

Nonlinear subunit models of neuronal receptive fields in the early visual pathway

Amol Gharat

Integrated Program in Neuroscience

McGill University, Montreal

August 2017

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

© Amol Gharat, 2017

DEDICATION

To Arya and Shadow.

ACKNOWLEDGEMENTS

This thesis would not have been possible without constant support and encouragement from Dr. Curtis Baker throughout my graduate school. I am thankful to Dr. Baker for providing me freedom to pursue my interests and opportunity to conduct research in his laboratory.

I would like to thank all my colleagues from Dr. Baker's lab for their support. Specially, I'd like to thank Guangxing Li for being always ready to help with any technical problem in the lab and for all his help with experiments and software. I'd like to thank Vargha Talebi for his software on system identification and assistance with the recording experiments. I am grateful to Philippe Nguyen for introducing me to convNets and providing the software for training them. I am thankful to the members of McGill Vision Research Unit for developing a very friendly and stimulating environment as well as providing useful feedback along the way.

My advisory committee members Dr. Amir Shmuel and Dr. Chris Pack provided me with necessary guidance and made sure I was on the right track. I would also like to thank former director of the IPN program Dr. Josephine Nalbantoglu for going out of her way and providing incredible help with preparing fellowship applications.

I am thankful to my parents and my family back in India for their constant support and encouragement. I am very grateful to my wife Pooja, for leaving behind her family and job in India and moving to Canada. This long journey would not have been possible without her unconditional love and support. I would like to thank all my friends in Montreal who made this journey enjoyable.

I am grateful for the financial assistance provided by Vanier Canada Graduate Scholarship from the Canadian Institutes of Health and Research (CIHR) and research grants to Dr. Curtis Baker from CIHR. I am also thankful to NVidia Corporation for the donation of a Tesla K40 GPU card through a NVidia Hardware Grant.

PREFACE

Contribution of the authors

This thesis is written in manuscript form as permitted by the McGill University. It is composed of five chapters: an introduction, one published manuscript (Gharat & Baker, 2017), two manuscripts to be submitted and a final discussion.

Chapter 2 contains a published manuscript: Gharat A and Baker CL (2017) Nonlinear Y-like receptive fields in the early visual cortex: An intermediate stage for building cue-invariant receptive fields from subcortical Y cells. The Journal of Neuroscience. 37(4): 998-1013. The study was designed by me and Dr. Baker. Recording experiments were performed by me with assistance from Dr. Baker and lab members, Guangxing Li and Vargha Talebi. Guangxing Li provided analysis software for analyzing plexon data files. Data analysis and model simulations were performed by me with guidance from Dr. Baker, using modeling software written by him and I. Finally, this manuscript was written by me and edited by Dr. Baker.

Chapter 3 contains a manuscript in preparation: Gharat A, Nguyen P and Baker CL. Estimating subunit receptive field models of thalamic neurons with deep learning. The study was designed by me and Dr. Baker. Recording experiments were performed by me with assistance from Dr. Baker and lab members, Guangxing Li and Philippe Nguyen. Philippe Nguyen provided software for estimation of convolutional models, which I adapted for use with my data, and helped with initial data analysis. I performed data analysis with guidance from Dr. Baker. Finally, this manuscript was written by me and edited by Dr. Baker.

Chapter 4 contains a manuscript in preparation: Gharat A and Baker CL. Separating ON and OFF pathway inputs to cortical simple cells reveal receptive fields with asymmetric push-pull. The study was designed by me and Dr. Baker. Recording experiments were performed by me with assistance from Dr. Baker and lab members, Guangxing Li and Philippe Nguyen. System identification software provided by Vargha Talebi was modified by me to estimate subunit receptive fields. I performed data analysis and wrote the manuscript with guidance from Dr. Baker.

Contributions to Original Knowledge

In Chapter 2 we found nonlinear Y-like receptive fields in the early visual cortex that could form an intermediate stage between subcortical Y cells and cortical orientation selective cue-invariant receptive fields. We proposed a novel neural network model that generates cue-invariant selectivity by combining Y pathway ON and OFF inputs in asymmetric manner. Through model simulations we demonstrated that its tuning properties to luminance and contrast-modulation gratings match the previously observed selectivity of cortical neurons.

In Chapter 3 we adapted a deep learning framework to quantitatively estimate receptive field models of LGN neurons based on known retinal circuitry. To my knowledge, it is the first study to show that for LGN Y cells, a multi-stage neural network model improves their predictive performance compared to a linear-nonlinear model on novel naturalistic stimuli. With this method we are able to infer computations happening at earlier stages than the LGN, since we could recover biologically interpretable subunit filters and their output nonlinearity. We showed that convolutional neural networks are powerful in modeling early stages of the visual pathway with limited neural data.

In Chapter 4 we devised a novel method to quantitatively measure both excitatory and inhibitory contributions of ON and OFF pathway inputs within individual cortical receptive fields, using naturalistic visual stimuli, from extracellularly recorded spiking signals. We recovered novel spatio-temporal asymmetries in the integration of ON and OFF inputs in cortical simple cells. These results challenge the standard model of a simple cell receptive field as a linear spatio-temporal filter. These asymmetries found in the inputs to cortical neurons could provide the neural mechanism for generating cue-invariant receptive fields from Y-pathway inputs.

In Summary, this thesis provides a new perspective on how spatial nonlinearities emerging from the retina can influence cortical processing of luminance- and texture-defined boundaries. These results strengthen the idea that luminance and texture signals are multiplexed through the early stages of the visual processing.

TABLE OF CONTENTS

DEDICATION	i
ACKNOWLEDGEMENTS	ii
PREFACE	iv
LIST OF FIGURES	xi
ABBREVIATIONS AND SYMBOLS	xiii
CHAPTER 1	1
1.1 Thesis Structure	2
1.2 First- and second-order information in visual scenes	2
1.3 Early visual pathways	4
1.4 Retinal Circuitry.....	6
1.5 Visual signal transmission from retina to cortex	8
1.6 Receptive field properties of cortical neurons	10
1.7 Neural mechanism for processing first- and second-order boundaries.....	13
1.7.1 Two-Stream Model	13
1.7.2 Single-Stream Model	14
1.8 Cortical Push-Pull Model.....	16
1.9 Quantitative Receptive Field Models.....	18
1.9.1 Spike-triggered methods	19
1.9.2 Generalized Linear Models (GLM)	21
1.9.3 Convolutional Neural Networks	23
1.10 Electrophysiology and spike sorting.....	24
1.11 Thesis Aims	26
CHAPTER 2	36
2.1 Abstract	37
2.2 Introduction.....	38
2.3 Materials & Methods	40
2.3.1 Animal Preparation	40
2.3.2 Visual Stimuli	41
2.3.3 Extracellular recording.....	41

2.3.4 Analysis.....	42
2.4 Results.....	44
2.4.1 Non-oriented receptive fields in cat Area 18.	44
2.4.2 Y-like spatial nonlinearities of non-ori receptive fields	48
2.4.3 Linear and nonlinear spatial frequency relationships of Y-like cortical neurons	49
2.4.4 Orientation tuning of linear and nonlinear responses of Y-like cortical neurons	50
2.4.5 Responses of Y-like cortical neurons to second-order stimuli	52
2.4.6 A possible cortical circuit utilizing Y-pathway inputs to build cue-invariant receptive fields.....	53
2.5 Discussion.....	56
2.5.1 Non-ori cells in cat Area 18	57
2.5.2 Nonlinear Y-like spatial summation	58
2.5.3 Neural mechanism for building cue-invariant receptive fields.....	58
2.5.4 Implications for second-order processing in other mammals	60
CHAPTER 3	79
3.1 Abstract.....	80
3.2 Introduction.....	81
3.3 Methods.....	83
3.3.1 Animal Preparation and Electrophysiological Recordings	83
3.3.2 Visual Stimuli and Recording Protocol	84
3.3.3 Receptive Field Model.....	86
3.3.4 Model Training and Performance	87
3.4 Results.....	89
3.4.1 Mapping nonlinear subunit receptive field models with deep learning	89
3.4.2 Predictive performance for novel stimuli	90
3.4.3 Intermediate nonlinearity (α).....	91
3.4.4 Two-branch model	93
3.5 Discussion	94
3.5.1 Relationship of receptive fields for LGN neurons vs. retinal ganglion cells.....	94
3.5.2 Visual stimuli for system identification.....	95
3.5.3 Convolutional Neural Networks	97

3.5.4 Future directions	99
CHAPTER 4	111
4.1 Abstract	112
4.2 Introduction	113
4.3 Methods	115
4.3.1 Animal Preparation and Electrophysiological Recordings	115
4.3.2 Visual Stimuli and Recording Protocol	115
4.3.3 Receptive Field Model Estimation	116
4.3.4 Linear-nonlinear receptive field model	117
4.3.5 ON-OFF subunit model	118
4.3.6 Receptive Field Analysis	119
4.4 Results	121
4.4.1 ON-OFF subunit model	121
4.4.2 ON-OFF spatial arrangement	122
4.4.3 ON-OFF temporal dynamics	124
4.5 Discussion	126
4.5.1 Limitations	126
4.5.2 Comparison to previous findings on ON-OFF integration by cortical neurons	129
4.5.3 Implications for second-order processing	131
CHAPTER 5	141
5.1 Summary of findings	142
5.2 Limitations and caveats	144
5.2.1 Anesthesia and brain state	144
5.2.2 Laminar organization	147
5.2.3 Sampling bias	148
5.2.4 Visual stimuli	149
5.2.5 Model architecture and training	150
5.3 Future directions	152
5.3.1 Multi-stage convolutional model	152
5.3.2 Two-photon imaging and neural network models	154
5.3.3 Second-order processing	155

REFERENCES	157
------------------	-----

LIST OF FIGURES

Figure 1.1 : First-order and second-order boundaries in visual scenes.....	29
Figure 1.2 : Filter-Rectify-Filter (FRF) model for detecting second-order boundaries.....	30
Figure 1.3 : Spatial frequency tuning for typical X- and Y-type cells measured with luminance gratings.....	31
Figure 1.4 : Subunit receptive field model of a Y-type retinal ganglion cell.	32
Figure 1.5 : Luminance Modulation and Contrast Modulation gratings	33
Figure 1.6: Spatial frequency tuning of a model cortical neuron.	34
Figure 1.7 : Two-stream processing model for first- and second-order boundaries.....	35
Figure 2.1 : Orientation tuning to drifting luminance gratings recorded with multielectrode arrays.....	63
Figure 2.2 : Receptive field properties of non-orientation selective neurons.	64
Figure 2.3 : Y-like non-orientation selective neurons in Area 18.....	66
Figure 2.4 : Linear and nonlinear spatial frequency tuning of cortical Y-like neurons.....	68
Figure 2.5 : Linear and nonlinear orientation responses of Y-like non-ori neurons.....	70
Figure 2.6 : Responses of Y-like non-ori neurons to contrast modulation (CM) gratings	72
Figure 2.7 : Neural circuit model for cue-invariant receptive fields constructed from Y-pathway.	74
Figure 2.8 : Spatial tuning of balanced and unbalanced push-pull model with Y-pathway inputs.	76
Figure 2.9 : Selectivity of the unbalanced push-pull model for relative spatial phase of luminance and contrast boundary.....	78
Figure 3.1 : Construction of naturalistic texture stimuli.....	101
Figure 3.2 : Model architecture of 1-filter convolutional neural network to model LGN receptive fields.....	102
Figure 3.3 : Receptive field model of an example Y cell.	103
Figure 3.4 : Trained neural networks for four example Y cells.....	104
Figure 3.5 : Comparing predictive power of subunit models to linear models in predicting neural responses.....	105
Figure 3.6 : Degree of spatial nonlinearity within receptive fields of X- and Y-type neurons. .	106
Figure 3.7 : Relationship between model performance improvement and nonlinearity.....	107

Figure 3.8 : Model architecture of a 2-filter convolutional neural network to model LGN receptive fields.	108
Figure 3.9 : 2-filter receptive field model for example neuron of Figure 3.3.	109
Figure 3.10 : Comparison of predictive performance across different receptive field models...	110
Figure 4.1 : Push-pull arrangement of ON and OFF pathway inputs within a simple cell receptive field.	133
Figure 4.2 : Subunit model architecture for simple type cells.	134
Figure 4.3 : Estimated ON-OFF subunit receptive field model for an example simple cell.	135
Figure 4.4 : Spatial ON-OFF weight maps for six example simple cells.	136
Figure 4.5 : Spatial arrangements of ON and OFF inputs within simple cell receptive fields of Area 17 and 18.	137
Figure 4.6 : Temporal dynamics of ON and OFF inputs within receptive fields of four example simple cells.	139
Figure 4.7 : Temporal dynamics of ON and OFF inputs within receptive fields of simple cells in Area 17 and 18.	140

ABBREVIATIONS AND SYMBOLS

CM, contrast modulation

CNN, convolutional neural network

DoG, difference of Gaussian

F1, first harmonic

F2, second harmonic

FRF, Filter-Rectify-Filter

GLM, generalized linear model

H-C, Horsley Clarke

LGN, lateral geniculate nucleus

LM, luminance modulation

LN, linear-nonlinear

MSE, mean squared error

OB, orientation bias

pReLU, parametric rectified linear unit

PSTHs, peri-stimulus time histograms

RGC, retinal ganglion cell

ReLU, rectified linear unit

RMS, root mean square

SF, spatial frequency

STA, spike triggered average

STC, spike triggered covariance

SVD, singular value decomposition

ABSTRACT

Our visual system is sensitive to boundaries defined by differences in cues such as luminance (first-order cue), as well as texture, contrast, or motion (second-order cues). Gradients in these cues can be utilized to perform tasks such as figure-ground segregation and 3D shape perception. A significant fraction of neurons in the early visual cortex of cats and monkeys have been shown to be selective to both first- and second-order boundaries. These neurons are thought to be the neural correlate for perceptual encoding of such boundaries. They are selective for the same boundary orientation irrespective of the cue (first- or second-order) that defines it (“form cue-invariance”), which makes these neurons powerful candidates for the task of segmentation. However, the neural circuitry that gives rise to this selectivity for the early stages of visual processing remains unclear. To address this question, I perform neurophysiological recordings at the early stages of the visual pathway in cats, and then build biologically inspired neural circuit models that can account for visual response properties of neurons at subcortical as well as early cortical stages.

In Chapter 2, I use multi-electrode recordings to demonstrate the presence of a significant fraction of neurons in cat Area 18 with nonlinear receptive fields like those of subcortical Y-type cells. These neurons have receptive field properties intermediate between subcortical Y cells and cortical orientation selective cue-invariant neurons. These are strong candidates for building cue-invariant orientation-selective neurons. Furthermore I present a novel neural circuit model that pools such Y-like neurons in an unbalanced “push-pull” manner, to generate orientation-selective cue-invariant receptive fields.

In Chapter 3, I estimate biologically constrained neural network models of cat LGN receptive fields using recent machine learning methods (deep learning). The receptive fields are modeled as arising from a two-stage convolutional neural network model. The first stage, corresponding to retinal bipolar cell subunits, is modeled as a convolutional filter layer, and the second stage is modeled as a pooling layer. These two layers are separated by an intermediate parametric nonlinearity. I train such a neural network model for each recorded LGN neuron, using its spiking responses to naturalistic texture stimuli. These models are not only better in comparison to the standard linear-nonlinear models at predicting response to arbitrary stimuli, but they also recover biologically interpretable subunit models.

In chapter 4, I evaluate the integration of ON- and OFF-pathway inputs by individual neurons in early cortical areas of the cat (Area 17 and Area 18). In this study, I model receptive fields of cortical simple cells as a linear weighted sum of rectified inputs from model ON- and OFF-center LGN afferents, with the weights estimated using a regression framework. The estimated models reveal significant asymmetries in spatiotemporal integration of ON and OFF signals within simple cell receptive fields. These observed asymmetries could provide the neural mechanism for generating cue-invariant receptive fields from Y-pathway inputs.

In summary, I put together our knowledge of retinal as well as early cortical processing to show how spatial nonlinearities emerging from the retina could provide an essential basis for cortical visual processing. I further evaluate these neural mechanisms by estimating single neuron receptive field models, using modern system identification methods. Finally I propose, and provide supportive evidence for, a novel neural circuit mechanism that could explain the

cue-invariant processing of luminance- and texture-defined boundaries through a common pathway.

RÉSUMÉ

Notre système visuel est sensible aux démarcations définies par des différences entre des indices tels que la luminance (indices de premier ordre) ainsi que la texture, le contraste ou le mouvement (indices de second ordre). Les gradients de ces indices peuvent être utilisés afin d'accomplir des tâches telles que l'identification d'un objet par rapport à l'arrière-plan et la perception de formes en 3D. Une proportion significative de neurones dans le cortex visuel précoce des chats et des singes a été identifiée comme étant sélective à la fois aux démarcations de premier et de second ordre. Ces neurones sont considérés comme étant le corrélât neuronal de l'encodage perceptif de ces démarcations. Ils sont sélectifs aux mêmes orientations de démarcation indépendamment de l'indice qui les définit (invariance d'indice de forme), ce qui fait de ces neurones des candidats compétents aux tâches de segmentation. Cependant, les circuits neuronaux qui génèrent cette sélectivité aux stades précoces du traitement visuel restent mal définis. Pour répondre à cette question, j'ai effectué des enregistrements neurophysiologiques aux stades précoces du système visuel chez le chat et ai ensuite construit des modèles de circuits neuronaux qui peuvent rendre compte des propriétés des réponses visuelles des neurones au niveau sous-cortical ainsi qu'aux stades corticaux précoces,

Dans le Chapitre 2, J'utilise des enregistrements multi-électrodes pour démontrer la présence d'une proportion significative de neurones dans l'aire 18 du chat présentant des champs récepteurs non-linéaires tels que ceux des cellules Y sous-corticales. Ces neurones ont des propriétés de champs récepteurs intermédiaires entre les cellules sous-corticales Y et les neurones corticaux sélectifs à l'orientation indifférents aux indices. Ce sont des candidats privilégiés pour le façonnement de neurones sélectifs à l'orientation indifférents aux indices. En

outre, je présente un nouveau modèle de circuit neuronal qui groupe ce type neurones de style Y avec un déséquilibre “pression-traction” pour générer des champs récepteurs sélectifs à l’orientation indifférents aux indices.

Dans le Chapitre 3, j’évalue des modèles de réseaux neuronaux contraints par la biologie des champs récepteurs du CGL du chat en utilisant des méthodes récentes d’apprentissage de machine (apprentissage profond). Les champs récepteurs sont modélisés tels qu’émergeant d’un modèle de réseau neuronal convolutionnel à deux niveaux. Le premier niveau, correspondant aux sous-unités des cellules bipolaires rétiniennes, est modélisé en tant que couche de filtrage convolutionnel et le second niveau est modélisé en tant que couche de groupement. Ces deux couches sont séparées par une non-linéarité paramétrique intermédiaire. Nous avons entraîné un tel modèle de réseau neuronal pour chaque neurone du CGL enregistré, en utilisant sa réponse supraliminaire à des stimuli de textures naturalistiques. Ces modèles ne sont pas seulement meilleurs en comparaison des modèles linéaires/non-linéaires standards pour la prédiction de la réponse à des stimuli arbitraires, mais ils expliquent également les modèles de sous-unités qui peuvent être interprétés biologiquement.

Dans le Chapitre 4, je planifie l’intégration des entrées des voies ON et OFF par les neurones individuels des aires corticales précoces du chat (aires 17 et 18). Dans cette étude, je modélise les champs récepteurs des cellules simples corticales comme une somme linéaire pondérée des entrées rectifiées des afférents à centre ON et OFF du CGL, avec des poids estimés dans le cadre d’une régression. Les modèles estimés révèlent des dissymétries significatives dans l’intégration spatio-temporelle des signaux ON et OFF dans les champs récepteurs des cellules

simples. Ces dissymétries pourraient sous-tendre les mécanismes neuronaux qui génèrent des champs récepteurs indifférents aux indices à partir des entrées des voies Y.

En résumé, j'ai combiné nos connaissances du traitement aussi bien rétinien que cortical pour montrer comment des non-linéarités spatiales émergeant au niveau de la rétine peuvent fournir une base essentielle au traitement visuel cortical. J'ai évalué en détails ces mécanismes neuronaux en estimant des modèles de champs récepteurs de neurones par des méthodes modernes d'identification de systèmes. Finalement, je propose et fournit des arguments en faveur d'un nouveau mécanisme neuronal qui pourrait expliquer le traitement indifférent aux indices des démarcations définies par la luminance ou la texture par une voie commune.

1

In Chapter 1, I briefly explain structure of the thesis, followed by a comprehensive literature review of the relevant background. Finally, I state the aims of the three data chapters and the rationales behind them.

CHAPTER 1

General Introduction

1.1 Thesis Structure

This thesis is written and organized in a manuscript based style containing five separate chapters. The first chapter provides a general overview of the topics pertinent to the three data chapters (2, 3 & 4) of this thesis which are written as three separate manuscripts. The overall theme of the thesis is to understand visual processing at the early stages of the visual pathway by quantitatively studying receptive field properties of single neurons at subcortical and cortical level. Since all three studies included in this thesis were conducted on cats, the introduction chapter is heavily focused on cat single-unit neurophysiology literature with some comparison to primates. In the final discussion chapter, I summarize and link the findings from three data chapters. Finally, I discuss limitations and caveats about the methodology and approach of these studies with possible future directions.

1.2 First- and second-order information in visual scenes

Our visual world is cluttered with a variety of objects and is highly dynamic. Despite this we seem to effortlessly identify objects and form a coherent perception of the world around us. One of the cues that we use for performing visual tasks is the luminance variation that occurs in natural scenes. Luminance cues are also referred to as "first-order" cues in the literature. Usually, different objects differ in their luminance, and this can help us to delineate an object from its

background ("figure-ground segregation"). Changes in luminance can also be used for interpreting the 3D shape of objects ("shape-from-shading"). However, relying only on luminance cues can introduce ambiguities. For example, luminance changes could be a result of material differences between an object and its background, or because of different illumination or shadows, or perhaps folds in a textured material.

Fortunately, our visual system is also sensitive to "second-order" cues such as texture, contrast or motion differences. Boundaries arising from second-order cues either occur independently or co-occur with first-order cues in natural scenes (Johnson and Baker, 2004). When an object occludes another object, very often the two objects not only have different luminance but also have different texture or local contrast. However, shadows cast on an object only generate luminance variations but other cues don't change. Thus, second-order cues can be helpful in disambiguating such false boundaries. Furthermore, such second-order cues have also been shown to help in interpreting 3D shape from shading (Schofield et al., 2010). Thus, second-order cues independently or in combination with first-order cues can provide vital information to our visual system.

Boundaries formed by luminance changes, as shown in Figure 1.1 between the tree and grass, can be detected by Gabor-shaped linear filters. These filters sum luminance within their subfields linearly and are selective for orientation of boundaries. For example, Gabor-like receptive fields in the primary visual cortex can encode such boundaries. However, boundaries formed by purely second-order cues, such as between the grass and its reflection in the water in Figure 1.1 may not contain any variation in mean luminance. Therefore, spatially linear filters cannot detect such boundaries. Hence it has been suggested that a specialized mechanism (filter-rectify-filter or FRF) is required for detecting such boundaries (Zhou and Baker, 1993), in which

a bank of small filters detects fine texture elements in the image, then the outputs of these filters are rectified and pooled together by a coarse scale filter as shown in Figure 1.2. A similar mechanism has been suggested by (Karklin and Lewicki, 2003) to efficiently code higher-order structure in natural images. However, the neural substrate for such processing remains unclear.

1.3 Early visual pathways

Visual signals from the photoreceptors in the retina diverge to generate two major divisions, i.e. the "ON" and the "OFF" pathways. Neurons in the ON pathway respond to the relatively bright regions in the visual scene and those in the OFF pathway to the dark regions. Receptive fields of retinal ganglion cells (RGCs) that form the output of retina are concentric with antagonistic centre-surround organization. ON-centre RGCs are excited by light stimuli in the centre of their receptive field and are inhibited by light stimuli in the surround region. On the other hand, the opposite happens for OFF-centre RGCs. These ON and OFF pathways have little interaction in the retina and LGN (but see Liang & Freed, 2010) until they are combined in the primary visual cortex (Schiller, 2010). Usually ON and OFF pathways are thought to be symmetric to one another in terms of their spatio-temporal receptive field properties. However, they have some interesting asymmetries that might give rise to some perceptual phenomenon. For example, visual signals are processed faster in the OFF pathway as compared to the ON pathway (Jin et al., 2011). This difference might explain why human subjects are faster at detecting dark regions compared to bright regions (Komban et al., 2014).

Based on receptive field properties, RGCs in the cat are further divided into different categories. X- and Y-type RGCs form two major fractions of cell types that send visual signals to

the LGN and the cortex. Cells are classified into these two types by measuring the linearity of spatial summation within their receptive fields (Enroth-Cugell & Robson, 1966). When tested with contrast-reversing gratings of high spatial frequencies, Y-type cells respond nonlinearly at a second harmonic (F2) of a temporally modulated stimulus, but X-type cells do not show such nonlinearity and respond linearly at the first harmonic. On the other hand, in response to drifting gratings, Y cells respond at the first harmonic to low spatial frequencies. Thus Y cells show dual spatial frequency tunings (Figure 1.3), one at low spatial frequencies corresponding to centre-surround organization of the receptive field and the other at high spatial frequencies corresponding to subunit structure (will discuss about this in detail in the following section) within the receptive field. At a given eccentricity Y cells have bigger receptive fields than X cells, and consequently Y cells are selective for lower spatial frequencies. Further, there is a close correspondence between functional X- and Y-type cells and morphologically defined β - and α -type retinal ganglion cells (Cleland et al., 1975; Wässle et al., 1975; Rodieck, 1979). β cells have small cell bodies and thinner axons while α cells have much bigger cell bodies and thicker axons. Both X- and Y-cells have ON- and OFF-centre receptive fields.

There is also another category of RGCs in the cat, W-type cells (Fukuda and Stone, 1973), which have not been well studied. They form a heterogeneous mixture of cells that are not classified into X- or Y-type. Some of the W cells are direction selective, while some have receptive fields giving mixed ON/OFF (excitation or inhibition) responses throughout their receptive field. W cells have large receptive fields like Y cells (Wilson et al., 1976; Dreher and Sefton, 1979).

1.4 Retinal Circuitry

In the retina, visual signals undergo multiple stages of processing through the cascade of various retinal cell types. Light is absorbed by the photoreceptors (rods and cones) and transduced into electrical signals, which are sent to ON- and OFF-type bipolar cells. Horizontal cells also integrate inputs from multiple photoreceptors and provide inhibition to bipolar cells. Horizontal cells are thought to be responsible for the surround mechanism of bipolar cell receptive fields. Then bipolar cells make direct synapses on to the retinal ganglion cells. Bipolar cells also make synapses with amacrine cells, which pool inputs from many bipolar cells and then provide inhibition to retinal ganglion cells, thus generating the surround mechanism of the retinal ganglion cells.

Interestingly, the bipolar cells that synapse onto Y ganglion cells provide half-wave rectified inputs (Demb et al. 1999, 2001a), and are thought to be the neural substrate for the nonlinear subunits within the Y cell receptive fields. This rectification is a consequence of a low spontaneous release rate of glutamate from the bipolar cells' synaptic terminal onto the RGC. Therefore, these bipolar cells cannot provide a negative output. This mechanism is contrary to the previously proposed mechanism in which the nonlinearity of the amacrine cells was thought to give rise to the Y cell spatial subunit nonlinearity (Hochstein and Shapley, 1976). Previously it was thought that Y cells contain two overlapping receptive fields, one with a linear centre-surround organization and the other with small nonlinear subunits covering both the centre as well as the surround. Demb et al. (1999) showed that in Guinea pig retina, even when the amacrine pathway was blocked pharmacologically, the spatial nonlinearity of the Y cells was preserved. Thus, the linear and nonlinear spatial properties of the Y cells were shown to both

arise from a single pathway due to a half-wave rectification of bipolar outputs rather than from two separate pathways.

Based on this evidence, receptive fields of Y cells have been modeled as a cascade of alternating spatially linear filters and simple nonlinearities (“subunit”) model, shown in Figure 1.4. Here the first-stage filters correspond to bipolar cells that are modeled as small linear filters with centre-surround organization. The outputs of these filters are rectified and then pooled, with synaptic weights that form the large-scale centre-surround organization of the RGC. A crucial aspect of the model is that the centre-surround strengths of the first-stage filters are imbalanced, allowing low spatial frequencies to pass the first stage and generate linear responses in the RGC (Rosenberg & Issa, 2011). Therefore this model can explain linear responses to low spatial frequency gratings as well as nonlinear responses to high spatial frequency gratings.

Interesting asymmetries have been demonstrated in the nonlinearities of ON- and OFF-centre Y cells in various species. For example, an *ex vivo* study in Guinea pig retina (Demb et al 2001a) showed that rectification of the signals from bipolar cells in the OFF-ganglion cell circuit is close to a half-wave rectification. However, in ON-centre ganglion cells, rectification was found to be much weaker. Despite weaker rectification in the ON cells, nonlinear responses in both types of cells were found to be similar. This discrepancy was later addressed in a study (Borghuis et al 2013) where they measured glutamate release at the bipolar-ganglion cell synapse, using 2-photon imaging in the mouse retina. They found that for ON-type cells, even though release of glutamate varied linearly with contrast of the visual input, the temporal dynamics for response increments and decrements were different. Hence due to this asymmetry when the responses of multiple bipolar cells are pooled together, responses to high SF gratings

would not cancel each other out, thus giving a relatively stronger nonlinear response than would be expected from just the weak rectification.

1.5 Visual signal transmission from retina to cortex

Retinal ganglion cells form the output layer of the retina, and send signals via a bundle of long axons to the lateral geniculate nucleus (LGN). The LGN in the left hemisphere receives input only from the right visual field and the LGN in the right hemisphere receives input only from the left visual field. The signals from the two eyes remain largely segregated in the LGN. The LGN in the cat has a laminated structure, with layers labelled as A, A1, C, C1, C2, C3. Neurons in layers A, C, and C2 receive retinal inputs from the contralateral eye, while neurons layers A1 and C1 receive retinal inputs from the ipsilateral eye. Layer C3 does not receive input from the retina, but from the superior colliculus (SC) (Torrealba et al., 1981). Unlike primates, in cat LGN there is no segregation of inputs from different retinal ganglion cell types. In primates, not only do the LGN layers receive inputs from only one eye, but also from one physiological type of RGC. However, in the cat LGN projections from both X- and Y-type cells are present in layers A and A1. In C layers inputs are mostly from Y and W cells. Despite receiving a mixture of inputs within a layer, there is little mixing of inputs from X and Y pathways within single LGN neurons (Bullier and Norton, 1979). Thus, LGN neurons in layers A and A1 have functional properties like either X- or Y-type RGCs. Different studies have reported slightly different proportions of cell types across LGN laminae. In layers A and A1 around 50-67% of neurons are X-type and 23-50% neurons are Y-type, while in layer C around 40-43% of neurons are Y-type and 43-52 % neurons are W-type, with very few X-type neurons

present in layer C. Layers C1-C3 contain only W-type neurons (Wilson et al., 1976; Cleland et al., 1976).

Some previous studies have compared receptive field properties of the RGC inputs and the LGN neurons by recording S-potentials. The S-potentials are tiny, slower monophasic spikes picked up by extracellular electrodes while simultaneously recording action potential spikes from LGN neurons. These S-potentials are thought to be the extracellularly recorded EPSPs in the LGN neuron (Kaplan and Shapley, 1984). Thus, S-potentials are a measure of input spikes from RGC afferents onto LGN neurons. These S-potentials either just precede the LGN spike or are present in isolation. Thus, not every input spike from a RGC triggers an LGN spike. So and Shapley (1981) found that spatial frequency tuning properties of both X- and Y-type LGN neurons were very similar to the receptive field properties of their S-potentials. Furthermore, an orientation bias observed in some LGN neurons was shown to be inherited from the retinal inputs rather a result of LGN processing (Soodak et al., 1987).

Even though spatial properties do not change much in the LGN from those in RGCs, the LGN neurons' receptive fields do have a "push-pull" arrangement: the principal neurons in the LGN not only receive direct excitatory inputs from RGCs but also receive feedforward inhibition via LGN interneurons (Martinez et al., 2014). This inhibition has been shown to drive LGN neurons in two modes of spiking (tonic firing and bursting) while viewing natural movies (Wang et al., 2007). This inhibitory mechanism in the LGN is thought to preserve information encoded in the spike times of RGCs, and effectively transmit information to the cortex (Wang et al., 2011). Thus inhibition in the LGN shapes temporal transmission of signals from retina to cortex, but does not alter the spatial receptive field properties.

LGN principal cells in cat send axons to the visual cortex, located at the occipital lobe of the neocortex. LGN cells project not only to the primary visual cortex (Area 17) but also to Areas 18 and 19 (Stone and Dreher, 1973; Humphrey et al., 1985). This is unlike the visual pathway of primates, where inputs from LGN project only to primary visual cortex (V1) and not to higher visual areas. There are interesting differences in the thalamo-cortical projections to these different visual areas in the cat with respect to their physiological types. Inputs to primary visual cortex (Area 17) are mainly from X-type LGN neurons, with only a small fraction of input from Y- and W-type neurons. Interestingly, inputs to secondary visual cortex (Area 18) are mainly from Y-type LGN neurons, with a small fraction of input from X and W cells. Area 18 also receives projections from Area 17 and vice-versa. Area 19 receives a majority of its input from the W pathway, and a small fraction from the Y pathway. However in primates, inputs from parvocellular and magnocellular pathways project to different layers ($4C\beta$ and $4C\alpha$ respectively) in the primary visual cortex. Thus there are important differences in the projections of thalamic inputs to visual cortex in cats and primates.

1.6 Receptive field properties of cortical neurons

Hubel and Wiesel (1959) discovered that receptive fields in the early visual cortex are very different from receptive fields in the retina and LGN. They found that receptive fields in primary visual cortex do not have concentric centre-surround receptive fields like subcortical neurons. Instead they are selective for a specific orientation. One class of cells that they classified as simple cells have ON and OFF regions like subcortical neurons, but instead of being concentric they are elongated and located side by side. Such receptive fields could best be

stimulated with black or white bars whose orientation matched that of the receptive field. The other type of cell they classified as complex cells, which also have similar orientation selectivity but do not have distinct ON and OFF regions in their receptive fields. Thus the complex cells are not sensitive to the position or luminance polarity of stimuli in their receptive fields. Furthermore, both simple and complex type cortical neurons have bandpass spatial frequency tuning (Movshon et al., 1978), unlike subcortical neurons that mostly have low pass tuning.

In addition to cortical neurons' selectivity to luminance changes (first-order boundaries), many neurons in the early visual cortex (~ one-quarter in Area 17, one-half in 18) of cat have been shown to also selectively respond to second-order boundaries (Zhou & Baker, 1993). The response properties of these neurons have been thoroughly characterized with contrast-modulated (CM) grating stimuli. This is a very simple kind of second-order stimulus in which the contrast of a high spatial frequency luminance grating (carrier) is modulated by a low spatial frequency grating (envelope), as shown in Figure 1.5. These "double duty" neurons have three kinds of spatial frequency tunings (Mareschal & Baker, 1999), as shown in Figure 1.6. Firstly, like the other cortical neurons they are tuned to the spatial frequency of luminance gratings. Secondly, these neurons are tuned to the spatial frequency of the carrier of CM gratings, which is outside the luminance passband (range of tuning to luminance grating) of the neuron. Thirdly, these neurons are also tuned for the envelope spatial frequency of the CM grating; the optimal envelope spatial frequency is similar to or slightly less than the optimal luminance spatial frequency.

Furthermore, these neurons show similar orientation tuning for luminance gratings and for the envelope of CM gratings ("form-cue invariance" - e.g. Mareschal & Baker, 1998a). On

the other hand, a significant fraction of these neurons are tuned for carrier orientation, though with no systematic relationship between optimal envelope and carrier orientation.

In addition, these neurons show three kinds of temporal frequency response. They show bandpass temporal frequency tuning for LM and the envelope of CM gratings, but the optimal temporal frequency for luminance gratings is usually slightly higher than for an envelope. Temporal frequency response for a drifting carrier is quite variable from one neuron to another, with some neurons preferring a stationary carrier (low pass) while others respond more vigorously to very high carrier temporal frequencies (Rosenberg and Issa., 2011; Gharat & Baker, 2012). In addition to responding to CM gratings, these neurons also respond selectively to other kinds of second-order stimuli such as illusory contours (Song & Baker, 2007) or motion-defined contours (Gharat & Baker, 2012) in a similarly cue-invariant manner. Neurons with similar response properties were recently described in monkey V2 as well (Li et al., 2011).

These neurons also respond in a selective manner when luminance and contrast boundaries are combined, i.e. they respond strongest when LM and CM gratings are superimposed in a phase-aligned manner, and weakest when they are in opposite phase (Hutchinson et al, 2016). Human psychophysics studies have shown that when LM and CM are in-phase they give a percept of corrugated 3D surfaces (Schofield et al 2006). Consequently these neurons could be suitable for detecting illumination changes over textured surfaces as luminance and contrast covaries (Schofield et al 2010; Schofield et al 2006), and thus could contribute to the perception of shape from shading.

1.7 Neural mechanism for processing first- and second-order boundaries

Responses of these neurons cannot be modeled with a linear receptive field, as a linear receptive field would fail to respond to second-order stimuli that contain energy outside the neurons' luminance passband. Instead a linear-nonlinear-linear (filter-rectify-filter, FRF) cascade has been proposed to detect second-order stimuli. Two possible neural circuit mechanisms will be discussed in the following section that could account for first- and second-order responses of these neurons.

1.7.1 Two-Stream Model

Earlier (Zhou & Baker, 1993; Mareschal & Baker, 1998a), a “two-stream” model, as shown in Figure 1.7, had been proposed to account for cortical neurons' responses to first- and second-order stimuli, based upon their tuning properties. According to this model two separate signal processing pathways act in parallel to process first- and second-order stimuli, prior to an Area 17/18 neuron's response. The first stream of the model consists of a linear, coarse-scale Gabor-like oriented filter (F0). This linear stream can respond to luminance gratings (first-order) but not to 2nd-order stimuli because the high carrier spatial frequency is outside the passband of the filter F0. On the other hand, the second stream consists of a bank of small Gabor filters (F1/early stage filters/ subunits) whose outputs are rectified and pooled by a late-stage large Gabor filter (F2). Neurons' tuning to the carrier of CM gratings corresponds to that of the early filters F1, while the envelope tuning corresponds to that of the late stage filter F2. Tuning of this late filter F2 is similar to that of the linear filter F0 of the first stream, leading to a cue-invariant response. This second stream can respond to second-order stimuli but not first-order stimuli,

since early filters (F1) are bandpass tuned for high SFs compared to first-order stimuli. The outputs of these two independent streams are then summed by a cortical neuron.

The early stage filters (F1) for the second stream were previously assumed to be Gabor filters because some of the cortical CM-responsive neurons have orientation tuning to the carrier grating (Mareschal & Baker, 1998a). Since orientation tuning was generally thought to first arise in the cortex, it appeared most likely that the early filters were high spatial frequency-selective Area 17 neurons. This model is successfully able to explain the selectivity of these neurons to luminance- (1st order) as well as contrast-, texture- & motion-defined (2nd order) boundaries (Zhou & Baker, 1993; Mareschal & Baker, 1999; Song & Baker, 2007; Gharat & Baker, 2012). However, one of the criticisms of this model is that it requires a set of neurons that are responsive only to 2nd-order stimuli and not responsive to 1st-order stimuli, corresponding to the late stage filter (F2) of the second stream. But to date such neurons have not been described. Also, some of CM-responsive neurons can respond to very high carrier temporal frequencies (Rosenberg & Issa, 2011; Gharat & Baker, 2012), which is inconsistent with the early filters that detect carrier gratings being cortical neurons (which are unresponsive to high temporal frequencies). So this raises the question whether the early filters might be subcortical, after all.

1.7.2 Single-Stream Model

Demb et al (2001b) instead suggested a single-stream model that can respond to both first- and second-order stimuli. According to this model (Figure 1.4), processing of both first- and second-order stimuli begins in the retina through the nonlinear Y cells. Retinal ganglion Y cells respond in a nonlinear fashion to contrast-reversing high spatial frequency luminance gratings, with a frequency-doubled response (Enroth-Cugell & Robson, 1966; Hochstein &

Shapley, 1976). This nonlinear receptive field property emerges due to rectified inputs from retinal bipolar cells, which act as nonlinear subunits (Demb et al 1999). Furthermore, due to these nonlinear subunits, Y cells also respond to contrast-modulated (CM) gratings (Demb et al 2001b). In addition to responding in a nonlinear fashion, Y cells respond linearly to low spatial frequency LM gratings (Enroth-Cugell & Robson, 1966; Hochstein & Shapley, 1976). This linear response can be accounted for by input from the same pool of bipolar cells, if their center-surround mechanism is not balanced (as discussed earlier). In this manner, a retinal circuit containing only a single stream can respond to both LM as well as CM gratings. Demb et al (2001b) proposed that Y cell input from retina to cortex via the LGN provides the basis for cortical neurons' response to both LM and CM gratings.

Anatomical studies have shown that Area 18 in the cat receives most of its input from LGN Y cells, while Area 17 receives only a small fraction of input from Y cells and a majority of its input from X cells. This is consistent with the results that around half of the neurons in Area 18 respond to 2nd-order stimuli while only about one-fourth of neurons in Area 17 respond to 2nd-order stimuli (Zhou & Baker, 1993). Furthermore, optimal carrier spatial frequencies of cortical neurons are in the same spatial frequency range (0.5 to 2 cpd) to which Y cells respond in a nonlinear manner (So and Shapley, 1981). However, earlier this model was not given a serious consideration because of CM carrier orientation selectivity in cortical neurons (Mareschal & Baker, 1998a). It was assumed that subcortical neurons couldn't show orientation selectivity. But subsequently it was shown that cat LGN Y cells do in fact show CM carrier orientation tuning like that of cortical neurons (Rosenberg et al 2010). Also like cortical neurons, LGN Y cells prefer lower CM envelope temporal frequencies compared to luminance gratings (Rosenberg and Issa, 2011). In addition, at least some LGN Y cells respond to very high carrier

temporal frequencies, like cortical neurons (Rosenberg & Issa, 2011). Thus, this model is biologically more plausible and a more parsimonious explanation of second-order response properties of cortical neurons.

In addition to cats, other mammalian species (macaque monkey, mouse, rabbit, guinea pig). have Y (α)-like retinal ganglion cells. In particular, in the macaque monkey all parasol (M) cells (one of the major retinal ganglion cell types) give a “Y-cell signature” (Figure 1.3) response (Crook et al, 2008a). Also, another RGC category called upsilon cells/smooth monostriated in macaque also show Y-cell signature response (Petrusca et al., 2007; Crook et al., 2008b). So this could be an underlying mechanism for processing 2nd-order stimuli in all mammalian species, including humans.

1.8 Cortical Push-Pull Model

Simple cells of cat striate cortex have been demonstrated to have a linear spatial summation. This was earlier shown by stimulating ON and OFF subregions of simple cells with bar-shaped stimuli while measuring their spiking responses extracellularly (Tolhurst and Dean, 1987). When bars matching the polarity of receptive field subregions were presented (ON region – white bar, OFF region – black bar), neurons gave strong excitatory responses. However when pairs of white bars (or black bars) were simultaneously presented in both ON and OFF regions, neurons would give little or no response. This result showed that subregions not only respond to stimuli of matching polarity but also provide covert inhibition for opposite polarities.

This phenomenon was directly demonstrated through intracellular recordings in simple cells, which could measure both excitation (EPSPs) as well as inhibition (IPSPs) along with

spikes in a recorded neuron. Ferster (1988) demonstrated that presenting a white bar in an ON subregion or a black bar in an OFF subregion generated a barrage of EPSPs (excitation or “push”), but the presence of a black bar in an ON subregion or a white bar in an OFF subregion generated a barrage of IPSPs (inhibition or “pull”). Thus, this study showed that EPSPs and IPSPs have spatially opponent receptive fields of comparable strength, giving rise to spatial linearity within the simple cell receptive field due to this "push-pull" arrangement.

Since thalamic inputs to the cortex are excitatory (Alonso et al., 2001), the excitation within the receptive fields of simple cells in thalamo-recipient layers is thought to arise from direct LGN afferents while inhibition (“pull”) is thought to arise from intracortical inputs (Hirsch, 2003). Reid and Alonso (1995) used simultaneous recordings in LGN and cortex to demonstrate that direct excitation from LGN afferents has the same contrast polarity as the recipient simple cell subregion - i.e. ON afferents excite ON subregions, and OFF afferents excite OFF subregions. Inhibition is thought to arise from simple cells with receptive fields having an opposite spatial polarity but otherwise identical receptive fields (Hirsch et al., 2003). A population of inhibitory interneurons with simple type receptive fields has been shown in layer 4 of cat striate cortex (Hirsch et al., 2003). Furthermore, simple cell receptive fields have been shown to have a varying degree of push-pull, but the distribution of a push-pull index has been shown to be like that of the LGN population (Martinez et al., 2005). This push-pull arrangement is not just restricted to the spatial domain, but also extends in the temporal domain of a given receptive field (Priebe and Ferster, 2005). In direction selective simple cells, inhibition as well as excitation are also tuned (spatiotemporally) for the same direction of motion (Priebe and Ferster, 2005).

These push-pull studies mentioned so far have been performed in cat striate cortex (Area 17). It remains unclear whether such an arrangement is also present in cat Area 18, which also receives direct LGN afferent input, and contains both simple and complex type cells. However, Area 18 receives the majority of its input from Y-type LGN cells (Stone and Dreher, 1973; Humphrey et al., 1985) that contain interesting spatial nonlinearities as discussed earlier. Hence, it is important to study the push-pull arrangement in cat Area 18, and see how it might impact processing of linear and nonlinear signals arising from the Y pathway.

1.9 Quantitative Receptive Field Models

In the early days of receptive field (RF) studies, visual RFs were mapped in the retina, LGN and primary visual cortex using hand projectors while listening to neural responses on an audio monitor (e.g. Hubel and Wiesel, 1959). More recently, over the past two decades, various system identification methods like Spike-Triggered Average (STA) (Chichilnisky, 2001), Spike-Triggered Covariance (STC) (Schwartz et al., 2006), Generalized Linear Model (GLM) (Wu et al., 2006), Phase separated Fourier model (David et al., 2004), or Neural network model (Prenger et al., 2004) have been used to estimate these RFs. These methods have enabled neuroscientists to not only visualize qualitatively the shape of these RFs but also to quantitatively measure their properties. Progress has been made to estimate these RFs under visual stimulation with artificial white noise stimuli as well as natural images (Wu et al., 2006). Receptive fields have usually been modeled as linear filters - however this approach is only appropriate for receptive fields with linear spatial summation, such as cortical simple type cells. More recently efforts have begun to capture nonlinearities arising from hierarchical visual processing, using multi-stage

models (e.g. Freeman et al., 2015; Mineault et al., 2012). These estimated quantitative models allow us to summarize receptive field properties across the population of neurons. They give insights into signal transformation, and allow us to test our understanding of the system by measuring predictive performance of these models on novel arbitrary visual stimuli. In the following section I will briefly summarize most common receptive field mapping methods with their advantages and disadvantages.

1.9.1 Spike-triggered methods

The Spike-Triggered Average (STA), or reverse correlation, has been very popular and one of the first methods to map complete 3D spatiotemporal (space-space-time) receptive fields of neurons in the early stages of the visual pathway (e.g. McLean and Palmer, 1989; Chichilnisky, 2001). This method estimates linear filter models by averaging over stimulus frames that triggered spiking activity in the neuron. One can add a parametric nonlinearity at the output of the linear filter to capture the non-negative firing rate of the neuron. However, to correctly estimate a linear-nonlinear (LN) system, this method requires the visual stimuli to be uncorrelated in space and time (Ringach and Shapley, 2004). Hence previous studies have mostly used white noise stimuli for STA estimation of receptive fields. Reverse correlation has been successfully used in mapping receptive fields of retinal as well as LGN neurons (Chichilnisky, 2001; Alonso et al., 1996). However, it is not so effective in early visual cortex, where receptive fields often have more elaborate stimulus selectivity than observed in subcortical neurons. White noise stimuli often do not drive strong neural responses in visual cortex neurons, so the neural responses have very low signal to noise ratio and the estimated models are noisy (Touryan et al, 2005; Felsen et al, 2005; Talebi & Baker, 2012). To circumvent this problem,

sinusoidal gratings drawn from a Hartley basis set (which is an orthogonal and complete basis for 2D images) have been used to map receptive fields in the visual cortex (Ringach et al., 1997). This can provide a robust estimate of receptive fields with a limited amount of neural data. However, this method probes receptive fields only in a limited stimulus space, based on apriori assumptions that might result in failure to reveal more complex RF properties. Laboratory stimuli such as gratings might drive neurons in a different manner than during natural stimulation. Reverse correlation has also been extended to use with natural images, by compensating for the correlations in the images (Willmore and Smyth, 2003). However, this compensation is approximate and can induce artifacts in the estimated receptive field models.

STA models receptive fields as linear filters with a rectified output, so it cannot capture spatial nonlinearities within the receptive field. For example, cortical complex type cells do not have distinct ON and OFF regions like simple cells, thus they are phase-insensitive. Complex cells are thought to sum a number of simple cells with similar spatio-temporal properties at adjacent spatial locations. Consequently the STA method was extended to also incorporate the covariance structure of stimuli that elicit spikes from a neuron (e.g., Touryan et al, 2005; Schwartz et al., 2006). This method, Spike-Triggered Covariance (STC), recovers multiple filters within a receptive field and thus can capture properties such as phase insensitivity of complex cells. However, STC involves optimization of a much larger number of parameters compared to STA, and thus requires a large amount of data. Also, the multiple filters recovered with this method are constrained to be orthogonal, which may not be biologically valid. Finally, classical STC has the same constraint of a white (uncorrelated) stimulus as previously described for STA. However, it can be used with natural image stimuli by compensating for the power spectrum (Touryan et al., 2005)

1.9.2 Generalized Linear Models (GLM)

More recently, receptive field estimation has been formulated as a regression problem in a GLM (general linear model) framework (Wu et al. 2006). In the simplest version, pixel intensities form the "features" of the model that are subjected to a linear weighted summation. The weights of these pixels can be estimated by minimizing the mean squared error between actual neuronal response and the model's predicted response, using iterative methods such as gradient descent. An important advantage of this regression method over reverse correlation is that it can be used with any arbitrary stimuli, such as natural images, since it does not introduce bias in the estimate from spatial correlations in the visual stimulus.

However, the number of model parameters to be learned is often of the same order as the neural data available to train these models, and in addition, neural data is very noisy. Training a model with limited and noisy neural data can lead to overfitting of the model, which captures the noise in the training data in addition to the signal (Wu et al., 2006). In order to avoid overfitting, various regularization methods are available, such as the "lasso" (Mairal & Yu, 2012), ridge regression (Hoerl & Kennard, 1970), early stopping (Yao et al., 2007) etc. However, the lasso and ridge regression methods require optimizing a hyperparameter for the regularization penalty term, using for example a grid search - that means the model needs to be trained several times at different values of the hyperparameter, which can be very time-consuming. The early stopping method does not require optimizing a hyperparameter - with every iteration, the model's predictive performance is measured on a separate holdback dataset (here called the regularization dataset). The training iterations are halted when the predictive performance stops improving. Usually the performance on the training data keeps improving with successive iterations - however the performance on the regularization data improves at first, but then starts to decline

when the model begins to overfit. At this point, model training is halted and model trained up to that iteration is selected. Finally, predictive performance of the model is measured on a "test" dataset that was not used for training or regularization.

The GLM framework described so far can be modified to capture multi-stage processing within receptive fields. Instead of using raw pixel intensities of the visual stimuli as input features to the model, we can perform a basis set transformation (pre-processing). For example, Nishimoto and Gallant (2011) modeled receptive fields of neurons in area MT as a linear weighted sum of rectified inputs from V1-like spatio-temporal filters. For estimating the model, stimulus images were first filtered with a bank of such spatio-temporal RFs, including many different combinations of spatial and temporal properties. Then weights of these filters were estimated through a GLM framework. Similarly Mineault et al. (2012) used the same approach to model receptive fields of MST neurons, using as a basis set a bank of direction selective models that mimic MT neurons.

Thus we can estimate more complicated subunit models based on known biological inputs to the neuron. However, we need to build fairly accurate models of the possible inputs to the neuron. Also, the more complicated the model for possible inputs, the more it increases the size of the filter bank and hence the number of weights to learn for the GLM. However for some situations this is a practical approach. In Chapter 4, I use this preprocessing/GLM approach to model receptive fields of neurons in cat Area 17 and 18, to the separate contributions of ON- and OFF-pathway inputs.

1.9.3 Convolutional Neural Networks

Convolutional neural networks (CNNs) have become popular over the past few years for previously very difficult machine learning problems, such as pattern recognition in images, speech understanding, language translation, etc. (LeCun et al., 2015). The availability of fast GPUs, large datasets and improvements in algorithms have made it possible to train very complicated CNNs that give a state of the art performance on highly challenging tasks. Interestingly, the architecture of these CNNs is inspired from the hierarchical nature of the mammalian visual system. They contain a cascade of spatially localized linear filters convolved across space, followed by static rectifying nonlinearities. Thus, it is a well-suited framework for modeling receptive fields of neurons in the visual pathway, and can capture spatial nonlinearities in receptive fields. A recent study (Yamins et al., 2014) has used the features learned by a CNN on an object recognition task to model receptive fields of neurons in higher areas in the ventral stream (V4 and IT). Interestingly, these models could accurately predict responses of single neurons to randomly selected natural images, thus suggesting that CNNs can capture the hierarchical processing of a biological visual system.

Neural networks have also been used to model receptive fields of neurons at the early stages of the visual pathway. Instead of using the features learned by deep CNNs from computer vision tasks, much simplified neural networks can be trained directly on experimentally recorded neural responses to visual stimuli. A recent study (Oliver, 2014) used regularized neural networks to model receptive fields of neurons in Area V1 and V2 in macaque monkeys. A 3-layer neural network with Dropout regularization (in which a fixed fraction of connections in the network are randomly removed) and added Poisson noise gave the best predictive performance and consistent feature selectivity. Unlike the preprocessing method described above, no

assumptions were made about the feature space, but the method was able to recover clear Gabor-like or Gaussian-like subunit filters. Also, these models were comparable in their predictive performance to those from the pre-processing method. Thus neural networks can provide a strong framework for modeling visual receptive fields and provide insights into the underlying neural computations with biologically interpretable models. In Chapter 3, I train CNNs to model the receptive fields of lateral geniculate nucleus (LGN) neurons to capture the subunit structure within their receptive fields.

1.10 Electrophysiology and spike sorting

Usually with single channel (metal or glass pipette) microelectrodes, neural activity from single neurons is measured extracellularly, then filtered and displayed on an oscilloscope. The position of the electrode is adjusted in small steps until spike amplitudes from a visually responsive single neuron are sufficiently distinct from the background activity to be isolated by a window discriminator. Then the event times of the detected spikes are saved on a hard drive, along with time registration information for the visual stimuli. While this method enables neuroscientists to precisely measure spiking activity of a single neuron even though it is densely packed together with other neurons in the brain, the yield of data from this method is very limited as one could measure responses of only a single neuron at a time. Due to the long durations of some recording protocols and the potential instability of maintaining isolation of single neurons, in practice useful data from only a handful of neurons can be recorded over the course of an acute experiment (~ 2 -4 days).

Also, the single channel approach may have a sampling bias for neurons with big spike amplitudes (large cell bodies). Furthermore, "search stimuli" used to stimulate receptive fields could also introduce bias, as it can stimulate only receptive fields optimal for those stimuli. Spikes from other neurons with low spontaneous activity that are poorly responsive to the search stimulus, might may not be detected. To circumvent these issues, over the past decade, the use of multi-channel electrodes has become much more popular.

In this thesis, I used a multi-channel recording system (Plexon Recorder) to record broadband neuronal signals from multi-electrodes. To record from visual cortex, I used 32-channel NeuroNexus multi-electrodes. For Chapter 2, most of the data from Area 18 was collected using NeuroNexus (A1x32) linear arrays with 100 μ m spacing between recording sites. Thus these electrode sites can span all cortical layers (2.5mm depth) in the cat visual cortex. With this electrode, signals from a neuron are usually picked up at only single electrode site. For Chapter 4, data from Area 17 and 18 were collected using either NeuroNexus linear arrays (A1x32), or NeuroNexus polytrodes (A1x32-Poly2). Unlike with linear arrays, the electrode sites of polytrodes are more densely packed, with 50 μ m spacing of sites arranged in two parallel columns. Consequently, these polytrodes can span only 750 μ m depth, but signals from single neurons often appear on multiple channels. This can help to get a higher yield of neurons with greater reliability from spike sorting.

Neuronexus probes that I used to record from cortex have relatively thin substrate (15 μ m) and cause minimal damage to the cortex as we could record visually responsive neurons from channels all along the length of these probes (Blanche et al., 2005). However, the longer probes that can reach much deeper brain structures use thicker substrate (50 μ m). From our preliminary attempts to record from LGN, we noticed that these probes were causing tissue damage and we

could record visually responsive neurons only from the channels located at the tip of the probe. So instead we used quartz-coated tungsten tetrodes (Thomas) to collect data from the LGN. Due to these tetrodes' conical-shaped tip and thin shaft, they cause minimal tissue damage, but give a much better yield of isolated neurons from spike sorting compared to single channel electrodes.

With Plexon Recorder, we acquired raw data signals (3 Hz to 8kHz; sampling rate, 40 kHz) from all electrode channels, which were streamed to a hard disk for later analysis. Single units were isolated from recorded multi-unit signals using SpikeSorter software (Swindale and Spacek, 2014; Swindale et al., 2017), and only clearly separated units were included in our analysis. Earlier, datasets were sorted using Plexon Offline Sorter software. But after the availability of software, SpikeSorter, I used it for sorting the rest of the data used in this thesis. SpikeSorter requires comparatively much less manual intervention and subjective judgements. It takes advantage of the geometry of the electrode sites - it accepts user-provided electrode site maps, and then compares spikes recorded across multiple nearby sites. An issue with longer recording durations (~1 hour or more) is that the shapes of spike waveforms can change over time due to relative displacement between neurons and electrode sites, when the electrode slips slightly through the brain tissue. SpikeSorter has an ability to track these changes over time, and recommends merging of units that would otherwise be classified as separate units.

1.11 Thesis Aims

In this thesis, I address three specific research aims as described in detail in Chapters 2, 3 and 4. Here I briefly describe these aims and the rationale behind them.

The aim of Chapter 2 is to understand how spatial nonlinearities in receptive fields emerging at a subcortical level are utilized at early cortical stages. Recent evidence has suggested that the nonlinear Y-pathway arising in the retina could contribute to the cortical processing of luminance- and texture-defined boundaries (Demb et al., 2001b; Rosenberg et al., 2010). To understand how these signals from the Y-pathway are pooled together in the cortex, I characterize receptive fields of cortical neurons using the same stimuli used to characterize subcortical Y-cells. Furthermore, I perform model simulations to test whether a novel neural circuit model utilizing cortical Y-like cells could account for known cortical receptive field properties for luminance- and texture-defined boundary stimuli.

In Chapter 3, I estimate biologically interpretable quantitative models of LGN (lateral geniculate nucleus) neuronal receptive fields that can predict responses to novel stimuli. Visual signals undergo multi-stage processing in the retina before reaching the LGN. In particular, the rectifying nonlinearity between bipolar cells and ganglion cells enables Y-type cells to encode both texture and luminance information in the visual scene. However, most previous studies modeled subcortical receptive fields as linear filters, and hence could not capture this important spatial nonlinearity. Therefore in this study, I model receptive fields as multi-stage convolutional neural networks.

In Chapter 4, I quantify the contributions of ON and OFF subcortical pathway inputs to cortical simple cell receptive fields. ON and OFF pathways emerge in parallel in the retina, with very little interaction between them until they reach visual cortex. These pathways are combined in cortical simple cells to build receptive fields that are selective for boundary orientation. Relative contributions of these complementary signals to individual receptive fields could affect functional properties of simple cells, such as selectivity for texture and motion. Here I estimate

subunit receptive field models of simple cells using a GLM approach with ON and OFF preprocessing inputs, to measure how ON and OFF signals are integrated within receptive fields over space and time.



Figure 1.1 : First-order and second-order boundaries in visual scenes.

Natural scene where the boundary between the tree and grass is formed by luminance (first-order) as well as texture (second-order) change, while the boundary between the grass and its reflection in the water is formed by only a difference in texture contrast (second-order). (Adapted from Li et al, 2014)

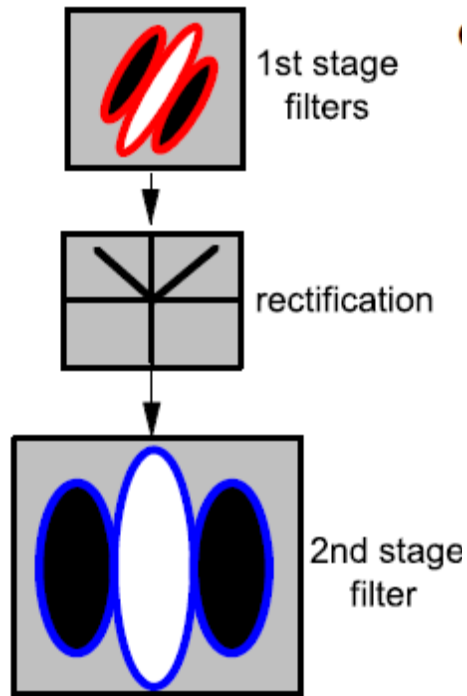


Figure 1.2 : Filter-Rectify-Filter (FRF) model for detecting second-order boundaries.

Two-stage FRF model for detecting second-order boundaries formed by changes in texture, or contrast. The first-stage filters detect fine texture elements in the image. The outputs of these filters are rectified and then summed by a second-stage filter of much coarser scale, that can detect changes in texture properties. This model can detect texture boundaries with vertical orientation corresponding to the second-stage filter. (Adapted from Li et al, 2014.)

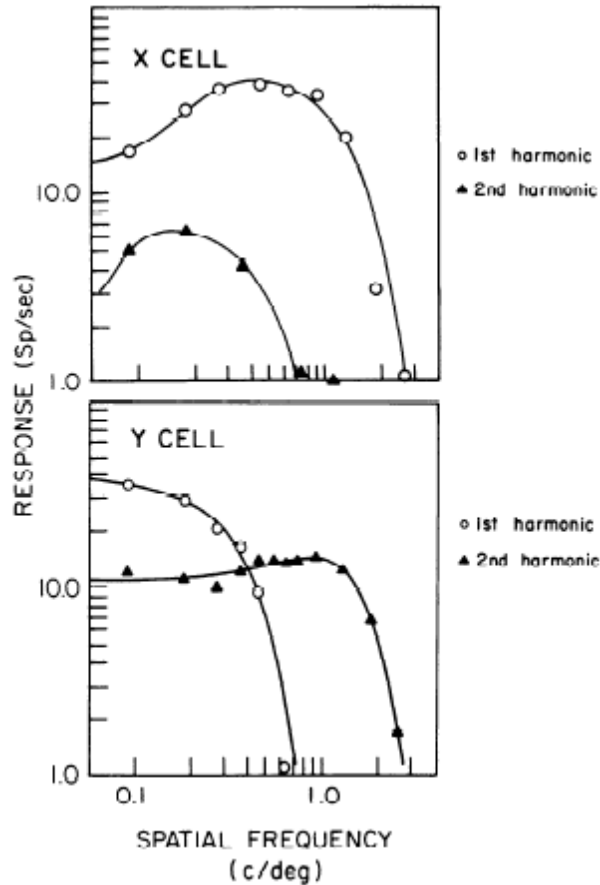


Figure 1.3 : Spatial frequency tuning for typical X- and Y-type cells measured with luminance gratings.

X-type cells respond linearly at all spatial frequencies, with 1st harmonic response much greater than 2nd harmonic response. The 2nd harmonic response for X cells arises from their final output rectification. Y-type cells respond linearly at low spatial frequencies with 1st harmonic response greater than 2nd harmonic. However, at high spatial frequencies, outside the luminance passband, Y cells respond nonlinearly at 2nd harmonic, with no 1st harmonic response. (Adapted from So and Shapley, 1981).

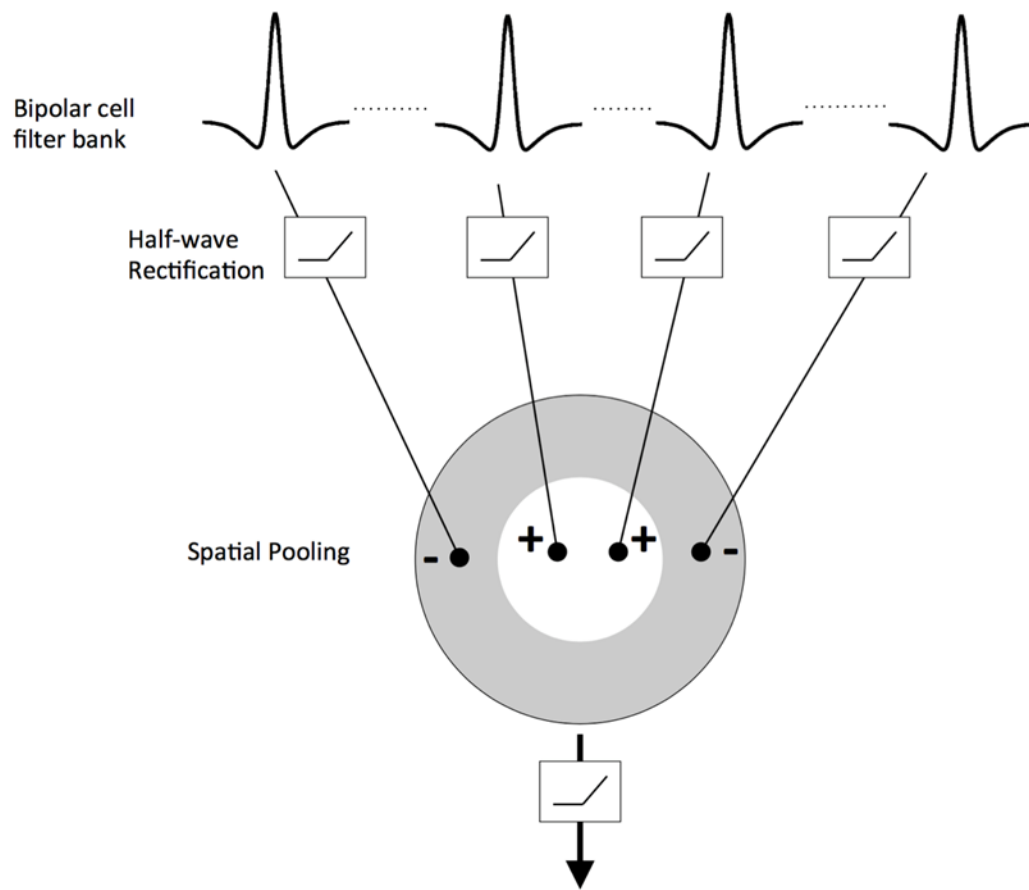


Figure 1.4 : Subunit receptive field model of a Y-type retinal ganglion cell.

Responses of Y cells are modeled as a two-stage filter model. The first stage consists of a bank of small circular subunit filters, corresponding to bipolar cells. Subunit outputs are half-wave rectified and then pooled in a centre-surround spatial layout.

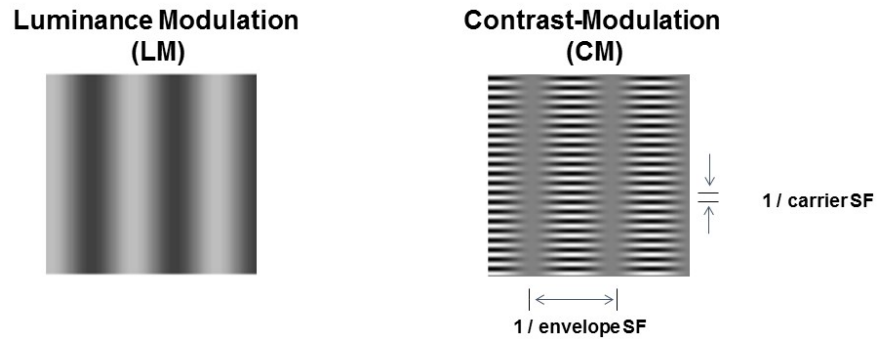


Figure 1.5 : Luminance Modulation and Contrast Modulation gratings

Two types of grating stimuli used for characterising receptive field properties of neurons. A luminance grating is constructed by sinusoidal modulation of luminance along one orientation. Contrast modulation gratings are constructed by modulating contrast of a high spatial frequency carrier grating by a low spatial frequency envelope grating.

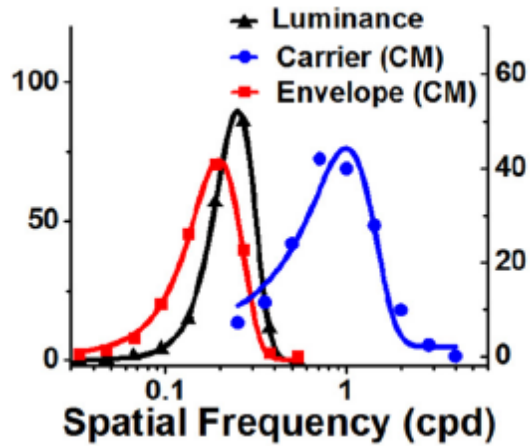


Figure 1.6: Spatial frequency tuning of a model cortical neuron.

A significant fraction of neurons in early visual cortex are selective for spatial frequency of luminance gratings (LM) as well as the carrier and envelope of contrast modulation (CM) gratings, as shown in this model simulation. Neurons are tuned for similar spatial frequencies for luminance and envelope of CM gratings, but very high spatial frequencies for carrier gratings. (Adapted from Gharat and Baker, 2017).

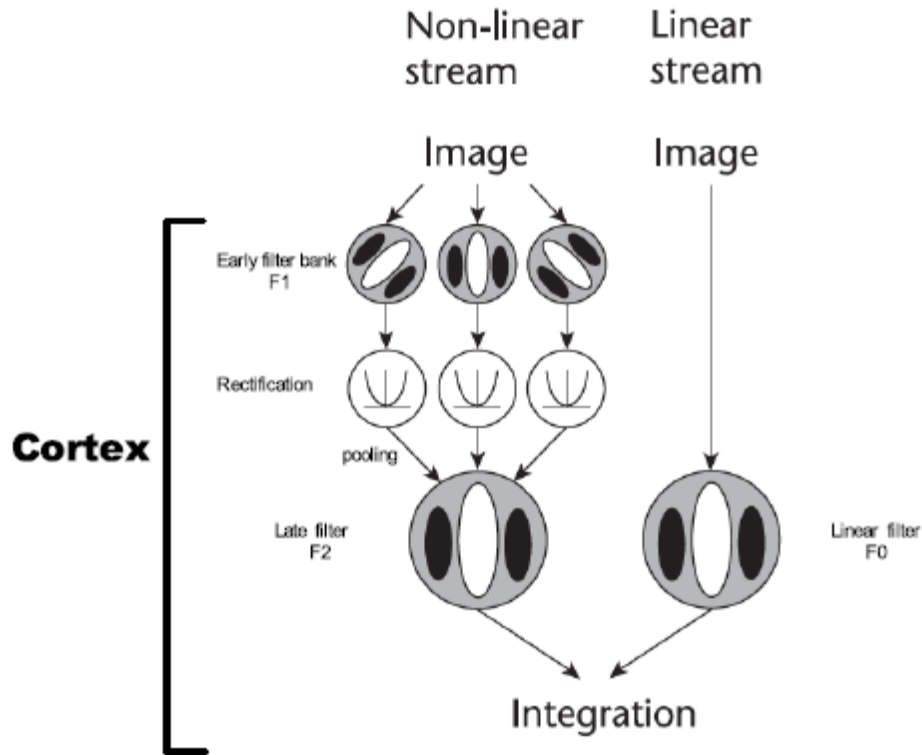


Figure 1.7 : Two-stream processing model for first- and second-order boundaries.

A Linear stream processes luminance with a linear filter F0. A non-linear stream processes 2nd-order stimuli with two-stage filtering. A first-stage filter bank F1 detects high spatial frequency texture and then the output of each filter is rectified and pooled by a late filter F2. The outputs of filters F0 and F2 are summed linearly by an Area 18 neuron. (Adapted from Mareschal and Baker, 1998a).

2

In Chapter 2, using multi-electrode recordings and sinewave grating stimuli, I characterize receptive fields in early visual cortex to understand how cue-invariant receptive fields are built from subcortical inputs. Here I demonstrate the presence of a significant fraction of neurons in cat Area 18 with nonlinear receptive fields like those of subcortical Y-type cells. These neurons have receptive field properties intermediate between subcortical Y cells and cortical orientation selective neurons that respond in a cue-invariant manner to luminance- and contrast-defined boundaries. These Y-like cells are strong candidates for building cue-invariant orientation-selective neurons. Furthermore, I present a novel neural circuit model that pools such ON- and OFF-center Y-like neurons in an unbalanced “push-pull” manner, to generate orientation-selective cue-invariant receptive fields. This chapter has been adapted from Gharat A and Baker CL (2017) Nonlinear Y-like receptive fields in the early visual cortex: An intermediate stage for building cue-invariant receptive fields from subcortical Y cells. *The Journal of Neuroscience*. 37(4): 998-1013.

CHAPTER 2

Nonlinear Y-like receptive fields in the early visual cortex: An intermediate stage for building cue-invariant receptive fields from subcortical Y cells

2.1 Abstract

Many of the neurons in early visual cortex are selective for orientation of boundaries defined by first-order (luminance) as well as second-order (contrast, texture) cues. The neural circuit mechanism underlying this selectivity is still unclear, but some studies have proposed that it emerges from spatial nonlinearities of subcortical Y cells. In order to understand how inputs from the Y cell pathway might be pooled to generate cue-invariant receptive fields, we recorded visual responses from single neurons in cat Area 18 using linear multi-electrode arrays. We measured responses to drifting and contrast-reversing luminance gratings as well as contrast-modulation gratings. We found that a large fraction of these neurons have non-oriented responses to gratings, similar to those of subcortical Y cells - they respond at the second harmonic (F2) to high spatial frequency contrast-reversing gratings and at the first harmonic (F1) to low spatial frequency drifting gratings (“Y-cell signature”). For a given neuron, spatial frequency tuning for linear (F1) and nonlinear (F2) response is quite distinct, similar to orientation-selective cue-invariant neurons. Also, these neurons respond to contrast modulation (CM) gratings with selectivity for the carrier (texture) spatial frequency and, in some cases, orientation. Their receptive field properties suggest that they could serve as building blocks for orientation selective cue-invariant neurons. We propose a circuit model that combines ON- and OFF-centre

cortical Y-like cells in an unbalanced push-pull manner, to generate orientation selective cue-invariant receptive fields.

2.2 Introduction

A substantial fraction of neurons in the early visual cortex (Area 18) of cats respond in a cue-invariant manner to boundaries formed by first-order (luminance) or second-order (contrast, texture, motion) differences (Zhou & Baker, 1993; Tanaka & Ohzawa, 2006; Song & Baker, 2007; Gharat & Baker, 2012). Recently, neurons in the early visual cortex (V2) of nonhuman primates were also shown to respond cue-invariantly to luminance- and contrast-defined boundaries (Li et al., 2014), with spatial selectivity to the carrier (texture) and envelope (modulator) of contrast boundaries very similar to previous findings in cat Area 18 (Mareschal & Baker, 1998a, Mareschal & Baker, 1999). Comparison of these primate V2 results with human psychophysics (Sutter et al., 1995; Dakin & Mareschal, 2000) suggests that these neurons could be the neural substrate for perception of 2nd order boundaries.

However the neural circuit underlying these highly specialized receptive fields, with cue-invariant selectivity for first- and second-order cues early in the visual pathway, is still unclear. The demonstration of carrier orientation-selectivity in cat Area 18 cells suggested a cortical substrate for carrier processing (Mareschal & Baker, 1998a). More recent evidence suggests that cortical neurons could achieve such receptive field properties by pooling inputs from the subcortical Y-pathway (Demb et al., 2001a, Rosenberg et al 2010; Rosenberg & Issa, 2011). Due to spatial nonlinearities, Y cells respond to first-order as well as second-order cues with selectivity for carrier (texture) spatial frequency and orientation similar to cortical neurons (Rosenberg et al 2010). Thus carrier processing for encoding 2nd order cues could take place in

the retina, with the cue-invariant envelope selectivity arising in the cortex from the Y-cell input to the cortical neurons. A similar mechanism is also plausible in the primate visual system, since the parasol and epsilon cells in the retina also have Y-like receptive field properties (Crook et al., 2008a,b; Petrusca et al., 2007). This challenges previous ideas that 1st and 2nd order cues are processed independently (Smith & Ledgesway, 1997) and that 2nd order cues are encoded in higher extrastriate areas (El-Shamayleh & Movshon, 2011; Smith et al., 1998; Pan et al., 2012; An et al., 2014). Previous studies have extensively analyzed the pooling of subcortical X-pathway inputs in cat Area 17 to generate simple cell (linear, Gabor-like) receptive fields with a "push-pull" combination of On- and Off-centre cells (Ferster, 1988; Hirsch et al., 1998; Martinez et al., 2005). However Area 18 receives a majority of its LGN input from the nonlinear Y-pathway, and it is unclear how these inputs are combined to generate receptive fields with precise selectivity for 1st as well as 2nd order cues.

In order to understand the cortical circuitry for second-order processing in the early visual pathway, we recorded single-unit activity from cat Area 18 using multi-electrode arrays that can span all cortical layers. To reduce possible sampling biases due to manual searching with bar-shaped stimuli, we employed a battery of grating measurements together with post-hoc spike sorting. We found that a significant fraction of Area 18 neurons have receptive field properties similar to LGN Y cells, suggesting that these neurons form an intermediate stage between subcortical Y cells and orientation selective cue-invariant neurons. Finally, we propose a cortical neural circuit model that combines signals from the ON and OFF cortical Y-like cells to generate receptive fields selective for orientation of both 1st and 2nd order boundaries in a cue-invariant manner. Unlike the balanced push-pull model proposed for Area 17 neurons, this model

has imbalanced push-pull, for example with ON inputs exerting a stronger effect than OFF inputs.

2.3 Materials & Methods

2.3.1 Animal Preparation

Our experimental procedures are explained in detail in our previous study (Gharat & Baker, 2012), and here are described briefly. Anesthesia was induced in adult cats of either sex with isoflurane/oxygen (3-5%) inhalation. Following intravenous cannulation, subsequent surgical anesthesia was obtained with i.v. propofol. A craniotomy and duratomy were performed (H-C A3/L4) for electrode placement in Area 18 (Tusa et al., 1979). During recording the animal was anesthetized and paralyzed with infusion of propofol ($5.3\text{mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$), fentanyl ($7.4\mu\text{g} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$) and gallamine triethiodide ($10\text{mg} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$), and a mixture of O₂ and N₂O (30:70 ratio) was delivered through a ventilator. Heart rate, EEG, body temperature, end-tidal CO₂, blood oxygen, and airway pressure were monitored, with adjustments in ventilator stroke volume and anesthesia level as indicated. Neutral contact lenses and artificial pupils were positioned, and spectacle lenses of appropriate power were selected using a slit retinoscope to bring visual stimuli in focus. Back-projection of the optic discs onto a tangent screen allowed estimation of area centralis positions. All of these procedures were approved by the Animal Care Committee of McGill University and are in accordance with the guidelines of the Canadian Council on Animal Care.

2.3.2 Visual Stimuli

Visual stimuli were presented on a gamma-corrected CRT monitor (NEC FP1350, 20", 640×480 pixels, 75 Hz, 36 cd/m²) at a viewing distance of 57 cm. Stimuli were generated with a Macintosh computer (MacPro, 2.66 GHz, 6 GB, MacOSX 10.6.8) using custom Matlab software with the Psychophysics Toolbox (Brainard 1997; Kleiner et al, 2007). Drifting sinusoidal luminance gratings with a Michaelson contrast of 30% were used to measure neurons' linear spatial frequency and orientation tuning.

Neurons were classified as X- or Y-like (see below) using contrast-reversing gratings, with a higher contrast (70%) since nonlinear responses are often lower in amplitude. These gratings were also used to measure spatial frequency and orientation tuning, and spatial phase dependence, of nonlinear responses. In some cases responses were also obtained to contrast modulation (CM) stimuli, composed from a stationary high spatial frequency sinusoidal grating (carrier, 70% contrast) whose contrast was modulated by a drifting low spatial frequency sinusoidal grating (envelope, 100% modulation depth).

2.3.3 Extracellular recording

Recordings were performed using multielectrodes (NeuroNexus), in most cases 32 channel (A1x32) linear arrays, but also sometimes 16 channel (A1x16) linear arrays and 16 channel (A4x4) tetrodes. Raw data signals were acquired with a Plexon Recorder (3Hz-8kHz, sampling rate 40kHz). Signals from a selected channel with visually responsive single- or multi-unit activity was used to guide the recording protocol. Spike times detected on this channel with a window discriminator were collected through a lab interface (ITC-18, Instrutech) and analyzed

online to get tuning curves and PSTHs (peristimulus time histograms). 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.

Manually controlled visual stimuli (bars, spots) were used to determine the approximate receptive field location for multi-unit activity on the monitored channel, so as to position the stimulus display to activate cells driven by the dominant eye (the non-dominant eye was occluded) - this procedure, rather than searching for single cells with bar-shaped stimuli, helped ensure a less biased sample including neurons lacking orientation selectivity. We attempted to insert multielectrodes perpendicular to the brain surface, so usually receptive field locations of neurons recorded on the other channels also fell on the display, enabling the simultaneous recording of useful visual responses of neurons on most channels. Drifting sinusoidal luminance gratings were presented to measure spatial frequency and orientation tuning. Each stimulus condition was interleaved with other conditions randomly, and repeated 5-10 times. Contrast-reversing luminance gratings were then presented to measure nonlinear spatial summation. For all the spatial frequencies tested, either grating spatial phase or orientation was also varied. In some cases we also measured responses to contrast modulation (CM) gratings. Multi-unit activity across all channels during the experiment was analyzed to check if recording sites were visually responsive. Once the recording protocol was finished, sometimes it was repeated on the non-dominant eye depending upon quality of spike amplitude across channels.

2.3.4 Analysis

Spike waveforms were carefully classified from the recorded data to isolate signals from single units, using Offline Sorter (version 3.3.3, Plexon) in earlier experiments, and later,

Spikesorter (Swindale & Spacek, 2014) for sorting multichannel electrode data. On some datasets sorting was done with both types of software and the results obtained were very similar. Only clearly sorted units were used for further analysis.

Responses of neurons to grating stimuli were accumulated as PSTHs (bin width 13.3 ms, duration of each frame), which were used to calculate first and second harmonic responses. Neurons were classified as simple or complex type cells by measuring the ratio of first harmonic modulation amplitude to mean, in response to the neuron's optimal drifting luminance grating (Skottun et al. 1991). For orientation and spatial frequency tuning curves, first harmonic response rate was used for simple type cells while mean response rate was used for complex type cells.

Neurons' orientation selectivity was characterized with an “Orientation Bias” (OB) index (Leventhal et al 2003):

$$OB = \frac{\left| \sum_k R_k \exp(i2\theta_k) \right|}{\sum_k R_k}$$

where R_k represents spontaneous-subtracted neuronal response at orientation θ_k . Orientation bias values range from zero (isotropic tuning) to unity (sharp tuning).

The degree to which neurons exhibited a binocular vs monocular response was summarized with a “binocularity index”, defined as the ratio of average response to optimal drifting gratings in the non-dominant eye to that in the dominant eye. The binocularity index ranges from zero (perfectly monocular) to unity (perfectly binocular).

To classify a neuron as X-like or Y-like we used a “nonlinearity index” (Hochstein & Shapley, 1976), defined as the maximum of the ratio of second-harmonic (F2) to first harmonic (F1) response. If at any spatial frequency, a neuron's second-harmonic response was significantly greater than its first harmonic component, it was classified as Y-like, otherwise as X-like. Note that only simple cells ($AC/DC > 1$) were further classified as X-like or Y-like, since complex cells respond nonlinearly (F2) within their luminance passband and their first harmonic (F1) is very weak or absent. Spatial frequency tuning curves of linear (F1) and nonlinear (F2) responses were fit with a Gaussian function (DeAngelis et al 1994):

$$R(sf) = ke^{-(sf-SF_{opt})/\alpha)^2} + R_o$$

where k , SF_{opt} , α are free parameters, R_o is spontaneous activity and R is neuronal response at spatial frequency sf , with SF_{opt} taken as the optimal spatial frequency.

Pearson's correlation coefficient between optimal linear and nonlinear spatial frequency was employed to assess any relationship between a neuron's spatial tuning for linear and nonlinear responses. The circular correlation (Berens, 2009) coefficient was used to assess the relationship between neurons' optimal orientation for drifting and contrast reversing gratings.

2.4 Results

2.4.1 Non-oriented receptive fields in cat Area 18.

Previous single-unit studies of cat Area 18, including those in our lab, have primarily reported orientation-selective neurons. However more recently, using multi-channel microelectrodes with which we simultaneously record spikes from multiple neurons and analyze

the data post hoc (see Methods), a significant fraction of neurons were found to have non-oriented receptive fields (Talebi & Baker, 2016).

Fig 2.1A shows example tuning curves of orientation-selective (left) and isotropic neurons (right), measured with drifting luminance gratings at each neuron's optimal spatial frequency - these two neurons were simultaneously recorded from the same site on a multielectrode. We quantified each neuron's orientation selectivity with an "Orientation Bias" (OB) index (see Methods) which ranged from zero (isotropy) to unity (perfect selectivity). Neurons were classified as "non-ori" cells if $OB < 0.2$, which is the range found for LGN neurons (Rosenberg et al, 2010). The tuning curves in Fig 2.1A show examples of neurons classified as orientation-selective (left, OB 0.54) and non-ori (right, OB 0.11).

Fig 2.1B shows an example of orientation tuning curves of neurons recorded simultaneously from a 32-channel linear array with recording sites separated by 100 μm . The array was inserted approximately perpendicular to the surface of the dura and lowered until most of the channels had spiking activity, so as to encompass all the cortical layers and to be approximately aligned with the columnar architecture. However due to curvature of the brain beneath the dura, such electrode penetrations were not necessarily confined within an orientation column. The penetration shown in Fig 2.1B is an example of an evidently somewhat oblique penetration, traversing different orientation columns. Note the span of depths with sorted neurons is 2.7 mm (28 channels), exceeding the anatomical thickness of grey matter in Area 18 (ca 2 mm - Tusa et al., 1979). Note that non-ori neurons (labeled with asterisks in Fig 2.1B) do not appear to be confined to particular layers, but rather are present at various depths spanning the gray matter, and are intermixed with orientation-selective neurons. This is consistent with Talebi & Baker (2016), who found neurons with non-oriented receptive field maps dispersed across all

depths of Area 18. Fig 2.1C shows the distribution of orientation selectivity (OB values) of all the neurons that were recorded - more than one-third (84 out of 208) were classified as non-ori. This histogram does not show a bimodal distribution indicating non-ori neurons as a separate class, which might seem in contradiction to the bimodal distribution seen in the similar histogram in Talebi and Baker (2016) of OB values of Area 18 simple cells (their Fig. 6A). However note that here we calculated OB values from orientation tuning curves constructed by measuring responses at only 13 discrete orientations (separated by 30 degrees), while Talebi & Baker (2016) measured OB values based on responses at a much larger number of orientations, simulated on a spatiotemporal receptive field map estimated by system identification. Their approach leads to much smoother tuning curves (see Fig 2D in Talebi & Baker, 2016) and much lower OB values. However the classical method of using responses to gratings can give high OB values due to limited sampling. So there is a strong possibility that even in our data non-ori cells might form a separate class from oriented receptive fields, but we fail to see it due to the limited sampling of orientations.

To assess whether these non-ori neurons behave like classical simple or complex type cells, we measured their AC/DC (modulated/mean response) ratio (Skottun et al, 1991 - see Methods) for responses to optimized drifting gratings. The distribution of AC/DC ratios of non-ori neurons (Fig. 2.2B) shows that most (75/84) are simple type (ratios greater than unity). This suggests that most non-ori neurons have isotropic receptive fields with distinct concentric ON and OFF regions similar to lateral geniculate nucleus (LGN) X and Y cells. We also find a few complex-like non-ori neurons (AC/DC ratios less than unity) - these could be receiving input from the W pathway, some of whose neurons have mixed On- and Off-responding receptive fields (Stone et al., 1979).

In some cases ($n = 38$) we also assessed the degree of binocular response of the non-ori neurons, by separately measuring responses to each eye and taking their ratio as a “binocularity index” (see Methods). A purely monocular neuron should have an index close to zero while a perfectly binocular neuron would have an index of unity. A histogram of these indices (Fig 2.2D) shows that most of the non-ori neurons are monocular (25/38), but about one-third are binocular with index values as high as 0.93. A scatterplot (Fig. 2.2C) comparing binocularity indices and AC/DC ratios shows that there is no relationship between these two parameters ($R = 0.0249$, $p = 0.882$, $n = 38$).

One might wonder if these non-ori neurons are actually terminals of LGN afferent fibers. However this is unlikely because we find them across all cortical depths (e.g. Fig.2.1B), whereas LGN inputs terminate in layers 4 and 6 (LeVay and Gilbert, 1976). In addition, some of the non-ori cells are binocular (Fig. 2.2), which is characteristic of visual cortex (Hubel and Wiesel, 1962). Another potential concern is that poor spike sorting might inadvertently combine signals from several neurons with differing preferred orientations, giving an apparent lack of orientation tuning. Fig 2.1D shows sorted raw spike waveforms of six example non-ori neurons recorded in the penetration shown in Fig 2.1B. These sorted waveforms are clearly from single-units, and hence the broad orientation tuning of these non-ori neurons is not due to contamination from multi-unit activity. Furthermore, most of these cells give simple type (modulated) responses (Fig. 2.2A), whereas a mixture of neurons tuned to different orientations would give complex-like (unmodulated) responses.

2.4.2 Y-like spatial nonlinearities of non-ori receptive fields

Area 18 in the cat receives a strong direct input from the LGN, predominantly from Y cells, with much less input from X and W cells (Dreher et al 1980; Stone & Dreher, 1973). Since these cortical non-ori neurons have orientation tuning similar to LGN cells, it seems likely that most of them receive direct or indirect input from LGN Y cells. Hence we hypothesized that most cortical non-ori neurons should show the nonlinear spatial summation that is characteristic of LGN (and retinal) Y cells. Similar to previous studies of Y type cells (Hochstein & Shapley, 1976; Demb et al., 2001a; Rosenberg et al 2010; Crook et al 2008a), we measured spatial properties of these neurons ($n=44$) using drifting and contrast-reversing gratings.

Both X and Y type cells respond to drifting sinusoidal gratings at their fundamental temporal frequency (F1), indicative of linear processing. With contrast-reversing gratings, X cells also respond linearly (F1), but Y cells give second harmonic (F2) responses (indicative of strong nonlinearity) at high spatial frequencies. We classified a neuron as Y-like if its second harmonic response component was significantly greater than the first harmonic to a contrast-reversing grating at any of the series of spatial frequencies tested (formalized as “nonlinearity index”, see Methods) - otherwise it was classified as X-like (Hochstein & Shapley, 1976).

Spatial frequency responses for a typical Y-like non-ori neuron are shown in Fig 2.3A. This neuron responded linearly (F1, black) to drifting gratings, with tuning for low spatial frequencies. But to contrast-reversing gratings the neuron responded nonlinearly (F2, blue), at high spatial frequencies outside the linear SF tuning range. This combination of results is the classic "Y-cell signature" (Hochstein & Shapley, 1976) for retinal and LGN Y cells.

Fig 2.3B shows PSTH responses of this neuron to contrast-reversing gratings at two spatial frequencies, one within the linear SF range and the other in the nonlinear range. At a low

SF (0.1 cpd, left) the neuron responded at the same temporal frequency as the grating (4Hz), and this response depended on the spatial phase of the grating relative to neuron's receptive field, with a minimum ("null") phase - all indicative of linear spatial summation. But at a higher SF (0.53 cpd, right) the neuron gave a frequency-doubled response (8 Hz) that was phase-independent, indicating nonlinear spatial summation. Fig 2.3C plots the first and second harmonic values calculated from the PSTHs in Fig 2.3B. The first harmonic values depend on spatial phase, with a clear null phase repeated in 180° intervals, but the second harmonic values are approximately constant with phase. Thus this neuron showed all the spatial characteristics of a typical Y cell (Hochstein & Shapley, 1976). The distribution of spatial nonlinearity indices for the simple type non-ori neurons (Fig 2.3D) were predominantly Y-like (36/44), but there were a few (8/44) X-like cells as well.

2.4.3 Linear and nonlinear spatial frequency relationships of Y-like cortical neurons

As shown in the previous section most of the cortical non-ori neurons have distinct linear and nonlinear SF tuning similar to those of retinal and LGN Y cells. Consequently it seems a likely possibility that Area 18 non-ori neurons may be involved in cortical processing of second-order as well as first-order (luminance) stimuli. To further explore this possibility we measured spatial tuning properties of non-ori neurons, to compare with previously studied orientation-selective CM-responsive cortical neurons (Mareschal & Baker, 1999). Fig 2.4A-F shows linear and nonlinear SF tuning plots of six non-ori cells - each has bandpass tuned nonlinear response (F2, blue) outside, and well above, the luminance passband (F1, black). We fitted the data points with Gaussian functions (see Methods) to derive optimal SF values for linear (F1) and for nonlinear (F2) tuning. A scatterplot of optimal SFs for linear vs. nonlinear responses (Fig 2.4G)

shows that optimal SFs for F2 are always substantially higher than for F1, with most of the neurons' values scattered around a 10:1 ratio line, and a weak correlation ($r = 0.34$) between optimal SFs for F1 and F2 for a given neuron. The distribution of F2/F1 ratios of optimal SFs (Fig 2.4H) shows ratios ranging from 4.6 to 28, with mean value of 11.3 (median = 8.7).

A previous study (Mareschal & Baker, 1999) of orientation-selective neurons in Area 18 with contrast modulation gratings found similar results for linear and nonlinear spatial tuning. In that study, the ratio of optimal SF for the carrier of CM gratings (nonlinear) and drifting luminance gratings (linear) varied from 5 to 30, with mean around 10. Similar ratios were also observed for CM response tunings in macaque V2 neurons (Li et al, 2014). Thus cortical non-ori neurons have a similar relationship between linear and nonlinear SF tuning to that of orientation-selective, CM-responsive neurons.

2.4.4 Orientation tuning of linear and nonlinear responses of Y-like cortical neurons

Some Area 18 neurons show pronounced orientation tuning for the high SF carrier of contrast modulation gratings (Mareschal & Baker, 1998a), which is independent of their orientation tuning for drifting luminance gratings. Hence it was previously thought that receptive field subunits that detect the carrier are cortical, for example orientation-selective Area 17 neurons having high SF selectivity. However Rosenberg et al (2010) showed that even though LGN Y cells exhibit little or no selectivity for orientation of drifting gratings, some of them show pronounced orientation tuning for the carrier of CM gratings as well as for the nonlinear response to contrast-reversing high SF gratings. Thus carrier orientation selectivity of CM-responsive Area 18 neurons might be inherited from afferent LGN Y cells. Therefore we measured orientation tuning of nonlinear (F2) responses of Y-like cortical non-ori neurons to see

whether some of them exhibit similar narrow tuning as found for cortical oriented CM-responsive cells (Mareschal & Baker, 1998a; Rosenberg et al 2010).

Fig 2.5A-B shows orientation tuning curves for the linear (F1, black) and nonlinear (F2, blue) responses of two non-ori Y-like neurons. The nonlinear (blue) tuning curves are symmetric because responses were collected for orientations from 0 to 180 deg, and the responses were then mirrored about the origin. For the neuron in Fig 2.5A the nonlinear response (blue) is not tuned (OB = 0.12) for orientation of contrast reversing gratings. For comparison, the same neuron's responses to drifting gratings (black) are also shown - note that these linear responses have very small orientation bias (OB = 0.02), and are not direction selective. On the other hand, for the neuron in Fig 2.5B the nonlinear response (blue) is sharply tuned (OB = 0.47) for orientation while the linear response is not tuned (OB = 0.02). The scatterplot in Fig 2.5C shows the orientation bias (OB) values of neurons' nonlinear against linear responses in this sample (n = 16). The linear responses (abscissa) all have OB values less than 0.2, as expected for non-ori neurons. However for the nonlinear responses (ordinate), some of these neurons (6/16) have substantial orientation selectivity (OB values greater than 0.2). We assessed the possibility of a systematic relationship between optimal orientation for linear (F1) responses and nonlinear (F2) responses. There was no significant circular correlation (Berens, 2009) between these optimal orientations for a given neuron ($R = 0.0075$, $p = 0.9719$, $n = 16$). The histogram in Fig 2.5D shows differences in preferred orientation for linear and nonlinear responses. The difference in preferred orientation for most (14/16) neurons was greater than 30 degrees. Thus in this regard orientation tuning for nonlinear responses of cortical Y-like non-ori neurons is similar to that for LGN Y cells (Rosenberg et al., 2010) and for CM carrier tuning of cortical orientation-selective neurons (Mareschal & Baker, 1998a).

2.4.5 Responses of Y-like cortical neurons to second-order stimuli

Previous studies (Demb et al 2001b; Rosenberg et al., 2010) demonstrated that retinal and LGN Y cells respond to contrast modulation (CM) gratings in addition to conventional luminance modulation gratings, suggesting that the Y-like non-ori cortical neurons might also be CM-responsive. Fig 2.6A-B shows example snapshot images of contrast modulation (CM) gratings with a vertically oriented low spatial frequency envelope that modulates the contrast of horizontal carrier gratings, the latter set at a lower carrier spatial frequency on the left (A), and higher on the right (B). For measuring responses to CM gratings we fixed the spatial frequency of the envelope at or near the optimal luminance SF (F1), and tested a series of carrier spatial frequencies outside the neuron's luminance passband.

Fig 2.6C-H shows six non-ori neurons' responses to CM gratings (orange) at a series of carrier SFs outside their luminance passbands (F1, black). These neurons show bandpass selectivity for the carrier of contrast modulation gratings (orange), which is similar to their nonlinear SF tuning (F2, blue). The scatterplot Fig 2.6I shows that optimal spatial frequency for the carrier is very similar to that for nonlinear (F2) tuning ($R = 0.9266$, $p = 0.0079$, $n = 6$). As shown in the scatterplot Fig 2.6J, the spatial frequency bandwidth for the carrier is often narrower than for nonlinear (F2) tuning. Furthermore, the optimal carrier spatial frequencies of these Y-like neurons fall within the same range, ca 0.5 to 2.0 cpd, as those of cortical ori-selective CM-responsive neurons (Zhou & Baker, 1993; Mareschal & Baker, 1999; Rosenberg et al., 2010). These results suggest that responses to CM gratings and nonlinear responses to contrast-reversing gratings are elicited by a common nonlinear mechanism.

2.4.6 A possible cortical circuit utilizing Y-pathway inputs to build cue-invariant receptive fields

We propose a cortical neural circuit model (Fig. 2.7B) that could generate cue-invariant orientation selective receptive fields from responses of cortical Y-like cells. In this model the responses of both ON- and OFF-centre cortical neurons are combined in a "push-pull" manner (Ferster, 1988; Hirsch et al., 1998; Martinez et al., 2005): the ON-subregions of an oriented receptive field receive excitatory input from ON-centre cells and also inhibitory input from OFF-centre cells, and vice versa for the OFF-subregions. It is straightforward to see that this receptive field would be selective for orientation of a luminance boundary. The centers of both ON- and OFF-type Y-cells contain subunits (Demb et al., 2001a) that are excited by increases in texture contrast (i.e. give ON-responses to contrast). Thus if the push-pull between ON- and OFF-pathways is balanced, then the nonlinear responses to texture contrast will cancel out, and the neuron will be unresponsive to contrast boundaries. However an imbalance of the ON- and OFF-pathways (w_{ON} not equal to w_{OFF} in Fig. 2.7B) would enable a contrast boundary response. For example if the ON-pathway is stronger than the OFF-pathway, then in the ON-subregion excitation from ON-subunits will be stronger than inhibition from OFF-subunits, and in the OFF-subregion inhibition from ON-subunits will outweigh excitation from OFF-subunits. Thus the ON-region would respond to an increase in texture contrast while the OFF-region would respond to a decrease in texture contrast - thus the receptive field as a whole would respond well to an oriented, periodic modulation of texture contrast.

In order to demonstrate the tuning properties of this unbalanced neural circuit model, we constructed a computer simulation using a cascade of spatial filters. We modeled Y-cells as summing rectified bipolar cell subunits (Enroth-Cugell and Robson, 1966; Demb et al., 2001a),

as shown in Fig. 2.7A. Outputs of ON- and OFF-type Y-cells were combined in a push-pull manner as shown in Fig. 2.7B. Thus this simulated model contains three filter stages corresponding to bipolar cells (ON- and OFF-centre), Y-cells (ON- and OFF-centre) and a cortical orientation selective simple cell, with half-wave rectification of each stage's responses. We implicitly assume that receptive field properties of Y-type retinal ganglion cells (RGC), LGN neurons and cortical Y-like cells are not significantly different in their spatial receptive field properties. Bipolar cells were modeled as Difference-of-Gaussian (DoG) filters with much wider surrounds compared to their centres, and with centre strengths outweighing surrounds (Dacey et al., 2000). Note that it is crucial for bipolar cell centres to be stronger than their surrounds, to enable a linear response to low spatial frequencies (Dacey et al., 2000; Rosenberg and Issa, 2011). Outputs of these bipolar cell filters were rectified and pooled with DoG weighting, corresponding to retinal ganglion cell (RGC) receptive fields. The centre size of this DoG was set to be several times (x10) larger than the centres of the bipolar cell filters. ON-centre Y-cells were built by pooling ON-centre bipolar cells, and OFF-centre Y-cells by pooling OFF-centre bipolar cells (Demb et al., 1999). Finally, outputs of ON- and OFF-centre Y-cells were summed in a push-pull manner to build a cortical orientation-selective simple cell.

We measured responses of this model with balanced as well as unbalanced push-pull, to luminance-modulation (LM) and contrast-modulation (CM) gratings, in order to compare the model's spatial selectivity to known cortical neurons' selectivity (e.g. Mareschal and Baker, 1998b, Mareschal and Baker, 1999, Li et al., 2014). As shown in Fig. 2.8A, B, C, the model with balanced push-pull responds selectively (spatial frequency and orientation) to LM gratings, but fails to respond to CM gratings having a higher carrier spatial frequency (matched to the bipolar cells' centre size). On the other hand, the model with unbalanced push-pull (Fig. 2.8D, E, F) not

only responds selectively to LM gratings but also to CM gratings. Spatial frequency tuning (Fig. 2.8D, red) for the envelope of CM gratings is similar (though not identical) to that for LM gratings, and the carrier spatial frequency tuning (blue) is well above the luminance passband. In addition, this unbalanced model is also selective for similar orientation of LM gratings (Fig. 2.8E) and the envelope of CM gratings (Fig. 2.8F) - i.e. form cue-invariance. Note that in this scheme carrier selectivity arises from retinal stage (bipolar cell) filters, while the envelope selectivity emerges from cortical stage circuitry.

Many CM-responsive neurons in cat Area 18 have broader envelope orientation tuning, and preference for lower envelope spatial frequencies, compared to their corresponding LM responses (Mareschal & Baker, 1999). In this model scheme these differences arise from the very wide surrounds of the bipolar stage filters compared to their centers (Dacey et al., 2000). These surrounds make Y cells' luminance spatial frequency tuning narrower by dampening responses to low spatial frequencies, thereby shifting the optimal spatial frequency slightly higher. However, for CM gratings at their optimal carrier spatial frequency (scale of bipolar cells' centers), the surrounds of bipolar cells are too wide to detect the carrier. So unlike the case with LM gratings, bipolar surrounds do not contribute to the selectivity for the envelope of CM gratings. This can result in subtle differences in spatial frequency tuning to LM gratings and envelopes of CM gratings in Y-cells, with preference for lower spatial frequencies of CM envelopes compared to LM gratings. These differences can be further increased by nonlinearities (expansive power-law) at the outputs of Y-cells and cortical ori cells, and thus can give a difference in selectivity for luminance gratings and envelopes of CM gratings as shown in Fig. 2.8 D, E, F.

Interestingly, CM-responsive Area 18 neurons show a pronounced selectivity for relative spatial phase between an LM grating and the envelope of a CM grating in a compound LM+CM stimulus (Hutchinson et al., 2016). Therefore we measured model responses to LM+CM stimuli (Fig. 2.9A) for comparison. In the compound stimuli, the spatial frequencies of the LM gratings, and envelope and carrier of the CM gratings, were set to optimal values, and the contrasts of the individual LM and CM gratings were adjusted such that model's responses to them were of equal strength - as in the experimental measurements of Hutchinson et al (2016). Then the model's responses were measured to LM+CM gratings that were added at varying relative phases. When the model is made unbalanced, with $w_{ON} > w_{OFF}$, its response (Fig. 2.9B) is selective for relative phase in the compound stimuli, with strongest response when the LM and CM are in phase (i.e. high luminance of LM aligned with high contrast of CM), in agreement with the results of Hutchinson et al (2016). This behavior arises because in the Y-driven push-pull model, ON- and OFF-subregions for contrast detection are phase-aligned with ON and OFF subregions for luminance detection.

2.5 Discussion

Our results have demonstrated that a large fraction of the sampled population of cat Area 18 neurons have non-oriented Y-like receptive fields, which are present at different cortical depths intermixed with orientation-selective neurons and not evidently clustered in particular layers. These Y-like cortical neurons respond at the second harmonic (F2) to high spatial frequency contrast-reversing gratings and at the first harmonic (F1) to low spatial frequency drifting gratings ("Y-cell signature", Enroth-Cugell and Robson, 1966; Hochstein and Shapley,

1976). A given neuron's SF tunings for linear and nonlinear responses are quite distinct, with on average about an 11-fold greater optimal SF for F2 than for F1. Furthermore, due to these neurons' nonlinearity at high spatial frequencies, they also respond to contrast modulation (CM) patterns (second-order stimuli), with high selectivity for the spatial frequency of the CM carrier grating (texture).

2.5.1 Non-ori cells in cat Area 18

Early visual cortical areas are conventionally described as characteristically comprised of orientation selective receptive fields. However there have been some reports also finding a substantial fraction of LGN-like non-oriented receptive fields in the early mammalian visual cortex. For example, non-ori neurons have been found in primary visual cortex of macaque (Livingston & Hubel, 1984; Ringach, Shapley, Hawken, 2002; Ringach, 2002), mouse (Bonin et al., 2011) and ferret (Chapman & Stryker, 1993), as well as in cat Area 17 (Dragoi et al., 2001; Hirsch et al., 2003). Earlier studies using single channel electrodes and bar-shaped search stimuli in cat Area 18 (Tanaka and Ohzawa, 2006; Ferster and Jagadeesh, 1991; Mareschal and Baker, 1998a) did not report non-oriented receptive fields. But a recent study (Talebi and Baker, 2016) in cat Area 18 using multi-channel electrodes, in conjunction with post hoc data analysis (spike sorting) similar to ours, has reported a large proportion of non-oriented receptive fields estimated using system identification methods. We believe that using multi-electrode arrays with post hoc spike sorting leads to less sampling bias compared to earlier approaches of sampling one neuron at a time with a single channel electrode. Furthermore, with earlier approaches, visual responsiveness of the neuron was typically assessed with moving bars. However we have noticed that a moving bar is not a good stimulus for driving responses from these non-ori neurons – they

are much better driven by flashing spots centered on their receptive fields, due to their comparatively strong surrounds. Thus previous studies might have rarely found such neurons or failed to recognize their visual responsivity.

2.5.2 Nonlinear Y-like spatial summation

Here we have demonstrated that a significant fraction of neurons in early visual cortex of the cat have spatial receptive field properties similar to those of subcortical Y cells. These cortical neurons exhibit both linear and nonlinear spatial response properties, which are tuned for quite distinct spatial frequencies ("Y cell signature" - Hochstein & Shapley, 1976). Optimal spatial frequencies of our non-ori cortical neurons for linear and nonlinear responses (Fig. 2.4) are similar to those reported for retinal and LGN Y cells (Hochstein & Shapley, 1976; So & Shapley, 1979).

Ferster & Jagadeesh (1991) also described harmonic responses of orientation selective simple cells in cat Area 18 to contrast-reversing gratings, and found around half of their neuronal population to have Y-like spatial nonlinearities. However they did not report the presence of non-ori Y-like cells. Spatial selectivity, such as the ratio of preferred spatial frequency of linear and nonlinear responses, of their cell population is similar to the non-ori cells reported here. However, orientation selectivity was not reported for their sample of neurons.

2.5.3 Neural mechanism for building cue-invariant receptive fields

A significant fraction of Area 18 orientation-selective neurons are responsive to both first- and second-order visual stimuli, with the same preferred orientation to both (Zhou & Baker,

1993; Song & Baker, 2006; Gharat & Baker, 2012) - i.e. they are "form cue-invariant" (Albright, 1992). Due to the additional selectivity of some of these neurons to carrier (texture) orientation, it was proposed that the neural substrate for subunits of Area 18 neurons was cortical in origin (Mareschal & Baker, 1998a). However more recent evidence suggests that subcortical Y cells could provide a substrate for the carrier selectivity of cortical neurons (Demb et al 2001; Rosenberg et al, 2010), with the envelope selectivity arising from cortical circuitry. The cortical Y-like neurons that we have described are probably driven by LGN Y cells, and could provide an intermediate stage for building cue-invariant orientation selective receptive fields. Firstly, they have carrier selectivity like cue-invariant neurons, but no orientation selectivity for drifting gratings, like Y cells. Unlike LGN cells, a significant fraction is binocular, which is also the case for some oriented CM-responsive cells (Tanaka & Ohzawa, 2006). Also, these Y-like neurons could provide both excitatory as well as inhibitory inputs to orientation selective neurons - since input from the LGN to the cortex is only excitatory (Alonso et al., 2001), some sort of inhibitory interneuron would be necessary to construct a push-pull architecture for cortical receptive fields. Furthermore, the presence of some of these Y-like neurons in the top cortical layers suggests that they could also be projecting to higher-tier cortical areas along with the orientation selective neurons.

Our model simulations predict that cortical neurons with unbalanced push-pull summation of Y-pathway inputs will be selective for orientation of both luminance and texture boundaries, while the neurons that sum Y-pathway inputs with conventional balanced push-pull will only be selective for luminance boundaries. Furthermore, the unbalanced push-pull model is able to predict previously shown (Mareschal and Baker, 1998b; Mareschal and Baker, 1999; Li

et al, 2014) spatial tuning properties of cortical neurons to LM and CM gratings, including systematic differences in tuning for LM gratings and envelopes of CM gratings.

This unbalanced push-pull model with a Y-pathway input is fundamentally different from the two-stream model proposed earlier (Zhou and Baker, 1993; Mareschal and Baker, 1998a) to explain cortical neurons' tuning properties. In the two-stream model, selectivity for luminance and contrast processing arises separately, and only at the final stage are the outputs from these two streams summed. However in this Y-pathway model, luminance and contrast cues are processed together all along the visual pathway beginning at the retina. In the two-stream model, the neural substrate for subunits that detect fine texture within contrast envelopes was thought to be Area 17 neurons (Mareschal & Baker, 1998a), but in this model it is retinal bipolar cells with rectified outputs. In Area 18, only about half of the orientation selective neurons are responsive to both LM and CM gratings, while the remainder are only responsive to LM but not CM gratings (Zhou and Baker, 1993). This has been accounted for in the previous scheme by the presence or absence of input from a second stream for processing contrast boundaries. However in this scheme a lack of response to contrast modulation would arise from a symmetrical push-pull, or from X- rather than Y-pathway inputs. Future studies could test this idea by assessing whether cortical neurons' CM responsiveness is correlated with their push-pull imbalance of Y-type inputs.

2.5.4 Implications for second-order processing in other mammals

While Y-type retinal ganglion cells were classically described in the cat, they have also been demonstrated in other mammals including mouse (Schwartz et al., 2012) and guinea pig (Demb et al., 2001a). There have been doubts about the presence of a cell type homologous to Y

cells in primates, as previous studies failed to clearly demonstrate “Y-cell signature” responses in retinal parasol cells (Petrusca et al., 2007). However Crook et al. (2008a) clearly demonstrated that macaque retinal parasol cells have Y-like spatial nonlinearities. In view of our results, it seems likely that many of the non-ori neurons in area V1 of both mouse (Bonin et al., 2011) and monkey (Livingston & Hubel, 1984; Ringach, Shapley, Hawken, 2002) might also have Y-like spatial nonlinearities inherited from subcortical Y pathway inputs - this would be a future avenue of investigation.

Li et al (2014) demonstrated that about one-third of neurons in macaque V2 respond to 2nd order stimuli in a form cue-invariant manner. Spatial tuning properties of these neurons to carriers and envelopes of CM gratings were qualitatively very similar to those in cat Area 18 neurons, differing principally in spatial scale. In addition, spatial frequency selectivity of V2 neurons (Li et al, 2014) for drifting luminance gratings and carriers of CM gratings is in a similar range to the spatial selectivity of retinal parasol cells (Crook et al, 2008a) to drifting (F1) and contrast-reversing (F2) gratings, respectively. So it is likely that, similar to cats, Y-like cortical cells are pooled to generate cue-invariant receptive fields in the early visual cortex of primates. Contrary to the view that 2nd order processing takes place in higher visual areas (El-Shamayleh & Movshon, 2011; Smith et al., 1998) and separate from 1st order processing (Larsson et al., 2006; Smith & Ledgesway, 1997), it seems possible that all mammals including primates might have a common mechanism for processing 2nd order stimuli, involving the Y-cell pathway providing an early substrate for carrier-tuning, and cortical circuitry with imbalanced push-pull for cue-invariant envelope tuning.

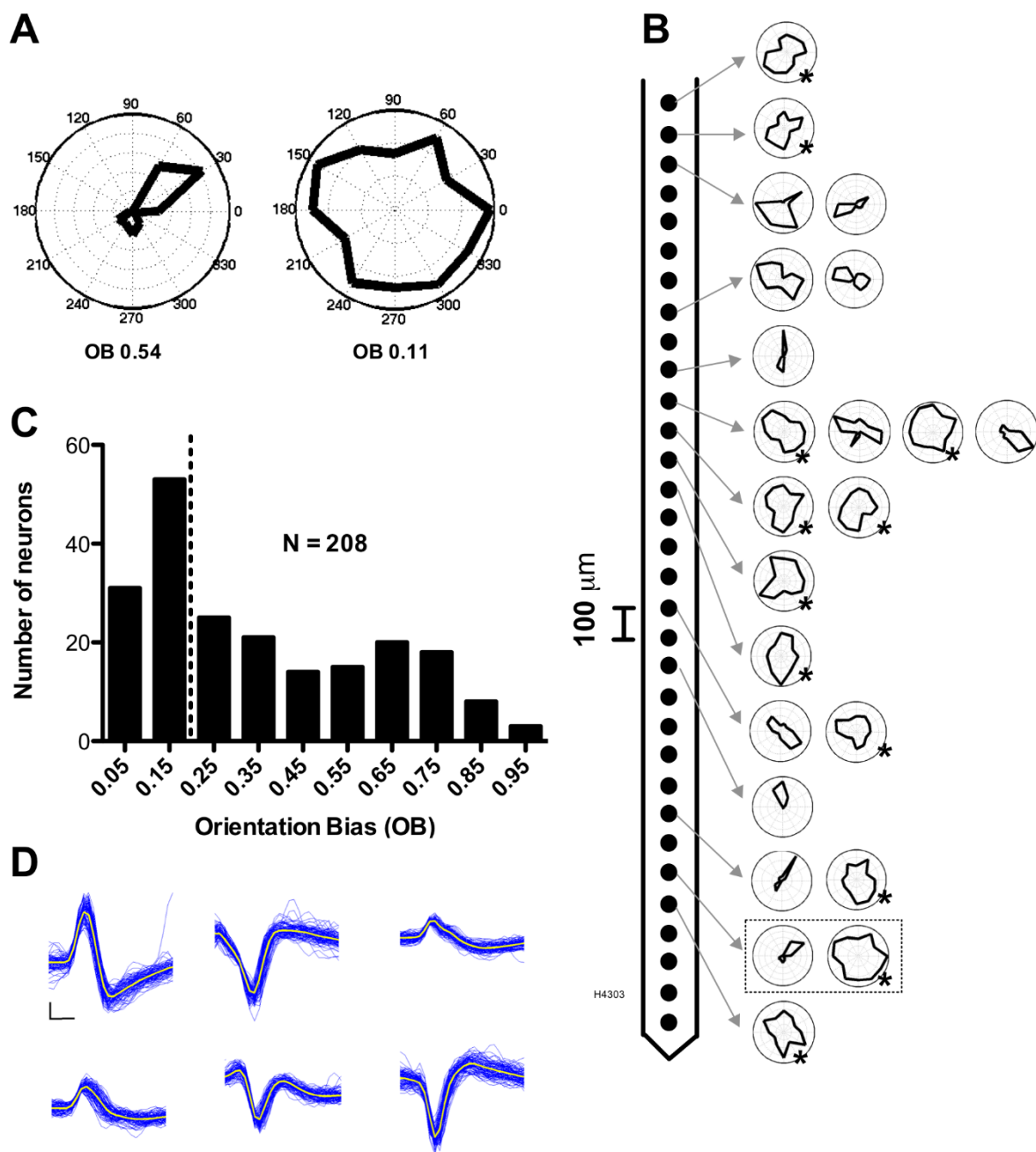


Figure 2.1 : Orientation tuning to drifting luminance gratings recorded with multielectrode arrays.

(A) Example tuning curves of an orientation-selective neuron (left) and a non-selective neuron (right), recorded simultaneously from the same site on a multielectrode. (B) Orientation tuning curves of neurons recorded simultaneously with a 32-channel linear array inserted almost orthogonal to, and spanning, the cortical layers. Neurons showed varying degrees of orientation selectivity, with a large fraction lacking significant orientation selectivity (denoted by asterisks). Dotted box indicates pair of neurons in (A). (C) Orientation selectivity of neurons is measured with an orientation bias (OB) index, with higher values indicating greater orientation-selectivity. Histogram shows distribution of OB values of all 208 neurons in our Area 18 sample. Neurons with $OB < 0.2$ are classified as non-orientation selective (LGN-like). More than one third (84/208) of these neurons are non-orientation selective. (D) Sorted spike waveforms for six example non-orientation selective neurons recorded simultaneously.

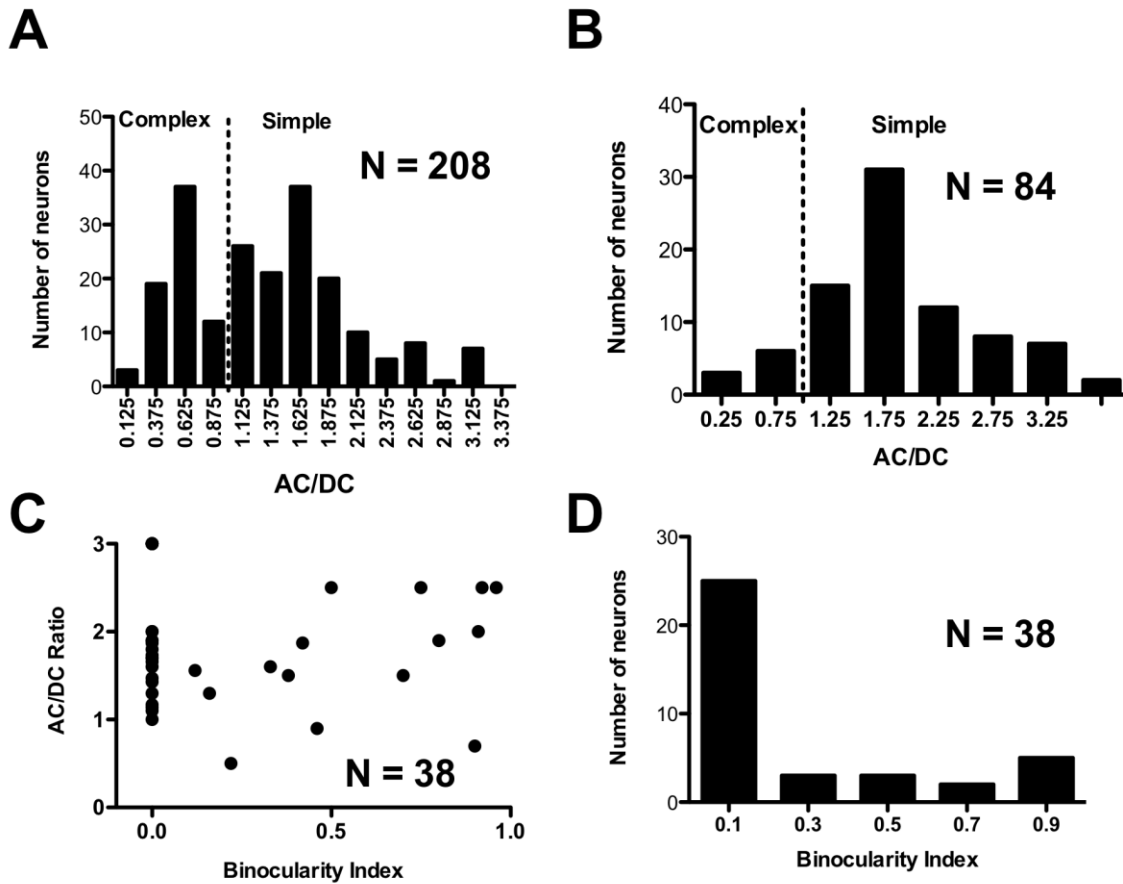


Figure 2.2 : Receptive field properties of non-orientation selective neurons.

(A) Histogram showing distribution of AC/DC (modulated/mean response) values of all 208 neurons in our Area 18 sample. (B) Histogram showing distribution of AC/DC (modulated/mean response) values for non-orientation selective neurons' responses to drifting gratings. The majority of these neurons are simple type (AC/DC > unity). (C) Scatterplot of binocularity index versus AC/DC ratio for non-orientation selective neurons. There is no clear relationship between these two parameters. (D) Histogram showing distribution of binocularity indices for non-orientation selective neurons. Most of these cells are monocular (index < 0.1), but about one-third are binocular.

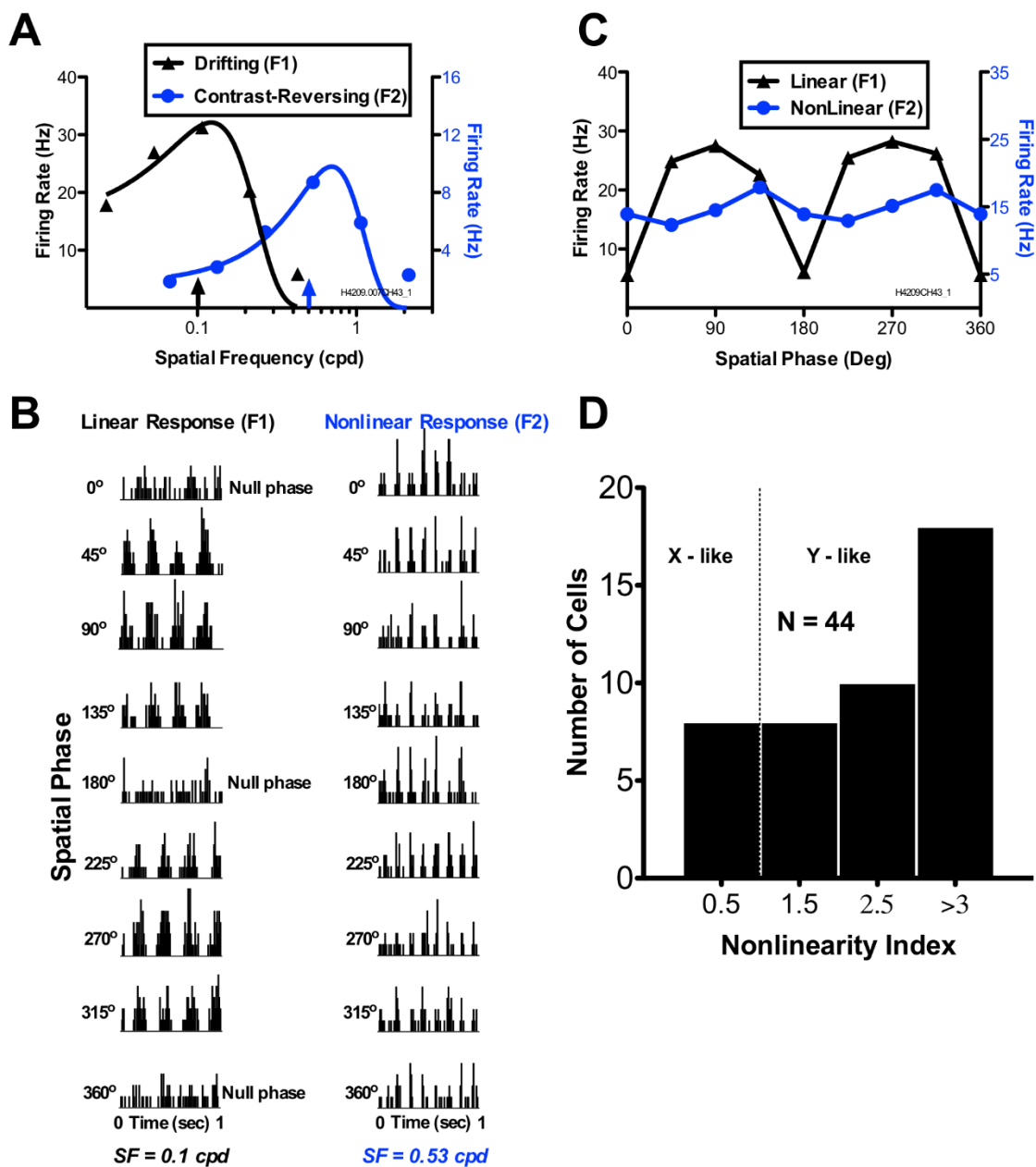


Figure 2.3 : Y-like non-orientation selective neurons in Area 18.

(A) Spatial frequency responses of a typical non-orientation selective neuron. First harmonic response to drifting gratings (F1, red) is bandpass to low spatial frequencies. Similar to subcortical Y cells, this neuron responds nonlinearly at the second harmonic (F2, blue) to contrast-reversing high spatial frequency gratings. **(B)** PSTHs (peristimulus time histograms) of the same neuron to contrast-reversing gratings (4Hz) of low (0.1cpd, left) and high (0.53cpd, right) spatial frequencies. At low spatial frequency the neuron responds at the first harmonic (4 Hz) with periodic phase dependence, while at high spatial frequency it exhibits a second harmonic (8 Hz) across the full range of phases. **(C)** Harmonic responses calculated from PSTHs in (B) as a function of spatial phase. First harmonic response (F1, red) is phase-dependent with a clear null phase repeated every 180° , while second harmonic response (F2, blue) is phase-independent. **(D)** Distribution of nonlinearity indices of non-orientation selective neurons. Most neurons are Y-like (nonlinearity index > 1.0).

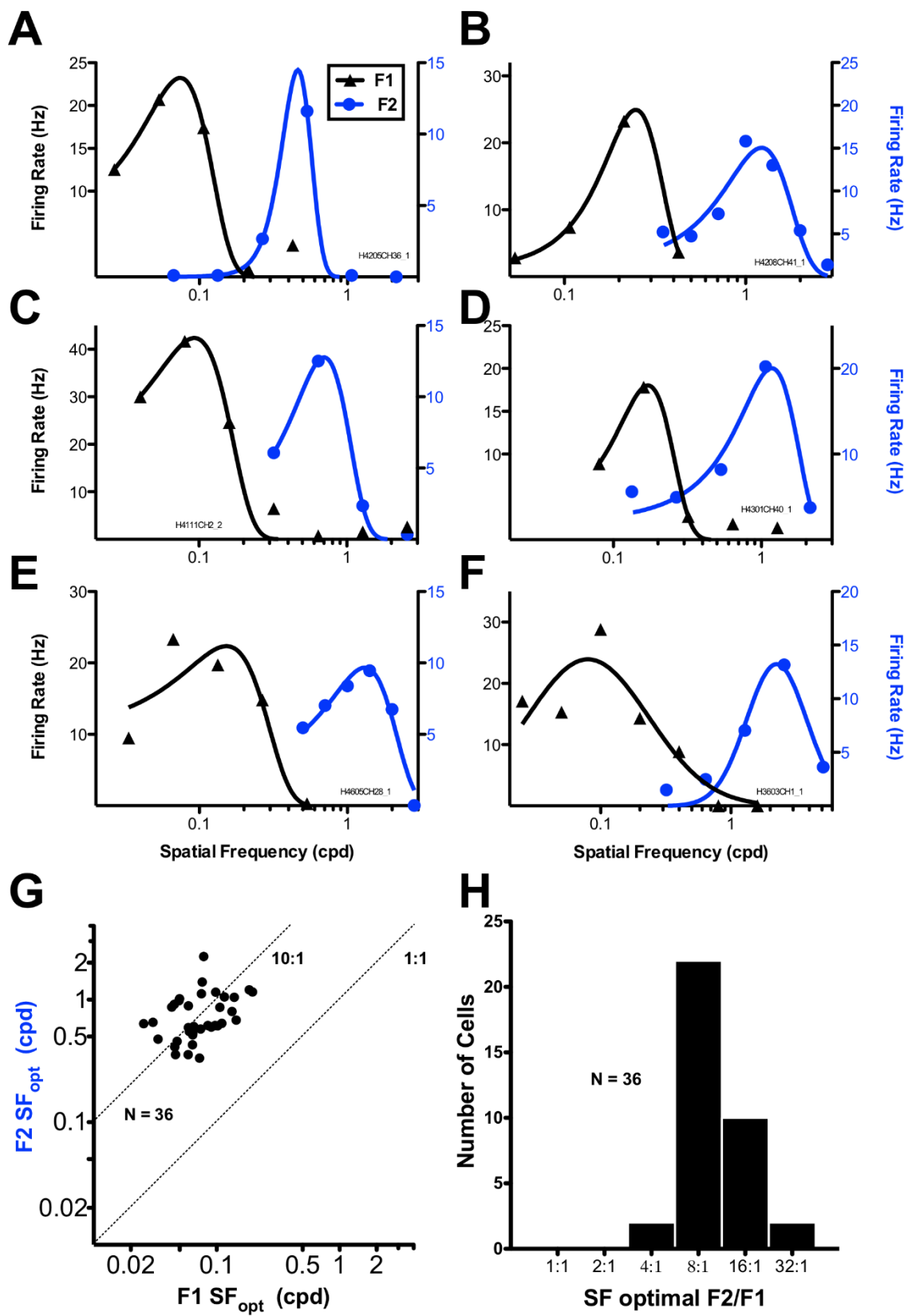


Figure 2.4 : Linear and nonlinear spatial frequency tuning of cortical Y-like neurons.

(A-F) Spatial frequency tuning for six cortical Y-like neurons. First harmonic (F1, red) responses of these neurons are band-pass tuned with selectivity for low spatial frequencies. Second harmonic (F2, blue) responses are band-pass tuned with selectivity for high spatial frequencies outside the luminance passband. **(G)** Scatterplot of optimal spatial frequency for second harmonic responses (F2) versus that for first harmonic responses (F1). All the points lie well above the 1:1 ratio line, indicating that a given neuron's optimal spatial frequency for F2 is substantially higher than for F1. **(H)** Histogram showing ratios of optimal spatial frequency for F2 vs. F1 (mean ratio = 11.3, median = 8.68).

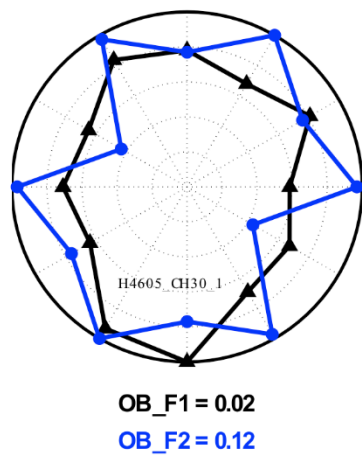
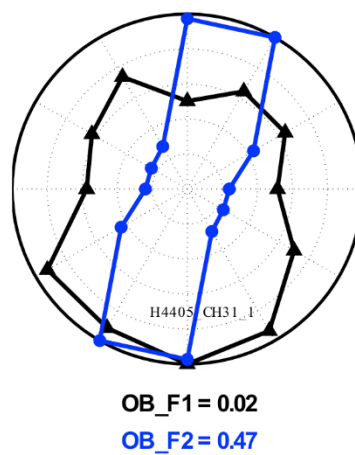
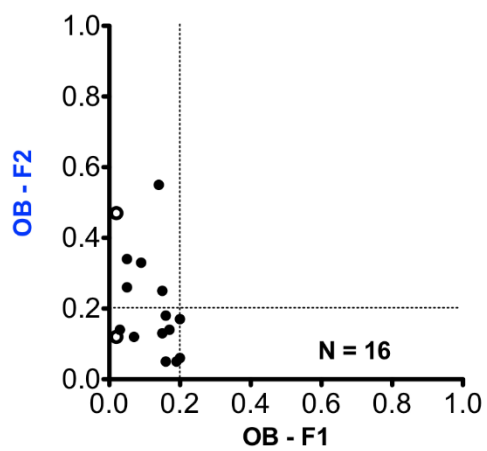
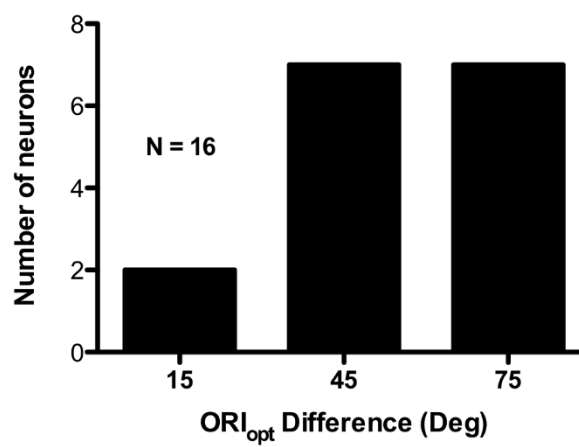
A**B****C****D**

Figure 2.5 : Linear and nonlinear orientation responses of Y-like non-ori neurons.

(A-B) Linear (F1, red) and nonlinear (F2, blue) orientation tuning plots of two Y-like non-ori neurons. One (A) is isotropic for orientation of nonlinear (F2) responses, while the other (B) has pronounced orientation tuning ($OB > 0.2$). **(C)** Scatterplot showing comparison of orientation tuning for low spatial frequency drifting grating (F1) and high spatial frequency contrast-reversing grating (F2). Higher OB values indicate greater selectivity. Open circles in the scatterplot correspond to the neurons in panel (A) & (B). **(D)** Histogram showing difference in optimal orientation for linear (F1) response and nonlinear (F2) response.

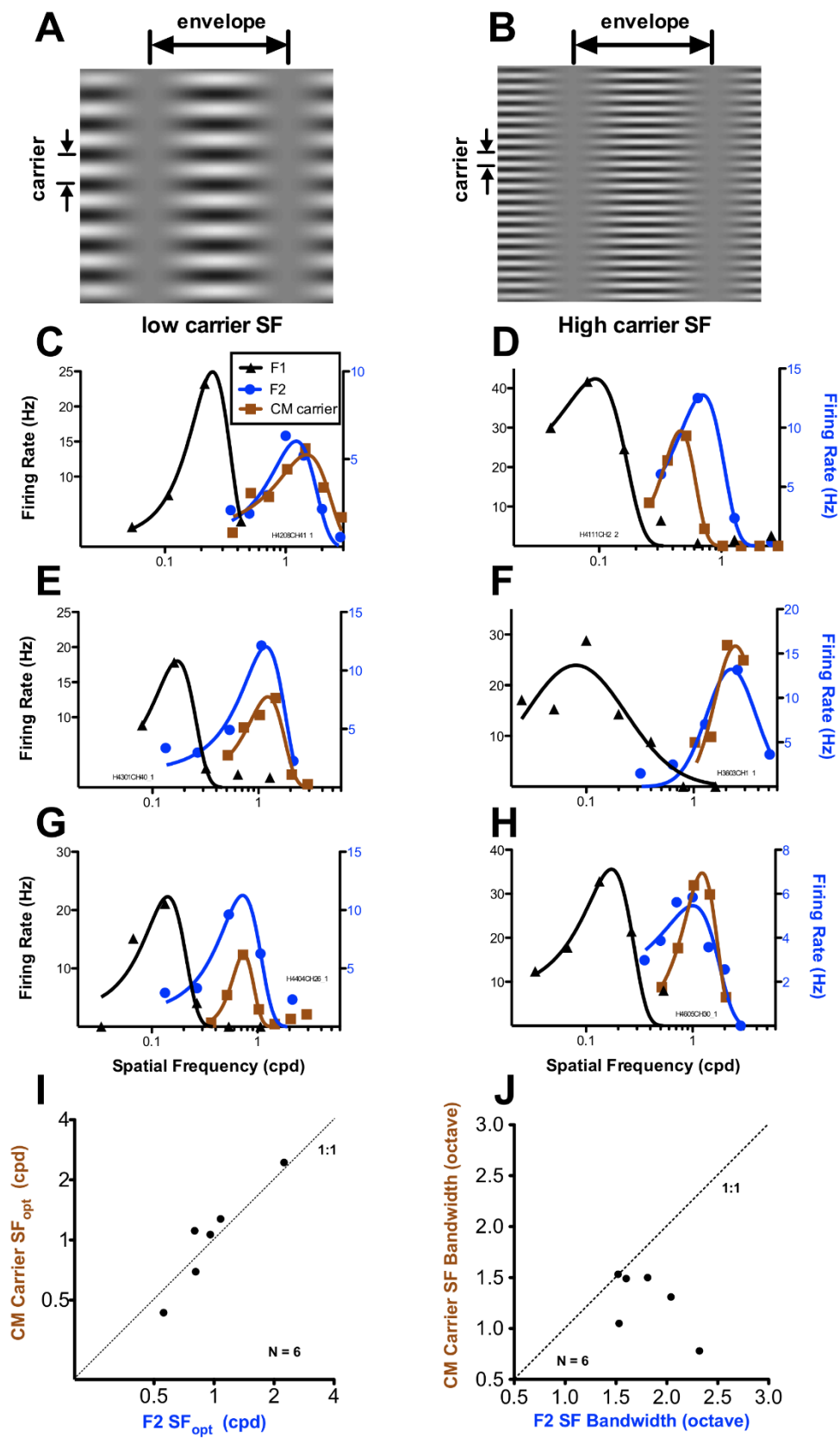
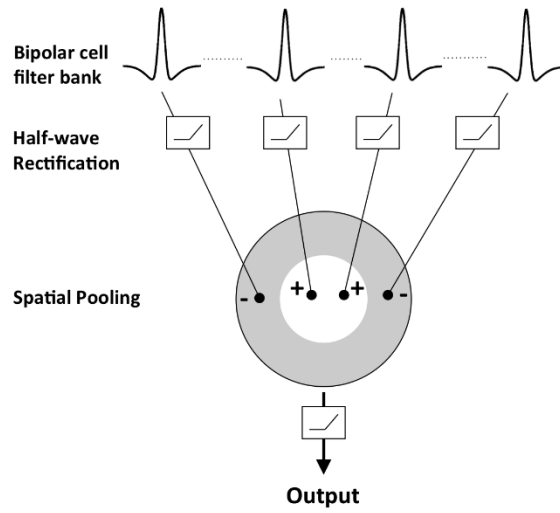


Figure 2.6 : Responses of Y-like non-ori neurons to contrast modulation (CM) gratings .

(A-B) Two examples of contrast modulation stimuli with vertically oriented low spatial frequency envelope that modulates contrast of horizontal carrier grating at low (A) or high (B) frequency. **(C-H)** Spatial frequency tuning plots of six neurons to contrast modulation and luminance gratings. For a given neuron, the CM envelope spatial frequency was fixed at a low value within the luminance passband (F1, red), and carrier spatial frequency was varied outside the luminance passband. Neurons show bandpass tuning to CM gratings (orange), similar to their second harmonic (F2) response to contrast-reversing gratings (blue). **(I)** Scatterplot of optimal spatial frequency of contrast reversing luminance gratings for second harmonic (F2) responses versus optimal spatial frequency of CM carrier grating. **(J)** Scatterplot showing spatial frequency bandwidth of second harmonic (F2) response versus bandwidth of CM carrier grating.

A

Y cell Model (Retinal Circuit)



B

Unbalanced Push-Pull (Cortical Circuit)

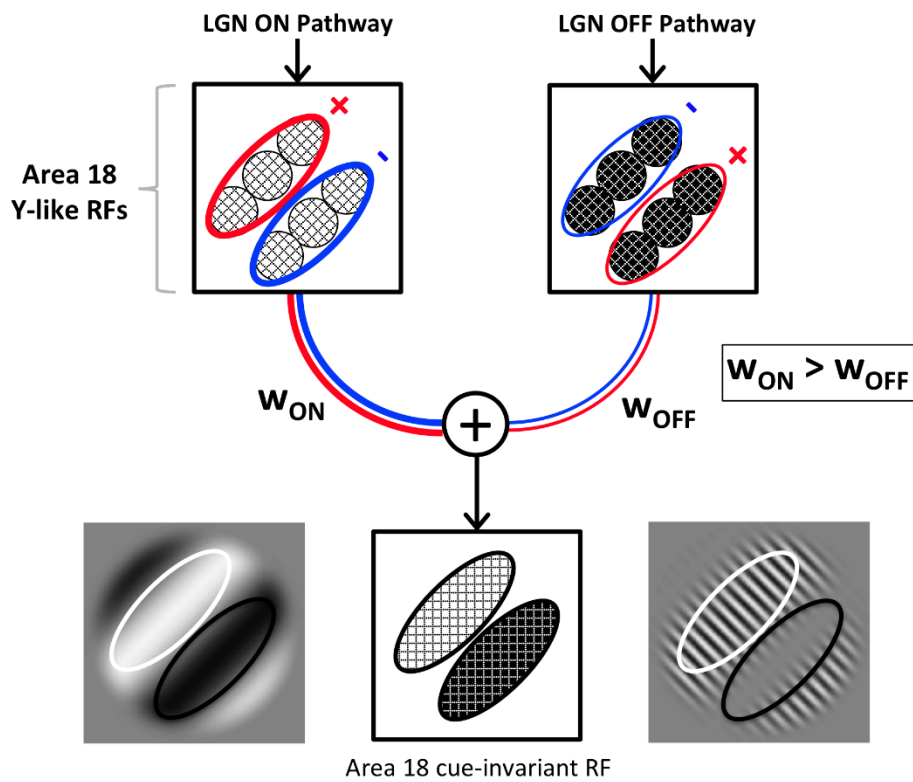
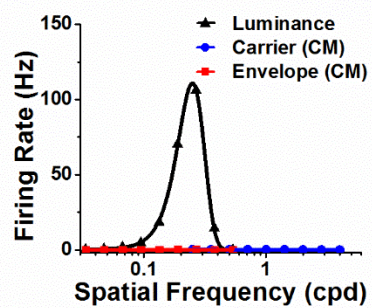


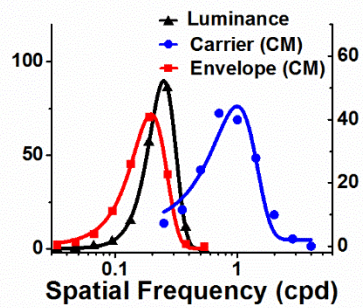
Figure 2.7 : Neural circuit model for cue-invariant receptive fields constructed from Y-pathway.

(A) The receptive field of a Y cell is modeled as a filter-rectify-filter (FRF) cascade. The first filter stage is comprised of a bank of small DoG (difference of Gaussian) receptive fields corresponding to bipolar cells. The rectified outputs of these subunits are linearly pooled with DoG weighting, at a much larger spatial scale (Demb et al., 2001a). **(B)** Receptive field of a cue-invariant cortical simple cell can be thought of as a pair of overlapping phase-aligned receptive fields, one constructed from summation of ON- and the other from OFF-center inputs. ON- and OFF-centre Y-like cortical neurons are combined in a push-pull arrangement, such that the ON region of the cortical neuron receives excitatory input from ON-centre cells and inhibitory input from OFF-centre cells, and vice-versa for the OFF region. This model will respond selectively to an oblique oriented luminance edge. But since the Y cells contain small nonlinear subunits, both ON and OFF types will be excited by the presence of texture, resulting in no net response. When the push-pull from ON- and OFF-pathways are unbalanced (e.g. stronger input from ON pathway), the nonlinear responses to texture no longer cancel, thereby enabling envelope orientation-selective responses to CM stimuli.

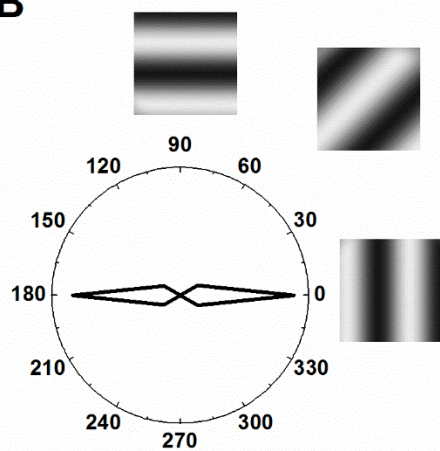
A Balanced Push-Pull



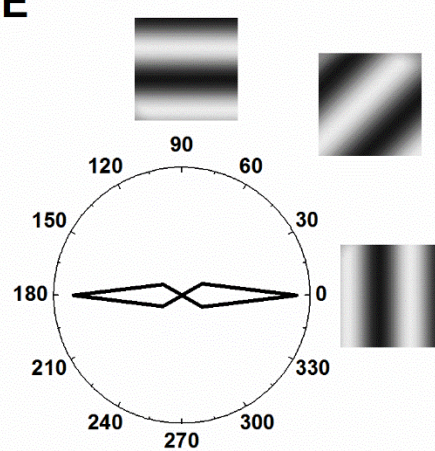
D Unbalanced Push-Pull



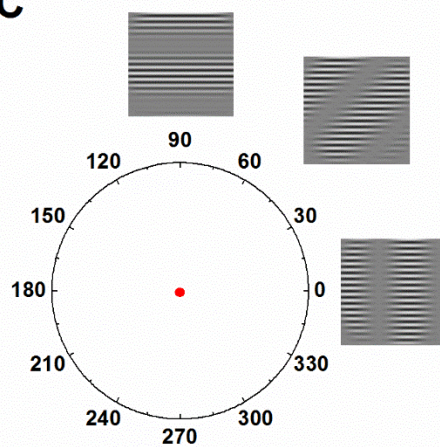
B



E



C



F

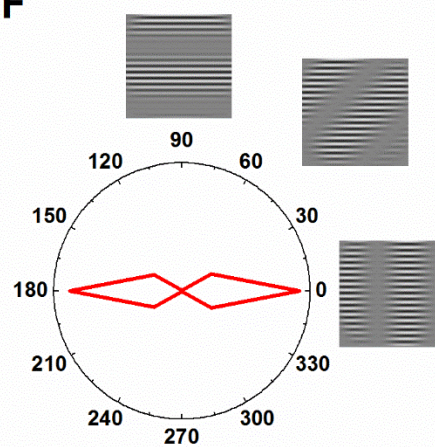
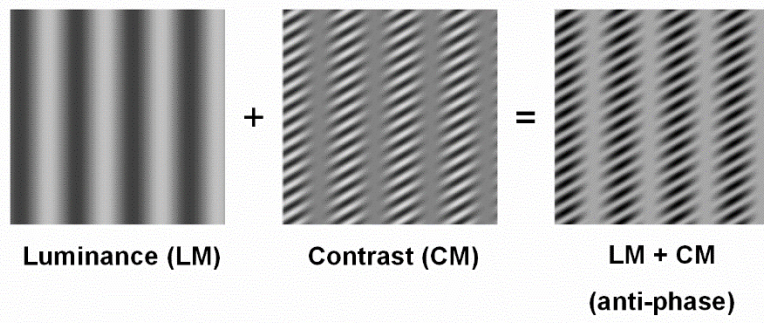
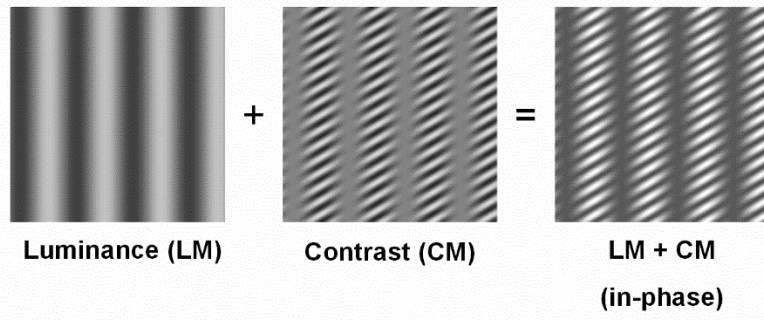


Figure 2.8 : Spatial tuning of balanced and unbalanced push-pull model with Y-pathway inputs.

(A) Spatial frequency response of the model with balanced push-pull. Responses to luminance gratings (black) are band-pass tuned to low spatial frequencies, but the model does not respond to CM gratings (red, blue) with carrier spatial frequencies outside the luminance pass-band. (B) Balanced push-pull model shows orientation selectivity to drifting luminance gratings. (C) Balanced push-pull model does not respond to CM gratings of any envelope orientation. (D) Spatial frequency response of the model with unbalanced push-pull. Responses to luminance gratings are band-pass tuned to low spatial frequencies as in A, but the model also responds to CM gratings with carrier spatial frequency selectivity (blue) outside luminance passband and envelope selectivity (red) similar to that for LM gratings. (E) Orientation tuning of the unbalanced push-pull model to drifting luminance gratings shows similar selectivity as B. (F) Similar orientation tuning of the unbalanced push-pull model to envelope of CM gratings.

A



B

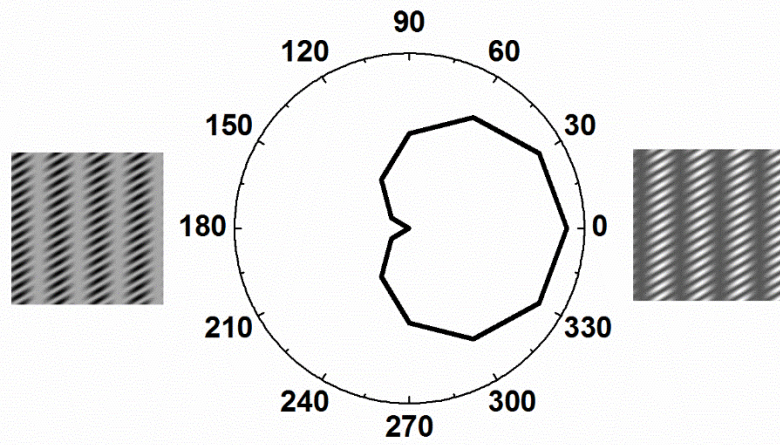


Figure 2.9 : Selectivity of the unbalanced push-pull model for relative spatial phase of luminance and contrast boundary.

(A). LM+CM compound gratings were constructed by linearly adding LM and CM gratings having identical spatial frequency and orientation. Compound stimuli are illustrated for "in-phase" condition (upper), where high luminance of LM grating is phase-aligned with high contrast of CM grating, and "anti-phase" (lower), where high luminance of LM grating is phase-aligned with low contrast of CM grating. **(B).** Responses of the unbalanced push-pull model to compound LM+CM gratings with varying relative spatial phase. The model, with $w_{ON} > w_{OFF}$, responds strongest when LM and CM gratings are phase-aligned and weakest when they are in anti-phase.

3

The previous chapter provided evidence supporting the role of the nonlinear Y pathway as an input for processing second-order stimuli in the visual cortex. Hence it is important to build quantitative receptive field models of subcortical Y cells that can account for their signal processing. So in this chapter, I estimate biologically plausible neural network models of cat LGN receptive fields, using recent machine learning methods (deep learning). I train convolutional neural network models for each recorded LGN neuron, using its spiking responses to naturalistic texture stimuli. I demonstrate that, these models are not only better in comparison to the standard linear-nonlinear models at predicting response to arbitrary stimuli, but they also recover biologically interpretable subunit models. This chapter is under preparation for submission to a peer reviewed journal.

CHAPTER 3

Estimating Subunit Receptive Field Models of Thalamic Neurons with Deep Learning.

3.1 Abstract

Spatial nonlinearities emerging at the level of the retina are thought to be important for processing texture boundaries in the visual cortex. However, previous studies have modeled receptive fields of subcortical neurons as linear filters with a static output nonlinearity. Although such "LN" models are relatively easy to estimate with reverse-correlation they cannot capture spatial nonlinearities within receptive fields. Here we model retinorecipient receptive fields of cat lateral geniculate nucleus (LGN) neurons as a two-layer convolutional neural network model with an intermediate parametric nonlinearity (pReLU). We train such a neural network model for each neuron, using its spiking responses to naturalistic texture stimuli. The convolutional filter is initialized with random weights, and no constraints are imposed on its shape. The learned models converge on to filters with clear Gaussian or DoG (Difference of Gaussians) like shapes, and often exhibit a high predictive performance on test datasets. The trained models of Y-type LGN neurons have a higher degree of nonlinearity compared to those for X-type neurons. In conclusion, a nonlinear two-layer convolutional model that is based on retinal neurobiology is better at predicting responses of Y-type neurons to novel test stimuli compared to an LN model.

3.2 Introduction

Our visual system utilizes both changes in luminance as well as higher order texture statistics for segmenting an object from its background. Even though processing of luminance signals is well understood at the early stages of the visual pathway, neural mechanisms for processing texture remain unclear. Recent studies (Rosenberg et al., 2010) have suggested that the subcortical Y pathway could be the neural substrate for joint processing of luminance as well as texture information. Y-type (alpha) ganglion cells in the mammalian retina have been classically distinguished from other retinal ganglion cell (RGC) types by the fine-scale spatial nonlinearities in their receptive fields (Enroth-Cugell and Robson, 1966; Hochstein and Shapley, 1976; Borghuis et al 2013; Gollisch, 2013; Crook et al., 2008a). The presence of a rectifying nonlinearity at the bipolar-ganglion cell synapse enables Y cells to detect fine scale textures as well as coarse-scale luminance changes (Demb et al, 1999; Demb et al., 2001a; Demb et al., 2001b). The resultant signals are relayed to the cortex via Y-type cells in the lateral geniculate nucleus (LGN), where recent evidence suggests they are utilized for building receptive fields that can encode orientation of boundaries formed by texture (second-order) as well as luminance (first-order) changes in a cue-invariant manner (Rosenberg et al., 2010; Rosenberg and Issa, 2011; Gharat & Baker, 2017).

A rigorous way to gauge our understanding of visual signal processing by a receptive field is to build quantitative models and test their ability to predicting responses to random stimuli. However, most previous efforts to fit quantitative receptive field models to RGC and LGN neural responses have assumed a linear filter followed by a static output nonlinearity (LN model) (Reid et al., 1997; Pillow et al., 2005, Wang et al., 2011). Obviously such linear filter models cannot capture selectivity to texture stimuli arising from multi-stage nonlinear processing

within the Y cell receptive fields. Recent studies have begun to incorporate these spatial nonlinearities by modeling receptive fields as multi-stage “subunit” models (Schwartz et al., 2012; Freeman et al., 2015; Turner and Rieke, 2016; Bolinger and Gollisch, 2012). However, it has been a challenge to develop a framework that can estimate parameters of robust and biologically interpretable models, using limited and noisy neural data (Vintch et al., 2011).

In this study, we use a deep learning framework to train multi-layer convolutional neural networks (CNNs) to model receptive fields of LGN neurons. Interestingly, the architecture of CNNs is inspired from the mammalian visual system, particularly its hierarchical multi-stage processing, containing a cascade of spatially localized linear filters convolved across space, separated by static rectifying nonlinearities (LeCun et al., 2015). Thus, CNNs are a well-suited framework for capturing spatial nonlinearities in Y cell receptive fields. Furthermore, a recent study (Yamins et al., 2014) obtained very promising results, in which they showed that deep CNNs trained on an object recognition task could accurately predict visual responses of neurons in higher visual areas of the ventral stream.

Here we estimate predictive receptive field models of single neurons in the cat LGN from their responses to high-resolution naturalistic synthetic textures. We train convolutional neural network models from responses of each neuron using backpropagation gradient descent methods. The architecture of the model (Figure 3.2) consists of a spatiotemporal convolutional filter layer, (“subunit filters”) followed by a parametric rectifying nonlinearity (pReLU). The rectified outputs of this layer are pooled by a densely-connected layer, and passed through a final output nonlinearity. We find that for Y-type cells, the convolutional model with intermediate nonlinearities performs significantly better than a conventional linear-nonlinear (LN) model at predicting neuronal responses to a holdback dataset not used for training the models. But for the

X-type (linear) cells, both models performed equally well. The learned convolutional filters resemble the receptive fields of retinal bipolar cells, having fine-scale, isotropic shapes. For a subset of neurons, adding a separate pathway to capture feedforward inhibition provided a big improvement in the model's predictive ability.

3.3 Methods

3.3.1 Animal Preparation and Electrophysiological Recordings

Animal procedures were as described in our previous study (Gharat & Baker, 2012). Briefly, intravenous cannulation was performed on adult cats of either sex under anaesthesia induced with isoflurane/oxygen. To record single-unit responses from lateral geniculate nucleus (LGN), a craniotomy and durotomy (HC A6/L9) were performed under anaesthesia with intravenous propofol. During electrophysiological recordings, the animal was anesthetised and paralyzed with infusions of propofol, remifentanyl, and gallamine triethiodide, and inhalation of N₂O/O₂. Vital signs (expired CO₂, blood O₂, heart rate, EEG, temperature) were monitored and maintained at appropriate levels throughout the experiment. Neutral contact lenses were provided for corneal protection, spectacle lenses for refraction, and artificial pupils for improved optical quality. Animal procedures were approved by the Animal Care committee of McGill University and are in accordance with the guidelines of the Canadian Council on Animal Care.

Neurons' responses in the LGN were recorded in most cases using quartz-coated tungsten tetrodes (Thomas Recording), and occasionally with single-channel glass-coated tungsten electrodes (Alpha Omega). The LGN was identified by strong monocular multi-unit responses to flashing light at around 13 mm depth, that switched with increased depth from the contralateral

eye (layer A) to ipsilateral (layer A1) and back to contralateral (layer C). All the single-units in this study were recorded in layers A and A1. Raw data signals were acquired with a Plexon Recorder (3 Hz to 8kHz; sampling rate, 40 kHz) and streamed to a hard disk for later analysis. A small photocell placed on a corner of the display screen provided signals for temporal registration of stimuli and spike times.

3.3.2 Visual Stimuli and Recording Protocol

Visual stimuli were generated with an Apple Macintosh computer (MacPro 4.1, 2.66 GHz/4 core, 6Gb, MacOSX 10.6.8, NVIDIA GeForce GT120) using custom MATLAB software with Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al. 2007) and presented on a gamma-corrected CRT monitor (NEC FP1350, 20 inches, 640 x 480 pixels, 150 Hz, 36 cd/m²) at a viewing distance of 57 cm. The display screen was roughly centred on the receptive field of multi-units, using manually controlled, drifting or flashing white or black bar or spot stimuli on a grey background. During recording, multi-unit signals from a single channel were used to guide the recording protocol. Responses of neurons were first recorded with conventional sinusoidal drifting gratings of varying spatial frequencies to measure spatial frequency tuning. Then contrast-reversing gratings were presented to assess spatial nonlinearities and classify neurons as X- and Y-type (see below).

After measuring responses to gratings, image sequences of naturalistic texture patterns (Kingdom et al, 2001; Zavitz and Baker, 2013) were presented to map receptive fields of neurons (Figure 3.1). Y cells respond linearly to low spatial frequencies, and the nonlinear subunit responses only become evident when presented with high spatial frequencies (Hochstein and Shapley, 1976). These spatial nonlinearities can enable responses to contrast-modulated texture

boundaries (Demb et al., 2001b). Our aim in this study was to map subunit structure within the receptive field using system identification, and hence we need a large set of independent images (~10,000) that can nonlinearly activate receptive fields - i.e. images rich in high spatial frequency textures. Instead of trying to manually select natural images with fine textures, we generated synthetic textures with naturalistic image statistics (Kingdom et al, 2001; Zavitz and Baker, 2013). These images were generated by modulating the contrast of a texture pattern (“carrier”) by an “envelope” as shown in Figure 3.1. Carrier images (Fig. 3.1A) were constructed with randomly positioned, high-density Gabors with high spatial frequencies ($\lambda = 0.25^\circ, 0.5^\circ, 1^\circ, 2^\circ$), with four different sizes and six different orientations ($0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ$). The sizes of the Gabors were chosen to cover the range of spatial frequencies (0.5cpd to 3cpd) that have been previously shown to activate Y cells nonlinearly (So and Shapley, 1979). The relative proportions of Gabors of various sizes was set such that the constructed image had a power spectrum with an approximately $1/f$ fall-off as in natural images (Field, 1987; Kingdom et al, 2001). Envelope images (Fig. 1B) were independently constructed in a similar manner but with coarse scale Gabors ($\lambda = 4^\circ, 8^\circ, 16^\circ, 32^\circ$). Then a carrier image and an envelope image were pixel-wise multiplied to generate a final modulated texture image (Fig. 1C) of 480×480 pixels ($30^\circ \times 30^\circ$). Stimulus movies were presented as 5-second sequences of 375 such images, refreshed every two CRT frames (i.e. at 75Hz), with the movie preceded and followed by a grey blank screen of the same mean luminance.

Three separate datasets were measured for training, regularization and testing of the receptive field models: the training dataset contains 20 movies repeated 5 times, the regularization and testing datasets each contain 5 movies repeated 20 times. Movies for training,

regularization and testing were randomly interleaved to minimize effects of slow changes in responsivity. Total recording time for the entire set of movies was around 40 minutes.

Single-units were isolated from recorded multi-unit signals offline using SpikeSorter software (Swindale and Spacek, 2014) and only clearly separated units were included in our analysis. After offline sorting, spike times of single neurons were binned in PSTHs (using bin width 13.33 ms, the duration of each frame), which were used to calculate first-harmonic (F1) and second-harmonic (F2) responses. Neurons were classified as Y-type if responses to contrast-reversing gratings of high spatial frequencies elicited significantly stronger F2 than F1 response, otherwise as X-type neurons (Hochstein and Shapley, 1976).

3.3.3 Receptive Field Model

The parameters of a convolutional neural network (CNN) model (LeCun et al., 2015) were optimized to fit naturalistic image responses of each neuron. The architecture of the CNN model used here (Figure 3.2) consists of a spatio-temporal convolutional filter layer, which models identical subunits tiled spatially on a square grid, followed by a spatially uniform static nonlinearity (PReLU, Parametric Rectified Linear Unit - He et al, 2015). Then these rectified filter outputs are linearly summed by a pooling layer (“Dense Layer”). Finally, the output of the pooling layer is passed through a pSoftplus (parametric softplus) nonlinearity. This pSoftplus nonlinearity has a shape similar to a ReLU (half-wave rectification) except that this function does not pass through the origin, and the output of the function is positive for both positive and negative inputs. This output nonlinearity was used to ensure that the final output of the neuron model never goes to exactly zero, because the Poisson loss function (see below) used for training the model takes the log of the output.

The PReLU nonlinearity is defined as:

$$f(x) = \alpha * x \quad \text{for } x < 0$$

$$f(x) = x \quad \text{for } x \geq 0$$

where α is a learned parameter.

The pSoftplus nonlinearity is defined as:

$$f(x) = \frac{1}{k} \log(1 + e^{kx}),$$

where k is a learned parameter.

The Poisson loss function is:

$$L = \hat{r} - r * \log(\hat{r})$$

where r is the actual firing rate and $\hat{r}$ is the predicted firing rate.

3.3.4 Model Training and Performance

Before training CNN models for each neuron, a Linear-Nonlinear (LN) model (Talebi & Baker, 2012) was estimated, to facilitate manually setting a cropping window to encompass the centre and surround of the receptive field. Then the cropped stimuli were downsampled to 40x40 pixels. The CNN models were trained using the Python Deep Learning library, Keras (Chollet, 2015), with the Theano (Bergstra et al, 2010) backend. The models were trained by minimizing a Poisson loss function with the Adam optimizer (Kingma & Ba, 2014). The weights of the convolutional filters and dense layer were randomly initialized using a Glorot normal initialization (where weights are drawn from a truncated normal distribution centred on zero), as it leads to faster convergence of neural networks (Glorot and Bengio, 2010). To avoid

overfitting, we used L2 weight regularization along with dropout ($p=0.5$) for both layers, and Poisson noise was applied to the input (Oliver MD, 2014; Lane et al, 2016). Poisson noise was used instead of Gaussian noise traditionally used for training neural networks, since a previous study (Oliver, 2014) found it to estimate more robust and less noisy subunit filters as well as improve the models' predictive performance. Each neural network model was trained for 1000 epochs with early stopping ("patience" parameter in Keras, set to 50).

Regularization datasets were used to optimize hyperparameters related to L2 regularization as well as architectural parameters (filter sizes). One of the challenges with training neural networks is the optimization of multiple hyperparameters. This was especially problematic here because we trained separate neural networks for >100 neurons. Consequently, instead of testing all possible combinations of all the hyperparameters on the datasets for every neuron, we used data from a small group of representative neurons (10) to test hyperparameters over a wider range. The results from these assessments indicated the relevant ranges over which to do more restricted grid searches for optimal hyperparameter values, for all the other datasets.

A separate holdback Test dataset, not used for training or regularization, was reserved for testing predictive performance of the model. Performance of each fitted model was quantified as percent variance accounted for (% VAF), calculated as the square of the correlation coefficient (r^2) between actual neuronal response and predicted response.

We also estimated a linear-nonlinear model (LN model) for comparison of predictive performance with that of the convolutional model. To be comparable, we used the exact same model architecture with the same number of parameters, except that the intermediate pReLU nonlinearity was removed, making it essentially an LN model. Furthermore we also used the same method (eg: optimizer, regularization etc) as used earlier for training the nonlinear model.

3.4 Results

3.4.1 Mapping nonlinear subunit receptive field models with deep learning

For this study, we recorded from 93 neurons (41 Y cells and 52 X cells) in 4 cats. For each neuron, a separate neural network model was trained using the neuron's responses to movies of naturalistic texture stimuli - see Methods for details of the model training. We constrained the weights in the pooling layer to be positive, which should capture the center mechanism of the receptive field. From here on we will refer to this architecture as a "1-branch" model, as it has only a single convolutional filter. (In a later section, we also estimate a "2-branch" model, i.e. having two distinct convolutional filters acting in parallel, one with positive pooling weights and the other with negative weights, that could potentially capture center as well as surround mechanisms of the receptive field.) No constraints are imposed on the shapes of the convolutional filter or the pooling layer weight map. In this framework, we can simultaneously learn the convolutional filter weights, the α parameter of the PReLU nonlinearity, and the weight map of the pooling layer.

Figure 3.3 shows an example of a trained neural network for one of the neurons in our sample. The model optimization recovers a subunit filter (upper left) that is spatially isotropic with a strong OFF-center (blue) and weak excitatory surround (red), and with a biphasic temporal filtering. The α parameter that corresponds to the degree of nonlinearity is 0.15, which is close to half-wave rectification. The 2D map of the pooling weights shows the spatial locations at which the subunits act, and their corresponding weights - note that pooling weights were constrained to be positive. The linear reconstruction map shown at the bottom is obtained by convolving the subunit spatial filter with the pooling weights.

Figure 3.4 shows trained neural networks of four additional example neurons. The neuron in panel A has an ON-center subunit spatial filter with a possible antagonistic surround organization, a monophasic temporal filter, and pooling layer weights whose layout does not appear to be clearly isotropic. The neuron in Figure 3.4B has a subunit that is OFF-center with little or no surround, and a biphasic temporal response, and a pooling layer that appears slightly oriented. Figure 3.4C shows a neuron whose subunit has an ON-center and surround inhibition, with an isotropic map layer. The neuron in Figure 3.4D shows a clear ON-center, OFF-surround for the subunit, with a biphasic temporal filter. The map layer is isotropic and substantially smaller than the other examples - taken together with the nearly linear α parameter for the PReLU. The α parameter of the nonlinearity is close to zero (half-wave rectification for the other examples (Fig. 3.4A,B,C)).

3.4.2 Predictive performance for novel stimuli

After estimating subunit receptive field models for each neuron, we tested their predictive ability for responses to novel test stimuli. This test dataset was not used for training or validating the model, so it gives a sense of how well the model generalizes to novel stimuli. The performance of the model was measured as the percentage of the neuronal response variance accounted for (VAF) - see Methods. For comparison, we also estimated a linear receptive field model for each neuron - for comparability we used exactly same model architecture and method of training, except that the intermediate nonlinearity (pReLU) was removed, making the model essentially a linear-nonlinear (LN) architecture.

Figure 3.5 shows scatterplots comparing neurons' VAF values for the 1-branch subunit model to those for a linear model, with each point denoting results from a single neuron. For Y-

type cells shown in Figure 3.5A, most of the points lie above the 1:1 line, suggesting that the subunit model performs better than the linear model at predicting responses of these neurons to novel stimuli. A Wilcoxon-signed rank test confirms that VAFs for the subunit model are significantly greater than for the linear model ($p= 1.4995e-04$, $N=41$). On the other hand, for X-type cells shown in Figure 3.5B most of the points lie close to the 1:1 line, suggesting that the subunit and linear models perform equally well at predicting responses to novel stimuli. A Wilcoxon-signed rank test confirms that the VAF values for the subunit model and for the linear model are not significantly different ($p= 0.5$, $N=52$). Thus having a spatial nonlinearity within the receptive field model improves its performance for the Y- but not X-type cells.

3.4.3 Intermediate nonlinearity (α)

One of the advantages of the deep neural network learning framework is that in some cases it can learn the intermediate nonlinearity between the subunit filter and pooling weights. We employ the PReLU nonlinearity (He et al., 2015), a simple piecewise-linear function having only one parameter – α – that denotes the slope of the negative half of the nonlinearity. It can vary from +1 (linear) to -1 (full-wave rectification). Thus the α parameter controls the degree of spatial nonlinearity within the receptive field. Neurons in our population were classified as X- or Y-type depending upon their responses to contrast-reversing gratings. Neurons that gave frequency-doubled nonlinear response to contrast-reversing gratings were classified as Y-type cells. So we wanted to examine whether there was a correspondence between neurons' α values and their cell type.

Figure 3.6 shows the distributions of α values for the Y- and X-type cells. For all the neurons in our sample, the α value ranged from 0 (half-wave rectification) to 1 (linearity). For large fraction of the Y-type cells (24/41, 59 %) α was less than 0.5, while α was greater than 0.5 for 39/52 (75 %) of X-type cells. The α values for Y-type cells were significantly less than for X-type cells ($p = 2.6335e-06$, Mann-Whitney U test). Thus with this modeling approach we could capture the varying degree of spatial nonlinearity within the receptive fields of different neurons, with trained models of Y-type cells having a higher nonlinearity than X-type cells.

Previous ex vivo retinal studies in guinea pigs (Demb et al., 2001a) and mice (Borghuis et al., 2013) have shown that OFF-type Y cells have a higher degree of rectification than ON-type cells. However in our sample, Y-type cells' α values were not significantly different between ON-type cells (mean = 0.45, N = 28) and OFF- type cells (mean = 0.52, N =13) (rank sum test: $p = 0.4085$).

We wanted to test if the improvement seen in the predictive power of subunit models compared to linear models could be related to the α parameter of the nonlinearity. To examine whether there is any systematic relationship, we constructed scatterplots (Figure 3.7) comparing α values of the estimated models to their improvement in predictive performance. Improvement was calculated as the ratio of subunit VAF to linear VAF. For Y-type cells shown in Figure 3.7A, neurons with α greater than 0.5 (towards linearity) did not show significant improvement in performance. But for neurons with α less than 0.5, performance of the neurons increased with decreasing α (increasing nonlinearity). This trend was statistically significant (Pearsons correlation coefficient $R = -0.7165$, $p = 1.3780e-07$, $N=41$). However, for X-type cells shown in

Figure 3.7B there was no significant correlation between neurons' α and improvement in performance ($R = -0.1278$, $p = 0.3667$, $N=52$).

3.4.4 Two-branch model

As mentioned earlier we further extended the 1-branch model to have a second branch with its own convolutional filter, PReLU and pooling layer, that could potentially capture the inhibition within the receptive field from the amacrine cell pathway. This was modeled by adding a second convolutional filter as shown in Figure 3.8. Outputs of this filter were also passed through a pReLU nonlinearity, combined in a separate Pooling Layer, and then passed through a half-wave rectification (ReLU). Pooling weights of the second branch were constrained to be negative to capture the feedforward inhibition. Half-wave rectification was included with this filter to capture the rectification introduced from the spiking responses of amacrine cells (Demb et al., 1999). Figure 3.9 shows the 2-filter receptive field model estimated for the example neuron previously shown in Figure 3.3. The spatial and temporal filters recovered for the excitatory filter in this framework is very similar to those recovered in the 1-branch analysis of Figure 3.3. Interestingly, for the inhibitory filter, the temporal filter has delayed dynamics compared to the excitatory filter. Also, the linear reconstruction for the inhibitory filter covers a wider spatial region. Consequently when the linear reconstructions for the two filters are summed, we get a centre-surround antagonistic organization as shown on the right. Adding the second filter improves predictive performance for this neuron, with the VAF increasing from 40.8% to 52.6%.

Figure 3.10A compares predictive performance of 2-branch versus 1-branch models for Y-type neurons. For a subset of neurons there is a clear improvement in VAF for the 2-branch

model. However for X-type neurons (Figure 3.10B) there was not a substantial improvement in VAF. Finally, Figure 3.10C compares the best subunit model (between 1- and 2-branch) versus the linear model for Y-type neurons. Clearly, the nonlinear subunit model performs better at predicting responses of Y-type neurons. This is confirmed by a Wilcoxon-signed rank test, which shows the VAF for the subunit model is significantly greater than for the linear model ($p=7.1601e-08$, $N=41$). However for X-type neurons (Figure 3.10D) there is only a modest improvement with the subunit model, which is still significant (Wilcoxon-signed rank test, $p=3.5165e-05$, $N=52$).

3.5 Discussion

We have demonstrated that convolutional neural network (CNN) models, fit to spiking responses from single LGN neurons, recover biologically plausible nonlinear subunit filters with Gaussian or Difference of Gaussian (DoG) shapes. CNN models of Y-type neurons show a higher degree of spatial nonlinearity within their receptive field compared to X-type neurons, and significantly improve predictive performance in comparison to linear-nonlinear models for Y-type neurons.

3.5.1 Relationship of receptive fields for LGN neurons vs. retinal ganglion cells

Recent studies have suggested that the subcortical Y pathway input to the cortex forms the basis for first- and second-order selectivity of early cortical neurons (Demb et al., 2001a; Rosenberg et al., 2010). In this study our aim was to capture, within an estimated model, the retinal nonlinear processing that makes Y-type subcortical neurons selective for texture as well as luminance changes. To model the signal processing in the Y-pathway neural circuit we

measured spiking responses from Y-type lateral geniculate nucleus (LGN) neurons, which are . comparatively much easier to record from than RGCs, in vivo. Although there are multiple ways of measuring RGC responses in vivo, for example intraretinally (Fukuda and Stone, 1974), from the optic tract (Bullier and Norton, 1979), or as LGN S-potentials (Kaplan and Shapley, 1984), all these methods are challenging, and give relatively low yields of data. However in the LGN we were able to use tetrodes with subsequent spike sorting, to simultaneously record from multiple neurons, substantially increasing our yield of neurons. Furthermore, in the LGN we could get good stability of recordings, enabling data collections for the long durations (~ one hour) needed for system identification.

The signals in the LGN do however reflect additional processing beyond that in the retina. For example, LGN neurons pool inputs from multiple RGC afferents (Martinez et al., 2014). Furthermore, there is feedforward inhibition within the LGN through inhibitory interneurons (Wang et al., 2011). In addition, the LGN also receives feedback from the visual cortex (Murphy & Sillito, 1996). However despite this additional processing, previous studies have shown there is little or no mixing of inputs between the X and the Y pathways, and most LGN cells, despite receiving convergent retinal inputs (Martinez et al., 2014), get predominant excitatory drive from single RGCs (Bullier and Norton, 1979; Mastronarde, 1992). Spatial receptive field properties of RGCs and neurons in the LGN are very similar (So and Shapley, 1981).

3.5.2 Visual stimuli for system identification

Receptive field properties of neurons can be stimulus dependent (David et al., 2004), so it is desirable to study receptive fields under the most naturalistic conditions in order for the

results to be most widely relevant (Touryan and Dan, 2005; Wu et al., 2006). However, Y cells give strong linear responses to the low spatial frequency content in natural images, even though there is a spatial nonlinearity present within their receptive fields. It is only when they are stimulated with high spatial frequencies in the range of the subunit size, that their spatial nonlinearity becomes evident (Enroth-Cugell and Robson, 1966). High spatial frequencies can differentially activate subunits within the receptive field, and the responses of subunits to opposite contrast polarity do not cancel out (Demb et al., 2001a). Here we wanted to resolve subunit structure within the receptive field using system identification, and hence we needed a large set of independent images that could nonlinearly activate the receptive fields. But at the same time, we wanted the visual stimuli to be naturalistic, since artificial stimuli may drive neurons in a non-ecological operating range.

To address the conflicting requirements, we generated random synthetic textures (Figure 3.1) having naturalistic image statistics (Kingdom et al, 2001; Zavitz and Baker, 2013). These textures contained high densities of randomly placed Gabor micropatterns, with their contrast modulated by randomly placed coarse-scale Gabors. An important advantage of these textures was that we could control the spatial frequency content in the images - the size of the texture elements was set to be within the expected range of nonlinear subunit selectivity for the Y-type cell population (So and Shapley, 1981). However, note that we did not customize these images for each neuron based on its spatial frequency selectivity - consequently for some of our sampled neurons these images could still activate relatively strong linear responses. Also, unlike sinusoidal gratings that contain energy only at one spatial frequency, these texture stimuli are broadband, containing texture elements of multiple sizes, so nonlinear responses will not be as strong, and consequently might not stand out from the noise in the estimated filters due to trial-

wise neuronal variability (Faisal et al, 2008; Harris & Thiele, 2011). In such cases our system identification method might fail to resolve the spatial nonlinearity in the receptive field and recover only a linear receptive field.

In this study, new images were presented at 75Hz, so there was no correlation between consecutive images. However during natural viewing the temporal structure of visual stimuli on the retina is controlled by eye, head and body movement as well as object motion. This stimulation is strongly correlated over time, unlike the stimuli here - therefore the temporal dynamics of the receptive fields might be different under natural stimulation. Kording et al (2001) generated natural movies by attaching a camera on a cat's head while the cat explored an outdoor environment. Even though this movie captures the cat's head and body movement, it cannot capture the eye movements. Baudot et al. (2013) used cats' eye movement statistics from the oculomotor literature, and simulated eye movements over static natural scenes to generate movies. They found neuronal responses to be sparse and highly precise. However such temporally correlated stimuli pose difficulties for system identification. Our preliminary attempts using such stimuli recovered receptive fields that were temporally "smeared", as if subjected to a temporal low-pass filter. The reason for this kind of biased estimate using simulated eye movement stimuli is that with temporally correlated image frames, the machine learning algorithm does not get responses to enough novel stimuli. Consequently much longer recording times would be needed to make effective use of such stimuli for experiments of this kind.

3.5.3 Convolutional Neural Networks

Convolutional neural networks with several layers and many filters have been very successful at tasks such as object recognition (LeCunn et al., 2015). These deep neural networks,

typically with millions of parameters, are trained using millions of labeled images of different object categories. However it is not practical to collect neural responses to millions of images in these kinds of experiments. Yamins et al. (2014) instead used the features of a neural network already trained on an object classification task to model receptive fields of neurons in the ventral stream. Neural data from individual neurons were used to learn pooling weights of these features. Interestingly, such models could predict responses of single neurons in area V4 and IT, suggesting that features encoded by artificial neural networks for a classification task might have a high similarity to encoding in the primate ventral stream. A recent study measured selectivity of hidden units in a deep CNN model using stimuli previously used in neurophysiological experiments to characterize shape-tuning, and found that tuning properties of a small number of units were similar to V4 neurons (Pospisil et al., 2016). However, it remains unclear to what extent the encoding at successive layers of these neural networks has any correspondence with hierarchical stages of the primate visual system.

Here we used a highly simplified convolutional neural network with only two layers and one or two filters to model receptive fields of LGN neurons. Instead of using features of a pre-trained neural network, we constructed a neural network architecture based on known biology, and used neuronal data to learn all the model parameters. Despite initializing the weights of the convolutional filters and pooling layers randomly, training always converged onto filters with clear structure, e.g. Gaussian or DoG shapes. Even though optimizing a neural network is not a convex problem (i.e. having multiple local minima), using various regularization methods to avoid overfitting (L2 weight, Poisson noise and Dropout) and the Adam optimizer, these models always converged on very similar filter shapes despite different random initializations.

An important aspect of our model architecture is the use of a parametric rectifier nonlinearity (He et al., 2015). The single parameter, α , of this "pReLU" nonlinearity, enabled us to model receptive fields of different cell types (X and Y) with the same model architecture. This approach is like that used to model simple and complex type visual cortex neurons (Nguyen, 2016). Thus convolutional neural networks are a promising candidate for modeling receptive fields of neurons at early stages of the visual pathway, by building model architectures with biological constraints and training with limited amounts of neural data.

3.5.4 Future directions

Qualitatively we observed that the spatial size of the recovered subunit filters is bigger than expected from the spatial frequency tuning (nonlinear F2-response) to contrast-reversing gratings. This discrepancy could be a result of using texture stimuli with a $1/f$ fall-off in spatial frequency, in which (as discussed earlier), due to the broadband nature of the stimuli, nonlinear responses to high spatial frequencies might be obscured by stronger responses to the low spatial frequencies and the neuronal noise. One way to improve this would be to collect data for longer durations with a larger number of images, and a higher number of repetitions. Such larger datasets might also enable satisfactory estimation of 2-filter models for a larger percentage of the neurons.

In vitro studies of the mouse retina have demonstrated that the intermediate nonlinearity for ON-type Y cells is temporally asymmetric (Borghuis et al., 2013). However, in this study we modeled this nonlinearity with the pReLU, a static nonlinearity. Future studies could attempt to model this dynamic nonlinearity, to better capture the processing in ON-type cells and improve model performance. We made preliminary attempts to capture additional processing at the LGN

stage (e.g. feedforward inhibition, Martinez et al., 2014) by adding an additional convolutional layer to the model (3-layer model). However we did not see any improvement in the predictive performance using this model compared to the 2-layer model. Larger datasets might help in capturing additional processing at the LGN stage.

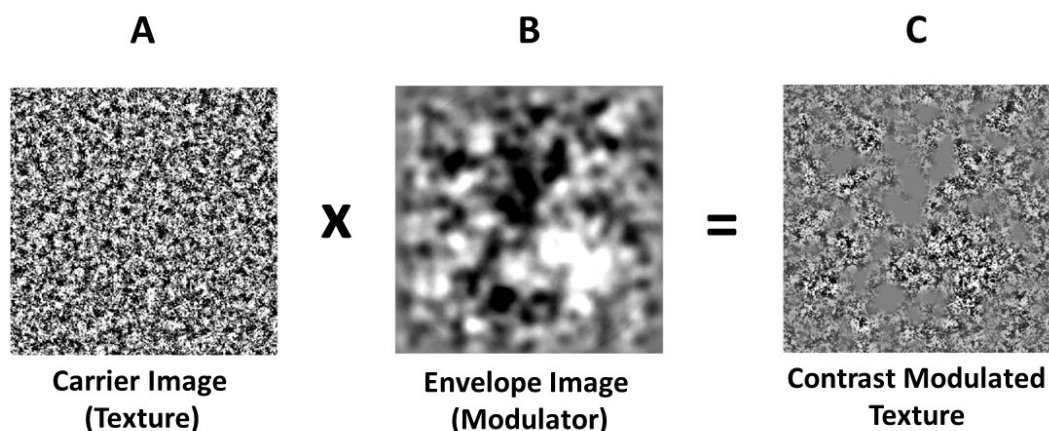


Figure 3.1 : Construction of naturalistic texture stimuli.

(A) Carrier texture patterns were constructed with a high density of randomly positioned Gabor micropatterns, to span all orientations with four different sizes to cover the range of spatial frequencies (0.5cpd to 3cpd) that have been previously shown to activate Y cells nonlinearly. The proportions of Gabors of various sizes was set so that the constructed image had a power spectrum with an approximately $1/f$ fall-off as in natural images. (B) Envelope patterns were constructed in a similar manner with coarse-scale Gabors (0.03 to 0.25 cpd). (C) Finally, the envelope pattern was applied to the carrier pattern to modulate the contrast of the texture. Ensembles of such contrast-modulated texture patterns, each independently constructed, were employed as frames of movie stimuli for system identification experiments.

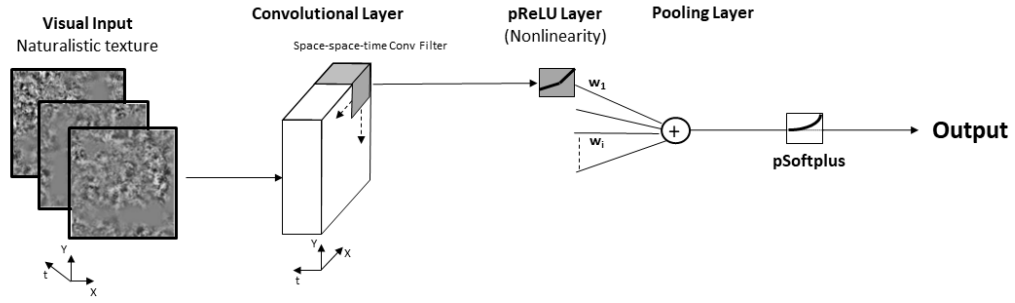


Figure 3.2 : Model architecture of 1-filter convolutional neural network to model LGN receptive fields.

Receptive fields are modeled as a two-stage convolutional neural network, with the first layer consisting of a space-time convolutional filter followed by a parametric rectifier (pReLU) nonlinearity. Outputs from all spatial locations are combined by a Pooling layer (dense layer) and passed through a final output nonlinearity (pSoftplus). Such a neural network was trained on each neuron's spiking responses to naturalistic texture stimuli.

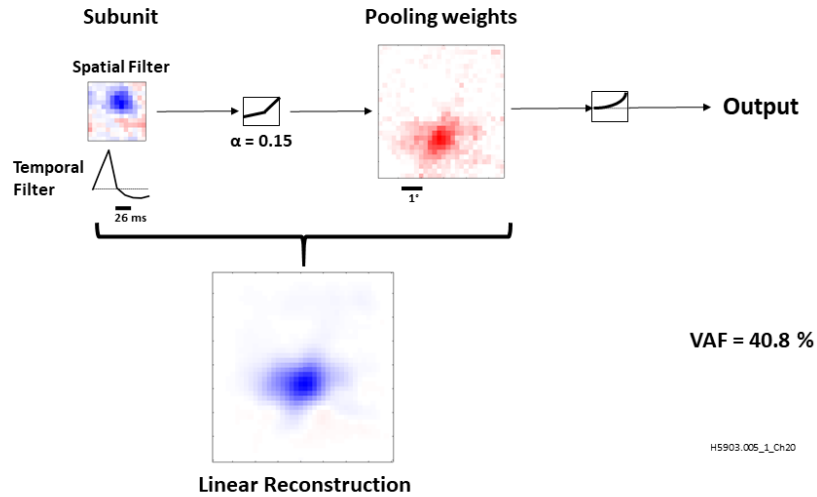


Figure 3.3 : Receptive field model of an example Y cell.

For the subunit filter (top left), the algorithm recovers an OFF-center isotropic spatial filter for the subunit, with a biphasic temporal filter. The α parameter that corresponds to degree of nonlinearity for this model is 0.15, which is close to half-wave rectification. Pooling weights show the spatial location of the subunits, with intensity of pixels indicating their corresponding weights. Since pooling weights were constrained to be positive they are all red in color. Linear reconstruction map (bottom), obtained by convolving subunit spatial filter with pooling weights.

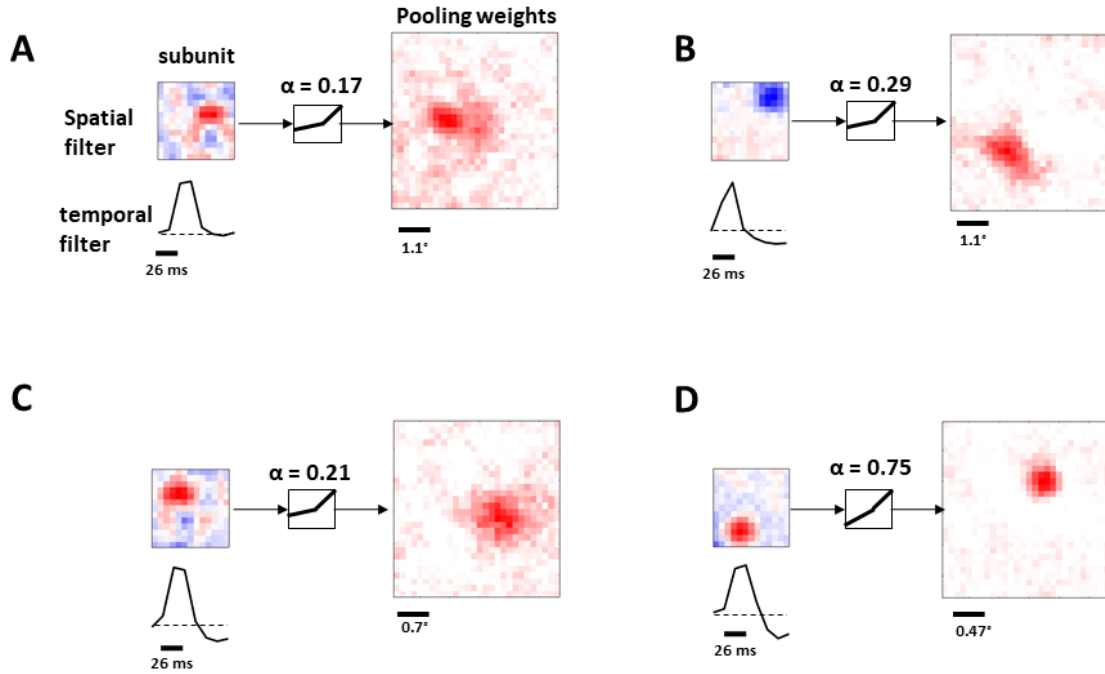


Figure 3.4 : Trained neural networks for four example Y cells.

Recovered subunit spatial filters have a clear isotropic structure, despite no algorithmic constraint on their shape. Some subunits have weak surrounds as in C. Temporal filters for the subunit are usually biphasic (B, C, D), but sometimes monophasic (A). The intermediate nonlinearity parameter α is close to half-wave rectification for some neurons (A, B, C), while for others it is close to linearity (D). VAFs for prediction of Test datasets, 44.44 % for (A), 37 % for (B), 28.5 % for (C), 41.23 % for (D).

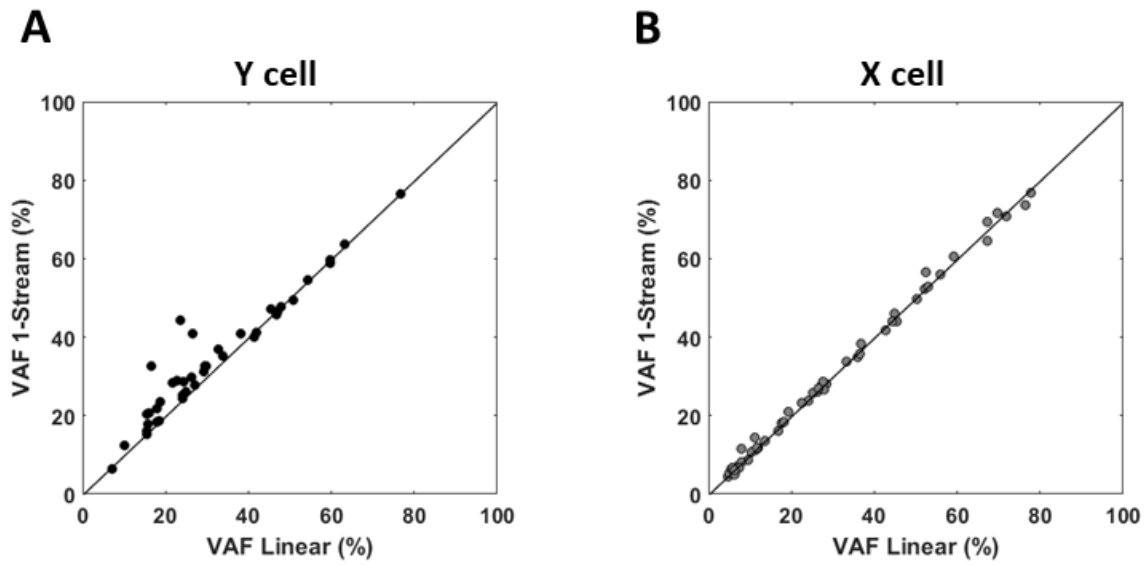


Figure 3.5 : Comparing predictive power of subunit models to linear models in predicting neural responses.

Scatterplots comparing neurons' VAF values for the subunit model and the linear model. (A) For Y-type cells most of the points lie above the 1:1 line (N=41), indicating that the subunit model performs better than the linear model at predicting responses to novel stimuli. (B) For X-type cells most of the points lie close to the 1:1 line (N=52), suggesting that subunit and linear models perform equally well at predicting responses to novel stimuli.

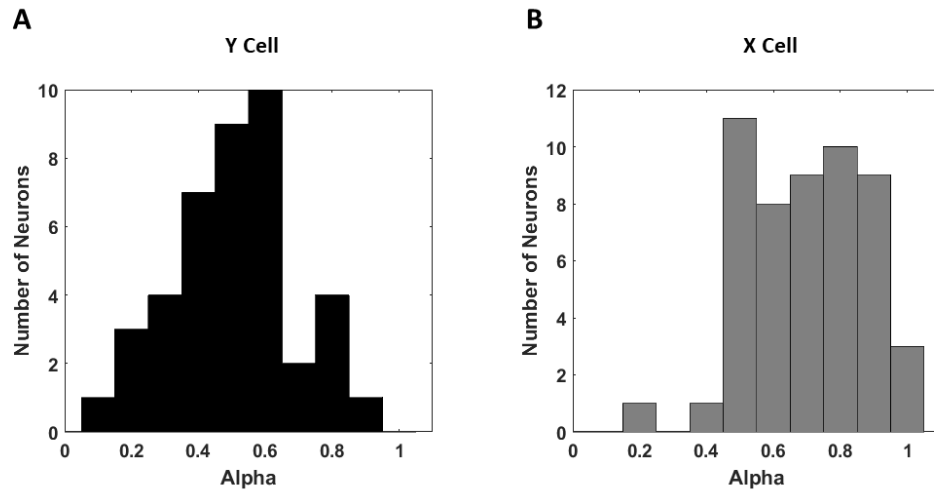


Figure 3.6 : Degree of spatial nonlinearity within receptive fields of X- and Y-type neurons.

Distributions of α values of parametric rectifier nonlinearity (pReLU) for Y- and X-type cells.

For all the neurons in our sample (N=93), α values fell between zero, corresponding to half-wave rectification, and a value close to one, indicating linearity. **(A)** For most of the Y-type cells, α was less than 0.5. **(B)** For X-type cells α was usually greater than 0.5. The α values for Y-type cells (mean = 0.5, SD = 0.19) were significantly less than for X-type cells (mean = 0.69, SD = 0.17) ($p = 2.6335e-06$, Mann-Whitney U test).

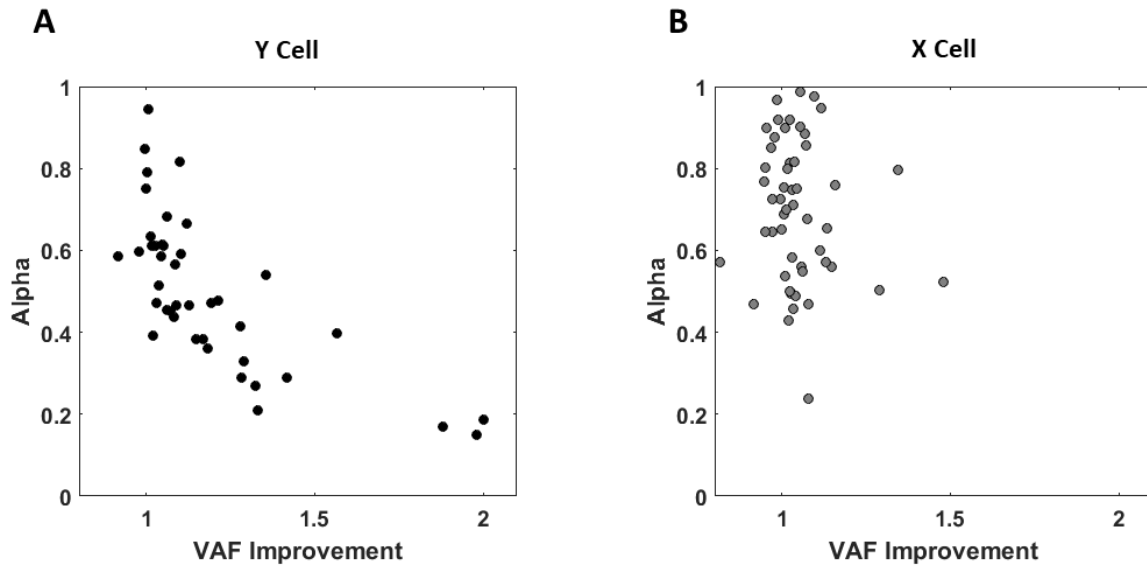


Figure 3.7 : Relationship between model performance improvement and nonlinearity.

Scatterplots comparing α values of the subunit models and their improvement in predictive performance compared to linear models. (A) For Y-type neurons there is a negative correlation between α and VAF improvement. Models with α closer to zero (more rectification) have higher improvement in their predictive performance. (B) For X-type neurons there is no such systematic relationship.

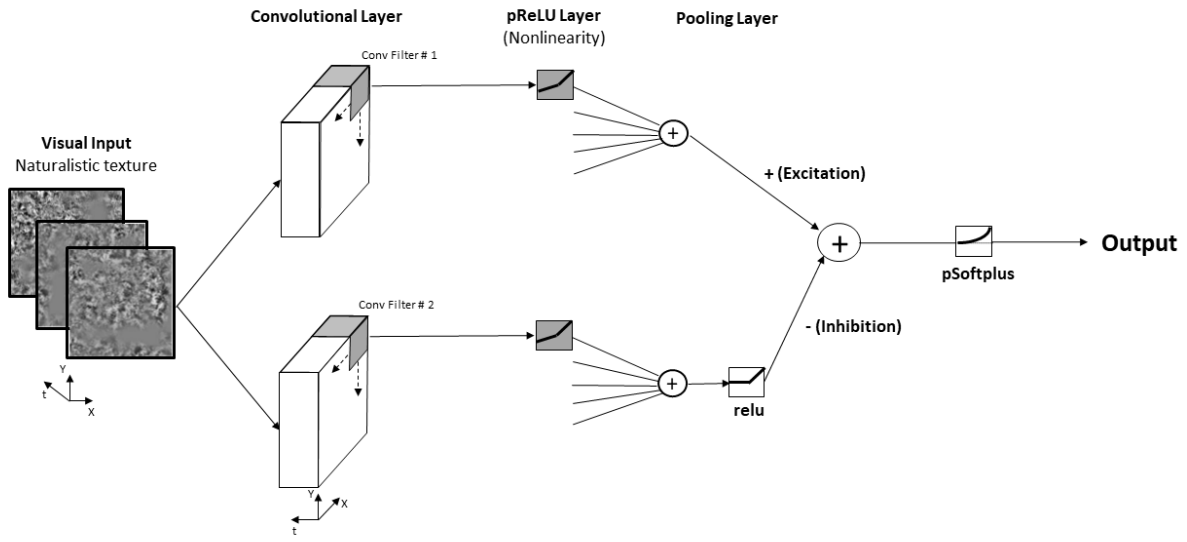


Figure 3.8 : Model architecture of a 2-filter convolutional neural network to model LGN receptive fields.

This model has an additional convolutional filter, with a distinct PReLU nonlinearity and pooling layer, acting in parallel with those of the model shown in Figure 3.2. Both filters are identical except one filter provides excitation and other provides inhibition. This model is based on known retinal circuitry, in which ganglion cells receive direct excitatory inputs from bipolar cells and indirect inhibitory inputs via spiking amacrine cells. Half-wave rectification (ReLU) in the inhibitory filter captures the rectification at the output of amacrine cells.

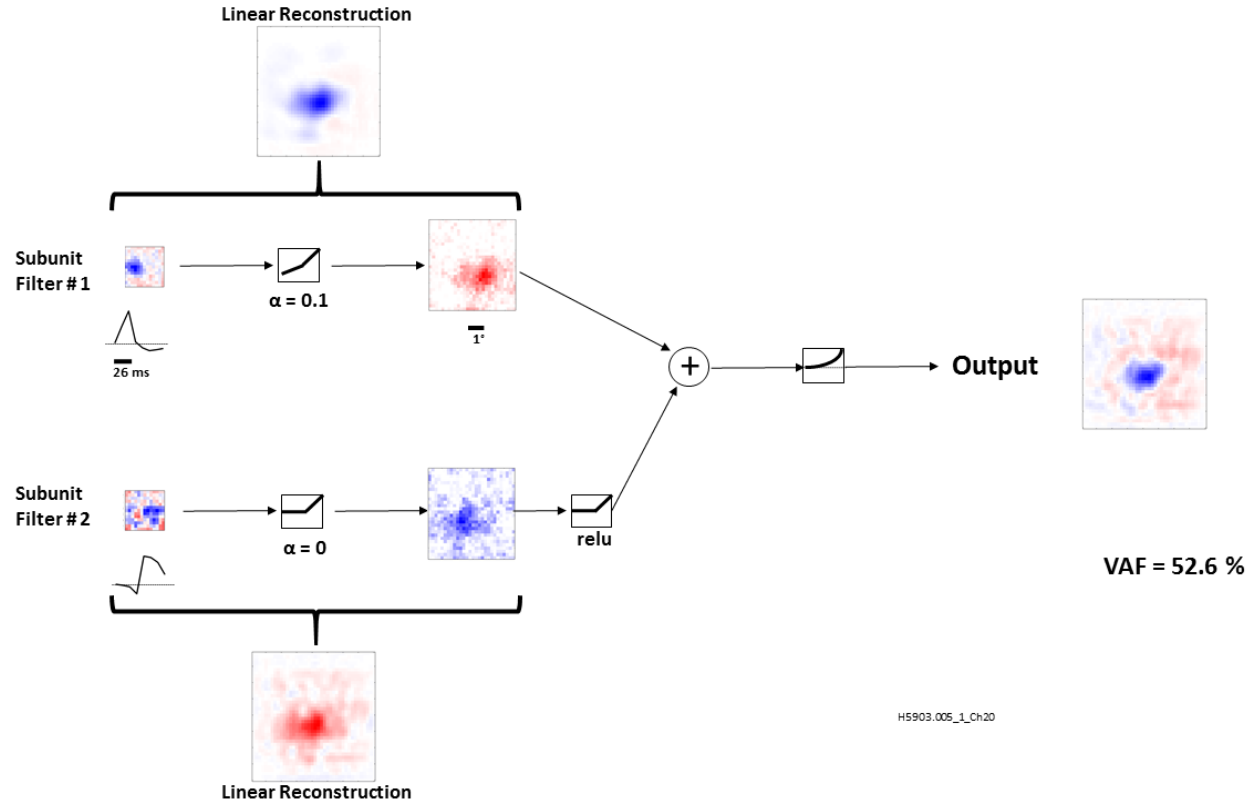


Figure 3.9 : 2-filter receptive field model for example neuron of Figure 3.3.

The spatial and temporal subunit filters recovered for the excitatory filter are very similar to those recovered in the 1-stream model for this neuron (Figure 3.3). Interestingly, for the inhibitory filter (#2) the temporal subunit filter has delayed dynamics compared to the excitatory filter. In addition, the linear reconstruction for the inhibitory filter covers a wider spatial region, so that the combined linear reconstruction (rightmost plot) shows an OFF-centre, ON-surround organization.

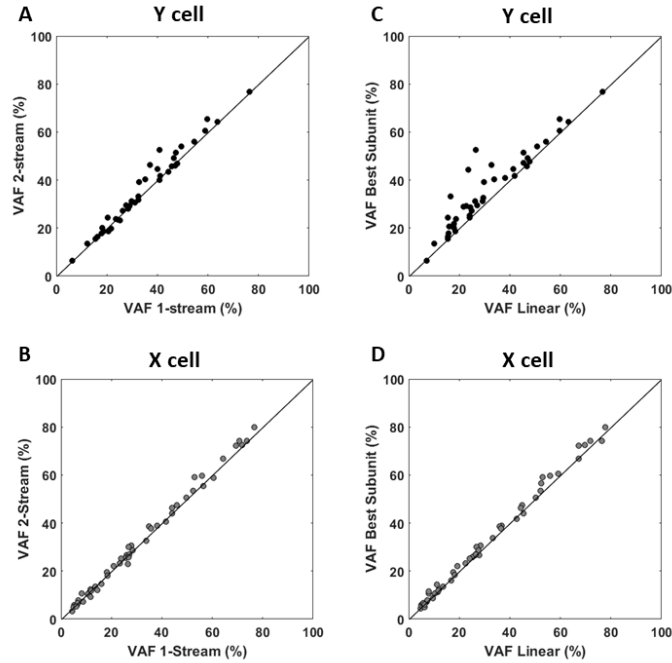


Figure 3.10 : Comparison of predictive performance across different receptive field models.

(A) Comparison of predictive performance of 2-filter versus 1-filter models for Y-type neurons (N=41). For a subset of these neurons there is a significant improvement in the VAF for the 2-filter model. **(B)** For X-type neurons (N=52) there is not a substantial improvement in VAF with the 2-filter model. **(C)** VAF values for best subunit model (between 1- and 2-filter) vs. linear model for Y-type neurons (N=41). Clearly, the nonlinear subunit model performs better at predicting responses of Y-type neurons. **(D)** For X-type neurons (N=52) there is only a slight improvement with the subunit model compared to the linear model.

4

Model simulations from Chapter 2 demonstrated that if inputs from ON and OFF-center Y cells are pooled in an unbalanced push-pull manner, then cortical receptive fields can have cue-invariant orientation selectivity. So in this chapter, I evaluate the integration of ON- and OFF-pathway inputs by individual neurons in early cortical areas of the cat (Area 17 and Area 18) using novel system identification method. The estimated models reveal spatial arrangements of ON and OFF inputs that are consistent with a push-pull model, but significant asymmetries in the relative strengths of ON and OFF signals, within simple cell receptive fields. These observed asymmetries could provide the neural mechanism for generating cue-invariant receptive fields from Y-pathway inputs. This chapter is under preparation for submission to a peer reviewed journal.

CHAPTER 4

Separating ON and OFF pathway inputs to cortical simple cells reveals receptive fields with asymmetric push-pull.

4.1 Abstract

Simple cells in the early visual cortex are conventionally thought of as linear Gabor-like spatial filters with an output nonlinearity. This linearity across the receptive field is thought to arise from a symmetric "push-pull" arrangement of inputs from ON and OFF pathways. However the generality of this finding to a wider sample of neurons, and to other early cortical areas, remains unclear. We used 32-channel multielectrodes (polytrodes and linear arrays) to record extracellular single-unit responses of cat Area 17 & 18 simple cells to natural image sequences. We estimated a two-stage (LNLN) receptive field model of individual simple cells using regularized gradient descent optimization. In this model, first stage filters correspond to receptive fields of ON- and OFF-center lateral geniculate nucleus afferents, while the second stage (weight map) corresponds to the spatial layout of their summation by the cortical neuron. This model enabled us to visualize the spatial arrangement of excitatory and inhibitory inputs from ON and OFF subcortical pathways to individual receptive fields. The estimated models reveal significant asymmetries in spatiotemporal integration of ON and OFF signals within simple cell receptive fields. These observed asymmetries could provide the neural mechanism for generating cue-invariant receptive fields from Y-pathway inputs.

4.2 Introduction

Light and dark parts of visual images are encoded separately in the retina by distinct populations of neurons having ON- and OFF-center receptive fields. These ON and OFF pathways that are thought to have very little interaction until they reach primary visual cortex, V1 (Schiller, 2010), where they are combined to build receptive fields selective for features such as orientation and spatial frequency. Simple cells in V1 are conventionally thought to combine inputs from these two pathways in a "push-pull" manner, i.e. at each spatial location in the receptive field, excitatory input from one pathway is complemented, or balanced, by inhibitory input from the other pathway (Ferster, 1988; Hirsch et al. 1998; Martinez et al, 2005). This arrangement allows simple cells to have spatially linear receptive fields despite integrating inputs from spiking neurons that have rectified outputs.

However, there exist spatial and temporal asymmetries between the ON and OFF pathways. For example, OFF pathway signals reach cortex approximately 3-6 ms before the ON pathway signals (Jin et al., 2011). ON ganglion cells have receptive fields with 20% larger diameter than those of OFF cells (Chichilnisky and Kalmar, 2002). OFF-dominated simple cells tend to have faster dynamics (Komban et al., 2014), and are in higher proportion in the superficial layers of V1 (Yeh and Shapley, 2009), than ON-dominant neurons. Most previous studies of ON/OFF pathways used white noise stimuli to map receptive fields with reverse-correlation. However natural scenes contain very different image statistics (Field, 1987; Thomson, 2001) that may drive visual neurons differently (Felsen et al., 2005).

Further it remains unclear whether these asymmetries in processing dark and light are present within individual receptive fields of cortical simple cells after integration of ON and OFF

inputs, or if these asymmetries are compensated by cortical mechanisms. Shariati & Freeman (2012) have suggested that ON/OFF temporal asymmetry within a receptive field could provide a mechanism for direction selectivity in cortical neurons. Gharat & Baker (2017) proposed that if ON and OFF inputs from subcortical Y-cells are combined in an unbalanced push-pull manner, it could result in receptive fields that have cue-invariant orientation selectivity for luminance as well as texture boundaries.

Here we separate the contributions of ON and OFF pathway inputs to cat Area 17 and 18 simple cells, while viewing natural images. We model a simple cell's receptive field as a weighted linear sum of simulated ON and OFF inputs, tiled across visual space. A gradient descent optimization method with regularization (to avoid overfitting) is used to estimate separate spatiotemporal maps of each simple cell's summation weights for these ON and OFF inputs, from its responses to natural images. With this method, we can map both excitatory as well as inhibitory contributions of the ON and OFF pathways within individual receptive fields. We find that for many neurons, maps of ON and OFF weights are approximately identical in spatial layout and magnitude, but are of opposite polarity (“push-pull”). However, there is a large diversity in the ratio of ON vs. OFF strengths, and a large subset of neurons have unbalanced push-pull. We also find asymmetries between temporal dynamics of ON and OFF weight maps. These results demonstrate spatiotemporal asymmetries in how visual cortex neurons integrate signals from light and dark regions within natural scenes, which might contribute to receptive field properties such as direction selectivity and cue-invariant second-order processing.

4.3 Methods

4.3.1 Animal Preparation and Electrophysiological Recordings

All the animal procedures for this study were approved by the Animal Care Committee of McGill University and are in accordance with the guidelines of the Canadian Council on Animal Care. Animal setup and electrophysiological recordings were identical to those of our previous study (Gharat & Baker, 2017). A craniotomy and durotomy were performed on cats for electrode placement in Area 17 (H-C P3-L1) and Area 18 (H-C A3/L4). Electrophysiological recordings were performed using 32-channel multielectrodes (NeuroNexus), either linear arrays (A1x32) with 100 μ m spacing or polytrodes (A1x32 poly2) with 50 μ m spacing, using Plexon Recorder for data acquisition (3 Hz to 8kHz; sampling rate, 40 kHz). Broadband signals from all 32 channels, along with the photocell data, were streamed to the hard drive and saved for later analysis. For this study we recorded from 97 simple cells (44 cells in Area 17 and 53 cells in Area 18) in 19 cats. On these cats data collection was also performed by other lab members for their projects.

4.3.2 Visual Stimuli and Recording Protocol

Visual stimuli were generated and presented as previously described (Gharat & Baker, 2017). Multi-electrodes were inserted perpendicular to the brain surface so that spatial locations of neuronal receptive fields at different channels were similar. Multi-unit spikes from one of the visually responsive channels of the electrode were used to guide the recording protocol. First, manually controlled flashing and drifting black or white bars were displayed on a grey-background CRT monitor (NEC FP1350, 20 inches, 640 x 480 pixels, 150 Hz, 36 cd/m², 57 cm viewing distance) to roughly map the spatial location of the multi-unit receptive field. The

dominant eye was identified, and the other eye then occluded. Subsequent stimuli were presented under computer control, centered on the location of the receptive field.

Greyscale natural images were presented on the CRT monitor. These images covered 30° of visual angle with a resolution of 480x480 pixels. These images were extracted from photographs in the McGill Calibrated Color Image Database (Olmos & Kingdom, 2004) and converted to greyscale. The mean intensity of each image was equated, and the images were normalized to have the same standard deviation. An ensemble of 375 randomly selected images was used to form a single movie clip, presented at 75 Hz for a 5 sec trial. Note that there was no temporal correlation between the consecutive frames of each movie clip. 30 such movie clips were generated, as previously (Talebi and Baker, 2016). 20 of these movies were repeated 5 times and responses to these movies were used for training the model. Responses to the other two sets of 5 movies, each repeated 20 times, were used for regularization and for testing of the model. Presentations of movies to be used for training, regularization and test datasets were all randomly interleaved, to minimize effects of neuronal gain change over time. It took around 45 minutes to collect these datasets

4.3.3 Receptive Field Model Estimation

Spike waveforms were carefully classified from these datasets to isolate responses from single neurons, using SpikeSorter software (Swindale & Spacek, 2014). Spike times were binned (using bin width equal to the duration of each image, i.e. 13.3 ms) to calculate spike rate of neurons. The training dataset was used for optimizing the model parameters by minimizing the mean-squared-error (MSE) between the actual neuronal response and the predicted response of the model, using gradient descent optimization. The regularization dataset was used to decide the

number of training iterations, using early stopping - i.e. the training was stopped when its predictive performance ceased to improve on the regularization dataset. The test dataset was used to evaluate the predictive performance of the estimated model. The MATLAB toolbox STRFlab (<http://strflab.berkeley.edu>) was used for performing optimization of the model.

4.3.4 Linear-nonlinear receptive field model

First a linear-nonlinear (LN) model was estimated, as described in Talebi & Baker (2012). Here the receptive field is modeled as a space-space-time linear spatiotemporal filter, followed by a static output nonlinearity. Since the measured neural responses were obtained to natural images, this model was estimated using a regression framework to avoid bias from the stimulus autocorrelation (Wu et al., 2006). To estimate the model, stimulus images were first down-sampled to 30x30 pixels. The linear filter part of the model was estimated for 8 time lags spaced by 13.3 msec (75Hz) - thus the estimated filter model had 30x30x8 (7200) parameters. After estimation of the linear filter, the output nonlinearity was estimated by fitting a half-rectified power law, which contained only one free parameter (Talebi & Baker, 2012). This estimated linear filter was further refined by defining a square cropping window around the filter and re-estimating the linear filter by repeating the above-mentioned procedure.

Only neurons whose estimated linear filter showed clear, orientated receptive fields with segregated positive (ON) and negative (OFF) regions in their receptive fields were classified as simple cells and included in further analysis. An optimal spatial frequency value was obtained for each estimated filter at the optimal temporal lag, from its Fourier transform (Talebi & Baker, 2016).

4.3.5 ON-OFF subunit model

Receptive fields of simple cells in the visual cortex are thought to have a "push-pull" arrangement of ON- and OFF-pathway inputs in their receptive fields, as shown in Figure 4.1. In such an arrangement, spatial locations in the receptive field with ON-pathway excitation are balanced by equal inhibition from OFF-pathway inputs, and vice-versa - thus generating linear receptive fields from nonlinear (i.e. rectified) ON- and OFF-pathway inputs (Ferster, 1988). Here, we test this balance between ON- and OFF-pathway inputs across both the space and time dimensions of the receptive field. The simple cell receptive fields were modeled as a linear weighted sum of ON- and OFF-centre LGN afferents as shown in Figure 4.2. The receptive fields of these afferents were modeled as 2D spatial Gaussian filters, whose outputs were half-wave rectified - the antagonistic surrounds were neglected, to simplify the model parameter estimation, since LGN cells typically have weaker surrounds in comparison to their centres (So and Shapley, 1981). Also, the spatial nonlinearity of the Y pathway was not included in the models of afferents. The size parameter of the Gaussian (σ) was set separately for each neuron, based on its estimated linear filter in the LN model, estimated as described earlier - the Gaussian σ was inversely proportional to the optimal spatial frequency, such that the subunit size was well within individual ON or OFF regions (Reid & Alonso, 1995). The banks of such simulated ON and OFF subunits were uniformly tiled on a square grid with spacing of 2σ (Chichilnisky, 2001; Ringach and Shapley, 2004, Anishchenko et al., 2010).

The weights of the ON and OFF subunits across space and time were estimated within a regression framework as described in the earlier section, with a Generalized Linear Model. The weights in this model were estimated by minimizing the mean square error between actual neuronal response and predicted response, on the training dataset. These estimated weights could

be positive (excitatory) or negative (inhibitory). Early stopping regularization was used to avoid overfitting. These estimated models were then evaluated on a separate holdback test dataset, which was not used for training or regularizing the model parameters. Performance of each fitted model was quantified as percent variance accounted for (% VAF), calculated as the square of the correlation coefficient (R^2) between actual neuronal response and predicted response.

4.3.6 Receptive Field Analysis

Figure 4.3 shows an example ON-OFF subunit receptive field estimated for a typical simple type cell. These colormaps indicate spatial maps of ON and OFF subunit weights across successive temporal lags, with each pixel denoting the weight of a Gaussian subunit centred at the pixel location. The colormaps in the top row indicate weights of ON subunits, whereas the bottom row indicates weights of OFF subunits - red indicates positive weights and blue indicates negative weights.

To assess spatial properties and conformity to the push-pull scheme, 2D spatial weight maps ("wMaps") were extracted for the ON and OFF inputs at optimal time lags in the space-space-time dependence. The chosen time lags were the ones with maximum variance. For only 15 out of 97 neurons optimal time lags for ON and OFF wMaps differed by more than one lag. In the example of Figure 4.3, this was the third lag (denoted by dotted squares) for both the ON and the OFF inputs.

For each neuron, the wMaps at these optimal lags were used to measure the strength of ON and OFF inputs. We defined an ON-OFF strength index 'S' to measure relative contribution of ON- and OFF-pathway inputs to a given neuron's estimated receptive field model:

$$S = \frac{Var(ON) - Var(OFF)}{Var(ON) + Var(OFF)}$$

where $Var(ON)$ is the variance of the ON wMap at the optimal time lag for ON inputs, and $Var(OFF)$ is variance of the OFF wMap at optimal time lag for OFF inputs. Note that this index takes into account both excitation as well as inhibition from each pathway.

To evaluate the relative spatial arrangement of excitation and inhibition of ON and OFF inputs within the receptive field, we calculated the Pearson's correlation (R) between ON and OFF wMaps at their respective optimal lags.

We also compared the temporal dynamics of ON and OFF inputs within individual neurons' receptive fields. To do this we decomposed the spatio-temporal (3D) weight maps into a 2D spatial map and 1D temporal filter, using singular value decomposition (SVD) (Mazer et al., 2002). For the example neuron in Figure 4.3, the extracted temporal dynamics for ON (red) and OFF (blue) inputs are shown on the right side. These plots were fitted with cubic spline interpolation using MATLAB's `csapi` function (Yeh et al., 2009). From these temporal functions, we obtained the peak times for the ON and OFF pathway inputs. To quantify the monophasic/biphasic nature of these temporal functions, we calculated an Amplitude Ratio and an Integral Ratio (Komban et al., 2014). The Amplitude Ratio is calculated by dividing the suppression amplitude by the peak amplitude. The Integral Ratio is calculated by dividing the suppression integral by the peak integral.

4.4 Results

4.4.1 ON-OFF subunit model

Receptive fields of oriented simple type cells were modeled as weighted linear sums of rectified ON- and OFF-centre LGN afferents. For each neuron, the weights of these afferent inputs were estimated using regularized linear regression. Spiking responses of each neuron to natural image sequences were used to optimize the model through gradient descent (see Methods section). This algorithm estimates space-space-time weights for both ON and OFF inputs, allowing us to compare spatiotemporal integration of ON and OFF signals within individual simple cell receptive fields.

Figure 4.3 shows spatial weight maps of ON and OFF inputs across eight time lags for an example simple cell receptive field. The color of each pixel in the spatial maps denotes the estimated weight of a subunit at that spatial location and time lag. Red colored pixels denote positive weights (excitation) and blue denotes negative weights (inhibition) - note that the model recovers both excitation as well as inhibition of both ON and OFF inputs within the receptive field. In this example, the weight maps show near-horizontally oriented regions for both the ON and OFF inputs. The two dotted squares highlight the spatial weight maps at their respective optimal time lags. Note that optimal lag for ON and OFF inputs was separately determined even though in this example the same (3rd) lag was optimal for both. For only 15 out of 97 neurons optimal lag differed for ON and OFF inputs by more than one. In this example, for ON inputs the elongated excitatory region is above the inhibitory regions - but for OFF inputs the symmetrically opposite pattern is evident, with the inhibitory region above the excitatory. Thus the ON and OFF inputs form a push-pull spatial arrangement within this receptive field.

From the spatio-temporal weight maps we also extracted temporal dynamics of the ON and OFF inputs using singular value decomposition (SVD) (Mazer et al., 2002). The graph on the right side in Figure 4.3 shows the temporal dynamics for ON and OFF inputs in red and blue, respectively. Temporal filters for both ON and OFF inputs in this example are quite similar, though the delay was clearly longer for ON (red) compared to OFF (blue) inputs. Both ON and OFF inputs have small undershoots, slightly more pronounced for the ON. Thus this model framework allows us to visualize the push-pull arrangement of ON and OFF inputs in the temporal dimension as well.

4.4.2 ON-OFF spatial arrangement

Spatial weight maps for ON and OFF inputs at their optimal time lags were used to compare spatial relationships and relative strengths of ON and OFF inputs. Figure 4.4 shows examples of ON and OFF spatial weight maps for six example simple cells. Like the example in Figure 4.3, these neurons also have similar shapes for ON and OFF weight maps, with opposite polarity (“push-pull”). However the relative strengths of ON and OFF inputs varies across these examples. To summarize the similarity in the spatial arrangement of ON and OFF subunits, we estimated a correlation index (r) between the two weight maps. Correlation (r) values for these examples are generally close to -1, indicating a strong negative correlation between ON and OFF weight maps as expected from a push-pull arrangement.

To summarize the relative strengths of ON and OFF inputs within a receptive field, we defined an index, the ON-OFF strength (S), having values ranging from -1 (OFF-dominated) to +1 (ON-dominated). The neurons in Figure 4.4A and 4B have strength index (S) values close to zero, suggesting equal contributions from ON and OFF inputs to their receptive fields. However,

the neurons in panels 4C and 4D have S values less than zero, suggesting stronger contributions from OFF inputs compared to ON inputs. On the other hand, the neurons in panels 4E and 4F have S values greater than zero suggesting stronger contributions of ON compared to OFF inputs.

In Figure 4.5, values for the r and S indices are summarized for the sample of neurons from Area 17 (N=44) and Area 18 (N= 53). Panel A shows the distribution of correlation values (r) between ON and OFF weight maps. Most of the neurons in our sample have strong negative correlation, consistent with a push-pull spatial arrangement. However, a minority of neurons (14/97) have a positive correlation between ON and OFF weight maps. This group of neurons also had a large difference between the temporal dynamics for ON and OFF inputs and will be discussed in the following section. The mean r value for the Area 17 sample is -0.53 and for Area 18, -0.36, but the difference between the means is not significantly different (two-sampled t-test, $p=0.1$).

The distribution of ON-OFF strength (S) values is shown in the histogram of Figure 4.5B. S is broadly distributed with both positive and negative values across our sample of neurons. The mean S values for the Area 17 and 18 samples (-0.07 and -0.09 respectively) are not significantly different (two sample t-test, $p = 0.8$). The mean S value for the total sample is -0.08, with a broad distribution of S values ($SD=0.35$). The mean value of S is significantly less than zero (one-tailed t-test, $p = 0.016$, $N=97$), suggesting a slight dominance of OFF inputs at the population level. These results indicate that even though most of the simple cells have a push-pull spatial arrangement (negative r values), the relative strength of ON and OFF inputs is highly variable across individual neurons.

Conceivably the observed imbalance in ON and OFF inputs could result from variance in the estimates of the receptive field models. One way to evaluate an estimated model is to measure its predictive ability on a test dataset not used for training - if a model has low predictive ability then its estimated parameters might be noisy. Figure 4.5C shows a scatterplot of ON-OFF strength (S) of each neuron and its predictive performance measured as Variance Accounted For (VAF). As can be seen in the scatterplot there was no systematic relationship between S and VAF ($r=0.06$, $p=0.6$, $N=97$) - neurons with high as well as low VAF have unequal ON-OFF strength, suggesting that the observed imbalance between ON and OFF inputs was not a result of noisy model estimates.

4.4.3 ON-OFF temporal dynamics

The above analysis described spatial asymmetries between ON and OFF inputs within simple cell receptive fields, but from the example already seen in Figure 4.3, some neurons also exhibit asymmetries between temporal dynamics of ON and OFF inputs. Figure 4.6 shows temporal profiles for four example neurons, extracted using singular value decomposition (SVD) on the space-space-time weight maps. The neuron in panel A has very similar temporal dynamics for both ON (red) and OFF (blue) inputs, although the OFF peaks slightly earlier than ON. The neuron in panel B has similar dynamics for the initial positive part, but a much greater subsequent undershoot for OFF than for ON inputs. The neuron in panel C has faster dynamics for OFF inputs, which are monophasic for OFF inputs but biphasic for ON inputs. A minority of neurons in our sample show large differences in peak times of the ON and OFF inputs, as shown in panel (D). This group of neurons are the ones pointed out earlier, that show positive

correlation between spatial weight maps for ON and OFF inputs - i.e. they do not show a push-pull arrangement.

The above examples show prominent differences between ON and OFF dynamics in two respects: different latencies and the degree of the undershoot. To summarize these differences in temporal dynamics of ON and OFF inputs across our sample we estimated three indices from these temporal profiles – peak time, amplitude ratio and integral ratio (see Methods). The peak time captures the latency to peak and the degree of the undershoot is captured by the amplitude and integral ratios. Neurons' sustained or transient response to step stimuli depends upon the integral of undershoot in comparison to the integral of peak (Schmid AM et al., 2009). Figure 4.7A compares peak times for ON and OFF inputs. Most of the points lie close to the 1:1 line, suggesting very similar peak times for ON and OFF. However a small number of neurons show large differences between their peak times, as pointed earlier for the neuron in Figure 4.6D. The sample means of ON and OFF peak times (38.99 and 36.33, respectively) are not significantly different (paired two-tail t-test, $p=0.35$, $N=97$).

Figure 4.7B shows a scatterplot comparing amplitude ratios for ON and OFF temporal profiles. An amplitude ratio close to zero indicates a small undershoot amplitude compared to the peak amplitude, while a ratio close to one indicates equal undershoot and peak amplitude. ON and OFF amplitude ratios for most of the neurons are less than 0.5 - i.e. the peak amplitude is substantially greater than the undershoot amplitude. There is a large scatter between ON and OFF amplitude ratios, and most points lie away from the 1:1 line. However, there is no systematic trend between ON vs OFF ratios. The sample means of the ON and OFF amplitude ratios (0.44 and 0.36, respectively) are not significantly different (paired two-tailed t-test, $p = 0.76$, $N=97$).

We also estimated an integral ratio, based on the areas of the initial positive vs subsequent negative parts of the temporal dynamics. The scatterplot in Figure 4.7C of integral ratios shows similar scatter to that in Figure 4.7B. These results suggest that even though peak times for ON and OFF inputs are similar, they often differ in subsequent undershoots within individual receptive fields. Thus most simple cells have substantial push-pull imbalance from the undershoots of their temporal profiles.

4.5 Discussion

In this study, we modeled receptive fields of simple cells in cat Area 17 and 18 as a linear weighted sum of simulated ON and OFF pathway afferents, to evaluate the extent to which they conform to a push-pull arrangement. Using the GLM framework, we could recover both excitation as well as inhibition within receptive fields from extracellularly recorded visual responses to natural scenes. We demonstrated that a majority of the simple cells have very similar spatial arrangement of ON and OFF pathway inputs, in a push-pull manner. However there was considerable diversity among neurons in terms of the relative strength of ON and OFF inputs within their receptive fields, with a slight bias for OFF inputs in our sample. In addition, we also found significant asymmetries in the temporal integration of ON and OFF inputs within receptive fields, especially during the undershoot of the temporal response.

4.5.1 Limitations

Here we modeled each simple cell as a linear weighted sum of simulated ON and OFF afferent inputs. However, there is a growing evidence that inputs might not be linearly integrated

in cortical neurons, for example due to dendritic nonlinearities (Wilson et al., 2016). Synaptic inputs clustered together on nearby parts of the dendrites tend to interact nonlinearly. For example, spatial clustering of synaptic inputs with similar orientation selectivity on the dendrites have been shown to nonlinearly affect a neuron's orientation selectivity (Wilson et al., 2016). In addition, dendritic spikes have been shown to make a significant contribution towards enhancing orientation selectivity of neurons (Smith et al., 2013). Such phenomena can lead to nonlinear interactions between synaptic inputs, and our assumption of linearity of summation will not hold. However it remains unclear how best to incorporate such nonlinearities in a model framework whose parameters can readily be estimated, for example with machine learning.

We modeled ON and OFF afferent inputs as Gaussian spatial filters, ignoring their surrounds to simplify parameter optimization. Therefore a part of the inhibition measured in the weight maps is presumably arising from the subcortical surround mechanisms, and in the approach used here we could not separate out these two sources of inhibition. Although, contribution of the surrounds may be somewhat limited as surrounds are typically weaker than the centers.

Instead of white noise stimuli often used in previous studies, here we used a rapid sequence of static natural images, which can better activate a wider range of neurons and yield models with more robust predictive power (Touryan et al., 2005; Talebi & Baker, 2012). However in the real world visual scenes are dynamic in more complex ways, due to motion of objects, eye movements, and the observer's movement through the world. So it is possible that the estimated temporal dynamics of the neurons and how they integrate ON and OFF signals might differ under such conditions. In addition, due to the lack of continuous motion in our stimuli, we might have undersampled neurons that are motion-selective. Future efforts might

benefit from providing more naturalistic visual stimuli to drive neurons in a more appropriate manner.

In this study, we separately evaluated push-pull along spatial and temporal dimensions by using spatial maps at optimal time lags and temporal filters extracted with SVD analysis. The SVD method decomposes a spatiotemporal receptive field matrix into multiple orthogonal components each containing 2D spatial filter and 1D temporal filter. For space-time separable RFs, the first principal component of SVD accounts for the majority of the power of the RF matrix. But for space-time inseparable RFs, higher-order components also contain a significant fraction of the power (Wolfe and Palmer, 1998). Here in this study, we compared temporal profiles from only the first component of ON and OFF weight maps. This comparison could be problematic for neurons that do not have space-time separable weight maps (e.g., direction selective neurons - Valois et al., 2000). On average in our sample neurons, the fraction of power accounted by the first component was 58 ± 16 % (mean $\pm$ SD) for ON weight maps and 64 ± 13 % for OFF weight maps. Along with nonseparability, noisy estimates of the weight maps also leads to less than 100% of power accounted. OFF maps were significantly more separable than ON maps (one tailed t-test, $p = 0.0012$). Since there is no clear way to extract a spatial and a temporal profile for inseparable weight maps, another possible way of evaluating push-pull between ON and OFF inputs could be to collapse 2D spatial weight maps along the optimal RF orientation at each time lag and generate 2D space-time maps (Valois et al., 2000). This method would be able to effectively handle neurons with space-time separable as well as inseparable weight maps.

4.5.2 Comparison to previous findings on ON-OFF integration by cortical neurons

Ringach and Mata (2005) estimated separate ON and OFF maps for receptive fields of neurons in macaque V1, to evaluate the simple vs. complex cell dichotomy. These ON and OFF maps were estimated from visual responses to sparse random white and black dots using reverse correlation, by separately correlating responses with white and black parts of the stimuli. From these maps, they were able to visualize a push-pull arrangement within the receptive fields of simple cells, and found a negative correlation between ON and OFF maps as also demonstrated by our method. The approach used here goes beyond this earlier work by comparing the relative strength of ON and OFF inputs, as well as their temporal dynamics. In addition, the ON and OFF maps recovered with this method were estimated from responses to natural images, which may be more effective in driving cortical neurons as well as producing estimated models that are more robust in predicting responses to other stimuli (Talebi & Baker, 2012).

Yeh et al. (2009) reported the presence of OFF-dominated neurons in layer 2-3 of macaque V1. They used reverse correlation with sparse noise stimuli containing only a single white or black square on each frame. However, with their method, they could not recover inhibition within the receptive fields, and the estimated temporal functions did not have an undershoot. In addition, they did not distinguish simple vs. complex cells in their sample. Nevertheless, they found neurons with ON or OFF dominance in equal numbers in layer 4, but strong OFF dominance in layer 2-3. Also, they found the imbalance to be stimulus dependent - the imbalance changed when measured with sinusoidal gratings. Thus it remains unclear if this OFF dominance observed with sparse noise or gratings holds true when measured with natural images. We do not have laminar analysis information to check if the imbalance we observe when viewing natural scenes is layer dependent.

The subcortical OFF pathway has been demonstrated to have faster temporal dynamics than the ON pathway in the cat (Jin et al., 2011). This temporal difference is thought to arise from different kinds of glutamate receptors in ON and OFF types of retinal bipolar cells (Nakajima et al., 1993). Komban et al. (2014) compared temporal dynamics between populations of ON-dominated and OFF-dominated cortical neurons, and found faster dynamics for OFF-dominated cortical neurons. However, they performed their ON/OFF dominance classification based on the sign of a single maximum-amplitude pixel in the receptive field. A neuron with a stronger ON region is not necessarily driven by stronger ON pathway inputs. The ON subregion of a receptive field receives both excitation from the ON and inhibition from the OFF pathway, consistent with a push-pull arrangement. Therefore it remains unclear if this difference in temporal dynamics is present after the integration of ON and OFF pathway signals within individual simple cells. The ON and OFF maps estimated in this study allowed us to disentangle the excitatory and inhibitory contributions of ON and OFF pathways within individual receptive fields. We did not see any bias towards faster integration of OFF signals compared to ON. There might be compensating mechanisms in the cortex to remove temporal latency differences emerging from the retina. One possible mechanism could be the distinct localization of ON and OFF afferent inputs on the dendrites (Lee et al., 2016). For example, localizing faster inputs on distal parts of the dendrites could delay their propagation to the soma.

We noticed a significant asymmetry in the undershoots of ON and OFF temporal dynamics. This imbalance in push-pull along the time dimension challenges models based on linearity of temporal summation (Priebe and Ferster, 2005), and might have implications for motion selectivity of cortical receptive fields (Shariati and Freeman, 2012). In future, it would be

worthwhile to check if the ON and OFF input dominant neurons observed here are segregated in separate domains in the cortex as shown previously for geniculate afferents (Jin et al., 2008).

4.5.3 Implications for second-order processing

In Chapter 2 (Gharat and Baker, 2017), I suggested that imbalance between "push" and "pull" in neurons that pool ON and OFF Y-pathway inputs could be the neural mechanism for building cue-invariant selectivity for edge detection. A previous study (Hutchinson et al., 2016) reported neurons in cat A18 with phase-dependent selectivity for compound stimuli containing luminance and contrast-modulated gratings, with maximum response for the phase-aligned condition. The unbalanced push-pull model could account for this specific phase selectivity if it contained stronger excitation as well as inhibition from the ON compared to the OFF pathway (Gharat and Baker, 2017).

In this study when we evaluated the push-pull within simple cell receptive fields we saw an imbalance of varying degree, but there was no systematic preference for the ON pathway inputs. Our sample contained neurons with both ON- as well as OFF-dominance, with a slight bias for OFF-dominance. Area 18 in particular, which receives a majority of its inputs from the Y pathway, contained neurons dominated by both ON as well as OFF inputs. So our results from this study would seem to predict that around half of the cue-invariant neurons (ON-dominant) should prefer in-phase combination of luminance and contrast modulated gratings and other half (OFF-dominant) should prefer anti-phase combination. However it seems very unlikely that Hutchinson et al. (2016) selectively sampled only ON-dominant neurons and missed the OFF-dominant neurons.

One possibility that might bear on this discrepancy is the asymmetry between ON and OFF Y cells in their rectification at the bipolar-ganglion cell synapse (Borghouis et al., 2013). In the Y cell model of Demb et al. (2001b), the amount of rectification controls the strength of a Y cell's response to CM gratings compared to LM gratings. In the model simulations from Chapter 2, I made a simplifying assumption that ON and OFF pathways contain equal amounts of rectification, and modeled the nonlinearity as a simple half-wave rectification (ReLU) for both. However it is possible that the CM responsiveness and CM-LM phase dependence of a cortical neuron might depend upon both imbalance in push-pull at the cortical level as well as asymmetric rectification in the retina. Results from Chapter 3 showed that estimated models of LGN Y cells contain varying degrees of rectification (alpha values). However, the studies presented in this thesis do not address whether there is any systematic relation between the push-pull arrangement in a simple cell and rectification in the ON and OFF Y pathway inputs to that cell. Future studies could address this question by training a convolutional model from cortical simple cell responses that can learn both retinal level rectification as well as cortical level push-pull. The model architecture described in Chapter 2 (Figure 2.7) could be used by replacing the ReLU nonlinearity in the Y cell model with a pReLU nonlinearity, then the fitted alpha values of the ON and OFF inputs could capture some of the asymmetry. However, given the complexity of this model architecture and its larger number of parameters, training such a model would require much larger datasets than used in Chapter 3.

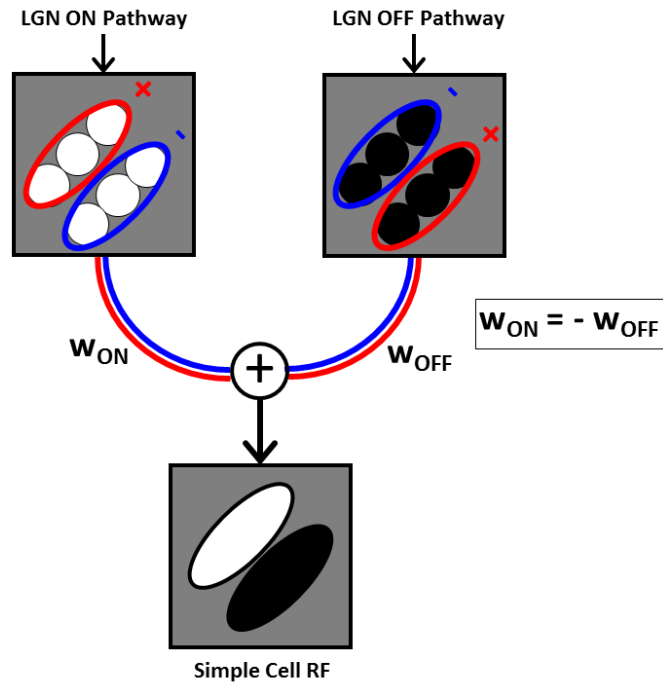


Figure 4.1 : Push-pull arrangement of ON and OFF pathway inputs within a simple cell receptive field.

Schematic model of how an obliquely oriented receptive field of simple cell is thought to be constructed from isotropic ON- and OFF-pathway inputs, with summation weights w_{ON} and w_{OFF} , respectively. At each spatial location an excitatory input (“push”) from one pathway is complemented by an inhibitory input (“pull”) of equal magnitude from the other pathway. In this manner, nonlinear (rectified) inputs can be combined to build a linear receptive field.

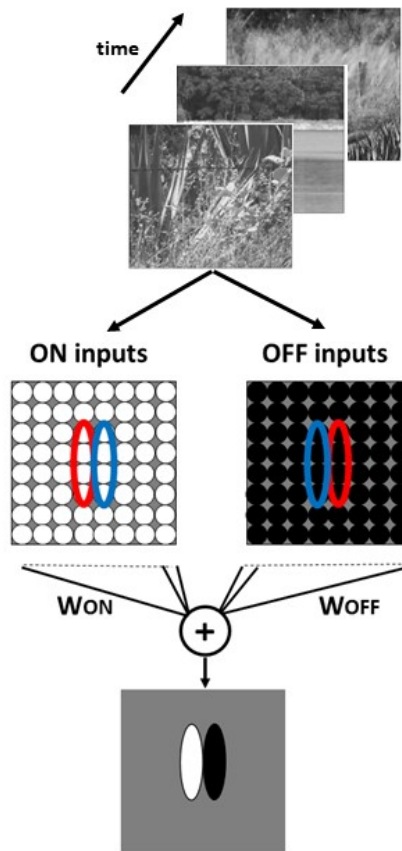


Figure 4.2 : Subunit model architecture for simple type cells.

Cortical neuron receptive field is modeled as a weighted linear sum of rectified outputs of simulated ON- and OFF-pathway inputs. Weights of these inputs are estimated from spiking responses of individual neurons to sequences of natural images. The sets of weights (W_{ON} and W_{OFF}) are estimated using regularized gradient descent optimization.

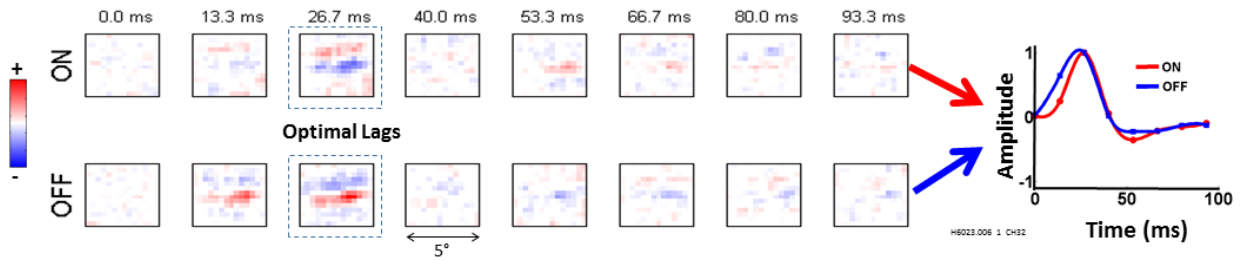


Figure 4.3 : Estimated ON-OFF subunit receptive field model for an example simple cell.

Space-space-time weight maps of ON- and OFF-pathway inputs to a simple cell (VAF for prediction of Test dataset, 44.6 %). Red and blue colors in the map denote positive weights (excitation) and negative weights (inhibition), respectively. At optimal time lag (3rd column) both weight maps have similar horizontal orientation but opposite polarity. Singular value decomposition was used to extract temporal dynamics of the ON and OFF weight maps. Plot on the right side shows temporal dynamics of the ON pathway in red and OFF pathway in blue.

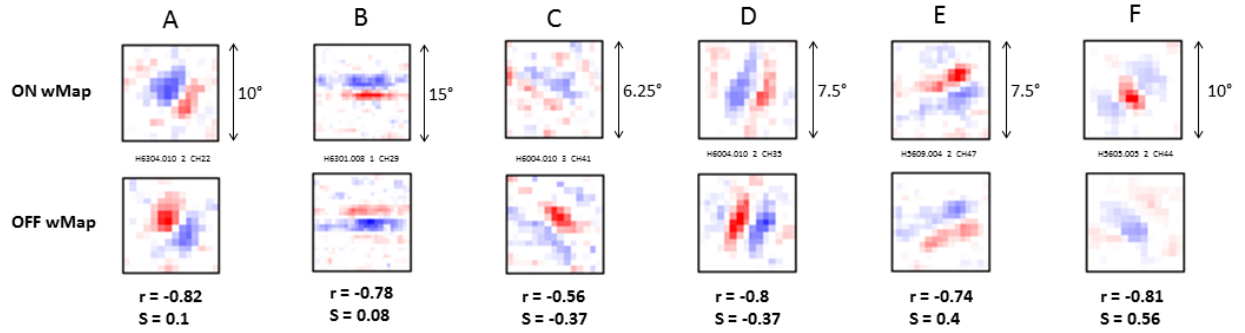


Figure 4.4 : Spatial ON-OFF weight maps for six example simple cells.

Estimated ON and OFF weight maps, shown at their optimal temporal lags. Comparison between ON and OFF maps is summarized with two indices – correlation (r) and ON-OFF strength (S). Neurons have negative correlation between ON and OFF wMaps with r close to -1, suggesting similar shapes but opposite polarity. Relative strengths of weight magnitudes for ON and OFF maps varied across neurons. Neurons in panel (A) and (B) have equal strength of ON and OFF inputs, with S close to 0. However, neurons in panel (C) and (D) have stronger OFF inputs compared to ON inputs, with S less than 0, while neurons in panel (E) and (F) have stronger ON than OFF inputs, with S greater than 0. VAFs for prediction of Test datasets, 12.3 % for (A), 14 % for (B), 32.5 % for (C), 26.7 % for (D), 21.9 % for (E), 39.84 % for (F).

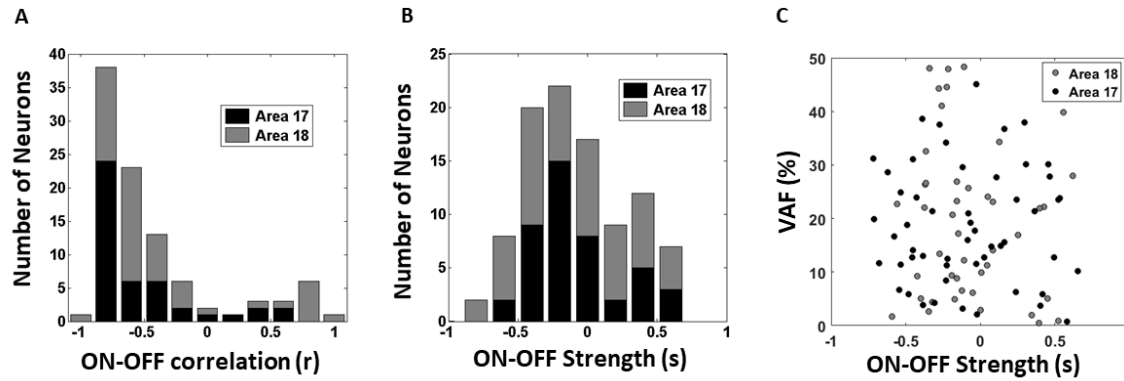


Figure 4.5 : Spatial arrangements of ON and OFF inputs within simple cell receptive fields of Area 17 and 18.

(A) Histogram showing distribution of correlation coefficients between spatial maps of ON and OFF inputs at optimal temporal lags. A majority of the sample neurons have a negative correlation, with r -values approaching -1. This suggests that wMaps for ON and OFF inputs have similar shapes but opposite polarity. However note that a small fraction of neurons have a positive correlation between wMaps for ON and OFF inputs. **(B)** Histogram showing distribution of ON-OFF strength (S), which quantifies the magnitude of ON versus OFF inputs. Neurons in our sample show a broad distribution for ON-OFF strengths (mean = -0.08, SD = 0.35, $N=97$). The mean of the ON-OFF distribution is significantly less than zero (one tailed t -test: $p=0.016$) suggesting an OFF dominance across the population. However, there was no significant difference between mean S indices for Area 17 ($N=44$) vs. 18 ($N=53$) neurons (two tailed t -test: $p=0.8$). **(C)** To see whether the observed imbalance between ON and OFF pathways within receptive fields was due to noisy estimates of the model, we compared ON-OFF strength

with predictive performance (VAF %) of the estimated model on test datasets. Scatterplot shows that there is no clear relationship between VAF and ON-OFF strength of neurons ($r = 0.056$, $p = 0.6$).

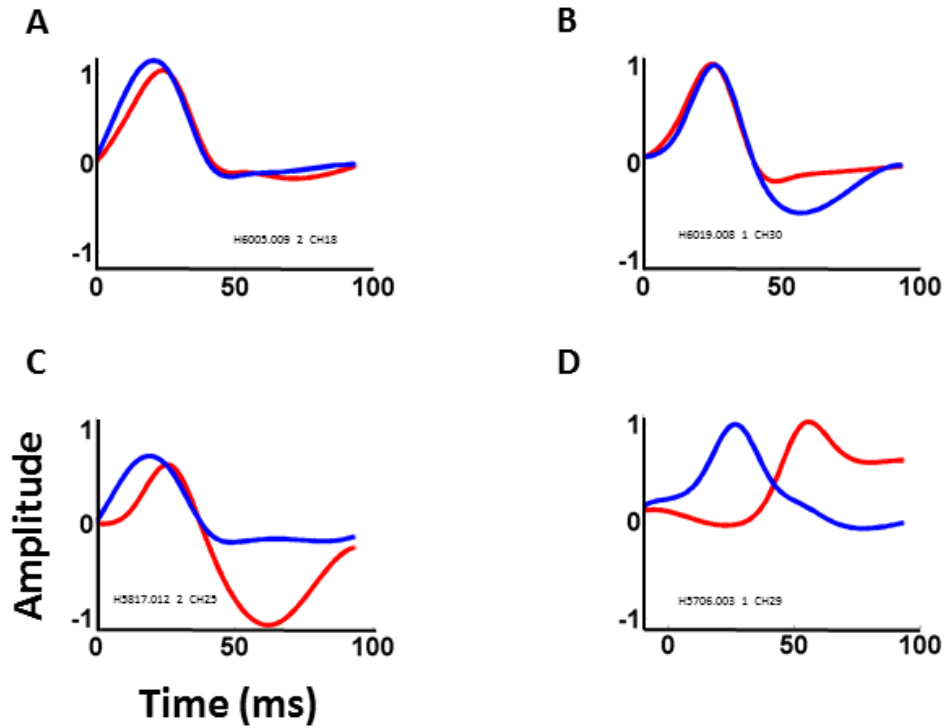


Figure 4.6 : Temporal dynamics of ON and OFF inputs within receptive fields of four example simple cells.

Neuron in panel A has very similar temporal dynamics for both ON (red) and OFF (blue) inputs. Note neuron in panel B has similar dynamics for positive half but a much greater undershoot for OFF inputs compared to ON. Neuron in panel C has faster dynamics for OFF inputs, and also the impulse function is monophasic for OFF inputs but biphasic for ON inputs. A minority of neurons in our sample show large differences in peak times of the ON and OFF impulse functions, as shown in panel (D). VAFs for prediction of Test datasets, 47.88 % for (A), 23.06 % for (B), 20.7 % for (C), 9.2 % for (D).

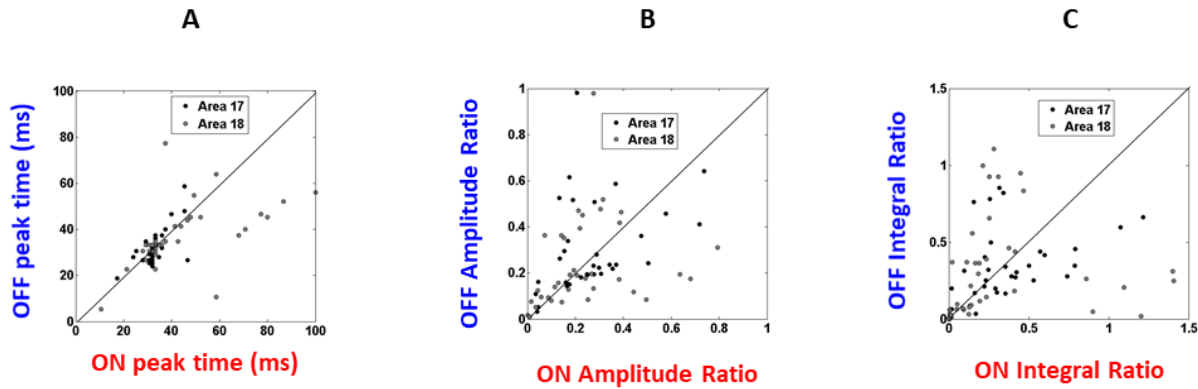


Figure 4.7 : Temporal dynamics of ON and OFF inputs within receptive fields of simple cells in Area 17 and 18.

(A) Scatterplot comparing peak times for ON vs. OFF inputs. Most of the points lie close to the 1:1 line, suggesting that both pathways have similar peak times within individual neurons.

Across the sample, these peak times are not significantly different (paired t-test: $p=0.35$). (B)

Scatterplot comparing Amplitude Ratios for ON vs. OFF inputs. Most of the points lie away from the 1:1 line, suggesting that the two pathways have different Amplitude Ratios. However across the sample, the mean values of Amplitude Ratios for ON and OFF were not significantly different (paired t-test: $p=0.76$). (C) Scatterplot comparing Integral Ratios for ON vs. OFF

inputs. Most of the points lie away from the 1:1 line suggesting that the two pathways have different Integral Ratios. However across the sample the means of the Integral Ratios for ON and OFF were not significantly different (paired t-test: $p=0.36$).

5

In this final chapter, I summarize the findings of three data chapters collectively. I discuss limitations and caveats of the overall approach, particularly in relation to the techniques used in these studies. I also present and discuss potential improvements or extensions of these studies that might be considered in future. Finally, I discuss implications of the results from this thesis on the models of second-order processing.

CHAPTER 5

General Discussion

5.1 Summary of findings

The aim of the thesis was to further understand how early stages of the visual system encode low level visual features like luminance and texture boundaries at a single neuronal level. Here the focus was on spatial nonlinearities emerging from the retina and their influence on processing different kinds of boundaries in the cortex. I employed an interdisciplinary approach using multielectrode recordings in different visual areas along the visual pathway, synthetic and naturalistic visual stimuli, model simulations of multi-layer convolutional neural networks and modern system identification methods including deep learning to study hierarchical processing at the early stages of the visual pathway.

In Chapter 2, to investigate how nonlinear Y pathway inputs could be pooled in cat Area 18 to build cue-invariant receptive fields, I performed multi-electrode recordings while stimulating receptive fields with grating stimuli. Using a relatively unbiased recording method, I found a large fraction of neurons in the early visual cortex with receptive fields not selective for orientation, that have spatial nonlinearities like those of subcortical Y cells. Such neurons responded to contrast-modulated (second-order) gratings with selectivity for their texture elements. This population of neurons form a strong candidate pool for building cue-invariant orientation selective neurons reported previously in the cortex. Finally, I proposed a novel neural circuit mechanism and showed through model simulations that pooling ON- and OFF-center Y-like cortical neurons in an unbalanced “push-pull” manner generates orientation-selective cue-

invariant receptive fields that can account for previously observed selectivity to luminance- and contrast-modulation gratings. According to this model, luminance (first-order) and texture (second-order) boundaries are processed jointly through a single pathway. This model is fundamentally different from the previously proposed model (Mareschal & Baker, 1998) that assumed processing of luminance and texture stimuli in two separate, parallel pathways with later summation of these signals in cortical neurons.

Results from Chapter 2 further supported the idea that spatial nonlinearities in the receptive fields of subcortical Y pathway neurons are important for texture processing in the cortex. So in Chapter 3, I estimated biologically interpretable quantitative receptive field models of LGN neurons that can account for their spatial nonlinearities. I trained a two-layer convolutional neural network model, based on previously known retinal circuitry, with extracellularly recorded spiking responses of individual neurons to naturalistic texture movies. Despite initializing convolutional filters' weights randomly without any shape constraints, the learned models converged to filters with clear Gaussian or DoG (Difference of Gaussians)-like shapes. The trained models of Y-type LGN neurons had a higher degree of rectifying nonlinearity compared to those of X-type neurons. Finally, this nonlinear two-layer convolutional model was better at predicting responses of Y-type neurons to novel test stimuli compared to a standard linear-nonlinear model without an intermediate nonlinearity. These results suggest that simplified convolutional networks with biologically interpretable components can be a powerful technique for modeling early stages of the visual pathway using limited and noisy neural data.

In Chapter 4, I separated the contributions of subcortical ON and OFF pathway inputs to receptive fields of cat Area 17 and 18 simple cells, when stimulated with natural scenes. I

modeled each simple cell receptive field as a weighted linear sum of simulated ON and OFF inputs, tiled across visual space. Regularized gradient descent regression was used to estimate separate spatiotemporal maps of each simple cell's summation weights for these ON and OFF inputs, from its responses to natural images. This method provided a way to measure both excitatory as well as inhibitory contributions of the ON and OFF pathways within individual receptive fields. I demonstrated that for most neurons, maps of ON and OFF weights are approximately identical in spatial layout and magnitude, but are of opposite polarity (“push-pull”). However, there was a large diversity in the ratio of strengths of ON vs. OFF inputs, and a large subset of neurons had unbalanced push-pull. In addition, I also demonstrated asymmetries between temporal dynamics of ON and OFF weight maps. These results show spatiotemporal asymmetries in how visual cortex neurons integrate signals from light and dark regions within natural scenes, challenging the model of simple cells as linear filters. This asymmetry in simple cell inputs could potentially contribute towards building cue-invariant receptive fields from Y-pathway inputs as suggested in Chapter 2.

5.2 Limitations and caveats

5.2.1 Anesthesia and brain state

All the experiments in this thesis were carried out on anesthetized and paralyzed cats. Anesthesia is known to have significant effects on visual processing both at the single neuron level (Lamme et al., 1998; Pack et al., 2001) as well as network levels (Sellers et al., 2015). Anesthesia and other sources of noise in the brain introduce trial-to-trial variability, such that repeated presentations of the same visual stimuli generate different neuronal responses (Faisal et

al., 2008; Harris et al., 2011). This neuronal variability increases in strength along the visual pathway (Goris et al., 2014). For example, the primary visual cortex of anesthetized cats has been shown to alternate between synchronous and asynchronous states while viewing natural scene movies, thus affecting spiking reliability of neurons (Spacek and Swindale, 2015). This neuronal variability due to brain state changes poses a significant challenge for receptive field estimation, since the neurons' visual responses can be obscured by the noise, and overfitting to the noise in neural data becomes a greater challenge for the training algorithm.

In order to minimize the effects of neuronal noise and slow state changes on receptive field model estimation, we repeated the stimuli multiple times, and randomized the presentation order of the stimulus movies. Repeating the stimulus and averaging the neuronal response for these repetitions helped to greatly reduce the noisiness in the neural data. This in turn helped to reduce over-fitting of the model to the noise in the data. Randomly intermixing the training and test data sets minimized any systematic difference in state while measuring training and test data, and helped to average out the effects of slow gain changes over the time of a recording session lasting around 45 minutes.

Another approach to accommodate cortical state-dependent effects on neuronal firing rate is to include signals from different LFP bands and multi-unit activity (MUA) recorded at other sites on the multi-electrode into the model architecture (Cui et al., 2016). Such a model is better able to predict neurons' responses across different trials. One concern, however, with such a model architecture, is that the recorded LFP and MUA signals are not independent of the stimulus, and thus can lead to the problem of collinearity (Wold et al., 1984). In that case, the estimated parameters of the model can be unstable.

It is conceivable that anesthesia might have differential effects on neuronal responses to first-order compared to second-order stimuli. According to the Y cell model from Demb et al. (2001a), rectification at the bipolar-ganglion cell synapse in the retina controls the strength of nonlinear second-order response compared to first-order response in Y cells. This rectification depends upon the spontaneous glutamate release rate from the bipolar cell's synaptic terminals. A lower release rate would result in a higher degree of rectification and thus stronger second-order response. It is possible that glutamate release rate depends upon type and depth of anesthesia used during the experiments.

To date all the studies looking at nonlinear responses in Y cells (Rosenberg et al., 2010) and cortical neurons (Zhou and Baker, 1993; Tanaka and Ohzawa, 2006; Rosenberg and Issa, 2011) have been performed under anesthesia, or in ex vivo retina (Demb et al., 2001b, Borghuis et al., 2013). But it remains unclear whether the nature of nonlinearity and second-order responsiveness of neurons is similar in awake behaving animals. One of the challenges in studying second-order processing in awake animals would be the necessity to monitor eye movements with high precision, since the second-order stimuli are at a very high spatial scale. This problem could be solved by synchronizing stimulus position on the screen with eye movements using high precision eye tracking. Such a retinal stabilization approach has been applied in human psychophysics to study the role of minute eye movements such as microsaccades (Poletti M et al. 2013; Rucci et al 2007). In the future, such methodology could be used to study the spatial nonlinearity of LGN and cortical neurons in awake and behaving monkeys.

5.2.2 Laminar organization

The cortex has a laminar structure with six layers, having specific arrangements of different cell types forming microcircuits. Furthermore, there are specific patterns of connections between neurons within a cortical area according to their locations in the laminae. Projections to, as well as inputs from, other cortical areas are also lamina-dependent. For example, in cat Area 17, the majority of the input from the LGN terminates in layer 4, while feedforward projection to higher areas originates from layers 2-3 and feedback to the LGN is from layer 6. Signal processing across these laminae and the functional roles of each layer mostly remains unknown.

Previous intracellular studies have tried to relate receptive field properties of neurons to their laminar location, by dye-filling and histological reconstructions (e.g. Martinez et al., 2005). However, making a connection between receptive field properties of extracellularly recorded neurons and their laminar location has been challenging. Recent studies have begun to take advantage of multi-electrodes that can span all the cortical layers, and to use current source density (CSD) analysis to localize the positions of electrode sites across laminae (e.g. Jin et al., 2011; Hansen et al., 2012). Even though most data in Chapter 2 and part of that in Chapter 4 was recorded with NeuroNexus linear arrays that can span the entire cortical grey matter depth, we were not able to obtain good enough CSD results to localize laminae. The small pad size ($177\ \mu\text{m}^2$) of the multi-electrodes used in these studies is good for isolating single units but not ideal for recording low frequency LFPs important for estimating CSD profiles. Preliminary attempts in our lab using larger pad size electrodes have provided improvements in the estimated CSD profiles. So it would be interesting to see in future if there is any systematic laminar dependency for the kind of ON-OFF input imbalance observed in Chapter 4, for example like that observed previously for black dominance in monkey V1 (Yeh et al., 2009).

5.2.3 Sampling bias

Using multi-electrodes, spike sorting and random naturalistic stimuli may help to greatly reduce sampling biases in the recorded neurons compared to single-channel electrode recordings with a search stimuli like bars (Talebi and Baker, 2016). Nevertheless, it is still difficult to know if we sampled from all types of neurons present in the LGN and cortex. It is thought that only around 10 % of neurons in the vicinity of a multi-electrode recording site are usually sampled (Shoham et al., 2006). For example, very small spike amplitudes originating from anatomically small neurons would be difficult to isolate with spike sorting. In addition, there is a possibility that these multi-electrodes with wider shanks might damage neurons' cell bodies or their processes (Blanche et al., 2005), and that some cell types might be more vulnerable to such damage than others. However, it is unlikely that the higher proportion of non-ori cells observed in Chapter 2 is a result of electrode damage to the neurons' dendrites. If that were the case, we would have seen a variety of receptive field maps with missing parts of receptive field subregions or with holes within receptive fields. In addition, a previous study from our lab (Talebi & Baker, 2016) as well as results from Chapter 4 did not reveal such damaged receptive fields.

2-photon calcium imaging provides a complementary way of measuring neural activity at single neuron resolution. In this method, activity from neurons in a small volume of neural tissue can be measured by imaging fluorescence signals from activity-related Calcium concentration. Since this method allows direct visualization of the neuronal cell bodies, one can ideally measure activity from all the neurons in a given block of neural tissue (Stosiek et al., 2003). However, 2-photon imaging lacks the temporal resolution of extracellular recording. Also, with this method

neural activity from only superficial cortical layers can be recorded. Currently, efforts are underway to measure activity from deeper layers (Na, 2017), multiple brain areas (Stirman et al., 2016), and deep brain structures (Bocarsly et al., 2015).

Ohki et al., 2006 with 2-photon imaging in cat Area 18 reported presence of only around 1 % neurons non-selective for orientation. Their results might look contradictory to our results in chapter 2, where close to one-third of the neurons were classified as non-Ori neurons. However, this discrepancy seems to arise from differences in the criteria used by Ohki et al. compared to ours for classifying the neurons. In our study, we used a vector based approach (Leventhal et al., 2003) to calculate Orientation Bias (OB) index and classified neurons with selectivity similar to cat LGN neurons ($OB < 0.2$) (Rosenberg et al., 2010) as non-Ori. But, Ohki et al. used ANOVA on neuronal responses at different orientation of gratings to classify neurons. So it is conceivable that most of our non-Ori would be classified as Ori-selective according to their criteria.

5.2.4 Visual stimuli

In this thesis, we have employed synthetic gratings in Chapter 2, synthetic textures with naturalistic properties in Chapter 3 and grayscale natural images in Chapter 4 as visual stimuli to study receptive field properties. However, ideally, we should be studying visual processing under completely natural stimulation, since receptive field properties of neurons can be stimulus-dependent (David et al., 2005). Even though I used natural images for estimating receptive field models in Chapter 4, the overall spatiotemporal stimuli were not truly natural - different randomly selected natural images were presented at 75Hz, with no correlation between images over time. But during natural stimulation, images formed on the retina are temporally correlated, with the temporal dynamics depending on head and body movements, eye movements, and

object motion. These dynamics add richness to the visual stimulation by providing important relative and global motion cues that can help in object segmentation, depth perception, guiding self-motion etc. However, a major challenge when using such data for system identification would be that the stimuli would be similar for a large number of frames, so the algorithm might not have enough novel information to fit the model parameters. Hence to utilize such temporally correlated stimuli, we would need much longer recording durations.

The visual stimuli used in this thesis were also unnatural in various other ways. Given the difficulty of eye alignment under anesthesia and paralysis, receptive fields were studied with monocular viewing, even though most of the cortical neurons could respond binocularly. Thus our model does not account for summation of inputs from the two eyes and the role of binocular disparity. In addition, the natural images used here were processed to remove color information, and mean luminance and r.m.s. contrast of the images was kept constant.

In future studies, efforts should be made to more closely mimic natural stimulation. This approach could lead to discovery of new kinds of feature selectivity within receptive fields.

5.2.5 Model architecture and training

In Chapters 3 and 4, I modeled receptive fields of LGN neurons and cortical simple cells as a two-stage feedforward subunit model with an intermediate nonlinearity. However, this model is an over-simplification of known neural circuits and cannot account for some of the known receptive field properties. For example, the LGN receives a majority of its input as a feedback connection from visual cortex, which was not included in our model architecture..

Similarly, in Chapter 4 the model architecture used for simple cell receptive fields did not include any feedback, even though cortical neurons receive feedback from higher visual areas.

In these models an important assumption is that rectified outputs of subunits are linearly summed by the receptive field. However, in the retina, amacrine cells not only provide inhibitory input to the ganglion cells but also control the gain of bipolar cells, thus giving a multiplicative interaction between excitatory and inhibitory inputs within the retinal circuit (Demb, 2008). Finally, nonlinearities between the subunit and pooling layers were modeled here as static nonlinearities (pReLU or ReLU). However, the nature of this nonlinearity has been demonstrated to be much more complicated within the retinal circuit of ON-type Y cells (Borghuis et al., 2013). Such a kind of dynamic nonlinearity is not captured with the currently used model framework. Nevertheless, despite using simplified version of the model we could account for a large fraction (up to 80 %, as shown in Figure 3.10) of neural response variance. Also, it remains a challenge to incorporate such complex models in a framework whose parameters can be readily estimated.

In Chapters 3 and 4, estimation of the receptive field models was framed as a supervised learning problem, with visual stimuli as input and spiking neural responses as output. The training algorithms acted to minimize an error function by updating the model parameters in an iterative manner using gradient descent back-propagation. Even though this method of model training has been very successful in the field of machine learning, it is not biologically plausible. Recently efforts have begun to develop algorithms that update synaptic weights in a manner inspired from spike timing-dependent plasticity (Bengio et al., 2016). It is possible that such training algorithms might improve model performance as well as require smaller amounts of neural data.

5.3 Future directions

5.3.1 Multi-stage convolutional model

In Chapter 3, I used a two stage convolutional model consisting of a convolutional filter layer and a pooling layer separated by a static nonlinearity to model LGN receptive fields. The advantages of this simplified model are that the parameters are easily interpretable, and the model requires relatively modest amounts of data for training. However, to model receptive fields of cortical neurons such that they account for both first- and second-order responses, we would need to use a much more complicated model architecture with at least 3 layers. But it is an open question whether such models trained on neural data from visual cortex neurons would have any correspondence with the biological neural circuitry. For example, the visual system has separate ON and OFF subcortical pathways that process visual information separately and only at the cortical stage these signals are mixed - but it is unclear whether an artificial neural network would be able to learn to have separate ON and OFF pathways from cortical neurons' data. Also, the deeper the neural network, the more difficult it becomes to interpret computations across the layers. It is unclear how to visualize receptive fields for neurons other than the first layer, due to the intermediate nonlinearities. One approach could be to have known biological constraints on the model architecture itself before training it. For example, front-end filters could be constrained to be circularly symmetric like retinal receptive fields. A model could be designed to have two separate pathways that are summed only at the final stage. Such a model would be easy to interpret and might be able to infer signal processing at the level of the retina (including nonlinear subunits) as well as the cortex (e.g. push-pull).

If the approach of training deeper neural networks is taken, it will require much larger datasets than used in this thesis. While recording with multi-electrodes over such long recording sessions, spike amplitudes of neurons often disappear from one channel and appear on other nearby channels, due to electrode slippage. Tracking movement of neuronal spike signals across multi-electrode channels could allow us to record visual responses for longer durations. In addition to tracking spike waveform shapes, visual responses to a repeated presentation of the same movie at different time intervals could be used to ascertain that we are measuring responses from same neuron over the recording session.

A very different approach could be to estimate a common neural network model for multiple neurons that are simultaneously recorded with multi-electrodes (McIntosh et al., 2016), as they probably share a majority of their inputs from lower stages of the visual pathway. This framework will have multiple outputs corresponding to each neuron recorded in the session, unlike the models in this thesis with only one output. It would be important to have the electrode penetrations perpendicular to the cortical surface so that the recorded neurons have large receptive field overlap, and therefore a higher proportion of shared inputs. This approach might require shorter recording times, and enable training deeper neural networks. One implicit assumption with this framework is that all the recorded neurons are at same stage of the visual hierarchy. This might not be true if the sample contains neurons across multiple laminae. Nevertheless, this might be a fruitful framework to explore, that could give novel insights into network level computations from multi-electrode recordings.

5.3.2 Two-photon imaging and neural network models

Previous studies as well as those in this thesis have modeled receptive fields of neurons in visual pathway as linear weighted sums of model subunits from the previous stage in the visual hierarchy. For example, receptive fields of MT have been modeled as a linear sum of direction selective V1-like receptive fields (Nishimoto and Gallant, 2011). Similarly, receptive fields of area MST neurons were modeled as a linear sum of model MT units (Mineault et al., 2012). However, integration of synaptic inputs by cortical neurons remains poorly understood. This question can be addressed in future by combining advancements in the methodologies of two-photon imaging and deep learning. It is now possible to simultaneously record visual responses across multiple visual areas using two-photon imaging (Stirman et al., 2016; Smith et al., 2017). Datasets collected with such a setup could be used to estimate transfer functions between two visual areas. In the neural network framework, responses of the neurons in the lower area could be used as the inputs and responses from the higher area neurons as the outputs of the neural network model.

Subunit integration can also be studied at the level of a single cortical neuron and its synaptic inputs, by performing two-photon imaging with sparse labeling (Wilson et al., 2016). With this method, it is possible to simultaneously measure activity of a neuron and its dendritic spines. Data from this experimental setup could also be used to train a neural network model that could predict responses of the neuron given its synaptic inputs. However, it should be noted that activity of only a small fraction of the synaptic boutons can be measured. Nevertheless, such neural network models could elucidate rules of connectivity between neurons in the visual hierarchy and provide a better understanding of how subunits are integrated within cortical receptive fields.

5.3.3 Second-order processing

In this thesis, I have proposed a novel neural computation that takes advantage of asymmetry between subcortical ON and OFF Y pathways for encoding texture stimuli simultaneously with luminance. This idea is contrary to the previous idea that texture/second-order stimuli are processed separately from luminance, in two independent parallel pathways (Baker, 1999). Several human psychophysical studies have supported the idea of separate processing of first- and second-order stimuli (e.g. Smith and Ledgeway, 1998; Schofield and Georgeson, 1999; Nishida et al., 1997). Furthermore, studies on patients with brain injuries to extrastriate cortex have demonstrated that processing of first- or second-order stimuli can be selectively impaired depending upon the damaged area (Vaina and Cowey, 1996; Vaina et al., 1998; Vaina et al., 1999). So how might we reconcile these seemingly conflicting findings from neurophysiology and psychophysics?

One possibility is that at the early stages of the visual pathway, first- and second-order signals are multiplexed, and only later in the extrastriate cortex might these signals be separately processed in distinct cortical areas. A similar idea has also been suggested for processing of color and luminance signals in the primate visual pathway (Bushnell et al., 2011; Gheiratmand M et al., 2013). However, there has been no neurophysiological evidence for the existence of "second-order only" responsive neurons in early cortical areas. Future studies should look for such second-order neurons in higher visual areas that might be only selective for purely texture stimuli. However, it is challenging to search for such neurons with conventional search stimuli such as simple bars or spots as these neurons would be unresponsive to such stimuli. System identification methods like the ones used in this thesis, with random naturalistic stimuli, could be a viable approach since it makes minimal prior assumptions about the receptive field properties.

At the population level, using intrinsic optical imaging it has been shown that second-order response strength is greater in higher visual area V4 than in V1 or V2 (Pan et al., 2012). Thus it would be worthwhile to test for the presence of purely second-order neurons in such brain areas.

REFERENCES

Albright T (1992) Form-cue invariant motion processing in primate visual cortex. *Science*. 255:1141-1143.

Alonso et al., (1996) Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*. 383: 815-819.

Alonso J, Usrey WM, Reid RC (2001) Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *The Journal of Neuroscience*. 21(11):4002-4015.

An X et al (2014) Orientation-cue invariant population responses to contrast-modulated and phase-reversed contour stimuli in macaque V1 and V2. *Plos One*. 9:e106753.

Anishchenko A et al., (2010) Receptive field mosaics of retinal ganglion cells are established without visual experience. *Journal of Neurophysiology*. 103:1856-1864.

Baker CL (1999) Central neural mechanisms for detecting second-order motion. *Curr Opinion in Neurobiology*. 9(4): 461-466.

Baudot et al., (2013) Animation of natural scene by virtual eye-movements evokes high precision and low noise in V1 neurons. *Frontiers in Neural Circuits*. 7(206).

Bengio Y et al., (2016) Towards biologically plausible deep learning. arXiv:1502.04156v3.

Berens P (2009) CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software*. 31(10): 1:21.

Bergstra, J., et al. (2010). Theano: A CPU and GPU math compiler in Python. Paper presented at the Proc. 9th Python in Science Conf.

Blanche TJ et al., (2005) Polytrodes: High-density silicon electrode arrays for large-scale multiunit recording. *J Neurophysiol* 93: 2987-3000.

Bocarsly ME et al., (2015) Minimally invasive microendoscopy system for in vivo functional imaging of deep nuclei in the mouse brain. *Biomedical Optics Express* 6(11): 4546-4556.

Bolinger D, Gollisch T (2012) Closed-loop measurements of Iso-response stimuli reveal dynamic nonlinear stimulus integration in the retina. *Neuron* 73:333-346.

Borghuis BG, Marvin JS, Looger LL, Demb JB (2013) Two-photon imaging of nonlinear glutamate release dynamics at bipolar cell synapses in the mouse retina. *J Neurosci* 33(27): 10972-85.

Bonin V, Histed MH, Yurgenson S, Reid RC (2011) Local diversity and fine-scale organization of receptive fields in mouse visual cortex. *The Journal of Neuroscience*. 31(50):18506-21.

Brainard DH (1997) The psychophysics toolbox. *Spatial Vision*. 10:433-436.

Bullier J and Norton T (1979) Comparison of receptive field properties of X and Y ganglion cells with X and Y lateral geniculate nucleus cells in the cat. *Journal of Neurophysiology*. 42(1):274-291.

Bushnell BN et al., (2011) Equiluminance cells in visual cortical area V4. *J Neuroscience*. 31(35): 12398-412.

Chapman B and Stryker MP (1993). Development of orientation selectivity in ferret visual cortex and effects of deprivation. *The Journal of Neuroscience*. 13(12):5251-62.

Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. *Network*, 12, 199-213.

Chichilnisky EJ and Kalmar RS (2002). Functional asymmetries in ON and OFF ganglion cells of primate retina. *The Journal of Neuroscience*. 22:2737-2747.

Chollet, F. (2015). Keras. GitHub repository: <https://github.com/fchollet/keras>.

Cleland B, Levick W, Wassle H (1975) Physiological identification of a morphological class of cat retinal ganglion cells. *Journal of Physiology*. 248:151-171.

Cleland et al., (1976) Lateral geniculate relay of slowly conducting retinal afferents to cat visual cortex. *Journal of Physiology* 255:299-320.

Crook JD, Peterson BB, Packer OS, Robinson FR, Troy JB, Dacey DM (2008a). Y-cell receptive field and collicular projection of parasol ganglion cells in macaque monkey retina. *J Neuroscience* 28:11277–11291.

Crook et al. (2008b). The smooth monostratified ganglion cell: Evidence for spatial diversity in the Y-cell pathway to the lateral geniculate nucleus and superior colliculus in the macaque monkey. *The Journal of Neuroscience* 28(48):12654-71.

Cui et al., (2016) Inferring cortical variability from local field potentials. *J Neurosci*. 36(14): 4121-4135.

Dacey D, Packer OS, Diller L, Brainard D, Peterson B, Lee B (2000). Center surround receptive field structure of cone bipolar cells in primate retina. *Vision Research* 40:1801-1811.

David S et al., (2004) Natural stimulus statistics alters the receptive field structure of V1 neurons. *J Neuroscience*. 24(31): 6691-7006.

DeAngelis GC, Freeman RD, Ohzawa I (1994) Length and width tuning of neurons in the cat's primary visual cortex. *J Neurophysiology* 71:347-374.

Demb J, Haarsma L, Freed MA, Sterling P (1999) Functional circuitry of the retinal ganglion cells nonlinear receptive field. *The Journal of Neuroscience*. 19(22):9756-9767.

Demb J, Zaghloul K, Haarsma L, Sterling P (2001a). Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (Y) ganglion cell in mammalian retina. *The Journal of Neuroscience*. 21(19): 7447-7454.

Demb J, Zaghloul K, Sterling P. (2001b) Cellular basis for the response to second order motion cues in Y retinal ganglion cells. *Neuron*. 32: 711-721.

Demb J (2008) Functional circuitry of visual adaptation in the retina. *J Physiology* 586(18):4377-84.

Dragoi V, Turcu CM, Sur M (2001) Stability of cortical responses and the statistics of natural scenes. *Neuron* 32:1181-1192.

Dreher B, Leventhal AG, Hale PT (1980) Geniculate input to cat visual cortex: a comparison of area 19 with areas 17 and 18. *Journal of Neurophysiology*.804-826.

Dreher B and Sefton J (1979) Properties of neurons in cat's dorsal lateral geniculate nucleus: a comparison between medial interlaminar and laminated parts of the nucleus. *J Comp Neurology* 183:47-64.

El-Shamayleh Y and Movshon JA (2011) Neuronal responses to texture-defined form in macaque visual area V2. *The Journal of Neuroscience* 31(23) 8543-8555.

Enroth-Cugell C & Robson JG (1966) The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology* 187:517-552.

Faisal AA, Selen LPJ, Wolpert DM (2008) Noise in the nervous system. *Nature Reviews* 9:292-303.

Felsen G et al. (2005) Cortical sensitivity to visual features in natural scenes. *PLoS Biol* 3(10): e342.

Ferster D. (1988) Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *The Journal of Neuroscience*. 8(4):1172-1180.

Ferster D and Jagadeesh B (1991) Nonlinearity of spatial summation in simple cells of areas 17 and 18 of cat visual cortex. *Journal of Neurophysiology*. 66(5): 1667-1679.

Field DJ (1987) Relations between the statistics of natural images and response properties of cortical cells. *J. Opt. Soc. Am.* 4(12): 2379-2394.

Freeman J et al. (2015) Mapping nonlinear receptive field structure in primate retina at single cone resolution. *eLife* 4:e05241.

Fukuda Y and Stone J (1974) Retinal distribution and central projections of Y-, X- and W-cells of the cat's retina. *Journal of Neurophysiology*: 37(4):749-772.

Gharat A and Baker CL (2012). Motion-defined contour processing in the early visual cortex. *Journal of Neurophysiology*. 108:1228-1243.

Gharat A and Baker CL (2017) Nonlinear Y-Like receptive fields in the early visual cortex: An intermediate stage for building cue-invariant receptive fields from subcortical Y Cells. *The Journal of Neuroscience*. 37(4):998-1013

Gheiratmand M et al., (2013) Blobs versus bars: Psychophysical evidence supports two types of orientation response in human color vision. *Journal of Vision*. 13,2.

Glorot X and Bengio Y. (2010) Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, PMLR 9:249-256.

Gollisch T (2013) Features and functions of nonlinear spatial integration by retinal ganglion cells. *Journal of Physiology* 107(5): 338-348.

Gorris R et al., (2014) Partitioning neuronal variability. *Nature Neuroscience*. 17(6):858-865.

Hansen et al., (2012) Correlated variability in laminar cortical circuits. *Neuron* 76: 590-602.

Harris KD, Thiele A (2011) Cortical state and attention. *Nature Reviews* 12:509-523.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision.

Hirsch JA, Alonso JM, Reid RC and Martinez LM. (1998) Synaptic integration in striate cortical simple cells. *The Journal of Neuroscience* 18(22):9517-9528.

Hirsch JA (2003) Synaptic physiology and receptive field structure in the early visual pathway of the cat. *Cerebral Cortex* 13:63-69.

Hirsch JA, Martinez LM, Pillai C, Alonso JM, Wang Q and Sommer FT (2003) Functionally distinct inhibitory neurons at the first stage of visual cortical processing. *Nature Neuroscience*. 6 (12):1300-1308.

Hochstein S, Shapley RM (1976) Linear and nonlinear spatial subunits in Y cat retinal ganglion cells. *J Physiology* 262:265–284.

Hoerl AE & Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. 12:55-67.

Hubel D and Wiesel T (1959) Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology* 148:574-591.

Humphrey A, Sur M, Uhlrich D J, Sherman S M. (1985) Termination patterns of individual X- and Y cell axons in the visual cortex of the cat: projections to area 18, to the 17/18 border region, and to both areas 17 and 18. *J. Comp. Neurol.*, 233: 190-212.

Hutchinson CV, Ledgeway T, Baker CL. (2016) Phase-dependent interactions in visual cortex to combinations of first- and second-order stimuli. *The Journal of Neuroscience*. 36(49): 12328-37.

Jin J et al., (2008) ON and OFF domains of geniculate afferents in cat primary visual cortex. *Nature Neuroscience*. 11(1): 88-94.

Jin et al., (2011) Population receptive fields of ON and OFF thalamic inputs to an orientation column in visual cortex. *Nature Neuroscience* 14(2): 232-238.

Jin J et al., (2011) Faster thalamocortical processing for dark than light visual targets. *The Journal of Neuroscience* 31(48): 17471-79.

Johnson A and Baker CL (2004) First- and second-order information in natural images: a filter-based approach to image statistics. *J. Opt. Soc. Am. A* 21(6):913-925.

Kaplan E and Shapley R. (1984) The origin of the S (slow) potential in the mammalian lateral geniculate nucleus. *Experimental Brain Research*. 55:111-116.

Kingdom, F.A.A., Hayes, A., and Field, D.J. (2001). Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Research* 41, 585–598.

Kingma DP and Ba J. (2014) Adam: A method for stochastic optimization.
arXiv:1412.6980[cs.LG]

Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C. (2007) What's new in Psychtoolbox-3. *Perception* 36(14),1.

Komban et al (2014) Neuronal and perceptual differences in the temporal processing of darks and lights. *Neuron*. 82:224-234.

Kording PK, Kayser C, Betsch BY, Konig P (2001) Non-contact eye-tracking on cats. *J Neurosci Methods* 110:103-111.

Lamme V et al. (1998) Figure-ground activity in primary visual cortex is suppressed by anesthesia. *PNAS* 95:3263-3268.

Larsson J, Landy MS, Heeger DJ (2006) Orientation-selective adaptation to first-and second-order patterns in human visual cortex. *J Neurophysiology* 95:862-881.

LeCun Y et al., (2015) Deep Learning. *Nature*. 521:436-444.

Lee KS, Wilson DE and Fitzpatrick D (2016) Synaptic organization of ON and OFF inputs within the dendritic field of individual layer 2/3 neurons in tree shrew primary visual cortex. *Soc Neurosci Abstract* 241.05/LL7.

LeVay S and Gilbert CD (1976) Laminar patterns of geniculocortical projection in the cat. *Brain Research* 113:1-19.

Leventhal AG, Wand Y, Pu M, Zhou Y, Ma Y (2003) GABA and its agonists improved visual cortical function in senescent monkeys. *Science* 300:812-815.

Li G, Yao Z, Wang Z, Yuan N, Talebi V, Tan J, Wang Y, Zhou Y, Baker CL Jr. (2014) Form-cue invariant second-order neuronal responses to contrast modulation in primate area V2. *The Journal of Neuroscience*. 34(36): 12081-12092.

Liang Z and Freed MA (2010) The ON pathway rectifies the OFF pathway of the mammalian retina. *The Journal of Neuroscience*. 30(16):5543-5533.

Livingstone MS and Hubel D (1984) Anatomy and physiology of a color system in the primate visual cortex. *The Journal of Neuroscience* 4(1): 309-356.

Mairal J & Yu B. (2012) Complexity analysis of the lasso regularization path. *arXiv:1205.0079*.

Mareschal I, Baker CL. (1998a) A cortical locus for the processing of contrast defined contours. *Nat Neurosci* 1: 150–154.

Mareschal, Baker CL. (1998b) Temporal and spatial response to second-order stimuli in cat A18. *Journal of Neurophysiology*. 80:2811-2823.

Mareschal I, Baker CL Jr. (1999) Cortical processing of second-order motion. *Vis Neuroscience* 16: 527–540.

Martinez LM, Wang Q, Reig RC, Pillai C, Alonso JM, Sommer FT and Hirsch JA. (2005) Receptive field structure varies with layer in the primary visual cortex. *Nature Neuroscience*. 8:372-379.

Martinez et al (2014) Statistical wiring of thalamic receptive fields optimizes spatial smapling of the retina image. *Neuron*. 81:943-956.

Mareschal I and Dakin SC (2000) Sensitivity of contrast modulation depends on carrier spatial frequency and orientation. *Vision Research* 40(3): 311-329.

Mastronarde DN (1992) Nonlagged relay cells and interneurons in the cat lateral geniculate nucleus: Receptive field properties and retinal inputs. *Visual Neuroscience*. 8(5): 407-441.

Mata ML and Ringach DL (2005) Spatial overlap of ON and OFF subregions and its relation to response modulation ratio in macaque primary visual cortex. *Journal of Neurophysiology*. 93:919-928.

Mazer JA et al. (2002) Spatial frequency and orientation tuning dynamics in area V1. *PNAS*. 99(3):1645-1650.

McIntosh L et al., (2016) Deep learning models of the retinal response to natural scenes.
Advances in Neural Information Processing Systems 29 (NIPS)

McLean J, Palmer LA (1989) Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. Vision Res 29:675-679.

Mineault PJ, Khawaja FA, Butts DA, Pack CC. (2012) Hierarchical processing of complex motion along the primate dorsal visual pathway. PNAS. 109(16):E972-80.

Movshon JA, Thompson ID and Tolhurst DJ. (1978) Spatial and temporal contrast sensitivity in area 17 and 18 of the cat's visual cortex. The Journal of Physiology. 283:101-120.

Murphy PC and Sillito AM (1996) Functional morphology of the feedback pathway from area 17 of the cat visual cortex to the lateral geniculate nucleus. J Neurosci 16(3): 1180-92.

Na J (2017) Adaptive optical fluorescence microscopy. Nature Methods 14: 374-380.

Nakajima Y, Iwakabe H, Akazawa C, Nawa H, Shigemoto R, Mizuno N, Nakanishi S (1993) Molecular characterization of a novel retinal metabotropic glutamate receptor mGluR6 with a high agonist selectivity for L-2-amino-4-phosphonobutyrate. J Biol Chem 268:11868 –11873.

Nguyen, P. (2016). Spatial Convolution Methods for Modeling Neurons in the Early Visual Cortex (McGill University Libraries).

Nishida S, Ledgeway T, Edwards M (1997) Dual multiple-scale processing for motion in the human visual system. *Vision Res.* 37:2685-2698.

Nishimoto S, Gallant JL (2011) A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. *The Journal of Neuroscience.* 31(41): 14551-64.

Ohki et al., (2006) Highly ordered arrangement of single neurons in orientation pinwheels. *Nature.* 442: 925-928.

Oliver MD (2014). Nonlinear modeling of the early and mid-level visual system. PhD Thesis, UC Berkeley: Vision Science. Retrieved from: <http://escholarship.org/uc/item/99p169h0>

Olmos A and Kingdom F. (2004) McGill calibrated colour image database.

Olsen SR et al. (2012) Gain control by layer six in cortical circuits of vision. *Nature* 000: 1-6

Pack C et al. (2001) Dynamic properties of neurons in cortical area MT in alert and anaesthetized macaque monkeys. *Nature* 414: 905-908

Pan Y et al. (2012) Equivalent representation of real and illusory contours in macaque V4. *The Journal of Neuroscience*. 32(20): 6760-6770

Pelli DG (1997) The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*. 10(4): 437-442.

Petrusca D, Grivich MI, Sher A, Field GD, Gauthier JL, Greschner M, Shlens J, Chichilnisky EJ and Litke AM. (2007) Identification and characterization of a Y-like primate retinal ganglion cell type. *The Journal of Neuroscience* 27(41):11019-11027.

Pillow et al., (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J Neurosci*. 25(47): 11003-11013.

Poletti M et al. (2013) Microscopic eye movements compensate for nonhomogeneous vision within the fovea. *Current Biology* 23: 1691-1695.

Pospisil, D., Pasupathy, A., & Bair, W. (2016). Comparing the brain's representation of shape to that of a deep convolutional neural network. In Suzuki, J. Nakano, T. & Hess H. (Eds.), *The first*

international workshop on computational models of the visual cortex: hierarchies, layers, sparsity, saliency and attention. New York: ACM.

Prenger R et al., (2004) Nonlinear V1 responses to natural scenes revealed by neural network analysis. *Neural Network*: 17(5):663-79.

Priebe NJ and Ferster D (2005) Direction selectivity of excitation and inhibition in simple cells of the cat primary visual cortex. *Neuron* 45:133-145.

Reid RC, Victor JD and Shapley RM (1997) The use of m-sequences in the analysis of visual neurons: linear receptive field properties. *Visual Neuroscience*. 14(6): 1015-27.

Reid C and Alonso J (1995) Specificity of monosynaptic connections from thalamus to visual cortex. *Nature* 378: 281-284.

Ringach DL, Sapiro G, Shapley R (1997) A subspace reverse-correlation technique for the study of visual neurons. *Vision Res* 37:2455-2464.

Ringach DL, Shapley RM and Hawken MJ (2002) Orientation selectivity in macaque V1: diversity and laminar dependence. *J Neuroscience*: 22(13):5639-5651.

Ringach DL (2002). Spatial structure and symmetry of simple-cell receptive fields in Macaque primary visual cortex. *Journal of Neurophysiology*: 88:455-463.

Ringach D and Shapley R (2004) Reverse correlation in neurophysiology. *Cognitive Science*, 28:147-166.

Rodieck RW (1979) Visual pathways. *Annual Review Neuroscience* 2:193-225.

Rosenberg A, Husson, Issa N (2010). Subcortical representation of non-Fourier image features. *J Neuroscience*. 30:1985-1993.

Rosenberg A and Issa N (2011) The Y cell visual pathway implements a demodulating nonlinearity. *Neuron*. 71: 348-361.

Rucci M et al. (2007) Miniature eye movements enhance fine spatial detail. *Nature* 447: 852-855.

Schiller P (2010) Parallel information processing channels created in the retina. *PNAS* 107(40):17087-94.

Schmid AM et al., (2009) Subpopulations of neurons in visual area V2 perform differentiation and integration operations in space and time. *Front Syst Neurosci*. 3:15.

Schofield AJ, Georgeson MA (1999) Sensitivity to modulations of luminance and contrast in visual white noise: separate mechanisms with similar behaviour. *Vision Res.* 39:2697-2716.

Schofield AJ et al., (2006) Local luminance amplitude modulates the interpretation of shape-from-shading in textured surfaces. *Vision Research.* 46(20): 3462-3482.

Schofield A et al., (2010) What is second-order vision for? Discriminating illumination versus material changes. *Journal of Vision* 10(9):2.

Schwartz O et al., (2006) Spike-triggered neural characterization. *Journal of Vision* 6:484-507.

Schwartz et al (2012) The spatial structure of a nonlinear receptive field. *Nature Neuroscience* 15:1572-1580.

Sellers KK et al., (2015) Awake vs anesthetized: Layer-specific sensory processing in visual cortex and functional connectivity between cortical areas. *Journal of Neurophysiology* 113: 3798-3815.

Shariati NH and Freeman AW. (2012) A multi-stage model for fundamental functional properties in primary visual cortex. *PLoS ONE* 7(4): e34466.

Shoham S et al., (2006) How silent is the brain: is there a “dark matter” problem in neuroscience? *Journal of comparative Physiology*. 192(8): 777-784.

Skottun BC, De Valois RL, Grosf DH, Movshon JA, Albrecht DG, Bonds AB (1991) Classifying simple and complex cells on the basis of response modulation. *Vision Research* 31:1079–1086.

Smith AT, Greenlee MW, Singh KD, Kraemer FM and Hennig J. (1998) The processing of first- and second-order motion in human visual cortex assessed by functional magnetic resonance imaging (fMRI). *The Journal of Neuroscience* 18(10):3816-3830.

Smith AT and Ledgeway T (1997) Separate detection of moving luminance and contrast modulations: fact or artifact? *Vision Research* 37:45-62

Smith AT, Ledgeway T (1998) Sensitivity to second-order motion as a function of temporal frequency and eccentricity. *Vision Res*, 38:403-410.

Smith SL et al., (2013) Dendritic spikes enhance stimulus selectivity in cortical neurons in vivo. *Nature* 503:115-120.

Smith IT et al., (2017) Stream-dependent development of higher visual cortical areas. *Nature Neuroscience*. 20:200-208.

So YT and Shapley R (1979) Spatial properties of X and Y cells in the lateral geniculate nucleus of the cat and conduction velocities of their inputs. *Experimental Brain Research* 36(3): 533-550.

So YT and Shapley R (1981) Spatial tuning of cell in and around lateral geniculate nucleus of the cat: X and Y relay cells and perigeniculate interneurons. *Journal of Neurophysiology*. 45:107-120.

Song Y and Baker CL Jr (2006) Neural mechanisms mediating responses to abutting gratings: Luminance edges vs. Illusory contours. *Visual Neuroscience* 23(2): 181-199.

Song Y and Baker CL Jr (2007) Neural responses to texture- and contrast-defined boundaries in early visual cortex. *Visual Neuroscience* 24(1): 65-77.

Soodak R, Shapley R, Kaplan E (1987) Linear mechanism of orientation tuning in the retina and lateral geniculate nucleus of the cat. *Journal of Neurophysiology* 58(2): 267-275.

Spacek MA and Swindale NV (2015) Cortical state and natural movie responses in cat visual cortex. Preprint at <http://www.biorxiv.org/content/early/2016/05/23/031765>.

Stirman JN et al., (2016) Wide field-of-view, multi-region, two photon imaging of neuronal activity in the mammalian brain. *Nature Biotechnology*. 34(8): 857-862.

Stone J & Dreher B (1973) Projection of X- and Y-cells of the cat's lateral geniculate nucleus to areas 17 and 18 of visual cortex. *Journal of Neurophysiology* 36(3):551-567

Stone J, Dreher B, Leventhal A (1979) Hierarchical and parallel mechanisms in the organization of visual cortex. *Brain Research Reviews* 1:345-394.

Stosiek C et al., (2003) In vivo two-photon calcium imaging of neuronal networks. *PNAS*. 100: 7319-24.

Sutter A, Sperling G, Chubb C (1995) Measuring the spatial frequency selectivity of second-order texture mechanisms. *Vision Research* 35(7): 915-924.

Swindale, N. V., Mitelut, C., Murphy, T. H., Spacek, M. A. (2017). A Visual Guide to Sorting Electrophysiological Recordings Using 'SpikeSorter'. *J. Vis. Exp.* (120), e55217, doi:10.3791/55217

Swindale NV & Spacek MA (2014) Spike sorting for polytrodes: a divide and conquer approach. *Frontiers in Systems Neuroscience* 8:6.

Talebi V and Baker CL (2012) Natural versus synthetic stimuli for estimating receptive field models: a comparison of predictive robustness. *J Neurosci.* 32(5): 1560-1576.

Talebi V & Baker CL Jr (2016). Categorically distinct types of receptive fields in early visual cortex. *Journal of Neurophysiology* 115(5): 2556-2576.

Tanaka H and Ohzawa I (2006). Neural basis for stereopsis from second-order contrast cues. *The Journal of Neuroscience* 26(16):4370-4382.

Thomson MGA (2001) Beats, kurtosis and visual coding. *Network: Comput. Neural Syst.* 12:271-287.

Tolhurst DJ and Dean AF (1987) Spatial summation by simple cells in the striate cortex of the cat. *Exp Brain Res* 66:607-620.

Torrealba F et al., (1981) Organization of the projection from the superior colliculus to the dorsal lateral geniculate nucleus of the cat. *Neuroscience* 6:1341-1360.

Touryan et al., (2005) Spatial structure of complex cell receptive fields measured with natural images. *Neuron* 46: 781-791.

Turner MH, Rieke F (2016) Synaptic rectification controls nonlinear spatial integration of natural visual inputs. *Neuron*. 90:1257-1271.

Tusa RJ, Rosenquist AC, Palmer LA (1979) Retinotopic organization of areas 18 and 19 in the cat. *Journal of Comparative Neurology*. 185:657-678

Vaina LM, Makris N, Kennedy D, Cowey A (1998) The selective impairment of the perception of first-order motion by unilateral cortical brain damage. *Vis Neurosci*. 15:333-348.

Vaina LM, Cowey A (1996) Impairment of the perception of second order motion but not first order motion in a patient with unilateral focal brain damage. *Proc Roy Soc Land [Biol]* 263:1225-1232.

Vaina LM, Cowey A, Kennedy D (1999) Perception of first- and second order motion: separable neurological mechanisms? *Human Brain Mapp*. 7:67-77.

deValois R et al., (2000) Spatial and temporal receptive fields of geniculate and cortical cells and direction selectivity. *Vision Research*. 40: 3685-3702.

Vintch B, Movshon JA, Simoncelli EP (2015) A convolutional subunit model for neuronal responses in macaque V1. *The Journal of Neuroscience*. 35(44): 14829-41.

Wang X et al., (2011) Thalamic interneurons and relay cells use complementary synaptic mechanisms for visual processing. *Nature Neuroscience* 14(2):224-231.

Want et al., (2011) Inhibitory circuits for visual processing in thalamus. *Current opinion in Neurobiology*. 21:726-733.

Wang X et al. (2007). Feedforward excitation and inhibition evoke dual modes of firing in the cat's visual thalamus during naturalistic viewing. *Neuron* 55:465-478.

Wassle H et al., (1975) The distribution of the alpha type of ganglion cells in the cat's retina. *The Journal of Comparative Neurology*. 159(3): 419-437.

Willmore B, Smyth D (2003) Methods for first-order kernel estimation: simple-cell receptive fields from responses to natural scenes. *Network* 14:553-577.

Wilson et al (1976) Properties of relay cells in cat's lateral geniculate nucleus: a comparison of W cells with X-and Y-cells. *Journal of Neurophysiology* 39:1193-1209.

Wilson DE et al. (2016) Orientation selectivity and the functional clustering of synaptic inputs in primary visual cortex. *Nature Neuroscience* 19(8): 1003-1009.

Wold S et al., (1984) The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. Sci and Stat Comput*. 5(3): 735-743.

Wolfe J and Larry PA (1998) Temporal diversity in the lateral geniculate nucleus of cat. *Visual Neuroscience*. 15:653-675.

Wu M et al., (2006) Complete functional characterization of sensory neurons by system identification. *Ann Rev Neurosci*. 29:477-505.

Yamins DL et al., (2014) Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 111(23):8619-8624.

Yao Y et al., (2007) On early stopping in gradient descent learning. *Constr. Approx* 26:289-315.

Yeh C, Xing D and Shapley RM (2009) “Black” responses dominate macaque primary visual cortex V1. *The Journal of Neuroscience*. 29(38):11753-11760.

Zavitz, E., & Baker C. L. (2013). Texture sparseness, but not local phase structure, impairs second-order segmentation. *Vision Research*, 91, 45–55.

Zhou Y & Baker CL (1993). A processing stream in mammalian visual cortex neurons for non-Fourier response. *Science* 261:98-101.