Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

Submitted in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

in

Neural Computation

TITLE: The structure and dimension of variability across multi-area cortical circuits

PRESENTED BY: Danielle M. Rager

ACCEPTED BY THE DEPARTMENT OF: Neuroscience Institute

Date Advisor

Date Department Head

Date Dean

The structure and dimension of variability across multi-area cortical circuits

THESIS

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Neural Computation

Danielle M. Rager

B.S., Bioengineering with Neuroscience and Music (Minors), University of Pittsburgh

Carnegie Mellon University

Pittsburgh, PA

April 2020

Abstract

Sensory and motor computations require tens of thousands of highly stochastic neurons in a cortical circuit to meaningfully coordinate their firing activity for a common goal. The trial-to-trial variability structure of neuronal population activity characterizes the coordinated neural dynamics underlying computation. Unsurprisingly, the dimension of the variability shared across neurons in one cortical population is generally orders of magnitude smaller than the number of neurons involved in a task. But how does this shared neuronal variability map across multiple cortical areas involved in the same computation?

In this thesis, I study the propagation of low dimensional shared variance across cortical regions as a means to understand the dynamics of multi-area brain computation. I first present a statistical model of movement encoding in human primary motor cortex that uncovers a one-dimensional trajectory of latent activity differentially modulated during movements in which the subject received somatosensory feedback. I then present new evidence that the dimension of shared variability increases from V4 to PFC during distributed processing of visual stimuli. I develop a multi-layer spiking network model with tuning-structured connectivity that, through non-linear recurrent dynamics, replicates the dimensionality expansion observed *in vivo*. Finally, I show evidence that my model's non-linear recurrent dynamics can be interpreted as time-sharing between multiple states of low-dimension, linear dynamics inherited from the upstream brain area. Together, these results aid our understanding of the subspaces of neuronal activity that are relevant across multiple brain areas during sensory and motor behaviors.

Acknowledgements

I am grateful to many people who helped make the completion of this thesis possible. I first acknowledge my committee members: Drs. Brent Doiron (chair), Valérie Ventura, Matthew Smith, Byron Yu, and Eric Shea-Brown. I have benefited from their expertise throughout my PhD. I feel fortunate to have been advised by both Valérie and Brent at different times in my graduate education, which provided me with a broad knowledge of computational neuroscience. Work with Valérie gave me an appreciation for the art and nuance involved with statistically characterizing and interpreting imperfect experimental data. Work with Brent over these last several years gave me a deep appreciation for the ground truths provided by mechanistic theories. Just as Brent said it would, my first match between network sims and network theory felt like a completed cycle and a major cause for celebration. I am grateful to have an advisor who was always prepared to celebrate those victories with a drink. I must also extend special thanks to Matt, who I feel provided careful guidance and kind encouragement beyond his duties throughout my work on Chapter 3 of this thesis.

Many students and post docs in the Doiron lab enriched my research experience and provided essential feedback on the presentation of my work. I have come to know Matt Getz particularly well over the last few years that we have been officemates. I appreciate the daily time we spend troubleshooting simulation errors and discussing plans of our next culinary undertaking. I frequently consulted Xinrou Yang about the linear response calculations in Chapter 3 of this thesis. I also first met Aubrey and Mike in the lab, and they have remained close friends throughout my graduate studies.

I received funding from the Department of Energy Computational Science Graduate Fellowship for several years of my graduate career. This program was instrumental in preparing me to undertake scientific research involving high performance computing, and it provided me with numerous academic resources that enabled my studies.

I owe immeasurable thanks to my family for my successes. My parents, both computer scientists, instilled in me a love of grappling with computational puzzles from a very young age. That is perhaps the most important gift you can bestow upon any scientist. They also worked to provide me with every resource and opportunity to thrive throughout my years of schooling, and they were present to support every achievement. My sister is an unwavering source of

friendship and inspiration. Her drive and scientific accomplishments at a young age motivate me in my own research. She is always the one on the other side of the chat providing laughs or advice at 4 am, when we are the only two night owls still awake. I am particularly indebted to family for keeping me afloat during a period of graduate school in which I had a serious back injury. Thank you to my aunts and uncles for transporting me to medical appointments and especially to my mom for watching over me during my weeks of post-surgical recovery.

Kevin has been my daily support structure in this last, hectic year of my PhD, making sure that I am always well fed, well caffeinated, and at least minimally rested in the times when I'm too stubborn to look out for myself. I cherish the way our times together are filled with music and joy. He reminds me of all life has to offer outside of work. I look forward to the next chapter, beyond this thesis, knowing that he will be a part of it.

Finally, I have many friends here in Pittsburgh who have made it home for essentially my entire adult life. It feels like I can't possibly acknowledge them all without forgetting important people whom I care about, and that is an incredible blessing.

Contents

List of Figures

List of Tables

1. *Introduction*

The mammalian brain consists of approximately 100 billion neurons [1] that communicate through the emission of timed action potentials, or spikes [2]. Patterns of spiking activity across neuronal populations give rise to the brain's ability to compute [3, 4, 5]. Contemporary systems neuroscience has thus been in a "technological arms race" [6] to simultaneously record single unit electrical activity from as many neurons as possible [7, 8, 9]. Calcium imaging techniques can record electrical activity from $\mathcal{O}(10^4)$ neurons (often of genetically-specified cell type) by measuring the fluorescence of proteins that bind to calcium ion channels [10, 11]. The resulting fluorescence signal represents a slower timescale convolution of the spiking activity of individual neurons. Multi-electrode arrays (MEAs) can record single unit activity from $\mathcal{O}(10^2)$ neurons, with the advantage that voltage traces are captured at the temporal resolution of single action potentials [12, 13, 14]. Systems neuroscience datasets, such as the two modeled in this thesis, increasingly consist of multiple, simultaneous MEA recordings. Datasets with simultaneous recordings from multiple regions of cortex are of particular interest, as they aid our understanding of neuronal communication both within and across cortical areas.

As advances in recording technologies continue to increase the scale of neural datasets, the focus of systems neuroscience has largely shifted from the study of individual neuron receptive field properties to the study of neuronal population dynamics – the temporal evolution of firing responses across entire neural ensembles [15]. Complex firing patterns across a neural ensemble can be summarized by the covariance structure of population spiking activity [16, 17, 18]. The covariance of a neural ensemble typically has latent structure of significantly lower dimensionality than the number of neurons in the ensemble, a signature of the ensemble's coordinated effort to perform a common computation [19, 20, 21, 8, 22, 23, 24]. The structure and dimensionality of spike count co-variability thus provides critical insight into the neural code [25, 26].

This thesis examines multi-area brain computation through the lens of neuronal co-variability. The introductory chapter that follows begins by defining neuronal variability and co-variability. I then provide an overview of statistical models and dimensionality reduction techniques that have been successfully used to characterize the latent structure of neuronal co-variability. Finally, I review the different physiological mechanisms that give rise to neuronal co-variability, distinguishable through network model studies of spiking neurons. I draw particular attention to the distinction between mechanisms in which co-variability is inherited from upstream neuronal populations and mechanisms by which co-variability is generated through recurrent network interactions. These distinctions are crucial to understand how neuronal co-variability propagates and transforms across the multiple recorded brain areas of our cortical circuits modeled in Chapters 2 and 3.

1.1 Definitions of spiking co-variability

Neural activity varies as a function of incoming sensory information. Two neurons that have similar receptive field properties, or functional *tuning*, are likely to respond to repeated presentations of a stimulus with similar spiking activity. I will refer to spiking co-variability arising from similarities in stimulus tuning as *signal correlations*. The pairwise signal correlation between the spiking activity of neuron i and neuron j can be defined as

$$\rho_{\text{signal}}(i,j,T) = \sum_{s\in\text{Stimuli}} \sum_{t\in\text{Trials}} \frac{\text{Cov}(N_i^T(s,t), N_j^T(s,t))}{\sqrt{\text{Var}(N_i^T(s,t))\text{Var}(N_i^T(s,t))}}, \tag{1.1}$$

where $N_i^T(s,t)$ is the spike count of neuron i over window length T [27].

The work in this thesis is focussed primarily on a different subset of spiking co-variability that is independent of changes to the measured environmental stimuli. Neurons exhibit irregular spiking responses to repeated presentations of the same stimulus, a phenomenon often referred to as *trial-to-trial variability*. Much of this trial-to-trial variability is private to each neuron and reflects individual cellular processes such as ion channel fluctuations and stochastic vesicle release [28, 29, 30]. A small fraction of trial-to-trial variability, however, is shared amongst cells in a cortical circuit and represents common fluctuations underlying the spiking responses of multiple neurons. These common fluctuations are due to unobserved sources of co-modulation, which may arise from cognitive processes such as arousal [31], attention [32, 33], learning [34, 35], working memory [36, 37], and other internal states, or may simply encompass modulations in network activity due to any unobserved common inputs [38, 39]. I will refer to this type of trial-to-trial co-variability as *shared variability*. The pairwise spike count correlation between neuron i and neuron j owing to shared variability can be defined as

$$\rho(i,j,s,T) = \sum_{t \in \text{Trials}} \frac{\text{Cov}(N_i^T(s,t), N_j^T(s,t))}{\sqrt{\text{Var}(N_i^T(s,t))\text{Var}(N_i^T(s,t))}}. \tag{1.2}$$

Here, $N_i^T(s,t)$ represents either neuron i's residual spike count in an evoked state, after subtraction of its mean response to a stimulus, or neuron i's spike count in a spontaneous state of activity.

Correlations defined by (1.2) are sometimes referred to as “noise correlations”. For brevity, this thesis more often adopts the standard of calling them simply *correlations*. The use of a modifier is reserved for discussions about signal correlations.

1.2 Characterizing low-dimensional shared variability

The full pairwise correlation matrix for a population of N neurons has $N(N+1)/2 - N$ unique entries, each defined by (1.2). However, analysis of neural data recorded *in vivo* consistently shows that population spiking activity exists in a subspace of significantly lower dimensionality [19, 20, 21, 8, 22, 23, 24]. The following portions this Introduction describe dimensionality reduction and statistical techniques that can be used to characterize the latent, low-dimensional structure of correlations. In addition, I will highlight example systems neuroscience studies in which the application of these techniques provided novel insight about neural computation.

1.2.1 Dimensionality reduction of neural data

Dimensionality reduction techniques reduce the population spiking activity from N neurons to a set of $K << N$ latent dimensions that still capture a majority of the population's shared variability. Most linear dimensionality reduction techniques are instances of a larger class of Linear Gaussian Models [40, 41]. I give more thorough treatment to this relationship in Appendix A.

Here, I will present a simple definition of a Linear Gaussian Model for i.i.d. observed data, defined by

$$\begin{aligned} \vec{x} &= \vec{w} & \vec{w} &\sim \mathcal{N}(0, Q) \\ \vec{y} &= C\vec{x} + \vec{\epsilon} & \vec{\epsilon} &\sim \mathcal{N}(0, R). \end{aligned} \tag{1.3}$$

In neuroscience applications, $\vec{y}$ is an observed population vector of activity from N neurons, $\vec{x}$ is a K-dimensional latent subspace that summarizes population activity, C is the matrix of model parameters that determine how the neural data is projected into the subspace, and $\vec{\epsilon}$ is a matrix of observation noise.

Principal component analysis (PCA), perhaps the most widely used dimensionality reduction technique, corresponds to a model of form (1.3) with the additional constraint $R = \lim_{\epsilon \to 0} \vec{\epsilon}\, \mathbb{I}$. In other words, PCA assumes all observed trial-to-trial co-variability of neural activity is due to shared variability. It lacks a model of observational noise, and by extension, a model of private trial-to-trial neuronal variability. It can then be shown that in PCA, the latent space $\vec{x}$ corresponds exactly to an eigendecomposition of $\vec{y}$, the observed trial-to-trial co-variability. Factor analysis (FA), another commonly used dimensionality reduction technique in neuroscience, corresponds to a model of form (1.3) with the much looser constraint $R = \text{diag}(R)$. This constraint represents the assumption that observation noise is un-correlated and due exclusively to private sources of trial-to-trial neuronal variability. Appendix A provides a detailed overview of Factor Analysis, which is used extensively in Chapter 3 of this thesis.

Some dimensional reduction techniques leverage the observation that spike trains are time-series data and neuronal firing activity is likely to evolve smoothly over time. The objective of this subclass of temporal dimensionality reduction techniques is to uncover latent dynamics [21]; specifically, these techniques assume trajectories of spiking activity in the K-dimensional latent subspace are signatures of an underlying dynamical structure controlling the temporal evolution of the network response. Dimensionality reduction techniques of this general form that have commonly been applied to neural data include Gaussian Process Factor Analysis (GPFA) [42] – in which the dynamics model of the latent space is stationary, but neural trajectories in the latent space are constrained to be smooth over time – and the family of Latent Linear Dynamical Systems (LDS) techniques, in which $\vec{x}$ from (1.3) has dynamics according to the full

Linear Gaussian Model definition presented in Appendix A's Equation (A.1). Temporal dimensionality reduction techniques like Switching Linear Dynamical Systems (SLDS) [43] and Latent Non-linear Dynamical Systems (NLDS) [44] expand upon this work by allowing the latent state to evolve with more flexible, non-linear dynamics. SLDS decomposes data into multiple states of latent linear dynamics. NLDS changes the form of the actual dynamics model to contain non-linear, recurrent terms.

1.2.2 Generalized Linear Model frameworks for capturing shared neuronal variability

Generalized Linear Models (GLMs) are an alternative, regression-based approach to capturing the structure of population spiking activity [45, 46]. GLMs flexibly extend the ordinary linear regression framework by allowing the response variable to take any distribution in the exponential family. A *link function* relates the linear predictor to the response variable. (Ordinary linear regression corresponds to a special case of the GLM in which the response variable is normally distributed, and the link function is simply the identity function.) Neuronal firing activity is commonly modeled with a GLM in which the response variable has a Poisson distribution, capturing the observation that spiking activity recorded *in vivo* has approximately equal mean and variance. The simplest GLM of sensory encoding posits that a neuron i's spiking activity $y_i(t)$ follows a Poisson distribution whose conditional intensity $\lambda_i(t)$ is a linear function of measured stimuli features:

$$\begin{aligned} \lambda_i(t) &= (y_i(t) \mid \theta) \\ &= f\left(\beta_{i0} + \sum_{j \in \text{Stim Feat}} \beta_{ij} S_j(t)\right) \end{aligned} \tag{1.4}$$

where β_{i0} is a constant bias capturing the magnitude of the neuron's baseline firing response, β_{ji} captures the neuron's receptive field properties for stimulus feature S_j, and f is the inverse of the link function. For a Poisson GLM, the canonical link function is the log of the response variable, and $f(u) = e^u$ is the inverse link function.

Multiple regression techniques can be used to simultaneously fit the spiking activity of many neurons according to the GLM in Equation (1.4). However, the resulting model of population spiking activity only characterizes signal correlations as defined in Equation (1.1). In the vocabulary of the Machine Learning community, encoding models like (1.4) constitute a *supervised learning* problem – the variability of the population spiking response is attributed entirely to known signals (the measured stimulus features). Shared neuronal variability as defined in Equation (1.2) is the portion of neuronal co-variability that we attribute to co-fluctuations in population activity from unobserved sources. Models involving latent (unobserved) covariates are labeled *unsupervised learning* problems. All dimensionality reduction techniques described in the previous section are thus unsupervised learning techniques. The GLM framework can be extended to include a mixture of signal covariates and latent covariates, simultaneously fitting a supervised learning model of stimulus encoding and an unsupervised learning model that captures shared variability as defined in Equation (1.2).

A Poisson GLM of neuron i's spiking activity that includes latent covariates may take the general form:

$$\lambda_i(t) = f\left(\beta_{i0} + \sum_{j \in \text{Stim Feat}} \beta_{ij} S_j(t) \; + \sum_{k \in \text{Latents}} C_{ik} X_k(t)\right), \tag{1.5}$$

where coefficient C_{ik} relates neuron i's conditional spiking activity $\lambda_i(t)$ to the kth dimension of the population latent space $X(t)$ and all other terms are consistent with Equation (1.4). Note the similarities in form between the latent covariate portion of the GLM model in Equation (1.5) and the Gaussian Linear Model generalizations of linear dimensionality reduction techniques defined in Equation (1.3). Neuro-statisticians have fit GLMs of the general form (1.5) with various implementations and interpretations of the latent space $X(t)$. In Rabinowitz et al. [23], X_k was a constant value per stimulus presentation that reflected the “shared gain” of visual neurons in hemisphere k. In Kulkarni & Paninski [39], the latent space $X(t)$ had only $K = 1$ latent dimension, but that univariate latent was a temporally evolving process (modeled as a Gauss-Markov autoregressive process) that represented unobserved common inputs. Models in which the latent space X evolves with temporal dynamics are known as *state-space models*. They have been applied generously to neural data [47, 48, 49, 50, 51], as summarized in a review from Paninski et al. [52].

I would now like to draw attention to the immense similarities between the GLM variety of state-space models discussed in the previous paragraph and temporal dimensionality reduction techniques. Both techniques adopt a Bayesian approach in which a generative probabilistic model links a state-evolving K-dimensional latent subspace to the N-dimensional population spiking activity, where $K << N$ [52]. The posterior distribution of each latent variable $k \in K$ is then recovered from the observed spike train data, as a function of the probabilistic model parameters, using a computational inference procedure [53]. In neuroscience literature, the nomenclature “state-space model” has most commonly been applied to regression approaches in which neural activity is modeled as a function of both observed stimulus features and a temporally evolving latent space. This remains true despite the observation that temporal dimensionality reduction techniques of the LDS form technically qualify as state-space models, even when they do not fit data to any observed covariates. A key distinction between the GLM approach to state-space models, described by Equation 1.5, and temporal dimensionality reduction techniques with linear dynamics, is the non-linearity and linearity, respectively, of the resulting models – the GLM approach singularly results in a non-linear mapping of the original spiking data because of its link function.

1.2.3 Low-dimensional shared variability in the brain

Several systems neuroscience studies have uncovered low-dimensional structure in the shared variability of neuronal population spiking activity. I will highlight a handful of these studies in which the results provided tantalizing scientific insight about the structure of neural computation. First, let us examine a collection of studies that used simple, rank-1 models to characterize the shared variability of activity in primary visual cortex (V1). Lin et al. [22] found that pairwise correlations in V1 were well-predicted by a population-wide model of shared variability; any heterogeneity in trial-to-trial responses across the population resulted only from each neuron’s multiplicative or additive gain onto a global factor. More importantly, this global factor constrained the population’s encoding of visual information. Relatedly, Schölvinck et al. [54] found that the majority of neuronal trial-to-trial variability in V1 could be explained with a global factor modeled as the sum of trial-to-trial variances over all recorded neurons. This global factor modulated with the presentation of different visual stimuli. In yet another analogous study, Okun et al. [55] found that the pairwise spike count correlations of V1 neurons in both mouse and monkey could be well-approximated merely by quantifying each neuron’s degree of coupling to the population firing rate. Population couplings subsequently predicted the tunedness of a neuron's response to sensory stimuli versus behavioral states such as motor intention. Population coupling was even well-correlated with a neuron's in-degree of synaptic connectivity. Together, this collection of studies links low-dimensional shared variability structure in V1 to observations about network anatomy and visual computation.

GLM methods of capturing latent shared variability (Section 1.2.2) have been used to improve the accuracy of stimulus encoding and decoding [46]. Vidne et al. [56] applied Kulkarni & Paninski [39]'s GLM model of (unobserved) common inputs to data from retinal ganglion cells (RGCs). The common input GLM successfully captured the spatiotemporal correlation structure in the spike trains of large populations of RGCs. The GLM simultaneously fit terms representing direct coupling between RGCs. In the absence of the univariate latent term for common input, the direct coupling strengths predicted by the model were significantly greater than the strength of synaptic connections measured *in vivo*; the addition of the latent term for common input produced biologically plausible coupling strengths. State-space GLMs have also been applied to brain-machine interfaces to improve the accuracy of decoded motor commands [57]. Finally, Rabinowitz et al. [23] used a GLM with latent shared gain terms to capture the effects of attentional modulation on a V4 neural population. Similar to the studies mentioned in the previous paragraph [22, 54, 55] only one latent gain modulator per V4 hemisphere was required to capture a significant fraction of variability in the population spiking activity, and the magnitude of the latent waxed and waned in the attended versus unattended state, respectively.

The field of systems neuroscience has employed dimensionality reduction techniques extensively and fruitfully to identify "neural manifolds" of sensory and motor coding [19, 21]. However, as noted in a recent review by Williamson et al. [58], dimensionality reduction techniques have, to date, most frequently been applied to either trial-averaged neural activity or single-trial neural responses to a stimulus presentation. This thesis is more interested in the burgeoning use of dimensionality reduction techniques for the identification of latent structure in the shared trial-to-trial variability of neural responses. Williamson et al. [59] used factor analysis (FA) to characterize the dimension of shared variability generated by Litwin-Kumar & Doiron [60]'s clustered spiking network model in the spontaneous state. Recently, Cowley et al. [61] used dimensionality reduction techniques to characterize drift in the shared variability of V4 and PFC activity. Huang et al. [24] used FA to uncover rank-1 shared variability in V4 and then developed a mechanistic spiking network model of internally-generated, low-dimensional shared-variability that was capable of producing similar FA results. This last study is an exciting example of how dimensionality reduction techniques applied to trial-to-trial neural responses can motivate and constrain neural circuit models.

1.3 Circuit mechanisms of shared variability

Dimensionality reduction techniques can identify the prominent latent modes of activity underlying shared neuronal variability, and GLMs can build expressions of neuronal spiking activity that account for these latent fluctuations and their contribution to neural encoding/decoding algorithms. However, neither of these statistical approaches supplies a mechanistic understanding of how correlations arise or propagate through the pairwise connectivity of individual neurons. Interpretable models of correlation transfer through neural circuits require us to consider the biophysical dynamics of individual neurons and the resultant dynamics of their interactions. In this section, I will first review the dynamics of spiking networks with random recurrent coupling obeying a balance between excitation and inhibition. *Balanced networks* generate trial-to-trial variability, but are incapable of generating shared variability. I then summarize circuit mechanisms by which shared variability is inherited from external neural populations and modulated by private sources of neuronal variability or the transfer functions of individual neurons. These mechanisms of variability propagation are reviewed in Doiron et al. [16]. Finally, I will tease circuit mechanisms in which shared variability is shaped by recurrent interactions, a topic that will be further dissected in Chapter 3 of this thesis.

1.3.1 Balanced networks for internally generated variability

van Vreeswijk & Sompolinsky [62, 63] introduced a *balanced network* theory in which neuronal variability is internally generated through the random, recurrent coupling of excitatory (E) and inhibitory (I) neurons. In the original formulation of the model, the spiking activity of neuron i is reduced to a binary state variable s_i, which thresholds the linear combination of its inputs:

$$s_i(t) = H\left(\mu + \sum_{j=1}^{N} J_{ij}^{\alpha\beta} s_j(t) - \theta\right). \tag{1.6}$$

Here, $H(\cdot)$ is the Heaviside step function, μ is a constant bias to each neuron, $s_j(t)$ is the spiking input from neuron j, J_{ij} is the synaptic strength of connection from neuron j in population β to neuron i in population α, where $\{\alpha, \beta\} \in \{E, I\}$, and θ is each neuron's threshold. Each neuron receives inputs from an average of $K = pN$ other neurons, where p is the probability of connection.

Balanced networks achieve spike count variability by scaling synaptic strengths according to $J^{\alpha\beta} \propto \mathcal{O}(1/\sqrt{K})$. If we denote the mean of the presynaptic input from population α to population β as $I^{\alpha\beta}$, the variance of $I^{\alpha\beta}$ is $\mathrm{Var}(I^{\alpha\beta}) \propto K(1/\sqrt{K})^2 = \mathcal{O}(1)$. However, this choice of scaling also implies that the mean synaptic input from each population β to α grows with network size according to $\langle I^{\alpha\beta}\rangle \propto K(1/\sqrt{K}) = \sqrt{K}$. To avoid mass excitation under their chosen scaling law, van Vreeswijk & Sompolinsky [62] introduced the balance condition, in which the mean synaptic inputs to each cell-type population exactly cancel. Assuming that the constant bias to each neuron also scales with K such that $\mu = \bar{\mu}\sqrt{K}$, the mean synaptic inputs to the E and I populations are, respectively,

$$\langle I^E\rangle = \sqrt{K}\left(\bar{\mu}^E + \bar{J}^{EE}\langle s^E\rangle - \bar{J}^{EI}\langle s^I\rangle\right) \tag{1.7}$$

$$\langle I^I\rangle = \sqrt{K}\left(\bar{\mu}^I + \bar{J}^{IE}\langle s^E\rangle - \bar{J}^{II}\langle s^I\rangle\right). \tag{1.8}$$

Equations 1.7 and 1.8 will diverge with K unless they satisfy

$$0 = \bar{\mu}^E + \bar{J}^{EE}\langle s^E\rangle - \bar{J}^{EI}\langle s^I\rangle \tag{1.9}$$

$$0 = \bar{\mu}^I + \bar{J}^{IE}\langle s^E\rangle - \bar{J}^{II}\langle s^I\rangle. \tag{1.10}$$

This leads to the following solutions for $\langle s^E\rangle$ and $\langle s^I\rangle$:

$$\langle s^E\rangle = \frac{\bar{\mu}^E\bar{J}^{II} - \bar{\mu}^I\bar{J}^{EI}}{\bar{J}^{EI}\bar{J}^{IE} - \bar{J}^{II}\bar{J}^{EE}} \tag{1.11}$$

$$\langle s^I\rangle = \frac{\bar{\mu}^E\bar{J}^{IE} - \bar{\mu}^I\bar{J}^{EE}}{\bar{J}^{EI}\bar{J}^{IE} - \bar{J}^{II}\bar{J}^{EE}}. \tag{1.12}$$

Without loss of generality, we can rescale all network weights such that $J^{EE} = J^{IE} = 1$ and only connections from I cells affect our solution. Positive values of $\langle s^E\rangle$ and $\langle s^I\rangle$ are then only satisfied through one of the following two conditions:

$$\bar{J}^{EI} > \bar{J}^{II}, \ \frac{\bar{\mu}^E}{\bar{\mu}^I} > \frac{\bar{J}^{EI}}{\bar{J}^{II}} \tag{1.13}$$

$$\bar{J}^{EI} < \bar{J}^{II}, \ \frac{\bar{\mu}^E}{\bar{\mu}^I} < \frac{\bar{J}^{EI}}{\bar{J}^{II}}. \tag{1.14}$$

The second condition allows $\langle I^E \rangle = \sqrt{K}\left(\bar{\mu}^E - \bar{J}^{EI}\frac{\bar{\mu}^I}{\bar{J}^{II}}\right) \to -\infty$ when $\langle s^E \rangle = 0$ and can thus be rejected. A balanced network solution therefore uniquely exists for the conditions of Equation 1.13.

These balance conditions can similarly be applied to networks of leaky-integrate-and-fire (LIF) model neurons. Figure 1.1 shows the raster of excitatory activity from a balanced network containing 4000 E and 1000 I LIF model neurons. We note that the balance condition produces an *asynchronous state* in which neural activity has near-zero shared variance. van Vreeswijk & Sompolinsky [62, 63] formally derived the asynchronous population dynamics of the balance condition under the assumption that network connectivity was sparse. That is, they assumed neuronal connection probabilities were very low such that $K << N$ and subsequently, that pairs of neurons had few common inputs. However, many noted that the balance condition (1.13) produced network activity in the asynchronous state even when networks were densely connected. Indeed, connection probabilities in our pictured network are $p^{EE} = 0.2, p^{IE} = p^{IE} = p^{II} = 0.5$, but population activity is still asynchronous.

Renart et al. [64] solved this puzzle when they demonstrated that shared E and I fluctuations cancel in densely connected balanced networks, producing net zero shared fluctuations and asynchronous activity. Let us define each subpopulation $\beta \in \{E, I\}$'s input current to neuron i of population α as

$$I_i^{\alpha\beta} = \sum_j J_{ij}^{\alpha\beta} s_j^{\beta}. \tag{1.15}$$

The total input current to population α is

$$I^{\alpha} = \mu^{\alpha} + I^{\alpha E} + I^{\alpha I}. \tag{1.16}$$

The shared variance between a pair of neurons i and j from population α is then

$$\text{Cov}(I_i^{\alpha}, I_j^{\alpha}) = \text{Cov}(I_i^{\alpha E}, I_j^{\alpha E}) + \text{Cov}(I_i^{\alpha I}, I_j^{\alpha I}) + 2\text{Cov}(I_i^{\alpha E}, I_j^{\alpha I}). \tag{1.17}$$

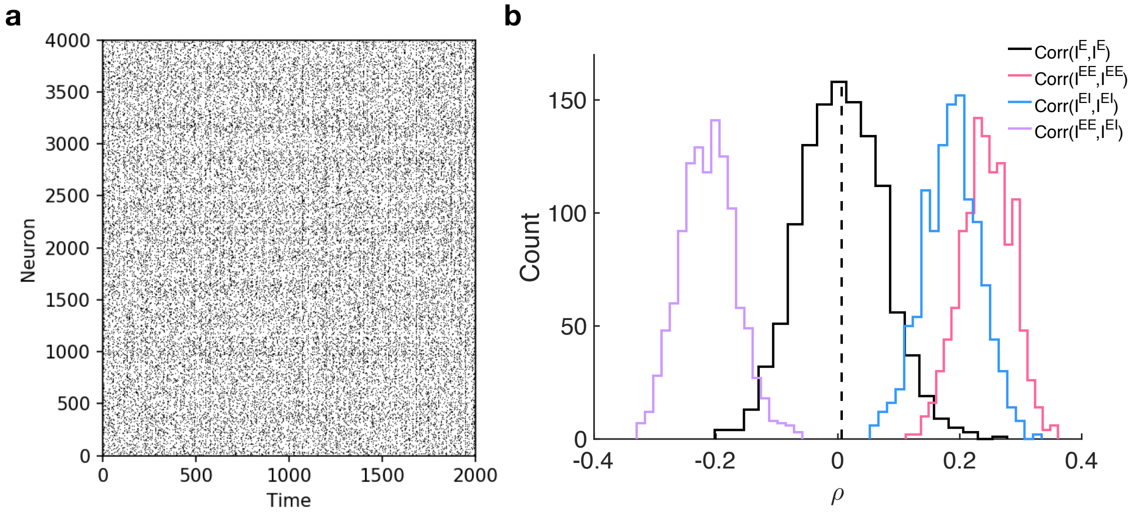

Figure 1.1: Asynchrony in balanced networks with dense connectivity **(a)** Spike times of $N^E = 4000$ neurons in a balanced network simulated over 2 seconds. The network contains $N^E = 1000$ neurons, and connection probabilities were $p^{EE} = 0.2, p^{IE} = p^{IE} = p^{II} = 0.5$. Spiking activity is asynchronous despite dense connectivity. **(b)** The total input currents received by pairs of excitatory neurons are, on average, uncorrelated (black), despite the fact that neuron pairs receive correlated input fluctuations from E neurons (red) and I neurons (blue). These positive input current correlations are cancelled by the anti-correlations between the E and I currents that pairs receive (purple). Pairwise correlations were computed for a sample of $n^E = n^I = 200$ neurons from the larger network.

When $I_i^{\alpha E}$ and $I_j^{\alpha I}$ are anti-correlated, terms on the right hand side of the above equation can cancel, making the total shared variance of our neuron pair zero. Figure 1.1b illustrates this phenomenon. Methods Section 3.3.10 of Chapter 3 sketches the formal derivation of a self-consistent solution, in which the asynchronous state, defined by $\mathcal{O}(1/N)$ shared variance, relies on the dynamic cancellation of subpopulation current fluctuations.

1.3.2 Shared variability inherited from upstream neuronal populations

What circuit mechanisms are capable of producing network activity with shared variability, since it cannot be achieved with classic E/I balance? To answer this question, we adopt the mechanistic framework of correlation transfer reviewed by Doiron et al. [16] and apply this framework to a pair of simultaneously recorded neurons. The neurons are themselves uncoupled, but can be correlated through the input each receives from an upstream neuronal population. We define the afferent inputs received by neuron i as $x_i(t)$ for $i = 1$ or 2 of our studied pair. The covariance of the activity across the two afferent populations is

$$\text{Cov}(\hat{x_1}, \hat{x_2}),$$

where $\hat{x_i}$ characterizes the integrated synaptic inputs to neuron i over time window T:

$$\hat{x_1} = \int_0^T x_i(t)dt.$$

When $\text{Cov}(\hat{x_1}, \hat{x_2}) \neq 0$, the presynaptic inputs are correlated. When these input correlations are small, the shared variance of our studied pair of output neurons can be approximated as a linear function of the input covariance,

$$\text{Cov}(y_1, y_2) \approx G^2\text{Cov}(\hat{x_1}, \hat{x_2}), \tag{1.18}$$

where y_i is the spiking activity of neuron i over time window T and $G_i = G$ is the gain of the target neuron. This gain represents the neuron's sensitivity to its inputs at the operational point of our approximation and is often referred to in statistical mechanics literature as the neuron's *linear response*. Equation (1.18) results from the assumption that the spiking activity of neuron i is a linear combination of its baseline activity and perturbations from weak common input fluctuations $s(t)$ such that

$$\langle y_i \rangle \approx \langle y_{i0} \rangle + G\hat{s}, \tag{1.19}$$

where $\langle . \rangle$ denotes an expectation over trials, $\langle y_{i0} \rangle$ is the mean spike count of neuron i at our operational point $s = 0$ and

$$\hat{s} = \int_0^T s(t)dt$$

represents the synaptic integration of the input fluctuations. Notably, Equation (1.18) can be used to measure state-dependent changes in shared variability structure, as modulations in network state will change the operating point of (1.18), and subsequently, alter the linear response G of our studied pair of neurons. See Appendix B for a more in-depth treatment of the assumptions of linear response theory, which will be used in Chapter 3.

The total integrated input to each neuron, $\hat{x}_i(t)$, is comprised of both a presynaptic component that we define as P_i, owing to the upstream neuronal population, and a postsynaptic component, owing to the receiving neuron's cellular sources of variability such as stochastic vesicle release. We assume that the postsynaptic component of the integrated input is uncorrelated between neuron 1 and neuron 2. Correlations inherited from the presynaptic component of the integrated input, occurring when $\text{Cov}(P_1, P_2) \neq 0$, can result from two sources. The first is common afferent projections, in which neuron 1 and neuron 2 receive direct synaptic input from the same neuron or neurons in the upstream neuronal population. The second mechanism of presynaptic input correlations owes to spike count correlations between the activity of the upstream neuronal ensembles projecting to neuron 1 and neuron 2. Notably, each of these presynaptic mechanisms can induce input correlations without the other. This implies, through Equation (1.18), that in a feed-forward network, shared variability between neuron 1 and neuron 2 can be inherited through common projections from an upstream neuronal population with asynchronous activity, disjoint projections from an upstream neuronal population with correlated activity, or common projections from an upstream neuronal population with correlated activity. These three circuit conditions capable of inducing presynaptic correlation will be indistinguishable from the vantage point of the postsynaptic neuron pair.

In a biological neural circuit, it is likely that total presynaptic activity is due to both excitatory E and inhibitory I upstream inputs, and the presynaptic shared variance is

$$\text{Cov}(P_1, P_2) = \text{Cov}(\hat{E}_1, \hat{E}_2) + \text{Cov}(\hat{I}_1, \hat{I}_2) + \text{Cov}(\hat{I}_1, \hat{E}_2) + \text{Cov}(\hat{E}_1, \hat{I}_2). \tag{1.20}$$

Doiron et al. [16] demonstrate that cell-type specific changes in drive to the upstream population can then have non-monotonic effects on the shared variance of neuron 1 and neuron 2. The network model we will develop in Chapter 3 uses a simpler upstream population structure in order to focus on the propagation of inherited variability; our input population consists only of excitatory cells, making the presynaptic shared variance

$$\text{Cov}(P_1, P_2) = \text{Cov}(\hat{E}_1, \hat{E}_2). \tag{1.21}$$

We modulate the strength of presynaptic shared variability (1.21) by way of modulating the strength of spike count correlations in the upstream ensemble activity, and we observe the result on the shared variability of neuron pairs in the output population (1.18). In a network without recurrent coupling, increasing (1.21) will monotonically increase (1.18), but Chapter 3 also studies less intuitive cases of shared variability propagation owing to recurrent interactions. We introduce recurrent mechanisms of shared variability in the subsection that follows.

As outlined in Doiron et al. [16], the presynaptic shared variability inherited in a feedforward network can be modulated by the postsynaptic noise of the receiving neurons or by changing the neurons' conductance-based input fluctuations. In the first modulation mechanism, inherited presynaptic shared variability is diluted by increasing levels of private postsynaptic variability owing to cellular properties like increased stochasticity of vesicle release or increased stochasticity of ion channel gating [28, 29, 30]. This modulation mechanism constitutes a change in background synaptic fluctuations. In the second modulation mechanism, increased variance of the presynaptic input Var(P_i) decreases the response gain G of a neuron (1.19),(1.18),[65, 66, 67].

1.3.3 Shared variability generated and modulated through recurrent interactions

Uniform balanced recurrent coupling produces correlations that scale inversely with network size N, resulting in near-zero correlations for networks with thousands of neurons (Figure 1.1). Networks with more structured recurrent coupling are capable of generating non-zero noise correlations; clustered [60, 68] and spatially-dependent [69, 70, 71, 72] recurrent coupling profiles can internally generate localized pockets of positive correlations between proximal pairs of neurons. Recent advances in spiking network theory show that networks with spatially-dependent coupling can even internally generate population-wide, low-rank shared variability [24].

Whether shared variability in a multi-layer network is internally generated through recurrent coupling or inherited through presynaptic inputs, we can study its propagation through recurrent interactions using an extension of the linear response framework introduced in the previous subsection (1.18) so long as the strength of shared input fluctuations is still much smaller than the strength of private synaptic fluctuations at our chosen operating point. When our studied output neuron pair exists in a network layer with recurrent coupling, the shared variance of presynaptic inputs Cov(P_1, P_2) will arise through common fluctuations $F(t)$ received via two circuit pathways. The first is a feedforward pathway, in which shared variance is inherited through direct projections of upstream inputs entrained to $F(t)$. We denote shared variance inherited through the feedforward pathway Cov_{FF}. The second is an indirect recurrent pathway, in which common fluctuations $F(t)$ are filtered through excitatory and inhibitory recurrent projections to the studied output neuron pair. The shared variance of presynaptic inputs is a function of the covariance owing to these two circuit pathways and their interactions such that

$$\text{Cov}(P_1, P_2) = f(\text{Cov}_{\text{FF}}, \text{Cov}_{\text{Rec}}, \text{Cov}_{\text{FFRec}}), \tag{1.22}$$

where Cov_{Rec} is the shared variance from common recurrent input to the studied output neuron pair and $\text{Cov}_{\text{FFRec}}$ represents covariance resulting from the interaction of the feedforward and recurrent circuit pathways.

1.4 Outline of this thesis

This thesis presents the work of two studies, both of which characterize the propagation of shared variance through a multi-area cortical circuit. In Chapter 2, I study the effects of somatosensory inputs to the primary motor cortex (M1) of a human subject using an intracortical brain-machine interface (BMI). I show that the somatosensory signal induces a low-rank shared variance of M1 activity that changes the appearance of M1 tuning; when unaccounted for, this signal disrupts the decoding performance of the BMI. I then develop an improved model of M1 encoding that accounts for latent shared variance due to sensory input. My model is a *generalized additive model*, which extends the class of GLMs discussed in Section 1.2.2 by including a smooth, non-parametric modulator representative of common sensory fluctuations. This work was done in collaboration with my advisor Valerie Ventura of the Carnegie Mellon University Department of Statistics, as well as former graduate student John Downey, Dr. Jennifer L. Collinger, Dr. Douglas J. Weber, Dr. Michael Boninger, and Dr. Robert Gaunt of the University of Pittsburgh Rehab Neural Engineering Laboratory.

In Chapter 3, I apply dimensionality reduction techniques (Section 1.2.1) to uncover an expansion of the dimension of shared variability between visual cortex (V4) and prefrontal cortex during a visual task. I then develop a mechanistic, spiking network model of this multi-area circuit to better understand how PFC's non-linear recurrent dynamics filter the low-dimensional shared variance inherited from multiple, tuned V4 inputs. My PFC model network is globally balanced (1.3.1) , but includes tuned assemblies that give rise to the non-linear recurrent dynamics. This work was done in collaboration with my advisor Dr. Brent Doiron from the University of Pittsburgh Department of Mathematics, as well as former graduate student Sanjeev Khanna from the University of Pittsburgh Department of Bioengineering and Dr. Matthew Smith from Carnegie Mellon University's Department of Biomedical Engineering.

2. *Proprioceptive feedback modulates motor cortical tuning during human brain-machine interface control*

2.1 Abstract

Loss of proprioception is known to severely impair motor control, but the neural mechanisms by which proprioception aids in the planning and execution of visually guided movements are not well understood. We investigated the impact of providing proprioceptive feedback to a human subject with tetraplegia and intact sensation who was implanted with two 100-channel microelectrode arrays in primary motor cortex (M1). BMI-assisted reach performance was highly dependent on the feedback sources provided during decoder training; if a decoder was trained with vision alone, adding proprioceptive feedback during a reach task degraded reach performance. The inability to mismatch decoder and task feedback conditions arises from a shift in M1 velocity tuning between the visual (V) and visual+proprioceptive (VP) feedback conditions. The VP condition was also marked by decreased modulation depth and increased variability of M1 velocity tuning, which resulted in degraded BMI-assisted reach performance. Because we do not believe that proprioception fundamentally degrades motor control in healthy individuals, we propose that M1 encodes proprioceptive information with dynamics unknown to our BMI decoder. We show evidence that M1 activity in the VP condition is better modeled with the inclusion of a smooth, time-dependent, modulator that is shared amongst the neural population. Our encoding model that includes this modulator improves M1 tuning the most when the subject receives somatosensory feedback, suggesting the modulator captures shared variability from somatosensory inputs to M1. Together, our results suggest that new decoders will need to

be developed for closed-loop BMIs that make efficient use of natural or surrogate somatosensory information.

2.2 Introduction

Complex limb movements like reaching and grasping are closed-loop motor programs that integrate feedback from multiple sensory modalities [73, 74]. A person attempting to grasp a cup must first visually locate the cup, plan a movement trajectory to the target [75], and issue that motor command through a pattern of motor cortex (M1) activity [76]. The motor command is read out by the spinal cord and executed by muscle motor neurons [73]. As the reach progresses, the person receives both visual and proprioceptive feedback to guide the movement and correct for perturbations [77]. Accordingly, motor control degrades significantly in human subjects with afferent pathway damage that results in the loss of proprioceptive information [78, 79, 80]. Recent studies have confirmed that somatosensory cortex (S1) communicates directly to M1 via monosynaptic projections that are capable of driving motor behavior [81, 82].

Brain machine interfaces (BMIs) restore motor function in patients with spinal cord injuries and damaged efferent pathways by transforming M1 activity to a control signal for a cursor or robotic arm [83]. Despite the known importance of proprioceptive information for upper extremity control, most current BMI implementations rely exclusively on visual feedback [84]. We had the rare opportunity to investigate the differential effects of providing proprioceptive feedback to a human subject during her BMI control. The subject, who has tetraplegia but intact sensation, used a BMI in which an exoskeleton moved her own arm congruently with a virtual arm or robotic arm. She thus received proprioceptive information from her muscle and tendon stretch receptors [85] and tactile information from the interaction between her arm and the exoskeleton [86].

Unlike the results of a previous study on proprioceptive feedback during non-human primate BMI control [87], our subject's performance of a BMI-assisted reaching task degraded when we allowed a decoder trained only with visual feedback to leverage the subject's proprioceptive signals. More generally, the feedback sources provided during decoder training and online BMI control could not be mismatched without a degradation of reach performance. This behavioral observation suggested that proprioception altered the decoder's known mapping between neural activity and kinematics. Analysis of neural activity confirmed that additive proprioceptive feedback shifted the velocity tuning curves of recorded M1 channels. Proprioceptive feedback also caused a decrease in the modulation depth of M1 velocity tuning curves, making neural activity non-exchangeable between reaches guided with somatosensory feedback and those guided with vision alone.

Contrary to the intuition that additive somatosensory feedback could only aid BMI performance, our subject's BMI assisted reach control degraded with online proprioceptive feedback, even when the decoder was trained using proprioception. The neural signature of this behavioral effect was reduced signal to noise ratio of M1 velocity tuning during trials when the patient had somatosensory feedback. This manifested as (1) a decrease in the modulation depth of velocity tuning curves, (2) an increase in the trial-to-trial variability of velocity tuning curves, and (3) an increase in the noise correlations of M1 activity.

Because we do not believe that proprioceptive information fundamentally degrades motor control in healthy individuals, we propose that M1 encodes proprioceptive information in ways incongruous with the simple endpoint velocity tuning leveraged by current BMI decoders. After the effects of velocity tuning were removed, neural activity in the proprioceptive feedback condition contained gain fluctuations that were coordinated across broad subpopulations of recorded M1 channels. The coordinated fluctuation of neural activity that we observed in the proprioceptive feedback condition is consistent with observations that M1 may integrate information from large regions of S1 [88]. Moreover, several recent studies have uncovered low-rank co-fluctuations of neural activity in cortex [55, 22, 23].

To account for co-fluctuations that might drive the M1 population when the subject had proprioceptive feedback, we developed a novel M1 encoding model in which the spiking activity of each recorded M1 channel is a function of channel-specific tuning for endpoint velocity and a time-varying waveform of activity that is shared across the M1 population. The addition of a global gain-modulating waveform improves velocity tuning in the proprioceptive feedback condition. We therefore propose that S1 activity might be encoded as a shared wave of activity that broadly modulates the gain of M1 neurons. BMIs seeking to make efficient use of natural or surrogate somatosensory feedback may need to account for such shared modulators.

2.3 Methods

The experimental methods that follow will be described in more detail in a forthcoming publication by Gaunt et al.

2.3.1 Study participant and regulations

The subject was a 52 year old woman with a diagnosis of spinocerebellar degeneration without cerebellar involvement, manifesting as complete tetraplegia with generally intact afferent innervation. While the subject did have some mild sensory deficits and hypersensitivity, clinical testing confirmed that her proprioception was robust enough to give her appropriate feedback.

The subject provided informed consent verbally for study participation, with documents signed by her legal proxy. The study was carried out under approval from the Institutional Review Boards of the University of Pittsburgh, and the Space and Naval Warfare Systems Center Pacific. Implanted devices were granted Investigational Device Exemption by the US Food and Drug Administration. The trial is registered on clinicaltrials.gov under identifier NCT01364480.

2.3.2 Behavioral tasks

The subject was trained to control either an adjacent, free-standing prosthetic arm (modular prosthetic limb [MPL], Johns Hopkins University, Applied Physics Laboratory, Baltimore, MD, USA), or a virtual representation of that arm, in a two-dimensional workspace. The subject completed two tasks that tested her BMI-assisted motor control. The first, a line-crossing (LC) task, required the subject to move the MPL medial-laterally across parallel lines spaced 20 cm apart as many times as possible in a 60 s period (Figure 2.1a). We used the number of line crossings achieved in the 60 s trial to measure the extent to which the subject could effectively control the MPL. Though the task was one-dimensional, MPL control was two-dimensional in the coronal plane.

The second task, a 2D pursuit (2DP), required the subject to move a virtual arm to one of five on-screen targets in the following positions of the coronal plane: center, above-center, left-of-center, right-of-center, or below-center (Figure 2.2a). The subject received a visual signal of the target that was to be pursued before the movement onset cue. The subject did not have to return to the center target after each reach.

The subject completed the two tasks described above using the typical BMI control paradigm of exclusively visual feedback (V) and a paradigm in which she received simultaneous visual and proprioceptive feedback (VP). In the VP paradigm, the subject's own arm was placed in an upper extremity exoskeleton (Armeo Power, Hocoma, Switzerland) and moved congruently with the MPL or virtual arm. In a small subset of initial LC task trials with VP feedback, prior to the introduction of the exoskeleton, an experimenter manually moved the subject's arm congruently with the MPL. The number of line crossings achieved in the VP condition did not differ significantly between trials in which the arm was moved manually and trials in which the arm was moved with an exoskeleton. The subject was always unable to view her own moving arm in the VP paradigm, disallowing multiple points of visual attention.

2.3.3 Electrophysiology

The subject had two 100-channel microelectorde arrays (4 x 4 mm^2 footprint, 1.5 mm shank length, Blackrock Microsystems, Salt Lake City, UT, USA) implanted into the somatotopic region of left M1 responsible for right arm and hand control [89]. The experiments analyzed in this study began 11 months post-implantation and continued until two years post-implantation. Signals from up to 192 microelectrodes were recorded. Each microelectrode recorded intracortical neuron depolarization using the NeuroPort data acquisition system (Blackrock Microsystems, Salt Lake City, UT, USA). Several parameters were saved from these depolarizations, including threshold crossings, timing and action potential waveform snippets. Due to the low firing rates observed in the recorded M1 activity, action potentials were not sorted to distinguish signals generated by individual neurons. Therefore, the neural data recorded on each channel of the array, which were used for decoding and all post-hoc analysis, may represent the activity of multiple neurons. Threshold crossing times were collected into 30 ms time bins to create a vector of spike counts $s_i(t)$ for each of the N total recorded channels.

2.3.4 Online decoding

Commands from motor cortex produce unique combinations of spiking activity [76]. The control scheme for a BMI thus requires a map between spiking activity and the kinematics of the MPL or virtual arm. To train the BMI decoder, the subject observed a computer program performing the 2DP task. She was visually cued to the appropriate reach target and then instructed to imagine performing the reach while she observed the virtual arm's movement. The spike count $s_i(t)$ of each recorded neural channel i was smoothed with a low pass filter in which the kernel was an exponential function of 450 ms width. A square root transform was performed on the smoothed vector to stabilize variance. Each channel's resulting firing rate vector $f_i(t)$ was then mapped to the virtual arm's movement velocity according to

$$f_i(t) = \beta_{i0} + \beta_{i1} v_x(t) + \beta_{i2} v_y(t) + \varepsilon_i(t) \tag{2.1}$$

where $v_x(t)$ and $v_y(t)$ are vectors of the x and y velocity of the virtual arm over the reach duration. In matrix form, this relationship is expressed as

$$F = VB + \varepsilon \tag{2.2}$$

where F is the $T \times N$ matrix of all channel firing rate vectors $f_i(t)$ over the total trial duration T, V is a T x 3 column matrix defined by $[\vec{1}, v_x(t)^\top, v_y(t)^\top]$, B is the $3 \times N$ matrix of regression coefficients $[\vec{\beta_0}^\top, \vec{\beta_1}^\top, \vec{\beta_2}^\top]$, and ε is the matrix of channel residuals $\sim \mathcal{N}(0, \sigma^2)$.

The maximum likelihood estimate of B with ridge regression regularization is:

$$\hat{B} = (V^T V + \lambda \mathbb{I})^{-1} V^T F \tag{2.3}$$

where λ is the regularization parameter optimized to minimize the prediction risk of F and $\mathbb{I}$ is a 3 $\times$ 3 identity matrix.

The mapping between firing activity and kinematics in Equation (2.2) was then inverted to predict the subject's reach intent during the behavioral tasks according to:

$$V = FW + \varepsilon \tag{2.4}$$

where the weight matrix W can be computed by taking the Moore-Penrose Pseudoinverse of $\hat{B}$ in Equation (2.4). Additional details about the decoder can be found in Collinger et al. [89] and Wang et al. [90].

We computed the mapping between velocity information and firing activity that is described above for two different feedback paradigms during 2DP training: one paradigm in which the subject had only visual feedback of the virtual arm (V), and one in which the subject had both visual and proprioceptive feedback (VP), provided by the exoskeleton. This resulted in two sets of decoder parameters, which we will refer to as W_V and W_{VP}.

2.3.5 Reach analysis

In the LC task, we evaluated the subject's ability to control her reaching movements by three metrics: (1) the number of line crossings she achieved in a 60 s period, (2) the mean path length per line crossing, and (3) the variance of the reach in the anterior-posterior dimension, which was the dimension of control orthogonal to the LC goal. In the 2DP task, we evaluated the subject's performance by the number of reaches that were successfully terminated at the designated target within a 3 s period. For each task, reach trajectories were smoothed with a Gaussian filter of width 200 ms. Kinematic observations were pooled across trials, and a two-way ANOVA with post hoc t-tests was used to determine whether performance differed across the feedback and decoder conditions.

2.3.6 Post-hoc tuning curve analysis

Our reach analysis suggested that the structure of the neural activity differed in the V and VP conditions. We therefore conducted a post-hoc analysis of M1 tuning, to compare M1 activity structure in the V and VP conditions. Previous studies have demonstrated that the preferred velocity directions of M1 neurons shift with prolonged BMI decoder in ways that optimize the movement readout of the decoder [91, 92, 93]. We wanted to study the natural structure of M1 activity in the V and VP feedback conditions, without the conflating influence of optimized decoder readout. We therefore focussed our analysis of M1 activity on the neural data recorded during the 2DP decoder training, when the subject imagined movement rather than affecting it with the neuroprosthetic.

The majority of recorded channels had very low, sub-5 Hz firing rates and Fano Factors of approximately 1. We therefore chose to re-analyze neural data using each channel's unsmoothed vector of spike counts $s_i(t)$. We fit new tuning curves, or mappings between each channel's spiking activity and the velocity of the virtual arm, according to the following Poisson regression:

$$\log(s_i(t)) = \beta_{i0} + \beta_{i1} v_x(t) + \beta_{i2} v_y(t) + \varepsilon_i(t) \tag{2.5}$$

where $v_x(t)$ and $v_y(t)$ are the x and y coordinates of the velocity vector $v(t)$ standardized by its Euclidean norm. Equation (2.5) can be restated as Equation (2.6), the cosine form of tuning commonly used to describe the velocity preferences of M1 neurons [76]:

$$\log(s_i(t)) = \alpha_{i0} + \alpha_{i1} \cos(\theta(t) - \theta_{iPD}) + \varepsilon_i(t) \tag{2.6}$$

where θ is the angle formed by the velocity vector $[v_x, v_y]$ and θ_{iPD} is the preferred direction of channel i. Channels for which the velocity tuning curve in Equation 2.6 did not explain significantly more variance than the spike count mean were excluded from all analyses that follow (Chi-squared test, $p < 0.05$ with Bonferroni correction).

The modulation depth of a tuning curve measures the strength of a channel's preference for θ_{iPD} over other velocity directions. Modulation depth was calculated as the difference between the channel's maximum and minimum fitted spike count standardized by the channel's maximum fitted spike count, and therefore always takes a unitless value between 0 and 1. Variability about fitted curves, or model dispersion, was estimated using the ratio between the residual deviance and the residual degrees of freedom of the model.

Correlations in the noise of the neural population describe the co-activity of channels once the effect of velocity tuning has been removed. Noise correlations were computed using the N x N pairwise correlation matrix Σ of each channel's tuning curve residual vector, $\varepsilon_i(t)$. Relatedly, the *mood* of the M1 population–which characterizes population-wide co-fluctuations of noise over time–was calculated by averaging residual vectors $\varepsilon_i(t)$ across all N channels. We characterized the degree to which a channel's non-velocity-tuned activity was coupled to the population mood by computing the Pearson correlation between the channel's residuals $\varepsilon_i(t)$ and the mood function.

2.3.7 A new, shared gain model of M1 tuning

Our post-hoc analysis of tuning curves revealed large co-fluctuations of noise in the VP condition. We therefore developed a novel model of M1 tuning in which we appended a time-varying signal that was shared amongst the neural population to the tuning model in Equation (2.5). The resulting tuning structure is described by the generalized additive model [94] below,

$$\log(s_i(t)) = \beta_{i0} + \beta_{i1} v_x(t) + \beta_{i2} v_y(t) + \gamma(t) + \varepsilon_i(t) \tag{2.7}$$

where $\gamma(t)$ is a single smooth function of time that is shared by all channels. Note that though $\gamma(t)$ is the same for every channel, the degree of coupling that each channel i has to the shared waveformform is allowed to vary through the intercept value β_{i0}. Also note that due to the log link function used in Poisson regression, $\gamma(t)$ modulates the gain of a channel's velocity tuning.

Equation (2.7) requires that all N channels be fit simultaneously. Our model design matrix $H \in \mathbb{R}^{TN\times(KN+1)}$ has the form

$$H = \begin{pmatrix} \tau_1 & V_1 & 0 & \dots & 0 \\ \tau_2 & 0 & V_2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \tau_N & 0 & \dots & 0 & V_N \end{pmatrix}, \tag{2.8}$$

where $V_i \in \mathbb{R}^{T\times K}$ is the matrix of K kinematic covariates to which each channel i is fit independently:

$$V_i = V = [\vec{1}, v_x(t)^\top, v_y(t)^\top], \tag{2.9}$$

and $\tau_i \in \mathbb{R}^{T\times 1}$ is a vector of time indices to which our shared function $\gamma(t)$ is simultaneously fit across each channel i:

$$\tau_i = \tau = [1, 2, ..T]^\top. \tag{2.10}$$

Because we did not have prior knowledge about the form of the shared gain waveform, we sought a smooth, flexible, and non-parametric transformation of the shared time index vector. We chose to model the gain waveform as a 1st degree local linear regression on the time index vector, corresponding to a kernel smoother in which each kernel is fitted with a weighted least squares linear regression between spiking activity and time index. Briefly, the general form of a local linear linear estimator $\hat{r}(x)$ of response variable Y with standard, Gaussian error distribution is

$$
\begin{aligned}
\hat{r}(x) &= \sum_i \ell_i(x) Y_i \\
\ell(x)^\top &= e_1^\top (X_x^\top W_x X_x)^{-1} X_x^\top W_x \\
X_x &= \begin{pmatrix} 1 & x_1 - x_0 \\ 1 & x_2 - x_0 \\ \vdots & \vdots \\ 1 & x_n - x_0 \end{pmatrix} \\
e_1 &= (1, 0, ..., 0)^\top \\
W_x &= w_i(x_0)\mathbb{I}
\end{aligned}
\tag{2.11}
$$

Our regression model is for a response variable with Poisson distributed error and requires a link function. See Fan et al. [95] for a generalized linear model treatment of local polynomial regression.

We used cross validation to select the kernel width of the smoother for each training session dataset independently. The kernel width selected by cross validation was always a physically interpretable number of time steps less than T, the duration of all reaches in the training session, despite that the total length of the time index column of H was TN. We fit the full model described by Equation (2.7) with the open source **gam** package in the scientific computing language **R**. The **gam** package fits generalized additive models using the backfitting algorithm [94], in which the partial residuals of each of the additive model terms are fit iteratively until the residual sum of squares of the full model satisfies the convergence criterion.

We evaluated the fit of the shared gain tuning model using the *deviance*, a scaled version of the log-likelihood ratio of two nested models. The total deviance of the shared gain model, which is equivalent to adjusted R^2, expresses the log-likelihood ratio between the shared gain model fit and a null model fit, where the null model is the mean spike count of a channel. We can also calculate the proportion of deviance explained by the added gain term $\gamma(t)$. This is equivalent to computing the log-likelihood ratio between the Equation (2.7) and Equation (2.5) model fits.

2.4 Results

2.4.1 M1 tuning changes in the presence of proprioceptive feedback

The subject first completed the 2DP decoder training with V and VP feedback, resulting in two sets of decoding parameters, W_V and W_{VP}. Each set of decoding parameters was used to perform the LC task, and the subject received either V or VP online feedback on a given trial (Figure 2.1a). This protocol allowed us to test the effect of mismatching decoder training protocol and task feedback condition. For example, the subject could perform the LC task with online VP feedback, even if the W_V decoder was being used to drive the movement of the MPL.

The subject's LC task performance degraded when we mismatched the decoder training protocol and the online task feedback. When using the W_V decoder, she completed significantly more line crossings in the V than the VP condition ($p < 0.05$), with an average of 27.9 ± 3.57 and 22.5 ± 3.37 line crossings (mean $\pm$ 1 SE, across all recording sessions) for the V and VP feedback conditions, respectively (Figure 2.1b). When using the W_{VP} decoder, she completed more line crossings in the VP (26.8 ± 3.01) than the V (24.3 ± 4.08) condition, though this effect was not significant, (mean $\pm$ 1 SE, across all recording sessions).

The quality of the subject's reaches during the LC task also degraded visibly when we mismatched the decoder training protocol and the online task feedback (Figure 2.1c-d). When the subject used the W_V decoder, the mean reach path length per line crossing was significantly smaller ($p < 0.01$) in the V (0.52 ± 0.009 m) than the VP (0.69 ± 0.045 m) condition (mean $\pm$ 1 SE, across all recording sessions). When the subject used the W_{VP} decoder, the mean reach path length per line crossing was significantly smaller ($p < 0.05$) in the VP (0.62 ± 0.019 m) than the V (0.67 ± 0.024 m) condition (mean $\pm$ 1 SE, across all recording sessions). The subject's decreased ability to control the reach when decoder training protocol and the online task feedback were mismatched was even more obvious when we analyzed the variance of her reaches in the task irrelevant dimension. Though the line crossing task was essentially one-dimensional, the subject had two-dimensional control of the MPL in the coronal plane. Optimal task performance would show negligible reach variability in the superior/inferior dimension, which was

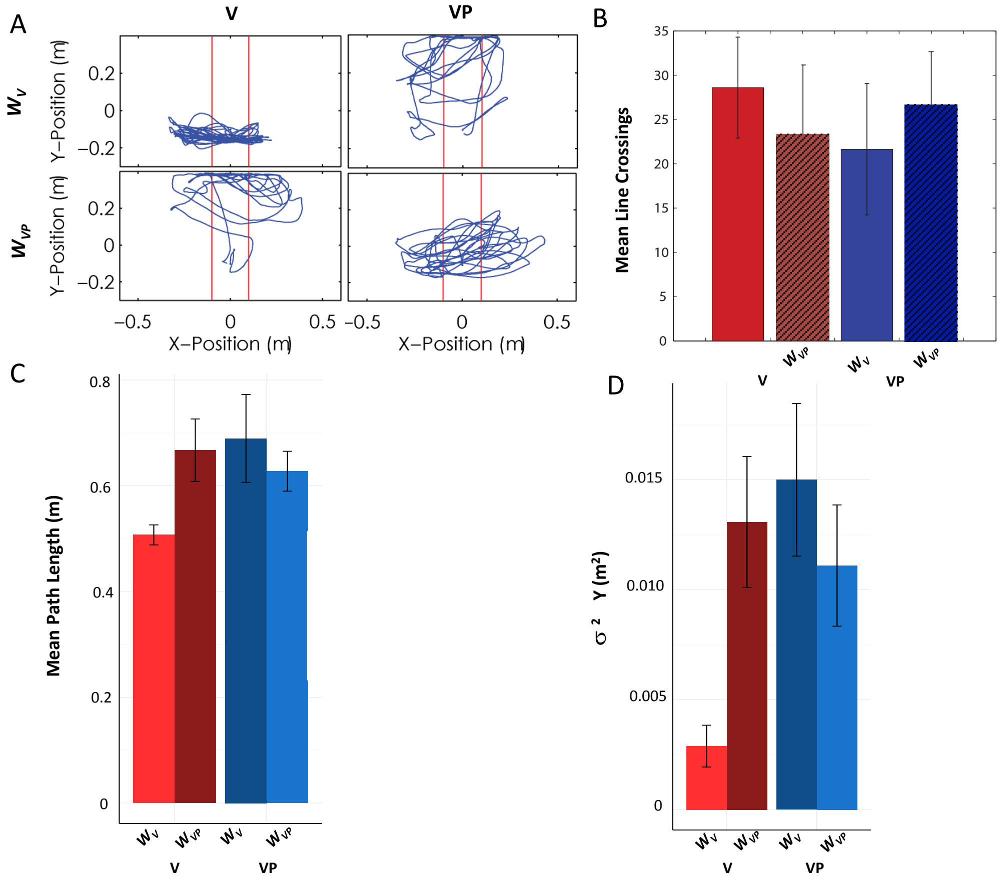

Figure 2.1: BMI-assisted 1D reaches suffer from proprioceptive feedback **(a)** Overhead view of single-trial reach trajectories in the BMI-assisted line-crossing task using the decoder training with vision alone (WV) or vision+proprioception (WV P) decoder. Online reach feedback was visual (V) or visual and proprioceptive (VP). **(b)** Mean number of line-crossings achieved in a 60 s period, pooled across four days of trials containing 5 runs each. Error bars represent 95% CIs. **(c)** Mean path length per line crossing, pooled across the same 4 days. Error bars represent 95% CIs. **(d)** Mean variance of reach in the task irrelevant dimension, pooled across the same four days. Error bars represent 95% CIs.

orthogonal to the LC goal. Variance in the task irrelevant dimension was significantly smaller ($p < 0.01$) in the V (0.0026 ± 0.00045 m^2) than the VP (0.015 ± 0.0018 m^2) condition when the subject used the W_V decoder (mean ± 1 SE, across all recording sessions). Similarly, variance in the task irrelevant dimension was significantly smaller ($p < 0.05$) in the VP (0.0118 ± 0.0011 m^2) than the V (0.0129 ± 0.0016 m^2) condition when the subject used the W_{VP} decoder (mean ± 1 SE, across all recording sessions).

The subject then completed the 2DP task (Figure 2.2a) with same 2 × 2 experimental block design that tested the interaction effects between decoder parameters and online task feedback

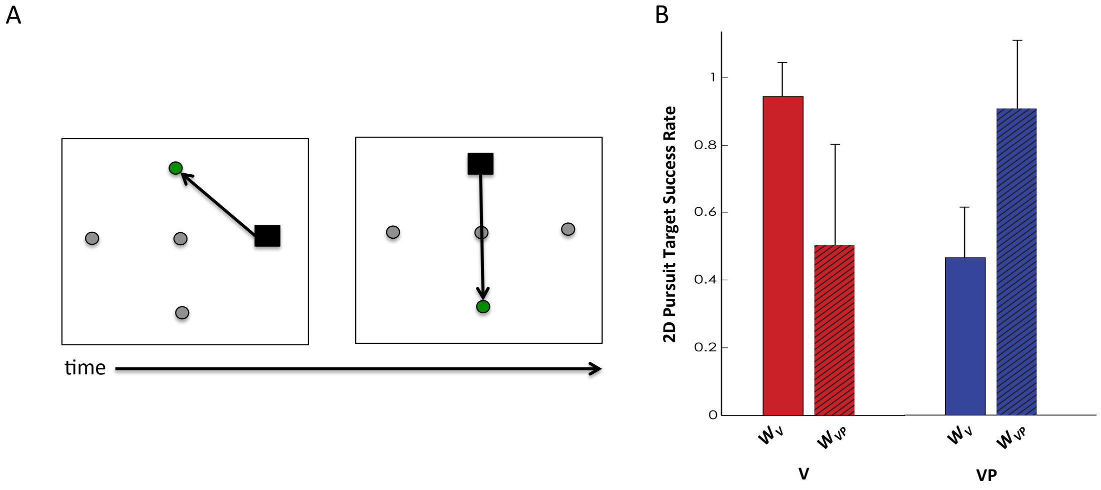

Figure 2.2: BMI-assisted 2D reaches suffer from mismatched feedback conditions **(a)** Schematic of the 2D pursuit task. The black box represents the virtual arm, the circles represent the 5 possible targets, and the green circle represents the target the subject was cued to pursue. Examples of 2 consecutive reaches are shown. **(b)** Average proportion of targets at which the subject successfully terminated a reach within a 3 s period during a single iteration of the 2D pursuit task, which contained approximately 30 reaches. Results are pooled across 5 iterations of the task in each condition. Error bars represent 95% CIs.

conditions. We quantified proficiency in the 2DP task by the fraction of times the subject successfully terminated a reach at the correct target within a 3 s period. As in the LC task, performance degraded significantly ($p < 0.01$) when we mismatched the decoder training protocol and the online task feedback (Figure 2.2b).

The subject's motor performance when using the W_V and W_{VP} decoders was not exchangeable, suggesting that the structure of M1 activity changed between the V and VP conditions. BMI decoders rely on a mapping between reach velocity and M1 activity, and failure of a BMI decoder necessitates that the mapping is no longer valid. We therefore sought to understand the difference between the map from kinematics to M1 activity in the V and VP conditions. Visualizations of channel firing rates as a function of velocity direction for the LC task revealed stable velocity tuning curves over multiple trials within a feedback condition, but not across feedback conditions (Figure 2.3a).

Tuning curves from the LC or 2DP tasks, when the subject affected movement with the BMI, might be a conflation of M1's natural tuning structure and a tuning structure M1 has learned to optimize the output a specific decoder [91, 92, 93]. We wished to investigate the natural state

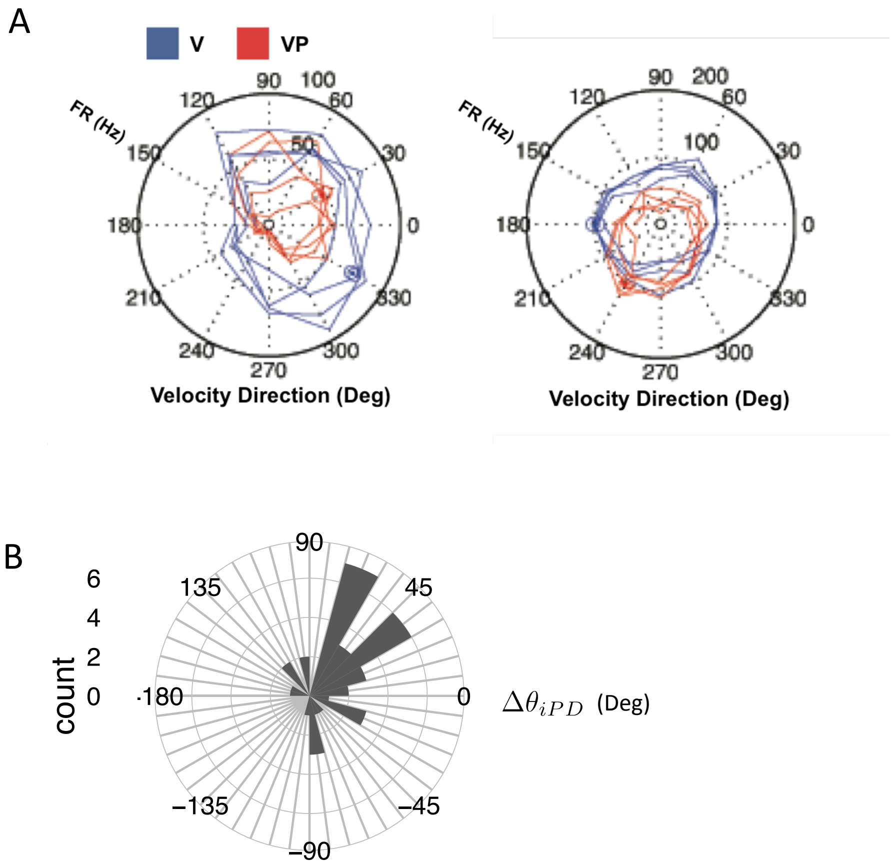

Figure 2.3: Shifts in preferred direction. **(a)** Firing rate as a function of velocity direction for 2 example M1 channels. Each curve represents one repetition of the line-crossing task, and four repetitions of the task are shown. Curves are stable within reach feedback type, but not across feedback types. Left: An example channel for which VP decreases modulation depth of velocity tuning. Right: An example channel for which there is a stable shift in θ_{iPD} between feedback conditions. **(c)** Distribution of the shift in preferred direction, $\Delta\theta_{iPD}$, between the V and VP conditions on the matched day.

of M1 velocity tuning in the V and VP conditions, not those potentially imposed by the W_V or W_{VP} decoder. To do this, we re-fit velocity tuning curves to the M1 activity recorded during W_V and W_{VP} training sessions, when the subject merely imagined movement. Our experiments were not designed with the purpose of detecting a shift in M1 velocity tuning between feedback conditions. This unexpected finding was only uncovered once we completed a more detailed analysis of the neural data, which was after the date when we were required to explant the subject's microelectrode arrays. Our data therefore contains only one session in which W_V and W_{VP} training was completed on the same day. This is the only day for which we can be reasonably certain that the activity of a channel i will reflect the activity of the same neuron or group of neurons for both the V and VP conditions. On this day, which we will henceforth refer

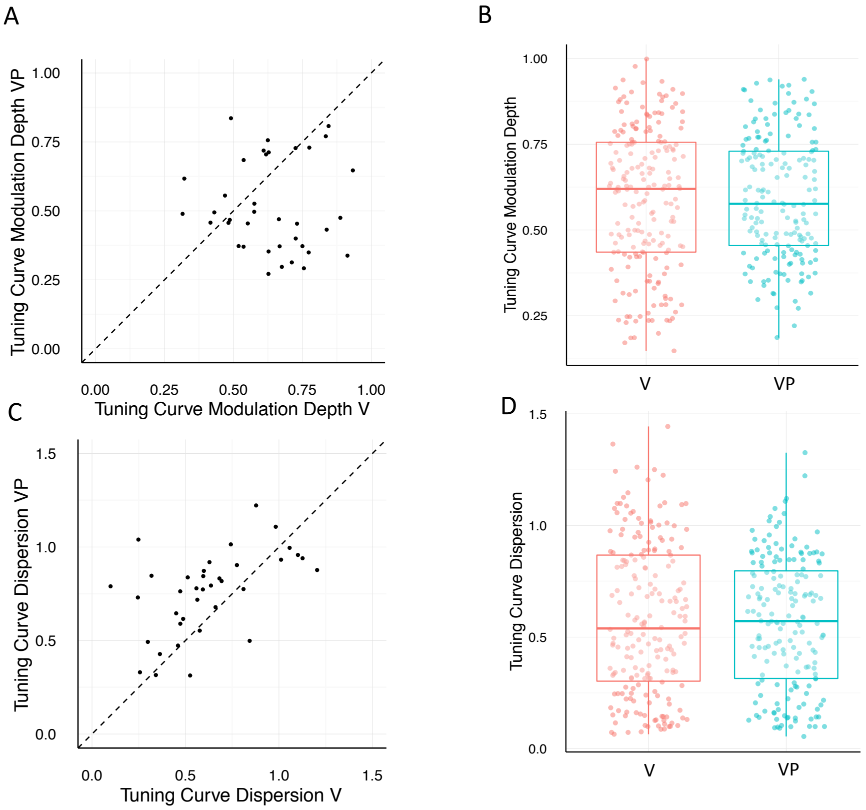

Figure 2.4: M1 responses have diminished velocity tuning in the VP condition. **(a)** Modulation depth of velocity tuning curves for each channel i on the paired day, compared across the V and VP conditions. Channels have larger modulation depth in the V condition. **(b)** Modulation depth of all analyzed channels in the V and VP conditions, pooled across four training sessions per condition. **(c)** The dispersion, or variability, about each channel i's fitted velocity tuning curve on the paired day, compared across the V and VP conditions. Velocity tuning is more variable in the VP condition. **(d)** Tuning curve dispersion of all analyzed channels in the V and VP conditions, pooled across the four training sessions per condition.

to as the *paired day*, there were roughly 60 tuned channels in the W_V and $W_V P$ conditions, and 37 channels that were tuned in both conditions. Of these 37 channels,$59.4 \pm 9\%$ (mean $\pm$ 1 SE, bootstrap) of channels exhibited $\geq 45°$ changes in the preferred velocity direction θ_{PD} (Eq. (2.5)) between the V and VP conditions (Figure 2.3b). The VP condition was also associated with a decrease in the modulation depth of the fitted velocity tuning curve. On the paired day, $37.8 \pm 6\%$ (mean $\pm$ SE, bootstrap) of channels exhibited $\geq 20\%$ decrease in modulation depth in the VP condition (Figure 2.4a).

2.4.2 Proprioceptive input results in shallower and more variable velocity tuning curves

One would intuitively think that additional sources of sensory feedback could only improve BMI performance. Contrary to that intuition, the quality of the subject's reaches in the LC task degraded strikingly in the VP condition even when the subject used the W_{VP} decoder. Reach variance in the task irrelevant dimension was significantly larger ($p < 0.001$) in the condition where VP feedback was provided to a W_{VP} decoder than the condition when V feedback was provided to a W_V decoder (Figure 2.1d). The mean path length per line crossing was also significantly greater ($p < 0.001$) in the condition where VP feedback was provided to a W_{VP} decoder than the condition when V feedback was provided to a W_V decoder (Figure 2.1c).

The behavioral result explained above cannot be accounted for by a shift in velocity tuning between the V and VP conditions, because we removed the effect of mismatched decoders. Instead, the behavior indicates that M1 activity must be less tuned to endpoint velocity in the VP condition (without accounting for latent sources of shared variability). We fit neural data from W_V and W_{VP} with velocity tuning models described by Equation 2.5 and calculated the dispersion of those tuning models, a measure of the spike count variability about the fitted tuning curve. Though both V and VP tuning curves had sub-Poisson variability (dispersion $<$ 1), dispersion was larger in the VP than the V condition. This was true to significant effect for the paired day training sessions (Figure 2.4c), for which we could directly compare the V and VP tuning curves of a channel i with confidence that the same neurons were being recorded in the two feedback conditions. VP tuning curves had greater dispersion on average when channel statistics were pooled across all four training sessions of each feedback type, but this effect was not significant (Figure 2.4d). As discussed previously, the modulation depth of tuning curves was also smaller in the VP condition than the V condition. Tuning curves were therefore broader (i.e., more overlapping) and more variable in the VP condition, which would result in more variable BMI output for a given reach intention.

Finally, we observed that neural data in the VP condition contained co-fluctuations of activity that were not tuned to velocity and that were shared by large subpopulations of channels. Correlations in noise–the neural activity once the mean effects of the velocity tuning were removed–were greater in the VP than the V condition on the paired day, when there is high

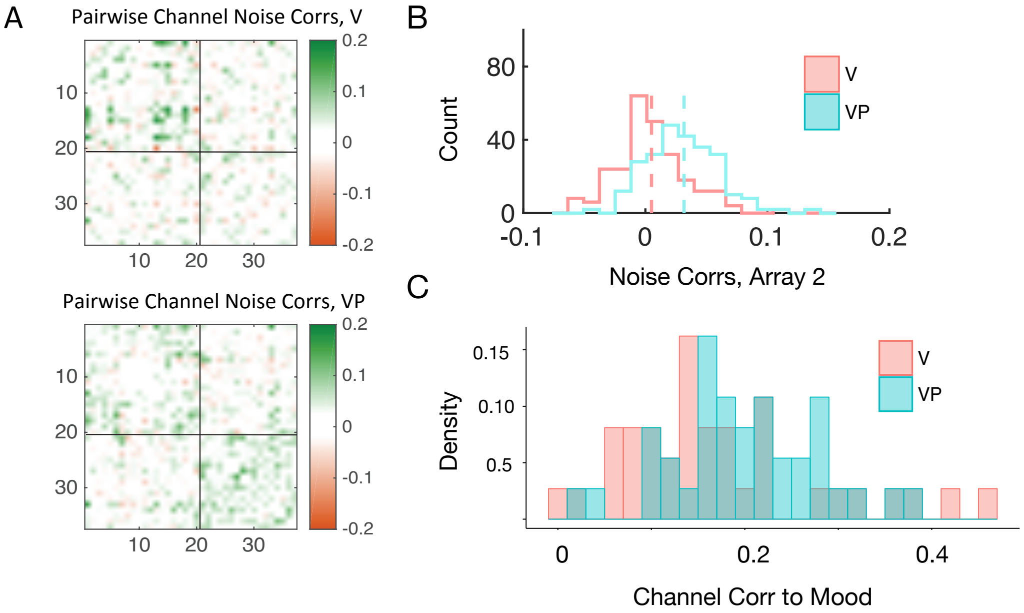

Figure 2.5: Low-dimensional shared variability in the VP condition **(a)** Pairwise noise correlations for paired day channels in the V (upper) and VP (lower) conditions. Channels are ordered according to their array locations, the black solid lines delineate the two electrode arrays. Array 2 consists of channels 21 through 38. **(b)** Distribution of noise correlations on Array 2. **(c)** Distribution of the correlation between the residuals of each channel on the paired day and the *mood* of the population, or the average residual activity as a function of time across channels. Channels are more correlated with the mood in the VP condition, indicating that noise fluctuations in the VP condition are shared by many channels.

probability that the same neurons were being recorded in feedback conditions (Figure C.1a-b). An anatomical mapping of noise correlations revealed that large, proximally-located subpopulations of channels on the microelectrode arrays co-fluctuated about their individual tuning curves in the VP condition. The VP condition was specifically marked by a significant increase in noise correlations on Array 2 of our MEAs (Figure C.1b, $p = 1.5 \times 10^{-19}$, two-sided Wilcoxon signed rank test). We characterized the *mood* of the neural population–the global, un-tuned activity as a function of time–using the average of the activity of the all N analyzed neurons after the spiking activity due to velocity tuning was removed. We then measured the correlation between each channel's residuals and the mood. Individual channels had greater correlation with the mood in the VP condition (Figure C.1c), indicating that there were global co-fluctuations in the noise in the VP condition.

2.4.3 A shared gain model of somatosensory feedback in M1

Our neural data analysis in Sections 2.4.1 and 2.4.2 revealed spiking co-variability in the VP condition that could not be explained by Equation 2.5, in which the spiking activity of each channel i was conditioned exclusively on its private velocity tuning. We developed a new model of M1 activity that quantifies global gain fluctuations in the VP condition. In the new model, the spiking activity of each channel i is conditioned on its private velocity tuning and a waveform of activity shared by every channel (Equation 2.7). The global waveform modulates the gain of each channel's private velocity tuning curve; the degree of that modulation is controlled by a term that represents the channel's coupling strength to the global activity. Examples of $\gamma(t)$ are shown in Figure 2.6. We were struck by the observation that $\gamma(t)$ seemed to fluctuate at time scales relevant to the duration of one imagined reach, despite the fact that we placed no constraint on our modulator's timescale (in our regression format, the temporal neighborhood of smoothing).

Moreover, when we compared the form of $\gamma(t)$ across reach epochs, we found that its latent, 1D dynamics were remarkably stereotyped across repeated reach trajectories (Figure 2.6b). Many potential factors could give rise to shared variability with stereotyped dynamics; proprioceptive information is one such factor that would be strongly correlated with reach type, as different subsets of muscle and tendon receptors are activated for reaches across body space. We are cautious of making bold claims about this observation, especially because we have very few repeated reaches for a given recording session. The form of $\gamma(t)$ fluctuations is probably not comparable across days, over which period it is unlikely that we are fitting our regression models to the same neural population across. Even if we were observing the same neurons over multiple recording sessions, we would likely need to account for population-wide slow drifts in neural activity to align $\gamma(t)$ over days [61, 96]. But slow drifts would share scaling with our formulation of $\gamma(t)$, making the two difficult to disambiguate. We hope future studies with more repeated reaching trials and statistical power might examine the latent dynamics of trial-to-trial variability in M1 to see if they have stereotyped form related to somatosensory feedback. Many studies have found evidence of latent manifolds that constrain M1's tuned activity during movement control [19, 97, 98], but few have isolated and examined the population structure of residual activity [58].

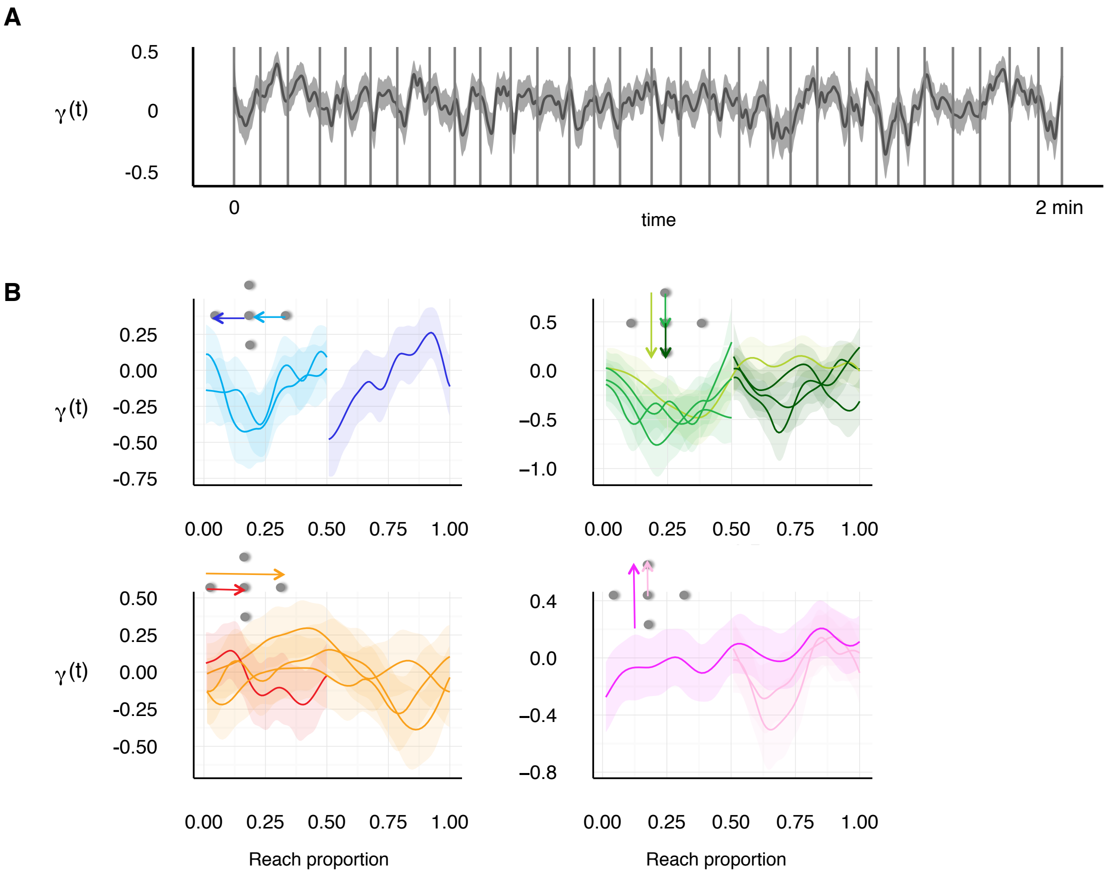

Figure 2.6: Latent shared gain trajectories $\gamma(t)$ in M1 trial-to-trial activity **(a)** Example of a shared gain signal $\gamma(t)$, fitted using the shared gain model in Equation 7. The entire timecourse of $\gamma(t)$ across all 30 reaches is shown, where reach epochs are delineated with vertical lines. The confidence band represents a 95% CI. The regression is fitted with a log-link function, so negative values of $\gamma(t)$ are interpreted as gain depression. **(b)** Shared gain signal $\gamma(t)$ segmented by reach type. Confidence bands represent 95% CIs. On a single plot, reaches of the same type are shown in the same color. The type of reach is indicated by the corresponding target diagram in the upper lefthand corner of the plot. $\gamma(t)$ traces are aligned using reach proportion (x axis), which the fraction of the reach completed, where the completion of a reach across two targets in either the x or y direction has a value of 1.

The addition of a shared gain term improved model fits dramatically for both the V and VP condition. Figure 2.7 shows the adjusted R^2 of the entire tuned population for the cosine tuning model (Equation 2.5) and the shared gain model (Equation 2.7) in both the V and VP conditions. The shared gain term explains 54.9% and 33.9% more spiking variability on average in the V and VP conditions, respectively, where averages were computed across training sessions with the same feedback conditions. However, variability in results across training sessions were large, especially in the V condition.

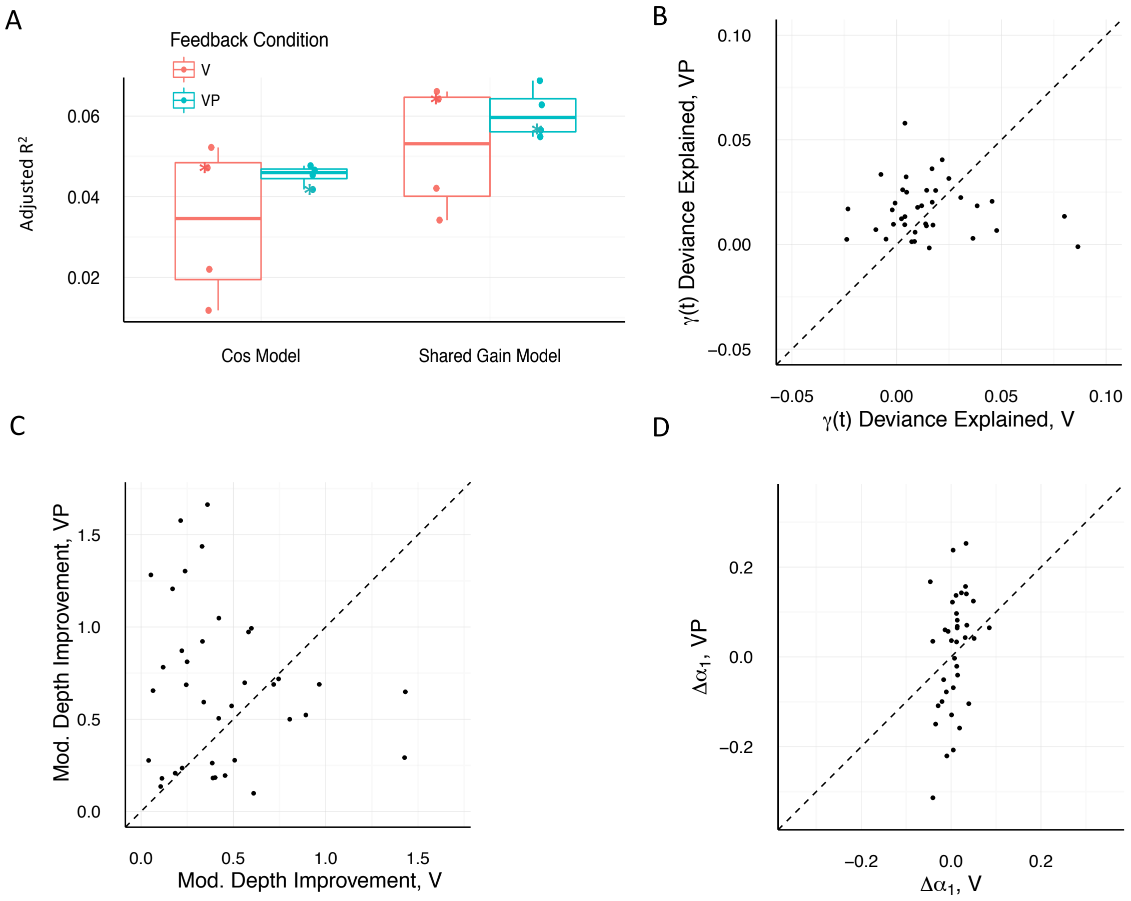

Figure 2.7: The shared gain model improves M1 tuning differentially in the VP condition. **(a)** Adjusted R^2 of the full population model (containing all N channels) for the V and VP conditions and the Cosine Tuning and Shared Gain models. Each point represents one analyzed training session, and there were four training sessions of each feedback condition. The paired day results are indicated with a *. Though the shared gain model assists tuning in both the V and VP conditions, the paired day shows more improvement for the VP than the V condition. **(b)** The proportion of deviance explained by $\gamma(t)$ for each channel i in the V and VP condition. Single channel values may be negative because all channels in the Shared Gain model are fit simultaneously (and the result will be the best fit for the entirety of the population). More deviance is explained by $\gamma(t)$ in the VP condition. **(c)** The improvement in the modulation depth of each channel i's velocity tuning after the addition of $\gamma(t)$, on the paired day. More channels show greater modulation depth improvement with the addition of $\gamma(t)$ in the VP condition. **(d)** Change in the amplitude (α_1, Eq. (2.6)) of velocity tuning for each channel i on the paired day with the addition of $\gamma(t)$. Channels show larger amplitude shifts in the VP condition, and therefore $\gamma(t)$ has greater effect.

We can also calculated the proportion of deviance explained explicitly by the shared gain term $\gamma(t)$ for each of the i fitted channels in a training session. On the paired day, the shared gain term explained a larger proportion of deviance in the VP condition than the V condition for 59.5% of channels (Figure 2.7b), indicating that $\gamma(t)$ aids the mapping between kinematics and spiking activity more in the VP than the V condition. Similarly, $\gamma(t)$ caused a differential increase in the modulation depth of channel tuning curves in the VP condition in 64.8% of channels (Figure 2.7c). This effect arises because the inclusion of a $\gamma(t)$ waveform causes much larger shifts in VP tuning curve parameters than V tuning curve parameters. With the addition of $\gamma(t)$, channels on the paired day showed larger standardized changes in the amplitude of their cosine velocity tuning curves (α_{i1}, Equation 2.6) in the VP condition than the V condition (Figure 2.7d). Together, these results suggest that global co-fluctuations of the neural population in the VP condition mask channels' private velocity tuning; once we condition spiking responses on a shared gain waveform $\gamma(t)$ in the VP condition, channels are more sharply tuned to endpoint velocity.

2.5 Discussion

We had the first known opportunity to study the neural differences between feedforward and closed-loop, somatosensory-assisted motor control in a human subject with a motor-specific neuropathy. We found that BMI-assisted motor control significantly degraded when there was a mismatch in the sensory feedback sources provided during decoder training and online BMI reaching performance. The change in motor behavior between the visual (V) and visual+proprioceptive (VP) conditions was marked neurally by significant shifts in the preferred directions, gains, and variability of M1 velocity tuning curves.

Our results are at odds with those of a previous study from Suminski et al. [87], which found that online proprioceptive feedback improved BMI-assisted reaching behavior in non-human primates even when the BMI decoder was trained with vision alone. The result from Suminski et al. [87] would require that M1 velocity tuning remain stable in the presence/absence of proprioceptive feedback, and that proprioception simply improve the signal-to-noise ratio of stable velocity tuning curves. Other studies have demonstrated instabilities in M1 tuning during sensory adaptation [99] and motor learning [100], lending credence to our conclusion that the mapping from kinematics to spiking activity shifts in the presence of proprioception.

There are several notable differences between our experiment and that published by Suminski et al. [87], which could account for the contrasting results. First, the neural data in their study was automatically sorted by the voltage waveform of the recorded spikes to ensure that tuning models were fit to the activity of single neurons. We chose not to spike sort the activity of our recorded channels, in part due to research demonstrating that BMIs have trivial reductions in performance when channel activity is merely thresholded rather than sorted, and successful spike sorting requires large quantities of data collected at high sampling rates [101]. We also observed that very few channels of our recorded data had supra-threshold single-unit activity; spike sorting activity for single units would therefore remove much of the information we recorded from the M1 population. It is therefore possible that the shift in tuning that we observed in the VP condition could be do to the activation of different subsets of cells recorded by a single electrode, and not to a shift in the velocity preference of single neuron. However, this explanation would be much more likely if we observed preferred direction shifts on only a few channels; instead, we observed shifts ≥ 45 degrees on the majority of recorded channels. The non-human primate subjects in the Suminski et al. [87] study also had years of experience using visually-trained BMI decoders of the form used in the experiment. We consider it more likely that the M1 activity of the non-human primates had adapted to optimize that decoder output [91, 92, 93]; this adaptation may have masked natural representations of somatosensory information.

More unexpectedly, we found that proprioception degraded reaching performance even when subject used a decoder trained with proprioceptive feedback. This is an unintuitive result given several proposed neural models of Bayesian multisensory integration, which posit that sensory state estimation is achieved through a linear combination of the sensory feedback sources, with weights that are inversely proportional to the variance of the feedback source [102, 103]. In this framework, state estimation can only be as bad as the most variable sensory cue. One would therefore think that motor responses to simultaneous visual and proprioceptive feedback would be at least as reliable as motor responses to visual feedback alone. However, the behavioral consequence of sensory cue integration is dependent on the read-out of the circuit. Circuits native to or downstream of M1 may be able to effectively integrate visual and somatosensory information such that additive proprioceptive feedback decreases the variability of motor output. Moreover, these read-out circuits may in fact be linearly summing basis patterns of activity encoded by visual and somatosensory feedback, but the basis patterns may not be of the tuning forms recognized by our current BMI decoder implementations.

Our results showed that a gain-modulating waveform of activity, which was shared across the M1 population, masked the effects of velocity tuning on spiking variability, particularly when M1 received somatosensory feedback. Our findings come amidst a series of recent publications that have discovered low-rank neural activity in several sensory regions of cortex [55, 22, 23]. These large network co-fluctuations have been shown to dominate and mask local circuit effects when their presence is not accounted for [72]. Though the co-fluctuations of activity uncovered by all of these studies are not due to the effects of the stimulus presented during the experiment, the mass coordination of the activity suggests that it is an effect of the circuit that shapes computation and not simply “noise”.

Though the source or circuit mechanism of the gain-modulating signal in our model remains uncertain, there are reasons to believe the shared gain term could represent communication from S1 to M1. First, the shared gain model aids the tuning of neural data in the VP condition more than the V condition. Additionally, sensorimotor circuit research has shown evidence that M1 neurons most responsive to sensory information are intermingled with motor-tuned neurons [104]. Furthermore, despite the fact that the S1 tonotopy is roughly preserved in S1 to M1 projections [105, 106], S1 tonotopic projections terminate in overlapping regions of M1, and wide M1 neuron dendritic activity results in single M1 neurons integrating information from broad somatotopic regions of S1 [88]. Finally, previous studies has found evidence of traveling waves of correlated fluctuations between sensory and motor areas, which would not be dissimilar in form from our modeling choice of $\gamma(t)$ [107]. We do not rule out the possibility that the shared gain term in our model represents converging somatosensory input as well as other sources of network co-variability [96]. This could explain why a shared gain signal still aided M1 tuning in the V condition.

In its current form, our shared gain model of M1 tuning presents a few challenges for translation to clinical BMI work. Though it is straightforward to invert our tuning model and create an Optimal Linear Estimator (OLE) [108] that predicts velocity kinematics from the neural data, doing so requires us to have advanced knowledge of the form of the shared gain signal over time. In other words, a decoder trained with one set of neural data might not have a shard gain signal that was generalizable to a new set of neural data encountered during online decoding. This challenge might be overcome if the single-trial activity of the shared gain signal has predictable dynamics, as suggested by our results in Figure 2.6. Additionally, the shared gain signal may be partially co-linear with velocity kinematics, meaning that neural data would contain less information about velocity once the effects of the shared gain were subtracted in the decoding process. Though just as others have espoused the benefits of using BMI as a tool for basic science research [109, 110], we view this study primarily as a rare opportunity to gain insight into human sensorimotor coding. Clinical BMI decoders that utilize somatosensory feedback will progress as we continue to gain a better understanding of the communication between S1 and M1.

3. *Assembly structure expands the dimension of shared variability in cortical networks*

3.1 Abstract

Cortical circuits often receive multiple inputs from upstream populations with non-overlapping stimulus tuning preferences. Both the feedforward and recurrent architectures of the receiving cortical layer will reflect this diverse input tuning. We study how population-wide neuronal variability propagates through a hierarchical cortical network receiving multiple, independent, tuned inputs. We present new analysis of *in vivo* neural data from the primate visual system showing that the number of latent variables (dimension) needed to describe population shared variability is smaller in V4 populations compared to those of its downstream visual area PFC. We successfully reproduce this *dimensionality expansion* from our V4 to PFC neural data using a multi-layer spiking network with structured, feedforward projections and recurrent assemblies of multiple, tuned neuron populations. We show that tuning-structured connectivity generates attractor dynamics within the recurrent PFC current, where attractor competition is reflected in the high dimensional shared variabilty across the population. Indeed, restricting the dimensionality analysis to activity from one attractor state recovers the low-dimensional structure inherited from each of our tuned inputs. Our model thus introduces a framework where high-dimensional cortical variability is understood as "time-sharing" between distinct low-dimensional, tuning-specific circuit dynamics.

3.2 Introduction

Contemporary recording technologies have enabled us to simultaneously monitor the single unit activities of large populations of neurons within and across cortical areas [7]. Neuroscientists have since sought to understand the patterns of activity across many neurons in a cortical circuit, as this provides insight into the population dynamics underlying sensory and motor computations. The trial-to-trial fluctuations of neural responses are one important measure of neural population dynamics that help characterize and differentiate private sources of neural variability from shared fluctuations in activity due to common afferent projections or recurrent network architectures [16, 17].

Network models offer a means to control the wiring rules governing network architecture and then observe the resulting structure and propagation of trial-to-trial variability in model neuron responses. As such, these models are a tool uniquely suited for the difficult challenge of relating network connectivity to network dynamics. Early spiking network models commonly employed uniform, random recurrent connections with balanced excitation and inhibition to internally generate the large variability observed in single neuron responses *in vivo* [111, 63, 112]. However, balanced network models with uniform connectivity produce asynchronous spiking dynamics with mean zero co-variability between the trial-to-trial responses of pairs of neurons [64]. These classical balanced models are thus inadequate to describe several datasets in which neurons have positive noise correlations, or shared variability [113, 114, 115, 54, 22]. More recent recurrent network models have used non-uniform connectivity to produce neural activity with positive, structured noise correlations. Recurrent networks with spatially-dependent coupling profiles can produce spatially-dependent noise correlations [71, 72, 116, 117]. Other wiring architectures that give rise to structured shared variability include local connectivity motifs [118, 119, 120, 121, 122], coupling described by connectivity matrices of constrained form [123, 124], and assembly structures [60].

A handful of studies have explicitly investigated how non-uniform recurrent connectivity architectures determine the dimension of shared variability. Huang et al. [24] internally generated low-dimensional shared variability in spiking networks with slow inhibitory kinetics and spatially-dependent coupling profiles. Recanatesi et al. [125] demonstrated that the rank of shared variance in recurrent networks depends on the local connectivity motifs, like chains and loops, that make up the network's full architecture. Mastrogiuseppe & Ostojic [123] were able to predict the minimum rank of a connectivity matrix required to implement a computation of specified complexity. Williamson et al. [59] discovered that the dimension of a spiking network's shared variance scaled linearly with the number of excitatory cell assemblies in Litwin-Kumar & Doiron [60]'s clustered spiking network framework. All of the above models provide mechanisms by which shared variability of determinate rank is internally generated through recurrent network interactions in one layer of a cortical circuit. However, this class of "internally-generated variability" models does not consider variability inherited from outside sources. Conversely, there are models that accurately capture the structure of neural data with low-dimensional shared variance by assuming that structure is inherited from fluctuations in an external brain area [16, 126, 127, 128, 129]. This class of "externally-inherited variability" models has historically not considered how recurrent network interactions might affect the rank of inherited co-variability.

Our study's central goal is to create a unifying theory of how structured recurrent connectivity can transform the dimension of the shared variability inherited from external brain regions. We are motivated by our finding that the dimension of shared variance expands between multiple, simultaneously recorded regions of a visual circuit. We first present novel data analysis of neural activity recorded *in vivo* showing that the dimension of shared variance in prefrontal cortex (PFC) is significantly greater than that of upstream visual area V4. We note that the significant dimensionality expansion between V4 and PFC activity either precludes linear dynamics in PFC or necessitates that PFC receives many additional, unobserved cortical inputs. While it is well known that PFC receives information from several sensory cortices to implement integrative brain functions [130, 131, 132, 133], we believe that appealing to unobserved data in order to explain our PFC activity "punts" the responsibility of developing a mechanistic theory of the cortical region's dynamics. We instead choose to develop a parsimonious model in which PFC can expand the dimensionality of shared variance inherited from V4 through non-linear recurrent interactions alone. We note that our parsimonious modeling choice still self-consistently allows for additional PFC inputs not observed in this study.

Our model accounts for PFC's laterally-tuned projections carrying visual inputs from both hemispheres of the brain. We show that a strongly coupled, balanced recurrent network exhibits multi-stable, non-linear dynamics when receiving disjoint projections from multiple independent, tuned upstream neuron populations. When the activity from each of these tuned upstream neuron populations is highly self-correlated, as is the neural activity in a single V4 hemisphere [32, 23], each attractor state of dynamics is marked by pathological anti-correlations between the activity of recurrent cells receiving opposing projections. Motivated by studies showing that cortical neurons have clustered recurrent architecture reflecting tuning preferences, we investigate the consequences of adding clustered excitatory (E) and inhibitory (I) connections to our model PFC network. We find that tuning-specific recurrent assemblies of E and I cells diffuse the pathological anti-correlations induced by lateralized input projections and temper winner-take-all recurrent dynamics. We identify a degree of clustering at which our recurrent network model successfully predicts the dimension of shared variance observed in our PFC neural data recorded *in vivo*. Finally, we show that the high-dimensional shared variance generated by our recurrent network model is the result of "time-sharing" between multiple states of activity, each of which independently has low-dimensional, linear dynamics reflective of our tuned inputs. Together, our results provide a new circuit-based model by which recurrent networks with structured connectivity can internally amplify the shared variability they inherit from upstream brain areas. Furthermore, our results provide a framework in which the high-dimensional dynamics commonly observed in sensory integration areas of the brain can be decomposed into multiple states of interpretable linear dynamics representing discretely tuned neural populations.

3.3 Methods

3.3.1 Experimental methods

All neural data were collected by the students and staff of the Smith Laboratory (formerly University of Pittsburgh Department of Opthalmology, currently Carnegie Mellon University Department of Biomedical Engineering.) All procedures were approved by the University of Pittsburgh Institutional Animal Care and Use Committee and complied with the National Institute of Health's *Guide for the Care and Use of Laboratory Animals*.

A 96-electrode "Utah" Array (Blackrock Microsystems, Salt Lake City, UT) was implanted in right 8Ar of the dorsolateral prefrontal cortex (PFC) of a male rhesus macaque (*Mucaca* *mulatta*). The PFC array was positioned on the pre-arcuate gyrus, medial to the principal sulcus and anterior to the arcuate sulcus. A second 96-electrode Utah Array was implanted in right V4.

The implanted monkey performed a memory guided saccade (MSG) task as follows. The monkey watched a 21” monitor with 1024x768 pixel resolution and 100 Hz refresh rate from 36 cm viewing distance. A 0.5 diameter dot appeared at the center of the screen. After fixation was established for 200 ms, a target appeared for 50 ms at one of forty 2D screen locations, defined by its coordinates at one of eight angular location (0° to 315° in increments of 45°) and one of five radial distances (5 degrees visual acuity (dva), 7.5 dva, 9.9 dva, 12.3 dva, 14.7 dva). When the target disappeared, the monkey was required to maintain fixation at the center of the screen for 500 ms. The disappearance of the central fixation point would then signal the monkey to make a saccade to the remembered target location. Each recording session consisted of blocks of 40 trials, in which a random 1 of the 40 possible target locations was presented on each trial. At least 40 blocks, or presentations of a given target condition, were collected during a single recording session. Further details about the array implantation, task, and neural data collection procedures can be found in Khanna et al. [134] and Cowley et al. [61].

3.3.2 Spike train statistics

Spike train statistics for both recorded neural data and the spiking network model realizations were computed as follows. A neuron i spikes at times $\{t_{i1}, t_{i2}, t_{i3}, \dots\}$. Neuron i’s spike train is then defined as $y_i(t) = \sum_k \delta(t - t_{ik})$, where $\delta(t - s)$ is a Dirac delta function centered at time point s. The number of spikes emitted by the neuron between times t and $t + \Delta t$ is

$$N_i(t, t + \Delta t) = \int_t^{t+\Delta t} y_i(t')dt'. \tag{3.1}$$

The firing rate of neuron i over interval $(t, t + \Delta t)$ is defined as:

$$f_i(t, t + \Delta t) = \frac{1}{\Delta t}\langle N_i(t, t + \Delta t)\rangle, \tag{3.2}$$

where $\langle\cdot\rangle$ denotes an expectation over trials. Finally, the Fano factor of neuron i is defined as

$$F_i(t, t + \Delta t) = \frac{\text{Var}\left(N_i(t, t + \Delta t)\right)}{\langle N_i(t, t + \Delta t)\rangle}. \tag{3.3}$$

The spike count correlation coefficient between neurons i and j is their covariance normalized by the geometric mean of their variances:

$$\rho_{ij}(t, t+\Delta t) = \frac{\text{Cov}\left(N_i(t, t+\Delta t), N_j(t, t+\Delta t)\right)}{\sqrt{\text{Var}\left(N_i(t, t+\Delta t)\right)\text{Var}\left(N_j(t, t+\Delta t)\right)}}. \tag{3.4}$$

Unless otherwise specified, spike count covariance and spike count correlation analyses of both the neural data recorded *in vivo* and the simulated data were performed on populations of excitatory neurons, and all spike train statistics were averaged across trials.

3.3.3 Neural data preparation

The neural data recorded were sorted into single unit activity. Units without at least a 2.5 signal-to-noise ratio (SNR), defined as the ratio of the average waveform amplitude to the standard deviation of the waveform noise, were disregarded. Remaining units were included in our analyses and taken to represent single neuron activity. Only neurons with a mean firing rates of at least 1 Hz ($\Delta t = 500$ ms,(3.2)) for all target locations relevant to a given analysis were used in that analysis. The analyzed neural population was further filtered to include only neurons showing evidence of spatial tuning. To test tuning specificity, PFC neural responses were calculated from 10 ms to 260 ms after the presentation of the target. These peak delay period responses were baseline-corrected by subtracting a neuron's average activity across all conditions in the 30 ms to 180 ms epoch after fixation. A Kruskal-Wallis one-way analysis of variance on a neuron's average firing activity across locations was then used to determine whether a neuron had significant ($p < .05$) spatial tuning.

Each neuron's full spatial response field was obtained by averaging its baseline-corrected response across multiple presentations of the same target condition. The resulting response portrait over target space was linearly interpolated to obtain 0.25 dva x 0.25 dva resolution response map, which was then convolved with a 2D gaussian filter of 1 dva variance for smoothing. A neuron's preferred spatial location was defined as the center of mass (COM) of all map points where the magnitude of the neural response was at least 75% that of the maximum response on the map. Responses that were suppressed below baseline were not considered in the COM calculation. Response fields presented are oriented such that the contralateral visual hemifield encodes the left half of the image displayed to the monkey.

3.3.4 Noise correlation analyses of neural data

Spike counts of PFC neurons were summed over the delay period of the task ($\Delta t = 500$ ms,(3.2)) for each trial. Neural responses for trials of the same target condition were then normalized such that each target condition had mean a spike count of zero and variance of one. The residual spiking activity around the baseline response of each target condition represents a neuron's response to "noise", or sources of fluctuation other than the stimulus tuning. Pearson correlation coefficients were computed between these noise responses of all pairs of neurons (3.4). Pairs of neurons were then organized according to the Euclidean distance between the COMs of the two neurons' spatial response fields, in bin increments of 2.25 dva.

3.3.5 Factor Analysis of neural data

To understand the dimensionality of the noise fluctuations in both layers of our network, we performed Factor Analysis (FA) [135, 136] on the simultaneously recorded PFC and V4 neural data. Target conditions on the right half of the visual field do not evoke meaningfully tuned responses in the recorded V4 hemisphere. All FA results (in both V4 and PFC neural populations) were therefore calculated using only the 25 stimulus conditions on the left visual hemifield, in which the target orientation was between 90 and 180 dva, inclusive of those boundaries.

PFC neural responses were analyzed over the task delay period, while V4 neural responses were analyzed during the period of target presentation. PFC spike counts were computed over a 0 to 540 ms interval following the target presentation, binned in windows of 180 ms. V4 spike counts were computed over a 0 to 90 ms interval beginning with the target presentation, binned in windows of 30 ms. Note that by binning spike counts with different window sizes for V4 and PFC data, we were able to instead hold constant the number of observations used in FA for each brain area. FA is sensitive to the number of data observations it receives [59]. Our bin width choices result in the same number of observations of PFC and V4 neural data, despite the fact that PFC data was measured surrounding a 500 ms delay period and V4 data could only be meaningfully measured surrounding the considerably shorter 50 ms target presentation period. In making this choice, we assume that the firing activity of PFC neurons is stationary over 180 ms periods, which is reasonable given the average firing rate of a neuron in the PFC population was < 5 Hz.

Neural data from both brain areas were normalized for each target condition such that the mean firing rate per target condition was zero. Spike count observations from all trial repetitions of all left hemifield target conditions were then concatenated into a single matrix of data observations per brain region in prepation for FA. Each matrix of data observations used in FA represented sorted single unit activity during a single recording session. We assume that data collected throughout a single recording session were from consistent V4 and PFC neuronal populations. FA was never performed on data pooled across multiple recording sessions, as we did not attempt to track single unit activity across days of recording.

Appendix A provides a detailed description of the computations underlying FA. In brief, FA finds a latent basis set that describes the shared variance of the analyzed neural data. We order the latent dimensions of this basis set by their percentage of shared variance explained. We define d_{shared} as the number of (ordered) latent dimensions required to cumulatively explain 95% of the neural data's shared variance. We calculated d_{shared} of the V4 and PFC neural data over each recording session. The FA loading matrix $\mathbb{L}$ describes how the activity of individual neurons loads onto each latent dimension. We computed the loadings of each PFC neuron onto the top 3 latent dimensions $\mathbb{L}_{1:3}$ to determine whether PFC neurons that preferred the ipsilateral versus contralateral visual hemifield were separable in the latent space.

3.3.6 Spiking network simulations

Layer 1 of our spiking network model represents two inputs from V4, encoding both the left (L) and right (R) visual hemifield. The neural activity of each model V4 hemifield was simulated as a doubly-stochastic process. V4 neurons with a given hemifield preference $h \in L, R$ shared a rank 1 fluctuation generated by the Ornstein-Uhlenbeck (OU) process

$$\tau \frac{d\lambda_h}{dt} = \bar{\lambda} - \lambda_h + \sqrt{\sigma^2 \tau} \xi_h(t), \tag{3.5}$$

where λ_h is the shared firing rate of hemifield h, $\bar{\lambda}$ is the baseline firing rate of that hemifield, $\xi_h(t)$ is the hemifield's shared white noise process, and τ and σ are the timescale and amplitude of the shared fluctuations, respectively. Each neuron i in hemifield h then emitted spikes according to the Poisson process

$$v_{i_h}(t) \sim \text{Pois}(\lambda_h) \tag{3.6}$$

Each V4 hemifield consisted of $N_L^{\text{FF}} = 2000$ excitatory neurons preferring the left visual hemifield and $N_R^{FF} = 2000$ excitatory neurons preferring the right visual hemifield, where the superscript FF denotes that these neurons make only feedforward projections.

Layer 2 of our model represents PFC and consists of excitatory (E) and inhibitory (I) populations of $N_E^{\text{Rec}} = 4000$ and $N_I^{\text{Rec}} = 1000$ neurons, respectively, where the superscript Rec denotes that these neurons make recurrent connections. Each Layer 2 neuron was modeled as a leaky integrate-and-fire unit obeying membrane potential dynamics given by:

$$\dot{V} = \frac{1}{\tau_{\text{mem}}} (\mu - V) + I_{syn}(t). \tag{3.7}$$

Neurons emit spikes when they reach the voltage threshold $V_{th} = 1$ in our non-dimensionalized units, at which time they are reset to $V_{re} = 0$ for an absolute refractory period of 5 ms. All other parameter values are specified in Table 3.1.

Synaptic currents were modeled as differences of exponentials according to the equation:

$$F^{\beta}(t) = \frac{H(t)}{\tau_2 - \tau_1} \left(e^{-t/\tau_1} - e^{-t/\tau_2} \right), \tag{3.8}$$

where synaptic timescale parameters are provided in Table 3.1. Here $H(t)$ is a Heaviside function and a pre-synaptic spike occurred at time $t = 0$.

The uncoupled network analyzed in Figure 3.3 consisted exclusively of feedforward projections from Layer 1 to Layer 2. The total synaptic input to Layer 2 neuron i of cell type $\alpha \in \{E, I\}$ was then:

$$I_{i,syn}^{\alpha}(t) = \sum_j J_{ij}^{\text{FF}} F^{\text{FF}} * v_j^{\text{FF}}(t), \tag{3.9}$$

J_{ij}^{FF} is the strength of the synaptic projections from neuron j in the feedforward population to neuron i in population α, F^{FF} is the synaptic filter for projections from neurons in the feedforward population, $*$ denotes convolution, and $v_j^{\text{FF}}(t) = \sum_k \delta(t - t_{jk})$ is the spike train of neuron j in population the feedforward population (t_{jk} is the k^{th} spike time from neuron j).

Connection probabilities $p^{\alpha\text{FF}}$ from neurons in the feedforward layer to neurons in the recurrent population were $p^{E\text{FF}} = 0.2$ and $p^{I\text{FF}} = 0.5$, respectively. Importantly, every neuron in Layer 2 was assigned a preference for the left or right visual hemifield, and Layer 1 neurons projected exclusively to Layer 2 neurons preferring their same visual hemifield. That is, $p^{\alpha_{h^{\text{Rec}}}\text{FF}_h} = p^{\alpha\text{FF}}$ when $h^{\text{Rec}} = h$ for $\{h^{\text{Rec}}, h\} \in \{L, R\}$ and $p^{\alpha_{h^{\text{Rec}}}\text{FF}_h} = 0$ when $h^{\text{Rec}} \neq h$. If a connection from neuron j in the feedforward population to neuron i in population α existed, $J_{ij}^{\alpha\text{FF}} = J^{\alpha\text{FF}}$; otherwise $J_{ij}^{\alpha\text{FF}} = 0$. Synaptic strengths $J_{ij}^{\alpha\text{FF}}$ for a strongly coupled network are provided in Table 3.1. These synaptic strengths are proportional to $1/\sqrt{K}$ where $K = p^{\alpha\text{FF}} N_{\alpha}^{\text{Rec}}$. Linear response theory for the weakly coupled network (Methods 3.3.9.2) assumes synaptic strengths proportional to $1/K$. Synaptic strengths describe the postsynaptic target's membrane potential deflection, neglecting leak, in the non-dimensional units of Eq. (3.7).

In Figures 3.4-3.6, we model Layer 2 of our network with recurrent connections. With the inclusion of these recurrent interactions, the total synaptic input to Layer 2 neuron i in population α is:

$$I_{i,syn}^{\alpha}(t) = \sum_{j} J_{ij}^{\text{FF}} F^{\text{FF}} * v_j^{\text{FF}}(t) + \sum_{k\beta} J_{ik}^{\alpha\beta} F^{\beta} * s_k^{\beta}(t), \tag{3.10}$$

where $\alpha, \beta \in \{E, I\}$, $J_{ik}^{\alpha\beta}$ is the strength of the synaptic projections from recurrent layer neuron k in population β to recurrent layer neuron i in population α, F^{β} is the synaptic filter for projections from neurons in recurrent population β, and $s_k^{\beta}(t)$ is the spike train of neuron k in recurrent population β, consisting of δ functions at each spike emission time.

Connection probabilities $p^{\alpha\beta}$ from neurons in the recurrent population β to neurons in the recurrent population α were $p^{EE} = 0.2$ and $p^{EI} = p^{IE} = p^{II} = 0.5$. If a connection from neuron k in population β to neuron i in population α existed, $J_{ik}^{\alpha\beta} = J^{\alpha\beta}$ in the uniformly connected recurrent network. In networks with clustered architecture, each recurrent layer neuron i, j was assigned membership to an assembly based on hemifield preference h^{Rec}. The ratio R dictated the gain on synaptic strengths of two neurons in the same assembly. See the Results section for further information on tuned assemblies. $J_{ik}^{\alpha\beta} = 0$ when there was no connection from neuron k to neuron i. All synaptic strengths $J_{ik}^{\alpha\beta}$ for a strongly coupled network are provided in Table 3.1. These synaptic strengths are proportional to $1/\sqrt{K}$ where $K = p^{\alpha\beta} N_{\alpha}^{\text{Rec}}$. Linear response theory for the weakly coupled network (Methods 3.3.9.2) assumes synaptic strengths proportional to $1/K$. Synaptic strengths describe the postsynaptic target's membrane potential deflection, neglecting leak, in the non-dimensional units of Eq. (3.7). All spiking network simulations were performed using Euler integration with a 0.1 ms timestep.

Symbol	Description	Value
N^{FF}	Number of feedforward (E) neurons	4,000
N_E^{Rec}	Number of E neurons, recurrent layer	4,000
N_I^{Rec}	Number of I neurons, recurrent layer	1,000
$p^{E\text{FF}}$	Connection probability, feedforward to E	0.2
$p^{I\text{FF}}$	Connection probability, feedforward to I	0.5
p^{EE}	Recurrent connection probability, E to E	0.2
p^{IE}	Recurrent connection probability, E to I	0.5
p^{EI}	Recurrent connection probability, I to E	0.5
p^{II}	Recurrent connection probability, I to I	0.5
$\bar{\lambda}$	baseline firing rate of feedforward neurons	4 Hz
τ_λ	timescale of input fluctuations	60 ms
σ	amplitude of input fluctuations	0-2.4 Hz
μ_{Rec}^E	membrane potential bias, E neurons, recurrent coupling	1.1-1.2
μ_{Rec}^I	membrane potential bias, I neurons, recurrent coupling	1.0-1.05
$\mu_{\text{Uncoupled}}^E$	membrane potential bias, E neurons, uncoupled network	-1.1
$\mu_{\text{Uncoupled}}^I$	membrane potential bias, I neurons, uncoupled network	-1.0
τ_{mem}^E	membrane time constant, E neurons	15 ms
τ_{mem}^I	membrane time constant, E neurons	10 ms
τ_1^E	Rise time for E synapses	1 ms
τ_2^E	Decay time for E synapses	3 ms
τ_1^I	Rise time for I synapses	1 ms
τ_2^I	Decay time for I synapses	2 ms
$J^{E\text{FF}}$	feedforward to E synaptic weight	0.0707
$J^{I\text{FF}}$	feedforward to I synaptic weight	0.0354
J^{EE}	E to E synaptic weight, recurrent layer, unclustered	0.0236
J^{IE}	E to I synaptic weight, recurrent layer, unclustered	0.0141
J^{EI}	I to E synaptic weight, recurrent layer, unclustered	-0.0453
J^{II}	I to I synaptic weight, recurrent layer, unclustered	-0.0566
R	recurrent clustering strength	1-2.5

Table 3.1: Spiking network parameters

3.3.7 Visualizing clustering

Visualizations of each network architecture in Figure 3.4 were performed on subsets of 250 neurons, sampled uniformly across all populations, using network visualization software Gephi [137]. Network nodes were visually distributed according to the strength of synaptic connections between neurons using Gephi's implementation of the Fruchterman-Reingold force-directed algorithm [138].

3.3.8 Variance of model V4 activity

Model V4 neurons in each hemifield were correlated through a common OU process as described by Eq. (3.5). The covariance of V4 activity across both hemifields is then

$$V = \left[\begin{array}{c|c} V_L & 0 \\ \hline 0 & V_R \end{array}\right], \tag{3.11}$$

where the matrix has block structure because spiking activity between the left and right V4 hemifields is uncorrelated.

The covariance of each visual hemifield V_h for $h \in \{L, R\}$ is:

$$\begin{aligned} V_h &= \mathrm{Cov}(v_h^{\mathrm{FF}}, v_h^{\mathrm{FF}}) \\ &= \begin{bmatrix} \bar{\lambda} + \frac{\sigma^2}{2} & & \dfrac{\sigma^2}{2} \\ & \ddots & \\ \dfrac{\sigma^2}{2} & & \bar{\lambda} + \frac{\sigma^2}{2} \end{bmatrix}, \end{aligned} \tag{3.12}$$

which is the variance of the stationary solution to the OU process described by Eq. (3.5).

3.3.9 Linear response approximations of model PFC shared variance

Linear approximations of our network dynamics assume that a recurrent neuron i linearly transforms its synaptic inputs to emit spiking response $y_i(t)$. See Appendix B for a more detailed review of the assumptions underlying this ansatz.

3.3.9.1 Uncoupled Network

When Layer 2 is uncoupled, the firing response of each neuron i has the linear approximation:

$$y_i^\alpha(t) \approx G_i^\alpha \left(\sum_j J_{ij}^{\text{FF}} F^{\text{FF}}(t) * v_j^{\text{FF}}(t) \right), \tag{3.13}$$

where $v_j^{\text{FF}}(t)$ is the firing response of a feedforward neuron that projects to $y_i^\alpha(t)$, $J_{ij}^{\alpha\text{FF}}$ is the strength of synaptic connection from neuron j in population FF to neuron i in population α (see Eq. (3.9)), $F^{\text{FF}}(t)$ is the synaptic filter (see Eq. (3.8)), and G_i^α is the gain of neuron i in population α, which quantifies its sensitivity to its inputs. We remark that Eq. (3.13) is only valid under an expectation, and when the system is observed at sufficient long time windows [120, 139] (See Appendix B). Under the second stated assumption, we need only consider the the effects of the synaptic filter $F^{\text{FF}}(t)$ integrated over all time:

$$\begin{aligned} y_i^\alpha(t) &\approx G_i^\alpha \int_0^\infty F^{\text{FF}}(t) dt \left(\sum_j J_{ij}^{\text{FF}} v_j^{\text{FF}}(t) \right) \\ &\approx G_i^\alpha \left(\sum_j J_{ij}^{\text{FF}} v_j^{\text{FF}}(t) \right) \end{aligned} \tag{3.14}$$

We chose the form of $F^{\text{FF}}(t)$ such that synaptic filter effects integrated to 1 (see Eq. (3.8)).

Using this linear approximation, the covariance of Layer 2 activity due to feedforward inputs is:

$$\begin{aligned} C_y &\approx \text{Cov}(\vec{y}, \vec{y}) \\ &\approx G J^{\text{FF}} V (G J^{\text{FF}})^\top, \end{aligned} \tag{3.15}$$

where $\vec{y}$ is the firing responses of all Layer 2 neurons, V is the shared variance of V4 activity (3.11), J^{FF} is the full feedforward connectivity matrix, and G is a diagonal matrix of all Layer 2 neuron gains. When we apply this linear response theory to our simulated data, we observe our network and compute spike count covariance statistics over a 50 ms time window. Though this breaks the previously stated condition that we observe our system over infinitely long time windows, we make the assumption that 50 ms is a sufficiently long observation period such that our network's trial averaged responses resemble those to static inputs and not to time-varying signals (see Appendix B).

Each entry of matrix C_y represents the covariance of a pair of Layer 2 E neurons, i and k, which inherit correlations from two feedforward mechanisms according to

$$C_{y_{ik}} = (J^{E\mathrm{FF}})^2 \left(Np^2 \left(\bar{\lambda} + \frac{\sigma^2}{2} \right) + N^2 p \left(\frac{\sigma^2}{2} \right) \right), \tag{3.16}$$

where $J^{E\mathrm{FF}}$ is the connection strength of projections from V4 neurons to Layer 2 E neurons, p is the probability of connection for projections from V4 neurons to Layer 2 E neurons, N is the total number of Layer 2 E neurons, and $\bar{\lambda}$ and $\sigma^2/2$ are the mean rate and fluctuation amplitude of the OU process underlying V4 spiking activity, respectively (3.5). The first term on the right hand side of Eq. (3.16) represents common projections that the Layer 2 neuron pair receives from the same V4 neuron. The second term of of Eq. (3.16) represents projections that the Layer 2 neuron pair receives from two (or more) different V4 neurons preferring the same visual hemifield, whose spiking activity is correlated through the underlying OU process (3.5). Recalling that synaptic connectivity strengths J^{FF} scale according to the balanced network condition such that $J^{\mathrm{FF}} \propto 1/\sqrt{N}$ [62], it becomes evident that each of the two feedforward mechanisms of correlation scales as

$$C_{y_{ik}} \propto \underbrace{p^2 \left(\bar{\lambda} + \frac{\sigma^2}{2} \right)}_{\mathcal{O}(1)} + \underbrace{Np \left(\frac{\sigma^2}{2} \right)}_{\mathcal{O}(N)}. \tag{3.17}$$

In the large N limit of neurons, correlations due to common projections (term 1) are negligible as compared to correlations arising from the spike count covariance of V4 activity (term 2). It is thus the shared variance, or off-diagonal terms of V (3.12) that predominantly contribute to

C_y in Equation (3.15). Defining the shared variance of V as

$$V_h^{\text{shared}} = \begin{bmatrix} 0 & & \frac{\sigma^2}{2} \\ & \ddots & \\ \frac{\sigma^2}{2} & & 0 \end{bmatrix}$$

$$V^{\text{shared}} = \left[\begin{array}{c|c} V_L^{\text{shared}} & 0 \\ \hline 0 & V_R^{\text{shared}} \end{array}\right], \tag{3.18}$$

the shared variance of Layer 2 activity is approximately

$$C_y \approx GJ^{\text{FF}} V^{\text{shared}} (GJ^{\text{FF}})^\top, \tag{3.19}$$

in the large N limit.

Notably, the rank of V^{shared} is 2. (Each block V_h^{shared} for $h \in \{L, R\}$ is rank 1.) The rank of Layer 2 shared variance C_y is thus restricted by the rank of V^{shared} through the Frobenius Inequality:

$$\begin{aligned} \text{rank}(C_y) &\leq \min\Big(\text{rank}(G), \text{rank}(J^{\text{FF}}), \text{rank}(V^{\text{shared}})\Big) \\ &\leq \text{rank}(V^{\text{shared}}) \\ &\leq 2. \end{aligned} \tag{3.20}$$

3.3.9.2 Recurrent Network

Linear response theory can be applied to recurrent networks so long as a single neuron's spiking response still scales linearly with sum of its synaptic inputs. This implies that neurons in the recurrent network are weakly coupled, such that it takes many inputs to one neuron to produce spiking activity in that neuron. Under these conditions, the linear approximation of Layer 2 neuron i's firing response is the scaled sum of its inputs from both feedforward projections and recurrent interactions:

$$y_i^\alpha(t) \approx G_i^\alpha \left(\sum_j J_{ij}^{\alpha \text{FF}} F^{\text{FF}}(t) * v_j^{\text{FF}}(t) + \sum_{k\beta} J_{ik}^{\alpha\beta} F^\beta(t) * y_k^\beta(t) \right), \tag{3.21}$$

where $y_k^\beta(t)$ is the firing response of a neuron k from recurrent population β that projects to neuron i from recurrent population α, $J_{ik}^{\alpha\beta}$ is the strength of synaptic connection from neuron k in recurrent population β to neuron i in recurrent population α, $F^\beta(t)$ is the synaptic filter of that connection (see Eq. (3.10)), and all other terms were included in our uncoupled linear approximation in Eq. (3.13). We showed previously that we need only consider the effect of our synaptic filter integrated over all time, and that our synaptic filter integrates to 1 (Eq. (3.14)). Eq. (3.21) therefore reduces to

$$y_i^\alpha(t) \approx G_i^\alpha \left(\sum_j J_{ij}^{\alpha \text{FF}} v_j^{\text{FF}}(t) + \sum_{k\beta} J_{ik}^{\alpha\beta} y_k^\beta(t) \right). \tag{3.22}$$

The shared variance of Layer 2 activity is then the shared variance of our network in the absence of coupling (3.19) filtered through our recurrent interactions such that:

$$\begin{aligned} C_y &\approx \text{Cov}(\vec{y}, \vec{y}) \\ &\approx (\mathbb{I} - GJ^{\text{Rec}})^{-1} GJ^{\text{FF}} V^{\text{shared}} (GJ^{\text{FF}})^\top (\mathbb{I} - (GJ^{\text{Rec}})^\top)^{-1}, \end{aligned} \tag{3.23}$$

where J^{Rec} is the full recurrent connectivity matrix in which each element is $J_{ik}^{\alpha\beta}$ and all other terms are taken from the definition of Layer 2 shared variance in the absence of coupling (3.19).

Notably, the rank of Layer 2 shared variance C_y is still bound by the rank of V^{shared} through the Frobenius Inequality:

$$\begin{aligned} \text{rank}(C_y) &\leq \min\Big(\text{rank}(G), \text{rank}(J^{\text{FF}}), \text{rank}(J^{\text{Rec}}), \text{rank}(V^{\text{shared}})\Big) \\ &\leq \text{rank}(V^{\text{shared}}) \\ &\leq 2. \end{aligned} \tag{3.24}$$

3.3.10 Mathematical analysis of correlated and asynchronous states in recurrent networks

This section aims to provide a theory linking our recurrent network's degree of clustering to shifts between strongly correlated/anti-correlated, multi-stable dynamics and balanced network dynamics. We now outline derivations from Rosenbaum et al. [72] and Baker et al. [140] and explain their novel relevance to our own, clustered network architectures.

3.3.10.1 Cross-spectral density as a measure of co-variability

In the derivations that follow, we will use the cross-spectral density (CSD) to measure co-variability, defined as:

$$\langle U, Z\rangle(f) = \int_{-\infty}^{\infty} C_{UZ}(\tau)e^{-2\pi i f_\tau} d\tau, \tag{3.25}$$

where

$$C_{UZ}(\tau) = \text{Cov}(U(t), Z(t+\tau)) \tag{3.26}$$

is the cross-covariance of U and Z. The CSD will simplify our co-variability calculations because many commonly used co-variability measures can be expressed as functions of the CSD. Note that cross-covariance (3.26) is the inverse Fourier transform of the CSD. Additionally, when we express spike count as the integral of a spike train over interval $[t, \Delta t]$ as in (3.1), we note that spike count covariances over long windows ($\Delta t \to \infty$) can be expressed as the zero-frequency CSD:

$$\lim_{\Delta t \to \infty} \frac{1}{\Delta t} \text{Cov}\left(\int_t^{\Delta t} U(t')dt' \int_t^{\Delta t} Z(t')dt'\right) = \langle U, Z\rangle(f=0). \tag{3.27}$$

The spike count covariance between neurons i and j can then be approximated (for large Δt) as:

$$\text{Cov}\left(N_i(t, t+\Delta t), N_j(t, t+\Delta t)\right) \approx \Delta t \langle y_i(t'), y_j(t')\rangle(f=0) \tag{3.28}$$

3.3.10.2 Balance conditions for strongly coupled recurrent networks with disjoint inputs

We define the population averaged cross-spectral matrix of input currents I_{syn}:

$$\langle I_{\text{syn}}, I_{\text{syn}}\rangle = \begin{bmatrix} \langle I^E_{\text{syn}}, I^E_{\text{syn}}\rangle & \langle I^E_{\text{syn}}, I^I_{\text{syn}}\rangle \\ \langle I^I_{\text{syn}}, I^E_{\text{syn}}\rangle & \langle I^I_{\text{syn}}, I^I_{\text{syn}}\rangle \end{bmatrix}, \tag{3.29}$$

where

$$\langle I^\alpha_{\text{syn}}, I^\beta_{\text{syn}}\rangle = \mathbb{E}_{i,k}\left(\langle I^\alpha_{\text{syn}_i}, I^\beta_{\text{syn}_k}\rangle\right) \tag{3.30}$$

is the expectation over pairwise CSDs between the total input current to recurrent layer neuron i from population α and recurrent layer neuron k from population β where $\{\alpha, \beta\} \in \{E, I\}$. Pairwise CSDs where $i = k$ and $\alpha = \beta$ are excluded from this expectation. We similarly define $\langle \text{FF}, \text{FF}\rangle$ and $\langle \vec{y}, \vec{y}\rangle$ as the 2 x 2 population averaged cross-spectral matrices of feedforward inputs FF and recurrent layer spiking activity $\vec{y}$, respectively. The asynchronous state, in which excitatory activity is balanced by inhibitory activity in the recurrent layer, is defined by the scaling laws:

$$\langle I_{\text{syn}}, I_{\text{syn}}\rangle, \langle \vec{y}, \vec{y}\rangle \propto \mathcal{O}(1/N) \tag{3.31}$$

$$\langle I_{\text{syn}}, \text{FF}\rangle, \langle \vec{y}, \text{FF}\rangle \propto \mathcal{O}(1/\sqrt{N}) \tag{3.32}$$

The population averaged cross-spectral matrix of input currents I_{syn} can then be restated in terms of FF and $\vec{y}$ such that

$$\langle I_{\text{syn}}, I_{\text{syn}}\rangle = \langle \text{FF}, \text{FF}\rangle + \sqrt{N}\left(J_{\text{Rec}}\langle \vec{y}, \text{FF}\rangle + \langle \text{FF}, \vec{y}\rangle J_{\text{Rec}}{}^*\right) + N J_{\text{Rec}}\langle \vec{y}, \vec{y}\rangle J_{\text{Rec}}{}^* + J_{\text{Rec}} A J_{\text{Rec}}{}^* + \mathcal{O}(1/\sqrt{N}), \tag{3.33}$$

where J^{Rec} is the recurrent connectivity matrix first defined in , * denotes a conjugate transpose, and A is defined as

$$A(f) = \begin{bmatrix} A^E(f)/q^E & 0 \\ 0 & A^I(f)/q^I \end{bmatrix}, \tag{3.34}$$

where

$$A^\alpha(f) = \mathbb{E}_k\langle y^\alpha_k(t), y^\alpha_k(t)\rangle \tag{3.35}$$

is the expectation over the power spectral densities of recurrent spiking activity from each neuron k from population $\alpha = \{E, I\}$ and q^{α} is the proportion of neurons that belong to population α. The $\mathcal{O}(1/\sqrt{N})$ term in (3.33) represents the diagonal elements omitted from $\langle \vec{y}, \text{FF} \rangle$.

The asynchronous state condition (3.32) necessitates

$$\sqrt{N} J_{\text{Rec}} \langle \vec{y}, \text{FF} \rangle = \sqrt{N} \langle \text{FF}, \vec{y} \rangle {J_{\text{Rec}}}^{*} = -\langle \text{FF}, \text{FF} \rangle + \mathcal{O}(1/\sqrt{N}). \tag{3.36}$$

Equation (3.33) can then be simplified to

$$\langle I_{\text{syn}}, I_{\text{syn}} \rangle \propto -\langle \text{FF}, \text{FF} \rangle + N J_{\text{Rec}} \langle \vec{y}, \vec{y} \rangle {J_{\text{Rec}}}^{*} + J_{\text{Rec}} A {J_{\text{Rec}}}^{*}. \tag{3.37}$$

We invoke the assumption that each neuron i's conversion of synaptic input I_{syn_i} to spiking activity $y_i(t)$ is $\mathcal{O}(1)$ such that

$$\langle I_{\text{syn}}, I_{\text{syn}} \rangle \propto \langle \vec{y}, \vec{y} \rangle. \tag{3.38}$$

Combining (3.37) and (3.38):

$$\langle \vec{y}, \vec{y} \rangle \propto -\langle \text{FF}, \text{FF} \rangle + N J_{\text{Rec}} \langle \vec{y}, \vec{y} \rangle {J_{\text{Rec}}}^{*} + J_{\text{Rec}} A {J_{\text{Rec}}}^{*}. \tag{3.39}$$

This apparent inconsistency can be resolved using the asynchronous state requirement $\langle \vec{y}, \vec{y} \rangle \propto \mathcal{O}(1/N)$ (3.31). The right hand side of (3.39) then cancels such that:

$$\lim_{N \to \infty} N J_{\text{Rec}} \langle \vec{y}, \vec{y} \rangle {J_{\text{Rec}}}^{*} = \langle \text{FF}, \text{FF} \rangle - J_{\text{Rec}} A {J_{\text{Rec}}}^{*}. \tag{3.40}$$

The asymptotic scaling of spike count correlations in the asynchronous state is then

$$\lim_{N \to \infty} N \langle \vec{y}, \vec{y} \rangle = (J_{\text{Rec}})^{-1} \langle \text{FF}, \text{FF} \rangle ({J_{\text{Rec}}}^{*})^{-1} - A. \tag{3.41}$$

Notably one of two conditions must be satisfied to obey these asynchronous state requirements. For (3.41) to be satisfied, J_{Rec} must be invertible. Alternatively (if J_{Rec} is singular), (3.40) can be satisified if $\langle \text{FF}, \text{FF} \rangle$ is in the column space of $A \mapsto J_{\text{Rec}} A {J_{\text{Rec}}}^{*}$.

In our network, every recurrent layer cell is assigned a preference for the left (L) or right (R) visual hemifield, corresponding to whether that neuron receives feedforward inputs from left or right V4. The full recurrent connectivity matrix is then

$$J_{\text{Rec}} = \begin{bmatrix} J^{E_L E_L} & J^{E_L E_R} & J^{E_L I_L} & J^{E_L I_R} \\ J^{E_R E_L} & J^{E_R E_R} & J^{E_R I_L} & J^{E_R I_R} \\ J^{I_L E_L} & J^{I_L E_R} & J^{I_L I_L} & J^{I_L I_R} \\ J^{I_R E_L} & J^{I_R E_R} & J^{I_R I_L} & J^{I_R I_R} \end{bmatrix} \tag{3.42}$$

In the network with uniform recurrent connectivity, $J^{E_h E_{h'}} = J^{EE}$, $J^{I_h I_{h'}} = J^{II}$, $J^{I_h E_{h'}} = J^{IE}$, and $J^{E_h I_{h'}} = J^{EI}$ for $\{h, h'\} \in \{L, R\}$. Thus, J_{Rec} is singular.

Moreover, the symmetry of our network with uniform recurrent connectivity implies that the average power spectral density is the same for populations E_L and E_R, as well as for populations I_L and I_R. Therefore,

$$A(f) = \begin{bmatrix} A^{E_L}(f)/q^{E_L} & 0 & 0 & 0 \\ 0 & A^{E_R}(f)/q^{E_R} & 0 & 0 \\ 0 & 0 & A^{I_L}(f)/q^{I_L} & 0 \\ 0 & 0 & 0 & A^{I_R}(f)/q^{I_R} \end{bmatrix} \tag{3.43}$$

where A^{α_h} is the average CSD of neurons in population α_h and q^{α_h} is the proportion of neurons in population α_h for $\alpha \in \{E, I\}$ and $h \in \{L, R\}$.

To determine whether $\langle \text{FF}, \text{FF} \rangle$ is in the column space of $A \mapsto J_{\text{Rec}} A J_{\text{Rec}}{}^*$, and subsequently, whether the asynchronous state is possible, we need to more closely examine the structure of $\langle \text{FF}, \text{FF} \rangle$. The average CSD from feedforward inputs $\langle \text{FF}, \text{FF} \rangle$ will be due to both spike count correlations between neurons in the V4 afferent populations and correlations owing to common projections from V4 to PFC. We reference the derivations in Baker et al. [140] and define the CSD due to feedforward inputs between recurrent neuron i and recurrent neuron k as

$$\langle \text{FF}_i^{\alpha}, \text{FF}_k^{\beta} \rangle = \left\langle \sum_j J_{ij}^{\alpha\text{FF}} (F^{\text{FF}} * v_j^{\text{FF}}(t)), \sum_{j'} J_{kj'}^{\beta\text{FF}} (F^{\text{FF}} * v_{j'}^{\text{FF}}(t)) \right\rangle \tag{3.44}$$

where $v_j^{\text{FF}}(t)$ is the spike train of feedforward neuron j, $v_{j'}^{\text{FF}}(t)$ is the spike train of feedforward neuron j', $J_{ij}^{\alpha\text{FF}}$ is the strength of projection from neuron j to recurrent neuron i of cell type

$\alpha \in \{E, I\}$, $J^{\beta \text{FF}}_{kj'}$ is the strength of projection from neuron j' to recurrent neuron k of cell type $\beta \in \{E, I\}$, and F^{FF} is the post-synaptic current waveform. The mean-field CSD due to feedforward inputs is then

$$\langle \text{FF}, \text{FF} \rangle = \underbrace{N J_{\text{FF}} \langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle J_{\text{FF}}^*}_{\mathcal{O}(N)} + \underbrace{q_{\text{FF}}^{-1} J_{\text{FF}} r_{\text{FF}} J_{\text{FF}}^* - q_{\text{FF}}^{-1} J_{\text{FF}} \langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle J_{\text{FF}}^*}_{\mathcal{O}(1)}, \tag{3.45}$$

where $\langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle$ describes spike count correlations between neurons in the feedforward layer, J_{FF} is the full feedforward connectivity matrix, q_{FF} is the proportion of neurons belonging to the feedforward layer, and U^* denotes the conjugate transpose of U. Note that the first term on the right hand side of Eq. (3.45), which scales according to $\mathcal{O}(N)$, represents feedforward correlations inherited through the spike count correlations in V4 activity. The second term of Eq. (3.45), which scales according to $\mathcal{O}(1)$, represents feedforward correlations inherited through common projections from the same V4 neuron to a Layer 2 neuron pair. These two mechanisms of feedforward correlation correspond exactly to those expressed in Eq. (3.16), our time-domain expression of the Layer 2 covariance owing to feedforward input.

For a network with feedforward correlations owing exclusively to shared projections from V4, $\langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle = 0$. This corresponds to $\sigma = 0$ from Figure 3.4. The mean-field CSD due to feedforward inputs is then reduced to

$$\langle \text{FF}, \text{FF} \rangle = q_{\text{FF}}^{-1} J_{\text{FF}} r_{\text{FF}} J_{\text{FF}}^*. \tag{3.46}$$

Notably, $\langle \text{FF}, \text{FF} \rangle$ will inherit all of its structure from connectivity matrix J_{FF}. Because V4 neurons make disjoint projections to PFC neurons with their same visual hemifield preference, J_{FF} has the block structure

$$J_{\text{FF}} = \begin{bmatrix} J^{E_L \text{FF}_L} & 0 \\ J^{I_L \text{FF}_L} & 0 \\ 0 & J^{E_R \text{FF}_R} \\ 0 & J^{I_R \text{FF}_R} \end{bmatrix} \tag{3.47}$$

$\langle \text{FF}, \text{FF} \rangle$ will also have block structure, and cannot be in the column space of $A \mapsto J_{\text{Rec}} A$. By extension, $\langle \text{FF}, \text{FF} \rangle$ cannot be in the column space of $A \mapsto J_{\text{Rec}} A J_{\text{Rec}}^*$. We therefore conclude that our network with disjoint V4 projections and uniform recurrent connectivity cannot achieve the asynchronous state.

In the network with clustered recurrent connectivity reflecting hemifield tuning, our recurrent connectivity matrix J_{Rec} (3.42) changes such that $J^{\alpha_h \beta_h} = RJ^{\alpha_h \beta_{h'}}$ for $\{\alpha, \beta\} \in \{E, I\}$ and $\{h, h'\} \in \{L, R\}$. Clustering therefore restores the asymmetry to J^{Rec} necessary to make it invertible, and the network with hemifield tuned assemblies can achieve the asynchronous state, even in the presence of disjoint inputs from V4 (Figure 3.4b, top left). Note that all our derivations in this section assume the large neuron limit ($N^{\text{Rec}} \to \infty$). According to these derivations, J^{Rec} is either invertible and asynchrony is possible, or J^{Rec} is singular and asynchrony is impossible. In our neural activity simulated from a network with uncorrelated V4 activity ($\sigma = 0$), the smooth transition from correlated recurrent activity to asynchrony that we observe as we increase R results from the finite size of our network simulations.

3.3.10.3 The correlated state

Our V4 neurons with a given visual hemifield preference h receive common fluctuations from the OU process described by Eq. (3.5). When $\sigma > 0$ (Figure 3.4), $\langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle > 0$. In the large N limit of this *correlated state*, only the $\mathcal{O}(N)$ correlations due to V4 spiking co-variability will have significant effect on $\langle \text{FF}, \text{FF} \rangle$ (3.45), which in turn reduces to

$$\langle \text{FF}, \text{FF} \rangle \approx N J_{\text{FF}} \langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle {J_{\text{FF}}}^*. \tag{3.48}$$

Balance is achieved in the correlated state when $\mathcal{O}(N)$ input correlations are reduced to $\mathcal{O}(1)$ spike count correlations in Layer 2 activity such that

$$\langle \text{FF}, \text{FF} \rangle \propto \mathcal{O}(N) \tag{3.49}$$

$$\langle \vec{y}, \vec{y} \rangle \propto \mathcal{O}(1). \tag{3.50}$$

Using Equation (3.48), Equation (3.39) can then be restated as

$$\underbrace{\langle \vec{y}, \vec{y} \rangle}_{\mathcal{O}(1)} \propto N \left(\underbrace{J_{\text{Rec}} \langle \vec{y}, \vec{y} \rangle {J_{\text{Rec}}}^* - J_{\text{FF}} \langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle {J_{\text{FF}}}^*}_{\mathcal{O}(1/N)} \right) + \underbrace{J_{\text{Rec}} A {J_{\text{Rec}}}^*}_{\mathcal{O}(1)}, \tag{3.51}$$

where the parenthetical terms must scale according to $\mathcal{O}(1/N)$ for self-consistency. Solving the parenthetical terms for $\langle \vec{y}, \vec{y} \rangle$, we find that Equation (3.51) is self-consistent if and only if J_{Rec} is invertible:

$$\langle \vec{y}, \vec{y} \rangle = (J_{\text{Rec}})^{-1} J_{\text{FF}} \langle \vec{v}_{\text{FF}}, \vec{v}_{\text{FF}} \rangle J_{\text{FF}}{}^{*} \left(J_{\text{Rec}}{}^{*} \right)^{-1} \tag{3.52}$$

Analogously to the previous section, uniform recurrent connectivity makes J_{Rec} singular, while clustered recurrent connectivity makes J_{Rec} invertible (3.42).

In conclusion, the asynchronous state can be restored in a network with uncorrelated, disjoint inputs by adding hemifield specific recurrent clustering. This makes the spatial scale of the inhibitory recurrent architecture commensurate to the spatial scale of each input, and allows the network to reduce $\langle \text{FF}, \text{FF} \rangle \propto \mathcal{O}(1)$ correlations due to feedforward inputs to $\langle \vec{y}, \vec{y} \rangle \propto \mathcal{O}(1/N)$ correlations in the recurrent layer. By analogous mechanisms, hemifield specific recurrent clustering is able to reduce $\langle \text{FF}, \text{FF} \rangle \propto \mathcal{O}(N)$ feedforward correlations from correlated, disjoint inputs to $\langle \vec{y}, \vec{y} \rangle \propto \mathcal{O}(1)$ correlations in the recurrent layer.

3.3.11 Partitioning model PFC activity into states

Our model networks with strong recurrent coupling and correlated, disjoint inputs exhibit multistable dynamics with alternating states of high firing activity from model PFC neurons preferring the left (State L) or right (State R) visual hemifield. To partition model PFC activity into State L or State R over time, we first projected each hemifield's population activity onto its mean activity and variance of activity over time. Let $Y_h = \{y^E_{1_h}(t), ..., y^E_{N_h}(t)\}$, where $y^E_{i_h}(t)$ is the vector of spike counts over time, binned in non-overlapping windows of $\Delta t = 50$ ms, of excitatory neuron i in model PFC with visual hemifield preference $h \in \{L, R\}$. $\mathcal{Y} = \{\langle Y_L \rangle, \text{Var}(Y_L), \langle Y_R \rangle, \text{Var}(Y_R)\}$ was then used as the feature matrix for State L versus State R classification by a Gaussian Mixture Model (GMM) [141], where angle brackets denote an expectation over PFC model neurons preferring hemifield $h \in \{L, R\}$. Note that a single observation of the feature matrix, which we will denote $\mathcal{Y}_t$, represents one timepoint of the original hemifield population activity. In brief, the GMM uses an Expectation-Maximization algorithm to learn without supervision $p(z_{tS} = 1 | \mathcal{Y}_t)$, or the probability that timepoint $\mathcal{Y}_t$ belongs to the cluster of activity of representing $S \in \{\text{State L}, \text{State R}\}$, where that cluster of activity is modeled as multivariate Gaussian $\mathcal{N}(\mu_S, \Sigma_S)$ in the feature space of $\mathcal{Y}$. Σ_S was constrained to be diagonal.

Each timepoint of neural activity $\mathcal{Y}_t$ in which $p(z_{\text{State L}} = 1|\mathcal{Y}_t) > 0.97$ was assigned to State L. Each timepoint of neural activity $\mathcal{Y}_t$ in which $p(z_{\text{State R}} = 1|\mathcal{Y}_t) > 0.97$ was assigned to State R. Remaining timepoints were considered to represent dynamics in which the neural activity was transitioning between the two states, and these timepoints were excluded from analyses on state partitions. State transitions never exceeded 20% of the total time over which we analyzed simulated network activity.

3.3.12 Factor Analysis of model PFC activity

We performed Factor Analysis (FA) (Appendix A) on the spike count activity of random subsets of 100 excitatory neurons from our PFC model simulations, sampled uniformly across hemifield preferences. Neurons whose firing rates were smaller than 1 Hz were excluded from analysis. Spike trains were binned in non-overlapping intervals of $\Delta t = 50$ ms.

For factor analysis of state partitioned data (Figure 3.8), we began our state partitioning procedure described in Methods 3.3.11 with 220 network simulations per connectivity matrices realization of length 10 s per simulation. Population activity was in a single state $S \in \{\text{State L, State R}\}$ for an average of 4.2 s per simulation. There were 20 non-overlapping sampling of neurons (10 sampling per realization of connectivity matrices, for 2 realizations of connectivity matrices). We applied FA on each sampling of neuron spike counts in state $S \in \{\text{State L, State R}\}$. FA was performed on non-state-partitioned activity by uniformly sampling network activity over both states $S \in \{\text{State L, State R}\}$ and transition timepoints for the same total duration as the average time spent in a single state S.

3.3.13 Linear response fits to simulated PFC activity

We fit our linear response theory from Methods 3.3.9 to our simulated Layer 2 data. We recall that a neuron's gain G is simply the derivative of the neuron's frequency-current (f-I) curve evaluated at its steady-state firing rate. We computed the mean input current $I^{\alpha_h}_{syn_i} = \mathbb{E}_t[I^{\alpha_h}_{syn_i}(t)]$ (3.9) and firing rate $f^{\alpha_h}_i = \mathbb{E}_t[f^{\alpha_h}_i(t)]$ of each neuron i in a population α_h, where $\alpha \in \{E, I\}$ denotes cell type and $h \in \{L, R\}$ denotes visual hemifield preference.

An f-I curve for the mean activity of all neurons in population α_h was then fit according to the piecewise regression model:

$$f_{\alpha_h} = \begin{cases} \beta_0 + \beta_1 I_{syn}^{\alpha_h} + \beta_2 \left(I_{syn}^{\alpha_h}\right)^2, & I_{syn}^{\alpha_h} \leq \mathcal{I} \\ \beta_0 + \beta_1 I_{syn}^{\alpha_h}, & I_{syn}^{\alpha_h} > \mathcal{I} \end{cases} \tag{3.53}$$

Knot location $\mathcal{I}$ was selected by comparing the cross-validated likelihood functions of models of the form (3.53) for varying values of $\mathcal{I}$. Once a mean-field f-I curve was fitted for population α_h, the gain of each neuron in the population was approximated by the derivative

$$\begin{aligned} G_{\alpha_h} &= \frac{df_{\alpha_h}}{dI_{syn}^{\alpha_h}} \\ &= \begin{cases} \beta_1 + 2\beta_2 \left(I_{syn}^{\alpha_h}\right), & I_{syn}^{\alpha_h} \leq \mathcal{I} \\ \beta_1, & I_{syn}^{\alpha_h} > \mathcal{I} \end{cases} \end{aligned} \tag{3.54}$$

evaluated at each neuron's mean firing rate.

These gains could then be used to compute a theoretical approximation of the full, pairwise covariance matrix of the network activity defined as C_y. We define the theoretical estimate of pairwise co-variability in the uncoupled network as $C_y^{\text{Uncoupled}}$, which was computed using Eq. (3.15). In the multi-stable network with strong recurrent coupling, C_y was computed according to Eq. (3.23) separately for State L and State R, where network activity was partitioned into states using the process described in Subsection 3.3.11. Linear response techniques provide accurate estimates of the pairwise co-variability of neural activity, but do not provide accurate estimates of each neuron's private variability [120]. We can re-write our state-partitioned linear response estimates of the network with strong recurrent coupling (3.23) as

$$\begin{aligned} C_y &= (\mathbb{I} - GJ^{\text{Rec}})^{-1} GJ^{\text{FF}} V (GJ^{\text{FF}})^\top (\mathbb{I} - (GJ^{\text{Rec}})^\top)^{-1} \\ &= (\mathbb{I} - GJ^{\text{Rec}})^{-1} C_y^0 (\mathbb{I} - (GJ^{\text{Rec}})^\top)^{-1}, \end{aligned} \tag{3.55}$$

where C_y^0 is the theoretical estimate of the network's pairwise co-variability in the absence of recurrent coupling. While we could replace the entirety of C_y^0 in this computation with the linear response estimate from our uncoupled network simulations $C_y^{\text{Uncoupled}}$, this would break a key linear response assumption that each neuron has one, stationary gain G_i at the fixed point of our linearization. Instead, we make only the substitution

$$\mathrm{diag}\left(C_y^0\right) = \mathrm{diag}\left(C_y^{\mathrm{Uncoupled}}\right) \tag{3.56}$$

to correct for linear response theory's flawed estimates of private neuronal variability.

3.4 Results

3.4.1 Characterizing variability in V4 and PFC

A non-human primate engaged in a memory-guided saccade (MSG) task in which a target was presented at one of forty locations in 2D screen space (Figure 3.1a). The primate had to make a saccade to the remembered location after a delay period. We analyzed simultaneous microelectrode array recordings from visual area V4 and visually-responsive [142] integration area PFC during this distributed visual task. Single neuron responses in PFC during the task's delay period were spatially tuned, demonstrating specificity for both the angular location and radial eccentricity of our dense mapping of target space (Figure 3.1b). We distilled each PFC neuron's 2D spatial response profile to a single preferred location, computed as the center of mass (COM) of the neuron's receptive field (black Xes, Figure 3.1b,d). PFC neurons recorded across all sessions ($N = 784$ total neurons, 19 recording sessions) exhibited preferences for a wide range of spatial eccentricities and radial locations that spanned the entire visual scene (Figure 3.1c-d). This is consistent with previous findings that PFC neurons show spatial tuning both contralateral and ipsilateral to the recorded hemisphere[143, 144, 142], owing to the converging projections that one PFC hemisphere receives from both hemispheres of upstream visual brain areas[145, 146, 147]. By contrast, V4 has retinotopic organization and only encodes visual information contralateral to the recorded hemisphere (Figure 3.1d)[148].

Having mapped V4 and PFC receptive fields, we sought to understand the coordinated population dynamics of both brain areas. We examined the structure of pairwise spike count covariability that was not due to stimulus tuning. These shared fluctuations underlying trial-to-trial variability are commonly referred to as noise correlations (see Methods). Consistent with previous studies demonstrating spatially-dependent correlations [72], noise correlations

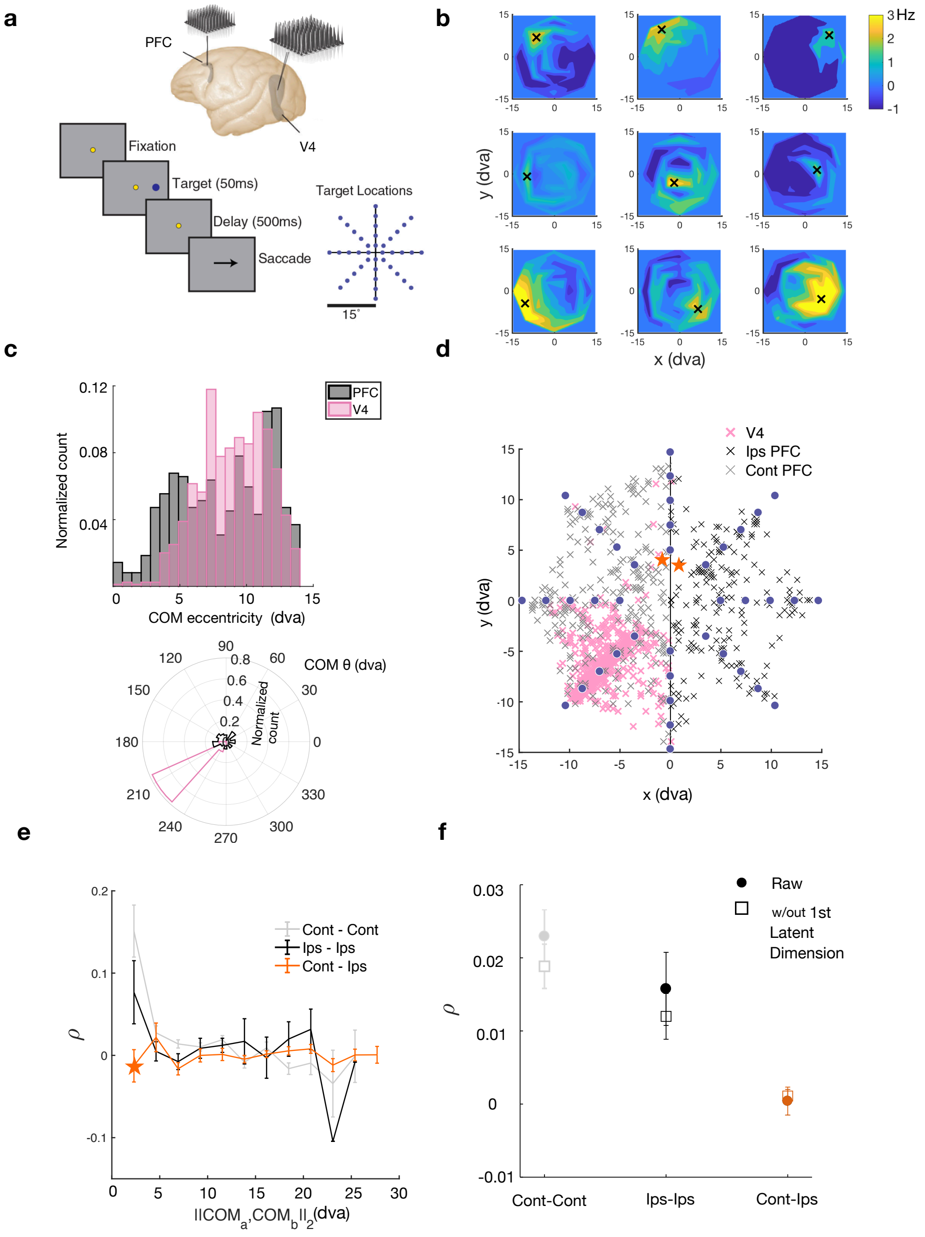

Figure 3.1: **(a)** Illustration of the visual task, in which a non-human primate made a saccade to a remembered target location in 2D space. Neural data were simultaneously recorded in V4 and PFC. **b-d:** Analysis of V4 and PFC tuning. **(b)** 2D spatial receptive fields of 9 example PFC neurons. A neuron's preferred location, calculated as the center of mass (COM) of its receptive field, is shown with a black X. Visual space is represented in units of degrees of visual acuity (dva). **(c)** Distribution of preferred eccentricity (top) and preferred angular location (bottom) of recorded V4 and PFC populations. Data from 747 V4 neurons and 487 PFC neurons shown, pooled across 19 recording sessions. **(d)** Preferred locations of all recorded neurons (see **c**) plotted in the 2D visual space. Targets shown in blue. **e-f:** Analysis of PFC noise correlations. Analyses were conducted on the responses of 487 PFC neurons across 19 recording sessions, 56 ± 3 trials per matched target condition in each session. Error bars are SEM. **(e)** Pairwise spike count correlation ρ as a function of the Euclidean distance between neurons' preferred spatial locations. Spike counts were summed over the 500 ms delay period. Pairs are organized by visual hemifield preference. Neurons preferring opposite visual hemifields are shown in orange. The orange star denotes pairs of neurons preferring similar spatial locations that still span the visual midline (example pair shown in **d**). **(f)** Pairwise spike count correlation by hemifield preference, averaged over all space. Raw correlations denoted by solid dots. Open squares denote residual correlations after subtracting the effects of the FA-identified top latent dimension of shared variance (Methods 3.3.5). Spike counts were binned in 180 ms windows.

between pairs of PFC neurons preferring the same visual hemifield were positive and decreased as a function of the Euclidean distance between the COMs of the neurons' spatial receptive fields (Figure 3.1e). However, pairs of neurons preferring opposite visual hemifields exhibited near-zero correlations, even when the distance between their COMs was very small. We sought to understand whether nearby neurons preferring opposite hemifields were truly uncorrelated or alternatively, exhibited competitive anti-correlations that were masked by shared common fluctuations inherited from outside sources. Using factor analysis, we tested whether we could recover anti-correlation structure between PFC neurons with opposite hemifield preferences after removing the dominate latent dimension of globally shared variability (see Methods 3.3.5). Residual correlations with the dominate latent dimension removed were still near-zero across all recorded pairs of PFC preferring opposite visual hemifields (Figure 3.1f). Thus, we concluded that PFC neurons with opposite hemifield preferences were not exhibiting strongly competitive dynamics.

We sought to further characterize the coordinated fluctuations underlying the noise correlations of our neuronal populations. We again used factor analysis (FA), which partitions the spike count co-variability of neuronal activity into a private variance component, representing the independent, Poisson-like firing variability of individual neurons, and a shared variance component, which represents the coordinated fluctuations of interest (Methods 3.3.5 and Appendix A). FA finds a latent basis set of dimensions that describe the neuronal population's shared variance. To assess the dimensionality of the coordinated fluctuations in both brain areas, we adopted metric d_{shared} from Williamson et al. [59], defined as the number of ordered latent dimensions required to explain 95% of the neuronal population's shared variance. Consistent with prior findings of low dimensional dynamics in V4 [24, 23], the average d_{shared} of V4 activity across recording sessions was one (Figure 3.2a). PFC exhibited much higher dimensional dynamics, with an average d_{shared} of five across recording sessions (Figure 3.2a). Possessing simultaneous recordings of our two brain areas, we were able to do a more direct comparison of the dimensionality expansion between V4 and PFC. Even when assessing the FA results of V4 and PFC activity from matched recording sessions, we found d_{shared} in PFC was typically ≥ 4 despite it consistently inheriting only one-dimensional shared variance from a single hemisphere of V4 (Figure 3.2b). We note once again that our recorded PFC hemisphere would have received information from both the left and right visual cortices. Assuming symmetric transmission of

V4 activity from both cortices, we would expect PFC to have at most two dimensions of shared variability if it directly reflected its V4 inputs. We thus concluded that PFC activity contained additional dimensions of shared variance from those possibly inherited from V4.

We pause now to consider whether it comes as a surprise that PFC has much higher dimensional shared variance than V4. PFC integrates information from multiple senses and is complicit in working memory and other executive functions [130, 131, 132, 133]. These high-level functions are made possible through the myriad of afferent projections that PFC receives from cortical areas other than V4. The most naive hypothesis of our dimensionality results would state that our observed high-dimensional, shared fluctuations in PFC reflect inputs from other brain regions unrecorded in this study. However, the logistics of confirming or refuting this hypothesis would be intractable with contemporary neural recording technologies [7]. Analyzing spike count co-variability with dimensionality reduction techniques like FA requires simultaneous recordings of single unit activity from many neurons. It is currently infeasible to collect single unit recordings of this scale across several brain areas simultaneously in an awake, behaving animal [7]. We instead chose to adopt a parsimonious modeling approach, in which we set out to determine whether PFC could expand the dimensionality of shared variance inherited from V4 through recurrent interactions alone. We note that our parsimonious model does not preclude the existence or influence of inputs to PFC from unobserved brain areas.

If it were true that PFC filtered our V4 input through complex recurrent dynamics rather than inheriting the structure of its activity directly, it would perhaps mean we would see no obvious hallmark of the V4 latent dimension in the shared fluctuations of our PFC activity. We set out to investigate this premise. Knowing that the visual system is lateralized and our recorded V4 activity was likely transmitted preferentially to PFC neurons encoding the same (contralateral) visual hemifield, we investigated whether PFC shared variability was dominated by any latent dimension that loaded differentially onto PFC neurons with opposite visual hemifield preferences (Figure 3.2c). A differentially-loaded latent would be quantified by a difference in the sign or polarity of that latent's loadings onto PFC neurons preferring the contralateral versus ipsilateral visual hemifield. Such a result would be evidence that the contralaterally-tuned PFC population directly inherits and trivially transforms the single latent dimension of shared variance observed in our V4 activity. To the contrary, we found that none of the latent dimensions describing PFC's shared variance cleanly exhibited differential loadings onto PFC neurons with

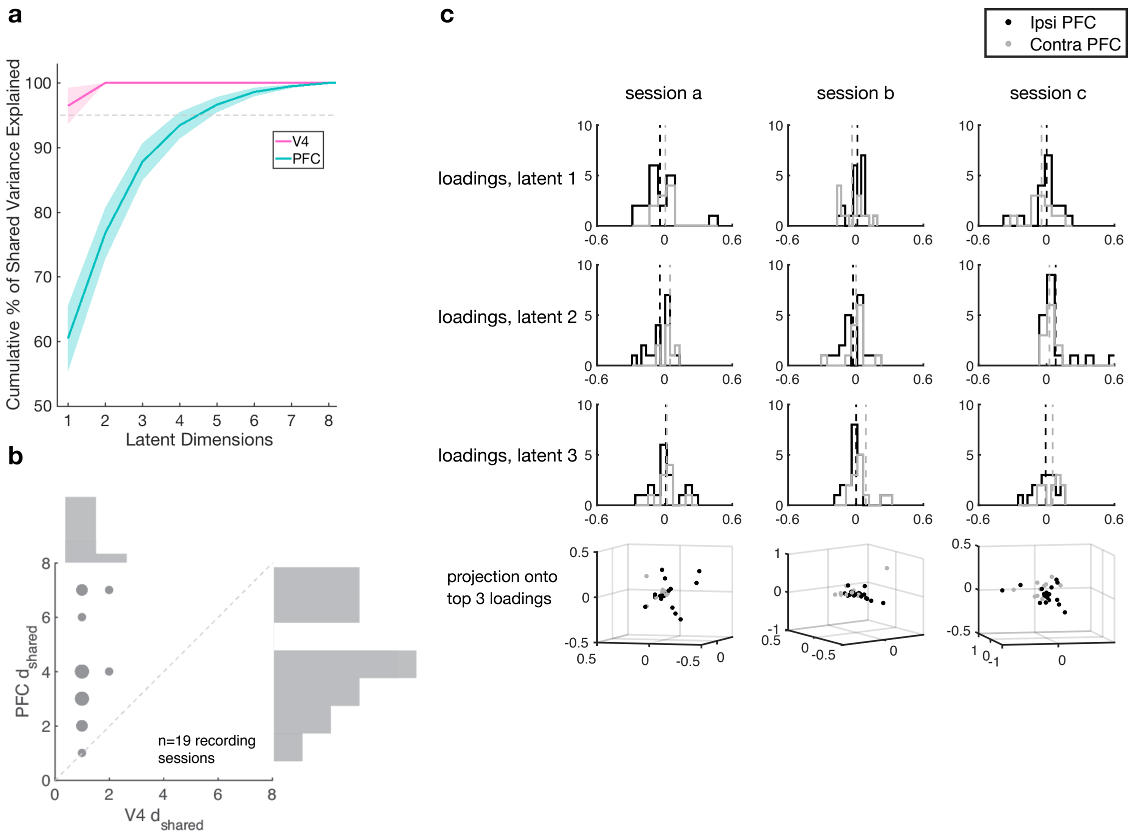

Figure 3.2: Factor analysis (FA) of residual activity (trial-to-trial variability) in V4 and PFC. Analyses performed on each of 19 recording sessions, where there were 39 ± 9 V4 neurons, 25 ± 4 PFC neurons, and 56 ± 3 trials per matched target condition in each session.**(a)** Cumulative percentage of shared variance explained by each ordered FA latent dimension. Results are pooled across sessions. Error bars are SEM. **(b)** Session by session comparison of d_{shared} (number of latent FA dimensions required to explain 95% of population shared variance) for simultaneously recorded V4 and PFC data. Marker size proportional to number of sessions represented at that datapoint (marginal data distributions shown). **(c)** Distribution of FA loadings onto the top 3 latent dimensions (Methods 3.3.5) for ipsilateral-preferring (black) and contralateral-preferring (grey) PFC neurons. 3 representative recording sessions are shown out of the 19 total sessions.

ipsilateral versus contralateral tuning. Figure 3.2c visualizes the distribution of loadings from the top three latent dimensions of shared variance onto ipsilateral and contralateral preferring PFC cells for 3 example recording sessions of the 19 analyzed in total. The polarity of these loadings did not cleanly decompose onto PFC neurons with opposite visual hemifield preferences. Moreover, ipsilateral and contralateral preferring PFC cells were not obviously separable in the multi-dimensional space of the loadings from all three top latents. It appeared that the strong, one-dimensional latent that would have been selectively-inherited by a PFC subpopulation had been transformed non-linearly across PFC's network and could no longer be extracted from the activity of its target subpopulation.

3.4.2 Linear network dynamics cannot expand dimensionality

Our findings in the previous section suggested that PFC expands the dimensionality of shared variance inherited from its inputs through recurrent interactions. The remainder of this chapter will use spiking network models to gain a mechanistic understanding of the recurrent connectivity architectures that are capable of dimensionality expansion.

We begin by modeling the laterally-tuned V4 inputs to PFC. In the previous section, we confirmed findings that the shared variance of activity in a single V4 hemisphere is approximately one-dimensional [24, 23]. We thus modeled our recorded V4 hemisphere as the activity of 2000 excitatory (E) neurons generated from a doubly-stochastic process, in which individual neurons had Poisson spiking statistics but were correlated through a common, one-dimensional fluctuation induced by an Ornstein-Uhlenbeck (OU) process (Methods, Equation 3.5). Previous studies indicate that noise correlations within a V4 hemisphere are positive, while spiking activity across V4 hemispheres is uncorrelated [32, 114]. We captured this effect by simulating two V4 populations, representing the two visual hemifields, each of which consisted of 2000 E cells correlated through two different realizations of the OU process (λ_L and λ_R, respectively, Figure 3.3a). The pairwise spike count covariance of our simulated V4 activity converged to the covariance of our underlying OU processes computed in Equation 3.12 (Figure 3.3a).

To first understand how PFC activity would reflect V4 inputs in the absence of recurrent interactions, we begin with the simplest PFC model architecture consisting of 4000 uncoupled E cells with leaky-integrate-and-fire (LIF) dynamics (Figure 3.3b). Noise correlation analyses of our PFC neural data showed that PFC cells preferring opposite visual hemifields lacked shared fluctuations (Figure 3.1e). This indicates that PFC neurons preferring opposite visual hemifields did not receive the same global fluctuations from a common afferent pool. Accordingly, we chose to model our feedforward connections from V4 to PFC (expressed by connectivity matrix J^{FF}) as disjoint projections reflective of hemifield tuning, where V4 model neurons preferring the left visual hemifield projected exclusively to PFC model neurons preferring the left visual hemifield (Figure 3.3b).

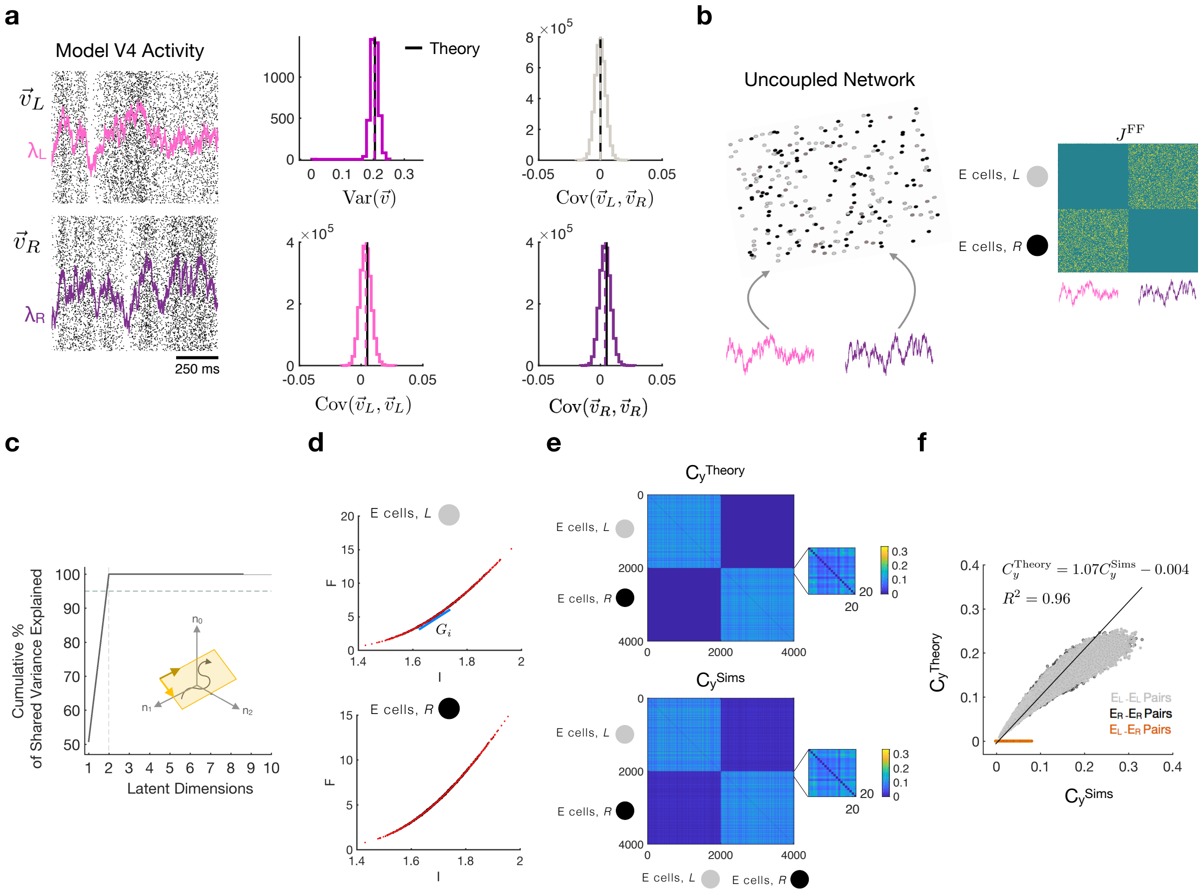

Figure 3.3: Propagation of shared variability through an uncoupled network with linear dynamics. **(a)** Left: Model network input consisting of two V4 hemifield populations of spiking neurons, $\vec{v}_L$ and $\vec{v}_R$. Spiking activity in each hemifield is correlated through a common, 1 dimensional fluctuation (OU process) specific to that hemifield (λ_L or λ_R). Right: Spike count variance and marginal spike count covariance of simulated V4 activity (colored distributions, mean denoted by dotted line) compared to the theoretical estimate of the underlying OU process (black line). Spike counts binned in 50 ms windows. **(b)** Schematic of a 2-layer network model in which the V4 inputs (a) project disjointly (weight matrix J^{FF} shown) to an uncoupled, downstream population of spiking neurons. **(c)** Factor analysis (FA) of the population activity from the output layer of the uncoupled network. Two FA latent dimensions capture 100% of the population shared variance because the network behaves with linear dynamics and directly inherits the variability structure of the inputs. (FA was performed on 1000 s of simulated activity, binned in 50 ms windows, from 10 samples of 100 excitatory neurons per network graph realization, for 2 realizations.) **d-f:** Linear response theory of the uncoupled network (activity from the full output layer population of 4000 E neurons, simulated over 1000 s, for 1 graph realization). Spike counts binned over 50 ms windows. **(d)** F-I curve for excitatory (E) neurons preferring the left (L) or right (R) visual hemifield (simulated data, black; subpopulation fits, red). The gain G_i of each neuron in the subpopulation was fit according to the slope of the F-I curve (blue). **(e)** Full pairwise covariance structure C_y of the network for the simulated data (bottom, Sims) and as predicted by linear response theory (Theory, top). Insets show that theory captures the microstructure of pairwise statistics. **(f)** Pairwise comparison of linear response predicted covariance C_y^{Theory} and covariance of simulated data C_y^{Sims}. Each datapoint is 1 matrix entry from **e** (E_L-E_L pairs, grey; E_R-E_R pairs, black; E_L-E_R pairs, orange).

Neurons in our uncoupled model directly inherit the spiking activity of their V4 inputs, with the exception of minor variability generated through each neuron's LIF dynamics. The spiking activity of each PFC neuron in the uncoupled model, defined here as $y_i(t)$, can thus trivially be approximated through a linear combination of its V4 inputs (Methods 3.3.9):

$$y_i(t) = G_i \left(\sum_j J_{ij}^{\mathrm{FF}} v_j^{\mathrm{FF}}(t) \right), \tag{3.57}$$

where $v_j^{\mathrm{FF}}(t)$ is the spiking activity of a V4 neuron j that projections to PFC neuron i with connection strength J_{ij}^{FF}. G_i is simply the gain by which each PFC neuron i scales its inputs, often denoted as the neuron's *linear response* [139, 120, 149]. When the network receives sufficiently slow perturbations and is observed over sufficiently long time windows such that its trial averaged response is static rather than locked to fast-timescale signal, G_i is well approximated by the slope of a neuron's frequency-current (F-I) curve, or average firing response to fixed input current. Given the linear response approximation in (3.57), the shared variance of PFC activity is:

$$\begin{aligned} C_y &= \mathrm{Cov}(\vec{y}, \vec{y}) \\ &= GJ^{\mathrm{FF}} V^{\text{shared}} (GJ^{\mathrm{FF}})^\top, \end{aligned} \tag{3.58}$$

where $\vec{y}$ contains the firing responses of all model PFC neurons. See Methods (3.19) for details. Importantly, the dimension of model PFC's shared variance C_y will be bounded by the two-dimensional spiking co-variability of V4 activity (V^{shared}) through the linear algebra Frobenius Inequality:

$$\begin{aligned} \mathrm{rank}(C_y) &\leq \min(\mathrm{rank}(G), \mathrm{rank}(J^{\mathrm{FF}}), \mathrm{rank}(V^{\text{shared}})) \\ &\leq \mathrm{rank}(V^{\text{shared}}) \\ &\leq 2. \end{aligned} \tag{3.59}$$

Factor analysis of the activity simulated from the uncoupled model reveals a d_{shared} of 2 (Figure 3.3c), confirming the bounded rank of share variance predicted by our linear response theory (3.59). Together, this exercise reveals that a network with linear dynamics cannot amplify dimensions; a linear network's shared variance will instead always be constrained by the minimum dimensionality of its input co-variability. So long as our PFC model network has linear dynamics and receives two-dimensional fluctuations from V4, its coordinated fluctuations will be confined to a two-dimensional subspace of neuronal activity (Figure 3.3c). We used our linear response theory for shared variance (3.58) to successfully predict the covariance structure, at the level of neuron pairs, of our simulated uncoupled network activity (Figure 3.3e-f). This involved first computing each neuron's gain G_i (3.57), which quantified the neuron's sensitivity to its inputs (Figure 3.3d). Individual neuron gains are well approximated by the slope of their population's frequency-current (f-I) curve. (See Methods Section 3.3.13 for details.)

A network with uniform recurrent coupling is also linearizable when that coupling is sufficiently weak such that a single neuron's spiking response is still linearly related to the sum of its synaptic inputs (in this case, of both the feedforward and recurrent variety). Linear response theory has been commonly applied to such weakly coupled recurrent networks [149, 139, 120]. Derivations of the linear response theory for the weakly coupled recurrent version of our model are contained in Methods Section 3.3.9.2. However, so long as linear response approximations are appropriate, the dimensionality constraint placed on the shared variance of our network by Equation (3.59) will still hold, and our population activity will still be confined to two-dimensional subspace (schematic, Figure 3.3c). This implies that even a PFC model network with weak recurrent coupling cannot generate dimensionality or expand the dimension of shared variance that it inherits from V4.

3.4.3 Metastable dynamics of recurrent networks with multiple, tuned inputs

The previous section demonstrated that linear dynamics are insufficient to explain the dimensionality expansion that we observe between the shared variance of V4 activity and PFC activity. We now move to studying strongly coupled networks, with the intuition that strong recurrent interactions are likely necessary to generate non-linear dynamics.

We being by studying a network model with strong, uniform recurrent connectivity. Strongly coupled networks require inhibition to prevent mass, pathological recurrent excitation and stabilize network dynamics [62, 64]. We therefore introduce 1000 inhibitory (I) cells to our recurrent network and make our network connections respect the relative synaptic strengths required

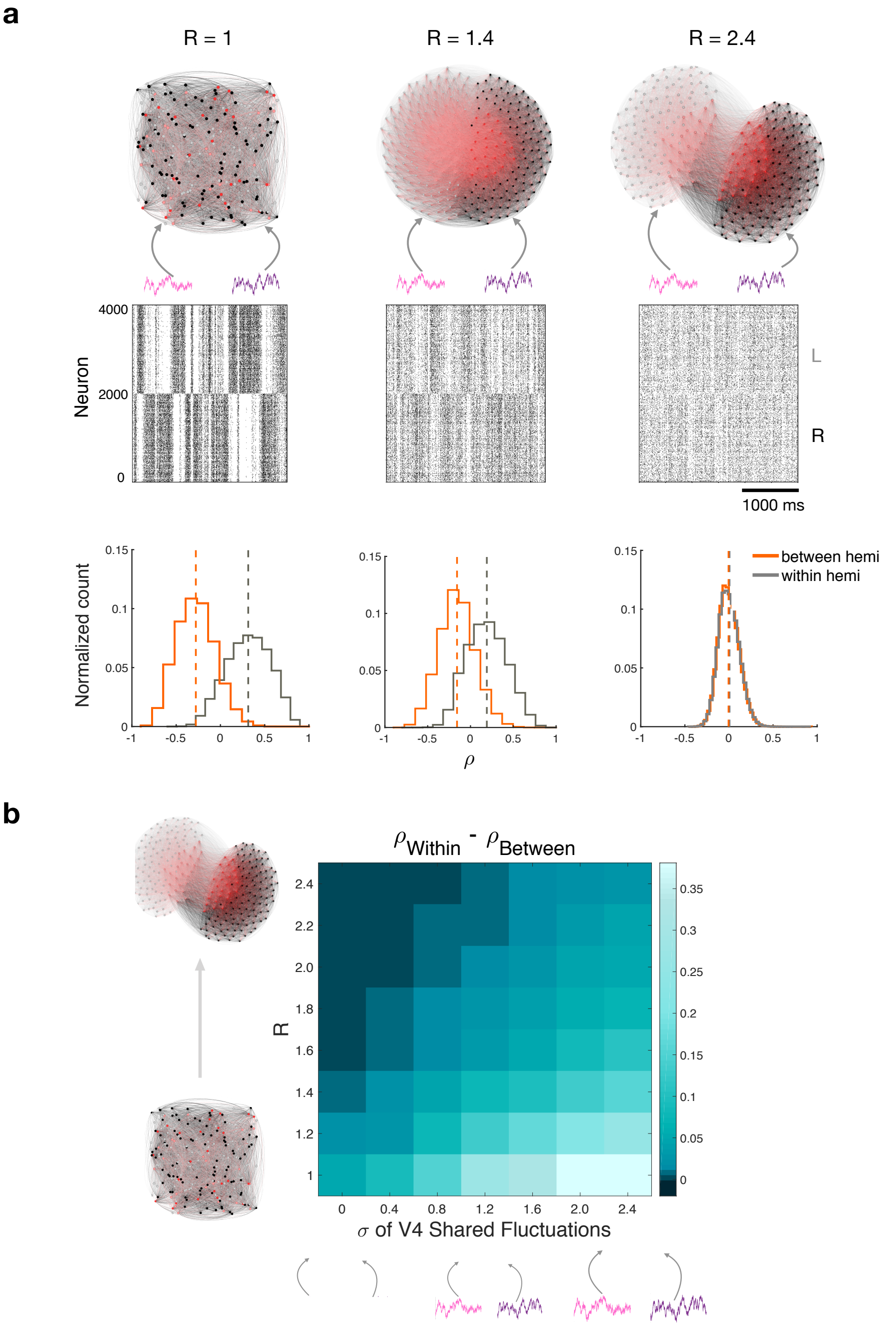

Figure 3.4: Dynamics of networks with assembly structure that inherit shared variability. **(a)** Top: Visualization of connectivity in recurrent networks, with increasing degrees of clustering (R), that inherit shared fluctuations from tuned, disjoint inputs (pink and purple). The $R = 1$ network has uniform, random recurrent connectivity. E cells are shown in grey (left hemifield preference) and black (right hemifield preference). I cells are shown in pink (left hemifield preference) and red (right hemifield preference). Networks of 250 neurons are visualized. Our simulated network contained 4000 E and 1000 I neurons. Middle: Spike rasters showing the spike times of all E neurons. Bottom: Distribution of spike count correlation ρ for pairs of E neurons within a hemifield (grey) or between hemifields (orange). ρ was computed over 150 ms windows. Dotted lines denote distribution means. **(b)** Heatmap of the difference between the mean within hemifield pairwise spike count correlation ρ_{within} and the mean between hemifield pairwise spike count correlations ρ_{between}, as a function of clustering strength R and amplitude of shared fluctuations σ. ρ was computed over 150 ms windows. Statistics for each network architecture were computed over 30 s of simulated data for 3 realizations of the network graph.

for E/I balance [62]. Our two V4 populations reflecting visual hemifield tuning still project disjointly to PFC layer cells preferring their same visual hemifield. Under this model architecture, pairs of neurons in PFC receiving inputs from the same V4 population now exhibit strong spike count correlations, while neuron pairs receiving inputs from opposing V4 populations exhibit strong anti-correlations ($R = 1$, Figure 3.4a). In fact, the anti-correlation mode of population activity with this model architecture is so strong that the network exhibits winner-take-all dynamics to the point of pathology – when PFC neurons tuned to one visual hemifield are active, neurons tuned to the opposite visual hemifield are nearly silent.

Networks in which multiple inputs disjointly project to a layer of neurons with strong, uniform, recurrent coupling cannot avoid this mass anti-correlation mode of activity [72, 117]. The intuition for this known result requires us to consider the relative spatial scales of our feedforward projections and recurrent interactions. Asynchronous dynamics are achieved in balanced networks when correlations due to feedforward inputs and recurrent inputs cancel [64]. A single, broad recurrent architecture cannot, however, dynamically balance multiple, spatially-localized pockets of correlated activity from tuned inputs with disjoint projections. (See Methods Section 3.3.10 for a formal derivation of this claim.) What results is multi-stable dynamics in our PFC network model, with alternating states of highly-correlated activity and silence from PFC neurons tuned for each visual hemifield.

3.4.4 Tuned recurrent assemblies counterbalance tuned inputs

Our pathological anti-correlations in the previous section resulted from a spatial imbalance of feedforward projections and recurrent interactions. To reduce these anti-correlations, we would need recurrent architecture with spatial structure similar to the hemifield-specific projections from our V4 populations [72]. We considered that clustered synaptic connections between neurons with similar functional tunings are commonly observed in cortex [150, 151, 152]. PFC neurons that we recorded *in vivo* preferring the same visual hemifield also showed strong evidence of increased covariability as compared to PFC neurons with opposite hemifield preferences (Figure 3.1e). We thus introduce recurrent assemblies that reinforce visual hemifield tuning. Similar to Litwin-Kumar & Doiron [60], we define a clustering parameter R used to control the degree of increased connectivity between two PFC neurons preferring the same visual hemifield:

$$R^{\alpha\beta} = \frac{J_{\text{in}}^{\alpha\beta}}{J_{\text{out}}^{\alpha\beta}}. \tag{3.60}$$

Here, subscript "in" denotes two PFC neurons in the same assembly, preferring the same visual hemifield. Subscript "out" denotes two PFC neurons in opposite assemblies. $J^{\alpha\beta}$ describes the synaptic strength of connections from neurons of cell type β to neurons of cell type α, where $\{\alpha, \beta\} \in E, I$. Note that, unlike the study by Litwin-Kumar & Doiron [60], this means both E and I connections in our network are clustered. Recent experimental studies support the existence and maintenance of clustered inhibition that is related to functional tuning [153, 154]. Moreover, inhibitory assemblies were shown to moderate firing rates of active excitatory assemblies in Litwin-Kumar & Doiron [60]'s model framework, tempering winner-take-all dynamics. We employ them for a related but different purpose – to provide recurrent connections with spatial scale commensurate to our input projections and help dynamically counterbalance the feedforward correlations arising from our disjoint V4 inputs.

For computational simplicity, we will induce a symmetric clustering constraint in all the work that follows such that $R = R^{EE} = R^{EI} = R^{IE} = R^{II}$. Note that $R = 1$ describes the uniform recurrent connectivity explored in the previous section. Competitive dynamics between PFC neurons preferring the left and right visual hemifield diluted when we introduced tuned assemblies to the recurrent architecture of our network model, corresponding to clustering coefficients of $R > 1$ (Figure 3.4a). This was measurable through shifts in the distributions of spike count correlations, both within and between assemblies. As the clustering coefficient R increased, PFC neurons within assemblies became less correlated, and PFC neurons in opposite assemblies became less anti-correlated. At $R = 2.5$ we see convergence of our two distributions such that there are near-zero mean correlations both within and between hemifields. Critically, the $R = 2.5$ network's ability to counterbalance input correlations arises exclusively from increased communication within assembly; there is no decrease in interactions between neurons in opposite assemblies as compared to recurrent network with uniform connections ($R = 1$), and we are not trivially restoring balance with two, independent networks.

We observed that our network now contained competing co-mechanisms of variability; strongly correlated, lateralized projections introduced competitive spiking dynamics with a strong anti-correlation mode, and strong recurrent assemblies ($R > 1$) appeared to dynamically re-balanced the anti-correlation mode. To better understand the interplay of these two mechanisms, we traversed the 2D model parameter space defined by ranging over the clustering coefficient R and the magnitude of our shared fluctuations inherited from V4 (Figure 3.4b). These V4 fluctuations are defined by the variance σ^2 of the OU process underlying our V4 spiking activity (Equation (3.5)). For each value of R and σ, we measured $\rho_{\text{within}} - \rho_{\text{between}}$, where ρ is the mean spike count correlation across neuron pairs, trials, and graph realizations ($N^E = 4000$ neurons, 30s of data per graph realization, 3 graph realizations) and the subscripts "within" and "between" denote pairwise correlations within the same hemifield assembly and between opposing hemifield assemblies, respectively.

The bottom left of the Figure 3.4b heatmap represents a uniformly connected recurrent network ($R = 1$) inheriting V4 correlations due exclusively to common projections; at $\sigma = 0$ spiking activity from our model V4 neurons is uncorrelated. This exact case is covered by Rosenbaum et al. [72], and correlations within hemifield are $\mathcal{O}(1)$ [140]. The top left of Figure 3.4b represents a network with the same input structure but strong recurrent assemblies. We have already presented the intuition for how assemblies restore the spatial scale of feedforward and recurrent connectivity and dynamically restore balance to the network. Derivations in Methods Section 3.3.10.2 show that the balance condition is dependent on the invertibility of our recurrent connectivity matrix. Assemblies restore the asymmetry to our recurrent connectivity matrix needed to make it invertible. When $\sigma = 0$, in the large N limit of neurons, recurrent assemblies can perfectly balance the feedforward correlations arising from our disjoint inputs and to produce asynchronous network dynamics.

As we move along the σ axis of Figure 3.4b, we increase the amplitude of the shared fluctuations induced by our OU process, and subsequently, the magnitude of spike count correlations in our V4 activity. Baker et al. [140] refer to this as the "correlated state", because the activity of the feedforward neuronal population has $\mathcal{O}(1)$ correlations. These correlations compound with correlations due to our common input projections, which are also $\mathcal{O}(1)$. The total feedforward correlations received by PFC model neurons are in turn $\mathcal{O}(N)$ (Methods, Equation (**??**)). Uniform recurrent connectivity ($R = 1$) cannot dynamically balance these spatially-localized feedforward correlations, producing model PFC spike count correlations also of $\mathcal{O}(N)$ (Bottom right, Figure 3.4b, and Methods 3.3.10.2). In this "correlated" state, strong recurrent assemblies still significantly dilute both within hemifield correlations ρ_{within} and between hemifield anti-correlations ρ_{between} (Top right, Figure 3.4b). In fact, in the large N limit of neurons, hemifield-tuned assemblies of sufficient clustering strength are able to dynamically restore balance (Methods 3.3.10.3). Assemblies cannot, however, restore truly asynchronous dynamics in the "correlated state", as model PFC spike count correlations/anti-correlations are then still at minimum $\mathcal{O}(1)$ (Methods 3.3.10.3).

3.4.5 Assembly networks expand the dimension of inherited shared variability

We established that a network model with correlated, disjoint V4 inputs and strong, uniform ($R = 1$) recurrent connectivity was sufficient to produce non-linear dynamics in the form of multi-stability. Identifying a network architecture that gave rise to non-linear recurrent interactions was our original goal, as non-linear dynamics are required for any system to intrinsically generate dimensionality; we specifically aimed to replicate a dimensionality expansion of shared variability observed between V4 and PFC. Though our $R = 1$ network had non-linear recurrent interactions, it also had pathological levels of anti-correlations that were not representative of the PFC neural activity we observed *in vivo*. We solved this problem in Results 3.4.4 by introducing hemifield-tuned recurrent assemblies, which could successfully dilute the pathological anti-correlations of the $R = 1$ model. We showed that in the large N limit of neurons and as $\sigma \to 0$, strongly clustered assemblies can even restore the balanced, asynchronous state of network activity.

Armed now with a model architecture exhibiting both non-linear recurrent dynamics and a biologically-plausible range of correlation outputs, we sought to test whether this architecture could indeed expand the two-dimensional shared variability of our model V4 activity. We performed dimensionality analyses on activity simulated from a network with a relatively strong clustering coefficient ($R = 2.3$) and relatively weak V4 input correlation parameter ($\sigma = 0.71$) (Figure 3.5). Working in this approximate parameter regime produced recurrent layer spiking activity that looked nearly asynchronous but contained subtle correlations, qualitatively similar to our PFC neural data recorded *in vivo*. Factor analysis (FA) confirmed that our chosen model architecture could successfully expand the two-dimensional shared variance inherited from V4

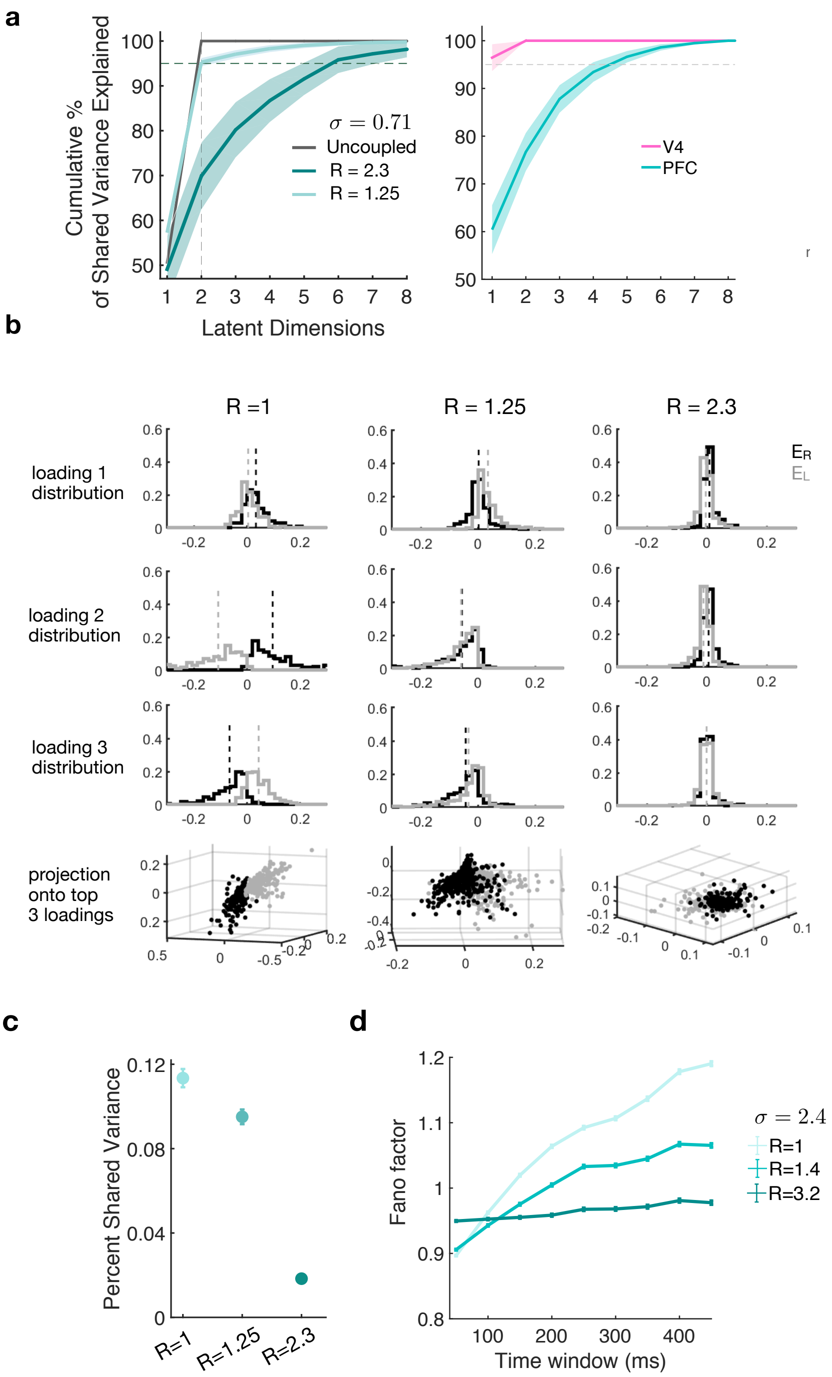

Figure 3.5: The dimension of shared variability in networks with assembly structure. **a-c** Factor analysis (FA) of population activity for networks with various strengths of recurrent clustering R, inheriting common fluctuations of amplitude $\sigma = 0.71$. Analyses of each network architecture were computed over 1000 s of simulated activity from 10 samples of 100 neurons each per network graph realization, for 2 realizations. Spike counts were binned in 50 ms windows. Error bars are SEM. **(a)** Cumulative percentage of shared variance explained by each FA-identified, ordered latent dimension in simulated data (left) and PFC neural data (right, see Fig. 3.2). Error bars are SEM across FA samples. **(b)** Distribution of FA loadings onto the top 3 latent dimensions (Methods 3.3.12) for model neurons of left (grey) or right (black) hemifield preference. **(c)** Percentage of each neuron's total variance that is shared amongst the population, averaged across all analyzed neurons. **(d)** Fano factor of simulated neural activity as a function of the time window over which spike counts are binned, for networks of varying R, $\sigma = 2.4$. Analyses of each network architecture were computed over 300 s of simulated activity for 2 realizations of the network graph.

(uncoupled network, control) to $d_{\text{shared}} = 6$ dimensions of shared variability in model PFC (Figure 3.5a). These results qualitatively replicated the shared variance that we observed from PFC activity recorded *in vivo*. Networks with smaller clustering parameters ($R = 1.25$ shown, $\sigma = 0.71$, Figure 3.5a) still showed slightly expanded shared variance from the uncoupled control. (Indeed, even our $R = 1$ network has multi-stable dynamics, albeit with with a highly anti-correlated population mode that swamps factor analyses.) Networks with weak clustering were not able to reproduce the magnitude of dimensionality expansion seen in the PFC data recorded in *in vivo*.

We demonstrated in Figure 3.4 that stronger recurrent clustering produced weaker anti-correlations between model PFC hemifield populations. Analogously, FA reveals that activity from the two PFC hemifield populations is more separable in the low-dimensional latent space for smaller network clustering coefficients. We analyzed the distribution of neuronal loadings onto the top three dimensions of the FA-identified latent space (Appendix A), where each PFC model neuron was categorized by its preference for the left or right visual hemifield (Figure 3.5b). In the $R = 1$ network, PFC neurons with opposite hemifield preferences loaded onto each of the top three latent dimensions with opposite polarity and were linearly separable in a three-dimensional latent space. Weak degrees of recurrent clustering made the hemifield populations load onto the top latent dimension with opposite polarities, but this hemifield separability was not apparent in the second or third latent dimension ($R = 1.25$, Figure 3.5b). In the $R = 2.3$ network chosen to model PFC neural data recorded *in vivo*, model neurons loaded onto the top three latent dimensions nearly indiscriminately, regardless of their hemifield preference. This result reproduces our FA findings from the PFC neural data recorded *in vivo*, in which neurons of opposite hemifield preference were similarly inseparable in the latent space (Figure 3.2c).

The dilution of correlations/anti-correlations with increasing degrees of recurrent clustering is also evident through the proportion of total variability that is shared across the population in networks with different clustering coefficients. We measured the percentage of each neuron's total variability that was explained by the latent space of shared variability. We refer to this measure as a neuron's "percent shared variance" [59]. The disjoint proportion of variance that is not shared across the population constitutes FA's estimate of a neuron's private "noise", or independent trial-to-trial variability. The $R = 1$ network had the largest percent shared variance, indicative of the strong, anti-correlating common fluctuations to which the network activity was entrained (Figure 3.5c). Neurons' percent shared variance decreased as function of recurrent clustering strength, meaning neural activity was less entrained to latent common fluctuations for stronger degrees of clustering.

Finally, we examined the neural population's average Fano factor as a function of recurrent clustering strength, where Fano factor is defined as the ratio between each neuron's trial-to-trial variance and mean spike count over a fixed time window (3.3). Neurons exhibit Poisson-like trial-to-trial variability, and Poisson processes of stationary rate have a Fano factor of 1. Fano factors of greater than 1 can therefore indicate fluctuations in a neuron's underlying firing rate. We measured Fano factor as a function of the window duration over which spike counts were binned to evaluate long-timescale firing rate variability in our model neurons (Figure 3.5d). Fano factors of all evaluated networks were sub-Poisson for time bins less than or equal to 100 ms. These small time bins primarily captured population activity within a single network state, in which the dominant mode of variability is within-hemifield spike count correlations. Unsurprisingly, for very small time bins ($\Delta t \leq$ 50 ms), neurons in the $R = 1$ network have lower Fano factors than neurons in clustered networks; within state, the $R = 1$ network activity is most correlated and most entrained to common latent fluctuations. For larger time bins ($\Delta t \geq$ 100 ms), however, neurons in clustered networks exhibit smaller Fano factors than neurons in the uniform network. This discrepancy magnifies as the size of the time bin increases. Large time bins capture fluctuations in firing rate across the two meta-stable states of network activity. Clustered networks exhibit weaker anti-correlations across hemifield populations in a single network state (Figure 3.4). We make the related observation that the spiking activity of a single neuron in a clustered network will then vary less between State R and State L. This observation manifests as lower Fano factors in clustered networks even when the networks are evaluated at timescales that capture state transitions. All Figure 3.5d analyses are performed on networks with larger inherited variability ($\sigma = 2.4$), as larger input correlations will drive firing rate fluctuations over state transitions. The qualitative trend reported in Figure 3.5d is robust across choice σ, but small input noise correlations never give rise to super-Poisson variability.

3.4.6 Metastability as time-sharing between states of low-dimensional, linear dynamics

Our networks with strong, recurrent coupling switch between Left and Right metastable states of activity. While this switching behavior clearly does not constitute linear dynamics, we asked whether it was possible to understand the single state dynamics of our network using the linear response frameworks from Results 3.4.2. We hypothesized that each of our two states of network activity exhibited approximately linear dynamics. Since linear systems are incapable of generating dimensionality, this hypothesis would imply that conditioned on each attractor state of activity, our network inherited two dimensions of shared variance from upstream V4 populations in each state of activity. We could then explain PFC's ability to expand dimensionality through its "time-sharing" between two states of network activity, each of which comprised two-dimensional, linear dynamics (Figure 3.6a). A perfect concatenation of these two linear states of activity (in which hops between states occur instantaneously) would be capable of producing up to four dimensions of shared variance. We proposed that the extra dimensions of shared variance observed in our PFC neural data and $R = 2.3$ model network ($d^{\text{PFC}}_{\text{shared}} = 5$, Figure 3.2a and $d^{R=2.3}_{\text{shared}} = 6$, Figure 3.5a, respectively) could arise through non-instantaneous state transitions, during which time the network activity would not behave according to the dynamics of either state.

To begin testing this hypothesis, we collected our model network's recurrent layer activity in 50 ms time bins. We used a Gaussian Mixture Model [141] to assign each time bin a probability of membership to the Left or Right attractor state of network dynamics. (See Methods 3.3.11 for details and Supplemental C for visualization.) Time bins with less than 0.97 probability of membership to either state were assumed to constitute a state transition and were disregarded from all subsequent analysis. We note from Figures 3.4a and 3.5d that the Left and Right activity states are less differentiable in networks with strong clustering. We therefore conducted all linear response analyses that follow on a model network with clustering coefficient $R = 1.25$. In this parameter regime, over 90% of the total analyzed population activity was assigned to the Left or Right state.

Using our state-partitioned population activity, we derived a linear response approximation (Methods 3.3.9) of the full population spike count covariance structure in either the Left and

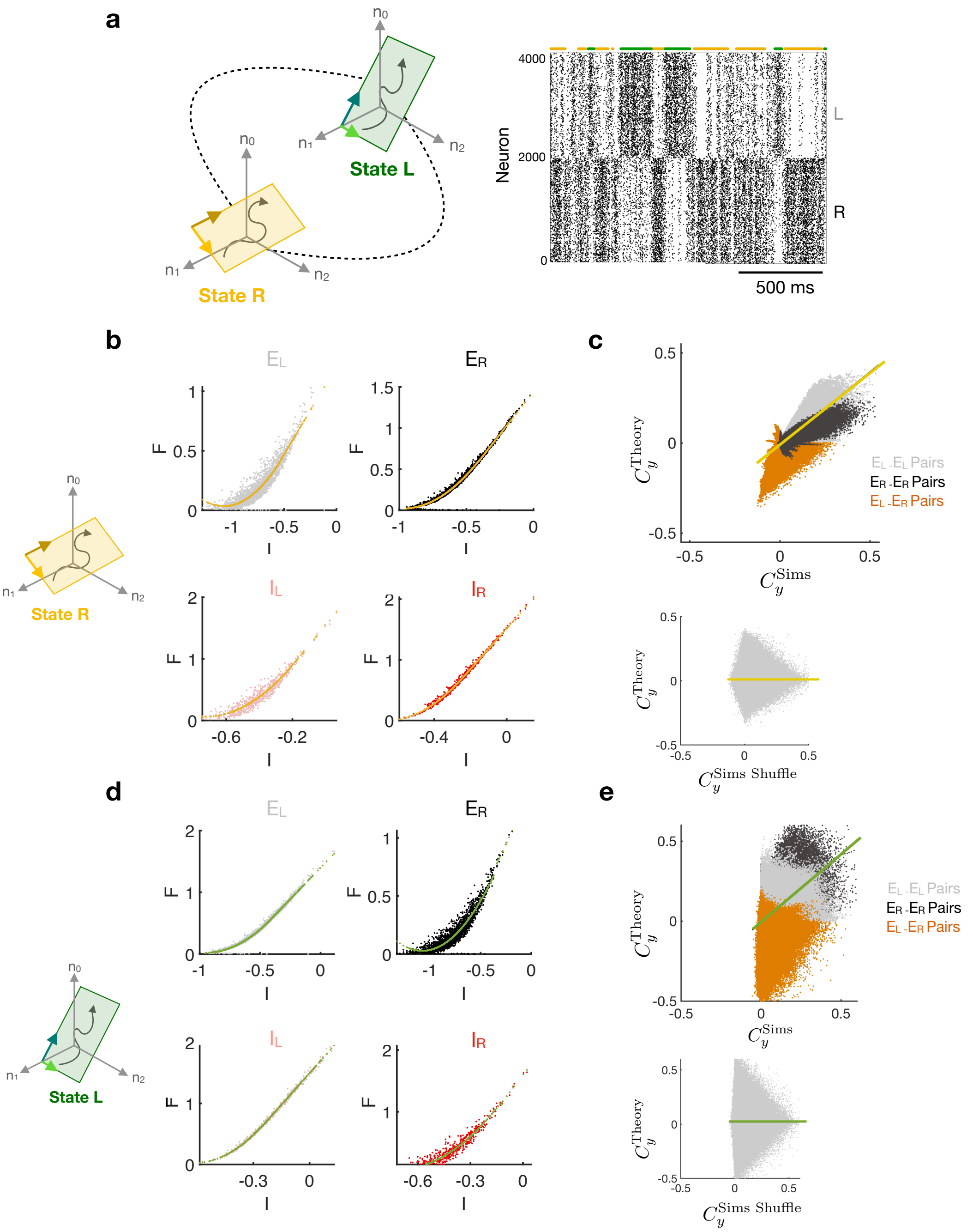

Figure 3.6: Attractor states of linear dynamics in networks with strong recurrent coupling and disjoint inputs. **(a)** Schematic of network activity “time-sharing” between two, low-dimensional attractor states, each of which has linear dynamics. State L (green) corresponds to high activity periods from neurons preferring the left visual hemifield. State R (yellow) corresponds to high activity periods from neurons preferring the right visual hemifield. **b-e:** Linear response theory applied to each attractor state of a network with $R = 1.25$ recurrent clustering strength that inherits $\sigma = 0.71$ amplitude shared fluctuations from disjoint inputs. Spike counts were binned in 50 ms windows. See Methods 3.3.11 and 3.3.13 for further details. **(b,d)** F-I curves fitted to each cell type and hemifield-tuned neural subpopulation. Simulated data shown as points with colors corresponding to subpopulation membership (network visualizations, Fig. 3.4). Fits shown in yellow and green for State R and State L, respectively. **(c,e)** Pairwise comparison of linear response predicted covariance C_y^{Theory} and covariance of state-partitioned simulated data C_y^{Sims} (E_L-E_L pairs, grey; E_R-E_R pairs, black; E_L-E_R pairs, orange). Results are compared against a control in which simulated data is shuffled. Lines represent linear regression summaries of the pairwise relationship between Theory and Sims across all pairs.

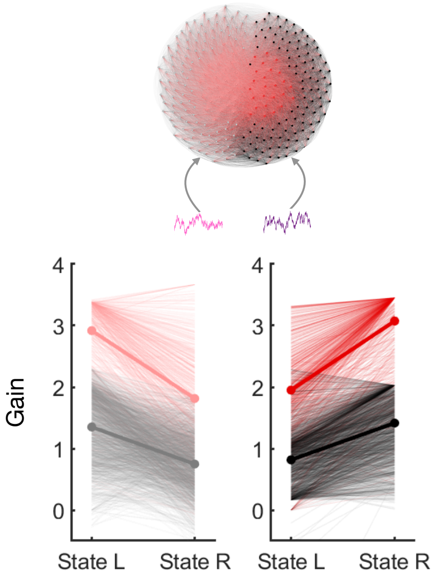

Figure 3.7: Neuronal gain shifts between State L and State R for E_L (grey), I_L (pink), E_R (black), and I_R (red) neurons, as predicted by the linear response theory in Fig. 3.6. Network schematic is shown above for subpopulation clarity.

Right state of activity. For each state, we computed every neuron's gain G_i in that state by differentiating the F-I curves fitted to each population α_h (Figure 3.6b and d for State R and State L fits, respectively). Here $\alpha \in \{E, I\}$ denotes cell type and $h \in \{L, R\}$ denotes visual hemifield preference. (See Methods 3.3.13 for details.)

We used these gains to compute a theoretical approximation of C_y, the state-conditioned shared variance of all neuron pairs, according to the linear response equations in Methods 3.3.9. In statistical mechanics literature, this is referred to as a linear response theory of the microcanonical ensemble. Theoretical approximations of shared variance in State R were predictive of the spike count covariance of the simulated data in State R at the level of individual neuron pairs (Figure 3.6c, $C_{y_{ij}}^{\text{Theory}} = 0.81 C_{y_{ij}}^{\text{Sims}} - 0.011$, $R^2 = 0.63$; excitatory neuron pairs only). We thus concluded that population activity in our network's State R behaved according to approximately linear underlying dynamics. Linear response predictions of shared variance were more heteroscedastic in State L, in which the theory tended to overestimate anticorrelations between neurons preferring opposite visual hemifields (Figure 3.6e). However, theory estimates of shared variance were still significantly more related to the shared variance of simulated activity than to the control, in which simulated data were shuffled (Figure 3.6e, $C_{y_{ij}}^{\text{Theory}} = 0.86 C_{y_{ij}}^{\text{Sims}} - 0.010$, $R^2 = 0.23$; $C_{y_{ij}}^{\text{Theory}} = -2.3 \times 10^{-5} C_{y_{ij}}^{\text{Shuffle}} + 0.012$; excitatory neuron pairs only). Linear response theory assumes that shared variance is shaped by the matrix of neuronal gains G. We therefore examined neuronal gains as a function of state to better understand the shift in shared variance structure between State R and State L (Figure 3.7). We remind the reader that all neurons participate in both states, i.e., model neurons preferring the left visual hemifield still participate in State R – they are simply the less active population. Neurons exhibited organized and significant gain shifts between State R and State L. Both model E and I neurons preferring the left visual hemifield had larger gains in State L than State R; analogously, model E and I neurons preferring the right visual hemifield had larger gains in State R than State L. Symmetries between E and I populations preferring the same visual hemifield result from the network activity's existence in a roughly balanced regime.

Our linear response analyses supported the hypothesis that each state of our metastable network activity obeyed roughly linear dynamics, with interpretable shifts in shared variance structure occurring between State R and State L. We concluded that if this interpretation of our network dynamics was indeed true, FA should uncover at most 2 dimensions of shared variance in state-partitioned population activity. We returned to the $R = 2.3$ model network used to capture PFC activity recorded *in vivo*, in which we saw significant dimensionality expansion ($d_{\text{shared}} > 4$) in the population activity across all states (Figure 3.5a). As predicted by our hypothesis, FA revealed $d_{\text{shared}} = 2$ latent dimensions of shared variance in State R of the $R = 2.3$ network activity. This constituted a highly significantly reduction from the $d_{\text{shared}} = 6$ latent dimensions of the $R = 2.3$ network activity across all states.

Factor analysis of State L activity in the same network determined that $d_{\text{shared}} = 4$ latent dimensions were required to explain 95% of shared variance. While this result is greater than the 2-dimensional shared variance that would be predicted in a system with truly linear dynamics, it still represents a statistically significant reduction from the $d_{\text{shared}} = 6$ latent dimensionality of the network activity over all time. We note that several sources of variability likely influence our imperfect State L Results. First, in our highly clustered network, the two states of network activity do not have drastically different population spiking statistics (Figure 3.4). Our Gaussian Mixture Model (GMM)'s unsupervised partitioning of activity into states can therefore produce variable results. Lacking a ground truth on network state, we cannot directly quantify the error in the GMM state partition, and all subsequent analyses are dependent upon this partition. This means it is possible that our State L partition of network activity contains instances of transitory dynamics. Second, our factor analyses of state-partitioned activity were performed on a single

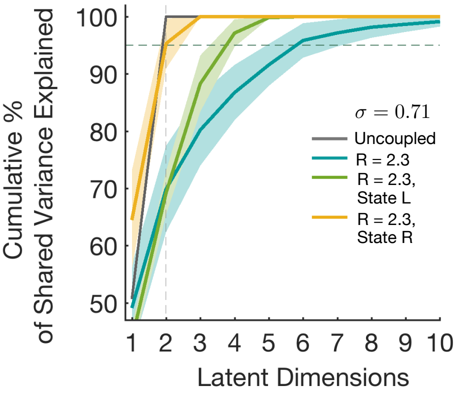

Figure 3.8: Factor Analysis of state-partitioned population activity in the biomimetic model network with $R = 2.3$ clustering strength and $\sigma = 0.71$ amplitude shared fluctuations from disjoint inputs (see Fig. 3.5a) Simulated activity was partitioned by a Gaussian Mixture Model (Methods 3.3.11). Spike counts were binned in 50 ms time windows. Analyses were performed over $\sim$ 1000 s of simulated activity from 10 samples of 100 neurons. Error bars are SEM. See Methods 3.3.12 for further details. The dimension of shared variance for state-partitioned activity (yellow and green for State R and L, respectively) is similar to that of the control network without recurrent coupling (grey), which has known linearizable dynamics (Fig. 3.3d-f).

instantiation of the network graph. Because our model network does not contain unlimited neurons, the micro-connectivity structure of our network in each graph instantiation can influence network dynamics. Third, we observe our network over $\Delta t = 50$ ms time windows, despite the fact that our theory is for time windows of infinite length (Appendix B). Finally, we note that linearization of each state of a network with strong, recurrent coupling and input correlations would be merely an approximation for even a theoretical network of infinite neurons and labeled states of network dynamics that was observed over infinitely long time windows.

3.5 Discussion

We have shown that multi-stable attractor dynamics arise in balanced networks receiving structured, feedforward inputs from upstream neural populations with non-overlapping tuning preferences (Figure 3.4). In networks with uniform, random recurrent coupling, this multi-stability is characterized by "winner-take-all" dynamics, in which neurons receiving different feedforward inputs have strongly anti-correlated activity (Figure 3.4). Recurrent assembly structures reflecting the input tuning help to dilute these pathological, winner-take-all dynamics, giving rise to population activity with smaller, biomimetic degrees of correlation (Figure 3.4). But even with de-correlating assembly architecture, recurrent network dynamics remain subtly metastable. Attractor competition between states of weakly correlated activity produces high-dimensional shared variability across the recurrent population (Figure 3.5). The result is a two-layer network in which low-dimensional shared variance inherited from multiple, tuned inputs is expanded through recurrent interactions. Using a model network with this connectivity architecture, we successfully reproduced *in vivo* neural data from the primate visual system in which V4's dimension of shared variance was smaller than that of downstream visual area PFC (Figure 3.2a and 3.5a). Finally, we showed that a single attractor state of recurrent activity reflected the low-dimensional structure inherited from our V4 inputs (Figure 3.6-3.8). We thus introduced a new framework in which high-dimensional cortical variability can be understood as“time-sharing” between low-dimensional, tuning-specific circuit dynamics.

3.5.1 The structure of shared variance across subpopulations

Analyses of our *in vivo* data revealed that PFC neurons preferring the same visual hemifield had positively correlated spiking activity, while PFC neurons preferring opposite visual hemifields had uncorrelated spiking activity (Figure 3.1e-f). This result constitutes an asymmetry in the shared variance of tuning-specific subpopulations. Our network model does not currently capture this feature of the analyzed neural data. For computational simplicity, our model placed a symmetric constraint on the clustering coefficient R such that in-cluster synaptic strength was scaled uniformly for E-to-E, E-to-I, I-to-E, and I-to-I recurrent connections. As a result, increasing R symmetrically reduced correlations within hemifield and anti-correlations between hemifields (Figure 3.4a).

We showed in Methods 3.3.10 that the balance condition relies on the invertibility of our recurrent weight matrix. We note that previously studied recurrent architectures of excitatory-only clusters [60] will not restore invertibility to the recurrent weight matrix and will not dynamically balance the correlations inherited from our disjoint inputs. Though some degree of inhibitory cell type clustering is required to prevent our recurrent weight matrix from being singular, the invertibility condition does not require exact weight symmetry between all clustered subpopulations. We therefore postulate that tuning-specific shared variance structure could be achieved by exploring the full space of recurrent connectivity architectures still satisfying the weight matrix invertibility condition. Characterizing this full connectivity space and relating it to network dynamics is an important topic for future study.

3.5.2 Long-timescale variability and co-variability through inheritance

Cortical neurons show firing rate fluctuations over long timescales [155, 156]. Fano factor is used as a measure of these firing rate fluctuations; since neurons are known to exhibit Poisson-like private spiking variability, Fano factors greater than 1 are thought to represent variability in a neuron's underlying firing rate. In this case, neural activity behaves according to a "doubly stochastic" process [157], in which spike count variability and slow timescale firing rate dynamics are separable. Litwin-Kumar & Doiron [60] reproduced long timescale rate fluctuations by introducing assembly structure to excitatory subpopulations of recurrent layer neurons. In this framework, competing pockets of excitatory activity in the recurrent layer internally gave rise to attractor dynamics. Rate fluctuations, as measured by large Fano factor values, reflect competition between attractor states. Therefore, in Litwin-Kumar & Doiron [60]'s framework, increasing the cluster density lengthened the timescale of rate fluctuations and increased Fano factors.

Our model introduces a completely opposing potential mechanism for firing rate variability, in which tuned, disjoint inputs give rise to attractor dynamics. In our framework, a network with balanced, uniform recurrent coupling can have metastable activity by inheriting structured input correlations from upstream brain areas. Our uniform, balanced recurrent network ($R = 1$) thus produces Fano factors greater than 1 at timescales capturing transitions between attractor states. Moreover, recurrent assemblies in our framework constitute neighborhoods of increased connectivity between all cell types. Our clustering parameter R strengthens E-to-E, I-to-I, E-to-I, and I-to$E$ connections. As such, our clusters are mechanisms to make the spatial scale of recurrent connections match the spatial scale of our disjoint inputs. In our framework, increasing the clustering strength R dilutes the metastable recurrent dynamics that arise through the input structure. Increased degrees of clustering thus result in reduced rate fluctuations and *smaller* Fano factors (Figure 3.5d).

Our framework also demonstrates an inverse relationship between Fano factor and the dimension of shared variability, as stronger clustering is associated with both smaller Fano factors and greater dimensionality expansion (Figure 3.5). Previous network models of internally-generated co-variaibility have been unable to decouple a direct relationship between the Fano factor magnitude and the rank of population-wide variability [24]. They have thus been unable to explain neural datasets with both low-rank shared variability and long timescale rate fluctuations [32]. Our framework might present a key to understanding such datasets.

These are two of the many ways in which our study highlights major differences in the dynamics of networks with internally generated versus inherited variability. We showed that structured inputs can significantly alter the dynamics of known recurrent architectures previously studied in isolation. We believe that studying the interplay between the mechanisms of inherited and internally-generated variability is an important direction for systems neuroscience, as it is widely acknowledged that integration areas of cortex receive common fluctuations from outside brain areas.

4. *Conclusion*

This thesis presented two studies on the propagation of shared variability through multi-layer cortical circuits. In Chapter 2, I uncovered low-rank shared variability in motor cortex likely owing to somatosensory inputs. In Chapter 3, I discovered that the dimension of shared variability expanded between V4 and PFC. I then produced a two-layer spiking network model that qualitatively captured the *in vivo* data, thus revealing one mechanism by which the visual circuit could have produced the observed dimensionality expansion. In this concluding chapter, I will briefly discuss relationships between these studies and present possible extensions of the work for future studies.

4.1 Treatment of shared variability across studies

Though both Chapters 2 and 3 studied circuits in which one cortical area inherited low-rank shared variability from another cortical area, the methods by which we uncovered low-rank co-variability structure from the inputs differed. In Chapter 3, we were fortunate to have simultaneous recordings of V4 and PFC, meaning we could directly observe the neural activity of the circuit's upstream population. This made it possible to disambiguate whether V4 activity had low-rank shared variability, or whether this latent mode of low-rank variability arose through common projections from V4 neurons onto the PFC population. Our analyses of V4 data showed that trial-to-trial activity in V4 was, itself, low-dimensional, and we inferred (through functional tuning properties of the circuit) that structured projections from V4 were likely an additional means of correlating the downstream PFC population. By contrast, in Chapter 2, we only had access to recordings from the downstream M1 population of our circuit; we were not able to directly observe S1 activity. It is thus impossible to distinguish whether the shared variability detected in M1 was due to low-dimensional activity in S1 or due to common projections from S1 neurons; As discussed in Section 1.3.2 of this thesis Introduction, these two circuit mechanisms of correlation look identical from the vantage point of the observable M1 population.

In Chapter 3, our dataset consisted of many repeated presentations of a finite set of visual stimuli. It was thus trivial to distinguish the recorded neural population's response to the stimulus from its trial-to-trial response; the trial-to-trial response was simply the residual neural activity about the NHP's mean response to each of the target stimuli. In Chapter 2, the task (minimally constrained 2D reaches or imagined reaches) was less stereotyped and had more degrees of freedom. An encoding (regression) model was required to determine the neural population's evoked response. In initial analyses, we took the neural population's trial-to-trial variability to be the residuals of this regression. We then developed a model of motor encoding that included a fluctuation shared across the M1 population. In this case, we fit an explicit model of M1's shared variance simultaneously with a model of M1's stimulus tuning, through a recursive least squares algorithm. Each of these descriptions of neural variability influenced the outcome of other. Goris et al. [96] titled their study *Partitioning Neural Variability*, and I believe this is an apt description of the frequent ambiguities that arise when attributing portions of a neuron's total spiking variability to "signal" or "noise".

The statistical models that we used to characterize the rank of shared variability differed between chapters, as well. In Chapter 2, our model of shared variability was constrained to be 1-dimensional by design. Fitting addition dimensions of variance would require the use of a true state space model with a latent state, rather than our chosen approach of regressing all response variables against a time vector. (If we were to fit multiple shared fluctuations using our current model format, those fluctuations would share scaling.) The form of our single shared modulator, however, was very flexible, and provided a non-parametric way to induce smoothness in the trajectory of our shared fluctuation over time. This approach has the general flavor of using a dimensionality reduction technique like GPFA [42] fitted with a single latent dimension, as GPFA enforces temporal smoothness of data in the latent space (albeit parametrically) while taking the dynamics of the latent space to be stationary. In Chapter 3 we used cross-validation to determine the dimensionality of the FA-identified latent space, but data was taken to be i.i.d. (by definition of the FA model) and was not required to have smooth trajectories in that multi-dimensional latent space.

Finally, in Chapter 3, we considered that the shared variability read out from our downstream neural population (PFC) was due to both inputs from upstream area V4 and PFC's own recurrent interactions. We had no such model of M1 recurrent interactions in Chapter 2. The effects of recurrent interactions are nearly impossible to study without a) direct, cell-type specific perturbations of a neural circuit through experimental techniques like optogenetics [158] or b) a mechanistic modeling approach to studying the circuit, in which the biophysical properties of individual neurons and pairwise interactions of those neurons are considered. Lacking these approaches in the Chapter 2 study, we in some senses assumed M1's readout of the shared fluctuation was either a direct reflection or trivial transformation of the S1's input to M1. Chapter 3 demonstrates how this approach to studying multi-area circuits can be fraught with error when the recurrent dynamics of the readout layer of the circuit are highly non-linear. Had we observed only PFC neural activity in Chapter 3 and assumed PFC's shared variability was directly inherited from our outside brain area V4, we would have come to the incorrect conclusion that V4 had had high dimensional shared variability. Alternatively we might have incorrectly assumed PFC's high dimensional shared variability could only be explained through the inheritance of additional sensory inputs. Many systems neuroscience studies implicitly assume this sort of one-to-one mapping between circuit inputs and outputs; this assumption may not be appropriate, particularly when studying integrative areas of the brain with complex dynamics.

4.2 Future directions

4.2.1 A theory of transitions between PFC attractor dynamics

In Chapter 3, I developed a circuit model in which PFC activity was metastable by way of transitions between two attractor states reflecting tuned inputs. I studied this system by linearizing the networks dynamics around an operating points in a single attractor State (L or R). The current network model thus provides no description of the dynamics of state transitions. Dimensionality reduction techniques like recurrent switching linear dynamical systems [159] might better characterize the structure of population activity across transitions between States L and R. Ultimately, it would be desirable to develop a full theory of our network's metastable dynamics, in which we could predict the transition rates of our neuronal ensembles. Such a theory remains an open problem in theoretical neuroscience.

4.2.2 PFC recurrent dynamics and mixed selectivity

Rigotti et al. [160] noted that single neurons in PFC have complex receptive field properties that seem to simultaneously encode mixtures of multiple task parameters. Populations of such "mixed-selectivity" neurons in PFC encode distributed information about all task-relevant sensory information, resulting in high-dimensional neural representations of the task. Rigotti et al. [160] characterized the advantage of such a neural code – non-linear receptive fields of individual neurons allow for neural ensembles to encode a larger range of stimuli than would be possible if each neuron exhibited linear selectivity. More over, high-dimensional representations of stimuli are more easily read out by linear classifiers, as finite data becomes increasingly separable by a hyper-plane when projected into higher dimensions. (Machine learning classifiers frequently exploit this observation with a technique known as the *kernel trick* [161].)

Though Rigotti et al. [160] characterized why mixed-selectivity might be important in an integrative and cognitive brain area like PFC, they did not propose any circuit mechanisms that might give rise to PFC's high-dimensional population dynamics or complex neural responses. I believe that the network model in Chapter 3 provides one mechanism through which mixed-selectivity might arise. In Chapter 3, we showed that multiple, tuned inputs to PFC give rise to highly non-linear population dynamics, in which the network "time-shares" between two attractor states representative of its tuned inputs. "Time-sharing" leads individual PFC neurons to have highly variable responses; switches between attractor states manifest as high firing rate variability in neuronal responses, measurable through larger neuronal Fano factors. Moreover, when the strongly correlating attractor dynamics of disjoint inputs are delicately counterbalanced with recurrent assemblies of excitatory and inhibitory neurons, the variability of neural responses across one tuned assembly can still be high. This gives rise to high dimensional dynamics akin to those described by Rigotti et al. [160].

I believe it would be interesting to extend Chapter 3's network model to include more than two tuned input populations. This would capture the many sensory inputs that PFC receives. All of these tuned inputs would likely have structured projections to specific subpopulations of PFC, and it is correlated, structured inputs that give rise to the attractor dynamics that we studied in Chapter 3. If we extended our model to include tuned inputs from additional sensory modalities, not all inputs would project to completely disjoint subpopulations of PFC. For example, a single PFC neuron might be a member of one assembly that receives tuned visual information, and a different, potentially overlapping assembly that receives tuned auditory information. I am interested in studying how this extension of our model might be capable of producing non-linear response properties in individual neurons, as they "time-share" between the dynamics of their multiple streams of sensory information. Such a circuit model might shed light on the mechanisms underlying multi-sensory integration.

A. *Factor Analysis and the general form of Linear Gaussian Models*

This appendix provides an overview of Factor Analysis (FA), a dimensionality reduction technique that has been used extensively on neural data (Introduction 1.2.1),[21] and is applied repeatedly to neural and simulated neural data in Chapter 3 of this thesis.

FA is part of a larger class of Linear Gaussian Models [40], which are discrete time linear dynamical systems of the general form

$$\begin{aligned} \vec{x}_{t+1} &= A\vec{x}_t + \vec{w} & \qquad \vec{w} &\sim \mathcal{N}(0, Q) \\ \vec{y}_t &= C\vec{x}_t + \vec{\epsilon} & \qquad \vec{\epsilon} &\sim \mathcal{N}(0, R). \end{aligned} \tag{A.1}$$

Here, $\vec{y}$ is a matrix of observable variables. When Equation A.1 is used to model latent dynamics in neural data, $\vec{y}$ is the zero-mean population vector of observed spike counts from N simultaneously recorded neurons. $\vec{x}$ is the state of the population activity in a K-dimensional latent subspace, where $K << N$. $\vec{x}$ evolves according to first-order Markov dynamics, governed by state transition matrix A. $\vec{w}$ is a random variable representing state evolution. C is the *generative matrix* of model parameters relating the latent space $\vec{x}$ to the observable data $\vec{y}$, and $\vec{\epsilon}$ is a matrix of observation noise.

If we assume that our neural dataset is i.i.d., the underlying state matrix $\vec{x}$ has no dynamics. In this case, $A = 0$, and the generative model reduces to

$$\begin{aligned} \vec{x} &= \vec{w} & \qquad \vec{w} &\sim \mathcal{N}(0, Q) \\ \vec{y} &= C\vec{x} + \vec{\epsilon} & \qquad \vec{\epsilon} &\sim \mathcal{N}(0, R). \end{aligned} \tag{A.2}$$

FA is a Linear Gaussian Model with stationary dynamics of form (A.2). FA places a constraint on the structure of the observation noise $\vec{\epsilon}$ such that $R = \mathrm{diag}(R) = \psi$. In neuroscience, this constraint represents the assumption that the observation noise is uncorrelated between neurons and instead represents private sources of trial-to-trial neuronal variability, such as stochastic vesicle release.

The marginal distribution of $\vec{y}$ in Equation (A.2) is a Gaussian of form

$$\vec{y} \sim \mathcal{N}(0, CQC^\top + \psi). \tag{A.3}$$

Because there is an arbitrary sharing of scaling between Q and C, we can assume $Q = \mathbb{I}$. In FA, the generative matrix of model parameters C are called *loadings* onto latent factors. We will therefore make the variable change $C = \mathbb{L}$, where $\mathbb{L}$ is a loading matrix. Thus, our FA model takes the form

$$\vec{y} \sim \mathcal{N}(0, \mathbb{L}\mathbb{L}^\top + \psi). \tag{A.4}$$

This is equivalent to saying that the spike count covariance of our neural data can be decomposed into a shared variability component $\mathbb{L}\mathbb{L}^\top$ and a private variability component ψ. The dimension of shared variability can then be understood as $\mathrm{rank}(\mathbb{L}\mathbb{L}^\top)$, which is equivalent to performing an eigendecomposition of the shared variability in the K dimensional latent space:

$$\mathrm{Cov}(\vec{y}, \vec{y}) = \sum_i^K \lambda_i \nu_i \nu_i + \psi. \tag{A.5}$$

Here, λ_i is the eigenvalue of the ith latent mode of neural population activity.

We will denote the cumulative percentage of shared variance explained by the top J latent dimensions p_{Cum}; using (A.5), this is

$$p_{\text{Cum}} = \frac{\sum_i^J \lambda_i}{\sum_i^K \lambda_i} \tag{A.6}$$

for $J \leq K$ and latent modes of neural activity in descending eigenvalue order. p_{Cum} is frequently visualized in Chapter 3. Associated metric d_{shared} is the number of latent modes of neural activity J required to explain 95% of the shared variance, or the number of latent modes J such that $p_{\text{Cum}} \geq 0.95$.

Chapter 3 also reports the percent of each neuron's total variance that is shared with other neurons in the population [59]. Using Eq. (A.4), we will denote p^{shared} of neuron n

$$p_n^{\text{shared}} = \frac{\mathbb{L}_n \mathbb{L}_n^\top}{\mathbb{L}_n \mathbb{L}_n^\top + \psi_k}, \tag{A.7}$$

where $\mathbb{L}_n$ is the nth row of the loading matrix $\mathbb{L}$ and ψ_k is neuron k's private trial-to-trial variability, which is the kth diagonal in diagonal matrix ψ.

Finally, in Chapter 3, we report the residual spike count covariance without the 1st latent dimension, expressed by

$$Q = \text{Cov}(\vec{y}, \vec{y}) - \mathbb{L}_1 \mathbb{L}_1^\top, \tag{A.8}$$

where $\mathbb{L}_1$ denotes loadings onto the first latent dimension when FA is only fitted with $K = 1$ latent dimension.

FA model parameters in Chapter 3 were fitted using an Expectation Maximization (EM) [141] algorithm. Fits were performed using two-fold cross validation, where the dimensionality of the latent space K was selected according to the cross-validated log-likelihood of the models.

B. *Linear response approximation of covariance*

This appendix reviews approximation methods [139, 120] of using network architecture to predict the pairwise spiking covariance of neurons. Consider a neuron with stochastically fluctuating membrane potential dynamics. We begin with a linear ansatz [162] that says the process by which a neuron integrates its inputs and produces a realization of a spike train is linear. This ansatz requires that the input $X(t)$ to the neuron be weak in relation to the neuron's underlying noise process $\xi(t)$, which drive its membrane potential fluctuations. Assuming this is true, a realization of a neuron's spiking output $y(t)$ can be defined

$$y_i(t) \approx y_{i0}(t) + (G_i * X)(t), \tag{B.1}$$

where $y_0(t)$ is the unperturbed point process, or realization of the neuron's spiking output due to its unperturbed membrane potential dynamics, $X(t)$ is a weak input with banishing temporal average over the window in which we observe the system's behavior, and $G(t)$ is the neuron's linear response, which measures its sensitivity to its inputs.

When the neuron is part of a larger network, input $X(t)$ comes from the neuron's recurrent network interactions. So long as those interactions are still sufficiently weak such that the neuron's spike train output is a linear transformation of its inputs, our linear ansatz still applies. In this case, we will replace $X(t)$ with the expression $(f_i(t) - \mathbb{E}_t[f_i])$, where $E_t[.]$ is an expectation over time and

$$f_i(t) = \sum_j J_{ij}(F_j * y_j)(t). \tag{B.2}$$

Here, $y_j(t)$ is the spiking response of neuron j that projects to neuron i with synaptic strength J_{ij}, and F_j is the synaptic filter applied to neuron j's output.

The convolution terms of Equations (B.1) and (B.2) become multiplicative relations in the

Fourier domain. We thus consider the Fourier transform of a spike train, $y_i(\omega) = \int_{-\infty}^{\infty} y_i(t)e^{-2\pi i\omega t}dt$, where ω is frequency. Combining Equations (B.1) and (B.2), the recurrent network formulation of the spike train of neuron i in the Fourier domain is

$$y_i(\omega) = y_i^0(\omega) + G_i(\omega)\left(\sum_j J_{ij}F_j(\omega)y_j(\omega)\right). \tag{B.3}$$

The linear response $G_i(\omega)$ now measures the degree to which synaptic currents at frequency ω are transferred into modulations about background spike train activity $y_i^0(\omega)$.

The cross-spectral density (CSD) of the activity of i and j – which is the equivalent to the Fourier transform of the pairwise spike train cross-covariances of i and j, $C_{ij}(s)$ – is written:

$$C_{ij}(\omega) = \langle y_i(\omega)y_j^*(\omega)\rangle, \tag{B.4}$$

where y^* is the conjugate transpose of y. The full pairwise CSD matrix of our network activity is then

$$C(\omega) = \left(\mathbb{I} - (J\cdot K(\omega))\right)^{-1} C^0(\omega)\left(\mathbb{I} - (J\cdot K(\omega))^*\right)^{-1}, \tag{B.5}$$

where $\cdot$ denotes element wise multiplication and $K(\omega)$ is an interaction matrix with entries $K_{ij}(\omega) = A_i(\omega)F_{ij}(\omega)$. $C^0(\omega)$ is the CSD in the absence of recurrent interactions.

We now note that spike count covariances over long windows ($\Delta t \to \infty$) can be expressed as the zero-frequency CSD:

$$\lim_{\Delta t\to\infty}\frac{1}{\Delta t}\text{Cov}\left(\int_t^{\Delta t} U(t')dt' \int_t^{\Delta t} Z(t')dt'\right) = \langle U, Z\rangle(\omega = 0). \tag{B.6}$$

For large Δt, the spike count covariance of a pair of neurons i and j can then be approximated as:

$$\text{Cov}\left(N_i(t, t+\Delta t), N_j(t, t+\Delta t)\right) \approx \Delta t\langle y_i(t'), y_j(t')\rangle(\omega = 0) \tag{B.7}$$

This means the linear response approximation of $C(\omega = 0)$ can be reduced to

$$C = \left(\mathbb{I} - (J\cdot K)\right)^{-1} C^0\left(\mathbb{I} - (J\cdot K)^\top\right)^{-1}, \tag{B.8}$$

where C is the spike count covariance of the network over an infinitely long window (B.7) and

C^0 is the spike count covariance of the network in the absence of recurrent interactions, also over an infinitely long window. Equation (B.8) is known as the zero-frequency linear response approximation. We use the zero-frequency linear response approximation throughout Chapter 3 rather than computing the population CSD over all frequencies ω (B.5). The zero-frequency approximation both dramatically simplifies the calculations required to characterize population covariance and makes our predictions of population covariance structure more comparable to the myriad of experimental neuroscience studies that have reported values of spike count covariance and noise correlations (computed using spike count covariance) in the brain [114].

We obviously do not observe our network activity over infinitely long time windows. Why is the approximation of Equation (B.8) then valid? The answer lies in the shape of neuronal linear response functions $G(\omega)$, which tend to be approximately constant from $\omega = 0$ to $\omega \approx 30$ Hz [65, 163]. This claim is equivalent to saying that neuron i has the same sensitivity to its inputs for synaptic current modulations of frequency 30 Hz or less.

In Chapter 3, we observe our network activity over windows of $\Delta t = 50$ ms. This is equivalent of observing our network's response to perturbations at frequency $\omega = 20$ Hz. While our choice of timescale is probably still within the range ω over which $G(\omega)$ is approximately constant, we do note that we are at the upper boundary of ω over which the zero-frequency response (B.8) might provide a good approximation of our second order network statistics. We made a conscious choice to observe our system at $\Delta t = 50$ ms windows because we needed to capture the spike count covariance of our population activity within one attractor state of our network, and our network tended to transition between attractor states at timescales $\tau < 300$ ms. However, we acknowledge that the $\Delta t = 50$ ms windows over which we compute spike count covariance might be another source of error in our reported linear response approximation of C in network State L (Chapter 3, Figure 3.6d-e).

C. *Chapter 3 Supplemental Figures*

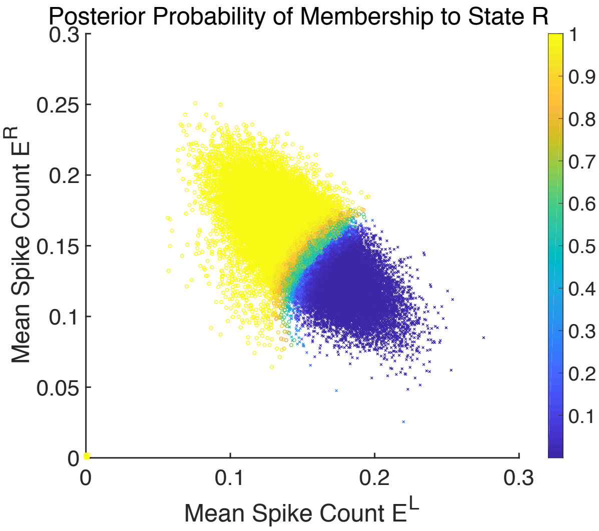

Figure C.1: Gaussian Mixture Model (GMM) state partitions of spiking network activity. Each data point here represents one time bin ($\Delta t = 50$ ms) of the activity of all $N^E = 4000$ excitatory cells in the $R = 2.3, \sigma = 0.71$ network, simulated for a total of $T = 2000s$. The N^E dimensional population activity is projected onto the 4D space described by the mean and variance of the firing activity of each hemifield. (Here we visualize the data in the 2D space described by the hemifield population means.) The GMM computes the posterior probability that each time bin of population activity belongs to State R. (The GMM was constrained to 2 clusters, and the posterior probability of membership to State R and State L sum to 1.) We accepted time bins for which the posterior probability of membership to either state was greater than 0.97. Other time bins (those that are not true yellow or royal blue in this visualization) were considered to represent dynamics in which our network was transitioning between the two attractor states.

Bibliography

[1] S. Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience* **3**, 31 (2009).

[2] F. Rieke, D. Warland, R. D. R. Van Steveninck, W. S. Bialek et al. *Spikes: exploring the neural code*, volume 7 (MIT press Cambridge, 1999).

[3] A. Pouget, P. Dayan & R. Zemel. Information processing with population codes. *Nature Reviews Neuroscience* **1**(2), 125–132 (2000).

[4] R. Yuste. From the neuron doctrine to neural networks. *Nature reviews neuroscience* **16**(8), 487–497 (2015).

[5] D. A. Ruff, A. M. Ni & M. R. Cohen. Cognition as a window into neuronal population space. *Annual review of neuroscience* **41**, 77–97 (2018).

[6] M. D. Humphries. Dynamical networks: finding, measuring, and tracking neural population activity using network science. *Network Neuroscience* **1**(4), 324–338 (2017).

[7] I. H. Stevenson & K. P. Kording. How advances in neural recording affect data analysis. *Nature neuroscience* **14**(2), 139 (2011).

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

[9] G. Buzsáki. Large-scale recording of neuronal ensembles. *Nature neuroscience* **7**(5), 446–451 (2004).

[10] T. Schrödel, R. Prevedel, K. Aumayr, M. Zimmer & A. Vaziri. Brain-wide 3d imaging of neuronal activity in caenorhabditis elegans with sculpted light. *Nature methods* **10**(10), 1013 (2013).

[11] M. Z. Lin & M. J. Schnitzer. Genetically encoded indicators of neuronal activity. *Nature neuroscience* **19**(9), 1142 (2016).

[12] M. E. J. Obien, K. Deligkaris, T. Bullmann, D. J. Bakkum & U. Frey. Revealing neuronal function through microelectrode array recordings. *Frontiers in neuroscience* **8**, 423 (2015).

[13] G. Hong & C. M. Lieber. Novel electrode technologies for neural recordings. *Nature Reviews Neuroscience* **20**(6), 330–345 (2019).

[14] N. A. Steinmetz, C. Koch, K. D. Harris & M. Carandini. Challenges and opportunities for large-scale electrophysiology with neuropixels probes. *Current opinion in neurobiology* **50**, 92–100 (2018).

[15] S. Saxena & J. P. Cunningham. Towards the neural population doctrine. *Current opinion in neurobiology* **55**, 103–111 (2019).

[16] B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker & K. Josić. The mechanics of state-dependent neural correlations. *Nature neuroscience* **19**(3), 383 (2016).

[17] R. E. Kass et al. Computational neuroscience: Mathematical and statistical perspectives. *Annual review of statistics and its application* **5**, 183–214 (2018).

[18] A. Kohn, R. Coen-Cagli, I. Kanitscheider & A. Pouget. Correlations and neuronal population information. *Annual review of neuroscience* **39**, 237–256 (2016).

[19] J. A. Gallego, M. G. Perich, L. E. Miller & S. A. Solla. Neural manifolds for the control of movement. *Neuron* **94**(5), 978–984 (2017).

[20] S. Ganguli & H. Sompolinsky. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu. Rev. Neurosci.* **35**(1), 485–508 (2012).

[21] J. P. Cunningham & M. Y. Byron. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience* **17**(11), 1500 (2014).

[22] I.-C. Lin, M. Okun, M. Carandini & K. D. Harris. The nature of shared cortical variability. *Neuron* **87**(3), 644–656 (2015).

[23] N. C. Rabinowitz, R. L. Goris, M. Cohen & E. P. Simoncelli. Attention stabilizes the shared gain of v4 populations. *Elife* **4**, e08998 (2015).

[24] C. Huang et al. Circuit models of low-dimensional shared variability in cortical networks. *Neuron* **101**(2), 337–348 (2019).

[25] B. B. Averbeck, P. E. Latham & A. Pouget. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**(5), 358–366 (2006).

[26] K. Josić, E. Shea-Brown, B. Doiron & J. de la Rocha. Stimulus-dependent correlations and population codes. *Neural computation* **21**(10), 2774–2804 (2009).

[27] A. Litwin-Kumar. *Relationship between neuronal architecture and variability in cortical circuits*. Ph.D. thesis, Carnegie Mellon University (2013).

[28] A. A. Faisal, L. P. Selen & D. M. Wolpert. Noise in the nervous system. *Nature reviews neuroscience* **9**(4), 292–303 (2008).

[29] R. Rosenbaum, J. E. Rubin & B. Doiron. Short-term synaptic depression and stochastic vesicle dynamics reduce and shape neuronal correlations. *Journal of neurophysiology* **109**(2), 475–484 (2013).

[30] A. D. Bird & M. J. Richardson. Long-term plasticity determines the postsynaptic response to correlated afferents with multivesicular short-term synaptic depression. *Frontiers in computational neuroscience* **8**, 2 (2014).

[31] J. F. A. Poulet & C. C. H. Petersen. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**(7206), 881–885 (2008).

[32] M. R. Cohen & J. H. R. Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**(12), 1594–1600 (2009).

[33] J. F. Mitchell, K. A. Sundberg & J. H. Reynolds. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**(6), 879–888 (2009).

[34] J. Jeanne, T. Sharpee & T. Gentner. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron* **78**(2), 352–363 (2013).

[35] Y. Gu et al. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron* **71**(4), 750–761 (2011).

[36] R. Salazar, N. Dotson, S. Bressler & C. Gray. Content-specific fronto-parietal synchronization during visual working memory. *Science* **338**(6110), 1097–1100 (2012).

[37] N. M. Dotson, R. F. Salazar & C. M. Gray. Frontoparietal correlation dynamics reveal interplay between integration and segregation during visual working memory. *Journal of Neuroscience* **34**(41), 13600–13613 (2014).

[38] M. Vidne et al. Modeling the impact of common noise inputs on the network activity of retinal ganglion cells. *Journal of computational neuroscience* **33**(1), 97–121 (2012).

[39] J. E. Kulkarni & L. Paninski. Common-input models for multiple neural spike-train data. *Network: Computation in Neural Systems* **18**(4), 375–407 (2007).

[40] S. Roweis & Z. Ghahramani. A unifying review of linear gaussian models. *Neural computation* **11**(2), 305–345 (1999).

[41] J. P. Cunningham & Z. Ghahramani. Linear dimensionality reduction: Survey, insights, and generalizations. *The Journal of Machine Learning Research* **16**(1), 2859–2900 (2015).

[42] M. Y. Byron et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. In *Advances in neural information processing systems*, 1881–1888 (2009).

[43] S. Linderman et al. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Artificial Intelligence and Statistics*, 914–922 (2017).

[44] M. Y. Byron et al. Extracting dynamical structure embedded in neural activity. In *Advances in neural information processing systems*, 1545–1552 (2006).

[45] R. E. Kass & V. Ventura. A spike-train probability model. *Neural computation* **13**(8), 1713–1720 (2001).

[46] L. Paninski, J. Pillow & J. Lewi. Statistical models for neural encoding, decoding, and optimal stimulus design. *Progress in brain research* **165**, 493–507 (2007).

[47] A. C. Smith & E. N. Brown. Estimating a state-space model from point process observations. *Neural computation* **15**(5), 965–991 (2003).

[48] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue & E. N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology* **93**(2), 1074–1089 (2005).

[49] L. Srinivasan, U. T. Eden, A. S. Willsky & E. N. Brown. A state-space analysis for reconstruction of goal-directed movements using neural signals. *Neural computation* **18**(10), 2465–2494 (2006).

[50] B. M. Yu et al. Mixture of trajectory models for neural decoding of goal-directed movements. *Journal of neurophysiology* **97**(5), 3763–3780 (2007).

[51] V. Lawhern, W. Wu, N. Hatsopoulos & L. Paninski. Population decoding of motor cortical activity using a generalized linear model with hidden states. *Journal of neuroscience methods* **189**(2), 267–280 (2010).

[52] L. Paninski et al. A new look at state-space models for neural data. *Journal of computational neuroscience* **29**(1-2), 107–126 (2010).

[53] S. W. Linderman & S. J. Gershman. Using computational theory to constrain statistical models of neural data. *Current opinion in neurobiology* **46**, 14–24 (2017).

[54] M. L. Schölvinck, A. B. Saleem, A. Benucci, K. D. Harris & M. Carandini. Cortical state determines global variability and correlations in visual cortex. *Journal of Neuroscience* **35**(1), 170–178 (2015).

[55] M. Okun et al. Diverse coupling of neurons to populations in sensory cortex. *Nature* **521**(7553), 511–515 (2015).

[56] M. Vidne et al. Inferring functional connectivity in an ensemble of retinal ganglion cells sharing a common input. In *Frontiers in systems neuroscience conference abstract: Computational and systems neuroscience 2009* (2009).

[57] W. Wu, J. E. Kulkarni, N. G. Hatsopoulos & L. Paninski. Neural decoding of hand motion using a linear state-space model with hidden states. *IEEE Transactions on neural systems and rehabilitation engineering* **17**(4), 370–378 (2009).

[58] R. C. Williamson, B. Doiron, M. A. Smith & M. Y. Byron. Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current opinion in neurobiology* **55**, 40–47 (2019).

[59] R. C. Williamson et al. Scaling properties of dimensionality reduction for neural populations and network models. *PLoS computational biology* **12**(12) (2016).

[60] A. Litwin-Kumar & B. Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**(11), 1498–1505 (2012).

[61] B. R. Cowley et al. Slow drift of neural activity as a signature of impulsivity in macaque visual and prefrontal cortex. *bioRxiv* (2020).

[62] C. van Vreeswijk & H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**(5293), 1724–1726 (1996).

[63] C. van Vreeswijk & H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Comput.* **10**(6), 1321–1371 (1998).

[64] A. Renart et al. The asynchronous state in cortical circuits. *Science* **327**(5965), 587–590 (2010).

[65] A. Litwin-Kumar, A.-M. M. Oswald, N. N. Urban & B. Doiron. Balanced synaptic input shapes the correlation between neural spike trains. *PLoS Comput. Biol.* **7**(12), e1002305 (2011).

[66] F. S. Chance, L. Abbott & A. D. Reyes. Gain modulation from background synaptic input. *Neuron* **35**(4), 773–782 (2002).

[67] J. A. Cardin, L. A. Palmer & D. Contreras. Cellular mechanisms underlying stimulus-dependent gain modulation in primary visual cortex neurons in vivo. *Neuron* **59**(1), 150–160 (2008).

[68] M. T. Schaub, Y. N. Billeh, C. A. Anastassiou, C. Koch & M. Barahona. Emergence of slow-switching assemblies in structured neuronal networks. *PLoS computational biology* **11**(7) (2015).

[69] B. Kriener, M. Helias, A. Aertsen & S. Rotter. Correlations in spiking neuronal networks with distance dependent connections. *J. Comput. Neurosci.* **27**(2), 177–200 (2009).

[70] A. Keane & P. Gong. Propagating waves can explain irregular neural dynamics. *Journal of Neuroscience* **35**(4), 1591–1605 (2015).

[71] R. Rosenbaum & B. Doiron. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X* **4**(2), 021039 (2014).

[72] R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin & B. Doiron. The spatial structure of correlated neuronal variability. *Nature neuroscience* **20**(1), 107 (2017).

[73] S. H. Scott. Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience* **5**(7), 532–545 (2004).

[74] J. Wright, V. G. Macefield, A. van Schaik & J. C. Tapson. A review of control strategies in closed-loop neuroprosthetic systems. *Frontiers in neuroscience* **10**, 312 (2016).

[75] N. Li, T.-W. Chen, Z. V. Guo, C. R. Gerfen & K. Svoboda. A motor cortex circuit for motor planning and movement. *Nature* **519**(7541), 51–56 (2015).

[76] A. P. Georgopoulos, J. F. Kalaska, R. Caminiti & J. T. Massey. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *The Journal of Neuroscience* **2**(11), 1527–1537 (1982).

[77] K. Pipereit, O. Bock & J.-L. Vercher. The contribution of proprioceptive feedback to sensorimotor adaptation. *Experimental Brain Research* **174**(1), 45 (2006).

[78] C. Ghez, J. Gordon & M. F. Ghilardi. Impairments of reaching movements in patients without proprioception. ii. effects of visual information on accuracy. *Journal of neurophysiology* **73**(1), 361–372 (1995).

[79] J. Cole & J. Paillard. Living without touch and peripheral information about body position and movement: Studies with deafferented subjects. *The body and the self* 245–266 (1995).

[80] J. Paillard. Body schema and body image-a double dissociation. *Motor control, today and tomorrow* 197–214 (1999).

[81] F. Matyas et al. Motor control by sensory cortex. *Science* **330**(6008), 1240–1243 (2010).

[82] S. Sachidhanandam, V. Sreenivasan, A. Kyriakatos, Y. Kremer & C. C. Petersen. Membrane potential correlates of sensory perception in mouse barrel cortex. *Nature neuroscience* **16**(11), 1671–1677 (2013).

[83] A. B. Schwartz, X. T. Cui, D. J. Weber & D. W. Moran. Brain-controlled interfaces: movement restoration with neural prosthetics. *Neuron* **52**(1), 205–220 (2006).

[84] S. J. Bensmaia & L. E. Miller. Restoring sensorimotor function through intracortical interfaces: progress and looming challenges. *Nature Reviews Neuroscience* **15**(5), 313–325 (2014).

[85] S. C. Gandevia, K. M. Refshauge & D. F. Collins. Proprioception: peripheral inputs and perceptual interactions. In *Sensorimotor control of movement and posture*, 61–68 (Springer, 2002).

[86] R. Johansson & Å. B. Vallbo. Detection of tactile stimuli. thresholds of afferent units related to psychophysical thresholds in the human hand. *The Journal of physiology* **297**, 405 (1979).

[87] A. J. Suminski, D. C. Tkach, A. H. Fagg & N. G. Hatsopoulos. Incorporating feedback from multiple sensory modalities enhances brain–machine interface control. *The Journal of neuroscience* **30**(50), 16777–16787 (2010).

[88] B. M. Hooks, J. Y. Lin, C. Guo & K. Svoboda. Dual-channel circuit mapping reveals sensorimotor convergence in the primary motor cortex. *The Journal of Neuroscience* **35**(10), 4418–4426 (2015).

[89] J. L. Collinger et al. High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet* **381**(9866), 557–564 (2013).

[90] W. Wang, S. S. Chan, D. A. Heldman & D. W. Moran. Motor cortical representation of position and velocity during reaching. *Journal of neurophysiology* **97**(6), 4258–4270 (2007).

[91] M. A. Lebedev et al. Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain-machine interface. *The Journal of neuroscience* **25**(19), 4681–4693 (2005).

[92] B. Jarosiewicz et al. Functional network reorganization during learning in a brain-computer interface paradigm. *Proceedings of the National Academy of Sciences* **105**(49), 19486–19491 (2008).

[93] K. Ganguly & J. M. Carmena. Emergence of a stable cortical map for neuroprosthetic control. *PLoS Biol* **7**(7), e1000153 (2009).

[94] T. Hastie & R. Tibshirani. Generalized additive models. *Statistical science* 297–310 (1986).

[95] J. Fan, N. E. Heckman & M. P. Wand. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association* **90**(429), 141–150 (1995).

[96] R. L. Goris, J. A. Movshon & E. P. Simoncelli. Partitioning neuronal variability. *Nature neuroscience* **17**(6), 858 (2014).

[97] G. F. Elsayed, A. H. Lara, M. T. Kaufman, M. M. Churchland & J. P. Cunningham. Reorganization between preparatory and movement population responses in motor cortex. *Nature communications* **7**(1), 1–15 (2016).

[98] P. T. Sadtler et al. Neural constraints on learning. *Nature* **512**(7515), 423–426 (2014).

[99] Y. Mandelblat-Cerf, R. Paz & E. Vaadia. Trial-to-trial variability of single cells in motor cortices is dynamically modified during visuomotor adaptation. *The Journal of Neuroscience* **29**(48), 15053–15062 (2009).

[100] U. Rokni, A. G. Richardson, E. Bizzi & H. S. Seung. Motor learning with unstable neural representations. *Neuron* **54**(4), 653–666 (2007).

[101] B. P. Christie et al. Comparison of spike sorting and thresholding of voltage waveforms for intracortical brain-machine interface performance. *Journal of Neural Engineering* **12**(1), 016009 (2015).

[102] W. J. Ma, J. M. Beck, P. E. Latham & A. Pouget. Bayesian inference with probabilistic population codes. *Nature neuroscience* **9**(11), 1432–1438 (2006).

[103] C. R. Fetsch, A. Pouget, G. C. DeAngelis & D. E. Angelaki. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature neuroscience* **15**(1), 146–154 (2012).

[104] D. Huber et al. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature* **484**(7395), 473–478 (2012).

[105] R. Aronoff et al. Long-range connectivity of mouse primary somatosensory barrel cortex. *European Journal of Neuroscience* **31**(12), 2221–2233 (2010).

[106] T. Mao et al. Long-range neuronal circuits underlying the interaction between sensory and motor cortex. *Neuron* **72**(1), 111–123 (2011).

[107] I. Ferezou et al. Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* **56**(5), 907–923 (2007).

[108] S. M. Chase, A. B. Schwartz & R. E. Kass. Bias, optimal linear estimation, and the differences between open-loop simulation and closed-loop performance of spiking-based brain–computer interface algorithms. *Neural networks* **22**(9), 1203–1213 (2009).

[109] M. D. Golub, S. M. Chase, A. P. Batista & M. Y. Byron. Brain–computer interfaces for dissecting cognitive processes underlying sensorimotor control. *Current opinion in neurobiology* **37**, 53–58 (2016).

[110] K. E. Schroeder & C. A. Chestek. Intracortical brain-machine interfaces advance sensorimotor neuroscience. *Frontiers in Neuroscience* **10**, 291 (2016).

[111] M. N. Shadlen & W. T. Newsome. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**(4), 569–579 (1994).

[112] T. P. Vogels & L. F. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.* **25**(46), 10786 –10795 (2005).

[113] M. R. Cohen & A. Kohn. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**(7), 811–819 (2011).

[114] A. Kohn, A. Zandvakili & M. A. Smith. Correlations and brain states: from electrophysiology to functional imaging. *Curr. Opin. Neurobiol.* **19**(4), 434–438 (2009).

[115] A. S. Ecker et al. State dependence of noise correlations in macaque primary visual cortex. *Neuron* **82**(1), 235–248 (2014).

[116] R. Pyle & R. Rosenbaum. Spatiotemporal dynamics and reliable computations in recurrent spiking neural networks. *Physical review letters* **118**(1), 018103 (2017).

[117] R. Darshan, C. Van Vreeswijk & D. Hansel. Strength of correlations in strongly recurrent neuronal networks. *Physical Review X* **8**(3), 031072 (2018).

[118] Y. Hu, J. Trousdale, K. Josić & E. Shea-Brown. Motif statistics and spike correlations in neuronal networks. *Journal of Statistical Mechanics: Theory and Experiment* **2013**(03), P03012 (2013).

[119] Y. Hu, J. Trousdale, K. Josić & E. Shea-Brown. Local paths to global coherence: cutting networks down to size. *Physical Review E* **89**(3), 032802 (2014).

[120] J. Trousdale, Y. Hu, E. Shea-Brown & K. Josić. Impact of network structure and cellular response on spike time correlations. *PLoS computational biology* **8**(3) (2012).

[121] V. Pernice, B. Staude, S. Cardanobile & S. Rotter. How structure determines correlations in neuronal networks. *PLoS computational biology* **7**(5) (2011).

[122] G. K. Ocker, K. Josić, E. Shea-Brown & M. A. Buice. Linking structure and activity in nonlinear spiking networks. *PLoS computational biology* **13**(6), e1005583 (2017).

[123] F. Mastrogiuseppe & S. Ostojic. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**(3), 609 – 623.e29 (2018).

[124] Y. Ahmadian, F. Fumarola & K. D. Miller. Properties of networks with partially structured and partially random connectivity. *Physical Review E* **91**(1), 012820 (2015).

[125] S. Recanatesi, G. K. Ocker, M. A. Buice & E. Shea-Brown. Dimensionality in recurrent spiking networks: global trends in activity and local origins in connectivity. *PLoS computational biology* **15**(7), e1006446 (2019).

[126] A. Ponce-Alvarez, A. Thiele, T. D. Albright, G. R. Stoner & G. Deco. Stimulus-dependent variability and noise correlations in cortical mt neurons. *Proceedings of the National Academy of Sciences* **110**(32), 13162–13167 (2013).

[127] K. Wimmer et al. Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area mt. *Nature communications* **6**(1), 1–13 (2015).

[128] T. Kanashiro, G. K. Ocker, M. R. Cohen & B. Doiron. Attentional modulation of neuronal variability in circuit models of cortex. *Elife* **6**, e23978 (2017).

[129] G. Hennequin, Y. Ahmadian, D. B. Rubin, M. Lengyel & K. D. Miller. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron* **98**(4), 846–860 (2018).

[130] D. Pandya & E. Yeterian. Comparison of prefrontal architecture and connections. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **351**(1346), 1423–1432 (1996).

[131] C. H. Donahue & D. Lee. Dynamic routing of task-relevant signals for decision making in dorsolateral prefrontal cortex. *Nature neuroscience* **18**(2), 295 (2015).

[132] M. Siegel, T. J. Buschman & E. K. Miller. Cortical information flow during flexible sensorimotor decisions. *Science* **348**(6241), 1352–1355 (2015).

[133] J. M. Fuster. The prefrontal cortex—an update: time is of the essence. *Neuron* **30**(2), 319–333 (2001).

[134] S. B. Khanna, J. A. Scott & M. A. Smith. Dynamic shifts of visual and saccadic signals in prefrontal cortical regions 8ar and fef. *bioRxiv* 817478 (2019).

[135] G. Santhanam et al. Factor-analysis methods for higher-performance neural prostheses. *Journal of neurophysiology* **102**(2), 1315–1330 (2009).

[136] B. Everett. *An introduction to latent variable models* (Springer Science & Business Media, 2013).

[137] M. Bastian, S. Heymann & M. Jacomy. Gephi: An open source software for exploring and manipulating networks (2009).

[138] T. M. J. Fruchterman & E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience* **21**(11), 1129–1164 (1991).

[139] B. Lindner, B. Doiron & A. Longtin. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys. Rev. E* **72**(6), 061919 (2005).

[140] C. Baker, C. Ebsch, I. Lampl & R. Rosenbaum. Correlated states in balanced neuronal networks. *Physical Review E* **99**(5), 052414 (2019).

[141] A. P. Dempster, N. M. Laird & D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22 (1977).

[142] S. Funahashi, C. J. Bruce & P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of neurophysiology* **61**(2), 331–349 (1989).

[143] H. Suzuki & M. Azuma. Topographic studies on visual neurons in the dorsolateral prefrontal cortex of the monkey. *Experimental brain research* **53**(1), 47–58 (1983).

[144] A. Mikami, S. Ito & K. Kubota. Visual response properties of dorsolateral prefrontal neurons during visual fixation task. *Journal of neurophysiology* **47**(4), 593–605 (1982).

[145] E. H. Yeterian, D. N. Pandya, F. Tomaiuolo & M. Petrides. The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* **48**(1), 58–81 (2012).

[146] L. G. Ungerleider, T. W. Galkin, R. Desimone & R. Gattass. Cortical connections of area v4 in the macaque. *Cerebral Cortex* **18**(3), 477–499 (2008).

[147] M. L. Leavitt, F. Pieper, A. J. Sachs & J. C. Martinez-Trujillo. A quadrantic bias in prefrontal representation of visual-mnemonic space. *Cerebral Cortex* **28**(7), 2405–2421 (2018).

[148] D. A. Pollen, A. W. Przybyszewski, M. A. Rubin & W. Foote. Spatial receptive field organization of macaque v4 neurons. *Cerebral Cortex* **12**(6), 601–616 (2002).

[149] B. Doiron, B. Lindner, A. Longtin, L. Maler & J. Bastian. Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Phys. Rev. Lett.* **93**(4), 048101 (2004).

[150] S. Song, P. J. Sjöström, M. Reigl, S. Nelson & D. B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**(3), e68 (2005).

[151] R. Perin, T. K. Berger & H. Markram. A synaptic organizing principle for cortical neuronal groups. *PNAS* **108**(13), 5419 –5424 (2011).

[152] H. Ko et al. Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**(7345), 87–91 (2011).

[153] J. L. Chen et al. Clustered dynamics of inhibitory synapses and dendritic spines in the adult neocortex. *Neuron* **74**(2), 361–373 (2012).

[154] R. D. D'Souza, P. Bista, A. M. Meier, W. Ji & A. Burkhalter. Spatial clustering of inhibition in mouse primary visual cortex. *Neuron* **104**(3), 588–600 (2019).

[155] M. M. Churchland et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* **13**(3), 369–378 (2010).

[156] A. K. Churchland et al. Variance as a signature of neural computations during decision making. *Neuron* **69**(4), 818–831 (2011).

[157] D. R. Cox & V. Isham. *Point processes* (CRC Press, 1980).

[158] C. K. Kim, A. Adhikari & K. Deisseroth. Integration of optogenetics with complementary methodologies in systems neuroscience. *Nature Reviews Neuroscience* **18**(4), 222 (2017).

[159] S. W. Linderman et al. Recurrent switching linear dynamical systems. *arXiv preprint arXiv:1610.08466* (2016).

[160] M. Rigotti et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**(7451), 585–590 (2013).

[161] T. Hofmann, B. Schölkopf & A. J. Smola. Kernel methods in machine learning. *The annals of statistics* 1171–1220 (2008).

[162] C. Gardiner. A handbook for the natural and social sciences. *Springer Series in Synergetics* **13** (2009).

[163] G. K. Ocker & B. Doiron. Kv7 channels regulate pairwise spiking covariability in health and disease. *Journal of neurophysiology* **112**(2), 340–352 (2014).