Neural Stochastic Dynamics of Perceptual Decision Making

Daniel Martí i Ortega

Tesi doctoral UPF / 2008

Supervisada pel
Prof. Dr Gustavo Deco
(Departament de les Tecnologies
de la Informació i les Comunicacions)

Barcelona, octubre de 2008

The text was typeset with LaTeX using the memoir class.

Dipòsit Legal:
ISBN:

In loving memory of my brother, Jaume

Abstract

Recent studies have identified some brain regions involved in the decision process for simple perceptual tasks, and have provided insight into how sensory information is evaluated before the commitment to a motor response. To understand the dynamical principles of the decision process, and how they emerge from the interaction of interconnected neurons, computational models have been proposed based on large-scale, neurobiologically inspired networks. In these models the decision is regarded as a transition between attractors, or stable states of the network. This transition is triggered by external inputs, which change the attractor configuration and drive the system from an initial resting state to one of the other stable states, associated with the categorical choices. We show that this transition can occur by two distinct mechanisms depending on the overall external input. The decision mechanism arising at low inputs, entirely driven by noise, leads to exponentially distributed decision times, with a mean that depends exponentially on the inverse of the amplitude of the noise present in the system. Moreover, both decision times and performances are monotonically decreasing functions of the overall external stimulation, thereby providing a plausible neurobiological basis of the speed-accuracy tradeoff observed in behavioral experiments. In a second project we propose a method to simplify the description of multistable neural systems under the influence of noise. The method allows us to transform the set of stochastic differential equations describing the activities of the neural units into a set of ordinary differential equations involving the first central moments of the distributions. In a third project, we show that the dynamics of models of perceptual detection based on bistable stochastic units, in which the percept arises as noise-induced activation triggered by stimulation, can be properly captured by a one-dimensional normal-form Langevin equation for inputs near the detection threshold.

Resum

Estudis recents han identificat algunes de les regions cerebrals que participen en el procés de decisió que té lloc durant tasques perceptives senzilles, i han permès comprendre millor com la informació sensorial s'avalua durant el procés de deliberació. A fi d'entendre els principis dinàmics del procés de decisió, i de com aquests sorgeixen de la dinàmica d'interacció entre neurones interconnectades, s'han proposat models computacionals basats en xarxes a gran escala d'inspiració neurobiològica. En aquests models la decisió s'associa a una transició entre atractors, o estats estables, de la xarxa. La transició entre estats és induïda pels inputs selectius associats a l'estímul, que canvien el paisatge d'atractors del sistema, tot afavorint la transició entre l'atractor neutre inicial a un atractor selectiu, associat a una elecció categòrica. En aquest treball mostrem que aquesta transició pot tenir lloc a través de dos mecanismes qualitativament diferents en funció de la intensitat mitjana dels inputs externs. També mostrem que el mecanisme de decisió que apareix a baixes intensitats, induït exclusivament pel soroll present a la xarxa, dóna lloc a temps de decisió distribuïts exponencialment, amb una mitjana que depèn exponencialment de la inversa de l'amplitud del soroll. A més, tant els temps de decisió com el rendiment són funcions monòtones decreixents de la intensitat mitjana dels inputs externs, fet que constitueix una possible base neurobiològica plausible al compromís entre rapidesa i rendiment observats en experiments conductuals. En un segon projecte, proposem un mètode per a simplificar la descripció de sistemes neurals multiestables subjectes a soroll. El mètode transforma el conjunt d'equacions diferencials estocàstiques que descriuen l'activitat de les unitats neurals en un conjunt d'equacions diferencials ordinàries per als primers moments centrals de la distribució d'activitats neurals. Finalment, en un tercer projecte mostrem que la dinàmica dels models de detecció perceptual basats en unitats biestables estocàstiques, en els quals la percepció apareix com un fenomen d'activació per soroll induït per estimulació, pot ser capturada per una equació de Langevin per a la forma normal de la bifurcació quan l'estimulació és pròxima al llindar de detecció.

Contents

Acknowledgements

The completion of this thesis would not have been possible without the contribution of many people, to whom I want to express my sincere gratitude.

First of all, I am very grateful to my advisor Gustavo Deco, who gave me the opportunity to join his group and carry out my thesis under his supervision. He has had a very friendly, enthusiastic, and supportive attitude along this eventful journey, and has always been there to give me a hand.

I am also very thankful to Paolo Del Giudice, Maurizio Mattia and Guido Gigante for all the inspiring and fruitful discussions, their very warm welcome in Rome, and the valuable feedback they have been giving me from the start.

I would like to express my gratitude to the members of the reading committee, Albert Compte, Jordi Garcia-Ojalvo, Paolo Del Giudice, Toni Guillamon, and Vicente López, as well as their substitutes Vicent Caselles, Albert Costa, Blas Echebarria, Hector Geffner, and Anders Ledberg, for their time and interest.

I appreciatively acknowledge the funding of the European Union research project EU FP6 IST, "Decisions in Motion".

The postdocs of our group. I wish to thank Ralph G. Andrzejak, for his many careful readings and his encouraging comments; I am also indebted to Anders Ledberg, for his many keen remarks and for his sharp and uncompromised critical eye I have come to learn (not without pain :) so much from. A big thank you goes to Alex Roxin for his sensible, constructive, and enlightening remarks, and for making our science chats such a rewarding experience. Thanks also to Libertad González for her punctual but really precious help. I am specially grateful to Rita Almeida, for being so patient and kind with me in the countless occasions I knocked on her door asking for help, advice, or plain common sense.

The students of our group. These years would not have been as plentiful without your camaraderie. I want to thank especially Andres Bühlmann for being such a kind, generous, and easy-going friend and officemate, and for solving all my problems with Linux. My gratitude also goes to Mar Pérez, for her charm and for cheering me up in my hard times, and to Marco Loh for his fine kindness, and for helping me with the last-time bureaucracy of the thesis. I wish to thank Saurabh Dhawan for catching quite a few typos in the final draft, and also to Johan Larsson and Laura Dempere, as well as the newcomers, Larissa Albantakis,

Joana Cabral, Pedro Ernesto García, Étienne Hugues, Axel Kammerer, Timothée Masquelier, and Mario Pannunzi, for being so patient and understanding with my lack of social skills during these past months.

All the people at the Department who contributed to create a friendly and stimulant atmosphere at Estació de França. Thanks to Pablo Arias, Marc Bernot, Juan Cardelino, Oscar Civit, Gabriele Facciolo, Sira Ferradans, Vicenç Gómez, Gloria Haro, Davinia Hernández Leo, Laura Igual, Anders Jonsson, Andreas Kaltenbrunner, Enric Meinhardt, Héctor Palacios, Leticia Peres, Miguel Ramírez, and Tomàs Winand. I want to thank also Anna Serra for her careful revisions, and for all the supportive and inspiring lunch-time conversations, which often gave a necessary humanistic counterpoint to my materialist bias.

La Pili, l'Anna, en Carlos, i en Julio, de Ca n'Aguilera, pels plats casolans d'entre setmana.

The free software community, for all the useful tools I have profited from all these years.

The Universitat Pompeu Fabra, for financing the printing of this thesis.

Oriol Pujolàs and Susagna Prat, for their kind hospitality in New York and for encouraging me from the distance.

Ernest Montbrió, for the all the help and support, and for all the memorable conversations about science and life we have had along the years.

La meva àvia, per les dosis setmanals de saviesa popular.

Els meus pares, Joan i Anna, per tot el suport, amor, generositat i paciència infinits que m'heu dedicat sempre.

Chapter 1

Introduction

Fluctuations, understood as inevitable random perturbations in measurable variables, are pervasive in the nervous system. At a subcellular level, processes such as the activation of ionic channels, the binding of molecules to receptors, or the release of neurotransmitters are inherently random due to thermal noise. Variability is also evident at a cellular level, where the temporal patterns of action potentials generated by individual neurons are highly irregular and, often, rarely reproducible even under identical stimulation conditions. At a higher level, animals and humans can exhibit behavior that appears to be unpredictable, a feature that may be advantageous from an evolutionary point of view.

Are these different manifestations of noise related? Is the indeterminacy observed in many behavioral responses linked to the variability observed at microscopic levels? While in the physical world fluctuations are typically uncorrelated and, therefore, averaged out, in the nervous system noise may lead to persistent changes that may be noticeable at larger scales. For instance, fluctuations in local calcium concentrations can activate proteins of NMDA receptor channels, inducing persistent modifications in the synaptic strength, which, in turn, alter the activity patterns of neurons and cortical networks. This sensitivity to fluctuations is partly due to the existence of multistability, by virtue of which the effect of transient perturbations can be maintained and can potentially determine the evolution of the system at higher scales.

In this work we address the possible mechanisms operating in cortical networks by which fluctuations can be amplified to give rise to macroscopic effects on the dynamics of the network and, ultimately, on behavior. We focus in particular on the neurobiological substrate for the variability of behavioral responses during simple decision tasks.

Decision-Making

The ability to make decisions is a hallmark of intelligent behavior. Thousands of years of evolutionary pressure have molded animal brains to convert them into powerful devices capable of generating different behavioral responses depending

on the context and the animal's desires or beliefs. Decisions vary from very simple decisions about the presence or identity of a stimulus, to elaborated plans of action that require the evaluation of goals and potential outcomes.

The behavioral and cognitive aspects of decision-making have been studied for decades, and considerable strides have been made in understanding the behavioral principles underlying goal-oriented, conscious choices (see, e.g., [96]). It has been been only recently, however, that the neural substrates of decision-making have begun to be elucidated. The progress has been particularly notable in the context of perceptual decisions, a simple type of decision that requires the transformation of sensory information into an appropriate action. By recording the activity of single cells within the brains of monkeys engaged in perceptual decision tasks, it is possible to relate behavior to neural activity. Although the perceptual task may vary, it always requires the experimental monkey to use specific stimuli as cues to produce a motor response.

Simple tasks like these allow the experimenter to have full control over the sensory inputs presented to the subject and to quantify behavior with direct measures such as the subject's reaction time or the response accuracy. Besides, the time needed to commit to a decision is in these tasks typically long, of several hundred milliseconds, making easier to identify and characterize the neural correlates of the decision formation process. Also, the fact that the sensory and motor pathways involved in the task are well known provides a hint as to where in the sensory-motor pathway the decision process is taking place.

Visual-saccadic tasks

A paradigmatic example of this approach has been the study of decision-making in visual-saccadic tasks. Trained monkeys have to watch a visual display showing a cloud of randomly moving dots, a small fraction of which drift in a coherent way to one of two predefined directions. The monkey has to decide to which one of the two possible directions this fraction of dots is actually moving, and report its choice with a rapid eye movement —a saccade— to the corresponding target choice. The difficulty of the task can be controlled by varying the fraction of dots that move coherently. While the monkey performs the task, the activity of LIP neurons is recorded. Neurons are chosen such that their receptive field encompasses one of the target choices, but that lies off the area where the random-dot stimulus is shown.

A good reason to study this apparently awkward task, called a "random-dot direction discrimination task" (Fig. 1.1), is that both the visual system and the brain stem circuits that control saccadic eye movements are probably the two most heavily studied neural systems in vertebrate brains. For this reason, the study of the particular pathways that connect the visual sensory system to the eye-movement control system has become a benchmark for the study of the general principles of sensory-motor processes.

It has been known since the late 1980s that there is a small area in the brains of primates, called the middle temporal area (area MT), which is highly specialized for processing moving visual stimuli, like those used in random-dot tasks. Cells in MT

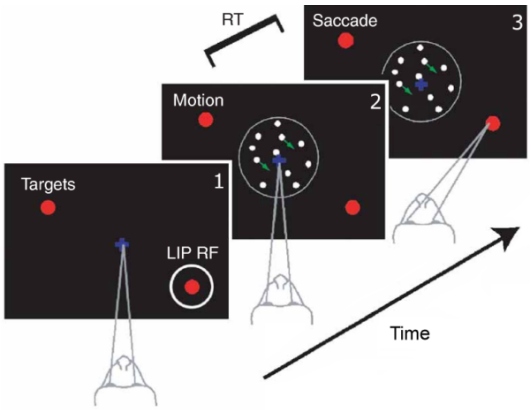


Figure 1.1: Reaction time version of the random-dot direction discrimination task. **1.** While the monkey is fixating its gaze at the center, two choice targets appear in the periphery, one of them in the response field (RF) of the LIP neuron being recorded. **2.** After a delay period of variable duration the random-dot stimulus appears. Both the fraction of dots moving coherently and the direction of coherent motion, which can be to either of the two possible choice targets, are chosen randomly on each trial. **3.** The monkey has to indicate the perceived direction of motion with a saccadic eye movement to the associated target at any time after the random-dot motion onset. The reaction time is the time elapsed between motion onset and saccade initiation. Figure adapted from [164].

are direction selective, meaning that their response is maximal when stimuli move in one particular direction. This maximal evoked response correlates well with the fraction of dots moving in the associated direction, known as the "motion strength" or "coherence" of the stimulus. Rather strikingly, the rate is also correlated with the probability that the monkey makes a saccade associated to that direction of motion [148]. Furthermore, when MT neurons are stimulated electrically, the probability that the monkey makes a particular saccade is altered [169]. These and subsequent studies suggested that cells in MT encode, in the firing rates, the instantaneous strength of the motion in their preferred direction [33, 34, 32, 172].

If cells in MT encode instantaneous motion strength, which clearly is the relevant sensory information to integrate in the random-dot task, the neural correlates of the perceptual decision process must occur somewhere downstream area MT. The neural correlates must also be positioned upstream the areas involved in saccade initiation, like the frontal eye fields [162, 35, 36] and the superior colliculus [198, 176]. An area that fulfills this requirements is the *lateral intraparietal area* (LIP). Neurons in LIP project to the frontal eye fields [21, 20] and superior colliculus [17], and receive projections from cells in MT [135]. Many neurons in LIP respond to visual stimuli that lie at the target of a planned saccadic eye movement [86, 50]. When the direction of motion of the random-dot stimulus instructs the monkey the choice of the saccadic-target, the average activity of LIP predicts the monkey's saccadic response [173]. In particular, the trial-averaged

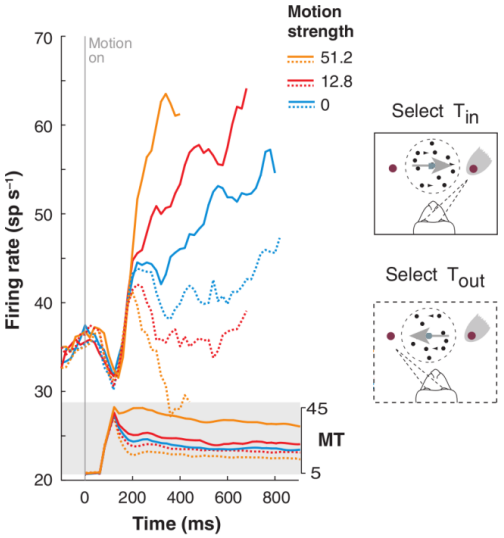


Figure 1.2: Average activity of LIP neurons during decision formation. Responses are obtained from 54 LIP neurons, averaged over trials, and grouped by stimulus coherence and direction of the saccadic choice, as indicated by the key. All responses are aligned to the onset of random-dot motion. The inset shaded in gray shows the average response of direction-selective neurons in MT to motion in the preferred and antipreferred directions. The activity of MT cells is approximately constant after a transient period. Adapted from [88].

activity of LIP cells builds up at a rate proportional to the coherence of the stimulus when dots move in the direction of the response field of the LIP neuron [173, 164] (Fig. 1.2). When dots move instead in the direction away the neuron's response field, the trial-averaged activity of the neuron ramps down at a rate also modulated by the coherence of the stimulus. The ramping of averaged activity of LIP neurons has been used to support the assertion that neurons in LIP would be integrating over time the sensory signals projected from area MT [173, 87].

Models

A biologically-inspired cortical model of decision that accounts for some prominent aspects of the decision-related neural activity of LIP was first proposed by Wang [191]. The cortical model implements a competition between two neural groups that encode the two possible choices in the task. Each group receives inputs proportional to the amount of sensory evidence for the choice they encode, mimicking the signal provided by MT neurons. The activation of these inputs forces the network to change its initial, spontaneously active state, in which both subpopulations show low firing activity, to a decision state where one of the neural groups fires at a significantly higher rate than the other. The outcome of the decision is the choice associated with the winning population, and is random by virtue of the noise present in the system.

The model by Wang et al. provides a plausible explanation for the slowness of the decision mechanism, characterized by reaction times of the order of hundreds of milliseconds. Long reaction times arise in their model as a result of long time constants of the NMDA receptor-mediated currents as well as the attractor configuration of the system. The network is driven upon stimulation to a competition regime by an increase of the selective external inputs, which destabilizes the initial

state. The decision process can thus be seen as the *relaxation* from an unstable stationary state [91] towards either of the two decision states of the network.

In this work we explore an alternative mechanism for slow decision that does not require long synaptic time scales. Unlike the regime studied in [191, 197], here we focus on those cases where stimulation does not destabilize the initial state, but rather increases the probability for a noise-induced transition to any of the decision states. We show that the proposed fluctuation-driven scenario for decision-making entails distinctive implications for the statistics of decision times. In particular, we show using numerical simulations that mean decision times tend to the Van't Hoff-Arrhenius exponential dependence on the amplitude of noise [84, 187] in the limit of vanishingly small noise. Furthermore, the distributions of decision times become in this regime more skewed as the noise amplitude is decreased, and tend to exponentials in the limit of vanishing noise, as expected from the theory of escape problems in gradient systems, where the notion of potential barrier is substantiated. We also show that mean decision times and performances are non-increasing functions of the average input to the competing units. The result lends more support to the analytical and numerical work done in [165], where the monotonic dependence on the input of both decision times and performances was rigorously derived at the critical point where the initial state loses its stability. Such monotonic dependence constitutes a plausible physiological substrate of the speed-accuracy tradeoff observed in behavioral experiments. Overall, the results suggest that noise-driven transitions among attractors constitute an alternative mechanism to describe the variability and wide range of decision times observed experimentally, which span from a few hundreds milliseconds to more than one second [123, 164].

Outline of the thesis

The work presented in this thesis is about the effect that fluctuations and multistability have on the dynamics of cortical network models, and, in particular, the possible implications that these effects may have on behavioral measures. Special emphasis is placed on biophysically realistic network models of decision-making, and on dimensional reductions of the original high-dimensional network models.

The thesis is structured as follows. Chapter 2 provides the general background on the modeling of cortical networks. It is basically a review of the neuronal, synaptic, and network models used in this work. The mean field approximation, which provides the first step toward a reduction in the dimensionality of the large-scale network models we use, is described in Chapter 3. Although these two chapters are collection of well known results, I have decided to include them to make the text as self-contained as possible. The reader familiarized with cortical network models may wish to skip them or use them as a refresher. The reader not interested in the technical aspects of the network models used in the thesis may wish to skip them too. Chapter 4 is a review on the existing network models of decision-making and serves as a nontechnical starting point for the next chapters. Chapter 5 is devoted to noise-driven transitions in attractor-based models of decision and constitutes the core of the thesis. In it we address, using a cortical

network model, the plausibility of noise-induced mechanisms operating at the network level in explaining the observed response times and performances, as well as the existing electrophysiological data. The following two chapters are more technical, and serve to illustrate two methods particularly relevant in the study of multistable systems with noise. Thus, Chapter 6 deals with the method of moments, by which we can describe the evolution of a stochastic system with a set of ordinary differential equations involving the first central moments of the density function. In Chapter 7 we illustrate how center manifold reduction can be applied to derive a one-dimensional nonlinear diffusion equation for detection. The method can be applied to relate psychophysical measures with the parameters of the network models.

Chapter 2

Cortical network models

In this chapter we summarize the theoretical framework used in the study of large-scale network models. We start giving a short survey on the physiology and anatomy of neurons and cortical networks, emphasizing the aspects that have constrained and inspired the network models described in this chapter. A short summary of experimental data on delay activity during short term memory tasks is also given. We then give an introduction to the theoretical framework of attractor neural networks. The rest of the chapter is devoted to the description of the realistic attractor models used in our work, based on large-scale networks of spike emitting units. We introduce the integrate-and-fire neuron, a simple description of neuronal dynamics that has become the choice of many theoretical studies on large-scale networks of spiking neurons. It follows a summary of the different models used to describe synaptic currents. Once the dynamical principles for the spike emitting units are established, we describe how these units interact when connected in a network.

The main references for the contents of this chapter have been the seminal papers [8, 7], the review [65], and the chapters "Network models of memory" in [40] and "Mean-Field Theory of Irregularly Spiking Neuronal Populations and Working Memory in Recurrent Cortical Networks" in [159]. Further references are given below.

2.1 Neurons

Neurons are cells specialized in processing information. Although we are still far from understanding the mechanisms by which information is processed in the brain, it is clear that these must rely on the ability of neurons to generate electrical signals in response to electrical or chemical inputs, and to transmit them to other cells. These signals travel in the form of changes in the electrical polarization of the cellular membrane. Under normal conditions the potential inside the cell is lower than the potential outside, which is 0 by convention. This voltage across the membrane arises from the permeability of the membrane to ions and the different ionic concentrations found at both sides of the membrane.

When the neuron is stimulated strongly enough, a sudden and brief reversal in the polarization, known as action potential or *spike*, occurs at one point on the cell's membrane. The action potential propagates down the membrane of the axon until cell's termination, where it triggers the release of neurotransmitters into the synaptic cleft. Neurotransmitters then diffuse across the intrasynaptic space and bind to receptors on the postsynaptic cell's membrane, which open up and allow for the influx of ions into the postsynaptic neuron. This in turn alters the membrane potential of the postsynaptic neuron by an amount called the postsynaptic potential (PSP). This PSP can be positive or negative, causing the membrane potential of the postsynaptic cell to depolarize (increase) or to hyperpolarize (decrease). The likelihood that the postsynaptic neuron emits a spike thus increases or decreases depending on the sign of the PSP, and we can then say that a presynaptic spike either *excites* or *inhibits* the postsynaptic neuron. If the effect of the incoming PSPs is such that the membrane potential of the postsynaptic neuron reaches a certain threshold, an action potential will be initiated at the postsynaptic cell, and the process will start again.

2.2 Cortical columns

Neurons are highly interconnected. A typical neuron has an elaborate branching structure, called the dendritic arbor, which allows it to receive inputs from many other neurons through synaptic connections. These inputs may come from very distant brain areas or even from anywhere in the body, as the axon from a single neuron can be as long as more than one meter. In the neocortex, neurons are distributed in six layers highly interconnected within cylindrical columns. Every column covers $\sim 1\,\mathrm{mm}^2$ of the cortex surface, and contains about 10^5 neurons, of which 80% are excitatory and 20% are inhibitory. Any single neuron in the module receives around 10^4 connections, most of which (50–80%) from neurons belonging to the same column [29]. Studies with cortical slices show that the probability that two neighboring neurons ($\lesssim 200\ \mu$m) are monosynaptically connected is about 10% [126, 104], and that the impact of a single presynaptic spike onto a postsynaptic neuron is small compared to the total excursion from the resting potential to the threshold [128, 174, 78]. Cells within a column show similar response properties, because they form a local processing network [146, 145]. This, together with the fact that the connectivity patterns within and among columns repeat quite stereotypically across the cortex, suggest that cortical columns may be the fundamental computational modules in the neocortex.

2.3 Delay activity

Many brain areas exhibit *delay activity*, by which neural populations that respond to a familiar stimulus can sustain elevated firing activity even after removal of the stimulus that elicited it. Delay activity was first discovered in the early seventies in experiments with awake monkeys [83, 116]. These experiments combined electrophysiological recordings with delayed response tasks, in which subjects

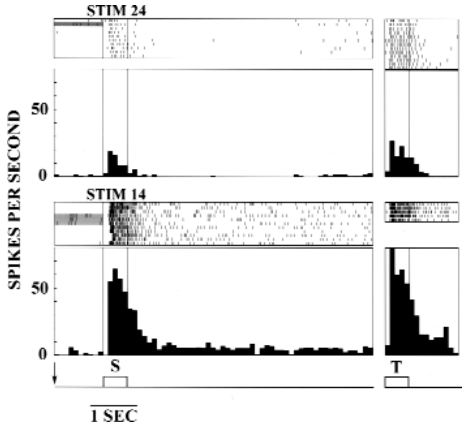


Figure 2.1: Delay activity in IT cortex during a delayed match-to-sample task. Top and bottom panels show the rastergram and the trial-averaged firing rate for the same cell, in trials in which two different sample stimuli were shown. The cell shows selective delay activity for stimulus #14, but not for #24. S, sample stimulus; T, test stimulus. Adapted by permission from Macmillan Publishers Ltd: Nature Neuroscience [201], copyright (1998).

have to remember a particular feature of an initial eliciting stimulus, such as its identity, color, or spatial location, in order to get a reward. Although there is a wide variety of experimental protocols, trials typically consist of three stages. In the first stage, called the *cue* period, an image (or *sample*) is shown to the subject for several hundred milliseconds. The image then disappears for a *delay* period that may span several seconds, to which follows the *test* period, in which another stimulus is shown, which the monkey has to compare with the sample in order to perform correctly and get a reward. For instance, during the test period two images may be shown: the sample image together with a distractor. The response is considered correct if the monkey chooses the sample image. Another version is the delayed match-to-sample task, in which only one image is shown, either the sample or a distractor. The monkey has to press a lever if the test image coincides with the sample image. In either case, the monkey has to hold an item in memory during several seconds.

While the monkey is performing the task, the neuronal activity is recorded using one or several electrodes. Before any stimulus is presented to the subject, the cells being recorded exhibit *spontaneous* (or *background*) activity, characterized by low firing activities of several spikes per second. Upon presentation of the sample during the cue period, some of the recorded cells start firing at higher rates, of several tens of spikes per second. Strikingly, these same cells keep firing during the delay period at rates significantly higher than the few Hz observed during the spontaneous state . Delay activity has been found in different cortical areas, such as inferior temporal (IT) cortex (see Figure 2.1), prefrontal cortex (PFC) and posterior parietal cortex (PPC) (for reviews see, e.g., [82] or [142]).

The precise properties of delay activity depend on the cortical area. In IT neurons, delay activity is often strongly stimulus selective, highly reproducible, and conveys information about the identity of the stimulus, rather than about simple features of it. The representation of the stimulus in the firing activity of IT cells is distributed and sparse: a single neuron may be selective to a few visual stimuli, and a single stimulus involves the activation of a small fraction of cells. This fraction is estimated to be around a few percent [140]. Moreover, the neural representation is robust against simple modifications of the stimulus, such as affine transformations, changes of color, or degradation with noise [141]. This suggests that delay activity is a collective phenomenon, arising from the properties of the neural network, rather than caused by changes in the state of the cell being recorded (see [5] or [65] for a discussion).

Delay activity in PFC and PPC is more complex, as it involves the representation of other properties of the stimulus, aside from its identity. Neurons in PFC and PPC exhibit delay activity during tasks in which the subject has to hold in memory not only the identity but also the location of a stimulus [46]. In these cases, delay activity may be selective for location, for object identity or even for both [155]. A classical example of a task requiring spatial short-term memory is the oculomotor delayed response task [80]. The monkey has to fix its gaze at a central spot on the screen. After a short time, a flash light is presented at a given, eccentric location. The monkey has to keep fixating at the central point throughout the delay period, which spans until the central spot disappears. At this moment, the monkey has to respond by making a saccade to the memorized position of the cue stimulus. Neurons in PFC and the PPC show location-tuned elevated activity during the delay period [80].

2.3.1 Biophysical substrate of multistability

The existence of delay activity indicates that the neural system exhibits multistability. The question arises as to what synaptic, cellular, or network mechanisms can explain delay activity. In general, multistability can be achieved with any mechanism that combines non-linearities with positive feedback. Several biophysical mechanisms have been proposed to explain the observed neural multistability (see [190] or [40] for details).

At the single cell level, multistability can arise from the positive feedback provided by active currents [125, 74, 147, 195] A large class of models show bistability in some range close to the firing threshold. Examples include uni- [102] and multicompartment [27, 28] models, with voltage-dependent currents [121], or with intrinsic regenerative calcium dynamics [122].

Another plausible hypothesis is that positive feedback loops arise from synaptic reverberations in recurrent circuits [89, 170, 53]. It is still debated whether the anatomical substrate for the recurrent excitatory loops are complex circuits involving cortical and subcortical areas, a network between cortical areas only, or are produced locally within a cortical module. The delay activity observed in a given area may reflect the delay activity somewhere else in the cortex, and might not be taking active part in the attractor dynamics. Whatever the substrate is, it

seems reasonable to assume that the dynamical principles of attractors do not depend much on whether or not the delay activity arises from the interaction of different cortical areas. Multistability with a large number of attractors can be obtained with recurrent network models in a robust way [10, 85, 8, 7, 65] In the models used in this work, multistability arises at a network level. To keep the model as economical as possible, the positive feedback loop is assumed to be provided by the collateral connections within the *local* module alone.

2.4 The attractor picture

Delay activity can be explained as the expression of attractor dynamics exhibited in cortical networks. By attractor we mean a stable configuration, or state, of the network, characterized by the stationary distribution of neuronal activities. The spontaneous activity patterns observed in the absence of stimulation, for example, may be regarded as a stable state of the cortical network. Delay activity patterns may also be regarded as attractors, characterized by the elevated firing of a subset of neurons, which persists even without external stimulation. These self-sustained patterns of activity, also referred to as *local reverberations*, may well be the internal representations of some features of the learned stimuli shown to the monkey during training (see [5]). Their stability results from the positive feedback provided by the recurrent connectivity and the nonlinearities in the response properties of neurons [8].

Experience shapes the set of all synaptic efficacies through a Hebbian mechanism, driven by the covariance of pre- and postsynaptic firing activities [97]. The process of learning a stimulus involves the potentiation of the synapses connecting the cells activated by the stimulus. The items being stored in memory thus leave an imprint in the synaptic structure, in such a way that when the cells driven by the original stimulus are re-activated, they cooperate to sustain their activity when the stimulus is removed [5]. Note that this involves two types of memory: a short-term (or *active*) memory, maintained by the sustained activation of a subset of neurons (the 'reverberatory activity of a cell assembly', using Hebb's words), and a long-term (or *passive*) memory in the form of persistent changes in synaptic efficacies. Synaptic plasticity allows for the formation of new activity patterns and enhances the stability of already created ones. The appearance of a familiar stimulus singles out one of the items imprinted in the network's synaptic structure, by activating the particular ensemble of cells associated with the stimulus.

When the stimulus presented is not exactly equal to any of the items previously learned, the synaptic structure will drive the initial activation pattern towards the 'closest' activation pattern it learned to sustain [5]. The network acts therefore as a *content addressable memory*: large classes of stimuli will elicit the same persistent activity for all the stimuli in a class. The different patterns of delay activity, or memory states, can be regarded in this sense as an internal representation of the class of stimuli that elicits it. Furthermore, the internal representations of all the stimuli leading to the same pattern of delay activity are considered to be in the same *basin of attraction*. On the other hand, when two stimuli

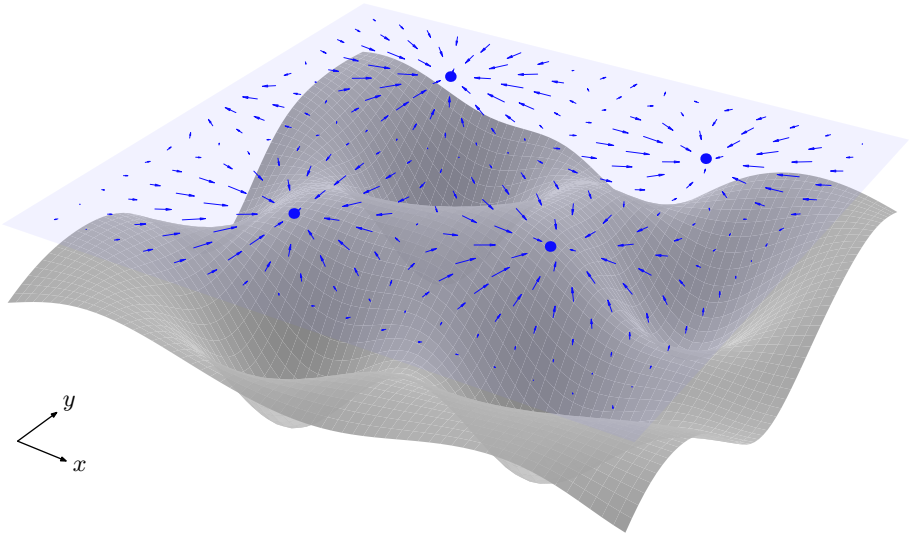


Figure 2.2: Illustration of the attractor landscape of a network. The landscape corresponds to an hypothetic energy or Lyapunov function defined on the high-dimensional space of all network configurations, formally represented in the two-dimensional x–y plane. Arrows point to the direction of evolution of the system at different points of the configuration space. All trajectories in the configuration space converge to one of the attractors (blue dots). Attractors correspond to the minima of the landscape, basins correspond to valleys, and basin boundaries correspond to ridges. Adapted from [9].

are different enough to induce different persistent activities, their corresponding internal representations are considered to be in different basins of attraction.

From a dynamical systems perspective, persistent activity states may be thought of as attractors of the network dynamics [194, 3, 105, 5, 204, 8]. The dynamics of the network are such that the system always evolves into one of the existing attractors. The configuration space can thus be partitioned into different basins of attraction, defined by the set of initial configurations leading to the same attractor. According to this point of view, the convergence of the network state from an initial configuration to the nearby attractor represents the retrieval of the stored memory on the basis of a partial cue. Unfortunately, only in some particular network models, as in the Hopfield model [105, 106], it is possible to find an energy function that decreases as the system evolves according to its dynamical rules. The existence of an energy function allows one to make use of the tools of statistical mechanics [12, 13, 14, 4]. Attractors correspond in this case to the local minima of this energy function defined on the configuration space. The dynamics of this system can be thought of as the motion of a particle on the energy surface under the influence of gravity and friction (see Figure 2.2).

In general there is no such energy function, and hence the notion of landscape is devoid of meaning, although the concept of network attractor is still valid. We will often refer to the attractor landscape metaphor due to its illustrative power.

Thus, for example, we shall say that the creation of new memory corresponds to the appearance of new attractor, and its corresponding basin of attraction. And, conversely, we shall say that forgetting a memory corresponds to the vanishing of both the attractor and its basin. We can say that changes in the synaptic structure of the network modify the attractor landscape in the sense that learning entails the creation, destruction, or modification of basins of attraction. It is important to remark that, in network models, the ability to restore an item from a degraded or partial cue arises from the collective nature of the persistent activity patterns, which involve the selective activation of a few thousands of neurons in the module. If the number of activated neurons happens to be too low, the module cannot sustain the selective activity, and system evolves to the background activity state.

2.5 Network models

In this section we summarize the theoretical framework of networks of spiking neurons with discrete attractors, which combines both biological plausibility and some degree of analytical tractability. These models bridge the gap between abstract models with networks of binary neurons, which have provided a deep understanding of the properties of systems with a large number of discrete attractors, and descriptive neurobiological models.

2.5.1 Neuron models: the integrate-and-fire neuron

A simple model for neuronal firing consists of considering the membrane voltage as a time-dependent state variable that tells us when an action potential is to be emitted. The evolution of the membrane potential is given by the time course of the input received by the neuron, which reflects in turn the activity of the presynaptic neurons. The integrate-and-fire model (IF) [120, 112, 161, 184] regards a neuron as a point-like element whose dynamical state is completely described by a single variable, the instantaneous value of its membrane depolarization $V(t)$. The whole real membrane is thus taken as equipotential, ignoring any effect caused by the complex spatial structure of the cell. Although this might seem a drastic reduction, a neuron of arbitrary geometry and described by a non-linear cable equation (see for example [113]) can be reduced to an IF neuron with properly rescaled parameters [16]. Moreover, the temporal dynamics of an IF neuron receiving noisy inputs do not differ much from those of a anatomically reconstructed complex pyramidal cell [175].

A simple version of the IF model is the *leaky* integrate-and-fire (LIF) model, in which the membrane potential $V(t)$ obeys the following differential equation

$$C_m \dot{V}(t) = -g_L\big(V(t) - V_L\big) - I_{\text{syn}}(t), \tag{2.1}$$

where C_m is the total membrane capacitance, g_L is the passive conductance, V_L is the resting potential, $I_{\text{syn}}(t)$ is the synaptic current that charges the neuron, and the dot over the variable $V(t)$ denotes the time derivative of the variable. We follow the convention that synaptic currents I_{syn} are positive-*outward*, and that

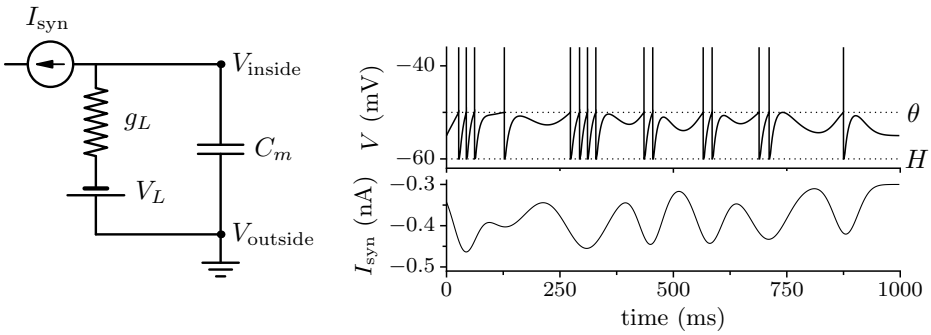


Figure 2.3: Passive integrate-and-fire model driven by a time-varying current $I_{\rm syn}(t)$. Left: Equivalent circuit of the passive integrate-and-fire cell. A capacitor C_m is charged by the contribution of the synaptic input of the neuron $I_{\rm syn}(t)$ and the leakage current due to the transmembrane conductance g_L between the interior and the exterior of the neuron. Note that synaptic currents are positive-outwards. In the absence of inputs, the membrane potential $V(t) = V_{\rm inside}(t)$ decays to the resting value V_L with time constant C_m/g_m. Right: trace of the membrane potential (upper panel) when driven by an arbitrary $I_{\rm syn}(t)$ (lower panel). Parameters: $C_m = 200\,\text{pF}$, $g_L = 20\,\text{nS}$, $V_L = -70\,\text{mV}$, $\theta = -50\,\text{mV}$, $H = -60\,\text{mV}$.

injected currents that enter the neuron through an electrode are positive-*inward*. The first term in the right side of (2.1), called the leakage term, accounts for the action of the passive ionic channels, which restore the resting potential V_L in the absence of input current. The LIF model as defined above is said to have *passive* currents because conductances do not depend on the voltage —the only conductance, g_L, is a constant. According to the differential equation (2.1) the neuron acts as a *leaky integrator* of the synaptic current or, using an electric analogy, as a simple resistor-capacitor (RC) circuit driven by a time-varying input source $I_{\rm syn}(t)$ (see Figure 2.3, left). The integrator capability of the neuron is complemented by the *fire* condition: when the depolarization $V(t)$ reaches a certain emission threshold θ, an action potential or *spike* is emitted and $V(t)$ is reset to a prescribed value $H < \theta$, following an absolute refractory period $\tau_{\rm rp}$ during which the neuron does not integrate the input.

The leaky integrate-and-fire model defined above is clearly a simplified description of the biophysics of neurons. Most notably, it does not describe how the action potential is generated; the fire condition, as well as the reset and the refractory period following it, constitute an *ad hoc* rule. The emphasis is put instead on the integration of synaptic inputs carried out by neurons in the sub-threshold regime. It is thus implicitly assumed that spikes constitute the minimal unit of information exchanged among neurons, in the form of stereotyped all-or-none responses. In this scheme, information is provided by the identity of the neuron emitting the spike and the time of emission. This level of description allows to reduce the number of the neuron's state variables down to one, and is specially suitable to study the collective dynamics of large-scale networks of

neurons due to its efficient implementation and its analytical tractability.

2.5.2 Synaptic inputs

Once we have provided a model for the neuron, we need a mathematical description of the inputs that any cortical neuron receives. The input depends on two factors: the firing activity of the afferent cells and the dynamics of synaptic activation. The first requires some assumptions about the network and the collective activity of the cells in it, and will be addressed in section 2.5.3. In this section we focus on the second, namely, on how the input current evolves in time after receiving a presynaptic spike. After a brief introduction about synaptic transmission in cortical cells, we introduce some of the available models of synaptic dynamics, from low to high complexity, and discuss the adequacy of each to describe the different types of postsynaptic currents found in cortical cells.

Neurons communicate mostly by chemical synaptic transmission. This transmission can be fast, of the order of tens of milliseconds, when transmitters bind to the synaptic channels and activate them directly. In such case, the transmission is called ionotropic, and will the one we will be concerned with. The other type of chemical synaptic transmission is slower, and involves the indirect activation of conductances through intracellular signaling pathway [77]. This type of transmission is called metabotropic, and has longer-lasting and much more diverse postsynaptic actions. Most of ionotropic transmission in the brain is mediated by two amino acid neurotransmitters, one excitatory and one inhibitory: glutamate and GABA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid), respectively. Glutamate-gated channels are divided into three subtypes, named after their selective agonists: AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid), kainate, and NMDA (N-methyl-D-aspartate) (see for example [77]). The AMPA and NMDA-gated channels mediate the bulk of ionotropic synaptic transmission in the brain. AMPA-receptor kinetics are fast, both in the activation and the deactivation phases, with characteristic times of a few milliseconds, while NMDA-receptor kinetics a considerably slower, specially in the deactivating phase, which can last several tens of milliseconds [73, 48]. Kainate receptors are also found throughout the brain and have similar properties to AMPA, but their function is not well understood. Finally, GABA receptors, in its subtype A form, mediate most of the inhibitory ionotropic transmission of vertebrates. The kinetics of GABA_A receptors are relatively fast both in the activation and the deactivation phases, with characteristic times of several milliseconds.

The current that flows into a neuron receiving a single presynaptic spike, called the postsynaptic current (PSC), depends therefore on the kinetics of the receptor channels to which neurotransmitters bind. This duration can be of the same order of magnitude or even larger than the membrane time constant. We can try to be as schematic as possible and model synaptic currents as delta pulses (instantaneous synapses) which, as we will see, offer some advantages on the analytical and computational side. Another approach would be the opposite: to provide an accurate description of the receptor gating kinetics based, for example, on Markov models [68], at the cost of losing tractability and computational speed. Here

we follow the intermediate solution initiated by [43], making some simplifying assumptions about the mechanisms of synaptic transmission in cortical cells and using a conductance-based description for the currents.

Instantaneous synapses

The simplest model of afferent input assumes that each presynaptic spike changes instantaneously the membrane polarization of the postsynaptic neuron. Each synaptic contact is associated with a given *synaptic efficacy* J which may be positive or negative depending on the excitatory or inhibitory nature of the presynaptic neuron; the value of J is related with the total charge entering the membrane upon spike arrival and hence determines the amplitude of the change in the membrane potential produced by the presynaptic spike. This simple model for afferent input is described by a sequence of delta pulses:

$$I_{\mathrm{syn},i}(t) = \sum_{j=1}^{C} J_{ij} \sum_{k} \delta(t - t_j^{(k)} - d_{ij}) \,, \tag{2.2}$$

where C is the number of synaptic sites, J_{ij} is the efficacy of each of them, $t_j^{(k)}$ is the time of the k-th spike emitted by the j-th neuron, and d_{ij} is the transmission delay between neuron j and i. Synaptic currents will be taken as injected currents, and hence positive-inward, until we introduce voltage-dependent synaptic currents, on p. 19.

The response of neuron i, initially at rest, after receiving a *single* presynaptic spike at time t_0 can be calculated by integrating (2.1)

$$V_i(t) = V_L + \frac{J_{ij}}{C_m} \exp\left(-\frac{t - t_0 - d_{ij}}{\tau}\right) \Theta(t - t_0 - d_{ij}) \,, \tag{2.3}$$

where $\Theta(t)$ is the Heaviside function ($\Theta(t) = 1$ if $t > 0$ and 0 otherwise). Thus, the PSP consists in a sudden jump in the membrane potential at the time of spike arrival, after which the potential decays exponentially to its resting value.

Synapses with instantaneous jump and exponential decay

Taking the synaptic currents as a train of delta functions is a rather crude approximation. The PSC produced by the arrival of a presynaptic spike can be more accurately described as an instantaneous rise followed by an exponential decay. This turns out to be a quite good approximation for fast synaptic currents like AMPA and $\mathrm{GABA_A}$-mediated currents [67]. The temporal evolution can be described by the following differential equation

$$\tau_s \dot{I}_{\mathrm{syn},i}(t) = -I_{\mathrm{syn},i}(t) + \sum_{j=1}^{C} J_{ij} \sum_{k} \delta(t - t_j^{(k)} - d_{ij}) \,, \tag{2.4}$$

with the same definitions used in Eq. (2.2), and where τ_s is the characteristic time constant of postsynaptic currents, which is about a few milliseconds for fast

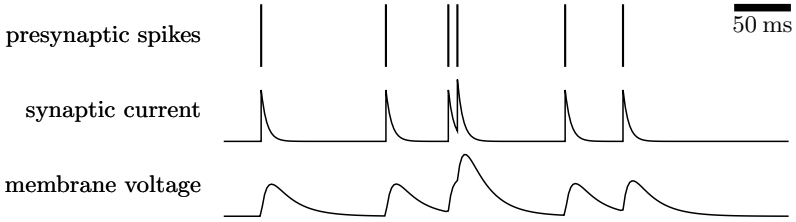


Figure 2.4: Evolution of the synaptic currents and the membrane potential for a LIF model neglecting the rise time constant of the synaptic current. Synaptic and membrane time constants are $\tau_s = 5$ ms and $\tau_m = 20$ ms, respectively.

synaptic currents like AMPA and GABA_A. Figure 2.4 illustrates the evolution of the synaptic currents and the membrane potential of a neuron receiving a sequence of spikes. Note that the solution of Eq. (2.4) is formally equivalent to that of the leaky integrator (2.1) taking the delta pulses (2.2) as synaptic current. The synapse acts thus as a filter that transforms a point process into a signal with finite jump discontinuities. Note also that taking synaptic currents as a sequence of deltas, as in Eq. (2.2), corresponds to the limit $\tau_s \to 0$ of the synaptic currents described by Eq. (2.4). Integration of the differential equation (2.4) yields

$$I_{\text{syn},i}(t) = \frac{1}{\tau_s} \sum_{j=1}^{C} J_{ij} \sum_k \exp\left[-\left(t - t_j^{(k)} - d_{ij}\right)/\tau_s\right] \Theta(t - t_j^{(k)} - d_{ij}) \,.$$

That is, the total current consists in a linear sum of single-spike contributions. For simplicity we will neglect in the following transmission delays, $d_{ij} = 0$. We will also drop the subindex i of the postsynaptic cell to make notation lighter.

Synapses with exponential rise and decay

A further step towards biological realism can be made by modeling the PSC as a difference of exponential functions

$$I_{\text{PSC}}(t) = \frac{J_{ij}}{\tau_\downarrow - \tau_\uparrow} \left[e^{-(t-t_0)/\tau_\downarrow} - e^{-(t-t_0)/\tau_\uparrow}\right] \Theta(t - t_0) \,, \tag{2.5}$$

where t_0 is the time of emission of the presynaptic spike, $\tau_\uparrow$ and $\tau_\downarrow$ are the synaptic time constants for the exponential rise and decay phases, and the prefactor ensures that the time integral of the current is J. The time course of the single-spike currents mediated by AMPA, NMDA, and GABA_A receptors can actually be modeled as a difference of exponentials [67]. Different experiments have shown that, although there is considerable variability among cells, the rise time constants for GABA_A and AMPA currents are short, of about 0.2 ms, while for NMDA currents the rise time constant can be as high as 2 ms. Decay time constants turn out to be longer than rise times, and have found to be in the range ∼1–10 ms in GABA_A and AMPA-mediated currents, while considerably longer (50–100 ms) in NMDA-mediated

Receptor	Rise phase, $\tau_\uparrow$	Decay phase, $\tau_\downarrow$	References
AMPA	0.2–0.4 ms	1.5–5.0 ms	[99, 178, 203, 18].
$GABA_A$	0.2–0.5 ms	5–10 ms	[168, 199, 90, 115, 23].
NMDA	1.5 ms	50–100 ms	[99, 167, 178, 203, 18].

Table 2.1: Estimated average values of the rising and decay time constants associated with the kinetics of AMPA, $GABA_A$, and NMDA receptors.

currents. The estimated average values of the observed time constants for all three channels are summarized in Table 2.1.

Since the rise times for $GABA_A$ and AMPA currents are one order of magnitude shorter than any other time scale, they are usually neglected in analytical studies. In such case, the first-order system (2.4) is an appropriate description of the dynamics of AMPA and $GABA_A$ currents. The same cannot be applied to NMDA currents, which have a rise time of the order of the fastest decay constant in the system ($\tau_{\mathrm{AMPA},\downarrow} \sim 1.5\,\mathrm{ms}$). In this case it is more appropriate to model the PSC with a difference of exponentials, Eq. (2.5).

The effect of a finite rise time can be included with a second variable in the system of differential equations describing synaptic dynamics:

$$\tau_\downarrow \dot{I}(t) = -I(t) + Jx(t)\,, \tag{2.6a}$$

$$\tau_\uparrow \dot{x}(t) = -x(t) + \tau_\uparrow \sum_{j=1}^{C} \sum_k \delta(t - t_j^{(k)}) \tag{2.6b}$$

In this description the total synaptic current generated by a train of presynaptic spikes is the linear sum of all single spike contributions. This is a plausible assumption as long as the presynaptic firing rates are low enough to not saturate the synaptic channels. Saturation occurs when the time elapsed between consecutive action potentials at the synapse is on average shorter the synaptic decay time constant. For the typical firing rates observed in the neocortex this implies that saturation can be safely neglected in AMPA and $GABA_A$-mediated currents, but cannot in NMDA currents. Even during spontaneous activity, during which neurons fire at a few spikes per second, the value of $\tau_{\mathrm{NMDA},\downarrow}$ is comparable to the average interspike interval (ISI). In this case the inputs saturate the postsynaptic NMDA receptors, making the actual impact of a presynaptic spike dependent on the value of the synaptic input. As a consequence, the total synaptic current is no longer a linear sum of the individual contributions. An accurate description of these effects can be obtained including a nonlinearity in Eq. (2.6a)

$$\tau_\downarrow \dot{I}(t) = -I(t) + Jx(t)\big(I_{\max} - I(t)\big) \tag{2.6a$'$}$$

where $I_{\max}$ is the maximal current attainable using a particular receptor.

Voltage dependent currents

There is an important difference between real synaptic currents and the currents we have described so far. In the models discussed above the current flows into the postsynaptic cell as mere injected current, with conductances playing no active role. They are *current*-based descriptions of the synaptic input. In contrast, real synaptic currents result from the ion permeability of the membrane and from the voltage across it. When neurotransmitters bind to receptors, the ion channels at the postsynaptic cell open, thereby increasing the membrane conductance. A realistic description of synaptic currents should include this explicit dependence of the current on the conductance and the driving force, which is the name commonly used for the difference between the voltage and the reversal potential. Models of this type are said to be *conductance*-based descriptions of the synaptic currents.

A possible description for the actual conductance is $g(V,t) = g_{\text{syn}}(V)s(t)$, where $g_{\text{syn}}(V)$ is the maximal conductance at a given voltage, and $s(t)$ is the fraction of open ion channels at the synapse. The synaptic current, understood as the current flowing *outwards* through the ion channels at the membrane, is then the product of the conductance and the driving force

$$I_{\text{syn}}(t) = g_{\text{syn}}\big(V(t)\big)s(t)\left(V(t) - V_0\right), \tag{2.7}$$

where V_0 is the synaptic reversal potential. According to Eq. (2.1), the synaptic current is depolarizing when $V_0 > V(t)$, and hyperpolarizing when $V_0 < V(t)$ [1].

The evolution of the gating variable $s(t)$ depends on the type of receptor. The time course of PSCs of AMPA and GABA_A receptors are modeled as a sudden jump followed by an exponential decay (see Eq. (2.4)). This time course can be described in terms of a first-order, linear differential equation

$$\tau_x \dot{s}_{\text{R},j}(t) = -s_{\text{R},j}(t) + \sum_{j=1}^{C}\sum_{k}\delta(t - t_j^{(k)}), \tag{2.8}$$

where R stands for either AMPA or GABA_A. The slow, NMDA-receptor mediated component of the synaptic current is better described in terms of another gating variable $s_{\text{NMDA}}(t)$ that obeys the nonlinear second-order differential equation (2.6a′)–(2.6b) [189]

$$\tau_\downarrow \dot{s}_{\text{NMDA},i}(t) = -s_{\text{NMDA},i}(t) + \alpha x_i(t)\big(1 - s_{\text{NMDA},i}(t)\big), \tag{2.9a}$$

$$\tau_\uparrow \dot{x}_i(t) = -x_i(t) + \tau_\uparrow \sum_{j=1}^{C}\sum_{k}\delta(t - t_j^{(k)} - d_{ij}), \tag{2.9b}$$

where α is a constant related to the synaptic efficacy. Note that although the system of equations (2.9a)–(2.9b) is formally equivalent to Eqs. (2.6a′)–(2.6b), the former represents the total synaptic input while the latter represents the gating variable. Figure 2.5 illustrates the effect of presynaptic spikes on the voltage-dependent synaptic currents. Note that the simpler models for synaptic inputs

[1] Recall that synaptic inputs enter in the resistor-capacitor equation with a minus sign: $C_m\dot{V}(t) = -g_L\big(V(t) - V_L\big) - I_{\text{syn}}(t)$.

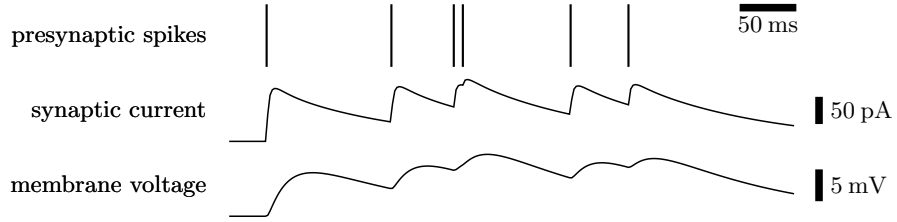


Figure 2.5: Evolution of the synaptic currents flowing into a given LIF neuron, as a result of the incoming of a presynaptic spike train, using a conductance-based description with non-negligible rise and decay synaptic time constants, and taking into account saturation effects (Eqs.(2.7)–(2.9b)). Parameters: $\tau_\uparrow = 2$ ms, $\tau_\downarrow = 5$ ms, $\tau_m = 20$ ms, $\alpha = 0.5\ (\text{ms})^{-1}$, $g_{\text{syn}} = 2$ nS, $C_m = 200$ pF, $V_0 = 0$.

described in previous sections neglect the driving force and assume implicitly that all active membrane conductances are constant in the subthreshold regime, which is not always a reasonable approximation to make.

Synaptic currents activated by NMDA-receptors are a non-trivial example of voltage-dependent current. These currents show, apart from the voltage dependence due to the driving force, a nonlinear dependence on the postsynaptic voltage because of the block of channels by extracellular magnesium ions. At the resting potential, the NMDA current is negligible even when neurotransmitters are bound to the receptor of the postsynaptic cell, because extracellular Mg^{2+} blocks the NMDA ion channel [149, 136]. The block is facilitated by the negative membrane potential of the cell. As the membrane potential increases, the electric force between Mg^{2+} and the cell decreases, and the block of channels by Mg^{2+} becomes less probable. This results in an increase of the conductance with the voltage. Using the notation introduced so far, this would amount to writing the maximal conductance $g_{\max}^{\text{NMDA}}(V)$ as a function of the voltage and the Mg^{2+} concentration. [108] have fit this dependence by

$$g_{\max}^{\text{NMDA}}(V) = g_{\text{NMDA}} \left(1 + \frac{[\text{Mg}^{2+}]}{\gamma} \exp(-\beta V)\right)^{-1}, \tag{2.10}$$

where $\gamma = 3.57$ mM and $\beta = 1/(16.13\ \text{mV}) = 0.062\ (\text{mV})^{-1}$. Typical values for extracellular magnesium concentration are around 1 mM. The main consequence of this voltage dependence is that NMDA receptors will open only when both presynaptic and postsynaptic cells are simultaneously depolarized. NMDA receptors act therefore as coincidence detectors, thereby providing the biophysical substrate for learning.

Excitatory and inhibitory cells in current- and conductance-based models

For simplicity, it is often assumed in network models that every neuron has either an excitatory or an inhibitory effect on all its postsynaptic targets (Dale's principle; but see also [182, 109]). The practical implementation of excitatory and inhibitory

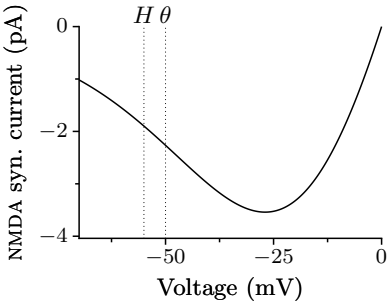


Figure 2.6: Relationship between the NMDA-mediated current and the membrane voltage. The functional dependence is given by the product of the driving force and the voltage-dependent maximal conductance, $I_{\text{NMDA}} = g_{\max}^{\text{NMDA}}(V)(V - V_E)$, where $g_{\max}^{\text{NMDA}}(V)$ is given by Eq. (2.10), and $[\text{Mg}^{2+}] = 1$ mM. The membrane voltage spends most of the time in the interval between the reset potential and the threshold (dotted vertical lines).

cells depends on the description of the inputs. In current-based descriptions, the synaptic efficacy J_{ij} takes positive values if j-th neuron is excitatory, and negative if it is inhibitory. The absolute value of J_{ij} provides the amplitude of the PSP. In conductance-based models the reversal potential V_0 in Eq. (2.7) is above threshold in excitatory synapses, $V_0 > \theta$, so that the driving force tends to depolarize (excite) the postsynaptic neuron when the (excitatory) presynaptic neuron emits a spike. In inhibitory synapses the reversal potential V_0 is below the reset potential, $V_0 < H$, and hence the driving force tends to hyperpolarize (inhibit) the postsynaptic neuron after receiving a presynaptic spike. The amplitudes of the EPSPs and IPSPs are given by the respective maximal conductances $g_{\max}$ (see Eq. (2.7)).

2.5.3 Connectivity structure

Recurrent and external connections

The synaptic current $I_{\text{syn}}(t)$ of a neuron contains contributions from all presynaptic cells. Given the modular structure of the cortical tissue, we can distinguish between recurrent and external currents according to the location of the presynaptic cell. Recurrent currents include all the currents generated by the spikes exchanged among cells in the same module, and can be either excitatory or inhibitory. External currents include the rest: non-specific background activity from cells not included in the module, as well as the coding of external stimuli provided by the projections from cells in distant brain areas. External contributions are only excitatory, as only excitatory cells have axons long enough to reach cells in other modules [29].

Fully and sparsely connected networks

Models of cortical networks mimic the connectivity patterns seen in cortical slices. They usually consist in a network of N LIF spiking neurons connected without any particular topological structure. The ratio of excitatory to inhibitory cells is 4:1, as observed experimentally. Synapses are of four generic types, as each of the presynaptic and postsynaptic cells can be either inhibitory or excitatory. We label synaptic quantities with $\alpha\beta$, where $\alpha, \beta = E, I$, to denote the connection

of a presynaptic cell of type β to a postsynaptic cell of type α. When there is no structure in the network, synaptic strengths J_{ij} depend only on the type of pre- and postsynaptic cells, namely $\{J_{\alpha\beta}\} = \{J_{EE}, J_{EI}, J_{II}, J_{IE}\}$. Structure is introduced in the network when synaptic strengths depend not only the type, but also on the identity of the cell. We describe structured networks in the next section.

How exactly these values for $J_{\alpha\beta}$ are chosen determines the connectivity of the network. In a fully connected network all synaptic strengths $J_{\alpha\beta}$ take nonzero values while in sparsely connected networks, the probability that two neurons are connected is lower than 1, and usually small. Synapses are in this case drawn at random, using either a *fixed* or a *variable* random connectivity scheme [7]. In a fixed scheme, every neuron of type α receives the same number of synaptic contacts from $c_{\alpha\beta}N_\beta$ cells selected at random. The fraction $c_{\alpha\beta} < 1$ is called the *sparseness level* and is usually independent of the type α, β, i.e., $c_{\alpha\beta} = c$, so that every cell is innervated by cN_E excitatory and cN_I cells chosen at random. In a variable random connectivity scheme, a post- and a presynaptic cell of type α and β are connected with probability $c_{\alpha\beta} < 1$. As a consequence, the number of connections that a neuron of type α receives from neurons of type β is no longer fixed, but is a binomially distributed random variable with $n = N_\beta$ and $p = c_{\alpha\beta}$. The mean and variance for this discrete random variable are therefore $c_{\alpha\beta}N_\beta$ and $c_{\alpha\beta}(1 - c_{\alpha\beta})N_\beta$, respectively.

Unless the model incorporates any type of activity-dependent synaptic plasticity (as in, e.g., [81], [66], or [15]), synaptic efficacies are fixed during the simulation. The connectivity is said to be *quenched*. The values for synaptic efficacies, when they are not zero, can be fixed or variable, in which case they are drawn from a Gaussian distribution of mean $\mu = J_{\alpha\beta}$ and variance $\sigma^2 = (J_{\alpha\beta}\Delta_{\alpha\beta})^2$. The probability density function $p(j)$ of synaptic efficacies can therefore be expressed as [130]:

$$p(j) = (1 - c_{\alpha\beta})\delta(j) + \frac{c_{\alpha\beta}}{J_{\alpha\beta}\Delta_{\alpha\beta}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{j - J_{\alpha\beta}}{J_{\alpha\beta}\Delta_{\alpha\beta}}\right)^2\right],$$

where $c_{\alpha\beta}$ is the sparseness level. Autapses, that is, synapses that a neuron makes onto itself, are excluded in the model in agreement with experimental observations. This amounts to set $J_{ii} = 0$ for all $i = 1 \ldots, N$.

Sparsely connected networks reflect more faithfully the real connectivity structure found in cortical networks, where the probability that two cells within a cortical column be connected is around 10% [128, 126, 174, 104]. However, simulating the dynamics of networks of sparsely connected cells is more computationally demanding than doing so with fully connected cells. The reason for that is that in sparsely connected networks, recurrent synaptic inputs differ from cell to cell and have to be computed individually, whereas in fully connected networks the recurrent synaptic input is the same for all cells. This may be an issue in large-scale models when synaptic currents are included in the description.

2.5.4 External inputs

The cortical column is an open system that interacts with other columns and brain areas. This interaction is mediated by excitatory projections from outside the column that provide external afferent currents. Estimates of the proportion of the excitatory synaptic inputs on a given neuron from external projections are in the range 50–80% (see Section 2.2). Hence a considerable part of the influx of spikes in the module comes actually from outside of it. Moreover, external currents are modeled as stochastic, as "an expression of our ignorance, not only of what the rest of the brain does, but also of the interactions of the brain with the outside world" [7]. External inputs can be either nonselective, reflecting the spontaneous activity of neurons outside the module, or selective, when they arise from the direct or indirect projections from sensory areas, which increase selectively their activity upon stimulus presentation.

Nonselective inputs

To model nonselective external currents, every neuron in the local network receives C_{ext} synaptic contacts from excitatory cells outside the module. These external cells emit spikes at rates similar, on average, to the spontaneous spike rate ν_0 of excitatory cells in the local network. The total influx of spikes $S_i^{\text{ext}}(t)$ converging from the external cells to cell i is modeled as a Poisson train (see, e.g., [58]) of rate $\lambda = C_{\text{ext}}\nu_0$:

$$S_i^{\text{ext}}(t) = \sum_k \delta(t - t_k)\,,$$

where t_k, $k = 1, 2, \ldots$, are the consecutive spike arrival times, which are Poisson random events. That is, the total number of spikes n that a cell i in the module receives from the external neurons during a period T is a random number $N_i(T)$, distributed according a Poisson distribution of mean $C_{\text{ext}}\nu_0 T$,

$$\Pr\big(N_i(T) = n\big) = \frac{(C_{\text{ext}}\nu_0 T)^n}{n!}\exp(-C_{\text{ext}}\nu_0 T)\,. \tag{2.11}$$

This random number $N_i(T)$ is drawn independently for each neuron in the local module. In simulations T is set to the timestep Δt, which is much smaller than the average ISI, $\Delta t/(C_{\text{ext}}\nu_0) \ll 1$. The probability of firing two or more spikes within Δt is then negligible, while the probability of firing one single spike is approximately given by $\Pr\big(N(\Delta t) = 1\big) \approx C_{\text{ext}}\nu_0\Delta t \ll 1$. Spike trains of rate $C_{\text{ext}}\nu_0$ can thus be generated by drawing at each timestep a random number x uniformly distributed over $[0, 1]$ and comparing it to $C_{\text{ext}}\nu_0 < 1$. A spike is emitted only if $x < C_{\text{ext}}\nu_0$.

Selective inputs

Selective external currents result from the presentation of a particular stimulus or from the selective activity in other modules. They are selective in the sense that only a fraction of cells in the module receive input upon presentation of the

stimulus. A set of p stimuli can be represented as N_E-bit words, $\xi_i^\mu = 0, 1$, where i runs over excitatory cells ($i = 1, \ldots, N_E$), μ over stimuli ($\mu = 1, \ldots, p$), and the 1-bits define the subpopulation of fN_E randomly chosen excitatory cells that are activated by the μ-th stimulus. The fraction $f < 1$ is the *coding level* of the stimulus and is assumed to be equal for all stimuli. Thus, external stimuli define p functional excitatory populations of fN_E neurons, each labeled by the external stimulus μ that activates it, and one population of $(1 - fp)N_E$ cells that do not respond to any stimulus [39].

When the stimulus μ is presented, the number of afferent spikes coming from outside the module increases for selective cells. This increase is parametrized by the factor $(1 + \tilde{\lambda})$, where the dimensionless parameter $\tilde{\lambda}$ is the selective *contrast* of the stimulus. In other words, during presentation of stimulus μ each cell i in the module receives at each timestep $N_i^{\rm st}(\Delta t)$ external spikes, where $N_i^{\rm st}(\Delta t)$ is a Poisson distributed random number with mean $C_{\rm ext}\nu_0 T(1 + \lambda\xi_i^\mu)$. The value for the selective contrast λ is typically low, around 0.1 or even less. This is because neurons are depolarized close to threshold during spontaneous activity, which make them very sensitive to stimuli with low contrast [183]. The synaptic efficacy of the external afferents is assumed to be the same, on average, as that of recurrent collaterals: $J_{\alpha,\rm ext} = J_{\alpha,E}$, where $\alpha = E, I$.

Overlapped representations of stimuli

To be consistent with experiments, where only a small fraction of cells respond to any stimulus, f is chosen very low, $f \ll 1$, and the subset of excitatory cells activated by the μ-th stimulus is chosen at random. As a first approximation, one can assume that for such low coding levels the probability that any neuron is stimulated by more than one stimulus is negligible, and therefore that populations of selective cells are disjoint. However, as also observed experimentally, there is a finite probability that any cell responds to more than one stimulus when selective cells are selected at random. If we denote by m the number of different stimuli to which a cell is responsive, the probability for any cell to have multiplicity higher than one is

$$\begin{aligned}\Pr(m > 1) &= 1 - \Pr(m = 0) - \Pr(m = 1) = 1 - \binom{p}{0}(1 - f)^p - \binom{p}{1} f(1 - f)^{p-1} \\ &= 1 - (1 - f)^p - pf(1 - f)^{p-1}\end{aligned}$$

Expanding the last expression for $\Pr(m > 1)$ in Taylor series about $f = 0$ leads to $\Pr(m > 1) = (pf)^2(1 - 1/p)/2 + O(f^3)$. The probability of having overlapping populations is thus negligible when $fp \ll 1$. Although it is possible to incorporate cells with multiplicity higher than one into network models [117, 57], we will neglect in the following the contribution of such cells and will assume that stimuli are represented in different, disjoint populations of cells. As we have just seen, this is a reasonable assumption as long as the coding level is low and the number of learned stimuli is not too high. On the other hand, the sparseness level f cannot be too low, as otherwise the selective activity state would be drowned by the noise coming from the spontaneous activity in background neurons [8].

2.5.5 Hebbian learning and structure in excitatory connections

According to the hypothesis of Hebbian plasticity, neural activity and training experience modify synaptic efficacies. Experimental observations actually show that neuronal activity affect synaptic strength through long-term potentiation (LTP, see e.g., [26, 25, 124]) and long-term depression (LTD, see e.g., [72, 47]). These findings have inspired models of unsupervised, Hebbian synaptic dynamics in attractor neural networks, which allow one to study quantitatively how learning protocols determine the statistical properties of the synaptic matrix [11, 6, 37]. Here we assume that the network has already been structured through repeated presentations of p different stimuli, taken in a random sequence [7]. The repeated presentation of stimuli induces the potentiation of synapses connecting neurons that are simultaneously activated by stimuli, while it weakens the synapses that connect cells with anticorrelated activities. Because external inputs project to excitatory cells only, and also for the sake of simplicity, plasticity is restricted to excitatory-to-excitatory synapses. Inhibitory synapses do not change, as the role of inhibition is to control the global activity of the network by responding proportionally to the mean firing activity of excitatory cells (shared inhibitory feedback).

The balance between synaptic potentiation and synaptic depression

The coexistence of potentiation and depression is essential to keep the network activity stable irrespective of changes in the synaptic matrix. In fact, although the nature of the regulatory mechanisms that contribute to maintain the overall activity is still subject of debate [for reviews, see [2], and [186]], it is clear that any homeostatic mechanism must rely on the interplay of both positive (LTP) and negative (LTD) feedback. In the network model we use, the regulatory mechanism is invoked *ad hoc*, requiring that the average excitatory-to-excitatory synaptic strength remains constant during learning [179, 11, 8]. To be more precise, if synapses between cells that belong to the same selective population are potentiated by a factor $w_+ > 1$ (i.e., have a strength $J_+ \equiv J_{EE} w_+$), synapses between cells belonging to different selective populations as well as synapses connecting nonselective to selective cells are depressed by an amount $w_- < 1$ related to w_+ (i.e., $J_- \equiv J_{EE} w_- < J_{EE}$) [39]. The remaining synapses do not change: synapses connecting excitatory cells to nonselective cells, as well as synapses involving inhibitory cells keep the same value as before learning. The structure of the synaptic matrix is depicted in Figure 2.7.

The exact dependence of w_- on w_+ is derived from the homeostatic condition. If we denoting by $\langle J_{EE} \rangle_0$ and $\langle J_{EE} \rangle_{\text{learn}}$ the average synaptic strength before and after learning, respectively, the condition reads

$$\langle J_{EE} \rangle_0 = J_{EE} \stackrel{!}{=} \langle J_{EE} \rangle_{\text{learn}} = J_{EE} \left(p f^2 w_+ + p f (1-f) w_- + (1 - pf) \right) . \quad (2.12)$$

The first term on the right hand side of the equation accounts for the synapses connecting cells in the same selective population; the second, for the synapses connecting excitatory cells *to* selective cells, and the third, for synapses connecting

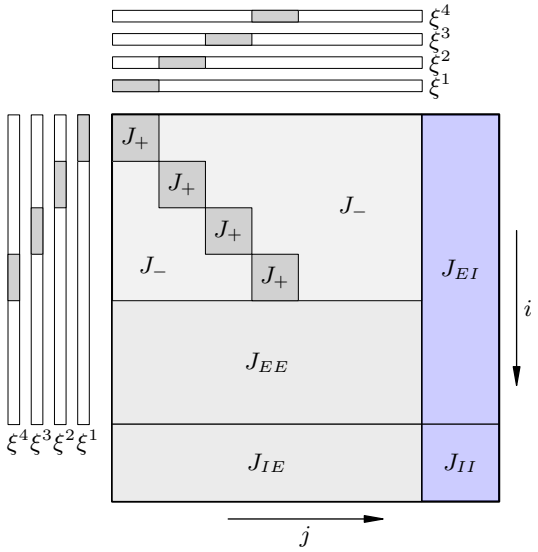


Figure 2.7: Representation of the synaptic matrix J_{ij} after learning. Indices $i, j = 1, \ldots, (N_E + N_I)$ specify post and presynaptic cells, respectively. Non-overlapping subpopulations of excitatory cells respond to $p = 4$ different stimuli $\xi^1, \ldots, \xi^4$ (shaded areas in upper and left boxes). Before learning the connection weights between excitatory cells are all the same, with value J_{EE}. As a result of Hebbian learning, the connections between cells selective to the same stimulus are potentiated with respect to the baseline J_{EE} ($J_+ < J_{EE}$), while connections between cells selective to different stimuli are depressed with respect to J_{EE} ($J_- < J_{EE}$). Synapses connecting a nonselective cell to a selective one are also depressed to the same strength J_-. The strengths of all excitatory projections to nonselective cells do not change, nor do connections involving inhibitory cells. The overall change in the synaptic strengths is zero when J_- is $J_{EE}(1 - fJ_+/J_{EE})/(1 - f)$. See text for details. Adapted from [130].

excitatory cells *to* nonselective cells (see also the synaptic matrix in Figure 2.7). Condition (2.12) implies

$$J_- = J_{EE} w_- = J_{EE} \frac{1 - f w_+}{1 - f} = J_{EE} \left(1 - f \frac{J_+/J_{EE} - 1}{1 - f} \right) \tag{2.13}$$

This balance condition depends exclusively on the synaptic matrix. In this case we have assumed that synapses from selective to nonselective cells do not change during learning. Another plausible choice would consist of taking these synapses as depressed, in the same way as synapses from nonselective to selective neurons are [8]. The balance condition would then read

$$J_{EE} \left[pf^2 w_+ + pf(1 - f) w_- + (1 - pf) pf w_- + (1 - pf)^2 \right] = J_{EE} \tag{2.14}$$

where the second term accounts for the projections from selective to excitatory cells, the third for the projections from nonselective to excitatory cells, and the

last for connections among nonselective cells. Equation (2.14) leads then to $w_- = (2 - f[p + w_+])/(2 - f[p + 1]) < 1$.

The structure in the matrix of synaptic efficacies is not necessarily associated with a topological or spatial structure in the real network, as cells that are selective to a particular stimulus are not necessarily close together. The structure induced by Hebbian learning depends on the correlation between pre- and postsynaptic firing. It is determined therefore by the selectivity properties of cells, regardless of their actual distance. However, it is plausible that the connectivity structure also reflects the fact that the distance between cells selective to the same stimulus is, on average, shorter than between cells selective to different objects, and hence that the connection probability cells selective to the same stimulus is higher [43]. The situation in the real cortex might be mixture of these two, with cells selective to the same stimulus tending to cluster in space, and the connectivity matrix shaped by the learning process.

2.5.6 Simulations

Once we know the models for the neurons, the synapses, and the connectivity structure, the dynamics of the network of spiking neurons are completely determined at a microscopic level. Computer simulations can be carried out to provide observables like those typically obtained from electrophysiological experiments. It is possible, for instance, to register spikes times and depolarization traces for samples of neurons, or to monitor population-averaged or spike-count firing rates.

We will use two types of network in this work. The first is the network model introduced by [8, 7], and consists of a network sparsely connected LIF neurons, with synaptic currents given by a linear sum of the presynaptic spikes, and synaptic matrix structured through Hebbian learning. Appendix B contains the parameters used in this work. The second model was proposed by [43], and is an extension of the first. It incorporates synaptic dynamics for the three main receptors involved in ionotropic synaptic transmission, namely GABA_A, NMDA, and AMPA-mediated currents. Since first network model can be regarded as a particular limit case of the second, we will describe in the following only Brunel & Wang model.

In the fully connected network with realistic currents introduced by [43], recurrent excitatory connections are mediated by both AMPA and NMDA receptors. Reverbatory excitations are balanced by the inhibitory feedback mediated by GABA_A interneurons, which presumably controls the excitability of the network. External currents are mediated by AMPA receptors only. The whole system is described by the set of coupled equations consisting in the RC-equation for the membrane potential of every neuron, Eq.(2.1), together with the synaptic input to every cell i, which depends in turn on the spiking activity of the rest of the network. The synaptic input is given by the sum of the glutamatergic and GABAergic components $I_i^{\text{syn}}(t) = I_i^{\text{AMPA ext}}(t) + I_i^{\text{AMPA rec}}(t) + I_i^{\text{NMDA}}(t) + I_i^{\text{GABA}}(t)$, where (see 2.5.2)

$$I_i^{\text{AMPA ext}}(t) = g_{\text{AMPA ext}}\big(V_i(t) - V_E\big) \sum_{j=1}^{N_{\text{ext}}} s_j^{\text{AMPA ext}}(t)\,, \tag{2.15a}$$

$$I_i^{\text{AMPA rec}}(t) = g_{\text{AMPA rec}}\big(V_i(t) - V_E\big) \sum_{j=1}^{N_E} w_j s_j^{\text{AMPA rec}}(t)\,, \tag{2.15b}$$

$$I_i^{\text{NMDA}}(t) = \frac{g_{\text{NMDA}}\big(V_i(t) - V_E\big)}{1 + \gamma \exp\big(-\beta V_i(t)\big)} \sum_{j=1}^{N_E} w_j s_j^{\text{NMDA}}(t)\,, \tag{2.15c}$$

$$I_i^{\text{GABA}}(t) = g_{\text{GABA}}\big(V_i(t) - V_I\big) \sum_{j=1}^{N_I} s_j^{\text{GABA}}(t)\,, \tag{2.15d}$$

and where the sums are over the synapses formed by presynaptic neurons j. Note that recurrent synaptic currents have two components, $I_i^{\text{AMPA rec}}(t)$ and $I_i^{\text{NMDA}}(t)$, and are modulated by the dimensionless factor w_j. Note also that the recurrent currents are identical for all neurons in the network, as they all involve a sum over all the excitatory or inhibitory cells in the network due to the all-to-all connectivity. The dynamics of the gating variables reproduce the time course of the PSC for each component, and are described by

$$\dot{s}_j^{\text{AMPA ext}}(t) = -s_j^{\text{AMPA ext}}(t)/\tau_{\text{AMPA}} + \sum_k \delta(t - t_j^{(k)})\,, \tag{2.16a}$$

$$\dot{s}_j^{\text{AMPA rec}}(t) = -s_j^{\text{AMPA rec}}(t)/\tau_{\text{AMPA}} + \sum_k \delta(t - t_j^{(k)})\,, \tag{2.16b}$$

$$\dot{s}_j^{\text{GABA}}(t) = -s_j^{\text{GABA}}(t)/\tau_{\text{GABA}} + \sum_k \delta(t - t_j^{(k)})\,, \tag{2.16c}$$

$$\dot{s}_j^{\text{NMDA}}(t) = -s_j^{\text{NMDA}}(t)/\tau_{\text{NMDA},\downarrow} + \alpha x_j(t)\big(1 - s_j^{\text{NMDA}}(t)\big)\,, \tag{2.16d}$$

$$\dot{x}_j(t) = -x_j(t)/\tau_{\text{NMDA},\uparrow} + \tau_\uparrow \sum_k \delta(t - t_j^{(k)})\,. \tag{2.16e}$$

The spike times $t_j^{(k)}$ in the equations for recurrent synaptic variables, Eqs. (2.16b), (2.16c), and (2.16e), are generated by the network itself through the threshold condition imposed on the N equations (2.1). Spike times from external connections, Eq. (2.16a), are, on the contrary, randomly generated as described in 2.5.4, and constitute the only source of *fast* noise in the system. Without this contribution, the system would be deterministic, so that it would be possible to know the values of all variables at any time, given some initial conditions. A different source of disorder is present when the connectivity matrix is sparse and random and synaptic efficacies and/or transmission delays are drawn from some probability distribution. This type of disorder is fixed or changes on time scales much longer than neuronal and synaptic time constants, so that can be considered effectively *quenched* [7].

The state of the system is specified by the value of the $6N$ dynamical variables present: N of each of $s^{\text{AMPA ext}}$, $s^{\text{AMPA rec}}$, s^{GABA}, s^{NMDA}, and x. Strictly speaking, the system (2.1),(2.15), (2.16) is a hybrid system: the state of a neuron evolves continuously according to some (smooth) differential equation, and incoming spikes trigger sudden changes in some of the state variables. The evolution of the system is thus described as a mixture of ordinary differential equations and

maps [31]. Maps are not explicit in Eqs. (2.1),(2.15), (2.16); we use instead the δ distribution to describe, keeping the formalism of differential equations, the sudden changes in the state variables induced by the arrival of a spike. In addition, there are threshold conditions on the state variables, and there is noise coming from synaptic inputs. Bearing these caveats in mind, we can loosely represent the evolution of the system as a set of differential equations

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}; t, \omega)\,, \tag{2.17}$$

where $\boldsymbol{x}$ is a vector whose components are the state variables of the system

$$\begin{aligned}\boldsymbol{x} &= \big(\{V\}, \{s^{\text{AMPA ext}}\}, \{s^{\text{AMPA rec}}\}, \{s^{\text{GABA}}\}, \{s^{\text{NMDA}}\}, \{x\}\big)\\ &= (V_1, V_2, \ldots, V_N, s_1^{\text{AMPA ext}}, \ldots, x_N)\,,\end{aligned}$$

and where the functions $\boldsymbol{F}$ are, up to a constant multiplicative factor, the right hand sides of the differential equations for the membrane potential and synaptic inputs, Eqs.(2.1) and (2.15). The symbol ω denotes a particular realization of the stochastic process, specified by the particular sequence of spike times coming from external afferents, Eq. (2.16a). In numerical simulations of the network, this sequence is completely determined by the seed we use to initialize the random number generator.

The numerical integration of (2.17), given the initial condition $\boldsymbol{x}(t=0) = \boldsymbol{x}_0$ and some particular realization ω is done as usual by discretizing time, $t_n = n\Delta t$, $n \in \mathbb{N}$, and applying Euler method

$$\boldsymbol{x}(t_{n+1}) \equiv \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta t\, \boldsymbol{F}\big(\boldsymbol{x}_n; t_n, \omega\big)\,.$$

That is,

$$V_i(t_{n+1}) = V_i(t_n) + \frac{\Delta t}{C_m}\Big[-g_L\big(V_i(t_n) - V_L\big) - I_i^{\text{syn}}(t_n)\Big],\ i = 1, \ldots, N$$

$$s_j^{\text{AMPA ext}}(t_{n+1}) = \begin{cases} s_j^{\text{AMPA ext}}(t_n) + 1 & \text{upon spike from ext. synapse } j \\ s_j^{\text{AMPA ext}}(t_n)(1 - \Delta t/\tau_{\text{AMPA}}) & \text{otherwise}\,, \end{cases}$$

and so on for the other state variables. In our simulations we have used the 2nd order Runge-Kutta scheme for the integration [see, e.g., 100]

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \Delta t\, \boldsymbol{F}\bigg(\boldsymbol{x}_n + \frac{\Delta t}{2}\, \boldsymbol{F}\big(\boldsymbol{x}_n; t_n, \omega\big); t_n + \frac{\Delta t}{2}, \omega\bigg).$$

Typically, all the state variables are updated simultaneously at every time step. The update is then called *synchronous* or *clock-driven*. This is the method commonly used for network models that incorporate synaptic dynamics, such as [43]. Another strategy takes advantage of the fact that the state variables evolve deterministically between any two consecutive spikes, and update the state of a neuron only when it receives or emits a spike. This is the approach of *event-driven* or asynchronous algorithms [131]. This type of update is suited for network models with instantaneous synapses, like that presented in [8].

After setting the initial values for the state variables, one can simulate the temporal evolution of all the variables of the system and extract from them potentially relevant observables. For example, one can monitor the membrane depolarization $V_i(t)$, or the spike times of an arbitrary subset of neurons. It is useful to represent the sequence of spikes emitted by the i-th neuron with its neural response function

$$\rho_i(t) \equiv \sum_k \delta(t - t_i^{(k)}) \,, \tag{2.18}$$

where $t_i^{(k)}$, $k = 1, 2, \ldots$ are the consecutive spike times [58]. The number of spikes emitted between times t_{n-1} and t_n in some arbitrary population x can also be tracked down,

$$n_x(t_n) = \sum_{i \in x}^{N_x} \int_{t_{n-1}}^{t_n} \rho_i(\tau) \, \mathrm{d}\tau \,,$$

where $\rho_i(t)$, is the neural response function of cell i, Eq. (2.18). From n_x we can obtain the population averages of the rates, defined as

$$\nu_x(t_n) = \frac{1}{N_x T \Delta t} \sum_{m=0}^{T-1} n_x(t_{n-m}) \,,$$

where T is the length of the time window, in units of Δt, over which we average the rates. Figure 2.8 illustrates the time course of some of these observables, as obtained from the cortical network model by [43].

Up to now we have not dealt with the problem of choosing the values for the large set of parameters present in the model. Many of the parameter values can be, and actually are, constrained to fall within experimental ranges. But even with the parameters constrained within plausible ranges, there is a rich variety of network states, most of which are not physiologically relevant. Although it is possible to use simulations to find, if it exists, the region of the parameter space where the activity of the network reproduces well experimental data, the computational cost associated to such scan is considerable and, more critically, the insight we gain into the dynamics of the network is rather limited. In the hypothetical case that we could narrow down the region of the parameter space indefinitely, reducing it to a point, and if we were lucky enough to have come up with a good model that reproduces experimental data, there would remain the essential question of what mechanisms give rise to the phenomena observed. This motivates the use of a description based on global, or collective variables, which we introduce in the next chapter.

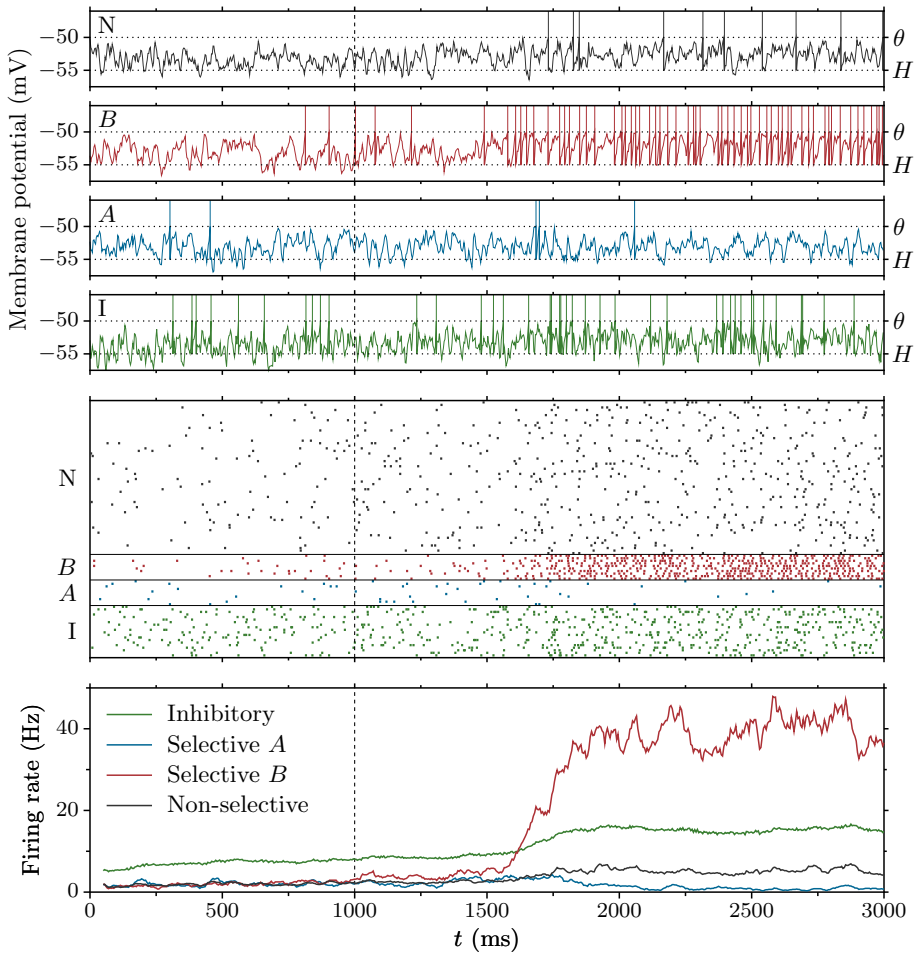


Figure 2.8: Simulation of a network of 2000 LIF neurons (1600 excitatory, 400 inhibitory), with synaptic currents mediated by AMPA, GABA, and NMDA receptors. Excitatory cells are subdivided into three populations: $p = 2$ selective (called A and B), and one nonselective (N). The population of inhibitory cells (I) has no structure. The neuronal and synaptic parameters are in Table A.1, p. 119. The synaptic matrix is that shown in Fig. 2.7, with $w_+ = 1.75$ and coding level $f = 0.15$. The average spike rate of external afferents is $\nu_{\text{ext}} = C_{\text{ext}}\nu_0 = 2400$ Hz for all cells and throughout the simulation, except for cells in A and B, which during the interval from 1000 to 3000 ms receive external spike trains at higher rate 2415 Hz (selective contrast 0.625%). Top panel: traces of the membrane potential of four cells belonging to the different functional subpopulations. Mid panel: rastergram showing the spike times of 60 nonselective cells, 20 selective cells (10 per stimulus), and 20 inhibitory cells. The depolarization traces shown in the top panel correspond in the rastergram to the first cell (starting from below) of each subpopulation. Bottom panel: population activity each of these four populations, in bins of 50 ms slided every 5 ms.

Chapter 3

The mean field approximation

The cortical network models presented in the previous chapter are valuable tools that help us gain insight into the dynamics of real cortical networks. However, both the number of parameters and the dimensionality of these models are large, rendering their analysis difficult. This problem motivates the need for a lower-dimensional description of large-scale networks, based on the use of a few global, collective variables. This is the essence of the so-called mean field approximation, with which it is possible to identify the stationary states of the network and analyze their stability as a function of the microscopic parameters of single-cells and synapses.

Several references have been found particularly useful for the material of this chapter. The articles [8, 7] are very readable and cover the first part of the chapter, devoted to the mean field formulation in networks with instantaneous synapses. For the second part, where realistic synaptic currents are incorporated in the mean field framework, the articles [43] and [79] are valuable references. The reviews [40] and [159] are also very comprehensible references.

3.1 Input-output transformation of single LIF neurons

3.1.1 Collective variables

Recordings in awake animals show that cortical cells emit action potentials at rates in the range 0–200 spikes s^{-1}, with very irregular spike trains. This irregularity can be characterized with the statistics of the time elapsed between consecutive action potentials, known as interspike interval (ISI). If the event 'emitting a spike' at a given time t were a pure Poissonian process, the ISI would be exponentially distributed. In this case the coefficient of variation (CV), defined as the standard deviation of the ISI divided by its mean, would be exactly 1. The CV measured in cortical cells is usually around 1 in all cortical areas [175], not only when cells discharge during background activity states but also during delay activity states [52]. The firing activity of cortical cells can thus be described as approximately Poissonian.

Moreover, as mentioned in Chapter 2, cortical columns are composed of functionally similar, highly interconnected neurons, with low probability of connection between pairs of cells. The inputs received by the neurons within a column are therefore expected to be similar, but weakly correlated. This implies that any postsynaptic cell can decode the activity of the neuronal population just by summing the thousands of weakly-correlated irregular spike trains, as fluctuations in the individual activities will average out when summed together. In other words, we can thus regard the set of columnar neurons as an ensemble of units undergoing independent realizations of the same random process.

This suggest a description of the network activity based on the mean rates of statistically homogeneous populations of neurons. By homogeneous we mean that the statistical properties of the currents and the connection strengths are identical for all neurons in the same population. We then make the assumption that all neurons in the same subpopulation emit spikes at the same rate. With this assumption, the statistics of the currents become homogeneous within each subpopulation, and we can determine the average emission rate in the subpopulation. The steady mean emission rate in each subpopulation is obtained by closing the loop: the output rates, which depend on the input currents via the activation function, must be equal to the input rates, which contribute to the input currents.

3.1.2 Statistical properties of the input currents

To formulate the mean-field description, we should first know the response properties of single cells. More specifically, we want to determine the average firing rate of a neuron as a function of its input currents. This is done in several steps. First we characterize the inputs that a single cell receives by the sole fact of being part of a network. For this, we make some assumptions about the connectivity and the firing activity of the neighboring cells. Once the inputs are specified, we are left with the problem of finding the statistics of the firing activity, such as the mean firing rate or the coefficient of variation of ISI intervals, of a LIF neuron.

We start by considering a population of excitatory and inhibitory neurons. As described in the previous chapter, every cell receives recurrent inputs from collaterals, and external inputs from regions outside the local module. We neglect for the moment any synaptic dynamics: synaptic inputs are modeled as a sum of delta pulses (section 2.5.2). Given the irregularity of spike emission seen in cortical cells, it is reasonable to assume that both excitatory and inhibitory cells emit spikes according to independent Poisson processes of rates $\nu_E(t)$ and $\nu_I(t)$.

The rates of excitatory and inhibitory cells are equal for all cells of the same type, and possibly time-dependent. External cells also emit spikes according to independent Poisson processes, at a constant rate ν_{ext}. Every cell in the network receives from recurrent connections C_E excitatory synaptic contacts of strength J_E, and C_I inhibitory synaptic contacts of strength $-J_I$, $J_I > 0$. It also receives C_{ext} excitatory contacts of strength J_E from outside. Synaptic strengths J_E and J_I are expressed in units of electric charge. In such network, the synaptic input

to any neuron is

$$I(t) = I^{\text{rec}}(t) + I^{\text{ext}}(t) = J_E \sum_{j=1}^{C_E} \rho_j(t-d) + J_I \sum_{j=1}^{C_I} \rho_j(t-d) + J_E \sum_{j=1}^{C_{\text{ext}}} \rho_j(t) \quad (3.1)$$

where $\rho_j(t) = \sum_k \delta(t - t_j^{(k)})$ is the neural response function of cell j, and d is the transmission delay in the recurrent connections, which we assume identical for all cells. The mean synaptic input of any neuron is obtained computing the mean rate of charge injection

$$\langle I(t) \rangle = J_E \big[C_E \nu_E(t-d) + C_{\text{ext}} \nu_{\text{ext}} \big] - J_I C_I \nu_I(t-d) \equiv \tilde{\mu}(t)\,, \quad (3.2)$$

where the angular brackets $\langle \cdot \rangle$ denote the average over realizations. Information about second order statistics is given by the two-point correlation function,

$$\begin{aligned} \langle\langle I(t) I(t') \rangle\rangle &\equiv \langle I(t) I(t') \rangle - \langle I(t) \rangle \langle I(t') \rangle \\ &= \big\{ J_E^2 \big[C_E \nu_E(t-d) + C_{\text{ext}} \nu_{\text{ext}} \big] + J_I^2 C_I \nu_I(t-d) \big\} \delta(t-t') \quad (3.3) \\ &\equiv \tilde{\sigma}^2(t) \delta(t-t')\,. \end{aligned}$$

Note that $\tilde{\mu}(t)$ has units of current and that $\tilde{\sigma}^2(t)$ has units of square of the current times time ($\text{A}^2 \cdot \text{s}$). As we did in the previous chapter, we will set the transmission delay d to zero.

3.1.3 Depolarization as a random walk

The input (3.1) is delta correlated, meaning that fluctuations in the current are independent at any time scale. The question now is, what is the firing rate of a LIF neuron that receives the noisy synaptic input (3.1), with mean $\tilde{\mu}(t)$ and variance $\tilde{\sigma}^2(t)$? The problem is a genuine example of a first-passage time problem, in which one has to find the average time it takes for a random variable to reach a given value. The mean rate is computed from the mean interspike interval, which is the average time spent for the voltage to reach the threshold from the reset potential, taking into account the leakage and the stochastic forcing due to the inputs. The membrane voltage obeys the LIF equation (2.1), which can be recast as the simpler expression

$$\tau \dot{V}(t) = -V(t) + I(t)\,, \quad (3.4)$$

by shifting V to $V + V_L$, defining τ as C_m/g_L, and absorbing g_L in the current $I(t)$, which is now expressed in voltage units. Doing so, the PSPs J are expressed in units of voltage per time. For J to be expressed in units of voltage only, we rewrite J as $J\tau$.

The membrane voltage is a function of a random variable, the synaptic current, and it is hence a random variable itself. Figure 3.1 illustrates a sample path of the membrane voltage in response to a particular realization of the stochastic input (3.1). As a random variable, the dynamics of the membrane potential should

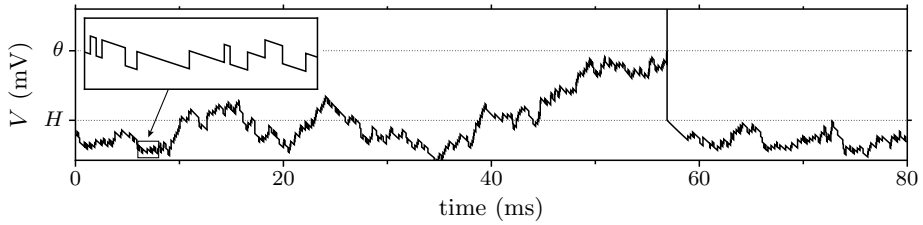


Figure 3.1: Depolarization of a LIF neuron in response to a stochastic point-process input. The input is a superposition of two spike trains, one with rate 5 kHz and PSP amplitude $J = 0.3$ mV, and another with rate 2 kHz and negative PSP amplitude $J = -0.3$ mV. The inset is a snapshot of the depolarization trace in the time interval from 6 to 8 ms. Parameters of the LIF neuron: $\tau = 20$ ms, $\theta = 20$ mV, $H = 15$ mV, $V_L = 0$. The refractory period is $\tau_{\rm rp} = 2$ ms.

be described in terms of the conditional probability density for $V(t) = v$ given that $V(t_0) = v_0$, which we denote by $P(v, t|v_0, t_0) \equiv Q(v, t)$. Because the input is delta correlated, the precise value of the input at any time does not convey any information about the values at later times. This is equivalent to saying that the conditional probability density $P(v, t|v_0, t_0)$ is not affected by any knowledge of the values of V at times earlier than t_0, and hence that $V(t)$ is a Markov process.

The conditional probability $Q(v, t)$ of a stationary Markov process must obey the so called differential form of the Chapman-Kolmogorov equation, also known as master equation (see, e.g., [187, 55, 84])

$$\frac{\partial}{\partial t}Q(v,t) = \int \big\{W(v|v')Q(v',t) - W(v'|v)Q(v,t)\big\}\,\mathrm{d}v'\,,$$

where $W(v'|v), W(v|v') \geq 0$ are the transition probabilities per unit time from v to v', and from v' to v, respectively. The master equation for the membrane voltage of a LIF neuron is

$$\begin{aligned}\tau\frac{\partial}{\partial t}Q(v,t) = \frac{\partial}{\partial v}\big[vQ(v,t)\big] \\ + \big[C_E\nu_E(t) + C_{\rm ext}\nu_{\rm ext}\big]\tau\big[Q(v-J_E,t) - Q(v,t)\big] \\ + C_I\nu_I(t)\tau\big[Q(v+J_I,t) - Q(v,t)\big]\,. \quad (3.5)\end{aligned}$$

The first term describes the change in $Q(v, t)$ due to the deterministic decay of the membrane potential, while the remaining terms account for the jumps in V provoked by the instantaneous injection of positive and negative charge due to synaptic inputs. More precisely, the terms in the second and third lines containing $Q(v - J_E, t)$ and $Q(v + J_I, t)$ represent the 'gain' of probability in $V = v$ due to transitions from $v - J_E$ and $v + J_I$, respectively. The accompanying negative terms, which contain $Q(v, t)$, represent the loss in $V = v$ due to transitions from v to $v + J_E$ or to $v - J_I$ caused by the arrival of an excitatory or inhibitory spike.

3.1.4 The diffusion approximation

Equation (3.5) is difficult to handle because it is non local. We can convert the equation to a more tractable partial differential equation using the diffusion approximation [see, in the context of neuronal activity, 184, 16]. The idea is to convert the original point process $I(t)$ to a continuous random process $\bar{I}(t)$ with the same mean $\langle I \rangle$ and two-point correlation function $\langle\langle I(t)I(t') \rangle\rangle$, so that the response of the potential to $\bar{I}(t)$ is continuous instead of jumpy, while the statistics of postsynaptic firing remain the same. This is achieved by taking the limits of vanishing PSPs and infinitely large number of connections. Given that every cortical neuron receives around 10^4 presynaptic contacts, each of which contribute with a small fraction of the total excursion from the reset to the threshold, the approximation is plausible. Moreover, since cortical cells emit at a few spikes per second, the number of spikes received by a cortical cell during the integration time scale is large, so that the evolution of the membrane voltage actually looks stochastic even at very short time scales.

The diffusion approximation is applied expanding the right hand side of (3.5) in a power series of Js, neglecting the terms of order higher than 2 (see [187] for a discussion about the validity of this truncation, and for a systematic account on power expansions of the master equation). This gives

$$\tau \frac{\partial}{\partial t} Q(v,t) = -\frac{\partial}{\partial v}\big[(\mu(t) - v)Q(v,t)\big] + \frac{\sigma^2(t)}{2}\frac{\partial^2}{\partial v^2} Q(v,t)\,, \tag{3.6}$$

where

$$\mu(t) = \tau\big\{J_E\big[C_E\nu_E(t) + C_{\text{ext}}\nu_{\text{ext}}\big] - J_I C_I \nu_I(t)\big\}\,, \tag{3.7a}$$

$$\sigma^2(t) = \tau\big\{J_E^2\big[C_E\nu_E(t) + C_{\text{ext}}\nu_{\text{ext}}\big] + J_I^2 C_I \nu_I(t)\big\}\,, \tag{3.7b}$$

are the mean and the variance of the current. Both $\mu(t)$ and the standard deviation $\sigma(t)$ are in voltage units.

Equation (3.6) is the Fokker-Planck equation describing the temporal evolution of the conditional probability density, $Q(v,t) = P(v,t|v_0,t_0)$, of a LIF neuron that receives a stochastic input $\bar{I}(t)$ composed of a pure deterministic component, $\mu(t)$, plus a pure stochastic component, a source of white noise with variance $\sigma^2(t)$. The system described by the Fokker-Planck equation (3.6), is mathematically equivalent to that described by the following differential equation

$$\tau \dot{V}(t) = -V(t) + \bar{I}(t)\,, \tag{3.8a}$$

where

$$\bar{I}(t) \approx \mu(t) + \sigma(t)\sqrt{\tau}\eta(t). \tag{3.8b}$$

The symbol $\eta(t)$ denotes a Gaussian random variable obeying

$$\langle \eta(t) \rangle = 0\,, \tag{3.9a}$$

$$\langle \eta(t)\eta(t') \rangle = \delta(t - t')\,. \tag{3.9b}$$

Equations (3.8) together with conditions (3.9) are the *Langevin description* of the stochastic system.

3.1.5 Mean firing rate of a LIF neuron

Both equation (3.6) and Eq. (3.8) describe the random motion of v under the influence of a stochastic perturbation of amplitude $\sigma(t)$ and a harmonic potential $U(v) = [v-\mu(t)]^2$ —an Ornstein-Uhlenbeck process [84]. The potential falls sharply at the firing threshold to represent an absorbing barrier at $\nu = \theta$. The potential is depicted in Fig. 3.2A (top). When the input has no stochastic component, that is, when $\sigma^2(t) = 0$, the membrane potential v evolves deterministically to the minimum of the potential μ. If $\mu < \theta$, the membrane potential cannot cross the threshold and the firing rate of the neuron is 0. The neuron can only fire if the mean intensity of the current is higher than the threshold, in which case the neuron fires regularly with a fixed ISI given by the refractory period plus the precise time it takes to decay from the reset potential to the threshold. For finite noise, $\sigma^2(t) > 0$, the neuron is able to fire even when $\mu < \theta$ because fluctuations can drive the membrane potential toward the threshold, surmounting the potential barrier $U(\theta) - U(\mu)$.

To calculate the firing rate, we need to determine the average time between consecutive spikes. We will also be interested in higher order statistics of the interspike interval, like the coefficient of variation. The ISI is a random variable given by the sum of the refractory period and the random time it takes for the voltage to reach the threshold θ, starting from the reset potential H. These quantities can be computed using the theory of first-passage times of an Ornstein-Uhlenbeck process. The mean firing rate is given by the inverse of the mean ISI, which can be computed with the following trick. The Fokker-Planck equation (3.6) can be rewritten as a continuity equation:

$$\frac{\partial}{\partial t}Q(v,t) = -\frac{\partial}{\partial v}S(v,t)\,,$$

where

$$S(v,t) = \frac{\mu(t)-v}{\tau}Q(v,t) - \frac{\sigma^2(t)}{2\tau}\frac{\partial}{\partial v}Q(v,t).$$

The quantity $S(v,t)$ is the probability current at $V = v$ at time t. The instantaneous firing rate, $\nu(t)$, of a neuron receiving the stochastic input $\bar{I}(t)$ is equal to the probability current at threshold

$$S(\theta,t) = -\left.\frac{\sigma^2(t)}{2\tau}\frac{\partial}{\partial v}Q(v,t)\right|_{v=\theta} = \nu(t)\,, \tag{3.10a}$$

where we have used the condition that $v = \theta$ is an absorbing barrier, $Q(\theta,t) = 0$. As it stands, Eq. (3.10a) is just a definition. For it to be useful we have to solve the partial differential equation (3.6), subject to a set of boundary conditions, particular to the LIF problem. First, since the membrane potential is reset, after a refractory period $\tau_{\rm rp}$, to $V = H$ whenever the potential crosses the threshold, the probability flux leaving the interval $(-\infty,\theta)$ through θ must be reinjected at the reset potential, after a refractory period $\tau_{\rm rp}$. As a consequence, the probability current has a jump discontinuity at $V = H$, with size equal to the instantaneous

firing rate at time $t - \tau_{\rm rp}$,

$$S(H^+, t) - S(H^-, t) = \nu(t) = -\frac{\sigma^2(t - \tau_{\rm rp})}{2\tau}\,, \tag{3.10b}$$

where we have used in the first equality the fact that $Q(v, t)$ is continuous (and hence $Q(H^+, t) = Q(H^-, t)$), and (3.10a) in the second. Since $Q(v, t)$ is, after all, a probability density function, it has decay sufficiently fast for $v \to -\infty$ so that it can be integrated. This is guaranteed imposing the limits

$$\lim_{v\to-\infty} Q(v, t) = 0, \qquad \lim_{v\to-\infty} vQ(v, t) = 0\,. \tag{3.10c}$$

In addition, $Q(v, t)$ has to be normalized. We have to take into account the probability $p_{\rm rp}(t)$ that the neuron is refractory at time t, and is hence missing in the sample space $(0, \theta)$. This probability is

$$p_{\rm rp}(t) = \int_{t-\tau_{\rm rp}}^{t} \nu(s)\,{\rm d}s\,.$$

The normalization condition reads then

$$\int_{-\infty}^{\theta} Q(v, t)\,{\rm d}v + p_{\rm rp}(t) = 1\,. \tag{3.10d}$$

The instantaneous mean firing rate $\nu(t)$ is obtained by solving (3.6) with the boundary conditions (3.10a)–(3.10d). Its value depends on the parameters of the input $\bar{I}(t)$, such as the number of connections or the PSP amplitudes (see Eqs. (3.7)), as well as on the parameters of the single-cell. We will be particularly concerned with the firing rate ν of a neuron under stationary conditions, that is, when $\mu(t) = \mu$, $\sigma^2(t) = \sigma^2$. In that case, the probability density function for $V = v$ is found from Eq. (3.6) by setting the time derivative to 0 and solving the resulting ordinary differential equation, with the boundary and continuity conditions enumerated above. The solution is

$$Q(v) = \frac{2\nu\tau}{\sigma} \exp\left(-\frac{(v-\mu)^2}{\sigma^2}\right) \int_{(v-\mu)/\sigma}^{(\theta-\mu)/\sigma} \Theta(u - H) \exp(u^2)\,{\rm d}u\,, \tag{3.11}$$

where $\Theta(x)$ is the step function, $\Theta(x) = 1$ for $x > 0$ and 0 otherwise. Figure 3.2A (bottom) shows the stationary solution of the Fokker-Planck equation (3.6), $Q(v)$. The firing rate is obtained from the normalization condition (3.10d), and is given by [161]

$$\nu = \frac{1}{\langle {\rm ISI}\rangle} \equiv \Phi(\mu, \sigma) = \left(\tau_{\rm rp} + \sqrt{\pi}\tau \int_{(H-\mu)/\sigma}^{(\theta-\mu)/\sigma} \exp\big(u^2\big)\big[1 + {\rm erf}(u)\big]\,{\rm d}u,\right)^{-1}, \tag{3.12}$$

where we have introduced the transduction function $\Phi(\mu, \sigma)$, which transforms the input current to a cell to output firing rate, and where we have used the definition

for the error function $\mathrm{erf}(x) = (2/\sqrt{\pi}) \int_0^x \exp(-u^2)\,\mathrm{d}u$. The transduction function is depicted in Fig. 3.2B for different noise amplitudes in the input. Note that noise smooths out the f-I curve given by $\Phi(\mu, \sigma)$, allowing the neuron to fire when the average membrane voltage is below the deterministic threshold —the subtreshold regime. Notice also that the firing rate increases supralinearly (is concave) for sub-threshold mean currents μ, while increases sublinearly (is convex) for supra-threshold currents. The change of curvature, from convex to concave, of $\Phi(\mu, \sigma)$ occurs around the threshold, $\mu = \theta$. The suprathreshold increase of the rate eventually saturates at $1/\tau_{\mathrm{rp}}$.

3.1.6 Higher order firing statistics of LIF neurons

Once we know the mean emission rate, we can compute the higher-order moments of the distribution of ISIs, μ_k, by integrating the recurrence relations for the first-passage times [84, 184]

$$-k\mu_{k-1}(x) = \frac{\sigma^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\mu_k(x) + (\mu - x)\frac{\mathrm{d}}{\mathrm{d}x}\mu_k(x).$$

In particular, we can derive from them the coefficient of variation of the ISI [184, 38]

$$\mathrm{CV}^2 = \frac{\langle\langle \mathrm{ISI}^2 \rangle\rangle}{\langle \mathrm{ISI} \rangle^2} = \frac{\mu_2(\theta) - \mu_1^2(\theta)}{\mu_1^2(\theta)} = 2\pi\nu_0^2 \int_{(H-\mu)/\sigma}^{(\theta-\mu)/\sigma} \exp(x^2)\psi(x)\,\mathrm{d}x,$$

where

$$\psi(x) = \int_{-\infty}^{x} \exp(y^2)\big[1 + \mathrm{erf}(y)\big]^2\,\mathrm{d}y.$$

The dependence of the coefficient of variation on the mean current μ is shown in Fig. 3.2D. For subthreshold currents, the voltage can decay to the steady state $v = \mu$ without crossing the threshold (see Fig. 3.2A). Were it not for the fluctuations in the current, the voltage would settle at $V = \mu$ and stay there forever. Fluctuations allow the voltage to escape from the local potential and reach the threshold, making the neuron fire. This is reflected in highly variable interspike intervals with coefficients of variation close to one. In contrast, for suprathreshold currents the voltage always crosses the threshold during its decay to the steady state $V = \mu$. A spike is then emitted, the voltage is set to the reset value, and the process starts over again. This process is entirely due to the deterministic component of the dynamics. Fluctuations only add some variability in the decay process, but are not needed for the neuron to fire. This is reflected in more regular spike trains, characterized by low values of the CV. The transition between the sub- and suprathreshold regimes of the CV is sharper the smaller is the amplitude of the fluctuations. In the limit of zero noise, the CV is a step function with value 1 for $\mu < \theta$ and 0 otherwise.

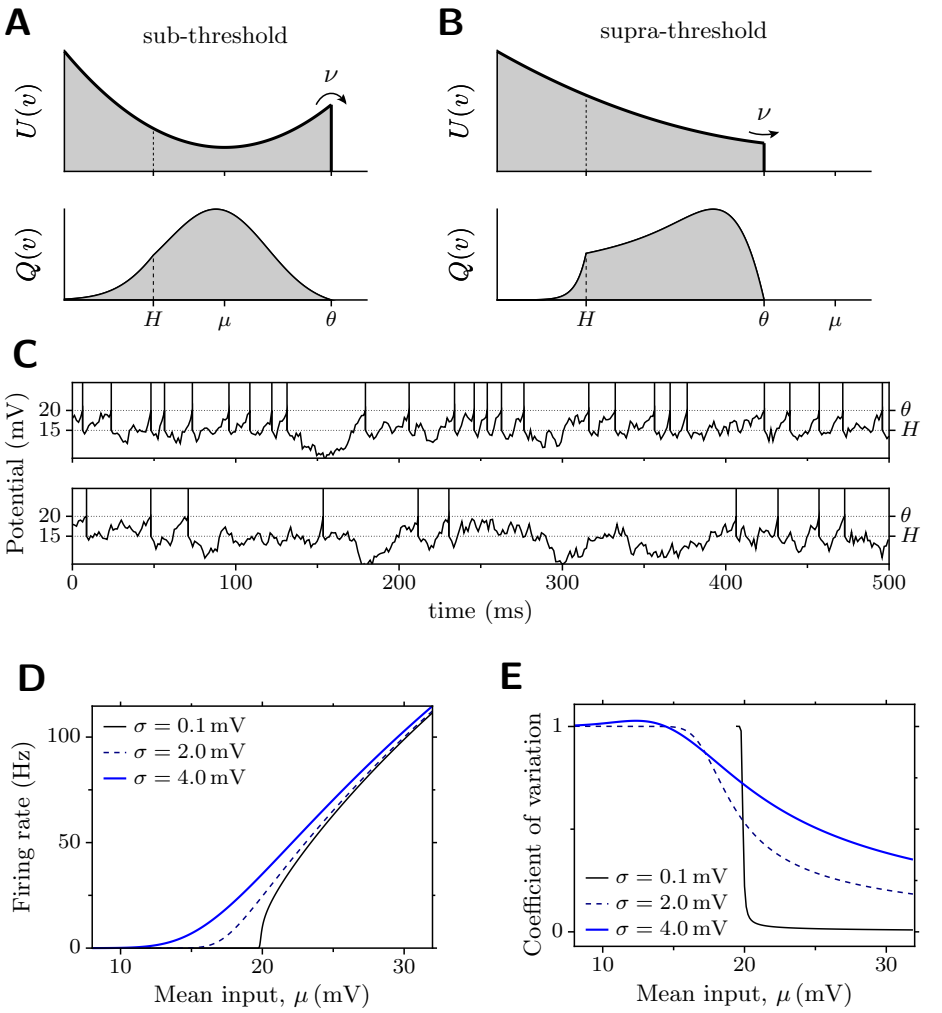


Figure 3.2: **A**: (Top): Dynamics of the depolarization of a LIF neuron receiving stochastic inputs. The sample path of the membrane depolarization is described by the stochastic motion in the harmonic potential $U(v) = (v - \mu)^2$, truncated at the firing threshold $v = \theta$ (Bottom): Stationary probability density of the membrane voltage, $Q(v)$. Note the (slight) discontinuity in the derivative of $Q(v)$ at $v = H$, due to the reset condition. The two plots correspond to the subthreshold regime, $\mu < \theta$. **B**: As in A, but for suprathreshold currents. The discontinuity in $Q'(v)$ at $v = H$ is evident. **C**: Membrane traces for subthreshold (bottom) and suprathreshold inputs (top). The threshold is $\theta = 20\,\mathrm{mV}$, and the inputs are $\mu = 16\,\mathrm{mV}$ (subthreshold) and $\mu = 22\,\mathrm{mV}$ (suprathreshold). In both cases $\sigma = 4\,\mathrm{mV}$. **D**: Single neuron transduction function $\Phi(\mu, \sigma) = \langle \mathrm{ISI} \rangle^{-1}$ for three different amplitudes of noise. **E**: Coefficient of variation of the ISI, for three different amplitudes of noise. The abrupt termination of the curve for $\sigma = 0.1$ is caused by a numerical artifact (it should be 1 for $\mu < \theta$). Single cell parameters: $\theta = 20\,\mathrm{mV}$, $H = 15\,\mathrm{mV}$, $\tau = 20\,\mathrm{ms}$, $\tau_{\mathrm{rp}} = 2\,\mathrm{ms}$.

3.2 Firing rates in homogeneous populations of neurons

3.2.1 Unstructured network

We now come back to the initial question about how to describe the activity of the network in terms of global variables. Due to the homogeneity and statistical independence of the inputs, the system can be described by the rates of the two existing populations, the excitatory and the inhibitory. In the derivation of the f-I curve for a single cell, we have assumed for simplicity that there were only two types of connections, namely, J_E and J_I, depending on the type of the presynaptic cell. It is straightforward to extend the results to the case where the synaptic strength depends also on the type of the postsynaptic cell. The synaptic weights J are then of the type J_{EE}, J_{EI}, J_{II}, or J_{IE} (see 2.5.3, on p. 21).

The means and variances received by excitatory (E) and inhibitory (I) cells are then

$$\mu_E(\nu_E,\nu_I) = \tau\big\{J_{EE}\big[C_E\nu_E + C_{\text{ext}}\nu_{\text{ext}}\big] - J_{EI}C_I\nu_I\big\}\,, \tag{3.13a}$$

$$\mu_I(\nu_E,\nu_I) = \tau\big\{J_{IE}\big[C_E\nu_E + C_{\text{ext}}\nu_{\text{ext}}\big] - J_{II}C_I\nu_I\big\}\,, \tag{3.13b}$$

$$\sigma_E^2(\nu_E,\nu_I) = \tau\big\{J_{EE}^2\big[C_E\nu_E + C_{\text{ext}}\nu_{\text{ext}}\big] + J_{EI}^2C_I\nu_I\big\}\,, \tag{3.13c}$$

$$\sigma_I^2(\nu_E,\nu_I) = \tau\big\{J_{IE}^2\big[C_E\nu_E + C_{\text{ext}}\nu_{\text{ext}}\big] + J_{II}^2C_I\nu_I\big\}\,, \tag{3.13d}$$

where ν_E and ν_I are the firing rates of the cells in the excitatory and inhibitory populations, respectively. These rates are determined self-consistently, by requiring them to be the output rates of the corresponding transfer function when inputs are Eqs. (3.13), that is,

$$\nu_E = \Phi_E(\mu_E,\sigma_E)\,, \tag{3.14a}$$

$$\nu_I = \Phi_I(\mu_I,\sigma_I)\,. \tag{3.14b}$$

The set of equations (3.13)–(3.14) can be solved numerically to find the firing rates of the stationary stable network states. The stability of these stationary states can also be analyzed [38]. It turns out that the asynchronous state of the network with low rates and irregular firing (coefficient of variation close to 1) is stable as long as recurrent inhibition is stronger than recurrent excitation, external inputs are suprathreshold, and the synaptic delays are sufficiently small compared to the membrane time constant τ.

3.2.2 Network structured by learning

As described in section 2.5.5, the effects of learning can be incorporated in the model by introducing structure in the matrix of excitatory synapses. We have seen that the 'strength' of memories can be parametrized by the level of recurrent excitation within selective populations —the parameter w_+ defined in Sec. 2.5.5, p. 25. When recurrent excitation is low, the only stable state of the network is the so-called 'spontaneous' state of the network, characterized by similar low rates for all cells in the network. When recurrent excitation is greater than some threshold

value, however, a new set of solutions appears. These networks states are called 'memory' states, and are characterized by the elevated firing of the cells in one of the selective sub-populations. The remaining cells, both nonselective cells and selective cells from other sub-populations, fire at low rates, not necessarily equal.

The mean field approximation allows us to compute the mean rates that characterize these memory states. Because of the structure in the population of excitatory cells, we have to distinguish between the rates and inputs of cells in the activated selective sub-population (denoted by $\nu_{\mathrm{sl}\uparrow}$, $\mu_{\mathrm{sl}\uparrow}$, and $\sigma_{\mathrm{sl}\uparrow}$), in non-activated selective sub-populations ($\nu_{\mathrm{sl}\downarrow}$, $\mu_{\mathrm{sl}\downarrow}$, $\sigma_{\mathrm{sl}\downarrow}$), and nonselective cells (ν_{nsl}, μ_{nsl}, σ_{nsl}). All these variables must obey the self-consistency relationships between the inputs and output rates. Let $\boldsymbol{\nu}$ denote the vector formed by all the different rates, $\boldsymbol{\nu} = (\nu_{\mathrm{sl}\uparrow}, \nu_{\mathrm{sl}\downarrow}, \nu_{\mathrm{nsl}}, \nu_I)$. The mean input for each of the four possible cell categories are

$$\begin{aligned}\mu_{\mathrm{sl}\uparrow}(\boldsymbol{\nu}) = \tau\big\{ & J_{EE}C_E\big[f w_+ \nu_{\mathrm{sl}\uparrow} + f(p-1)w_-\nu_{\mathrm{sl}\downarrow} + (1-pf)w_-\nu_{\mathrm{nsl}}\big] \\ & + J_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J_{EI}C_I\nu_I\big\}\,,\end{aligned} \tag{3.15a}$$

$$\begin{aligned}\mu_{\mathrm{sl}\downarrow}(\boldsymbol{\nu}) = \tau\big\{ & J_{EE}C_E\big[f w_- \nu_{\mathrm{sl}\uparrow} + f\big(w_+ + (p-2)w_-\big)\nu_{\mathrm{sl}\downarrow} \\ & + (1-pf)w_-\nu_{\mathrm{nsl}}\big] + J_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J_{EI}C_I\nu_I\big\}\,,\end{aligned} \tag{3.15b}$$

$$\begin{aligned}\mu_{\mathrm{nsl}}(\boldsymbol{\nu}) = \tau\big\{ & J_{EE}C_E\big[f \nu_{\mathrm{sl}\uparrow} + f(p-1)\nu_{\mathrm{sl}\downarrow} + (1-pf)\nu_{\mathrm{nsl}}\big] \\ & + J_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J_{EI}C_I\nu_I\big\}\,,\end{aligned} \tag{3.15c}$$

$$\begin{aligned}\mu_I(\boldsymbol{\nu}) = \tau\big\{ & J_{IE}C_E\big[f \nu_{\mathrm{sl}\uparrow} + f(p-1)\nu_{\mathrm{sl}\downarrow} + (1-pf)\nu_{\mathrm{nsl}}\big] \\ & + J_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J_{II}C_I\nu_I\big\}\,.\end{aligned} \tag{3.15d}$$

The expressions for the variances are similar, with the only difference that synaptic strengths, including modulatory factors w_+ and w_-, are squared

$$\begin{aligned}\sigma^2_{\mathrm{sl}\uparrow}(\boldsymbol{\nu}) = \tau\big\{ & J^2_{EE}C_E\big[f w^2_+ \nu_{\mathrm{sl}\uparrow} + f(p-1)w^2_-\nu_{\mathrm{sl}\downarrow} + (1-pf)w^2_-\nu_{\mathrm{nsl}}\big] \\ & + J^2_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J^2_{EI}C_I\nu_I\big\}\,,\end{aligned} \tag{3.15e}$$

$$\begin{aligned}\sigma^2_{\mathrm{sl}\downarrow}(\boldsymbol{\nu}) = \tau\big\{ & J^2_{EE}C_E\big[f w^2_- \nu_{\mathrm{sl}\uparrow} + f\big(w^2_+ + (p-2)w^2_-\big)\nu_{\mathrm{sl}\downarrow} \\ & + (1-pf)w^2_-\nu_{\mathrm{nsl}}\big] + J^2_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J^2_{EI}C_I\nu_I\big\}\,,\end{aligned} \tag{3.15f}$$

$$\begin{aligned}\sigma^2_{\mathrm{nsl}}(\boldsymbol{\nu}) = \tau\big\{ & J^2_{EE}C_E\big[f \nu_{\mathrm{sl}\uparrow} + f(p-1)\nu_{\mathrm{sl}\downarrow} + (1-pf)\nu_{\mathrm{nsl}}\big] \\ & + J^2_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J^2_{EI}C_I\nu_I\big\}\,,\end{aligned} \tag{3.15g}$$

$$\begin{aligned}\sigma^2_I(\boldsymbol{\nu}) = \tau\big\{ & J^2_{IE}C_E\big[f \nu_{\mathrm{sl}\uparrow} + f(p-1)\nu_{\mathrm{sl}\downarrow} + (1-pf)\nu_{\mathrm{nsl}}\big] \\ & + J^2_{EE}C_{\mathrm{ext}}\nu_{\mathrm{ext}} - J^2_{II}C_I\nu_I\big\}\,.\end{aligned} \tag{3.15h}$$

On the other hand, cells emit spikes in response to the inputs given by (3.15). The mean emission spike rates for each of the four possible categories are

$$\nu_{\mathrm{sl}\uparrow} = \Phi_E(\mu_{\mathrm{sl}\uparrow}, \sigma^2_{\mathrm{sl}\uparrow})\,, \tag{3.16a}$$

$$\nu_{\mathrm{sl}\downarrow} = \Phi_E(\mu_{\mathrm{sl}\downarrow}, \sigma^2_{\mathrm{sl}\downarrow})\,, \tag{3.16b}$$

$$\nu_{\mathrm{nsl}} = \Phi_E(\mu_{\mathrm{nsl}}, \sigma^2_{\mathrm{nsl}})\,, \tag{3.16c}$$

$$\nu_I = \Phi_I(\mu_I, \sigma^2_I)\,. \tag{3.16d}$$

The state of the network is described by the vector of rates $\boldsymbol{\nu}$, which can be determined solving numerically the equations (3.15)–(3.16).

Multi-memory states

The potentiation parameter w_+ can be increased beyond the threshold value $w_+^{(1)}$ necessary to sustain single memory states. At some value $w_+^{(2)}$, a new type of network state will appear, in which two selective sub-populations are active, i.e., fire at rates considerably higher than background values. Increasing even further the value of w_+ gives rise to new stable states with an increasing number of active sub-populations. These are the so-called multi-item memory states or l-memory states, l being the number of coactivated subpopulations. In general, the threshold values of the potentiation parameter $w_+^{(l)}$ at which the l-memory states appear form an ordered sequence, i.e., $w_+^{(1)} < w_+^{(2)} < w_+^{(3)} < \cdots$. We will find 2-memory states in the analysis of the decision-making network presented in chapter 5.

3.3 Treatment of realistic synaptic dynamics

The analytical results reviewed in previous sections are based on a very simple model for the input currents, that of instantaneous synapses. Real synaptic currents differ from those described above in at least three respects. First, synapses are not instantaneous, but have a characteristic time course that depends on the kinetics of the receptor, as seen in Section 2.5.2. Second, synaptic activation results from an increase in the synaptic conductance, rather than from a plain injection of charge. Third, receptors saturate, which implies that in general postsynaptic currents do not necessarily sum linearly. We show in the following how these features are dealt with in the mean field formalism.

3.3.1 Synaptic filtering

In the previous section, the noise carried by the inputs was assumed to be delta correlated In real neurons, however, the inputs entering the soma have been previously filtered by synapses, which introduce temporal correlations in the synaptic currents. This can be easily seen with the simple model for fast synapses we have introduced in 2.5.2, in which currents obey an equation of the form

$$\tau_s \dot{I}(t) = -I(t) + S(t)\,, \tag{3.17}$$

where τ_s is the relaxation time of the synapse and $S(t)$ is a sum of delta pulses,

$$S(t) = \sum_{\alpha} \sum_{j=1}^{C_\alpha} J_{\alpha\, j} \sum_k \delta(t - t_j^{(k)})\,.$$

The first sum is over the different types of neuron, $\alpha \in \{E, I, \text{ext}\}$. Using the diffusion approximation, we can describe the fluctuations around the mean of the

input train as pure Gaussian white noise

$$S(t) \approx \tilde{\mu}(t) + \tilde{\sigma}(t)\eta(t)\,,$$

where $\tilde{\mu}(t)$ and $\tilde{\sigma}^2(t)$ are proportional to the mean and the variance of the synaptic input. To be more specific, $\langle S(t)\rangle = \tilde{\mu}(t)$, $\langle\langle S(t)S(t')\rangle\rangle = \tilde{\sigma}^2(t)\delta(t-t')$, where $\tilde{\mu}(t)$ and $\tilde{\sigma}^2(t)$ are given in Eqs. (3.2) and (3.3). The quantity $\eta(t)$ is a source of delta-correlated Gaussian random variable (i.e., it satisfies Eqs. (3.9)). From the integration of Eq. (3.17) it follows

$$\begin{aligned}\langle I(t)\rangle &= \tilde{\mu}(t)\,,\\ \langle\langle I(t)I(t')\rangle\rangle &= \frac{\tilde{\sigma}^2}{2\tau_s}\exp\left(-\frac{|t-t'|}{\tau_s}\right).\end{aligned}$$

The current has thus an exponentially decaying correlation function, with the time scale given by the synaptic time constant. The main consequence of this nonzero width of the two-point correlation function is that the membrane voltage V can no longer be regarded as a Markov process. However, the joint process that results from taking the voltage V together with the current I, *is* a Markov process —a *bivariate* one. In general, non-Markov processes may be reduced to Markov processes by introducing new random variables. It is possible therefore to derive the associated two-dimensional Fokker-Planck equation and, in principle, solve it to obtain the joint probability density for V and I, from which the mean firing rate can be obtained. The calculation can be done using singular perturbation theory in the small τ_s limit [92, 42]. It can be shown that the synaptic decay time effectively introduces an effective threshold $\theta + \Delta$ and an effective reset potential $H + \Delta$, where $\Delta = \sigma(\alpha/2)\sqrt{\tau_s/\tau}$, with $\alpha = \sqrt{2}|\zeta(1/2)| \approx 3.479$, τ being the membrane time constant and $\zeta(x)$ being the Riemann zeta function [79]. This correction is valid to first order in $k = \sqrt{\tau_s/\tau}$. The main consequence of synaptic filtering is a reduction in the post-synaptic firing rates. This reduction is specially manifest in sub-threshold regimes, where the neuronal discharge is driven by fluctuations and is hence more sensitive to synaptic filtering.

3.3.2 Conductance-based description of the currents

The amount of current flowing through ion channels is proportional to the product of the number of open channels times the driving force (see Eq. (2.7)). All postsynaptic currents are thus voltage-dependent. In addition to this natural dependence on the voltage, there may be other voltage-dependencies, like that of the maximal conductance in NMDA-mediated currents due to magnesium blockage.

In conductance-based descriptions of currents, stochasticity is introduced by the synaptic gating variables. Assuming stationary inputs, the time course of a gating variable $s_{\rm syn}(t)$ can be split in a constant component $\bar{s}_{\rm syn}$ and a fluctuating component $\delta s_{\rm syn}(t)$. If we ignore for the moment any dependence on the voltage of the maximal conductance, the equation (2.7) for the unitary synaptic current

reads now

$$\begin{aligned} I_{\rm syn}(t) &= g_{\max} s_{\rm syn}(t)\big(V(t)-V_{\rm syn}\big) \\ &= g_{\max}\bar{s}_{\rm syn}\big(V(t)-V_{\rm syn}\big) + g_{\max}\delta\bar{s}_{\rm syn}(t)\big(V(t)-V_{\rm syn}\big) \end{aligned} \tag{3.18}$$

The fluctuation term depends on the voltage. To avoid the difficulties associated to multiplicative noise, the voltage is replaced by its average in the driving force for the fluctuating component, i.e., $g_{\max}\delta s_{\rm syn}(t)\big(V(t)-V_{\rm syn}\big) \approx g_{\max}\delta s_{\rm syn}\big(\overline{V}-V_{\rm syn}\big)$. The deterministic part of the current is on the other hand easy to handle. The factor $g_{\max}\bar{s}_{\rm syn}$ can be absorbed by a redefinition of the leak conductance, $g_L \to g_L + g_{\max}\bar{s}_{\rm syn}$, leading to an effective decrease of the membrane time constant from $\tau_m = C_m/g_L$ to

$$\tau_m^{\rm eff} \equiv C_m/(g_L + g_{\max}\bar{s}_{\rm syn})\,.$$

The neuron is therefore leakier by a factor $1 + g_{\max}\bar{s}_{\rm syn}/g_L$. From (3.18) we see that the resting membrane potential is also renormalized through the constant $g_{\max}\bar{s}_{\rm syn}V_{\rm syn}$ and the effective leak conductance, leading to

$$V_L \to \frac{g_L V_L + g_{\max}\bar{s}_{\rm syn}V_{\rm syn}}{g_L + g_{\max}\bar{s}_{\rm syn}}\,.$$

The maximal conductance for NMDA-currents depends on the voltage (see Eq. (2.10) and Fig. 2.6). To simplify the analytical treatment, the voltage dependence of NMDA currents is linearized around the mean value of the voltage $\overline{V}$, which will be calculated self-consistently below. The linearization reads

$$\begin{aligned} \frac{V(t)-V_E}{1+[{\rm Mg}^{2+}]\exp(-\beta V)/\gamma} &\equiv \frac{V(t)-V_E}{J(V)} \\ &\approx \frac{V(t)-V_E}{J(\overline{V})} + \big(V(t)-\overline{V}\big)\frac{J(\overline{V})-\beta(\overline{V}-V_E)\big(1-J(\overline{V})\big)}{J^2(\overline{V})}\,. \end{aligned}$$

This linear approximation is very accurate for the range of values between the reset potential and the threshold, which is where the voltage stays most of the time (see Fig. 2.6). As we will see below, the mean value of the voltage turns out to be in this range. Note that the current that results from the linearization can be written as

$$I_{\rm NMDA}(t) = g_{\rm NMDA}^{\rm eff} s_{\rm NMDA}(t)\big(V(t)-V_E^{\rm eff}\big)\,,$$

where

$$g_{\rm NMDA}^{\rm eff} = g_{\rm NMDA}\frac{J(\overline{V})-\beta(\overline{V}-V_E)\big(1-J(\overline{V})\big)}{J^2(\overline{V})}\,, \tag{3.19a}$$

$$V_E^{\rm eff} = \overline{V} - \frac{g_{\rm NMDA}}{g_{\rm NMDA}^{\rm eff}}\left(\frac{\overline{V}-V_E}{J(\overline{V})}\right)\,. \tag{3.19b}$$

To complete the derivation, we must specify the average voltage $\overline{V}$. Its value can be computed using the probability density function $Q(V)$ for the potential,

Eq. (3.11), and taking into account the stay at the reset potential during the refractory period

$$\begin{aligned}\overline{V} &= \int_{-\infty}^{\theta} V\big(Q(V) + \nu\tau_{\rm rp}\delta(V-H)\big)\,{\rm d}V \\ &= \mu - (\theta - H)\nu\tau_m^{\rm eff} - (\mu - H)\nu\tau_{\rm rp},\end{aligned}$$

where μ is the mean synaptic current, which will be specified below, and ν is the *post*synaptic rate. Note that the average voltage is a linear function of ν, which implies that the postsynaptic rate ν cannot be simply computed as a function of the synaptic input, but it has to be instead determined in a self-consistent way.

3.3.3 Fluctuations in the synaptic current with multiple synaptic time scales

The derivation presented above can be applied when synaptic currents have one single time scale, which is certainly not the case when fluctuations are meditated by AMPA, NMDA, and GABA receptors. There is no rigorous analytical treatment for fluctuations with multiple time scales, so one has to resort to approximations. A simple one can be made by noting that there is a hierarchy of time scales, the time constant of AMPA receptors being much smaller than those of GABA and NMDA receptors. The longer the synaptic time constants, the more fluctuations are filtered out. From this it follows that the amplitude of the fluctuations of the AMPA synaptic variables is much larger than those in GABA and NMDA synaptic variables. The approximation neglects fluctuations due to GABA and NMDA receptors. Another approximation can be made if $g_{\rm AMPA,rec}$ is much smaller than $g_{\rm AMPA,ext}$, neglecting the former. Figure 3.3 summarizes the architecture, the population structure, and the different types of synaptic current of the cortical network model.

3.3.4 Nonlinear sum of inputs

The temporal dynamics of AMPA and GABA are fast compared to the typical interspike-intervals, so that in normal conditions these receptors can recover during the time between consecutive spikes. As a consequence, the fraction of open channels is a linear function of the presynaptic rate when these are in physiological ranges. This is easily seen by integrating the dynamical equations for these receptors, Eq.(2.8), replacing the presynaptic train of spikes by the corresponding mean ν, giving

$$\langle s_{\rm AMPA}(t)\rangle \equiv \bar{s}_{\rm AMPA} = \nu\tau_{\rm AMPA} < 1\,, \tag{3.20a}$$

$$\langle s_{\rm GABA}(t)\rangle \equiv \bar{s}_{\rm GABA} = \nu\tau_{\rm GABA} < 1\,. \tag{3.20b}$$

NMDA receptors have, on the contrary, decay times of the order of hundred milliseconds, and hence they saturate whenever the incoming spike has a rate of around 10 Hz. The gating variable becomes then a nonlinear function of the

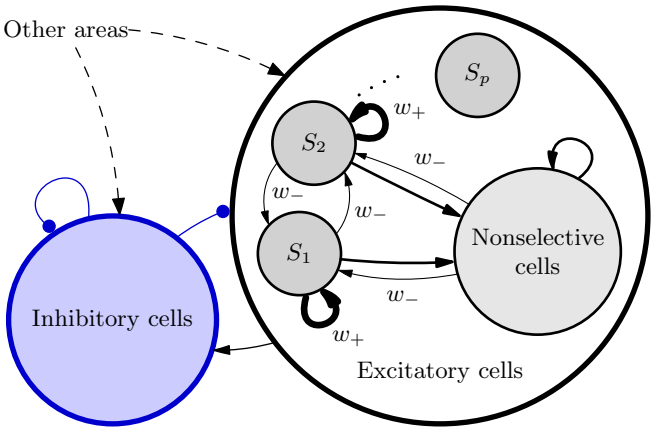


Figure 3.3: Architecture of the cortical network model. The population of excitatory neurons is subdivided in non-overlapping populations selective to p different stimuli. Black solid arrows: NMDA and AMPA-mediated recurrent excitatory connections. Black dashed arrows: AMPA-mediated external excitatory connections. Blue circle-headed arrows: GABA-mediated inhibitory connections. There are three possible synaptic strengths for recurrent excitatory connections: potentiated (by a relative factor $w_+ > 1$, thick arrows), depressed (by a relative factor $w_- < 1$, thin arrows), and unchanged (baseline level $w = 1$, connections without labels). The dots stand for the missing $S_3, \ldots, S_{p-1}$ populations and the corresponding connections. Connections from and to population S_p are not shown for clarity.

presynaptic rate. As a first approximation, this nonlinear function can be derived replacing the incoming spike train $S(t) = \sum_k \delta(t - t^{(k)})$ in Eqs. (2.9) by the corresponding mean discharge rate $\langle S(t) \rangle = \nu$, and finding the stationary solution. The average value of the gating variable would be in this case

$$\langle s_{\text{NMDA}}(t) \rangle = \frac{\tilde{\tau}\nu}{1 + \tilde{\tau}\nu} \,,$$

where $\tilde{\tau} \equiv \tau_{\text{NMDA}\uparrow}\tau_{\text{NMDA}\downarrow}\alpha$. The approximation can be refined if $S(t)$ is taken as a pure Poisson spike train. The calculation is in this case more involved [43], and we only sketch the ideas behind the derivation. The starting point is the formal solution for the gating variable $s_{\text{NMDA}}(t)$ in Eqs. (2.9)

$$s_{\text{NMDA}}(t) = \alpha \int_{-\infty}^{t} x(v) \exp\left(-\frac{t-u}{\tau_{\text{NMDA}\downarrow}} - \alpha \int_{v}^{t} x(u)\, \mathrm{d}u \right) \mathrm{d}v \,, \tag{3.21}$$

Since $s_{\text{NMDA}}(t)$ is a function of the random variable $x(t)$, it is a random variable itself. We split the variable $x(t)$ in Eqs. (2.9) into a deterministic component, equal to the stationary solution $\langle x \rangle = \tau_{\text{NMDA}\uparrow}\nu$, and a fluctuating component δx, of zero mean, $\langle \delta x(t) \rangle = 0$. The n-point correlation function of x can be obtained

from the properties of Poisson processes

$$\langle \delta x(t_1) \cdots \delta x(t_n) \rangle = \langle x \rangle \exp\Big(- \sum_{j=2}^{n} (t_j - t_1)/\tau_{\text{NMDA}\uparrow} \Big) . \tag{3.22}$$

The average of s_{NMDA} can then be calculated from the expansion in power series of the exponential in Eq. (3.21), separating the deterministic from the fluctuating component and using (3.22). The final result is

$$\langle s_{\text{NMDA}}(t) \rangle \equiv \psi(\nu) = \frac{\tilde{\tau}\nu}{1+\tilde{\tau}\nu} \Big(1 + \frac{1}{1+\nu\tilde{\tau}} \sum_{n=1}^{\infty} \frac{1}{(n+1)!} (-\alpha \tau_{\text{NMDA}\uparrow})^n T_n \Big), \tag{3.23}$$

where

$$T_n \equiv \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{\tau_{\text{NMDA}\uparrow}(1+\nu\tilde{\tau})}{\tau_{\text{NMDA}\uparrow}(1+\nu\tilde{\tau}) + k\tau_{\text{NMDA}\downarrow}} .$$

3.3.5 Firing statistics of a neuron with realistic synaptic inputs

Here we derive the final mean field equations for networks with realistic synaptic inputs mediated by AMPA, GABA, and NMDA receptors. We go along the same lines as the derivation followed for the network with instantaneous synapses. Thus, we assume that the synaptic conductances and the mean firing rates of all presynaptic inputs from the same subpopulation are identical, so that the network can be partitioned into subpopulations.

In the spontaneous activity state, the network consists functionally of one population of C_E excitatory cells and one population of C_I cells. The subthreshold dynamics of a single cell, Eq. (2.1), with the synaptic currents given by Eqs. (2.15) is then

$$\begin{aligned} C_m \dot{V}(t) = &- g_L \big(V(t) - V_L\big) - g_{\text{AMPA ext}} \big(V(t) - V_E\big) S_{\text{AMPA ext}}(t) \\ &- g_{\text{AMPA rec}} \big(V(t) - V_E\big) S_{\text{AMPA rec}}(t) - g^{\text{eff}}_{\text{NMDA}} \big(V(t) - V^{\text{eff}}_E\big) S_{\text{NMDA}}(t) \\ &- g_{\text{GABA}} \big(V(t) - V_I\big) S_{\text{GABA}}(t) , \end{aligned} \tag{3.24}$$

where $g^{\text{eff}}_{\text{NMDA}}$ and V^{eff}_E are given by Eqs. (3.19). The symbols $S_{\text{AMPA ext}}$, $S_{\text{AMPA rec}}$, S_{NMDA}, and S_{GABA}, denote the *total* fraction of open channels for each type of receptor. By appealing to the diffusion approximation, gating variables can be

replaced with a sum of a deterministic tonic component and a fluctuation term

$$\begin{aligned}
S_{\text{AMPA ext}}(t) &\equiv \sum_{j=1}^{C_{\text{ext}}} s_j^{\text{AMPA ext}} = C_{\text{ext}}\big(\bar{s}_{\text{AMPA ext}} + \delta s_{\text{AMPA ext}}(t)\big) \\
&= C_{\text{ext}}\big(\tau_{\text{AMPA}}\nu_{\text{ext}} + \delta s_{\text{AMPA ext}}(t)\big)\,, \\
S_{\text{AMPA rec}}(t) &\equiv \sum_{j=1}^{C_E} w_j s_j^{\text{AMPA rec}} = C_E\big(\bar{s}_{\text{AMPA rec}} + \delta s_{\text{AMPA rec}}(t)\big) \approx C_E\tau_{\text{AMPA}}\nu_E\,, \\
S_{\text{NMDA}}(t) &\equiv \sum_{j=1}^{C_E} w_j s_j^{\text{NMDA}} = C_E\big(\bar{s}_{\text{NMDA}} + \delta s_{\text{NMDA}}(t)\big) \approx C_E\psi(\nu_E)\,, \\
S_{\text{GABA}}(t) &\equiv \sum_{j=1}^{C_I} s_j^{\text{GABA}} = C_I\big(\bar{s}_{\text{GABA}} + \delta s_{\text{GABA}}(t)\big) \approx C_I\tau_{\text{GABA}}\nu_I\,.
\end{aligned}$$

As explicitly stated in these equations, fluctuations are assumed to be introduced by external AMPA receptors only. The correlation function of these fluctuations is, from the discussion following Eq. (3.17),

$$\big\langle \delta s_{\text{AMPA ext}}(t)\,\delta s_{\text{AMPA ext}}(t')\big\rangle = \nu_{\text{ext}}\tau_{\text{AMPA}}\exp\big(-|t-t'|/\tau_{\text{AMPA}}\big)\,.$$

Given that the deterministic components of the current are all linear in the voltage, the dynamical equation (3.24) for the membrane potential can be rewritten in terms of the effective parameters introduced above. The equation then reads

$$\tilde{\tau}_m\dot{V} = V_L - V + \mu + \sigma\sqrt{\tilde{\tau}_m}\eta(t)\,, \tag{3.25}$$

where we have defined the renormalized time constant

$$\begin{aligned}
\tilde{\tau}_m &\equiv C_m/\tilde{g}_L\,, \\
\tilde{g}_L &\equiv g_L S\,, \\
S &\equiv 1 + g_L^{-1}\big(G_{\text{AMPA ext}} + G_{\text{AMPA rec}} + G_{\text{NMDA}} + G_{\text{GABA}}\big)\,,
\end{aligned} \tag{3.26}$$

with

$$\begin{aligned}
G_{\text{AMPA ext}} &\equiv g_{\text{AMPA ext}}C_{\text{ext}}\tau_{\text{AMPA}}\nu_{\text{ext}}\,, & G_{\text{NMDA}} &\equiv g_{\text{NMDA}}^{\text{eff}}C_E\psi(\nu_E)\,, \\
G_{\text{AMPA rec}} &\equiv g_{\text{AMPA rec}}C_E\tau_{\text{AMPA}}\nu_E\,, & G_{\text{GABA}} &\equiv g_{\text{GABA}}C_I\nu_I\tau_{\text{GABA}}\,.
\end{aligned}$$

The synaptic current is specified by the following parameters

$$\begin{aligned}
\mu = \frac{1}{\tilde{g}_L}\Big\{&(V_E - V_L)\big(G_{\text{AMPA ext}} + G_{\text{AMPA rec}}\big) \\
&+ (V_E^{\text{eff}} - V_L)G_{\text{NMDA}} + (V_I - V_L)G_{\text{GABA}}\Big\}\,,
\end{aligned} \tag{3.27}$$

$$\sigma^2 = C_{\text{ext}}\nu_{\text{ext}}g_{\text{AMPA ext}}^2\big(\overline{V} - V_E\big)^2\tau_{\text{AMPA}}^2\tilde{\tau}_m/C_m^2\,, \tag{3.28}$$

The noise source $\eta(t)$ in Eq. (3.25)

$$\begin{aligned}\langle \eta(t) \rangle &= 0\,, \\ \langle \eta(t)\eta(t') \rangle &= \tau_{\text{AMPA}}^{-1} \exp\bigl(-|t-t'|/\tau_{\text{AMPA}}\bigr)\,.\end{aligned}$$

From Eq. (3.26), the effective membrane time constant is lower than the bare membrane time constant C_m/g_L by a shunting factor S. This decrease in the membrane time constant results from the increase in the overall conductance caused by the different types of synaptic bombardment. In fact, measures on *in vivo* preparations have shown that the effective leak conductance can be up to 5 times larger than the bare conductance g_L [69], and hence the effective membrane time constant $\tilde{\tau}_m$ can be as short as 2 ms. The mean current μ corresponds to the asymptotic value of the membrane potential in the absence of fluctuations and in the absence of spiking [43].

To obtain the mean firing rate of a LIF cell whose potential is described by Eq. (3.25) we follow the same procedure used in Section 3.1.5, with the caveat that inputs have exponentially decaying correlations due to synaptic filtering. As mentioned in section 3.3.1, such correlations introduce an effective threshold and an effective reset potential [79]

$$\begin{aligned}\theta &\to \theta + \Delta\,, \\ H &\to H + \Delta\,,\end{aligned} \qquad \text{where} \qquad \Delta \equiv \sigma \frac{\alpha}{2} \sqrt{\frac{\tau_{\text{AMPA}}}{\tilde{\tau}_m}}, \quad \alpha = \sqrt{2}\left|\zeta\Bigl(\frac{1}{2}\Bigr)\right|.$$

Thus, the mean discharge rate of a cell is

$$\nu = \phi(\mu,\sigma) = \left(\tau_{\text{rp}} + \sqrt{\pi}\tau \int_{\beta(\mu,\sigma)}^{\alpha(\mu,\sigma)} \exp\bigl(u^2\bigr)\bigl[1+\operatorname{erf}(u)\bigr]\,\mathrm{d}u\right)^{-1}, \tag{3.29}$$

where

$$\begin{aligned}\alpha(\mu,\sigma) &= (\theta + \Delta - V_L - \mu)/\sigma\,, \\ \beta(\mu,\sigma) &= (H + \Delta - V_L - \mu)/\sigma\,.\end{aligned}$$

Note that in [43] only the effective threshold appeared. The discrepancy is due to the slightly different assumptions made in [79] with respect to [42] regarding the stationary probability density function of synaptic currents (see [79] for details).

3.3.6 Delay activity states

The network is able to sustain persistent activity states when recurrent inputs are high enough, in a completely analogous way to the network with instantaneous synapses. The population of excitatory cells breaks into three subpopulations, depending on whether or not cells are selective and, if they are, on whether or not the cells are activated by the stimulus. The set of nonlinear equations characterizing the persistent activity state of the network are derived along the lines shown in Section 3.2.2. The difference with respect to Eqs. 3.15 is that now

the mean and variance of the inputs should include the different contributions from AMPA, NMDA, and GABA receptors.

It is convenient to introduce the following definitions

$$T_{\rm ext} \equiv (g_{\rm AMPA\,ext}/g_L) C_{\rm ext} \tau_{\rm AMPA}\,, \qquad \kappa \equiv (g_{\rm NMDA}/g_L) C_E \psi(\nu_E)\,,$$
$$T_{\rm AMPA} \equiv (g_{\rm AMPA\,rec}/g_L) C_E \tau_{\rm AMPA}\,, \qquad T_I \equiv (g_{\rm GABA}/g_L) C_I \tau_{\rm GABA}\,,$$

as well as

$$\Delta V_E \equiv V_E - V_L\,, \qquad \Delta V_E^{\rm eff} \equiv V_E^{\rm eff} - V_L\,, \qquad \Delta V_I \equiv V_I - V_L\,.$$

Using the same notation of Section 3.2.2, the populations are described by

$$\mu_{{\rm sl}\uparrow} = \frac{1}{S_{{\rm sl}\uparrow}} \Big\{ \Delta V_E \big(T_{\rm ext}\nu_{\rm ext} + T_{\rm AMPA} n_{{\rm sl}\uparrow} \big) + \Delta V_E^{\rm eff} \kappa N_{{\rm sl}\uparrow} + \Delta V_I T_I \nu_I \Big\}\,, \quad (3.30a)$$

$$\mu_{{\rm sl}\downarrow} = \frac{1}{S_{{\rm sl}\downarrow}} \Big\{ \Delta V_E \big(T_{\rm ext}\nu_{\rm ext} + T_{\rm AMPA} n_{{\rm sl}\downarrow} \big) + \Delta V_E^{\rm eff} \kappa N_{{\rm sl}\downarrow} + \Delta V_I T_I \nu_I \Big\}\,, \quad (3.30b)$$

$$\mu_{\rm nsl} = \frac{1}{S_{\rm nsl}} \Big\{ \Delta V_E \big(T_{\rm ext}\nu_{\rm ext} + T_{\rm AMPA} n_{\rm nsl} \big) + \Delta V_E^{\rm eff} \kappa N_{\rm nsl} + \Delta V_I T_I \nu_I \Big\}\,, \quad (3.30c)$$

where we have defined

$$S_{{\rm sl}\uparrow} = 1 + T_{\rm ext}\nu_{\rm ext} + T_{\rm AMPA} n_{{\rm sl}\uparrow} + \kappa N_{{\rm sl}\uparrow} + T_I \nu_I\,,$$
$$S_{{\rm sl}\downarrow} = 1 + T_{\rm ext}\nu_{\rm ext} + T_{\rm AMPA} n_{{\rm sl}\downarrow} + \kappa N_{{\rm sl}\downarrow} + T_I \nu_I\,,$$
$$S_{\rm nsl} = 1 + T_{\rm ext}\nu_{\rm ext} + T_{\rm AMPA} n_{\rm nsl} + \kappa N_{\rm nsl} + T_I \nu_I\,,$$

and

$$n_{{\rm sl}\uparrow} = f w_+ \nu_{{\rm sl}\uparrow} + f(p-1) w_- \nu_{{\rm sl}\downarrow} + (1-fp) w_- \nu_{\rm nsl}\,,$$
$$N_{{\rm sl}\uparrow} = f w_+ \psi(\nu_{{\rm sl}\uparrow}) + f(p-1) w_- \psi(\nu_{{\rm sl}\downarrow}) + (1-fp) w_- \psi(\nu_{\rm nsl})\,,$$
$$n_{{\rm sl}\downarrow} = f w_- \nu_{{\rm sl}\uparrow} + f\big(w_+ + (p-2)w_-\big) \nu_{{\rm sl}\downarrow} + (1-fp) w_- \nu_{\rm nsl}\,,$$
$$N_{{\rm sl}\downarrow} = f w_- \psi(\nu_{{\rm sl}\uparrow}) + f\big(w_+ + (p-2)w_-\big) \psi(\nu_{{\rm sl}\downarrow}) + (1-fp) w_- \psi(\nu_{\rm nsl})\,,$$
$$n_{\rm nsl} = f \nu_{{\rm sl}\uparrow} + f(p-1) \nu_{{\rm sl}\downarrow} + (1-fp) \nu_{\rm nsl}\,,$$
$$N_{\rm nsl} = f \psi(\nu_{{\rm sl}\uparrow}) + f(p-1) \psi(\nu_{{\rm sl}\downarrow}) + (1-fp) \psi(\nu_{\rm nsl})\,.$$

The mean firing rates of the different subpopulations are found from

$$\nu_{{\rm sl}\uparrow} = \phi(\mu_{{\rm sl}\uparrow}, \sigma)\,, \quad (3.31a)$$
$$\nu_{{\rm sl}\downarrow} = \phi(\mu_{{\rm sl}\downarrow}, \sigma)\,, \quad (3.31b)$$
$$\nu_{\rm nsl} = \phi(\mu_{\rm nsl}, \sigma)\,, \quad (3.31c)$$
$$\nu_I = \phi(\mu_I, \sigma)\,, \quad (3.31d)$$

where $\phi(\mu, \sigma)$ is given by Eq. (3.29), and the mean and variance of the synaptic inputs are given by Eqs. (3.30) and (3.28), respectively. The stable solutions of this self-consistency condition are the stable attractors of the structured network.

Chapter 4

Biophysically realistic models of decision-making

In this chapter we summarize the attractor-based model for decision-making first proposed by Wang in [191], and further extended in subsequent articles. We describe in section 4.1 the first version of the model, which is essentially a particular implementation of the network models introduced in the previous chapter. The network model can account for some salient aspects of both behavioral and neurophysiological data, which we enumerate in section 4.2. A two-variable rate model that captures the main dynamical properties of the original network model was introduced in [197] to provide a simpler and more manageable description of the decision process. It is described in section 4.3.

4.1 Decision-making network

As explained in the introduction (Ch. 1), neurons in area LIP exhibit decision-related activity during visual-saccadic perceptual tasks. These cells are innervated mostly by neurons from area MT, which are known to encode the instantaneous motion strength of the random-dot stimulus and provide, therefore, the sensory evidence relevant for the perceptual decision. When averaged over trials, the activity of LIP neurons builds up gradually during the presentation of the stimulus at a rate proportional to the fraction of dots moving coherently in the cell's preferred direction. The ramping activity observed in LIP cells has been hypothesized to be the neural correlate of the "integration of evidence" provided by the activity of MT cells [173, 164].

The idea that the activity of LIP reflects the accumulation of evidences necessary for a perceptual decision task was specially appealing due to its resemblance with sequential sampling models, a class of abstract models proposed to describe most of the behavioral aspects of two-choice decision tasks, such as speed-accuracy tradeoffs or reaction time distributions [123, 158]. The diffusion model [154] is perhaps the most conspicuous version of sequential sampling models. According to it, the moment-by-moment evidence provided by the sensory inputs is modeled as a

one-dimensional stochastic variable that is integrated over time. The accumulated signal plays the role of a decision variable, which quantifies the amount of evidence for or against a particular choice. When the decision variable crosses one of two predefined boundaries, a decision is considered to be made with choice determined by the boundary being reached. The model is simple and tractable, and fits surprisingly well not only behavioral data, including chronometric and psychometric curves, but also neuronal data [173, 137, 156].

However, while the model describes the data, it does not explain it. The model is heuristic and as such it does not address the possible cellular and synaptic mechanisms underlying the observed dynamics in LIP. It is unclear, for example, how a one-dimensional decision variable can arise from a high-dimensional and highly nonlinear system like a network of cortical neurons. Also relevant is the question about the neurobiological mechanisms underlying the non-leaky integration of inputs inherent in the diffusion model. In order to elucidate the biophysical mechanisms of the decision-related activity in LIP and its possible relationship with the existing behavioral models, Wang proposed a biophysically inspired network model for the visual discrimination task [191]. The observation that LIP neurons exhibit, apart from decision-related activity, direction-selective persistent activity during delay periods led Wang to investigate whether network models of working memory could also perform the integration of stimuli and the formation of categorical choices observed experimentally.

4.1.1 Architecture

Wang's cortical model of LIP consists of a fully connected recurrent network of integrate-and-fire neurons with realistic synaptic excitation, mediated by AMPA and NMDA receptors, and inhibition, mediated by GABA_A receptors (see the previous two chapters for details). The network is structured in a set of statistically homogeneous neural populations. In particular, the statistical properties of the synaptic currents and the connection strengths are identical for all the cells from the same population. There is one population of inhibitory cells and one population of excitatory cells, which is partitioned into three subpopulations. Two of these, A and B, represent the ensemble of LIP excitatory neurons selective to either one of the two possible directions of motion —say, cells in A are selective to leftward motion and cells in B are selective to rightward motion. The third subpopulation includes the remaining nonselective cells. Recurrent connections between cells from the same selective subpopulation A or B are potentiated by a factor $w_+ > 1$ with respect to the baseline connectivity level, while connections between cells from different selective subpopulations are weakened by a factor $0 < w_- < 1$, related to w_+ (see Sec. 2.5.5). A schematic representation of the network architecture is shown in Figure 4.1.

4.1.2 Stimulation and decision formation

Stimulation is modeled as the activation of external inputs to neurons in populations A and B, mimicking the motion-selective inputs provided by the projections

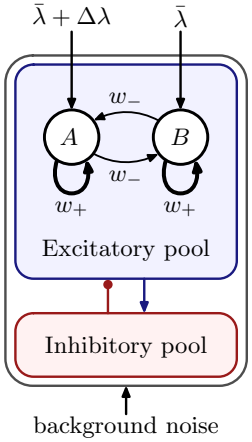


Figure 4.1: Architecture of the decision-making network [191]. Two populations A and B of excitatory neurons encode in their activity the two possible choices in the decision process. They are endowed with strong recurrent connections (parametrized by w_+), and inhibit each other through shared feedback from the inhibitory population. All cells receive nonselective background input, modeled with independent Poisson spike trains of rate ν_0. When recurrent connections are strong enough, the network operates as a winner-take-all. Stimulation is modeled as an increase, with respect to the background input ν_0, in the rate of the Poisson spike train arriving to selective cells. This selective increase is $\bar{\lambda} + \Delta\lambda$ and $\bar{\lambda} - \Delta\lambda$ for neurons in populations A and B, respectively. See text for details.

from area MT. Neurons in MT are tuned to a particular direction of visual motion, and their firing rates are given by an approximately linear increasing (decreasing) function of the coherence when the motion is in the preferred (null) direction of the cell [33]. The selective inputs due to the random dot stimulus are modeled accordingly [191], with greater imbalances for higher motion coherences.

The recurrent strength w_+ within selective populations plays an important role in the decision-related capabilities of the network. When w_+ is large enough the network can sustain, apart from the spontaneous activity state, a selective state that persists after stimulus withdrawal. The two populations engage into a competition for higher activity when stimulation is applied, increasing first their activity together until at some point the recurrent inputs are strong enough and the population rates start to diverge from each other (Fig. 4.2A). The network decays then into a delay activity state in which the activity of either one of the populations exceeds significantly the activity of the other —an *asymmetric state*. The choice made by the network is then said to be A or B according to the population firing at highest rate, i.e., according to the population *winning* the competition. Based on the observation that the activity of LIP neurons is correlated with the behavioral outcome, the activity of the two neural populations is assumed to select the monkey's planned motor response. That is, the possible behavioral outcomes of the task are identified with the possible asymmetric states of the model. To stress this correspondence, delay activity states are called in this context *decision* states, while the spontaneously active state of the network is called *neutral*, *resting*, or *undecided* state.

The temporal evolution of the system can be represented as a trajectory in the phase plane of the rates of the two competing populations, (ν_A, ν_B). Before stimulus onset the network is in the neutral state, with both populations firing at low and approximately equal rates. Were it not for the noise present in the system, the neutral state would be correspond to a single point in the lower left region in the phase plane, around the diagonal. Any trajectory starting at the vicinity of the fixed point would relax to it and stay there forever. Because of

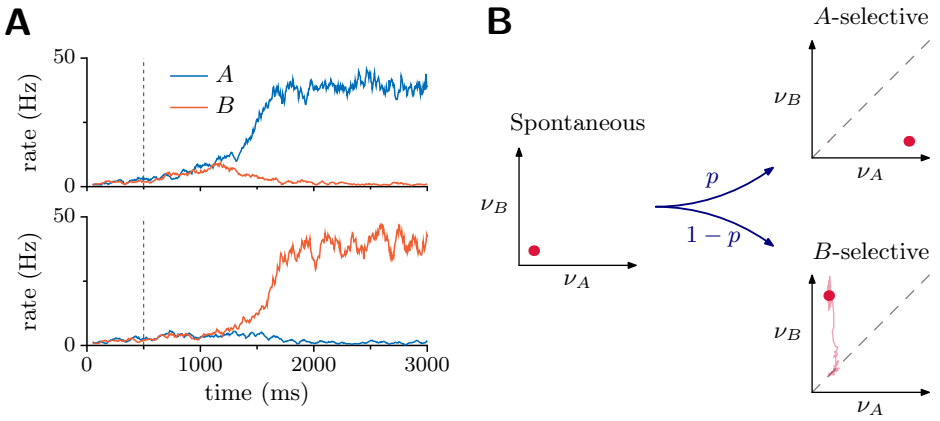


Figure 4.2: Decisions as stimulus-triggered transitions from the spontaneous (neutral) to either one of the two selective (decision) states. **A**: Average rates in selective populations as a function of time, for two different trials, one in which the network choses A (top) and another in which it choses B (bottom). Between $t = 500$ ms and $t = 3000$ ms, the selective input to both populations is increased by $\bar{\lambda} = 20$ Hz. **B**: Trajectories in the phase-plane (ν_A, ν_B) for the same two trials shown in **A**. Dashed line: separatrix $\nu_A = \nu_B$ between the two basins of attraction. Due to the noise present in the system, the outcome of this transition is random, with probability p that the final attractor is A and probability $1 - p$ that the final attractor is B. Crimson dots represent the peaks of the stationary distributions of rates associated with each attractor —hypothetically, the location of the fixed points in the limit of vanishingly small noise.

the stochasticity present in the network through background and selective inputs, attractors are not longer fixed points but rather distributions of neural activity centered around the stable fixed points that would arise in the limit of an infinitely large network. Incidentally, another consequence of the presence of noise is that the system can escape spontaneously (without stimulation) from one attractor to another. We will address this issue thoroughly in the next chapter.

The increase in selective inputs reconfigures the attractor landscape of the network, destabilizing the neutral state and partitioning the phase plane into two basins of attraction associated with the two decision states. When the selective inputs to both populations are identical, the separatrix between the two basins coincides with the diagonal $\nu_A = \nu_B$. The system follows then a jagged trajectory along the separatrix, until goes deep into one of the basins of attraction and eventually decays to the corresponding attractor (Fig. 4.2B). Due to the symmetry of the attractor landscape and of the initial configuration, the network chooses randomly one of the two possible attractors with equal probability. When external inputs are not identical the symmetry of the system is lost, and the basin of attraction associated with the correct choice gets larger at the expense of the other. This results in an increase of the probability of choosing correctly. The larger the difference between the inputs to A and B is, the larger more decisions will be biased toward the correct choice. We will make this claims more precise in

Sec. 4.3, when we discuss the two-dimensional rate model of decision.

The model operates in this situation as a stochastic winner-take all [98]. All trajectories converge to one or the other attractor depending on the initial condition and on the realization of the noise. Trajectories starting in the region $\nu_A > \nu_B$ are more likely to converge to the A-selective attractor due to the deterministic drift of the dynamics, while trajectories starting in the region $\nu_B > \nu_A$ are likely to converge for the same reason to the B-selective attractor. The probability of falling into one or another attractor is closer to chance when the system is around the separatrix between the two basins of attraction, where trajectories that start very close together may have very different dynamical fates. In this region, the system is very sensitive to initial conditions, and it is noise, not the deterministic drift of attractor dynamics, which determines ultimately the successive states of the system. In the limit of vanishing noise, there is no possible indeterminacy, and the fate of any trajectory depends solely on its initial condition.

4.1.3 Winner-take-all behavior

In the model, the competition between the two neural groups arises from the negative feedback of inhibitory neurons. Upon stimulation, the input to both groups increases, causing selective cells to fire at higher rates. This activation leads to an increase of the synaptic drive to inhibitory cells, which respond discharging at a higher rate. As a result, excitatory cells receive more nonselective inhibitory inputs, which tend to compensate the initial increase of excitatory activity. When the amplitude of recurrent currents is large enough, stimulation induces a symmetry breaking mechanism that forces the system to undergo a transition from the neutral state to one of the two decision states.

In a decision state, the cells from the winning population receive suprathreshold currents due to the increase of recurrent excitatory currents. In contrast, cells from the losing population receive less recurrent excitatory currents because they fire at lower rates and because the connections they receive from activated cells are weak. This diminished excitatory drive cannot compensate the strong inhibitory feedback caused by the activation of the winning population. As a result, the input feeding non-activated selective cells is effectively subthreshold, and cells fire at low rates [43]. It is important to remark that for this state to be stable it is necessary that feedback inhibition acts sufficiently fast, otherwise oscillatory solutions may appear [75, 51, 44]. This is actually the case in the model, as GABA_A-mediated inhibitory currents have a synaptic time scale one order of magnitude shorter than that of NMDA receptors.

4.2 Agreement with experimental data

The decision model by Wang reproduces some prominent features of the behavioral and neurophysiological data obtained during visual-saccadic decision tasks. First, the model is capable of forming a categorical choice. The elevated activity of the winning neural population reaches a stereotypical level that is independent of the stimulus, and that persists after the removal of it, in accordance with the delay

activity patterns observed in LIP neurons during delayed-response versions of the random dot task [164]. These properties stem naturally from the multistability between discrete attractors of the network, and are consistent with the view that the categorical representation of the choice is independent of the details of the inputs used in the deliberation. Second, the trial-averaged activity of the neurons selective for the correct saccadic target grows at a rate proportional to the coherence of the stimulus.

Another virtue of the attractor model is that replicates the violation of time shift invariance [196]. If LIP neurons are truly integrating —in the mathematical sense— the inputs provided by MT, the effect of a brief perturbation in the random dot stimulus should be independent of the time at which this perturbation is applied. To test this hypothesis, Huk and Shadlen analyzed the effect that such brief perturbation had on behavioral and neuronal responses [107]. The perturbation consisted in the presentation with some variable delay of a brief pulse of coherent motion superimposed on the random dot stimulus. The direction of motion of the brief pulse could be either the same or the opposite to the coherent motion of the random dots. Although the subject was instructed to attend the random dot stimulus and ignore the brief pulse, both the response of LIP neurons and the behavioral performances were significantly modulated by the sign of the pulse. Moreover, the impact of the pulse was larger when its onset time was earlier, a clear violation of the time shift invariance and completely opposite to what a leaky integrator model would predict [196].

On the behavioral side, the model replicates some of the most prominent features of the performances and reaction times observed experimentally. First, the psychometric functions derived from the model agree with those obtained experimentally, with similar parameters. The psychometric function is defined as the fraction p of correct trials as a function of the coherence, and is usually fit with a Weibull function

$$p(c) = 1 - 0.5 \exp\left(-[c'/\alpha]^{\beta}\right), \tag{4.1}$$

where c' is the coherence, α is the coherence threshold, defined as the value for c' for which the fraction of correct trials is 82%, and β is related with the slope of the psychometric curve at zero coherence. With the set of parameters for spiking network model used in [197], the coherence threshold is $\alpha = 8.4\%$ and the slope, $\beta = 1.6$. These values are comparable to the experimental ones $\alpha = 7.4\%$ and $\beta = 1.25$ [164].

Second, the coherence threshold decreases with the duration of the stimulus presentation in a similar way to that observed in experiments [33], with coherence thresholds around 50% for stimuli 250 ms long to coherence thresholds lower than 10% for stimuli 2 s long [191]. Third, decision times are decreasing functions of the coherence. This observation is compatible with the integration of sensory evidence that hypothetically LIP neurons perform during the deliberation, before the decision is made. Higher coherences lead, according to this hypothesis, to steeper ramping activities, and therefore to faster accumulation of evidences that result in shorter decision times. The model also reproduces the experimental observation that decisions are slower in incorrect trials than in correct trials [164].

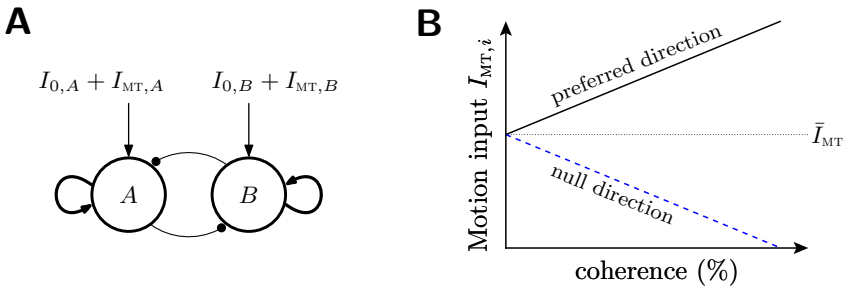


Figure 4.3: **A**: Two competing neural units are selective for leftward (A) and rightward (B) direction. Self coupling is strong and excitatory (arrows), while cross coupling is inhibitory (circleheaded arrows). The external inputs consist of motion-selective signals from area MT, $I_{\text{MT},i}$, and nonselective background input $I_{0,i}$ ($i = A, B$). **B**: The motion-selective inputs from MT are linearly increasing (decreasing) functions of the motion coherence when motion is in the preferred (null) direction of the cell.

This is a nontrivial feature of the attractor model that the simplest diffusion model fails to replicate, unless some trial-to-trial variability is introduced in the drift term of the model [157].

4.3 Two-variable rate model of decision

To understand better the dynamics of the original decision network, the authors of [197] derived from the network model a two-variable neural model using semianalytical arguments. The reduced model captures the main features of the original spiking neural network model, and is more amenable to analysis than the mean field approximation of the spiking network. Although the contents of this section lie somewhat outside the main line of discussion, we have decided to include them because the reduced model provides a simple yet rigorous picture of many aspects of the dynamics of models we will encounter in the next chapters.

The reduction consists in three steps. First, since nonselective neurons fire at constant and low rates under a wide range of parameter conditions, their dynamics can be safely ignored. Second, the transfer function of inhibitory neurons is assumed to be linear, which spare us the self-consistency condition for the inhibitory population, simplifying therefore the equations. And third, the evolution of the system is governed by the slow dynamics of NMDA gating variables, which are the dynamical variables with the longest effective time constant. We refer to the original article for the details of the derivation.

Two competing neural populations, $i = A, B$ are selective for leftward (A) and rightward (B) direction of motion 4.3. The firing rate of the neural population i receiving synaptic input I_i obeys the following current-frequency relation [1]

$$\nu_i = f(x_i) = \frac{x_i}{1 - \exp(-dx_i)}\,, \qquad \text{where} \qquad x_i = aI_i - b\,.$$

Parameters a, b, and c are fitted to the Ricciardi formula for first-passage time of a LIF neuron receiving noisy input, Eq. (3.12). Their values are $a = 270\,\mathrm{Hz/nA}$, $b = 108\,\mathrm{Hz}$, and $d = 0.145\,\mathrm{s}$. Given that in the original network model recurrent excitation is mostly mediated by NMDA receptors, synaptic drive is approximated to be given by the total NMDA gating variable S_i. The total synaptic current to each neural population is then

$$I_A = w\big(J_+ S_A - J_- S_B\big) + I_{\mathrm{MT},A} + I_0\,, \tag{4.2a}$$

$$I_B = w\big(J_+ S_B - J_- S_A\big) + I_{\mathrm{MT},B} + I_0\,, \tag{4.2b}$$

where $w = 1$ is a scale parameter of recurrent coupling, and $J_+ = 0.3725\,\mathrm{nA}$ and $J_- = 0.1137\,\mathrm{nA}$ are the self- and cross-coupling strengths, respectively. The inputs $I_{\mathrm{MT},i}$ represent the signal carried by the projections from area MT, and are modeled as linear functions of the coherence

$$I_{\mathrm{MT},i} = \bar{I}_{\mathrm{MT}}\Big(1 \pm f\frac{c'}{100\%}\Big)\,, \quad i = A, B\,, \tag{4.3}$$

where $\bar{I}_{\mathrm{MT}} = 0.0292\,\mathrm{nA}$ is the overall motion input, $f = 0.45$, and c' is the coherence of the motion stimulus. The sign $+$ or $-$ refers to the neural population for which the motion stimulus is in the preferred or null direction, respectively. The input $I_0 = 0.3297\,\mathrm{nA}$ is a tonic background input. The response of NMDA gating variables to inputs is described by

$$\dot{S}_i = -S_i/\tau_{\mathrm{NDMA}} + (1 - S_i)\gamma f(I_i)\,, \quad i = A, B\,, \tag{4.4}$$

where $\gamma = 0.641$ and $\tau_{\mathrm{NDMA}} = 60\,\mathrm{ms}$ is the time constant of NMDA receptors. Note that the gating variables S_i are the truly dynamical variables, from which the rates are computed via the transduction function f, $\nu_i = f\big(I_i(S_A, S_B)\big)$.

The system (4.4) of 2-coupled 'rate' equations can be analyzed using the standard tools from nonlinear dynamics theory. In particular, since the system is two-dimensional, a phase-plane analysis can be carried out [180]. Let us first determine the nullclines of the two-dimensional system. The S_A-nullcline is the curve $\dot{S}_A \equiv G_A(S_A, S_B) = 0$. It partitions the phase plane (S_A, S_B) into two regions where the dynamical variable S_A moves in opposite directions. The S_B-nullcline is defined analogously. The intersections of the S_A and S_B nullclines are the fixed points of the system which can be either stable or unstable depending on how the nullclines intersect with each other. The nullclines and fixed points of the *rates* ν_A and ν_B are qualitatively similar to the nullclines and fixed points of the gate variables S_A and S_B[1]

[1] The nullclines for ν_A and ν_B are obtained from those for S_A and S_B applying the chain rule

$$\begin{pmatrix}\dot{\nu}_A\\ \dot{\nu}_B\end{pmatrix} = \begin{pmatrix}f'(I_A) & 0\\ 0 & f'(I_B)\end{pmatrix}\begin{pmatrix}J_{AA} & J_{AB}\\ J_{BA} & J_{BB}\end{pmatrix}\begin{pmatrix}\dot{S}_A\\ \dot{S}_B\end{pmatrix} \overset{!}{=} 0. \tag{*}$$

The first square matrix is nonsingular because $f(I)$ is a monotonically increasing function. The second matrix is also nonsingular with the given values of the coupling strengths. Therefore, the equality (*) holds only if $(\dot{S}_A, \dot{S}_B) = (0, 0)$. That is, the ν_A- and ν_B-nullclines consist of a coordinate-dependent scaling (first matrix in Eq. (*)) times and a coordinate-independent scaling (second matrix) of the S_A- and S_B-nullclines.

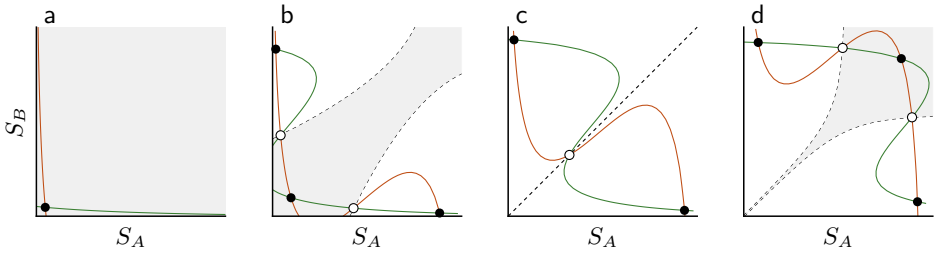


Figure 4.4: Reconfiguration of the phase portrait by the external input. The red and green curves are the nullclines $\dot{S}_A = 0$ and $\dot{S}_B = 0$, respectively. Filled circles are stable fixed points, and empty circles, unstable fixed points. The area shaded in gray is the basin of attraction of the symmetric stable fixed point, with separatrices marked with dashed curves. The recurrent coupling parameter is $w = 0.4$.

The 'landscape' of fixed points of the system, or phase portrait, depends most crucially on two parameters: the common input I_0 and the recurrent coupling w. For low inputs and low recurrent coupling, the system has one single stable fixed point, characterized by equal and low firing rates in both neural groups (Fig. 4.4a). This fixed point is refered to as low-activity symmetric state, and is the equivalent to the neutral, or spontaneously active, state of the spiking neural network model. If recurrent coupling is sufficiently strong, an increase in the external input gives rise to a pair of stable fixed points that lie off the diagonal in the phase plane, i.e., with rates ν_A and ν_R differing considerably from each other (Fig. 4.4b). Accordingly, these two states are called *asymmetric* states, and are associated with the two possible categorical choices.

The appearence of the pair of asymmetric states reduces the basin of attraction of the low-activity symmetric state. This basin shrinks further when inputs increase, until it eventually disappears due to the loss of stability of the symmetric fixed point, which becomes a saddle (Fig. 4.4c). In this range, the network exhibits bistability between the two asymmetric fixed points, and acts as a winner-take all. This is not surprising, as it is known that systems with fast cross-inhibition, described by equations of the form $\dot{r}_i = F(r_i, I, \sum_{j \neq i} r_j)$, where F is an increasing function of its second argument and a decreasing function of its third argument, can generate winner-take-all behavior [76]. A high-activity symmetric state will appear if the inputs are increased beyond the winner-take-all regime (Fig. 4.4d). In this case, external inputs counteract the effect of mutual inhibition between both neural groups and the network can sustain a non-decision state, which will necessarily be a high-activity state due to the strong excitatory drive fed by external inputs. The high-activity symmetric state is the rate analogous to the 2-memory state described in 3.2.2.

The phase portrait depends also on the recurrent coupling. For low recurrent excitation, the network cannot sustain asymmetric states and hence no decisions can be made. In that case the activity of the symmetric fixed point grows smoothly and monotonically with the input feeding the network (Fig. 4.5a). When recurrent

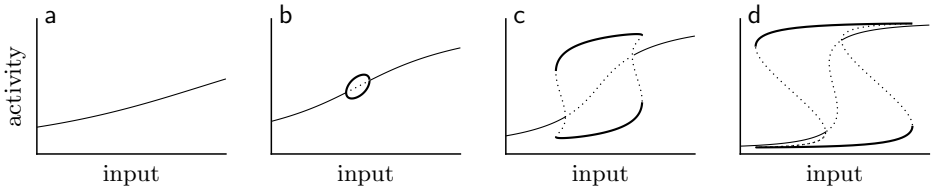


Figure 4.5: Dependence of the activity of the two neural populations on the overall external input, for 4 different values of the recurrent coupling (increasing from left to right, $w = 0.1$, 0.23, 0.3, 0.5). Thin solid curves represent the rate of the symmetric states, in which $\nu_A = \nu_B$. Thick solid curves represent the rates of selective (asymmetric) states, in which one population fires at high rate and the other fires at low rate.

coupling is increased beyond some threshold value, a new pair of stable fixed points appear through a supercritical pitchfork bifurcation as the input increases (Fig. 4.5b). If the recurrent coupling is increased even further, the two asymmetric fixed points appear instead through a pair of saddle-node bifurcations as the input is varied (Fig. 4.5c,d). The unstable branches emerging from the pair of saddle-node bifurcations, together with the stable branch of the symmetric state, coalesce as the input increases into a symmetric unstable branch, through a subcritical pitchfork bifurcation. There is therefore a range of inputs where the symmetric state coexists with the asymmetric states, within which the system can subserve working-memory (hysteresis). In that range, a transient input can push the system to the winner-take-all region of the bifurcation diagram, forcing the system to decay to either of the two asymmetric states. After the transient fades away, the system stays in the asymmetric state —the choice is 'maintained internally in working memory' [197]. This picture remains essentially unchanged when all the set of mean field equations is taken into account (i.e., when no approximations are made).

Chapter 5

Fluctuation-driven mechanism for decision

The model by Wang presented in the previous chapter was the first attempt to describe, using a large-scale cortical network model, the decision-related neuronal activity at LIP. In their subsequent articles, Wang and co-workers reduced the model and provided a dynamical description of the neuronal activity in LIP. In all cases, stimulation is supposed sufficiently strong to set the system in a winner-take-all regime, where the neutral state is no longer a stable state of the network. In this chapter we explore a different mechanism by which the system can undergo a transition to a decision state, based on a noise-driven escape from the basin of attraction of the neutral network state.

5.1 Noise-driven transitions between network states

In the decision network models seen in the previous chapter, the application of the stimulus reconfigures the phase space so that the spontaneous state, initially a stable fixed point, becomes unstable (a saddle, in the two-dimensional description). As a result, the system decays to one of the two decision states. The noise present in the network perturbs the trajectory of the system during the decay, but it is not essential for the system to undergo the transition. Thus, if it were possible to eliminate noise completely, the system would follow a deterministic trajectory that would end in the asymmetric fixed point associated with the basin of attraction in which the initial condition happened to lie. In other words, the system is driven mostly by the deterministic component of the dynamics [197] —the transition is *mean-driven*.

A different transition mechanism arises when the stimulus strength is below the critical value at which the spontaneous state becomes unstable. In that case, transitions from the spontaneous to a decision state are possible by virtue of the fluctuations present in the system. While the deterministic component of the dynamics tends to keep the system in the locally stable spontaneous state, fluctuations tend to 'liberate' the system by driving it out of the associated

basin of attraction. Once the system escapes from the basin of attraction of the spontaneous state, the deterministic part of the dynamics takes over and the system eventually decays to a decision state. Transitions are thus *fluctuation-driven*. If noise were completely eliminated from the system, the transitions from the spontaneous to the decision states would cease, and the system would stay in the spontaneous state indefinitely. Since a complete suppression of the noise present in the system cannot be done in practice, we expect that fluctuation-driven transitions play a primary role when the spontaneous state is a metastable state of the network. As we will see, this is the situation that arises when the external inputs are low.[1]

Many important processes in science involve the noise-driven escape of a system from a potential trap. The problem has been studied for more than one century in many different disciplines, giving rise to various theories of noise-activated escape (for a review see, e.g., [95]). These theories aim to provide answers to the typical problems arising in stochastic systems with multiple stable states, such as determining the mean time needed for a bistable system to switch from one state to another, or specifying the dynamics of relaxation from an unstable state, to mention just a few. In this chapter we will be concerned with the consequences on behavioral measures of noise-induced transitions in attractor-based decision models. Before we go into the network models, we illustrate the problem of noise-induced transitions as it was first posed in the physics literature.

5.2 Kramers' escape problem

The problem of escape from metastable states dates back to the 1880s, when Van't Hoff and Arrhenius found that the dissociation rates k of various electrolytes follow

$$k = A \exp(-\beta \Delta U)\,, \tag{5.1}$$

where A is a constant, $\beta = (k_B T)^{-1}$ is the inverse temperature, with k_B being the Boltzmann constant, and ΔU is the *activation* energy, the energy necessary for the reactants to dissociate. Relation (5.1) is known today as Van't Hoff-Arrhenius law [188, 19], and states that the dissociation rate is exponentially suppressed by the ratio of the activation energy to the amplitude of the thermal fluctuations. It was soon realized that many different systems involving metastable states could be well described by such law, giving birth to the discipline known as reaction rate theory [see, e.g., 192, 95].

[1] Incidentally, the distinction between mean- and fluctuation-driven regimes for decision is equivalent, yet in a completely different context, to the distinction between sub- and suprathreshold regimes of the integrate-and-fire neuron described in Section 3.1.5. The membrane voltage has a single fixed point given by the mean current $V = \mu$ (see Eq. (3.4), and Figures 3.2A–B). For subthreshold currents, $\mu < \theta$, the fixed point $V = \mu$ is a metastable state and the membrane potential reaches the threshold by means of the fluctuations in the current. This is reflected in the high variability of the ISI, with coefficients of variation close to one. As the mean input increases, the deterministic component of the dynamics takes over, the firing becomes more regular, and the CV of the ISI decreases. As we will see, decision-making models exhibit an analogous dependence on mean inputs.

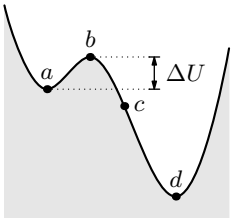


Figure 5.1: Energy potential near a metastable state. The system is initially caught in a potential well a, from which can escape by surmounting the energy barrier ΔU. The point c represents an arbitrary point of no-return: the particle can cross it in one direction but not in the other.

Kramers proposed studying the escape problem as a diffusive process in a potential [114]. A sketch of his derivation is the following. Consider an overdamped Brownian particle moving in a free energy potential with two minima a, and d, as illustrated in Fig. 5.1. By overdamped we mean that the inertia of the particle, given by $m\ddot{x}$ in the Newton equation, is negligible compared to the friction term $b\dot{x}$, so that the velocity may be taken proportional to the force

$$f(x) = -\frac{\mathrm{d}U(x)}{\mathrm{d}x} = m\ddot{x} + b\dot{x} \approx b\dot{x}\,.$$

Equivalently, $\dot{x} = \mu f(x)$, where $\mu = 1/b$ is the mobility of the particle. The Brownian motion of the particle is specified by the diffusion constant D, given by the Einstein relation

$$D = \lim_{\Delta t \to 0} \frac{\langle (\Delta X)^2 \rangle}{2\Delta t}\,,$$

and is related to the temperature and the friction through $D = k_B T/b = \mu k_B T$. The free energy at d is so low that it can be effectively regarded as an abyss, or absorbing well.

The particle, initially at a, wanders around the minimum a due to thermal fluctuations of size D. If these fluctuations are small compared to the activation energy $\Delta U = U(b) - U(a)$, the particle will spend most of its time in the vicinity of a, until the particle eventually jumps over the energy barrier and decays to the minimum at d. For convenience, we introduce a point of no-return c somewhere between the potential hill at b and the global minimum at d. The particle can cross c when it decays to the minimum d, but it cannot climb back from d to b because the energy barrier is unsurmountable. This is rigorously true only if $U(b) - U(d) = \infty$, but we expect that whenever $U(b) - U(d) \gg U(b) - U(a)$ transitions from d to a are extremely more rare than transitions from a to d. The escape problem consists of determining the rate of escape events in terms of the parameters of the system.

Kramers formulated the escape problem in terms of the Fokker-Planck equation. The probability density for the position x of the particle obeys

$$\frac{\partial}{\partial t} P(x,t) = \mu \frac{\partial}{\partial x}\big(U'(x) P(x,t)\big) + D \frac{\partial^2}{\partial x^2} P(x,t). \tag{5.2}$$

We assume that the particle is lost when it reaches point c. As we saw in Section 3.1.4, the Fokker-Planck equation can be regarded as a continuity equation

for the probability current, which in this particular case reads

$$\begin{aligned} J(x,t) &= -\mu U'(x)P(x,t) - D\frac{\partial}{\partial x}P(x,t) \\ &= -D\exp\left(-\mu\frac{U(x)}{D}\right)\frac{\partial}{\partial x}\left\{\exp\left(\mu\frac{U(x)}{D}\right)P(x,t)\right\} \\ &= -D\exp\left(-\frac{U(x)}{k_BT}\right)\frac{\partial}{\partial x}\left\{\exp\left(\frac{U(x)}{k_BT}\right)P(x,t)\right\}, \end{aligned} \tag{5.3}$$

where in the last equation we have used $\mu k_BT = D$. This current cannot be zero, as the particle eventually escapes from the potential trap. We make two further assumptions: that the current is stationary and homogeneous, namely, independent of t and x. That the current is stationary is justified by the smallness of the current itself, which does not produce appreciable changes in the probability distribution within the typical time scales of the problem. Multiplying both sides of (5.3) with $\exp\big(U(x)/k_BT\big)$ and integrating from a to c we obtain

$$D\exp\left(\frac{U(a)}{k_BT}\right)P(a,t) = J\int_a^c \exp\left(\frac{U(x)}{k_BT}\right)\mathrm{d}x\,,$$

where we have used the fact that the point c is an absorbing barrier, and therefore that $P(c,t) = 0$. The escape rate is obtained dividing the probability current J by the probability p that the particle is in the vicinity of a. We estimate p by assuming that the probability density around a is approximately given by the stationary distribution

$$P(x,t) \approx P(a,t)\exp\left(-\frac{U(x)-U(a)}{k_BT}\right).$$

This probability is now integrated in the vicinity of a, between some arbitrary, but irrelevant, limits $a_l < a$ and $a < a_t < b$

$$p = \int_{a_l}^{a_t} P(x,t)\,\mathrm{d}x = P(a,t)\exp\left(\frac{U(a)}{k_BT}\right)\int_{a_l}^{a_t}\exp\left(-\frac{U(x)}{k_BT}\right)\mathrm{d}x\,.$$

The inverse of the escape rate, called the mean escape time, reads then

$$k^{-1} \equiv T_{a\to c} = \frac{p}{J} = \frac{1}{D}\int_a^c \exp\left(\frac{U(x)}{k_BT}\right)\mathrm{d}x\int_{a_l}^{a_t}\exp\left(-\frac{U(x)}{k_BT}\right)\mathrm{d}x$$

These integrals can be approximated by a simple expression using the so-called "parabolic approximation". The factor $\exp\big(U(x)/k_BT\big)$ in the first integral is large when x is in the vicinity of b, and is exponentially smaller otherwise. Likewise, the factor $\exp\big(-U(x)/k_BT\big)$ in the second integral is large when x is around a, and exponentially smaller otherwise. We thus have

$$\begin{aligned} \exp\left(\frac{U(x)}{k_BT}\right) &\approx \exp\left(\frac{U(b)}{k_BT}\right)\exp\left(-\frac{|U''(b)|(x-b)^2}{2k_BT}\right) \quad \text{near } b\,, \\ \exp\left(-\frac{U(x)}{k_BT}\right) &\approx \exp\left(-\frac{U(a)}{k_BT}\right)\exp\left(-\frac{U''(a)(x-a)^2}{2k_BT}\right) \quad \text{near } a\,. \end{aligned}$$

Extending the limits of integration to $(-\infty, \infty)$ and evaluating the Gaussian integrals we finally get

$$T_{a\to c} = \frac{1}{\mu}\frac{2\pi}{\sqrt{U''(a)|U''(b)|}}\exp\left(\frac{\Delta U}{k_B T}\right).$$

Note that the escape time only depends on the energy gap and the curvature of the potential at the extrema a and b. The precise form of the potential is, therefore, irrelevant. Another interesting feature is that a new time scale is introduced by the exponential dependence on the height of the energy barrier, in units of the amplitude of noise. The times of escape can therefore be much longer than the natural time scales of the system. In other words, escape events are rare events, as they will only take place in the improbable case that a large fluctuation occurs. The derivation followed here has been somewhat handwaving. A more elegant derivation of the Kramers' escape rate can be done using the theory of first-passage times [187, 84]. It should be said, though, that there is no analytical solution for the problem, and one has always to rely on approximations to gain insight.

5.3 Noise-induced transitions in a decision network

A semianalytical treatment of the escape problem in the decision-making network model is impossible for two reasons. First and foremost, the system does not have an energy function and hence the sole idea of an energy barrier with a well-defined height is devoid of meaning. And second, even if the system had an energy function, we should face the problem of a noise-induced escape in a high-dimensional system. Although, there exist approximate solutions to the Kramers' escape rate problem for systems with more than one dimensions [118, 119, 129, 171], these generalizations always assume that such systems admit a potential energy, from which the equilibrium distribution can be easily derived and the escape rates can be readily estimated (see, e.g., [84], p. 363). In general, though, the dynamics of high-dimensional systems cannot be derived from a potential, and they should be studied either numerically or making additional approximations. The decision network is not an exception.

5.3.1 Effective reduction of dimensions at bifurcations

The description of the high-dimensional system can be simplified under some circumstances. All the networks models for decision discussed so far were, in essence, winner-take-all networks sharing the same bifurcation structure: when stimulation is applied, the system undergoes a bifurcation through which the neutral state becomes unstable, leaving decision states as the only stable states of the system. According to these models, stimulation should not be too strong, i.e., should not be much higher than the critical value at which this bifurcation occurs, as otherwise the network would settle in a 2-memory state (high-activity symmetric state, in the reduced model) and, therefore, no categorical choice would

be formed. Stimulation should not be too weak either, because in that case the neutral state would be too stable for the system to leave it, preventing the formation of a choice. The relevant dynamics for decision occur thus near the bifurcation point where the neutral state loses stability.

Interestingly, close to bifurcation points the dynamics of the system are confined on a low dimensional subset on the phase space, called the center manifold (see, e.g., [56, 193], and Ch. 7), where trajectories exhibit a so-called critical slowing down, evolving much more slowly than usual. The slow manifold is tangent to the *center subspace* of the critical fixed point, which is the linear space spanned by the generalized eigenvectors of the linearized system that have eigenvalues of zero real part[2]. Because the characteristic time scales of the different modes are given by the inverse of the real part of their associated eigenvalue, the evolution of the critical system is fast everywhere except along the center manifold. Moreover, the center manifold is invariant, so that all trajectories starting in the manifold stay in the manifold. The system is thus effectively decoupled from the rest of the phase space, and the dimension of the relevant dynamics, reduced to the dimension of the center space.

It is typically the case that the center manifold at a bifurcation point has dimension one. For example, in all the decision models based on winner-take-all networks described so far, the neutral state destabilizes through a pitchfork bifurcation. Since the center subspace at a bifurcation point of this type has dimension one, we can describe the system near the bifurcation point using a one-dimensional dynamical equation, with some additive noise sources if the original system was stochastic. In such cases, an energy function can be trivially defined, and one can directly apply the well-known results of first-passage times theory in one-dimensional systems to obtain the mean escape time in terms of the parameters of the original model. This was actually done in [165], where the authors applied this reduction to describe the dynamics of several winner-take-all models in terms of simpler, one-dimensional nonlinear diffusion equations, with coefficients determined by the parameters of the original model. They showed that the dynamics near the bifurcation point could be captured by the noise-driven motion in a quartic potential,

$$\dot{x} = -\frac{\mathrm{d}E(x)}{\mathrm{d}x} + L(t), \tag{5.4}$$

where

$$E(x) = \alpha x + \mu x^2 \pm x^4 \,,$$

and where $L(t)$ is a noise source obeying $\langle L(t) \rangle = 0$ and $\left\langle L(t)L(t') \right\rangle = \sigma^2 \delta(t - t')$. Although the coefficients α, μ, and σ depend in general on the parameters of each

[2] Given a dynamical system $\dot{x} = V(x, \mu)$, $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}$ with fixed point at $(x, \mu) = (0, 0)$, i.e., $V(0, 0) = 0$, the *linearized system* is the Taylor expansion of the dynamical system to first (linear) order in the dependent variable: $\dot{x} = D_x V(0, 0)x$, where $D_x V(0, 0)$ is a n-dimensional matrix with elements $D_x V(0, 0)|_{ij} = \partial V_i / \partial x_j (0, 0)$. The center subspace is $E^c = \mathrm{span}\{v \mid v \in E_\lambda \text{ and } \mathrm{Re}\,\lambda = 0\}$, where E_λ denotes the subspace spanned by the generalized eigenvectors with eigenvalue λ (see, e.g., [101] for details).

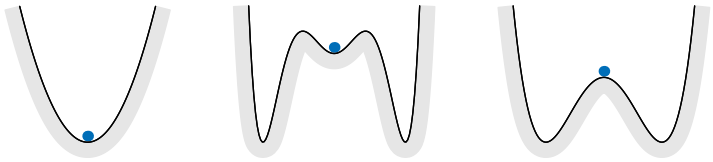


Figure 5.2: Illustration of the three possible one-dimensional decision landscapes for a generic winner-take-all network with symmetric inputs. For low recurrent excitation, the system has only one stable state, that associated with the spontaneous activity state of the network (left, one local minimum). When recurrent excitation is sufficiently high, the large positive feedback in the system allows the network to sustain two more stable states, associated with the two possible outcomes of the decision. For values of the external input lower than a certain critical value I_c the spontaneous state coexists with these two decision states (center, three local minima). For values of the external input higher than I_c, the spontaneous state is unstable (right, two local minima).

particular model, they show a universal dependence on the critical parameters. The coefficient α of the linear term is in all cases proportional to the imbalance in the inputs received by the competing populations, i.e., $\alpha \propto I_A - I_B$, whereas the coefficient of the quadratic term is proportional to the distance to criticality, $\mu \propto I - I_c$, where $I = (I_A + I_B)/2$. This parameter determines the stability of the neutral state (Fig. 5.3.1). With the explicit form for equation (5.4), the authors could derive analytical expressions for the reaction times and the performances in terms of the parameters of the original models, including the external inputs. They also showed that both the mean decision time and the performance are nonincreasing functions of $\mu \sim (I - I_c)$.

Although the center manifold reduction is strictly valid only at bifurcation points, one often sees a good agreement between the reduced and the original high-dimensional system far beyond the bifurcation occurs. In addition, the reduction also reveals the fundamental bifurcation structure of a dynamical system near the bifurcation point, and hence provides the essential information about the qualitative structure of the trajectories in the phase space. Another virtue of the method is its generality, since it can be applied to any imaginable dynamical system of any dimension. In Chapter 7 we will illustrate the reduction for a very simple model of detection.

Unfortunately, a derivation of the nonlinear diffusion equation (5.4) from the original decision network by Wang ([191]) is extremely cumbersome, and one has to rely on numerical simulations to analyze the effects of the inputs on behavioral measures. Our goal in the following sections is to address the plausibility of noise-driven mechanisms of decision in explaining behavioral and neurophysiological data. We will start analyzing the original network model by Wang with realistic synaptic currents. The analysis consists, first, in locating the regions of the parameter space susceptible to accommodate noise-induced transitions relevant for decision. Once the regions of multistability between the neutral and the decision states are identified, we run numerical simulations of the full network using the parameters selected. Decision times and performances are then estimated from

a large sample of simulated trials. To assess the generality of the noise-driven mechanism, we use in Section 5.6.2 a network which differs from the original in that the connectivity is sparse and synapses are instantaneous [8, 7].

5.4 Mean field analysis

In order to identify the attractors accessible to the network and to study their stability as a function of the parameters of the model, we use the mean field approximation derived in [43] (see Chapter 3 for a detailed description).

5.4.1 Existence and stability of network states

The mean field approximation gives rise to a set of n nonlinear equations describing the mean rate of the different populations

$$\nu_l = \phi_l(\nu_1, \ldots, \nu_n), \tag{5.5}$$

where $l = 1, \ldots, n$ labels the different neural populations, and where ϕ_l is the transduction function of population l, which gives the output firing rate of population l in terms of the inputs, see also Eqs. (3.29) and (3.31). These inputs depend in turn on the rates of all the populations. The system of equations (5.5) is just the self-consistency condition that neurons in every population produce an output that is compatible with their inputs. To solve the system (5.5), we integrate numerically the set of differential equations

$$\dot{\nu}_l = -\nu_l + \phi_l(\nu_1, \ldots, \nu_n), \quad l = 1, \ldots, n, \tag{5.6}$$

which have the same fixed point solutions as equations (5.5). This set of differential equations is sometimes referred to as the *fake-dynamics* of the system, to make clear that it is not meant to describe the real time course of the population rates.

The set of all possible fixed-points that coexist for a given parameter setting is found integrating equations (5.6) with different initial conditions [43]. To find the spontaneous activity solution, we use an initial condition of the type $\nu_{\mathrm{sl}\uparrow} = \nu_{\mathrm{sl}\downarrow} = \nu_{\mathrm{nsl}} \sim 1$ Hz, and check that the trajectory converges to a fixed point with spontaneous-like activity. Likewise, selective activity solutions were found using an initial condition $\nu_{\mathrm{sl}\uparrow} \gg \nu_{\mathrm{sl}\downarrow} = \nu_{\mathrm{nsl}} \sim 1$ Hz, and checking whether the trajectory ended up in a selective fixed-point. We also check for the existence of 2-memory states, using initial conditions of the form $\nu_{\mathrm{sl}\uparrow} = \nu_{\mathrm{sl}\downarrow} \gg \nu_{\mathrm{nsl}} \sim 1$ Hz (note that, since the network has only two populations of selective neurons, we expect to find at most 2-memory states, see p. 44).

The different stable fixed points of Eqs. (5.6) are the stationary population rates of the network. They are thus the mean rates characterizing the different attractors the network is able to sustain for a given choice of parameters. In the following we will be concerned with the concept of metastability of network states and the relative size of their basins of attraction, for which we need more information than just the fixed points of the system. This motivates the need for a phase plane description involving the rates of the two competing populations.

5.4.2 Effective population response

The total number of neural populations in the decision network is four, and therefore the fixed points characterizing the neural rates are 4-tuples of positive real numbers, $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_4) \in \mathbb{R}^4_+$. Although much lower than the dimension of the original network model, the mean field approximation involves still too many variables to get some intuition about the qualitative dynamics of the network. One possibility is to use the reduction derived in [197] and presented in the previous chapter, which makes use of some simplifying assumptions. Another possibility is to derive numerically the effective population response of the two competing populations, using the method presented in [127]. This method does not rely on any other approximation than those assumed in the mean field approximation, and is the one we use here to obtain the nullclines of the system as well as the effective flow representation. As we will see, the results are qualitatively similar to those given by the reduced model of Wong & Wang (see Sec.4.3).

The basic idea is to consider l of the n population rate variables as parameters; while keeping these l parameters fixed, we find the stationary points of the remaining $n - l$ rate variables by using Eqs. (5.6). That is, we allow the system to adapt to the stationary state induced by the l frozen variables.

Nullclines The nullclines shown in Fig. 5.3 are calculated by quenching one of the rates associated with either of the two selective populations. For example, the nullcline $\dot{\nu}_2 = 0$ is obtained by taking $\nu_1 = \bar{\nu}$ as a parameter of the system and calculating, for every value of this parameter, the solutions of

$$\begin{aligned} \nu_2 &= \phi_2\big(\mu_2(\bar{\nu}, \nu_2, \ldots, \nu_n), \sigma_2(\bar{\nu}, \nu_2, \ldots, \nu_n)\big), \\ &\vdots \\ \nu_n &= \phi_n\big(\mu_n(\bar{\nu}, \nu_2, \ldots, \nu_n), \sigma_n(\bar{\nu}, \nu_2, \ldots, \nu_n)\big). \end{aligned} \tag{5.7}$$

The values $(\nu_2, \ldots, \nu_n)$ that satisfy Eqs. (5.7) are the fixed-points of the $(n-1)$-dimensional map defined by the equations. Note that for a given value of the parameter $\bar{\nu}$ there may exist different fixed-points due to the nonlinearity of the transfer functions ϕ. The nullcline $\dot{\nu}_2 = 0$ is then obtained by plotting the values of ν_2 that one gets after solving (5.7), against the value of $\nu_1 = \bar{\nu}$, for all the values of $\bar{\nu}$ in the range considered. The nullcline $\dot{\nu}_1 = 0$ is obtained in a completely analogous way, taking ν_2 as quenched variable.

Flow The rate-flow diagrams were plot following the same principle. We covered a part of the ν_1-ν_2 plane with a grid. At each point $(\bar{\nu}_1, \bar{\nu}_2)$ of this grid we solved the $(n-2)$-dimensional fixed-point equation:

$$\begin{aligned} \nu_3 &= \phi_3\big(\mu_3(\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}'), \sigma_3(\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}')\big), \\ &\vdots \\ \nu_n &= \phi_n\big(\mu_n(\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}'), \sigma_n(\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}')\big). \end{aligned} \tag{5.8}$$

where $\boldsymbol{\nu}' = (\nu_3, \ldots, \nu_n)$ is the rate vector formed by the dynamical (not quenched) variables. For fixed rates ν_1 and ν_2, therefore, the solutions of (5.8) are the stationary points of the remaining populations induced by the rates quenched at $\nu_1 = \bar{\nu}_1$ and $\nu_2 = \bar{\nu}_2$ and by the full feedback among all the other populations. The solution depends on $\bar{\nu}_1$ and $\bar{\nu}_2$, i.e., $\boldsymbol{\nu}' = \boldsymbol{\nu}'(\bar{\nu}_1, \bar{\nu}_2)$. The currents afferent to neurons in populations 1 and 2 tend to drive them to new rates $\nu_{1,\text{out}}$ and $\nu_{2,\text{out}}$ that are in general different from the quenched values $\bar{\nu}_1$ and $\bar{\nu}_2$:

$$\nu_{1,\text{out}} = \phi_1\Big(\mu_1\big[\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}'(\bar{\nu}_1, \bar{\nu}_2)\big], \sigma_1\big[\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}'(\bar{\nu}_1, \bar{\nu}_2)\big]\Big),$$
$$\nu_{2,\text{out}} = \phi_2\Big(\mu_2\big[\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}'(\bar{\nu}_1, \bar{\nu}_2)\big], \sigma_2\big[\bar{\nu}_1, \bar{\nu}_2, \boldsymbol{\nu}'(\bar{\nu}_1, \bar{\nu}_2)\big]\Big).$$

The rate-flow diagram was obtained by drawing at each point of the grid an arrow from the point (ν_1, ν_2) to $(\nu_{1,\text{out}}, \nu_{2,\text{out}})$. For clarity we represented every arrow with a length given by $\log(1 + m/2)$, where m is the original length of the arrow, and excluded arrows whose modules were larger than 8.0.

5.4.3 Multistability between neutral and decision states

As in the reduced model, there are essentially two classes of network states: symmetric and asymmetric. The symmetric states include the spontaneously active state as well as the 2-memory state arising at high inputs. The asymmetric states encompass the two 1-memory (selective) states A and B associated with the categorical choices (see Sec. 4.1). When the inputs to both populations is identical, the system is completely symmetric with respect to the transformation $A \leftrightarrows B$, and the decision states always appear and disappear in pairs as we vary the parameters.

Figure 5.3 shows the regions where the different states are found according to the mean field approximation in the space of the specific input $\bar{\lambda}$ and the recurrent excitation w_+. Note that the conditions for existence and stability of the neutral and decision states are qualitatively the same as in Wong & Wang's model [197], where some additional approximations were made to reduce the system to two dimensions. We want to delimit the region of parameter space $(\bar{\lambda}, w_+)$ where the network shows tristability among the spontaneous and the two decision states. We will later confirm with numerical simulations that the network of spiking neurons shows tristability in approximately the same region.

Note that there are no decision states when recurrent excitation is too low, no matter how strong the input is (see the S strip on the left of the phase diagram, and the phase portrait at lower left). The network lacks in this case the minimal degree of structure to sustain decision states. Figure 5.3 also shows that the minimal amount of recurrent excitation needed to have decision states depends nonmonotonically on the input, as a consequence of the greater recruitment of shared inhibitory feedback for higher input strengths [43]. Importantly, there are regions of tristability (labeled S, A, B in the phase diagram), where the two decision states coexist with the either the neutral state or with the 2-mixed state. Note also that, although we distinguish between two different (unconnected) regions of

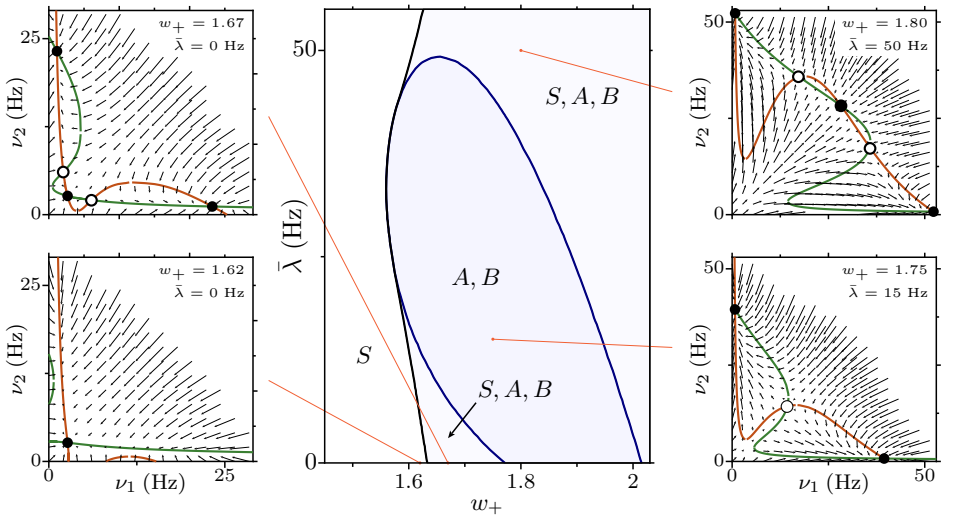


Figure 5.3: Central panel: Phase diagram of the system as given by the mean field approximation, for balanced inputs ($\Delta\lambda = 0$). In each region of the diagram, the presence of the different stable states is indicated by initials S (spontaneous and high-activity symmetric state), A (A-selective state), and B (B-selective state). In regions labeled with S, A, B all three states are simultaneously stable. Boundaries between regions correspond to bifurcation points at which either the symmetric state or the asymmetric states disappear (blue and black thick curves, respectively). Lateral panels show the fixed points, the flows, and the nullclines of the effective 2-dimensional reduction of the system (see Sec. 5.4.2), for different representative points in the phase space. Filled and empty circles denote the stable and unstable fixed points of the reduced system. Red and green curves are the nullclines $\dot{\nu}_1 = 0$ (vertical flow) and $\dot{\nu}_2 = 0$ (horizontal flow), respectively.

tristability in the phase diagram, they actually are portions of a connected region, as one would see if negative values for $\bar{\lambda}$ were included in the phase diagram. The stable symmetric states found at high enough $\bar{\lambda}$ and w_+ (rightmost S, A, B region in the phase diagram; see also the upper right figure ($w_+ = 1.80$, $\bar{\lambda} = 50\,\text{Hz}$) in the lateral panels) are characterized by firing rates considerably higher ($\gtrsim 20\,\text{Hz}$) than those associated with the spontaneous activity measured in the cortex. For this reason, we exclude this region from our analysis and concentrate on the S, A, B regime found between $w_+ = 1.6$ and $w_+ = 1.8$, for $\bar{\lambda} < 20\,\text{Hz}$ (lower center part of the phase diagram).

The average firing rates of the spontaneous and selective states as a function of the selective input $\bar{\lambda}$ are shown in Figure 5.4, for two different values of the w_+ lying in the region S, A, B aforementioned. Solid curves in the figure are calculated from the mean field approximation, while data points are obtained from simulations of the full network. The discrepancies between simulations and mean field approximation are significant close to the bifurcation values, where fluctuations around the mean rates are expected to depart from gaussianity. Yet,

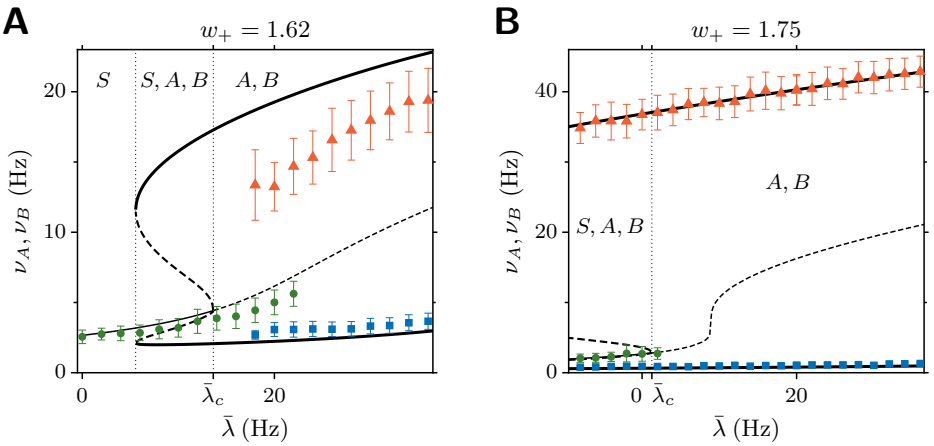


Figure 5.4: Dependence of the network activity on overall external input, as obtained from the mean field approximation (black solid curves) and from network simulations (symbols), for $w_+ = 1.62$ (**A**) and $w_+ = 1.75$ (**B**), and using balanced inputs ($\Delta\lambda = 0$). Thin solid curve: firing rates of both populations in a stable symmetric state (both populations at equal rates); thick solid curves: firing rates of the two selective populations in an asymmetric state (one population at high rate, the other at low rate). Thick dashed lines show the position of the unstable fixed point. Dotted vertical lines indicate the boundaries of the three different regimes present in the system, as predicted by the mean field approximation. At $\bar{\lambda} = \bar{\lambda}_c$ the spontaneous state loses its stability. Error bars indicate the sample standard deviation of the firing rates.

both the mean field description and the simulations show that the network is able to sustain three different states for a given range of parameters $\bar{\lambda}$ and w_+.

5.5 Finite-size noise

In this model, network states are not really stable because of finite-size effects. When recurrent conductances are scaled as $1/N$, the mean recurrent input remains constant as the size of the network is varied. The variance of recurrent inputs, in contrast, tend to zero when the size of the network is increased, so that the only source of noise in the system is provided by the external currents when $N \to \infty$. The number of spikes emitted by all neurons in a given population x during the time interval Δt is a Poisson random variable with mean and variance $N_x \nu_x \Delta t$, where N_x is the size of the population and ν_x is the corresponding mean firing rate, independent of N. For large N this random variable is well approximated by a Gaussian random variable of the same mean and variance. Thus, the estimated rate for the neural population can be written as $\nu_x(t) = \nu_x + \sqrt{\nu_x/N_x}\eta(t)$, where $\eta(t)$ is a source of standard Gaussian white noise. The fluctuating part of this process is felt by all the neurons in the network due to all-to-all connectivity. It constitutes a source of *internal* noise. There are other types of fluctuations in the network, such as those carried by external inputs or, when the network is

sparse, due to the quenched randomness of the synaptic matrix. However, these fluctuations are uncorrelated from neuron to neuron, and are already taken into account by the mean field approximation through the variance of the inputs [7] (see also Section 3.1.3).

Finite-size fluctuations induce transitions between the different network states and affect the collective dynamics of the network (see, e.g., [41]). We will use network simulations to capture the effect of finite-size noise on the stability of the network states. Although it is possible to incorporate finite-size effects in a mean field treatment [41, 38, 132, 134, 177] the description becomes too cumbersome when applied to complex architectures involving more than two populations recurrently interconnected, like those used in decision-making networks (see [71] for an example using feed-forward architectures). The amplitude of finite-size effects can be controlled by using different network sizes and scaling proportionally the recurrent conductances, so that the average input current is kept constant. In the simulations shown in Sect. 5.6.2, with sparse connectivity, both the mean and the variance of the input current are kept constant as N varies.

5.6 Simulations of the network of spiking neurons

Once the ranges of parameters $\bar{\lambda}$ and w_+ for which the network shows tristability were found, we studied the statistical properties of transition times and their dependence on the network parameters. To this end we simulated, given some fixed values for the parameters $(\bar{\lambda}, w_+, N)$, 4000 trials with different random seeds, which determined the initial values for the membrane potentials and the synaptic gate variables, as well as the random realization of the external currents. With the first two parameters we controlled the regime of operation of the network (i.e., tristable or not), as well as the distance to the boundaries of the tristability range (S, A, B). By using different network sizes we modulated the amount of noise in the system.

To make the analysis simpler, and to mimic experimental conditions, we kept the value w_+ of recurrent connectivity fixed and varied only the external input $\bar{\lambda}$. The selected value of w_+ was such that the spontaneous state was stable when $\bar{\lambda} = 0$ and it was high enough to provide acceptable signal-to-noise ratios, the signal being the difference between the rate of the winning population and the rate of the spontaneous state, and the noise being the amplitude of the rate fluctuations in the winning population. The value $w_+ = 1.75$ fulfilled these two requirements. While keeping w_+ fixed, we used $\bar{\lambda}$ as a control parameter that allowed us to drive the system from the tristable regime (S, A, B) to the competition regime (A, B) as well as to control the distance to the bifurcation point.

Every simulated trial consisted of two stages. During the first (pre-stimulus) stage, spanning from 0 to 500 ms, every neuron in the network received only the baseline background input. The network remained in the spontaneous state at that stage. After this period, neurons in both selective populations received an additional signal of magnitude $\bar{\lambda}$ (see lower panel in Figure 5.5 for a representation of the protocol used and the stimulation applied). This increase in input strength

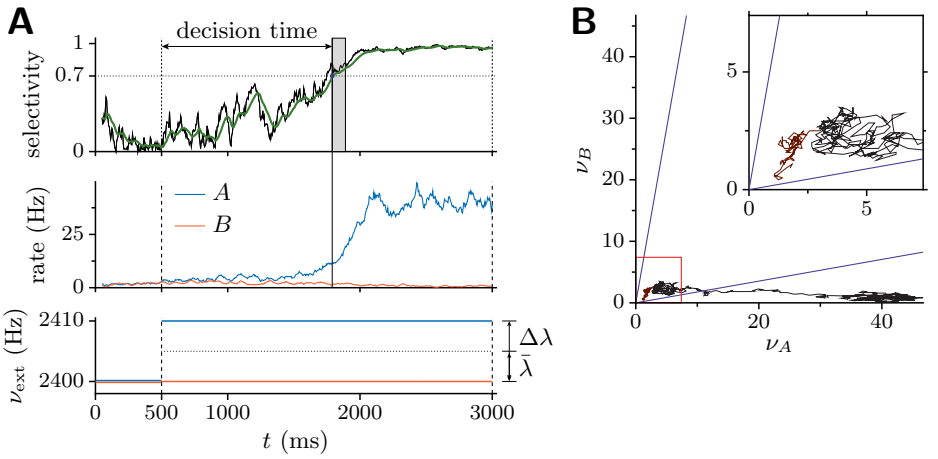


Figure 5.5: Decision neural dynamics. **A**: Evolution of the selectivity index (top), the average activity of populations A and B (middle), and stimulation applied (bottom), along a single trial. The green line in the top panel is the low-pass filtered selectivity with $\tau = 50$ ms. From 0 to 500 ms no stimulation is applied ($\bar{\lambda} = \Delta\lambda = 0$). From 500 ms to the end of the trial, $\bar{\lambda} = 5$ Hz and $\Delta\lambda = 5$ Hz ($\lambda_A = 10$ Hz, $\lambda_B = 5$ Hz). The *decision time* (DT) was the time elapsed between stimulus onset and the time at which the low-pass filtered selectivity index crossed the threshold 0.7 and stayed above it for at least 100 ms. The shaded area shows the time window within which the signal (green line) is required to be greater than the threshold. $N = 2000$, $w_+ = 1.75$. **B**: Trajectory in the phase plane (ν_A, ν_B) for the same realization represented in **A**. The trajectory from 0 to 500 ms (prestimulus stage) is plotted in brown. Blue straight lines: decision boundary $X_{\text{thr}} = |\nu_A - \nu_B|/(\nu_A + \nu_B) = 0.7$. Inset: blowup of region marked in red.

may either destabilize completely the spontaneous state or facilitate noise-induced transitions to the decision states.

The occurrence of a transition in the simulated trial was determined with the selectivity index defined as $X = |\nu_A - \nu_B|/(\nu_A + \nu_B)$. This variable provides a measure of the asymmetry between the two rates and allows to describe with a single variable the transition from the spontaneous state ($X \gtrsim 0$) to a decision state ($X \lesssim 1$, see top panel in Fig. 5.5). The selectivity index X can thus be thought of as the 'decision variable', or weight of evidence supporting one alternative over the other in the decision problem [87, 88]. Furthermore, to take into account occasional high fluctuations transiently bringing the selectivity index X above threshold, we applied a first-order low-pass filter with $\tau = 50$ ms and considered that a decision was properly formed if the filtered signal crossed the threshold $X_{\text{thr}} = 0.7$, and remained above it for at least 100 ms. We name decision time (DT) the time elapsed between stimulation onset and threshold crossing. The criterion used differs from the 'hard threshold' methodology used in [191, 197], but it leads to qualitatively similar results and it has the advantage of avoiding the use of a particular level of activity as threshold.

According to the bifurcation diagram in Fig. 5.4B, given $w_+ = 1.75$ the

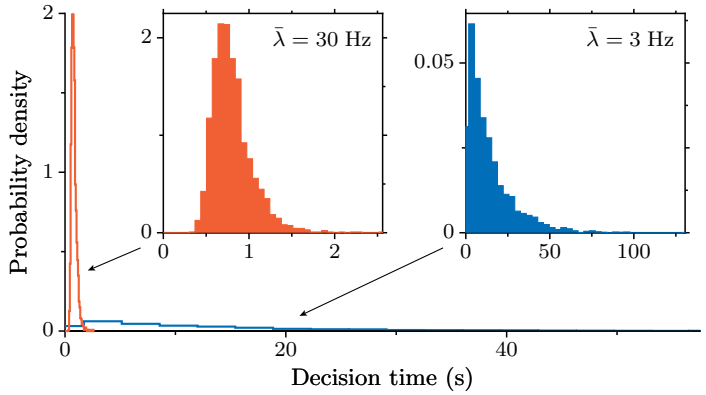


Figure 5.6: Distributions of decision times for a regime with spontaneous stable state (blue, $\bar{\lambda} = 3$ Hz) and without spontaneous stable state (red, $\bar{\lambda} = 30$ Hz), from a sample of 4000 trials each. The two insets show the distributions separately (note the different scales). $N = 4500$, $w_+ = 1.75$.

spontaneous state is stable for values of $\bar{\lambda}$ below the value $\bar{\lambda}_c = 2$Hz, approximately. Figure 5.6 shows the distribution of DTs for two values of $\bar{\lambda}$: one below (blue) and one above (red) the critical value $\bar{\lambda}_c$. For low input intensities ($\bar{\lambda} < \bar{\lambda}_c$) transitions between network states are fluctuation-driven, and the distribution of transition times is very skewed right, close to an exponential or a gamma with very low shape parameter. In contrast, high enough input intensities lead to transition times that are significantly shorter, more narrowly distributed, and less right skewed as a consequence of the dominant deterministic mechanism underlying the transition [91]. The transition from a fluctuation-driven to a relaxation regime is more abrupt the lower is the presence of noise in the system. This is shown in both panels in Figure 5.7, where the mean value and coefficient of variation (CV) of decision times obtained from a simulated sample are represented as a function of the control parameter $\bar{\lambda}$ for different levels of noise. Mean decision times grow as the external input is reduced, regardless of the regime in which the network operates. Decision times are however much more sensitive to the value of $\bar{\lambda}$ in the fluctuation-driven regime than in the relaxation regime.

Second order statistics of decision times also show distinctive properties depending on the regime. The variability of decision times around the mean is measured with the coefficient of variation, $\text{CV} = \sigma_{\text{DT}}/\langle \text{DT} \rangle$, and is plotted in Fig. 5.7A. The CV of DTs tends for sufficiently large N (small noise) to the value 1 as $\bar{\lambda}$ is decreased below the bifurcation value. This asymptotic value, together with the histogram in Fig. 5.6 (blue), suggest that in this regime and in the limit of vanishing noise decisions are essentially Poisson processes, with exponentially distributed decision times. This Poissonian character is gradually lost as the external input increases and the deterministic component of the dynamics takes over the stochastic one, leading to more peaked, gamma-like DT distributions and hence to lower CV values. From Fig. 5.7A it is also seen that for $\bar{\lambda} < \bar{\lambda}_c$ the value

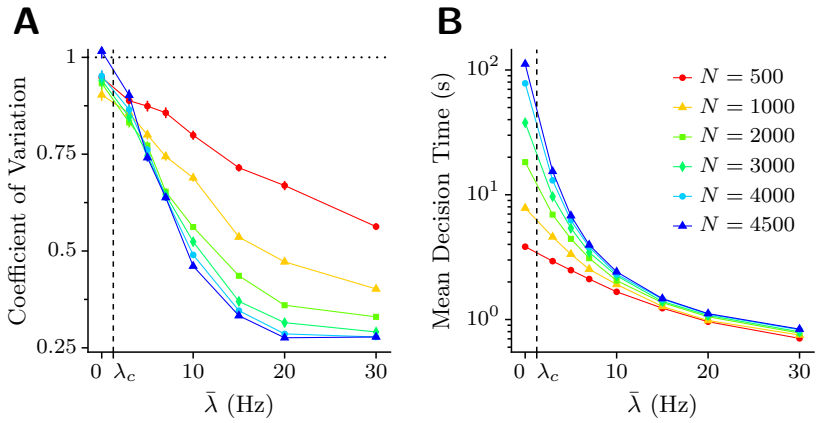


Figure 5.7: Coefficient of Variation (**A**) and mean value (**B**) of decision times versus the external input intensity $\bar{\lambda}$, for different sizes of the network, as indicated in the key. $w_+ = 1.75$.

of CV of DT is essentially insensitive to the amount of noise (while the mean value of DT strongly depends on N in the same region), consistently with the picture of an approximate Poisson statistics for the noise-driven decision process. For $\bar{\lambda} > \bar{\lambda}_c$ the converse is observed, the strong dependence of CV on $\bar{\lambda}$ being due to the fact that for increasing noise (decreasing N) the representative point in the (ν_A, ν_B) plane drops off the symmetric ridge down from the unstable spontaneous state at more widely distributed times.

As seen in section 5.2, the average escape time from a metastable state in a one-dimensional system depends exponentially on the inverse of the variance σ^2 of the fluctuations (Van't Hoff-Arrhenius law): $\langle T \rangle \sim \exp(\Delta U/\sigma^2)$, where ΔU is height of the potential barrier the system has to jump over to escape from the basin of attraction of the initial state. The law is strictly valid only in the limit of vanishing noise. For multidimensional systems it may even be impossible to define a potential function, but the general dependence on σ^2 is still of the type $\sim \exp(K/\sigma^2)$ [84, 187]. In any case, since σ^2 scales as $1/N$, decision times grow exponentially with the size of the network. As Fig. 5.8B shows, the mean DT grows approximately exponentially with N for $\bar{\lambda} < \bar{\lambda}_c$, consistent with the theory of noise-driven escape processes. Furthermore, the CV tends to one as $N \to \infty$ for $\bar{\lambda} < \bar{\lambda}_c$, while it slowly decreases with N when $\bar{\lambda} > \bar{\lambda}_c$ (Figure 5.8A). In the thermodynamic limit $N \to \infty$ the CV would decay to 0 whenever $\bar{\lambda}$ is high enough to destabilize the spontaneous state, as the transition would consist in this case on a deterministic relaxation from an unstable to a stable state.

The dynamics unfolded in this regime is compatible with the ramping-like activity observed in LIP when neuronal activity is averaged over trials [173, 164]. In the noise-driven regime, but also in the mean-driven regime, single trial activity exhibits a rather sharp transition between the spontaneous and a decision state. Such transitions have durations of about one hundred milliseconds, and reflect

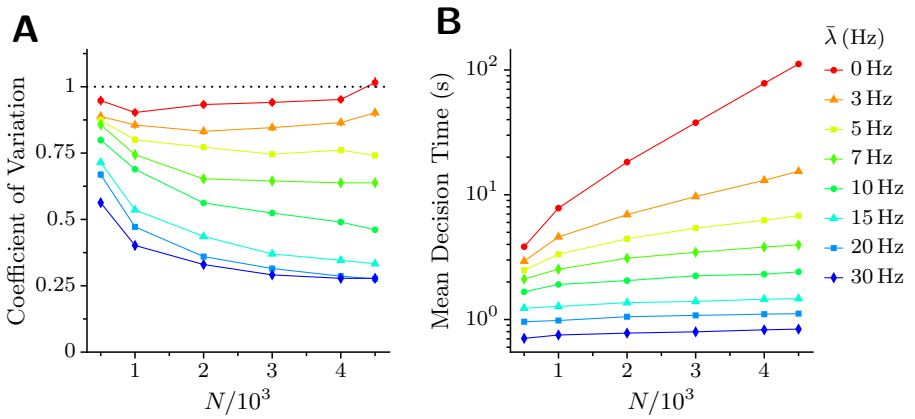
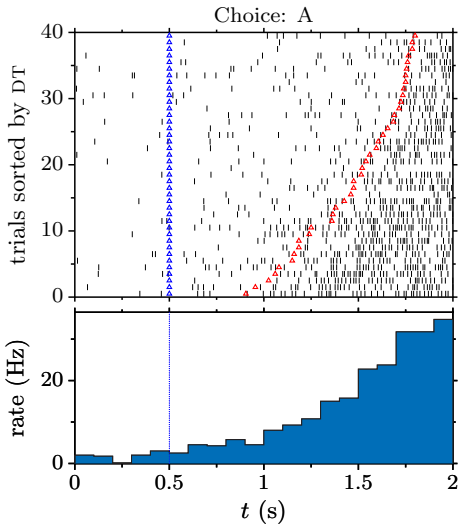


Figure 5.8: Coefficient of Variation (**A**) and mean value (**B**) of decision times versus the size of the network, for different external input intensities, as indicated in the key.

Figure 5.9: Upper row: rasters of a single cell in population A, across 40 different trials and sorted by decision time. Bottom row: trial-averaged firing rate. In all these trials the network chose option A —using the experimental analogue, the target choice was in the response field of the recorded cell. Blue triangles: stimulus onset. Red triangles: decision time, determined according to the criterion illustrated in Figure 5.5. Parameters: $N = 1000$, $w_+ = 1.75$, $\bar{\lambda} = 5$ Hz, $\Delta\lambda = 0$.



the decay to the final decision attractor with time constant given by the NMDA synaptic time scale. The stereotyped nature of the transitions can be seen at a neuronal level in Fig. 5.9, which shows simulated single-cell raster activity from different trials (top) and the corresponding trial-averaged activity (bottom; see also Fig. 5.5 for the population averaged activity).

It should be remarked that even if transitions are thought of as sharp, random jumps between two stereotyped levels of activity, smooth ramping activities can be obtained averaging over trials [143, 204, 151, 111]. To be specific, if we model the activity $r(t)$ as a step function with random threshold T, i.e., $r(t) = \Theta(t - T)$, where $\Theta(t)$ is the step function and T is a random variable with probability density $p(T)$, the average over trials gives rise to the cumulative density function of the

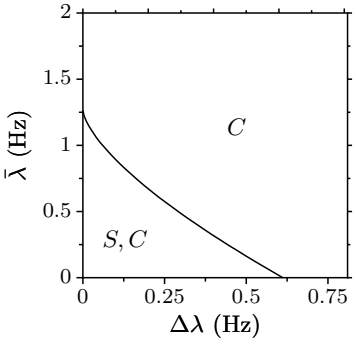


Figure 5.10: Phase diagram of the system as a function of the overall selective input $\bar{\lambda}$ and the bias $\Delta\lambda$, as given by the mean field approximation with $w_+ = 1.75$. Labels are as in Fig. 5.3: S labels regions where the spontaneous state is stable, while A and B label the regions of stability of the respective selective states. Note that the spontaneous state coexists with both selective states when $\bar{\lambda}$ and $\Delta\lambda$ are small.

decision times,

$$\langle r(t)\rangle = \int_0^\infty p(s)\Theta(t-s)\,\mathrm{d}s = \Pr(T < t).$$

The fact that realistic cumulative density functions are smooth, monotonically increasing functions would explain in this case the ramping activity observed in trial-averaged activities (see also Discussion).

5.6.1 Biased inputs

So far, we have considered only identical inputs to both neural populations. This is the situation one may expect when the coherence of the stimulus is zero, in which case the MT cells selective to either of the two possible directions of motion fire at identical rates. For nonzero coherences, the activity of MT neurons increase (decrease) if the coherent motion of the dots is in the preferred (antipreferred) direction of the cell. Indeed, it has been shown that the coherence of the stimulus is encoded approximately linearly in the rates of the direction-selective neurons in MT [34]. Based on these observations, we assume that the difference in the inputs received by populations A and B depends linearly on the coherence c', i.e., $\Delta\lambda \propto c'$. Apart from motion-selective inputs, cells in populations A and B also receive inputs for the visual signals used to indicate the monkey the two possible saccadic targets.These target inputs are equal for both neural groups and can be absorbed into the overall selective input $\bar{\lambda}$ [196].

We now follow the same line of analysis done in the previous section. First, we determine the region in the parameter space $(\bar{\lambda}, \Delta\lambda)$ where the network exhibits, according the mean field approximation, multistability between the neutral and the decision states (Fig. 5.10). The recurrent strength w_+ is assumed to be fixed. We then choose two different values for the overall input $\bar{\lambda}$. For one of these values, $\Delta\lambda$ can sweep a finite range of the tristability region (S, A, B), within which the network is expected to operate driven by noise. A possible value satisfying this condition is $\bar{\lambda} = 0$ Hz. For the other value of $\bar{\lambda}$, the network operates exclusively in the winner-take-all regime (A, B) at all values of $\Delta\lambda$. We chose $\bar{\lambda} = 10$ Hz. We then simulated different sets of trials with different values $\Delta\lambda$ for both values of λ. The results are summarized in Fig. 5.11.

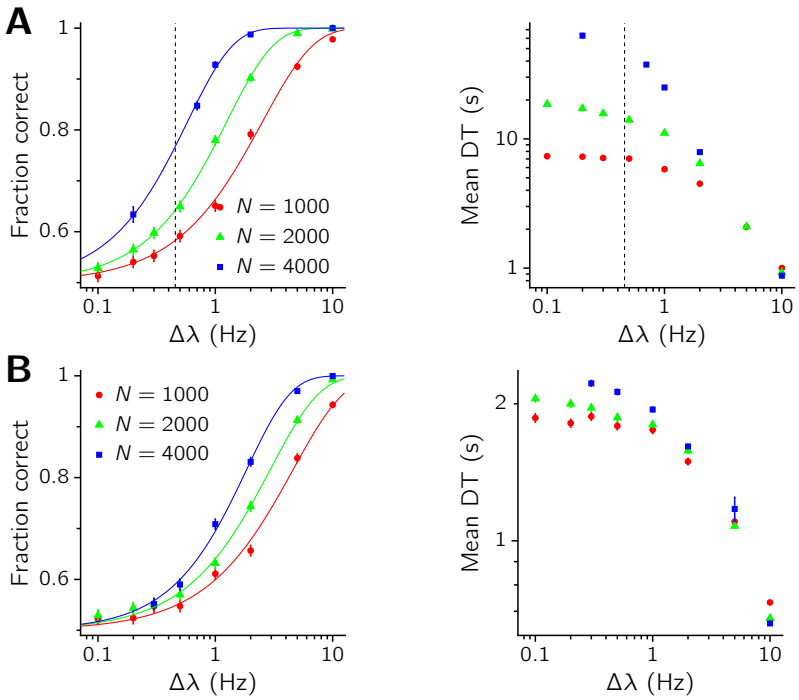


Figure 5.11: Performances and mean decision times as a function of the bias in the inputs, $\Delta\lambda$, for $w = 1.75$. **A**: The overall selective input is $\bar{\lambda} = 0$. Dashed vertical lines indicate the value of $\Delta\lambda$ where the spontaneous state becomes unstable according to the mean field approximation (see also Fig. 5.10). **B**: $\bar{\lambda} = 10$ Hz, which is far above the critical value $\bar{\lambda}_c$ at which the spontaneous state destabilizes.

Remarkably, performances in the noise-driven regimes (low $\bar{\lambda}$) are significantly better than in the mean-driven counterpart. The bias threshold $\Delta\lambda_{\rm thr}$, defined as the value of $\Delta\lambda$ at which the Weibull fit reaches $p = 1 - 0.5\exp(-1) \approx 81.6\%$ of correct choices, is systematically lower in the noise-driven regime than in the mean-driven regime (see Fig. 5.12) for all network sizes, although a more systematic check should be certainly carried out to assess the generality of this claim. The ability of the decision network to discriminate differences in the selective inputs seems therefore to improve when it operates in a noise-driven regime. Moreover, simulations suggest that performance improves when the amplitude of noise decreases, both for the mean- and the noise-driven regimes. On the other hand, decision times decrease as a function of the overall input, as seen in the previous section. The mono-

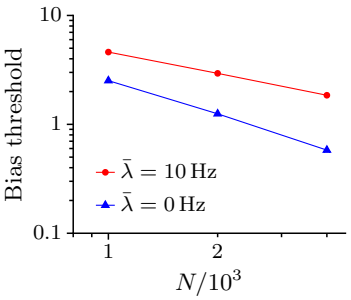


Figure 5.12

tonic increase of both performance and decision times with the overall input has been proved analytically for values of the input close to the bifurcation point where the neutral state destabilizes [165], and constitutes a plausible neurobiological mechanism for the tradeoff between speed and accuracy observed in behavioral experiments (see also Discussion).

We show in the next section that the picture emerging from Figs. 5.7 and 5.8 is recovered in simulating sparse networks of simpler synaptic and neural elements, where larger N intervals have been explored.

5.6.2 Sparse network

In this section we briefly discuss the results of an analysis similar to the one performed in the previous sections, but carried out in the context of a simpler network model. Specifically, the network is again composed of four populations of leaky integrate-and-fire neurons, with the same architecture as in the previous sections, (for N neurons, 12% of N belong to each of the selective, A and B populations, 20% of N are inhibitory neurons, and 56% of N are background, nonselective excitatory neurons) with the following differences:

1. synaptic transmission is instantaneous. Thus, the dynamics of AMPA, NMDA and GABA receptors are completely ignored.
2. the connectivity is sparse. Every neuron in the network receives spikes from a fixed number of presynaptic neurons, randomly chosen at the beginning of the simulation. Hence, no topology is imposed on the network structure. This random choice of synaptic connectivity provides a source of 'quenched' noise, such that simulations run for the same set of parameters and the same stimulation protocol embody different realizations of the statistical distribution of synaptic contacts.
3. spikes are propagated to their postsynaptic targets with a delay δ. The values of δ are drawn from an exponential distribution with mean value $\langle\delta\rangle = 11.3$ms for spikes generated by excitatory neurons, and $\langle\delta\rangle = 1.2$ ms for inhibitory spikes. A distribution of spike transmission delays is a physiologically plausible feature to incorporate, and contributes to make the states of asynchronous activity of the network more stable, tempering the propensity to ignite global oscillations [133]. We remark that the longest delays between excitatory neurons are much smaller than the characteristic time of NMDA conductances.

The values of the synaptic efficacies are chosen such that the unstructured network ($w_+ = w_- = 1$) possesses a stable state of spontaneous activity with $\nu_E = 3$ Hz and $\nu_I = 6$ Hz for the excitatory and the inhibitory neurons, respectively. Along the lines of the previous sections, the symmetry between the recurrent excitation and the cross-excitation in the populations A and B is broken by choosing $w_+ > 1$ and $w_- < 1$ in such a way as to support three fixed points in the network (S, A, B).

The purpose of this stage of analysis is to illustrate how the noise-driven mechanism envisaged is able per se to account for slow decision processes in the

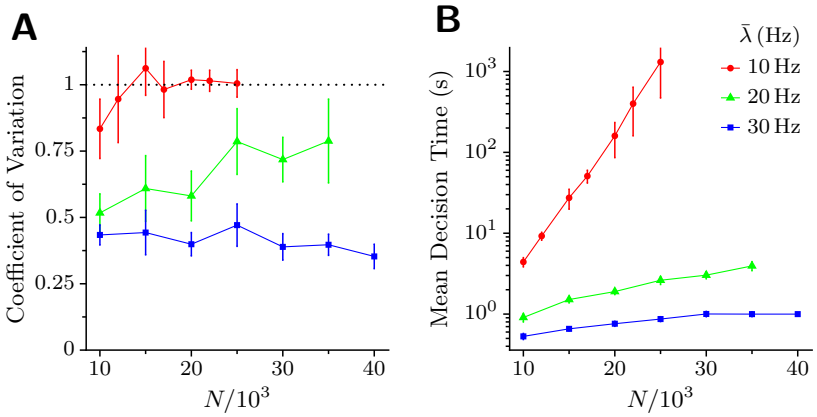


Figure 5.13: **A**: Coefficient of variation CV of Decision times, DT vs size of the sparse network N. **B**: Mean DTs vs N.

simplest network model, implicitly checking whether the characteristic times of the synaptic transmission, included as realistic features in the Brunel-Wang model adopted in the previous sections, are essential in allowing the network to exhibit such a wide range of DTs. We will show that, indeed, mean DTs obtained from the simplified network studied in this section also extend to very high values compared to all synaptic times, including those associated with NMDA.

The stimulation protocol and the estimate of DT are the same as in the previous sections. For a given set of network parameters, statistics on the DT is accumulated over 200 simulations. Consistently with theoretical expectation, the long tail of the distribution of DTs is well fitted by an exponential. Results are summarized in Fig. 5.13 in terms of the mean decision times and their coefficient of variation with respect to the network size N, for three values of $\bar{\lambda} = 10$ Hz, 20 Hz, and 30 Hz ($w_+ = 1.41$). According to the mean field approximation the spontaneous state is stable for the first two values of $\bar{\lambda}$ and unstable for the third. From Fig. 5.13B it is seen that when the stimulus intensity $\bar{\lambda}$ is such as to keep the spontaneous state stable, the mean DT is very close to be exponential in N, confirming the scenario of noise-driven transitions reported in the previous section. For strong stimuli (see the case $\bar{\lambda} = 30$ Hz in the figure) mean DTs show a very mild dependence on N, confirming the quasi-deterministic nature of the motion. Furthermore, as Fig. 5.13A shows, the CV of the noise-driven escape events ($\bar{\lambda} = 10$ Hz and 20 Hz) tends to 1 for sufficiently small noise (large N) thereby signaling an asymptotic Poisson behavior, which is consistent with known theoretical results on the distribution of residence times in noisy bistable systems. For $\bar{\lambda} = 30$ Hz CV stays approximately constant for the whole of the range explored for N, with a slight decrease for high N. Indeed, when the stimulus destabilizes the initial spontaneous state and the noise is very small, intuition suggests that the CV of DT should tend to zero (for a symmetric landscape in the phase space).

The simplification introduced in the present section aims to show that in the mechanism we adopted, of a noise-driven escape from the spontaneous state as the dynamical underpinning of the decision process, very slow decision times can be easily obtained as a result of the interaction between the finite-size noise and the cooperative-competitive dynamics of the system.

5.7 Discussion

A prominent problem in psychology is to explain the duration and variability of response times [138, 123, 94, 164], with distributions exhibiting long tails that can be approximated as exponential (see e.g. [45]). In this work we have shown that stimulus-triggered, noise-driven mechanisms among stable, asynchronous network states operating in simple neural architectures can accommodate such variability without the need for fine-tuning the parameters of the neuron and synapse models, and constitute a plausible and robust mechanism for spanning behavioral times further beyond synaptic and cellular time scales. Predictions amenable to experimental check include those related to the skewness of the distribution of decision times for both the relaxation and the noise-driven regimes upon presentation of zero-coherence random-dot stimuli.

The idea that noise-induced transitions among attractors could underlie some aspects of cognitive process is not new. It has been proposed to account for the variability of dominance times in perceptual rivalry [93, 160, 30, 144], and also to explain the neurophysiological and behavioral signatures of Weber's law, in the context of a two-forced choice vibrotactile task [64] or timing [150, 151]. The present work addresses the implications that noise-induced mechanisms entail on the dynamics of decision formation.

Attractor models of decision-making describe the decision process as a transition from a neutral (spontaneous) network state to another state associated with a given categorical choice —a 'decision' state. Noise, a constitutive ingredient of these models, is responsible for the variability in response times and for the probabilistic outcome of the decision process. The effect that noise has on the system, however, depends strongly on the regime of operation of the system. Previous works on attractor-based models of decision have focused on a particular dynamical regime, that associated to strong stimulation, in which the network acts as a winner-take-all and decision states are the only stable states of the system. In such situation, the initial neutral state is destabilized upon stimulation, and the system is forced to relax to either one of the decision states. Noise does not induce transitions by itself but rather introduces some randomness in the decision process by perturbing the deterministic drift towards the either one of the decision states. The system is specially sensitive to the effects of noise when it is close to the separatrix delimiting the basin of attraction of the decision states. Once the system leaves this region, the deterministic component of the dynamics takes over and makes the system decay to the final attractor.

A different mechanism for decision arises when the mean input is low and does not destabilize the neutral state of the network. In this case, decision and neutral

states coexist together, and the transition dynamics between states are genuinely different from those associated with the relaxation mechanism. The existence of such multistable regime is a plausible assumption supported by the observation of delay activity during delay-response versions of the random dot discrimination task [173, 164]. This delay-activity patterns are clear evidence that the network can sustain a spontaneous activity state as well as a decision state when stimulation is turned off. It is also reasonable to assume that this multistability is not destroyed when the inputs are low enough, because of structural stability. Under these conditions, noise plays a primary role in the decision process by letting the system escape from the initial basin of attraction of the neutral spontaneous state to either of the existing decision states.

When the external inputs are no longer balanced, the attractor favored by external stimulation is unambiguously assigned to the correct choice. In such case, the basin of the attractor associated with the correct response grows at the expense of the other basins. This change in the attractor landscape results in a greater probability of choosing the option favored by the inputs [197]. Regardless of these modifications in the attractor landscape, the spontaneous state will remain stable as long as the inputs (balanced or not) are low enough not to destabilize it. Our numerical simulations suggest that performances improve as the overall selective input is decreased. That is, the more the system is set to operate in the regime where transitions are mainly noise-driven, the better is the network in discriminating imbalances in the selective inputs. A plausible, heuristic explanation for this improvement in sensitivity is that in noise-driven regimes the probability of transition to the incorrect attractor is exponentially suppressed with respect to the probability of transition to the correct attractor, due to the Van't Hoff-Arrhenius activation factor. In fact, the reduction made by [165] around the bifurcation point where the neutral state destabilizes shows that for subcritical currents, the decision process can be described as the motion of a Brownian particle in an inverted double-well potential (see Fig. 5.3.1, center). When selective inputs are asymmetric, the potential is tilted and the two barriers have no longer the same height, so that the probability of escaping through the lower barrier becomes exponentially larger than the probability of escaping through the other[3]. This exponential dependence renders the system extremely sensitive to differences in the inputs, yet at the price of long decision times. The fact that both speed and accuracy are monotonically decreasing functions of the overall input [165] constitutes a plausible mechanism accounting for the speed-accuracy tradeoff observed behaviorally.

The main prediction that follows from this work is the existence of two distinguishable decision behaviors depending on the mean input feeding the decision network. For low inputs, the dynamics governing the decision process are mainly noise-driven and characterized, in the limit of vanishing noise, by exponentially distributed decision times, with coefficients of variation close to

[3]In a one-dimensional problem, the probability that a diffusive system in a potential $U(x)$ starting at x_0 reaches first a one of the two limits of the interval $I = [l, r]$, with $x_0 \in I$ is given by the *splitting probabilities* [187, 84], $\pi_l(x_0)$ and $\pi_r(x_0)$, whose ratio can be shown to be of the form $\pi_l(x_0)/\pi_r(x_0) \sim \exp[U(x_r) - U(x_l)]$.

1, and mean values that can be substantially larger than neuronal and synaptic time scales. As inputs increase decision times are less variable and decrease monotonically. These predictions become more exact the smaller is the amplitude of the noise present in the system.

Several factors may contribute to the modulation of the overall afferents to LIP. The neuronal activity in LIP is affected not only by the motion information provided by the projections from MT, but also by the temporal structure of the task [173]. When the trial is short and the subject has to make a rapid decision, the neuronal activity in LIP evolves more rapidly than when trials are longer. Thus, the expectations of the subject about the duration of the trial influence the evolution of the neuronal activity in LIP. Also the behavioral value associated with each choice has been found to modulate the activity in LIP cells [153]. All these modulations are likely to be attributable to changes in the afferent activity to LIP. Interestingly, a recent analysis of the generic dynamical properties of winner-take-all networks shows that modulations in the input common to both populations can account for the speed-accuracy tradeoff observed in behavioral experiments [165]. From the analysis it follows that the pre-stimulus average activity in LIP should be higher when the subject has to respond more rapidly, and it should decrease when it has to respond more accurately. If this the prediction is correct, the overall input to the decision-making network could be manipulated by instructing subjects to respond within a given time interval [166, 152]. Thus, a shortening of the time interval would result in an increase of the overall input feeding both populations, and the network would operate in a relaxation regime. Similarly, long time intervals allow for very accurate responses and presumably entail low common inputs to the decision network. If two different mechanisms for decision exist, one should observe that reaction time distributions tend to be more skewed in trials where the subject is instructed to be as accurately as possible.

Furthermore, since the final response time is a sum of the decision time and some residual (transduction, transmission, etc.) latencies, we expect response time distributions to reflect only in part the time devoted to decision formation [123]. Although these residual latencies are relatively short compared to the decision times when the discrimination is difficult, they introduce an additional source of variability in the final response time, whose distribution necessarily reflects the indeterminacies of both decision and non-decision contributions. In this respect, exponentially distributed decision times cannot be ruled out on experimental grounds. In fact, it has been suggested that the long right tails observed in empirical response-time distributions may result from the contribution of some exponentially distributed random variable in the response time [123]. A simple description that explores this exponential contribution is the ex-Gaussian model [103, 123, 202]. In this model, response times are the sum of two independent random variables: one, exponentially distributed, represents the decision stage, while the other, normally distributed, represents the nondecision stage. The distribution of response times is given in this case by the convolution of an exponential and a normal distribution, which is an ex-Gaussian. This distribution turns out to fit surprisingly well behavioral data.

We have also showed that, for a noise-driven decision scenario, the widely

distributed decision times can give rise to a ramping profile of the trial-averaged firing rate. Such profiles have been observed in experiments involving perceptual decisions, and we suggest that sharp firing rate transitions sparsely occurring in time, as implied in the present work, might also contribute to the explanation of these observations. This is not meant to exclude ramping firing activity at the single trial level; indeed, published data would seem to provide partial support to both scenarios [107, 139]. Perceptual decisions leading to motor responses such as saccades would plausibly involve a multi-stage process, first accumulating perceptual evidence, to be later read out by downstream neurons. In a noise-driven decision scenario, ramping activities observed in peristimulus time histograms would be an artifact of averaging single trials characterized by sharp firing rate transitions for the first stage, and a genuine reflection of single trial features for the second stage.

In this work the noise source is explicitly identified as the finite-size fluctuation of the network spiking activity. As such, it does not affect the dynamics as an additional preset external random signal (as in several analysis previously proposed), but rather as a re-entrant effect of the network recurrent dynamics. While the effective number N of neurons involved in the various stages of a decision process is obviously unknown, and the predictions shown for the N-dependence of the decision times statistics cannot be directly checked, a qualitative hint might come from experiments in which different stimulation/performance conditions are thought to involve neural populations of different sizes in the same brain areas. For example, it has been suggested that the 'oblique effect', by which subjects discriminate better visual stimuli with horizontal and vertical rather than oblique orientations, may result from the overrepresentation of cardinal (horizontal/vertical) orientations in MT cells [200]. If a similar anisotropic representation of orientations is found in LIP, one could devise an experiment showing different distributions of reaction times depending on the orientation of the opposing targets in a random-dot direction discrimination task. For instance, a subject can be instructed to respond within different time intervals so that one can manipulate the speed-accuracy tradeoff [152]. The average reaction time in long duration trials should be longer for choices involving cardinal orientations than for choices in oblique orientations, by virtue of the different number of cells involved in their representation. The higher amount of noise associated with the representation of oblique orientations would also account for the 'oblique effect' itself, as larger noise amplitudes give rise to poorer performances.

Chapter 6

The method of moments

In this chapter we briefly discuss the method of moments and apply it to a simple competition rate model. The method can be applied to Markovian systems where fluctuations are small compared to the scales of the dynamical variables, in which the probability density function is a sharply peaked function at all times, and the evolution of the system can be described in an approximate way by a set of differential equations involving the first two central moments of the distribution.

The method of moments has been used to describe the time course of the dynamical variables of single-cell and spiking network models [163, 185]. Here we use the method to characterize the different quasi-stationary distributions that typically arise in a competition model. This work has been published in two different articles [61, 62]. The presentation given here differs slightly from the one given in the published work in that, instead of using Itō calculus, here we use an equivalent and hopefully more intuitive formulation based on the master equation [187].

6.1 Stochastic rate models

As we have seen in Chapter 3, it is possible to derive rigorously a population rate model from a network model of spiking neurons. For the decision network model, the description can be reduced further making some plausible approximations, giving rise to a simple 2-rate model (see Sec. 4.3). In many situations, though, rigorous derivations from the original spiking network model are not really necessary, as one may only be interested in a qualitative description of the dynamical properties of the system. This is the approach we take in this chapter. We will be concerned with simple models that retain two important ingredients of the decision models discussed in Chapters 4 and 5: the winner-take-all property and the presence of noise.

While the dynamical properties of decision models based on winner-take-all networks have been addressed in Ch. 4, here we focus on the effect of noise on the dynamics of such models. As in the reduced model discussed in Sec. 4.3, noise terms are added to the dynamical equations to model the time-varying

fluctuations that naturally arise in spiking neural networks but that are absent in the mean field formulation. The resulting equations describe a stochastic process, and as such it is completely specified by its probability density.

6.1.1 General setting

Suppose we have a system of coupled neuronal populations. Let Y be an n-dimensional Markov variable representing the firing rates of the different populations, and let y_0 be the value of Y at time $t = 0$. The evolution of the system is given by a stochastic differential equation, which can be formulated in its Langevin form (see, e.g., [84])

$$\dot{y} = A(y) + L(t), \tag{6.1}$$

where $A(y)$ is an $\mathbb{R}^n$-valued nonlinear function on $\mathbb{R}^n$, and where we have added an n-valued Langevin term $L(t)$ to describe the fluctuations in the system. These fluctuations consist of Gaussian noise sources, of zero mean and two-point correlation function $\langle L(t)L(t')\rangle = \beta^2 \mathbb{I}\delta(t-t')$, where the symbol $\mathbb{I}$ denotes the identity in the n-dimensional space.

The non-linear function $A(y)$ in Eq. (6.1) is of the type $A(y) = -y+\Phi(Wy+E)$, where W is a $n \times n$ symmetric matrix of connection weights, E is a diagonal matrix of inputs, and Φ is a vector with components $\Phi_i(\sum_{k=1}^{n} W_{ik}y_k + E_i)$; each $\Phi_i(x)$ is a monotonically increasing activation function. The form of $A(y)$ satisfies the conditions for the Cohen-Grossberg theorem [49] (see also [76]), which can therefore be applied to prove that all trajectories decay to fixed points —i.e., there are no oscillatory solutions. Moreover, we assume that the synaptic matrix W has positive entries in the diagonal and negative entries off the diagonal to ensure that the network operates as a winner-take-all (see Sec. 4.1). When fluctuations are turned off, $L(t) = 0$, the system evolves deterministically toward the stable fixed point y^* of the basin of attraction where the initial value y_0 is.

6.2 Dynamical equations for the mean and the covariance

The addition of the fluctuation term renders Y a stochastic variable, and the evolution of Y must be described in terms of its probability density function $P(y,t)$. At time $t = 0$ the quantity Y has the precise value y_0, and hence the probability density is initially $P(y,t_0) = \delta(y - y_0)$. As t increases, $P(y,t)$ will tend to the stationary distribution $P^*(y)$. The main assumption of the method of moments is that fluctuations remain small during the whole decay process. In other words, $P(y,t)$ is a sharply peaked function of y for all t. The location of this peak can, although not necessarily should, be identified with the mean value

$$\langle Y_i\rangle_t \equiv \int y_i P(y_1,\ldots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n\,.$$

The subscript t indicates that the averages are over realizations at time t. The width of the distribution is given by the diagonal elements of the covariance matrix

$$\begin{aligned}\left\langle\!\left\langle Y_i Y_j \right\rangle\!\right\rangle_t &\equiv \left\langle \left[Y_i - \langle Y_i\rangle_t\right]\left[Y_j - \langle Y_j\rangle_t\right]\right\rangle_t \\ &= \int \left[y_i - \langle Y_i\rangle_t\right]\left[y_j - \langle Y_j\rangle_t\right] P(y_1,\ldots,y_n,t)\,\mathrm{d}y_1\cdots\mathrm{d}y_n\,.\end{aligned}$$

To simplify the notation, we shall omit the subscripts whenever there is no ambiguity. So, we will write

$$\langle Y\rangle_t = \int y P(y,t)\,\mathrm{d}^n y\,,$$

and

$$\left\langle\!\left\langle Y^2 \right\rangle\!\right\rangle_t \equiv \left\langle \left[Y - \langle Y\rangle_t\right]^T \left[Y - \langle Y\rangle_t\right]\right\rangle_t = \int \left[y - \langle Y\rangle_t\right]^T \left[y - \langle Y\rangle_t\right] P(y,t)\,\mathrm{d}^n y\,,$$

and it will be understood that the mean $\langle Y\rangle_t$ is an n-dimensional vector, the covariance $\left\langle\!\left\langle Y^2 \right\rangle\!\right\rangle_t$ an $n\times n$ matrix, and integrals are over the n-dimensional sample space. Being Y a Markov random variable, its evolution is fully specified by the associated master equation, which describes the temporal evolution of the probability density of Y. Instead of deriving the master equation for Y, we will be concerned with the time evolution of the first two moments. Let us first consider the one-dimensional case. The location of the peak must obey, by virtue of the Markov character of Y,

$$\begin{aligned}\frac{\partial}{\partial t}\langle Y\rangle_t &= \int y \frac{\partial}{\partial t} P(y,t)\,\mathrm{d}y = \iint y\left\{W(y|y')P(y',t) \;- W(y'|y)P(y,t)\right\}\,\mathrm{d}y'\mathrm{d}y \\ &= \iint (y'-y) W(y'|y) P(y,t)\,\mathrm{d}y'\mathrm{d}y = \int a_1(y) P(y,t)\,\mathrm{d}y \\ &= \langle a_1(Y)\rangle_t, \end{aligned} \tag{6.2}$$

In the last equality we have introduced definition for the jump moments

$$a_\nu(y) \equiv \int (y'-y)^\nu\, W(y'|y)\,\mathrm{d}y' = \int (\Delta y)^\nu\, W(y'|y)\,\mathrm{d}(\Delta y),$$

where we have redefined y' as $y+\Delta y$. In n-dimensions, the ν-th jump moment is defined as

$$\begin{aligned} a_\nu^{j_1,\ldots,j_\nu}(y_1,\ldots,y_n) &\equiv \int (y'_{j_1} - y_{j_1})\cdots(y'_{j_\nu} - y_{j_\nu})\, W(y'|y)\,\mathrm{d}^n y' \\ &= \int \Delta y_{j_1}\ldots\Delta y_{j_\nu}\, W(y'|y)\,\mathrm{d}^n(\Delta y), \end{aligned} \tag{6.3}$$

which is nothing but

$$a_\nu^{j_1\ldots j_\nu} = \lim_{\Delta t\to 0} \langle \Delta y_1\cdots\Delta y_\nu\rangle/\Delta t\,.$$

With this definition, the n-dimensional version of equation 6.2 reads

$$\frac{\partial}{\partial t}\langle Y_i\rangle_t = \langle a_1^i(Y)\rangle_t \tag{6.4}$$

The second order moment obeys

$$\begin{aligned}\frac{\partial}{\partial t}\big\langle Y_iY_j\big\rangle_t &= \iint (y_i'y_j' - y_iy_j)\big\{W(y|y')P(y',t) \; - W(y'|y)P(y,t)\big\}\,\mathrm{d}^n y'\mathrm{d}^n y \\ &= \iint \big\{(y_i'-y_i)(y_j'-y_j) + y_i(y_j'-y_j) + y_j(y_i'-y_i)\big\} \\ &\qquad\qquad \times W(y'|y)P(y,t)\,\mathrm{d}^n y'\mathrm{d}^n y \\ &= \big\langle a_2^{ij}(Y)\big\rangle_t + \big\langle y_j a_1^i(Y) + y_i a_1^j(Y)\big\rangle_t\,. \end{aligned} \tag{6.5}$$

The equation describing the temporal evolution of the covariance $\langle\!\langle Y_iY_j\rangle\!\rangle_t = \langle Y_iY_j\rangle_t - \langle Y_i\rangle_t\langle Y_j\rangle_t$ follows from Eqs. (6.4) and (6.5) and reads

$$\frac{\partial}{\partial t}\big\langle\!\big\langle Y_iY_j\big\rangle\!\big\rangle_t = \big\langle [Y_i - \langle Y_i\rangle]a_1^j(Y)\big\rangle_t + \big\langle [Y_j - \langle Y_j\rangle]a_1^i(Y)\big\rangle_t + \big\langle a_2^{ij}(Y)\big\rangle_t. \tag{6.6}$$

6.2.1 Calculation of the jump moments

Both Eq. (6.4) and Eq. (6.6) are exact identities. To derive the dynamical equation of the moments of our particular system we need to compute the jump moments. The jump moments are computed from the definition (6.3) and the exact identity

$$\Delta y_i = \int_t^{t+\Delta t} \Big\{A_i\big[y(s)\big] + L_i(s)\Big\}\,\mathrm{d}s\,.$$

Thus, the average of Δy_i with initial value $y_i(t)$ is

$$\langle \Delta y_i\rangle = A_i\big[y(t)\big]\Delta t + o(\Delta t)\,.$$

Similarly, the mean square displacement is

$$\begin{aligned}\big\langle \Delta y_i \Delta y_j\big\rangle &= \left\langle \int_t^{t+\Delta t} \Big\{A_i\big[y(s)\big] + L_i(s)\Big\}\,\mathrm{d}s \int_t^{t+\Delta t} \Big\{A_j\big[y(s')\big] + L_j(s')\Big\}\,\mathrm{d}s' \right\rangle \\ &= \left\langle \int_t^{t+\Delta t} A_i\big[y(s)\big]\,\mathrm{d}s \int_t^{t+\Delta t} A_j\big[y(s')\big]\,\mathrm{d}s' \right\rangle \\ &\quad + \int_t^{t+\Delta t}\int_t^{t+\Delta t} \big\langle A_i\big[y(s)\big]L_j(s')\big\rangle\,\mathrm{d}s'\mathrm{d}s \\ &\quad + \int_t^{t+\Delta t}\int_t^{t+\Delta t} \big\langle A_j\big[y(s')\big]L_i(s)\big\rangle\,\mathrm{d}s'\mathrm{d}s \\ &\quad + \int_t^{t+\Delta t}\int_t^{t+\Delta t} \big\langle L_i(s)L_j(s')\big\rangle\,\mathrm{d}s'\mathrm{d}s\,. \end{aligned} \tag{6.7}$$

The first term on the right hand side is of order $(\Delta t)^2$. This can be seen expanding $A\big[y(s)\big]$ in the integrand around the initial value $A\big[y(t)\big]$,

$$\begin{aligned} A_i\big[y(s)\big] &= A_i\big[y(t) + \big(y(s) - y(t)\big)\big] \\ &\approx A_i\big[y(t)\big] + \sum_{k=1}^{n} \partial_k A_i\big[y(t)\big]\big(y_k(s) - y_k(t)\big) + \cdots , \end{aligned} \tag{6.8}$$

where ∂_j denotes partial derivative with respect to y_j. Since $y(s) - y(t)$ is a well-behaved function of s, in the sense that it contains no delta functions, the integral in the first term of Eq. (6.7) must grow linearly Δt in the limit of small time increments. To see how the second and third terms in Eq. (6.7) depend on Δt we proceed analogously. If we plug the expansion (6.8) of $A_i[y(s)]$ into the second term of Eq. (6.7) we obtain

$$\begin{aligned} \int_t^{t+\Delta t}\int_t^{t+\Delta t} \big\langle A_i\big[y(s)\big] L_j(s')\big\rangle \,\mathrm{d}s'\mathrm{d}s &= A_i\big[y(t)\big] \int_t^{t+\Delta t} \big\langle L_j(s)\big\rangle \,\mathrm{d}s \\ &+ \sum_{k=1}^{n} \partial_k A_i\big[y(t)\big] \int_t^{t+\Delta t}\int_t^{t+\Delta t} \big\langle \big(y_k(s) - y_k(t)\big) L_j(s)\big\rangle \,\mathrm{d}s'\mathrm{d}s + \cdots . \end{aligned}$$

The first term is zero from the properties of $L(t)$. The second term involves a double integral of a well-behaved function multiplied by the stochastic forcing. It must therefore be of order $(\Delta t)^2$. Going back to Eq. (6.7), only the third term depends linearly on Δt, with value $\delta_{ij}\beta^2\Delta t$. The Kronecker delta arises from the orthogonality between components of $L(t)$.

To summarize, the first two jump moments for our neurodynamical model Eq. (6.1) are

$$a_1^i(y) = \lim_{\Delta t\to 0} \frac{\langle \Delta y_i\rangle}{\Delta t} = A_i\big(y\big), \tag{6.9a}$$

$$a_2^{ij}(y) = \lim_{\Delta t\to 0} \frac{\langle \Delta y_i \Delta y_j\rangle}{\Delta t} = \beta^2\delta_{ij} \,. \tag{6.9b}$$

6.3 The macroscopic approximation

The dynamical equations for the mean and the covariance can be simplified with the assumption that fluctuations are small. In that case, the values of the stochastic variable Y distribute narrowly around $\langle Y\rangle$, so that the first jump moment appearing in Eq. (6.4) can be approximated by the lowest order terms of the expansion around $\langle Y\rangle$,

$$\begin{aligned} a_1^i(y) = a_1^i\big(\langle Y\rangle\big) &+ \sum_{k=1}^{n} \left[\frac{\partial}{\partial y_k} a_1^i\big(\langle Y\rangle\big)\right] \big[y_k - \langle Y_k\rangle\big] + \\ &\sum_{k=1}^{n}\sum_{l=1}^{n} \frac{1}{2}\left[\frac{\partial}{\partial y_k}\frac{\partial}{\partial y_l} a_1^i\big(\langle Y\rangle\big)\right] \big[y_k - \langle Y_k\rangle\big]\big[y_l - \langle Y_l\rangle\big] + \cdots . \end{aligned}$$

Therefore,

$$\langle a_1^i(Y)\rangle = a_1^i\big(\langle Y\rangle\big) + \sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{2}\left[\frac{\partial}{\partial y_k}\frac{\partial}{\partial y_l}a_1^i\big(\langle Y\rangle\big)\right]\big\langle\!\big\langle Y_k Y_l\big\rangle\!\big\rangle_t + \cdots . \tag{6.10}$$

The dynamical equation for the first moment, Eq. (6.2), reads then

$$\begin{aligned}\frac{\partial}{\partial t}\langle Y_i\rangle_t &= \langle a_1^i(Y)\rangle_t, \\ &= a_1^i\big(\langle Y\rangle_t\big) + \sum_{k=1}^{n}\sum_{l=1}^{n}\frac{1}{2}\left[\frac{\partial}{\partial y_k}\frac{\partial}{\partial y_l}a_1^i\big(\langle Y\rangle_t\big)\right]\big\langle\!\big\langle Y_k Y_l\big\rangle\!\big\rangle_t + \cdots\end{aligned} \tag{6.11}$$

While the dynamical equation for the covariance, Eq. (6.6), is

$$\begin{aligned}\frac{\partial}{\partial t}\big\langle\!\big\langle Y_i Y_j\big\rangle\!\big\rangle_t &= \big\langle [Y_i - \langle Y_i\rangle_t] a_1^j(Y)\big\rangle_t + \big\langle [Y_j - \langle Y_j\rangle_t] a_1^i(Y)\big\rangle_t + \big\langle a_2^{ij}(Y)\big\rangle_t \\ &= \sum_{k=1}^{n}\big\langle\!\big\langle Y_i Y_k\big\rangle\!\big\rangle_t \frac{\partial}{\partial x_k}a_1^j\big(\langle Y\rangle_t\big) + \sum_{k=1}^{n}\big\langle\!\big\langle Y_j Y_k\big\rangle\!\big\rangle_t \frac{\partial}{\partial x_k}a_1^i\big(\langle Y\rangle_t\big) \\ &\quad + a_2^{ij}\big(\langle Y\rangle_t\big) + \cdots\end{aligned} \tag{6.12}$$

Note that, unless the jump moments are linear functions of y, equations (6.11) and (6.12) do not form a closed system of equations, as they involve all the moments of the probability density. The macroscopic approximation assumes that the higher order terms can be neglected because they are small.

To keep the notation uncluttered we define $\mu_i(t) \equiv \langle Y_i\rangle_t$ and $\gamma_{ij}(t) \equiv \big\langle\!\big\langle Y_i Y_j\big\rangle\!\big\rangle_t$. We invoke the macroscopic approximation to truncate the expansion at second order in the expansion parameter $\Delta y = y - \langle Y\rangle$. Taking the explicit forms for the jump moments, Eqs. (6.9), and using $A(y) = -y + \Phi(Wy + E)$, we have

$$\begin{aligned}\dot{\mu}_i &= A_i(\mu) + \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}\gamma_{lk}\partial_k\partial_l A_i(\mu) \\ &= -\mu_i + \Phi_i(u_i) + \frac{1}{2}\sum_{k=1}^{n}\sum_{l=1}^{n}\gamma_{lk}W_{ik}W_{il}\Phi_i''(u_i)\,,\end{aligned} \tag{6.13}$$

where in the first line $\mu \equiv (\mu_1, \mu_2)$ and in the second line we have defined $u_i = \sum_{k=1}^{n} W_{ik}\mu_k + E_k$. Equation (6.13) states that the temporal course of the mean depends not only on the mean itself but also on the fluctuations around it.

The dynamical equation for the covariance matrix is obtained analogously,

and reads

$$\begin{aligned}\dot{\gamma}_{ij} &= \delta_{ij}\beta^2 + \sum_{k=1}^{n}\Big\{\gamma_{ik}\partial_k A_j(\mu) + \gamma_{jk}\partial_k A_i(\mu)\Big\} \\ &= \delta_{ij}\beta^2 + \sum_{k=1}^{n}\Big\{\gamma_{ik}\big(-\delta_{jk} + W_{jk}\Phi_j'(u_j)\big) \\ &\qquad + \gamma_{jk}\big(-\delta_{ik} + W_{ik}\Phi_i'(u_i)\big)\Big\} \\ &= \delta_{ij}\beta^2 - 2\gamma_{ij} + \sum_{k=1}^{n}\Big\{\gamma_{ik}W_{jk}\Phi_j'(u_j) + \gamma_{jk}W_{ik}\Phi_i'(u_i)\Big\}.\end{aligned} \tag{6.14}$$

Equations (6.13) and (6.14) form a closed set of ordinary differential equations. Note that the equations are in general nonlinear due to the factors $\Phi'\big(u_i(t)\big)$ and $\Phi''\big(u_i(t)\big)$. By describing the original stochastic system in terms of the first two moments of the probability density, we end up dealing with a system of deterministic differential equations, which can be analyzed straightforwardly. One may, for example, find the fixed points of Eqs. (6.13)–(6.14), which yield the mean and variances of the quasi-stationary distributions centered at the different fixed points of the original nonlinear system (6.1), in the limit of small noise.

It should be noted, however, that the approximation is based on the assumption that fluctuations are small, and the equations derived above should be consistent with the assumption. To be specific, one should check that all the elements in the covariance matrix remain small along the trajectory. We see from Eq. (6.14) that the variances, given by the diagonal elements of γ, tend to increase at a rate given by the diffusive term β^2, but this tendency might be balanced by the terms linear in γ present on the right hand side of Eq. (6.14). At the fixed point Eqs. (6.14) read

$$0 = \delta_{ij}\beta^2 - 2\gamma^*_{ij} + \sum_{k=1}^{n}\Big\{\gamma^*_{ik}W_{jk}\Phi_j'(u^*_j) + \gamma^*_{jk}W_{ik}\Phi_i'(u^*_i)\Big\} \equiv \delta_{ij}\beta^2 - L(t)\gamma^*_{ij} \tag{6.15}$$

where $u^* = \sum_{k=1}^{n} W_{ik}\mu^*_k + E_k$, and $L(t)$ is a linear operator whose elements depend on μ^* only. The components γ^*_{ij} are thus given by $\gamma^*_{ij} = L^{-1}\beta^2|_{ij}$ as long as L is nonsingular. Incidentally, we see from Eq. (6.12) that the form of L is basically a symmetrized version of the Jacobian matrix $DA(y)$, and hence the solution for γ^* exists only for hyperbolic points, where $\det DA(y)|_{y=y^*} \neq 0$. The condition is actually more restrictive as it requires not only the fixed point but also the whole trajectory to the fixed point to be hyperbolic (see, e.g., [193]).

6.4 Example: two-dimensional decision neural model

Let us take two neural populations with sigmoidal activation functions

$$\Phi_1(x) = \Phi_2(x) = \big(1 + \exp[-\alpha(x-\theta)]\big)^{-1},$$

and symmetric synaptic matrix

$$W = \begin{pmatrix} w_+ & w_- \\ w_- & w_+ \end{pmatrix}, \quad \text{with } w_+ > 0,\ w_- < 0\,.$$

The dynamical equations are

$$\tau \dot{y}_1 = -y_1 + \Phi(w_+ y_1 + w_- y_2 + E_1) + \sqrt{\tau} L_1(t)\,, \tag{6.16a}$$
$$\tau \dot{y}_2 = -y_2 + \Phi(w_+ y_2 + w_- y_1 + E_2) + \sqrt{\tau} L_2(t)\,. \tag{6.16b}$$

For the sake of concreteness, let us assume that $w_+ = 1.5$, $w_- = -1$, $\alpha = 4$, $\theta = 1$, and $\tau = 1$. We also assume for simplicity that the external inputs to both neural groups are identical, $E_1 = E_2 \equiv E$.

6.4.1 Noiseless limit

The system is qualitatively similar to Wong & Wang's 2-rate equation [197, 196] (see also Sec. 4.3). When the recurrent coupling is sufficiently strong, the system may exhibit multistability between symmetric and asymmetric stable fixed points depending on the common external input E. Figure 6.1A shows the dependence of the stationary rates on the input E for the deterministic system, $\beta = 0$ (cf. Figs. 4.5c–d). The two-coupled system (6.16) exhibits also winner-take-all behavior for an intermediate range of the common external input. In that range, the phase plane (y_1, y_2) is partitioned into two basins of attraction, with separatrix at the diagonal, and all trajectories decay to either of the two asymmetric states depending on which basin of attraction the system is initialized in (Fig. 6.1B). The winner-take-all regime is destroyed for sufficiently low or sufficiently high external inputs, which cause asymmetric states lose their stability. For these limit cases, symmetric states are the only stable fixed points of the system. The dependence of the phase portrait on the external input is basically the same as that seen in Fig. 4.4 (not shown).

6.4.2 Finite noise: derivation of the moment equations

When noise is finite, we can apply the method of moments and transform the original Langevin equations (6.16) to a system of five coupled deterministic differential equations relating the means and covariances. In this particular example, the dynamical equations for the two components of the mean, $\mu(t) = (\mu_1(t), \mu_2(t))$ are:

$$\dot{\mu}_1 = -\mu_1 + \frac{1}{2}\Phi''(u_1)\big[w_+^2\gamma_{11} + w_-^2\gamma_{22} + w_+ w_-(\gamma_{12} + \gamma_{21})\big]\,, \tag{6.17a}$$
$$\dot{\mu}_2 = -\mu_2 + \frac{1}{2}\Phi''(u_2)\big[w_+^2\gamma_{22} + w_-^2\gamma_{11} + w_+ w_-(\gamma_{12} + \gamma_{21})\big]\,, \tag{6.17b}$$

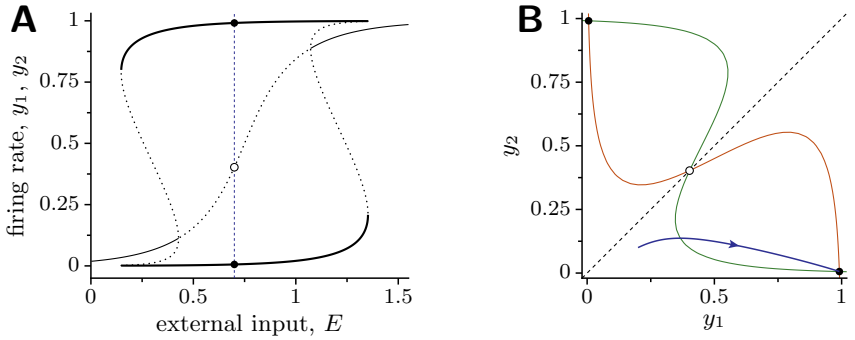


Figure 6.1: Bifurcation diagram and phase portrait of the decision neural model. Parameters: $w_+ = 1.5$, $\beta = 0$ (no noise), other parameters as in the text. **A**: Dependence of the firing rate of the two neural groups on the common input $E_1 = E_2$. Thin solid curves: rates of both groups in a symmetric solution ($y_1 = y_2$). Thick solid curves: rates of both groups in asymmetric solutions ($y_1 \gg y_2$ and vice versa). Dotted black curves: rates of the unstable solutions. Dashed blue curve: $E = 0.7$, used in B. **B**: Phase plane of the rate model for $w_+ = 1.5$ and $E = 0.7$. Red and green curves: y_1- and y_2-nullclines, respectively. Filled dots are the stable fixed points of the system, and the empty dot is a saddle. The blue curve is the deterministic trajectory starting at $(y_1, y_2) = (0.2, 0.1)$ and ending at the y_1-choice fixed point.

where we have defined $u_1 \equiv w_+\mu_1 + w_-\mu_2$ and $u_2 \equiv w_+\mu_2 + w_-\mu_1$. The dynamical equations for the elements of the covariance matrix read

$$\dot{\gamma}_{11} = \beta^2 - 2\gamma_{11} + 2\Phi'(u_1)\big(w_+\gamma_{11} + w_-\gamma_{12}\big)\,, \tag{6.18a}$$

$$\dot{\gamma}_{12} = -2\gamma_{12} + \Phi'(u_1)\big(w_+\gamma_{21} + w_-\gamma_{22}\big) + \Phi'(u_2)\big(w_+\gamma_{12} + w_-\gamma_{11}\big)\,, \tag{6.18b}$$

$$\dot{\gamma}_{21} = -2\gamma_{21} + \Phi'(u_2)\big(w_+\gamma_{12} + w_-\gamma_{11}\big) + \Phi'(u_1)\big(w_+\gamma_{21} + w_-\gamma_{22}\big)\,, \tag{6.18c}$$

$$\dot{\gamma}_{11} = \beta^2 - 2\gamma_{22} + 2\Phi'(u_2)\big(w_+\gamma_{22} + w_-\gamma_{21}\big)\,, \tag{6.18d}$$

As expected, the dynamical equations for γ_{12} and for γ_{21} are identical, and hence their trajectories coincide when their initial value is the same, as it should by definition of covariance matrix. The number of independent dynamical variables is therefore five.

The set of five differential equations (6.17)–(6.18) can be integrated numerically for a given initial mean and covariance matrix. The solution can be regarded as a trajectory $m(t)$ in the five-dimensional vector space $\mathbb{R}^5$, spanned by the dynamical variables

$$m(t) \equiv \big(\mu_1(t), \mu_2(t), \gamma_{11}(t), \gamma_{12}(t), \gamma_{22}(t)\big)\,.$$

The solution represents the time course of the parameters of the bivariate normal distribution describing the stochastic process $(y_1(t), y_2(t))$ specified by the Langevin equations (6.16). The description will be more accurate the more peaked is the density function for $(y_1(t), y_2(t))$. The existence of such bump is guaranteed by the condition that the elements of the covariance matrix remain small at all times, which will be satisfied as long as the trajectory falls far from separatrices.

6.4.3 Comparison of analytical results with simulations

To see that the solutions of the deterministic system (6.17)–(6.18) describe well the evolution of the ensemble of solutions of the original Langevin Eqs. (6.16), we integrated numerically both systems from an arbitrary initial point $y_0 = (y_{1,0}, y_{2,0})$. The initial conditions for the two-dimensional Langevin system were therefore $y(t_0) = (y_1(t_0), y_2(t_0)) = y_0$, while the initial conditions for the five-dimensional moment equations were those corresponding to a delta distribution centered at y_0, namely, $m_0(t_0) = (y_{1,0}, y_{2,0}, 0, 0, 0)$.

The parameters were chosen to set the system operate in the winner-take-all regime, although the choice was irrelevant for the validity of the method. We estimated the mean trajectory $\langle Y \rangle_t = \mu(t)$ from $N = 1000$ simulated trials with same initial conditions and different random seeds,

$$\bar{y}(t_n) = \frac{1}{N} \sum_{k=1}^{N} y^{(k)}(t_n) = \frac{1}{N} \sum_{k=1}^{N} \big(y_1^{(k)}(t_n), y_2^{(k)}(t_n)\big),$$

where

$$y^{(k)}(t_0) = y_0 \text{ for all } k.$$

The estimated mean trajectory was then compared with the solution of the moment equations, $(\mu_1(t), \mu_2(t))$ (see Fig. 6.2A). When the initial condition y_0 is not too close to the separatrices, both the estimated trajectory $\bar{y}(t)$ and the trajectory $\mu(t)$ computed from Eq. (6.13) are hardly distinguishable from one another, and they both converge to the same locally stable attractor. When y_0 is close to a separatrix, however, there is a finite probability that fluctuations carry the system out of its initial domain of attraction, driving the system toward a different attractor from that expected according to the deterministic limit. In such case, the probability density describing the evolution of the ensemble splits in two autonomous bumps, and the small noise assumption we invoked to apply the macroscopic approximation cannot be applied.

We have also checked that the stationary solutions of the equations of moments (6.17)–(6.18) fit well the stationary distribution centered at an attractor. Using the Langevin dynamical equations (6.16), we simulated 40 random trajectories in the range $t \in [0, 20]$ and assumed that the system was both self-averaging (ergodic) and quasi-stationary in that time interval. The initial conditions were randomly drawn from a normal distribution of variance 0.1 and centered at the asymmetric fixed point of the noise-free version of Eqs. (6.16), $y_{\text{asym1}} = (0.9911, 0.0059)$, and the network parameters were the same used in Fig. 6.1, $w = 1.5$, $E = 0.7$. The histogram obtained from the simulated trials was then fit with a normal distribution using maximum likelihood estimators. The estimated values were compared with the fixed points of Eqs. (6.17)–(6.18) associated with the quasi-stationary distribution centered at y_{asym1}. Figure 6.2B shows the bivariate normal distribution estimated from the simulated data, as well as the corresponding error ellipse, defined as the curve of constant density equal to $\exp(-1/2)$ of the value of the density at the mean. The figure also shows the error ellipse of the distribution derived from the equations of moments, described

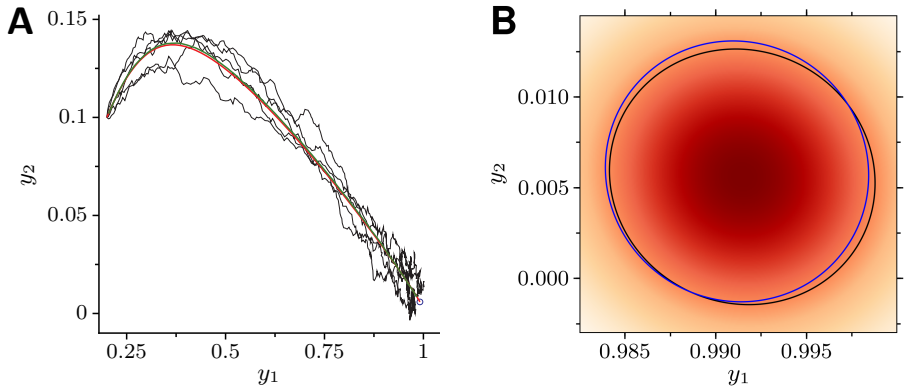


Figure 6.2: Moment equations capture the temporal evolution of the stochastic variable $Y = (y_1, y_2)$. Parameters: $w_+ = 1.5$, $E = 0.7$, $\beta = 0.01$. **A**: Sample trajectories starting at $(y_1, y_2) = (0.2, 0.1)$ for 6 different noise realizations (thin black curves). Red: trajectory for the mean $\big(\mu_1(t), \mu_2(t)\big)$, obtained solving the system (6.17)–(6.18) with initial condition. Green: trajectory for the estimated mean, obtained by a sample average over 1000 realizations of the Langevin dynamical equation. Note that in some parts of the trajectory the red and green curves are indistinguishable. To avoid cluttering the figure with random walks around the y_1-choice fixed point, trajectories have been integrated for $t \in [0, 6]$. **B**: Estimated stationary probability density for a sample of simulated trajectories wandering around the y_1-decision state (see text for details). Black: error ellipse for the density estimated from the simulated trials. Blue: error ellipse for the density computed with the method of moments (also shown in A).

by the condition $(x-\mu)^T\gamma^{-1}(x-\mu) = 1$, where μ is the two-dimensional vector of the means and γ is the 2-dimensional covariance matrix. The ellipse is centered at (μ_1^*, μ_2^*) and has major axes defined by the eigenvectors of the covariance matrix evaluated at the fixed point. The lengths of the major axes are given by the square root of the corresponding eigenvalues.

The predictions from the reduced deterministic system are consistent with the simulations of the stochastic system for a wide range of initial conditions, network parameters, and noise amplitudes. The method of moments provides thus an equivalent description based on a set of deterministic differential equations for the first- and second-order moments of the state variables, which implicitly include information about the fluctuations present in the system. An extension of the method to bistable systems, where the distribution of state variables is bimodal, was developed in [62]. The method allows us to investigate the role of random fluctuations in systems of coupled rate equations like those commonly used in neurodynamical modeling, and without the need for simulated random realizations.

Chapter 7

Nonlinear diffusion models of detection

In this chapter we formally reduce the dynamics of a minimal rate model of detection to the corresponding one-dimensional nonlinear diffusion equation using multiscale analysis, along the lines followed in [165] for decision-making models. The reduction allows us to regard the detection problem as a noise-driven motion along a one-dimensional energy function with shape determined by the parameters of the model. Putative behavioral outcomes of the model such as psychometric curves can then be related analytically with the model parameters by using one-dimensional rate theory (see Sec. 5.2). This framework is generic and can be applied to different models, such as firing rate descriptions or networks of spiking neurons, as long as they share the same bifurcation structure near the detection threshold.

7.1 Perceptual detection

Behavioral and neuronal responses in simple detection tasks vary across repetitions under the same stimulus conditions. The probabilistic nature of both types of response is specially manifest when the intensity of the stimulus is close to the detection threshold. Recent studies have analyzed the neuronal responses in several cortical areas during a vibrotactile detection task [59, 60], in which trained monkeys have to report the presence or absence of a mechanical vibration applied to one fingertip. Responses in medial premotor cortices (MPC) were found to be strongly correlated with the monkey's perceptual report, and only weakly correlated with the stimulus intensity, showing an all-or-none response after the perceptual decision was made, with a latency determined by the stimulus strength.

Some aspects of the activity in MPC during perceptual detection tasks can be modeled using attractor network models exhibiting multistability. In the simplest version of such models, the perceptual event is encoded in the activity of one neural population receiving external inputs proportional to the intensity of the probe stimulus. The system can be either in a silent state, associated with the negative

percept ("no" response), or in an active state, associated with the positive percept ("yes" response). The model system is assumed to be bistable for inputs near the detection threshold, with stimulation raising the probability of noise-induced transitions from the default silent state to the active state [63]. We illustrate in the next section a very simple model with these features.

7.2 Two-dimensional stochastic rate model of detection

We consider a simple rate model describing the activity of one excitatory population and one inhibitory population coupled together

$$\begin{aligned} \dot{r}_{\mathrm{E}} &= -r_{\mathrm{E}} + \Phi(J_{\mathrm{EE}} r_{\mathrm{E}} - J_{\mathrm{EI}} r_{\mathrm{I}} + I) + \sigma_{\mathrm{E}} \eta_{\mathrm{E}}(t)\,, \\ \tau_{\mathrm{I}} \dot{r}_{\mathrm{I}} &= -r_{\mathrm{I}} + \Phi_{\mathrm{I}}(J_{\mathrm{IE}} r_{\mathrm{E}} - J_{\mathrm{II}} r_{\mathrm{I}}) + \sigma_{\mathrm{I}} \sqrt{\tau_{\mathrm{I}}} \eta_{\mathrm{I}}(t)\,, \end{aligned} \tag{7.1}$$

where r_{E} and r_{I} are the firing rates of the excitatory and inhibitory neural populations, respectively, J_{EE}, J_{IE}, J_{EI}, and J_{II} are the coupling strengths, all positive, and I is the external input. The activation functions are sigmoids,

$$\Phi_{\mathrm{E,I}}(x) = \big(1 + \exp\big[-\alpha(x-\theta)\big]\big)^{-1}\,,$$

Fluctuations in the firing rates are modeled with additive Gaussian white noise sources of zero mean and variances σ_{E}^2 and σ_{I}^2 for the excitatory and inhibitory populations, respectively (last term in both Eqs. (7.1), where $\langle \eta_\alpha(t) \rangle = 0$ and $\langle \eta_\alpha(t) \eta_\beta(t') \rangle = \delta_{\alpha\beta}\delta(t-t')$, $\alpha, \beta = \mathrm{E}, \mathrm{I}$). Note that time is expressed in units of the time constant of the excitatory neural population. The architecture of the model is illustrated in Figure 7.1A.

Let us first consider the noiseless case, $\sigma_{\mathrm{E}} = \sigma_{\mathrm{I}} = 0$. The system exhibits bistability when the positive feedback arising from excitatory recurrent coupling, J_{EE}, is sufficiently strong (Fig. 7.1B), so that a high-activity (H) a low-activity (L) stable states coexist for some range of the inputs. As in the models of decision discussed so far, multistability results from the interplay between the nonlinearity in the transfer function and the positive feedback, provided in this case by the self-coupling. A minimum amount of positive feedback is therefore needed to sustain a high-activity stable state. Beyond this critical value, the system is bistable for some range of the external input, with the bistability range broadening as the self-coupling increases (region L, H in Fig. 7.1C). Note that phase diagram has a *cusp* at the start of the bistability range.

The branch of high-activity stable solutions appears jointly with a branch of saddle (unstable) solutions via a saddle-node bifurcation as the external input I increases (Fig. 7.1B). The unstable branch coalesces with the low-activity stable branch input increases and disappears through a second saddle-node bifurcation at some critical value $I = I_c$. If the system is initially in the low-activity stable state, the application of an external current I greater than I_c will force the system jump to the high-activity state, while for $I < I_c$ the system will remain in the low-activity branch.

In the model, the low- and the high-activity states are identified with the negative and positive perceptual reports, respectively. The system is initialized in

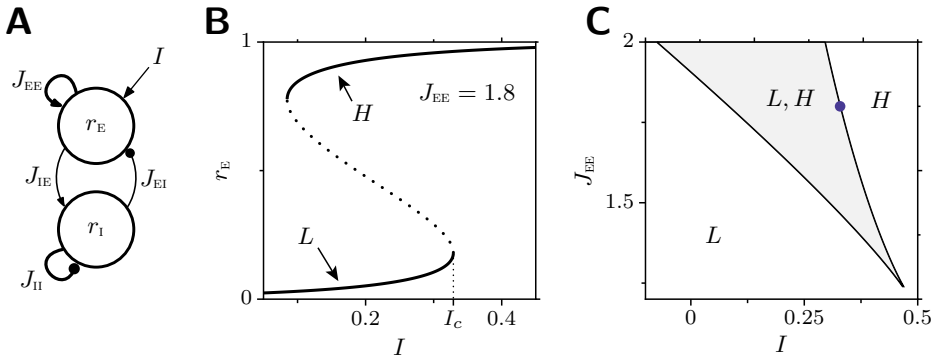


Figure 7.1: Bistability in a simple two-dimensional rate model. **A**: Architecture. The circuit is composed of one excitatory and one inhibitory units, which are self- and cross-coupled. Only the excitatory unit receives external input. **B**: Dependence of the activity r_E of the excitatory population as a function of the external input I, for $J_{EE} = 1.8$. Thick solid curves: stable fixed point solutions (H: high-activity branch; L: low-activity branch). Parameters: $J_{II} = J_{EI} = J_{IE} = 1$, $\tau_I = 1$, $\Phi_E(x) = \Phi_I(x)$, with $\alpha = 4$ and $\theta = 1$. **C**: Phase diagram of the rate model. In regions labeled with L (H) there exists a stable fixed point with low (high) excitatory activity, r_E. The area shaded in gray and labeled L, H represents the region of bistability between high- and low-activity solutions. Blue dot: bifurcation point where the low-activity branch represented in B disappears, i.e., having coordinates $(I, J_{EE}) = (I_c, 1.8)$. All parameters except J_{EE} as in B.

the low-activity state before stimulation to mimic the low, spontaneously active activity of MPC neurons before stimulus onset [59]. In the absence of fluctuations, the excitatory unit will become active, and hence a positive percept will be reported, whenever $I > I_c$. Otherwise, the unit remains silent. The psychometric curve, representing the fraction of positive perceptual reports as a function of the stimulus intensity, would be in this case be a step function with threshold at $I = I_c$.

The situation changes when random fluctuations are included in the model. Noise-driven transitions between the low- and the high-activity stable states may occur when the input is below the critical value I_c. The hypothetical psychometric curve that would result from such stochastic perceptual activation is a smeared out step function, namely, a sigmoid, whose exact shape will depend in general on the parameters of the model —notably, the amplitude of the fluctuations— as well as on the duration of the trial. The model predicts that the fraction of positive percepts must increase with the duration of the trial, as the probability of transition from one to the other state tends to one in the limit of trials of infinite duration. In a general detection problem, though, the temporal window within which the sensory signal is supposed to be integrated is not specified at all, and some mechanism to frame temporally the integration of evidence needs to be invoked [88, 181]. We ignore in the following this problem by assuming that such temporal window is set beforehand, which corresponds to the experimental situation where the presentation of the stimulus is cued.

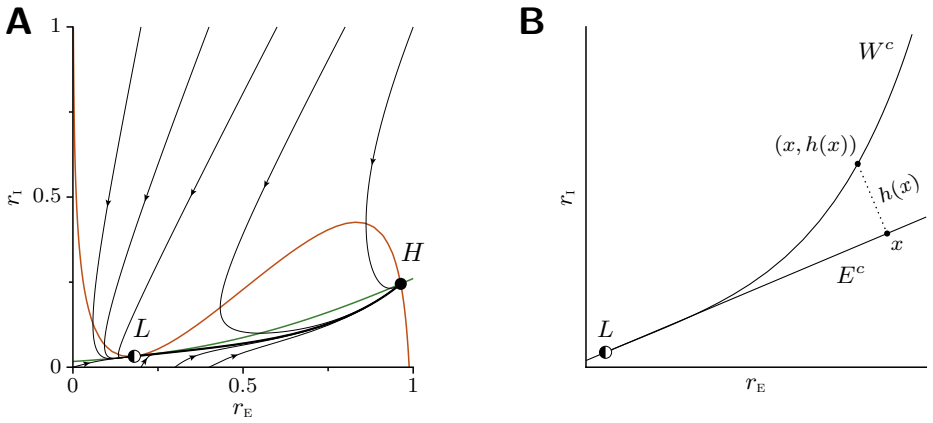


Figure 7.2: Phase portrait of the two-dimensional rate model at the bifurcation value $I = I_c$, for $J_{EE} = 1.8$. **A**: Red and green curves are the r_E- and r_I-nullclines, respectively, and thin black curves are trajectories with different initial conditions. Black thick curve: one-dimensional center manifold. Parameters as in Fig. 7.1. **B**: Inset. Near the critical fixed point L, dynamics are slow and confined to the center manifold W^c, which can be locally represented by the graph of a function $h(x)$, $x \in E^c$, from the center manifold E^c to its orthogonal space $E^s \oplus E^u$. In this system the orthogonal space is simply E^s.

The identification of the detection threshold with the saddle-node bifurcation implies a critical slowing down of the dynamics near threshold, with a subsequent reduction in the dimension of the system because of the separation of time scales. Near the bifurcation point at $I = I_c$, dynamics are fast except along the one-dimensional unstable manifold of the saddle, which is, strictly speaking, the center manifold of the critical fixed point. Trajectories relax rapidly onto the manifold, and then evolve slowly on it. Figure 7.2A depicts the phase portrait for the bifurcation value $I = I_c$ at which the low-activity state disappears. Trajectories starting at different initial conditions over the phase space get trapped on the slow manifold of the critical fixed point L. Our objective is to derive the equation of motion along this one-dimensional slow manifold.

7.3 Dynamics on the center manifold

Let $r^* = (r_E^*, r_I^*)$ be a fixed point of the system (7.1), which we represent succinctly as $\dot{r} = f(r) = -r + \Phi(\boldsymbol{r}; I, \tau_I, \{J\})$. The stability of the solution $\boldsymbol{r}^*$ is determined by the eigenvalues of the linearized system at the fixed point,

$$Df(r^*) = \begin{pmatrix} -1 + J_{EE}\Phi'_E & -J_{EI}\Phi'_E \\ J_{IE}\Phi'_I & -1 - J_{II}\Phi'_I \end{pmatrix} \tag{7.2}$$

where the derivatives Φ'_E and Φ'_I are evaluated at the fixed point,

$$\Phi'_E \equiv \Phi'_E(J_{EE}r_E^* - J_{EI}r_I^* + I)\,,$$
$$\Phi'_I \equiv \Phi'_I(J_{IE}r_E^* - J_{II}r_I^*)\,.$$

A bifurcation occurs when at least one of the eigenvalues of the linearized system, a *critical* eigenvalue, has zero real part. In a saddle-node bifurcation the critical eigenvalue is 0, and we must have therefore $\det Df(r^*) = 0$. This condition gives rise to a functional relationship G between the parameters, $G(I, \{J\}, \tau_{\mathrm{I}}) = 0$, which is precisely the relationship that the two curves shown in the phase diagram Fig. 7.1C satisfy. In the following we will take coupling strengths, time constants, and noise amplitudes as fixed parameters, and consider the input I as the only control parameter.

The center subspace The eigenspace associated with the critical eigenvalue 0 is found from $Df(r^*)x = 0$ and is called the center subspace. Note that it is the kernel of the Jacobian $Df(r^*)$. In general, any vector r of the original space can be decomposed as a direct sum of the center space E^c and the stable and unstable subspaces, E^s and E^u, of the fixed point, spanned by the eigenvectors associated with the eigenvalues with negative and positive real part, respectively [101, 193]. In short, any vector r of the phase space, $r \in \mathbb{R}^n$, can be represented as

$$r = (x, y), \text{ where } x \in E^c \text{ and } y \in E^s \oplus E^u\,, \tag{7.3}$$

and where, if E_λ denotes the eigenspace with eigenvalue λ,

$$\begin{aligned} E^c &= \mathrm{span}\{v \mid v \in E_\lambda \text{ and } \operatorname{Re}\lambda = 0\}\,, \\ E^s &= \mathrm{span}\{v \mid v \in E_\lambda \text{ and } \operatorname{Re}\lambda < 0\}\,, \\ E^u &= \mathrm{span}\{v \mid v \in E_\lambda \text{ and } \operatorname{Re}\lambda > 0\}\,. \end{aligned}$$

The direct sum of the stable, unstable, and center subspaces, of dimensions n^s, n^u, and n^c, respectively, is the whole phase space: $\mathbb{R}^n = E^s \oplus E^u \oplus E^c$. Using the decomposition (7.3), any dynamical system can be cast into

$$\dot{x} = Ax + N_1(x, y)\,, \tag{7.4a}$$
$$\dot{y} = By + N_2(x, y)\,, \tag{7.4b}$$

where A is an $n^c \times n^c$ matrix whose eigenvalues lie all on the imaginary axes, and B is an $(n^s + n^u) \times (n^s + n^u)$ matrix with eigenvalues off the imaginary axis. The terms N_1 and N_2 are nonlinear in (x, y) variables.

The center manifold The critical fixed point r^* possesses a center manifold W^c that passes through r^* and which is tangent to the center subspace E^c at r^*. The tangency at r^* allows us to express the manifold W^c as the graph of a function $y = h(x)$ (see Fig. 7.2B)

$$\begin{aligned} h : E^c &\longrightarrow E^s \oplus E^u \\ x &\longmapsto y\,, \end{aligned}$$

where, for x sufficiently small, the point $r = (x, h(x))$ belongs to W^c [56]. Moreover, if $\bar{r}(t) = (\bar{x}(t), \bar{y}(t))$ is a trajectory on the center manifold with sufficiently small

amplitude (not too far from the fixed point), one has that $\bar{y}(t) = h(\bar{x}(t))$; this curve defines a vector field $\dot{\bar{y}}(t) = Dh(\bar{x}(t))\dot{\bar{x}}(t)$, which has to be equal to the vector field in Eq. (7.4b). The function $h(x)$ that satisfies this "invariance condition",

$$Dh(x)\dot{x} = Dh(x)[Ax + N_1\big(x, h(x)\big)] \stackrel{!}{=} Bh(x) + N_2\big(x, h(x)\big), \tag{7.5}$$

defines the center manifold. Equation (7.5) is a nonlinear partial differential equation and cannot be solved in general. It can, however, be solved perturbatively using power series expansions.

As mentioned in Chapter 5, the usefulness of center manifold theory is that it allows us to reduce the dimensionality of the problem down to the dimension of the center subspace. The function function $h(x)$ makes this reduction possible by decoupling the flow described by Eq. (7.4a) from that given in Eq. (7.4b)

$$\dot{x} = Ax + N_1\big(x, h(x)\big). \tag{7.6}$$

This equation describes the evolution of the system in the vicinity of the bifurcation. When the unstable subspace is empty ($n^u = 0$) trajectories rapidly converge to n^c-dimensional manifold $(x, h(x))$, and then evolve according to the dynamical equation (7.6). The dynamics of the system on the center manifold are slow due to the infinite long time constants associated with the eigenvalues with zero real part.

7.3.1 Multiscale analysis

We will now derive the dynamical equation of the system (7.1) on the center manifold. In principle we should first determine $h(x)$ from the invariance condition (7.5) to the desired order and then plug the approximate solution into the dynamical equation (7.6). Instead, we will take advantage of the separation of time scales in the center manifold and will apply multiscale analysis [24]. The idea is to compute the perturbative terms beyond linear order in an expansion around the critical fixed point. We will ignore noise for the moment.

Let $r^* = (r^*_{\text{E}}, r^*_{\text{I}})$ be the critical fixed point in the lower branch, at $I = I_c \equiv I_0$. We introduce a small perturbative parameter ϵ which measures the distance to the bifurcation, and expand the solution r and the control parameter I in power series of ϵ:

$$r = r^* + \epsilon r_1 + \epsilon^2 r_2 + \cdots, \tag{7.7a}$$

$$I = I_0 + \epsilon I_1 + \epsilon^2 I_2 + \cdots. \tag{7.7b}$$

We also assume there are two timescales, one fast, t, and one slow, $T = \epsilon t$, and that the perturbations r_i, $i = 1, \ldots$, evolve according to the slow time scale T. The two time coordinates are treated as independent variables, so

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{dT}{dt}\frac{\partial}{\partial T} = \partial_t + \epsilon\partial_T. \tag{7.8}$$

Plugging Eqs. (7.7) and (7.8) into Eqs. (7.1), and collecting terms order by order in ϵ we get a series of linear equations for the higher-order terms r_k. At

every order in ϵ the equations to solve have the form $Lr_k = N(r_{l<k})$, where L is a linear operator and N include all the inhomogeneous terms, which may depend on the solutions found for lower orders. The 0-th order of the expansion is simply the fixed point condition $\dot{r} = V(r^*, I_c) = 0$. Let us compute the first and second order corrections.

First order in ϵ Collecting all terms linear in ϵ, we have

$$0 = \begin{pmatrix} -1 + J_{\text{EE}}\Phi'_{\text{E}} & -J_{\text{EI}}\Phi'_{\text{E}} \\ J_{\text{IE}}\Phi'_{\text{I}} & -1 - J_{\text{II}}\Phi'_{\text{I}}, \end{pmatrix} \begin{pmatrix} r_{1\text{E}} \\ r_{1\text{I}} \end{pmatrix} + \begin{pmatrix} I_1\Phi'_{\text{E}} \\ 0 \end{pmatrix} \equiv -Lr_1 + N_1$$

where we have defined L as a shorthand notation for the linear operator consisting in the Jacobian matrix with sign reversed, $L = -DV(r^*, I_c)$. Since at the critical point $\det L = 0$, we cannot invert L to solve for r_1. A solution for r_1 only exists if N_1 is in the range of L. According to Fredholm alternative theorem, the equation $Lx = b$ has a solution if and only if $b^T n = 0$ for every vector n satisfying $L^T n = 0$. That is to say, the kernel of L^T is the orthogonal complement of the range of L. We can express this solvability condition finding a left null eigenvector of L, i.e., a vector n such that $n^T L = 0$, and imposing $n^T N_1 = 0$. It leads to

$$n = \begin{pmatrix} J_{\text{IE}}\Phi'_{\text{I}} \\ 1 - J_{\text{EE}}\Phi'_{\text{E}} \end{pmatrix}, \text{ so that } n^T N_1 = 0 \;\Rightarrow\; J_{\text{EI}}\Phi'_{\text{E}}\Phi'_{\text{I}} I_1 = 0 \;\Rightarrow\; I_1 = 0,$$

where in the last equality we have used the fact that Φ'_{E} and Φ'_{I} are both nonzero at the bifurcation point. We are then left with the equation $Lr_1 = 0$, which states that r_1 must be in the kernel of the linear operator L. Since $L = -DV(r^*, I_c)$, the kernel of L is by definition the center subspace of the fixed point (r^*, I_c) (Sec. 7.3). Solving for r_1 we obtain

$$r_1 = \begin{pmatrix} r_{1\text{E}} \\ r_{1\text{I}} \end{pmatrix} = A \begin{pmatrix} J_{\text{EI}}\Phi'_{\text{E}} \\ J_{\text{EE}}\Phi'_{\text{E}} - 1 \end{pmatrix}, \tag{7.9}$$

where the arbitrary coefficient A depends in general on the slow time scale, $A = A(T)$. The first order approximation states therefore that perturbations grow along the critical (center) eigenvector, as expected.

Second order in ϵ We have

$$\begin{pmatrix} \partial_T \\ \tau_{\text{I}}\partial_T \end{pmatrix} r_1 = Lr_2 + \frac{1}{2}\Phi'' r_1^2 + \begin{pmatrix} \Phi'_{\text{E}} I_2 \\ 0 \end{pmatrix},$$

which, after rearranging terms, reads

$$Lr_2 = -\partial_T \begin{pmatrix} r_{1\text{E}} \\ \tau_{\text{I}} r_{1\text{I}} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \Phi''_{\text{E}}[J_{\text{EE}} r_{1\text{E}} - J_{\text{EI}} r_{1\text{I}}]^2 \\ \Phi''_{\text{I}}[J_{\text{IE}} r_{1\text{E}} - J_{\text{II}} r_{1\text{I}}]^2 \end{pmatrix} + \begin{pmatrix} \Phi'_{\text{E}} I_2 \\ 0 \end{pmatrix} \equiv N_2(r_1).$$

Again, we must have $n^T N_2 = 0$, where n is the same left null eigenvector of L used above. It gives

$$\begin{pmatrix} J_{\text{IE}}\Phi'_{\text{I}} \\ 1 - J_{\text{EE}}\Phi'_{\text{E}} \end{pmatrix}^T \begin{pmatrix} -J_{\text{EI}}\Phi'_{\text{E}}\partial_T A + \Phi''_{\text{E}} J_{\text{EI}}^2 A^2/2 + \Phi'_{\text{E}} I_2 \\ \tau_{\text{I}}[1 - J_{\text{EE}}\Phi'_{\text{E}}]\partial_T A + \Phi''_{\text{I}}[J_{\text{IE}} J_{\text{EI}}\Phi'_{\text{E}} - J_{\text{II}}(J_{\text{EE}}\Phi'_{\text{E}} - 1)]^2 A^2/2 \end{pmatrix} = 0,$$

which can be recast into

$$\partial_T A \left(J_{\text{IE}} J_{\text{EI}} \Phi'_{\text{E}} \Phi'_{\text{I}} - \tau_{\text{I}} [1 - J_{\text{EE}} \Phi'_{\text{E}}]^2 \right) = J_{\text{IE}} \Phi'_{\text{I}} \Phi'_{\text{E}} I_2 + \frac{A^2}{2} \left(J_{\text{IE}} J_{\text{EI}}^2 \Phi'_{\text{I}} \Phi''_{\text{E}} + \Phi''_{\text{I}} (1 - J_{\text{EE}} \Phi'_{\text{E}}) [\Phi'_{\text{E}} (J_{\text{IE}} J_{\text{EI}} - J_{\text{II}} J_{\text{EE}}) + J_{\text{II}})]^2 \right).$$

This differential equation has the form

$$\partial_T A(T) = \mu + b A(T)^2, \tag{7.10}$$

where the coefficients μ and b depend on the parameters of the original system, Eq. (7.1), and are given by

$$\alpha = (J_{\text{EE}} \Phi'_{\text{E}} - 1) \left[1 + J_{\text{II}} \Phi_{\text{I}} + \tau_{\text{I}} (1 - J_{\text{EE}} \Phi'_{\text{E}}) \right], \tag{7.11a}$$

$$\alpha \mu = J_{\text{IE}} \Phi'_{\text{I}} \Phi'_{\text{E}} I_2 \equiv \hat{\mu} I_2 = \hat{\mu} (I - I_c), \tag{7.11b}$$

$$\alpha b = \left(J_{\text{IE}} J_{\text{EI}}^2 \Phi'_{\text{I}} \Phi''_{\text{E}} + \Phi''_{\text{I}} (1 - J_{\text{EE}} \Phi'_{\text{E}}) [\Phi'_{\text{E}} (J_{\text{EI}} J_{\text{IE}} - J_{\text{II}} J_{\text{EE}}) + J_{\text{II}})]^2 \right) / 2, \tag{7.11c}$$

We have used $\det L = (1 - J_{\text{EE}} \Phi'_{\text{E}})(1 + J_{\text{II}} \Phi'_{\text{I}}) + J_{\text{IE}} J_{\text{EI}} \Phi'_{\text{E}} \Phi'_{\text{I}} = 0$ to simplify the expression for α. Note that the monotonicity of the activation functions implies $\Phi'_{\text{E,I}} > 0$ and hence $\hat{\mu} > 0$, so that the sign of the constant μ depends on whether the control parameter is above or below the critical value. The value μ is proportional to the gains of both neural units.

Noise sources The contribution to the amplitude equation (7.10) of the noise terms in the original system is calculated to first order in ϵ, projecting on the space of the left null eigenvector n,

$$\text{noise terms} = n^T \begin{pmatrix} \sigma_{\text{E}} \eta_{\text{E}}(t) \\ \sigma_{\text{I}} \eta_{\text{I}}(t) \end{pmatrix} = J_{\text{IE}} \Phi'_{\text{I}} \sigma_{\text{E}} \eta_{\text{E}}(t) + (J_{\text{EE}} \Phi'_{\text{E}} - 1) \sigma_{\text{I}} \eta_{\text{I}}(t).$$

These two terms are then divided by the common factor α arising from the projection —as we did for μ and b, in Eqs. (7.11b), (7.11c). The variance of the noise in the amplitude equation is therefore

$$\sigma^2 = \frac{1}{\alpha^2} \left[(J_{\text{IE}} \Phi'_{\text{I}})^2 \sigma_{\text{E}}^2 + (J_{\text{EE}} \Phi'_{\text{E}} - 1)^2 \sigma_{\text{I}}^2 \right]. \tag{7.11d}$$

7.4 Nonlinear diffusion equation

With the reduction derived above we arrive at a one-dimensional stochastic differential equation

$$\partial_T A(T) = \mu + b A(T)^2 + \sigma \eta(t), \tag{7.12}$$

with coefficients μ, b, and σ given by Eqs. (7.11). Equation (7.12) can be thought of as describing the motion of an overdamped Brownian particle under the influence of a potential $E(A)$ (see Sec. 5.2):

$$\partial_T A(T) = -\frac{\mathrm{d}E(A)}{\mathrm{d}A} + \sigma \eta(t), \tag{7.13}$$

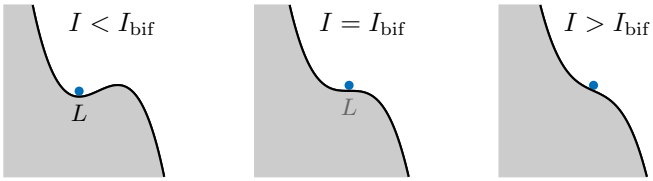


Figure 7.3: One-dimensional energy function E in the vicinity of the low-activity state L for inputs below, equal, and above the critical value.

where

$$E(A) = -\mu A - \frac{b}{3}A^3(T). \tag{7.14}$$

The Langevin equation (7.12) is referred to in this context as amplitude equation, as it describes the time course of the amplitude along the center subspace (see Eq. (7.9)). We will also refer to it as nonlinear diffusion equation.

When the input is below the critical value, the constant term μ is negative and the energy has a local minimum, which corresponds to the low-activity stable fixed point L of the original system (Fig. 7.3). The state becomes marginally stable when the input reaches the critical value I_c ($\mu = 0$), to finally disappear for inputs higher than I_c ($\mu > 0$). In the latter case, the system evolves going 'downhill' the energy function, thereby increasing A, until it decays to the high-activity state H (not represented in the figure).

We stress that, although we have used a particular model to illustrate the reduction to a one-dimensional nonlinear diffusion equation, the same argumentation can be followed for other systems having the same local bifurcation structure. To be precise, we expect the same type of noise-driven motion in a cubic potential, Eq. (7.13), whenever a system undergoes a saddle-node bifurcation through which a 'low-activity state' disappears when inputs are increased. Of course, the coefficients μ, b, and σ will depend on the specific details of the model, but the dependence on the amplitude A will remain the same.

7.4.1 Comparison with the original system

Equation (7.10) captures the dynamics of the system near the bifurcation point $(r, I) = (r^*, I_c)$. It is basically the normal form of a saddle-node bifurcation [180], with a pair of stable and unstable fixed point solutions appearing at $A^* = \pm\sqrt{-\mu/b}$. According to the first-order correction to the critical fixed point, Eq. (7.9), near the bifurcation point at $I = I_c$, the fixed point solutions will be

$$\begin{pmatrix} r_{\text{E}}^* \\ r_{\text{I}}^* \end{pmatrix} = \pm\sqrt{-\frac{\mu(I)}{b}} \begin{pmatrix} J_{\text{EI}}\Phi'_{\text{E}} \\ J_{\text{EE}}\Phi'_{\text{E}} - 1 \end{pmatrix} + O(\epsilon^2),$$

The dependence on the input I of the fixed point solutions for the original system and for the amplitude equation is shown in Figure 7.4A. The solution of the amplitude equation provides the second order approximation in I of the

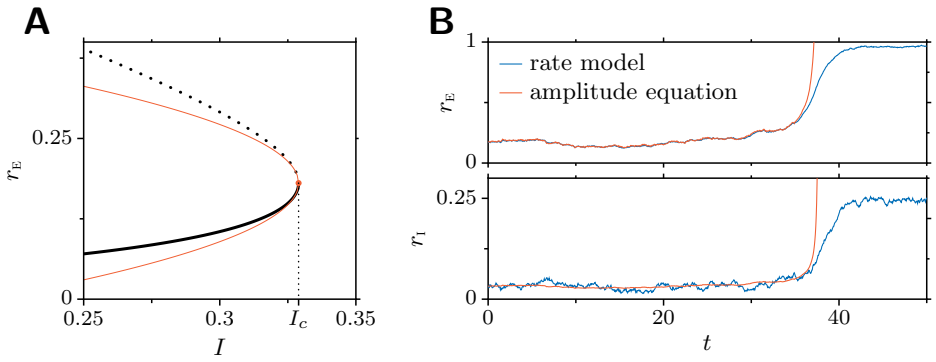


Figure 7.4: The dynamics described by the amplitude equation and by the original rate equations match at the bifurcation point (r^*, I_c). **A**: Bifurcation diagram for the original system (black) and the amplitude equation (red) with coefficients given by Eqs. (7.11). **B**: Sample trajectories of the original system (blue) and the one-dimensional amplitude equation (red), for $I \approx I_c - 9.56 \cdot 10^{-4}$. $\sigma_E = \sigma_I = 0.01$. Parameters: $J_{EE} = 1.8$, $J_{EI} = J_{IE} = J_{II} = 1$, $\tau_I = 1$.

stable and unstable branches around the bifurcation point, and becomes a better approximation the closer is I to the bifurcation value. The match between both solutions can be further improved including the ϵ^2 correction r_2 to the critical fixed point (see Eq. (7.7a)), although this correction would not add any new qualitative feature to the description.

The evolution of the original system is correctly captured by the nonlinear diffusion equation. Figure 7.4B shows the time course in a single trial of the rate variables r_E and r_I, as given by the original two-dimensional model, Eqs. (7.1) and the one-dimensional system, Eq. (7.12). In the latter, the rates r_E and r_I are related to the amplitude variable A through Eq. (7.9). The input was chosen slightly below the critical value to have the low-activity state L metastable. The nonlinear diffusion equation matches better the evolution of the variable r_E than it does for r_I, because the center subspace onto which the dynamics are projected lies mostly on the subspace spanned by r_E (see Fig. 7.2A). Thus, since the center manifold is practically orthogonal to r_I, fluctuations in this transversal dimension hardly affect the dynamics on the manifold and hence on the formation of a percept, according to the model. This is indirectly reflected in the suppressed variability of the r_I variable derived from the amplitude equation when compared to the original counterpart (lower panel of Fig. 7.4B).

The trajectories described by the one-dimensional nonlinear diffusion equation and by the original rate equations match closely until both systems are driven out by noise from the domain of attraction of the low-activity metastable state. When this transition occurs, the trajectory of the original two-dimensional system decays to the high-activity fixed point H, while the trajectory of the one-dimensional system blows up to infinity due to the positive quadratic term in Eq. (7.12). We have estimated the distributions of transition times, defined as the time necessary

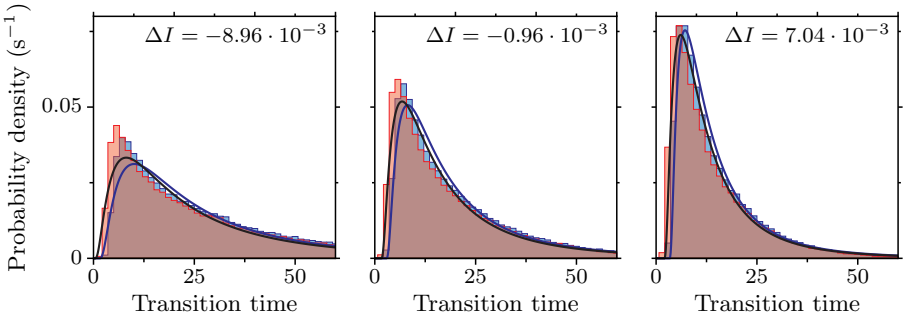


Figure 7.5: Distribution of transition times for the original and the one-dimensional system, for three values of the input. Transition times are defined as the time elapsed before r_{E} reaches the value 0.8, and $\Delta I \equiv I - I_c$. Blue histogram: transition times for the original system. Red histogram: transition times for the one-dimensional system. Blue and black curves: lognormal fits of the transition times for the original and the reduced systems, respectively. The parameters of the lognormal distribution were found applying an adapted Levenberg-Marquardt algorithm on the sum squared distances between the expected cumulative density and the observed cumulative probability [110]. Each histogram has been obtained from a sample of 10^5 trials. Other parameters as in Fig. 7.4.

for r_{E} to cross the threshold $r_\theta = 0.8$ when the system is initialized near the low-activity state, are shown in Fig. 7.5 for sub-, near-, and super-critical values of the input I. Even when I differs considerably from the critical value, the distribution of escape times obtained from the nonlinear diffusion equation (7.12) closely matches that predicted by the full-dimensional system. The distribution corresponding to the one-dimensional system is slightly more skewed to low transition times than that associated with the two-dimensional system. The difference become significant for subcritical currents (Fig.7.5 left). This may reflect a dimensional bottleneck effect caused by the fewer available dimensions through which the system can escape from its metastable state.

7.5 Summary and perspectives

Experimental studies on detection show that behavioral responses during simple perceptual detection tasks are highly variable for stimulations near the detection threshold (see, e.g., [22, 33, 54, 59]). Based on this experimental fact, we have proposed a simple stochastic bistable model for detection and have reduced its dynamics onto the slow critical manifold that arises at the bifurcation where the inactive state disappears, and which we identify as the detection threshold. The reduction provides a useful tool to derive the dependence of behavioral measures on the model parameters like the stimulus intensity or the connection strengths. Such dependence is expected to be generic for different classes of models based on noise-driven, bistable units sharing the same bifurcation structure. Our next step

will consist in assessing whether the universal properties exhibited by this broad class of models can explain the electrophysiological and behavioral data obtained during detection tasks.

Chapter 8

Results and Conclusions

The work presented in this thesis addresses the putative role that noise-induced transitions in multistable neural systems may play in both the variability and duration of behavioral responses during simple perceptual decision tasks. Our results rely the assumption that neurons of the brain regions involved in the decision process engage in a competition for higher activity upon stimulation, as suggested by recent electrophysiological experiments [173, 164, 70]. It is also assumed that cortical networks exhibit multistability, an assumption supported by the existence of persistent activity states during working memory tasks [83, 116]. We have focused on the regimes that potentially arise for values of the control parameters (mainly, the inputs) below the onset of winner-take-all behavior, where the interplay between noise and multistability is expected to have the strongest effects. We summarize in the following the main results of this thesis.

Project 1 (Chapter 5)

1. Noise-induced transitions among asynchronous, stable network states constitute a neurobiologically plausible mechanism for spanning times beyond synaptic and cellular time scales, without the need for fine-tuning.

2. Noise-induced activation mechanisms lead in a natural way to exponential distributions of transition times, which can potentially account for the long tails commonly observed in the distribution of response times of different mental processes.

3. We observe a monotonic dependence of response times and performance on the average input feeding the competing units, in agreement with the analytical results of [165], valid for systems operating near the onset of winner-take-all behavior. This monotonic dependence constitutes a plausible mechanism to explain physiological substrate of the speed-accuracy tradeoff observed behaviorally.

4. Our results predict an that the skewness of the response time distributions must increase when subjects are instructed to respond more accurately.

Project 2 (Chapter 6)

1. The time course of a stochastic neural rate model, described by a set of nonlinear Langevin equations, can be properly captured by a reduced set of ordinary differential equations, called moment equations, for the first central moments of the distribution of the original stochastic variables.

2. The method of moments can be applied also to characterize the quasistationary distributions associated with each of the stable fixed points of the deterministic limit of the rate model.

3. The agreement of the moment equations with the simulated stochastic system is very good when the system is far from criticality.

Project 3 (Chapter 7)

1. The dynamics of many stochastic models exhibiting bistability between an active and an nonactive state, as those proposed to account for the neural activity in premotor areas during vibrotactile detection tasks, can be captured by a one-dimensional Langevin equation in a cubic potential.

2. The reduction, valid in the vicinity of the bifurcation point where the nonactive state disappears, provides the functional dependence of the putative behavioral outcomes on the parameters of the model.

Appendix **A**

Brunel-Wang model

Here we give a short description of the network introduced by [43], in its particular implementation as a binary decision making model [191]. Please refer to Chapter 2 for a motivation and a more detailed description of the models.

Network

The network consists of N neurons, of which $N_E = 0.8N$ are pyramidal cells (excitatory) and $N_I = 0.2N$ are interneurons (inhibitory). Every neuron receives from the network N_E excitatory synaptic contacts and N_I inhibitory synaptic contacts; the network is thus fully connected. The whole set of neurons is partitioned into different populations, with all neurons in a population sharing the same single-cell parameters and the same statistical properties of the afferent inputs. The set of all excitatory neurons is in turn structured in three different populations: two populations formed by neurons that encode one or the other choice, and a third population formed by the remaining excitatory neurons. The former two constitute the two disjoint *selective* populations, of fN_E ($f = 0.15$) neurons each. The other $(1 - 2f)N_E$ excitatory neurons do not encode any information about the choices, and constitute the *nonselective* population.

External inputs

To simulate the background input from other brain regions, every neuron in the network receives 800 excitatory connections from external neurons, each of which fires according to an independent Poisson process with rate 3 Hz. Every cell receives therefore an independent Poisson train of spikes of rate 2.4 kHz. On top of this background signal, neurons in selective populations receive external inputs encoding stimulus specific information. Stimulation is modeled with an increase λ in the rate of external incoming spikes received by selective cells (see also 2.5.4, on p. 23).

Neurons

Neurons are modeled as leaky integrate-and-fire units, with resting potential $V_L = -70\,\mathrm{mV}$, firing threshold $V_{\mathrm{thr}} = -50\,\mathrm{mV}$, reset potential $V_{\mathrm{reset}} = -55\,\mathrm{mV}$, and refractory period $\tau_{\mathrm{rp}} = 2\,\mathrm{ms}$ for excitatory cells and $\tau_{\mathrm{rp}} = 1\,\mathrm{ms}$ for inhibitory cells. The subthreshold dynamics of the membrane potential $V(t)$ of every neuron is described by

$$C_m \frac{dV(t)}{dt} = -g_L(V(t) - V_L) - I_{\mathrm{syn}}(t),$$

where C_m is the membrane capacitance, of value 0.5 nF for excitatory neurons and 0.2 pF for inhibitory neurons; g_{m} is the membrane conductance, set to 25 nS for excitatory cells and to 20 nS for inhibitory cells. The total afferent postsynaptic current (PSC) $I_{\mathrm{syn}}(t)$, is a sum of recurrent (coming from the local module) and external (background activity and stimuli) contributions, described in detail below.

Currents

The afferent synaptic current includes glutamatergic excitatory components, mediated by AMPA and NMDA receptors, and inhibitory components, mediated by $\mathrm{GABA_A}$ receptors. External cells contribute to the current only trough AMPA receptors. The total current is given by

$$I_{\mathrm{syn}}(t) = I_{\mathrm{AMPA\,ext}}(t) + I_{\mathrm{AMPA\,rec}}(t) + I_{\mathrm{NMDA}}(t) + I_{\mathrm{GABA}}(t). \tag{A.1}$$

where the different currents are described as the product of the maximal conductance, the driving force, and the fraction of open channels:

$$I_{\mathrm{AMPA\,ext}}(t) = g_{\mathrm{AMPA\,ext}}(V(t) - V_E) \sum_{j=1}^{N_{\mathrm{ext}}} s_j^{\mathrm{AMPA\,ext}}(t)$$

$$I_{\mathrm{AMPA\,rec}}(t) = g_{\mathrm{AMPA\,rec}}(V(t) - V_E) \sum_{j=1}^{N_E} w_j s_j^{\mathrm{AMPA\,rec}}(t)$$

$$I_{\mathrm{NMDA}}(t) = \frac{g_{\mathrm{NMDA}}(V(t) - V_E)}{1 + [\mathrm{Mg}^{2+}] \exp\big(-\beta V(t)\big)/\gamma} \sum_{j=1}^{N_E} w_j s_j^{\mathrm{NMDA}}(t)$$

$$I_{\mathrm{GABA}}(t) = g_{\mathrm{GABA}}(V(t) - V_I) \sum_{j=1}^{N_I} s_j^{\mathrm{GABA}}(t)$$

where s_j^x is the fraction of open channels for receptor $x \in \{\mathrm{AMPA}, \mathrm{NMDA}, \mathrm{GABA}\}$, and g^x is the corresponding maximal synaptic conductance. The values for the synaptic conductances are given in table A.1. To keep the mean recurrent input constant as we vary the size N of the network, all recurrent conductances are rescaled by $1/N$. The reversal potentials are $V_E = 0\,\mathrm{mV}$ and $V_I = -70\,\mathrm{mV}$. The dimensionless parameters w_j found in excitatory recurrent currents are the inter-population synaptic weights, described in section "Connectivity Structure". The maximal synaptic conductance of NMDA currents is voltage dependent and

controlled by the intracellular magnesium concentration $[Mg^{2+}] \approx 1\,\text{mM}$, with coefficients $\gamma = 3.57\,\text{mM}$ and $\beta = 0.062\,(\text{mV})^{-1}$.

The fraction of open AMPA (external and recurrent) channels, s_j^{AMPA} in neuron j obeys the following differential equation:

$$\dot{s}_j^{\text{AMPA}}(t) = -s_j^{\text{AMPA}}(t)/\tau_{\text{AMPA}} + \sum_k \delta(t - t_j^{(k)} - d)\,,$$

where $\tau_{\text{AMPA}} = 2.0\,\text{ms}$, and the sum over k represents a sum over spikes emitted by presynaptic neuron j at time $t_j^{(k)}$, and received after a transmission delay d. In the case of external AMPA currents, spikes are fired following independent Poisson processes of rate $\nu_{\text{ext}} = 2.4\,\text{kHz}$, except for stimulated cells, which receive a Poisson train of spikes at rate $\nu_{\text{ext}} = 2.4\,\text{kHz} + \lambda_i$, $i = A, B$. The dynamics for the NMDA synaptic currents are described by

$$\begin{aligned}\dot{s}_j^{\text{NMDA}}(t) &= -s_j^{\text{NMDA}}(t)/\tau_{\text{NMDA},\downarrow} + \alpha \dot{x}_j(t)\big(1 - s_j^{\text{NMDA}}(t)\big)\,,\\ \dot{x}_j(t) &= -\dot{x}_j(t)/\tau_{\text{NMDA},\uparrow} + \sum_k \delta(t - t_j^{(k)} - d)\,,\end{aligned}$$

where characteristic rise and decay times are $\tau_{\text{NMDA},\uparrow} = 2.0\text{ms}$ and $\tau_{\text{NMDA},\downarrow} = 100\text{ms}$, and $\alpha = 0.5\,(\text{ms})^{-1}$. The GABA synaptic component of the current obeys

$$\dot{s}_j^{\text{GABA}}(t) = -s_j^{\text{GABA}}(t)/\tau_{\text{GABA}} + \sum_k \delta(t - t_j^{(k)} - d)\,,$$

with $\tau_{\text{GABA}} = 5\,\text{ms}$. The rise time for GABA_A currents is, like for AMPA currents, neglected. The transmission delay d is 0.5 ms in all channels.

Neuronal, synaptic, and network parameters are summarized in table A.1. The values for synaptic conductances are chosen so that the firing rate of excitatory and inhibitory cells during spontaneous activity is 3 Hz and 9 Hz, respectively. Additional constraints are needed to determine the values of the 8 conductances. They are the following. First, the average external excitatory input equals the average recurrent excitatory input. Second, the mean recurrent inhibition is three times as high as the mean recurrent excitation, as measured by the charge entry per presynaptic spike. Third, the gating variable of NMDA receptors is 0.95 during the spontaneous activity state. These three conditions are imposed on both excitatory and inhibitory cells, giving rise to six constraints that, together with two conditions imposed on the rates, determine uniquely the 8 conductances.

Connectivity structure

The connection weights between cells determine the structure of the network. Weights are given by the dimensionless parameters w_j (see equations following Eq. (A.1)), which denote the relative strength of the modified synapses with respect to the baseline, to which there corresponds the value $\langle w \rangle = 1$. Note that only recurrent currents contain weights, whose precise value is set according to the average inter-population synaptic efficacies that would result from a Hebbian

plasticity mechanism. According to this prescription, the connection weights would be stronger than average when the activity of pre and postsynaptic neurons is correlated, lower than average when it is anticorrelated, and unaltered when it is uncorrelated. In a selective population, where neurons tend to be coactivated, connections are strengthened above the baseline. The connection weight between cells from the same selective population is denoted by $w_+ > 1$. Analogously, since the activity of the two selective populations is anticorrelated, the two populations are weakly connected, with a value denoted by $w_- < 1$. All other weights are set to the baseline value 1. To ensure that the average excitatory synaptic efficacy is not changed in the learning process, w_- must depend on w_+ as $1 - f(w_+ - 1)/(1 - f)$.

Simulations

We have used a second order Runge-Kutta routine to integrate the system of coupled differential equations that describe the dynamics of all cells and synapses. The time step used was 0.02 ms. To calculate the firing rate of a population we divided the number of spikes emitted in a 50 ms window by the number of neurons in the population and by the window size. The time window was slided with a time step of 5 ms. Every trial was simulated until a decision was made. For a given parameter set, we estimated decision times from a block of 4000 trials.

Initial values

The network is initialized as follows. The initial depolarizations of the N neurons are drawn from a uniform distribution of interval $[H, \theta]$, where H is the reset potential and θ the threshold. The initial values for the $5N$ gating variables are drawn from a uniform distribution of interval $[0, 1]$. For the baseline levels of connection strengths and external inputs, the state of the network decays to the spontaneous state in about one hundred of milliseconds, which is of the order of the longest neuronal or synaptic time scale of the system. Choosing different distributions leads to the same results, up to a transient period of the same time scale.

Table A.1: Parameters used in the integrate-and-fire simulations

Parameter	**Value**	
Network parameters:		
N: number of cells in the network	500–4500 cells	
N_E: number of excitatory cells	$0.8N$	
N_I: number of inhibitory cells	$0.2N$	
N_{ext}: number of cells in the external module	800	
c: sparseness (connectivity level)	1.0	
p: number of selective populations	2	
f: fraction of exc. cells in each selective population	0.15	
w_+: relative strength of single potentiated synapses	1.5–2.0	
w_-: relative strength of single depressed synapses	$1 - f(w_+ - 1)/(1 - f)$	
ν_{ext}: spike rate at external synapse	2.4 kHz	
Neuronal parameters:	*excitatory*	*inhibitory*
V_L: resting membrane potential	-70 mV	
θ: firing threshold	-50 mV	
H: reset potential	-55 mV	
C_m: membrane capacitance	0.5 nF	0.2 pF
g_L: membrane leak conductance	25 nS	20 nS
V_E/V_I: reversal potential (excitatory/inhibitory)	0 mV	-70 mV
τ_{rp}: refractory period	2 ms	1 ms
Synaptic parameters:	*excitatory*	*inhibitory*
d: transmission delay (fixed)	0.5 ms	0.5 ms
$g_{\text{AMPA ext}}$: external AMPA synaptic conductance	2.08 nS	1.62 nS
$g_{\text{AMPA rec}}$: recurrent AMPA synaptic conductance	104 nS/N	81 nS/N
g_{NMDA}: recurrent NMDA synaptic conductance	327 nS/N	258 nS/N
g_{GABA}: recurrent GABA synaptic conductance	1250 nS/N	973 nS/N
$[Mg^{2+}]$: extracellular magnesium concentration	1 mM	
γ: modulatory factor of magnesium block	3.57 mM	
β: gain factor in magnesium block	0.062 $(\text{mV})^{-1}$	
τ_{AMPA}: decay time of AMPA currents	2 ms	
τ_{GABA}: decay time of GABA currents	10 ms	
$\tau_{\text{NMDA},\downarrow}$: decay time of NMDA currents	100 ms	
$\tau_{\text{NMDA},\uparrow}$: rise time of NMDA currents	2 ms	
α: normalization factor for NMDA PSCs	0.5 $(\text{ms})^{-1}$	

Appendix **B**

Amit-Brunel model

In this section we summarize the network model by [7], in the particular implementation as a binary decision making model, and enumerates the parameter values we used. Here we highlight the differences of this model with respect to the model by [43] described in Appendix A, to be referred to as BW.

Network

The network is composed of N neurons, from which $N_E = 0.8N$ are excitatory and $N_I = 0.2N$ are inhibitory. Each cell in the network has a probability c of having a direct synaptic contact to any other neuron in the network, either excitatory or inhibitory. The number of contacts per neuron is therefore a random number binomially distributed, with $p = c$ and $n = N$ (variable random connectivity, p. 22). We assume that the connectivity level c is independent of the type of pre- and postsynaptic cell. Excitatory neurons are divided three different subpopulations: $p = 2$ populations, of fN_E cells each, selective to either one stimulus, and another population of cells not responding to any (of size $(1 - fp)N_E$). Apart from being sparse, the synaptic matrix is slightly different with respect to the model in A. As in the BW model, synaptic efficacies between connected cells that belong to the same selective subpopulation are stronger than the baseline level, while connections between cells from different selective populations are weaker. Unlike the BW model, *all* connections between selective and nonselective cells are weaker than the baseline level. Only connections between nonselective cells remain to the baseline level. The synaptic matrix is show in Fig. B.1

Neurons and synapses

Excitatory and inhibitory cells are modeled as leaky integrate-and-fire cells. Neuron number i of type $\alpha \in \{E, I\}$, where $i = 1 \ldots, N_\alpha$, has at time t two dynamical variables associated with it: the depolarization of the membrane potential $V_i(t)$ and the afferent postsynaptic current $I_i^{\text{syn}}(t)$. The membrane potential obeys:

$$\tau_m V_i(t) = -(V_i(t) - V_L) + I_i^{\text{syn}}(t)$$

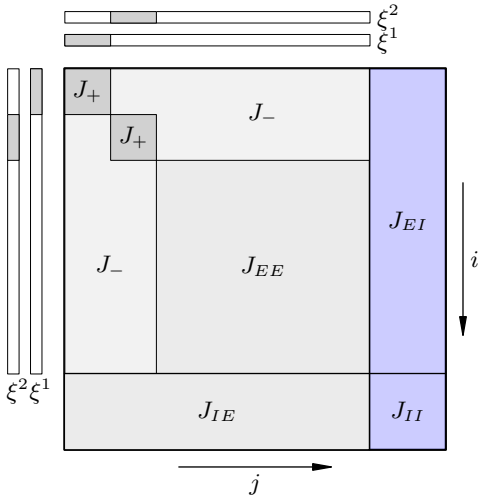


Figure B.1: Synaptic matrix of the sparse network model. The overall change in the synaptic strengths is zero if $J_- = (2 - f(p + J_+/J_{EE}))/(2 - f(p+1))$. Compare with the synaptic matrix used for the BW model.

where the synaptic current $I_i^{\rm syn}(t)$ is expressed in units of the potential. The ith neuron emits a spike when $V_i(t)$ reaches the threshold θ. Right after the emission of a spike $V_i(t)$ is reset to H, following an absolute refractory period $\tau_{\rm rp}$. $I_i^{\rm syn}(t)$ is the sum of a recurrent and an external contribution. Both synaptic currents are modeled as instantaneous injections of charge into the cell (see 2.5.2)

$$\begin{aligned} I_i^{\rm syn}(t) &= I_i^{\rm rec}(t) + I_i^{\rm ext}(t) \\ &= \sum_{j=1}^{N} J_{ij} c_{ij} \sum_k \delta(t - t_j^{(k)} - d_{ij}) + \sum_{j=1}^{N_{\rm ext}} J_{i\,\rm ext} \sum_k \delta(t - t_j^{(k)} - d_{ij}) \end{aligned}$$

where c_{ij} is a Bernoulli random variable, which is 1 if there exists a synaptic contact between neuron j and neuron i, and 0 otherwise. The probability of synaptic contact is by definition the sparseness level c, i.e., $\Pr(c_{ij} = 1) = c$. The symbol J_{ij} denotes the synaptic strength between cell j and i, whose value is drawn from a Gaussian distribution with mean $J_{\alpha\beta}$ and standard deviation $J_{\alpha\beta}\Delta$; $t_j^{(k)}$ is the time of the k-th spike emitted by the j-th presynaptic neuron, while d_{ij} is the corresponding transmission delay, which is drawn from an exponential distribution of mean $d_{i\alpha}$. The weights J_{ij}, the connectivity c_{ij}, and the transmission delays d_{ij} are fixed throughout a trial. The values used for this work are shown in Table B.1.

Table B.1: Parameters used in the integrate-and-fire simulations, using a sparse network.

Parameter	**Value**	
Network parameters:		
N: number of cells in the network	10000 cells	
N_E: number of excitatory cells	$0.8N$	
N_I: number of inhibitory cells	$0.2N$	
N_{ext}: number of cells in the external module	800	
c: sparseness (connectivity level)	0.1	
p: number of selective populations	2	
f: fraction of exc. cells in each selective population	0.15	
w_+: relative strength of single potentiated synapses	1.3–1.5	
w_-: relative strength of single depressed synapses	$[2 - f(p + w_+)]/[2 - f(p + 1)]$	
ν_{ext}: spike rate at external synapse	2.4 kHz	
Neuronal parameters:	*excitatory*	*inhibitory*
V_L: resting membrane potential	-70 mV	
θ: firing threshold	-50 mV	
H: reset potential	-55 mV	
τ_m: membrane time constant	20 ms	10 ms
τ_{rp}: refractory period	2 ms	1 ms
Synaptic parameters:	*excitatory*	*inhibitory*
d: transmission delay (random, exponential)	11.3 ms	1.2 ms
Efficacies $J_{\alpha\beta}$ are drawn from a Gaussian distribution $\rho_{\alpha\beta}$		
• external (excitatory) innervations:		
$J_{\alpha\,\text{ext}}$: mean of $\rho_{\alpha\,\text{ext}}$	0.25 mV	0.50 mV
$\Delta_{\alpha\,\text{ext}}$: relative s.d. of $\rho_{\alpha\,\text{ext}}$	0.71	0.50
• recurrent collaterals to excitatory cells:		
$J_{E\beta}$: mean of $\rho_{E\,\alpha}$	0.35 mV	-1.00 mV
$\Delta_{E\beta}$: relative s.d. of $\rho_{E\,\alpha}$	0.71	0.25
• recurrent collaterals to inhibitory cells:		
$J_{I\beta}$: mean of $\rho_{I\,\alpha}$	0.50 mV	-1.00 mV
$\Delta_{I\beta}$: relative s.d. of $\rho_{I\,\alpha}$	0.5	0.25

List of abbreviations

AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
CV	coefficient of variation
DT	decision time
EPSP	excitatory postsynaptic potential
GABA	γ-aminobutyric acid
IF	integrate-and-fire
IPSP	inhibitory postsynaptic potential
ISI	interspike interval
IT	inferior temporal
LDP	long-term depression
LIF	leaky integrate-and-fire
LIP	lateral intraparietal area
LTP	long-term potentiation
MT	middle temporal (visual area V5)
NMDA	N-methyl-D-aspartate
PFC	prefrontal cortex
PSC	postsynaptic current
PSP	postsynaptic potential
PPC	posterior parietal cortex
WTA	winner-take-all

Bibliography

1. Abbott LF, Chance FS (2005) "Drivers and modulators from push-pull and balanced synaptic input." *Prog Brain Res*, **149**, pp. 147–155. doi:10.1016/s0079-6123(05)49011-1. 4.3
2. Abbott LF, Nelson SB (2000) "Synaptic plasticity: taming the beast." *Nat Neurosci*, **3**, pp. 1178–1183. doi:10.1038/81453. 2.5.5
3. Amari S (1977) "Dynamics of pattern formation in lateral-inhibition type neural fields." *Biological Cybernetics*, **27**(2), pp. 77–87. doi:10.1007/bf00337259. 2.4
4. Amit DJ (1989) *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press. ISBN 978-0-521-42124-9. 2.4
5. Amit DJ (1995) "The hebbian paradigm reintegrated: local reverberations as internal representations." *Behav Brain Sci*, **18**(4), pp. 617–626. http://www.bbsonline.org/documents/a/00/00/04/21/bbs00000421-00/bbs.amit.html. 2.3, 2.4
6. Amit DJ, Brunel N (August 1995) "Learning internal representations in an attractor neural network with analogue neurons." *Network-Comp Neural*, **6**(3), pp. 359–388. doi:10.1088/0954-898x/6/3/004. 2.5.5
7. Amit DJ, Brunel N (1997) "Dynamics of a recurrent network of spiking neurons before and following learning." *Network-Comp Neural*, **8**(4), pp. 373–404. doi:10.1088/0954-898x/8/4/003. 2, 2.3.1, 2.5.3, 2.5.4, 2.5.5, 2.5.6, 2.5.6, 3, 5.3.1, 5.5, B
8. Amit DJ, Brunel N (1997) "Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex." *Cereb Cortex*, **7**(3), pp. 237–252. doi:10.1093/cercor/7.3.237. 2, 2.3.1, 2.4, 2.5.4, 2.5.5, 2.5.5, 2.5.6, 2.5.6, 3, 5.3.1
9. Amit DJ, Del Giudice P, Fusi S (1999) "Apprendimento dinamico della memoria di lavoro: una realizzazione in elettronica analogica." *Frontiere della vita*, **3**, p. 599. 2.2
10. Amit DJ, Evans M, Abeles M (1990) "Attractor neural networks with biological probe records." *Network-Comp Neural*, **1**(4), pp. 381–405. doi:10.1088/0954-898x/1/4/001. 2.3.1
11. Amit DJ, Fusi S (1994) "Learning in neural networks with material synapses." *Neural Comput*, **6**(5), pp. 957–982. doi:10.1162/neco.1994.6.5.957. 2.5.5, 2.5.5
12. Amit DJ, Gutfreund H, Sompolinsky H (August 1985) "Spin-glass models of neural networks." *Phys Rev A*, **32**(2), pp. 1007–1018. doi:10.1103/physreva.32.1007. 2.4
13. Amit DJ, Gutfreund H, Sompolinsky H (March 1987) "Information storage in neural networks with low levels of activity." *Phys Rev A*, **35**(5), pp. 2293–2303. doi:10.1103/physreva.35.2293. 2.4

14. Amit DJ, Gutfreund H, Sompolinsky H (1987) "Statistical mechanics of neural networks near saturation." *Ann Phys*, **173**(1), pp. 30–67. doi:10.1016/0003-4916(87)90092-3. 2.4

15. Amit DJ, Mongillo G (2003) "Spike-driven synaptic dynamics generating working memory states." *Neural Comput*, **15**(3), pp. 565–596. http://neco.mitpress.org/cgi/content/abstract/15/3/565. 2.5.3

16. Amit DJ, Tsodyks M (1992) "Effective neurons and attractor neural networks in cortical environment." *Network-Comp Neural*, **3**(2), pp. 121–137. doi:10.1088/0954-898x/2/3/003. 2.5.1, 3.1.4

17. Andersen R, Asanuma C, Essick G, Siegel R (1990) "Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule." *J Comp Neurol*, **296**(1), pp. 65–113. 1

18. Angulo MC, Rossier J, Audinat E (1999) "Postsynaptic glutamate receptors and integrative properties of fast-spiking interneurons in the rat neocortex." *J Neurophysiol*, **82**(3), pp. 1295–1302. http://jn.physiology.org/cgi/content/abstract/82/3/1295. 2.5.2

19. Arrhenius SA (1889) "Über die Reaktionsgeschwindigkeit bei der Inversion von Rohrzucker durch Säuren." *Zeitschrift für Physik Chemie*, **4**, pp. 226–248. 5.2

20. Asanuma C, Andersen R, Cowan W (1985) "The thalamic relations of the caudal inferior parietal lobule and the lateral prefrontal cortex in monkeys: divergent cortical projections from cell clusters in the medial pulvinar nucleus." *J Comp Neurol*, **241**(3), pp. 357–381. doi:10.1002/cne.902410309. 1

21. Barbas H, Mesulam M (1981) "Organization of afferent input to subdivisions of area 8 in the rhesus monkey." *J Comp Neurol*, **200**(3), pp. 407–431. doi:10.1002/cne.902000309. 1

22. Barlow H, Levick W (1969) "Changes in the maintained discharge with adaptation level in the cat retina." *The Journal of Physiology*, **202**(3), pp. 699–718. http://jp.physoc.org/cgi/content/abstract/202/3/699. 7.5

23. Bartos M, Vida I, Frotscher M, Geiger JRP, Jonas P (2001) "Rapid Signaling at Inhibitory Synapses in a Dentate Gyrus Interneuron Network." *J Neurosci*, **21**(8), p. 2687. http://www.jneurosci.org/cgi/content/abstract/21/8/2687. 2.5.2

24. Bender CM, Orszag SA (1999) *Advanced Mathematical Methods for Scientists and Engineers I*. Springer Verlag. doi:978-0387989310. 7.3.1

25. Bliss TVP, Collingridge GL (1993) "A synaptic model of memory: long-term potentiation in the hippocampus." *Nature*, **361**(6407), pp. 31–39. doi:10.1038/361031a0. 2.5.5

26. Bliss TVP, Lømo T (1973) "Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path." *J Physiol*, **232**(2), pp. 331–356. http://jp.physoc.org/cgi/content/abstract/232/2/331. 2.5.5

27. Booth V, Rinzel J (1995) "A minimal, compartmental model for a dendritic origin of bistability of motoneuron firing patterns." *J Comput Neurosci*, **2**(4), pp. 299–312. doi:10.1007/bf00961442. 2.3.1

28. Booth V, Rinzel J, Kiehn O (1997) "Compartmental model of vertebrate motoneurons for Ca^{2+}-dependent spiking and plateau potentials under pharmacological treatment." *J Neurophysiol*, **78**(6), pp. 3371–3385. http://jn.physiology.org/

cgi/content/full/78/6/3371. 2.3.1

29. Braitenberg V, Schütz A (1991) *Anatomy of the Cortex*. Springer Verlag, Berlin. ISBN 978-0-387-53233-2. 2.2, 2.5.3
30. Brascamp JW, van Ee R, Noest AJ, Jacobs R, van den Berg AV (2006) "The time course of binocular rivalry reveals a fundamental role of noise." *J Vision*, **6**(11), p. 8. doi:10.1167/6.11.8. 5.7
31. Brette R, Rudolph M, Carnevale T, Hines M, Beeman D, Bower JM, Diesmann M, Morrison A, Goodman PH, Harris FC, Zirpe M, Natschläger T, Pecevski D, Ermentrout B, Djurfeldt M, Lansner A, Rochel O, Vieville T, Muller E, Davison AP, Boustani SE, Destexhe A (2007) "Simulation of networks of spiking neurons: A review of tools and strategies." *J Comput Neurosci*, **23**(3), pp. 349–398. doi:10.1007/s10827-007-0038-6. 2.5.6
32. Britten K, Newsome W, Shadlen M, Celebrini S, Movshon J (1996) "A relationship between behavioral choice and the visual responses of neurons in macaque MT." *Vis Neurosci*, **13**(1), pp. 87–100. 1
33. Britten K, Shadlen M, Newsome W, Movshon J (1992) "The analysis of visual motion: a comparison of neuronal and psychophysical performance." *J Neurosci*, **12**(12), pp. 4745–4765. http://www.jneurosci.org/cgi/content/abstract/12/12/4745. 1, 4.1.2, 4.2, 7.5
34. Britten K, Shadlen M, Newsome W, Movshon J (1993) "Responses of neurons in macaque MT to stochastic motion signals." *Vis Neurosci*, **10**(6), pp. 1157–69. 1, 5.6.1
35. Bruce C, Goldberg M (1985) "Primate frontal eye fields. I. Single neurons discharging before saccades." *J Neurophysiol*, **53**(3), pp. 603–635. http://jn.physiology.org/cgi/content/abstract/53/3/603. 1
36. Bruce C, Goldberg M, Bushnell M, Stanton G (1985) "Primate frontal eye fields. II. Physiological and anatomical correlates of electrically evoked eye movements." *J Neurophysiol*, **54**(3), pp. 714–734. http://jn.physiology.org/cgi/content/abstract/54/3/714. 1
37. Brunel N (1996) "Hebbian learning of context in recurrent neural networks." *Neural Comput*, **8**(8), pp. 1677–1710. doi:0.1162/neco.1996.8.8.1677. 2.5.5
38. Brunel N (2000) "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons." *J Comput Neurosci*, **8**(3), pp. 183–208. doi:10.1023/a:1008925309027. 3.1.6, 3.2.1, 5.5
39. Brunel N (2000) "Persistent activity and the single-cell frequency-current curve in a cortical network model." *Network-Comp Neural*, **11**(4), pp. 261–280. doi:10.1088/0954-898x/11/4/302. 2.5.4, 2.5.5
40. Brunel N (2003) "Network models of memory." In "Methods and Models in Neurophysics, Volume Session LXXX: Lecture Notes of the Les Houches Summer School," pp. 407–476. ISBN 978-0-444-51792-0. 2, 2.3.1, 3
41. Brunel N, Hakim V (1999) "Fast global oscillations in networks of integrate-and-fire neurons with low firing rates." *Neural Comput*, **11**(7), pp. 1621–1671. doi:10.1162/089976699300016179. 5.5
42. Brunel N, Sergi S (1998) "Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics." *J Theor Biol*, **195**(1), pp. 87–95. 3.3.1, 3.3.5
43. Brunel N, Wang XJ (2001) "Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition." *J Comput*

Neurosci, **11**(1), pp. 63–85. doi:10.1023/a:1011204814320. 2.5.2, 2.5.5, 2.5.6, 2.5.6, 2.5.6, 3, 3.3.4, 3.3.5, 3.3.5, 4.1.3, 5.4, 5.4.1, 5.4.3, A, B, B, B.1

44. Brunel N, Wang XJ (2003) "What Determines the Frequency of Fast Network Oscillations With Irregular Neural Discharges? I. Synaptic Dynamics and Excitation-Inhibition Balance." *J Neurophysiol*, **90**(1), pp. 415–430. doi:10.1152/jn.01095.2002. 4.1.3
45. Burbeck SL, Luce RD (1982) "Evidence from auditory simple reaction times for both change and level detectors." *Percept Psychophys*, **32**(2), pp. 117–33. http://www.psychonomic.org/search/view.cgi?id=13446. 5.7
46. Chafee MV, Goldman-Rakic PS (1998) "Matching Patterns of Activity in Primate Prefrontal Area 8a and Parietal Area 7ip Neurons During a Spatial Working MemoryTask." *J Neurophysiol*, **79**(6), pp. 2919–2940. http://jn.physiology.org/cgi/content/abstract/79/6/2919. 2.3
47. Christie BR, Kerr DS, Abraham WC (1994) "Flip side of synaptic plasticity: long-term depression mechanisms in the hippocampus." *Hippocampus*, **4**(2), pp. 127–135. doi:10.1002/hipo.450040203. 2.5.5
48. Clements J (1996) "Transmitter timecourse in the synaptic cleft: its role in central synaptic function." *Trends Neurosci*, **19**(5), pp. 163–171. doi:10.1016/S0166-2236(96)10024-2. 2.5.2
49. Cohen M, Grossberg S (1983) "Absolute stability of global pattern formulation and parallel memory storage by competitive neural networks." *IEEE T Syst Man Cyb*, **13**, pp. 815–826. 6.1.1
50. Colby C, Duhamel J, Goldberg M (1996) "Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area." *J Neurophysiol*, **76**(5), pp. 2841–2852. http://jn.physiology.org/cgi/content/abstract/76/5/2841. 1
51. Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) "Synaptic Mechanisms and Network Dynamics Underlying Spatial Working Memory in a Cortical Network Model." *Cereb Cortex*, **10**(9), pp. 910–923. doi:10.1093/cercor/10.9.910. 4.1.3
52. Compte A, Constantinidis C, Tegnér J, Raghavachari S, Chafee MV, Goldman-Rakic PS, Wang XJ (2003) "Temporally Irregular Mnemonic Persistent Activity in Prefrontal Neurons of Monkeys During a Delayed Response Task." *J Neurophysiol*, **90**(5), pp. 3441–3454. doi:10.1152/jn.00949.2002. 3.1.1
53. Compte A, Sanchez-Vives MV, McCormick DA, Wang XJ (2003) "Cellular and Network Mechanisms of Slow Oscillatory Activity (<1 Hz) and Wave Propagations in a Cortical Network Model." *J Neurophysiol*, **89**(5), pp. 2707–2725. doi:10.1152/jn.00845.2002. 2.3.1
54. Cook EP, Maunsell JHR (2002) "Dynamics of neuronal responses in macaque MT and VIP during motion detection." *Nat Neurosci*, **5**(10), pp. 985–994. doi:10.1038/nn924. 7.5
55. Cox DR, Miller HD (1977) *The Theory of Stochastic Processes*. Chapman & Hall/CRC. ISBN 978-0-412-15170-5. 3.1.3
56. Crawford JD (October 1991) "Introduction to bifurcation theory." *Rev Mod Phys*, **63**(4), pp. 991–1037. doi:10.1103/RevModPhys.63.991. 5.3.1, 7.3
57. Curti E, Mongillo G, La Camera G, Amit DJ (2004) "Mean field and capacity in realistic networks of spiking neurons storing sparsely coded random memories."

Neural Comput, **16**(12), pp. 2597–2637. doi:10.1162/0899766042321805. 2.5.4

58. Dayan P, Abbott LF (2001) *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press. ISBN 978-0-262-04199-7. 2.5.4, 2.5.6
59. de Lafuente V, Romo R (2005) "Neuronal correlates of subjective sensory experience." *Nat Neurosci*, **8**, pp. 1698–1703. doi:10.1038/nn1587. 7.1, 7.2, 7.5
60. de Lafuente V, Romo R (2006) "Neural correlate of subjective sensory experience gradually builds up across cortical areas." *Proc Nat Acad Sci*, **103**(39), pp. 14266–14271. 7.1
61. Deco G, Martí D (2007) "Deterministic analysis of stochastic bifurcations in multi-stable neurodynamical systems." *Biological Cybernetics*, **96**(5), pp. 487–496. doi:10.1007/s00422-007-0144-6. 6
62. Deco G, Martí D (2007) "Extended method of moments for deterministic analysis of stochastic multistable neurodynamical systems." *Phys Rev E*, **75**(3), p. 31913. doi:10.1103/PhysRevE.75.031913. 6, 6.4.3
63. Deco G, Pérez-Sanagustín M, de Lafuente V, Romo R (2007) "Perceptual detection as a dynamical bistability phenomenon: A neurocomputational correlate of sensation." *P Natl Acad Sci USA*, **104**(50), pp. 20073–20077. doi:10.1073/pnas.0709794104. 7.1
64. Deco G, Scarano L, Soto-Faraco S (2007) "Weber's Law in Decision Making: Integrating Behavioral Data in Humans with a Neurophysiological Model." *J Neurosci*, **27**(42), p. 11192. doi:10.1523/jneurosci.1072-07.2007. 5.7
65. Del Giudice P, Fusi S, Mattia M (2003) "Modelling the formation of working memory with networks of integrate-and-fire neurons connected by plastic synapses." *J Physiology-Paris*, **97**(4-6), pp. 659–681. doi:10.1016/j.jphysparis.2004.01.021. 2, 2.3, 2.3.1
66. Del Giudice P, Mattia M (2001) "Long and short-term synaptic plasticity and the formation of working memory: A case study." *Neurocomputing*, **38**(40), pp. 1175–1180. 2.5.3
67. Destexhe A, Mainen ZF, Sejnowski TJ (1994) "Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism." *J Comput Neurosci*, **1**(3), pp. 195–230. doi:10.1007/bf00961734. 2.5.2, 2.5.2
68. Destexhe A, Mainen ZF, Sejnowski TJ (1998) "Kinetic models of synaptic transmission." In C Koch, I Segev, editors, "Methods in Neuronal Modeling," pp. 1–25. MIT press. ISBN 978-0-262-11231-4. 2.5.2
69. Destexhe A, Paré D (1999) "Impact of Network Activity on the Integrative Properties of Neocortical Pyramidal Neurons In Vivo." *J Neurophysiol*, **81**(4), pp. 1531–1547. http://jn.physiology.org/cgi/content/abstract/81/4/1531. 3.3.5
70. Ditterich J, Mazurek ME, Shadlen MN (2003) "Microstimulation of visual cortex affects the speed of perceptual decisions." *Nat Neurosci*, **6**(8), pp. 891–898. doi:10.1038/nn1094. 8
71. Doiron B, Rinzel J, Reyes A (2006) "Stochastic synchronization in finite size spiking networks." *Phys Rev E*, **74**(3), p. 30903. doi:10.1103/physreve.74.030903. 5.5
72. Dudek SM, Bear MF (1992) "Homosynaptic long-term depression in area CA1

of hippocampus and effects of N-methyl-D-aspartate receptor blockade." *P Natl Acad Sci USA*, **89**(10), pp. 4363–4367. doi:10.1073/pnas.89.10.4363. 2.5.5

73. Edmonds B, Gibb A, Colquhoun D (1995) "Mechanisms of activation of glutamate receptors and the time course of excitatory synaptic currents." *Annual Reviews in Physiology*, **57**(1), pp. 495–519. doi:10.1146/annurev.ph.57.030195.002431. 2.5.2

74. Egorov A, Hamam B, Fransen E, Hasselmo M, Alonso A (2002) "Graded persistent activity in entorhinal cortex neurons." *Nature*, **420**(6912), pp. 173–8. doi:0.1038/nature01171. 2.3.1

75. Ermentrout B (1992) "Complex dynamics in winner-take-all neural nets with slow inhibition." *Neural networks*, **5**(3), pp. 415–431. doi:10.1016/0893-6080(92)90004-3. 4.1.3

76. Ermentrout B (1998) "Neural networks as spatio-temporal pattern-forming systems." *Reports on Progress in Physics*, **61**(4), pp. 353–430. doi:10.1088/0034-4885/61/4/002. 4.3, 6.1.1

77. Fain GL (1999) *Molecular and cellular physiology of neurons*. Harvard University Press. ISBN 978-0-674-58155-5. 2.5.2

78. Feldmeyer D, Lübke J, Sakmann B (2006) "Efficacy and connectivity of intracolumnar pairs of layer 2/3 pyramidal cells in the barrel cortex of juvenile rats." *J Physiol*, **575**(2), p. 583. doi:10.1113/jphysiol.2006.105106. 2.2

79. Fourcaud N, Brunel N (2002) "Dynamics of the firing probability of noisy integrate-and-fire neurons." *Neural Comput*, **14**(9), pp. 2057–2110. doi:10.1162/089976602320264015. 3, 3.3.1, 3.3.5, 3.3.5

80. Funahashi S, Bruce CJ, Goldman-Rakic PS (1989) "Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex." *J Neurophysiol*, **61**(2), pp. 331–349. http://jn.physiology.org/cgi/content/abstract/61/2/331. 2.3

81. Fusi S, Annunziato M, Badoni D, Salamon A, Amit DJ (2000) "Spike-driven synaptic plasticity: theory, simulation, VLSI implementation." *Neural Comput*, **12**(10), pp. 2227–2258. http://neco.mitpress.org/cgi/content/abstract/12/10/2227. 2.5.3

82. Fuster JM (1995) *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. MIT Press. ISBN 978-0-262-56124-2. 2.3

83. Fuster JM, Alexander GE (1971) "Neuron Activity Related to Short-Term Memory." *Science*, **173**(3997), pp. 652–654. doi:10.1126/science.173.3997.652. 2.3, 8

84. Gardiner CW (1985) *Handbook of Stochastic Methods: for Physics, Chemistry and the Natural Sciences*. Springer-Verlag New York. ISBN 978-3-540-20882-2. 1, 3.1.3, 3.1.5, 3.1.6, 5.2, 5.3, 5.6, 3, 6.1.1

85. Gerstner W, van Hemmen J (1992) "Associative memory in a network of spiking neurons." *Network-Comp Neural*, **3**(2), pp. 139–164. doi:10.1088/0954-898x/3/2/004. 2.3.1

86. Gnadt J, Andersen R (1988) "Memory related motor planning activity in posterior parietal cortex of macaque." *Exp Brain Res*, **70**(1), pp. 216–20. doi:10.1007/bf00271862. 1

87. Gold JI, Shadlen MN (2002) "Banburismus and the brain decoding the relationship between sensory stimuli, decisions, and reward." *Neuron*, **36**(2), pp. 299–308.

doi:10.1016/S0896-6273(02)00971-6. 1, 5.6

88. Gold JI, Shadlen MN (2007) "The neural basis of decision making." *Annu Rev Neurosci*, **30**, pp. 535–574. doi:10.1146/annurev.neuro.29.051605.113038. 1.2, 5.6, 7.2
89. Goldman-Rakic PS (1995) "Cellular basis of working memory." *Neuron*, **14**(3), pp. 477–485. doi:10.1016/0896-6273(95)90304-6. 2.3.1
90. Gupta A, Wang Y, Markram H (2000) "Organizing Principles for a Diversity of GABAergic Interneurons and Synapses in the Neocortex." *Science*, **287**(5451), p. 273. doi:10.1126/science.287.5451.273. 2.5.2
91. Haake F, Haus JW, Glauber R (June 1981) "Passage-time statistics for the decay of unstable equilibrium states." *Phys Rev A*, **23**(6), pp. 3255–3271. doi:10.1103/physreva.23.3255. 1, 5.6
92. Hagan PS, Doering CR, Levermore CD (1989) "Mean Exit Times for Particles Driven by Weakly Colored Noise." *SIAM J Appl Math*, **49**(5), pp. 1480–1513. 3.3.1
93. Haken H (1994) "A brain model for vision in terms of synergetics." *J Theor Biol*, **171**(1), pp. 75–85. doi:10.1006/jtbi.1994.1213. 5.7
94. Hanes DP, Schall JD (1996) "Neural control of voluntary movement initiation." *Science*, **274**(5286), p. 427. doi:10.1126/science.274.5286.427. 5.7
95. Hänggi P, Talkner P, Borkovec M (1990) "Reaction-rate theory: fifty years after Kramers." *Rev Mod Phys*, **62**(2), pp. 251–341. 5.1, 5.2
96. Hastie R (2001) "Problems for judgment and decision making." *Annu Rev Psychol*, **52**(1), pp. 653–683. doi:10.1146/annurev.psych.52.1.653. 1
97. Hebb DO (1949) *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley. ISBN 978-0-8058-4300-2. 2.4
98. Hertz J, Krogh A, Palmer RG (1991) *Introduction to the Theory of Neural Computation*. Perseus Books. ISBN 978-0-201-51560-2. 4.1.2
99. Hestrin S, Sah P, Nicoll RA (1990) "Mechanisms generating the time course of dual component excitatory synaptic currents recorded in hippocampal slices." *Neuron*, **5**(3), pp. 247–53. doi:10.1016/0896-6273(90)90162-9. 2.5.2
100. Hildebrand FB (1987) *Introduction to Numerical Analysis*. Dover, 2nd edition. ISBN 978-0-486-65363-1. 2.5.6
101. Hirsch MW, Smale S (1974) *Differential equations, dynamical systems, and linear algebra*. Academic Press. ISBN 978-0-12-349550-1. 2, 7.3
102. Hodgkin AL, Huxley AF (1952) "A quantitative description of membrane current and its application to conduction and excitation in nerve." *J Neurophysiol*, **117**, pp. 500–544. doi:10.1007/bf02459568. 2.3.1
103. Hohle R (1965) "Inferred components of reaction times as functions of foreperiod duration." *J Exp Psychol*, **69**, pp. 382–6. 5.7
104. Holmgren C, Harkany T, Svennenfors B, Zilberter Y (2003) "Pyramidal cell communication within local networks in layer 2/3 of rat neocortex." *J Physiol*, **551**(1), pp. 139–153. doi:10.1111/j.1469-7793.2003.00139.x. 2.2, 2.5.3
105. Hopfield JJ (1982) "Neural networks and physical systems with emergent collective computational properties." *P Natl Acad Sci USA*, **79**, pp. 2554–2588. doi:10.1073/pnas.79.8.2554. 2.4
106. Hopfield JJ (1984) "Neurons with graded response have collective computational properties like those of two-state neurons." *P Natl Acad Sci USA*, **81**(10), pp.

3088–3092. doi:10.1073/pnas.81.10.3088. 2.4

107. Huk AC, Shadlen MN (2005) "Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making." *J Neurosci*, **25**(45), p. 10420. doi:10.1523/jneurosci.4684-04.2005. 4.2, 5.7
108. Jahr C, Stevens C (1990) "Voltage dependence of NMDA-activated macroscopic conductances predicted by single-channel kinetics." *J Neurosci*, **10**(9), p. 3178. http://www.jneurosci.org/cgi/content/abstract/10/9/3178. 2.5.2
109. Jonas P, Bischofberger J, Sandkühler J (1998) "Corelease of two fast neurotransmitters at a central synapse." *Science*, **281**(5375), p. 419. doi:10.1126/science.281.5375.419. 2.5.2
110. Jones E, Oliphant T, Peterson P, et al. "SciPy: Open source scientific tools for Python." http://www.scipy.org/. 2001–2008. 7.5
111. Kitano K, Okamoto H, Fukai T (2003) "Time representing cortical activities: two models inspired by prefrontal persistent activity." *Biological Cybernetics*, **88**(5), pp. 387–394. doi:10.1007/s00422-002-0390-6. 5.6
112. Knight BW (1972) "Dynamics of Encoding in a Population of Neurons." *J Gen Physiol*, **59**(6), pp. 734–766. doi:10.1085/jgp.59.6.734. 2.5.1
113. Koch C (1999) *Biophysics of Computation: Information Processing in Single Neurons*. Oxford University Press, New York. ISBN 978-0-19-518199-9. 2.5.1
114. Kramers HA (1940) "Brownian motion in a field of force and the diffusion model of chemical reactions." *Physica*, **7**(4), pp. 284–304. 5.2
115. Kraushaar U, Jonas P (2000) "Efficacy and stability of quantal GABA release at a hippocampal interneuron-principal neuron synapse." *J Neurosci*, **20**(15), p. 5594. http://www.jneurosci.org/cgi/content/abstract/20/15/5594. 2.5.2
116. Kubota K, Niki H (1971) "Prefrontal cortical unit activity and delayed alternation performance in monkeys." *J Neurophysiol*, **34**(3), pp. 337–347. http://jn.physiology.org/cgi/reprint/34/3/337/. 2.3, 8
117. La Camera G (1999) *Apprendimento di stimoli sovrapposti in una rete di neuroni impulsanti*. Ph.D. thesis, Università di Roma, La Sapienza. 2.5.4
118. Landauer R, Swanson JA (March 1961) "Frequency factors in the thermally activated process." *Physical Review*, **121**(6), pp. 1668–1674. doi:10.1103/PhysRev.121.1668. 5.3
119. Langer JS (September 1968) "Theory of nucleation rates." *Phys Rev Lett*, **21**(14), pp. 973–976. doi:10.1103/PhysRevLett.21.973. 5.3
120. Lapicque L (1907) "Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarization." *Journal de Physiologie et Pathologie Générale*, **3**, pp. 620–635. 2.5.1
121. Lisman JE, Fellous JM, Wang XJ (1998) "A role for NMDA-receptor channels in working memory." *Nat Neurosci*, **1**, pp. 273–275. doi:10.1038/1086. 2.3.1
122. Loewenstein Y, Sompolinsky H (2003) "Temporal integration by calcium dynamics in a model neuron." *Nat Neurosci*, **6**(9), p. 961. doi:10.1038/nn1109. 2.3.1
123. Luce RD (1991) *Response Times: Their Role in Inferring Elementary Mental Organization*. Oxford University Press US. ISBN 978-0-19-507001-9. 1, 4.1, 5.7
124. Malenka RC, Nicoll, A R (1999) "Long-term potentiation–A decade of progress?" *Science*, **285**(5435), pp. 1870–1874. doi:10.1126/science.285.5435.1870. 2.5.5

125. Marder E, Abbott L, Turrigiano GG, Liu Z, Golowasch J (1996) "Memory from the dynamics of intrinsic membrane currents." *P Natl Acad Sci USA*, **93**(24), pp. 13481–13486. doi:10.1073/pnas.93.24.13481. 2.3.1

126. Markram H, Lübke J, Frotscher M, Sakmann B (1997) "Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex." *J Physiol*, **500**(2), pp. 409–440. http://jp.physoc.org/cgi/reprint/500/pt_2/409. 2.2, 2.5.3

127. Mascaro M, Amit DJ (1999) "Effective neural response function for collective population states." *Network-Comp Neural*, **10**(4), pp. 351–373. doi:10.1088/0954-898x/10/4/305. 5.4.2

128. Mason A, Nicoll A, Stratford K (1991) "Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro." *J Neurosci*, **11**(1), p. 72. http://www.jneurosci.org/cgi/content/abstract/11/1/72. 2.2, 2.5.3

129. Matkowsky BJ, Schuss Z (1977) "The exit problem for randomly perturbed dynamical systems." *SIAM J Appl Math*, **33**(2), pp. 365–382. ISSN 00361399. http://www.jstor.org/stable/2100884. 5.3

130. Mattia M (1996) *Dinamica di una rete di neuroni impulsivi con depolarizzazione lineare*. Ph.D. thesis, Università di Roma, La Sapienza. 2.5.3, 2.7

131. Mattia M, Del Giudice P (2000) "Efficient event-driven simulation of large networks of spiking neurons and dynamical synapses." *Neural Comput*, **12**(10), pp. 2305–2329. doi:10.1162/089976600300014953. 2.5.6

132. Mattia M, Del Giudice P (November 2002) "Population dynamics of interacting spiking neurons." *Phys Rev E*, **66**(5), p. 051917. doi:10.1103/physreve.66.051917. 5.5

133. Mattia M, Del Giudice P (2003) "A distribution of spike transmission delays affects the stability of interacting spiking neurons." *Scientiae Mathematicae Japonicae*, **58**(2), pp. 335–342. http://neural.iss.infn.it/papers/scmj.pdf. 3

134. Mattia M, Del Giudice P (November 2004) "Finite-size dynamics of inhibitory and excitatory interacting spiking neurons." *Phys Rev E*. doi:10.1103/physreve.70.052903. 5.5

135. Maunsell J, van Essen D (1983) "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey." *J Neurosci*, **3**(12), p. 2563. http://www.jneurosci.org/cgi/content/abstract/3/12/2563. 1

136. Mayer ML, Westbrook GL, Guthrie PB (1984) "Voltage-dependent block by Mg^2 of nmda responses in spinal cord neurones." *Nature*, **309**(26), pp. 1–263. doi:10.1038/309261a0. 2.5.2

137. Mazurek ME, Roitman JD, Ditterich J, Shadlen MN (2003) "A role for neural integrators in perceptual decision making." *Cereb Cortex*, **13**, pp. 1257–1269. doi:10.1093/cercor/bhg097. 4.1

138. Meyer D, Osman A, Irwin D, Yantis S (1988) "Modern mental chronometry." *Biol Psychol*, **26**(1-3), pp. 3–67. doi:10.1016/0301-0511(88)90013-0. 5.7

139. Mitz A, Godschalk M, Wise S (1991) "Learning-dependent neuronal activity in the premotor cortex: activity during the acquisition of conditional motor associations." *J Neurosci*, **11**(6), pp. 1855–1872. http://www.jneurosci.org/cgi/content/abstract/11/6/1855. 5.7

140. Miyashita Y (1988) "Neuronal correlate of visual associative long-term memory

in the primate temporal cortex." *Nature*, **335**(6193), pp. 817–820. doi:10.1038/335817a0. 2.3

141. Miyashita Y, Chang HS (1988) "Neuronal correlate of pictorial short-term memory in the primate temporal cortex." *Nature*, **331**, pp. 68–70. doi:10.1038/331068a0. 2.3
142. Miyashita Y, Hayashi T (2000) "Neural representation of visual objects: encoding and top-down activation." *Curr Opin Neurobiol*, **10**(2), pp. 187–194. doi:10.1016/s0959-4388(00)00071-4. 2.3
143. Mongillo G, Amit DJ, Brunel N (2003) "Retrospective and prospective persistent activity induced by Hebbian learning in a recurrent cortical network." *Eur J Neurosci*, **18**(7), pp. 2011–2024. doi:10.1046/j.1460-9568.2003.02908.x. 5.6
144. Moreno-Bote R, Rinzel J, Rubin N (2007) "Noise-induced alternations in an attractor network model of perceptual bistability." *J Neurophysiol*, **98**(3), p. 1125. doi:10.1152/jn.00116.2007. 5.7
145. Mountcastle V (1997) "The columnar organization of the neocortex." *Brain*, **120**(4), pp. 701–722. doi:10.1093/brain/120.4.701. 2.2
146. Mountcastle VB (1957) "Modality and topographic properties of single neurons of cat's somatic sensory cortex." *J Neurophysiol*, **20**(4), pp. 408–434. http://jn.physiology.org/cgi/content/citation/20/4/408. 2.2
147. Nelson AB, Krispel CM, Sekirnjak C, du Lac S (2003) "Long-Lasting Increases in Intrinsic Excitability Triggered by Inhibition." *Neuron*, **40**(3), pp. 609–620. doi:10.1016/s0896-6273(03)00641-x. 2.3.1
148. Newsome WT, Britten KH, Movshon JA (1989) "Neuronal correlates of a perceptual decision." *Nature*, **341**(6237), pp. 52–54. doi:10.1038/341052a0. 1
149. Nowak L, Bregestovski P, Ascher P, Herbet A, Prochiantz A (1984) "Magnesium gates glutamate-activated channels in mouse central neurones." *Nature*, **307**(5950), pp. 462–465. doi:10.1038/307462a0. 2.5.2
150. Okamoto H, Fukai T (2000) "A model for neural representation of temporal duration." *BioSystems*, **55**(1-3), pp. 59–64. doi:10.1016/S0303-2647(99)00083-0. 5.7
151. Okamoto H, Fukai T (2001) "Neural mechanism for a cognitive timer." *Phys Rev Lett*, **86**(17), pp. 3919–3922. doi:10.1103/physrevlett.86.3919. 5.6, 5.7
152. Palmer J, Huk AC, Shadlen MN (2005) "The effect of stimulus strength on the speed and accuracy of a perceptual decision." *J Vision*, **5**(5), pp. 376–404. doi:10.1167/5.5.1. 5.7
153. Platt ML, Glimcher PW (July 1999) "Neural correlates of decision variables in parietal cortex." *Nature*, **400**(6741), pp. 233–238. 5.7
154. Radcliff R (1978) "A theory of memory retrieval." *Psychol Rev*, **85**, pp. 59–1080. 4.1
155. Rao SC, Rainer G, Miller EK (1997) "Integration of What and Where in the Primate Prefrontal Cortex." *Science*, **276**(5313), p. 821. doi:10.1126/science.276.5313.821. 2.3
156. Ratcliff R, Cherian A, Segraves M (2003) "A Comparison of Macaque Behavior and Superior Colliculus Neuronal Activity to Predictions From Models of Two-Choice Decisions." *J Neurophysiol*, **90**(3), pp. 1392–1407. doi:10.1152/jn.01049.2002. 4.1
157. Ratcliff R, Rouder JN (1998) "Modeling Response Times for Two-Choice

Decisions." *Psychological Science*, **9**(5), pp. 347–356. doi:10.1111/1467-9280.00067. 4.2

158. Ratcliff R, Smith PL (2004) "A Comparison of Sequential Sampling Models for Two-Choice Reaction Time." *Psychol Rev*, **111**(2), p. 333. 4.1
159. Renart A, Brunel N, Wang XJ (2003) *Computational Neuroscience: A Comprehensive Approach*, chapter 15, pp. 431–490. Chapman & Hall/CRC. ISBN 978-1-58488-362-3. 2, 3
160. Riani M, Simonotto E (May 1994) "Stochastic resonance in the perceptual interpretation of ambiguous figures: A neural network model." *Phys Rev Lett*, **72**(19), pp. 3120–3123. doi:10.1103/PhysRevLett.72.3120. 5.7
161. Ricciardi LM (1977) *Diffusion Processes and Related Topics in Biology*, volume 14 of *Lecture Notes in Biomathematics*. Springer-Verlag New York Inc. 2.5.1, 3.1.5
162. Robinson D, Fuchs A (1969) "Eye movements evoked by stimulation of frontal eye fields." *J Neurophysiol*, **32**(5), pp. 637–648. http://jn.physiology.org/cgi/content/long/32/5/637. 1
163. Rodriguez R, Tuckwell HC (1996) "Statistical properties of stochastic nonlinear dynamical models of single spiking neurons and neural networks." *Phys Rev E*, **54**(5), pp. 5585–5590. doi:10.1103/PhysRevE.54.5585. 6
164. Roitman JD, Shadlen MN (November 2002) "Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task." *J Neurosci*, **22**(21), pp. 9475–9489. http://www.jneurosci.org/cgi/content/abstract/22/21/9475. 1.1, 1, 1, 4.1, 4.2, 4.2, 5.6, 5.7, 8
165. Roxin A, Ledberg A (March 2008) "Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation." *PLoS Comput Biol*, **4**(3), p. e1000046. doi:10.1371/journal.pcbi.1000046. 1, 5.3.1, 5.6.1, 5.7, 7, 3
166. Ruthruff E (1996) "A test of the deadline model for speed-accuracy tradeoffs." *Percept Psychophys*, **58**(1), pp. 56–64. 5.7
167. Sah P, Hestrin S, Nicoll R (1990) "Properties of excitatory postsynaptic currents recorded in vitro from rat hippocampal interneurones." *The Journal of Physiology*, **430**, p. 605. http://jp.physoc.org/cgi/content/abstract/430/1/605. 2.5.2
168. Salin PA, Prince DA (1996) "Spontaneous $GABA_A$ receptor-mediated inhibitory currents in adult rat somatosensory cortex." *J Neurophysiol*, **75**(4), pp. 1573–1588. http://jn.physiology.org/cgi/content/abstract/75/4/1573. 2.5.2
169. Salzman CD, Britten KH, Newsome WT (1990) "Cortical microstimulation influences perceptual judgements of motion direction." *Nature*, **346**(6280), pp. 174–177. doi:10.1038/346174a0. 1
170. Sanchez-Vives MV, McCormick DA (2000) "Cellular and network mechanisms of rhythmic recurrent activity in neocortex." *Nat Neurosci*, **3**(1), pp. 027–1034. 2.3.1
171. Schuss Z (1980) "Singular perturbation methods in stochastic differential equations of mathematical physics." *SIAM Review*, **22**(2), pp. 119–155. ISSN 00361445. http://www.jstor.org/stable/2029958. 5.3
172. Shadlen MN, Newsome WT (1996) "Motion perception: Seeing and deciding." *P Natl Acad Sci USA*, **93**, pp. 628–633. doi:10.1073/pnas.93.2.628. 1
173. Shadlen MN, Newsome WT (October 2001) "Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey." *J Neurophysiol*, **86**, pp.

1916–1936. http://jn.physiology.org/cgi/content/abstract/86/4/1916. 1, 4.1, 5.6, 5.7, 8

174. Sjöström PJ, Turrigiano GG, Nelson SB (2001) "Rate, Timing, and Cooperativity Jointly Determine Cortical Synaptic Plasticity." *Neuron*, **32**(6), pp. 1149–1164. doi:10.1016/s0896-6273(01)00542-6. 2.2, 2.5.3

175. Softky WR, Koch C (1993) "The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs." *J Neurosci*, **13**, pp. 334–350. http://www.jneurosci.org/cgi/content/abstract/13/1/334. 2.5.1, 3.1.1

176. Sparks D, Holland R, Guthrie B (1976) "Size and distribution of movement fields in the monkey superior colliculus." *Brain Reseaarch*, **113**(1), pp. 21–34. doi:10.1016/0006-8993(76)90003-2. 1

177. Spiridon M, Gerstner W (1999) "Noise spectrum and signal transmission through a population of spiking neurons." *Network-Comp Neural*, **10**(3), pp. 257–272. doi:10.1088/0954-898x/10/3/304. 5.5

178. Spruston N, Jonas P, Sakmann B (1995) "Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons." *J Physiol*, **482**(2), pp. 325–352. http://jp.physoc.org/cgi/content/abstract/482/Pt_2/325. 2.5.2

179. Stanton PK, Sejnowski TJ (1989) "Associative long-term depression in the hippocampus induced by hebbian covariance." *Nature*, **339**(6221), pp. 215–218. doi:10.1038/339215a0. 2.5.5

180. Strogatz SH (1994) *Nonlinear Dynamics and Chaos*. Addison-Wesley Reading, MA. ISBN 978-0-7382-0453-6. 4.3, 7.4.1

181. Stüttgen MC, Schwarz C (August 2008) "Psychophysical and neurometric detection performance under stimulus uncertainty." *Nat Neurosci*, **11**(9), pp. 1091–1099. doi:10.1038/nn.2162. 7.2

182. Todd AJ, Watt C, Spike RC, Sieghart W (1996) "Colocalization of GABA, glycine, and their receptors at synapses in the rat spinal cord." *J Neurosci*, **16**(3), p. 974. http://www.jneurosci.org/cgi/content/abstract/16/3/974. 2.5.2

183. Tsodyks M, Sejnowski TJ (1995) "Rapid state switching in balanced cortical network models." *Network-Comp Neural*, **6**(2), pp. 111–124. doi:10.1088/0954-898x/6/2/001. 2.5.4

184. Tuckwell HC (1988) *Introduction to Theoretical Neurobiology: Volume 2, Nonlinear and Stochastic Theories*. Cambridge Studies in Mathematical Biology. Cambridge University Press, Cambridge. ISBN 978-0-521-35217-8. 2.5.1, 3.1.4, 3.1.6

185. Tuckwell HC, Rodriguez R (1998) "Analytical and simulation results for stochastic fitzhugh-nagumo neurons and neural networks." *J Comput Neurosci*, **5**(1), pp. 91–113. doi:10.1023/a:1008811814446. 6

186. Turrigiano GG, Nelson SB (2004) "Homeostatic plasticity in the developing nervous system." *Nat Rev Neurosci*, **5**(2), pp. 97–107. doi:10.1038/nrn1327. 2.5.5

187. van Kampen NG (1992) *Stochastic Processes in Physics and Chemistry*. North-Holland Amsterdam, 3rd edition. ISBN 978-0-444-52965-7. 1, 3.1.3, 3.1.4, 5.2, 5.6, 3, 6

188. van't Hoff JH (1884) "Études de Dynamique Chimique." *Müller & Co, Amsterdam*, **214**, p. 114. Translated by T. Ewan as "Studies in Chemical Dynamics" (London 1896). 5.2

189. Wang XJ (1999) "Synaptic basis of cortical persistent activity: the importance of

NMDA receptors to working memory." *J Neurosci*, **19**(21), pp. 9587–9603. 2.5.2

190. Wang XJ (2001) "Synaptic reverberation underlying mnemonic persistent activity." *Trends Neurosci*, **24**(8), pp. 455–463. doi:10.1016/s0166-2236(00)01868-3. 2.3.1
191. Wang XJ (2002) "Probabilistic decision making by slow reverberation in cortical circuits." *Neuron*, **36**, pp. 955–968. doi:10.1016/s0896-6273(02)01092-9. 1, 4, 4.1, 4.1, 4.1.2, 4.2, 5.3.1, 5.6, A
192. Weiss GH (1986) "Overview of theoretical models for reaction rates." *Journal of Statistical Physics*, **42**(1), pp. 3–36. doi:10.1007/bf01010838. 5.2
193. Wiggins S (2003) *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer. ISBN 0-387-00177-8. 5.3.1, 6.3, 7.3
194. Wilson H, Cowan J (1973) "A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue." *Biological Cybernetics*, **13**(2), pp. 55–80. doi:10.1007/bf00288786. 2.4
195. Winograd M, Destexhe A, Sanchez-Vives MV (2008) "Hyperpolarization-activated graded persistent activity in the prefrontal cortex." *P Natl Acad Sci USA*, **105**(20), pp. 7298–7303. doi:10.1073/pnas.0800360105. 2.3.1
196. Wong KF, Huk AC, Shadlen MN, Wang XJ (2007) "Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making." *Front Comput Neurosci*, **1**(6). doi:10.3389/neuro.10.006.2007. 4.2, 5.6.1, 6.4.1
197. Wong KF, Wang XJ (2006) "A recurrent network mechanism of time integration in perceptual decisions." *J Neurosci*. doi:10.1523/jneurosci.3733-05.2006. 1, 4, 4.2, 4.3, 4.3, 5.1, 5.4.2, 5.4.3, 5.6, 5.7, 6.4.1
198. Wurtz R, Goldberg M (1972) "Activity of superior colliculus in behaving monkey. III. Cells discharging before eye movements." *J Neurophysiol*, **35**(4), pp. 575–586. http://jn.physiology.org/cgi/reprint/35/4/575/. 1
199. Xiang Z, Huguenard JR, Prince DA (1998) "$GABA_A$ receptor-mediated currents in interneurons and pyramidal cells of rat visual cortex." *J Physiol*, **506**(3), pp. 715–730. http://jp.physoc.org/cgi/content/abstract/506/3/715. 2.5.2
200. Xu X, Collins CE, Khaytin I, Kaas JH, Casagrande VA (2006) "Unequal representation of cardinal vs. oblique orientations in the middle temporal visual area." *P Natl Acad Sci USA*, **103**(46), p. 17490. doi:10.1073/pnas.0608502103. 5.7
201. Yakovlev V, Fusi S, Berman E, Zohary E (1998) "Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations." *Nat Neurosci*, **1**, pp. 310–317. doi:10.1038/1131. 2.1
202. Zandt TV (2000) "How to fit a response time distribution." *Psychonomic Bulletin & Review*, **7**(3), pp. 424–465. http://ecl.ucsd.edu/van_Zandt_2000.pdf. 5.7
203. Zhou FM, Hablitz JJ (1997) "Rapid Kinetics and Inward Rectification of Miniature EPSCs in Layer I Neurons of Rat Neocortex." *J Neurophysiol*, **77**(5), pp. 2416–2426. http://jn.physiology.org/cgi/content/abstract/77/5/2416. 2.5.2
204. Zipser D, Kehoe B, Littlewort G, Fuster JM (1993) "A spiking network model of short-term active memory." *J Neurosci*, **13**(8), pp. 3406–20. http://www.jneurosci.org/cgi/content/abstract/13/8/3406. 2.4, 5.6