The Transformation of Visual Information Across Cortical and Subcortical Synapses

By

PRESCOTT COPPELL ALEXANDER
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Kenneth H. Britten, Chair

W. Martin Usrey

Marie E. Burns

Mark S. Goldman

Friedrich T. Sommer

Committee in Charge

2023

Contents

Abstract

Phenomenological models of neural computations have achieved considerable success within modern visual neuroscience; however, the biological details of how neurons and neural circuits might perform these computations remains considerably less well understood. One of the key components in uncovering the biological basis of particular computations is a detailed knowledge of how the response properties of pre-synaptic populations differ from their post-synaptic targets. Armed with a precise understanding of such input-output relations, candidate circuit architectures and cellular or synaptic attributes that could mediate such transformations can then be explored within computational models to generate specific predictions that can then be experimentally tested. The research projects outlined in this work aim to apply this framework to two micro-circuits within the mammalian visual system, the retinogeniculate circuit in carnivores, and the thalamo-cortical circuit in primates. Ultimately, beyond primary sensory receptors, neurons do not respond directly to external stimuli but to patterns of input from their pre-synaptic network. Thus, understanding how information is integrated and transformed across synapses is a critical step in the broader endeavor to understand how organisms sense and respond to the world around them.

1 General introduction

The goal of neuroscience, generally speaking, is to understand the brain, and likewise, the goal of visual neuroscience is to understand vision in biological systems. However, how one defines what precisely constitutes an "understanding", and how one goes about building such an understanding, is not straightforward. In his seminal work, David Marr (Marr, 2010) laid out a framework for understanding vision, and complex systems more generally, that divides understanding into three, largely independent levels of description: 1) the computational, or problem solving level, 2) the algorithmic or representational level, and 3) the implementation or physical realization level. To understand how a cash register works (Marr, 2010), in Marr's view, we need to understand the computational "purpose" of the device (i.e., what problem does the device solve), which essentially amounts to "the adding of numbers". We likewise need to understand how numbers are represented, abstractly speaking, within the device (e.g., it might use a base 10 or a base 2 encoding scheme), and what algorithm is then used to add those representations. Finally, we must also understand how those representations are physically realized (e.g., mechanical wheels, or digital transistors) and how those physical entities are manipulated.

To understand the visual system within this framework we would need to understand (1) what computational problems the system is solving. For example, we might discover, or deduce, that the system must be estimating the depth of objects in order to guide adaptive movements. We would then need to figure out (2) the algorithmic approach that the system is using. For example, it might be inferring depth from binocular disparity, the difference in the relative locations of corresponding objects in the images from each eye. We would then need to explain how objects are represented and matched, and how their locations are compared between the two retinal images. Finally, we would need to describe (3) how those representations are implemented, and how operations on those representations are performed in terms of the patterns of activity of, and connectivity between, the neurons within the visual system that are involved with estimating depth (Marr and Poggio, 1976; Marr, 2010).

While Marr's framework might not apply perfectly to every visual phenomenon that we might wish to explain, for example it may not make sense to talk about the computational role of visual illusions,

it has nonetheless proven to be both useful and influential in informing the approach that visual neuroscience research has taken over the last 40 years. Indeed, the field has made tremendous progress in understanding the visual system at all three levels of explanation (Carandini et al., 2005; Olshausen and Field, 2005; DiCarlo et al., 2012; Yamins and DiCarlo, 2016; Trenholm and Krishnaswamy, 2020; Rust and Cohen, 2022). However, while Marr maintained that the three levels of description are "nearly independent" (Marr and Poggio, 1976), it has become increasingly clear that explanations that make links between the levels are not only desirable, but likely critical in building a holistic understanding of how the visual system works. For example, systems visual neuroscience has historically focused primarily on the algorithmic level of description, where an understanding of the image processing steps performed, and the representational strategy used, by each layer of the system (retina, thalamus, primary and then extrastriate visual areas) is sought by means of receptive field mapping or related techniques (Kuffler, 1953; Hubel and Wiesel, 1961, 1962; Movshon et al., 1978; Adelson and Bergen, 1985; Optican and Richmond, 1987; Olshausen and Field, 1996; Reid et al., 1997; Rust et al., 2005; Rust et al., 2006; Bashivan et al., 2019; Ponce et al., 2019). While this method of explaining neural responses in terms of direct processing of the retinal image is quite powerful, in that it can illustrate neural computations visually via receptive field maps or preferred stimuli, its major drawback is that all earlier stages of the system are either omitted or are approximated in very simple terms. While omitting or coarsely approximating earlier stages is helpful in focusing the analysis on the area under study, it introduces a disconnect between the explanation, given in terms of image processing operations, and the biology, in which visual neurons are, of course, only responding to their direct inputs and do not directly process the visual scene.

The historical focus on explanations in terms of image processing operations (an algorithmic focus), instead of explanations in terms of the processing of neural inputs (more of a neural implementation focus), is due in large part to experimental accessibility. In order to explain neural responses in terms of operations performed on their direct inputs, something about the activity of those direct inputs must be known. While simultaneously measuring pre- and post-synaptic activity is typically not feasible, it has been achieved in very limited circumstances using electrophysiology (Reid and Alonso, 1995; Alonso et al., 1996; Usrey et al., 1998; Rathbun et al., 2010; Fisher et al.,

2017) and/or imaging techniques (Wertz et al., 2015; Wilson et al., 2016, 2018). When available, these simultaneous pre- and post-synaptic neural recordings provide invaluable insight into the computations that individual neurons are performing in terms that can be related to both circuit and cellular mechanisms, as well as the visual scene. Thus, such data can provide a bridge between the vast knowledge of the image processing operations that neurons are known to perform and the biological features of cells and circuits that actually implement those operations.

Simultaneous pre- and post-synaptic recordings in the visual system have most frequently been performed within the retinogeniculate pathway, where relay cells of the lateral geniculate nucleus (LGN) receive very strong connections from at most a few retinal ganglion cells (RGCs) (Cleland et al., 1971; Usrey et al., 1999). Given that the cell bodies of RGCs reside within a thin layer in the inner retina they are relatively accessible to electrophysiological recording techniques by inserting a microelectrode through the sclera (Cleland et al., 1971; Mastronarde, 1983; Usrey et al., 1998) and approaching the inner surface of the retina (just above where the RGC cell bodies reside). While simultaneously recording from an LGN cell, the two-dimensional surface of the retina can then be searched until a RGC with a direct, synaptic connection to the LGN cell under study can be found. Using this technique, it was demonstrated that LGN cells preferentially respond to RGC spikes that arrive in quick succession (i.e., have short, preceding inter-spike intervals (ISIs)) in a manner consistent with temporal summation at the retinogeniculate synapse (Usrey et al., 1998; Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008; Rathbun et al., 2010).

In chapter 2, we reexamine this question by analyzing a collection of simultaneously recorded, synaptically connected RGC-LGN pairs that we believe to be the largest such data set that has been published. Specifically, we approach the question of how LGN relay cells transform their retinal inputs by using three models that try to predict which RGC spikes were relayed on to cortex by the LGN, what we call the "relay status" of RGC spikes. We start by using the well documented relationship between retinal ISI and relay probability, or efficacy, (Usrey et al., 1998; Rathbun et al., 2010) to build a simple model that predicts relay status from just the preceding ISI (Wang et al., 2010), and represents the historical view that temporal summation can explain much or most of how the LGN processes its retinal inputs. However, the basic biophysics underlying the ISI-efficacy relation, namely LGN temporal summation of relatively slowly decaying excitatory

post-synaptic potentials, suggests that the single ISI model is likely too simplistic. The membrane potential of a relay cell at a given moment in time, which represents how likely an arriving retinal spike would be to elicit a response, is influenced by all retinal spikes that have recently arrived, among other factors. In light of this, our work extends the ISI approach by using generalized linear models (GLMs) to predict relay status based on the full pattern of RGC spikes within a given time window, and in a second, two-component model incorporate the full pattern LGN spikes as well. Importantly, this approach allows us to test the sufficiency of the retinal ISI model in explaining how LGN cells integrate their retinal inputs, as the flexibility of GLMs can uncover more complex dynamics on longer timescales if such dynamics are predicatively useful. We find that under most conditions the single ISI model performs as well, or nearly as well, as the GLMs, suggesting that the preceding ISI is indeed the largest factor influencing how LGN cells integrate their retinal inputs. However, further analyses suggest that the form of LGN integration (effective time constant, etc.) is not static, and changes with the level of activity within the early visual circuit. While an ISI mechanism can certainly accommodate this finding, the traditional ISI-efficacy approach, by its very structure, cannot be used to explore this kind of behavior, thus demonstrating the utility of more flexible approaches such as the GLM approach that we use here.

One major impediment to learning neural computations from pre- and post-synaptic recordings is that, on the post-synaptic side where the input-output relation is being learned, the observed input is not the only source of variability (ignoring true intrinsic noise). Even in the retinogeniculate circuit, a relatively simple circuit compared to (for example) those found in cortex, there are non-retinal contributions that modulate the activity of relay cells (Usrey and Alitto, 2015). Most directly, there are resident inhibitory interneurons within the LGN that receive retinal input and synapse on local relay cells (Hirsch et al., 2015). In addition, many relay cell axons, whose principal targets reside in V1, send a collateral branch into the thalamic reticular nucleus (TRN) which drives inhibitory cells that then project back into the LGN (Hirsch et al., 2015; Usrey and Alitto, 2015). Likewise, cortico-thalamic cells within layer 6 of V1 synapse on LGN relay cells and TRN cells, providing both direct excitatory and indirect inhibitory input (Hirsch et al., 2015; Usrey and Alitto, 2015). Several neuromodulatory systems are also known to provide input to the LGN, including the noradrenergic, serotoninergic, and cholinergic systems (Usrey and Alitto, 2015). Despite the

wide array of diverse inputs, the LGN still follows the activity of the retina very closely (Cleland et al., 1971; Mastronarde, 1987; Usrey et al., 1998; Rathbun et al., 2010; Fisher et al., 2017).

The situation in V1 is quite different; compared to retinogeniculate synapses, geniculo-cortical synapses are much weaker (Reid and Alonso, 1995; Hirsch et al., 1998; Usrey et al., 2000), and V1 receives input from an immense number of cortical and subcortical areas (Vanni et al., 2020). As one might expect from the breadth of connectivity, V1 has also been reported to exhibit a higher degree of response variability than the LGN (Vogels et al., 1989; Snowden et al., 1992; Gur et al., 1997). These observations suggested that neural activity in V1 could be usefully partitioned (at a conceptual level) into visually driven activity and ongoing (i.e., not directly visually driven) activity. Arieli and colleagues (Arieli et al., 1996) examined this idea directly, and found that ongoing population activity could account for a great deal of the observed variation of V1 responses to visual stimuli. Thus, the state of V1 at the moment when incoming sensory signals arrive affects how it responds to those signals. This suggests that in order to understand the processing taking place within a visual area, one must not only consider the activity patterns of anatomically earlier areas and how those patterns are transformed upon arrival, but one must also consider the ongoing dynamics within the area of interest itself.

In chapter 3, we investigate how ongoing activity within V1 affects the way in which the different layers of V1 respond to visual and direct electrical stimulation. While their are countless measures of ongoing activity that might account for some amount of the response variability seen in V1, we chose to use a simplified version of a metric developed by Zerlaut et al. (2022) that they found correlates most closely with fluctuations in membrane potential (i.e., the degree of depolarization), namely delta oscillation amplitude. Our reasoning was simply that the amplitude of stimulus driven responses within V1, and more specifically layer 4, should relate, at least in part, to the average depolarization state of the the thalamo-recipient cells. When more layer 4 cells are in a relatively depolarized state, that population (and the larger circuit in which they reside) is expected to show a stronger response. Thus, we sought to validate delta amplitude as an indicator of the excitability of V1 columns as a first step towards building a more detailed understanding of how the more internally driven (or, perhaps, driven by an unchanging stimulus) dynamics within V1 affect the processing of visual information.

Overall, this work aims to further our understanding of how visual neurons and visual circuits process information in the service of adaptive behaviors. While our findings are narrow and often indirect, they nonetheless move us closer to bridging the gap between algorithm and implementation in the context of neural circuits for vision.

2 Dynamics of temporal integration in the lateral geniculate nucleus

This work has been published as:

> Alexander PC, Alitto HJ, Fisher TG, Rathbun DL, Weyand TG, Usrey WM (2022) Dynamics of Temporal Integration in the Lateral Geniculate Nucleus. eNeuro 9. 10.1523/ENEURO.0088-22.2022.

2.1 Abstract

Before visual information from the retina reaches primary visual cortex, it is dynamically filtered by the lateral geniculate nucleus (LGN) of the thalamus, the first location within the visual hierarchy at which non-retinal structures can significantly influence visual processing. To explore the form and dynamics of geniculate filtering we used data from monosynpatically connected pairs of retinal ganglion cells (RGCs) and LGN relay cells in the cat that, under anesthetized conditions, were stimulated with binary white noise and/or drifting sine-wave gratings to train models of increasing complexity to predict which RGC spikes were relayed to cortex, what we call "relay status". In addition, we analyze and compare a smaller data set recorded in the awake state to assess how anesthesia might influence our results. Consistent with previous work, we find that the preceding retinal inter-spike interval is the primary determinant of relay status with only modest contributions from longer patterns of retinal spikes. Including the prior activity of the LGN cell further improved model predictions, primarily by indicating epochs of geniculate burst activity in recordings made under anesthesia, and by allowing the model to capture gain control-like behavior within the awake LGN. Using the same modeling framework, we further demonstrate that the form of geniculate filtering changes according to the level of activity within the early visual circuit under certain stimulus conditions. This finding suggests a candidate mechanism by which a stimulus specific form of gain control may operate within the LGN.

2.2 Significance

The LGN is a dynamic, tunable filter, transforming information as it flows from the retina to primary visual cortex. In this work we utilize a large data set of monosynaptically connected RGC and LGN cell pairs to model the filtering function performed by individual LGN neurons in the anesthetized or awake state. We demonstrate that, while much of the filtering that the LGN performs can be accounted for by temporal summation, other factors, such as the bursting activity of relay cells, also play a role. Additionally, we show that the time scale of summation is dynamic under certain stimulus and network conditions and that the integration dynamics are largely similar between the anesthetized and awake states.

2.3 Introduction

There are two primary dimensions along which relay cells of the lateral geniculate nucleus (LGN) might transform the visual information that they receive from the retina, namely space and time. In the spatial dimensions, a substantial body of evidence suggests a limited transformation, most notably an increase in the strength of the antagonistic surround of the center/surround receptive field (Usrey et al., 1998, 1999; Wang et al., 2010). On the other hand, data demonstrating substantial temporal transformations by LGN relay cells of their direct retinal inputs abound (Usrey et al., 1998; Carandini et al., 2007; Sincich et al., 2007; Sincich et al., 2009; Babadi et al., 2010; Wang et al., 2010; Rathbun et al., 2016). Prior work has demonstrated that the temporal transformation performed by the LGN results in an increased encoding efficiency in the signals sent by the LGN to primary visual cortex (V1) compared to the signals received from the retina (Sincich et al., 2009; Uglesich et al., 2009; Wang et al., 2010), and that this increased efficiency can be explained by temporal summation within relay cells (Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008) and a selective filtering out of less informative retinal spikes (Rathbun et al., 2010). Furthermore, it has recently been shown that temporal summation within the LGN changes with stimulus contrast (Alitto et al., 2019a), suggesting that geniculate filtering is dynamic and can adapt to the statistics of the visual environment. The aim of this work is to investigate this filtering process by modeling the input-output relation of LGN cells using generalized linear models (GLMs), and

to further examine whether the input-output relation changes under different stimulus or network conditions.

In order to investigate the input-output relation of LGN relay cells, we first assembled a large database of simultaneous, extracellular recordings of monosynaptically connected retinal ganglion cell (RGC) - LGN cell pairs from previously published work in anesthetized cats (Usrey et al., 1998; Rathbun et al., 2010; Fisher et al., 2017). Although these data offer a near optimal level of spatial and temporal resolution with which to examine input-output relations in single neurons, they only capture a single RGC input to each relay cell, which are thought to receive input from between two and five RGCs in the cat (Cleland et al., 1971; Usrey et al., 1999). Thus, instead of focusing our analyses on the full spike train produced by relay cells, which contains contributions from all RGC inputs, we instead focus specifically on trying to model the process that determines which spikes from the recorded RGC are relayed, that is elicit a spike in their geniculate partner, and which are not. We begin by considering the simplest model of temporal summation, the often used inter-spike interval (ISI) model (Usrey et al., 1998; Sincich et al., 2007; Weyand, 2007; Wang et al., 2010; Rathbun et al., 2016; Alitto et al., 2019a) whereby the relay probability of each retinal spike is predicted based on the elapsed time since the last retinal spike. We then show how the ISI model can be conceptually extended using GLMs, allowing the full pattern of retinal spikes, within a given window of time, to be used in the predictions. We then introduce a two component GLM that includes the pattern of LGN spikes preceding each retinal spike to investigate whether the LGN spike train contains additional information about the relay probability of future retinal spikes. Finally, we explore whether high levels of activity within the retino-thalamo-cortical circuit influences how LGN relay cells integrate their retinal inputs, and whether this change might explain the dynamic temporal filtering within relay cells that has been previously reported (Rathbun et al., 2016; Alitto et al., 2019a).

While this approach allows the computations performed by individual LGN relay cells to be examined with a level of detail unmatched by any existing method, it does require anesthesia to record the spiking activity of individual RGCs within the eye. In order to complement this approach, and to offer more general findings, we additionally analyze a smaller data set recorded from awake cats in which S-potentials, the extracellular record of excitatory post-synaptic potentials driven by

the dominant retinal input (Kaplan and Shapley, 1984), were recorded simultaneously with the LGN spikes that they frequently elicit (Weyand, 2007). Given the small size of the awake data set, we cannot make quantitative comparisons between the anesthetized and awake state. However, we can use the awake data set to qualitatively confirm, or refute, whether our findings from the anesthetized state are generally applicable.

2.4 Methods

2.4.1 Data sources

The data analyzed in this study contributed to previous reports on the retinogeniculate pathway in both anesthetized (Usrey et al., 1998, 1999; Rathbun et al., 2010, 2016; Fisher et al., 2017; Alitto et al., 2019b) and awake (Weyand, 2007) cats. All experimental procedures conformed to NIH and USDA guidelines and were approved by the Institutional Animal Care and Use Committee at the University of California, Davis or LSU Health Sciences Center.

2.4.2 Code Accessibility

All data and code used in this study are available at: https://github.com/scottiealexander/relaygl m_paper.

2.4.2.1 Computing and software resources All analyses were performed on a Dell Precision T3610 desktop with an Intel Xenon processor (E5-1620) running the Lubuntu 18.04.6 operating system.

All analyses were performed using custom written code in the Julia programming language version 1.6.1 (Bezanson et al., 2017). Visualizations were created using the Julia interface (Johnson, 2020) to the Matplotlib graphics package (Hunter, 2007).

2.4.3 Anesthetized recordings

2.4.3.1 Surgery and preparation Twenty-three adult cats of either sex contributed to this data set. As previously described, anesthesia was initiated with ketamine (10 mg/kg, i.m.) or ketamine and thiopental sodium (20 mg/kg, i.v.) and maintained with either sodium pentothal (2-3 mg/kg/h, i.v.), or isoflurane (0.7-2%). Administration rate of the anesthetic agent was increased when physiological monitoring indicated low levels of anesthesia. A tracheotomy was performed and animals were placed in a stereotaxic apparatus and mechanically respired. Body temperature, ECG, EEG, and expired CO_2 were monitored for the duration of the experiment. All wound margins were infused with lidocaine. The cortical surface overlying the LGN was exposed by a craniotomy and durotomy and then protected with a layer of agarose. To minimize eye movements and facilitate retinal recordings, the sclera beneath the lateral margin of each eye was glued to a rigid ring that was mounted to the stereotaxic frame. The posterior chamber of each eye was accessed via a trans-scleral guide tube inserted through the ring. Upon completion of surgical procedures, animals were paralyzed with either vecuronium bromide (0.2 mg/kg/hr, i.v.) or gallium triethiodide (6-8 mg/kg/h). The nictitating membranes of the eye were retracted with 10% phenylephrine and pupils were maintained in a dilated state with 1% atropine sulfate and flurbiprofen sodium (1.5 mg/h). The eyes were then refracted, fitted with contact lenses, and focused on a tangent screen in front of the animal.

2.4.3.2 Electrophysiological recording and visual stimuli Extracellular recordings of RGCs were made using single, parylene-coated microelectrodes (AM Systems) inserted though the trans-scleral guide tube into the posterior chamber of the eye via a custom-made manipulator. Extracellular recordings of LGN cells in the A laminae were made using a seven-channel multielectrode array (Thomas Recording). Neural signals were amplified, filtered (AM Systems, Thomas Recording) and recorded by either a computer running Brainwave software (Datawave Systems) or a 1401 data acquisition system connected to a computer running the Spike2 software package (Cambridge Electronic Design). Single-neuron isolation was based on waveform analysis and the presence of a refractory period in the auto-correlogram.

Visual stimuli were generated by either a Pepper Graphics System video card (Number Nine Computer Corporation) and presented on a CRT monitor at 80 or 100 Hz (NEC Multisync), or a VSG

2/5 visual stimulus generator (Cambridge Research Systems) and presented on a gamma-calibrated CRT monitor at 140 Hz (Sony). Drifting sinewave gratings that varied in either contrast or diameter were presented at a temporal frequency of 4 Hz and at the optimal spatial frequency for the RGC-LGN pair under study. Binary white-noise stimuli were comprised of a 16x16 grid of squares where the brightness of each square (black or white) on each stimulus frame was governed by a $2^{15} - 1$ frame long pseudorandom sequence, the "m-sequence", Reid et al. (1997). The stimulus frame was updated either on every or every other monitor frame (7 - 25 ms stimulus frame duration).

2.4.4 Awake recordings

Four adult cats of either sex contributed to this data set. Details of surgical procedures, training, and recording have been described previously (Weyand and Gafka, 1998; Weyand, 2007). In brief, animals underwent an initial implant surgery to allow for head-fixed training and eye tracking, followed by a training period in which animals learned to maintain fixation to within 1.5° of a small spot (0.2°) for 1-3 seconds to receive a food reward. Following the training period animals underwent a second surgery in which a cannula was introduced into the brain (~6 mm deep) through a small craniotomy and fixed in place allowing a microelectrode to access the LGN for awake, extracellular, recordings (the orientation of the cannula could be adjusted, see (Weyand, 2007) for details). Signals from microelectrodes (1 - 1.5 MΩ at 1 kHz) were amplified (x100 - 1,000), filtered (0.001 - 10 kHz), and digitized at 22.5 kHz by a modified VCR (A. R. Vetter, Rebersberg, PA) and transferred to a computer for storage using hardware and software from National Instruments (Austin, TX). S-potentials and action potentials were identified and sorted offline using Mini-Analysis (Athens, GA) (for details see (Weyand, 2007)). As S-potentials are thought to be the extracellular record of excitatory post-synaptic potentials driven by the dominant retinal input (Cleland et al., 1971; Kaplan and Shapley, 1984; Weyand, 2007), the delay between a successful S-potential (reflecting a relayed RGC spike) and the triggered LGN spike is substantially shorter than the analogous delay between an RGC spike recorded within the eye and the LGN spike that it triggers. Thus, for the analyses presented in this paper the timing of the S-potentials for a given pair were shifted "backwards" in time relative to the paired LGN spikes by 2.4 ms which ensured that the median

delay between S-potentials and triggered LGN spikes (which was 0.4 ms prior to shifting) matched the median delay observed between RGC and triggered LGN spikes in the anesthetized data set (2.8 ms). This shift helps to minimize any contribution from the different recording approaches to any differences in timing that may be observed between the awake and anesthetized data sets and allows S-potentials to be identified as relayed or not using the same criteria as those used for RGC spikes recorded within the eye (see Identification of monosynaptically connected pairs and relayed RGC spikes). For simplicity, throughout this paper we refer to both RGC spikes recorded within the eye as well as time-shifted S-potentials as "RGC spikes".

Given the difficulty of recording S-potentials in an awake animal, the duration over which individual pairs could be recorded was often quite limited and thus most of the pairs analyzed in this study (7 of 8) were not presented with a controlled stimulus but were instead stimulated by whatever features of the well-lit room fell within their receptive field (see (Weyand, 2007) for details). The one exception, pair 200001250, was stimulated with a sinewave grating (see Supporting Figure 6.2).

2.4.5 Data analysis

2.4.5.1 Identification of monosynaptically connected pairs and relayed RGC spikes

Simultaneously recorded RGC and LGN cells that showed a prominent, short-latency peak in their spike time cross-correlograms were considered to be monosynaptically connected pairs (Mastronarde, 1987; Usrey et al., 1999; Rathbun et al., 2010; Fisher et al., 2017). All S-potential-LGN pairs from (Weyand, 2007) met this criterion by definition. Cross-correlograms, LGN spiking relative to each RGC spike, were constructed for all pairs (from both the anesthetized and awake data sets) using 0.1 ms bins. Peaks were considered prominent if at least one bin exceeded a threshold of $\mu_{\text{baseline}} + 3\sigma_{\text{baseline}}$, where μ_{baseline} is the mean of the baseline period spanning 30-50 ms on either side of the peak bin, and σ_{baseline} is the standard deviation of the baseline period. All bins adjacent to the peak bin that also exceeded the threshold were considered part of the peak. Peaks were considered short latency if they occurred within 2-6 ms of $t = 0$, the time of each retinal spike. All retinal spikes that were followed by a LGN spike that fell within the peak bins of cross-correlograms were considered "relayed", all other retinal spikes were considered "non-relayed". Retinal efficacy (or simply efficacy) is the number of relayed spikes divided by the total number of retinal spikes.

Likewise, all LGN spikes that were preceded by a retinal spike within the monosynaptic window (defined as above) were considered "triggered". Retinal contribution (or simply contribution) is the number of triggered LGN spikes divided by the total number of LGN spikes.

2.4.5.2 Modeling framework All models discussed in the paper generally take the form:

$$\lambda = \sigma(f(t|\boldsymbol{\theta})) \tag{1}$$

where t are retinal spike times, $\boldsymbol{\theta}$ are the model parameters, and λ are the predicted relay probabilities in (0, 1). For the ISI model, f is a nonlinear map between the interval $t_{\mathrm{i}} - t_{\mathrm{i}-1}$ and a conditional intensity. For GLMs, f is a linear function of t represented as a binary vector over an n millisecond interval prior to each t_{i}. For two component GLMs, f also takes as input the LGN spike times t_{LGN}, $f(t; t_{\mathrm{LGN}}|\boldsymbol{\theta})$. For all models, σ is the logistic function:

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{2}$$

that maps a conditional intensity to a relay probability.

2.4.5.3 Assessing model performance All models were assessed in a train-on-90%, test-on-10% ten-fold cross-validation procedure. In each fold, 90% of the data was used to fit the model and the remaining 10% was used only to assess model performance. This procedure was performed ten times such that all data appear in the test set exactly once. Data partitioning across folds was performed such that all test sets contained approximately the same number of relayed spikes. This balancing helped reduce the variability in mean efficacy across folds for a given pair, which serves to stabilize the performance metric that we used (see below) especially for pairs with relatively low mean efficacy. Model performance is the mean performance across folds.

As all models presented in this paper produce a relay probability (0,1) for each retinal spike, we use the cross-validated single-event Bernoulli information ($\mathcal{I}_{\mathrm{Bernoulli}}$) to assess model performance. $\mathcal{I}_{\mathrm{Bernoulli}}$ is the Bernoulli analog of the cross-validated single-spike information used for Poisson

GLMs (Williamson et al., 2015) and can be calculated from the Bernoulli log-likelihood function $\mathcal{L}$ (Truccolo et al., 2005; Williamson et al., 2015):

$$\mathcal{L}(\lambda; y(t)) = \sum y(t)\ \log(\lambda) + (1 - y(t))\ \log(1 - \lambda) \tag{3}$$

where λ are the predicted relay probabilities (as above), and $y(t)$ indicates whether each retinal spike was relayed as {0,1}, what we call "relay status". Using $\mathcal{L}$ we can calculate $\mathcal{I}_{\text{Bernoulli}}$:

$$\mathcal{I}_{\text{Bernoulli}} = \frac{1}{n_{\text{test}} \log(2)} (\mathcal{L}(\lambda_{\text{train}}; y_{\text{test}}) - \mathcal{L}(y_{\text{test}})) \tag{4}$$

where $\lambda_{\text{train}} = \lambda(t_{\text{test}}|\boldsymbol{\theta}_{\text{train}})$ are the predicted relay probabilities for test-set retinal spikes (t_{test}) given the parameters ($\boldsymbol{\theta}_{\text{train}}$) learned from the training-set. $y_{\text{test}} = y(t_{\text{test}})$ is the observed relay status for t_{test}, and $n_{\text{test}} = \text{length}(t_{\text{test}})$ is the number of retinal spikes in t_{test}. $\mathcal{L}(y_{\text{test}})$ represents the log-likelihood of a homogeneous Bernoulli model where the mean efficacy of the test-set is predicted for every spike:

$$\mathcal{L}(y_{\text{test}}) = r_{\text{test}}\ \log(\bar{\lambda}_{\text{test}}) + (n_{\text{test}} - r_{\text{test}})\ \log(1 - \bar{\lambda}_{\text{test}}) \tag{5}$$

where $r_{\text{test}} = \sum y_{\text{test}}$ and $\bar{\lambda}_{\text{test}}$ is the mean efficacy across t_{test} (i.e., $\frac{r_{\text{test}}}{n_{\text{test}}}$).

In this construction, $\mathcal{I}_{\text{Bernoulli}}$ has units of bits/spike and for well fit models will take on values between ~0 (no better than a homogeneous model) and 1 (perfect performance). In practice, poorly fit models can result in negative $\mathcal{I}_{\text{Bernoulli}}$ due to separate training and testing data sets (i.e., cross-validation). Conceptually, $\mathcal{I}_{\text{Bernoulli}}$ quantifies how informative model predictions are about the relay status of the test-set relative to a homogeneous model with the same mean efficacy as the test-set. While somewhat elaborate compared to metrics like accuracy, for a binary process like relay status it is important to take into account the fact that as $\bar{\lambda}_{\text{test}}$ approaches 0 or 1, correctly predicting the outcome becomes trivial. Quantifying model performance relative to a homogeneous model ensures that as $\bar{\lambda}_{\text{test}}$ approaches 0 or 1, $\mathcal{I}_{\text{Bernoulli}}$ approaches 0 for a model with perfect predictions (and values less than 0 for lesser performing models). While this behavior is necessary

to accurately quantify model performance on this kind of classification task (i.e.,where the number of relayed and non-relayed spike cannot be matched), it entails that the maximum achievable $\mathcal{J}_{\text{Bernoulli}}$ depends in part on the mean efficacy of the RGC-LGN pair being modeled. For example, for a pair with an efficacy of 0.05 the maximum $\mathcal{J}_{\text{Bernoulli}}$ for a perfect performing model is 0.286 bits/spike.

2.4.5.4 Inter-spike interval models Usrey et al. (Usrey et al., 1998) described the effect of retinal inter-spike interval (ISI) on efficacy, using the term "paired spike enhancement". They observed that retinal spikes following short ISIs have a higher efficacy than those following long ISIs. Following (Wang et al., 2010), we recast that observation as a simple model for predicting which retinal spikes were relayed based on the elapsed time since the last retinal spike. This model was constructed by creating a histogram of the ISIs preceding all relayed retinal spikes and dividing the count in each bin by the total number of retinal ISIs that fell within that bin. We used a bin width of 1 ms and the resulting histograms were smoothed with a unit-area Gaussian (the standard deviation of which was chosen separately for each pair, see Hyperparameter optimization) to produce a function relating ISI to efficacy (ISI-efficacy function) which we denote as $P(t|t_{\text{ISI}})$ where t is the time of a retinal spike and $t_{\text{ISI}} = t_{\text{i}} - t_{\text{i}-1}$ is the ISI preceding t for ISIs up to a maximum (ISI_{MAX}) that was chosen separately for each pair (see Hyperparameter optimization). For any retinal spikes with ISIs greater than ISI_{MAX}, the model predicted the average efficacy across all ISIs in the corresponding data set. For example, if the ISI of a retinal spike within a given test-set is greater than ISI_{MAX}, the model would predict the mean efficacy of that test-set (i.e.,$\frac{r_{\text{test}}}{n_{\text{test}}}$).

After building $P(t|t_{\text{ISI}})$ (abbreviated as P below for clarity) for a given pair, the fitting algorithm then found a linear transform $f(P) = \beta P + \alpha$ such that the Bernoulli log-likelihood of the resulting predictions

$$\lambda_{\text{ISI}} = \sigma(f(P)) \tag{6}$$

was maximized. This allows a shifting and rescaling of the predictions such that the mean of λ_{ISI}

matches the mean efficacy of the data being used for model fitting. Omitting this step would penalize the ISI model quite significantly in the calculation of $\mathcal{I}_{\text{Bernoulli}}$ because λ_{ISI} may be incorrectly scaled relative to the mean efficacy (due to the ISI cutoff) and thus the likelihood of the homogeneous model, $\mathcal{L}(\bar{\lambda})$ above, would be expected to be large relative to $\mathcal{L}(\lambda_{\text{ISI}})$, yielding potentially negative values for $\mathcal{I}_{\text{Bernoulli}}$ that would incorrectly indicate poor performance.

As with all models discussed herein, for quantifying performance all parameters and hyperparameter were determined from an independent subset of the data from that used to assess performance (see Assessing model performance).

2.4.5.5 Generalized linear models

2.4.5.5.1 General In order to generalize the ISI based model to consider all activity within a period of time preceding each retinal spike, we used a generalized linear model (GLM) framework (Truccolo et al., 2005; Paninski et al., 2007; Pillow et al., 2008; Babadi et al., 2010). GLMs are a generalization of ordinary linear regression in which the to-be-predicted, or "response", variable need not be normally distributed, and the predictor variables and response variable need not be linearly related (Nelder and Wedderburn, 1972). Similarly, GLMs can be thought of as a particular class of linear-nonlinear (LN) cascade models in which the nonlinearity, or activation function, is fixed and invertible (Chichilnisky, 2001; Paninski et al., 2004). GLMs generally take the form:

$$\mathbf{y} = g(\mathbf{X}\boldsymbol{\theta}) \tag{7}$$

where y is the response variable, $\mathbf{X}$ is a matrix of predictors, $\boldsymbol{\theta}$ is a vector of model parameters, and g is the activation function (formally, the inverse link function). Given an assumed or known error distribution of $\mathbf{y}$ and an appropriate choice of g, the parameters $\boldsymbol{\theta}$ can be efficiently fit by maximum likelihood methods (Paninski et al., 2004; Babadi et al., 2010).

In the present context, the response ($\mathbf{y}$) that we are trying to predict is the (binary) relay status of each retinal spike. Thus, a natural choice for the error distribution of $\mathbf{y}$ is the Bernoulli distribution, and a natural choice of activation function is the Logistic function (i.e., logistic regression). The

Bernoulli-Logistic GLM is given by:

$$\mathbf{y} = \lambda(t|\boldsymbol{\theta}) = \sigma(\mathbf{X}\boldsymbol{\theta}) \tag{8}$$

where t are the retinal spike times, and the predictor matrix $\mathbf{X}$ is derived from the retinal spike times alone (retinal history model) or using both the retinal and LGN spikes times (combined history model). The relay status, y, of a set of retinal spikes, t, is then modeled as:

$$y(t) \sim \text{Bernoulli}(\lambda(t|\boldsymbol{\theta})) \tag{9}$$

The parameter vector $\boldsymbol{\theta}$ that minimized the negative log-likelihood (i.e.,$-\mathcal{L}$) for each model instance was found using Newton's method (Nocedal and Wright, 2006) as implemented in (Mogensen and Riseth, 2018).

2.4.5.5.2 Retinal history models Within the GLM framework used here, $\mathbf{X}$ is an m by $n+1$ matrix where m is the number of retinal spikes being used to fit the model (typically 90% of the retinal spikes recorded under a given stimulus condition, see Assessing model performance) and n is the number of temporal components. The additional column is the additive offset or "y-intercept" term. In the "retinal history only" version (RH) of the model, whose predictor matrix, sans-offset, we will refer to as $\mathbf{X}_\text{R}$, the "temporal components" are simply n 1 ms time bins representing the retinal spike train, as a binary vector, during the n milliseconds preceding each retinal spike. In this form, summing over the m rows of $\mathbf{X}_\text{R}$ would yield the autocorrelogram of the retinal spike train over an n millisecond window. The hyperparameter n was optimized separately for each pair (see Hyperparameter optimization). The n parameters corresponding to the n time bins of a fitted model can be thought of as a linear kernel or filter that reflects the extent to which retinal spikes occurring at a given time prior to the "target spike" influence the likelihood that the target spike will be relayed.

Given previous work suggesting that LGN temporal filters are likely to be smooth functions in this context (Usrey et al., 1998; Rathbun et al., 2010), and to help prevent overfitting, we introduce

a smoothing prior on $\boldsymbol{\theta}$ (excluding the y-intercept term), yielding a maximum *a-posteriori* (MAP) estimator for $\boldsymbol{\theta}$:

$$\mathcal{L}_{\mathrm{MAP}}(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta}) - \eta \sum (\theta_{\mathrm{i}} - \theta_{\mathrm{i-1}})^2 \tag{10}$$

where the prior weighting term η is optimized separately for each pair (see Hyperparameter optimization).

2.4.5.5.3 Combined history models The combined history (CH) model extends the RH model by introducing another set of predictors derived from the activity of the LGN cell. In the RH model, the LGN cell only contributes by classifying each retinal spike as relayed or non-relayed, whereas in the CH model the recent activity of the LGN cell can also contribute to the relay prediction (via spikes not "caused" by the recorded RGC). For the CH model, the predictor matrix $\mathbf{X}_{\mathrm{C}}$ can be thought of as the column-wise concatenation of $\mathbf{X}_{\mathrm{R}}$ with an analogous binary matrix, $\mathbf{X}_{\mathrm{L}}$, of size m by p where each row of length p is the LGN cell's binary spike train (1 ms bin size) during the p milliseconds preceding each retinal spike. Thus, summing over the rows of $\mathbf{X}_{\mathrm{L}}$ would yield the cross-correlogram of the LGN activity relative to the RGC spike times for negative time lags. Importantly, the time window in which the LGN cell could respond to a given RGC spike was not included; the model could only consider events preceding a retinal spike in predicting whether or not it was relayed.

To help prevent overfitting we introduce a Gaussian prior on the coefficients of the CH model (θ_{C}) to penalize large coefficient values (i.e., ridge regression), yielding a MAP estimator for θ_{C}:

$$\mathcal{L}_{\mathrm{MAP}}(\theta_{\mathrm{C}}) = \mathcal{L}(\theta_{\mathrm{C}}) - \eta \sum {\theta_{\mathrm{C}}}^2 \tag{11}$$

where, as above, η is optimized separately for each pair (see Hyperparameter optimization).

As the CH model is an extension of the RH model, the time window spanned by $\mathbf{X}_{\mathrm{R}}$ was fixed, for each pair, at the value derived from RH model fitting (see Retinal history models above). The time window spanned by $\mathbf{X}_{\mathrm{L}}$ was optimized for each pair in an analogous manner (see Hyperparameter

optimization).

In order to help mitigate the cost of increasing the number of free parameters (which would otherwise increase quite dramatically), for CH models $\mathbf{X}_\mathrm{R}$ and $\mathbf{X}_\mathrm{L}$ were represented in a basis of raised-cosine functions following common practice (Pillow et al., 2005; Pillow et al., 2008; Ghanbari et al., 2017):

$$b_\mathrm{k}(t) = \frac{\cos(q_\mathrm{k}(t)) + 1}{2} \tag{12}$$

$$q_\mathrm{k}(t) = (\log(t + \Psi) - \log(\phi_\mathrm{k} + \Psi))\frac{\pi}{2\gamma} \tag{13}$$

such that $q_\mathrm{k}(t) \in [\text{-}\pi, \pi]$, where ϕ_k is the center of the "raised bump" of the k-th basis vector, Ψ is a constant hyperparameter (see Hyperparameter optimization) that controls the linearity of the spacing between bumps, and γ is a scaling factor that controls the width of the bumps such that they tile the time axis (i.e.,γ is a function of the number of basis vectors and the duration they need to cover). This representation greatly reduces the number of parameters while still allowing good temporal resolution around the time of the retinal spike (by setting Ψ closer to ~1) at the cost of forcing the kernels to be smooth. However, this smoothness assumption is well supported by Usrey et al. (1998) and Rathbun et al. (2010) and loosely resembles, in its effects, the smoothing prior used in fitting RH models.

In practice $\mathbf{X}_\mathrm{C}$ is a m by $n_\mathrm{R} + n_\mathrm{L} + 1$ matrix where n_R and n_L are the number of basis vectors used to represent $\mathbf{X}_\mathrm{R}$ and $\mathbf{X}_\mathrm{L}$ respectively. Here, n_R was set to 16 and n_L was optimized separately for each pair (see Hyperparameter optimization).

In a manner analogous to thinking of $\mathbf{X}_\mathrm{C}$ as $[\mathbf{X}_\mathrm{R}\ \mathbf{X}_\mathrm{L}]$, we can separate the retinal and LGN filters learned by the model as $\boldsymbol{\theta}_\mathrm{C} = [\boldsymbol{\theta}_\mathrm{R},\ \boldsymbol{\theta}_\mathrm{L}]$ (ignoring the additive offset term). For clarity, throughout this paper when we refer to $\boldsymbol{\theta}_\mathrm{R}$ (or $\boldsymbol{\theta}_\mathrm{L}$) we are referring to $\boldsymbol{\theta}_\mathrm{R}$ transformed back into the time-domain:

$$\boldsymbol{\theta}_\mathrm{R} = \mathbf{B}_\mathrm{R}{\boldsymbol{\theta}^*}_\mathrm{R} \tag{14}$$

$$\mathbf{B}_\mathrm{R} = [b_{\mathrm{R},1}(t)\ b_{\mathrm{R},2}(t)\ ...\ b_{\mathrm{R,k}}(t)] \tag{15}$$

where $b_{\mathrm{R,k}}(t)$ is the k-th basis for $\mathbf{X}_{\mathrm{R}}$ (as above) and $\boldsymbol{\theta}^{*}_{\mathrm{R}}$ are the coefficients on $\mathbf{X}_{\mathrm{R}}$ learned by the model. The analogous set of relations apply to $\boldsymbol{\theta}_{\mathrm{L}}$, $\mathbf{B}_{\mathrm{L}}$ etc.

2.4.5.5.4 Optimization error For RH models, in which data used for fitting were represented in the standard temporal basis, the standard error of the estimate for each parameter was computed from the Hessian of the log-likelihood function ($\bigtriangledown^{2}\mathcal{L}(\boldsymbol{\theta}_{\mathrm{ML}})$) at the maximum likelihood estimate ($\boldsymbol{\theta}_{\mathrm{ML}}$) following standard practice (Paninski et al., 2004; Truccolo et al., 2005; Paninski et al., 2007; Babadi et al., 2010).

$$stderr(\boldsymbol{\theta}_{\mathrm{ML}}) = \mathrm{diag}([\bigtriangledown^{2}\mathcal{L}(\boldsymbol{\theta}_{\mathrm{ML}})]^{-1})^{\frac{1}{2}} \tag{16}$$

The standard error of parameter estimates for models fit to data represented in the raised-cosine basis are omitted from visualizations as the standard error cannot be validly transformed back into the time-domain as the parameter estimates can.

2.4.5.6 Hyperparameter optimization Hyperparameter values that could be chosen based on the literature or reasonable assumptions (in cases where a hyperparameter has little impact on the model overall) were fixed for all pairs at the values specified below. For quantifying model performance (Figures 4 & 5), all other hyperparameters were chosen for each pair from a predefined set based on which value yielded the highest cross-validated $\mathcal{I}_{\mathrm{Bernoulli}}$ in a nested cross-validation procedure. On each fold of the main cross-validation loop the training-set (consisting of 90% of the total data from a pair) was further partitioned into sub-training and sub-testing sets (again a 90-10 split); the combination of model parameters and hyperparameters that yielded the highest cross-validated $\mathcal{I}_{\mathrm{Bernoulli}}$ on the sub-testing set across sub-folds were then used to quantify the model's performance on the main testing-set.

2.4.5.6.1 ISI efficacy model The maximum ISI for which the ISI-efficacy model would predict a value other than the mean was chosen for each pair from a set of eight logarithmically spaced values (base 10) between 0.03 and 0.5 seconds. The standard deviation of the Gaussian kernel used

to smooth ISI-efficacy functions was chosen from a set of seven logarithmically spaced values (base 10) between 0.002 and 0.03 seconds and the value 0 (i.e., no smoothing).

2.4.5.6.2 RH model The RH model contains two hyperparameters: the temporal span, which is the length of the time window preceding each target retinal spike that is used to train the model, and the prior weighting term, η, that controls the magnitude of the smoothness constraint (see Retinal history models). The temporal span was chosen from a set of eight logarithmically spaced values (base 10) between 0.03 and 0.5 seconds (rounded to the nearest millisecond). The prior weighting term was chosen from a set of five logarithmically spaced values (base 2) between 4 and 4096.

2.4.5.6.3 CH model As the CH model is an augmented version of the RH model, the temporal span of the retinal component of each pair's CH model was fixed at the value derived from RH model fitting. Thus, seven hyperparameters remained: the temporal span of the LGN component, the number of basis vectors (in the raised-cosine basis, see Combined history models) used to represent each component (one hyperparameter per component), the weight given to the l^2 penalty for each component, and the linearity of the basis vector spacing, Ψ, (see Combined history models, again one per component). The temporal span of the LGN component was chosen from a set of eight logarithmically spaced values (base 10) between 0.04 and 0.6 seconds (rounded to the nearest millisecond). The number of basis vectors for retinal components was fixed at 16 for all pairs. For LGN components the number of basis vectors was chosen from the set {8, 12, 18, 24, 32}. The weight of the l^2 penalty was chosen from a set of five logarithmically spaced values (base 2) between 0.125 and 8.0. The linearity of basis vector spacing, Ψ, was fixed at 10 and 8 for retinal and LGN components respectively. The six-dimensional grid defined by the specified sets of values for the six non-fixed hyperparmeters was searched exhaustively.

2.4.5.7 Filter visualization For visualizing (Figures 2 & 3) and analyzing (Figures 6-8) temporal profiles of the filters learned by the models, data from all pairs were fit with a fixed set of hyperparameters: temporal span for both RH and CH components was fixed at 200 ms, and the number of basis vectors in RH (CH) components was fixed at 16 (24). As they more directly affect

the shape of the learned filters, prior weighting terms were chosen individually for each pair using ten-fold cross-validation from the same range specified in Hyperparameter optimization. When displaying averaged filters for a population or condition filters for each pair were scaled to have unit norm prior to averaging and, unless specified otherwise, error shading reflects the 95% confidence interval of the mean (see Statistics).

2.4.5.8 Burst spike definition Geniculate bursts were identified by the criteria established by (Lu et al., 1992): a geniculate burst must be preceded by at least 100 ms of quiescence and contain two or more spikes each separated by no more than 4 ms. The relaxed definition reduced the quiescence duration to 50 ms and increased the maximum ISI to 6 ms (Figure 2.4C & D). Non-cardinal burst spikes were defined as all spikes that were part of an identified burst, except the first or "cardinal" spike of each burst.

2.4.5.9 Classification of retinal spikes by activity level In order to assess how the level of activity of the early visual network might alter the integration dynamics of LGN cells we partitioned all retinal spikes from a given pair into four quartiles based on the LGN spike count in a 100 ms window preceding each retinal spike. RH models were then fit separately to data from each quartile. Differences between filters learned from data from distinct quartiles were quantified by taking the integral of the absolute difference between the two filters: $\int \mid \theta_{\mathrm{N}} - \theta_{\mathrm{M}} \mid$. Where θ_{N} and θ_{M} are the filters learned the N'th and M'th quartiles respectively. We refer to this metric as the "absolute difference" metric.

2.4.5.10 Simulating GLMs Given a retinal spike train and a set of learned filter coefficients $\boldsymbol{\theta}$, a GLM can be used to simulate the relay status of the retinal spike train by constructing a predictor matrix $\mathbf{X}$ from the retinal spike train as described above (see Retinal history model), multiplying $\mathbf{X}$ by the learned coefficients and passing the result through the logistic function $\sigma(\mathbf{X}\boldsymbol{\theta})$ to attain the predicted relay probability for each retinal spike. Relay status $\mathbf{y}$ can then be simulated by drawing a random number for each retinal spike from a uniform distribution on (0,1); if the random number is less than the predicted probability for a given retinal spike the spike is considered relayed (this is equivalent to flipping a coin whose probability of heads, or in this case "relayed", is given by the

predicted relay probability). A GLM can then be fit to the retinal spike train and simulated relay status just as is done for real data (see Retinal history model). Due to the stochastic nature of simulating relay status, for all simulations presented here the final two steps (simulate relay status and fit GLM) are repeated 50 times for each pair and the resulting coefficients are then averaged.

2.4.5.11 Statistics Unless otherwise noted in the text, data are reported as the median (or paired median difference) and the median absolute deviation (MAD) defined as: median($|\ \mathbf{x} - \text{median}(\mathbf{x})\ |$). Confidence intervals are derived from bootstrap estimation with 5000 re-samples, and are bias corrected and accelerated (Efron, 1987) using the Bootstrap.jl software package (Gehring et al., 2021). For the awake data set, the small sample size prevents the valid use of bootstrap-derived confidence intervals; thus we report the range of values ([min, max]) instead. For model comparisons, p-values are calculated from paired samples permutation tests with 5000 re-samples, where the permutation is performed within pair. For example, if comparing model A to model B, on each iteration the model performance values for each pair are randomly reassigned (i.e., swapped or not between A and B with probability 0.5) and the resulting paired median difference is calculated. After 5000 iterations the observed paired median difference is compared to the permuted differences distribution to yield a p-value. Computed p-values are then corrected so that they cannot be exactly zero (which would otherwise be possible given the discrete nature of the permuted differences distribution) using the method proposed in Phipson and Smyth (2010).

2.5 Results

To investigate the factors that contribute to how the LGN filters retinal spike trains, we analyzed data from 45 monosynaptically connected RGC-LGN cell pairs from anesthetized cats and 8 pairs from awake cats. For the recordings under anesthesia, neurons were stimulated with binary white noise (n=40) and/or drifting sinewave gratings (n=33) and connectivity was assessed by cross-correlation of the spike times from the two simultaneously recorded neurons. Figure 2.1 shows data from an example pair. The top row (Figure 2.1, A & B) shows receptive field (RF) maps of the RGC (left) and LGN neuron (right) derived from the spike-triggered average of the binary white-noise frames. The one standard deviation contour of a circularly symmetric Gaussian fit to

the LGN (RGC) RF is overlayed in white (black) on the RGC (LGN) RF, demonstrating the high degree of spatial overlap between the two RFs. The bottom row (Figure 2.1, C & D) shows the cross-correlograms, LGN spike times relative to each RGC spike, for the two stimulus conditions for this pair. Using a monosynaptic latency derived from the time lag at which the cross-correlogram peaks, we identified each retinal spike as being relayed (i.e., evoked a spike in its LGN partner) or not, and each LGN spike as being triggered (i.e., was evoked by a RGC spike) or not. Retinal efficacy, the proportion of retinal spikes that were relayed (see Identification of monosynaptically connected pairs and relayed RGC spikes), for this example pair during binary white noise (drifting grating stimuli) was 0.316 (0.473); retinal contribution, the proportion of LGN spikes that were triggered, was 0.812 (0.760). Across the population, for binary white noise data median retinal efficacy was 0.097 (median absolute deviation (MAD) 0.070, 95% confidence interval (CI) [0.054, 0.173]) and median retinal contribution was 0.247 (MAD 0.156, 95% CI [0.136, 0.394]); for drifting grating data median retinal efficacy was 0.161 (MAD 0.098, 95% CI [0.088, 0.229]) and median retinal contribution was 0.347 (MAD 0.197, 95% CI [0.205, 0.490]).

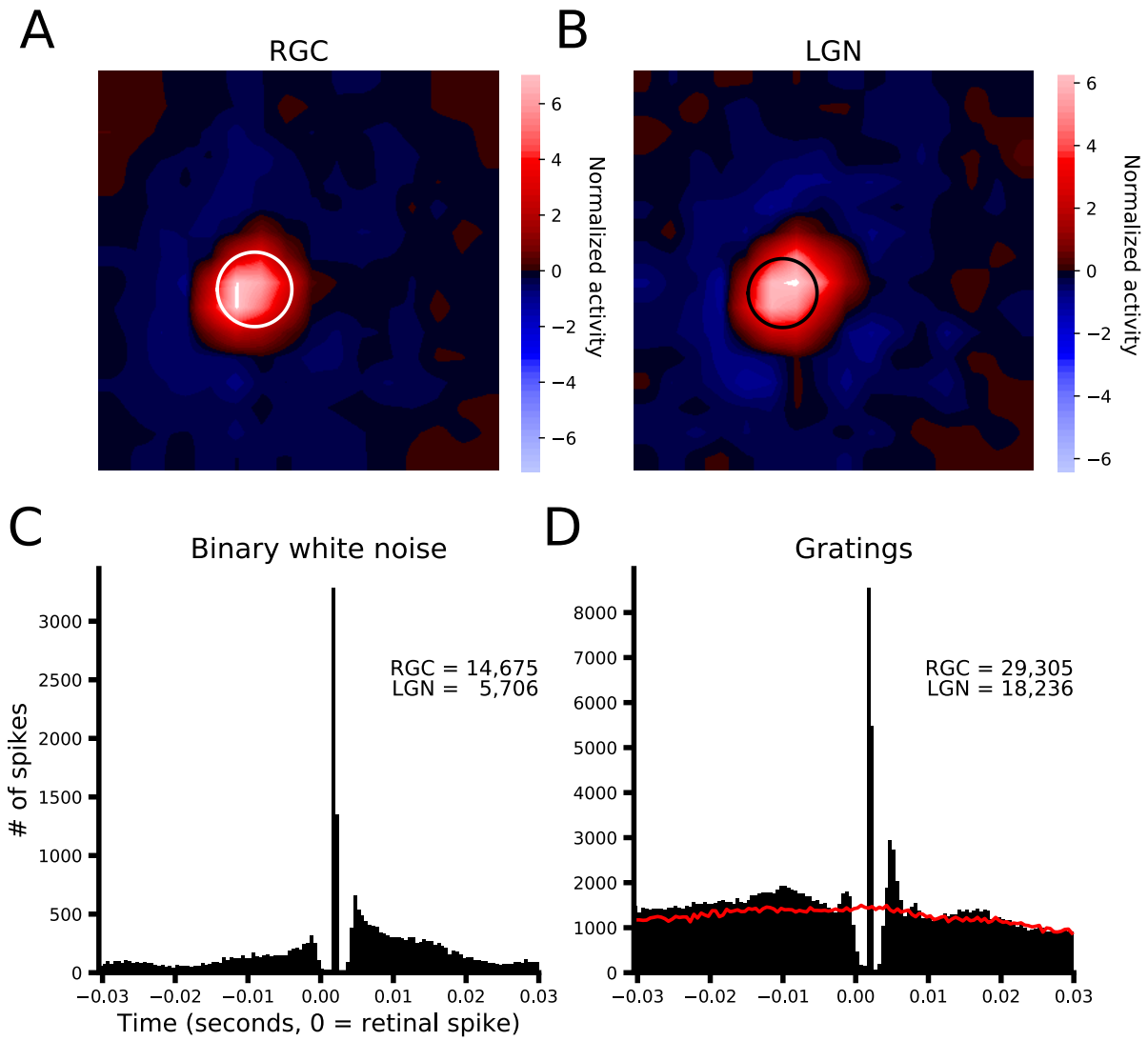


Figure 2.1: Data from an example pair (pair 214). **A and B**: Receptive field (RF) maps derived from reverse correlation between recorded spike trains and binary white noise stimulus. Red (blue) denotes regions of the RF that were excited by brighter (darker) pixels. White (black) circle in **A** (**B**) is the 1 SD contour of a circular Gaussian fit to RF of the LGN cell (RGC) overlayed on the RGC (LGN) RF to illustrate the high degree of spatial overlap. **C and D**: Cross-correlation between RGC and LGN spike trains for binary white noise (**C**) and drifting sinewave grating (**D**) stimuli. The inset text indicates the number of spikes recorded from each of the two neurons (**C**: 14,675 retinal spikes, 5,706 LGN spikes; **D**: 29,305 retinal spikes, 18,236 LGN spikes). The red line in **D** shows the correlation due to the stimulus that is attained if the spike train of the RGC is shifted in time by one stimulus cycle.

Additionally, we analyzed data from a smaller set of eight RGC-LGN cell pairs from awake cats in which the spike train of the connected RGC was inferred from the presence of S-potentials that could be isolated, along with the LGN cell's spikes, from the extracellular voltage trace recorded

within the LGN (Weyand, 2007). Across the population, median retinal efficacy was 0.519 (MAD 0.133, range [0.154, 0.724]) and median retinal contribution was 0.935 (MAD 0.037, range [0.604, 0.997]).

Importantly, these data confirm the well documented finding that not every retinal spike is relayed by the LGN ((Cleland et al., 1971; Kaplan and Shapley, 1984; Kaplan et al., 1987; Usrey et al., 1998; Sincich et al., 2007; Weyand, 2007) among others). Taken together with the generally accepted notion that every non-burst relay cell spike is triggered by the retina (Kaplan and Shapley, 1984; Sincich et al., 2007; Weyand, 2007), this finding suggests that the primary role of LGN relay cells is to edit the incoming retinal spike train by selective deletion. Thus, we sought to investigate the factors that determine which retinal spikes are relayed and which are not, what we term "relay status". Given this goal, we consider models of retinogeniculate transmission that focus specifically on predicting the relay status of retinal spikes rather than trying to predict the LGN spike train directly (i.e., we do not attempt to predict LGN spikes that were not triggered by the recorded RGC).

2.5.1 ISI efficacy model

Previous work has clearly demonstrated that one of the primary factors that determines which retinal spikes are relayed is the elapsed time since the last retinal spike (i.e., retinal interspike interval (ISI), (Usrey et al., 1998; Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008; Sincich et al., 2009; Wang et al., 2010)). This is often visualized by plotting retinal efficacy as a function of the preceding retinal ISI ((Usrey et al., 1998), see Inter-spike interval models). Figure 2.2, left column, shows the ISI-efficacy relation for an example pair of cells from the anesthetized data set (A, pair ID 208), the population as a whole (C, anesthetized data set), and the relations for each pair in the awake data set (E), where the data from each pair in C and E were normalized to their mean before averaging. The ISI-efficacy functions follow the typical decay pattern (shorter ISIs in general show higher efficacies) that has been reported previously (Usrey et al., 1998; Weyand, 2007; Rathbun et al., 2010). Interestingly, the drifting grating data (Figure 2.2C, red line) do show a slight increase in efficacy for ISIs greater than 150 ms, potentially caused by the release from a slow acting suppressive influence such as synaptic depression. Implicitly, the ISI-efficacy relation is a simple

model for predicting which retinal spikes were relayed based on the preceding retinal ISI (Wang et al., 2010), thus we formalized the model to quantitatively access its decoding performance. We utilized a 10-fold cross-validation procedure in which ISI-efficacy functions were constructed using 90% of retinal spikes (training set), and performance was assessed on the remaining 10% (test set) by looking up the expected efficacy of each spike in the test set from the training-set-derived ISI-efficacy function (see Inter-spike interval models). This procedure was repeated ten times such that each retinal spike was included in the test set once and model performance was evaluated by the cross-validated, single-spike Bernoulli information ($\mathcal{I}_{\text{Bernoulli}}$) which quantifies how informative model predictions are about the relay status of test-set retinal spikes relative to a homogeneous model that always predicts the mean efficacy (see Assessing Model Performance). For binary, white-noise data, median $\mathcal{I}_{\text{Bernoulli}}$ was 0.019 bits/spike (MAD 0.018, 95% CI [0.004, 0.041]). For drifting grating data, median $\mathcal{I}_{\text{Bernoulli}}$ was 0.026 bits/spike (MAD 0.017, 95% CI [0.010, 0.030]). For the awake data set, median $\mathcal{I}_{\text{Bernoulli}}$ was 0.177 bits/spike (MAD 0.085, range [0.075, 0.439]). This demonstrates that the ISI-efficacy model was able to predict the relay status of retinal spikes significantly better than the homogeneous model regardless of the stimulus condition or the state of the animal (anesthetized or awake).

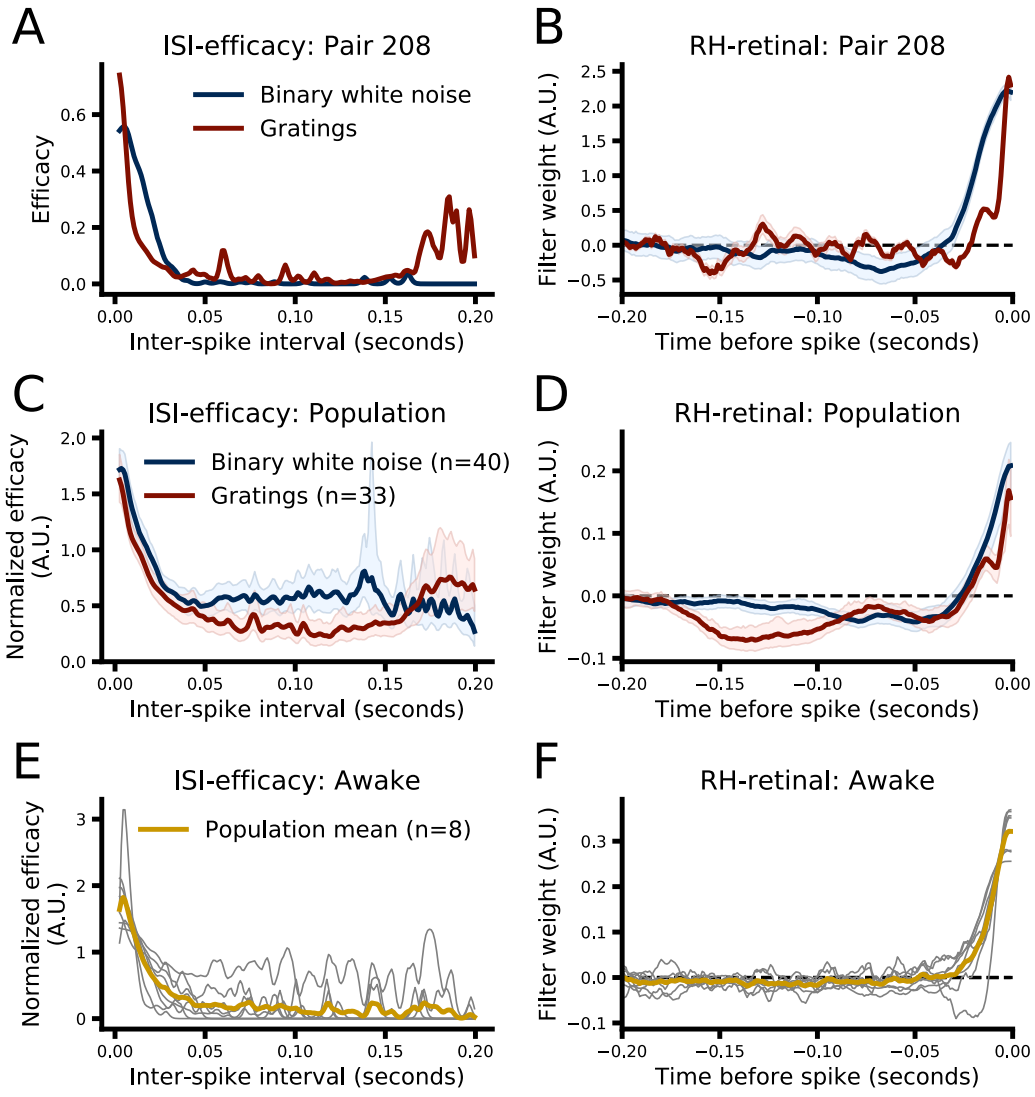


Figure 2.2: Comparison of ISI efficacy (left column) and retinal history (RH) models (right column). **A**: Relationship between retinal ISI and retinal efficacy for binary white noise (blue) and drifting grating data (red) for pair 208. **B**: Retinal filters learned by the RH model fit to binary white noise (blue) and drifting grating (red) data from pair 208. Shading indicates ± 1 standard error of the optimization (see Methods). The time base for GLM filters is always relative to the retinal spike about which a prediction (relayed or non-relayed) is being made (i.e., the "target spike"). **C**: Normalized ISI-efficacy relation averaged across the population. Efficacies for each pair were normalized to the mean efficacy across all ISIs for that pair before averaging. Shading represents the 95% confidence interval (CI) across pairs from 5000 bootstrap resamples (see Statistics). **D**: Same as **B** but showing the average filters across pairs. Filters fit to the data from each pair were scaled to have a unit norm prior to averaging. Shading represents the 95% CI across pairs. **E**: Normalized ISI-efficacy relations for all eight pairs from the awake data set (thin grey lines) and the population average (thick gold line). Normalization was performed as in **C**. **F**: Retinal filters learned by RH models fit to data from each pair in the awake data set (thin grey lines) and the population average (thick gold line). Filters were scaled to have unit norm (as in **D**) to aid visualization.

2.5.2 Retinal history model

While retinal ISI is a strong predictor of relay status, its influence is a natural consequence of the temporal integration that occurs within relay cells. This fact suggests that the history dependence of relay probability is likely to extend beyond the most recent spike and might be better captured by considering all retinal spikes that occur within a given window of time. Thus, we sought to extend the ISI-efficacy model by using generalized linear models (GLM) to predict the relay status of retinal spikes based on the patterns of retinal activity preceding each spike, what we call the retinal history (RH) model. Historically, GLMs have been used to predict the activity of visual neurons based on the changing pattern of a visual stimulus (Chichilnisky, 2001; Paninski et al., 2004; Truccolo et al., 2005; Pillow et al., 2008; Babadi et al., 2010); here we instead use the pattern of activity recorded simultaneously from a monosynaptic input (see Generalized linear models). In brief, the GLM predicts the relay status of a retinal spike by convolving the pattern of recent activity with a learned temporal filter, the output of which is then passed through a static nonlinearity to produce a relay probability. Specifically, we use Bernoulli-Logistic GLMs (i.e.,logistic regression) to model retinogeniculate transmission as a binary parsing (Wang et al., 2010) or coin-flip process where the probability of a positive outcome (relayed retinal spike) varies continuously over time as a function of the pattern of recent retinal spikes (see Retinal history models).

Figure 2.2, right column, shows the temporal filters learned from drifting grating (red) and binary, white-noise data (blue) for an example pair (Figure 2.2B, pair ID 208), the population recorded under anesthesia (Figure 2.2D), and the population recorded in the awake state (Figure 2.2F), where filters from each pair were scaled to have unit norm before averaging in D and F. For visualization purposes, the time span preceding each retinal spike that the model could consider (temporal span) was set to 0.2 seconds for all pairs (see Filter visualization and Hyperparameter optimization). Much like the ISI-efficacy functions, the temporal filters show relatively large positive values in the time window just prior to the target retinal spike (at t=0), indicating that retinal spikes falling within this time window increase the likelihood that the target retinal spike will be relayed. Retinal spikes that occurred earlier relative to the target spike ($>$ 0.02-0.04 seconds) were less informative about relay status, as shown by the smaller magnitude of the filter values, and in general tended to slightly decrease the probability that the target retinal spike would be relayed (i.e., filter values slightly $<$

0) for pairs recorded under anesthesia. Interestingly, the filters learned from drifting grating data tended to have larger negative values during earlier pre-spike time windows (> ~0.08-0.18 seconds pre-spike) and show a slight oscillation at ~10Hz, which is unlikely to be due solely to the periodic nature of the drifting grating, which had a temporal frequency of 4Hz (see Electrophysiological recording and visual stimuli). As with the ISI-efficacy model, the performance of the RH model was assessed using a train-on-90% test-on-10%, 10-fold cross-validation procedure, in which overall performance was computed as the average $\mathcal{I}_{\text{Bernoulli}}$ across folds. For binary, white-noise data median $\mathcal{I}_{\text{Bernoulli}}$ was 0.024 bits/spike (MAD 0.022, 95% CI [0.007, 0.042]). For drifting grating data median $\mathcal{I}_{\text{Bernoulli}}$ was 0.068 bits/spike (MAD 0.047, 95% CI [0.036, 0.099]). For the awake data set $\mathcal{I}_{\text{Bernoulli}}$ was 0.154 bits/spike (MAD 0.072, range [0.061, 0.452]).

2.5.3 Combined history model

Although the recorded RGC could account for the majority of LGN spikes in many cell pairs (i.e., a retinal contribution > 0.5), a considerable number of LGN spikes could not be directly attributed to (i.e., were not triggered by) the recorded RGC. These non-triggered spikes likely represent the activity of other RGC inputs to the recorded relay cell (Usrey et al., 1999), and might provide a non-redundant source of information that could aid predictions about which RGC spikes were relayed and which were not. Thus, we built an augmented version of the RH model that included an additional filter that acted on the spiking history of the recorded LGN relay cell, what we call the combined history (CH) model. Two attributes of this additional filter are worth noting: 1) the activity of the LGN cell only contributes to the RH model by identifying which retinal spikes were relayed. Thus, for any pair with a retinal contribution less than one, the LGN activity may contribute additional information that the model can take advantage of, and 2) the LGN filter is aligned relative to the time of the target retinal spike just as the retinal filter is, so only LGN spikes that occurred prior to target retinal spike are included (see Combined history models). This construction is distinct from those commonly used to represent spike history effects in GLM models (Pillow et al., 2008; Babadi et al., 2010) and reflects our focus on predicting the relay status of retinal spikes and not the activity of the LGN cell per se. As a result, the LGN filter can capture some features of LGN activity, such as bursting in certain circumstances, but not others, such as a

refractory period, which is not relevant for predicting retinal relay status.

Figure 2.3 shows the retinal (Figure 2.3A) and LGN (Figure 2.3B) filters for an example pair (pair ID 208) and the population as a whole (Figure 2.3C & D, filters from each pair were scaled to have unit norm before averaging). For visualization purposes, the temporal span of both retinal and LGN filters was set to 0.2 seconds for all pairs (see Filter visualization). Two aspects of the filters learned by the CH model are worth noting. First, the shape of the retinal filters are nearly identical to the shape of the retinal filters learned by the RH model as expected (compare Figure 2.3D and Figure 2.3C), despite using far fewer parameters (see Combined history models), suggesting that the addition of the LGN filter has not fundamentally changed how the model is weighting retinal spikes in making predictions. Second, much like the retinal filters, the LGN filters show a strong positive component immediately preceding the target spike that rapidly declines (~-0.015 seconds) followed by a lower amplitude negative component that decays to near zero fairly quickly for drifting grating data (~0.04 seconds, Figure 2.3C red) and more slowly for binary white noise data (~0.1 seconds, Figure 2.3C blue). The strong, positive weights assigned by the model to the time window immediately preceding the target spike suggests that retinal spikes that follow LGN spikes at very short latencies are more likely to be relayed. This pattern of LGN-RGC-LGN spiking is expected to be particularly likely when a retinal spike arrives during a geniculate burst (Llinás and Jahnsen, 1982; Huguenard and McCormick, 1992; Alitto et al., 2019b). To test whether this filter component was in fact due to LGN bursting, we repeated the CH model fitting procedure after removing all non-cardinal burst spikes (i.e., removing all spikes that comprise a burst except the first spike, see Burst spike definition). Interestingly, while the resulting filters do show a strongly attenuated early positive component for the drifting grating data, removing all non-cardinal burst spikes only minimally altered the LGN filters learned from binary white noise data (Figure 2.4B). However, relaxing the definition of bursts somewhat to include more high-frequency events reduced the early positive component for binary white noise data (Figure 2.4D), suggesting that the early positive component of LGN filters may reflect both burst as well as high-frequency, non-burst events (Alitto et al., 2019b).

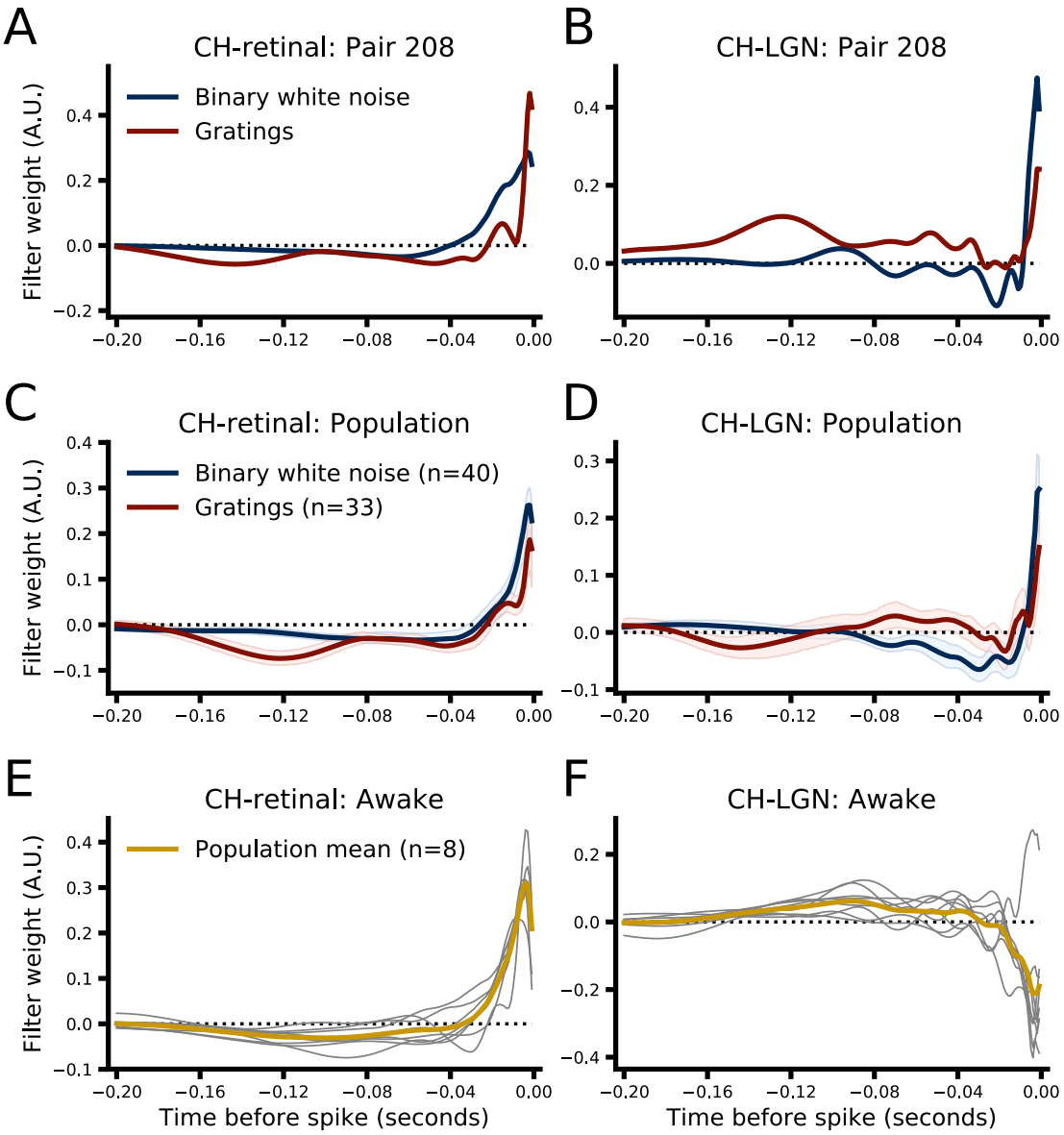


Figure 2.3: Summary of filters learned by the two-component, combined history (CH) model. The left column shows the retinal filters, and the right column shows the LGN filters for example pairs and the population for each data set. **A**: Retinal filters learned by the CH model for binary white noise (blue) and drifting grating (red) data from pair 208. **B**: Same as **A** but showing the LGN filters learned by the CH model. The time base for retinal and LGN filters is the same (0 is the time of the "target" retinal spike), but LGN filters operate on the prior activity of the LGN cell. **C**: Same as **A** but for the population. Filters fit to the data from each pair were scaled to have a unit norm prior to averaging. Shading represents 95% CI across pairs. **D**: Same as **C** but for LGN filters. **E**: Same as **C** but showing retinal filters learned from the awake data set (thin grey lines show filters from each pair, the thick gold line shows the mean across pairs). **F**: same as **E** but showing LGN filters.

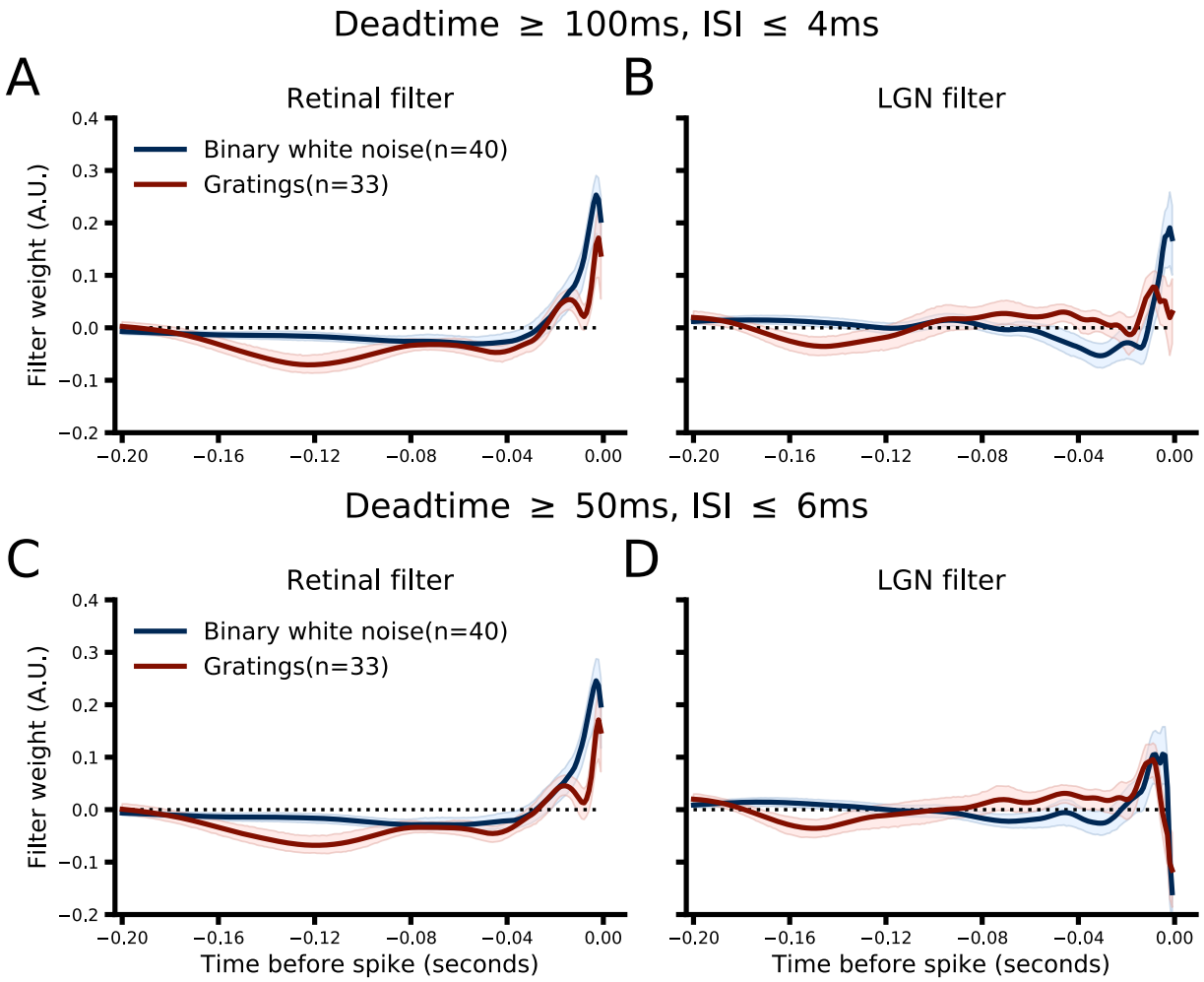


Figure 2.4: Retinal (**A** & **C**) and LGN (**B** & **D**) filters from the CH model fit to data where non-cardinal burst spikes were first removed. The first row (**A** & **B**) use the classic burst spike definition of Lu et al. (1992): a quiescent period $\geq$ 100 ms followed by two or spikes with ISIs $\leq$ 4 ms. The second row (**C** & **D**) use a more relaxed criteria: a quiescent period $\geq$ 50 ms followed by two or more spikes with ISIs $\leq$ 6 ms.

The retinal filters learned by the CH model from the awake data (Figure 2.3E) closely resembled those learned from the anesthetized data, as expected from the RH model results (Figure 2.2, D & F). However, the LGN filters learned from the awake data show a very different pattern. Instead of the short latency, positive component that appears to be due in large part to LGN bursting (see above), the LGN filters for seven of the eight pairs of the awake data set show a clear, negative component over the same time span (~-0.03 - 0.0 seconds preceding the target spike). Two aspects of this observation are worth noting. First, LGN cells in the awake data set produced very few bursts. Averaged across pairs only 0.235% (median 0%, range [0.0, 1.52]) of LGN spikes were part

of bursts, with five of the eight producing no bursts at all by the accepted definition ((Lu et al., 1992), see Burst spike definition). In comparison, across pairs from the anesthetized data sets the median percentage of spikes that were part of bursts was much higher: 14.203% (MAD 9.698, 95% CI [9.223, 17.496]) for the binary white noise data set, and 18.264% (MAD 14.805, 95% CI [9.792, 27.777]) for the drifting grating data set. Thus, the lack of the positive component seen in the anesthetized data is expected. Second, the negative component of the LGN filters suggests that some form of gain control or normalization is occurring. This follows from the construction of the model, negative LGN filter weights over some time interval indicate that LGN spikes that occur during that interval will push the model towards predicting that the target spike will not be relayed, thus lowering the activity of the LGN cell itself and producing a gain control or normalization-like effect (i.e., the same retinal input produces a smaller magnitude response when the LGN has just been active compared with when it has just been quiescent (Shapley and Enroth-Cugell, 1984)).

As with previously discussed models, the performance of the CH model was assessed using 10-fold cross-validation procedure. For binary white noise data median $\mathcal{I}_{\text{Bernoulli}}$ across pairs was 0.033 bits/spike (MAD 0.030, 95% CI [0.015, 0.064]), and for drifting grating data median $\mathcal{I}_{\text{Bernoulli}}$ was 0.073 bits/spike (MAD 0.051, 95% CI [0.051, 0.134]). For the awake data set, median $\mathcal{I}_{\text{Bernoulli}}$ was 0.263 bits/spike (MAD 0.083, range [0.086, 0.489]). Consistent with the idea that CH-LGN filters may be capturing the effect of LGN bursts in the anesthetized data set, we observed that the gain in performance of CH models compared to RH models across pairs was fairly well correlated with the "burstiness" of the LGN cell of each pair. The Spearmann correlation between $\mathcal{I}_{\text{Bernoulli}}$ difference (CH - RH) and percentage of LGN spikes that were part of bursts (not including cardinal spikes, see Burst spike definition) was 0.500 (95% CI [0.182, 0.726], $p < 0.01$) for binary white noise data, and 0.286 (95% CI [-0.087, 0.569], $p \approx 0.1$) for drifting grating data (Supporting Figure 6.1C).

2.5.4 Model comparison

In order to illustrate how well each model performed relative to the others we first examined how well the model-predicted efficacies correlated with the observed efficacies. To do this we grouped the retinal spikes from each pair according to their predicted efficacy (normalized by the mean efficacy of that pair), calculated the observed efficacy for each group (also normalized within-pair by that

pair's mean efficacy), and then plotted the normalized, observed efficacy against the normalized, predicted efficacy. Efficacies, both predicted and observed, for each pair were normalized by the observed mean efficacy of that pair (across all spikes) to account for the large difference in efficacy across pairs as is typically done (e.g (Alitto et al., 2019a, 2019b)). In such a framework, a well performing model will produce a "unity" line with a slope of one and y-intercept of zero (i.e., predicted efficacy and observed efficacy match). The left column of Figure 2.5 shows, for each data set, the median relationship between observed and predicted efficacy for each model (error bars represent the median absolute deviation (MAD) across pairs). While all models appear to perform quite well within this framework, there is a systematic trend for the ISI-efficacy model to perform worse for the spikes that it predicts to have the highest efficacy within the drifting grating and binary white noise data sets. Given that the highest efficacy spikes should follow short ISIs (see e.g.,Figure 2.5), this suggests that the ISI-efficacy model may be performing worse than the GLMs specifically for short ISI spikes. Consistent with this suggestion, the right column of Figure 2.5 shows that the difference in $\mathcal{I}_{\text{Bernoulli}}$ between the GLM and ISI-efficacy models is most pronounced for retinal spikes with the shortest ISIs within the drifting grating and binary white noise data sets. Interestingly, within this comparison framework the ISI model appears to perform as well as the GLMs on the awake data set.

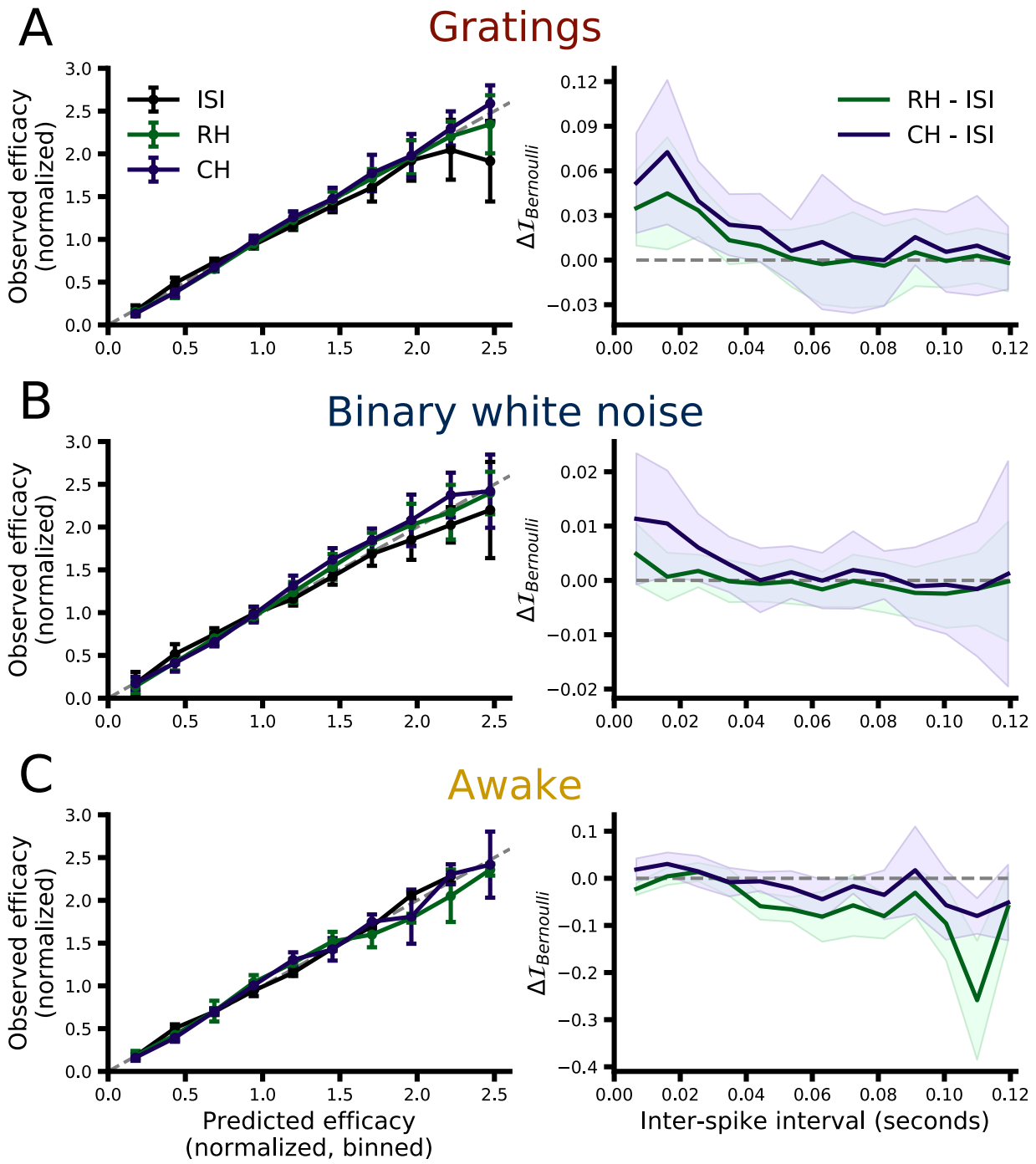


Figure 2.5: Qualitative comparison of model performance. **A left**: The predicted efficacies from each model were used to group retinal spikes into bins, and the observed efficacy for each group (median across pairs) is plotted against the corresponding bin label (error bars represent the MAD across pairs). Both predicted and observed efficacies from each pair were normalized by the mean efficacy of that pair prior to calculating the median and MAD. **A right**: The performance ($\mathcal{I}_{\text{Bernoulli}}$) of the GLMs relative to the ISI-efficacy model is shown as a function if ISI. Lines show the median performance difference across pairs; shading represents the MAD. **B & C**: Same as **A** but for the binary white noise (**B**) and awake (**C**) data sets.

While Figure 2.5 provides a helpful overview of model performance, given the present context the most rigorous way to assess the performance of the models presented here is using $\mathcal{I}_{\text{Bernoulli}}$, the cross-validated single-spike Bernoulli information, which quantifies the accuracy of model predictions on a spike-by-spike basis. Figures 2.6 and 2.7 summarize the results of a direct model comparison analysis for the binary white noise and drifting grating data respectively, in which all hyperparameters for all models were optimized individually for each pair (see Hyperparameter optimization). The top row of each figure shows the cross-validated $\mathcal{I}_{\text{Bernoulli}}$ for each pair and each model, where points corresponding to the same pair are connected, and the bottom row shows a bootstrap estimation of the paired median difference in $\mathcal{I}_{\text{Bernoulli}}$ between models (see Statistics). For the binary white noise data (Figure 2.6), the paired median difference between ISI-efficacy and RH models was 0.002 bits/spike (MAD 0.003 95% CI [0.000, 0.003], $p \approx 0.0092$)[a], between ISI-efficacy and CH models was 0.009 bits/spike (MAD 0.008 95% CI [0.004, 0.015], $p \approx 0.0002$)[b], and between RH and CH models was 0.004 bits/spike (MAD 0.004 95% CI [0.003, 0.009], $p \approx 0.0002$)[c]. For the drifting grating data (Figure 2.7), the paired median difference between ISI-efficacy and RH models was 0.030 bits/spike (MAD 0.020 95% CI [0.012, 0.047], $p \approx 0.0002$)[d], between ISI-efficacy and CH models was 0.049 bits/spike (MAD 0.033 95% CI [0.032, 0.080], $p \approx 0.0002$)[e], between RH and CH models was 0.020 bits/spike (MAD 0.014 95% CI [0.006, 0.027], $p \approx 0.0002$)[f].

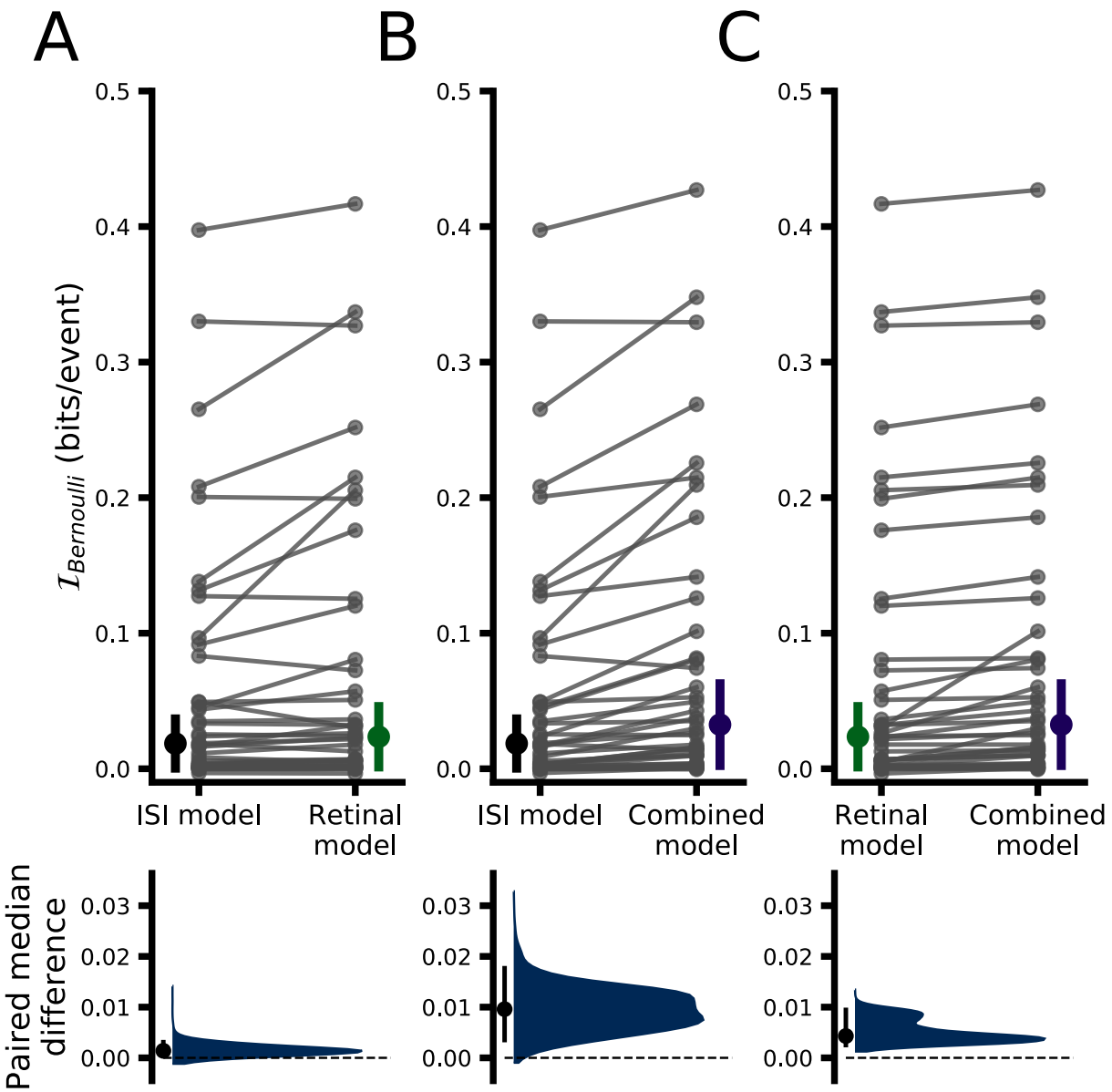


Figure 2.6: Performance comparison of all models for binary white noise data. **A upper**: Comparison of ISI efficacy and RH models. Each dot indicates the mean $\mathcal{I}_{\text{Bernoulli}}$ for a given pair and model; lines connect data belonging to the same pair across models (thus the slope of the lines depicts the change in $\mathcal{I}_{\text{Bernoulli}}$). The height of the vertical, colored bars indicates the median absolute deviation (MAD) of $\mathcal{I}_{\text{Bernoulli}}$ across pairs for a given model, with the filled circle indicating the median value. **A lower**: Estimated paired median difference $\mathcal{I}_{\text{Bernoulli}}$ between ISI efficacy and RH models. The black dot indicates the observed paired median difference and the vertical black line indicates the 95% confidence interval (CI) of the bootstrap distribution (5000 samples) shown in blue. **B**: Same as **A** but comparing ISI efficacy and CH model performance. **C**: Same as **A** but comparing RH and CH model performance.

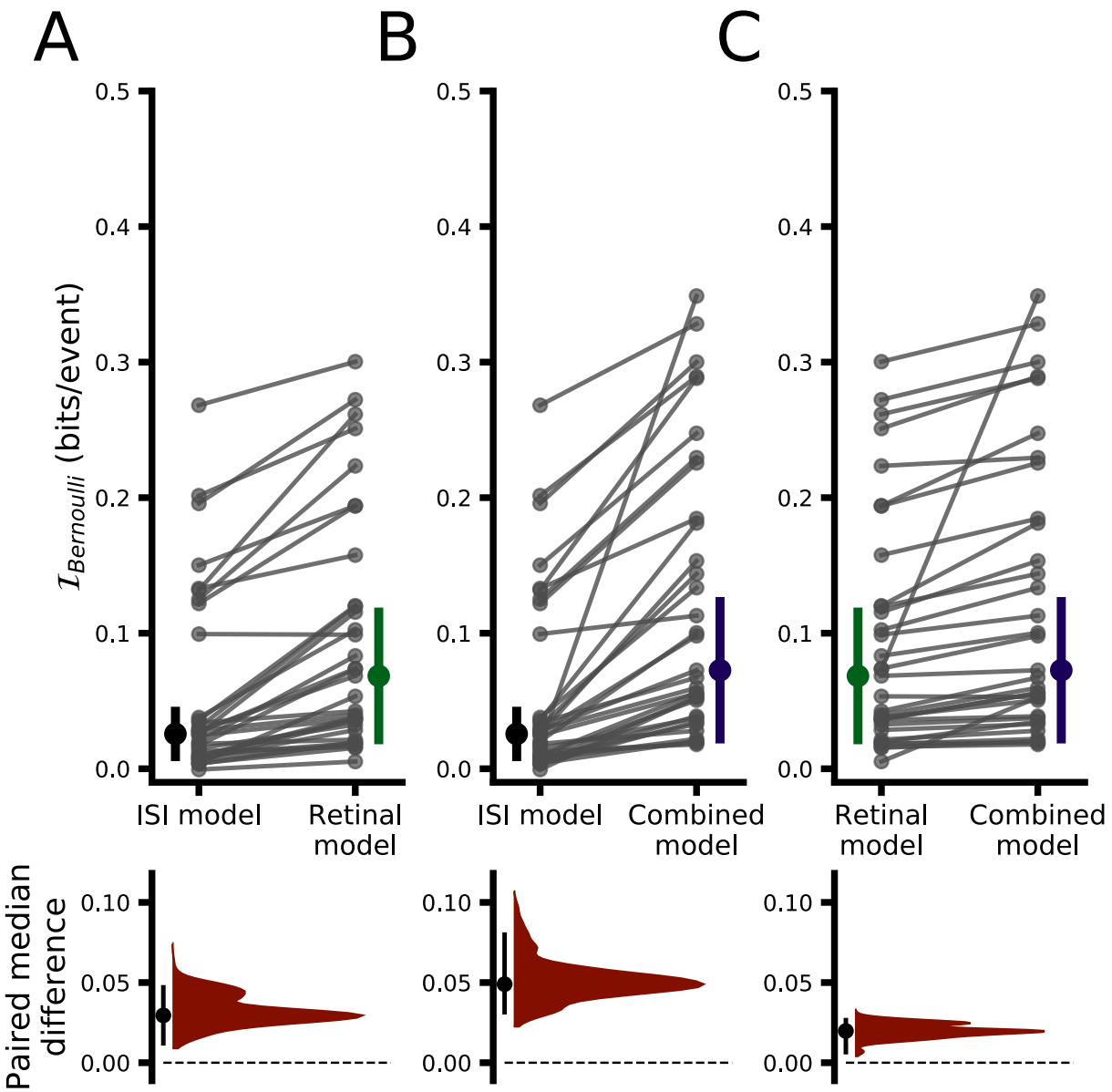


Figure 2.7: Performance comparison of all models for drifting grating data. All conventions exactly follow those from Figure 6. Correlates of model performance are shown in Supporting Figure 6.1. Model performance for the awake data set is shown in Supporting Figure 6.2.

While the small size of the awake data set precludes a statistical comparison of model performance, a qualitative assessment shows largely the same pattern as seen in the anesthetized data. Supporting Figure 6.2A illustrates the pairwise difference in model performance between the three models (ISI, RH and CH) which suggests that although no difference between the performance of the ISI and RH models is evident, the inclusion of the LGN filter in the CH model may substantially

improve performance (median pairwise difference in $I_{\text{Bernoulli}}$ between RH and CH models was 0.058 bits/spike, range [-0.004, 0.170]).

Overall, while RH models do show significantly better performance than ISI-efficacy models, and CH models significantly outperform RH models, the magnitude of the performance gain is rather modest, suggesting that, overall, retinal ISI is the dominant factor in determining which retinal spikes are relayed. However, while both stimulus conditions showed this trend, the magnitude of the performance gain associated with RH and CH models over the ISI-efficacy model was substantially larger when pairs were stimulated with drifting gratings, suggesting that some subtler aspects of LGN integration may differ between the two stimulus conditions (e.g.,Figure 2.2D & 2.3C, (Usrey et al., 1998)).

2.5.5 Integration dynamics depend on firing rate

One potential drawback of using GLMs in the present context is that by fitting a single set of filters to all spikes (or a random subset), we are asking the fitting algorithm to find what amounts to the average integration behavior of relay cells during the recording period. The analysis is, by design, insensitive to any changes in relay cell integration that may occur within a stimulus condition. While this implicit assumption of stationarity may be largely valid for the binary white noise stimulus, it may not hold during drifting grating stimulation due to the high degree of spatial and temporal correlations present in drifting gratings, which are of course absent from the binary white noise. The strong correlations present in drifting gratings may result in larger fluctuations in activity for both the RGC-LGN cell pair being recorded as well as the wider network (including e.g., the thalamic reticular nucleus, V1, etc.) and thus may alter LGN integration dynamics in a more significant manner. Consistent with this idea, we observed higher RGC firing rate variability during drifting grating stimulation in the 200 ms period immediately preceding each retinal spike (the same time period that the model could consider): median pairwise difference in firing rate standard deviation (gratings minus binary white noise) was 3.549 spikes/second (MAD 4.581, 95% CI [0.152, 5.844]) (median 16.768 and 11.587 spikes/second for gratings and binary white noise respectively). The models presented thus far are not sensitive to these potential within-condition changes, as each model is fit to all spikes (or a randomly selected subset) from a single stimulus

condition. Thus, we sought to investigate specifically whether LGN integration dynamics might differ based on the level of activity by assigning each retinal spike to one of four "quartile" subsets (Q1 - Q4) of the data based on the quartile into which the LGN spike count in a 100 ms window preceding each retinal spike fell (see Classification of retinal spikes by activity level). We then fit separate GLMs to the data from each quartile for each stimulus type. For this analysis we consider only RH models, as the quartile partitioning results in too few LGN spikes in the lowest activity quartile to reliably fit CH models. Additionally, the binary white noise data from one pair (pair ID 102) did not contain enough spikes to reliably fit RH models for each quartile and was excluded from activity level analyses. Figure 2.8 shows the filters learned by the model for each activity level and stimulus condition averaged across pairs (filters from each pair were scaled to have unit norm before averaging) where the shaded regions represent the 95% confidence interval (CI) across pairs (see Filter visualization). For binary white noise (Figure 2.8A), there is an apparent trend towards a small difference between approximately 40 and 120 ms preceding the target spike (at time=0) such that retinal spikes during that window may have a somewhat stronger negative influence on relay probability (i.e.,push the model to predict "not relayed") during epochs of heightened activity (Q3 and Q4); however, the magnitude and variability of this effect (as seen in the overlapping CI shading) suggest little qualitative difference between activity levels. On the other hand, filters learned from drifting grating data show a much clearer difference between activity levels, specifically within a time window approximately 5 ms to 20 ms before the target retinal spike (Figure 2.8C and inset), such that the filters learned from high activity data (Q3 & Q4) show a faster decay towards zero from the initial positive peak immediately preceding the target retinal spike. This difference suggests a narrowing of the effective integration window of LGN relay cells during epochs of elevated activity. Importantly, this difference is unlikely to be due to differences in the ability of the model to fit the different data sets (Figure 2.8D), as median paired difference in $\mathcal{I}_{\text{Bernoulli}}$ between models fit to data from the highest (Q4) and lowest (Q1) activity levels was -0.005 bits/spike (Q4 - Q1, MAD 0.041, 95% CI [-0.047, 0.003], $p \approx 0.353$)[g]. Model performance was also not significantly different between Q4 and Q1 subsets for the binary white noise data set: median paired difference in $\mathcal{I}_{\text{Bernoulli}}$ was 0.001 bits/spike (MAD 0.008, 95% CI [-0.001, 0.004], $p \approx 0.396$)[h].

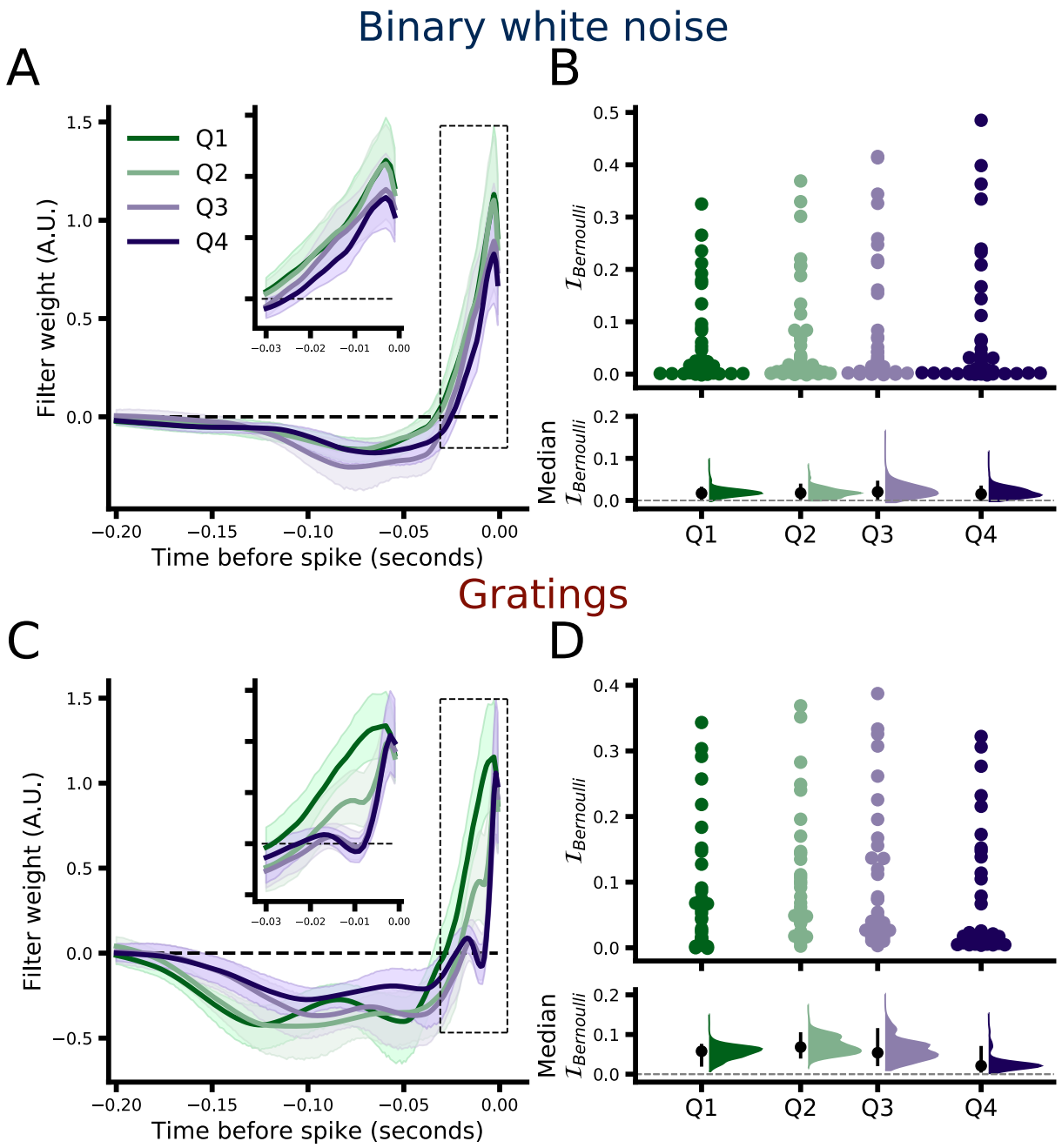


Figure 2.8: Comparison of RH models fit separately to subsets (quartiles) of the data grouped by LGN activity level. **A**: Average retinal filters from RH models fit to each quartile of the binary white noise data set from low (Q1, green) to high (Q4, purple) based on the activity level of the LGN neuron within a 100 ms period directly preceding the target retinal spike at $t = 0$. Shading represents 95% CI across N=38 pairs. **B Upper**: Comparison of model performance ($\mathcal{J}_{\text{Bernoulli}}$) across all activity subsets. Each dot represents the model performance for a single pair (the spread along the x-axis is to aid visualization). **B Lower**: Bootstrap estimation of median model performance for each subset. Black dots indicate the median across pairs and black vertical lines indicate the 95% CI of the bootstrap distribution (shown in color, 5000 samples). **C, D** Same as **A, B** but for the drifting gratings data set (N=33). Results from a control analysis wherein relay status was simulated via GLMs is shown in Supporting Figure 6.3 (see main text for details). Results of changing the spike quartile classification window are shown in Supporting Figure 6.4.

One potential concern with the above analysis is that the data used to train the model differed considerably between quartiles. Although the quartiles are defined based on LGN firing rates, retinal firing rates will of course be highly correlated. Thus, the observed difference in LGN integration dynamics could be due entirely to differences in the training data. To control for this possibility we use a single, fixed filter learned from all the data from a given pair (i.e.,the filters shown in Figure 2.2) to simulate the relay status of each retinal spike (i.e.,the pattern of retinal spikes preceding each target spike is convolved with the learned filter, the output of which is passed through the logistic function and relay status is determined by a coin flip, see Simulating GLMs). We then performed the quartile subsetting and model fitting exactly as for Figure 2.8. The learned filters for each stimulus type and activity quartile are shown in Supporting Figure 6.3. Importantly, in this case the training data have exactly the same quartile related differences as for the original analysis, the only difference is that the integration dynamics of the LGN cell are fixed via the simulation. Thus, the fact that the filters learned from all the quartile subsets are highly overlapping suggests that the differences observed in Figure 2.8C are not due to differences in the training data alone. The overlap in the learned filters is especially apparent through the first ∼30-50 ms where the most striking difference in Figure 2.8C can be seen.

The finding that LGN integration dynamics depend on firing rate proved to be robust to the precise time window used to classify activity levels (tested over a range spanning 50 - 200 ms, see Supporting Figure 6.4C-D); however, using time windows close to the cycle duration of the drifting grating (i.e., around 250 ms) is likely to produce a severe underestimate of the real difference as it would effectively average over the preferred and non-preferred phases of the drifting grating (which is the likely cause of the higher variability in firing rate observed during drifting grating stimulation). Consistent with this idea, repeating the analysis using a 250 ms time window to partition the data into quartiles substantially reduced the difference between filters learned from the drifting grating data (filters learned from binary white noise data continued to show no difference, see Supporting Figure 6.4A-B).

The awake data set did not contain a sufficient number of spikes to perform the quartile subsetting procedure that we used for the anesthetized data set (median number of retinal spikes per-pair in the awake data set was 2,017.0 (MAD 427.5), while anesthetized data sets had a median of

12,303.5 (MAD 5,624.0) and 38,425.0 (MAD 18,150.0) for the binary white noise and drifting grating data sets respectively). Thus, we used a median split to assign each retinal spike from each pair to a low or high activity subset. The filters learned from low and high subsets showed little difference (Supporting Figure 6.2B), similar to what was seen in the binary white noise (anesthetized) data although lacking the prolonged negative component (between approximately -90 to -120 ms). Interestingly, the one pair that does appear to show a more substantial difference between filters learned from low and high activity data (pair 200001250) was stimulated with gratings during recording (see Discussion).

To quantify the apparent differences in filters learned from the highest (Q4) and the lowest (Q1) activity data (Figure 2.9C), we calculated the integral of the absolute difference between the Q1 and Q4 filters for each pair (see Classification of retinal spikes by activity level). The distribution of the paired absolute differences, along with kernel density estimates, for each stimulus condition are shown in Figure 2.9A with the corresponding estimation of the median of each distribution show in Figure 2.9B. For the binary white noise data set the median absolute difference between Q4 and Q1 was 0.024 (MAD 0.010, 95% CI [0.018, 0.028]), and for drifting grating it was 0.055 (MAD 0.026, 95% CI [0.033, 0.067]). For the filters learned from simulated data (Supporting Figure 6.3), the median absolute difference was 0.002 (MAD 0.001, 95% CI [0.002, 0.003]) and 0.005 (MAD 0.003, 95% CI [0.003, 0.007]) for binary white noise and drifting grating data respectively.

A paired permutation test including only pairs for which both binary white noise and drifting grating data were available (N=27) confirmed the differences between the two stimulus conditions: paired median difference (drifting gratings minus binary white noise) in absolute difference was 0.031 (MAD 0.016 95% CI [0.018, 0.039], $p \approx 0.0002$)[i]. Repeating the analysis when only including the 30 ms period preceding the target retinal spike yielded similar results (paired median difference of 0.008, MAD 0.007 95% CI [0.002, 0.012], $p \approx 0.0004$)[j].

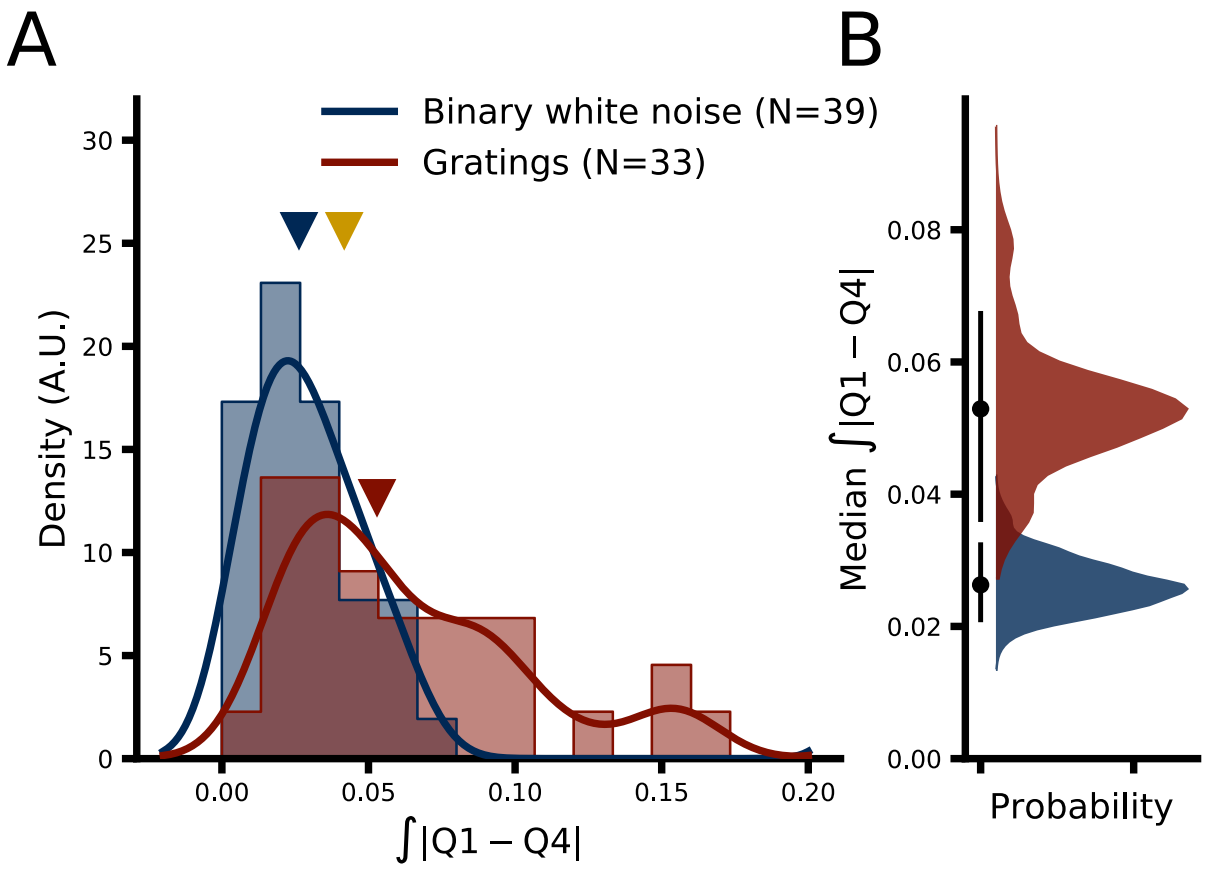


Figure 2.9: Quantification of differences between filters learned from highest (Q4) and lowest (Q1) activity data sets. **A**: Population distributions (filled bars) and kernel density estimates (thick lines) of absolute differences between Q4 and Q1 filters for binary white noise (blue) and drifting grating (red) data. Filled triangles denote the median of each distribution. The gold triangle indicates the median difference for the awake data set for reference (where "high" and "low" were defined by a median split due to fewer spikes in that data set). **B**: estimation of population medians from A. Filled black dots indicate the median and black vertical lines indicate the 95% CI of the bootstrap distributions of population medians shown in blue (red) for binary white noise (drifting grating) data.

2.6 Discussion

The aim of this study was to investigate how LGN relay cells integrate their retinal inputs over time, and how the integration process changes under different stimulus and network conditions, by using computational models to predict which retinal spikes were relayed on to V1 and which were not. We model retinogeniculate transmission as a coin flip (or Bernoulli) process where the primary quantity of interest is the probability, p, that each incoming retinal spike will be relayed. In the simplest

	Data set	Metric	Conditions	Paired median difference	MAD	95% CI	p-value
Figure 2.6							
a	Binary Noise (N=40)	$\mathcal{I}_{\text{Bernoulli}}$	RH - ISI	0.002 bits/spike	0.003	[0.000, 0.003]	0.0092
b	Binary Noise (N=40)	$\mathcal{I}_{\text{Bernoulli}}$	CH - RH	0.009 bits/spike	0.008	[0.004, 0.015]	0.0002
c	Binary Noise (N=40)	$\mathcal{I}_{\text{Bernoulli}}$	CH - ISI	0.004 bits/spike	0.004	[0.003, 0.009]	0.0002
Figure 2.7							
d	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	RH - ISI	0.030 bits/spike	0.020	[0.012, 0.047]	0.0002
e	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	CH - ISI	0.049 bits/spike	0.033	[0.032, 0.080]	0.0002
f	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	CH - RH	0.020 bits/spike	0.014	[0.006, 0.027]	0.0002
Figure 2.8							
g	Gratings (N=33)	$\mathcal{I}_{\text{Bernoulli}}$	Q4 - Q1	-0.005 bits/spike	0.041	[-0.047, 0.003]	0.353
h	Binary Noise (N=39)	$\mathcal{I}_{\text{Bernoulli}}$	Q4 - Q1	0.001 bits/spike	0.007	[-0.001, 0.004]	0.396
Figure 2.9							
i	Anesthetized (N=27)	Absolute difference	Gratings - Noise 100ms	0.031	0.016	[0.018, 0.039]	0.0002
j	Anesthetized (N=27)	Absolute difference	Gratings - Noise 30ms	0.008	0.007	[0.002, 0.012]	0.0004

Table 2.1: Statistical table of results. Confidence intervals are derived from 5,000 bootstrap resamples and are bias corrected and accelerated. p-values are derived from paired-permutation tests with 5,000 permutations. See Methods for details.

possible model p is a constant given by the mean efficacy across all retinal spikes recorded from a given RGC-LGN cell pair. This constant p model (or homogeneous Bernoulli model) forms the basis of comparison for all other models that we considered, as the constant p model captures the fact that as mean efficacy approaches the extremes (0 or 1) predicting relay status becomes trivial (simply guessing the mean will approach perfect performance). Thus, we chose to quantify model performance in terms of the cross-validated single-spike Bernoulli information ($\mathcal{I}_{\text{Bernoulli}}$) which quantifies how informative model predictions are about the relay status of retinal spikes (that were not "seen" during model fitting) relative to a homogeneous model. In our construction, $\mathcal{I}_{\text{Bernoulli}}$ has units of bits/spike and can take on values between ~0 and 1, where 0 represents performance no better than a constant p model and 1 represents perfect performance (see Assessing model

performance for details).

The fact that $\mathcal{I}_{\text{Bernoulli}}$ quantifies model performance relative to a homogeneous model is critical given the present context of trying to predict the relay status of retinal spikes. This follows from the fact that the difficulty of predicting relay status varies with mean efficacy: relay status is trivially easy to predict for pairs with a mean efficacy close to zero or one, and is maximally difficult for pairs with a mean efficacy of ~0.5. Thus, an optimal performance metric needs to take into account both the quality of the predictions as well as the difficulty of the task for a given pair. $\mathcal{I}_{\text{Bernoulli}}$ does exactly this. However, as a result the maximum $\mathcal{I}_{\text{Bernoulli}}$ achievable for pairs with very low or very high mean efficacy is substantially less than one. This fact accounts in part for the low $\mathcal{I}_{\text{Bernoulli}}$ values achieved by the models considered here, especially on the anesthetized data sets where many pairs have low mean efficacies (9% and 25% of pairs from the drifting grating and binary white noise data sets, respectively, have a mean efficacy less than 0.05). It should be noted that this behavior is not a deficiency in the $\mathcal{I}_{\text{Bernoulli}}$ metric, rather it reflects an inherent difficulty in predicting relay status.

We first considered a model where p varies in time according to the elapsed interval since the last retinal spike (ISI) based on extensive evidence that retinal spikes following shorter ISIs are more likely to be relayed due to temporal summation (Usrey et al., 1998; Sincich et al., 2007; Weyand, 2007; Casti et al., 2008; Sincich et al., 2009; Wang et al., 2010; Rathbun et al., 2016; Alitto et al., 2019a). Following the framework of Wang et al. (2010), we formalize this observation as a simple model, $p = f(ISI)$, where the relation between ISI and relay probability (i.e., efficacy) is learned from a subset of the data (training set) and the performance of the model is tested on a separate subset (testing set, see Assessing model performance).

We further considered a model where p is a function of the pattern of retinal spikes that an LGN cell receives within a given window of time, what we call the retinal history (RH) model. Conceptually, this can be seen as an extension of the ISI-efficacy model that additionally takes into account the notion that the influence of retinal activity on the current state of a relay cell (i.e., its propensity to relay a retinal input should one arrive) is unlikely to be limited to just the most recent retinal spike. Thus, allowing a model to consider the full pattern of recent spikes from the recorded RGC should improve predictions of relay status and provide a less constrained view of the temporal integration

dynamics of retinogeniculate interactions. To that end we utilized Bernoulli-Logistic generalized linear models to predict the relay status of each retinal spike based on the convolution of a learned temporal filter (retinal filter) with the pattern of recent retinal activity, the output of which is then mapped to a predicted relay probability (or equivalently, predicted efficacy).

In comparing the parameters learned by the ISI-efficacy and RH models, one critical difference between the models is worth noting. For the ISI-efficacy model, relay probability is modeled as a univariate, nonlinear function of ISI, while the RH model is a linear function of the multivariate pattern of retinal spikes over a given time window (which is then passed through a logistic nonlinearity). Thus, the similarity of the ISI-efficacy functions and RH retinal filters presented in Figure 2.2 should be interpreted carefully. However, the rapid decay of both functions does tell a consistent story, namely that the time windows over which retinal spikes positively interact (i.e., promote a relay probability above the mean) is approximately 20 - 30 ms regardless of the stimulus (gratings or binary white noise) or the state of the animal (anesthetized or awake). This likely accounts for the observation that the RH model only outperforms the ISI-model by the smallest of margins in the anesthetized data (Figures 2.5, 2.6 and 2.7), and not at all in the awake data (though the small size of the awake data set should be noted).

The final model that we considered was a further augmented version of the RH model that included a second, learned temporal filter (LGN filter) that operated on the recent activity history of the LGN cell, what we call the combined history (CH) model. As stated previously, for RH models the LGN activity is only used to identify the relay status of each RGC spike, and thus the LGN spike train (and, in particular the LGN spikes not triggered by the recorded RGC) may provide additional information that can help predict the relay status of retinal spikes. While the CH model did outperform the other two models for all data sets tested here, further analysis of the correlates of performance and consideration of the shape of the learned filters suggests that the improvement may be based on different features within the anesthetized and awake data sets. In particular we found that, for the anesthetized data set the improvement in performance between RH and CH models was correlated with the degree of "burstiness" (i.e., the percentage of LGN spikes that were part of bursts) of the LGN cells of the pairs (Supporting Figure 6.2). Furthermore, the shape of the LGN filters, large positive values at very short pre-target-spike latencies, suggests that the

model is capturing the increase in retinal efficacy that occurs during geniculate bursts (Alitto et al., 2019b), and this component of the LGN filters was specifically attenuated when non-cardinal burst spikes were removed from the data prior to CH model fitting (Supporting Figure 6.1). In contrast, the LGN filters learned from the awake data cannot be accounted for by bursts, as burst were extremely rare in the awake data set. Instead, the negative component seen between ~-40 and 0 ms (Figure 2.3) likely reflects the influence of a gain control or normalization mechanism that could result from intrathalamic negative feedback through the thalamic reticular nucleus (TRN) (or perhaps the longer LGN → V1 → TRN → LGN loop). Across the analyses that we performed, this was the only clear difference between the awake and anesthetized data sets.

Lastly, we asked whether relay cell temporal integration dynamics might differ depending on the level of activity within the retinogeniculate circuit, and whether that difference is seen for both stimulus conditions in the anesthetized data. To that end we assigned each retinal spike to one of four data subsets based on the quartile of LGN activity during the preceding 100 ms (see Classification of retinal spikes by activity level) and fit RH models separately to each data subset. We specifically chose to use LGN activity to partition retinal spikes as, although retinal and geniculate activity levels are highly correlated, LGN activity is likely to be more indicative of the activity level of the wider retino-thalamo-cortical circuit. For binary white noise data, learned temporal filters showed little difference between subsets (Figure 2.8A), while for drifting grating data a substantial difference is observed between approximately 5 and 20 ms (Figure 2.8C) such that filters learned from the highest activity subsets (Q3 and Q4) show a shorter effective temporal integration window (i.e., the duration of time preceding a target spike where the arrival of another retinal spike will increase the likelihood that the target spike is relayed). For the awake data set, most pairs showed little difference between epochs of higher and lower activity when analyzed in a similar manner (albeit using a simpler median split as there were not enough spikes to reliably fit model to quartile subsets). Interestingly, the one apparent exception (pair 200001250, Supporting Figure 6.2B) was also the only pair that was stimulated with gratings during recordings. While this is a single example and so should be considered only the slimmest of evidence, it is nonetheless consistent with the idea that the effective integration window of LGN cells, in both the awake and anesthetized states, is dynamically regulated in a manner that is inversely proportional to the ongoing firing rate (i.e.,

shorter integration windows during periods of higher activity).

While there are several cellular and circuit mechanisms that could underlie the shortening of the temporal integration window, such as spike rate adaptation within relay cells, short-term depression at the retinogeniculate synapse, feedforward inhibition from geniculate interneurons, feedback inhibition (direct or indirectly via cortex) from the thalamic reticular nucleus, or a change in oscillatory activity coming from the retina (Koepsell et al., 2009), the functional consequence of this process is a form a gain control wherein the specificity of geniculate filtering scales with activity level. The idea being that, under lower levels of activity the LGN behaves more permissively and relays patterns of retinal spikes that under higher activity conditions, where the LGN is less permissive, would not be relayed. This process might offer an explanation for several observations about retinogeniculate transmission, such as the finding from Alitto et al. (2019a) that retinal efficacy following ISIs in the ~5 to ~25 ms range is higher under low contrast (and thus low activity) than high contrast (and thus high activity) stimulus conditions. Likewise it could potentially explain the finding from Rathbun et al. (2016) that as the contrast of a drifting grating stimulus increases, responses of LGN cells shift to progressively earlier phases of the stimulus cycle and that the rate of this "phase advance" is higher in relay cells compared to their direct retinal inputs. Further work is needed to address whether the magnitude of the integration widow shortening that we observe here quantitatively matches the observations listed above.

2.6.1 Relationship to previous work

A considerable amount of effort has been put into modeling the computations performed by relay cells of the LGN, due in large part to the fact that simultaneous recordings of both a dominant input (from RGCs) and the output (LGN spiking) is possible. Prior work on modeling retinogeniculate interactions can be coarsely grouped into two approaches: those that focus on LGN processing of retinal spike trains in the absence (Casti et al., 2008; Heiberg et al., 2013), or presence (Norheim et al., 2012) of extra-retinal input, and those that include an additional channel for processing the visual stimulus directly (Babadi et al., 2010; Butts et al., 2016). The logic of including the additional stimulus channel is that it enables models to capture stimulus driven effects that are not mediated by the direct retinal input, so that "indirect" effects (e.g., from cortical or TRN feedback) might

be uncovered. While this is a powerful approach to studying geniculate computations generally, we instead chose to focus our efforts more narrowly on modeling how LGN cells process individual retinal inputs by trying to predict which retinal spikes were relayed and which were not. This approach is particularly well suited to our data, which consists primarily of recordings of RGC-LGN cell pairs in which the RGC spikes were recorded within the eye. This entails that 1) we can be confident that few, if any, RGC spikes went undetected, and 2) that most of our recordings were made from non-dominant RGC inputs. The second point follows from the observation that most relay cells in the cat receive input from two to five RGCs (Cleland et al., 1971; Hamos et al., 1987; Usrey et al., 1999; Martinez et al., 2014), and thus landing an extracellular electrode in the vicinity of the dominant input should be somewhat rare. Conversely, S-potential recordings are likely to reflect just the dominant input (Kaplan and Shapley, 1984; Weyand, 2007). Consistent with this idea, we observed considerably higher mean efficacies in the awake data set (on average ~0.52) compared to either the drifting grating (~0.16) or binary white noise (~0.1) data sets from the anesthetized animal. Given the above, we reasoned that the most fruitful approach would be to focus on predicting the relay status of the retinal spikes that we did record and avoid making predictions about LGN spikes that were not triggered by the RGC under study.

Overall, this approach emphasizes the computations being performed by relay cells on individual retinal inputs. Previous work has proposed that the core of these computations is well approximated by linear filtering with an exponential kernel (Casti et al., 2008; Heiberg et al., 2013) as suggested by the strong relationship between retinal efficiency and retinal ISI (Usrey et al., 1998; Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008; Sincich et al., 2009; Uglesich et al., 2009; Rathbun et al., 2010; Wang et al., 2010). The strength of taking a statistical approach, as we do here, is that the form of the linear filter is directly learned by the model. Our results confirm that an exponential filter is indeed a good model of relay cell temporal integration and, given the relatively short apparent time constants (on the order of 10 - 20ms, consistent with Casti et al. (2008)), suggest that the retinal ISI is likely to be the strongest single influence on whether a given retinal spike is relayed or not.

2.7 Conclusion

Overall, our results suggest that the dominant factor that determines whether or not a given RGC spike is relayed to cortex by the LGN is the retinal ISI, confirming previous findings (Usrey et al., 1998; Carandini et al., 2007; Sincich et al., 2007; Casti et al., 2008; Sincich et al., 2009; Uglesich et al., 2009; Rathbun et al., 2010; Wang et al., 2010). However, quantitatively smaller, yet still likely important, contributions were observed for retinal activity further into the past, as well as LGN activity patterns indicative of periods of burst firing. Furthermore, we have demonstrated that the time scale over which the LGN integrates its retinal inputs changes as a function of the level of activity within the retino-thalamo-cortical circuit. This finding raises the possibility that gain control (Shapley and Enroth-Cugell, 1984), a core visual function of the LGN (Alitto et al., 2019a), could be achieved in part by modulating the temporal integration window of LGN relay cells. The source of this modulation remains an open question for future work to explore.

3 Delta oscillations modulate V1 responses in a layer and time dependent manner

3.1 Introduction

The brain is an immense, dynamical system with ongoing, complex internal dynamics that are stable, yet also exquisitely sensitive to incoming information from the external (and internal) environment (McKenna et al., 1994; Cocchi et al., 2017). The sensory pathways are particularly clear examples of this, where extensive evidence indicates that incoming information both shapes, and is shaped by, the patterns of neural activity moving through sensory circuits at the moment when stimulus information arrives (Haenny and Schiller, 1988; Arieli et al., 1996; Cano et al., 2006; Gilbert and Sigman, 2007; Issa and Wang, 2011; Bennett et al., 2013; Polack et al., 2013; McCormick et al., 2015; McGinley et al., 2015; Schreck et al., 2022). In particular, numerous studies have shown that the amplitude and phase of ongoing neural oscillations in sensory cortex modulates how those circuits respond to stimuli and shape the concomitant perceptions (Busch et al., 2009; Benwell et al., 2017; Iemi et al., 2017; Iemi et al., 2019; Dou et al., 2022). While a great deal of this work has focused specifically on the link between the amplitude and phase of alpha oscillations (8 - 12 Hz) and coarse, population scale measures of neural activity, such as electroencephalography (EEG) (Ergenoglu et al., 2004; Klimesch et al., 2007; Romei et al., 2008; Busch and VanRullen, 2010; Mathewson et al., 2011), less is known about how ongoing oscillations in other frequency bands modulate neural activity at the scale of circuits or a cortical column. Recent work has shown that in the awake mouse cortex the amplitude of ongoing delta oscillations (2 - 5 Hz) in the local field potential (LFP) correlate quite strongly with fluctuations in the membrane potential of nearby neurons, and may provide a robust, accessible index of time-varying cortical excitability (Zerlaut et al., 2022).

While Zerlaut et al. (2022) derived a more complex measure of the oscillatory state of the cortex based on the LFP (i.e., more complex than just delta amplitude), in general their findings demonstrate that neural populations within somatosensory and visual cortex of awake mice are more active (i.e., fire more spikes) during periods where the cortical LFP is arrhythmic and delta

oscillations are absent or of low amplitude. More specifically, their data suggest a somewhat weak "U-shaped" relationship such that during periods of especially strong delta oscillations, neural activity is increased relative to periods of intermediate delta. However, the stronger effect showed that spiking activity was substantially higher when the LFP was arrhythmic and delta oscillations were weak or undetectable.

Based on these findings, we sought to investigate whether the amplitude of ongoing delta oscillations modulates how primary visual cortex (V1) responds to thalamic input driven by indirect visual, or direct electrical, stimulation in primates. Based on the work of Zerlaut et al. (2022), we hypothesized that stimuli that arrive during periods of low delta would result in stronger responses than those that arrived during periods of high delta. While much of the previous work looking at the effect of ongoing oscillations on stimulus processing utilized low-resolution recording techniques in humans (e.g., EEG) (Ergenoglu et al., 2004; Romei et al., 2008; Busch et al., 2009; Benwell et al., 2017; Iemi et al., 2017), or high-resolution techniques in mice (Bennett et al., 2013; Polack et al., 2013; McGinley et al., 2015; Zerlaut et al., 2022), here we report results using high-density electrode arrays (Neuropixel probes, (Jun et al., 2017)) to record neural activity from V1 of anesthetized primates. Importantly, the Neuropixel probes are able to span the full depth of primate V1, allowing us to simultaneously record ongoing delta oscillations and the LFP and spiking responses to visual and electrical stimuli from all cortical layers. Furthermore, the high channel-density of the Neuropixel probes enabled automatic identification of laminar compartments (putative cortical layers) based on the current source density (CSD), allowing for a more fine-grained, circuit-focused analysis (Mitzdorf, 1985; Klein et al., 2021). In addition, by using electrical stimulation to directly drive the lateral geniculate nucleus (LGN) and the axons that connect the LGN with V1, we were able to highlight the locations along the depth of the probe that have direct thalamic connections, and thus corroborate the boundaries of the functionally defined compartments and their correspondence with anatomically defined cortical layers. Using these techniques, we find that the amplitude of ongoing delta oscillations does modulate V1 responses, but does so in a manner that is time and layer dependent. Consistent with previous work, our data demonstrate that, for the range of delta synchronization we observed, stronger cortical responses occur during periods of weaker delta oscillations.

3.2 Methods

3.2.1 Animal preparation

Two adult macaque monkeys (Macaca mulatta, 14 kg, 13 year-old male; 9.3 kg, 14 year-old female) contributed data to this study. All experimental procedures conformed to NIH and USDA guidelines and were approved by the Institutional Animal Care and Use Committee at the University of California, Davis.

Anesthesia was initiated with ketamine (10 mg/kg, i.m.) and maintained with sufentanil citrate (0.1-25 μg/kg/hr, i.v.) and propofol (4-6 mg/kg/hr, i.v.) throughout the experiment. Animals were placed in a stereotaxic apparatus, and body temperature, expired CO_2, oxygen saturation, electrocardiogram (ECG), and electroencephalogram (EEG) were monitored continuously to access the health and depth of anesthesia of the animal. If the ECG, EEG, or expired CO_2 indicated a reduced depth of anesthesia the rate of sufentanil and/or propofol infusion was increased. A midline scalp incision was made and wound margins were infused with lidocaine. A small craniotomy was made over the LGN to allow access for recording and stimulation with single electrodes, and a large craniotomy and durotomy were made over opercular V1 to allow access for a the Neuropixel probe. Once all surgical procedures were complete, animals were paralyzed with vecuronium bromide (0.2 mg/kg/hr, i.v.). The pupils were dilated with 1% atropine sulfate, and the eyes were fitted with contact lenses and focused on a tangent screen.

Stimulating electrodes were placed near the dorsal boundary of the LGN (i.e., above or just within the LGN) targeting either the optic radiations (above the LGN) or the relay cell bodies within the LGN.

Prior to insertion of the Neuropixel probe, the pia overlying the insertion site was pricked with a microelectrode under visual guidance using a surgical microscope to allow the probe to smoothly enter the cortex at an angle as close to 90 degrees as possible.

3.2.2 Visual stimulus display

Visual stimuli were generated by a ViSaGe visual stimulus generator (Cambridge Research Systems, Rochester, UK) and presented on a gamma-calibrated CRT monitor at 140 Hz (Sony, Toyko, Japan). For most recordings, the visual stimulus consisted of a full-field, drifting sinewave grating (4 Hz temporal frequency) displayed at 100% contrast. Using the responses of one highly active neuron, the spatial frequency and orientation of the grating were chosen to approximately match the preference of the V1 column under study. On a minority of runs the contrast, orientation, or diameter of the drifting grating were varied across trials while all other parameters were kept at the values stated above. Additionally, a few runs from several of the recording sites used a full-field, alternating black - white stimulus (typically with a 0.25 second period).

3.2.3 Electrical stimulus delivery

On approximately half of the trials on each run one or more electrical shocks (200 μA, mono or bi-phasic, 0.3 - 1.0 ms duration) were delivered through a platinum/iridium stimulating electrode (FHC, Bowdoin, ME) located above the LGN via an A-M Systems Model 2100 isolated pulse stimulator (Sequim, WA). Parameters of electric stimulation were chosen based on previous work involving electrical stimulation of the LGN in primates (Briggs et al., 2013).

3.2.4 Recording apparatus

The timing of visual and electrical stimuli was controlled by custom software written in the Spike2 programming language utilizing a Power 1401 acquisition and control system (Cambridge Electronic Design, Cambridge, UK). Feedback signals indicating the precise timing of visual and electrical stimuli were also recorded by the Spike2 and Power 1401 system for offline analysis.

Neural data was acquired by a Neuropixel Phase 3A (option 3) probe (Jun et al., 2017) and Xilinx Kintex-7 FPGA board (San Jose, CA) and recorded by the SpikeGLX acquisition software (Janelia, Ashburn, VA). Feedback signals indicating the precise timing of visual and electrical stimuli were acquired alongside the Neuropixel data steam by the Xilinx board, allowing for offline

synchronization of SpikeGLX and Spike2 generated files.

3.2.5 Electrophysiological data

Neuropixel phase 3A, option 3 probes contain 960 contacts spread across a 20 μm x 70 μm x 10 mm shank. Of the 960 contacts, 384 can be recorded from simultaneously. All data presented here were recorded from the 384 contacts closest to the probe tip (i.e., the contacts that were deepest in the brain). The data steam from each contact is split into two bands: an LF band (300 Hz lowpass filter, 250x gain, 2.5 kHz sampling rate) and an AP band (300 Hz highpass filter, 500x gain, 30 kHz sampling rate). Ten of the 384 contacts are used for referencing and do not contain brain signals. In addition to data and reference channels, each band contains an unfiltered "sync" channel that carries stimulus timing information (about both electrical and visual stimuli) originating directly from either the pulse stimulator or ViSaGe.

The resulting data set was comprised of 22 recordings from 8 recording sites in opercular V1.

3.2.6 Data analysis

3.2.6.1 Preprocessing

3.2.6.1.1 Shock artifact suppression Prior to any processing, an electrical shock artifact suppression procedure was run on both the AP and LF band data. The general approach used a matching pursuit algorithm (Mallat and Zhang, 1993) to selectively suppress the features of the shock artifact. Matching pursuit is a sparse, greedy, decomposition algorithm that iteratively build a representation of a signal in terms of a linear combination of an overcomplete set (or dictionary) of basis functions (or atoms) that are generally chosen *a priori* (or in some cases learned from the data). Given a matrix (or dictionary) $\mathbf{D}$, whose columns are the atoms, and a signal x_0, the algorithm proceeds iteratively by finding the atom in $\mathbf{D}$ with the largest inner product with x_0, while keeping track of the atom (a_0) and its inner product (c_0) (i.e., $c_0 = a_0 x_0$). The weighted atom is then subtracted from the signal, yielding a new, residual signal:

$$x_1 = x_0 - c_0 a_0 \tag{17}$$

which then serves as the input to the next iteration. These steps are repeated until a stopping criteria is met.

In using matching pursuit for artifact suppression, the two key choices are the atoms and the stopping criteria. As the electrical shocks we used were mono or biphasic square wave pulses, we chose a dictionary consisting of square wave atoms that varied in width (i.e., frequency), number of cycles, and their temporal onset relative to the shock timing (as reported by the sync channel). As an example, within this framework a delta function is a monophasic (half-cycle), positive, width 1 (i.e., 1 sample) square wave. Note that in this context "width" refers to the number of samples in a half-cycle.

The dictionary for suppressing the LF band artifact consisted of 8 monophasic, width 2 (0.8 ms) square waves with temporal onsets varying between 0 and 7 samples (0.0 and 2.8 ms) relative to shock onset, and 8 tri-phasic, width 1 (0.4 ms) square waves with the same set of temporal onsets (i.e., 16 atoms total).

The dictionary for suppressing the AP band artifact consisted of 81 mono-phasic, width 2 square waves with temporal onsets varying between -5 and 75 samples (-0.167 to 2.5 ms) relative to shock onset; 81 tri-phasic, width 1 square waves located at the same temporal positions; 77 bi-phasic square waves comprised of 7 widths between 27 and 33 samples (0.9 and 1.1 ms) and 11 temporal onsets between -5 and 5 samples (-0.167 and 0.167 ms) (i.e., 239 atom total).

The pursuit terminated when the percent change in the l^2 norm of the residual signal between successive iterations fell below a criteria of 10^{-2} and 10^{-3} for LF and AP data, respectively.

Figure 1 shows data from an example shock from an example recording before (left) and after (right) artifact suppression for all channels from the LF (top) and AP (bottom) bands.

3.2.6.1.2 Spike sorting Spike sorting was performed using Kilosort3 (Pachitariu et al., 2023) with all parameters left at their default values as specified in the standard configuration file provided

with the Kilosort3 software. The results of spike sorting were automatically curated using a modified version of the ecephys_spike_sorting package originally developed by the Allen Institute. The automatic curation criteria (Lecoq et al., 2021) specified that spikes were removed when a larger spike was detected within 0.16 ms or within 5 channels (~50 μm), and units were removed if their template had unrealistic spatial spread (only 1 channel, or more than 25 channels), lacked a peak-trough shape, or had multiple spatial peaks.

3.2.6.1.3 LFP preprocessing LF band data was preprocessed by 1) median subtraction to suppress common mode noise, 2) filtering to reduce monitor refresh artifacts and line noise, 3) reference channel interpolation, and 4) averaging channels with the same vertical position. This resulted in a depth-by-time data matrix with 192 rows (20 μm row spacing) spanning 3.840 mm along the probe shank.

Median subtraction

Following Lecoq et al. (2021), the median subtraction procedure grouped every 24th channel together (excluding reference channels) and subtracted from each channel the median signal across the group (i.e., group k consisted of channels $k : 24 : 384 - (24 - k)$ where $k = [1, 2, 3, ..., 24]$).

Filtering

The filtering procedure was performed using zero-phase, forward-reverse filtering as implemented in the DSP.jl software package. A 120 Hz lowpass, Butterworth (4th order) filter to remove artifacts from the CRT monitor refresh (which ran at 140 Hz), and a 60 Hz infinite-impulse response notch filter (2 Hz bandwidth) to reduce line noise.

Reference interpolation & depth averaging

After filtering, data from reference channels were replaced by the mean of the 2 nearest (non-reference) channels, and signals from contacts at the same vertical position along the probe shank were averaged together.

3.2.6.2 Identification of cortical compartments Cortical compartments were identified by running a simple image segmentation algorithm on current source density (CSD) data.

The CSD is given by the negative, second spatial derivative of the voltage data, and reflects the extent to which positive charge flows into (sink) or out of (source) a given point in space (Mitzdorf, 1985). A point in space with a lot of depolarizing neural membrane will thus appear as a sink in the CSD. For primary, sensory cortical areas like V1, the thalamic input layers (most prominently layer 4C) will display a strong, early sink in the CSD data when aligned to the onset of an adequate driving stimulus (Mitzdorf and Singer, 1979; Mitzdorf, 1985; Maier et al., 2011).

The CSD was calculated by downsampling the preprocessed LF data in time to 250 Hz (as implemented in DSP.jl), and then averaging the data aligned to the onset of the visual stimulus. Where available, data from the full-field, black-white stimulus was used (i.e., data were aligned to luminance transitions), otherwise data were aligned to the onset of a drifting grating (trials with electrical shock were omitted). Aligned, averaged data were then smoothed in space with a Gaussian ($\sigma = 200\ \mu m$), and the second spatial derivative was calculated by second-order, central finite differencing to yield the CSD.

For each recording site, the "CSD image" (a depth x time matrix of CSD values) for a time window spanning 0.1 seconds before to 0.156 seconds after stimulus onset was processed with a watershed image segmentation algorithm (following Klein et al. (2021)) to identify candidate cortical compartments. The vertical extent and onset set latency for each candidate compartment was calculated, and candidates with extents less than 200 μm were removed. The vertical overlap between remaining candidates was then calculated, and for any pair of candidates with an overlap greater than 33% the candidate with the later onset latency was removed. Finally, overlap and unassigned space (in the vertical dimension) between adjacent candidate compartments was removed by placing the boundary between the compartments at the vertical location within the overlapping or unassigned region where the absolute value of the CSD was at a minimum. Figure 2 illustrates the steps in this image processing pipeline.

The resulting compartment boundaries were then manually checked to ensure that 1) they correctly identified the short latency, prominent sink corresponding to layer 4C, and 2) qualitatively captured the large scale structure of the CSD. Only very minor, manual adjustments of the compartment boundaries were required for some recording sites.

3.2.6.3 Calculation of frequency envelope

The time-varying frequency envelope (TVFE) was calculated based on an approach developed by Zerlaut et al. (2022). First, the preprocessed LF data from each channel was convolved with N complex Morlet wavelets with center frequencies spanning a specified range (e.g., 2 - 5 Hz for the delta band), resulting in N complex signals (per channel). The magnitude of each sample of each resulting signal was calculated, and the maximum at each time point (over the N signals) was taken to produce a single signal for each channel that represents the time course of the oscillatory amplitude for the given frequency band.

3.2.6.4 Classification of trials based on oscillatory activity

For all main analyses, each trial (defined by the onset of either a visual or electrical stimulus, depending on the particular analysis) was classified as starting during a period of low or high oscillatory activity by 1) calculating the time-averaged amplitude of the TVFE within putative layer 4C (i.e., the compartment corresponding to the initial, prominent sink in the CSD) during a pre-stimulus time epoch, and 2) performing a median split on the pre-trial time-averaged TVFE to identify each trial as below or above the median.

The pre-trial delta TVFE was calculated using 5 equally spaced frequencies between 2 and 5 Hz, while pre-trial alpha used 9 equally spaced frequencies between 8 and 12 Hz. For analyses of responses to visual stimuli, a pre-trial classification window between -1.05 and -0.05 seconds relative to stimulus onset was used; for analyses of responses to electrical stimuli, a pre-trial window between -0.55 and -0.05 seconds was used.

3.2.6.5 Quantifying V1 responses

3.2.6.5.1 Event related potentials

Event related potentials (ERPs) were calculated from the preprocessed LF data time-aligned to stimulus onset (either visual or electrical). For visual (electrical) stimulus analyses a time window of -0.05 to 1.0 (-0.05 to 0.5) seconds relative to stimulus onset was used. Each trial was smoothed with a 2d Gaussian (σ_{depth} = 100 μm, σ_{time} = 0.002 seconds) prior to averaging.

3.2.6.5.2 Gamma-band envelope The gamma-band TVFE (TVGE) was calculated from the preprocessed LF data following the approach used by Zerlaut et al. (2022) using 20 equally spaced frequency bands between 40 and 120 Hz and then smoothed in time with a Gaussian ($\sigma = 42\ ms$). Trials were defined exactly as for ERPs: for visual (electrical) stimulus analyses a time window of -0.05 to 1.0 (-0.05 to 0.5) seconds relative to stimulus onset was used. Each trial was smoothed with a 2d Gaussian ($\sigma_{depth} = 100\ \mu m$, $\sigma_{time} = 0.002$ seconds) prior to averaging.

3.2.6.5.3 Single and multi-unit spiking Following automatic curation (see Spike sorting above), each unit was assigned a depth based on the depth of the channel for which that unit's template (i.e., characteristic waveform) had the largest peak-to-trough amplitude. Units were then binned into 49 equally spaced bins between 0 and 3840 μm (80 μm bin size) and the peri-stimulus time histogram (PSTH, using 2 ms time bins) of all units within the same depth bin were averaged together, yielding a depth-by-time matrix for each trial, analogous to the ERP and TVGE data. For visual (electrical) stimulus analyses a time window of -0.05 to 1.0 (-0.02 to 0.08) seconds relative to stimulus onset was used, and trial was smoothed with a 2d Gaussian prior to averaging (for visual trials $\sigma_{depth} = 160\ \mu m$, and $\sigma_{time} = 0.008$ seconds; for electrical trials $\sigma_{depth} = 80\ \mu m$, and $\sigma_{time} = 0.002$ seconds).

3.2.6.5.4 Random sampling normalization Visualizations of average responses to visual and electrical stimuli include a z-scoring procedure, applied to each recording separately, in order to convey a sense of magnitude of the response relative to what would be expected from a "noise" distribution specific to that recording and stimulus. As visual stimuli appear "on top of" a uniform gray background (i.e., mean luminance), the corresponding noise (or null) distribution should reflect the mean and variance observed during that background. Thus, to generate a noise distribution for the visual data, "noise trials" of the same length as visual trials (i.e., 1.05 seconds) were randomly sampled from all time epochs during which no stimulus was present (i.e., the monitor was a uniform gray) such that no portion of any noise trial overlapped with a stimulus of any kind. All time points that satisfied these criteria were equally likely to be chosen, and sampling was performed with replacement such that a given time point could contribute to more than one noise trial (time was discretized at 30,000 Hz to match the time sampling of event information in the recorded

data). Noise trials were generated in sets whose size matched the number actual trials of the recording being considered. Each set was then averaged over trials and smoothed identically to the actual data. The noise trial selection and averaging procedure was repeated 240 times to create an estimate of the full noise distribution for a given recording (i.e., a time x channels x 240 array). The mean (μ_{noise}) and standard deviation (σ_{noise}), computed over iterations, was then taken, and the observed average data (time x channels) was Z-scored by the mean and standard deviation of the noise distribution:

$$\mathrm{Z}_{noise} = \frac{observed - \mu_{noise}}{\sigma_{noise}} \tag{18}$$

In this context, Z_{noise} essentially represents a signal-to-noise ratio in standard deviation units, such that a $|\mathrm{Z}_{noise}| >$ ~1.96 would correspond to a significant ($p < 0.05$) deviation from noise.

A analogous procedure was used to Z-score the shock-driven data, except the "background" epochs from which the noise distribution was estimated were restricted to time periods that contained a visual stimulus but no electric shocks. As shocks always occurred during visual stimulation, visually driven activity is the proper noise background from which the signal-to-noise of shock responses should be assessed.

3.2.6.6 Compartmental responses across recording sites In order to combine data across recording sites, the time course data from all depths within each identified compartment (putative layer or sub-layer) were averaged together within recording, producing one signal per compartment per recording (typically, this was in fact a difference signal: below median trials minus above median trials). Compartment signals for each response type (ERP, TVGE, PSTH) were then averaged across recordings (and recording sites). In all time course figures, error shading reflects one standard error of the mean across recordings.

3.2.6.7 Statistical approach As described above, each response measure that we used results in a depth-by-time matrix for each trial, such that trials can then be grouped according to the level of pre-trial oscillatory activity (or any other measure), and then the group mean and difference

between group means can be calculated. In this case, the natural statistical approach is to then test whether the magnitude of the observed difference between trial groups (i.e., above median trials v.s. below median trials) is greater than what one would expect from taking the difference between randomly grouped trials. This was performed via a permutation test where on each of 5,000 iterations trials were randomly assigned to "above" and "below" groups and the difference between groups was taken and stored. The mean (μ) and standard deviation (σ) of the permuted differences were then used to Z-score the observed difference:

$$\mathrm{Z} = \frac{observed - \mu_{permuted}}{\sigma_{permuted}} \tag{19}$$

to create a depth-by-time matrix that represents, at each depth-time location, the deviation of the observed difference, in units of standard deviations, from the expected difference under the null-hypothesis of random trial grouping. Thus, assuming that the permuted differences are normally distributed (which is guaranteed by the central limit theorem for a sufficiently large number of permutations), any Z-score magnitude (i.e., |Z|) greater than 1.96 corresponds to a p-value of 0.05 or less.

3.3 Results

In this work, we examine the way in which ongoing activity within primary visual cortex affects how it responds to thalamic input. While the LGN comprises the largest source of driving input to V1 (Peters et al., 1994; Latawiec et al., 2000; Garcia-Marin et al., 2019), activity within V1 is also shaped by many other factors, including feedback from subsequent visual areas (V2, V4, etc. Felleman and Van Essen (1991)), input from non-visual areas (cholinergic system, etc. Krueger and Disney (2019)), and internal dynamics within and between columns in V1 itself (Angelucci and Bressloff, 2006). Thus, the LGN and V1 can be thought of as distinct, though strongly coupled, dynamical systems with their own internal state; V1 responds strongly to, but does not simply follow the activity of, the LGN, and how exactly V1 responds to a given thalamic input depends on both the structure of that input and the state of the V1 circuit when the LGN input arrives. Our goal in this work is to explore how ongoing cortical oscillations modulate how the cortex

responds to thalamic input generated "naturally" by a visual stimulus, as well as thalamic input generated artificially by direct, electrical stimulation of thalamo-cortical cells and fibers. Because we are primarily interested in how ongoing V1 activity affects how it responds to geniculate input, reducing the variability of that input, which would otherwise comprise a substantial source of variability in the cortical response, serves to both better isolate the system of interest as well as increase our ability to detect any influence that may be present. Furthermore, studying the dynamics of the system under artificial, more isolated conditions can aid in the interpretation of observations made using visual stimuli in which the LGN input will also vary.

In order to address this question we made electrophysiological recordings from high-density, Neuropixel probes from opercular V1 of anesthetized rhesus macaques in response to either visual stimulation (4 Hz drifting sinewave gratings, see Visual stimulus display), or electrical stimulation of the LGN and/or optic radiations above the LGN (see Electrical stimulus delivery). In both conditions, V1 is responding to input from the LGN, but the degree of temporal and spatial structure is assumed to be quite different as electrical stimulation leads to a much more synchronous input with little spatial structure among the activated neurons.

Our data set consisted of 22 recordings from 8 distinct recording sites from 2 animals (see Electrophysiological data, where each recording consisted of a series of trials (80 to 200 trial of 2 - 8 seconds duration depending on recording) during which a drifting sinewave grating (4 Hz, near preferred spatial frequency and orientation) appeared on the screen, followed by a "blank" period where the screen was a uniform gray (2 - 8 seconds). On approximately half of trials one or more electrical shocks were delivered to the LGN (see Electrical stimulus delivery). Thus, for the purposes of analysis, trials can be aligned either to the onset of the visual stimulus (excluding trials on which an electrical shock occurred) or to the onset of a shock.

3.3.1 Shock artifact suppression and identification of cortical compartments

In order to analyze V1 responses of any kind, the electrical shock artifact needed to be suppressed so as not to contaminate the local field potentials (LFPs) or spike sorting results. Suppression of electrical shock artifacts was performed using a matching pursuit (Mallat and Zhang, 1993)

approach in which time epochs of the continuous voltage data that contained a shock were iteratively decomposed using a set of basis functions tailored to the shape of the artifact, with each iteration subtracting the projection of the data onto the best matching basis function. Thus, over iterations the contribution of the shock is reduced until a stopping criterion indicating success is met (see Shock artifact suppression). Figure 3.1 illustrates the result of running this process on both the local-field (LF) and action-potential (AP) band data for an example recording: Figure 3.1A & C show the raw data with the prominent shock artifact for 64 channels spanning the full length of the probe (for visual clarity, not all channels are shown), along with the recorded output of the stimulator (black dashed line) for comparison. Two of the reference channels can be seen among the group, as these do not display the shock artifact. Figure 3.1B & D show the data after the artifact suppression procedure has been run. Note that the square wave structure of the AP (and to a lesser extent LF) signals indicates that the amplifier was saturated for ~2 ms, preventing shock-induced spikes from being recorded within that window. However, the vast majority of thalamo-recipient cells in V1 show shock response latencies longer than 2 ms (Briggs and Usrey, 2007), so few spikes should be "lost" due to amplifier saturation or artifact suppression.

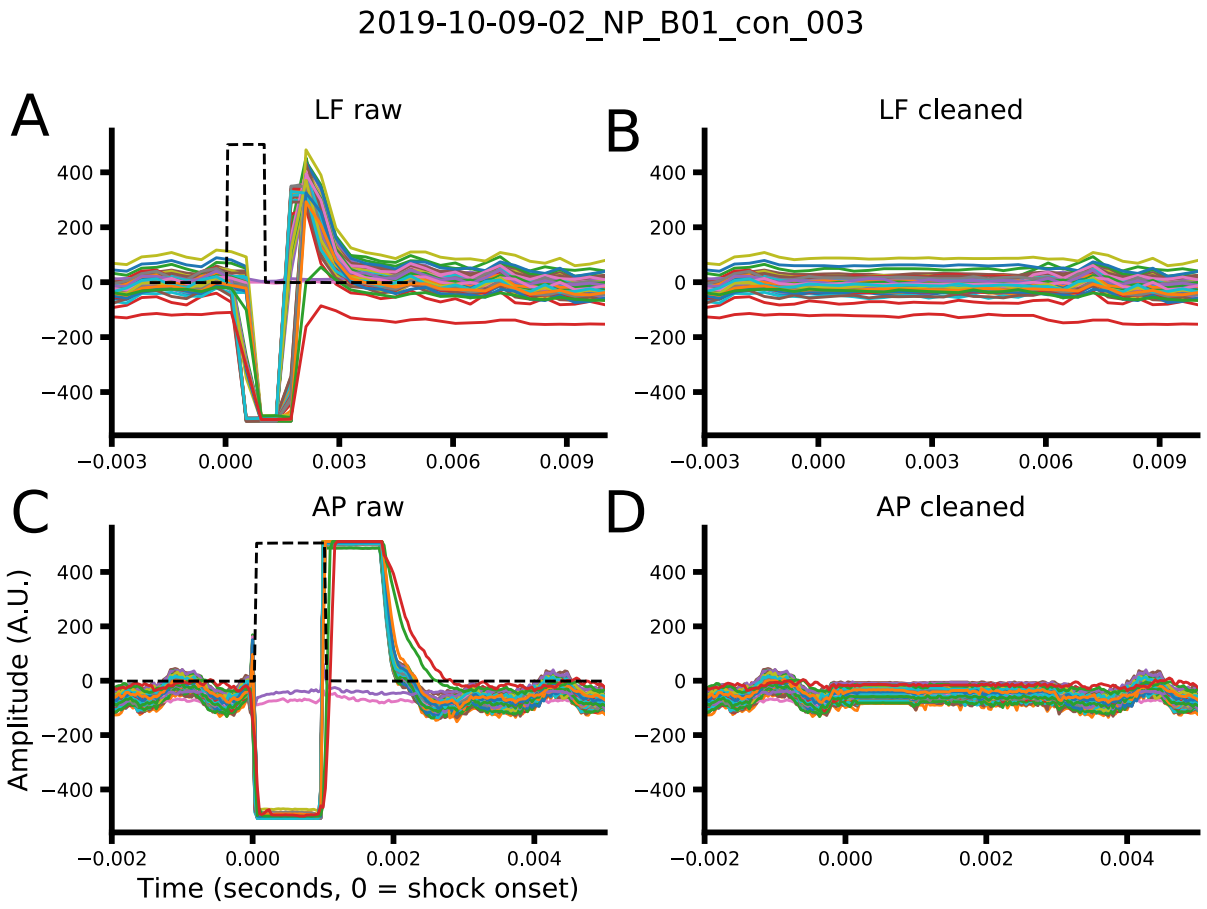


Figure 3.1: Results of artifact suppression for an example recording for a subset (64, every 6th along the full length of the probe) of channels (384 total). **A** & **C**: Raw voltage traces for a short time epoch surrounding an example shock for LF (**A**) and AP (**C**) data. Black dashed lines show the simultaneously recorded, unfiltered output of the stimulator for reference. **B** & **D**: Cleaned voltage traces following artifact suppression for both data types.

The second major analysis step that needed to be performed in order to interpret the results of the main analyses reported here was to identify the depth of the probe (and thus the depth of each electrode contact) relative to the cortical layers. This can be done using current source density (CSD) analysis which estimates the net density of positive current flow into (sink) or out of (source) a given point in space (within the purview of the probe) from preprocessed LFP data (see Identification of cortical compartments). Thus, locations with a large amount of synchronously depolarizing neural membrane will appear as sinks in the CSD. Within primary, sensory cortices, such as V1, the thalamic input layers (in particular layer 4C) can be readily identified from the prominent, short latency sinks that develop there following a strong stimulus. For each of the 8

recording sites in our data set, we calculated the CSD response to either luminance transitions from an alternating black-white full-field "flash" stimulus or the onset of a drifting grating. The resulting CSD "images" (depth by time) were then processed using a simple image segmentation algorithm (see Identification of cortical compartments) to identify layer 4C as well as other cortical compartments that are likely to share some correspondence with the other cortical layers (2/3, 4AB, 5, and 6). However, without anatomical confirmation we cannot verify the relationship between CSD defined compartments and the anatomically defined layers (layer 4C being the only exception due to its well established, unique, prominent contribution to the CSD (Mitzdorf and Singer, 1979; Mitzdorf, 1985). Figure 3.2 illustrates the processing pipeline that calculates the CSD and identifies the cortical compartments. Figure 3.2A shows the average ERP image, aligned to full-field luminance transitions, for an example recording site. The dashed white box indicates the time epoch that the segmentation algorithm operated upon (-0.1 - 0.156 seconds relative to luminance transitions / stimulus onset). The CSD is then calculated from the smoothed second spatial derivative of the ERP (Figure 3.2B), and the absolute value is taken (|CSD|, Figure 3.2C). A watershed image segmentation algorithm is then run on the |CSD|, resulting in a number of candidate compartments (Figure 3.2D). Candidate compartments are then evaluated based on their temporal onset and spatial extent, and only the shortest latency components that span at least 200 μm in depth are kept. The boundaries of the resulting candidates are then adjusted so that they tile the depth dimension (Figure 3.2E, see Identification of cortical compartments). Figure 3.2F shows the final compartment boundaries and putative layer correspondence drawn over the CSD (laminar labels represent the "best guess" given the general structure of the data, and the locations of layer 4C, the pial surface, and the gray matter - white matter boundary).

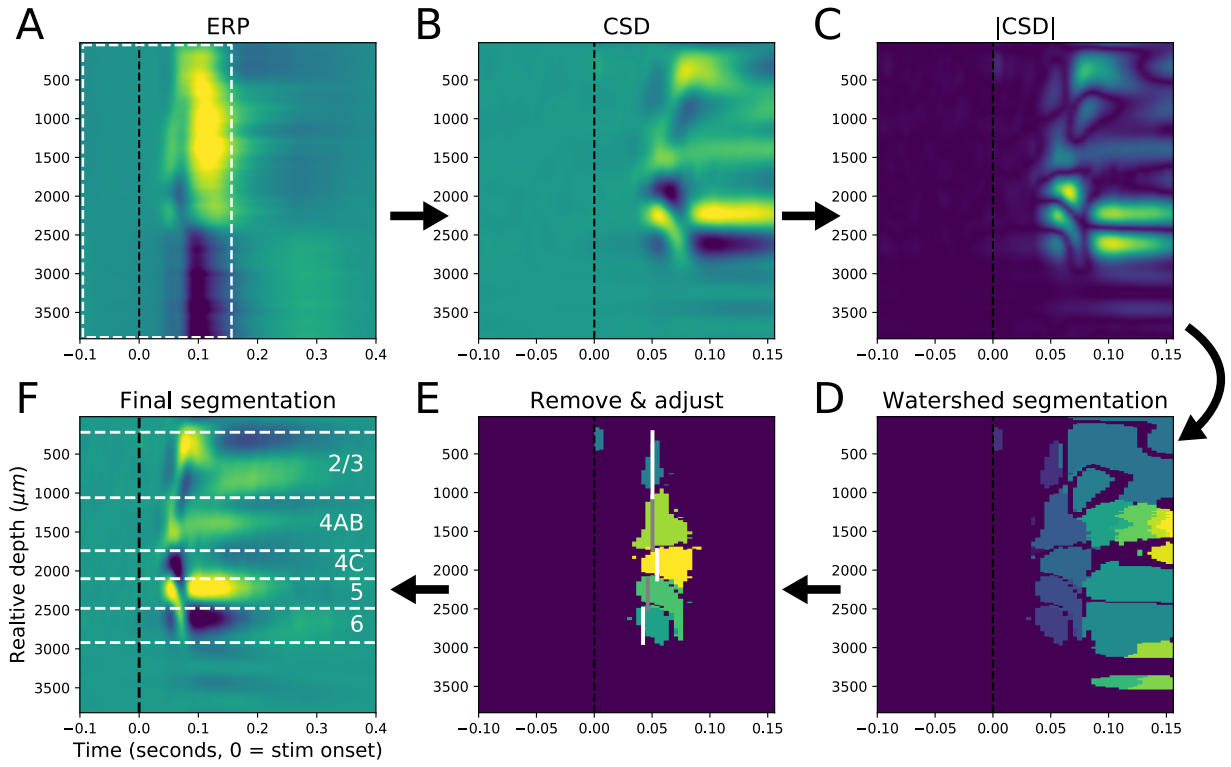


Figure 3.2: Illustration of cortical compartment identification pipeline. **A**: ERP image (color indicates response amplitude as a function of depth (y-axis) and time (x-axis)) aligned to full-field, black-white transitions for an example recording site. Dashed, black line indicates the timing of the luminance transition; dashed, white box indicates the time epoch used for compartment identification. **B**: CSD calculated from the 2nd spatial derivative of the ERP in **A**. **C**: Absolute value of the CSD. **D**: Results of watershed segmentation, note that this results in an over-segmentation due to temporal dynamics and noise. **E**: Results of the "blob" removal and boundary adjustment procedure (see Identification of cortical compartments). **F**: Final segmentation result with layer boundaries (white, dashed lines) drawn over the CSD image. Note that compartment labels represent a "best guess" given the general structure of the data and the locations of layer 4C, the pial surface, and the grey matter - white matter boundary.

3.3.2 Structure of V1 responses across the full depth of cortex

Once the electrical shock artifact was suppressed, and cortical compartments were identified, the LF data were preprocessed and the AP data was spike sorted (see LFP preprocessing). From the LF data we computed two types of V1 responses: the event related potential (ERP) response, which reflects broad band cortical activity that is phase-locked to stimulus onset, and the gamma-band, time-varying frequency envelope (TVGE), which reflects the time-varying amplitude of the high-frequency components (40 - 120 Hz) of the LFP and has been shown to be highly correlated with

fluctuations in the membrane potential of nearby cells (Zerlaut et al., 2022). Also, unlike ERP data, TVGE signals with opposite phases will not cancel out during averaging, thus average TVGE reflects the gamma-band activity of sources that respond to the stimulus regardless of phase. From the spike sorted AP data we computed depth-resolved peri-stimulus time histograms (PSTHs) by averaging together the single trial spiking activity of units (including both single and multi units) located at nearby depths (see Single and multi-unit spiking). Importantly, all three of these response types result in data that are organized by depth, time, and trials (see Quantifying V1 responses) so that the trial averaged time x depth "images" can be readily compared. Note however that the depth resolution of the PSTH data are lower than the LF derived response types (80 μm for PSTH data vs. 20 μm for ERP and TVGE data).

Figure 3.3 shows data from each of the three response types for one recording from an example recording site (Experiment 2019-10-09, site B01). The left column (A, C, E) shows average data aligned to the onset of visual stimuli for trials without any electric shocks, and the right column shows the corresponding data aligned to shock onset (B, D, F). In all panels, color indicates the amplitude of the response Z-scored relative to a noise distribution obtained via a constrained random sampling data generation procedure (see Random sampling normalization). In this framework Z-scores represent the deviation of the observed data from what would be expected from a background noise distribution derived from the data itself. Values whose magnitude exceeds ~1.96 would be considered statistically significant at the $p < 0.05$ level. The white dashed lines in all figures indicate the compartment boundaries estimated by the procedure illustrated in Figure 3.2. The data displayed in Figure 3.3 came from the same recording site as the data shown in Figure 3.2, but from a separate recording using different stimuli (Figure 3.2 used full-field black-white luminance transitions, Figure 3.3 used a drifting grating with (right column) or without (left column) electric shocks). Thus, the general alignment between the estimated boundaries and the data in Figure 3.3 suggests that the boundary estimation procedure is capturing meaningful structure in the data. Specifically, the compartment that likely corresponds to layer 4C (at ~2,000 μm relative depth, see Figure 2F) shows the earliest response to the visual stimulus across all response types (Figure 3.3A, C, and E), and the strongest response in the spiking data and the strongest negative response in the ERP data (Figure 3.3A & E). Likewise, the putative 4C compartment shows the strongest and

fastest response to the shock for all three response types (Figure 3.3B, D, & F), and is the only compartment that shows a clear response to the shock in the spiking data (Figure 3.3F, beginning at ~5 ms post shock), consistent with receiving direct, synaptic input from shock-driven fibers.

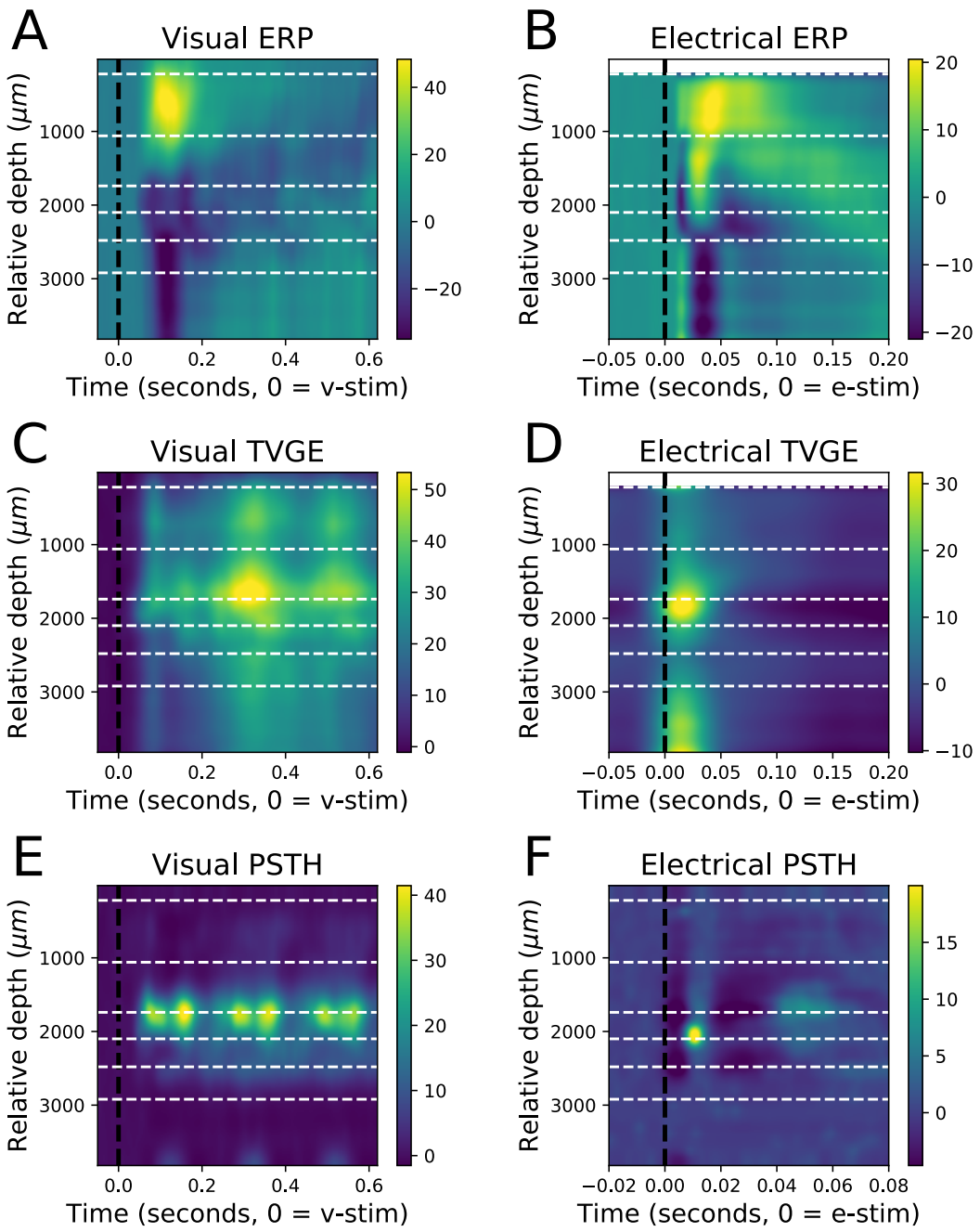


Figure 3.3: V1 responses to visual (left column) and electrical (right column) stimuli for an example recording site. **A & B**: average ERP images aligned to visual (**A**) or electrical (**B**) stimulus onset (at t=0). Color indicates ERP response amplitude z-scored relative to a null distribution obtained via a random sampling surrogate data generation procedure (see Random sampling normalization). **C & D**: average z-scored TVGE of the LFP response (see Gamma-band envelope). Same conventions as **A & B**. **E & F**: average z-scored spiking response. Same conventions as **A & B**. Black dashed lines indicate the timing of stimulus onset (t=0); white dashed lines indicate estimated laminar compartment boundaries (see Identification of cortical compartments). In panels B & D data from depths that are estimated to be near or above the pial surface are "blanked out" due to incomplete suppression of the shock artifact at the most superficial depths.

The general depth-time structure of the visual data is also worth noting. The ERP data shows a strong positive-over-negative pattern that is highly consistent across the data set (i.e., positive ERP components dominate the superficial compartments while negative components dominate in the middle and deep compartments, Figure 3.3A). The TVGE data are more ambiguous in their depth structure, but display a much stronger stimulus-locked periodicity than the ERP data do (the three initial cycles of the 4 Hz drifting grating are readily apparent Figure 3.3C). The spiking data (Figure 3.3E) are dominated by the middle layers, which show clear phase locking to the drifting grating, although at twice the stimulus frequency. This could result from recording from a mixed population of both ON-preferring and OFF-preferring cells in close proximity within the middle layers (and thus their responses get averaged together in our analysis). It could also result from a population of complex cells that lack a distinct phase preference (i.e., they respond to both white and black phases of the grating). However, simple cells (that have a distinct phase preference) are known to outnumber complex cells within layer 4C, thus the mixed population is more likely (Ringach et al., 2002). In this recording, the superficial compartments (above ~100 μm relative depth, likely corresponding to layer 2/3) show almost no response in the spiking data despite clear responses in the ERP and TVGE data. We suspect this may be due to anesthesia, which is known to affect the superficial and deep layers more than the middle layers (Sellers et al., 2015).

Figure 3.4 shows the data from another example recording site from a different subject (Experiment 2019-06-05, site C01, all other figure conventions follow those of Figure 3.3 exactly). While much of the data from this recording site show the same general features as those from Figure 3.3, several differences are worth noting as they illustrate the variability observed within the data set as a whole. First, the visual responses of this recording site show more activity within the superficial compartments than that seen in Figure 3.3, although the compartment that likely corresponds to layer 4C (~2,200-3,000 μm relative depth, Figure 3.4A, C, & E) still shows the shortest response latency and the strongest shock-response in the spiking data (Figure 3.4F). However, this recording showed much stronger spiking visual responses in the superficial (putative later 2/3, ~900 - 1700 μm) and deeper (putative layer 5, ~3000 - 3500 μm). Additionally, while the visual ERP data show the typical positive-over-negative structure, the shock ERP data show the reverse (Figure 3.4B) unlike that seen in Figure 3.3B. While the data do not indicate a definitive cause for the

sign reversal (it is not due to the z-scoring procedure or other analysis choices), it may be due to the waveform of the electrical shock that was used. The recording shown in Figure 3.4B, D, and F used a monophasic, negative shock (0.3 ms duration) while that shown in Figure 3.3 used a bi-phasic shock (positive leading, 0.5 ms total duration). Finally, the visual TVGE data shown in Figure 3.4C shows a strong but delayed (~0.2 seconds) response spanning the full depth of cortex with little stimulus driven periodicity, while the electrical TVGE data (Figure 3.4D) appears to be contaminated by artifact from the shock, although the shock ERP data appears clean.

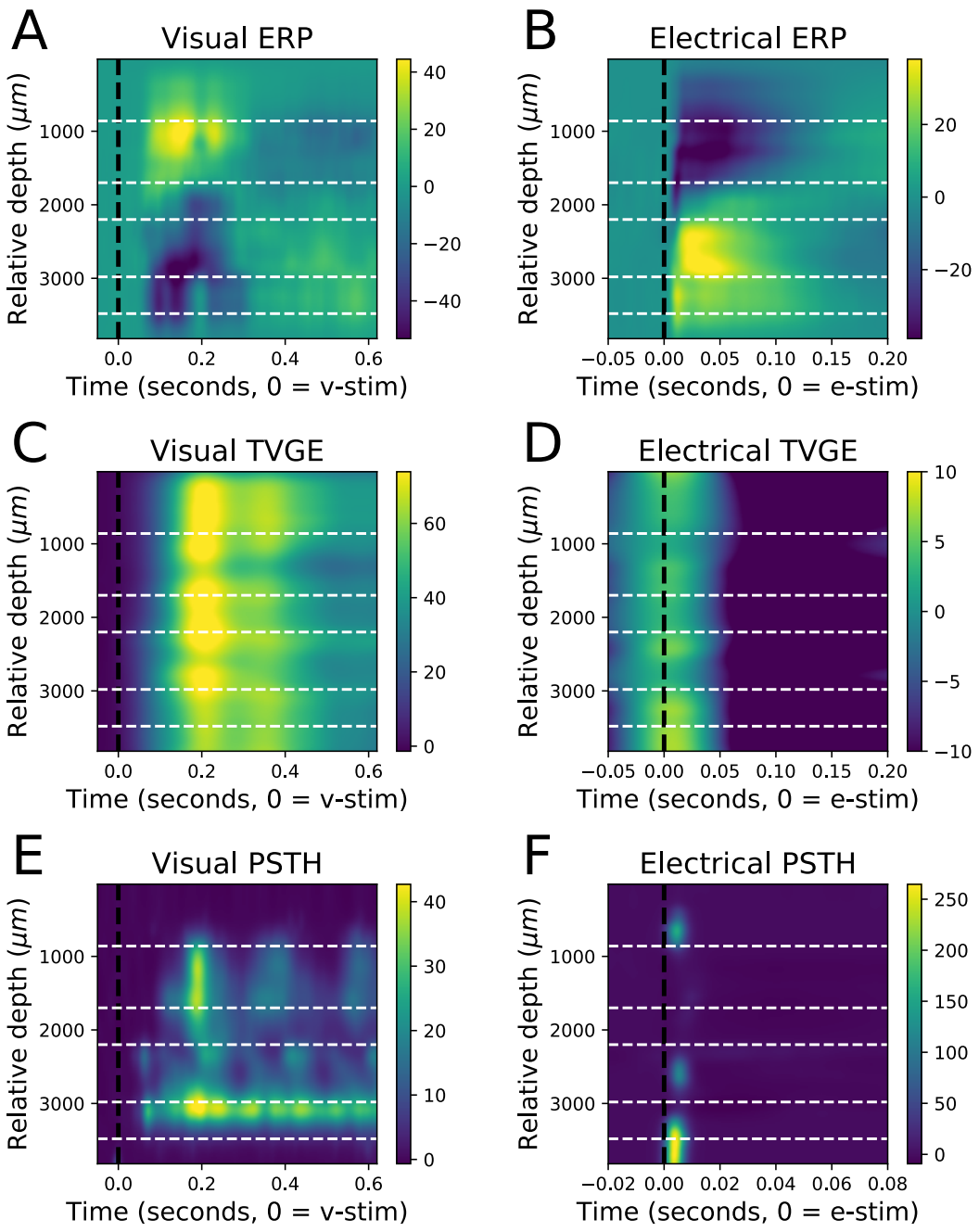


Figure 3.4: V1 responses to visual (left column) and electrical (right column) stimuli for another example recording site from a different subject. All conventions exactly follow those used in Figure 3.3

3.3.3 Pre-stimulus delta amplitude modules V1 responses: example recordings

In order to investigate how ongoing activity within V1 affects how V1 responds to thalamic input, we classified each trial according to the amplitude of ongoing delta oscillations (2 - 5 Hz) at the moment of stimulus onset (see Classification of trials based on oscillatory activity). We reasoned that if the cortex was in a period of strong delta oscillations when the stimulus driven thalamic input arrived, that input would be relatively less effective at driving cortical responses compared to input that arrived during periods of weaker delta (Zerlaut et al., 2022). To achieve this, we calculated the average amplitude of delta oscillations during a 1 second long epoch prior to stimulus onset (0.5 seconds for electrical stimuli) within the laminar compartment identified as likely corresponding to layer 4C (see Identification of cortical compartments). Each trial was then classified as being either above or below the median pre-trial delta amplitude, and "above trials" and "below trials" were averaged separately and their difference calculated. Figure 3.5 shows the distribution of delta oscillation amplitude during the pre-trial epoch used for classification for an example recording (Experiment 2019-10-09, site B01). The distribution of pre-trial delta amplitudes for below (above) median trials is shown in blue (orange), and the delta distribution from all length-matched epochs across the full recording is shown in black. In this example recording, the full delta distribution is clearly unimodal, suggesting that the broader network dynamics are continuous within the delta range, whereas a more multi-modal distribution would suggest more discrete or "state-like" dynamics. Additionally, the distribution of pre-trial delta amplitude appears to match the full distribution fairly well, although for visual trials the distribution is flatter and exhibits heavier tails (Figure 3.5A), indicating that the pre-visual-trial epochs have a higher variability in delta amplitude than the recording as a whole, though not hugely so.

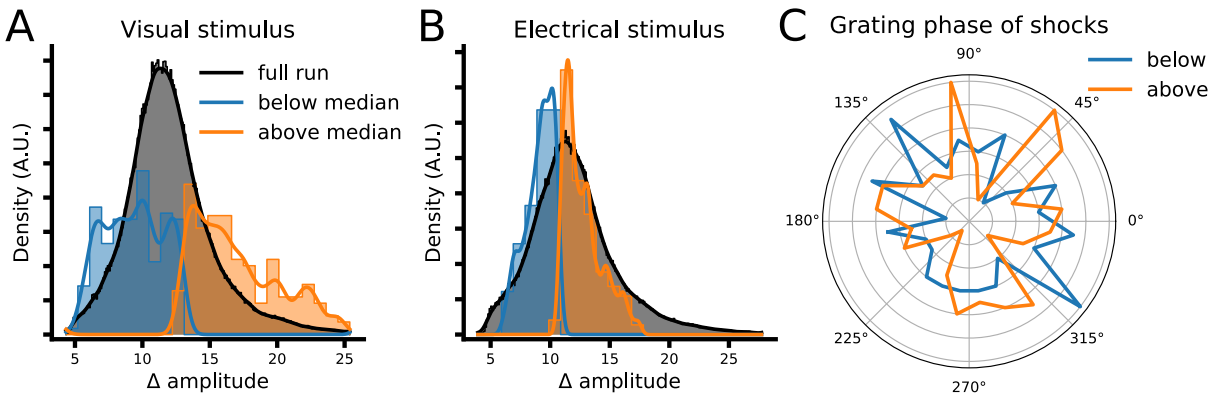


Figure 3.5: Distributions of average delta oscillation amplitude and drifting grating phase at which electrical shocks occurred for an example recording (Experiment 2019-10-09, site B01). **A**: Distribution of delta amplitude for 1 second long below median (blue) and above median (orange) pre-trial epochs for visual trials. The distribution of average delta amplitude for all 1 second epochs during the entire recording is shown in black. **B**: same as **A** but for electrical stimuli and utilizing 0.5 second epochs (see main text). **C**: distribution of drifting grating (i.e., visual stimulus) phase at which electrical shocks occurred for below median and above median trials for the same example recording as **A** and **B**. The radial dimension indicates probability density in arbitrary units.

Figure 3.6 shows the ERP response for the same recording site during below (Figure 3.6A & D) and above (Figure 3.6B & E) median trials following visual (top row) and electrical (bottom row) stimuli (conventions follow those from Figure 3.3 and fig. 3.4). Figure 3.6C and F show the difference between below and above trials (color) as well as statistical contours (where the gray-to-black gradient indicates large-to-small p-values) resulting from a permutation test that estimates the degree of difference that would be expected from random assignment of trials into above and below conditions (the lightest gray corresponds to $0.05 \geq p > 0.01$, while black corresponds to $p < 0.001$, see Statistical approach). While the statistical contours are useful insofar as they indicate deviations from what is essentially a noise model (i.e., random assignment to above and below median trial groups), they should be interpreted cautiously, as large numbers of trials will result in even very small differences being found to be statistically significant. In general, for this recording the largest difference in the visual ERP response was seen in superficial compartments at long latencies (0.2+ seconds post stimulus onset) where below median trials show a larger negative response following the strong initial positivity (which did not prove to be consistently difference between above and below trials). While these differences were statistically robust (as indicated by the dark contours), the long latencies make interpretation somewhat complicated, as the bulk

of the visual ERP onset response is over by ~0.2 seconds (Figure 3.6A & B, Figure 3.3A) and the differences begin only after ~0.2 seconds (Figure 3.6C). However, given that the stimulus was a 4 Hz drifting grating it appears that the level of pre-trial delta is modulating the response to later grating cycles, despite the absence of any detectable response to later stimulus cycle in the all-trial averaged ERP (Figure 3.3A). In this recording it's clear that stimulus onsets that occur during periods of relatively low delta amplitude result in stronger (more negative) responses to later stimulus cycles in the most superficial compartment.

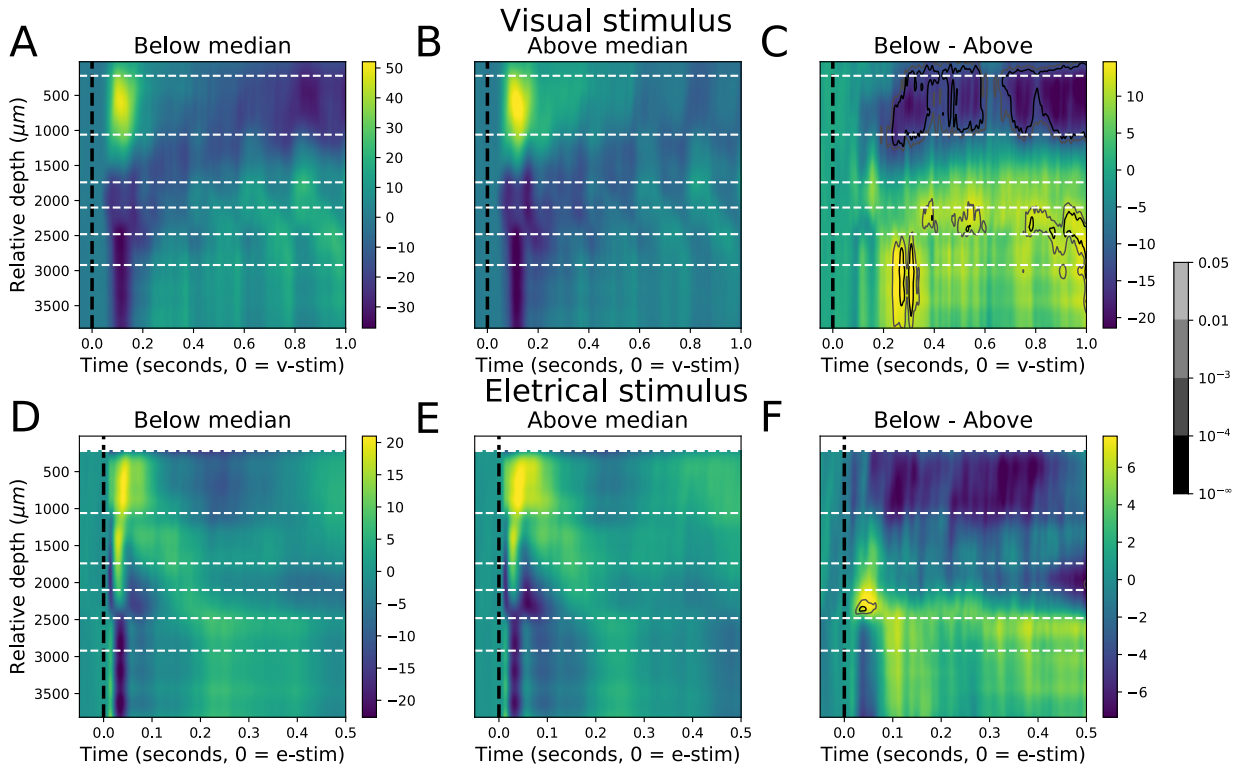


Figure 3.6: The modulation of V1 ERP responses to visual and electrical stimuli by ongoing delta amplitude for an example recording. **A & B**: Average ERPs during trials classified as below (**A**) and above (**B**) the median of pre-trial delta amplitude. Color indicates the z-score relative to a noise distribution as in Figure 3.3 (see Random sampling normalization). White dashed line indicate estimated compartment boundaries. **C**: Difference image between below and above median trials (below - above). As in **A** and **B**, color indicates z-score difference. Grayscale contours reflect statistical significance of the observed difference relative to a permuted, null distribution of differences that estimates the expected difference if trial are random assigned to conditions (see Statistical approach).

The difference between above and below trials for electrical stimuli are subtler, but a short latency difference is clear within the putative layer 5 compartment (i.e the compartment immediately

below the presumptive layer 4C compartment), with a stronger negative response on above median trials (Figure 3.6F). While the timing of this difference coincides quite well with the shock-induced response (within ~0.1 seconds of shock onset), it's not clear why only putative layer 5, being in general two or more synapses away from any direct, shock-driven inputs (Douglas et al., 1989; Douglas and Martin, 1991; Bastos et al., 2012), would exhibit a robust difference.

Figures 3.7 and 3.8 show the results of the same analysis applied to the TVGE (Figure 3.7) and spiking (Figure 3.8) data. In general, the visual TVGE and spiking data show the same pattern as the ERP data, with below median trials showing stronger responses for later cycles of the drifting grating. This is especially clear in the superficial compartment of the TVGE data (Figure 3.7A-C). Interestingly, the spiking data shows more robust differences in the non-4C middle and upper compartments (Figure 3.8C), deeper than the effects seen in the TVGE and ERP data. However, interpretations of these data with respect to precise compartment boundary locations needs to be done carefully, as the relationship between compartments defined from the CSD and anatomical layers is uncertain, and imprecise. While portions of the shock response do show significant differences in the TVGE and spiking data (Figure 3.7 and fig. 3.8 D-F), the differences are long latency compared to the shock response and, like the shock, are non-periodic and thus do not on their own suggest the presence of a practically meaningful effect.

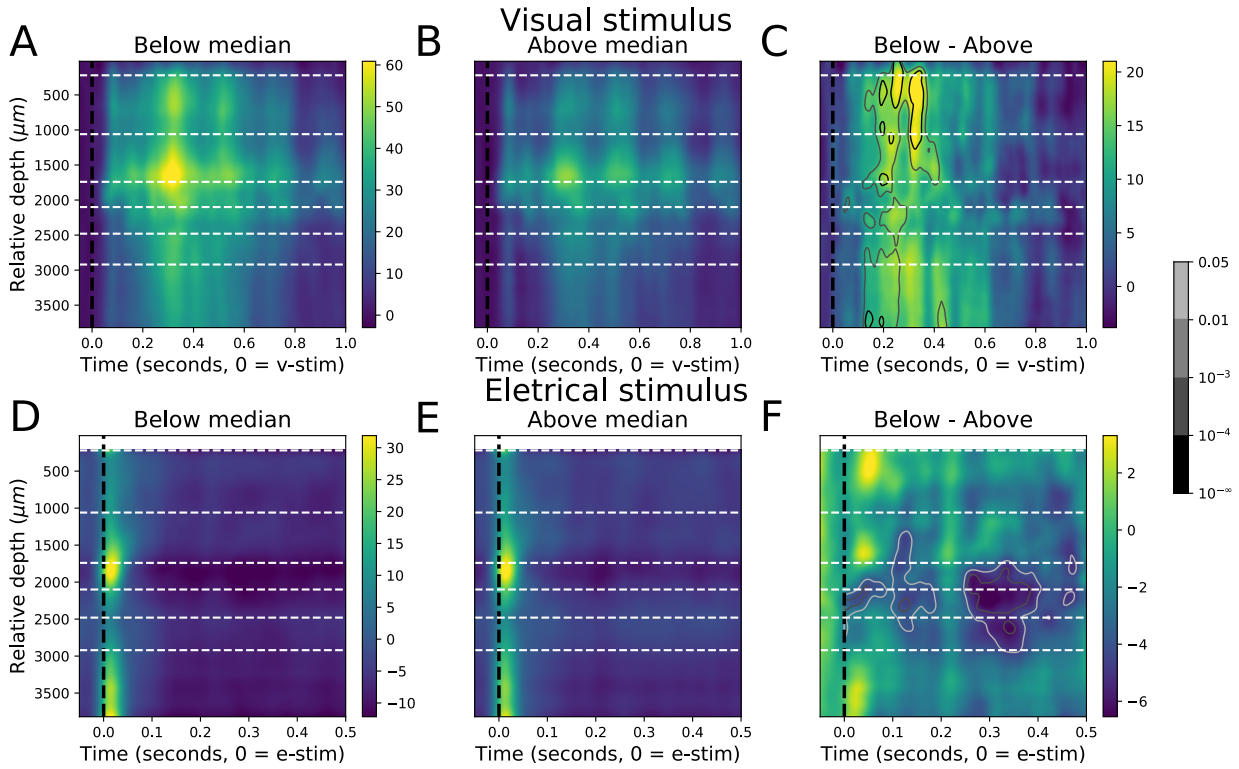


Figure 3.7: The modulation of V1 TVGE responses to visual and electrical stimuli by ongoing delta amplitude for an example recording. All conventions exactly follow those used in Figure 3.6.

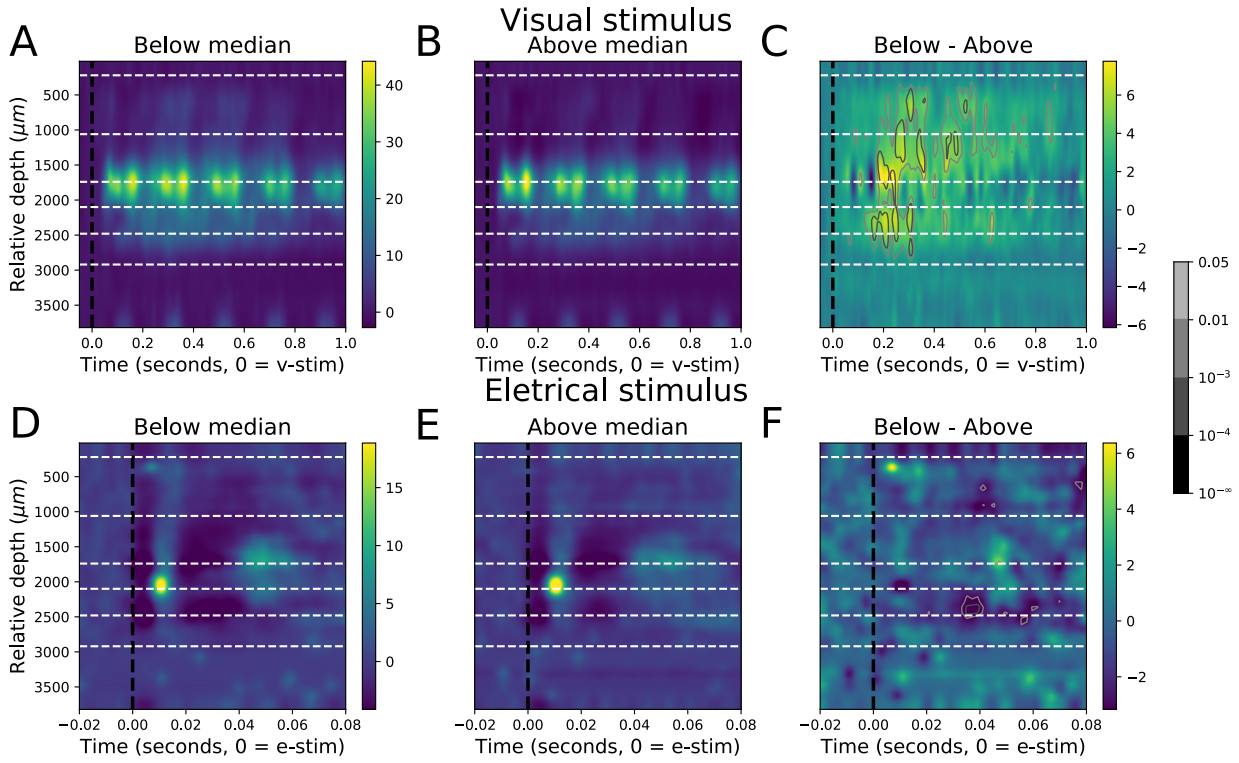


Figure 3.8: The modulation of V1 spiking responses to visual and electrical stimuli by ongoing delta amplitude for an example recording. All conventions exactly follow those used in Figure 3.6 and 3.7

3.3.4 Pre-stimulus delta amplitude modulates V1 responses: population analysis

In order to combine data across recordings and recording sites, for each recording we averaged together all difference signals (below median trials minus above median trials) whose depth corresponded to the same laminar compartment as defined by the CSD segmentation (see Identification of cortical compartments), and then calculated the mean (and standard error) across recordings, within compartment (i.e., each within-compartment averaged signal from each recording that corresponded to putative layer 4C was then averaged across recordings). The resulting signals, one per compartment, are shown in Figure 3.9 for the ERP data, where the left column (Figure 3.9A & C) shows data aligned to visual stimulus onset, and the right column (Figure 3.9B & D) shows data aligned to electrical stimulus onset. The bottom row of subpanels (C & D) show the same data as the top row, but with the time axis magnified around stimulus onset to aid visualization of the shortest latency components, which are especially prominent in the shock-aligned data (Figure

3.9B & D). In this and following figures the compartments are labeled with letters according to depth, with compartment A (C-A) being the most superficial and compartment E (C-E) being the deepest, in an attempt to avoid confusion with the numbering scheme used to identify anatomical layers (i.e., 1-6) and to reinforce the fact that the correspondence between compartments and layers is only approximate, and the uncertainty in the agreement between compartment and laminar boundaries is not uniform across depth (i.e., the C-C - C-D compartment boundary very likely corresponds to the anatomical layer 4C - 5 boundary, while the relationship between the C-A - C-B boundary and the layer 2/3 - 4AB boundary has much higher uncertainty, see Identification of cortical compartments and Caveats).

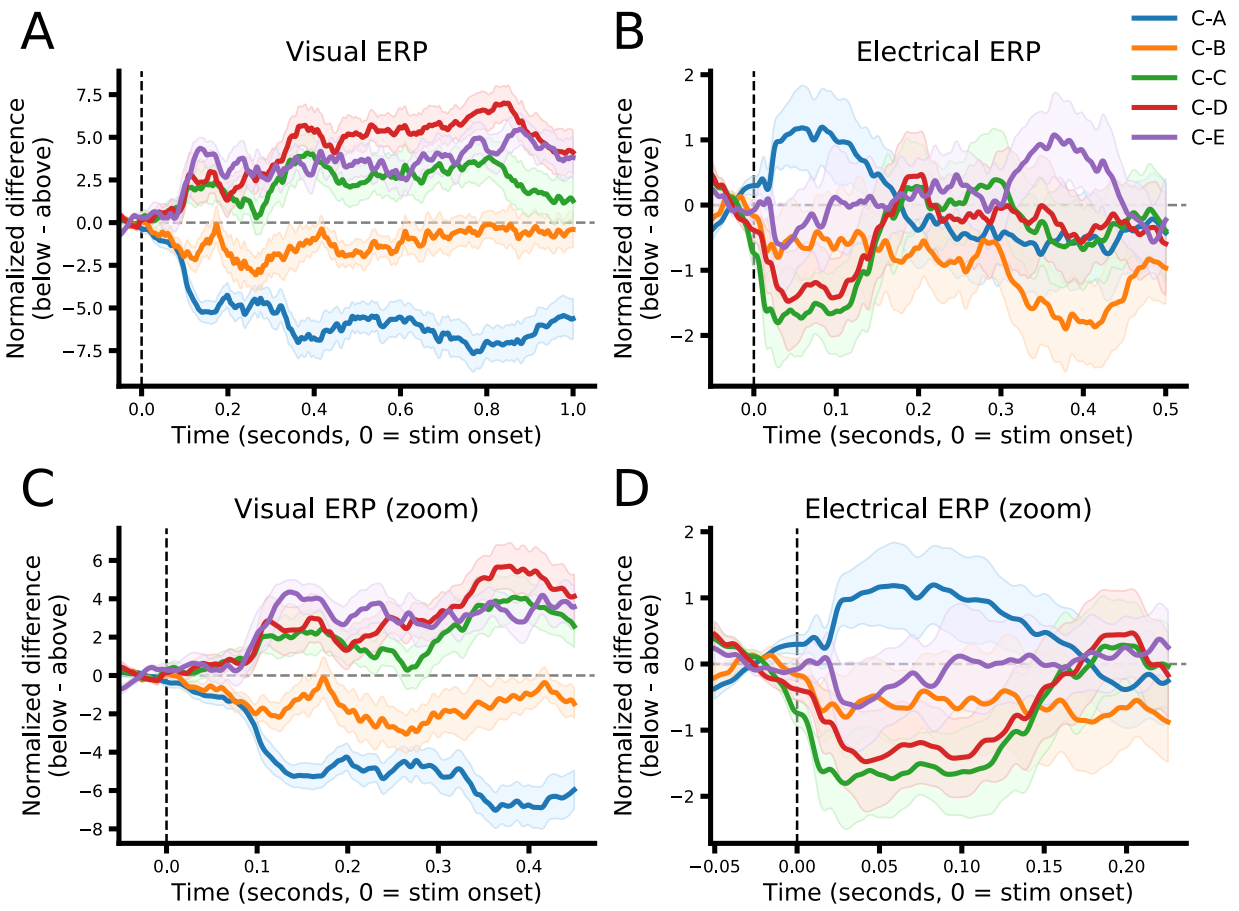


Figure 3.9: The modulation of V1 ERP responses to visual and electrical stimuli by ongoing delta amplitude, averaged across all recordings, for five depth compartments. **A**: Visually aligned ERP differences (below median trials minus above median trials) averaged within compartment for each recording, and then averaged across recordings (n=22). Signals from each recording are z-scored relative to a noise distribution prior to averaging, thus the y-axis reflects deviation from noise in standard deviation units (see Random sampling normalization). Color indicates putative compartment identity, and shading indicates ± 1 SEM. **B**: same as **A** but for shock-aligned data. **C & D** show the same data as **A** and **B** but with the time axis magnified so that the shortest latency components are more visible.

Interestingly, the size and sign of the difference between below and above trials in the visual data quite clearly sorts the compartments by depth, especially within the first 0.25 seconds following stimulus onset (Figure 3.9C). The most shallow compartment (C-A) shows the largest negative effect (i.e above > below), with the neighboring compartment (C-B) showing a similar but smaller negative effect. The deeper compartments (C-C though C-E) all show positive effects (i.e., below > above) with C-C showing a weaker effect than the two deepest compartments (C-D and C-E).

The pattern within the electrical data is rather different, although they also appear structured with respect to depth (Figure 3.9B & D). The electrical ERP data show the largest effect in the middle compartments, with C-C and C-D showing the largest negative effect, followed by C-B. The deepest compartment, C-E, appears minimally affected by pre-trial delta amplitude, while the most shallow, C-A, shows a slight positive effect (i.e., below > above). However, the variability across recordings is substantial, as seen in the vertical spread of the standard error shading, thus this pattern should be considered at most a potentially interesting trend that needs further investigation (see Responses to electrical stimuli). However, the timing of the effect onset (beginning within ~10 ms post shock) is consistent with what would be expected from the known thalamo-cortical anatomy if in fact different amplitudes of pre-trial delta lead to different amplitude V1 ERP responses. Furthermore, the prominence of the effect within the putative thalamo-recipient compartment C, where the shock response is expected to be strongest and thus the least susceptible to floor effects, is also consistent with this idea, though indirectly so (see Caveats).

Figure 3.10 shows the results of the same analysis as Figure 3.9 applied to the TVGE data. Across all compartments, the visual TVGE data show a stronger response on below median than on above median trials, with little difference seen between compartments. There is perhaps a slight trend towards a larger effect in compartment E (Figure 3.10C), but certainly nothing conclusive. The electrical TVGE data show a small, quite variable effect of pre-trial delta amplitude (below > above), with again no discernible difference between compartments (Figure 3.10B & D).

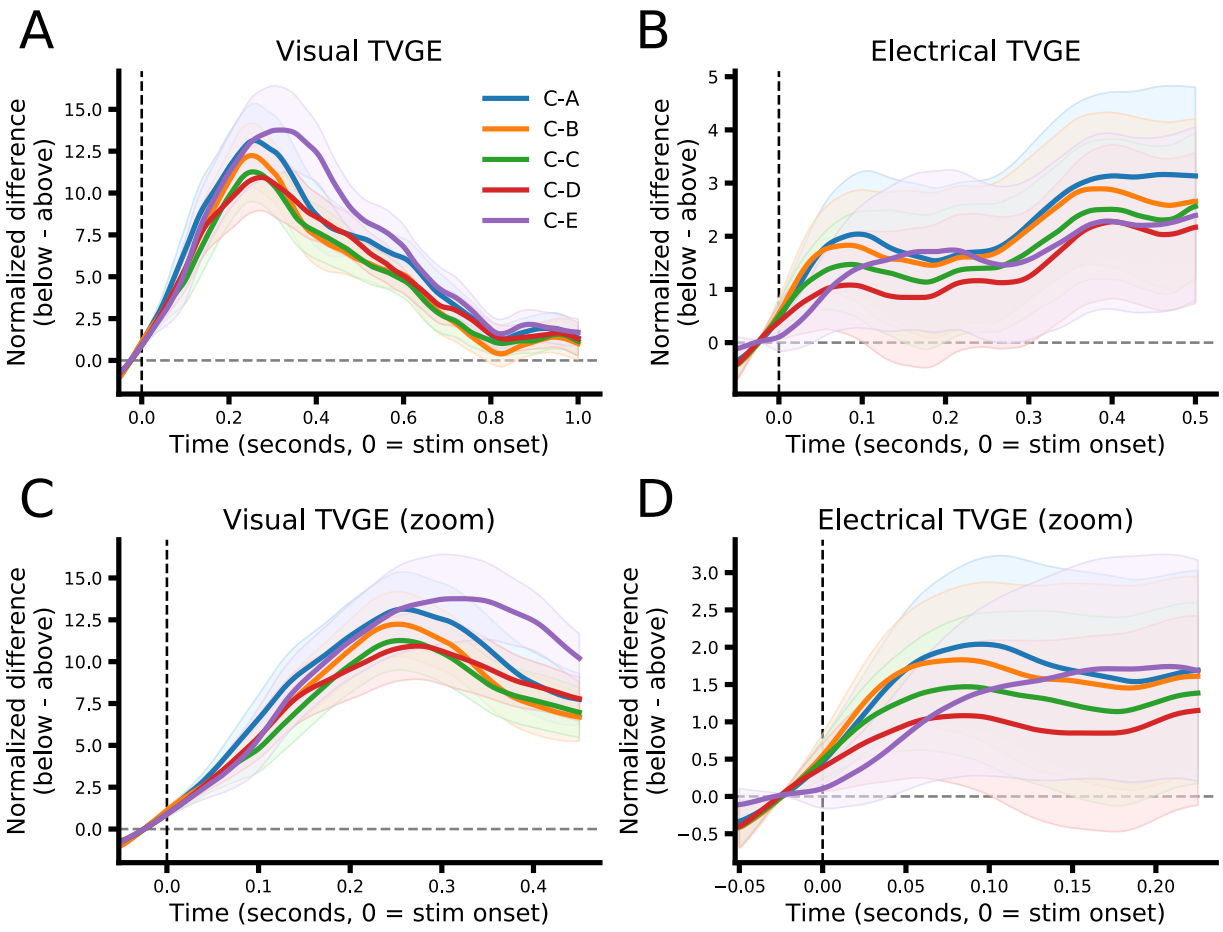


Figure 3.10: The modulation of V1 TVGE responses to visual and electrical stimuli by ongoing delta amplitude, averaged across all recordings, for five depth compartments. All conventions exactly follow those used in Figure 3.9.

Figure 3.11 shows the analogous results for the spiking data. As seen in the TVGE data, all compartments responded more strongly on below than on above median trial. The correlation between the TVGE and spiking results is quite striking (compare Figure 3.10A and Figure 3.11A), with a relatively fast onset (~0.2 seconds) and the slower offset (at ~0.8 seconds) of the delta effect can be clearly seen in both response types. While unsurprising, this similarity provides further support of the idea that high gamma-band activity is a good proxy for local spiking activity.

In the visual spiking data, compartment B showed the largest effect (below > above), with the other compartments showing a similar, but smaller, effect that was weakest in C-E (Figure 3.11A & C). This pattern is rather different from that seen in the ERP and TVGE data, with a clearly

larger effect in C-B than C-E (Figure 3.11C). While in general the spiking activity was less strong in C-B than C-C or C-D (Figure 3.3E), and occasionally weaker than C-A (Figure 3.4E), the spiking activity was consistently weakest in the deepest, C-E, compartment (Figures 3.3E and fig. 3.4E). This inconsistency between the spiking and LFP derived data may be due in part to the effect of anesthesia, which is known to affect some layers more than others (see Caveats, (Sellers et al., 2015)). While the electrical spiking data do show a small effect of pre-trial delta at very short latencies (within ~3-10 ms of shock onset) in the same direction as the visual data (i.e., below > above), the effect does not appear to be specific to any particular compartment, or have any apparent depth structure. In general, the spiking response to electrical shocks was quite weak in our data set, especially outside of the putative thalamo-recipient compartment (C-C, see Figures 3.3F, fig. 3.4F, fig. 3.8D-F). Thus, these data (Figure 3.11B & D) should not be taken as evidence against depth-related effects, only that we failed to detect any in the weak, shock-driven spiking responses that we recorded.

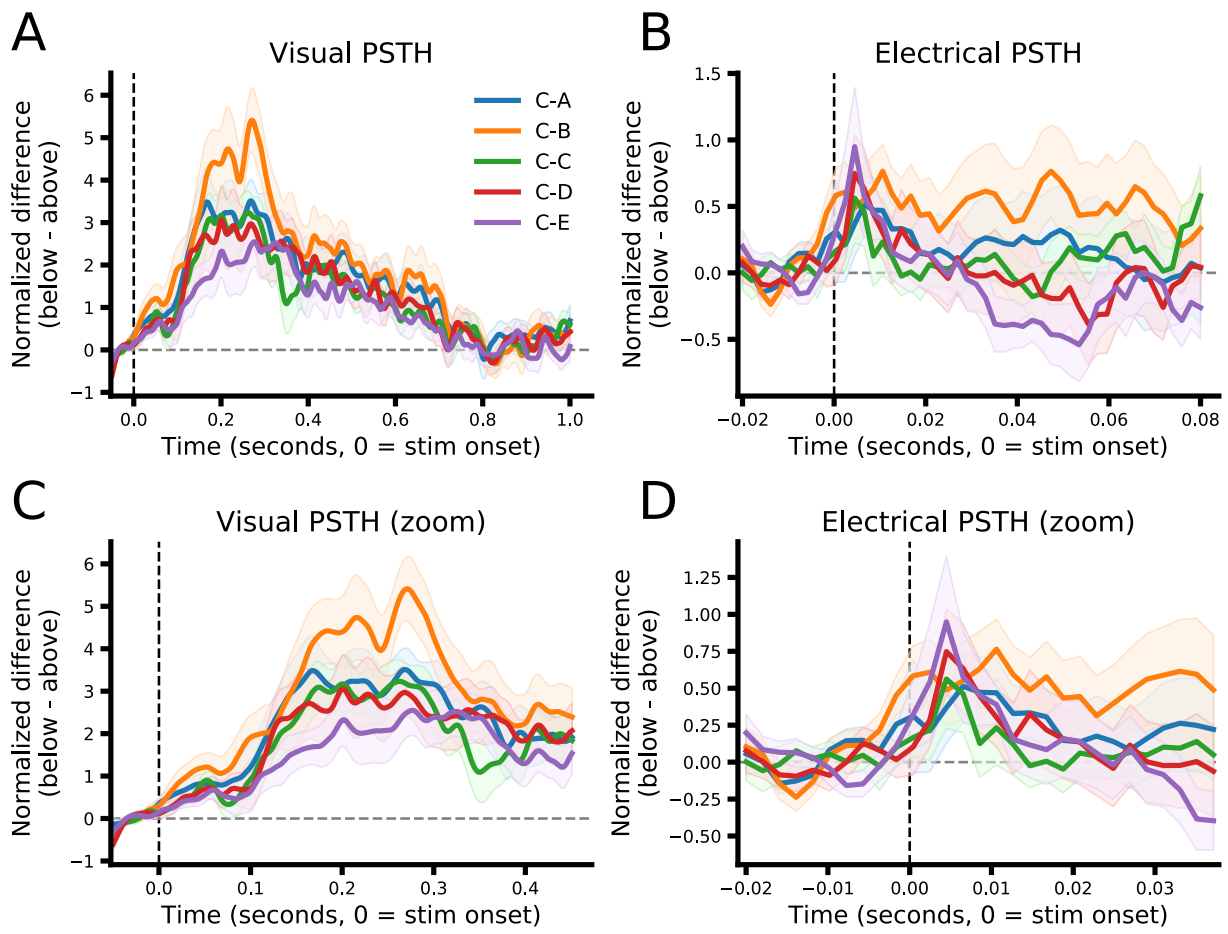


Figure 3.11: The modulation of V1 spiking responses to visual and electrical stimuli by ongoing delta amplitude, averaged across all recordings, for five depth compartments. All conventions exactly follow those used in Figure 3.9 and 3.10.

3.4 Discussion

3.4.1 General summary

In this work, we have investigated whether the amplitude of ongoing delta oscillations within V1 modulates how V1 responds to thalamic input. We utilized two forms of thalamic input, natural input driven by a visual stimulus (drifting sinewave grating) and highly synchronized artificial input generated by electrical stimulation of the LGN and/or the axons connecting the two areas. Note that electrical stimulation will excite both geniculo-cortical and cortico-geniculate fibers, producing orthodromic (i.e., synaptic) and antidromic (i.e., non-synaptic "backfiring") stimulation of cells in

V1. We measured V1 responses using a high-density Neuropixel probe comprised of 384 contacts spanning ~3.8 mm in depth, allowing us to record from the full, or nearly full, extent of individual columns in opercular V1. From the electrophysiological signals recorded by the probe we derived three types of V1 responses: two complementary measures based on the local field potential, the broadband ERP and the time-varying gamma envelope, as well as the single and multi-unit (combined) spiking activity (see (Figure 3.3 and 3.4). To quantify the effect of ongoing delta oscillations, we performed a median split of trials from each stimulus type (visual and electrical) based on the amplitude of LFP delta oscillations within the putative thalamo-recipient layer 4C over a short time window prior to stimulus onset (1 second for visual and 0.5 seconds for electrical stimuli). We then compared responses between the below and above median trial groups by simply computing the response difference (below minus above), focusing on whether any effects we saw differed between different depth compartments that were defined based on current source density analysis of visually driven LFP signals (Figure 3.2, see Identification of cortical compartments).

3.4.1.1 Responses to visual stimuli Across 22 recordings from 8 recording sites from 2 subjects, the effect of pre-trial delta amplitude on visual responses was relatively robust, and varied considerably by depth compartment (Figures 3.9, 3.10 & 3.11 subpanels A & B), with the clearest depth structure seen in the ERP data. The ERP data showed a clear depth ordering with the superficial compartment (C-A) showing the most negative difference (i.e., above > below) and the deepest compartment (C-E) showing the most positive difference (i.e., below > above), with intermediate compartments showing intermediate effects (see Figure 3.9). Note that, unlike the other activity measures, ERPs can show positive or negative responses to a stimulus (TVGE and spiking data can only be zero or positive), thus the ERP results are a little trickier to interpret. However, the example recordings shown in Figure 3.4 and Figures 3.3 and 3.6 - 3.8, are representative of the data set as a whole, with superficial compartments showing smaller, positive responses early (within ~0.2 seconds of stimulus onset), followed by larger, negative responses later (beyond ~0.2 seconds) on below median trials (see Figure 3.6 A-C). Deeper compartments generally show a stronger negative response throughout the stimulus duration, although some recordings mostly show an early difference, and others mostly a late difference. Overall, the early portion (< ~0.2 seconds) of the ERP response indicates stronger responses on above median trials, while the later portion (> ~0.2

seconds) indicates stronger responses on below median trials. This somewhat complex arrangement results in the clear depth ordering seen in Figure 3.9A and C.

The results of the TVGE data are generally clearer, with stronger responses seen on below median trials across all depth compartments (Figure 3.10A & C). While it might be tempting to speculate that the difference between the ERP and TVGE results could be due to the fact that the TVGE is a rectified LFP measure, while the ERP is an "unrectified" measure, that is likely at most a partial explanation, as the ERP is dominated by low-frequency components of the LFP (Nunez, 1990), while the TVGE only includes high-frequency components (40 - 120 Hz, see Gamma-band envelope). In fact, we chose to include both ERP and TVGE measures due to the complementary features of the LFP that they highlight.

The results of the spiking data were somewhat different than those of the LFP derived measures. Like the TVGE data, all compartments showed a stronger response on below median trials (Figure 3.11A & C), and like the TVGE data the depth structure was relatively weak. However, compartment B (which putatively corresponds to layer 4AB, see Caveats below) clearly showed the strongest effect. Why the effect of pre-trial delta would be largest in compartment B is not clear. Layer 4B, which is expected to comprise the majority of the 4AB region (Callaway, 1998), receives its principal input from layer 4Cα which primarily carries information from the magnocellular relay cells of the LGN that exhibit high-contrast sensitivity, prefer low spatial and high temporal frequency, and are often direction selective (Hawken et al., 1988; Movshon and Newsome, 1996). As one might expect given the properties of its inputs, layer 4B sends a major projection to the middle temporal area that is well known for its role is analyzing visual motion (Born and Bradley, 2005). However, why ongoing delta oscillations would specifically modulate layer 4B (or 4A) is not clear. Another interesting trend in the spiking data that was different than what was observed in the TVGE or ERP data is that the deepest compartment, C-E (putative layer 6) showed the weakest effect. While the reason for this also difference are not clear, the spiking responses from the C-E were generally quite weak across all recordings. While the ERP and TGVE responses were more robust, they were still generally weaker in C-E than in other compartments (compare Figures 3.3 and 3.4, subpanels A and C with subpanel E). This could well be due to the use of anesthesia, and anesthesia is known to greatly affect layer 6 (Sellers et al., 2015). Thus, the weak effect of pre-trial

delta in compartment C-E seen in Figure 3.11 A and C could be a floor effect. Further work in awake, behaving animals is needed to resolve this issue.

3.4.1.2 Responses to electrical stimuli

Our goal in utilizing electrical stimuli was to provide V1 with a less variable (trial to trial) and more synchronous input. However, the responses to electrical stimuli proved to be more variable than the responses to visual stimuli. The shock-driven ERP response was the most robust (Figures 3.3B, 3.4B, 3.6D-F), and showed a modest effect of pre-trial (i.e., pre-shock) delta (Figure 3.9B & D), and a modest degree of variation in that effect with depth (Figure 3.9B & D). However, the shock-driven TVGE and spiking responses were weaker (Figures 3.3D & F, 3.4D & F, 3.7 & 3.7D-F) showing a small, positive effect that was indistinguishable across compartments (Figure 3.10B & D, Figure 3.10B & D). Overall, the electrical data certainly do not contradict the visual data, and could be reasonably considered "weakly consistent"; however, the electrical data are in general too ambiguous to draw much of a conclusion from. On the other hand, the electrical data do provide independent corroboration of the CSD derived compartment boundary locations, as the depth pattern seen in the shock-aligned responses of all three response types conforms quite well to the compartment boundaries, which were estimated from wholly independent data (see Figures 3.3 & 3.4).

3.4.2 Caveats

Indeed, one of the main potential weaknesses with this work is the uncertainty surrounding the compartment boundary locations and the relationship between compartment boundaries and anatomical layers. In this study, compartment boundaries were derived from the application of an image segmentation algorithm to the CSD depth map (see Figure 3.2 and Identification of cortical compartments). The segmentation algorithm did incorporate some prior knowledge of what anatomical layers "look" like, such as the typical thickness of layers within V1, but in general the compartments were defined largely on the basis of the time-varying current density (Figure 3.2), and so the boundaries are only as precise as the electrical (i.e., local field) response is distinct between compartments. As the result, the boundary between layer 4C and 5 could be estimated very reliably due to the prominence of the short latency sink within layer 4C (Figure 3.2F), and likewise the

correspondence between the electrically defined "putative layer 4C" (referred to as compartment C or C-C in the [Reults]) and anatomically defined layer 4C could be made with relatively high confidence. Given the 4C-5 boundary, layer 5 could be easily identified as the compartment immediately deep to 4C, and layer 6 as the next, and final, deepest compartment. Likewise, layer 4AB must be superficial-adjacent to layer 4C, and layer 2/3 the next and most superficial to that. Thus, the full depth segmentation could be bootstrapped quite effectively from just a few strong spatial components of the CSD.

However, the precise boundaries for 5-6 and 4AB-2/3 were harder to estimate as those CSD components were in practice much less prominent (see Figure 3.2), making the mapping between electrical compartments and anatomical layers more uncertain. Though in practice the estimated compartment boundaries proved to be surprisingly robust, in that independent recordings often utilizing different stimuli, showed a clear depth organization corresponding to the estimated boundaries. Figure 3.3 is a clear example of this, as the compartment boundaries were estimated from full-field luminance transition data (see Identification of cortical compartments), and the depth-structure of the ERP, TVGE, and spiking data in response to drifting gratings (Figure 3.3A, C, E) and electrical shocks (Figure 3.3B, D, F) all conform quite well to them. The depth structure shown in Figure 3.4 corresponds somewhat more weakly to the estimated boundaries in the TVGE data, but the correspondence in the ERP and spiking data is quite clear. Overall, the correspondence is apparent both in the general structure of the data that may not necessarily reflect strong anatomical priors, such as the visual TVGE data (Figure 3.3C) or electrical ERP (Figure 3.3B and fig. 3.4B), but also in the structure that does, such as the localization of shock-driven spiking and TVGE responses to putative layer 4C (Figure 3.3D & F, (Figure 3.4F). Thus, while precise statements about anatomical layers cannot be made without direct anatomical confirmation of the probe's position, we strongly suspect that our findings would remain largely, if not entirely, unchanged were that confirmation available.

Three further points are worth considering in the interpretation of these data. First, all data were collected under anesthetized conditions, which tends to depress cortical activity in general, but most strongly so in the superficial and deep layers (Sellers et al., 2015; Sorrenti et al., 2021). Thus, we cannot be sure that our results will generalize to the awake, behaving state. However, that work

is currently underway, and any findings that replicate between the anesthetized and awake state will be made all the more robust for it, and any discrepancies that may arise can further deepen our understanding of how anesthesia affects cortical and cortico-thalamic circuits.

Second, with respect to the shock-driven data, it is important to keep in mind that, as all shock occurred during visual stimulation with a drifting sinewave grating, the grating phase at which the shocks occurred can have a significant impact on the observed responses. This can happen if the shocks all occur at the same (or similar) grating phase, in which case the visual response will not average out and might mask, or otherwise complicate the interpretation of, shock-driven effects. Similarly, variation across recordings or recording sites in the timing of the shock with respect to the preferred grating phase of the population of cells being recorded from could lead to variability in shock-driven effects. For example, if the shocks all arrived at the grating phase that best drives the cortical column under study at one recording site, but at a non-preferred phase at another recording site, the shock-driven effects could average out (in the worst case) or end up being highly variable between the recording sites.

Finally, variability in the retinotopic alignment between the stimulating electrode and the Neuropixel probe across recording sites could further increase the variability of shock-driven effects. For example, large retinotopic separation is likely to result in weak stimulation of V1, and thus weak, if any, effects of (e.g., pre-trial delta. In addition, intermediate separation could result in some degree of "surround effects" that could suppress activity, or elicit a complex mixture of excitation and suppression, at the recording site. Given the technical challenge of retinotopically aligning recording and stimulating devices between the LGN and V1, we were unable to control for, or systematically vary, the degree of retinotopic separation in our data. Our goal in placing the stimulting electrode was to reach as much of the LGN and/or optic radtions as possible in hopes of stimulating some fibers (as many as possible) that terminate in, or emmanate from, our recording site in V1. All of these scenarios are likely present in our data, and are likely to contribute to the high degree of variability seen in the shock responses.

3.4.3 Relationship to previous work

Our main finding, that strong pre-stimulus delta oscillations lead to weaker evoked responses (both visual and electrical), is consistent with the majority of previous studies looking at the effect of pre-stimulus oscillations on neural activity and visual performance. The largest body of evidence relates specifically to alpha oscillations (8 - 12 Hz), where it has been well documented that stronger pre-stimulus alpha activity typically leads to weaker evoked responses in the EEG (Mathewson et al., 2011; Rajagovindan and Ding, 2011; Iemi et al., 2017; Iemi et al., 2019), and increased visual performance (Thut et al., 2006; Hanslmayr et al., 2007; Händel et al., 2011; Chaumon and Busch, 2014). Similar results have been reported for other modalities (Haegens et al., 2011; Kubetschek and Kayser, 2021). In this work, we examined the effect of pre-stimulus delta on visual responses, following the work of Zerlaut et al. (2022) showing that an index of network state derived from delta oscillation amplitude strongly correlated with membrane potential fluctuations of nearby neurons and spiking activity within the local cortical network. Their network state index (NSI) is essentially a delta-gamma ratio (i.e., 8 - 12 Hz amplitude relative to 40 - 120 Hz amplitude), and their data show that the spiking activity of nearby neurons exhibits a "U-shaped" relation with their NSI, with somewhat higher activity during periods of high delta-gamma ratio (negative NSI) compared to intermediate values (NSI near zero), and much higher activity during periods where the delta-gamma ratio was low (positive NSI) (Zerlaut et al., 2022). Two aspects of their finding are worth noting in the the context of the present work. First, the "U-shaped" function that Zerlaut et al. (2022) report is highly asymmetric, indicating that in general stronger delta oscillations correlate with weaker cortical responses, except at very high level of delta where overall activity increases slightly, but in a manner that is strongly linked to the phase of the delta oscillation (Zerlaut et al., 2022), reminiscent of what has been reported for alpha oscillations (Haegens et al., 2011). Second, their index was developed using data from the somatosensory cortex awake mice and validated with data from the visual cortex of awake mice. Thus, the index may not generalize perfectly to V1 of anesthetized primates. In particular, in our data the gamma oscillations were generally so weak, likely due to anesthesia, that using a delta-gamma ratio was either unstable (i.e., became very large during periods of extremely low gamma) or, if stabilized by using the NSI of Zerlaut et al. (2022) or some similar approach, was not appreciably different from simply using delta amplitude.

4 General discussion

The principal aim of this work was to investigate the cellular and circuit functions underlying interareal communication in the early visual system. While a great deal of work has been done that seeks to uncover the visual features represented, or imaging processing operations performed, by neurons within the visual system, we instead chose to focus on how neurons respond to the activity of their principal, anatomical inputs, and how the ongoing activity within visual circuits changes how neural populations respond to their principal inputs. In many ways this is a subtle change in focus, from thinking about what neurons compute from images, to thinking about what they compute from their synaptic inputs. The hope is that by narrowing the scope of what we want to understand, from the broader spikes-from-pixels goal to a more narrow spikes-from-spikes goal, a more implementation-focused understanding, and hopefully more accurate models, can be built that represent more "atomic" neural computations. These atomic computations could then be combined and connected, following anatomical motifs, and driven by the output of models of phototransduction and retinal image processing (Guo et al., 2014; Pan et al., 2019; Schröder et al., 2020). The goal of these models would be to reproduce the results of spikes-from-pixels models (Bashivan et al., 2019; Cadena et al., 2019; Ponce et al., 2019; Shah et al., 2020; Mounier et al., 2021), but in a manner that allows for explanations in terms of the collective behavior of simpler, more atomic (and thus hopefully better understood) components.

In a small step towards that broad, ambitious goal, we used a general, flexible modeling approach to approximate the computations that LGN relay cells perform on their direct retinal inputs. In many ways, the retinogeniculate pathway is an ideal system in which to study these kinds of neural input-output relations, as although the circuit is complex (see General introduction), it is much simpler than any other stage of the mammalian visual system within the central brain. We found that, at least in the anesthetized state, the computations that LGN relay cells perform on their direct retinal inputs are fairly simple and can be well accounted for by temporal summation due to (e.g.) the slow decay of excitatory post-synaptic potentials (EPSPs, see Chapter 2: Results). Importantly, prior work suggesting that temporal summation is sufficient to explain LGN filtering of retinal spike trains focused largely or exclusively on the immediately preceding retinal ISI. Here,

we employ a more flexible approach, namely GLMs, that allows the data to settle the question of whether ISI filtering specifically, and temporal summation of EPSPs more generally, is sufficient to explain how the LGN transforms retinal spike trains. Our analyses demonstrate that knowing the full pattern of retina spikes only marginally improves predictions about which retinal spikes were relayed compared with only knowing the immediately preceding ISI (see Figure 2.6 and 2.7). This suggests that ISI filtering is indeed the dominant mechanism by which the LGN transforms retinal spike trains, at least within the system that we studied. However, further analyses showed that the form of this ISI filtering is not static, and that for stimuli that induce significant fluctuations in network activity (e.g., correlated stimuli such as drifting gratings or, less extremely, natural images) LGN integration time is dependent on the activity level in a manner consistent with gain control (i.e., low gain during high activity and vice versa).

However, two important, outstanding questions remain, the first being to what degree this result applies to the awake state, where the non-retinal inputs to the LGN are thought to play a more significant role in modulating LGN activity (Hirsch et al., 2015; Usrey and Alitto, 2015). While the awake data that we had acces too showed the same pattern of results, in fact more strongly so (see Figure 2.2 and Supporting Figure 6.2), the data set is too small to draw any statistical conclusions and thus further work is needed. The second question is to what degree, if at all, does this result generalize to neurons in other areas of the visual pathway. For example, if one could simultaneously record from a significant fraction of the LGN relay cells that synapse upon a single cortical cell in layer 4C of V1, would temporal summation, both within and between inputs, provide an equally good account of the V1 cell's activity as it does for LGN relay cells? Methods that allow for simultaneous recordings of the activity of individual neurons and many of the pre-synaptic cells that feed them will be critical in answering these questions and clarifying the computations performed by single neurons more generally.

We took a complementary approach in Chapter 3, where we investigated how the state of circuits within V1 affects how those circuits respond to input from the LGN. The general idea driving this line of work is that, although one would expect the algorithmic operations (by Marr's definition Marr (2010)) performed by V1 circuits to be stable in the face of non-visually driven fluctuations in the network, the details of how those operations are implemented by the circuits that comprise the

network may change substantially. We had hoped that the use of direct, artificial stimulation (via electric shocks) of thalamocortical fibers would provide a more consistent input to V1, and thus allow us to measure changes in V1 response due to changes in its own internal state more readily. Ultimately, the cortical response to electrical stimulation proved to be too small and variable to provide any additional insights beyond those afforded by visual responses. Our findings, consistent with previous work, showed that cortical synchronization at low frequencies generally weakens V1 visual responses. Interestingly, we observed an augmentation of the earliest component of the ERP response (i.e the onset response) with increasing low-frequency synchronization. The ERP response to later (i.e., non-onset) cycles of the drifting grating stimulus was reduced following periods of stronger delta synchronization, similar to what we observed for high-frequency LFP and spiking activity (see Chapter 3: Results).

Within the broader spikes-from-spikes approach, Chapter 2 and Chapter 3 focus on complementary aspects of the same underlying scientific problem of trying to understand how neural circuits compute. In Chapter 2 we focused on the specific operations that individual neurons utilize to transform their inputs into outputs, and how those operations might depend on the state of the network. We showed that, in the case of the LGN, the dominant operation appears to be dynamic, or "state sensitive", ISI filtering, in that the integration time of the filtering is much shorter during period of intense activity. In Chapter 3 we focused on how ongoing activity within the network at the moment when visual input arrives changes how the network responds to that input. In broad agreement with Chapter 2, we found that V1 visual responses are also sensitive to the state of the network, in that the magnitude of V1 spiking and local field responses varies with the strength of ongoing delta oscillations, and does so in a layer dependent manner (Figure 3.10). In a sense, one can consider three factors that together determine how a visual area (or any sensory area) responses to sensory input: 1) the structure of the input itself, 2) the transformation that is being applied to that input, and 3) the ongoing state of the receiving neural circuit at the moment when the input arrives. In Chapter 2 we were able to examine all three factors within a single data set, a testament to the power of paired cell recordings, and in Chapter 3 we focused on just the third factor. However, by utilizing the unprecedented resolution of modern high-density neural probes we were able to resolve variations in state dependent response modulation between the cortical

layers.

The downside to the broader program of trying to understand visual processing in terms of simpler, more "atomic" neural computations (i.e., at the implementation level in Marr's framework) is that if one succeeds in building such a bottom-up model that can predict neural responses as well as visual behavior, such a model may nonetheless fail to offer an explanation of how the atomic operations carried out by neurons, and the further interactions of those neurons, actually produces the collective computational function of the network. Indeed, this is the same problem that arises in the study of deep artificial neural networks (DNNs), where a perfect understanding of network architecture, neural operations and interactions, and computational purpose fail to adequately explain the abilities that DNNs have been shown to possess (Saxe et al., 2019; Bahri et al., 2020; Nayebi et al., 2020). This could reasonably be seen as an incomplete understanding at Marr's algorithmic level (Marr and Poggio, 1976; Marr, 2010). Although the algorithms for building and training such DNN models are perfectly understood, the algorithms that the DNNs effectively implement (or approximate) in order to accomplish their given task are not. Thus, much like biological neural networks, DNNs remain a "black box" (Bahri et al., 2020).

However, regardless of whether one prefers a spikes-from-pixels or a spikes-from-spikes approach (or perhaps more accurately, a spikes-from-spikes-…-from-pixels approach), understanding vision requires an explanation of how the collective, computational function of neural networks emerges from the activities and wiring of individual cells and circuits. The spikes-from-spikes approach simply addresses this problem from the bottom up, whereas the spikes-from-pixels framework takes a more top-down (or perhaps middle-down) approach.

The great promise of DNNs for neuroscience is that, because of their accessibility and relative simplicity, their investigation could produce general-purpose analytical and theoretical tools, and perhaps even a full mathematical framework, for understanding how neural networks, both biological and artificial, function. However, even a complete explanation for how DNNs work will at most help to explain the function of biological neural networks at the algorithmic level. It will still be up to experimental neuroscience to figure out how biology has managed to implement (or more specifically, evolve implementations of) those solutions.

5 References

Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. Journal of the Optical Society of America A, Optics and Image Science 2:284–299.

Alitto HJ, Rathbun DL, Fisher TG, Alexander PC, Usrey WM (2019a) Contrast gain control and retinogeniculate communication. European Journal of Neuroscience 49:1061–1068.

Alitto HJ, Rathbun DL, Vandeleest JJ, Alexander PC, Usrey WM (2019b) The Augmentation of Retinogeniculate Communication during Thalamic Burst Mode. Journal of Neuroscience 39:5697–5710.

Alonso J-M, Usrey WM, Reid RC (1996) Precisely correlated firing in cells of the lateral geniculate nucleus. Nature 383:815–819.

Angelucci A, Bressloff PC (2006) Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. Progress in Brain Research 154:93–120.

Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. Science (New York, NY) 273:1868–1871.

Babadi B, Casti A, Xiao Y, Kaplan E, Paninski L (2010) A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus. Journal of Vision 10:22.

Bahri Y, Kadmon J, Pennington J, Schoenholz SS, Sohl-Dickstein J, Ganguli S (2020) Statistical Mechanics of Deep Learning. Annual Review of Condensed Matter Physics 11:501–528.

Bashivan P, Kar K, DiCarlo JJ (2019) Neural population control via deep image synthesis. Science (New York, NY) 364:eaav9436.

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical Microcircuits for Predictive Coding. Neuron 76:695–711.

Bennett C, Arroyo S, Hestrin S (2013) Subthreshold Mechanisms Underlying State-Dependent Modulation of Visual Responses. Neuron 80:350–357.

Benwell CSY, Tagliabue CF, Veniero D, Cecere R, Savazzi S, Thut G (2017) Prestimulus EEG Power Predicts Conscious Awareness But Not Objective Visual Performance. eNeuro 4.

Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A Fresh Approach to Numerical Computing. SIAM Review 59:65–98.

Born RT, Bradley DC (2005) Structure and Function of Visual Area Mt. Annual Review of Neuroscience 28:157–189.

Briggs F, Mangun GR, Usrey WM (2013) Attention Enhances Synaptic Efficacy and Signal-to-Noise in Neural Circuits. Nature 499:476–480.

Briggs F, Usrey WM (2007) A fast, reciprocal pathway between the lateral geniculate nucleus and visual cortex in the macaque monkey. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 27:5431–5436.

Busch NA, Dubois J, VanRullen R (2009) The phase of ongoing EEG oscillations predicts visual perception. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29:7869–7876.

Busch NA, VanRullen R (2010) Spontaneous EEG oscillations reveal periodic sampling of visual attention. Proceedings of the National Academy of Sciences of the United States of America 107:16048–16053.

Butts DA, Cui Y, Casti ARR (2016) Nonlinear computations shaping temporal processing of precortical vision. Journal of Neurophysiology 116:1344–1357.

Cadena SA, Denfield GH, Walker EY, Gatys LA, Tolias AS, Bethge M, Ecker AS (2019) Deep convolutional models improve predictions of macaque V1 responses to natural images. PLoS computational biology 15:e1006897.

Callaway EM (1998) Local circuits in primary visual cortex of the macaque monkey. Annual Review of Neuroscience 21:47–74.

Cano M, Bezdudnaya T, Swadlow HA, Alonso J-M (2006) Brain state and contrast sensitivity in the awake visual thalamus. Nature Neuroscience 9:1240–1242.

Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC (2005) Do We Know What the Early Visual System Does? Journal of Neuroscience 25:10577–10597.

Carandini M, Horton JC, Sincich LC (2007) Thalamic filtering of retinal spike trains by postsynaptic summation. Journal of Vision 7:20.

Casti A, Hayot F, Xiao Y, Kaplan E (2008) A simple model of retina-LGN transmission. Journal of Computational Neuroscience 24:235–252.

Chaumon M, Busch NA (2014) Prestimulus neural oscillations inhibit visual perception via modulation of response gain. Journal of Cognitive Neuroscience 26:2514–2529.

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network (Bristol, England) 12:199–213.

Cleland BG, Dubin MW, Levick WR (1971) Simultaneous recording of input and output of lateral geniculate neurones. Nature: New Biology 231:191–192.

Cocchi L, Gollo LL, Zalesky A, Breakspear M (2017) Criticality in the brain: A synthesis of neurobiology, models and cognition. Progress in Neurobiology 158:132–152.

DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? Neuron 73:415–434.

Dou W, Morrow A, Iemi L, Samaha J (2022) Pre-stimulus alpha-band phase gates early visual cortex responses. NeuroImage 253:119060.

Douglas RJ, Martin KA (1991) A functional microcircuit for cat visual cortex. The Journal of Physiology 440:735–769.

Douglas RJ, Martin KAC, Whitteridge D (1989) A Canonical Microcircuit for Neocortex. Neural Computation 1:480–488.

Efron B (1987) Better Bootstrap Confidence Intervals. Journal of the American Statistical Association 82:171–185.

Ergenoglu T, Demiralp T, Bayraktaroglu Z, Ergen M, Beydagi H, Uresin Y (2004) Alpha rhythm of the EEG modulates visual detection performance in humans. Brain Research Cognitive Brain Research 20:376–383.

Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cerebral Cortex (New York, NY: 1991) 1:1–47.

Fisher TG, Alitto HJ, Usrey WM (2017) Retinal and Nonretinal Contributions to Extraclassical Surround Suppression in the Lateral Geniculate Nucleus. Journal of Neuroscience 37:226–235.

Garcia-Marin V, Kelly JG, Hawken MJ (2019) Major Feedforward Thalamic Input Into Layer 4C of Primary Visual Cortex in Primate. Cerebral Cortex 29:134–149.

Gehring J, Widmann D, Kleinschmidt D, Finnegan R, Kamiński B, Bowers C, Molignano M, Bouchet-Valat M, Mogensen PK, Kelman T, Ignatiadis N (2021) Juliangehring/Bootstrap.jl: Bootstrap v2.3.3.

Ghanbari A, Malyshev A, Volgushev M, Stevenson IH (2017) Estimating short-term synaptic plasticity from pre- and postsynaptic spiking. PLOS Computational Biology 13:e1005738.

Gilbert CD, Sigman M (2007) Brain States: Top-Down Influences in Sensory Processing. Neuron 54:677–696.

Guo T, Tsai D, Bai S, Morley JW, Suaning GJ, Lovell NH, Dokos S (2014) Understanding the retina: A review of computational models of the retina from the single cell to the network level. Critical Reviews in Biomedical Engineering 42:419–436.

Gur M, Beylin A, Snodderly DM (1997) Response Variability of Neurons in Primary Visual Cortex (V1) of Alert Monkeys. The Journal of Neuroscience 17:2914–2920.

Haegens S, Nácher V, Luna R, Romo R, Jensen O (2011) α-Oscillations in the monkey sensorimotor network influence discrimination performance by rhythmical inhibition of neuronal spiking. Proceedings of the National Academy of Sciences of the United States of America 108:19377–19382.

Haenny PE, Schiller PH (1988) State dependent activity in monkey visual cortex. Experimental Brain Research 69:225–244.

Hamos JE, Van Horn SC, Raczkowski D, Sherman SM (1987) Synaptic circuits involving an individual retinogeniculate axon in the cat. The Journal of Comparative Neurology 259:165–192.

Händel BF, Haarmeier T, Jensen O (2011) Alpha oscillations correlate with the successful inhibition of unattended stimuli. Journal of Cognitive Neuroscience 23:2494–2502.

Hanslmayr S, Aslan A, Staudigl T, Klimesch W, Herrmann CS, Bäuml K-H (2007) Prestimulus oscillations predict visual perception performance between and within subjects. NeuroImage 37:1465–1473.

Hawken MJ, Parker AJ, Lund JS (1988) Laminar organization and contrast sensitivity of direction-selective cells in the striate cortex of the Old World monkey. The Journal of Neuroscience 8:3541–3548.

Heiberg T, Kriener B, Tetzlaff T, Casti A, Einevoll GT, Plesser HE (2013) Firing-rate models capture essential response dynamics of LGN relay cells. Journal of Computational Neuroscience 35:359–375.

Hirsch JA, Alonso J-M, Reid RC, Martinez LM (1998) Synaptic Integration in Striate Cortical Simple Cells. The Journal of Neuroscience 18:9517–9528.

Hirsch JA, Wang X, Sommer FT, Martinez LM (2015) How Inhibitory Circuits in the Thalamus Serve Vision. Annual Review of Neuroscience 38:309–329.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology 160:106–154.2.

Hubel DH, Wiesel TN (1961) Integrative action in the cat's lateral geniculate body. The Journal of Physiology 155:385–398.

Huguenard JR, McCormick DA (1992) Simulation of the currents involved in rhythmic oscillations in thalamic relay neurons. Journal of Neurophysiology 68:1373–1383.

Hunter JD (2007) Matplotlib: A 2D Graphics Environment. Computing in Science Engineering 9:90–95.

Iemi L, Busch NA, Laudini A, Haegens S, Samaha J, Villringer A, Nikulin VV (2019) Multiple mechanisms link prestimulus neural oscillations to sensory responses. eLife 8:e43620.

Iemi L, Chaumon M, Crouzet SM, Busch NA (2017) Spontaneous Neural Oscillations Bias Perception by Modulating Baseline Excitability. Journal of Neuroscience 37:807–819.

Issa EB, Wang X (2011) Altered Neural Responses to Sounds in Primate Primary Auditory Cortex during Slow-Wave Sleep. Journal of Neuroscience 31:2965–2973.

Johnson SG (2020) JuliaPy/PyPlot.jl.

Jun JJ et al. (2017) Fully integrated silicon probes for high-density recording of neural activity. Nature 551:232–236.

Kaplan E, Purpura K, Shapley RM (1987) Contrast affects the transmission of visual information through the mammalian lateral geniculate nucleus. The Journal of Physiology 391:267.

Kaplan E, Shapley R (1984) The origin of the S (slow) potential in the mammalian Lateral Geniculate Nucleus. Experimental Brain Research 55:111–116.

Klein N, Siegle JH, Teichert T, Kass RE (2021) Cross-population coupling of neural activity based on Gaussian process current source densities. PLOS Computational Biology 17:e1009601.

Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: The inhibition-timing hypothesis. Brain Research Reviews 53:63–88.

Koepsell K, Wang X, Vaingankar V, Wei Y, Wang Q, Rathbun DL, Usrey WM, Hirsch JA, Sommer FT (2009) Retinal oscillations carry visual information to cortex. Frontiers in Systems Neuroscience 3:1–18.

Krueger J, Disney AA (2019) Structure and function of dual-source cholinergic modulation in early vision. The Journal of comparative neurology 527:738–750.

Kubetschek C, Kayser C (2021) Delta/Theta band EEG activity shapes the rhythmic perceptual sampling of auditory scenes. Scientific Reports 11:2370.

Kuffler SW (1953) Discharge Patterns and Functional Organization of Mammalian Retina. Journal of Neurophysiology 16:37–68.

Latawiec D, Martin KA, Meskenaite V (2000) Termination of the geniculocortical projection in the striate cortex of macaque monkey: A quantitative immunoelectron microscopic study. The Journal of Comparative Neurology 419:306–319.

Lecoq J, Oliver M, Siegle JH, Orlova N, Ledochowitsch P, Koch C (2021) Removing independent noise in systems neuroscience data using DeepInterpolation. Nature Methods 18:1401–1408.

Llinás R, Jahnsen H (1982) Electrophysiology of mammalian thalamic neurones in vitro. Nature 297:406–408.

Lu SM, Guido W, Sherman SM (1992) Effects of membrane voltage on receptive field properties of lateral geniculate neurons in the cat: Contributions of the low-threshold Ca2+ conductance. Journal of Neurophysiology 68:2185–2198.

Maier A, Aura CJ, Leopold DA (2011) Infragranular Sources of Sustained Local Field Potential Responses in Macaque Primary Visual Cortex. Journal of Neuroscience 31:1971–1980.

Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Transactions on Signal Processing 41:3397–3415.

Marr D (2010) Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Cambridge, MA: The MIT Press.

Marr D, Poggio T (1976) From Understanding Computation to Understanding Neural Circuitry.

Martinez LM, Molano-Mazón M, Wang X, Sommer FT, Hirsch JA (2014) Statistical Wiring of Thalamic Receptive Fields Optimizes Spatial Sampling of the Retinal Image. Neuron 81:943–956.

Mastronarde DN (1987) Two classes of single-input X-cells in cat lateral geniculate nucleus. II. Retinal inputs and the generation of receptive-field properties. Journal of Neurophysiology 57:381–413.

Mastronarde DN (1983) Correlated firing of cat retinal ganglion cells. I. Spontaneously active inputs to X-and Y-cells. Journal of Neurophysiology 49:303–324.

Mathewson KE, Lleras A, Beck DM, Fabiani M, Ro T, Gratton G (2011) Pulsed out of awareness:

EEG alpha oscillations represent a pulsed-inhibition of ongoing cortical processing. Frontiers in Psychology 2:99.

McCormick DA, McGinley MJ, Salkoff DB (2015) Brain state dependent activity in the cortex and thalamus. Current Opinion in Neurobiology, SI: Brain rhythms and dynamic coordination 31:133–140.

McGinley MJ, David SV, McCormick DA (2015) Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron 87:179–192.

McKenna TM, McMullen TA, Shlesinger MF (1994) The brain as a dynamic physical system. Neuroscience 60:587–605.

Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and EEG phenomena. Physiological Reviews 65:37–100.

Mitzdorf U, Singer W (1979) Excitatory synaptic ensemble properties in the visual cortex of the macaque monkey: A current source density analysis of electrically evoked potentials. Journal of Comparative Neurology 187:71–83.

Mogensen P, Riseth A (2018) Optim: A mathematical optimization package for Julia. Journal of Open Source Software 3:615.

Mounier E, Abdullah B, Mahdi H, Eldawlatly S (2021) A deep convolutional visual encoding model of neuronal responses in the LGN. Brain Informatics 8:11.

Movshon JA, Newsome WT (1996) Visual Response Properties of Striate Cortical Neurons Projecting to Area MT in Macaque Monkeys. The Journal of Neuroscience 16:7733–7741.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Receptive field organization of complex cells in the cat's striate cortex. The Journal of Physiology 283:79–99.

Nayebi A, Srivastava S, Ganguli S, Yamins DLK (2020) Identifying Learning Rules From Neural Network Observables.

Nelder JA, Wedderburn RWM (1972) Generalized Linear Models. Journal of the Royal Statistical Society Series A (General) 135:370–384.

Nocedal J, Wright S (2006) Numerical optimization. Springer Science & Business Media.

Norheim ES, Wyller J, Nordlie E, Einevoll GT (2012) A minimal mechanistic model for temporal signal processing in the lateral geniculate nucleus. Cognitive Neurodynamics 6:259–281.

Nunez PL (1990) Physical principles and neurophysiological mechanisms underlying event-related

potentials In: Event-Related Brain Potentials: Basic Issues and Applications, pp19–36. New York, NY, US: Oxford University Press.

Olshausen BA, Field DJ (2005) How close are we to understanding V1? Neural Computation 17:1665–1699.

Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609.

Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. Journal of Neurophysiology 57:162–178.

Pachitariu M, Sridhar S, Stringer C (2023) Solving the spike sorting problem with Kilosort.

Pan G, Tan J, Guo Y (2019) Modeling and simulation of phototransduction cascade in vertebrate rod photoreceptors. BMC Ophthalmology 19:55.

Paninski L, Pillow J, Lewi J (2007) Statistical models for neural encoding, decoding, and optimal stimulus design. Progress in Brain Research 165:493–507.

Paninski L, Pillow JW, Simoncelli EP (2004) Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. Neural computation 16:2533–2561.

Peters A, Payne BR, Budd J (1994) A numerical analysis of the geniculocortical input to striate cortex in the monkey. Cerebral Cortex (New York, NY: 1991) 4:215–229.

Phipson B, Smyth GK (2010) Permutation P-values Should Never Be Zero: Calculating Exact P-values When Permutations Are Randomly Drawn. Statistical Applications in Genetics and Molecular Biology 9.

Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and Decoding of Retinal Ganglion Cell Responses with a Probabilistic Spiking Model. Journal of Neuroscience 25:11003–11013.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454:995–999.

Polack P-O, Friedman J, Golshani P (2013) Cellular mechanisms of brain state–dependent gain modulation in visual cortex. Nature Neuroscience 16:1331–1339.

Ponce CR, Xiao W, Schade PF, Hartmann TS, Kreiman G, Livingstone MS (2019) Evolving Images

for Visual Neurons Using a Deep Generative Network Reveals Coding Principles and Neuronal Preferences. Cell 177:999–1009.e10.

Rajagovindan R, Ding M (2011) From prestimulus alpha oscillation to visual-evoked response: An inverted-U function and its attentional modulation. Journal of Cognitive Neuroscience 23:1379–1394.

Rathbun DL, Alitto HJ, Warland DK, Usrey WM (2016) Stimulus Contrast and Retinogeniculate Signal Processing. Frontiers in Neural Circuits 10:8.

Rathbun DL, Warland DK, Usrey WM (2010) Spike timing and information transmission at retinogeniculate synapses. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 30:13558–13566.

Reid RC, Alonso JM (1995) Specificity of monosynaptic connections from thalamus to visual cortex. Nature 378:281–284.

Reid RC, Victor JD, Shapley RM (1997) The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. Visual Neuroscience 14:1015–1027.

Ringach DL, Shapley RM, Hawken MJ (2002) Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. The Journal of Neuroscience 22:5639–5651.

Romei V, Brodbeck V, Michel C, Amedi A, Pascual-Leone A, Thut G (2008) Spontaneous fluctuations in posterior alpha-band EEG activity reflect variability in excitability of human visual areas. Cerebral Cortex (New York, NY: 1991) 18:2010–2018.

Rust NC, Cohen MR (2022) Priority coding in the visual system. Nature Reviews Neuroscience 23:376–388.

Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. Nature Neuroscience 9:1421–1431.

Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal Elements of Macaque V1 Receptive Fields. Neuron 46:945–956.

Saxe AM, McClelland JL, Ganguli S (2019) A mathematical theory of semantic development in deep neural networks. Proceedings of the National Academy of Sciences of the United States of America 116:11537–11546.

Schreck MR, Zhuang L, Janke E, Moberly AH, Bhattarai JP, Gottfried JA, Wesson DW, Ma M (2022) State-dependent olfactory processing in freely behaving mice. Cell Reports 38:110450.

Schröder C, Klindt D, Strauss S, Franke K, Bethge M, Euler T, Berens P (2020) System Identification with Biophysical Constraints: A Circuit Model of the Inner Retina In: Advances in Neural Information Processing Systems, pp15439–15450. Curran Associates, Inc.

Sellers KK, Bennett DV, Hutt A, Williams JH, Fröhlich F (2015) Awake vs. Anesthetized: Layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. Journal of Neurophysiology 113:3798–3815.

Shah NP, Brackbill N, Rhoades C, Kling A, Goetz G, Litke AM, Sher A, Simoncelli E, Chichilnisky E (2020) Inference of nonlinear receptive field subunits with spike-triggered clustering. eLife 9:e45743.

Shapley R, Enroth-Cugell C (1984) Visual adaptation and retinal gain controls. Progress in Retinal Research 3:263–346.

Sincich LC, Adams DL, Economides JR, Horton JC (2007) Transmission of Spike Trains at the Retinogeniculate Synapse. Journal of Neuroscience 27:2683–2692.

Sincich LC, Horton JC, Sharpee TO (2009) Preserving information in neural transmission. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29:6207–6216.

Snowden RJ, Treue S, Andersen RA (1992) The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. Experimental Brain Research 88:389–400.

Sorrenti V, Cecchetto C, Maschietto M, Fortinguerra S, Buriani A, Vassanelli S (2021) Understanding the Effects of Anesthesia on Cortical Electrophysiological Recordings: A Scoping Review. International Journal of Molecular Sciences 22:1286.

Sutter EE (1987) A practical non-stochastic approach to nonlinear time-domain analysis In: Advanced Methods of Physiological Systems Modeling (Marmarelis VZ ed), Los Angeles, California: Biomedical Simulations Resource.

Thut G, Nietzel A, Brandt SA, Pascual-Leone A (2006) Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 26:9494–9502.

Trenholm S, Krishnaswamy A (2020) An Annotated Journey through Modern Visual Neuroscience. Journal of Neuroscience 40:44–53.

Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A Point Process Framework for Relating Neural Spiking Activity to Spiking History, Neural Ensemble, and Extrinsic Covariate

Effects. Journal of Neurophysiology 93:1074–1089.

Uglesich R, Casti A, Hayot F, Kaplan E (2009) Stimulus size dependence of information transfer from retina to thalamus. Frontiers in Systems Neuroscience 3.

Usrey WM, Alitto HJ (2015) Visual Functions of the Thalamus. Annual review of vision science 1:351–371.

Usrey WM, Alonso J-M, Reid RC (2000) Synaptic Interactions between Thalamic Inputs to Simple Cells in Cat Visual Cortex. The Journal of Neuroscience 20:5461–5467.

Usrey WM, Reppas JB, Reid RC (1999) Specificity and Strength of Retinogeniculate Connections. Journal of Neurophysiology 82:3527–3540.

Usrey WM, Reppas JB, Reid RC (1998) Paired-spike interactions and synaptic efficacy of retinal inputs to the thalamus. Nature 395:384–387.

Vanni S, Hokkanen H, Werner F, Angelucci A (2020) Anatomy and Physiology of Macaque Visual Cortical Areas V1, V2, and V5/MT: Bases for Biologically Realistic Models. Cerebral Cortex (New York, NY) 30:3483–3517.

Vogels R, Spileers W, Orban GA (1989) The response variability of striate cortical neurons in the behaving monkey. Experimental Brain Research 77:432–436.

Wang X, Hirsch JA, Sommer FT (2010) Recoding of Sensory Information across the Retinothalamic Synapse. Journal of Neuroscience 30:13567–13577.

Wertz A, Trenholm S, Yonehara K, Hillier D, Raics Z, Leinweber M, Szalay G, Ghanem A, Keller G, Rózsa B, Conzelmann K-K, Roska B (2015) Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. Science (New York, NY) 349:70–74.

Weyand TG (2007) Retinogeniculate Transmission in Wakefulness. Journal of Neurophysiology 98:769–785.

Weyand TG, Gafka AC (1998) Activity of neurons in area 6 of the cat during fixation and eye movements. Visual Neuroscience 15:123–140.

Williamson RS, Sahani M, Pillow JW (2015) The Equivalence of Information-Theoretic and Likelihood-Based Methods for Neural Dimensionality Reduction. PLOS Computational Biology 11:e1004141.

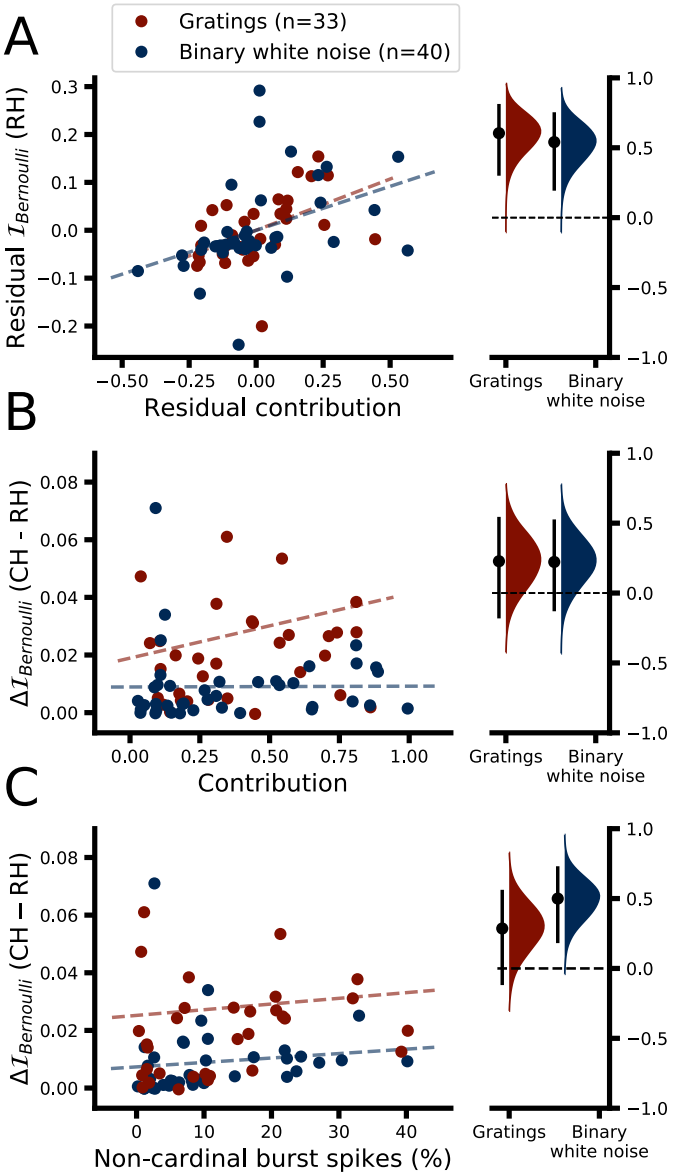
Wilson DE, Scholl B, Fitzpatrick D (2018) Differential tuning of excitation and inhibition shapes direction selectivity in ferret visual cortex. Nature 560:97–101.

Wilson DE, Whitney DE, Scholl B, Fitzpatrick D (2016) Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. Nature Neuroscience 19:1003–1009.

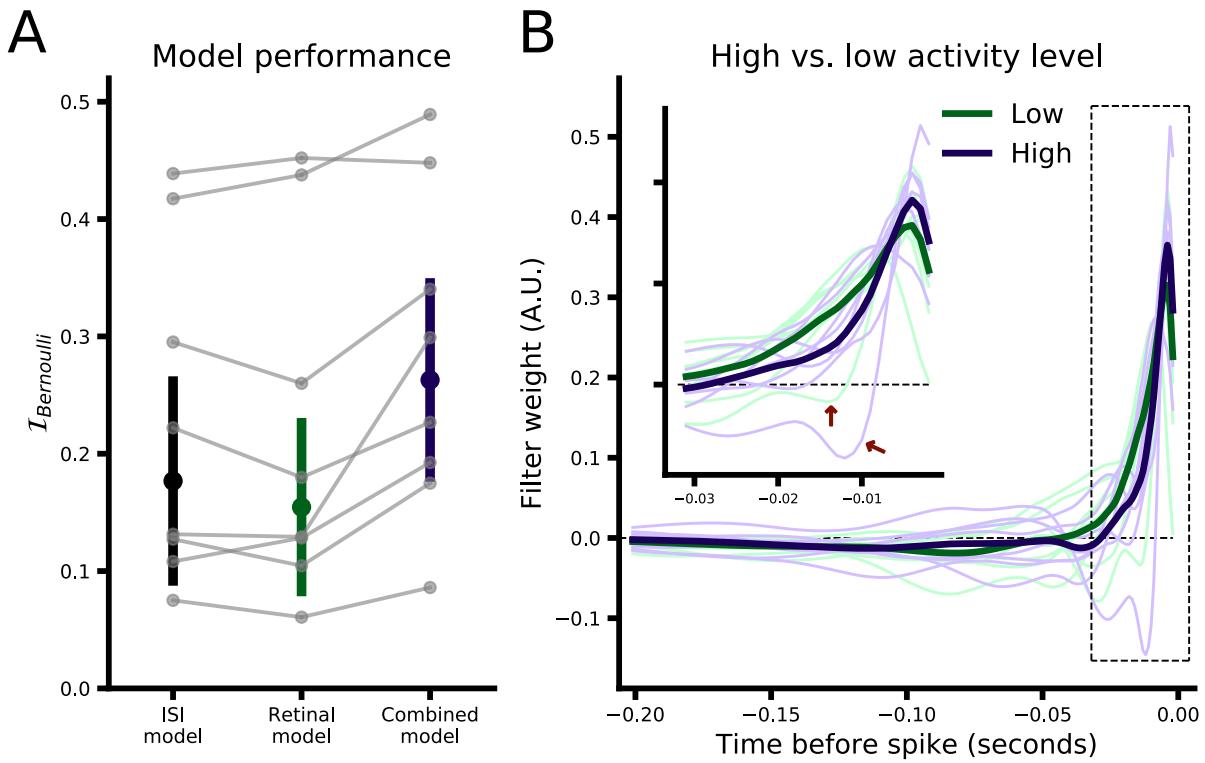
Yamins DLK, DiCarlo JJ (2016) Using goal-driven deep learning models to understand sensory cortex. Nature Neuroscience 19:356–365.

Zerlaut Y, Zucca S, Fellin T, Panzeri S (2022) Network States Classification based on Local Field Potential Recordings in the Awake Mouse Neocortex. eNeuro 9.

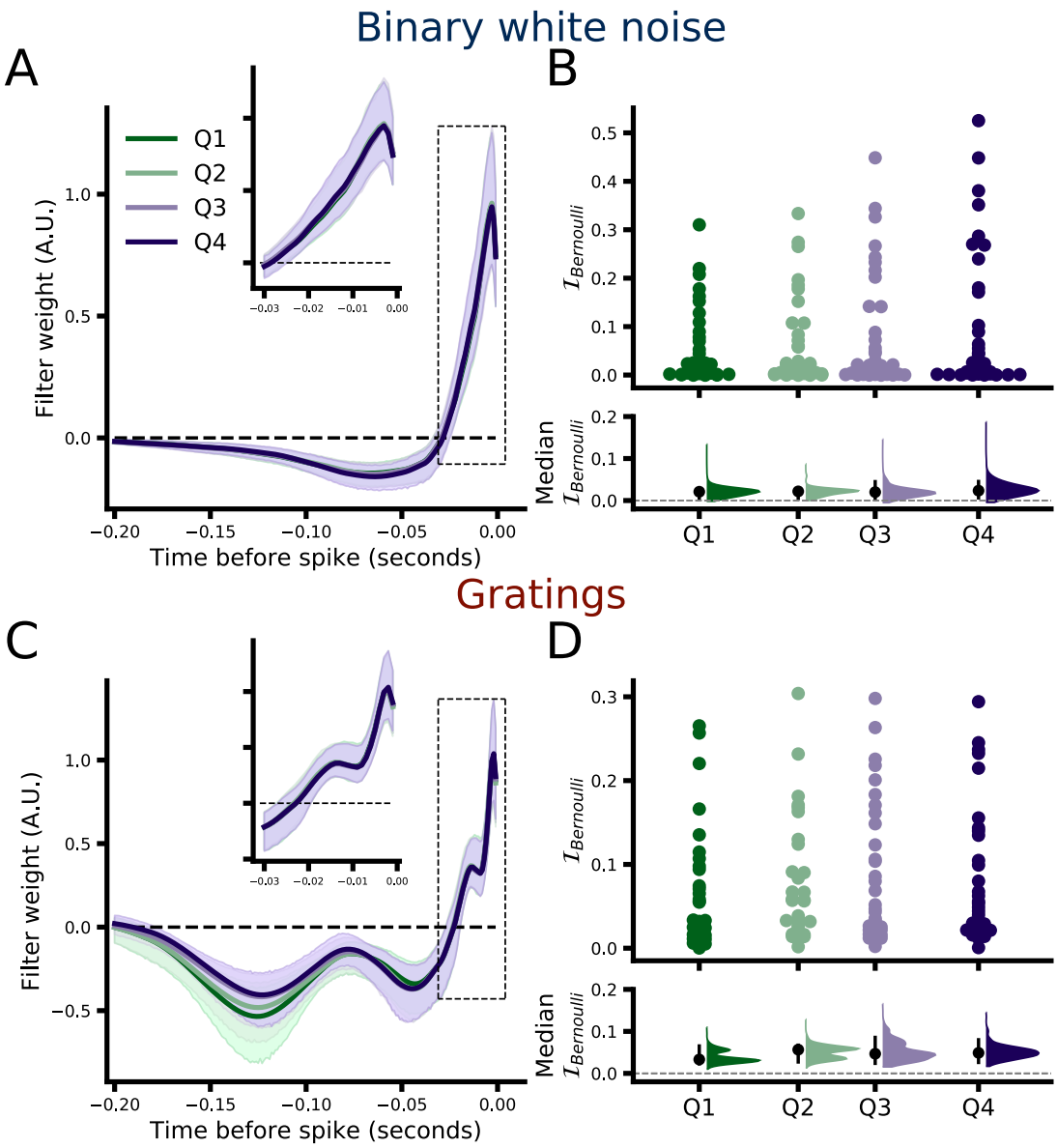
6 Appendix A



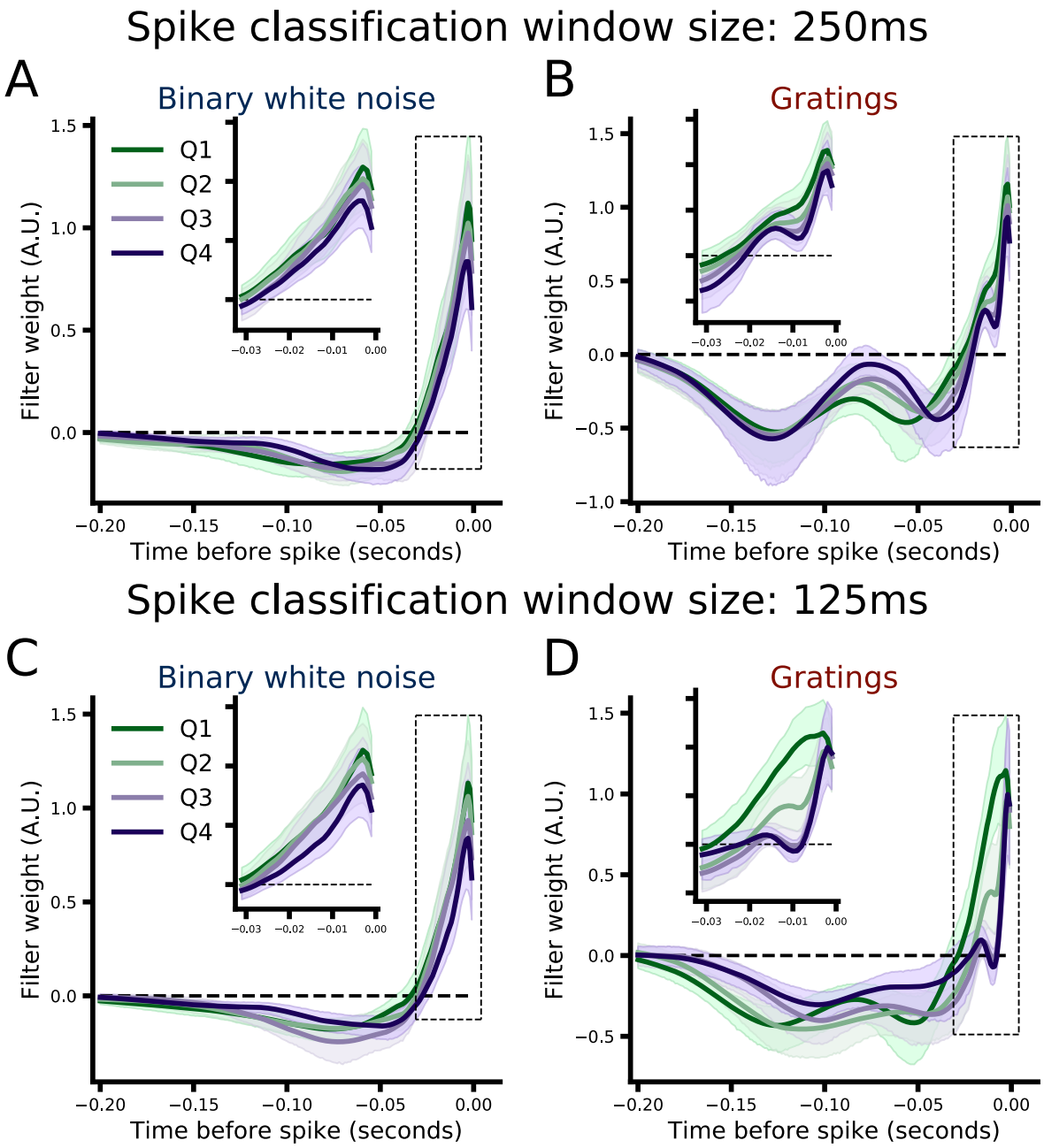
Supporting Figure 6.1: Correlates of model performance. **A**, left: residual Spearmann correlation between $\mathcal{I}_{\text{Bernoulli}}$ from RH models and retinal contribution where the effect of retinal efficacy on each variable has been removed prior to the analysis. Right, estimation of correlation coefficient using 5000 bootstrap resamples. Black dots denote point estimates, vertical black lines denote 95% CI, and filled distributions summarize the results of the resampling. **B**, left: Spearmann correlation between model performance improvement ($\Delta\mathcal{I}_{\text{Bernoulli}}$) between CH and RH models and retinal contribution. As retinal efficacy is not correlated with $\Delta\mathcal{I}_{\text{Bernoulli}}$ regular Spearmann correlation was used. Right, estimation analysis for correlation show at left. **C**, left: Spearmann correlation between $\Delta\mathcal{I}_{\text{Bernoulli}}$ and the percent of LGN spikes that were part of identified bursts (using the traditional criteria of Lu et al. (1992)) excluding the cardinal spike of each burst. Right, estimation analysis for correlation show at left.



Supporting Figure 6.2: Model comparison and activity level analysis for awake data. **A**: Model performance (Mean $\mathcal{I}_{\text{Bernoulli}}$ across folds) for each model and pair (grey points) where grey lines connect points that correspond to the same pair. Large, solid color circles indicate the median, and solid-color vertical lines show the MAD, across pairs for a given model (Black: ISI model, green: retinal (RH) model, purple: combined (CH) model). **B**: Retinal filters learned by RH models from low (green) and high (purple) activity data sets (similar to Figure 6 but utilizing a median split to assign each retinal spike to a data set). Filters from individual pairs are show in less saturated, thin lines while thick saturated lines indicate the mean across pairs (all filters are scaled to have unit norm to aid visualization). Inset axis highlights the boxed region corresponding to the 30 ms immediately preceding each "target spike" (at t=0). The red arrows indicate the filters learned from pair 200001250, which is the only pair of the awake data set that was stimulated with gratings during recording.



Supporting Figure 6.3: Comparison of RH models fit separately to subsets (quartiles) of simulated data grouped by LGN activity level. The relay status of each retinal spike was determined by simulating a RH GLM with a fixed retinal filter (i.e., the filter did not change with activity level). **A**: Average retinal filters from RH models fit to each quartile of the binary white noise data set from low (Q1, green) to high (Q4, purple) based on the activity level of the LGN neuron within a 100 ms period directly preceding the target retinal spike at $t = 0$. Shading represents 95% CI across N=38 pairs. **B Upper**: Comparison of model performance ($\mathcal{I}_{\text{Bernoulli}}$) across all activity subsets. Each dot represents the model performance for a single pair (the spread along the x-axis is to aid visualization). **B Lower**: Bootstrap estimation of median model performance for each subset. Black dots indicate the median across pairs and black vertical lines indicate the 95% CI of the bootstrap distribution (shown in color, 5000 samples). **C, D** Same as **A, B** but for the drifting gratings data set (N=33).



Supporting Figure 6.4: Activity level analysis utilizing different time windows for partitioning retinal spikes. **A**: RH model filters learned from lowest (Q1) to highest (Q4) activity level subsets for binary white noise data where retinal spike assignment is based on a quartile partitioning of LGN spike count within a 250ms window preceding each retinal spike. **B**: Same as **A** but for drifting grating data. **C & D**: Same as **A & B** but using a 125ms window for partitioning.