Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

Vijay Singh

Towards the understanding of Network Information Processing in Biology

By

Vijay Singh
Doctor of Philosophy
Physics

Ilya Nemenman, Advisor

Stefan Boettcher, Committee Member

Fereydoon Family, Committee Member

Sergei Urazhdin, Committee Member

Paul Goldbart, Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James. T. Laney School of Graduate Studies

Date

Towards the understanding of Network Information Processing in Biology

By

Vijay Singh
M.Sc., Indian Institute of Technology, Kanpur, 2010
B.Sc., St. Stephen's College, Delhi University, 2008

Advisor: Ilya Nemenman, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Physics
2015

Abstract

Towards the understanding of Network Information Processing in Biology

By
Vijay Singh

Living organisms perform incredibly well in detecting a signal present in the environment. This information processing is achieved near optimally and quite reliably, even though the sources of signals are highly variable and complex. The work in the last few decades has given us a fair understanding of how individual signal processing units like neurons and cell receptors process signals, but the principles of collective information processing on biological networks are far from clear. Information processing in biological networks, like the brain, metabolic circuits, cellular-signaling circuits, etc., involves complex interactions among a large number of units (neurons, receptors). The combinatorially large number of states such a system can exist in makes it impossible to study these systems from the first principles, starting from the interactions between the basic units. The principles of collective information processing on such complex networks can be identified using coarse graining approaches. This could provide insights into the organization and function of complex biological networks. Here I study models of biological networks using continuum dynamics, renormalization, maximum likelihood estimation and information theory. Such coarse graining approaches identify features that are essential for certain processes performed by underlying biological networks. We find that long-range connections in the brain allow for global scale feature detection in a signal. These also suppress the noise and remove any gaps present in the signal. Hierarchical organization with long-range connections leads to large-scale connectivity at low synapse numbers. Time delays can be utilized to separate a mixture of signals with temporal scales. Our observations indicate that the rules in multivariate signal processing are quite different from traditional single unit signal processing.

Towards the understanding of Network Information Processing in Biology

By

Vijay Singh
M.Sc., Indian Institute of Technology, Kanpur, 2010
B.Sc., St. Stephen's College, Delhi University, 2008

Advisor: Ilya Nemenman, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Physics
2015

Acknowledgements

Prof. Ilya Nemenman is a great advisor. He is an even better human being. I thank him for teaching me different aspects of science and research. This work would not have been possible without his encouragement and support.

Prof. Stefan Boettcher has been very inspirational in doing science and keeping a positive attitude. I thank him for the time he spent in patiently answering all my questions.

Martin Tchernookov has been a colleague, a mentor and a friend. He was the "go to" person for all questions, scientific and non-scientific. He gave me the guidance I needed whenever I went into an existential crisis. I thank him for his help and support in my research.

I would like to thank my thesis committee members Prof. Fereydoon Family, Prof. Sergei Urazhdin and Prof. Paul Goldbart for their support and encouragement and keeping me aware of the bigger picture.

I would like to thank the faculty, students, postdocs and staff at Emory Physics department, especially Prof. H.G.E. Hentschel, Dr. Thomas Bing, Dr. Jakub Otwinowski, Dr. Trent Brunson, Dr. Kenneth Desmond, Dr. Adonis Bovell, Dr. Neil Anthony, Dr. Justin Pye, Skanda Vivek, Dr. Andrew Mugler, Joseph Natale, Mr. Calvin Jackson, Mr. Jason Boss, Mr. Art Kleyman and all members of Young Emory Physicists.

Contents

List of Figures

Chapter 1

Introduction

Sensory systems in all biological organisms perform incredibly well in processing the information coming from the environment and identify the sources of signals reliably [1–4] and optimally [5, 6]. The sensory systems seem to achieve this even though the signals from the environment are noisy and corrupted by the complex interactions between multiple signal sources. Living organisms utilize a variety of signal processing units (neurons, cell receptors, genetic networks, etc.) to perform the required information processing. The number of such signal processing units can range from hundreds (receptors on individual cells) [7] to about a hundred billion (human brain) [8]. Just like the sources, these sensors interact in a complex manner among themselves through direct and indirect physical connections. While a lot is known about information processing in individual information processing units [9–15], the same can not be said about information processing on complex biological networks [16]. Collective information processing that utilizes population codes [17–21] and interaction among the sensors [22–25] has been considered only recently. The combinatorial possibilities arising from the complex interactions between large number of processing units and the limited knowledge of the interactions between these units makes it difficult to understand the rules of information processing starting from the first principles. Employing coarse graining techniques from physics [26, 27], one could utilize the limited experimental knowledge we have, to identify the essential features of these systems and to understand the principles of information processing on these networks. Here we use some of these techniques to study information processing on some biological networks and identify the features that are relevant to perform a certain task.

One of the most fascinating sensory systems that performs rapid information processing is the human brain [28]. The brain processes gigantic amounts of data at a fast rate [29, 30] with incredible accuracy [31]. This is done through the 100 billion neurons present in the brain that are interconnected in a complex manner [32]. Object recognition in the visual cortex is one of the many functions that are performed by the brain using these complex interactions [33]. An understanding of object recognition can provide insights to how sensory input is processed in the brain to make a map of the environment [34]. Such understanding has wide scope of practical applications in computational vision [35] and medical research [36]. In Chapter 2 we attempt to understand contour detection in the primary visual cortex using a novel continuum dynamics approach to study the activity of cortical neurons. The visual input coming from the retina is processed in the visual cortex part of the brain to form a representation of the visual field [34]. This representation in terms of the spiking activity of the neurons lasts for time scales that are larger than that of individual spikes [37] and covers spatial extents that are much larger than the size of individual neurons [38]. Moreover the number of neurons in the visual cortex [39], although smaller than the whole brain, is still quite large. This suggests that the functional aspects of the neurons in the visual cortex that are important for tasks like object recognition might be an emergent property of interactions among large number of cortical neurons. To study this emergent behavior of cortical neurons, we model the dynamics of the neurons in the primary visual cortex as a continuum field. This continuum model aims at identifying the interactions in primary visual cortex neurons that enables the neurons in collective detection of the objects contours present in the visual field. Our model suggests that neurons in the primary visual cortex utilize anisotropic long range interaction patterns to identify the global scale features present in the visual field and bind them together to detect an object contour.

The long range connections in the cortical neurons not only provide the scope to identify the global features present in the stimulus, but also provides the ability to have long range information transmission. Several experiments studying the spontaneous activity in a network of cortical neurons have observed bursts of action potentials [40–43], where a single action potential initiated spontaneously at a point results in a cascade of action potentials in the network with the size distribution of these action potential cascades approximating a power law. The power law distribution of the size suggests that cortical neurons organize themselves in a critical state where they can transmit information at all possible scales [44]. The observed criticality and power law distribution [45] suggests an equivalence of these neuronal avalanches to the phenomenon of forest fires, earthquake models and percolation in physics [46]. This criticality in the cortical

networks could be a result of the mixture of local short range and hierarchical long range synaptic connections present in the cortex. To what extent does the topology of the underlying network dictate the rules of information transmission on cortical networks is a question that can be studied using the standard real-space renormalization group (RG) techniques developed to study percolation on hierarchical networks. Chapter 3 discusses the techniques through which the self-similarity of regular hierarchical networks is utilized to study percolation. We show that a mixture of long range small world bonds and short range bonds results in novel phenomenon on complex networks, one of them being the existence of discontinuity in cluster size distribution in ordinary percolation. We show that this discontinuity is generic for networks with a hierarchical structure of long range bonds. This suggest that the critical phase [45] and optimal connectivity [47] observed in cortical networks might result form the topology of the underlying complex network. Utilizing such RG approaches on model cortical networks can provide insights into understanding large scale communication in brain [48] and phenomenon like synchronization in epileptic seizures [44].

The complexity in biological networks might also result from the complex structure in the signal itself. Be it the olfactory system, the immune system, cellular signal transduction, vision or auditory system, the signal that is received by these system is often a combination of signals from many individual sources. Identifying the different sources from such a combination is task that is performed very well by sensory systems. The neural mechanism for such multi-signal identification has been an active area of research [49, 50], but the signal estimation in molecular systems has mainly focused on one-to-one interactions [51–57]. Research in neuroscience suggests that neurons utilize not only the average rate of spiking but also the temporal structure of the spike timing to represent a signal [58, 59]. In chapter 4, we explore whether these ideas from neuroscience can be applied to cellular receptor systems. We discuss how a system of cellular receptors could utilize the temporal structure present in the signal to infer multiple ligands present in the environment. Here we study the effect of cross-talk between multiple ligands and receptors on the concentration estimation of ligands. In particular we ask the question: Is it possible to utilize the sequence of precise timings of binding and unbinding events on a receptor to estimate the concentration of multiple ligands using a single receptor? We show that if the entire sequence of binding and unbinding on the receptor is accessible, one could utilize this information to estimate the concentration of multiple ligands using the statistics of bound and unbound duration of the receptor. We describe how a complex, and yet common, cellular network can be employed for such multiple ligand estimation.

In chapter 5, we ask how the interaction among multiple receptors affects the information processing and concentration estimation in ligand-receptor systems. We use ideas from information theory to study the effects of interactions between receptors on the information between the ligand concentration and the response of receptors. We present simple biophysically realistic models of ligand-receptor systems where the receptors are coupled through a diffusive interaction. We show that this diffusive interaction results in a correlation structure in this system where the signal induced and noise induced covariances of responses are orthogonal. An interaction with such correlation structure is believed to result in higher information about the signal compared to a system with independent units [17–21, 25]. This is the well-known "sign-rule of correlations" from neuroscience literature [17, 60]. We show that this rule does not hold for a simple receptor system coupled with diffusive interactions. Although the extrinsic signal induced correlation and intrinsic noise induced correlations are orthogonal in this system, the mutual information is not higher for interacting receptors compared to independent receptors. This happens because the interactions don't just change the covariance but also the variance of the responses, an effect ignored in prior work [18, 21, 60]. These ideas were typically developed in context of neural systems where the covariance matrix could not be calculated but had to be imposed externally. Our analysis is a simple special case where the interaction changes the entire covariance matrix in a such a way that the total information remains constant with or without the interactions.

Chapter 2

A continuum model of primary visual cortex for contour detection

(This chapter is based on: *Director Field Model of the Primary Visual Cortex for Contour Detection.* V Singh, M Tchernookov, R Butterfield, and I Nemenman, *PLoS One*, 2014 [61].)

2.1 Introduction

To recognize an object in a visual scene, humans and other primates process visual signals relayed through the retina [62] in the ventral stream of the cortex. Contour detection is a crucial part of this process (Fig. 2.1). It is carried out at early stages of the processing in the primary visual cortex (V1) of the brain [63]. V1 consists of hundreds of millions of neurons organized topographically into columns of $\sim 10^4 \ldots 10^5$ neurons each. Neurons in each column receive inputs from a localized part of the visual field (called classical, or feed-forward receptive field). They are directionally selective, responding primarily to oriented edges within their receptive fields [64, 65]. Computational vision models that account for such receptive fields of individual neurons [66–71] typically incorporate them within feedforward hierarchical structures similar to the cortex [34, 72, 73]. Such feedforward models account for the visual processes on short time scales, and achieve error rates as low as $\sim 10 - 20\%$ on typical object detection tasks [71, 74].

It is believed that, *in vivo*, the error rate is reduced by orders of magnitude by contextual information that influences local processing, which may not be captured fully in such models [38, 75]. These collective, recurrent dynamics span large spatiotemporal scales and are mediated through thousands of axons laterally connecting distant columns [76]. These interactions are believed to suppress the clutter present in the visual field, while simultaneously binding edges into contours [77].

Our goal here is to build *the simplest model of the primary visual cortex* that simultaneously achieves two contradictory tasks: clutter removal and occlusion filling. While occlusion filling helps in binding an incomplete contour together for object recognition, clutter removal gets rid of the parts that con not belong to the object contour. It is believed that detecting the complete contour of an object is important in order to recognize the object. Here we aim at making identifying the interactions that are important for long range contour detection. We do not aim at the state of the art performance on complex natural images, but rather ask what is the smallest set of computational primitives that must be implemented in a model to achieve such detection and integration of long contours in a nontrivial setting. For this, we focus on a proposal of a specific lateral connectivity among V1 neurons [78, 79], which incorporates the Hebbian constraint that neurons that are excited simultaneously by the same long, low-curvature contours should activate each other [78]. However, in our model, we do not reproduce the complexity of V1, which has ~ 100 million neurons, with each neuron having $\gtrsim 10^3$ connections, some extending for many millimeters. Instead, unlike most individual neuron based discrete models, we represent the activity as a *coarse-grained, continuous neural field*, which we model as a complex-valued field on the complex plane, $W(z)$. The magnitude and the phase of W represents the level of excitation and the orientation of the dominant contour element at point z in 2 dimensional space, respectively. This coarse graining helps us to identify the minimal features of the neural structure and dynamics that are essential for contour recognition.

Importantly, our complex field approach is significantly simpler than most other coarse-grained models, thus pushing the limits in identification of the minimal set of the required computational primitives. Indeed, typically the neural firing rate is represented as a real function of three variables (position in the visual plane and the directional sensitivity) [80, 81]. In our model, the firing rate is represented as a complex function of a complex variable (or, equivalently, two real variables), which, manifestly, has fewer degrees of freedom. Previous approaches that used a similar complex field representation [82, 83] have focused on development, rather than on the visual performance of the cortex. Thus

it has been unclear if the simplified, lower-dimensional model can solve complex visual tasks. Here we answer this question affirmatively.

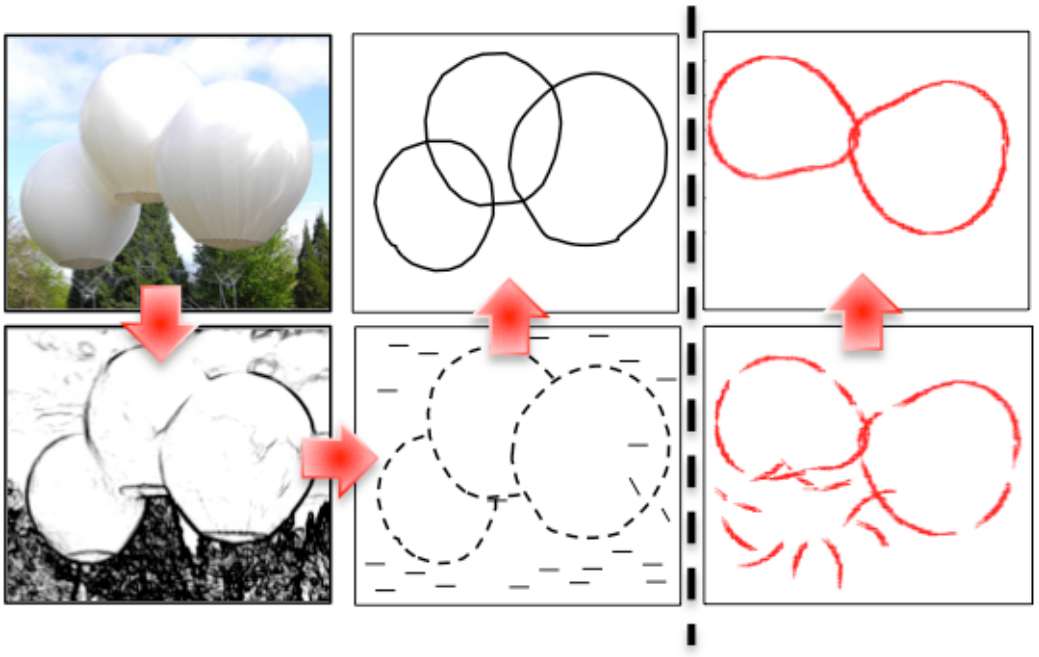

FIGURE 2.1: **Contour Reconstruction Task**

A 2d image (left top; credit: `'Pont de Singe', Olivier Grossetete. Photo: Thierry Bal`) is recorded as a field of contrast by the retina and the lateral geniculate nucleus (LGN) (left bottom). V1 neurons respond to regions of contrast changes in a direction-selective manner, performing edge detection (middle bottom). The information from edges is integrated to reconstruct long contours (middle top). In this paper, we model the visual process starting from edges in V1; sample input (bottom) and output (top) to our model are on the right.

2.2 Continuum model of contour detection

We define the dynamical variables in our coarse-grained model as the neural firing rate $s(x, y)$, $s \geq 0$, over the two dimensional plane $\mathcal{R}(x, y)$, and the orientation preference Θ of neurons, both averaged over a microscopic patch of the cortex, which still contains many thousands of neurons. Such averaging is traditional in, for example, fluid dynamics, where continuous dynamics is sought from discrete agents. The neural activity is invariant under parity (i. e., an edge or its π rotation results in the same activity). Further, two equal edges at one point oriented $\pi/2$ apart lead to cross orientation suppression, not forming a dominant orientation at the point [84]. Thus the fields s and Θ are combined into a time varying complex field $W(z, t)$ in a somewhat uncommon way,

forming an object called a *director* [85]: $W(z,t) = s \times e^{i2\Theta}$. The magnitude of this field is the average firing rate, and the argument is twice the average dominant orientation preference of neurons at a point $z = |z|e^{i\phi} = x + iy$ [82, 83]. We similarly coarse-grain the input images, identifying the dominant orientation at every point (see *Methods*). This orientation field serves as the input to the model. Note the crucial reduction in the number of degrees of freedom in going from a more traditional description $s = s(x, y, \Theta)$ to $W = W(z)$. One of the costs of the simplification is the lost ability to represent multiple different orientations at the same point, which happens when contours intersect. Correspondingly, one of our goals is to verify that this loss does not make it impossible to perform non-trivial visual tasks.

Neurophysiological and psychophysical experiments [38, 75, 77, 86, 87] and theoretical considerations [78] suggest that neurons in V1 are laterally connected such that active neurons excite nearby neurons with collinear or large-radius co-circular directional preference. Conceptually, simultaneous input from several collinear or co-circular neurons can excite other neurons that might otherwise not be getting enough excitation from the visual field due to occlusion or noise, cf. Fig. 2.2**A**. At the same time, neurons responding to high spatial frequency clutter elements do not get sufficient lateral excitation, and their activity decays. These collective dynamics integrate information over large spatial scales.

We represent these phenomena in a traditional linear-nonlinear model, where the neural field at a point z is affected by a combination of lateral synaptic inputs:

$$\frac{dW(z,t)}{dt} = F_{\delta_{\text{th}}}[I(z,t)] - r(z,t) + j(z,t). \tag{2.1}$$

Here $F_{\delta_{\text{th}}}$ is some sigmoidal function of the excitatory input $I(z,t)$, $r(z,t)$ describes the inhibitory contribution to the field, and $j(z,t)$ is the stimulus.

The excitatory input, $I(z,t)$, combines synaptic input from all points z' in its interaction region 'Ex'

$$I(z,t) = \int_{\text{Ex}} \mathrm{d}^2 z' \; K[z - z'|W(z,t)] \; W^*(z',t), \tag{2.2}$$

where $K[z - z'|W(z,t)]$ is the excitatory interaction kernel between the fields at point z' and z, when the field at z is $W(z,t)$. The kernel for an arbitrary orientation of $W(z)$ can be defined by an appropriate rotation of the kernel defined for $W = 1$ (parallel to

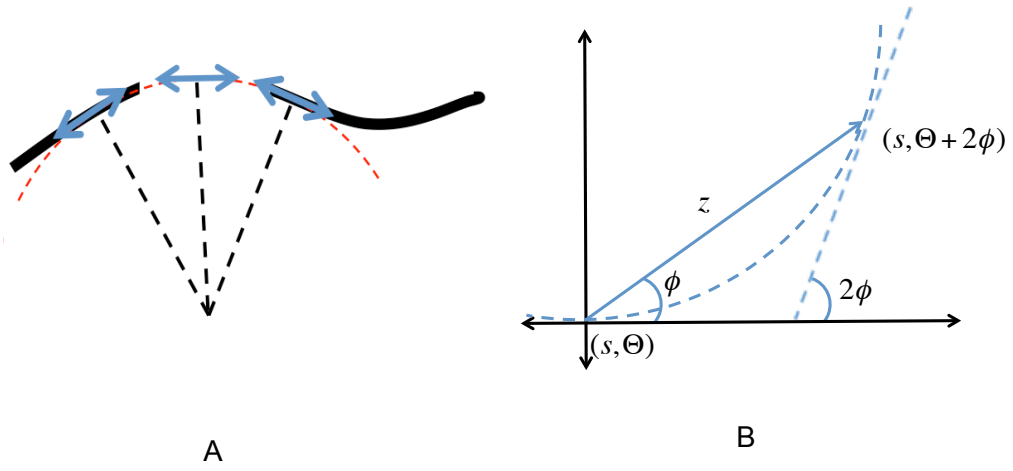

FIGURE 2.2: **Co-circularity condition**

(A) Neurons send excitatory signals along approximately co-circular directions. Thus neurons in occluded gaps may get enough excitatory input along smooth contours to get excited without direct visual input. (B) The orientation at two points is said to be co-circular if they are tangential to the circle connecting the two points. If the orientation preference at the origin is along the real axis, the co-circular edge at a point $z = |z|e^{i\phi}$ has the orientation 2ϕ. Multiplication by $e^{i2\phi}$ can be written as: $e^{i2\phi} = (e^{i\phi})^2 = (z/|z|)^2 = (z/z^*)$.

the real axis):

$$K[z - z'|W] = K\left[(z - z')\, e^{-i\frac{\arg(W)}{2}}|1\right]. \tag{2.3}$$

Co-circular excitation (Fig.2.2 B) may be represented as

$$K[z|1] = \left(\frac{z}{z^*}\right)^2 \times \exp\left\{-\frac{|z|^2}{2\,\sigma^2} - \mu\frac{|\mathrm{Im(z)}|}{[\mathrm{Re(z)}]^2}\right\}. \tag{2.4}$$

The first term, derived in Fig. 2.2, determines the field direction at z that is co-circular to the field at z'. Since we are rotating a director by *phi*, the argument of z, we need to square the term derived in Fig. 2.2. The σ term in the exponent determines the spatial range of the excitation. The μ term determines the smallest radius for which substantial co-circular excitations still exist, giving the kernel and hence the induced dynamics their characteristic bow-tie shapes [78], see Fig. 2.3. Note again the reduced complexity of this model, where the kernel is defined by just two real-valued parameters, instead of being inferred empirically from the data in a form of a multi-dimensional matrix, as in Ref. [79] and references therein.

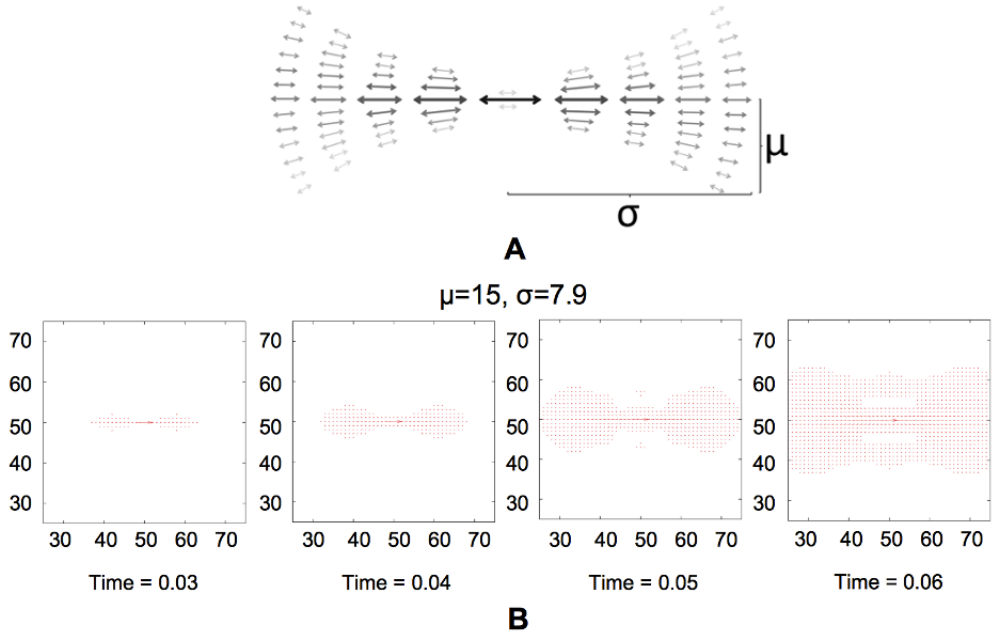

FIGURE 2.3: **Shape of the interaction kernel**

(A)Schematic shape of the interaction kernel $K[z - z'|1]$. Arrows represent the orientation preference and darkness and size represent the magnitude. (B) Results of dynamics with the kernel K with the current $j(z,t) = \delta(z)\delta(t)$. Here, as everywhere in this work, we use $A = 5, \delta_{\text{th}} = 5, \sigma = 7.9, \mu = 15, \gamma_{\text{g}} = 0.012, \gamma_{\text{l}} = 1$, which optimizes the performance according to a genetic algorithm search over the parameter space, see *Methods*.

We define the input nonlinearity using a complex step function:

$$F_{\delta_{\text{th}}}(I) = \frac{I}{|I|} \times \; A\, H(|I| - \delta_{\text{th}}), \tag{2.5}$$

where H is the Heaviside step function and A determines the maximum excitation strength. Smoother sigmoidal nonlinearities were tried as well, but this had little effect on the results presented below. If the total excitatory input is higher than the threshold δ_{th}, then the field $W(z)$ gets a positive increment in the direction of the total input. For this, the excitatory contribution from a large part of the neural field must align in the same direction, representing coincidence detection. While importance of this coincidence detection phenomenon in vision is unclear, it is crucial in the context of auditory signal processing [88]. Thresholding also suppresses clutter-induced spurious excitations, as it is unlikely that the excitatory input from short clutter elements becomes higher than the threshold in the absence of contextual support from long contours.

The inhibition term r represents two distinct phenomena: local relaxation, which depends on the local field magnitude [89], and global inhibition [90], which keeps the

activity of the entire neural field in check (presumably through intermediate inhibitory neurons, not modeled explicitly). In the spirit of writing the simplest possible model, we represent inhibition as linear, resulting in:

$$r(z) = \gamma_l W(z) + \gamma_g H(|W(z)|)\frac{W(z)}{|W(z)|}\int_{\text{In}} \mathrm{d}^2 z' |W(z')|. \tag{2.6}$$

Here γ_l and γ_g determine the rates of local and global inhibition, and 'In' stands for the range of global inhibitory interactions. Combined with the non-linear excitation, this linear inhibition produces bimodal asymptotic field values. Hence, neurons can be defined as 'active' or not.

2.3 Methods

2.3.1 Image generation

Since our focus is not on practical image processing algorithms, we focus on synthetic images in this work, as in [79]. This makes it easier to analyze effects of various image properties on the performance.

Targets – The "amoeba" objects (long closed contours with gaps) are generated by choosing a center at a random point in the image, and then drawing the amoeba around this point in polar coordinates, with the radius as a superposition of periodic functions with different radial frequencies, $\rho(\phi) = \Sigma_{k=0}^{n} a_k \sin(k\phi + \phi_k)$. The Fourier coefficients a_k are generated randomly from a normal distribution ($\sigma = 1$), with $k \leq n = 3$, and the phases ϕ_k are uniformly distributed between 0 and 2π. To create amoebas that are about the same size, the coefficients are further constrained such that the minimum and the maximum radii of the resulting amoeba and their ratios obey $0.2L < R_{\min} < R_{\max} < 0.3L,\ 0.4 < \frac{R_{\min}}{R_{\max}} < 0.6$, where L is the image size. The input current then is $j(z) = \delta(z - z_e)e^{i2\Theta}$, for every point z_e within 1 lattice spacing away from any point on the amoeba contour, where Θ is tangential to the contour at that point. While generating an amoeba, we also determine an exclusion region around it of 8 lattice sites. Clutter elements (see below) with orientations parallel to the closest amoeba segment are not allowed in these regions. Without such exclusion, a nearby clutter edge could help amoeba detection, which would artificially elevate the measured performance. We prefer to err on the side of underestimating the performance, and hence we remove these ambiguous cases.

Occlusions – We simulate occlusions and noise in real-world images by removing parts of amoebas. A random number of 2-4 segments with random angular length combining to the total of $\sim 25\%$ of the amoeba length are chosen at random positions along the amoeba contour. Within the chosen segments, the input current $j(z)$ is then set to zero.

Clutter – We need the clutter to be indistinguishable from the targets by curvature, brightness, and other local statistics, so that object detection is impossible without long-range contextual contour integration afforded by co-circular connectivity. Thus clutter is generated by first generating an amoeba as described above, partitioning it into segments, and then randomly shuffling and rotating the segments to break long-range contour continuity. Specifically, the model cortex (100 square lattice used for simulation) is divided into 5×5 square regions, which are then randomly permuted. The center-of-mass (CoM) of an image within each region is computed, and the dominant angular orientation is determined. Then each region is rotated around its CoM by a random angle, subject to a constraint that the resulting dominant orientations of neighboring regions are different. The constraint ensures that the clutter does not form long range target-like structures.

Combined images – One or two targets and clutter resulting from breakup of one or two additional targets were then superimposed together to form test images, see Fig. 4, for an example. Clutter in the exclusion zones along the amoeba contours was then removed, as described above.

Transforming pixel images – Images used previously in psychophysics experiment (Fig. 4) were imported into MATLAB and then converted to grey scale using `rgb2grey`. The resulting matrix was then thresholded and converted into a binary matrix. A 2D Gabor filter was used to find edges in this bitmap image. For each point in the image, we find the convolution of a Gabor filter ($\sigma_{\text{smaller}} = 10$ pixel, $\sigma_{\text{longer}} = 100$ pixel, convolution range = 20 pixel$\times$20 pixel) with the image at $(360/n)$ angles where $n = 100$. The direction with the maximum convolution is taken as the orientation of the visual field at the point, and the result of the convolution as the field magnitude. The image thus processed is presented as an input for simulations.

2.3.2 Simulations

The time evolution of the model is studied on a square lattice of a linear size $L = 100$ with periodic boundary conditions using Euler iteration method. The lattice discretization is

done for simulation purposes, and should not be viewed as a representation of discrete neurons; we are not aware of numerical algorithms able to simulate our model dynamics without discretizing the space first.

In each iteration cycle we first calculate the total input I at each point z from all other points z' in the excitation region 'Ex' using a precomputed interaction kernel $K[z-z'|1]$ on a $4L \times 4L$ *kernel lattice*. Square discretization destroys the angular symmetry of the kernel evaluated at an arbitrary z. The following procedure restored the symmetry. First, to calculate the contribution from z' to $I(z)$, the kernel lattice is superimposed on the image lattice with the origin of the kernel lattice at point z of the image lattice. Next the kernel lattice is rotated by $\frac{\arg(W(z'))}{2}$ with respect to the image. Then the contribution from the point z' to $I(z)$ is $W^*(z') \times K(0, z'')$, where z'' is the point on kernel lattice closest to z'. The total input $I(z)$ is then the sum of contributions from all points z' in the excitatory interaction region 'Ex'. After the input is calculated, if $|I(z)| > \delta_{\text{th}}$, then the field is incremented $W(z, t+\Delta t) \leftarrow W(z,t) + A\frac{I(z)}{|I(z)|}\Delta t$, where Δt is the time step. To account for degradation, we finally set $W(z, t+\Delta t) \leftarrow W(z, t+\Delta t) \times \exp[-r(z) \times \Delta t / W(z, t+\Delta t)]$, where $r(z)$ is as in Eq. (2.6). To the first order in Δt, this is equivalent to the dynamics in Eq. (2.1). However, this exponential form removes the large fluctuations in $r(z)$ when $W(z) \approx 0$.

In our simulations, the excitation range 'Ex' is 3σ, where σ is the effective spatial range of the kernel $K[z - z'|W(z)]$. For global inhibition range 'In' is the entire lattice. The model is easily modified to restrict the suppression to a smaller inhibition region.

We first chose the parameter μ to be similar to the curvature of a typical amoeba. Next σ was chosen such that it was larger than the typical extent of the occluded amoeba segments. The initial values of γ_{l} and γ_{g} were determined using steady state analysis of the model, which leads to $(N\gamma_{\text{g}} + \gamma_{\text{l}})W_0 \sim F_\delta(\infty)$, where N is the typical number of points with non zero field, and F is the thresholding function as defined in Eq. (1). Setting $\gamma_{\text{l}} = 1$ and $W_0 = 1$, we thus constrain all other parameters. Using these initial values, some coarse parameter optimization was done by simply observing the simulations while the parameters were varied. After that genetic algorithm was used to optimize the model for maximum simultaneous precision and recall (see Results for definitions). We used the area under the precision-recall curve as our fitness function. Parameters were changed by a percentage drawn from a uniform distribution (from -1% to 1%) and the fitness function was recalculated for the new parameters. Then, the new parameters were either accepted or rejected according to whether $1/[1+\exp(new\ area - old\ area)/0.005] >$ random variable drawn from uniform distribution on (0,1). The parameter 0.005 acts

as the temperature. The final optimized values of the parameters used for simulations presented here were: $A = 5, \delta_{\text{th}} = 5, \sigma = 7.9, \mu = 15, \gamma_{\text{g}} = 0.012, \gamma_{\text{l}} = 1$.

The code was implemented in C, compiled with the gcc v. 4.7, and optimized with OpenMP libraries. Simulations were performed on a computer with Intel i7 2600k (clock speed 3.4 GHz). The simulation time for 250 iteration cycles for one image took about 10s. All model dynamics times were measured in units of $1/\gamma_{\text{l}}$, which was set to 1 in our simulations.

2.4 Results

Figure 2.4 (top and middle) shows the time evolution of the neural field $W(z,t)$ in our coarse-grained model for a sample input image, generated as described in *Methods*, where a large contiguous contour with gaps (an amoeba) is superimposed on clutter. The gaps model occlusion of contours by other objects and noise in the earlier stages of visual processing. Similarly, Figure 2.4 (bottom) illustrates the model output for an image previously used in psychophysics experiments with human subjects [91]. Its simplicity notwithstanding, the model performs qualitatively similar to humans in that long contours implied by collinearity of nearby edge segments are easily detected. The gaps in amoeba targets get filled, while the clutter decays with time, resulting in emergence of long contours. Note also that spurious activity appears around contours at large simulation times. Even though such hallucinations rarely happen in human vision, they are not of a big concern here since, at large times, the dynamics would be affected by feedback from higher cortical areas and eye movements, which we are not modeling. Importantly, these observation suggests that the model performance must be evaluated at finite, but not asymptotically large times.

We quantified the performance in terms of precision, P, and recall, R. Precision determines the fraction of the total field activity integrated over the image that matches the actual target contour (visible and occluded/invisible). Recall gives the fraction of the target contour that has been recovered. $P = 1$ means that there is no clutter, and $R = 1$ means that all parts of the contour have been identified. For a successful contour detection, we must have $R, P \to 1$ simultaneously. Both P and R depend on the cutoff used to decide which neurons are considered active (larger cutoff degrades clutter faster, but slows down occlusion filling), and on the time of the simulation (Fig. 2.5). Hence different cutoffs and times must be explored.

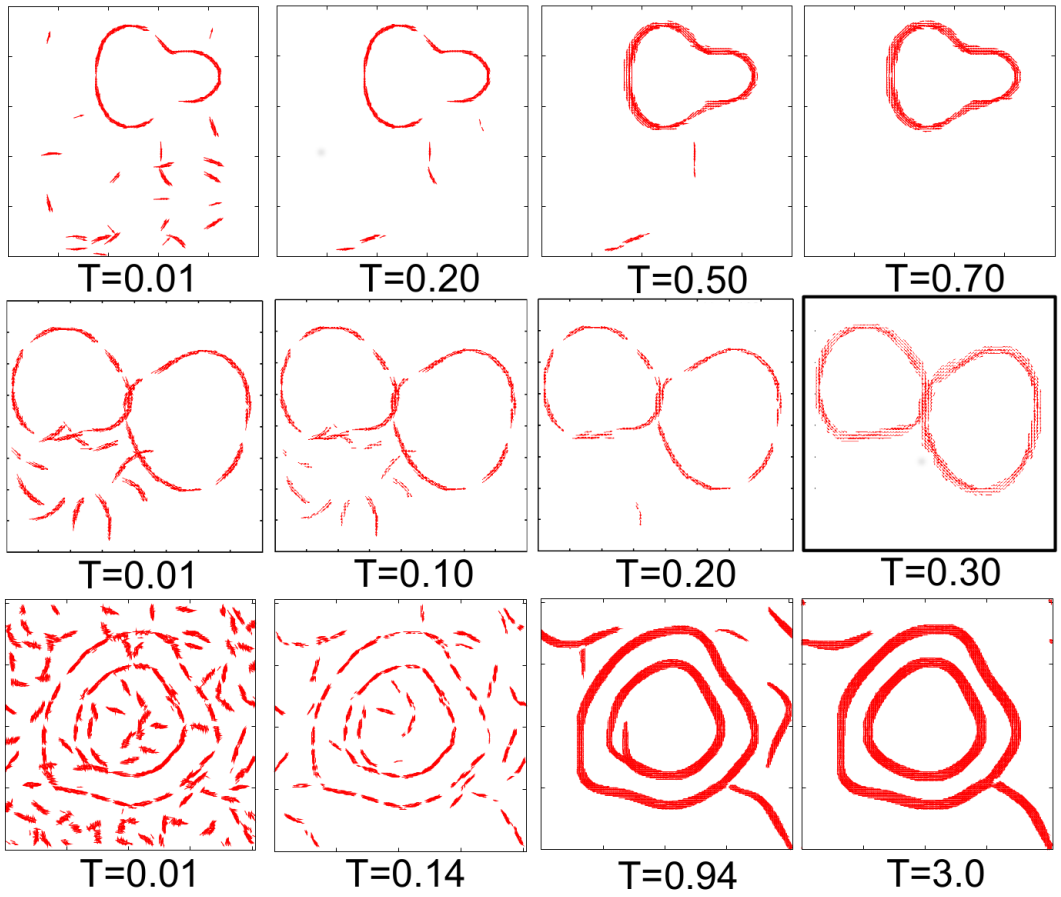

FIGURE 2.4: **Neural field dynamics**

(top and middle) Time evolution of the neural field for sample images. The magnitude (line width) and the direction of the field are plotted at every point where the strength of the field is higher than a cutoff (0.35). The parameters of the dynamics are as in Fig. 2.3. Dynamics removes the clutter and fills in the occlusion gaps. However, spurious activity (widening lines) appears for large simulation times, so that the best performance is obtained for intermediate times. (bottom) Performance of the model on an image used in psychophysics experiments [91]. Like human subjects, the model can identify, complete, and bind together long punctuated contours.

Figure 2.6**A** gives the variation of precision and recall at various cutoffs at particular times during the simulation. At $t = 0$, $(R, P) = (0.75, 0.5)$ on average, i. e., initially about 25% of the target is invisible and the total lengths of the clutter and the target segments are nearly equal. At t as small as 0.25 (with $\Delta t = 0.01$), P, R are above 0.9 simultaneously for a large set of cutoff values $(1\% - 42\%)$. Since we present the stimulus instantaneously only, its effect eventually decreases with time. Thus there is a time that optimizes performance, at which the precision vs. recall curve majorates the same curves for other times. For the data-set in Fig. 2.6, this optimal time is $t = 0.40 \times 1/\gamma_1$ (40

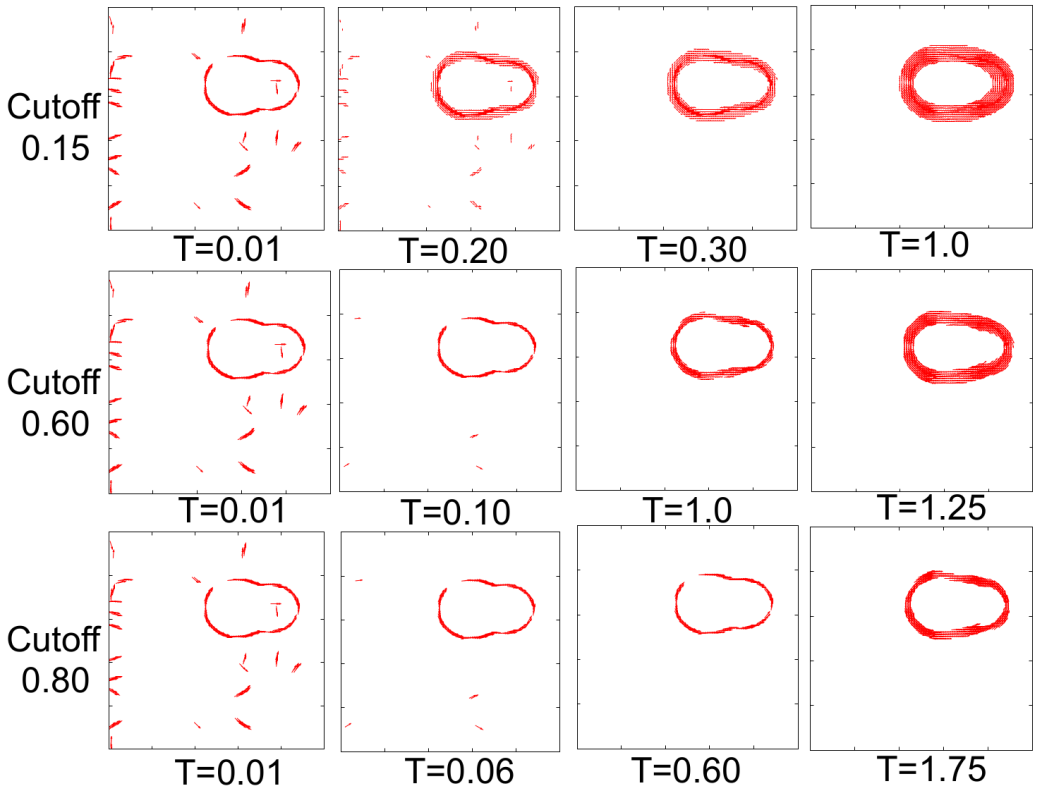

FIGURE 2.5: **Neural dynamics at different cutoffs**

Time evolution of a sample image at different cutoff values. At a lower cutoff the occlusions fills rapidly, but it takes longer to suppress the clutter. At higher cutoffs clutter removes quickly, while it takes longer to fill the gaps. Notice the spurious activity around the contours at longer times. This spurious activity is dominant at lower cutoffs.

numerical iterations), where the curve reaches $R \approx 0.97$ and $P \approx 0.95$ simultaneously.

Performance depends only weakly on the *ad hoc* details of the simulations and the data. For example, defining the threshold parameter not as an absolute value, but as a fraction of the maximum activity of the field at a given time point did not change the precision-recall curves much (Fig. 2.7). Similarly, different amounts of initial clutter had only a moderate effect if the length of the clutter elements remained the same (Fig. 2.6**B**). This is because the time scale of the clutter decay depends on the size of the segments, and not on their number. For longer segments, the decay takes longer, and hence the optimal processing time increases. The optimal processing time also increases with the linear dimension of the occlusions present in the target amoebas and with the number of occlusions (Fig. 2.6**B**). However, for all of these cases, the maximum precision and recall remain simultaneously high.

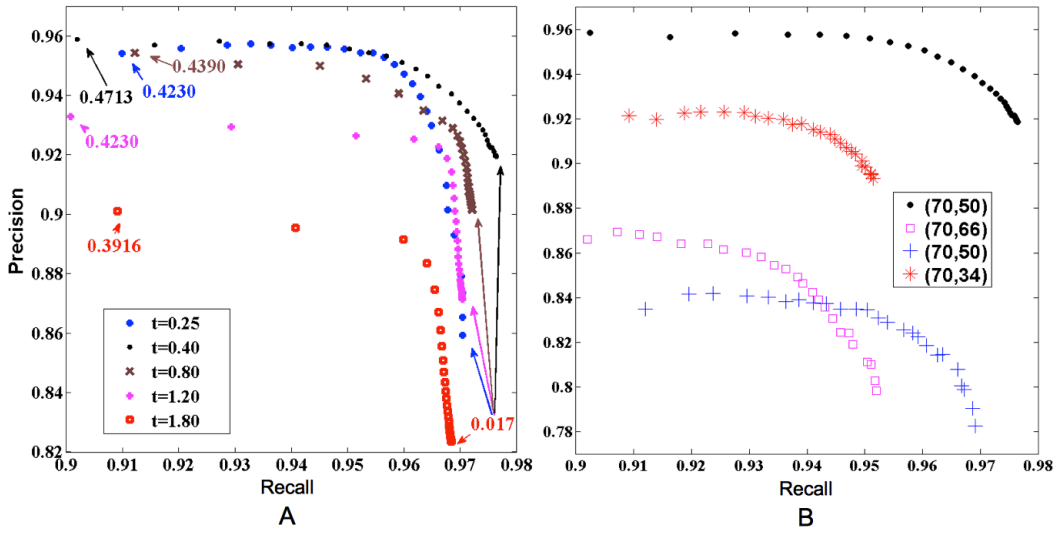

FIGURE 2.6: **Precision vs Recall with an absolute cutoff**

(A) P vs R averaged over 500 randomly generated images at various simulation times starting with $(R, P) = (0.75, 0.5)$. The numbers indicate cutoff values for a specific data point at the corresponding simulation time. Note the weak dependence on the cutoff. The simulation lengths of $t = 0.40 \times 1/\gamma_{\rm l}$ (black dots) produces the curve with the best precision and recall combination. (B) P vs R with different starting values of precision and recall averaged over 100 randomly generated images, but with the same model parameters. Legend indicates the initial (R, P). The black dots are the same as in the top panel. Red $*$'s correspond to a lower initial precision (more clutter), compared to the black dots. Blue $+$'s stand for the same initial (R, P) as black, but with the target partitioned into more shorter segments (a larger number of occlusions). Pink $\square$'s correspond to higher initial precision (less clutter), but the clutter elements are longer and harder to suppress.

2.5 Discussion

We developed a *continuum, coarse-grained model* of V1 to study contour detection in complex images, which is substantially simpler than other models in the literature, and yet still performs nontrivial visual computation. While borrowing heavily from previous research, our model differs from most previous approaches by forgoing individual neurons and describing the neural activity as a parity-symmetric continuous director field, which makes expressions for Hebbian connectivity and solutions of the model dynamics expressible in the closed form. We incorporate some experimentally observed properties of the visual neural dynamics, namely non-linear excitation, thresholding, cross orientation suppression, local relaxation, global suppression, and, crucially, co-circular excitatory connectivity [78], which brings long-range context to local edge detection.

The model identifies long object contours in computer-generated images with simultaneous recall and precision of over 90% for many conditions. It happens even though

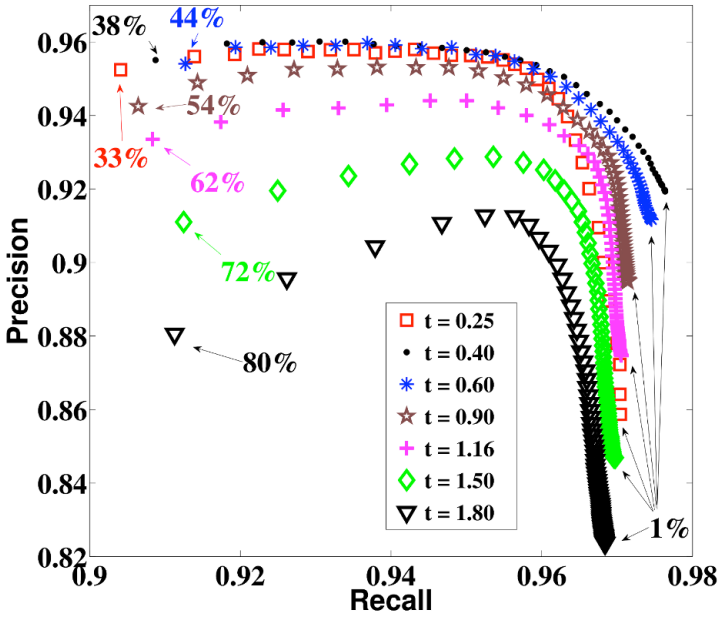

FIGURE 2.7: **Precision vs Recall with a relative cutoff**

P vs R averaged over 500 randomly generated images at various simulation times starting with $(R, P) = (0.75, 0.5)$. The numbers indicate cutoff values for a specific data point in terms of the percentage of the maximum activity of the field at the corresponding time. Note the similarity with the results in case of absolute cutoff values (Figure 6**A**).

initially large parts of objects are invisible (potentially lowering recall), and clutter is present (decreasing precision). The model fills in the occlusions and filters out the clutter based on the presence or absence of co-circular contextual edge support. In addition to the substantial simplification, this ability to *fill in the occlusions* particularly distinguishes our approach from the previous work on co-circular excitatory feedback [78, 79]. It remains to be seen to which extent the performance is affected by more natural statistics of images, and by the presence of stochasticity and synaptic plasticity in neural dynamics.

The model performs on par or better than agent-based three-dimensional models (two spatial dimensions and one orientation preference dimension), with complex, empirically specified co-circular interaction kernel [79]. This illustrates that discreteness of neurons, existence of the orientation preference as an independent variable, and intricate details of the kernel are *not crucial* for the studied visual processing function. The reduced complexity is not only conceptually appealing, but also can result in more efficient computational implementations. For example, it should be possible to augment practical feedforward models of object detection, such as [71], with the laterally connected layer

developed in this work. We expect this to lead to improvements in object recognition performance.

The model makes predictions that can be tested experimentally, such as regarding the amount of neural excitation in V1 as a function of the computation time and the duration of exposure to an image. Additionally, it predicts that the neural activity localizes to long contours with time, which can be tested with various imaging technologies. Finally, it can be used to predict the dependence of the contour detection performance on the statistical structure of images and on the exposure time. Testing such predictions in psychophysics experiments [79] will be a subject of the future work.

Finally, we notice that the neural field $W(z) = s(x, y) \times e^{i2\Theta(x,y)}$ can be mapped exactly onto the Landau - de Gennes order parameter for a two-dimensional nematic liquid crystal

$$Q_{\mu\nu} = \frac{1}{2} s \begin{pmatrix} -\sin(2\Theta) & \cos(2\Theta) \\ \cos(2\Theta) & \sin(2\Theta) \end{pmatrix}. \tag{2.7}$$

This may help solve a crucial difficulty in implementing an artificial laterally-interacting neural model: the computational cost of long-range communication. Indeed, one can think of materials with symmetry and dynamical properties such that the neural computation and the communication are performed by the intrinsic dynamics of the material itself. Potential implementations can include polarizable liquid crystals with long-range magnetic interactions, polar colloidal materials, or heterogenous solid state materials with long-range connectivity. The liquid crystal analogy suggests the use of the well-developed repertoire of theoretical physics to understand the impact of different terms in the model neural dynamics, Eq. (2.1). In particular, one can hope that the future renormalization group treatment of this dynamics will reveal the terms in the interaction kernel K that are relevant for its long-time, long-range aspects.

Chapter 3

Coarse-graining hierarchical networks

(This chapter is based on: *Scaling of Clusters near Discontinuous Percolation Transitions in Hyperbolic Networks.* V Singh, S Boettcher *Physical Review E*, 90(1):012117, 2014 [92].)

3.1 Introduction

There are two approaches to study biological information processing in large realistic complex networks. One of them utilizes real world data to infer the structural organization of networks and builds statistical models based on the inference [93, 94]. While such models have better predictive power for specific networks, they lack generalizability. The second approach uses artificial hierarchical networks that resemble real networks in only a few features, but can provide prediction about the behavior through exact results that depend on the network topology [95–97]. In the absence of detailed experimental knowledge of the connectivity of biological networks, the second approache can reveal many features of biological information processing from simple rules of network organization [98–100]. As we saw in the previous chapter, for large scale networks like the visual cortex it is possible that a lot of the details of the networks can be ignored when one attempts to understand phenomena that take place on larger scales. Coarse graining approaches like the renormalization group can provide insights about the structural and functional organization of such networks. Since we are quite far from knowing of

the exact connectivity pattern of the cortical networks [101, 102], model networks that incorporate the limited experimental knowledge we have about biological networks can be utilized to understand the information processing through coarse graining techniques.

Cortical networks in the brain have been observed to organize themselves in structures that allows large scale information processing [45, 48]. The spontaneous activity of neurons on these networks produces "avalanches" of action potentials [40, 43] that span the entire network and show a power law scaling in their size distribution and temporal duration distribution[45]. This suggests that the cortex exists in a critical state to facilitate large scale information transmission [103] making an optimal use of the synaptic connections [47]. This power law scaling of the avalanches and the large scale connectivity with few synapses shows a clear resemblance of these avalanches to criticality observed in models of percolation from statistical physics. Since cortical networks resemble small world hierarchical networks in terms of short and long range connectivity [75, 76] and the hierarchy in structural and functional organization [62, 104, 105], model hierarchical networks with small world properties can be utilized to study the organization principles of cortical networks.

Small-world hierarchical networks have generated much interest as models for the prevalent hierarchical organization in complex networks [106–110]. The hierarchical organization in the brain and long range connectivity has attracted many models utilizing the topology of these networks to study the brain [95, 96]. The recursive structure of hierarchical networks provide deeper insights into the nonlinear behavior caused by small-world connections because they yield exact results for statistical models, compared to some presumed network ensemble that often requires approximate or numerical methods. Work on percolation [111–115], the Ising model [107, 116–118], and the Potts model [119, 120] have shown that critical behavior once thought to be exotic and model-specific [110] can be universally described near the transition point [121, 122] for a large class of hierarchical networks with hyperbolic properties. In a hyperbolic structure, sites are typically randomly connected but possess a hierarchical organization of sites that allows to identify a few sites harboring many small-world bonds as central while an extensive portion of sites with less access resides on the periphery [123, 124]. Such structures are common in disordered materials [125, 126], human organizations[106], information and communication networks [124, 127], or neural networks [128, 129].

The topological structure that is required for long range information processing [40, 48], synchronization [130] and sustained cortical activity [131] share similarity with that of hierarchical small world networks. Here, we study percolation on hierarchical networks

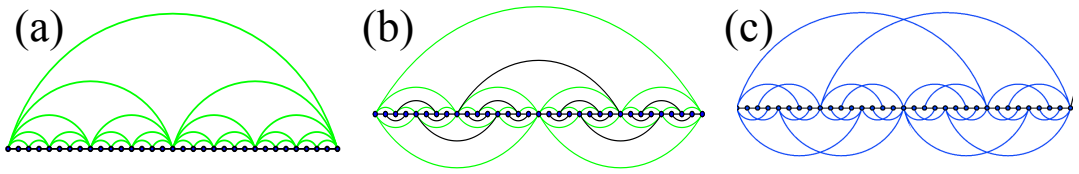

FIGURE 3.1: **Depiction of hierarchical networks**

(a) MK1, (b) HN5, and (c) HNNP. For all networks the recursive pattern that scales to the thermodynamic limit is evident. Each network features regular geometric structures, such as a one-dimensional backbone, and a distinct set of small-world links. While MK1 and HN5 are planar, HNNP is non-planar.

with long range small world connections. We extend the discussion of universality on such networks by studying the emergence of the discontinuous transition recently found in ordinary percolation [113]. Due to the discovery of percolation transitions that first appeared to be "explosive" [132–134], the dynamics of cluster formation at the onset of such a transition has been the focus of much research [135–139]. While details of the cluster size distribution $\rho(s)$ remain accessible only to simulations, we can use the renormalization group (RG) to determine the exact large-N scaling of the average size of the largest cluster,

$$\langle s_{\max} \rangle \sim N^{\Psi(p)}, \tag{3.1}$$

near the onset of the transition. Analyzing a number of different networks for site and bond percolation, we find that the behavior observed in Ref. [113] appears to be generic for hyperbolic networks. By "hyperbolic" we mean a hierarchical network with small-world properties. The hierarchy ensures the distinction between an extensive set of peripheral nodes of low centrality and ever sparser bulk nodes of increasing centrality, while small-world bonds reduce average distances to scale logarithmically with system size. In all cases, here or in related work [123, 140], it is found that within hyperbolic networks the cluster size exponent $\Psi(p)$ defined in Eq. (3.1) depends on the percolation parameter p in a nontrivial manner and has only quadratic or higher-order corrections in its approach to an extensive cluster, $\Psi \to 1$, at the transition, $p \to p_c$. This would suggest the emergence of a dominant, albeit sub-extensive, cluster long before the transition is reached.

Such a non-linear approach towards the transition contrasts with the behavior of the equivalent exponent, defined via the susceptibility, on the same networks near the critical temperature for the Ising model, and also with the predictions of the universal theory for these transitions [121], which would obtain a linear correction generically.

3.2 Small-world hyperbolic networks

The models we are studying here are familiar hierarchical networks that have become popular because they provide exact results for complex processes by way of the real-space renormalization group. MK1, depicted in Fig. 3.1(a), is the one-dimensional version of the small-world Migdal-Kadanoff hierarchical diamond lattice [107], which has been used previously to prove the existence of the discontinuous transition in ordinary percolation [113]. MK1 is recursively generated starting with two sites connected by a single edge at generation $n = 0$. Each new generation recursively combines two sub-networks of the previous generation and adds single edge connecting the end sites. As a result, the nth generation contains $2^n + 1$ vertices, 2^n backbone bonds, and $2^n - 1$ small-world bonds.

To show that this discontinuity persists for more complicated but hierarchical structures, we consider here also the Hanoi networks HN5 and HNNP, also shown in Fig. 3.1(b-c). A similar recursive procedure as described above for MK1 is also applied to obtain each new generation, however, due to their more complicated structure their basic building block at $n = 0$ consists of a triangle of three sites. For these Hanoi networks, the existence of a non-trivial bond-percolation transition has been demonstrated previously [112]. HN5 is similar to MK1 but requires a coupled system of RG-recursions. It also can be easily adapted to complement previous investigations of site-percolation [141] in a non-trivial fashion. HNNP is special in that it is a non-planar graph, and aspect that is missing from other hierarchical networks.

FIGURE 3.2: **Diagrammatic definition of generating functions**

Diagrams of $T_n(x)$ and $S_n(x, y)$ in Eqs. (3.5) for MK1 in Fig. 3.1. End sites are represented by open circles and clusters by shaded areas. $T_n(x)$ consists of one spanning cluster, labeled x, which connects both end-sites and $S_n(x, y)$ consists of two non-spanning clusters, x and y, each connected to one end-site. Isolated clusters not containing either of the end sites are ignored.

3.3 Review of cluster renormalization in bond percolation

Before we apply it to calculate exact expressions for the scaling of the average cluster size for HN5 and HNNP in the next section, we first review briefly the formalism needed to analyze the average cluster size near the bond-percolation transition, as used for MK1 in Ref. [113]. While a full understanding the dynamics of cluster formation near the discontinuous percolation transition requires knowledge of the entire cluster-size distribution, already the average size of the largest cluster $\langle s_{\max} \rangle_n$ at generation n provides profound insights. In particular, we will be focused on the system-size scaling of $\langle s_{\max} \rangle_n$ for $p \to p_c$. In the following, we derive $\langle s_{\max} \rangle_n$ using cluster generating functions.

3.3.1 Cluster generating function for MK1

We review briefly the procedure described in Ref. [113] for MK1. There, the generating functions were obtained by introducing merely two quantities: the probability $t_i^{(n)}(p)$ that both end-sites are connected to the same cluster of size i, and the probability $s_{i,j}^{(n)}(p)$ that the left end-site is connected to a cluster of size i and the right end-site to a different cluster of size j. The generating functions, as depicted in Fig. 3.2, are defined as

$$T_n(x) = \sum_{i=0}^{\infty} t_i^{(n)}(p)\, x^i \tag{3.2}$$

$$S_n(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} s_{i,j}^{(n)}(p)\, x^i\, y^j. \tag{3.3}$$

The recursion relations for these generating functions can be obtained by considering all possible configurations on three sites, as shown in Fig. 3.3, taking into account the cluster sizes as described in Ref. [113]. The graphlets on three sites are assigned to the correct two-site graphlet in the next generation, and the weights of all the graphlets that contribute to the same higher-generation graphlet are added together to get the

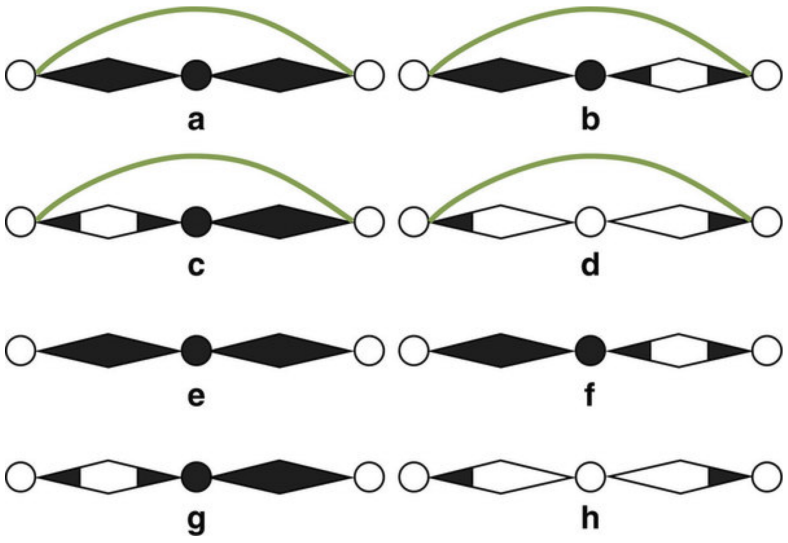

FIGURE 3.3: **Diagrammatic evaluation of generating functions for MK1**

All graphlets contributing to $T_{n+1}(x)$ and $S_{n+1}(x+y)$ in the nth generation. Graphlets (**a-e**) have end-to-end connections and contribute to $T_{n+1}(x)$ while (**f-h**) contribute to $S_{n+1}(x,y)$. The contribution of each graphlet is (**a**) $xpT_n^2(x)$ (**b**) $xpT_n(x)S_n(x,x)$ (**c**) $xpT_n(x)S_n(x,x)$ (**d**) $pS_n(x,1)S_n(1,x)$ (**e**) $x(1-p)T_n^2(x)$ (**f**) $x(1-p)T_n(x)S_n(x,y)$ (**g**) $y(1-p)T_n(y)S_n(x,y)$ (**h**) $(1-p)S_n(x,1)S_n(1,y)$. The recursion can be obtained by adding weights (**a-e**) for $T_{n+1}(x)$ and (**f-h**) for $S_{n+1}(x,y)$ resulting in Eq. (3.5). See Appendix A.1 for an algorithm to automate the evaluation.

recursion relations,

$$\begin{aligned} T_{n+1}(x) &= xT_n^2(x) \\ &\quad + p\left[2xT_n(x)S_n(x,x) + S_n(x,1)S_n(1,x)\right], \end{aligned} \tag{3.4}$$

$$\begin{aligned} S_{n+1}(x,y) &= (1-p)\left[xT_n(x)S_n(x,y) + yT_n(y)S_n(x,y)\right. \\ &\quad \left. + S_n(x,1)S_n(1,y)\right], \end{aligned} \tag{3.5}$$

as indicated in Fig. 3.3 and discussed in more detail in Appendix A.1.1.

3.3.2 Fixed point analysis for average cluster size

The recursion equations in Eq. (3.5) can be simplified by combination them into a vector $\vec{V}_n(x) = [T_n(x), S_n(x,x), S_n(x,1)]$ of distinct observables, where we focus on the largest cluster x only. The RG can now be written as

$$\vec{V}_{n+1}(x) = \vec{F}\left(\vec{V}_n(x), x\right) \tag{3.6}$$

for the nonlinear vector-function $\vec{F}$ that derives from Eqs. 3.5. As Eq. (3.2) suggest, the average size of a spanning cluster (which dominate in the cluster-size distribution) is generated by $\langle s \rangle \sim T'_n(x=1)$; any form of S_n does not affect to the spanning cluster and its contributions prove subdominant. We obtain $T'_n(x=1)$ in terms of $T_n = T_n(x=1)$ and p by linearizing the recursion relation in Eq. 3.6

$$\frac{\partial \vec{V}_{n+1}}{\partial x} = \frac{\partial \vec{F}}{\partial \vec{V}}\left(\vec{V}_n\right) \cdot \frac{\partial \vec{V}_n}{\partial x} + \frac{\partial \vec{F}}{\partial x}\left(\vec{V}_n\right), \tag{3.7}$$

near $x = 1$. Eq. (3.6) itself at $x = 1$ (where $S_n = 1 - T_n$) reduces for MK1 in each component of $\vec{V}$ to

$$T_{n+1} = p + (1-p)\, T_n^2 \qquad (T_0 = p) \tag{3.8}$$

with fixed point $T_\infty = \lim_{n\to\infty} T_n$

$$T_\infty(p) = \begin{cases} \frac{p}{(1-p)} & 0 \le p < \frac{1}{2} \\ 1 & \frac{1}{2} \le p \le 1, \end{cases} \tag{3.9}$$

providing the critical point $p_c = \frac{1}{2}$, where any spanning cluster also becomes extensive, see Fig. 3.4(a).

Ignoring the subdominant inhomogeneity in Eq. (3.7), the remaining homogeneous linear system gives the dominant contribution for V'_∞, i.e. T'_∞, S'_∞. The largest eigenvalue λ of the coefficient-matrix $\frac{\partial \vec{F}}{\partial \vec{V}}\left(\vec{V_\infty}\right)$ at the fixed point $T_\infty(p)$ becomes for MK1

$$\lambda = \begin{cases} \frac{1+3p-4p^2}{2(1-p)} + \sqrt{\frac{1-p(1-4p)^2}{4(1-p)}} & 0 \le p < \frac{1}{2} \\ 2 & \frac{1}{2} \le p \le 1. \end{cases} \tag{3.10}$$

Finally, we obtain the order parameter P_∞ as

$$P_\infty = \frac{\langle s_{\max} \rangle}{N} \sim \frac{T'_\infty}{N} \sim N^{\Psi(p)-1} \tag{3.11}$$

with the fractal exponent(3.12)

$$\Psi(p) = \log_2 \lambda. \tag{3.12}$$

Note that this implies that the largest cluster below the transition is already diverging with a non-zero power of the system size, although in a sub-extensive manner, $\Psi < 1$ for $p < p_c$, such that $P_\infty \to 0$ for $N \to \infty$. These spanning, sub-extensive clusters exist, albeit with finite probability given by $T_\infty(p)$ in Eq. (3.9), for all $0 < p < p_c$.

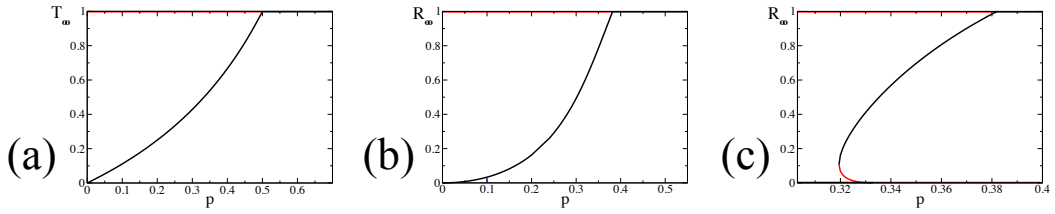

FIGURE 3.4: **Phase diagram for the probability of a spanning cluster**

(a) T_∞ for MK1 in Eq. (3.9), (b) R_∞ for HN5 in Eq. (A.14), and (c) R_∞ for HNNP in Eq. (A.10) (for $x = 1$), all as a function of bond probability p. Black lines mark stable fixed points, and red-shaded lines are unstable fixed point solutions. The critical transition, at which the probability of any site to belong to the largest cluster becomes finite and that cluster becomes extensive, occurs exactly when the probability of a spanning cluster becomes unity, at $p_c = \frac{1}{2}$ for MK1 and $p_c = 2 - \phi = 0.38197\ldots$ for both, HN5 and HNNP [112]. However, in all cases, there is a non-zero probability for a spanning cluster, albeit sub-extensive, even below p_c, due to the hyperbolic nature of these hierarchical networks. For MK1 and HN5, such a cluster can exist for all $0 < p < p_c$, while for HNNP it disappears below the branch-point singularity at $p_l = 0.31945\ldots$. Note that in each case the transition occurs at the intersection of two lines of *stable* fixed points.

This behavior for hyperbolic systems contrasts with that of regular lattices, where such sub-extensive clusters with fractal scaling only exist for $p = p_c$ and $\Psi(p) \equiv 0$ for $p < p_c$ such that all clusters remain finite or at most diverge logarithmically in N.

In Fig. 3.5(a), we show a plot of $P_\infty(p)$ for MK1 evaluated after $n = 10^k$ iterations using Eq. (3.7) displayed for $k = 1, ..., 5$ corresponding to system sizes up to $N \simeq 2^n \sim 10^{3010}$ sites. P_∞ converges slowly to zero for $p < p_c = \frac{1}{2}$. At and above p_c, it can be shown using Eq. (3.7) that T'_n is monotonically increasing with n while being bounded above by 1, thus the order parameter is positive definite for $\frac{1}{2} \leq p < 1$. The order parameter P_∞ changes discontinuously from 0 to 0.609793... at $p = p_c$ and converges to 1 for $p \to 1$. A more detailed discussion, including a proof of the discontinuity, is provided in Ref. [113].

3.3.3 Scaling behavior near the transition

From Eqs. (3.10-3.12) it is now easy to determine the scaling behavior for the average cluster size near the transition. By expanding the eigenvalue λ in Eq. (3.10) for $p \to p_c$ from below, we find that the leading behavior only has quadratic corrections, and inserting into Eq. (3.12) results in

$$\Psi(p) \sim 1 - \frac{8}{\ln 2}(p - p_c)^2, \qquad p \nearrow p_c = \frac{1}{2}, \tag{3.13}$$

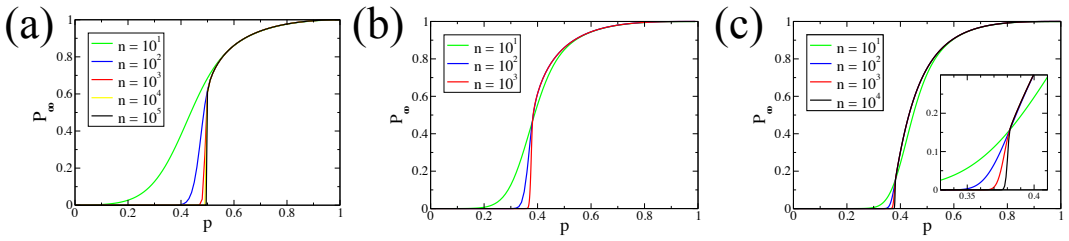

FIGURE 3.5: **Discontinuity in the percolation order parameter $P_\infty(p)$**

(a) MK1, (b) HN5, and (c) HNNP, each for $n = 10^k$ iterations for some integer k. In each case, P_∞ converges slowly to zero just below p_c, and at p_c, P_∞ changes discontinuously. The discontinuity decreases left to right, and is barely visible for HNNP, see inset.

which rapidly approaches unity. This implies that the largest (spanning) cluster that dominates the distribution is nearly extensive already much before the discontinuous transition is reached. RG can only determine the probability T_∞ and average size $\langle s_{\max}\rangle \sim T'_\infty$ of the spanning cluster. Their sub-extensive nature for $p < p_c$ would allow in principle for a diverging number of such clusters. Our simulations show that already for small systems the largest cluster is almost certainly connected to at least one end-site near p_c. (In fact, for MK1 we could have just as well defined $\langle s_{\max}\rangle \sim T'_\infty + \mathcal{S}'_\infty$ to account not just for spanning but all end-site connected clusters, without affecting the scaling.) However, as we will see for HNNP, the non-extensive clusters further below p_c may well be purely internal, with zero probability of spanning between any end-sites.

In light of the discussion regarding universal behavior in hyperbolic networks [121, 140], it is interesting to also explore the scaling behavior of the order parameter on its approach to the discontinuity from above the transition. Numerically, with the RG, we find that a fit to

$$P_\infty(p) \sim P_\infty(p_c) + A\,(p - p_c)^\beta \qquad (p \searrow p_c) \tag{3.14}$$

is quite consistent with a simple, linear approach, i.e., $\beta = 1$, see Fig. 3.6(a).

3.4 Cluster-size scaling for hanoi networks

In the following, we will apply the formalism from Sec. 3.3 to the Hanoi networks HN5 and HNNP in Fig. 3.1(b-c). Their phase diagram, as shown in Fig. 3.4(b-c), has already been discussed in Ref. [112]. To obtain their average cluster size requires the automated algorithm developed in the Appendix, due to the substantial combinatorial effort to

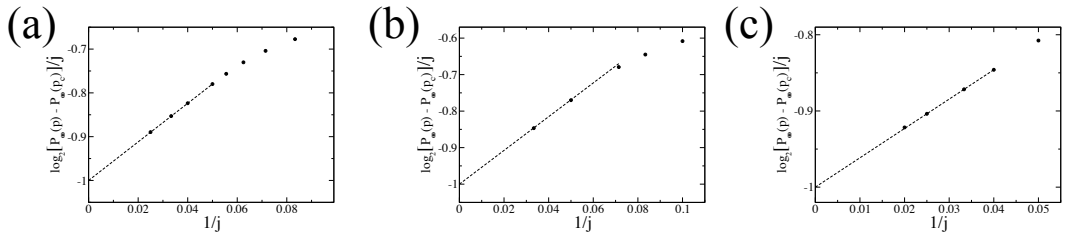

FIGURE 3.6: **Scaling of the order parameter $P_\infty(p)$**

$P_\infty(p)$ for $p \searrow p_c$ according to Eq. (3.14) for (a) MK1, (b) HN5, and (c) HNNP. In each case, taking $p - p_c = \frac{1}{2^j}$, we plot $\log_2 \left[P_\infty(p) - P_\infty(p_c)\right]/j$ vs. $1/j$ which linearly extrapolates to $\beta \sim 1$ as the intercept at $j \to \infty$, i.e., $p \to p_c$.

enumerate their conformations. We will focus here on the more interesting case of HNNP first and then merely report equivalent results for HN5, without the details.

Despite of the added complexity, we find remarkably similar results near the transition for these networks, as compared to MK1, and only some distinctly interesting features for HNNP in the "patchy" regime below p_c. Such robust behavior suggests universal features [121, 140], which can be traced back to the fundamental phase diagram shared by all three networks, as is evident from Fig. 3.4. For comparison, this bond-percolation behavior is not shared by another hierarchical network, MK2, which mutatis mutandis has quite a distinct phase diagram [112, 142], leading instead to a BKT transition. See Ref. [140] for an interpolation between both cases.

In the Appendix, Sec. A.1.2, we show how to obtain the RG-recursions for the cluster generating functions. While otherwise similar to the discussion in Sec. 3.3.1, HNNP (as well as HN5) requires four such functions to account for all possibilities, of having clusters linking any combination of three end-sites or remain isolated, even after accounting for all symmetries of the network. The resulting recursions, Eqs. (A.10), are similar to those for MK1 in Eqs. (3.5), although rather more involved. In the end, we only care for the dominant cluster, which we label x, and consider each possible contribution from one RG-step to the next while disregarding sub-dominant clusters by setting $y = z = 1$. Note that even clusters that are disconnected from any end-site at one step could significantly contribute at the next via the small-world bonds that are linking graphlets between consecutive RG-steps. In the end, we can identify *ten* distinct observables that form a

closed set of recursions. When combined into a single vector,

$$\begin{aligned}\vec{V}_n(x) \;=\; & [R_n(x), S_n(x,x), S_n(x,1), U_n(x,x), \\ & U_n(x,1), N_n(x,x,x), N_n(x,x,1), \\ & N_n(x,1,x), N_n(x,1,1), N_n(1,x,1)]\,,\end{aligned} \tag{3.15}$$

these satisfy the equivalent recursion in (3.6), with the nonlinear RG-flow given by Eqs. (A.10).

To zeroth order, at $x = 1$, Eq. (3.6) gives the recursion relation for percolation of the HNNP graph as derived in Ref. [112]. The coupled recursion relations in (R_n, S_n, U_n, N_n) result in the roots of a sextic polynomial, which can be solved numerically to get the probability of, say, the spanning cluster R_∞ between the end-sites. Fig. 3.4(c) gives the phase diagram for HNNP representing the solutions of the sextic equation, which correspond to the probability R_∞ for $0 < p < 1$. HNNP provides a unique example of a network in which the probability of the dominant cluster to touch *any* end-site vanish below some finite value $0 < p_l < p_c$. In Ref. [112] this was interpreted as a second, lower, critical point, where below p_l neither a spanning nor an extensive cluster exists while between p_l and p_c at least a spanning cluster exists that does not need to be extensive, due to the hyperbolic structure of the network. That spanning cluster becomes extensive only above p_c, the true critical percolation point with non-zero order parameter, $P_\infty > 0$. However, as was shown in Ref. [114], even below the non-zero p_l in HNNP a diverging cluster remains and $\Psi(p)$ defined in Eq. (3.1) remains positive for all $p > 0$. At p_l, $\Psi(p)$ merely jumps discontinuously to a lower but finite value, yet, diverging clusters that connect end-sites are almost certainly absent. Any diverging cluster is fully contained inside HNNP.

The nature of the largest cluster can be studied by looking at the first-order term in the Taylor expansion, Eq. 3.7, of the vector $\vec{V}_n(x)$ in Eq. 3.15. For HNNP the Jacobian $\frac{\partial \vec{F}}{\partial \vec{V}}\left(\vec{V}_n\right)$ at $x = 1$ consists now of a 10×10 matrix and the inhomogeneity is a 10×1 matrix. For large system sizes $(n \to \infty)$ at $x = 1$, it can be shown that the inhomogeneity is subdominant, leaving a homogeneous equations. As before, the largest eigenvalue of the Jacobian gives the scaling exponent $\Psi(p)$ for the largest cluster in the network from Eq. (3.12), as shown in Fig. 3.7. It shows that $\Psi(p) < 1$ for $p_l < p < p_c$, but $\Psi(p)$ drops to zero discontinuously at p_l and vanishes for $p < p_l = 0.31945\ldots$, since the cluster measured by the RG is conditioned on being rooted at an end-site. The RG misses

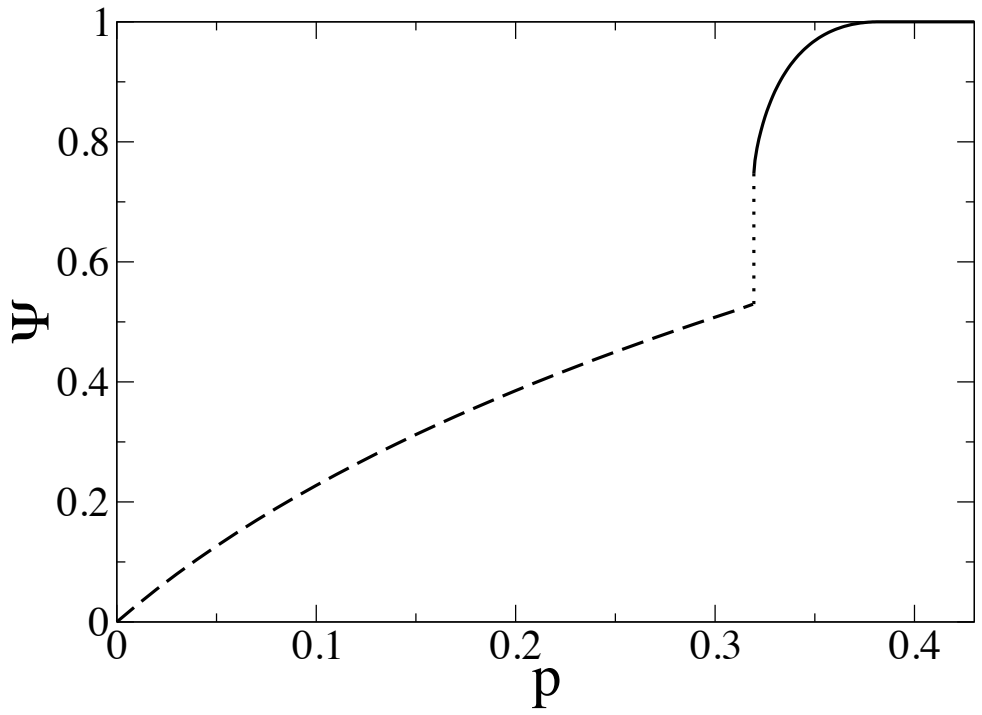

FIGURE 3.7: **Plot of the fractal exponent $\Psi(p)$ for HNNP**

The behavior of $\Psi(p)$ for $p_l < p < p_c = 0.38197\ldots$ (full line) is obtained by exact evaluation of the Jacobian matrix, which develops a branch-point singularity at $p_l = 0.31945\ldots$. Ref. [114] has provided a lower bound, $\Psi(p) = \log_2\left(1+\sqrt{1+8p}\right) - 1$ for $p < p_l$ (dashed line), suggesting a discontinuity in the scaling of the largest cluster at p_l (dotted line) when spanning clusters emerge.

diverging clusters that that do not span the network which apparently dominate below p_l [140]. In any case, since $\Psi(p) < 1$, Eq. (3.11) ensures that $P_\infty \equiv 0$ for all $0 \leq p < p_c$.

Near $p_c = 2 - \phi$, where $\phi = \left(\sqrt{5}+1\right)/2$ is the "golden section", we again find a percolation transition with a discontinuous jump in the order parameter P_∞. By evolving the recursion equations (3.7) for V'_n , the order parameter can be rigorously shown to have monotone convergence to non-zero values *at* and above p_c, see Fig. 3.5(c). For $p \nearrow p_c$, the way $\Psi(p)$ approaches unity can be found through considering the secular equation

$$0 = \det\left\{V'_\infty - \left(2 - a_1\epsilon + a_2\epsilon^2 + \ldots\right) \times \mathbf{I}\right\}, \tag{3.16}$$

expanded in terms of $\epsilon = p_c - p \ll 1$, where $\mathbf{I}$ is the identity matrix. Note that at p_c, the largest eigenvalue of V'_∞ is $\lambda = 2$, around which we expand. Since the percolation

probabilities at p_c are given by $R_\infty = 1, S_\infty = U_\infty = N_\infty = 0$, we assume an expansion of the percolation probabilities as $R_\infty = 1-\rho_1\epsilon+\rho_2\epsilon^2$, $S_\infty = \sigma_1\epsilon+\sigma_2\epsilon^2$, $U_\infty = \nu_1\epsilon+\nu_2\epsilon^2$, and $N_\infty = \eta_1\epsilon + \eta_2\epsilon^2$. To satisfy Eq. (3.16), each coefficient in powers of ϵ should be zero. As a result, we find that linear corrections to the eigenvalue λ vanish, i.e., $a_1 = 0$. Using conservation of probability, $\rho_i+\sigma_i+\upsilon_i+\eta_i = 0$, for each $i \geq 1$ at $p = p_c$, we find a non-vanishing quadratic correction, $a_2 = a_2(\rho_1, \sigma_1, \nu_1, \eta_1) = -\frac{5}{16}\left(38 + 17\sqrt{5}\right)$, for which the second-order corrections in the percolation probabilities proved irrelevant. Hence, Eq. (3.12) yields

$$\Psi_{\text{HNNP}}(p) \sim 1 - \frac{5\left(38 + 17\sqrt{5}\right)}{32\log_e(2)}\left(p_c - p\right)^2 + \ldots, \quad p \nearrow p_c. \tag{3.17}$$

For HN5, by using the same cluster generating functions as for HNNP in the Appendix, we obtain their RG recursions in (A.14). Again, the resulting equations for the cluster size are too complicated to express or solve in closed form. But it is easy to evaluate their phase diagram in Fig. (3.4)(b) for R_∞, as well as the order parameter P_∞ in Fig. (3.5)(b) to any desired accuracy. Here, the same local analysis near p_c as for HNNP yields for HN5:

$$\Psi_{\text{HN5}}(p) \sim 1 - \frac{5\left(677 + 304\sqrt{5}\right)}{484\log_e(2)}\left(p_c - p\right)^2 + \ldots, \quad p \nearrow p_c. \tag{3.18}$$

As for MK1 and HNNP, almost extensive clusters in HN5 emerge well before the transition, with $\Psi(p)$ varying quadratically. It suggests that the quadratic dependence below p_c might be universal for hierarchical networks with discontinuous percolation transitions. Above p_c, the scaling of P_∞ in Eq. (3.14) for both, HN5 and HNNP, also provides $\beta \sim 1$, as shown in Fig. 3.6(b-c).

3.5 Discussion

In this chapter we have shown that a hierarchical network of one-dimensional lattice with small-world type bonds can result in an explosive percolation transition. This discontinuous transition found in the hierarchical lattices are unique, as alternative models based on correlated bond additions have been proven to fail [134]. Our investigation of properties of the cluster formation near the discontinuous percolation transition in hyperbolic networks affirms the robustness of the observed finite-size scaling of the largest cluster in the system. Although the precise conditions for such discontinuous behavior on these

hierarchical networks are not clear, we observe that such behavior is quite general for networks where small-world bonds convert an initially finitely-ramified network into an infinitely ramified network to provide $p_c < 1$. The resulting"explosive transition" indicates that large scale communication is possible on such networks even at small value of bond probability p. This suggests that large scale emergent phenomenon observed in the brain [48], like synchronization [130], avalanches [40, 43] and sustained activity [131] could be a result of the small-world hierarchical structure in brain connectivity [95, 97]. It would be interesting to investigate neural networks similar to the ones discussed in this chapter with firing neurons at each node and recursively generated synaptic connections. An RG approach, similar to one explained here, can then be utilized to understand large scale emergent properties on such model-neural networks.

Chapter 4

Accurate sensing of multiple ligands with a single receptor

4.1 Introduction

Cells obtain information about their environment by capturing ligand molecules with receptors on their surface and estimating the ligand concentration from the receptor activity. Limits on the accuracy of such estimation have been a subject of interest since the seminal work of Berg and Purcell [143], with several substantial extensions found recently [51–57]. All of these assume one ligand species coupled to one receptor species. However, cells carry many types of receptors and have many species of ligands around them. The same ligands can bind to many receptors, albeit with different affinities, and vice versa. This is commonly referred to as *cross-talk*, and represents a complex network of ligand-receptor interactions. Here we study signal processing on a simple ligand-receptor network with cross-talk.

In traditional deterministic chemical kinetics, one cannot estimate concentrations of more ligands than there are receptor types. Further, even a weak cross-talk prevents determination of concentrations of individual chemical species since activity of a receptor is a function of a weighted sum of concentrations of all ligands that can bind to it. In contrast, here we argue that, with cross-talk, concentration of more than one chemical species can be inferred from the activity of one receptor, provided that the entire stochastic temporal sequence of receptor binding and unbinding events is accessible instead of its mean occupancy. This surprising result can be understood by noting that

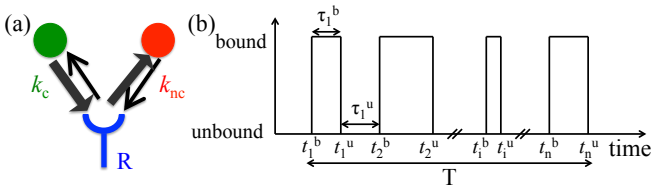

FIGURE 4.1: **Model of ligand-receptor cross-talk**

(a). Two lignads, cognate and non-cognate, bind to a receptor R with binding rates k_c and k_{nc}, respectively. The cognate unbinding rate is defined as lower than the non-cognate one ($r_c < r_{nc}$). (b) Time series of receptor occupancy is used to determine both on-rates.

a typical duration of time that a ligand remains bound to the receptors depends on its unbinding rate. Thus observing the statistics of the receptor's unbound time durations allows estimation of a weighted average of all chemical species that interact with it [54], and then observing the statistics of the bound time durations allows to tell how common each ligand is.

Here, we derive these results for the simplest problem of the class, namely one receptor interacting with two ligand species. While the exact solution of the inference problem for finding both ligand concentrations is hard to implement using common biochemical machinery, we show that an accurate approximation is possible using the familiar kinetic proofreading mechanism [144, 145].

4.2 Model of ligand-receptor cross-talk

Consider a single receptor estimating concentrations of a cognate and a non-cognate ligand, Fig. 4.1. The ligands bind to the receptor with on-rates k_c and k_{nc}. These are proportional to the ligand concentrations with known coefficients of proportionality. Thus estimating $k_{c,nc}$ is equivalent to estimating the concentrations themselves. The unbinding, or off-rates, r_c and r_{nc}, distinguish the two ligands: $r_{nc} > r_c$, and a cognate molecule typically stays bound for longer. Following Ref. [54], we estimate k_c and k_{nc} from the time-series of binding, $\{t_i^b\}$, and unbinding, $\{t_i^u\}$ events of a total duration T using Maximum Likelihood techniques. The numbers of binding and unbinding events are different by, at most, one, which is insignificant since we consider $T \to \infty$. Thus without loss of generality, we assume that the first event was a binding event at t_1^b, and the last one was the unbinding at t_n^u. We write the probability distribution of observing the sequence $\{t_1^b, t_1^u, \ldots, t_n^b, t_n^u\}$, or alternatively the sequence of binding and unbinding

intervals $\tau_i^{\mathrm{b}} = t_i^{\mathrm{u}} - t_i^{\mathrm{b}}$, and $\tau_i^{\mathrm{u}} = t_{i+1}^{\mathrm{b}} - t_i^{\mathrm{u}}$:

$$P \equiv P(\{\tau_i^{\mathrm{b}},\ \tau_i^{\mathrm{u}}\}|k_{\mathrm{c}}, k_{\mathrm{nc}}) = \frac{1}{Z}\prod_{i=1}^{n}\left[e^{-\tau_i^{\mathrm{u}}(k_{\mathrm{c}}+k_{\mathrm{nc}})}\left(k_{\mathrm{c}}\ r_{\mathrm{c}}\ e^{-\tau_i^{\mathrm{b}} r_{\mathrm{c}}} + k_{\mathrm{nc}}\ r_{\mathrm{nc}}\ e^{-\tau_i^{\mathrm{b}} r_{\mathrm{nc}}}\right)\right]. \quad (4.1)$$

Here the first term under the product sign is the probability of the receptor staying unbound for τ_i^{u}. The second term, which we from now on denote by $D(k_{\mathrm{c}}, k_{\mathrm{nc}}, \tau_i^{\mathrm{b}})$, is proportional to the probability of staying bound for τ_i^{b}, which has contributions from being bound to the cognate and the noncognate ligands, with odds of $k_{\mathrm{c}}/k_{\mathrm{nc}}$. Finally, Z is the normalization. Note that here we define $\tau_n^{\mathrm{u}} = t_1^{\mathrm{b}} + (T - t_n^{\mathrm{u}})$, so that the n'th unbound interval includes the "incomplete" unbound intervals before the first binding and after the last unbinding.

4.3 Maximum Likelihood estimate of concentrations

The log-likelihood of $k_{\mathrm{c,nc}}$ is the logarithm of P, Eq. (4.1). Taking the derivatives of the log-likelihood w. r. t. k_{c} and k_{nc} and setting them to zero gives the Maximum Likelihood (ML) equations for the two concentrations. This approach is similar to minimizing the action to get the classical path in Lagrangian mechanics. Denoting by $T^{\mathrm{u}} = \sum_{i=1}^{n} \tau_i^{\mathrm{u}}$ the total time the receptor is unbound, these are

$$-T^{\mathrm{u}} + \sum_{i=1}^{n} \frac{r_{\mathrm{c}} e^{-\tau_i^{\mathrm{b}} r_{\mathrm{c}}}}{D(k_{\mathrm{c}}^*, k_{\mathrm{nc}}^*, \tau_i^{\mathrm{b}})} = 0, \quad (4.2)$$

$$-T^{\mathrm{u}} + \sum_{i=1}^{n} \frac{r_{\mathrm{nc}} e^{-\tau_i^{\mathrm{b}} r_{\mathrm{nc}}}}{D(k_{\mathrm{c}}^*, k_{\mathrm{nc}}^*, \tau_i^{\mathrm{b}})} = 0, \quad (4.3)$$

where $*$ denotes the ML solution. Multiplying Eqs. (4.2, 4.3) by k_{c}^* and k_{nc}^*, respectively, and adding them gives

$$k_{\mathrm{c}}^* + k_{\mathrm{nc}}^* = \frac{n}{T^{\mathrm{u}}}, \quad (4.4)$$

which determines the sum of the two concentrations, showing that the estimates are negatively correlated. As in Ref. [54], the total on-rate (the weighted average of the external concentrations) is determined only by the average duration of the unbound interval, $(n/T^{\mathrm{u}})^{-1}$, because no binding is possible when the receptor is already bound.

In general, the ML equations cannot be solved analytically, requiring numerical approaches. However, as all ML estimators, they are unbiased to the leading order in n. The standard errors of the ML estimates can be obtained by inverting the Hessian

matrix,

$$\left.\frac{\partial^2 \log P}{\partial k_\cdot \partial k_\cdot}\right|_{k_\mathrm{c}^*, k_\mathrm{nc}^*} = \sum_{i=1}^{n} \left[\frac{-1}{D(k_\mathrm{c}, k_\mathrm{nc}, \tau_i^\mathrm{b})^2} \begin{pmatrix} r_\mathrm{c}^2 e^{-2\tau_i^\mathrm{b} r_\mathrm{c}} & r_\mathrm{c} r_\mathrm{nc} e^{-\tau_i^\mathrm{b}(r_\mathrm{c}+r_\mathrm{nc})} \\ r_\mathrm{c} r_\mathrm{nc} e^{-\tau_i^\mathrm{b}(r_\mathrm{c}+r_\mathrm{nc})} & r_\mathrm{nc}^2 e^{-2\tau_i^\mathrm{b} r_\mathrm{nc}} \end{pmatrix} \right], \quad (4.5)$$

where $\cdot$ stands for $\{\mathrm{c}, \mathrm{nc}\}$. The inverse of $\frac{\partial^2 \log P}{\partial k_\cdot \partial k_\cdot}$, which scales as $\propto 1/n$, sets the minimum variance of any unbiased estimator according to the Cramer-Rao bound. It has straightforward analytical approximations in various regimes. For example, for $k_\mathrm{c}/k_\mathrm{nc} \gg 1$ and $r_\mathrm{c}/r_\mathrm{nc} \ll 1$, when the noncognate ligand is almost absent, and its few molecules do not bind for long, one gets $\sigma^2(k_\mathrm{c}^*) \approx \left(\partial^2 \log P/\partial k_\mathrm{c}^2\right)^{-1}_{k_\mathrm{c}=k_\mathrm{c}^*} \approx 1/n$, matching the accuracy of sensing one ligand with one receptor [54]. A regime relevant for detection of a rare, but highly specific ligand [146, 147]) can be investigated as well. Instead, we focus on how the receptor estimates (rather than detects) concentrations of *both* ligands simultaneously, which requires us to investigate the full range of on-rates.

To study the variability of the ML estimator, we define its error as $E_\mathrm{c,nc} = n\sigma^2(k_\mathrm{c,nc}^*)/k_\mathrm{c,nc}^2$, the squared coefficient of variation, multiplied by n, which has a finite limit at $n \to \infty$. $E = 1$ corresponds to the accuracy that a receptor measuring a single ligand would obtain [54]. We show $\log_{10} E$ for different on- and off-rates in Fig. (4.2). If the two ligands are readily distinguishable, $r_\mathrm{c} \ll r_\mathrm{nc}$, then the ligand with the dominant k has $E \sim 1$. When $k_\mathrm{c} \sim k_\mathrm{nc}$, $E_\cdot \sim 4 \ldots 5$, and it grows to $10 \ldots 30$ for a ligand with a very small relative on-rate. Emphasizing the importance of the time scale separation, $E > 100$ if the ligands are hard to distinguish, $r_\mathrm{c} \sim r_\mathrm{nc}$. Here, in addition, the correlation coefficient ρ of the two estimates reaches -1 because the same binding event can be attributed to either ligand. Finally, the asymmetry of the plots w. r. t. the exchange of k_c and k_nc is because the cognate ligand can generate short binding events, while long events from the noncognate ligand are exponentially unlikely. In summary, it is possible to infer two ligand concentrations from one receptor, with the error of only $1 \ldots 10$ times larger than for ligand-receptor pairs with no cross talk, as long as the two off-rates are substantially different.

4.4 Approximate solution

Solving Eqs. (4.2, 4.3) to find the ML on-rates would be hard for the cell. Luckily, an approximate solution exists. To find it, we notice that most of the long binding events come from the cognate ligand since the noncognate one dissociates faster. Defining long

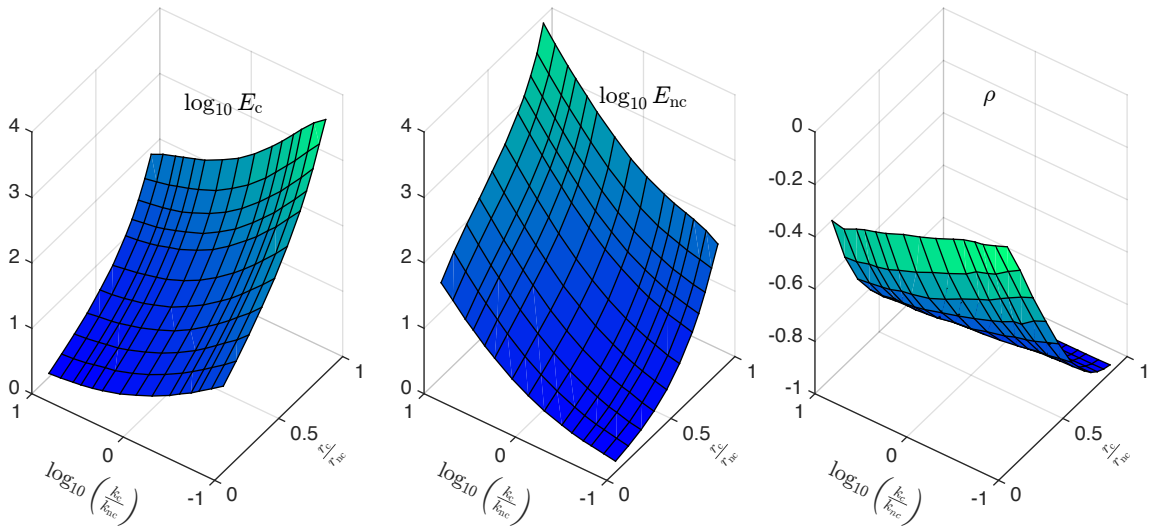

FIGURE 4.2: **Variability of the ML estimators**

Variability represented by $\log_{10} E_c$ (left), $\log_{10} E_{nc}$ (center), and the correlation coefficient ρ between k_c^* and k_{nc}^* (right) as functions of $k.$ and $r.$. Here we use $r_{nc} = k_c + k_{nc} = 1$. The plotted quantities are estimated as averages over $30,000$ randomly generated binding/unbinding sequences for each combination of the rates. Each sequence consists of $n = 30,000$ binding events, simulated using the Gillespie algorithm. Standard errors are too small to be represented.

events as $\tau_i^b \geq T^c$, we rewrite Eqs. (4.2, 4.4) as

$$\frac{n}{k_c^* + k_{nc}^*} = \left(\sum_{\tau_i^b \geq T^c} + \sum_{\tau_i^b < T^c} \right) \frac{r_c e^{-\tau_i^b r_c}}{D(k_c^*, k_{nc}^*, \tau_i^b)} \tag{4.6}$$

Assuming that almost all long events are cognate, $T^c \gg 1/r_{nc}$, this gives

$$\frac{n}{k_c^a + k_{nc}^a} = \frac{n_l}{k_c^a} + \sum_{\tau_i^b < T^c} \frac{r_c e^{-\tau_i^b r_c}}{D(k_c^a, k_{nc}^a, \tau_i^b)}, \tag{4.7}$$

where n_l is the number of long events, and the superscript "a" stands for the *a*pproximate solution. If further T is long enough so that there are many short events, and a single binding duration hardly affects k_c^*, then the sum in Eq. (4.7) can be approximated by the expectation value:

$$\frac{n}{k_c^a + k_{nc}^a} = \frac{n_l}{k_c^a} + (n - n_l) \int_0^{T_c} \frac{r_c e^{-\tau^b r_c} P(\tau^b | k_c^a, k_{nc}^a) d\tau^b}{D(k_c^a, k_{nc}^a, \tau^b)}, \tag{4.8}$$

where $P(\tau^{\mathrm{b}}|k_{\mathrm{c}}^{\mathrm{a}}, k_{\mathrm{nc}}^{\mathrm{a}})$ is the probability of observing a binding event of the duration τ^{b} for the given binding rates,

$$P(\tau^{\mathrm{b}}|k_{\mathrm{c}}^{\mathrm{a}}, k_{\mathrm{nc}}^{\mathrm{a}}) = \frac{D(k_{\mathrm{c}}^{\mathrm{a}}, k_{\mathrm{nc}}^{\mathrm{a}}, \tau^{\mathrm{b}})}{k_{\mathrm{c}}^{\mathrm{a}} + k_{\mathrm{nc}}^{\mathrm{a}}}. \tag{4.9}$$

Plugging Eq. (4.9) into Eq. (4.8), we obtain

$$\frac{1}{k_{\mathrm{c}}^{\mathrm{a}} + k_{\mathrm{nc}}^{\mathrm{a}}} = \frac{n_{\mathrm{l}}}{n k_{\mathrm{c}}^{\mathrm{a}}} + \left(1 - \frac{n_{\mathrm{l}}}{n}\right) \frac{1 - e^{-r_{\mathrm{c}} T^{\mathrm{c}}}}{k_{\mathrm{c}}^{\mathrm{a}} + k_{\mathrm{nc}}^{\mathrm{a}}}. \tag{4.10}$$

Finally, since $n_{\mathrm{l}} \ll n$, using Eq. (4.4), we get:

$$k_{\mathrm{c}}^{\mathrm{a}} = \frac{n_{\mathrm{l}}}{T^{\mathrm{u}}} e^{r_{\mathrm{c}}\, T^{\mathrm{c}}}, \tag{4.11}$$

$$k_{\mathrm{nc}}^{\mathrm{a}} = \frac{n}{T^{\mathrm{u}}} - \frac{n_{\mathrm{l}}}{T^{\mathrm{u}}} e^{r_{\mathrm{c}}\, T^{\mathrm{c}}}. \tag{4.12}$$

In other words, the approximate cognate ligand concentration is proportional to the number of long events.

We can estimate the bias and the variance of $k_{\mathrm{c,nc}}^{\mathrm{a}}$ in a limiting case. If r_{c} and r_{nc} are not very different from each other, then T^{c} must be much larger than the inverse of either of them, $T^{\mathrm{c}} \gg \{r_{\mathrm{nc}}^{-1}, r_{\mathrm{c}}^{-1}\}$, and $n_{\mathrm{l}} \ll n$. Then most of the variance of $k_{\mathrm{c,nc}}^{\mathrm{a}}$ in Eqs. (4.11, 4.12) comes from variability of n_{l}, but not T^{u}. Thus we write $\langle k_{\mathrm{c}}^{\mathrm{a}} \rangle \approx \frac{\langle n_{\mathrm{l}} \rangle}{\langle T^{\mathrm{u}} \rangle} e^{r_{\mathrm{c}} T^{\mathrm{c}}}$. Further, the individual unbound periods are independent, so that $\langle T^{\mathrm{u}} \rangle = n \langle \tau^{\mathrm{u}} \rangle = n/(k_{\mathrm{c}} + k_{\mathrm{nc}})$ (notice the use of k rather than k^{a} in this relation). Further, $\langle n_{\mathrm{l}} \rangle = n\, P(\tau^{\mathrm{b}} > T^{\mathrm{c}}) = \frac{n}{k_{\mathrm{c}} + k_{\mathrm{nc}}} \left(k_{\mathrm{c}} e^{-r_{\mathrm{c}} T^{\mathrm{c}}} + k_{\mathrm{nc}} e^{-r_{\mathrm{nc}} T^{\mathrm{c}}}\right)$. Combining these expressions, we get

$$\langle k_{\mathrm{c}}^{\mathrm{a}} \rangle \approx k_{\mathrm{c}} + k_{\mathrm{nc}} e^{-(r_{\mathrm{nc}} - r_{\mathrm{c}}) T^{\mathrm{c}}}. \tag{4.13}$$

Thus for large T^{c}, the bias of the approximate estimator, $k_{\mathrm{nc}} e^{-(r_{\mathrm{nc}} - r_{\mathrm{c}}) T^{\mathrm{c}}}$, grows with the relative number of noncognate long bindings events. In turn, the latter is proportional to k_{nc}, but decreases exponentially with T^{c}.

Within the same approximation, the variance of the estimator is $\sigma^2(k_{\mathrm{c}}^{\mathrm{a}}) \approx \frac{\sigma^2(n_{\mathrm{l}})}{\langle T^{\mathrm{u}} \rangle^2} e^{2 r_{\mathrm{c}} T^{\mathrm{c}}}$. But long binding events are rare, independent of each other, and hence obey the Poisson statistics. Thus $\sigma^2(n_{\mathrm{l}}) = \langle n_{\mathrm{l}} \rangle$, so that

$$\sigma^2(k_{\mathrm{c}}^{\mathrm{a}}) \approx \langle k_{\mathrm{c}}^{\mathrm{a}} \rangle \frac{k_{\mathrm{c}} + k_{\mathrm{nc}}}{n} e^{r_{\mathrm{c}} T^{\mathrm{c}}}. \tag{4.14}$$

The variance obviously grows with T^{c}.

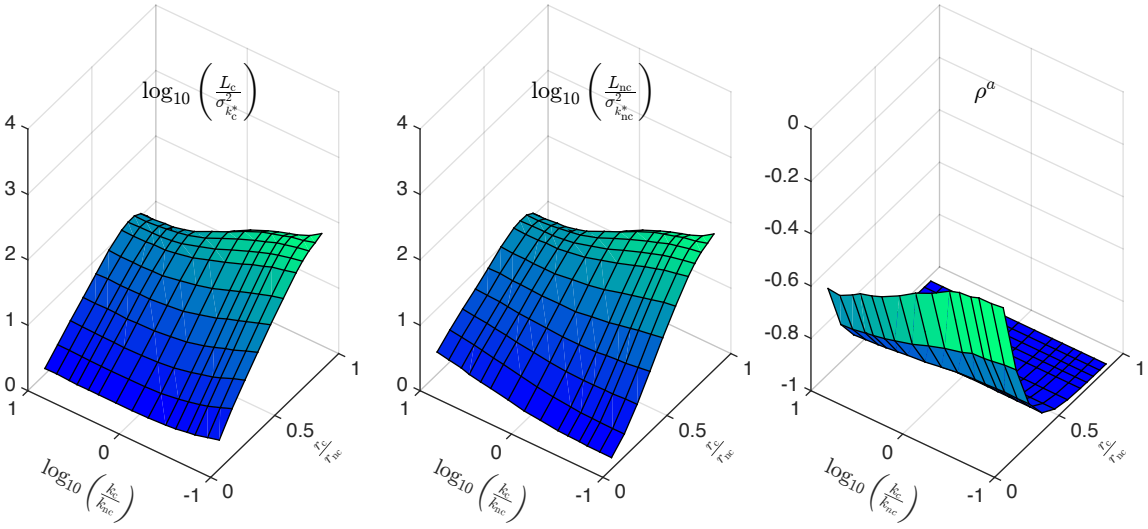

FIGURE 4.3: **Comparison of errors of the approximate and the ML solutions**

We plot $\log_{10}(L_{\mathrm{c}}(T_0)/\sigma^2_{k^*_{\mathrm{c}}})$ (left), $\log_{10}(L_{\mathrm{nc}}(T_0)/\sigma^2_{k^*_{\mathrm{nc}}})$ (center) and the covariance of the approximate estimates (right) as functions of on- and off-rates. Simulations are performed in the same way as in Fig. 4.2.

Knowing that the bias and the variance of the approximation change in opposite directions with T^{c}, we can find the optimal cutoff by minimizing the overall error, or, in other words, solving the bias-variance tradeoff:

$$T^{\mathrm{c}}_* = \arg\min_{T^{\mathrm{c}}} L = \arg\min_{T^{\mathrm{c}}} \left[(k_{\mathrm{c}} - \langle k^{\mathrm{a}}_{\mathrm{c}} \rangle)^2 + \sigma^2(k^{\mathrm{a}}_{\mathrm{c}}) \right], \tag{4.15}$$

where L is the sum of the squared bias and the variance of the estimator. Near the optimal cutoff, the bias is small, and we use k_{c} instead of $k^{\mathrm{a}}_{\mathrm{c}}$ for the variance of the estimator, Eq. (4.14). Then solving Eq. (4.15) gives:

$$T^{\mathrm{c}}_* = \frac{1}{(2r_{\mathrm{nc}} - r_{\mathrm{c}})} \log \left[2T^{\mathrm{u}} \left(\frac{r_{\mathrm{nc}}}{r_{\mathrm{c}}} - 1 \right) \frac{k^2_{\mathrm{nc}}}{k_{\mathrm{c}}} \right]. \tag{4.16}$$

Plugging this into Eqs. (4.13, 4.14), we can get the minimal error of the estimator, which we omit here for brevity.

The optimal cutoff is $\propto 1/r_{\mathrm{nc}}$ if $r_{\mathrm{nc}} \gg r_{\mathrm{c}}$, and it grows with r_{c}, allowing for better disambiguation of cognate and noncognate events. Crucially, the off-rates are specified with the ligand identities. In contrast, the on-rates, $k_{\mathrm{c,nc}}$, are what the receptors measures. Therefore, it is encouraging that T^{c} depends only logarithmically on the on-rates (and also on the duration of the measurement, T^{u}): fixing T^{c} as T^{c}_* at some fixed values of $k_{\mathrm{c,nc}}$ remains near-optimal for a broad range of on-rates. To illustrate this, we use

$T^{\mathrm{c}} = T^{\mathrm{c}}_*(k_{\mathrm{c}} = k_{\mathrm{nc}} = 1/2) \equiv T_0$ and analyze the quality of the approximation in Fig. 4.3, where we plot the ratio $L_{\mathrm{c,nc}}(T_0)/\sigma^2_{k_{\mathrm{c,nc}}}$. Since the ratio approaches 1 when $r_{\mathrm{c}}/r_{\mathrm{nc}} \to 0$ (specifically, for $r_{\mathrm{c}}/r_{\mathrm{nc}} = 0.1$, $L_{\mathrm{c}}(T_0)/\sigma^2_{k_{\mathrm{c}}} \approx 1.47$, and $L_{\mathrm{nc}}(T_0)/\sigma^2_{k_{\mathrm{nc}}} \approx 1.21$), we conclude that the approximation is accurate even at fixed $T^{\mathrm{c}} = T_0$ when its assumptions are satisfied. In contrast, when the ligands are nearly indistinguishable, $L_{\mathrm{c,nc}}(T_0)/\sigma^2_{k_{\mathrm{c,nc}}} \sim 100$, but here one would not use one receptor to estimate two concentrations since even the ML solution is bad (cf. Fig. 4.2). Note also that both L_{c} and L_{nc} are smaller for $r_{\mathrm{c}} \sim r_{\mathrm{nc}}$ if $k_{\mathrm{c}} \gg k_{\mathrm{nc}}$. This is because our main assumption (that almost all long events are cognate) holds better when cognate ligands dominate. Finally, the correlation coefficient between the approximate estimates, ρ^{a} (right panel) reaches -1 earlier than in Fig. 4.2. This is a direct consequence of Eqs. (4.11, 4.12).

4.5 Kinetic Proofreading Mechanism for approximate estimation

The approximate solution can be computed by cells using the well-known kinetic proofreading (KPR) mechanism [144, 145, 148, 149]. In the simplest model of KPR [150], intermediate states between an inactive and an active state of a receptor delay the activation. Thus bound ligands can dissociate before the receptor activates, at which point it quickly reverts to the inactive state. Since $r_{\mathrm{c}} > r_{\mathrm{nc}}$, cognate ligands dominate among bindings that actually lead to activation. The resulting increase in specificity in various KPR schemes has led to their exploration in the context of *detection* of rare ligands [146, 147, 149], and here we extend them to *measurement* of concentration of cognate and noncognate ligands simultaneously.

Consider a biochemical network in Fig. 4.4: the receptor (R) activates two messenger molecules (A) and (B). The first one is activated with the rate k_{A} whenever the receptor is bound. The second one is activated only if the receptor stays bound for longer than a certain T^{c} (with the delay achieved using the KPR intermediate states). The activation rate after the delay is k_{B}. The molecules deactivate with the rates r_{A} and r_{B}, respectively, and all activations/deactivations are first-order reactions. Then the mean concentrations

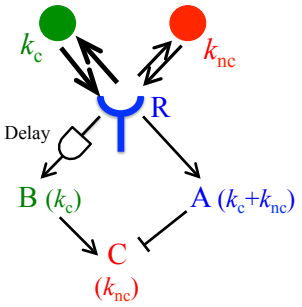

FIGURE 4.4: **Kinetic Proofreading for estimating multiple concentrations.**

Molecules A and B are produced when the receptor is bound, but A is produced only for long bindings. Another chemical species C subtracts A from B, so that A approximates k_c and C approxiates k_{nc}.

of the messenger molecules are:

$$\bar{A} = \frac{k_c/r_c + k_{nc}/r_{nc}}{1 + k_c/r_c + k_{nc}/r_{nc}} \frac{k_A}{r_A} \tag{4.17}$$

$$\bar{B} = \frac{k_c/r_c e^{-r_c T^c} + k_{nc}/r_{nc} e^{-r_{nc} T^c}}{1 + k_c/r_c + k_{nc}/r_{nc}} \frac{k_B}{r_B}, \tag{4.18}$$

Assuming again that most bindings longer than T^c are cognate, we solve Eqs. (4.17, 4.18) for the on-rates

$$k_c = \frac{\bar{B} e^{r_c T^c} r_c r_B}{k_B} \left(1 + \frac{\bar{A}}{k_A/r_A - \bar{A}}\right), \tag{4.19}$$

$$k_{nc} = \left[\frac{\bar{A}}{k_A/r_A - \bar{A}} - \frac{\bar{B} e^{r_c T^c} r_B}{k_B} \left(1 + \frac{\bar{A}}{k_A/r_A - \bar{A}}\right)\right] r_{nc}. \tag{4.20}$$

The corrections of the form $\bar{A}/(k_A/r_A - \bar{A})$ appear because bindings only happen to unbound receptors, as emphasized in Ref. [54]. However, these nonlinear relations are still hard to implement with simple biochemical components. We solve this by further assuming $\epsilon = \bar{A}/(k_A/r_A) \ll 1$, which is true if the receptor is mostly unbound (both on-rates are small compared to the respective off-rates). This gives

$$k_c^{KPR} \approx \frac{\bar{B} e^{r_c T^c} r_c r_B}{k_B}, \tag{4.21}$$

$$k_{nc}^{KPR} \approx \left(\frac{r_A \bar{A}}{k_A} - \frac{\bar{B} e^{r_c T^c} r_B}{k_B}\right) r_{nc}. \tag{4.22}$$

These equations are analogous to Eqs. (4.11, 4.12). They are easy to realize biochemically (cf. Fig. 4.4): k_c is related to the concentration of the proofread species B by a rescaling, and k_{nc} comes from subtracting rescaled versions of A and B from each other. The

subtraction can be done by the third species C, activated by A and suppressed by B. Since $\epsilon \ll 1$, then $\bar{A}$ and $\bar{B}$ are small, and many such activation-suppression schemes are linearized as the subtraction [57].

The bias of $k^{\mathrm{a}}_{\mathrm{c,nc}}$ due to long, but noncognate binding events, Eq. (4.13), carries over to $k^{\mathrm{KPR}}_{\mathrm{c,nc}}$. However, there is an additional contribution since the time to traverse the intermediate states is random. Thus T^{c} has some variance $\sigma^2_{T^{\mathrm{c}}}$ [150, 151]. This variability changes the rate of occurence of long biding events, but they are still rare, nearly independent, and Poisson-distributed. Denoting by $\langle\cdot\rangle$ the averaging at a fixed T^{c}, and by $\bar{\cdot}$ the averaging over T^{c}, we get

$$\frac{\overline{\langle n_{\mathrm{l}}\rangle}}{k_{\mathrm{c}}} \approx \frac{n}{k_{\mathrm{c}}+k_{\mathrm{nc}}} e^{-r_{\mathrm{c}}\bar{T}^{\mathrm{c}}+\frac{1}{2}r_{\mathrm{c}}^2\sigma^2_{T^{\mathrm{c}}}}. \tag{4.23}$$

Thus $\sigma^2_{T^{\mathrm{c}}}$ effectively renormalizes the cutoff to $\bar{T}^{\mathrm{c}} - \frac{1}{2}r_{\mathrm{c}}\sigma^2_{T^{\mathrm{c}}}$, which is independent of the on-rates. Replacing T^{c} in Eqs. (4.21, 4.22) by its renormalized value, which is an easy change in the scaling factors, removes this additional bias due to the random T^{c} in the KPR scheme.

Since long bindings are rare, the variance of the KPR estimator is dominated again generally by $\bar{B}$, but not $\bar{A}$. The intrinsic stochasticity in production of molecules of B contributes to the variance. However, this contribution can be made arbitrarily small by increasing k_{B}, and we neglect it here. A larger contribution comes from the random number of long bound intervals and a random duration of each of them. To calculate this, in the limit of rare long binding events, we use well-known results in the theory of noise propagation in chemical networks [152]

$$\frac{\sigma^2_{\mathrm{B}}}{\bar{B}^2} \approx \frac{(1+k_{\mathrm{c}}/r_{\mathrm{c}}+k_{\mathrm{nc}}/r_{\mathrm{nc}})\, e^{r_{\mathrm{c}}T^{\mathrm{c}}-\frac{1}{2}r_{\mathrm{c}}^2\sigma^2_{T^{\mathrm{c}}}}}{k_{\mathrm{c}}(1/r_{\mathrm{c}}+1/r_{\mathrm{B}})} = \frac{e^{r_{\mathrm{c}}T^{\mathrm{c}}-\frac{1}{2}r_{\mathrm{c}}^2\sigma^2_{T^{\mathrm{c}}}}}{k_{\mathrm{c}}(1/r_{\mathrm{c}}+1/r_{\mathrm{B}})} + O(\epsilon). \tag{4.24}$$

This is a direct analog of Eq. (4.14).

4.6 Discussion

The realization of Refs. [54, 153] and others that the detailed temporal sequence of binding and unbinding events carries more information about the ligand concentration than the mean receptor occupancy is a conceptual breakthrough. It parallels the realization in the computational neuroscience community that precise timing of spikes carries more

information about the stimulus than the mean neural firing rate [11, 154–158], and it has a potential to be equally impactful. This extra information when measuring one ligand concentration with one receptor [54] amounted to increasing the sensing accuracy by a constant prefactor, or, equivalently, getting only a finite number of additional bits from even a very long measurement [159]. In contrast, here we show that two concentrations can be measured with one receptor with the variance that decreases inversely proportionally to the number of observations, n, Eq. (4.14), or to the integration time, $1/r_{\mathrm{B}}$, Eq. (4.24), so that the accuracy is only a (small) prefactor lower than would be possible with one receptor per ligand species. Asymptotically, this doubles the information obtained by the receptor [159].

In principle, one can measure more than two concentrations similarly, as long as all species have sufficiently distinct off-rates. While the error (the variance for the ML estimator, and both the bias and the variance for the approximate and the KPR estimators) would grow with a larger number of ligand species, this would still represent a dramatic increase in the information gained by the receptor that keeps track of its precise temporal dynamics, rather than just the average binding state.

Crucially, such improvement would not be possible without the cross-talk, or binding among noncognate ligands and receptors. Normally, the cross-talk is considered a nuisance that must be suppressed [160, 161]. Instead we argue that cross-talk can be beneficial by recruiting more receptor types to measure concentration of the same ligand. In particular, this allows having fewer receptor than ligand species, potentially illuminating how cells function reliably in chemically complex environments with few receptor types. Further, the cross-talk can increase the dynamic range of the entire system: a ligand may saturate its cognate receptor, preventing accurate measurement of its (high) concentration, but it may be in the sensitive range of non-cognate receptors at the same time. Finally, the increased bandwidth may lead to improvements in sensing a time-dependent ligand concentration [146, 153]. We will explore such many-to-many sensory schemes, extending ideas of Ref. [162] to tracking temporal sequences of activation of receptor and to varying environments in forthcoming publications.

While the exact maximum likelihood inference of multiple concentrations from a temporal binding-unbinding sequence is rather complex, we showed that when the cognate and the non-cognate off-rates are substantially different, there is a simpler, approximate, but accurate inference procedure. In various immune system problems, $r_{\mathrm{nc}}/r_{\mathrm{c}} \sim 5$, which would allow the approximation to work. Moreover, when the receptor is not saturated

and spends most of its time unbound, this inference can be performed by biochemical motifs readily available to the cell. Namely, one needs two branches of activation downstream of the receptor, with one of them having a kinetic proofreading (KPR) time delay, and then an estimate of the difference of activities of the branches. This suggests a possible signal estimation role for the KPR scheme in addition to the more traditional signal detection one [146, 147, 153]. Such branching and merging of signaling pathways downstream of a receptor is common in signaling [14, 161, 163, 164]. Thus exploring the function of such complex organization in the context of estimation of multiple signals with cross-talk is in order.

In summary, monitoring precise temporal sequences of receptor activation/deactivation opens up new and exciting possibilities for environment sensing by cells.

Chapter 5

Extrinsic and intrinsic correlations in molecular information transmission

5.1 Introduction

The concentration estimation of external ligand molecules through cell surface receptor has been an active area of research for a long time. Starting with Berg and Purcell [143], a lot of work has been done to estimate how well cell receptors can perform in estimating the ligand concentration both individually[52–56] and collectively [51, 165]. However, little is known about the effects of interactions between receptors on the information that the receptors contain about the ligands. Similar questions about representing signals using activity of multiple response units have been explored in details under the name of "population coding" [17–25, 60]. Transplanting the accumulated knowledge to molecular sensing domain is bound to be fruitful. Here we focus on the simplest problem of the kind and analyze the ability of two identical receptors to measure the concentration of an external ligand to understand the effects of interactions among receptors and the corresponding correlations on the information processing by the receptors. We quantify how useful it is to keep track of which receptor absorbed a ligand molecule, rather than counting the total number of captured ligands. We analyze the nonlinear coupling between the receptors, which emerges because a molecule absorbed on one receptor cannot be absorbed on the other. Where naive expectations suggest a synergy, we show a

reduction of the information gathered about the stimulus compared to the noninteracting receptors case.

5.2 Background

The following simple, and yet instructive model, borrowed from the computational neuroscience literature [17, 60], serves as a starting point for our analysis. Imagine a Gaussian signal s with the mean $\bar{s}$ and the variance σ_s^2. It is read out and represented by two responses, r_1 and r_2 (firing rates of neurons in neuroscience, or receptor activity here). For simplicity, these are assumed linearly and equivalently dependent on s (or the response to small fluctuations is linearized), such that

$$r_1 = as + \eta_1, \tag{5.1}$$

$$r_2 = as + \eta_2, \tag{5.2}$$

where a is the gain, and $\eta_{1,2}$ are Gaussian noises with $\langle \eta_i \rangle = 0$, and $\operatorname{var} \eta_i \equiv \langle \eta_i^2 \rangle = \sigma_\eta^2$.

One can estimate the signal from the two responses as $s_{\text{est}} = (r_1 + r_2)/(2a)$. Then the variance of the estimate is

$$\operatorname{var}(s_{\text{est}} - s) \equiv \sigma_{\text{err}}^2 = \frac{\sigma_\eta^2(1+\rho_\eta)}{2a^2}. \tag{5.3}$$

Here $\rho_\eta \sigma_\eta^2 = \operatorname{cov}(\eta_1, \eta_2)$ stands for the covariance of the two noises, or the *noise-induced* covariance [17], and ρ_η is the corresponding correlation coefficient. By analogy with the intrinsic noise in cell biology literature [166], ρ_η can also be called the *intrinsic noise correlation*. When $\rho_\eta = 0$, Eq. (5.3) reduces to the usual decrease of the error variance by a factor of two for two independent measurements. However, when $\rho_\eta < 0$, the error variance is smaller. In particular, if $\rho_\eta \to -1$, the signal can be estimated with no error. Generalizing this simple observation, one can define the *stimulus-induced response covariance* [17] or the *extrinsic noise covariance* [166], as the covariance between mean responses to stimuli, averaged over all stimuli, $\operatorname{cov}(\bar{r}_1, \bar{r}_2) \equiv \rho_s a^2 \sigma_s^2$. Then our simple example illustrates the well-known *sign rule* [60]: if ρ_s and ρ_η are of opposite signs, then the stimulus can be inferred from the two responses with a smaller error compared to the (conditionally) independent responses, $\rho_\eta = 0$. The same result can also be written in terms of the *mutual information* between the pair of responses (r_1, r_2) and the stimulus

s [15, 167]

$$I[r_1, r_2; s] = \frac{1}{2} \log \left[1 + \frac{a^2 \sigma_s^2}{(1 + \rho_\eta)\sigma_\eta^2} \right]. \tag{5.4}$$

For Eqs. (5.1, 5.2), $\rho_s = 1 > 0$, and then $\rho_\eta < 0$ corresponds to increase in the information.

In the case of a chemical ligand being absorbed by two identical receptors, the mean values of r_1 and r_2 change in the same way with the ligand concentration, so that $\rho_\mathrm{s} = 1 > 0$. At the same time, a molecule absorbed at one receptor cannot be absorbed at the other, which should give $\rho_\eta < 0$, and hence will increase the measured information according to the sign-rule. However, in computational neuroscience, where these ideas originated, noise (co)-variances are inferred empirically and are, in principle, unconstrained. In contrast, in cell biology, intrinsic noises are generated from the discreteness and stochasticity of individual chemical reaction events [152, 168, 169], which constrains relations among these quantities. In particular, ρ_η may depend on σ_η, and then it is unclear if the sign rule would hold in Eq. (5.4). Indeed, the main goal of this work is to show that, in contrast to neurobiology and to a molecular setup where negative intrinsic correlations are a result of an explicit mutual inhibition [25], measuring concentration with two identical receptors does not obey the sign rule.

5.3 Model of two diffusively coupled receptors

We consider two identical receptors that can bind ligand molecules with a rate k_in, cf. Fig. 5.1. No more than one molecule can be bound to each receptor at the same time. The bound molecule can be absorbed/deactivated with the rate k_abs, freeing the receptor (such absorbing receptors collect more information about the stimulus compared to the receptors where the ligand can unbind and rebind [53]). Alternatively, it can unbind and leave the vicinity of receptors with the rate k_off. Finally, it can leave one receptor and diffuse to the other. We model this as a transition between the receptors with the rate k_hop, which in reality would depend on the diffusion constant, the distance between the receptors, and the binding rate. The number of molecules absorbed on both receptors over time t, $\{Q_1(t), Q_2(t)\}$, carries information about the binding rate k_in. In its turn, k_in is proportional to the ligand concentration, so that counting absorbed molecules measures the concentration.

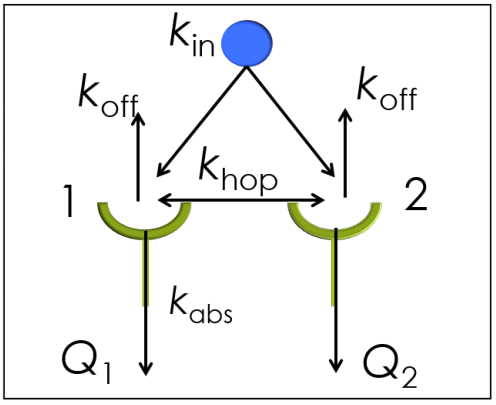

FIGURE 5.1: **Model of two diffusively coupled receptors**

Receptors 1 and 2 can bind ligands with rate k_{in}, and the bound molecules can detach and diffuse away to infinity with the rate k_{off}. The bound ligand also can be absorbed with the rate k_{abs}, or they can dissociate and diffuse to the other receptor (hop) with the rate k_{hop}. $Q_{1/2}$ is the number of ligands absorbed through the receptor 1 or 2.

Within this setup, we aim to understand how the hopping coupling between the two receptors affects the information about the concentration, $I[Q_1, Q_2; k_{\text{in}}]$. Note that the hopping can change the conditional distribution $P(Q_1, Q_2|k_{\text{in}})$, which could affect the information, but it cannot change the conditional distribution of the total number of captured molecules $Q_+ = Q_1 + Q_2$. Thus the change in the information, if any, can come only from the dependence between $Q_- = Q_1 - Q_2$ and k_{in}. This is in contrast to much of the molecular sensing literature [52, 53, 143], where one estimates k_{in} based only on the integrated number of observed ligands. Therefore, together with our main question, we need to quantify if the set of individual responses of all receptors is more informative about the concentration than the integrated response.

5.4 Solution

5.4.1 Linear Interactions

We first analyze the case where there is no restriction on the number of molecules that can occupy the receptor at a time. If n and m are the occupancies of receptors 1 and 2 respectively, then the master equation for the occupancy of each receptor $(1, 2)$ can be

written as

$$\begin{aligned}\dot{P}_{n,m}(t) =& k_{\text{in}}\left[P_{n-1,m}+P_{n,m-1}-2P_{n,m}\right]+k_{\text{hop}}\left[(m+1)P_{n-1,m+1}+(n+1)P_{n+1,m-1}\right.\\ &\left.-(m+n)P_{n,m}\right]+k_{\text{abs}}\left[(n+1)P_{n+1,m}+(m+1)P_{n,m+1}-(m+n)P_{n,m}\right]\end{aligned} \tag{5.5}$$

where n and m are the number of ligand molecules on receptor 1 and 2, respectively. To find the distribution of (Q_1, Q_2) the master equation is modified by multiplying the terms corresponding to absorption with e^{χ_n} and e^{χ_m}:

$$\begin{aligned}\frac{\partial \tilde{P}_{n,m}(t)}{\partial t} &= k_{\text{in}}\left(\tilde{P}_{n-1,m}+\tilde{P}_{n,m-1}-2\tilde{P}_{n,m}\right)+k_{\text{hop}}\left((m+1)\tilde{P}_{n-1,m+1}+(n+1)\tilde{P}_{n+1,m-1}\right.\\ &\left.-(m+n)\tilde{P}_{n,m}\right)+k_{\text{abs}}\left[(n+1)\tilde{P}_{n+1,m}e^{\chi_n}+(m+1)\tilde{P}_{n,m+1}e^{\chi_m}-(m+n)\tilde{P}_{n,m}\right].\end{aligned} \tag{5.6}$$

$\tilde{P}(\{\chi_n, \chi_m\})$, gives the cumulant generating function for the distribution of molecules absorbed through each receptor. To solve Eq. (S5.6) we define the generating function as $f(x, y) = \sum_{n,m} y^m x^n \tilde{P}_{n,m}$.

Multiplying both sides of Eq. (S5.6) by $x^n y^m$ and summing over all values of n and m from 0 to ∞, we get:

$$\begin{aligned}\frac{\partial f}{\partial t}-[k_{\text{abs}}e^{\chi_n}+yk_{\text{hop}}-(k_{\text{abs}}+k_{\text{hop}})x]\partial_x f-[k_{\text{abs}}e^{\chi_m}+xk_{\text{hop}}-(k_{\text{abs}}+k_{\text{hop}})y]\partial_y f\\ =k_{\text{in}}(x+y-2)f\end{aligned} \tag{5.7}$$

To solve Eq. (S5.7), we use the method of characteristics and set the coefficients of the terms corresponding to $\partial_x f$ and $\partial_x f$ as $\partial_t x$ and $\partial_t y$ respectively,

$$\begin{aligned}-[k_{\text{abs}}e^{\chi_n}+yk_{\text{hop}}-(k_{\text{abs}}+k_{\text{hop}})x] &= \partial_t x\\ -[k_{\text{abs}}e^{\chi_m}+xk_{\text{hop}}-(k_{\text{abs}}+k_{\text{hop}})y] &= \partial_t y\end{aligned} \tag{5.8}$$

which give x and y as,

$$(x+y) = (e^{\chi_n}+e^{\chi_m})+(x_0+y_0-(e^{\chi_n}+e^{\chi_m}))e^{k_{\text{abs}}t} \tag{5.9}$$

$$(x-y) = \left(x_0-y_0-\frac{k_{\text{abs}}(e^{\chi_n}-e^{\chi_m})}{(2k_{\text{hop}}+k_{\text{abs}})}\right)e^{(2k_{\text{hop}}+k_{\text{abs}})t}+\frac{k_{\text{abs}}(e^{\chi_n}-e^{\chi_m})}{(2k_{\text{hop}}+k_{\text{abs}})} \tag{5.10}$$

Now, along the plane defined by these x and y, the equation for the generator can be written as a total derivative $df/dt = k_{\rm in}(x+y-2)f$, which has the solution:

$$\begin{aligned} f(t) &= f(0)\exp\left[k_{\rm in}\int(x+y-2)dt\right] = f(0)\exp\left[k_{\rm in}\int(x_0+y_0-2)e^{-k_{\rm abs}t}dt\right] \\ &= f(0)\exp\left[k_{\rm in}(x_0+y_0-2)/k_{\rm abs}(1-e^{-k_{\rm abs}t})\right] \end{aligned} \tag{5.11}$$

In steady state the occupancy of each receptor is Poisson and hence

$$f(0) = \sum_{n,m} x^n y^m P_{n,m} = \sum_{n,m} x^n y^m \lambda_1^n \lambda_2^m \frac{\exp[-\lambda_1-\lambda_2]}{(n!m!)} \tag{5.12}$$

The generating function becomes:

$$f(\mathrm{x},\mathrm{y},\mathrm{t}) = \sum_{m,n=0}^{\infty}\left[\frac{\exp\left(-\lambda_1-\lambda_2\right)\lambda_2^{\mathrm{m}}\lambda_1^{\mathrm{n}}}{n!m!}\exp\left(\frac{\mathrm{k_{in}(x+y-2)\left(1-e^{t(-k_{abs})}\right)}}{\mathrm{k_{abs}}}\right)\mathrm{x^n y^m}\right] \tag{5.13}$$

where x, y are given by Eq.[5.9-5.10]. The mean occupancy at the two sites and the total number of molecules absorbed through each receptor are given as:

$$\bar{n} = \frac{\partial f(\mathrm{x},\mathrm{y},\mathrm{t})}{\partial x}|_{x,y\to1,\chi_+,\chi_-\to0} \tag{5.14}$$

$$\bar{m} = \frac{\partial f(\mathrm{x},\mathrm{y},\mathrm{t})}{\partial y}|_{x,y\to1,\chi_+,\chi_-\to0} \tag{5.15}$$

$$\bar{Q}_{\rm N} = \frac{\partial \mathrm{Log[f(x,y,t)]}}{\partial \chi_+}|_{x,y\to1,\chi_+,\chi_-\to0} \tag{5.16}$$

$$\bar{Q}_{\rm M} = \frac{\partial \mathrm{Log[f(x,y,t)]}}{\partial \chi_-}|_{x,y\to1,\chi_+,\chi_-\to0} \tag{5.17}$$

which are obtained as ($\lambda_1 = \lambda_2 = \lambda$):

$$\bar{n} = \bar{m} = \frac{k_{\rm in}}{k_{\rm abs}} + \left(\lambda - \frac{k_{\rm in}}{k_{\rm abs}}\right)e^{-tk_{\rm abs}} \tag{5.18}$$

$$\bar{Q}_{\rm N} = \bar{Q}_{\rm M} = k_{in}t + \left(\lambda - \frac{k_{in}}{k_{\rm abs}}\right)(1-e^{-tk_{\rm abs}}) \tag{5.19}$$

The covariance $\langle Q_n Q_m\rangle - \langle Q_n\rangle\langle Q_m\rangle$ can be obtained as

$$\frac{\partial^2 \mathrm{Log[f(x,y,t)]}}{\partial\chi_+\partial\chi_-}|_{x,y\to1,\chi_+,\chi_-\to0} \tag{5.20}$$

which is zero for this case. Similarly it can be shown that $(\langle nm\rangle - \langle n\rangle\langle m\rangle) = 0$ at all times.

In case of linear interactions the covariance between the number of ligand molecules absorbed through the receptors is zero at all times, as if these receptors were independent. What this means is that the identity of the ligand in terms of whether it got absorbed at receptor 1 or receptor 2 does not provide any more information compared to just the total. Naively one would expect that the information about the total concentration would get contribution from both the sum and the difference of the molecules absorbed through each receptor. Surprisingly, in case of linear interactions the difference does not contribute to the information about the ligand concentration.

5.4.2 Non-Linear Interactions

Now we analyze a biophysically realistic situation where the number of ligand molecules that can occupy the receptor at a time is finite. Such a restriction results in a non-linear interaction between the receptors. To understand the effects of such non-linear interactions we consider here the simplest possible case that only one ligand can occupy the receptor at a time.

To calculate the distribution $P(Q_1, Q_2|k_{\rm in})$, we start with the master equation describing the dynamics of the vector of probabilities of having 0 or 1 molecules bound to each of the receptors, $\mathbf{P} = \{P_{nm};\, n, m = 0, 1\}^T = \{P_{00}, P_{01}, P_{10}, P_{11}\}^T$,

$$\dot{\mathbf{P}}(t) = -H\,\mathbf{P}(t). \tag{5.21}$$

Here the generator matrix is

$$H = \begin{bmatrix} 2k_{\rm in} & -k_{\rm off} - k_{\rm abs} & -k_{\rm off} - k_{\rm abs} & 0 \\ -k_{\rm in} & k_{\rm tot} & -k_{\rm hop} & -k_{\rm abs} - k_{\rm off} \\ -k_{\rm in} & -k_{\rm hop} & k_{\rm tot} & -k_{\rm abs} - k_{\rm off} \\ 0 & -k_{\rm in} & -k_{\rm in} & 2k_{\rm off} + 2k_{\rm abs} \end{bmatrix}, \tag{5.22}$$

with $k_{\rm tot} = k_{\rm in} + k_{\rm off} + k_{\rm abs} + k_{\rm hop}$.

To find the probability distribution of the numbers of molecules absorbed by both receptors, we use the standard generating functional technique [170–174]. Namely, we

separate out the parts of H that correspond to the absorption events

$$H \equiv H_0 + H_{\text{abs},1} + H_{\text{abs},2}, \tag{5.23}$$

$$H_{\text{abs},1} = \begin{bmatrix} 0 & -k_{\text{abs}} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -k_{\text{abs}} \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{5.24}$$

$$H_{\text{abs},2} = \begin{bmatrix} 0 & 0 & -k_{\text{abs}} & 0 \\ 0 & 0 & 0 & -k_{\text{abs}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5.25}$$

Then we tag the terms corresponding to the absorption reactions by counting fields e^{χ_1} and e^{χ_2}, forming the tagged generator matrix,

$$\tilde{H}(\chi_1, \chi_2) \equiv H_0 + H_{\text{abs},1} e^{\chi_1} + H_{\text{abs},2} e^{\chi_2}. \tag{5.26}$$

Finally we realize that the vector of moment generating functions (or the Laplace transforms) of $P(Q_1, Q_2 | k_{\text{in}}, n, m)$, denoted as $\mathbf{Z}(\chi_1, \chi_2, t) = \{Z_{00}, Z_{01}, Z_{10}, Z_{11}\}$, satisfies the tagged master equation

$$\dot{\mathbf{Z}}(\chi_1, \chi_2, t) = -\tilde{H}(\chi_1, \chi_2)\mathbf{Z}(\chi_1, \chi_2, t). \tag{5.27}$$

We are interested in the long-time asymptotic, where each receptor has had many absorption events, $Q_1, Q_2 \gg 1$. Then the solution of Eq. (5.27) can be approximated as

$$\mathbf{Z}(\chi_1, \chi_2, t) \approx \mathbf{Z}(0) \exp[-\tilde{\lambda}_{\text{min}}(\chi_1, \chi_2)\, t], \tag{5.28}$$

where $\tilde{\lambda}_{\text{min}}$ is the smallest real part eigenvalue of $\tilde{H}$. From here, one can read off the cumulant generating functions conditional on the occupancy of the receptors, to the leading order in t, $F_{mn}(\chi_1, \chi_2, t) \approx -\tilde{\lambda}_{\text{min}}(\chi_1, \chi_2)\, t$. As expected, the leading order behavior is the same for any value of m, n. Thus the mean values and the (co)variances of the numbers of absorbed molecules, conditional on k_{in} all scale linearly with time. They can be obtained by differentiating $\tilde{\lambda}_{\text{min}}(\chi_1, \chi_2)$ with respect to χ_1 and χ_2. Denoting

by $\langle \dots | k_{\text{in}} \rangle$ expectations conditional on k_{in}, we write:

$$\langle Q_m | k_{\text{in}} \rangle = t \left. \frac{\partial \tilde{\lambda}_{\min}(\chi_1, \chi_2, t)}{\partial \chi_m} \right|_{\chi_1, \chi_2 = 0}, \tag{5.29}$$

$$\langle \delta Q_m \delta Q_n | k_{\text{in}} \rangle = t \left. \frac{\partial^2 \tilde{\lambda}_{\min}(\chi_1, \chi_2, t)}{\partial \chi_m \partial \chi_n} \right|_{\chi_1, \chi_2 = 0}. \tag{5.30}$$

In its turn, the eigenvalue $\tilde{\lambda}_{\min}$ can be obtained using standard non-Hermitian perturbation theory considering χ_m as the perturbation parameters around the eignevalue $\lambda_{\min} = 0$ of the unperturbed Hamiltonian. For compactness of notation, we define $k_{\text{ioa}} = k_{\text{in}} + k_{\text{off}} + k_{\text{abs}}$. This allows us to write:

$$\langle Q_n | k_{\text{in}} \rangle = \frac{k_{\text{in}} k_{\text{abs}}\, t}{k_{\text{ioa}}}, \tag{5.31}$$

$$\begin{aligned} \langle \delta Q_n \delta Q_n | k_{\text{in}} \rangle &= \langle Q_n | k_{\text{in}} \rangle \\ &\times \left(1 - \frac{2 k_{\text{in}} k_{\text{abs}}}{k_{\text{ioa}}^2} + \frac{2 k_{\text{hop}} k_{\text{in}} k_{\text{abs}}}{k_{\text{ioa}}^2 (k_{\text{tot}} + k_{\text{hop}})} \right), \end{aligned} \tag{5.32}$$

$$\langle \delta Q_1 \delta Q_2 | k_{\text{in}} \rangle = -2 \langle Q_n | k_{\text{in}} \rangle \frac{k_{\text{hop}} k_{\text{in}} k_{\text{abs}}}{k_{\text{ioa}}^2 (k_{\text{tot}} + k_{\text{hop}})}. \tag{5.33}$$

These expressions fully define the conditional distribution $P(Q_1, Q_2 | k_{\text{in}})$ to the leading, Gaussian order. Notice that $\langle \delta Q_1 \delta Q_2 | k_{\text{in}} \rangle < 0$ as long as $k_{\text{hop}} \neq 0$, and thus, according to the sign rule, we expect more information from the two correlated receptors than the two independent ones (i. e., $k_{\text{hop}} = 0$).

In the basis of $Q_\pm = Q_1 \pm Q_2$, the covariance matrix diagonalizes, and we get

$$\langle Q_+ | k_{\text{in}} \rangle = \frac{2\, k_{\text{in}} k_{\text{abs}}\, t}{k_{\text{ioa}}}, \tag{5.34}$$

$$\langle Q_- | k_{\text{in}} \rangle = 0, \tag{5.35}$$

$$\left\langle \delta Q_+^2 | k_{\text{in}} \right\rangle = \langle Q_+ | k_{\text{in}} \rangle \frac{\left[k_{\text{ioa}}^2 - 2 k_{\text{in}} k_{\text{abs}} \right]}{k_{\text{ioa}}^2}, \tag{5.36}$$

$$\left\langle \delta Q_-^2 | k_{\text{in}} \right\rangle = \langle Q_+ | k_{\text{in}} \rangle \frac{\left[k_{\text{ioa}}^2 - 2 k_{\text{in}} k_{\text{abs}} + 2 k_{\text{hop}} k_{\text{ioa}} \right]}{k_{\text{ioa}} (k_{\text{tot}} + k_{\text{hop}})}, \tag{5.37}$$

$$\langle \delta Q_+ \delta Q_- | k_{\text{in}} \rangle = 0. \tag{5.38}$$

Since neither $\langle Q_+ | k_{\text{in}} \rangle$ nor $\left\langle \delta Q_+^2 | k_{\text{in}} \right\rangle$ depend on k_{hop}, these expressions clearly show that the total number of molecules absorbed by the two receptors is not affected by the interaction parameter k_{hop}, as we alluded to previously. The coupling between the

receptors only affects the variance of the difference of the number of molecules coming from each receptor.

We now define the absorption currents $J_\pm = Q_\pm/t$, so that $\langle J_\pm|k_{\rm in}\rangle = \langle Q_\pm|k_{\rm in}\rangle/t$, and $\langle \delta J_\pm^2|k_{\rm in}\rangle = \langle \delta Q_\pm^2|k_{\rm in}\rangle/t$. Now assuming a Gaussian marginal distribution of $k_{\rm in}$, with the mean $\bar{k}_{\rm in}$ and the variance $\sigma^2_{k_{\rm in}}$, we can write down the marginal distribution of absorption currents averaged over all possible external signal concentrations

$$P(J_+,J_-) = \int \frac{dk_{\rm in}}{\sqrt{2\pi}\sigma_{k_{\rm in}}} \exp\left[-\frac{(k_{\rm in}-\bar{k}_{\rm in})^2}{2\sigma^2_{k_{\rm in}}}\right] \frac{t\exp\left[-\frac{t(J_+-\langle J_+|k_{\rm in}\rangle)^2}{2\langle \delta J_+^2|k_{\rm in}\rangle} - \frac{tJ_-^2}{2\langle \delta J_-^2|k_{\rm in}\rangle}\right]}{2\pi\sqrt{\langle \delta J_+^2|k_{\rm in}\rangle \langle \delta J_-^2|k_{\rm in}\rangle}}. \tag{5.39}$$

Note that $\langle \delta J_\pm^2|k_{\rm in}\rangle \propto 1/t$ for large t. This is the usual manifestation of the law of large numbers, so that the ratio of the standard deviation of the currents to their means decreases as $\propto 1/t^{1/2}$.

Both $\langle J_+|k_{\rm in}\rangle$ and $\langle \delta J_\pm^2|k_{\rm in}\rangle$ depend on $k_{\rm in}$. We assume that $\sigma^2_{k_{\rm in}}$ is small, so that this dependences can be written to the first order in $\delta k_{\rm in} = k_{\rm in} - \bar{k}_{\rm in}$. Then the dependence of the mean currents on $k_{\rm in}$ still preserves the Gaussian form of Eq. (5.39), while the dependence of the variance manifests itself in sub-Gaussian orders. To the leading order in small $\sigma^2_{k_{\rm in}}$, the marginal distribution of the currents is still a product of two Gaussians,

$$P(J_+,J_-) = \frac{\exp\left[-\frac{(J_+-\langle J_+\rangle)^2}{2\sigma_+^2} - \frac{J_-^2}{2\sigma_-^2}\right]}{2\pi\sigma_+\sigma_-} \tag{5.40}$$

with

$$\langle J_+\rangle = \frac{2\,\bar{k}_{\rm in}k_{\rm abs}}{\bar{k}_{\rm ioa}}, \tag{5.41}$$

$$\sigma_+^2 = \frac{\langle \delta J_+^2|\bar{k}_{\rm in}\rangle}{t}\left(1 + \frac{\left(\frac{\partial\langle J_+\rangle}{\partial k_{\rm in}}\right)^2 \sigma^2_{k_{\rm in}}}{\langle \delta J_+^2|\bar{k}_{\rm in}\rangle/t}\right), \tag{5.42}$$

$$\sigma_-^2 = \frac{\langle \delta J_-^2|k_{\rm in}\rangle}{t}. \tag{5.43}$$

The mutual information we are seeking is $I[Q_1,Q_2;k_{\rm in}] = S(Q_1,Q_2) - \langle S(Q_1,Q_2|k_{\rm in})\rangle_{k_{\rm in}}$, where S are the marginal and the conditional entropies. In the limit of small σ_k^2 we

obtain:

$$\mathrm{I}(\{Q_1, Q_2\}; k_{\mathrm{in}}) = \frac{1}{2} \ln \left[1 + \frac{\left(\frac{\partial \langle J_+ \rangle}{\partial k_{\mathrm{in}}} \right)^2 \sigma_k^2}{\left\langle \delta J_+^2 | k_{\mathrm{in}} \right\rangle / t} \right]_{\bar{k}} , \tag{5.44}$$

which is independent of k_{hop}. Thus, to the leading order in $\delta k_{\mathrm{in}} = k_{\mathrm{in}} - \bar{k}_{\mathrm{in}}$, the mutual information is independent of the interaction between the receptors. In terms of Eq.(5.4), although the covariance coefficient (ρ) is negative, the quantity $(1+\rho)\sigma^2$ ($= \left\langle \delta J_+^2 | k_{\mathrm{in}} \right\rangle / t$) is independent of k_{hop} and so is the mutual information. To see the effect of interactions, one needs to take into account the higher order corrections.

Before we go further, it is instructive to notice that the total absorption current J_+ is bounded by $2 \times k_{\mathrm{abs}}$. This bound ($J_+ = 2k_{\mathrm{abs}}$) would be attained when both the receptors are saturated ($k_{\mathrm{in}} = \infty$) and in such a situation hopping would have no effect. The effect of hopping can only be seen at values of k_{in} lower than or comparable to k_{abs}.

To compare the effects of coupling to independent receptors we look at the quantity $\Delta I(k_{\mathrm{in}}, k_{\mathrm{abs}}, k_{\mathrm{hop}}) = I(k_{\mathrm{in}}, k_{\mathrm{abs}}, k_{\mathrm{hop}}) - I(k_{\mathrm{in}}, k_{\mathrm{abs}}, 0)$. Our Gillispie simulations (Fig. 5.2) shows that this quantity is less than zero, which means that diffusive coupling reduces the mutual information. In Fig. 5.2, we plot the difference $\Delta I(k_{\mathrm{in}}, k_{\mathrm{abs}}, k_{\mathrm{hop}})$ for different values of hopping rate k_{hop}. As can be seen, the difference in mutual information goes to zero at long times indicating that the diffusive coupling does not provide any extra information compared to independent receptors at long times. At $t = 0$, the mutual information is zero. This makes sense, since at $t = 0$ no ligand molecule has been absorbed.

At intermediate times the difference in mutual information is small but negative, meaning that diffusive coupling reduces mutual information between the signal and the response. As we proved above the interactions only affect the difference between the number of molecules absorbed through each receptor. The diffusive coupling adds an extra noise to this difference and "washes-out" any spatial informations the molecules carry about each individual receptor. This extra noise reduces the mutual information compared to the case where the receptors are independent and there is no "wash-out" effect.

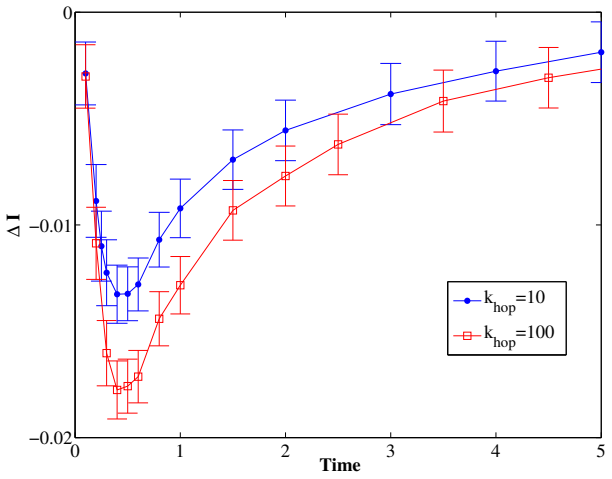

FIGURE 5.2: **Difference between Mutual Information for coupled and independent receptors**

The difference between the mutual information of two independent receptors and the mutual information of two receptors coupled through diffusion. We use Gillispie algorithm to simulate the model and NSB entropy estimator to calculate the mutual information and error-bars. Here $k_{\text{in}} = k_{\text{abs}} = 10, k_{\text{off}} = 0$. The difference in information ΔI is negative, which means that diffusively coupled receptors have lower information compared to independent receptors.

5.5 Discussion

In this chapter we have analyzed a simple model of two identical receptors that are coupled through a diffusive interaction. We study the effect of the coupling between the receptors on the concentration estimation of an external ligand. We find that, in the steady state, the diffusive coupling has no effect on the mutual information between the ligand concentration and the number of molecule absorbed through the receptors. This means that the interactions amongst receptors do not improve the concentration estimation apart from the 1/N reduction in the variance of the estimation of concentration by having multiple receptors. This is true even though the extrinsic signal correlations and the intrinsic noise correlations resulting from the interactions between receptors are orthogonal suggesting a higher mutual information according to the well known "sign-rule" [17].

Our analysis also shows that in the transient period the mutual information between the interacting receptors is lower than that of independent receptors. Although non intuitive at first, this makes sense as the diffusive interactions between the receptors add another source of randomness in the estimation and hence reduce the accuracy. This suggest

that unbinding and rebinding of ligands results in lower accuracy, as has been noted previously [54].

In the model of cellular receptors discussed here, the correlation structure between the response of the receptors emerges out from biophysically plausible interactions. We show that such interactions effect both the variance in the response of the receptors and their covariances. In previous work, analyzing the effects of correlations between decoding units on the informations the variance in the estimates is assumed to be independent of the interactions between the units [17–21]. Our model clearly shows that biophysical interactions do not necessarily obey such restrictions. This observation opens up new avenues in the understanding of effects of interactions in multivariate complex systems on the mutual information between signal and response.

Chapter 6

Conclusion

Biological systems are complex. They are made of a large number of individual functional units that interact among themselves in a complex manner. Understanding the design and functional aspects of these systems from the first principle presents challenges, both experimental and theoretical. Coarse graining approaches that identify the important features of these complex systems and leave out the details, reveal interesting properties of these networks. Here we attempted to identify some of these features using various approaches such as continuum modeling, real space renormalization, maximum likelihood inference and information theory. These approaches revealed the essential features that are required in biological networks for large scale emergent phenomena. Global information present in the signal can be identified through long range anisotropic connections. Long range connections also fill-in the missing parts of noisy signal for a complete reconstruction of the signal. Hierarchical small-world type connections provide large scale end-to-end connectivity even at low connection probability. Time delays introduced through signaling networks can be employed to utilize the temporal structure in a mixture of multiple signals to estimate each one of them accurately. These emergent properties of biological networks provides clues to the general rule of their structural and functional organization.

Although we were able to identify some of the design principles of sensing and information processing on biological networks, our work reveals new question in the structure of biological systems and opens new areas of research. The renormalization approach of chapter 3 gives insight to how the hierarchical small world type long range connections might provide large scale efficient connectivity and criticality in the cortex. It would be interesting to see how such connections might give rise to persistent activity and

dynamical switching between different states [175] by studying a model neural network with spiking neurons with a hierarchy of synaptic connections.

The existence of temporal codes in neural systems has been an active area of research [2–4, 58, 59], but not much work has been done to understand the how the temporal structure receptor-ligand systems can be utilized to estimate concentrations. The realization of Enders and Wingreen [54] that recording a temporal sequence of receptor binding provide more information has farther reaching consequences than originally thought. We show that an access to such sequence not only improves the accuracy of estimation, but also allows for a greater information as one can estimate the concentration of multiple ligands using just one receptor. This observation provides an insight into the questions related to olfaction, immunology and cellular signaling where each of these systems perform discrimination of signals much more than the number of receptors available in the system.

The observation of chapter 5 that a biophysically realistic model of receptor coupled through diffusive interaction does not follow the "sign-rule" has posed more question than answers about the effects of correlations on information. Most of the theoretical work in neuroscience has focused on the basic assumption that interactions do not affect the variance of decoding units (neurons) [17–21, 25]. It would be interesting to revisit the ideas about effects of correlations on population coding and discover new rules of information transduction on complex biological networks.

Our approach to identify the important features of biological information processing on complex networks that are essential to perform a certain task has not just helped in understanding particular phenomena, but also provided insights into broad classes of similar tasks. We have made some small steps towards building an understanding of multivariate information processing in biological systems, but we are still far from a complete understanding of the design principles of information processing on biological networks. It will take a long time to reveal Nature's complete code, an this thesis shows just few of my first steps.

Appendix A

Appendix

A.1 Automated graph counting

The recursion relations (3.5) for MK1 are obtained by a process of graph counting depicted in Fig. (3.3). As the number of possible graphlets increases exponentially for more complicated hierarchical networks (e.g. HN5 and HNNP), automating the graph enumeration process *insilico* makes it easier to obtain their recursion equations. Key to this process is the adjacency matrix A_{ij}, which gives the information about the presence of single bonds between two sites in a graph.

A.1.1 Counting MK1 graphlets

In the MK1-graphlet in Fig. 3.3a,

$$A_a = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \tag{A.1}$$

is an example of an adjacency matrix when all possible bonds are present. The bonds are bi-directional, which results in a symmetric matrix, and the diagonal elements are zero, since there are no bonds that loop back to a site. In the case where two ends are not connected by a single bond, the adjacency matrix effectively searches for alternate paths to connect the two end-sites. In Fig 3.3e, for example, the small-world bond is missing, and sites 1 and 3 are not connected via a single bond. The adjacency matrix is

thus,

$$A_e = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{A.2}$$

By itself, the adjacency matrix gives the number of one-step end-site connections. To find the number of two-step end-site connections for a graphlet, the adjacency matrix must be squared. The off-diagonal elements of A^2 give the number of possible paths between two sites that are exactly two hops long. Squaring the adjacency matrix in Fig. 3.3a (Eq. A.1) gives

$$A_e^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}. \tag{A.3}$$

Since matrix element $A^2_{e,13} = 1$, there exists only one possible path in which two-steps can be made to connect the end-sites. Since the maximum path length for the simple case of MK1 is two, only $A_{e,13}$ (one step) and $A^2_{e,13}$ (two steps) need to be checked for finding end-to-end connections.

The graphlets are classified as contributing to $T_{n+1}(x)$ or $S_{n+1}(x,y)$ depending on whether an end-to-end connection exists. The weights of the graphlets are calculated by first labeling the end-sites as x and y. Both end-sites are labeled x in fully-connected graphs contributing to $T_{n+1}(x)$, and unconnected graphs contributing to $S_{n+1}(x,y)$ contain the left end-site labeled x and the right end-site labeled y.

For each graphlet in the n^{th} generation, x or y is assigned to each site and $T_n(x)$ or $S_n(x,y)$ to each bond, depending on whether the end sites are attached. Isolated sites/clusters are assigned a weight of 1. The contribution of each graphlet in the $(n+1)^{th}$ generation is set as the product of the value assigned to the bonds and intermediate sites. For example, the two shaded backbone bonds of Fig. 3.3a indicate that the graphlet has two bonds of type $T_n(x)$. The small-world bond exists with probability p, and all the sites are connected to the same cluster. Therefore, the graphlet contributes to $T_{n+1}(x)$ in the next generation with weight $p\,x\,T_n^2(x)$. Similarly, for the graphlet in Fig. 3.3f, the backbone bonds are of the types $T_n(x)$ and $S_n(x,y)$. The small-world bond is absent with probability $1-p$, and the end-sites are connected to separate clusters, x

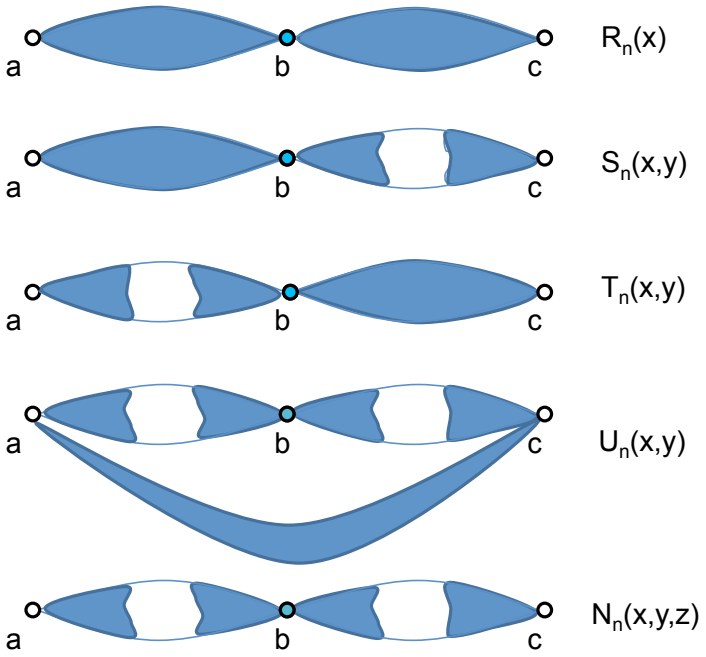

FIGURE A.1: **Diagrammatic definition of generating functions for HNNP and HN5**

Sites a, b and c represent the end-sites of the network. $R_n(x)$ consist of one cluster spanning all three end-sites, $S_n(x,y)$, $T_n(x,y)$ and $U_n(x,y)$ two clusters, one of which spanning two end-sites, and $N_n(x,y,z)$ represents non-spanning clusters which connect to at most one end-site.

and y. Hence, this graphlet contributes to $S_{n+1}(x,y)$ in the next generation with weight $(1-p)\,x\,T_n(x)S_n(x,y)$.

A.1.2 Cluster generating function for HNNP:

The generating functions for the Hanoi network HNNP in Fig. 3.1 can be calculated using the same principles described for MK1. As in Sec. 3.3.1, we define the generating

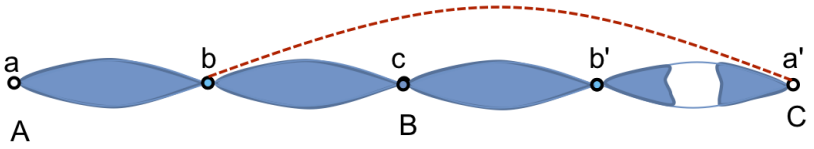

FIGURE A.2: **Example graphlet for HNNP**

By looking at the elements of A^2_{13} & A^2_{13} of A^2, one can see that all the three end-sites are connected. So this graph contributes to $R_{n+1}(x)$ in the next generation. In fact all the sites are connected to the same cluster in this case, which can be verified by looking other element of A, A^2, A^3 & A^4. Since all sites are connected to the same cluster (say of size x) and there is only one long range small world bond is present, the weight of the graphlet is $p(1-p)x^2R_n(x)S_n(x,x)/4$.

functions for HNNP depicted in Fig. A.1:

$$R_n(x) = \sum_{k=0}^{\infty} r_k^{(n)}(p)x^k, \tag{A.4}$$

$$S_n(x,y) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} s_{k,l}^{(n)}(p)x^k y^l, \tag{A.5}$$

$$U_n(x,y) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} u_{k,l}^{(n)}(p)x^k y^l, \tag{A.6}$$

$$N_n(x,y,z) = \sum_{k=0}^{\infty} n_{k,l,m}^{(n)}(p)x^k y^l z^m, \tag{A.7}$$

where we introduce the probabilities

- $r_k^n(p)$ that sites a, b and c are all connected within the same cluster of size k;
- $s_{k,l}^n(p)$ that a and b are mutually connected within a cluster of size k, and c is connected to a separate cluster of size l;
- $t_{k,l}^n(p)$ that a is connected to a separate cluster of size k, and b and c are mutually connected within cluster of size l;
- $u_{k,l}^n(p)$ that a and c are mutually connected within a cluster of size k, and b is connected to a separate cluster of size l;
- $n_{k,l,m}^n(p)$ that a is connected to a cluster of size k, b is connected to a cluster of size l, and c is connected to a cluster of size m, but all mutually disconnected.

The symmetry of $s_{k,l}^n$ and $t_{k,l}^n$ are included in the definition of $S_n(x,y)$ [112]. As for MK1, the three end-notes themselves are not counted in the cluster size.

We want to obtain the system of RG recursions for generating functions, where $(R_{n+1}, S_{n+1}, U_{n+1}, N_{n+1})$ are functions of $(R_n, S_n, U_n, N_n; p)$. The algorithm first generates the adjacency matrices corresponding to all possible $(2^8 = 256)$ graphlets for the HNNP network. For each one of these graphlets the possibility of their contribution to one of $(R_{n+1}, S_{n+1}, U_{n+1}, N_{n+1})$ in the next generation is checked using the adjacency matrices.

As an example of our graph counting algorithm for HNNP, we consider the graphlet in Fig. A.2. At first glance it appears that there are two separate clusters of sizes k and l. The adjacency matrix for this graphlet is

$$A = \begin{array}{c} \text{Node} \\ a \\ b \\ c \\ b' \\ a' \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & b' & a' \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{A.8}$$

where the disconnect between sites a' and b' is indicated by $A_{4,5} = A_{5,4} = 0$. After the sites b and b' in Fig. A.2 are decimated in the RG step, the remainder is matched with one of the graphlets in the generating function diagram in Fig. A.1. Thus, only the matrix elements in Eq. A.8 that connect end sites a to c, a to a', and c to a' contribute to the recursion equations for the generating functions. In general, the matrix elements for A^4 must be checked for a five-point HNNP graphlet, since the maximum number of steps required to connect all end-sites is four. In our example,

$$A^4 = \begin{bmatrix} 3 & 0 & 4 & 0 & 3 \\ 0 & 10 & 0 & 4 & 0 \\ 4 & 0 & 6 & 0 & 4 \\ 0 & 4 & 0 & 2 & 0 \\ 3 & 0 & 4 & 0 & 3 \end{bmatrix}. \tag{A.9}$$

Elements A^4_{13}, A^4_{15}, and A^4_{53} are non-zero, indicating that the end sites (a, c, and a') form a contiguous cluster, where a' becomes connected by way of the small-world bond. The graphlet therefore renormalizes into an R-type bond. To determine its weight, we note that the sites a, b, and c are connected via an R_n-type bond and the sites c, b', and a' form an S_n-type bond. Only the right-hand one of the small-world bonds is present. Hence, the total weight of this graphlet in the next generation is $p(1-p)\ x^2\ R_n(x)S_n(x,x)/4$.

Here, S_n becomes a function of x in both arguments, since the small-world bond merges the previously disconnected clusters x and y. The factor 1/4 is due to the symmetry explained in Ref. [112].

This process is repeated for all 256 graphlets with our automated counting algorithm, where each graphlet is attributed to its appropriate next-generation graphlet. After adding the weights, the generating function recursion relations are found to be:[1]

$$\begin{aligned}
R'(x) =& \{xR(x) + pxU(x,x) + (1-p)U(x,1)\}^2 + \frac{3}{4}p^2x^2S(x,x)^2 \\
&+ 2pxR(x)\{pxN(x,x,x) + (1-p)N(x,1,x)\} \\
&+ pxS(x,x)\{(1-p)[xR(x) + U(x,1)] + 2xR(x) + pxU(x,x)\}, \qquad \text{(A.10)} \\
S'(x,y) =& \frac{1-p}{2}S(x,y)\left\{px^2S(x,x) + py^2S(y,y) + x^2R(x) + y^2R(y) + \right. \\
&(1-p)xy[R(x) + R(y)] + [x + (1-p)y]U(x,1) + [y + (1-p)x]U(y,1) \\
&\left. +p(x+y)^2U(x,y) + pxN(x,1,x) + pyN(y,1,y)\right\} \\
&+ \frac{p^2}{2}xyS(x,y)\{2U(x,y) + N(x,y,x) + N(y,x,y)\} \\
&+ (1-p)N(x,1,y)\{p[x+y]U(x,y) \\
&+ (1-p)[xR(x) + yR(y) + U(x,1) + U(y,1)]\} \\
&+ pxN(x,x,y)\{(1-p)[xR(x) + U(x,1)] + pyU(x,y)\} \\
&+ pyN(x,y,y)\{(1-p)[yR(y) + U(y,1)] + pxU(x,y)\} \qquad \text{(A.11)} \\
U'(x,y) =& \frac{1}{4}px[(2-p)x + 2(1-p)y]S(x,y)^2 \\
&+ pxS(x,y)^2\{(1-p)N(x,1,y) + pxN(x,x,y)\} \qquad \text{(A.12)} \\
N'(x,y,z) =& \frac{1}{4}(1-p)^2[x+y][y+z]S(x,y)S(y,z) \\
&+ \frac{1-p}{2}[x+y]S(x,y)\{(1-p)N(x,1,z) + pxN(y,x,z)\} \\
&+ \{(1-p)N(x,1,y) + pzN(x,z,y)\}\{(1-p)N(x,1,z) + pxN(y,x,z)\} \\
&+ \frac{1-p}{2}[y+z]S(y,z)\{(1-p)N(x,1,y) + pzN(x,z,y)\} \qquad \text{(A.13)}
\end{aligned}$$

Note that for $x = y = z = 1$, i.e., when graphlets are counted irrespective of cluster sizes, these equations revert back to those previously listed in Ref. [112].

[1] Primed quantities correspond to index $n+1$ and unprimed to n.

A.1.2.1 Cluster generating function for HN5:

The discussion on how to obtain the RG recursion equations for the cluster generating functions of HN5 parallels that for HNNP above. The definition of the generating functions in Eqs. A.4, as illustrated in Fig. A.1, equally apply to HN5. The main difference originates with the structure of small-world bonds, which leads to a planar graph for HN5 and a non-planar graph for HNNP. Then, our graph counting algorithm results in the following RG recursions:

$$\begin{aligned}
R'(x) =& \{U(x,1) + xR(x)\}^2 + \frac{1}{2}p^2x^2S(x,x)^2 + 2p\{N(x,1,x)U(x,1) \\
&+ xS(x,x)\left[(1-p)U(x,1) + pxU(x,x)\right]\} + pxR(x)\{2(1-p)N(x,1,x) \\
&+ 2pxN(x,x,x) + (3-p)xS(x,x) - 2U(x,1) + 2xU(x,x)\} \qquad \text{(A.14)} \\
S'(x,y) =& (1-p)N(x,1,y)\{U(x,1) + U(y,1) + (1-p)\left[xR(x) + yR(y)\right]\} \\
&+ p(1-p)\left\{x^2R(x)N(x,x,y) + y^2R(y)N(x,y,y)\right\} \\
&+ \frac{1-p}{4}S(x,y)\left\{px^2S(x,x) + py^2S(y,y)\right\} \\
&+ \frac{p(1-p)}{2}\left\{x^2\left[U(x,y) + U(x,x)\right] + y^2\left[U(x,y) + U(y,y)\right]\right\} \\
&+ \frac{(1-p)^2}{2}\left[x+y\right]\left\{U(x,1) + U(y,1)\right\} \\
&+ \frac{1-p}{2}\left\{xR(x)\left[-py + x + y\right] + yR(y)\left[-px + x + y\right]\right\} \qquad \text{(A.15)} \\
U'(x,y) =& p\left\{N(x,1,y) + \frac{1}{2}(1-p)\left[x+y\right]S(x,y)\right\}^2 \\
&+ p^2S(x,y)\left\{x^2N(x,x,y) + y^2N(x,y,y)\right\} \\
&+ \frac{p}{4}S(x,y)^2\left\{\left(1 + p - p^2\right)x^2 + 2p(1-p)xy + (2-p)py^2\right\} \qquad \text{(A.16)} \\
N'(x,y,z) =& \frac{p(1-p)}{2}\left\{S(x,y)\left[x^2N(y,x,z) + y^2N(y,y,z)\right]\right. \\
&\left. + S(y,z)\left[y^2N(x,y,y) + z^2N(x,z,y)\right]\right\} \\
&+ \frac{(1-p)^2}{2}\left\{\left[x+y\right]N(y,1,z)S(x,y) + \left[y+z\right]N(x,1,y)S(y,z)\right\} \\
&+ \frac{(1-p)}{4}S(x,y)S(y,z)\left\{(1-p)\left[xy + xz + yz\right] + y^2\right\} \\
&+ (1-p)N(x,1,y)N(y,1,z) \qquad \text{(A.17)}
\end{aligned}$$

Again, these equations revert back to those previously listed in Ref. [112] for $x = y = z = 1$.

Bibliography

[1] K Klemm and S Bornholdt. Topology of biological networks and reliability of information processing. *PNAS*, 102(51):18414–18419, 2005.

[2] J Lisman. Bursts as a unit of neural information: making unreliable synapses reliable. *Trends in neurosciences*, 20(1):38–43, 1997.

[3] W Bialek and F Rieke. Reliability and information transmission in spiking neurons. *Trends in Neurosciences*, 15(11):428–434, 1992.

[4] Z F Mainen and T Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503–1506, 1995.

[5] V Sourjik and N Wingreen. Responding to chemical gradients: bacterial chemotaxis. *Current Opinion in Cell Biology*, 24(2):262–268, 2012.

[6] R Bogacz. Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, 11(3):118–125, 2007.

[7] Y Chen, A Munteanu, Y Huang, J Phillips, Z Zhu, M Mavros, and W Tan. Mapping receptor density on live cells by using fluorescence correlation spectroscopy. *Chemistry-A European Journal*, 15(21):5327–5336, 2009.

[8] S Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3, 2009.

[9] W Bialek, F Rieke, R de Ruyter van Steveninck, and D Warland. Reading a neural code. *Science*, 252:1854–1857, 1991.

[10] G Tkačik, A Walczak, and W Bialek. Optimizing information flow in small genetic networks. *Physical Review E*, 80(3):031920, 2009.

[11] A Fairhall, E Shea-Brown, and A Barreiro. Information theoretic approaches to understanding circuit function. *Current Opinion in Neurobiology*, 22:653–659, 2012.

[12] F Tostevin and P-R ten Wolde. Mutual information between input and output trajectories of biochemical networks. *Physical Review Letters*, 102:218101, 2009.

[13] G Tkacik and AM Walczak. Information transmission in genetic regulatory networks: a review. *Journal of Physics: Condensed Matter*, page 153102, 2011.

[14] R Cheong, A Rhee, CJ Wang, I Nemenman, and A Levchenko. Information transduction capacity of noisy biochemical signaling networks. *Science*, 334:354–358, 2011.

[15] I Nemenman. Information theory and adaptation. In M Wall, editor, *Quantitative Biology: From Molecular to Cellular Systems*, page 73. CRC Press, 2012.

[16] G Tkačik and W Bialek. Information processing in living systems. *arXiv preprint arXiv:1412.8752*, 2014.

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

[18] I Ginzburg and H Sompolinsky. Theory of correlations in stochastic neural networks. *Physical Review E*, 50(4):3171, 1994.

[19] S Seung and H Sompolinsky. Simple models for reading neuronal population codes. *PNAS*, 90(22):10749–10753, 1993.

[20] H Yoon and H Sompolinsky. The effect of correlations on the fisher information of population codes. In *Advances in Neural Information Processing Systems 11: Proceedings of the 1998 Conference*, volume 11, page 167. MIT Press, 1999.

[21] L Abbott and P Dayan. The effect of correlated variability on the accuracy of a population code. *Neural computation*, 11(1):91–101, 1999.

[22] A Walczak, G Tkacik, and W Bialek. Optimizing information flow in small genetic networks. ii. feed-forward interactions. *Physical Review E*, 81:041905, 2010.

[23] G Tkacik, J Prentice, V. Balasubramanian, and E Schneidman. Optimal population coding by noisy spiking neurons. *PNAS*, 107:14419–14424, 2010.

[24] R da Silveira and M Berry. High-fidelity coding with correlated neurons. *PLoS Computational Biology*, 2014. arXiv q-bio.NC/1307.3591.

[25] S Hormoz. Cross talk and interference enhance information capacity of a signaling pathway. *Biophysical Journal*, 104:1170–1180, 2013.

[26] M Saunders and G Voth. Coarse-graining methods for computational biology. *Annual Review of Biophysics*, 42:73–93, 2013.

[27] N Sinitsyn, N Hengartner, and I Nemenman. Adiabatic coarse-graining and simulations of stochastic biochemical networks. *PNAS*, 106(26):10546–10551, 2009.

[28] L Swanson. *Brain architecture: understanding the basic plan.* Oxford University Press, 2012.

[29] S Thorpe, D Fize, C Marlot, et al. Speed of processing in the human visual system. *Nature*, 381(6582):520–522, 1996.

[30] T Poggio and W Reichardt. Visual control of orientation behaviour in the fly: Part ii. towards the underlying neural interactions. *Quarterly Reviews of Biophysics*, 9 (03):377–438, 1976.

[31] E Moser, E Kropff, and M Moser. Place cells, grid cells, and the brain's spatial representation system. *Annual Review of Neuroscience*, 31:69–89, 2008.

[32] T Behrens and O Sporns. Human connectomics. *Current Opinion in Neurobiology*, 22(1):144–153, 2012.

[33] J DiCarlo, D Zoccolan, and N Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.

[34] M Riesenhuber and T Poggio. Models of object recognition. *Nature Neuroscience*, 3:1199, 2000.

[35] T Moeslund and E Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81(3):231–268, 2001.

[36] N Ayache. Medical computer vision, virtual reality and robotics. *Image and Vision Computing*, 13(4):295–313, 1995.

[37] M Smith and A Kohn. Spatial and temporal scales of neuronal correlation in primary visual cortex. *The Journal of Neuroscience*, 28(48):12591–12603, 2008.

[38] A Angelucci, J Levitt, E Walton, J Hupe, J Bullier, and J Lund. Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience*, 22(19), 2002.

[39] G Leuba and R Kraftsik. Changes in volume, surface estimate, three-dimensional shape and total number of neurons of the human primary visual cortex from midgestation until old age. *Anatomy and Embryology*, 190(4):351–366, 1994.

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

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

[42] J Beggs and D Plenz. Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures. *The Journal of Neuroscience*, 24(22):5216–5229, 2004.

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

[44] J Milton. Neuronal avalanches, epileptic quakes and other transient forms of neurodynamics. *European Journal of Neuroscience*, 36(2):2156–2163, 2012.

[45] Nir Friedman, Shinya Ito, Braden AW Brinkman, Masanori Shimono, RE Lee DeVille, Karin A Dahmen, J M Beggs, and T C Butler. Universal critical dynamics in high resolution neuronal avalanche data. *Physical Review Letters*, 108(20): 208102, 2012.

[46] D Stauffer and A Aharony. *Introduction to percolation theory*. CRC press, 1994.

[47] D Chklovskii, T Schikorski, and C Stevens. Wiring optimization in cortical circuits. *Neuron*, 34(3):341–347, 2002.

[48] S Laughlin and T Sejnowski. Communication in neuronal networks. *Science*, 301 (5641):1870–1874, 2003.

[49] D Rokni, V Hemmelder, V Kapoor, and V Murthy. An olfactory cocktail party: figure-ground segregation of odorants in rodents. *Nature Neuroscience*, 2014.

[50] S Haykin and Z Chen. The cocktail party problem. *Neural Computation*, 17(9): 1875–1902, 2005.

[51] D Bray, M Levin, and C Morton-Firth. Receptor clustering as a cellular mechanism to control sensitivity. *Nature*, 393(6680):85–88, 1998.

[52] W Bialek and S Setayeshgar. Physical limits to biochemical signaling. *PNAS*, 102: 10040–10045, 2005.

[53] R Endres and N Wingreen. Accuracy of direct gradient sensing by single cells. *PNAS*, 105:15749–15754, 2008.

[54] R Endres and N Wingreen. Maximum likelihood and the single receptor. *Physical Review Letters*, 103(15):158101, 2009.

[55] B Hu, W Chen, W Rappel, and H Levine. Physical limits on cellular sensing of spatial gradients. *Physical Review Letters*, 105(4):048104, 2010.

[56] K Kaizu, W de Ronde, J Paijmans, K Takahashi, F Tostevin, and P ten Wolde. The berg-purcell limit revisited. *Biophysical Journal*, 106(4):976–985, 2014.

[57] A Mugler, A Levchenko, and I Nemenman. Limits to the precision of gradient sensing with spatial communication and temporal integration. *arXiv:1505.04346*, 2015.

[58] W Bialek, F Rieke, R Van Steveninck, and D Warland. Reading a neural code. *Science*, 252(5014):1854–1857, 1991.

[59] S Panzeri, R Petersen, S Schultz, M Lebedev, and M Diamond. The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, 29 (3):769–777, 2001.

[60] Y Hu, J Zylberberg, and E Shea-Brown. The sign rule and beyond: boundary effects, flexibility, and noise correlations in neural population codes. *PLoS Computational Biology*, 10:e1003469, 2014.

[61] V Singh, M Tchernookov, R Butterfield, and I Nemenman. Director field model of the primary visual cortex for contour detection. *PLoS One*, 2014.

[62] D Felleman and D Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1:1, 1991.

[63] O Creutzfeldt and H Nothdurft. Representation of complex visual stimuli in the brain. *Naturwissenschaften*, 65:307, 1978.

[64] D Hubel and T Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1), 1968.

[65] D Hubel and T Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160:106, 1962.

[66] G Wallis, E Rolls, et al. Invariant face and object recognition in the visual system. *Progr Neurobiol*, 51:167, 1997.

[67] L Itti, C Koch, and E Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Trans*, 20:1254, 1998.

[68] Y LeCun, L Bottou, Y Bengio, and P Haffner. Gradient-based learning applied to document recognition. *Proc IEEE*, 86:2278, 1998.

[69] S Thorpe, A Delorme, and R van Rullen. Spike-based strategies for rapid processing. *Neural Networks*, 14:715, 2001.

[70] Y-L Boureau, F Bach, Y LeCun, and J Ponce. Learning mid-level features for recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, page 2559, 2010.

[71] T Serre, A Oliva, and T Poggio. A feedforward architecture accounts for rapid categorization. *PNAS*, 104:6424, 2007.

[72] M Riesenhuber and T Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2:1019, 1999.

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

[74] Y LeCun, F Huang, and L Bottou. Learning methods for generic object recognition with invariance to pose and lighting. In *Computer Vision and Pattern Recognition, Proc 2004 IEEE Comp Soc Conf*, volume 2, page II, 2004.

[75] D Stettler, A Das, J Bennett, and C Gilbert. Lateral connectivity and contextual interactions in macaque primary visual cortex. *Neuron*, 36:739, 2002.

[76] M Colonnier, J O'Kusky, et al. Number of neurons and synapses in the visual cortex of different species. *Revue Canadienne de Biologie/editee par l'Universite de Montreal*, 40(1), 1981.

[77] D Field, A Hayes, and R Hess. Contour integration by the human visual system: Evidence for a local "association field". *Vision Res*, 33:173, 1993.

[78] P Parent and S Zucker. Trace inference, curvature consistency, and curve detection. *Pattern Analysis and Machine Intelligence, IEEE Trans*, 11:823, 1989.

[79] V Gintautas, M Ham, B Kunsberg, et al. Model cortical association fields account for the time course and dependence on target complexity of human contour perception. *PLoS Comp Biol*, 7:e1002162, 2011.

[80] J Zweck and Lance R Ws. Euclidean group invariant computation of stochastic completion fields using shiftable-twistable functions. *Journal of Mathematical Imaging and Vision*, 21(2):135–154, 2004.

[81] P Bressloff and J Cowan. The functional geometry of local and horizontal connections in a model of v1. *Journal of Physiology-Paris*, 97(2):221–236, 2003.

[82] F Wolf and T Geisel. Spontaneous pinwheel annihilation during visual development. *Nature*, 395:73, 1998.

[83] F Wolf and T Geisel. Universality in visual cortical pattern formation. *Journal of Physiology-Paris*, 97(2):253–264, 2003.

[84] G DeAngelis, J Robson, I Ohzawa, and R Freeman. Organization of suppression in receptive fields of neurons in cat visual cortex. *Journal of Neurophysiology*, 68: 144, 1992.

[85] P de Gennes, J Prost, and R Pelcovits. The physics of liquid crystals. *Physics Today*, 48:70, 1995.

[86] C Gilbert and T Wiesel. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9:2432, 1989.

[87] I Kovacs and B Julesz. A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *PNAS*, 90:7495, 1993.

[88] L Jeffress et al. A place theory of sound localization. *Journal of Comparative and Physiological Psychology*, 41:35, 1948.

[89] P Dayan and L Abbott. *Theoretical Neuroscience*. MIT Press, 2005.

[90] T Miconi and R VanRullen. The gamma slideshow: object-based perceptual cycles in a model of the visual cortex. *Frontiers in Human Neuroscience*, 4:205, 2010.

[91] C Altmann, H Bülthoff, and Z Kourtzi. Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, 13(4):342–349, 2003.

[92] V Singh and S Boettcher. Scaling of clusters near discontinuous percolation transitions in hyperbolic networks. *Physical Review E*, 90(1):012117, 2014.

[93] D Sussillo and L Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron*, 63(4):544–557, 2009.

[94] A Margolin, I Nemenman, K Basso, C Wiggins, G Stolovitzky, R Favera, and A Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7(Suppl 1):S7, 2006.

[95] D Bassett and E Bullmore. Small-world brain networks. *The Neuroscientist*, 12 (6):512–523, 2006.

[96] E Bullmore and O Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

[97] O Sporns and J Zwi. The small world of the cerebral cortex. *Neuroinformatics*, 2 (2):145–162, 2004.

[98] E Bullmore and O Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, 2012.

[99] S Achard, D Bassett, A Meyer-Lindenberg, and E Bullmore. Fractal connectivity of long-memory networks. *Physical Review E*, 77(3):036104, 2008.

[100] A Roxin, H Riecke, and S Solla. Self-sustained activity in a small-world network of excitable neurons. *Physical Review Letters*, 92(19):198101, 2004.

[101] J Lichtman, H Pfister, and N Shavit. The big data challenges of connectomics. *Nature Neuroscience*, 17(11):1448–1454, 2014.

[102] S Seung. Reading the book of memory: sparse sampling versus dense mapping of connectomes. *Neuron*, 62(1):17–29, 2009.

[103] W Shew, H Yang, T Petermann, R Roy, and D Plenz. Neuronal avalanches imply maximum dynamic range in cortical networks at criticality. *The Journal of Neuroscience*, 29(49):15595–15600, 2009.

[104] T Coogan and A Burkhalter. Hierarchical organization of areas in rat visual cortex. *Journal of Neuroscience*, 13:3749–3749, 1993.

[105] C Wessinger, J VanMeter, B Tian, J Van Lare, J Pekar, and J Rauschecker. Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, 13(1):1–7, 2001.

[106] Ala Trusina, Sergei Maslov, Petter Minnhagen, and Kim Sneppen. Hierarchy measures in complex networks. *Physical Review Letters*, 92(17):178702, Apr 2004.

[107] M. Hinczewski and A. N. Berker. Inverted Berezinskii-Kosterlitz-Thouless singularity and high-temperature algebraic order in an Ising model on a scale-free hierarchical-lattice small-world network. *Phys. Rev. E*, 73:066126, 2006.

[108] S. Boettcher, B. Gonçalves, and H. Guclu. Hierarchical regular small-world networks. *J. Phys. A: Math. Theor.*, 41:252001, 2008.

[109] A. Clauset, C. Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453:98–101, 2008.

[110] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80:1275–1335, 2008.

[111] H Rozenfeld and D ben Avraham. Percolation in hierarchical scale-free nets. *Phys. Rev. E*, 75:061102, Jun 2007. doi: 10.1103/PhysRevE.75.061102.

[112] S. Boettcher, J. L. Cook, and R. M. Ziff. Patchy percolation on a hierarchical network with small-world bonds. *Phys. Rev. E*, 80:041115, 2009.

[113] Stefan Boettcher, Vijay Singh, and Robert M. Ziff. Ordinary percolation with discontinuous transitions. *Nature Communications*, 3:787, 2012.

[114] T Hasegawa and T Nogawa. Absence of the nonpercolating phase for percolation on the nonplanar Hanoi network. *Phys. Rev. E*, 87:032810, 2013.

[115] P Minnhagen and S Baek. Analytic results for the percolation transitions of the enhanced binary tree. *Phys. Rev. E*, 82:011113, 2010.

[116] M. Bauer, S. Coulomb, and S Dorogovtsev. Phase transition with the Berezinskii-Kosterlitz-Thouless singularity in the Ising model on a growing network. *Physical Review Letters*, 94(20):200602, May 2005.

[117] S. Boettcher and C. T. Brunson. Fixed point properties of the Ising ferromagnet on the Hanoi networks. *Phys. Rev. E*, 83:021103, 2011.

[118] S Baek, H Mäkelä, P Minnhagen, and B Kim. Ising model on a hyperbolic plane with a boundary. *Phys. Rev. E*, 84:032103, 2011.

[119] E Khatjeh, S Dorogovtsev, and J Mendes. Berezinskii-Kosterlitz-Thouless-like transition in the Potts model on an inhomogeneous annealed networks. *Phys. Rev. E*, 75:041112, 2007.

[120] T. Nogawa, T. Hasegawa, and K. Nemoto. Criticality governed by the stable renormalization fixed point of the ising model in the hierarchical small-world network. *Phys. Rev. E*, 86:030102, 2012.

[121] S Boettcher and T Brunson. Classification of critical phenomena having a parameter-dependent renormalization. *arXiv preprint arXiv:1209.3447*, 2012.

[122] T Nogawa, T Hasegawa, and K Nemoto. Generalized scaling theory for critical phenomena including essential singularities and infinite dimensionality. *Physical Review Letters*, 108:255703, 2012.

[123] T. Hasegawa, T. Nogawa, and K. Nemoto. Profile and scaling of the fractal exponent of percolations in complex networks. *EuroPhys. Lett.*, 104:16006, 2013.

[124] D Krioukov, F Papadopoulos, M Kitsak, A Vahdat, and M Boguñá. Hyperbolic geometry of complex networks. *Phys. Rev. E*, 82(3):036106, Sep 2010.

[125] D Wales. *Energy landscapes*. Cambridge University Press, Cambridge, 2003.

[126] A. Fischer, K. H. Hoffmann, and P. Sibani. Intermittent relaxation in hierarchical energy landscapes. *Phys. Rev. E*, 77:041120, 2008.

[127] M Boguñá, D Krioukov, and K Claffy. Navigability of complex networks. *Nature Physics*, 5:74 – 80, 2009.

[128] D Meunier, D Lambiotte, A Fornito, K Ersche, and E Bullmore. Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics*, 3 (0):37, 2009.

[129] P Moretti and M Muñoz. Brain architecture, Griffiths phases, and the stretching of criticality. *ArXiv e-prints*, 2013.

[130] C Gray, P König, A Engel, W Singer, et al. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–337, 1989.

[131] C Stam, B Jones, G Nolte, M Breakspear, and P Scheltens. Small-world networks and functional connectivity in alzheimer's disease. *Cerebral Cortex*, 17(1):92–99, 2007.

[132] D Achlioptas, R D'Souza, and J Spencer. Explosive percolation in random networks. *Science*, 323(5920):1453–1455, 2009.

[133] R da Costa, S Dorogovtsev, A Goltsev, and J Mendes. Explosive Percolation Transition is Actually Continuous. *Physical Review Letters*, 105:255701, 2010.

[134] O Riordan and L Warnke. Explosive percolation is continuous. *Science*, 333(6040): 322–324, 2011.

[135] Eric J. Friedman and Adam S. Landsberg. Construction and analysis of random networks with explosive percolation. *Physical Review Letters*, 103(25):255701, 2009. doi: 10.1103/PhysRevE.54.3135.

[136] P Grassberger, C Christensen, G Bizhani, S Son, and M Paczuski. Explosive Percolation is Continuous, but with Unusual Finite Size Behavior. *Physical Review Letters*, 106:225701, 2011.

[137] N Araujo and H Herrmann. Explosive Percolation via Control of the Largest Cluster. *Physical Review Letters*, 105:035701, 2010.

[138] Y Cho, S Kim, J Noh, B Kahng, and D Kim. Finite-size scaling theory for explosive percolation transitions. *Phys. Rev. E*, 82:042102, 2010.

[139] W Chen and R D'Souza. Explosive Percolation with Multiple Giant Components. *Physical Review Letters*, 106:115701, 2011.

[140] T Nogawa and T Hasegawa. Transition-type change between an inverted Berezinskii-Kosterlitz-Thouless transition and an abrupt transition in bond percolation on a random hierarchical small-world network. *Phys. Rev. E*, 89:042803, Apr 2014.

[141] T Hasegawa and K Nemoto. Hierarchical scale-free network is fragile against random failure. *Phys. Rev. E*, 88:062807, 2013.

[142] A Berker, M Hinczewski, and R Netz. Critical percolation phase and thermal Berezinskii-Kosterlitz-Thouless transition in a scale-free network with short-range and long-range random bonds. *Physical Review E*, 80(4):041118, Oct 2009. doi: 10.1103/PhysRevE.80.041118.

[143] H Berg and E Purcell. Physics of Chemoreception. *Biophysical Journal*, 20:193–219, 1977.

[144] J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *PNAS*, 71(10):4135–4139, 1974.

[145] J Ninio. Kinetic amplification of enzyme discrimination. *Biochimie*, 57(5):587–595, 1975.

[146] J-B Lalanne and P Francois. Chemodetection in fluctuating environments: receptor coupling, buffering, and antagonism. *PNAS*, 112(6):1898–1903, 2015.

[147] P Francois, G Voisinne, E Siggia, G Altan-Bonnet, and M Vergassola. Phenotypic model for early T-cell activation displaying sensitivity, specificity, and antagonism. *PNAS*, 110(10):E888–97, 2013.

[148] T McKeithan. Kinetic proofreading in T-cell receptor signal transduction. *PNAS*, 92(11):5042–5046, 1995.

[149] B Goldstein, D Coombs, J Faeder, and W Hlavacek. Kinetic proofreading model. *Advances in Experimental Medicine and Biology*, 640:82–94, 2008.

[150] G Bel, B Munsky, and I Nemenman. The simplicity of completion time distributions for common complex biochemical processes. *Physical Biology*, 7:0610003, 2010.

[151] X Cheng, L Merchan, M Tchernookov, and I Nemenman. A large number of receptors may reduce cellular response time variation. *Physical Biology*, 10(3): 035008, 2013.

[152] J Paulsson. Models of stochastic gene expression. *Physics of Life Reviews*, pages 157–175, 2005.

[153] E Siggia and M Vergassola. Decisions on the fly in cellular sensory systems. *PNAS*, 110(39):E3704–E3712, 2013.

[154] S Strong, R Koberle, R van Steveninck, and W Bialek. Entropy and information in neural spike trains. *Physical Review Letters*, 80(1):197–200, 1998.

[155] P Reinagel and R Reid. Temporal coding of visual information in the thalamus. *Journal of Neuroscience*, 20(14):5392–5400, 2000.

[156] R Liu, S Tzonev, S Rebrik, and K Miller. Variability and information in a neural code of the cat lateral geniculate nucleus. *J Neurophysiol*, 86:2789–2806, 2001.

[157] I Nemenman, G Lewen, W Bialek, and R van Steveninck. Neural coding of natural stimuli: information at sub-millisecond resolution. *PLoS Computational Biology*, 4(3):e1000025, 2008.

[158] C Tang, D Chehayeb, K Srivastava, I Nemenman, and S Sober. Millisecond-scale motor encoding in a cortical vocal area. *PLoS Biology*, 12(12):e1002018, 2014.

[159] W Bialek, I Nemenman, and N Tishby. Predictability, complexity, and learning. *Neural Computation*, 13:2409, 2001.

[160] M McClean, A Mody, J Broach, and S Ramanathan. Cross-talk and decision making in map kinase pathways. *Nature Genetics*, 39(3):409–414, 2007.

[161] M Laub and M Goulian. Specificity in two-component signal transduction pathways. *Annual Review of Genetics*, 41:121–145, 2007.

[162] J Tsitron, A Ault, J Broach, A Morozov, and C Rao. Decoding complex chemical mixtures with a physical model of a sensor array. *PLoS Comput. Biol*, 7:e1002224, 2011.

[163] L Chylek, D Holowka, B Baird, and W Hlavacek. An interaction library for the fcεri signaling network. *Frontiers in Immunology*, 5, 2014.

[164] L Chylek, L Harris, C Tung, J Faeder, C Lopez, and W Hlavacek. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 6(1):13–36, 2014.

[165] A Mugler, Filipe Tostevin, and Pieter Rein ten Wolde. Spatial partitioning improves the reliability of biochemical signaling. *PNAS*, 110(15):5927–5932, 2013.

[166] P Swain, M Elowitz, and E Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *PNAS*, 99:12795–12800, 2002.

[167] C Shannon and W Weaver. *The mathematical theory of communication*. University of Illinois Press, Urbana, 1998.

[168] M Elowitz, A Levine, E Siggia, and P Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

[169] N Van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, 2011.

[170] D Bagrets and Y Nazarov. Full counting statistics of charge transfer in Coulomb blockade systems. *Physical Review B*, 67:085316, 2003.

[171] A Jordan, E Sukhorukov, and S Pilgram. Fluctuation statistics in networks: A stochastic path integral approach. *Journal of Mathematical Physics*, 45:4386, 2004.

[172] I Gopich and A Szabo. Theory of the statistics of kinetic transitions with application to single-molecule enzyme catalysis. *The Journal of Chemical Physics*, 124: 154712, 2006.

[173] N Sinitsyn and I Nemenman. The Berry phase and the pump flux in stochastic chemical kinetics. *Europhysics Letters*, 77:58001, 2007.

[174] N Sinitsyn and I Nemenman. Universal geometric theory of mesoscopic stochastic pumps and reversible ratchets. *Physical Review Letters*, 99:220408, 2007.

[175] M Tsodyks, T Kenet, A Grinvald, and A Arieli. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286 (5446):1943–1946, 1999.