UNIVERSITY OF CALIFORNIA SAN DIEGO

The Topology of Neural Population Activity in the Songbird Auditory System

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Neurosciences

with a Specialization in Computational Neurosciences

by

Bradley Harold Theilman

Committee in charge:

Professor Timothy Gentner, Chair
Professor Tatyana Sharpee, Co-Chair
Professor Takaki Komiyama
Professor Saket Navlakha
Professor Gabriel Silva

2022

The dissertation of Bradley Harold Theilman is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

This dissertation is dedicated to:

My Mother, Regina.

My Aunt Linda and Uncle Tom.

My Grandparents, Harold and Margaret.

EPIGRAPH

Dem Höchsten Gott allein' zu Ehren,
Dem Nechsten, draus sich zu belehren.

-Johann Sebastian Bach, *Orgelbüchlein*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am extremely grateful to Professor Timothy Gentner for providing continuous encouragement and the opportunity to pursue the ideas presented in this dissertation. I could not have asked for a better mentor.

I would like to give special thanks to Ezequiel Arneodo. I could not imagine going through graduate school without our regular discussions about science, life, the world, and everything in between.

I would like to thank the members of the Gentner Lab, especially Michael, Krista, Marvin, Justin, Anna, Srihita, and Lorraine for all of their help and inspiration over the years. I have learned a tremendous amount from each of you.

I especially would like to acknowledge the members of my cohort in the Neurosciences Graduate Program, especially Barbara, Catie, Mark, Megan, and Peter, for some of the best memories that I could ever ask for.

Finally, I would like to acknowledge all of my teachers. Each and every one has taught me something that influenced this dissertation in some way.

Chapter 2, in full, is a reprint of the material as it appears in Spike Train Coactivity Encodes Learned Behavioral Invariances in Songbird Auditory Cortex, Theilman, B., Perks, K., and Gentner, T.Q., The Journal of Neuroscience, 2021. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full is currently being prepared for submission for publication and will include Professor Timothy Q. Gentner as the senior author and Michael Turvey and Srihita Rudraraju as co-authors. The dissertation author was the primary investigator and author of this material.

VITA

2014 Bachelor of Science, Biomedical Engineering and Mathematics
University of Cincinnati, Cincinnati, OH, USA

2022 Doctor of Philosophy, Neurosciences
with a Specialization in Computational Neurosciences
University of California San Diego, La Jolla, CA, USA

PUBLICATIONS

Theilman, B., Perks, K.E., and Gentner, T.Q. (2021) Spike Train Coactivity Encodes Learned Natural Stimulus Invariances in Songbird Auditory Cortex. Journal of Neuroscience 41:73-88.

Sainburg, T., Theilman, B., Thielk, M., Migliori, B., and Gentner, T.Q. (2019). Parallels in the
Sequential Organization of Birdsong and Human Speech. Nature Communications 10, 3636.

Bradley, T.D., Lewis, M., Master, J., Theilman, B. (2018). Translating and Evolving: Towards a Model of Language Change in DisCoCat. Electronic Proceedings in Theoretical Computer Science 283:50-61. https://arxiv.org/abs/1811.11041v1 (authors listed alphabetically)

Sainburg, T., Thielk, M., Theilman, B., Migliori, B., and Gentner, T.Q. (2018) Generative adversarial interpolative autoencoding: adversarial training on latent space interpolations encourage convex latent distributions. ArXiv preprint. https://arxiv.org/abs/1807.06650

Todd, J.B., Theilman, B., and Wendell, D.W. (2012). Detailed kinetic analysis of the Φ29 DNA packaging motor providing evidence for coordinated intersubunit ATPase activity of gp16. Virology. 432:370-675

ABSTRACT OF THE DISSERTATION

The Topology of Neural Population Activity in the Songbird Auditory System

by

Bradley Harold Theilman

Doctor of Philosophy in Neurosciences

with a Specialization in Computational Neurosciences

University of California San Diego, 2022

Professor Timothy Q. Gentner, Chair

Perception depends upon the coordinated activity of populations of neurons. How neural populations represent structure in the outside world through their activity is an important open question. One approach to this problem is the concept of the receptive field, which quantifies how external stimulation modulates the activity of individual neurons. While

receptive fields are a powerful concept for the experimenter, the brain itself does not have access to its own receptive fields. This dissertation applies methods from the field of algebraic topology to characterize the structure of neural population activity in the secondary auditory region NCM of the European starling. Chapter 2 demonstrates that the simplicial complex associated to population activity in NCM carries behaviorally-relevant information about learned categories. Along the way, a new similarity measure for population activity is developed, called the Simplicial Laplacian Spectral Entropy. It is shown that this measure quantifies the similarity of simplicial complexes associated with neural activity in a way that depends upon their global topological structure. Chapter 3 explores the connection between neural topology and classical receptive fields, by showing that the intrinsic geometry of the population activity matches the geometry of the receptive fields. This shows that the temporal coativity structure of the population contains a direct representation of the stimulus structure without requiring explicit computation of receptive fields. This validates a previously described theoretical mechanism in a sensory system for the first time *in-vivo*, and provides a new understanding for population activity in sensory regions. This chapter also introduces a technique for reconstructing acoustic spectrograms of complex, natural vocal signals from neural activity, which will be a powerful technique for exploring population level representations in future studies. The final chapter of this dissertation describes a new mathematical description of the phenomenon of polychronization in spiking neural networks. This serves as a bridge between the experimental results of this dissertation and a theoretical understanding of the emergence of spatiotemporal structure in the activity of neural populations.

Chapter 1 - Introduction

Perception is the capability of an organism to extract, through its senses, information about an environment that is useful for the organism's behavior and survival. For the purposes of this thesis, an animal is said to perceive a feature if it is capable of making behavioral distinctions that depend on knowledge of that feature. Perception depends on the organism's nervous system, since perturbations to the nervous system yield perturbations in perception. A major goal of neuroscience is to understand how perception arises from the organization and activity of the nervous system. What is an understanding of perception? There must be four components to any understanding of perception:

- What in the environment constitutes information for the animal?
- How the organism's senses transduce that information into nervous system activity.
- How changes in the environment correspond to changes in the organism, in particular, to changes in the activity in the organism's nervous system.
- How these nervous system activity changes influence the organism's behavior.

There are two high-level approaches to understanding perception: direct and indirect. Each approach agrees on the need to explain perception in terms of nervous system activity, but they differ strongly in their underlying conceptual frameworks. The conceptual tension between these approaches has led to an extensive debate over many decades. This debate is important because it informs the underlying assumptions behind every experiment in perceptual neuroscience, and each side has far-reaching implications for our understanding of nervous systems as a whole. Current trends in perceptual neuroscience favor approaches falling under indirect perception, but the debate is far from settled. Advances in experimental and theoretical neuroscience mean that the conceptual distinctions that differentiate these approaches are now

subject to empirical tests. We begin with an overview of indirect and direct perception, and then discuss their implications for neuroscience research.

Indirect Perception

Indirect perception began with a consideration of the phenomenological aspects of perception compared to the burgeoning scientific understanding of sense organs. As science was evolving to place emphasis on empirically measured quantities and the relationships between them, scientists began describing the physical structure and operation of sense organs with the same "objective" terminology used to describe the features of inanimate objects in the world. From this emerged the consensus that the only relevant description of stimulation for an animal's perception was based on physically measurable quantities. Perceptual qualities, such as color or timbre, became secondary to primary qualities, such as wavelength and amplitude.

Describing stimuli purely with physical quantities revealed a disparity between stimulation and perception. Namely, there are features to percepts that have no direct manifestation in the physical variables. A prototypical example of this is three-dimensional depth perception. A physical description of light impinging on the retina consists of a specification of a "retinal image" - a two dimensional surface with a function describing the wavelength and amplitude of light hitting each point on the surface.

Indirect perception starts from a set of assumptions motivated by dissimilarities between the nature of sensations and the nature of our subjective knowledge of the environment. Apparently, our knowledge of the environment contains details that are not immediately present in the underlying sensations. Stimuli, the physical antecedent to sensations, are described as would a physicist describe them, consisting of physical quantities such as wavelength, intensity,

or frequency. Animals have the capability to make distinctions between (that is, produce alternative behaviors depending on) features of the environment that are not present in the physical description of the stimulation informing these distinctions.

A prototypical example is three-dimensional depth perception. Animals, including humans, can distinguish the positions of objects in a three dimensional space. The stimulation informing these distinctions is the pattern of light falling on the retina: without such stimulation, this distinction cannot be made, and this is the level closest to the sensory epithelia that such a dependence manifests. The physical description of light on the retina consists of a map of wavelengths and intensities of light on a two-dimensional surface. There is nothing in this physical description that directly corresponds to the x,y, and z coordinates of the objects. Thus, perception of the objects' positions (i.e. knowledge that informs this distinction) is indirectly related to the stimulation. This implies that some process must elaborate (Marr, 2010; Michaels & Carello, 1981) the stimulation to produce the perception of position in three-dimensional space .

Indirect approaches treat perceptions as discrete products of computational processes that begin with the senses. This forces the question of the relation between perception and neural activity to factor into three parts, most famously articulated by David Marr:

- What is the computation that must be performed on the sensory inputs to yield perceptions?
- What is the algorithm that accomplishes this computation?
- How is this algorithm implemented in neural systems?

Thus, indirect approaches require a strict, sequential information-processing point of view because they assume that raw stimulation transduced by the sensory epithelia does not contain

enough information on its own to account for the richness of information present in the resultant perceptions (Bickhard & Richie, 1983).

Direct Perception

Direct perception begins with the observation that each animal occupies a particular niche in its environment. This niche was not present a priori, but came to be just as the animal itself did: through evolution. Since the animal's niche contains the most relevant aspects of its environment, on evolutionary timescales, the animal determines its environment as the environment determines the animal (Corris, 2020). Direct perception postulates that this intertwining of animal and environment exists at the level of a single individual. Since perception is the acquisition of knowledge about the environment, this postulate has strong implications for the nature of perception.

Both the animal and environment are dynamic entities. If the animal and environment mutually determine each other, any knowledge about the environment the animal has must manifest over a continuous span of time. This means that, fundamentally, perception cannot be considered as a series of static percepts. Instead, knowledge of the environment manifests as systems of relationships that hold between the animal and the environment, over varying spans of time. Since perception is the acquisition of this information, direct perception requires the animal to be sensitive to these relationships.

Indirect/Direct perception and neuroscience

Acquiring an understanding of the nervous system's role in perception looks different under indirect and direct perception. Indirect perception's requirement of successive processing

stages maps well onto the observed hierarchy of sensory structures in most nervous systems (Riesenhuber & Poggio, 1999; Yamins et al., 2014). The task for neuroscientists is to identify which perceptual computations correspond to each level of the hierarchy and which representations exist at the intermediate steps. To do this, neuroscientists take advantage of the fact that neural activity is correlated with variables in the world outside the animal that the experimenter themself can measure. Because, according to indirect perception, such physical variables are the only relevant specification for the stimulus, for the system to perceive, the experimenter must be able to extract or "decode" these physical variables from the neural activity. Thus, a significant fraction of the work in sensory systems neuroscience is involved in establishing maps (i.e. models) between external physical variables and neural activity. There are a variety of approaches to accomplish this, but most can be described as single neuron/population and statistical/dynamical.

Single neuron, statistical models attempt to define the map as a statistical relationship between two random variables: the external physical variable(s) and the neuron's activity. The parameters for such models can be found by matching the statistical correlations between the experimentally observed neural response to a distribution of physical stimuli. With these parameters, a neuron is said to represent a physical variable if changes in that variable correspond to changes in the probability of that neuron to produce a spike or increase its firing rate.

Population statistical models operate similarly to the single neuron case. An experimenter seeks a statistical model of the response of multiple neurons in such a way that correlation between the model-predicted activity distribution and the stimulus distribution match. In the population case, activity from multiple cells can be thought of as a point in a high

dimensional “neural activity space”; postulated to be a linear space in which dimensions correspond to the activity of individual cells. This “extra room” allows for more complex interpretations of the modulations in neural activity due to changes in stimulus variables, leading to ideas such as mixed selectivity (Fusi et al., 2016; Rigotti et al., 2013). Purely statistical models of population activity suffer from the curse of dimensionality, meaning that the data required to estimate model parameters grows exponentially with the number of neurons in the population. This places severe limits on the order of statistical correlations that can be adequately constrained by the available data (Fitzgerald et al., 2011; Ganguli & Sompolinsky, 2012; Sharpee, 2013). It may be that higher order correlations are not even necessary, as pairwise correlations can capture a surprising fraction of the variance (Barreiro et al., 2014; Leen & Shea-Brown, 2015; Schneidman et al., 2006).

Recent technological advances in electrophysiology allow for recording the activity of many neurons simultaneously. In these large populations of cells, it is commonly found that a significant fraction of the variance in the activity across the population can be accounted for by far fewer dimensions than the total number of cells. This leads to the slogan that neural activity is low-dimensional, and forms the motivation for a large class of models characterized as population dynamical models. The idea is that neural activity can be described as a noisy realization of a low-dimensional dynamical system. Borrowing the concept of neural activity space from the preceding discussion, this dynamical system manifests as a lower-dimensional subset (a ‘submanifold’) of the higher-dimensional neural space (Cunningham & Yu, 2014; Gallego et al., 2017). Again, the point of these models is to provide maps between neural activity and physical dimensions of the external world. The purpose of the geometric description of neural activity is to again provide “extra room” to define and classify these maps. Once

neural activity is cast into this high dimensional space, such maps may take the form of linear decoders or other functions operating on the trajectories of the dynamical system (Aghagolzadeh & Truccolo, 2014; Aoi et al., 2020; Mante et al., 2013; Vyas et al., 2020).

More recent approaches combine the ideas of statistical and dynamical population models (Pandarinath et al., 2018; Yu et al., 2009). The goal with these approaches is to account for the bulk modulations of neural population activity by using a low-dimensional dynamical system. On top of this dynamical model, a stochastic process introduces variability in such a way to capture the residual statistical relationships between neurons and between neurons and external variables.

In all these cases, the point is to define a space to serve as the domain of some map between neural activity and external physical variables. If such a map exists and adequately preserves the correlations between neural activity and external variables, then the neuron or population is said to represent the variable that is specified by the map. Following the logic of indirect perception, these representations serve as intermediate steps between the computations that elaborate sensations into perceptions.

Neuroscientists have found exquisite and robust examples of maps between neural activity and external variables (Hafting et al., 2005; Kim et al., 2019). Nevertheless, the underlying logic of this avenue of investigation has structural issues that limit its ability to provide a complete understanding of the neural basis of perception.

Most importantly, the maps sought by neuroscientists that define the representations are entities that exist only for the experimenter. This is because only the experimenter has simultaneous access to the neural activity and the external physical variables necessary to define the map (Brette, 2018). There is no way for the nervous system to know its own mapping. Such

maps may (and in fact do) exist, but the nervous system itself cannot make use of them. This is also demonstrated by the nature of physical measurement underlying the physical variables that indirect perception postulates as the substrate of perception: in all cases, the value of a physical variable is established by bringing some other object (e.g. a ruler) in contact with the system under test, and the experimenter, who has access to both the system under test and the ruler, reads the correlation between the two. The measurement of a physical variable requires an experimenter external to both system and ruler. Since the organism's nervous system cannot serve the roles of ruler and observer at the same time, physical variables cannot be the substrate for perception. This demonstration does not contradict the existence of robust maps between neural activity and external physical variables, but shows that an understanding of the neural basis of perception cannot be built on representations of those variables and computations on these representations (Bickhard & Richie, 1983). By postulating information that an animal cannot have access to, any understanding of perception based on the assumptions of indirect perception falls into contradiction in answering the question, "What in the environment constitutes information for the animal?" To fix this, neuroscientists must carefully consider what information is available to the animal (Brette, 2013).

Sensory systems neuroscience research under the assumptions of direct perception is comparatively impoverished with respect to research under indirect approaches. Direct perception recognizes that since the animal is embedded in a continuous stream of interaction with the external world, the animal defines its environment and the environment defines the animal (Corris, 2020). Since perception is the acquisition of knowledge about the environment, this implies that the knowledge about the environment consists of knowledge about regularities in the animal-environment interaction. In contrast to indirect perception, these regularities, or

relationships, are directly available to the animal; hence the term direct perception. Such relationships are features of the animal environment interaction that remain stable or invariant over periods of time. These invariants lie at the level of the interaction between animal and environment, and do not necessarily correspond to features in the animal or environment considered separately. A classic example is optic flow (Gibson, 2014). When an animal is moving in a fixed direction, its visual experience consists of constantly changing patterns of light falling on the retina: there are no static features. Nevertheless, there is an invariant structure in this continuous change. The flow of patterns on the retina specifies a unique point in the visual field, and this point corresponds to the direction in which the animal is heading. By detecting this optic flow, the animal in principle knows its heading exactly.. The flow is specified directly by the sensory information, and does not require elaborating this information into a representation of 3-dimensional space and then computing a heading vector from this representation.

Direct perception claims that this example generalizes to more abstract relationships between the animal and environment. The set of all possible sensory/motor interactions is constrained by the natural laws of physics and specifies all of the knowledge an animal can know about its environment. Thus, the aim of research in sensory neuroscience under direct perception is to identify these relationships and the neural mechanisms for their detection. Because the relationships sought manifest invariant structure in the environment/animal interaction, one strategy for their identification is to search for invariants in neural activity.

Neural invariants through Topology

Requiring that neural invariants be specified by information that is entirely internal to the animal places strong constraints on the forms they can take. A powerful approach to this problem was put forward by Curto and Itskov (Curto & Itskov, 2008). In their highly original work, they reasoned that in the hippocampal place cell system, the place fields that tiled a physical environment were analogous to the concept of open covers, a central concept in the mathematical field of topology. This insight led directly to their definition of a novel method for capturing invariants in population spiking activity. By interpreting *temporal coactivation* as instructions for constructing a mathematical object called a simplicial complex, they were able to show that the temporal coactivation structure in a population manifests topological invariants in this simplicial complex. What's more, this structure does not depend on any information outside of the animal, since it is defined by the relative spike times of neurons in the population.

Curto and Itskov showed that, in the case of simulated hippocampal place cell activity, the information contained in the simplicial complex was sufficient to reconstruct the physical environment the simulated animal explored. This powerful result opened up the possibility of relating concepts from indirect and direct perception through neural physiology. From the direct perception side, the simplicial complexes can be thought of as manifestations of relationships between the receptive fields of individual neurons. From the direct perception side, the simplicial complex can be thought of as manifestations of network constraints that produce specific temporal coactivation patterns through the dynamic interaction between the animal and its environment. By expressing both sides in a commensurate mathematical language, the Curto-Itskov construction provides a point of contact between the disparate approaches to perception, and allows for the relative merits of each approach to be weighed empirically.

While interesting in its own right, their original work was restricted to a simulated case, namely, 2-dimensional flat physical environments. In this case the relevant spatial variables and structure are known precisely and are analytically tractable. One of the advantages of the Curto-Itskov construction is that it provides a way to *define* an abstract "sensory space" when an appropriate space is unknown a priori. A sensory space is defined by the relevant features (dimensions) along which variation produces perceptible changes, and the relations between these features (their geometry). Receptive field techniques find relevant sensory features, but their geometry is constrained by a priori specification of the receptive field model. Additionally, as mentioned previously, receptive fields are quantities external to the animal. The Curto-Itskov approach reconstructs the system of relationships between features (through the topology of the simplicial complex) directly from neural activity, without a priori assumptions on the geometry, but does not associate these relata with stimulus features. If the geometry defined by the receptive fields and by the simplicial complex correspond, it shows that the receptive fields are not necessary, because every question one could ask about the receptive field features that is meaningful for the animal is contained in the system of relationships recovered by the simplicial complex.

The central insight of their original paper, that topological (i.e. relational) structure in spaces of natural stimuli is present in the combinatorial structure of neural population activity, should apply to any sensory system. It is the central aim of this thesis to show that the ideas of Curto-Itskov (2008) are applicable to higher-order sensory systems and provide valuable insight into the neural underpinnings of higher order perception.

To accomplish this, the bulk of this thesis rests on two projects. The first is a demonstration that the abstract sensory space defined by the topology of the simplicial complex

associated to population spiking activity contains invariants that are relevant to behavior. The second is a demonstration that, when receptive fields of higher order sensory neurons are definable, the interpretation of the simplicial complex as manifesting relationships between receptive fields is appropriate. Similar to the original Curto and Itskov work, this is demonstrated through reconstructions of natural stimuli, and showing that the metric relationships that enable these reconstructions are obtainable directly from the population.

Songbird auditory system

The specific perceptual system studied in this thesis is the songbird auditory system. The European starling (*Sturnus vulgars)* is a species of songbird that, like other songbirds, is a vocal learner that uses complex, extended acoustic signals (songs) as a part of its natural behavior (Eens, 1997). Starlings are capable of fine perceptual distinctions in their song (null Gentner & Hulse, 1998; T. Q. Gentner & Hulse, 2000), indicating that their cortical auditory pathways are adapted for representing complex acoustic information.

The anatomy of the songbird auditory systems shares several analogies with mammalian auditory cortices. In particular, the system is arranged hierarchically. In the songbird, auditory input arrives at the forebrain in a region called Field L (Vates et al., 1996). Regions NCM and CMM/CLM receive projections from field L and are anatomically higher order sensory regions, analogous to the superior temporal gyrus in mammals (Bolhuis et al., 2010). This characterization is supported by their functional properties. In particular, NCM is known to be involved in song memory, song recognition, and learning (Bailey et al., 2002; Chew et al., 1996; George et al., 2008; Thompson & Gentner, 2010)

Due to its anatomical relationships and functional properties, NCM is a prime target for identifying neural mechanisms underlying the perception of higher-order features in natural acoustic signals.

The songbird auditory system is also appealing for contrasting indirect and direct perception. Acoustic signals are, by nature, temporally extended. There is no acoustic analog of a static image, a commonly used idealization in vision research upon which many results favoring indirect perception rest. Perceptually relevant acoustic modulations happen at multiple timescales (Narayan et al., 2006). Furthermore, auditory responses are often seen to be temporally precise (Sen et al., 2001; Wehr & Zador, 2003). This challenges indirect approaches, because for these modulations to affect behavior, the neural dynamics underlying this perception must occur on similar timescales. This places constraints on algorithmic processes that are postulated to produce these perceptions.

Polychronization

The two main chapters of this thesis show that neural activity is structured at the population level. This structure must emerge from the physiological structure of the network that produces this activity. Currently, starlings lack comparable genetic and molecular techniques for teasing apart fine detail in the structure of the avian connectome. This limits the amount that can be known concerning the biophysics that produces the organized population activity patterns seen in these studies. Nevertheless, judicious theoretical considerations combined with what is known about avian networks yields promising hypotheses about the emergence of structure population activity. The final chapter of this thesis describes a

mathematical approach to understanding the phenomenon of polychronization in spiking networks.

Polychronization (Izhikevich, 2006) is the appearance of self-reinforcing time-locked but asynchronous patterns of spiking activity distributed across a neural population. The phenomenon emerges from two fundamental observations of cortical networks. First, for most cortical cells, individual presynaptic spikes are not sufficient to trigger postsynaptic action potentials *in-vivo*. However, fine temporal coincidences in presynpatic input can significantly modulate postsynaptic spiking (Rossant et al., 2011). Despite the observation of significant variability in neural responses (Shadlen & Newsome, 1994; Softky & Koch, 1993), neural spike generation under physiological conditions is remarkably temporally precise (Mainen & Sejnowski, 1995), especially in systems where the relevant temporal modulations occur on fast timescales (Bair & Koch, 1996; Bale et al., 2015; Kayser et al., 2010; Mackevicius et al., 2012; Reinagel & Reid, 2002). Second, there is a non-zero temporal propagation delay along axons, even when presynaptic and postsynaptic partners are anatomically located in the same brain region (Egger et al., 2019; Ferraina et al., 2002; Swadlow, 1990). What's more, this delay is heterogeneous and varies between presynaptic partners of a given postsynaptic neuron. The effect of this delay is to force temporal coincidence to be a *relative* notion that depends on the neuron. For example, two spikes may arrive coincidentally to postysnaptic neuron A but arrive with a temporal offset to neuron B, due to the heterogeneous delays. Spike patterns that conform to these temporal offsets form self-reinforcing patterns of activity, called *polychronous groups.* Because of their discreteness, these groups have been discussed as a substrate for neural computation (Brette, 2015; Izhikevich, 2006).

Polychronization requires few biological commitments beyond long-established and repeatable experimental observations, and so is an eminently plausible component of neural computation. Recently, next-to-direct evidence for polychronization was found in the avian song production pathway (Egger et al., 2019). Furthermore, the results of the two main studies in this thesis square nicely with a view of neural population activity based on polychronization, because temporally-precise, information-containing distributed patterns of spiking activity are central to both these studies and the concept of polychronization. Thus, polychronization is a promising approach to relate structure population activity to the biophysics of the spiking network.

One of the advantages of polychronization as a mechanism for neural computation is that relatively small networks have the potential to produce exponentially large numbers of polychronous groups. This is advantageous for computation but makes systematic study of polychronization difficult. The final chapter of this thesis develops a mathematical, compositional approach to polychronization describing how larger polychronous groups are built from smaller polychronous groups. This provides a tentative step in the theory of polychronization beyond the brute force approaches for studying polychronization in spiking networks (Chrol-Cannon et al., 2017; Hereld et al., 2014) to quantitative theories linking network structure and neural topology.

This thesis

Taken together, the chapters of this thesis demonstrate that the biological spiking networks that underlie sensory perception organize their activity into temporal patterns at the population level. This organization is not arbitrary, but manifests specific topological and geometric structure that coincides with structure in the outside world. This provides a

mechanism for the direct representation of stimulus structure by spiking networks, without the conceptual commitments of indirect perception and their associated limitations. Along the way, several new techniques are introduced for analyzing population spiking activity. The first is a new similarity measure for population spike trains based on the similarity of their associated simplicial complexes. This measure takes into account the temporal organization of activity across the whole population and across time. The second is a technique for reconstructing acoustic spectrograms from neural activity. Unlike previous methods for reconstructing spectrographic stimuli, this method does not require knowledge of the stimulus autocorrelation (Mesgarani et al., 2009). The final chapter of this thesis introduces a new mathematical description of polychronization that centers on the compositional nature of polychronous groups. This connects with the previous chapters by providing a route to a plausible explanation of the origin of the temporal patterns characterized by the first two studies.

References

Aghagolzadeh, M., & Truccolo, W. (2014). Latent State-Space Models for Neural Decoding. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2014*, 3033–3036. https://doi.org/10.1109/EMBC.2014.6944262

Aoi, M. C., Mante, V., & Pillow, J. W. (2020). Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nature Neuroscience*, *23*(11), 1410–1420. https://doi.org/10.1038/s41593-020-0696-5

Bailey, D. J., Rosebush, J. C., & Wade, J. (2002). The hippocampus and caudomedial neostriatum show selective responsiveness to conspecific song in the female zebra finch. *Journal of Neurobiology*, *52*(1), 43–51. https://doi.org/10.1002/neu.10070

Bair, W., & Koch, C. (1996). Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural Computation*, *8*(6), 1185–1202. https://doi.org/10.1162/neco.1996.8.6.1185

Bale, M. R., Campagner, D., Erskine, A., & Petersen, R. S. (2015). Microsecond-Scale Timing Precision in Rodent Trigeminal Primary Afferents. *Journal of Neuroscience*, *35*(15), 5935–5940. https://doi.org/10.1523/JNEUROSCI.3876-14.2015

Barreiro, A., Gjorgjieva, J., Rieke, F., & Shea-Brown, E. (2014). When do microcircuits produce beyond-pairwise correlations? *Frontiers in Computational Neuroscience*, *8*, 10. https://doi.org/10.3389/fncom.2014.00010

Bickhard, M. H., & Richie, D. M. (1983). *On The Nature Of Representation: A Case Study Of James Gibson's Theory Of Perception*. Ny: Praeger.

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews. Neuroscience*, *11*(11), 747–759. https://doi.org/10.1038/nrn2931

Brette, R. (2013). Subjective physics. *ArXiv:1311.3129 [q-Bio]*. http://arxiv.org/abs/1311.3129

Brette, R. (2015). Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00151

Brette, R. (2018). Is coding a relevant metaphor for the brain? *The Behavioral and Brain Sciences*, *42*, e215. https://doi.org/10.1017/S0140525X19000049

Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proceedings of the National Academy of Sciences*, *93*(5), 1950–1955. https://doi.org/10.1073/pnas.93.5.1950

Chrol-Cannon, J., Jin, Y., & Grüning, A. (2017). An Efficient Method for online Detection of Polychronous Patterns in Spiking Neural Network. *Neurocomputing*, *267*, 644–650. https://doi.org/10.1016/j.neucom.2017.06.025

Corris, A. (2020). Defining the Environment in Organism–Environment Systems. *Frontiers in Psychology*, *11*, 1285. https://doi.org/10.3389/fpsyg.2020.01285

Cunningham, J. P., & Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, *17*(11), 1500–1509. https://doi.org/10.1038/nn.3776

Curto, C., & Itskov, V. (2008). Cell Groups Reveal Structure of Stimulus Space. *PLOS Computational Biology*, *4*(10), e1000205. https://doi.org/10.1371/journal.pcbi.1000205

Eens, M. (1997). *Understanding the complex song of the European starling: An integrated ethological approach*. https://doi.org/10.1016/S0065-3454(08)60384-8

Egger, R., Tupikov, Y., Katlowitz, K. A., Benezra, S. E., Picardo, M. A., Moll, F., Kornfeld, J., Jin, D. Z., & Long, M. A. (2019). Local axonal conduction delays underlie precise timing of a neural sequence. *BioRxiv*, 864231. https://doi.org/10.1101/864231

Ferraina, S., Paré, M., & Wurtz, R. H. (2002). Comparison of cortico-cortical and cortico-collicular signals for the generation of saccadic eye movements. *Journal of Neurophysiology*, *87*(2), 845–858. https://doi.org/10.1152/jn.00317.2001

Fitzgerald, J. D., Sincich, L. C., & Sharpee, T. O. (2011). Minimal Models of Multidimensional Computations. *PLOS Computational Biology*, *7*(3), e1001111. https://doi.org/10.1371/journal.pcbi.1001111

Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, *37*, 66–74. https://doi.org/10.1016/j.conb.2016.01.010

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025

Ganguli, S., & Sompolinsky, H. (2012). Compressed Sensing, Sparsity, and Dimensionality in Neuronal Information Processing and Data Analysis. *Annual Review of Neuroscience*, *35*(1), 485–508. https://doi.org/10.1146/annurev-neuro-062111-150410

Gentner, null, & Hulse, null. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris. *Animal Behaviour*, *56*(3), 579–594. https://doi.org/10.1006/anbe.1998.0810

Gentner, T. Q., & Hulse, S. H. (2000). Perceptual classification based on the component structure of song in European starlings. *The Journal of the Acoustical Society of America*, *107*(6), 3369–3381. https://doi.org/10.1121/1.429408

George, I., Cousillas, H., Richard, J.-P., & Hausberger, M. (2008). A Potential Neural Substrate for Processing Functional Classes of Complex Acoustic Signals. *PLOS ONE*, *3*(5), e2203. https://doi.org/10.1371/journal.pone.0002203

Gibson, J. J. (2014). *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press. https://doi.org/10.4324/9781315740218

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. https://doi.org/10.1038/nature03721

Hereld, M., Suresh, J., Radojicic, M., Pesce, L. L., & van Drongelen, W. (2014). Connectivity from spike trains of neocortex neuron populations. *BMC Neuroscience*, *15*(1), P18. https://doi.org/10.1186/1471-2202-15-S1-P18

Izhikevich, E. M. (2006). Polychronization: Computation with Spikes. *Neural Computation*, *18*(2), 245–282. https://doi.org/10.1162/089976606775093882

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Millisecond encoding precision of auditory cortex neurons. *Proceedings of the National Academy of Sciences*, *107*(39), 16976–16981. https://doi.org/10.1073/pnas.1012656107

Kim, S. S., Hermundstad, A. M., Romani, S., Abbott, L. F., & Jayaraman, V. (2019). Generation of stable heading representations in diverse visual scenes. *Nature*, *576*(7785), 126–131. https://doi.org/10.1038/s41586-019-1767-1

Leen, D. A., & Shea-Brown, E. (2015). A Simple Mechanism for Beyond-Pairwise Correlations in Integrate-and-Fire Neurons. *The Journal of Mathematical Neuroscience (JMN)*, *5*(1), 17. https://doi.org/10.1186/s13408-015-0030-9

Mackevicius, E. L., Best, M. D., Saal, H. P., & Bensmaia, S. J. (2012). Millisecond Precision Spike Timing Shapes Tactile Perception. *The Journal of Neuroscience*, *32*(44), 15309–15317. https://doi.org/10.1523/JNEUROSCI.2161-12.2012

Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science (New York, N.Y.)*, *268*(5216), 1503–1506. https://doi.org/10.1126/science.7770778

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

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

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex. *Journal of Neurophysiology*, *102*(6), 3329–3339. https://doi.org/10.1152/jn.91128.2008

Michaels, C. F., & Carello, C. (1981). *Direct perception*. Prentice-Hall.

Narayan, R., Graña, G., & Sen, K. (2006). Distinct time scales in cortical discrimination of natural sounds in songbirds. *Journal of Neurophysiology*, *96*(1), 252–258. https://doi.org/10.1152/jn.01257.2005

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., Henderson, J. M., Shenoy, K. V., Abbott, L. F., & Sussillo, D. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, *15*(10), 805–815. https://doi.org/10.1038/s41592-018-0109-9

Reinagel, P., & Reid, R. C. (2002). Precise firing events are conserved across neurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *22*(16), 6837–6841. https://doi.org/20026633

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. https://doi.org/10.1038/14819

Rigotti, M., Barak, O., Warden, M. R., Wang, X.-J., Daw, N. D., Miller, E. K., & Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. *Nature*, *497*(7451), 585–590. https://doi.org/10.1038/nature12160

Rossant, C., Leijon, S., Magnusson, A. K., & Brette, R. (2011). Sensitivity of Noisy Neurons to Coincident Inputs. *Journal of Neuroscience*, *31*(47), 17193–17206. https://doi.org/10.1523/JNEUROSCI.2482-11.2011

Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*(7087), 1007–1012. https://doi.org/10.1038/nature04701

Sen, K., Theunissen, F. E., & Doupe, A. J. (2001). Feature Analysis of Natural Sounds in the Songbird Auditory Forebrain. *Journal of Neurophysiology*, *86*(3), 1445–1458. https://doi.org/10.1152/jn.2001.86.3.1445

Shadlen, M. N., & Newsome, W. T. (1994). Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, *4*(4), 569–579. https://doi.org/10.1016/0959-4388(94)90059-0

Sharpee, T. O. (2013). Computational Identification of Receptive Fields. *Annual Review of Neuroscience*, *36*(1), 103–120. https://doi.org/10.1146/annurev-neuro-062012-170253

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, *13*(1), 334–350. https://doi.org/10.1523/JNEUROSCI.13-01-00334.1993

Swadlow, H. A. (1990). Efferent neurons and suspected interneurons in S-1 forelimb representation of the awake rabbit: Receptive fields and axonal properties. *Journal of Neurophysiology*, *63*(6), 1477–1498. https://doi.org/10.1152/jn.1990.63.6.1477

Thompson, J. V., & Gentner, T. Q. (2010). Song Recognition Learning and Stimulus-Specific Weakening of Neural Responses in the Avian Auditory Forebrain. *Journal of Neurophysiology*, *103*(4), 1785–1797. https://doi.org/10.1152/jn.00885.2009

Vates, G. E., Broome, B. M., Mello, C. V., & Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *The Journal of Comparative Neurology*, *366*(4), 613–642. https://doi.org/10.1002/(SICI)1096-9861(19960318)366:4<613::AID-CNE5>3.0.CO;2-7

Vyas, S., Golub, M. D., Sussillo, D., & Shenoy, K. V. (2020). Computation Through Neural Population Dynamics. *Annual Review of Neuroscience*, *43*(1), 249–275. https://doi.org/10.1146/annurev-neuro-092619-094115

Wehr, M., & Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, *426*(6965), 442–446. https://doi.org/10.1038/nature02116

Yamins, D. L. K., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., & Sahani, M. (2009). Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *Journal of Neurophysiology*, *102*(1), 614–635. https://doi.org/10.1152/jn.90941.2008

Chapter 2 – Spike Train Coactivity Encodes Learned Natural Stimulus Invariances in Songbird Auditory Cortex

Abstract

The capacity for sensory systems to encode relevant information that is invariant to many stimulus changes is central to normal, real-world, cognitive function. This invariance is thought to be reflected in the complex spatiotemporal activity patterns of neural populations, but our understanding of population-level representational invariance remains coarse. Applied topology is a promising tool to discover invariant structure in large datasets. Here, we use topological techniques to characterize and compare the spatiotemporal pattern of coactive spiking within populations of simultaneously recorded neurons in the secondary auditory region NCM of European starlings (*Sturnus vulgaris*). We show that the pattern of population spike train coactivity carries stimulus specific structure that is not reducible to that of individual neurons. We then introduce a topology-based similarity measure for population coactivity that is sensitive to invariant stimulus structure and show that this measure captures invariant neural representations tied to the learned relationships between natural vocalizations. This demonstrates one mechanism whereby emergent stimulus properties can be encoded in population activity and shows the potential of applied topology for understanding invariant representations in neural populations.

Significance Statement

Information in neural populations is carried by the temporal patterns of spikes. We applied novel mathematical tools from the field of algebraic topology to quantify the structure of these temporal patterns. We found that in a secondary auditory region of a songbird, these

patterns reflected invariant information about a learned stimulus relationship. These results demonstrate that topology provides a novel approach for characterizing neural responses that is sensitive to invariant relationships that are critical for the perception of natural stimuli.

Introduction

Functional sensory systems must express a degree of representational invariance in order to convey relationships or abstract properties, such as category membership, between otherwise distinguishable stimuli. At the neural level, invariant representations may manifest as robustness to "noise", or consistency across trials or learned stimuli in a single animal, or across animals. For high-dimensional natural signals in particular, whose representation is distributed, invariant representations likely involve the coordinated spiking activity of neuronal populations across time. The nature of this coordination and how it supports invariance is an important open question in neuroscience.

Most current methods for quantifying neuronal population structure rely on converting spiking activity into a vector of instantaneous firing rates (Cunningham et al., 2009; Churchland et al., 2012; Gallego et al., 2018). Over time, these vectors trace out a path in a high-dimensional space. The dynamics of this path can be correlated with simultaneous information about the stimulus or behavior. However, these methods require averaging activity over time or space to overcome putative noise in single neuron firing rates. This is problematic for populations that must represent intrinsically time-varying objects, such as vocal communication signals, on fast time scales. Auditory populations that do so often exhibit lifetime-sparse firing, sometimes complicating the estimation of instantaneous firing rates. Thus, while firing rate vectors may be decodable in some cases (Ince et al., 2013; Rigotti et al., 2013), they belie the near instantaneous and continuous nature of natural perception.

Alternatively, information may be carried in patterns of coincident spiking, or a "coactiviation code" (Curto and Itskov, 2008; Brette, 2012). Curto and Itskov developed an algorithm for converting neural coactivity into a kind of topological space called a simplicial complex (Curto and Itskov, 2008). The spatial structure of this complex is determined by the relative timing of spikes in a population. Their simulations showed that invariant topological measures of the spike-derived simplicial complex matched several measures of the space of stimuli driving the neural response. This powerful property allowed them to reconstruct some stimulus space properties solely from the population activity without having to correlate physical positions to population firing rates, i.e. compute receptive fields. To our knowledge, this topological approach has not been used to examine the invariant properties of the temporal coactivation patterns of in-vivo sensory system responses.

Here, we apply the Curto-Itskov construction to extracellular responses from the caudal medial neostriatum (NCM), a secondary auditory cortical region, in European starlings, a species of songbird. Starlings adeptly learn complex acoustic signals, and provide an important model for the neural basis of vocal learning and perception (Kiggins et al., 2012). The NCM, an auditory forebrain region analogous to secondary auditory cortex in mammals (Karten, 1967; Vates et al., 1996; Wang et al., 2010), is involved in processing complex behaviorally relevant acoustic signals (birdsongs) (Sen et al., 2001; Grace et al., 2003; Thompson and Gentner, 2010). NCM neurons respond selectively to subsets of conspecific songs (Bailey et al., 2002; De Groof et al., 2013), and single neuron responses display complex receptive fields (Theunissen et al., 2000; Kozlov and Gentner, 2016). How the responses of single NCM neurons combine to represent invariant structure from natural vocal signals at the population level remains unknown. We hypothesized that invariant song representations in NCM are carried in the coactivation

pattern of spikes in the population, and that these patterns manifest as a topologically invariant structure. Our results support this hypothesis. We show that the coactivity-based topology of NCM neural activity is non-trivial and carries stimulus specific information. Using a novel mathematical approach to compare simplicial complexes, we find evidence that temporal coactivity patterns in NCM represent both stimulus identity and learned relationships between stimuli. We suggest that understanding the topology of sensory-driven neural population coactivity offers novel insight into the nature of how invariant representations are constructed from complex natural signals by neural populations.

Materials and Methods

Lead Contact and Materials Availability

Requests for resources, reagents, and further information is available from the Lead Contact, Timothy Gentner (tgentner@ucsd.edu). This study did not generate new unique reagents.

Experimental Model and Subject Details

All protocols were approved by the UC San Diego IACUC.

Subjects

Four birds were used in the study, identified as B1083, B1056, B1235, and B1075. All birds were wild-caught in southern California and housed communally in large flight aviaries prior to training and physiological testing. During behavioral training, birds were housed in isolation along with a custom operant apparatus in sound-attenuation chambers (Acoustic Systems, Austin, TX) and maintained on a light schedule that followed local sunrise and sunset

times. Birds had unrestricted access to water at all times. During behavioral training, birds received all food (chick chow) through the operant apparatus, contingent on task performance. All subjects were at least one year of age or older at the start of the experiment. We did not control for the sex of the subjects.

Method Details

Stimuli

Starlings (both male and female) produce long, spectro-temporally rich, and individualized songs composed of repeating, shorter acoustic segments known as 'motifs' that are learned over the bird's lifespan. Motifs are on the order of 0.5 to 1.5 seconds in duration. In these experiments, a library of 16 six-second long "pseudo-songs" were used as stimuli. Each pseudo-song was constructed from six one-second-long motifs manually extracted from a larger library of natural starling song produced by several singers. The first motif of each pseudo-song was an introductory whistle motif, as commonly occurs in natural starling songs. The same whistle was used for all 16 pseudo-songs. After the introductory whistle, all motifs were unique to one pseudo-song (i.e none of the other motifs occurred in more than one song). The amplitude of each pseudo-song was normalized to 65 dB SPL.

Behavioral Training

We trained four birds on a two-alternative forced-choice task to recognize four naturalistic pseudo-song stimuli, following previously described operant conditioning procedures (Knudsen and Gentner, 2013). Briefly, the operant apparatus afforded access to three separate response ports. Subjects learned to peck at the center response port to trigger the playback of one of four pseudo-songs associated with two different responses: a peck to the left response port or a peck to the right response port. Pecking at the left response port following the presentation of

two songs, or at the right response port following the other two songs, allowed for brief (1-2 sec) access to a food hopper. Incorrect responses – pecking the response port not associated a given song – resulted in a short time out during which the house light was extinguished and the bird could not feed from the hopper. The twelve remaining pseudo-songs were never presented during operant training for that bird. Assignment of songs for behavioral training was counterbalanced across subjects. The accuracy of the birds over the last 500 trials prior to neural recording was B1083: 86.2%, B1056: 95.4%, B1235: 97.6%, B1075: 91.2%. All birds exceeded 90% accuracy during training. Acquisition of the behavior was fast; the number of 500 trial blocks required to first exceed 90% accuracy was B1083: 30, B1056: 14, B1235: 5, B1075: 13.

Electrophysiology

Once trained, we prepared the birds for physiological extracellular recording from caudo-medial nidopallium (NCM). We anesthetized birds (20% Urethane, 7 ml/kg IM), affixed a metal pin to their skull with dental cement, removed an approximately 1mm square window from the top layer of skull, and placed a small craniotomy dorsal to NCM. We inserted a 16- or 32-channel silicon microelectrode (NeuroNexus Technologies, Ann Arbor, MI) through the craniotomy into the NCM of the head-fixed bird positioned on a foam couch within a sound attenuation chamber (Acoustic Systems). Auditory stimuli were presented from a speaker mounted ~20 cm above the center of the bird's head, at a mean level of 60 dB SPL measured at 20 cm. The microelectrode was lowered until all recording sites were within NCM, and then allowed to sit for approximately 15 minutes to allow any tissue compression that might have occurred during insertion to relax and stabilize before stimulus presentations began. Auditory activity was confirmed by exposing the birds to non-task-related sounds and observing sound evoked responses in the raw voltage signal on one or more channels. The 16 pseudo-song stimuli

in the block were presented pseudo-randomly until a total of 20 repetitions for each stimulus. For B1083, after the completion of the first block, the electrode was driven deeper into NCM by a distance greater than the electrode length, to obtain another block from a disjoint population. Raw voltages recorded from each microelectrode site were buffered and amplified (20X gain) through a headstage (TBSI, Durham, NC), bandpass filtered between 300 Hz and 5 kHz and amplified (AM-systems model 3600, Sequim, WA) and sampled digitally at 20 kHz (16 bit; CED model 1401, Cambridge, UK) via Spike2 software (CED). We saved the full waveforms for offline analysis.

Experimental Design and Statistical Analysis

Spike Sorting

Extracellular waveforms were spike sorted offline using the Phy spike sorting framework (Rossant et al., 2016). After automatic clustering, the sort was completed manually using the KwikGUI interface. Clusters with large signal-to-noise ratio and less than 1% refractory period violations (taken to be 1 ms) were labelled as "Good" clusters and included in the analysis. Because the analyses are meant to quantify population level structure, sorting priority was given to extracting as many neural signals as possible. Clusters identified automatically were mostly kept separate unless obvious duplicates were observed. Duplicate clusters were defined by overlapping distributions in PCA space, similar waveform shape, and cross-correlations that respected refractory periods. Despite these measures, the set of clusters labelled "Good" are likely to contain multi-unit activity as well as single units.

Data Processing

Data processing began by converting the neural population activity on each trial into an n x m matrix comprising all the spikes across all the isolated clusters and binned into 10 ms

windows with a 5 ms overlap between windows. We used the Perseus persistent homology software for computing Betti numbers from simplicial complexes (Mischaikow and Nanda, 2013). For computing samples from the simplicial configuration model, we used the code described in (Young et al., 2017). All other computations were performed using custom-written Python and C code available at: http://github.com/theilmbh/NeuralTDA

Mathematical Background

Outside of neuroscience, invariance has been the subject of intense mathematical investigation. In particular, the mathematical field of topology is dedicated to studying the properties of arbitrary spaces that are invariant to different classes of transformations. This research has produced a wealth of techniques and methods for defining and quantifying invariant structures. Recently, these techniques have been increasingly applied to the problem of finding invariant structure in large datasets. Topology takes a global point of view, and the "quantities" of interest are entire spaces, rather than single numerical measures. By converting neural population activity into a topological space instead of reducing it to a series of numerical measures, we gain access to the library of techniques topology offers for characterizing invariant structure.

Here, we introduce the basics of the mathematics used for our analysis. More thorough introductions are available elsewhere (Hatcher, 2002), including for the many applications of algebraic topology (Ghrist, 2014). Our analyses work by constructing mathematical objects known as simplicial complexes. Simplicial complexes are topological spaces that are built from discrete building blocks known as simplexes. For each dimension, there is only one prototype simplex. For example, 0-dimensional simplexes are points, 1-dimensional simplexes are line segments, and 2-dimensional simplexes are triangles. Simplexes exist in every dimension,

including dimensions greater than 3 where visualization is not possible. For every simplex of dimension *d*, there is a collection of *d-1* simplexes called faces. Consider the tetrahedron, which is the prototype 3-dimensional simplex (or 3-simplex). The faces of the tetrahedron are the 4 triangles, which themselves are 2-simplexes. The "faces" of the triangles are the 6 edges, which are 1-simplexes, and so on.

Simplicial complexes are built by "joining" simplexes along common faces. That information is carried through the *boundary operators*, which are linear maps encoding which *d*-simplexes are attached to which *(d-1)*-simplexes. Given a simplicial complex, we construct a series of vector spaces, one for each dimension of simplex in the complex. These vector spaces are called "chain groups." The basis vectors of these vector spaces are taken to be, formally, the simplexes in the various dimensions. For example, a graph has a chain group in dimension 0 consisting of all formal linear combinations of vertices, and a chain group in dimension 1 consisting of all formal linear combinations of edges. The boundary maps are linear maps (matrices) between chain groups of adjacent dimensions, mapping the dimension *d* chain group to the *d-1* chain group. These matrices contain all the structure of the simplicial complex, which allows for all the computations described in this work to be performed using the standard, highly optimized numerical linear algebra routines available in scientific programming packages like SciPy.

Constructing Simplicial Complexes for Neural Data

To construct the simplicial complex associated to a population spike train, we follow methods similar to those developed by Curto and Itskov (Curto and Itskov, 2008). For each presentation of a stimulus, the neural response was divided into time bins of width *dt* and a percentage overlap between adjacent bins, both free parameters for the analysis. Unless

otherwise noted, for all analyses, *dt* was 10 ms with 50% overlap, to preserve continuity of the spike trains. Each spike in the time-period was assigned to its associated time bins. The result is an N_cells x N_bins x N_trials multidimensional array *D*, representing the population activity, where each element is a spike count in the time window *dt*. The spike counts were then divided by the window width to determine a "firing rate" for that cell in that time bin. The time-average was taken to yield an N_cells x N_trials matrix of average instantaneous firing rates for each cell for each trial. Then, the multidimensional array *D* was thresholded by determining which bins had a firing rate greater than some threshold value times the average firing rate. This yielded an N_cells x N_bins x N_trials binary array *B* that represents significantly active cells. Unless otherwise noted, the threshold value of 4 times the average firing rate of the unit was used. Since the average firing rate of the units was low, the threshold value did not make a significant impact on the results.

For each trial and for each time bin in *B*, the N_cell-long binary vector V was used to create a list of cells that were active in that time bin. This was repeated across all time bins to create a list of "cell groups" active during the trial, much like the cell groups defined by Curto and Itskov (Curto and Itskov, 2008). This list of cell groups was taken to define the maximal simplexes for the simplicial complex. Using algorithms derived from published work (Kaczynski et al., 2006) this list of cell groups was turned into a list of lists of basis vectors for the chain complex. The n-th sublist represents the chain complex basis vectors corresponding to the n-dimensional simplexes in the simplicial complex. We also computed the matrices that represent the boundary operators in the chain complex.

Betti Curves

Homology is one of the most basic topological invariants one can use to describe a topological space. Loosely speaking, homology counts the number of n-dimensional holes in the space. These holes are described purely algebraically by computing how n-dimensional cycles (such as a circle) are not the boundary of an n+1-dimensional simplex. For each dimension n, the number of such holes is called the nth Betti number. The reason that Betti numbers are a useful characterization of the topological spaces we construct is that any two topological spaces which differ in their Betti numbers cannot be topologically equivalent. Furthermore, topological holes interpreted neurally represent "gaps" in the coactivation pattern (which neurons *do not* fire together) and thus carry information about how the population activity is structured. We summarized the topology of neurally derived simplicial complexes by computing these Betti numbers across time to yield Betti curves.

Figure 2.1d illustrates the temporal filtration. From the start of the stimulus, cell groups from each time bin are added to the growing simplicial complex. The Betti numbers for the entire complex are computed to yield the value of the Betti curve for that bin. Then, the next bin is added, and the Betti numbers are recomputed. Simplicial complexes were fed into the Perseus persistent homology program (Mischaikow and Nanda, 2013; Nanda, 2013), which computed the sequence of Betti numbers in the evolving simplicial complex. In the language of applied topology, we computed the homology of the filtered simplicial complex using time as the filtration parameter (Giusti et al., 2016). Betti curves were computed for each dimension and for each trial of a given stimulus. Because the Betti numbers are discrete, we interpolated the output from Perseus using step functions. This allowed us to average the Betti curves across trials. We varied the time bin width *dt* between 5-250ms and observed that the shape of the Betti curves was consistent for values of *dt* between 5-50ms. For values of *dt* longer than 50ms, the Betti

curves looked like low-pass filtered versions of the small-dt curves. Varying the percentage overlap from 0 to 50% did not appear to significantly alter the Betti curves.

The Betti curves reflect the accumulated effects of moment-to-moment spatiotemporal structure in the population activity. To serve as a control, we used various shuffling procedures to break this structure. We computed "fully shuffled" Betti curves by taking the neural response matrix for each trial and shuffling each row independently across columns. Since each row corresponds to a cell's response, this approach shuffles each cell's response independently across trials and across cells. The effect is to break the temporal correlations between cells while preserving the total number of spikes from a given cell in a given trial. We computed trial-shuffled Betti curves by employing a similar shuffling procedure but instead of shuffling a cell's response across time independently for each trial and each cell, we shuffled each response across trials independently for each cell and each time bin.

Spatiotemporal structure in population responses to stimuli reflects two kinds of correlation. The first is correlation between individual cell spike trains and individual stimuli. We call this the stimulus specificity of single unit responses. The other kind of correlation is the set of n-wise correlations between neurons on a single trial across time (sometimes referred to as the noise correlation). The fully shuffled condition destroys both these forms of correlation, by assigning different single unit spike trains, and therefore a random n-wise correlation structure, to each neuron on each presentation of each stimulus. To dissociate these two correlation sources, we created an additional shuffle of the population data, termed the "within-stimulus mask shuffle", in which each stimulus is associated with a unique time bin-to-time bin permutation mask. We then shuffle the spiking response of each neuron in time individually and independently within a single trial for each stimulus, applying the same bin-to-bin permutation to

all trials of that same stimulus. This creates a new, independent, stimulus-specific response for each neuron, and replaces the empirical n-wise population correlations with a random pattern of co-activity that repeats across repetitions of each stimulus.

Stimulus Specificity

We analyzed the stimulus specificity of the empirical and shuffled Betti curves by first computing the distribution of the correlations between the Betti curves of pairs of trials. Then we compared the distributions of correlations from a pair of trials within a single stimulus to correlations from pairs of trials between different stimuli. The significance of these differences was computed using a linear mixed effects model (Lindstrom and Bates, 1988). The model predicted the correlation between Betti curves on pairs of trials treating the within-stimulus versus between-stimulus distinction as a random effect and treating the average firing rate and an interaction of firing rate and pair type as fixed effects. We included firing rate effects since the shuffles preserve average firing rate but destroy precise temporal correlations.

Simplicial Configuration Model

To test whether the topology of the neurally-derived simplicial complexes is likely to be seen in a random simplicial complex, we employed the simplicial configuration model (Young et al., 2017). Due to the computational constraints of the configuration model, for each stimulus, we averaged the neural activity over trials to yield an average population spike train. The simplicial complex for this spike train was computed and seeded the simplicial configuration model software (Young et al., 2017). This algorithm produces a Markov chain that yields samples from the uniform distribution over all simplicial complexes with local connectivity structure identical to the seed complex derived from neural activity. For each sample, the Betti numbers were computed using Perseus. The distribution of these Betti numbers serves as a null

distribution against which Betti numbers derived from the empirical data can be compared. We interpret data having Betti numbers that fall well outside the null distribution as containing non-random topological structure. To quantify how far outside the null distribution an observed Betti number falls, we calculated empirical p-values using the formula(Davison and Hinkley, 1997)

$$p = \frac{1+\sum(b_i>b_e)}{N+1} \quad (1)$$

Where b_i represents the Betti number for the i-th sample from the simplicial configuration model, N is the number of samples from the simplicial configuration model, and b_e is the empirically observed Betti number.

Simplicial Laplacian

We generalized the methods of (De Domenico and Biamonte, 2016) to define information-theoretic quantities related to simplicial complexes. Equipped with the boundary operator matrices ∂_i, the simplicial Laplacians (Horak and Jost, 2013) were computed as:

$$L_i = \partial_i^* \partial_i + \partial_{i+1} \partial_{i+1}^* \quad (2)$$

where * indicates the adjoint, which in these analyses means the matrix transpose. Given a simplicial Laplacian L_n, the related density matrix was computed as:

$$\rho_n = \frac{e^{-\beta L_n}}{Tr\ e^{-\beta L_n}} \quad (3)$$

where beta is a free parameter. Given a density matrix expressed in its eigen-basis, the eigenvalues represent the probability distribution over eigenstates with maximum entropy under the Hamiltonian given by the Laplacian. We can define the Kullback-Leibler (KL) divergence between any two such density matrices, ρ and σ, by first diagonalizing both ρ and σ and then sorting each vector of eigenvalues by magnitude. The KL divergence is then defined by:

$$D_{KL}(\rho, \sigma) = \sum_i \rho_i (\log \rho_i - \log \sigma_i) \quad (4)$$

Where ρ_i is the i-th eigenvalue. The Jensen-Shannon Divergence was defined using:

$$M = \frac{\rho + \sigma}{2} \quad (5)$$

$$D_{JS}(\rho, \sigma) = \frac{1}{2}\left(D_{KL}(\rho, M) + D_{KL}(\sigma, M)\right) \quad (6)$$

The advantage of sorting the eigenvalues is that the divergence becomes invariant to the labels assigned to individual neurons. This is desirable because these measures should depend only the neural activity and not the specific labels assigned to neurons, analogous to how different maps of the world describe the same geography with different coordinates. This definition of JS and KL divergence ignores the eigenvectors of rho and sigma. Indeed, rho and sigma may not be simultaneously diagonalizable (or even of the same dimension). Rectifying these issues requires a choice in the definition of the metrics, which like any other metric is justified by its ultimate use. We explain our choices in the following paragraph. Regardless, the eigenspectra of rho and sigma do form discrete probability distributions, and the definitions of KL and JS divergence here agree with the usual definitions of these divergences on discrete probability distributions.

One limitation in the naive generalization of the foregoing metrics to neural data is that they rely on the density matrices being square and of the same dimension. For real data, however, different stimuli or repetitions of the same stimulus often evoke different numbers of active neurons, giving rise to chain groups, boundary operators, and density matrices of different dimension. To address this, given two Laplacians of different dimensions, we expanded the dimension of the smaller Laplacian by padding with zeros. An alternative approach is to collapse

the two simplicial complexes together along their common simplexes and derive "masked" boundary operators that only operate on the simplexes within the original, individual complexes, and then define the Laplacians using these masked boundary operators. Both the "zero-padding" and the "masking" approaches yielded similar results for our data sets. We report results from the computationally simpler method of zero-padding. In all analyses, the simplicial Laplacian was computed from the final simplicial complex constructed from the population response to an entire trial.

Beta Parameter. The free parameter, beta, that appears in the expression for the density matrix originates in statistical physics, where it is interpreted as "inverse temperature", i.e. beta is proportional to 1 / temperature. The significance of beta in the context of network theory is actively under study (Nicolini et al., 2018). Here we interpret beta by noting that it acts to renormalize the Laplacian matrix, and by extension, the eigenvalues of the Laplacian. In this sense, beta acts as a "scale parameter" that sets the scale of the spectral features of interest. The eigenvalues of the Laplacian are always non-negative, and if beta is large, then large eigenvalues are significantly damped by the negative exponential. Heuristically, increasing beta decreases the proportion of larger eigenvalues that contribute to the quantities of interest. In our case, if beta is too large, too little of the spectrum will be relevant for the JS or KL divergences, and we reasoned the results would be less interpretable. To confirm this exponential damping, we computed the proportion of eigenvalues above a threshold of 1e-14 as beta varied from 0.1 to 100. Generally, this proportion started decreasing from 100% with beta $\cong$ 1, to below 20% when beta > 10. We set beta = 1 in all of our analyses, but the main conclusions do not change as along as beta was below the upper bound of 10.

Simulated Spiking Populations

We used synthetic spike trains to examine how well the proposed metrics reveal invariant properties between neural populations. For "target" spike trains, we simulated a population of 20 Poisson spiking neurons for 1000 time-steps with a rate parameter selected from the set {0.01, 0.02, 0.03, 0.04}. We generated "test" spike trains by choosing 50 rate parameters evenly in the range of [0.001, 0.1] and simulating 25 population spike trains for each rate. We computed the simplicial KL divergence between each test spike train and the target spike train, then computed the average divergence between the target and the model at the given rate parameter.

We applied the same procedure to simulate heterogeneous Poisson populations, except that half of the neurons were simulated with rate parameter 0.02 and the other with 0.05. The assignment of neurons to rate-parameter subgroups was random. The above-described parameter sweep procedure was repeated over the two population rate parameters.

Simulated Environments

We simulated physical environments similarly to the environments used in Curto and Itskov's original work(Curto and Itskov, 2008). We randomly placed 0 to 4 holes of radius 0.3 meters inside a 2 x 2 m arena, ensuring hole centers did not overlap. We then randomly distributed 100 place fields of radius 0.2 meters in the environment using an algorithm that ensured that the environment was fully tiled. To model movement, we constructed a random walk trajectory simulating a 10-minute traverse of the environment sampled at 10 samples per second to ensure the whole arena was explored. We constructed spike trains by simulating Poisson spike trains at a constant rate for each neuron while the random walk trajectory was within the cell's place field, and silent outside of it.

We constructed five different environments with each specified number of holes. For each environment, 10 random walks were simulated. The pairwise 1-JS divergence between pairs

of individual trials was computed by taking the simulated spike trains' 1-Laplacians and computing the spectral JS divergence as defined above. We averaged the 1-JS divergences over all walks for each environment to yield the plot in Figure 2.5 panel a. Then, the JS divergences for each of the 25 environment pairs with given numbers of holes were averaged to yield the plot in Figure 2.5 panel b.

***In-vivo* JS Divergences**

Data from trained, anesthetized birds was recorded and preprocessed to yield population spike trains as described earlier. The JS divergence analyses were restricted to the 4 learned stimuli and 4 other unfamiliar stimuli. For an arbitrary pair of responses from two separate stimulus presentations, the 1-Laplacian of each response was computed by first converting the neural data into a simplicial complex as described above, using a bin size of 10 ms with 5 ms overlaps. Then, the JS divergences between the 1-Laplacians were computed as described above. It is important to note that the JS divergences were computed using the final simplicial complex from each trial. As a result, it may "miss" homological features that do not persist through the whole trial. We do not consider this a limitation in the present study, since the final simplicial complex reflects the aggregate effects of temporal coactivations through the entire trial, and thus tells us something about the population response to the entire stimulus considered as a unit. Nevertheless, it will be valuable for future work to explore the properties of these features that may be missed by the current analysis.

Correlation-based population response similarity

As an alternative to the Simplicial Laplacian Spectral Entropy measure of population similarity, we also compute a correlation-based measure to quantify the trial-to-trial similarity in the response of a population, $C(A,B)$, where A and B are the responses of the same population

on different trials responding to either the same stimulus or to different stimuli. We first smooth the population spike trains with a gaussian kernel with a width of 10 ms, then compute the correlation coefficients between the spike train of a neuron in response A and the spike train from the same neuron in response B, then average across all the correlation coefficients to a yield a raw scalar estimate of similarity, *C*, between the two populations. We use 1-C in place of the raw correlation measure so that smaller values of both correlation and JS divergence indicated more similar spike trains, such that $1 - C = 0$ when the response of every single neuron in the population for population response A is perfectly correlated with its corresponding response for population response B. Comparing this similarity measure to the SLSE measures allows us to test whether physically dissimilar responses (in terms of the actual spike trains) may evince similar topologies.

We also compute the distance between pairs of population responses using covariance matrices. For each trial, we calculated the pairwise covariance matrix between gaussian-smoothed spike trains, again with a 10 ms window. This yields an Ncell x Ncell matrix for each trial in which each *i,j* entry is the temporal covariance between the smoothed spike trains from neurons i and j on that trial. We obtained a distance between pairs of trials by computing the Frobenius norm of the difference between each trial's covariance matrix. The distance matrix was normalized by its maximum value over all pairs of trials to yield a normalized distance measure. This measure places two trials close to each other if the second order (pair-wise) correlation structure in the population for response A is similar to the second order correlation structure for response B.

The correlation between population responses with relabeled neurons was obtained by shuffling the rows of the gaussian smoothed population response matrix and repeating the population correlation computation described above on the shuffled matrices.

JS divergence and correlation measures were compared using the relative accuracy of logistic regressions. We fit logistic regressions to predict whether a pair of responses to different stimuli came from stimuli that belonged to either the same or different behaviorally trained class based on the value of either the JS divergence or the specific correlation measure of interest between the responses. Regressions were fit to a random subset of 80% of the data and tested on the remaining 20%. A total of 240 separate regressions were performed for each condition.

We measured the similarity of topological structure across disjoint sub-populations by randomly splitting the full dataset from B1083 population 1 into two disjoint subgroups each with 50 neurons. Then we conducted an identical analysis to the logistic regression described in the preceding paragraph but using the JS divergence between responses from the two disjoint populations. We repeated this 40 times with a different (randomly chosen) split of the original population. We assessed the difference between the means of the accuracies of the regressions fit to the JS divergence or correlations with a gaussian GLM from the statsmodels python library with the type of measure (JS divergence / correlation) and the type of shuffle (shuffle / no shuffle) as regressors, and included an interaction term between the shuffle and measure.

Linear Mixed Effects Model

We used a linear mixed effects model to examine the stimulus specificity of the empirically derived Betti curves associated with each stimulus, and those derived from the various shuffled versions of the spiking responses. The regression models predicted the correlation between Betti curves in a given dimension between pairs of trials, as a function of

whether the pair came from two trials of a single stimulus or two trials of different stimuli. The model included population identity as a random effect. The model was implemented using MixedLM from the statsmodels python library. We included firing rate (and its interaction with stimuli) as a separate fixed effect in the model and exclude this from all reported effects on Betti curve correlations tied to stimulus specificity. Overall, the mean firing rate on a pair of trials tends to increase the strength of the correlation between Betti curves on those trials, in part because it tends to smooth over small moment-to-moment variations in a manner equivalent to computing simplicial complexes from spike trains quantized into larger time bins.

Data and Code Availability

Data is available upon request. All of the custom code used for the topological analyses described here is available at https://github.com/theilmbh/NeuralTDA

Results

Stimulus Specific Topological Structure in Auditory Neural Populations

We first asked whether auditory stimulus-driven NCM population activity can be well-described by its spike-based topological structure. We recorded song-evoked responses from five, distinct, large populations of neurons in the NCM of adult starlings (323 neurons total, distributed across 4 birds; B1083: 101, 95 neurons in two separate populations; B1056: 54; B1235: 40; B1075: 33). For each presentation of each song stimulus within each population, we re-represented the full population spike train as a simplicial complex, then computed the Betti curves (Methods) associated with each song stimulus. Each Betti curve describes structure in the population coactivity pattern across time. More specifically, each curve captures the evolution of the homology of the simplicial complex in a single dimension, as defined by the time-resolved spiking pattern across each neural population for a given stimulus (Figure 2.1). Visual inspection

of the trial-averaged Betti curves (Fig. 1e) suggests consistent stimulus-specific temporal dynamics in the homology of dimensions 0, 1, and 2. Although the simplicial complexes sometimes display three-dimensional homology (Fig.1e, bottom row), these are not consistent across birds or stimuli. We note that the trajectories of the Betti curves in dimensions 0-2 tended to overlap during the first motif for each stimulus, consistent with their stimulus specificity, as the first motif was the same for all songs.

Betti numbers provide one measure of topological properties present in the underlying neural population coactivity, and so can be interpreted as abstract proxies for structure. They also provide insight into more concrete aspects of the underlying population activity. Betti 0 counts the number of connected components in the simplicial complex. Thus for neural data, increasing Betti 0 suggests multiple independent subsets of neurons firing coincidently within a group but not across groups. Higher dimensional Betti numbers, 1 to N, count the number of n-dimensional 'holes' in the simplicial complex. For neural population coactivity patterns, the Betti *n* count corresponds loosely to the number of gaps (i.e., periods of simultaneous inactivity among neurons, or "missing" patterns of coactivity). As these gaps fill in, due to coincident activity, the corresponding Betti numbers may decrease.

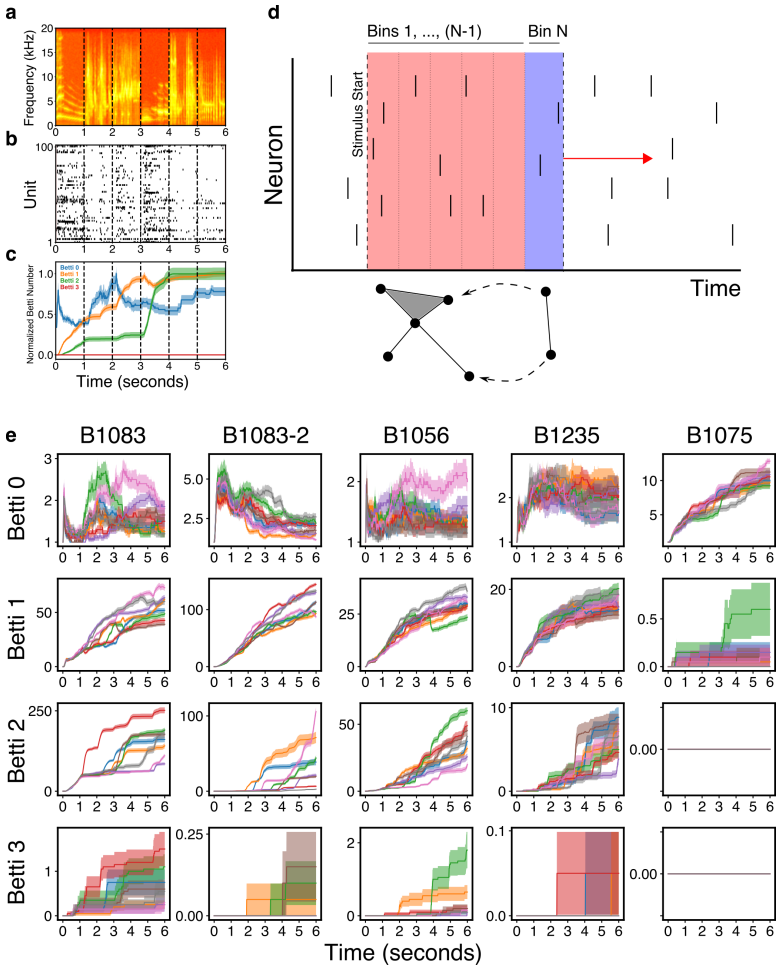


Figure 2.1: NCM population activity produces non-trivial topological features. (a) Spectrogram from a typical 6-second stimulus with motif boundaries marked with black dotted lines. (b) Spike raster from 100 units from region NCM during one trial of the presentation of the stimulus in (a). (c) Normalized average Betti curves from the population in (b) for the stimulus in (a). The values of the Betti numbers are normalized to their maximum during the stimulus presentation, to facilitate visualization of all dimensions on the same plot. The Betti curves are averaged over 20 presentations of the same stimulus. (d) Schematic of the temporal filtration. Beginning with the start of the stimulus, the elementary simplex from each time bin is added to the growing simplicial complex. At each time bin, the Betti numbers for the entire complex up to that point in time are computed. Then, the next bin is added, and the values of the Bettis are recomputed. (e) Betti curves for Betti numbers 0-3 for all stimuli for all populations. Each color represents a different stimulus, and identical colors across panels represent the same stimulus. Shaded regions indicate standard error of the mean. Individual stimuli produce unique Betti curves. Betti curves for different stimuli overlap during the first 1 second of the stimulus, consistent with the first 1 second of all stimuli being identical.

To quantify stimulus specificity, we computed all the correlations between Betti curves from pairs of trials with the same stimulus and pairs of trials from different stimuli. We defined the population as being stimulus specific if the distribution of within-stimulus correlations is significantly larger (ie. closer to 1, more similar) than the distribution of between-stimulus correlations. By this measure, all Betti curves show significant stimulus specificity (linear mixed effects model, $Z \leq -4.083$ $p \leq 4.45e\text{-}5$ for all dimensions), with the coefficient significantly less than zero (Table 1), indicating that between-stimulus correlations are significantly lower than within-stimulus correlations. That is, the topology of auditory stimulus-driven population activity in NCM is stimulus specific.

Table 2.1: Stimulus Specificity of Empirical and Shuffled Betti Curves. A linear mixed-effects model was used to compute the dependence of the Betti curve correlation on the type of trial pair (within stimulus or between stimuli). For all Betti numbers, the empirical curves show significant stimulus specificity, and this specificity is abolished by both fully shuffling the spikes and performing the within-stimulus mask shuffle (Methods).

Data set	Betti Number	Coefficient	Z	P-Value	97.5% CI
Empirical	0	-0.031	-4.083	4.45e-05	[-0.046, -0.016]
Shuffled	0	-0.010	-1.275	0.202	[-0.024, 0.005]
Within-Mask	0	-0.003	-0.33	0.741	[-0.017, 0.012]
Empirical	1	-0.032	-5.571	2.53e-08	[-0.043, -0.021]
Shuffled	1	0.000	0.303	0.762	[-0.003, 0.004]
Within-Mask	1	0.002	1.432	0.152	[-0.001, 0.005]
Empirical	2	-0.149	-15.603	6.96e-55	[-0.167, -0.130]
Shuffled	2	-0.030	-2.993	0.003	[-0.050, -0.010]
Within-Mask	2	-0.008	-0.857	0.391	[-0.027, 0.011]

To further characterize the stimulus specific dynamics of the Betti curves and examine the source of the stimulus specificity in the empirical population response, we compared the original Betti curves to those obtained after shuffling the spiking response of each neuron in time individually and independently, within each trial (Methods). This shuffle, which we call the full-shuffle, preserves the total number of spikes and spike rate per trial, but destroys all the original spike-time coincidences between neurons in the population. Because the topological structure of the population depends on the spike-time coincidences, any stimulus specificity tied explicitly to the spike-time based topology (rather than the overall spike rate) should be abolished in the full-shuffled data. This is turn should yield Betti curves that are more similar across different stimuli and birds compared to the original Betti curves. Figure 2.2 shows the Betti curves that result from the full-shuffle (in orange) for a single stimulus from each of the populations. In each dimension, the fully shuffled curves show a degeneration to a stereotypical trajectory across populations. The same pattern is observed for all other stimuli. We quantified this explicitly by again comparing correlations between Betti curves from pairs of trials within and between stimuli. For Betti 0 and 1, the shuffled curves were not stimulus specific (linear mixed effects model, $p > 0.202$). For Betti 2, some marginal stimulus specificity remained (linear mixed effects model, $Z = -2.993$, $p = 0.003$), but the magnitude of the specificity was significantly smaller than that for the empirical data (linear mixed effects model, $Z=-21.24$, $p < 1e\text{-}13$). Based on these results, we conclude that the topological structure of NCM population activity reveals stimulus specific coincident patterns of neuronal firing.

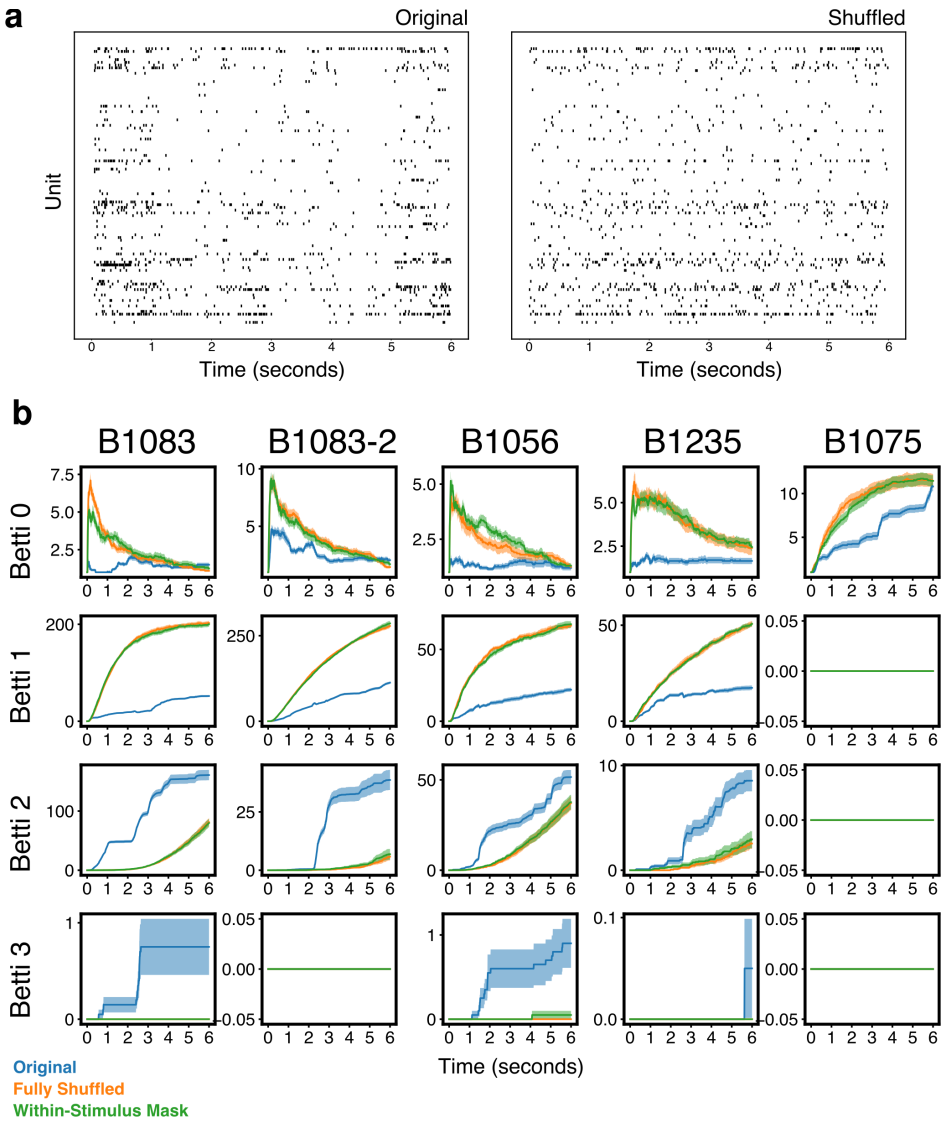


Figure 2.2: Shuffled Betti Curves. (a) Example empirical and shuffled population responses from an example trial. (b) The average Betti curve for dimensions 0-3 from an example stimulus is plotted for all populations. The empirically observed Betti curves (blue) are obtained from the raw neural data. Control Betti curves are computed from "fully shuffled" (orange) neural data in which the responses of individual cells are independently shuffled in time (Methods), or the same neural data with the single trial spike responses permuted by either a within-stimulus mask (green) or across trials (red). A different within-stimulus mask was generated for each stimulus, and the same mask was used for all trials of a given stimulus.

How is it that the temporal coactivity pattern of the population response, as measured by the topology, carries stimulus specific information? One possibility is that the population simply inherits specificity from single unit responses. Individual NCM neurons have complex receptive fields (Kozlov and Gentner, 2016) and their responses are likely to be stimulus specific (Thompson and Gentner, 2010). The collective responses of many such neurons is therefore also likely to be stimulus specific, but the specific pattern of coactivations among individual neurons that defines the topology may or may not be determined by chance. That is, the topology may reflect random coactivity between independent spike trains driven by individual stimuli (i.e., the stimulus specificity of single unit responses), and/or a unique non-random *n*-wise pattern of coactivity between neurons that is associated with a given stimulus, where *n* ranges from 2 to the number of recorded neurons. The full-shuffle destroys both these forms of co-activity by permuting single unit spike trains on each presentation of each stimulus, which in turn yields a unique *n*-wise coactivity structure on each trial. To further investigate these two possible sources of population co-activity structure, we created an additional shuffle of the original population data using a "shuffle mask" that describes how to permute the time bins for a given neuron on a single trial. We randomly generated a shuffle mask independently for each stimulus, and then applied the same mask to all trials from a single stimulus. This creates a new random correlation pattern between the spike trains of individual neurons and individual stimuli, and a random *n*-wise co-activity pattern between individual neurons, that are both preserved across separate trials from the same stimulus. If the observed stimulus specific topologies emerge from a random alignment of spiking responses repeated across trials of the same stimulus, the mask-shuffled data should resemble an empirical response to a novel stimulus. If, however, the empirical n-

wise correlation structure is privileged in some way, then the mask-shuffled data should resemble the full-shuffle responses.

We recomputed the Betti curves for the within-stimulus mask-shuffled data. Figure 2.2 shows the various shuffled Betti curves from an arbitrarily chosen stimulus for each population. Like the full-shuffle, the Betti curves for the mask-shuffled data do not appear to be stimulus specific. To test this, we again measured stimulus specificity by computing the distributions of correlations between pairs of Betti curves from trials with either the same or different stimuli. The mask-shuffled Betti curves did not show significant stimulus specificity (linear mixed effects model, $P > 0.152$ for all Betti numbers). Table 1 gives the results of the linear mixed effects model (Methods) that assesses the significance of stimulus specificity in the Betti curves from the within-stimulus mask shuffle. Figure 2.3 presents histograms of the Betti curve correlations for within-stimulus and between-stimulus pairs of trials, for both the empirical and shuffled data, for population B1083. The lack of stimulus specificity in both the full- and mask-shuffled data supports the conclusion that stimulus evoked coactivity in NCM carries stimulus specific information that is not a trivial product of randomly aligned stimulus-specific single neuron responses.

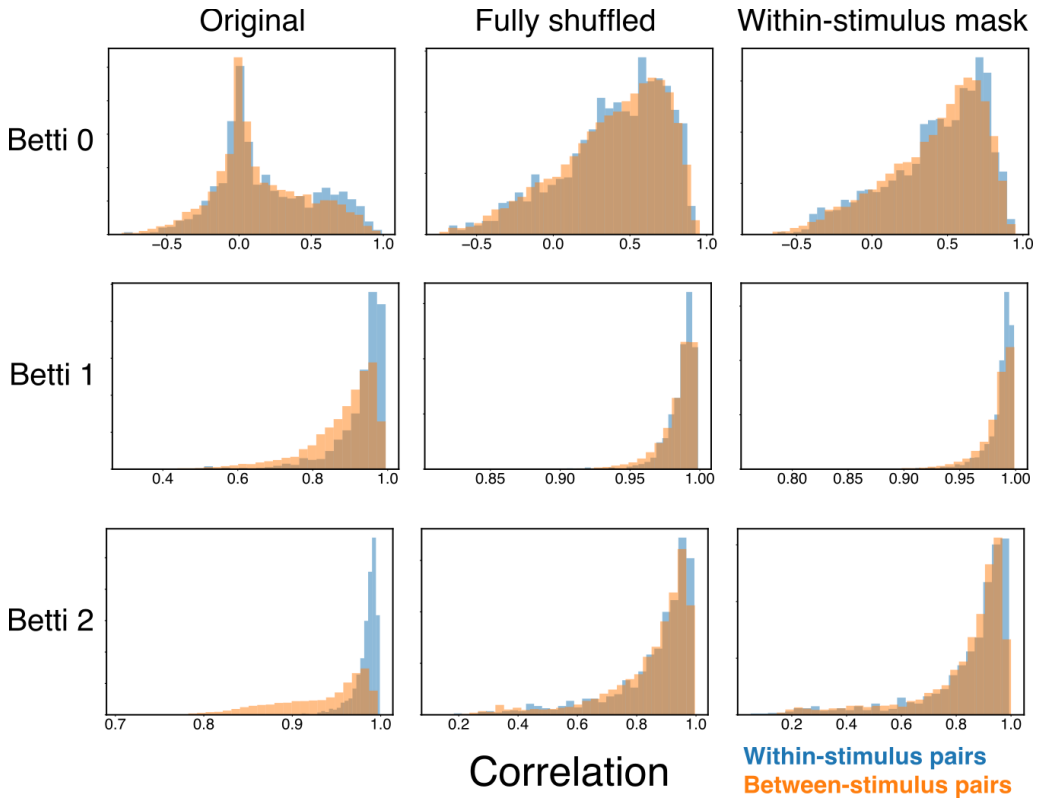


Figure 2.3: Example Betti curve correlation distributions. Stimulus selectivity of the Betti curves is assessed by computing the distributions of the pairwise correlations between Betti curves on pairs of trials. Shown are distributions from bird B1083. The blue distributions come from trial pairs that belonged to the same stimulus. The orange distributions come from pairs that belong to different stimuli. The original, empirical distributions show significant differences, indicating stimulus selectivity of the Betti curves. The distributions computed from shuffled data show little difference among within or between stimulus pairs. The significance of the differences in these distributions was assessed using a linear mixed effects model. Table 1 lists the significance for the empirical and shuffled conditions.

The differing shapes of the Betti curves for the empirical and both the full- and mask-shuffled spike trains suggests that the NCM population activity is structured non-randomly at scales above the single neuron. To test this idea directly, we asked how likely it is that the observed topological structure in a given population might occur by chance. We constructed a null model for simplicial complexes that produces samples from the uniform distribution over all simplicial complexes that match the "local structure" of a given "seed" complex (Young et al., 2017), and used this model as the basis to compare the final Betti numbers for each stimulus (Figure 2.4). For each bird, the majority of the stimuli showed responses with at least one Betti number significantly outside the null distribution (B1083: 8/8 stimuli, B1056: 8/8 stimuli, B1235: 8/8 stimuli, B1075: 7/8 stimuli). Thus, the stimulus specific topology in NCM population spiking is not produced by chance spike coincidences between neurons.

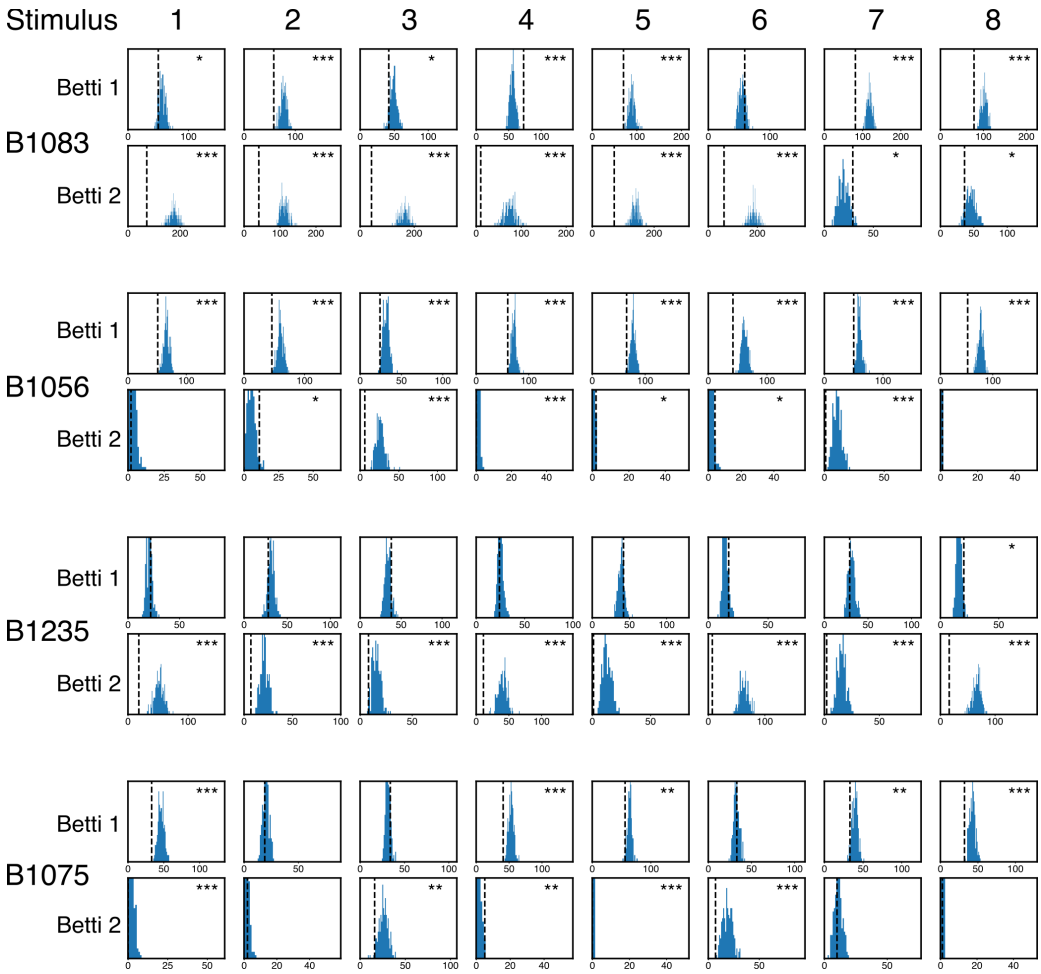


Figure 2.4: Most Betti numbers lie outside their null distributions. Each plot shows the null distribution of values of a given Betti number for a given stimulus and neural population generated by the simplicial configuration model (Methods). Black dashed lines denote the observed value of the Betti numbers in dimensions 1 and 2 at the end of the trial-averaged response to the indicated stimulus. Asterisks indicate the degree to which the dashed lines lie outside of the null distributions, interpreted as an empirical p-value (***: $p < 0.01$; **: $p < 0.05$; *: $p < 0.1$).

As noted, the topological analyses we detail differ in fundamental ways from traditional correlation-based measures of population co-activity. Nonetheless, it is helpful to understand our results within the context of classically defined noise correlations (i.e. the stimulus independent covariance in simultaneous firing rates between neurons). To do this, we again shuffled the empirical data, this time by permuting spikes across trials rather than time. The Betti curves for these trial-shuffled data show similar trajectories compared to those for the empirical data (Figure 2.5), but with larger magnitudes. This suggests that the relative temporal dynamics of the co-activity pattern are governed largely by what would be referred to, classically, as the 'signal correlation', whereas the absolute number of "holes" in a given topological dimension is constrained by the 'noise-correlation'. Recall that the magnitude of Betti n corresponds roughly to the number gaps in the population responses that are bounded by co-active cell groups of order n, and that as these gaps fill in, Betti n decreases. In other words, noise correlations increase the numbers of co-active cell groups. Overall, both the signal and noise correlation contribute to the topological structure of coactivity on each trial. This is consistent with previous studies (Jeanne et al., 2013) that show NCM signal and noise correlations are not independent.

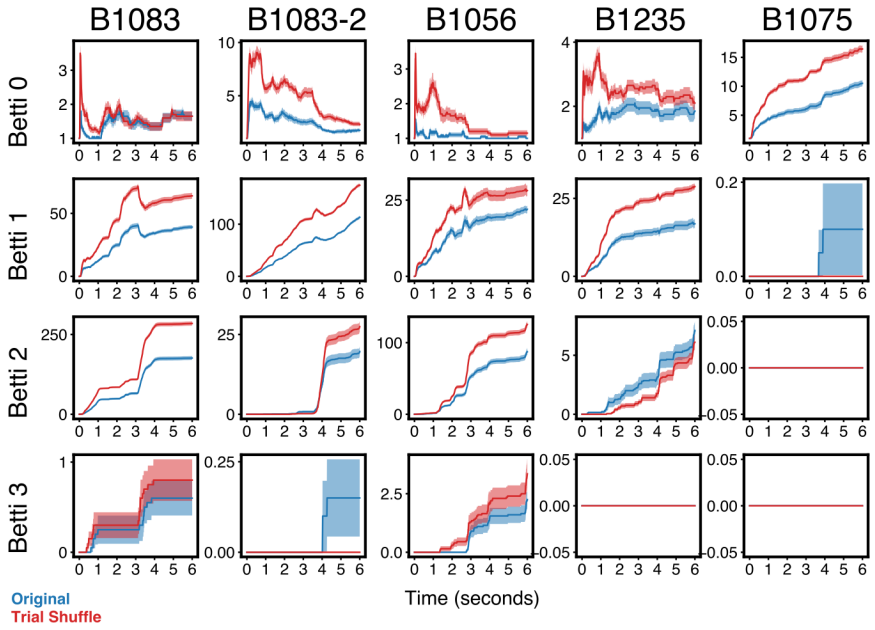


Figure 2.5: Trial Shuffled Betti Curves. The red curves result from shuffling spikes across trials rather than across time. Betti curves show similar trajectories but different magnitudes. A single stimulus from each population is shown, and this stimulus is different than that displayed in Figure 2.2.

Topological Tools for Comparing Population Spiking Activity

Representing population spiking activity with a simplicial complex has the advantage of providing a single mathematical object that encodes the entirety of the spatiotemporal structure of the population response on a single trial. This entity is non-numerical, however, and so to facilitate numerical comparisons we sought to define a numerical measure of similarity between simplicial complexes. To do this, we generalized recent advances in network theory to define information-theoretic measures of simplicial complex structure and similarity. A detailed description of these measures is provided in the Methods; Table 2 gives the principle formulas in this analysis. We refer to this general approach of computing information theoretic quantities from simplicial complexes as Simplicial Laplacian Spectral Entropy (SLSE).

Fitting Spiking Models Using Simplicial Laplacian Spectral Entropies

The Kullback-Leibler divergence is often used as a cost function for statistical model fitting, in which one attempts to choose the parameters of a statistical model such that the model distribution is as close to the data distribution as possible. The Kullback-Leibler divergence can be defined to allow for this same sort of model fitting, but for graphs (De Domenico and Biamonte, 2016). We extended the Kullback-Leibler divergence to simplicial complexes (Methods) with the reasoning that it could be used similarly to fit the parameters of spiking neural network models. To test this reasoning, we began with a proof-of-concept: fitting the rate parameter of a simulated population of Poisson-spiking neurons. The population consisted of 20 neurons each firing independently and with a rate parameter that was constant across the population. We simulated a single trial of 1000 time samples to serve as the "target" population spike train. We then performed a parameter sweep by simulating the population with a rate parameter that varied in some range. For each choice of rate parameter, 25 "test" trials from the

model were simulated and the 1-KL divergence between the "target" population spike train and each "test" trial population spike train was computed. The 1-KL divergences from each of the 25 trials of the chosen parameter were averaged together. Figure 2.6a displays the 1-KL divergence as a function of the test rate parameter for four different models with different "target" rate parameters. The dotted lines indicate the true value of the rate parameter used to generate the target population spike train for each model, separated by color. For each model, the minimum of the KL divergence closely approximates the true value of the rate parameter.

Table 2.2: Summary of the main formulas for Simplicial Laplacian Spectral Entropy. Included are the names, formulas, and the most important properties of each quantity. The formulas should be taken as definitions.

Quantity	Formula	Major Property
Simplicial Laplacian	$L_i = \partial_i^* \partial_i + \partial_{i+1} \partial_{i+1}^*$	Dimension of kernel is i-th Betti number
Density Matrix	$\rho = \frac{e^{-\beta L}}{Tr\, e^{-\beta L}}$	Eigenvalues form discrete probability distribution
Kullback-Leibler (KL) Divergence	$D_{KL}(\rho, \sigma) = \sum_i \rho_i (\log \rho_i - \log \sigma_i)$	Measure of distinction between density matrices
Jensen-Shannon (JS) Divergence	$M = \frac{\rho + \sigma}{2}$ $D_{JS}(\rho, \sigma) = \frac{1}{2}\left(D_{KL}(\rho, M) + D_{KL}(\sigma, M)\right)$	Symmetrized version of KL divergence

We next tested a slightly more complicated model that had two distinct populations of neurons. Each population was again Poisson, but the two populations had differing rate parameters. Figure 2.6b shows a heat map plotting the 1-KL divergence as a function of the two rate parameters. The white dots indicate the true parameters. There is a degeneracy in this model in that it doesn't matter which subpopulation is labeled A and which is labeled B. Again, the minima of the 1-KL divergence closely approximate the true values of the parameters for the heterogeneous population.

Importantly, for all of the Poisson spike train simulations, the "test" and "target" spike trains never coincided. Instead, similarities quantified by the simplicial KL divergence rely on the global topological structure of the population spike trains, not on the match between specific spike trains of single neurons. The SLSE is sensitive to how the activity of an individual neuron relates to the simultaneous activity of all the other neurons in the population in which it is embedded.

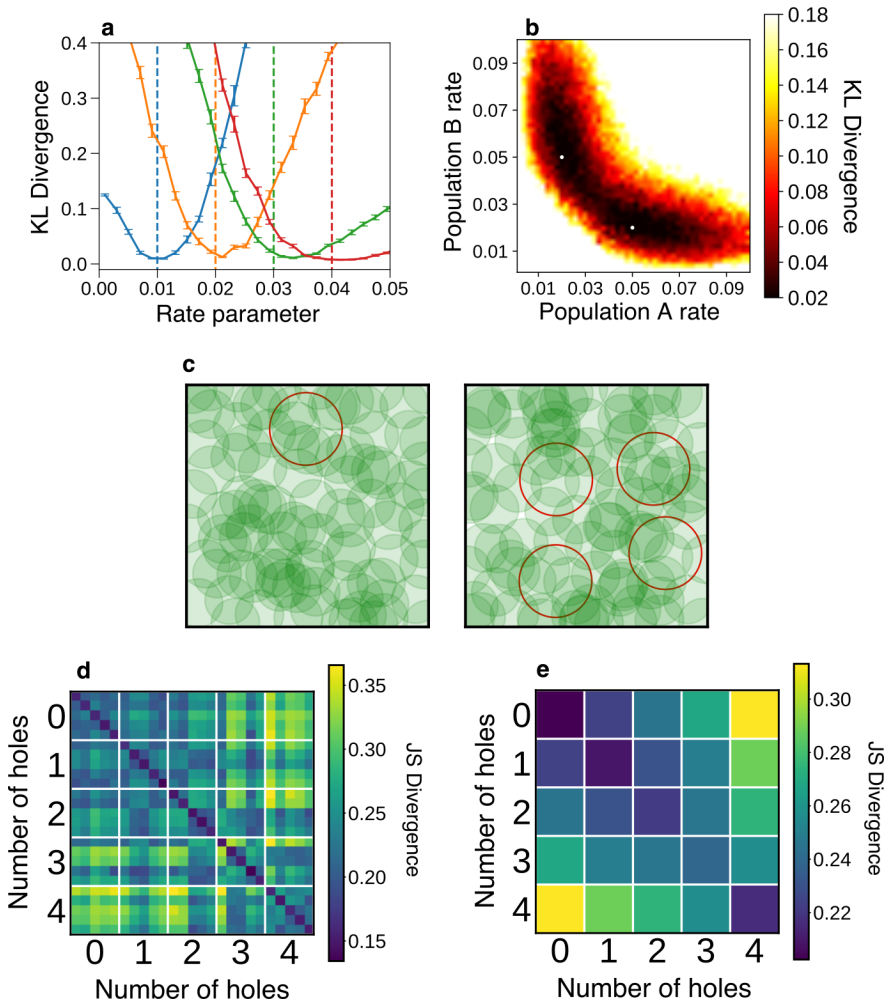


Figure 2.6: Simplicial Laplacian Spectral Entropy captures invariant population activity structure. (a) KL divergence as a function of the rate parameter for four different models (denoted by color), show that the minimum aligns with the target parameter (dotted vertical line) in each model. Error bars show SEM. (b) Heat map showing KL divergence as a function of rate for a model consisting of two independent Poisson subnetworks with different rate parameters (white dots show the true values for each network). (c) Two example simulated environments with 1 and 4 holes (red circles) tiled by place fields (filled green circles). (d) Average JS divergence between simplicial complexes derived from simulated place cell activity in environments with the indicated number of holes. (e) As in (d), but averaged over all similar environments, showing that the JS divergence grows with the differences between environments.

Reconstructing Latent Stimulus Relationships using Simplicial Laplacian Spectral Entropies

Our observation that NCM populations produce non-random, stimulus-specific, topologies implies that Poisson spike trains are not a good source of biologically relevant topologies, and so the above evaluation of these metrics is limited. As a more biologically relevant validation of the SLSE measures, we turned to the place cell system used in the original Curto and Itskov (2008) study of neural population-derived simplicial complexes and began by replicating these earlier results. We simulated the activity of hippocampal place cells during free exploration of a 2-dimensional rectangular arena, within which we randomly placed different numbers of circular "holes." A virtual animal explored the space through a random walk trajectory that was excluded from the holes in the simulated environment. Population spike trains were generated by covering the simulated environment with place fields and for each field, generating spikes from simulated place cells as a function of the random walk trajectory. Environments differed in the number and position of the holes in each space, and the position of the simulated place fields.

The functional utility of any neural representation rests on its ability to encode the relationships between stimuli. Consider, for example, a pair of our simulated environments that happen to share the same number of holes. This invariance in the stimulus is independent of how different sets of place fields might cover either environment, or the paths that different animals might take when navigating through either environment. If the neural topology mirrors the topology of physical stimulus structure, as is thought to be the case (Curto and Itskov, 2008), then the relationships between our simulated environments should be reflected in some of the relationships between the associated neural topologies. Thus, we hypothesize that the simplicial

JS divergence on simulated population spike trains from these environments should capture the invariant relationships between environments. That is, population spike trains from an environment with N holes should be most similar to those from other environments with N holes, and the simplicial JS divergence should increase as the difference between the number of holes in the environments increases. The results of our simulations (Figure 2.6 c-e) support this hypothesis and show that the simplicial JS divergence can detect similarities in the neural representation induced by relationships between the stimuli. As the difference in the number of holes between a pair of environments grows, the 1-JS divergence between spike trains from each environment grows. Importantly, these similarities persist across trials with unique paths through the arena, across populations that sample from different receptive fields, and across environments with different arrangements (but similar numbers) of holes. In each case, the SLSE technique detects similar representations of similar environments. Even though two spike trains may manifest wildly different spatiotemporal spike patterns, the structure encoded in their coactivity patterns can encode invariant relationships between the stimuli they represent. Because stimuli have no intrinsic value to the neural population, we argue that these relationships define the stimuli themselves.

Spectral Entropy-based Divergences between *in-vivo* Neural Population Activities

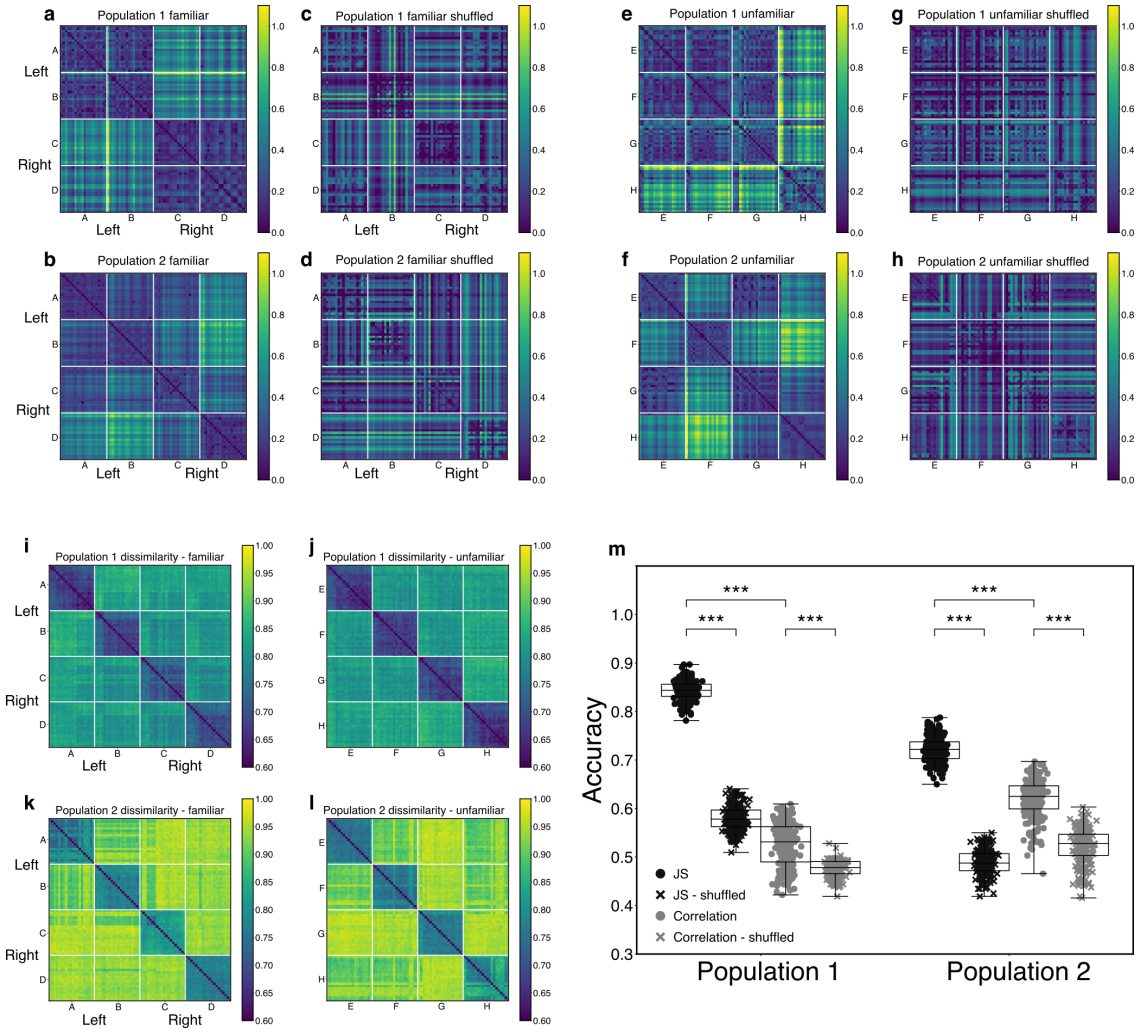
Having demonstrated that SLSE-based divergences are useful in quantifying invariant representations in simulated neural populations, we next applied these techniques to activities recorded from the brains of anesthetized birds trained to perform a two-alternative choice task (Methods). Briefly, birds were trained using established techniques to peck at assigned locations on an operant panel in response to different acoustic stimuli, responding “left” when presented

with two pseudo-songs and to respond “right” when presented with another two. From this training, the birds learned an arbitrary categorical structure (invariance) independent of any similarities in physical acoustics of the stimuli. We hypothesized that since the simplicial JS and KL divergences are designed to capture latent invariant structure in population activity, these divergences should reveal the learned categorical structure imposed on the stimuli by the behavior and presumably reflected in the population-level representation of the stimuli. Figure 2.7 shows the JS divergence between 1-Laplacians computed from the single-trial responses to both learned and unlearned songs in B1083, the bird with the most units recorded. Panels a and b show the divergences for learned stimuli in two disjoint populations in the same bird. Population 2 was located deeper than population 1 by more than one probe length, in this case 600 um. The stimuli are organized so that the two left-signaling stimuli are adjacent, followed by the two right-signaling stimuli. In both populations, the block structure of the JS divergence matrix shows that temporal coactivity patterns elicited by stimuli in different behavioral classes are “farther apart” (less similar) than the coactivity patterns elicited by different stimuli in the same behavioral class. This allows for the reconstruction of the learned latent categorical structure of the stimuli, supporting our hypothesis. Shuffling individual neurons’ spike trains in time abolishes the categorical structure (panels c and d). As a control, computing the JS divergences between trials for four unfamiliar stimuli shows no consistent categorization of the stimuli among the two populations (panels e-h).

To see whether learned behavioral categories could be encoded by the physically similarity of the population spike trains, we computed the correlations (Methods) between the population spike trains on single trials. Figure 2.7 panels i-l show the dissimilarity (1 – Correlation) for each pair of trials. Under this measure, physically similar population responses

result in lower values, like the JS divergence. Not surprisingly, the single-trial population responses to individual stimuli are physically similar (Figure 2.7 i-l), as the population correlation reflects the trial-to-trial repeatability of the stimulus-locked timing of spikes in individual neurons. In contrast to the JS-divergence, which captures the relative timing of spikes across neurons, the learned behavioral categories are not observable in the correlation (Figure 2.7 i, k). Thus, behaviorally relevant invariances captured by the coactivity spiking pattern are not observable in a measure of trial-to-trial repeatability that considers each neuron as independent.

Figure 2.7: JS divergence between in-vivo neural populations. (a-h) Each pixel represents a JS divergence between two trials. White lines demarcate trials belonging to different stimuli, identified by letters A-H. (a,b) JS divergence between trials for familiar stimuli for populations 1 and 2, respectively, on which the bird was trained to peck either left or right as indicated. Both populations show categorical structure reflecting the learned behavioral categories. (c,d) Same data as (a,b) but each neuron's spike train has been shuffled in time, breaking correlations but keeping absolute firing rates, and abolishing categorical structure. (e,f) JS divergence between trials for untrained stimuli. No overarching categorical pattern is observed. (g,h) Results from shuffling the data in (e,f, respectively. (i,l) Correlation dissimilarity between population spike trains (Methods), where 1 or 0 indicate perfect or no correlation, respectively, and behavioral categories are not observable for either population. (m) Accuracy of predicting same/different behavioral classes from the two different populations using either the JS divergence between simplicial complexes (black) or spike train correlation (gray; ***: $p < 0.001$).


a
Population 1 familiar
b
Population 2 familiar
c
Population 1 familiar shuffled
d
Population 2 familiar shuffled
e
Population 1 unfamiliar
f
Population 2 unfamiliar
g
Population 1 unfamiliar shuffled
h
Population 2 unfamiliar shuffled
i
Population 1 dissimilarity - familiar
j
Population 1 dissimilarity - unfamiliar
k
Population 2 dissimilarity - familiar
l
Population 2 dissimilarity - unfamiliar
Left
Right
m
Accuracy
***
JS
JS - shuffled
Correlation
Correlation - shuffled
Population 1
Population 2

To quantify the foregoing observation, we trained logistic regressions, using either the JS-divergence scores or the raw population correlations, to predict whether the population responses to two different stimuli belonged to either the same or different learned classes, then tested the accuracy of each decoder on held-out data (Methods). Figure 2.7m shows the accuracy of each decoder. For both populations, the accuracy of the decoded using JS divergence was significantly higher that the decoders using correlations, in both populations (t-test, t=95.36, 34.6; p=3.1e-313, 2.96e-132). Both population response measures outperformed decoders trained on shuffled datasets (JS-divergence: t-test; t=122.4, 44.49; $p < 1e\text{-}313$, =4.62e-172; correlation: t-test; t=12.42, 29.62; p=6.36e-31, 2.87e-110) indicating the even the raw correlations contain some information about the learned invariances.

If neurons are not independent, then trial-to-trial repeatability may not manifest in the stimulus-locked spiking of individual neurons but may instead be present in their statistical covariances. To test this, we computed the neuron-to-neuron pairwise covariance matrix for each trial (Methods), then found the normalized distance between these covariance matrices for pairs of trials (Methods). Figure 2.8c shows the distances between these covariance matrices. As in the independent neuron correlations (Fig. 7 i-l), we did not observe a strong signature of learned behavioral categories in the similarity between pairwise covariance matrices. We again quantified this observation by comparing the accuracy of logistic regression-based decoders trained to detect the learned behavioral class from the distance between covariance matrices on pairs of trials (Figure 2.8e) to that for the JS-divergence. Accuracy using the JS-divergence was significantly higher than that for the covariance matrix distance (t-test, t=102.43, 89.62; $p < 1e\text{-}313$, =5.0e-301; Fig. 7e).

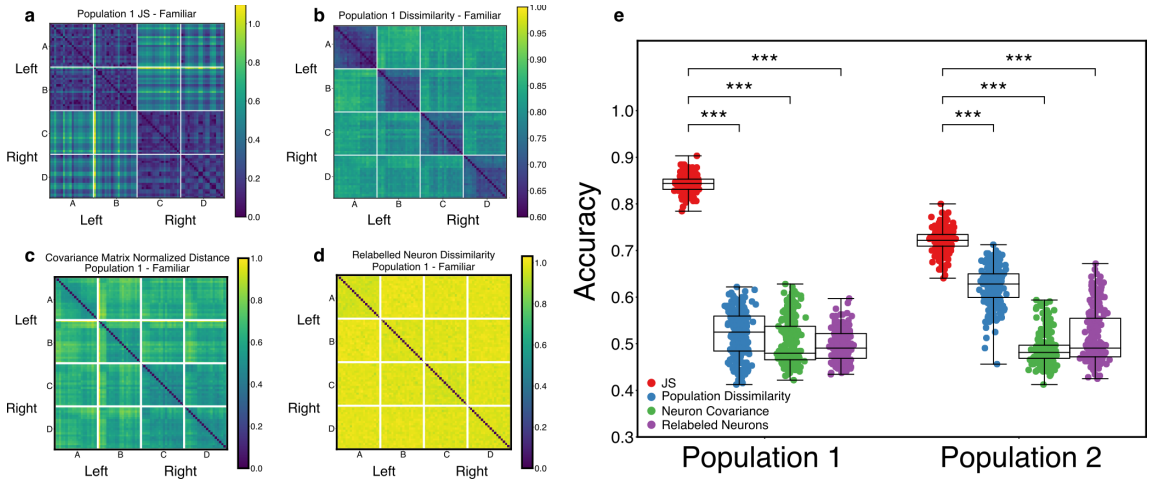


Figure 2.8: JS Divergence outperforms correlational measures of coactivity. (a, b) Distance matrices determined using the JS divergence and population dissimilarity (as in Fig. 6). (c) Normalized distance between neuron-neuron covariance matrices on pairs of trials, showing no segregation into behavioral classes. (d) Correlation between trials with neuron identities randomly relabeled on each trial, showing no segregation by behavioral class. (e) Accuracies of a logistic decoder trained to determine whether a pair of trials belongs to the same or different behavioral classes for each measure of population similarity. The JS divergence significantly outperforms all other measures tested, for both populations (***: $p < 0.001$).

To understand why the JS-divergence might capture information in the population representation that measures of correlation do not, it is helpful to note that statistical measures of coactivity, like correlation, assume coactivity is tied to fixed collections of neurons. Correlations at the population level (as in Fig. 7) assume each neuron is independent and simply track response reliability within each neuron across trials. Pairwise (or higher order) correlations (as in Fig. 8) require one to sample a defined pair (or larger set) of neurons multiple times, either across time within a trial or across trials to estimate variance. Moreover, because most pair-wise responses are not correlated, these effects can wash out as population sizes increase. The JS divergence measure is, by definition, invariant to the arbitrary labelling of neurons, can be computed over any interval, and is immune to addition of non-coactive neurons. Thus, the population coactivity pattern is different from a correlation. To highlight this, we computed the quantity 1-Correlation between population responses on each trial, having relabeled the neurons independently for each trial. The resulting distance matrix is shown in Figure 2.8d, again demonstrating a lack of segregation by behavioral class. Comparing this measure of correlation to the JS divergence using logistic regression (Figure 2.8e) again revealed a significant difference between the two measures (t-test, t=144.62, 51.06; $p < 1e\text{-}313$, =1.12e-195).

We performed the same logistic regression analysis between the 1-JS and correlation measures of response similarity on the neural populations obtain from the remaining birds. B1056 showed a statistically significant difference between JS and correlation (t-test; t=6.09, p=2.3e-09) and between JS and covariance matrix distance (t-test; t=7.06, p=5.7e-12). B1235 also showed a statistically significant difference between JS and correlation (t-test; t=11.8, p=1.4e-28) and between JS and covariance matrix distance (t-test; t=5.85, p=9.25e-9), but the regressions on the JS from fully shuffled data slightly outperformed the JS on empirical data (t-

test; t=-8.36; p=6.5e-16). B1075 did not have enough neurons in the population to robustly compute simplicial Laplacians on individual trials. In total, the JS divergence outperformed correlations in all of the 5 populations for which the JS divergence was computable.

The results so far suggest that the topological structure within a population is specific to behavioral classes. We next asked whether this topological structure was similar across disjoint but simultaneously recorded populations. To test this we took population 1 from B1083 and split it into two disjoint populations containing equal numbers of neurons. Then, we computed both the JS divergence and correlation between individual trials both within and between these subpopulations. Finally, we computed the accuracy of a logistic regression model to predict whether trials from physically distinct stimuli nevertheless belonged to the same behavioral class. We repeated this measurement for 40 random, independent splits of the original population. Figure 2.9 shows that the JS Divergence outperforms correlation (glm; z=2.501; p=0.012), indicating there is some similarity between the topological structures produced by simultaneous but independent populations that reflects the learned behavioral class. Furthermore, both the JS divergence and correlation outperform the shuffled versions of their own measures (t-test; t=3.876, 4.61; p=0.0003, 1.57e-5). The large variance in the JS divergence we attribute to the fact that the split populations have half as many neurons as in Figure 2.7, leading to a "lower resolution" reconstruction of the topology.

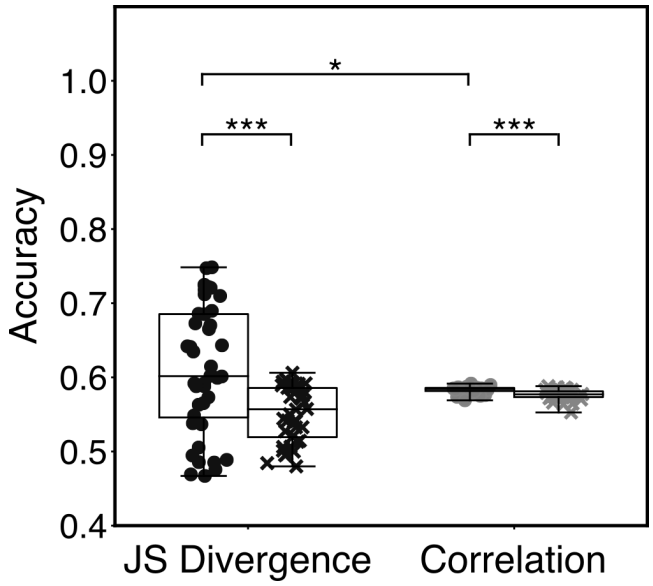


Figure 2.9: JS Divergence shows behavioral selectivity across disjoint populations. Accuracy of logistic regression for predicting whether two trials from different stimuli belong to the same behavioral class, for either JS divergence (black) or correlation (grey). Each dot corresponds to a random split of the population into two disjoint subpopulations. Xs correspond to shuffled data. Included in the regression is the JS divergence or correlation between the responses from the joint populations, indicating that simultaneous disjoint populations share some topological similarity.(*: $p < 0.05$; ***: $p < 0.001$).

Discussion

We have shown that the population activity in a secondary auditory region of a songbird contains non-trivial structure in the temporal relationships between the spiking activities of individual neurons. The population coactivity structure can be captured directly using the Curto-Itskov construction. We demonstrated that the population coactivity, as captured by its topology, is not due solely to the aggregation of independent, stimulus specific responses from individual neurons. Instead, the structure emerges intrinsically at the population level. Using novel mathematical tools to compare simulated topological population structures, we revealed an invariant structure common to distinct neural populations which corresponds to learned invariances between sensory stimuli. Using these same tools to examine in-vivo neural population activity, we showed that learned categorial relationships between natural stimuli yield invariant relationships in the topological population structure of avian auditory cortex, despite low pair-wise correlations between single-trial spike trains. From these results, we conclude that algebraic topology offers a valuable tool for understanding invariant representations in neural populations.

One concern with computing Betti curves from spike coactivations aggregated across time (temporal filtration) is that "noise" in the responses may lead to spurious coactivations that fill in (and thus destroy) the holes that the Betti curves measure. While valid, we did not consider it a severe limitation in this case because declaring any particular spikes to be noise amounts to an assumption of what constitutes "signal" – an assumption that we did not want to make. Instead, we intended for the actual observed topology to define the signal. We wanted to see how individual trials related to each other through their topology, agnostic to the precise origins of the coactivations in the population. This informed our decision to use the simplicial

complex that results from coactivations aggregated across the whole trial, rather than the most persistent homological features. This also motivated our use of time as a parameter in the Betti curves, to characterize how the simplicial complex evolved to this final state. Nevertheless, the results presented necessarily include contributions from possible "spurious" spikes, and exploration and comparison of other filtration strategies on similar datasets would be an important addition to the literature.

Our analysis differs from other approaches to population activity in that we do not begin by separating stimulus induced structure (signal correlations) from within-trial covariation (noise correlations). Our approach does not consider the population responses as variations around a mean response from individual neurons, but rather seeks to understand how a single event is represented by the population on a single 'trial'. There is increasing evidence that there exists stimulus-specific noise correlations (Gu et al., 2011; Jeanne et al., 2013; Ramalingam et al., 2013; Bondy et al., 2018) and separating these sources of variance may be an artificial convenience of the way responses have traditionally been measured. Separating these two forms of correlation at the outset of an analysis is liable to miss behaviorally relevant structure in the population response. Our results suggest that structures in the population response do carry information about behaviorally-relevant stimuli. Furthermore, these structures are not reducible to collections of individual neuron responses. Thus, these coincidence patterns carry meaning of their own.

This approach is a conceptual alternative to neural representations centered on the concept of receptive fields. By definition, computing a receptive field requires access to both neural activity and the ground-truth stimulus. Since neural networks do not have access to the stimulus outside of their own stimulus-induced activity, neural networks cannot, in principle,

perform this computation. In other words, they do not have access to their own receptive fields (Curto and Itskov, 2008). However, because natural stimuli are not random, structure in the stimulus will induce structure in the firing responses of neurons (Brette, 2015). One way this structure manifests is as temporal structure of spike coactivations in populations of neurons. Neural networks can in principle access this structure, because it requires only the ability to detect temporal coincidences among spikes within the neural population itself, and no information from outside the animal. Thus, the approach avoids some of the conceptual difficulties with receptive fields. At the same time, the approach is consistent with the idea of receptive fields. Indeed, a key insight of Curto and Itskov is that relationships between spikes captured in the topology of the population response are sufficient to reconstruct some features of the physical environment without knowledge of explicit place fields (Curto and Itskov, 2008). In other words, while receptive fields can be computed, they are not required.

Receptive field-based analyses also require strong a priori assumptions about the basis feature space for external stimulus representation. The Curto-Itskov construction allows for an alternative definition of stimulus space, as the topological space constructed from neural activity. This alternative definition offers advantages in higher order sensory systems. Like other natural signals, the relevant stimulus space for birdsong does not easily lend itself to simple (low-dimensional) parameterizations, such as, for example, the orientation angle of drifting gratings that one might use to parameterize simple stimuli for the visual system. While useful in quantifying neural responses, these experimenter-imposed parameterizations are not guaranteed to align with the dimensional axes for neural "tuning". In contrast, the intrinsic relationships between neuronal responses, regardless of their receptive field tuning, are precisely the response properties that are parameterization invariant. The topological constructions used here directly

capture some of these relationships. As for most natural stimuli, we do not have access to the "correct" parameterized stimulus space for bird song. Using the topological methods described here allows us to avoid this problem altogether, measure invariant properties of the neural representation, and reconstruct a neurally-defined stimulus space intrinsically defined by secondary auditory neurons in songbirds.

Understanding invariance is central to understanding perception, and the general notion features prominently in a longstanding debate regarding the nature of sensory systems and neural computation. The classical view, identified most strongly with Marr (Marr, 1982), conceives of sensory systems as computational pipelines, taking discrete stimuli (or stimulus features) as inputs and implementing algorithms to produce more complex representations as output. Accordingly, a class of algorithms may be specified to extract similarities from multiple stimulus representations and output an invariant representation (Riesenhuber and Poggio, 1999). A contrasting view, articulated by Gibson (Gibson, 1986), posits that sensory systems encode invariances directly, through a sensitivity to the patterns of information in the environment that define the relationships (i.e., changes or the lack thereof) between stimuli. Neuroscience has the potential to inform these deep debates on the nature of perception, but to do so we must improve the currently limited methods for quantifying invariant representations. Searching for invariants in neural activity, however, by correlating aggregate measures of neural activity with stimulus properties is problematic. As noted, the animal cannot perform the same computation (Curto and Itskov, 2008), because it does not have direct access to the stimulus. Gibson used this "closed" property of sensory systems to infer that invariant representations must be encoded directly from patterns of information in the environment (Gibson, 1986). We have demonstrated how to extract potential sensory representational invariants using only the information directly

available to the nervous system – the spatiotemporal relationships between spikes. These techniques provide a route to understanding representational invariances and their foundational role in perception.

The precise computational mechanisms for how neural populations might make use of this topological structure remain a topic for future research. However, previous studies illuminate possible avenues of investigation. In the dynamical regime of excitation-inhibition (E-I) balance, neurons can be highly sensitive to these temporal coincidences, with a sensitivity surpassing the limits of experimentally measurable statistical correlations between spike trains (Rossant et al., 2011). A state of E-I balance is commonly observed throughout the cortex of mammals and has been observed in songbird NCM neurons (Rossant et al., 2011; Perks and Gentner, 2015), and could support the use of temporal coincidences as a substrate for computation (Brette, 2012). As a general phenomenon, coincidence detection is a nearly ubiquitous capability in neurophysiological systems.

A growing body of literature investigates neural systems from the point of view of topology. Recent demonstrations that driven and spontaneous activity produce characteristic dynamics in the functional topology (Reimann et al., 2017) agree qualitatively with our findings. Likewise, persistent homology applied to spontaneous and evoked activity in cortical area V1 of monkeys is able to reconstruct the spherical topology of orientation selective cells (Singh et al., 2008), and theoretical work demonstrates how topological structure in the hippocampus can persist in the face of noise (Babichev and Dabaghian, 2017). Our work represents the first use algebraic topology to examine representational invariance in a sensory system. Combined with the earlier theoretical and empirical studies, this work helps to establish neural topology as a useful tool to investigate neural activity.

Acknowledgment

Chapter 2, in full, is a reprint of the material as it appears in Spike Train Coactivity Encodes Learned Behavioral Invariances in Songbird Auditory Cortex, Theilman, B., Perks, K., and Gentner, T.Q., The Journal of Neuroscience, 2021. The dissertation author was the primary investigator and author of this paper.

References

Babichev A, Dabaghian Y (2017) Transient cell assembly networks encode stable spatial memories. Sci Rep 7:3959.

Bailey DJ, Rosebush JC, Wade J (2002) The hippocampus and caudomedial neostriatum show selective responsiveness to conspecific song in the female zebra finch. J Neurobiol 52:43–51.

Bondy AG, Haefner RM, Cumming BG (2018) Feedback determines the structure of correlated variability in primary visual cortex. Nat Neurosci 21:598–606.

Brette R (2012) Computing with neural synchrony. PLoS Comput Biol 8:e1002561.

Brette R (2015) Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain. Front Syst Neurosci 9:151.

Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, Shenoy KV (2012) Neural population dynamics during reaching. Nature 487:51–56.

Cunningham JP, Gilja V, Ryu SI, Shenoy KV (2009) Methods for estimating neural firing rates, and their application to brain-machine interfaces. Neural Netw 22:1235–1246.

Curto C, Itskov V (2008) Cell groups reveal structure of stimulus space. PLoS Comput Biol 4:e1000205.

Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge: Cambridge University Press.

De Domenico M, Biamonte J (2016) Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison. Phys Rev X 6:041062.

De Groof G, Poirier C, George I, Hausberger M, Van der Linden A (2013) Functional changes between seasons in the male songbird auditory forebrain. Front Behav Neurosci 7:196.

Gallego JA, Perich MG, Naufel SN, Ethier C, Solla SA, Miller LE (2018) Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. Nat Commun 9:4233.

Ghrist R (2014) Elementary Applied Topology. Createspace Independent Pub.

Gibson JJ (1986) The Ecological Approach to Visual Perception. Psychology Press.

Giusti C, Ghrist R, Bassett DS (2016) Two's company, three (or more) is a simplex : Algebraic-topological tools for understanding higher-order structure in neural data. J Comput Neurosci 41:1–14.

Grace JA, Amin N, Singh NC, Theunissen FE (2003) Selectivity for conspecific song in the zebra finch auditory forebrain. J Neurophysiol 89:472–487.

Gu Y, Liu S, Fetsch CR, Yang Y, Fok S, Sunkara A, DeAngelis GC, Angelaki DE (2011) Perceptual learning reduces interneuronal correlations in macaque visual cortex. Neuron 71:750–761.

Hatcher A (2002) Algebraic Topology, illustrated, reprint. Cambridge University Press.

Horak D, Jost J (2013) Spectra of combinatorial Laplace operators on simplicial complexes. Adv Math (N Y) 244:303–336.

Ince RAA, Panzeri S, Kayser C (2013) Neural codes formed by small and temporally precise populations in auditory cortex. J Neurosci 33:18277–18287.

Jeanne JM, Sharpee TO, Gentner TQ (2013) Associative learning enhances population coding by inverting interneuronal correlation patterns. Neuron 78:352–363.

Kaczynski T, Mischaikow K, Mrozek M (2006) Computational Homology, illustrated. Springer Science & Business Media.

Karten HJ (1967) The organization of the ascending auditory pathway in the pigeon (Columba livia) I. Diencephalic projections of the inferior colliculus (nucleus mesencephali lateralis, pars dorsalis). Brain Res 6:409–427.

Kiggins JT, Comins JA, Gentner TQ (2012) Targets for a comparative neurobiology of language. Front Evol Neurosci 4:6.

Knudsen DP, Gentner TQ (2013) Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. J Neurophysiol 109:1690–1703.

Kozlov AS, Gentner TQ (2016) Central auditory neurons have composite receptive fields. Proc Natl Acad Sci USA 113:1441–1446.

Lindstrom MJ, Bates DM (1988) Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data. J Am Stat Assoc 83:1014.

Marr D (1982) Vision. Henry Holt and Co., Inc.

Mischaikow K, Nanda V (2013) Morse theory for filtrations and efficient computation of persistent homology. Discrete Comput Geom 50:330–353.

Nanda V (2013) Perseus, the Persistent Homology Software. http://www.sas.upenn.edu/~vnanda/perseus: Vidit Nanda.

Nicolini C, Vlasov V, Bifone A (2018) Thermodynamics of network model fitting with spectral entropies. Phys Rev E 98:022322.

Perks KE, Gentner TQ (2015) Subthreshold membrane responses underlying sparse spiking to natural vocal signals in auditory cortex. Eur J Neurosci 41:725–733.

Ramalingam N, McManus JNJ, Li W, Gilbert CD (2013) Top-down modulation of lateral interactions in visual cortex. J Neurosci 33:1773–1789.

Reimann MW, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, Dłotko P, Levi R, Hess K, Markram H (2017) Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function. Front Comput Neurosci 11:48.

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2:1019–1025.

Rigotti M, Barak O, Warden MR, Wang X-J, Daw ND, Miller EK, Fusi S (2013) The importance of mixed selectivity in complex cognitive tasks. Nature 497:585–590.

Rossant C, Kadir SN, Goodman DFM, Schulman J, Hunter MLD, Saleem AB, Grosmark A, Belluscio M, Denfield GH, Ecker AS, Tolias AS, Solomon S, Buzsaki G, Carandini M, Harris KD (2016) Spike sorting for large, dense electrode arrays. Nat Neurosci 19:634–641.

Rossant C, Leijon S, Magnusson AK, Brette R (2011) Sensitivity of noisy neurons to coincident inputs. J Neurosci 31:17193–17206.

Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. J Neurophysiol 86:1445–1458.

Singh G, Memoli F, Ishkhanov T, Sapiro G, Carlsson G, Ringach DL (2008) Topological analysis of population activity in visual cortex. J Vis 8:11.1-18.

Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20:2315–2331.

Thompson JV, Gentner TQ (2010) Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. J Neurophysiol 103:1785–1797.

Vates GE, Broome BM, Mello CV, Nottebohm F (1996) Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. J Comp Neurol 366:613–642.

Wang Y, Brzozowska-Prechtl A, Karten HJ (2010) Laminar and columnar auditory cortex in avian brain. Proc Natl Acad Sci USA 107:12676–12681.

Young J-G, Petri G, Vaccarino F, Patania A (2017) Construction of and efficient sampling from the simplicial configuration model. Phys Rev E 96:032312.

Chapter 3 - Direct Representation of Stimulus Space Geometry by Temporal Coactivity Patterns in Neural Populations

Abstract

Perception depends only on representations carried in the collective activity of neural populations. Yet, our understanding of these representations requires direct correspondence to stimuli external to the neural system. Recent trends in neuroscience have highlighted the importance of the geometry of neural representations of natural stimuli by a population. In higher-order sensory regions, the representational geometry is harder to capture because neuronal responses in these regions are driven by complex stimulus features. Here, recording from neurons in NCM, a secondary auditory region of the songbird, we show that the intrinsic geometry of the temporal coactivation of spiking activity carries a direct representation of complex, natural vocal communication signals. Previous work has demonstrated this correspondence in simulation, but to our knowledge this is the first demonstration of this correspondence in a sensory system *in vivo.* This validates a theoretical mechanism for representing stimulus structure in population activity consistent with individual cell receptive fields but without requiring their explicit computation.

Introduction

The collective activity of populations of neurons underlies the perception of the natural world. A central task of neuroscience is to understand how structure in neural activity relates to structure in the stimulation. These relationships are commonly approached by correlating neural activity with measures of external variables made by the experimenter. This yields the concept

of the receptive field, which is defined as the regions of sensory space that are associated with a response from a neuron (Meyer et al., 2017; Sharpee, 2013).

The receptive field concept is of fundamental importance to neuroscience. Nevertheless, receptive fields are quantities that are only meaningful for the experimenter. This is because only the experimenter has access to both the neural activity and the external stimulus needed to define the field. The brain itself does not have access to its own receptive fields (Curto & Itskov, 2008); its knowledge of the external world has to be completely contained within its own activity patterns. Therefore, to understand the neural basis of perception it is necessary to understand how the structure of external stimulation is manifested directly by the neural activity, without the requirement of explicit receptive fields.

Curto and Itskov (Curto & Itskov, 2008) recognized this important point and developed a technique for reconstructing information about the external world directly from neural activity. In the context of the hippocampal place cell system, they used temporal coincidences between neurons in a population to define a mathematical object called a simplicial complex. Using simulated neural activity, they showed in principle that the geometry of the external environment could be reconstructed from the simplicial complex associated with the neural activity. Because the simplicial complex is defined solely by temporal coincidences, this demonstrates a potential mechanism for the direct representation of stimulus information in a neural population without a receptive field intermediary. However, whether such a relationship holds for sensory systems *in vivo* is an important open question.

To test whether the temporal coactivation structure of neural activity directly represents stimulus structure, we measured population activity in region NCM, a secondary auditory region of the European starling while passively listening to natural birdsong. We computed the

simplicial complex associated to the population following the methods of Curto and Itskov. The neural topology of this population has previously been shown to carry information about learned acoustic categories (Theilman et al., 2021). To compare the geometry of the associated simplicial complexes with the structure of the acoustic stimulation, we first introduce a technique for reconstructing acoustic spectrograms from population activity and the receptive fields given by the Maximum Noise Entropy (MNE) model (Fitzgerald et al., 2011), which has been used previously to characterize the responses of individual NCM cells (Kozlov & Gentner, 2016). These reconstructions allow us to link metric relationships between receptive fields to the structure of stimulus spectrograms and vice versa. Finally, we report the relationship between these distinct topology- and receptive field- derived geometries, showing that the Curto-Itskov insight that the coactivation structure of the population contains direct information about the stimulus space holds for this sensory population.

Results

Curto-Itskov graph associated to population responses

We recorded distinct subpopulations of NCM neurons from anesthetized birds using silicon microelectrodes while presenting audio recordings of natural starling song. We binned the response of each cell to agree with the stimulus spectrogram time bins (~21 ms, Methods)) and used these binned spikes to construct the simplicial complex associated to the population. For each bin, the cells with firing rates above a certain multiple of their own lifetime average firing rate defined what is called a *cell group*. Each cell group defines an elementary simplex: pairs of coactive cells correspond to line segments, triplets to triangles, quadruplets to tetrahedra, etc. These elementary simplices are attached along their common faces, forming a larger structure

called a simplicial complex. The simplicial complex is a manifestation of the temporal coactivation structure in the population across all stimuli. With population sizes on the order of 100 units, there are ~2^100 possible cell groups. The vastness of this number means that only a fraction of the possible cell groups will be observed in an actual population, and a significant amount of information is carried in the identity of the cell groups that actually appear. The simplicial complex is a higher-order structure that keeps track of this information and its interrelations (Giusti et al., 2016).

We define a metric on cell groups using a modified form of Dijkstra's algorithm on the simplicial complex (Methods). To each simplicial complex we associate an ordinary graph, where each simplex in the complex is associated to a vertex in the graph. There is an edge between two vertices in the graph precisely when one simplex is the face of the other. All edges are weighted with unit weight. Applying Dijkstra's algorithm to this graph yields the length of the shortest path between any two simplices, and hence, cell groups.

MNE Receptive fields of NCM neurons

If the simplices correspond to the pattern of overlaps of receptive fields, then the simplicial complex should reconstruct the underlying acoustic stimulus space topologically. The cell group metric defined above adds natural geometric information to the complex. In their simulations, Curto and Itskov showed that this geometry matched the geometry of the physical environment. Similarly, we want to show that NCM cell group geometry corresponds to the geometry of the space of physical stimuli.

To do this, we computed receptive fields for NCM neurons to relate acoustic spectrograms to neural activity. Previous work has shown that the Maximum Noise Entropy

(MNE) model well-describes the sensitivity to higher-order features exhibited by NCM cells (Kozlov & Gentner, 2016).

For each single unit in each simultaneously recorded population, we independently computed the optimal parameters a, h, and J for the MNE model of the neuron's response, taken as a binary value, using 80% of the stimulus spectrograms for training. Example MNE features for a single MNE cell are shown in Figure 3.1b. After obtaining the MNE parameters for each cell on the training data, we evaluated the MNE model by predicting the response of the unit on the remaining test stimulus spectrogram segments. We evaluated the prediction by computing the correlation between the predicted response and the empirical response for that cell. Figure 3.1d shows an example response prediction for a single cell and stimulus.

To evaluate the quality of the MNE model, we again computed MNE parameters but from shuffled spike trains. Using the MNE parameters obtained from shuffled spike trains, we evaluated the correlation of the predicted shuffled response to the actual shuffled response.

Figure 3.1e shows the distribution of correlations between predicted and actual responses for both empirical and shuffled data. Consistent with previous results, we found that the MNE model accurately predicted individual cell's responses to held-out stimuli, and that individual neurons had composite receptive fields (Kozlov & Gentner, 2016).

The receptive fields obtained from shuffled datasets show poor prediction accuracy, demonstrating that the MNE model is picking up the statistical relationships between individual cell's activity and the spectrographic features of the stimulus.

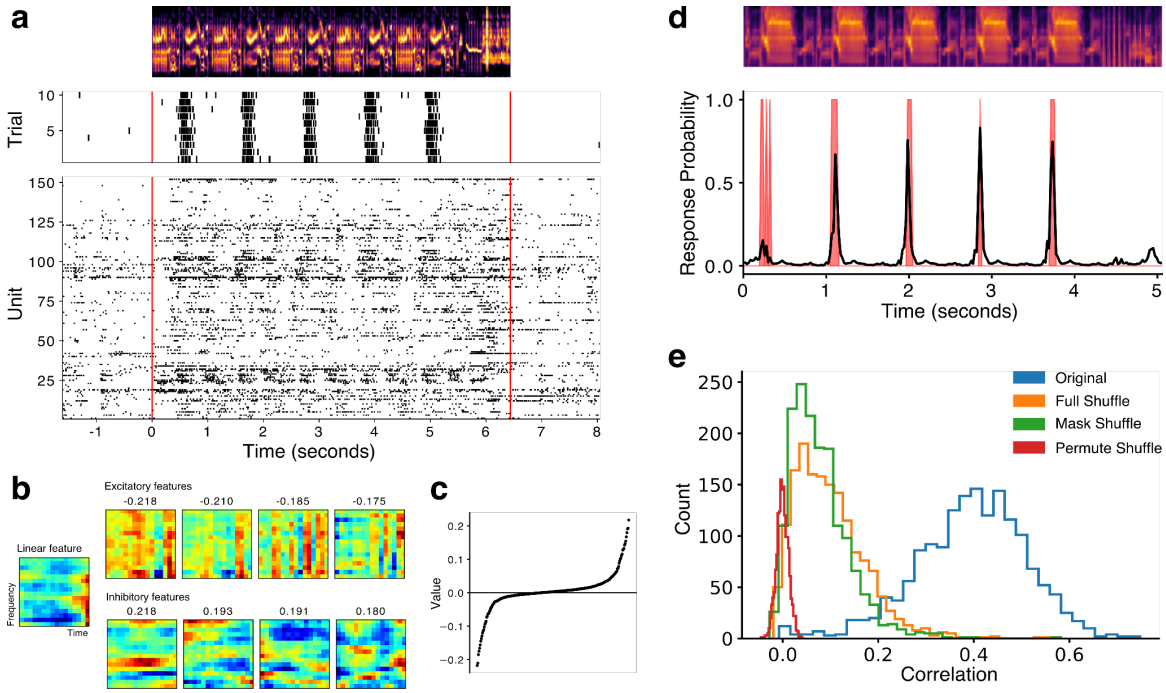


Figure 3.1: MNE receptive fields for NCM populations. a) top: Spectrogram of an acoustic stimulus composed of segments of natural starling song. Middle: Spike raster showing the response of a single NCM cell to multiple presentations of the stimulus above. NCM cells typically show sparse, highly selective responses to natural birdsong. b) MNE composite receptive field for a single NCM cell. The linear feature corresponds to the MNE h parameter. The excitatory quadratic features (top row) are the eigenvectors associated with the four most negative eigenvalues of the MNE J matrix. Likewise, the inhibitory features are associated with the four most positive eigenvalues. c) The eigenvalue spectrum of the MNE J matrix for a typical cell. NCM cells display significant positive and negative eigenvalues. d) Response prediction by the MNE model. The black curve shows the predicted firing probability to the lower resolution stimulus shown above. The red curve shows the binned, binarized, trial averaged response of the cell. e) Distribution of correlations between predicted and actual response across all cells and all birds (N=1,235 units). Each color represents a different shuffle condition (Methods). The MNE better predicts the responses for original data compared to shuffled data (Mann-Whitney U Test, $p<1e-20$ for all conditions).

Reconstructing spectrograms using a spherical subspace

The receptive field for a neuron allows the experimenter to know which features in the stimulus are associated with that neuron's response. In principle, knowledge of these receptive fields allows for the stimulus to be reconstructed from the response. For NCM cells, this is more complicated because the MNE model reveals that multiple features drive or inhibit the response from individual NCM cells. Thus, to reconstruct the acoustic spectrogram from NCM activity, some method for disambiguation must be specified. One approach is to aggregate the features from multiple neurons, and use the population itself for disambiguation.

The product of the response probabilities of each neuron in a population gives the probability of the population response under the assumption of conditional independence between the neurons. Since these probabilities depend on the stimulus, the population response probability equivalently defines the likelihood of a stimulus given a population response. Maximizing this likelihood over the space of spectrograms yields a reconstructed stimulus.

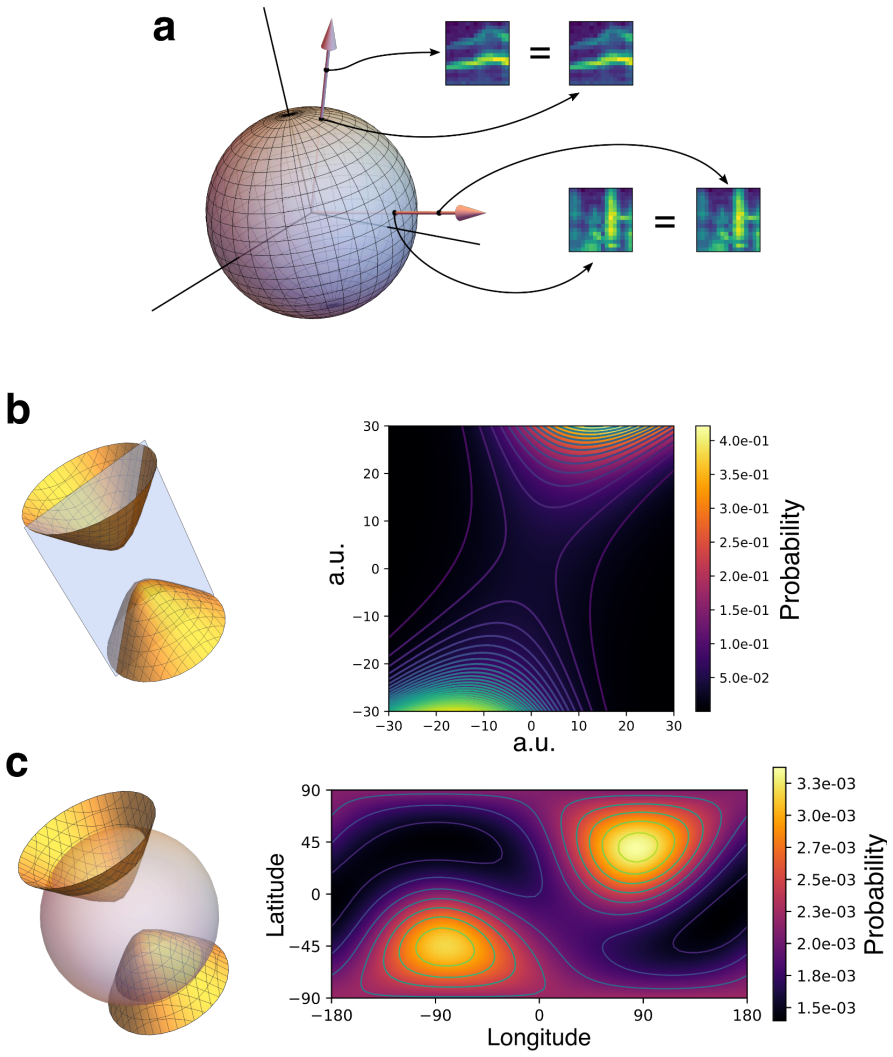


Figure 3.2: Spherical projection of MNE models yields classic, convex receptive fields. a) Along any ray in the 156 dimensional space of spectrogram segments, two points are related by a scalar multiplication. This transformation preserves the ratios of spectrotemporal power in different bins. Thus, these two spectrograms have the same "shape". b) The quadratic kernel of the MNE model implies that surfaces of constant firing probability for NCM cells are high-dimensional hyperboloids. Evaluating the probability on a two-dimensional slice (for visualization) reveals a probability density that never attains a maximum, preventing optimization of the likelihood. The axis of the hyperboloid is aligned with a ray in spectrogram space, corresponding to an equivalence class of spectrogram segments. c) Evaluating the model instead on the sphere reveals convex regions of high firing probability that resemble class receptive fields or place fields. The model shows selectivity to the antipodal point because the MNE kernel is quadratic. This implies the further identification of antipodal points, reducing the sphere to real projective space and implying the need for the Fubini-Study metric.

Naive maximization of the product of MNE models over an NCM population fails, however, because the likelihood function never attains its maximum over the full space of spectrograms. Careful consideration of the geometry of the MNE model reveals that the full space of spectrograms is, in fact, not the appropriate space for the optimization.

For NCM cells, the observation of significant positive and negative eigenvalues in the J matrix spectrum (Figure 3.1c) means that surfaces of constant probability are high-dimensional hyperboloids. Projecting these hyperboloids onto planes reveals that the quadratic kernel is unbounded, spoiling optimization (Figure 3.2b). Instead, projecting the MNE model to a sphere in the high-dimensional spectrogram space reveals convex regions of high spiking probability, resembling classic receptive fields seen in other sensory domains (Figure 3.2c). Mathematically, the likelihood function is a continuous function. Restricting to a compact domain, such as the sphere, guarantees it obtains its maximum.

Restricting spectrograms to a sphere has a natural interpretation. For all spectrograms, the unit used to measure energy in a particular time-frequency dimension is arbitrary. The brain itself cannot depend on any particular system of units. Thus, the relative values of energy in different bins define the 'shape' of the spectrogram. This means that two spectrograms should be regarded as equivalent if they are positive scalar multiples of each other. The sphere is precisely the space of equivalence classes of spectrogram segments under this equivalence relation (Figure 3.2a).

For each binary response vector from a population of neurons, applying the maximum likelihood approach described above yields a reconstructed spectrogram segment. The reconstructed point occupies the region of overlap of the receptive fields corresponding to the spiking units in the response vector. A sequence of population response vectors yields a

sequence of reconstructed spectrogram segments (Figure 3.3a). By overlapping and adding the elements of this sequence, we obtain the reconstruction of the full spectrogram. Figure 3.3b shows example reconstructions of full spectrograms compared to the actual spectrograms. By pooling the responses of a population of cells, the reconstruction fidelity is high. We evaluated the quality of the reconstructions by computing the cosine distance between the reconstructed and original spectrogram segments (Methods). Figure 3.3c shows the reconstruction performance across populations compared to the shuffled MNE models, showing that this method reliably allows for high-fidelity spectrographic reconstructions from neural population activity.

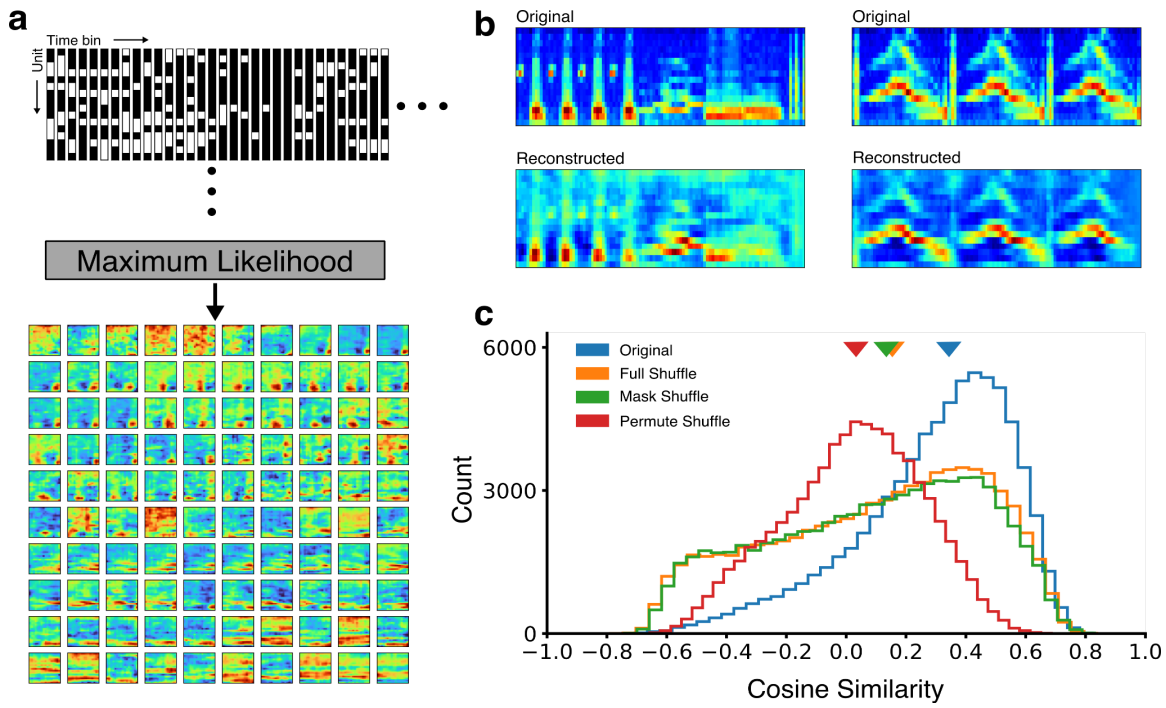


Figure 3.3: Maximum likelihood stimulus reconstructions. a) Binary response vectors from the population define a likelihood function over spectrogram segments. Maximizing this function constrained to the sphere yields reconstructed spectrogram segments (below). Each binary response is associated with a single reconstruction, even though individual cells have composite receptive fields. The reconstructions are shifted in time by the temporal offset of their response vector and added together. b) Example reconstructed full spectrograms. c) Reconstruction performance measured by the cosine similarity of reconstructed segments to the original segments for the different shuffle conditions, on held-out test data, across all populations (n = 11). Arrows indicate the median of each distribution. The reconstructions on original data outperform all shuffle conditions (Mann-Whitney U Test, p<1e-18 for all conditions).

Relationship between Receptive Fields and Neural Topology

The maximum likelihood procedure associates a spectrogram segment to each cell group. These spectrograms live on the surface of a high-dimensional sphere. The metric on the surface of this sphere defines an alternative metric on cell groups, where the distance between cell groups is defined as the distance between their reconstructed spectrograms. However, there is an additional symmetry that affects the definition of distance between spectrograms. Antipodal points on the sphere have the same spectral power ratios that define the "shape" of the spectrogram, but inverted in sign. Because the MNE model is quadratic, it cannot distinguish between these points. This is also demonstrated by the appearance of two peaks in Figure 3.2c Thus, antipodal points of the sphere must be identified. This yields what is known as real projective space (Hatcher, 2002), and the spherical metric is replaced with the Fubini-Study metric on projective space (Study, 1905). We used the Fubini-Study metric to compute the distances between pairs of cell groups induced by their spectrographic reconstructions. This metric is analogous to the physical Euclidean distance between points in space used by Curto and Itskov (Curto & Itskov, 2008).

Our hypothesis that the temporal coactivation pattern in the NCM population corresponds to relationships between receptive fields implies that the topological and spectrographic metrics on cell groups should be strongly related. This would show that the geometry intrinsic to the population defined by temporal coactivations directly represents stimulus structure without the explicit mappings to external stimuli carried by receptive fields. Figures 3.4a and 3.4b show distance matrices for a subset of 32 cell groups computed using the graph distance (3.4a) or Fubini-Study metric (3.4b). We observed a consistent relationship between these two metrics, shown in the example in Figure 3.4c. To compare these two metrics, we measured the correlation

between the distances between pairs of cell groups given by the two metrics. We found strong correlation between the two distance matrices across different populations and birds, and that this correlation was weakened under the shuffle conditions (Mann-Whitney U Test, $p<1e-80$ in all cases) (Figure 3.4d).

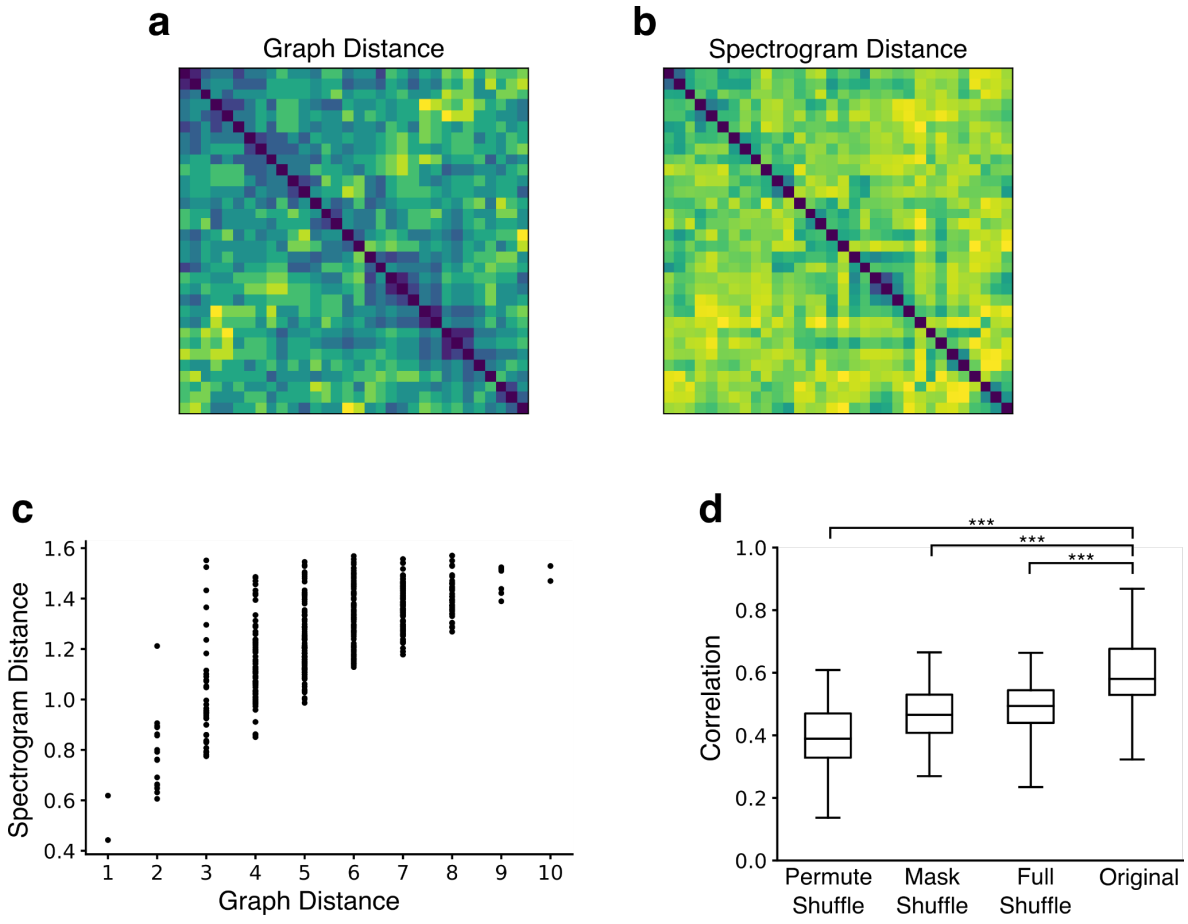


Figure 3.4: Cell group geometry corresponds to spectrogram geometry. a) Distance matrix between 32 randomly chosen cell groups computed using the shortest path length in the Curto-Itskov simplicial complex. b) Distance matrix between the reconstructed spectrogram segments corresponding to the cell groups in a, computing using the Fubini-Study metric. c) Scatter plot of spectrogram distance vs. graph distance showing the correlation between the two geometries. The maximum possible distance between points under the Fubini-Study metric is pi/2. d) Distribution of correlations between the graph distance matrices and spectrogram distance matrices across all populations. The correspondence is strongest for the original data, which preserves both the temporal coactivity structure in the population and the relationship between that structure to the stimulus (Mann-Whitney U Test, $p<1e-80$ in all cases).

Discussion

We have shown that the temporal coactivation structure of neural population activity in a secondary auditory region contains a direct manifestation of stimulus structure that does not rely on the computation of receptive fields. The intrinsic geometry of the population given by the simplicial complex and associated graph metric correlates with the geometry of the reconstructed spectrograms under the Fubini-Study metric. These reconstructions depend on the relationships between receptive fields, and the receptive fields of individual neurons induce the representational geometry (Kriegeskorte & Wei, 2021). The correspondence of the two geometries shows that the relationships between receptive fields correspond to the relationships between simplices in the simplicial complex. This validates the central insight of Curto and Itskov (2008), for the first time in a sensory system *in vivo.* While we have demonstrated the intrinsic geometry of the population coactivity structure relates to the geometry of the receptive fields, we note that this relationship is nonlinear and complex for an auditory sensory system. A deeper investigation into the mathematics of the reconstruction procedure will reveal the precise nature of this relationship.

This demonstration opens up promising new lines of theoretical and experimental investigation. Because the temporal coactivation structure of population activity is determined by the biophysics of the underlying network, we have demonstrated a mechanism by which stimulus information can be processed directly by spikes (Brette, 2015). What is missing is an understanding of the emergence of precise temporal patterns in neural populations. Recent work finding evidence for polychronization (Izhikevich, 2006) in the songbird motor nucleus HVC (Egger et al., 2020) suggests a plausible theoretical route for understanding how stimulus-

evoked, temporally-precise patterns of activity are generated by neural populations. The relationship between neural topology and polychronization is a priority for future research.

Careful consideration of the relevant symmetries of the space of acoustic spectrograms revealed that the appropriate stimulus space was not the full high-dimensional space of spectrograms, but a spherical subspace. This yielded a well-defined optimization procedure for reconstructing complex, natural, acoustic stimuli from population spiking activity. The approach we implemented is an improvement over other methods for reconstructing acoustic stimuli as it does not require knowledge of the stimulus autocovariance (Mesgarani et al., 2009). We expect that this procedure will facilitate future experiments probing how population level representations of complex acoustic stimuli change during processes such as learning or sleep.

Methods

Electrophysiology

All protocols were approved by the UC San Diego IACUC. Five birds were used in this study. All birds were wild-caught in southern California. We did not control for the sex of the subjects. Naive European starlings were anesthetized with urethane (0.7 mg/kg) and head-fixed in an acoustic isolation chamber (Acoustic Systems). A small craniotomy was opened over the region NCM. We placed multi-channel silicon electrodes (Masmanidis 128DN or 64F) (Yang et al., 2020) in NCM until auditory-evoked activity was observed on the majority of channels. After placing the electrode, we left the bird in silence for 30-60 minutes prior to starting trials. We repeated each stimulus 20 times in a random order with a random inter-trial interval of 2-5 seconds. Once a block of 1200 trials finished, we advanced the electrode by one probe-length

further into NCM to record another block from an independent population of cells. Recording blocks were obtained from both hemispheres.

Spike Sorting

Simultaneously recorded blocks were spike sorted with Kilosort2 (Stringer et al., 2019). After automatic sorting, putative units were manually curated using the phy interface. Units were manually sorted into putative single units and multiunit activity. We considered units well-isolated if they showed fewer than 1% refractory period violations with a period of 2 ms. Only putative single units were used for further analyses.

Preprocessing acoustic stimuli

A library of 60 stimuli were compiled for these experiments. The stimuli consisted of 10 full starling songs (30s - 1 min in length) and 50 song segments (3 - 6 seconds in length). Stimuli were downsampled to 24 kHz and then the STFT spectrogram was computed with an NFFT=128, a Hanning window of length 128, and 50% window overlap. We excluded the DC component and the spectrogram magnitudes were log scaled. Spectrograms were averaged as in (Kozlov & Gentner, 2016): time bins were averaged pairwise 3 times and frequency bins were averaged pairwise twice. This yielded spectrograms with 16 frequency bins and with time bins approximately 21 ms in length. Spectrograms were divided into overlapping segments of 16 time bins in length (~300 ms) to form 16 x 16 (256) dimensional spectrogram segments that serve as the stimulus for the MNE algorithm. We projected each stimulus segment to a sphere of fixed radius R = 10 by normalizing each stimulus vector.

Preprocessing spiking data for MNE

For each single unit and for each trial, the unit's spike train was binned in time to correspond with the output of the STFT spectrogram described above. Likewise, the temporal spike bins were added pairwise along with the spectrograms to yield bins ~21 ms in length indicating the number of spikes in from this neuron in that bin. We computed the trial average spike count in each bin. Finally, responses were binarized by setting bins with a non-zero average spike count to 1, leaving the remaining bins at 0.

Shuffling spiking responses

To isolate the effects of different levels of spatiotemporal organization, we computed shuffled versions of the population responses that broke different levels of spatiotemporal correlation between individual cell's responses and between cells and the stimuli.

The first shuffle we term the "full shuffle" condition. For each cell and each trial, we permuted the cell's spike times. This preserves the total number of spikes from each cell on each trial, but destroys cell-stimulus and cell-cell correlations.

Next, we computed the "masked shuffle" by generating a random "mask" for each cell that assigns spikes from that cell to new time bins. We applied the same cell-specific masks across all trials. This shuffle breaks inter-neuronal coordination but generates a new stimulus-specific spike train for each cell.

Finally, we computed the "permuted shuffle" by shuffling each time bin across the entire stimulus set, while preserving the spiking coactivity across all cells within a single time bin and trial. This shuffle destroys stimulus-locked responses but preserves the temporal coactivation structure of the population response.

Fitting MNE models

MNE models were fit according to the methods described in (Fitzgerald et al. 2011). We divided the data for each population into training and test sets, with the training sets consisting of 80% of the data. We estimated parameters using a jackknife procedure; averaging estimates from two subsets of the training datasets to yield the final parameters. To regularize the estimates, we used early stopping.

Predicting responses from MNE parameters

To predict the response of a unit with the MNE model, the a, h, and J parameters computed from the training set were applied to test stimulus spectrogram segments. This generated a time series of predicted spiking probability. To evaluate the prediction, we computed the correlation between the predicted spiking probability and the empirical response given by the binarized trial-average spike counts (Figure 3.1d).

Reconstruction of stimulus spectrograms from MNE models

For a population of N units, we write the MNE model associated to unit i as (Fitzgerald et al., 2011):

$$p_i(r_i|s) = \frac{\exp(a + h \cdot s + s^T J s)^{1-r_i}}{1 + \exp(a + h \cdot s + s^T J s)}$$

Where r_i is either 1 or 0 if unit i spikes or is silent, respectively. Assuming conditional independence of the units, the probability of an arbitrary response vector from the population as the product of individual cell MNE models:

$$p(r|s) = \prod_i p_i(r_i|s)$$

This gives the probability of a response pattern given a stimulus, but is also interpreted as the likelihood of a stimulus given a response pattern.

Taking the logarithm of this expression yields the log likelihood:

$$\log p(r|s) = \sum_i \left[(1 - r_i)(a + h \cdot s + s^T J s) - \log(1 + \exp(a + h \cdot s + s^T J s))\right]$$

Given a fixed binary response vector, maximizing this function with respect to the stimulus dimensions, with the stimulus constrained to a sphere of radius R, yields the maximum likelihood estimate, or reconstruction, of the stimulus spectrogram segment associated to this response vector.

$$\hat{s} = \underset{s}{\operatorname{argmax}} \log p(r|s)$$

$$|s| = R$$

We implemented this constrained optimization using TensorFlow and the RiemOpt library for optimization on Riemanninan manifolds (Smirnov, 2021).

Given a sequence of population response vectors, applying the above maximum likelihood procedure yields a sequence of spectrogram segments. The full reconstructed spectrogram results from overlapping and adding the elements of this sequence of spectrogram segments, with each segment shifted according to the temporal position of the associated response vector.

Evaluation of spectrogram reconstructions

Because we consider spectrograms the same if they are scalar multiples, our evaluation of reconstruction quality must be invariant to this transformation. We used the cosine distance to

evaluate the similarity between two spectrograms. The cosine distance between spectrograms u and v is given by:

$$1 - \frac{u \cdot v}{||u||_2 ||v||_2}$$

Constructing the simplicial complex associated to population activity

Binary spiking responses were binned to correspond with the spectrogram segments described previously. Within each time bin, the co-active cells define the cell group associated with that bin. Each cell group is assigned an elementary simplex corresponding to the number of cells in the cell group: line segments represent pairs of cells, triangles represent triplets, tetrahedra represent quadruplets, and so on. The vertices of each elementary simplex were labelled with the corresponding cell ID. Elementary simplices were glued along their common faces/vertices.

Computing the topological graph-distance metric on cell groups

To define a metric on cell groups, we follow (Curto & Itskov, 2008) and define a graph from the simplicial complex. Each simplex in the complex is assigned a vertex in the graph, and an edge links two graph vertices if the two simplices associated to each vertex are faces of one another.

We computed distances in this graph using Dijkstra's algorithm to find the length of the shortest path that connects any two vertices. Since each vertex in the graph corresponds to a cell group, this yields a (discrete) metric on the cell groups. The population sizes we obtained produced simplicial complexes too large to compute in their entirety. To overcome this limitation, we ranked cells according to the spectral radius of their MNE J-matrix as a measure for the strength of stimulus modulation. We computed distances for the Curto-Itskov graph

associated with the top 15% of cells in each population by spectral radius. This yielded a graph that was small enough to fit in memory but large enough to demonstrate the existence of the geometric relationships. Still, these graphs were large (~5000 cell groups) and dense, making computing the shortest path between all pairs of simplices computationally expensive. To overcome this limitation, we selected random subsamples of 32 cell groups from the full set and computed the 32 x 32 pairwise distances between the elements of these subsets.

Computing the Fubini-Study metric on cell groups

On a sphere, the great circle distance is the shortest distance between any two points. It is related to the angle between two points:

$$d_{\text{sphere}}(x, y) = \arccos\left(\frac{x \cdot y}{|x||y|}\right)$$

When antipodal points of the sphere are identified under the symmetry x = -x, this spherical metric restricts to the Fubini-Study metric:

$$d_{\text{Fubini-Study}}(x, y) = \arccos\left(\frac{|x \cdot y|}{|x||y|}\right)$$

The Fubini-Study metric has a maximum value of pi/2.

To compute the Fubini-Study metric on cell groups, we reconstructed the spectrogram segments associated to each cell group in the subsets taken from the full Curto-Itskov graph, and used the above expression to compute the distance between spectrogram segments.

Comparing graph and spectrogram geometries

Once we computed the distance matrices for each geometry and each subset, we computed the correlation between the independent components of each matrix excluding the

main diagonal. We pooled these correlations across populations for each shuffle condition to assess the significance of the correlation between the geometries.

Acknowledgements

Chapter 3, in full is currently being prepared for submission for publication and will include Professor Timothy Q. Gentner as the senior author and Michael Turvey and Srihita Rudraraju as coauthors. The dissertation author was the primary investigator and author of this material.

References

Brette, R. (2015). Philosophy of the Spike: Rate-Based vs. Spike-Based Theories of the Brain. *Frontiers in Systems Neuroscience*, *9*. https://doi.org/10.3389/fnsys.2015.00151

Curto, C., & Itskov, V. (2008). Cell Groups Reveal Structure of Stimulus Space. *PLOS Computational Biology*, *4*(10), e1000205. https://doi.org/10.1371/journal.pcbi.1000205

Egger, R., Tupikov, Y., Elmaleh, M., Katlowitz, K. A., Benezra, S. E., Picardo, M. A., Moll, F., Kornfeld, J., Jin, D. Z., & Long, M. A. (2020). Local Axonal Conduction Shapes the Spatiotemporal Properties of Neural Sequences. *Cell*, *183*(2), 537-548.e12. https://doi.org/10.1016/j.cell.2020.09.019

Fitzgerald, J. D., Sincich, L. C., & Sharpee, T. O. (2011). Minimal Models of Multidimensional Computations. *PLOS Computational Biology*, *7*(3), e1001111. https://doi.org/10.1371/journal.pcbi.1001111

Giusti, C., Ghrist, R., & Bassett, D. S. (2016). Two's company, three (or more) is a simplex. *Journal of Computational Neuroscience*, *41*(1), 1–14. https://doi.org/10.1007/s10827-016-0608-6

Hatcher, A. (2002). *Algebraic Topology*. Cambridge University Press.

Izhikevich, E. M. (2006). Polychronization: Computation with Spikes. *Neural Computation*, *18*(2), 245–282. https://doi.org/10.1162/089976606775093882

Kozlov, A. S., & Gentner, T. Q. (2016). Central auditory neurons have composite receptive fields. *Proceedings of the National Academy of Sciences*, *113*(5), 1441–1446. https://doi.org/10.1073/pnas.1506903113

Kriegeskorte, N., & Wei, X.-X. (2021). Neural tuning and representational geometry. *Nature Reviews Neuroscience*, *22*(11), 703–718. https://doi.org/10.1038/s41583-021-00502-3

Mesgarani, N., David, S. V., Fritz, J. B., & Shamma, S. A. (2009). Influence of Context and Behavior on Stimulus Reconstruction From Neural Activity in Primary Auditory Cortex. *Journal of Neurophysiology*, *102*(6), 3329–3339. https://doi.org/10.1152/jn.91128.2008

Meyer, A. F., Williamson, R. S., Linden, J. F., & Sahani, M. (2017). Models of Neuronal Stimulus-Response Functions: Elaboration, Estimation, and Evaluation. *Frontiers in Systems Neuroscience*, *10*, 109. https://doi.org/10.3389/fnsys.2016.00109

Sharpee, T. O. (2013). Computational Identification of Receptive Fields. *Annual Review of Neuroscience*, *36*(1), 103–120. https://doi.org/10.1146/annurev-neuro-062012-170253

Smirnov, O. (2021). TensorFlow RiemOpt: A library for optimization on Riemannian manifolds. *ArXiv:2105.13921 [Cs]*. http://arxiv.org/abs/2105.13921

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, *364*(6437), eaav7893. https://doi.org/10.1126/science.aav7893

Study, E. (1905). Kürzeste Wege im komplexen Gebiet. *Mathematische Annalen*, *60*(3), 321–378. https://doi.org/10.1007/BF01457616

Theilman, B., Perks, K., & Gentner, T. Q. (2021). Spike Train Coactivity Encodes Learned Natural Stimulus Invariances in Songbird Auditory Cortex. *Journal of Neuroscience*, *41*(1), 73–88. https://doi.org/10.1523/JNEUROSCI.0248-20.2020

Yang, L., Lee, K., Villagracia, J., & Masmanidis, S. C. (2020). Open source silicon microprobes for high throughput neural recording. *Journal of Neural Engineering*, *17*(1), 016036. https://doi.org/10.1088/1741-2552/ab581a

Chapter 4 - An Algebraic Approach to Polychronization

4.1 Introduction

Natural behavior depends upon the temporal dynamics of neural activity. These dynamics ultimately depend upon the physical structure of the neural network. One of the central aims of neuroscience is to understand how the structure of a neural network determines its temporal patterns of neural activity, and how these patterns of activity determine behavior.

An important example of a complex, natural behavior depending on precise temporal patterns of activity is the songbird vocal motor system. In the avian brain, nucleus HVC is critical for the production of natural birdsong. Numerous experiments have shown that during song production, HVC neurons exhibit sparse, temporally precise bursts of activity (Hahnloser et al., 2002; Long and Fee, 2008; Markowitz et al., 2015). Many competing models of the HVC song circuit have been proposed to explain the emergence of these temporal patterns, but precisely how the HVC circuit generates these patterns remains a mystery (Armstrong and Abarbanel, 2016; Long et al., 2010; Verduzco-Flores et al., 2012; Yildiz and Kiebel, 2011). Recently, experimental observations have shown that song-related population activity in HVC is better explained by a neural network model that exhibits *polychronization*, a mechanism by which spikes in a neural network interact with long, heterogeneous conduction delays between neurons to reinforce specific temporal patterns of activity over others (Egger et al., 2019). Polychronization is a promising mechanism to explain the origin of temporal patterns in neural activity in a variety of systems, and warrants deeper theoretical and experimental investigations.

Polychronization was introduced by Izhikevich (Izhikevich, 2006) as a mechanism by which networks generate time-locked but asynchronous patterns of activity. These patterns arise from heterogeneous conduction delays between neurons. Heterogeneous delays cause specific temporal orderings of presynaptic spikes to arrive simultaneously at a postsynaptic neuron. This large coincident synaptic drive produces a spike in the postsynaptic neuron. This postsynaptic spike interacts with its own axonal delays and other spikes in the network to continue the pattern. These self-reinforcing time-locked patterns of spiking activity are called *polychronous groups*.

The study of polychronization is challenging because in even modestly sized networks, the number of polychronous groups grows immensely, making the enumeration of individual polychronous groups difficult (Chrol-Cannon et al., 2017; Izhikevich, 2006; Pauli et al., 2018). This property is often used to support the representational capacity of polychronous networks, but also makes such networks resistent to theoretical analysis. What is needed is a precise mathematical characterization of polychronization, polychronous groups, and their relationship to network structure. Such a characterization links physical network structures with their spiking repertoires and provides a foundation for relating temporal patterns of activity to representation and ultimately, behavior.

This work introduces an algebraic description of polychronous groups as spaces of solutions of systems of polynomial equations. We apply ideas from category theory to formulate the compositional structure of polychronous groups, which allows us to construct a category-theoretic description of all the possible polychronous groups that make up the spiking repertoire of a given network. The mathematical description we provide formalizes the concept of

polychronization and will aid in further theoretical investigations of the phenomenon itself and its relationship to spiking dynamics in biological neural networks.

4.2 Polychronous Groups

Intuitively, a polychronous group is a specific temporal pattern of spikes that interacts with the physical structure of a spiking neural network to sustain itself. In Figure 4.6b, spikes from neurons 1 and 4 arrive simultaneously at neuron 2, causing neuron 2 to spike. This spike travels to neuron 3, arriving at the same time the original spike from neuron 1 reaches neuron 3. The simultaneous arrival of these spikes causes neuron 3 to spike. This spike may go on to interact with more neurons, continuing the pattern. Any time neurons 1 and 4 spike with that same temporal offset, the rest of the sequence will follow with the same temporal relationships, as long as the conduction delays and synaptic weights remain the same. In this way, we can think of this pattern as a unified whole, a "polychronous group." Thus, the spikes defining a polychronous group are locked together by the physical structure of the spiking network into precise temporal relationships. However, these groups have an important degree of freedom: their occurrence in time. The pattern of spikes defining a polychronous group specifies when spikes must occur in relation to each other, but not their absolute time. Given a set of n spikes, we need n real numbers to specify when each spike occurs. Another way to specify the spikes is to arbitrarily pick one spike as a "reference" spike. Every other spike can be specified by its time of occurrence relative to this reference spike. These relationships require $n - 1$ real numbers. By specifying the absolute time of occurrence of the reference spike, we automatically specify the absolute times of the other spikes. By leaving its absolute time unspecified,

there is a 1-dimensional space of possible times for the reference spike. Each one of these possible reference spike times yields a set of absolute spike times for the rest of the spikes in the group that respect the temporal relations between spikes. Thus, these spikes behave as a cohesive pattern that can occur anywhere in time: a polychronous group. This consideration allows us to give a precise definition of a polychronous group, namely, a particular kind of one-dimensional subset of a space of spike times.

The set of possible spike times for a set of n spikes is in bijective correspondence with the points of $\mathbb{R}^n$. A polychronous group is a set of spikes with fixed offsets relative to a reference spike. For concreteness, consider 3 spikes and label their spike times as t_1, t_2, t_3. Arbitrarily choose t_1 to be the reference spike. Then, the fixed temporal offsets imply

$$t_2 = t_1 + d_{21}$$

$$t_3 = t_1 + d_{31}$$

for some real numbers d_{21} and d_{31}. Choosing a parameter t, interpreted as the time of occurrence of the first spike, these relationships can also be expressed as

$$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ d_{21} \\ d_{31} \end{bmatrix}$$

This is the equation for a line in $\mathbb{R}^3$. We could use many different parametrizations of this line, but it is the geometry of this line, not the parameterization, that defines the polychronous group.

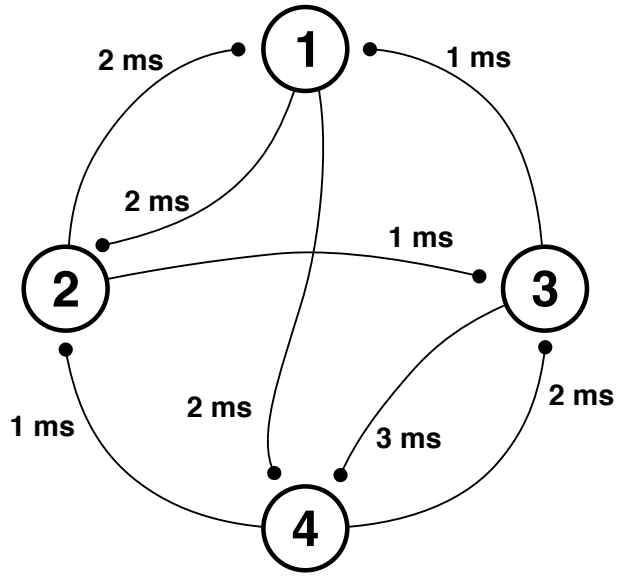


Figure 4.1: An example minimal polychronous network. The times next to connections represent axonal conduction delays. While given in milliseconds here, the unit is arbitrary. This network is supplemented by the spiking condition that any two coincident spikes on a neuron cause that neuron to spike.

4.3 Minimal Polychronous Model

In this section we give a formal definition of the minimal spiking network model in (Izhikevich, 2006) that is used to demonstrate polychronization. This model is the simplest model that exhibits polychronization, and so we refer to it as the Minimal Polychronous Model.

A spiking neural network consists of neurons and one-way synaptic connections between them. Thus, specifying the model starts by specifying a directed graph, consisting of a set V of neurons (vertices), and a set E of synapses (edges). The orientation of the synapses is specified by a pair of functions between these sets, $s, t : E \to V$. The function s maps each edge to its source vertex, and the function t maps each edge to its target vertex. In neuroscience parlance the function s maps a synapse to its presynaptic neuron, and the function t maps a synapses to its postsynaptic neuron. The physical structure of the graph is determined entirely by these sets and functions. The spiking dynamics are determined by additional structures defined over the directed graph.

The first additional structure is the set of synaptic delays. These delays are represented by positive real numbers associated to each synapse. Mathematically this requires an additional function $d : E \to \mathbb{R}^+$ that assigns a delay to each synapse in the directed graph.

The next additional structure is a set of rules for transforming presynaptic spikes into postsynaptic spikes. In most spiking neural networks, this is given by differential equations that describe the evolution in time of the various state variables associated to each neuron, along with conditions for spike initiation, reset, and the influence of synaptic conductances. There is much flexibility here, as the numerous spiking network models present in the liter-

ature can attest. However, what all spiking models have in common is that certain temporal patterns of presynaptic activity causally drive a postsynaptic neuron to produce a spike at a specific time, while other patterns do not. This is the essential property that the minimal polychronous model captures.

To capture this property, for each neuron in the model we associate a set of *spiking conditions*. A spiking condition is a subset of the set of synapses onto a given neuron. The interpretation of a spiking condition is that the coincident *arrival* of spikes from the presynaptic neurons in a spiking condition drives a spike from the postsynaptic neuron. Because coincident arrival is what counts, the synaptic delays imply that specific temporal patterns of presynaptic activity yield the required coincident arrivals of spikes to elicit a spike from the postsynaptic neuron.

The requirement of coincident arrival is related to the idea of temporal integration in spiking models. In these models, synaptic inputs drive contributions to the membrane potential that are filtered by the RC properties of the postsynaptic membrane. Due to the time constant of this filter, synaptic inputs that arrive within a certain window of one another will sum together, while those outside of these windows will not, because the contribution of the first will have decayed. Thus, the minimal polychronous model can be seen as a limit of an integrate and fire model in which these integration windows become vanishingly small. Physiological neurons are subject to thousands of synaptic inputs, causing them to enter high-conductance states (Destexhe et al., 2003). In these states, the membrane time constant is short, meaning that neural activity is significantly driven by coincident activity (Rossant et al., 2011). This motivates the simplification inherent in the definition of the spiking conditions.

Another aspect of the minimal polychronous model that differs from most other spiking models is the lack of explicit synaptic weights. Conceptually, synaptic weights are real values that quantify the degree of causal influence of presynaptic spiking activity on a postsynaptic neuron. The minimal polychronous model captures this causal relationship by directly specifying the input patterns that drive the postsynaptic neuron to spike, so a real-valued synaptic weight is not needed. However, specifying synaptic weights in typical spiking networks has ultimately the same effect: selecting particular input patterns that drive a postsynaptic spike. The spiking conditions in the minimal polychronous model are a generalization of the effect of synaptic weights. In biological neural networks, the synapse is itself a complex dynamical system, and reducing its influence to a single numerical value might introduce unnecessary or unrealistic constraints on the kinds of models and theories that are produced to account for this synaptic influence. The minimal polychronous model captures the causal structure underlying both biological synapses and numerical synaptic weights, and this generality allows it to fit between both frameworks.

We end this section with a summary of the formal definition of a minimal polychronous model.

Definition. A minimal polychronous network consists of

1. A directed graph $G = (V, E, S, T)$ with vertices (neurons) V, edges (synapses) E, a source (presynaptic) function $S : E \to V$, and a target (postsynaptic) function $T : E \to V$.

2. A function $D : E \to \mathbb{R}^+$ specifying the propagation delays along each connection.

3. For each $u \in V$, a subset $C_u \subset \{0,1\}^{|\mathbf{Pre}(u)|}$ of binary vectors that specify *spiking conditions*, or *presynaptic coincidence codes*, which are patterns of coincident presynaptic activity that cause neuron u to spike.

4.3.1 Example

Here we introduce an explicitly defined minimal polychronous model that will serve as a running example of the techniques in the rest of this paper. Figure 4.1 illustrates the model. There are four neurons, with directed connections illustrated by the arrows. The numbers adjacent to each neuron give the delay associated to that synapse, in arbitrary time units. This network structure is supplemented with the global spiking condition that any two or more coincident inputs to any neuron drive it to spike.

4.4 Synaptic Polynomials

The structure of a spiking neural network determines the dynamics of the network by constraining the possible spiking patterns. Even under these constraints, the space of possible patterns is enormous. We refer to this space of possible spiking patterns of a given network as the network's *spiking repertoire*. Since the constraints determine the repertoire, studying spiking neural networks by way of constraints is a fruitful way of managing their combinatorial enormity.

The constraints imposed by the minimal polychronous model are the finite signal propagation delays and the spiking conditions. Constraints are often synonymous with equations, and solutions of these equations satisfy the constraints. Here we write the constraints imposed

by the minimal polychronous model as special systems of polynomial equations. This links algebraic geometry to the study of polychronous spiking networks.

We associate a polynomial to each synapse in the minimal polychronous model. For an edge connecting neuron j to neuron i, we associate

$$p_{ij} = t_i - (t_j + d_{ij})$$

where d_{ij} is the conduction delay from neuron j to neuron i.

This polynomial is an element of the ring $\mathbb{R}[t_i, t_j]$. Roots of this polynomial, such that $p_{ij} = 0$, are spike times t_i and t_j that are consistent with the conduction delay between neurons i and j. Thus, this polynomial picks out possible spike times for these two neurons that satisfy a constraint imposed by the structure of the polychronous network.

Polynomials of this form are fundamental to the algebraic description of polychronization, so we give them a formal definition.

Definition. A *synaptic polynomial* is a polynomial in $\mathbb{R}[t_i, t_j]$ of the form $t_i - (t_j + d_{ij})$ with $d_{ij} \in \mathbb{R}^+$. Given a minimal polychronous model M, to each edge $e \in E$ with presynaptic neuron $i = S(e)$ and postsynaptic neuron $j = T(e)$, we associated the synaptic polynomial $t_i - (t_j + D(e))$ Given a synaptic polynomial $p_{ij} = t_i - (t_j + d)$, we say that p has *target* i and *source* j.

Since each synapse has an associated synaptic polynomial, and each spiking condition is a collection of synapses with common postsynaptic neuron, each spiking condition yields a system of polynomial equations. For each such system, the number of indeterminates is equal

to the number of presynaptic neurons, plus one for the postsynaptic neuron. However, there is only one equation for each presynaptic neuron. This means the space of solutions to this system has one degree of freedom and is geometrically a line. Specifying the absolute spike time of any neuron in the system determines the absolute spike times of all the others. Thus, this solution space is a polychronous group in the sense of the discussion in section 2. The polychronous groups emerging from individual spiking conditions constitute the most fundamental polychronous groups. All other polychronous groups must be built from the spiking conditions, because the spiking conditions determine the spiking dynamics of the network. As such, we call these "elementary polychronous groups" and give them a precise definition.

Definition. An *elementary polychronous group* on neuron i, E, is a set of synaptic polynomials $E = \{p_1, \ldots, p_k\}$ each with target i.

4.4.1 Example

Because these elementary polychronous groups identify all the ways that individual neurons can spike, every spike train necessarily decomposes into elementary polychronous groups. In Figure 4.2 we give the elementary polychronous groups associated to the minimal polychronous model depicted in Figure 4.1. There are four elementary polychronous groups in this model, since each neuron has two presynaptic partners, and a neuron is triggered by the simultaneous arrival of spikes from these partners. . Each elementary polychronous group consists of two synaptic polynomials.

Neuron 1	Neuron 2
$p_{11} = t_1 - (t_3 + 1)$	$p_{21} = t_2 - (t_1 + 2)$
$p_{12} = t_1 - (t_2 + 2)$	$p_{22} = t_2 - (t_4 + 1)$
Neuron 3	**Neuron 4**
$p_{31} = t_3 - (t_2 + 1)$	$p_{41} = t_4 - (t_1 + 2)$
$p_{32} = t_3 - (t_4 + 2)$	$p_{42} = t_4 - (t_3 + 3)$

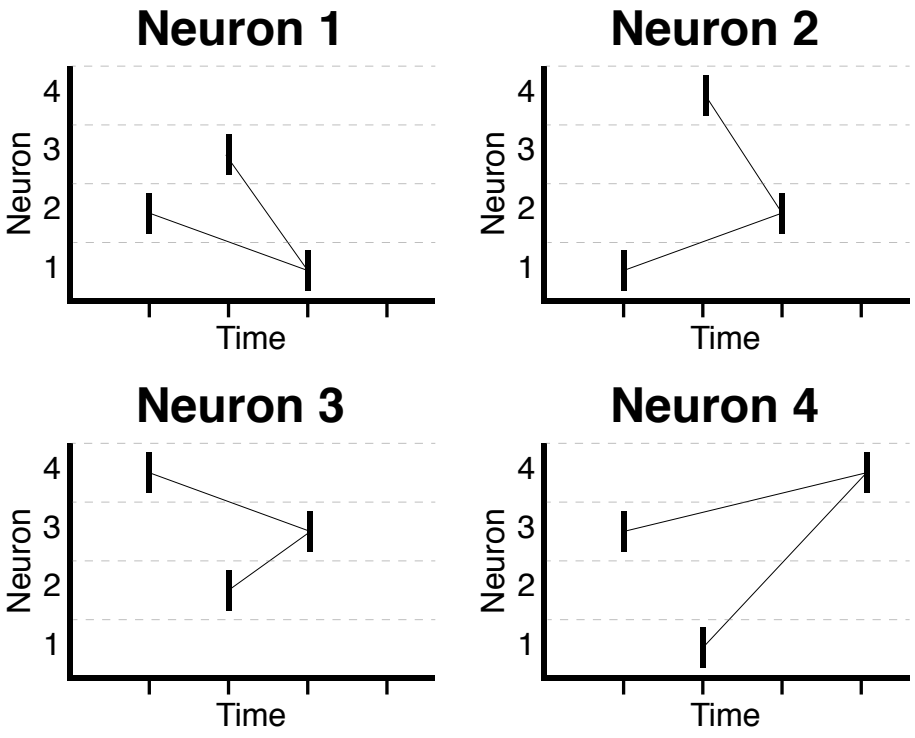


Figure 4.2: Elementary polychronous groups defined by the example network. Black lines connecting spikes represent the causal interactions leading to the final spike. The absolute times on the time axis are unspecified because the polychronous groups could happen at any time.

4.5 Polychronous Rings

We have established that polychronous groups arise from the constraints placed on the dynamics by the physical structure of the spiking network. These constraints are expressed mathematically with systems of polynomial equations. The solution spaces of these systems are geometric entities that represent the polychronous groups. We exploit the duality between algebra and geometry to give an algebraic description of polychronous groups.

Given an elementary polychronous group E_i on neuron i, we can form the set of indeterminates appearing in the system: $\{t_i, t_{j_1}, \ldots, t_{j_k}\}$. Thus, the polynomials in E_i live in the polynomial ring $\mathbb{R}[t_i, t_{j_1}, \ldots, t_{j_k}]$. The polynomials in E_i generate an ideal in this ring,

$$I_i = \{\alpha_1 p_{i1} + \cdots + \alpha_k p_{ik} | \alpha_t \in \mathbb{R}[t_i, t_{j_1}, \ldots, t_{j_k}]\}$$

Because this is an ideal, we can form the quotient ring $A_i = \mathbb{R}[t_i, t_{j_1}, \ldots, t_{j_k}]/I_i$. This quotient ring is an algebraic representation of the geometric polychronous group associated to E_i. To fix our terminology, we give definitions of elementary polychronous ideals and rings

Definition. An *elementary polychronous ideal* is an ideal generated by the polynomials defining an elementary polychronous group.

Definition. An *elementary polychronous ring* is the quotient ring R/I where $I \subset R$ is an elementary polychronous ideal.

By virtue of being quotients of commutative polynomial rings, elementary polychronous rings also have the structure of commutative algebras.

4.5.1 Example

The elementary polychronous rings obtained from the model of Figure 4.1 yield the following elementary polychronous rings

$$A_1 = \mathbb{R}[t_1, t_3, t_4]/\langle p_{11}, p_{12}\rangle$$

$$A_2 = \mathbb{R}[t_2, t_1, t_4]/\langle p_{21}, p_{22}\rangle$$

$$A_3 = \mathbb{R}[t_3, t_2, t_4]/\langle p_{31}, p_{32}\rangle$$

$$A_4 = \mathbb{R}[t_4, t_1, t_3]/\langle p_{41}, p_{42}\rangle$$

4.6 Composing elementary polychronous groups

The interesting features of polychronization come from their compositionality. If a polychronous group contains a subpattern of spikes that matches a subpattern in another polychronous group, then activation of the first polychronous group can lead to activation of the second, and the polychronous groups may be chained together. A single neuron may belong to multiple polychronous groups, and a spike from a neuron may be part of several groups simultaneously. To understand polychronization, it is necessary to have a mathematical formulation of this compositional structure. Category theory provides a language for formalizing and understanding compositionality, and in this section use category theoretic ideas to characterize the compositional structure of polychronous groups via their algebraic representations.

Given two systems with inputs and outputs, composing them means identifying ('gluing') the outputs of one system with the inputs of the other. In spiking networks, polychronous

groups are glued together via spikes: a set of spikes with a particular pattern in polychronous group A is time-aligned with and locked (glued) to a matching pattern of spikes in polychronous group B.

The ring associated to a polychronous group represents a 1-dimensional space of possible spike times that respect the pattern that defines the polychronous group. Now, the solution space for two *independent* polychronous groups must be two dimensional, as there is one free parameter for each polychronous group: the time times at which each pattern occurs. The two polychronous groups can slide past each other in time, and, importantly, overlap. Algebraically, this operation on polychronous groups is described by the *tensor product* of commutative algebras. Gluing two polychronous groups along common spikes is achieved by identifying the common spikes by means of an equivalence relation defined on this tensor product.

Definition. Given two commutative $\mathbb{R}$-algebras A and B, the *tensor product* $A \otimes_{\mathbb{R}} B$ is the commutative algebra made up of elements of the form $a \otimes b$, $a \in A$ and $b \in B$. The structure of this algebra is defined by

- $1_{A \otimes_{\mathbb{R}} B} = 1_A \otimes 1_B$
- $a \otimes b + a' \otimes b = (a + a') \otimes b$
- $a \otimes b + a \otimes b' = a \otimes (b + b')$
- $(a \otimes b)\dot{(}a' \otimes b') = aa' \otimes bb'$

There are obvious homomorphisms $A \to A \otimes_{\mathbb{R}} B$ and $B \to A \otimes_{\mathbb{R}} B$ given by $a \mapsto a \otimes 1_B$ and $b \mapsto 1_A \otimes b$. For polynomial rings A_1 and A_2, the tensor product can be obtained by forming all formal products of polynomials from each ring ensuring each ring has unique indeterminates.

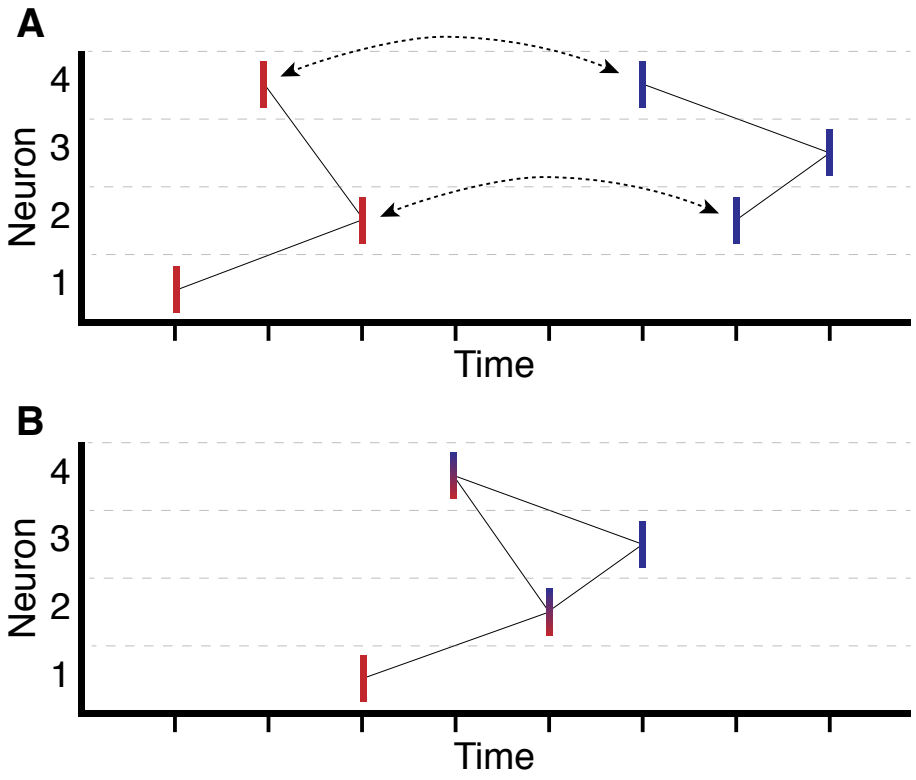


Figure 4.3: Gluing polychronous groups A_2 and A_3. Group A_2 is in red and A_3 in blue. A: Unconstrained, the two polychronous groups are free to slide past each other in time. However, regardless of their absolute timing they share common relative timing for neurons 2 and 4 (dotted arrows). The polychronization principle states that these groups can be combined into a larger polychronous group. B: The combined polychronous group obtained by constraning the spikes from neurons 2 and 4 to overlap between the two subgroups. Algebraically, this gluing is achieved by taking a quotient by another polynomial ideal. Constrained in this way, the two groups are not free to slide past each other in time, but must remain locked together, as a group.

4.6.1 Gluing constraints

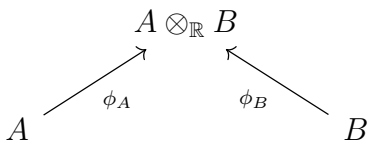
If two polychronous groups share a subpattern of spikes, then when the polychronous groups are aligned relative to each other in time such that this common pattern overlaps, the absolute spike times from the spikes in each group's matching subpattern will be identical. To lock these two groups together, we add the constraint that the times for the spikes in the common subpatterns are identical. In our running example, polychronous groups A_2 and A_3 share a common pattern defined by the spikes from neurons 2 and 4. This situation is depicted in Figure 4.3a with the dotted arrows. Because they share the same temporal pattern between the groups, the spikes from group A_2 can trigger group A_3, and result in a larger polychronous group. This gluing is depicted in Figure 4.3b, where the spikes from neurons 2 and 4 are shaded both red and blue to indicate they are shared by the two separate elementary polychronous groups.

Constraining separate spikes from neurons 2 and 4 to align amounts to taking the quotient by the ideal $\langle t_2 - s_2, t_4 - s_4 \rangle$. Thus the ring corresponding to this composed polychronous group is given by

$$A_2 \otimes_{\mathbb{R}} A_3 / \langle t_2 - s_2, t_4 - s_4 \rangle \simeq \mathbb{R}[t_2, t_1, t_4, s_3, s_2, s_4] / \langle p_{21}, p_{22}, p_{31}, p_{32}, t_2 - s_2, t_4 - s_4 \rangle$$

Iterating this process creates algebraic representations of ever larger polychronous groups by attaching smaller polychronous groups along common subpatterns. This process embodies the essence of the idea of polychronization. In the next sections, we will use category theory to systematize this process.

Colimits in the category of commutative $\mathbb{R}$-algebras contain the idea of forming the quotient of a tensor product. In fact, the bare tensor product of such algebras is a colimit of a discrete pair of $\mathbb{R}$-algebras as in the following diagram.



where ϕ_A sends $a \in A$ to $a \otimes 1_B$ and ϕ_B sends $b \in B$ to $1_A \otimes b$. The coproduct (tensor product) of two polychronous rings

$$A = \mathbb{R}[t_{i_1}, \ldots, t_{i_k}]/\langle\{p_i(t)\}\rangle$$

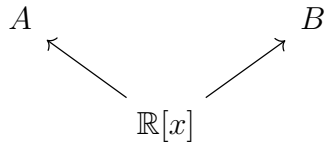
and

$$B = \mathbb{R}[t_{i_{k+1}}, \ldots, t_{i_{k+l}}]/\langle\{q_i(t)\}\rangle$$

is given by

$$A \otimes_{\mathbb{R}} B = \mathbb{R}[t_{i_1} \ldots, t_{i_k}, s_{i_{k+1}}, \ldots, s_{i_{k+l}}]/\langle\{p_i(t)\}, \{q_i(s)\}\rangle$$

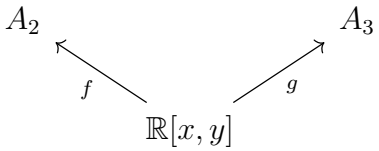
That is, the indeterminates are extended and the constraints that define each polychronous group are applied concurrently, with the indeterminates appropriately renamed. As we have seen, the coproduct corresponds to the case where two polychronouos groups are free to slide past one another in time. Now, consider a diagram such as



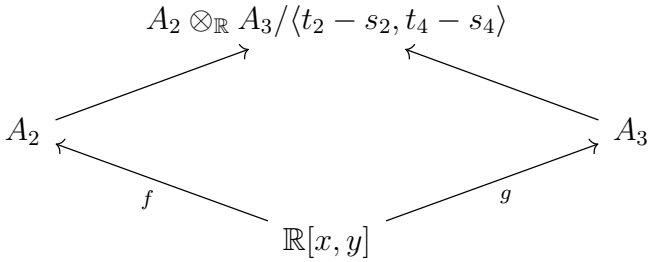
It is generally true that colimits can be described via coproducts and coequalizers. The colimit

of this diagram corresponds to the coequalizer of (diagram) The coequalizer is given by quotienting the coproduct by the ideal generated by relations of the form $f(r) - g(r)$ for $r \in \mathbb{R}[x]$.

As mentioned, gluing polychronous groups together along common spikes amounts to quotienting the tensor product of their corresponding rings by ideals generated by relations of the form $t_i - s_i$, where i indicates the neuron with spikes that match up in both groups. Consider the process of gluing groups A_2 and A_3 in the running example above. We form the diagram



The $\mathbb{R}$-algebra morphisms f and g are defined by their actions on x and y in $\mathbb{R}[x, y]$. We define f through the assignment $f(x) = t_2$, $f(y) = t_4$ and g through the assignment $g(x) = s_2$, $g(y) = s_4$, and then extend by linearity. Under this definition, the colimit for this diagram becomes



This is precisely the ring corresponding to the glued polychronous group illustrated in Figure 4.3.

From this example, we see that pushouts produce rings that result from gluing two polychronous groups along common spikes, where the common spikes are identified by the morphisms f and g. No matter which homomorphisms are used as f and g in this diagram, the colimit will exist. However, if the two polychronous groups do not align along the spikes

identified by f and g, then the resulting constraints will have no common solution. Thus, the ideal generated by these constratins will be the whole polynomial ring $\mathbb{R}[t_{i_1}, \ldots, t_{i_k}]$, and so the colimit will be the zero ring. This is the first indication of a separation between the syntax of polychronous groups and their semantics. We can always write down a hypothetical combination of polychronous groups, but it may turn out that such a combination does not actually exist, because of the constraints put on the dynamics by the structure of the network. This is similar to how we can always construct an arbitrary syntactically correct computer program in some programming language, but when compiled it may not even run.

4.6.2 Presheaves and Free Cocompletions

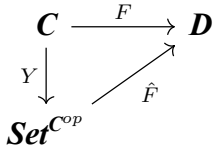
We know now that gluing polychonronous groups along common spikes can be expressed as a colimit. All polychronous groups can be built from elementary polychronous groups by constructing all the possible ways of gluing the elementary polychronous groups together. In this section, we describe constructions for freely gluing together elementary polychronous groups. This will yield a categorical description of all possible polychronous groups for a given network.

To motivate these constructions, we first consider directed graphs. As we have already seen in the definition of a minimal polychronous model, a directed graph consists of two sets and two functions between these sets. The two sets are identified with sets of edges and vertices, and the functions identify the source and target vertex associated with each edge. In some sense, all directed graphs are built from one common building block: an arrow (Vigna, 2003). Larger directed graphs are made by gluing arrows together.

Now, a *presheaf* G on this category is simply a contravariant functor from this category to the category of sets. What this means is that a presheaf is equivalent to the data of a set $G(V)$, a set $G(E)$, and two functions $G(s), G(t) : G(E) \to G(V)$. This is precisely the data that defines a directed graph. So directed graphs are equivalent to presheaves on this base category.

Changing the base category allows us to form more general structures than directed graphs. In all cases, the mechanics of the construction is the same: we use the base category to define our "building blocks" and the rules by which they join together. Then we construct every possible structure consistent with those rules. Since this construction is the same regardless of the base category used, it is an example of a universal property. Specifically, it is known as the *free cocompletion*. It should be thought of as freely adding all the possible colimits of diagrams in the base category.

Theorem 4.6.1. *Given a small category* $\boldsymbol{C}$*, a cocomplete category* $\boldsymbol{D}$*, and a functor* $F : \boldsymbol{C} \to \boldsymbol{D}$*, there is a unique functor* $\hat{F} : \boldsymbol{Set}^{\boldsymbol{C}^{op}} \to \boldsymbol{D}$ *such that the following diagram commutes*



Where Y *is the Yoneda embedding,* $Y(c) = Hom(-, c)$

Constructing a formula for the free cocompletion begins by defining the category of elements of a presheaf P, denoted $\int P$.

Definition. Given a presheaf $P : \mathbf{C}^{op} \to \mathbf{Set}$, the *category of elements* $\int P$ is the category

consisting of

- Objects are pairs (c, p) for each c an object of $\mathbf{C}$ and each $p \in P(c)$
- Morphisms $(c', p') \to (c, p)$ are morphisms $u : c' \to c$ such that $p' = Pu(p)$.

There is an important projection functor $\pi_P : \int P \to \mathbf{C}$ given by

$$\pi_P(c, p) = c$$

The category of elements of a presheaf P serves as the index category for a diagram whose colimit is isomorphic to P. It is an important theorem that every presheaf of sets on a small category can be expressed as a certain colimit:

Theorem 4.6.2. *For $\boldsymbol{C}$ a small category, every presheaf $P \in \boldsymbol{Set}^{\boldsymbol{C}^{op}}$ is isomorphic to the colimit*

$$P = \mathit{colim}\left(\int P \xrightarrow{\pi_P} \boldsymbol{C} \xrightarrow{Y} \boldsymbol{Set}^{\boldsymbol{C}^{op}} \right)$$

Now, if we define the extension functor using the formula

$$\hat{F}(P) = \operatorname{colim}\left(\int P \xrightarrow{\pi_P} \mathbf{C} \xrightarrow{F} \mathbf{D} \right)$$

then it turns out that $\hat{F}$ is a left adjoint (Mac Lane and Moerdijk, 1994), and so preserves colimits. Since every presheaf is a colimit, and colimits are unique up to isomorphism, $\hat{F}$ is unique up to isomorphism. This establishes the universal property of the free cocompletion.

General polychronous groups are built from elementary polychronous groups by gluing the elementary polychronous groups together subject to certain syntactic rules. We can use

the universal property of the free cocompletion to study all the possible polychronous groups in a given spiking network. This requires defining a new base category that captures the elementary polychronous groups and their rules for composition.

Definition. Given a minimal polychronous model, the polychronous base category **PG** is defined to have

- Objects x_i for each neuron i.
- Objects A_α for each elementary polychronous group.
- Morphisms $f_{i\alpha} : x_i \to A_\alpha$ if neuron i spikes in elementary polychronous group A_α.
- Requisite identity morphisms on each object.

Figure 4.4 illustrates the category **PG** for the running example network

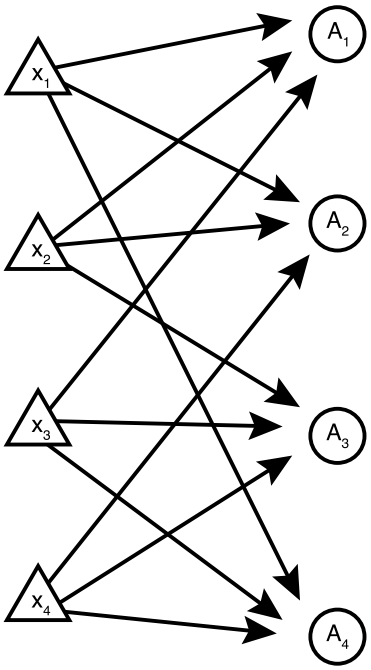


Figure 4.4: Illustration of the category **PG** for the example polychronous network.

Using the Yoneda embedding we can pass to the category of presheaves of sets on **PG**. This is exactly analogous to the construction of directed graphs, but now we have many different kinds of "edges" and "vertices" and many non-trivial relationships between them. This category is denoted $\mathbf{Set}^{\mathbf{PG}^{op}}$ and a presheaf $F \in \mathbf{Set}^{\mathbf{PG}^{op}}$ is defined by the data

- A set $F(x_i)$ for each neuron i.
- A set $F(A_\alpha)$ for each elementary polychronous group A_α.
- For each morphism $f_{i\alpha} : x_i \to A_\alpha$, a function $F(f_{i\alpha}) : F(A_\alpha) \to F(x_i)$.

Note that some of these sets may be empty. The interpretation is that the set associated to each elementary polychronous group is the set of instances of that elementary polychronous group in the larger, glued polychronous group. The sets associated to each neuron represent the spikes from that neuron along which multiple elementary polychronous groups may be joined. The functions mapping the elementary polychronous group sets to sets of neurons define how different instances of polychronous groups are matched up along common spikes. Figure 4.5 presents an example of such a presheaf, in this case depicting the gluing of elementary polychronous groups A_2 and A_3 as described above.

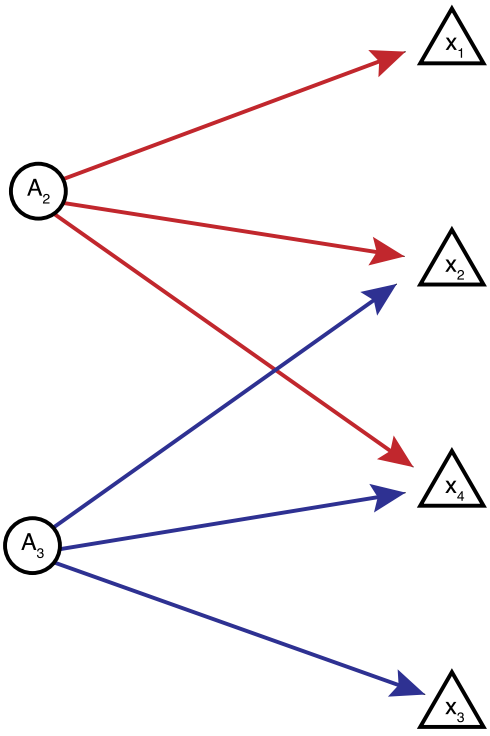


Figure 4.5: A presheaf in $\mathbf{Set}^{\mathbf{PG}^{op}}$ that represents stitching elementary polychronous groups A_2 and A_3 along spikes from neurons 2 and 4.

Having defined the PG base category and its associated category of presheaves of sets, we can begin to add semantics. To do this we define a functor $S : \mathbf{PG} \to \mathbf{CAlg}_{\mathbb{R}}$

Definition. Given a minimal polychronous model, the spiking semantics functor $S : \mathbf{PG} \to \mathbf{CAlg}_{\mathbb{R}}$ is defined by

- $S(A_\alpha) = A_\alpha$, the elementary polychronous ring associated with elementary polychronous group A_α.
- $S(x_i) = \mathbb{R}[x_i]$, the polynomial ring on a single indeterminate x_i.
- For each morphism $f : x_i \to A_\alpha$, the morphism $S(f) : \mathbb{R}[x_i] \to A_\alpha$ is the unital $\mathbb{R}$-algebra homomorphism that sends x_i to $[t_i] \in A_\alpha$.
- The implicit identity morphisms in **PG** are sent to identity morphisms in $\mathbf{CAlg}_{\mathbb{R}}$

The category $\mathbf{Set}^{\mathbf{PG}^{op}}$ satisfies the universal property of the free cocompletion, meaning given the functor S into the cocomplete category $\mathbf{CAlg}_{\mathbb{R}}$ there exists a unique functor $\hat{S}$ making the following diagram commute

$$\begin{array}{ccc} \mathbf{PG} & \xrightarrow{S} & \mathbf{CAlg}_{\mathbb{R}} \\ {\scriptstyle Y}\downarrow & \nearrow_{\hat{S}} & \\ \mathbf{Set}^{\mathbf{PG}^{op}} & & \end{array}$$

In fact, the functor $\hat{S}$ is given explicitly by

$$\hat{S}(P) = \operatorname{colim}\left(\int P \xrightarrow{\pi_P} \mathbf{PG} \xrightarrow{S} \mathbf{CAlg}_{\mathbb{R}} \right)$$

where $P \in \mathbf{Set}^{\mathbf{PG}^{op}}$ is a presheaf and $\int P$ is the category of elements of P.

This is exactly the colimit that corresponds to gluing together elementary polychronous groups indexed by $\int P$. The extension $\hat{S}$ maps presheaves to commutative algebras, and in the language of polychronization it assigns the appropriate polychronous ring to each presheaf interpreted as a syntactic combination of elementary polychronous groups. Because $\mathbf{Set}^{\mathbf{PG}^{op}}$ is universal, $\hat{S}$ it is uniquely defined up to isomorphism. This is another way of saying that the structure of the neural network uniquely determines the possible polychronous groups and hence the spiking dynamics.

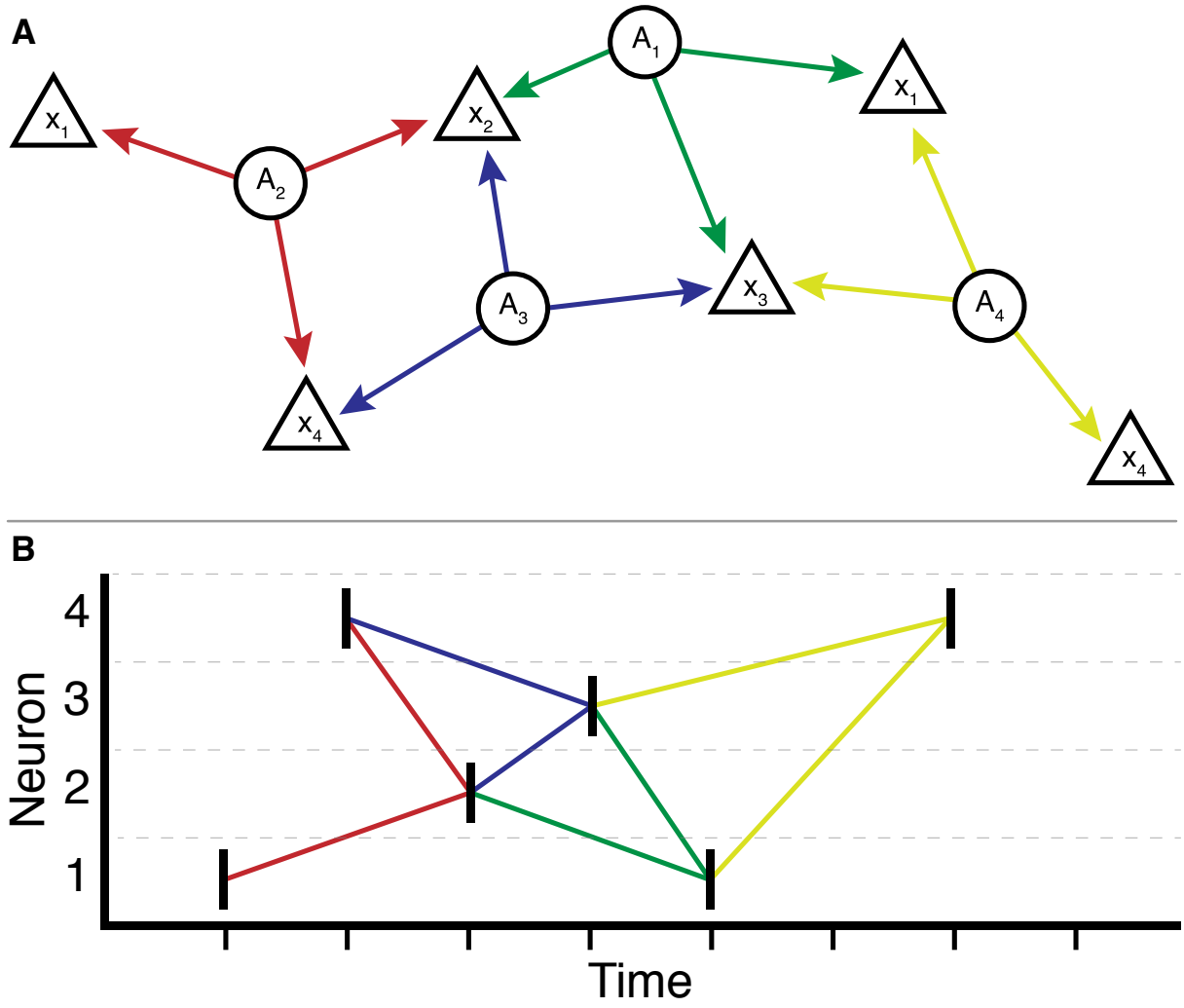


Figure 4.6: A larger polychronous graph and its corresponding polychronous group generated by the example network. A: The polychronous graph is a depiction of a functor (presheaf) $S : \mathbf{PG}^{op} \rightarrow \mathbf{Set}$. B: This functor corresponds to a colimit with spiking semantics illustrated by this spike raster. Colors label corresponding elementary polychronous groups from which the larger group is built.

4.7 Discussion

This chapter introduces a compositional, categorical theory of polychronization in spiking neural networks. Here we summarize the main points of the approach.

- A polychronous spiking network defines a category **PG** that inventories the elementary polychronous groups and their relationships to the individual neurons in the network.
- These elementary polychronous groups are the atoms from which larger polychronous groups and whole spike trains are constructed
- A polychronous spiking network also defines a functor $F : \mathbf{PG} \rightarrow \mathbf{CAlg}_{\mathbb{R}}$ that defines the semantics associated to the elementary polychronous groups. Each elementary polychronous group is associated to a polynomial ring that is interpreted as the coordinate ring obtained from the ideal of constraints yielded by the temporal pattern of spikes in the elementary polychronous group.
- The Yoneda embedding of **PG** into $\mathbf{Set}^{\mathbf{PG}^{op}}$ gives the free colimit completion of **PG**. This presheaf category consists of all the possible syntactical combinations of elementary polychronous groups. That is, a presheaf consists of elementary polychronous groups putatively glued together via shared spikes.
- By the universal property of the free colimit completion, we automatically obtain the unique Yoneda extension of the semantics functor F: $\hat{F} : \mathbf{Set}^{\mathbf{PG}^{op}} \rightarrow \mathbf{CAlg}_{\mathbb{R}}$. This functor gives the polynomial ring associated to every syntactic combination of elementary polychronous groups.

- If some syntactic combination is forbidden by the semantics of temporal pattern of activity implied by the combination, $\hat{F}$ will yield the zero ring.

Expressing polychronization in the domain of category theory provides a mathematically precise formulation of the phenomenon that exposes many theoretical "handles" which will support deeper investigations of the topic. For example, as a category of presheaves of sets on a small category, $\mathbf{Set}^{\mathbf{PG}^{op}}$ has the structure of the topos. As a result, there is a rich internal logical language defined precisely by the properties of polychronous spiking activity. This is potentially connected to the idea of "sensory invariants" in subjective physics (Brette, 2013), which are logical laws that embody the invariant structure in the environment that Gibson (Gibson, 2014) postulated as fundamental to perception. The algebraic side of this theory of polychronization connects spiking neural networks with many powerful tools from computational algebra (Cox et al., 2015). Future research may show these algorithms solving fundamental computational questions such as enumerating all the possible polychronous group given a particular network structure, or enumerating all the possible network structures consistent with a particular set of polychronous groups. Such algorithms would allow for direct contact with experimental data and significantly enhance our understanding of spiking dynamics in biological neural networks.

References

Armstrong, E., & Abarbanel, H. D. I. (2016). Model of the songbird nucleus HVC as a network of central pattern generators. *Journal of Neurophysiology*, *116*(5), 2405–2419. https://doi.org/10.1152/jn.00438.2016

Brette, R. (2013). Subjective physics. *arXiv:1311.3129 [q-bio]*. Retrieved July 6, 2020, from http://arxiv.org/abs/1311.3129

Chrol-Cannon, J., Jin, Y., & Grüning, A. (2017). An efficient method for online detection of polychronous patterns in spiking neural network. *Neurocomputing*, *267*, 644–650. https://doi.org/10.1016/j.neucom.2017.06.025

Cox, D. A., Little, J. B., & O'Shea, D. (2015). *Ideals, varieties, and algorithms: An introduction to computational algebraic geometry and commutative algebra* (Fourth edition) [OCLC: 915334372]. Springer.

Destexhe, A., Rudolph, M., & Paré, D. (2003). The high-conductance state of neocortical neurons in vivo [Number: 9 Publisher: Nature Publishing Group]. *Nature Reviews Neuroscience*, *4*(9), 739–751. https://doi.org/10.1038/nrn1198

Egger, R., Tupikov, Y., Katlowitz, K. A., Benezra, S. E., Picardo, M. A., Moll, F., Kornfeld, J., Jin, D. Z., & Long, M. A. (2019). Local axonal conduction delays underlie precise timing of a neural sequence [Publisher: Cold Spring Harbor Laboratory Section: New Results]. *bioRxiv*, 864231. https://doi.org/10.1101/864231

Gibson, J. J. (2014). *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press. https://doi.org/10.4324/9781315740218

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, *419*(6902), 65–70. https://doi.org/10.1038/nature00974

Izhikevich, E. M. (2006). Polychronization: Computation with spikes [Publisher: MIT Press]. *Neural Computation*, *18*(2), 245–282. https://doi.org/10.1162/089976606775093882

Long, M. A., & Fee, M. S. (2008). Using temperature to analyze temporal dynamics in the songbird motor pathway. *Nature*, *456*(7219), 189–194. https://doi.org/10.1038/nature07448

Long, M. A., Jin, D. Z., & Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. *Nature*, *468*(7322), 394–399. https://doi.org/10.1038/nature09514

Mac Lane, S., & Moerdijk, I. (1994). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer New York. https://doi.org/10.1007/978-1-4612-0927-0

Markowitz, J. E., Iii, W. A. L., Guitchounts, G., Velho, T., Lois, C., & Gardner, T. J. (2015). Mesoscopic patterns of neural activity support songbird cortical sequences [Publisher: Public Library of Science]. *PLOS Biology*, *13*(6), e1002158. https://doi.org/10.1371/journal.pbio.1002158

Pauli, R., Weidel, P., Kunkel, S., & Morrison, A. (2018). Reproducing polychronization: A guide to maximizing the reproducibility of spiking network models [Publisher: Frontiers]. *Frontiers in Neuroinformatics*, *12*. https://doi.org/10.3389/fninf.2018.00046

Rossant, C., Leijon, S., Magnusson, A. K., & Brette, R. (2011). Sensitivity of noisy neurons to coincident inputs [Publisher: Society for Neuroscience Section: Articles]. *Journal of Neuroscience*, *31*(47), 17193–17206. https://doi.org/10.1523/JNEUROSCI.2482-11.2011

Verduzco-Flores, S. O., Bodner, M., & Ermentrout, B. (2012). A model for complex sequence learning and reproduction in neural populations. *Journal of Computational Neuroscience*, *32*(3), 403–423. https://doi.org/10.1007/s10827-011-0360-x

Vigna, S. (2003). A guided tour in the topos of graphs. *arXiv:math/0306394*. Retrieved September 24, 2020, from http://arxiv.org/abs/math/0306394

Yildiz, I. B., & Kiebel, S. J. (2011). A hierarchical neuronal model for generation and online recognition of birdsongs. *PLoS computational biology*, *7*(12), e1002303. https://doi.org/10.1371/journal.pcbi.1002303