Transient inhibition to light explains stronger V1 responses to dark stimuli

David St-Amand

Integrated Program in Neuroscience
McGill University
Montreal, Québec, Canada

December 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science
# Table of contents

List of Figures ................................................................. iv  
Abstract .............................................................................. v  
Résumé ................................................................................. vi  
Acknowledgements ............................................................ viii  

Chapter I: General introduction ........................................... 1  
I. Retina and LGN  
   a. Circuitry ........................................................................ 2  
   b. Receptive field properties .............................................. 2  
   c. ON/OFF pathways ...................................................... 4  
II. Primary visual cortex  
   a. Circuitry ......................................................................... 5  
   b. Receptive field properties of simple cells ..................... 5  
   c. Time dynamics ............................................................ 6  
   d. Simple and complex cells .............................................. 6  

IV. Dark-dominance in V1  
   a. Excitatory inputs from LGN ........................................... 8  
   b. Dark-dominance across laminae ..................................... 9  
   c. Alternative explanations ............................................... 11  
   d. Dark-dominance varies across spatial frequencies .......... 11  
   e. Responses are faster to dark than to light stimuli .......... 13  

V. Intracortical inhibition and orientation selectivity .............. 13  

VI. Electrophysiology experiments  
   a. Animal preparation ...................................................... 19  
   b. Neural recording and analysis ....................................... 19  

VII. Machine learning  
   a. Linear regression ........................................................ 21  
   b. Gradient descent ....................................................... 23  
   c. Overfitting ................................................................. 24  
      1. Early-stopping ............................................................ 25  
      2. L2-regularization ..................................................... 27  
      3. Dropout ................................................................. 27  

Chapter II: Transient inhibition to light explain stronger V1 responses to dark stimuli  
I. Introduction ................................................................. 28
II. Methods
   a. Animal preparation ................................................. 30
   b. Extracellular recordings ........................................... 31
   c. Visual stimuli ........................................................ 32
   d. Model architecture .................................................. 33
   e. Optimization and regularization ................................... 36
   f. Two-pass training procedure ...................................... 37
   g. Output non-linearity ................................................ 38
   h. Estimating excitation and inhibition ............................... 38
   i. Orientation selectivity .............................................. 39

III. Results
   a. Example neurons ..................................................... 40
   b. Population responses ............................................... 42
   c. Light/dark balance .................................................. 43
   d. Time dynamics ........................................................ 45
   e. Orientation selectivity .............................................. 48

IV. Discussion
   a. Inferring neural circuitry from machine learning ............ 50
   b. Dark-dominance caused by inhibition to light ................ 51
   c. Time dynamics of dark-dominance ................................. 52
   d. Orientation selectivity .............................................. 53

Chapter III: General discussion
I. Model architecture and V1 circuitry ............................... 55
II. Measuring the dark-dominance effect .............................. 57
III. Time dynamics .......................................................... 61
IV. Future directions ....................................................... 62
V. Concluding remarks ................................................... 64

References ................................................................. 65
List of figures

Figure 1: Receptive fields of ON-center and OFF-center retinal ganglion cells ........3
Figure 2: Blockage of retinal ON-center cells using APB ........................................... 5
Figure 3: Example of a simple cell receptive field ....................................................... 6
Figure 4: Geniculocortical current sinks in the ON and OFF pathways .................9
Figure 5: Dark-dominance in V1 across laminae ..........................................................10
Figure 6: Dark-dominance is stronger at low spatial frequencies .........................12
Figure 7: Contrast-invariant orientation tuning .............................................................14
Figure 8: Orientation selectivity is lower in fast response latencies ....................16
Figure 9: Orientation selectivity of simple and complex inhibitory interneurons ......17
Figure 10: Training and validation error over training epochs .......................26
Figure 11: Schema of the model architecture ...............................................................33
Figure 12: Model fit of two example neurons ...............................................................41
Figure 13: Input strength at the optimal time lag ......................................................43
Figure 14: Dark-dominance across time lags ...............................................................44
Figure 15: Input strength over response latencies ....................................................46
Figure 16: Orientation selectivity across time lags ....................................................49
Abstract

Neurons in the primary visual cortex (V1) receive excitation and inhibition from two different pathways processing lightness (ON) and darkness (OFF). V1 neurons overall respond more strongly to dark than light stimuli (Yeh, Xing and Shapley, 2010; Kremkow et al., 2014), consistent with a preponderance of darker regions in natural images (Ratliff et al., 2010), as well as human psychophysics (Buchner & Baumgartner, 2007). However, it has been unclear whether this "dark-dominance" is due to more excitation from the OFF pathway (Jin et al., 2008) or more inhibition from the ON pathway (Taylor et al., 2018). To understand the mechanisms behind dark-dominance, we record electrophysiological responses of individual simple-type V1 neurons to natural image stimuli and then train biologically inspired convolutional neural networks to predict the neurons' responses. Analyzing a sample of 65 neurons (in anesthetized, paralyzed cats) has revealed their responses to be more driven by dark than light stimuli, consistent with previous investigations (Yeh et al., 2010; Kremkow et al., 2013). We show this asymmetry to be predominantly due to inhibition from the ON pathway rather than by excitation from the OFF pathway. Consistent with V1 responses being faster to dark than to light stimuli (Komban et al., 2014), we find dark-dominance to only occur in the early latencies of neurons' responses. Neurons that are strongly dark-dominated also seems to be less orientation selective. This novel approach gives us new insight into the dark-dominance phenomenon and provides an avenue to address new questions about excitatory and inhibitory integration in cortical neurons.
Résumé

Les neurones du cortex visuel primaire (V1) reçoivent de l'excitation et de l'inhibition provenant de deux voies différentes ; l'une communique la lumière (ON) et l'autre l'obscurité (OFF). Les neurones de V1 produisent plus de décharges lorsque les images sont sombres que lorsqu’elles sont claires (Yeh, Xing et Shapley, 2010; Kremkow et al., 2014). Ces résultats concordent avec une prépondérance de régions plus sombres dans les images naturelles (Ratliff et al., 2010), ainsi qu’avec des résultats similaires chez les humains (Buchner & Baumgartner, 2007). Cependant, on ne sait pas si ce biais pour les images sombres est dû à plus d'excitation de la voie OFF (Jin et al., 2008) ou à plus d'inhibition de la voie ON (Taylor et al., 2018). Pour mieux comprendre les mécanismes derrière ce phénomène, nous enregistrons les réponses électrophysiologiques des neurones V1 à des images naturelles, puis entraînons des réseaux de neurones artificiels inspirés de la neurobiologie pour prédire les réponses des neurones. L'analyse d'un échantillon de 65 neurones (chez des chats anesthésiés et paralysés) a révélé que leurs réponses étaient davantage dictées par des images sombres que lumineuses, ce qui est conforme avec les enquêtes précédentes (Yeh et al., 2010; Kremkow et al., 2013). Nous montrons que cette asymétrie est principalement due à l'inhibition de la voie ON plutôt que par l’excitation de la voie OFF. Nous aussi constatons que ce biais pour les images obscures ne se produit que dans les premières latences des réponses des neurones. Les neurones qui sont fortement dominées par les images obscures semblent également être moins sélectives pour l'orientation de l'image.
Cette nouvelle approche nous permet de mieux comprendre pourquoi les neurones ont des réponses plus fortes pour les images sombres et fournit une avenue pour aborder de nouvelles questions sur l'intégration de l'excitation et l'inhibition dans les neurones corticaux.
Acknowledgements

This work could not have been done without the help of many people. I would like to thank my supervisor Curtis Baker for his great guidance. He not only taught me most of what I know about the visual system, but also was a patient and understanding supervisor who has been a great help whenever technical difficulties occurred. I would also like to thank Guang-Xing Ling for putting a lot of efforts into building the RDP2 software used for data-preprocessing, and for being so friendly and accommodating whenever I had problems with Linux or needed technical support. I would also like to thank both Phillipe Nguyen and Amol Gharat for collecting the electrophysiological data used in this thesis.
Chapter I: General introduction

This thesis’ research is about dark-dominance in the cat’s primary visual cortex. I use a machine learning approach to predict the recorded electrophysiological responses of V1 neurons to natural images. This approach shows V1 neurons tend to respond more to dark than light stimuli on average. This dark-dominance is driven, on average, more by inhibition to light than by excitation to dark. Dark-dominance occurs in early but not in later time lags, especially in neurons that are less orientation selective.

The introduction of this thesis is divided into two different parts. In the first part, I will review the literature relevant to this thesis. I will explain what we already know about the circuitry and receptive field properties of retinal ganglion cells, lateral geniculate nucleus (LGN), and primary visual cortex (V1) neurons. Next, I will delve into the literature more specifically to this thesis. I will discuss the known literature behind the dark-dominance effect, followed by the literature on the role of inhibition in orientation selectivity.

In the second part of the introduction, I will give some background on the methods used in this research. I will first explain how the electrophysiology experiments work, the different types of stimuli used, and how we analyze the neural data afterward. I will then give some background on the machine learning approaches relevant to this work. Finally, I will end with an overview highlighting the key points of this introduction.
Retina and LGN

Circuitry

The visual system first starts in the retina, where photoreceptors receive light and convert it into a neural signal. These photoreceptors then excite bipolar cells, which then excite retinal ganglion cells. Inhibition happens due to horizontal and amacrine cells, which inhibit the photoreceptors and retinal ganglion cells, respectively. Retinal ganglion cells then send their axons through the optic nerve and synapse outside of the retina, in the Lateral geniculate nucleus (LGN). RGN cells also send information to other brain structures, such as the superior colliculus in the midbrain. Next, LGN neurons send feedforward excitatory connections to the primary visual cortex, which in turn projects feedback connections back to the LGN.

Receptive field properties

Retinal ganglion cells have center-surround receptive fields, which are usually either ON-center or OFF-center (see Figure 1). ON-center retinal ganglion cells increase their firing rate when you shine light in the center of their receptive field but decrease their firing rate when you shine light in their surround. OFF-center do the opposite; they decrease their firing rate when there is light in the center but increase their firing rate when there is light in the surround. The inhibition from the surround is much weaker than the excitation from the center: if you shine a light on both the center and surround of an Off-center retinal ganglion cell, the neuron will still fire more than baseline. This increase in firing rate can either be transient or sustained, where transient
neurons increase their firing rate for a short period after the onset of the stimulus, while sustained neurons remain at a higher firing rate for a long period after onset. In both types of retinal ganglion cells, the center part of the receptive field is thought to be due to excitatory connection from bipolar cells, while the surrounding part of the receptive field is thought to be due to inhibition from inhibitory interneurons. LGN neurons also have a center-surround organization very similar to retinal ganglion cells, with the difference that the surround is smaller than in the retina.

Figure 1: Receptive field of ON-center and OFF-center retinal ganglion cells. Taken from Gao, Yang, Chen, and Ni (2009).
ON/OFF pathways

A remarkable property of the early visual system is its segregation into ON and OFF pathways. These two pathways are complementary and roughly symmetric. For example, ON retinal ganglion cells project to ON but not OFF LGN neurons, and OFF retinal ganglion cells project to OFF but not ON LGN neurons. In the retina, ON retinal ganglion cells receive inputs from ON-bipolar cells, while OFF retinal ganglion cells receive inputs from OFF bipolar cells. These two pathways are not completely segregated but do interact with each other. For example, amacrine cells from one pathway can inhibit cells in the other pathway, a phenomenon known as ‘cross-over inhibition’ (Molnar et al., 2009). The two pathways are also segregated within the LGN, where each layer distinctly receives inputs from either the ON or OFF pathway.

ON and OFF pathway neurons in the retina have different characteristics, such as different synaptic receptors. ON retinal bipolar cells mostly rely on metabotropic glutamate receptors such as mGluR6 to increase their firing rate in response to light, while OFF retinal bipolar cells instead rely on ionotropic glutamate receptors such as AMPA and kainate receptors (Nelson & Connaughton, 2012). Ionotropic receptors are faster than metabotropic receptors (Reiner & Levitz, 2018), which suggests the faster responses of OFF than ON LGN cells (Jin et al., 2011) might originate in the retina. Injection of DL-2-amino-4-phospho-nobutyricacid (APB) specifically blocks mGluR glutamate receptors and therefore blocks responses of neurons in the ON but not in the OFF pathway.
(Miller & Slaughter, 1986; see Figure 2). These findings further support the idea that the early system is largely composed of two distinct pathways.

![Graph showing effect of DL-2-amino-4-phospho-nobutyricacid (APB) on responses of bipolar cells in the ON (top panel) and OFF (bottom panel) pathways. The x-axis shows time, while the y-axis shows neural responses to repeated diffuse light stimuli. Taken from Miller and Slaughter (1986).]

**Primary visual cortex**

*Circuitry*

The LGN sends most of its excitatory inputs to layer IV of the primary visual cortex, which integrates inputs from both the ON and OFF pathways. Neurons in Layer IV then send excitatory connections to layers 2/3, 5, and 6. While no inhibitory inputs are coming directly from the LGN (Montero, 1986), there are both excitatory and inhibitory connections within V1.

*Receptive field properties of simple cells*

V1 neurons have receptive fields that are more complex than in the LGN. Instead of being tuned for spots of light or dark, V1 neurons are tuned to bar and
edges. There are two common types of neurons in V1: Simple and complex cells. Simple cells have distinct regions that respond to light and dark just like LGN neurons, except that these regions are elongated (see Figure 3). The receptive field of simple cells can be modeled as two-dimensional Gabor functions, a multiplication between a sinewave and a Gaussian function (Jones & Palmer, 1987).

![Figure 3](image.png)

Figure 3: Example of the receptive field of a simple cell. The neuron increases its firing rate when there is light in the + regions, and when there is dark in the – regions.

These Gabor-like receptive fields respond well to bars and edges, but only if the bar or edge is in a specific orientation. Hence, V1 neurons are usually orientation-selective and fire in response to a limited array of orientations.

**Time dynamics**

Visual neurons have spatio-temporal receptive fields and are responsive to stimuli presented within a recent window of time. For example, the temporal aspect of the receptive fields is what makes some simple-type cortical cells direction-selective. These neurons respond best to a drifting bar or edge because different time lags have different optimal phases.

The spatiotemporal receptive fields of simple cells also tend to reverse in polarity over greater time lags. For example, if a neuron responds best to a light
bar in the center of its receptive field in its early time lags, it might respond best to a dark bar in the same location in later time lags. This phenomenon happens because the responses of simple cells are usually transient, which means for example that a neuron might respond more strongly immediately after a light bar appears than if it has been there for a while. These transient responses are what makes neurons respond more strongly to a stimulus if another stimulus was not previously present.

*Simple and complex cells*

A characteristic of the receptive fields of simple cells is their distinct light and dark-driven regions. However, in complex cells, light and dark regions overlap. While simple cells will greatly vary in their response depending on whether a light bar is within the light or dark-driven region of its receptive field, complex cells will not vary their firing rate much depending on where the bar is within their receptive field. Due to their non-linear nature complex cells are harder to model than simple cells, which is why this thesis focuses more on simple rather than complex cells.

**Dark-dominance in V1**

In textbooks, V1 neurons are usually thought of as having relatively balanced light-driven and dark-driven regions. This is consistent with the idea of Gabor-like receptive fields for simple cells. Such a circuitry would not necessarily be optimal, since natural images have been shown to have more darkness than
lightness (Cooper & Norcia, 2015). Indeed, recent findings show a large majority of V1 neurons to respond more to dark than to light stimuli (Jin et al., 2008; Yeh et al., 2009; Jansen et al., 2019), even though the cortical circuitry that underlies this “dark-dominance” is still unclear. In the next section, I will discuss the literature behind the dark-dominance phenomenon.

*Excitation from lateral geniculate nucleus*

A possible explanation for this “dark-dominance” phenomenon might be a higher number of excitatory synapses from the LGN to V1 in the ON than the OFF pathway. Jin et al. (2008) have done simultaneous electrophysiological recordings in both the LGN and V1 of the cat and mapped receptive fields using reverse correlation. They found that in the area centralis but not in the periphery, there is a higher number of excitatory connections between V1 neurons in layer 4 and OFF LGN neurons than with ON LGN neurons (see Figure 4). Even though the number of LGN neurons in both pathways was approximately the same, they found neurons in the OFF pathway to connect with neurons spanning a horizontal distance across the cortical surface approximately 1.8 times greater than in the ON pathway. This suggests the average LGN neuron from the OFF pathway to have synapses connecting to a larger number of V1 neurons, which might explain why most V1 neurons show dark-dominance.
Figure 4: Near the cortical representation of the area centralis (< 5 degrees eccentricity), pairs between LGN and V1 (current sinks) were more frequently found with off-center geniculate afferents than with on-center geniculate afferents ($P = 0.02$, Chi-square test, data obtained from 9 cats). This difference was not found outside of the area centralis. Taken from Jin et al., (2008).

**Dark-dominance across laminae**

If the dark-dominance effect in V1 was solely due to a higher number of thalamocortical excitatory connections from the OFF than the ON pathway, you would expect the dark-dominance effect in layer 4 to be just as strong – or stronger - in layer 4 compared to the other layers. Surprisingly, layers 2/3 show a stronger dark-dominance effect than layer 4. Yeh, Xing, and Shapley (2009) did electrophysical recordings of area V1 in the monkey. They showed neurons
either white or black sparse noise on a gray background, which allows them to classify neurons as either light-dominated or dark-dominated depending on which type of stimuli excited them most. As you can see in Figure 5, they replicated the dark-dominance effect in layer 4, with approximately 65% of neurons being dark-dominated. However, the dark-dominance effect was most prominent in layers 2/3, with more than 90% of neurons being dark-dominated. This suggests the dark-dominance effect to not only be due to the stronger excitatory thalamocortical inputs from the OFF pathway but also to be strongly enhanced by intracortical circuitry. Two possible mechanisms as to how intracortical circuitry enhances dark-dominance could either be excitatory inputs from other dark-dominated neurons or because of inhibition from light-dominated neurons.

Figure 5: There are more dark-dominant than light dominant neurons across all layers, especially in layers ⅔. Taken from Yeh et al. (2009)
Alternative explanations

As seen above, there is evidence linking dark-dominance with the higher number of thalamocortical excitatory inputs from the OFF pathway. One limitation of these studies is that they are unable to infer the types of inputs behind the dark and light-driven regions of a receptive field. Since there does not seem to be direct inhibitory connections between LGN and V1 (Montero, 1986), it seems unlikely for dark-dominance to be caused by greater thalamocortical inhibition from the ON pathway. However, dark-dominance could also arise due to other sources. For example, it could come from a higher number of excitatory inputs from other dark-dominated cortical neurons. Another possibility is for dark-dominance to be caused by strong inhibitory inputs from light-dominated cortical interneurons. These light-dominated interneurons could transmit or integrate thalamocortical excitatory inputs from the ON pathway, or receive excitatory inputs from nearby light-dominated cortical neurons. As these possibilities have yet to be investigated, it might be too early to conclude dark-dominance to be exclusively due to the greater thalamocortical excitatory inputs from the OFF pathway.

Dark-dominance varies across spatial frequencies

The strength of this dark-dominance depends on the spatial frequency of the stimulus. Janse et al. (2018) recorded from the macaque primary visual cortex and mapped the receptive fields of neurons using sinewave gratings of varying
spatial frequencies and phases. While the dark-dominance effect was strong when mapping receptive fields with low spatial frequency gratings, the size of light-driven and dark-driven regions was approximately similar when mapping receptive fields with higher spatial frequency gratings (Figure 6). This suggests dark-dominance to be most pronounced at lower spatial frequencies (Janse & al., 2018). Kremkow et al. (2014) suggested a possible mechanism for this phenomenon to be the nonlinear saturating response function of the ON pathway. They recorded the responses of V1 and LGN neurons to white or black rectified sinewave gratings as they varied contrast. While they found neurons in the OFF pathway to respond linearly as a function of contrast, the responses of LGN neurons in the ON pathway saturated at higher contrasts. This saturating function could explain why neurons are more responsive to dark at lower but not at higher spatial frequencies.
Figure 6: The number of OFF-dominated RFs in V1 decreases as the grating size decreases and the spatial frequency increases. Taken from Janse et al. (2018).

*Responses are faster to dark than light stimuli*

Another interesting characteristic of dark-dominance is its temporal dynamics. Jin et al. (2011) recorded from the cat’s LGN and found that geniculate neurons in the OFF pathway had, on average, response latencies 3.8 ms faster than neurons in the ON pathway. This temporal difference might possibly originate from the retina and be explained by different glutamate receptors in ON and OFF bipolar cells. OFF bipolar cells have ionotropic glutamate receptors, which are faster than the metabotropic glutamate receptors of ON bipolar cells (Nelson & Connaughton, 2012; Reiner & Levitz, 2018). This time difference has also been found by Komban et al. (2014) in the primary visual cortex, with dark-dominated neurons having response latencies on average 3.45 ms faster than light-dominated neurons. These faster responses might possibly originate from the retina,

**Intracortical inhibition and orientation selectivity**

Inhibitory connections play key roles within V1. For example, V1 intracortical inhibition has played key roles in theories of contrast-invariant orientation tuning. An orientation tuning curve can be measured using a neuron’s response to drifting sinewave gratings at the optimal spatial and temporal
frequencies. Contrast-invariant orientation tuning refers to how the orientation tuning curve widths of simple cells do not vary as a function of contrast (Sclar & Freeman, 1982; Skottun et al., 1987; see figure 7).

Figure 7: Responses of a simple (A) and a complex (B) neurons in cat’s area A17 to sinewave gratings at different orientations (top-left) and different spatial frequencies (top-right). The half-width half-height of the orientation (bottom-left) and spatial frequency (bottom-right) tuning curves as a function of contrast has been plotted on the two bottom panels. Taken from Skottun et al. (1987).

This is a result difficult to explain from a feedforward model with non-existent or even weak inhibition since higher contrast leads to stronger LGN responses (Cheng et al., 1995) which should depolarize V1 neurons above threshold even at non-optimal or null orientations (Ferster & Miller, 2000). Troyer
et al. (1998) simulated excitatory and inhibitory LGN inputs organized in a push-pull manner, where excitation from one pathway is complemented by inhibition from the other. They showed inhibition being stronger than excitation could explain the contrast-invariant orientation tuning of V1 neurons. This inhibition most likely does not arise from direct LGN inhibitory inputs, since inhibitory neurons are known to generally not form connections over long distances, and no physiological evidence for direct thalamocortical inhibitory connections has been found (Ferster & Lindström 1983, Martin & Whitteridge 1984). Instead, inhibition comes from inhibitory interneurons (Isaacson & Scanziani, 2011), with layer 4 inhibitory interneurons receiving strong thalamocortical excitatory inputs (Bruno & Simons, 2002). In the model from Troyer et al. (1998), the simulated inhibitory interneurons do not receive inhibition and do not themselves show contrast-invariant orientation tuning. If these inhibitory neurons were inhibited in a push-pull manner, they would be too weakly responsive to a null or cross-oriented stimulus to counter LGN excitation (Ferster & Miller, 2000). To obtain contrast-invariant tuning, a neuron would need to be inhibited from orientations different from the one it is tuned to.

Using intracellular recordings, Monier et al. (2003) have shown inhibition to be more broadly tuned than excitation. Moreover, they have shown that inhibition can be tuned to a different orientation than excitation, a phenomenon that occurred in roughly 40% of their sampled cells. Shapley, Hawken, and Ringach (2003) have shown the early responses of cat’s V1 neurons to be broadly tuned to orientation, and for responses to non-preferred orientation to
become suppressed over time. This phenomenon is thought to be due to inhibition being slower than excitation, since excitation can be provided directly by the thalamus while inhibition needs to go through inhibitory interneurons. In most neurons, the later suppression of response happened to all non-preferred orientation, suggesting the inhibition received by a simple cell to be unselective to orientation. However, consistent with Monier et al. (2003), they did find a subset of neurons whose suppression in response over time was orientation-selective (see figure 8). Shapley et al. (2003) also found a strong negative correlation between the total amount of inhibition and orientation selectivity, suggesting the possibility of a causal relationship between the two.

Figure 8: Orientation tuning curves at different times for two different neurons in macaque V1. The black line is the orientation tuning curve at the optimal time lag. The red line represents the tuning curve slightly before the optimal time lag, while the blue line is the tuning curve slightly after. Taken from Shapley et al. (2003).
Hirsch et al. (2003) studied the receptive field properties of layer IV inhibitory smooth cells using whole-cell clamp recordings. Out of 10 recorded smooth cells, 6 were simple and 4 were complex. The smooth simple cells varied in their orientation selectivity within ranges similar to excitatory spiny neurons. This suggests push-pull might not be sufficient to fully explain contrast-invariance, as proposed by Troyer et al. (1998). On the other hand, none of the smooth complex cells showed more than a weak orientation selectivity bias (see figure 9); these interneurons fired nearly as strongly (> 75%) to stimuli of all orientations. This makes them strong candidates as the cause for contrast-invariant orientation tuning and could also help explain gain control non-linearities.

![Figure 9: Averaged orientation tuning curves (gray) for simple (left) and complex (right) smooth inhibitory interneurons, with their Gaussian fits (black). Orientation tuning curves for depolarization is shown using dotted traces, and hyperpolarization using dashed traces. Taken from Hirsch et al. (2003).](image-url)
Overall, these findings suggest inhibitory interneurons vary immensely in the nature of their receptive fields and in how they can inhibit V1 simple cells. As predicted by push-pull models, some neurons receive inhibition that is tuned to the neuron’s optimal orientation but tuned to the opposite phase. Inhibition can also happen in cross or even null orientations (Monier et al., 2003; Shapley et al., 2003). Simple inhibitory interneurons have similar orientation tuning curves and do have receptive fields with push-pull receptive inputs (Hirsch et al., 2003). On the other hand, there are also complex inhibitory interneurons which are not orientation selective and might help provide inhibition over a wide range of orientations.

In conclusion, a better understanding of how inhibition affects V1 responses can help us understand the trademark of V1 receptive fields that is orientation selectivity.

Electrophysiology experiments

This thesis uses electrophysiology experiments in the cat to better understand dark-dominance in the primary visual cortex. The cat is one of the species whose primary visual cortex has been studied the most extensively, making it a good animal model to relate our work with the literature. The cat’s primary visual cortex also shares many similarities with macaques and humans, like its laminar organization and how many characteristics of receptive fields – such as preferred orientation and ocular dominance – are organized in columns.
In the following section, I will explain general knowledge about electrophysiology experiments. I will start by discussing the animal preparation necessary to conduct the experiments. I will then end by explaining how the electrophysiological data is obtained and how it is processed after the experiment is over.

**Animal preparation**

There is a procedure that needs to be followed before we can record from a cat’s primary visual cortex. The experiment begins with a strong dose of anesthetic to proceed with surgery. Surgery involves craniotomy and durotomy over A17, where we remove a section of the skull and the dura. This craniotomy and durotomy are necessary so we can later insert electrodes in the primary visual cortex to record neural activity. The surgery also involves a cannula being inserted through one of the animal’s veins to be able to give a continuous injection of anesthesia and paralytic throughout the experiment. Continuous anesthesia is necessary to prevent the animal from being conscious and feeling pain through the experiment. Paralysis is necessary to prevent the cat’s eyes from moving during recording, which would make it much harder for us to characterize a neuron’s receptive field.

**Neural recording and analysis**

How we record neural activity and how we analyze it has significantly changed throughout history. Many decades ago, each electrode only had a single recording channel. Today, we use multi-channel electrodes which have
many recording sites. In this research project, we used 32-channel electrodes with each channel being 100 microns apart and recorded raw data signals with a Plexon Recorder (3 Hz to 8 kHz; sampling rate, 40 kHz). More recent electrodes can even record from more than 300 channels at the same time. Such a technical revolution means researchers can record from a much large number of neurons with fewer experiments.

How we isolate spikes in neural recordings has also greatly changed throughout the years. Researchers used to isolate spikes by looking at the high-pass filtered multi-unit activity and manually set a threshold for the neuron’s spikes. This method was simple but also had its fair share of limitations. The first issue is that whenever a channel recorded spikes from more than a single neuron, it was difficult to identify which spikes came from which neuron. The second limitation is how overwhelming it has become to manually sort spikes with the large amounts of data collected from multi-channel electrodes.

Thankfully, spike-sorting algorithms have now become more advanced. Most of the recent spike-sorting algorithms use Principal-Component-Analysis (PCA), which can automatically cluster spikes by extracting the spike waveforms that can explain most of the variance in the electrophysiology data. This approach still requires some human intervention but has greatly improved the reliability and time required to do spike-sorting. An example of a spike-sorting algorithm using PCA is Spikesorter (Swindale & Spacek, 2014), which is the algorithm we used in this thesis.
Machine learning

We can use different machine learning methods to understand the receptive field of a neuron’s responses to natural images. Using rich stimuli such as natural images makes the learned models more generalizable (Talebi & Baker, 2012) and is helpful when trying to disambiguate excitation and inhibition to light and dark stimuli. Machine learning can then be used to predict a V1 neuron’s response from the visual stimuli. We can then infer the receptive field of the neuron based on the parameters of the machine learning model. In this section, I will be introducing the general concepts behind machine learning. I will begin by introducing simple and multivariate linear regression. I will then explain overfitting, a common issue in machine learning, and the different methods we can use to reduce overfitting.

Linear regression

Simple linear regression predicts the output variable y using only a single predictor variable, where the relationship between the predictor and output variables is assumed to be linear:

\[
\hat{y}_i = \alpha + \beta x_i
\]  

(1)

Where \(\hat{y}_i\) is the predicted response of the \(i^{th}\) example, \(x_i\) is the predictor value of the \(i^{th}\) training example, \(\beta\) is the estimated slope of the relationship between \(x\) and \(y\), and \(\alpha\) is the intercept of the estimated function. Simple linear regression is limited to predicting \(y\) using a single variable, which often is not enough to predict...
y accurately. Thankfully, more advanced methods such as multivariate linear regression allow us to incorporate many variables into a single model:

\[ \hat{y}_i = \alpha + \sum_{j=1}^{p} \beta_j x_{ij} \]  

(2)

Where \( \beta_j \) is the slope of the \( j^{th} \) predictor variable, and \( x_{ij} \) the value of the \( j^{th} \) predictor variable for the \( i^{th} \) example. Multivariate linear regression is popular in both statistics and machine learning as it is one of the simplest and most robust ways to predict \( y \) from multiple variables. Its main limitation is that it assumes the relationship between \( x \) and \( y \) to be linear. In scenarios where this relationship is more complex, other statistical or machine learning methods might be more appropriate.

In both simple and multivariate linear regression, the \( \alpha \) and \( \beta \) parameters are estimated from the data. These values are chosen to minimize the prediction error between \( \hat{y} \) and \( y \), which can be expressed as follow:

\[ L(\beta, \alpha) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]  

(3)

There are two different methods to estimate \( \alpha \) and \( \beta \) from this loss function. The first one is the closed-form algebraic solution, which directly finds the optimal parameter values to minimize the above function. While this method is proven to minimize the error, it has two main downsides. The first downside is that the closed-form solution is especially prone to overfitting, which is when the model learns the noise in the data rather than the true relationship between \( x \) and \( y \). The second downside is that many non-linear machine learning algorithms, such as neural networks, often do not have a closed-form solution. A third
downside is that computing the closed-form solution is very computationally expensive when the dataset is large. For these reasons, we instead use another method to estimate $\alpha$ and $\beta$: gradient descent.

**Gradient Descent**

Stochastic gradient descent is an iterative method to estimate parameter values step-by-step using one training example at a time. This method usually involves randomly choosing initial parameter values drawn from a statistical distribution (e.g. Uniform or Gaussian). This parameter is then slightly updated on each training epoch to further reduce the error. To infer how strongly and in which direction does a parameter needs to change to reduce error, we can compute the derivative of the loss function with respect to parameter $\beta$:

$$
\beta_{w+1} = \beta_w + \eta \frac{d}{d\beta} (y_i - \hat{y}_i)^2
$$

(4)

Where $\beta_w$ is the parameter $\beta$ at training epoch $w$ and $\eta$ is the learning rate. In the case of simple linear regression:

$$
\frac{d}{dB_j} (y_i - \hat{y}_i)^2 = -2x_i(y_i - (\alpha + \beta x_i))
$$

(5)

$\beta$ is updated from the previous step based on both the derivative of the cost function and the learning rate. The learning rate influences how quickly the model learns from specific examples, and its value needs to be chosen from the user. The gradient influences how quickly the parameters are updated and varies depending on the prediction error of specific training examples.
The above formula shows a gradient that is computed one example at a time. Mini-batch gradient descent is very similar to gradient descent with the difference that the cost function is computed for an arbitrary number of training examples each time. Updating the gradient from many training examples at the same time is both less noisy and computationally faster, making mini-batch gradient descent often more advantageous than stochastic gradient descent.

Overfitting

Too much overfitting is a common issue in machine learning. When training a machine learning model on some data, we usually want the model to predict the output (y) from the inputs (x). However, the model might instead focus on predicting individual training examples and their inherent noise. Overfitting is a normal phenomenon in machine learning. However, too much overfitting is problematic since the model is learning noise rather than the true relationship between x and y. The amount of overfitting a model strongly depends on two different factors: The number of training examples and the number of parameters in the model. The fewer training examples we can feed to the model, the easier it will be for the model to focus on these specific examples and miss the true relationship between x and y. When a model has more parameters, it is also more flexible, which means it can better learn specific training examples. Hence, strong overfitting usually occurs when the number of parameters is relatively high compared to the number of training examples.

Unfortunately, it is not always possible to reduce overfitting by significantly increasing the number of training examples. Reducing overfitting often requires
drastically increasing the amount of collected data, which researchers cannot always do due to limited resources. For example, in electrophysiology, the amount of recording time we have per neuron is limited due to technical reasons. Unfortunately, it is also not always preferable to reduce the number of parameters. Making the model less complex is often at the cost of reducing the predictive power of the model. Within the context of this thesis, reducing the number of parameters would mean either decreasing the spatial resolution of the input image or reducing the number of time lags.

Thankfully, computer scientists have designed regularization methods that allow us to reduce overfitting without making the dataset larger or the number of parameters smaller. In the following section, I will explain three regularization methods that are popular and were used in this research project: Early-stopping, L2-regularization, and dropout.

*Early-stopping*

Overfitting means what the model learns does not fully generalize outside of the data it has been trained on. One way to check the amount of overfitting is to compare performance on the training set with performance on a separate validation dataset. Performance can be measured using the loss function shown in equation (3), where higher performance means smaller error. While performance on the training set usually keeps increasing over the training epochs, performance on the validation set usually peaks after a certain number of epochs and then starts to decrease. A schematic version of this phenomenon can be seen in Figure 10. This decrease happens because overfitting becomes
worse as the number of epochs increases, to the point where a significant proportion of what the model learns is the noise within the training examples. A solution to this issue is early-stopping, where we stop training when the performance on the validation set starts to decrease. Early-stopping has proven itself to be powerful in preventing overfitting and has become a standard practice in machine learning.

![Figure 10: Schema of the error in training and validation datasets (y-axis) over epochs (x-axis).](https://zhuanlan.zhihu.com/p/60197273)

The error in the training set keeps decreasing, while the error in the validation set first decreases then starts increasing again due to overfitting. Taken from https://zhuanlan.zhihu.com/p/60197273

Unfortunately, when we use early-stopping, the performance on the validation set becomes a biased estimate of the true performance of the model. There is a certain amount of noise in our estimate of the validation performance, and this noise becomes biased towards high positive values if we stop training whenever validation performance is at its highest. It has, therefore, become good
practice to have a third ‘testing’ dataset used to obtain an unbiased estimate of the model’s performance.

*L2-regularization*

To be able to overfit a model to specific training examples, the training algorithm must set the model’s parameters to very high positive or negative values. L2-regularization helps reduce overfitting by adding a penalty for the square of the parameters’ values to the loss function (Chavent & Kunisch, 1997):

\[
L(\beta, \alpha) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2
\]

\(\lambda\) is the L2-regularization hyper-parameter which controls how strong the L2 penalty is. Since there is no direct way to optimize \(\lambda\), standard practice is to try different \(\lambda\) values and choose the one which gives the highest performance on the validation set. L2-regularization has proven itself to be very efficient in reducing overfitting.

*Dropout*

Dropout is another approach useful in reducing overfitting. When a model overfits, its ability to predict each training example relies on certain parameters. Dropout reduces the number of parameters a model is using on each training epoch by randomly setting a certain amount of the model’s parameters to zero. The parameters which are set to zero changes from epoch to epoch, and the proportion of parameters set to zero is a hyper-parameter which is usually set to 50% for theoretical and experimental reasons (Srivastava et al., 2014).
Now that I have given the background knowledge related to my thesis, the next chapter will be about the research I did during my Master’s. This work is presented in journal article format and includes an introduction, methods, results, and discussion sections.

Chapter II: Transient inhibition to light explain stronger V1 responses to dark stimuli

Introduction

The early visual system is separated into two distinct pathways: The ON pathway which encodes light and the OFF pathway which encodes dark. These two pathways merge when they reach the primary visual cortex (V1). How this integration occurs is crucial in understanding how vision works.

V1 neurons are known to receive asymmetrical inputs from the two pathways, since they are on average more responsive to dark than light stimuli (Jin et al., 2008; Yeh et al., 2009), especially at low spatial frequencies (Kremkow et al., 2014; Jansen et al., 2018) and shorter time latencies (Komban et al., 2014). This is presumably adaptive due to the preponderance of dark regions in natural images (Ratliff et al., 2010), which is also more pronounced at lower spatial frequencies (Cooper & Norcia, 2015). These asymmetries may influence human perception, since dark stimuli are processed faster and more reliably than light stimuli (Buchner & Baumgartner, 2007; Komban, Alonso & Zaidi, 2011).
This “dark-dominance” effect has been attributed to more OFF than ON excitatory inputs from the lateral geniculate nucleus (LGN) to V1 input layer 4 (Jin et al., 2008). However, this does not explain why more dark-dominant neurons are found in layers 2/3 than in layer 4 (Yeh et al., 2009). This discrepancy could be explained by stronger ON than OFF intracortical inhibition within V1 (Taylor et al., 2018). Hence, whether dark-dominance is mostly due to excitation to dark stimuli or inhibition to dark stimuli is still unclear. Here we develop a novel machine learning algorithm to disambiguate excitation from inhibition in extracellular recordings, which allows us to make quantitative inferences about how cortical neurons integrate ON and OFF inputs.

To better understand what drives V1 responses, we predict the responses of recorded neurons to natural images with a biologically-inspired convolutional neural network. This neural network processes the natural images' light (ON) and dark (OFF) information in separate pathways. The first layer of each pathway consists of a convolution with a parametrized 2D gaussian, which represents the responses of LGN neurons (omitting the weaker surrounds; Croner & Kaplan, 1995). The second layer is non-parametrized with weights representing the excitatory or inhibitory contribution to the model’s output. From these estimated weights, we can infer how much excitation and inhibition arises from each pathway, at every spatial location and temporal lag of a V1 simple cell’s receptive field.

Using this approach, we show dark-dominance in V1 neurons to be mainly driven by inhibition from the ON pathway rather than excitation from the OFF
pathway. This effect mostly occurs at the earliest measured time lag (0 to 13.3 ms), which explains why most neurons respond faster to dark than light. These findings suggest that faster inhibition to light than to dark plays a crucial role in the dark-dominance found in primary visual cortex.

**Methods**

*Animal preparation*

Anesthesia in adult cats was induced by isoflurane-oxygen (3–5%) inhalation, followed by intravenous (iv) cannulation and bolus iv injection of thiopental sodium (8 mg/kg) or propofol (5 mg/kg). Surgical anesthesia was maintained with supplemental doses of propofol or thiopental sodium. Atropine sulfate (0.05 mg/kg iv) or glycopyrrolate (30 μg im) and dexamethasone (0.2 mg/kg iv or 1.8 mg im) were administered and a tracheal cannula or intubation tube was inserted. Throughout the surgery, body temperature was thermostatically maintained and heart rate was monitored (Vet/Ox Plus 4700).

The animal was then positioned in a stereotaxic apparatus and connected to a respirator (Ugo Basile 6025). A craniotomy (P3/L1) over cortical A17 was performed, followed by a small durotomy. The cortical surface was protected with 2% agarose capped with petroleum jelly. Local injections of bupivacaine (0.50%) were administered at all surgical sites. Expired CO2, EEG, ECG, body temperature, blood oxygen, heart rate, and airway pressure were monitored and maintained at appropriate levels. During recording, the animal was anesthetized and paralyzed by the infusion of propofol (5.3 mg kg⁻¹ h⁻¹) and fentanyl (7.4 g kg⁻¹ h⁻¹) after a bolus injection (2.5 μg/kg), with a mixture of O2 and N2O (30:70
ratio) delivered through a ventilator. Paralysis was produced with a bolus iv injection of gallamine triethiodide (to effect), followed by infusion (10 mg·kg⁻¹·h⁻¹).

Corneas were initially protected with topical carboxymethylcellulose (1%) and subsequently with neutral contact lenses. Spectacle lenses were selected with slit retinoscopy to produce emmetropia at 57 cm, and artificial pupils (2.5 mm) were provided. Topical atropine sulfate (1%) and phenylephrine hydrochloride (2.5%) were administered daily, and topical altropine sulfate has been replaced by cyclopentolate in more recent experiments. Intramuscular glycopyrrolate (16 μg) and dexamethasone (1.8 mg) were also applied daily.

All animal procedures were approved by the McGill University Animal Care Committee and are in accordance with the guidelines of the Canadian Council on Animal Care.

Extracellular recording

Recordings were performed using multielectrodes (NeuroNexus), in most cases 32 channel (A1 32) linear arrays with each channel being 100 microns apart. Raw data signals were acquired with a Plexon Recorder (3 Hz to 8 kHz; sampling rate, 40 kHz). Signals recorded from a small photocell placed over one corner of the CRT were used for temporal registration of stimuli and spikes, and to verify the absence of dropped frames. Electrode penetrations were made approximately perpendicular to the brain surface, yielding similar RF locations for recording sites along the length of a linear-array multielectrode. Electrodes were advanced with a stepping motor microdrive (M. Walsh Electronics, uD-800A).
Spike waveforms were carefully classified from the recorded data to sort multichannel electrode data into single units, using Spikesorter (Swindale & Spacek, 2014). Only clearly sorted units were used for further analysis.

**Visual stimuli**

Visual stimuli are presented on a gamma-corrected CRT monitor (20 inches, 640x480 pixels, 150 Hz, 36 cd/m², NEC FP1350) at a viewing distance of 57 cm. Stimuli are presented with an Apple Macintosh computer (MacPro, 2.66 GHz, 6 GB, MacOSX ver. 10.6.8, NVIDIA GeForce GT 120 graphic card) using custom software written in MATLAB (ver. 2012b) with the Psychophysics Toolbox (ver. 3.0.10; Pelli, 1997; Brainard, 1997; Kleiner et al., 2007). We select a channel having MUA with good visual responses to hand-held bar stimuli. We then use this channel’s responses to select the dominant eye and close the non-dominant one. We also use this channel’s responses to position the CRT monitor to be approximately centered around the population receptive field.

Visual stimuli are natural images taken from the McGill Calibrated Colour Image Database (Olmos & Kingdom, 2004). The 2560x1920 images are cropped to 480x480, converted to monochrome with luminance values stored as 8-bit integers. We randomly present the images at a rate of 75 images per second (i.e. every 13.33 ms) in short movies of 5 seconds each. We have separated the stimuli into three sets, to evaluate predictive performance independently from overfitting. The training set has 20 movies which are presented 5 times each, while the validation and testing sets both have 5 movies which were presented
20 times each. The validation and testing sets are presented more often to provide a less noisy estimate of the fitted model's predictive performance.

**Model architecture**

To better understand differences between the ON and OFF pathways, we use a model architecture inspired from visual circuitry (Figure 11). The model predicts the mean recorded spiking response of the neuron from the natural images it was presented. To do so, it estimates the receptive fields of both ON and OFF lateral-geniculate inputs, and how these subunits excite or inhibit the V1 neurons we record from. We characterize LGN receptive fields as parametrized 2D gaussians with estimated parameters. The connections between LGN gaussians and the recorded neuron are estimated as free parameters.

Figure 11: Schematic representation of the machine learning algorithm used in this study.
The inputs to the model are the pixel luminance of the natural image stimuli, which are downsampled to 40x40. The mean of the inputs is centered at zero by subtracting the overall mean across all images. To obtain an estimate of the neurons’ spatio-temporal processes, the inputs are composed of the last 7 images, i.e. the time dynamics across the preceding 7 x 13.33 = 93.33 ms. The output is the responses of a neuron to grating stimuli, which were accumulated as PSTHs (bin width, 13.3 ms; duration of each frame).

The images are first fed to the ON pathway through a convolution with a positive parametrized 2D gaussian followed with a half-wave rectification (ReLU). The images are in parallel fed to the OFF pathway through a convolution with a negative parametrized 2D gaussian followed with a ReLU. The 2D gaussians represent receptive fields of LGN neurons in which the weaker surrounds (Croner & Kaplan, 1995) were omitted and are expressed as follow:

\[ g(h, v, p) = \frac{\alpha_p}{\sqrt{2\pi\sigma_p^2}} e^{-\left(\frac{h^2 + v^2}{\sigma_p^2}\right)} \]  

(7)

With h and v being the horizontal and vertical distances between a pixel and the center of the gaussian, respectively. \( \sigma \) represents the standard deviation (i.e. width) and \( \alpha \) the amplitude (i.e. the height) of the 2D gaussian. \( \sigma \) and \( \alpha \) are estimated separately for each pathway (p). For the ON pathway to process light information, \( \alpha \) is always positive (and negative in the OFF pathway). The convolution of the 2D gaussian with the inputs is followed with a ReLU, mimicking the half-wave rectified spiking responses of LGN neurons (Persi et al., 2011):
\[ c(i, j, t, p, k) = \max(0, \sum_{h=-6}^{6} \sum_{v=-6}^{6} x_{i+h,j+v,k,t} \ast g(v, h, t, p)) \]  

With \( i \) and \( j \) being the horizontal and vertical coordinates of the center of the 2D gaussian. \( x \) is the luminance of a specific downsampling pixel, \( t \) is the number of bins between the shown image and the recorded response, and \( k \) is the time bin of the neuron’s response. The convolution with the 2D gaussians has a kernel size of 12x12 pixels, is implemented with zero-padding and a "stride" of 1. Due to the rectification, the ON pathway processes information above the mean luminance, and the OFF pathway information below the mean luminance.

We then take a linear weighted sum of the convolution output from each pathway, with each weight representing the excitatory or inhibitory inputs from a LGN gaussian to the recorded neuron. This dense layer is followed by a rectified power law output non-linearity, which is the final output of the model and the prediction of the neuron’s mean spiking response:

\[ \hat{y}_{\text{pre}}(k) = \max(0, \sum_{p=1}^{2} \sum_{t=1}^{7} \sum_{i=1}^{40} \sum_{j=1}^{40} c(i, j, p, t, k) \ast w_{i,j,p,t}) \]  

\[ \hat{y}(k) = \hat{y}_{\text{pre}}^b(k) \]  

Where \( w \) represents the dense layer weights. \( \hat{y}(k) \) is the prediction of a neuron’s response for the \( k^{th} \) time bin. \( b \) is the estimated exponent of the rectified power nonlinearity and is estimated separately from the other weights (see below).
Optimization and regularization

To characterize the neuron’s receptive field, we find the model parameters which minimize the difference between recorded single neuron responses and the responses predicted by the model. There is a total of $2 \times 40 \times 40 \times 7 = 22,400$ dense layer weights, and 2 parameters for the 2D gaussians to estimate. Due to the high number of parameters, there is a risk for the model to exhibit overfitting. L2-regularization is a well-established approach to reduce overfitting by including a penalty for the squared amplitude of the weights (Hoel & Kennard, 1970). We implement it by minimizing the following loss function:

$$
\sum_{k=1}^{n} (y_k - \hat{y}_k)^2 + \lambda \sum_{p=1}^{\gamma} \sum_{t=1}^{\delta} \sum_{i=1}^{40} \sum_{j=1}^{40} \beta_{i,j,t,p}^2
$$

(11)

where $y_k$ is the neuron’s response and $\hat{y}_k$ the model’s predicted response to the $k^{th}$ time bin of the responses, $\beta$ the dense layer weights and $\lambda$ the L2-regularization parameter. The first term is the squared error of the model prediction and the second term the regularization penalty. $\lambda$ is set to $5 \times 10^{-5}$ in the first pass and $1 \times 10^{-5}$ in the second pass.

This loss function is minimized using the Adam optimization algorithm (Kingma & Ba, 2014) with mini-batch gradient descent (Li et al., 2014). To further reduce overfitting, we apply dropout during training to both the convolutional and final dense layers with a probability of 50% (Srivasta et al., 2014).

The data is separated into training, validation, and test sets, corresponding to the three sets of stimulus movies. The model parameters are fit
to the training set using a mini-batch size of 100 stimulus-response pairs. We evaluate the performance of the models across training epochs on a separate validation set, which allows us to evaluate the performance of the model independently from overfitting. We save a neuron’s model whenever it reaches its best validation performance and stop it from further training if the model has not reached its best validation performance for the last 50 epochs. The real performance of the model is evaluated on a third, separate test set to prevent any bias in measured predictive performance.

**Two-pass training procedure**

Because V1 receptive fields usually only occupy a small subset of the visual field, it would be detrimental to optimize each neuron’s models based on the full stimuli. Doing so would require us to either have a very high number of parameters or for the model to receive a low-resolution version of the image as its inputs. To solve this issue, we use a two-pass training procedure. In the first-pass, we optimize the models using the full 480x480 stimuli downsampled to 40x40. We then identify the receptive field of each neuron and manually designate a square cropping window region that encloses the receptive field. Next, we rescale the image within the cropping window we crop the image within the square and rescale the image within it to 40x40. This cropped image is then used to train the models in the second-pass. This procedure drastically increases the performance of the model and allows us to characterize a neuron’s receptive field with higher resolution.
Output non-linearity

As a neuron’s input increases, its firing rate can be modeled as a rectified power law (Persi et al., 2011). It has proven problematic to simultaneously estimate the power law exponent with the other parameters using backpropagation. This problem is most likely due in part to the exploding gradient problem (Pascanu, Mikolov & Bengio, 2012). To solve this issue, we initially set a power value of 1 and wait 50 epochs to get a rough estimate of the parameter values. We then bin the predicted responses into 100 bins of 75 responses each and compute the mean measured response for each bin. We then fit the exponent b to minimize the difference between the binned predicted response \( \hat{y} \) and the measured spike rate y. Optimization of the exponent b is done with the ‘optimize.curve_fit’ function from python’s scipy package, and is then fixed to the same value for the rest of the training epochs.

Estimating excitation and inhibition

The fitted model responses depend on both the dense layer weights and the estimated 2D gaussians. To incorporate both in our analysis, we convolve the 2D gaussians with the dense weights separately for the ON and OFF pathways. This convolution shows the “reconstructed” spatio-temporal receptive field of each pathway, which has a size of 40x40 and is used for further analyses. We estimate the overall amount of excitation and inhibition from each pathway and time lag by taking the sum of all positive or negative values in either the ON or OFF reconstructed receptive field. This procedure allows us to infer the total
amount of ON excitation, ON inhibition, OFF excitation and OFF inhibition in each neuron.

Orientation selectivity

We simulated measurement of orientation selectivity from sinewave gratings. We computed the fitted models' responses to a combination of sinewave gratings with 36 different orientations (with increments of 5 degrees), 56 different spatial frequencies (equidistant from 0.0667 to 0.143 cycles per image) and 36 different phases (increments of 5 degrees). These responses were used to compute the orientation selectivity of each neuron using a vector summation method (Bonhoeffer et al. 1995):

\[ L = \frac{(a^2 + b^2)^{1/2}}{\sum_{i=0}^{N-1} R(x_i)} \]  

(12)

Where \[ a = \sum_{i=0}^{N-1} R(x_i) \cos (2x_i) \]  

(13)

and \[ b = \sum_{i=0}^{N-1} R(x_i) \sin (2x_i) \]  

(14)

N represent the number of sinewave gratings \( x_i \), while \( R(x_i) \) represents the simulated responses to such gratings. The orientation selectivity \( L \) was computed separately for each time lag in every neuron.
Results

Example neurons

To estimate how a neuron is excited and inhibited by light and dark, we optimize the convolutional neural network model parameters to best predict the responses of a single neuron. The estimated parameters of two example neurons can be seen in Figure 12, with neuron #1 on the left (Figure 12A - D) and neuron #2 on the right (Figure 12E - H). Figure 12A shows the estimated 2D gaussian for the ON pathway, for neuron #1. Figure 12B shows the dense layers for the ON pathway, where red represents excitatory inputs and blue inhibitory inputs. Note that the ON pathway of neuron #1 responds most strongly at the 26.67 ms time lag. At this latency, this neuron has a receptive field horizontally gabor-like with a dark-driven center and light-driven surround. However, at later time lags the receptive field of this neuron reverses in sign: The center becomes light-driven with a slight dark-driven surround. The OFF pathway (Figures 12C – D) shows similar behavior, with the difference that both the 2D gaussian and dense weights are of the opposite sign. At the 26.67 ms time lag, the dark and light regions are similar in magnitude; however, this is not the case for all neurons. Neuron #2 (Figures 12E - H) is most responsive at the 13.33 ms lag. The ON pathway has more inhibition than excitation, and the OFF pathway has more excitation than inhibition. Hence, at the 13.33 ms time lag this neuron is unbalanced and responds more to dark than to light stimuli.
Figure 12: Estimated spatiotemporal filters and weight maps of two example neurons. (A) to (D) represent about the first example neuron, while (E) to (F) represent the second example neuron. (A) shows the estimated ON gaussian filter for the first neuron, (B) shows the estimated dense weights (excitatory in red and inhibitory in blue) for the ON pathway, (C) shows the estimated OFF gaussian filter and (D) the estimated dense weights (excitatory in red and inhibitory in blue) for the OFF pathway. E, F, G, and H show the same information as A, B, C, D but for the second example neuron.

To assess the strength of dark-dominance, we take the sum of all positive (red) and negative (blue) weights in each pathway as a measure of this neuron’s excitation and inhibition (see Methods). At the 13.33 ms lag, neuron #2 has its dark responses (OFF excitation and ON inhibition) being 56.7% greater than light responses (ON excitation and OFF inhibition). Dark-dominance in neuron #2 is especially prominent in the ON pathway, where ON inhibition is 77% stronger than ON excitation. The dark-dominance effect is also present but weaker in the OFF pathway, where OFF excitation is only 22% stronger than inhibition. Hence, the dark-dominance of this neuron is more strongly driven by ON inhibition than
by OFF excitation. This example raises the question of whether dark-dominance across the population usually arises from relatively more ON inhibition.

*Population responses*

To investigate whether dark-dominance is more due to ON inhibition or OFF excitation across the population of V1 neurons, we computed the sum of the four types of inputs for each neurons’ optimal time lag (see Methods). All neurons had an optimal time lag of either 13.33 ms (35 neurons) or 26.67 ms (34 neurons). We classified neurons as dark-dominated or light-dominated depending on whether a neuron was more responsive to dark (OFF excitation and ON inhibition) or light (ON excitation and OFF inhibition) at its optimal time lag. Across our population of 69 neurons, we found 48 neurons (69.57%) to be dark-dominated and 21 neurons (30.43%) to be light-dominated, similar to Yeh et al. (2009). The amount of excitation and inhibition from both pathways for all neurons can be seen in Figure 13A. Both types of dark inputs (OFF excitation and ON inhibition) are stronger than their light counterparts (ON excitation and OFF inhibition). OFF excitation is on average 30.7% stronger than ON excitation (Figure 13B; paired t-test, \( t = 3.77, df = 68, p = 0.00034 \)), consistent with Jin et al. (2008). ON inhibition also plays an important role in the dark-dominance effect by being on average 32.85% stronger than OFF inhibition (Figure 13C; \( t = 4.02, df = 68, p = 0.00014 \)). ON inhibition is on average slightly (13.70%) stronger than OFF excitation (Figure 13D; \( t = 2.69, df = 68, p\text{-value} = 0.0089 \)), which suggests ON inhibition contributes slightly more to the dark-dominance effect than OFF
excitation at the optimal time lag. These findings may help explain the laminar
distribution reported in primate V1 (Yeh et al., 2009; see discussion).

![Figure 13](image1.png)

Figure 13: Strength of excitation and inhibition from the ON and OFF pathways at each neuron’s optimal time lag. (A) shows the strength of the four types of inputs, where each neuron is a different point. Orange points are dark-dominated neurons and blue points are light-dominated neurons. (B) shows the relationship between ON excitation and OFF excitation for each neuron, (C) shows the relationship between OFF inhibition and ON inhibition and (D) the relationship between OFF excitation and ON inhibition.

**Light/dark balance**

To better understand how much individual neurons are light or dark dominated, we designed a light-dark balance index (LDB) to indicate how balanced a neuron is in its light and dark weights:
LDB = (Light – dark) / (Light + dark) \hspace{1cm} (15)

where

Light = ON excitation + OFF inhibition \hspace{1cm} (16)

Dark = OFF excitation + ON inhibition \hspace{1cm} (17)

This index varies from -1 to 1. Positive LDB values mean a neuron is light-dominated, and negative values mean the neuron is dark-dominated. A balanced neuron would have an LDB of approximately 0. The histogram of each neuron’s LDB at their optimal time lag can be seen in Figure 14A. The peak of the histogram is slightly below 0, showing most neurons to be dark-dominated.

Figure 14: Light-dark balance index across the population of neurons for each latency. (A) shows the histogram of light-dark balance index values at optimal latency for each neuron. (B) does the same but at the 0 ms latency, (C) at 13.33 ms and (D) at 26.67.
Figure (E) shows the distribution of light-dark balance at all latencies for each neuron, with the gray bars representing the average LDB at a specific latency.

*Time dynamics*

The above results used each neuron’s optimal time lag, which was usually either 13.33 or 26.67 ms (Figure 14), consistent with Komban et al. (2014). Note that since our neural data is binned into segments of 13.33 ms, these latency bins includes spikes up until 13.33 ms later. For example, the 0 ms latency bin includes spikes up until 13.33 ms, while the 13.33 ms latency bin includes spikes up until 26.67 ms. Because V1 neurons respond on average 3.4 ms faster to dark than to light stimuli (Komban et al., 2014), we suspected dark-dominance might vary as a function of time. To study whether dark-dominance varies as a function of time, we separately compute the LDB of each latency for every neurons. As shown in Figure 14C, the dark-dominance effect is especially predominant at the 13.33 ms latency ($t = -3.742$, $df = 68$, $p$-value = 0.00038), where 72.46% of neurons are dark-dominated. Dark-dominance is weaker but also present at the 0 ms latency (Figure 14B; $t = -5.63$, $df = 68$, $p = 4.88 \times 10^{-7}$). The 26.67 ms latency (Figure 14D) is on average slightly light-dominated ($t = 2.00$, $df = 68$, $p$-value = 0.0497), but this effect is small. While LDB values are skewed towards dark-dominance at the 0 and 13.33 ms latencies (Figures 14B and 14C), the 26.67 ms latency is mostly balanced (Figure 14D). These findings suggest that while V1 neurons are biased towards dark responses in their short latencies, longer latencies are relatively balanced between light and dark responses.
To understand why neurons are dark-dominated in their early latencies, we investigate how the strength of each input type varies as a function of time. As can be seen in Figure 15A and 15B, responses are weaker at the 13.33 than at the 26.67 ms time lag for both ON excitation (t = 4.22, df = 68, p = 7.48 x 10^{-5}) and OFF inhibition (t = 6.56, df = 68, p = 8.54 x 10^{-9}). This delay does not happen for OFF excitation (t = 0.14, df = 68, p = 0.89) and ON inhibition (t = 0.64, df = 68, p = 0.53) who have similar strength at these two latencies (Figure 15C and 15D). Therefore, dark-dominance is stronger at the 13.33 ms lag because both types of light inputs are delayed while both types of dark inputs do not.

Figure 15: Input strength over latencies. (A), (B), (C) and (D) show how the strength ON excitation, OFF excitation, ON inhibition and OFF inhibition change over latencies, respectively. (E), (F) and (G) show the relative strength of each type of input at latencies of 0, 13.33 and 26.67 ms, respectively.
To better understand how dark-dominance occurs at a latency of 13.33 ms, we looked at the strength of each type of input across all neurons. The latency of 13.33 ms (Figure 15 F) exhibits dark-dominance with ON inhibition being on average 53% stronger than the other types of inputs. ON inhibition is significantly stronger than OFF inhibition \((t = 4.8, \text{ df} = 68, p = 8.85 \times 10^{-6})\), ON excitation \((t = 4.6, \text{ df} = 68, p = 1.47 \times 10^{-5})\) and OFF excitation \((t = 3.7, \text{ df} = 68, p = 0.0004)\). On average, ON inhibition is 53.6% stronger than OFF inhibition, while OFF excitation is only 32% stronger than ON excitation. At the 0 ms time lag, the dark-dominance instead is mostly due to OFF excitation, which is significantly stronger than both ON excitation \((t = 4.44, \text{ df} = 68, p = 3.91 \times 10^{-5})\), OFF inhibition \((t = 9.04, \text{ df} = 68, p = 7.37 \times 10^{-13})\), and ON inhibition \((t = 4.97, \text{ df} = 68, p = 5.84 \times 10^{-6})\). However, at longer latencies, the dark-dominance effect disappears. At a latency of 26.67 ms, neurons are slightly light-dominated (Figure 4B) with inhibitory inputs being stronger than excitatory inputs in the OFF pathway \((t = 3.86, \text{ df} = 68, p = 0.00025)\), but this difference is non-significant in the ON pathway \((t = 1.8, \text{ df} = 68, p = 0.076)\). ON and OFF inhibition are not significantly different from each other at this time lag \((t = 1.7, \text{ df} = 68, p = 0.087)\). As early as the 26.67 ms lag, OFF inhibition becomes as strong as ON inhibition, which in turn dampens the dark-dominance effect. These results suggest the dark-dominance to only happen at latencies of 0 and 13.33 ms.
Orientation selectivity

So far, we have found the dark-dominance phenomenon to only happen in early response latencies. However, it is still unknown how these findings relate to the spatial properties of the receptive fields such as orientation selectivity. To infer these properties on our sample of neurons, we simulated a gratings experiment and extracted the responses of the neurons’ fitted models to a variety of gratings stimuli. Static gratings were briefly presented at a single time lag at a time and varied in their orientation, spatial frequency, and phase. We then selected the sinewave grating the neuron was most responsive to at a specific time lag, and inferred properties of that neuron’s time lag based on the best sinewave grating. Orientation selectivity was computed using a vector summation method (Bonhoeffer et al. 1995; see Methods). Orientation selectivity for each neuron across time lags can be seen in Figure 16. We find orientation selectivity to peak at a latency of 26.67 ms, which coincidentally happens to be the latency at which neurons are the most balanced in their light and dark responses (Figure 14). This correlation suggests the dark-dominance effect might possibly reduce orientation selectivity in the early response latencies of a subset of neurons.
Discussion

An interesting characteristic of the primary visual cortex is that most of its neurons respond more strongly to dark than to light stimuli. However, the mechanisms and properties of this dark-dominance effect are still poorly understood. Here we use a machine learning approach combined with natural image stimuli to infer the excitatory and inhibitory inputs to V1 neurons. Using this technique, we show dark-dominance to be driven more strongly by ON inhibition than OFF excitation. We also show the dark-dominance effect to only occur in the early time lags of neurons’ responses, with neurons becoming on
average balanced starting from the 26.67 ms time lag. Orientation selectivity is also lower at shorter latencies, possibly due to the dark-dominance effect.

*Inferring neural circuitry from machine learning*

A novel aspect of this study is how we use machine learning to infer the inputs to a recorded V1 neuron. A strength of the model architecture is how it is based on known neurobiology. Like the LGN, the model is separated into ON and OFF pathways, each composed of non-oriented receptive fields which act somewhat like Gaussian linear filters followed by half-wave rectifiers ($$ Mention how I neglect the surrounds). This model architecture makes our model better suited for system identification of V1 neurons.

More specifically, separating the inputs into two different pathways allows the machine learning algorithm to infer how neurons increase and decrease their firing rate in response to light and dark stimuli. In this thesis, we refer to this increase and decrease as excitation and inhibition. However, what we estimate does not necessarily reflect the synaptic excitatory or inhibitory inputs a neuron directly receives. For example, a neuron could decrease its firing rate in response to light because its excitatory inputs are inhibited by light. In general, the excitation and inhibition we estimate might reflect circuitry more complex than direct synaptic inputs. We should instead interpret the average strength of each input for a neuron as reflecting what is happening at the level of both the neuron and its inputs. Inhibition is stronger to light than dark, but it is possible for this inhibition to occur at the level of the inputs to the recorded neuron rather than directly to the neuron itself.
Another strength of the model architecture is its simplicity. The model architecture is like, but different from, a linear model followed by a half-wave power-law output nonlinearity. The first difference is the separation of the model into two different pathways that separately process light and dark information. This separation is a core aspect of this study because it is what allows us to disambiguate excitation from inhibition. The second difference is the 2D gaussian filters. If the model architecture were more complex, the results we obtain from the analysis might be more dependent on the sort of model we use. Since the model we use is close to a linear model, we are confident our findings are not an artifact of our choice of the model architecture.

Dark-dominance caused by inhibition to light

Dark-dominance in V1 is usually thought to originate from greater lateral geniculate excitation from the OFF than from the ON pathway (Jin et al., 2008). However, recent findings suggest dark-dominance might also be caused by stronger intracortical inhibition to light than dark stimuli (Taylor et al., 2018). Our results support both hypotheses and show ON inhibition to be much (53.6%) stronger than OFF inhibition, while OFF excitation is only slightly (32%) stronger than ON excitation. These results suggest the dark-dominance effect to be mostly driven by intracortical inhibition to light stimuli, differently from what was originally proposed by Jin et al. (2008).

These findings help explain why dark-dominance is strongest in layer 2/3 of primate V1 (Yeh et al., 2009). If dark-dominance were almost only due to stronger lateral geniculate excitation from the OFF pathway, we would expect
dark-dominance to be at least just as strong in layer IV than in the other layers, since this is where most of LGN neurons synapse. While layer IV does show dark-dominance with two-thirds of its neurons being dark-dominant, this effect is much stronger in layers 2/3 where almost every neuron is dark-dominant (Yeh et al., 2009). This laminar difference might be due to pyramidal neurons in layers 2/3 receiving extensive inhibition as has been shown in the mouse (Kätzel et al., 2011), with inhibition being stronger to light than dark stimuli (Taylor et al., 2018). Parvalbumin-positive interneurons (PVs) could be behind this transient inhibition to light since these neurons are fast-spiking (Hu et al., 2014), directly target cell bodies and are the most common interneurons in sensory cortex (Wood et al., 2017).

*Time dynamics of dark-dominance*

Our results show the dark-dominance effect to only occur at the shorter latencies. This is consistent with previous research showing faster responses by LGN neurons in the OFF than in the ON pathway (Jin et al., 2011), and dark-dominant V1 neurons having faster responses (Komban et al., 2014). Dark responses in V1 are thought both stronger and faster than light responses (Yeh et al., 2009; Komban et al., 2014; Mazade et al., 2019). However, we only observe the dark-dominance effect in the early but not in the later time lags. These results suggest dark-dominance might occur because of the faster responses to dark stimuli.
Since V1 neurons can receive excitation directly from LGN afferents, while inhibition has to first go through a cortical inhibitory interneuron (Eccles, 1976; Montero, 1986), it is somewhat surprising for V1 dark-dominance to be predominantly due to inhibition. A possible mechanism for this strong early ON inhibition might be the strong transient response from inhibitory interneurons in layers 2/3 of V1 (Tucker & Fitzpatrick, 2006). This transient inhibition might reduce the firing rate of downstream neurons in their early time lags. Because inhibition of cortical neurons is stronger to light than dark stimuli (Taylor et al., 2018), this transient inhibition might in turn leads to dark-dominance in the early response latencies.

**Relationship with orientation selectivity**

This study also brings a new perspective on the intracortical mechanisms of orientation selectivity. A surprising characteristic of orientation selectivity is that V1 neurons are less orientation-selective in their early time lags (Shapley et al., 2003). A suggestion explanation for this phenomenon is excitation from the LGN being faster than intracortical inhibition (Shapley et al., 2003). While it is likely for slower inhibition to contribute to low orientation selectivity in the early time lags, we find that neurons still lack orientation selectivity at the 13.33 ms time lag when inhibition and excitation are roughly balanced but dark-dominance is still present. Neurons tend to reach their peak orientation selectivity at the 26.67 ms time lag, which is when the dark-dominance effect disappears. It is interesting how low orientation selectivity and the dark-dominance effect have
similar latencies. This correlation suggests there might potentially be a causal relationship between the two, where stronger dark responses might reduce orientation selectivity. These findings suggest more work might be required to better understand the relationship between dark-dominance and orientation selectivity.

In conclusion, this research brings a new perspective on the dark-dominance phenomenon in the primary visual cortex. First, we find dark-dominance to be more strongly driven by inhibition to light than by excitation to dark. Second, we show the dark-dominance effect to only occur in the early time lags. Finally, we also show how dark-dominance might be related to the lower orientation selectivity of V1 neurons in their early time lags.

Now that I’ve presented my research, I will discuss about more general concepts and ideas related to my findings.

Chapter III: General discussion

This thesis uses a machine learning approach to better understand the responses of extracellularly recorded V1 neurons to natural images. We estimate the spatio-temporal excitation and inhibition from the ON and OFF pathways for each neuron, which allows us to replicate previous findings and to better understand the asymmetries between responses to light and dark stimuli. Using this method, we replicate stronger OFF excitation than ON excitation (Jin et al., 2008), as well as stronger average responses to dark than light stimuli (Yeh et
al., 2009). We find this “dark-dominance” effect to be especially strong in the early time lags, consistent with dark-dominant neurons having faster responses than light-dominant neurons (Komban et al., 2014). We also replicate broad orientation-selectivity of V1 neurons in the early time lags, that sharpens in later time lags (Shapley et al., 2003).

**Model architecture and V1 circuitry**

This research offers a new perspective on how we can fit models to predict a neuron’s response. The machine learning algorithm we used was heavily inspired by known neurobiology, such as the split between ON and OFF pathways as well as lateral geniculate nucleus (LGN) afferents being modeled as 2D gaussians (from which we omit the weaker surrounds). Basing our model on known neurobiology rather than using a deep neural network architecture allows us to answer questions about how neurons work more efficiently. For example, one of the strengths of this analysis is that we can estimate how V1 neurons are excited and inhibited by different inputs.

While the neural circuitry within V1 is not yet fully understood, we know it to be different and much more complicated than the architecture of our model. The model assumes the inputs are ON and OFF 2D Gaussians, which are meant to represent the receptive fields of LGN neurons (neglecting their weaker surrounds). A limitation of the current study is that we cannot directly relate the estimated weights to the synaptic inputs received by the cortical neurons we
record. For example, it is well-known that V1 neurons receive not only excitation from the LGN, but also intracortical excitation from other cortical neurons, especially within the same column where neurons tend to share many properties such as orientation preference and ocular dominance. Because of this intracortical excitation, it is likely for the excitation we estimate via our model-fitting to reflect excitation not only from the LGN but also from other V1 neurons. Our model is also unlikely to reflect how V1 neurons receive inhibitory inputs, since V1 neurons do not receive direct synaptic inhibition from the LGN (Montero, 1986). Instead, V1 neurons receive inhibitory inputs from intracortical inhibitory interneurons (Eccles, 1976), which themselves receive most of their excitatory inputs from either the LGN or other V1 neurons depending on whether they are PV+ or SST+ interneurons (Versendaal & Levelt, 2016). These inhibitory inputs may synapse on the neuron we record from but might also synapse on the inputs of the recorded neuron. Unfortunately, our approach is unable to distinguish between these two possibilities. We cannot tell whether a neuron is directly inhibited by an inhibitory interneuron or whether it is the inputs to the neuron that are inhibited. More generally, the circuitry of the primary visual cortex is much more complicated than our model architecture and we must not assume our results to reflect direct excitatory and inhibitory inputs to the recorded neurons.
Measuring the dark-dominance effect

To understand what our results represent, it is important to understand what is learned by the machine learning model. The model optimizes its weights to best predict the recorded neuron’s firing rate in response to natural image stimuli. Positive weights in the ON and OFF pathways indicate the neuron increases its firing rate in response to light or dark, respectively. Negative weights in either pathway suggest the neuron decreases its firing rate in response to light or dark stimuli. Since V1 neurons usually have a very low baseline firing rate, they do not usually fire less than baseline when inhibited. Instead, inhibition seems to play the role of canceling out excitation. For example, imagine a neuron fires when it is shown a high amount of light in its light-driven region. If this is the case, simultaneously showing light in its dark-driven region might prevent it from firing. This is the type of interaction these inhibitory weights we estimate represent. For a V1 neuron to reduce its firing rate to light or dark, there needs to be some sort of inhibition somewhere within V1. The inhibition might not directly synapse onto our recorded neuron but needs to synapse onto at least one of its inputs. Hence, while our estimated models might not represent the synaptic inputs a neuron directly receives, our results indirectly represent the overall amount of excitation and inhibition to light and dark a neuron and its inputs receive.

Natural image stimuli play a key role in the model’s ability to distinguish between excitation from one pathway and inhibition from the other pathway. Natural images are rich in information and similar to the visual stimuli animals
see in everyday life. This richness allows neurons to simultaneously receive visual stimuli that both increase and decrease their firing rate in different parts of their receptive fields. Because neurons are simultaneously presented with an array of luminance intensities, the machine learning algorithm can distinguish between excitation from one pathway and inhibition from the other. Dense white noise stimuli combined with our machine learning approach might also allow us to make the distinction between excitation and inhibition. However, dense white noise performs poorly for system identification (Talebi & Baker, 2012). Sparse noise on a gray background (Yeh et al., 2009) is well-suited to measure the dark-dominance effect but cannot distinguish excitation from inhibition, since it does not show black and white stimuli simultaneously. However, simultaneously showing both white and black checks on a gray background might fix this issue.

Another advantage of natural images is that they are scale-invariant and can map receptive fields of different sizes. Other types of stimuli like sparse noise on a gray background require to approximately match the stimuli to the receptive field size. Sampled neurons in multi-neuron recordings might vary quite a lot in their receptive field sizes, which makes natural images especially suited for multi-neuron recordings.

The dark-dominance effect in V1 can be measured using a variety of different stimuli. Because these stimuli are so different from each other, it is important to consider how the dark-dominance effect we found with natural images and machine learning relates to the dark-dominance found by other researchers. Jin et al. (2008) used white noise and reverse correlation analysis to
estimate the receptive fields of V1 neurons, and the first researchers to report the
dark-dominance effect. Reverse correlation analysis with white noise stimuli
assumes the neuron to have linear responses with a push-pull organization of
rectified inputs, in which excitation and inhibition are equal. This differs from our
method which does not assume equal excitation and inhibition. Because of this
linearity assumption, reverse correlation works best on simple but not on
complex cells. Therefore, the dark-dominance effect they found might have been
based on simple cells. Unfortunately, Jin et al. (2008) only showed the dark-
dominance effect on a subset of neurons. Even though they claimed that neurons
responded more to dark than light stimuli across the population, they did not
report population results.

Yeh et al. (2009) did a more objective and thorough analysis of the dark-
dominance effect. They used sparse noise with both light and dark noise
elements on a gray background, which allowed them to separately evaluate
responses to light and dark stimuli. This method does not make the strong linear
assumption present in reverse correlation, which might have allowed their
analysis to include both simple and complex cells. They classified neurons as
either light-dominant or dark-dominant based on whether they responded more
strongly to white checks on a gray background or black checks on a gray
background. They found almost all neurons to be dark-dominant in layers 2/3.
They also found about 66% of neurons to be dark-dominant in layers 4, 5, and 6,
which is similar to our results. Since V1 neurons have low baseline firing rates
and only one type of check (light or dark) was shown at a time, sparse noise on a
gray background might probe excitation more than inhibition. However, because
the checks were relatively large compared to the receptive field sizes, it is likely
for the checks to not perfectly match the receptive field and for inhibition to have
influenced the results as well. Since layers 2/3 have a high density of inhibitory
interneurons that are interconnected (Tremblay, Lee & Rudy, 2017), strong
inhibition to light (Taylor et al., 2018) might be a plausible explanation for their
results. In conclusion, it seems likely for the dark-dominance effect found from
sparse noise on a gray background by Yeh et al. (2009) to be due to a
combination of stronger excitation to dark and inhibition to light.

Sinewave gratings (Jansen et al., 2019) have also been used to measure
the dark-dominance effect. Sinewave gratings can be shown at a variety of
orientations and spatial frequencies sampled from a Hartley function (Ringach et
al., 1997) to do receptive field mapping and estimate the dark-dominance effect.
Since light parts of the stimulus are often presented in dark-driven regions and
vice-versa for light-driven regions, sinewave gratings at non-optimal orientations
most likely recruit both excitation and inhibition. This method implies the neuron’s
responses are linear, which suggests the sampled neurons with clear receptive
fields are most likely simple cells. Half-wave rectified sinewave gratings
(Kremkov et al., 2014) do not assume the neuron’s response to be linear and
have also been used to measure dark-dominance. When shown at the optimal
orientation, spatial frequency and phase, both sinewave gratings and their half-
wave rectified version fail to show dark stimuli in light-driven regions and light
stimuli in dark-driven regions. It is therefore likely for these stimuli to recruit
excitation but not much inhibition, and for the dark-dominance effect estimated from these stimuli to be somewhat different from the dark-dominance effect we estimate.

In summary, I think the dark-dominance effect found by other researchers using different stimuli most likely recruits a combination of excitation and inhibition. The only exception is when sinewave or half-wave rectified sinewave gratings are at the optimal orientation, spatial frequency and phase. Another difference might be how we measure dark responses by adding OFF excitation and ON inhibition. Excitation and inhibition may interact in a more complicated manner when measuring the dark-dominance effect using other stimuli.

*Time dynamics*

An interesting finding of this study is how the dark-dominance effect is only present in the early time lags of V1 neurons. In later time lags the dark-dominance effect disappears and the population of neurons responds, on average, as strongly to dark as to light stimuli. To our knowledge, we are the first study to report such a discovery. Previous researchers only considered each neuron’s optimal time lag when reporting population effects. As we show in our results, we find a dark-dominance effect at the optimal time lag only because many neurons have their optimal time lag at 13.33 ms, and this is a latency at which the dark-dominance effect occurs. We do not find a dark-dominance effect at latencies longer than 13.33 ms. These results are consistent with Komban et al. (2014) who found the latency of dark-dominated neurons to be shorter than that of light-dominated neurons. These findings suggest it might be better to
analyze each time lag separately when looking at the dark-dominance effect. It might also be worthwhile to analyze each time lag separately when investigating other phenomenon within V1, such as orientation selectivity.

*Future directions*

In the following section, I will address future directions that could be taken to extend this study and gain a deeper understanding of the dark-dominance effect. I will start by discussing potential experiments that could help validate the current results and relate them to the literature. I will then propose new analysis methods that could allow us to better understand the time dynamics of the dark-dominance effect.

Previously I have discussed how our dark-dominance findings might relate to the rest of the literature. More specifically, it might be worthwhile to check whether our new method gives the same results as methods used in previous studies. To do so, we could do a new experiment where we record responses of neurons to both natural images and white noise (Jin et al., 2008), sparse noise on a gray background (Yeh et al., 2009) or sinewave gratings (Komban et al., 2014). We could then test whether the neurons that are dark-dominant according to the natural image analysis are also found to be dark-dominant when using other visual stimuli.

Another question that would be interesting to answer is whether the neurons in our sample are simple or complex. The main difference between our
model architecture and a linear system is how we separate light and dark stimuli into two distinct pathways. Our model architecture is relatively close to being linear and might not be suited to estimate complex cells. However, it might be possible for simple and complex cells to be on a continuum rather than being two distinct categories (Mechler & Ringach, 2002), and for our method to capture the linear aspect of neurons that also have complex responses. To investigate whether this is the case, we could use drifting sinewave gratings and compute each neuron’s F1/F0 ratio to classify them as simple or complex cells (Skottun et al., 1991). We could then investigate whether the dark-dominance effect is similar across these two types of neurons.

An improvement that could be made to this study would be to increase the time resolution, which would allow us to gain a better understanding of how the dark-dominance effect evolves across latencies. The time resolution of 13.33 ms in our analysis is not limited by the sampling rate of the multi-electrode. The spike times are collected into bins of 13.33 ms because this happens to be the duration of each natural image frame. We could instead use time bins that are two or three times smaller, and a proportionally higher frame rate for the stimuli, to substantially increase the temporal resolution of the analysis. This increase in resolution would come with the cost of producing more noisy results. Even so, the tradeoff might be worthwhile, since increasing the temporal resolution would allow us to learn more about the time dynamics of OFF excitation and ON inhibition.
Concluding remarks

In conclusion, this thesis brings a new perspective on the dark-dominance effect. We used a machine learning approach combined with natural images to infer the excitation and inhibition to light and dark stimuli of the primary visual cortex. First, we first found stronger dark than light responses to be more strongly due to inhibition from light stimuli than to excitation from dark stimuli. Secondly, we showed the dark-dominance effect to only occur at the early latencies of V1 responses. Lastly, our findings suggest dark-dominance might be related to lower orientation selectivity in the early latencies. This research brings us one step closer to understanding the intricate information processing achieved by the primary visual cortex.
References


