

**An unsupervised approach to extracellular electrophysiology with biologically validated
cell type classification**

Éloïse Valérie Giraud

Thesis submitted to the Faculty of Social Sciences
in partial fulfillment of the requirements for the
Doctorate in Psychology

Department of Psychology, Faculty of Social Sciences

University of Ottawa

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
LEGEND	viii
ABSTRACT	xi
ACKNOWLEDGEMENTS.....	xii
CHAPTER 1: INTRODUCTION	1
1. ELECTROPHYSIOLOGICAL PROPERTIES OF THE MAMMALIAN BRAIN.....	1
2. ELECTROPHYSIOLOGY AND NEURONAL RECORDINGS	2
<i>The action potential.....</i>	2
2.1 The intracellular action potential	5
2.2 The extracellular action potential	5
3. COMPREHENDING MULTIELECTRODE ARRAY SIGNALS	6
3.1 IN VITRO RECORDING COMPLEMENTARY METAL-OXIDE-SEMICONDUCTOR RECORDING PLATFORMS	6
<i>Graphical depiction of the recording platform with an image of a prefrontal cortex slice (mice)</i>	
<i>mounted on the array.....</i>	7
3.2 SIGNAL COMPOSITION.....	7
4. THE PREFRONTAL CORTEX	8
5. DISTINCT CELL TYPES AND THEIR KNOWN ROLE IN NEURAL DYNAMICS.....	9
<i>Excitatory neurons.....</i>	10
<i>Inhibitory interneurons.....</i>	10
6. CURRENT CHALLENGES IN UNDERSTANDING THE BRAIN’S ARCHITECTURE USING ELECTROPHYSIOLOGY..	11
6.1 Need for automatized spike sorting of distinct cell types.....	11
6.2 Need to probe neuronal activity in large, intact networks, with light.....	12
6.3 GOAL OF THE THESIS	13
CHAPTER 2: AN AUTOMATED APPROACH TO LARGE-SCALE SPIKE SORTING	14
1. OVERVIEW OF SPIKE SORTING	14

1.1 Filtering.....	14
1.2 Spike detection	15
1.3 Single cell identification.....	16
Dimensionality reduction	17
Clustering	18
Linear discriminant analysis	19
1.4 Soma localization	19
2. CHALLENGES IN SPIKE SORTING	20
2.1 Overlapping spikes	20
2.3 No cell type specific sorting	20
2.4 No ground truth.....	21
3. IMPLEMENTING AN OPEN-SOURCE, SCALABLE PIPELINE	21
3.1 Data acquisition: High-Density Multielectrode Arrays	21
Methodology	21
3.2 Data processing: Implementing a spike sorting pipeline	23
3.2.1 Filtering.....	24
3.2.2 Spike detection	25
3.2.3 Single cell identification.....	28
3.2.4 Soma localization	32
CHAPTER 3: CELL TYPE CLASSIFICATION THROUGH VOLTAGE WAVEFORM ANALYSIS.....	35
1. ELECTROPHYSIOLOGICAL PROPERTIES OF DISTINCT CELL TYPES IN CORTEX	36
Regular spiking excitatory cells	36
Fast spiking inhibitory cells.....	37
2. SPIKE SORTING OF DISTINCT CELL TYPES	38
3. A METHOD UTILIZING WAVEFORM PROPERTIES IN LARGE NETWORK.....	40
3.1 Time resolution	40
3.2 Feature extraction.....	41
3.3 Identification of putative excitatory and inhibitory cells using k-mean clustering	42

CHAPTER 4: A STEP-FUNCTION OPSIN PROTOCOL FOR CELL CLASSIFICATION	48
1. PROBING NEURONAL ACTIVITY ON MEA PLATFORMS.....	48
<i>Optogenetics</i>	48
<i>Stabilized step-function opsin</i>	49
<i>Probing activity on MEA platforms: shortfalls</i>	49
2. CELL TYPE SPECIFIC PROBING THROUGH OPTOGENETICS ON MEA PLATFORM	50
<i>Methodology</i>	51
3. PATCH-CLAMP VALIDATION OF VIRAL STRATEGY	54
<i>Protocol established</i>	57
CHAPTER 5: OPTOGENETIC VALIDATION OF E/I CLASSIFICATION ON MULTIELECTRODE ARRAYS.	58
1. QUANTIFYING POPULATION-LEVEL ACTIVITY CHANGES TO STEP FUNCTION OPSIN ACTIVATION OF PARVALBUMIN INHIBITORY INTERNEURONS	58
2. A DIVERSITY OF INDIVIDUAL NEURON RESPONSES TO STEP-FUNCTION OPSIN ACTIVATION OF PARVALBUMIN INHIBITORY INTERNEURONS	60
3. CELL TYPE-SPECIFIC DYNAMICAL INTERACTIONS IN LARGE-SCALE NETWORKS	63
<i>Probing activity in intact networks</i>	63
CHAPTER 6: DISCUSSION.....	69
6.1 <i>Summary of results</i>	69
6.2 <i>Challenges in large-scale neural recordings</i>	71
6.3 <i>Limits of this work and future directions</i>	72
CONCLUDING REMARKS.....	80
APPENDICES	81
APPENDIX A.....	82
APPENDIX B.....	91
APPENDIX C	102
BIBLIOGRAPHY	121

LIST OF TABLES

Table 1: Comparison of available spike sorting methods	77
--	----

LIST OF FIGURES

Figure 1: Multielectrode array recording platform with prefrontal cortex slice	7
Figure 2: Action potential propagation in a neuron's compartments	17
Figure 3: Multielectrode array recording set up	23
Figure 4: Workflow diagram of the proposed method	24
Figure 5: A python package to process voltage and extract spikes	25
Figure 6: Spike detection on the extracellular voltage of a single electrode and raw voltage detection around the detected spikes	26
Figure 7: The firing rate per electrode depends on the threshold selected	27
Figure 8: Optimal threshold selection allows to part cellular activity from normally distributed noise	28
Figure 9: K-mean clustering of detected spikes on a single electrode reveals cluster with distinct average waveforms	29
Figure 10: Ten-fold cross-validation shows better than chance performance	30
Figure 11: Voltage amplitude decays as a function of distance from the electrode where the highest peak in activity was detected	31
Figure 12: The current sink signal from a single neuron	32
Figure 13: The center of mass location of all spikes for a single soma captured by one recording electrode	33
Figure 14: Spline interpolation	41
Figure 15: Action potential waveform features	41
Figure 16: Distinct waveform kinetics for E and I cells	44
Figure 17: Respective cell location	45
Figure 18: Raster plot of E and I activity	46
Figure 19: Fano factor for putative E and I cells	46
Figure 20: Mean correlation as a function of spatial distance	47

Figure 21: Light artefact	50
Figure 22: Visually identified EYFP expressing PV neuron	52
Figure 23: Viral injection expression	54
Figure 24: Nearby pyramidal neuron receiving inhibitory currents following PV opto-activation	55
Figure 25: Inhibitory post-synaptic currents onto pyramidal cell depends on rest period	56
Figure 26: Optogenetic protocol	56
Figure 27: Average firing rate of putative excitatory and inhibitory neurons for a single slice....	59
Figure 28: Quantification of change in firing rate for 3 slices.....	60
Figure 29: Delta change in firing rate (on – drugs)	61
Figure 30: Opto-evoked change in firing rate	62
Figure 31: Average firing rate for putative excitatory and inhibitory neurons following opto- activation when no blockers are applied	64
Figure 32: Change in firing rate following optogenetic activation.....	65
Figure 33: Inhibition-stabilized network architecture	67

LEGEND

aCSF	artificial cerebrospinal fluid solution
AIS	axon initial segment
AP	action potential
APos	anterior-posterior
APS	active pixel sensor
AHp	afterhyperpolarization
AHP	afterhyperpolarizing potential
ADp	afterdepolarization
ChR	channelrhodopsin
CT	corticothalamic
CMOS	complementary metal-oxide-semiconductor
CNQX	cyanquixaline
CV	cross-validation
DV	dorsal-ventral
E	excitatory
EAP	extracellular action potential
FR	firing rate
FS	fast spiking
FRB	fast repetitive bursting
GABA	gamma-aminobutyric acid
Glu	glutamate
HAW	half-amplitude width
HD-MEAs	high-density multi electrode arrays
Hz	hertz
I	inhibitory

ISN	inhibition-stabilized network
I_h	hyperpolarization-activated current
IPSCs	inhibitory postsynaptic currents
IAP	intracellular action potential
IB	intrinsically bursting
IT	intratelencephalic
L	layer
LDA	linear discriminant analysis
LFP	local field potential
LTS	low threshold spikes
MEA	multielectrode array
ML	medial-lateral
mPFC	medial prefrontal cortex
Ms	millisecond
PCA	principal component analysis
PFC	prefrontal Cortex
PP	peak-to-peak
PT	pyramidal tract
PTX	picrotoxin
PV	parvalbumin
rrSVD	rank-reduced singular value decomposition
RS	regular spiking
S	second
SEM	standard error of the mean
SD	standard deviation
SOM	somatostatin
V	volts

VIP	vasoactive intestinal peptide
VRC	variance ratio criterion

ABSTRACT

Large-scale extracellular recording techniques represent a major advance in interrogating the structure and function of neuronal circuits. However, the field lacks methods that can resolve cell type identity in a principled way, while simultaneously scaling to thousands of neurons. Here, I introduce a pipeline for the analysis of large-scale recordings of in vitro cortical activity that not only allows for the detection of spikes produced by single neurons (spike sorting), but also allows for the reliable distinction between genetically determined cell types by utilizing viral and optogenetic strategies to provide a ground-truth validation. Importantly, this novel methodology tightly integrates the analysis pipeline to an experimental protocol allowing for the dynamical probing of distinct cell types while simultaneously recording from large populations. The novelty of the proposed approach is to combine a stream of analysis, and experimental techniques in an end-to-end fashion as follows. First, individual spike waveforms are fitted by spline interpolation to estimate their half-amplitude and peak-to-peak durations. These values are then entered in a principal component analysis with k-means clustering to identify uncorrelated signals from single channels on the array. Optimal separability of clusters is assessed by linear discriminant analysis. Finally, each channel's source location is identified using spatiotemporal characteristics of spike waveforms across the array. I validated this pipeline using activity monitored from mouse prefrontal cortex in vitro slices with an array of 4,096 closely-spaced channels, in conjunction with optogenetic, viral and pharmacological strategies. I show an effective distinction of regular-spiking excitatory neurons from genetically validated fast-spiking inhibitory interneurons using measures of action potential waveform, Fano factor, and spatially-dependent correlations. In sum, the proposed approach allows for a comprehensive characterization of neuronal activity obtained from different cell types in high-density multielectrode recordings. This provides a scalable approach to investigate the interplay between distinct cell types in microcircuits of the brain.

ACKNOWLEDGEMENTS

The work presented in this thesis was conducted over several years and represents the most educational chapter of my life. It has been punctuated by challenges, and a steep learning curve which allowed me to develop a new kind of creativity and resilience.

First and foremost, I would like to thank my Ph.D. advisor, Dr. Jean-Philippe Thivierge for believing in me and taking me in his computational neuroscience laboratory knowing I had no formal training in programming. Your mentorship, guidance and support not only made this adventure possible, but pleasant.

I would like to thank Dr. Jean-Claude Béïque for his support over the years, and for always finding time to entertain fascinating science discussions regardless of how late it was. I am grateful for your support leading up to talks I had to give, your encouragement never went unnoticed and helped me become more comfortable to give presentations.

Thank you to my thesis committee members, Dr. Jean-Claude Béïque, Dr. Sylvain Chartier, and Dr. Denis Cousineau. Reading and evaluating a Ph.D. thesis represents a lot of work. Thank you for taking the time to evaluate my work.

I am extremely grateful to have been surrounded by exceptional lab mates, Sébastien Maillé, Emerson Harkin, Léa Caya-Bissonnette, Philippe Vincent-Lamarre, Megan Boucher-Routhier, Camille Godin, and many more. Your unwavering support and camaraderie have been invaluable throughout this educational journey we picked.

To Michael, whose presence has been a driving force behind my academic journey. Thank you for being a constant source of inspiration, for believing in my abilities in challenging moments, and for your unconditional support.

To my best friends, Sophiane, Marie-Pier, Valéry and Élizabeth, thank you for always believing in me, for cheering me on through the highs and lows, and for always pushing me to be the best version of myself. Your unconditional love never goes unnoticed.

To my parents, Nathalie, and Jean-François, for their unwavering support, sacrifices, and endless love. Thank you for instilling in me a passion for learning, and for always believing in my potential. To my little brother, Anthony, for the incredible childhood, and now adulthood, filled with adventures that we embarked on. Your adventurous spirit, and curiosity, have shaped my approach to life and learning.

COPYRIGHTED CONTENTS

Appendix A: This article was published in *Cognitive Computation* in 2023.

“Articles are published open access under a CC BY licence (Creative Commons Attribution 4.0 International licence). The CC BY licence is the most open licence available and considered the industry 'gold standard' for open access; it is also preferred by many funders. This licence allows readers to copy and redistribute the material in any medium or format, and to alter, transform, or build upon the material, including for commercial use, providing the original author is credited.”

Appendix B: This article was published in *Chaos* in 2022.

“Copyright (2022) Jean-Philippe Thivierge, Eloïse Giraud, Michael Lynn, Annie Thériège Charbonneau. This article is distributed under a Creative Commons Attribution (CC BY) License.” “AIP Publishing grants to the author(s) of papers submitted to *Chaos* the right to post their accepted manuscript anywhere on the Web immediately after acceptance by AIP Publishing. In addition, 12 months after publication, the final AIP Publishing version may be posted on the author's personal website, the author's institutional website, or in an institutional or funder-designated repository.”

Appendix C: This article was published in *JNeurosci* in 2017.

“Authors grant *JNeurosci* a license to publish their work and copyright remains with the author. For articles published after 2014, the Society for Neuroscience (SfN) retains an exclusive license to publish the article for 6 months; after 6 months, the work becomes available to the public to copy, distribute, or display under the terms of the Creative Commons Attribution 4.0 International License (CC-BY). This license allows data and text mining, use of figures in presentations, and posting the article online, provided that the original article is credited.”

Chapter 1: Introduction

1. Electrophysiological properties of the mammalian brain

The mammalian brain is a remarkably intricate and complex structure that lays the foundation of our cognitive processes and bodily functions. It is composed of billions of neurons, each with elaborate interconnections through an extensive network of electrical and chemical synapses. Neurons are the fundamental building blocks of the brain. They are the biological basis of our brain's processes, allowing us to experience the world around us. The intricate dynamics that arise from their meticulously coordinated communication lays the foundation of our actions, thoughts, and internal states. Thanks to their unique properties, they are an integral part of local and global connectivity, actively participating in the integration and transmission of multiple synaptic inputs. Such interactions lead to the complex behaviors that define us.

The history of neuroscience lays in a collective attempt to reconcile the precision of molecular mechanisms, and the web of their connections, with the complexity of emergent brain functions; it is the understanding of a complex and precisely regulated neural code. To succeed in encoding various sensory and cognitive information, the brain must extract and analyze neural activity from a multitude of distinct units. This raises a fundamental question regarding neural and synaptic mechanisms: what properties of a network dictate how brain activity underlies our behavior?

The complexity of the brain arises not only from the number of neurons, and their sophisticated communication, but also from its diverse cell types. A fundamental objective of modern neuroscience is to understand the physiological composition of neural circuits and bridge it with its functional implications. Yet, understanding the unique contribution of each cell type to emerging network dynamics remains a challenging endeavour. At present, available

electrophysiological technologies have limited researchers' abilities to simultaneously stimulate and record the electrical activity of identified cell types in large, interconnected populations of neurons.

In the remainder of this introduction, I present concepts of electrophysiology and neuronal recordings more in depth (Section 2), spanning the action potential and its intricate properties. I introduce multielectrode array recordings and how we can comprehend recorded signals (Section 3), discuss the prefrontal cortex and its properties (Section 4), and present some of the distinct cell types that make up our brains and their known role in neural dynamics (Section 5). Finally, I outline two key challenges in understanding the brain's architecture using electrophysiology: the need for automatized spike sorting of distinct cell types (Section 6.1) and the need for large-scale optogenetic probing methods (Section 6.2). Lastly, I present the main goal of my thesis: to develop an integrated method to study distinct cell types in large, in vitro neural networks (Section 6.3).

2. Electrophysiology and neuronal recordings

The remarkable properties of neurons allow them to retain an electric potential across their membrane, which incoming signals can modulate. Due to an uneven distribution of ions, the inside of a neuron's cell membrane has a negative charge at rest. This polarization is key for electrical signalling: a stimulation will lead a neuron to depolarize, making the cell's membrane more positive. This, in time, will lead to the triggering of an action potential (AP).

The action potential

In the brain, information is carried to, and from, neurons by both electrical and chemical signalling. When a neuron is not active, it remains at resting membrane potential. The resting

potential is established by the distribution of ions across the membrane, and its permeability to each ion type. Brief electric signals are the product of transient changes in the electric current, flowing in and out of neurons, bringing the cell's membrane electric potential away from its resting potential. This, in time, will lead to an action potential. When this event happens, the change in electrical potential travels along a neuron's axon, powered by ionic channels opening and closing in a meticulously coordinated fashion. At the neuron's terminal, the arrival of the electrical signal will lead to the release of neurotransmitters, dispatching the signal to the next neuron.

The AP is triggered by a change in membrane potential. At rest, a neuron has a voltage across its membrane called the resting membrane potential which is determined by both the concentration gradients of ions across the membrane and by the membrane permeability to each type of ion. In a resting neuron, there are concentration gradients across the membrane for sodium and potassium. Ions move down their gradients via channels, causing a separation of charge that creates the resting potential. The membrane being more permeable to potassium than to sodium, the resting potential is closer to the equilibrium potential of potassium.

A depolarization in the membrane changes this balance towards the threshold for an action potential. As the membrane potential approaches threshold, voltage-gated sodium channels open rapidly. The increase in membrane conductance to sodium causes the sodium influx to exceed the potassium efflux, creating a net influx of positive charge causing further depolarization. The increase in depolarization causes still more voltage-gated sodium channels to open, resulting in a greater influx of sodium, which accelerates the depolarization even more. In some neurons increasing the membrane potential by 10 mV (from -65 to -55 mV) makes the membrane much more permeable to sodium than to potassium. The resultant influx of positively charged sodium neutralizes the negative charge inside the cell and causes a brief

and explosive change in membrane potential to +40 mV. The membrane potential never quite reaches equilibrium for sodium because potassium efflux continues throughout the depolarization. So many voltage-gated sodium channels open during the rising phase of the action potential that the cell membrane's sodium conductance is much greater than the conductance of either chloride or potassium. So, at the peak of the action potential the membrane potential approaches the sodium equilibrium potential and just as at rest (when permeability to potassium is predominant) the membrane potential tends to approach the potassium equilibrium potential.

Action potentials are initiated at a specialized trigger region near the origin of the axon called the *initial segment* from which they propagate down the axon, without failure or distortion, at speeds of about 1 to 100 meters per second (s). The amplitude of an AP traveling down the axon remains constant at 100 mV because the AP is an all-or-none impulse that is regenerated at regular intervals along the axon. The AP is conducted down the cell's axon to the axon's terminal, where it initiates an elaborate chemical communication with other neurons (or cells). An action potential typically lasts approximately 1 millisecond (ms), after which the membrane returns to its resting state, with its normal separation of charges and higher permeability to potassium than to sodium.

The electrical properties of neurons allow one to measure its activity using recording electrodes. This is possible because the generation of an action potential leads to the rapid flow of electric currents in, and out, of a neuron. When an action potential is fired, the fast voltage deflection it causes is recorded as a "spike". Single electrodes can be used to perform intracellular recordings and precisely record from a finite number of carefully probed neurons. This allows for good electrical coupling with the neuron, providing an accurate readout of its electrical voltage.

2.1 The intracellular action potential

The intracellular action potential (IAP) has a waveform that rapidly rises in voltage (from about -70 mV, to about $+10$ mV), and then drops, a little below the resting membrane potential before returning to its resting membrane potential. This typical sequence is due to the precise time activation of distinct ionic channels (see: The action potential). As such, IAPs start with a positive spike, and later show a negative voltage deflection. IAPs typically last around 1 ms, and measure tens of millivolts in amplitude (Buzsáki et al., 2012, Obien et al. 2015). IAPs require direct access to the inside of a neuron or its membrane.

When a neuron fires an action potential, the associated change in voltage can be captured in the extracellular space surrounding the neuron. This is due to the movement of ions, flowing in and out of the cellular membrane in response to the neuron's activity for about one millisecond. This current creates an electrical field which can be recorded by placing electrodes nearby. These electrodes capture the asymmetrical movement of charges in, and out, of a neuron's membrane. By doing so, they record the currents, associated with the electrical activity of a neuron, gaining a readout of their intrinsic activity.

2.2 The extracellular action potential

In the extracellular space, action potentials are typically identified as voltage deflection that exceeds an established threshold. At the beginning of an action potential, the rapid influx of sodium ions creates a sink, leading to a large negative spike in the extracellular action potential (EAP). After this, the slower efflux of potassium results in a source, leading to a small, and positive spike in voltage. EAPs typically last around 1 ms, just like IAPs, but measure around tens to hundreds of microvolts in amplitude (Buzsáki et al., 2012, Obien et al. 2015). EAPs can be detected by placing electrodes near (at about ~ 100 μm) the origin of a spike, typically in

proximity to the soma or the axon's initial segment (AIS). Since a microelectrode senses the changes in the extracellular field stemming from the movement of ions in and out of the intracellular space of cells, it has allowed researchers to look at the detailed propagation, in space, of the voltage associated with an action potential in a single neuron's compartments (Obien et al. 2015, Henze et al. 2000).

3. Comprehending multielectrode array signals

3.1 In vitro recording Complementary Metal-Oxide-Semiconductor recording platforms

Previously, extracellular recordings were conducted with single electrodes, or tetrodes (made with four microwires). Recent technological advances have seen innovative development propel the capabilities of neural recording systems. As such, the integration of neural probes and electronic circuits in Complementary Metal-Oxide-Semiconductor (CMOS) technology has allowed for denser (more electrodes), and more precise (lower noise) recording platforms. In such platforms, an array of electrodes is distributed over a large area to record (and sometimes, stimulate) the electrical activity of thousands of cells (Figure 1). These recording chip designs are based on the design of Active Pixel Sensor (APS) technology commonly used as light sensors in digital cameras, but instead of detecting light changes, detects small voltage variations resulting from cellular activity. These have advanced our understanding of how populations of neurons may represent and process information, offering crucial insights into complex network dynamics (Riquelme et al., 2023; Hemberger et al., 2019; Shein-Idelson et al., 2017).

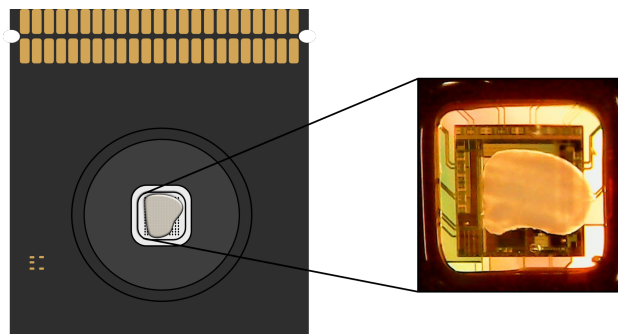


Figure 1: Multielectrode array recording platform with prefrontal cortex slice

Graphical depiction of the recording platform with an image of a prefrontal cortex slice (mice) mounted on the array.

3.2 Signal composition

The electrical signal recording using extracellular platforms is composed of several elements, all contributing differently to the overall measured voltage. Below, the overall signal's main components, the neuronal activity, and noise are detailed.

Neuronal activity

The component of the electrical signal, generally of most interest to researchers is the neuronal activity. By the opening of distinct ionic channels, currents around neurons are generated in response to the ionic movement provoked by their activity (see: The Action Potential). The placement of electrodes nearby neurons allow for the recording of what is commonly known as “local field potentials” (LFPs), which is the electric signal surrounding neurons. This term refers to the electrical potential, or V_e , measured in Volts (V) with respect to a reference potential), generated by the superimposed contributions of all active cellular processes at a given location in the extracellular space (Buzsáki et al. 2012). The difference in V_e between two distinct spatial location leads to an electric field (measured in Volts per

distance) defined as the negative spatial gradient of V_e (Buzsáki et al. 2012). This signal reflects the collective activity of a population of neurons, rather than the activity of a single neuron (Kajikawa et al. 2011). By using thorough signal acquisition, processing, and analysis, the processed LFP can be used to examine the underlying neuronal activity. Single neurons can be identified using what we know about the electrical properties of action potential, and its propagation in the extracellular space surrounding the neuron (See: Chapter 2: Spike Sorting).

Noise

All types of recordings are prone to noise acquisition. In multielectrode array recording systems, three types of noise are most frequently observed. The first type is biological noise. This main source of noise is due to the combined electric activity from neurons, or other cells, located near the recording electrode. For example, this can include APs from other nearby neurons, and synaptic noise. These can be affected by several factors, such as the number of proximal neurons, their size, their morphology, and their respective activity. In addition to biological noise, the contact between the recording electrodes and the recording solution generates noise. A third source of noise found in recordings stems from the system's device, termed device noise, which is used to record, amplify, and digitize the signals.

4. The prefrontal cortex

The prefrontal cortex (PFC) plays a key role in higher brain functions, such as cognition, motivation, reward, and emotion (Miller et al. 2001; Robbins et al. 2009).

The cerebral cortex is composed of several layers, in which neurons interact locally and receive long-range inputs from other brain regions. The traditional laminar view of the sensory cortex divides it into 6 layers. Layer (L) 1 contains the apical dendrites of pyramidal neurons,

and populations of GABAergic interneurons (Schuman et al. 2019; Anastasiades et al. 2021). L2/3 contains intratelencephalic (IT) neurons which project to other cortical areas (Ueta et al. 2019). L4 being a major input layer, contains pyramidal cells which make local connections within cortex. L5 being an output layer, contains IT and pyramidal tract (PT) cells targeting subcortical regions (Harris and Shepherd, 2015; Baker et al. 2018). L6 is also an output layer, containing neurons such as corticothalamic (CT) cells projecting to thalamic regions (Baker et al. 2018). Several types of interneurons are distributed across all layers, including parvalbumin (PV), somatostatin (SOM), and vasoactive intestinal peptide (VIP) cells (Harris and Shepherd, 2015, Tremblay et al. 2016) (Anastasiades & Carter, 2021).

In the medial prefrontal cortex (mPFC) of mice, a few things are notably different. First, the mPFC is agranular, meaning it does not have a L4. This layer typically receives canonical thalamic projections, but instead thalamic inputs are found to be distributed across L1, L2/3, L5 and L6. A second key difference is that projection neurons are found across L2 to L6, layers which also form reciprocal circuits with other regions through a variety of interconnected cell types including IT, PT, and CT cells communicating through local excitation (Harris and Shepherd, 2015, Ueta et al. 2019). A third distinction resides in the differences in the types, quantity, and location of inhibitory interneuron subtypes, which mediate local inhibition (Kim et al. 2017). As such, there seems to be fewer PV cells, and more SOM cells, which could lead to distinct local computations. A fourth difference is that long-range inputs from other cortical region bringing signals relevant for action, cognition, reward, and emotion arrive across all layers (Anastasiades & Carter, 2021).

5. Distinct cell types and their known role in neural dynamics

The brain is composed of a wide variety of cell types, each contributing differently to overall emerging network dynamics, in both physiological, and pathological, conditions. The last years

have witnessed a growing appreciation for the study of distinct cell types, and their specific contribution. Broadly, neurons can be divided into two main categories: excitatory cells, which constitute about ~80% or more of cortical neurons, and inhibitory cells, which make up the remaining ~20% of cells in the prefrontal cortex (Anastasiades & Carter, 2021)

Excitatory neurons

Excitatory neurons, also referred to as glutamatergic neurons, work by releasing the neurotransmitter glutamate (Glu). The released glutamate binds to receptors on the postsynaptic cell and will depolarize it, resulting in the cell being excited. Glutamate is the main excitatory transmitter of pyramidal cells. Essentially, these projection neurons are the excitatory counterpart of the inhibitory neurons. Importantly, glutamate is found in high abundance in the prefrontal cortex.

Inhibitory interneurons

Inhibitory neurons, also referred to as GABAergic interneurons, work by releasing the neurotransmitter gamma-aminobutyric acid (GABA), which binds to receptors on the postsynaptic cell and will hyperpolarize it. This, hyperpolarization decreases the chance of an action potential being fired, resulting in the cell being inhibited. The amino acid GABA is the prevailing inhibitory neurotransmitter in the central nervous system. It largely acts locally, serving interneurons and acting on neighbouring cells. GABA is just as prevalent in the prefrontal cortex as it is elsewhere. An estimate of one quarter of all neurons within the cerebral cortex uses GABA. It is also the most abundant of all the neurotransmitters that have been identified so far in the central nervous system and the prefrontal cortex is no exception to this (Fuster, 2015). GABA is mostly an intrinsic neurotransmitter that mediates activation of close-by neighbouring cells within local circuitry where inhibition is involved. GABA neurons are distributed all over throughout the cortex but are most abundant in layers II and IV (Hirsch, 1962). GABA is mostly present in non-pyramidal cells like granules and stellate cells (Houser,

1983). According to its even distribution and high concentration in intrinsic cells, it is highly plausible that in the prefrontal cortex GABA is at the basis of local inhibitory functions, potentially lateral inhibition that may accentuate the salience and contrast of excitatory patterns of prefrontal neurons (Emson, 1979; Fuster, 2015).

6. Current challenges in understanding the brain's architecture using electrophysiology

In the sections below, I introduce two of the main obstacles researchers face when trying to make sense of the brain's activity, namely the need for automatized spike sorting of distinct cell types (Section 6.1) and the need for large-scale optogenetic probing methods (Section 6.2). Lastly, I summarize the goal of this thesis and my Ph.D. work (Section 6.3).

6.1 Need for automatized spike sorting of distinct cell types

The recording of extracellular potentials is a widespread method to investigate neural activity, offering a high spatiotemporal resolution whilst monitoring action potentials from thousands of cells (Magland et al., 2020; Buzsáki, 2004). The advent of high-density, large-scale multielectrode array recordings has advanced our understanding of how populations of neurons may represent and process information, offering crucial insights into complex network dynamics (Riquelme et al., 2023; Hemberger et al., 2019; Shein-Idelson et al., 2017). Furthermore, the use of silicon recording probes such as Neuropixels probes in vivo, exhibiting similar spatiotemporal characteristics, has helped uncover foundational principles of neural circuits in awake, behaving mice (Lak et al., 2020; Peters et al., 2021). However, it remains a challenge to distinguish recorded cell types based on their physiological properties alone, despite a growing appreciation for the importance of heterogeneous cell types in circuit-level computations (Sylwestrak et al., 2022; Liu et al., 2021; Bugeon et al., 2022). Thus, the scalability of large-scale electrophysiological recording techniques depends not only on the

detection of spikes produced by single neurons (spike sorting), but also on the reliable distinction between the different cell types involved.

Modern extracellular recording techniques allow researchers to monitor the simultaneous activity of large, intact networks of neurons. However, in dense preparation, the activity of a single cell may be captured by several electrodes, and each electrode may report the activity of several nearby cells. To segregate the activity of single cells, effective spike-sorting techniques must be employed. A common lacking feature in currently available methods is the ability to identify distinct cell types in a biologically validated fashion. Despite evidence showing differences in action potential kinetics for distinct cell types, there exists no large-scale validation efforts, to my knowledge, showing that extracellular waveforms can be used to reliably distinguish cell types on high density multielectrode array (MEA) recordings. Methods leveraging genetic differences between cell types offer statistical measures predictive of cell types, often relying solely on characteristics of the action potential waveform such as the peak-to-peak distance and the half-amplitude width, and sometimes with metrics like the Fano factor (Becchetti et al., 2012). Such methods have provided researchers with the ability to infer, in a post-hoc fashion, the most probable cell types that were sampled. Yet, a fully integrated suite offering an open-source data analysis pipeline paired with a validated experimental protocol allowing for high-density cell type-specific activation during recordings is needed. Such a method will enable the investigation of dense and intricate circuits in a flexible, scalable, and generalizable manner.

6.2 Need to probe neuronal activity in large, intact networks, with light

The neuroscience field evolved with the requirement to precisely control neurons to decode neural circuitry. Optogenetics emerged in the mid-2000 (Deisseroth et al., 2006) as an approach allowing for the precise activation and monitoring of neuronal activity. This elegant

biological technique takes advantage of light to activate light-sensitive ion channels expressed in targeted cells. While multielectrode arrays have allowed the simultaneous recording of thousands of electrodes, the use of light-sensitive CMOS circuits, allowing such high density, has limited the ability of researchers to use them in conjunction with optogenetics. As the ability to rapidly perturb distinct populations in large network provides valuable insight as to how local circuits represent information (Lemon et al., 2021; Buzsáki, 2010), it is of broad interest to validate methodologies that allow for the optogenetic manipulation of defined cell populations in multielectrode arrays.

6.3 Goal of the thesis

The overarching goal of my Ph.D. work has been to develop an integrated method to overcome the two challenges outlined above to study distinct cell types in large, in vitro neural networks. More specifically, I worked on designing an open-source data analysis pipeline performing spike sorting on large-scale platforms. I worked on consolidating knowledge of action potential properties to extract statistical measures reflective of cell type identity. By means of developing this method, I worked on designing an experimental protocol to be used in conjunction with the data analysis pipeline. The designed experimental protocol aims to take advantage of several methodologies available in neuroscience, combining them in a way that enables researchers to dissect network properties. Crucially, I tightly integrate our analysis with an experimental protocol utilizing a stabilized step function opsin to dynamically probe distinct populations and validate sorting while avoiding the light-induced artefacts common in CMOS-based chip designs (Fiscella et al., 2012).

Chapter 2: An automated approach to large-scale spike sorting

A first step in addressing the need for automatized spike sorting of distinct cell types presented above (Chapter 1, Section 6.1) lies in designing an efficient, automatized, and scalable data analysis pipeline to making sense of neuronal data. In the first part of this chapter, I present a brief overview of the key elements (Section 1) and challenges (Section 2) that make up spike sorting analysis. Finally, I present an open-source, unsupervised and scalable spike sorting analysis pipeline which performs cell type classification which I demonstrate using multielectrode array data sampled from mice mPFC slices with 4096 electrodes (Section 3).

1. Overview of spike sorting

Observing the activity of single neurons lays the foundation for understanding how brain mechanisms underlay complex behaviors. Recent technological advances have allowed electrophysiologists to record thousands of densely packed microelectrodes, permitting the sampling of large populations of neurons with individual cell precision. Comprehending such data relies heavily on the detection of spikes produced by single neurons, achieved through spike sorting.

1.1 Filtering

The extracellular voltages recorded by multielectrode arrays typically reflect the dynamics of multiple cells in proximity to a given electrode (Rey et al., 2015), raising the problem of how to isolate individual units. As a necessary first step to isolate individual units, the extracellular voltage is commonly filtered using a second order Butterworth bandpass filter (300-5,000 Hz) (Yger et al. 2018; Rossant et al., 2016) With this, high pass filtering reduces the magnitude of

the low frequency components in the signal, and low pass filtering is used for extracting the LFP signal.

1.2 Spike detection

Once the raw data has been filtered, action potentials (spikes) can be distinguished from background noise and be detected using amplitude threshold crossings. This method is one of the simplest, and by far the most common way to measure neural activity (Lewicki, 1998). A crucial step to extract meaningful activity lies in selecting the proper threshold. A threshold value which is set to a value that is too small will likely lead to noise fluctuations being detected as events, introducing false-positive events. Alternately, using a threshold that is too large can result in the inability to detect spikes of low amplitude (Rey, 2015). Historically, thresholds were often set manually. This approach, however, has been shown to lead to user-biased results (Wood, 2014). Furthermore, selecting an optimal threshold manually requires a time-intensive intervention which does not scale to a constantly growing number of recording electrodes.

For most neurons, the most prominent feature of the action potential's shape is the amplitude (or "height") of the spike (Lewicki, 1998). This voltage deflection can be detected by an automatic threshold, set as a multiple of the standard deviation from the mean of the filtered signal,

$$V_{\text{thresh}} = x \sigma \quad (1)$$

To identify the proper threshold, the filtered voltage histogram of each channel can be fitted by a Gaussian curve to determine the standard deviation value at which non-normal, putative spike-related activity is found.

In data acquisition systems with a high sampling frequency, a single action potential commonly passes this threshold several times. To avoid detecting a single spike more than once, an absolute refractory period of 2 ms is applied, and successive events are discarded. Following event detection, a time window covering the full duration (5 ms) of each putative action potential is extracted, yielding a series of spike waveforms which will be the input to the spike sorting algorithm.

1.3 Single cell identification

In dense preparations, a single electrode can report the activity of more than one neuron. Like so, a single neuron's activity can be captured by more than one electrode. Although nearby neurons could have the same peak amplitude, the shape of the action potential waveform is likely to vary as it propagates throughout a neuron's compartment (Figure 2). Reports in the literature largely identify action potential initiation at the distal AIS (Obien et al. 2015). Once triggered, the action potential travels both towards the axon and backwards via the proximal AIS and towards the soma. The voltage deflection associated with an action potential is at its largest peak near the AIS, in close proximity to the neurons' soma. Through filtering and cable properties, the voltage deflection caused by an action potential gradually dampens as it travels towards the dendrites of the neuron.

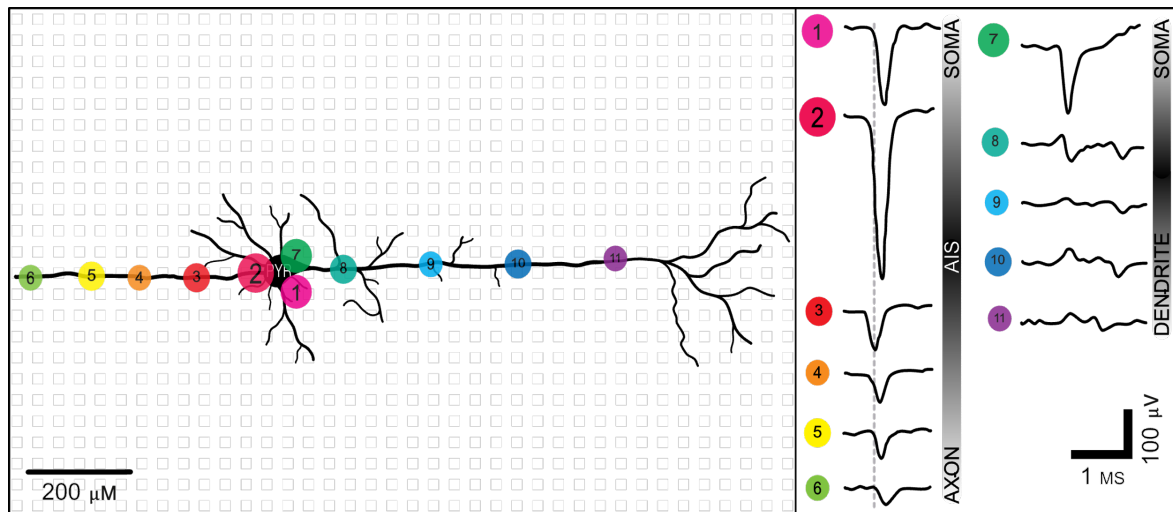


Figure 2: Action potential propagation in a neuron's compartments

Graphical depiction of the voltage propagation of an action potential in the compartments of a neuron based on current knowledge of the extracellular action potential (See: Chapter 1: Section 2.3.2 The extracellular action potential).

To identify and localize single neurons, the captured features (waveform) can be analyzed using dimensionality reduction.

Dimensionality reduction

When recording from dense populations, electrophysiological experiments only manage to sample sparsely, at random, from a system containing millions of neurons. A crucial step in spike sorting is to resolve the dimensionality of the dataset, thus going from the number of recording electrodes to the actual number of single neurons present in a recording. Depending on the number of neurons that are present, this can either expand or reduce the overall dimensionality of the data (Trautmann, 2019).

A method which allows for automatic feature selection is the principal component analysis (PCA). Using PCA, the data can be linearly transformed to a new coordinate system where the variation in the data is described with fewer dimensions than the initial data. This is achieved by finding an ordered set of orthogonal basis vectors which captures the data's greatest variation (Lewicki, 1998). Using this method, the aligned detected waveforms of each electrode can be used as input to identify the optimal number of components capturing the most variance in the data. For this, waveforms are first aligned around the largest voltage peak deflection to minimize variability in the action potential shape. Each action potential is represented by the principal components which are scaled, using a *score*, and added together. By scaling the scores according to the variability they capture, the first x components will represent most of the variation in the data. Subsequent components offer progressively smaller corrections until the action potential is described accurately. To obtain the principal component vectors, the eigenvectors of the covariance matrix are computed (Lewicki, 1998). The scores on the components representing most of the data's variation can next be used as features for clustering to resolve distinct neuron's spikes.

Clustering

A key element to make sense of complex neural recordings lies in reliably identifying meaningful clusters amongst data points. In a multidimensional dataset, this step allows for the identification of clusters, and for the classification of data points based on these clusters. This assumes that the data observed is from distinct, independent classes (Lewicki, 1998). On a single electrode, each cluster may represent the contribution of a distinct neuron.

Although cluster boundaries can be set by hand, large datasets benefit from an unbiased and automated approach. One of the most widely used approaches to clustering is the k-means clustering algorithm. This algorithm works by finding a set of centroids, for which the

total squared distance between each data point and the nearest centroid is minimized. Data points are assigned to clusters based on their distance to each centroid (Kauffmann, 2021).

Linear discriminant analysis

Next, to assess cluster separation in an unsupervised and unbiased fashion, a linear discriminant analysis (LDA) is then applied to the labelled principal components (Hill et al. 2011) to determine how well the data were separated by this approach. Linear discriminant analysis is one of the most widely employed classification techniques (Mai, 2013), allowing to estimate the probability that a given input belongs to each class. It provides an estimate of the linear classification error between the two clusters. One key limitation with the nature of our data lies in LDA's assumption that the data follows a Gaussian distribution, and that the covariance matrix is identical for each class (equal class covariance).

1.4 Soma localization

As voltage propagates through the arborization of a neuron, several electrodes may capture its activity. A substantial challenge in dense preparations sampled by closely space electrodes is to ensure the activity of each neuron is only reported once. For this, a radius of 250 μm , corresponding to the approximate spread of extracellular activity around the soma of individual neurons, can be applied (Jia et al. 2019). Within this radius, the channel with the largest negative deflection in the mean waveform can be identified. Previous work observed that the largest negative amplitude of extracellular signal originates from the axon initial segment (AIS) (Bakkum et al. 2019). Utilizing spatial localization of the AIS, the channel showing the greatest negative deflection can be considered to be the closest to the putative soma of a given cell (Jia et al. 2019).

2. Challenges in spike sorting

2.1 Overlapping spikes

A key challenge in spike sorting occurs when action potential from at least two distinct neurons overlap in time. This can introduce two types of error: missed spikes (type 2 error) and false spikes (type 1 error). First, missed spikes may happen when the peak voltage of a neuron is aligned with the dip in the voltage of another nearby neuron. Secondly, false spikes could be detected when two distinct action potential, from different neurons, occurs simultaneously.

Furthermore, as technological progress has allowed the sampling of dense tissue by a growing number of electrodes which are tightly packed, a growing number of neurons captured by a single electrode may result in more frequent overlap. This highlights the need for optimized method scaling to a constantly growing ratio of cells per electrode.

2.3 No cell type specific sorting

Neurons are normally divided into two main categories, based on the effect they have on other neurons. These are excitatory neurons, which produce an excitatory post-synaptic potential and inhibitory neurons, which trigger an inhibitory post-synaptic potential. Diverse subtypes of neurons have been identified as playing a unique and important role in neural computations. Yet, to understand how such an elaborate repertoire emerges in circuits, we must first reliably identify distinct neural cell types.

As discussed above (see: Spike-sorting and interneurons), several characteristics allow for the identification of cell types, including their morphology, gene expression, network

connectivity and their electrical properties. However, current spike sorting method aimed for extracellular recordings do not allow for the distinction between cell types.

2.4 No ground truth

In extracellular electrophysiology, ground-truth information of spiking activity is unavailable. This is also true for the identity of the cell types recorded. Taken in conjunction with the unsupervised nature of spike sorting, this highlights the need for a reliable and validated spike sorting pipeline through several quality control metrics.

In the section below, I present the application of an open-source, unsupervised and scalable spike sorting analysis pipeline which performs cell type classification on data acquired from High-Density Multielectrode Arrays (HD-MEAs).

3. Implementing an open-source, scalable pipeline

3.1 Data acquisition: High-Density Multielectrode Arrays

Methodology

The extracellular voltage was recorded using High-Density Multielectrode Arrays from 3 Brain, Switzerland (www.3brain.com). A detailed description of the technology behind this system can be found in Imfeld et al. (2008) and in Berdondini et al. (2009). High density recordings from the mouse's prefrontal cortex were performed in vitro using the BioCAM X platform with HD-MEA Arena chips (3Brain GmbH, Switzerland). This system records using 4,096 square microelectrodes (pitch = 42 μm) positioned in a 64 x 64 layout (2.67mm x 2.67mm area) with a sampling rate f_s of 18 kHz when recording from the entire 64 x 64 array.

Activity was recorded at 12 bits resolution per channel. The data was visualized and acquired with the BrainWaveX software provided by 3Brain.

mPFC slices were prepared from P21 old wild-type mice. The animals were kept on a 12:12-h light/dark cycle, with access to food and water ad libitum. All experiments and procedures were performed in accordance with approved procedures and guidelines set forth by the University of Ottawa Animal Care and Veterinary Services (protocol # 3471). For slice preparation, mice were anesthetized by inhalation of isoflurane (Baxter Corporation) and killed by decapitation. The brain was removed, and coronal brain slices were sectioned while immersed in an ice-cold choline chloride-based cutting solution of the following composition: 119 mM choline-Cl, 2.5 mM KCl, 1 mM CaCl_2 , 4.3 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM NaH_2PO_4 , 1.3 mM sodium L-ascorbate, 26.2 mM

NaHCO_3 , and 11 mM glucose, and equilibrated with 95% O_2 and 5% CO_2 . Slices were recovered in a chamber containing standard artificial cerebrospinal fluid solution (aCSF) of the following composition: 119 mM NaCl, 2.5 mM CaCl_2 , 1.3 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM NaH_2PO_4 , 26.2 mM NaHCO_3 , and 11 mM glucose, at a temperature of 37 °C, continuously bubbled with a mixture of 95% O_2 and 5% CO_2 . Slices were recovered for 1 h in the recovery chamber and equilibrated to a temperature of ~25 °C before the recordings were performed. The tissue was continuously perfused with fresh aCSF.

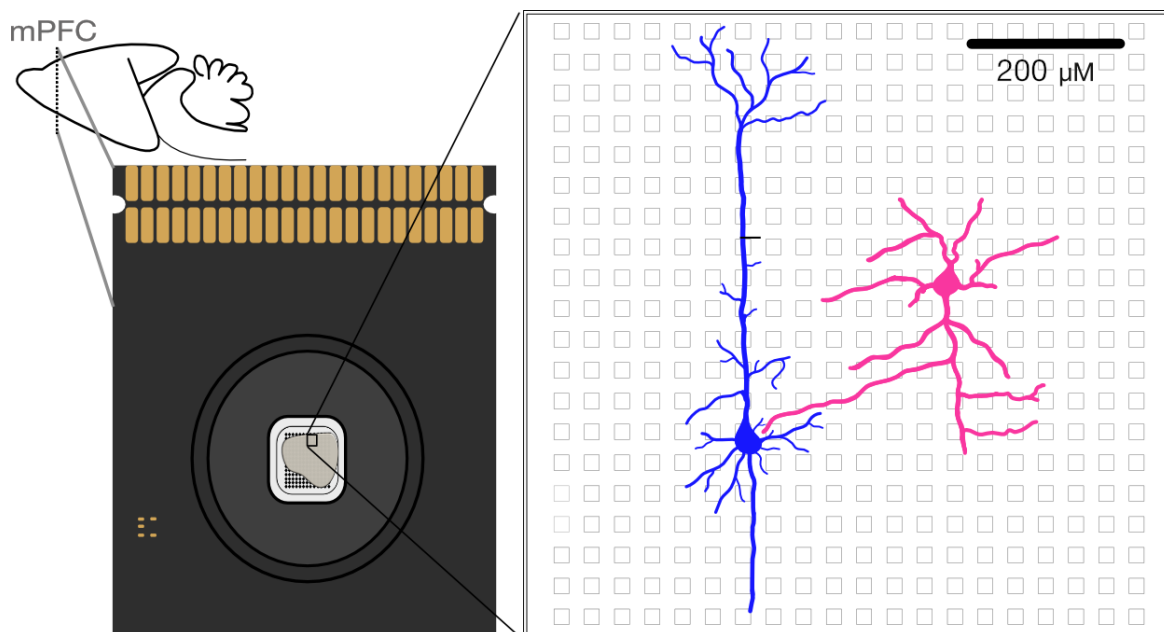


Figure 3: Multielectrode array recording set up

Graphical depiction of the multielectrode array recording platform. Note that the size of the electrodes and neurons depicted are to scale.

3.2 Data processing: Implementing a spike sorting pipeline

The full package and the github documentation associated with it will be available at <https://github.com/evgiraud>

The spike sorting pipeline I implemented performs four main tasks: 1. the detection of voltage deflections associated with somatic spikes, 2. the identification of individual putative neurons using spike clustering, 3. the spatial localization of single cells in high-density recordings, and 4. the classification of distinct cell types based on action potential waveform properties (Figure 4).

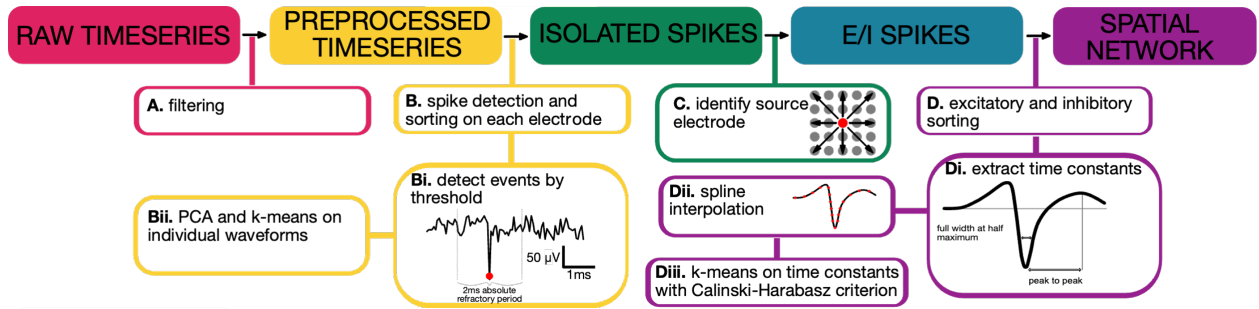


Figure 4: Workflow diagram of the proposed method

(A) Filtering of the data is achieved through bandpass filtering (300 to 5000 Hz). Spike detection is done on filtered voltages, however raw voltages around each spike to avoid signal distortion. **(B)** Spike sorting on each electrode. **(Bi)** Events are detected using a threshold method (4.5 standard deviations from the mean of the filtered signal). **(Bii)** PCA and k-means on individual waveforms ($k = 2$) **(C)** Precise soma location on the array. **(D)** Excitatory and inhibitory sorting. **(Di)** Spike timing kinetics are extracted at the half-amplitude width and peak-to-peak distances. **(Dii)** Interpolated voltage for increased resolution **(Diii)** k-means to distinguish cell type based on their waveform kinetics.

I demonstrate the use of this pipeline through targeted high-density recording from mouse medial prefrontal cortex (Figure 3).

3.2.1 Filtering

As a necessary first step to isolate individual units, the extracellular voltage was filtered using a second order Butterworth bandpass filter (300-5,000 Hz) (Figure 5) to remove slowly changing field potentials and high frequency noise (Bullman et al., 2019). Recording artefacts (voltage deflection $> \pm 500$ μV) were removed and set to the mean voltage.

3.2.2 Spike detection

Putative spike times were detected using a voltage-threshold method (Lewicki, 1998), identifying negative signal peaks below a threshold of τ standard deviation (SD) from the mean of the filtered signal. Here, τ is calculated for each recording, by plotting the filtered voltage histogram for each channel and fitting it with a Gaussian curve to determine the SD value at which non-normal, putative spike-related activity is found. To avoid multiple detections of the same event, an absolute refractory period was applied and successive events within <2 ms are discarded. Next, the unfiltered voltage around the time of each spike (from -5 to $+5$ ms) was extracted (Figure 6), yielding a series of high-frequency (18 kHz) waveforms for each channel. For the data shown here, $\tau = 4.5$.

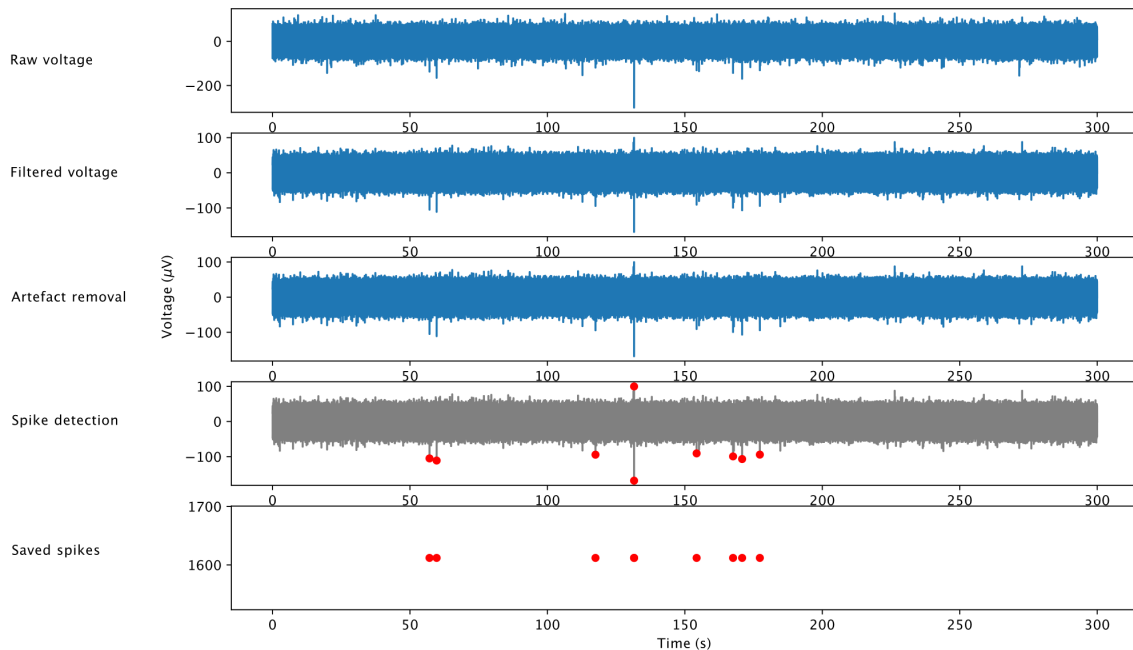


Figure 5: A python package to process voltage and extract spikes

An example plot depicting the voltage of a single recording electrode. **Raw voltage:** The raw signal of each electrode is extracted. **Filtered voltage:** Voltage following filter application.

Artefact removal: Voltage upon artefact removal and absolute refractory period. **Spike**

detection: Negative signal below threshold are detected. **Saved spikes:** Event times are saved and a window around each action potential is extracted.

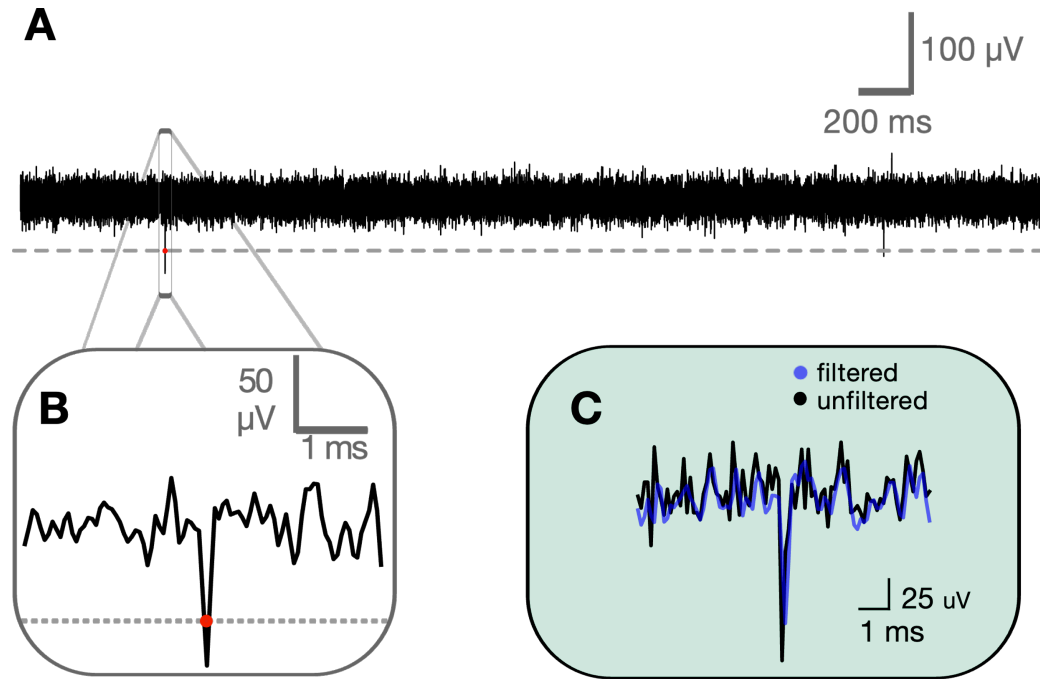


Figure 6: Spike detection on the extracellular voltage of a single electrode and raw voltage detection around the detected spikes

(A) Extracellular voltage on a single electrode (B) Spike detection on filtered voltage (C) Raw voltage extraction around detected spike

Single cell detection

For each channel, waveforms were stored in a matrix $\mathbf{X} \in \mathbb{R}^{[N \cdot T]}$ where N is the total number of spikes and $T=180$ are time points surrounding each spike. This matrix was analyzed using rank-reduced singular value decomposition (rrSVD) with $k=3$ dimensions.

Optimal threshold choice is key in discriminating spikes from noise, as the number of spikes detected per electrode is highly dependent on the threshold used (Figure 7).

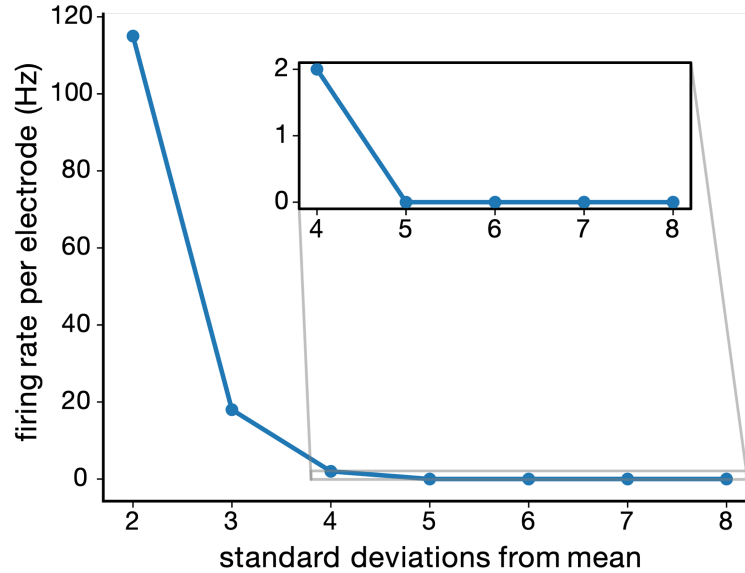


Figure 7: The firing rate per electrode depends on the threshold selected

An example plot depicting the overall firing rate (Hz) obtained per electrode as a function of the threshold selected (standard deviation from mean) for spike detection. This is for a single recording with 4096 recording channels.

Here, the optimal threshold for detection was selected for each recording by fitting the voltage histogram of each channel with a Gaussian distribution to separate spiking activity from normally distributed noise (Figure 8). Here, the signal starts to deviate from Gaussian noise around 3 SD. At 4.5 SD, there is minimal involvement of Gaussian noise, allowing the neuronal activity to be parsed out from the noise, making it an ideal threshold for spike detection.

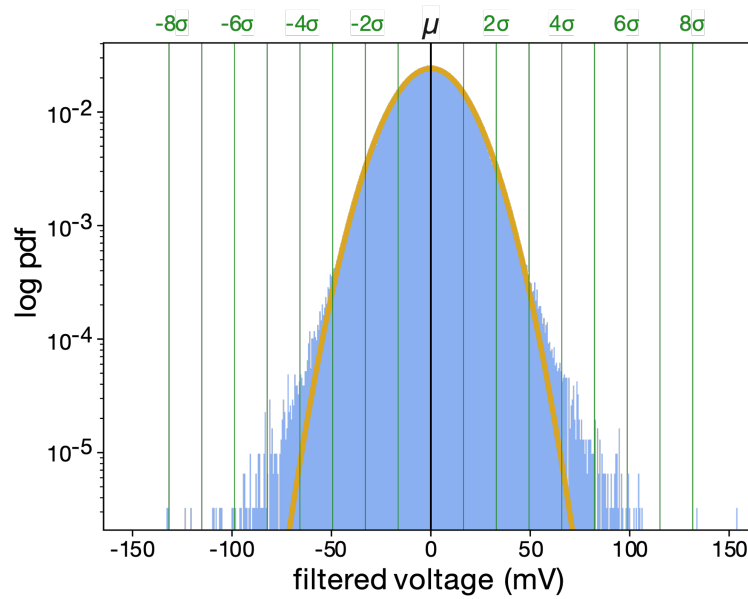


Figure 8: Optimal threshold selection allows to part cellular activity from normally distributed noise

An example plot depicting the filtered voltage of a recording on 30 seconds of activity. The voltage histogram is fitted with a Gaussian distribution (yellow) to separate spiking activity from normally distributed noise. The logarithmic probability density function (log pdf) (bin = 200) of the filtered voltage (mV) is shown with the number of standard deviations from the mean (green).

3.2.3 Single cell identification

In high density preparations, a single electrode can report the activity of multiple cells. To isolate single-cell activity, I employed a principal component analysis on the extracted spike waveforms from individual channels (Rossant et al., 2016), which revealed distinct clusters putatively corresponding to individual cells (Figure 9), indicating that more than one neuron likely contributes to the voltage recorded on each channel.

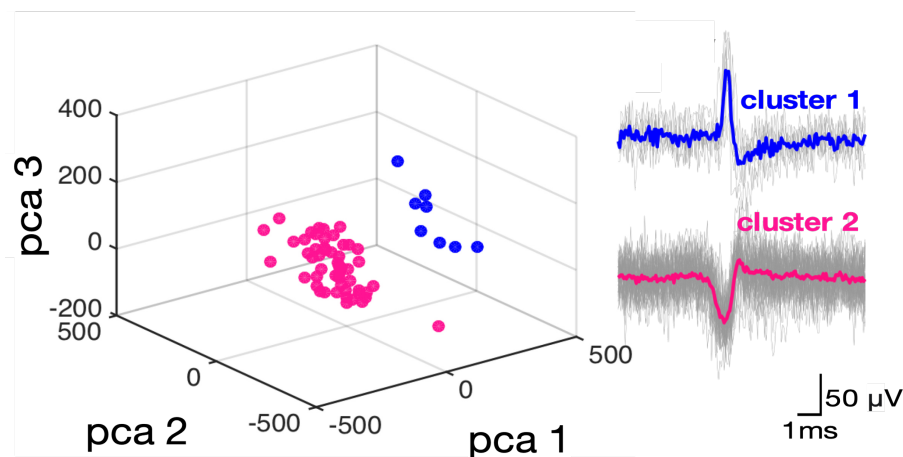


Figure 9: K-mean clustering of detected spikes on a single electrode reveals cluster with distinct average waveforms

An example showing each detected spike on a single electrode. Two distinct clusters (pink and blue) were detected on this recording channel. Note the average mean waveform revealing two distinct action potential waveform kinetics.

Next, clusters amongst the principal components of rrSVD were extracted using a k-means cluster analysis ($c=2$) (Xu and Wunsch, 2005). A linear discriminant analysis (LDA) was then applied to the labeled principal components (Hill et al., 2011) to determine how well the data were separated by this approach. LDA provides an estimate of linear classification error between the two clusters assuming equal class covariance. Ten-fold cross-validation was employed, where chance corresponds to an error estimate of 0.5 (Figure 10). Clusters with error lower than chance were merged by applying the same label to all waveforms. Each cluster was considered as a putative neuron.

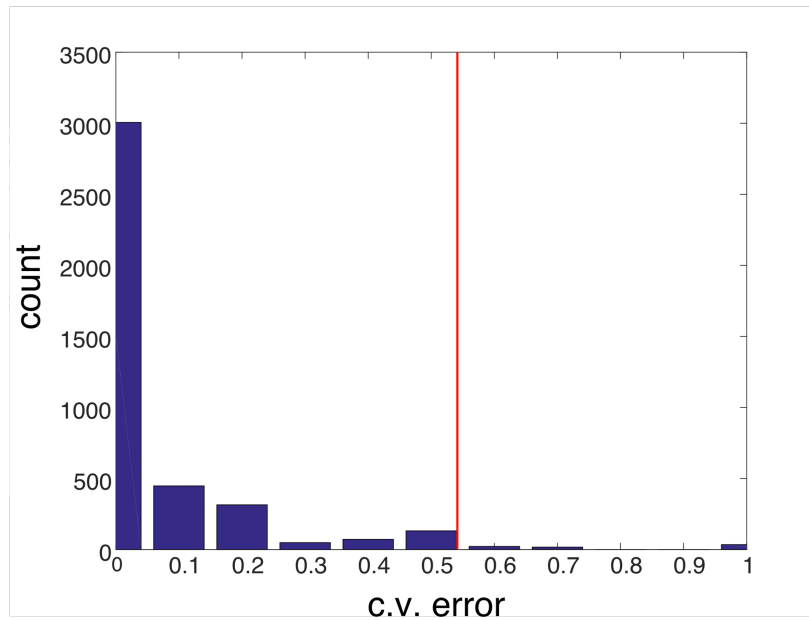


Figure 10: Ten-fold cross-validation shows better than chance performance

An example for a single recording (4096 recording electrodes) where ten-fold cross-validation is applied. Note that most clusters are well separate and few display high values of CV error.

To ensure the activity of each single cell was only reported once, I set a radius of 250 μm around the electrode where a given cell was detected. This radius corresponds to the approximate spread of extracellular activity around the soma of individual neurons (Jia et al., 2019). Within this radius, the channel with the largest negative deflection in mean waveform was identified. Previous work observed that the largest negative amplitude of extracellular signal originates from the axon initial segment (AIS) (Bakkum et al., 2019; Obien et al., 2015). Utilizing spatial localization of the AIS, the channel showing the greatest negative deflection was the closest to the putative soma of a given cell (Jia et al., 2019). Activity at this given channel was consistent with the somatic activity, showing a decay in peak voltage as we moved away from the soma (Figure 11; Ruz et Schultz, 2014; Pettersen and Einevoll, 2008; Linden et al., 2011; Mechler et al., 2011).

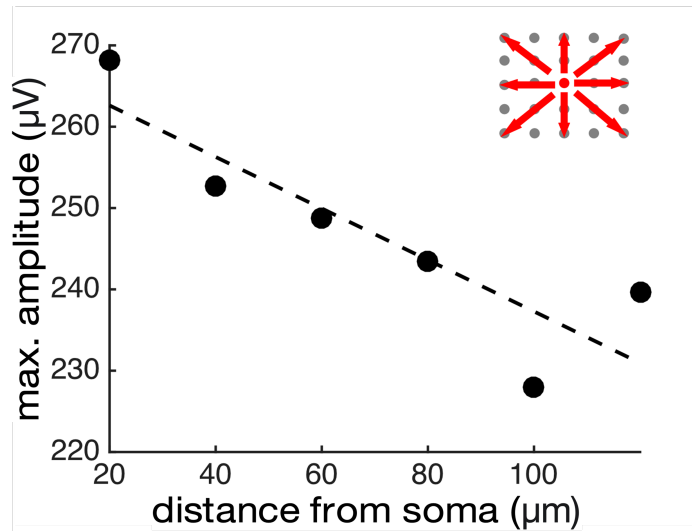


Figure 11: Voltage amplitude decays as a function of distance from the electrode where the highest peak in activity was detected

An example of one recording showing the average spatial decay in the maximum amplitude (μV) of the negative peaks of action potentials from all putative cells as a function of distance (μm) from the soma.

Visualization of adjacent electrodes revealed a complex structure of voltage activity, possibly due to activity propagation within the compartments of one neuron, or to the interactions between distinct neurons and their arborizations (Figure 12; 4 different neurons are shown) (Shein-Idelson et al., 2017; Obien et al., 2015).

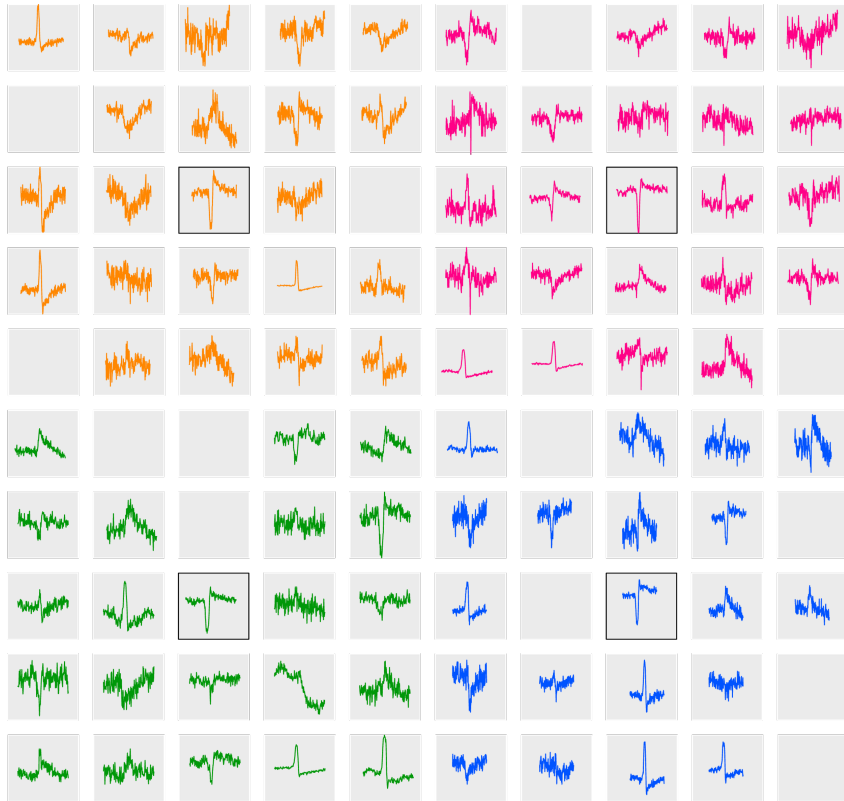


Figure 12: The current sink signal from a single neuron

An example showing 4 distinct putative neurons (soma location identified as the center square of each respective color) where the extracellular voltage associated with action potentials spans multiple electrodes and likely reflects the voltage propagation in the different compartment of the neuron.

3.2.4 Soma localization

In order to precisely estimate soma locations between each recording electrode, we used a center-of-mass analysis (Muthmann et al., 2015; Hilgen et al., 2017). For each putative soma identified previously, a 3-by-3 matrix with elements $a_{i,j}$ was constructed by extracting the waveforms on adjacent electrodes. Location (2,2) of this matrix was the amplitude of the negative peak in mean waveform of the somatic electrode. Surrounding locations were

populated with the amplitude of the negative peaks of all electrodes immediately adjacent to the somatic electrode.

The row of the center-of-mass is given by

$$r = \frac{\sum_i \sum_j (i \cdot a_{ij})}{\sum_{ij} a_{ij}} \quad (2)$$

and the column is

$$c = \frac{\sum_i \sum_j (j \cdot a_{ij})}{\sum_{ij} a_{ij}} \quad (3)$$

This revealed subtle changes in position, reflected by distinct weights on each nearby electrodes (Figure 13; mean indicated in red, $n = 155$ spikes).

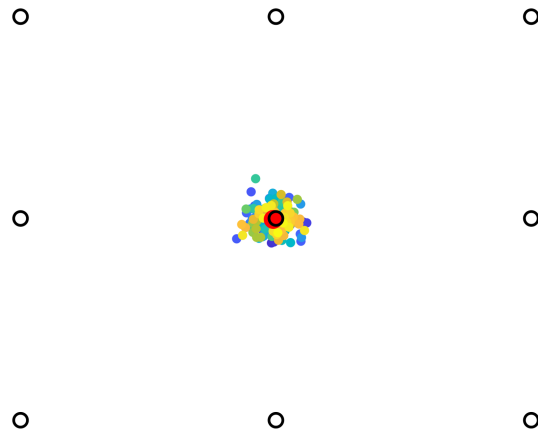


Figure 13: The center of mass location of all spikes for a single soma captured by one recording electrode.

An example showing each spike (different colored dots) for a single electrode for the duration of an entire recording. The average (in red) represents the mean location of all spikes.

In summary, this Chapter presents the main steps of spike sorting and implements them in an open-source, scalable data-analysis pipeline composed of a series of Python and Matlab code I implemented to automatically perform spike sorting on large datasets. I demonstrated this method on data I acquired from the prefrontal cortex of mice where first, individual spike waveforms are fitted by spline interpolation to extract action potential kinetics (half-amplitude and peak-to-peak durations). These values are analyzed using a principal component analysis with k-means clustering to identify uncorrelated signals from single channels. Optimal separability of clusters is assessed with a linear discriminant analysis. Lastly, each channel's source location is identified using spatiotemporal characteristics of spike waveforms across the array.

Chapter 3: Cell type classification through voltage waveform analysis

A second step in addressing the need for automatized spike sorting of distinct cell types presented above (Chapter 1, Section 6.1) lies in efficiently classifying distinct cell types. In the first part of this chapter, I present a brief overview of the electrophysiological properties of distinct cell types in cortex (Section 1). Next, I discuss the spike sorting of distinct cell types (Section 2). Lastly, building on the previous chapter's spike sorting protocol, I present an analysis method utilizing waveform properties in large network (Section 3).

The ability to classify cell types in distinct electrophysiological categories is fundamental in understanding network properties, offering insights on the role each unit plays in complex computations. In the mammalian brain, several classes of electrophysiological cells exist (see Chapter 1). The biophysical properties of individual cells dictate their firing properties, playing an important role in the transformation of synaptic input into spike output (Contreras, 2004). The last decades of work have seen an increased interest in the roles played by distinct cell types, but much remains to be uncovered to understand how they affect computations in large networks. In this chapter, we will look at the different cell types present in cortical networks. More specifically, we will appreciate how their identity translates into distinct electrophysiological features, focusing on two main categories of neurons: regular spiking excitatory neurons and fast spiking inhibitory neurons.

1. Electrophysiological properties of distinct cell types in cortex

Initial work characterizing the electrophysiological properties of cortical cells can be traced back to Connors, Gutnick and Prince (1982) and McCormick, Connors, Lighthall and Prince (1985), but other articles in prior years since the start of intracellular recordings had looked at cortical cells in vivo (Albe-Fessard & Buser, 1953; Calvin & Sypert, 1976; Li, 1959; Phillips, 1956; Takahashi, 1965; Tasaki, Polley, & Orrego, 1954).

Based on their firing patterns in response to stimulation, cortical cells can be divided into five electrophysiological categories: regular spiking (RS), intrinsically bursting (IB), fast spiking (FS), fast repetitive bursting (FRB) and low threshold spikes (LTS) (Contreras, 2004).

Regular spiking excitatory cells

RS cells are the most prevalent cell type in cortex, accounting for about 80% of all neurons (Anastasiades & Carter, 2021). They were named “regular spiking” by Mountcastle et al. (1969), because of the regular trains of action potentials they generated in monkey cortical neurons. Intracellularly, they fire a single spike at threshold and respond to current injection with sustained, repetitive firing (Connors et al., 1982; McCormick et al., 1985). Their spiking frequency as a function of injected current (the $f - I$ curve), is largely linear and they can fire at frequencies up to 200-300 Hz (Stafstrom et al., 1984). The action potential of RS cells typically lasts 1 ms, and is then followed by an afterhyperpolarization (AHp), or afterdepolarization (ADp). RS cells are almost always pyramidal neurons, releasing glutamate as a neurotransmitter, making them excitatory (Amitai & Connors, 1995). They have a uniform morphology, with a triangular cell-body measuring around 20-70 μm in diameter. They have an apical dendrite pointing in the direction of layer 1 with ramification along the way, as well as a complex set of basal dendrites distributed horizontally in the layer of the cell body.

Fast spiking inhibitory cells

FS cells are another type of cell that was described by Mountcastle et al. in 1969. They were named “fast spiking” because of their thin (short in duration) action potential. These cells were also detected during extracellular recordings performed by Simons (1978) in the barrel cortex of rats. Intracellularly, these cells stand out due to the rapid repolarization of their spikes and their short AHP (McCormick et al., 1985). As such, FS cells have a very short action potential duration of around 0.4 to 0.6 ms. In response to current injections, these cells can generate regular trains of action potentials at high frequencies (up to about 800 Hz), with almost no accommodation. Their fast spiking behaviour means they can fire with little decrease in frequency over a prolonged period (Connors and Gutnick, 1990). Their spiking frequency as a function of injected current (the $f - I$ curve), is very linear. FS cells are almost exclusively non-pyramidal neurons releasing GABA, making them inhibitory interneurons. GABAergic interneurons can be further subdivided through immunohistochemistry by looking at their calcium-binding protein properties, revealing that FS cells are almost always parvalbumin (PV) positive (Kawaguchi, 1995; Kawaguchi & Kubota, 1998). Other types of GABAergic interneurons displaying either LTS or RS electrophysiological properties are negative for PV. Cortical interneurons also differ in their neuropeptides, with PV cells not expressing neuropeptides. Other types of GABAergic interneurons which are positive for neuropeptides include VIP, which usually express calretinin as their calcium-binding protein and SOM positive cells expressing calbindin. Despite this variety, parvalbumin cells represent more than 50% of cortical interneurons.

In summary, the waveform of action potentials differs substantially amongst various cell types in the brain. In the cortex, GABA interneurons generally have narrower (faster) action potential than glutamatergic pyramidal neurons. This is most evident in intracellular recordings, as the shape of the action potential can be resolved more clearly (Nowak et al., 2003;

McCormick et al. 1985), however, the difference in action potential width is also present during in vivo extracellular recordings (Mountcastle et al. 1969). Classification based on action potential waveform is useful and informative, however there has been discrepancies in the literature, including work reporting fast-spiking excitatory neurons (Vigneswaran et al., 2011), highlighting the need for thorough classification and biological validation. But why is there such differences in the action potential of distinct cell types?

Action potentials in cell bodies encode information in their frequency and pattern. This behavior is made possible by the expression of a wide array of voltage-dependent ion channels. The variety of channels present in neurons has been appreciated since the 1970s, and subsequent interest has provided insights on the voltage conductance expressed by neurons in the brain. These conductances include several components of sodium currents, voltage-dependent calcium currents, voltage-activated potassium currents, calcium-activated potassium currents, hyperpolarization-activated current (I_h), and others (Bean, 2007). Their combined and coordinated action leads to a complexity that allows distinct cell types to have slightly different action potentials. Below, we will look at the shape distinction in the firing of action potentials between various types of neurons.

2. Spike sorting of distinct cell types

The use of action potential properties to distinguish distinct cell types has been documented in previous studies. As such, it has been proposed that excitatory and inhibitory neurons in particular can be distinguished based on the width of their action potential waveforms (Hafizi et al., 2022; Ren et al., 2020; Bathó et al., 2004). Narrow and broad action potentials have been distinguished from one another in several neocortical areas, including prefrontal cortex (Mitchell et al., 2007; Wilson et al., 1994; Constantinidis and Goldman-Rakic,

2002; Gonzalez-Burgos et al., 2005; Hasenstaub et al., 2005, Bartho et al., 2004, Tamura et al., 2004).

To do this, methodologies typically extract the temporal features of action potential waveforms. Two measures are commonly used, the peak-to-peak (PP) distance, and the half-amplitude width (HAW) distance. The peak-to-peak distance refers to the distance, in units of time (milliseconds), between the point where the voltage is at its most negative value (peak 1) and the subsequent point where the voltage is at its most positive value (peak 2) (Figure 15). The half-amplitude width is the distance, also measured in milliseconds, of the first negative voltage peak measured at half the maximum amplitude (width of the spike) (Figure 15).

The rationale for using these features as a proxy for cell type identity stems from the electrophysiological properties mentioned above, highlighted in intracellular recording studies, which report that inhibitory cells, such as parvalbumin-expressing GABAergic interneurons, have short-duration action potentials (Nowak et al., 2003; Peyrache and Destexhe, 2019; Kawaguchi, 1993; Connors and Gutnick, 1990), whereas excitatory pyramidal neurons have longer-duration action potentials (McCormick et al., 1985). Thus, pyramidal excitatory neurons will usually have longer time kinetics on both the peak-to-peak distance and half-amplitude width.

These differences have been hypothesized to result from the fact that repolarization of the membrane potential following a spike is slower among broad-spiking pyramidal neurons (McCormick et al., 1985; Nowak et al., 2003; Hasenstaub et al., 2005). In the extracellular signal, which resembles the negative derivative of the intracellular membrane potential, this gives both a broader and shallower peak following the initial trough (Henze et al., 2000).

3. A method utilizing waveform properties in large network

Despite technological advances, the ability to identify distinct cell types based solely on the extracellular voltage has remained challenging. Current data processing methods often rely on waveform kinetics (Bathó et al., 2004) to return a classification of contributing excitatory and inhibitory units. As technological progress continuously allows for denser arrays, waveform spike analyses are increasingly sensitive to noise generated by the signal acquisition of proximal units.

Here, as the first step in characterizing cell type identity from extracellular recordings from large networks, I first implemented a classification method using current knowledge about cell type-specific waveform kinetics to distinguish putative excitatory and inhibitory neurons in large-scale networks. This was done using the same data as in Chapter 2 (See Chapter 2: Methodology).

3.1 Time resolution

Action potentials happen over an exceedingly fast timescale, with the average action potential only lasting for about 1 ms. This poses a great challenge for the extraction of precise time features of an action potential, despite the technological advances allowing the sampling at higher rates. To provide a more precise description of the shape of each spike, I first fitted each waveform using piecewise cubic spline interpolation (Matlab R2015b), a method allowing for the inference of new points within the boundaries of a set of known points. Here, using query points at 90 kHz I computed the mean fitted waveform (Blanche, 2006) for each putative single cell. For an AP lasting roughly 21 ms, this means that the initial signal acquired at 18 kHz (one data point every 0.05 ms) contained 20 data points in a 2 ms period. At 90 kHz (one data point every 0.01 ms), the same 1 ms AP would contain 100 data points. Here, samples are

added whilst maintaining the length in respect to time, preserving the frequency content. No content has been added to the signal so there would be no aliasing issues (Figure 14).

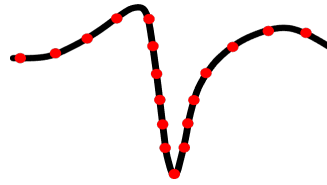


Figure 14: Spline interpolation

Graphical depiction of an action potential with extrapolated data points

3.2 Feature extraction

Next, I extracted the two time constants mentioned above (PP and HAW) from the mean spike waveforms at somatic locations (Figure 15). (Sirota et al., 2008; Sakata & Harris, 2009, Barthó et al., 2004; Insel & Barnes, 2015).

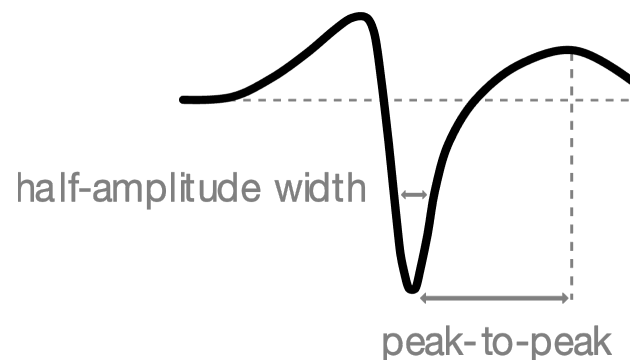


Figure 15: Action potential waveform features

Graphical depiction showing the two metrics extracted to classify excitatory and inhibitory neurons: The half-amplitude width and the peak to peak distance. The half-amplitude width represents the distance (in ms) in the middle of the negative portion of the trough of an action potential. The peak to peak distance represents the distance (in ms) between the largest negative point and the highest subsequent voltage value of the action potential.

Some work has cast doubt on the accuracy of spike-width based classification (Moore and Wehr, 2013). Even though narrow spikes are typical of PV+ inhibitory neurons, they can be prone to misclassification errors for two main reasons. First, not all PV+ neurons display narrow spikes. Like so, not all narrow spikes cells are from PV+ neurons (Gray and McCormick, 1996; Cauli et al., 2000; Markram et al., 2004). Secondly, the reliability of their spiking properties can be affected by the relative position from the recording electrode (Henze et al., 2000; Gonzalez-Burgos et al., 2005; Gold et al., 2006).

Despite these challenges, several studies have shown good classification accuracy using waveform properties, but these largely rely on the use of intracellular recording (Nowak et al., 2003; Kawaguchi and Kubota, 1997; Barthó et al., 2004; Madisen et al., 2012; Stark et al., 2013; Cardin et al., 2009; Cohen and Mizrahi, 2015).

3.3 Identification of putative excitatory and inhibitory cells using k-mean clustering

To extract groups of data points with a high level of similarity in their action potential waveform kinetics (here, hypothesized to be due to cell type identity), k-means clustering was performed on values of FW and PP obtained across all simultaneously recorded somatic electrodes. k-means clustering is a partitioning method used to partition data into k clusters, which are mutually exclusive. To do so, each data point is considered as an object that has a location in space. k-mean clustering finds a partition in the data where objects in each cluster are as close as possible to one another, but also as far as possible from objects in other clusters (MathWorks, 2023). Like most clustering methods, k-means requires the user to specify the number of clusters before clustering. To allow the method to remain unsupervised and unbiased, the optimal number of clusters was chosen with a Calinski-Harabasz criterion

(Calinski and Harabasz, 1974; Vendramin et al., 2009). The Calinski-Harabasz criterion is also referred to as the variance ratio criterion (VRC).

The Calinski-Harabasz index is defined as:

$$VRC_k = 1 + \frac{SS_B}{SS_W} \times \frac{N - k}{k - 1} \quad (4)$$

where SS_B is the overall between-cluster variance, SS_W is the overall within-cluster variance, k is the number of clusters, and N is the number of observations.

The overall between-cluster variance SS_B is defined as

$$SS_B = \sum_{i=1}^k n_i \| m_i - m \|^2 \quad (5)$$

where k is the number of clusters, n_i is the number of observations in the cluster i , m_i is the centroid of cluster i , m is the overall mean of the sample data, and $\|m_i - m\|$ is the Euclidean distance (L^2 norm) between the two vectors.

The overall within-cluster variance SS_W is defined as

$$SS_W = \sum_{i=1}^k \sum_{x \in c_i} \| x - m_i \|^2 \quad (6)$$

where k is the number of clusters, x is a data point, c_i is the i th cluster, m_i is the centroid of cluster i , and $\|x - m_i\|$ is the Euclidean distance (L^2 norm) between the two vectors.

Clusters that are well-defined tend to have a large between-cluster variance (SS_B) and a small within-cluster variance (SS_W). As such, the larger the VRC_k ratio, the better is the data partition. To determine the optimal number of clusters, this criterion maximize VRC_k with respect to k . Therefore, the optimal number of clusters corresponds to the solution with the highest Calinski-Harabasz index (VRC_k) value (MathWorks, 2023; Calinski and Harabasz, 1974).

Clusters corresponding to putative excitatory (E) and inhibitory (I) neurons were identified based on HAW and PP values (Figure 16).

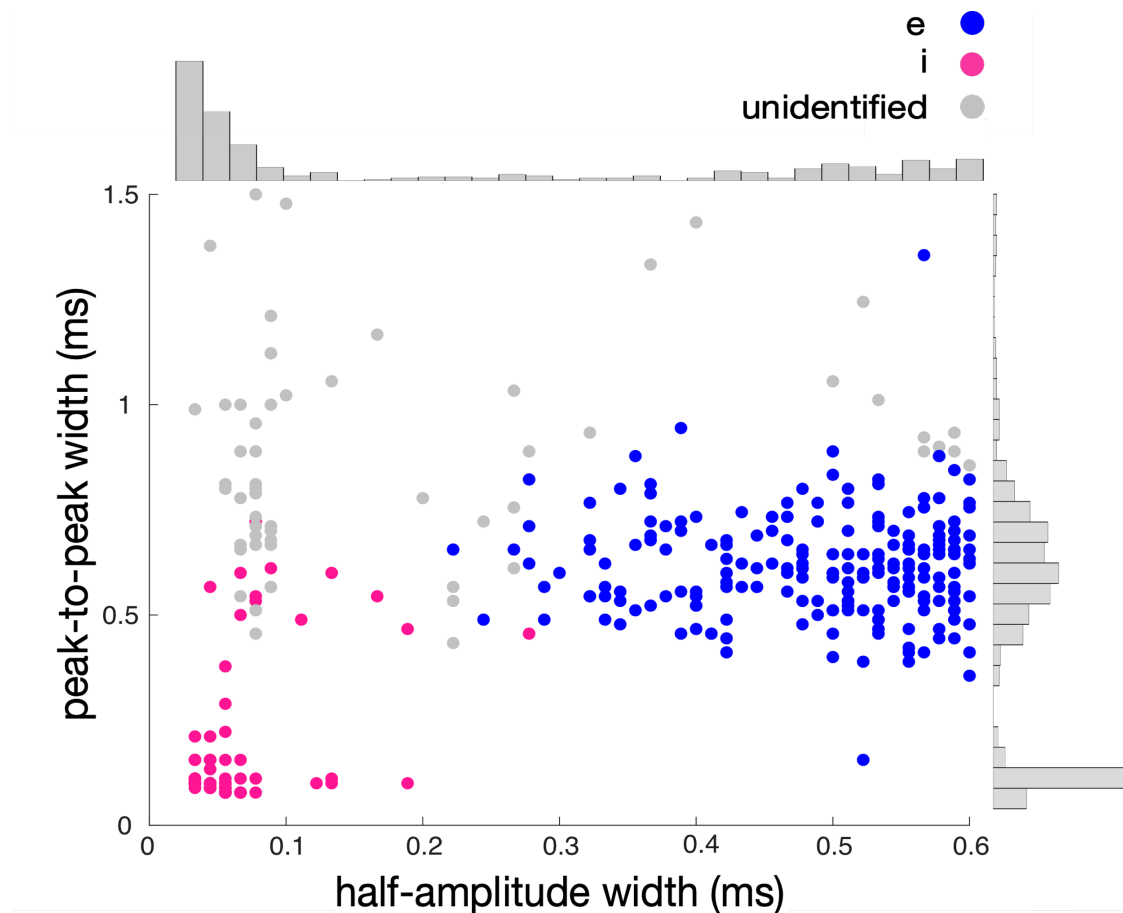


Figure 16: Distinct waveform kinetics for E and I cells

An example showing the clustering of all putative neurons based on their respective mean value for both the half-amplitude width and the peak-to-peak distance (ms). Inhibitory neurons (pink) are identified as a single cluster with small values on both metrics reflective of the fast

time kinetics expressed by fast-spiking inhibitory interneurons. This contrasts with neurons identified as excitatory (blue) neurons which display higher values, in particular for the half-amplitude width, exhibiting longer time kinetics.

I cells were identified as a single cluster with rapid time constants for both the HAW and PP (Barthó et al., 2004). I extracted the putative soma locations of excitatory and inhibitory neurons on the MEA (Figure 17), and visualized firing activity in a raster plot (Figure 18).

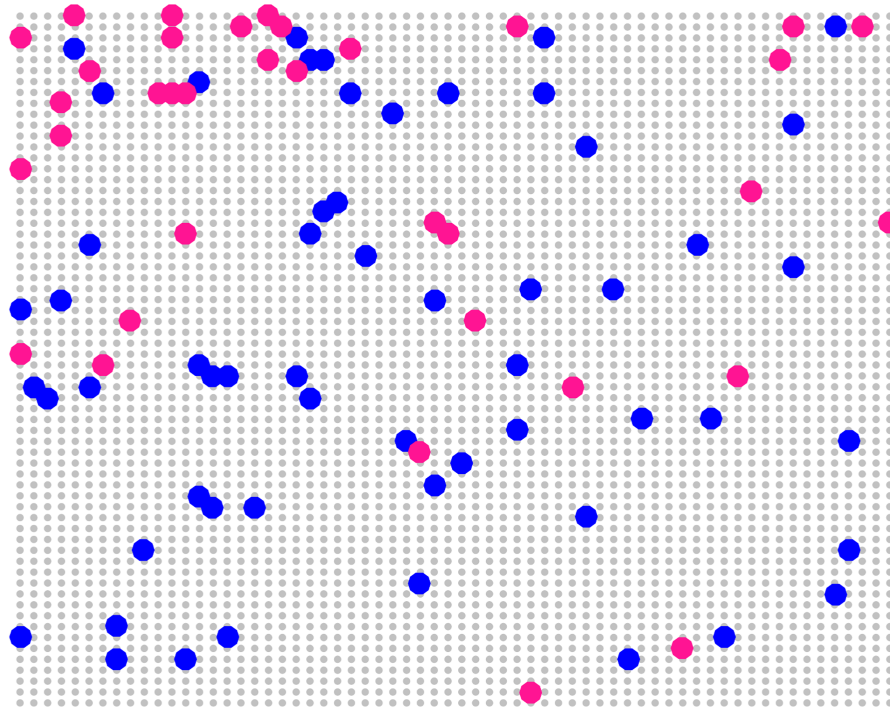


Figure 17: Respective cell location

The location of each putative inhibitory and excitatory neuron is superimposed on the multi-electrode array for spatial visualization.

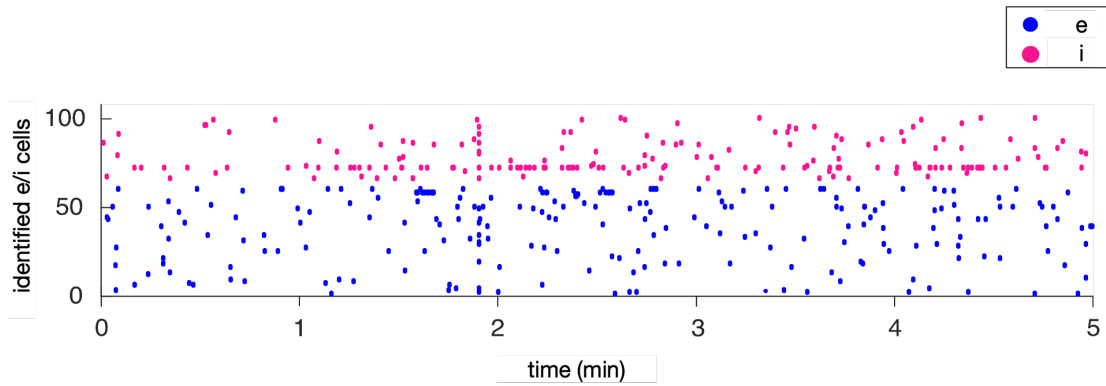


Figure 18: Raster plot of E and I activity

An example raster plot showing the spiking activity of classified excitatory and inhibitory neurons during a 5 min recording.

Other properties of putative I cells were consistent with past work, including a higher Fano factor (Figure 19; Becchetti et al., 2012) and a rapid spatial decay of pairwise correlation (Figure 20; Peyrache et al., 2012). The Fano factor, which is the variance-to-mean ration of spike counts in a given time window is often used to assess the variability of neuronal spike trains (Rajdl et al. 2020).

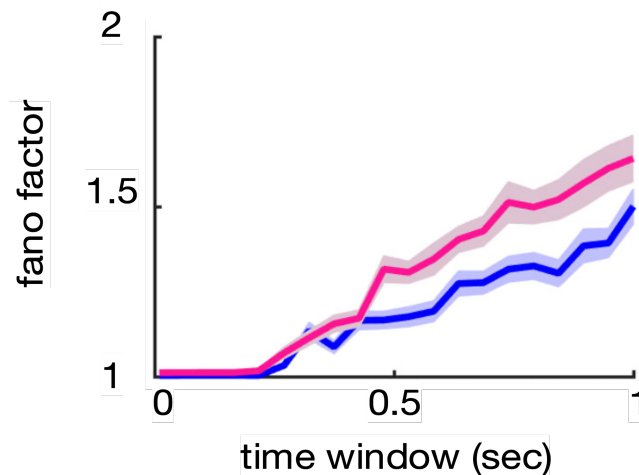


Figure 19: Fano factor for putative E and I cells

An example showing the respective fano factor of excitatory (blue) and inhibitory (pink) putative neurons for a single recording. Note that the fano factor is higher for putative inhibitory neurons

indicating a higher variability of neuronal spike trains. The shaded area represents the standard error of the mean (SEM).

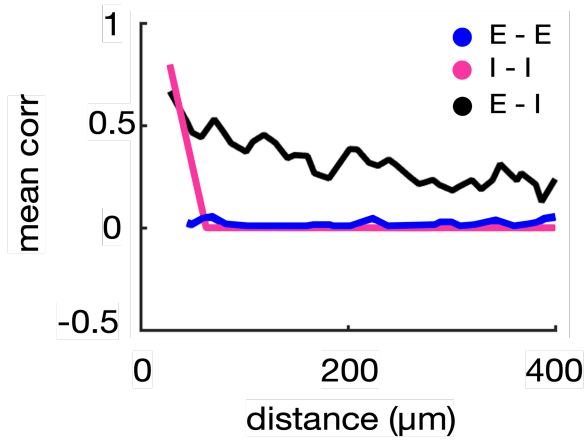


Figure 20: Mean correlation as a function of spatial distance

An example showing the respective mean correlation, as a function of distance, of excitatory (blue) and inhibitory (pink) putative neurons for a single recording. Note the rapid spatial decay of pairwise correlation of inhibitory-to-inhibitory (I-I) in contrast with excitatory-to-excitatory (E-E) and excitatory-to-inhibitory (E-I).

These results extend previous findings to large-scale recordings, yet a ground-truth validation of cell types is still required to accept this method as a good proxy of biological identity, as examined next.

Chapter 4: A step-function opsin protocol for cell classification

1. Probing neuronal activity on MEA platforms

Multielectrode arrays have allowed the simultaneous recording of neuronal activity using thousands of densely packed electrodes. The ability to sample from such a high number of distinct recording sites has been made possible by the use of CMOS circuits. Using this technology has allowed researchers to sample at high frequency from a constantly increasing number of close-by electrodes. This has permitted the recording of neuronal activity in large biophysical circuits with an unprecedented spatial resolution, capturing the activity of a single neuron's compartments on several nearby electrodes. This technological advance has provided researchers with the ability to monitor spontaneous activity happening in neuronal networks.

Traditionally, stimulation of neurons was achieved electrically. Some MEA manufacturers (3Brain) have designed MEA platforms which allow the electrical stimulation of recorded neurons using stimulating electrodes located on the MEA platform. Additionally, certain experimental designs allow for the use of an external stimulating electrode to probe parts of the network.

Optogenetics

Modern neuroscience has evolved in part through the ability to manipulate neural activity using optogenetics. Optogenetics was first introduced by Deisseroth et al. in 2006, referring to the ability to utilize a carefully designed combination of genetic engineering and optical technology to both control, and monitor biological cell activity. Using photosensitive proteins, cellular biological function can be manipulated, resulting in the modulation of the electrical excitability of the targeted population of genetically defined neurons (Boyden et al. 2005).

Voltage manipulation of individual neurons using light is made possible by Channelrhodopsin (ChR), first discovered in algae (Joshi et al., 2020). Two forms of ChR are encoded by the green algae: ChR1 and ChR2, both expressing light mediated responses. ChR2 became of increasing interest due to its ability to form functional ChR2 in the mammalian system (Joshi et al. 2020). ChR2 displays fast onset kinetics (in the order of milliseconds), and has the ability to reliably produce action potentials from high frequency light pulses (milliseconds). This form of optogenetics operates on a fast timescale, and the associated change in voltage match the timescale of fast action potentials (milliseconds).

Stabilized step-function opsin

Another type of channelrhodopsin later developed (Berndt et al., 2009) termed “bi-stable” channelrhodopsin, or “stabilized step-function opsin”, operates slightly differently. Rather than causing a transient change in voltage, this channelrhodopsin converts a short pulse of light into a stable step in the membrane’s voltage potential. These photo currents can be triggered, and stopped using two different color wavelengths offering a voltage change persisting over a longer time scale. This form of optogenetics provides a substantial advantage when traditional ChR2 currents cannot be monitored adequately due to their fast time kinetics.

Probing activity on MEA platforms: shortfalls

The use of CMOS technology has limited researchers’ ability to use them in conjunction with optogenetics, one of the most widely used tool to manipulate neural activity. This is due to CMOS circuits being exceedingly sensitive to light (see Chapter 1). As such, the application of a brief (1 ms) pulse of blue light leads to a signal artefact captured as saturation of the entire array lasting around (5 ms) (Figure 21). This design property has restricted the ability to use them in conjunction with optogenetics, one of the most widely used tools to investigate

neuronal spiking properties. As such, researchers have not been able to selectively activate a genetically defined population with light, while simultaneously recording from the entire population. As the ability to rapidly perturb distinct populations in large network provides valuable insight as to how local circuits represent information (Lemon et al., 2021; Buzsáki, 2010), it is of broad interest to utilize methodologies that allow for optogenetic manipulation of defined cell populations in multielectrode arrays.

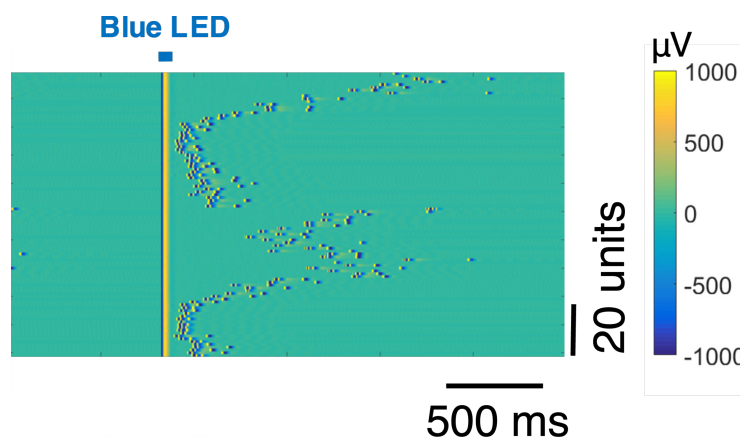


Figure 21: Light artefact

Recording from mouse prefrontal cortex showing that the application of a brief (1 ms) pulse of light leads to saturation lasting about 5 ms post-stimulation.

2. Cell type specific probing through optogenetics on MEA platform

To address these challenges, I propose and validate an experimental protocol using a stabilized step-function opsin (Bi-stable neural state switches; Berndt et al., 2009), which can be activated in a step-like fashion for prolonged periods (minutes) by brief photostimulation with blue light and can be inactivated by green-yellow light. This viral and optogenetic strategy, used in conjunction with CMOS-based multielectrode arrays, allows for the monitoring of induced spiking beyond the initial light-generated artefacts. I demonstrated this by inducing spiking selectively in targeted PV inhibitory neurons, as more than 90% of them have been

reported to be fast-spiking (Stringer et al., 2016), to verify whether cells previously identified as putative inhibitory interneurons indeed correspond to light-activated cells.

Methodology

For this, PV-CRE mice were again kept on a 12:12-h light/dark cycle, with access to food and water ad libitum. All experiments and procedures were performed in accordance with approved procedures and guidelines set forth by the University of Ottawa Animal Care and Veterinary Services (protocol # 3471).

We used a combination of cell type-specific viral strategies and optogenetics to address cell type distinction. We injected P21 PV-CRE mice (B6.Cg-Pvalb^{tm1.1(cre)}Aibs^{>/J}) with a stabilized step-function opsin (pAAV-Ef1a-DIO hChR2 (C128S/D156A)-EYFP) bilaterally in the medial prefrontal cortex (mPFC) (Berndt et al., 2009). This allows for selective, prolonged activation of parvalbumin (PV) interneurons with a 100 ms pulse of 465 nm blue light and deactivation with a 10 s pulse of 545-580 nm green-yellow light. For stereotaxic injections, mice were first injected with 0.05 mg/kg buprenorphine and anesthetized by inhalation of isoflurane. Injections were performed using a 10- μ L Hamilton syringe with a 33-gauge needle, and 1 μ L was injected per coordinate. Stereotaxic coordinates are as follows, from the bregma landmark: mPFC [anterior–posterior (APos), +2.4 $\rightarrow$ +2.6; medial–lateral (ML), $\pm$ 0.42; dorsal–ventral (DV), -1.5 $\rightarrow$ -1.8]. The DV coordinate was calculated from the surface of the pia. For ChR2 expression, 400 nL of pAAV-Ef1a-DIO hChR2 (C128S/D156A)-EYFP was injected, per site (bilaterally) and animals were left to recover for 2–3 weeks before ex vivo electrophysiological recordings.

mPFC slices were prepared from 35 to 42 day old PV-CRE mice, 2–3 weeks poststereotaxic injections. For slice preparation, mice were anesthetized by inhalation of

isoflurane (Baxter Corporation) and killed by decapitation. The brain was removed, and coronal brain slices were sectioned while immersed in an ice-cold choline chloride-based cutting solution of the following composition: 119 mM choline-Cl, 2.5 mM KCl, 1 mM CaCl_2 , 4.3 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM NaH_2PO_4 , 1.3 mM sodium L-ascorbate, 26.2 mM NaHCO_3 , and 11 mM glucose, and equilibrated with 95% O_2 and 5% CO_2 . Slices were recovered in a chamber containing standard artificial cerebrospinal fluid solution (aCSF) of the following composition: 119 mM NaCl, 2.5 mM CaCl_2 , 1.3 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 mM NaH_2PO_4 , 26.2 mM NaHCO_3 , and 11 mM glucose, at a temperature of 37 °C, continuously bubbled with a mixture of 95% O_2 and 5% CO_2 . Slices were recovered for 1 h in the recovery chamber and equilibrated to a temperature of ~25 °C before the recordings were performed. The tissue was continuously perfused with fresh aCSF.

PFC neurons were visualized using an upright microscope: BX61WU upright microscope (60X, 1.0 NA objective; Olympus, Melville, NY). PV neurons were visually identified by EYFP expression (Figure 22). We next performed whole-cell recordings from pyramidal neurons located near identified PV interneurons using in vitro slice preparations.

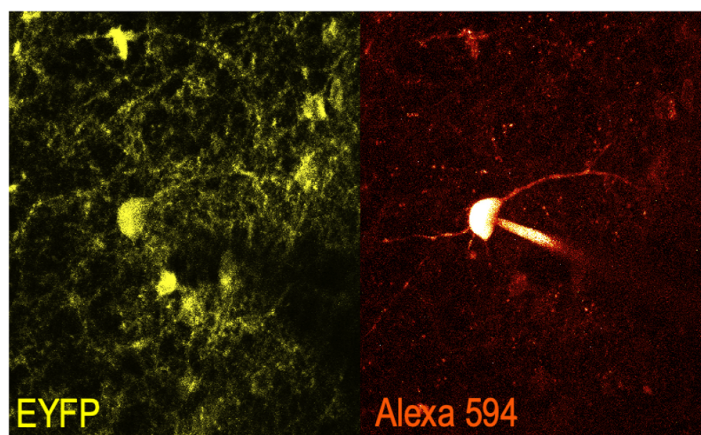


Figure 22: Visually identified EYFP expressing PV neuron

Whole-cell recordings were carried out using an Axon Multiclamp 700B amplifier, sampled at 10 kHz, digitized with an Axon Digidata 1440A digitizer, and filtered at 2 kHz. Whole-cell recordings were performed using borosilicate glass patch electrodes (3–6 M Ω ; World Precision Instruments) pulled on a Narishige PC-10 pipette puller. All experiments were performed at room temperature in Ringer's solution containing (in mM) 119 NaCl, 2.5 KCl, 1.3 MgSO₄, 2.5 CaCl₂, 1.0 NaH₂PO₄, 11 glucose, and 26.2 NaHCO₃ saturated with 95% O₂ and 5% CO₂ (pH 7.3, 295–310 mOsm/L). Synaptic currents were recorded using an intracellular solution of the following compositions (in mM): 115 K-gluconate, 20 KCl, 10 HEPES, 4 ATP-Mg, 0.5 GTP, 10 Na-phosphocreatine (pH = 7.2–7.3; 270–290 mOsm/L).

LED Photostimulation

LED photostimulation was delivered using a PlexBright LED module and controller (465 nm; Plexon). Light was delivered through a 200- μ m patch fiber cable with a bare fiber tip with ~1 cm of glass exposed. Photodeactivation was delivered using a yellow activation wideband 545–580 nm filter (BP545–580; Olympus).

Immunohistochemistry

Brain slices used for immunohistochemistry were prepared from 35 to 42 day old mice with 1 $\times$ PBS used as the solvent for all solutions. Mice were anesthetized by inhalation of isoflurane. They were then intracardially perfused with 5mL of PBS pH=7.4 followed by 10mL of 4% paraformaldehyde/ 0.1M PBS, pH 7.4 for 10 min. Postfixation in 4% PFA was carried out for 2 h before sequential cryoprotection steps in 15% and 30% sucrose over the course of 2 d. Brains were frozen in –30 °C isopentane, mounted in cryomatrix, and cut to 40- μ m slices on a cryostat (CM 3050S; Leica). Free-floating sections were blocked in 0.5% PBS-Triton-X-100 (TX-100) for 45 min, and incubated in primary anti-body against GFP (GFP-1020; 1:1000,

AvesLabs) in blocking solution at 20 °C. On the following morning, slices were washed three times, for 5 min each time, with 0.5% PBS-Triton-X-100 before incubation in the secondary antibody goat anti-chicken–conjugated AlexaFluor 488 (Jackson Labs, 703-546-155) for 3 h at room temperature. Slices were washed with 0.1% Tx-100 three times, for 5 min each time, before mounting onto slides with Vectashield (Vector Labs). Additionally, all slices used for recordings were saved, kept immersed in Formalin 10% in PBS pH=7.4, for immunohistochemistry to verify viral expression (Figure 23). The protocol is the same as above.

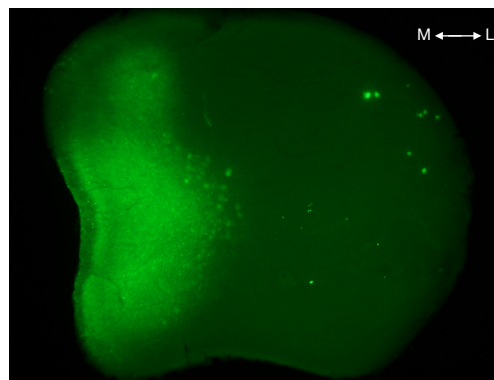


Figure 23: Viral injection expression

Imaging of a slice where PV inhibitory interneurons express EYFP. Note that the injection site targets the medial prefrontal cortex and that expressing interneurons are largely located in the infralimbic and prelimbic cortical subregions.

3. Patch-clamp validation of viral strategy

Photostimulation with blue light was sufficient to activate PV interneurons, which provided light-induced inhibitory currents to proximal pyramidal neurons that were consistent between trials (Figure 24). Conversely, pulses of yellow light reduced the frequency of inhibitory postsynaptic currents (IPSCs) onto recorded pyramidal neurons, indicating that the step-function opsin could be reliably activated and inactivated in a step-like manner by phasic delivery of light (Figure 24). Upon PV photoactivation, recorded pyramidal neurons reliably

received high-frequency barrages of 10 to 30 pA inhibitory currents for >20 seconds, followed by elevated rates of inhibitory input until deactivation with green-yellow light (Figure 24).

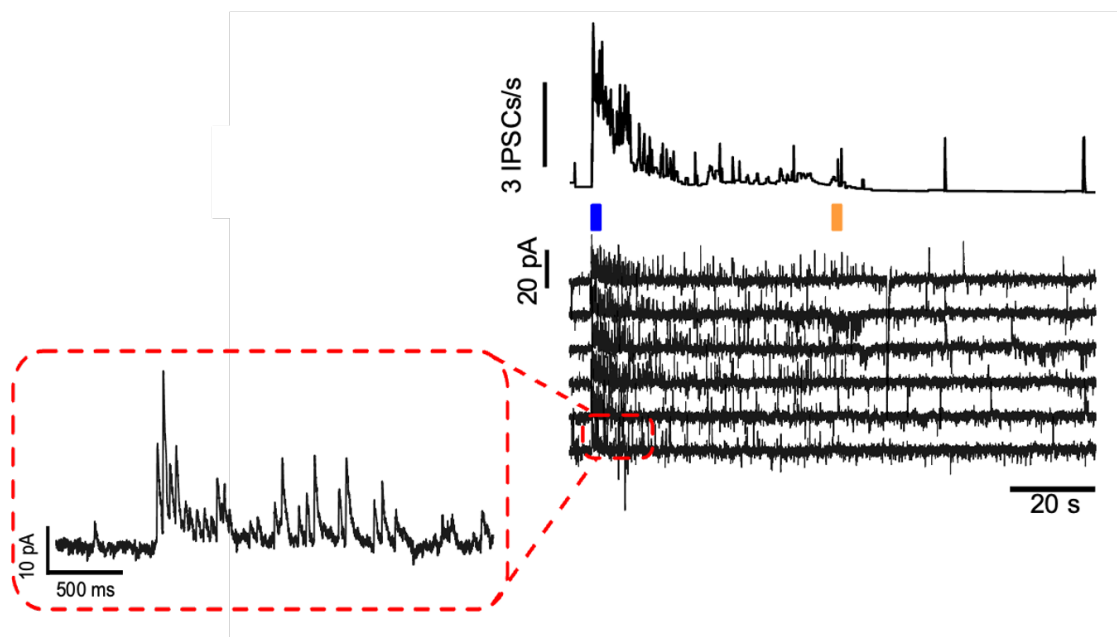


Figure 24: Nearby pyramidal neuron receiving inhibitory currents following PV opto-activation. Patch clamp recording figure indicates that PV interneurons provided light-induced inhibitory currents (left box) to proximal pyramidal neurons consistently between trials (traces on the right). Top trace indicates the number of inhibitory post-synaptic currents per second.

To characterize the refractory period of the step-function opsin, I delivered repeated photostimulation at various inter-trial intervals and measured steady-state activation (Figure 25). Longer inter-trial intervals (120 seconds) (Figure 25) allowed transfected cells to recover and prevented the reduction in steady-state activation observed with 20 and 60 second inter-trial intervals (Figure 25).

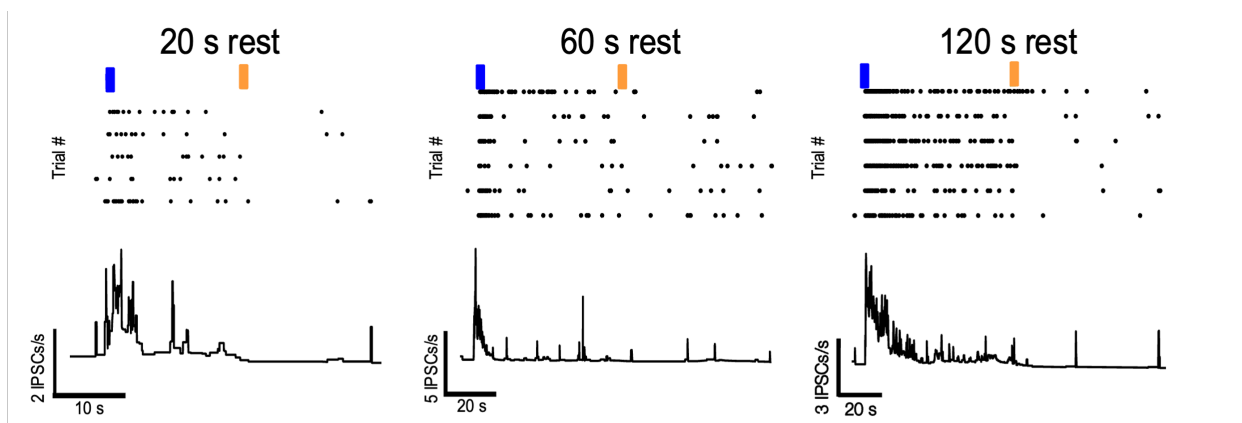


Figure 25: Inhibitory post-synaptic currents onto pyramidal cell depends on rest period

The duration, in seconds, of the rest intervals (of 20s, 60s and 120s, respectively) between trial indicates that to reach full activation potential a longer trial interval must be used.

Furthermore, the reduction in firing rate upon opsin deactivation was more prominent for longer interval trials (Figure 25). These results provided key insights to inform multielectrode array recording protocol, allowing us to selectively activate and inactivate PV interneurons in a temporally precise fashion using tools compatible with multielectrode array technology. The protocol designed is as follow (Figure 26):

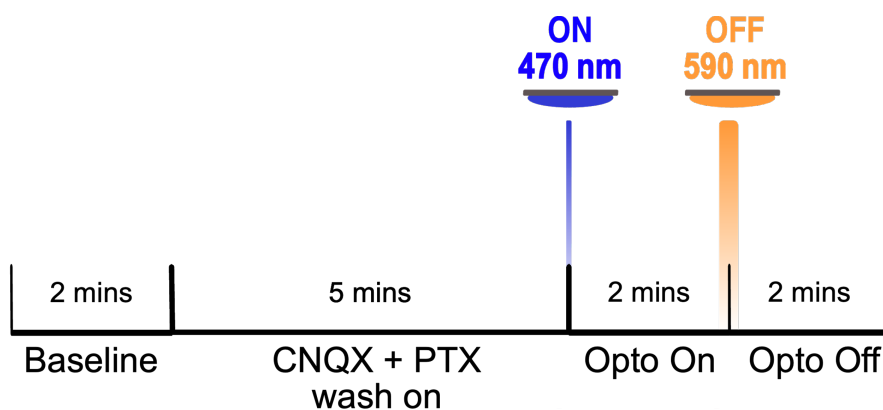


Figure 26: Optogenetic protocol

Protocol established

Upon assessing the opsin, the following protocol was established for subsequent multielectrode array recordings. The recordings were 24 minutes long and were as follow: baseline (2 min), opto-on (2 min), opto-off (2 min), cyanquixaline (CNQX) wash-on (5 min), CNQX + opto-on (2 min), CNQX + opto-off (2 min), picrotoxin (PTX) wash-on (5 min), CNQX and PTX + opto-on (2 min), CNQX and PTX + opto-off (2 min) (Figure 26). Additionally, the MEA-chip baseline (no slice) was recorded (30 secs) as control prior to adding each slice.

Chapter 5: Optogenetic validation of E/I classification on multielectrode arrays

1. Quantifying population-level activity changes to step function opsin activation of parvalbumin inhibitory interneurons

In this chapter, I aimed to validate the prior waveform-based cell type classification (Chapter 3), while conducting multielectrode array recordings at the population level, using an intersection of genetic tools and optogenetic constructs.

In mouse prefrontal cortex slices, transfected PV interneurons were selectively activated with the step-function opsin while performing multielectrode recordings. In baseline conditions, we observed low mean spontaneous firing rates ($\sim 0.5\text{Hz}$, Figure 27). However, this may be due to our stringent detection criterion. To prevent the propagation of optogenetic activation throughout the network, we applied 10 mM cyanquixaline (CNQX) and 100 mM picrotoxin (PTX) to block AMPA receptors and GABA A receptors, respectively. The application of this cocktail of drugs led to a slight reduction in firing rate (FR) of 0.0937Hz ($p < 0.001$, Figure 27). Photostimulation was delivered using brief 100 millisecond pulses of blue light, leading to a rapid increase in firing rate for cells classified as inhibitory of 0.3Hz . ($p < 0.001$) This change persisted for minutes after the stimulation, consistent with previous patch-clamp recording findings (Figure 24). Neurons classified as excitatory, on the other hand, showed little light-specific responses during the entire post-stimulation period with a reduction of 0.0125Hz ($p < 0.001$). Deactivation of photostimulation was delivered using a 10 second pulse of yellow (545-580 nm) light. This resulted in the firing rate of neurons identified as inhibitory returning to near baseline firing rates.

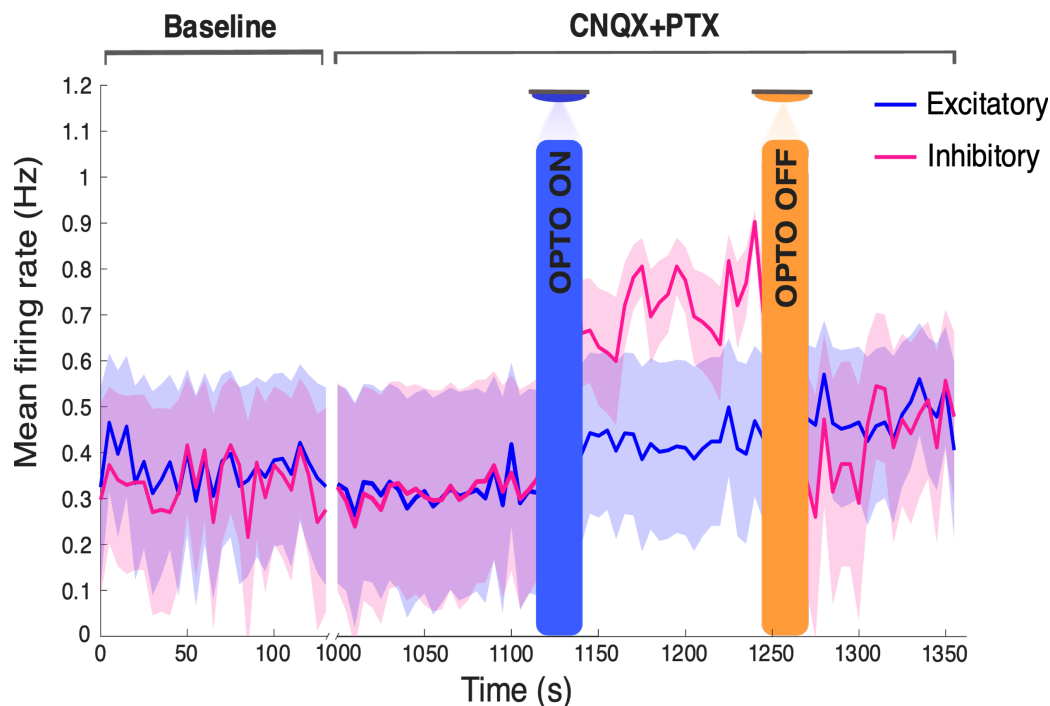


Figure 27: Average firing rate of putative excitatory and inhibitory neurons for a single slice

An example showing the mean firing rate of neurons ($n = 1170$) classified as excitatory (blue, $n = 1138$) and inhibitory (pink, $n = 32$) for a single recording (duration 22.5 min). To note is a time-lock increase in firing rate for neurons identified as inhibitory. The shaded area represents the SEM.

To quantify the increase in firing rate following optogenetic activation, we pooled neurons across 3 different recordings ($n = 8189$, 3 slices, from 3 mice) and post-stimulation responses of all identified neurons were examined (Figure 28). In the presence of synaptic blockers, optogenetic activation of targeted PV interneurons led to a significant increase in mean firing rate for the population of cells that had been classified as inhibitory based on their waveform properties ($n = 430$, $p < 0.003$). Neurons classified as excitatory neurons based on their waveform kinetics also showed a slight decrease in the mean firing rate ($n = 7759$, $p < 0.001$) (Figure 28). While our pharmacological manipulations attempted to largely eliminate network effects, this slight reduction in the firing rate of putative excitatory neurons may have reflected factors such as GABA B receptor-mediated local interactions.

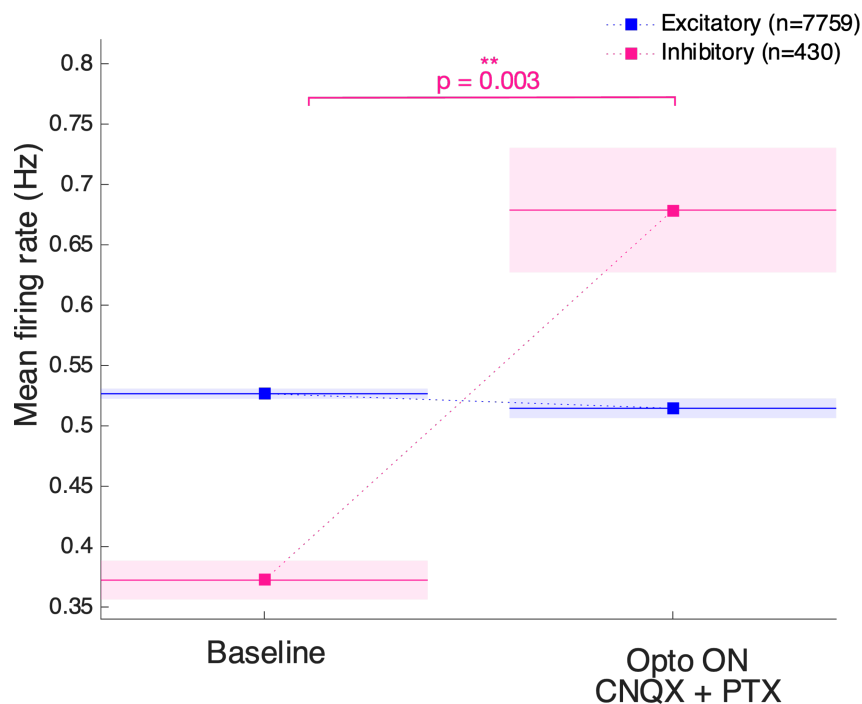


Figure 28: Quantification of change in firing rate for 3 slices

Quantification of the overall change in firing rate for 3 distinct slices (recording duration: 22.5 mins each). Here, 8189 individual neurons were identified, 7759 were classified as excitatory (pink), and 430 were classified as inhibitory. A significant two-fold increase in firing rate for neurons identified as inhibitory ($p < 0.003$) is observed. Shaded area represents the standard error of the mean (SEM).

2. A diversity of individual neuron responses to step-function opsin activation of parvalbumin inhibitory interneurons

Next, the response profiles of individual neurons to photostimulation were examined in further detail to validate our algorithm at the single cell level. To isolate the effect of photostimulation on the firing rate of each cell, we subtracted each cell's activity during the drugs epoch from light-on epoch activity. Individual neurons displayed a diverse range of responses to light stimulation, including neurons showing no change in firing rate, or delta firing

rate ($\Delta FR = 0$), neurons showing an increase in firing rate ($\Delta FR > 0$), and neurons showing a decrease in firing rate ($\Delta FR < 0$) (Figure 29). Fitting the data with Gaussian curves highlighted that a greater fraction of cells identified as inhibitory displayed an increase in ΔFR upon light application compared with putative excitatory neurons. For neurons identified as excitatory, the Gaussian curve remained centered near zero, indicating little to no light-induced changes on average.

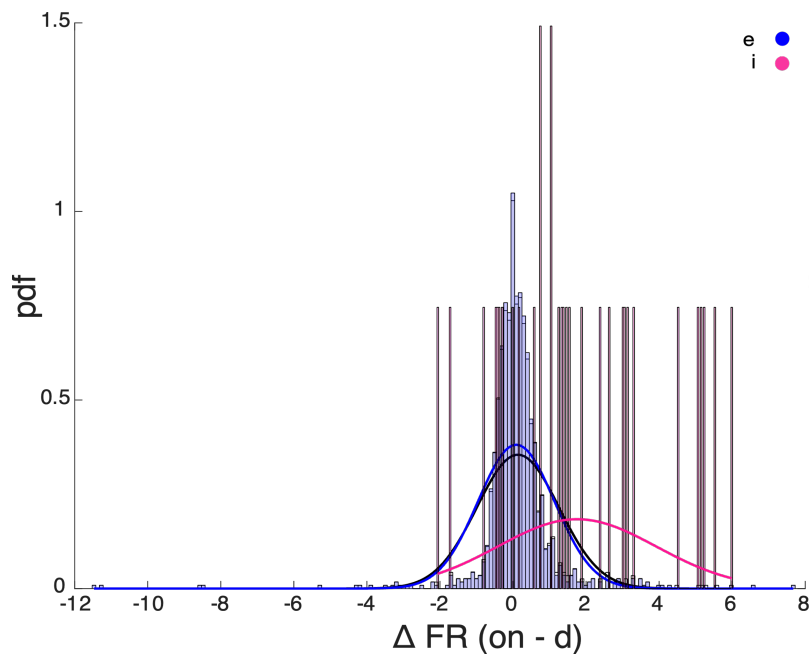


Figure 29: Delta change in firing rate (on – drugs)

Examination of the response profiles of individual neurons to photostimulation. To highlight cell type-specific differences, curves are plotted for each population (blue = excitatory, pink = inhibitory, black = average). The baseline activity of each cell (drugs epoch) is subtracted from light-on epoch activity. To note diverse range of responses to light stimulation, including neurons showing no change in firing rate ($\Delta FR = 0$), neurons showing an increase in firing rate ($\Delta FR > 0$), and neurons showing a decrease in firing rate ($\Delta FR < 0$).

Next, we computed a z-score for each neuron corresponding to photostimulation-evoked change in firing rate, expressed in units of SD from the baseline mean (Figure 30). A strict criterion was applied to separate neurons showing a statistically robust increase in firing rate (z score >3). A total of ($n= 7610$, 3 slices, from 3 mice) were classified in this fashion. Overall, 6819 (89.6 %) were reliably classified as excitatory and 791 (10.4 %) were reliably identified as inhibitory. Using this method, 56% of neurons that had been classified as inhibitory displayed a light-induced increase in firing rate. Conversely, only 8% of neurons which had been classified as excitatory showed an increase in firing rate after photostimulation (Figure 30). These could be misclassified excitatory neurons, or alternately, be due to network effects not blocked by the application of CNQX and PTX, like the inhibition of inhibitory neurons.

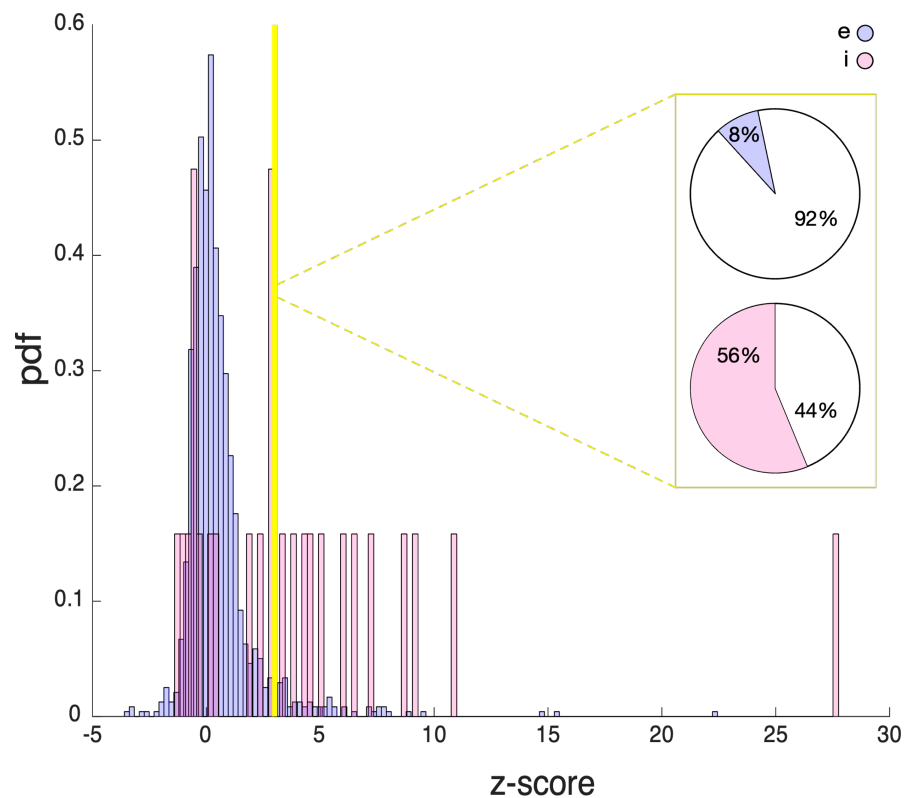


Figure 30: Opto-evoked change in firing rate

Computed z-score for each neuron, in units of SD from baseline mean. Delta change in firing rate is calculated as on, minus drugs (on – d). Excitatory neurons are shown in blue, and

inhibitory neurons in pink. Yellow line indicates criterion (z score >3) to determine a statistically robust increase in firing rate.

While the majority of I cells increased their firing rate in response to photostimulation, there are several potential reasons why not all I cells were responsive. First, the percentage of I cells responsive to light could be explained by our viral strategy targeting only a subset of PV-expressing interneurons largely located in the infralimbic and prelimbic cortical subregions (Figure 23). Thus, it is possible that more PV interneurons were recorded and classified as inhibitory but did not express the opsin and therefore did not respond to light stimulation. Additionally, non-PV inhibitory interneurons, and some excitatory neurons, could have been initially classified as inhibitory based on their waveform properties. These experiments, however, unambiguously identified a subset of genetically validated PV neurons: those that showed a transient photostimulation-induced increase in firing rate. Through a comparison with FWHM and PTP features, these results demonstrate that a step-function opsin approach is a viable and scalable method for validating spike waveform classification on multielectrode arrays.

3. Cell type-specific dynamical interactions in large-scale networks

Probing activity in intact networks

Next, building on our validated cell type-specific sorting, I asked whether network-wide interactions played a role in shaping population responses. For this, PV neurons were photostimulated without the presence of synaptic blockers, in physiological conditions, and the response of each cell type population was recorded on protracted timescales. Without drugs, network interactions between cells and between cell types remained intact, allowing them to dynamically modulate each other's activity.

Photostimulation was again delivered using brief 100 millisecond pulses of blue light (n = 3276, 1 slice, from 1 mouse) (Figure 31).

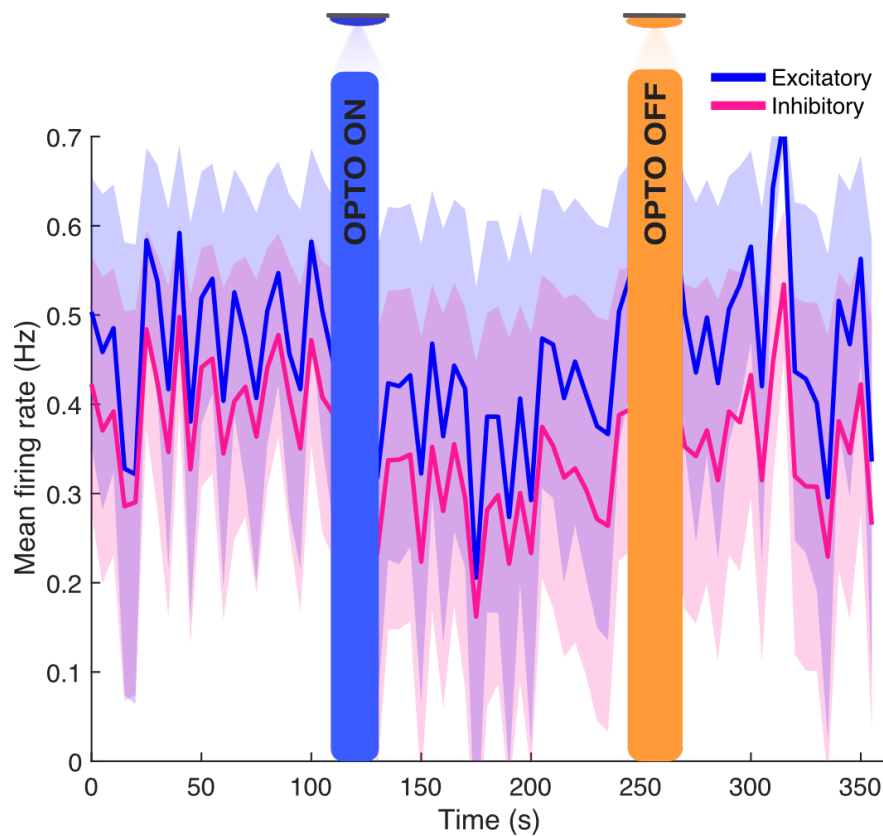


Figure 31: Average firing rate for putative excitatory and inhibitory neurons following opto-activation when no blockers are applied

An example showing the mean firing rate of neurons (n = 3276) classified as excitatory (blue, n = 2928) and inhibitory (pink, n = 348) for a single recording (duration 6 min). To note is a time-lock decrease in firing rate for neurons identified as excitatory, and for neurons identified as inhibitory. The shaded area represents the SEM.

Interestingly, this led to a significant decrease in overall firing rate for both putative excitatory neurons (from 0.425Hz to 0.3Hz, n = 2928, $p < 0.003$), and putative inhibitory neurons (from 0.5Hz to 0.375Hz, n = 348, $p < 0.003$) (Figure 32). More specifically, 65.5% (n = 1918) of neurons classified as excitatory, and 62.4% (n = 217) of neurons classified as inhibitory showed a decrease in firing rate after photostimulation (Figure 32).

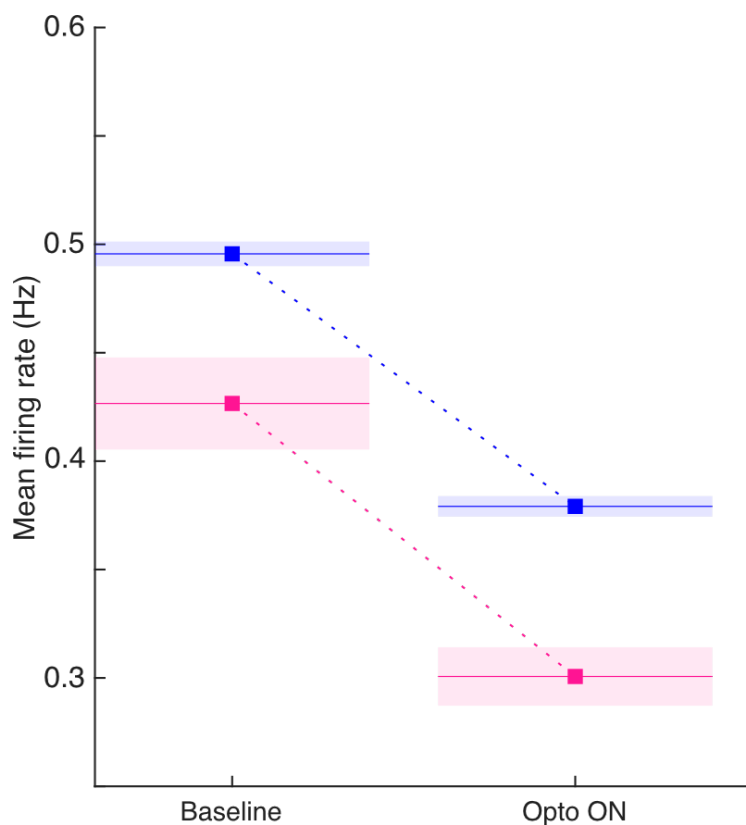


Figure 32: Change in firing rate following optogenetic activation

Quantification of the overall change in firing rate. To note is a decrease in firing rate for both excitatory and inhibitory neurons. Shaded area represents the standard error of the mean (SEM).

This counterintuitive suppression of inhibitory firing persisted past photostimulation (Figure 31), during the entire epoch that followed. Deactivation with a 10 second pulse of yellow (545-580 nm) light led to a return to near baseline values for both populations.

Although seemingly paradoxical, it has been reported that in networks with strong recurrent coupling, excitatory cells could be self-amplifying unless stabilized by inhibition (Sanzeni et al., 2020; Thivierge et al., 2022). A predominant feature of the cerebral cortex is its extensive

recurrent connectivity between nearby neurons. Both theoretical (Sanzeni et al., 2020; van Vreeswijk and Sompolinsky, 1996; van Vreeswijk and Sompolinsky, 1998), and experimental (Sanzeni et al., 2020) work has demonstrated that the computational properties of networks composed of excitatory and inhibitory neurons could be impacted by the strength of recurrent coupling. The functional role of dynamics relying on strong excitatory recurrent coupling has been linked faster response to external stimuli (Vreeswijk and Sompolinsky, 1996; van Vreeswijk and Sompolinsky, 1998), the ability to sustain persistent activity (Amit and Brunel, 1997), provide better memory capacity and storage (Rubin et al., 2017), allow networks to amplify particular input patterns (Murphy and Miller, 2009), and to generate complex activity patterns to brief inputs (Hennequin et al., 2018).

Such dynamics with strong recurrent coupling need to be stabilized by inhibition as to prevent unstable dynamics. This means that in conditions where recurrent connectivity is weak, the firing rates of excitatory neurons can remain stable independently of the activity of inhibitory neurons (Sanzeni et al., 2020). When recurrent connections are strong, however, excitatory neurons need the right amount of inhibitory input to remain in a stable regime due to self-amplifying excitatory-to-excitatory connections. In these conditions, neurons require inhibition to remain in a stable regime and are called inhibition-stabilized networks (ISNs) (Sanzeni et al. 2020) (Figure 33).

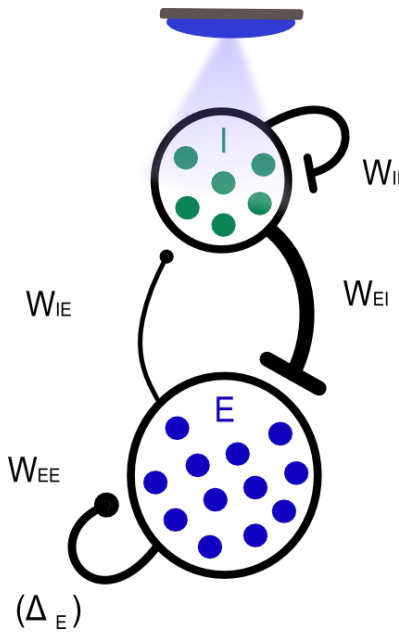


Figure 33: Inhibition-stabilized network architecture

Graphical depiction of a network composed of both excitatory and inhibitory subpopulations. Connection strength (weight) is represented as W , and is specific to each population (E or I) and their respective anatomical connections.

Despite the growing interest in favor of this network regime, it remains unclear whether cortical networks operate in an ISN regime, whether they always operate in such regimes (during spontaneous states or only in response to specific input), which cortical areas operate in this regime, and the conditions in which they might operate in this regime (Sanzeni et al., 2020).

As such, it is possible that in some instances, our recordings captured hallmarks of networks displaying inhibition stabilized properties. A defining feature of ISN activity is that inhibitory activation leads to a decrease in steady state inhibitory firing rates (Litwin-Kumar et al. 2016; Tsodyks et al., 1997) This paradoxical effect is not reported in non-ISN networks (although it can be observed in unstable networks (Mahrach et al. 2020)), where a perturbation applied to inhibitory neurons causes an increase in their firing rate.

Interestingly, in ISN networks stimulation of PV neurons leads to an increase in PV activity, but a decrease in overall inhibition (Thivierge et al., 2022). In some of our recordings (Figure 31 and 32), it is possible that a greater proportion of other types of inhibitory neurons (such as SOM, or VIP) are being tracked, thus resulting in a decrease in overall inhibition. We further explored the parameter space in which this is possible, by implementing a structured reservoir model incorporating a diversity of cell types and their known connection. The manuscript for this work can be found in Appendix B. Future work using this model could generate predictions which could then be tested experimentally using the method introduced in this thesis. Used in conjunction, these methods would provide valuable insights on how distinct types of neurons contribute to network dynamics, and their respective role in creating and maintaining a stable regime.

Chapter 6: Discussion

6.1 Summary of results

What do distinct cell types do in the brain? In this thesis, I have made the case that this can only be answered by first having the tools required to properly collect and analyze this complex data. How has this been achieved? By tightly integrating the analysis with an experimental protocol to dynamically probe and sort recorded neurons reliably.

In Chapter 1, I presented an overview of the electrophysiological properties of the mammalian brain. From the distinction between the action potential's electrical footprint in the intracellular and extracellular space, to multielectrode array systems allowing us to sample their activity (and sometimes noise too), making our way to a particularly interesting brain region (the prefrontal cortex) and its landscape composed of distinct cell types (excitatory and inhibitory neurons), this chapter ends by introducing the main challenges researchers face when trying to study all of those pieces together.

In Chapter 2, I studied the features of spike sorting for large-scale multielectrode arrays. In the first part of this chapter, the main steps of spike sorting are presented and key challenges are highlighted. Next, I presented an open-source, scalable data-analysis pipeline composed of a series of Python and Matlab code I implemented to automatically perform spike sorting on large datasets. I demonstrated this method on data I acquired from the prefrontal cortex of mice.

In Chapter 3, I dived deeper into the action potential, describing the waveform features that distinguish distinct cell types between them. I integrated this knowledge into the data analysis

pipeline, and further classified identified single neurons as either inhibitory or excitatory based on their action potential features.

In Chapter 4, I introduced the simultaneous manipulation and recording of a large-scale cortical network. For this, I tested and developed a protocol taking advantage of a unique viral strategy to activate neurons with light over a protracted timescale. Intracellular recordings were first acquired to characterize the response profile of this opsin, and then a protocol to be used whilst conducting large-scale multielectrode array recordings is presented.

Finally, in Chapter 5, I recorded the individual neuronal responses to step function opsin activation of a particular subtype of neurons (parvalbumin). With the established protocol, I extracted a subclass of neurons which were responsive to light stimulation and showed that these were a subset of neurons which had been classified as inhibitory with the prior spike sorting of the action potential waveform. Response to light stimulation without the presence of synaptic blockers, in some cases, led to a counter-intuitive decrease in firing rate for neurons labelled as inhibitory. Such dynamics are seen in networks operating in inhibitory stabilized regime, and evidence in favor of this regime had not previously been recorded in large-scale multielectrode recordings containing such a high number of recording electrodes with labelled excitatory and inhibitory neurons.

Together, these chapters provide tools to study spiking activity in large populations, with a consideration for the distinct units that compose it and allow for complex network dynamics to emerge. Such methods are necessary to accurately study how neuronal composition is involved in biological networks. By constructing an experimentally validated cell type-specific spike sorting analysis pipeline, I arrived at the conclusion that a combination of these tools is necessary to accurately track the complex activity of neuronal networks.

6.2 Challenges in large-scale neural recordings

High-density extracellular electrophysiological recordings have allowed the spiking of thousands of neurons to be recorded simultaneously, in largely intact networks. Their versatility has enabled the collection of large datasets across many brain regions. Indeed, pioneering work over five years ago established the use of fully integrated silicon probes in awake behaving animals (Jun et al., 2017), in tandem with advances in techniques used to functionally separate the spiking activity of neurons (e.g. Pachitariu et al., 2016; Pachitariu et al., 2023). However, challenges have remained as these technologies and analysis tools evolve in tandem. One example is the extreme difficulty in obtaining ground-truth data for either the actual grouping of spikes into clusters emanating from single neurons, or in cell type-specific measurements. In principle, this problem can be solved by the use of detailed biophysical simulations to generate synthetic datasets (e.g. Hagen et al., 2015) or, for example, simultaneous juxtacellular recordings from neurons. Here, by presenting a method for optogenetic stimulation of specific cell types while simultaneously recording their activity, I present one possible solution towards obtaining ground-truth validation for spike sorting.

It is to be noted that other challenges remain for conducting large-scale neural recordings from awake behaving animals that are not a problem for the multi-electrode arrays used here. One example is that brain movement can distort the waveforms of spikes or shift their apparent locations up and down the probe shaft; signal registration is one possible solution to this. Such registration is not a technical issue for the multi-electrode array recordings presented in this thesis but would be a potential source of error when applying the spike-sorting methods here to *in vivo* recordings, and would have to be carefully integrated into existing pipelines that incorporate registration. Another challenge is the localization of probes in the brain, which might complicate efforts to obtain anatomically precise viral expression of opsins in regions close to the probe, as well as to photostimulate regions of the brain close to the probe.

6.3 Limits of this work and future directions

Together, the results and method presented raise questions about the overall implications of not just one, but the many more distinct cell types that make up neuronal networks and contribute to emerging dynamics. This work has selected to dynamically probe a single cell type (parvalbumin) to prototype recording, stimulating, and analyzing action potential waveforms in a cell type-specific manner. Cortical inhibitory neurons are heterogeneous, and the fact that fast-spiking PV interneurons represent the largest population exhibiting landmark features of narrow-spike made them an attractive candidate to verify our classification. However, future work including different inhibitory interneurons would prove useful for the classification of all inhibitory cell types. Given that spike sorting features in vitro can differ from properties exhibited in vivo (Henze et al., 2000; Nowak et al., 2003; Gonzalez-Burgos et al., 2005), future work should address if these classification/waveform properties can be generalized to behaving organisms, for example with tetrodes, optrodes or implanted electrode arrays (see Chapter 6: 6.2 Challenges in large-scale neural recordings).

This could be done for the many more cell types that make up our brains. In line with the study of excitatory and inhibitory neurons, this work could be extended to somatostatin and vasoactive intestinal peptide interneurons, or even to distinct subtypes of each kind of inhibitory interneurons. Importantly, such characterization would allow researchers to probe each population independently to simultaneously study the interplay of distinct populations, and the underlying large-scale network-dynamics effects associated.

To further investigate the role of individual cell types in network dynamics, one could adapt this method to target genetically defined populations of neurons releasing key neuromodulators (such as dopamine, serotonin, acetylcholine, norepinephrine, etc.). For example, serotonin neurons are characterized by a particularly large afterhyperpolarizing potential (AHP) as well as by broad spike waveforms, and these features have been reliably

identified from extracellular recordings (Cohen et al., 2015; Harkin et al; 2023). Thus, spike features other than peak-to-peak and half-amplitude width (as employed here) could be used to reliably identify cell types, using a similar formalism to that described in this thesis for optogenetic validation.

Further work could take advantage of the method proposed in this thesis to gather population-specific data and create template models of the extracellular action potential associated with distinct cell types. This would prove useful in studying large networks composed of many distinct cell types and could eventually enable cell type identification without the need to dynamically probe them for validation once reliable, validated templates are achieved.

Building on this idea, creating a model predictive of cell type identity could eventually allow for the online identification of distinct cell types while recording from neuronal populations. With network composition information on hand while acquiring recordings of large and interconnected populations, targeted and spatially constrained stimulation of some selected neurons could be delivered. Allowing researchers to study complex networks with reliably identified cell types that can be probed would assist in understanding or distinct units of complex networks interact.

One of the many obstacles in creating template-base models of each and every cell type in the brain lies in the temporal precision required to extract action potential features and distinguish cell types. Current technology and multielectrode recording systems have a limited sampling rate (approximately 30 KHz for Neuropixels probes, or around 0.03 ms sampling interval). The method presented in this thesis introduces the first application of interpolation to statistically attain a finer precision of the recorded waveform of neurons. Yet, to reliably

distinguish some cell types, or sub-types of cells may require a greater sampling rate than currently possible with both the sampling rate of current systems and statistical methods.

In this line of ideas, it is worth noting that an important tradeoff lies in the sampling rate used to collect experimental data on multielectrode systems with thousands of recording channels which is the recording data size. As such, increasing the sampling rate considerably increases the data size, and processing time, quickly becoming an important bottleneck for the extraction of meaningful activity and data analysis. Utilizing statistical methods, closely paired with an experimental protocol to provide experimental validation, as in this method, aims to be a first step in bridging this tradeoff.

Another such tradeoff is the number of electrodes on the recording platform and the distance between them. In dense populations, many neurons and their respective compartments are tightly packed around each recording site. As can be the case in a room full of people, where, for example, a single microphone is dropped in the middle of the room. An audio recording of this microphone would contain multiple streams of information (everyone's voice), each modulated by a distinct set of factors (such as the distance to the microphone, or even voice volume). In such a scenario, resolving the speech of everyone can be challenging. Alternatively, if we were to record the same room full of people with 1000 (or say 4096), equally spaced microphones, resolving and separating the speech of everyone would be easier. The number of people in the room, and their spatial distance helps in determining the optimal number of microphones to accurately capture everyone's contribution to the room. This analogy highlights how the use of numerous recording sites helps to resolve and assign activity to distinct units. This, however, is computationally costly and too many recording sites, say more than the number of units sampled, would only increase the difficulty in identifying and tracking the respective activity of each individual unit.

This work is also impacted by other limits inherent to multielectrode arrays recordings. These devices offer several advantages over traditional intracellular recording methods, such as the ability to sample from thousands, or more, of interconnected neurons spanning entire brain regions simultaneously. However, these systems also have several disadvantages. One key challenge is the interpretation of the extracellular signal, and the lack of ground truth without having access to the internal space of neurons. To ensure optimal interpretation, the spike sorting method introduced here was designed cautiously by introducing features such as a spike detection threshold analysis, an independent criterion to determine the number of clusters, and the experimental validation of the cell type classification. However, these strict criteria make it possible that some meaningful activity was discarded.

Another important challenge pertains to the light-sensitive nature of multielectrode array recording platforms. Exposing the recording array to light generates an artefact, thus constraining the temporal precision with which the underlying activity can be resolved following opto-activation. The use of a stabilized step function opsin allows for the activity to be recorded past the light-generated artefact. While this methodology allows for the prolonged activation required to see activity on this system, this methodology also presents some challenges. One such challenge pertained to achieving deactivation of the opsin using a pulse of yellow light. For this, several Plexon LED modules spanning a broad spectrum of wavelengths (525 nm to 660 nm) were unsuccessful in achieving while conducting single-cell patch clamp recordings. Initial work (Chapter 4), allowed for the identification of a yellow activation wideband 545-580 nm filter (BP545-580; Olympus) as required for photodeactivation. For this, sufficient power output was needed to photodeactivate the step-function opsin. In multielectrode array recordings, the recording electrodes are located at the bottom of the recorded slice, requiring light to penetrate through the slice to deactivate the recorded neurons. In the case of the deactivating opsin, where power is critical to reach complete deactivation, it is possible that not all neurons were successfully deactivated (Figure

27). Another possible challenge with the use of a stabilized step function opsin is the possibility that neurons eventually enter depolarization block following prolonged activation.

It is also worth noting that the use of strict threshold for spike detection, as in this method, minimizes the chance of capturing noise and assist in isolating reliable spiking activity. However, this could be a modulating factor in studying neuron dynamics, such as inhibitory stabilized networks as it could increase the likelihood of approaching a floor effect, where the overall activity is attenuated from the strict criteria applied to isolate activity. Thus, it is possible that some meaningful activity involved in broader network dynamics is missed.

Several methodologies have been developed over the past years to perform spike sorting on recordings of extracellular voltage. A few of the most used in the field include Kilosort4 (Pachitariu et al., 2023), SpyKING CIRCUS (Yger et al., 2018), MountainSort (Chung et al., 2017), Herding Spikes (Hilgen et al., 2017), YASS (Lee et al., 2017), and JRCLUST (Jun et al., 2017). Each method has specialized features, making them more or less suited to certain types of neuronal recordings. A detailed comparison of these methods is presented in Table 1. A common lacking feature in currently available methods is the ability to classify excitatory and inhibitory neurons based on their action potential waveforms. The method introduced in this thesis not only allows for cell type classification, but biologically validates this classification and provides an experimental protocol to dynamically probe a genetically defined population. This can further be extended by researchers to other cell types of interest. The method presented here also distinguishes itself by being unsupervised, making it easy to use for large-scale implementations. For those with specific needs, the source code is available and can be customized.

	lang.	# elect.	data preproc.	event/ channel loc.	wavef orm extrac.	e/i sorting	valid ation	experim . prot.	unsu per.	source code available
<i>Giraud</i> (This method)	Python and Matlab	4096	+	+	+	+	+	+	+	+
Kilosort4 (Pachitariu et al., 2023) github.com/cortex- lab/KiloSort	Python	5120	+	+	-	-	-	-	-	+
SpyKING CIRCUS (Yger et al., 2018) spyking- circus.rtfld.org	Python	4225	-	+	+	-	(+)	-	+	+
MountainSort (Chung et al., 2017) github.com/ flatironinstitute/ mountainsort	Matlab and C++	16	+	+	-	-	-	-	+	+
Herding Spikes (Hilgen et al., 2017) github.com/ mhhennig/HS2	Python	4096	+	+	+	-	(+)	+	+	+
YASS (Lee et al., 2017) yass.readthedocs.io	Python	512	-	+	-	-	-	-	+	+
JRCLUST (Jun et al., 2017) jrclust.org	Matlab and CUDA	120 (1000)	+	+	-	-	(+)	-	-	+

Table 1: Comparison of available spike sorting methods

lang.: Programming language, **# elect.:** Maximum number of electrodes tested on, **data**

preproc.: Data preprocessing (formatting, filtering, artefact removal) included, **event/channel**

loc.: Localization of single events, **waveform extrac.:** Spike features (such as peak-to-peak

distance and half-amplitude width) are extracted from waveforms, **e/i sorting:** excitatory and

inhibitory cell sorting, **validation:** validation of cell type sorting (+) = ground truth validation via

intracellular recordings, but not e/i validation, **experim. prot.:** integrates an experimental

protocol to allow for network manipulation, **unsuper.:** unsupervised, requiring no manual user-

intervention, **source code available:** entire code is open-source and available.

To reconcile the precision of a complex architecture of intricately connected neurons with the emergence of a precisely regulated neural code, we must first develop accurate methodologies capable of capturing the physiological composition of neural circuits and bridge it with its functional implications. This work provides evidence that extracellular recordings can offer insights on how distinct neuronal populations function. This pipeline encourages the integration of viral strategies, optogenetics and electrophysiological recordings to manipulate complex networks to better understand the contribution of each cell type. It opens the door to studying how these complex networks function and relate to behavior and decision-making in awake, behaving animals.

6.4 Inhibition-stabilized networks and dynamics

Here, dynamically probing a single cell type (parvalbumin) in physiological conditions without the presence of synaptic blockers led to a paradoxical decrease in firing rates for both putative excitatory and inhibitory neurons. Although seemingly paradoxical, it has been reported that in networks with strong recurrent coupling, such as in the cerebral cortex, excitatory cells could be self-amplifying unless stabilized by inhibition (Litwin-Kumar et al. 2016; Tsodyks et al., 1997). A key finding presented in Thivierge et al. 2022 (Appendix B) is that distinct subtypes of inhibitory interneurons, such as somatostatin, play a key role in shaping the response of an entire network. Future work targeting, and probing, different inhibitory interneurons would prove useful in studying the involvement of distinct cell types, and within other regions of the brain.

To further investigate the role of distinct cell types in network dynamics, the methodologies presented in this thesis could be used for large-scale reconstruction of functional networks of genetically identified neurons during spontaneous activity. The integrated suite with an open-source data analysis pipeline paired with a validated

experimental protocol would enable high-density and cell type-specific activation during recordings. The insights of this work will allow for the investigation of dense and intricate circuits in a flexible, scalable, and generalizable manner, including the contribution of specific cell types in defined temporal windows to generating large-scale neural dynamics.

Functional networks are defined as graphs where nodes (neurons or channels) are linked by weighted connections (statistical dependencies between pairs of nodes). Several lines of evidence show a non-random organization of functional networks recorded on MEAs (Vincent et al., 2012; Thivierge, 2014; Vincent et al., 2013; Ren et al., 2020). However, few papers have examined functional networks where the nodes are labelled according to their respective cell type (E or I) (Ren et al., 2020). Given that our work generated recordings of neural activity where individual neurons are labelled, it would be straightforward to construct functional networks showing the interactions between E and I cells. Such networks would provide valuable information on the structure of these interactions during normal brain functioning, as well as how these interactions are modulated by external activity and pathological states.

Concluding remarks

Broadly, this thesis has focused on describing an open-source, unsupervised, and scalable suite of computational tools that perform spike sorting and cell type classification on large-scale, high-density recordings. By further integrating the analysis pipeline with an experimental protocol, this method allows for the validation of the statistical method and for the optogenetic manipulation of distinct cell types in dense networks on large-scale multielectrode arrays. These tools collectively provide a formalism for investigating large-scale neural dynamics of genetically defined neural populations.

Appendices

Appendix A

Published paper:

Thivierge, J. P., **Giraud, É.** & Lynn, M. (2023). Toward a Brain-Inspired Theory of Artificial Learning. *Cognitive Computation*, 1–8. <https://doi.org/10.1007/s12559-023-10121-y>

Author Contributions:

The authors contributed equally to the work.



Toward a Brain-Inspired Theory of Artificial Learning

J. P. Thivierge^{1,2} · Éloïse Giraud^{1,2} · Michael Lynn^{2,3}

Received: 24 March 2022 / Accepted: 21 January 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Despite the rapid progress made by artificial neural networks (ANNs), the design of these models makes it unlikely that incremental improvements will eventually bring them on par with human-level cognitive capabilities. Here, three fundamental shortcomings of ANNs are described, namely the strictly statistical nature of the learning process, the inability to handle universal mappings, and the lack of key structural brain features that constitute the building blocks of behavior and cognition. Solutions to these issues are discussed, including the use of few-shot learning paradigms, network architectures inspired by cytoarchitectural features of the brain, and neuromodulator-derived rules for learning and updating environmental variables. We cast these solutions in the broader context of recent discoveries about the brain. Their implementation in ANNs, however, will require a deeper understanding of the cognitome—the map between elementary cognitive functions and the patterns of neural connections that support them. We contend that to reach true human-level cognitive capabilities, ANNs require both a principled approach to extending their architecture and learning rules and a deep understanding of the cognitome linking brain structures to mental operations.

Keywords Artificial neural networks · One-shot learning · Neuromodulators · Universal mappings · Cognitome

Introduction

The past century has witnessed the gradual emergence of neuroscience as a field of study, from sparse observations of synaptic inputs in single neurons [1] to the recent capability of scientists to record from thousands of cells while animals—or humans—are engaged in a task [2]. Along with these innovations, theories of brain function have made headway into linking neural activity to cognitive and motor processes and their associated learning mechanisms [3, 4].

Because descriptions of the brain range in scales from molecules [5] to synaptic networks [6] and ultimately whole-brain activity [7], there is a central consideration in brain theories: what is the appropriate level of description? A scale too fine (for instance, single ions) would be hard to

link to behavior, while a scale too coarse (such as the activation of the entire frontal lobes) would not capture the computations performed by neural circuits.

Converging evidence dating back to the 1940s has established that a key candidate for describing the computations performed by neural circuits is cell assemblies [8]. These assemblies form groups of roughly 10^4 – 10^5 neurons, infinitesimally small in comparison to the scale of the whole brain but powerful enough to carry out the canonical tasks of learning, memory, navigation, motor control, sensory perception, and countless more. For instance, consider the functional subgroup of neurons activated when an image of an apple is presented. A subset of these neurons may also be activated when a pineapple is presented (phonetic similarity) or when an image of Newton is presented (conceptual link) (Fig. 1).

Without a doubt, the most convincing way to show how a cell assembly works is to build one a vulnus—that is, from scratch. Indeed, building a neural circuit that captures the building blocks of living brain networks and performs computationally useful tasks is widely accepted as a necessary and sufficient condition for understanding the brain [9]. This goal has pervaded the field of theoretical neuroscience since its inception and forms the core of the

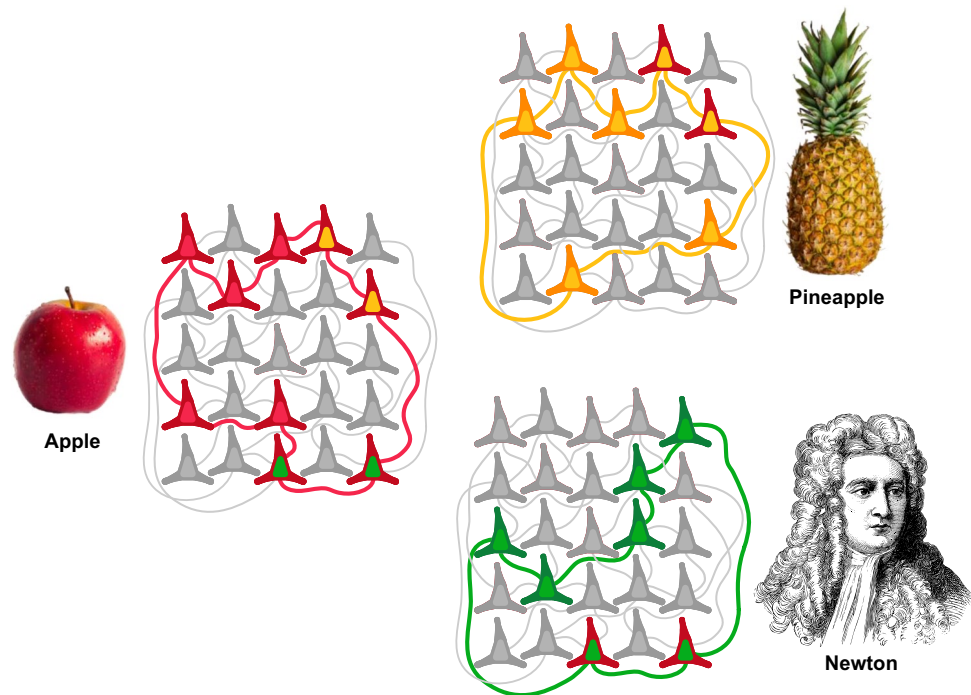
✉ J. P. Thivierge
jean-philippe.thivierge@uottawa.ca

¹ School of Psychology, University of Ottawa, 136 Jean Jacques Lussier, Ottawa K1N 6N5, Canada

² Brain and Mind Research Institute, University of Ottawa, Ottawa, Canada

³ Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, Canada

Fig. 1 Cell assemblies: the coordinated activation of local groups of neurons underlies sensory perception and behavior



connectionist paradigm—the idea that cognitive abilities can be explained by ANNs [10]—extending to modern deep learning approaches.

Connectionist models, however, have been criticized in many respects, including for their seeming inability to represent key attributes that underlie human cognition [11]. Granted, it is easy to find fault in models that are deliberately designed as an oversimplification of both neural and cognitive processes. As these models become more refined, it is tempting to keep “moving the goal posts” (to paraphrase Dan Rockmore in a piece for the *New Yorker*) toward more complex challenges in our understanding of the brain. Nevertheless, here we will argue that the goal posts were misguided, to begin with—ANNs are unfortunately not on the gradual path toward a complete characterization of how thoughts and actions emerge from neural circuitry. We describe three major roadblocks and provide suggestions for moving forward.

Rare and Unusual Data

Perhaps the most prominent success story in the field of ANNs is the ability of these models to correctly classify novel exemplars of a class once trained on a dataset of labeled inputs [12]. For instance, a network trained on several images of cats will adequately classify a novel image as belonging to that category. However, there is an underlying assumption of the learning process that is not often disclosed, namely that the training data follows a distribution

roughly centered around “typical” exemplars (Fig. 2). In this way, the most common instances will be the ones that best represent the category. Departures from this assumption have debilitating effects.

As a telling illustration, consider what happens when we input a picture of an average-looking tabby cat into Google reverse image search: the output provides a definition of domestic short-haired cats, as one might expect. However, when presented with an atypical picture of a cat in a nobleman’s clothes, Google reverse image search identifies it as Maximilien II de Béthune, despite the image lacking human-like facial features. Here, the system cannot adequately process the rare illustration that was presented. As a result, its output misses the mark and fails to reflect the key characteristics of the image.

The point is that ANNs perform poorly on datasets containing rare and unusual instances. This is due to the statistical nature of these systems—neural networks are designed to capture regularities in a set of data and form boundaries around well-defined categories. Humans, on the other hand, can readily identify the portrait of a cat, even though most of us have had little or no exposure to similar images.

Arguably, ANNs could eventually learn to classify rare examples if given enough data and sufficient learning time. However, even the most sophisticated models would need more data than the average school-aged child to become accurate at classifying images [13]. A dataset of 10^7 examples, for instance, corresponds roughly to one labeled image per second for an entire year. (Worse still, a disquieting amount of real-world data is unlabeled.) Even a smaller

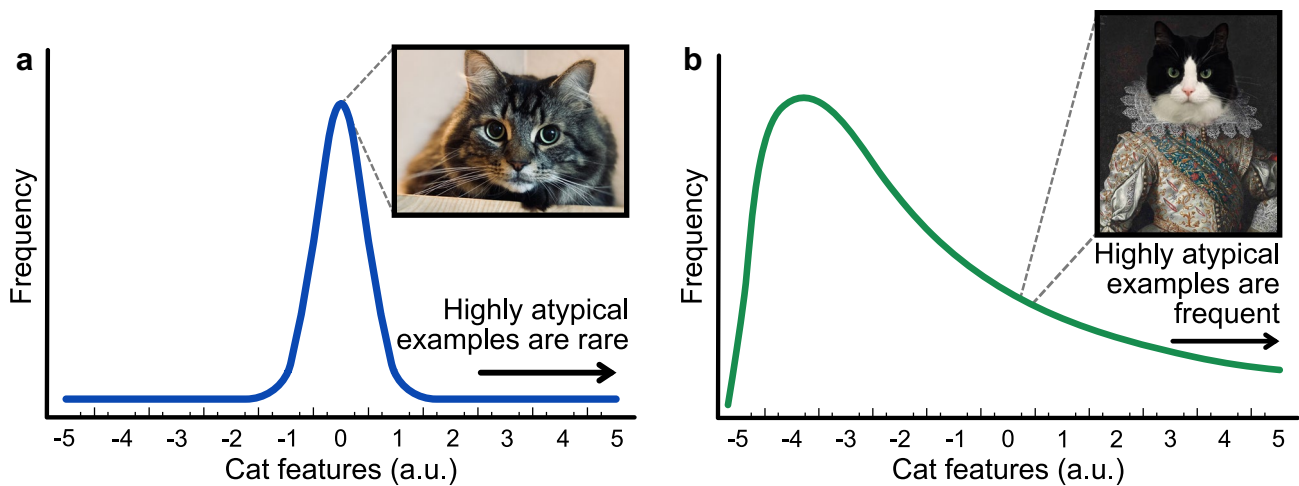


Fig. 2 The impact of data distribution on learning and representation. **a** Normally distributed data with few highly unusual cases of cat features in arbitrary units (a.u.). **b** Skewed distribution where unusual examples are located along the heavy tail to the right of the mean

dataset like the popular MNIST handwritten characters [12] contains 60,000 examples, and good performance from a deep neural network is only achieved after hundreds of iterations that cycle through these images [14].

There are, of course, several avenues for addressing the issue of long training times in neural networks, notably the use of prior knowledge [15] and rapid, one-shot learning of natural images [16, 17]. However, these approaches are still sharply vulnerable to the problem of rare and unusual exemplars. As ANN pioneer Yann Lecun stated, “Human children are very quick at learning human dialogue and learning common sense about the world. We think there is something we have not uncovered yet—some learning paradigm that we have not figured out” [18].

Clearly, biological brains have developed sophisticated algorithms over long evolutionary timescales to navigate this set of challenges, including a sensitivity to rare examples, an ability to parse unlabeled data, and rapid few- or one-shot learning. What are the underlying neural architectures and computational principles that imbue biological networks with these capabilities? A widespread principle across brain structures is that learning is under the strict control of neuromodulators, such as serotonin, dopamine, and norepinephrine, that exert multifaceted effects on behavior and cognition [19]. For instance, serotonin has been shown to regulate brain dynamics [7], indicate uncertainty [20], represent cue-outcome contingencies [21], and guide selective attention [22].

Emerging results paint a picture where environmental statistics are tracked by neuromodulators and employed to guide learning in a “meta-learning” fashion where the learning process is itself subjected to plasticity—which could imbue networks with heightened sensitivity to learning from rare or dynamic examples. As an illustration, consider

a scenario where mice are presented with a dynamic foraging task with noisy outcome contingencies that are altered periodically (for example, varying the probability of water delivery to two licking spouts). Recent work has suggested that the meta-learning characteristics observed behaviorally are controlled by levels of serotonin and are successfully captured by a meta-learning model where the learning rate is modulated by unexpected uncertainty (i.e., how fast the task statistics are changing), allowing rapid behavioral adaptation to changing outcome contingencies [20]. It appears, in short, that the brain has evolved a finely tuned neuromodulation machinery to efficiently contend with precisely the types of real-world inferences, as well as rare and unlabeled data, that pose a challenge to many ANNs.

Universal Mappings

Universal mappings (UMs) are computations that connect features of the input to features of the output of a neural circuit [23]. A simple example is the “identity” function, where each input feature is faithfully reproduced at the output (Fig. 3). As an illustration, consider a task where the learner is presented with pairs of inputs and outputs that are arrangements of the letters A and B. Afterwards, the learner is provided a test pattern (“B B B”) and asked to generate an output. A vast majority of able-minded individuals will provide “B B B” as their answer to the test pattern, inferring that the task is to perform an identity mapping. However, most ANNs will produce “A B B” instead of the human answer. The reason is rooted in the statistical nature of artificial learning: notice that the first letter of all training examples is always “A.” ANNs exploit this regularity, while humans seem to rely on other factors.

Universal Mapping for the Identity Relation

	INPUT	OUTPUT
Training patterns	ABA	ABA
	AAA	AAA
	ABB	ABB
	AAB	AAB
Testing patterns	BBB	?
Human:	BBB	✓
ANN:	ABB	✗

Fig. 3 A learning task with a universal mapping representing the identity function

While this task appears trivial, it raises what has been deemed one of the most central issues in cognitive science [24], namely, the inability of ANNs to learn anything other than pure statistical regularities. Solutions have been proposed [25, 26], but rely heavily on the injection of prior knowledge in the network (specifically, a bias toward the identity function).

The fact that human solvers do not rely on strict statistical rules when solving this type of problem suggests that other mechanisms—largely elusive to ANNs—are at play. Inherent biases may constitute one such mechanism, based on neural circuits that are hard-wired for pattern detection involving symmetry [23, 27]. In any case, the complexity and computational power of ANNs are utterly powerless in capturing a basic human ability such as the UM of identity. The title of a recent talk by Alessio Plebe and Nicole Cilia captures the essence of this challenge: “La difficoltà nel simulare la semplicità,” or the “difficulty in simulating simplicity [28].”

The Complexity Fallacy

Merely adding more features (including a diversity of connections, transfer functions, and learning objectives) to a model does not render that model more plausible from a biological standpoint. Lines of reasoning that fall prey to the “complexity fallacy” can be summarized as follows: given that the brain is a complex system and an ANN is also complex, it must be a good model of the brain. Although it may seem trite to argue that more complex models of the brain are not necessarily better ones, by doing so we open the door

to a broader discussion of what constitutes a “good” model and what is an appropriate level of description.

While it may be useless to build a model as complex as the brain, as it would yield no mechanistic insights, the expectation is that a successful model is one that captures the complex structure and computations of the brain with only a minimum number of adjustable parameters. Yet, many ANNs exhibit little of the dynamical richness seen in the brain. Specifically, neural circuits generate patterns of activity that evolve over time and are directly linked to thoughts and behavior.

A key feature of these patterns is what we term “hierarchical embeddedness”: the idea that interactions amongst elements at one scale influence interactions at other scales (Fig. 4). In this way, the coordination of large-scale traveling waves across the surface of cortex is accompanied by the local synchronization of local field potentials within spatially delimited circuits. Further down into microscopic scales, the synchronized activation of clustered synaptic inputs is associated with calcium amplification and propagation in single dendritic compartments, a process that ultimately affects local plasticity [29].

Because of the intrinsically chaotic nature of brain activity, changes at one spatial scale can reverberate at all other scales. This rich dynamical interaction across scales is key to the integration and segregation of information across the brain [30] and is one of the most important concepts in linking neural activity to higher cognitive function [31, 32]. For instance, the interactions between connectivity, synaptic dynamics, and subcellular compartments provide an efficient computational solution for representing independent streams of information in the same neural population, as well as for demultiplexing signals in downstream areas [33]. Rather than falling prey to the complexity fallacy when designing ANNs, we posit that the key features of biologically realistic architectures are first, hierarchical embeddedness across multiple spatial scales, and second, the presence of the kinds of dynamics exhibited by real brains. A renewed focus on these features may help ANNs approach general-purpose intelligence.

What We Need Moving Forward

The current paper raised three fundamental problems with ANNs that prevent them from attaining a full description of the loop between the brain, mental processes, and behavior. We now consider what key ingredients are missing from these models.

Firstly, ANNs would benefit from new learning paradigms that can learn from one (or few) examples instead of large datasets, which may involve implementing hierarchical learning rules inspired by biological systems. Instances of

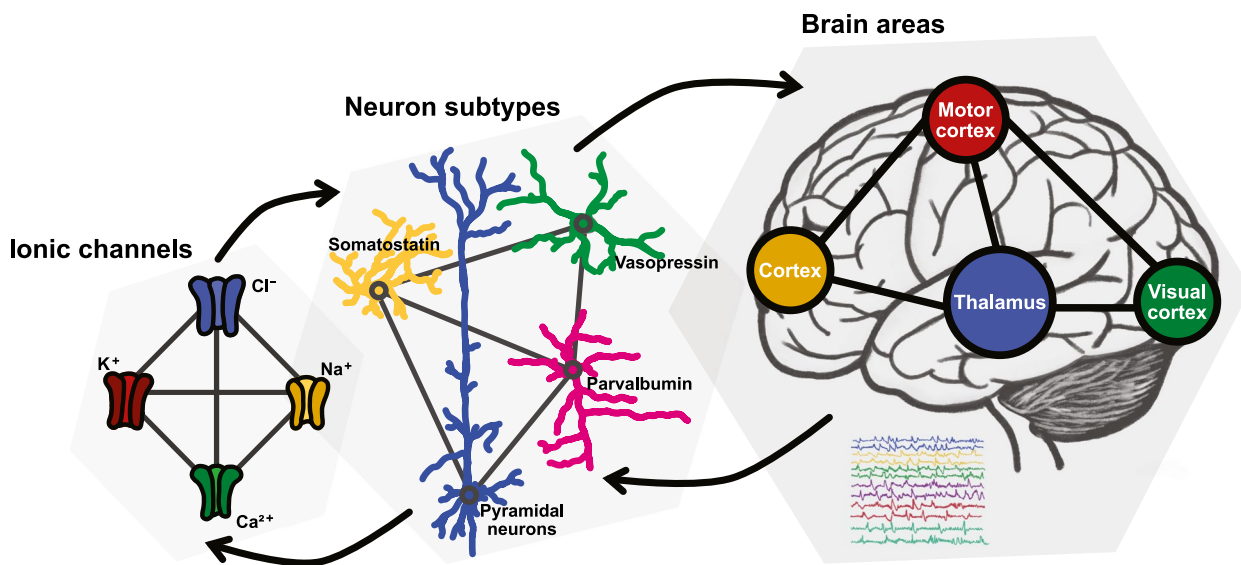


Fig. 4 Hierarchical embeddedness across spatial scales of brain activity

learning rules based on one-shot learning are currently being developed [16, 34]. In one such model, weights are adjusted in a single iteration of the learning rule by taking a “snapshot” of network activity and applying a least-squares criterion [19]. It remains controversial, however, what constitutes one-shot learning [35], and many models do not follow the same standards, making direct comparisons difficult.

Furthermore, one-shot learning rules remain largely based on a normative approach where an objective function is optimized under a set of constraints. Humans and animals, by comparison, often exhibit “cognitive laziness” and nonoptimal shortcuts during learning and decision-making. These shortcuts, while perhaps reflecting failures in engaging brain states involved in optimal decision-making [36], may also account for the highly efficient way in which sensory information is categorized and dynamically routed. This may occur by selectively attending to limited features of an input or even by calibrating inferences about uncertain sensory stimuli based on past history [37]. We suggest that by taking a broader view of cognitive shortcuts and behavioral dynamics shown by humans during decision-making and learning, we might be able to infer architectures and learning rules that support dynamic decision-making and one-shot learning in artificial systems.

It is worth remarking that modern deep neural networks are beginning to integrate mechanisms for hierarchical learning that bear some conceptual similarities to those employed by neuromodulatory systems in the brain. Such mechanisms can allow networks to efficiently play classic ATARI video games, for example, by breaking the task down into policies

over intrinsically generated goals and actions [38]. Furthermore, advances in few-shot learning have relied on a meta-learning model to first learn the optimization algorithm, followed by classifier training [34]. More specifically, a meta-learner receives as input a particular data set and outputs one of several possible classifiers (different neural networks) that is predicted to achieve high classification accuracy on these data. These advances rely on novel hierarchical architectures, typically composed of multiple stages of learning and control. In a similar fashion, the brain’s neuromodulatory systems could be argued to implement hierarchical control of downstream regions by tracking higher-order environmental and internal variables and communicating these to other neural structures.

We contend that the architecture, learning rules, and division of labor present in biological neuromodulatory systems could act as inspiration to solve the challenges faced by ANNs around parsing rare and unusual examples from rich streams of data. However, unlike traditional control theory, neuromodulatory control does not require knowledge of the state of the system as a whole, as synaptic plasticity rules are primarily local. Furthermore, neuromodulatory control can handle neural structures that are constantly changing over the course of development, while basic control theory has trouble with systems whose state space is itself reconfiguring over time.

Furthermore, mammalian brains are the product of haphazard evolution and not optimal engineering. Evolution employs strategies that differ markedly from optimized solutions. For instance, evolution often co-opts

Table 1 Comparison between traditional connectionist models and an approach based on neuromodulatory control of neural activity

Connectionist models	Neuromodulatory control	Examples	Refs
Strictly statistical learning	Inclusion of nonstatistical learning	Learning rare examples	[22]
Fixed learning parameters	Parameters evolving with meta-learning	Classifier selection	[34]
		Learning rates	[20]
Slow and gradual learning	Few- and one-shot learning	Perceptual encoding	[16]
		Classification	[34]
No prior knowledge	Learning biases and priors	Symmetry bias	[23, 27]
Layered network architecture	Hierarchical embeddedness	Across spatial scales	[30, 31]
		Across temporal scales	[33]
Gradient descent optimization	Nonoptimal learning	“Cognitive laziness”	[36]
		Recycling of brain motifs	[23]

older structures to address new goals. As phylogenetic adaptation sprouted new neural pathways in the neocortex of primates and early humans, it may have recycled some of the same “building blocks” found in older brain structures. As a striking illustration, motifs of lateral inhibition are found in peripheral circuits of the retina [39] and olfactory bulb [40], as well as within the prefrontal cortex [41], where they support key roles of neural suppression and competition. By comparison, ANNs are often trained “de novo,” with no predefined motifs. Incorporating some of the known building blocks of the brain—including not only lateral inhibition but also synaptic convergence, divergence, and feedforward inhibition—offers an exciting avenue to speed up training and incorporate architectural features of the brain into artificial models.

As an added benefit, ANNs built on cytoarchitectural features of the brain may generate more dynamical richness than their predecessors. Vanilla ANNs trained with backpropagation are typically based on a feedforward architecture that severely limits dynamical richness given their inability to propagate activity internally. As an alternative, recurrent architectures can generate diverse forms of activity that closely capture living brain circuits [42]. However, it remains challenging to train these networks. One suggestion emerges from models with feedback connections from the output to the input neurons of the network to facilitate learning [43].

Hybrid models that strike a balance between powerful feedforward networks and more biologically grounded recurrent networks provide a fertile ground for new discoveries [44]. Recent successes in blending artificial and brain-inspired neural networks are beginning to emerge [33, 45]. A summary of comparisons between traditional connectionist models and a more biologically-inspired approach based on neuromodulatory control is provided in Table 1. This list, while non-exhaustive, provides a launching point for incorporating some of the elements discussed above in models of artificial learning.

Conclusion

Despite decades of progress in advancing the computational capabilities of ANNs, it is hard to conceive that a series of incremental improvements, building on past successes, will eventually attain the full capacities of the human mind. Rather, we propose that a change of paradigm will be required to shift away from de novo, statistical, and dynamically poor learning approaches to models that are built upon known neuronal “building blocks,” are nonstatistical, and offer dynamically rich patterns of neural activity that capture those of living neural circuits.

With the incorporation of neuronal building blocks into ANNs, however, there remains the question of how these blocks can be assembled to generate cognitive functions. This question will be best addressed by the emerging field of research into the human cognitome—the collection of elementary cognitive functions and their link to neural assemblies [46]. As an example of how cognitomes may be cataloged, consider a task of action selection between two alternatives (for instance, choosing to turn left or right at an intersection). This task likely involves building blocks such as a recurrent circuit (memory maintenance of the goal), lateral inhibition (suppression of competing choices), etc. [47]. A comprehensive description of cognitomes would allow researchers not only the ability to decompose complex tasks into a series of basic primitives but also the reverse—namely, to assemble primitives into broader neural systems that perform complex operations [48]. We hope that the coming years will witness the burgeoning of novel approaches that take advantage of neural motifs and neuromodulation to attain human-level learning.

Acknowledgements This work was supported by a discovery grant to J.P.T. from the Natural Sciences and Engineering Council of Canada (NSERC Grant No. 210977) and a doctoral scholarship to E.G. from the Fonds de recherche du Québec en Nature et Technologies.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

References

- Castillo J, Katz B. Quantal components of the end-plate potential. *J Physiol*. 1954;124:560–73.
- Jun J, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature*. 2017;551:232–6.
- Miller EK, Lundqvist M, Bastos AM. Working memory 2.0. *Neuron* 2018;100:463–475.
- Shine JM, Breakspear M, Bell PT, Martens KE, Shine R, Koyejo O, et al. Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. *Nat Neurosci*. 2019;22:289–96.
- Gaiteri C, Mostafavi S, Honey CJ, De Jager PL, Bennett DA. Genetic variants in Alzheimer disease — molecular and brain network approaches. *Nat Rev Neurol*. 2016;12:413–27.
- Kato HK, Asinof SK, Isaacson JS. Network-level control of frequency tuning in auditory cortex. *Neuron*. 2017;95:412–23.
- Marquez JC, Li M, Schaak D, Robson DN, Li JM. Internal state dynamics shape brainwide activity and foraging behaviour. *Nature*. 2020;577:239–43.
- Hebb DO. The organization of behavior. New York: Wiley; 1949.
- MacLennan BJ. Connectionist approaches. In: Smelser NJ, Baltes PB, editors. *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier; 2001. p. 2568–73.
- Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psych Rev*. 1958;65:386–408.
- Pinker S, Prince A. On language and connectionism: analysis of a parallel distributed processing model of language acquisition. *Cognition*. 1988;28:73–193.
- Rawat W, Wang Z. Deep convolutional neural networks for image classification: a comprehensive review. *Neural Comput*. 2017;29:2352–449.
- Zador AM. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat Comm*. 2019;10:3770.
- Ciresan DC, Meier U, Gambardella LM, Schmidhuber J. Deep, big, simple neural nets for handwritten digit recognition. *Neural Comput*. 2010;22:3207–20.
- Guo J, Yang C. Constructing deep neural networks with a priori knowledge of wireless tasks. *Arxiv*. 2020. [arxiv:2001.11355](https://arxiv.org/abs/2001.11355).
- Boucher-Routhier M, Zhang BLF, Thivierge JP. Extreme neural machines. *Neural Netw*. 2021;144:639–47.
- Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D. Matching networks for one shot learning. *Adv Neur In*. 2016;29.
- Knight W. Facebook heads to Canada for the next big AI breakthrough. In: MIT Technology Review. 2017. <https://www.technologyreview.com/2017/09/15/149146/facebook-heads-to-canada-for-the-next-big-ai-breakthrough/>. Accessed 11 Jan 2021.
- Mei J, Muller E, Ramaswamy S. Informing deep neural networks by multiscale principles of neuromodulatory systems. *Trends Neurosci*. 2022;45:237–50.
- Grossman CD, Bari BA, Cohen JY. Serotonin neurons modulate learning rate through uncertainty. *Curr Biol*. 2022;32:1–14.
- Cohen JY, Amoroso MW, Uchida N. Serotonergic neurons signal reward and punishment on multiple timescales. *eLife*. 2015;4:e06346.
- Thiele A, Bellgrove MA. Neuromodulation of attention. *Neuron*. 2018;97:769–85.
- Thivierge JP, Marcus GF. The topographic brain: from neural connectivity to cognition. *Trends Neurosci*. 2007;30:251–9.
- Seidenberg MS, Elman JL. Networks are not ‘hidden rules.’ *Trends Cogn Sci*. 1999;3:288–9.
- O’Reilly RC. Generalization in interactive networks: the benefits of inhibitory competition and Hebbian learning. *Neural Comput*. 2001;13:1199–241.
- Smolensky P. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artif Intell*. 1990;46:159–216.
- Huang Y, Xue X, Spelke E, Huang L, Zhang W, Peng K. The aesthetic preference for symmetry dissociates from early-emerging attention to symmetry. *Sci Rep*. 2018;8:6263.
- Plebe A. La difficoltà nel simulare la semplicità. *Proceedings of the 13th Annual Conference of the Italian Association for Cognitive Sciences*. Mind the gap: brain, cognition and society; 2016 Nov 24–26; Turin, Italy.
- Lee KFH, Soares C, Thivierge JP, Béique JC. Correlated synaptic inputs drive dendritic calcium amplification and cooperative plasticity during clustered synapse development. *Neuron*. 2016;89:784–99.
- Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. *Nat Rev Neurosci*. 2018;19:17–33.
- Breakspear M. Dynamic models of large-scale brain activity. *Nat Neurosci*. 2017;20:340–52.
- Marshall JH, Kim YS, Machado TA, Quirin S, Benson B, Kadmon J, et al. Cortical layer-specific critical dynamics triggering perception. *Science*. 2019. <https://doi.org/10.1126/science.aaw5202>.
- Naud R, Sprekeler H. Sparse bursts optimize information transmission in a multiplexed neural code. *PNAS*. 2018. <https://doi.org/10.1073/pnas.1720995115>.
- Ravi S, Larochelle H. Optimization as a model for few-shot learning. 5th International Conference on Learning Representations; 2017 April 24–26; Toulon, France.
- Lake BM, Salakhutdinov R, Tenenbaum JB. The Omniglot challenge: a 3-year progress report. *Behav Sci*. 2019;29:97–104.
- Taghia J, Cai W, Ryali S, Kochalka J, Nicholas J, Chen T, et al. Uncovering hidden brain state dynamics that regulate performance and decision-making during cognition. *Nat Comm*. 2018. <https://doi.org/10.1038/s41467-018-04723-6>.
- Jazayeri M, Shadlen MN. Temporal context calibrates interval timing. *Nat Neurosci*. 2010;13:1020–6.
- Kulkarni TD, Narasimhan KR, Saeedi A, Tenenbaum JB. Hierarchical deep reinforcement learning: integrating temporal abstraction and intrinsic motivation. *Adv Neur In* 2016;30.
- Beckwith-Cohen B, Holzhausen LC, Nawy S, Kramer RH. Controlling horizontal cell-mediated lateral inhibition in transgenic zebrafish retina with chemogenetic tools. *eNeuro*. 2020. <https://doi.org/10.1523/ENEURO.0022-20.2020>.
- Hellmer CB, Bohl JM, Hall LM, Koehler CC, Ichinose T. Dopaminergic modulation of signal processing in a subset of retinal bipolar cells. *Front Cell Neurosci*. 2020. <https://doi.org/10.3389/fncel.2020.00253>.
- Krimer LS, Goldman-Rakic PS. Prefrontal microcircuits: membrane properties and excitatory input of local, medium, and wide arbor interneurons. *J Neurosci*. 2001;21:3788–96.
- Miconi T. Biologically plausible learning in recurrent neural networks reproduces neural dynamics observed during cognitive tasks. *eLife*. 2017. <https://doi.org/10.7554/eLife.20899>.
- Achler T. Symbolic neural networks for cognitive capacities. *Biol Inspired Cog Arch*. 2014;9:71–81.
- Spoerer CJ, McClure P, Kriegeskorte N. Recurrent convolutional neural networks: a better model of biological object recognition. *Front Psychol*. 2017. <https://doi.org/10.3389/fpsyg.2017.01551>.

45. Lilicrap TP, Cownden D, Tweed DB, Cakerman CJ. Random synaptic feedback weights support error backpropagation for deep learning. *Nat Commun*. 2016. <https://doi.org/10.1038/ncomms13276>.
46. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J. Task representations in neural networks trained to perform many cognitive tasks. *Nat Neurosci*. 2019. <https://doi.org/10.1038/s41593-018-0310-2>.
47. Calderini M, Zhang S, Berberian N, Thivierge J-P. Optimal read-out of correlated neural activity in a decision-making circuit. *Neural Comput*. 2019. https://doi.org/10.1162/neco_a_01083.
48. Marcus G, Marblestone A, Dean T. The atoms of neural computation. *Science*. 2014;6209:551–3.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Appendix B

Published paper:

Thivierge, J.-P., **Giraud, E.**, Lynn, M. & Charbonneau, A. T. (2022). Key role of neuronal diversity in structured reservoir computing. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(11), 113130. <https://doi.org/10.1063/5.0111131>

Author Contributions:

J. P. Thivierge: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Eloïse Giraud:** Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – review & editing (equal). Michael Lynn: Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Writing – review & editing (equal). Annie Theberge Charbonneau: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – original draft (equal).

RESEARCH ARTICLE | NOVEMBER 10 2022

Key role of neuronal diversity in structured reservoir computing

Jean-Philippe Thivierge   ; Eloïse Giraud; Michael Lynn  ; Annie Théberge Charbonneau



Chaos 32, 113130 (2022)

<https://doi.org/10.1063/5.0111131>



CrossMark

AIP Advances

Why Publish With Us?



25 DAYS
average time
to 1st decision



740+ DOWNLOADS
average per article



INCLUSIVE
scope

[Learn More](#)



Key role of neuronal diversity in structured reservoir computing

Cite as: Chaos 32, 113130 (2022); doi: 10.1063/5.0111131

Submitted: 18 July 2022 · Accepted: 24 October 2022 ·

Published Online: 10 November 2022



View Online



Export Citation



CrossMark

Jean-Philippe Thivierge,^{1,2,a)}  Eloïse Giraud,² Michael Lynn,^{1,3}  and Annie Th  berge Charbonneau²

AFFILIATIONS

¹University of Ottawa Brain and Mind Research Institute, 451 Smyth Rd., Ottawa, Ontario K1H 8M5, Canada

²School of Psychology, University of Ottawa, 156 Jean-Jacques Lussier, Ottawa, Ontario K1N 6N5, Canada

³Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth Rd., Ottawa, Ontario K1H 8M5, Canada

^{a)}Author to whom correspondence should be addressed: jthivier@uottawa.ca

ABSTRACT

Chaotic time series have been captured by reservoir computing models composed of a recurrent neural network whose output weights are trained in a supervised manner. These models, however, are typically limited to randomly connected networks of homogeneous units. Here, we propose a new class of structured reservoir models that incorporates a diversity of cell types and their known connections. In a first version of the model, the reservoir was composed of mean-rate units separated into pyramidal, parvalbumin, and somatostatin cells. Stability analysis of this model revealed two distinct dynamical regimes, namely, (i) an inhibition-stabilized network (ISN) where strong recurrent excitation is balanced by strong inhibition and (ii) a non-ISN network with weak excitation. These results were extended to a leaky integrate-and-fire model that captured different cell types along with their network architecture. ISN and non-ISN reservoir networks were trained to relay and generate a chaotic Lorenz attractor. Despite their increased performance, ISN networks operate in a regime of activity near the limits of stability where external perturbations yield a rapid divergence in output. The proposed framework of structured reservoir computing opens avenues for exploring how neural microcircuits can balance performance and stability when representing time series through distinct dynamical regimes.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0111131>

The brain is composed of a broad variety of cell types that form complex patterns of connections and play distinct roles in neural computation. While artificial systems known as reservoir computers are successful at capturing complex time series including chaotic attractors, they fail to account for the diversity of neuronal types that support learning in the brain. We propose a novel reservoir computing framework where the interactions between two classes of inhibitory neurons and one class of excitatory neurons control the dynamical regime and stability of the model. In a regime with strong excitation, the network accurately learned to encode a Lorenz attractor but was sensitive to small perturbations applied to a subset of its neurons. By contrast, a regime with weak excitation attained lower accuracy but was more robust to perturbations. Taken together, these results show that interactions among various cell types can control the dynamical regime and performance of a reservoir computer trained on chaotic time series.

I. INTRODUCTION

Reservoir computing models have been trained to generate complex time-evolving dynamics such as chaotic attractors.^{1–6} In these models, a recurrent network sends projections to a set of trainable weights.^{7–10} The recurrent network is typically composed of a homogeneous population of neurons connected in a random fashion, with some exceptions where models exhibit structured connectivity.^{11,12}

However, reservoir models overlook the diversity of neuron types reported in the brain. In the cerebral cortex, computations rely on a diversity of precisely connected GABAergic interneurons that control the activity of principal excitatory cells. Between 80% and 90% of inhibitory neurons belong to one of three subtypes, namely, parvalbumin (PV), somatostatin (SOM), and vasointestinal peptide (VIP) expressing cells.^{13–17} Together with pyramidal (PYR) neurons, these interneurons form a canonical circuit linked to a wide range of sensory,^{18–20} attentional,²¹ and associative²² roles.^{23–25}

Within this circuit, the coordination of excitatory and inhibitory neurons is key to the ability of the cortex to control its regime of activity—from highly synchronized^{26,27} to asynchronous,^{28,29} and from stable firing rates³⁰ to irregular fluctuations.³¹ Capturing the diversity of cortical neurons and their interactions in artificial neural networks is a critical step to understanding their role in neural computation.³²

Here, a novel class of reservoir models termed “structured reservoir computing” is introduced where we replaced the traditional random network with a canonical circuit composed of PYR, PV, and SOM neuronal subtypes^{33–38} as well as feedforward inhibition originating from VIP cells.^{13–17} By controlling the relative strength of excitatory and inhibitory inputs to PYR neurons, the model transitioned between a state of weak excitation to a regime where unstable recurrent excitation was brought under control by strong inhibition,^{39–43} also known as an inhibition-stabilized network (ISN).^{41,42,44–50}

The proposed framework was tested on a computational task where the network learned to relay and generate a Lorenz attractor. Analytical and numerical results show that interactions between PV, SOM, and PYR neurons play a key role in producing temporal sequences and prevent recurrent networks from falling into an unstable regime where learning is compromised.

II. RESULTS

A. Canonical circuit with multiple neuronal types

A modeled circuit, including PYR, PV, and SOM, neurons was designed to capture the known interactions between these populations (Fig. 1). This circuit was composed of a recurrent network (the “reservoir”) with output weights projecting from all neurons to a readout layer trained to generate a desired target output.⁵¹ In this framework, weights within the recurrent network remained frozen and only the output connections were subject to plasticity.

Before training the network, we sought to characterize the regimes of activity produced by the recurrent units. Toward this goal, the model was studied in a simplified scenario with a single mean-rate equation for each subpopulation. The mean firing rate r_X at a given time t across a given neural population X evolved

according to a linear integrator,^{52–54}

$$\tau \frac{dr_X(t)}{dt} = -\delta r_X(t) + v_X + \mu + \sum_{j \in X} g_{Xj} r_j(t) + \vartheta_{\text{noise}} \xi_X(t), \quad (1)$$

where X denotes a population of either pyramidal (E), SOM (S), or PV neurons (P), τ is an integration time constant, δ is a leak term, v_X is a constant external stimulation, μ is a bias input, ϑ_{noise} is a parameter that scales the noise, and $\xi_X(t)$ is zero-mean Gaussian noise with unit variance. This simplified model ignores synaptic nonlinearities⁵⁰ and focuses on the linear impact of interneuron subtypes on pyramidal cells as done previously.⁵⁵ In the absence of coupling between populations, Eq. (1) follows an Ornstein–Uhlenbeck process with solution

$$r_X(t) = \frac{v_X + \mu}{\delta} + \left(r_{X0}(t) - \frac{v_X + \mu}{\delta} \right) e^{-\frac{\delta}{\tau}t} + \frac{\vartheta_{\text{noise}}}{\tau} \int_0^t e^{-\frac{\delta}{\tau}(t-s)} dB(s). \quad (2)$$

Neural couplings $g_{Xj} \in \mathbf{G}$ between the three neuronal populations were introduced by the matrix

$$\mathbf{G} = \begin{bmatrix} -W_{PP} & W_{PE} & -W_{PS} \\ -\Delta_I W_{EP} & \Delta_E W_{EE} & -\Delta_I W_{ES} \\ 0 & W_{SE} & 0 \end{bmatrix}, \quad (3)$$

where $W_{XX} = w$. As a result, differences among inhibitory subtypes were solely due to their pattern of connectivity and not their relative strength. Equation (2) is based upon evidence of reciprocal connections between PV and PYR neurons,^{42,55–59} largely unidirectional connections between SOM and PV,^{15,16,59,60} and reciprocal connections between SOM and PYR.^{57,58,60} This connectivity ignores SOM–SOM interactions given their weak impact on circuit dynamics.^{15,16} Feedforward inhibition from VIP to SOM neurons¹⁶ was not included given that the influence of VIP can be subsumed in the model by the external stimulus $v_S = -2$ (unless otherwise noted) received by SOM neurons in Eq. (1).

In this reduced model, the strength of inhibitory contributions from PV and SOM to PYR was modulated by a single parameter (Δ_I). Similarly, the strength of recurrent PYR–PYR interactions was modulated by one parameter (Δ_E). The interaction of the parameters Δ_I and Δ_E determines the stability of neuronal activity. Formally, the stability of the model is provided by the two non-trivial eigenvalues of its Jacobian matrix $\mathbf{G} - \mathbf{I}$,

$$\lambda = \frac{1}{2} \left(\pm w \sqrt{\Delta_E^2 + 2\Delta_E^2 - 8\Delta_I + 1} + \Delta_E w - w - 2 \right). \quad (4)$$

Based on these eigenvalues, the *critical value* of Δ_E was defined as the point where the largest eigenvalue crosses from negative to positive, thus generating instability in the system. This critical value is obtained by setting inhibition $\Delta_I = 0$, yielding the eigenvalues

$$\begin{aligned} \lambda_1 &= -w - 1, \\ \lambda_2 &= \Delta_E w - 1. \end{aligned} \quad (5)$$

We sought Δ_E such that $\lambda_2 > 0$, assuming a fixed connection strength w . For the numerical examples and analyses discussed below, setting $w = 1$ leads to a critical value of $\Delta_E = 1$.

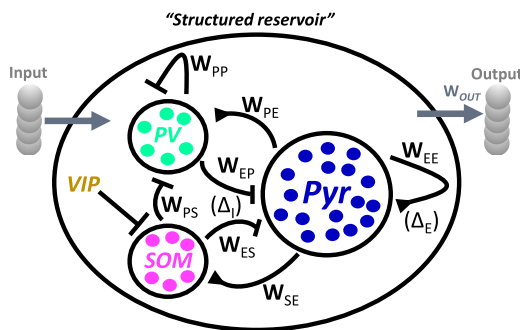


FIG. 1. Circuit of a structured reservoir network with multiple neuronal subtypes.

The interplay between inhibitory (Δ_I) and excitatory (Δ_E) inputs to PYR neurons generated stable and unstable regions of activity [Fig. 2(a)]. For ease of analysis, we considered two regimes of activity where the strength of inhibition remained the same, but excitation is modulated to generate either a state of low activity (non-MSN) or a state of high activity that is stabilized by inhibition (MSN). When the strength of excitation fell below the critical value of Δ_E , the model remained stable regardless of inhibition, given that excitation was too weak to generate runaway activity. This region of the parameter space is characteristic of a non-MSN regime. Conversely, above the critical value of Δ_E , stability was compromised if inhibition was too low to compensate for excitation. This scenario is consistent with an MSN regime where strong excitation must be compensated by strong inhibition to ensure stability.

Increasing the overall strength of connections in the model (controlled by w) resulted in a smaller stable region [Fig. 2(b)]. Hence, a simple circuit with connections between PYR, PV, and SOM neurons exhibited a range of behaviors consistent with MSN and non-MSN regimes reported in recent theoretical^{35,44} and experimental^{42,49} work.

These results were compared to a simpler model with a single inhibitory subtype [Fig. 3(a)], in which case the connectivity becomes

$$J = \begin{bmatrix} -W_{II} & -\Delta_I W_{EI} \\ W_{IE} & \Delta_E W_{EE} \end{bmatrix}, \quad (6)$$

with eigenvalues

$$\lambda = \frac{1}{2} \left(\pm \sqrt{w^2(-4\Delta_I + \Delta_E^2 + 2\Delta_E + 1) + \Delta_E w - w - 2} \right). \quad (7)$$

The state space of this simplified model [Fig. 3(b)] revealed a markedly smaller stable region than the model with multiple inhibitory subtypes [Fig. 2(a)], resulting in a transition to instability at lower values of excitation (Δ_E). Hence, the presence of multiple inhibitory subtypes promoted a broader region of stability than a model with a single inhibitory population.

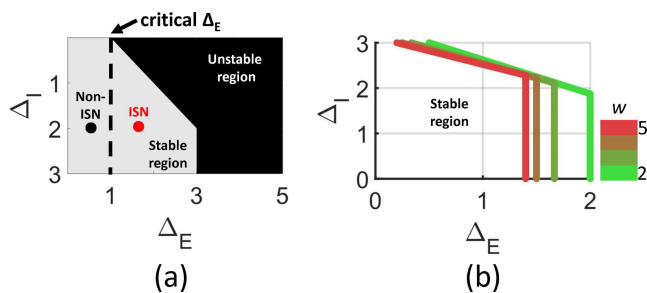


FIG. 2. Regions of stability and instability based on the strength of excitatory (Δ_E) and inhibitory (Δ_I) inputs into pyramidal neurons. (a) Within the stable region, MSN and non-MSN networks fall on different sides of the critical Δ_E . (b) Impact of coupling strength (w) on network stability.

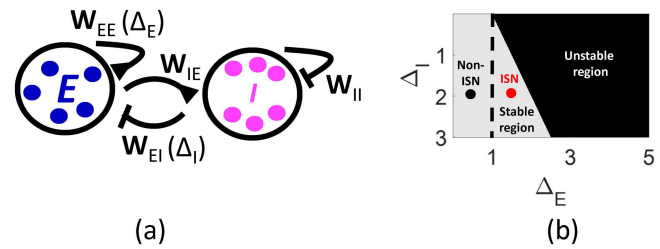


FIG. 3. (a) A modeled circuit with a single inhibitory subtype yields a state space (b) with stable and unstable regions.

B. Paradoxical response to PV stimulation

A defining feature of MSN activity is that inhibitory activation leads to a *decrease* in steady state inhibitory firing rates.^{35,40,41,46,47,61–63} This paradoxical effect is not reported in non-MSN networks (although it can be observed in unstable networks⁶⁴), where a perturbation applied to inhibitory neurons causes an increase in their firing rate. Following experiments showing an MSN-specific paradoxical effect obtained by optogenetically activating PV neurons,⁴² numerical simulations were performed where a constant external input ($v_p = 20$) was applied to the PV population of the model in the MSN state ($\Delta_E = 1.2$, $\Delta_I = 2$). This stimulus altered the rate of all populations in the model, leading to an increase in PV activity and a decrease in both SOM and PYR [Fig. 4(a)].

Mean inhibitory activity, defined as $r_{inhib}(t) = (r_p(t) + r_s(t))/2$ decreased in response to PV activation after a brief transient increase, resulting in a paradoxical effect whereby the activation of inhibitory neurons (PV) yielded a sustained decrease in inhibition.³⁵ Intuitively, this decrease is explained by the marked reduction in SOM activity following stimulation, accompanied by a comparably weaker increase in PV.

These results were compared to simulations of a non-MSN network ($\Delta_E = 0.8$, $\Delta_I = 2$), where inhibition was not required to stabilize excitation [Fig. 2(a)]. In this network, the activation of PV neurons resulted in a steady state increase in mean inhibition [Fig. 4(b)]. This non-paradoxical effect was caused by a smaller reduction in SOM activity compared to the MSN state. Thus, a

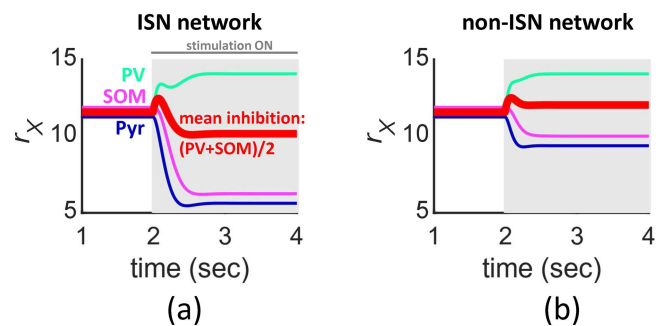


FIG. 4. Paradoxical decrease in inhibition following PV stimulation in MSN (a) but not in non-MSN (b) networks.

defining feature of ISN networks, in both our results and related work,^{35,40,61} was a paradoxical decrease in inhibition following PV activation. Conversely, PYR suppression was common to both ISN and non-ISN networks^{19,65} due to the propagation of feedforward inhibition from PV neurons.

C. Steady state activity and gain of inhibition

The above observations were formalized by calculating the steady states ($dr_X/dt = 0$) of PV, SOM, and PYR neurons with $\vartheta_{noise} = 0$ and applying a PV stimulation such that $v_E = 0$, $v_S = 0$, $v_P > 0$,

$$\begin{aligned} r_P^* &= \frac{v_P(\Delta_I w^2 - \Delta_E w + 1) + \mu(w-1)(w(\Delta_E - 1) - 1)}{w^2(2\Delta_I - \Delta_E) - \Delta_E w + w + 1}, \\ r_S^* &= \frac{\mu(-2\Delta_I w + w + 1) - \Delta_I v_P w}{w^2(2\Delta_I - \Delta_E) - \Delta_E w + w + 1}, \\ r_E^* &= \frac{\Delta_I v_P w^2 + \mu(w+1)(w(\Delta_E - 1) - 1)}{w^2(\Delta_E - 2\Delta_I) + w(\Delta_E - 1) - 1}, \end{aligned} \quad (8)$$

where non-trivial solutions are admitted when $\Delta_I + 1 \neq \Delta_E$. These expressions provide the firing rates of PV, SOM, and PYR neurons when they are not impacted by transient fluctuations.

To examine the change in inhibition resulting from stimulation, a measure Δr_{inhib} was computed as the difference between steady state mean inhibition, $(r_P^* + r_S^*)/2$, before and after PV stimulation. Across a range of Δ_E and Δ_I values, regions with $\Delta r_{inhib} < 0$ indicated a decrease in inhibition consistent with an ISN regime (Fig. 5). A narrow band of negative Δr_{inhib} values showed that paradoxical responses were restricted to a specific range of excitation and inhibition. In other words, paradoxical responses did not span the entire stable region of the parameter space but were confined to a combination of Δ_E and Δ_I values. This result is consistent with a continuous transition between ISN and non-ISN regimes, as reported in both experimental⁴² and theoretical⁶⁶ work. In this respect, the analytical results shown above (Fig. 2) report a discontinuous boundary that approximates the underlying continuous transition between the two regimes. While previous theoretical work has reported paradoxical responses in ISN networks,³⁵ our results demonstrate that only a delimited portion of the state space is associated with such responses.

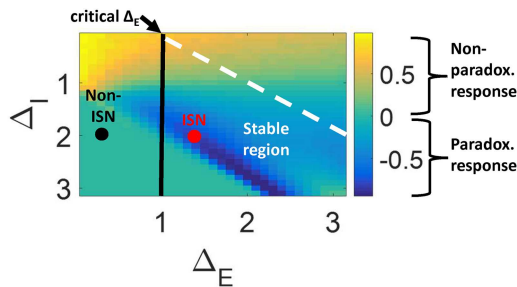


FIG. 5. Change in steady state firing rate of mean inhibition (Δr_{inhib}) induced by PV stimulation.

Formally, the change in inhibition relative to an external stimulation to PV neurons is termed the “gain” of inhibition⁶⁶ and is given by

$$\begin{aligned} h_{inhib}[v_P] &= \frac{\partial(r_P^* + r_S^*)}{\partial v_P} \\ &= \frac{\Delta_E w - 1}{w^2(\Delta_E - 2\Delta_I) + (\Delta_E - 1)w - 1}. \end{aligned} \quad (9)$$

In the limit of Eq. (9) as $\Delta_I \rightarrow \infty$,

$$\lim_{\Delta_I \rightarrow \infty} \frac{\partial(r_P^* + r_S^*)}{\partial v_P} = 0,$$

which can be expressed as a Laurent series,

$$\begin{aligned} \lim_{\Delta_I \rightarrow \infty} \frac{\partial(r_P^* + r_S^*)}{\partial v_P} &= \frac{1 - \Delta_E w}{2\Delta_I w^2} - \frac{(w+1)(\Delta_E w - 1)^2}{4\Delta_I^2 w^4} \\ &\quad - \dots + \mathcal{O}\left(\frac{1}{\Delta_I^6}\right). \end{aligned} \quad (10)$$

As the strength of inhibition increases, the gain of PV neurons approaches zero for both ISN and non-ISN networks. However, for intermediate values of Δ_I , the negative gain of ISN networks ensures that global inhibition decreases under PV activation; the opposite effect is found for non-ISN networks [Figs. 6(a) and 6(b)]. This explains the paradoxical effect of stimulating PV neurons observed specifically in the ISN regime.

By comparison, the gain of PYR neurons is given by

$$\begin{aligned} h_E[v_P] &= \frac{\partial r_E^*}{\partial v_P} \\ &= -\frac{\Delta_I w}{w^2(2\Delta_I - \Delta_E) - \Delta_E w + w + 1}, \end{aligned} \quad (11)$$

with limit

$$\lim_{\Delta_I \rightarrow \infty} \frac{\partial r_E^*}{\partial v_P} = -\frac{1}{2w}. \quad (12)$$

The negative gain of PYR neurons shows that PV stimulation reduces the activity of these neurons in both the ISN and non-ISN states.^{19,65}

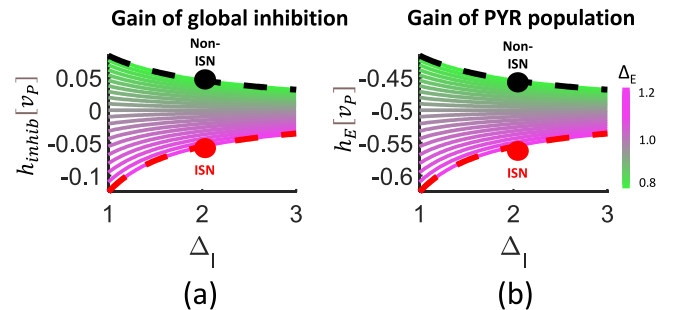


FIG. 6. Gain of global (PV + SOM) inhibition (a) and PYR neurons (b) under PV stimulation.

D. Integrate-and-fire model

The paradoxical effect obtained with the mean-rate model was replicated in a more complex integrate-and-fire network where sub-populations PV, PYR, and SOM were each simulated with M interconnected neurons whose membrane potential evolved according to

$$\tau \frac{dV_i(t)}{dt} = -\ell(V_i(t) - E) + R \left(\sum_{j \in M} g_{ij} k_j(t) + I_i^{STIM}(t) + I^{TONIC} \right) + \xi_i(t), \quad (13)$$

where ℓ is a leak conductance, E is a reversal potential, R is the membrane resistance, $g_{ij} \in \mathbf{G}$ is the connectivity matrix, I^{TONIC} is a constant current, $I_i^{STIM}(t)$ is an input current, τ is the membrane time constant, and $\xi_i(t)$ is zero-mean Gaussian noise with unit variance. The postsynaptic potential of a neuron j is

$$k_j(t) = U \cdot \sum_{s \in S} \exp\left(\frac{t_{j,s}^{spike} - t}{\tau_{fall}}\right) - \exp\left(\frac{t_{j,s}^{spike} - t}{\tau_{rise}}\right), \quad (14)$$

where $t_{j,s}^{spike}$ is the time of a spike $s \in S$ with S denoting all spikes emitted up to time t . The time constants of the rise and fall of postsynaptic potential are τ_{rise} and τ_{fall} and U is a scaling constant. The coupling matrix $\mathbf{G}$ was designed such that connections between neuronal types followed Eq. (3) and connectivity within each type was drawn at random with a probability of 0.5.

Numerical simulations of the integrate-and-fire model were performed with either strong (ISN) or weak (non-ISN) connectivity between PYR neurons. With strong connectivity, the model yielded a paradoxical decrease in mean inhibition in response to PV stimulation [Fig. 7(a)]. With weaker PYR-to-PYR connectivity, PV stimulation yielded an increase in mean inhibition and, hence, no paradoxical response [Fig. 7(b)]. In response to stimulation, PV neurons increased more dramatically in the non-ISN regime compared to ISN, where PV activity was closer to saturation prior to the stimulation. A broad parameter exploration of the model revealed that a narrow region of inhibition (Δ_I) and excitation (Δ_E) yielded $\Delta r_{inhib} < 0$ associated with a paradoxical response (Fig. 8). Thus,

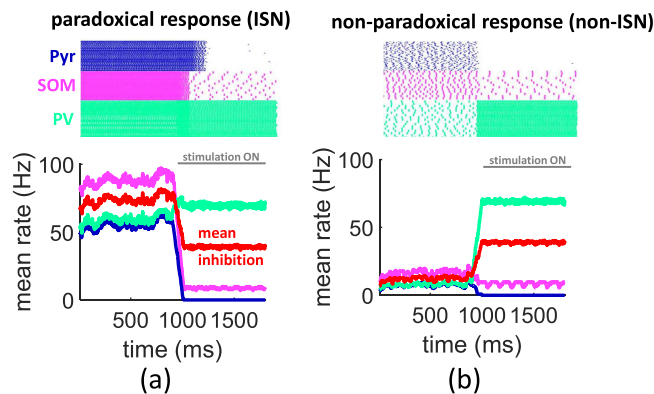


FIG. 7. Response of the integrate-and-fire network in ISN (a) and non-ISN (b) states to PV stimulation.

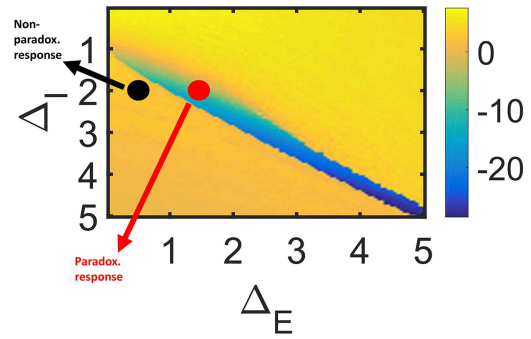


FIG. 8. Change in mean inhibition (Δr_{inhib}) induced by PV stimulation in the integrate-and-fire network.

the integrate-and-fire network replicated the responses found analytically in a network of mean-rate units. These results served as a starting point to examine the contribution of inhibitory and excitatory cell types to training a structured reservoir framework with spiking neurons, as examined next.

E. Learning a Lorenz attractor

All integrate-and-fire neurons (PYR, PV, and SOM) of the recurrent network were connected to output weights that were trained to follow a three-dimensional target signal $\mathbf{Y}$ defined as a chaotic Lorenz attractor,

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y - x), \\ \frac{dy}{dt} &= x(\rho - z) - y, \\ \frac{dz}{dt} &= xy - \beta z, \end{aligned} \quad (15)$$

with parameters $\rho = 28$, $\sigma = 10$, and $\beta = 8/3$ [Fig. 9(a)], dropping the dependence of x, y, z on time t for conciseness. The input to the network was $I^{STIM} = \mathbf{Z}\mathbf{Y}$, where $\mathbf{Z} \in \mathbb{R}^{M \times N}$ was a zero-mean random matrix of i.i.d. elements and $\mathbf{Y} \in \mathbb{R}^{N \times T}$ given M recurrent units and $N = 3$ dimensions from the Lorenz attractor. Output weights $\mathbf{W}_{OUT}$ were adjusted according to a minimum norm least squares rule,⁵¹

$$\mathbf{W}_{OUT} = (\mathbf{K}')^* \mathbf{Y}', \quad (16)$$

where $\mathbf{K}$ is a matrix of postsynaptic potentials [Eq. (14)] from all neurons in the network and $(\cdot)^*$ is the Moore–Penrose inverse operator. Once the output weights were adjusted, the activity of the output layer was computed by

$$\mathbf{B} = (\mathbf{K}' \mathbf{W}_{OUT})'. \quad (17)$$

This rule is not meant to reflect biological mechanisms but makes it possible to examine what network structure and dynamics in the model best support the task performed. The performance of the network was assessed by computing the pairwise correlation between the output of the network $\mathbf{B}$ and target signal $\mathbf{Y}$.

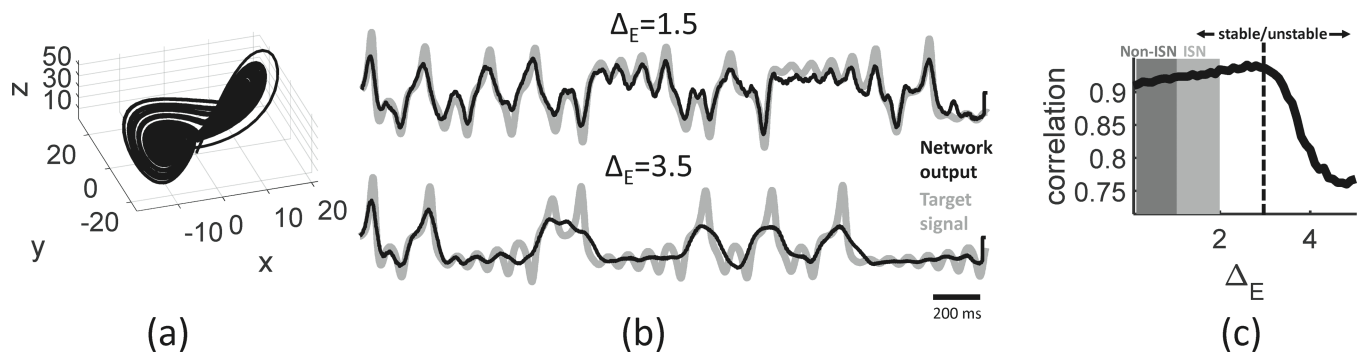


FIG. 9. Performance of the structured reservoir model on Lorenz time series. (a) 3D Lorenz attractor. (b) Network output after training (smoothed using a 20 ms rolling window). The target signal (x) is compared to the corresponding network output. (c) Correlation between the network output and the target signal. ISN and non-ISM regions of Δ_E were identified based on paradoxical responses [Fig. 8]. Stable/unstable regimes were identified based on mean-rate analysis [Eqs. (4) and (5)].

A network in the ISN state ($\Delta_E = 1.5, \Delta_I = 2$) with $M = 500$ spiking neurons replicated the target signal [Fig. 9(b), top]. However, increasing recurrent excitation to an unstable regime ($\Delta_E = 3.5$) reduced performance [Fig. 9(b), bottom]. In a series of simulations, ISN and non-ISM regimes were compared by altering the strength of excitation (Δ_E) while keeping inhibition fixed. Performance increased gradually as Δ_E increased from a non-ISM to an ISN regime and declined sharply with $\Delta_E > 3$, where the network became unstable [Fig. 9(c)]. Thus, the network was robust to changes in the regime of activity so long as the activity operated under a stable state.

After training, output weights and initial conditions of the network were frozen, and the network was rerun to assess its ability to replay the Lorenz time series. An example of network output in the ISN regime ($\Delta_E = 1.5$) is shown in Fig. 10(a). The network was able to replay the Lorenz attractor with high precision given that initial conditions and network weights were the same as during training. However, the network was sensitive to a time-delimited

perturbation ($I_i^{STIM}(t) = 10$ nA) applied as a constant input lasting 10 ms to PV neurons during replay [Fig. 10(a)].

To further examine this effect, we numerically generated distributions of firing rate by pooling all neurons (bins of 0.1 Hz) for unperturbed (P) and perturbed (Q) outputs of the network over 10 000 ms and computed the Kullback–Leibler divergence across firing rates (r),

$$D_{KL}(P \parallel Q) = \sum_{r \in T} P(r) \log \left(\frac{P(r)}{Q(r)} \right), \quad (18)$$

where T is the total number of bins. Divergence was low for non-ISM networks but increased in the ISN state as the network approached instability ($\Delta_E = 3$)^{7,8,10} [Fig. 10(b)]. A similar effect of perturbations was observed for PV, PYR, and SOM neurons. In ISN networks, small time-delimited perturbations have been reported to yield a long-lasting impact on neural performance during learning and memory.^{67–69} This was confirmed in our simulations by

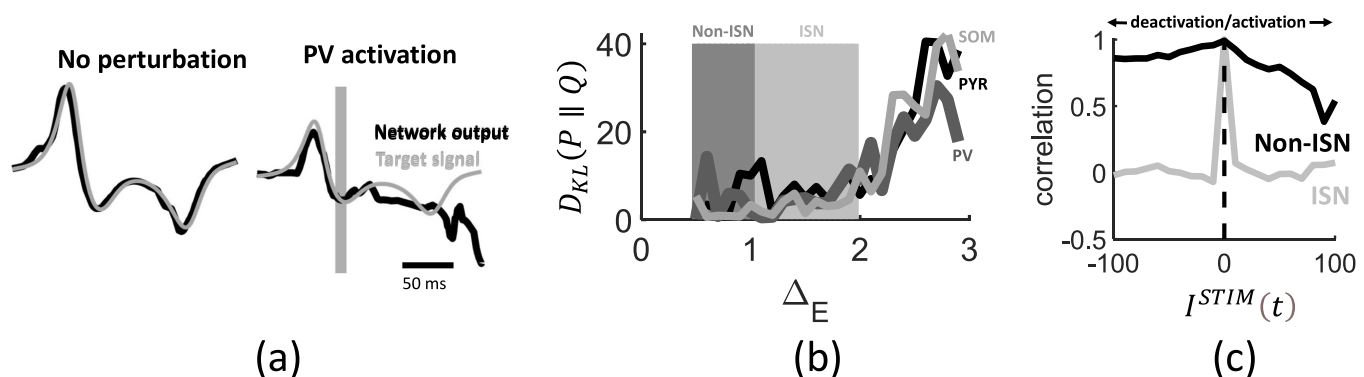


FIG. 10. (a) Time-delimited perturbation of PV, SOM, and PYR neurons. The shaded area indicates the time of perturbation. (b) Kullback–Leibler divergence [Eq. (18)] comparing the distribution of unperturbed and perturbed network outputs. The impact of perturbations on network performance (c) depends on the intensity of the stimulus delivered to neurons ($I^{STIM}(t) = [-100, 100]$ nA) as well as the regime of activity.

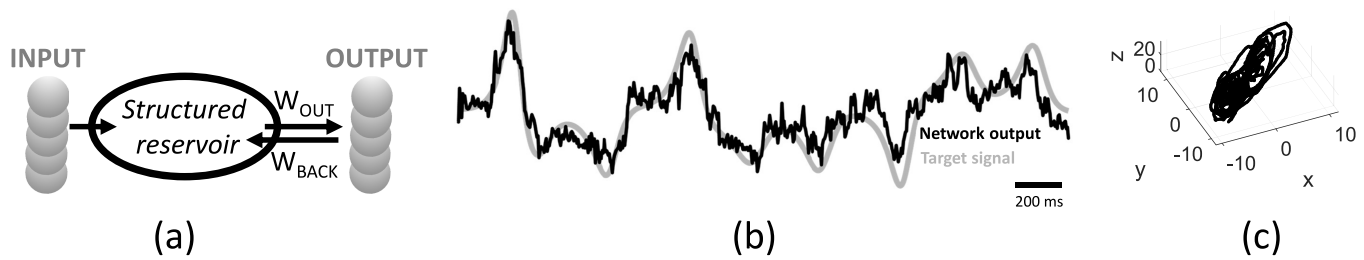


FIG. 11. A network with recurrent connections from the output to the reservoir (a) generated autonomous activity that approximated the Lorenz dynamics (b) and state space (c).

comparing the performance of ISN and non-ISN networks while varying the amplitude of the perturbation [Fig. 10(c)]. The performance of ISN networks dropped precipitously with PV perturbation, while non-ISN networks were comparatively more robust.

To examine the ability of the network to autonomously produce a Lorenz series, we designed a model where the output was fed back to the reservoir via a set of fixed random connections $\mathbf{W}_{\text{BACK}} \in \mathbb{R}^{M \times N}$ [Fig. 11(a)]. The training phase proceeded as above [Eqs. (16) and (17)]. Afterward, an autonomous phase involved feeding the output activation back to the reservoir,

$$\mathbf{I}^{\text{STIM}} = \mathbf{W}_{\text{BACK}} \mathbf{B}. \quad (19)$$

Then, reservoir activity was generated according to Eq. (13) and weights were adjusted autonomously by using the output of the network itself instead of the target signal [Eq. (16)],

$$\mathbf{W}_{\text{OUT}} = (\mathbf{K}')^* \mathbf{B}', \quad (20)$$

and an autonomous output was obtained as

$$\mathbf{B}_{\text{AUTO}} = (\mathbf{K}' \mathbf{W}_{\text{OUT}})'. \quad (21)$$

The correlation between the target sequence and the autonomous output ($\mathbf{B}_{\text{AUTO}}$) was high (0.94) [Fig. 11(b)] and the network generated an approximation of the Lorenz strange attractor in state space [Fig. 11(c)]. This is expected because the output of the network obtained after training ($\mathbf{B}$) was close to the desired target and, therefore, functioned as a reasonable surrogate during the autonomous phase. Alternative strategies exist for generating autonomous reservoirs, for instance, by training the network to output on each given time step a prediction of the input on the next step and feeding this prediction back to the network.¹ This would require an iterative learning rule, as opposed to the batch approach employed here. Such a strategy will be worth exploring in future versions of this framework.

In sum, an ISN regime close to instability was beneficial to learning a Lorenz series; however, this regime was highly sensitive to short PV perturbations delivered during the recall phase. Thus, while dynamical regimes with multiple neuronal types can support reservoir computing, proximity to instability is a limiting factor for both performance and recall accuracy of complex time series.

III. DISCUSSION

In this work, a novel framework was developed to perform reservoir computing with multiple cell types found in the cerebral cortex and their pattern of connections. Stability analysis showed that the model exhibited two distinct regimes of activity that depend on the relative strength of excitation and inhibition. An ISN regime was characterized by strong excitation that was stabilized by inhibition, while a non-ISN regime had weak excitation that did not require inhibition. When trained on a task where the network learned to relay and generate a Lorenz series, accuracy increased as the network transitioned from a non-ISN to an ISN regime but declined sharply in an unstable regime with strong excitation. Because of its proximity to instability, perturbations in the ISN regime resulted in a divergence of dynamics that hindered the recall performance of the learned task.

This work builds on previous multi-class neural networks^{20,34–36,70,71} by incorporating multiple neuronal types in a reservoir computing framework. This approach complements and extends optimization models whose architecture is inspired by brain connectomes.⁷² An advantage of this approach is that it allows one to directly control the regime of the recurrent network through tunable parameters (Δ_E and Δ_I). As results show, controlling the state of the network by balancing excitation and inhibition reveals a trade-off between the higher performance of ISN networks and the more robust behavior of non-ISN networks in the presence of perturbations. These results add to recent work showing the benefits of an ISN regime for improving response gain at the cost of compromised network stability.⁶⁶ The finding that reservoir computing attains optimal performance at the edge of stability is supported by recent work⁷³ showing that in many instances, the transition to dynamical instability (and not necessarily chaos⁷⁴) impacts training accuracy.

We speculate that the cerebral cortex may adjust excitation and inhibition to take advantage of higher training accuracy in the ISN regime but maintain a safe margin away from an unstable regime where performance is diminished. Accordingly, transitions between regimes are a central feature of neural activity and underlie the brain's response during selective attention,^{75–78} arousal,^{79–81} and sensory stimulation.⁸²

Conflicting experimental evidence suggests that the activation of PV neurons leads to either impaired task performance,^{19,83} improved processing,⁸⁴ or no change.⁸⁵ Further, PV suppression is linked to both improvements⁸⁶ and impairments in neural

processing.^{85,87} To reconcile these results, our model suggests that the effect of PV perturbation is state-dependent [Fig. 10(c)]. The ISN regime is highly susceptible to perturbations, and both activation and suppression of PV neurons can lead to impairments in processing.⁸⁸ Further work will be needed to elucidate the state-dependent activity of recurrent networks in response to the activation of specific neural subtypes.

The proposed model was simplified in many ways that could be explored in future implementations. For instance, all connections between the various cell types had the same value, whereas experiments have revealed differences in the strength of connections across cell types as well as cortical layers.⁸⁹ Further, the model did not incorporate plasticity between the various neural types.⁹⁰ Caution is warranted, however, when applying plasticity within reservoir networks, as it may lead to instability when learning an output function.⁸

While our model captured the paradoxical reduction in inhibition associated with the ISN regime (Fig. 4), the potential consequences of this effect on computation remain to be elucidated. Despite the computational advantages of an ISN regime highlighted here and elsewhere,³⁵ a specific advantage of reducing global inhibition has yet to be explored. Speculatively, this reduction may allow for a greater dynamical range of responses by pushing SOM neurons away from a state of high activity that is near saturation [Fig. 7(a)].

In sum, the structured reservoir model we described took advantage of a variety of cell types to capture known regimes of activity in the cerebral cortex and learned to produce chaotic attractors when output weights were trained. This model is uniquely disposed to address questions on the computational role of connectivity, cellular diversity, and operating regimes of the brain in tasks that involve complex sequences of neural activity.

ACKNOWLEDGMENTS

This work was supported by a Discovery grant to J.-P.T. from the Natural Sciences and Engineering Council of Canada (NSERC Grant No. 210977).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

J. P. Thivierge: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Funding acquisition (equal); Investigation (equal); Methodology (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Eloïse Giraud:** Conceptualization (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – review & editing (equal). **Michael Lynn:** Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Writing – review

& editing (equal). **Annie Theberge Charbonneau:** Conceptualization (equal); Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Writing – original draft (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: SIMULATION PARAMETERS

Default parameters for the mean-rate model were as follows: $\tau = 10$ ms, $\vartheta_{noise} = 0.1$, $\mu = 10$, and $\delta = 1$. Parameters for the spiking network were as follows: $\ell = 0.01$ pS, $E = -65$ mV, $I^{TONIC} = 4$ nA for PV and PYR neurons, $I^{TONIC} = -2$ nA for SOM neurons, $\tau = 10$ ms, $R = 10$ M Ω , $U = 0.4$ nA, $\tau_E^{rise} = 3$ ms, $\tau_E^{fall} = 40$ ms, $\tau_I^{rise} = 1$ ms, and $\tau_I^{fall} = 5$ ms. Spikes were triggered when the membrane potential [Eq. (13)] of a neuron reached -15 mV from a value below the threshold. At that time, the potential was set to 100 mV for 1 ms and then reset to -65 mV for an absolute refractory period lasting 3 ms. PV stimulation (Fig. 7) was delivered using $I^{STIM}(t) = 60$ nA.

The coupling matrix $\mathbf{G}$ was set such that 30% of neurons were inhibitory and 50% of all possible weights were randomly set to zero. Non-zero weights included self-connections and were drawn from a zero-mean Gaussian distribution and scaled by $\sqrt{M}$. The spiking and mean-rate models were simulated using a forward Euler method with a time resolution of 0.1 ms. The number of neurons in the recurrent network followed a ratio of 0.8:1 for SOM:PV cells.

REFERENCES

- D. J. Gauthier, E. Bollt, A. Griffith, and W. A. S. Barbosa, *Nat. Commun.* **12**, 5564 (2021).
- H. Jaeger and H. Haas, *Science* **304**, 78, (2004).
- W. Maass, T. Natschlager, and H. Markram, *Neural Comput.* **14**, 2531 (2002).
- J. Pathak, Z. Lu, B. R. Hunt, M. Girvan, and E. Ott, *Chaos* **27**, 121102 (2017).
- P. R. Vlachas, J. Pathak, B. R. Hunt, T. P. Sapsis, M. Girvan, E. Ott, and P. Koumoutsakos, *Neural Netw.* **126**, 191 (2020).
- S. Bompas, B. Georgeot, and D. Guéry-Odelin, *Chaos* **30**, 113118 (2020).
- R. Laje and D. V. Buonomano, *Nat. Neurosci.* **16**, 925 (2013).
- P. Vincent-Lamarre, G. Lajoie, and J.-P. Thivierge, *J. Comput. Neurosci.* **41**, 305 (2016).
- P. Vincent-Lamarre, M. Calderini, and J.-P. Thivierge, *Front. Comput. Neurosci.* **14**, 78 (2020).
- D. Sussillo and L. F. Abbott, *Neuron* **63**, 544 (2009).
- A. Maes, M. Barahona, and C. Clopath, *PLoS Comput. Biol.* **16**, e1007606 (2020).
- Y. Kawai, J. Park, and M. Asada, *Neural Netw.* **112**, 15 (2019).
- H. Markram, M. Toledo-Rodriguez, Y. Wang, A. Gupta, G. Silberberg, and C. Wu, *Nat. Rev. Neurosci.* **5**, 793 (2004).
- B. Rudy, G. Fishell, S. Lee, and J. Hjerler-Leffler, *Dev. Neurobiol.* **71**, 45 (2011).
- C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, and M. Scanziani, *Nat. Neurosci.* **16**, 1068 (2013).
- R. Tremblay, S. Lee, and B. Rudy, *Neuron* **91**, 260 (2016).
- A. Kepecs and G. Fishell, *Nature* **505**, 318 (2014).
- S. El-Boustani and M. Sur, *Nat. Commun.* **5**, 5689 (2014).
- L. L. Glickfeld, M. H. Histed, and J. H. R. Maunsell, *J. Neurosci.* **33**, 19416 (2013).
- A. K. Moore, A. P. Weible, T. S. Balmer, L. O. Trussell, and M. Wehr, *Neuron* **97**, 1341 (2018).

- ²¹H. Kim, S. Åhrlund-Richter, X. Wang, K. Deisseroth, and M. Carlén, *Cell* **164**, 208 (2016).
- ²²A. T. Lee, S. M. Gee, D. Vogt, T. Patel, J. L. Rubenstein, and V. S. Sohal, *Neuron* **81**, 61 (2014).
- ²³M. Xue, B. V. Atallah, and M. Scanziani, *Nature* **511**, 596 (2014).
- ²⁴H.-J. Pi, B. Hangya, D. Kvitsiani, J. I. Sanders, Z. J. Huang, and A. Kepecs, *Nature* **503**, 521 (2013).
- ²⁵Y. Fu, J. M. Tucciarone, J. S. Espinosa, N. Sheng, D. P. Darcy, R. A. Nicoll, Z. J. Huang, and M. P. Stryker, *Cell* **156**, 1139 (2014).
- ²⁶A. S. Ecker, P. Berens, R. J. Cotton, M. Subramaniam, G. H. Denfield, C. R. Cadwell, S. M. Smirnakis, M. Bethge, and A. S. Tolias, *Neuron* **82**, 235 (2014).
- ²⁷I. Lampl, I. Reichova, and D. Ferster, *Neuron* **22**, 361 (1999).
- ²⁸M. R. Cohen and A. Kohn, *Nat. Neurosci.* **14**, 811 (2011).
- ²⁹A. Y. Y. Tan, Y. Chen, B. Scholl, E. Seidemann, and N. J. Priebe, *Nature* **509**, 226 (2014).
- ³⁰G. Maimon and J. A. Assad, *Neuron* **62**, 426 (2009).
- ³¹M. N. Shadlen and W. T. J. Newsome, *J. Neurosci.* **18**, 3870 (1998).
- ³²L. Hertäg and C. Clopath, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115699119 (2022).
- ³³J. Gwak and J. Kwag, *Chaos* **30**, 053102 (2020).
- ³⁴A. Garcia del Molino, G. R. Yang, J. F. Mejias, and X.-J. Wang, *eLife* **6**, e29742 (2017).
- ³⁵A. Litwin-Kumar, R. Rosenbaum, and B. Doiron, *J. Neurophysiol.* **115**, 1399 (2016).
- ³⁶K. V. Kuchibhotla, J. V. Gill, G. W. Lindsay, E. S. Papadoyannis, R. E. Field, T. A. Hindmarsh Sten, K. D. Miller, and R. C. Froemke, *Nat. Neurosci.* **20**, 62 (2017).
- ³⁷G. R. Yang, J. D. Murray, and X.-J. Wang, *Nat. Commun.* **7**, 12815 (2016).
- ³⁸X.-J. Wang and G. R. Yang, *Curr. Opin. Neurobiol.* **49**, 75 (2018).
- ³⁹C. van Vreeswijk and H. Sompolinsky, *Science* **274**, 1724, (1996).
- ⁴⁰M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton, *J. Neurosci.* **17**, 4382 (1997).
- ⁴¹H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster, *Neuron* **62**, 578 (2009).
- ⁴²A. Sanzeni, A. Bradley, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed, *eLife* **9**, e54875 (2020).
- ⁴³S. Sadeh and C. Clopath, *Nat. Rev. Neurosci.* **22**, 21 (2021).
- ⁴⁴P. E. Latham, B. J. Richmond, P. G. Nelson, and S. Nirenberg, *J. Neurophysiol.* **83**, 808 (2000).
- ⁴⁵J. Veit, R. Hakim, M. P. Jädi, T. J. Sejnowski, and H. Adesnik, *Nat. Neurosci.* **20**, 951 (2017).
- ⁴⁶H. Adesnik, *Neuron* **95**, 1147 (2017).
- ⁴⁷H. K. Kato, S. K. Asinof, and J. S. Isaacson, *Neuron* **95**, 412 (2017).
- ⁴⁸M. Tsodyks and C. Gilbert, *Nature* **431**, 775 (2004).
- ⁴⁹S. Sadeh, R. A. Silver, T. D. Mrsic-Flogel, and D. R. Muir, *J. Neurosci.* **37**, 12050 (2017).
- ⁵⁰D. B. Rubin, S. D. Van Hooser, and K. D. Miller, *Neuron* **85**, 402 (2015).
- ⁵¹M. Boucher-Routhier, B. L. F. Zhang, and J.-P. Thivierge, *Neural Netw.* **144**, 639 (2021).
- ⁵²M. Calderini and J.-P. Thivierge, *J. Math. Neurosci.* **11**, 6 (2021).
- ⁵³M. Calderini, S. Zhang, N. Berberian, and J.-P. Thivierge, *Neural Comput.* **30**, 1573 (2018).
- ⁵⁴N. Berberian, A. MacPherson, E. Giraud, L. Richardson, and J.-P. Thivierge, *J. Neurophysiol.* **117**, 738 (2017).
- ⁵⁵B. V. Atallah, W. Bruns, M. Carandini, and M. Scanziani, *Neuron* **73**, 159 (2012).
- ⁵⁶D. D. Bock, W.-C. A. Lee, A. M. Kerlin, M. L. Andermann, G. Hood, A. W. Wetzel, S. Yurgenson, E. R. Soucy, H. S. Kim, and R. C. Reid, *Nature* **471**, 177 (2011).
- ⁵⁷M.-S. Safari, J. Mirnajafi-Zadeh, H. Hioki, and T. Tsumoto, *Sci. Rep.* **7**, 12764 (2017).
- ⁵⁸M. Kecskés, N. Henn-Mike, A. Agócs-Laboda, S. Szócs, Z. Petykó, and C. Varga, *Commun. Biol.* **3**, 754 (2020).
- ⁵⁹T. Kamigaki and Y. Dan, *Nat. Neurosci.* **20**, 854 (2017).
- ⁶⁰J. C. H. Cottam, S. L. Smith, and M. Häusser, *J. Neurosci.* **33**, 19567 (2013).
- ⁶¹S. Sadeh and C. Clopath, *eLife* **9**, e52757 (2020).
- ⁶²N. Li, S. Chen, Z. V. Guo, H. Chen, Y. Huo, H. K. Inagaki, G. Chen, C. Davis, D. Hansel, C. Guo, and K. Svoboda, *eLife* **8**, e48622 (2019).
- ⁶³H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, and M. Scanziani, *Nature* **490**, 226 (2012).
- ⁶⁴A. Mahrach, G. Chen, N. Li, C. van Vreeswijk, and D. Hansel, *eLife* **9**, e49967 (2020).
- ⁶⁵D. R. Sparta, N. Hovelsø, A. O. Mason, P. A. Kantak, R. L. Ung, H. K. Decot, and G. D. Stuber, *J. Neurosci.* **34**, 3699 (2014).
- ⁶⁶H. Bos, A.-M. Oswald, and B. Doiron, *BioRxiv* (2020).
- ⁶⁷J. A. Cardin, M. Carlén, K. Meletis, U. Knoblich, F. Zhang, K. Deisseroth, L.-H. Tsai, and C. I. Moore, *Nature* **459**, 663 (2009).
- ⁶⁸V. S. Sohal and J. L. R. Rubenstein, *Mol. Psychiatry* **24**, 1248 (2019).
- ⁶⁹G. Buzsáki, *Hippocampus* **25**, 1073 (2015).
- ⁷⁰M. Ter Wal and P. H. E. Tiesinga, *Biol. Cybern.* **115**, 487 (2021).
- ⁷¹L. Hertäg and H. Sprekeler, *eLife* **9**, e57541 (2020).
- ⁷²G. Yan, P. E. Vértés, E. K. Towson, Y. L. Chew, D. S. Walker, W. R. Schafer, and A.-L. Barabási, *Nature* **550**, 519 (2017).
- ⁷³T. L. Carroll, *Chaos* **30**, 121109 (2020).
- ⁷⁴N. Bertschinger and T. Natschlager, *Neural Comput.* **16**, 1413 (2004).
- ⁷⁵M. R. Cohen and J. H. R. Maunsell, *Nat. Neurosci.* **12**, 1594 (2009).
- ⁷⁶J. F. Mitchell, K. A. Sundberg, and J. H. Reynolds, *Neuron* **63**, 879 (2009).
- ⁷⁷T. A. Engel, N. A. Steinmetz, M. A. Gieselmann, A. Thiele, T. Moore, and K. Boahen, *Science* **354**, 1140, (2016).
- ⁷⁸G. H. Denfield, A. S. Ecker, T. J. Shinn, M. Bethge, and A. S. Tolias, *Nat. Commun.* **9**, 2654 (2018).
- ⁷⁹J. F. A. Poulet and C. C. H. Petersen, *Nature* **454**, 881 (2008).
- ⁸⁰M. J. McGinley, M. Vinck, J. Reimer, R. Batista-Brito, E. Zagha, C. R. Cadwell, A. S. Tolias, J. A. Cardin, and D. A. McCormick, *Neuron* **87**, 1143 (2015).
- ⁸¹K. A. Ferguson and J. A. Cardin, *Nat. Rev. Neurosci.* **21**, 80 (2020).
- ⁸²M. A. Smith and A. J. Kohn, *J. Neurosci.* **28**, 12591 (2008).
- ⁸³J. J. Cone, M. D. Scantlen, M. H. Histed, and J. H. R. Maunsell, *eNeuro* **6**, ENEURO.0337 (2019).
- ⁸⁴S.-H. Lee, A. C. Kwan, S. Zhang, V. Phoumthipphavong, J. G. Flannery, S. C. Masmanidis, H. Taniguchi, Z. J. Huang, F. Zhang, E. S. Boyden, K. Deisseroth, and Y. Dan, *Nature* **488**, 379 (2012).
- ⁸⁵E. A. Phillips and A. R. Hasenstaub, *eLife* **5**, e18383 (2016).
- ⁸⁶R. V. Rikhye, M. Yildirim, M. Hu, V. Breton-Provencher, and M. Sur, *J. Neurosci.* **41**, 8761 (2021).
- ⁸⁷Y. Zhu, W. Qiao, K. Liu, H. Zhong, and H. Yao, *Nat. Commun.* **6**, 6802 (2015).
- ⁸⁸A. Kumar, I. Vlachos, A. Aertsen, and C. Boucsein, *Trends Neurosci.* **36**, 579 (2013).
- ⁸⁹L. Campagnola, S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli *et al.*, *Science* **375**, eabj5861 (2022).
- ⁹⁰M. A. Gainey, J. W. Aman, and D. E. Feldman, *J. Neurosci.* **38**, 4749 (2018).

Appendix C

Published paper:

Berberian, N., MacPherson, A., **Giraud, E.**, Richardson, L. & Thivierge, J.-P. (2017). Neuronal pattern separation of motion-relevant input in LIP activity. *Journal of Neurophysiology*, 117(2), 738–755. <https://doi.org/10.1152/jn.00145.2016>

Author Contributions:

N.B., A.M., L.R., and J.-P.T. conception and design of research; N.B., A.M., L.R., and J.-P.T. performed experiments; N.B., A.M., **E.G.**, L.R., and J.-P.T. analyzed data; N.B., A.M., L.R., and J.-P.T. interpreted results of experiments; N.B. and J.-P.T. prepared figures; J.-P.T. drafted manuscript; N.B., A.M., **E.G.**, L.R., and J.-P.T. edited and revised manuscript; N.B., A.M., **E.G.**, L.R., and J.-P.T. approved final version of manuscript.

RESEARCH ARTICLE | *Higher Neural Functions and Behavior*

Neuronal pattern separation of motion-relevant input in LIP activity

Nareg Berberian,¹ Amanda MacPherson,² Eloïse Giraud,¹ Lydia Richardson,¹ and J.-P. Thivierge¹

¹Center for Neural Dynamics and School of Psychology, University of Ottawa, Ottawa, Ontario, Canada; and ²Department of Neuroscience, McGill University, Montréal, Québec, Canada

Submitted 17 February 2016; accepted in final form 18 November 2016

Berberian N, MacPherson A, Giraud E, Richardson L, Thivierge JP. Neuronal pattern separation of motion-relevant input in LIP activity. *J Neurophysiol* 117: 738–755, 2017. First published November 23, 2016; doi:10.1152/jn.00145.2016.—In various regions of the brain, neurons discriminate sensory stimuli by decreasing the similarity between ambiguous input patterns. Here, we examine whether this process of pattern separation may drive the rapid discrimination of visual motion stimuli in the lateral intraparietal area (LIP). Starting with a simple mean-rate population model that captures neuronal activity in LIP, we show that overlapping input patterns can be reformatted dynamically to give rise to separated patterns of neuronal activity. The population model predicts that a key ingredient of pattern separation is the presence of heterogeneity in the response of individual units. Furthermore, the model proposes that pattern separation relies on heterogeneity in the temporal dynamics of neural activity and not merely in the mean firing rates of individual neurons over time. We confirm these predictions in recordings of macaque LIP neurons and show that the accuracy of pattern separation is a strong predictor of behavioral performance. Overall, results propose that LIP relies on neuronal pattern separation to facilitate decision-relevant discrimination of sensory stimuli.

NEW & NOTEWORTHY A new hypothesis is proposed on the role of the lateral intraparietal (LIP) region of cortex during rapid decision making. This hypothesis suggests that LIP alters the representation of ambiguous inputs to reduce their overlap, thus improving sensory discrimination. A combination of computational modeling, theoretical analysis, and electrophysiological data shows that the pattern separation hypothesis links neural activity to behavior and offers novel predictions on the role of LIP during sensory discrimination.

pattern separation; decision making; computational model; lateral intraparietal cortex

HOW DOES THE BRAIN EXTRACT information from its environment to guide behavior? One mechanism by which brain signals categorize stimuli is a process termed pattern separation. In brain circuits, neuronal pattern separation is proposed to decrease the similarity between input patterns and allow the disambiguation of sensory stimuli (Santoro 2013; Yassa and Stark 2011). Pattern separation is reported in early sensory areas including the olfactory bulb (Gschwend et al. 2015) as well as in CA3 hippocampus (Leutgeb et al. 2007), dentate gyrus (McHugh et al. 2007), and piriform cortex (Chapuis and Wilson 2012). However, there is no a priori reason to believe

that it is unique to these regions. Indeed, pattern separation is a measurable property of any neuronal network where signal correlations are lower than input correlations (Santoro 2013). Here, signal correlations refer to the correlation of mean spike rates for pairs of neurons across several distinct stimuli (Cohen and Kohn 2011). Input correlations, on the other hand, refer to the degree of similarity between distinct stimuli. Here, these stimuli represent either the visual displays presented to monkeys (in the case of experimental data) or inputs to the model (in the case of simulated data).

Neuronal pattern separation has not been directly examined in the lateral intraparietal area (LIP), a region known for its involvement in the discrimination of visual motion information (Shadlen and Kiani 2013). Visual discrimination is a process that requires the disambiguation of incoming stimuli and is known to be improved by pattern separation (Gschwend et al. 2015). In one task of motion discrimination involving rapid decision making, rhesus monkeys viewed a display of dots moving in two opposite directions and performed a saccade to a target indicating where a majority of dots were moving (Gold and Shadlen 2000). Neurons in the LIP region show firing rates that correlate with decision making during this task. Cells that fire preferentially in the direction of the correct decision show a gradual increase in firing rate preceding the saccade. Furthermore, the speed of this accumulation is related to the strength of evidence in favor of the target decision (Roitman and Shadlen 2002).

According to the pattern separation hypothesis, performance in this task is enhanced when signal correlations in LIP are lower than input correlations. Testing this hypothesis would require an estimate of input correlations obtained from neuronal populations upstream of LIP; as an alternative, computational simulations of mean-rate neurons provide a way to examine the relationship between input and signal correlations. A key feature of these simulations is the heterogeneous nature of the model, where each unit produces a unique pattern of neural activity, in agreement with the idiosyncratic response profiles of LIP units combined with reliable activity across trials (Jun et al. 2010). The presence of neural heterogeneity in LIP suggests rich neural dynamics that go beyond a simple accumulation of population activity (Meister et al. 2013; Park et al. 2014). This heterogeneity is rooted in the complex role of LIP in integrating visual, cognitive, and motor information, a hallmark feature of associative cortex.

In this study, we propose that the heterogeneity of neural responses promotes accurate decision making by allowing cells

Address for reprint requests and other correspondence: J.-P. Thivierge, School of Psychology and Center for Neural Dynamics, Univ. of Ottawa, 136 Jean Jacques Lussier, Vanier Hall, Office 2064, Ottawa, Ontario, Canada K1N 6N5 (e-mail: jthivier@uottawa.ca).

with low signal correlation to combine their responses to discriminate among visual stimuli. We begin by illustrating, through a simple population model of balanced integrator units, how the mean accumulation of population activity observed in LIP recordings can be obtained by summing individual heterogeneous units. Then, we show conditions under which heterogeneous network activity generates signal correlations that are lower than input correlations, in turn leading to accurate stimulus discrimination. Finally, we test several predictions of the population model in recordings of LIP data obtained from macaque monkeys performing a task of dot-motion discrimination (Roitman and Shadlen 2002). Converging evidence from computational, analytical, and experimental analyses suggests that neuronal pattern separation based on heterogeneous activity may underlie the accurate discrimination of visual stimuli.

MATERIALS AND METHODS

Balanced population model. As a starting point, we began with a simple population model of persistent neural activity, a necessary building block of information integration over time (Cain et al. 2013). Although more complex models are available (Chaudhuri et al. 2015), the simplicity of the current approach allows us to perform a detailed parametric exploration as well as formal stability analysis (see APPENDIX) and focus on the specific role of pattern separation.

In the simplest possible scenario, and ignoring all effects of noise, we can represent activity as a neural integrator using a linear mean rate equation (Goldman 2009):

$$\tau \frac{dr_i}{dt} = -\alpha r_i + I_i(t), \quad (1)$$

where r_i is the average activity of unit i , $I_i(t)$ is an external input delivered at time t , $\tau = 0.01$ is a constant of integration, and α is a parameter that controls the leak of the model (set as a default to $\alpha = 1$). Although this linear account is a simplification of underlying biological processes, it nonetheless represents a useful approximation that is amenable to formal analysis (see APPENDIX; Dayan and Abbott 2001; Ganguli et al. 2008; Miri et al. 2011; Murphy and Miller 2009). Furthermore, many subtypes of neurons emit firing rates that are well-represented by a linear function of their input within a given range (Chance et al. 2002).

Although some work has argued that neural accumulation in LIP is more compatible with a perfect integrator than a model with leaky activity (Mazurek et al. 2003), we have nonetheless chosen to include a leak term in the model, by setting the parameter $\alpha > 0$, such that activity decays back to rest in the absence of input. In the case of a perfect integrator ($\alpha = 0$), population activity (described below) becomes unstable. This can be seen when performing a linear stability analysis of the model (see APPENDIX). Furthermore, emerging evidence suggests that motion signals may be processed with a strong leak (Carland et al. 2015; Cisek et al. 2009; Ditterich 2006). Therefore, we opted for a scenario where the leak term insured stability of the neural activity. This version of the model captured experimental data on neural heterogeneity and allowed us to study its role on stimulus discrimination, as detailed below.

In the model, unit r_i receives an external input and acts as a low-pass filter on this input with a constant of integration τ . The time integration of the model (i.e., the time required to induce a change in activity corresponding to $\sim 63.2\%$ of baseline) is 100 ms, in line with neural integrator models of LIP activity (Mazurek et al. 2003). To draw a parallel with experimental data, we defined the preferred (P) units as those for which external inputs correspond to orientation-

selective activation (i.e., normalized activation > 0.01 Hz for a given stimulus orientation) and nonpreferred (NP) units as all others.

This model, however, does not account for the characteristically noisy nature of neural activity. Experimental recordings show that neuronal variability is markedly greater than expected by classic attractor models (Brody et al. 2003; Shafi et al. 2007). In fact, persistent activity during tasks of working memory is typically irregular and Poisson-like (Baeg et al. 2003; Compte et al. 2003). We account for this variability (again, in as simplified a manner as possible) by adding a noise term to Eq. 1:

$$\tau \frac{dr_i}{dt} = -\alpha r_i + I_i(t) + \xi_i(t), \quad (2)$$

where $\xi_i(t)$ is drawn from a Gaussian distribution with mean $\mu = 0$ and standard deviation (SD) $\mu = 0.05$. This added noise term represents both stimulus and neural noise, which have a similar phenomenological effect (Burak and Fiete 2012). It is henceforth referred to as the input noise of the model.

Next, we expanded the above model by considering the dynamics of a synaptically interacting population of neurons:

$$\tau \frac{dr_i}{dt} = -\alpha r_i + I_i(t) + \xi_i(t) + \sum_j w_{ij} r_j(t - \delta_{ij}), \quad (3)$$

where w_{ij} is the connection strength from unit i to unit j and δ_{ij} is a transmission delay set randomly between 1 and 5 ms using a uniform distribution, in agreement with synaptic conductance delays observed in cortex (Swadlow 1985).

Motion-relevant input in the population model. In a well-studied motion discrimination task, dots move in one of two opposite directions on a screen. This visual input is held constant during a viewing phase of ~ 500 ms, then the stimulus is removed for the remainder of the trial, corresponding to a delay phase (Meister et al. 2013). In simulations, we represented visual stimuli coming into LIP in a highly simplified manner. These stimuli were meant to represent a combination of visual input as well as urgency, thus reflecting the integration of multiple sources of information in LIP. Here, urgency refers to the subjective haste of subjects to produce a response over the course of a given trial (Cisek et al. 2009). The purpose of this simplification is to allow us to compute input correlations in a tractable manner, which would be difficult if we considered independent sources of input. In consequence, we do not propose our framework as a complete model of decision-making circuits but rather as a model of one particular computation, namely pattern separation, that may be performed by these circuits.

Inputs in the population model represented firing rates of incoming activation into LIP units. These inputs were dependent on the strength of stimulus evidence as well as elapsed time. This aspect of our simulations is in keeping with experiments in LIP relating the slope of neural accumulation to the strength of visual evidence (Roitman and Shadlen 2002), an aspect that has been replicated in related work through an urgency signal causing a gradual buildup of activity over time originating from extrastriate visual cortex (Thura and Cisek 2014). Other aspects of LIP dynamics, including movement-relevant activation, were not captured here.

In keeping with the idea of divisive normalization in sensory input (Ma et al. 2014), the maximum strength of inputs had a constant sum (set to 20 Hz) across motion coherences. Thus only the relative amplitude of inputs was altered. Coherences employed here matched those of experimental work (0, 3.2, 6.4, 12.8, 25.6, and 51.2%). To obtain the relative proportion (p) of amplitudes, we solved for coherence = $12p - 100\%$. A coherence of 3.2%, for instance, yielded relative amplitudes of 48.4 and 51.6%.

Inputs in the model consisted of Gaussian functions centered at a given stimulus orientation. The amplitude of these functions corresponded to stimulus coherence, whereas their variance controlled the orientation specificity of the stimulus. In this way, a narrow Gaussian

input would activate units that respond only to the specific orientation of the stimulus. By default, Gaussian inputs had a standard deviation of 30 and were centered at either 90° (target) or 270° (opposite orientation). An example of stimulus is shown in Fig. 1B (discussed below).

Across different simulations, we varied the amplitude of Gaussian inputs to capture the motion coherence of the behavioral task. This was achieved by multiplying Gaussian inputs by a constant factor initially set to 0 and increasing at a rate proportional to motion coherence. For 0% coherence, experimental values show a slope of approximately ± 0.01 Hz/ms (Roitman and Shadlen 2002). We used this value as a starting point to compute the slope (s) of other coherences c , $s = 0.01 + 0.01c$ (where we altered the sign of s to ensure that $s > 0$ for P units and $s < 0$ for NP units).

Orientation-selective tuning in the population model. Each unit in the population model was assigned a distinct tuning curve that describes the relationship between stimulus orientation and neuronal response (Fig. 1A). This tuning curve was defined by a Gaussian function with a mean corresponding to the stimulus orientation associated with maximal response:

$$f(\theta) = r_0 + r_{\max} \exp\left[-\frac{1}{2\sigma^2}(\theta - \theta_{\text{pref}})^2\right], \quad (4)$$

where r_0 is the baseline amplitude of tuning curves, $r_{\max}$ is the maximum amplitude, σ^2 is the variance, θ_{pref} is the preferred orientation of the unit, and θ is a stimulus orientation. An advantage of this formulation over wrap-around von Mises functions (Ecker et al. 2011) is that it allows a direct and independent control of the amplitude and variance of individual tuning curves.

We developed both a homogeneous and a heterogeneous version of the population model. In the homogeneous version, tuning-curve parameters (r_0 , $r_{\max}$, and σ^2) were the same across all units, and the preferred orientation of these units (θ_{pref}) was spaced evenly according to $360/N$. In the heterogeneous version, these parameters were chosen from Gaussian distributions (see Table 1). Importantly, the means of these Gaussian distributions were taken directly from the

homogeneous model. We normalized all peak amplitudes to a sum of 1. Tuning curves were designed to wrap around 360° of orientation such that curve segments extending beyond 360° were added to curve segments between 1 and 360°.

Inputs injected into the network were multiplied by the tuning curve of all units. For instance, if the network received a Gaussian stimulus centered at a 90° orientation and 20-Hz amplitude, the tuning curve of each unit would be multiplied by that value, yielding distinct firing rates across a subpopulation of units (Fig. 1B). These firing rates were then stored in $I_i(t)$, corresponding to the input of the population model in Eq. 3. Examples of $I_i(t)$ values for 10 units receiving an input centered at 90° are shown in Fig. 1B.

Connectivity weights of the population model. In the population model, we connected together $N = 1,000$ units using sparse, balanced connectivity (Fig. 1, C and D). This connectivity followed several rules. First, over the whole network, 80% of units were randomly chosen to be excitatory, whereas the remainder 20% were inhibitory. Following Dale's law, a given excitatory/inhibitory unit only had positive/negative outgoing connection weights.

Second, network connectivity was sparse, with only 20% of all possible connections present (chosen randomly among all possible connections). Self-connections were not permitted ($w_{ii} = 0$). This rule is in line with cortical neurons having numerous lateral connections yet few autapses (Braitenberg and Schüz 1998). Related work shows that models with sparse intracortical connectivity offer a robust fit to experimental data in LIP (Ganguli et al. 2008).

Third, we balanced the sum of excitation and inhibition coming into each unit. The result was an approximate cancellation of the sum of synaptic inputs as evidenced in cortical networks (Haider et al. 2006; Shu et al. 2003).

Fourth, connectivity followed a “sunken Mexican hat” whereby units with a similar orientation preference interacted via excitatory connections ($w > 0$) and units with a dissimilar orientation preference interacted via inhibitory connections ($w < 0$). This connectivity was obtained by subtracting one Gaussian function (amplitude: 0.45; SD: 0.38) from another (amplitude: 0.4; SD: 0.43), where the means of

Fig. 1. Linear integrator network of orientation-selective units. **A**: tuning curves for 10 units of the population model (each unit is shown in a different color). The x-axis shows the orientation (in degrees) of the visual stimulus. The y-axis is the normalized (norm.) activation of each unit in response to stimuli of different orientations. **B**: activation of a subpopulation of units receiving a stimulus. Black: Gaussian stimulus centered at 90° of orientation (SD, 20). **C**, solid line: the connectivity between units is dependent on orientation preference and follows a sunken Mexican hat. Filled circles: correlation between pairs of units connected via reciprocal weights $-w$ and w (values of the weights indicated on the y-axis). The stimulus orientation is indicated on the x-axis. **D**: map of connections between 100 units of the population model. The sum of entering synaptic weights for each unit is shown to the right of the map and is close to 0 across all units. Top: sum of outgoing weights for the unit shown by gray arrow.

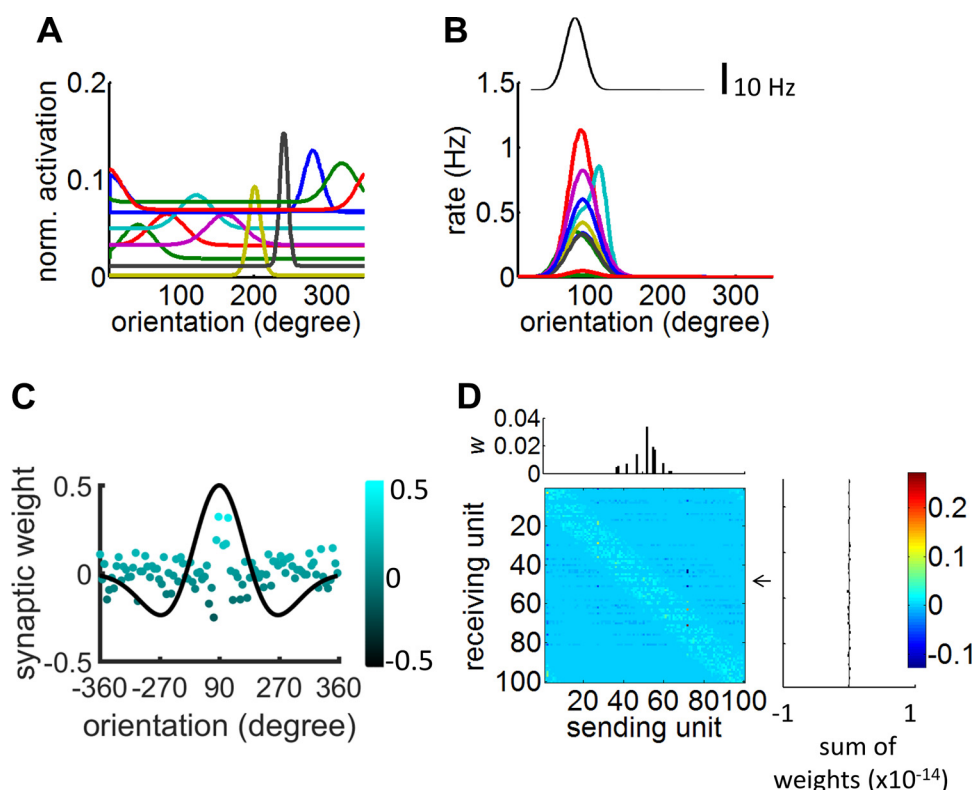


Table 1. *Tuning curve parameters in the homogeneous and heterogeneous mean-rate model*

	Homogeneous Model	Heterogeneous Model
Baseline (r_0)	0.05	$\mu = 0.05, \sigma = 0.02$
Maximum amplitude ($r_{\max}$)	1	$\mu = 1, \sigma = 0.2$
Variance (σ^2)	20	$\mu = 20, \sigma = 5$
Spacing between preferred orientations	360/ N	$\mu = 360/N, \sigma = 2$
Correlation (c_{ij})	0.2	$\mu = 0.2, \sigma = 0.05$

Values denoted by μ and σ are drawn from a Gaussian distribution with given mean and SD.

both functions corresponded to the preferred orientation of each neuron.

Finally, all connection weights were scaled as $1/\sqrt{k}$, where k is the total number of synaptic inputs. The resulting network connectivity was thus a sparse, balanced matrix where P and NP units competed for activation from the input (Fig. 1, C and D).

We performed all simulations using a fourth-order Runge-Kutta integration in the MATLAB programming language (The MathWorks, Natick, MA). Custom code in MATLAB is available from the authors by request.

Fisher linear discriminant analysis. Can the neural activity produced by the above population model be read by downstream neurons to perform stimulus discrimination? We employed a Fisher linear discriminant analysis to address this question. Although a principal components analysis could have been employed here, its ability to extract a linear boundary between two classes is limited to cases where this boundary is aligned with the direction of maximum variance in the data. In the linear discriminant analysis, if a linear boundary is found that can successfully discriminate the neural responses, then a plausible rule for synaptic plasticity such as a linear Hebbian rule could also be designed to extract such a boundary (Buonomano and Maass 2009). Many examples in neuroscience show that a linear readout is sufficient to decode stimulus information based on neural activity (Klompf et al. 2012; Meyers et al. 2008) and select a relevant motor outcome for decision making based on population activity (Nienborg and Cumming 2010; Shadlen et al. 1996).

Let us assume we have a set of N units $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$, where each sample $\mathbf{x}$ is a vector corresponding to the activity of an individual unit over a time window of $T = 1,000$ ms (total trial duration). Here, N_1 of these samples belong to class ω_1 and N_2 to class ω_2 . We wish to obtain a scalar y by projecting the samples x onto a line:

$$y = \mathbf{w}^T \mathbf{x}. \quad (5)$$

Of all the possible lines, we seek the one that maximizes separation between the two classes (McLachlan 2004). To measure how well we are separating the classes, we compute the mean population response to stimulus i for each class:

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^N \mathbf{x}_j. \quad (6)$$

We then apply Fisher linear discriminant as our objective function (Iatan 2010). First, we compute the class covariance:

$$\sum_i = \frac{1}{N_i - 1} \sum_{j=1}^N (\mathbf{x}_j - \mu_j)(\mathbf{x}_j - \mu_j)^T, \quad (7)$$

and the within-class variance,

$$S_W = \sum_{i=1}^c (N_i - 1) \sum_i, \quad (8)$$

where c is the total number of classes. The between-class variance (S_B) is

$$S_B = \sum_{i=1}^c N_i (\mu_i - \bar{\mathbf{x}})(\mu_i - \bar{\mathbf{x}})^T, \quad (9)$$

where $\bar{\mathbf{x}}$ is the grand mean of the data. The Fisher linear discriminant is the linear function $\mathbf{w}^T \mathbf{x}$ that maximizes the criterion function:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}. \quad (10)$$

The closed-form solution is given by:

$$\mathbf{w}^* = S_W^{-1}(\mu_1 - \mu_2). \quad (11)$$

We assessed the performance of the Fisher linear discriminant analysis using a method of 10-fold cross-validation. This method divides the data set to be classified into 10 bins of equal size. The linear discriminant analysis is trained on 9 of the 10 bins and tested on the 10th bin. This procedure is repeated 10 times, each time selecting 9 bins for training and the remaining bin for testing.

Electrophysiology. The data analyzed here are taken from previous work where the experimental procedures are described in detail and were approved by the University of Washington Animal Care Committee (Roitman and Shadlen 2002). Briefly, two rhesus monkeys were trained on a task of dot-motion discrimination. In this task, dots were presented on a visual display and moved in one direction or the opposite. Monkeys indicated their choice by a saccade in the direction where a majority of dots were moving. Different trials varied the orientation of stimuli as well as their motion strength (i.e., the proportion of dots moving in either direction). Spike times were extracted from single-electrode recordings in LIP. Neurons were classified according to their preferred motion, corresponding to the stimulus orientation where firing rate was maximal (P) versus the opposite (nonpreferred) orientation (NP).

LIP data analysis. To reduce the amount of data entered in linear discriminant analysis, we downsampled single-cell activity on a trial-by-trial basis by averaging activity using nonoverlapping windows of 10 ms, from the onset of stimulus motion to 1,000 ms afterward. We then entered a gradually increasing number of randomly chosen cells into a linear discriminant analysis (LDA) classifier. The input to this classifier consisted in a set of N neurons $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$. Each sample $\mathbf{x}$ was a vector corresponding to the activity of an individual neuron recorded over a fixed time window. A portion of the samples corresponded to neuronal activity in the target direction, whereas the remainder of the samples corresponded to activity in the opposite direction (P and NP neurons were identified in the data set). For each cell that was recorded over repeated trials (with same coherence and behavioral decision), we averaged these trials together before performing LDA. An alternative would be to treat individual neurons recorded over several trials as different entries in the linear discriminant analysis. However, given that neurons have similar response profiles across trials (Meister et al. 2013), this could lead to an overestimation of the number of neurons needed to perform discrimination.

RESULTS

This section is structured as follows. First, we describe conditions under which the population model produces heterogeneous responses that capture LIP activity during the dot-motion discrimination task. Then, we examine how these heterogeneous responses may contribute to pattern separation. Finally, we describe the contribution of various aspects of the population model to the accurate discrimination of visual information. Throughout, we test predictions of the model

against data collected in LIP during a visual motion discrimination task.

Pattern separation by heterogeneous neural activity. We consider a population of neural integrators that interact through balanced connectivity. In the model, a balance of excitation and inhibition allows neuronal activity to produce complex yet globally stable activity, as shown below and in related work (Goldman 2009).

At the level of single units, the pattern of activity produced in the population model forms an intricate motif, where the combination of excitatory and inhibitory synapses forces some units to increase (or decrease) their activity in response to the stimulus and produces nonlinear dynamics over time, as described below.

Population activity in the model, when averaged across units and trials (separately for P and NP units), reflected the coherence of the input stimulus. With low coherence, the mean activity of P and NP units (i.e., units responding to stimuli in the target and opposite direction, respectively) remained close in value; with higher coherence, however, mean P activity was higher, as expected by the nature of the stimuli (Fig. 2A). When averaging activity across all six coherences, the activity of P units showed a well-known accumulation over time, whereas the activity of NP units decreased gradually (Fig. 2B). These responses are qualitatively similar to experimental recordings of LIP neurons (Meister et al. 2013).

This simple pattern of accumulation, however, was not present across all single-unit activity, again reflecting experimental findings (Fig. 2C). Rather, single units displayed a wide heterogeneity of responses. Furthermore, these responses were stable across trials, as seen from the limited spread of SE relative to the range of firing rates across P and NP units.

Next, we consider whether neurons can take advantage of this heterogeneous activity when performing stimulus discrimination. Based on the pattern separation hypothesis, stimulus discrimination is enhanced when signal correlations are lower than input correlations. In this way, network activity may reformat highly similar inputs, leading to enhanced discrimination accuracy. We tested this idea in the population model by directly comparing input and signal correlations across all six values of stimulus coherence. To compute input correlations, we took the vector I_i (from $t = 1, \dots, T$) corresponding to the input of each unit and computed its mean pairwise correlation across all units. Similarly, to compute signal correlations, we took the vector r_i (from $t = 1, \dots, T$) corresponding to the activity of each unit and computed its mean pairwise correlation across all units. We employed Pearson correlations throughout. Consistent with pattern separation, we found markedly lower values of signal correlations than input correlations (Fig. 3A, open circles). This result, however, was dependent on the variance of the noise in the input patterns; higher variance reduced input correlations until they became comparable with

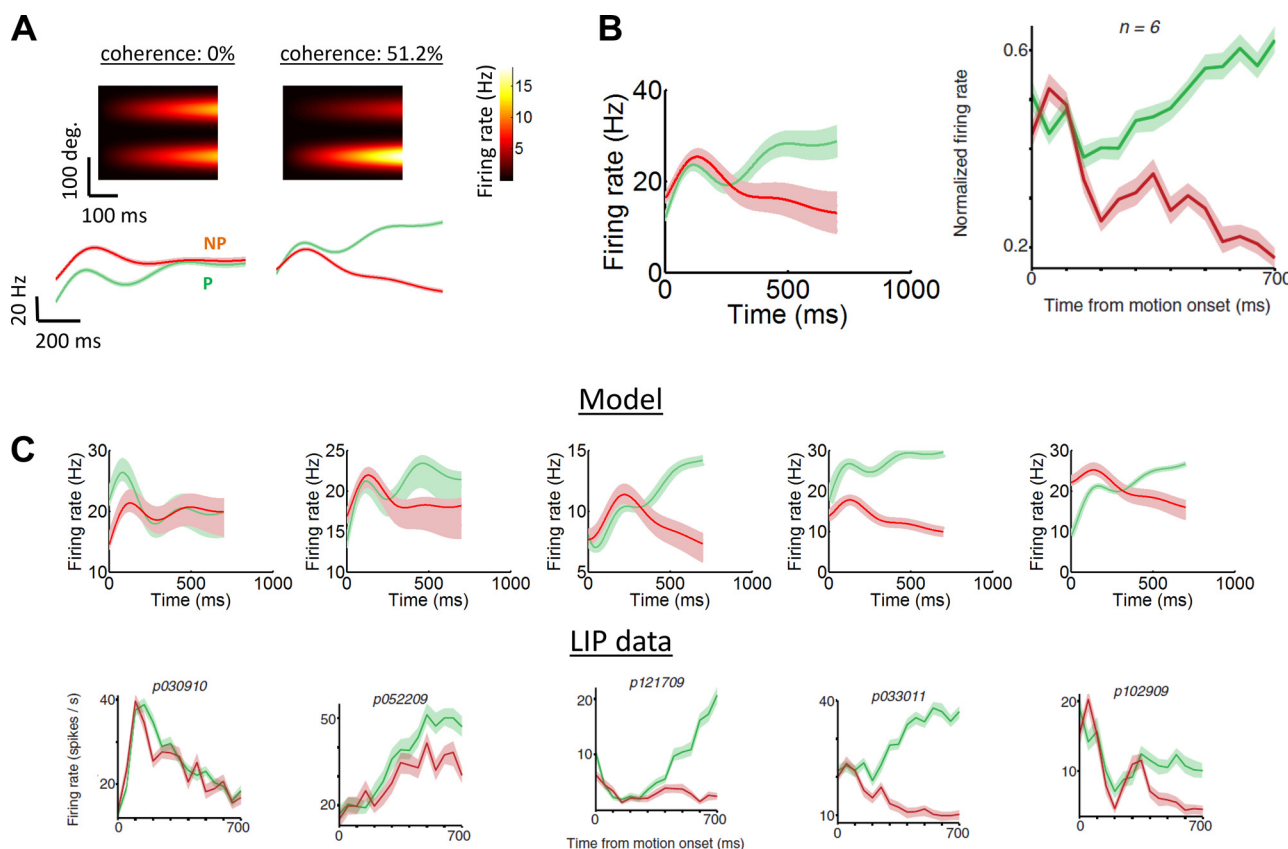


Fig. 2. Neuronal heterogeneity in LIP. **A**, top: examples of input across all units for different levels of coherence. Bottom: mean population activity across units and trials for P (green) and NP units (red) in response to different levels of coherence. deg., Degrees. **B**, left: mean population activity in the population model (Eq. 3) averaged across all 6 levels of coherence. Stimulus onset corresponds to time 0. Right: mean population activity obtained from recordings of LIP neurons (Meister et al. 2013). **C**: examples of single-unit activity in the model (top row) and in LIP data (bottom row). Responses are averaged across all 6 levels of coherence. Within each plot, the same unit is displayed for both preferred and nonpreferred stimuli. Shaded area: standard error of the mean.

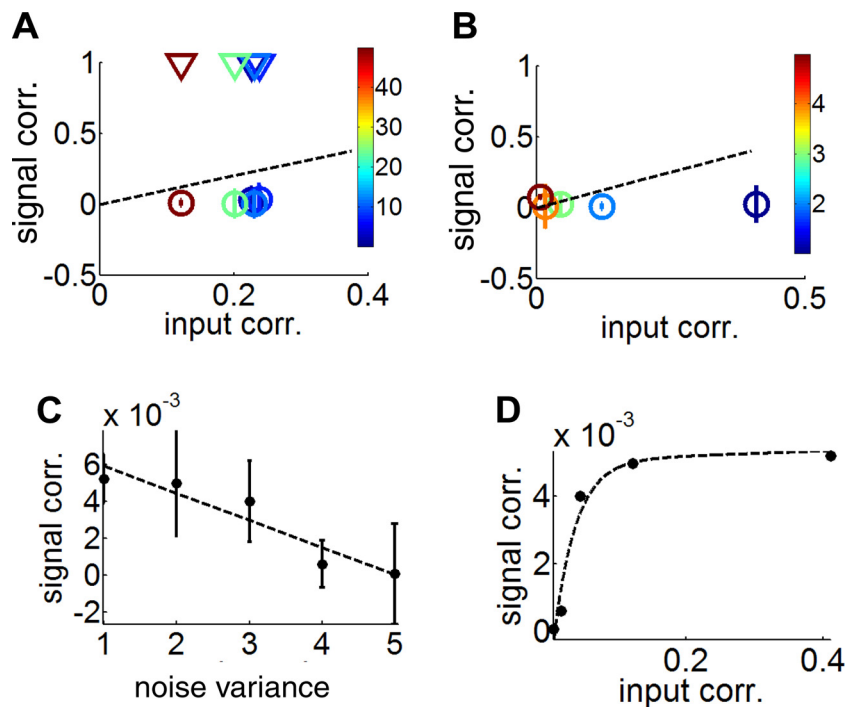


Fig. 3. Neuronal pattern separation. A: relation between input and signal correlations (corr.) in the heterogeneous (circles) and homogeneous model (triangles). Scale bar shows stimulus coherence. Dashed line is unity. Variance of input noise is set to 2.0 (Eq. 3). B: input vs. signal correlations with different values of input noise variance (scale bar). Vertical bars: SE. C: relation between noise variance of the input and signal correlations produced by network activity. Dashed line: best-fitting line of regression. Vertical bar: SE. D: relation between input and signal correlations. Solid black circles are obtained by altering the input noise variance between values of 1 and 5. Dashed line: best-fitting exponential function.

signal correlations (Fig. 3B). At such point, network activity provides no advantage over input patterns in terms of pattern separation.

To investigate further the relationship between input and signal correlation, we performed simulations where we altered the variance of the input noise and examined the impact on signal correlations. We found a correlation between noise variance of the input and signal correlations ($r^2 = 0.89$, $P < 0.02$; Fig. 3C). Thus signal correlations diminished in a linear fashion when noise variance of the input was increased. Furthermore, we examined the relation between input correlations and signal correlations when noise variance was altered. We found an exponentially increasing relation between input and signal correlations, resulting in a least-mean-squares fitting of the function $f(x) = 0.005e^{0.14x} - 0.007e^{-31.49x}$ (Fig. 3D). Signal correlations thus increased rapidly up to a maximum plateau with increasing values of input correlations. In sum, network activity in the population model was sensitive to the statistics of input patterns; overall, however, heterogeneous network activity generated signal correlations that were markedly lower than input correlations (Fig. 3, A–B).

We compared the above results to those obtained with a homogeneous model that eliminated neural heterogeneity by replacing all single-unit activity with mean population activity. This model resulted in markedly higher signal correlations than the heterogeneous model (Fig. 3A, open triangles). Hence, neural heterogeneity was a key factor in driving pattern separation in the model.

A common pitfall when considering the discrimination of input patterns is overfitting, given that network activity separates inputs based on both signal and noise. To address this issue, we turned to a rigorous estimate of discrimination based on Fisher linear discriminant analysis (LDA), in which the goal is to find a decision boundary that yields minimal discrimination error (McLachlan 2004; MATERIALS AND METHODS). To examine whether the population model could discriminate

among trials on the basis of their corresponding target decision, we simulated 1,000 trials of activity in the model using each of 6 coherence values as in related experiments (0, 3.2, 6.4, 12.8, 25.6, and 51.2%). We analyzed the resulting network activity with LDA, starting with 2 randomly chosen units and adding progressively more until all units were included ($N = 1,000$). Incorporating >100 units led to no further reduction in cross-validated discrimination error (Fig. 4A, solid red line). The performance of the population model was markedly superior to that of a homogeneous model where we replaced single-unit activity by mean population activity (Fig. 4A, solid green line). Performance of this homogeneous model was comparable to that of a null model where we replaced all activity by Gaussian noise with same mean and SD as the original data (Fig. 4A, solid black line). The poor performance of the homogeneous model is attributable to high signal correlations, providing similar responses to distinct inputs, as described above (Fig. 3A). Discrimination was also superior when pooling P and NP units together compared with either subpopulation alone.

A question that arises naturally from the above results is how a linear model (Eqs. 1–3) can serve as basis for a linear classification of inputs, given that such classification is known to require a nonlinear model (Vapnik 2000). To address this question, recall that classification relies on the temporal dynamics of activity (see MATERIALS AND METHODS). In the model, this activity evolves in a nonlinear fashion. As an illustration, consider a simplification of Eq. 3 with only two units, x and y , receiving reciprocal weights $[w, -w]$ and a constant input, and ignoring the effect of noise (see APPENDIX):

$$\tau \frac{dx}{dt} = -\alpha x + I + wy(t - \delta), \quad \tau \frac{dy}{dt} = -\alpha y + I - wx(t - \delta). \quad (12)$$

The solution to this equation for $x(t)$ is:

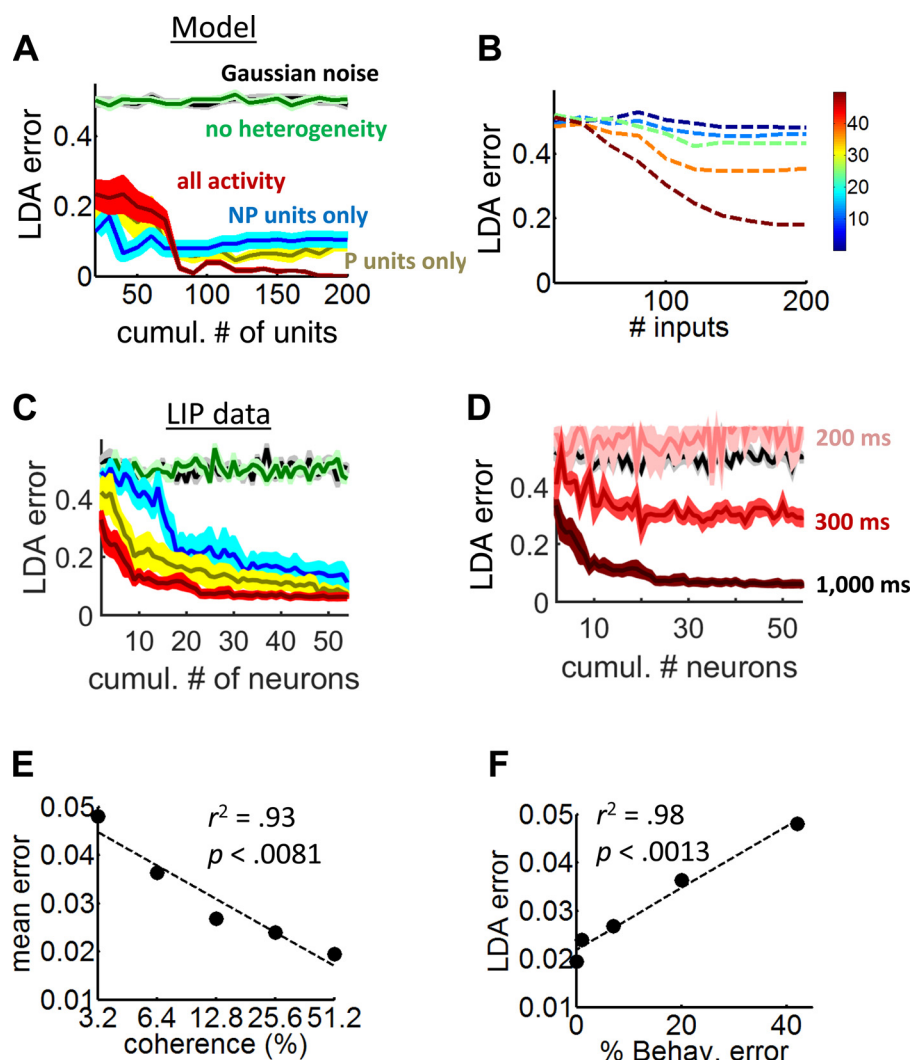


Fig. 4. Discrimination of visual stimuli in the population model and LIP data. *A*: discrimination error in the model. Red: all units. Yellow: P units only. Blue: NP units only. Black: all activity replaced by Gaussian noise with same mean and SD as original activity. Green: replacing the activity of each unit by the mean activity of the population. cumul. #, Cumulative number. *B*: discrimination error with LDA applied directly to the stimuli. Different values of coherence shown in scale bar. *C*: stimulus discrimination with LIP data. Colors are the same as in *A*. *D*: linear discriminant analysis of LIP data using only a 200- or 300-ms time window from stimulus onset. Discrimination error is averaged over all 6 motion coherences. Solid black line: all activity replaced by Gaussian noise with same mean and SD as original activity. *E*: relation between stimulus coherence and mean LDA discrimination error in LIP. Dashed line: best-fitting linear regression. *F*: relation between mean percentage behavioral (Behav.) errors and mean LDA discrimination error.

$$x(t) = \frac{\tau w y}{\alpha^2} + x_0 \exp\left(-\frac{\alpha t}{\tau}\right) - \frac{\delta w y}{\alpha} + \frac{I}{\alpha}, \quad (13)$$

given an initial value x_0 . The term $x_0 \exp(-\alpha t/\tau)$ results in transient nonlinear dynamics of x over time. We address the contribution of temporal dynamics in greater detail below.

To evaluate further the degree to which network activity contributed to stimulus discrimination, we repeated the above LDA using input patterns directly, as opposed to network activity. Discrimination error across all values of input coherence was markedly higher than error obtained from network activity (Fig. 4*B*). In sum, simulations suggest that heterogeneous network activity was a driving force in the discrimination of ambiguous stimuli. Network activity disambiguated input patterns by generating low signal correlations, as consistent with neuronal pattern separation. Next, we examine whether neural heterogeneity plays a role in the accurate discrimination of visual stimuli using experimental data from LIP neurons obtained during a task of motion discrimination (Roitman and Shadlen 2002).

Motion discrimination with LIP activity. The data analyzed here consisted of $N = 54$ neurons recorded over repeated trials with different degrees of stimulus coherence. We began by extracting all single-cell activity (both correct and incorrect

decisions, across all 6 coherences, with both preferred and nonpreferred movement directions). We analyzed these data in a similar fashion as the population model, using LDA. As in the model, the cumulative addition of neurons in LDA resulted in a gradual decrease in discrimination error (Fig. 4*C*). Approximately 10 neurons were needed before error dropped below 20% discrimination error (averaged over all coherences). As predicted by the model, eliminating neural heterogeneity (by replacing all activity with mean population activity) led to chance-level performance. Discrimination accuracy in these homogeneous data was comparable to a null model that replaced all neural activity with Gaussian noise with same grand mean and SD as the original activity. These results highlight the importance of neural heterogeneity for stimulus discrimination. An analysis of specific characteristics of neural heterogeneity (rate vs. temporal code as well as temporal precision) are addressed below and show that temporal fluctuations in neural activity play a key role in stimulus discrimination.

How quickly within each trial can LIP classify visual motion information? We examined this question with a linear discriminant analysis that considered only the 1st 200 or 300 ms of LIP activity after stimulus onset. These smaller time windows led to higher error (Fig. 4*D*), and although above-chance perfor-

mance could be reached with a window of 300 ms, it could not be reached with a window of 200 ms. Therefore, LIP neurons may not need to accumulate sensory evidence for an entire trial duration to guide accurate behavioral decisions but do require at least a few hundred milliseconds to perform visual stimulus discrimination, roughly in line with the time course of rapid decision making based on visual information (DiCarlo et al. 2012).

Mean discrimination error obtained via the above LDA was strongly related to the coherence of visual stimuli [Pearson correlation, $r^2(4) = 0.93$, $P < 0.0081$; Fig. 4E]. Hence, stimulus ambiguity was directly reflected in the ability of a linear classifier to discriminate among input based on heterogeneous neural activity.

Finally, we examined whether discrimination error in LDA was directly predictive of behavioral performance during the task. The mean percentage behavioral error across stimulus coherence was strongly related to the mean linear discrimination error of LIP [$r^2(4) = 0.98$, $P < 0.0013$; Fig. 4F]. In sum, as predicted by the population model, analyses of LIP activity were consistent with a driving role of neural heterogeneity in the discrimination of ambiguous visual stimuli. A linear classifier based on heterogeneous activity was not only sensitive to the coherence of visual stimuli, but also predicted mean behavioral errors during decision making.

We compared the above results with those obtained with the accumulation of mean population activity (Fig. 5A; Georgopoulos et al. 1986; Lee et al. 1988). In this approach, we computed the mean of all P neurons as well as that of all NP neurons. We set a decision threshold of 50 Hz on mean population activity in agreement with the approximate firing rate corresponding to a decision in the behavioral data (Roitman and Shadlen 2002). A decision was reached when either mean P or NP activity crossed the threshold. Importantly, results obtained from LDA as well as accumulation of population activity rely on the same model; the only difference lies in the way the activity of the model is employed to discriminate

stimuli. In this approach, mean error was only weakly related to stimulus coherence [$r^2(4) = 0.45$, $P > 0.15$; Fig. 5B]. To be sure, accumulation of mean population activity does convey stimulus-relevant information (see below); however, it conveyed limited information about stimulus coherence compared with a heterogeneous account. Furthermore, this approach yielded higher discrimination error than a pattern separation based on a population model of heterogeneous activity, suggesting that further information about the visual input can be gained by the latter approach than by the accumulation of mean population activity in LIP (Fig. 5C). Whereas computational work shows that many results obtained with a single accumulator model can be replicated with a population of weakly correlated accumulators (Zandbelt et al. 2014), our results show a clear advantage of heterogeneous accumulators in terms of discrimination error.

To illustrate the key difference between a mean population model and our LDA approach, consider three putative units (x , y , and z) responding to an input oriented either in the preferred or nonpreferred direction (Fig. 6A). The temporal fluctuations in these units leads to a mean population activity that is transiently higher for responses in the nonpreferred direction, which may result in a behavioral error (Fig. 6B). LDA, by comparison, takes the same activity and represents it in a multidimensional space where each dimension is a single unit (Fig. 6, C and D). Here, the activity of preferred and nonpreferred orientations can be accurately separated with a linear boundary. Despite the simplicity of this example, it illustrates that the representation of neural activity as either a population average or a multidimensional signal markedly affects the decision that is ultimately read out.

Next, we examine several aspects of neural activity in the population model that contribute to stimulus discrimination, including linear accumulation of activity from the onset of the stimulus onward, temporal fluctuations in neural activity over the course of a trial, variability in neuronal activity across

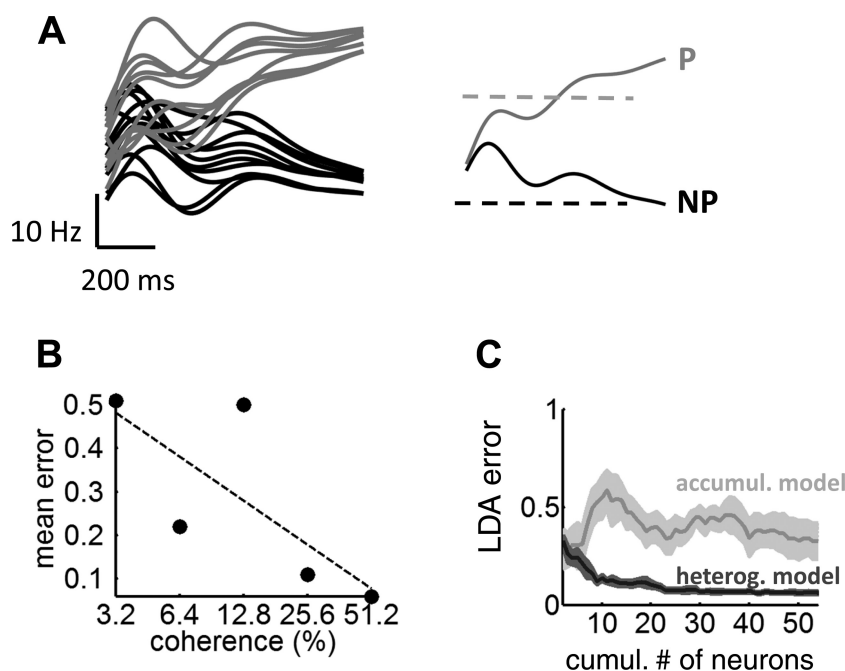


Fig. 5. Linear accumulator model of LIP activity. *A*: cartoon illustration of a simple accumulator model. *Left*: a population of individual mean-rate units that respond to a stimulus in the preferred (P; gray) or nonpreferred (NP; black) directions over the course of a trial. *Right*: mean population activity. Dashed lines are thresholds when a decision is reached. *B*: relation between stimulus coherence and mean discrimination error of an accumulator model based on mean population activity in LIP. Dashed line: best-fitting linear regression. *C*: comparison between the heterogeneous (heterog.) and accumulation (accumul.)-based classification of LIP population activity. Discrimination error is averaged over all 6 motion coherences.

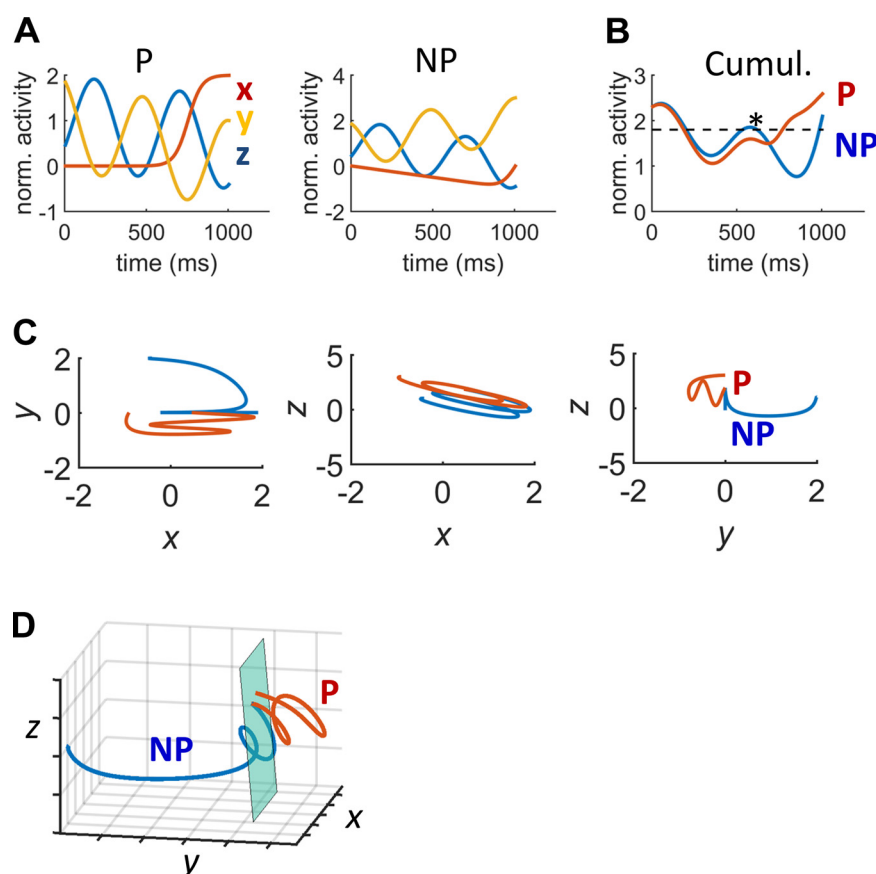


Fig. 6. Mean population accumulation compared with LDA. *A*: a cartoon illustration of 3 units (x , y , and z) responding to P vs. NP stimuli. *B*: sum of activity across the 3 units. Dashed line: decision threshold. Asterisk: decision corresponding to a behavioral outcome. *C*: 2-dimensional plots showing the activity of units x , y , and z for P vs. NP stimuli. *D*: same activity as *C* but plotted in 3-dimensional coordinates where each dimension is a single unit. Values on y -axis are reversed for ease of visualization. Green: separating plane.

stimuli, heterogeneity of tuning curves across units of the model, and noise correlations.

Impact of linear accumulation. One aspect of neural responses that may contribute to stimulus discrimination is the gradual accumulation of activity over the course of individual trials (Cain and Shea-Brown 2012). We tested this aspect by performing a linear regression of all single-unit activity. This regression used time as the predictor and was computed independently for each cell (where cells recorded over repeated trials were averaged together; see MATERIALS AND METHODS). The residuals of the regression were entered in LDA. As a result of removing linear accumulation, the population model no longer performed above-chance discrimination (Fig. 7, *left*). Neural accumulation therefore conveyed stimulus-relevant information that serves as a basis for discrimination.

A similar effect was observed in LIP data. Removing neural accumulation led to a marked increase in discrimination error,

with discrimination accuracy hovering around a random performance (0.5 proportion of misclassified trials; Fig. 7, *right*). Therefore, linear accumulation was a key requirement for stimulus discrimination in LIP (Hanks et al. 2015).

Temporal dynamics. A second aspect of neural activity that may contribute to stimulus discrimination is the diversity of responses across units in the population model (Fig. 2C). There are two potential aspects of this diversity that may contribute to discrimination: 1) the diversity in mean firing rates across units; and 2) unique temporal fluctuations in individual units over the course of a trial. We term the first aspect “rate-based heterogeneity” and the second aspect “temporal heterogeneity.”

In simulations, we isolated the contribution of these two aspects to stimulus discrimination. First, to assess the benefits of rate-based heterogeneity, we replaced the activity of each unit in the population model by a constant over time corresponding to its mean activity. As a result, the activity of

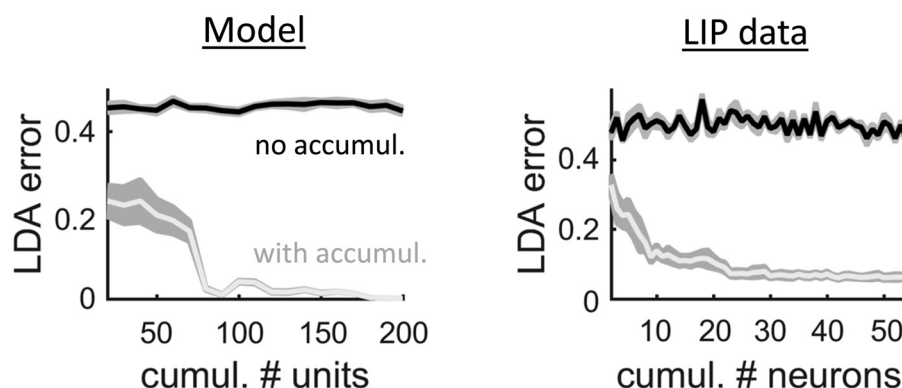


Fig. 7. Linear discrimination with LIP activity. *Left*: discrimination accuracy in the population model with neural accumulation (“with accumul.”) compared with discrimination after the removal of linear accumulation (“no accumul.”). *Right*: discrimination error obtained from LIP data before and after removing neural accumulation by linear regression.

individual units remained constant over time (Fig. 8A). Discrimination performance remained at chance in this scenario.

Second, we assessed the benefits of temporal heterogeneity by computing a “response motif” across all units, corresponding to the mean pattern of activity of each unit across time, taken over all simulated trials (total of 1,000). We then shifted the response motif of each unit by a constant (obtained through a brute-force search procedure) to reduce the absolute difference between the mean response motif across units and the response motif of each unit. As a result, all units had a similar firing rate but differed in their temporal dynamics (Fig. 8B). Classification performance in this scenario attained above-chance accuracy (Fig. 8C, black line). Taken together, these simulations show that heterogeneity in temporal dynamics (but not in mean firing rates) drives accurate stimulus discrimination.

We compared the above simulation results with analyses of LIP data. To characterize what aspects of neural activity contribute to discrimination in LIP, we investigated two versions of the experimental data, namely rate-based and temporal heterogeneity, as was done with the population model. Results are in line with predictions of the model: temporal heterogeneity (but not rate-based heterogeneity) yielded above-chance discrimination (Fig. 8D).

Going further, we examined what temporal resolution of the data was required to obtain an accurate discrimination. We

divided the data into equal-sized, nonoverlapping time bins and averaged the firing rate of each neuron within these bins. In this analysis, using a time bin equal to the total length of a trial would be equivalent to a rate-based discrimination, whereas a time bin of 1 ms would be equivalent to a temporal discrimination. Results show that low ($\sim 20\%$) discrimination error benefitted from precise temporal resolution (Fig. 8E).

In a follow-up analysis, we examined what temporal resolution of the LIP data was most compatible with behavioral performance. We began by binning the data with nonoverlapping windows of 10 ms and computed LDA for each time window from the onset of stimulus onward. We then examined the correlation between behavioral errors (Fig. 4F) and LDA error in the last time window of trials. Results show a gradual decrease in error over the time course of a trial (Fig. 8F), with a correlation to behavioral errors of $r = 0.98$. Crucially, this correlation varied with the size of the time window and was maximal for windows of 10–40 ms (Fig. 8F, *inset*). This time scale is shorter than the time constant corresponding to the decay of autocorrelation in LIP (Murray et al. 2014), raising the possibility that behavioral outcomes in the task may rely on a shorter integration window than the full time course of LIP activity, a question that is beyond our current scope.

Neuronal variability. A third aspect of neural activity that may contribute to stimulus discrimination is the variability in

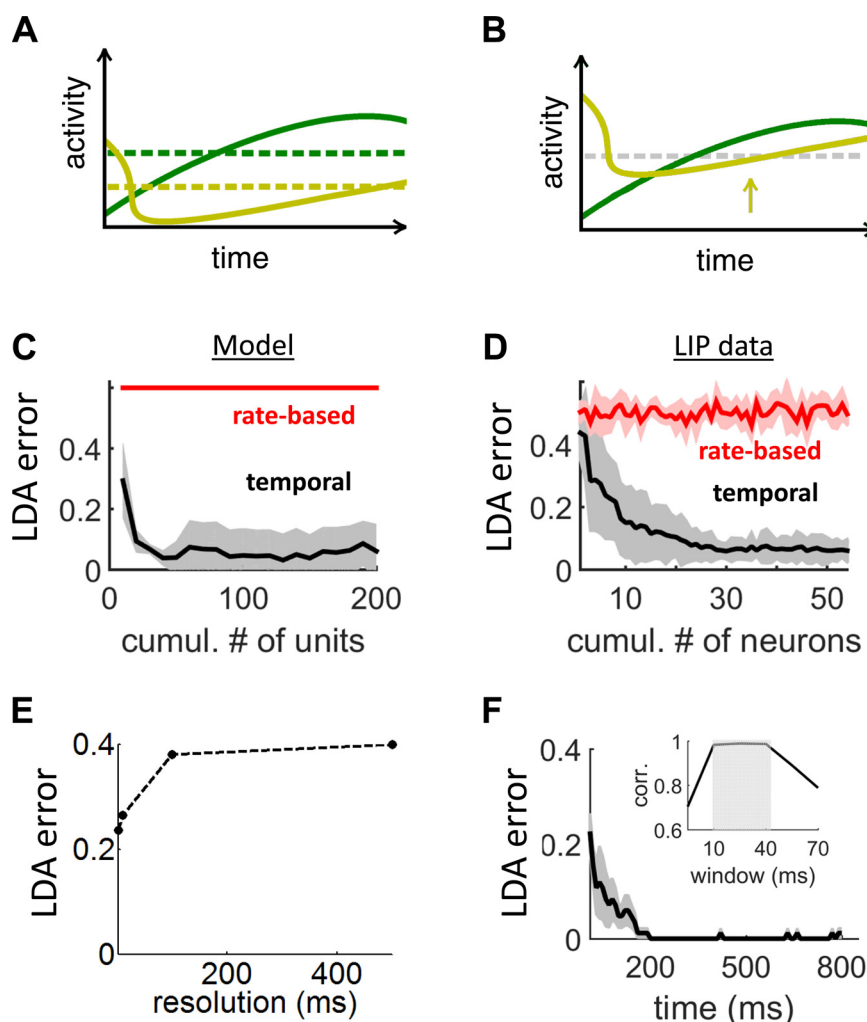


Fig. 8. Rate-based versus temporal discrimination. **A:** illustration of rate-based activity, where we replace the activity of individual units (solid lines) by the mean over time (dashed lines). **B:** illustration of temporal-based activity, where the activity of some units (solid lines) is shifted by a constant (indicated by arrow) such that the mean activity of all units (dashed line) is close in value. **C:** in simulations, stimulus discrimination that considers temporal dynamics (solid black line) outperforms discrimination with rate-based dynamics (solid gray line). **D:** rate-based vs. temporal dynamics in LIP data. **E:** in LIP, a high resolution of temporal dynamics is necessary to achieve a high discrimination performance. Each dot is an average discrimination error using all neurons. **C–E** are averaged over all 6 stimulus coherences. **F:** LDA classification of nonoverlapping time windows of 10 ms in LIP activity. *Inset:* correlation between end-of-trial LDA error and percentage behavioral error for time windows of varying sizes. Shaded area: highest correlations.

responses across different stimuli. This aspect is represented in the denominator of the LDA criterion function (Eq. 10) and is termed within-class variance. Neural variability across stimuli may account for the gradual decrease in discrimination error as further units are incorporated in LDA (Fig. 4A). Indeed, incorporating further units in LDA leads to a gradual reduction in within-class variance (Fig. 9A, left, solid black line). In turn, this gradual reduction shrinks the denominator of the LDA criterion function to be maximized (Eq. 10), thus improving stimulus discrimination.

What aspects of neural activity explain this gradual reduction in within-class variance? To address this question, we turn to a simplified account of neural activity based on a combination of homogeneous Gaussian processes with mean and variance obtained from the global mean and variance of the population model (Fig. 9A, left, solid gray line). The variance of the homogeneous Gaussian processes reduces as $\sim 1/\sqrt{n}$ units, yet discrimination error remains near chance because the mean difference between these processes remains constant (Fig. 9A, right, solid gray line). Assuming that the correlation between these Gaussian processes is 0, the exact reduction in variance as a function of n units can be expressed as follows, assuming $n > 1$:

$$\text{Var}\left(\sum_{i=1}^n X_i/n\right) = \sum_{i=1}^n \text{Var}(X_i/n), \quad (14)$$

where X_i are homogeneous Gaussian processes with finite mean and $\text{Var}(\cdot)$ returns the variance of a given vector. With positive correlation, the above expression becomes:

$$\text{Var}\left(\sum_{i=1}^n X_i/n\right) = \sum_{i=1}^n \text{Var}(X_i/n) + 2 \sum_{i < j} \text{Cov}(X_i/n, X_j/n), \quad (15)$$

where $\text{Cov}(\cdot)$ returns the covariance of two vectors. In this case, the variance obtained from Eq. 15 remains asymptotically above 0 at a value proportional to the covariance of the mean of X_i and X_j .

The lack of fit between homogeneous Gaussian processes and the behavior of the population model led us to develop an

alternative approach based on a combination of heterogeneous Gaussian processes with means and variances taken from each individual unit of the model. With these processes, discrimination error offered a close approximation to the population model. Within-class variance, however, behaved in a similar way as the homogeneous Gaussian process (Fig. 9A, dashed black line).

The poor correspondence in within-class variance between the population model and heterogeneous Gaussian processes may be attributed to a low degree of pairwise correlation between these processes. We tested this possibility by first computing pairwise correlations in the model (Fig. 9B). These pairwise correlations were comparable in range with experimental values (Cohen and Newsome 2008; Zohary et al. 1994). Then, we employed a Cholesky factorization that allowed us to control the mean pairwise correlation of the heterogeneous Gaussian processes and match it to the population model.

This factorization worked as follows. We defined a matrix $\mathbf{S}$ of size $p \times p$, with elements $S_{ij} = 1$ for $i = j$ and r otherwise (where r is the desired correlation between patterns of activity). Then, we performed a Cholesky factorization $\mathbf{S} = \mathbf{L}\mathbf{L}^*$, where $\mathbf{L}^*$ indicates a conjugate transpose. We performed an element-wise multiplication of $\mathbf{L}$ by a matrix of same dimensions with Gaussian elements for which mean and variance matched the population model. The within-class variance of the resulting correlated processes offered a close fit to the model (Fig. 9A, dashed gray line).

We repeated the above analysis using LIP data. As with the population model, we found a poor fit of homogeneous Gaussian processes to both within-class variance and discrimination error (Fig. 9C). We also found a poor fit of the heterogeneous Gaussian processes to within-class variance. As with the model, the variance of LIP data saturated at a markedly higher value than Gaussian processes as further neurons were incorporated in LDA. Unfortunately, because each LIP unit was acquired during an individual recording session, we cannot estimate pairwise correlations. However, we strongly suspect that taking such correlations into account would offer a good fit

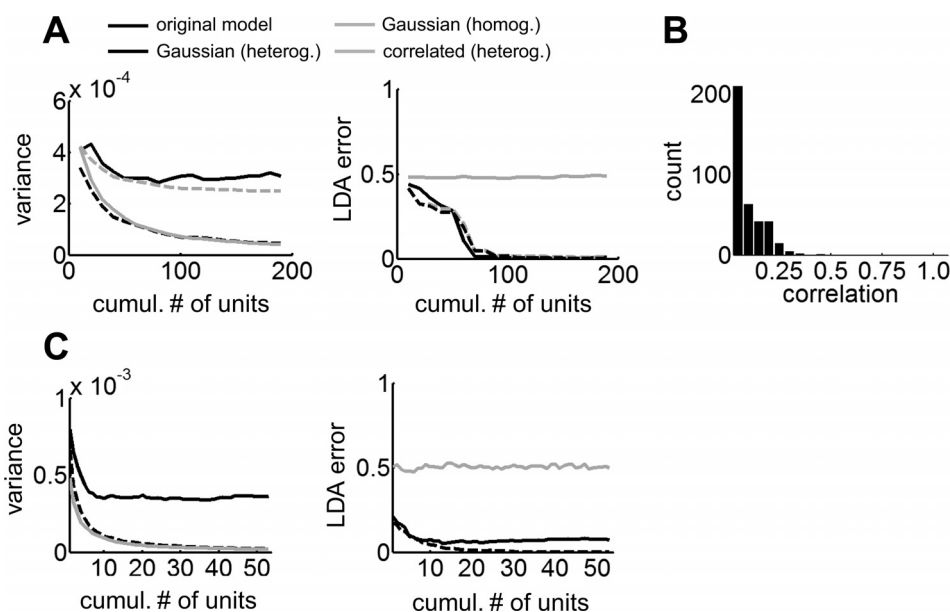


Fig. 9. Neuronal variability and discrimination error. A: within-class variance (left) and discrimination error (right) obtained for the population model as well as different Gaussian processes. Values are averaged over all 6 coherences. homog., Homogeneous. B: distribution of pairwise correlations in the model. C: within-class variance and discrimination error in LIP data and Gaussian processes. Legend follows from A.

to both variance and discrimination error, as was found with the population model.

Tuning-curve diversity. A fourth aspect of neural activity that may contribute to stimulus discrimination is the heterogeneity of tuning curves across units of the population model. Naturalistic tuning curves present heterogeneous amplitudes, widths, and baselines that may contribute to the processing of sensory input (Pitkow et al. 2015). In the model, we found superior discrimination performance with heterogeneous tuning curves compared with a homogeneous account (Fig. 10, A and B).

To analyze further the influence of tuning-curve heterogeneity in the discrimination of motion stimuli, we turned to a measure of Fisher information (Ecker et al. 2011; Shamir and Sompolsky 2006). This measure reflects an upper bound on stimulus encoding by a population of neurons as extracted by an unbiased linear estimator (Abbott and Dayan 1999; Moreno-

Bote et al. 2014). A simplifying assumption here is that the readout mechanism can utilize all of the information that is encoded to guide stimulus discrimination. Linear Fisher information is computed as follows:

$$J(\theta) = \sum_{ij} f'_i(\theta) C_{ij}^{-1} f'_j(\theta), \quad (16)$$

where the summation is taken over pairs of neurons i and j for a given stimulus orientation θ , $f'_i(\theta)$ is the derivative of the tuning-curve function evaluated at the stimulus orientation, and C_{ij} is the correlation given by:

$$C_{ij} = c_{ij} \exp\left(-\frac{\|\phi_i - \phi_j\|}{p}\right), \quad (17)$$

where $\|\cdot\|$ is the relative angle difference between preferred orientations ϕ_i and ϕ_j (ranging from 0 to π), c_{ij} is the noise

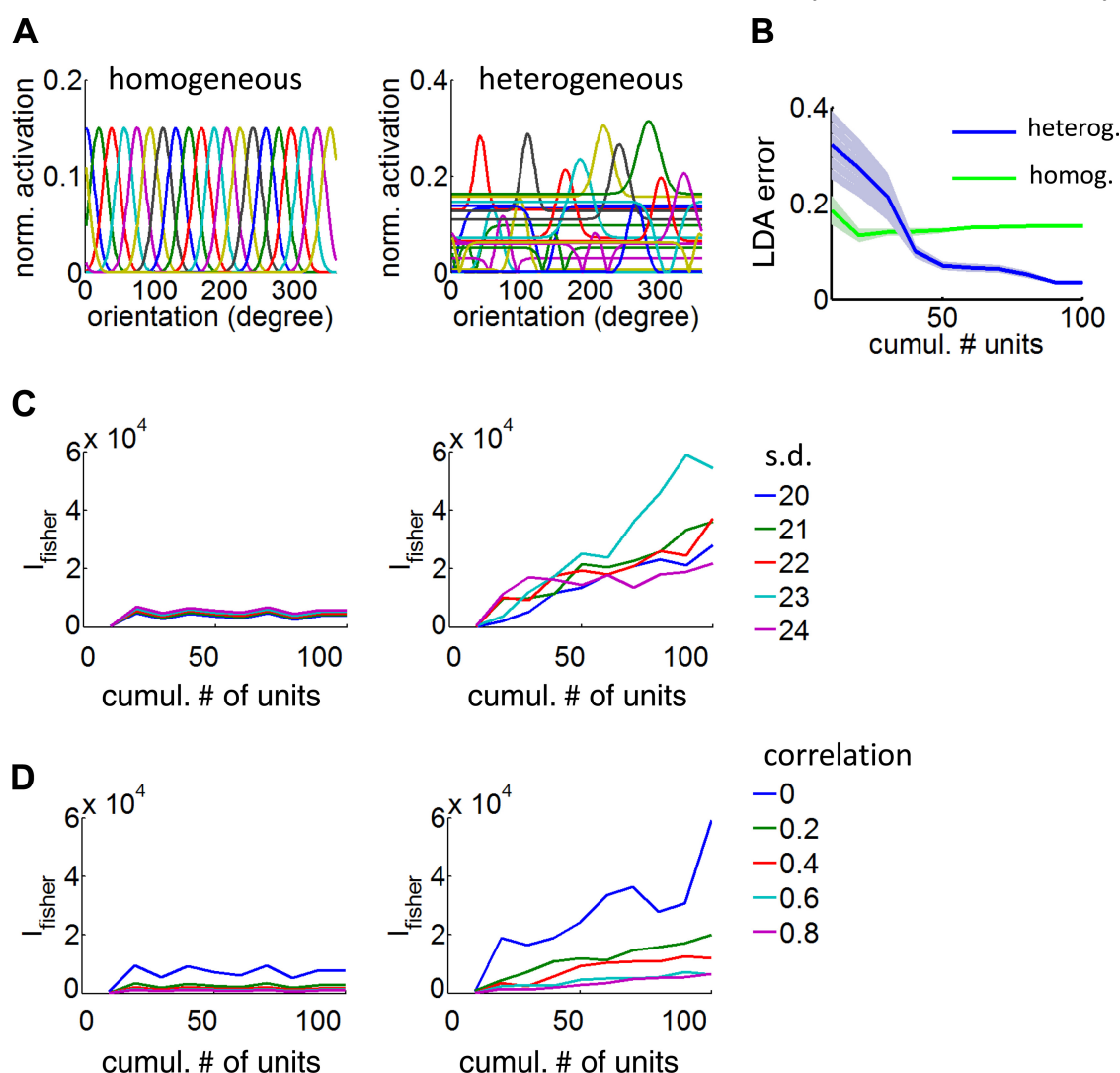


Fig. 10. Impact of tuning-curve diversity. *A, left:* homogeneous tuning curves. For ease of visualization, only 20 curves are displayed. Parameters are as follows: maximum amplitude (0.01), variance (20), and spacing between tuning curves ($360/N$). *Right:* heterogeneous parameters chosen from a Gaussian distribution: amplitudes (mean 0.01, SD 0.001), variance (mean 20, SD 0.5), and spacing between tuning curves (mean corresponding to preferred orientation, SD 2°). *B:* mean discrimination error taken over all 6 coherences for heterogeneous (“heterog.”) and homogeneous (“homog.”) networks. Shaded area: SE. *C:* Fisher information (I_{fisher}) for homogeneous (*left*) and heterogeneous (*right*) tuning functions ($N = 100$ units). Legend shows different values for the variance of tuning curves (in the case of heterogeneous curves, these values are the mean of a parameter distribution). For heterogeneous curves, correlations were chosen from a Gaussian distribution (mean 0.2, SD 0.05). *D:* Fisher information with different values of correlation (shown in legend to the right of row) for homogeneous (*left*) and heterogeneous (*right*) tuning curves.

correlation between neighboring cells (i.e., the correlation between responses to repeated presentations of the same stimulus), and p is a decay length (set by default to 1.0 as in related work; Ecker et al. 2011). Based on the mean pairwise signal correlation calculated from our default population model (Fig. 10B), we set $c_{ij} = 0.2$ for all pairs i, j for the homogeneous model and drew c_{ij} values from a Gaussian distribution for the heterogeneous model (Table 1).

We computed Fisher information for a range of stimuli between 1 and 360° at intervals of 1°, where each stimulus represented a single orientation along one dimension. Furthermore, we assumed that noise correlations only depended on the differences in preferred direction between pairs of tuning functions. Our goal here was limited to examining the contribution of naturalistic tuning curves to stimulus encoding and comparing them to homogeneous tuning curves. Although naturalistic tuning curves show variability in amplitude, baseline, variance, as well as spacing in terms of preferred orientation, only the amplitude variability is analytically tractable (Ecker et al. 2011), and therefore we turned to numerical simulations.

Using simulations, we compared the mean Fisher information of homogeneous networks (where all parameters have the same value across units) and heterogeneous networks (where baselines, amplitudes, variance, intervals, and correlations of tuning functions were chosen from Gaussian distributions; Table 1). Heterogeneous networks attained higher Fisher information than homogeneous networks across a range of values of variance of tuning curves (Fig. 10C) and correlation (parameter c_{ij} in Eq. 17; Fig. 10D). Note that, due to the data-processing inequality (Cover and Thomas 1991), Fisher information will eventually saturate in heterogeneous networks as further neurons are added; this saturation will occur at a point corresponding to a greater number of neurons than the homogeneous model (Kanitscheider et al. 2015).

The latter figure suggests that highest Fisher information is obtained with correlations of 0. Indeed, this result is confirmed when examining Fisher information across a range of correlation values (Fig. 11A, *left*). However, a more nuanced portrait emerges in heterogeneous networks when we consider the mean and SD of the Gaussian distribution from which we picked these correlations. In a series of simulations, we drew correlations (i.e., the value of c_{ij} in Eq. 17) from a Gaussian

distribution with different means and SDs and examined the impact on Fisher information. With a mean correlation of $c_{ij} = 0.1$, the value of SD yielding the highest Fisher information was ~ 0.4 (Fig. 11A, *middle*). With a mean correlation to $c_{ij} = 0.3$, however, the value of SD yielding the highest Fisher information shifted to ~ 0.6 (Fig. 11A, *right*).

We repeated the above analysis using different values of c_{ij} (from 0 to 1.0). For each value of c_{ij} , we identified the value of SD yielding the highest Fisher information. Surprisingly, we found that maximal Fisher information was obtained with a mean correlation value of $c = 0.3$, hence markedly above a value of 0 correlation (Fig. 11B). Therefore, higher Fisher information may be obtained with positive (as opposed to 0) correlation values in heterogeneous networks. This result is shown elsewhere to depend on several factors, including noise entropy (Ecker et al. 2011) and the stimulus itself (Zylberberg et al. 2016).

Next, we examined to what extent the heterogeneity of tuning-curve parameters contributed to Fisher information. First, we considered the variance of tuning curves (σ^2 in Eq. 4). We ran a number of simulations where we varied either the mean variance or SD of the variance across units and found that, for both mean and SD, a non-0 value was associated with highest Fisher information (Fig. 11C). Thus some degree of heterogeneity may be ideal at promoting Fisher information in the orientation selectivity of a neuronal population.

We repeated the above procedure by varying the SD of Gaussian distributions from which the distance (in degrees) between the preferred stimulus orientations of tuning curves was drawn. We kept the mean of these distributions fixed at $360/N$. We found that Fisher information was maximal for non-0 values of SD (Fig. 11D). Thus introducing heterogeneous distances between the preferred stimulus orientations of a population of neurons may lead to improved Fisher information compared to a homogeneous account.

Taken together, the above results show that heterogeneity in tuning curves improves Fisher information as well as the discrimination of dot-motion stimuli. These results corroborate the role of neural heterogeneity in pattern separation (Fig. 3, A and B) and stimulus discrimination (Fig. 4, A–C).

Noise correlations. A fifth aspect of neural activity that contributes to stimulus discrimination is the presence of noise correlation between units, defined as the correlation between

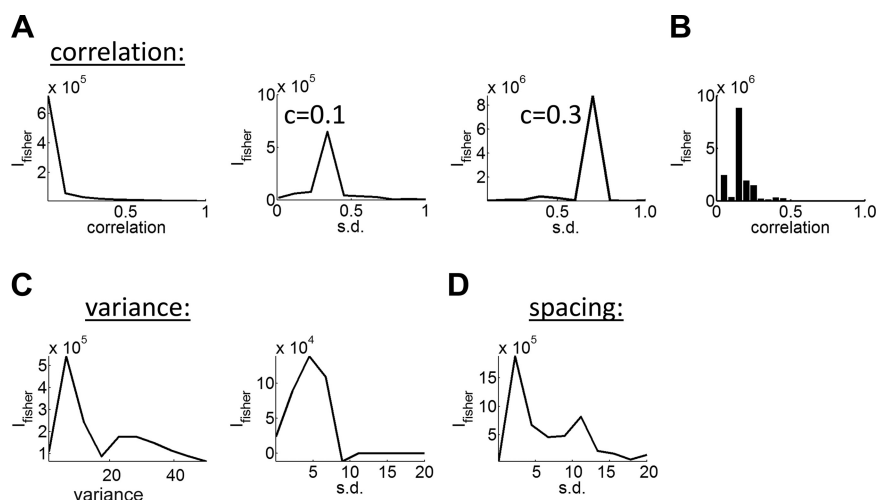


Fig. 11. Impact of correlated tuning curves on Fisher information. A, *left*: in heterogeneous networks, low correlation values (parameter c_{ij} in Eq. 17) are associated with higher Fisher information. *Middle and right*: with a given fixed value of mean correlation ($c = 0.1$ or $c = 0.3$), a narrow range of SD values (~ 0.4 and ~ 0.6 , respectively) is associated with maximal Fisher information. B: maximal value of Fisher information obtained by varying the SD in simulations with different mean correlations. C: impact of varying the standard deviation of tuning curves (*left*: mean; *right*: SD) on Fisher information. D: the degree of spacing between preferred orientation influences Fisher information.

responses of pairs of units to repeated presentations of the same stimulus. The effect of noise correlation on discrimination depends on the direction of the class separation (Averbeck et al. 2006) and the orientation of the stimuli relative to the neuronal tuning curves (Franke et al. 2016; Zylberberg et al. 2016).

To illustrate these points, we considered two reciprocally connected linear units (Eq. 3) that receive a constant input ($I_i = 10$ for $t = 1, \dots, T$) at a given orientation (Fig. 12A, left). We

began with a scenario where the two units are connected via reciprocal positive weights (as is expected with units that have similar tuning preference; Fig. 1C). A return map of these two units shows the relation between their tuning curves as well as the presence of positive noise correlations (Fig. 12A, middle). These correlations are orthogonal to the return map of the tuning curves and therefore promote the discrimination of similarly oriented stimuli (Franke et al. 2016). Put differently, a negative relation between signal and noise correlations pre-

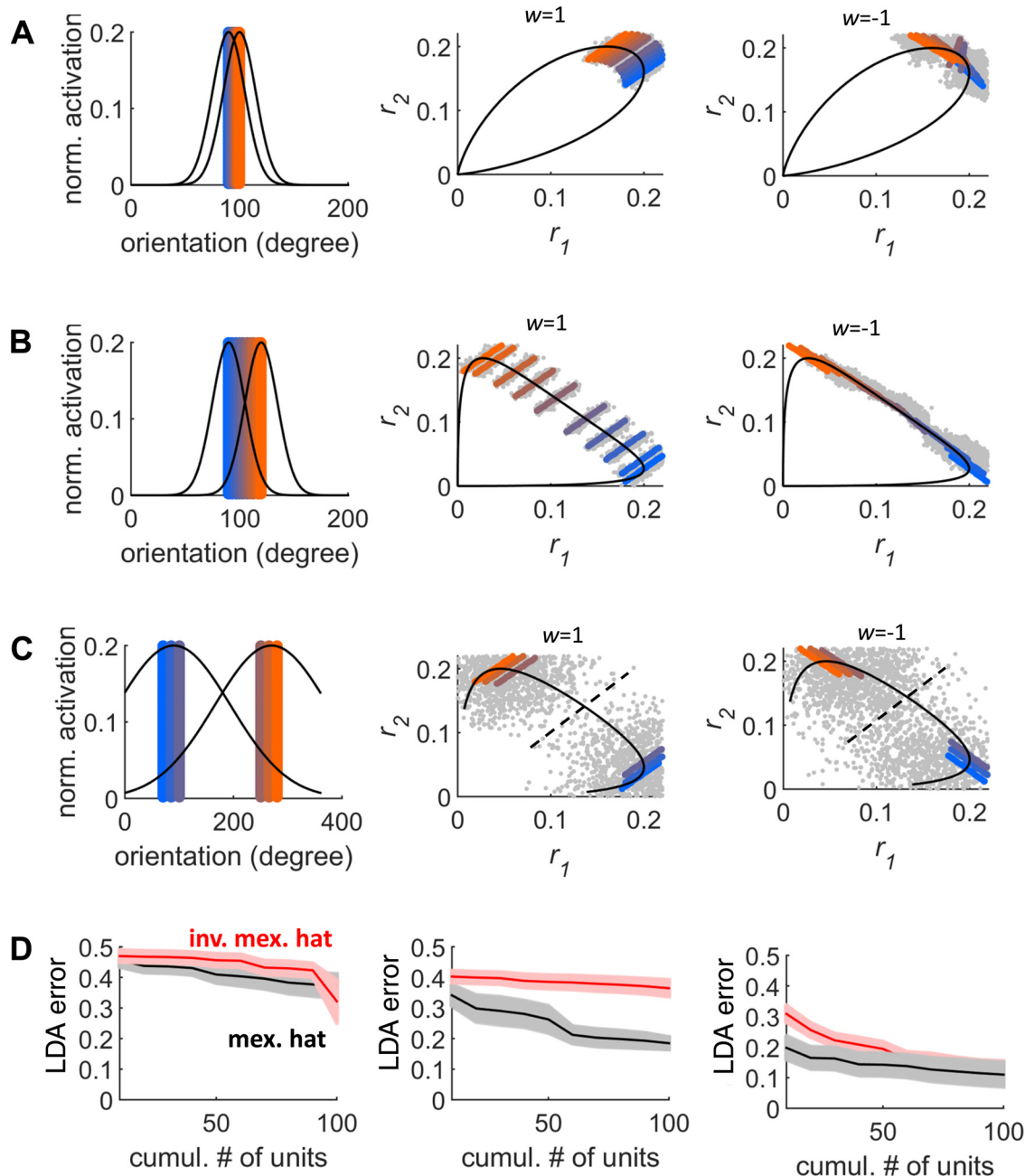


Fig. 12. Relation between signal and noise correlation shapes stimulus discrimination. **A, left:** tuning curves of 2 units in the model (solid black lines). Color shows stimulus orientation. **Middle and right:** return map showing the relation between tuning curves (solid black line) of 2 units (r_1 and r_2). Solid colored lines: orientation of noise correlations for individual stimulus orientations corresponding to the left. Gray: normalized responses of the 2 units. " $w=1$ " and " $w=-1$ " indicate reciprocal positive and negative connections, respectively. **B and C:** increasing the distance between preferred orientations of the tuning curves. For ease of visualization, variance of the tuning curves was increased to 60 in **C**. Input noise had SD of 0.05 except in **C** where it was increased to 0.2. Solid dashed lines: arbitrary separation between stimuli oriented at 90 and 270°. **D:** linear discriminant analysis of stimuli at 90 vs. 100° orientation (left), 90 vs. 120° (middle), and 90 vs. 270° (right). Input noise had SD of 0.2. Solid black line: Mexican hat ("mex. hat") connectivity (w_{ij} ; Fig. 1D). Solid red line: inversed Mexican hat ("inv. mex. hat") connectivity set to $w_{ij} = -w_{ji}$ for all pairs of units i, j . Error is averaged over all stimulus coherences.

vents neural responses to nearby stimuli from overlapping, therefore improving discrimination.

By comparison, we considered two units connected via reciprocal negative weights (Fig. 12A, *right*). In this case, negative noise correlations create a marked overlap between the neural responses to nearby stimuli on the return map, therefore hindering discrimination.

Increasing the distance between tuning curves (Fig. 12B, *left*) leads to an increase in the size of the region where positive noise correlations are orthogonal to the return map (Fig. 12B, *middle*). In turn, this leads to a greater region where these positive correlations benefit stimulus discrimination. These results are in contrast to those obtained with negative weights (Fig. 12B, *right*).

To examine noise correlations in the context of the motion discrimination task described earlier, we examined two units with orientation preferences of 90 and 270°, respectively (Fig. 12C, *left*). With positive reciprocal weights, stimuli presented near the orientation preference of either unit fall outside of the “flat” region of the return map where positive correlations improve discrimination (Fig. 12C, *middle*). To obtain noise correlations that lie orthogonal to the return map, negative weights are required (Fig. 12C, *right*). This result, however, is strongly modulated by input noise. In a scenario where we increased the SD of input noise from its default value of 0.05 to 0.2, negative noise correlations will yield an overlap between orthogonal stimuli (Fig. 12C, gray dots).

To examine noise correlations in the context of a broader population of units, we took advantage of the fact that the sign of weights is related to noise correlations (Fig. 1C). In this way, Mexican hat connectivity yields positive noise correlations for nearby stimuli and negative noise correlations otherwise (Fig. 1C). We compared stimulus discrimination under two conditions: 1) using a standard Mexican hat; or 2) using a “mirror” connectivity scheme with $w_{ij}^* = -w_{ij}$ for all pairs of units i, j . Discrimination was poor when stimuli were close together in orientation (90 vs. 100°), regardless of the type of connectivity employed (Fig. 12D, *left*), given the high input noise employed. However, performance improved with stimuli that were farther apart (90 vs. 120° as well as 90 vs. 270°; Fig. 12D, *middle* and *right*, respectively). Importantly, lower error was attained with Mexican hat connectivity than mirror connectivity. Thus connectivity shaped noise correlations and was instrumental in driving stimulus discrimination.

In sum, stimulus discrimination is sensitive to the structure of signal correlations, noise correlations, and input noise. The relation between these variables outlines conditions under which different configurations of connection weights contribute to discrimination.

DISCUSSION

This work examined whether neural activity in LIP during a task of decision making is consistent with a pattern separation hypothesis where ambiguous inputs are accurately classified by network activity. A combination of experimental, computational, and analytical results shows that the heterogeneity of responses in LIP provides the substrate for neuronal pattern separation where motion-relevant input may be reformatted and adequately classified. Together, findings suggest a role for pattern separation outside of hippocampal areas where it is

traditionally reported (Leutgeb et al. 2007; McHugh et al. 2007) and raise the prospects of uncovering similar functions in other circuits of the brain.

The ability of the population model to capture experimental findings on visual stimulus discrimination despite few assumptions of the underlying neural processes suggests that pattern separation based on heterogeneous activity may form a general principle of stimulus discrimination across various modalities. Given the remarkable heterogeneity of the mammalian neocortex (Braitenberg and Schüz 1998), neuronal pattern separation may be found across several regions, particularly where low signal correlations are present, as may be expected in several areas of cortex (Brody et al. 2003; Singh and Eliasmith 2006).

The proposed population model relies on a handful of key biological principles. First, a balance of excitation and inhibition prevented runaway activation in network activity in response to input (see APPENDIX). Second, because of randomness in the connectivity and orientation selectivity of the model, each unit received a unique combination of input that contributed to the heterogeneity of neural responses. In turn, this heterogeneity was instrumental to the maintenance of low signal correlations in the responses of individual units.

We have shown that several aspects of the population model contributed to its ability to classify stimuli, including neural heterogeneity, linear accumulation, temporal dynamics, variability across stimuli, and noise correlations. However, it is important to consider that neural heterogeneity may not always lead to increased stimulus discrimination. Indeed, alterations in the model that lead to unstable neural activity may result in decreased discrimination accuracy. Furthermore, the link between neural heterogeneity and discrimination may depend on the correlated nature of neuronal activity. In the model, units that show strong signal correlations contributed little to stimulus discrimination (Fig. 3; Orhan and Ma 2015). This raises a problem with neuronal circuits where correlated activity is often observed (Averbeck et al. 2006). In the model, signal correlations arise in part because of balanced connectivity. The reason for the stability of mean population dynamics despite fluctuations in single units is that excitation and inhibition “track” each other and thus cancel out. This tracking, in turn, creates statistical dependencies among the units, which reduce stimulus discrimination.

Despite the fact that some of our computational results (Figs. 10D and 11A) as well as those of others (Abbott and Dayan 1999; Sompolinsky et al. 2001; Zohary et al. 1994) show a negative impact of neural correlations on information processing and stimulus discrimination, more subtle effects were observed when we considered correlations among the tuning curves of a heterogeneous model (Fig. 11, A and B). Surprisingly, under some conditions, positive noise correlations improved Fisher information. We examined this finding with pairs of units and showed that positive noise correlations yielded better separation of responses along a return map of responses based on tuning curves. Furthermore, under defined conditions of input noise and signal correlations, positive noise correlations improved stimulus discrimination. The relation between neural correlation and information processing is thus not straightforward and depends on several factors, including neural heterogeneity (Ecker et al. 2011).

It is worth considering whether neuronal pattern separation is indeed required for stimulus discrimination, given that pop-

ulation LIP activity is linearly separable in tasks of visual dot-motion discrimination (Roitman and Shadlen 2002). Perhaps a simpler alternative based on the accumulation of mean population activity (Fig. 5) could capture behavioral performance on successful versus unsuccessful trials for different degrees of coherence (Smith and Ratcliff 2004). However, we showed that such an approach does not fare so well in this respect (Fig. 7). By comparison, a framework based on neuronal pattern separation puts the issue of linear separability at the center of the debate and considers whether the accuracy of a linear classifier reflects behavioral performance. With this framework, we show a strong relationship between linear discrimination error and stimulus coherence (Fig. 4E) as well as between discrimination and behavioral errors (Fig. 4F). Therefore, evidence suggests that it is appropriate to consider the relationship between neuronal pattern separation and task performance during visual motion discrimination.

By transforming the neural activity of LIP data as well as the linear model, we found that heterogeneity in temporal dynamics, but not firing rates, contributed to stimulus discrimination (Fig. 8). One aspect of temporal dynamics that drives discrimination is the linear accumulation of neural activity over the course of a trial; indeed, removing such feature from the data by linear regression resulted in chance-level discrimination (Fig. 7). This result, however, does not rule out alternative strategies by which LIP may perform stimulus discrimination without relying on gradual linear accumulation, for instance through discrete processing steps (Latimer et al. 2015).

One limitation of our approach is that the simplicity of the proposed population model does not address how a neural code based on spikes (in the form of either spike rate or precise timing) could achieve pattern separation. Recent work on asynchronous networks has begun to address this question, and further work is needed (Ostojic 2014). In addition, our work does not consider the distinct role of recurrent versus feedforward pathways in performing pattern separation, a question that is beyond the scope of our results. We emphasize, however, that recurrent networks may be designed to mimic the behavior of feedforward networks, thus blurring the lines between these two architectures (Goldman 2009; Vincent et al. 2012). Finally, our model does not consider the role of synaptic plasticity in the disambiguation of incoming stimuli. Plasticity rules including spike timing-dependent plasticity and the BCM rule (Bienenstock et al. 1982) may contribute to pattern separation by decorrelating (Linster and Cleland 2010) and stabilizing (El Boustani et al. 2012) incoming patterns of activity.

Although our results have focused on the role of LIP in neuronal pattern separation, it is likely that other regions are involved in this process in a distributed fashion. It would be worthwhile to extend our approach to other pathways, both cortical and subcortical, that contribute to visual motion discrimination and rapid decision making. It remains an open question whether (and how) pattern separation may be distributed across these various pathways.

Although we acknowledge alternative views on the role of LIP, including the integration of information over time (Roitman and Shadlen 2002), the computation of likelihood ratios (Kira et al. 2015), and the control of a speed-accuracy tradeoff (Hanks et al. 2014), our proposal of a pattern separation mechanism in LIP offers a close fit to both input statistics (Fig. 4E) and behavioral performance (Fig. 4F). Although alterna-

tive models cannot be ruled out, our results provides viable predictions to refine or amend the pattern separation hypothesis. These predictions characterize the variance of neural responses (Fig. 9) and the relation between input and noise correlations (Fig. 12), both of which will require simultaneous recordings in a population of LIP neurons.

Although our model relies on the heterogeneous nature of responses to perform pattern separation, it is plain that other candidate mechanisms should also be considered. Nonlinear dendritic integration, for instance, may help project synaptic input into a space where linear discrimination is possible (Branco and Hausser 2011; Lee et al. 2016). Several other factors could also be taken into account, including short-term plasticity (Leibold and Bendels 2009), long-term plasticity (Dan and Poo 2004), and behavioral state (DiCarlo et al. 2012). Each of these factors may either enhance or decrease the heterogeneity of neural responses, in turn helping or hindering stimulus discrimination.

APPENDIX

Linear stability analysis. It is straightforward to show that the population model described in Eq. 3 (main text) eventually decays back to a stable state when the stimulus is removed (Thivierge et al. 2014). As a starting point, let us simplify the model by considering only a single excitatory (x) and inhibitory (y) unit, reciprocally connected with synapses w and $-w$, each having transmission delay δ :

$$\tau \frac{dx}{dt} = -\alpha x + I - wy(t - \delta), \quad \tau \frac{dy}{dt} = -\alpha y + I + wx(t - \delta). \quad (A1)$$

Although decision-making models typically set $w_{ij} = w_{ji}$, here we consider a mean-rate simplification of the proposed balanced network where afferent excitation and inhibition cancels out over the whole population. The fixed points of the above system (i.e., the values of x and y for which the derivatives are 0) are given by the solutions of the following equations:

$$x^* = \alpha x^* + I - wy^*, \quad y^* = \alpha y^* + I + wx^*. \quad (A2)$$

We define $X = x - x^*$ and $Y = y - y^*$. Centering x and y around the fixed points yields:

$$x^* = I \frac{(\alpha - w)}{\alpha^2 + w^2}, \quad y^* = I \frac{(\alpha - w)}{\alpha^2 + w^2}. \quad (A3)$$

By substitution into Eqs. A1 and A2, we obtain:

$$\tau \frac{dX}{dt} = -\alpha X - wY(t - \delta), \quad \tau \frac{dY}{dt} = -\alpha Y + wX(t - \delta). \quad (A4)$$

Setting $A(T) = X(t)$, $B(T) = Y(t)$, and $T = t/\tau$:

$$A(T - 1) = X(t - \tau), \quad B(T - 1) = Y(t - \tau). \quad (A5)$$

Given the above as well as:

$$\frac{d}{dt} = \frac{1}{\tau} \frac{d}{dT}, \quad (A6)$$

Eq. A4 becomes:

$$\begin{aligned}\frac{d}{dT}A(T) &= -\alpha A(T) - wB(T-1), \\ \frac{d}{dT}B(T) &= -\alpha B(T) - wA(T-1).\end{aligned}\quad (A7)$$

A linear stability analysis of Eq. A7 yields eigenvalues:

$$\begin{aligned}0 &= (\lambda_n + \alpha)^2 + we^{-\lambda_n\delta} \\ 0 &= (\lambda_n + \alpha - we^{-\lambda_n\delta})(\lambda_n + \alpha + we^{-\lambda_n\delta}) \\ \lambda_n &= -\alpha \pm we^{-\lambda_n\delta}.\end{aligned}\quad (A8)$$

When $w \neq 0$, the characteristic equation admits the solution:

$$\lambda_n = -\alpha + \frac{W_n(\pm w\delta e^{\alpha\delta})}{\delta}, \quad (A9)$$

where $W_n(z)$ is the analytic continuation of the product log function (also termed Lambert W), which gives the principal solution for w in $z = we^w$. The stability conditions leading to $\text{Re}(\lambda_n) < 0$ are given by:

$$|w| < \alpha \text{ and } \delta < \frac{\cos^{-1}(\pm \alpha/w)}{\sqrt{w^2 - \alpha^2}}. \quad (A10)$$

Because our model assumes that $w > 0$, a direct implication is that $\alpha = 0$ yields an unstable solution. In simple terms, this is because such a perfect integrator accumulates network activity in an uncontrolled fashion. More generally, the term α imposes strict limits on the range of connection weights that will lead to stable dynamics. Generalizing the above analysis to more than two units, any unit in a larger network will be stable so long as the absolute sum of its incoming weights is lower than α . This condition is met in our network because the balance of excitation and inhibition results in a sum of weights that is tuned to be near 0 (Fig. 1D). In the case of an unbalanced network, however, population activity grows unchecked.

ACKNOWLEDGMENTS

We are thankful to Guillaume Lajoie, Alex Huk, Paul Cisek, André Longtin, and Denis Cousineau for useful comments.

GRANTS

This work was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC Grants 210977 and 210989), operating funds from the Canadian Institutes of Health Research (CIHR Grant 6105509), and the University of Ottawa Brain and Mind Institute (uOBMI).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

N.B., A.M., L.R., and J.-P.T. conception and design of research; N.B., A.M., L.R., and J.-P.T. performed experiments; N.B., A.M., E.G., L.R., and J.-P.T. analyzed data; N.B., A.M., L.R., and J.-P.T. interpreted results of experiments; N.B. and J.-P.T. prepared figures; J.-P.T. drafted manuscript; N.B., A.M., E.G., L.R., and J.-P.T. edited and revised manuscript; N.B., A.M., E.G., L.R., and J.-P.T. approved final version of manuscript.

REFERENCES

- Abbott LF, Dayan P. The effect of correlated variability on the accuracy of a population code. *Neural Comput* 11: 91–101, 1999.
- Averbeck BB, Latham PE, Pouget A. Neural correlations, population coding and computation. *Nat Rev Neurosci* 7: 358–366, 2006.
- Baeg EH, Kim YB, Huh K, Mook-Jung I, Kim HT, Jung MW. Dynamics of population code for working memory in the prefrontal cortex. *Neuron* 40: 177–188, 2003.
- Bendat JS, Piersol AG. *Random Data: Analysis and Measurement Procedures*. Hoboken, NJ: Wiley, 2010.
- Bienenstock EL, Cooper LN, Munro PW. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J Neurosci* 2: 32–48, 1982.
- Braitenberg V, Schüz A. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin: Springer, 1998.
- Branco T, Häusser M. Synaptic integration gradients in single cortical pyramidal cell dendrites. *Neuron* 69: 885–892, 2011.
- Brody CD, Hernandez A, Zainos A, Romo R. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cereb Cortex* 13: 1196–1207, 2003.
- Buonomano DV, Maass W. State-dependent computations: spatiotemporal processing in cortical networks. *Nat Rev Neurosci* 10: 113–125, 2009.
- Burak Y, Fiete IR. Fundamental limits on persistent activity in networks of noisy neurons. *Proc Natl Acad Sci USA* 109: 17645–17650, 2012.
- Cain N, Barreiro AK, Shadlen MN, Shea-Brown E. Neural integrators for decision making: a favorable tradeoff between robustness and sensitivity. *J Neurophysiol* 109: 2542–2559, 2013.
- Cain N, Shea-Brown E. Computational models of decision making: integration, stability, and noise. *Curr Opin Neurobiol* 22: 1047–1053, 2012.
- Carland MA, Marcos E, Thura D, Cisek P. Evidence against perfect integration of sensory information during perceptual decision making. *J Neurophysiol* 115: 915–930, 2015.
- Chance FS, Abbott FS, Reyes AD. Gain modulation from background synaptic input. *Neuron* 35: 773–782, 2002.
- Chapuis J, Wilson DA. Bidirectional plasticity of cortical pattern recognition and behavioral sensory acuity. *Nat Neurosci* 15: 155–161, 2012.
- Chaudhuri R, Knoblauch K, Gariel MA, Kennedy H, Wang XJ. A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron* 88: 419–431, 2015.
- Cisek P, Puskas GA, El-Murr S. Decisions in changing conditions: the urgency-gating model. *J Neurosci* 29: 11560–11571, 2009.
- Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011.
- Cohen MR, Newsome WT. Context-dependent changes in functional circuitry in visual area MT. *Neuron* 60: 162–173, 2008.
- Compte A, Constantinidis C, Tegner J, Raghavachari S, Chafee MV, Goldman-Rakic PS, Wang XJ. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J Neurophysiol* 90: 3441–3454, 2003.
- Cover TM, Thomas JA. *Elements of Information Theory*. New York: Wiley, 1991.
- Dan Y, Poo MM. Spike timing-dependent plasticity of neural circuits. *Neuron* 44: 23–30, 2004.
- Dayan P, Abbott LF. *Theoretical Neuroscience*. Cambridge, MA: MIT Press, 2001.
- DiCarlo JJ, Zoccolan D, Rust NC. How does the brain solve visual object recognition? *Neuron* 73: 415–434, 2012.
- Ditterich J. Evidence for time-variant decision making. *Eur J Neurosci* 24: 3628–3641, 2006.
- Ecker AS, Berens P, Tolias AS, Bethge M. The effect of noise correlations in populations of diversely tuned neurons. *J Neurosci* 31: 14272–14283, 2011.
- El Boustani S, Yger P, Frégnac Y, Destexhe A. Stable learning in stochastic network states. *J Neurosci* 32: 194–214, 2012.
- Franke F, Fisel M, Sevel M, Roska B, Hierlemann A, da Silveira RA. Structures of neural correlation and how they favor coding. *Neuron* 89: 409–422, 2016.
- Ganguli S, Bisley JW, Roitman JD, Shadlen MN, Goldberg ME, Miller KD. One-dimensional dynamics of attention and decision making in LIP. *Neuron* 58: 15–25, 2008.
- Georgopoulos AP, Schwartz AB, Kettner RE. Neuronal population coding of movement direction. *Science* 233: 1416–1419, 1986.

- Gold JI, Shadlen MN. Representation of a perceptual decision in developing oculomotor commands. *Nature* 404: 390–394, 2000.
- Goldman MS. Memory without feedback in a neural network. *Neuron* 61: 621–634, 2009.
- Gschwend O, Abraham NM, Lagier S, Begnaud F, Rodriguez I, Carleton A. Neuronal pattern separation in the olfactory bulb improves odor discrimination learning. *Nat Neurosci* 18: 1474–1482, 2015.
- Haider B, Duque A, Hasenstaub AR, McCormick DA. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci* 26: 4535–4545, 2006.
- Hanks TD, Kiani R, Shadlen MN. A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *Elife* 3, 2014.
- Hanks TD, Kopec CD, Brunton BW, Duan CA, Erlich JC, Brody CD. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* 520: 220–223, 2015.
- Iatan IF. The Fisher's linear discriminant. In: *Advances in Intelligent and Soft Computing. Combining Soft Computing and Statistical Methods in Data Analysis*. Berlin: Heidelberg, Germany: Springer-Verlag Berlin Heidelberg, 2010, vol. 77, p. 345–352.
- Jun JK, Miller P, Hernandez A, Zainos A, Lemus L, Brody CD, Romo R. Heterogeneous population coding of a short-term memory and decision task. *J Neurosci* 30: 916–929, 2010.
- Kanitscheider I, Coen-Cagli R, Pouget A. Origin of information-limiting noise correlations. *Proc Natl Acad Sci USA* 112: E6973–E6982, 2015.
- Kira S, Yang T, Shadlen MN. A neural implementation of Wald's sequential probability ratio test. *Neuron* 85: 861–873, 2015.
- Klampff S, David SV, Yin P, Shamma SA, Maass W. A quantitative analysis of information about past and present stimuli encoded by spikes of A1 neurons. *J Neurophysiol* 108: 1366–1380, 2012.
- Latimer KW, Yates JL, Meister ML, Huk AC, Pillow JW. Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science* 349: 184–187, 2015.
- Lee C, Rohrer WH, Sparks DL. Population coding of saccadic eye movements by neurons in the superior colliculus. *Nature* 332: 357–360, 1988.
- Lee KF, Soares C, Thivierge JP, Béique JC. Correlated synaptic inputs drive dendritic calcium amplification and cooperative plasticity during clustered synapse development. *Neuron* 89: 784–799, 2016.
- Leibold C, Bendels MH. Learning to discriminate through long-term changes of dynamical synaptic transmission. *Neural Comput* 21: 3408–3428, 2009.
- Leutgeb JK, Leutgeb S, Moser MB, Moser EI. Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science* 315: 961–966, 2007.
- Linster C, Cleland T. Decorrelation of odor representations via spike timing-dependent plasticity. *Front Comput Neurosci* 4: 1–11, 2010.
- Ma WJ, Husain M, Bays PM. Changing concepts of working memory. *Nat Neurosci* 17: 347–356, 2014.
- Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. A role for neural integrators in perceptual decision making. *Cereb Cortex* 13: 1257–1269, 2003.
- McHugh TJ, Jones MW, Quinn JJ, Balthasar N, Coppari R, Elmquist JK, Lowell BB, Fanselow MS, Wilson MA, Tonegawa S. Dentate gyrus NMDA receptors mediate rapid pattern separation in the hippocampal network. *Science* 317: 94–99, 2007.
- McLachlan G. *Discriminant Analysis and Statistical Pattern Recognition*. Hoboken, NJ: John Wiley & Sons, 2004.
- Meister ML, Hennig JA, Huk AC. Signal multiplexing and single-neuron computations in lateral intraparietal area during decision-making. *J Neurosci* 33: 2254–2267, 2013.
- Meyers EM, Freedman DJ, Kreiman G, Miller EK, Poggio T. Dynamic population coding of category information in inferior temporal and prefrontal cortex. *J Neurophysiol* 100: 1407–1419, 2008.
- Miri A, Daie K, Arrenberg AB, Baier H, Aksay E, Tank DW. Spatial gradients and multidimensional dynamics in a neural integrator circuit. *Nat Neurosci* 14: 1150–1159, 2011.
- Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, Pouget A. Information-limiting correlations. *Nat Neurosci* 17: 1410–1417, 2014.
- Murphy BK, Miller KD. Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron* 61: 635–648, 2009.
- Murray JD, Bernacchia A, Freedman DJ, Romo R, Wallis JD, Cai X, Padoa-Schioppa C, Pasternak T, Seo H, Lee D, Wang XJ. A hierarchy of intrinsic timescales across primate cortex. *Nat Neurosci* 17: 1661–1665, 2014.
- Nienborg H, Cumming B. Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? *Curr Opin Neurobiol* 20: 376–381, 2010.
- Orhan AE, Ma WJ. Neural population coding of multiple stimuli. *J Neurosci* 35: 3825–3841, 2015.
- Ostojic S. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat Neurosci* 17: 594–600, 2014.
- Park IM, Meister ML, Huk AC, Pillow JW. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat Neurosci* 17: 1395–1403, 2014.
- Pitkow X, Liu S, Angelaki DE, DeAngelis GC, Pouget A. How can single sensory neurons predict behavior? *Neuron* 87: 411–423, 2015.
- Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22: 9475–9489, 2002.
- Santoro A. Reassessing pattern separation in the dentate gyrus. *Front Behav Neurosci* 7: 96, 2013.
- Shadlen MN, Britten KH, Newsome WT, Movshon JA. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* 16: 1486–1510, 1996.
- Shadlen MN, Kiani R. Decision making as a window on cognition. *Neuron* 80: 791–806, 2013.
- Shafi M, Zhou Y, Quintana J, Chow C, Fuster J, Bodner M. Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience* 146: 1082–1108, 2007.
- Shamir M, Sompolinsky H. Implications of neuronal diversity on population coding. *Neural Comput* 18: 1951–1986, 2006.
- Shu Y, Hasenstaub A, McCormick DA. Turning on and off recurrent balanced cortical activity. *Nature* 423: 288–293, 2003.
- Singh R, Eliasmith C. Higher-dimensional neurons explain the tuning and dynamics of working memory cells. *J Neurosci* 26: 3667–3678, 2006.
- Smith PL, Ratcliff R. Psychology and neurobiology of simple decisions. *Trends Neurosci* 27: 161–168, 2004.
- Sompolinsky H, Yoon H, Kang K, Shamir M. Population coding in neuronal systems with correlated noise. *Phys Rev E Stat Nonlin Soft Matter Phys* 64: 051904, 2001.
- Swadlow HA. Physiological properties of individual cerebral axons studied in vivo for as long as one year. *J Neurophysiol* 54: 1346–1362, 1985.
- Thivierge JP, Comas R, Longtin A. Attractor dynamics in local neuronal networks. *Front Neural Circuits* 8: 22, 2014.
- Thura D, Cisek P. Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron* 81: 1401–1416, 2014.
- Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer, 2000.
- Vincent K, Tauskela JS, Thivierge JP. Extracting functionally feedforward networks from a population of spiking neurons. *Front Comput Neurosci* 6: 86, 2012.
- Yassa MA, Stark CE. Pattern separation in the hippocampus. *Trends Neurosci* 34: 515–525, 2011.
- Zandbelt B, Purcell BA, Palmeri TJ, Logan GD, Schall JD. Response times from ensembles of accumulators. *Proc Natl Acad Sci USA* 111: 2848–2853, 2014.
- Zohary E, Shadlen MN, Newsome WT. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* 370: 140–143, 1994.
- Zylberberg J, Cafaro J, Turner MH, Shea-Brown E, Rieke F. Direction-selective circuits shape noise to ensure a precise population code. *Neuron* 89: 369–383, 2016.

Bibliography

- Albe-Fessard, D. (1953). Exploration of certain activities of the motor cortex in cat with microelectrodes; endo-somatic leads. *Journal of Physiology (Paris)*, 45, 14-16.
- Amit, D. J. & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, 7(3), 237–252. <https://doi.org/10.1093/cercor/7.3.237>
- Amitai, Y., & Connors, B. W. (1995). Intrinsic physiology and morphology of single neurons in neocortex. In *The barrel cortex of rodents* (pp. 299-331). Boston, MA: Springer US.
- Anastasiades, P. G. & Carter, A. G. (2021). Circuit organization of the rodent medial prefrontal cortex. *Trends in Neurosciences*, 44(7), 550–563. <https://doi.org/10.1016/j.tins.2021.03.006>
- Anastasiades, P. G., Collins, D. P. & Carter, A. G. (2021). Mediodorsal and Ventromedial Thalamus Engage Distinct L1 Circuits in the Prefrontal Cortex. *Neuron*, 109(2), 314-330.e4. <https://doi.org/10.1016/j.neuron.2020.10.031>
- Baker, A., Kalmbach, B., Morishima, M., Kim, J., Juavinett, A., Li, N. & Dembrow, N. (2018). Specialized Subpopulations of Deep-Layer Pyramidal Neurons in the Neocortex: Bridging Cellular Properties to Functional Consequences. *The Journal of Neuroscience*, 38(24), 5441–5455. <https://doi.org/10.1523/jneurosci.0150-18.2018>
- Bakkum, D. J., Obien, M. E. J., Radivojevic, M., Jäckel, D., Frey, U., Takahashi, H. & Hierlemann, A. (2019). The Axon Initial Segment is the Dominant Contributor to the

Neuron's Extracellular Electrical Potential Landscape. *Advanced Biosystems*, 3(2), 1800308. <https://doi.org/10.1002/adbi.201800308>

Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D. & Buzsáki, G. (2004).

Characterization of Neocortical Principal Cells and Interneurons by Network Interactions and Extracellular Features. *Journal of Neurophysiology*, 92(1), 600–608. <https://doi.org/10.1152/jn.01170.2003>

Becchetti, A., Gullo, F., Bruno, G., Dossi, E., Lecchi, M. & Wanke, E. (2012). Exact distinction of excitatory and inhibitory neurons in neural networks: a study with GFP-GAD67 neurons optically and electrophysiologically recognized on multielectrode arrays. *Frontiers in Neural Circuits*, 6, 63. <https://doi.org/10.3389/fncir.2012.00063>

Berdondini, L., Imfeld, K., Maccione, A., Tedesco, M., Neukom, S., Koudelka-Hep, M. & Martinoia, S. (2009). Active pixel sensor array for high spatio-temporal resolution electrophysiological recordings from single cell to large-scale neuronal networks. *Lab on a Chip*, 9(18), 2644–2651. <https://doi.org/10.1039/b907394a>

Berndt, A., Yizhar, O., Gunaydin, L. A., Hegemann, P. & Deisseroth, K. (2009). Bi-stable neural state switches. *Nature Neuroscience*, 12(2), 229–234. <https://doi.org/10.1038/nn.2247>

Blanche, T. J., & Swindale, N. V. (2006). Nyquist interpolation improves neuron yield in multiunit recordings. *Journal of neuroscience methods*, 155(1), 81-91. <https://doi.org/10.1016/j.jneumeth.2005.12.031>

Bugeon, S., Duffield, J., Dipoppa, M., Ritoux, A., Prankerd, I., Nicoloutsopoulos, D., Orme, D., Shinn, M., Peng, H., Forrest, H., Viduolyte, A., Reddy, C. B., Isogai, Y., Carandini, M.,

- Harris, K. D. A transcriptomic axis predicts state modulation of cortical interneurons. (2022). *Nature* **607**, 330–338. <https://doi.org/10.1038/s41586-022-04915-7>
- Bullmann, T., Radivojevic, M., Huber, S. T., Deligkaris, K., Hierlemann, A. & Frey, U. (2019). Large-Scale Mapping of Axonal Arbors Using High-Density Microelectrode Arrays. *Frontiers in Cellular Neuroscience*, 13, 404. <https://doi.org/10.3389/fncel.2019.00404>
- Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, 7(5), 446–451. <https://doi.org/10.1038/nn1233>
- Buzsáki G. (2010). Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron* 68:362–385. doi:10.1016/j.neuron.2010.09.023
- Buzsáki, G., Anastassiou, C. A. & Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, 13(6), 407–420. <https://doi.org/10.1038/nrn3241>
- Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1), 1-27.
- Calvin, W. H. & Sypert, G. W. (1976). Fast and slow pyramidal tract neurons: an intracellular analysis of their contrasting repetitive firing properties in the cat. *Journal of Neurophysiology*, 39(2), 420–434. <https://doi.org/10.1152/jn.1976.39.2.420>
- Cauli, B., Porter, J. T., Tsuzuki, K., Lambolez, B., Rossier, J., Quenet, B. & Audinat, E. (2000). Classification of fusiform neocortical interneurons based on unsupervised clustering.

Proceedings of the National Academy of Sciences, 97(11), 6144–6149. <https://doi.org/10.1073/pnas.97.11.6144>

Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H. & Moore, C. I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, 459(7247), 663–667. <https://doi.org/10.1038/nature08002>

Carlson, D. & Carin, L. (2019). Continuing progress of spike sorting in the era of big data. *Current Opinion in Neurobiology*, 55, 90–96. <https://doi.org/10.1016/j.conb.2019.02.007>

Chung, J. E., Magland, J. F., Barnett, A. H., Tolosa, V. M., Tooker, A. C., Lee, K. Y., Shah, K. G., Felix, S. H., Frank, L. M. & Greengard, L. F. (2017). A Fully Automated Approach to Spike Sorting. *Neuron*, 95(6), 1381–1394.e6. <https://doi.org/10.1016/j.neuron.2017.08.030>

Cohen, J.Y., Amoroso, M.W. & Uchida, N. (2015). Serotonergic neurons signal reward and punishment on multiple timescales. *eLife* 4:e06346.

Cohen, L. & Mizrahi, A. (2015). Plasticity during Motherhood: Changes in Excitatory and Inhibitory Layer 2/3 Neurons in Auditory Cortex. *The Journal of Neuroscience*, 35(4), 1806–1815. <https://doi.org/10.1523/jneurosci.1786-14.2015>

Connors, B. W., Gutnick, M. J. & Prince, D. A. (1982). Electrophysiological properties of neocortical neurons in vitro. *Journal of Neurophysiology*, 48(6), 1302–1320. <https://doi.org/10.1152/jn.1982.48.6.1302>

Connors, B. W. & Gutnick, M. J. (1990). Intrinsic firing patterns of diverse neocortical neurons. *Trends in Neurosciences*, 13(3), 99–104. [https://doi.org/10.1016/0166-2236\(90\)90185-d](https://doi.org/10.1016/0166-2236(90)90185-d)

Constantinidis, C. & Goldman-Rakic, P. S. (2002). Correlated Discharges Among Putative Pyramidal Neurons and Interneurons in the Primate Prefrontal Cortex. *Journal of Neurophysiology*, 88(6), 3487–3497. <https://doi.org/10.1152/jn.00188.2002>

Contreras, D. (2004). Electrophysiological classes of neocortical neurons. *Neural Networks*, 17(5–6), 633–646. <https://doi.org/10.1016/j.neunet.2004.04.003>

Deisseroth, K., Feng, G., Majewska, A. K., Miesenböck, G., Ting, A. & Schnitzer, M. J. (2006). Next-Generation Optical Technologies for Illuminating Genetically Targeted Brain Circuits. *The Journal of Neuroscience*, 26(41), 10380–10386. <https://doi.org/10.1523/jneurosci.3863-06.2006>

Emson, P.C., Lindvall, O., 1979. Distribution of putative neurotransmitters in the neocortex. *Neuroscience* 4, 1–30.

Fiscella M, Farrow K, Jones IL, Jäckel D, Müller J, Frey U, Bakkum DJ, Hantz P, Roska B, Hierlemann A. 2012. Recording from defined populations of retinal ganglion cells using a high-density CMOS-integrated microelectrode array with real-time switchable electrode selection. *J Neurosci Meth* **211**:103–113. <https://doi.org/10.1016/j.jneumeth.2012.08.017>

Fuster, J. (2015). The prefrontal cortex. Academic Press.

Gold, C., Henze, D. A., Koch, C. & Buzsáki, G. (2006). On the Origin of the Extracellular Action Potential Waveform: A Modeling Study. *Journal of Neurophysiology*, 95(5), 3113–3128.

<https://doi.org/10.1152/jn.00979.2005>

González-Burgos, G., Krimer, L. S., Povysheva, N. V., Barrionuevo, G. & Lewis, D. A. (2005). Functional Properties of Fast Spiking Interneurons and Their Synaptic Connections With Pyramidal Cells in Primate Dorsolateral Prefrontal Cortex. *Journal of Neurophysiology*, 93(2), 942–953. <https://doi.org/10.1152/jn.00787.2004>

Gray, C. M. & McCormick, D. A. (1996). Chattering Cells: Superficial Pyramidal Neurons Contributing to the Generation of Synchronous Oscillations in the Visual Cortex. *Science*, 274(5284), 109–113. <https://doi.org/10.1126/science.274.5284.109>

Hafizi, H., Nigam, S., Barnathan, J., Ren, N., Stevenson, I. H., Masmanidis, S. C., Newman, E. L., Sporns, O. & Beggs, J. M. (2022). Inhibition-Dominated Rich-Club Shapes Dynamics in Cortical Microcircuits. *bioRxiv*, 2021.05.07.443074. <https://doi.org/10.1101/2021.05.07.443074>

Hagen, E., Ness, T.V., Khosrowshahi, A., Sorensen, C., Fyhn, M., Halfting, T., Franke, F., Einevoll, G.T. (2015). ViSAPy: A Python tool for biophysics-based generation of virtual spiking activity for evaluation of spike-sorting algorithms. *Journal of Neuroscience Methods* 245:182-204.

Harkin, E.F., Lynn, M.B., Payeur, A., Boucher, J.-F., Caya-Bissonnette, L., Cyr, D., Stewart, C., Longtin, A., Naud, R. & Beique, J.-C. (2023). Temporal derivative computation in the dorsal raphe network revealed by an experimentally-driven augmented integrate- and-fire modeling framework. *eLife* 12:e72951.

- Harris, K. D. & Shepherd, G. M. G. (2015). The neocortical circuit: themes and variations. *Nature Neuroscience*, 18(2), 170–181. <https://doi.org/10.1038/nn.3917>
- Hasenstaub, A., Shu, Y., Haider, B., Kraushaar, U., Duque, A. & McCormick, D. A. (2005). Inhibitory Postsynaptic Potentials Carry Synchronized Frequency Information in Active Cortical Networks. *Neuron*, 47(3), 423–435. <https://doi.org/10.1016/j.neuron.2005.06.016>
- Hemberger, M., Shein-Idelson, M., Pammer, L. & Laurent, G. (2019). Reliable Sequential Activation of Neural Assemblies by Single Pyramidal Cells in a Three-Layered Cortex. *Neuron*, 104(2), 353–369.e5. <https://doi.org/10.1016/j.neuron.2019.07.017>
- Hennequin, G., Ahmadian, Y., Rubin, D. B., Lengyel, M. & Miller, K. D. (2018). The Dynamical Regime of Sensory Cortex: Stable Dynamics around a Single Stimulus-Tuned Attractor Account for Patterns of Noise Variability. *Neuron*, 98(4), 846–860.e5. <https://doi.org/10.1016/j.neuron.2018.04.017>
- Henze, D. A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K. D. & Buzsáki, G. (2000). Intracellular Features Predicted by Extracellular Recordings in the Hippocampus In Vivo. *Journal of Neurophysiology*, 84(1), 390–400. <https://doi.org/10.1152/jn.2000.84.1.390>
- Hilgen, G., Sorbaro, M., Pirmoradian, S., Muthmann, J.-O., Kepiro, I. E., Ullo, S., Ramirez, C. J., Encinas, A. P., Maccione, A., Berdondini, L., Murino, V., Sona, D., Zanacchi, F. C., Sernagor, E. & Hennig, M. H. (2017). Unsupervised Spike Sorting for Large-Scale, High-Density Multielectrode Arrays. *Cell Reports*, 18(10), 2521–2532. <https://doi.org/10.1016/j.celrep.2017.02.038>

- Hill, D. N., Mehta, S. B. & Kleinfeld, D. (2011). Quality Metrics to Accompany Spike Sorting of Extracellular Signals. *The Journal of Neuroscience*, 31(24), 8699–8705. <https://doi.org/10.1523/jneurosci.0971-11.2011>
- Hirsch, H. E. & Robins, E. (1962). Distribution of γ -aminobutyric acid in the layers of the cerebral and cerebellar cortex. Implications for its physiological role. *Journal of Neurochemistry*, 9(1), 63–70. <https://doi.org/10.1111/j.1471-4159.1962.tb07494.x>
- Houser, C. R., Hendry, S. H. C., Jones, E. G. & Vaughn, J. E. (1983). Morphological diversity of immunocytochemically identified GABA neurons in the monkey sensory-motor cortex. *Journal of Neurocytology*, 12(4), 617–638. <https://doi.org/10.1007/bf01181527>
- Imfeld, K., Neukom, S., Maccione, A., Bornat, Y., Martinoia, S., Farine, P.-A., Koudelka-Hep, M. & Berdondini, L. (2008). Large-Scale, High-Resolution Data Acquisition System for Extracellular Recording of Electrophysiological Activity. *IEEE Transactions on Biomedical Engineering*, 55(8), 2064–2073. <https://doi.org/10.1109/tbme.2008.919139>
- Insel, N. & Barnes, C. A. (2015). Differential Activation of Fast-Spiking and Regular-Firing Neuron Populations During Movement and Reward in the Dorsal Medial Frontal Cortex. *Cerebral Cortex*, 25(9), 2631–2647. <https://doi.org/10.1093/cercor/bhu062>
- Jia, X., Siegle, J. H., Bennett, C., Gale, S. D., Denman, D. J., Koch, C. & Olsen, S. R. (2019). High-density extracellular probes reveal dendritic backpropagation and facilitate neuron classification. *Journal of Neurophysiology*, 121(5), 1831–1847. <https://doi.org/10.1152/jn.00680.2018>

- Joshi, J., Rubart, M. & Zhu, W. (2020). Optogenetics: Background, Methodological Advances and Potential Applications for Cardiovascular Research and Medicine. *Frontiers in Bioengineering and Biotechnology*, 7, 466. <https://doi.org/10.3389/fbioe.2019.00466>
- Jun, J. J., Mitelut, C., Lai, C., Gratiy, S. L., Anastassiou, C. A. & Harris, T. D. (2017). Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction. *bioRxiv*, 101030. <https://doi.org/10.1101/101030>
- Jun, J.J., Steinmetz, N., Siegle, J. et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236.
- Kajikawa, Y. & Schroeder, C. E. (2011). How Local Is the Local Field Potential? *Neuron*, 72(5), 847–858. <https://doi.org/10.1016/j.neuron.2011.09.029>
- Kawaguchi, Y. & Kubota, Y. (1993). Correlation of physiological subgroupings of nonpyramidal cells with parvalbumin- and calbindinD28k-immunoreactive neurons in layer V of rat frontal cortex. *Journal of Neurophysiology*, 70(1), 387–396. <https://doi.org/10.1152/jn.1993.70.1.387>
- Kawaguchi, Y. (1995). Physiological subgroups of nonpyramidal cells with specific morphological characteristics in layer II/III of rat frontal cortex. *The Journal of Neuroscience*, 15(4), 2638–2655. <https://doi.org/10.1523/jneurosci.15-04-02638.1995>
- Kawaguchi, Y. & Kubota, Y. (1998). Neurochemical features and synaptic connections of large physiologically-identified GABAergic cells in the rat frontal cortex. *Neuroscience*, 85(3), 677–701. [https://doi.org/10.1016/s0306-4522\(97\)00685-4](https://doi.org/10.1016/s0306-4522(97)00685-4)

- Kim, Y., Yang, G. R., Pradhan, K., Venkataraju, K. U., Bota, M., Molino, L. C. G. del, Fitzgerald, G., Ram, K., He, M., Levine, J. M., Mitra, P., Huang, Z. J., Wang, X.-J. & Osten, P. (2017). Brain-wide Maps Reveal Stereotyped Cell type-Based Cortical Architecture and Subcortical Sexual Dimorphism. *Cell*, 171(2), 456-469.e22. <https://doi.org/10.1016/j.cell.2017.09.020>
- Lak, A., Okun, M., Moss, M. M., Gurnani, H., Farrell, K., Wells, M. J., Reddy, C. B., Kepecs, A., Harris, K. D. & Carandini, M. (2020). Dopaminergic and prefrontal basis of learning from sensory confidence and reward value. *Neuron*, 105(4), 700-711.
- Lee, J., Carlson, D., Shokri, H., Yao, W., Goetz, G., Hagen, E., Batty, E., Chichilnisky, E., Einevoll, G. & Paninski, L. (2017). YASS: Yet Another Spike Sorter. *bioRxiv*, 151928. <https://doi.org/10.1101/151928>
- Lemon, R. N., Baker, S. N. & Kraskov, A. (2021). Classification of Cortical Neurons by Spike Shape and the Identification of Pyramidal Neurons. *Cerebral Cortex*, 31(11), bhab147-. <https://doi.org/10.1093/cercor/bhab147>
- Lewicki, M. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4), R53–R78. <https://doi.org/10.1088/0954-898x/9/4/001>
- Li, C.-L. (1959). Cortical intracellular potentials and their responses to strychnine. *Journal of Neurophysiology*, 24(4).

- Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M. & Einevoll, G. T. (2011). Modeling the Spatial Reach of the LFP. *Neuron*, 72(5), 859–872. <https://doi.org/10.1016/j.neuron.2011.11.006>
- Litwin-Kumar, A., Rosenbaum, R. & Doiron, B. (2016). Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes. *Journal of Neurophysiology*, 115(3), 1399–1409. <https://doi.org/10.1152/jn.00732.2015>
- Liu, Y. H, Smith, S., Mihalas, S. & Sumbul, U. (2021). Cell type–specific neuromodulation guides synaptic credit assignment in a spiking neural network. *PNAS*, 118, e2111821118.
- Madisen, L., Mao, T., Koch, H., Zhuo, J., Berenyi, A., Fujisawa, S., Hsu, Y.-W. A., Garcia, A. J., Gu, X., Zanella, S., Kidney, J., Gu, H., Mao, Y., Hooks, B. M., Boyden, E. S., Buzsáki, G., Ramirez, J. M., Jones, A. R., Svoboda, K., ... Zeng, H. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neuroscience*, 15(5), 793–802. <https://doi.org/10.1038/nn.3078>
- Magland, J., Jun, J. J., Lovero, E., Morley, A. J., Hurwitz, C. L., Buccino, A. P., Garcia, S. & Barnett, A. H. (2020). SpikeForest, reproducible web-facing ground-truth validation of automated neural spike sorters. *ELife*, 9, e55167. <https://doi.org/10.7554/elife.55167>
- Mahrach, A., Chen, G., Li, N., Vreeswijk, C. van & Hansel, D. (2020). Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. *eLife*, 9, e49967. <https://doi.org/10.7554/elife.49967>

- Mai, Q. & Zou, H. (2013). A Note On the Connection and Equivalence of Three Sparse Linear Discriminant Analysis Methods. *Technometrics*, 55(2), 243–246. <https://doi.org/10.1080/00401706.2012.746208>
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G. & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10), 793–807. <https://doi.org/10.1038/nrn1519>
- Marre, O., Amodei, D., Deshmukh, N., Sadeghi, K., Soo, F., Holy, T. E. & Berry, M. J. (2012). Mapping a Complete Neural Population in the Retina. *The Journal of Neuroscience*, 32(43), 14859–14873. <https://doi.org/10.1523/jneurosci.0723-12.2012>
- McCormick, D. A., Connors, B. W., Lighthall, J. W. & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, 54(4), 782–806. <https://doi.org/10.1152/jn.1985.54.4.782>
- Mechler, F., Victor, J. D., Ohiorhenuan, I., Schmid, A. M. & Hu, Q. (2011). Three-dimensional localization of neurons in cortical tetrode recordings. *Journal of Neurophysiology*, 106(2), 828–848. <https://doi.org/10.1152/jn.00515.2010>
- Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1), 167–202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. (2007). Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4. *Neuron*, 55(1), 131–141. <https://doi.org/10.1016/j.neuron.2007.06.018>

- Moore, A. K. & Wehr, M. (2013). Parvalbumin-Expressing Inhibitory Interneurons in Auditory Cortex Are Well-Tuned for Frequency. *The Journal of Neuroscience*, 33(34), 13713–13723. <https://doi.org/10.1523/jneurosci.0663-13.2013>
- Mountcastle, V. B., H., T. W. & H., S. (1969). Cortical neuronal mechanisms in flutter-vibration studied in unanesthetized monkeys. Neuronal periodicity and frequency discrimination. *Journal of Neurophysiology*, 32(3), 775–806. <https://doi.org/10.1152/jn.1963.26.5.775>
- Murphy, B. K. & Miller, K. D. (2009). Balanced Amplification: A New Mechanism of Selective Amplification of Neural Activity Patterns. *Neuron*, 61(4), 635–648. <https://doi.org/10.1016/j.neuron.2009.02.005>
- Muthmann, J.-O., Amin, H., Sernagor, E., Maccione, A., Panas, D., Berdondini, L., Bhalla, U. S. & Hennig, M. H. (2015). Spike Detection for Large Neural Populations Using High Density Multielectrode Arrays. *Frontiers in Neuroinformatics*, 9, 28. <https://doi.org/10.3389/fninf.2015.00028>
- Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M. & McCormick, D. A. (2003). Electrophysiological Classes of Cat Primary Visual Cortical Neurons In Vivo as Revealed by Quantitative Analyses. *Journal of Neurophysiology*, 89(3), 1541–1566. <https://doi.org/10.1152/jn.00580.2002>
- Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J. & Frey, U. (2015). Revealing neuronal function through microelectrode array recordings. *Frontiers in Neuroscience*, 8, 423. <https://doi.org/10.3389/fnins.2014.00423>

- Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & D., H. K. (2016). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, 061481. <https://doi.org/10.1101/061481>
- Pachitariu, M., Sridhar, S. & Stringer, C. (2023). Solving the spike sorting problem with Kilosort. *bioRxiv*, 2023.01.07.523036. <https://doi.org/10.1101/2023.01.07.523036>
- Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M. & Harris, K.D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. *Advances in Neural Information Processing Systems*, 29.
- Peters, A. J., Fabre, J. M., Steinmetz, N. A., Harris, K. D., & Carandini, M. (2021). Striatal activity topographically reflects cortical activity. *Nature*, 591(7850), 420-425.
- Pettersen, K. H. & Einevoll, G. T. (2008). Amplitude Variability and Extracellular Low-Pass Filtering of Neuronal Spikes. *Biophysical Journal*, 94(3), 784–802. <https://doi.org/10.1529/biophysj.107.111179>
- Peyrache, A., Dehghani, N., Eskandar, E. N., Madsen, J. R., Anderson, W. S., Donoghue, J. A., Hochberg, L. R., Halgren, E., Cash, S. S. & Destexhe, A. (2012). Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. *Proceedings of the National Academy of Sciences*, 109(5), 1731–1736. <https://doi.org/10.1073/pnas.1109895109>
- Peyrache, A. & Destexhe, A. (2019). Electrophysiological monitoring of inhibition in mammalian species, from rodents to humans. *Neurobiology of Disease*, 130, 104500. <https://doi.org/10.1016/j.nbd.2019.104500>

- Phillips, C. G. (1956). Intracellular records from Betz cells in the cat. *Quarterly Journal of Experimental Physiology and Cognate Medical Sciences*, 41(1), 58–69. <https://doi.org/10.1113/expphysiol.1956.sp001163>
- Prentice, J. S., Homann, J., Simmons, K. D., Tkačik, G., Balasubramanian, V. & Nelson, P. C. (2011). Fast, Scalable, Bayesian Spike Identification for Multi-Electrode Arrays. *PLoS ONE*, 6(7), e19884. <https://doi.org/10.1371/journal.pone.0019884>
- Rajdl, K., Lansky, P. & Kostal, L. (2020). Fano Factor: A Potentially Useful Information. *Frontiers in Computational Neuroscience*, 14, 569049. <https://doi.org/10.3389/fncom.2020.569049>
- Ren, N., Ito, S., Hafizi, H., Beggs, J. M. & Stevenson, I. H. (2020). Model-based detection of putative synaptic connections from spike recordings with latency and type constraints. *Journal of Neurophysiology*, 124(6), 1588–1604. <https://doi.org/10.1152/jn.00066.2020>
- Rey, H. G., Pedreira, C. & Quiroga, R. Q. (2015). Past, present and future of spike sorting techniques. *Brain Research Bulletin*, 119(Pt B), 106–117. <https://doi.org/10.1016/j.brainresbull.2015.04.007>
- Riquelme, J. L., Hemberger, M., Laurent, G. & Gjorgjieva, J. (2023) Single spikes drive sequential propagation and routing of activity in a cortical network *eLife* 12:e79928. <https://doi.org/10.7554/eLife.79928>.
- Robbins, T. W. & Arnsten, A. F. T. (2009). The Neuropsychopharmacology of Fronto-Executive Function: Monoaminergic Modulation. *Annual Review of Neuroscience*, 32(1), 267–287. <https://doi.org/10.1146/annurev.neuro.051508.135535>

- Rossant, C., Kadir, S. N., Goodman, D. F. M., Schulman, J., Hunter, M. L. D., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., Tolia, A. S., Solomon, S., Buzsaki, G., Carandini, M., Harris K. D. 2016. Spike sorting for large, dense electrode arrays. *Nat Neurosci* **19**:634–641. doi:10.1038/nn.4268
- Rubin, R., Abbott, L. F. & Sompolinsky, H. (2017). Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proceedings of the National Academy of Sciences*, 114(44), E9366–E9375. <https://doi.org/10.1073/pnas.1705841114>
- Ruz, I. D. & Schultz, S. R. (2014). Localising and classifying neurons from high density MEA recordings. *Journal of Neuroscience Methods*, 233, 115–128. <https://doi.org/10.1016/j.jneumeth.2014.05.037>
- Sakata, S. & Harris, K. D. (2009). Laminar Structure of Spontaneous and Sensory-Evoked Population Activity in Auditory Cortex. *Neuron*, 64(3), 404–418. <https://doi.org/10.1016/j.neuron.2009.09.020>
- Sanzeni, A., Akitake, B., Goldbach, H. C., Leedy, C. E., Brunel, N. & Histed, M. H. (2020). Inhibition stabilization is a widespread property of cortical networks. *ELife*, 9, e54875. <https://doi.org/10.7554/elife.54875>
- Schuman, B., Machold, R. P., Hashikawa, Y., Fuzik, J., Fishell, G. J. & Rudy, B. (2019). Four Unique Interneuron Populations Reside in Neocortical Layer 1. *The Journal of Neuroscience*, 39(1), 125–139. <https://doi.org/10.1523/jneurosci.1613-18.2018>

- Shein-Idelson, M., Pammer, L., Hemberger, M. & Laurent, G. (2017). Large-scale mapping of cortical synaptic projections with extracellular electrode arrays. *Nature Methods*, 14(9), 882–890. <https://doi.org/10.1038/nmeth.4393>
- Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M. & Buzsáki, G. (2008). Entrainment of Neocortical Neurons and Gamma Oscillations by the Hippocampal Theta Rhythm. *Neuron*, 60(4), 683–697. <https://doi.org/10.1016/j.neuron.2008.09.014>
- Stafstrom, C. E., Schwindt, P. C. & Crill, W. E. (1984). Repetitive firing in layer V neurons from cat neocortex in vitro. *Journal of Neurophysiology*, 52(2), 264–277. <https://doi.org/10.1152/jn.1984.52.2.264>
- Stark, E., Eichler, R., Roux, L., Fujisawa, S., Rotstein, H. G. & Buzsáki, G. (2013). Inhibition-Induced Theta Resonance in Cortical Circuits. *Neuron*, 80(5), 1263–1276. <https://doi.org/10.1016/j.neuron.2013.09.033>
- Stringer, C., Pachitariu, M., Steinmetz, N. A., Okun, M., Bartho, P., Harris, K. D., Sahani, M., Lesica, N. A. (2016). Inhibitory control of correlated intrinsic variability in cortical networks. *Elife*, 5, e19695. <https://doi.org/10.7554/eLife.19695>
- Sylwestrak, E. L., Jo, Y. J., Vesuna, S., Wang, X., Holcomb, B., Tien, R. H., Kim, D. K., Fenno, L., Ramakrishnan, C., Allen, W. E., Chen, R., Shenoy, K. V., Sussillo, D. & Deisseroth, K. (2022). Cell type-specific population dynamics of diverse reward computations. *Cell* 185, 3568–3587.
- Tamura, H., Kaneko, H., Kawasaki, K. & Fujita, I. (2004). Presumed Inhibitory Neurons in the Macaque Inferior Temporal Cortex: Visual Response Properties and Functional

Interactions With Adjacent Neurons. *Journal of Neurophysiology*, 91(6), 2782–2796.

<https://doi.org/10.1152/jn.01267.2003>

Takahashi, K. (1965). Slow and fast groups of pyramidal tract cells and their respective membrane properties. *Journal of Neurophysiology*, 5(28), 908–924.

Tasaki, Polley, I. & F, O. (1954). Action potentials from individual elements in cat geniculate and striate cortex. *Journal of Neurophysiology*, 5(17), 454–474. <https://doi.org/10.1152/jn.1949.12.4.277>

Thivierge, J.-P. (2014). Scale-free and economical features of functional connectivity in neuronal networks. *Physical Review E*, 90(2), 022721. <https://doi.org/10.1103/physreve.90.022721>

Thivierge, J. P., Giraud, E., Lynn, M., & Théberge Charbonneau, A. (2022). Key role of neuronal diversity in structured reservoir computing. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 32(11), 113130.

Tremblay, R., Lee, S. & Rudy, B. (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron*, 91(2), 260–292. <https://doi.org/10.1016/j.neuron.2016.06.033>

Tsodyks, M. V., Skaggs, W. E., Sejnowski, T. J. & McNaughton, B. L. (1997). Paradoxical Effects of External Modulation of Inhibitory Interneurons. *The Journal of Neuroscience*, 17(11), 4382–4388. <https://doi.org/10.1523/jneurosci.17-11-04382.1997>

Ueta, Y., Sohn, J., Agahari, F. A., Im, S., Hirai, Y., Miyata, M. & Kawaguchi, Y. (2019). Ipsi- and contralateral corticocortical projection-dependent subcircuits in layer 2 of the rat frontal

cortex. *Journal of Neurophysiology*, 122(4), 1461–1472. <https://doi.org/10.1152/jn.00333.2019>

Vendramin, L., Campello, R. J., & Hruschka, E. R. (2009, April). On the comparison of relative clustering validity criteria. In Proceedings of the 2009 siam international conference on data mining (pp. 733-744). Society for Industrial and Applied Mathematics.

Vigneswaran, G., Kraskov, A. & Lemon, R. N. (2011). Large Identified Pyramidal Cells in Macaque Motor and Premotor Cortex Exhibit “Thin Spikes”: Implications for Cell Type Classification. *The Journal of Neuroscience*, 31(40), 14235–14242. <https://doi.org/10.1523/jneurosci.3142-11.2011>

Vincent, K., Tauskela, J. S. & Thivierge, J.-P. (2012). Extracting functionally feedforward networks from a population of spiking neurons. *Frontiers in Computational Neuroscience*, 6, 86. <https://doi.org/10.3389/fncom.2012.00086>

Vincent, K., Tauskela, J. S., Mealing, G. A. & Thivierge, J.-P. (2013). Altered Network Communication Following a Neuroprotective Drug Treatment. *PLoS ONE*, 8(1), e54478. <https://doi.org/10.1371/journal.pone.0054478>

Vreeswijk, C. van & Sompolinsky, H. (1996). Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science*, 274(5293), 1724–1726. <https://doi.org/10.1126/science.274.5293.1724>

- Vreeswijk, C. van & Sompolinsky, H. (1998). Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*, 10(6), 1321–1371. <https://doi.org/10.1162/089976698300017214>
- Wilson, F. A., O'Scalaidhe, S. P. & Goldman-Rakic, P. S. (1994). Functional synergism between putative gamma-aminobutyrate-containing neurons and pyramidal neurons in prefrontal cortex. *Proceedings of the National Academy of Sciences*, 91(9), 4009–4013. <https://doi.org/10.1073/pnas.91.9.4009>
- Xu, R. & D. Wunsch. (2005). Survey of Clustering Algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–678. <https://doi.org/10.1109/tnn.2005.845141>
- Yael, D. & Bar-Gad, I. (2017). Filter based phase distortions in extracellular spikes. *PLOS ONE*, 12(3), e0174790. <https://doi.org/10.1371/journal.pone.0174790>
- Yger, P., Spampinato, G. L., Esposito, E., Lefebvre, B., Deny, S., Gardella, C., Stimberg, M., Jetter, F., Zeck, G., Picaud, S., Duebel, J. & Marre, O. (2018). A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *eLife*, 7, e34518. <https://doi.org/10.7554/elife.34518>