Difference that makes the mind

Nishant Joshi

his project has received funding from the European Union' s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 860949.

Author: Nishant Joshi
Title: Working Title

Radboud Dissertations Series ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl

Design: Nishant Joshi
Cover: TBD
Printing: DPN Rikken/Pumbo

ISBN: 9789493296374
DOI: 10.54195/ 9789493296374
Free download at: www.boekenbestellen.nl/radboud-university-press/dissertations

© Nishant Joshi, 2025

This is an Open Access book published under the terms of Creative Commons Attribution-Noncommercial-NoDerivatives International license (CC BY-NC-ND 4.0). This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator, see http://creativecommons.org/licenses/by-nc-nd/4.0/.

Diversity creates the mind

Proefschrift ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.M. Sanders,

volgens besluit van het college voor promoties

in het openbaar te verdedigen op

maandag 15 april 2024

om 16:30 uur precies

door

Nishant Joshi

geboren op 26 January 1994

te Varanasi, India

Promotor:

Prof. dr. T. Celikel

Copromotoren:

Dr. F. Zeldenrust

Manuscriptcommissie:

TBD.

TBD.

TBD.

TBD.

TBD.

Funding

This project has received funding from the European Union' s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 860949.

Table of Contents

Chapter 1

General Introduction

Take a look at your palm, go ahead and examine it. What do you see? Five fingers (I hope?). Are they all the same? Now, look around you and observe the trees, the leaves on them, each one with a unique shape and size, isn't it? The people surrounding you have different skin colors, facial features, and heights. Have you ever wondered why nothing is exactly the same? Variance is an essential feature of nature; it permeates everything, from our thoughts and dreams to our language and lives. Without this diversity, life would be sterile, perhaps even impossible.

Throughout my PhD journey, I have sought to understand the origins and implications of these differences at the most fundamental level: the neurons inside our brains. By exploring the variability among these building blocks of thought, ideas, dreams, language, and personality, I aim to shed light on how diversity shapes not only our individual experiences but also the very fabric of existence.

Neuronal heterogeneity through the lens of the past

The mammalian brain is an extraordinarily complex organ, composed of an immense variety of neuronal cell types. This diversity is evident across multiple dimensions, exploring this diversity has been a central tenet in neuroscience, since its inception, dating back to Ramon y Cajal (Fishell and Heintz (2013); Huang and Paul (2019); Markram et al. (2004); Mukamel and Ngai (2019); Nelson et al. (2006); pet (2008); Sanes and Masland (2015); Seung and Sümbül (2014); Somogyi and Klausberger (2005); Yuste et al. (2020); Zeng (2022); Zeng and Sanes (2017)). I would like to take the reader through what we already know about neuronal heterogeneity via different modalities, ranging from shapes to molecular composition.

Morphology

Neurons differ dramatically in their shapes and structures. Some, like pyramidal neurons, exhibit long apical dendrites and a characteristic triangular soma, while others,

such as inter-neurons, display more compact and intricate branching patterns. These morphological differences are closely linked to the specific roles neurons play within neural circuits (Liu et al. (2024)).

Electrophysiology

Neuronal diversity is also exhibited in the ways neurons generate and propagate electrical signals. Neurons exhibit distinct firing patterns, action potential shapes, adaptation and selectivity to synaptic input. For instance, some neurons are fast-spiking, while others display adapting or bursting firing patterns. These electrophysiological properties are determined by the unique composition of ion channels and receptors expressed by each neuron (Huang and Paul (2019); Markram et al. (2004); Mukamel and Ngai (2019); Fishell and Heintz (2013); Masland (2012); Tasic et al. (2018); Zeng and Sanes (2017)).

Gene Expression

Advances in molecular biology have revealed that neurons can be distinguished by their gene expression profiles. Single-cell RNA sequencing has enabled researchers to catalog the transcriptomes of individual neurons, uncovering a rich landscape of molecular identities. These profiles often correlate with, but are not strictly determined by, morphological and electrophysiological features (Wagner et al. (2016); Shapiro et al. (2013); Trapnell (2015)).

Connectivity

Neurons are further defined by their patterns of connectivity—both the sources of their inputs and the targets of their outputs. Some neurons form long-range projections across brain regions, while others participate in local microcircuits. The connectivity matrix of the brain is thus shaped by the diversity of its neuronal components. (Wang et al. (2021); Zhang et al. (2024); Gollo et al. (2020); Goldman et al. (2023); Tripathy

et al. (2017); Sagner and Briscoe (2019))

Patch-seq

The state of the art in recording neuronal data is a technique known as patch-seq (Cadwell et al. (2016); Fuzik et al. (2016)), whereby it is possible to simultaneously record morphological, electrophysiological and molecular properties of neurons. This technique is really powerful as it provides a multi-modal perspective on neuronal heterogeneity. This technique has been the linchpin in some of the biggest classification efforts such by the Allen Brain database (Gouwens et al. (2019, 2020)).

Large-Scale Taxonomy Initiatives

In the past decade, large-scale collaborative efforts have sought to systematically map and classify the full diversity of neurons in the brain. Notable among these are the Allen Institute for Brain Science's Cell Types Program and the BRAIN Initiative Cell Census Network (BICCN). These projects leverage cutting-edge techniques—including high-throughput single-cell transcriptomics, large-scale electrophysiological recordings, and high-resolution imaging—to build comprehensive taxonomies of neuronal types Gouwens et al. (2019, 2020).

While these initiatives have greatly expanded our understanding of neuronal diversity, they often operate under the assumption that neuronal identity is static and can be captured by a fixed set of features. However, emerging evidence suggests that neuronal identity may be more dynamic and context-dependent than previously thought, raising important questions about how best to define and classify the brain's myriad cell types.

Traditional Classification Approaches

Historically, neuroscientists have sought to classify neurons into discrete "types" based on intrinsic, relatively stable characteristics. Early classification schemes focused on

observable features such as soma size, dendritic arborization, and axonal projections. With the advent of intracellular recording techniques, electrophysiological properties like spike shape, firing rate, and synaptic integration became central to neuronal taxonomy. More recently, molecular markers—such as the expression of specific neurotransmitters, calcium-binding proteins, or transcription factors—have been used to further refine neuronal classifications.

The underlying assumption in many of these approaches is that each neuron possesses a fixed identity: a stable set of features that persists across time and context. This has led to the widespread use of terms like "cell type" and "canonical neuron," suggesting a degree of invariance in neuronal identity.

Limitations of Static Neuronal Classification

- **The Influence of Input and Network State on Neuronal Function** Recent experimental and computational efforts have increasingly pointed towards the idea that a neuron's functional role is not solely determined by its intrinsic static properties such as morphology, gene expression, or ion channel dynamics but also by the nature of the input it receives and its dynamic state within a circuit. Neurons operate within continuously changing environments, receiving temporally structured synaptic input that reflects sensory stimuli, behavioral demands, and ongoing internal activity. These inputs interact with intrinsic biophysical parameters in complex ways, such that the same neuron may perform different computational roles depending on the input regime or network context. Additionally, neuromodulatory systems (e.g., dopaminergic and cholinergic pathways) further reconfigure neuronal function in a cell-type- and receptor-specific manner, modulating excitability, gain, adaptation, and stimulus selectivity. This context-dependence challenges the classical view of neurons as fixed computational units.

- **Limitations of Static Characterization Protocols** Conventional approaches to

neuronal classification typically rely on static stimulation protocols, such as step-and-hold current injections, to derive electrophysiological signatures. While these protocols offer insights into baseline excitability and passive membrane properties, they do not capture the complex temporal filtering or nonlinear input-output transformations that neurons perform under more realistic, time-varying conditions. In naturalistic settings, neuronal input is dynamic, high-dimensional, and often stochastic features that are absent in traditional measurements. As a result, static protocols risk underestimating or mis-characterizing the functional capabilities of neurons, potentially leading to oversimplified or misleading classifications. This disconnect highlights the need for stimulus-rich paradigms, such as frozen noise inputs or in vivo recordings, that better approximate the computational demands faced by neurons in situ.

- **Neuronal Identity as an Emergent, Dynamic Construct** Taken together, these observations point toward a new conceptual framework in which neuronal identity is not a fixed property, but rather a context-sensitive, emergent phenomenon. If neuronal function can be reshaped by synaptic input, neuromodulatory state, and network dynamics, then identity must be understood as fluid and multidimensional, rather than static and categorical. This dynamic view aligns with recent findings showing that neurons shift their classification depending on the input stimulus or modulatory condition. It also suggests that functional heterogeneity in neural populations is not simply biological variability, but may reflect adaptive specialization to a range of computational roles.

  In this thesis, I explore this emergent view by analyzing how neurons reorganize their functional attributes under different input regimes and neuromodulatory conditions, using high-dimensional clustering and integrative analysis across multiple feature domains. This work contributes to a growing shift in neuroscience:

from static taxonomies to dynamic, functionally grounded models of neuronal identity.

- **Rich high-dimensional functional space remains unexplored** Neurons functional space is high dimensional but most electrophysiological classification only consider one feature at a time for separability. This leads to an incomplete picture of functional heterogeneity. As we have discussed before, stimulus protocols have a strong role to play in features that can actually be extracted. For example a static input protocol doesn't expose the subthreshold potential dynamics of a neuron, similar properties related to action potential are a function of input protocol. Clustering based on these features typically involves looking at low dimensional (1-3 dimensions) feature space using a method such as K-means clustering. This approach underestimates the rich functional space in which neurons function and thus clustering based on functional properties needs method that utilize multiple features pertaining to function simultaneously and provide a richer overview of functional heterogeneity.

- **No consensus on properties that are the most informative about heterogeneity** While neurons have been categorized based active and passive features there is no consensus on properties or a set of properties that are the most informative about heterogeneity. While most classification studies focus on passive properties such as capacitance or conductance of the cell or active properties such as firing rates and inter-spike intervals, there are no studies that compare neuronal heterogeneity based on different sets of properties and provide a consensus for the field. Moreover, the properties extracted are limited by the stimulation protocol, therefore we must first establish the input protocol that is most suitable for studying a neuronal population and then compare properties extracted in each protocol to provide a consensus.

Neuromodulation: Reconfiguring Neural Computation

Neuromodulatory systems particularly those involving dopamine and acetylcholine play a critical role in shaping brain states associated with attention, learning, memory, and pathology. Rather than directly triggering spikes, these modulators reshape how neurons respond to input, modulating intrinsic excitability and synaptic integration. Their effects are often receptor and cell-type-specific, suggesting a finely tuned mechanism for reconfiguring the computational landscape of cortical networks.

Despite extensive behavioral and molecular work, the functional impact of neuromodulation on neuronal computation remains poorly understood, especially at the level of high-dimensional feature interactions. Are neuromodulatory effects isolated to individual electrophysiological features, or do they act in a coordinated manner to restructure how neurons encode information?

- **Effect of neuromodualtion on computation** Much is known about molecular and excitability changes caused by dopamine and acetylcholine modulation in single neurons. Though it is still unclear how this alteration changes what neurons compute. Since neurons relay information within a network, it is unclear how dopamine and acetylcholine changes information transfer in neurons. Since activity of individual neurons determine that state of a network, it is clear that understanding alteration in transferred information due to neuromodualtion would explain a lot about how cortical circuits are modulated.

- **Hetero modulation in neuronal population** A given population of neuron comprises of individuals or sub-groups of neurons that differ from each other in terms of their shape and and ion-channel type and distribution. It is still unclear how this individual variability manifests itself on a circuit scale. While it is known that this heterogeneity provides flexibility to the circuit (cite), the effect of neuromodualtion on this individuality is still unclear. More precisely, it

is unknown if all functional parameters are altered uniformly or there exists a heterogeneity in terms of properties that are altered due to specific modulation and what are population level effects of neuromodualtion.

- **Reconfiguration under the influence of dopamine and acetylcholine** Most studies that focus on neuromodulation of individual neurons, focus on how individual active or passive properties such as firing rate or conductance is altered due to a specific receptor activation. Neuronal functional space is high dimensional, while it is important to study how individual properties are altered due to neuromodulation, it is also important to study how this high-dimensional functional space is altered due to neuromodulation, that requires looking at correlation structure between active and passive properties and how this structure changes due to neuromodulation.

Research Objectives and Key Questions

This thesis addresses two overarching questions:

1. **How does input context influence functional classification of neurons?**
   In Chapter 2, we demonstrate that neuronal identity is *not static*, but *input-dependent*, with classification outcomes varying significantly under dynamic (frozen noise) vs static (step-and-hold) stimuli. This challenges the notion of fixed electrophysiological types and highlights the role of stimulus structure in shaping neural function.

2. **How does neuromodulation alter the functional landscape of neurons?**
   In Chapter 3, we explore how activation of dopaminergic (D1-R, D2-R) and cholinergic (M1-R) receptors *reconfigures information encoding*, shifting the balance between shared and private variance across feature domains. We show that *neuromodulatory effects are highly cell-type and receptor-specific*, modulating

not just individual attributes but also their coordination.

Experimental Framework and Approach

For answering the research questions stated above, this thesis utilizes single unit in-vitro recordings form the somatosensory cortex layer 2/3 da Silva Lantyer et al. (2018). These recordings were performed in tandem, first using a static step and hold input protocol and later with a frozen noise protocol. Within each input protocol, neurons were first recorded in a vehicle control (artificial cerebral spinal fluid - aCSF) condition and a drug condition, where a specific receptor agonist is added to the bath and the recording is performed again. Some times multiple control and drug trials are performed for a single neuron.

Input Protocols

Step and Hold

The step-and-hold protocol is a widely used electrophysiological method for characterizing neuronal properties through controlled current injections. In this protocol, a neuron's membrane potential is maintained at a baseline value (commonly around -70 mV), and then a series of incremental current steps are injected, each lasting for a fixed duration—typically 500 ms—with recovery periods (e.g., 5.5 seconds) between steps. The current amplitude is increased in defined increments (e.g., 40 pA per step), allowing researchers to observe how the neuron responds to increasing levels of depolarizing input, including changes in firing rate, spike threshold, and other action potential characteristics. This approach enables the classification of neurons (such as distinguishing excitatory from inhibitory cells) and the assessment of properties like maximum firing frequency, spike latency, and after-hyperpolarization. The step-and-hold protocol thus provides a standardized way to probe intrinsic excitability and firing dynamics, serving as a foundational tool in cellular neurophysiology.

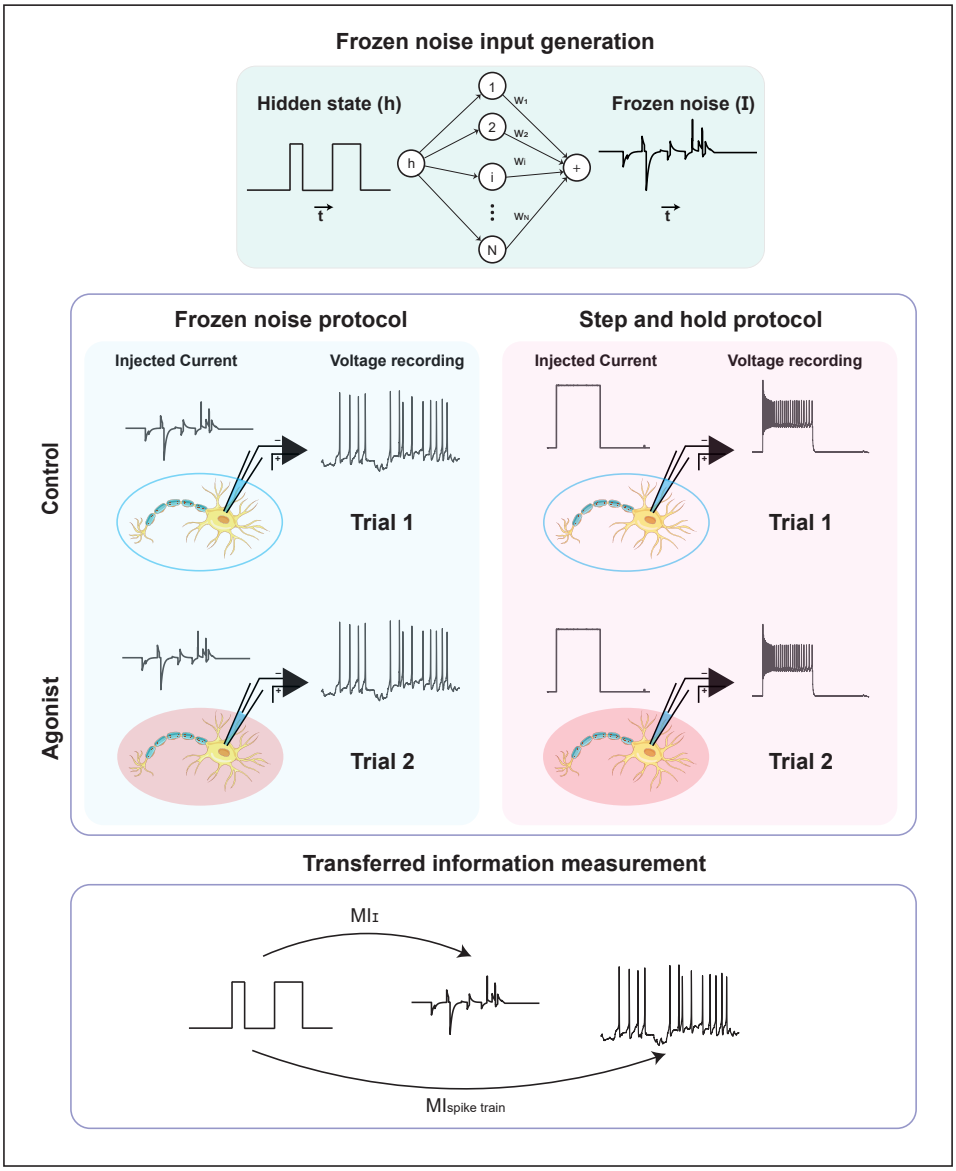

Figure 1.1. **Data collection and input protocol**

Frozen Noise

The frozen noise protocol Zeldenrust et al. (2017), is a method designed to quantify the mutual information between a neuron's input and its spike train output in electrophysiological experiments. This protocol generates a time-varying input current by simulating the activity of a presynaptic neural network of 1,000 neurons, each firing Poisson spike trains in response to a binary "hidden state" (a Markov process representing the presence or absence of an external stimulus). The injected current is "frozen," meaning the same input sequence is used across trials or conditions, enabling direct comparison of neuronal responses. By analyzing how the recorded neuron transforms this structured input into spikes, researchers can calculate the information-theoretic relationship between the hidden state and the output spike train. This approach overcomes limitations of traditional step-and-hold protocols by mimicking naturalistic synaptic input patterns while maintaining experimental control, allowing efficient bias-free information quantification with short (3–6 minute) recordings. The protocol's output includes the injected current trace, hidden state timeline, and voltage response, facilitating both forward modeling of neuronal dynamics and reverse-engineering of coding principles.

Extracted features

We analyze neuronal function across four distinct attribute sets:

- **Action Potential (AP) Dynamics**
- **Passive Biophysical (PB) attributes**
- **Adaptation Currents (AC)**
- **Input Feature Selectivity**, estimated using **Spike-Triggered Averages (STA)**

To dissect the specificity and dynamics of neuronal encoding depending on input and neuromodulation, we apply unsupervised high-dimensional clustering Lee et al.

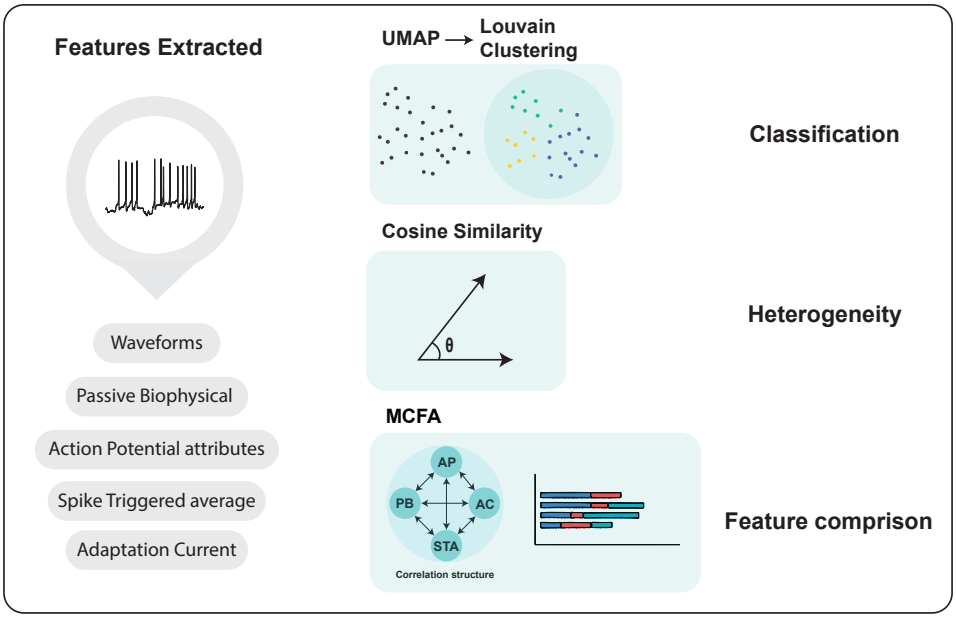

Figure 1.2. **Overview of extracted features and analysis tools**

(2021), cosine similarity analysis, and Mult-iset Correlation and Factor Analysis (MCFA) Brown et al. (2023). These methods allow us to examine both within-domain variance and cross-domain coordination of features, offering a system-level view of functional reconfiguration.

Summary and Outlook

This introduction has outlined the motivation and rationale for investigating neuronal identity and function through the dual lenses of input-dependence and neuromodulatory flexibility. Traditional classification approaches, though informative, fall short in capturing the high-dimensional, dynamic nature of neuronal computation. By leveraging frozen noise protocols, multivariate electrophysiological features, and modern unsupervised analysis techniques, this thesis proposes a data-driven framework to redefine neuronal identity as an emergent, context-sensitive construct. The following

chapters present experimental results and theoretical insights that collectively support this paradigm shift.

Thesis Structure

- **Chapter 2**: Neuronal Identity is Not Static—An Input-Driven Perspective
  Demonstrates how different stimulation protocols yield distinct classifications, showing that neuronal identity is dynamic and shaped by input.
- **Chapter 3**: Neuromodulatory Control of Cortical Function
  Examines how neuromodulators reshape the computational roles of neurons, altering both encoding capacity and feature interdependence.
- **Chapter 4**: General Discussion and Future Directions
  Integrates findings across studies, discusses implications for neuroscience and artificial neural networks, and proposes directions for future research.

References

(2008). Petilla terminology: nomenclature of features of gabaergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, 9(7):557–568.

Brown, B. C., Wang, C., Kasela, S., Aguet, F., Nachun, D. C., Taylor, K. D., Tracy, R. P., Durda, P., Liu, Y., Johnson, W. C., et al. (2023). Multiset correlation and factor analysis enables exploration of multi-omics data. *Cell Genomics*, 3(8).

Cadwell, C. R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K. F., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using patch-seq. *Nature biotechnology*, 34(2):199–203.

da Silva Lantyer, A., Calcini, N., Bijlsma, A., Kole, K., Emmelkamp, M., Peeters, M., Scheenen, W. J., Zeldenrust, F., and Celikel, T. (2018). A databank for intracellular electrophysiological mapping of the adult somatosensory cortex. *GigaScience*, 7(12):giy147.

Fishell, G. and Heintz, N. (2013). The neuron identity problem: form meets function. *Neuron*, 80(3):602–612.

Fuzik, J., Zeisel, A., Máté, Z., Calvigioni, D., Yanagawa, Y., Szabó, G., Linnarsson, S., and Harkany, T. (2016). Integration of electrophysiological recordings with single-cell rna-seq data identifies neuronal subtypes. *Nature biotechnology*, 34(2):175–183.

Goldman, M. S. et al. (2023). Intrinsic neural diversity quenches the dynamic volatility of balanced neural networks. *Proceedings of the National Academy of Sciences*, 120(28):e2218841120.

Gollo, L. L. et al. (2020). A diversity of intrinsic timescales underlie neural computations. *Frontiers in Neural Circuits*, 14:615626.

Gouwens, N. W., Sorensen, S. A., Baftizadeh, F., Budzillo, A., Lee, B. R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K., et al. (2020). Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. *Cell*, 183(4):935–953.

Gouwens, N. W., Sorensen, S. A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S. M., Feng, D., Anastassiou, C. A., Barkan, E., et al. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature neuroscience*, 22(7):1182–1195.

Huang, Z. J. and Paul, A. (2019). The diversity of gabaergic neurons and neural communication elements. *Nature Reviews Neuroscience*, 20(9):563–572.

Lee, E. K., Balasubramanian, H., Tsolias, A., Anakwe, S. U., Medalla, M., Shenoy, K. V., and Chandrasekaran, C. (2021). Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *Elife*, 10:e67490.

Liu, Y., Jiang, S., Li, Y., Zhao, S., Yun, Z., Zhao, Z.-H., Zhang, L., Wang, G., Chen, X., Manubens-Gil, L., et al. (2024). Neuronal diversity and stereotypy at multiple scales through whole brain morphometry. *Nature Communications*, 15(1):10269.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience*, 5(10):793–807.

Masland, R. H. (2012). The neuronal organization of the retina. *Neuron*, 76(2):266–280.

Mukamel, E. A. and Ngai, J. (2019). Perspectives on defining cell types in the brain. *Current opinion in neurobiology*, 56:61–68.

Nelson, S. B., Sugino, K., and Hempel, C. M. (2006). The problem of neuronal cell types: a physiological genomics approach. *Trends in neurosciences*, 29(6):339–345.

Sagner, A. and Briscoe, J. (2019). Establishing neuronal diversity in the spinal cord: a time and a place. *Development*, 146(22):dev182154.

Sanes, J. R. and Masland, R. H. (2015). The types of retinal ganglion cells: current status and implications for neuronal classification. *Annual review of neuroscience*, 38(1):221–246.

Seung, H. S. and Sümbül, U. (2014). Neuronal cell types and connectivity: lessons from the retina. *Neuron*, 83(6):1262–1272.

Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nature Reviews Genetics*, 14(9):618–630.

Somogyi, P. and Klausberger, T. (2005). Defined types of cortical interneurone structure space and spike timing in the hippocampus. *The Journal of physiology*, 562(1):9–26.

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. *Genome research*, 25(10):1491–1498.

Tripathy, S. J. et al. (2017). Transcriptomic correlates of neuron electrophysiological diversity. *PLoS Computational Biology*, 13(10):e1005814.

Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nature biotechnology*, 34(11):1145–1160.

Wang, Q., Mitra, A., Snyder, A., et al. (2021). The diversity and specificity of functional connectivity across spatial scales and brain states. *NeuroImage*, 237:118193.

Yuste, R., Hawrylycz, M., Aalling, N., Aguilar-Valles, A., Arendt, D., Armañanzas, R., Ascoli, G. A., Bielza, C., Bokharaie, V., Bergmann, T. B., et al. (2020). A community-based transcriptomics classification and nomenclature of neocortical cell types. *Nature neuroscience*, 23(12):1456–1468.

Zeldenrust, F., de Knecht, S., Wadman, W. J., Denève, S., and Gutkin, B. (2017). Estimating the information extracted by a single spiking neuron from a continuous input time series. *Frontiers in computational neuroscience*, 11:49.

Zeng, H. (2022). What is a cell type and how to define it? *Cell*, 185(15):2739–2755.

Zeng, H. and Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530–546.

Zhang, Z., Wang, Q., Luo, L., et al. (2024). Diverse and asymmetric patterns of single-neuron projectome in the mouse brain. *Nature Communications*, 15:3496.

Chapter 2

Neuronal Identity is Not Static—An Input-Driven Perspective

Based on: Willemsen, Y., Beijers, R., Arias Vásquez, A., de Weerth, C., Do Breastfeeding History and Diet Quality Predict Inhibitory Control at Preschool Age? *Nutrients*, 2021, 13, 2752.

Abstract

Neuronal classification based on morphology, electrophysiology, and molecular markers is often considered static. Here, we challenge this view, showing that functional classification depends on input patterns. Using single-cell recordings from layer 2/3 barrel cortex neurons in mice, we compared responses to step-and-hold versus dynamic frozen noise inputs that mimic presynaptic activity. Action potential and waveform-based classifications varied significantly, highlighting the dynamic nature of neuronal identity. To assess the contribution of input versus neuronal attributes toward classification, we analyzed four attribute sets, namely action potential, passive biophysical, adaptation currents, and linear input filters derived via spike-triggered averages (STA). Our findings revealed that the STA, which captures a neuron' s selective responsiveness to presynaptic activity, explained the most variance within the population. This highlights input-driven dynamics as key to functional identity, emphasizing the need for physiologically relevant inputs in defining neuronal classes and shifting the focus from static properties to dynamic functional diversity.

Introduction

Neural circuits are composed of diverse neuronal populations that exhibit variability in morphology, molecular composition, and electrophysiological properties. These neurons interact dynamically to process sensory information, support cognition, and drive behavior Harris and Shepherd (2015). A long-standing challenge in neuroscience has been to classify neurons into meaningful functional groups, with traditional approaches relying on intrinsic features such as molecular markers, morphological characteristics, and electrophysiological properties. However, despite significant advances, a consensus on the most informative features for neuronal classification remains elusive Huang and Paul (2019); Markram et al. (2004); Mukamel and Ngai (2019); Fishell and Heintz (2013); Masland (2012); Tasic et al. (2018); Zeng and Sanes (2017). An often-overlooked dimension in this classification challenge is the role of input dynamics in shaping neuronal function.

Neurons act as spatiotemporal filters, transforming incoming synaptic inputs into output firing patterns. This transformation is governed by an interplay between the structure and dynamics of the input a neuron receives, and the intrinsic membrane processes this input dynamically recruits.. Traditional classification approaches, with a focus on static properties, may therefore miss critical aspects of functional diversity that emerge from the interaction between neurons and their presynaptic partners. Recent studies suggest that electrophysiological identity is context-dependent Gouwens et al. (2019); Scala et al. (2021); Gouwens et al. (2020), varying as a function of the stimulation protocol used to probe neuronal function. However, a direct comparison of neuronal classification under physiologically realistic input conditions remains largely unexplored.

Recent advances in patch sequencing Fuzik et al. (2016); Cadwell et al. (2016); Scala et al. (2019) enable simultaneous extraction of transcriptomic, morphological,

and physiological properties, improving neuronal classification. The Allen Brain Institute dataset Gouwens et al. (2019); Scala et al. (2021); Gouwens et al. (2020) provides a broad classification of visual and motor cortex neurons using multi-modal features (MET types). However, classification based on electrophysiology remains challenging. Neuronal properties exist on a continuum Markram et al. (2004), with unclear boundaries between classes. Moreover, molecularly defined types exhibit overlapping electrophysiological properties Gouwens et al. (2019, 2020), and intrinsic electrophysiology varies with stimulation protocols. While multi-modal techniques improve classification, they overlook the influence of synaptic input. Neurons receive inputs from thousands of presynaptic neurons shaping their firing properties, necessitating classification under physiologically realistic conditions Connors and Gutnick (1990); Steriade (2000).

To address this gap, we investigated how neuronal identity is shaped by the nature of the input a neuron receives. We recorded from layer 2/3 neurons in the barrel cortex and compared their responses under two different stimulation paradigms: a standard step-and-hold (SH) stimulus, which provides a static, artificial input, and a frozen noise (FN) stimulus Zeldenrust et al. (2017), which simulates natural presynaptic activity. We hypothesized that functional classification of neurons would be stimulus-dependent, with the FN protocol revealing a distinct organization of neuronal diversity that is not captured by the SH protocol.

We analyzed four commonly used sets of electrophysiological attributes—action potential features, passive biophysical properties, adaptation currents, and linear input filters estimated via spike-triggered averaging (STA)—to assess their contributions to neuronal classification under dynamic input conditions. Using multiset correlation and factor analysis, we found that the linear input filter, which characterizes a neuron's sensitivity to specific input features, was the most informative attribute for understanding neuronal functional variance, i.e, it is the most relevant attribute to be used for clustering neurons. This finding challenges the traditional view that neuronal iden-

tity is a static property, emphasizing the importance of input dynamics on functional diversity.

By demonstrating that neuronal classification is highly dependent on input dynamics, our study highlights the need to incorporate physiologically relevant stimuli when defining neuronal types. Our findings suggest that neurons should not be categorized based solely on static features but rather on how they process and respond to dynamic synaptic input. This perspective has broad implications for both experimental and computational neuroscience, urging a paradigm shift toward input-dependent models of neuronal function.

Materials and methods

Ethics statement The data used in this research was previously published and made freely available to the community da Silva Lantyer et al. (2018) and Yan et al. (2022). All the experimental work, as outlined in the cited articles, were carried out in compliance with the European directive 2010/63/EU, the national regulations of the Netherlands, and international standards for animal care and use.

Slice electrophysiology Data acquisition procedures, the details of the in vitro slice preparation, intracellular access to anatomically targeted neurons, data digitization, and preprocessing have been described in detail elsewhere Kole et al. (2020); Kole and Celikel (2019); da Silva Lantyer et al. (2018); kol (2017); Miceli et al. (2017). In short, Pvalbtm1(cre)Arbr (RRID:MGI:5315557) or Ssttm2.1(cre)Zjh/J (RRID:IMSR_JAX: 013044) mice, including both females and males, were obtained from local breeding colonies and studied after the maturation of evoked neurotransmitter release in the primary somatosensory cortex (Martens et al. (2015)).

Mice were anesthetized with Isoflurane (1.5 mL/mouse) before extracting tissue, and coronal slices of the primary somatosensory cortex (barrel subfield) were prepared.

The brain was removed, and 300 µm-thick coronal slices were made. Slices were then incubated in artificial cerebrospinal fluid (aCSF) (120 mM NaCl, 3.5 KCl, 10 glucose, etc.), aerated with 95% O_2/5% CO_2 at 37°C, and then at room temperature after 30 minutes.

Whole-cell electrophysiological recordings were performed with continuously oxygenated aCSF. The barrel cortex was localized, and cells in the supragranular layers were patched under 40x magnification using HEKA EPC 9 and EPC10 amplifiers with Patch Master software. Patch-clamp electrodes were pulled from glass capillaries (1.00 mm external diameter, 0.50 mm internal diameter) and used with 5–10 MOhm resistance, filled with intracellular solution (130 mM K-Gluconate, 5 KCl, 1.5 $MgCl_2$, etc., pH adjusted to 7.22 with KOH). Data were band-pass filtered at 0.1–3000 Hz before storage for offline analysis.

Step and hold (SH) protocol The Step and Hold protocol was set up in a current clamp configuration, the resting membrane potential was set to -70 mV before current injection into the soma of the neuron. In total, 10 current step injections, each 500 ms long, were performed. The steps ranged from 40 pA to 400 pA with an inter-sweep interval of 6.5s. The stimulus was repeated 1 to 3 times for neurons. The drifts encountered were not corrected for.

Frozen Noise (FN) protocol The Frozen Noise input protocol consisted of injecting a somatic current that is the result of an artificial neural network of 1000 neurons responding (firing Poisson spikes) to random stimuli i.e., the hidden state, the membrane potential response to the somatic input is recorded with a sampling rate of 20 kHz for a total length of 360 seconds and saved. Each raw data file consisted of a vehicle control trial (artificial Cerebrospinal fluid i.e. aCSf) and a drug trial (a specific neuromodulatory receptor agonist or antagonist was added to the bath and the recording was repeated). Some files consisted of multiple control and drug trials. See

Zeldenrust et al. (2017); da Silva Lantyer et al. (2018) for more details.

Method details

Feature Extraction

Feature extraction is performed separately for SH and FN protocols to classify neurons using different inputs and to determine cell class with a realistic stimulus. For the first part, which is comparing classification across protocols, we collect waveforms and Action potential features for comparing SH and FN protocols using a subset of the data, (186 neurons). As not all cells were recorded with both types of input protocols, this subset was chosen to match the cell IDs across the two protocols. For the second part, which compares different physiological attributes for the FN protocol, we collect waveforms, spike event-based features, biophysical features, and the STA for each control (aCSF) trial for each cell in the FN dataset. We discard trials that contain recording artifacts (observed distortions or high-frequency noise), in total 11 neurons. Since there were multiple drug and control trials available for each neuron, we always took the first control trial for this study as this prevented any residual effect from the drug condition after washing. The total number of neurons used for this part of the study was 312.

Spike waveforms Spike waveforms were extracted for each neuron from both SH and FN protocol trials (this could not be done for all the neurons as some neuron IDs could not be matched between SH and FN protocols due to missing metadata). Firstly, we identified the spike times in each trial in each protocol. This was done by identifying peaks in the membrane potential trace using the Scipy Findpeaks function Virtanen et al. (2020) with $height = 20mV$, and $distance = 80ms$ as the chosen hyperparameters. Next, we cut the spike waveform by defining a window of 2 ms before and 3 ms after each spike peak to get a 5 ms long waveform for each spike. We ignored the spikes that had an ISI lower than 3 ms as we were interested in

non-bursting type spike shapes. We then average over all the waveforms for each trial to get an average waveform shape in both FN and SH protocols respectively.
Extracted Waveforms for the second part were 10 ms long (5 ms before and after the spike peak), to incorporate the subthreshold dynamic before reaching the threshold of the neuron in the waveform classification as well. We observed some variability in the baseline membrane potential values as well as the slope of membrane potential before the threshold value was reached.

Action potential features Action potential features were extracted to study the variability in spike-related dynamics across individual neurons. These features were designed based on their suitability for comparing neurons in FN and SH protocols. These Action potential features were divided into three categories: Spiking Dynamics, spike threshold, and Action Potential height and width.

Spiking Dynamics

This includes the following features:

- **Current at the first spike**: We take the current amplitude when the membrane potential crosses the threshold for the first time in the trial for FN protocol. For SH protocol, it is the current step that produces the spike for the first time.
- **AP count**: We count the number of spikes over the entire trial length of 360 seconds for the FN trial and 500 milliseconds for the SH trial which is the duration for the current onset.
- Inter spike interval: Inter spike interval was measured as the time interval between two spikes in milliseconds. We calculate mean, median, maximum, and minimum values for each trial in SH and FN protocols. For the SH case, we measure the ISI for the maximum current amplitude (400 pA) trial.

- **Time to first spike**: We measure the time (in milliseconds) it takes for the neuron to fire the first action potential. For the SH case, we take the lowest amplitude step where a spike is observed.

- **Firing rate**: We calculate the firing rate as the number of spikes per second. For the FN case, we take the entire length of 360 seconds of the trial and for the SH case, we take the duration of the current onset which was 500 milliseconds for the highest current step (400 pA).

$$fr = \frac{N_{spikes}}{T} \tag{2.1}$$

  where N_{spikes} is the total number of spikes in the trial and T is the total length of the trial.

- **Interspike Interval**: Interspike interval is the duration between two spikes $t_{spike_{n+1}} - t_{spike_n}$, where $t_{spike_{t+1}}$ and t_{spike_t} are spike times for spikes n+1 and n. We take the interspike interval for all the spikes in the trial for both FN and SH protocols and calculate the mean, median, minimum, and maximum values.

- **Instantaneous rate**: Instantaneous rate is defined as the reciprocal of the average of the Interspike interval.

$$inst.firing\_rate = \frac{1}{t_{spike_{n+1}} - t_{spike_n}} \tag{2.2}$$

Spike threshold

- We measure the threshold of each spike from the trial as described in Fontaine et al. (2014). The threshold is defined as the voltage V at the spike onset when the first derivative of the membrane potential dV/dt reaches 25 mV/ms for the first time. We take the first threshold value of the trial as well as the mean,

median, maximum, and minimum values of the thresholds from the entire trial length.

Action Potential height and width

- **Width**: Spike width is calculated as the time it takes between when the membrane potential reaches the AP threshold and when the membrane voltage goes below the threshold after the spike peak. We calculate the width for each spike in the trial for both FN and SH protocols. We calculate the mean, median, maximum, and minimum for all the values obtained from a trial.
- **Amplitude**: Spike amplitude is calculated as the difference between AP peak and AP threshold value. We calculate the amplitude for each spike in the trial for both FN and SH protocols. We calculate the mean, median, maximum, and minimum for all the values obtained from a trial.

Biophysical Feature extraction using GLIF model

Since it is not possible to empirically observe the biophysical properties of the cell just using the membrane potential, we fit a Generalized Leaky Integrate and fire model (GLIF) to the recordings, which can capture universal spiking and sub-threshold dynamics Pozzorini et al. (2015). The following equations define the GLIF model:

$$C\dot{V}(t) = -g_L(V(t) - E_L) - \sum_{\hat{t}_j < t} \eta(t - \hat{t}_j) + I(t), \tag{2.3}$$

where $V(t)$ is the membrane potential, C is the membrane capacitance, gL is the leak conductance, EL is the resting potential, and $\eta(t)$ is the adaptation current triggered by a spike event. Spikes are stochastically produced by a point process that represents conditional firing intensity $\lambda(t|V, V_t)$ that is dependent on the instantaneous difference between the membrane potential and voltage threshold given by:

$$\lambda(t|V, V_t) = \lambda_0 * exp\Big(\frac{V(t) - V_T(t)}{\Delta V}\Big), \tag{2.4}$$

where λ_0 is the base firing rate in Hz, V(t) is the membrane potential, and V_T(t) is the moving spike threshold and ΔV controls the sharpness of the exponential threshold function. The probability of a spike $\hat{t}$ between a time interval t and Δt is given by the following equation (based on Gerstner et al. (2014)):

$$P(\hat{t} \in [t, t + \Delta t]) = 1 - exp(-\int_{t}^{t+\Delta t} \lambda(s)ds) \approx 1 - exp\Big(-\lambda(t)\Delta t\Big), \tag{2.5}$$

The dynamics of the firing threshold $V_T(t)$ are given by:

$$V_T(t) = V_T^* + \sum_{\hat{t}_j < t} \gamma(t - \hat{t}_j), \tag{2.6}$$

where γ is the stereotypical movement of the spike threshold after a spike and V_T^* is the threshold baseline.

The method for fitting this neuron model to a membrane potential recording is divided into the following steps:

Preparation step: A 100-second window from the initial part of the trial is taken as the training set for the fitting, Pozzorini et al. (2015) shows that a longer trial length doesn't improve the fit. Spike times and waveforms are also extracted for the preparation step of the fitting procedure.

Step 1. Fitting the reset voltage: The waveforms extracted in the preparation step were averaged, and then the Reset voltage V_{reset} was extracted using the averaged waveform by setting an arbitrary refractory period t_{ref} and taking the membrane potential value at $t_{\hat{i}} + t_{ref}$, where $t_{\hat{i}}$ is the spike peak. The refractory

period is chosen to be always lower than the minimum inter-spike interval, we chose the refractory period of $t_{ref} = 4ms$ in this case.

Step 2. Fitting sub-threshold dynamics : The voltage dynamics in eq (2.3) are given by parameter set $\theta_{sub} = \{C, g_L, \eta, \text{and } E_L\}$, by fitting the temporal derivative of the data $\dot{V}_{data}$ in the model, we can extract the set of passive parameter set θ_{sub} for the data. Firstly, we can write the adaptation current η as a linear sum of rectangular basis functions Mensi et al. (2012).

$$\eta(t) = \sum_{k=1}^{K} a_k f^{(k)}(t), \tag{2.7}$$

Using the fact that the voltage dynamics are approximately linear in the sub-threshold regime, θ_{sub} parameter set can be extracted using a multi-linear regression between $\dot{V}_{data}$ and $\dot{V}_{model}$. For this, we created a training set V^{sub}_{data} where we removed the spike waveforms from V_{data}, $V^{sub}_{data} = \{V_{data}(t) | t \notin (t_{\hat{i}} - 5ms, t_{\hat{i}} + t_{ref})\}$, where $t_{\hat{i}}$ is the spike times. The regression problem can be stated as:

$$\theta_{sub} = (X^T X)^{-1} X^T \dot{V}^{sub}_{data}, \tag{2.8}$$

where X^T is a matrix representing parameter values at different time points, its row elements x_t^T are of the following form

$$x_t^T = [V^{sub}_{data}, 1, f^{(1)}(t), f^{(2)}(t), ..., f^{(K)}(t)], \tag{2.9}$$

Step 3. Fitting the spike probability: For fitting the spiking probability to the data, we need to extract parameters defining the dynamics of the threshold eq (2.6). The stereotypical shape of the adaptation current threshold movement can be

expanded as a sum of rectangular basis functions as follows (Mensi et al. (2012)):

$$\gamma(t) = \sum_{p=1}^{P} \gamma_p f^{(p)}(t), \tag{2.10}$$

We use the parameters obtained in the previous steps to compute the subthreshold membrane potential of the model using numerical integration of eq (2.3). We set $\lambda_0 = 1$ Hz and all the threshold parameters $\theta_{thr} = \{\Delta V,\ V_T^*\ ,$ and $\gamma(t)\ \}$ are extracted by maximizing the likelihood function of the following form based on the experimental spike train:

$$\hat{\theta}_{thr} =_{\theta_{thr}} \Big[\sum_{t \in \{\hat{t}_j\}} y_t^T \theta_{thr} - \Delta T \sum_{t \in \Omega} exp(y_t^T \theta_{thr}) \Big], \tag{2.11}$$

Where $\Omega = \{\mathrm{t}|\mathrm{t} \notin (\hat{t}_j, \hat{t}_j + t_{ref})\}$,

The subthreshold fit is examined by comparing the variance explained R^2 of the subthreshold membrane potential trace V between the data and the model. All the models chosen for clustering had an R^2 value > 0.7. The sets $\theta_{clustering} = \{g_L,\ \Delta V,$ C, $V_T^*,\ E_L,\ V_{reset}\}$ are the parameters that are extracted from the model that is used in the clustering procedures.

Spike Triggered Average The spike-triggered average (STA) is the average shape of the stimulus that precedes each spike. We extracted the STA using the following equation given by Schwartz et al. (2006):

$$STA = \frac{1}{N} \Sigma_{n=1}^{N} \vec{s}(t_n), \tag{2.12}$$

where t_n is the n^{th} spike time, s is the stimulus vector preceding the spike for a fixed time window of 100 ms, and N is the total number of spikes. Before clustering, we

standardize (i.e. z score) and then normalize the STA vector with an L_2 norm. We didn't use any kind of whitening or regularization to calculate the STA.

UMAP + Louvain clustering

Conventional clustering algorithms such as K-means do not perform well in high dimensional spaces (p>>N, where p is the dimension of data and N is the number of samples) due to the curse of dimensionality Aggarwal et al. (2001), and therefore need a pre-processing dimensionality reduction step. Addressing this issue, a non-linear dimensionality reduction algorithm such as UMAP McInnes et al. (2018a) creates a high-dimensional graph representation of the data which can be utilized for clustering using a graph-based clustering method such as Louvain clustering Blondel et al. (2008) or ensemble clustering Poulin and Théberge (2019). This method utilizes the high dimensional space of the data for clustering. As shown by Lee et al. (2021) using the WaveMAP algorithm, the UMAP+Louvain community detection algorithm has been successful in finding neuron types based on extracellular recordings.

UMAP

Universal Manifold Approximator (UMAP) is a non-linear dimensionality reduction technique that preserves local and global relationships between data in high dimensional space McInnes et al. (2018a). It is divided into two steps, the first step is creating a k-nearest neighbor graph and the second step is to generate a low-dimensional representation that is similar to the high-dimensional graph structure.

We used the Scikit-learn UMAP-learn software package McInnes et al. (2018b) to extract the embedding and the graphs.

Louvain Community detection

The Louvain community detection algorithm Blondel et al. (2008) maximizes modularity amongst the identified groups in a graph. Modularity can be defined by the following equation:

$$Q = \frac{1}{w}\sum_{i,j}\left[A_{i,j} - \gamma\frac{d_i^+, d_i^-}{w}\right]\delta(c_i, c_j) \tag{2.13}$$

where $A_{i,j}$ is the adjacency matrix, $k_i = \sum_j A_{ij}$ is the sum of the weights of the edges attached to the vertex i, c_i is the community for vertex i, d_i is the degree of node i, d_i^+ and d_i^- are the in degree and out degree for node i,$\delta(u, v)$, Kronecker symbol, is 1 if u=v and 0 otherwise, w = $\sum_{i,j} A_{i,j}$ and $\gamma > 0$ is the resolution parameter. For Louvain graph-based clustering we used the implementation from the Scikit-network software package Bonald et al. (2020).

The clustering approach can be summarized in the following steps:

1. The high dimensional k-neighbor graph is obtained by the first step of the UMAP algorithm using data vectors that are first standardized and then normalized using the L_2 norm. The nearest neighbor and distance parameters for UMAP were 20 and 0.1 respectively. This is to ensure a compact embedding and a clear clustering.

2. Using the graph obtained in the first step, we perform Lovain community-based clustering, using the resolution parameter γ that maximizes the modularity score, and the corresponding community/partition is chosen as the final cluster labels.

3. Using the cluster labels found in the second step, we color the individual points in distinct colors on the low-dimensional UMAP representation.

This unsupervised clustering approach was effective in capturing the global structure of the high dimensional space across attributes such as waveforms, adaptation

current, and Spike Triggered average. The clusters found using this method were robust even when clustering was repeated with a sub-sample of the data and while iteratively removing the features.

Cluster Stability and Parameter Selection Cluster stability is tested by clustering a 90% sub-sample of the data chosen at random and repeating the procedure 25 times for each resolution parameter, varying from 0 to 5 with a step of 0.5. The modularity score is calculated for each resolution parameter and finally, the resolution parameter is chosen for which the modularity score is maximal. Variation in the number of clusters is observed for each resolution parameter and is contrasted with the modularity score. Clustering robustness is also tested by repeating the procedure above while excluding one feature at a time for Action potential feature clustering and passive biophysical clustering.

Cluster Likelihood comparison Cluster likelihood between two sets of labels was calculated in two steps, firstly we created a contingency matrix such that $C_{i,j}$ contained the number of times neurons classified in cluster i in the SH protocol classified as cluster j in the FN protocol, such that each row contains the division of elements of cluster i in the SH protocol into all the clusters of the FN protocol. Secondly, to get the likelihood, we divided each row by the total count of neurons in cluster i in the SH protocol.

$$P(j|i) = \frac{C_{i,j}}{\sum_j C_{i,j}} \tag{2.14}$$

Here $P(j|i)$ is the probability of the neuron classifying into FN cluster j given it is classified in SH cluster i.

Cluster Similarity Measures We measured the similarity between two given cluster assignments using the Adjusted Random Index (ARI) and Adjusted Mutual

Information (AMI) Score. Both of these measures amount to change agreements between two clusters. The ARI (range -1-1) measures the pairwise relationships between clusters and the AMI (range 0-1) measures the overall information shared between the two clusters.
We used the scikit-learn Python package Pedregosa et al. (2011) to calculate the ARI and AMI measures.

Ensemble Clustering for Graphs (ECG)

We used Ensemble Clustering for Graph algorithm Poulin and Théberge (2019) to validate the clusters found using the Louvain Community detection algorithm. It is a two-step algorithm, the first step called the generation step consists of producing a $k-$level 1 partition $P = \{P_1, P_2, ...P_k\}$ by running the first pass of the Louvain clustering algorithm with random vertices on the initial graph G= (V, E). The second step, also known as the integration step, consists of running the Louvain algorithm on a weighted version of the initial graph. Where the weights are the weight of an edge given by

$$W_p(u,v) = \begin{cases} w_* + (1 - w_*) * \left(\frac{\sum_{i=1}^{k} v_{p_i}(u,v)}{k} \right) & \text{, if (u,v) is in 2-core of G} \\ w_* & \text{, otherwise} \end{cases} \tag{2.15}$$

where $0 < w_* < 1$ is the minimum ECG weight and $v_{p_i}(u,v) = \sum_{j=1}^{l_i} 1_{C_i^j}(u) \cdot 1_{C_i^j}(v)$ shows if the vertices u and v co-cluster in the same cluster of P_i or not. Thus it takes advantage of multiple instances of the Louvain clustering algorithm to make a clustering based on consensus. We used the implementation provided by (Lee et al. (2021)) for comparing the waveform clustering for SH and FN protocols. The graph used for Ensemble clustering was the same as in the original clustering using the

UMAP algorithm.

Quantification and statistical analysis

MANOVA We measured the significance between excitatory and inhibitory action potential and passive biophysical feature vectors using a one-sided MANOVA. We present the following statistics: Wilk's lambda, Pillai's trace, Hotelling-Lawley trace, and Roy's greatest root using the Stats Model python package Perktold et al. (2024).

Canonical Correlation Analysis To perform a post-hoc analysis on Action potential and Passive biophysical features, to find the importance of each feature, we used a Canonical Correlation analysis. Which is a method to find a linear combination of features between two datasets that maximizes the correlation between them. It is a deterministic method that results in canonical variates of the two datasets that are maximally correlated. Since the excitatory/inhibitory populations were different in numbers, we repeated the SHA procedure 10 times with random sampling from the larger group to make the population size equal between excitatory and inhibitory groups. We then obtain loadings for each dataset by averaging over the 10 repetitions, which represents the correlation of a feature with the canonical variate. We used the Scikit learn python package (Pedregosa et al. (2011)) for performing the SHA procedure.

Welch's ANOVA To compare the significance of cosine similarity measures for excitatory and inhibitory populations, we first calculated the cosine similarity matrix comparing the excitatory and inhibitory populations separately and a third matrix comparing excitatory with inhibitory populations. We take the upper triangular part of the excitatory and inhibitory within-population comparison matrix and the entire excitatory vs inhibitory matrix for the significance test. The Welch's ANOVA test was used to compare the three groups of cosine similarity for each attribute with one

categorical variable namely, excitatory or inhibitory. This test was chosen because the populations for each attribute were heterodrastic and of unequal size. We performed a post-hoc Games-Howell test to determine which group was significantly different, this was done because the variances across groups were heterodrastic. The significance levels are reported in the figures based on the p-values obtained by the post-hoc test. We used the Stats model python package Perktold et al. (2024) for calculating Welch's ANOVA.

Multi-set Correlation and Factor Analysis

The Multi-set Correlation and Factor Analysis (MCFA) was performed using the procedure as described by Brown et al. (2023), and the accompanying software was used for the analysis Brown and collinwa (2023). The problem can be stated as follows, let $\{Y^m\}_{m=1}^M$ be a set of M attributes extracted from the electrophysiological data, each with dimension $N \times P_m$, where N is the number of samples and P_m is the dimension of each attribute (Action potential parameters (P_1 =22), passive biophysical parameters (P_2 =6), Spike triggered Average (P_3 =2000), and Spike triggered current (η) (P_4 =10000)). Each attribute set can be modeled as having a contribution from two factors, a shared and a private factor respectively, as shown below.

$$z_n \sim N(0, I_d) \tag{2.16}$$

$$x_n^m \sim N(0, I_{k_m}) \tag{2.17}$$

$$y_n^m \sim N(W_m z_n + L_m x_n^m, I_{k_m}, \Psi_m) \tag{2.18}$$

Here z_n is the shared factor of dimension d, and W_m is the shared space loading matrix of shape $P_m \times d$. k_m is the dimension of each private mode. x^m is the private space for each attribute m of dimension k_m, L_m is the private space loading of shape

$P_m \times k_m$. $\Psi_m = diag(\psi_m^1, ..., \psi_m^{p_m})$ are the diagonal residual covariance matrices. Given Y, d, and k_m the goal is to find hidden factors z_n, x_n^m, and loading matrices W_m, L_m. This is achieved by the Expectation Minimization (EM) method Brown et al. (2023).

We center and scale all variables as in Brown et al. (2023) and initialize the loading matrices similarly to the original method, using the pairwise correlations with average variance constraint initialization. To model the shared latent space z_n, we chose the most informative PCA components based on Marchenko Pasteur Law to control over-fitting Brown et al. (2023), which states for any normalized dataset ($\mu = 0$, $\sigma = 1$), the principal components with eigenvalues above $\lambda_m = 1 + \sqrt{p_m/N}$ are considered non-noise. We set the size of the private space $k_m = 1$ for passive biophysical parameters due to its relatively lower dimension and $k_m = 2$ for the other attributes. After initializing all the variables, we run the expectation minimization (EM) algorithm to obtain z_n, x_n^m, W_m, and L_m matrices.

For feature j of mode m, the variance explained by a shared feature d is given by $W_m^{(j,d)2}$. Similarly, the variance explained by the k^{th} private factor of feature j of mode m is given by $L_m^{(j,k_m)2}$. The total variance explained for a mode by a shared factor is given by $\sum_j W_m^{(j,d)2}$, similarly, the total variance explained by the private factor k_m is $\sum_j L_m^{(j,k_m)2}$. Hence, the total variance explained by all shared and all private factors is given by $\sum_{j,d} W_m^{(j,d)2}$ and $\sum_{j,k_m} L_m^{(j,k_m)2}$.

The relative feature importance is given by the cross-correlation of columns of the posterior mean of Z on observing a single mode denoted by $\hat{Z}_m$ Brown et al. (2023)

$$\hat{Z}_m = E[Z|W_m, \Psi_m, L_m, Y_m] = Y_m(W_m W_m^T + L_m L_m^T \Psi_m)^{-1} W_m \tag{2.19}$$

$$S_d = cor(\hat{Z}_1^{(:,d)}, ..., \hat{Z}_m^{(:,d)}) \tag{2.20}$$

Results

We aimed to observe the contribution of the input type in explaining functional classification and to explore what a physiologically realistic stimulus reveals about the functional classification of a neuron. For the first aim, we researched the stimulus dependence of neural classification using two different sets of classification features: 1) action potential waveforms and 2) other action potential attributes, such as the spike threshold and spiking dynamics. Next, to research the second aim, we investigated which attributes are the most informative about neuronal heterogeneity under a physiologically realistic stimulation protocol. Therefore, We performed classifications based on four different attribute sets that capture 1) action potential attributes, 2) Passive biophysical attributes, 3) Adaptation current, and 4) linear input filter through a Spike Triggered Average (STA). Finally, we used a method known as Multi-set correlation and Factor Analysis (MCFA) to compare the variance explained by the shared structure across these four attribute sets.

Stimulus dependence of neural classification

We first aim to understand the role of a neuron's input in the functional classification of neural populations. For this, we analyze single-cell patch-clamp recordings (da Silva Lantyer et al. (2018); Yan et al. (2022)) recorded with two different input conditions: 1) a Step and Hold (SH) and 2) a Frozen Noise (FN) protocol (see Methods). We want to understand the influence of a physiologically realistic FN input on classification using commonly used features in contrast to SH input. To contrast the heterogeneity between SH and FN input conditions, we compare compare classification in these two input conditions using the neuronal spike waveforms and action potential attribute sets and also measure the similarity in the waveforms and action potential attributes between each cluster across the two input protocols.

Stimulus dependence of neural classification using intracellular action potential waveforms

To understand if waveform-based classification differs under FN and SH input conditions, we analyzed control (aCSF) trials from a total of 186 in-vitro whole-cell patch-clamp neural recordings that consist of the same cell recorded under 2 different input protocols, namely Step and Hold (SH) and Frozen Noise (FN) (Fig. 2.1 **c-d**, see methods). Using extracted waveforms of the same length (5 ms) from both FN and SH trails, we standardize (subtracting the mean and scaling to unit variance) the waveforms and then normalize the data using an L2 norm. Next, we apply an unsupervised high dimensional clustering algorithm that combines UMAP and Louvain Community detection (see Methods) and found 7 clusters for SH trials as well as 7 clusters for FN trials (Fig. 2.1 **a-b**) respectively. We also measured the stability of the clusters (see Methods) against changing the hyperparameter (resolution parameter) for the unsupervised method. We found the clusters to be stable (low standard deviation) for the chosen resolution parameter (chosen at a value of 1.0) (Fig. 2.1**g**).

From inspecting the two-dimensional UMAP projection, we found that the manifold representing the waveforms in the FN protocol (Fig. 2.1**a**) consists of two broad subdivisions, whereas the SH protocol manifold seems much more connected and spread out (Fig. 2.1**b**). We compared waveform shapes across clusters by creating a pairwise cosine similarity matrix between FN and SH protocols and then taking the average of the submatrix for each SH and FN cluster pair. The average similarity across clusters is summarized using a heatmap (Fig. 2**e-f**). We first perform this comparison for each protocol separately and find that some cluster waveform shapes are similar (cosine similarity > 0.9) to their immediate neighbors (e.g. cluster 7, 3, and 4 in SH and cluster 6, 3, and 4 in FN in (Fig. 2.1**e**) for both FN and SH protocols. Similarly, comparing the waveform shapes between SH and FN, we find that narrow-width clusters for the SH

protocol (4, 3, and 7) are highly similar to narrow-width clusters for the FN protocol (6, 2, and 3). We also project all the SH and FN waveforms together on a single embedding space and observed they were distinct (Appendix. 6.1**a**), suggesting that waveform shapes are not fixed across protocols. We then test if there exists a drift in the recordings by dividing the 360-second trial into two parts and comparing the waveforms between the first and the second trial by projecting the waveforms on the same 2D UMAP space. Both halves are overlapping suggesting that waveform shapes are similar between the two halves of the trail.

Finally, we calculate the likelihood for neurons in one of the SH clusters to be clustered together in one of the FN clusters (see methods and Fig. 2**h**). We find that for each SH cluster, the likelihood of clustering in one of the FN clusters is spread without a strong majority, suggesting that neurons grouping in the SH protocol do not group in the FN protocol. We further quantify the clustering agreement between the SH and FN clusters using an adjusted random score (ARI) and adjusted mutual information score (AMI) between FN and SH clusters. Both measures show low values ($ARI = 0.085$, $AMI = 0.133$). This confirms that neurons cluster differently between SH and FN protocols. To verify if the found clusters don't have a bias as a result of the method that we use, we repeat the analysis using the Ensemble clustering method (Appendix. 6.2) to cluster the SH and FN waveforms. Taking the average of the highest values for each row in the co-clustering matrix, we find 84% correspondence for FN waveforms between wave map and ensemble clustering and 96% correspondence for SH waveforms between wave map and ensemble clustering. Comparing the ensemble clustering with Louvain clustering using an adjusted mutual information score (AMI Louvain vs Ensemble (FN) = 0.765, AMI Louvain vs Ensemble (SH) = 0.736), we find a high level of similarity for both SH and FN protocols. In conclusion, we find that clustering neurons into cell classes based on their waveforms results in different cell classes in the SH and FN input protocols, resulting from differences in the waveform

shapes due to the stimulation protocol. This shows that waveform-based neuronal identity is stimulus-dependent.

Stimulus dependence of neural classification using action potential attributes

It has been found that excitability measures such as total spike count and AMPA conductance threshold (dynamic) vs rheobase (static) have a low correlation between static and dynamic stimulus conditions Hernáth et al. (2019); Szabó et al. (2021). As we found in the previous section that cell clustering based on waveforms depend critically on the stimulation protocol. This intrigued us to investigate if neuronal classification based on action potential attributes (see Methods), that incorporate among others spiking dynamics, spike threshold and action potential height and width (a total of 22 attributes), is also input-specific. We designed a set of features to allow for comparisons across input protocols. This choice of attributes is based on previous literature that clusters neurons based on electrophysiological attributes. The attribute list is not exhaustive but rather confined as features can be input specific. The properties included in the spiking dynamics set incorporate the firing statistics such as inter-spike interval (ISI), firing rate, and action potential count among others that are commonly studied. Next, we design a spike threshold attribute set to capture spike threshold-related parameters for the trial. The spike thresholds were calculated for each action potential in each trial for the analysis (see Methods). Finally, we measure the action potential height and width parameters, even though we have clustered based on average waveforms in the previous section. We do this for two reasons, first, height and width parameters are electrophysiological parameters commonly used in the clustering literature and second, the average waveforms studied previously do not capture the change in the height and width of the waveform within each trial. We used descriptive statistics such as mean, median, minimum and maximum values for some of the features to capture the distribution of these features in each trial.

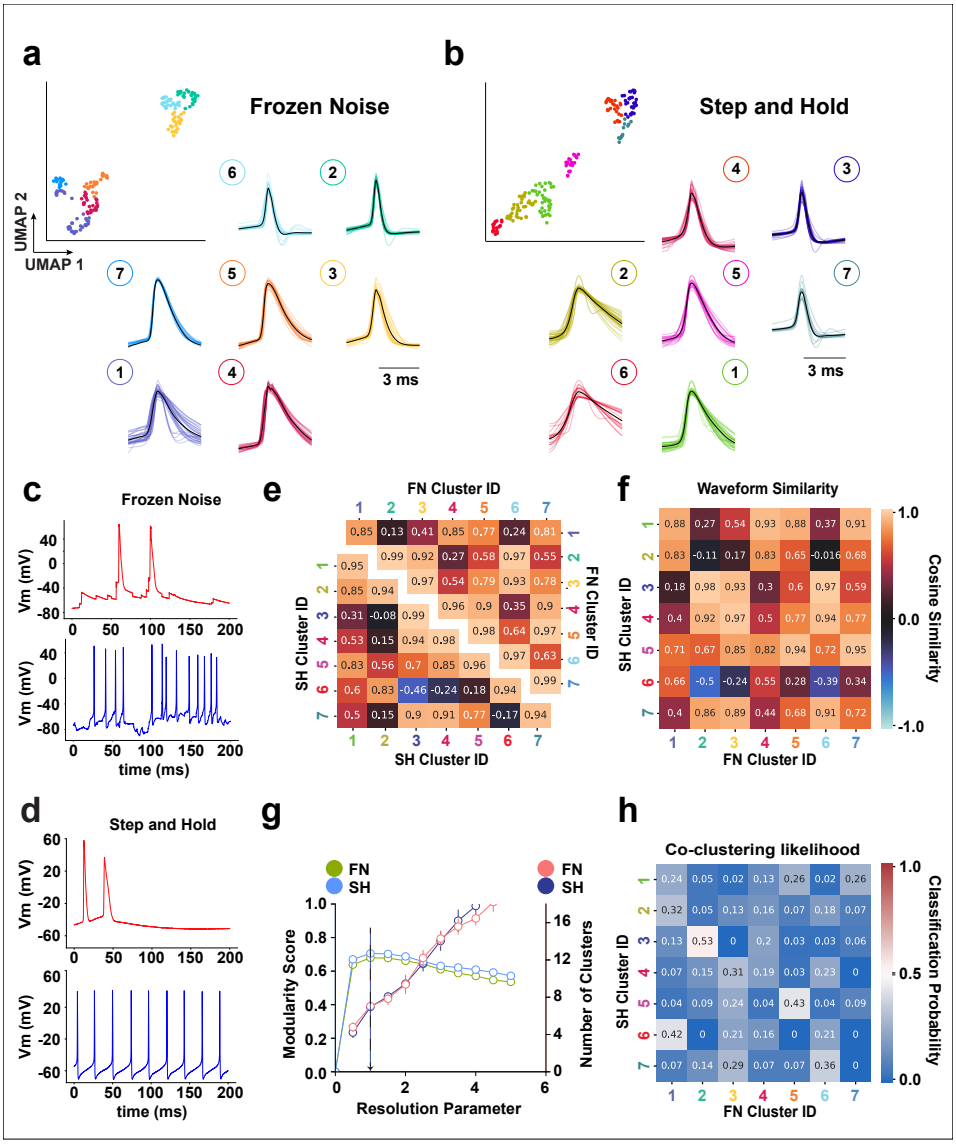

Figure 2.1. **Intracellular waveform-based neural identity (cluster membership) depends on the stimulus type.** Each of the 186 cortical neurons was studied using two distinct stimulation protocols: a step and hold (SH) protocol and a frozen noise (FN) protocol. **(a-b)** UMAP representation of spike waveforms from the **FN** and **SH** protocols, colored with their respective cluster identities (top) and their corresponding waveform shapes (bottom). **(c)**200 ms of a FN recording of a neuron with a broad-width spike shape (red) and narrow-width spike shape (blue). **(d)**200 ms SH trial of a neuron with a broad-width spike shape (red) and a narrow-width spike shape (blue). (**continued on the next page**)

(e) The confusion matrix shows an inter-class waveform comparison based on the cosine similarity across SH and FN classes. **(f)** The confusion matrix shows the cosine similarity-based comparison between waveform shapes for each SH class with all FN classes (row), and each SH class shows a similarity with at least 2 FN classes (> 0.8). **(g)** The mean and standard deviation of the modularity score for SH and FN protocols (left axis) and the number of clusters (right axis) as a result of running the clustering algorithm 25 times over 90% of the data sampled randomly each time, with changes in the resolution parameter ranging from 0 to 5 with increments of 0.5. The resolution parameter that maximizes the modularity score (black arrow) was chosen. (h) The heat map shows the likelihood of waveforms for each SH class being classified in one of the FN classes. Class identification between FN and SH protocols is inconsistent, as they show low values for the Adjusted Random Score (0.085) and Adjusted Mutual Information score (0.133).

We use the unsupervised UMAP+Louvain clustering method (explained in the previous section) on the 22-dimensional feature set that we extracted from the SH and FN trials for 186 neurons. We find 7 classes for SH and 7 classes for FN protocol respectively (Fig. 2.2**a-b**). Unlike in the waveform-based clustering, the UMAP representation is more continuous for the FN trials than for the SH protocols (Fig. 2.2 **a and b**). We visualize the differences attributes between the classes using a radar plot to investigate this further. We consider features based on spiking dynamics (such as firing rate, ISI, etc.), spike threshold, and action potential height and width separately for both SH and FN protocols (Fig. 2.2 **d-e** and methods for a list of all features and their sub-classification). We plot these features for each trial (thin line) along with the mean for each class (thick line). On visual inspection, we observe that the means for each cluster (thick line) in both SH and FN protocols are non-overlapping for all the 3 attribute sets (spiking dynamics, spike threshold, and AP height and width) respectively. Suggesting that each cluster has distinct action potential attributes. We measure the stability of the clusters (see methods) as a result of changing the hyperparameters (resolution parameter) for the unsupervised clustering method. We find the clusters to be stable (low standard deviation) for the chosen hyperparameter (resolution parameter $= 1.0$) (Fig. 2.2**c**).

We want to compare action potential attribute sets across SH and FN trials, so we quantify the differences across clusters for the FN and SH protocols using the cosine similarity between the feature vectors. We calculate the pairwise cosine similarity matrix between all SH and FN trials and take the means of the sub-matrix for neurons in each SH and FN cluster pair. This gives a cosine similarity matrix comparing each SH cluster with each FN cluster. We observe that none of the SH clusters show a high similarity (> 0.9) with FN clusters, suggesting that action potential attributes differ drastically as a result of input protocol (Fig. 2.2**f**). We also overlay the UMAP embedding for each feature vector for the SH and FN protocol (Appendix. 6.1**c**) and found that the SH and FN feature manifolds are entirely different.

To understand the role of input in overall action potential attributes based cluster assignment, we calculate the likelihood (see Methods) for a neuron in one of the SH clusters to be clustered in one of the FN clusters (Fig. 2.2**g**). We find that for each SH cluster, the likelihood for clustering in one of the FN clusters is spread over the FN clusters without a strong majority for a specific FN cluster. We further quantify the correspondence between SH and FN clusters using an adjusted random index score (ARI) and adjusted mutual information score (AMI), both values were found to be low (ARI $= 0.149$ and AMI $= 0.206$). This shows neurons cluster differently between SH and FN protocols based on Spike Event-based features. These results suggest that action potential attributes based clustering of neuronal populations is input dependent.

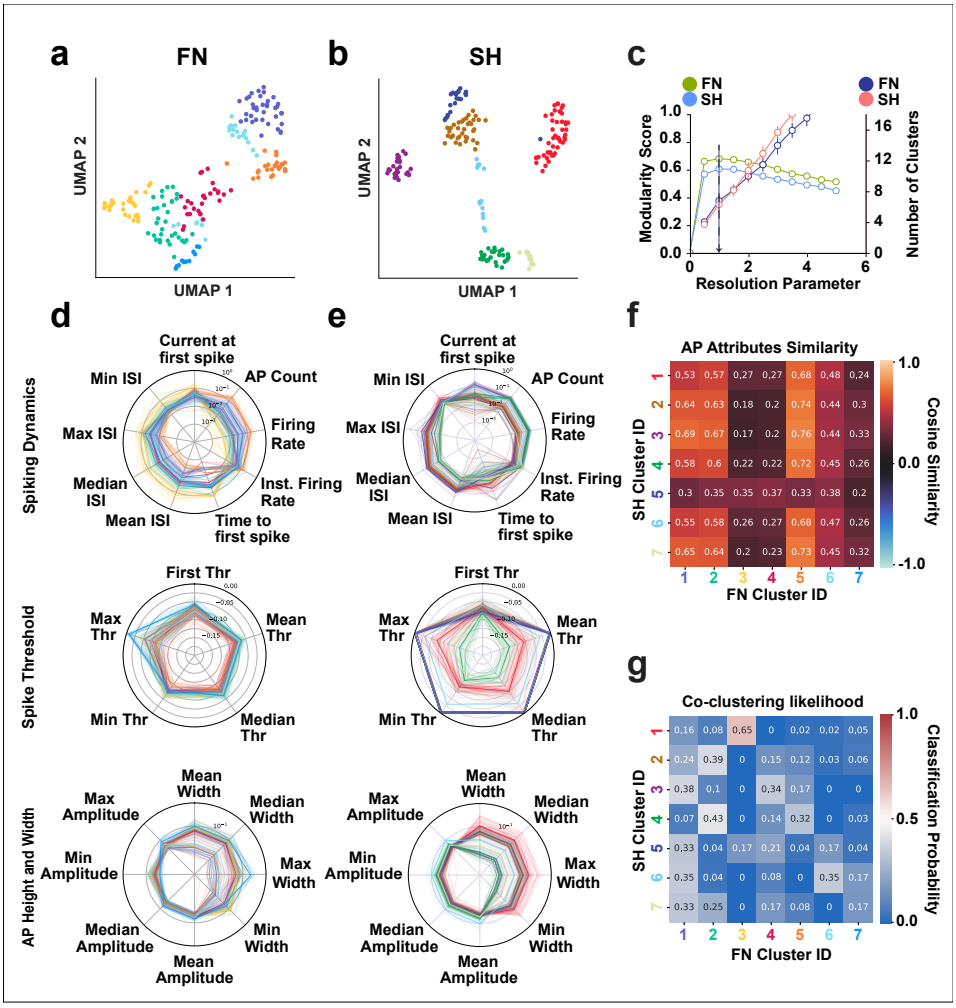

Figure 2.2. **Neural identity based on Action Potential attributes depends on the stimulus type.** (**a**-**b**) UMAP representation of Action Potential attributes for neurons was recorded with FN and SH protocols, and each color corresponded to a cluster identity. (**c**) shows the mean and standard deviation of the Modularity Score for SH and FN protocols (left axis) and the number of clusters (right axis) as a result of running the clustering algorithm 25 times over 80% of the data with changes in the resolution parameter ranging from 0 to 5 with an increment of 0.5, the resolution parameter is chosen that maximizes the Modularity score (black arrow). (**Continued on the next page**)

(**d-e**) The radar plot shows action potential attributes divided into spiking dynamics, spike threshold, and AP height and width features for each neuron, recorded with SH and FN protocols with each cluster in their respective color. Thick lines represent the mean over trials for each cluster. (**f**) Heatmap comparing action potential attributes using cosine similarity between SH and FN clusters. It can be seen that none of the values in the heatmap is greater than 0.8 suggesting that SH and FN action potential attributes are dissimilar. (**g**) Heatmap showing the likelihood of neurons to classify in one of the FN classes for each SH class. Note that the class identification between FN and SH protocols is inconsistent. The adjusted Random Score was 0.149 and the Adjusted Mutual Information score was 0.206, suggesting inconsistency between SH and FN clusters.

Functional classification of neurons stimulated by the FN protocol based on different feature sets

We have shown in the previous section that neuronal classification based on waveforms and action potential attributes is input-dependent. We want to expand our understanding of which features result in distinctive functional classifications within the FN-stimulated neurons. To understand the variance in the neuronal population captured by different attribute sets, we perform classifications based on four different attribute sets that capture 1) the commonly used action potential attributes, 2) passive biophysical attributes, 3) adaptation attributes, and 4) linear input filters approximated using the Spike Triggered Average (STA), to assess input feature selectivity.

Ample experimental evidence suggests that cortical neurons can be divided into two broad functional categories Avermann et al. (2012), namely excitatory (glutamatergic) and inhibitory (GABAergic), based on the type of effect (either excitation or inhibition) they have on their post-synaptic neurons Zeng and Sanes (2017). Excitatory and inhibitory neurons have also been found to have distinct electrophysiological properties and thus are known to perform different functions. We therefore subdivide our data into excitatory and inhibitory groups to study the diversity within and across populations. Previous studies have associated neuronal waveform shapes with functional identity Trainito et al. (2019); Lee et al. (2021). The broad and narrow spike-width

neurons have also been found to have a characteristic firing statistic, i.e. neurons with narrow-width waveforms were found to have a high firing rate (putatively inhibitory), and neurons with broad-width waveforms were found to have lower firing rates (putatively excitatory) Connors and Gutnick (1990); Bean (2007); Kiritani et al. (2023) in the barrel cortex. This suggests that inhibitory neurons can be putatively characterized by narrow width and high firing rate and that excitatory neurons can be putatively characterized by broad width and low firing rate. Also, barrel cortical excitatory neurons are more adaptive compared to inhibitory neurons Heiss et al. (2008). Based on this reasoning, we partitioned our data into a putative excitatory and an inhibitory population.

We extract the average intracellular waveforms from the entire dataset, 312 cells in total. Note that the number of neurons under analysis is much larger than in the previous sections because more FN than SH experiments were performed. We then apply the UMAP+Louvain algorithm to classify intracellular waveforms and find 8 clusters. (Fig. 2.3**a**) shows the UMAP projection of all waveforms with their corresponding cluster label colors. Next, we plot the distribution of the firing rates and half-widths for each cluster with matching colors using a violin plot (Fig. 2.3**b-c**). We observe clusters 1,5 and 6 to have a narrow width and high firing rate relative to the rest of the clusters; therefore, we categorize these clusters as putatively inhibitory (I) and the rest as putatively excitatory (E). We compare the average cosine similarity between E/I populations. We find that the excitatory and inhibitory populations are significantly different (Welch's ANOVA, ($F(2,49135) = 19148.00$, $p = 0.0$; Post-hoc Games-Howell test, E vs I ($p = 0.001$), I vs ExI ($p = 0.001$), and E vs ExI ($p = 0.001$), Fig. 2.3**d**)), but the excitatory population is more heterogeneous than the inhibitory one, based on the average similarity of the waveforms within the E/I population (Fig. 2.3**d**).

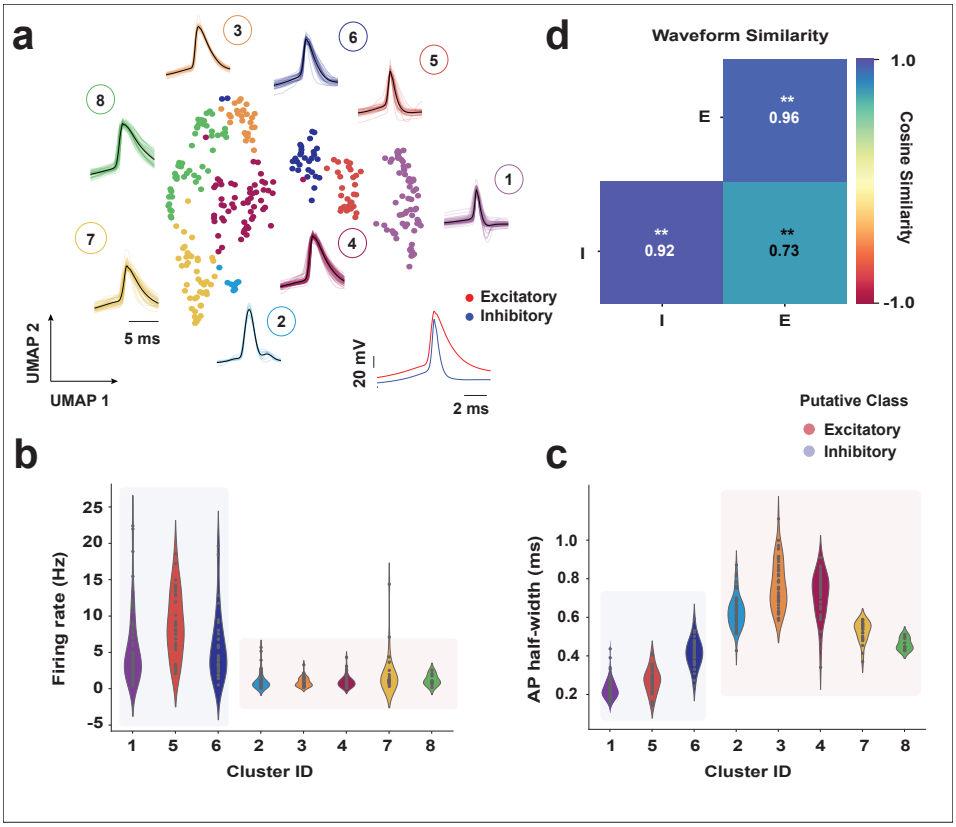

Figure 2.3. **Putative classification of excitatory and inhibitory neurons based on waveform and firing characteristics.** 312 cortical neurons were recorded using the FN protocol. Before studying neuronal classification based on different functional attributes, a putative excitatory and inhibitory classification of neurons was performed based on firing rate and AP half-width attributes. (**a**) The UMAP representation of the clustered average waveform along with their corresponding shape represented with the same color from 312 neurons recorded from layer 2/3 in the somatosensory cortex. (inset) average waveforms for putative excitatory and inhibitor populations. (**b**) Violin plot showing the distribution of firing rate across clusters identified in (**a.**), cluster IDs [1, 5, 6; n= 66, 30, 30] were found to have higher firing rates and are labeled as putatively inhibitory (n=126) and putatively excitatory (n=186). (**c**) shows the AP-half width across clusters identified in (a), cluster IDs [2, 3, 4, 7, 8; n= 58, 48, 44, 27, 9] were found to have lower half-width. (**d**) The average cosine similarity between putative excitatory and inhibitory classes. On average the putative Excitatory waveform group shows higher heterogeneity and excitatory and inhibitory waveforms differ. Cosine similarity measures between in-population excitatory (E), in-population inhibitory (I), and across excitatory-inhibitory populations (ExI) show significant differences. Welch's ANOVA, F(2, 49135) = 19148.00, ***p= 0.0); Post-hoc Games-Howell test, E vs I (**p= 0.001), I vs ExI (**p= 0.001), and E vs ExI (**p= 0.001).

Action potential attributes based neuron profiles using the FN protocol

As pointed out in the previous section, we aimed to understand and compare the usefulness of commonly used physiological attributes in uncovering functional classification in neuronal populations when neurons receive a physiologically realistic FN input. For that aim, we study commonly used action potential attributes for excitatory and inhibitory populations separately to discern if action potential attributes sufficiently capture within-population heterogeneity in excitatory and inhibitory groups. We also aim to unravel the differences between excitatory and inhibitory populations regarding their action potential attributes. For this, we extract 22 attributes (see Methods) subdivided into spiking dynamics, Spike threshold, and action potential height and width attributes with their descriptive statistics incorporating the mean, median, minimum, and maximum values.

We cluster the E/I populations separately based on the features and find 7 clusters for excitatory cells (shaded in red) and 6 clusters for inhibitory cells (shaded in blue). The UMAP representation of the spike-based properties for E/I populations with unique colors for each cluster for the chosen cluster parameter is shown in (Fig. 2.4**a-b**). We show the cluster stability (see Methods) in (Fig. 2.4**c**), the number of clusters is stable (low standard deviation) for the chosen resolution parameter (black arrow). We also test the stability of the clusters by excluding one attribute at a time and repeating the stability analysis (see Methods) and find the inhibitory clusters to be more stable to attribute exclusion (Appendix. 6.3**a-b**). We compare the means of action potential attributes simultaneously between excitatory and inhibitory populations and find them to be significantly different (one-sided MANOVA (see Methods), Wilks' lambda; $F(21, 290.0) = 35.1841$; $p = 0.000$, Pillai' s trace; $F(21, 290.0) = 35.1841$; $p = 0.000$, Hotelling-Lawley trace; $F(21, 290.0) = 35.1841$; $p = 0.000$, Roy' s greatest root; $F(21, 290.0) = 35.1841$; $p = 0.000$). To identify the relative importance of

each attribute in the separation of excitatory and inhibitory populations, we perform a canonical correlation analysis (CCA) (see Methods) between the excitatory and inhibitory action potential attributes and calculate the loading (structure correlation) for each attribute (Fig. 2.4**d**). We find that spiking dynamics attributes have the most influence on the inhibitory canonical variates suggesting that these variables have the most influence on separating the inhibitory population apart from the excitatory population. Alternatively, AP height and width attributes strongly influence the excitatory population canonical variate, suggesting that AP height and width are the most discriminatory. These results also show that individual action potential attributes contribute differently to the canonical variate (i.e., the latent structure) for excitatory and inhibitory populations respectively.

Next, we want to compare the level of heterogeneity within excitatory (E-E) and inhibitory (I-I) populations as well as across (I-E) populations using the action potential attributes. We visualize the action potential attributes using a radar plot shown in (Fig. 2.4**e-f**). It shows the diversity of spiking dynamics, spike threshold, and AP height and width for each neuron in each cluster (with the same color as in the UMAP projection (Fig. 2.4**a-b**)) along with the mean for the cluster (thick line). We observe that spiking dynamics attributes differ between excitatory and inhibitory clusters. The spiking threshold profiles were also found to be different across E/I clusters. Comparing the AP height and width attributes, however, show a similar profile across E/I clusters. We quantify the heterogeneity in the E/I clusters using a cosine similarity measure. The excitatory population had a significantly higher cosine similarity measure within the population than the inhibitory population (Welch's ANOVA, $F(2,49135)=3626998.65$, $p= 0.0$; Post-hoc Games-Howell test, E-E vs I-I ($p=0.001$), I-I vs I-E ($p=0.001$), and E-E vs I-E ($p=0.001$), (Fig. 2.4**g**)), suggesting that inhibitory action potential attributes are more heterogeneous than their excitatory counterparts. The mean cosine similarity score between excitatory and inhibitory populations are significantly lower than the

within-population cosine similarity score for both excitatory and inhibitory populations respectively, suggesting that excitatory and inhibitory feature action potential attribute vectors are different from each other, reiterating the MANOVA results. The results above demonstrate that action potential attributes are different between excitatory and inhibitory populations. The inhibitory population is more heterogeneous in its action potential attributes than the excitatory population.

Passive biophysical feature and adaptation-based profile using the FN protocol

As we describe in the last two sections, we aim to compare neuronal attribute sets to find the attributes(s) most informative about heterogeneity in excitatory and inhibitory populations and try to understand how neurons cluster based on these properties. The passive biophysical attributes (i.e., membrane resistance, capacitance, etc.) shape the response properties of a neuron. Moreover, it has been reported that adaptation current which captures the adaptation properties of the membrane is another important neuron property that can determine cell classes with noisy input Mensi et al. (2012). We, therefore, consider the passive biophysical attributes and the adaptation current as potential candidates that capture heterogeneity within the neuronal population. To understand the heterogeneity of passive biophysical properties and adaptation parameters across the excitatory and inhibitory populations, we extract both a set of passive parameters (see methods) and an adaptation current by fitting a Generalized Leaky Integrate and Fire (GLIF) model on the first 100 seconds of each FN trial, using the automated method described by Pozzorini et al. (2015) (see methods). (Fig. 2.5**a**) shows an example of a 10s instance of a GLIF-fitted model and one of the original recordings. We characterized the goodness of fit by measuring the explained variance (EV) between the subthreshold membrane potential of the original data and the model. (Fig. 2.5**b**) shows the distributions of the EV and Γ for the entire dataset. We eliminate the models with an EV value below 0.7 and obtain a set of 307 samples (out of 312

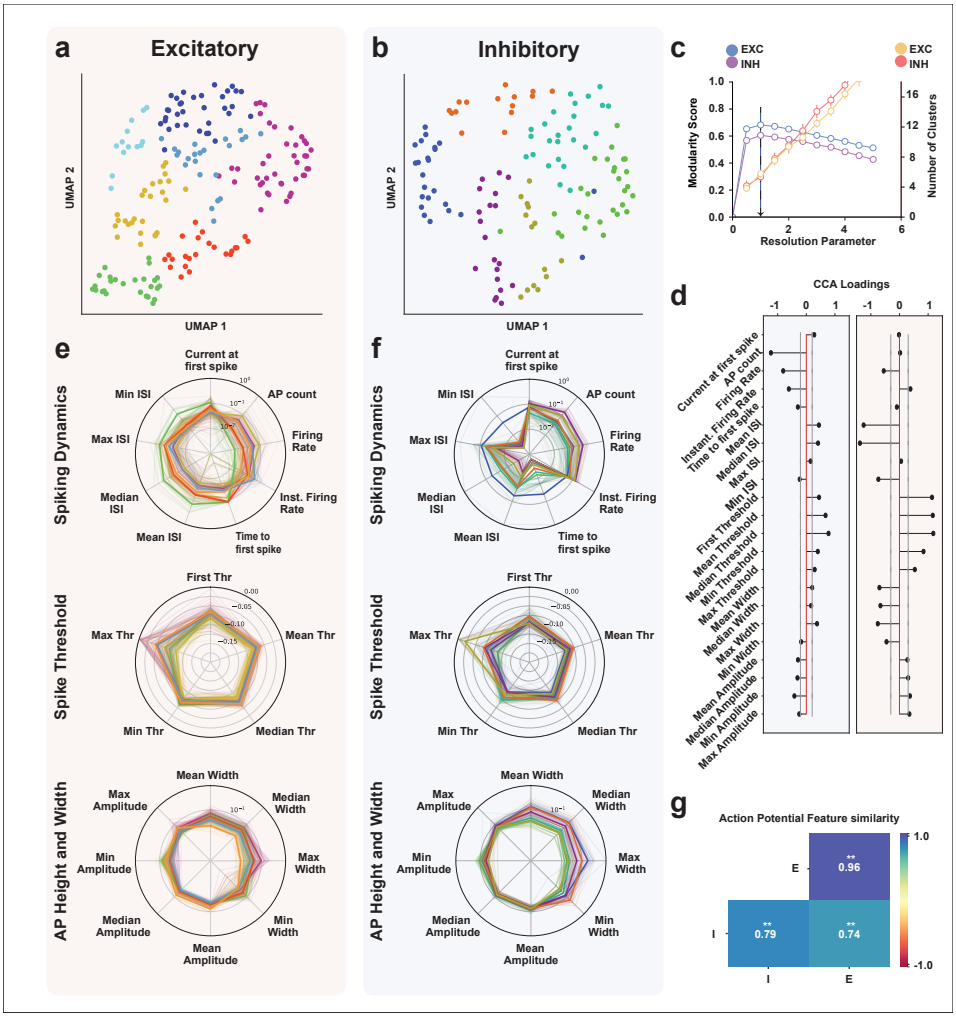

Figure 2.4. **Cell type-specific heterogeneity and the principal discriminatory attributes of the action potential.** NB Excitatory/Inhibitory results are grouped with red and blue backgrounds respectively. (**a-b**) The clustering procedure groups E/I neuron populations into 7 and 6 clusters respectively. (**c**) shows the chosen resolution parameters corresponding to the highest modularity and the corresponding number of clusters. (**d**) The means of excitatory-inhibitory action potential attributes were found to be significantly different (one-sided MANOVA, Wilks' lambda; $F(21, 290.0) = 35.1841$; $p = 0.000^{***}$, Pillai's trace; $F(21, 290.0) = 35.1841$; $p = 0.000^{***}$, Hotelling-Lawley trace; $F(21, 290.0) = 35.1841$; $p = 0.000^{***}$, Roy's greatest root; $F(21, 290.0) = 35.1841$; $p = 0.000^{***}$). We performed a post-hoc CCA analysis between the excitatory and inhibitory populations and showed the loadings for each attribute with a 95% confidence interval (grey line). (**e-f**) action potential attributes profile broken down into spiking dynamics, spike threshold, and AP height and width parameters plotted with the average for each cluster (thick line) and individual neurons (thin lines) and their corresponding cluster colors. Spiking dynamics parameters show the most variability across clusters for both excitatory and inhibitory groups.(**Continued on the next page**)

(**g**) The cosine similarity between E/I feature vectors. Inhibitory neurons are more heterogeneous in their action potential attributes than excitatory neurons. Welch's ANOVA shows a significant difference between cosine similarity measures between within-population excitatory (E-E), in-population inhibitory(I-I), and across excitatory-inhibitory populations (I-E) ($F(2, 49135) = 3626998.65$; $p = 0.0$); Post-hoc Games-Howell test, E-E vs I-I (**p= 0.001), I-I vs I-E ($**p = 0.001$), and E-E vs I-E ($**p = 0.001$).

samples).

We cluster the recordings into cell classes based on a set of passive biophysical parameters along with threshold adaptation constants (in total 6 features, see Fig. 2.5**a-b**) using the UMAP+Louvain clustering method explained in the previous sections, we find 7 clusters for the excitatory population and 6 clusters for the inhibitory population. We examine the stability of the clusters (Fig. 2.5**c**) with different hyperparameters (resolution parameters, see Methods) and choose (black arrow) the resolution parameter with the maximum modularity score. The clusters for the chosen hyperparameter for both excitatory and inhibitory populations were stable (low standard deviation). We also test the stability of the clusters by excluding one attribute at a time and repeating the stability analysis (see Methods) and found the inhibitory clusters to be more stable to attribute exclusion (Appendix. 6.3**c-d**). (Fig. 2.5 **d-e**) shows the UMAP representation of the feature set (inset) for excitatory (red background) and inhibitory (blue background) populations, the radar plot shows the individual passive biophysical parameter values for each neuron and each thin line in the radar plot represents a single neuron and are color matched to respective clusters, the mean for each cluster is plotted with a thick line. We compare the means of all the passive biophysical attributes simultaneously between excitatory and inhibitory populations and find them to be significantly different (one-sided MANOVA, Wilks' lambda; $F(6, 301) = 23.2629$; $p = 0.000$, Pillai's trace; $F(6, 301) = 23.2629$; $p = 0.000$, Hotelling-Lawley trace; $F(6, 301) = 23.2629$; $p = 0.000$, Roy's greatest root; $F(6, 301) = 23.2629$; $p = 0.000$). To identify the relative importance of each passive biophysical attribute in the separation between ex-

citatory and inhibitory populations, we perform a canonical correlation analysis (CCA) (see Methods) between the excitatory and inhibitory passive biophysical attributes and calculate the loading (structure correlation) for each attribute (Fig. 2.5**h**). We find dissimilar contributions from excitatory and inhibitory passive parameters towards their respective canonical variate, suggesting that each passive biophysical attribute contributes differently to the excitatory-inhibitory latent structure and therefore sets the excitatory-inhibitory populations apart. Next, we compare the level of heterogeneity between the passive feature vectors across the E/I populations (Fig. 2.5**i**) using the cosine similarity measure within and across excitatory populations and find no significant difference between the mean cosine similarity measure within excitatory and inhibitory populations as well as across excitatory-inhibitory population (due to low effect size) (Welch' s ANOVA, $F(2,66792) = 0.16777$; $p = 0.84$, Fig. 2.5**i**). These results suggest that even though the mean of the excitatory and inhibitory passive biophysical attributes are significantly different from each other (measured using MANOVA), the level of heterogeneity within excitatory-inhibitory populations respectively based on passive biophysical parameters is quite low. Therefore the passive biophysical attributes are not so informative about the neuronal heterogeneity.

As we saw above, passive biophysical attributes are not informative about population heterogeneity, this result is consistent with Mensi et al. (2012), which finds that passive biophysical attributes do not distinguish between neuron types. The adaptation current is more useful in putting neuron types apart. Therefore we classify the neurons using the adaptation current (η) extracted from the fitted GLIF model for the E/I population separately. We find 6 classes for the excitatory population (red) and 5 classes for the inhibitory population (blue). In (Fig. 2.5**f**-**g**) we show the UMAP representation of the adaptation current and the corresponding average normalized shapes for each cluster (inset) each cluster has its respective color in the UMAP as well as shape plots. We observe that the excitatory adaptation currents have a stronger negative amplitude

than their inhibitory counterparts. Moreover, the adaptation currents of the inhibitory population relax back to their resting values at earlier times than for the excitatory population.

We quantify the heterogeneity of adaptation currents within excitatory and inhibitory populations using the cosine similarity measure. The heatmap in (Fig. 2.5j) shows that the excitatory adaptation currents have a significantly higher average cosine similarity compared to the inhibitory adaptation currents, suggesting that the inhibitory adaptation current is more heterogeneous than the excitatory population (Welch's ANOVA; $F(2,46968) = 2796.77$, $p = 0.0$; Post-hoc Games-Howell test E vs I ($p = 0.001$), I vs ExI ($p = 0.75$), and E vs ExI ($p = 0.001$), Fig. 2.5**j**). The similarity measure is not significantly different between the inhibitory (I) population and across the inhibitory and excitatory (I vs E) population. This suggests that inhibitory adaptation currents are as different from each other as they are different from the excitatory adaptation currents. These results suggest that adaptation current profiles are different between excitatory and inhibitory populations. These results also show that adaptation currents are useful for understanding the heterogeneity within the inhibitory population but not so much for the excitatory populations.

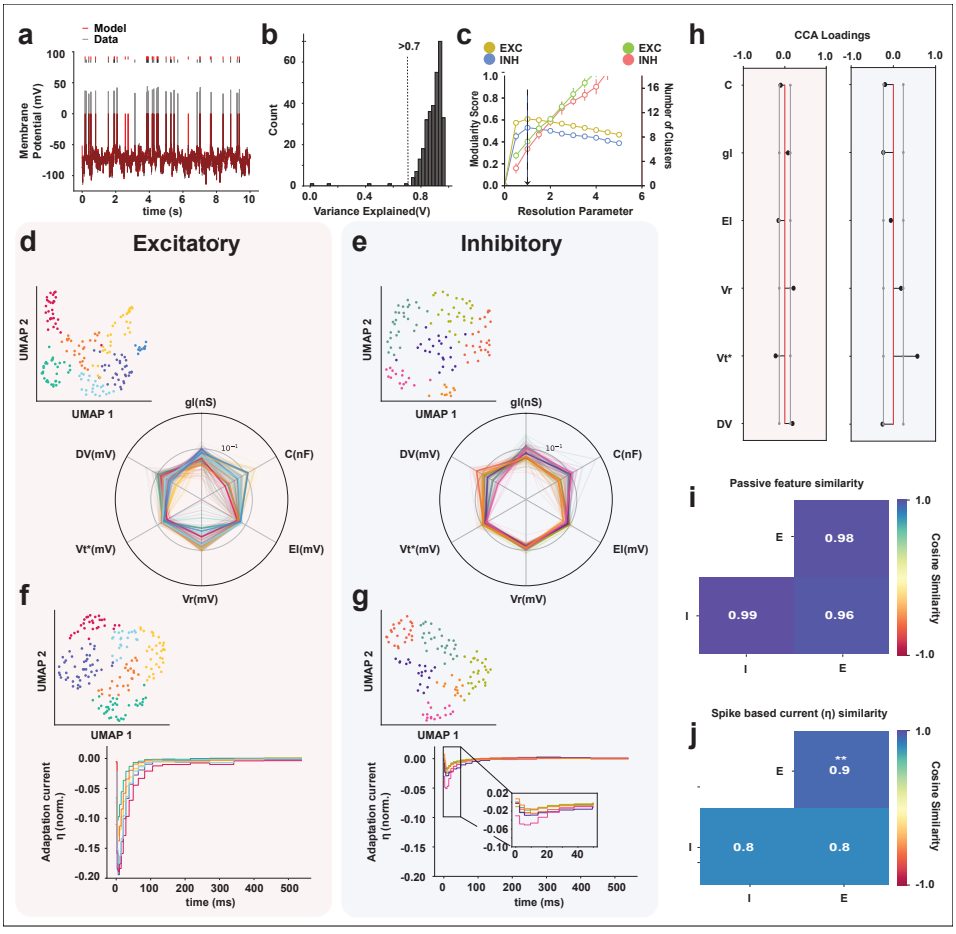

Figure 2.5. **Discriminatory power of passive biophysical and adaptive membrane properties for neuronal classification.** A GLIF model was fitted to the membrane potential recordings of neurons stimulated using the FN protocol. (**a**) An example of a 10 s long snippet of the GLIF model (red) with the original trial (grey). (**b**) Histogram showing the variance explained by the subthreshold membrane potential of the GLIF model. We included only models that exceeded a subthreshold variance explained higher than 0.7 (dashed line) for clustering purposes. (**c**) The chosen resolution parameters correspond to the highest modularity and the corresponding number of clusters for passive parameters. (**d-e**) UMAP representation of 7 clusters for excitatory (left) and 6 clusters for inhibitory (right) found using the passive biophysical parameters (inset). (**continued on the next page**)

The radar plot shows the passive biophysical profile of excitatory (left) and inhibitory cells (right), thick lines show the mean for each cluster and thin lines show the passive biophysical properties for each neuron colored in its respective cluster color. (**f-g**) UMAP representation of 6 clusters for excitatory (left) and 6 clusters for inhibitory (right) neurons found using the adaptation current (inset) and the corresponding adaptation current shape. (**h**) Post-hoc CCA between excitatory and inhibitory population and the loadings for each attribute with a 95% confidence (grey line). The excitatory-inhibitory passive biophysical attributes means were found to be significantly different (one-sided MANOVA, Wilks' lambda; $F(6,301) = 23.2629$; $p = 0.000^{***}$, Pillai's trace; $F(6,301) = 23.2629$; $p = 0.000^{***}$, Hotelling-Lawley trace; $F(6,301) = 23.2629$; $p = 0.000^{***}$, Roy's greatest root; $F(6,301) = 23.2629$; $p = 0.000^{***}$). (**i**) Heatmap showing passive biophysical attributes vectors compared within the excitatory (E-E) and inhibitory (I-I) population as well as across excitatory and inhibitory populations (I-E) using the cosine similarity measure, the values are averaged for I-I, E-E, and I-E comparisons. E/I population passive parameters were not significantly different for both within (I-I, E-E) and across (I-E) population comparisons (Welch's ANOVA, F(2, 66792) = 0.16777; p= 0.84). (**j**) Heatmap showing the cosine similarity-based heterogeneity measurement of normalized adaptation currents for E-E, I-I, and I-E populations. The average cosine similarity values of adaptation current currents were significantly higher for the E-E group than the I-I group suggesting that the inhibitory population is more heterogenous. The I-I and I-E mean cosine similarity values were not significantly different. (Welch's ANOVA, F(2,46968)=2796.77, **p= 0.0); Post-hoc Games-Howell test I-I vs E-E (**p=0.001), I-I vs I-E (p=0.75), and E-E vs I-E (**p=0.001)).

Neuronal classification based on linear input filter approximated using a Spike Triggered Average (STA) in the FN protocol

In last two sections, we explored neuronal heterogeneity based on action potential, passive biophysical, and adaptation attributes in neurons while responding to an FN input. We found that except for passive biophysical attributes, other properties show both within population across population heterogeneity for both excitatory and inhibitory populations. We observe a higher level of heterogeneity for inhibitory populations. Since we aim at understanding the extent to which various neuronal properties help uncover neuronal diversity, in this section we focus on the linear input filter of neurons and aim to understand the neuronal diversity based on this attribute. We consider this attribute important as neurons respond strongly to specific features in the input Sharpee (2014), making linear input filter an important property to study the functional diversity of the neuronal population.

The spike-triggered average (STA) method, as described in Schwartz et al. (2006), estimates a neuron's linear input filter by identifying the features in the input that trigger the neuron to spike. It does this by averaging the input signal over a specific time window preceding each spike, using data from all observed spikes. This makes the STA a useful method to approximate the linear input filter of a neuron. We want to investigate the STA diversity across the E/I population, for this, we extract the STA from all the neurons (see Methods). For calculating the STA, we use the injected current which is the result of shifting a theoretical dimensionless input with a constant baseline and scaling it by a factor as explained by Zeldenrust et al. (2017); da Silva Lantyer et al. (2018).

We cluster the excitatory and inhibitory population STAs separately using an unsupervised UMAP+Louvain clustering method (see Methods). We find 7 clusters for the excitatory population STAs (red) and 8 for the inhibitory population STAs. The

STAs were normalized and standardized before clustering (see Methods). The UMAP representation and the corresponding averaged and normalized STAs for each cluster are shown in (Fig. 2.6**a-b**). A visual inspection shows that inhibitory and excitatory STA shapes differ in their peak amplitudes and in the maximum slope of their initial rise (Fig. 2.6**a-b** inset), this difference is shown more clearly in (Fig. 2.6**d**). The stability of the clustering algorithm for the chosen parameters is shown in (Fig. 2.6**c**). To quantify the heterogeneity within the excitatory and inhibitory population as a result of that linear input filters, we first calculate the STA shape similarity between clusters, using a cosine similarity between excitatory and inhibitory STA clusters (Fig. 2.6**e-f**). The values shown in the heatmap in Fig. 2.6**e-f** were calculated by averaging the cosine similarity matrices between STAs of each cluster pair and within each cluster (for the within-cluster comparison, only the upper triangular values were included in the average). The average cosine similarity value between excitatory STA clusters shows higher values (> 0.9) with each other except for clusters IDs 2 and 6, which show a low similarity with every other cluster except for itself. On average, each excitatory cluster shows a high similarity (> 0.9) with more than 2 other clusters. On the other hand, the average cosine similarity values between inhibitory clusters is low (< 0.9), with some exceptions such as cluster IDs (2, 8), (3, 6), (5, 4), (6, 5) and (8, 6). On average, the STA of each inhibitory cluster is highly similar to approximately 1 other cluster. The average cosine similarity value is high between STAs within each inhibitory cluster. We summarize the STA similarities within the excitatory (E-E) and inhibitory populations (I-I) as well as across the two populations (I-E) in the heatmap in (Fig. 2.6**g**). This is done by averaging the cosine similarity matrices between STAs within excitatory (E-E) and inhibitory (I-I) populations as well as across excitatory and inhibitory (I-E) populations, these matrices are shown in (Appendix. **??b**). For within-population averages (E-E and I-I), we only take the upper-triangular portion of the matrix. The average cosine similarity value within the inhibitory population is significantly lower

than that of within the excitatory population (Fig. 2.6**g** Welch' s ANOVA, F(2,48825) = 1005.90, p = 0.0; Post-hoc Games-Howell test, E vs I (p = 0.001), I vs ExI (p = 0.001), and E vs ExI (p = 0.001)), suggesting that inhibitory STAs are more heterogeneous than excitatory ones. Also, the STA shape similarity between excitatory and inhibitory populations is significantly lower than the within excitatory and inhibitory similarity, suggesting that excitatory and inhibitory STAs are different. We can see that on average inhibitory neurons respond to more diverse features in the input than excitatory neurons.

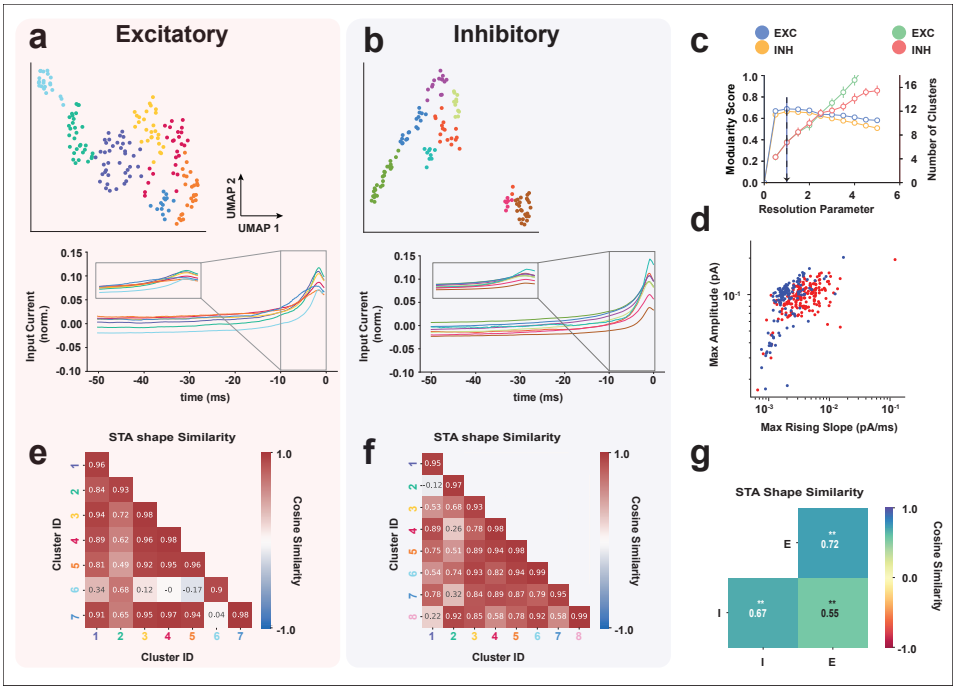

Figure 2.6. **Linear Input Filter measured using a Spike Triggered Average (STA)** (**a-b**) UMAP cluster representation for excitatory (left) and inhibitory (right) populations and the corresponding normalized average STA shape for each cluster in the 50 ms before a spike. Excitatory and inhibitory neuron populations were found to have 7 and 8 clusters respectively. (**c**) The chosen resolution parameters correspond to the highest modularity and the corresponding number of clusters. (**d**) Scatter plot showing the maximum slope vs maximum amplitude extracted from each STA (excitatory neurons depicted in red and inhibitory in blue). (**e-f**) Average cosine similarity across excitatory (e) and inhibitory (f) STA clusters. The excitatory clusters show a higher similarity with each other compared to inhibitory clusters. Each excitatory cluster on average shows high similarity (>0.9) with ≈ 2 other clusters, the inhibitory clusters on the other hand show a high similarity with ≈ 1 other cluster on average. (**g**) Heatmap showing the average cosine similarity for both within E/I and across E/I population STAs. The inhibitory population (I-I) shows a significantly higher heterogeneity (lower similarity) compared to the excitatory population (E-E). Welch's ANOVA shows a significant difference between the cosine similarity measures between excitatory (E), inhibitory (I), and across excitatory-inhibitory populations (ExI) $F(2,48825)=1005.90$, ***p= 0.0); Post-hoc Games-Howell test, E vs I (**p=0.001), I vs ExI (**p=0.0.001), and E vs ExI (**p=0.001).

Comparing Physiological Heterogeneity across attribute sets using Multi-set Correlation and Factor Analysis

We aim to explore what a physiologically realistic stimulus reveals about the functional heterogeneity of a neural population. For this aim, we need to compare FN-based neuronal attribute sets and find out which attribute set is the most informative about neuronal heterogeneity under the dynamic FN stimulation protocol. This is a complicated problem, as the four attribute sets are of varying dimensions. To perform this comparison, we use a method known as Multi-set Correlation and Factor Analysis (MCFA) Brown et al. (2023), which is an unsupervised multi-set integration method based on probabilistic Principal Component Analysis (pPCA) and Factor Analysis (FA), that can help to understand the common and shared factors across multi-modal data (see methods & Brown et al. (2023)). We use this method to compare the private and shared variance explained by the 4 attribute sets we examined in the previous sections (i.e., action potential attributes, passive biophysical attributes, adaptation currents, and linear input filters (STA)). We use the pPCA space to model the shared structure across attributes (see Methods), and the residuals are modeled as a private structure for each attribute set using factor analysis (see methods).

Comparing the variance explained by the shared structure across attributes, we found that the excitatory population which is inferred by a 4-dimensional shared structure, explains almost 50% of the variance in action potential and passive biophysical attributes (Fig. 2.7**a** top), this is higher than linear input filter (STA) (32.9%) and adaptation current attribute (19.7%). Most importantly, we see that the linear input filters (STAs) explain the highest private variance (57.2%), followed by adaptation current (41.2%). The action potential and passive biophysical attributes have a relatively lower private variance (35.7%). We can further investigate the contribution to the most important shared factor by each attribute, through which we find that action

potential and passive biophysical attributes explain the most variance for the most important shared dimension (Fig. 2.7**a**(middle, bottom)). Similarly, for the inhibitory population, explained by a 3 dimensional shared structure, we find that the linear input filter (STA) explains the most amount of private variance (80.6%), much higher than the excitatory population (Fig. 2.7**b** (top)), followed by adaptation current (29.4%), action potential attributes (35.7%), and passive biophysical attributes (17.1%) respectively. The action potential attributes and passive biophysical attributes explain the most and almost equal amount of shared variance (33.8% and 33. 85% respectively) (Fig. 2.7b (middle)), followed by adaptation current (29.4%) and linear input filter (9.9%) (Fig. 2.7**b** (top)). We find that the adaptation current and passive features explain the largest amount of variance for the most important shared dimension (Fig. 2.7**b** (bottom)) for the inhibitory population. It is important to observe that the linear input filter explains more than 85% of the total variances (shared+private) for both excitatory and inhibitory populations, higher than the other 3 attribute sets.

The high values of private variance explained by the linear input filters (STAs) in both excitatory and inhibitory cases show that passive biophysical, action potential, and adaptation currents are not highly correlated with the linear input filter of a neuron. Therefore, the linear input filter contains unique information about the excitatory and inhibitory population that is not shared with the other 3 attribute sets. Most importantly, the high value of private variance for the linear input filter, along with low residual values compared to other attribute sets indicates that linear input filters are likely to be the most informative attributes for explaining neuronal heterogeneity compared to passive biophysical, action potential, and adaptation current attributes. On the contrary, high values of shared variance for passive biophysical attributes and action potential attributes show that these attributes share a common structure, and can be a good predictor of one another. It is also important to observe that the total variance explained (shared + private) by adaptation currents is lower than action po-

tential attributes for both excitatory and inhibitory populations, suggesting that action potential attributes are more informative about neuronal heterogeneity than adaptation currents.

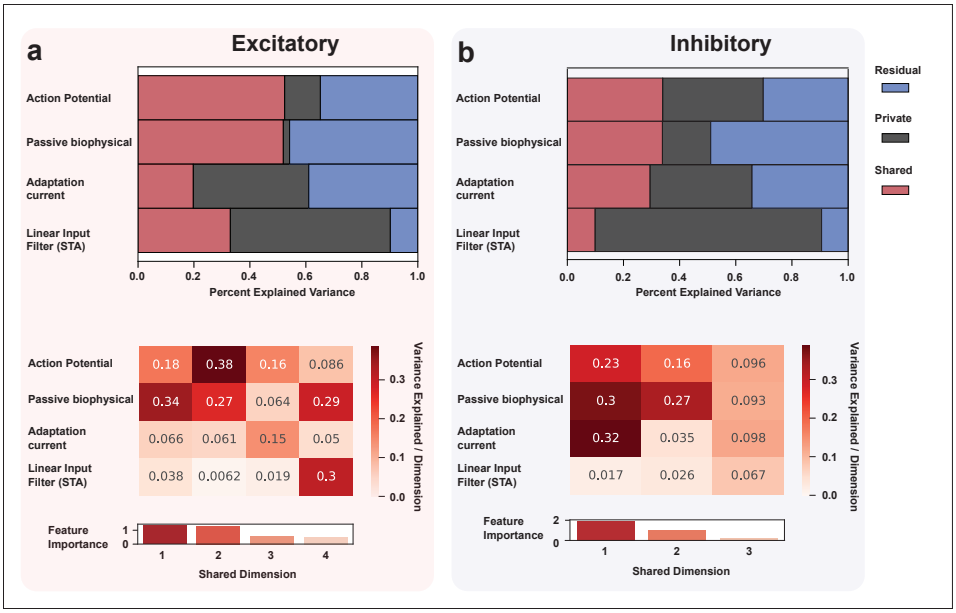

Figure 2.7. **MCFA analysis comparing shared and private variance explained by Action potential, Passive biophysical, Adaptation current, and Linear input filter (STA)**: (**a**) (top) Histogram showing the percentage of shared and private variance explained by each attribute for the excitatory population. action potential parameters (shared = 52.3%, private = 12.7%, residual = 34.8%) and passive biophysical parameters (shared = 51.9%, private = 2.2%, residual = 45.8%) explain most of the shared variance between attributes suggesting that these properties are not uniquely informative about population heterogeneity, on the other hand, linear input filter (STA) contains most of the unique information (shared = 32.9%, private = 57.2%, residual = 9.8%) about the excitatory population and the least amount of un-modelled residual, adaptation current explains more private variance (shared = 19.7%, private = 41.2%, residual = 38.9%) than action potential and passive biophysical features but less than linear input filter (STA). (middle) Heatmap showing the contribution of each shared dimension (total 4 based on selection criteria (methods)) in explaining variance across each attribute and their relative importance. (bottom) passive biophysical and action potential parameters have contributed to the most important shared dimensions. (**b**) (top) Histogram showing the percentage of shared and private variance explained by each attribute for the Inhibitory population. (**continued on the next page**)

Action potential parameters (shared = 33.9%, private = 35.7 %, residual = 30.2%) and passive biophysical parameters (shared = 33.85%, private = 17.1%, residual = 48.9%) explain the highest amount of shared variance between attributes suggesting that these properties are not uniquely informative about population heterogeneity on the other hand, passive biophysical parameters have the most amount of un-modeled (residual) variance. STA contains most of the unique information (private variance) about the inhibitory population (shared = 9.9%, private = 80.6%, residual = 9.4%) and the least amount of un-modeled (residual) variance followed by the adaptation current (private = 29.4%, shared = 36.29%, residual = 34.2%). Heatmap (middle) showing the contribution of each shared dimension (total 3, based on selection criteria (methods)) in explaining variance across each attribute and their relative importance. (bottom) Most important shared dimensions have contributions from passive biophysical, action potential parameters, and adaptation current. Based on this, a strong suggestion can be made for using linear input filter(STA) as the marker for heterogeneity in electrophysiological studies for both, excitatory and inhibitory populations. This demonstrates the importance of input in determining the difference across neural populations.

Discussion

In this work, we aimed to study the effect of input protocol (SH vs FN) on neuronal functional classification as well as to draw consensus on the attribute that are most informative about functional heterogeneity. For this, we first looked at the effect of the input protocol on neuronal classification based on two feature sets: waveforms and action potential features, while neurons were stimulated using two stimulus protocols: a static Step and Hold (SH) and a physiologically realistic Frozen Noise (FN) input. We found neuronal classification based on FN and SH protocols to be inconsistent when comparing the similarity in the cluster assignment in both input conditions using the co-cluster likelihood and cluster similarity indices (ARI and AMI). This highlights the importance of using a physiologically realistic input for studying functional heterogeneity of a neuronal population. We then aimed to determine which attribute(s) is/are the most informative about the neuronal functional heterogeneity when neurons are stimulated with a physiologically realistic FN input. We first explored the functional diversity of cells using 4 attribute sets (action potential, passive biophysical, adaptation current, and linear input filter (STA) attributes) for putative excitatory and inhibitory

populations. To infer which attribute(set) is most informative about neuronal heterogeneity, we compared the private and shared variance across the four attribute sets. We found that the linear input filters (STAs) explain the most variance, especially private variance across the excitatory and inhibitory neurons, and thus, contains unique information about neuronal functional heterogeneity compared to other attributes and is the most informative about neuronal functional heterogeneity.

The role of input in neuronal functional diversity

In the present work, we found that neuronal functional classification changes as a function of the input the neurons receive, this point is still undiscussed in neuronal clustering literature. Our results fill an important gap in the literature, the effect of input on physiological classification, by showing that waveforms and electrophysiological attributes-based classifications are different when neurons are stimulared with static versus dynamic inputs. Previous studies that combined electrophysiological, morphological, and molecular attributes for neuronal classification (Gouwens et al. (2019); Scala et al. (2019); Gouwens et al. (2020)), hypothesized that electrophysiological classification of a neuron can potentially change when using a different stimulus protocol. Our classification results confirm this hypothesis, by showing that clustering based on intracellular waveforms and action potential attributes are inconsistent across SH and FN-based inputs.

Our results showcase the importance of considering the functional dimension of neuronal identity that emerges from the interaction between input characteristics and intrinsic neuronal properties. These results do not only showcase how neurons cluster differently as a result of changing input protocol, which is an important and novel finding in the context of neuronal classification, but they also show that spiking dynamics, spike threshold, and AP height and width attributes were distinct between FN and SH protocols, as shown by the consistently low cosine similarities between FN

and SH clusters and the radar plots comparing the spiking dynamics, spike threshold, and AP height and width attributes for each neuron. This is in agreement with earlier results of Szabó et al. (2021); Hernáth et al. (2019); Szücs and Huerta (2015), which have already shown that firing intensity measures (AP count and AMPA conductance) are weakly correlated between static and dynamic stimuli (dynamic clamp) conditions. We extended these results by showing that the commonly studied action potential attributes span different latent manifolds as a result of input protocol (Appendix 6.1). Therefore, these classical action potential features do not encode a signature of the neurons that is invariant to the input, but are susceptible to the input the neurons receive. Since the conventionally used step and hold protocol does not represent an input a neuron receives in vivo, our findings advocate for using a physiologically realistic dynamic inputs such as the FN input for studying functional diversity. Our results therefore establish that input dynamics and its effects on functional classification and need to be considered even before molecular and morphological markers.

We used an unsupervised UMAP+Louvain clustering method in this study which has already been shown to distinguish neuronal classes based on extracellular waveforms (Lee et al. (2021)) and also has been shown to improve upon classifications based on a low number of features extracted from waveforms. We capitalized on this idea by using UMAP+Louvain clustering to cluster SH and FN intracellular waveforms and action potential attributes and then compared the cluster assignments across the two protocols. Since the Louvain community method clusters on the high-dimensional graph structure provided by the UMAP algorithm, rather than projecting the data on a low dimensional space Druckmann et al. (2013); Gouwens et al. (2019), which often leads to a loss of information about the latent structure Druckmann et al. (2013), the clusters using the UMAP graphs were more robust than found either using a dimensional feature set or a low dimensional projection of the original high dimensional feature set. A second important advantage of using a non-linear dimensionality reduction technique

like UMAP is that it allows for a manifold comparison between the two input conditions (see Appendix 6.1**(a-b)**).

Neuronal classification using frozen noise input-based attributes

We have established that neuronal functional classification is a result of the type of input that neurons receive, therefore a physiologically realistic input is important to understand neuronal functional diversity. A new question emerges: which attribute sets are the most informative about neuronal functional heterogeneity when neurons are presented with a dynamic stimuli? A similar question has been raised by (Zeng and Sanes (2017)). We attempted to provide a schema for answering this question. We did not intend to suggest a definitive number of classes of neurons, nor did we want to match our findings with previously established MET type classification, but we rather aimed to delineate a framework for drawing a consensus about which attribute(s) are the most informative about neuronal functional heterogeneity. An array of previous classification studies have relied upon variants of action potential and passive biophysical features Contreras (2004); Halabisky et al. (2006); McGarry et al. (2010); Casale et al. (2015), without any consensus about the informativeness of the attributes selected. Also, these studies have conventionally been reliant on feature extraction based on step stimulus protocols, which we demonstrate to produce cell classes that are different from when the neurons are stimulated with a dynamic input.

Although Hernáth et al. (2019); Szabó et al. (2021) provide a comparison between neurons responding to static and dynamic inputs, an active recommendation for which parameters are the most discriminatory was still missing. We supersede this by offering an alternative classification paradigm, based on the physiologically realistic frozen noise input. We tried to fill this gap by dividing the neuronal population into putative excitatory and inhibitory groups, then clustering neurons using the unsupervised UMAP+Louvain method, using four different sets of attributes separately. We found

that neuronal clusters are subjective to feature selection. There was little consistency in comparing cluster assignments across the four attribute sets (see Appendix 6.6**a-b**): the number of classes as well as the cluster assignments were found to be inconsistent across the four attribute sets.

We divided the neuronal population into putative excitatory and inhibitory classes based on the waveform shapes and firing rates, which is rather different from using molecular and morphological labels for E/I classification. Still, the goal of this study is not to align the neuronal identities to their ME-type markers, but rather to highlight the importance of input stimuli in functional classification. We show this by demonstrating that the firing properties of the narrow-width and broad-width neurons, which are conventionally categorized into inhibitory and excitatory classes respectively, change as a result of changing input type.

Parallel to investigating how putative E/I populations cluster based on different features, we also estimated the within-population variance for each feature set within E/I populations, which is representative of population heterogeneity based on a said feature. For action potential and passive biophysical attribute sets, we calculated the differences in the means of cosine similarity between E/I populations. We found that action potential attributes significantly differed between excitatory and inhibitory populations, consistent with previous findings Connors and Gutnick (1990). We also found that the action potential attributes for the inhibitory population was more heterogeneous than the excitatory population, which is also consistent with previous findings Gouwens et al. (2019, 2020). It is important to highlight that the number of clusters found in our analysis is similar across excitatory and inhibitory populations, which might result from our dataset containing more excitatory than inhibitory neurons. The clustering based on the second feature set, that of passive biophysical properties extracted by fitting a GLIF model to the neural recordings, showed that passive parameters were also significantly different between excitatory and inhibitory neurons, but within-population

heterogeneity was not significantly different between excitatory and inhibitory neurons. This suggests that passive parameters do not drive functional heterogeneity. This result is consistent with Mensi et al. (2012), which suggests that passive properties are not sufficiently discriminatory within E and I populations. Moreover, the authors claim that adaptive properties extracted using the GLIF model, such as the adaptation current, are more discriminatory than passive parameters. Our findings confirm this: adaptation currents show low similarity between excitatory and inhibitory populations, we found that the absolute maximum amplitude of all the adaptation current of the inhibitory neurons is smaller than that of the excitatory ones. We also found that inhibitory neurons have a significantly more heterogeneous adaptation profile than the excitatory population. This result provides data for designing and studying heterogeneous adaptive network models to further enhance our understanding of neural circuits' functional underpinnings.

Neurons can be functionally classified based on their input response features Famulare and Fairhall (2010); Sharpee (2014); Chéreau et al. (2020). We analyzed the linear input filters of neurons using the Spike Triggered Averages (STAs). This technique has been extensively used in studying the stimulus preference of neurons in the visual cortex Chichilnisky (2001); Rathbun et al. (2018). Our results show that the STA features are effective for estimating neuronal functional heterogeneity. We found that on average the STA shapes between the excitatory and inhibitory populations were different, based on comparing the average cosine similarity values between E and I populations, showing the difference in linear input filter between the two populations. We also found that the STA heterogeneity within the inhibitory population is significantly higher than the excitatory population. It has been shown Cardin et al. (2007) that visual cortical fast-spiking (putatively inhibitory) and regular-spiking (putatively excitatory) neurons have distinct levels of feature selectivity due to differences in passive biophysical attributes, such as the membrane time constant and input resistance.

Since fast-spiking neurons have higher membrane leak conductance, therefore lower resistance it results in sharpening of neuronal selectivity to its preferred input as shown in previous studies such as Li et al. (2020). Conversely, regular spiking neurons were found to have lower conductance and thus, lower sensitivity to preferred stimuli. Previous computational studies have found that physiological and passive biophysical have degenerative relationships with STA kernels Jain and Narayanan (2020), this study has shown that the same STA shape can be achieved by multiple difference values of the passive biophysical properties. This kind of degeneracy has been observed at the ion channel level as well Prinz et al. (2004); Marder and Taylor (2011). On the contrary, we observed a limited variability in biophysical features giving rise to a higher heterogeneity (comparing the within-population similarity matrices) in the STAs, i.e. in the functional linear input filter. We expect that the measured passive biophysical parameters from our data would be a good starting point to study the relationship between STA shape differences and the the range of passive biophysical attributes, this needs further computational modeling efforts. Since the diversity of feature preference by single neurons in the barrel cortex is not completely understood, a quantification of the functional heterogeneity observed in the linear input filter provided by our results is important for creating biophysically realistic models of cortical circuits and for a better understanding of circuit characteristics.

MCFA-based variance comparison across attribute sets

To determine which attribute set is most informative about neuronal functional heterogeneity, we compared the amount of neuronal population heterogeneity showcased by each attribute set by comparing the amount of private and shared variance explained by each attribute set. To our knowledge, the multi-attribute set comparison has not been done with physiological attributes before. This method provides a structured pathway to understand the limitations of commonly used electrophysiological features in cluster

studies and helps to reach a consensus about the choice of attributes to be used for functional classification. We found that linear input filter explained the highest amount of private variance of all the attribute sets, for both excitatory and inhibitory neuronal populations. This is a clear indication of the usefulness of a linear input filter (STA) as an attribute to explore functional heterogeneity. Contrarily, we found that passive biophysical attributes and action potential attributes explain the most shared variance for excitatory and inhibitory populations, suggesting that these attributes are correlated and contain similar latent structures. This is an important result that can aid in the debate around selecting a feature that is the most informative about neuronal functional heterogeneity. We expect our approach to provide a framework for comparing heterogeneity across other brain regions as well.

Limitations

There are several limitations to our study. Firstly, the sample size for the shared FN and SH comparison set was 186 cells, which might not be enough to capture all variability across the somatosensory cortex layer 2/3. For the second part of the study, the sample size was 312 neurons. We expect an increased sample size would increase confidence in functional neuronal clusters. It would be insightful to sample from all layers of the barrel cortex and compare the linear input filter across layers, to gain a complete insight into the barrel cortical functional diversity. It has been argued that the activity of a large population of neurons that captures a certain behavior can be approximated by a low-dimensional representation from a few neurons (Gao and Ganguli (2015)) and therefore that the number of recorded neurons should depend on the neural task complexity. Even though the Frozen Noise is a better approximation of the synaptic input than a Step-and-Hold stimulus, it might not be representative of the full dimensionality of the input that a neuron receives in vivo, as the FN input is based on somatic current injection. We understand that a somatic current injection doesn't

represent the full range of non-linear dendritic integration but it is quite helpful in estimating the linear input filter of a neuron which is found to be the most varying across neural populations and is sufficient for understanding the functional heterogeneity. We also lack the morphological and transcriptomic labels of the neurons, which makes the clusters found in this study incomparable to the commonly known types Zeng (2022). We expect a dynamic clamp-like setup using the frozen noise input along with morphological and transcriptomic labels to provide more clarity. Our study is also limited in providing a mechanistic relationship between the attribute sets we use, such as how the input changes the passive biophysical features and ultimately the linear input filter. Although the point GLIF model is quite helpful for extracting passive features, it does not provide a mechanistic description at ion channel level resolution for the variability in adaptation current and the feature selectivity. As explained above, we expect that a more detailed single neuron model, when studied with a physiologically realistic input, would provide a more elaborate picture of how the passive biophysical properties give rise to action potential and adaptation properties, which eventually result in the linear input filter. However, more detailed neuron models are difficult to fit and often do not produce unambiguous model properties Gonçalves et al. (2020); Nandi et al. (2022).

Conclusion

In conclusion, we show that neuronal functional classification is a function of the input protocol and therefore, a physiologically realistic input should be preferred for functional classification. We also established that linear input filters are the most distinguishing property, compared to action potential, passive biophysical, and adaptation current attributes, for understanding the functional diversity of neurons when stimulated with a physiologically realistic input. These results provide an important recommendation for neural taxonomists and electrophysiologists: to consider a neuron's physiological

input when defining neural identity. We expect computational single neuron as well as cortical network modeling efforts to discover the implications of the heterogeneity found in the 4 attribute sets of the excitatory and inhibitory neuronal populations we studied.

References

(2017). Proteomic landscape of the primary somatosensory cortex upon sensory deprivation. *Gigascience*, 6(10):gix082.

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *Database theory—ICDT 2001: 8th international conference London, UK, January 4–6, 2001 proceedings 8*, pages 420–434. Springer.

Avermann, M., Tomm, C., Mateo, C., Gerstner, W., and Petersen, C. C. (2012). Microcircuits of excitatory and inhibitory neurons in layer 2/3 of mouse barrel cortex. *Journal of neurophysiology*, 107(11):3116–3134.

Bean, B. P. (2007). The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451–465.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Bonald, T., de Lara, N., Lutz, Q., and Charpentier, B. (2020). Scikit-network: Graph analysis in python. *Journal of Machine Learning Research*, 21(185):1–6.

Brown, B. and collinwa (2023). collinwa/mcfa: First pypi release.

Brown, B. C., Wang, C., Kasela, S., Aguet, F., Nachun, D. C., Taylor, K. D., Tracy, R. P., Durda, P., Liu, Y., Johnson, W. C., et al. (2023). Multiset correlation and factor analysis enables exploration of multi-omics data. *Cell Genomics*, 3(8).

Cadwell, C. R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K. F., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using patch-seq. *Nature biotechnology*, 34(2):199–203.

Cardin, J. A., Palmer, L. A., and Contreras, D. (2007). Stimulus feature selectivity in excitatory and inhibitory neurons in primary visual cortex. *Journal of Neuroscience*, 27(39):10333–10344.

Casale, A. E., Foust, A. J., Bal, T., and McCormick, D. A. (2015). Cortical interneuron subtypes vary in their axonal action potential properties. *Journal of Neuroscience*, 35(47):15555–15567.

Chéreau, R., Bawa, T., Fodoulian, L., Carleton, A., Pagès, S., and Holtmaat, A. (2020). Dynamic perceptual feature selectivity in primary somatosensory cortex upon reversal learning. *Nature communications*, 11(1):3245.

Chichilnisky, E. (2001). A simple white noise analysis of neuronal lightresponses. *Network: computation in neural systems*, 12(2):199.

Connors, B. W. and Gutnick, M. J. (1990). Intrinsic firing patterns of diverse neocortical neurons. *Trends in neurosciences*, 13(3):99–104.

Contreras, D. (2004). Electrophysiological classes of neocortical neurons. *Neural Networks*, 17(5-6):633–646.

da Silva Lantyer, A., Calcini, N., Bijlsma, A., Kole, K., Emmelkamp, M., Peeters, M., Scheenen, W. J., Zeldenrust, F., and Celikel, T. (2018). A databank for intracellular electrophysiological mapping of the adult somatosensory cortex. *GigaScience*, 7(12):giy147.

Druckmann, S., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2013). A hierarchical structure of cortical interneuron electrical diversity revealed by automated statistical analysis. *Cerebral Cortex*, 23(12):2994–3006.

Famulare, M. and Fairhall, A. (2010). Feature selection in simple neurons: how coding depends on spiking dynamics. *Neural computation*, 22(3):581–598.

Fishell, G. and Heintz, N. (2013). The neuron identity problem: form meets function. *Neuron*, 80(3):602–612.

Fontaine, B., Peña, J. L., and Brette, R. (2014). Spike-threshold adaptation predicted by membrane potential dynamics in vivo. *PLoS computational biology*, 10(4):e1003560.

Fuzik, J., Zeisel, A., Máté, Z., Calvigioni, D., Yanagawa, Y., Szabó, G., Linnarsson, S., and Harkany, T. (2016). Integration of electrophysiological recordings with single-cell rna-seq data identifies neuronal subtypes. *Nature biotechnology*, 34(2):175–183.

Gao, P. and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155.

Gerstner, W., Kistler, W. M., Naud, R., and Paninski, L. (2014). *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press.

Gonçalves, P. J., Lueckmann, J.-M., Deistler, M., Nonnenmacher, M., Öcal, K., Bassetto, G., Chintaluri, C., Podlaski, W. F., Haddad, S. A., Vogels, T. P., et al. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *Elife*, 9:e56261.

Gouwens, N. W., Sorensen, S. A., Baftizadeh, F., Budzillo, A., Lee, B. R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K., et al. (2020). Integrated morphoelectric and transcriptomic classification of cortical gabaergic cells. *Cell*, 183(4):935–953.

Gouwens, N. W., Sorensen, S. A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S. M., Feng, D., Anastassiou, C. A., Barkan, E., et al. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. *Nature neuroscience*, 22(7):1182–1195.

Halabisky, B., Shen, F., Huguenard, J. R., and Prince, D. A. (2006). Electrophysiological classification of somatostatin-positive interneurons in mouse sensorimotor cortex. *Journal of neurophysiology*, 96(2):834–845.

Harris, K. D. and Shepherd, G. M. (2015). The neocortical circuit: themes and variations. *Nature neuroscience*, 18(2):170–181.

Heiss, J. E., Katz, Y., Ganmor, E., and Lampl, I. (2008). Shift in the balance between excitation and inhibition during sensory adaptation of s1 neurons. *Journal of Neuroscience*, 28(49):13320–13330.

Hernáth, F., Schlett, K., and Szücs, A. (2019). Alternative classifications of neurons based on physiological properties and synaptic responses, a computational study. *Scientific reports*, 9(1):13096.

Huang, Z. J. and Paul, A. (2019). The diversity of gabaergic neurons and neural communication elements. *Nature Reviews Neuroscience*, 20(9):563–572.

Jain, A. and Narayanan, R. (2020). Degeneracy in the emergence of spike-triggered average of hippocampal pyramidal neurons. *Scientific reports*, 10(1):374.

Kiritani, T., Pala, A., Gasselin, C., Crochet, S., and Petersen, C. C. (2023). Membrane potential dynamics of excitatory and inhibitory neurons in mouse barrel cortex during active whisker sensing. *Plos one*, 18(6):e0287174.

Kole, K. and Celikel, T. (2019). Neocortical microdissection at columnar and laminar resolution for molecular interrogation. *Current Protocols in Neuroscience*, 86(1):e55.

Kole, K., Zhang, Y., Jansen, E. J., Brouns, T., Bijlsma, A., Calcini, N., Yan, X., Lantyer, A. d. S., and Celikel, T. (2020). Assessing the utility of magneto to control neuronal excitability in the somatosensory cortex. *Nature neuroscience*, 23(9):1044–1046.

Lee, E. K., Balasubramanian, H., Tsolias, A., Anakwe, S. U., Medalla, M., Shenoy, K. V., and Chandrasekaran, C. (2021). Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *Elife*, 10:e67490.

Li, B., Routh, B. N., Johnston, D., Seidemann, E., and Priebe, N. J. (2020). Voltage-gated intrinsic conductances shape the input-output relationship of cortical neurons in behaving primate v1. *Neuron*, 107(1):185–196.

Marder, E. and Taylor, A. L. (2011). Multiple models to capture the variability in biological neurons and networks. *Nature neuroscience*, 14(2):133–138.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience*, 5(10):793–807.

Martens, M. B., Celikel, T., and Tiesinga, P. H. (2015). A developmental switch for hebbian plasticity. *PLoS Computational Biology*, 11(7):e1004386.

Masland, R. H. (2012). The neuronal organization of the retina. *Neuron*, 76(2):266–280.

McGarry, L. M., Packer, A. M., Fino, E., Nikolenko, V., Sippy, T., and Yuste, R. (2010). Quantitative classification of somatostatin-positive neocortical interneurons identifies three interneuron subtypes. *Frontiers in neural circuits*, 4:1284.

McInnes, L., Healy, J., and Melville, J. (2018a). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

McInnes, L., Healy, J., Saul, N., and Grossberger, L. (2018b). Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861.

Mensi, S., Naud, R., Pozzorini, C., Avermann, M., Petersen, C. C., and Gerstner, W. (2012). Parameter extraction and classification of three cortical neuron types reveals two distinct adaptation mechanisms. *Journal of neurophysiology*, 107(6):1756–1775.

Miceli, S., Nadif Kasri, N., Joosten, J., Huang, C., Kepser, L., Proville, R., Selten, M. M., van Eijs, F., Azarfar, A., Homberg, J. R., et al. (2017). Reduced inhibition within layer iv of sert knockout rat barrel cortex is associated with faster sensory integration. *Cerebral Cortex*, 27(2):933–949.

Mukamel, E. A. and Ngai, J. (2019). Perspectives on defining cell types in the brain. *Current opinion in neurobiology*, 56:61–68.

Nandi, A., Chartrand, T., Van Geit, W., Buchin, A., Yao, Z., Lee, S. Y., Wei, Y., Kalmbach, B., Lee, B., Lein, E., et al. (2022). Single-neuron models linking electrophysiology, morphology, and transcriptomics across cortical cell types. *Cell reports*, 40(6).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Perktold, J., Seabold, S., Sheppard, K., ChadFulton, Shedden, K., jbrockmendel, j grana6, Quackenbush, P., Arel-Bundock, V., McKinney, W., Langmore, I., Baker, B., Gommers, R., yogabonito, s scherrer, Zhurko, Y., Brett, M., Giampieri, E., yl565, Millman, J., Hobson, P., Vincent, Roy, P., Augspurger, T., tvanzyl, alexbrc, Hartley, T., Perez, F., Tamiya, Y., and Halchenko, Y. (2024). statsmodels/statsmodels: Release 0.14.2.

Poulin, V. and Théberge, F. (2019). Ensemble clustering for graphs. In *Complex Networks and Their Applications VII: Volume 1 Proceedings The 7th International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2018 7*, pages 231–243. Springer.

Pozzorini, C., Mensi, S., Hagens, O., Naud, R., Koch, C., and Gerstner, W. (2015). Automated high-throughput characterization of single neurons by means of simplified spiking models. *PLoS computational biology*, 11(6):e1004275.

Prinz, A. A., Bucher, D., and Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nature neuroscience*, 7(12):1345–1352.

Rathbun, D., Ghorbani, N., Shabani, H., Zrenner, E., and Hosseinzadeh, Z. (2018). Spike-triggered average electrical stimuli as input filters for bionic vision—a perspective. *Journal of neural engineering*, 15(6):063002.

Scala, F., Kobak, D., Bernabucci, M., Bernaerts, Y., Cadwell, C. R., Castro, J. R., Hartmanis, L., Jiang, X., Laturnus, S., Miranda, E., et al. (2021). Phenotypic variation of transcriptomic cell types in mouse motor cortex. *Nature*, 598(7879):144–150.

Scala, F., Kobak, D., Shan, S., Bernaerts, Y., Laturnus, S., Cadwell, C. R., Hartmanis, L., Froudarakis, E., Castro, J. R., Tan, Z. H., et al. (2019). Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas. *Nature communications*, 10(1):4174.

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *Journal of vision*, 6(4):13–13.

Sharpee, T. O. (2014). Toward functional classification of neuronal types. *Neuron*, 83(6):1329–1334.

Steriade, M. (2000). Corticothalamic resonance, states of vigilance and mentation. *Neuroscience*, 101(2):243–276.

Szabó, A., Schlett, K., and Szücs, A. (2021). Conventional measures of intrinsic excitability are poor estimators of neuronal activity under realistic synaptic inputs. *PLOS Computational Biology*, 17(9):e1009378.

Szücs, A. and Huerta, R. (2015). Differential effects of static and dynamic inputs on neuronal excitability. *Journal of Neurophysiology*, 113(1):232–243.

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, 563(7729):72–78.

Trainito, C., von Nicolai, C., Miller, E. K., and Siegel, M. (2019). Extracellular spike waveform dissociates four functionally distinct cell classes in primate cortex. *Current Biology*, 29(18):2973–2982.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Yan, X., Calcini, N., Safavi, P., Ak, A., Kole, K., Zeldenrust, F., and Celikel, T. (2022). A whole-cell recording database of neuromodulatory action in the adult neocortex. *bioRxiv*, pages 2022–01.

Zeldenrust, F., de Knecht, S., Wadman, W. J., Denève, S., and Gutkin, B. (2017). Estimating the information extracted by a single spiking neuron from a continuous input time series. *Frontiers in computational neuroscience*, 11:49.

Zeng, H. (2022). What is a cell type and how to define it? *Cell*, 185(15):2739–2755.

Zeng, H. and Sanes, J. R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, 18(9):530–546.

Chapter 3

Neuromodulatory Control of Cortical Function: Cell-Type Specific Reshaping of Neuronal Information Transfer

Based on: Willemsen, Y., Beijers, R., Gu, F., Arias Vásquez, A., Schols, H.A., de Weerth, C. Fucosylated Human Milk Oligosaccharides during the First 12 Postnatal Weeks are Associated with Better Executive Functions in Toddlers. *Nutrients*, 2023, 15, 1463.

Abstract

Neuromodulatory systems are crucial for cognitive flexibility, dynamically altering neuronal and circuit properties. However, how these systems control information transfer, particularly in a cell-type and receptor-specific manner, remains poorly understood. Here, we investigated how dopaminergic (D1-R, D2-R) and cholinergic (M1-R) receptor activation reconfigures the input-output processing of excitatory and inhibitory neurons in layer 2/3 of mouse somatosensory cortex. Using in-vitro whole-cell recordings under a Frozen Noise stimulus protocol, we characterized neurons by four sets of functional attributes: action potential dynamics, passive biophysical properties, adaptation currents, and input feature selectivity (via spike-triggered averages). We found that neuromodulator receptor activation alters information transfer between stimulus and spike train in a manner dependent on both cell type and receptor. Unsupervised clustering (UMAP+Louvain) of attribute sets revealed that neuromodulation dynamically reshapes neuronal functional identities. Critically, multi-set correlation and factor analysis (MCFA) demonstrated that neuromodulators reorganize the covariance structure between these attribute sets. For instance, D1-R and D2-R activation increased the independence of input feature selectivity (STA) in excitatory neurons while enhancing coordination among output-related features. Conversely, inhibitory neurons generally exhibited increased coordination across functional domains. These findings demonstrate that neuromodulators do not merely tune individual neuronal properties but orchestrate a complex, coordinated reconfiguration of the functional landscape, thereby dynamically reshaping the computational capabilities of cortical circuits.

Introduction

Neuromodulators such as dopamine and acetylcholine play a critical role in shaping circuit development as well as brain states and cognitive functions, including attention, learning, memory, and sleep Dalley et al. (2004); Hasselmo et al. (2006); Sarter et al. (2009). These molecules do not directly evoke neural activity but also modulate how neurons respond to synaptic input, tuning the input–output relationships of individual neurons. Disruptions in neuromodulatory systems have been implicated in neurological and psychiatric disorders such as Parkinson's disease and schizophrenia Durstewitz and Seamans (2008); Arnsten and Rubia (2012); Winterer and Weinberger (2004).

Despite extensive work on their cellular and behavioral effects, it remains unclear how neuromodulators influence what neurons compute, specifically, how they encode and transmit information. Traditional approaches have primarily focused on measuring the effects of neuromodulation on single variable measures such as firing rate or spike-frequency adaptation Nadim and Bucher (2014); Marder (2012); Shine et al. (2021). While informative, these scalar metrics fail to capture the complexity of neuronal computation, which emerges from the dynamic interaction of passive membrane properties, intrinsic excitability, synaptic integration, and stimulus selectivity. Furthermore, little is known about whether neuromodulatory effects act independently on each property or reshape the correlation between properties in a structured way, particularly in a receptor- and cell-type-specific manner.

Neurons are high-dimensional dynamical systems. Understanding how neuromodulation alters the relationships among functional attributes such as action potential dynamics, adaptation, and input filtering is critical for revealing how neurons adjust their computational roles within a circuit. Dopaminergic (via D1-R and D2-R) and cholinergic (via M1-R) signaling pathways are known to modulate ion channel activity and membrane excitability Bargmann (2012); Marder (2012); Taghert and Nitabach

(2012), thereby influencing both spike generation and input selection Dascal (2001); Seong and Carter (2012); Cousineau et al. (2020). These effects are mediated through receptors, expressions of which varies across cell types Nusser (2009), suggesting that neuromodulators may reconfigure cortical circuits in a subtype-specific manner. Moreover, cell properties themselves are independent. For instance, passive properties such as membrane conductance shape downstream attributes like spike waveform and filtering dynamics Häusser (2000); Ferguson and Cardin (2020); Salinas and Sejnowski (2001), this raises the possibility that neuromodulatory effects may produce coordinated, rather than isolated, functional changes. To test this, we asked: How does receptor-specific neuromodulation reorganize the functional encoding landscape of cortical neurons?

We addressed this question using in-vitro whole-cell recordings from excitatory and inhibitory neurons in layer 2/3 of the mouse somatosensory cortex. Each neuron was stimulated with a biophysically realistic time-varying current generated by the Frozen Noise protocol (FN) Zeldenrust et al. (2017), and recordings were obtained under control (aCSF) and receptor-specific agonist conditions (D1-R, D2-R, M1-R). This protocol allowed us to measure both the information transferred by each neuron and the detailed physiological attributes underlying that computation.

Firstly, we quantify how neuromodulation affects the amount of information that is transferred from the input of a neuron to its output Zeldenrust et al. (2017). We find that neuromodulation changes information transfer in a cell-type specific manner. Secondly, we analyze the effects of neuromodulation in high-dimensional electrophysiological property space. From each recording, we extracted four sets of functional attributes: (1) action potential (AP) dynamics, (2) passive biophysical (PB) properties, (3) adaptation currents (AC), and (4) input filtering via spike-triggered average (STA). These feature sets each capture a different aspect of neuronal function, from passive to active (input-driven). We then examined whether receptor activation reorganized the functional identity of neurons. Specifically, for each set of attributes, we

were interested to see whether neurons that would cluster together in unsupervised high-dimensional clustering (UMAP + Louvain) under the aCSF condition, would also cluster together under neuromodulation. Finally, we applied multi-set correlation and factor analysis (MCFA) to assess whether neuromodulation alters the correlation between attribute sets, reflecting a global reconfiguration of a neuronal population's computational architecture.

Together, these analyses reveal how dopaminergic and cholinergic modulation reshape the functional landscape of cortical neurons. By linking receptor-specific activation to changes in how neurons encode stimuli through shifts in high-dimensional functional attribute sets, this study provides a systems-level framework for understanding neuromodulation in both single neuron and circuit level which is crucial for studying disease such Parkinson's and schizophrenia as well development of biologically realistic neural networks.

Methods

Ethics statement The data used in this research was previously published and made freely available to the community da Silva Lantyer et al. (2018) and Yan et al. (2022). All experimental work, as described in the articles cited, was carried out in accordance with the European directive 2010/63/EU, the Dutch national regulations and international standards for animal care and use.

Slice electrophysiology The experimental procedures were described before da Silva Lantyer et al. (2018); Yan et al. (2022). In short, slices were prepared from adult mice expressing Cre recombinase under the control of either the parvalbumin promoter (RRID: MGI:5315557) or the somatostatin promoter (RRID: IMSR JAX:013044), backcrossed to the C57BL/6 background. The animals were housed in a 12-hour light / dark cycle with ad libitum access to food and water and kept in family cages until the day of the experiment.

The experimental procedures followed previously published protocols (**??????**). Coronal slices (thickness: 300 microns) were cut from the barrel cortex subregion of the primary somatosensory cortex, and individual neurons were visualized using differential interference contrast (DIC) optics at 40× magnification before whole-cell access using pipettes with a resistance of 5-9 $M\Omega$. The internal (pipette) solution contained (in mM): 130 K-gluconate, 5 KCl, 1.5 $MgCl_2 \cdot 6H_2O$, 0.4 Na_3GTP, 4 Na_2ATP, 10 HEPES, 10 Na-phosphocreatine, and 0.6 EGTA. The pH was adjusted to 7.22 with KOH. Recordings were performed using HEKA EPC10 amplifiers and PatchMaster software (v2x90.2).

Targeted pharmacological activation of the select receptors was performed using agonists for the serotonin $5-HT_1F$ receptors (LY334370, $5{\cdot}\mu M$), muscarinic acetylcholine M1 receptors (McN-A-343, $20{\cdot}\mu M$), dopamine D1 (SKF38393, $1{\cdot}\mu M$) and D2 receptors (Quinpirole, $10{\cdot}\mu M$). All compounds were purchased from Sigma-Aldrich and dissolved in ACSF Yan et al. (2022).

Each neuron served as its own control, with baseline recordings acquired in ACSF prior to drug application. Drug effects were assessed beginning 3 minutes after compound introduction into the continuously perfused bath. Following each experiment, the recording chamber was thoroughly cleaned, and dedicated containers were used for each drug solution to prevent cross-contamination Yan et al. (2022).

Frozen Noise (FN) protocol The Frozen Noise input protocol consisted of injecting a somatic current that is the result of an artificial neural network of 1000 neurons responding (firing Poisson spikes) to random stimuli i.e., the hidden state, the membrane potential response to the somatic input is recorded with a sampling rate of 20 kHz for a total length of 360 seconds and saved. Each raw data file consisted of a vehicle control trial (artificial Cerebrospinal fluid i.e. aCSf) and a drug trial (a specific neuromodulatory receptor agonist or antagonist was added to the bath and the recording was repeated). Some files consisted of multiple control and drug trials. See

Zeldenrust et al. (2017); da Silva Lantyer et al. (2018) for more details. A schema for the protocol is shown in Fig. 3.1. In total 288 neurons (Table. 3.1) were analyzed, we discarded recording sets with high levels of noise.

Condition	Cell-type	Trials
aCSF	Excitatory	183
aCSF	Inhibitory	121
D1	Excitatory	50
D1	Inhibitory	36
D2	Excitatory	40
D2	Inhibitory	19
M1	Excitatory	19
M1	Inhibitory	20

Table 3.1. Number of recording set for control and agonist conditions

Analysis

Feature Extraction

We extracted waveform, action potential, passive biophysical and spike triggered average from single neuron recordings recorded under Frozen Noise input protocol, these recordings were performed first under a vehicle control (aCSF) and then repeated with a receptor agonist added to the bath.

Spike waveforms As explained in Joshi et al. (2024), we identified peaks from the membrane potential traces and kept the hyperparameters and ISI threshold criteria the same as in Joshi et al. (2024). The length of the waveforms used in this study is 10ms (5ms before and after the peak).

Action potential attributes The action potential attributes were extracted to study the dynamics, threshold and waveforms related attributes throughout a trial via descriptive statistics. The action potential attributes were extracted for aCSF and agonist trails as described in Joshi et al. (2024). A summary of all the Action potential attributs is provided in Table. 3.2.

Feature Group	**Feature**	**Description**	**Summary Statistics**
Spiking Dynamics	Current at first spike	Current amplitude at which the neuron first crosses threshold to fire a spike.	Single value per trial.
	AP count	Total number of action potentials generated during a trial.	Single value per trial.
	Time to first spike	Time (in ms) from trial start to first spike.	Single value per trial.
	Firing rate	Total number of spikes divided by trial duration (spikes/sec).	Single value per trial.
	Interspike Interval (ISI)	Time between successive spikes ($t_{n+1} - t_n$).	Mean, median, max, min calculated.
	Instantaneous rate	Reciprocal of ISI ($1/(t_{n+1} - t_n)$).	Mean, median, max, min calculated.
Spike Threshold	Spike threshold	Voltage at which dV/dt first exceeds 25 mV/ms. Calculated per spike.	Mean, median, max, min computed.
AP Height & Width	Width	Time from threshold crossing to return below threshold after AP peak.	Mean, median, max, min calculated.
	Amplitude	Voltage difference between AP peak and threshold.	Mean, median, max, min calculated.

Table 3.2. Summary of Action Potential Attributes and Their Descriptions

Passive Biophysical Feature extraction using GLIF model

In order to extract passive biophysical attributes as well as adaptation current from aCSF and agonist trials, we fit a Generalized Leaky Integrate and Fire (GLIF) neuron model Pozzorini et al. (2015) as described in detail in Joshi et al. (2024). We take the first 100 second from the recording as the training set to extract passive attributes as well as adaptation current from the recording. A summary of all the passive biophysical features extracted is provded in Table. 3.3. Using the same GLIF model fitted to neural recordings, we also extracted the adaptation current $\eta(t)$ triggered by a spike event, see Joshi et al. (2024); Pozzorini et al. (2015).

Feature	**Description**
Membrane Capacitance (C)	The cell membrane' s ability to store charge; influences how quickly voltage changes in response to current.
Leak Conductance (g_L)	Governs the passive flow of ions across the membrane; contributes to the rate of membrane potential decay.
Resting Potential (E_L)	The membrane voltage the cell settles at in the absence of input; baseline membrane potential.
Sharpness of Spike Threshold (ΔV)	Controls how sharply the firing probability increases as membrane potential approaches threshold.
Threshold Baseline (V_T^*)	The baseline value of the dynamic spike threshold; determines the average voltage required to trigger a spike.
Reset Voltage (V_{reset})	The membrane potential value the neuron resets to after a spike and refractory period.

Table 3.3. Passive Biophysical Features Extracted from the GLIF Model

Spike Triggered Average The spike-triggered average (STA) is the average shape of the stimulus that precedes each spike. We extracted the STA using the following

equation given by Schwartz et al. (2006):

$$STA = \frac{1}{N}\Sigma_{n=1}^{N}\vec{s}(t_n), \tag{3.1}$$

where t_n is the n^{th} spike time, s is the stimulus vector preceding the spike for a fixed time window of 100 ms, and N is the total number of spikes. Before clustering, we standardize (i.e. z score) and then normalize the STA vector with an L_2 norm. We didn't use any kind of whitening or regularization to calculate the STA.

UMAP + Louvain clustering

Universal Manifold Approximator (UMAP) is a non-linear dimensionality reduction algorithm which is advantageous for preserving global structure of the data McInnes et al. (2018) in lower dimensions, this makes it more suitable for visualization especially for high dimensional datasets ($p >> N$, where p is the dimensionality of the data and N is number of samples) compared to other methods such as PCA which fail to perform due to curse of dimensionality Aggarwal et al. (2001). As explained in Lee et al. (2021); Joshi et al. (2024) the high-dimensional graph obtained during the intermediate step in the UMAP algorithm can be exploited to perform clustering using Louvain community detection Blondel et al. (2008). We chose the hyperparameter based on cluster stability criteria and the corresponding number of clusters based on the same heuristic as explained in Joshi et al. (2024).

For measuring clustering similarity between aCSF and drug conditions, we calculate a cluster similarity matrix as explained in Joshi et al. (2024). For quantifying the similarity between labels assigned to aCSF versus drug trials, we used cluster similarity measure such as adjusted random index and adjusted mutual information score using scikit-learn Python package Pedregosa et al. (2011).

Information transfer protocol

We employed the information transfer protocol detailed in Zeldenrust et al. (2017, 2024) to quantify how much information a single neuron extracts from its inputs. This protocol assumes that neurons respond to the presence or absence of a preferred stimulus, which is modeled as a binary hidden state ($\mathbf{x}$) that switches between 0 and 1 according to a memoryless Markov process. Importantly, the neuron does not directly observe this hidden state; instead, it receives input from a large population of simulated presynaptic neurons. Each presynaptic neuron fires a Poisson spike train, with firing rates q^i_{on} when the hidden state is ON ($x = 1$), and q^i_{off} when it is OFF ($x = 0$).

To generate the input current ($\mathbf{I}$), the spike train of each presynaptic neuron is convolved with a 5 ms exponential kernel and then weighted by $w_i = \log\left(\frac{q^i_{on}}{q^i_{off}}\right)$. The weighted signals are summed to produce the total input current, which is then scaled and injected as somatic input during in vitro patch-clamp recordings. The neuron's membrane potential and spike times are recorded in response to this current (Fig. 3.1).

Information transfer is quantified in several steps. First, the entropy of the hidden state (H_{xx}) is calculated. Next, the mutual information between the hidden state and the input current (MI_I), and between the hidden state and the neuron's spike times ($MI_{\text{spike times}}$), are computed. The fraction of information (FI) transferred by the neuron is then defined as:

$$FI = \frac{MI_{\text{spike times}}}{MI_I} \tag{3.2}$$

Since the entropy of the hidden state is always greater than or equal to the mutual information, FI values range between 0 and 1. This approach assumes ergodicity (so time averages equal ensemble averages) and that spike trains are approximately Poissonian, though minor deviations from Poisson statistics do not significantly affect the results. By using a binary hidden state, this protocol allows for reliable estimation of mutual information even from relatively short recordings, providing a robust measure

of how effectively a neuron's spikes encode information about dynamic stimuli.

Quantification and statistical analysis

We used the non-parametric Wilcoxon test to test the significance of the observed difference in information transfer (FI) between paired aCSF and drug trials . We also used one sided Student's T-test to test the significance of the observed change between aCSF and agonist condition for passive biophysical and action potential attributes. Significance value was set to $p < 0.05$ in both cases. We performed a Kruskal-Wallis H test and a post-hoc Mann-Whitney U test with Bonferroni correction for multiple comparisons for comparing the cosine similarity between aCSF and agonist trials. All statistical tests were performed using scipy-stats package Virtanen et al. (2020).

Multi-set Correlation and Factor Analysis

To investigate how neuromodulation alters the relationship between different functional attribute sets (AP, PB, AC, and STA), we used **Multi-set Correlation and Factor Analysis (MCFA)** Brown et al. (2023), an unsupervised integration method designed to model both shared and private variance across multiple high-dimensional data types from the same samples. This method combines elements of canonical correlation analysis and factor analysis to produce low-dimensional representations of common and dataset-specific structure.

Each attribute set was z-scored (mean-centered and variance-scaled), and MCFA was applied separately to aCSF and agonist conditions. For the aCSF condition, we fixed the number of principal components to 2 for all attribute sets, due to the relatively low dimensionality of AP and PB features (e.g., PB: 6 dimensions). Private latent dimensionality k_m was also set to 1 for each set, based on model stability and convergence tests. Attempts to use higher latent dimensions resulted in non-convergent fits. These same dimensionality parameters (PCs $= 2$, $k_m = 1$) were applied to agonist conditions to maintain consistency and due to smaller sample sizes in drug trials.

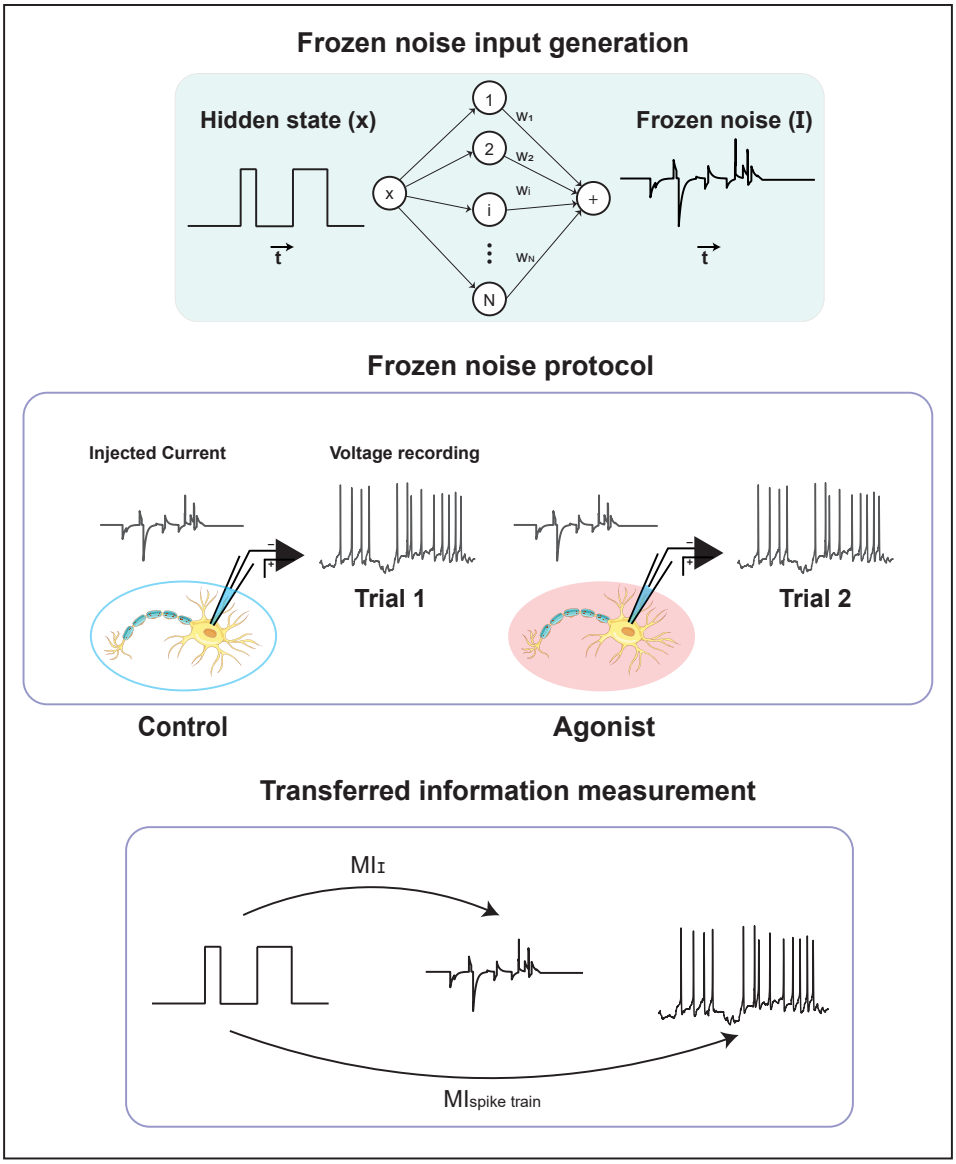

Figure 3.1. **Frozen noise protocol**: The top figure provides an illustration of how input is generated, the middle figure illustrates how this input is injected into the soma of a neuron for both aCSF and agonist conditions, the bottom figure illustrates the schema for calculating mutual information.

Since the number of aCSF recordings exceeded the number of agonist trials, we bootstrapped the control condition by randomly sampling subsets matched in size to the agonist group with the least number of recording sets, i.e., M1-R. This process was repeated 100 times, and MCFA was run on each subset. The resulting shared and private variances were averaged across runs to produce stable estimates for the control condition.

The MCFA fitting procedure used an expectation-maximization (EM) algorithm as described in Brown et al. (2023), initialized according to Joshi et al. (2024). Convergence was assessed by monitoring the log-likelihood and change in loading matrices across iterations.

Results

Neuronal identity is shaped by both intrinsic properties, such as gene expression and ion channel makeup Schultz (2007); Fishell and Heintz (2013), and external influences like synaptic input Joshi et al. (2024) and neuromodulatory signals. While previous work has studied excitability effects, it is still unclear how specific neuromodulators reconfigure a neuron's functional role across different physiological domains.

This study investigates how selectively activating dopaminergic (D1-R, D2-R) and cholinergic (M1-R) receptors modifies the input–output properties of excitatory and inhibitory neurons. We used the single unit somatic in-vitro recording dataset (Yan et al. (2022)), in which the neurons were recorded using the Frozen Noise (FN) protocol Zeldenrust et al. (2017) (see Methods and Fig. **3.1**) to capture responses under biologically realistic input. The dataset includes recordings from 312 individual neurons, each recorded under control (aCSF) and again after bath application of a specific neuromodulatory agonist (e.g., D1-R, D2-R, M1-R, etc.). Some neurons were recorded multiple times, by rinsing and repeating the aCSF, followed by agonist application. We refer to each such pair of recordings from a single neuron, including both aCSF

and agonist conditions as a recording pair and each individual recording as trial. After excluding low-quality trials, 296 such recording sets were retained for analysis (see Methods). For this study, we only focus of D1-R, D2-R and M1-R agonist recording sets.

We extracted a large set of intrinsic and input dependent attributes and grouped them into four functional attribute sets from each recording namely, action potential (AP) attributes, passive biophysical (PB) attributes, adaptation currents (AC) using a GLIF model, and spike-triggered averages (STAs) (see Methods) and analyzed how these changed under control (aCSF) versus agonist conditions. Neurons were grouped as excitatory or inhibitory following the same labeling protocol as Joshi et al. (2024) using waveform shapes and firing rates. We then analyzed how neuromodulation affected these attribute sets individually and in combination, assessing changes in encoding, intrinsic properties, input filtering, and their inter-dependencies.

Functional Attribute sets	**Description**
Action Potential (AP)	An ensemble of descriptive statistics of action potential shapes and dynamics.
Passive Biophysical (PB)	Attributes related to the non-active properties of cells, such as membrane capacitance and resistance.
Adaptation Current (AC)	Refers to the ionic currents in neurons that change in response to prolonged stimuli. Extracted via fitting a GLIF model.
Spike Triggered Average (STA)	The average of signal features occurring before neuron spikes, used to understand stimulus-response relations. It is an approximation of the linear input filter of a neuron.

Table 3.4. Functional Attributes sets and their Descriptions

Neuromodulation alters information transfer in a cell-type and agonist-specific manner

To determine how neuromodulators influence the encoding capabilities of individual neurons, we analyzed changes in fractional information (FI) (see Methods and Fig. 3.1)

and firing rate (FR) across agonist and control conditions using two complementary approaches.

First, we quantified the per-cell change in FI and FR between drug and control trials:

$$\Delta\mathrm{FI} = \frac{\mathrm{FI}_{\mathrm{agonist}} - \mathrm{FI}_{\mathrm{aCSF}}}{\mathrm{FI}_{\mathrm{aCSF}}}, \quad \Delta\mathrm{FR} = \frac{\mathrm{FR}_{\mathrm{agonist}} - \mathrm{FR}_{\mathrm{aCSF}}}{\mathrm{FR}_{\mathrm{aCSF}}}$$

We compared these values between excitatory and inhibitory neurons to assess whether receptor-specific neuromodulation affects cell types differently. To account for intrinsic trial-to-trial variability, we also computed ΔFI and ΔFR across two control (aCSF) trials from the same recording sets.

$$\Delta\mathrm{FI} = \frac{\mathrm{FI}_{\mathrm{aCSF2}} - \mathrm{FI}_{\mathrm{aCSF1}}}{\mathrm{FI}_{\mathrm{aCSF1}}}, \quad \Delta\mathrm{FR} = \frac{\mathrm{FR}_{\mathrm{aCSF2}} - \mathrm{FR}_{\mathrm{aCSF1}}}{\mathrm{FR}_{\mathrm{aCSF1}}}$$

In a second, complementary analysis, we compared FI distributions between control and drug conditions separately for excitatory and inhibitory neurons using a Wilcoxon rank-sum tests. This allowed us to ask not just whether individual neurons changed, but whether entire populations shifted their encoding capacities in response to neuromodulation.

Cell type specific changes in information transfer and firing rate relative to baseline variability

We looked at the effects of neuromodulation in a cell by cell manner. We compared the mean of change in fractional information ΔFI and firing rate ΔFR between control and agonist conditions. Next, we looked at the variability in the change of fractional information ΔFI and the firing rate ΔFR between aCSF and agonist conditions. We first examined the variability in information transfer and firing rate between two control trials (aCSF trial 1 vs. aCSF trial 2). Excitatory neurons showed a significant baseline

shift for both, ΔFI ($t = -2.57$, $p = 0.0118$, $d = -0.60$) and ΔFR ($t = -3.94$, $p = 0.0017$, $d = -0.92$). Inhibitory neurons showed no significant change in ΔFI ($t = 1.29$, $p = 0.2054$, $d = 0.32$), and only a modest change in ΔFR ($t = 2.28$, $p = 0.0244$, $d = 0.56$, see Fig. 3.2A). These values define the baseline against which agonist-driven effects are interpreted.

Following D1 receptor activation, excitatory neurons showed a greater increase in the variability of change between aCSF and agonist condition compared to inhibitory neurons (see Fig. **3.2C**). Compared to inhibitory neurons, excitatory neurons significantly decrease fractional information ΔFI ($t = -3.17$, $p = 2.7 \times 10^{-3}$, $d = -0.94$) and firing rate ΔFR ($t = -3.94$, $p = 1.7 \times 10^{-3}$, $d = -0.92$, see Fig. **3.2C**) on average. These effects were substantially larger than those observed in the aCSF–aCSF baseline (Fig. **3.2A** and **C**), indicating a strong and specific effect of D1-R activation on excitatory encoding and spiking.

In contrast, D2 receptor activation showed more modest variance in changes. While fractional information was significantly lowered in excitatory neurons as a result of D2-R activation ΔFR ($t = -2.49$, $p = 0.0157$, $d = -0.71$) than inhibitory neurons, fractional information ΔFI did not significantly differ between cell types ($t = -1.54$, $p = 0.129$, $d = -0.44$), and its magnitude was comparable to the control variability (see Fig. **3.2E**).

M1 receptor activation led to the strongest modulation of all in terms of the average value of change but not in terms of variability of change for fractional information and firing rate. Excitatory neurons showed a significant decrease in both fractional information ΔFI ($t = -2.95$, $p = 0.0054$, $d = -0.96$) and firing rate ΔFR ($t = -3.20$, $p = 0.0027$, $d = -1.04$) compared to inhibitory neurons. These changes clearly exceeded baseline variability, supporting a robust, cell-type-specific effect of M1-R on information processing (see Fig. **3.2G**).

These results demonstrate that D1-R and M1-R activation significantly enhance

both information transfer and firing rate in excitatory neurons, while inhibitory neurons remain largely unaffected except for the firing rate. D2-R has weaker or more variable effects. These findings suggest that neuromodulatory influence on encoding is not uniform but shaped by both receptor type and cell identity.

Neuromodulation shifts FI distributions in a receptor and cell-type-specific manner

After looking at the single cell level, we now look at the population level: does neuromodulation change the distribution over neurons of the firing rate and fraction of transferred information? We first established a baseline variability for each cell type. As expected from the per-cell analysis, the distribution of fractional information FI in excitatory neurons was significantly altered across aCSF control trials ($Z = 340.0$, $p = 1.1 \times 10^{-3}$, $d = 0.32$), while inhibitory neurons showed no significant change ($Z = 170.0$, $p = 0.205$, $d = -0.32$) (see Fig. **3.2B**).

Under D1-R activation, the mean of FI distribution increased significantly in inhibitory neurons ($Z = 110.0$, $p = 0.0190$, $d = -0.24$), but not in excitatory neurons ($Z = 84.0$, $p = 0.10$, $d = 0.54$), despite a moderate effect size (see Fig. 3.2**B** and **D**).

D2-R activation led to a significant decrease in the mean of FI for excitatory neurons ($Z = 225.0$, $p = 0.0120$, $d = 0.31$), but the magnitude of this change was similar to baseline difference between two aCSF trials, making interpretation hard. Inhibitory FI remained stable between aCSF and agonist condition ($Z = 87.0$, $p = 0.768$, $d = -0.03$) (see Fig. **3.2B** and **F**).

In contrast, M1-R activation caused a pronounced and significant decrease in mean FI for excitatory neurons ($Z = 22.0$, $p = 0.0020$, $d = 0.48$), clearly exceeding baseline variability. Inhibitory neurons again showed no significant change ($Z = 106.0$, $p = 0.5235$, $d = -0.12$) (see Fig. **3.2B** and **H**).

Summary and interpretation

These results demonstrate that neuromodulators exert both cell-type and receptor-specific effects on both firing rate and information transfer. D1-R and M1-R activation robustly diminish encoding in excitatory neurons, with effects that exceed intrinsic fluctuations seen across control trials. Inhibitory neurons remain largely stable, except for a small but significant reduction in FI under D1-R activation. D2-R produces more ambiguous effects, suggesting a more nuanced or weaker modulation. D2-R effect, although weak, is in the opposite direction as that of D1-R. Taken together, these findings indicate that neuromodulators do not uniformly shift excitability, but instead selectively reshape how neurons encode dynamic input based on cell identity and receptor types. A summary of statistical results is provided in Table. 6.1.

Specific receptor activation alters functional classification

Neuromodulators are known to affect the intrinsic properties that may or may not depend on the input. For instance D2-R activation modulates excitability in motor cortex Cousineau et al. (2020). Neuronal functional classification which is classically studied using action potential and passive biophysical features, has been shown to change as a function of input to the neuron Joshi et al. (2024); Hernáth et al. (2019); **?**. Also, we have shown in the previous section that neuromodulation (D1, D2, and M1 agonists) changes computation by changing amount of information transferred to a postsynaptic neuron in a cell-type specific manner. Therefore it is important to understand how functional attributes (passive, as well as input dependent) are altered by D1-R, D2-R and M1-R activation. Moreover, as functional attributes change, their assignment into different classes might also chance. Therefore, we look at how neural clustering based on functional attributes change as a result of neuromodulation.

We first wanted to check if there is a drift present in the data as result of experimental setup while performing multiple trial (see Frozen Noise protocol section in Methods), for this we extracted the 4 functional attribute sets (AP, PB, AC and STA, see Table. **3.4**) along with their waveforms from experiments with multiple aCSF trials and compared clustering the cells into subclasses (see Louvain+UMAP section in Methods) based on the acsf trial 1 and on acsf trial 2. The histogram in Fig. 6.8 show that there is a high level of correspondence between clustering based on aCSF trial 1 and trial 2 for waveform and passive biophysical features. The low level of correspondence between the AP, AC, and STA results from inherent trial to trial variability present in the cell. We also show the manifold overlap between clustering based on trial 1 and trial 2 for all properties (see Fig. **6.8**).

After establishing the baseline for clustering similarity, we clustered D1-R, D2-R and M1-R agonist trials as well as their corresponding aCSF trials separately using

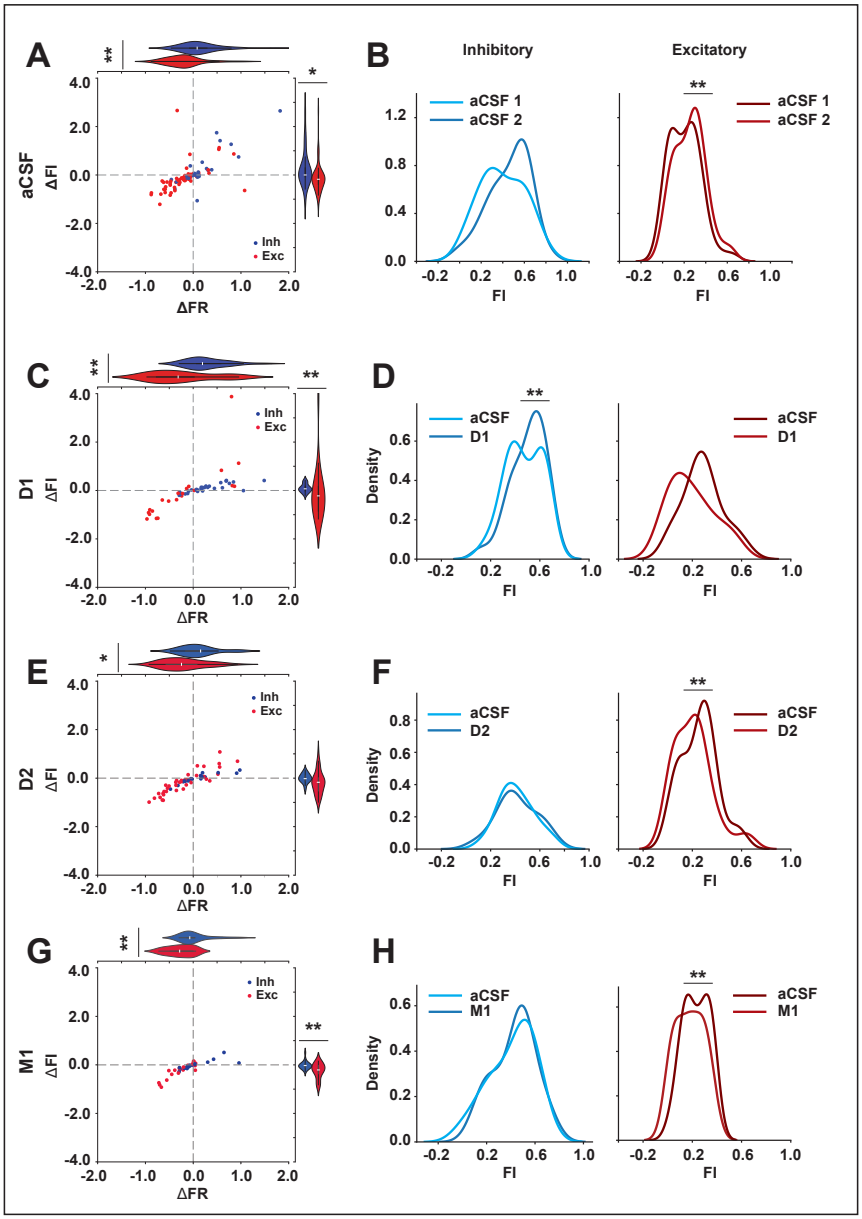

Figure 3.2. **Neuromodulation changes the amount of information transferred about the stimulus in a cell specific manner.**(**A**) The scatter plot shows the relationship between change in firing rate and information transfer between two aCSF trials of the same neuron. It can be seen that majority of Excitatory cells have decreases firing rate and decreased FI values as a result of consecutive recordings. In contrast, inhibitory cells shows a much more diverse behavior. (**B**) KDE plot showing the distribution of measured FI for aCSF trial 1 and aCSF trial 2 conditions respectively. (**C**) Scatter plot (left) showing the normalized change in firing rate vs fractional information as a result of D1 receptor activation.

(**D**) KDE plot showing the distribution of measured FI for aCSF and D1 conditions respectively. (**E**) Scatter plot (left) showing the normalized change in firing rate vs fractional information as a result of D2 receptor activation. (**F**) KDE plot showing the distribution of measured FI for aCSF and D2 conditions respectively. (**G**) Scatter plot (left) showing the normalized change in firing rate vs fractional information as a result of M1 receptor activation. (**H**) KDE plot showing the distribution of measured FI for aCSF and M1 conditions respectively. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

UMAP+Louvain clustering (see Louvain+UMAP section in Methods as well as Joshi et al. (2024)) and measured the similarity between cluster labels for aCSF and agonist trials using adjusted mutual information score (AMI) see Louvain+UMAP section in Methods. We summarized the reliability of clusters between aCSF and agonist trials for excitatory neurons in case of D1-R activation in histogram (Fig. **3.3 A.1**), it can be seen that AMI scores are consistently low for AP, PB, AC and STA, suggesting that functional attribute based clustering is altered as a result of D1-R activation, detailed co-classification matrices and average similarity within each cluster pair are provided in Fig. **6.7 A**.

We further explored how each attribute in the AP and PB attribute sets is altered as a result of D1-R activation, for passive biophysical properties (Fig. **3.3 A.2**) we found that conductance (gL) (One Sided t-test: $p < 0.001$) and reset voltage (Vr) (One Sided t-test: $p < 0.05$) are significantly reduced as a result of D1-R activation. We then explored how adaptation current and linear input filter (STA) are altered as a result of D1-R modulation shown in Fig. **3.3 A.3** and Fig. **3.3 A.4**, the red curves represent D1 trials and black curves represent aCSF trials. To quantify the differences between D1 and aCSF trials, we calculated the rise time and peak values (see Methods **??**) for both adaptation currents and STA. The joint plot Fig. 6.11 shows the rise time and peak differences between D1 and aCSF trials for adaptation current and Fig. 6.12 shows rise time and peak value for STA. The peak and rise time for adaptation current (AC paired t-test (decay time): $p = 0.4341$, AC paired t-test (peak): $p = 0.2444$) and STA (STA paired t-test (rise time): $p = 0.758$, STA paired t-test (peak): $p = 0.0514$)

were found to not be significantly different between aCSF and D1 trials for excitatory neurons. This suggest that STA and adaptation currents are not altered significantly as a result of D1-R activation.

We also performed a cosine similarity measurement within (between agonist-agonist or aCSF-aCSF) and across the aCSF and D1 trials for AC curves see Fig. 6.13, we performed a Kruskal–Wallis H test on three groups: within-aCSF, within-agonist, and across-aCSF and agonist conditions. The test revealed a significant effect of group based on similarity distributions ($H(2) = 15.3953,\ p = 4.53e{-}4$). Post hoc comparisons using Mann–Whitney U tests (Bonferroni-corrected for multiple comparisons) showed that:

- The mean cosine similarity score was significantly higher for within aCSF compared to across-pair (aCSF-D1) comparisons ($U = 120654.00,\ p = 3.17e-4$),
- Within D1 mean cosine similarity was also found to be significantly lower than across-pair (D1-aCSF) mean cosine similarity ($U = 153267.00,\ p = 0.0217$),
- We didn't observe a significant difference between the mean cosine similarity between within aCSF and D1 cosine similarity distributions ($U = 135299.00,\ p = 1.0$).

These findings indicate that neural representations are more consistent within conditions than across conditions, suggesting that D1 modulation significantly alters adaptation current consistently across the population.

Similar to adaptation current, we also performed a cosine similarity measurement within and across the aCSF and D1 trials for STA curves, we performed a Kruskal–Wallis H test on three groups: within-aCSF, within-agonist, and across-aCSF and agonist conditions. The test revealed a significant effect of group based on similarity distributions ($H(2) = 14.3199,\ p = 7.77e-4$).

Post hoc comparisons using Mann–Whitney U tests (Bonferroni-corrected for multiple comparisons) showed that:

- Similarity scores were significantly higher within aCSF compared to across-pair comparisons ($U = 142428.00,\ p = 7.44e-3$),
- D1 also showed significantly lower similarity than across-pair comparisons ($U = 134285.50,\ p = 1.95e-3$),
- We didn't observe a significant difference between aCSF and D1 mean cosine similarity distributions ($U = 142204.00,\ p = 1.0$).

These findings indicate that neural representations are more consistent within conditions than across conditions, suggesting that D1 modulation significantly alters the STA consistently across the population.

Finally, we assessed the effect of D1-R activation on action potential attributes (Fig. **3.3 A.5**) which incorporates spiking dynamics, spiking threshold and AP height and width attributes for the excitatory population. We found that maximum ISI (one sided t-test: $t = 3.148,\ p = 4.49e-3$), mean ISI (one sided t-test: $t = 2.283,\ p = 0.0319$), instantaneous firing Rate (one sided t-test: $t = -3.826,\ p = 8.64e-4$) from the spiking dynamics are significantly altered as a result of D1-R activation. We performed similar analysis for D2 and M1 agonist trials and summarized the results in Fig. 6.9 and Fig. 6.10. This was done to provide a template for understanding effect of the modulation on individual properties in a cell-type specific manner. Also, the D1-R are shown because they have the most number of samples.

For inhibitory neurons, the effect on clustering in case of D1-R activation is shown in histogram (Fig. **3.3 B.1**). It can be seen that AMI scores are consistently low for AP, PB, AC and STA, suggesting that functional attribute based clustering is altered as a result of D1-R activation, detailed co-classification matrices and average similarity within each cluster pair for each parameters is provided in Fig. **6.7 A**. We further

explored how each attribute set is altered as a result of D1-R activation. For passive biophysical properties (Fig. **3.3 B.2**), we found that the conductance (gL) and the reset voltage (Vr) are significantly reduced (One sided t-test: $p<0.001$, $p<0.05$) as a result of D1-R activation. We then explored how adaptation current and linear input filter (STA) are altered as a result of D1-R modulation shown in Fig. **3.3 B.3** and Fig. **3.3 B.4**. As for the excitatory population, we calculated the rise time and peak values (see **??**) for both adaptation currents and STA. The joint plot Fig. 6.11 shows the rise time and peak differences between D1 and aCSF trials for adaptation current and Fig. 6.12 shows rise times and peak values for STA. The peak (paired t-test: $t = -1.195,\ p = 0.244$) and decay times (paired t-test: $t = 0.7966,\ p = 0.4341$) for adaptation currents were not found to be significantly different. For STA, the rise time was found to be significantly different between aCSF and D1 trials (paired t-test: $t = 2.454,\ p = 0.0208$) but the peak current was not significantly different for inhibitory neurons (paired t-test: $t = 1.834,\ p = 0.077$). This suggest that STAs and adaptation currents are not altered significantly as a result of D1-R activation.

We also performed a cosine similarity measurement within and across the aCSF and D1 trials for AC curves and a Kruskal–Wallis H test on three groups to check for significant grouping: within-aCSF, within-agonist, and across-aCSF and agonist conditions. The test revealed a no significant effect between aCSF and D1 trials on similarity distributions ($H(2) = 2.0098,\ p = 0.3660$). These findings indicate that the AC is not altered as a result of D1 modulation. Similarly, we performed a cosine similarity measurement within and across the aCSF and D1 trials for STA curves. The test revealed a significant effect of group based on similarity distributions ($H(2) = 3.2486,\ p = 0.1970$). These findings indicate that the AC is not altered as a result of D1 modulation. Finally, we assessed the effect of D1-R activation on action potential attributes (Fig. **3.3 B.5**) which incorporates spiking dynamics, spike threshold and AP height and width attributes for inhibitory population. We found that firing rate (one sided t-test: $t = 3.6275,\ p = 0.0011$) from the spiking dynamics subset is significantly altered, and also the mean threshold (one sided t-test: $t = 4.1205$, p $= 3.03e-4$), median threshold (one sided t-test: $t = 4.361,\ p = 1.58e-4$) and minimum threshold (one sided t-test: $t = 2.443,\ p = 0.021$) from the spiking threshold set were significantly altered as result of D1-R activation.

We wanted to further understand if there are sub groups of neurons that are altered differently in their action potential and passive biophysical attributes as a result of D1-R activation. For this, we clustered the change in AP and PB attributes between aCSF and D1 trials for both excitatory and inhibitory neurons and summarized our finding using polar plots with each set of attributes in Fig. **3.4**. Each curve represents a single neuron, colored with its respective cluster identity and the mean is represented with a thick line. It can be seen that there are 3 clusters of AP attributes for excitatory neurons and 4 clusters for inhibitory neurons. Similarly, we find 3 clusters each of PB attributes for both excitatory and inhibitory neurons. We also compared the overall similarity between clustering results based on aCSF trials and clustering the difference

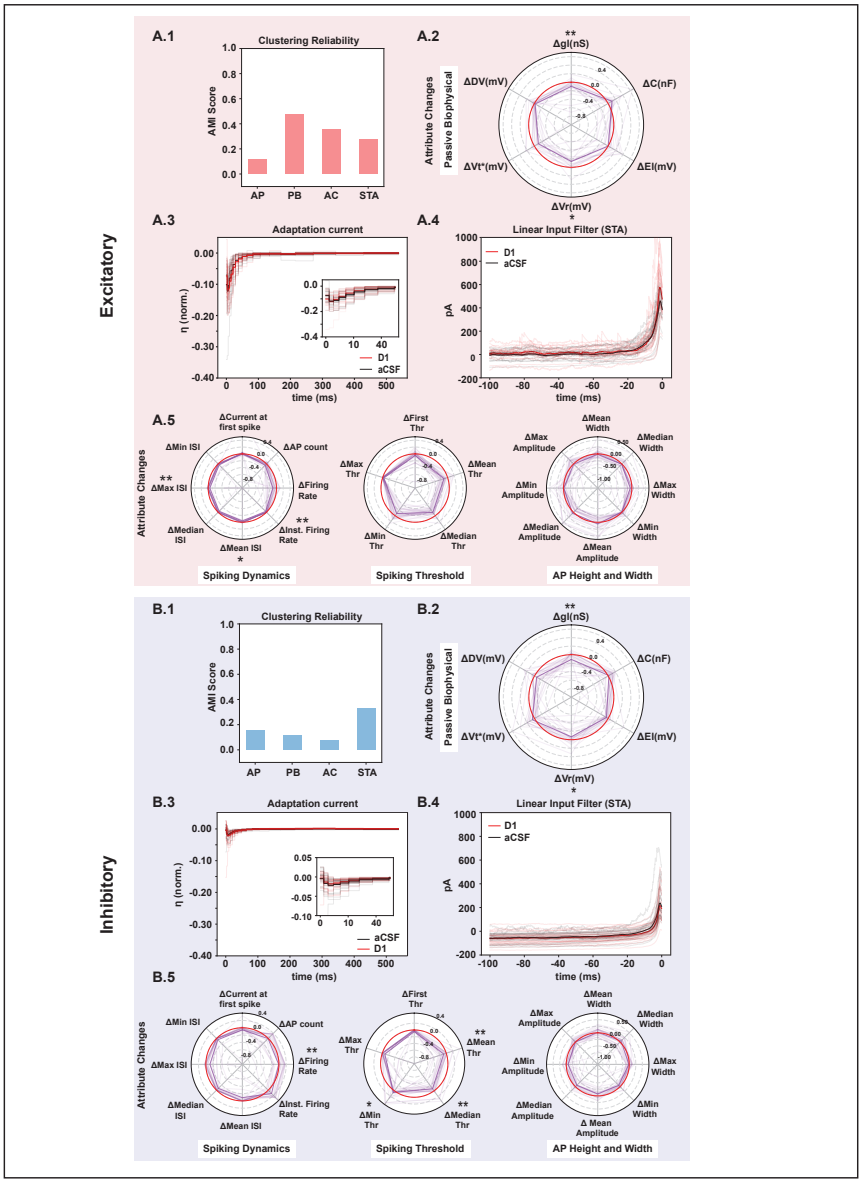

Figure 3.3. **D1-R activation changes functional clustering and attributes for both excitatory and inhibitory neurons.**(**A.1**) Histogram shows the adjusted mutual information score between the clustering labels obtained for aCSF and D1 trials, the histogram shows a shift in cluster identities as a result of D1 receptor activation across four attributes. (**A.2**) shows the change in passive parameters with respect to control as a result of D1 receptor activation, normalized by the control trial values. (**A.3**) shows the adaptation current for control (black) and D1 (red), the mean is represented with darker curves. (**A.4**) shows the spike triggered average profile for aCSF (black) and D1 (red) trials, the mean is represented with a thick line.

(**A.5**) shows the change in action potential attribute sets with respect to control as a result of D1 receptor activation, normalized by the control trial values. With the mean marked represented with a thick line and the zero line is coloured in red. (**B.1-5**) Same as **A.1-5** but for inhibitory neurons. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

between aCSF and D1 trials for the 4 attribute sets using the AMI score between cluster labels. We summarized our findings in histogram shown in Fig. 3.4. It can be seen that AMI score for both excitatory and inhibitory neurons are consistently low for all 4 attributes for both excitatory and inhibitory neurons. We performed a similar analysis for D2 and M1 agonist trials and summarized our findings in Fig. **6.14** and Fig. **6.15**.

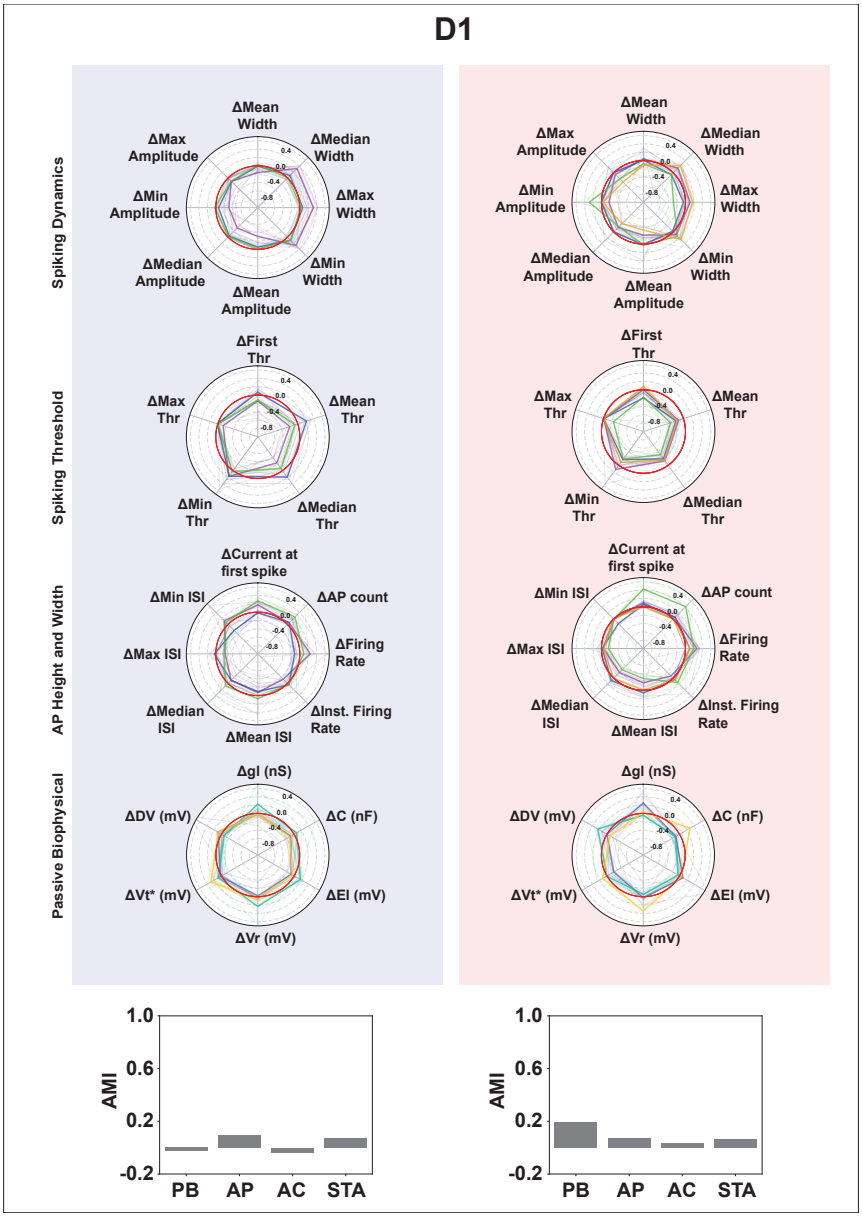

Figure 3.4. **Clustering based on differences between control and D1 trials for action potential and passive biophysical attributes reveals subgroups of neurons getting modulated differently as a result of D1-R activation.**

The polar plots show the clusters based on difference values between D1 and aCSF trials for action potential (subdivided into spiking dynamics, spiking threshold and AP height and width) and passive biophysical properties for both, excitatory (red background) and inhibitory (blue background). Each neuron is represented with a thin line and colored with their respective cluster label. The mean for each cluster is represented with a thick line. The histogram at the bottom shows the AMI score between cluster results using aCSF trials and the cluster results based on the difference between D1 and aCSF trials.

Neuromodulation changes structured correlation between functional attributes in a cell-type as well as receptor type specific manner

After showing that neuromodulation alters stimulus encoding in a cell-type specific manner and reorganizes functional subpopulations in cortical circuits, we further asked how neuromodulation (dys)regulates specific sets of neuronal properties (AP, PB, AC, and STA) which together span the high-dimensional correlated structure of functional properties neurons. While clustering revealed how neuromodulation reshapes functional subtypes, it did not address whether these changes reflect coordinated shifts in the underlying feature sets. Understanding how these attributes shift in response to receptor-specific activation provides a more mechanistic insight into how computational capacities are reconfigured. Specifically, we asked: are these properties modulated in isolation, or are they co-regulated in a coordinated fashion that might explain observed changes in encoding? To establish this, we applied multi-set correlation and factor analysis (MCFA) to our dataset, first on aCSF (control) trials and each agonist trial set respectively. We summarized the results of shared, private and residual variance using a stacked histogram (Fig. **3.5**). In this analysis, when individual feature sets exhibit a high shared variance, this means that the features are functionally coordinated: changes in one set tend to co-vary with changes in others. In contrast, when feature sets exhibit a high private variance, it suggests that the features are relatively independent i.e, un-

coordinated, meaning changes in one set do not systematically correspond to changes in the others, or a feature with a high private variance varies independently from the other features. A high residual signifies either noise or a complex non-linear relationship not captured by a linear method such as MCFA (Summarized in Table.3.5). To account

Variance type	Description
Shared variance	Variance correlated across all feature sets (indicating coordinated modulation)
Private variance	Variance unique to each set (indicating functional independence)
Residual variance	Unexplained variance (potential noise or non-linear effects)

Table 3.5. A brief summary of MCFA interpretation and meaning of different variances and residuals.

for the sample size imbalance, we performed a bootstrapped random sub-sampling of aCSF trials equal to the agonist set with minimum number of trials, and repeated MCFA a 100 times using random subsets. The reported aCSF values reflect the mean shared/private/residual variance across bootstraps. Standard deviations (or confidence intervals) are included in Tables. 6.2 & 6.3 to reflect the variability in control estimates. We present our findings using a stacked histogram for each condition in a cell type specific manner.

D1-R agonist

First we looked at the effect of D1-agonist. It can be seen in Fig. 3.5 that the private variance for excitatory population decreases for AP, PB and AC sets except for STA, for which it sharply increases (aCSF: 16.7742% → D1: 73.9663%) compared to aCSF trials (see Table.6.2 and Figure 3.5 **A**-**B**) upon D1 receptor activation. The shared variance reduces sharply compared to aCSF for AC (aCSF: 15.8369% → D1: 6.9056%) and STA (aCSF: 23.6072% → D1: 9.5260%) sets for the excitatory population. This suggests that D1-R activation makes the STA more independent compared to the other

attributes. The shared variance for AP an PB has slightly increased, suggesting that PB and AP attributes become more coordinated. We observe a dramatically different effect in the inhibitory population compared to the excitatory counterpart. The shared variance increases sharply for the STA set (aCSF: 23.6595% → D1: 46.2345%) and also for the AC set (aCSF: 26.9763% → D1: 44.4749%) (see Figure. 3.5**A**-**B** and Table. 6.3), while the private variance decreases for these attributes, suggesting that the AC and STA become functionally coordinated with the latent space, therefore with AP and PB properties. It is also worth noticing that the private variance decreases sharply for the AP (aCSF: 14.9272% → D1: 0.8297%) and PB (aCSF: 13.1228% → D1: 3.9880%). Similarly, the shared variances decrease for both the AP and PB (see Figure. 3.5**A**-**B** and Table. 6.3) suggesting the interaction between these two sets show a weakened linear correlation with other feature sets.

In summary, the effect of D1-R is cell-type specific and nuanced, increasing the coordination between STA and AC properties in inhibitory neurons, while decreasing coordination for excitatory STA with the rest of the attribute sets.

D2-R agonist

In order to study the effect of D2-R activation, we analyzed D2-R trials using MCFA in a similar manner to the D1-R activation and present the results in comparison with aCSF trials (see Fig. 3.5**A**-**C** and Table. 6.2).

We observe that D2-R modulates the coordination between functional attributes in a more subtle manner than D1-R for excitatory neurons: the private variance for STA increases with D2-R activation (aCSF: 57.7380% → D2: 71.2200%), and at the same time the shared variance between AC and AP attributes increases (see Fig. 3.5**A**-**C** and Table. 6.2). Private variance on the other hand decreases drastically for AP, AC and PB attributes as also observed in case of D1. The stark increase in private variance in case of STA needs to be highlighted. This suggests that D2-R activation makes

AP and AC properties more coordinated, while making STA and PB more functionally decoupled from other attributes.

In the case of the inhibitory neurons, the modulation is rather straight-forward. The private variance for all attributes except for the AC decreases and the shared variance increases for all the attributes except for AP attributes. This suggests that D2-R overall makes the functional properties more coupled for inhibitory population.

In summary, the effect of D2-R activation is cell-type specific as in the case of D1-R case. Increasing the coordination between AP and AC properties in the excitatory population while making the STA independent and making the PB, STA and AC sets coordinated for the inhibitory population.

Muscarinic M1-R agonist

In the case of M1 receptor activation, we observe that the shared variance increases for AP (aCSF: 30.8543% → M1: 66.0431%) an AC (aCSF: 15.8369% → M1: 35.0802%) properties similar to D2 for excitatory neurons and decreases for PB and STA (see Fig. 3.5**A**-**D** and Table. 6.2). This suggests that AC and AP attributes become coordinated and PB and STA properties become more independent as a result of M1-R activation. On the other hand, the private variance decreases for all the attributes except for PB. Another important thing to notice is that the residual for STA increases sharply, we suspect that this is due to a decrease in firing rate as a result of M1-R activation, that increases the noise in STA measurement.

For inhibitory neurons, the shared variance increases for all sets except for AC, suggesting a stronger functional coordination (see Fig. 3.5**A**-**D** and Table. 6.3). Surprisingly, the private variance increases sharply for the AC, suggesting a decoupling with the rest of the attributes set.

In summary, the effect of Muscarinic M1-R activation is cell-type specific as well but different from D1-R and D2-R, increasing coordination between AP and AC properties

Receptor	Cell-type	Effect
D1-R	Excitatory	Increases coordination between AP and PB and makes STA independent.
	Inhibitory	Increases coordination between AC and STA and reduces linearity. between AP and PB sets.
D2-R	Excitatory	Increases coordination between AP and AC and makes STA independent.
	Inhibitory	Makes PB, AC and STA coordination stronger.
M1-R	Excitatory	Makes AP and AC coordination stronger while making PB more independent.
	Inhibitory	Makes coordination between AP, PB and STA stronger while making AC independent.

Table 3.6. A summary of findings from MCFA for each receptor type.

and making PB set independent in excitatory population and making STA independent while making AP, STA and PB sets coordinated (increasing independence for AC) and for inhibitory population.

Summary

In summary, our MCFA analysis reveals that neuromodulation alters the coordination structure between key functional attributes – action potential dynamics, passive biophysics, adaptation currents, and input filtering in a receptor and cell-type-specific manner. In excitatory neurons, D1-R and D2-R activation selectively increase the independence of input feature selectivity (STA), while promoting a greater coupling between output-related features such as action potentials and adaptation currents. In contrast, inhibitory neurons exhibited a consistent increase in coordination across all functional domains under D1-R, D2-R, and M1-R activation, with a marked reduction in private variance, suggesting a shift towards more integrated and cohesive functional states. Notably, M1-R activation produced a unique pattern: excitatory neurons exhibited strengthened AP–AC coupling alongside increased independence in passive and

input features, while inhibitory neurons showed a widespread coordination (except for adaptation currents, which became more decoupled). Overall, the ratio of shared to private variance increased across all receptor conditions relative to control, indicating that neuromodulators generally enhance the global coordination of neuronal features. These findings highlight the diverse strategies by which neuromodulators reshape the functional architecture of neurons—either by reinforcing coordinated, stable encoding in inhibitory populations or by promoting selective flexibility and specialization in excitatory neurons. A complete summary is shown in Table. 3.6.

Discussion

In this study, we aimed to investigate how neuromodulation, specifically through dopamine (D1R, D2R) and acetylcholine (M1R) receptor activation, alters the functional properties of cortical neurons beyond traditional measures of excitability. Using frozen noise stimulation based single-cell in-vitro somatic recordings from layer 2/3 of the mouse somatosensory cortex, we extracted four functional feature sets comprising action potential (AP), passive biophysical (PB), adaptation currents (AC), and linear input filter via a spike-triggered average (STA). This approach enabled us to study the impact of neuromodulation across multiple physiological domains within and across neuronal subtypes.

Receptor-Specific Neuromodulation Reshapes Functional Architecture in Distinct Cell Types

Our analyses revealed that dopaminergic and cholinergic receptor activation altered the correlation structure among functional attributes, and these effects were excitatory/inhibitory cell-type specific. For inhibitory neuron**s**, D1R and D2R activation increased inter-attribute correlations, suggesting a convergence of functional attributes under neuromodulatory influence. In contrast, excitatory neurons displayed decreased correlations under D1R activation, indicating a decoupling of intrinsic and encoding features. This suggests that dopaminergic modulation sharpens functional coherence in inhibitory neurons while increasing functional independence in excitatory neurons.

These results extend previous findings that D1R activation increases firing in both excitatory and inhibitory neurons in prefrontal and motor cortices Seamans and Yang (2004); Tritsch and Sabatini (2012); Anastasiades et al. (2019), by showing that in the sensory cortex, such modulation also reorganizes functional coupling between key electrophysiological domains. Inhibitory neurons, under D1R and D2R activation, exhibited

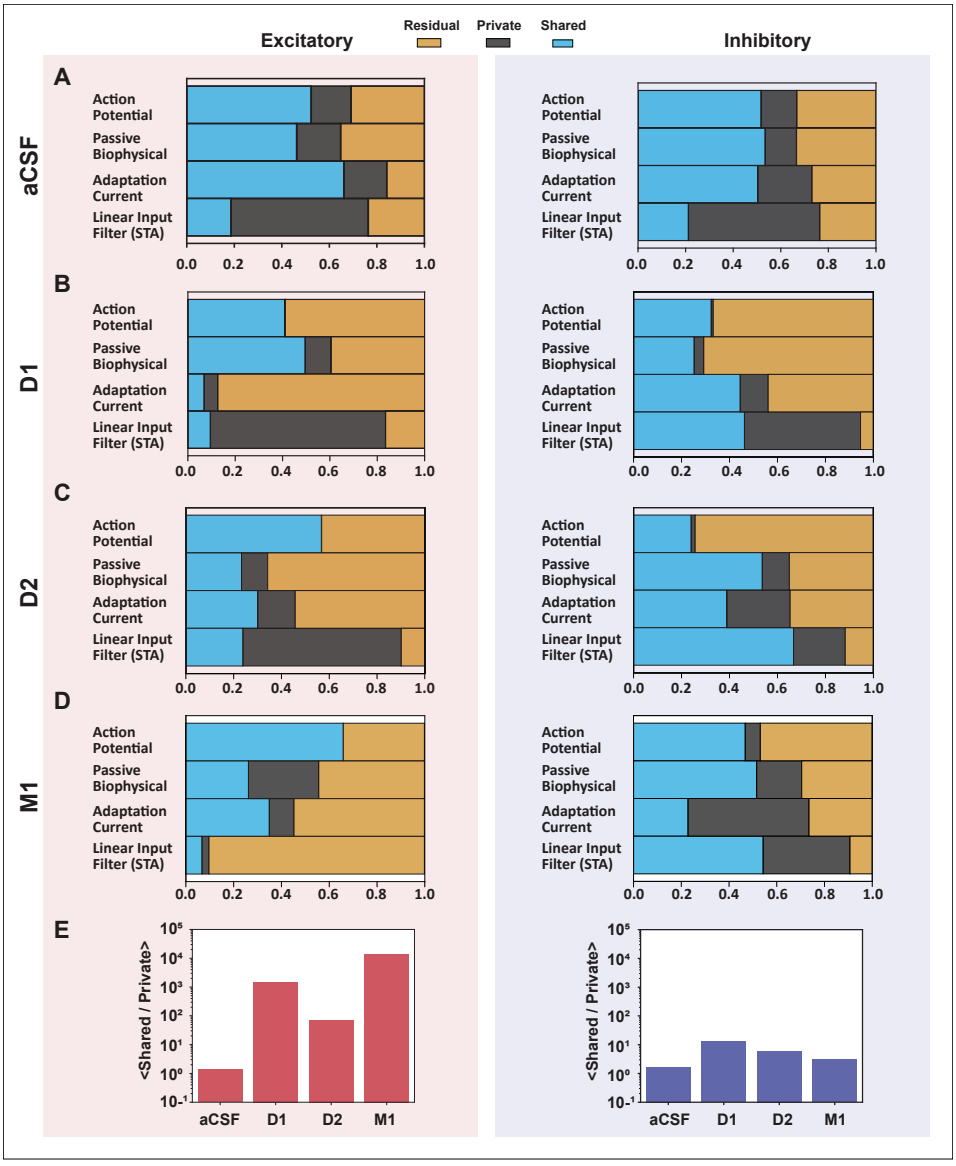

Figure 3.5. **Correlation between functional attributes changes as a result of D1, D2 and M1 receptor activation compared to control in both excitatory and inhibitory neurons.** **A** Left stacked bar-plot shows the MCFA results for aCSF-excitatory trials and on the right is the aCSF-inhibitory trials. **B** Left stacked bar-plot shows the MCFA results of D1-excitatory trials and the corresponding inhibitory trials is on the right. **C-D** show the excitatory and inhibitory MCFA results for D2 and M1 agonist trials respectively. Excitatory trials are on the left and inhibitory on the right. **E** Left histogram shows the ratio of shared and private variance for control and each agonist conditions in the excitatory population. Similarly right histogram shows the ratio of shared over private variance for inhibitory population for each agonist condition.

stronger coupling between passive properties, AP dynamics, and linear input filtering; suggesting that neuromodulation may reduce heterogeneity and impose a more unified computational role. This aligns with the hypothesis that dopamine can increase synchronization and gain control within inhibitory networks, potentially sharpening their influence on local circuits Gao and Goldman-Rakic (2003); Morozova et al. (2016); Seamans et al. (2001).

Interestingly, we observed similar coupling effects under M1R activation, pointing to a convergent mechanism across dopaminergic and cholinergic systems in shaping the internal structure of inhibitory neuron function. The decoupling observed in excitatory neurons, particularly under D1R, may reflect a shift toward greater computational flexibility or specialization, allowing for more diverse input-output transformations. These effects may be a substrate for dynamic control of cortical processing modes, such as switching between attentive and exploratory states.

Functional Clustering and Heterogeneous Modulation of Neuronal Identity

To examine how these neuromodulatory changes reorganize functional neuron types, we applied UMAP-Louvain clustering to the high-dimensional feature sets before and after receptor activation. Our results show that agonist application drastically alters functional clustering, suggesting that receptor-specific modulation reshapes neuronal identities in both excitatory and inhibitory populations.

Importantly, the modulatory effects were not uniform, but heterogeneous within cell types, implying that neuromodulation acts in a subtype-specific manner. This is consistent with recent work showing volume transmission and receptor expression gradients as mechanisms for differential modulation across cell types Özçete et al. (2024). By clustering neurons based on their change vectors (difference between agonist and control), we identified distinct subpopulations exhibiting coordinated shifts in AP and passive properties, reinforcing the idea that neuromodulation reconfigures the func-

tional state space of neurons rather than simply scaling their excitability.

We also found that adaptation currents and STAs became more homogeneous in excitatory neurons compared to inhibitory neurons following neuromodulation. This suggests that although both populations undergo modulation, excitatory neurons may be pushed toward a more constrained encoding regime, possibly to provide stability in coding under varying network conditions. Nonetheless, the core distinction between excitatory and inhibitory populations remained intact, highlighting the robustness of intrinsic identity despite substantial functional plasticity. We speculate that a modeling effort superimposing neuromodulatory affects on a balanced networks would reveal the implication of neuromodulatory alteration on the balance excitation-inhibition and therefore the shift that neuromodulation can cause.

Neuromodulation Selectively Alters Information Transfer

While neuromodulators are known to affect neuronal excitability, less is known about how such changes impact information transmission. Using the frozen noise protocol Zeldenrust et al. (2017), we estimated fractional information (FI) between stimulus and spike train for each neuron under different receptor conditions.

Our key finding is that neuromodulation alters information transfer in a cell-type and receptor-specific manner. Specifically, D1R and M1R activation significantly decreased FI in excitatory neurons, while D2R activation increased FI in inhibitory neurons. This suggests that excitatory and inhibitory neurons are differentially engaged by neuromodulators to redistribute information processing roles. Moreover, FI variance increased in excitatory neurons under D1R, implying a diversification of encoding strategies, whereas inhibitory neurons exhibited more stable and coordinated changes.

These findings underscore an important result that modulation of biophysical and spike-generating properties has direct consequences on the information transfer and encoding of neurons. Neuromodulators do not simply shift the gain of neurons, they

reconfigure how inputs are integrated and transformed into outputs, thus shaping the flow of information through cortical networks. This raises exciting new questions about how neuromodulatory systems shape perceptual inference, attention, and learning by dynamically allocating information processing across cell types.

Limitations and Future Directions

Our study, while comprehensive, has several important limitations. First, our recordings were limited to the soma, whereas neuromodulation often targets synaptic and dendritic compartments, which could substantially influence neuronal input-output transformations. Second, our analysis is constrained by sample size and cortical region, necessitating broader recordings across layers and brain areas to validate the generality of our findings.

Most critically, neuromodulatory systems are highly degenerate: the same ion channel can be targeted by multiple modulators, and a single neuromodulator can affect many channels and cellular processes. Thus, our receptor-specific results represent only a partial view of the underlying modulatory landscape. Future work should incorporate multi-receptor interactions, consider temporal dynamics of modulation, and examine how these single-cell changes propagate to network-level phenomena such as synchrony, attractor stability, and behavioral output.

Conclusion

Together, our findings reveal that D1, D2, and M1 receptor activation differentially reorganizes the biophysical, adaptive, and encoding properties of excitatory and inhibitory neurons in layer 2/3 of the somatosensory cortex. Neuromodulation can couple or decouple domains of neuronal function, alter functional classification, and modulate information transmission in a subtype-specific manner. These effects likely support context-dependent reconfiguration of cortical computation and offer a new window into

how global modulatory signals dynamically orchestrate diverse local circuit functions.

References

Aggarwal, C. C., Hinneburg, A., and Keim, D. A. (2001). On the surprising behavior of distance metrics in high dimensional space. In *Database theory—ICDT 2001: 8th international conference London, UK, January 4–6, 2001 proceedings 8*, pages 420–434. Springer.

Anastasiades, P. G., Boada, C., and Carter, A. G. (2019). Cell-type-specific d1 dopamine receptor modulation of projection neurons and interneurons in the prefrontal cortex. *Cerebral cortex*, 29(7):3224–3242.

Arnsten, A. F. and Rubia, K. (2012). Neurobiological circuits regulating attention, cognitive control, motivation, and emotion: disruptions in neurodevelopmental psychiatric disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(4):356–367.

Bargmann, C. I. (2012). Beyond the connectome: how neuromodulators shape neural circuits. *Bioessays*, 34(6):458–465.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

Brown, B. C., Wang, C., Kasela, S., Aguet, F., Nachun, D. C., Taylor, K. D., Tracy, R. P., Durda, P., Liu, Y., Johnson, W. C., et al. (2023). Multiset correlation and factor analysis enables exploration of multi-omics data. *Cell Genomics*, 3(8).

Cousineau, J., Lescouzères, L., Taupignon, A., Delgado-Zabalza, L., Valjent, E., Baufreton, J., and Le Bon-Jégo, M. (2020). Dopamine d2-like receptors modulate intrinsic properties and synaptic transmission of parvalbumin interneurons in the mouse primary motor cortex. *eneuro*, 7(3).

da Silva Lantyer, A., Calcini, N., Bijlsma, A., Kole, K., Emmelkamp, M., Peeters, M., Scheenen, W. J., Zeldenrust, F., and Celikel, T. (2018). A databank for intracellular electrophysiological mapping of the adult somatosensory cortex. *GigaScience*, 7(12):giy147.

Dalley, J. W., Theobald, D. E., Bouger, P., Chudasama, Y., Cardinal, R. N., and Robbins, T. W. (2004). Cortical cholinergic function and deficits in visual attentional performance in rats following 192 igg–saporin-induced lesions of the medial prefrontal cortex. *Cerebral Cortex*, 14(8):922–932.

Dascal, N. (2001). Ion-channel regulation by g proteins. *Trends in Endocrinology & Metabolism*, 12(9):391–398.

Durstewitz, D. and Seamans, J. K. (2008). The dual-state theory of prefrontal cortex dopamine function with relevance to catechol-o-methyltransferase genotypes and schizophrenia. *Biological psychiatry*, 64(9):739–749.

Ferguson, K. A. and Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, 21(2):80–92.

Fishell, G. and Heintz, N. (2013). The neuron identity problem: form meets function. *Neuron*, 80(3):602–612.

Gao, W.-J. and Goldman-Rakic, P. S. (2003). Selective modulation of excitatory and inhibitory microcircuits by dopamine. *Proceedings of the national academy of sciences*, 100(5):2836–2841.

Hasselmo, M. E., Giocomo, L. M., et al. (2006). Cholinergic modulation of cortical function. *Journal of Molecular Neuroscience*, 30(1-2):133–136.

Häusser, M. (2000). The hodgkin-huxley theory of the action potential. *Nature neuroscience*, 3(11):1165–1165.

Hernáth, F., Schlett, K., and Szücs, A. (2019). Alternative classifications of neurons based on physiological properties and synaptic responses, a computational study. *Scientific reports*, 9(1):13096.

Joshi, N., van der Burg, S., Celikel, T., and Zeldenrust, F. (2024). Understanding neuronal diversity: Role of input dynamics and filtering. *bioRxiv*, pages 2024–10.

Lee, E. K., Balasubramanian, H., Tsolias, A., Anakwe, S. U., Medalla, M., Shenoy, K. V., and Chandrasekaran, C. (2021). Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. *Elife*, 10:e67490.

Marder, E. (2012). Neuromodulation of neuronal circuits: back to the future. *Neuron*, 76(1):1–11.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

Morozova, E. O., Zakharov, D., Gutkin, B. S., Lapish, C. C., and Kuznetsov, A. (2016). Dopamine neurons change the type of excitability in response to stimuli. *PLoS Computational Biology*, 12(12):e1005233.

Nadim, F. and Bucher, D. (2014). Neuromodulation of neurons and synapses. *Current opinion in neurobiology*, 29:48–56.

Nusser, Z. (2009). Variability in the subcellular distribution of ion channels increases neuronal diversity. *Trends in neurosciences*, 32(5):267–274.

Özçete, Ö. D., Banerjee, A., and Kaeser, P. S. (2024). Mechanisms of neuromodulatory volume transmission. *Molecular Psychiatry*, pages 1–14.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Pozzorini, C., Mensi, S., Hagens, O., Naud, R., Koch, C., and Gerstner, W. (2015). Automated high-throughput characterization of single neurons by means of simplified spiking models. *PLoS computational biology*, 11(6):e1004275.

Salinas, E. and Sejnowski, T. J. (2001). Book review: gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. *The Neuroscientist*, 7(5):430–440.

Sarter, M., Parikh, V., and Howe, W. M. (2009). Phasic acetylcholine release and the volume transmission hypothesis: time to move on. *Nature Reviews Neuroscience*, 10(5):383–390.

Schultz, W. (2007). Multiple dopamine functions at different time courses. *Annu. Rev. Neurosci.*, 30(1):259–288.

Schwartz, O., Pillow, J. W., Rust, N. C., and Simoncelli, E. P. (2006). Spike-triggered neural characterization. *Journal of vision*, 6(4):13–13.

Seamans, J. K., Durstewitz, D., Christie, B. R., Stevens, C. F., and Sejnowski, T. J. (2001). Dopamine d1/d5 receptor modulation of excitatory synaptic inputs to layer v prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, 98(1):301–306.

Seamans, J. K. and Yang, C. R. (2004). The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Progress in neurobiology*, 74(1):1–58.

Seong, H. J. and Carter, A. G. (2012). D1 receptor modulation of action potential firing in a subpopulation of layer 5 pyramidal neurons in the prefrontal cortex. *Journal of Neuroscience*, 32(31):10516–10521.

Shine, J. M., Müller, E. J., Munn, B., Cabral, J., Moran, R. J., and Breakspear, M. (2021). Computational models link cellular mechanisms of neuromodulation to large-scale neural dynamics. *Nature neuroscience*, 24(6):765–776.

Taghert, P. H. and Nitabach, M. N. (2012). Peptide neuromodulation in invertebrate model systems. *Neuron*, 76(1):82–97.

Tritsch, N. X. and Sabatini, B. L. (2012). Dopaminergic modulation of synaptic transmission in cortex and striatum. *Neuron*, 76(1):33–50.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Winterer, G. and Weinberger, D. R. (2004). Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends in neurosciences*, 27(11):683–690.

Yan, X., Calcini, N., Safavi, P., Ak, A., Kole, K., Zeldenrust, F., and Celikel, T. (2022). A whole-cell recording database of neuromodulatory action in the adult neocortex. *bioRxiv*, pages 2022–01.

Zeldenrust, F., Calcini, N., Yan, X., Bijlsma, A., and Celikel, T. (2024). The tuning of tuning: How adaptation influences single cell information transfer. *PLOS Computational Biology*, 20(5):e1012043.

Zeldenrust, F., de Knecht, S., Wadman, W. J., Denève, S., and Gutkin, B. (2017). Estimating the information extracted by a single spiking neuron from a continuous input time series. *Frontiers in computational neuroscience*, 11:49.

Chapter 4

Heterogeneity in delay and timescales improves task Performance

Based on: Willemsen, Y.*, Ou, Y.*, Belzer, C., Arias Vásquez, A., Smidt, H., Beijers, R., and de Weerth, C. A longitudinal study of the gut microbiota during the first three years of life: Links with problem behavior and executive functions at preschool age. *Development and Psychopathology*, 2023, 1-17. *Authors contributed equally

Abstract

Biological neural systems exhibit significant heterogeneity in temporal dynamics, with neurons operating on a wide range of intrinsic timescales. In contrast, most artificial recurrent networks, including those in the reservoir computing (RC) framework, use homogeneous decay parameters across units. In this work, we investigate whether introducing decay heterogeneity improves task performance in Echo State Networks (ESNs) and Distance-based Delay Networks (DDNs). We evaluate four heterogeneity configurations:homogeneous network, homogeneous cluster, heterogeneous network, and heterogeneous cluster, across two benchmark tasks: NARMA-30 and Mackey-Glass. Networks are optimized using Covariance Matrix Adaptation Evolution Strategy (CMA-ES), and evaluated on task performance, we studied the dynamical stability (Lyapunov exponents), representational dimensionality, and linear memory capacity of these networks. Our results show that DDNs consistently outperform ESNs, with homogeneous DDNs achieving the highest task performance. Timescale heterogeneity enhances memory flexibility in DDNs without compromising stability, while ESNs remain relatively unaffected. These findings suggest that structured heterogeneity in intrinsic timescales significantly improves the computational capabilities of delay-based recurrent systems.

Introduction

Physical systems such as brains consist of heterogeneous neurons; not just in terms of morphology and molecular composition, but also in their temporal properties Koch and Laurent (1999); Habashy et al. (2024). In some cases, temporal differences among neuron types span several orders of magnitude London et al. (2010); Brunel and Wang (2003); Wang et al. (2020). This diversity serves multiple functional roles, including supporting motor control Cavanagh et al. (2020), preserving remote memories Runyan et al. (2017b), and enabling spatial memory formation McNaughton et al. (2006). In particular, variation in neuronal time constants and decay dynamics is thought to be critical for temporal processing and memory formation in the brain Hasson et al. (2008); Chu et al. (2020).

In contrast, artificial neural networks, especially recurrent neural networks and their variants like the reservoir computing (RC) framework, are typically designed with homogeneous units that share identical intrinsic parameters. RC is a powerful approach for modeling temporal sequences Lukoševičius and Jaeger (2009), using high-dimensional dynamical systems such as Echo State Networks (ESNs) Jaeger (2001b). These networks rely on a fixed, randomly connected recurrent reservoir whose states are linearly combined by a trainable readout layer. This setup simplifies training while maintaining strong performance on memory-intensive and nonlinear tasks. A key hyperparameter in RC systems is the leak or decay rate, which controls how much past information is taken into consideration in each unit, shaping the network's temporal memory Schrauwen et al. (2007).

Recently, an extension of ESNs called Distance-based Delay Networks (DDNs) Iacob et al. (2022) has been introduced. DDNs incorporate non-uniform delays between reservoir units, motivated by the fact that axonal delays are ubiquitous in the brain and that information in physical systems like the brain is constrained by spatial structure

and finite signal transmission speed Izhikevich (2006). Prior work has shown that these delays lead to a redistribution of memory across timescales and can improve task performance compared to standard ESNs Iacob et al. (2022); Soriano et al. (2014); **?**.

However, while the role of inter-unit delays has been explored, it remains unclear whether heterogeneity in intrinsic parameters, specifically, decay rates—can further enhance performance in both ESNs and DDNs. Most prior studies assume a single, homogeneous decay parameter across all units, effectively enforcing a single timescale on the reservoir. This may limit the system's ability to capture rich, multi-timescale dynamics, especially in tasks that demand both, short and long-term memory. In contrast, biological neural systems inherently operate with a range of time constants London et al. (2010); Nam et al. (2017); Zhou et al. (2023), suggesting that incorporating decay heterogeneity could yield computational benefits.

To investigate this, we design and analyze two sets of reservoir networks, ESNs and DDNs, each implemented in four different configurations. These include a homogeneous setup where all units share a fixed decay value, a fixed-cluster configuration where the reservoir is divided into modules, each with its own fixed decay, a network-heterogeneous configuration where each unit's decay is sampled independently from a common distribution; and a cluster-heterogeneous configuration where different modules sample their decays from different distributions. We evaluate all configurations on two standard benchmark tasks: the NARMA-30 task, which tests nonlinear memory capacity, and the Mackey-Glass task, which involves chaotic time-series prediction. Performance is optimized using Covariance Matrix Adaptation Evolution Strategy (CMA-ES), with task evaluation based on normalized root mean square error (NRMSE) for NARMA and prediction horizon for Mackey-Glass.

Beyond performance, we analyze the networks' internal dynamics to understand the mechanisms underlying any observed differences. We compute Lyapunov exponents to assess dynamical stability and chaotic behavior, measure representational dimen-

sionality using SVD-based metrics, and evaluate linear memory capacity. Our results show that DDNs consistently outperform ESNs across all heterogeneity types, with fixed-cluster DDNs performing best. These findings suggest that structured temporal heterogeneity, particularly when organized in modular clusters, significantly enhances the computational capabilities of reservoir networks for tasks involving memory and temporal dynamics. But the heterogniety requirement is task dependent.

Results

The aim of this study was to study the effect of intrinsic heterogeneity on task performance of ESNs and DDNs. The network design and dynamics are explained in detail in Methods section 5.1. For the stated aim we designed 4 heterogeneity/homogeneity configurations explained as follows:

1. **Homogeneous Network**: All units share a single decay parameter.
2. **Homogeneous Cluster**: The reservoir is divided into clusters, each with its own fixed decay parameter.
3. **Heterogeneous Network**: Each unit samples its decay from a shared distribution.
4. **Heterogeneous Cluster**: Each cluster samples decay parameters from different distributions.

These configurations are tested on two benchmark tasks namely NARMA-30 and Mackey-Glass tasks (see Methods section 5.2):

- **NARMA-30**: A nonlinear auto-regressive moving average task designed to test long-range memory and nonlinear dynamics. Parameters for NARMA-30 task is given in Methods section 5.2.1.
- **Mackey-Glass**: A chaotic time-series prediction task that evaluates a network's ability to generate stable yet complex temporal outputs. Parameters for Mackey-Glass task is given in Methods section 5.2.2.

A key reason to optimize network for these two tasks is to understand if inherent heterogeneity is utilized based on task performance. We optimized all networks using

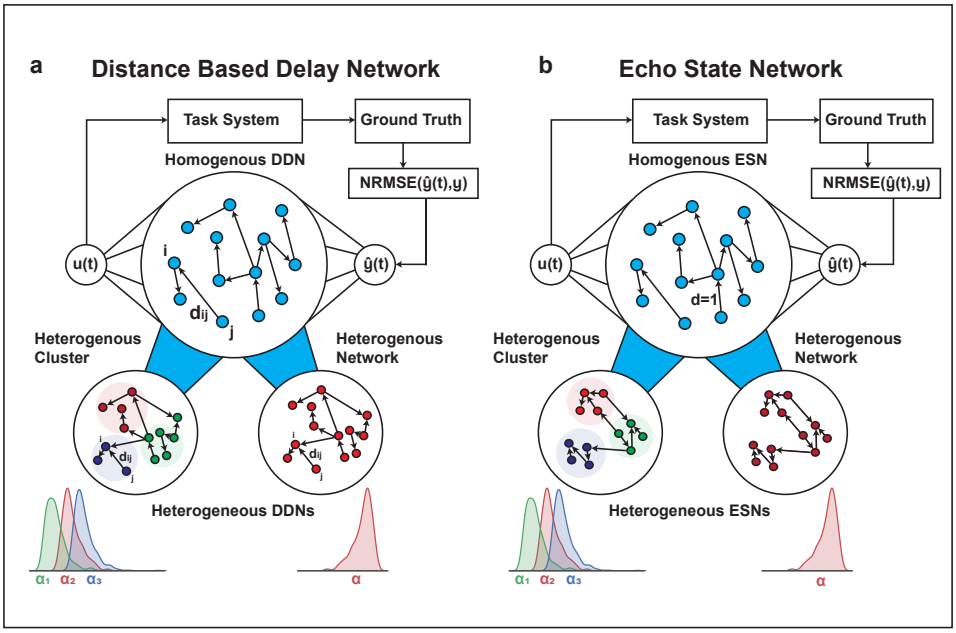

Figure 4.1. **Reservoir architecture and heterogeneity design in DDNs and ESNs: (a)** DDNs were made heterogeneous by introducing heterogeniety in delays (α), cluster heterogeneity is designed by sampling delays from different distributions for each cluster, similarly, network heterogeneity is designed by sampling delays from a single distribution. **(b)** same heterogeneity designed for ESNs. It can also be seen that the key difference between ESNs and DDNs are the introduction of delays between units in case of DDNs.

CMA-ES learning algorithm (see Method section 5.3). We evaluated the task performance for NARMA-30 and Mackey-Glass tasks using NRMSE and prediction horizon measures respectively (see Method section 5.1).

Task Performance

NARMA 30

In order to study the effect of heterogeneity in leak parameters in ESNs and DDNs, we evaluated the networks using their validation scores for each generation of CMA-ES algorithm shown in **Fig. 4.2a** while these networks learn to perform NARMA-30 task (see Methods). It can be seen that as the networks are optimized over generations, the

validation scores lower for both ESNs and DDNs, but the NRMSE scores for DDNs, especially DDNs with network wide fixed decay, per cluster distributed decay and per cluster fixed decay are much lower than ESNs. It can also be seen that all ESNs are optimized to similar NRMSE scores. Interestingly, it can be seen that DDNs with per-cluster fixed decay out perform, DDNs with higher levels of heterogeneity, while the DDNs with network wide distributed decay showed similar performance as the ESNs. We also evaluated the performance of the best optimized networks on a test set shown in **Fig. 4.2b**, it can be seen that DDNs with network wide fixed decay has lower NRMSE score than ESNs. For per-cluster case, the performance of DDNs and ESNs are indistinguishable.

Mackey-Glass

Similar to the NARMA-30 task we optimized ESNs and DDNs with varying levels of heterogeneity for Mackey-Glass time series prediction task (see Methods section), we evaluated the performance of the networks using prediction horizon metric (see Methods section), it can be seen from **Fig. 4.2b** that prediction horizons for validation run over generation of CMA-ES optimization for DDNs are consistently higher than the ESNs. It can also be seen that DDNs with network wide distributed decay performs the best among all the DDNs, reaching the prediction horizon of 500 steps. It is also noticeable that ESNs with per cluster distributed decay and network wide fixed decay perform the worse compared to all other variants. ESNs with network wide distributed decay and per cluster fixed decay perform objectively better, reaching a prediction horizon value of 300 steps. Similar to the NARMA-30 taks, we evaluated the performance of the best optimized networks on a test set shown in **Fig. 4.2b**, it can be seen that DDNs with network wide fixed decay has lower NRMSE score than ESNs.

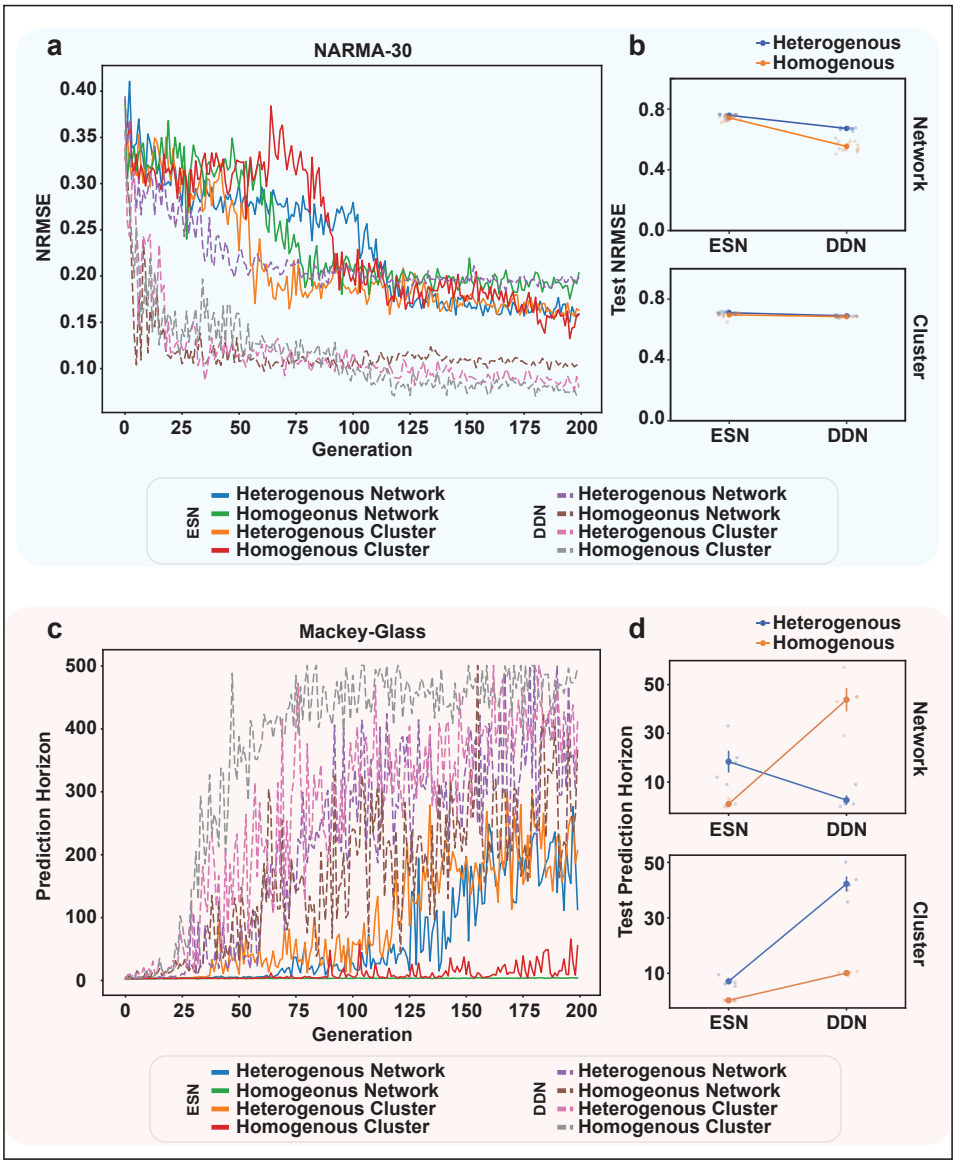

Figure 4.2. **Network performance for DDNs and ESNs with or without heterogeneity**: **(a)** Line plot shows the perforamce in terms of NRMSE over generation during CMA-ES optimization for DDNs and ESNs during NARMA 30 task.

(b) Test NRMSE for different network configurations for both ESNs and DDNs. It can be seen that DDNs with/without heterogeneity perform better than ESNs. **(c)** Performance of ESNs and DDNs with or without heterogeneity optimized for Mackey-Glass prediction task over generations of CMA-ES optimization. **(d)** Test performance of ESNs and DDNs optimized for Mackey-Glass timeseries prediction task, separated based on heterogeneity types.

Stability and Dimensionality

Maximum Lyapunov exponents

In order to evaluate the advantage of intrinsic heterogeneity on ESNs and DDNs we calculated the largest lyapunov exponents of the sampled best networks while they perform the task (see Methods section). This was done in order to examine the stability of the network dynamics while the networks perform the task, we hypothesized that intrinsic decay heterogeneity might cause the networks to be more chaotic than their homogeneous counterparts. We evaluated the Maximum Lyapunov exponents for each network and heterogeneity types for different perturbation magnitudes ranging from $1e^{-1}$ to $1e^{-9}$. We perform this analysis for NARMA-30 and Mackey-Glass tasks separately, the results are summarized in **Fig. 4.3a-b**.

For NARMA-30 task (**Fig. 4.3a**), it can be seen that for each perturbation magnitude, the largest lyapunov exponent are negative for all ESNs variants and are lower than DDNs, suggesting that ESNs have more contracting dynamic than DDNs despite the heterogeneity. The outlier to this observation is the ESN with network wide fixed delay, for which the lyapunov exponent is negative but invariant to the perturbation magnitude and comparable to DDNs. It is also important to observe that for all perturbation magnitudes, the highest perturbation yields the lowest exponent. In case of DDNs, the per cluster fixed decay configuration the maximum lyapunov exponent was invariant to the perturbation magnitude. The MLE for DNNs with network wide distributed decay and per cluster distributed decay were comparable, on the other hand,

DDNs with network wide fixed decay shows a positive MLE for highest magnitude perturbation ($1e^{-1}$) suggesting chaotic dynamics.

For Mackey-Glass timeseries prediction task (**Fig. 4.3b**), it can be seen that for each perturbation magnitude, the largest lyapunov exponent are negative for all ESNs variants and are lower than DDNs, suggesting that ESNs have more contracting dynamic than DDNs for this task as well, despite the heterogeneity. It is also important to observe that for all perturbation magnitudes, the highest perturbation yields the lowest exponent just as in case of NARMA-30. The ESNs with network wide distributed decay were found to have the highest MLE among ESNs and the ESNs with per cluster fixed decay were found to have the lowest MLE. In case of DDNs, the per cluster fixed decay configuration shows a positive exponent for the magnitude of $1e^{-}9$ and $1e^{-}7$. The MLE for DNNs with all other variants had comparable negative MLE for all each perturbation magnitude suggesting a stable contracting dynamic.

Dimensionality

In order to further evaluate the usefulness of intrinsic heterogeneity in terms of recruiting reservoir units for performing the task, we calculated the dimensionality of the network using the network states while the network is driven by the task input. We hypothesized that intrinsic decay heterogeneity might increase the dimensionality of the network, improving the task performance. We drove the network with NARMA-30 and Mackey-Glass input for 2000 steps, and used the network state after 300 steps of warm-up period. For NARMA-30 task, we observed low dimensionality for all networks, the highest for ESN with fixed decay per cluster (D=15) and lowest for DDNs with network wide fixed decay (D=8), distributed decay per cluster (D=8) and fixed decay per cluster (D=8). Similarly, for Mackey-glass time series prediction task, we observe low dimensionality for all networks except for ESNs with distributed decay per cluster (D=98). The network with lowest dimension was found to be DDNs with fixed decay

per cluster. Overall, this shows that both ESNs and DDNs do not recruit a high number of units in order to perform both NARMA-30 and Mackey-glass tasks.

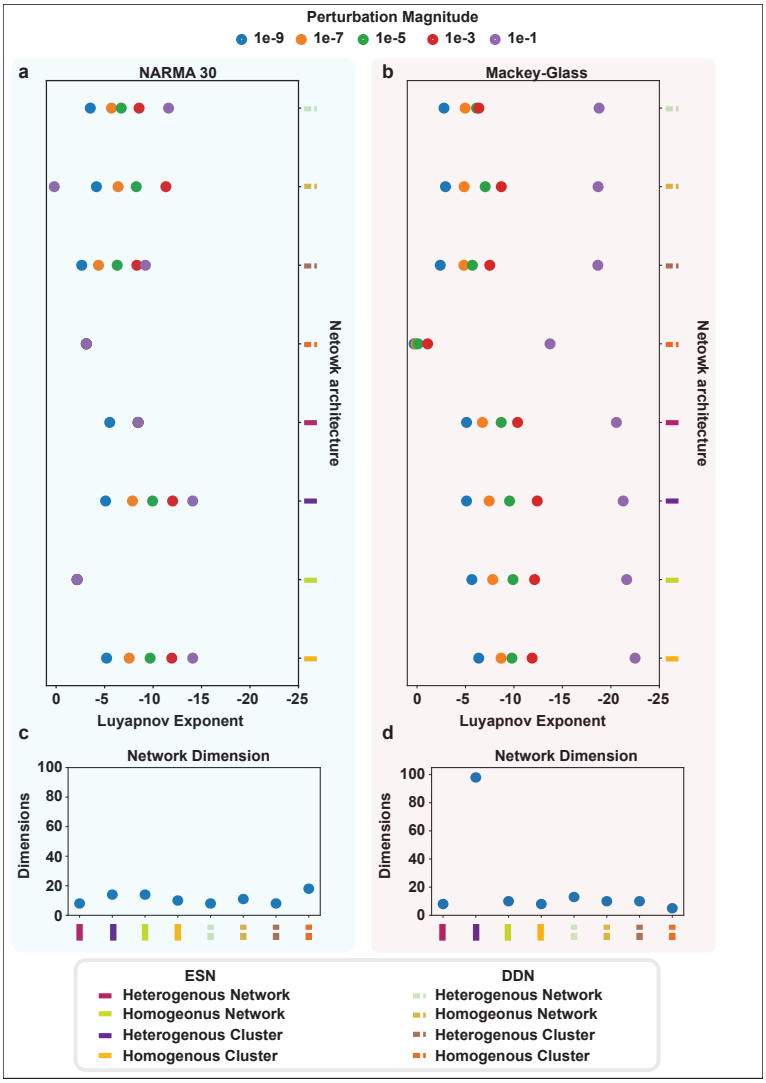

Figure 4.3. **Dynamics based on Lyapunov exponent and dimensionality of optimized networks: (a-b)** shows the Lyapunov exponents of the optimized networks while performing the NARMA 30 (blue background) and Mackey-Glass (red background). **(c-d)** shows the network dimensionality while performing the NARMA 30 (blue background) and Mackey-Glass timeseries prediction tasks (red background).

Linear Memory capacity

In order to understand if intrinsic decay heterogeneity has an effect on the memory of ESNs and DDNs, especially to see if networks with heterogeneous decay parameters change the way networks store temporal information, especially the number of time steps these networks can recall information from, we calculated the linear memory capacity (see methods section). We used the maximum delay of 40 steps. The observed linear memory capacity for each network variant is summarized on **Fig. 4.4**.

For the NARMA-30 task (**Fig. 4.4 left**), it can be seen that every ESN variant has the capacity to recall from upto 20 time steps ago and the capacity is distributed over all the steps between 0 and 20, we do not observe a stark improvement in the memory capacity as a result of heterogeneity. On the other hand, DDNs show a much more nuanced difference in terms of memory distribution for NARMA-30 task, DDNs with fixed decay per cluster show a concentration of capacity at 3 different time steps, while DDNs with distributed decay per cluster show a prominent concentration at two delay steps. It is also clearly observed that DDNs do not utilize the time window closer to the input, creating a except for DDNs with network wide fixed decay. This suggests that ESNs have limited mobility to recall the past steps in order to perfrom the task, and decay heterogeneity doesn't provide any improvement on this. While, DDNs scatter the capacity more sparsely, thus utilizing the past states more strategically.

For the Mackey-Glass time series prediction task (**Fig. 4.4 left**), it can be seen that every ESN variant ustilizes just a few past states from when the input is received, most of the capacity is recruited from just the previous step, we do not observe a stark improvement in the memory capacity as a result of heterogeneity for the case of Mackey-Glass as well. On the other hand, DDNs show a much more variability, DDNs with network wide distributed decay and DDNs with distributed decay per cluster, show a wide spread capacity over multiple delay steps, while DDNs with fixed values show

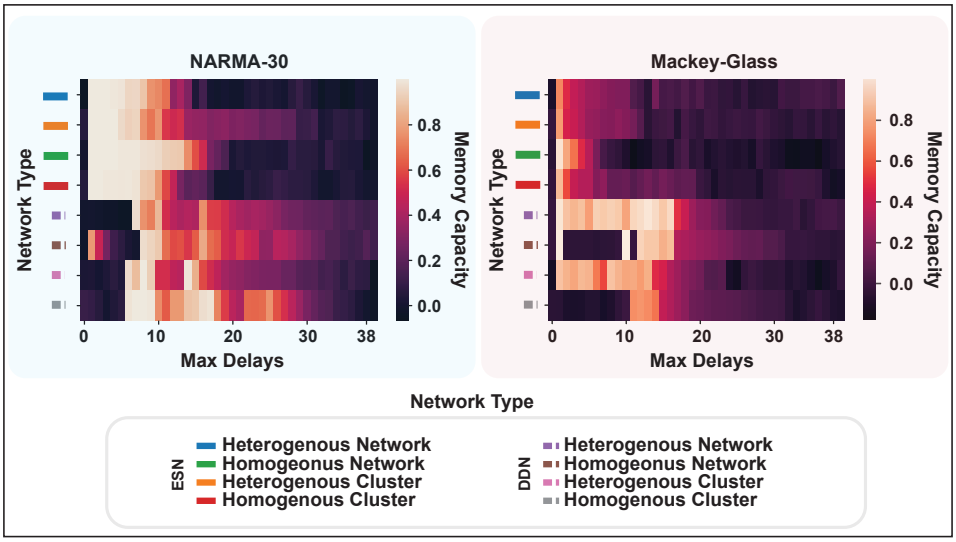

Figure 4.4. **Linear Memory Capacity of optimized networks**: **(Left)** Linear memory capacity of the networks optimized for NARMA 30 task for different delays. **(Right)** Linear memory capacity of the networks optimized for Mackey-Glass time series prediction task for different delays.

a concentration of capacity around 10 time steps ago.

This goes to show that while for ESNs, the decay heterogeneity doesn't change drastically, for DDNs, the memory capacity is altered due to decay heterogeneity, therefore aiding in task performance.

Discussion

This study aimed at studying the effect of timescale (decay parameter) heterogeneity of individual units in reservoir networks on their computational performance, particularly in ESNs and DDNs. Set across two benchmark tasks namely, NARMA-30 and Mackey-Glass, and across four heterogeneity configurations. We found that across all heterogeneity configurations, DDNs outperform ESNs in both tasks. This confirms previous work showing that delay-based architectures are better suited for memory-intensive tasks Iacob et al. (2022), this can be because the introduction of time delays enables more flexible temporal dynamics, even more so than inherent decay, rendering inherent decay redundant. We observed that ESNs trained on NARMA-30 task, homogeneous cluster performs the best. But not for the Mackey-Glass task, where ESNs with Heterogeneous cluster, shows the best performance. Surprisingly, homogeneous network DDNs outperform heterogeneous network ESNs, suggesting that architectural delay mechanisms alone offer substantial computational benefits extending the hypothesis Tanaka et al. (2022) that diverse timescales enhance network performance in both DDNs and ESNs.

The ultimate best-performing configuration across both tasks was the homogeneous cluster DDNs. This suggests that structured, modular heterogeneity in decay parameters is more beneficial than randomly distributed heterogeneity. One likely reason is that this setup creates functionally specialized subpopulations with different integration windows, which quite common in biological neural circuits. It allows the reservoir to simultaneously retain short and long term dependencies spread across multiple units. These results support the hypothesis rooted in neuroscience that the diversity of timescales is functionally beneficial for temporal processing and memory Lundqvist et al. (2016); Runyan et al. (2017a); Cavanagh et al. (2022).

Contrary to our initial hypothesis, networks with decay heterogeneity did not show

increased instability. In fact, all ESNs remained in a contracting regime, with strongly negative Lyapunov exponents across perturbation magnitudes. DDNs were more variable, with some configurations (e.g., Homogeneous network) approaching neutral or even slightly positive exponents under high perturbation. This suggests that DDNs trade off stability for richer dynamics, and that heterogeneity does not necessarily lead to chaotic or unstable behavior. Interestingly, for all network types, the largest perturbation magnitude yielded the most negative Lyapunov exponent. This is likely due to saturation or nonlinear contraction in state space, and highlights the importance of evaluating stability in the local (infinitesimal) regime to interpret true Lyapunov behavior.

We hypothesized that decay heterogeneity would increase the effective dimensionality of the network enabling richer representations. This was not supported. Most networks exhibited relatively low-dimensional dynamics, with little difference between configurations. This could mean that the tasks at hand do not require high-dimensional embeddings, or that the reservoirs learn to compress relevant dynamics efficiently regardless of heterogeneity. It may also suggest that stability and memory, not dimensionality, are the primary contributors to performance in this setting.

Linear memory capacity analysis revealed a more nuanced story. While ESNs showed limited changes across heterogeneity types, DDNs exhibited significant differences in how memory was distributed. Notably, DDNs with distributed decay showed wider memory spread, while fixed-decay variants concentrated memory at specific lags. This suggests that decay heterogeneity allows DDNs to strategically allocate memory a critical property for tasks like NARMA, where both recent and delayed inputs matter. This behavior was absent in ESNs, which tended to have more rigid, shallow memory profiles regardless of heterogeneity.

This study is limited to two benchmark tasks and a specific class of recurrent models. Future work should explore: Broader tasks, including those involving categorical

sequence prediction or noisy real-world data, other forms of heterogeneity (e.g., in connectivity, input weights), how decay heterogeneity interacts with delay heterogeneity under more complex learning rules or online adaptation. It would also be valuable to explore the Information processing capacity of these systems, and how dimensionality evolves over time or under different input statistics.

Our results suggest that introducing modular heterogeneity in decay dynamics can meaningfully improve performance and memory organization in reservoir networks particularly when paired with architectural delays. From a neuroscience perspective, this supports the idea that diverse time constants in biological systems are not incidental, but serve computational roles such as memory stratification and temporal coding.

From an engineering viewpoint, these findings argue for designing non-uniform, structured reservoirs in artificial systems, especially for tasks involving long-range dependencies or multi-timescale dynamics. Rather than tuning a single decay or leak parameter, practitioners could benefit from multi-cluster architectures with targeted decay profiles.

Methods

Network Design

Echo State networks

Echo state networks (ESN) are essentially recurrent networks with fixed weights and randomly connected units given by the following equation:

$$x(n+1) = (1-\alpha)x(n) + \alpha f(W_{res}x(n) + b_{res} + W_{in}v(n)) \tag{4.1}$$

Contrary to RNNS, only the linear readout layer is trained for an ESN in order to optimize it for a task. In the given formulation above, x(n) represents the state of the reservoir at time step n. The size of the reservoir is given by the number of units N. W_{res} represents the $N \cdot N$ weight matrix of the reservoir. Symbol W_{in} represents a $N_{in} \cdot N$ input weight matrix. b_{res} represents the bias reservoir weights of size N. The symbol $v(n)$ is the input to the reservoir at time step n. α is the leak parameter, which decides the importance of previous states on the current state, this acting as memory. This leak parameter can be fixed or distinct for each unit. $f(\cdot)$ is the non-linear activation function, which is usually sigmoid or hyperbolic tangent.

Distance based delay networks

The architecture of Distance based delay networks are similar to ESNs except for the introduction of delays, the delays are implemented by assuming that each unit lies on a 2-D Euclidean space and the distance between two units determine the delay between them. A distance matrix D is computed, where each element D_{ij} represents the distance between units i and j scaled by signal propagation velocity and simulation timestep. A new masked weight matrix $W^{res}_{D=d}$ is created accounting for delays, each element in this matrix is given by the following equation:

$$W_{i,j,D=d} = \delta_{i,D_{ij}} \cdot W_{i,j} \tag{4.2}$$

here, d is ∈ [1, D_{max}] and δ is the Kronecker delta operator. D_{max} is the maximum value delay can take. This method allows for all the weights other than at delay of d to be zero. The final update equation for DDNs is given by:

$$i(n) = \sum_{d=0}^{D_{max}} \left(W^{res}_{D=d} x(n-d) + W^{in}_{D=d} v(n-d) \right) + b_{res} \tag{4.3}$$

$$x(n) = (1-\alpha)x(n-1) + \alpha f(i(n)) \tag{4.4}$$

All the parameters in the above set of equations are identical except for the addition of delays. As described in Iacob et al. (2022), the delay between individual units are treated as hyperparameters to be optimized. The coordinates are sampled using a Gaussian Mixture model, with K clusters, instead of finding the optimal coordinates for each unit, the GMM model is optimized to find the fitting distribution for each cluster.

Task

NARMA-30

The Nonlinear auto-regressive moving average (NARMA) is a popular benchmark task used to evaluate the performance of RC system. It is a system of equation that is highly non-linear and has clear temporal dependencies, defined by the order parameter. The general system of equation is given by the following equation:

$$y(t+1) = c_1 \cdot y(t) + c_2 \cdot \sum_{i=0}^{p} y(t-i) + c_3 \cdot u(t-p)u(t) + c_4 \tag{4.5}$$

Where $c_1 - c_4$ are task parameters, and p is the NARMA order. $y(t)$ is the state of the system at time t, $u(t)$ is the input to the system at time t, which is uniformly

sampled between 0.0 and 0.5. Essentially the task entails predicting the next state given the serialized previous states, we used the NARMA-30 task to train our networks, the parameters are $p = 29,\ c_1 = 0.2,\ c_2 = 0.004,\ c_3 = 1.5,$ and $c_4 = 0.1$.

Evaluation

We evaluated the NARMA-30 task using Normalized root mean squared error (NRMSE) given as follow:

$$NRMSE = \frac{1}{\bar{y}}\sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \tag{4.6}$$

where $\bar{y}$ is the mean of the ground truth, y_i and $\hat{y}_i$ are ground truth and network readout respectively, n is the total number of elements in the time series.

Mackey-Glass

The discretized Mackey-Glass timeseries is given by the following equation:

$$x(t+1) = x(t) + \beta\frac{x(t-\tau)}{1 + x(t-\tau)^n} - \gamma x(t) \tag{4.7}$$

The parameters the following $\tau = 17,\ n = 10,\ \beta = 0.2$ and $\gamma = 0.1$. The aim of this task is to accurately generate the Mackey-Glass sequence as long as possible using the previous step and not receiving any external input or correction. This is done by training the readout layers of the networks for one step ahead prediction. The performance is measured by a metric called prediction horizon, which measures how many time steps into the future can be predicted by the network accurately within an error margin. The error margins are $\pm 0.1\sigma_l^2$, where σ_l^2 is the label variance.

Evaluation

We evaluated the network performance using prediction horizon, which the number of time steps the network can predict the time series within the error margins.

Training

CMA-ES Optimization

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) Hansen et al. (2006) is a widely used evolutionary algorithm for hyperparameter optimization, especially effective for non-convex and high-dimensional problems. CMA-ES consists of a multivariate Gaussian search distribution defined by a mean vector, step-size, and covariance matrix. At each iteration, it samples a population of candidate solutions from this distribution, evaluates their fitness, and updates the mean and covariance matrix to sample future samples from more fitting regions of the search space. The mean vector represents the current best candidate based on the population's progress, while the covariance matrix spans the space of the search distribution. This iterative process enables CMA-ES to efficiently explore complex solution spaces and is robust due to its invariance properties and ability to handle parallel evaluations. In case of ESN and DDN optimization, the mean vectors are the hyperparameters of DDNs and ESNs. Initially, all hyperparameters are roughly made to fall in the same range, after every iteration, a new hyperparameter vector is generated based on the evolution of the multi-variate distribution. Using this new solution vector, networks are generated and are used to perform validation scores. Both ESN and DDN optimizations are run for 200 generations with a population size of 20 for Mackey Glass task and population size 25 for NARMA 30 task.

Readout training

The randomly sampled network is not directly able to perform the task, it is essentially it acts as a temporal kernel that increases the dimensionality of the input and thus increases the linear separability, the activity from a network driven by a random input can be recorded and can be treated as a feature matrix X of size $N \times T$, where N is the

number of neurons and T is the number of time steps of the simulation. The output labels $\hat{y}$ and and the corresponding input vector $v(n)$ are ordered vectors of size T. The objective function J given by the following equation below is solved using ridge regression trained on network activities:

$$J(W_{readout}) = ||XW_{readout} - \hat{y}||_2^2 + \lambda||W_{readout}||_2^2 \tag{4.8}$$

Optimizing the above equation gives the $W_{readout}$, where λ is the regularizing parameter and $W_{readout}$ is the readout.

Lyapunov exponent

Dynamics of the RC system can be studied by estimating the Lyapunov exponent of the network trajectory, we calculated the Lyapunov exponent of ESNs and DDNs using numerical method, explained by Boedecker et al. (2012). Lyapunov exponent is given by the following equation:

$$\lambda = \lim_{t \to +\infty} ln\left(\frac{\gamma_t}{\gamma_0}\right) \tag{4.9}$$

The steps involved in the numerical calculation as described by Boedecker et al. (2012) is as follows:

1. Two identical networks X_1 and X_2 were created taking the best sampled networks, a reservoir unit from one of the network X_1 was given a small perturbation and the network X_2 was kept as is. The initial $L2$ normalized distance between X_1 and X_2 is therefore γ_0.

2. The network is driven by the input that the respective network is trained on, that is either NARMA-30 or Mackey-Glass inputs. At each simulation step t a normalized $L2$ distance between networks $\gamma_t = ||X1(t) - X2(t)||^2$ is calculated.

3. The perturbed network X_1 is reset to the initial distance γ_0, $X_1 \to X_2(t) +$

$(\gamma_0/\gamma_t)(X_2(t) - X_1(t))$ once it crosses the upper or lower thresholds, i.e, $\gamma_t \notin [1e^{-5}, 1e^{-14}]$, this step is essential to prevent numerical overflows.

4. The simulation is run for 200 steps, and is repeated for multiple perturbation intensities ranging from $1e^{-1}$ to $1e^{-9}$. Finally the Lyapunov exponent is given by the equation 4.9.

The above steps are repeated for all the networks for both the tasks. All the numerical simulations were performed using python and Tensorflow 1.x.

Dimensionality

The dimensionality of a given network while performing a task was calculated using Singular Vector decomposition of the activity matrix and counting the top singular values that explains 99% of the variance. The SVD is given as follows:

$$A = U\Sigma V^T \tag{4.10}$$

where U of and V^T are left and right singular matrices of dimensions $N \times N$ and $T \times T$ respectively, Σ is the diagonal matrix of dimensions $N \times T$. The dimensionality is given by:

$$dim = \frac{\sum_{i=1}^{n} \sigma_{i,i}}{\sum_{i=1}^{N} \sigma_{i,i}}; n < N \tag{4.11}$$

where $\sigma_{i,i}$ are the diagonal elements of the singular value matrix Σ, N is the number of reservoir units in ESNs and DDNs.

Linear Memory Capacity

Linear Memory Capacity measures short term memory distribution of a dynamical system Jaeger (2001a). In order to calculate the linear memory capacity, the ENSs and DDNs are driven with a uniform random input and the readout layer is trained to reproduce input from l time lag ago. The linear memory capacity at a specific lag l

shows how well the reproduction by a network correlates with the input l time steps ago. Total memory capacity is given by summing all the memory capacity for each lag. The idea is formulated as follows:

$$MC_l = r(u(n-l), \hat{u}(n))^2 \tag{4.12}$$

$$MC_{total} = \sum_{l=1}^{\infty} MC_l \tag{4.13}$$

Where $r(\cdot)$ is Pearson correlation, $u(n)$ is the input at timestep n and $\hat{u}(n)$ is the readout reproduction of the input at timestep $n-l$.

References

Boedecker, J., Obst, O., Lizier, J. T., Mayer, N. M., and Asada, M. (2012). Information processing in echo state networks at the edge of chaos. *Theory in Biosciences*, 131:205–213.

Brunel, N. and Wang, X.-J. (2003). Firing frequency, spike timing and synaptic plasticity of the hippocampal neurons: modeling and experiments. *Journal of Neurophysiology*, 90(1):415–430.

Cavanagh, S., Towers, J., Wallis, J., Hunt, L., and Kennerley, S. (2022). Reconciling persistent and dynamic hypotheses of working memory coding in prefrontal cortex. *Nature Communications*, 13:1–18.

Cavanagh, S. E., Hunt, L. T., and Kennerley, S. W. (2020). A diversity of intrinsic timescales underlie neural computations. *Frontiers in neural circuits*, 14:615626.

Chu, J. C., Wang, Q., Baldassano, C., and Guo, N. D. (2020). Long-timescale processing in human auditory and visual cortex supports short-term memory. *Nature Neuroscience*, 23:1686–1695.

Habashy, K. G., Evans, B. D., Goodman, D. F., and Bowers, J. S. (2024). Adapting to time: Why nature may have evolved a diverse set of neurons. *PLOS Computational Biology*, 20(12):e1012673.

Hansen, N., Müller, S. D., and Koumoutsakos, P. (2006). The cma evolution strategy: A comparing review. In Lozano, J., Larrañaga, P., Inza, I., and Bengoetxea, E., editors, *Towards a New Evolutionary Computation. Advances in the Estimation of Distribution Algorithms*, pages 75–102. Springer.

Hasson, U., Yang, E., Vallines, I., Heeger, D., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, 28(10):2539–2550.

Iacob, S., Freiberger, M., and Dambre, J. (2022). Distance-based delays in echo state networks. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 211–222. Springer.

Izhikevich, E. M. (2006). Polychronization: Computation with spikes. *Neural Computation*, 18(2):245–282.

Jaeger, H. (2001a). Short term memory in echo state networks.

Jaeger, H. (2001b). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *GMD Technical Report 148*.

Koch, C. and Laurent, G. (1999). Complexity and the nervous system. *Science*, 284(5411):96–98.

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466:123–127.

Lukoševičius, M. and Jaeger, H. (2009). Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149.

Lundqvist, M., Rose, J., Herman, P., Brincat, S., Buschman, T., and Miller, E. (2016). Gamma and beta bursts underlie working memory. *Neuron*, 90(1):152–164.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M.-B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678.

Nam, L., Kang, R., Kim, D., and Marder, E. (2017). Diversity matters: Neural variability promotes learning and generalization. *Neuron*, 96(4):795–807.

Runyan, C., Piasini, E., Panzeri, S., and Harvey, C. (2017a). Distinct timescales of population coding across cortex. *Nature*, 548:92–96.

Runyan, C. A., Piasini, E., Panzeri, S., and Harvey, C. D. (2017b). Distinct timescales of population coding across cortex. *Nature*, 548(7665):92–96.

Schrauwen, B., Verstraeten, D., and Van Campenhout, J. (2007). An overview of reservoir computing: theory, applications and implementations. *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*.

Soriano, M. C., Ortín, S., Keuninckx, L., Appeltant, L., Danckaert, J., Pesquera, L., and Van der Sande, G. (2014). Delay-based reservoir computing: noise effects in a combined analog and digital implementation. *IEEE transactions on neural networks and learning systems*, 26(2):388–393.

Tanaka, G., Matsumori, T., Yoshida, H., and Aihara, K. (2022). Reservoir computing with diverse timescales for prediction of multiscale dynamics. *Physical Review Research*, 4(3):L032014.

Wang, P. X., Farzadfard, F., and Pehlevan, C. (2020). Heterogeneity of time constants improves memory performance in recurrent neural networks. *arXiv preprint arXiv:2012.13074*.

Zhou, W., Fiete, I., and Lampl, I. (2023). Temporal diversity in neural populations enhances robustness and memory capacity in recurrent networks. *Science Advances*, 9(10):eaaz8693.

Chapter 5

General Discussion

Chapter 6

Supplementary Materials

Supplementary materials of chapter 2

Waveform and Action Potential properties manifold comparison

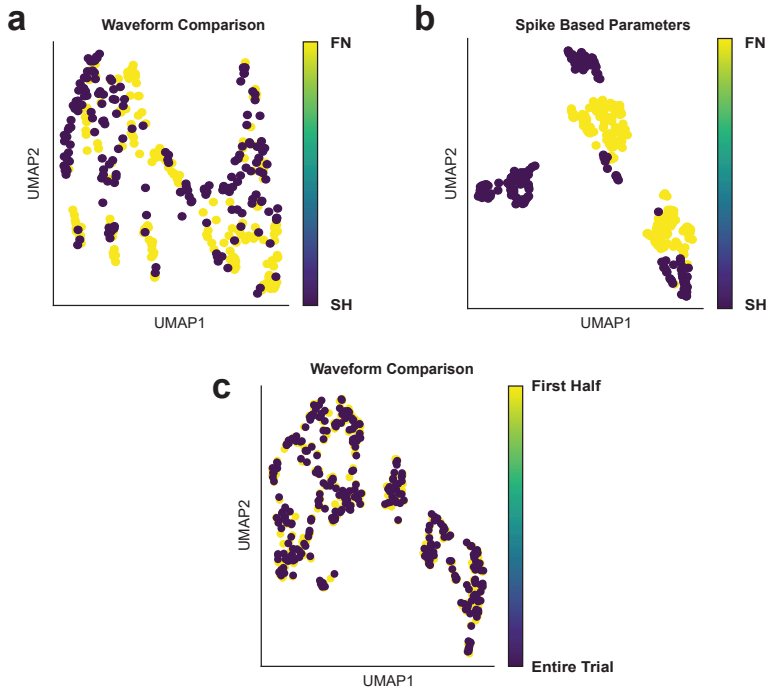

Figure 6.1. (**a**) Overlaid UMAP representation FN and SH waveforms from 186 neuron used in classification. The waveform shapes are different between SH and FN protocols.(**b**) FN and SH Action potential parameters used for classification projected together on the same space. The Action potential properties are different between FN and SH protocols. The SH Action potential properties show a bigger spread than FN properties. (**c**) UMAP projection of averaged Waveform comparison between the first half and the entire trial. The average waveform shape doesn't change as a result of trial length.

Louvain vs Ensemble Clustering for Graphs (ECG) algorithm comparison

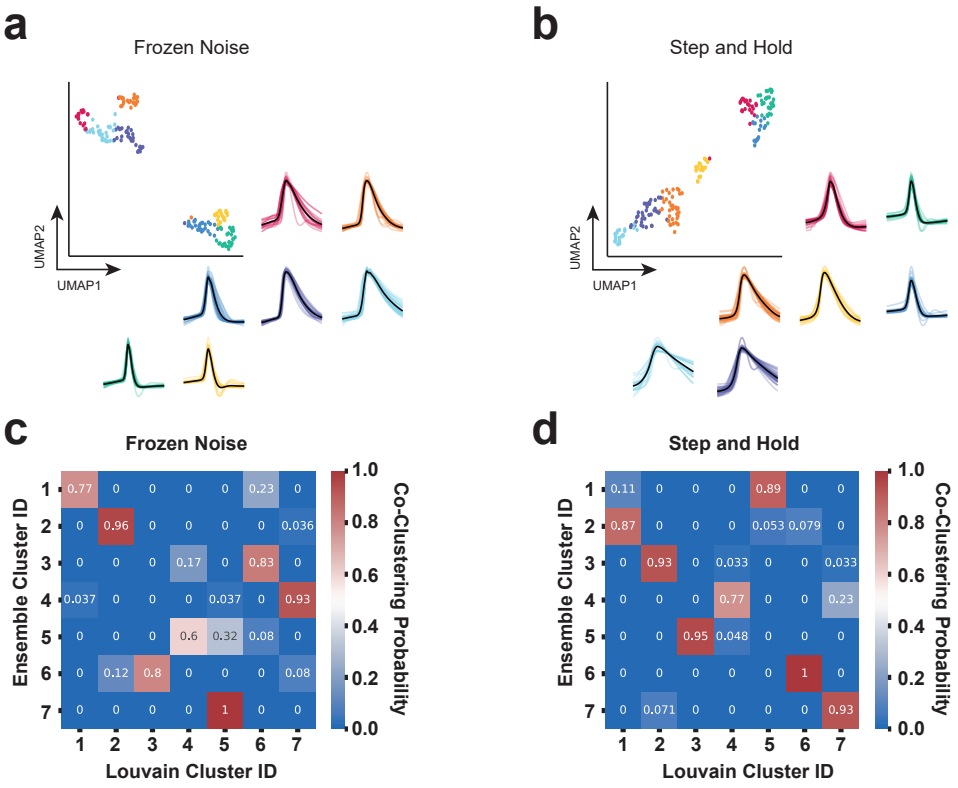

Figure 6.2. (**a**) UMAP embedding of FN waveforms colored with cluster labels found using Ensemble clustering, with original waveforms in the same color as the respective color. 7 clusters were observed, the same number as in the case of the Louvain community method. (**b**) UMAP embedding of SH waveforms colored with cluster labels found using Ensemble clustering, with original waveforms in the same color as the respective color. 7 clusters were observed, the same number as in the case of the Louvain community method.

(**c**) Heatmap showing the correspondence between Louvain and Ensemble clustering for graph on FN waveforms. Clusters using the Louvain community detection algorithm show a high correspondence with the clusters obtained using Ensemble clustering for the graph method. (**d**) Heatmap showing the correspondence between Louvain and Ensemble clustering for graph on SH waveforms. Clusters using the Louvain community detection algorithm show a high correspondence with the clusters obtained using Ensemble clustering for the graph method.

Cluster stability after leaving one attribute out at a time

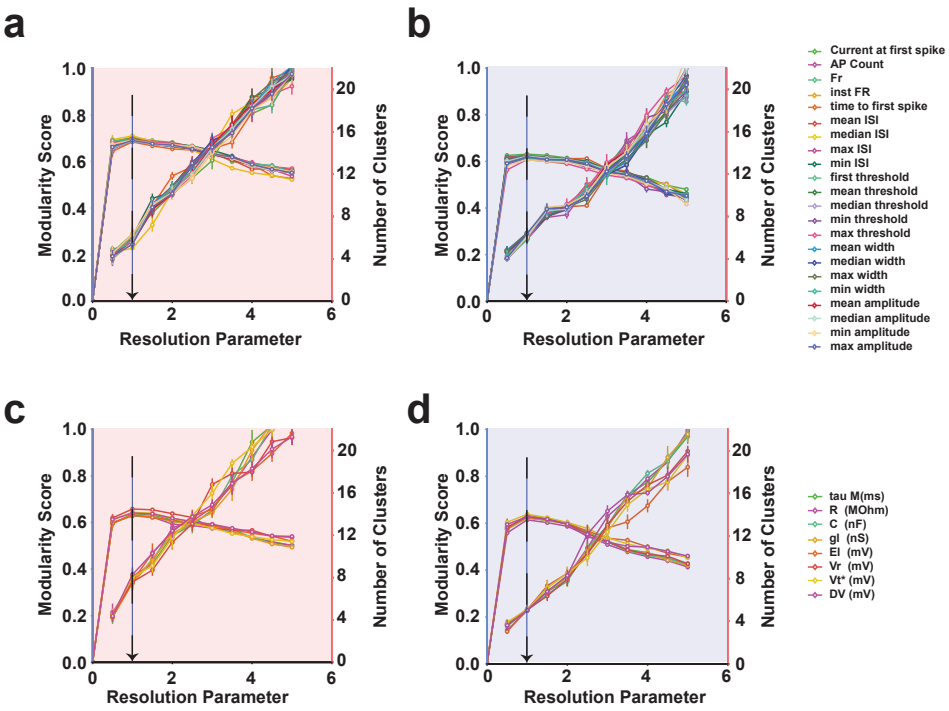

Figure 6.3. (**a-b**) Stability of clusters for action potential attributes leaving one attribute out for excitatory and Inhibitory sets. The clustering was performed 25 times with random 80% samples for resolution attributes ranging from 0.0 to 1, the mean and standard deviation of the resulting number of clusters and modularity score are plotted. For the chosen resolution parameter (1.0), the cluster numbers fluctuate between 5-7. The cluster number fluctuates more for the excitatory set (left) than the Inhibitory set (right) (**c-d**) Stability of clusters for biophysical attributes leaving one attribute out for excitatory and Inhibitory sets. The clustering was performed 25 times with random 90% samples for resolution parameters ranging from 0.0 to 1, the mean and standard deviation of the resulting number of clusters and modularity score are plotted. For the chosen resolution parameter (1.0), the excitatory population fluctuates between 6-8. On the contrary, the Inhibitory population doesn't show any fluctuation.

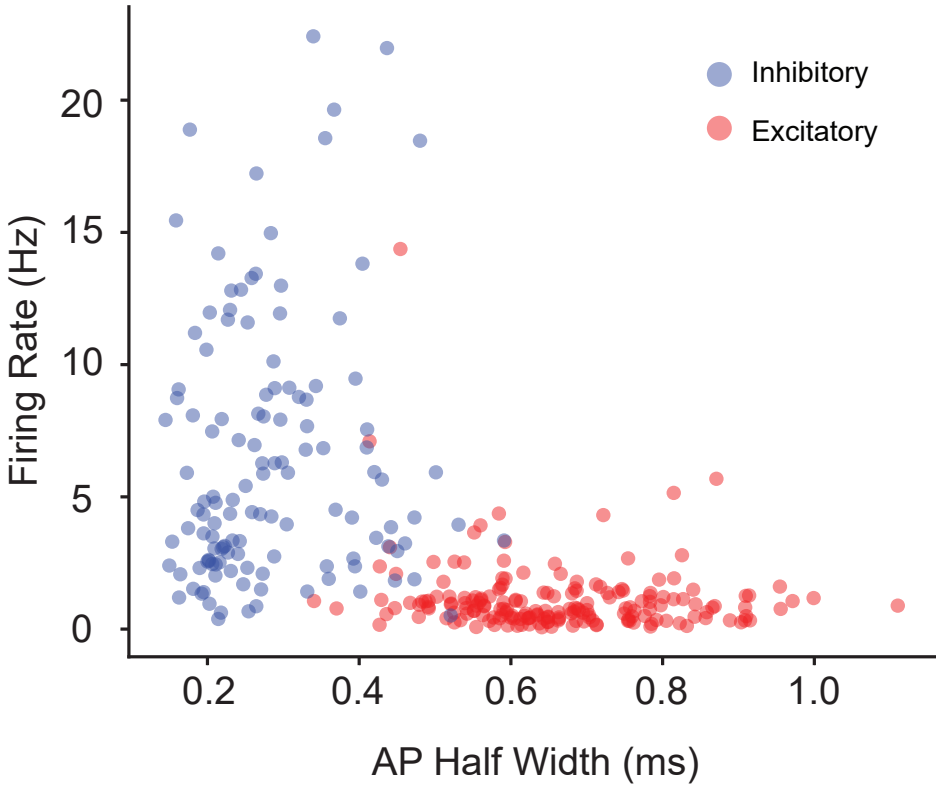

Figure 6.4. Firing rate vs AP half-width between excitatory (red) and inhibitory (blue) populations. The excitatory population has a lower firing rate and higher AP width. The Inhibitory population has a lower AP width and higher firing rate.

STA Heterogeneity

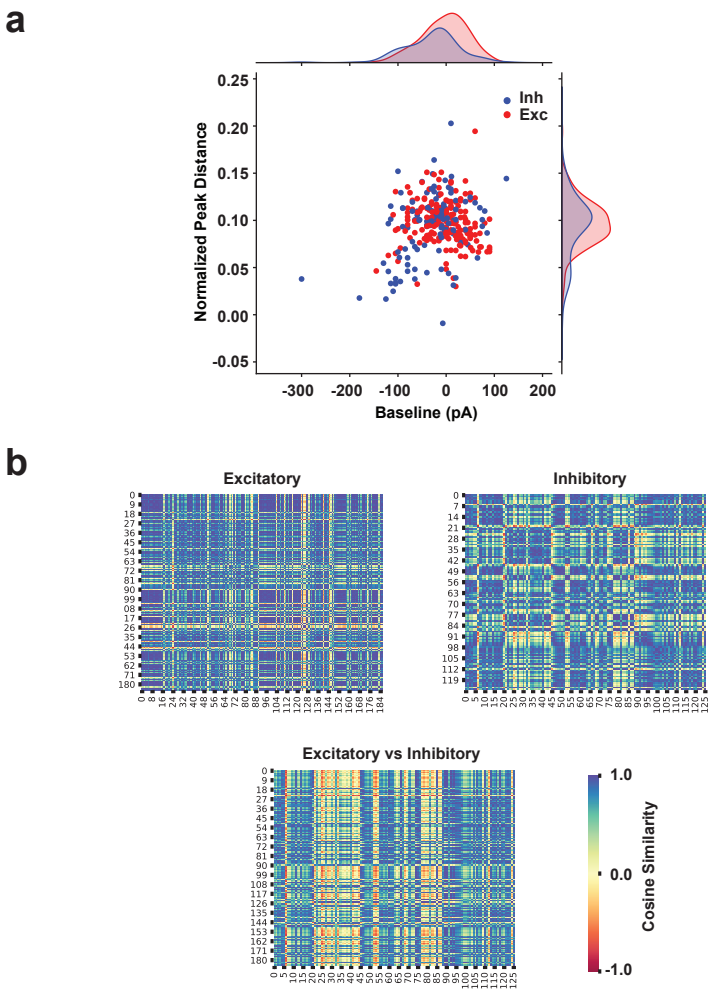

Figure 6.5. (**a**) The scatter plot shows the baseline added to the theoretical input vs the Peak distance of the STA. The excitatory population has a higher baseline and higher peak distance. The variance for both baseline and peak distance seems higher for the excitatory population than inhibitory population. (**b**) The STA cosine similarity is for the excitatory population (top left), the STA cosine similarity is for the inhibitory population (top right), and the STA cosine similarity is between the excitatory and inhibitory populations. The excitatory population has higher similarity than the inhibitory population.

Cluster assignment likelihood across attributes

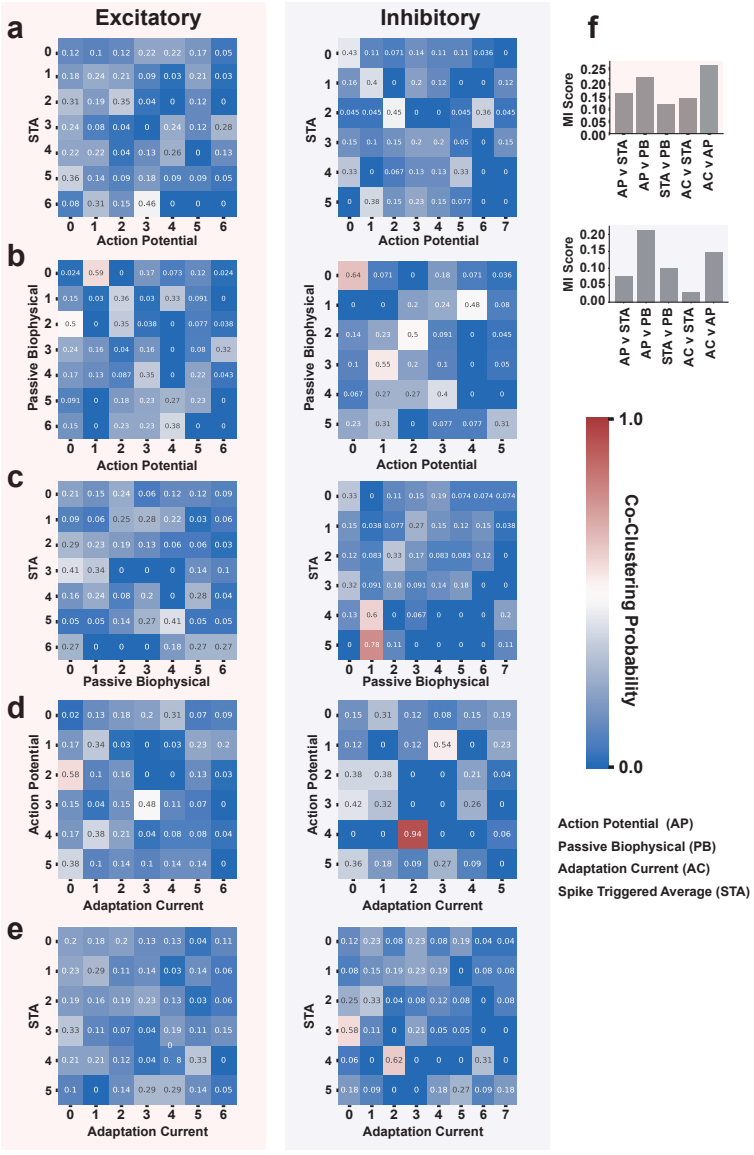

Figure 6.6. (**a-e**) Heatmap showing the likelihood for neuron clustering in attribute 1 clustering

in one of the clusters in attribute 2 for excitatory (left) and inhibitory (right) population. For each attribute (Action potential, Passive biophysical, STA, Adaptation current (η), none of the clusters show a high likelihood to be clustering in another attribute. (f) Cluster label comparison between modularity pairs using Mutual information score for excitatory and inhibitory population. None of the pairs reach the modularity of 0.25 showing that neurons cluster differently across attribute sets.

Supplementary materials of chapter 3

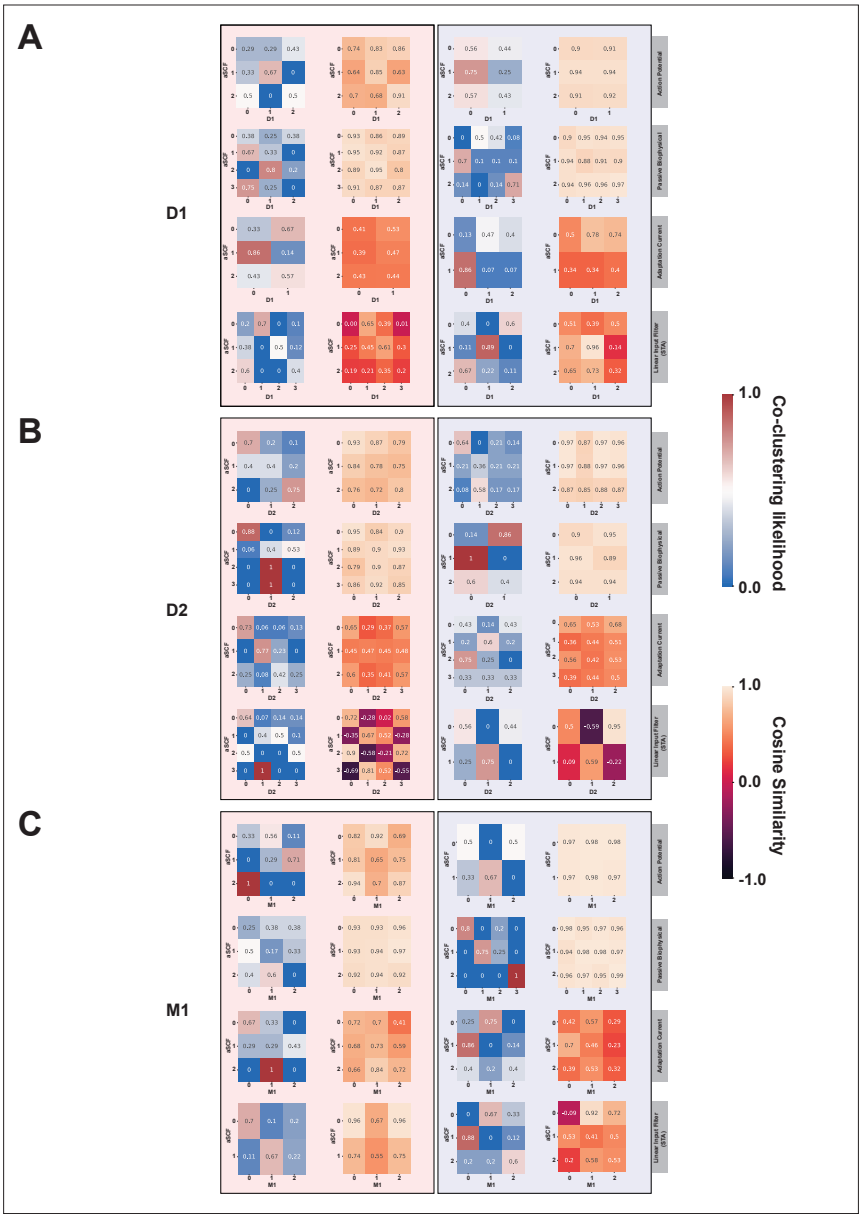

Figure 6.7. **Clustering likelihood matrix and cosine similarity across labels shifts for all four properties as a result of neuromodulation for D1-R, D2-R and M1-R**

Top panel shows the cluster likelihood and cosine similarity for D1 vs aCSF trials. Middle shows the cluster likelihood and cosine similarity for D2 vs aCSF trials. Bottom shows the cluster likelihood and cosine similarity for M1 vs aCSF trials.

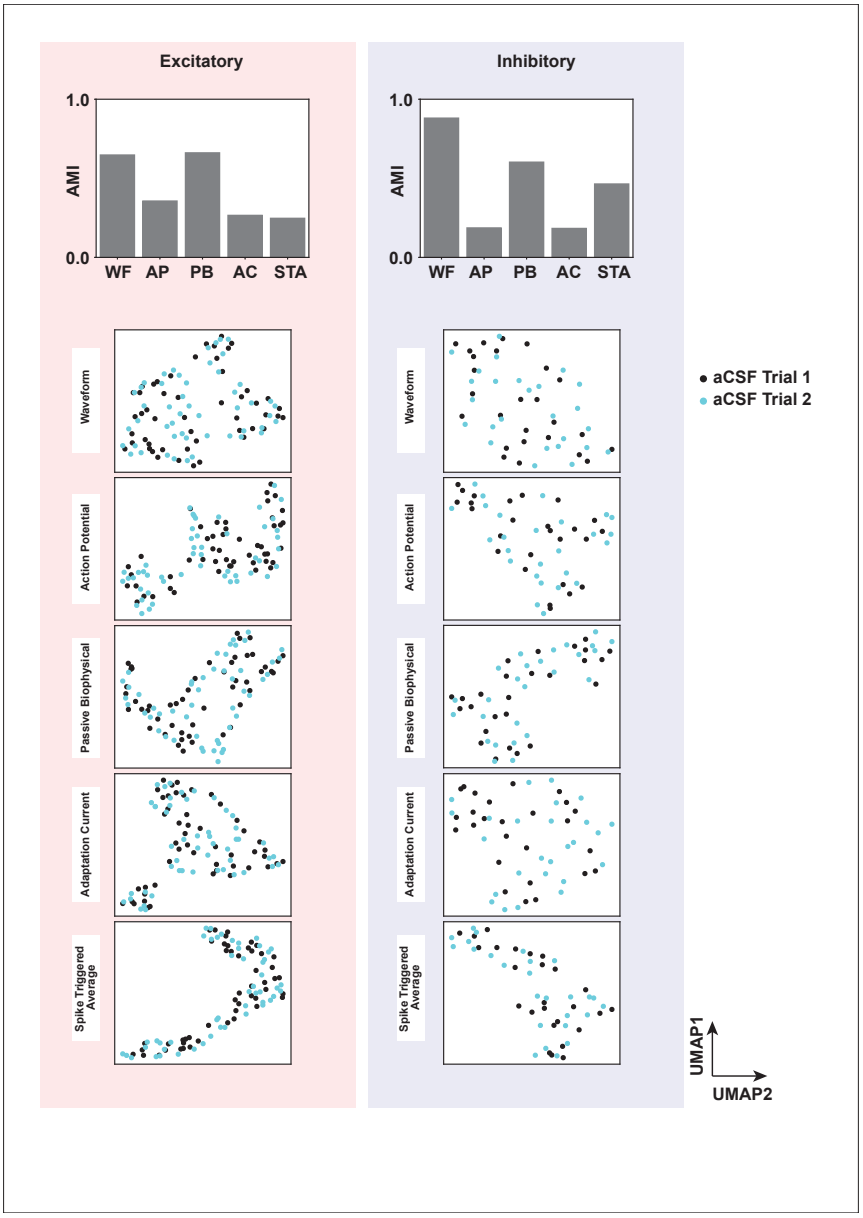

Figure 6.8. **Comparison of aCSF trial 1 vs trial 2 clustering and manifold**

Left and right histograms on top the AMI score between cluster IDs found for aCSF trial 1 vs aCSF trial 2. Waveforms and passive biophysical properties were found to be consistent across the two trials as these properties are not strongly dependent input dynamics. The UMAP plots show the alignment for aCSF trial 1 vs aCSF trial 2.

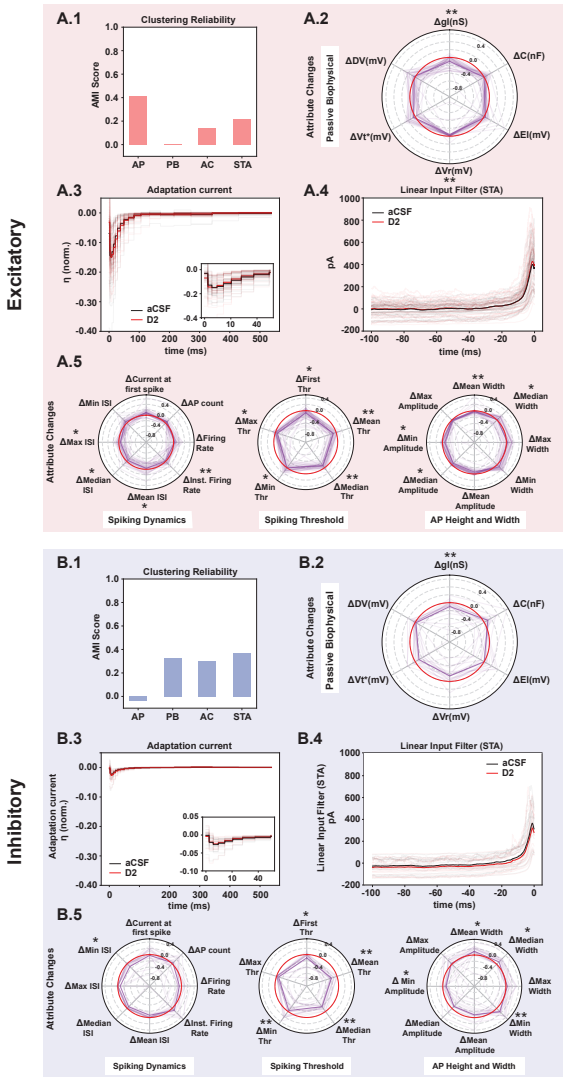

Figure 6.9. **D2 receptor activation changes functional clustering and attributes for both excitatory and inhibitory neurons.**

(**A.1**) Histogram shows the adjusted mutual information score between the clustering labels obtained for aCSF and D2 trials, the histogram shows a shift in cluster identities as a result of D2 receptor activation across four attributes. (**A.2**) shows the change in passive parameters with respect to control as a result of D2 receptor activation, normalized by the control trial values. With the mean marked represented with a thick line. (**A.3**) shows the adaptation current for control (black) and D2 (red), the mean is represented with darker curves. The adaptation current profiles between D2 and aCSF trials were found to be similar. The rise time as well the peak adaptation current between aCSF and D2 were found to be non-significant (see Appendix 6.11). (**A.4**) shows the spike triggered average profile for aCSF (black) and D2 (red) trials, the mean is represented with a thick line. (**A.5**) shows the change in action potential attribute sets with respect to control as a result of D2 receptor activation, normalized by the control trial values. With the mean marked represented with a thick line and the zero line is colored in red. The significant features are marked. **B.1-5** Same analysis repeated for inhibitory neurons. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

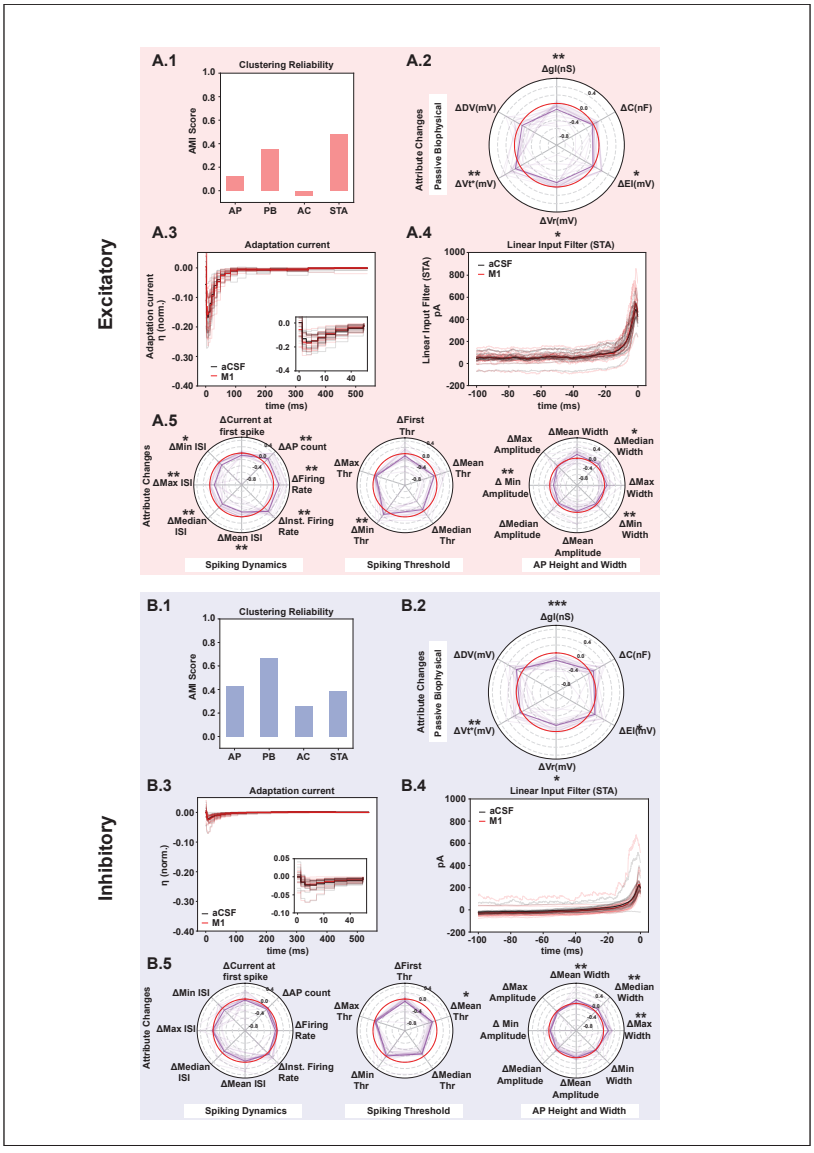

Figure 6.10. **M1 receptor activation changes functional clustering and attributes for both excitatory and inhibitory neurons.**

(**A.1**) Histogram shows the adjusted mutual information score between the clustering labels obtained for aCSF and M1 trials, the histogram shows a shift in cluster identities as a result of M1 receptor activation across four attributes. (**A.2**) shows the change in passive parameters with respect to control as a result of M1 receptor activation, normalized by the control trial values. With the mean marked represented with a thick line. (**A.3**) shows the adaptation current for control (black) and M1 (red), the mean is represented with darker curves. The adaptation current profiles between M1 and aCSF trials were found to be similar. (**A.4**) shows the spike triggered average profile for aCSF (black) and M1 (red) trials, the mean is represented with a thick line. (**A.5**) shows the change in action potential attribute sets with respect to control as a result of M1 receptor activation, normalized by the control trial values. With the mean marked represented with a thick line and the zero line is colored in red. **B.1-5** Same analysis repeated for inhibitory neurons. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

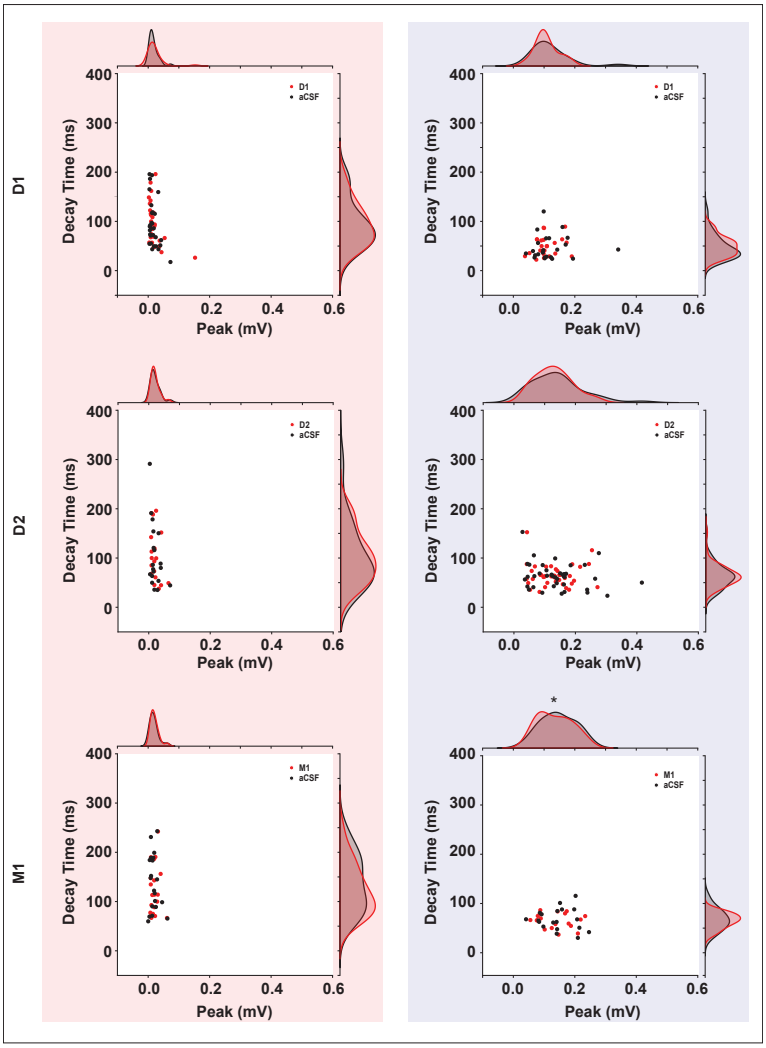

Figure 6.11. **Adaptation current peaks and decay times are not strongly altered by D1, D2 and M1 modulation** The scatter plots show the normalized peaks of the AC curve on the x-axis and the decay time for D1, D2 and M1 vs aCSF trials. The red panel represents excitatory and blue panel represents inhibitory. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$.

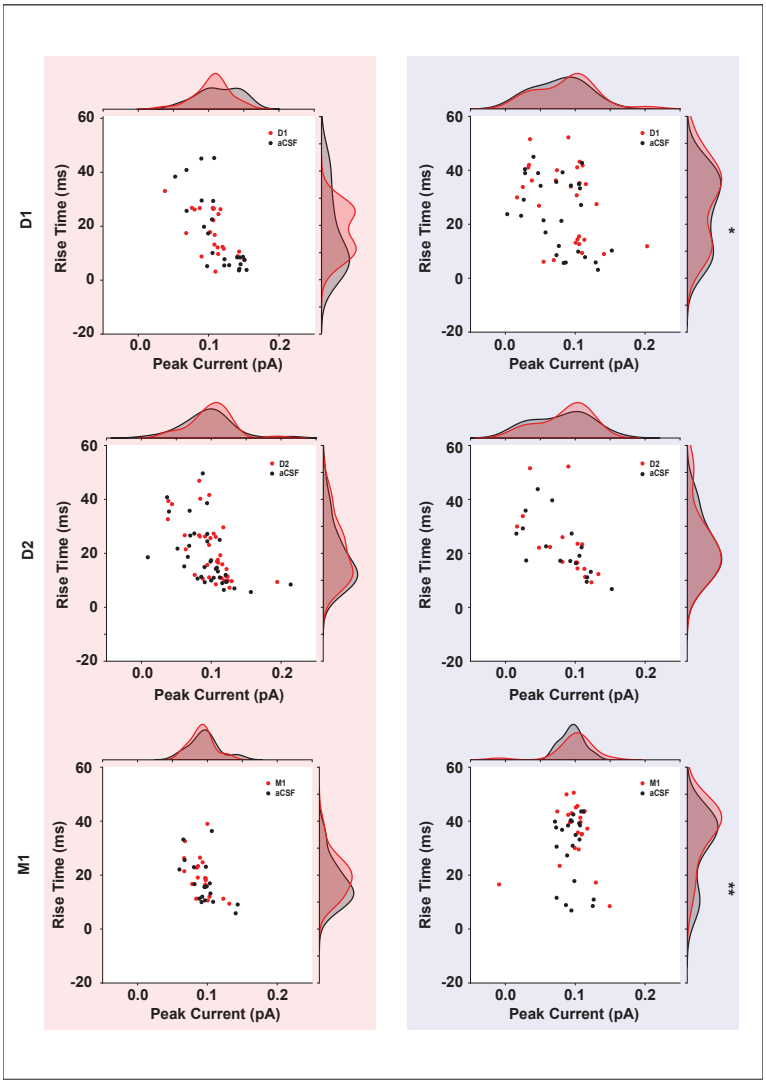

Figure 6.12. **Spike triggered average peaks and decay times are modestly altered by D1, D2 and M1 modulation** The scatter plots show the normalized peaks of the STA curve on the x-axis and the decay time for D1, D2 and M1 vs aCSF trials. The red panel represents excitatory and blue panel represents inhibitory population. * $p < 0.05$, ** $p < 0.001$, *** $p < 0.0001$

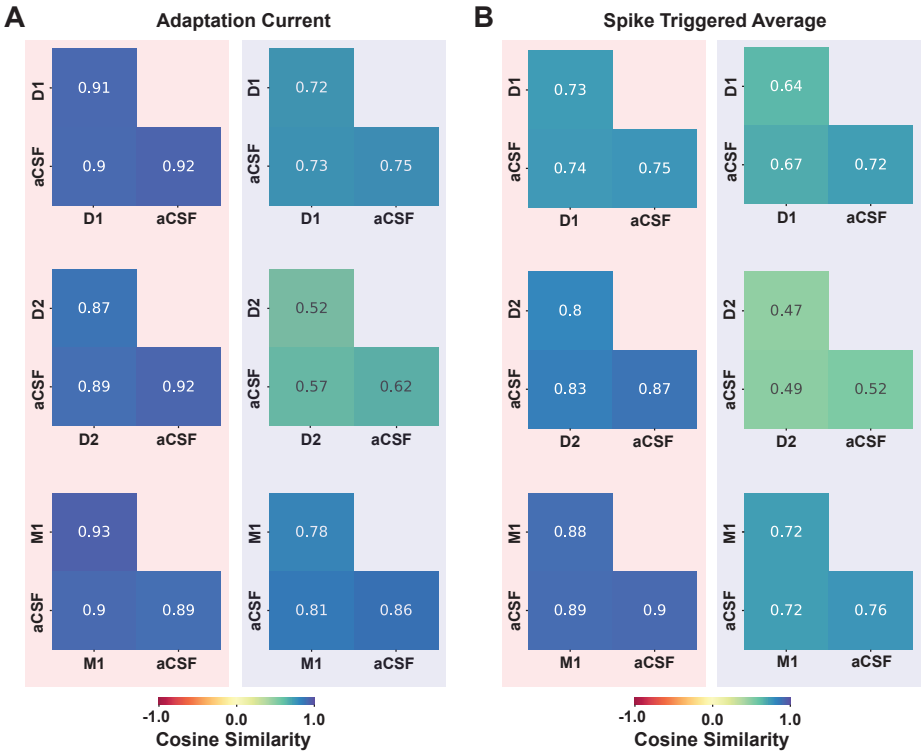

Figure 6.13. **Average cosine similarity across and within aCSF-Agonist trials vary in a cell-type specific manner.** (**A**) Heatmaps show the average AC similarity for excitatory (red background) and inhibitory (blue background) for D1, D2 and M1 vs aCSF trials. (**B**) Heatmaps show the average STA similarities for excitatory (red background) and inhibitory (blue background) for D1, D2 and M1 vs aCSF trials.

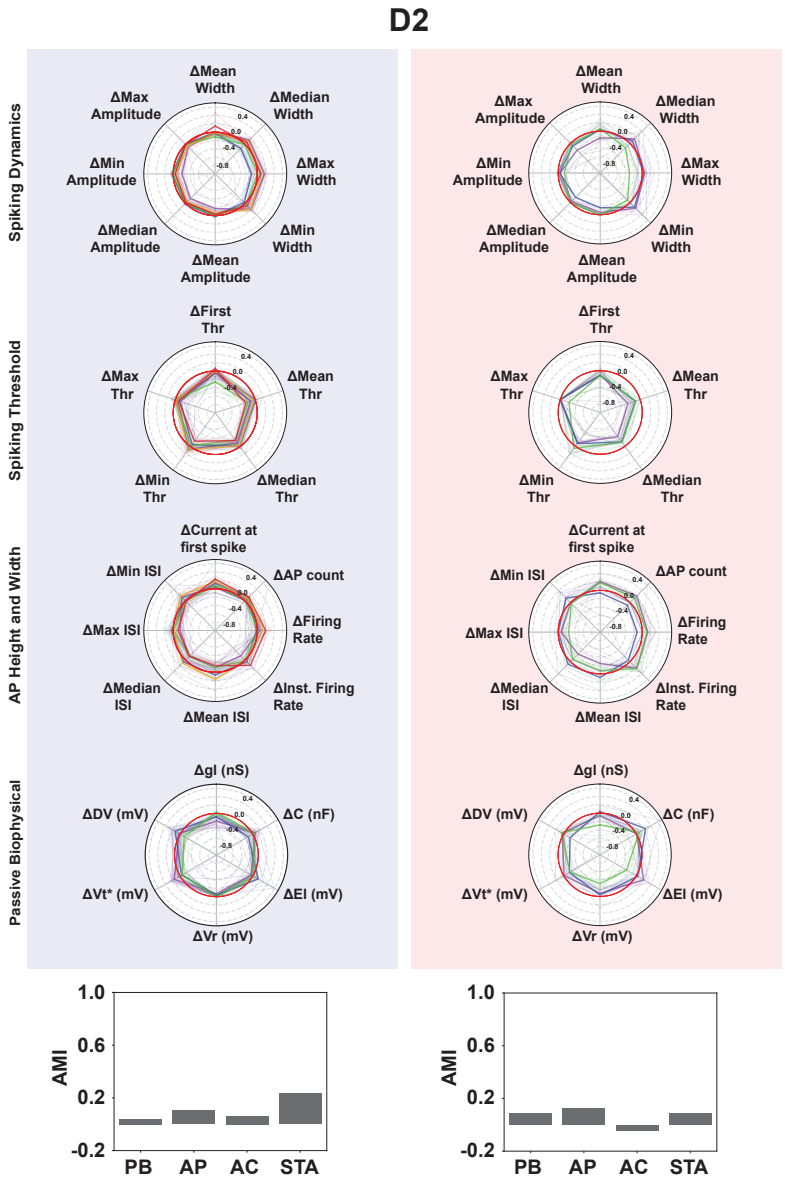

Figure 6.14. **Clustering based on differences between control and D2 for action potential and passive biophysical attributes reveals subgroups of neurons getting modulated differently as a result of D2-R activation.**

The polar plots show the clusters based on difference values between D2 and aCSF trials for action potential (subdivided into spiking dynamics, spiking threshold and AP height and width) and passive biophysical properties for both, excitatory (red background) and inhibitory (blue background). Each neuron is represented with a thin line and Coloured with their respective cluster label. The mean for each cluster is represented with a thick line. The histogram at the bottom shows the AMI score between cluster labels using aCSF trials and the cluster labels based on difference between D2 and aCSF trials.

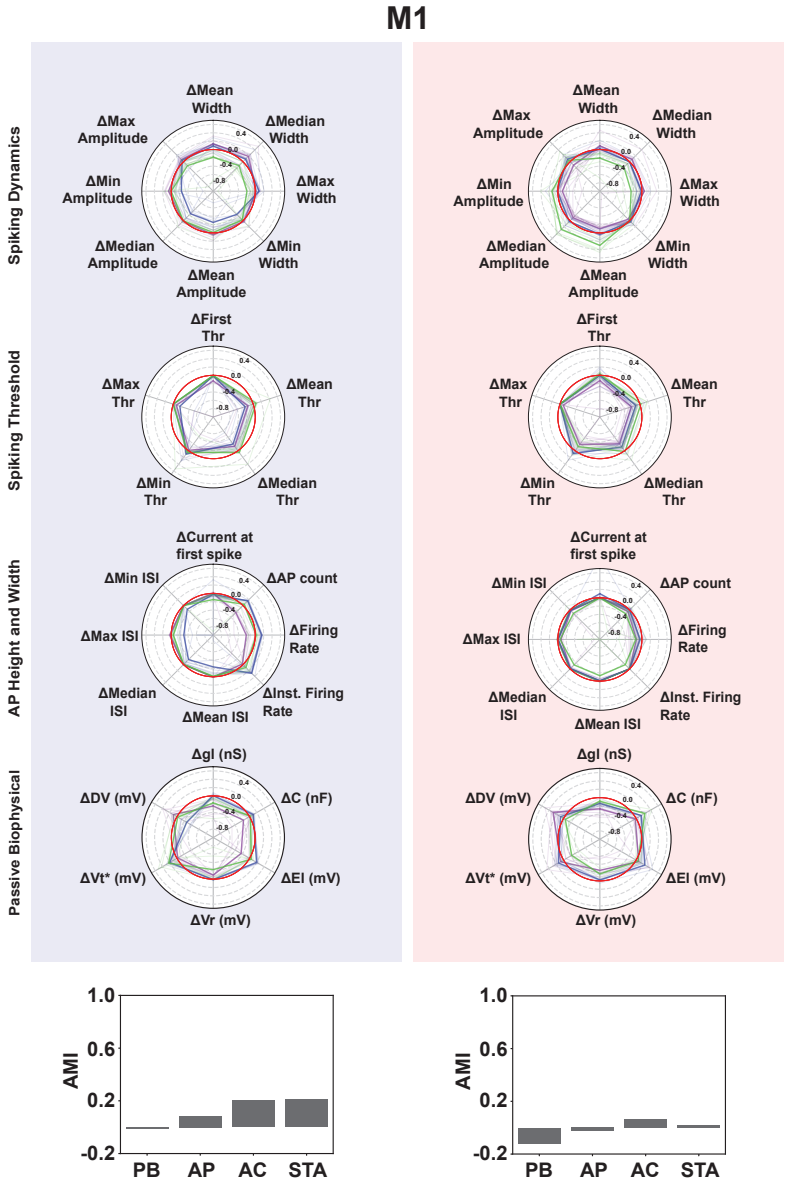

Figure 6.15. **Clustering based on differences between control and M1 for action potential and passive biophysical attributes reveals subgroups of neurons getting modulated differently as a result of M1-R activation.**

The polar plots show the clusters based on difference values between M1 and aCSF trials for action potential (subdivided into spiking dynamics, spiking threshold and AP height and width) and passive biophysical properties for both, excitatory (red background) and inhibitory (blue background). Each neuron is represented with a thin line and Coloured with their respective cluster label. The mean for each cluster is represented with a thick line. The histogram at the bottom shows the AMI score between cluster labels using aCSF trials and the cluster labels based on difference between M1 and aCSF trials.

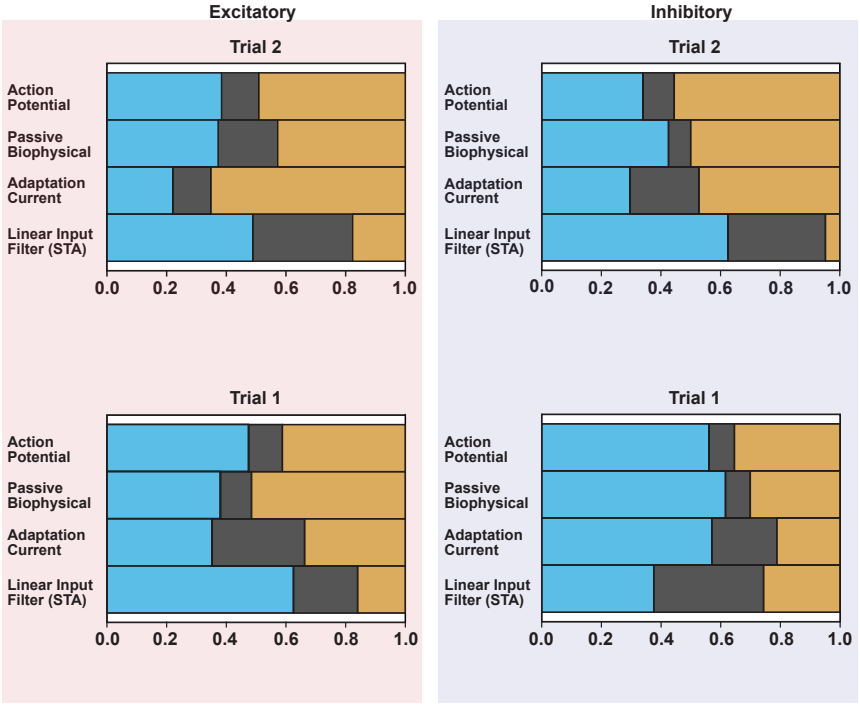

Figure 6.16. **MCFA for 1st and 2nd aCSF trials.**

Condition	Cell Type	P-value	Statistic	Cohen' s d
D1–aCSF	Inhibitory	0.019044	110	–0.2409
D1–aCSF	Excitatory	0.104543	84	0.5385
D2–aCSF	Inhibitory	0.768005	87	–0.0287
D2–aCSF	Excitatory	0.012016	225	0.3053
M1–aCSF	Inhibitory	0.523511	106	–0.1181
M1–aCSF	Excitatory	0.002022	22	0.4790
$aCSF_1$–$aCSF_2$	Inhibitory	0.205410	170	–0.3161
$aCSF_1$–$aCSF_2$	Excitatory	0.000111	340	0.3231

Table 6.1. Statistical comparison of fractional information (FI) changes under receptor activation.

Attribute	Variance type	aCSF	D1	D2	M1
AP	Private	16.7742 (+/- 11.11)	0.0062	0.1991	1.18E-03
AP	Shared	30.8543 (+/- 13.21)	41.0578	55.0319	66.0431
AP	Residual	52.3714 (+/- 12.18)	58.9360	44.7689	33.9556
PB	Private	18.5046 (+/- 13.50)	10.4084	14.4461	29.4316
PB	Shared	35.2320 (+/- 15.97)	49.5931	22.2190	26.3528
PB	Residual	46.2633 (+/- 15.84)	39.9985	63.2628	44.2155
AC	Private	18.0218 (+/- 17.14)	5.7216	16.8436	10.3262
AC	Shared	15.8369 (+/- 11.35)	6.9056	30.6685	35.0802
AC	Residual	66.1412 (+/- 21.22)	87.3728	44.7689	54.5935
STA	Private	57.7380 (+/- 24.75)	73.9663	71.2200	2.8344
STA	Shared	23.6072 (+/- 19.86)	9.5260	18.0047	6.9094
STA	Residual	18.6546 (+/- 18.49)	16.5077	10.7753	90.2561

Table 6.2. **Excitatory MCFA results**. Values that are reduced compared to bootstrapped aCSF are colored in red and values that are increased compared to aCSF are colored in blue.

Attribute	Variance type	aCSF	D1	D2	M1
AP	Private	14.9272 (+/- 09.5664)	0.8297	1.6361	6.3213
AP	Shared	33.3099 (+/- 12.0733)	32.3923	24.0248	46.7960
AP	Residual	51.7627 (+/- 10.2972)	66.7777	74.3390	46.8826
PB	Private	13.1228 (+/- 10.4672)	3.9880	11.2795	18.8554
PB	Shared	33.4015 (+/- 14.8031)	25.2733	53.6689	51.5773
PB	Residual	53.4755 (+/- 15.0643)	70.7389	35.0514	29.5671
AC	Private	22.7058 (+/- 16.5534)	11.6295	26.3594	50.6683
AC	Shared	26.9763 (+/- 14.6503)	44.4749	38.9088	22.8306
AC	Residual	50.3177 (+/- 21.2265)	43.8957	35.0514	26.5009
STA	Private	55.1458 (+/- 25.8562)	48.3895	21.5490	36.3625
STA	Shared	23.6595 (+/- 17.2281)	46.2345	66.6929	54.3020
STA	Residual	21.1946 (+/- 21.2975)	5.3761	11.7580	9.3353

Table 6.3. **Inhibitory MCFA results**. Values that are reduced compared to bootstrapped aCSF are colored in red and values that are increased compared to aCSF are colored in blue.

Attribute	Variance type	Variance 1st Trial	Variance 2nd Trial
AP	Private	11.2873 (+/- 11.81)	12.4735 (+/- 08.39)
AP	Shared	47.5009 (+/- 15.28)	38.4021 (+/- 10.44)
AP	Residual	41.2118 (+/- 10.25)	49.1244 (+/- 08.84)
PB	Private	10.3859 (+/- 08.94)	19.9091 (+/- 12.32)
PB	Shared	38.0565 (+/- 14.78)	37.3036 (+/- 14.24)
PB	Residual	51.5576 (+/- 18.05)	42.7873 (+/- 13.55)
AC	Private	31.0247 (+/- 16.07)	12.7529 (+/- 09.86)
AC	Shared	35.2139 (+/- 15.73)	22.0695 (+/- 11.62)
AC	Residual	33.7614 (+/- 14.71)	65.1776 (+/- 16.40)
STA	Private	21.4966 (+/- 16.40)	33.4581 (+/- 23.06)
STA	Shared	62.5138 (+/- 17.43)	48.8704 (+/- 25.72)
STA	Residual	15.9897 (+/- 08.43)	17.6716 (+/- 15.15)

Table 6.4. **Excitatory MCFA results for second aCSF trial**. Values that are reduced compared to bootstrapped aCSF trial 1 are colored in red and values that are increased compared to aCSF trial 1 are colored in blue.

Attribute	Variance type	Variance 1st Trial	Variance 2nd Trial
AP	Private	08.4805 (+/- 07.68)	10.5232 (+/- 07.13)
AP	Shared	56.1049 (+/- 11.24)	33.9479 (+/- 17.05)
AP	Residual	35.4145 (+/- 07.97)	55.5289 (+/- 18.73)
PB	Private	08.3097 (+/- 06.94)	07.4436 (+/- 09.55)
PB	Shared	61.6040 (+/- 12.22)	42.5329 (+/- 16.00)
PB	Residual	30.0864 (+/- 09.39)	50.0235 (+/- 15.53)
AC	Private	21.8111 (+/- 10.96)	23.1401 (+/- 15.18)
AC	Shared	56.9907 (+/- 12.05)	29.6210 (+/- 10.40)
AC	Residual	21.1982 (+/- 09.24)	47.2389 (+/- 15.30)
STA	Private	36.8157 (+/- 23.86)	32.6709 (+/- 21.86)
STA	Shared	37.5418 (+/- 17.49)	62.5230 (+/- 22.45)
STA	Residual	25.6425 (+/- 22.69)	04.806 (+/- 02.85)

Table 6.5. **Inhibitory MCFA results for second aCSF trial**. Values that are reduced compared to bootstrapped aCSF trial 1 are colored in red and values that are increased compared to aCSF trial 1 are colored in blue.

Chapter 7

Appendices

Nederlandse samenvatting

Research data management statement

Acknowledgements

Curriculum Vitae

Donders Graduate School

Nederlandse samenvatting

Belangrijkste bevindingen van dit proefschrift:

Research Data Management Statement

Ethics

FAIR principles

1. Findable

2. Accessible

3. Interoperable

4. Reusable

Privacy

Acknowledgements

Curriculum Vitae

Donders Graduate School for Cognitive Neuroscience

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g., Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit: http://www.ru.nl/donders/graduate-school/phd/