Coordinated neural activity: Mechanistic origins and impact on stimulus coding

N. Alex Cayco Gajic

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2015

Reading Committee:
Eric Shea-Brown, Chair
Bernard Deconinck
Adrienne Fairhall

Program Authorized to Offer Degree:
Department of Applied Mathematics

University of Washington

Abstract

Coordinated neural activity: Mechanistic origins and impact on stimulus coding

Natasha Cayco Gajic

Chair of the Supervisory Committee:
Professor Eric Shea-Brown
Department of Applied Mathematics

How does the activity of populations of neurons encode the signals they receive? Since neurons in vivo are inherently variable, each fixed input to a population will elicit not a deterministic response, but rather a probability distribution of states of the individual neurons. Traditional theories of neural coding rely on single-cell tuning curves that describe the average response of each neuron to stimulus features. Adding complexity to this neuron-by-neuron view is the fact that neural activity is not independent: it is often correlated, reflecting shared input and connectivity. Such "coordinated" activity can have diverse and potentially strong impacts on how neural circuits encode stimuli.

In this dissertation, we combine dynamical and statistical tools to examine how single-cell and network properties dynamically generate coordinated neural spiking, and how this affects stimulus coding in populations of cells. First, we show how feedforward connectivity leads to the emergence of a neutrally stable subspace that allows information about input rates to be transmitted through layers. At this critical parameter regime, neural activity is characterized by *higher-order* interactions, meaning that the activity cannot be described by minimal models including only the lower-order moments (mean and pairwise interactions).

Interestingly, recent experiments have also demonstrated the existence of higher-order correlations in the neural activity patterns in retina and cortex. Using maximum entropy techniques, we show that in general populations, higher-order correlations can facilitate the encoding of stimulus information in neural activity patterns. We propose a statistical model for fitting neurophysiological data that incorporates only the most significant higher-order interactions. We apply this model to analyze the statistics of population firing patterns in the lateral geniculate nucleus of awake mice. Finally, we analyze dendritic nonlinearities as a novel mechanism by which intrinsic cell properties can generate higher-order correlations. Together, these results work towards determining the origins of coordinated spiking, understanding its impact on neural coding, and building better tools for quantification in electrophysiological data.

"And thus you will gain knowledge, guided by a little labour,
For one thing will illuminate the next, and blinding night
Won't steal your way; all secrets will be opened to your sight,
One truth illuminate another, as light kindles light."
– Lucretius, *On the Nature of Things*

Acknowledgements

I am deeply thankful for the unfailing support of my advisor, Eric Shea-Brown, whose personal dedication and infectious enthusiasm were constant sources of inspiration. My committee members Adrienne Fairhall, Bernard Deconinck, and Fred Rieke have also been instrumental to the maturation of my scientific and mathematical outlook. I have been very lucky for the opportunity to collaborate with Joel Zylberberg and Severine Durand, and am also grateful for years of invaluable feedback from colleagues at the University of Washington and beyond, especially the Shea-Brown research group, the Allen Institute for Brain Science, and the students and faculty of the Okinawa, Woods Hole, and Friday Harbor summer courses. The material in this dissertation is based upon work supported by the National Science Foundation under Grants No. DGE-0718124 and DGE-1256082, and by the Achievement Rewards for College Scientists (ARCS) Foundation.

More personally, I had the pleasure of engaging in countless scientific debates with Ben Lansdell, Eric Johlin, Braden Brinkman, Susan Massey, Pedro Maia, and Yann Sweeney. Such discourse has been one of the highlights of my graduate career, and I will miss it greatly.

Finally, my love of science and mathematics flourished as a youth thanks to the encouragement of several female mathematicians at San Jose State University, especially: Joanne Becker, Linda Valdes, and most importantly, my mother, the first mathematician I ever knew.

Contents

CHAPTER 1

Introduction

Each historical period has struggled to understand brain function using the language of its most advanced technology: clockwork automata during the Enlightenment, hydraulic machines in the Victorian era, and the computer today [40]. Viewed as input-output black boxes, brains and computers share several defining properties. Not unlike a computer, the central nervous system takes sensory input from the natural world and, via a cascade of electrical activity, performs a series of complex computations that leads to the appropriate behavior. Adding to the strength of the metaphor is the fact that information is represented in the brain as binary units signaling the presence or absence of neuronal *spikes*.

Spikes, also called action potentials, are fast fluctuations in the membrane potential of neurons that are transmitted to adjacent cells. Because of their all-or-nothing responses and stereotyped waveforms, the activity of populations of neurons can be described by sequences of spike times $I_i(t)$ corresponding to a 1 if neuron i fired an action potential in a small time window $[t, t + \Delta t]$ and a 0 otherwise; these are called *spike trains*. In a prototypical neuroscience experiment, an animal is presented a stimulus while the scientist records the spike trains of a population of neurons (as in Figure 1.1). Different stimuli – e.g., varying the direction of moving bars – elicit different spiking activity. This underlies the classic question of *neural coding*: how does the spiking activity of populations of neurons represent key features about the stimulus?

Complicating this question (and destroying the brain-as-computer metaphor) is the fact that, even for the same stimulus, neural activity varies from trial to trial. Because of this variability, neural populations represent stimuli not as specific responses but as *distributions* of responses.

Neural variability can originate internally or externally [46]. External sources of variability include contextual cues and nuisance parameters – for example, in response to images of the same river, neural activity in visual cortex may vary depending on lighting condition, angle of view, etc. Equally important are internal sources of variability, which may reflect the intrinsic state of the animal, such as levels arousal or attention [33, 48, 58, 75, 47], up/down states in anesthetized cortical tissue [43], or suboptimal bayesian inference [19]. Lower-level sources of neural variability also play a role. These

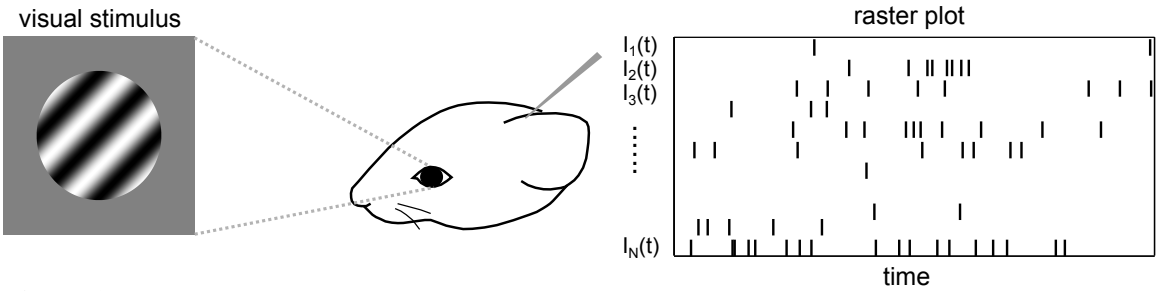

Figure 1.1: Schematic of a prototypical neuroscience experiment. A mouse is presented with visual stimuli (gratings moving in a particular direction) while an electrode records the membrane potential of a population of neurons. The spike train of the ith cell is denoted by $I_i(t)$. The panel on the right shows the *raster plot* of the recorded population. Each line at coordinate (t, i) represents a spike by neuron i at time t.

include: thermal noise [59], ion channel noise [149, 146, 138], probabilistic neurotransmitter release [26, 7], and chaotic dynamics [78, 102].

Whatever the source of this variability, downstream populations must learn to accurately decode which stimulus was presented to the animal based on neural activity from a single trial. If neurons were statistically independent, this variability could be averaged by summing over the full population. However, neural activity is far from independent, as it is often coordinated from neuron to neuron in ways that may affect how stimulus information is represented in neural populations [32].

In this dissertation, we examine how fundamental properties (such as network architecture and intrinsic nonlinearities) dynamically generate coordinated neural spiking, and how this affects stimulus coding in populations of cells. Chapter 1 tackles this topic in a simple, mathematically tractable setting: layered feedforward networks of binary units. "Complex" neural dynamics (i.e., non-independent spiking that avoids runaway synchronization by supporting a variety of firing patterns) emerges at a critical level of connectivity between layers. We develop and analyze a mean-field model describing the activity at each layer in the network, using a combination of Markov chain theory and discrete dynamical systems methods to predict when complex signal propagation is possible in these networks.

Interestingly, the stable modes of asynchronous activity show *higher-order* statistical structure, meaning that the dynamics can not be captured by a minimal model with knowledge of only the mean activity and pairwise statistics (the lower-order moments). Moreover, when compared with a model that only reproduces the lower-order moments, the higher-order interactions reduce the encoded information about the input rates.

For the second part of the dissertation, we use maximum entropy statistical models [124] to study higher-order correlations between the spiking activity of cells in general populations. Higher-order correlations have recently been observed in neural populations in retina [52, 141] and cortical areas [92, 100, 109, 79]. In the literature, examples exist of when higher-order correlations improve the ability to decode stimuli based on population activity [99, 52], as well as examples in which they have little effect on stimulus information

[109, 100]. While much work has been dedicated to understanding the role of correlations between cell pairs [154, 2, 13, 114, 112, 135, 36, 70, 127], less is understood about when and how higher-order correlations affect neural coding.

In Chapter 2, we study the impact that varying triplet correlations – the most frequently-observed higher-order correlation [52] – can have on stimulus encoding. We vary the effect of triplet correlations, while fixing the lower-order statistics to prescribed values typical of those reported in physiology experiments. We find significant increases in encoded stimulus information when triplet correlations among similarly-tuned cells are stimulus-dependent. On the other hand, stimulus-independent triplet correlations had little effect. We show that these findings can be explained geometrically as either positively or negatively skewing the distribution of the summed population activity in short time windows.

These results emphasize the importance of measuring higher-order correlations in neurophysiology experiments, as they may have a significant impact on the coding performance of neural systems. However, higher-order correlations can be difficult to quantify experimentally for two reasons. First, because synchronous spiking of groups of three or more cells is relatively infrequent, long recordings or many trials are required to measure higher-order correlations accurately. In Chapter 2, we give an estimate of how much data is necessary to measure higher-order moments within a certain expected precision.

Second, the number of higher-order interactions increases exponentially with system size. Fitting all higher-order interactions is not only computationally infeasible, but because of the finite number of samples in a given experiment, it can also lead to overfitting models to noise in the data. How can we determine which higher-order spiking features are needed to describe the structure of spiking activity in neural populations? In Chapter 3, we develop a higher-order maximum entropy model building off of [52] that identifies and fits only the most frequently occurring spiking interactions. Applying this model to extracellular recordings in mouse lateral geniculate nucleus, we find that the correlations between pairs of cells – but not higher-order correlations – shape the distribution of firing patterns.

Finally, many features of neural populations can create higher-order correlations, such as common input [17, 90, 153] and connectivity structure [69, 30]. Intrinsic properties, such as adaptation or nonlinear input summation, also play a role in shaping spiking statistics. Chapter 4 illustrates how higher-order correlations can be modulated by intrinsic cell properties – in particular, nonlinear dendrites. We close with a discussion of open questions in the field.

CHAPTER 2

Signal propagation in feedforward networks

Many basic questions remain unresolved in understanding how simple features of network connectivity determine the statistical structure of their outputs. In particular, as we vary the average connectivity strength between model neurons, what kinds of transitions occur in model dynamics? The first dynamical property we might study at a transition is neutral stability of trajectories. Intuitively, it appears that neutral stability could favor signal transmission, because it suggests that input patterns (and their noisy perturbations) will retain their original separation in state space, neither diverging nor converging towards some fixed attractor [23, 85]. The second, allied property that could occur as networks transition from weak to strong connectivity is the production of a wide range of output states – that is, a mix of firing patterns that induce a broad distribution with high response entropy. If responses are tallied via total network output, this could require statistical correlations of all orders [10]; thus, higher-order correlations are another statistical property of interest at network transitions. Finally, an assay that involves all of these properties is the decodability of input patterns based on network outputs.

But how are all of these properties related? Do networks ever exhibit all of them simultaneously, and if so, when? Developing the complete picture is a formidable challenge; in this chapter, we progress by answering these questions in what is probably the most tractable class of systems in which they can be studied. These are noisy, feedforward networks of thresholding neurons [107] .

Several prior studies of signal propagation in feedforward networks inform our approach. These suggest that a wide range of network responses fails to occur in broad parameter regimes: rather, the only outputs produced are all cells firing or being silent simultaneously. This is due to the buildup of correlations among neural activity at each layer, even when inputs drive the cells to fire independently in the first layer. In particular, for iteratively constructed *in vitro* feedforward networks, neurons displayed a marked tendency to synchronize [120]. Subsequent simulations and analyses with thresholding neurons have corroborated these findings, arguing that any initial spike count distribution

becomes strongly bimodal after a few layers [107]. Integrate-and-fire neurons similarly fail to transmit activity without decaying or saturating to a point independent of the input rate, but are able to support stable propagation of highly-synchronized volleys of spikes even in the presence of asynchronous background noise with low firing rates [42, 82, 87, 140]. These cascades of coherent activity, called *synfire chains*, have been observed experimentally and are hypothesized to encode stimuli through precise spatiotemporal spike timing [3]. However, in embedded feedforward networks, even synchronous propagation dies out if background network activity is sufficiently synchronous or has high firing rate [82]. If synapses are strengthened and connectivity sparse, rate propagation but not synfire activity can be sustained [148]. Moreover, different studies demonstrate a "critical" regime with broadly distributed output patterns and significant higher-order interactions [20, 153].

As we further explore here, the key difference among these studies turns out to be the *threshold* number of excitatory inputs that each cell must receive in order to fire [83]. This threshold is low (a single spike) in the work of [20] and [148] but much higher for [107], [120], and [87]. As reviewed in [83], densely connected feedforward networks with synapses that are weak relative to threshold tend to produce more synchrony than their sparsely connected counterparts, due to the neurons having more shared inputs [122]. Thus, as synaptic inputs are weak compared to spike thresholds in many biological settings, it may appear that synchrony is inevitable. However, noise "local" to each neuron decreases synchrony, and can do so without damaging the capacity to transmit signals, at least those defined by firing rates within each network layer ([147], but see also [107, 120]).

Here, we undertake to unify these results through a common mathematical framework, and extend them by treating multiple properties of network outputs. In particular, we show when and how neutral stability, broad response distributions, higher-order correlations, and the transmission of firing rate signals coexist and when these properties fail to coexist. For any level of spike threshold, we find that there is always an intermediate value of network connectivity characterized by neutral stability and higher order correlations. High response entropy and transmission of firing rates, however, only occur at this point when neurons have low thresholds or added noise.

2.1 Model of stochastic feedforward networks

Network: Following [107], we examine a network of binary [97] neurons in a feedforward architecture with probabilistic synapses and input (see Figure 2.1A for a schematic). Each layer consists of N identical neurons. In general, we will illustrate $N = 20$; results hold for larger N as well, as we summarize in the Discussion. The neurons are thresholding units that spike if they receive at least θ synaptic inputs from neurons in the previous layer and are otherwise quiescent (i.e., silent). The connectivity structure between layers is random and spatially homogenous; each neuron upstream is connected to C postsynaptic neurons chosen uniformly from the downstream layer. Connections between neurons have a fixed probability p of synaptic transmission. In this chapter we are interested in signal propa-

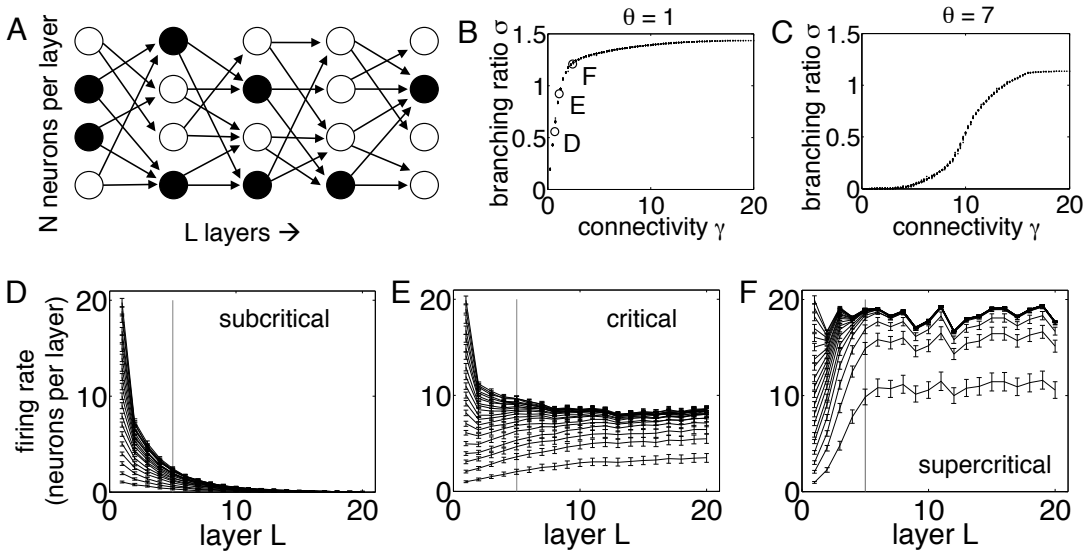

Figure 2.1: Average rate transmission in feedforward networks. (A) A schematic of a feedforward network. Filled circles indicate spikes, hollow circles quiescence (i.e., absence of firing). In this example, $N = 4$, $L = 5$, $C = 2$, $p = 0.5$, and $\theta = 1$. (B, C) Branching ratio σ as a function of connectivity strength γ for $N = 20$, (B) $\theta = 1$ and (C) $\theta = 7$. Each data point is the branching ratio of a network of a particular connectivity structure. (D - F) Simulated propagation of firing rates shown for three sample networks with $\theta = 1$ and $C = 3$, $p = 0.25, 0.43, 0.85$, respectively. These parameters are also indicated by the markers in (B). Noisy, uncorrelated input is injected into the first layer, and the resulting firing rates are averaged over 1000 trials plotted over multiple layers. Error bars indicate standard deviation scaled by a factor of 1/10 to facilitate comparison. Vertical grey bars are shown at $L = 5$ to emphasize that henceforth we will primarily be concerned with shallow layers. (D) In subcritical networks ($\sigma < 1$), activity tends to die out. (F) In supercritical ($\sigma > 1$) networks, activity saturates. (E) Critical networks ($\sigma \approx 1$) reveal greatest fidelity in propagating Poisson input rates through layers; however, while this picture is qualitatively true for networks of low-threshold neurons, when θ reaches higher values, networks tend to transmit only high or low rates (see Section 2.5).

gation through a fixed number of layers, as we expect biological feedforward networks to be shallow (we will in general take $L = 5$). The number of layers through which a rate signal can be transmitted in similar feedforward networks before losing information has been addressed by [142, 86, 51].

We concentrate solely on networks with a feedforward connectivity structure. However, these networks are equivalent to synchronously-updated, discrete-time systems with random recurrent connections (including "autapse" connections to from cells to themselves), under the annealing approximation [23]. Thus, to the extent that these assumptions hold, the results of this work may also be applied to the evaluation of persistent activity in recurrent networks.

Stimuli: The networks are driven by a stimulus to elicit an average spike count of $S \in \{0, \dots N\}$ firing neurons in the first layer at that time step. Unless otherwise specified, this stochastic input is injected *independently* so that each neuron in the first layer responds as an independent (0,1) Bernoulli random variable with biased probability S/N of spiking (taking value 1). This results in a binomially distributed spike count in the first layer.

Propagation: The state of the Lth layer is denoted by $\mathbf{x}_L$, an N-vector of zeros and ones, and the connectivity matrix between layers L and $L+1$ by E_L. (Henceforth we will use E to refer to the $N \times N \times L-1$ connectivity matrix of the entire network.) Since the connections between neurons are stochastic, in a given trial each synapse fails with probability $1 - p$. It will be useful to call the realization of E_L according to the probability of synaptic transmission the "effective" connectivity matrix $\hat{E}_L$, keeping in mind that different trials will yield different $\hat{E}_L$ yet E_L will remain fixed. The state at layer L of a realization of a given network can now be expressed as

$$\mathbf{x}_{L+1} = \Theta(\hat{E}_L \mathbf{x}_L - \theta), \tag{2.1.1}$$

where Θ is the elementwise Heaviside step function. The key parameter in this system is the connectivity strength $\gamma = Cp$.

Limitations and simplifications: We note several important facts about the model setup and analysis. First, this model has no time in its dynamics; each trial can be thought of as a wave of activity evolving from a particular initialization in the first layer, and is independent of the next. Because of this, the phenomenon of synchrony in the usual temporal sense is not applicable. The corresponding concept of synchrony is when neurons in a layer tend to fire, or be quiescent, together in a given trial; this is what we will mean in the remainder of this chapter when we refer to synchrony or synchronous coding. Second, because of the assumption of spatial homogeneity both in inputs and in network architecture, this model is not well-suited to study spatial modes of activity.

Third, and most importantly, our analysis henceforth focuses on the total activity within each layer. That is, rather than quantify network responses in the full space of 2^N firing patterns that can occur in each layer, we restrict our description to the *number* of cells that fire in that layer: the $N + 1$ different values of the (layer-averaged) firing rate.

2.2 The branching ratio

To understand the qualitative dynamics and average firing rate transmission through multiple layers, we borrow a useful tool from the criticality literature [20]. A "critical" transition regime is often experimentally defined via the *branching ratio* σ, the ratio between the number of cells in a population firing at a particular time step and the number of cells firing at the previous timestep, averaged over time. To avoid decay or growth of activity, the system must produce firing rate dynamics which are neutrally stable, satisfying $\sigma \approx 1$; such networks are labeled critical.

In our feedforward framework, the relevant measurement is the branching ratio averaged over trials and layers rather than time. To quantify the general capacity of a particular network with fixed connectivity structure E to maintain activity in a one-to-one manner, we will also average this layerwise branching ratio over repeated trials with the same network, each with different stimulus rates as well as different (random) inputs $\mathbf{x}_1$ to the first layer. The result is:

$$\sigma = \left\langle \left\langle \left\langle \frac{S_L}{S_{L-1}} \right\rangle_L \right\rangle_{\hat{E},\mathbf{x}_1} \right\rangle_S , \tag{2.2.1}$$

where S_L is the number of neurons spiking in layer L on a given trial. Throughout this chapter, when we refer to the branching ratio we will mean σ.

We conducted Monte Carlo simulations to compute how this quantity changes with connectivity level γ. In detail, at a fixed γ, we first chose one example of a network structure E for every $C \geq \lceil \gamma \rceil$, the constrained value ensuring that $p < 1$. For each E, we then input a deterministic rate of exactly $S = \theta + 1, \ldots, N$ spiking neurons in the first layer with 100 random instantiations of $\mathbf{x}_1$, evolve the network, and measure the ratios S_L/S_{L-1} for each layer until either the neural activity dies out or until the last layer is reached. Finally, σ is computed as the average over the 100 random network realizations and instantiations at the first layer, and subsequently over all stimulus levels greater than θ spiking neurons per layer (as any input less than that is guaranteed to be extinguished at the next layer.)

Figure 2.1BC shows results over five layers. Each of the tight scatter of points at each value of connectivity γ is the branching ratio of a particular network with that value of connectivity and a specific architecture E. (The fact that there is very little variation at a given level γ supports our choice of this combined parameter as the principal one in our study.[1]) Note that as we sweep connectivity γ from small to large values, we pass through a critical value (which we shall denote by γ_{obs}) at which $\sigma \approx 1$. Thus, we find that the transition (critical) branching parameter is consistently found in our networks at some intermediate connectivity level.

We next illustrate the implications of the branching parameter for propagation of firing rates across network layers. For many different networks, we compute rate trajectories averaged over 1000 trials for input rates ranging from 0 to $N = 20$ neurons firing in the first layer. In each trial, the input rate S and E are fixed, yet $\mathbf{x}_1$ and $\hat{E}$ change probabilistically. The evolution of the firing rate over 20 layers is shown for three representative networks with threshold $\theta = 1$ in Figure 2.1D-F. In subcritical networks (Figure 2.1D) neural activity dies after a few layers regardless of stimulus. The supercritical, i.e. $\sigma > 1$, network (Figure 2.1E) inflates rates to nearly maximal values, and as in the subcritical case it is difficult to distinguish between two inputs based on output rate alone. In critical networks, however, rate trajectories remain separated at each layer (Figure 2.1F). This result is in

[1] In more detail: in following sections we reduce the two parameters C and p dictating network connectivity to the single connectivity parameter γ; this is supported by the observation that variation in σ for fixed γ but varying values of C and p is has a negligible impact on the branching ratio, as shown by the tight scatter of points at each γ in Figure 2.1BC. Moreover, in the mean-field theory we develop, it is also the case that γ, rather than C and p separately, enters.

agreement with other findings in the literature regarding information transmission of critical networks. Overall, these simulations confirm the expected picture that the average firing rate statistic is best propagated through networks when $\sigma \approx 1$.

Beyond preservation of firing rates from one layer to the next, we are interested in networks that can produce a broad distribution of responses, and avoiding the limitations of strong synchrony. To assess this, in the next section we introduce a tool to describe propagation of firing rate distributions across layers via a mean-field approximation.

2.3 Mean-field Markov chain model and spectral analysis

Since the state of each layer depends solely on the state of the previous layer and the synaptic connections between layers, our feedforward networks are Markov chains [107]. Furthermore, as we aim to describe only the propagation of layer-averaged firing rates, rather than particular firing patterns (or binary "words"), our Markov chain has $N + 1$ states. We proceed to derive a mean-field description of the Markov chain for each connectivity level γ by averaging over possible connection matrices, yielding a discrete-time dynamical system describing the evolution of the spike counts or firing rate from layer to layer. To do this, we make two assumptions: conditional independence and permutation symmetry. The former assumes that neurons in the downstream layer are conditionally independent given the number of neurons firing in the previous layer. This condition effectively ignores correlations that would arise due to inhomogeneous connectivity architectures, e.g., divergent motifs. However, conditional independence neither implies nor is implied by unconditional independence; therefore, activity between neurons may still be correlated ultimately.

The second assumption states that neurons within a layer are permutation symmetric, meaning that each group of k neurons in a layer have the same statistical properties. Both approximations (and therefore the mean-field model) are exact in the special case of all-to-all connectivity ($C = N$), for which [107] developed the same description. When $C < N$ the mean-field and true models differ; the latter would be obtained by first finding the spike count distributions for fixed connectivity and then averaging over all E. However, for the excitatory networks considered in the main part of the chapter (Sections 2.1-2.5), we have verified numerically that the mean-field model is a good predictor of the true spike count dynamics except in the sparse limit of large p and small C (see Appendix).

The mean-field transition matrix A – i.e., the matrix whose (n, m)th element is the probability that there are $S_L = m$ neurons spiking at a layer given $S_{L-1} = n$ spiking in the previous layer – is given by:

$$A_{nm} = \binom{N}{m} q_n^m (1 - q_n)^{N-m} \tag{2.3.1}$$

if $n \geq \theta$. Here, q_n is the probability that a downstream neuron will fire assuming n spiking

neurons in the previous layer:

$$q_n = \sum_{k=\theta}^{n} \binom{n}{k} \left(\frac{\gamma}{N}\right)^k \left(1 - \frac{\gamma}{N}\right)^{n-k}. \tag{2.3.2}$$

If $n < \theta$, then $q_n < 0$ and $A_{nm} = 0$. In principle, A could be derived from the transition matrix of the original Markov chain (see Appendix for details).

Using the transition matrix, the spike count probability distribution P_L at layer L (the vector of length $N + 1$ whose jth component is the probability that j neurons are firing in layer L) is simply given by matrix-vector multiplication: $P_L = P_{L-1}A$.

The long-term behavior of these feedforward networks can be predicted using the eigenvalues and eigenvectors of the mean-field transition matrix. To illustrate this, assume A is diagonalizable, so that the input probability distribution $P_{\text{input}} = P_1$ can be decomposed into a linear combination of the eigenvectors of A:

$$P_{\text{input}} = \alpha_0 v_0 + \alpha_1 v_1 + \cdots + \alpha_N v_N. \tag{2.3.3}$$

The spike count probability distribution at the Lth layer is simply $P_L = P_{\text{input}} A^L$:

$$P_L = \alpha_0 \lambda^L v_0 + \alpha_1 \lambda_1^L v_1 + \cdots + \alpha_N \lambda_N^L v_N. \tag{2.3.4}$$

The persistence of different eigenmodes through layers is determined by the size of their corresponding eigenvalues.[2] If $\lambda_i \ll 1$ then after a few layers the contribution ith eigenmode will decay to become negligible. On the other hand, eigenmodes whose eigenvalues are near 1 will survive through many layers.

We analyze the eigenstructure of A through a combination of mathematical analysis and computational argument. First, it can be proven that A has one unique stationary state corresponding to all neurons being quiescent: $v_{\text{off}} = (1, 0, \ldots, 0)$ (Proposition 1 in Appendix). The existence of a stable quiescent state is expected, as any nonzero probability of synaptic failure will eventually attenuate any signal. Second, if A is well-behaved in the sense that its eigenvectors have limits as $\gamma \to N$ (an assumption that is supported by numerics, see Figure 2.2AC), then the second largest eigenvalue λ^* of A converges to 1 as $\gamma \to N$, indicating the emergence of an additional dimension of persistent activity. This, too, is intuitive, as a maximal connectivity strength should result in a stationary state in which all neurons fire at each layer. The catch, however, is that the corresponding eigenvector v^* converges to a vector in the subspace of bimodal or synchronous distributions – that is, to the span of v_{off} and v_{on} where $v_{\text{on}} = (0, ..., 0, 1)$ corresponds to all neurons firing (Proposition 2 in

[2] In Markov chains with very non-normal transition matrices, transient activity can persist past that expected solely by the spectrum; these matrices can be analyzed through their prominent pseudospectral sets, which are the eigenvalues of small perturbations of the matrix. When we pursued this type of analysis of A, we did not find significant pseudospectral sets that described the persistent activity of our networks beyond expectations from the spectral analysis (results not shown; see [143] for more details on pseudospectral analysis).

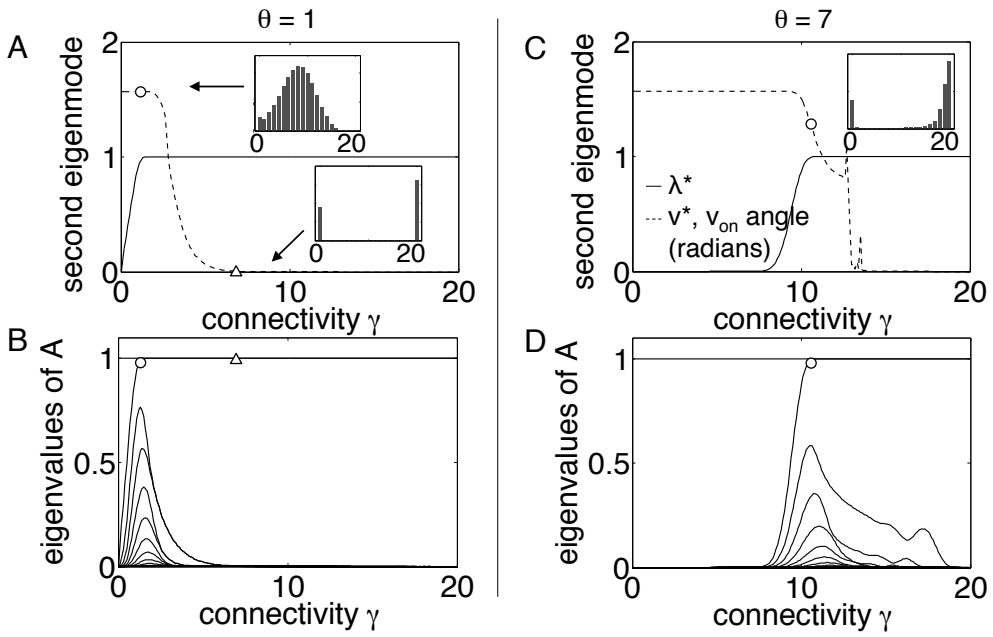

Figure 2.2: Spectral analysis of the mean transition matrix for networks with (A, B) $\theta = 1$ and (C, D) $\theta = 7$. (A, C) The second largest eigenvalue λ^* (solid line) effectively converges to 1 while the angle between v^* and the vector corresponding to full synchrony v_{on} (dashed line, plotted in radians) maintains significant value for a range of γ, indicating that the second eigenmode is both persistent and far from bimodal. This is also illustrated by the insets, which show typical histograms on the line quasi-attractor either at γ_{eig} (circle markers on dashed line in A, C and on the second dominant eigenvalue in B, D) or when γ is too high to support broadly distributed eigenmodes, resulting in bimodal distributions (triangular marker in A, B). (B, D) Also at the emergence of the line quasi-attractor (circle markers), all eigenvalues are near-maximal compared to their values over the entire connectivity range.

Appendix). All other eigenmodes must converge to 0 as $\gamma \to N$. So despite the emergence of this extra persistent dimension, activity becomes synchronous as connectivity strength increases. Ideally, what we want is for λ^* to be practically 1 yet for the span of v^* and v_{off} to be far from the plane of bimodal distributions.

Intriguingly, numerical calculations reveal that this does occur for an intermediate level of connectivity γ_{eig} (Figure 2.2AC), implying the emergence of a plane spanned by the two principal eigenmodes v^* and v_{off} that, due to increased persistence, effectively acts as an attractor in sufficiently shallow layers: because of this, we will call this plane *quasi-attracting*. Once the vectors are normalized to represent probability distributions, this means that at γ_{eig}, there exists a line quasi-attractor that is far from the space of bimodal distributions, and hence that the network can support broadly distributed, incompletely synchronized firing states. At this same intermediate γ_{eig}, we also observe significant values of all eigenmodes (Figure 2.2BD), showing further persistent activity contributed by other

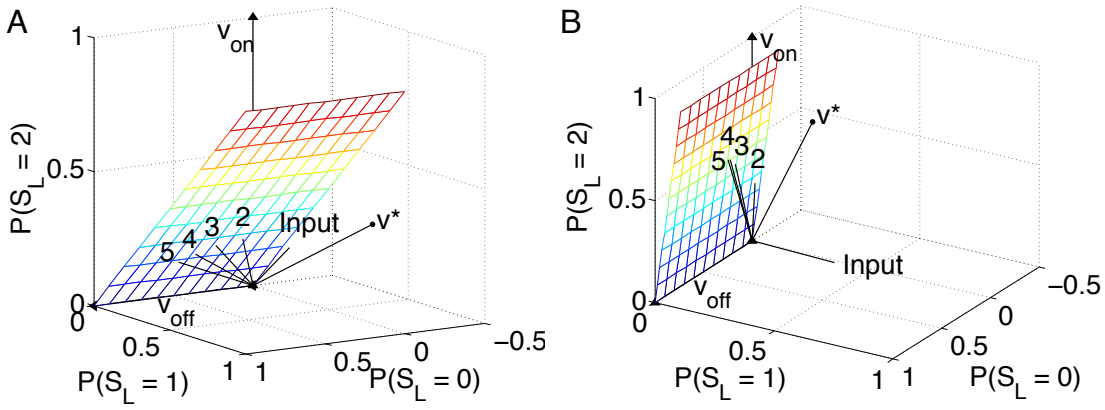

Figure 2.3: Geometrical interpretation of spike count propagation for $N = 2$, $\theta = 1$. The spike count histogram evolves through layers according to the mean-field model via iterated matrix multiplication, corresponding to a discrete trajectory in the space of spike count probability distributions. (A) For low connectivity strengths, the input distribution quickly converges to the plane spanned by the first two eigenmodes (the gray plane shown, although histograms will be constrained to the line embedded in the plane satisfying $\sum_i P_i = 1$) as the network encodes the signal. The distribution then slowly converges to the true stationary state v_{off}, and the signal decays. In this example convergence to quiescence occurs in only a few layers. (B) For high connectivity levels, activity persists through deeper layers, but the line quasi-attractor has rotated closer to the space of bimodal distributions spanned by v_{on} and v_{off}. The ideal network lies between these two figures. See text for a more detailed discussion.

eigenmodes, at least for the initial network layers.

We pause to give a geometrical view of the mean-field dynamics described above. This is illustrated in Figure 2.3 for $N = 2$, although the following description holds for arbitrary N. Consider the $(N + 1)$-dimensional space of the spike count probability distribution at a layer. Starting with any input probability vector P_{input}, the layer-to-layer mean-field dynamics of the network can be visualized as iterated mappings of the input vector P_{input} in the space of spike-count distributions, constrained to the simplex $\sum_i |P_i| = 1$. In Figure 2.3, repeated applications of A are enumerated. In the first couple iterations, the spike count distribution converges towards the line quasi-attractor spanned by v_{off} and v^* as smaller eigenmodes decay. This may be interpreted as the encoding of the input distribution in the lower-dimensional quasi-attractor after a few layers. Eventually, the system drifts to the stationary state where all neurons are silent, v_{off}. This of course represents a final state in which the network has "forgotten" the input. If $\gamma < \gamma_{\text{eig}}$, then the convergence to v_{off} happens within a few layers (as in Figure 2.3A). When $\gamma > \gamma_{\text{eig}}$, although activity persists through many layers as expected, the line quasi-attractor has rotated nearer to the span of v_{on} and v_{off}, so that the persistent activity is nearly synchronous (see Figure 2.3B). It is only when $\gamma \approx \gamma_{\text{eig}}$ that activity is persistent while resisting synchrony. In this sense, γ_{eig} represents the existence of a persistent mode of activity characterized by a balance of

principal eigenmodes that are broadly distributed, avoiding firing patterns being limited to synchrony or quiescence. In fact, as we will explore in the following section, $\gamma \approx \gamma_{\text{eig}}$ also predicts further interesting statistical features of network responses.

2.4 Complexity of network responses

Next we study two measures of the statistical complexity of the network responses over the range of connectivity strengths and threshold levels. First, we take an information theoretic approach to the question of *higher-order interactions*, which asks: when do neurons fire in a way that cannot be predicted from their firing rates and pairwise spike correlations alone? Beyond their basic role in characterizing the degree of coordinated spiking in networks [129, 124, 92, 137], higher-order statistical interactions have been shown to be necessary to produce broad distributions spiking activity [10] (for recent applications, see [90, 153]), and to significantly impact coding of stimuli [52, 100].

To calculate the extent of higher-order moments in the response distributions, we utilize maximum entropy models [129, 124, 109, 153, 72]. The pairwise maximum entropy fit of a probability distribution is defined as the distribution that has maximal entropy while being constrained to match the first and second moments of the true distribution. Thus, this fit makes the fewest additional assumptions on the structure of the probability distribution – any additional structure is attributed to higher-order moments. For the permutation-symmetric networks at hand, the pairwise maximum entropy distribution takes the form of a one-dimensional discretized Gaussian:

$$P^{\text{ME}}(n) = \frac{1}{Z}\exp\{\lambda_1 n + \lambda_2 n^2\} \tag{2.4.1}$$

where Z is a normalizing factor and λ_1, λ_2 are the parameters chosen to match the first two moments. As a measure of the impact that higher-order correlations have in shaping the response distributions, we follow [129, 124, 109, 153] and compute the stimulus-averaged Jensen-Shannon (JS) divergence between the true distribution and its maximum entropy fit:

$$\begin{aligned} D_{JS}(P_L, P_L^{ME}) =& \frac{1}{2}\sum_{m=0}^{N} P_L(m)\log_2\left(\frac{P_L(m)}{\frac{1}{2}(P_L(m)+P_L^{ME}(m))}\right) \\ &+ \frac{1}{2}\sum_{m=0}^{N} P_L^{ME}(m)\log_2\left(\frac{P_L^{ME}(m)}{\frac{1}{2}(P_L(m)+P_L^{ME}(m)}\right). \end{aligned} \tag{2.4.2}$$

This quantity assumes values $0 \leq D_{JS}(P_L, P_L^{ME}) \leq 1$; it can be thought of as a symmetrized version of the Kullback-Leibler divergence.

This comparison of the spike count distribution with the maximum entropy model will describe how well (over all possible networks of that particular gamma) the first two moments can describe the responses. This is a complimentary view from directly calculating

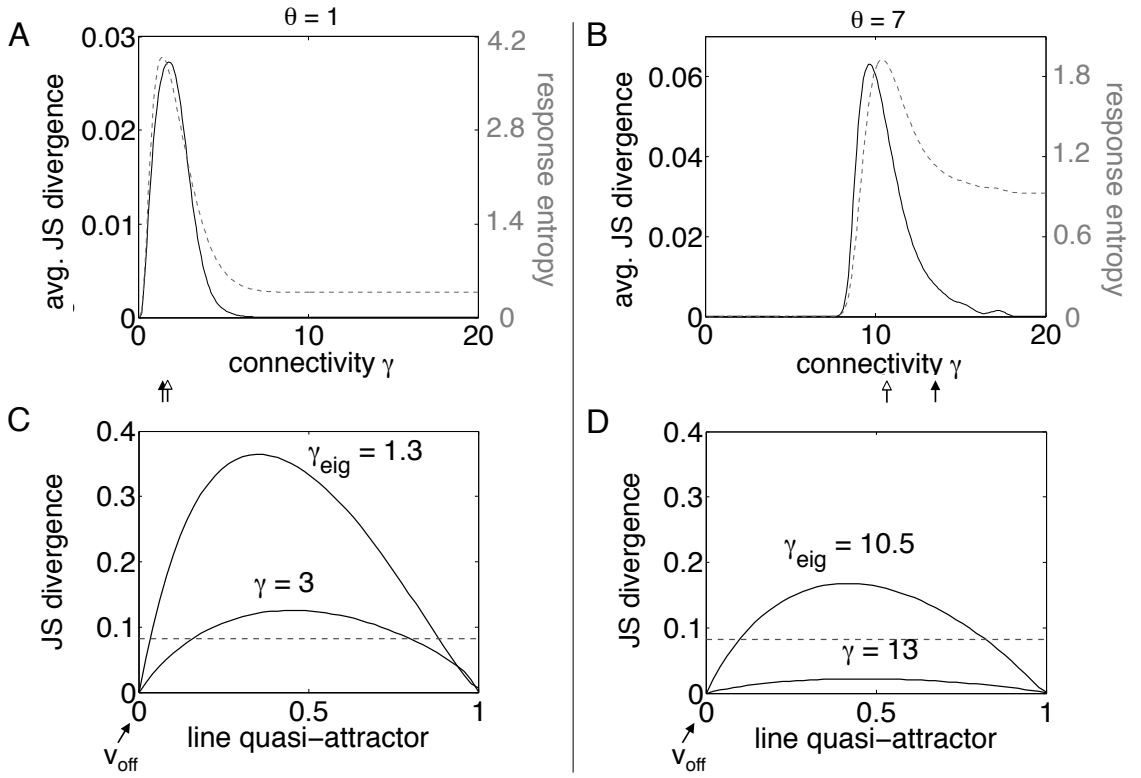

Figure 2.4: Statistical complexity of network responses for (A, B) $\theta = 1$ and (C, D) $\theta = 7$. (A, C) Response entropy (dashed grey line), and the stimulus-averaged JS divergence between the true distribution at layer 5 and the pairwise maximum entropy fit (solid black line) plotted as a function of γ. Also shown are γ_{eig} (hollow arrow below panel) and γ_{obs} (solid arrow). (A) When $\theta = 1$, peaks in both curves line up with γ_{eig}, as does γ_{obs} (arrows offset for visibility). (B) For higher threshold networks, γ_{obs} doesn't align with γ_{eig} or other assays (see Section 2.5). (C, D) JS divergence between spike count histograms and their maximum entropy fits along the line quasi-attractor (solid lines; axis parameterizes distributions along the line quasi-attractor starting at v_{off}). Compare with the divergence averaged over the entire space of histograms (dashed lines).

higher-order cumulants [136], directly quantifying how distinct the full distribution is from one with the minimal possible statistical assumptions. In the special case of fully connected layers ($C = N$), the mean-field description is exact, and hence this will correspond to the tendency of groups of any group of $k > 2$ neurons to fire in the same bin (since the network is homogenous). For $C < N$, this quantity approximates the amount of coincident spiking averaged over groups of k neurons.

Recall that each neuron in the first layer is independently stimulated so that their firing is a Bernoulli trial, so no correlations are injected into the network. All correlations, pairwise and higher-order alike, emerge solely from the network interactions. We computed the JS divergence between spike count distributions distributions at layer 5 and their pairwise

maximum entropy conditioned on input rate, then average this over all possible stimuli. Through this assessment, we note significant complexity already by the fifth layer at γ_{eig} (Figure 2.4AB, solid line). Beyond γ_{eig}, as firing rates approach maximal values, the JS divergence decreases to zero. This is because the JS divergence infers the effects of the higher-order moments in the response distribution *beyond* that which would be expected from the first two moments alone, so high firing rates alone don't necessarily imply large JS divergence (e.g., if neurons are independent with maximal firing rate they will fire together frequently but exhibit no correlation).

How can we understand the origin of such statistical complexity arises? We next show that they can be predicted from the spectral analysis of the previous section, without the need for simulation. Figure 2.4CD plots the JS divergence between the spike count histograms on the line quasi-attractor and their pairwise maximum entropy fit. Here, we plot this quantity as a function of their position along the line, parameterized so that v_{off} is at position 0. This can be compared to an average JS divergence of approximately 0.08 (dashed lines; calculated by averaging over 10,000 random sample distributions so that the mean had converged) over the entire space of possible response histograms. In particular, the eigenvectors at γ_{eig} produce significantly larger statistical complexity than the average. This is because at this level of connectivity, the response distribution is a mixture of two distributions: a large component of quiescent neurons corresponding to v_{off}, and a broader component corresponding to the contribution of v^*. As θ increases, the level of the JS divergence decreases on the line quasi-attractor, as higher thresholds reduce the breadth of v^*.

The second statistical feature of note is the response entropy of the spike count distribution:

$$H(P(S_L)) = \sum_{n=0}^{N} P(S_L = n) \log_2 P(S_L = n). \tag{2.4.3}$$

Larger response entropies indicate broader response distributions. The response entropy at the 5th layer peaks at γ_{eig}, indicating that the emergence of the line quasi-attractor corresponds to the broadest distribution of activity across all values of γ (Figure 2.4, dashed grey line). However, the peak response entropy decreases for higher values of θ; this is the result of the fact that as θ increases, v^* produces less broad response distributions due to the high threshold and hence the silencing of weak inputs, preventing them from eliciting any firing in the subsequent layer.

In sum, we have shown that at γ_{eig}, networks display an emergence of statistical complexity – both through maximal response entropy and significant contributions from higher-order moments within responses – directly because of the contribution of other eigenmodes at that level of connectivity.

2.5 Combining neutral stability and broad response distributions

In order to maintain averaged levels of activity without succumbing to synchrony, a network must simultaneously satisfy two criteria. The first is that it must be able to preserve averaged firing rates from layer to layer without succumbing either to runaway excitation and maximal firing rates in deeper layers, or to decaying network activity. Second, a network must exhibit a broad spike count distribution at each layer in order to prevent the buildup of correlations and synchrony [83, 120, 87]. We refer to these properties, taken together, as *asynchronous rate coding*.

For which parameter regimes can such asynchronous rate coding occur? To quantify this we need an assay that captures how well networks are able to propagate broad response distributions from one layer to the next. We base this on the propagation of binomial spike count distributions, as these correspond to fully independent activity in each layer. Specifically, we define the spike count JS divergence D to be the JS divergence (Equation 2.5) between the binomial input distribution P_1 and the Lth layer spike count distribution P_L averaged over all stimuli S:

$$D(\gamma, \theta) = \frac{1}{N+1} \sum_{n=0}^{N} D_{JS}(P(S_L|S = n), P(S_1|S = n)). \tag{2.5.1}$$

Networks will exhibit good performance, measured by low values D, when they maintain the broad (independent) spike count distribution and the averaged firing rate that occurs in the first layer.

Plots of the spike count JS divergence over γ are shown in Figure 2.5 for increasing values of the spike threshold θ. For each fixed θ, there is an optimal, intermediate value of γ at which networks are best able to satisfy both of our criteria. However, as threshold level increases, the best value of the spike count JS divergence also increases, showing that high-threshold networks fail to produce asynchronous rate coding.

This failure follows from the requirements of neutral dynamics and broad response distributions described in previous sections. First, from Section 2.2, γ_{obs} captures the first criterion of complex signal coding outlined above: that is, networks demonstrate neutral stability and average one-to-one rate transmission when they average a branching ratio of $\sigma \approx 1$. On the other hand, Section 2.3 shows that γ_{eig} reflects when the network supports persistent, broad response distributions, providing an assay of the second criterion. Complex signal propagation can therefore only occur in these systems when $\gamma_{\text{obs}} \approx \gamma_{\text{eig}}$. Comparing Figure 2.2 with the previous Monte Carlo simulations in Figure 2.1BC reveals both criteria can indeed be simultaneously satisfied when few inputs are required to cause a spike, but a gap between these required values of connectivity γ appears with increasing θ. To be precise, for $N = 20$, $\theta = 1$, Monte Carlo simulations and spectral analysis both yield $\gamma_{\text{obs}} \approx \gamma_{\text{eig}} \approx 1.3$. When $\theta = 7$, however, simulations show $\gamma_{\text{obs}} \approx 13.75$ yet $\gamma_{\text{eig}} \approx 10.5$. As shown through the eigenstructure, at γ_{obs}, only bimodal activity is supported after a few

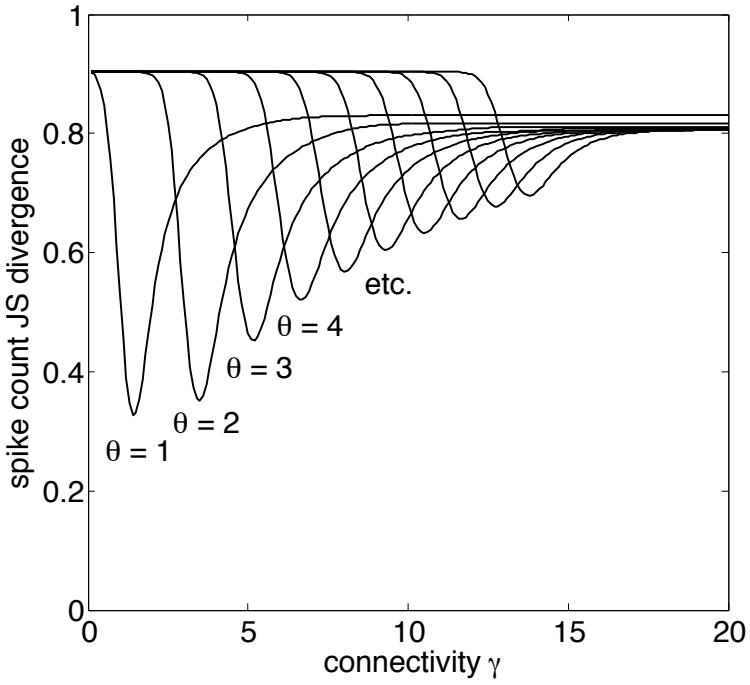

Figure 2.5: Spike count JS divergence plotted as a function of γ for increasing θ. Optimal performance for each threshold (the minimum value of the curve) occurs near γ_{eig}.

layers. In fact, because of the inevitable synchrony in deep layers, optimal performance under the JS divergence tends to fall nearer to γ_{eig} than to γ_{obs}. Networks of high-threshold neurons are therefore unable to simultaneously satisfy both requirements of complex signal propagation outlined at the beginning of this section.

Intuitively, the reason that $\gamma_{\text{obs}} > \gamma_{\text{eig}}$ is that networks with high-threshold neurons reject inputs of low firing rate, so that when θ is large there is an increased likelihood that connectivity structure and stochasticity will conspire to silence all activity in the next layer. Geometrically speaking, as θ increases so does the nullity of A, resulting in a larger and larger subspace that trajectories must avoid lest they risk susceptibility to network quiescence; in order to reach an average of one-to-one rate transmission, it is necessary to provide a buffer for the coding subspace from the nullspace by inflating the connectivity into the regime of bimodality.

Also of practical importance is the question of robustness to parameters. The delicate nature of γ_{eig} and γ_{obs} constrains networks that produce asynchronous rate coding to finely-tuned connectivity strengths; one requires that the branching ratio lie at a critical value $\sigma \approx 1$, while the other relies on a precise balance between persistent yet broadly supported eigenmodes. This sensitivity is reflected in the sharp troughs in the JS divergence (Figure 2.5); for even higher N, these troughs become even sharper and robustness is a more important goal to obtain. As we will see in the next section, however, this sensitivity can

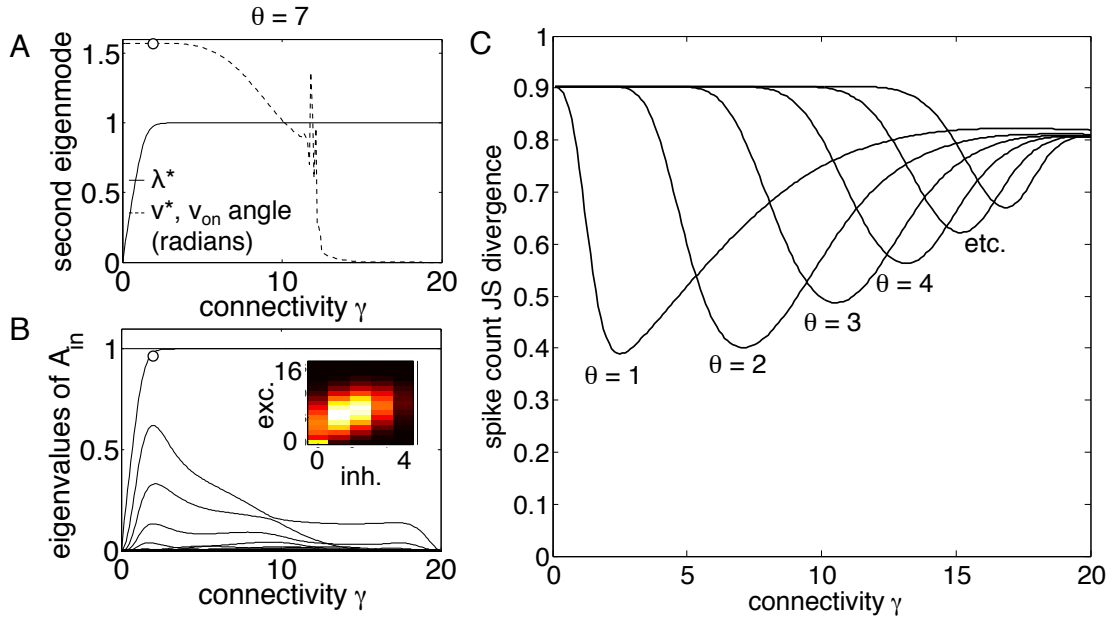

Figure 2.6: Excitatory-inhibitory networks display increased robustness, $N_E = 16$ and $N_I = 4$. (A) The second largest eigenvalue λ^* and the angle between v^* and v_{on} (dashed line) overlap over a larger parameter space, indicating robustness at γ_{eig}; (B) similarly, all eigenvalues A of have broader peaks. Inset shows a typical broadly distributed histogram at γ_{eig} (indicated by the marker in A, B). (C) The spike count JS divergence has a wider minimum for all values of θ, showing that inhbition also allows for more robust asynchronous rate propagation.

be mitigated by adding an inhibitory population to each layer.

To summarize results thus far, we evaluate networks on two criteria: $\sigma \approx 1$ and broad response distributions. Low-threshold networks can always satisfy broad response distributions and maintained average rate transmission at the same γ. High-threshold networks are able to somewhat support broad distributions, although the preserved aspects of network responses and their lower values of response entropy indicate less broad distributions as compared to their low-threshold counterparts. They also can satisfy $\sigma \approx 1$, however this is due to averaging: because of the increasing nullity of the mean transition matrix, these networks cannot propagate weak input stimuli, so they must overcompensate by inflating γ. Because of this, no high-threshold network of a fixed γ can simultaneously satisfy both criteria, and hence they cannot propagate rates asynchronously through layers. This appears to be a significant limitation for high-threshold networks – and, importantly, for many biological neural networks, in which many inputs are required to elicit a spike. In the following sections we will incorporate additional biophysical features – inhibition and noise – and study whether this provides a resolution so that high-threshold networks can support persistent, broadly distributed activity.

2.6 Excitatory-inhibitory networks display increased robustness

How can asynchronous rate propagation emerge in high-threshold networks? Intuitively, we might expect an added inhibitory population to prevent runaway excitation and saturation of firing rates to high values, thus preventing synchrony. To test this, we added an inhibitory population of N_I neurons to each layer of $N_E = N - N_I$ excitatory neurons, and further impose $N_E - N_I > \theta$ (otherwise no activity could be transmitted due to the homogeneity in network connectivity – even if only the excitatory population is active in layer L, the random connectivity imposed will cause the same proportion of the excitatory and inhibitory populations in layer $L+1$ to fire). Network parameters are assumed to be homogenous among the inhibitory and excitatory populations. Because of this assumption, it is straightforward to calculate the new, four-dimensional mean-field transition matrix A_{in}:

$$P(S_L^E = m_E, S_L^I = m_I | S_{l-1}^E = n_E, S_{l-1}^I = n_I) = \binom{N_E}{m_E} q_{n_E,n_I}^{m_E} (1 - q_{n_E,n_I})^{N_E - m_E} \times \binom{N_I}{m_I} q_{n_E,n_I}^{m_I} (1 - q_{n_E,n_I})^{N_I - m_I}, \tag{2.6.1}$$

where S_L^i is the number of cells spiking in the excitatory ($i = E$) or inhibitory ($i = I$) population at the Lth layer, and q_{n_E,n_I} is the probability that a downstream neuron spikes given n_E spiking excitatory neurons and n_I spiking inhibitory neurons in the upstream layer:

$$q_{n_E,n_I} = \sum_{k_E=\theta}^{n_E} \sum_{k_I=0}^{\min(n_I, k_E-\theta)} \binom{n_E}{k_E} \left(\frac{\gamma}{N}\right)^{k_E} \left(1 - \frac{\gamma}{N}\right)^{n_E - k_E} \times \binom{n_I}{k_I} \left(\frac{\gamma}{N}\right)^{k_I} \left(1 - \frac{\gamma}{N}\right)^{n_I - k_I}. \tag{2.6.2}$$

The binomial input distributions now take the following form:

$$P(S_1^E = m_E, S_1^I = m_I | S = n) = \binom{N_E}{m_E} \left(\frac{n}{N}\right)^{m_E} \left(1 - \frac{n}{N}\right)^{N_E - m_E} \times \binom{N_I}{m_I} \left(\frac{n}{N}\right)^{m_I} \left(1 - \frac{n}{N}\right)^{N_I - m_I}. \tag{2.6.3}$$

The expression for the transition matrix for the excitatory-inhibitory networks has a similar form to that of the purely excitatory networks, so the eigenstructure of A_{in} (Equation 2.6.1) is similar to that of A (Equation 2.3.1): it has a unique stationary state corresponding to all neurons being quiescent, and as $\gamma \to N$, the second largest eigenvalue converges to 1 and its eigenvector corresponds to bimodality (Proposition 3 in Appendix). There also is an intermediate state of connectivity at which $\lambda^* \approx 1$ and v^* is far from bimodal (Figure 2.6AB). Here we consider $\theta = 7$, as well as $N_E = 16$, $N_I = 4$ to simulate ~20% inhibition, as typically used in, for example, cortical modeling (cf. [25]).[3] This

[3] We emphasize that results in this section are not particular to these specific choices of N_I and N_E. As long as $N_E - N_I > \theta$, the intermediate γ_{eig} producing broad, persistent distributions will continue to exist.

yields $\gamma_{\text{eig}} \approx 16.9$. However, according to Monte Carlo simulations, the branching ratio is always less than 1 for all $\gamma < N$. Firing rates thus fail to be maintained in this network, as reflected in the spike count JS divergence in Figure 2.6C. The reason is that A_{in} is so structurally similar to A: as in the purely excitatory networks, the high threshold still rejects weak inputs and sends them to the stationary state of quiescence, ν_{off}. This is in agreement with [120], who found that adding a homogenous inhibitory population to each layer does not help networks avoid synchrony.

Inhibition does, however, increase the robustness of JS divergence to perturbations in connectivity strength γ. Specifically, the troughs of minimal JS divergence widen compared to those of purely excitatory networks (Figure 2.6C). This is reflected as well in the eigenstructure: the intermediate state of persistent, broadly distributed distributions is now stretched to cover a wider range of γ (Figure 2.6AB). This robustness further grows as the size of the inhibitory population is increased, so long as $N_E - N_I > \theta$ (results not shown). Intuition for this effect can be obtained by comparing to the purely excitatory case for large N. On average, each neuron in this case has synaptic input of γ. To produce a broad range of responses and avoid either too many inputs (resulting in maximal firing rates) or too few (resulting in quiescence), γ must hover near some critical value that depends on parameters. Now suppose we add a population of N_I inhibitory neurons: then each neuron has on average $1/N(N_E - N_I)n$ net synaptic input. Since $N_E - N_I < N$, this slope is shallower than that for purely excitatory networks, so the networks are more robust to chance perturbations around the critical value of inputs, and hence to connectivity strength. Simulations agree with this scaling argument (see Appendix, esp. Figure 2.14).

Increased robustness to connectivity parameters in the presence of inhibition is interesting as it addresses a major concern regarding the plausibility of dynamics at critical transition values of connectivity (as discussed in the previous section). In sum, inhibition may help resolve the need for fine-tuning by enhancing robustness to fluctuations in network connectivity.

2.7 Impact of background noise

The next attempt to recover asynchronous rate propagation follows from [147], in which a noisy background current was shown to enhance the preservation of firing rates in feedforward networks of integrate-and-fire neurons; see also [107], [120], and [87]. We inject background noise in the form of independent, zero-mean Gaussian independent noise current χ to each neuron, $\chi \sim \mathcal{N}(0, \sigma_\chi^2)$. This transforms the heaviside-like thresholding into a smoother, sigmoidal operation. The probability that a neuron will spike given n cells firing

Other results regarding robustness, and limitations on asynchronous rate propagation high θ, also continue to hold.

in the upstream layer is now

$$q_n = \int_{-\infty}^{\infty} Pr(I > \theta - x) Pr(\chi = x) dx, \tag{2.7.1}$$

where I is the synaptic input from the upstream layer without the additional noise component. If $x > \theta$ then the neuron fires with probability 1 because the noise alone is enough to elicit a spike. If $x < \theta - n$, then the neuron can never fire as even the addition of all upstream neurons delivering input would be insufficient to cross threshold. We can then rewrite q_n as:

$$\begin{aligned} q_n = &\int_{\theta-n}^{\theta} \frac{1}{\sqrt{2\pi}\sigma_\chi} \exp\left(-\frac{x^2}{2\sigma_\chi^2}\right) \left[\sum_{k=\lceil\theta-x\rceil}^{n} \binom{n}{k} \left(\frac{\gamma}{N}\right)^k \left(1-\frac{\gamma}{N}\right)^{n-k}\right] dx \\ &+ \int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\chi} \exp\left(-\frac{x^2}{2\sigma_\chi^2}\right) dx. \end{aligned} \tag{2.7.2}$$

[107] consider a similar expression (Equation 7 in their paper), although both the exact form of their expression, and their conclusion that it has little effect on the transition matrix, differ from ours. We denote by A_{noisy} the transition matrix describing these networks with added noise, generated by the new q_n.

Figure 2.7 plots the spike count JS divergence (Equation 1) as a function of θ and σ_χ. The main result is that adding noise produces lower values of JS divergence – and thus more consistent propagation of asynchronous inputs – at larger values of θ. This result agrees with the findings of [147] (cf. their Figure 2B and see Appendix for further comparisons with this study). Our result is also in agreement with [120], who finds that adding white noise as a background current reduces the amount of synchrony present in networks.

For each of the threshold values θ shown, there is an optimal σ_χ for asynchronous rate propagation (i.e., that minimizes the JS divergence). This amount of noise gives spontaneous firing rates of less than 12%, as measured by the probability $Pr(\chi > \theta)$. For the remainder of this section, we will take the optimal value of noise for each value of threshold. Figure 2.8C uses these values to provide another view of optimized JS divergence, which reveals the improvement in comparison with noise-free networks (Figure 2.5). Moreover, γ_{obs} and γ_{eig} do coincide in the noisy case, even for high values of θ (at about 14.25 in for $\theta = 7$, branching ratio figure not shown).

One might expect the optimal level of noise to decrease for networks with more layers. Counter to this intuition, however, we find that the optimal noise level does not depend strongly on the number of layers.[4] This is because the optimal noise levels have the effect to elicit at *each* layer the amount of spontaneous background firing that will precisely balance the probability that the signal will terminate due to synaptic failure or chance redundant

[4] For these parameters, we find an optimal noise level of $\sigma_\chi = 0.3$ for 5 layers, and $\sigma_\chi = 0.4$ for 20 layers (for $\theta = 1$). For the high-threshold case of $\theta = 7$, we find that the optimal $\sigma_\chi = 4.6$ for both 5 and 20 layer networks.

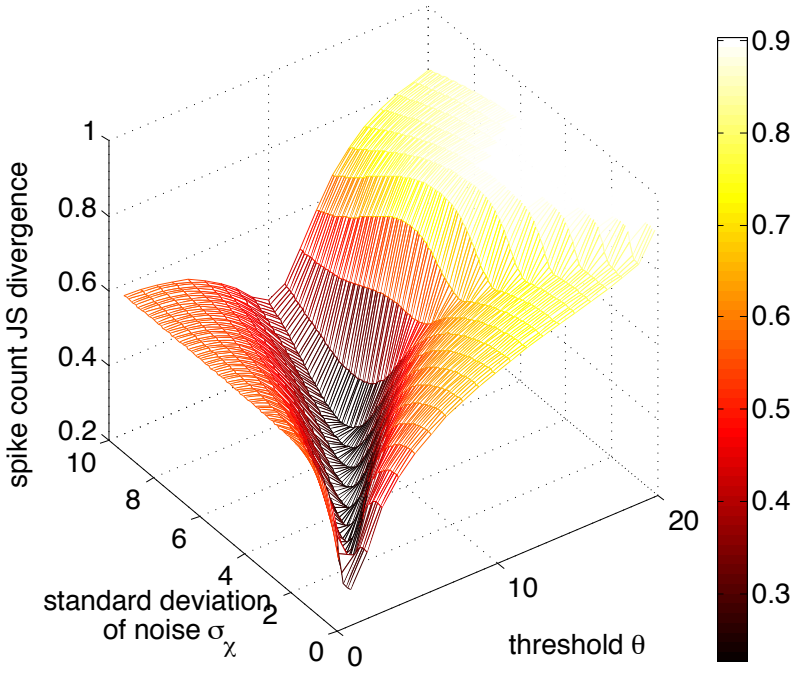

Figure 2.7: Impact of noise on input propagation. Surface shows spike count JS divergence as a function of θ and the standard deviation of noise added to each neuron, σ_χ. For each θ, there is a σ_χ that optimizes asynchronous rate propagation. For $\theta < 10$ the relationship between θ and optimal σ_χ is linear.

(i.e., highly convergent) connections. Therefore, optimal noise levels reduce an inherent asymmetry of the model that allows transitions from high firing rates to quiescence, but not the other way around.

In contrast to the effects of inhibition, the addition of background noise does produce substantial changes in the structure of the transition matrix. For example, comparing A_{noisy} with A, spontaneous activity is now possible, as v_{off} is no longer the stationary state. Instead, the stationary state v_{SS} is now a function of γ. In particular, the noise contributes a nonzero probability from transitioning from any state to any other state, so the components of A_{noisy} are strictly positive. By the Perron-Frobenius theorem, this means the system has a unique stationary state v_{SS} whose components are all strictly positive (so it can never be v_{on} or v_{off}). Computationally we find that the second largest eigenvalue now does not converge to 1 as $\gamma \to N$; it does, however, attain a peak value near 1 at an intermediate γ_{eig}, and at this point v_{SS} and v_{off} are also far from bimodal (Figure 2.8AB). Thus despite the differences in eigenstructure between A and A_{noisy}, the predominant features that define the existence of a persistent set of broad firing distributions are still apparent: there is an intermediate connectivity level γ_{eig} at which all eigenvalues are significant, the second largest eigenvalue in particular is close to 1, and both the stationary distribution and the second eigenmode

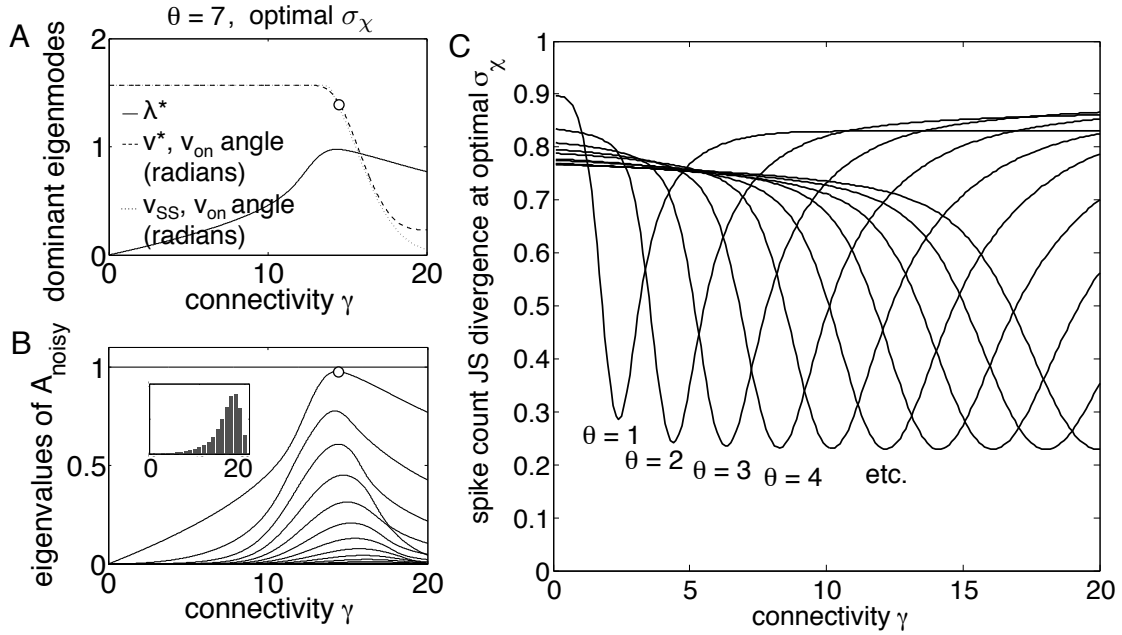

Figure 2.8: Properties of noisy networks at optimal background noise levels. (A) The second largest eigenvalue λ^* peaks very close to 1 at an intermediate γ_{eig}. The angle between v^* and v_{on} (dashed line, plotted in radians) and between v_{SS} and v_{on} (dotted line, plotted in radians) have large values at γ_{eig}. (B) At this same value γ_{eig}, all eigenmodes have significant contribution. Inset shows a typical broadly distributed histogram at γ_{eig} (indicated by the marker in A, B). (C) The spike count JS divergence, taking the optimal value of σ_χ for each θ. With optimal noise values added, asynchronous rate propagation is dramatically improved for high-threshold networks.

are far from bimodal.

Finally, to put the role of noise to a more demanding test, we test its impact on the capacity of networks to discriminate between different input stimuli. For this, we calculate the *rate discriminability* by measuring the error rate given by the maximum likelihood estimator on T trials. Specifically, suppose the network produces output spike counts $S_L^1, \ldots, S_L^T$ under some fixed input stimulus level S. Since the trials are independent, the maximum likelihood estimator (MLE) chooses between two stimuli S and S' by selecting the one that is likelier to result in the given data, via the likelihood ratio:

$$\prod_{j=1}^{T} \frac{P(S_L^j|S)}{P(S_L^j|S')}. \tag{2.7.3}$$

If this product is greater than 1 the MLE chooses stimulus S; less than 1 and the MLE chooses stimulus S'. Assuming S and S' are equally likely *a priori*, the error rate is given

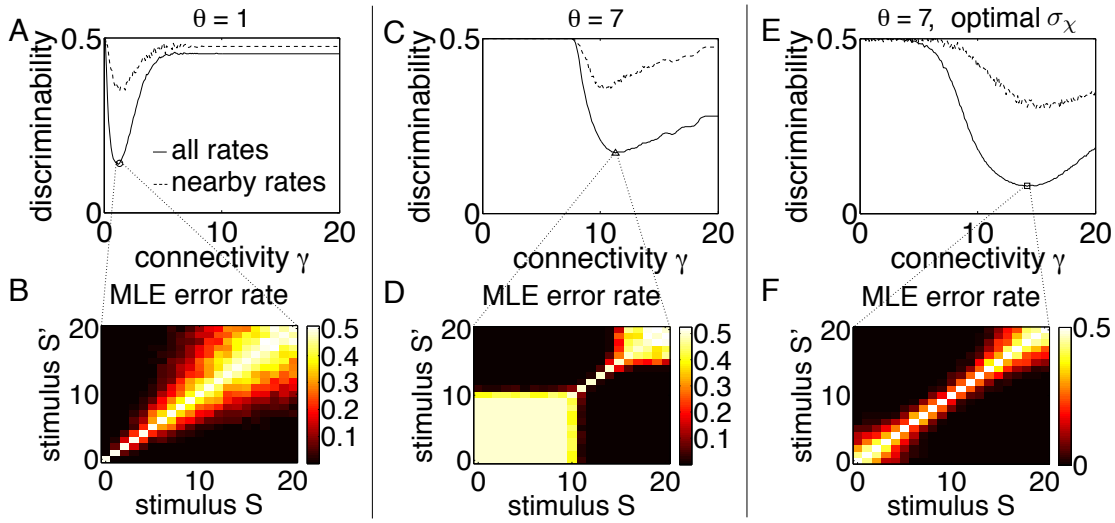

Figure 2.9: Rate discriminability for noise-free networks with (A, B) $\theta = 1$, and (C, D) $\theta = 7$, as well as (E, F) a network with $\theta = 7$ and optimal noise, $\sigma_\chi = 4.2$. (A, C, E) Nearby discriminability (dashed line) and average discriminability (solid line) for $T = 25$ trials plotted as a function of connectivity level γ. (A) For low-threshold networks, rate discriminability is optimal at γ_{eig}. (C) For high-threshold networks, nearby discriminability is best near γ_{eig}, but this minimum is shifted for average discriminability. (E) Adding noise improves rate discrimination in high-threshold networks. (B, D, F) Maximum likelihood error rate plotted for every possible pair of input stimuli before averaging. The chosen networks are those that minimize average discriminability, as indicated by the markers in (A, C, E).

by

$$\text{ER}(S, S') = \frac{1}{2}\mathbb{E}\left[\mathbb{I}\left(\prod_{j=1}^{T}\frac{P(S_L^j|S)}{P(S_L^j|S')} > 1\right)\middle|S'\right] + \frac{1}{2}\mathbb{E}\left[\mathbb{I}\left(\prod_{j=1}^{T}\frac{P(S_L^j|S)}{P(S_L^j|S')} > 1\right)\middle|S\right], \tag{2.7.4}$$

where the first expectation is taken over the distribution $P(\cdot|S')$ and the second over $P(\cdot|S)$. This produces an $(N+1)\times(N+1)$ matrix describing the MLE error rate for distinguishing S from S'. We then average over either the entries in the entire matrix to give either the average discriminability, or we average the entries in the superdiagonal to give the *nearby* discriminability, i.e, the discriminability between nearby rates.

Figure 2.9AC first summarizes discriminability in the absence of noise. Rate discriminability reaches its minimal value at $\gamma_{\text{eig}} \approx \gamma_{\text{obs}}$ when $\theta = 1$; when $\theta = 7$, the minimal discriminability does not exactly coincide with either γ_{eig} or γ_{obs}. A glance at the MLE error rates sans averaging reveals the particular type of computation performed in each case: Figure 2.9BD shows the error rates at the values of γ that yield the lowest average discriminability, as indicated by the markers in Figure 2.9AC. Low-threshold networks are able to accurately discriminate between rates over the entire stimulus space, including nearby

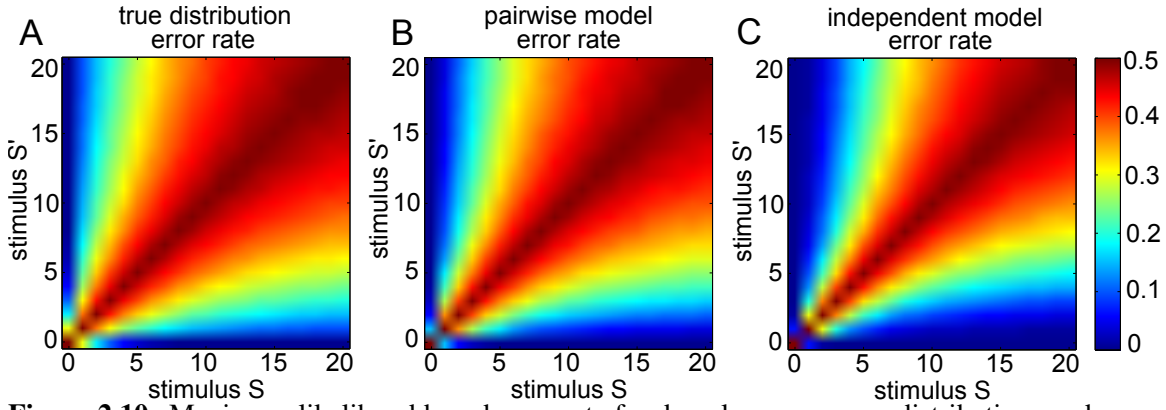

Figure 2.10: Maximum likelihood based error rate for deep-layer response distributions and maximum entropy fits. (A) Error rate at layer $L = 5$ of a critical network ($\theta = 1$), plotted for discrimination between every possible pair of input stimuli S and S'. Also shown are the error rates for the (B) pairwise maximum entropy and (C) independent fits of the spike count response distributions.

rates. High-threshold networks, on the other hand, although able to perfectly distinguish a few rates in a limited intermediate range, cannot at all distinguish between nearby high rates or low rates. Rather, these networks are better suited to classifying input rates into two bins: low and high.

Interestingly, the added background noise promotes better discriminability between rates in high-threshold networks, dropping the minimal level to values even below that of noise-free, low-threshold networks (Figure 2.9E). Moreover, the MLE error rates (Figure 2.9F) show a marked improvement in the ability to distinguish between nearby rates at γ_{eig}, as revealed by the tightly banded matrix structure. Not only, then, does noise improve rate propagation in neurons – it also changes the computation from a coarse-grained classifier to one with more resolution. This is a specific example of the more general phenomenon of stochastic resonance (see, e.g. [98, 89]).

2.8 Higher-order statistics and stimulus coding

We have shown that transmission of rate signals through feedforward layers best occurs at the critical connectivity range near γ^*. In this same parameter regime, the structure of the response distributions are most shaped by higher-order interactions, meaning that they cannot be fit by a pairwise maximum entropy model. This observation led us to investigate how higher-order statistics might contribute to signal transmission. In particular, how accurately can the input signal be determined based on the network response from one trial? In this case the MLE error rate reduces to the following:

$$\text{ER}(S,S') = \frac{1}{2}\mathbb{E}\left[\mathbb{I}\left(P(S_L^j|S) > P(S_L^j|S')\right)\middle|S'\right] + \frac{1}{2}\mathbb{E}\left[\mathbb{I}\left(P(S_L^j|S) > P(S_L^j|S')\right)\middle|S\right]. \quad (2.8.1)$$

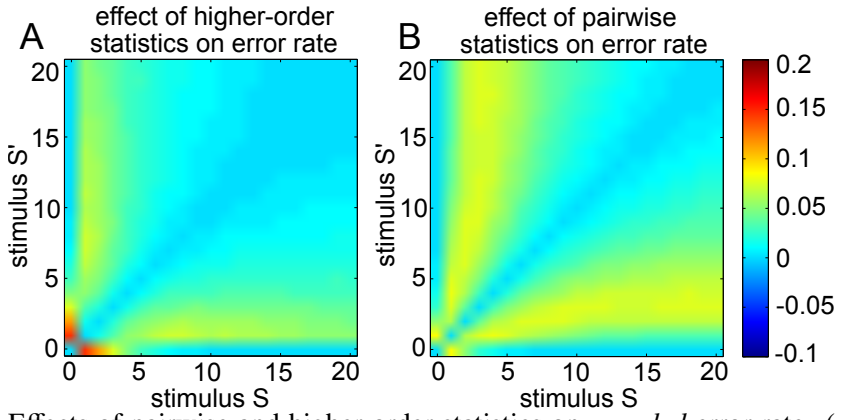

Figure 2.11: Effects of pairwise and higher-order statistics on *encoded* error rate. (A) Difference between the encoded error rate for the true distribution (i.e., the matrix shown in Figure 2.10A) and the encoded error rate for the pairwise maximum entropy models (Figure 2.10B) for all pairs of input stimuli. This shows the increase in the error that is due to higher-order statistics. (B) Difference between the encoded error rate for the pairwise models (Figure 2.10B) and the encoded error rate for the independent models (Figure 2.10C), showing the increase in error that is due to pairwise statistics.

Figure 2.10A shows the error rate in the 5th layer of a network at critical connectivity strength, shown for every possible pair of stimuli (cf. Figure 2.9). Next, in order to ascertain the effect of pairwise and higher-order statistics on signal transmission, we fit the spike count distribution in response to each stimulus by either a pairwise or an independent maximum entropy model. Then, we repeated the calculation of the MLE error rate as before, now replacing the distribution $P(\cdot|S)$ with the model distribution: for example, the error rate in the pairwise maximum entropy distribution is:

$$\mathrm{ER}(S,S') = \frac{1}{2}\mathbb{E}\left[\mathbb{I}\left(P^{ME}(S_L^j|S) > P^{ME}(S_L^j|S')\right)\middle|S'\right] + \frac{1}{2}\mathbb{E}\left[\mathbb{I}\left(P^{ME}(S_L^j|S) > P^{ME}(S_L^j|S')\right)\middle|S\right]. \tag{2.8.2}$$

The results are shown in Figure 2.10BC. This shows that overall error rates for both the pairwise and independent models follow the same trends as the full network responses.

To determine how the higher order statistics affect discrimination, we subtracted the error rates from the pairwise model (i.e., Figure 2.10B) from the true error rates (Figure 2.10A). The result, plotted in Figure 2.11A, shows the increase in error rate that is due to the higher-order statistics in the spike count distributions (i.e., statistical interactions among groups of three or more cells). Across all stimulus pairs, the higher-order statistics increased error rate by an average of 2.12%; for specific stimulus pairs the increase was as great as a 14.78% increase. Similarly, pairwise statistics increased error rates by up to 8.44%. This is shown in Figure 2.11B, which plots the difference between the error rates from the independent model and those from the pairwise model.

Does this mean that the higher-order statistics that emerge at the critical coupling $\gamma*$

are actually detrimental to rate discrimination? Figure 2.11 shows that this is the case, from the perspective known as "stimulus encoding." But, do higher-order statistics matter for a downstream population that must decode which stimulus was presented based on the activity of the last layer of the network? This perspective is called "stimulus *de*coding".

These two concepts (stimulus *encoding* versus *decoding*) give complimentary views of population coding. In both, MLE discrimination is based on the model of the true distribution: for example, when probing the effect of higher-order statistics, this means the pairwise model distribution P^{ME}. The subtle distinction is that for *encoding*, the MLE discrimination is applied to the pairwise model; for *decoding*, it is applied to the true distribution [84, 105, 106, 37, 151]. Philosophically, *encoding* performance describes whether the generation or presence of higher-order interactions in the response distributions contribute to the intrinsic representation of stimuli, while *decoding* performance reveals whether a decoder (such as a readout neuron or downstream population) requires knowledge of the higher-order statistics in order to discriminate between different stimuli. Mathematically, this translates into a difference in how the averaging is done in Equation 2.8.2: when measuring encoding performance, the expectation $\mathbb{E}[\cdot|S]$ is calculated over the model distribution $P^{ME}(\cdot|S)$, whereas measuring decoding performance requires calculating the expectation over the true distribution $P(\cdot|S)$.

Figure 2.12A shows the difference between the error rate in the true distribution and the decoded error rate in the pairwise maximum entropy distribution. Including higher-order interactions increased decoding performance, but only by very small amounts of up to 1.34%. Similarly, inclusion of pairwise statistics had little effect on discrimination (Figure 2.12B). Therefore, while the presence of pairwise and higher-order interactions in the response distributions impeded stimulus discrimination, they had little effect on the information that could be decoded from the true responses. Even though the higher-order interactions emerged with critical connectivity γ^*, they themselves did not contribute to stimulus discrimination. The richness of these surprising results led us to ask: when, if ever, do higher-order statistics facilitate coding? This is the topic of the next chapter in this dissertation, in which we investigate how higher-order spiking statistics impact stimulus encoding in more general populations.

2.9 Discussion

In this chapter, we study the transitions in feedforward network dynamics that occur as connectivity strength and firing threshold are varied. We characterize these transitions via critical branching, neutral stability, higher-order correlations, and broad firing distributions. After quantifying critical branching by computing the branching ratio, we show that neutral stability (persistence of firing patterns from one network layer to the next), together with statistical properties of the persistent patterns, can be predicted via a spectral analysis of the underlying mean-field transition matrix. Throughout most of the parameter space, persistent activity is restricted to highly bimodal, synchronous responses, as found

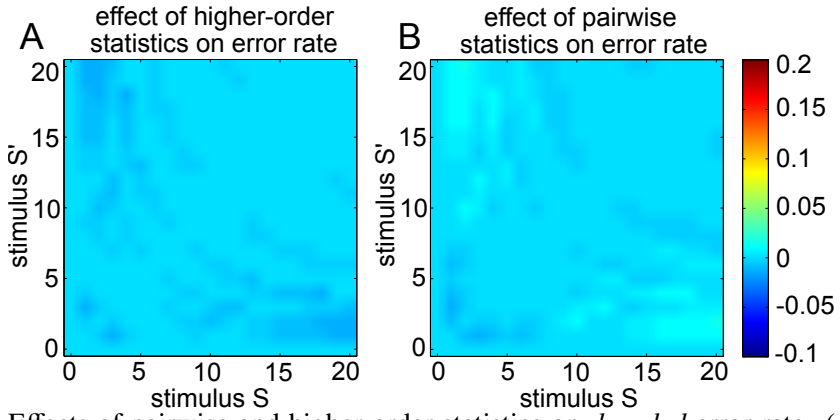

Figure 2.12: Effects of pairwise and higher-order statistics on *decoded* error rate. (A) Difference between the decoded error rate for the true distribution (i.e., the matrix shown in Figure 2.10A) and the decoded error rate for the pairwise maximum entropy models (not shown) for all pairs of input stimuli. (B) Difference between the decoded error rate for the pairwise models and the decoded error rate for the independent models.

by [120], [107], and [87]. However, there are "transition" connectivity levels that yield persistent, broadly-distributed spike count histograms with higher-order correlations and large response entropy. For low threshold networks this occurs simultaneously with (approximately) critical branching, revealing that such networks are well-suited to transmitting rates without synchronization. On the other hand, high-threshold networks do not produce both critical branching and broad response distributions at the same connectivity strength; when the former is satisfied, these networks tend to produce synchronous responses.

Interestingly, adding further biologically-motivated features increased the robustness of transitions in high-threshold networks. In particular, simulations and spectral analysis show that including an inhibitory cell population extended the connectivity range that yields asynchronous propagation of inputs. Adding zero-mean noise to each neuron had a similar effect and also improved the discriminability of inputs, echoing the findings of [147] in integrate-and-fire networks.

We conclude that networks with low firing thresholds, or those in which intrinsic noise elevates firing probabilities, exhibit a set of dynamical and statistical signatures associated with "critical" transitions in network activity.

2.9.1 Connections with the criticality literature

We now discuss links with the broader literature on criticality, which suggests that the brain may operate at a state characterized by complex dynamics, significant higher-order correlations, and enhanced computational properties. This is often described as operating on the boundary between ordered and irregular (or chaotic) activity. In particular, such systems can flexibly perform a wide range of operations on time-dependent inputs when their recurrent networks lie near the "critical" state, which is defined by calculating the

expected neutral separation of trajectories using a mean-field model [23, 85].

Along these lines [20] motivates a feedforward model based on array recordings.[5] Here, the authors compute the mutual information between the 2^N possible binary "words" at the first and last layers. Intriguingly, they numerically show – for the low threshold case $\theta = 1$ – that the mutual information is maximized for the same parameters at which critical branching occurs. Our finding in the averaged, mean-field setting echoes this result. An interesting extension of our work would be to explain the findings of [20] via the spectral properties of the allied layer-to-layer transition matrix between binary words. In principle, such an approach could predict the occurrence of cascades of such words over multiple network layers and their role in encoding stimuli.[6]

2.9.2 Verifying and extending the model

We imposed a number of simplifications to achieve analytical tractability. The most prominent of these is that our neurons are modeled as simple thresholding units with no intrinsic properties or time dependence [107]. An important extension is to consider how results might change if neurons included a finite integration time. For example, in excitatory-inhibitory networks, timing of inhibitory spikes will effect action potential generation [50]. While this can increase the selectivity of synfire chain propagation [80], the effect on rate transmission is unknown. More generally, if both an integration timescale and a refractory period are included, the collapsing of C and p into a single parameter – a key to our mean-field approach – may not be accurate in certain cases. Specifically, this could occur when C is high and p is low, as interaction between different waves of propagation through the chain will allow for synaptic facilitation, whereas the opposite relationship (low C and high p) will allow combinations of synaptic facilitation with refractory periods. It is hard to say for certain how these details will affect our results without direct simulation or analysis. However, our results agree with those in networks of more realistic neurons [147, 120, 122, 83]. We therefore believe that our findings give a reasonable description of rate propagation in feedforward networks and provide a good intuition for these systems in a tractable manner.

Another possible limitation is that the numerical studies presented above utilize a fixed value of $N = 20$ neurons. However, our analytical results on spectral properties of the transition operator are independent of this choice. Moreover, we verified that our main qualitative results are preserved, e.g., for the larger value $N = 100$ (taking $\theta = 1, 5, 10, 20, 35$); data not shown. In more detail, as with the smaller network, the system at $N = 100$ remains well-described by a mean-field transition matrix (in fact, due to the larger population size, it is even better fit). The eigenstructure of these matrices reveals an intermediate γ_{eig} at which the second dominant eigenmode is both persistent and broadly-distributed, and there is significant contribution from all eigenvalues as well as maximal response entropy. For

[5] The authors argue that a feedforward model is appropriate in this context as electrode sites are rarely active more than once during the cascades of neural activity that they study.

[6] Such cascades are called *neuronal avalanches* and have been the focus of a number of experimental and theoretical studies [20, 76, 61, 115, 67, 101].

$\theta = 1$, this value overlaps with γ_{obs}, but as threshold increases, the gap between the two widens; accordingly, the spike count JS divergence increases. As for the $N = 20$ case, while inhibition does continue to increase this range, the optimal performance is not improved. The addition of noise in large networks, however, has similar beneficial effects: an optimal amount of noise lowers the minimum JS divergence to around 0.32 for high values of θ. This amount of background noise required generates less than 10% probability of spontaneous firing, similar to that obtained at $N = 20$. However, one difference at $N = 100$ is that the optimal performance under the JS divergence metric D is lower: when $\theta = 1$, the optimal network attains at best a score of 0.58, compared to 0.33 for $N = 20$. Moreover, in the larger network the "well" in D values near the optimal γ value is even narrower, requiring a finer tuning of γ. These findings suggest that, while our findings remain qualitatively similar for larger networks, there may be interesting new phenomena in the continuum limit of large N – an interesting subject of future study.

On another note, we focused on only a few of the many metrics of signal propagation and coding that could be applied to the networks at hand. We note further results on one of these in the Appendix, that used by [147] to measure propagation of firing rates via trial-to-trial variance of responses in deep layers. This showed similar results to our measure D of JS divergence between input and output distributions over intermediate firing rates; the two measures showed distinctions at extreme firing rates, assessing the quiescent or saturating patterns that occur there differently (see Appendix).

We close by noting experimental predictions of our work, as could be tested directly in *in-vitro* feedforward networks (using the techniques of [120]), or, with the considerations above, could predict dynamics in recurrent systems as well. First, asynchronous rate propagation should become possible when the membrane potentials of neurons are biased upwards (equivalent to decreasing the spike-generation threshold). Second, this should also occur when sufficient noise is added local to each cell (some white noise has already been shown to reduce synchrony in [120]), and the optimal amount of white background noise should be independent of the number of layers. Finally, adding an inhibitory population at each layer should increase the robustness of asynchronous propagation to network connectivity and synaptic strength.

2.10 Appendix

2.10.1 Table of important symbols and their descriptions

Symbol	Description
N	number of neurons per layer
L	number of layers
θ	threshold level
C	number of downstream neurons connected to each neuron in the previous layer
p	probability of synaptic transmission
S	input spike count
γ	connectivity strength $\gamma = Cp$
$\mathbf{x}_L$	activity pattern in layer L (2^N possible values)
E_L	connectivity matrix between layers L and $L+1$
$\hat{E}_L$	effective connectivity matrix (including instantiations of p)
E	$N \times N \times L$ connectivity tensor of the entire network
σ	branching ratio averaged over E, p, and S
S_L	spike count at layer L
γ_{obs}	connectivity strength at which $\sigma \approx 1$ is observed via simulation
A	$(N+1) \times (N+1)$ mean-field transition matrix
q_m	probability that a neuron will fire given n spiking neurons in the previous layer
P_L	$(N+1)$-vector whose elements are the probabilities that $S_L = 0, \ldots, N$
λ^*	second largest eigenvalue of A
v_{off}	stationary state and dominant eigenmode of A, corresponding to quiescence
v_{on}	histogram of all neurons firing (not an eigenvector except for $\gamma = N$)
v^*	second dominant eigenmode of A (not a probability distribution)
γ_{eig}	connectivity strength at which $\lambda^* \approx 1$ yet v^* is far from bimodal
P_L^{ME}	pairwise maximum entropy fit of P_L
D_{JS}	Jensen-Shannon divergence
D	spike-count JS divergence, i.e., $D_{JS}(P_L, P_1)$ averaged over stimuli
N_E	number of excitatory neurons per layer
N_I	number of inhibitory neurons per layer
A_{in}	mean-field transition matrix for excitatory-inhibitory networks
σ_χ	standard deviation of background noise
A_{noisy}	mean-field transition matrix for excitatory networks with background noise
v_{SS}	(non-quiescent) stationary state, dominant eigenmode of A_{noisy}

2.10.2 Derivation of mean-field Markov chain

In this appendix, we outline how to obtain a formula for the mean-field Markov chain starting from the stochastic equations. The "derivation" is pedagogic in nature; that is, it is

intended to describe in principle the assumptions inherent in the mean-field model rather than provide a practical method of calculation of Equations 2.3.1 and 2.3.2. The transition probabilities are found through combinatorial arguments and were originally studied in [107].

As stated in the text, the main assumption here is that neurons within a layer are permutation symmetric. This appendix will describe how to obtain the mean-field Markov chain on spike counts (averaged over all connectivity architectures of connectivity strength γ) from the exact Markov chain on spiking patterns given a particular connectivity architecture E. The first step is to determine the probability that, given true connectivity matrix E_L, the instantiated "effective" connectivity matrix is $\hat{E}_L$:

$$P(\hat{E}_L|E_L) = p^{K(\hat{E}_L)}(1-p)^{K(E_L-\hat{E}_L)}, \tag{2.10.1}$$

where $K(M)$ is the number of nonzero elements in matrix M. Each element of the transition matrix in pattern space is then given by

$$P(\mathbf{x}_L = x|\mathbf{x}_{L-1} = \tilde{x}, E_L) = \sum_{\hat{E}_L} P(\hat{E}_L|E_L) \cdot \delta\left(x - \Theta(\hat{E}_L\tilde{x} - \theta)\right). \tag{2.10.2}$$

where δ denotes the Kronecker delta. We will in two steps reduce the dimension of the system to condense pattern space into rate space. First, summing over x:

$$P(S_L = m|\mathbf{x}_{L-1} = \tilde{x}, E_L) = \sum_{x} P(\mathbf{x}_L = x|\mathbf{x}_{L-1} = \tilde{x}, E_L) \cdot \delta\left(m - \sum_{i=1}^{N} x(i)\right). \tag{2.10.3}$$

Another sum gives the $(N+1) \times (N+1)$ transition matrix conditioned on the connectivity matrix E_L:

$$P(S_L = m|S_{L-1} = n, E_L) = \sum_{\tilde{x}} P(S_L = m|\mathbf{x}_{L-1} = \tilde{x}, S_{L-1}, E_L) \cdot P(\mathbf{x}_{L-1} = \tilde{x}|S_{L-1}, E_L) \tag{2.10.4}$$

$$= \sum_{\tilde{x}} P(S_L = m|\mathbf{x}_{L-1} = \tilde{x}, E_L) \cdot P(\mathbf{x}_{L-1} = \tilde{x}|S_{L-1}). \tag{2.10.5}$$

Finally, averaging over every possible E_L for the fixed γ, we obtain the elements of the mean field transition matrix, i.e.,

$$A_{nm} = P(S_L = m|S_{L-1} = n) \tag{2.10.6}$$

$$= \sum_{E_L} P(S_L = m|S_{L-1} = n, E_L) \cdot P(E_L), \tag{2.10.7}$$

which gives the elements of the mean-field transition matrix. Through these steps, the explicit derivation of the mean-field model (Equations 2.3.1 and 2.3.2) from the original

setup is demonstrated.

Finally, we can calculate the specific values of A_{nm} in Equation 2.3.1 as follows. Suppose m neurons are firing in the upstream layer. Assuming conditional independence of neurons within a layer, the probability of pattern x conditioned on the previous layer's spike count can be written as

$$\begin{aligned} P(\mathbf{x}_L = x | S_{L-1} = n) &= \prod_{i=1}^{N} q_n^{x(i)} (1 - q_n)^{1-x(i)} \\ &= q_n^{\sum_i x(i)} (1 - q_n)^{N - \sum_i x(i)}, \end{aligned} \tag{2.10.8}$$

where q_n represents the probability of any neuron firing given n neurons firing in the previous layer (as in Equation 2.3.2). Then, the transition probability $A_{nm} = P(S_L = m | S_{L-1} = n)$ can be computed from Equation 2.10.8 by summing over all possible combinations of patterns x such that $\sum_i x(i) = m$ i.e.,

$$P(S_L = m | S_{L-1} = n) = \binom{N}{m} q_n^m (1 - q_n)^{N-m}, \tag{2.10.9}$$

which is exactly Equation 2.3.1.

2.10.3 Validity of the mean-field Markov chain model

In this section we investigate the validity of the mean-field Markov chain model. Specifically, for a fixed network connectivity structure, we first estimate the true spike count distribution P_5 in response to input rate S through Monte Carlo simulation. We then compare this to the distribution predicted by the mean-field Markov chain $P_5^{\text{MF}} = P_{\text{input}} A^4$ by computing the Jensen-Shannon divergence between these two distributions. Finally, we average the JS divergence (Equation 2.5) over 100 instantiations of all possible input rates and over 20 random networks for that particular C and p.

Overall, the mean-field distribution approximates the true spike count distribution quite accurately, as shown in Figure 2.13A. The white curve overlain on the figure indicates the level set $\gamma \approx \gamma_{\text{eig}}$. Note that agreement is perfect for fully connected networks. The only major challenge to the accuracy of the mean-field approximation is in the sparse limit of low C and high p. Since C is low there are few trials for the stochastic synapses, and the high p additionally ensures that over repetitions of the same stimulus S the activity follows a nearly deterministic trajectory, resulting in P_5 having a narrower distribution than the mean-field predicts. Example histograms are shown in Figure 2.13BC to give an interpretation of values for the JS divergence.

When repeated for $\theta = 7$ (data not shown), the mean-field model even better captured the true distributions, with a maximal JS divergence of 0.15 in the region of inaccuracy in the limit of $p \approx 1$ and $C \approx \gamma_{\text{obs}}$. As a final check, we also compared the *means* of the response distributions and found that, as expected, the averaged error was below machine

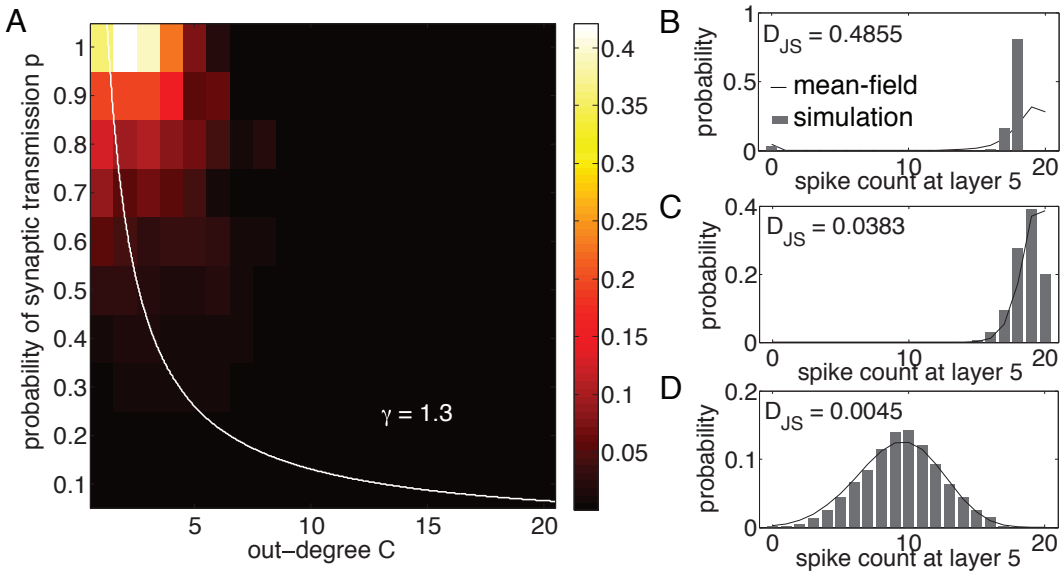

Figure 2.13: Investigating the validity of the mean-field model. (A) Average JS divergence between the distribution after simulation through five layers and that predicted by the mean-field model for varying C, p. The mean-field model breaks down in the sparse limit of small C and high p. The white curve represents $\gamma \approx \gamma_{\text{eig}} \approx 1.3$. (B - D) Example spike count distributions from 1000 Monte Carlo simulations (grey bars) and their mean-field predictions (black line) for three orders of magnitude of the JS divergence. Parameters are (B) $C = 3$, $p = 1$, $S = 3$ for the worst fit, (C) $C = 6$, $p = 0.5$, $S = 9$ for the intermediate fit, and (D) $C = 5$, $p = 0.26$, $S = 11$ for the best fit.

epsilon (results not shown).

2.10.4 Analytical results for the eigenstructure of the mean-field transition matrix

Proposition 1. *For any threshold $\theta \geq 1$ and connectivity $0 < \gamma < N$, the transition matrix A possesses a unique stationary state $\pi = v_{\text{off}}$ such that $\pi A = \pi$.*

Proof: Let $\pi = (p_0, \ldots, p_N)$. Then from direct matrix multiplication with Equation 2.3.1, the mth component of the vector πA is

$$(\pi A)_m = \sum_{n=0}^{N} p_n \binom{N}{m} q_n^m (1 - q_n)^{N-m}. \tag{2.10.10}$$

The stationary state requires $p_m = (Ap)_m$ for all m, i.e.

$$p_m = \sum_{n=0}^{N} p_n \binom{N}{m} q_n^m (1 - q_n)^{N-m} \tag{2.10.11}$$

for all $m = 0, \ldots, N$. In particular, when $m = 0$, this becomes

$$p_0 = p_0 + \sum_{n=1}^{N} p_n \binom{N}{0} q_n^0 (1 - q_n)^N. \tag{2.10.12}$$

The summed term on the right hand side must be zero. However, note that each of the components of this sum is nonnegative, so they each must be zero, i.e., for each $n = 1, \ldots, N$ either $p_n = 0$ or $(1 - q_n)^N = 0$. We could have $(1 - q_n)^N = 0$ for a particular n if $q_n = 1$. However, q_n can never be 1 for sensible parameter values of $\theta > 0$ and $0 < \gamma < N$. Therefore, we must have $p_n = 0$ for all $n = 1, \ldots, N$, and thus $p_0 = 1$. The resulting stationary state is therefore unique and precisely equal to v_{off}. □

Proposition 2. *Suppose the eigenvectors of A have limits as $\gamma \to N$. Then, A has an eigenvalue $\lambda^* \to 1$ as $\gamma \to N$ with corresponding eigenvector v^* that converges to a vector in the span of v_{on} and v_{off}.*

Proof: First consider (following Equation 2.3.2)

$$q_n = 1 - \sum_{k=0}^{\theta-1} \binom{n}{k} \left(\frac{\gamma}{N}\right)^k \left(1 - \frac{\gamma}{N}\right)^{n-k} \tag{2.10.13}$$

as $\gamma \to N$. For $n \leq \theta$, $q_n = 0$ by definition. For $n > \theta$, the sum on the right side of this equation approaches 0 since $n > k$, so $q_n \to 1$. Below we summarize for various m and n the limit of $q_n^m (1 - q_n)^{N-m}$ as $\gamma \to N$:

$$\begin{aligned}
n > \theta: \quad m = 0: &\quad q_n^0 (1 - q_n)^N \to 0 \\
0 < m < N: &\quad q_n^m (1 - q_n)^{N-m} \to 0 \\
m = N: &\quad q_n^N (1 - q_n)^0 \to 1 \\
n \leq \theta: \quad m = 0: &\quad q_n^0 (1 - q_n)^N \to 1 \\
0 < m < N: &\quad q_n^m (1 - q_n)^{N-m} \to 0 \\
m = N: &\quad q_n^N (1 - q_n)^0 \to 0.
\end{aligned}$$

Now suppose λ is an eigenvalue of A with corresponding eigenvector v for some γ. Then, λ and v satisfy

$$\sum_{n=0}^{N} v_n \binom{N}{m} q_n^m (1 - q_n)^{N-m} = \lambda v_m \tag{2.10.14}$$

for all $m = 0, \ldots, N$. In particular, for $m = 0$, we have:

$$\sum_{n=0}^{\theta} v_n \binom{N}{0} q_n^0 (1 - q_n)^N + \sum_{n=\theta+1}^{N} v_n \binom{N}{0} q_n^0 (1 - q_n)^N = \lambda v_0, \tag{2.10.15}$$

which, taking $\gamma \to N$, reduces to the following:

$$\sum_{n=0}^{\theta} \tilde{v}_n = \tilde{\lambda}\tilde{v}_0, \tag{2.10.16}$$

all other terms having vanished. Here, $\tilde{v}$ is the limit of v, which exists by assumption, and $\tilde{\lambda}$ is the limit of λ, which exists by the continuity of eigenvalues. For $m = N$, a similar expression is obtained:

$$\sum_{n=\theta+1}^{N} \tilde{v}_n = \tilde{\lambda}\tilde{v}_N. \tag{2.10.17}$$

Finally, for $0 < m < N$:

$$0 = \tilde{\lambda}\tilde{v}_m \tag{2.10.18}$$

This last equation reveals two possibilities: either $\tilde{\lambda} = 0$ or $\tilde{v}_m = 0$ for $0 < m < N$. The latter case implies that $\lambda = 1$, thus the second largest eigenvalue of A converges to 1 with limiting eigenvector in the span of v_{on} and v_{off}. All other eigenvalues converge to 0. □

Proposition 3. *Suppose $N_E - N_I > \theta$. Then, the (four-dimensional) transition matrix A_{in} has a unique (two-dimensional) stationary state π corresponding to no inhibitory and no excitatory neurons spiking at a layer. Moreover, the second largest eigenvalue converges to 1, and assuming the eigenvectors of A_{in} have limits as $\gamma \to N = N_E + N_I$, then its corresponding eigenvector converges to the space spanned by the vector corresponding to all inhibitory and excitatory neurons firing, and the vector corresponding to all inhibitory and excitatory neurons being quiescent.*

Proof: Because of the structure of A_{in}, this proposition follows similarly to those of the previous two propositions. □

2.10.5 Verification of scaling argument for the robustness of γ in excitatory-inhibitory networks

Section 2.6 explains that an added inhibitory population increases the robustness of optimal propagation to perturbations in γ, by arguing that the net synaptic input to each neuron is scaled by a factor of $(N_E - N_I)/N$. This scaling argument is exact in the limit of large N (minimizing fluctuation in synaptic inputs), and only treats perturbations near the critical γ that gives optimal rate propagation (as measured by the spike count JS divergence). The value of this critical γ, as well as the minimal value of the spike count JS divergence, need not be the same with and without inhibition under the scaling argument.

To verify that this is the case numerically, we calculated the spike count JS divergence for two different networks: the purely excitatory network (E) with $N_E = 200$, and the excitatory-inhibitory network (EI) with $N_E = 200$, $N_I = 50$. We then scaled each curve horizontally by a factor of $(N_E - N_I)/N$ (this is the identity transformation in the case of

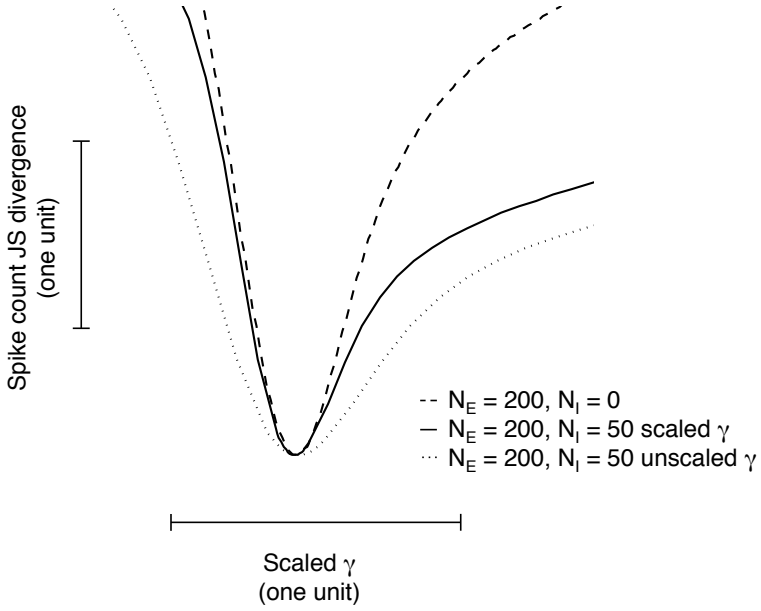

Figure 2.14: Spike count JS divergence of E networks with $N_E = 200$ (dashed curve) and EI networks with $N_E = 200$, $N_I = 50$ (solid curve) plotted against scaled connectivity $1/N(N_E - N_I)\gamma$. All curves have been translated horizontally and vertically to align the minimum at the same value. Note that near the minimum, the E network and the scaled EI network overlap. Compare with the spike count JS divergence for EI networks plotted against unscaled γ (dotted curve).

the E network) and translated the curves both vertically and horizontally to match their minimal values. If our scaling argument is correct, these curves should overlap near the minimal value. This is exactly what is seen in simulations (Figure 2.14, compare with the unscaled curve for the EI network).

2.10.6 Another metric for rate propagation

In addition to the measures described in the main text, we also considered the metric for rate propagation following [147]. Define the *rate dissimilarity* between the input rate S/N and the rate at the Lth layer S_L/N via:

$$\mathrm{RD}(\gamma, \theta) = \mathbb{E}_S \left[\mathbb{E}_{\text{trials}} \left[(S_L/N - S/N)^2 \middle| S \right] \right]. \tag{2.10.19}$$

There are two potential sources of poor performance according this quantification: (1) if the mean value of S_L is far from S, or (2) if S_L has large variance. As we see in Figure 2.15A, when $\theta = 1$ the rate dissimilarity reaches its minimal value at critical connectivity $\gamma_{\text{eig}} \approx \gamma_{\text{obs}}$, suggesting that for low-threshold neurons, these networks are best able to propagate rates through the network. Outside of this intermediate connectivity range, the dissimilarity between input and output returns to high values.

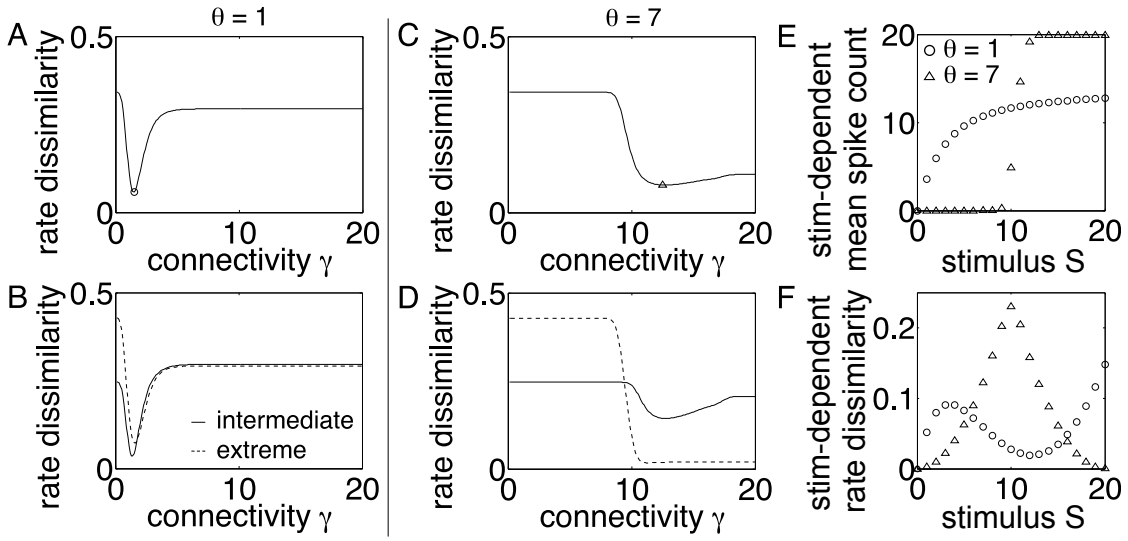

Figure 2.15: Rate dissimilarity for (A, B) $\theta = 1$, and (C, D) $\theta = 7$. (A, C) Rate dissimilarity plotted as a function of connectivity level γ. For (A) low-threshold networks, rate propagation is optimal at γ_{obs}. For (C) high-threshold networks, this is no longer the case. (B, D) Rate dissimilarity averaged over intermediate values ($S = 6, \ldots, 15$, solid line) and extreme rates ($S = 0, \ldots, 5$ and $16, \ldots, 20$, dashed line). (E) Mean of spike count and (F) rate dissimilarity for high-threshold (triangles) and low-threshold (circles) networks plotted as a function of input stimulus. The networks shown in (E) and (F) are those that minimize the stimulus-averaged rate dissimilarity, as indicated by the markers in (A, C).

When threshold is raised, the dissimilarity curve changes shape and no longer has a sharp minimum at γ_{eig} (Figure 2.15C); instead, there is a robust minimum. Moreover, the minimal rate dissimilarity values for the low- and high-threshold networks are at comparable values. This may at first seem surprising, given that the high-threshold networks produce strong synchrony, and this should lead to large response variance. What is actually happening is an effect of both the increasing nullity of A and averaging over all stimuli. In Figure 2.15E the stimulus-dependent mean of the output at the 5th layer is plotted as a function of the stimulus for the networks that minimize average rate dissimilarity, indicated by the markers in Figure 2.15AC, for both $\theta = 1$ (circles) and $\theta = 7$ (triangles). It is immediately clear that the low-threshold network better propagates intermediate rates as compared to the high-threshold network. By calculating the stimulus-dependent rate dissimilarity, rather than taking the uniform average, we see in Figure 2.15F the difference between these two networks. While high-threshold networks can propagate low and high rates better than low-threshold networks, only the latter can propagate intermediate rates. This is because high-threshold networks produce bimodal responses at the connectivity value required to propagate rates. To make this point more apparent, in Figure 2.15BC we have crudely separated the rate dissimilarity averaged over intermediate rates (solid lines) and extreme (either high or low) rates (dashed lines). This reveals that low-threshold

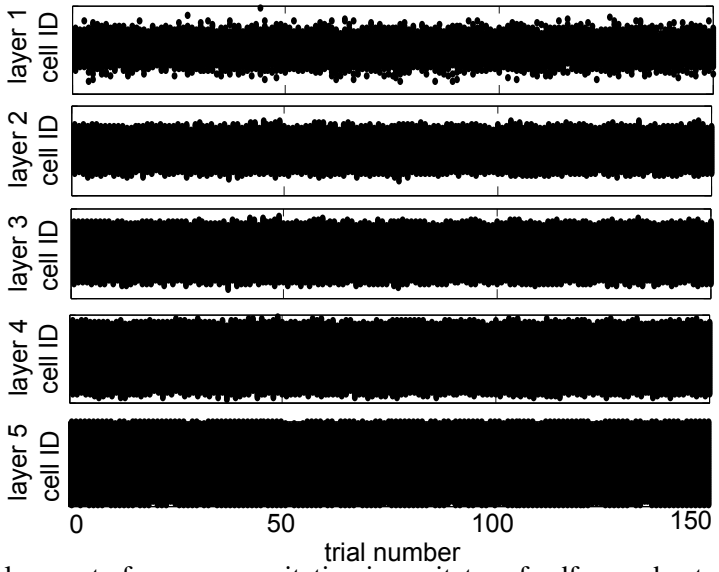

Figure 2.16: Development of runaway excitation in excitatory feedforward networks with spatially structured connectivity. Raster plots show the activity of every cell in five layers of an example network with input $\alpha = 50$ over 150 trials (each black dot indicates a spike). Network parameters: $N_E = 500$, $\sigma_E = .07$, $\theta = 5$.

networks perform better than high-threshold networks for intermediate rates.

Faced with the subtlety of these results, in the main text we use the spike count JS divergence in order to unambiguously reveal network properties that support the propagation of asynchronous input distributions.

2.11 Stable bump propagation in networks with localized connectivity

The work thus far described focused on feedforward networks with spatially homogenous connectivity between layers: that is, any two neurons within a layer had the same probability of connection to a neuron in the previous layer. In this section, we extend our analysis to study networks with localized connectivity between layers: in particular, we study the propagation of "bumps" of neural activity (i.e., localized regions of firing cells) through feedforward networks such that the width of the bump does not change. Computational models of working memory often use broad inhibitory and structured excitatory connections to stabilize bumps of activity in neural attractor models [34, 29, 9, 74, 55]. We extend these ideas to investigate the interplay of excitatory and inhibitory connectivity in models of bump transmission in feedforward networks.

2.11.1 Stochastic binary networks

As in the spatially homogenous case, we assume that neurons are binary thresholding units in a feedforward, layered network. We first consider purely excitatory networks, and subsequently add an inhibitory population at each layer.

Purely excitatory networks

We begin with a system for which each layer consists of a population of N_E excitatory neurons equally spaced on the interval $[0, 1]$ (with a ring topology). The location of neuron i is denoted as x_i. The excitatory connectivity is described by the function $f_E(x_i, x_j)$, which is the probability of a synaptic connection from neuron i in the upstream layer to neuron j in the next layer. Assuming that the connectivity kernel is translationally symmetric over the layer, we can replace $f_E(x_i, x_j)$ with the simpler notation $f_E(x_i - x_j)$. In the previous section, $f_E(\cdot)$ was spatially uniform; here, we examine the effect of localized connectivity by assuming a Gaussian connectivity profile:

$$f_E(x - y) = \exp\{-(x - y)^2/\sigma_E^2\} \tag{2.11.1}$$

where σ_E represents the width of the localized connectivity. The input to the network is a bump localized at position x^* and with fixed width:

$$s(x_i) = \exp\{-\alpha(x_i - x^*)^2\}. \tag{2.11.2}$$

The function $s(x)$ defines the probability that a first-layer neuron at position x fires. The parameter α tunes the width of the input: larger values indicate narrower bumps. Given such a spatially-structured network, is localized activity able to be transmitted from layer to layer? Figure 2.16 shows a typical example of a simulation of the propagation of a bump of activity through five layers of a network with $\sigma_E = 0.07$. While the activity at the first layer is localized within a subset of the layer, by the fifth layer activity has spread to all neurons. This shows that, in this example, purely excitatory networks were unable to stably propagate localized activity.

Excitatory-inhibitory networks

We now add N_I inhibitory neurons to each layer. The function f_I describes the probability that an upstream inhibitory neuron is connected to a downstream neuron (either excitatory or inhibitory):

$$f_I(x - y) = \frac{1}{k} \exp\{-(x - y)^2/\sigma_I^2\}, \tag{2.11.3}$$

where σ_I represents the width of the localized connectivity, and k sets the relative strength of the excitatory versus inhibitory populations. Here, we take $k = 2$. The effective connectivity between layers is the difference between excitatory and inhibitory connectivity

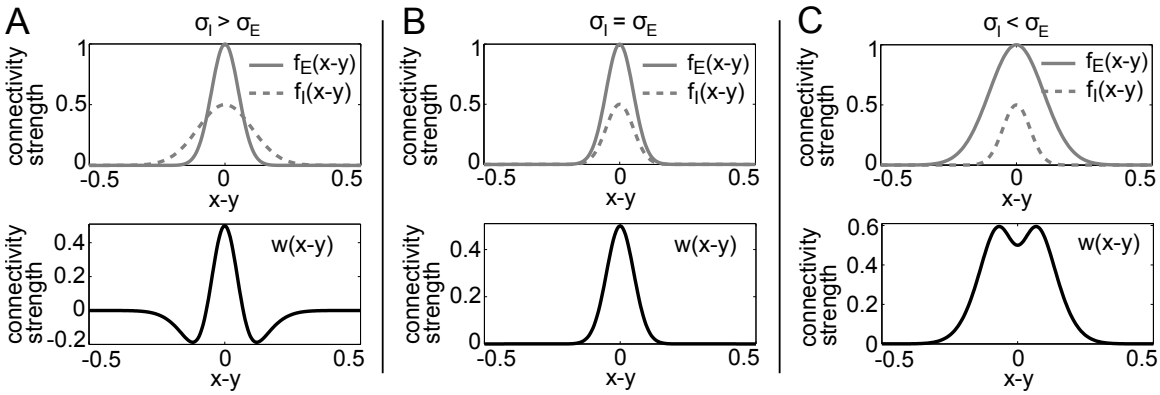

Figure 2.17: Example connectivity kernels for three representative cases. (A) Broader inhibition, $\sigma_E = .05$, $\sigma_I = .1$; (B) Equal inhibition and excitation, $\sigma_E = \sigma_I = .05$; (C) Broader excitation, $\sigma_E = .1$, $\sigma_I = .05$. Within each column, the top panel plots the excitatory connectivity kernel $f_E(x, y)$ (solid line) and the inhibitory connectivity kernel $f_I(x, y)$ (dashed line), while the bottom panel plots the effective connectivity kernel $w(x - y) = f_E(x - y) - f_I(x - y)$.

kernels:

$$w(x - y) = f_E(x - y) - f_I(x - y). \tag{2.11.4}$$

There are three distinct possibilities for the relationship between $f_E(x-y)$ and $f_I(x-y)$. The first possibility is that the excitatory and inhibitory connectivity have the same spatial scale: $\sigma_I = \sigma_E$ (Figure 2.17B). In this case, the effective connectivity is simply a Gaussian. If excitation is broader than inhibition ($\sigma_I < \sigma_E$; Figure 2.17C), the effective connectivity is a double-lobed function, but is still always positive. Finally, if inhibition is broader ($\sigma_I > \sigma_E$; Figure 2.17A), the effective connectivity has a shape similar to the Wicker wavelet, characterized by a large Gaussian-like component near 0 that is flanked by two smaller negative lobes. Given our choice of k, this is the only scenario in which $w(x - y)$ is not always positive.

In these excitatory-inhibitory populations, stable propagation is now possible: in particular, if the inhibitory connectivity is broader than the excitatory connectivity. Figure 2.18A shows an example of stable bump propagation in an excitatory-inhibitory network; the parameters here are identical to the Figure 2.16, except that we have added inhibitory neurons with $\sigma_I = 0.95$. With this broad inhibition, the bump of activity is preserved through even the fifth layer.

This is demonstrated in Figure 2.18B, which plots $H_L(x_i)$, the probability that cell i in layer L spikes. Using $H_L(x)$, we can now quantify the propagation of the bump of activity from layer to layer. Let R_L be the "excited region" of layer L, that is, the fraction of cells that have over 50% chance of firing:

$$R_L = \frac{1}{N_E} \sum_{i=1}^{N_E} \mathbb{I}[H_L(x_i) > 0.5] \tag{2.11.5}$$

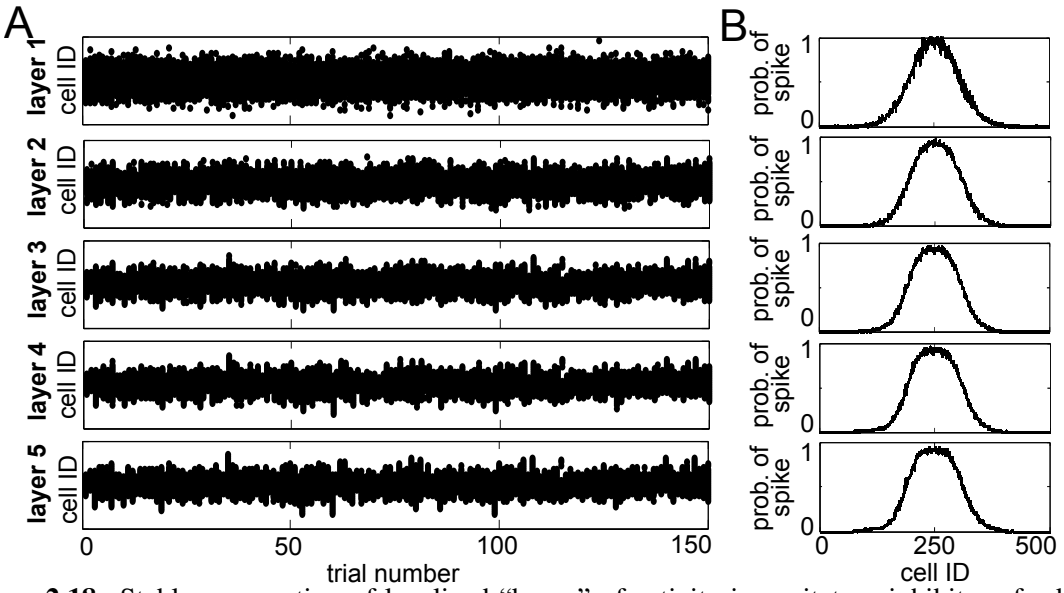

Figure 2.18: Stable propagation of localized "bump" of activity in excitatory-inhibitory feedforward networks. (A) Raster plots for an example network with input $\alpha = 50$. Broad inhibitory connectivity stabilizes bump propagation, preventing runaway excitation in the fifth layer. (B) Probability of spiking for each neuron in each layer, computed by averaging over activity in the 150 trials in panel A. Note how the bump of activity remains localized without growing or shrinking in size. In this example, the ratio of bump widths in the last layer to the ratio in the first layer is: $R_5/R_1 = 1.066$. Network parameters: $N_E = 500$, $N_I = 500$, $\sigma_E = .07$, $\sigma_I = .12$, $\theta = 5$.

The stable propagation in Figure 2.18 can now be quantified by R_L/R_1: for $L = 5$, this ratio is 1.066, revealing that the activity hardly changes width over the five layers.

What kinds of spatial profiles for excitatory and inhibitory connectivity are able to preserve bump activity through layers? To answer this question, we simulated the propagation of activity in networks with varying σ_I and σ_E for a particular choice of input: $\alpha = 50$ (results did not change qualitatively for different choices of α). Figure 2.19A shows R_5/R_1 plotted for the different excitatory and inhibitory connectivity widths. The networks that were able to stably propagate bump activity without it either decaying or saturating — that is, those with R_5/R_1 close to 1 — were characterized by a linear relationship between σ_E and σ_I. In general, networks with broader inhibition than excitation were best suited for keeping activity localized at each layer (i.e., had R_5/R_1 nearest 1, shown in solid white line), in agreement with the literature on modeling working memory as stabilized bumps in recurrent attractor networks [34, 29, 9, 74, 55]. In the remainder of these sections, we will determine whether a mean-field model is able to capture the same behavior of these stochastic simulations.

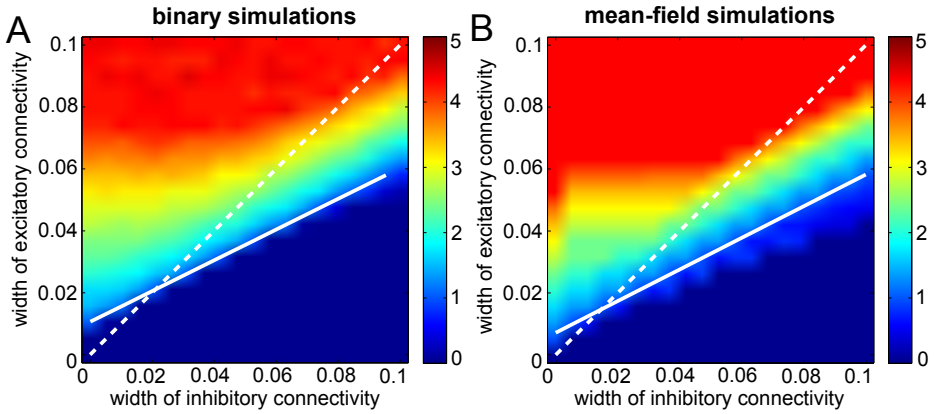

Figure 2.19: Propagation of bump of activity for various excitatory and inhibitory widths. (A) Stochastic binary simulations. The ratio of the widths of the excited region at the fifth and first layers, R_5/R_1, is plotted as a function of σ_E and σ_I. Also shown are the linear fit between σ_E and σ_1 such that $R_5/R_1 = 1$ (solid white line), and, for comparison, the identity line (dashed white line). Network parameters: $N_E = 500$, $N_I = 500$, $\sigma_E = .07$, $\sigma_I = .095$, $\alpha = 50$, $\theta = 5$. (B) Mean-field model. Parameters as in A, but with appropriate scaling factors (see text) and $r = .1$. The ratio of the widths of the excited region at the fifth and first layers, w_5/w_1, is plotted as a function of σ_E and σ_I. Also shown are the linear fit between σ_E and σ_1 such that $R_5/R_1 = 1$ (solid white line), and the identity line (dashed white line).

2.11.2 Mean-field model to describe localized connectivity

We now derive a mean-field model to describes the propagation of localized activity through layers. Here, we adapt the model studied in [9], which describes the time-dependent activity in a single continuous layer of cells with recurrent lateral connectivity, to study the propagation of activity in layered feedforward networks.

Let $u_L(x,t)$ denote the average membrane potential of the neuron at position x, time t, and layer L. We denote the input to the first layer as $\tilde{s}(x)$, a scaled version of the input to the binary model:

$$\tilde{s}(x) = A \exp\{-\alpha(x - x^*)^2\}. \tag{2.11.6}$$

The membrane potential at the first layer is modeled as a leaky integration of $\tilde{s}(x)$:

$$\tau \frac{\partial u_1(x,t)}{\partial t} = -u_1(x,t) + \tilde{s}(x) - r, \tag{2.11.7}$$

where τ is a time constant and r is the resting potential.

The input to subsequent layers can be represented as a convolution between the activity of the spiking cells in the previous layer and the weighting kernel, which is a scaled version of the effective connectivity kernel $w(x-y)$ for binary networks:

$$\tilde{w}(x-y) = B\left(f_E(x-y) - f_I(x-y)\right). \tag{2.11.8}$$

As in the binary network, we assume that neurons are thresholding units that are active only when the membrane potential is above a certain threshold, here taken to be 0 (without loss of generality, since we can adjust the resting potential r to set the effective threshold). The activity at each layer is then given by the following integro-differential equation:

$$\tau \frac{\partial u_L(x,t)}{\partial t} = -u_L(x,t) - r + \int_0^1 \tilde{w}(x-y) H[u_{L-1}(y,t)] dy \tag{2.11.9}$$

where $H[\cdot]$ is the Heaviside step function. Later in this section, we will derive values for the scaling factors A and B in order to quantitatively compare the mean-field model with the stochastic binary simulations.

Steady-state and time-dependent solutions to the mean-field equations

Because of the feedforward architecture, the time-dependent solution can be solved exactly using the integrating factor $e^{t/\tau}$. After some algebra, we arrive at the following equations:

$$u_1(x,t) = -re^{-t/\tau} + (\tilde{s}(x) - r)(1 - e^{-t/\tau}) \tag{2.11.10}$$

$$u_L(x,t) = -re^{-t/\tau} + \int_0^t \frac{1}{\tau} e^{-(t-s)/\tau} \left(-r + \int_0^1 \tilde{w}(x-y) H[u_{L-1}(y,s)] dy \right) ds \tag{2.11.11}$$

However, for the remainder of this chapter, we will consider only the steady-state solutions. At the first layer, setting $\frac{\partial u_1}{\partial t}$ to zero yields the following steady state solution:

$$u_1^{SS}(x) = \tilde{s}(x) - r \tag{2.11.12}$$

Similarly, the steady state solution in deeper layers reduces to:

$$u_L^{SS}(x) = -r + \int_0^1 \tilde{w}(x-y) H[u_{L-1}^{SS}(y)] dy. \tag{2.11.13}$$

These stationary states are stable, and the membrane potential of any neuron x will simply decay exponentially to its stationary state value. Because of the simplicity of the time-dependent dynamics, in the remainder of this section we assume that all layers have converged to their stationary states; for clarity, we will replace $u_L^{SS}(x)$ with simply $u_L(x)$.

Relationship to homogenous mean-field model

Earlier in this chapter, we developed a mean-field Markov chain model of spatially homogenous networks. How does that mean-field model relate to the one in Equation 2.11.13?

Suppose we were to take the same approach here that we took for the homogenous case. For simplicity we will treat the case of purely excitatory networks, but the equations for the excitatory-inhibitory networks follow the same logic. Let u_i^L be the state of the ith neuron in the Lth layer of the network, and suppose $\mathbf{u}^{L-1}$, the binary vector representing

activity at the previous layer, is known. Let S be the set of spiking cells in layer $L-1$: $S = \{i | u_i^{L-1} = 1\}$. In order for a downstream neuron to spike, it must receive inputs from at least θ of these cells. Therefore, the probability that any downstream cell spikes given the activity at the previous layer is:

$$P[u_i^L = 1|\mathbf{u}^{L-1}] = \sum_{S_1 \in 2^S} \mathbb{I}[|S_1| \geq \theta] \prod_{j=1}^{|S_1|} f_E(x_i^L - x_{k_j}^{L-1}) \prod_{j=1}^{|S \setminus S_1|} (1 - f_E(x_i^L - x_{h_j}^{L-1})), \quad (2.11.14)$$

where the first term indicates summation over all subsets $S_1 = \{k_1, k_2, \ldots, k_{|S_1|}\}$ of S of size θ or greater. The first product represents the probability that the neuron at position x_i^L receives synaptic inputs from all cells in S_1. Similarly, the second product term represents the probability that the neuron does not receive input from the remaining spiking cells: $S \setminus S_1 = \{h_1, h_2, \ldots, h_{|S \setminus S_1|}\}$. For homogenous networks, Equation 2.11.14 can be reduced significantly due to the permutation symmetry and the homogeneity of the probability of connectivity. But for spatially structured networks, these simplifications fail.

Instead, the mean-field description of Equation 2.11.13 takes a different approach. Let W^{L-1} be an instantiation of a binary adjacency matrix from layer $L-1$ to layer L with probabilities of connection $f_E(x_i - x_j)$. The average activity of cell i at layer L, conditioned on the activity at the previous layer, is given by:

$$\mathbb{E}[u_i^L|\mathbf{u}^{L-1}] = P[u_i^L = 1|\mathbf{u}^{L-1}] \quad (2.11.15)$$

$$= P\left(\sum_{j=1}^{N_E} W_{ij}^{L-1} u_j^{L-1} \geq \theta\right) \quad (2.11.16)$$

$$= \mathbb{E}\left[\Theta\left(\sum_{j=1}^{N_E} W_{ij}^{L-1} u_j^{L-1} - \theta\right)\right]. \quad (2.11.17)$$

At this step, we make the approximation of switching the Heaviside function and the expectation:

$$\mathbb{E}[u_i^L|\mathbf{u}^{L-1}] \approx \Theta\left(-\theta + \sum_{j=1}^{N_E} \mathbb{E}[W_{ij}^{L-1}] u_j^{L-1}\right). \quad (2.11.18)$$

This can now be simplified to the following equations:

$$\hat{u}_L(i) = -\theta + \sum_{j=1}^{N_E} f_E(x_i^L - x_j^{L-1}) u_j^{L-1} \quad (2.11.19)$$

$$u_i^L = \Theta(\hat{u}_i^L). \quad (2.11.20)$$

Finally, taking a continuum limit results in the mean-field model in Equation 2.11.13 in the previous section.

When is it appropriate to go from Equation 2.11.17 to Equation 2.11.18: that is, when

does $\mathbb{E}[\Theta(X)] \approx \Theta(\mathbb{E}[X])$? In general, this is an extremely poor approximation (for instance, one side is continuous-valued while the other is binary). However, when the support of X is either fully positive or fully negative, it is exact. This means that when the probability a cell spikes is either near 1 or 0, the mean-field model described here is accurate. For modeling the propagation of bumps of activity, this turns out to be a reasonable approximation for most stable locations except for a rapid transition region at the edges of the bump (see Figure 2.18). This explanation underlies the difference between Equation 2.11.13 and the mean-field transition matrix for homogenous networks: the latter was able to describe graded rate transmission through layers, while the former can only model the stable propagation of highly active but spatially localized activity.

Equilibrium Solution in Layers

Now instead of the equilibrium solution in time, we consider the equilibrium solution from layer-to-layer. This corresponds to a function $\tilde{u}(x)$ such that:

$$\tilde{u}(x) = -r + \int_0^1 w(x-y)H[\tilde{u}(y,t)]dy. \tag{2.11.21}$$

The activity in any layer described by membrane potential function $u_L(x)$ can be defined succinctly by the "excited region" [9], which is the subset of the layer in which the cells are firing: $R[u_L] = \{x|u_L(x) > 0\}$. Denote by w_L the width of $R[u_L]$. Now, supposing that the excited region for the layer-equilibrium solution $R[\tilde{u}]$ is the interval $[a, b]$. Following the calculations in [9], we can rewrite the equation above as:

$$\tilde{u}(x) = -r + \int_{R[\tilde{u}]} w(x-y)dy \tag{2.11.22}$$

$$= -r + \int_a^b w(x-y)dy, \tag{2.11.23}$$

$$= -r + \int_{a-x}^{b-x} w(y)dy, \tag{2.11.24}$$

where we have used the assumption that $w(x-y)$ is symmetric in the third step. For consistency, the following conditions on $\tilde{u}(x)$ must hold:

- $\tilde{u}(a) = 0$, or equivalently:
$$\int_0^{b-a} w(y)dy = r \tag{2.11.25}$$
- $\tilde{u}(b) = 0$, or equivalently:
$$\int_{a-b}^0 w(y)dy = r \tag{2.11.26}$$
Because of the symmetry of $w(y)$, this is equivalent to the previous condition.

- $\tilde{u}(x) > 0$ for $x \in [a, b]$, or equivalently:

$$\int_{a-x}^{b-x} w(y)dy \geq r \tag{2.11.27}$$

Trivially, these conditions hold if $H[\tilde{u}(x)]$ is either identically 0 or 1 for all x. It is more difficult to determine when $\tilde{u}(x)$ supports localized activity that is neither completely quiescent or completely active across the entire layer.

An interesting consequence of these conditions is that that multiple bumps can be supported stably by certain connectivity profiles, even though the input consists of only one bump. In the case that $\sigma_E \geq \sigma_I$, there can only be on localized bump because the connectivity kernels are always positive. However, networks with broad inhibitory connectivity ($\sigma_E < \sigma_I$) can support multiple localized bumps of activity if σ_I is sufficiently large and if the input is broad enough. This is because the negative lobes of $w(x - y)$ (seen in Figure 2.17A) make it possible to have two bumps of activity with negative membrane potential between them. We will return to the possibility of multiple bumps in the following section.

2.11.3 Matching the mean-field model and stochastic models

In order to compare the continuous, deterministic mean-field model with our discrete, stochastic binary simulations, we must find the appropriate scaling factors on the input $\tilde{s}(x)$ and synaptic connectivity $\tilde{w}(x-y)$ that will quantitatively match the neural activity for the two models. Towards this end, in this section we derive values for A and B to match the effective strength of the external and synaptic input in each model.

First, for any input width α, the two models should have the same fraction of active cells in the first layer. Therefore our first constraint is that the width of the excited region in the first layer of the mean-field model must match the average number of cells that fire in the first layer of a binary network. Recall that in the mean-field model, the input into the first layer is given by:

$$\tilde{s}(x) = A \exp\{-\alpha(x - x^*)^2\} \tag{2.11.28}$$

From Equation 2.11.12, the steady state distribuion is $u_1(x) = \tilde{s}(x) - r$. The width of the excited region in the first layer, w_1, is determined by finding all x such that $u_1(x) = 0$, or equivalently, $\tilde{s}(x) = r$:

$$A \exp\{-\alpha(x - x^*)^2\} = r \tag{2.11.29}$$

$$-\alpha(x - x^*)^2 = \log\left(\frac{r}{A}\right) \tag{2.11.30}$$

$$x - x^* = \pm\sqrt{-\frac{1}{\alpha}\log\left(\frac{r}{A}\right)} \tag{2.11.31}$$

which means the width of the excited region is:

$$w_1 = 2\sqrt{-\frac{1}{\alpha}\log\left(\frac{r}{A}\right)}. \tag{2.11.32}$$

Since the cells are distributed over the interval $[0, 1]$, the excited region width w is also the fraction of cells active in the first layer of the mean-field model.

In the binary network, the input $s(x_i)$ describes the probability that neuron x_i in the first layer spikes in a particular trial; thus, the average number of active cells is simply

$$\mathbb{E}[\#\text{ cells on}] = \sum_i s(x_i). \tag{2.11.33}$$

In the continuum limit of large N, we can use a Riemann sum approximation to get:

$$\mathbb{E}[\#\text{ cells on}] \approx \frac{1}{\Delta x}\int_0^1 s(x)dx. \tag{2.11.34}$$

Here, $\Delta x = 1/N_E$. To match the fraction of cells firing in the first layer in the binary and mean-field models, we set the equation above equal to $N_E w_1$, or:

$$\int_0^1 s(x)dx = 2\sqrt{-\frac{1}{\alpha}\log\left(\frac{r}{A}\right)}. \tag{2.11.35}$$

Solving for A yields:

$$A = \frac{r}{\exp\left\{-\frac{\alpha}{4}\left(\int_0^1 s(x)dx\right)^2\right\}}. \tag{2.11.36}$$

Similarly, in deeper layers, we seek to match the fraction of threshold attained by the synaptic input to a single cell, assuming that all neurons in the previous layer are firing. In the mean-field model, this means that $H[u_{L-1}(x)] = 1$, so the steady state at the next layer simplifies to:

$$u_L(x) = -r + \int_0^1 \tilde{w}(x-y)dy \tag{2.11.37}$$

$$= -r + \int_0^1 Bw(x-y)dy \tag{2.11.38}$$

which is independent of the choice of downstream neuron x. Letting $I = \int_0^1 w(x-y)$, fraction of threshold reached by the synaptic input to x is BI/r.

In the binary network, the expected number of synaptic inputs to a downstream cell x_j is

$$\mathbb{E}[\#\text{ inputs to } x_j] = \sum_i (f_E(x_j, y_i) - f_I(x_j, y_i)). \tag{2.11.39}$$

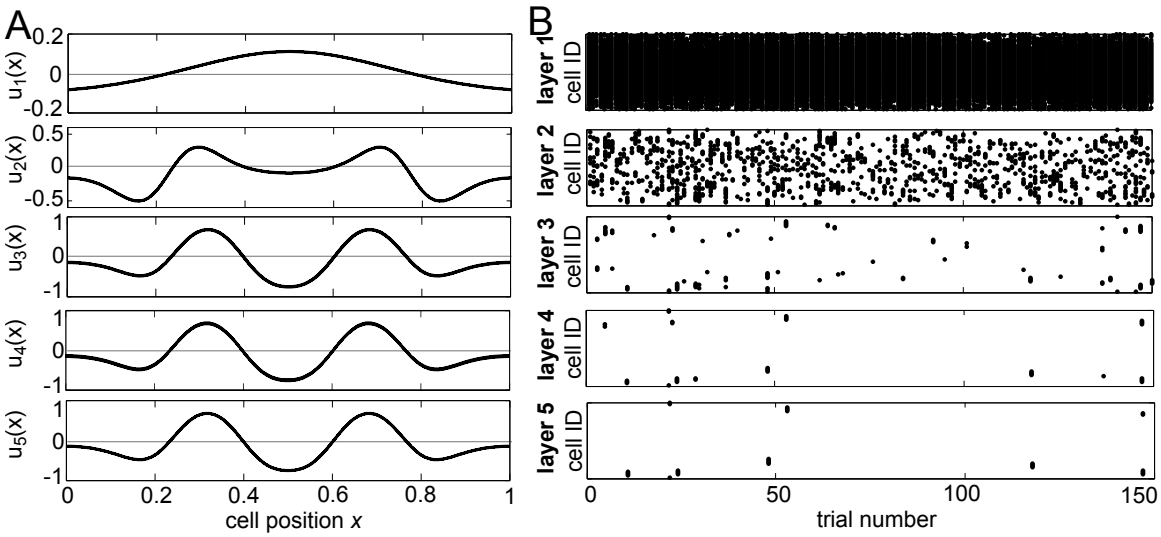

Figure 2.20: Examples of a single bump input splitting into multiple bumps in deeper layers. (A) Example of two-bump steady-state solution $u(x)$ in the mean-field model. The membrane potential $u_L(x)$ is plotted against cell position x for five layers. Any cell with positive membrane potential (i.e., $u_L(x)$ is above the grey line) is considered to be active. The single, broad input bump already separates into two distinct bumps by the second layer. This example could not be reproduced in the corresponding binary network. Parameters: $\sigma_E = .05$, $\sigma_I = .1$,$\alpha = 10$, $r = .1$. (B) Raster plots for an example binary network in which two localized bumps of activity are visible in layer 3. However, the multiple bumps are not stable, and they quickly decay. Network parameters: $N_E = 500$, $N_I = 500$, $\sigma_E = .03$, $\sigma_I = .079$, $\alpha = 10$, $\theta = 5$.

Again, as $N \to \infty$ we use a Riemann sum approximation:

$$\mathbb{E}[\#\text{ inputs to } x_j] \approx \frac{1}{\Delta x}\int_0^1 w(x-y)dy = \frac{I}{\Delta x} \tag{2.11.40}$$

so the fraction of threshold reached by the synaptic input to neuron x_j is $I/(\theta\Delta x)$. Matching these two fractions:

$$\frac{I}{\theta\Delta x} = \frac{BI}{r}. \tag{2.11.41}$$

Solving for B yields:

$$B = \frac{r}{\theta\Delta x}. \tag{2.11.42}$$

With Equations 2.11.36 and 2.11.42 we can now approximately match the binary and the mean-field models. Under the mean-field model, stable bump propagation is measured by calculating the ratio of the width of the excited region at layer L to the width of the excited region at the first layer: w_L/w_1. To see how well the mean-field approximates the simulations, in Figure 2.19B we calculated this ratio while varying σ_E and σ_I. Comparing panels A and B in Figure 2.19, we can see that the mean-field model matches the stochastic simulations extremely well. But, as remarked previously, the mean-field model is not able

to describe all qualitative behavior of the networks.

Thus, surprisingly, it appears that interactions that develop through layers do not have a significant effect on bump propagation. But is the mean-field model able to perfectly describe the qualitative behavior of binary networks?

Recall that in the previous section we found that it is possible for a single bump at the input to separate into multiple bumps in the equilibrium solution, in particular if $\sigma_I > \sigma_E$ and if the input bump is sufficiently wide. An example of this is shown in Figure 2.20A, in which a broad input to the first layer is separated into two stable bumps in deeper layers. In the binary network with the same σ_E and σ_I, the initial bump does not separate into two bumps. However, Figure 2.20B shows an example with different parameters in which it is possible to see traces of two localized bumps in the third layer; this said bumps are not stable, and they decay quickly in the next couple of layers.

2.11.4 Discussion

In sum, we simulated the propagation of localized neural activity in stochastic, binary feedforward networks with spatially structured excitatory and inhibitory activity. We found that, in general, broad inhibition was necessary to stabilize bump propagation through multiple layers; specifically, stable propagation was predicted by a linear relationship between the inhibitory and excitatory connectivity widths. These results were very accurately reproduced in a deterministic, mean-field model. However, the mean-field model failed to accurately describe activity with intermediate rates, and conversely, stochastic networks were not able to support multiple bumps in deeper layers. Future work will focus on when spatially structured feedforward networks can support localized bumps of graded rates, as well as develop analytical tools to model them.

CHAPTER 3

Impact of triplet correlations on stimulus coding

The brain transforms sensory inputs into spiking activity that is distributed across neural populations and is variable from trial to trial. What are the key statistical features of this activity that determine the amount of sensory information encoded by such a population? Much can be learned by quantifying the mean responses as well as the trial-to-trial variability of spikes emitted by individual cells. However, this variability is often coordinated across the population. Significant correlations between the spikes emitted simultaneously by pairs of cells have been observed across the brain, e.g. in visual cortex [62, 93, 77] (but see [44]), auditory cortex [41], motor cortex [95], prefrontal cortex [35], the lateral geniculate nucleus [8], and retina [94] – possibly reflecting circuit mechanisms such as recurrent connectivity and common input [145, 77, 24, 119, 126, 27]. Such pairwise spike correlations can have a wide range of impacts on stimulus encoding. In principle, pairwise correlations can interfere with population-wide averaging that would otherwise damp noise; conversely, they may play a more positive role, allowing variability to be cancelled or even acting as an extra conduit of information independent of firing rates. Thus, a large body of theoretical work has been dedicated to understanding the precise relationship between pairwise correlations and population coding, e.g., [154, 2, 13, 114, 112, 135, 36, 70, 127].

Intriguingly, recent experiments suggest that knowing the correlations between pairs of neurons is not enough to characterize collective activity across a neural population. This implies the existence of "higher-order correlations" (HOCs): that is, correlated firing between groups of three or more cells that is either more or less than what would be expected from the firing rates and pairwise correlations alone [52, 109, 100, 128, 79, 141]. Results to date illustrate that, as for pairwise correlations, HOCs can have a range of positive to negative effects on stimulus encoding. This is assessed by comparing coding fidelity based on the "ful" responses recorded simultaneously in a population, with coding fidelity based on a model population that has the same firing rates and pairwise correlations but no HOCs. In vertebrate retina, higher-order correlations among retinal ganglion cells improved coding efficiency – specifically, they increased the speed with which the identity of two types

of visual stimulus could be decoded from the population response [52]. Meanwhile, HOCs in somatosensory cortex decreased mutual information between neural activity and the frequency of whisker stimulation [100].

These findings raise two important questions. First, when should we expect HOCs to have a significant impact on population coding? Second, a common rule of thumb for pairwise correlations is that the encoded information increases when the noise correlations cancel out out signal correlations [13]. Are there similar simple rules that predict when HOCs will facilitate versus hinder the population code? These questions remain largely unexplored, but the answers may lead to new perspectives on neural coding, as many studies to date have used measures of coding accuracy, such as the optimal linear estimator [123] that do not incorporate the effects of HOCs.

We approach these questions by investigating the effect that triplet correlations — the most frequently-observed higher-order correlation [52] — can have on two-stimulus discrimination tasks. Throughout, we use maximum-entropy statistical models [124] that isolate the effect of triplet correlations, while fixing the lower-order statistics (i.e., mean activity of each neuron and correlations among each pair) to prescribed values typical of those reported in physiology experiments. Positive (or negative) triplet correlations signify that triplets of cells spike together more (respectively, less) frequently than expected from the lower-order statistics. We find that triplet correlations can indeed strongly improve stimulus encoding, if they have a stimulus-dependent structure. Specifically, if triplet correlations among cells with similar stimulus tuning are larger for their nonpreferred versus their preferred stimulus (or, to a lesser extent, vice-versa), then the triplet correlations will separate the distributions of population spiking patterns produced by each stimulus. As a result, the stimuli can be better discriminated. Comparable statistical models with stimulus-independent triplet correlations show relatively little effect on coding. We show that these findings can be explained geometrically as either positively or negatively skewing the distribution of the summed population activity in short time windows. Our results emphasize the importance of quantifying higher-order correlations in neurophysiology experiments, as they may have a significant impact on the coding performance of neural systems. Finally, a major challenge of measuring correlated spiking is the large amount of data that is required for accurate detection. We give a simple calculation that estimates the length of recordings necessary to identify such triplet correlations experimentally.

3.1 Material & Methods

We investigate the effect of higher-order spike correlations (HOCs) on the level of stimulus information that a neural population encodes about pairs of stimuli: a preferred stimulus (eliciting a higher firing rate), and a non-preferred stimulus. Each stimulus elicits a different distribution of spike patterns characterized by firing rates, pairwise correlations, and HOCs. We vary the triplet statistics separately for each stimulus, and calculate the amount of information that spiking patterns contain about the stimulus identity. In order to iso-

late the effect of HOCs, we keep the lower-order statistics (i.e., firing rates and pairwise correlations) fixed during this process. We do this by using a popular class of statistical models called maximum entropy models, which are able to match any given statistics of a population of neurons while minimally constraining other features of the spike distribution.

3.1.1 The maximum entropy model

Consider the spikes emitted by N cells in response to stimulus $S^{(m)}$, where $m = 1$ or 2. Binning these spikes in small windows yields a sequence of spiking patterns $\vec{\sigma}$, each of which is a vector of 1s and 0s representing whether a given neuron spiked or not within that time window. Assuming that the population is at a stationary state, each pattern $\vec{\sigma}$ can be viewed as a random sample from a probability distribution that describes the simultaneous, population-wide response of the neurons to a particular stimulus. These are the probability distributions that we will study in this paper.

If the i^{th} neuron spikes with probability μ_i in each time window (i.e., the firing rate of the i^{th} neuron is $\mu_i/\Delta t$), then the (simultaneous) pairwise spike correlations for cell i and j are:

$$\rho_{ij} = \frac{Pr(\sigma_i, \sigma_j = 1) - \mu_i \mu_j}{\sqrt{\text{var}(\sigma_i)\text{var}(\sigma_j)}}. \tag{3.1.1}$$

In other words, to quantify the correlation between pairs of neurons, one must subtract from the observed probability of simultaneous paired spiking the probability of simultaneous paired spiking in a "null" model (in this case, assuming all activity is independent).

Similarly, quantifying higher-order correlations requires comparing against a null model. In this case, we use the pairwise maximum entropy model, which matches the observed lower-order statistics while making the fewest additional assumptions about the structure of the data [125, 124]. Under this model, the probability of firing pattern $\vec{\sigma}$ under stimulus $S^{(m)}$ is given by:

$$P_{PW}(\vec{\sigma}|S^{(m)}) = \frac{1}{Z} \exp\left[\sum_i h_i^{(m)} \sigma_i + \sum_{i>j} J_{ij}^{(m)} \sigma_i \sigma_j\right]. \tag{3.1.2}$$

Here, the interaction terms $h_i^{(m)}$ and $J_{ij}^{(m)}$ are tuned so that the distribution matches the prescribed lower-order statistics, that is, firing rates and pairwise correlations. Z is a normalization factor. Thus equipped, we define a measure of triplet correlations as the probability of three neurons firing simultaneously, relative to what would be expected from the pairwise maximum entropy model:

$$\kappa_{ijk} = Pr(\sigma_i, \sigma_j, \sigma_k = 1) - Pr_{PW}(\sigma_i, \sigma_j, \sigma_k = 1). \tag{3.1.3}$$

We refer to this quantity as the "excess triplet probability."

Throughout this paper, the phrase "triplet correlations" is synonymous with the excess triplet probability: the probability of simultaneous spiking in triplets of cells beyond what

is expected from the lower-order statistics under a maximum entropy assumption.

In order to explore the effects of HOCs, we add a triplet interaction term $G^{(m)}$ to the previous distribution:

$$P(\vec{\sigma}|S^{(m)}) = \frac{1}{Z}\exp\left[\sum_i h_i^{(m)}\sigma_i + \sum_{i>j} J_{ij}^{(m)}\sigma_i\sigma_j + G^{(m)}\sum_{i>j>k}\sigma_i\sigma_j\sigma_k\right]. \tag{3.1.4}$$

Increasing (or decreasing) $G^{(m)}$ increases (or decreases) the excess triplet probability κ_{ijk}. For simplicity, we set the triplet interaction term to be the same for all triplets of neurons; however we have also added heterogeneity by adding zero-mean noise to the triplets $G^{(m)}$ terms for each triplet i, j, k, and we found the same qualitative results that we will report here, as long as the $G_{ijk}^{(m)}$ have the same sign for each triplet (data not shown).

The approach we have described is useful, because it allows us to isolate the effects of triplet correlations: for each triplet interaction $G^{(m)}$, we re-fit the single-cell and pairwise interactions $h_i^{(m)}$ and $J_{ij}^{(m)}$ to maintain the same firing rates and pairwise correlations. However, this is computationally demanding, and limits the size of the populations that we can study systematically to around $N = 10$ neurons. We return to the issue of population size in the Discussion.

3.1.2 Fitting the maximum entropy models

To fit maximum entropy models [72], we used improved iterative scaling (IIS), an algorithm that maximizes the average log-likelihood of the parameterized model to find the interaction parameters such that the moments of the resulting distribution match prescribed values [21, 39]. For homogeneous populations, the interaction parameters $h_i^{(m)}$ and $J_{ij}^{(m)}$ are identical for each neuron and neuron pair. Fitting is thus sped up considerably, as we are reduced to a three-parameter search. To explore the full range of possible triplet statistics that are consistent with prescribed single-cell and pairwise statistics, we varied the probability of synchronous triplet firing in steps of 0.001 and found the values for which the lower-order statistics and the probability of triplet firing converged within an average relative error of 1% in 1000 steps. For heterogenous populations, we implemented a slight variant of this algorithm. We fixed the triplet interaction terms $G^{(m)}$, and then used IIS to tune the first and second order interaction terms so that the lower-order statistics converged to the specified values within an average 5% error.

3.1.3 Mutual information between stimuli and firing patterns

To quantify encoded stimulus information, we compute the mutual information between the binary firing pattern $\vec{\sigma}$ and stimulus $S^{(m)}$. This is given by the following difference in entropies:

$$I = H(\vec{\sigma}) - H(\vec{\sigma}|S). \tag{3.1.5}$$

The first term denotes the entropy in the full distribution of firing patterns:

$$H(\vec{\sigma}) = -\sum_{\vec{\sigma}} P(\vec{\sigma}) \log_2 P(\vec{\sigma}). \tag{3.1.6}$$

The second term, sometimes called the noise entropy, is the average entropy of the firing patterns conditioned on a particular stimulus (each of which we assume is equally likely):

$$H(\vec{\sigma}|S) = -\sum_{m} \sum_{\vec{\sigma}} \frac{1}{2} P(\vec{\sigma}|S^{(m)}) \log_2 P(\vec{\sigma}|S^{(m)}). \tag{3.1.7}$$

Thus, the mutual information quantifies how much entropy (or uncertainty) in the firing patterns is reduced given knowledge of the stimulus identity. The benefit of using mutual information is that it is not specific to a particular neural decoder. Instead, it can be thought of as an upper bound for how much information any decoder can extract from the spiking activity of the population. Throughout this paper, we calculate mutual information exactly, without requiring any entropy estimators.

To quantify the effect of beyond-pairwise statistics, we first calculate I_{PW}, the mutual information between the stimulus and the firing patterns of the pairwise maximum entropy models. This we compare to the information in populations that include triplet statistics with the following equation:

$$\text{relative } \Delta I = \frac{I - I_{PW}}{I_{PW}}. \tag{3.1.8}$$

This quantifies the factor of increase in mutual information that is gained by populations that include triplet statistics.

3.1.4 Homogenous populations

As described above, we prescribe the firing rates and pairwise correlations in our neural populations and hold these statistics fixed while we vary triplet correlations. We first consider populations with homogenous statistics: i.e., all neurons have the same firing rates, all pairwise correlations are the same, etc. We consider various choices for the firing rates of our cells, in the range of 0.1 to 0.35 spikes per bin, with step sizes of 0.05. For spikes counted in 20 ms bins, this corresponds to spiking at 5-17 Hz, a range similar to that of average stimulus-evoked firing rates under different preparations in rodent sensory cortex [18]. We denote the difference in firing rates between the preferred and non-preferred stimulus by $\Delta\mu$, and use values of $\Delta\mu = \mu^{(2)} - \mu^{(1)} = 0.05, 0.10, 0.15$ (2.5-7.5 Hz); larger values gave highly discriminable responses regardless of the choice of higher-order correlations. We take pairwise noise correlations fixed at various values between 0 and 0.25, a range corresponding to values typically reported in, e.g., sensory and motor cortex [32]. For simplicity, we use the same values of pairwise correlations for both stimuli.

3.1.5 Heterogenous populations

For populations with heterogenous spiking statistics, we make the following choices. For concreteness, we choose firing rates and pairwise correlations from distributions reported in anesthetized cat visual cortex in response to natural movies [93]. Under the non-preferred stimulus, firing rates were taken to be exponentially distributed [14] with a median firing rate of 5 Hz as indicated in [93]. The activity under the preferred stimulus was given by adding to each cell's firing rate a Gaussian random variable with mean $\Delta\mu$ and standard deviation 0.02, where $\Delta\mu$ ranged from 0.1 to 0.15. The probability of spiking (or of two neurons spiking together) was constrained to be no less than 0.05 (2.5 Hz) to avoid convergence problems with tuning the maximum entropy models.

Spike correlations between pairs of cells were drawn from a Gaussian distribution with mean and interquartile length of approximately 0.05 each, as reported for 20 ms time bins in [93]. These values were used as the elements in the spike count covariance matrix as long as they formed a positive semidefinite matrix; if the matrix were not positive semidefinite, another random draw of values was taken. Since larger correlations have been observed in other areas and preparations [32], we also repeated this study with average noise correlations of 0.1 and 0.2 and the same variance as before. For simplicity, in all cases we continue to use the same noise correlation matrix for both stimuli.

All calculations were averaged over 24 random populations, i.e., 24 random draws from the same distributions of lower-order statistics.

3.1.6 Calculation of T_{est}

Here we calculate the length of recordings that would be required in order to estimate a key quantity in our study: the frequency with which three neurons fire within the same time bin. In particular, based on a particular experiment lasting T time bins, we want to bound the 95% confidence intervals of the relative error of the sample estimate of the frequency of cells i, j, k firing within the same time bin in the data. Suppose we want the relative error between the estimated frequency $\hat{p}$ and the true frequency $p = Pr(\sigma_i, \sigma_j, \sigma_k = 1)$ to be at most α, which means the raw error must be bounded by αp. Assuming the time bins are independent, the variance of the estimated frequency is $\text{var}(\hat{p}) = p(1-p)$. Under a normal approximation, the 95% confidence interval for the true probability p is within two standard errors above or below $\hat{p}$. This means that, in order to bound the relative error $(p - \hat{p})/p$ by α with 95% confidence, we must set the following inequality:

$$2\sigma_{\text{SEM}} \leq \alpha p, \tag{3.1.9}$$

Using the definition of the the standard error as $\sigma_{\text{SEM}} = \sqrt{\text{var}(\hat{p})/T} = \sqrt{p(1-p)/T}$, this can be rearranged into the following equation for the desired length of the experiment:

$$T \geq \frac{1-p}{p\left(\frac{\alpha}{2}\right)^2}. \tag{3.1.10}$$

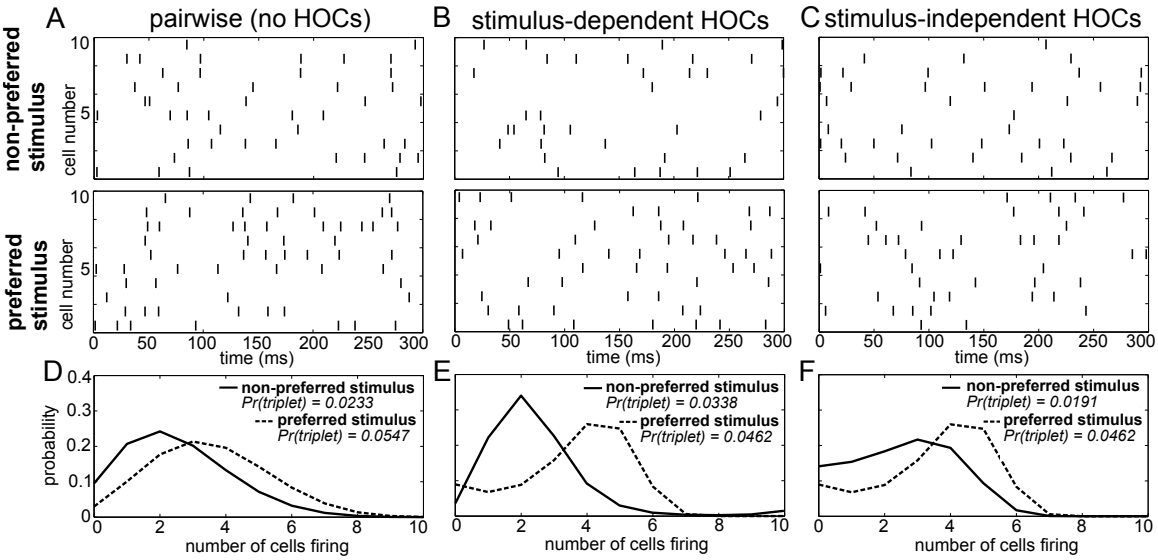

Figure 3.1: Population spike responses in three examples with different higher-order correlations. (A-C) Raster plots for three sample populations in response to two stimuli (parameters are indicated in Figure 3.3A). All three populations have identical firing rates and pairwise correlations, and differ solely in the level of higher-order correlations. (A) The "pairwise" model, which can be fully described by the firing rates and pairwise correlations. In (B), the probability of three neurons spiking simultaneously has been increased (decreased) compared to the pairwise model in response to the non-preferred (preferred) stimulus. In (C), the probability of such triplet spiking is decreased for both stimuli. (D-F) Histograms of population spike count within 20 ms time bins for the three populations. Note how triplet correlations impact the skew of these response distributions (see text).

The inequality above provides a lower bound on how many time bins are needed to estimate any triplet spike of probability of p or greater within a relative accuracy of α. In the text we call this lower bound T_{est}.

3.2 Results

Firing rates of individual neurons, and correlations between spiking activity in pairs of neurons, are the properties that are typically used in assessing neural variability and population coding. Far less is known about the role of higher-order correlations (HOCs). When and how should we expect HOCs to affect the fidelity of the neural code?

As an example, Figure 3.1 shows spike trains of three sample populations in response to two different stimuli: a preferred stimulus, eliciting relatively high firing rates, and a nonpreferred stimulus. Importantly, all three of these populations have the same firing rates and pairwise correlations for each stimulus (i.e., the same "lower order statistics"). The sole difference is in the HOCs within each population. In Figure 3.1A, the first and second order statistics are sufficient to fully characterize the responses. That is, the responses follow a

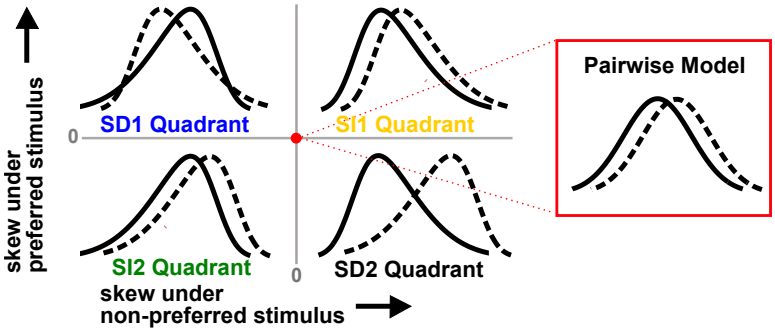

Figure 3.2: Schematic illustrating how triplet correlations skew population spike count distributions. Each quadrant corresponds to a different case of stimulus-dependent (SD) or stimulus-independent (SI) triplet spike correlations. The means and variances of the distributions are the same for all four quadrants; only the skew differs (and higher moments). In particular, note that the distributions are pulled away from each other when the non-preferred response (solid line) is positively skewed and the preferred stimulus (dashed line) is negatively skewed (i.e., the SD2 quadrant). This case gives the largest coding advantage (see text).

pairwise maximum entropy distribution [72, 125]. We refer to this simply as the "pairwise" model; it is the null case against which we compare the responses of populations with other HOCs. In Figure 3.1B, we modified the probability of three neurons firing within a short time window, keeping the lower-order statistics fixed. In particular, we changed triplet spiking probabilities in a stimulus-dependent way, so that the frequency of synchronous triplets is decreased under the preferred stimulus and increased under the non-preferred one. We refer to this difference between the true triplet spiking probability and what is predicted by the pairwise model as *triplet correlations*.

It is difficult to visualize the difference in population spiking from the raster plots alone (e.g. comparing Figures 3.1A and 3.1B). However, the implications for stimulus coding become apparent from distributions of the spike count, that is, the number of cells spiking within short time windows. For the pairwise model, these response distributions overlap strongly (Figure 3.1D). Changing the triplet correlations significantly reduces this overlap by skewing the spike count histograms away from each other (Figure 3.1E). Note that the stimulus dependence of the triplet correlations is crucial; simply changing the triplet correlations identically under each stimulus skews the spike count histograms in the same direction, preserving much of the overlap in the pairwise distributions (Figure 3.1CF). This is the key observation from this example: increasing (or decreasing) the frequency of triplets of neurons firing together corresponds to increasing (decreasing) the skew of the spike count distribution, which can shape the response distributions to significantly improve stimulus encoding. Moreover, the largest improvements arise when triplet correlations for the two stimuli are distinct.

These observations are illustrated by the schematic in Figure 3.2. The labeled regions show the four possible types of skewed distributions for the preferred and non-preferred stimulus. If the signs of the triplet correlations are the same under each stimulus, we say

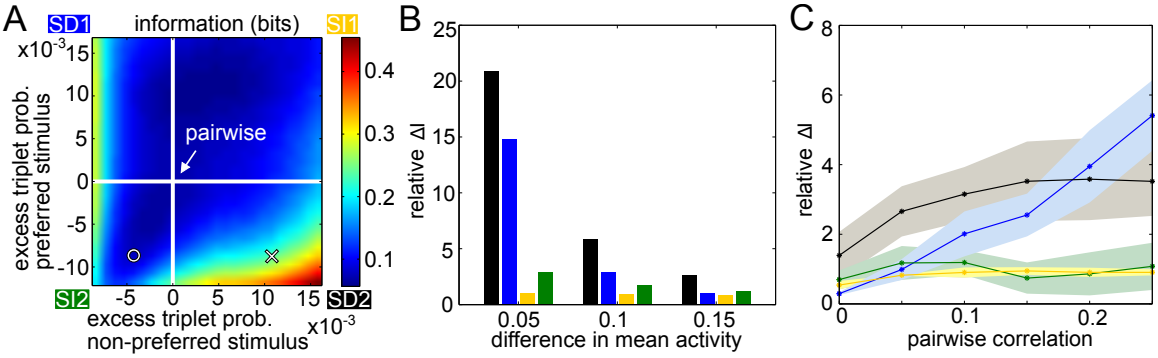

Figure 3.3: Populations with homogenous statistics. (A) Mutual information in bits as the excess triplet probability is varied for responses to the preferred and non-preferred stimuli. White lines indicate the pairwise maximum entropy model under each stimulus (shown in Figure 3.1A). The cross marker indicates the population in Figure 3.1B; circular marker for Figure 3.1C. Quadrants are labeled corresponding to the different stimulus-dependent triplet correlations (see Figure 3.2). In this example, the firing rate $\mu_1 = 0.25$ for the non-preferred stimulus, $\mu_2 = 0.35$ for the preferred stimulus, and the pairwise correlation $\rho = 0.05$ for both stimuli. (B) Relative increase in mutual information for the full model compared to the pairwise fit (see text), averaged over populations with firing rates between $0.1 - 0.35$ but keeping $\Delta\mu$ fixed to 0.05, 0.10, or 0.15. Pairwise correlations are fixed to $\rho = 0.05$. Colors correspond to the corners of the quadrants indicated in A (blue, SD1; yellow, SI1, etc.). (C) Relative increase in mutual information as a function of pairwise noise correlations, averaged over different firing rates. Shading represents standard deviation over single-cell activity, ranging from $0.1 - 0.35$ with step sizes of 0.05.

they are stimulus-independent (SI). The skews of the spike count distributions then can either be larger compared to the pairwise model (which we call the SI1 quadrant), or smaller (SI2). Alternatively, the triplet correlations may be stimulus-dependent (SD), in which case they have opposite sign for the two stimuli (SD1 and SD2). Figure 3.2 shows that stimulus-dependent triplet correlations give a greater coding benefit than stimulus-independent ones. Moreover, the greatest benefit occurs in the SD2 quadrant, where the skewed distributions are the most strongly separated; if neural populations produce responses of this type, ignoring HOCs may lead to a significant underestimation of encoded information.

Guided by this intuition, we studied the range of effects that triplet correlations can have on encoded information in populations of $N = 10$ neurons. We first considered populations with homogenous firing rates and correlations for all cells, and then moved to the heterogeneous case, where we took lower-order statistics consistent with those observed in anesthetized cat V1 [93]. In each case, we used maximum entropy models to manipulate the triplet correlations while keeping the lower-order moments fixed (see Methods).

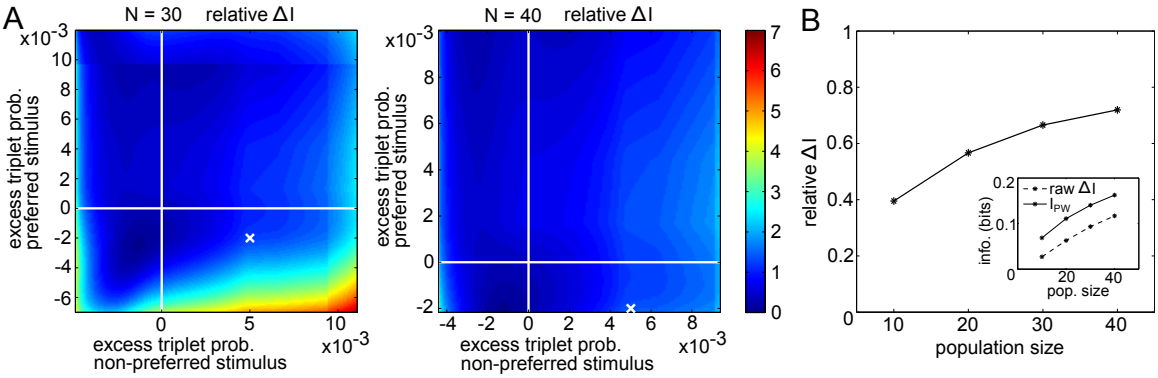

Figure 3.4: Effect of population size on mutual information in homogenous populations. (A) Relative increase in information as the excess triplet probability is varied for responses to the preferred and non-preferred stimuli, shown for populations of $N = 30$ (left) and $N = 40$ (right) cells. Note that the range of triplet correlations is smaller for the larger population. This is because, as N increases, there are tighter constraints on the possible values of triplet correlations that can be attained homogeneously across every triplet in the population, while still maintaining the same (low) predefined firing rates and pairwise correlations.Still, in the region of overlap, the strength of the impact on mutual information is similar in magnitude in both plots. Here, firing rates and pairwise correlations are fixed to: $\mu_1 = 0.25, \mu_2 = 0.35, \rho = 0.05$. Compare with the plot of raw mutual information (as opposed to the relative increase in information) in 10-cell populations that is shown in Figure 3A. (B) Relative increase in information (black curve) for fixed triplet correlations and lower-order statistics, for increasing population size. Specifically, the values of the triplet correlations were: $\kappa = .005$ for the non-preferred stimulus, and $\kappa = -.002$ for the preferred stimulus, corresponding to the cross in panel A. The impact of the triplet statistics on mutual information grows with population size. Inset shows I_{PW}, the mutual information between the pairwise distribution and the stimuli (solid line) and ΔI, the raw increase in information due to triplet correlations (dashed line) for varying population size.

3.2.1 Populations with homogenous statistics

We first investigated populations with homogenous firing statistics (i.e., equal firing rates $\mu_i^{(m)} = \mu^{(m)}$, pairwise correlations $\rho_{ij} = \rho$, etc.). This simple case illustrates how the information in neural populations can vary with triplet firing statistics, and is used as a basis for studying more realistic populations in the next section. As described above, we fixed the firing rates and pairwise correlations elicited by each stimulus, and independently varied triplet spike probabilities over the entire range for which the models can be tuned (see Methods for details). For each value of triplet correlation, we calculated the mutual information between the stimuli and the spike responses in the population. Because the population is homogenous, this process simplifies: a histogram of the total number of spikes produced in response to a stimulus (the spike count histogram) gives a complete representation of the population activity. For example, the firing patterns 1010000100 and 0011001000 are equally likely to occur because they have the same number of active neurons.

Figure 3.3A summarizes how triplet correlations can affect the level of encoded infor-

mation in a homogenous population. The axes of this plot are given by κ (Equation (3.1.3)), the excess probability of a triplet spike versus that expected in the corresponding pairwise model; they differ in scale because the range of realizable triplet spiking probabilities varies depending on the prescribed lower-order statistics. Within this plot, the cross indicates the population illustrated in Figure 3.1BE, while the circle marker represents that in Figure 3.1CF. The pairwise distributions occur along the white lines; at their intersection is the case shown in Figure 3.1AD. The asymmetry between quadrants SD1 and SD2 is due to the difference in the average firing rate evoked by each stimulus.

The overall trends in mutual information agree with the intuition developed in Figure 3.2. Mutual information is largely increased with the presence of oppositely signed triplet correlations that skew the response distributions away from each other, whereas simply increasing or decreasing the triplet correlations independent of stimulus identity does not have a significant effect. This is especially true in the SD2 quadrant. In general, the relative effects on mutual information are strongest when the population activity is noisy relative to the difference in firing rates, i.e., when firing rates are similar under the two stimuli or when the correlation between pairs of cells is large (Figure 3.3BC).

One concern is that our results for $N = 10$ neurons may not hold for larger populations. To test this, we repeated our calculations of mutual information with fixed lower-order and triplet statistics, for increasing population size (up to $N = 40$; see Figure 3.4). We found that, for fixed κ, the relative increase in information can be stable across a range of population sizes, at least for homogenous populations; in fact, it increases slightly with N. We return to the question of population size in the discussion.

3.2.2 Populations with heterogeneous statistics

To test the effect of triplet correlations on stimulus encoding in a more realistic setting, we next considered populations with heterogenous statistics. For concreteness, we chose distributions of firing rates and pairwise correlations that have been observed in mammalian V1 (see Methods, Heterogeneous Populations). The difference in the average firing rates under each stimulus is a free parameter that determines the baseline level of encoded information in the pairwise models. If the stimulus-evoked firing rates are very different, higher-order correlations would have little room to improve discrimination. We therefore set $\Delta\mu$ so that stimulus discrimination was 60% accurate on average for the pairwise models; later in this section this parameter was increased to correspond to up to 75% accuracy.

We first considered a population in which all neurons have similar stimulus tuning and hence fire preferentially to the same stimulus. This is often referred to as positive stimulus correlation [53]. As above, we varied the triplet interaction parameters of a third-order maximum entropy model (Equation (3.1.4)), re-tuning the lower-order interaction parameters each time to keep constant the population's mean activity and pairwise correlations. Specifically, triplet interaction parameters were increased or decreased to explore each of the four quadrants in Figures 3.2 and 3.3.

Since the spiking statistics are heterogeneous across the population, mutual information

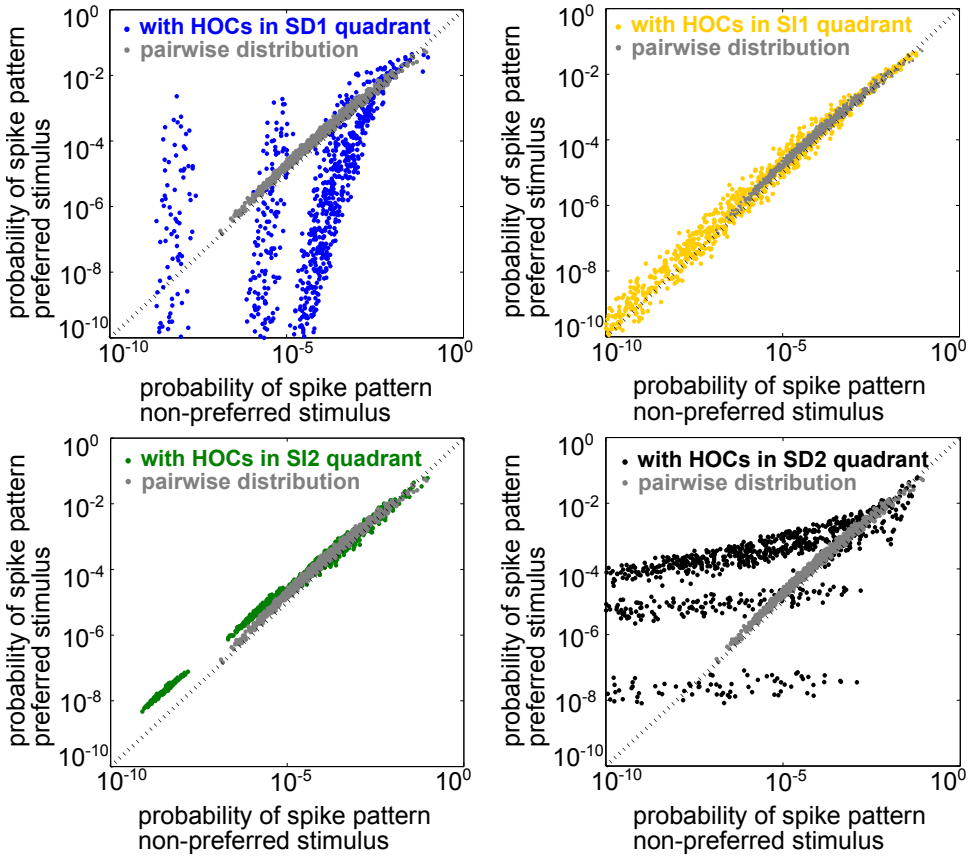

Figure 3.5: Illustration of stimulus discriminability based on spike patterns in a heterogeneous neural population. Each point represents a different spiking pattern either for the pairwise model (grey, same model for all panels) or one with triplet correlations from one of the four quadrants in Figure 3.3A. The firing rates and pairwise correlations are identical for all five populalations. The axes represent the probability of that spiking pattern under each stimulus. The triplet statistics drawn from quadrants SD1 and SD2 lead to better stimulus discrimination, since the points lie far from the identity line (see text).

must be computed using the response distributions over all spiking patterns rather than simply over spike counts, as in the homogeneous case. In this setting, the two stimuli are the most discriminable when the population spike patterns have the most different frequencies under each stimulus. To illustrate this, Figure 3.5 shows scatter plots of the probability of every firing pattern under the preferred versus the non-preferred stimulus. Good discriminability between the stimuli therefore corresponds to points lying far from the identity line. The figure shows probabilities for four example populations, each having the same lower-order statistics but differing in triplet interaction terms. The four populations correspond to the four stimulus-dependent (SD) and stimulus-independent (SI) cases introduced for homogeneous populations above. For comparison, grey points show responses for the pairwise model. The presence of triplet correlations changed spike pattern probabilities in each

case. However, these changes only significantly improved discriminability when they are stimulus-dependent. Stimulus-independent triplet correlations failed to significantly affect discrimination because they change the probabilities in a similar way for each stimulus. In sum, it appears that the same rule of thumb that we found for the homogeneous populations also applies here: stimulus-dependent triplet correlations can significantly improve population coding in cases where stimulus-independent correlations will have little effect.

To test this idea, we next computed the coding effect of triplet correlations in population models with a range of spiking statistics. Figure 3.6A shows the relative increase in encoded information compared to the pairwise maximum entropy models (Equation (3.1.8)). Because of our focus on small populations, we are able to calculate mutual information exactly without need for entropy estimators. Results were averaged over 24 random draws of firing rates and pairwise correlation matrices (see Methods, Heterogeneous Populations). Stimulus-dependent triplet correlations produced a significant effect while stimulus-independent triplet correlations did not, and again the optimal strategy that we found was to increase triplet spiking for the non-preferred stimulus and decrease triplet spiking for the preferred stimulus (region SD2). Figure 3.6B verifies that the triplet interaction term ($G^{(m)}$ in Equation (3.1.4)) has the expected effect on the averaged excess triplet spike probability (κ_{ijk}, from Equation (3.1.3)).

Example rasters from a population in region SD2 (red box in Figure 3.6A) and the corresponding pairwise model are shown in Figure 3.6C. Despite the fivefold increase in mutual information, the effect of the added triplet correlations on spike rasters appears subtle to the eye. The similarity of the pairwise firing pattern distributions and the triplet distributions can be measured by the Kullback-Leibler (KL) divergence, which calculates the average difference between the log-likelihood of each firing pattern under the triplet and pairwise distributions. A large KL divergence indicates that the pairwise model would fit the neural data poorly if the triplet model were the "true" distribution of firing patterns. The inset in Figure 3.6A shows that even a population with a fourfold increase in mutual information has a relatively low KL divergence of only 0.2, which is approximately the KL divergence between the experimental recordings and pairwise fit in salamander retina reported in [52]. Note that large KL divergence does not necessarily correlate with a large increase in information. For example, populations in region SI1 have a KL divergence of up to 0.4 but minimal effect on discrimination. This fact is also illustrated in Figure 3.5: triplet correlations modify the firing pattern probabilities (yellow points) so that they are very different from the pairwise models (gray points), but they lie distributed around the identity line, showing that the firing pattern probabilities are similar between stimuli.

Over a variety of parameter choices, stimulus-dependent triplet statistics continued to have a strong effect on information. Figure 3.6D shows the relative increase in information as the difference between the stimulus-conditioned firing rates increases, averaged over networks with different firing rates (see Methods, Heterogenous Populations). The effect of triplet correlations decreased as the stimulus-conditioned means become more different because the response distributions are less overlapping; however, region SD2 continued to strongly enhance correlations while other regions have smaller effect. Finally, panel

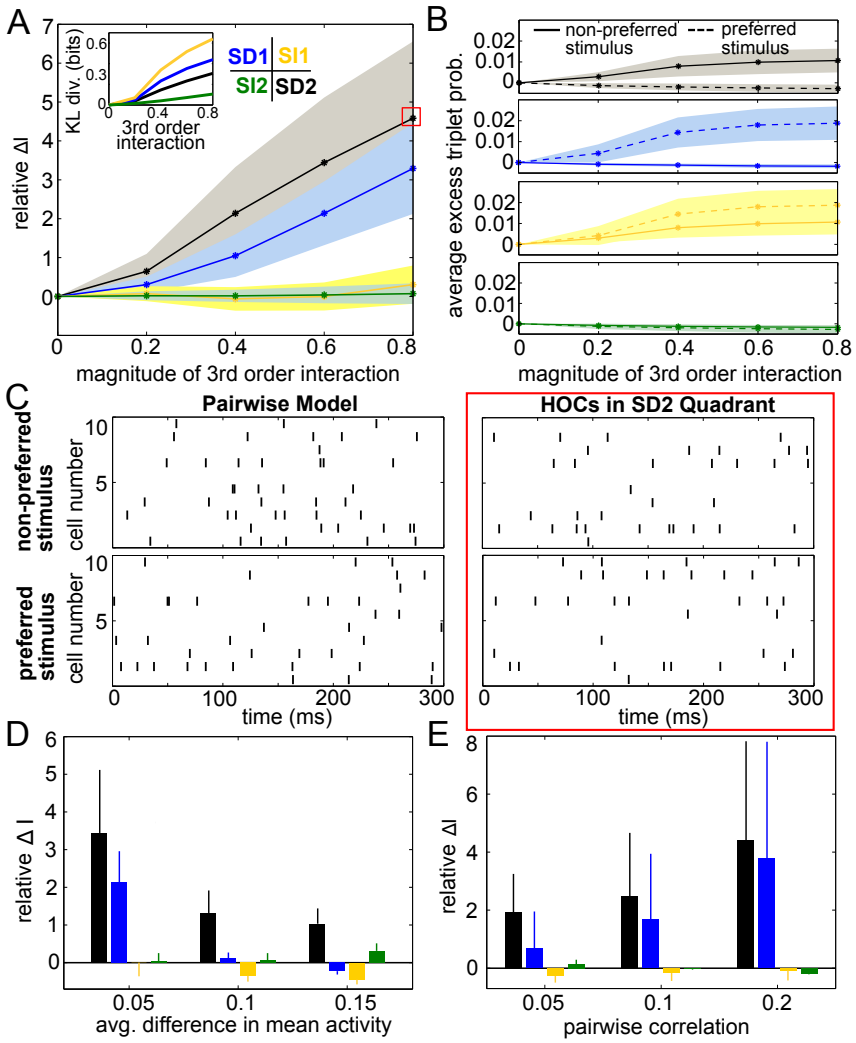

Figure 3.6: Impact of triplet correlations on stimulus coding for populations with heterogeneous spiking statistics and similar stimulus tuning for all cells. (A) Relative increase in information ΔI, averaged across 24 populations with different single-cell and pairwise statistics. ΔI is plotted against the magnitude of the third order interactions $G^{(m)}_{ijk}$, as the magnitudes of these interactions increase within the four different quadrants (see text). Colors correspond to the quadrants indicated in Figure 3.3A. Average discrimination accuracy over the 24 pairwise models is 60%. The average correlation coefficient is 0.05 and the average difference between the probability of a spike under each stimulus is 0.05. The inset shows the average Kullback-Leibler divergence in bits between the triplet models and their pairwise maximum entropy fits. (B) Excess triplet probability for the non-preferred (solid lines) and preferred (dashed lines) stimuli, averaged over all triplets. (C) Raster plots for the population marked with a red box in A, and the pairwise model. Note that the triplet correlations do not create large population-wide events immediately apparent by eye. (D) Relative increase in information over varying $\Delta\mu$ with average correlation of $\rho = 0.05$. The average baseline firing rate (to the non-preferred stimulus) was fixed to 0.05. (E) Relative increase in information as a function of average pairwise correlation. Here, the triplet interaction term is fixed to a magnitude of 0.6. Values are averaged over all firing rates (see Methods, Heterogeneous Populations). All error bars and shading represent standard deviation.

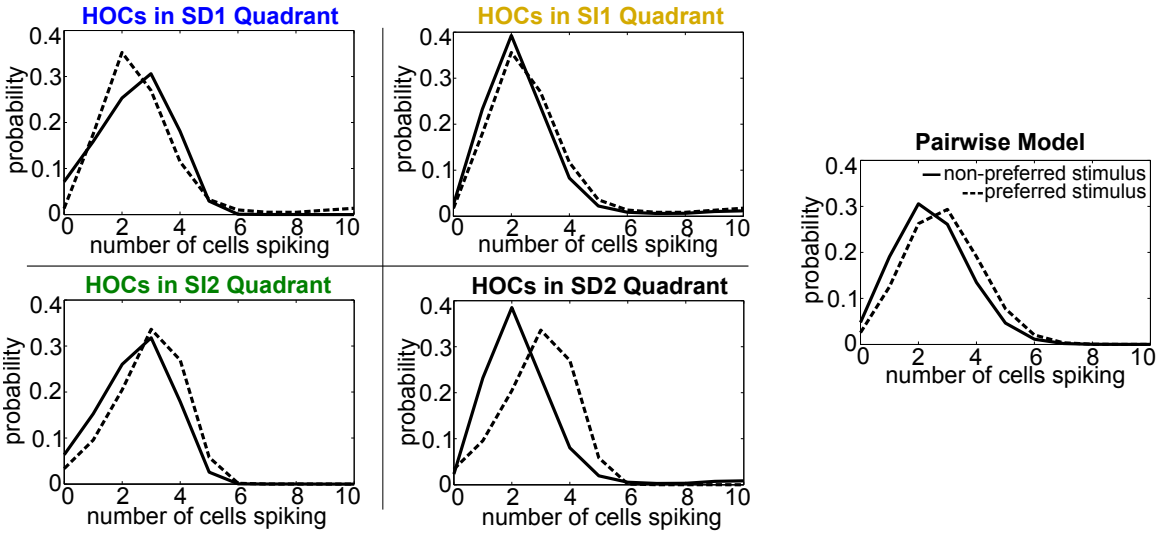

Figure 3.7: Spike count histograms for five sample populations, all of which share the same heterogeneous lower-order statistics. Panels show the pairwise model (in which $G^{(m)} = 0$) and the four different quadrants of triplet interactions ($G^{(m)} = \pm 0.8$). Parameters are taken from the red box in Figure 3.6 but are reduced from probabilities of spiking patterns to distributions of spike counts. The average pairwise correlation coefficient is $\rho = 0.05$ and the average difference between the probability of a spike under each stimulus is $\Delta\mu = 0.05$.

E shows the relative increase in information for networks with increasing pairwise correlations. In highly correlated networks, any stimulus-dependent triplet correlations (both region SD1 and region SD2) strongly increased information.

Intuitively, these effects follow the predictions from the schematic in Figure 3.2 that stimulus-dependent triplet correlations enhance discrimination by skewing the response distributions. Illustrating this, Figure 3.7 shows a reduction of the distributions to the population spike count distributions for the four quadrants in one population from Figure 3.6A. The spike count response distributions are skewed away from each other in region SD2 (and to a lesser extent in SD1), whereas stimulus-independent statistics (in SI1 or SI2) shape the distributions in the same direction. Even though the intuition in Figure 3.2 describes the effects of skewing distributions of the population spike *count*, the findings here agree with the trends shown for stimulus information based on spike *patterns*. In fact, Figure 3.8A shows the strong correlation between the raw increase in mutual information in the individual firing patterns (abscissa) and in the population-wide spike count (ordinate) for all populations in Figure 3.6A. This correlation is only guaranteed when the triplet correlations are all within the same quadrant (defined in Figure 3.2) and is not generally true for randomly generated population statistics (Figure 3.8B).

Finally, we tested whether the same effects of triplet correlations on stimulus information would occur in populations with more diverse stimulus tuning. Towards this end, we split the populations into two groups of cells, each preferring a different stimulus. Within

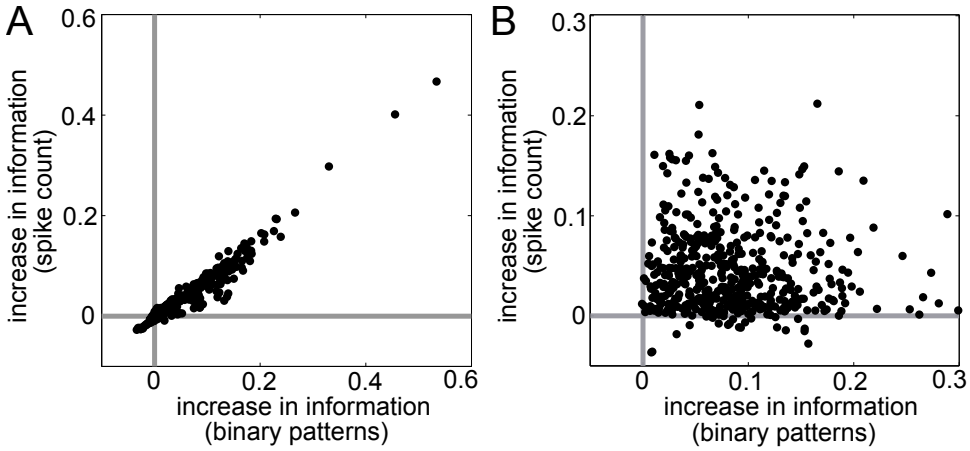

Figure 3.8: Raw increase in information relative to pairwise ($I - I_{PW}$) for (A) all populations shown in Figure 3.6A and (B) 500 populations with random interaction parameters. Abscissa represents the increase in information over all firing patterns, while the ordinate shows the increase in information in the spike count distributions.

each subgroup, all triplets had the same interaction parameter $G^{(m)}_{ijk}$. The magnitude of this triplet interaction term was varied while the sign was fixed in accordance with the four quadrants in Figure 3.3A. For example, in region SD1, $G^{(m)}_{ijk}$ for a particular triplet is positive under the preferred stimulus for neurons i, j, and k, and is negative under the non-preferred stimulus for those neurons. The triplet interaction terms for triplets composed of cells drawn from both subgroups were set to zero. That is, nonzero triplet interactions only occurred for cells with similar stimulus tuning, a choice consistent with empirical observations of triplet correlations being localized to nearby cortical microcolumns [109]. We also tried manipulating all triplets regardless of subgroup, and saw a similar increase in information for stimulus-dependent triplet correlations, but the scale of the effect was significantly smaller (data not shown).

Results were qualitatively the same as before (Figure 3.9AD). Stimulus-independent triplet correlations made little difference on the discriminability of the stimuli. Meanwhile, the largest increase in information occurred in region SD2, when the frequency of triplet spikes within each subgroup was depressed under the preferred stimulus and enhanced under the non-preferred stimulus. The changes in triplet spiking from case to case continued to have only a subtle impact on the raster plots (Figure 3.9C). Finally, stimulus-dependent correlations in region SD2 continued to have a strong effect on networks with different stimulus-conditioned firing rates and different average pairwise correlations (Figure 3.9DE).

Finally, we asked whether the same intuition that we have developed throughout this paper, about how triplet correlations impact stimulus encoding by skewing distributions of population spike counts, also applies here. Because the two subgroups differ in stimulus selectivity, we did not group their spikes into a single count; instead, we considered the

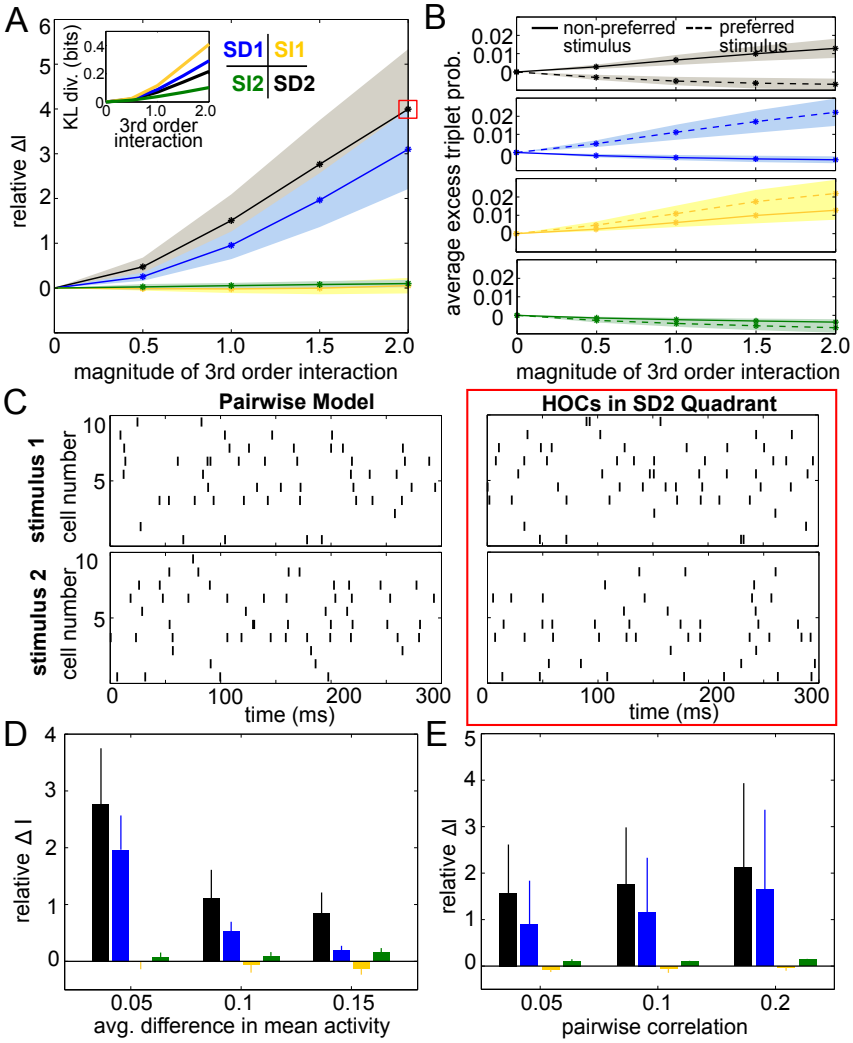

Figure 3.9: Impact of triplet correlations on stimulus coding for populations with heterogeneous spiking statistics and different stimulus tuning for subgroups of cells. (A) Relative increase in information ΔI, averaged across 24 populations with different single-cell and pairwise statistics. ΔI is plotted against the magnitude of the third order interactions $G^{(m)}_{ijk}$, as the magnitudes of these interactions increase within the four different quadrants (see text). Colors correspond to the quadrants indicated in Figure 3.3A. Average discrimination accuracy over the 24 pairwise models is 60%. The average correlation coefficient is 0.05 and the average difference between the probability of a spike under each stimulus is 0.05. The inset shows the average Kullback-Leibler divergence in bits between the triplet models and their pairwise maximum entropy fits. (B) Excess triplet probability for the non-preferred (solid lines) and preferred (dashed lines) stimuli, averaged over all triplets. (C) Raster plots for the population marked with a red box in A, and the pairwise model. Note that the triplet correlations do not create large population-wide events immediately apparent by eye. (D) Relative increase in information over varying $\Delta\mu$ with average correlation of $\rho = 0.05$. The average baseline firing rate (to the non-preferred stimulus) was fixed to 0.05. (E) Relative increase in information as a function of average pairwise correlation. Here, the triplet interaction term is fixed to a magnitude of 1.5. Values are averaged over all firing rates (see Methods, Heterogeneous Populations). All error bars and shading represent standard deviation.

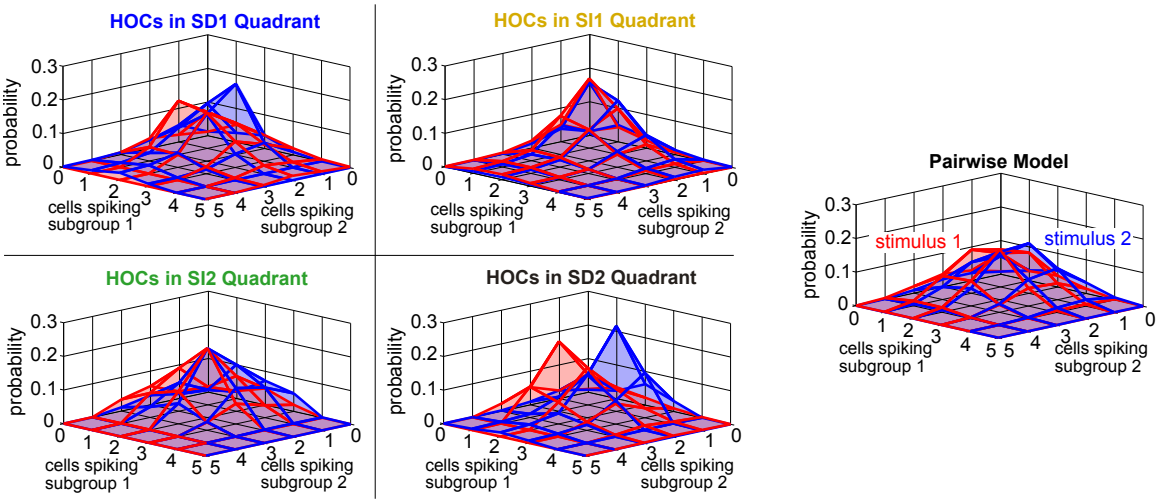

Figure 3.10: Spike count histograms for five sample populations with dissimilar stimulus tuning, all of which share the same inhomogenous lower-order statistics. Panels show the pairwise model, right, (in which $G^{(m)} = 0$) and the four different quadrants of triplet interactions, left, ($G^{(m)} = \pm 2.0$). Parameters are taken from the red box in Figure 3.9 but are reduced from probabilities of spiking patterns to spike counts. In particular, the average pairwise correlation coefficient is 0.05 and the average difference between the probability of a spike under each stimulus is 0.05.

spike counts of the two subgroups separately. The resulting two-subgroup spike count histograms are shown in Figure 3.10. These provide insight into how the triplet correlations shape the response distributions. The triplet correlations in region SD2 skew the two-dimensional response distributions away from each other, allowing the stimuli to be better distinguished. Stimulus-independent triplet correlations, however, again shape the distributions in the same way for both stimuli. We conclude that, even for our inhomogenous populations with diverse stimulus tuning, the intuition developed in Figure 3.2 describes how triplet correlations can affect the encoding of preferred versus non-preferred stimuli.

3.2.3 How much data is necessary to estimate HOCs?

Above, we have seen when and how triplet spiking statistics can have a significant impact on discrimination in neural populations. To characterize the effect of higher-order correlations in data, accurate measurements of the frequencies of spiking patterns are crucial. An essential source of difficulty in observing HOCs is the amount of data required. Since synchronous spiking events are relatively infrequent, they require longer recordings or many trials to measure. We estimated the amount of data that is required to measure the likelihood of a triplet of neurons spiking synchronously within a relative error of α by bounding the 95% confidence interval of any triplet of probability larger than $p_{\min}$ (see Methods for

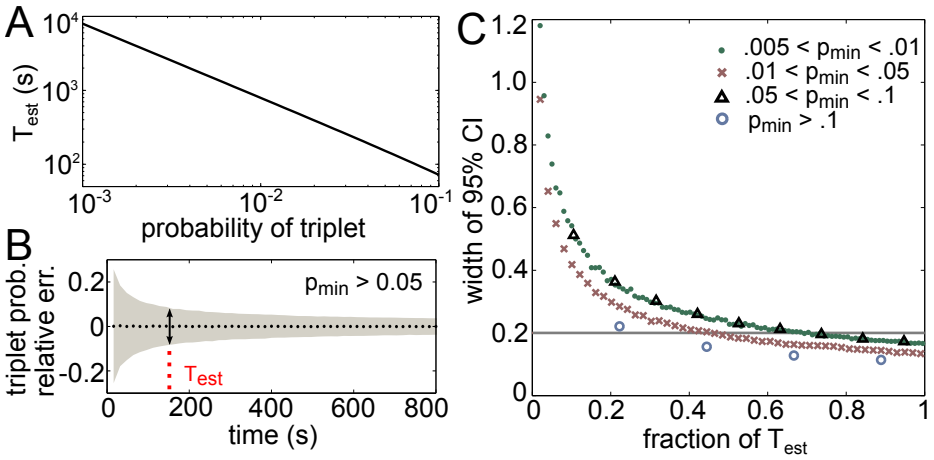

Figure 3.11: Amount of data necessary to accurately estimate triplet frequencies. In all panels, $\alpha = 0.1$ (this represents the prescribed relative error; see text). (A) T_{est} scales as a power law for small probabilities; here it is plotted in seconds, assuming 20 ms time bins. (B) Triplet probabilities were estimated from samples of 1000 triplet maximum entropy models with randomly chosen interaction parameters using different amounts of simulated data. Black dotted line shows the average relative error for all triplets with frequency greater than $p_{\min} = 0.05$ as a function of time used in the estimation (assuming 20 ms time bins). Grey funnel represents the 95% confidence interval. Dotted red line shows T_{est} calculated from Equation (3.2.1). As expected, the width of the confidence interval here (denoted by the arrows) is $2\alpha = 0.2$. (C) Width of 95% confidence interval (CI) plotted as a fraction of T_{est} for four choices of $p_{\min}$. All widths are below 2α by time $T = T_{\text{est}}$.

details). This gives the following equation:

$$T_{\text{est}} = \frac{1 - p_{\min}}{p_{\min}\left(\frac{\alpha}{2}\right)^2}. \tag{3.2.1}$$

T_{est} provides a lower bound on the number of binned activity patterns that are necessary to measure all triplets with frequencies of $p_{\min}$ or greater within a relative error of α. The choice of bin size is an important issue that we do not address here, as it does not affect these results. Figure 3.11A illustrates the dependence of T_{est} on $p_{\min}$ for a relative error of 10%, or $\alpha = 0.1$ (plotted in seconds assuming time bins of 20 ms). Note the logarithmic scaling on the axes: for example, only 220 seconds of data would be necessary to estimate the average triplet probabilities in Figure 3.9C (right panels), but over two hours are needed to estimate the least frequent triplets.

To test the tightness of the bound, we generated third order maximum entropy distributions with random interaction parameters and calculated the probability of three neurons firing synchronously from independent samples from the distribution. Figure 3.11B shows the mean relative error (black dots) and two standard errors of the mean (gray funnel) for all triplets with sample probability greater than $p_{\min} = 0.05$. At the estimate T_{est}, the width of the 95% confidence interval is around 2α, as predicted. The estimate is shown to be ac-

curate for several ranges of p in Figure 3.11C; in fact, the estimate is conservative, because probabilities larger than $p_{\min}$ will require even less data. This formula can be helpful for designing experiments to detect infrequent spiking events; or alternately, given a data set, this formula specifies which spiking patterns have sample frequencies that are large enough to be relatively accurately determined.

3.3 Discussion

The spiking patterns that neural populations produce in response to a given stimulus are variable, and this variability is correlated from cell to cell. There has been extensive work on how these correlations impact the fidelity with which a population encodes its stimuli, but the majority of this work has focused on correlations between pairs of cells. Here, we held such pairwise correlations as well as firing rates fixed and explored the impact of triplet correlations, which have recently been observed in multiple brain areas, on discriminating between preferred versus non-preferred stimuli in small populations of neurons.

Starting with homogeneous populations and working through those with progressively more diverse properties, we found that a common set of principles governed the impact of triplet correlations on the discrimination of stimuli. When triplet spike correlations were either increased or decreased relative to the level occurring in a null "pairwise model," and this increase or decrease occurred similarly for both stimuli, there was little impact on coding accuracy. However, stimulus-dependent triplet correlations significantly enhanced coding, by shaping the response distributions to reduce their overlap. In particular, when pairwise correlations were low, the greatest improvements were found when triplet spike correlations were decreased for their preferred stimulus, and increased for their non-preferred stimulus. Despite the fact that these triplet correlations are constrained by experimentally-observed lower order statistics, they were able to have a significant impact on coding. These effects can be understood intuitively as skewing the stimulus-conditioned spike count distributions away from or towards each other (as in Figure 3.2). We showed that this intuition is fruitful even when considering the information encoded in spiking patterns of heterogenous populations with more diverse tuning properties.

Thus, if triplet correlations are modulated by stimuli, models that only take pairwise statistics into account could significantly underestimate the information represented in neural populations, at least in the cases we study here. On the other hand, if triplet correlations are similarly shaped for different stimuli, we found that pairwise models were able to capture the amount information encoded relatively well. Importantly, as we have illustrated, the presence of triplet correlations can be easily overlooked despite their potentially large impact on stimulus encoding: for example, some measures of coding accuracy, such as the optimal linear estimator, do not incorporate higher-order correlations. Second, higher-order spiking statistics are difficult to observe from raster plots alone (as in Figure 3.1). Finally, even direct measurements may be impractical in some cases as long recordings are necessary to reliably sample infrequent spiking events. With an eye toward future experiments,

we provide an estimate in Equation (3.2.1) of how much data is required to accurately measure higher-order statistics within a given relative error.

The ability of stimulus-dependent triplet correlations to facilitate stimulus encoding is not guaranteed *a priori*. In fact, pairwise correlations between similarly tuned neurons enhance coding *not* when they are not stimulus-dependent, but when they are negative independent of stimulus identity [13]. This suggests a general trend for how varying kth order interactions away from the $(k-1)$th order model can affect information in populations of similarly tuned neurons: since odd moments (such as triplets correlations) shape probability mass around a distribution's mean in an asymmetric way, inducing stimulus-dependent values increases information by shaping distributions away from each other. Conversely, even moments (such as pairwise correlations) shape probability mass symmetrically, and therefore information can increase when they are negative, independent of the stimulus.

Whether neural circuits actually exploit our finding that stimulus-dependent triplet correlations can strongly improve coding remains unknown. At the level of pairs of cells, correlations in cortex are modulated by task relevance [73] and attention [33]; beyond-pairwise correlations can be modulated during motion preparation in motor cortex of awake macaques [128]. On the other hand, in [109], higher-order spiking correlations in anesthetized macaque visual cortex were found to be negative regardless of stimulus (as in region SI2 in Figure 3.3A). In agreement with our general theory, these triplet correlations had no measurable effect on encoded information.

A natural question that arises from our findings is the mechanistic origin of stimulus-dependent higher-order correlations. While common input is a prime candidate for the generation of HOCs in general, stimulus-dependence might stem from intrinsic nonlinearities such as thresholding or spike generation [90, 17, 155]. On the other hand, if triplet correlations act similarly under differing stimuli, they may have no impact on coding; intriguingly, however, they may serve a complimentary purpose such as sparsifying the neural code [109]. Moving forward, one could test experimentally how higher-order correlations are modulated during learning in animals that are trained to discriminate between similar stimuli. If the population spiking statistics adapt so that triplet correlations are strongly stimulus-dependent after training, this would be an indicator that neural systems can use higher-order correlations to their advantage to better discriminate between similar stimuli.

Our study had a number of simplifications and limitations that will be addressed in future work. First, we chose to study discrimination between pairs of stimuli, but the approach could be extended to encoding of multiple stimuli. Second, because we were interested in isolating the effect of triplet correlations, we held pairwise statistics constant from one stimulus to the next. Our intuition may generalize, however, to cases where these pairwise correlations also change with stimuli. In the schematic of Figure 3.2, increasing correlations between pairs of neurons will change the variances of the population spike count, but will not change the effect of oppositely-skewing the spike count distributions once the lower-order moments are fixed. However, it would be interesting to study varying pairwise and higher-order statistics together.

Furthermore, because maximum entropy models assume that responses are stationary

in time, they are generally used to characterize zero-lag correlations rather than more complicated temporal dependencies. While the models can in theory be extended to include spatiotemporal patterns [91], the added dimensionality is a major hurdle to overcome.

This leads to perhaps the strongest limitation of our study — we study only relatively small population sizes. This is due in part to the computational expense of tuning maximum entropy models with order N^2 parameters, while varying triplet interaction terms systematically and averaging over multiple realizations of random populations. Exact calculations of mutual information also become intractable in large populations, as the probabilities of 2^N states must be enumerated. For certain sensory coding problems, population sizes close to the $N = 10$ we used may be the relevant order of magnitude. For instance, only eight directionally selective ganglion cells encode motion at each retinal location [11]. In other applications, this number is insufficient.

We expect the intuition we developed based on the skewness of response distributions to hold for larger populations, as long as the triplet interaction parameters within clusters of similarly-tuned cells are restricted to fall squarely in one of the four quadrants in Figure 3.3A (and are therefore relatively homogenous across the subgroup). We have confirmed that, for fixed triplet statistics (excess triplet spiking κ) the relative increase in information due to triplet correlations can remain stable as N increases, at least for homogenous populations (see Figure 3.4). However, for the setting of this paper — in which we fix pairwise correlations and firing rates to relatively low values and assume that triplet correlations exist among every triplet within the population — the range of possible triplet correlations is likely to decrease with N, and this may limit their possible impact on encoded information. Thus, the present work is best thought of as investigating the impact of triplet correlations in small subpopulations sharing similar tuning preferences, as for the localized triplet correlations found in primate cortex [109].

To fully understand encoding in neural circuits, it is essential to characterize the functional interactions between different groups of neurons, and how they change with external stimuli. With this work, we have made a first step toward extending this program to incorporate beyond-pairwise spike correlations. With ongoing advances in high-density recordings and large-scale data analysis, we can look forward to an increasingly unified theory of how neural covariability at all orders impacts coding.

CHAPTER 4

Modeling reliable spiking features in neural data

What is the collective statistical structure of neural activity, and how does this structure vary from stimulus to stimulus? Answers to these questions can provide insight as to how ensembles of neurons modulate their spiking structure to flexibly perform different computations. In the previous chapter, for example, we showed how triplet spiking correlations could have a significant impact on population coding using maximum entropy techniques.

Maximum entropy models are a useful tool for determining which spiking features characterize the distribution of spiking activity. Recently, they have become a widely used method to fit neural spiking data [92, 124, 1, 129, 63, 52, 91, 141]. But, in order to probe higher-order correlations, the number of parameters increases exponentially. Perhaps this may be avoided by selecting a subset of spiking features are most significant for the neural code, but what is the criterion for inclusion in such a model?

Building off of [52], we develop in this chapter an adaptive maximum entropy model that fits only the interactions among neurons that spike together more than a fixed threshold frequency. As a consequence, the "Reliable Moment" model is less likely to fit to noise in the data, while capturing the higher-order spiking features that significantly shape the distribution of population spiking patterns. We apply this technique to extracellular recordings from mouse lateral geniculate nucleus (LGN) during presentations of visual stimuli to analyze neural covariability and how it contributes to neural coding. Visual stimuli consisted of grating patterns, i.e., movies of parallel bars drifting in a particular direction (see, e.g., Figure 1.1). We find that pairwise correlations significantly shape population activity, both sparsifying the neural code and also affecting stimulus decoding. In particular, correlations increase the decoding accuracy of grating direction. Finally, we find that higher-order correlations do not significantly shape population activity.

4.1 The Reliable Moment maximum entropy model

Any stationary distribution of binary firing patterns $\vec{\sigma}$ can be written in the following Gibbs distribution form [52]:

$$P(\vec{\sigma}) = \frac{1}{Z} \exp\left\{ \sum_i {}_i\sigma_i + \sum_{i<j} \beta_{ij}\sigma_i\sigma_j + \sum_{i<j<k} \gamma_{ijk}\sigma_i\sigma_j\sigma_k + \sum_{i<j<k<l} \delta_{ijkl}\sigma_i\sigma_j\sigma_k\sigma_l + \dots \right\}. \tag{4.1.1}$$

The above model fits all spiking features of the population. By "spiking feature," we mean the subsets of synchronously spiking cells that are present within a firing pattern. For example, the five-cell firing pattern $\vec{\sigma} = [11001]$ contains the following seven spiking features: $f(\vec{\sigma}) = \sigma_1,\ \sigma_2\ ,\ \sigma_5,\ \sigma_1\sigma_2,\ \sigma_1\sigma_5,\ \sigma_2\sigma_5$, and $\sigma_1\sigma_2\sigma_5$.

In order to determine whether various spiking features contribute to neural firing patterns, Equation 4.1.1 is usually fit only up to a fixed order of interactions (pairwise, triplet, etc.), while setting all higher terms to zero. Although useful for isolating interactions of various orders, this approach has several disadvantages. First, fitting $O(N^k)$ parameters to data becomes very computationally expensive as the order k of the models increases. Second, since we are considering binary random variables, the kth order moment is equal to the corresponding marginal probability:

$$\langle \sigma_{i_1}\sigma_{i_2}\cdots\sigma_{i_K}\rangle = Pr(\sigma_{i_1} = 1, \sigma_{i_2} = 1, \dots, \sigma_{i_K} = 1). \tag{4.1.2}$$

This means that higher-order moments are necessarily smaller than lower-order ones (i.e., $\langle\sigma_1\sigma_2\rangle \leq \langle\sigma_1\rangle$). As a result, empirical estimates of higher-order spiking features are susceptible to significant sampling error, hence maximum entropy models that have been fit to these noisy estimates may not generalize to held-out data. In general, stopping at a fixed order seems rather arbitrary, as it is possible that some highly interacting neural ensembles may act collectively, while other neurons may spike independent of the activity of their neighboring cells.

Instead, we adopt the following principle for higher-order maximum entropy models, originally proposed in [52]: we fit only the spiking features (of any order) that are "reliable" in shaping the distribution of population spiking patterns. In particular, the kth order spiking feature $f_i(\vec{\sigma}) = \sigma_{n_1}\sigma_{n_2}\cdots\sigma_{n_k}$ is reliable if its corresponding sample moment $\langle\sigma_{n_1}\sigma_{n_2}\cdots\sigma_{n_k}\rangle$ (as estimated from the data) is larger than some threshold frequency. Equivalently:

$$Pr(\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{n_k} = 1) \geq p_{\min}. \tag{4.1.3}$$

The threshold probability $p_{\min}$ is a parameter that implicitly controls the number of parameters in the model. As derived in Chapter 3, it can also be related more concretely to the accuracy of the sample estimates of the moments from the dataset. Specifically, given T samples (i.e., firing patterns within a small time window), in order for 95% of the sample probabilities to lie within a relative error α of their true values, the threshold probability

must be:

$$p_{\min} = \frac{1}{1 + T\left(\frac{\alpha}{2}\right)^2}. \tag{4.1.4}$$

Once the reliable spiking features $\{f_i(\vec{\sigma})\}_{i=1}^{w}$ have been determined from the data, we model the probability of population spiking pattern $\vec{\sigma}$ the in the following maximum entropy form:

$$P_{RM}(\vec{\sigma}) = \frac{1}{Z} \exp\left\{\sum_i h_i f_i(\vec{\sigma})\right\}. \tag{4.1.5}$$

The interaction parameters h_i are fit in order to match the expected values of f_i to their observed frequencies in the data. Because our criterion for inclusion in the model depends on the frequency of occurrence, this model will fit neural data with a sparse structure that incorporates select interactions of varying orders, rather than all interactions of a fixed order.

We call this model the Reliable Moment model. While building off of the Reliable *Interaction* model established in [52], we make two key modifications to the parameter fitting and selection criteria in order to improve the model's ability to accurately fit neural data using sparse spiking features. We will discuss the relationship between the Reliable Moment and Reliable Interaction models later in this chapter.

4.1.1 Fitting maximum entropy models with Minimum Probability Flow learning

To quickly fit the interaction parameters in the Reliable Moment model, we used a newly developed method called Minimum Probability Flow (MPF) learning [134]. In this section, we briefly explain the general idea of MPF as derived in [134] before describing how we applied it to maximum entropy models with arbitrary interaction terms, such as the Reliable Moment model.

Suppose $\mathcal{D}$ represents a finite number of samples of the underlying distribution. In the case of neural data, this typically represents a set of binary population-wide firing patterns that have been counted over small time bins (of usually 10-20 ms). MPF learning defines the following dynamics describing the deterministic evolution of distribution $P^{(t)}$ from the empirical distribution $P^{(0)}$:

$$\dot{P}^{(t)} = \Gamma P^{(t)}, \tag{4.1.6}$$

where the coupling matrix Γ has been chosen specifically so that the system has an attractor $P^{(\infty)}$ guaranteed to lie within a specific class of models (see [134] for details).

At this point, the distinction between MPF and other techniques becomes clear. Maximum likelihood learning, such as Iterative Scaling (discussed and used in Chapter 3), finds the interaction parameters that maximize the log-likelihood of the data. This is mathematically equivalent to minimizing the KL divergence between the empirical distribution $P^{(0)}$ and the attractor $P^{(\infty)}$. MPF, on the other hand, minimizes the KL divergence between the

empirical distribution and the distribution $P^{(\epsilon)}$, i.e., that distribution after an *infinitesimal* amount of time. In practice, this modification can lead to very accurate solutions with a dramatic reduction in computation time.

Taylor expanding in ϵ, the KL divergence between $P^{(0)}$ and $P^{(\epsilon)}$ can be approximated in the following close form [134]:

$$K(\vec{h}) = \frac{\epsilon}{|\mathcal{D}|} \sum_{j \in \mathcal{D}} \sum_{i \notin \mathcal{D}} g_{ij} \exp\left\{ \frac{1}{2} \left(E_i(\vec{h}) - E_j(\vec{h}) \right) \right\}, \tag{4.1.7}$$

where $E_i(\vec{h})$ denotes the energy of the jth state $\vec{\sigma}^{(j)}$ with interaction parameter vector $\vec{h}$:

$$E_j(\vec{h}) = \sum_i h_i f_i(\vec{\sigma}^{(j)}), \tag{4.1.8}$$

and g_{ij} is a binary coupling parameter that describes which non-data state $i \notin \mathcal{D}$ to be considered. The choice of g_{ij} will be discussed in the following section. The gradient is therefore given by:

$$\frac{\partial K}{\partial h_k} = \frac{\epsilon}{|\mathcal{D}|} \sum_{j \in \mathcal{D}} \sum_{i \notin \mathcal{D}} \frac{1}{2} \left[\frac{\partial E_i(\vec{h})}{\partial h_k} - \frac{\partial E_j(\vec{h})}{\partial h_k} \right] g_{ij} \exp\left\{ \frac{1}{2} \left(E_i(\vec{h}) - E_j(\vec{h}) \right) \right\}. \tag{4.1.9}$$

The minimum of $K(\vec{h})$ can be found very quickly using existing advanced gradient descent techniques implemented in MATLAB or Python toolkits; following [134], we use the minFunc package in MATLAB. Additionally, we include a small $L1$ regularization term to prevent overfitting. For exponential distributions (such as models of the form 4.1.1), $K(\vec{h})$ is convex; and also assuming infinite data, MPF will converge to the true solution.

Implementation of MPF for Reliable Moment model

In order to quickly fit the parameters of the Reliable Moment model, we implemented MPF so that it could flexibly fit maximum entropy models with any arbitrary set of w spiking features $\{f_i(\vec{\sigma})\}$. Here we describe this implementation. For clarity, throughout this section, matrices are denoted in bold (e.g., $\mathbf{A}$), vectors with arrows ($\vec{A}$), and scalars in plain font (A).

Let $\mathbf{X}$ be the $N \times T$ raster matrix, i.e., the matrix whose columns are the population firing patterns $\vec{\sigma}$ that have been observed. The set of spiking features can be summarized by the $N \times w$ matrix $\mathbf{F}$, whose ith column is the spiking pattern corresponding to feature $f_i(\vec{\sigma})$. Then the active spiking features in each observed pattern can be described by the $w \times T$ matrix $\mathbf{W^X}$, whose (i, j)th element is a 1 if interaction i is active in pattern $\vec{\sigma}^{(j)}$, and a 0 if not. In matrix form,

$$\mathbf{W^X} = \mathbb{I}[\mathbf{F^T} \cdot \mathbf{X} = \mathbf{F^T} \cdot \mathbf{1}], \tag{4.1.10}$$

where $\mathbf{1}$ is the $N \times N$ matrix of ones and $\mathbb{I}[\cdot]$ represents the indicator function (i.e., the

function that yields 1 if its argument is true, and 0 otherwise). Finally, the energies of the different observed spiking patterns are given by the following $1 \times T$ vector:

$$\vec{E}^{\mathbf{X}} = \vec{h} \cdot \mathbf{W}^{\mathbf{X}}, \tag{4.1.11}$$

where $\vec{h}$ is the $1 \times w$ vector of interaction parameters corresponding to each spiking feature. Using the notation of the previous section, the jth element of the vector $E^{\mathbf{X}}$ is equal to $E_j(\vec{h})$ from Equation 4.1.8 (i.e., the energy of the jth spiking pattern).

In order to calculate K (Equation 4.1.7), we must choose the coupling term g_{ij} to determine which non-data states are to be considered. Following [134], we set $g_{ij} = 1$ if the spiking pattern vectors i and j differ either by a single bit flip or by flipping all bits; otherwise, $g_{ij} = 0$. The single bit flip states are a standard choice to allow probability to flow from any state to any other state along an N-dimensional lattice. For neural data, which is typically sparse, the all bit flip state is added to expedite probability flow to sparser patterns.

First, consider the contribution of the single bit flip states: in particular, the states corresponding to neuron l being bit-flipped, denoted by the matrix $\mathbf{X}^l$. All rows of $\mathbf{X}^l$ are identical to $\mathbf{X}$, except for row l, which has been bit-flipped. Then the contribution to K from flipping the state of cell l for each of the observed patterns is given by the following $1 \times T$ vector:

$$\vec{K}^l_{\text{single}} = \exp\left\{\frac{1}{2}(\vec{E}^{\mathbf{X}^l} - \vec{E}^{\mathbf{X}})\right\} \tag{4.1.12}$$

$$= \exp\left\{\frac{1}{2}\vec{h} \cdot (\mathbf{W}^{\mathbf{X}^l} - \mathbf{W}^{\mathbf{X}})\right\}. \tag{4.1.13}$$

Each element in the vector corresponds to a different observed state in Equation 4.1.7 (i.e., the first summation in that equation), and each l corresponds to a non data state (the second summation). Averaging over all elements in $\vec{K}^l_{\text{single}}$ and summing over all l gives the contribution to $K(\vec{h})$ in Equation 4.1.7 that comes from single bit-flipped states.

Following Equation 4.1.9, the gradient corresponding to each term is given by the following $1 \times w$ vector:

$$D\vec{K}^l_{\text{single}} = \frac{1}{2T}\vec{K}^l_{\text{single}} \cdot (\mathbf{W}^{\mathbf{X}^l} - \mathbf{W}^{\mathbf{X}})^{\mathbf{T}} \tag{4.1.14}$$

Similarly, now consider the contribution of the full bit flip states. The matrix of patterns that are attained by flipping all of the bits in the observed patterns is $\bar{\mathbf{X}} = 1 - \mathbf{X}$. As before:

$$\vec{K}_{\text{all}} = \exp\left\{\frac{1}{2}\vec{h} \cdot (\mathbf{W}^{\bar{\mathbf{X}}} - \mathbf{W}^{\mathbf{X}})\right\} \tag{4.1.15}$$

and the gradient is given by:

$$D\vec{K}_{\text{all}} = \frac{1}{2T}\vec{K}_{\text{all}} \cdot (\mathbf{W}^{\bar{\mathbf{X}}} - \mathbf{W}^{\mathbf{X}})^{\mathbf{T}}. \tag{4.1.16}$$

Finally, the KL divergence is given by summing over all non-data states (i.e., N single bit-flipped states and one fully bit-flipped state) for each observed state, then averaging over all observed states in Equation 4.1.7:

$$K = \left\langle \sum_{l=1}^{N} \vec{K}^{l}_{\text{single}} + \vec{K}_{\text{all}} \right\rangle, \tag{4.1.17}$$

The gradient is given by:

$$D\vec{K} = \sum_{l=1}^{N} D\vec{K}^{l}_{\text{single}} + D\vec{K}_{\text{all}}. \tag{4.1.18}$$

4.1.2 Relationship to the Reliable Interaction model

The Reliable Moment model builds off of the Reliable *Interaction* model, which also identifies the frequent spiking features in population spiking data to approximate maximum-entropy models [52]. While sharing the same motivation, we have made two key modifications.

The first modification concerns how the two models are fit to data. As the number of parameters grows, maximum likelihood learning of maximum entropy models to binned spiking activity quickly becomes very costly and, therefore, slow. In the Reliable Moment model, slow parameter fitting is ameliorated by applying MPF learning, which is significantly faster than maximum likelihood learning, to models with arbitrary features. In contrast, [52] uses an arithmetic trick to allow interaction terms to be calculated directly from the observed frequencies of spiking patterns. The trick relies on the fact that in neural data, since the spiking pattern 000...0 (corresponding to no spiking activity) is relatively frequent, the partition function can be estimated accurately as $\hat{Z} = 1/P(000...0)$. From there, the subsequent interaction terms can be calculated recursively:

$$\hat{Z} = 1/P(000...0) \tag{4.1.19}$$

$$\hat{\alpha}_1 = \log P(100...0) + \log \hat{Z} \tag{4.1.20}$$

$$\hat{\beta}_{12} = \log P(110...0) + \log \hat{Z} - \hat{\alpha}_1 - \hat{\alpha}_2, \text{ etc.} \tag{4.1.21}$$

Specific interaction terms are chosen to be fit if the corresponding spiking pattern occurs above a fixed threshold frequency (the consequences of which will be discussed later in this section).

While this method allows extremely fast estimation of the interaction parameters for any N, as noted by the authors [52], it does not result in a normalized distribution. To see this, consider the toy example of a population of N cells that never spike. The probability of any spiking pattern $\vec{\sigma}$ is:

$$P(\vec{\sigma}) = \mathbb{I}\left(\sum_i \sigma_i = 0 \right). \tag{4.1.22}$$

Under the Reliable Interaction model, the partition function is:

$$Z = \frac{1}{P(0...0)} = 1. \tag{4.1.23}$$

Because only the silent state occurs, all other interactions are set to 0. As a result, the probability of any spiking pattern is:

$$P_{RI}(\vec{\sigma}) = \frac{1}{Z}\exp(0) = 1, \tag{4.1.24}$$

therefore the probabilities sum to $\sum_{\vec{\sigma}} P(\sigma) = 2^N$. Although $\hat{Z}$ is an accurate estimate of the partition function for true underlying distribution, it is not the right partition function for the model that is fit using Equations 4.1.19-4.1.21. But since the interaction terms are found using $\hat{Z}$, the model also cannot simply be renormalized; in fact, doing so would yield a uniform distribution, a poor fit for the delta function in Equation 4.1.22. This toy example clarifies why the Reliable Interaction model fails to normalize: it use the right normalizing factor for the wrong model. In contrast, since the Reliable *Moment* model is solved by direct parameter fitting techniques, it is always normalized.

The second modification changes how the significant interaction terms are selected. The Reliable Moment model includes spiking features if the corresponding moment is greater than a threshold value $p_{\min}$. For example, in a 10 cell population, the 4th order feature $f_i(\vec{\sigma}) = \sigma_1\sigma_3\sigma_5\sigma_{10}$ would be included if the probability $P(\sigma_1 = 1, \sigma_3 = 1, \sigma_5 = 1, \sigma_{10} = 1) \geq p_{\min}$. In contrast, the Reliable Interaction model determines that a particular interaction is reliable if the corresponding spiking *pattern* frequency exceeds this threshold; for example, $f_i(\vec{\sigma})$ would be included if the pattern 1010100001 occurs with frequency $p_{\min}$ or greater. The difference is whether the criterion for inclusion in the model is based on the full probability of that particular pattern, or on the *marginal* probability that cells 1, 3, 5, and 10 fire together.

While this may seem like a subtle difference, the marginal probabilities satisfy an important property: the probability of a set of cells firing is necessarily less than the probability of any subset of those cells firing, e.g.,

$$\langle\sigma_1\sigma_3\sigma_5\sigma_{10}\rangle \leq \langle\sigma_1\sigma_5\sigma_{10}\rangle \leq \langle\sigma_1\sigma_5\rangle \leq \langle\sigma_5\rangle. \tag{4.1.25}$$

This hierarchy of probabilities underlies the fact if the Reliable Moment model includes any particular spiking feature, it must also include the features corresponding to each of its subsets of cells. The spiking patterns, on the other hand, do not create such a hierarchy. As a consequence, the Reliable Interaction model may fit higher-order interactions without fitting the subgroup interactions, potentially enabling the finding of spurious higher-order interactions.

As an example, consider the following pairwise maximum entropy distribution for $N =$

3 cells:

$$P(\vec{\sigma}) = \frac{1}{Z}\exp\{-2\sigma_1 - 2\sigma_2 - 2\sigma_3 + 2\sigma_1\sigma_2 + 2\sigma_1\sigma_3 + 2\sigma_2\sigma_3\}. \tag{4.1.26}$$

The probability of each pattern can be found analytically:

$$P(\vec{\sigma}) \propto \begin{cases} 1 & \text{0 or 3 cells fire} \\ \exp\{-2\} & \text{1 or 2 cells fire} \end{cases} \tag{4.1.27}$$

with a partition function of $Z = 2 + 6e^{-2}$. Therefore the probability of all or no neurons firing is 0.3556, while the probability of one or two neurons firing is 0.0481. We generated 10000 samples from Equation 4.1.26 to test how well the Reliable Moment and Reliable Interaction models could infer the underlying interaction parameters. For each model, we set the threshold probability to $p_{\min} = 0.05$.

First, consider the Reliable Interaction model. Since the frequency of spiking patterns with one or two spikes falls below $p_{\min}$, the first and second-order interaction parameters are set to zero. The triplet interaction term then reduces to:

$$\hat{g}_{123} = \log P(111) + \log \hat{Z}. \tag{4.1.28}$$

Therefore the values of the interaction parameters are given in the following table:

$f_i(\vec{\sigma})$	σ_1	σ_2	σ_3	$\sigma_1\sigma_2$	$\sigma_1\sigma_3$	$\sigma_2\sigma_3$	$\sigma_1\sigma_2\sigma_3$
h_i	0	0	0	0	0	0	0.0493

The resulting model is able to fit the frequency of the patterns 000 and 111, but is a poor fit for any of the patterns that were not included in the fitting process (Figure 4.1; again, note that this model is unnormalized). And because the first and second order parameters are incorrectly assumed to be zero, a spurious triplet interaction term is inferred.

On other hand, the moments of the distribution are all greater than $p_{\min}$, so *all* interactions are fit under the Reliable Moment model. MPF finds the following parameters:

$f_i(\vec{\sigma})$	σ_1	σ_2	σ_3	$\sigma_1\sigma_2$	$\sigma_1\sigma_3$	$\sigma_2\sigma_3$	$\sigma_1\sigma_2\sigma_3$
h_i	-2.0289	-1.9785	-2.0225	2.0184	2.0373	1.9869	0.0000

Not only are the first and second order terms accurate as compared to Equation 4.1.26, but this triplet term is correctly set to zero – even though this term was included in the fitting. This contrasts the Reliable Interaction model, which incorrectly inferred triplet interactions in the data. Moreover, in this example, the Reliable Moment model also provided an accurate fit of the probabilities of all spiking patterns (Figure 4.1).

In sum, the Reliable Moment model builds off of the Reliable Interaction model while modifying how the model is fit to data and how interactions are chosen for inclusion. These modifications result in two advantages of using the Reliable Moment model to fit neural population data to higher-order maximum entropy models: (1) it is normalized, and (2) it is less likely to infer spurious higher-order terms.

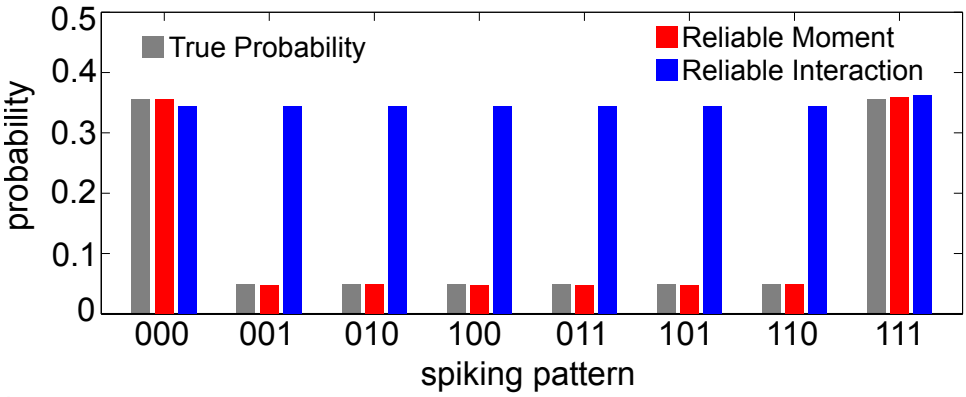

Figure 4.1: Example comparing the Reliable Interaction and Reliable Moment models. Probabilities for different spiking patterns of the three-cell system defined by Equation 4.1.26 (gray), as well as the Reliable Moment (red) and Reliable Interaction (blue) model fits to sampled data. While the Reliable Interaction model only fits the spiking patterns that occurred with frequency over $p_{\min}$, the Reliable Moment model accurately fits the whole distribution. Also note that the Reliable Interaction model is not normalized.

4.2 The statistical structure of spiking patterns in mouse lateral geniculate nucleus

We used the Reliable Moment (RM) model and other maximum entropy techniques to study collective population spiking activity in the lateral geniculate nucleus (LGN). The LGN is a structure in thalamus that relays visual information from the retina to visual cortex for more complex information processing. Although its response properties are modulated by significant cortical feedback [6, 96, 103], the activity of LGN is theorized to be an efficient recoding of retinal activity [38, 12]. As in retina [60, 130, 124], correlated spiking is observed between pairs of LGN cells [71]. These precise temporal correlations lead to a 20% increase in information about stimulus pixel strength in the output of cell pairs [37]. Less is known about the population-wide spiking structure in LGN, although there have been some hints of coordinated spiking in adjacent areas: higher-order correlations, for example, have been observed in both retina [124, 52] and cortex [109, 100, 79].

To better understand this collective activity, we analyzed population spiking patterns of neurons in the LGN of awake mice during the presentation of grating stimuli drifting in different directions at multiple spatial and temporal frequencies. We fit independent, pairwise, and RM models to the distributions of spiking patterns in order to determine which spiking features shape the population activity. We found that pairwise correlations have a large impact on the distribution of spiking patterns, with implications for stimulus decoding; on the other hand, although we find evidence of some higher-order spiking interactions, they had little impact on population activity.

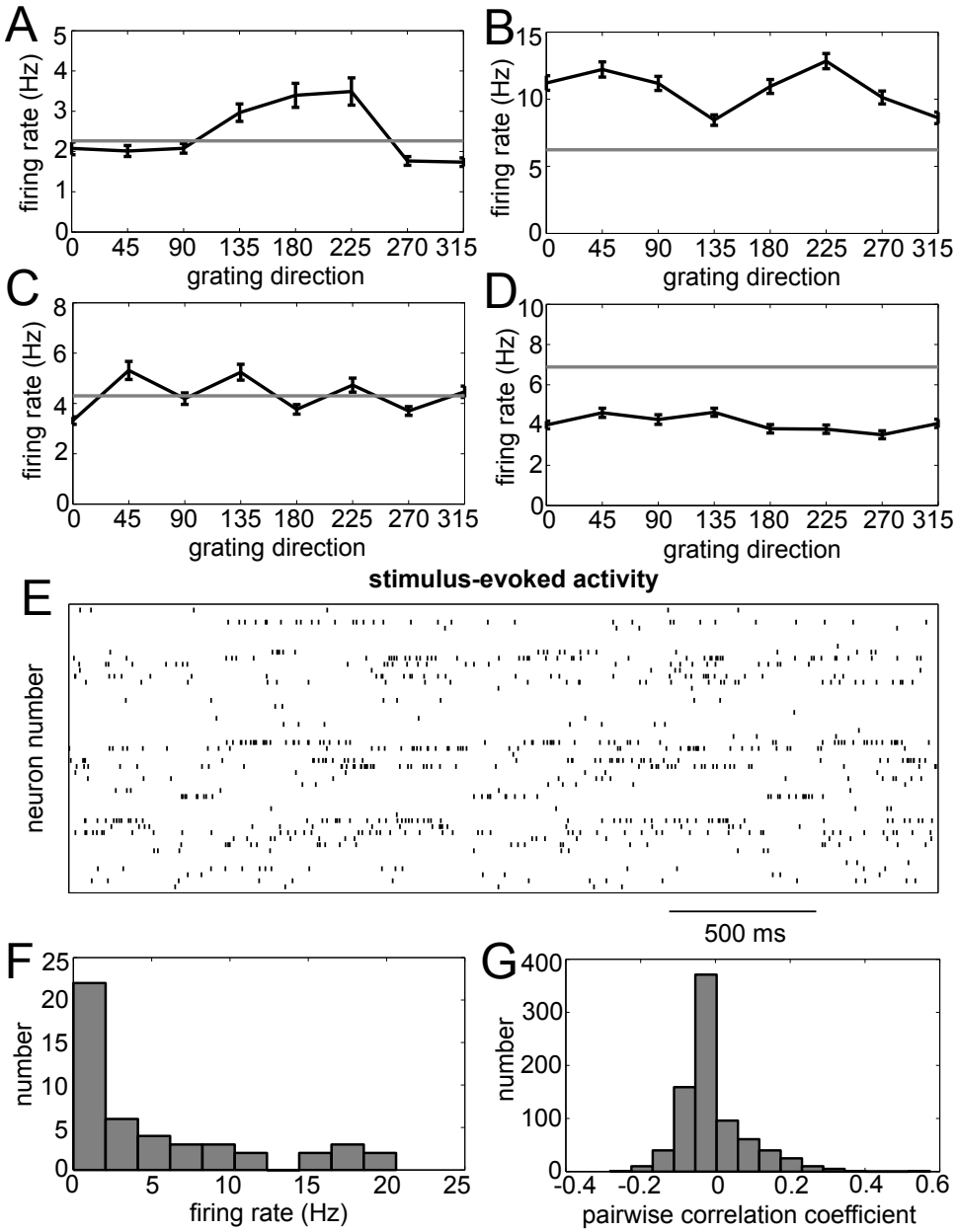

Figure 4.2: Tuning curves and raster plots of recorded LGN cells. (A-D) Stimulus direction tuning of four example LGN cells. Spontaneous firing rates (i.e., in response to zero-contrast stimulus) are shown in grey, while trial-averaged firing rates for different grating directions are plotted in black (errors depict standard error of the mean over all trials). These examples illustrate (A) direction-tuning and (B) orientation-tuning, as well as weaker tuning in other cells (CD). Also shown are raster plots of spiking activity during (E) one trial of 45 degree gratings with spatial frequency of 0.025 deg^{-1} and temporal frequency of 1 Hz (bottom panel). (F) Histogram of the firing rates over the entire population for the same trial. (G) Histogram of correlation coefficients for all cell pairs (calculated for 20 ms windows), also for same trial. All data from mouse M34.

4.2.1 Experimental procedures

All data was collected by Severine Durand of the Allen Institute for Brain Science. Extracellular currents were recorded from the lateral geniculate nucleus of awake, behaving mice using a NeuroNexus 6-shank probe. Each shank contained 8 channels spanning 140 microns, and shanks were spaced 200 microns apart. Spike waveforms were sorted using the graphical cluster cutting package Klusters [65].

During recordings, mice were presented with visual stimuli for intervals of three seconds, alternated with one second of zero-contrast (gray) stimuli. After every 15 stimuli, 4 seconds of zero-contrast stimuli were presented. Visual stimuli comprised gratings moving in directions from 0 to 360 degrees, with 45 degree spacings, and at six spatial frequencies (0.025 – 0.8 Hz) and five temporal frequencies (1 – 15 deg^{-1}). For all grating stimuli, the contrast was fixed to 0.8.

After spike sorting, neurons were included in our analysis only if they spiked significantly above their baseline firing rate (in response to the zero-contrast input) for any orientation and spatiotemporal frequency; this corresponds to 47 cells.

4.2.2 Direction tuning and population activity in LGN

Traditional theories of population coding rely on the "tuning" of individual neurons, i.e., which stimulus properties drive each neuron to spike significantly above its baseline firing rate. As an illustration, Figure 4.2A-D shows the average firing rates of four example LGN cells in response to different grating directions (black curve) as well as the spontaneous firing rate in response to a zero-contrast gray screen (grey curve). Some neurons can be well-described by their tuning properties: Figure 4.2A shows an example of a direction-selective neuron, which spikes preferentially in response to gratings moving at around 180 degrees, and Figure 4.2B shows an orientation-tuned neuron that fires when the stimulus is moving at either 45 or 225 degrees. However, some neurons show weak direction tuning properties that are difficult to interpret (Figure 4.2CD).

Instead of studying single-cell responses, we use maximum entropy methods to analyze the spiking activity of the population as a whole. Figure 4.2E shows an example raster plot over a three-second trial during which a 45 degree grating with spatial and temporal frequencies of 0.025 deg^{-1} and 1 Hz, respectively. Pairwise correlation coefficients ranged over strong values between -0.2 and 0.3 (Figure 4.2FG). To determine which spiking features shape such population activity, we binned rasters in 20 ms windows (following [52]), and modeled the distribution of simultaneous population spiking patterns using independent, pairwise, and Reliable Moment maximum entropy models. For each direction of the grating stimulus, the spiking patterns were randomly partitioned into either a training set or a test set. However, the test and training sets mixed responses to stimuli of different spatial and temporal frequencies. This is equivalent to treating the frequencies of the gratings as *nuisance parameters*, that is, examining how information about stimulus direction is generalized across spatial and temporal frequencies.

Model parameters were fit to the "training" half of the data of each stimulus direction through MPF learning. The remaining half of the data (the "test" data) was reserved for cross-validation of the models. Specifically, we estimated the Kullback-Leibler divergence (Equation 4.2.9) between the empirical probabilities and the predicted model probabilities of the spiking patterns in the test data. This technique allows us to test whether the models fit the true distribution of spiking patterns underlying the data, or rather if they overfit to noise within the training data.

4.2.3 Estimating the partition function

Through MPF learning, we can rapidly fit interaction parameters for arbitrary maximum entropy models. But in order to predict the probability of a particular firing pattern, we additionally need the normalizing factor or partition function Z. For large N, it is impossible in practice to compute Z directly as it requires a summation over all 2^N possible spiking patterns.

To get around this, we follow [63] and use a technique to estimate the partition function called the "Good-Turing" estimate. Originally discovered by Alan Turing and I. J. Good as part of their cryptanalysis of the Enigma code in World War II, the Good-Turing estimate is a useful tool of frequency estimation with broad applications [57, 63]. This unbiased estimator [57] gives a simple approximation of the probability mass M of all unobserved states as the fraction of distinct spiking patterns that occur exactly once in the dataset $\mathcal{D}$:

$$M = \frac{\#\text{ states only observed once}}{\text{total }\#\text{ observations}}. \tag{4.2.1}$$

The Good-Turing estimate is accurate for heavy-tailed distributions in which the number of states observed only once tends to be large [57], a common statistical feature of neural data. The estimator's applicability to Ising-like models of neural population spiking patterns has been directly tested in [63].

Following [63], we can rewrite the partition function as follows:

$$Z = \sum_{\vec{\sigma}} \exp\left\{\sum_i h_i f_i(\vec{\sigma})\right\} \tag{4.2.2}$$

$$= \sum_{\vec{\sigma}\in\mathcal{D}} \exp\left\{\sum_i h_i f_i(\vec{\sigma})\right\} + \sum_{\vec{\sigma}\notin\mathcal{D}} \exp\left\{\sum_i h_i f_i(\vec{\sigma})\right\} \tag{4.2.3}$$

$$= X + Y. \tag{4.2.4}$$

Here, X represents the contribution of the observed states to the partition function, and Y the contribution of the non-observed states. Using these definitions, M can also be written

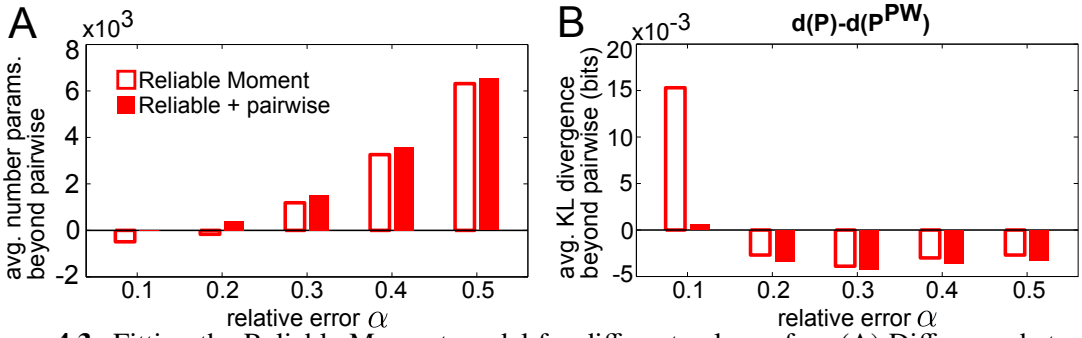

Figure 4.3: Fitting the Reliable Moment model for different values of α. (A) Difference between the number of parameters in the higher-order model and the pairwise model, averaged over stimuli, for the pure Reliable Moment (hollow red bars) and the hybrid Reliable Moment + pairwise model (solid red bars). (B) Difference between approximate KL divergences for higher-order versus pairwise models: i.e., $d(P^{RM}) - d(P^{PW})$, averaged over stimuli.

as the sum of the probabilities of unobserved states:

$$M = \sum_{\vec{\sigma} \notin \mathcal{D}} P(\vec{\sigma}) \tag{4.2.5}$$

$$= \frac{1}{Z} \sum_{\vec{\sigma} \notin \mathcal{D}} \exp\left\{ \sum_i h_i f_i(\vec{\sigma}) \right\} \tag{4.2.6}$$

$$= \frac{Y}{Z} = \frac{Y}{X+Y}. \tag{4.2.7}$$

Rewriting gives $Y = \frac{M}{1-M} X$, and combining with Equation 4.2.4 results in the following equation for the partition function:

$$Z = \frac{1}{1-M} X. \tag{4.2.8}$$

With this formula, we are able to estimate the probabilities of spiking patterns for various maximum entropy models without needing brute force calculation, which would require enumerating over all 2^N states.

4.2.4 Choosing a Reliable Moment model

Fitting the Reliable Moment model requires an arbitrary choice of threshold: in particular, the cutoff probability $p_{\min}$ determines which interaction terms will be included in the model. Large values of $p_{\min}$ may result in a model that is too underdetermined to accurately fit the population activity. On the other hand, smaller values of $p_{\min}$ increase the number of parameters, which risks overfitting the model to noisy estimates of infrequent spiking features that may not generalize in cross-validation.

To find the model that optimizes this tradeoff, we fit the Reliable Moment model to the training data for a range of values of the expected relative error α (which can be directly related to $p_{\min}$ using Equation 4.1.4). For comparison, we also fit the pairwise maximum entropy models. Figure 4.3AB shows the difference between the number of parameters in the Reliable Moment model and the pairwise maximum entropy model (hollow red bars) for both mice.

The ability of any model distribution $\tilde{P}(\cdot)$ to fit the empirically-observed distribution $P^{\text{emp}}(\cdot)$ is quantified by the KL divergence:

$$D_{KL}(P^{\text{emp}}, \tilde{P}) = \sum_{\vec{\sigma}} P^{\text{emp}}(\vec{\sigma}) \log\left(\frac{P^{\text{emp}}(\vec{\sigma})}{\tilde{P}(\vec{\sigma})}\right). \tag{4.2.9}$$

The empirical distribution $P^{\text{emp}}(\cdot)$ simply counts the fraction of times each spiking pattern occurs in the held-out test data, rather than the training data. This cross-validates the model, which is necessary to ensure that $\tilde{P}(\cdot)$ is not overfitting to noise within the training data.

But because we have a finite amount of data, the empirical distribution is a noisy approximation to the underlying true distribution. In particular, spiking patterns that occur only once in the data will have the most biased empirical probabilities. Therefore, instead of directly calculating the KL divergence, we used the following approximation:[1]

$$d(\tilde{P}) = \sum_{\vec{\sigma}\in\mathcal{D}'} P^{\text{emp}}(\vec{\sigma}) \log\left(\frac{P^{\text{emp}}(\vec{\sigma})}{\tilde{P}(\vec{\sigma})}\right) \tag{4.2.10}$$

$$\approx D_{KL}(P^{\text{emp}}, \tilde{P}), \tag{4.2.11}$$

where $\mathcal{D}'$ is the set of spiking patterns that occur more than once in the held-out data; $d(\tilde{P})$ provides an estimate for how well the different maximum entropy models are able to capture the data. We also calculated the full KL divergence using Equation 4.2.9; while the overall magnitude was larger, the qualitative relationship between the fit of the independent, pairwise, and RM models did not change (data not shown).

Figure 4.3CD shows the difference between the pairwise and Reliable Moment model KL divergences, $d(P^{RM}) - d(P^{PW})$, averaged over all stimuli (hollow red bars). The RM model improved the fit compared to the pairwise distribution for any $\alpha > 0.1$. We further hypothesized that some pairwise interactions, although small, may still contain important information about the collective activity of the population, a point we shall return to in the discussion. We therefore also fit a variant of the RM model that includes *all* pairwise interactions and only the reliable higher-order ones (solid red bars in Figure 4.3). This hybrid Reliable Moment + pairwise model was able to best capture the spiking activity for

[1] A variant of this equation was originally used as an approximation of the KL divergence in [52] when testing the Reliable Interaction model. The only difference is that [52] takes the absolute value of the logarithm in Equation 4.2.10. Their justification is that since the Reliable Interaction model is not normalized, the absolute value is necessary to penalize models that underestimate probabilities. Since our model is normalized, we used the Equation 4.2.10 to approximate the KL divergence.

$\alpha = 30$. Therefore this is the higher-order model that we will use throughout the remainder of this section. (For clarity we will refer to it as simply the RM model.)

As a comparison, we also analyzed the fit of the triplet maximum entropy model to the training data. Although it had more than 13,000 extra parameters, the triplet model actually performed worse than both the pairwise and RM models on the held-out data (data not shown). As mentioned before, this is because the triplet model fits noisy triplet spiking features in the training data that do not accurately reflect the test data. This fact highlights the strength of the RM model in identifying the higher-order interactions that are sufficiently accurately estimated to generalize to more data.

4.2.5 Statistical interactions in mouse LGN

To determine which spiking features shape the collective population activity, we compared the independent, pairwise, and RM maximum entropy models to the empirically observed frequencies of the held-out spiking patterns. Figure 4.4A shows the cross-validated fits of the models to the test data for all stimulus directions. The independent model is a poor fit, having a relative high KL divergence of 0.4., and the large reduction in the KL divergence in the pairwise models reveals that pairwise correlations are necessary to capture population activity in LGN.

How do pairwise correlations shape population activity? Figure 4.4B shows the empirical versus model probabilities of each observed spiking pattern under 90 degree stimuli. The independent model systematically underestimates the states that are most frequently observed in the data, which correspond to three or fewer cells spiking synchronously. This illustrates one way that pairwise correlations shape population activity: by sparsifiyng spiking responses so that fewer cells spike at the same time [109].

On the other hand, while the RM model does fit the held-out data better than the pairwise model, the size of the improvement is slim (Figure 4.4A). The RM model for this dataset includes approximately 1500 parameters more than the pairwise model; these correspond to a thousand triplets, hundreds of quadruplets, and a few quintuplets (Figure 4.4C.This represents a small fraction of all *potential* higher-order interactions. For example, the 1240 triplets in Figure 4.4C makes up just under 10% of all 13,244 possible triplets in the recorded population (Figure 4.4D). These extra 1500 higher-order parameters accounts for a reduction of approximately 1.5% of the KL divergence, whereas the 946 pairwise interactions reduce the KL divergence by 70% compared to the independent model. Therefore, we found that higher-order interactions did not significantly shape the spiking structure of population spiking in mouse LGN.

4.2.6 Correlated spiking impacts direction decoding

In the previous section, we found that pairwise (but not higher-order) correlations significantly shape the statistical structure of population activity in mouse LGN. These pairwise correlations contribute in part to sparser population responses. Much work has shown that

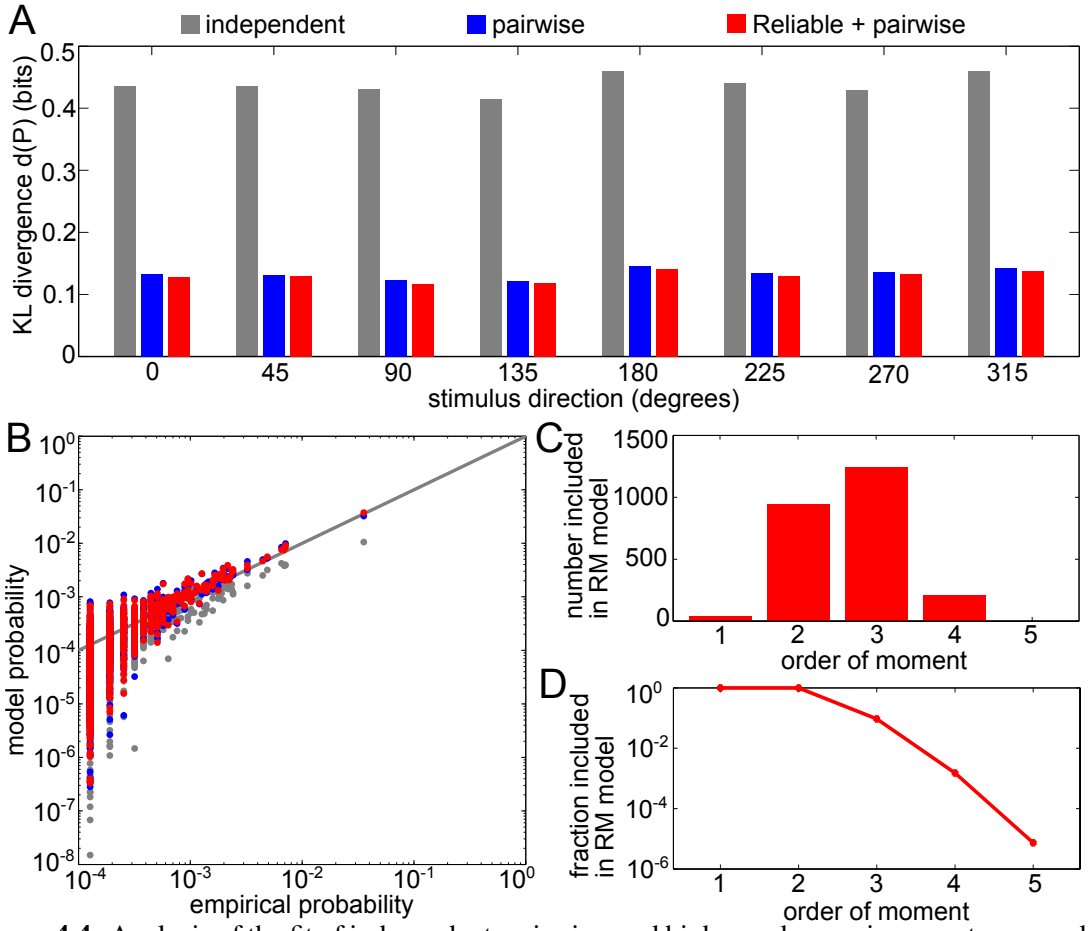

Figure 4.4: Analysis of the fit of independent, pairwise, and higher-order maximum entropy models to LGN recordings: mouse M34. (A) Cross-validated KL divergence $d(P)$ of independent (gray), pairwise (blue), and Reliable Moment (red) models, plotted for different stimulus directions. (B) Characteristic example of the word-by-word fit of the independent, pairwise, and RM models (here, the stimulus direction is 90 degrees). Each point plots the empirical probability of a binary firing pattern in the held-out data set against the probability predicted by each of the models. (C) Number of interaction terms included in the RM model given a 90 degree stimulus, from single cell terms to quintuplets. (D) Fraction of all possible kth order interactions that are included in the RM model under a 90 degree stimulus.

correlated activity can also have a significant effect on how much stimulus information can be encoded into neural populations [154, 2, 13, 114, 112, 135, 36, 70, 127].

Towards this end, we compared stimulus decoding based on the pairwise versus independent maximum entropy models. For each pair of stimuli, we used maximum likelihood estimation (as described in Chapter 2) to decode stimulus direction from each spiking pattern in the held-out data. That is, for each spiking pattern in the test data, we computed the likelihood of that spiking pattern based on the independent, pairwise, or Reliable Moment models (which have been fit to the training data). We then calculated the accuracy of

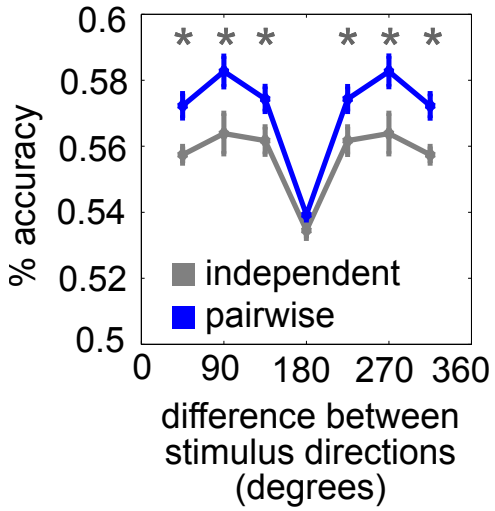

Figure 4.5: Decoding stimulus direction from spiking patterns. (A) Percent accuracy of two-stimulus maximum likelihood decoding shown for both the independent (gray) and pairwise (blue) maximum entropy models. Each point represents the average over all stimulus direction pairs at a fixed difference. Error bars indicate standard error. Asterisks represent statistical significance across stimulus pairs (i.e., a p-value of 0.05 or less).

the maximum likelihood estimator, which "chooses" the stimulus that gives greater likelihood of each sample, compared to the stimulus that actually elicited that spiking pattern. This is the *decoding* perspective of Section 2.8, which is analogous to asking whether a downstream circuit requires knowledge of spiking correlations in the population in order to accurately decode stimulus direction.

Figure 4.5 shows the average percent decoding accuracy as a function of the difference between the directions of the stimuli, plotted for both the independent (gray) and pairwise (blue) models. The coding accuracy dips for stimuli moving in opposite directions (180 degrees), likely due to the presence of cells that are selective to direction regardless of orientation (Figure 4.2C). For all other stimulus pairs, the presence of the pairwise correlations improves decoding of spiking patterns by a few percentage points (asterisks indicate statistical significance), showing that the pairwise correlations are marginally beneficial for coding.

4.3 Discussion

How "collective" is the spiking activity of populations of neurons beyond their individual tuning properties? And how does collective activity contribute to population-based stimulus decoding? To answer these questions, we developed an adaptive maximum entropy model to identify and fit the significant higher-order interactions of neural spiking. Unlike related maximum entropy models that impose an arbitrary limit on the order of included interactions, the Reliable Moment (RM) model selects the sparse subset of spiking features

that are most frequently observed in the data. Not only does this approach require fewer parameters, but it also ensures that the model is informed by the most accurately measured features in the data, and thereby less susceptible to overfitting to noise.

The RM model builds off of the Reliable Interaction model [52] with two key modifications. First, we adapted Minimum Probability Flow learning [134] to fit arbitrary interaction parameters extremely quickly. This, in combination with the Good-Turing approximation of the partition function [57, 63], guarantees that the resulting model is a normalized distribution. Second, the RM model uses moments, or equivalently the marginal probabilities of subsets of cells spiking together, instead of firing pattern frequencies as a criterion for inclusion in the model. Because moments are naturally hierarchical (i.e., $P(A, B) \leq P(A)$), this means that for any higher-order spiking feature fit by an RM model, all sub-features will be fit as well. As we showed in a simple example, this second modification reduces the likelihood of finding spurious higher-order correlations in the data.

We tested the RM model on extracellular recordings in awake mouse LGN upon repeated presentations of gratings of varying direction and spatial and temporal frequencies. By comparing the fits of independent, pairwise, and RM models to held-out test data separately for each stimulus direction, we were able to determine which spiking features were necessary to describe the structure of the synchronous population spiking activity. We found that, although the RM model was consistently able to better fit the spiking data than models that did not include higher-order correlations, the magnitude of the improvement was minimal. However, this may be due to mixing neural responses from different spatiotemporal frequencies, which could average away higher-order correlations if they vary for different frequencies. We are currently working to collect more data with a fixed spatiotemporal frequency in order to further explore this possibility.

Still, the lack of evidence for higher-order correlations in our data was surprising, given that they have been observed both in retina [52] and in cortex [109, 100, 79]. As we discussed in the previous paragraph, one potential reason for this discrepancy may be due to the fact that we treated spatiotemporal frequency as nuisance parameters. Or, it may be a result of methodological differences in how higher-order correlations are quantified in different studies. An alternate and intriguing possibility is that LGN may actively decorrelate the higher-order structure for further information processing. Although *temporal* decorrelation (i.e., flattening of the autocorrelation function) during natural movies has been linked to increasing the efficiency of the retinal code [38, 12], it is unclear what the benefit would be for decorrelation of the higher-order *population* spiking activity in short time windows. This will be the topic of future research.

On the other hand, pairwise correlations strongly shaped the distribution of spiking patterns for all stimuli. This covariability made population responses more sparse than predicted by the independent model, an effect that has also been observed in cortex [109]. We found that correlations improved the decoding accuracy of stimulus direction based on spiking patterns by a few percent. Future work will continue the characterization of the covariability of populations of LGN neurons and how it relates to spatial position and response properties of individual cells.

Finally, the main strength of the RM model is that it fits only the most frequent – and hence most reliable – spiking features. Importantly, this prevents overfitting to noise in the data. But in certain scenarios, this may also prevent the identification of the coordinate *absence* of spiking features, which may indicate *negative* correlations. However, it is difficult to identify with higher-order spiking features are important to include. Therefore, future work will aim to develop models that are able to identify which infrequent spiking features may be important for the spiking structure of the population activity.

CHAPTER 5

Dendritic nonlinearities modulate higher-order correlations

Previously, we investigated how higher-order spiking correlations can impact stimulus encoding in neural populations, as well as the quantification, measurement, and fitting of higher-order statistical features in maximum entropy models. However, these approaches are agnostic about the origins of correlated activity, raising the following question: which biophysical mechanisms generate HOCs in neural circuits? Several factors are known to contribute to beyond-pairwise statistics, including common input [90, 17] and motif structures within the circuit architecture [69]. In this chapter, we discuss a novel cellular mechanism that can shape HOCs: dendritic nonlinearities.

Dendrites are the branched tree-like structures of neurons that transport signals from synapses to the soma, or cell body. They come in a variety of shapes and sizes, from the labyrinthine arbors of Purkinje cells to more austere unipolar brush cells, each of which has only one dendrite; accordingly, their unique morphologies make them useful for classifying distinct neural cell types. Once believed to be simple passive cables, more recent work has identified a rich set of active dendritic properties that allow for a variety of intracellular computations [45, 117, 116, 152, 4, 144, 88, 64]. For example, dendrites have been shown to play a crucial role in direction selectivity in retina [132, 15, 16, 118, 45], and even orientation selectivity in mouse visual cortex [133]. Also in retina, nonlinear dendritic integration of synaptic inputs from electrical and chemical synapses has been shown to drive correlated spiking between nearby ganglion cells [144]. Active dendritic properties have also been implicated in: coincidence detection in auditory brainstem [5], identification of looming stimuli in the locust visual system [49], and backpropagation of action potentials in cortex, hippocampus, and other brain regions [139, 150, 22, 28]. More theoretical modeling studies have demonstrated that dendritic properties may potentially enhance information processing by increasing the memory capacity [56] and dynamic range of individual neurons [56], as well as enabling computation of linearly non-separable functions [31].

To see how this connects to HOCs, we return to the third-order maximum entropy model. Under this formulation, the probability that neuron i spikes conditioned on the

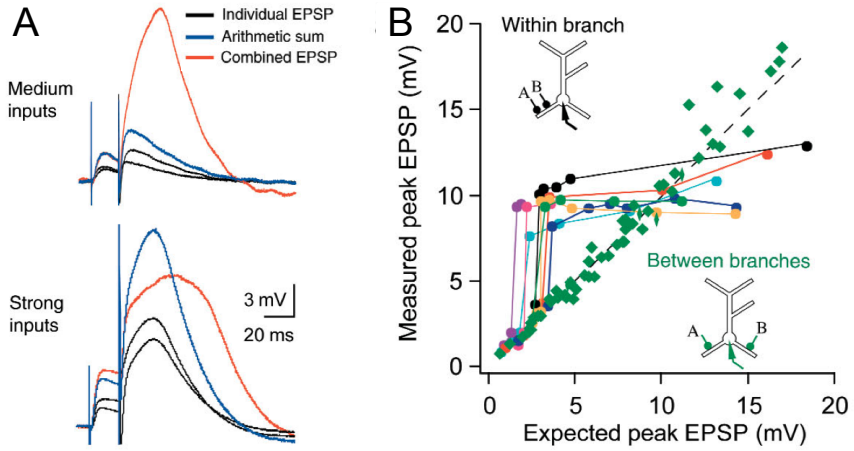

Figure 5.1: Dendritic nonlinearities in cortical pyramidal cells. Adapted from Figure 2 in [117]. (A) Excitatory postsynaptic potentials (EPSPs) measured at the soma in response to individual paired-pulse stimuli (black traces) or simultaneous stimulation of both paired pulses (red trace) on the same dendritic branch. The blue traces shows the linear sum of the individual EPSPs. Medium-sized inputs elicit a larger response at the soma than expected from the individual EPSPs (top panel), while strong inputs elicit a smaller response (bottom panel). (B) Peak amplitude of the measured EPSP in response to both inputs, versus the expected peak value from the arithmetic sum of individual EPSPs. When separate branches were stimulated (green diamonds), somatic responses were always linear. However, stimuli on the same branch were summed sublinearly or superlinearly at the soma depending on size (circles; colors indicate different dendritic branches).

activity of the rest of the population is given by the following formula [155, 141]:

$$P(\sigma_i = 1|\{\sigma_{j\neq i}\}) = f\left(h_i + \sum_j J_{ij}\sigma_j + \sum_{j<k} G_{ijk}\sigma_j\sigma_k\right), \tag{5.0.1}$$

where $f(\cdot)$ is a sigmoidal function. In the absence of triplet interactions ($G_{ijk} = 0$), Equation 5.0.1 can be interpreted as a linear-nonlinear model neuron, with h_i representing external input and J_{ij} representing the connectivity between cells. The addition of nonzero G_{ijk} (or even higher-order interaction terms) modifies this interpretation by adding a nonlinear synaptic contribution. For example, suppose cells j and k spike simultaneously. If $G_{ijk} > 0$, this amounts to a positive contribution to the argument of Equation 5.0.1, signifying a *superlinear* summation of the synaptic inputs. Similarly, negative G_{ijk} induces a *sublinear* summation of the inputs.

Intriguingly, this observation turns out to have key similarities to a particular type of dendritic nonlinearity discovered in the basal dendrites of cortical pyramidal cells [117]. The authors of the study gave paired pulse stimuli to either the same dendritic branch or to separate branches, and measured the change in the membrane potential at the soma. When stimulating the same branch, the two paired pulse inputs were found to be summed

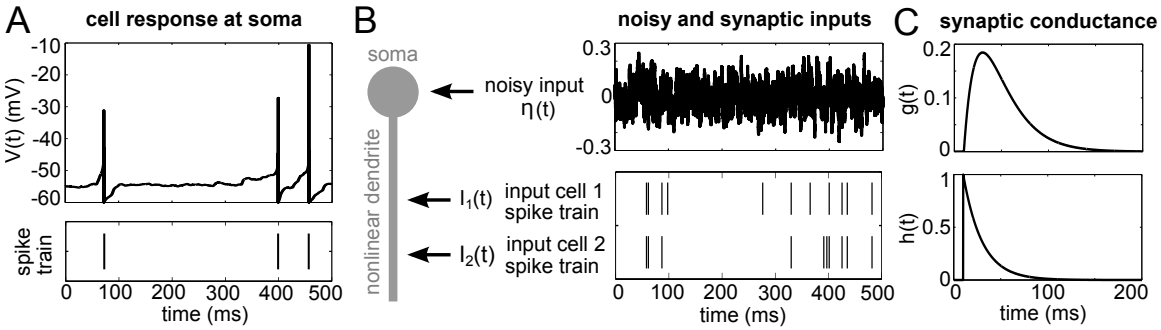

Figure 5.2: Illustration of the exponential integrate-and-fire (EIF) neuron model. (A) Example voltage trace (top) and corresponding spike train showing the times of all action potentials (bottom). (B) Noisy and synaptic inputs. Noise (modeled by an Ornstein-Uhlenbeck process; top right) is injected directly to the soma. Synaptic input is filtered through a dendritic nonlinearity (described in text). Bottom right panel shows example spike trains from the presynaptic cells. (C) Synaptic conductance alpha function shows the response to a single synaptic input. Top panel plots the conductance $g(t)$, while the bottom panel plots the auxiliary function $h(t)$ (see text).

nonlinearly at the soma. Figure 5.1A shows the excitatory postsynaptic potentials (EPSPs) in response to either of the two inputs individually (black traces) and the EPSPs in response to both inputs simultaneously (red trace). Medium-sized pulse inputs elicited responses that were much larger than predicted by the linear sum of the individual EPSPs (shown in blue), indicating a superlinear summation of the inputs; meanwhile, strong inputs were combined sublinearly. In contrast, inputs to separate dendritic branches were summed linearly at the soma (Figure 5.1B).

Combining the observation that higher-order interactions can be interpreted as nonlinear summation of synaptic inputs with the findings from [117] that active dendritic processes induce such nonlinearities, we hypothesize that dendritic nonlinearities may shape HOCs in neural populations. Here, we demonstrate this mechanism through spiking simulations of a single neuron receiving input from two correlated spike trains. We show that changing the curvature of the dendritic nonlinear filter modulates the triplet correlations between the neuron and its presynaptic cells. As we saw in Chapter 3, such modulation may have a significant impact on stimulus coding in neural populations.

5.1 Nonlinear summation in single cells

In order to isolate the effect of dendritic nonlinearities in generating higher-order correlations, we model a single neuron receiving synaptic activity from two presynaptic cells as well as zero-mean noise representing other sources of input. The membrane potential $V(t)$ of the postsynaptic cell is given by the following differential equation:

$$C_m \frac{\partial V}{\partial t} = -g_L(V(t) - E_L) + g_L \Delta_T \exp\left\{\frac{V - V_T}{\Delta_T}\right\} + I(t). \tag{5.1.1}$$

C_m is the membrane capacitance, g_L is the leak conductance, and E_L is the leak potential (all parameter definitions and values in this section are given in Table 5.1).

The second term on the right-hand side is the spike-generating current, so called because it causes the membrane potential to diverge, approximating the shape of an action potential. Specifically, when $V(t)$ is near E_L, the cell acts as a leaky integrator. If the external input $I(t)$ is able to drive the membrane potential sufficiently high, the exponential form of the spike-generating current becomes larger and larger, dominating the leak term and causing the voltage to diverge or spike. The shape of the action potential is controlled by parameters Δ_T. V_T sets the excitability of the cell. Once the membrane potential reaches the value V_{Th}, we say the cell has spiked, and the neuron is hyperpolarized by setting $V(t)$ to the reset potential V_{Re}. An example is shown in Figure 5.2A, which plots a typical voltage trace over 500 ms (top panel) and the resulting spike train (bottom panel).

The input to the cell, $I(t)$, is a combination of synaptic inputs and zero-mean noise:

$$I(t) = I_{\text{syn}}(t) + \eta(t) \tag{5.1.2}$$

The noise is modeled as an Ornstein-Uhlenbeck process, defined by the following stochastic differential equation:

$$d\eta(t) = -\theta\eta(t)dt + \sigma dW(t). \tag{5.1.3}$$

The parameters σ and θ are chosen to generate quickly-fluctuating subthreshold input, reflecting background noise and other potential synaptic inputs to the soma (Figure 5.2B, top panel).

Finally, the synaptic input is the crucial component of the model for creating higher-order correlations. For simplicity, we consider a single dendrite receiving input from two presynaptic cells (as depicted by the schematic in Figure 5.2B). The presynaptic spike trains, denoted by $I_1(t)$ and $I_2(t)$, are assumed to be correlated but marginally Poisson with firing rates λ. To generate these spike trains, we first generate spiking times for three independent Poisson processes: $I_C(t)$, which spikes at a rate of $\lambda_C = \rho\lambda$ (for some $0 \leq \rho \leq 1$), and $I_1^{\text{ind}}(t)$ and $I_2^{\text{ind}}(t)$, each of which spikes at a rate of $\lambda_I = (1-\rho)\lambda$. The presynaptic cell spike trains are then given by the sums of the common and the independent spikes:

$$I_1(t) = I_C(t) + I_1^{\text{ind}}(t) \tag{5.1.4}$$

$$I_2(t) = I_C(t) + I_2^{\text{ind}}(t) \tag{5.1.5}$$

The resulting spike trains are Poisson with firing rate λ and a zero-lag correlation of approximately ρ (this becomes exact for small time bins). Figure 5.2B (bottom panel) shows an example with $\lambda = 10$ Hz and $\rho = 0.5$.

With the statistics of the presynaptic spike trains thus set, the synaptic input to the neuron is given by a sum of the postsynaptic conductances $g_i(t)$ passed through a nonlinearity:

$$I_{\text{syn}}(t) = f_{\text{dend}}(g_1(t) + g_2(t)). \tag{5.1.6}$$

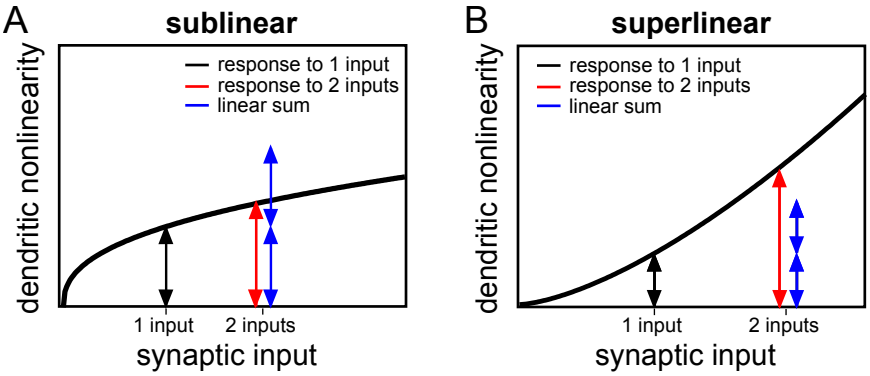

Figure 5.3: Schematic illustrating the effect of curvature on nonlinear summation of synaptic inputs. Shown are the contribution to the soma of a single synaptic input (black), two synaptic inputs (red), and the expected linear sum of the response to two inputs (blue). (A) If $f_{\text{dend}}(x)$ has negative curvature, the contribution of two synaptic inputs is less than expected from the linear sum. (B) Conversely, if $f_{\text{dend}}(x)$ has positive curvature, the contribution of two synaptic inputs is greater than the linear sum.

The function $f_{\text{dend}}(\cdot)$ defines the form of dendritic nonlinearity and will be described in detail in the following section. The conductances $g_i(t)$ are transient pulse-like functions that represent increased excitability at the postsynaptic terminal each time presynaptic cell i spikes (see Figure 5.2C). They are modeled as an alpha function:

$$\frac{\partial g_i}{\partial t} = -\frac{1}{\tau^{(1)}} g_i(t) + \frac{1}{\tau^{(2)}} h_i(t), \tag{5.1.7}$$

where $h_i(t)$ is an auxiliary function that increases by 1 every time presynaptic cell i spikes, and leaks towards zero otherwise:

$$\frac{\partial h_i}{\partial t} = -\frac{1}{\tau^{(1)}} h_i(t) + \delta(t - t_i^{\text{spike}}). \tag{5.1.8}$$

The timescales $\tau^{(1)}$ and $\tau^{(2)}$ parameterize the shape of the conductances. We set $\tau^{(2)} = (\tau^{(1)})^2$ for the remainder of this section; this ensures that the integral of the conductance in response to a single spike is 1.

5.1.1 Linear and nonlinear dendrites

We compared the effect of linear and nonlinear dendritic filters on the spiking statistics of the simulated cell with its presynaptic inputs. The crucial component that determines whether synaptic inputs will be integrated nonlinearly is the curvature of the dendritic filter $f_{\text{dend}}(x)$. In particular, if $f_{\text{dend}}(x)$ has negative curvature then the contribution from multiple synaptic inputs will be smaller than expected from a linear sum (as illustrated in Figure 5.3A). Mathematically, this is due to the fact that if $\frac{\partial^2 f_{\text{dend}}}{\partial x^2} < 0$ then the following

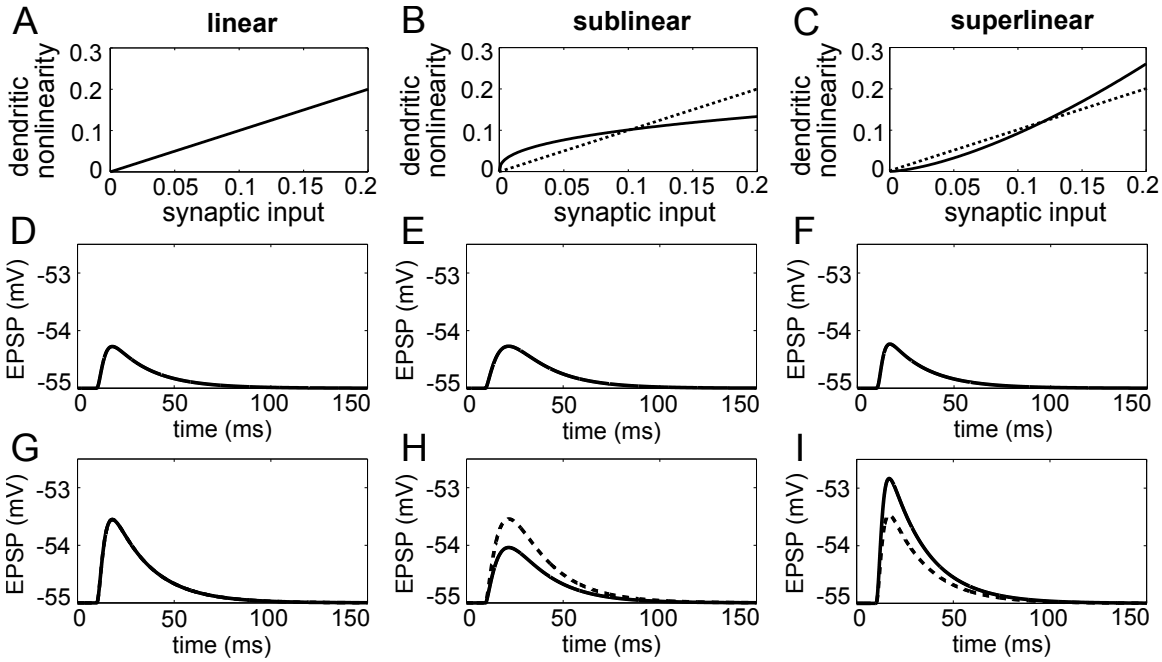

Figure 5.4: Dendritic nonlinearities and effect on EPSPs. Left, middle, and right columns depict linear, sublinear, and superlinear dendrites. (A-C) Dendritic nonlinearity as a function of synaptic input (solid line; dashed indicates identity). (D-F) EPSPs in response to a single synaptic input. Note that in each of the three cases, the amplitude of the response is identical. (G-I) EPSPs in response to two simultaneous synaptic inputs (solid lines). Also shown are the linear sum of two individual EPSPs (dashed lines). In the linear model, the two-input EPSP is exactly the sum of two individual EPSPs. Sublinear (or superlinear) dendrites decrease (respectively, increase) the magnitude of the response from two inputs compared to the linear sum. EPSPs are generated by integrating Equation 5.1.1 in response to either a single input spike or to two simultaneous input spikes, with noise and spike-generating current removed for clarity.

inequality holds:

$$f_{\text{dend}}(x+y) < f_{\text{dend}}(x) + f_{\text{dend}}(y). \tag{5.1.9}$$

Conversely, if $\frac{\partial^2 f_{\text{dend}}}{\partial x^2} > 0$ then:

$$f_{\text{dend}}(x+y) > f_{\text{dend}}(x) + f_{\text{dend}}(y), \tag{5.1.10}$$

hence positive curvature increases the contribution from multiple inputs relative to the linear sum (Figure 5.3B).

To induce such super- or sublinear summation of synaptic inputs, we therefore tested

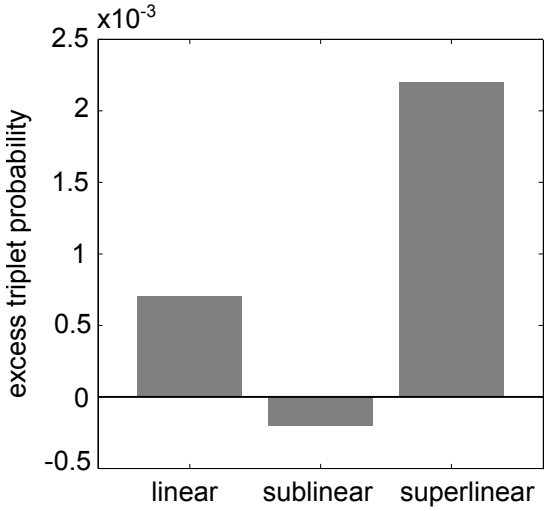

Figure 5.5: Triplet correlation among a simulated EIF cell and its two presynaptic inputs for models with linear, sublinear, or superlinear dendrites. As predicted, sublinear dendrites decrease the triplet correlations, while superlinear dendrites increase triplet correlations. Input spike train parameters: $\rho = .5$, $\lambda = 10$ Hz. Models were simulated for 100 s.

nonlinear functions with positive or negative curvature, specifically:

$$f_{\text{dend}}(x) = \begin{cases} x & \text{linear} \\ K^{\text{sub}} x^{\alpha} & \text{sublinear, } 0 < \alpha < 1 \\ K^{\text{sup}} x^{\beta} & \text{superlinear, } \beta > 1 \end{cases} \tag{5.1.11}$$

In particular, we chose $\alpha = 0.4$ and $\beta = 1.5$. For a fair comparison, the constants K^{sub} and K^{sup} were chosen so that the amplitude of the individual EPSPs (i.e., in response to a single incoming spike) were identical for each model. This was accomplished by scaling each nonlinearity to match the integral of $f_{\text{dend}}(x)$ up to g^*, the maximum conductance attained in response to a single cell, i.e.:

$$\int_0^{g^*} f_{\text{dend}}(x)dx = \int_0^{g^*} xdx, \tag{5.1.12}$$

which simplifies to the following expressions:

$$K^{\text{sub}} = \frac{(g^*)^2}{2\int_0^{g^*} x^{\alpha}dx} \qquad \text{and} \qquad K^{\text{sup}} = \frac{(g^*)^2}{2\int_0^{g^*} x^{\beta}dx}. \tag{5.1.13}$$

The resulting linear, sublinear, and superlinear dendritic functions are shown in Figure 5.4A-C. As constructed, the EPSPs in response to a single synaptic input have approximately the same magnitude for all three cases (Figure 5.4D-F). In Figure 5.4, we show the EPSPs in

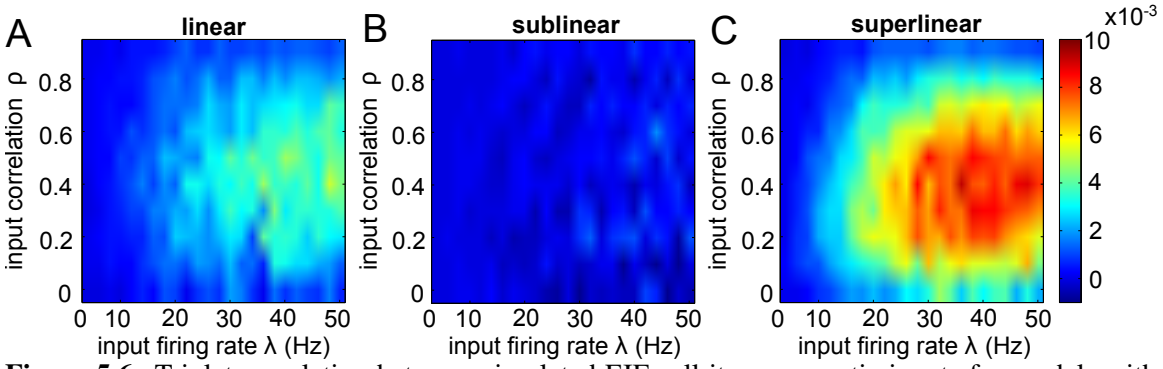

Figure 5.6: Triplet correlation between simulated EIF cell its presynaptic inputs for models with linear, sublinear, or superlinear dendrites for varying presynaptic input statistics. In general, sublinear dendrites decreased the triplet correlation while superlinear dendrites increased them. This effect was most pronounced for large firing rates and intermediate input correlations. Models were simulated for 500 s.

response to two simultaneous synaptic inputs in solid (spike-generating current removed for clarity). The dashed line in these panels represents the linear sum of the single-input EPSPs. In the linear neuron, these curves lie on top of each other, as the response is perfectly linear. However, in sublinear dendrites, the somatic response to two inputs is less than the linear sum of two individual EPSPs; similarly, superlinear dendrites cause a larger somatic response than expected. These effects are comparable to the magnitude of superlinear or sublinear summation of individual EPSPs that have been observed in the dendritic branches of cortical pyramidal cells [117] (shown in Figure 5.1A).

5.1.2 Triplet spiking correlations are modulated by the dendritic filter

Having verified the effect of $f_{\text{dend}}(x)$ on the nonlinear summation of inputs, we tested the effect on the triplet correlation κ for each of the three models. As detailed in the previous chapters, the triplet correlation was measured by: (1) simulating Equation 5.1.1 for 100 seconds, (2) binning the spiking activity into time steps of 20 ms, (3) fitting the resulting raster plot with a pairwise maximum entropy model, and (4) calculating the excess probability of the cell spiking simultaneously with its two presynaptic inputs that was not accounted for by the maximum entropy model.

The results are plotted in Figure 5.5 for an example spike train with presynaptic input spike correlations of $\rho = 0.5$ and firing rates of $\lambda = 10$ Hz. Due in part to intrinsic nonlinearities within the EIF model, the linear model does not have nonzero triplet correlation; still, the addition of superlinear dendrites increases the triplet correlations beyond this baseline value, while the addition of sublinear dendrites decreases them, as predicted by our reasoning based on the curvature of $f_{\text{dend}}(x)$.

We repeated the simulations while varying the presynaptic input statistics: correlation coefficients ranged from 0 to 1 and presynaptic cell firing rates ranged up to 50 Hz.

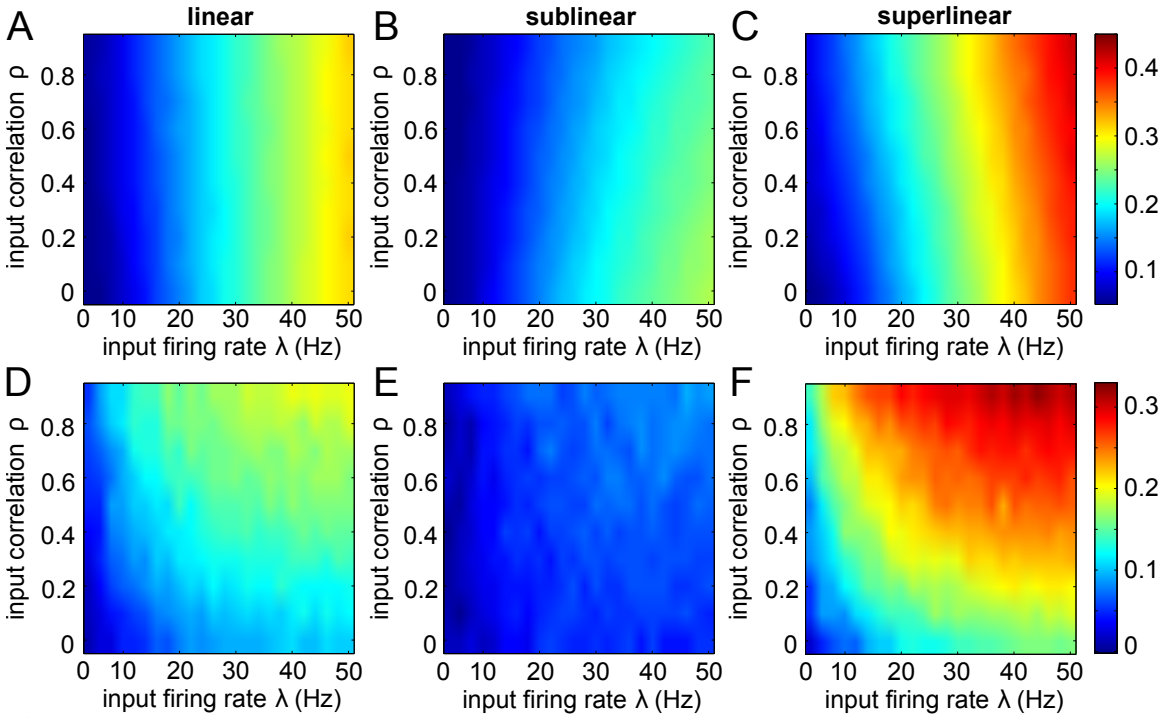

Figure 5.7: Lower-order spiking statistics for models with linear, sublinear, or superlinear dendrites for varying presynaptic input statistics. Dendritic nonlinearities have the same effect on the lower-order statistics as on triplet correlations. (A-C) Probability of of the postsynaptic neuron spiking (within a 20 ms time bin). Note that the input correlation has little effect on the single-neuron spiking statistics. (D-F) Average pairwise correlation between the postsynaptic cell and either of its inputs. The effect is stronger for larger input correlations. Models were simulated for 500 s.

Generally, triplet spiking statistics decreased with sublinear dendrites and increased with superlinear dendrites (Figure 5.6). This effect was most pronounced for large input firing rates and intermediate input correlations. This is because when input correlations are very small, simultaneous spiking – and therefore the effect of nonlinear dendritic summation – is infrequent. On the other hand, if the inputs are near fully correlated, the postsynaptic cell is frequently driven by the the three cells, and the system is well described by a pairwise maximum entropy model with very strong pairwise interactions. But when input correlations are in an intermediate range, dendritic nonlinearities have a strong effect on simultaneous triplet spiking.

How do dendritic nonlinearities affect other spiking statistics of the three cells? As we have argued in this chapter, nonzero triplet interactions in Equation 5.0.1 increase or decrease the probability of a cell spiking in response to synchronous firing in its presynaptic inputs; but this should affect not only the triplet correlations, but also the firing rate of the postsynaptic cell and its pairwise correlation with either of its inputs. In fact, Figure 5.7 shows that sublinear and superlinear dendritic nonlinearities have the same effect on the lower-order statistics that it had on the triplet correlation of the system.

5.2 Stimulus-dependent HOCs

In Chapter 3, we showed that significant facilitation of stimulus coding occurred when triplet correlations change sign under different stimuli. Are such stimulus-dependent HOCs able to be produced by dendritic nonlinearities? One indication of this possibility was given in [117], in which differentially sized stimuli were found to elicit different dendritic nonlinearities: that is, smaller pulses were summed superlinearly at the soma while larger pulses were summed sublinearly. Given the findings in this chapter, this is consistent with triplet correlations in the SD2 region in Chapter 3, which accounted for the largest increase in encoded information.

One potential way to generate stimulus-dependent HOCs is through a sigmoidal nonlinearity, which incorporates regions of both positive and negative curvature (Figure 5.8A). Then, stimulus-dependent switching between the two regimes of the nonlinearity will induce either superlinear or sublinear summation of synaptic inputs, and hence stimulus-dependent triplet statistics. In fact, comparing Figure 5.1B with the equivalent depicted in Figure 5.8B, sigmoidal nonlinearities appear to be a good qualitative fit of the pulse-size dependent nonlinear summation described in [117]. In principle, such a mechanism may be able to generate stimulus-dependent modulation of higher-order statistics and will be a topic of further research.

5.3 Discussion

In this chapter, we studied how dendritic nonlinearities can generate triplet statistics in exponential integrate-and-fire models. The curvature of the nonlinearity induces either superlinear or sublinear summation of synaptic inputs, which increases or decreases the probability of the postsynaptic neuron spiking, and as a result generates triplet correlations. This effect was strongest for intermediate spiking correlations between the presynaptic cells. By incorporating more synaptic inputs on each branch, we could extend these results to even higher-order statistics. For simplicity, we used a very simple model of a single dendrite to illustrate our results; however the rich variety of dendritic morphology, which is now understood to have many active properties, may create more complex statistical structure. Future work will investigate how dendritic nonlinearities can shape higher-order correlations within large spiking networks and how this affects population coding.

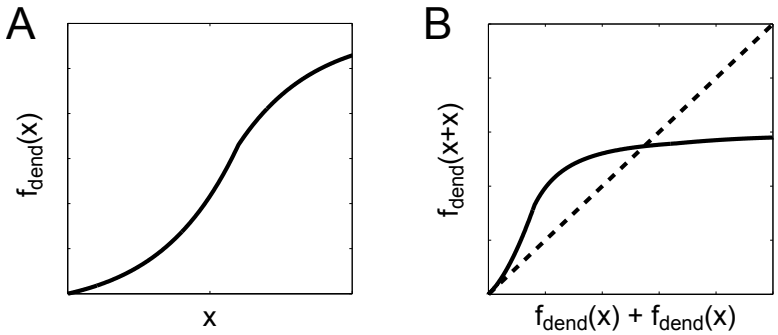

Figure 5.8: Stimulus-dependent dendritic nonlinearity. (A) Schematic of a sigmoidal dendritic nonlinearity, which incorporates both positive and negative curvature. (B) Comparison of nonlinear summation of sigmoidal dendritic linearity (oordinate) and linear summation (abscissa) shows regions of both superlinear and sublinear summation (solid line; dashed line indicates the identity). Compare with Figure 5.1B.

Symbol	Value	Description
C_m	0.2 nF	Membrane capacitance
g_L	.01 μS	Leak conductance.
E_L	-55 mV	Leak potential.
Δ_T	1.4 mV	Action potential shape parameter.
V_T	-53 mV	Soft threshold potential, at which spike-generating current activates.
V_{Re}	-60 mV	Reset membrane potential.
V_{Th}	20 mV	Hard threshold potential, at which membrane potential is reset to V_{Re}.
$\tau^{(1)}$	2 ms	Synaptic conductance time constant (controls exponential decay).
$\tau^{(2)}$	4 ms	Synaptic conductance time constant (controls exponential rise).
σ	0.1	Relaxation parameter for Ornstein-Uhlenbeck process.
θ	1	Drift parameter for Ornstein-Uhlenbeck process.
λ	varies	Firing rate (in Hz) of presynaptic neurons.
c	varies	Correlation between the two presynaptic neurons.

Table 5.1: Parameters used in EIF simulation.

CHAPTER 6

Conclusion

The spiking activity of neural populations in different brain regions is often far from statistically independent; rather, neural activity tends to be coordinated, exhibiting synchronous spiking and correlations among groups of two or more cells. In this dissertation, we addressed three fundamental questions regarding coordinated spiking. How does coordinated neural firing arise from simple cellular and network properties? How does coordinated spiking impact the representation of external stimuli in neural population activity? And how can we detect and model coordinated spiking among groups of cells in electrophysiological recordings?

For the first question, we focused on two specific properties that can generate coordinated spiking: feedforward dynamics and intrinsic cell nonlinearities. Chapter 2 studied the emergence of synchrony and higher-order interactions in the summed population activity of layered networks. In Chapter 5, we showed how dendritic nonlinearities can modulate higher-order correlations among the activity of a cell and its presynaptic inputs. These chapters discuss only two potential origins of coordinated spiking, but the interaction between these and other sources has yet to be explored. In particular, higher-order correlations can also be generated by common fluctuating input [90, 153, 17] and network motifs [69]. Given the spike trains of a population of cells, is it possible to distinguish between different sources of coordinated spiking? While this has been studied for pairwise correlations [108, 66, 113, 121], the broader question remains unsolved.

The second question has been a major source of research and debate at the level of pairs of cells. However, the effect of higher-order correlations on neural coding is a largely unexplored question. In Chapter 3, we demonstrated how and when triplet correlations facilitate the encoding of stimuli in population spiking patterns, while building intuition with a geometrical explanation of how triplet correlations skew the distributions of summed population responses. This is just a first step towards understanding the full impact of higher-order correlations on brain functionality. Other theoretical studies show that higher-order correlations also play a role in plasticity [54], the activity a downstream readout neuron [81], or inferring binocular disparities [99]. Future work is needed to ultimately understand the rich variety of effects that higher-order correlations may have on neural

computation. Equally important is whether learning or adaptation mechanisms change the statistical properties of neural spiking to benefit stimulus coding, as has recently been shown for pairwise correlations in songbird auditory cortex [73]. If this is also true for higher-order correlations, it would be strong evidence of their role in neural coding.

Finally, because of the relative infrequency of synchronous spiking in subgroups of cells, higher-order correlations are difficult to detect in data. In particular, long recordings or many repeated trials are necessary to generate enough statistical power to accurately measure them. In Chapter 4, we developed the Reliable Moment model, an adaptive statistical model modified from [52] that identifies and fits the most frequently occurring spiking features in the data. This minimizes the chance of overfitting to noise, while our modifications make the model less likely to capture spurious higher-order correlations. However, one weakness of our model is that it cannot capture the coordinated *absence* of spiking, which corresponds to strong negative correlations. Although difficult to estimate accurately because of their rarity, these features may still carry important information about the spiking activity. Therefore future work will attempt to identify which spiking features are necessary for the most parsimonious model of neural firing patterns.

When we applied the Reliable Moment model to electrophysiological data from mouse lateral geniculate nucleus (LGN), we found that the spiking activity was significantly shaped by pairwise correlations; however, we did not find evidence of the presence of higher-order correlations. This result was puzzling in light of recent observations of higher-order correlations both in retina [52, 141] and in visual cortex [109, 92], which are major sources of input to LGN (and in the case of cortex, also the primary destination of LGN output). One intriguing possibility is that LGN decorrelates the higher-order interactions in its inputs. While there is evidence that LGN decorrelates temporal activity at the pairwise level as a strategy to improve coding efficiency [38, 12], it is not immediately clear what computational benefits would underlie the removal of higher-order correlations among cells.

Another potential cause of this apparent discrepancy is the lack of standardization in how higher-order correlations are measured and inferred in data. The term "higher-order correlations" is not precisely defined, and as such, many studies measure slightly different quantities that are difficult to compare. Here, we list a few of the differences between our work and existing studies that could explain why we did not observe significant higher-order correlations:

1. Higher-order correlations are quantified differently across the literature. For example, [100] assumed population homogeneity and looked at the higher-order interactions in the spike count distributions (as we did in Chapter 2, but differing from the remaining chapters). While interesting in their own right, it is unclear when higher-order interactions in the population activity correspond to higher-order correlated spiking in heterogenous populations. On the other hand, [92] directly measured the value of the higher-order interaction terms in maximum entropy models (e.g., the γ_{ijk} and δ_{ijkl} terms in Equation 4.1.1). However, the effect of these nonzero interaction terms on the distribution of spiking patterns depends on the lower-order interaction terms as well [104]. Therefore a more direct measurement of higher-order correlations would be to calculate the Kullback-Leibler di-

vergence or to use moments or cumulants.

2. Because of the potential to overfit to noise, higher-order models must be cross-validated (i.e., tested on a held-out set data that was not used for training) for verification. Therefore, in studies that observe higher-order correlations without cross-validating them [109, 100, 92], it is difficult to know if the observed statistics will generalize to more data.

3. A distinction is often drawn between correlated spiking that reflects correlations within the stimulus ("signal" correlations) versus correlated spiking in response to a fixed stimulus ("noise" correlations) [13]. In this dissertation, we have mainly been considering the latter: for example, in Chapter 4 we fit the spiking patterns in response to each grating direction separately (however, we mix spatial and temporal frequencies, the implications of which we discuss in a few paragraphs). Several studies in retina [141, 52] and in visual cortex [79] calculate the statistics of neural firing patterns while the animal is presented with natural movies, without conditioning on fixed frames. In fact, natural images tend to have significant higher-order correlations between pixel strengths, which account for the texture of images [131]. Therefore it is unclear to what extent the beyond-pairwise statistics observed in [141, 52, 79] reflect noise correlations as opposed to signal correlations.

4. Spurious higher-order correlations could potentially be observed if the firing rates fluctuate over the period of the recording. This may be particularly confounding when measuring correlations in anesthetized animals because of the presence of cortical up/down states. For example, up/down states in visual cortex have been found to create pairwise correlations that are not representative of neural activity in the awake animal [43]. A similar effect could take part for higher-order correlations observed in anesthetized cortex [109, 100, 92].

5. Finally, the presence of hidden units (i.e., neurons embedded within a population that are not observed) may create the appearance of higher-order correlations even in a population that has none. Hidden units are an unfortunate reality in any experiment, but their effect becomes stronger as more and more cells are marginalized from the observed population. Therefore studies that observe higher-order correlations in very small subgroups of 3-6 cells [109, 92] may actually be observing strong drive from a common presynaptic cell.

At the same time, there are several reasons why our analysis of mouse LGN may have missed higher-order correlations within that region. First, we only recorded populations of 40 neurons, but some studies indicate that around a hundred neurons were necessary to observe higher-order correlations [52]. Moreover, [109] only found evidence of higher-order interactions in populations within 300 microns. In our probes, the six shanks were 200 microns apart; perhaps focusing on more localized populations would reveal higher-order correlated structure. Finally, in our analysis, we averaged all data over different spatial and temporal frequencies, treating them as nuisance parameters. If there were higher-order correlations between LGN cells that changed structure for different frequencies, their presence may have been obscured by this averaging. We are currently in the process of collecting more data to continue our analysis into each of these potential directions.

The range of issues that we have discussed in this section showcases the potential diffi-

culties inherent in determining the structure of higher-order correlations and their relevance for brain function. These remain open and important questions in neuroscience. Therefore, an important direction for future research is to create more accurate statistical tests for the presence of higher-order correlations, and more sophisticated tools for understanding how they affect neural spiking. We have used maximum entropy models and the Kullback-Leibler divergence to infer higher-order spiking correlations [124, 17]. But other techniques to infer higher-order correlations in spiking activity have recently been explored, ranging from Monte Carlo statistical tests [110] to cumulant-based methods [137, 136]; and other models to fit higher-order models have been proposed, including copula families [111] and restricted Boltzmann machines [79, 68]. Developing these techniques will be an important part of characterizing the full range of coordinated activity in neural populations in the future, and in understanding what this variability means for brain function.

References

[1] Tang. A., D. Jackson, J. Hobbs, W. Chen, J.L. Smith, H. Patel, A. Prieto, D. Petrusca, M.I. Grivich, A. Sher, Hottowy. P., W. Dabrowski, A.M. Litke, and J.M. Beggs. A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *Journal of Neuroscience*, 28(2):505–518, 2008.

[2] L. Abbott and P. Dayan. The effect of correlated variability on the accuracy on a population code. *Neural Computation*, 11:91–101, 1999.

[3] M. Abeles, H. Bergman, and W. Margalit, E. a nd Vaadia. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, 70:1629–1638, 1993.

[4] T. Abrahamsson, L. Cathala, K. Matsui, R. Shigemoto, and D. DiGregorio. Thin dendrites of cerebellar interneurons confer sublinear synaptic integration and a gradient of short-term plasticity. *Neuron*, 73:1159–1172, 2012.

[5] H. Agmon-Smir, C.E. Carr, and J. Rinzel. The role of dendrites in auditory coincidence detection. *Nature*, 393:268–272, 1998.

[6] H.J. Alitto and W.M. Usrey. Corticothalamic feedback and sensory processing. *Current Opinion in Neurobiology*, 13:440–445, 2003.

[7] C. Allen and C.F. Stevens. An evaluation of causes for unreliability of synaptic transmission. *PNAS*, 91:10380–10383, 1994.

[8] J. Alonso, M. Usrey, and C. Reid. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, 383:815–819, 1996.

[9] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[10] S. Amari, H. Nakahara, S. Wu, and Y. Sakai. Synchronous firing and higher-order interactions in neuron pool. *Neural Computation*, 15:127–142, 2003.

[11] F.R. Amthor and C.W. Oyster. Spatial organization of retinal information about the direction of image motion. *PNAS*, 92:4002–4005, 1995.

[12] J. Atick and A.N. Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308–320, 1990.

[13] B. Averbeck, P. Latham, and A. Pouget. Neural populations, population coding and computation. *Nature Reviews Neuroscience*, 7:358–366, 2006.

[14] R. Baddeley, L. Abbott, M. Booth, F. Sengpiel, T. Freeman, E. Wakeman, and E. Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London B*, 264:1775–1783, 1997.

[15] H.B. Barlow, R.M. Hill, and W.R. Levick. Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *Journal of Physiology*, 173:377–407, 1964.

[16] H.B. Barlow and W.R. Levick. The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology*, 178:477–504, 1965.

[17] A. Barreiro, J. Gjorgjieva, F. Rieke, and E. Shea-Brown. When do microcircuits produce beyond-pairwise correlations? *Frontiers in Computational Neuroscience*, 8:e00010, 2014.

[18] A. Barth and J. Poulet. Experimental evidence for sparse firing in the neocortex. *Trends in Neurosciences*, 35:345–355, 2012.

[19] J.M. Beck, W.J. Ma, X. Pitkow, P.E. Latham, and A. Pouget. Not noisy, just wrong: the role of suboptimal inferences in behavioral variability. *Neuron*, 74(1):17, 2012.

[20] J. Beggs and D. Plenz. Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35):11167–11177, 2003.

[21] A. Berger, S. Della Pietra, and V. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71, 1996.

[22] C. Bernard and D. Johnston. Distance-dependent modifiable threshold for action potential back-propagation in hippocampal dendrites. *Journal of Neurophysiology*, 90(3):1807–1916, 2003.

[23] N. Bertschinger and T. Natschlager. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, 16(7):1413–1436, 2004.

[24] Marc Binder and Randall Powers. Relationship between Simulated Common Synaptic Input and Discharge Synchrony in Cat Spinal Motoneurons. *J Neurophysiol*, 86(5):2266–2275, 2001.

[25] V. Braitenberg and A. Schüz. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, 1998.

[26] T. Branco and K. Staras. The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Reviews Neuroscience*, 10:373–383, 2009.

[27] R. Bruno. Synchrony in sensation. *Current Opinion in Neurobiology*, 21:701–708, 2011.

[28] G. Buzsaki and A. Kandel. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *Journal of Neurophysiology*, 79:1587–1591, 1998.

[29] M. Camperi and X.J. Wang. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *Journal of Computational Neuroscience*, 5:383–405, 1998.

[30] A. Cayco-Gajic and E. Shea-Brown. Neutral stability, rate propagation, and critical branching in feedforward networks. *Neural Computation*, 25(7):1768–1806, 2013.

[31] R. Caze, M. Humphries, and B. Gutkin. Passive dendrites enable single neurons to compute linearly non-separable functions. *PLOS Computational Biology*, 9(2):1002867, 2013.

[32] M. Cohen and A. Kohn. Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14:811–819, 2011.

[33] M. Cohen and J. Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12:1594–1600, 2009.

[34] A. Compte, N. Brunel, P. Goldman-Rakic, and X.J. Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, 10:910–923, 2000.

[35] C. Constantinidis and P. Goldman-Rakic. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of Neurophysiology*, 88:3487–3497, 2002.

[36] R. da Silveira and M. Berry. High-fidelity coding with correlated neurons. *PLOS Computational Biology*, 10:e1003970, 2014.

[37] Y. Dan, J.M. Alonso, W.M. Usrey, and R.C. Reid. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neuroscience*, 1:501–507, 1998.

[38] Y. Dan, J. Atick, and R.C. Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *Journal of Neuroscience*, 16(10):3351–3362, 1996.

[39] J. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *Annals of Mathematical Statistics*, 43(5):1470–1480, 1972.

[40] J. Daugman. Brain metaphor and brain theory. In W.P. Bechtel, P. Mandik, J. Mundale, and R.S. Stufflebeam, editors, *Philosophy and the Neurosciences*, pages 23–36. Blackwell, 2001.

[41] R.C. deCharms and M. Merzenich. Primary cortical representation of sounds by the coordination of action-potential timing. *Nature*, 381:610–613, 1996.

[42] M. Diesmann, M. Gewaltig, and A. Aertsen. Stable propagation of synchronous spiking in cortical neural networks. *Nature*, 402:529–533, 1999.

[43] A. Ecker, P. Berens, J. Cotton, M. Subramaniyan, G.H. Denfield, C.R. Cadwell, S.M. Smirnakis, M. Bethge, and A. Tolias. State dependence of noise correlations in macaque primary visual cortex. *Neuron*, 82:235–248, 2014.

[44] A. Ecker, P. Berens, G. Keliris, M. Bethge, N. Logothetis, and A. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327:584–587, 2010.

[45] T. Euler, P.B. Detwiler, and W. Denk. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature*, 418:845–852, 2002.

[46] A.A. Faisal, L.P.J. Selen, and D.M. Wolpert. Noise in the nervous system. *Nature Reviews Neuroscience*, 9:292–303, 2008.

[47] A. Fontanini and D.B. Katz. Behavioral states, network states, and sensory response variability. *Journal of Neurophysiology*, 100:1160–1168, 2008.

[48] M.D. Fox, A.Z. Snyder, J.L. Vincent, and M.E. Raichle. Intrinsic fluctuations within cortical systems account for intertrial variability in human behavior. *Neuron*, 56(1):171–184, 2007.

[49] F. Gabbiani, H.G. Krapp, C. Koch, and G. Laurent. Multiplicative computation in a visual neuron sensitive to looming. *Nature*, 420:320–324, 2002.

[50] L. Gabernet, S. P. Jadhav, D. E. Feldman, M. Carandini, and M. Scanziani. Somatosensory integration controlled by dynamic thalamocortical feed-forward inhibition. *Neuron*, 48:315–327, 2005.

[51] S. Ganguli, D. Huh, and H. Sompolinsky. Memory traces in dynamical systems. *PNAS*, 105:18970–18975, 2008.

[52] E. Ganmor, R. Segev, and E. Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *PNAS*, 108:9679–9684, 2011.

[53] T.J. Gawne and B.J. Richmond. How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience*, 13(7):2758–2771, 1993.

[54] J. Gjorgjieva, C. Clopath, J. Audet, and J.P. Pfister. A triplet spike-timing-dependent plasticity model generalizes the bienenstock-cooper-munro rule to higher-order spatiotemporal correlations. *PNAS*, 108(48):19383–19388, 2011.

[55] P.S. Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14:477–485, 1995.

[56] L. Gollo, O. Kinouchi, and M. Copelli. Active dendrites enhance neuronal dynamic range. *PLOS Computational Biology*, 5(6):e1000402, 2009.

[57] I.J. Good. The population frequencies of species and the estimation of population parameters. *Biometrika*, 40(3/4):237–264, 1953.

[58] R. Goris, J.A. Movshon, and E.P. Simoncelli. Partitioning neuronal variability. *Nature Neuroscience*, 17:858–865, 2014.

[59] S. Gozem, I. Schapiro, N. Ferre, and M. Olivucci. The molecular mechanism of thermal noise in rod photoreceptors. *Science*, 337(6099):1225–1228, 2012.

[60] M. Greschner, J. Shlens, C. Bakolitsa, G.D. Field, J.L. Gauthier, L.H. Jepson, A. Sher, A.M. Litke, and E.J. Chichilnisky. Correlated firing among major ganglion cell types in primate retina. *Journal of Physiology*, 589:75–86, 2011.

[61] G. Hahn, T. Petermann, M. Havenith, S. Yu, D. Plenz, W. Singer, and D Nikolic. Neuronal avalanches in spontaneous activity in vivo. *Journal of Neurophysiology*, 104(6):3312–3322, 2010.

[62] B. Hansen, M. Chelaru, and V. Dragoi. Correlated variability in laminar cortical circuits. *Neuron*, 76(3):590–602, 2012.

[63] R. Haslinger, D. Ba, R. Galuske, Z. Williams, and G. Pipa. Missing mass approximations for the partition function of stimulus driven ising models. *Frontiers in Computational Neuroscience*, 7(96):e00096, 2013.

[64] M. Hausser and B. Mel. Dendrites: bug or feature? *Current Opinion in Neurobiology*, 13:372–383, 2003.

[65] L. Hazan, M. Zugaro, and G. Buzsaki. Klusters, neuroscope, ndmanager: A free software suite for neurophysiological data processing and visualization. *Journal of Neuroscience Methods*, 155:207–216, 2006.

[66] M. Helias, T. Tetzlaff, and M. Diesmann. The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. *PLOS Computational Biology*, 10(1):e1003428, 2014.

[67] M. Hennig, C. Adams, D. Willshaw, and E. Sernagor. Early-stage waves in the retinal network emerge close to a critical state transition between local and global functional connectivity. *Journal of Neuroscience*, 29(4):1077–1086, 2009.

[68] G.E. Hinton and R.R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[69] Y. Hu, J. Trousdale, K. Josic, and E. Shea-Brown. Local paths to global coherence: Cutting networks down to size. *Physical Review E*, 89:032802, 2014.

[70] Y. Hu, J.. Zylberberg, and E. Shea-Brown. The sign rule and beyond: Boundary effects, flexibility, and optimal noise correlations in neural population codes. *PLOS Computational Biology*, 10(2):e1003469, 2014.

[71] H. Ito, P.E. Maldonado, and C.M. Gray. Dynamics of stimulus-evoked spike timing correlations in the cat lateral geniculate nucleus. *Journal of Neurophysiology*, 104:3276–3292, 2010.

[72] E. Jaynes. Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630, 1957.

[73] J. Jeanne, T. Sharpee, and T. Gentner. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, 78:352–363, 2013.

[74] J. Johnson, J. Spencer, S. Luck, and G. Schoener. A dynamic neural field model of visual working memory and change detection. *Psychological Science*, 20(5):568–577, 2009.

[75] M.A. Kisley and G.L. Gerstein. Trial-to-trial variability and state-dependent modulation of auditory-evoked responses in cortex. *Journal of Neuroscience*, 19(23):10451Ð10460, 1999.

[76] M.G. Kitzbichler, M.L. Smith, S.R. Christensen, and E. Bullmore. Broadband criticality of human brain network synchronization. *PLoS Computational Biology*, 5(3):1000314, 2009.

[77] A. Kohn and M. Smith. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14):3661–3673, 2005.

[78] H. Korn and P. Faure. Is there chaos in the brain? ii. experimental evidence and related models. *Comptes Rendus Biologies*, 326(9):787–840, 2003.

[79] U. Köster, J. Sohl-Dickstein, C. Gray, and B. Olshausen. Modeling higher-order correlations within cortical microcolumns. *PLOS Computational Biology*, 10:1003684, 2014.

[80] J. Kremkow, L. U. Perrinet, G. S. Masson, and A. Aertsen. Functional consequences of correlated excitatory and inhibitory conductances in cortical networks. *Journal of Computational Neuroscience*, 28:579–594, 2010.

[81] A. Kuhn, A. Aertsen, and S. Rotter. Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation*, 15:67–101, 2003.

[82] A. Kumar, S. Rotter, and A. Aertsen. Conditions for propagating synchronous spiking and asynchronous firing rates in a cortical network model. *Journal of Neuroscience*, 28:5268–5280, 2008.

[83] A. Kumar, S. Rotter, and A. Aertsen. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nature Reviews Neuroscience*, 11:615–627, 2010.

[84] P. Latham and S. Nirenberg. Synergy, redundancy, and independence in population codes, revisited. *Journal of Neuroscience*, 25:5195–5206, 2005.

[85] R. Legenstein and W. Maass. What makes a dynamical system computationally powerful? In S. Haykin, J.C. Principe, T.J. Sejnowski, and J.G. McWhirter, editors, *New Directions in Statistical Signal Processing: From Systems to Brains*, pages 127–154. MIT Press, 2007.

[86] S. Lim and M. Goldman. Noise tolerance of attractor and feedforward memory models. *Neural Computation*, 24:332–390, 2012.

[87] V. Litvak, H. Sompolinsky, I. Segev, and M. Abeles. On the transmission of rate code in long feedforward networks with excitatory-inhibitory balance. *Journal of Neuroscience*, 23:3006–3015, 2003.

[88] M. London and M. Hausser. Dendritic computation. *Annual Reviews of Neuroscience*, 28:503–532, 2005.

[89] A. Longtin. Stochastic resonance in neuron models. *Journal of Statistical Physics*, 70:309–327, 1993.

[90] J. Macke, M. Opper, and M. Bethge. Common input explains higher-order correlations and entropy in a simple model of neural population activity. *Physical Review Letters*, 106(20):e208102, 2011.

[91] O. Marre, S. El Boustani, Y. Fregnac, and A. Destexhe. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Physical Review Letters*, 102:e138101, 2009.

[92] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia. Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies. *Neural Computation*, 12(11):2621–2653, 2000.

[93] K. Martin and S. Schröder. Functional heterogeneity in neighboring neurons of cat primary visual cortex in response to both artificial and natural stimuli. *Journal of Neuroscience*, 33(17):7325–7344, 2013.

[94] D. Mastronarde. Correlated firing of cat retinal ganglion cells. i. spontaneously active inputs to x- and y-cells. *Journal of Neurophysiology*, 49:303–324, 1983.

[95] E. Maynard, N. Hatsopoulos, C. Ojakangas, B. Acuna, J. Sanes, R. Normann, and J. Donoghue. Neuronal interactions improve cortical population coding of movement direction. *Journal of Neuroscience*, 19(18):8083–8093, 1999.

[96] J.W. McClurkin, L.M. Optican, and B.J. Richmond. Cortical feedback increases visual information transmitted by monkey parvocellular lateral geniculate nucleus neurons. *Visual Neuroscience*, 11:601–617, 1994.

[97] W. S. McCulloch and W. H. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7:115–133, 1943.

[98] M. McDonnell and D. Abbott. What is stochastic resonance? definitions, misconceptions, debates, and its relevance to biology. *PLoS Computational Biology*, 5(5):1000348, 2009.

[99] M.M. Michel and R.A. Jacobs. The costs of ignoring high-order correlations in populations of model neurons. *Neural Computation*, 18(3):660–682, 2006.

[100] F. Montani, R. Ince, R. Senatore, E. Arabzadeh, M. Diamond, and S. Panzeri. The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. *Philosophical Transactions of the Royal Society A*, 367:3297–3310, 2009.

[101] T. Mora and W. Bialek. Are biological systems poised at criticality? *Journal of Statistical Physics*, 144(2):268–302, 2011.

[102] G.J. Mpitsos, R.M. Burton, H.C. Creech, and S.O. Soinila. Evidence for chaos in spike trains of neurons that generate rhythmic motor patterns. *Brain Research Bulletin*, 21(3):529–538, 1988.

[103] P.C. Murphy, S.G. Duckett, and A.M. Sillito. Feedback connections to the lateral geniculate nucleus and cortical response properties. *Science*, 286(5444):1552–1554, 1999.

[104] H. Nakahara and S. Amari. Information-geometric measure for neural spikes. *Neural Computation*, 14:2269–2316, 2002.

[105] S. Nirenberg, S.M. Carcieri, A.L. Jacobs, , and P. Latham. Retinal ganglion cells act largely as independent encoders. *Nature*, 411:698–701, 2001.

[106] S. Nirenberg and P. Latham. Decoding neuronal spike trains: How important are correlations? *PNAS*, 100:7348–7353, 2003.

[107] T. Nowotny and N. Huerta. Explaining synchrony in feedforward networks: Are mcculloch-pitts neurons good enough? *Biological Cybernetics*, 89:237–241, 2003.

[108] D. Nykamp. Reconstructing stimulus-driven neural networks from spike times. *Advances in Neural Information Processing Systems 15*, pages 306–316, 2003.

[109] E. Ohiorhenuan, F. Mechler, K. Purpura, A. Schmid, Q. Hu, and J. Victor. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466:e012019, 2010.

[110] A. Onken, V. Dragoi, and K. Obermayer. A maximum entropy test for evaluating higher-order correlations in spike counts. *PLOS Computational Biology*, 8:e1002539, 2012.

[111] A. Onken and K. Obermayer. A frank mixture copula family for modeling higher-order correlations of neural spike counts. *Journal of Physics: Conference Series*, 197:617–621, 2009.

[112] M. Oram, P. Foldiak, D. Perrett, and F. Sengpiel. The 'ideal homunculus': decoding neural population signals. *Trends in Neurosciences*, 21:259–265, 1998.

[113] S. Ostojic, N. Brunel, and V. Hakim. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *Journal of Neuroscience*, 29:10234–10253, 2009.

[114] S. Panzeri, S. Schultz, A. Treves, and E. Rolls. Correlations and the encoding of information in the nervous system. *Proceedings of the Royal Society of London B*, 266:1001–1012, 1999.

[115] T. Petermann, T. C. Thiagarajan, M. A. Lebedev, D. R. Nicolelis, M. A. andChialvo, and D. Plenz. Spontaneous cortical activity in awake monkeys composed of neuronal avalanches. *PNAS*, 106(37):15921–15926, 2009.

[116] P. Poirazi, T. Brannon, and B. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37:989–999, 2003.

[117] A. Polsky, B. Mel, and J. Schiller. Computational subunits in thin dendrites of pyramidal cells. *Nature Neuroscience*, 7(6):621–627, 2004.

[118] W. Rall. Theoretical significance of dendritic trees for neuronal input-output relations. In R. Reiss, editor, *Neural Theory and Modeling*, pages 73–97. Stanford University Press, 1967.

[119] C. Reid. Divergence and reconvergence: multielectrode analysis of feedforward connections in the visual system. *Progress in Brain Research*, 130:141–154, 2001.

[120] A. Reyes. Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro. *Nature Neuroscience*, 6(6):593–599, 2003.

[121] R. Rosenbaum, T. Tchumatchenko, and R. Moreno-Bote. Correlated neuronal activity and its relationship to coding, dynamics, and network architecture. *Frontiers in Computational Neuroscience*, 27:e00102, 2010.

[122] R. Rosenbaum, J. Trousdale, and K. Josic. Pooling and correlated neural activity. *Frontiers in Computational Neuroscience*, 4(9):e00009, 2010.

[123] E. Salinas and L.F. Abbott. Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, 1:89–107, 1994.

[124] E. Schneidman, M. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007–1012, 2006.

[125] E. Schneidman, S. Still, M. Berry, and W. Bialek. Network information and connected correlations. *Physical Review Letters*, 91:e238701, 2003.

[126] M. Shadlen and W. Newsome. The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18:3870–3896, 1998.

[127] M. Shamir. Emerging principles of population coding: in search for the neural code. *Current opinion in neurobiology*, 25:140–148, 2014.

[128] H. Shimazaki, S. Amari, E. Brown, and S. Grün. State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. *PLOS Computational Biology*, 8:e1002385, 2012.

[129] J. Shlens, G.D. Field, J.L. Gauthier, M.I. Grivich, D. Petrusca, A. Sher, A.M. Litke, and E.J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, 26(32):8254–8266, 2006.

[130] K.D. Simmons, J.S. Prentice, G. Tkacik, J. Homann, H. Yee, S. Palmer, P. Nelson, and V. Balasubramanian. Transformation of stimulus correlations by the retina. *PLOS Computational Biology*, 10:e1003344, 2013.

[131] E.P. Simoncelli and B.A. Olshausen. Natural image statistics and neural representation. *Annual Reviews in Neuroscience*, 24:1193–1216, 2001.

[132] B. Sivyer and S. Williams. Direction selectivity is computed by active dendritic integration in retinal ganglion cells. *Nature Neuroscience*, 16(12):1848–1856, 2013.

[133] S. Smith, I. Smith, T. Branco, and M. Hausser. Dendritic spikes enhance stimulus selectivity in cortical neurons *in vivo*. *Nature*, 503:115–120, 2013.

[134] J. Sohl-Dickstein, P. Battaglino, and M. DeWeese. New method for parameter estimation in probabilistic models: minimum probability flow. *Physical Review Letters*, 107:e220601, 2011.

[135] M. Sompolinsky, H. Yoon, K. Kang, and M. Shamir. Population coding in neuronal systems with correlated noise. *Physical Review E*, 64:e051904, 2001.

[136] S. Staude, S. Grün, and S. Rotter. Higher-order correlations and cumulants. In S. Grün and S. Rotter, editors, *Analysis of Parallel Spike Trains*, pages 253–280. Springer, 2010.

[137] S. Staude, S. Rotter, and S. Grün. Cubic: cumulant based inference of higher-order correlations in massively parallel spike trains. *Journal of Computational Neuroscience*, 19:s10827–009–0195–x, 2010.

[138] P.N. Steinmetz, A. Manwani, C. Koch, M. London, and I. Segev. Subthreshold voltage noise due to channel fluctuations in active neuronal membranes. *Journal of Computational Neuroscience*, 9:133–148, 2000.

[139] G. Stuart, N. Spruston, B. Sakmann, and M. Hausser. Action potential initiation and backpropagation in central neurons. *Trends in Neuroscience*, 20:125–131, 1997.

[140] T. Tetzlaff, T. Geisel, and M. Diesmann. The ground state of cortical feed-forward networks. *Neurocomputing*, 44-46:673–678, 2002.

[141] G. Tkacik, O. Marre, D. Amodei, E. Schneidman, W. Bialek, and M. Berry. Searching for collective behavior in a large network of sensory neurons. *PLOS Computational Biology*, 10:e1003408, 2014.

[142] T. Toyoizumi. Nearly extensive sequential memory lifetime achieved by coupled nonlinear neurons. *Neural Computation*, 24:2678–2699, 2012.

[143] L.N. Trefethen and N. Embree. *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*. Princeton University Press, 2005.

[144] S. Trenholm, A. McLaughlin, D. Schwab, M. Turner, R. Smith, F. Rieke, and G. Awatramani. Nonlinear dendritic integration of electrical and chemical synaptic inputs drives fine-scale correlations. *Nature Neuroscience*, 17:1759–1766, 2014.

[145] Philipp Khuc Trong and Fred Rieke. Origin of correlated activity between parasol retinal ganglion cells. *Nature Neuroscience*, 11(11):1343–1351, 2008.

[146] M. van Rossum, B.J. O'Brien, and R.G. Smith. Effects of noise on the spike timing precision of retinal ganglion cells. *Journal of Neurophysiology*, 89(5):2406–2419, 2003.

[147] M. van Rossum, G. Turrigiano, and S. Nelson. Fast propagation of firing rates through layered networks of noisy neurons. *Journal of Neuroscience*, 22(5):1956–1966, 2002.

[148] T. Vogels and L.F. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *Journal of Neuroscience*, 25(46):10786–10795, 2005.

[149] J.A. White, J.T. Rubinstein, and A.R. Kay. Channel noise in neurons. *Trends in Neurosciences*, 23(3):131–137, 2000.

[150] S.R. Williams and G.J. Stuart. Action potential backpropagation and somato-dendritic distribution of ion channels in thalamocortical neurons. *Journal of Neuroscience*, 20(4):1307Ð1317, 2000.

[151] S. Wu, H. Nakahara, N. Murata, and S. Amari. Population decoding based on an unfaithful model. *Advances in Neural Information Processing Systems 15*, pages 167–173, 2000.

[152] N. Xu, M. Harnett, S. Williams, D. Huber, D. O'Conner, K. Svoboda, and J. Magee. Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature*, 492:247–251, 2012.

[153] S. Yu, H. Yang, H. Nakahara, G. Santos, D. Nikolic, and D. Plenz. Higher-order interactions characterized in cortical activity. *Journal of Neuroscience*, 31(48):17514–17526, 2011.

[154] E. Zohary, M. Shadlen, and W. Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370:140–143, 1994.

[155] J. Zylberberg and E. Shea-Brown. Input nonlinearities shape beyond-pairwise correlations and can improve information transmission by neural populations. *arXiv*, 2013.