomly generate $x$- and $y$-coordinate positions of every neuron by sampling from
  the uniform distribution.
while $i < k$ do
    1. Randomly choose two neurons and calculate distance.
    2. With probability calculated from $\Gamma(\alpha, \beta)$, establish a connection.
    3. Determine connection strength that is proportional to $\frac{1}{\text{distance}}$.
    4. $i = i + 1$.
end
Form $G$.
```

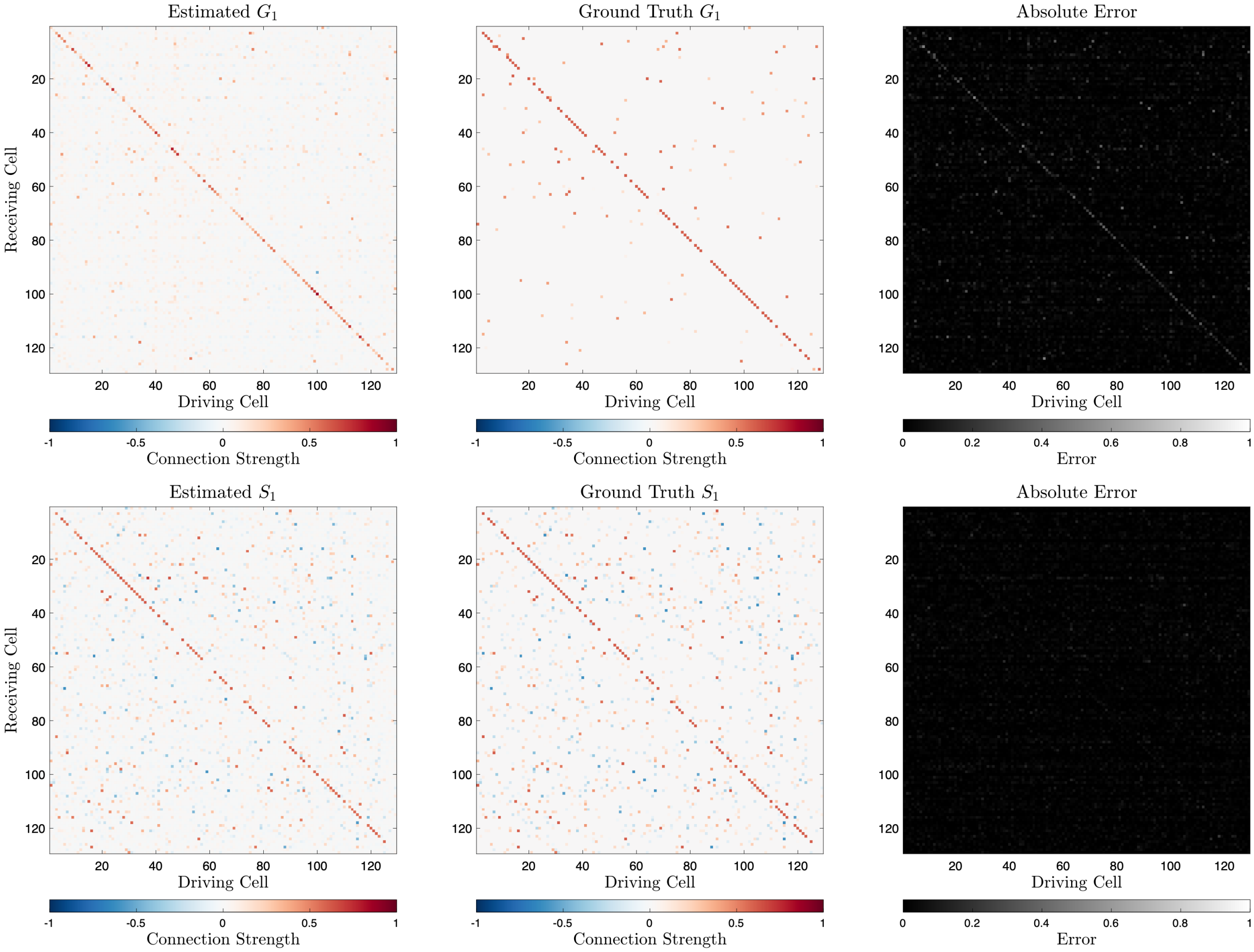

Figure 6.2: In silico limited field-of-view experiment results. Estimated, ground truth, and absolute error matrices for both the selected autoregressive and photostimulus matrices are shown.

Chapter 7

In vivo Results and Validations

We test and validate the photo-stimulation method for circuit estimation in V1 of both awake and anesthetized mice. To establish the feasibility of our approach in vivo, we conducted 3D-SHOT photo-stimulation and recording of the mouse primary visual cortex (V1) neurons in L2/3. We used transgenic mice expressing GCaMP6s in all forebrain excitatory neurons (camk2a-tTA;tetO-GCaMP6s) [46], and expressed microbial opsins via adeno-associated viruses (AAV) injected intravenously using PhP.eB. This expression paradigm helps ensure even, stable and widespread co-expression of the calcium sensor and opsin in the great majority of cortical excitatory neurons. In the future, we plan to leverage newly developed transgenic lines for ChroME2.0 and GCaMP8. In some experiments, we plan to additionally label and photo-stimulate GABAergic interneurons by co-injecting an AAV driving ChroME fused to a blue fluorescent protein and co-expressing GCaMP with the mDlx enhancer (AAV-mDlx-ChroME-tagBFP2-p2A-GCaMP8f) [15]. Trials with excessive brain motion, high run speed, and low stimulation success rate, which all could lead to spatially mistargeted photo-stimulation were discarded. In the same session, we characterized the visual responses properties of the neurons in the volume, especially their tuning to orientation with conventional drifting grating stimuli. We use this physiological data to better interpret the AR model. In each experimental session we designed 100-120 unique ensembles of neurons for co-stimulation and repeated each ensemble photo-stimulation across 15-30 trials (total experimental time approximately 2-3 hours). In the future, we will scale up both the number of trials and the number of ensembles by optimizing experimental design and additionally aggregating data from the same network over days by precisely returning to the same imaging volume. This will also allow us to empirically validate our approach by computing photo-stimulation patterns offline (between experimental sessions), and then presenting those that are predicted to have specific effects on the network (such as net excitation or net inhibition). Finally, to maximize throughput we stimulated more neurons in each photo-stimulus by leveraging ChroME2.0 opsins which dramatically increased the number of L2/3 pyramidal neurons that we can simultaneously photo-stimulated. Such an increase in scale is necessary to probe more of the putative input network at higher rates, which will become particularly important as we include ever larger field-of-views. We collected data in

7 sessions from 3 mice virally expressing ChroME and transgenically expressing GCaMP6s. The virus also drives a separate nuclear red fluorescent protein (H2B-mRuby3), which serves to identify opsin expressing neurons and facilitates their automatic segmentation in vivo. We targeted opsin expression to excitatory neurons specifically (via Cre-dependent recombination in the emx1-IRES-Cre line). Adult mice of both sexes were prepared with standard cranial windows and mounted under a custom microscope equipped with both volumetric laser-scanning 2P imaging (Coherent Ultra II) and scanless volumetric holographic photostimulation (Coherent Monaco and a Meadowlark HD SLM).

7.1 Results

To generate the training data, we photo-stimulated randomly generated ensembles of 30 neurons throughout the 3D volume. Figure 7.1 canonically shows the mean data across the trials from one such experiment.

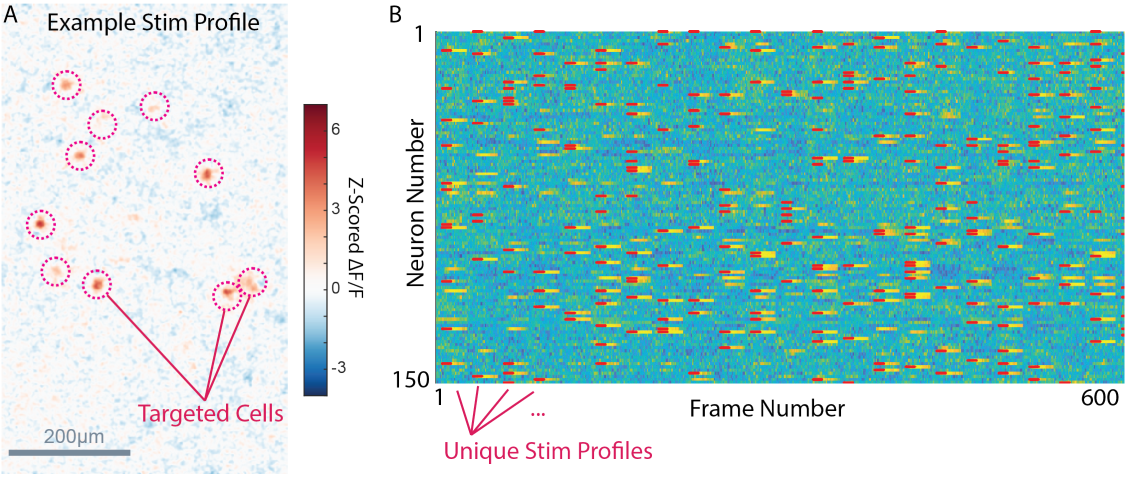

Figure 7.1: Across trial mean data matrix from one random stimulation experiment using 3D-SHOT with settings tailored to the AR model parameter estimation routine pipeline.

In each experiment, we tested 100 unique ensembles, resulting in 15 trials/ensemble. In the following particular experiment, we conducted the random stimulation experiment and performed calcium imaging in the same field-of-view comprised of the same set of neurons in the same mouse, but across two different brain states: awake and under mild concentration of chlorprothixene and isoflurane. The correlation matrices resulting from the training data from the two different brain states: under anesthesia and awake, are shown in Figure 7.2 and Figure 7.3, respectively.

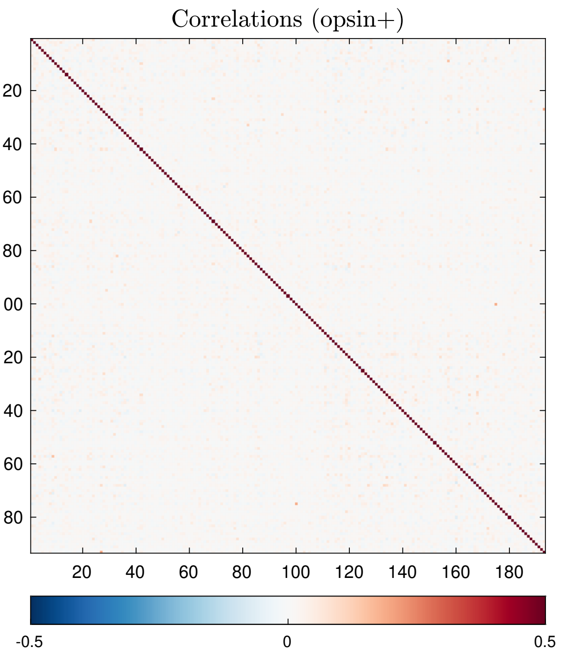

Figure 7.2: Correlation matrix calculated from the training data from mouse V1 under mild anesthesia.

The correlation matrices are symmetric and dense; and they do not provide any meaningful insight into the underlying model-driven circuit mechanism. We can however conclude that the mouse brain has changed drastically from the application of anesthesia.

The parameters of the autoregressive biophysical model: the autoregressive effective connectivity and photostimulus effect matrices among the stimulated, opsin-expressing cells corresponding to the mouse V1 under anesthesia and awake conditions are shown in Figure 7.4 and Figure 7.5, respectively.

Compared to simple correlation matrices, we see that the recovered matrices are sparse. The disappearing diagonal connection weights, which characterize a neuron's connection to itself, or the decay rate dynamic, for higher order Gs and Ss are consistent with the physics and the observed data. Higher order Gs, which characterize the effect of neurons' activity from farther back in history should be smaller in magnitude. Effects from photostimulation should also decay over time, which is reflected in higher order S matrix having less overall energy, especially in its diagonal weights. The photostimulation effects matrices Ss especially show sparsely populated driving cells that exhibit strong, directed connectivity influence that effectively drives the one specified receiving cell.

Another interesting observation is that there exist connections among the same pair of neurons that survive the application of anesthesia. While there are multiple evidences of brain "re-wiring," there do exist few strong excitatory connections that do not change

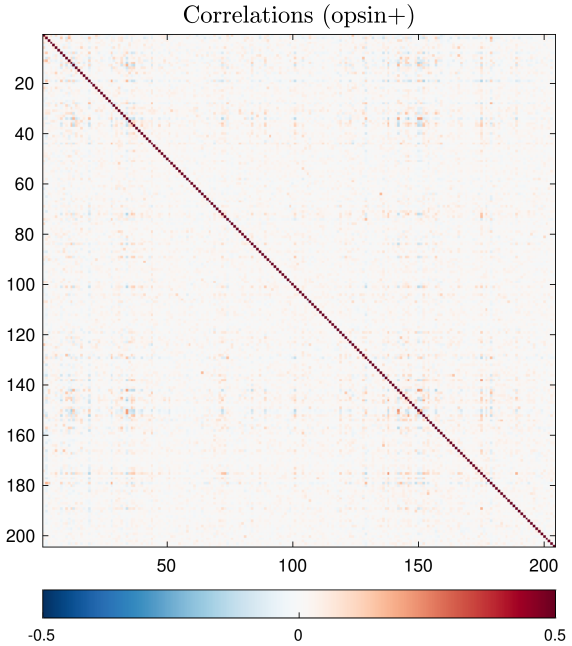

Figure 7.3: Correlation matrix calculated from the training data from awake mouse V1.

even though the brain state has changed extensively. It is also interesting to note that the matrices corresponding to the neurons under anesthesia have weaker weights and less significant connections overall. The autoregressive biophysical model obtained from in vivo experiments easily yield conclusive population-level signatures associated with specific brain states.

7.2 Cross-Validations

To test the reliability of the recovered model parameters, we perform cross-validations using the in vivo experimental data from the mouse V1 under anesthesia. We first split the entire experiment data in two approximately equal halves in chronological order. Direct comparison of the connection weights from G_0, G_1, S_0, and S_1 obtained from the first half of the experiment and those obtained from the last half of the experiment are shown in Figure 7.6 below.

We then split the entire experiment data by collecting odd-only and even-only trials. Direct comparison of the connection weights from G_0, G_1, S_0, and S_1 obtained from the odd-numbered trials of the experiment and those obtained from the even-numbered trials of the experiment are shown in Figure 7.7 below.

We see that despite the different sparsity hyperparameters for different splits of the experimental data, we obtain similar corresponding matrices in comparison. The slopes of the regression lines for the elements of the photostimulation effects matrices S_0 and S_1 calculated using the above cross-validation methods are not perfectly one. This difference may have been due to the system-level differences in the optimized sparsity hyperparameters during the estimation steps or simply insufficient number of trials to overcome the signal-to-noise ratio caused by significant spontaneous activity, animal behavior, and/or experiment conditions. They could also signify a biological constraint that we cannot ignore. Although these results show that both strong and weak connections recovered from different training set data exhibit consistent trends, there exist signs that the brain may have developed minor plasticity across the duration of the lengthy experiment. This finding emphasizes the need for shorter experiment session duration. Our estimation algorithm pipeline is nevertheless robust to slight variations and noise in experimental data.

7.3 Prediction

One of the strengths of our approach is our model's predictive power. To evaluate the AR model's predictive power, we held out test set data containing 25% of the entire experimental data. This data set was not used for training, i.e., the estimation algorithm did not see this data and only used the training data set comprised of 75% of the entire experimental data. With the ratio between the training and test sets set to 3:1, 80% of the test set stimuli were used to make up the validation data set. The validation data were used to tune, or optimize, the sparsity hyperparameters λ_G and λ_S via K-fold cross-validation, where K was set to the number of filtered trials. The remaining 20% of the test set data were used for the purposes of evaluating the AR model's predictive power. Figure 7.8 is the graphical illustration of how the training and test data sets were created using the in vivo experimental data.

Using the AR model with the model parameters fitted to experimental data, we can generate and predict network responses to specific optogenetic stimuli. We use the biophysical model with model parameters determined using training data from the mouse V1 under anesthesia to generate two versions of predictions for held-out test set random stimulations. First, we generate the response of the neural network purely from the model, starting from zero activity for every neuron under consideration. We call this prediction the generated response. Second, we test the predictive power of just the connectivity matrices by taking the past test set neural activity that the autoregressive model requires to generate predictions. Instead of generating these past observations the model requires from scratch, we apply the held-out past test set neural activity to generate predictions. We call this prediction the predicted response. The trial-to-trial generated, predicted, and observed network responses to the held-out test set stimulation profiles are shown in Figure 7.9 below.

Figure 7.10 below shows the generated response to one canonical example test set stimulation and compares the response to the mean observed response for comparison. We see that the biophysical model's prediction accurately matches the responses of the stimulated cells as well as cells that have not been directed stimulated. The model correctly predicts one strong indirect response, a response arising from connection rather than direct stimulation.

Furthermore, the generated network response from the shown test set stimulus show that the AR model accurately forecasts the relative excitation strengths of neurons that have been targeted for direct photostimulation. These observations demonstrate that the biophysical AR model can achieve substantial predictive power, and we can conclude that the model accurately predicts circuit dynamics to photo-stimulation data that were not used to train the model.

7.4 Jackknife Resampling

Results from the leave-one-out jackknife resampling across the trial dimension of in vivo experiment data suggest that the AR model accurately captures reliable connections among the neurons in the field-of-view. Complete Jackknife means and standard error measures of the parameters of the autoregressive biophysical model: the autoregressive effective connectivity and photostimulus effect matrices among all cells in the field-of-view corresponding to the mouse V1 under anesthesia and awake conditions are shown in Figure 7.11 and Figure 7.12, respectively. Only driving influences of stimulated cells can be recovered using our model-driven estimation approach in the photostimulation effects matrices S_js. Nevertheless, the effective driving connections from the stimulated cells to all cells (entire columns of S_js) can be computed.

In both the anesthesia and awake conditions, the Jackknife mean estimates of the effective connectivity matrices shows that both strong excitatory and inhibitory connections are associated with low corresponding standard error metrics. Therefore, we can conclude that reliable, meaningful connections can be procured using our AR model and the associated estimation algorithm pipeline described in the previous chapters.

7.5 Biological Validations

We further validate the connections we recovered using the model-driven estimation approach by comparing the connection features to what is known in the V1 circuitry. We expect connection probability, and thus connection weights to decrease with increasing distance between each pair of neurons in the V1 circuitry [40]. We also expect neurons sharing similar tuning properties, orientation tuning specifically, to share stronger connection strengths we know connection probability is higher for neurons with the same preferred orientation [22, 23]. Because much of the literature on the mouse V1 circuit properties among excitatory L2/3 pyramidal cells focuses on connection probabilities, connection strengths may exhibit different patterns. However, we expect connection strengths and probabilities to be correlated.

Distance

Figure 7.13 and Figure 7.14 show the calculated excitatory and inhibitory connectivity weights of G_0, G_1, S_0, and S_1 obtained using experimental data from mouse V1 under anesthesia as a function of distance between the connected neurons, respectively. Diagonal connectivity weights, i.e., a neuron's influence onto itself, have been discarded. We generally see a monotonically decreasing trend for both excitatory and inhibitory connection weights for all matrices.

Figure 7.15 and Figure 7.16 show the calculated excitatory and inhibitory connectivity weights of G_0, G_1, S_0, and S_1 obtained using experimental data from awake mouse V1 as a function of distance between the connected neurons, respectively. Diagonal connectivity weights, i.e., a neuron's influence onto itself, have been omitted. We generally see a monotonically decreasing trend for both excitatory and inhibitory connection weights for all matrices.

The patterns we observe are consistent with our expectations. The high standard error metrics associated with connection strengths between neurons placed at great distances from one another are due to small sample size. Despite insufficient number of data points, the distance-based validations above provide a powerful physiological support that the model parameters we estimate are biologically meaningful.

Orientation Tuning

It is difficult to conclude that neurons that share orientation tuning preference share a higher degree of interconnectivity due to the small number of "tuned" neurons that have also been stimulated. It is known that in the mouse V1, similarly tuned cells, i.e., cells that respond to specifically oriented visual grating stimuli with statistical significance, share higher connection probability [22, 23]. This trend is not immediately obvious in both our awake mouse V1 and anesthetized mouse V1 data. In the particular mouse used to generate the above results, only 15 cells in the field-of-view were both stimulated and orientation tuned. Therefore, data from multiple experiments across multiple mice are needed to make meaningful claims and conclusions regarding orientation tuning selectivity and effective connectivity strengths. Figure 7.17 and Figure 7.18 show the excitatory and inhibitory connection weights, with the diagonal weights excluded, of orientation tuned and stimulated neurons obtained from mouse V1 under anesthesia and awake mouse V1, respectively, as a function of degree difference in preferred orientation.

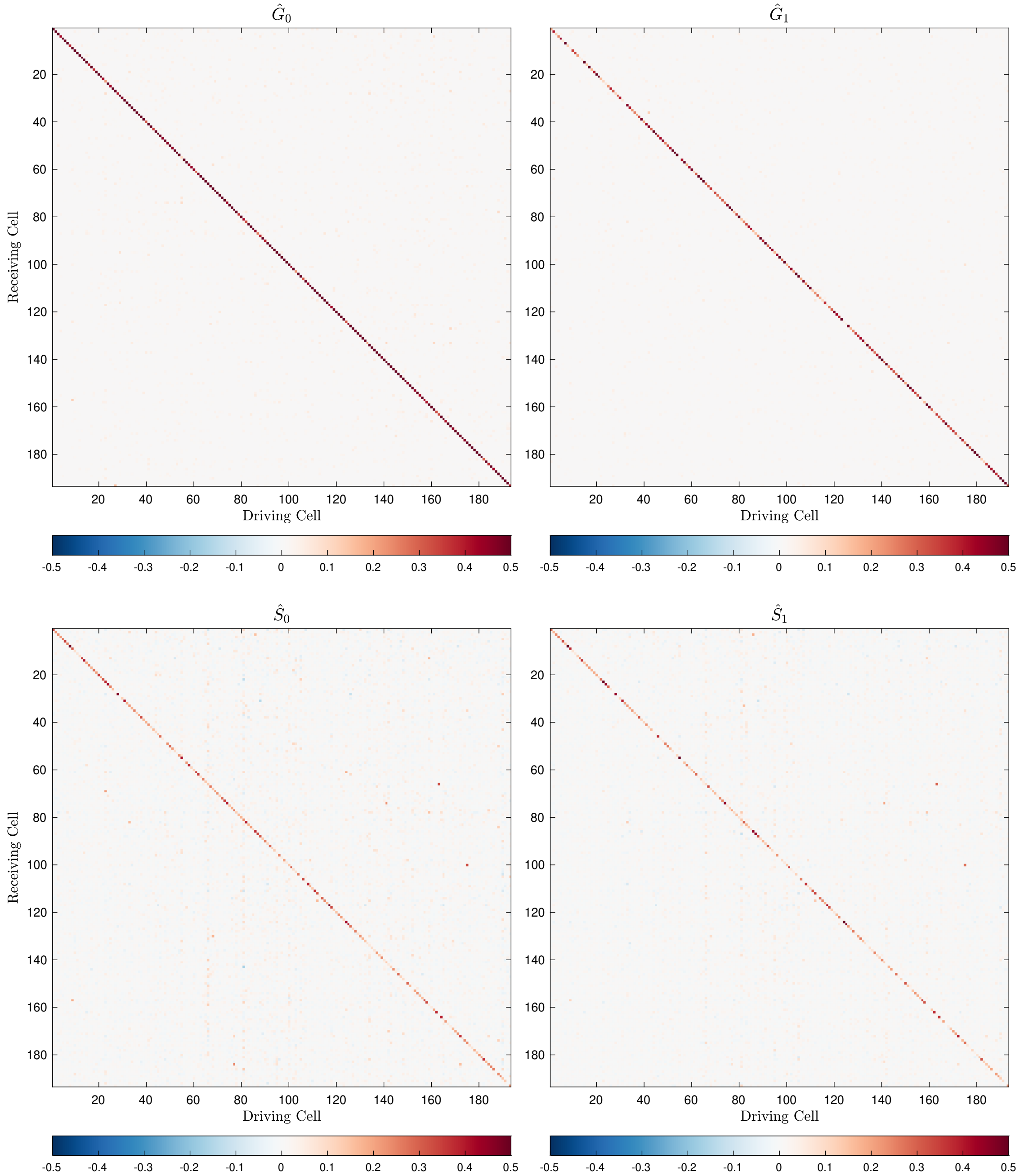

Figure 7.4: Autoregressive effective connectivity and photostimulus effect matrices among the stimulated cells recovered from random stimulation experiment in mouse V1 under anesthesia.

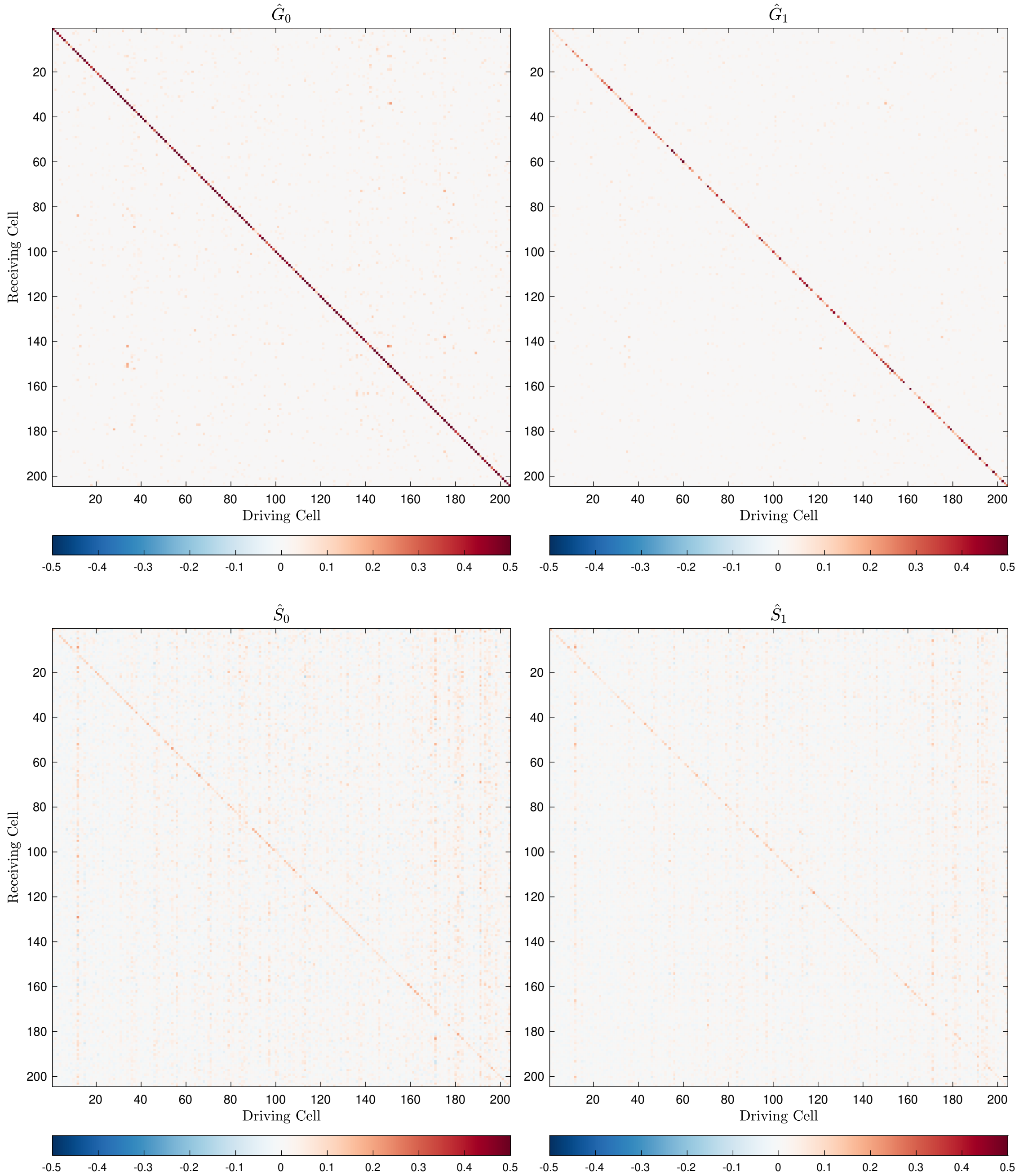

Figure 7.5: Autoregressive effective connectivity and photostimulus effect matrices among the stimulated cells recovered from random stimulation experiment in awake mouse V1.

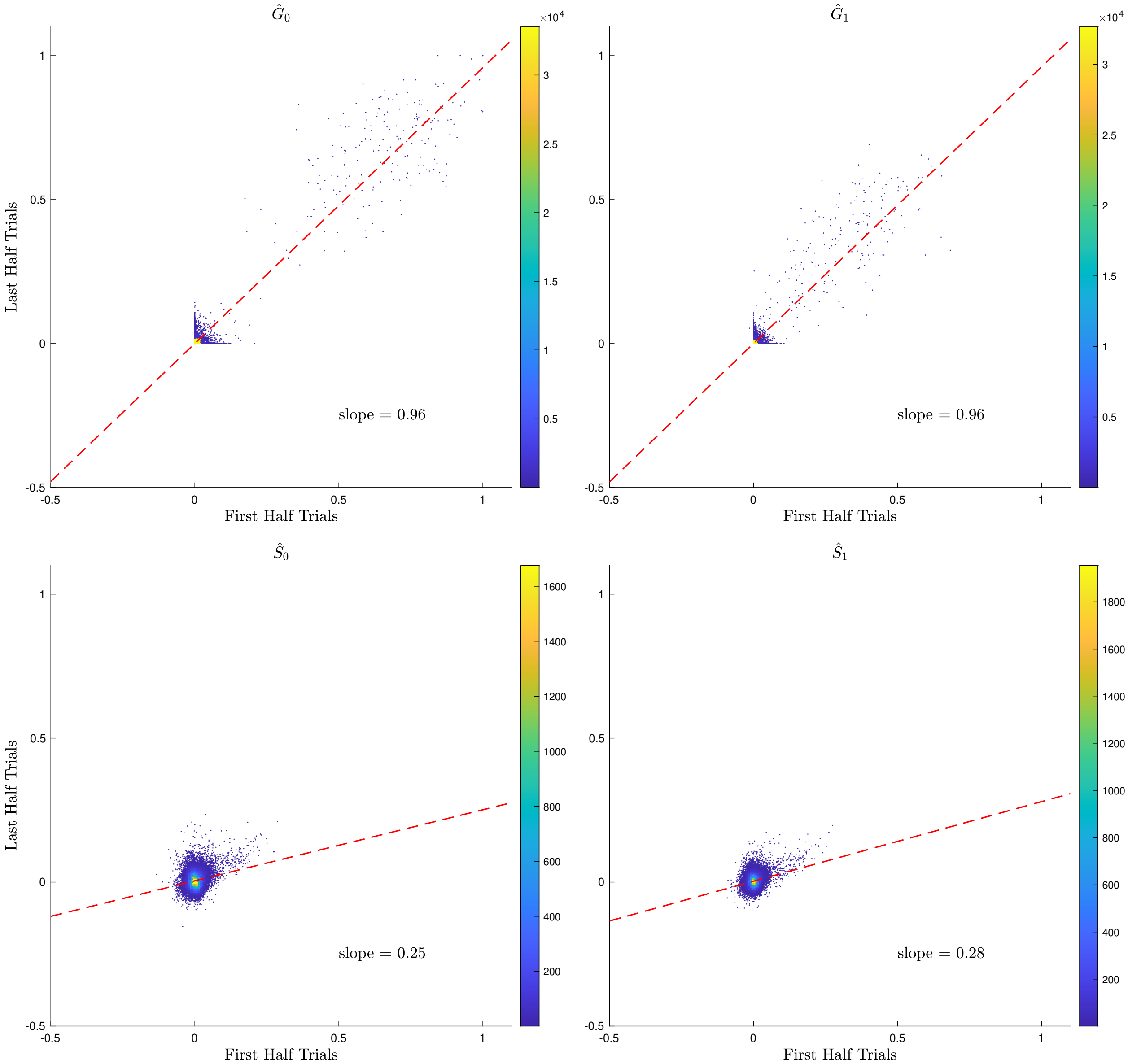

Figure 7.6: Heatscatter comparison plot of the autoregressive model connectivity matrices weights arising from two different (first and last) halves of the experiment.

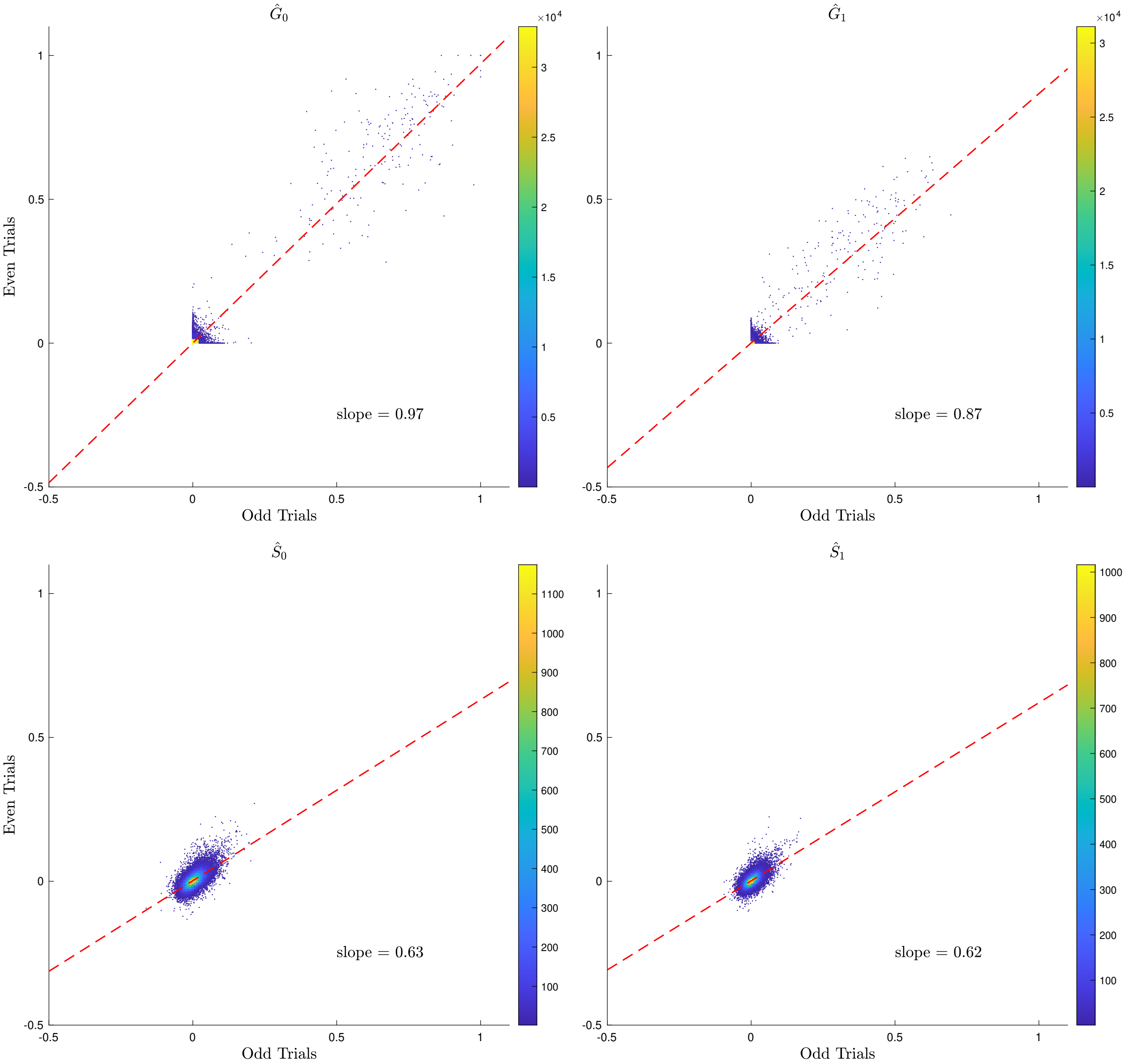

Figure 7.7: Heatscatter comparison of the autoregressive model connectivity matrices weights arising from odd and even trials of the experiment.

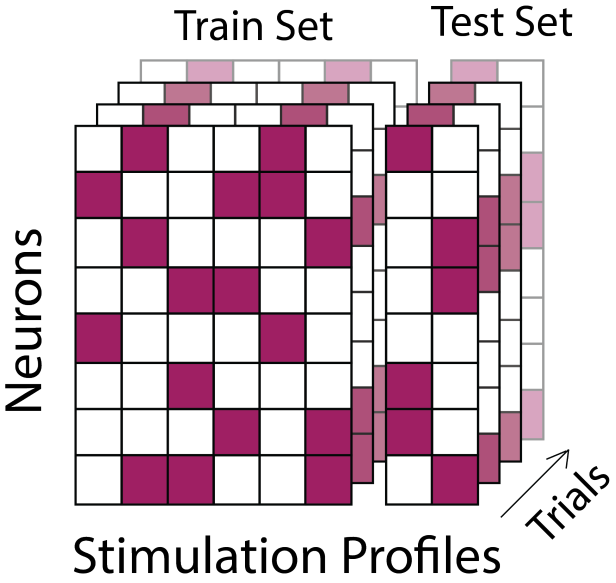

Figure 7.8: Illustration of the in vivo data set arising from one experiment. The ratio between the training and test sets was 3:1. Of the test set data, 80% of the test set stimuli were used as validation data sets, which were used to tune the sparsity hyperparameters. The remaining 20% of the test set data were used for the purposes of evaluating the AR model's predictive power.

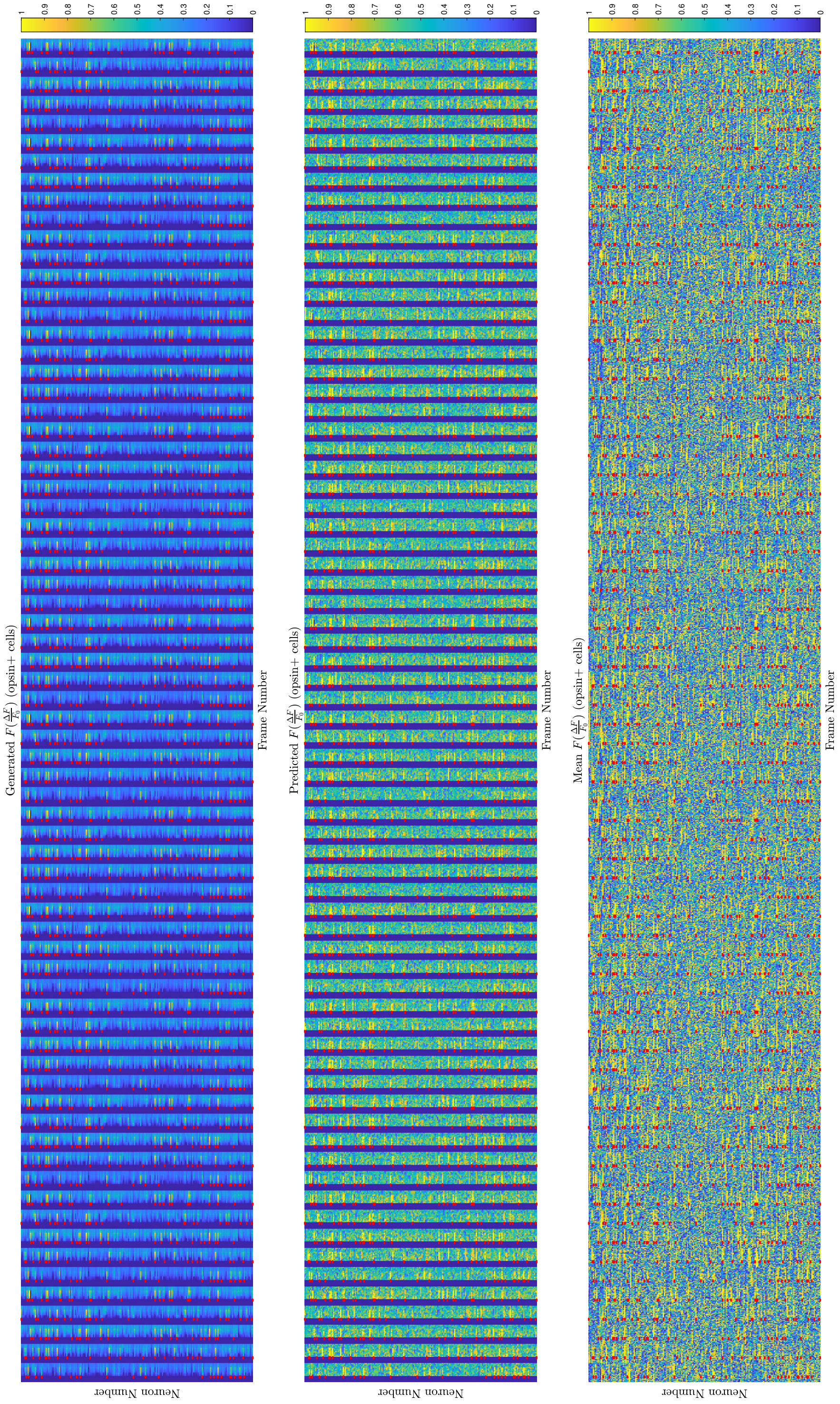

Figure 7.9: Trial-by-trial generated, predicted, and observed neural network responses to held-out test set stimulations.

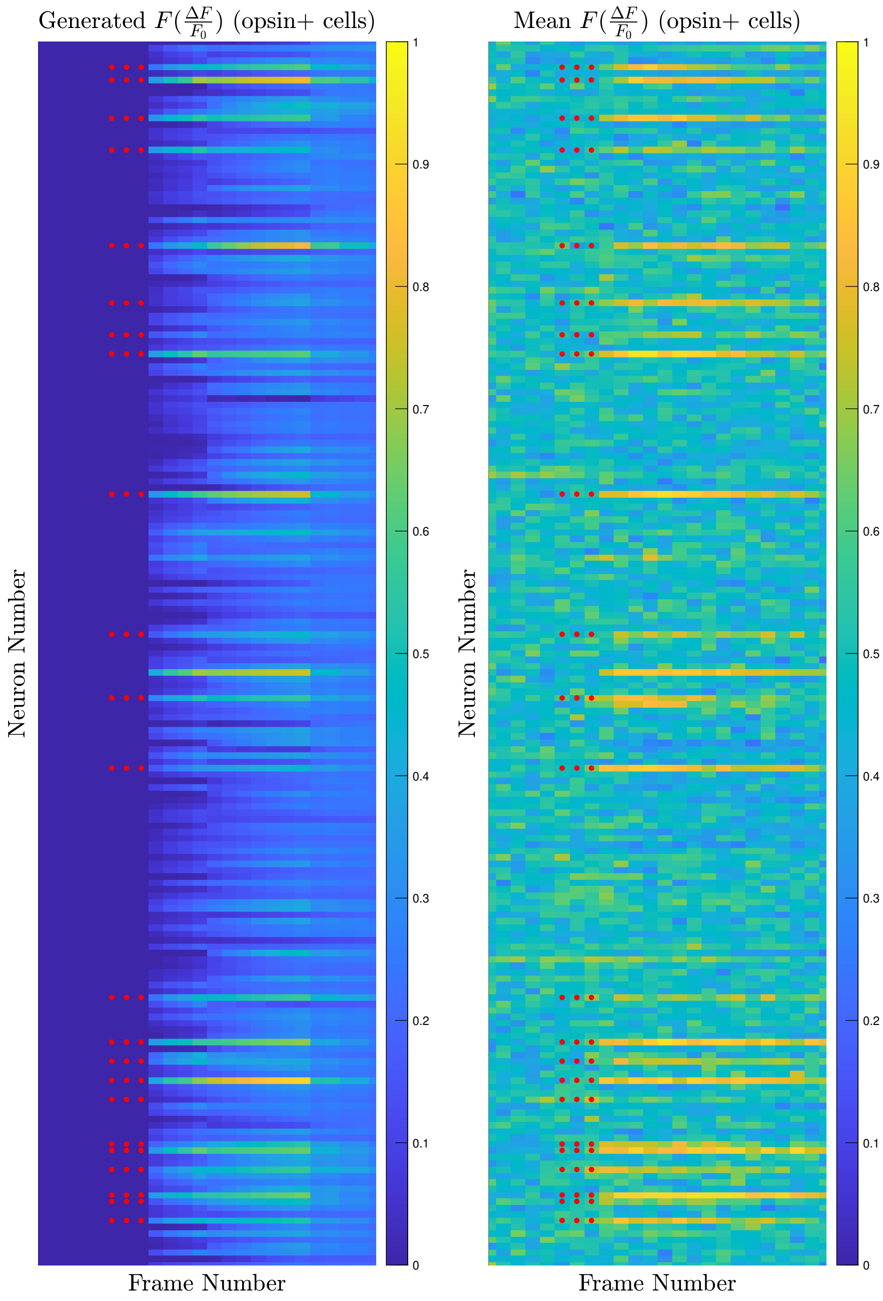

Figure 7.10: Comparison between the biophysical model generated response and the observed mean response to one canonical example test set stimulation. The model generated network response correctly predicts the activation of neurons that were not targeted for photostimulation. The relative strengths of excitation from photostimulation are also accurately forecasted.

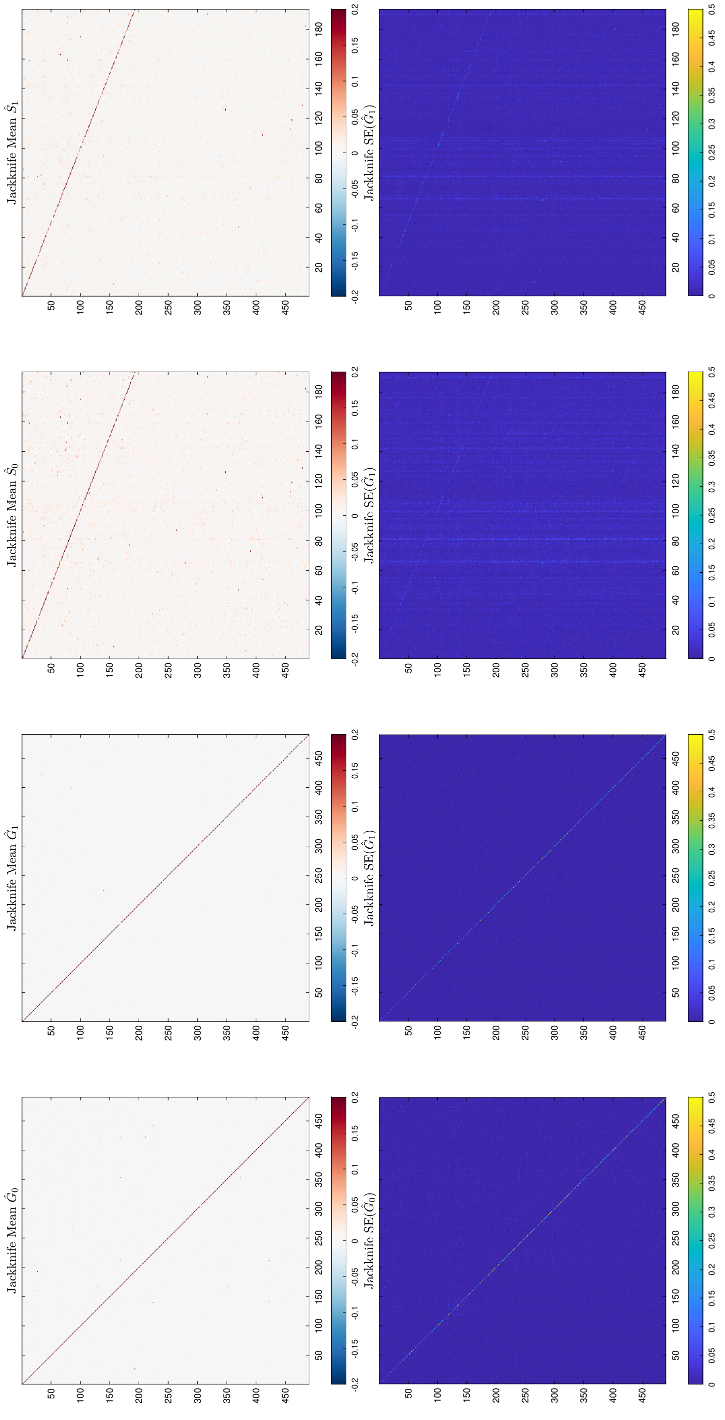

Figure 7.11: Autoregressive effective connectivity and photostimulus effect matrices among all cells recovered from random stimulation experiment in mouse V1 under anesthesia.

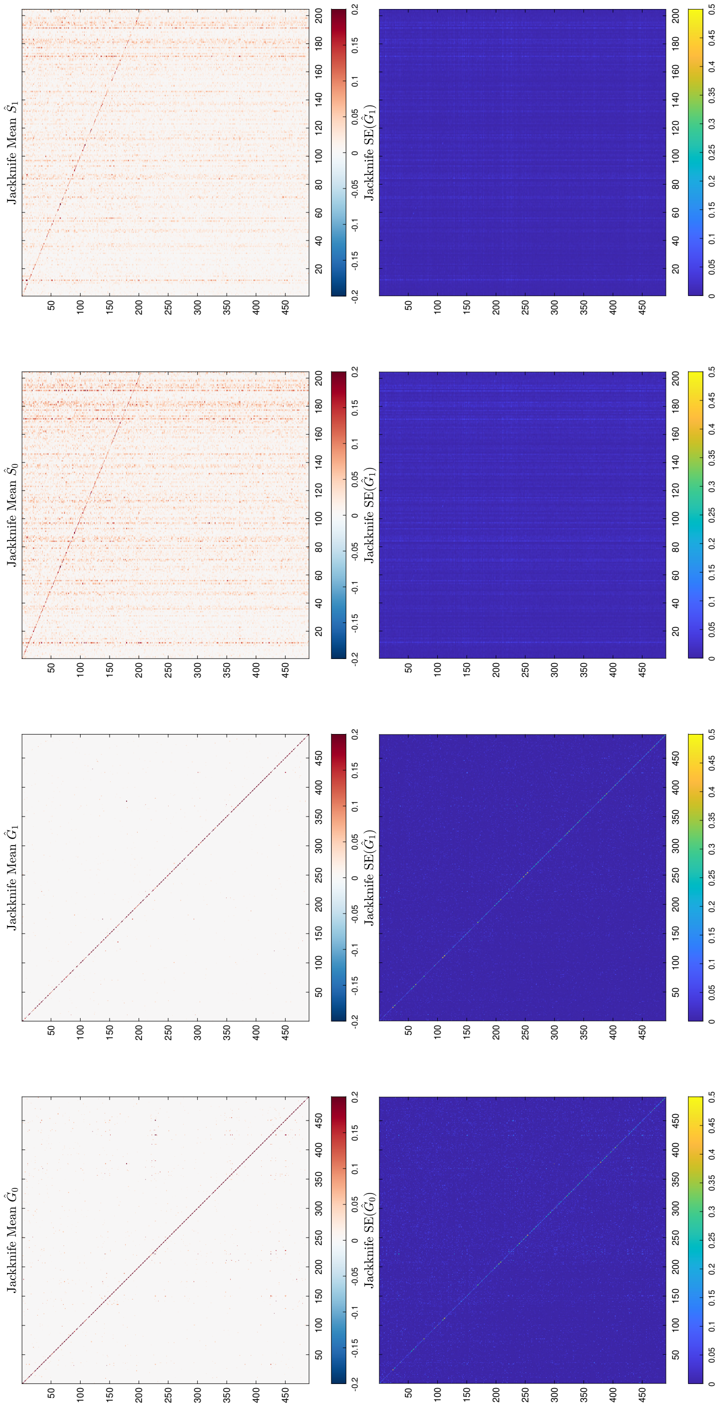

Figure 7.12: Autoregressive effective connectivity and photostimulus effect matrices among all cells recovered from random stimulation experiment in awake mouse V1.

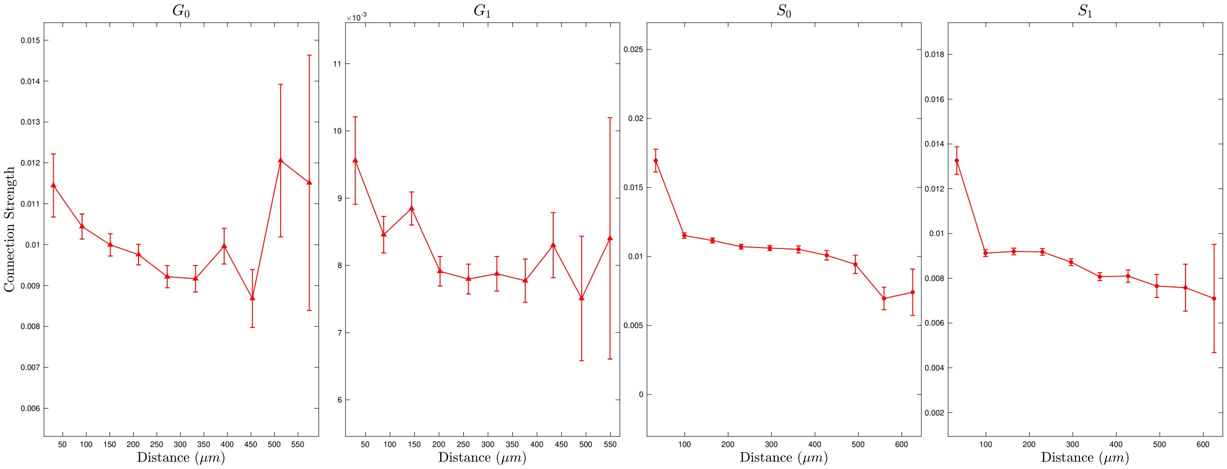

Figure 7.13: Excitatory connection weights of G_0, G_1, S_0, and S_1 obtained using experimental data from mouse V1 under anesthesia as a function of distance between the connected neurons.

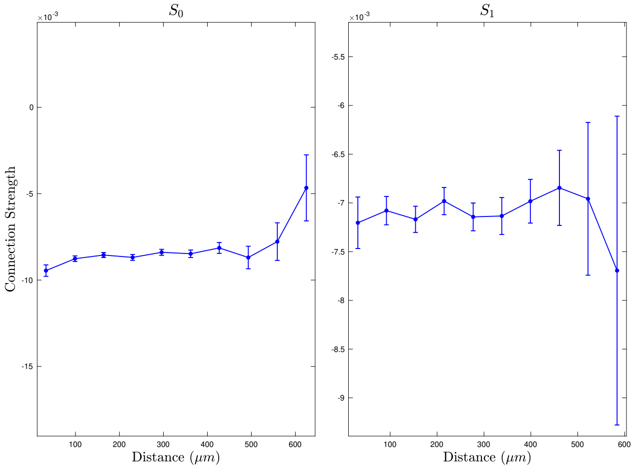

Figure 7.14: Inhibitory connection weights of S_0 and S_1 obtained using experimental data from mouse V1 under anesthesia as a function of distance between the connected neurons.

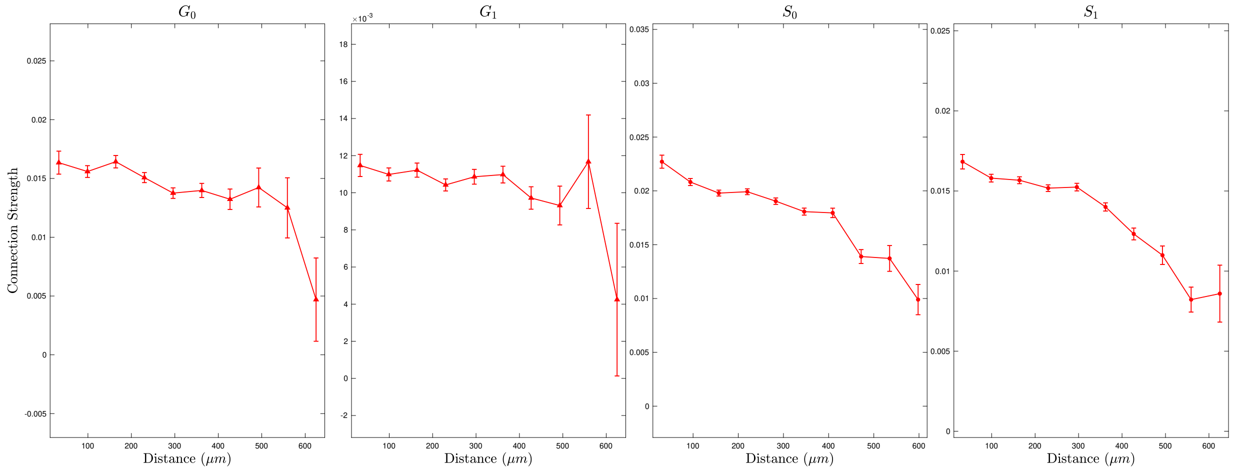

Figure 7.15: Excitatory connection weights of G_0, G_1, S_0, and S_1 obtained using experimental data from awake mouse V1 as a function of distance between the connected neurons.

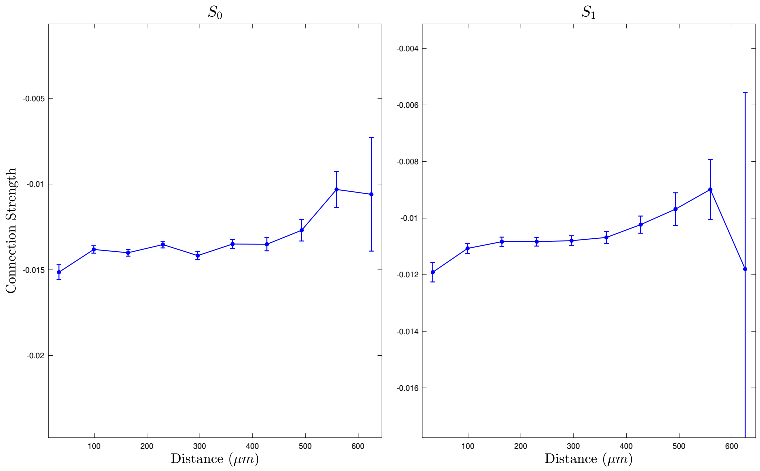

Figure 7.16: Inhibitory connection weights of S_0 and S_1 obtained using experimental data from awake mouse V1 as a function of distance between the connected neurons.

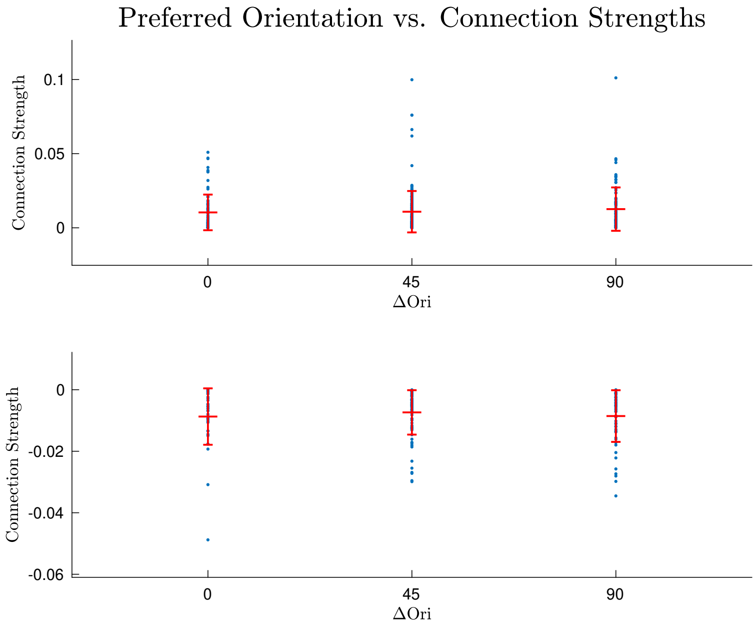

Figure 7.17: Excitatory and inhibitory connection weights of orientation tuned neurons obtained from mouse V1 under anesthesia as a function of degree difference in preferred orientation.

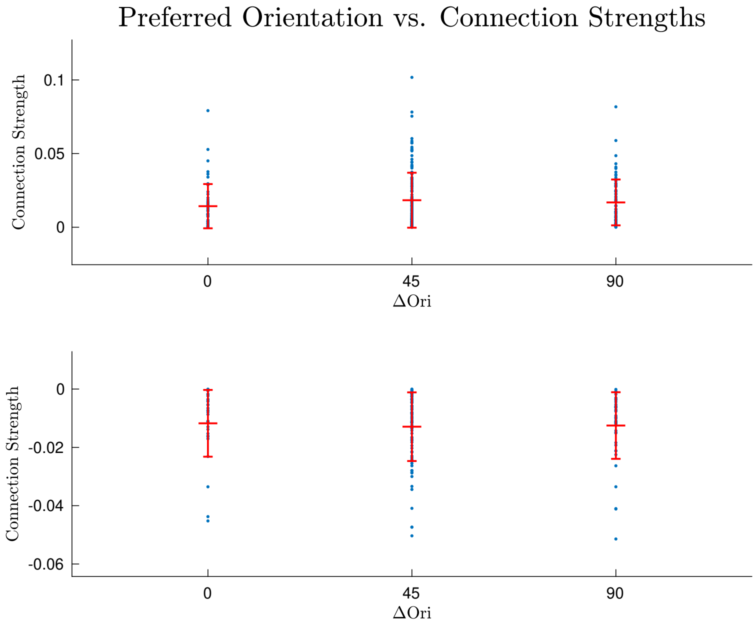

Figure 7.18: Excitatory and inhibitory connection weights of orientation tuned neurons obtained from awake mouse V1 as a function of degree difference in preferred orientation.

Chapter 8

Summary Statistics

The objective of optogenetics experimentation here is to estimate the effective connectivity of the neural network inside one region of the brain. We described an algorithm that allows us to estimate the neural network under investigation within a field-of-view using optogenetics. Building on these estimated connections and applying the ideas in interference effects from experimental design, we now present a framework for estimating the individual direct, as well as indirect (spillover), causal effect of selected neurons as higher-fidelity effective connectivity metrics.

8.1 Exposure Mapping

We apply the formulation introduced in [1] and adapt the exposure mapping framework to fit our setting. Recall U, the finite population of neurons indexed by $n = 1, \dots, N$. Randomized stimulation according to the aforementioned sensing matrix is performed on U. Recall the treatment assignment vector $d := \begin{bmatrix} d_1 & \cdots & d_N \end{bmatrix}^\top$, which is the stimulation profile, where $d_n \in \{0, 1\}$. We restricted treatment assignments to the support $\Omega := \{n : \text{Prob}(d_n) > 0\}$, i.e. we direct our attention towards opsin-positive neurons that are available for stimulation. Focusing on one autoregressive effective connectivity matrix G, the exposure mapping is the function

$$\mathcal{M} : \Omega \times \Theta^{(e)} \times \Theta^{(i)} \mapsto \Delta, \tag{8.1}$$

where Δ is the set of all possible exposures d_k; and

$$\theta_n^{(e)} \in \Theta^{(e)} = \begin{bmatrix} \theta_{n1}^{(e)} & \cdots & \theta_{nN}^{(e)} \end{bmatrix}^\top, \tag{8.2}$$

$\theta_{nj}^{(e)} = \mathbb{I}(g_{nj} > 0)$, and

$$\theta_n^{(i)} \in \Theta^{(i)} = \begin{bmatrix} \theta_{n1}^{(i)} & \cdots & \theta_{nN}^{(i)} \end{bmatrix}^\top, \tag{8.3}$$

$\theta_{nj}^{(i)} = \mathbb{I}(g_{nj} < 0)$, are respectively the unit-specific excitatory and inhibitory connectivity traits estimated using the aforementioned compressed sensing algorithm. Of course, we can choose to only consider strong excitatory and inhibitory connections of neuron n by defining $\theta_{nj}^{(e)} := \mathbb{I}(g_{nj} > \phi)$ and $\theta_{nj}^{(i)} := \mathbb{I}(g_{nj} < -\phi)$, where ϕ satisfies $0 < \phi < 1$.

We can view the exposure d_k of neuron n as changes in the standardized $\frac{\Delta F}{F_0}$, $F\left(y_n\right)$. Then, the exposure mapping can be written

$$\mathcal{M}(d, \theta_n^{(e)}, \theta_n^{(i)}) = \begin{cases} \text{(Direct + Indirect Excitatory \& Inhibitory Exposure)} \\ \delta_{111} : d_n \mathbb{I}(d^\top \theta_n^{(e)} > 0)\mathbb{I}(d^\top \theta_n^{(i)} > 0) = 1 \\ \text{(Direct + Indirect Excitatory Exposure)} \\ \delta_{110} : d_n \mathbb{I}(d^\top \theta_n^{(e)} > 0)\mathbb{I}(d^\top \theta_n^{(i)} = 0) = 1 \\ \text{(Direct + Indirect Inhibitory Exposure)} \\ \delta_{101} : d_n \mathbb{I}(d^\top \theta_n^{(e)} = 0)\mathbb{I}(d^\top \theta_n^{(i)} > 0) = 1 \\ \text{(Isolated Direct Exposure)} \\ \delta_{100} : d_n \mathbb{I}(d^\top \theta_n^{(e)} = 0)\mathbb{I}(d^\top \theta_n^{(i)} = 0) = 1 \\ \text{(Indirect Excitatory \& Inhibitory Exposure)} \\ \delta_{011} : (1 - d_n)\mathbb{I}(d^\top \theta_n^{(e)} > 0)\mathbb{I}(d^\top \theta_n^{(i)} > 0) = 1 \\ \text{(Indirect Excitatory Exposure)} \\ \delta_{010} : (1 - d_n)\mathbb{I}(d^\top \theta_n^{(e)} > 0)\mathbb{I}(d^\top \theta_n^{(i)} = 0) = 1 \\ \text{(Indirect Inhibitory Exposure)} \\ \delta_{001} : (1 - d_n)\mathbb{I}(d^\top \theta_n^{(e)} = 0)\mathbb{I}(d^\top \theta_n^{(i)} > 0) = 1 \\ \text{(No Exposure)} \\ \delta_{000} : (1 - d_n)\mathbb{I}(d^\top \theta_n^{(e)} = 0)\mathbb{I}(d^\top \theta_n^{(i)} = 0) = 1 \end{cases} . \tag{8.4}$$

Estimands of Interest

We are particularly interested in estimating the following interference effects for selected neurons n:

- (Isolated Direct Effect of Treatment on Neuron n)
  $\mathbb{E}\left[y_n(\delta_{100}) - y_n(\delta_{000})\right]$,
- (Spillover Effect from Excitatory and Inhibitory Connections to n)
  $\mathbb{E}\left[y_n(\delta_{011}) - y_n(\delta_{000})\right]$,
- (Spillover Effect from Excitatory Connections to Neuron n)
  $\mathbb{E}\left[y_n(\delta_{010}) - y_n(\delta_{000})\right]$,
- (Spillover Effect from Inhibitory Connections to Neuron n)
  $\mathbb{E}\left[y_n(\delta_{001}) - y_n(\delta_{000})\right]$,

where the expectation can be estimated using the sample average taken over the trials. The aforementioned estimands: isolated direct effect of treatment, spillover effect from entire connections, spillover effect from excitatory connections, and spillover effect from inhibitory connections for selected neurons n provide additional insights into the effective connectivity of the neurons, and are more robust to random biological, physical, and cross-trial variations. These estimands therefore serve as higher-fidelity summary statistics that build on the more crude estimates of the effective connectivity to provide more robust estimates of connectivity.

Chapter 9

Closed-loop Control

The ability to modulate circuit dynamics in a defined manner is extremely useful to test candidate mechanisms governing cortical activity patterns and their link to modulation of behavioral outcomes. The AR model yields optimized strategies to efficiently modulate circuit dynamics with optogenetic stimulation, which demonstrates the model's utility and the estimated connectivity matrices' worths. The effective connectivity matrices, especially the photostimulation effects matrices can be used to determine the optimal photo-stimulation patterns to generate desired circuit dynamics, such as driving the network into specific activity states. Possible target dynamics include recreating sensory evoked patterns, amplifying certain sub-networks of neurons while suppressing others, and deliberately generating maximal population activity levels or deliberately generating maximal suppression of the network under study. We focus on the latter problem in the next few sections.

We can formulate the question as solving an unconstrained inverse problem based on the learned AR model using stochastic control and optimization. We can consider a control cost function based on a descriptive statistics of the neural activity including GCaMP response or firing rate over a given time period. The control input variables are stimulation patterns over space and time both with and without sparsity constraints. By generating and testing such predictions empirically (in the same mice) we can reveal basic mechanisms that drive circuit dynamics. For example, there may exist "hub-like" neurons that play outsize roles in cortical dynamics by exhibiting higher levels of interconnectivity or especially strong connections strengths. There may exist strong evidence for highly recurrently connected subnetworks that amplify their own activity while suppressing the activity of others. By further relating these predicted and validated patterns to the visual response properties, spatial locations, and cell types of the neurons in the predicted photo-stimulus the AR model can provide insights into cortical coding strategies. One particularly simple example based on recent work [28] is that we expect the AR model to predict co-stimulating an ensemble of iso-oriented excitatory neurons, which should preferentially drive activity across other iso-oriented neurons. But a key power of this approach is that we can test many hypotheses in silico and then select those photo-stimulation paradigms for in vivo validation that are predicted to generate outsize effects on the network.

We present two different algorithms to determine the spatiotemporal stimulation patterns for controlling circuit dynamics. The most naïve approach is to attempt to control each and every neuron in the network through direct optogenetic stimulation. This approach however is neither the most efficient nor effective. Given the sheer number of neurons in any cortical network, the aforementioned approach is neither feasible nor informative. Instead, we rely on the AR model's predictive power. We can use the photostimulation effect matrix to find subsets of neurons that we can photo-stimulate to drive the network as a whole into the desired state very efficiently. By capitalizing on the recovered connectivity matrix to specifically target neuronal subsets that show highly levels of interconnectivity, we can achieve the desired objectives.

Prior approaches have used purely observational data to achieve a similar goal [8], but the AR model is far more flexible and appropriate as it relies on a physically meaningful understanding of network connectivity. Two simple closed-loop control algorithms to address the tasks of generating maximal population activity levels and maximal suppression of networks are outlined below. One utilizes a specific centrality measure from graph theory and another directly uses the AR model.

9.1 Ranked Katz Centrality Algorithm

The Katz centrality measure of a node in a graph describes the importance of that node in a graph. In our setting, a node is a neuron and the graph is the network we estimate, characterized by the connectivity matrix. Katz centrality of one neuron, defined by Equation 9.1 below, takes into account contributions from all adjacent neurons as well as the all neurons connected k hops away for every k.

$$C_K(x) = \sum_{k=1}^{\infty} \sum_{y \xrightarrow{w^k} x} w^k \alpha^k. \tag{9.1}$$

In practice, we limit k because neurons that are not adjacently connected do not influence the neuron under investigation. In simulated small-world networks with both excitatory and inhibitory connections, Katz centrality is quite predictive of the firing rates of neurons [18]. Using this fact, we can use the photostimulation effect matrix to calculate the Katz centrality measures of every neuron under investigation. We can then rank the cells based on their Katz centrality measures. With the limited laser power of the 3D-SHOT setup constraining the total number of cells we can photostimulate simultaneously, we can then select a hologram that includes as many cells as the 3D-SHOT system permits with the highest Katz centrality measures to drive the network to a maximally excited state, and select a hologram that includes as many cells as the system allows with the most negative Katz centrality measures to drive the network to a maximally suppressed state. The pseudocode of this approach is detailed in Algorithm 4 below.

```
Algorithm 4: Ranked Katz centrality algorithm
    Input : $S_0$, the lowest order photostimulation effects matrix, and $N_{\max}$, the
            maximum number of neurons in a hologram
    Output: $d^*$, the optimal hologram
    Initialize: $\alpha < \frac{1}{\lambda_{\max}(S_0)}$.
    for $n = 1, \ldots, N$ do
        Compute the Katz Centrality measure corresponding using the $n^{\text{th}}$ column:
            $$C_K(n) = \sum_{k=1}^{\infty} \sum_{i \xrightarrow{s_{in}^k} n} s_{in}^k \alpha^k$$
    end
    Rank the neurons according to their Katz centrality measures in
      ascending/descending order for maximally suppressed/excited brain state,
      respectively.
    Define the threshold $\gamma \coloneqq C_K(n^*)$, where $C_K(n^*)$ is the Katz Centrality measure of
      the neuron $n^*$ ranked at ${N_{\max}}^{\text{th}}$ place.
    Define $\mathcal{I} \coloneqq \{n \ : \ C_K(n) \geq \gamma\}$.
        $$d_n^* = \begin{cases} 1 & \text{if } n \in \mathcal{I} \\ 0 & \text{otherwise} \end{cases} \quad \forall \; n = 1, \ldots N.$$
```

9.2 AR Model Holograms

We can also use the AR model to predict photostimulation patterns that will maximize or minimize the overall network activity. In simulation, we can identify specific photostimulus patterns that are predicted to generate maximally or minimally excited states. Similar to the Ranked Katz Centrality Algorithm described above, we can rank neurons based on their abilities to excite or suppress the network from stimulations. We can then choose as many highest ranked neurons (according to our network-level objective) as the laser power allows. The selection of these neurons constitutes the optimal hologram that can achieve the objective of achieving the maximally excited or suppressed brain states.

9.3 Potential Problems

A practical concern is that the theoretically most informative stimulation patterns may not be experimentally practical due to the technical limitations on the number neurons that we can stimulate at a time. Nonetheless, the technical advances with more potent opsins and optimized photostimulation protocols will mitigate these concerns, and we can develop algorithms to determine the optimal stimulation patterns based on the AR model under specific experimental constraints mentioned above. The two algorithms detailed above computes highly efficient photostimulation patterns that can achieve the desired network dynamics while requiring stimulation of the fewest neurons. It remains as future work to test these “optimal” stimulation patterns in vivo to see if they produce the predicted circuit dynamics. It is probable that the mathematically optimal solution for driving the network into a specific activity state may not be what occurs biologically, or at least in a condition we can observe. Even so, we expect this approach to inform on candidate mechanisms that govern network activity states. More importantly, they should allow us to test hypotheses for what patterns may optimally alter behavior.

Chapter 10

Conclusion

Our explainable artificial intelligence approach combined behavior and neurophysiological data: the machine learning model we developed was interpretable due to its base in real biological and physical principles of the neurons and networks. Furthermore, the interpretability was greatly enhanced because of the model's sparsity, modular nature, and stimulatability. The sparsity originated from the sparse connectivity of neocortical neurons, which permitted our compressed sensing approach, greatly increasing throughput. The governing equations allowed us to simulate the circuit dynamics in response to arbitrary perturbations. Our approach was also modular as it was comprised of two independent processes: the circuit-to-dynamics model and the dynamics-to-behavior model, both of which can be tuned to any circuit or brain region by training with the appropriate data. Our model was therefore generalizable. This modular approach overcame a significant limitation of "black box" methods or post-hoc analysis of "black box" approaches that attempt to fit ensemble neural activity directly to behavior data.

10.1 Significance

The causal relationships between precise features of neural activity and behavior are largely unknown. Purely observational approaches can generate hypotheses but cannot test them. Conventional perturbation paradigms, such as microstimulation or optogenetics, can relate the activity of specific brain areas or cell types to a behavior, but typically cannot reveal the causal features of the underlying neural codes. Precisely patterned optogenetics can address this challenge but given the large number of possible perturbations it quickly becomes uncertain which perturbations will yield the most insight into the problem. Recent studies have stimulated ensembles of neurons based on their co-activation during a sensory task [28, 8, 7, 11], but the choice of such perturbations is biased by the narrow set of network states that are observed in a given experiment.

Machine learning and artificial intelligence techniques offer an alternative strategy to model a circuit and design perturbations in order to obtain key insight into neural circuit mechanisms of behavior. “Black box” machine learning approaches so far have generated limited insight [35, 17], only superficially categorizing and predicting neural and behavioral outcomes. Without knowing the biological and physical variables critical to a machine learning-based outcome, it is difficult to achieve a deep mechanistic understanding of brain functions and the pathophysiology of psychiatric disorders. Our proposed solution was to fuse machine learning and artificial intelligence techniques with biologically informed network models that are trained by large-scale patterned optogenetic perturbations in behaving animals. The new approaches use precise optogenetic perturbations in an unbiased manner to train an autoregressive (AR) model of the neural circuit (mouse primary visual cortex, V1) that estimates a set of dynamic effective connectivity matrices of the neurons under study. We then leverage these matrices and the biophysical model to design “optimal” perturbations that make testable predictions about circuit connectivity, neural dynamics and ultimately behavior. Analysis of these impactful perturbations will yield insight into the underlying computations. They will highlight which specific neurons, connections, and spatial features of network activity can explain the neural processes that drive perception, cognition, or action. Our approach thus leveraged explainable artificial intelligence to advance our understanding of the neural codes underlying behavior.

10.2 Limitations

We detail some potential problems of our approach, and provide solutions and alternative approaches.

There is significant trial-to-trial variability in the neural activity of behaving mice with or without optogenetic stimulation. Even when a mouse repeats the same behavior experiments, the observed neural activity exhibits large variations. We hence do not expect that our model will be able to predict and control individual calcium traces perfectly.

Exact targeting in vivo is never perfect. Many factors affect the targeting of neurons, e.g., physiological point spread function, motion, etc. We minimize off-target effects with careful system calibration and exclude trials with significant motion (¿3 μm).

There exists a tradeoff between the accuracy of our estimates and speed. The accuracy of our estimates increases with more experiment (more unique stimulation profiles) and further enhances with more computation time as more algorithm iterations can be devoted for the parameter estimation task. Online and closed-loop schemes do not become feasible if time is not spent wisely. To circumvent the bottleneck in time for online and closed-loop experiments, we reduce the number of unique stimulations and do not run the algorithm iterations to convergence. This shortcuts ensure that the computation is fast enough for online experiments. We still expect to recover the strong connections even with these sacrifices.

GCaMP only indirectly reports action potentials. A crucial concern is whether our measurements taken via calcium imaging are accurate enough estimations of the underlying firing rates of neurons in the FOV. GCaMP6s reports spike rates approximately linearly across a broad range of spike frequencies [20, 26]. The calcium trace indicator’s low signal-to-noise compared to electrophysiology means that spike count estimations on individual trials are not as accurate. However, our model operates on the average calcium response across many repeats of individual photo-stimulation patterns, mitigating this concern.

The slow decay of GCaMP signals limits the number of unique stimulations we can perform per unit time, requiring about one second between trials to avoid ambiguity on the source of a change in a postsynaptic neuron's response. To address this in part, we expect to conduct most future experiments with GCaMP8m (or 'f'), which is much more sensitive and shows substantially faster dynamics than GCaMP6s [48]. If needed, we can maximize sampling rate from individual neurons albeit with fewer imaging planes (scanning rates up to 60 Hz albeit in one plane is possible). The faster rise and decay of GcaMP8m and its single spike sensitivity may allow us to employ deconvolution strategies to reconstruct spike trains of imaging neurons with much higher accuracy than has been achieved with prior GCaMP sensors. If and when genetically encoded voltage sensors that can meet our experimental requirements become available, we will explore 2P voltage imaging as an alternative to calcium.

Our photo-stimulation regimen could generate long-term, undesired synaptic plasticity. We addressed this nuisance by randomizing the stimulation patterns as much as possible, which should minimize the correlated activity that is known to be critical for inducing such plasticity.

External inputs are always a source of concern. By sampling as many of the putative inputs to the targeted neurons, we can minimize the impact external inputs have on our estimation procedure. The contributions from the remaining neurons that will necessarily lie outside the field-of-view will be treated as external contributions in the parameter estimation scheme detailed above. If the same local connections respond to external contributions consistently across multiple trials, the algorithm will recover the neural network's signature circuit dynamics associated with the brain state under investigation.

10.3 Future Work

One of the most impactful use cases of the methods described in this dissertation is to conduct the AR model-driven closed-loop estimation and control experiments online. There are situations when online and closed-loop circuit estimation and modulation are necessary and beneficial. For example, an offline computed stimulation patterns may not produce accurate dynamics due to an inaccurately estimated AR model or if plasticity develops during the experiments. The online approach presently under development aims to estimate the stimulation patterns adaptively. Using an iterative estimation scheme that performs the aforementioned model parameter estimation in a batch processing manner, the model parameters can be updated as new batches are acquired. For 1000 neurons, one cycle of closed-loop estimation and stimulation can be accomplished within 30 minutes during a behavior experiment using a single server with 4 NVIDIA TITAN Xp GPU and with cold start. Since our algorithm is inherently parallelized as estimation is done for each neuron independently, order of magnitude compute speed acceleration can readily be achieved with more servers or cloud computing. According to compressed sensing (CS) theory, random ensemble photostimulation is optimal. With high probability, the sensing matrix that arises from this stimulation pattern satisfies the restricted isometry property (RIP) condition guaranteeing unique recovery. Nonetheless, the initial selection of stimulation patterns may not be ideal for parameter estimation, e.g., experimental and computational effort may be wasted on recovering zeros in the connectivity matrix, i.e., the weights of disconnected neurons. We can optimize both the incoherence of the CS measurement matrix for optimal satisfaction of the RIP condition as well as the experiment and computation time necessary to estimate the model parameters. We will develop and test the following two approaches to improve the estimation. Approach 1: estimate the model parameters based on random stimulation profiles, and then refine the stimulation patterns based on the estimated connectivity. Use the refined stimulation profiles to estimate the model parameters. Approach 2: partially estimate the connectivity matrix and perform matrix completion by constructing a low-rank matrix without the diagonal elements [5, 4]. We expect the standard errors of the model parameters obtained from resampling procedures to decrease with Approach 1. With Approach 2, we expect the experiment as well as the computation time to decrease significantly.

Additionally, we hope to extend our approach to the mesoscale. By scaling our approach from the local level (within mouse V1) to include higher cortical areas downstream via a novel 2P holographic mesoscope, we can both train our AR model on much more of the relevant inputs and tests its predictions on the outputs of V1. We expect that modeling and identifying perturbations to V1 that optimally drive downstream cortical regions should facilitate the identification of activity patterns to potently impact behavior. A key consideration for the AR model is obtaining data on as many potential input neurons as possible to each neuron within the FOV. We plan to maximize this number in the future by sampling most of each L2/3's putative presynaptic cells in V1 by employing fast and dense volumetric imaging using a holographic 2P mesoscope that can simultaneously sample and stimulate retinotopically aligned presynaptic zones in higher visual areas. With these further technological and algorithmic developments, it is our hope that the AR model and the estimation pipeline can be leveraged to gain further insight into the field of systems neuroscience and to ultimately understand, predict, modulate, and possibly control the neural codes underlying animal percept and behavior.

Bibliography

[1] Peter M Aronow and Cyrus Samii. “Estimating average causal effects under general interference, with application to a social network experiment”. In: *The Annals of Applied Statistics* 11.4 (2017), pp. 1912–1947.

[2] Amir Beck and Marc Teboulle. “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”. In: *SIAM journal on imaging sciences* 2.1 (2009), pp. 183–202.

[3] Edward S Boyden et al. “Millisecond-timescale, genetically targeted optical control of neural activity”. In: *Nature neuroscience* 8.9 (2005), pp. 1263–1268.

[4] Emmanuel J Candès and Yaniv Plan. “Matrix completion with noise”. In: *Proceedings of the IEEE* 98.6 (2010), pp. 925–936.

[5] Emmanuel J Candès and Benjamin Recht. “Exact matrix completion via convex optimization”. In: *Foundations of Computational mathematics* 9.6 (2009), pp. 717–772.

[6] Emmanuel J Candès, Justin Romberg, and Terence Tao. “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information”. In: *IEEE Transactions on information theory* 52.2 (2006), pp. 489–509.

[7] Luis Carrillo-Reid et al. “Controlling visually guided behavior by holographic recalling of cortical ensembles”. In: *Cell* 178.2 (2019), pp. 447–457.

[8] Luis Carrillo-Reid et al. “Imprinting and recalling cortical ensembles”. In: *Science* 353.6300 (2016), pp. 691–694.

[9] I-Wen Chen, Eirini Papagiakoumou, and Valentina Emiliani. “Towards circuit optogenetics”. In: *Current opinion in neurobiology* 50 (2018), pp. 179–189.

[10] Tsai-Wen Chen et al. “Ultrasensitive fluorescent proteins for imaging neuronal activity”. In: *Nature* 499.7458 (2013), pp. 295–300.

[11] Selmaan N Chettih and Christopher D Harvey. “Single-neuron perturbations reveal feature-specific competition in V1”. In: *Nature* 567.7748 (2019), pp. 334–340.

[12] Kayvon Daie, Karel Svoboda, and Shaul Druckmann. “Targeted photostimulation uncovers circuit motifs supporting short-term memory”. In: *Nature Neuroscience* 24.2 (2021), pp. 259–265.

[13] Abhranil Das and Ila R Fiete. “Systematic errors in connectivity inferred from activity in strongly recurrent networks”. In: *Nature Neuroscience* 23.10 (2020), pp. 1286–1296.

[14] Karl Deisseroth. “Optogenetics: 10 years of microbial opsins in neuroscience”. In: *Nature neuroscience* 18.9 (2015), pp. 1213–1225.

[15] Jordane Dimidschstein et al. “A viral strategy for targeting and manipulating interneurons across vertebrate species”. In: *Nature neuroscience* 19.12 (2016), pp. 1743–1749.

[16] David L Donoho. “Compressed sensing”. In: *IEEE Transactions on information theory* 52.4 (2006), pp. 1289–1306.

[17] Jean-Marc Fellous et al. “Explainable artificial intelligence for neuroscience: Behavioral neurostimulation”. In: *Frontiers in neuroscience* 13 (2019), p. 1346.

[18] Jack McKay Fletcher and Thomas Wennekers. “From structure to activity: Using centrality measures to predict neuronal activity”. In: *International journal of neural systems* 28.02 (2018), p. 1750013.

[19] Andreas Holzinger et al. “Current advances, trends and challenges of machine learning and knowledge extraction: from machine learning to explainable AI”. In: *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer. 2018, pp. 1–8.

[20] Lawrence Huang et al. “Relationship between spiking activity and simultaneously recorded fluorescence signals in transgenic mice expressing GCaMP6”. In: *BioRxiv* (2019), p. 788802.

[21] Leo Katz. “A new status index derived from sociometric analysis”. In: *Psychometrika* 18.1 (1953), pp. 39–43.

[22] Ho Ko et al. “Functional specificity of local synaptic connections in neocortical networks”. In: *Nature* 473.7345 (2011), pp. 87–91.

[23] Ho Ko et al. “The emergence of functional microcircuits in visual cortex”. In: *Nature* 496.7443 (2013), pp. 96–100.

[24] Christian Ledig et al. “Photo-realistic single image super-resolution using a generative adversarial network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 4681–4690.

[25] Jingwei Liang, Tao Luo, and Carola-Bibiane Schönlieb. “Improving” Fast Iterative Shrinkage-Thresholding Algorithm”: Faster, Smarter and Greedier”. In: *arXiv preprint arXiv:1811.01430* (2018).

[26] Evan H Lyall et al. “Synthesis of higher order feature codes through stimulus-specific supra-linear summation”. In: *BioRxiv* (2020).

[27] Alan R Mardinly et al. “Precise multimodal optical control of neural ensemble activity”. In: *Nature neuroscience* 21.6 (2018), pp. 881–893.

[28] James H Marshel et al. "Cortical layer–specific critical dynamics triggering perception". In: *Science* 365.6453 (2019).

[29] Alexander Naka et al. "Complementary networks of cortical somatostatin interneurons enforce layer specific control". In: *Elife* 8 (2019), e43696.

[30] Adam M Packer et al. "Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo". In: *Nature methods* 12.2 (2015), pp. 140–146.

[31] Adam M Packer et al. "Two-photon optogenetics of dendritic spines and neural circuits". In: *Nature methods* 9.12 (2012), pp. 1202–1205.

[32] Shir Paluch-Siegler et al. "All-optical bidirectional neural interfacing using hybrid multiphoton holographic optogenetic stimulation". In: *Neurophotonics* 2.3 (2015), p. 031208.

[33] Eirini Papagiakoumou et al. "Scanless two-photon excitation of channelrhodopsin-2". In: *Nature methods* 7.10 (2010), pp. 848–854.

[34] Nicolas C Pégard et al. "Three-dimensional scanless holographic optogenetics with temporal focusing (3D-SHOT)". In: *Nature communications* 8.1 (2017), pp. 1–14.

[35] Blake A Richards et al. "A deep learning framework for neuroscience". In: *Nature neuroscience* 22.11 (2019), pp. 1761–1770.

[36] Yaniv Romano, Matteo Sesia, and Emmanuel Candès. "Deep knockoffs". In: *Journal of the American Statistical Association* 115.532 (2020), pp. 1861–1872.

[37] L Federico Rossi, Kenneth D Harris, and Matteo Carandini. "Spatial connectivity matches direction selectivity in visual cortex". In: *Nature* 588.7839 (2020), pp. 648–652.

[38] Hasim Sak, Andrew W Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling". In: (2014).

[39] Jürgen Schmidhuber, Sepp Hochreiter, et al. "Long short-term memory". In: *Neural Comput* 9.8 (1997), pp. 1735–1780.

[40] Stephanie C Seeman et al. "Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex". In: *Elife* 7 (2018), e37349.

[41] Savitha Sridharan et al. "High performance microbial opsins for spatially and temporally precise perturbations of large neuronal networks". In: *Biorxiv* (2021).

[42] Carsen Stringer et al. "High-dimensional geometry of population responses in visual cortex". In: *Nature* 571.7765 (2019), pp. 361–365.

[43] Carsen Stringer et al. "Spontaneous behaviors drive multidimensional, brainwide activity". In: *Science* 364.6437 (2019).

[44] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288.

[45] Mai-Anh T Vu et al. “A shared vision for machine learning in neuroscience”. In: *Journal of Neuroscience* 38.7 (2018), pp. 1601–1607.

[46] Joseph B Wekselblatt et al. “Large-scale imaging of cortical dynamics during sensory perception and behavior”. In: *Journal of neurophysiology* 115.6 (2016), pp. 2852–2866.

[47] Ian R Wickersham et al. “Monosynaptic restriction of transsynaptic tracing from single, genetically targeted neurons”. In: *Neuron* 53.5 (2007), pp. 639–647.

[48] Y Zhang et al. “jGCaMP8 fast genetically encoded calcium indicators”. In: *Ashburn, VA: Janelia Research Campus* (2020).