Università di Pisa
Facoltà di Scienze Matematiche Fisiche e Naturali
Corso di Laurea Specialistica in Scienze Fisiche

Anno Accademico 2015/2016

Tesi di Laurea Specialistica

Memory retrieval in Balanced Neural Networks with dynamical synapses

Candidato
Enrico Di Nardo

Relatori
Paolo Rossi
Gianluigi Mongillo

È piacevolissima ancora, per le sopradette cagioni, la vista di una moltitudine innumerabile, come delle stelle, o di persone ec. un moto molteplice, incerto, confuso, irregolare, disordinato, un ondeggiamento vago ec. che l'animo non possa determinare, né concepire definitamente e distintamente ec., come quello di una folla, o di un gran numero di formiche o del mare agitato ec.

Giacomo Leopardi
Zibaldone, 1821

Contents

Introduction

Single-cell recordings from various areas of the brain cortex in living animals have revealed a remarkably close relationship among sensory stimuli, neural activity, and perceptual states. One of the prominent phenomena that has been reported in many behavioral experiments on monkeys since the seventies [1, 2, 3, 4] is the phenomena of *selective delayed activity*: upon the presentation to the animal of some particular visual stimuli, the activity of some neurons increases and after the stimulus is removed the activity remains elevated for a long period, until the animal perform the specific behavioral task he has been trained for (fig 1 left). It is believed that the persistent activity in absence of stimulation reflects the ability of the brain to form and retain an internal representation of an item of information even when this information is not available to the senses, i.e. the neural correlate of working memory [5, 6, 7, 8].

Together with neuroscientists, during the past decades physicists have been actively involved in elaborating a theoretical framework to understand the origin of the observed neural activity [9]. A commonly accepted hypothesis is that persistent activity can be self-sustained by a feedback mechanism within sub-networks of neurons (selective for a particular stimulus) which have stronger mutual interactions [6, 8], consistently with the theory on activity dependent synaptic plasticity and cell-assembly formation proposed by Donald Hebb [10] to explain how memory traces are encoded and stored in the brain [11, 12].

According to this idea, a neural network system with a proper interaction structure should be able, in principle, to support the existence of a stable state of homogeneous activity - corresponding to the ongoing background activity in absence of stimulation, and several "memory" stable states, in which a subset of neurons (corresponding to a specific memory object) is active at a higher rate than the background state while other neuron's activity is barely unchanged. The class of models that are capable of displaying such multistability between memory states is termed *Attractor Neural Networks* and has been the object of intense studies since the seventies [13, 14, 15, 16]. Despite the successful realization of a theoretical paradigm for the delayed activity phenomena, the construction of models that realize this multista-

bility has proved problematic in reproducing realistically the characteristics of persistent activity that are observed, in particular the activity rates in both the background and the memory states are quite low (below ~ 50 Hz) compared to the maximum firing rate of a neuron ($\sim 500Hz$) and their difference is small too, with typical background rate ranging from 1 to 10 Hz and memory state roughly between 10 and 40 Hz.
Achieving two regimes of activity at realistic low rates has proven to require a significant amount of fine-tuning [17, 18, 19, 20, 21] which is a manifestation of a general problem connected to the non-linear shape of the single neuron response to a noisy input: sketchily, the equilibrium rates ν are obtained requiring self-consistency between the average input and the average output activity of a neural population, which translates into equations of the form $\nu = \phi(\nu)$ where ϕ is the input/output relation of a neuron receiving a, and since in general ϕ is a sigmoid-shaped function it's easy to get a stable low-rate solution representing the background activity, while the elevated rate state corresponding to memory occurs in a region where the activity is near saturation (see fig.1 right).

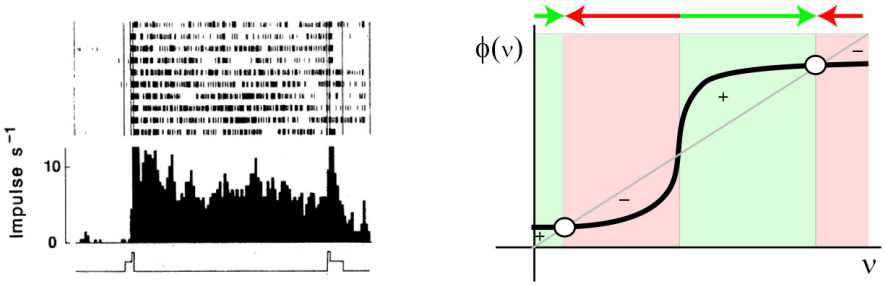

Figure 1: Left: single neuron spiking rate in a selective delayed activity experiment, adapted from [22]. Right: typical configuration of the equilibrium solution in a bistable network, adapted from [23]

In this thesis we have explored the possibility of obtaining multistability at low rates by introducing an additional nonlinearity in the system through an activity-dependent adaptation mechanism of the input received by neurons. This choice is motivated by the well established fact that the magnitude of the signals transmitted by a neuron is far from being static, instead, it is modulated on relatively short time scales as a function of the activity of the neuron itself, a phenomenon which goes under the name of *short-time synaptic plasticity* [24]. In particular, we examined wether short-term depression - a gradual reduction of the transmitted signal amplitude whenever a neuron stays active - could represent a physiologically plausible mechanism capable of lowering the exceedingly high rate of the memory solution in an attractor networks model.

The starting point will be the standard balanced network model of binary neurons [25, 19] which has the appealing property of reproducing robustly,

under mild conditions, the characteristic irregularity of neural activity as an entirely self-generated network effect (i.e. without including external source of noise). In the same spirit, the study of the attractor network model with short-term depression here presented has been conducted in the infinite size limit; the rationale of this approach is that since local cortical circuits are composed by several thousands of neurons each receiving between 10^3 and 10^4 connections, taking the limit of infinite number of neurons and connections is a strategy to obtain exact results which can be taken as reasonable approximation to the finite-size solutions, but above all, it's a way to clean the system from finite size effects and highlight its intrinsic properties.

Thesis outline

The first chapter is about background notions. After giving some essential facts of neural physiology, we review the classical theory of unstructured balanced networks exposed in [25] to set the formalism and the theoretical context in which the thesis study is set.

In chapter 2 we give an overview about the issue of exceedingly high-rate memory solution in current models of attractor neural network to put into frame the kind of the solution proposed here.

In chapter 3 we introduce a model that mimics short-term depression of chemical synaptic transmission and we derive an original mean-field theory for an unstructured balanced network of binary neurons equipped with this dynamical synaptic interaction. The prediction of the theory are tested through comparison with simulations.

In the fourth and final chapter is dedicated to examination of the central hypotheses of the thesis: by adding a structured coupling to the model presented in chapter 3 we obtain an attractor memory network model with dynamical synapses and extend the MF description to this case accordingly. It will be shown with both theory and simulations that, within certain conditions, the model is able to display bistability between a background state and a memory state with physiologically plausible low average activity.

The significance of the results and their implications are discussed in the conclusion.

Chapter 1

Modeling neural networks

"The brain is the most complex object in the universe and it's just behind the nose"

Stephane
The Science of Sleep, 2007

In this chapter we give the experimental and theoretical background of thesis. After recalling some basic notion of neurophysiology [26], we introduce the standard binary-neuron neural network model on which the thesis work is build upon. The derivation of the mean field equations for the macroscopic order parameters of the system is illustrated and it will be discussed how different scalings of the synaptic strengths yields different dynamical properties of the network in the limit of infinite number of neurons; in particular it will be shown that under certain conditions the network can have a regime of activity -the *balanced* regime- in which neurons activity exhibit a high degree of variability, consistently with observations, as an entirely dynamical effect, i.e. without any external noise source.

1.1 Neurons, synapses and local cortical circuitry

Neurons exchange electro-chemical signals with one another through specialized connections sites called synapses. Input signals are collected at *dendrites*, branched anatomical structures which receive synaptic connections from other neurons, which are then termed *pre-synaptic* neurons. The "output" part of the neural cell is a ramified wire-like structure called *axon* which makes synaptic contacts on the dendrites of other neurons, thus called *post-synaptic* neurons (see figure 1.1).
Like any other cell, a quiescent neuron maintains a difference in ions concentration (by means of ionic pumps) between its interior and the exterior, which results in an electric potential difference of about -70mV termed *resting potential* V_r. This potential can be varied by transient currents that are

generated when synapses are triggered by an input signal from a presynaptic cell. When the cell potential is depolarized beyond a certain *threshold* $V_\theta \simeq -50$mV, a series of fast bioelectrical processes provoke an explosive rise of the membrane potential (depolarization) peaking to around $+50mV$ followed by a drop below threshold in about 2msec; this sort of electrical shock-wave is termed *action potential* or "*spike*" and propagates rapidly along the axon. The arrival of the action potential at synaptic terminals of the axon trigger the release of particular molecules called neurotransmitters which bind to specialized receptors located on the membrane of the postsynaptic neuron, inducing the flow of electric current through the postsynaptic neuron membrane.
In general, synapses can be divided in two classes: *excitatory* or *inhibitory*, whether they provoke an increase or a decrease of the membrane potential. These transient variations, termed *excitatory(inhibitory) postsynaptic potential* (EPSP/IPSP), have peak amplitude of order 0.1mV and rise/decay times of order 10msec but the exact values can be quite variable.

It is a well established neurophysiological fact (Dale's law) that a neuron performs the same action (excitatory or inhibitory) at all of its output synaptic connections to other cells, regardless of the identity of the target cell, thus excitatory and inhibitory neurons can be distinguished accordingly.

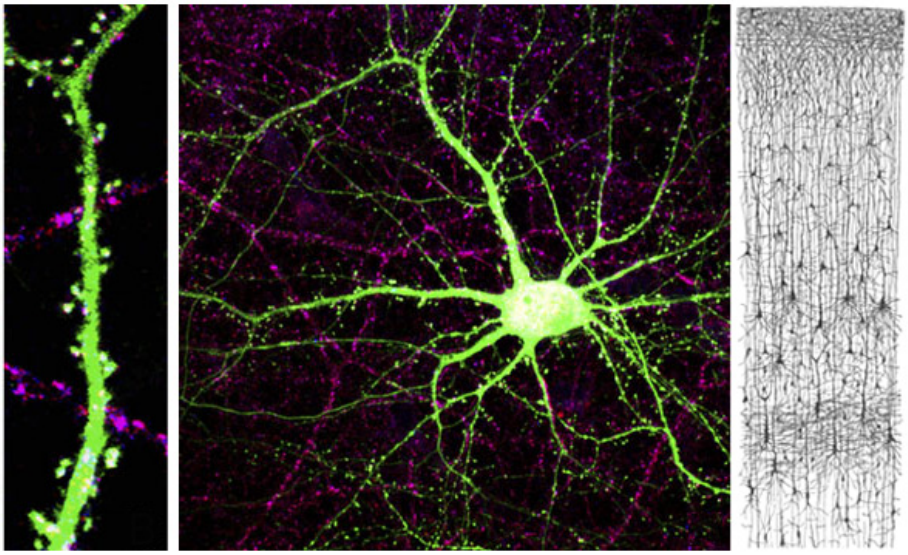

Figure 1.1: Left and Center: Mouse cortical neuron (hippocampus) marked with fluorescent dye [Courtesy of Dr. C. Hoogenraad - Erasmus MC, Rotterdam]. *Left*: Details of two axons with their presynaptic terminals (purple), making connection with the postsynaptic sites protruding from a dendritic branch (green). *Center*: the dendrites emerging from the cellular body of a pyramidal neuron (green) receiving synaptic contact from the axons of other neurons (purple) outside the field of view. *Right*: Drawings of a vertical cross-section of a human brain cortex where a random subset of neurons are stained with Golgi's method [Ramón y Cajal].

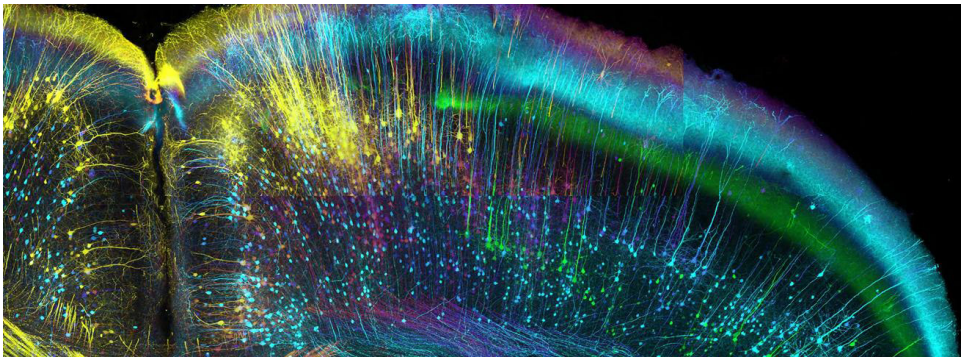

Figure 1.2: Mouse brain, coronal view (Luis de la Torre-Ubieta, Geschwind Laboratory, UCLA - Wellcome Library Ref.b23981763)

It is estimated that there are between 10^{11} and 10^{12} neurons in the human brain. On a global scale, it appears that this huge number of cells are organized and connected through communications pathways in a highly structured and complex architecture [27, 28, 29], on the other hand, since the first neuroanatomical observations of the cerebral cortex (the 3-4 mm thick outermost sheet of neural tissue of the mammalian brain) it was apparent that the structural organization of the networks on a smaller scale is remarkably similar across various cortical areas [30].

Neuroanatomy revealed that cortical neurons receive $10^3 - 10^4$ synaptic contacts, more than half of which comes from neurons within a $100 - 200$ μm radius, dropping to 10% from cells at distance of about 1 mm [31]. The majority of excitatory neurons in the cortex is composed by pyramidal neurons, a kind of cell which can have long axons making contacts both onto neighboring and distant neurons; inhibitory neurons, instead, project their axonal connection only to the cells that are within short distance, hence the name local circuit interneurons. A typical cortical neuron, then, receives excitatory inputs from both nearby cells as well as from long-range axons originating from distant cortical or subcortical regions, while inhibitory inputs are always coming from neighboring cells.

Anatomical and physiological studies [32, 33] suggested that agglomeration of about $10^4 - 10^5$ excitatory and inhibitory neurons receiving excitatory connections from a common set of distant cells, can be considered as a relatively homogeneous functional unit. The cell bodies and the connections of these assemblies of neurons are tightly packed within about 1 mm^3 of cortical tissue and are organized in layered arrays perpendicularly to the direction of the cortical surface (fig.1.1, right), hence the name *cortical columns* [32]. On this ground, it has been proposed [34, 35, 36] that cortical columns constitutes the fundamental processing module of the cortex, which can thus be regarded as a mosaic of many such local circuits with similar internal design and functioning.

Neural activity: temporal and spatial irregularity

Since the first single-neuron recording in living animals a great wealth of experiments [37, 38, 39, 40, 41, 42, 43, 44] have shown that, independently of the behavioral state and the cortical area examined, a strikingly common characteristic of neural activity is temporal irregularity.

Many studies [45, 46] have analyzed the statistics of the inter-spike intervals, and reported that the temporally irregular structure of neuron spiking is very similar to the output of Poisson point process (figure 1.6 right); moreover, across repeated trials of identical behavioral experiments, the temporal structure of spikes sequence is never the same (figure 1.6 left). Also, intracellular single-neuron recordings show that the membrane potential of the cell is strongly fluctuating (figure 1.4).

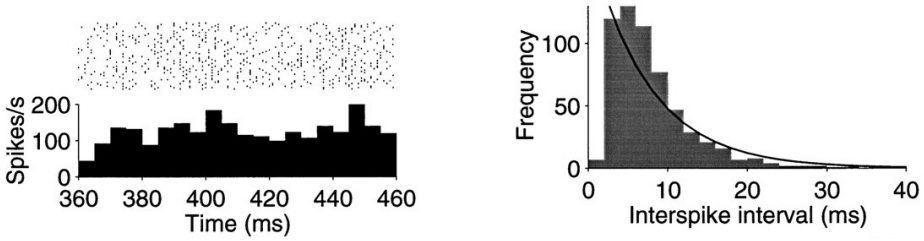

Figure 1.3: Recordings from a neuron of the middle temporal visual area of an alert monkey, adapted from [46]. *Left (top)*: raster plot of spiking sequence representing the occurrence of action potentials depicting response during 50 presentations of an identical random dot motion stimulus. *Left (Bottom)* the average spike rate, computed in 5 msec bins from the above spiking patterns. *Right*: histogram depicting the distribution of inter-spike intervals in *Left(top)* . The solid line is the best fitting exponential - the expected distribution of inter-events times for a Poisson process.

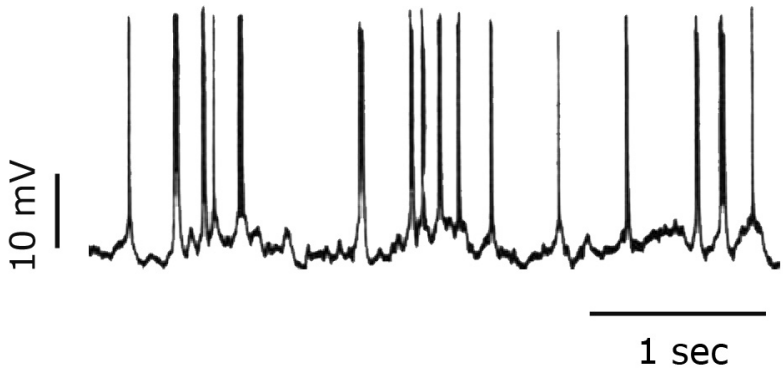

Figure 1.4: Intracellular recording of the membrane potential of a pyramidal neuron in the visual cortex of the cat. Adapted form [41]

In addition to single-cell temporal variability of the spiking patterns, the simultaneous registration of multiple neurons within local assemblies show a high degree of variability among the spiking frequency of the various neurons. Recordings in several cortical areas [37, 47, 48, 49] have reported that the typical distribution of average spiking rates in a population of neurons spans a wide interval (from zero up to several tens of hertz), is significantly long tailed, and markedly peaked at very low frequencies, meaning that most of the neurons are almost silent [50] (figure 1.5 left). In particular, the distribution seems to be well fitted by a log-normal distribution [48, 51] (figure 1.5 right).

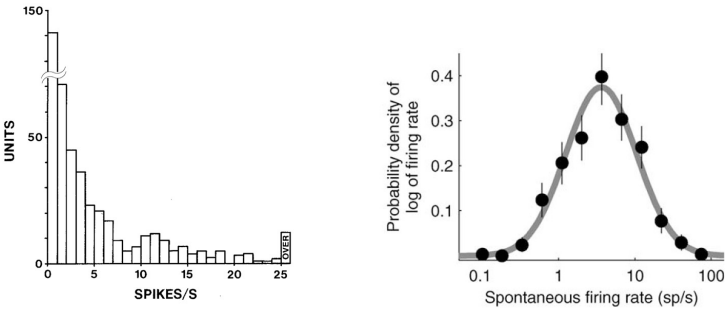

Figure 1.5: *Left*: Distribution according to average frequency of discharge of 456 neurons from the parietal cortex of a monkey, recorded simultaneously during spontaneous activity, adapted from [47]. *Right*: fit with a log-normal distribution (gray line) of the distribution of spontaneous firing rates (dots) recroded from 145 neurons of the rat auditory cortex, adapted from [48].

Why neurons spiking is highly irregular?

A long-standing problem in cortical dynamics is finding a convincing explanation to the manifest irregularity of neuron spiking seen in many experiments.

It can be hypothesized that the spiking irregularity might be due to the thermal noise at the molecular level affecting the ion channels and thus the spike ignition reactions [52, 53], but in vitro experiments show that cortical neurons fire in a highly repeatable manner in response to repeated injections of the same current, revealing that the spike generation mechanism is fairly accurate (fig. 1.6 left).

Thus, the irregularity of the in vivo neuronal activity seems to arise from the fluctuations of the membrane potential which are determined by the barrage of the signals coming from the presynaptic neurons [54, 55, 43]. However, if we consider that a typical cortical neuron receives about 10^3 to 10^4 contacts, and that a single excitatory input provoke a membrane depolarization of roughly 3-10% of the gap between resting and threshold potential which decays in about 10-20 msec [56, 57], it means that the arrival

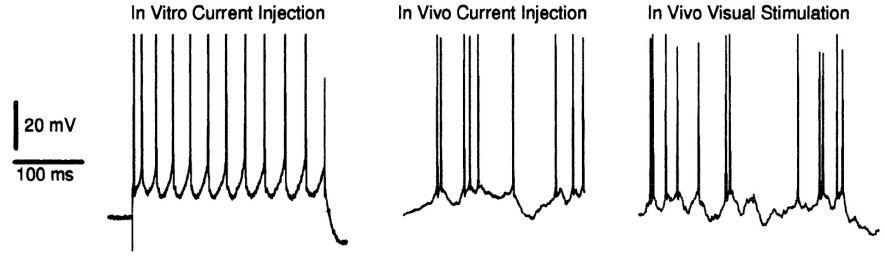

Figure 1.6: Comparison of primary visual cortex cells from adult cats in slice and in vivo: sample traces from 2 pyramidal cells, 1 from a slice (left) and 1 from an intact animal that was stimulated by current injection (middle) and by a visual stimulation (right). The lack of a large difference between spiking variability in response to current or visual stimulation in vivo is intriguing. Adapted from [43]

of a few tens of excitatory inputs within a time interval of 10-20 msec would be sufficient to bring the membrane potential above threshold, leading the neuron to spike constantly at it maximal frequency of about 500 Hz, which is by far above any mean spiking frequency recorded in living animals brain. It is a wonder, then, that the neuron can produce any variable output at all despite the large number of inputs it actually receives.

From a statistical point of view the problem can be stated as follow:

> Suppose that a set of different events results in various numbers of input being delivered to a neuron over a short period of time. The incoming action potentials appear on C separate presynaptic afferents. We would expect the number of incoming action potentials averaged across events to be proportional to C. Being an average over all events, this order-C input contains no information about the identity of any individual event, such identity is contained in the event-to-event variance of each input about its average value (note that this event-related variance is not noise; it is what make possible to discriminate among the different events).
>
> Assuming that the inputs react independently to the different events (spiking activities of neurons appears to be weakly correlated [58, 59, 60]), for central limit theorem argument the variances of the single input can be summed, producing a total event-related variance that is, like the average, proportional to C, thus the amplitude of the fluctuations that can provide information about the event identity is proportional to $\sqrt{C}$.
>
> It follows that the ratio between the component of the input that is useful for discriminating between different events and the average input is of order $\sqrt{C}/C = 1/\sqrt{C}$: for $C \sim 10^5$ this ratio is only 1%.

In other words, when integrating a large number of incoming synaptic inputs, the fluctuations will fall always far above threshold because of the large average, making impossible for a neuron to detect the informative event-to-event variations of the afferent signals.
This scenario, not only is in open contrast with the experimentally reported levels of irregular spiking and strongly fluctuating membrane potentials, but in this light the high connectivity typical of the cortex would seem to be a liability rather than an advantage. How can this paradox be solved? A number of features of neuronal circuitry that have been studied by both experimental and theoretical neuroscientists over the past 30 years can be seen as responses to this basic problem [61].
Proposed solutions can be divided into three broad classes:

- One way is to reduce the effective value of C, assuming that at any given time the inputs are effectively coming from a small fraction of the presynaptic neurons, thus making the number of the *active* synapses quite small. This hypotheses, which goes under the name of *Sparse coding* [62, 63] is a way of reducing C by keeping the number of neurons responding to any given event or feature small, thus making larger the ratio between fluctuation and mean of the input signal. Sparse coding has been proposed as an important principle of neuronal sensory processing for reasons independent of the ones given here[48, 64].

- Another possible solution is to assume that the inputs received by a neuron can be substantially correlated [65, 66] or synchronized [45, 67] and therefore their fluctuations are not averaged out: if for example is considered the extreme case in which all the inputs carry the same information, the event-related variance is of order C^2 (rather than C) and the ratio of informative to noninformative signal would be of order 1 rather than $1/\sqrt{C}$.
  Solutions of this form are seen at early stages of sensory processing, most dramatically in olfaction, where receptors with similar response characteristics project to common targets. Away from the periphery, the idea can still be applied by using correlation or synchrony to boost the efficacy of informative sets of inputs [68]. Indeed, the spike trains of pairs of neurons in cortex and in thalamus are often correlated in a time window of the order of 10 msec[69, 70]. However, the observed size of these correlations indicates that in general, only a small fraction of the incoming input is tightly correlated [71].

- Another possible solution is that (negative) inhibitory inputs to a neuron can cancel out the mean (positive) excitatory input in such a way that the mean input and the fluctuations are of the same order of magnitude, so that if the input fluctuates somewhat below threshold, highly irregular intervals between subsequent spikes can be obtained [72, 73].

> This interesting hypothesis have been examined theoretically in a seminal study [25] by van Vreeswijk and Sompolinsky where it has been shown that, under fairly mild conditions, a balance between excitatory and inhibitory inputs occurs automatically, achieving the cancelation of the noninformative part of the input to a neuron and producing temporally irregular firing as an intrinsic effect of the network dynamic, i.e. without the need of adding any external noise source. Moreover, as a result of the variability in the number of connections received by neurons, the model accounts also for another typical feature of neural activity activity seen in most experiments: the diversity of mean firing rates among different neurons in a population.

Despite a conclusive explanation of the irregularity of neural spiking is still lacking, the balancing between excitation and inhibition seems to be one of the most convincing mechanism, not only for its simplicity and effectiveness in capturing a good deal of observations but also because increasing experimental evidences [74, 75, 76, 77] are supporting the idea that the role of a dynamical balance of excitation and inhibition is fundamental in the regulation of local cortical circuits activity [78, 42, 72, 25, 46, 79].

The network model analyzed by van Vreeswijk and Sompolinsky in their landmark study [25] is capable of reproducing the spatio-temporal irregularity of neural activity with relatively simple assumptions and has the advantage to allow an analytic description. The next sections will be dedicated to review this model and its properties.

1.2 Binary model of a local cortical network

Since cortical columns are thought to be the basic module of the cerebral cortex, it is reasonable to consider these functional entities as the object of reference for neural circuitry modeling.
We will refer to the well established tradition of spin-like elements neural networks models [80, 81, 82, 16]. Despite being very schematic, this type of models can successfully reproduce many qualitative feature of biological networks with a minimal degree of complexity.

Neuron model

In a spin-like neuron model, the state of the neuron is represented by a binary variable

$$S(t) = 0, 1$$

which we will take as a representation of the activation state of the cell and is determined by the input received through some input/output relation.
All the bioelectrical processes occurring in real neurons are inherently stochastic; fundamentally, this is due to the fact that the movements of ions across the cell membrane are inevitably subject to thermal noise [52, 53]. In particular, this means that spike generation is affected, to some extent, by these fluctuations [83, 84].
We can account for this intrinsic noise at the single cell level by saying that if the net synaptic input is h, the neuron will be active with a probability

$$\mathcal{F}(h) = \frac{1}{1 + e^{-2(h-\theta)/T}} \tag{1.1}$$

where the parameter T expresses the magnitude of the microscopic noise, we call $\mathcal{F}(h)$ the *activation function* (see figure 1.7).
In the noiseless limit $T \to 0$, we get the Heaviside step function $\Theta(h - \theta)$ and the neuron state depends deterministically on the input: the neuron is active if $h > \theta$, silent otherwise.

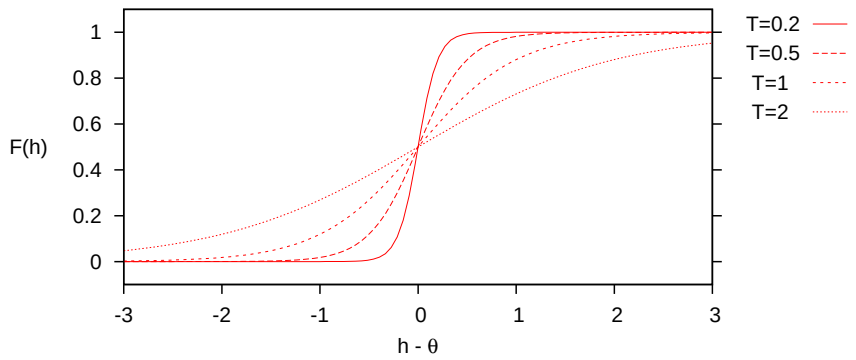

Figure 1.7: Activation function for different values of the intrinsic neuronal noise.

Network model

We consider a model network composed of N_E excitatory and N_I inhibitory neurons. In the following, every quantity relative to the i^{th} neuron of population $A = E, I$ will be indexed by a subscript i and a superscript A.
The architecture of the synaptic connections among the neurons of the network is specified by the *connectivity matrix*: c_{ij}^{AB} is 1 if neuron j of population B projects a connection to neuron i of population A, is 0 if there is no synapse between the cells.

We consider a model in which neurons are randomly connected so that every neuron has a probability c_E of receiving a connection from a cell in the excitatory population and a probability c_I of receiving a connection from a cell in the inhibitory population. In this case the elements of the *connectivity matrix* are binary i.i.d random variables

$$c_{ij}^{AB} = \begin{cases} 1 & \text{with probability } c_B \\ 0 & \text{with probability } 1 - c_B \end{cases} \tag{1.2}$$

the (quenched) structure of the network connectivity is completely specified once a particular realization of the matrix c_{ij}^{AB} is given.
In this model, the number of connections received by a generic neuron from population A is a binomial random variable k with average $C_A = c_A N_A$

$$\mathcal{B}(k, N_A, c_A) = \binom{N_A}{k} (c_A)^k (1 - c_A)^{N_A - k} \tag{1.3}$$

In general, the input received by a neuron is a function of the state of all the other neurons in the network that synapse into it. Consistently with experimental observations [85, 86] the standard assumption is that neurons operate a *linear* combination of the presynaptic inputs, therefore the total input current to the i^{th} neuron of population A can be expressed as

$$h_i^A(\{S\}) = \sum_{j=1}^{N_E} c_{ij}^{AE} J_{ij}^{AE} S_j^E - \sum_{j=1}^{N_I} c_{ij}^{AI} J_{ij}^{AI} S_j^I + h_i^{Aex} \tag{1.4}$$

where $\{S\}$ indicates the state of all other neurons in the network.
The first and second term in (1.4) represent respectively the total excitatory and inhibitory signals coming from the other neurons of the network; J_{ij}^{AB} is the *synaptic strength matrix* which determine the amplitude of the postsynaptic current pulse received by neuron i in population A when the presynaptic neuron j of population B is active.
The last term accounts for all the excitatory inputs coming from neurons which are outside the network.

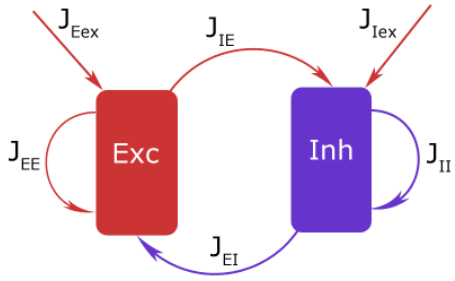

Figure 1.8: Schematic representation of the the model network considered.

We make the simplifying assumption that synaptic strengths are constants and depend only on the type of population of the pre and postsynaptic neurons (see figure 1.8), namely

$$J_{ij}^{AB} = \mathcal{J}_{AB}$$

It is also assumed that the external excitatory input is constant: each neuron in population A receives exactly C synapses from outside of strength J_{Aex} and the strength of the external input is indicated with the constant m_0.

Under our hypotheses, the instantaneous input to the neuron labeled i of population A has the following form

$$h_i^A(t) = \mathcal{J}_{AE} \sum_{j=1}^{N} c_{ij}^{AE} S_j^E(t) - \mathcal{J}_{AI} \sum_{j=1}^{N} c_{ij}^{AI} S_j^I(t) + C\mathcal{J}_{Aex} m_0 \tag{1.5}$$

1.3 Network dynamics

Our network model is a system of $N_E + N_I$ binary units that thus can assume $2^{N_E+N_I}$ possible configuration, called *microstates*, each of which can be represented as a string whose elements are the states of each neurons:

$$X = (S_1^E, ..., S_{N_E}^E, S_1^I, ..., S_{N_I}^I) \qquad \text{where} \quad S_i^A = 0, 1$$

A time evolution of the system would correspond to a journey through subsequent microstates within the configuration space $\Gamma = \{X\}$. To make this possible, a rule of transition among microstate must be specified.

If the transition rule is deterministic, given any initial microstate X_0, the path of the system in the space Γ is unique and is completely described by a certain sequence of consecutive microstates.
With a stochastic dynamical rule, instead, transitions among states are probabilistic, thus multiple realization of the time-evolution starting from the same microstate X_0 can yield different paths in Γ; for this reason the evolution of the system has to be described by a (discrete) probability distribution function $P(X, t)$ expressing the likelihood that at a certain time t the system configuration corresponds to X.

A random sequential dynamic rule in discrete time (Glauber dynamics) is considered for the network update:

> At every time step the population is selected at random: the excitatory with probability p_E, the inhibitory with probability $p_I = 1 - p_E$; within the chosen population, a neuron is selected at random and its state at the following time-step is updated according to
>
> $$S_i^A(t+1) = \begin{cases} 1 & \text{with probability} \quad \mathcal{F}[\, h_i^A(t)\,] \\ 0 & \text{with probability} \quad 1 - \mathcal{F}[\, h_i^A(t)\,] \end{cases} \tag{1.6}$$

Given that at time-step t the configuration of the network is the microstate $X = (S_1^A, ..., S_{N_A}^A, S_1^B, ..., S_{N_B}^B)$, we want to calculate the probability that at the following time-step $t+1$ the network make a transition to the microstate $X_i^A = (S_1^A, .., 1 - S_i^A, .., S_{N_A}^A, S_1^B, ..., S_{N_B}^B)$ in which neuron S_i^A has flipped state. This is given by the probability that population A is updated (p_A) times the probability that neuron i is updated ($1/N_A$) times the probability that the neuron flips its state

$$\begin{aligned} \mathcal{W}(X \to X_i^A) =& \frac{p_A}{N_A} \Big| \, S_i^A - \mathcal{F}\left[\, h_i^A(X) \,\right] \Big| \\ =& \frac{p_A}{N_A} \Big\{ S_i^A + (1 - 2S_i^A)\, \mathcal{F}\left[\, h_i^A(X) \,\right] \Big\} \end{aligned} \tag{1.7}$$

Passing from a discrete to a continuous time description we should introduce a time scale to define the transition rates: for any neuron in population A the average number of time-step between two updating is $n_A = N_A/p_A$, if we set τ_A as the continuous-time equivalent [1] of n_A we can derive that a single time step correspond to a time interval of

$$\Delta = \frac{\tau_A}{n_A} = \tau_A \frac{p_A}{N_A}$$

thus we can write the transition rate as

$$w(X \to X_i^A) = \frac{\mathcal{W}(X \to X_i^A)}{\Delta} = \frac{1}{\tau_A} \left\{ S_i^A + (1 - 2S_i^A)\, \mathcal{F}\left[\, h_i^A(X) \,\right] \right\} \quad (1.8)$$

Since the probability of moving to next state depends only on the present state through the input, the time evolution of the network in its configuration space is a Markov process with non-vanishing transition probabilities only between microstates which differ by the state of a single neuron, and the time evolution for the probability distribution $P(X,t)$ can be described by the master equation

$$\frac{dP(X,t)}{dt} = \sum_A^{E,I} \left[\sum_{i=1}^{N} P(X_i^A, t)\; w(X_i^A \to X) - \sum_{i=1}^{N} P(X,t)\; w(X \to X_i^A) \right] \quad (1.9)$$

with the transition rates given in (1.8).

Macrostates and Ergodicity

For any observable defined over the points of the phase space $Q(X)$ we can define a *time average* along a path $X(t)$, assuming discrete time:

$$\overline{Q} = \lim_{T \to \infty} \frac{1}{T} \sum_{n=0}^{T} Q(X(t_n))$$

For any given initial condition, the system can follow different paths depending on the random sequence of updates. Therefore, we define the *ensemble average* of the observable

$$\langle Q \rangle(t) = \sum_{X \in \Gamma} P(X,t) Q(X)$$

When the system reaches the equilibrium we have that $\partial_t P(X,t) = 0$, thus the ensemble average of any observable is constant as well and serves as a

[1] Actually, the exact value of the two time constant is immaterial, the relevant quantity is their ratio, expressing the relative frequency of updating of one population with respect to the other:

$$\frac{\tau_I}{\tau_E} = \frac{p_E}{p_I} \frac{N_I}{N_E}$$

macroscopic parameter. A stationary state of the system associated to a set of macroscopic parameters is called *macrostate*.
We can say that a discrete dynamical system is *ergodic* if its time evolution is such that every point of the phase space is eventually visited. In terms of probability distribution, it means that regardless of the initial distribution

$$\lim_{n\to\infty} P(X, t_n) \neq 0 \qquad \forall X \in \Gamma$$

If a system is ergodic, the time average of a quantity is equal to the ensemble averages $\overline{Q} = \langle Q \rangle$.
In particular, it can be demonstrated that irreducible Markov chains with finite state spaces are ergodic systems. When considering systems in the thermodynamic limit $N \to \infty$ ergodicity breaking can occur: the phase space is partitioned into mutually disconnected regions and, at equilibrium, the system dynamics is confined in one of those regions; in such case the time average is equivalent to the ensemble average of the distribution restricted to the subspace where the system evolves.

Local average activity dynamics

If we take S_i^A as an observable, its average over the probability distribution corresponds to the probability that neuron i of population A is in the active state at time t, we define it *local average activity*:

$$m_i^A(t) \equiv \langle S_i^A \rangle(t)$$

taking the time derivative of the expression and using the definition of ensemble average we have

$$\frac{d}{dt}\, m_i^A(t) = \sum_{X\in\Gamma} S_i^A(X) \frac{dP(X,t)}{dt} \tag{1.10}$$

where $\sum_X$ represents the sum over all possible configurations.
If we substitute the master equation (1.9) in (1.10) we get:

$$\begin{aligned}\frac{dm_i^A}{dt} =& \sum_X \sum_{j=1}^{N_A} S_i^A P(X_j^A, t)\; w(X_j^A \to X) - \sum_X \sum_{j=1}^{N_A} S_i^A P(X,t)\; w(X \to X_j^A) \\ &+ \sum_X \sum_{j=1}^{N_B} S_i^A P(X_j^B, t)\; w(X_j^B \to X) - \sum_X \sum_{j=1}^{N_B} S_i^A P(X,t)\; w(X \to X_j^B)\end{aligned}$$

Since the above summations are taken over all possible configurations, for each configuration X in which $S_j^A(X) = 1$ there is a complementary configuration $\overline{X}$ in which $S_j^A(\overline{X}) = 0$ and vice versa, therefore we can make the substitution

$$P(X_j^C, t)\; w(X_j^C \to X) = P(X,t)\; w(X \to X_j^C)$$

which holds for any j if $C = B$ and for all $j \neq i$ if $C = A$. For the terms with $j = i$ in the sum relative to population A we can take

$$S_i^A P(X_i^A, t)\, w(X_i^A \to X) = (1 - S_i^A) P(X, t)\, w(X \to X_i^A)$$

It follows that the last two terms cancel and the r.h.s. reads

$$\sum_X P(X,t) \left[(1 - S_i^A)\, w(X \to X_i^A) + \sum_{j \neq i}^{N} S_i^A\, w(X \to X_j^A) - \sum_{j=1}^{N} S_i^A\, w(X \to X_j^A) \right]$$

The terms in the first summation cancel the corresponding terms in the second summation and we are left with

$$\begin{aligned} \frac{d}{dt} m_i^A(t) &= \sum_X P(X,t)(1 - 2S_i^A)\, w(X \to X_i^A) \\ &= \langle (1 - 2S_i^A)\, w(X \to X_i^A) \rangle \end{aligned} \tag{1.11}$$

substituting the expression for the transition rate (1.8) and using the fact that $S_i^2 = S_i$ we get a differential equation for the time evolution of the local average activity of neuron i in population A:

$$\tau_A \frac{d}{dt} m_i^A(t) = -m_i^A(t) + \left\langle \mathcal{F}[\, h_i^A \,] \right\rangle (t) \tag{1.12}$$

Population average activity dynamic

For a description of the network in term of a macroscopic quantity it is natural to consider the mean over the entire population A of the local average activities, *population average activity*

$$m_A = \frac{1}{N_A} \sum_{i=1}^{N_A} S_i^A \tag{1.13}$$

The dynamical equation for the statistical average $\langle m \rangle(t) = \sum_\Gamma m(X) P(X,t)$ is obtained simply by taking the over the populations of (1.12), which gives:

$$\boxed{\tau_A \frac{d}{dt} \langle m_A \rangle(t) = -\langle m_A \rangle(t) + \frac{1}{N_A} \sum_{i=1}^{N_A} \left\langle \mathcal{F}[\, h_i^A \,] \right\rangle (t)} \tag{1.14}$$

The calculation of the average in the last term is not trivial, nevertheless, in the limit of $N \to \infty$ and $C \to \infty$, the term can be calculated exactly yielding a mean field equation for $\langle m \rangle(t)$ which in the limit can be taken as an equation for the non averaged quantity $m(t)$ itself, since in the infinite size limit its fluctuations are vanishingly small and its value converge to $\langle m \rangle(t)$.

To simplify notation from now on we take $N_E = N_I = N$ and $C_E = C_I = C$.

1.4 Mean Field theory

As is commonly done in the study of large systems of many interacting elements we can adopt a mean field approach to obtain a reduced description of the system in terms of the macroscopic order parameters, which in our case are the population average activities m_E and m_I.
The idea is to substitute the different input to each neuron with a self consistent average input field, in order to carry out the average in the last term of (1.14). This approximation become exact in the thermodynamical limit, namely when the number of neurons and connections goes to infinity, and when taking this limit it's necessary to assume a scaling relationship between the number of neurons in the network and the number of connections each neurons receive.

In conventional fully connected network models, for instance, the symmetry of interactions allows to analyze the properties of the system with the tools of equilibrium statistical mechanics in the $N \to \infty$ limit [87]. Although it's impossible to use the same approach if interactions are not symmetric, as in a randomly connected network, in [88] it has been shown that an exact analytic solution can be found also in this case, provided that the number of connections per neurons is much smaller than the neurons in the network: $C \ll N$, a situation which is referred to as *sparse connectivity*.

Analogously, the mean field theory (MFT) for the model here presented has been derived under the assumption of *sparse connectivity* limit [25] and eventually it was considered the *high connectivity* limit, namely when the average number of connections received by a neuron is very large (consistently with the biological reality where $C \sim 10^3 - 10^4$). The resulting mean field theory applies then to a network such that connectivity is *high* but still *sparse*, i.e. the mean number of connection per unit is large, but much smaller than the number of cells in the network:

$$1 \ll C \ll N$$

Technically, this is done by taking the limit $N \to \infty$ while keeping fixed the number of connections per neurons C and then subsequently taking the limit $C \to \infty$.
The detailed derivation of the mean field equations as done in [25] is retraced in what follows.

> (From now on, we drop all population index and consider as if the network is made of a single population to avoid cumbersome notation)

Sparse connectivity limit: uncorrelated neurons activities

Let's write explicitly the statistical average in the last term of (1.14).

$$\frac{1}{N}\sum_{i=1}^{N}\langle\mathcal{F}[\ h_i\ ]\rangle\,(t)=\frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{X\in\Gamma}P(X,t)\mathcal{F}[\ h_i(X)\ ]\right\}\equiv \mathrm{F}(t) \tag{1.15}$$

Let's consider that neuron i receives a number K_i of connections from presynaptic neurons labeled $j_1,...,j_{K_i}$, the net input at time t reads

$$h_i(t)=J\sum_{r=1}^{K_i}S_{j_r}(t)$$

There are K_i+1 possible input values, each associated with a set of the possible configuration of the K_i presynaptic neurons. Thus, the probability of having a certain input depends on the probability of having the corresponding presynaptic neuron configuration.
In principle, the activity of any two presynaptic neurons $S_{j_a}(t)$ and $S_{j_b}(t)$ can be correlated because they might have common afferents input, but if $N\gg C$, then any two couple of neurons have negligible probability to share an input coming from the same neuron, even indirectly, therefore their activity are *uncorrelated* to a very good approximation[2]. It can be proved that this hold rigorously when $C\ll\log N$ [90, 91]:

> At time-step t a neuron will typically receive inputs which have arrived from C^t neurons (ancestors) in the initial configuration at $t=0$. The states of two neurons at time t will be uncorrelated if the two trees of ancestor neurons connecting them to the initial state have no neurons in common. Two sets of Q randomly chosen neurons, out of the N available ones, will not have neurons in common [3] is $Q\ll\sqrt{N}$. The condition that neurons will remain uncorrelated after a time t is therefore $C^t\ll\sqrt{N}$. This, in turn, implies that C must be smaller than any power of N, namely
>
> $$C\ll\log N \tag{1.16}$$

[2]This situation is consistent both with anatomical studies showing that cortical neurons in the same cortical column receive quite distinct sets of inputs [89], and functional experiments that measures weak correlations among pairs of neurons in the same column [58, 59, 71]

[3]The probability for none of the neurons in the second set of Q randomly chosen neurons not to belong to the first set is

$$Pr=\left(1-\frac{Q}{N}\right)^{Q}\simeq\exp\left(-\frac{Q^2}{N}\right)$$

where the approximation hold for $Q\gg 1$. If $Q=aN^b$ then $Pr\simeq\exp(-a^2N^{2b-1})$ Consequently, when taking the limit $N\to\infty$, if $b<1/2$ we have that $Pr\to 1$, i.e. the two sets will certainly have no neurons in common.

The absence of correlations between the afferent signals permits to factorize the joint probability distribution of the system into the product of the probability distribution for the activity of the individual neurons

$$P(X,t) \;=\; P(S_1,...,S_N,t) \;=\; \prod_{i=1}^{N} p\,(S_i,t) \tag{1.17}$$

where, according to the definition of local average activity, the single neuron probability distribution can be written as

$$p\,(S_i,t) \;=\; S_i(2m_i(t)-1)+1-m_i(t) \;=\; \begin{cases} m_i & S_i=1 \\ 1-m_i & S_i=0 \end{cases}$$

Putting the factorized probability distribution (1.17) in the expression (1.15) we can marginalize all the factors relative to $n-K_i$ neurons that are not included in the presynaptic tree of neuron i so that we are left with a sum over the 2^{K_i} possible configurations of the K_i presynaptic neurons of i

$$\mathrm{F}(t) = \frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{\{S_r\}}\prod_{r=1}^{K_i}\left[\,S_{j_r}(2m_{j_r}(t)-1)+1-m_{j_r}(t)\,\right]\;\mathcal{F}\left[J\sum_{r}^{K_i}S_r\right]\right\}$$

At this point is convenient to take the *average over the possible realization of the network architecture* which can be considered equivalent to an average over all the possible shapes of pre-synaptic trees of a generic neuron.
We proceed in two steps:

- first, with K fixed, we take the average over all possible sets of K presynaptic neurons. Since S_r are dummy variables, the only part that depends on the connectivity structure is the marginalized probability distribution, which depends on the set of the local average activity of the presynaptic neurons $(m_{j_1},...,m_{j_K})$, we then have

$$\begin{aligned} &\prod_{r=1}^{K}\frac{1}{N}\sum_{j_r=1}^{N}\left[\,S_{j_r}(2m_{j_r}(t)-1)+1-m_{j_r}(t)\,\right] \\ &=\;\left[\,S_{j_r}(2m(t)-1)+1-m(t)\,\right]^K \end{aligned} \tag{1.18}$$

  upon averaging over the randomness in the connectivity, the averaged activity of all the presynaptic neurons is equal to the population activity $m(t)$, therefore the number of active presynaptic neurons becomes the sum of independent identically distributed (binary) random variables, thus a binomial random variable, which allow us to write

$$\begin{aligned} &\sum_{\{S_r\}}\left[\,S_r(2m(t)-1)+1-m(t)\,\right]^K\mathcal{F}\left[J\sum_{r}^{K}S_r\right] \;= \\ &=\sum_{n=0}^{K}\binom{K}{n}(m(t))^n(1-m(t))^{K-n}\;\mathcal{F}\left[\,Jn\,\right] \end{aligned}$$

- If we rewrite the full expression we have

$$\mathrm{F}(t) = \frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{n=0}^{K_i}\binom{K_i}{n}(m(t))^n(1-m(t))^{K_i-n}\,\mathcal{F}\left[\,Jn\,\right]\right\} \tag{1.19}$$

and the last step consists in taking the average over the population, thus over the possible number of connection K_i that a neuron can receive, which has a binomial distribution (1.3) but in the sparse connectivity limit $N \gg C$ can be approximated by a Poisson distribution with average C, therefore we have:

$$\begin{aligned}\mathrm{F}(t) &= \sum_{K=0}^{\infty}\frac{C^K}{K!}\,e^{-C}\left\{\sum_{n=0}^{K}\binom{K}{n}(m(t))^n(1-m(t))^{K-n}\,\mathcal{F}\left[\,Jn\,\right]\right\}\\ &= \sum_{n=0}^{\infty}\left[\sum_{K=n}^{\infty}\frac{C^K}{K!}\,e^{-C}\binom{K}{n}(m(t))^n(1-m(t))^{K-n}\right]\mathcal{F}\left[\,Jn\,\right]\\ &= \sum_{n=0}^{\infty}\frac{(mC)^n}{n!}e^{-mC}\,\mathcal{F}\left[\,Jn\,\right]\end{aligned} \tag{1.20}$$

which tells us that the number of active inputs n is a random variable with Poisson statistics with mean and variance equal to Cm.

High connectivity limit: Gaussian distributed input

From the central limit theorem we have that if the neurons receives a large number of inputs the Poisson distribution in (1.20) can be approximated by a Gaussian distribution with mean Cm and variance Cm, therefore in the high connectivity limit we can replace n with a continuous random variable x and write

$$\mathrm{F}(t) = \frac{1}{\sqrt{Cm2\pi}}\int e^{-\frac{(x-Cm)^2}{2Cm}}\mathcal{F}\left[\,J\,x\,\right]\mathrm{d}x \tag{1.21}$$

$$= \frac{1}{\sqrt{2\pi}}\int e^{-\frac{x^2}{2}}\mathcal{F}\left[\,JCm + x\sqrt{J^2Cm}\,\right]\mathrm{d}x \tag{1.22}$$

where in the argument of the activation function $\mathcal{F}$ we can read the expression for the "mean field" in the sparse and high connectivity limit: a Gaussian distributed random variable with mean equal to JCm and variance equal to J^2Cm corresponding to the average input.
For the complete network model the mean and the average of the input are

$$h_A(t) = C\left[\,\mathcal{J}_{AE}m_E(t) - \mathcal{J}_{AI}m_I(t) + \mathcal{J}_{Aex}m_0\,\right] \tag{1.23a}$$

$$\sigma_A^2(t) = C\left[\,J_{AE}^2m_E(t) + \mathcal{J}_{AI}^2m_I(t)\,\right] \tag{1.23b}$$

Network with fixed number of connection per neuron

It is useful to derive the MF equations for network architecture in which all the neurons receive exactly C input. In this case, the population average in 1.19 is straightforward because $K_i = C$ for all i and we get

$$\mathrm{F}^{\mathrm{C}}(t) = \sum_{n=0}^{C} \binom{C}{n} (m(t))^n (1-m(t))^{C-n} \, \mathcal{F}\,[\,Jn\,] \tag{1.24}$$

in the high connectivity limit $C \ll 1$ (for central limit theorem argument) the binomial distribution is approximated by a Gaussian with mean $\mu = Cm$ and variance $\sigma^2 = C(m - m^2)$ thus the input are also Gaussian distributed random variables with mean and variance given by

$$h_A(t) = C\Big[\, \mathcal{J}_{AE} m_E(t) - \mathcal{J}_{AI} m_I(t) + \mathcal{J}_{Aex} m_0 \,\Big] \tag{1.25a}$$

$$\sigma_A^2(t) = C\left[\mathcal{J}_{AE}^2 m_E(t)(1-m_E(t)) + \mathcal{J}_{AI}^2 m_I(t)(1-m_I(t))\right] \tag{1.25b}$$

We highlight that the variance of the input, compared to the case of randomly connected network, is lessened by a factor m^2 because in this case there is no variability associated to the difference in the number of input received by neurons in the population, the only source of variability in the input is the temporal fluctuations in the activity of the presynaptic neurons.

Mean Field Equations

To summarize, the statistical independence of the neurons activities in the sparse connectivity limit $N \gg C$, allows the statistics of the input to converge to a Gaussian statistics in the $C \gg 1$ limit. Under these hypotheses the resulting MF dynamic equations for the two order parameters of the network are

$$\begin{cases} \tau_E \dfrac{dm_E(t)}{dt} = -m_E(t) + \mathrm{F}\big(\, h_E(t), \sigma_E(t) \,\big) \\ \tau_I \dfrac{dm_I(t)}{dt} = -m_I(t) + \mathrm{F}\big(\, h_I(t), \sigma_I(t) \,\big) \end{cases} \tag{1.26}$$

$$\mathrm{F}(\, h, \sigma \,) \;=\; \int \mathcal{F}\,[\, h + x\,\sigma \,] \, \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \, \mathrm{d}x$$

where the arguments of F are given by (1.23) (or (1.25)).
The description of the network dynamics has been reduced to a closed system for the population average activities $m_E(t)$ and $m_I(t)$ as if the two populations where a pair of mutually coupled (and self-coupled) neurons receiving a stochastic input whose statistics is Gaussian and is completely determined by the coupling strength and the average activities.

We stress that in the $N \to \infty$ and $C \to \infty$ limits, with $C/N \to 0$, this description of the population activities dynamics is *exact*.

1.5 Interaction scaling and Balanced state

To find the network activities at equilibrium we look for steady state solutions of the mean field equations (1.26):

$$\begin{cases} m_E = \int \mathcal{F}\left[\, h_E(m_E,\, m_I) + x\, \sigma_E(m_E, m_I) \,\right] \dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\, \mathrm{d}x \\ m_I = \int \mathcal{F}\left[\, h_I(m_E,\, m_I) + x\, \sigma_I(m_E, m_I) \,\right] \dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\, \mathrm{d}x \end{cases} \tag{1.27}$$

$$h_A(m_E,\, m_I) = C\Big[\, \mathcal{J}_{AE} m_E - \mathcal{J}_{AI} m_I + \mathcal{J}_{Aex} m_0 \,\Big] \tag{1.28a}$$

$$\sigma_A^2(m_E,\, m_I) = C\Big[\, \mathcal{J}_{AE}^2 m_E + \mathcal{J}_{AI}^2 m_I \,\Big] \tag{1.28b}$$

As pointed out in section 1.1, the crucial aspect of neural activity is that neuron's output is irregular and far below saturation despite each cell receives a large number of inputs $C \sim 10^3 - 10^4$ whose amplitude, relative to the resting-threshold potential gap, is such that a neuron's output saturates whenever a few tens of excitatory presynaptic inputs are active during an integration time interval. The central idea of [25] to solve the issue is that the observed neural activity is the result of a situation in which the magnitude of the input fluctuations are comparable with the distance of the input from threshold.

It is then useful to analyze the network in the $C \to \infty$ limit, not only because the equations (1.27) becomes an exact analytical description of the system and their solution are a good approximation of the finite $C \gg 1$ solution, but above all because considering the $C \to \infty$ is a way to clean the system from finite size effects and the qualitative features of the specific model network can be highlighted.
When considering the $C \to \infty$ limit, one is forced to normalize the connection strengths by (some power of) the number of connections per cell C, in order to keep the total afferent input within the (presynaptic) dynamic range of the neuron (whose order of magnitude is given by the distance between reset and threshold). As we show below, different scaling schemes of J with C lead to different relative magnitudes of the mean and fluctuations of the afferent current into the cells in the extensive limit, and this in turn determines the type of steady-state solutions for the network. For instance if we assume the generic scaling

$$\mathcal{J}_{AB} = \frac{J_{AB}}{C^a} \qquad \mathcal{J}_{Aex} = \frac{J_{Aex}}{C^a} \qquad a > 0 \tag{1.29}$$

where J_{AB} are constant, the mean and the variance of the input will be

$$h_A(t) = C^{1-a}\{J_{AE} m_E - J_{AI} m_I + J_{Aex} m_0\} \tag{1.30}$$

$$\sigma_A^2(t) = C^{1-2a}\{J_{AE}^2 m_E + J_{AI}^2 m_I\} \tag{1.31}$$

and when taking the $C \to \infty$ we have different possible scenarios:

- $a > 1$. Both the mean and the variance go to zero.
- $a = 1$. The mean input stays finite but the variance goes to zero
- $1 > a > 1/2$. The mean input diverges and the variance goes to zero.
- $a = 1/2$. Mean input diverges but variance stays finite.
- $a < 1/2$. Both mean and variance diverge

We examine the two significative cases in which either the mean or the variance are finite and independent on network size

$\mathbf{J \sim 1/C}$

This scaling is the common to most MF theories of large, highly connected neural networks, yielding:

$$h_A(t) = \; J_{AE} m_E - J_{AI} m_I + J_{Aex} m_0 \tag{1.32}$$

$$\sigma_A^2(t) = \; \frac{1}{C} \left[J_{AE}^2 m_E + J_{AI}^2 m_I \right] \tag{1.33}$$

This implies that when taking the $C \to \infty$ limit, the variance vanishes, thus the equations depend only on the average input:

$$\begin{cases} m_E = \mathcal{F}\big(h_E(m_E,\, m_I) \big) \\ m_I = \mathcal{F}\big(h_I(m_E,\, m_I) \big) \end{cases} \tag{1.34}$$

Such networks converge either to either static states or globally coherent limit cycles [92, 93, 94].
In particular, if we consider noiseless neurons taking the limit $T \to 0$ in (1.1), we would obtain that $\mathcal{F}(h) \to \Theta(h - \theta)$ and only trivial solutions are possible with $m_A = 0$ or $m_A = 1$ if, respectively, the self-consistent equilibrium input is below threshold ($h_A < \theta_A$) or above it ($h_A > \theta_A$).

$\mathbf{J \sim 1/\sqrt{C}}$

If $a = 1/2$ the variance is of order 1 and the mean input of order $\sqrt{C}$:

$$\boxed{\begin{aligned} h_A(t) &= \; \sqrt{C} \; \left[\, J_{AE} m_E - J_{AI} m_I + J_{Aex} m_0 \, \right] && \text{(1.35a)} \\ \sigma_A^2(t) &= \; J_{AE}^2 m_E + J_{AI}^2 m_I && \text{(1.35b)} \end{aligned}}$$

Clearly, one might think that for large C values the distance of the input from the threshold would be much greater than the amplitude fluctuations

and, depending on the sign of the net average input, the network activity would be either saturated of zero, analogously to the previous case.
However, if the population rates are such that *the inhibitory input cancels the net excitatory input to leading order*, the total input (1.35a) can stay finite in the infinite C limit as well and we can have the so called

> ***balanced state***: a $C \to \infty$ solution of the MF equations (1.61) in which the distance of the input from the threshold is of the same order of magnitude of the input fluctuations.

If we write m_A as a power series in $1/\sqrt{C}$

$$m_A = m_A^{(0)} + \frac{m_A^{(1)}}{\sqrt{C}} + \frac{m_A^{(2)}}{C} + ... \qquad\qquad m_A^{(n)} \in \mathbb{R} \tag{1.36}$$

and then substitute it into (1.35a) to obtain an expansion in power of $1/\sqrt{C}$ for the average input:

$$h_A = \sqrt{C}\left(J_{AE}\left[m_E^{(0)} + \frac{m_E^{(1)}}{\sqrt{C}} + ...\right] - J_{AI}\left[m_I^{(0)} + \frac{m_I^{(1)}}{\sqrt{C}} + ...\right] + J_{Aex}m_0\right)$$

$$= \sqrt{C}\left(J_{AE}m_E^{(0)} - J_{AI}m_I^{(0)} + Am_0\right) \; + \; (J_{AE}m_E^{(1)} - J_{AI}m_I^{(1)}) \; + \; ...$$

it can be seen that the coefficient of n-th order in the input series depends on the $(n+1)$-th order coefficient of the activity series (1.36), namely:

$$h_A = h_A^{(-1)}\sqrt{C} + h_A^{(0)} + \frac{h_A^{(1)}}{\sqrt{C}} + \frac{h_A^{(2)}}{C} + ... \tag{1.37}$$

$$\begin{aligned} h_A^{(-1)} &= J_{AE}m_E^{(0)} - J_{AI}m_I^{(0)} + J_{Aex}m_0 \\ h_A^{(n)} &= J_{AE}m_E^{(n+1)} - J_{AI}m_I^{(n+1)} \qquad\qquad \text{for } n = 0, 1, 2... \end{aligned}$$

To have a balanced state we require that the total input does not diverge when C goes to infinity. This can occur only if the coefficient of the leading term in (1.37) vanishes for both population; thus, in the $C \to \infty$ limit, the balancing condition is

$$\begin{cases} h_E^{(-1)} = 0 \\ h_I^{(-1)} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} J_{EE}m_E^{(0)} - J_{EI}m_I^{(0)} + J_{Eex}m_0 = 0 \\ J_{IE}m_E^{(0)} - J_{II}m_I^{(0)} + J_{Iex}m_0 = 0 \end{cases} \tag{1.38}$$

Solving this linear system we obtain the leading term in the expansion of m, which in the $C \to \infty$ limit corresponds to the actual population activity

$$m_E^\infty \equiv \lim_{C\to\infty} m_E = m_E^{(0)} = \frac{J_{Eex}J_{II} - J_{Iex}J_{EI}}{J_{EI}J_{IE} - J_{EE}J_{II}}\ m_0 \ \equiv\ \Omega_E\, m_0 \tag{1.39a}$$

$$m_I^\infty \equiv \lim_{C\to\infty} m_I = m_I^{(0)} = \frac{J_{Eex}J_{IE} - J_{Iex}J_{EE}}{J_{EI}J_{IE} - J_{EE}J_{II}}\ m_0 \ \equiv \Omega_I\, m_0 \tag{1.39b}$$

For this solution to be meaningful we require that rates should be positive, therefore that network parameters have to respect one of the following chain of inequalities:

$$\frac{J_{Eex}}{J_{Iex}} < \frac{J_{EI}}{J_{II}} < \frac{J_{EE}}{J_{IE}} \qquad\qquad \frac{J_{Eex}}{J_{Iex}} > \frac{J_{EI}}{J_{II}} > \frac{J_{EE}}{J_{IE}} \tag{1.40}$$

The notable result of (1.39) is that in the $C \to \infty$ limit the average activities become *linear* functions of the external input m_0, with a coefficient of proportionality Ω_A depends solely on the synaptic strength.

We point out that the synaptic strength scaling $J \sim 1/\sqrt{C}$ is just for convenience in thinking about the problem, it's a way to consider "strong" synaptic interactions (compared to the $J \sim 1/C$ case); the physiologically relevant assumptions are only that excitatory and inhibitory inputs are separately much larger than their difference and that the latter is of the same order as their fluctuations.

Mean input in the balanced state

The balancing condition (1.38) requires that the $O(\sqrt{C})$ term of the average input must vanish, so that in the $C \to \infty$ limit only the the $O(1)$ term is left, thus for the average input we have

$$\begin{aligned} h_E^\infty &\equiv \lim_{C\to\infty} h_E = h_E^{(0)} = J_{EE}m_E^{(1)} - J_{EI}m_I^{(1)} \\ h_I^\infty &\equiv \lim_{C\to\infty} h_I = h_I^{(0)} = J_{IE}m_E^{(1)} - J_{II}m_I^{(1)} \end{aligned} \tag{1.41}$$

But we only know $m_A^\infty = m_A^{(0)}$, not $m_A^{(1)}$. Nevertheless, the balanced solution allows us to calculate the input variance $\sigma_A^\infty = m_0\,(J_{AE}^2\Omega_E + J_{AI}^2\Omega_I)$, so we can consider the MF equation as implicit equations for the average input:

$$m_A^\infty \equiv \Omega_A m_0 = \int \left\{1 + \exp\left[\frac{-2(h_A^\infty - \theta_A + x\sigma_A^\infty)}{T}\right]\right\}^{-1} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\mathrm{d}x \tag{1.42}$$

More properly, the equation (1.42) determines implicitly the distance of the mean input from threshold

$$u_A^\infty \equiv h_A^\infty - \theta_A \tag{1.43}$$

in fact, in the $C \to \infty$ limit the value of the threshold is immaterial and the network will organize its activity in order to have that the *distance* between the mean input and the threshold is of order 1.

Unbalanced and partially balanced states

If the leading term in the neuron input expansion (1.37) does not vanish, in the $C \to \infty$ the input can diverge to $+\infty$ or $-\infty$, depending on the sign of the coefficient $h_A^{(-1)}$, which means that the corresponding populations will be always active ($m_A = 1$) or silent ($m_A = 0$). We call **unbalanced states** the self-consistent solutions in which one or both populations are either active at maximum rate or inactive. Here, we examine all possible solutions of this type and determine the corresponding condition of existence:

- $(m_I = m_E = 0)$
  This solution is trivially excluded because it would require that both inputs are negative when the entire network is off, and this can never happen because the external input is always positive.

- $(m_E = 1;\, m_I = 0)$
  This solution cannot exist because it would require that $h_I < 0$ with a purely excitatory (positive) input.

- $(m_E = 0;\, m_I = 1)$
  The inputs are consistent with the rates if:

$$\begin{cases} h_E < 0 \\ h_I > 0 \end{cases} \Rightarrow \begin{cases} -J_{EI} + J_{Eex} m_0 < 0 \\ -J_{II} + J_{Iex} m_0 > 0 \end{cases} \Rightarrow \begin{cases} m_0 < \dfrac{J_{EI}}{J_{Eex}} \\ m_0 > \dfrac{J_{II}}{J_{Iex}} \end{cases}$$

  The interval in the external input m_0 for which this solution is possible is not empty if

$$\frac{J_{Eex}}{J_{Iex}} < \frac{J_{EI}}{J_{II}} \tag{1.44}$$

- $(m_I = m_E = 1)$
  here we require that both inputs are positive

$$\begin{cases} h_E > 0 \\ h_I > 0 \end{cases} \Rightarrow \begin{cases} J_{EE} - J_{EI} + J_{Eex} m_0 > 0 \\ J_{IE} - J_{II} + J_{Iex} m_0 > 0 \end{cases} \tag{1.45}$$

$$m_0 > \max\left\{ \frac{J_{EI} - J_{EE}}{J_{Eex}}; \frac{J_{II} - J_{IE}}{J_{Iex}} \right\}$$

  since m_0 has to be positive, this solution exists for any $m_0 > 0$ if

$$(J_{EI} < J_{EE}) \quad \text{and} \quad (J_{II} < J_{IE}) \tag{1.46}$$

There are also solutions in which one population is "unbalanced", i.e. off or saturated, while the other is 'balanced" in the sense that the average activity is between 0 and 1. We refer to this class of solutions as **partially balanced states**. Let's consider, for instance, a solution of the form

$$m_E = 0 \quad ; \quad 0 < m_I < 1$$

the inputs are self-consistent if

$$\begin{cases} -J_{EI} m_I + J_{Eex} m_0 < 0 \\ -J_{II} m_I + J_{Iex} m_0 = 0 \end{cases}$$

solving the balancing condition for m_I and substituting in the unbalancing condition we get that this type of solution can exist only if

$$\frac{J_{Eex}}{J_{Iex}} < \frac{J_{EI}}{J_{II}}$$

which turns out to be the same condition of existence for the unbalanced solution ($m_E = 0$; $m_I = 1$).

Balancing conditions

If we want meaningful (positive) balanced rates solutions, and exclude any unbalanced solution[4], we require that the network parameter satisfy the inequalities (1.40) and, simultaneously, brake the conditions (1.44, 1.46). This implies that we must have the following constraints

$$\boxed{\frac{J_{Eex}}{J_{Iex}} > \frac{J_{EI}}{J_{II}} > \frac{J_{EE}}{J_{IE}} \qquad \text{and} \qquad J_{EI} > J_{EE}} \tag{1.47}$$

Linear perturbation analysis (appendix A.2) shows that the stability with respect to perturbation of order $\delta m \sim 1/\sqrt{C}$ depends essentially on the ratio between the time constants τ_E/τ_I. Depending on this ratio the balanced state can be a fixed point attractor or a stable limit cycle. Stability to larger perturbation is also guaranteed for appropriate values of the ratio τ_E/τ_I (see [25] for details).

[4] at least in some interval of the external input

1.6 Spatial and Temporal variability

The population average activity m_A does not distinguish between spatial and temporal fluctuations of the activity level. For instance, a certain population average rate m might be the outcome of a state where all the individual cell average activity is m or, on the other extreme, it can results from a state in which a fraction m of the cells in the network is always active while all others are always silent.
To characterize the statistics of both temporal and spatial fluctuations in the activities in the balanced state we observe that the variability of a neuron output is determined by the variability of the input which in turn is influenced by two factors:

- *Fast noise*: the temporal fluctuation of the presynaptic inputs coming from the stochastic flipping of the presynaptic neurons
- *Quenched noise*: the variability in the number and identity of the afferent connection to each neuron

We have seen that in the sparse and high connectivity limit the statistics of the inputs becomes Gaussian. We can then decompose the total input fluctuation as the sum of two fluctuating part: a quenched spatially fluctuating part and a temporally fluctuating part, both having gaussian statistics[5] for central limit theorem arguments.
The input to a generic neuron at equilibrium in the balanced state can thus be written as

$$\begin{aligned} h_i^A(t) &= h_A + \sigma^A x_i^A = \\ &= h_A + s_A \xi_i^A + \sqrt{\sigma_A^2 - s_A^2}\, \eta_i^A(t) \end{aligned} \tag{1.49}$$

with both ξ_i and $\eta_i(t)$ gaussian variable with zero mean and unit variance.

The fluctuation of the time-averaged rates across the population s_A are called *quenched fluctuations*. They can be calculated taking the variance of the temporally averaged input at equilibrium

$$s_A^2 = \frac{1}{N} \sum_{i=1}^{N} \left[h_A - \langle h_i^A \rangle \right]^2$$

[5] If X and Y are independent random variables that are normally distribute their sum is also normally distributed with mean and average given by the sum of the mean and average of the two

$$\begin{matrix} X \sim G(\mu_X, \sigma_X^2) \\ Y \sim G(\mu_Y, \sigma_Y^2) \end{matrix} \qquad X + Y \sim G(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2) \tag{1.48}$$

Following the same step of the calculation of (A.6) it is easily seen that

$$s_A^2 = J_{AE}^2 q_E + J_{AI}^2 q_I \tag{1.50}$$

where the order parameter q is defined as

$$\boxed{q_A = \frac{1}{N}\sum_{i=1}^{N} m_i^2} \tag{1.51}$$

since $m \in [0:1]$ the quantity q is bounded $m_A^2 \leqslant q_A \leqslant m_A$ and its relation to the average population activity gives information about the degree of *freezing* in a network, i.e. wether a part of the neurons have non-fluctuating activity, being permanently active or silent. For instance in the extreme case in which all neurons are frozen we would have that $q = m$ since $m_i = m_i^2 = 0$ or 1, on the other hand for non-frozen neuron $m \in (0;1)$; thus the more q is closer to its upper bound, the more is likely to have a certain degree of neuron freezing.

Distribution of average local activities

Separating the total input fluctuation into the temporal and the quenched part as in (1.49) we have that the time averaged *local* activity at equilibrium can be rewritten as

$$\begin{aligned} m_i^A = \langle S_i^A(t)\rangle &= \left\langle \mathcal{F}\left[\, h_A + \, s_A\,\xi_i \, + \sqrt{\sigma_A^2 - s_A^2}\;\eta_i(t)\right]\right\rangle \\ &= \int \mathcal{F}\left[\, h_A + \, s_A\,\xi_i \, + \sqrt{\sigma_A^2 - s_A^2}\;\eta\,\right]\frac{e^{-\eta^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\eta \end{aligned} \tag{1.52}$$

Therefore the spatial inhomogeneity of the input induces a spatial inhomogeneity in the local activities, which can be regarded as a random variable resulting from the transformation of the spatial fluctuation of the input ξ :

$$m(\xi) = \int \mathcal{F}\left[\, h_A + \, s_A\,\xi \, + \sqrt{\sigma_A^2 - s_A^2}\;\eta\,\right]\frac{e^{-\eta^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\eta \tag{1.53}$$

The distribution of m is readily determined recalling that given an invertible function f and a random variable ξ whose distribution is $G(\xi)$, the random variable $m = f(\xi)$ has probability distribution $\rho(m)$ given by

$$\rho(m) = G\big(\xi(m)\big)\frac{1}{\left|\frac{df}{d\xi}\right|_{\xi(m)}} \tag{1.54}$$

where $\xi(m) = f^{-1}(m)$.
In general, to calculate the average of any function of the local rates $g(m)$

it is not necessary to know the analytic form of $\rho(m)$, in fact we can use the function (1.53) to change variable in the integrand and obtain an average over the explicit (gaussian) distribution of the quenched disorder:

$$\bar{g} = \int_0^1 g(m)\rho(m)dm \;=\; \int_{-\infty}^{+\infty} g(m(\xi))\,\frac{e^{-\xi^2/2}}{\sqrt{2\pi}}\,d\xi \tag{1.55}$$

It is straightforward to see that the order parameter of the system m_A and q_A corresponds to the first two moments of the probability distribution of the activity

$$m_A = \frac{1}{N}\sum_{i=1}^{N} m_i^A \qquad \to \int_0^1 m\,\rho_A(m)\,dm \tag{1.56a}$$

$$q_A = \frac{1}{N}\sum_{i=1}^{N} \left(m_i^A\right)^2 \qquad \to \int_0^1 m^2\,\rho_A(m)\,dm \tag{1.56b}$$

Calculating (1.56a) using (1.55) we recover (1.27) as expected:

$$m_A = \int \left[\int \mathcal{F}\left[\, h_A + s_A\,\xi + \sqrt{\sigma_A^2 - s_A^2}\,\eta \,\right] \frac{e^{-\eta^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\eta \right] \frac{e^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \tag{1.57}$$

$$= \int \mathcal{F}\left[\, h_A + \sigma_A x \,\right] \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,\mathrm{d}x \tag{1.58}$$

the last identity is an obvious consequence of the fact that the distribution of the sum of two random variables is the convolution of their distribution.

Equation (1.56b) gives us the system of equations for the order parameters q_A at equilibrium:

$$\begin{cases} q_E = \int \left\{ \int \mathcal{F}\left[\, h_E + s_E\,\xi + \sqrt{\sigma_E^2 - s_E^2}\,\eta \,\right] \frac{e^{-\eta^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\eta \right\}^2 \frac{e^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \\ q_I = \int \left\{ \int \mathcal{F}\left[\, h_I + s_I\,\xi + \sqrt{\sigma_I^2 - s_I^2}\,\eta \,\right] \frac{e^{-\eta^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\eta \right\}^2 \frac{e^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \end{cases} \tag{1.59}$$

$$s_A^2 = J_{AE}^2 q_E + J_{AI}^2 q_I$$

which is a closed system for (q_E, q_I) after the stationary rates m_A have been determined resolving (1.27) (or (1.39) for $C \to \infty$) and used to calculate σ_A and h_A with (1.35) (or (1.42) for $C \to \infty$)

1.7 Noiseless neurons

As we have seen, in the previous sections, the solutions for the populations activities are found through the balance condition which does not involve the single neuron response function (1.1). The specific form of the single neuron response function comes into play for the determination of the residual inputs (1.42) and of the order paramenter q. Therefore, if we consider the noiseless $T \to 0$ limit in (1.1) (so that $\mathrm{F}(h)$ becomes a step function) we keep the qualitative picture while making calculations more simple.

For noiseless neurons, we have that (1.27) simplifies to

$$\begin{aligned} \mathrm{F}(\,h,\sigma\,) \;&=\; \int_{-\infty}^{+\infty} \Theta\,[\,h\,-\theta+x\,\sigma]\;\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\;\mathrm{d}x \\ &=\; \int_{\theta}^{\infty}\frac{e^{-\frac{(x-h)^2}{2\sigma}}}{\sigma\sqrt{2\pi}}\;\mathrm{d}x \;=\; \int_{\frac{\theta-h}{\sigma}}^{\infty}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\;\mathrm{d}x \end{aligned}$$

Defining $H(z)$ the complementary cumulative distribution function of a Normal Distribution with zero mean and unit variance

$$H(z) \equiv \int_z^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx \;=\; \frac{1}{2}\left[1-\mathrm{erf}\left(\frac{z}{\sqrt{2}}\right)\right] \;=\; \frac{1}{2}\,\mathrm{erfc}\left(\frac{z}{\sqrt{2}}\right) \qquad (1.60)$$

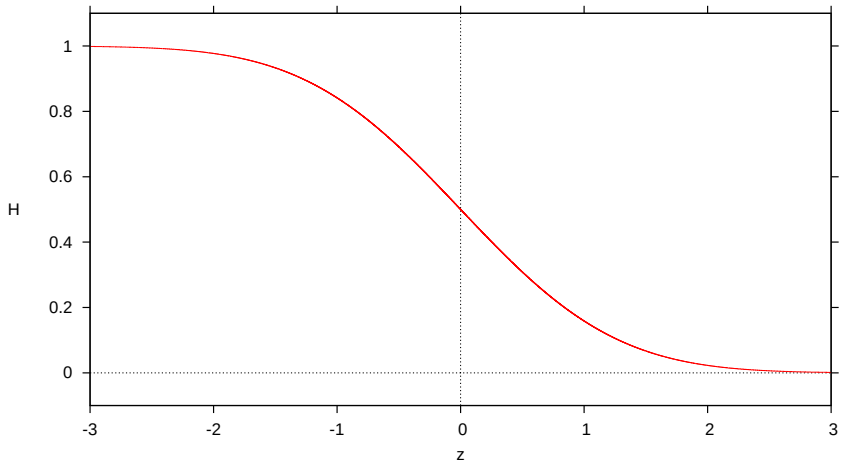

Figure 1.9: Function $H(z)$

MF equations for the order parameters

We write explicitly the systems for the stationary values of the population average activity (1.27) in the noiseless neurons limit

$$\begin{cases} m_E = H\left(\dfrac{\theta_E - h_E(m_E, m_I)}{\sigma_E(m_E, m_I)}\right) \\ m_I = H\left(\dfrac{\theta_I - h_I(m_E, m_I)}{\sigma_I(m_E, m_I)}\right) \end{cases} \tag{1.61}$$

$$h_A(t) = \sqrt{C}\ \left[\ J_{AE} m_E - J_{AI} m_I + J_{Aex} m_0\ \right] \tag{1.62a}$$

$$\sigma_A^2(t) = J_{AE}^2 m_E + J_{AI}^2 m_I \tag{1.62b}$$

The temporal average in (1.53) can be expressed explicitly as

$$m^A(\xi) = H\left(\frac{\theta_A - h_A + s_A\,\xi}{\sqrt{\sigma_A^2 - s_A^2}}\right) \tag{1.63}$$

and the system (1.64) read:

$$\begin{cases} q_E = \displaystyle\int \left[H\left(\frac{\theta_E - h_E + s_E\,\xi}{\sqrt{\sigma_E^2 - s_E^2}}\right)\right]^2 \frac{e^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \\ q_I = \displaystyle\int \left[H\left(\frac{\theta_I - h_I + s_I\,\xi}{\sqrt{\sigma_I^2 - s_I^2}}\right)\right]^2 \frac{e^{-\xi^2/2}}{\sqrt{2\pi}}\,\mathrm{d}\xi \end{cases} \tag{1.64}$$

$$s_A^2 = J_{AE}^2 q_E + J_{AI}^2 q_I \tag{1.65}$$

Distribution of local activities

In the case of a noiseless neuron the calculation of the distribution of local activities (1.54) can be done explicitly by means of (1.63) and gives:

$$\rho_A(m) = \frac{1}{\beta_A} \exp\left[\ \frac{-\xi^2 + (\alpha_A + \beta_A \xi)^2}{2}\ \right]\Bigg|_{\xi=\xi(m)} \tag{1.66}$$

$$\alpha_A = \frac{\theta_A - h_A}{\sqrt{\sigma_A - s_A}} \qquad\qquad \beta_A = \frac{s_A}{\sqrt{\sigma_A - s_A}}$$

It can be demonstrated (appendix A.5) that in the low input limit $m_0 \ll 1$ the second moment of the distribution $\rho(m)$ can be approximated to leading order as

$$q_A \ = \ m_A^2 + O(m_A^3 |\log m_A|) \tag{1.67}$$

therefore, at very low input, for the variance of the population distribution of activity $\rho(m)$ is $q^2 - m^2 \simeq O(m^3|\log m|)$ and the distribution is narrowly peaked around the mean value (see figure 1.10, left panel $m_E = 0.02$) namely, *at very low level of activity there is a low spatial variability in the local rates*. In fact from the ratio of the temporal fluctuation to the total fluctuation

$$\frac{\sigma^2 - s^2}{\sigma^2} \sim 1 - m + O(m_A^2|\log m_A|)$$

it appears that for $m \ll 1$ the input fluctuations comes mainly from the temporal fluctuations while quenched fluctuations become becomes much less prominent: the network behavior is close to that of a network where all the neurons receive the same number of contacts.

Increasing the external input the average activity grows, the distribution becomes more skewed showing longer tail, meaning that there is an increased degree of spatial variability in the local activities across the network (see figure 1.10 to compare the shape of the distribution for different levels of average activity).

Increasing the level of activity in the network further more, a state in which temporal and quenched fluctuations are equal $s_A^2 = \sigma_A^2/2$ is eventually reached; let us call m_0^* the value of the external input for which this occurs. It can be demonstrated (see A.3) that for $m_0 > m_0^*$ the population distribution of activity diverges in $m = 0$ and $m = 1$ (see right panel in figure 1.10).

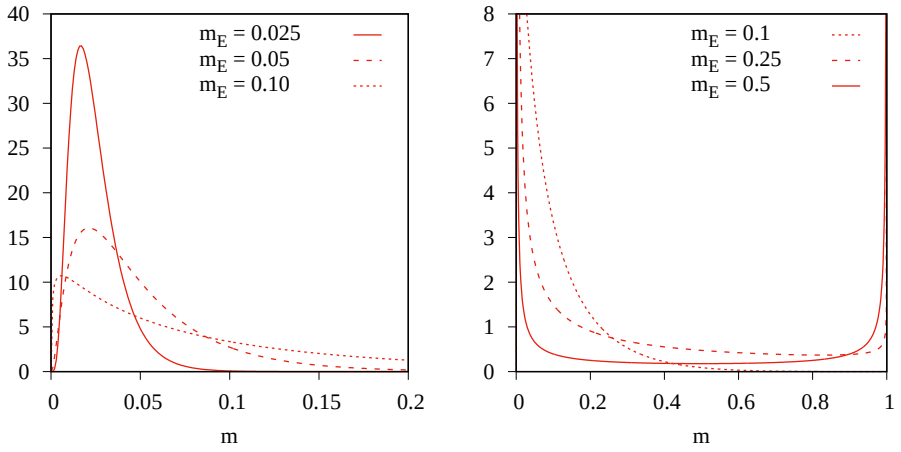

Figure 1.10: Qualitative differences between the distributions of local excitatory activities $\rho_E(m)$for different mean value m_E. Calculated from (1.66) in the $C \to \infty$ limit with parameters 1.70

In traditional disordered spin system the source of the fast (temporal) fluctuations is an external heath bath: the high temperature state is dominated by temporal fluctuations while at low temperature partial freezing of

the spins occurs. Attempting to make an analogy we can say that states of low activity of the balanced network, dominated by temporal fluctuations, would correspond to a "high temperature phase" while the high activity (partially frozen state), instead, would be analogous to a "low temperature phase". What makes balanced networks model peculiar with respect to traditional spin systems, is that temporal stochasticity in neuron flipping is self-generated by the network dynamic itself and not by an external source.

It can be argued that the irregular activity of the model is due to the stochasticity of the update times of the model neurons, but it can be demonstrated [25] that a model with a complete deterministic update rule lead to exactly the same mean fields equations of the model with stochastic updating rule.

1.8 Time-delayed Autocorrelations

We have already seen that the temporal fluctuations of the input, $\eta_i(t)$ in (1.49), obey gaussian statistics with variance given by $\sqrt{\sigma^2 - s^2}$, to fully characterize the input statistics we needs to determine the input autocorrelation:

$$R_A(\tau) \;=\; \frac{1}{N}\sum_{i=1}^{N}\Big\langle \Big[h_i^A(t) - h_A\Big]\,\Big[h_i^A(t+\tau) - h_A\Big]\Big\rangle$$

It can be shown that:

$$R_A(\tau) \;=\; J_{AE}^2\, Q(\tau) + J_{AI}^2\, Q(\tau) \tag{1.68}$$

Where $Q(\tau)$ is the time-delayed autocorrelation of local activities

$$Q_A(\tau) = \frac{1}{N}\sum_{i=1}^{N}\langle S_i^A(t) S_i^A(t+\tau)\rangle \tag{1.69}$$

Note that $Q_A(0) = m_A$ whereas $Q_A(\tau \to \infty) = q_A$ and, consistently, $R_A(0) = \sigma_A^2$ and $R_A(\tau \to \infty) = s_A^2$.

With arguments analogous to those that were used to derive the MF equations it can be shown [25] that the following self-consistent equation for $Q_A(\tau)$ (with $\tau > 0$) is obtained:

$$\tau_A \frac{dQ_A(\tau)}{d\tau} \;=\; -Q_A(\tau) \;+\; \int_0^\infty \frac{\mathrm{d}t}{\tau_A}\, e^{-\frac{t}{\tau_A}} \int \left[H\left(\frac{\theta_A - h_A + x\sqrt{R_A(t+\tau)}}{\sqrt{\sigma_A^2 - R_A(t+\tau)}}\right)\right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}}\mathrm{d}x$$

where the integral over t results from averaging over the distribution of update time intervals. The solution of this integral equation yields a function $Q_A(\tau)$ which decays to its equilibrium value q_A with a time constant of the order of the mean update time τ_E.

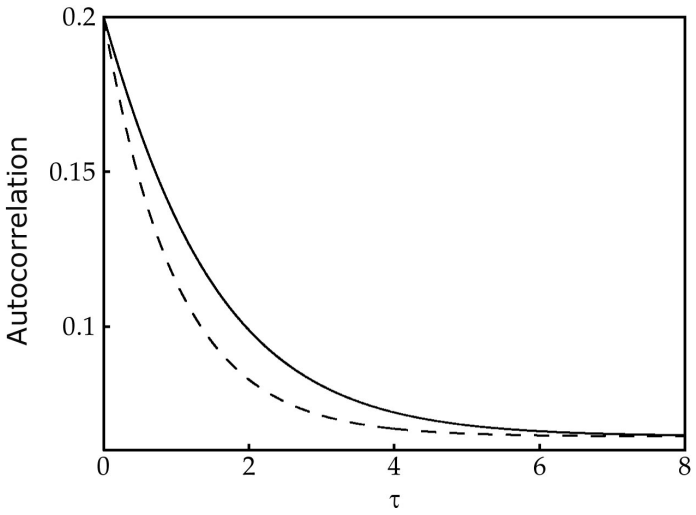

Figure 1.11: (Taken from [25]) *Solid line*: Population-averaged autocorrelation for the excitatory population in the large C limit. *Dashed line*: autocorrelation for a population of cells with the same rate distribution but Poisson updating.

A solution of the equation for $Q_A(\tau)$ is shown for example in figure 1.11, where for comparison it is shown the autocorrelation for a population with the same rate distribution but with Poisson updating (at each time step neuron i is set to the active state with probability m_i). This enhancement of short-time correlation is the manifestation of the refractoriness of the activities of the input cells, i.e. the input to a neuron between two subsequent time-step can change at most by the value of a single presynaptic neuron contribution.

1.9 Balanced Network numerical simulations

Simulations of a model network have been realized with two populations of $N = 30'000$ randomly connected noiseless binary neurons with mean number of connections per neurons $C = 1000$ and this set of parameters:

$J_{EE} = 1$	$J_{EI} = 2$	$J_{Eex} = 2.5$	$\theta_E = 1$	$\tau_E = 1$
$J_{IE} = 1$	$J_{II} = 1.8$	$J_{Iex} = 2.15$	$\theta_I = 1$	$\tau_I = 0.5$

(1.70)

Figure 1.12 (bottom) shows the input received by a generic neuron during a simulation; the excitatory (red) and inhibitory (blue) part of the total input (black) makes visible the balancing mechanism: the positive and negative part of the input are individually very large compared to the distance between resting and threshold but their algebraic sum yields a net input whose average is slightly subthreshold and whose fluctuations are of the same order of magnitude of the threshold.
It should be recalled that a change of state can take place only when the neuron happens to be selected in the random updating sequence and not whenever the net input crosses the threshold.

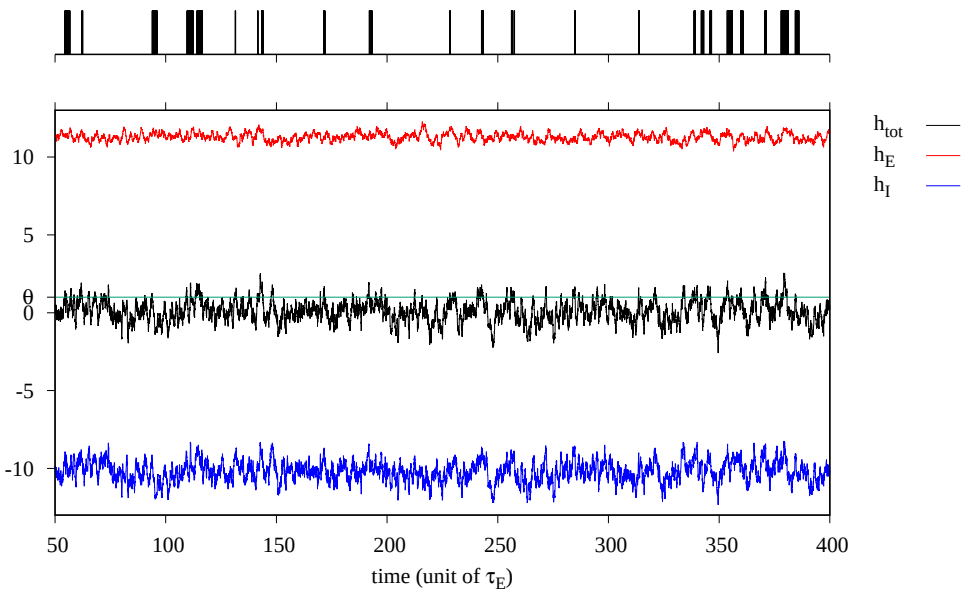

Figure 1.12: Example of a neuron activity (top) and input (bottom) in a simulation. Total excitatory inhibitory input are respectively the red and blue trace, net input is the black trace. The horizontal line is the threshold

Convergence to the balanced stationary state is quite rapid: roughly a pair of network update cycles are sufficient to settle into equilibrium (fig. 1.14).

Figure 1.13 show the stationary values of the order parameters of the network m and q as a function of the external input m_0. The dashed lines represent the solutions for finite $C = 10^3$, 10^4 while the continuous lines are the asymptotic solution in the $C \to \infty$ limit: at low input $m_0 \lesssim 0.2$ the solutions for $C = 10^4$ are quite well approximated by the $C \to \infty$ limit. Recalling that the central hypothesis at the base of the MF equations derivation is that $1 \ll C \ll N$, simulations shows that a network with $C = 10^3$ and $N = 3 \cdot 10^4$ appears to be already well within this condition as can be seen from the remarkably good agreement between simulations (crosses) and MF solutions.

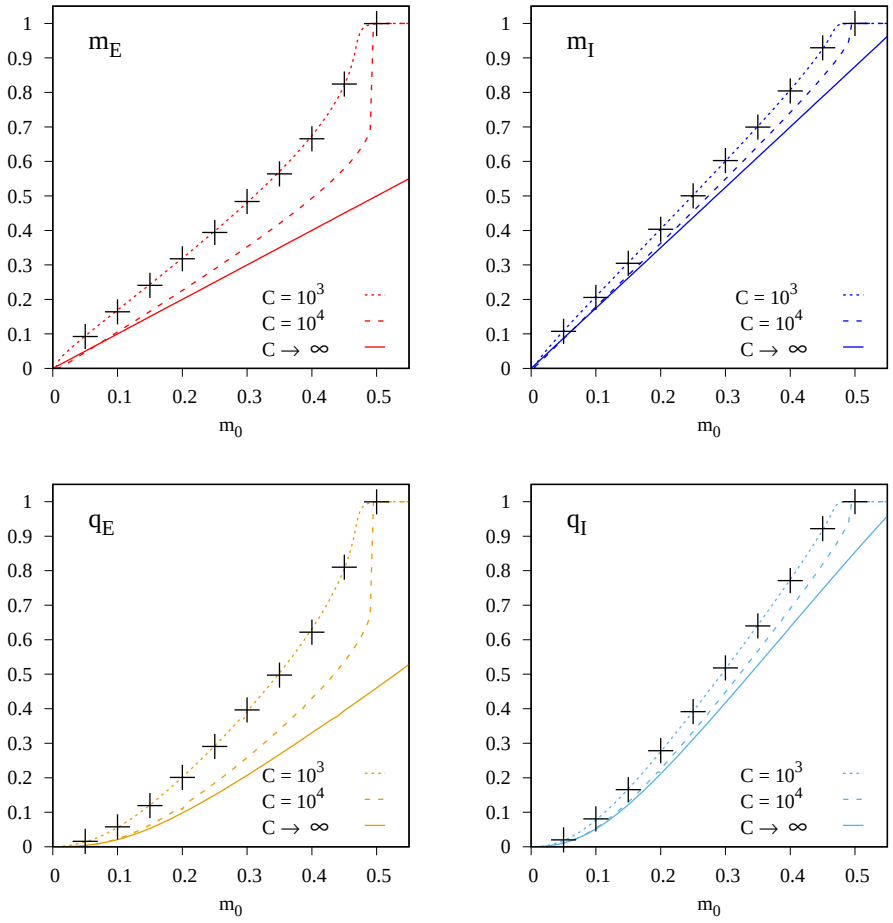

Figure 1.13: Equilibrium values of the network order parameters m, q as a function of the input m_0 for $C = 10^3$, $C = 10^4$ and in the $C \to \infty$ limit obtained from the solution of (1.61) and (1.64) with parameters (1.70). Crosses: simulation of a network with $N = 10^4$.

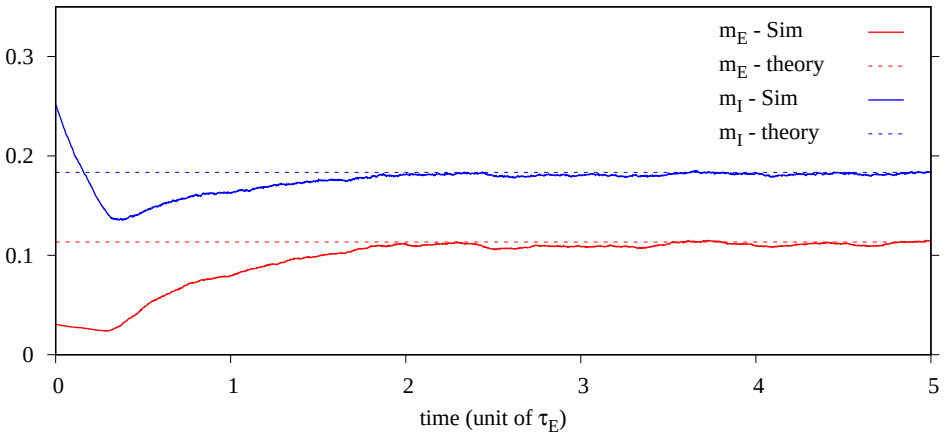

Figure 1.14: Time course of the population averaged activity $m_A(t) = \sum_i^N S_i^A(t)$ during a simulation initialized with $m_E(0) = 0.03$ and $m_I(0) = 0.25$

Theoretical predictions concerning spatial variability of the local average activity are also very well matched by the simulations: fig. 1.15 shows the distribution of the excitatory population activities $\rho_(m)$ measured from a simulation (histogram) at low input[6], superimposed to the theoretical distribution calculated from (1.66). The plot on log scale in the inset is given for comparison with the experimentally measured distribution in fig. 1.5 right.

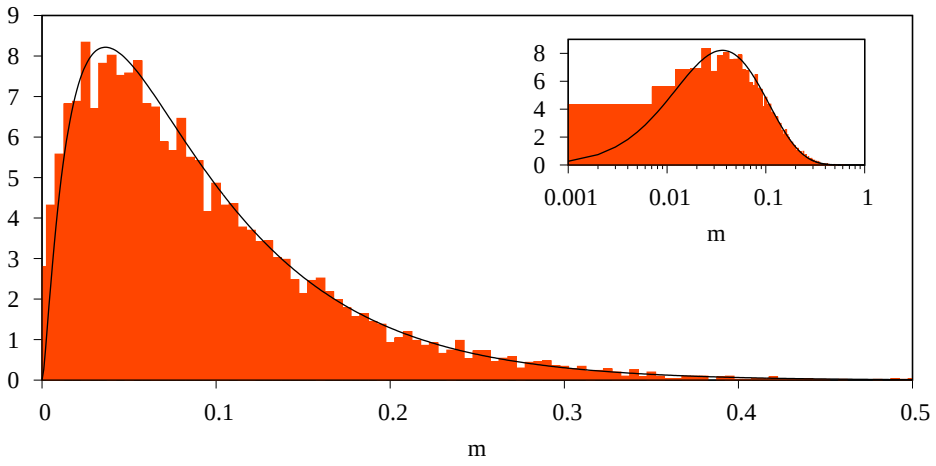

Figure 1.15: Example of distribution of local average activity in the excitatory population. Histogram: simulation with parameters (1.70) and $m_0 = 0.05$. Black curve: theoretical distribution. Inset graph: same plot with log scale on the x axis.

[6]The shape here displayed is not general but typical for low input, at higher input the distribution accounts for the existence of silent neurons. See section 1.7 for details

Chapter 2

Memory function and multi-stable neural networks

Ogni ramo della scienza sembra ci voglia dimostrare che il mondo si regge su entità sottilissime: come i messaggi del DNA, gli impulsi dei neuroni, i quarks, i neutrini vaganti nello spazio dall'inizio dei tempi...

Italo Calvino
Lezioni Americane, 1985

In this chapter we review the experimental evidences from which the theories on the neural mechanism underlying memory have been formulated. We then illustrate the characteristics of attractor neural networks, a successful and widely accepted paradigm to explain the observed phenomenology, and show that, when trying to apply the same principles in the context of the balanced networks, the model produces a behavior which is inconsistent with experimental observation. Last section introduce and motivate the solution proposed in the present thesis to resolve the inconsistencies of these class of models.

2.1 Persistent activity and memory

Many experiments [95, 41, 96] have shown that when a certain sensory stimulus (image, sound, odor...) is presented to an animal, some neurons maintain their firing activity practically unaltered (figure 2.1A). Other neurons, instead, respond with a marked enhancement of the activity: soon after the stimulus onset, the spiking rate rises up to some tens of spikes/sec (figure 2.1B,C see vertical gray band) however remaining far below the theoretical maximum firing frequency sustainable by a neuron of ~ 500 spikes/sec.

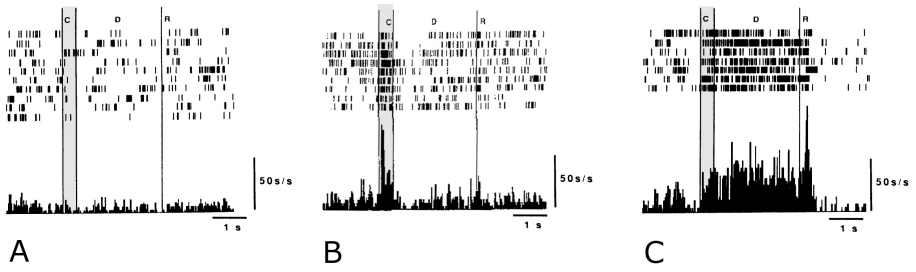

Figure 2.1: Different type of response to visual stimulation in neurons of the monkey prefrontal cortex (adapted from [3]). The upper part of each panel shows the spike rasters of repeated trials of the same stimulation protocol, in the lower part the trial-averaged spiking frequency. The vertical gray band indicates the period in which the visual stimulus is presented to the monkey. *Panel A*: spontaneous, pre-stimulus activity remains practically unchanged during and after the stimulation. *Panel B*: spiking activity is enhanced during stimulus presentation only. *Panel C*: spiking rate is increased during stimulus presentation and remains high after stimulus removal until the execution of the behavioral task associated with the stimulus.

When the stimulus is turned off, either the activity drops back to the pre-stimulus spontaneous regime (figure 2.1B) or, for certain specific stimuli, the enhanced spiking activity is maintained (figure 2.1C), a phenomenon called *selective persistent activity* [1, 2, 3, 97], meaning that the enhanced activity is maintained for certain stimuli and not for others.

When multiple trials are conducted with the same stimulus, a certain neuron responds in the same qualitative manner displaying selectivity or not. Nevertheless, the spiking patterns during stimulation and persistent activity show a high degree of temporal variability with Poisson-like distributed inter-spike intervals [98, 99] and considerable trial-to-trial variability in the spiking pattern is observed[1] (see spike rasters in fig. 2.1 and fig.2.2). This supports the idea that accurate timing of spike patterns is not essential to the brain for the representation of information, rather, it seems that stimulus-related information can be reliably represented in the short-time average spiking rate.

Wether the representation of information corresponds to the activity of single, individually meaningful cells (grandmother cells), or it is the global activity pattern across a whole cell population, has been largely debated. Experimental evidences point to a compromise between these extreme; in fact, the typical stimulus-evoked response within a local patch of cortex appears to be *sparse*: a small, stimulus-specific, subset of neurons becomes highly active while the vast majority of the neurons remains spontaneously active at very low rates [48, 101, 64].

The experimental results accumulated since the seventies suggest that selective persistent activity of sparse populations of neurons can be inter-

[1] It is observed that the variance of spike count can be between 1 and 2 times the mean count [100, 40].

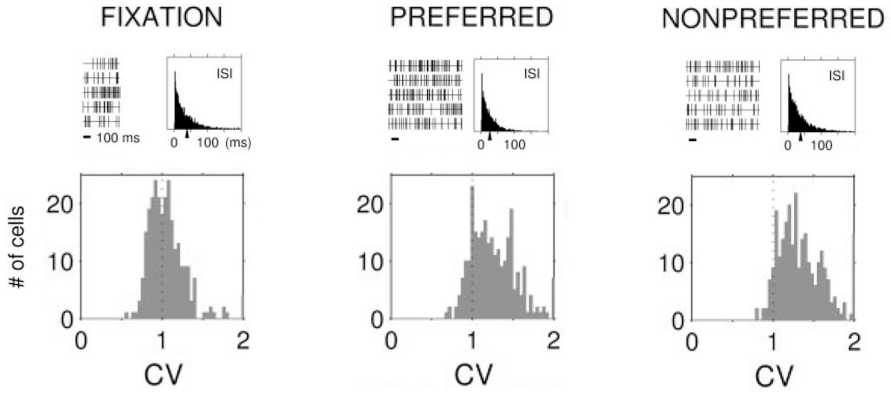

Figure 2.2: Temporal irregularity in working memory task: fixation epoch (last 500 ms before cue onset), delay period after preferred cues, delay period after presentation of non-preferred cue. *Top*: sample spike trains and inter-spike interval (ISI) histograms (arrowheads on x-axis indicate mean value). *Bottom*: measure of ISI variability in the neuron's population during the different experiment phases. Database of 229 neurons recorded from monkey's prefrontal cortex. Adapted from [98]

preted as the neural correlate of working memory [5, 6, 7, 8], i.e. the ability of the brain to form and retain an internal representation of an item of information even when this information is not available to the senses.

Modeling memory function: Attractor Neural Networks

A lot of theoretical work has been carried out in the last years in trying to understand how the ability to actively hold in mind previously learned information, i.e. memory, is implemented in neuronal networks [9, 23].
Since the elevated neural activity can be triggered by a brief input but outlasts it for many seconds, persistent activity cannot be explained by a feedforward mechanism, instead, it has been hypothesized that persistent activity can be self-sustained by a feedback mechanism within a local network with strengthened mutual connections [6, 8], a widely assumed mechanism by which memory traces are encoded and stored in the brain [11, 12] dating back to Donald Hebb's theory on activity dependent synaptic plasticity and cell-assembly formation [10]:

> *Let us assume then that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. [...] When an axon of cell A is near enough to excite a cell B and, repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.*

These concepts have been formalized into precise mathematical models inspired by spin-glass physics in some seminal papers written between the seventies and the eighties by Little [14], Hopfield [82], Amit et al. [16], laying the foundation of one of the currently standard paradigm to explain working memory: *Attractor Neural Networks* [87, 102, 103].
In these models, the interactions between neurons are symmetric and the properties of the system can be studied in terms of free-energy; each stored memory is represented by a specific configuration of the neural network and a fictitious "temperature" parameter is assumed to represent synthetically every source of stochastic disturbance and fluctuation. The dynamics is a relaxation towards minima of the free-energy, the retrieval of a stored memory occurs when the system falls in the corresponding basin of attraction; statistical-mechanics inspired calculation makes possible the estimation of global quantities such as memory storage capacity and stability in function of the noise level [87].

Although spin-glass neural networks model can be an advantage for framing the main principles in an analytically manageable form, their simplicity is the price to pay for a series of aspects that makes these kind of models quite unrealistic: full connectivity between neurons, symmetric interactions, high rates activity, the nonexistence of a stable background activity state, no distinction between excitatory and inhibitory cells; in fact soon after the first studies a lot of work have been dedicated to include in these models features that could bring them closer to the biological reality [88, 104, 105].

Successively, attractor networks composed of more realistic single neuron spiking models have been analyzed with both theoretical and numerical tools [106, 17, 18, 107, 108], extending to a more biologically plausible setting the attractor paradigm.
These models are able to reproduce qualitatively the experimental observations with a multi-stability mechanism between a non-selective attractor state (low background activity) and selective attractor states in which subgroups of neurons are active at rates which are higher (but not much higher) than background rates. Bistability between a background and a memory state is made possible by the nonlinearity of the response function of the single neuron; in fact, the general form of the mean field equations for the activity of the selective subpopulation ν is

$$\nu \;=\; \phi(h + \beta\nu) \tag{2.1}$$

where $\phi(x)$ is the neuron response function to a noisy input, $\beta\nu$ represents the part of the recurrent input coming from strengthened synapses and h the rest of the input. Since $\phi(x)$ has a stereotypical sigmoidal shape[109], a graphical solution of MF equation (2.1) shows that, for suitable choice of the parameter β, there can be *two stable solutions* (fig. 2.3).

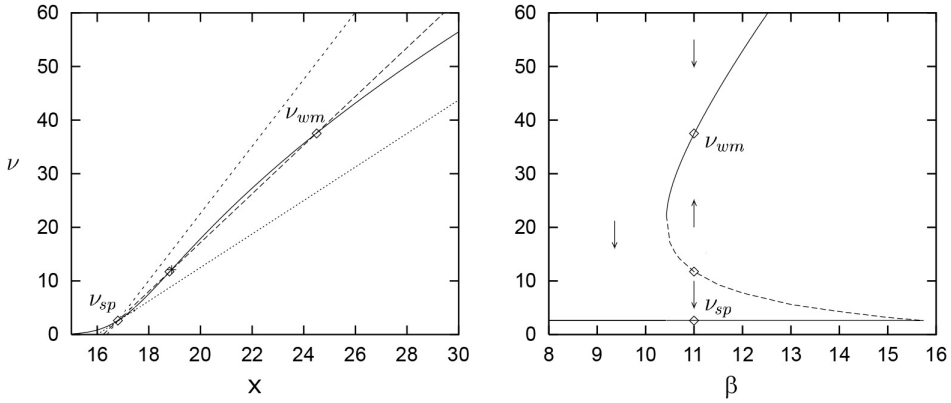

Figure 2.3: *Left*: with the variable change $x = h + \beta\nu$ in (2.1), the solutions of the equation are found graphically as intersection of the sigmoidal input/output single neuron function $\phi(x)$ and the line $\nu(x) = (x - h)/\beta$. *Right*: bifurcation diagram, bistability appears above a certain value of the synaptic reinforcement β. Adapted from [109]

A complete mean field description of attractor network can been derived for finite size network [17, 109, 18] but it turns out that the solutions can present, to various extent, two problematic aspects[110]:

- memory solution occurs at rates which are too high [17, 109, 18] compared to the low (and not much higher than background) rates seen in vivo [2, 111, 99].
- temporal irregular activity is obtained in the background state but persistent activity states typically exhibit much more regular firing [109], in contrast with the experiments which suggest that the degree of irregularity in persistent activity states is equal or greater than in the spontaneous activity state [112, 99].

Since irregular firing of cortical neurons is naturally explained if one assumes that networks operate in the balanced regime (net input and fluctuations of the same order of magnitude), some studies [19, 20, 21] considered to extend the balanced network approach (making the scaling explicit and taking the $C \to \infty$ limit) to the attractor framework. Nevertheless, as we will illustrate in the next section, it turns out that to realize bistability between two *balanced* states the problem of unrealistically high rate for the memory solution is unavoidable.

2.2 Multistability in the balanced regime

The balanced regime in an unstructured network model reviewed in chapter 1 provides a convincing explanation to the observed irregularity in the spontaneous activity, but it is incompatible with multistability since the single neuron nonlinearity becomes irrelevant in the $C \to \infty$ limit and rates depend linearly on the external inputs (1.39), i.e. there is a single solution for a fixed external input.
In the following we review the extension of the balanced model to the attractor framework given by [19] and illustrate the problem connected with the bistability in the context of balanced regime.

A simple memory model for synaptic connections

Persistent activity is supported if synaptic connections between sub-populations of neurons are reinforced in a Hebbian-like way. The population of selective neurons for a certain stimulus is called *foreground neurons*, the others *background neurons*. The *coding level* is the fraction of foreground neurons in a certain stimulus. A standard way to build a set of P *independent*, memory patterns with average coding level f in a network of N neurons, is to associate to each memory $\mu = 1...P$ a binary vector ξ^μ whose elements are equal to 1 for foreground neurons and 0 for background neurons:

$$\begin{matrix}\forall i = 1,...,N \\ \forall \mu = 1,...,P\end{matrix} \qquad \xi_i^\mu = \begin{cases} 1 & \text{with probability } f \\ 0 & \text{with probability } 1-f \end{cases}$$

There are several ways to store in the excitatory-to-excitatory connections this set of memory patterns, an extremely simple version of Hebb's rule was proposed by Willshaw[113, 114] in one of the earliest neural network model of associative memory:

> a synapse between two excitatory neurons is present if the two neuron are simultaneously active in at least one of the patterns, otherwise is zero, yielding the (symmetric) connectivity matrix

$$J_{ij}^{EE} = \Theta\left[\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\nu\right]$$

As in [19], we consider a simple modification of the Willshaw model in which the synaptic architecture has the following expression

$$J_{ij}^{EE} = c_{ij}^{EE}\left\{1 + (a-1)\Theta\left[\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\nu\right]\right\}\frac{J_{EE}}{\sqrt{C}} \qquad a > 1 \qquad (2.2)$$

with this rule, if two connected neurons i and j participate both to *at least* one pattern ($\sum_\mu \xi_i^\mu \xi_j^\mu > 0$), the synapse is strengthen $J_{ij} = aJ_{EE}$, while if ($\sum_\mu \xi_i^\mu \xi_j^\mu = 0$) the synaptic efficacy is unmodified J_{EE}.

Mean field equations

For each pattern ν it's straightforward to define the population average activity in the foreground and background population

$$m_+^\nu(t) \equiv \frac{\sum_{i=1}^{N} \xi_i^\nu m_i^E(t)}{\sum_{i=1}^{N} \xi_i^\nu} \qquad m_-^\nu(t) \equiv \frac{\sum_{i=1}^{N} (1-\xi_i^\nu) m_i^E(t)}{\sum_{i=1}^{N} (1-\xi_i^\nu)}$$

For each pattern, the memory system is described macroscopically by three order parameters m_+^ν, m_-^ν and m_I. Under the hypotheses of large network with high but diluted connectivity $1 \ll C \ll N$ the net input received by neurons has Gaussian statistics to a very good approximation, and we can write a system of self consistent MF equations once the mean and the variance of the input to each of the three population has been found.

The input received by the generic excitatory neuron is:

$$\begin{aligned} h_i^E(t) \; &= \frac{J_{EE}}{\sqrt{C}} \sum_{j=1}^{N} c_{ij}^{EE} \left\{ 1 + (a-1)\Theta \left[\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\nu \right] \right\} S_j^E(t) \\ &- \frac{J_{EI}}{\sqrt{C}} \sum_{j=1}^{N} c_{ij}^{EI} S_j^I(t) \; + \; m_0 E \sqrt{C} \end{aligned} \tag{2.3}$$

and we define the average input to foreground and background neurons in pattern ν, respectively, as

$$h_+^\nu(t) \equiv \frac{\sum_{i=1}^{N} \xi_i^\nu h_i^E(t)}{\sum_{i=1}^{N} \xi_i^\nu} \qquad h_-^\nu(t) \equiv \frac{\sum_{i=1}^{N} (1-\xi_i^\nu) h_i^E(t)}{\sum_{i=1}^{N} (1-\xi_i^\nu)}$$

The average of the structured part of the connections in (2.3) can be calculated heuristically with statistical considerations: for a certain pattern ν, a neuron in the background population receives on average fC contacts from foreground neurons and $(1-f)C$ contacts from background neurons; some of these connections can be stronger if are received from neurons which belongs, together with the considered neurons, to other patterns $\mu \neq \nu$, otherwise have baseline strength J_{EE}. Since patterns are random and independent, the probability that two generic cells are not in a particular pattern is $1-f^2$, thus the probability that the two cells are not both in any of the $P-1$ patterns $\mu \neq \nu$ is $(1-f^2)^{P-1}$. The complementary probability

$$\Pi = 1 - (1-f^2)^{P-1}$$

is then the probability that a background neurons receive a strong synapse aJ_{EE} from any other neuron in the network. With analogous consideration the average input to foreground, background and inhibitory neurons can be calculated[2] and we get:

$$h_- = \sqrt{C}\{J_{EE}(a\Pi + 1 - \Pi)[fm_+ + (1-f)m_-] - J_{EI}m_I + Em_0\} \quad (2.4a)$$

$$h_+ = \sqrt{C}\{J_{EE}[afm_+ + (a\Pi + 1 - \Pi)(1-f)m_-] - J_{EI}m_I + Em_0\} \quad (2.4b)$$

$$h_I = \sqrt{C}\{J_{IE}\,[fm_+ + (1-f)m_-] - J_{II}m_I + Im_0\} \quad (2.4c)$$

The computation of the input variance gives:

$$\sigma_-^2 \;=\; J_{EE}^2\,(\,a^2\Pi + 1 - \Pi\,)[\,fm_+ + (1-f)m_-\,] \;+ J_{EI}^2 m_I \quad (2.5a)$$

$$\sigma_+^2 \;=\; J_{EE}^2\,[\,a^2 fm_+ + (a^2\Pi + 1 - \Pi)(1-f)m_-\,] \;+ J_{EI}^2 m_I \quad (2.5b)$$

$$\sigma_I^2 \;=\; J_{IE}^2\,[\,fm_+ + (1-f)m_-\,] + J_{II}^2 m_I \quad (2.5c)$$

The variances are all independent from C and thus of order 1, while average input are proportional to $\sqrt{C}$. To have a solution in which all three populations are balanced (average input of order 1 like the fluctuations) in the $C \to \infty$ limit, all the three terms between curly brackets in 2.4 should vanish. But if we take the difference between the average input in foreground (2.4b) and background (2.4a) we have that

$$h_+ \;-\; h_- \;=\; \sqrt{C}\; J_{EE}(a-1)(1-\Pi)fm_+ \quad (2.6)$$

is a quantity of order $\sqrt{C}$, which means that foreground and background populations cannot be *both* in a balanced state because if one of the two excitatory subpopulation is in the balanced state its input will be $O(1)$ and automatically the input to the other population will be $O(\sqrt{C})$ thus unbalanced, bringing the neuron in the population to saturation or quiescence, depending on the sign.

Simultaneous balance of both foreground and background neurons can be achieved if (2.6) remains of order 1 in the $C \to \infty$ limit, which means that the term $J_{EE}(a-1)(1-\Pi)fm_+$ has to be $O(1/\sqrt{C})$. In particular, we can consider the case in which one among synaptic potentiation $(a-1)$, memories non-overlap probability $(1-\Pi)$ or coding level f scale like $1/\sqrt{C}$, while the other quantities are $O(1)$. Again, as for the choice of scaling synaptic strength $J \sim 1/\sqrt{C}$, the rationale of assuming a certain scaling with C for a parameter of the system is a way to make a certain assumption about a physiological characteristic of the system and then derive its consequences straightforwardly when taking the $C \to \infty$ limit.

To have that the difference between the mean input to foreground and background neurons dose not become too large in a highly connected network, one of the following conditions must hold:

[2]Since all patterns are equivalent, from now on we drop the superscript ν

- *Weak potentiation.* $(a-1) \sim \frac{1}{\sqrt{C}}$ means that the hebbian-like structuring of the excitatory-excitatory connections produce an increase in synaptic strength which is small with respect to the baseline: $J_{EE}^{+} = J_{EE} + J_{EE}(a-1)/\sqrt{C}$.
- *High storage.* $(1-\Pi) = (1-f^2)^{P-1} \sim \frac{1}{\sqrt{C}}$ given a constant coding level $f < 1$, this mean that the number of stored memories P should be large. This condition depends essentially from the Williwaw model, where only two synaptic strength are possible, it's essentially a way to make h^+ and h^- similar because for large P a large fraction of the input to a neurons in both sub-populations will be potentiated.
- *Sparse coding.* $f \sim \frac{1}{\sqrt{C}}$ which mean that each the sub-populations of neurons which are selective for a certain stimulus are made up of a small number of units compared to the network size.

The solution adopted by [21] is weak potentiation and finite coding level, while in [19] potentiated synapses are of the same order of magnitude of the baseline ones and sparse coding is assumed.
In what follows, to illustrate the general principle of bistability in the balanced regime we will make reference to the latter study and consider sparse coding. In fact, several experimental evidences [115, 62, 48, 101, 64] show that typical stimulus-evoked response within a local patch of cortex trigger elevated activity in small number of neurons while the vast majority of the other cells in the assembly remains spontaneously active at very low rates. Although we are far from making conclusive statement about the representation of information in cortical networks, sparse coding is considered a reasonable physiological hypotheses and has been included in several models of attractor neural networks [114, 108, 19, 109].

Bistability and balanced regime in the sparse coding limit

Assuming sparse coding, we define

$$f = \frac{F}{\sqrt{C}}$$

with F a constant parameter. If this is the case, if a finite number of patterns P is stored, the probability Π becomes vanishingly small for $C \to \infty$, therefore we consider the classical scaling of the number of stored patterns $P = \alpha C$ where α is termed *load parameter* so that $\lim_{C\to\infty} \Pi = 1 - e^{-\alpha F^2}$.

With these definition, for the average inputs (2.4) we obtain:

$$h_- = \sqrt{C}\{J^{\star}_{EE} m_- - J_{EI} m_I + E m_0\} + J^{\star}_{EE} F(m_+ - m_-) \tag{2.7a}$$

$$h_I = \sqrt{C}\{J_{IE} m_- - J_{II} m_I + I m_0\} + J_{IE} F(m_+ - m_-) \tag{2.7b}$$

$$h_+ = \; h_- \; + \; (a-1)(1-\Pi) J_{EE} F m_+ \tag{2.7c}$$

where $J^{\star}_{EE} = J_{EE}[a\Pi + 1 - \Pi]$ is the average excitatory-excitatory synaptic strength. Thus, if the input coming from the background and the inhibitory neurons are balanced, in the mean input of all three population will be of order 1 as the variances, which in the sparse coding limit becomes:

$$\sigma^2_- = \sigma^2_+ \equiv \sigma^2_E = J^2_{EE}\left(a^2\Pi + 1 - \Pi\right) m_- + J^2_{EI} m_I \tag{2.8a}$$

$$\sigma^2_I = J^2_{IE}\, m_- + J^2_{II} m_I \tag{2.8b}$$

We notice that the input variance is equal in both excitatory subpopulations. The balance condition is completely analogous to the case of he unstructured network (section 1.5): provided that

$$\frac{E}{I} > \frac{J_{EI}}{J_{II}} > \frac{J^{\star}_{EE}}{J_{IE}} \qquad \text{and} \qquad J_{EI} > J^{\star}_{EE}$$

the average rates m_- and m_I are given by the balancing condition:

$$\begin{cases} J^{\star}_{EE} m_- - J_{EI} m_I + E m_0 = 0 \\ J_{IE} m_- - J_{II} m_I + I m_0 = 0 \end{cases} \quad \Longrightarrow \quad \begin{aligned} m_- &= \tfrac{EJ_{II} - IJ_{EI}}{J_{EI}J_{IE} - J^{\star}_{EE}J_{II}}\, m_0 \\ m_I &= \tfrac{EJ_{IE} - IJ^{\star}_{EE}}{J_{EI}J_{IE} - J^{\star}_{EE}J_{II}}\, m_0 \end{aligned}$$

The intuitive picture is the following: if the relative size f of the selective populations is taken to be small compared to the rest of the network, the global activity of the network is set by the balance condition between the inhibitory population and the large background population, the activity in the small selective excitatory sub-population becomes essentially uncoupled from the rest of the network. Since the corrections to the leading order of the input are different between the selective and non-selective populations, the activity foreground neurons m_+ is a solution of the mean field equation

$$m_+ = H\left(\frac{\theta - h_- - \beta m_+}{\sigma_E}\right) \tag{2.9}$$

where $\beta = J_{EE}(a-1)Fe^{-\alpha F^2}$. The solution of m_- and m_I derived from the global balance condition allow to calculate straightforwardly σ_E (2.8a) and h_- by inverting the mean field equation for the background population

$$m_- = H\left(\frac{\theta - h_-}{\sigma_E}\right) \qquad \Rightarrow \qquad h_- = \theta - \sigma_E H^{-1}(m_-)$$

Equation (2.9) can be solved graphically after defining the (invertible) function $x(m_+) = \frac{\beta}{\sigma_E} m_+ \Rightarrow m_+(x) = x\frac{\sigma_E}{\beta}$ so that the solution for m_+ corresponds to the ordinate of the intersection between the straight line and the sigmoidal neuron transfer function:

$$x\frac{\sigma_E}{\beta} = H\Big(H^{-1}(m_-) - x\Big) \tag{2.10}$$

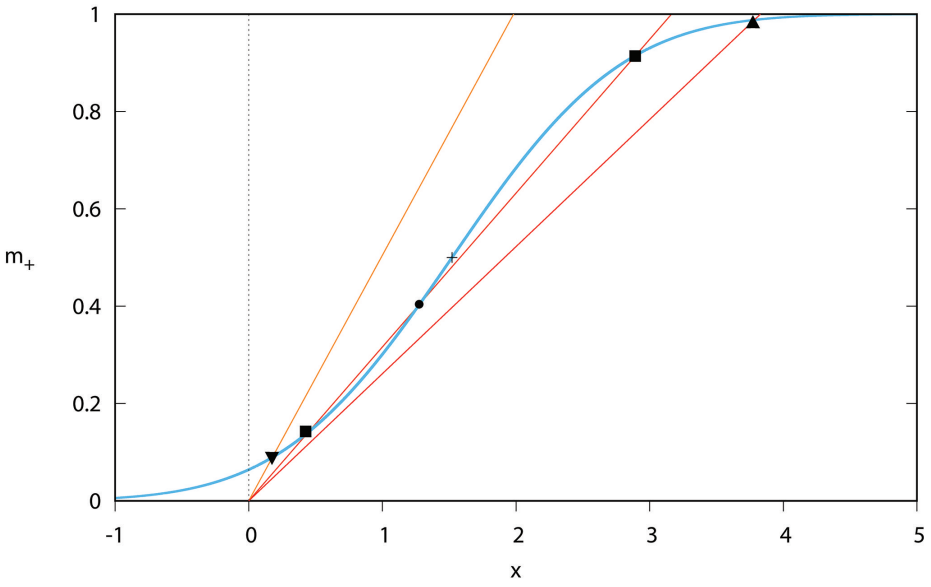

Figure 2.4: Graphical solutions of (2.10): depending on the parameters the curves can have from one (triangles) to three intersection, of which two are stable states (squares) and one unstable (circle). Connectivity parameters are take from (1.70) and $m_0 = 0.06$, $\alpha = 0.003$, $F = 1.5$, the lines, from left to right corresponds to $a = 1.65, 2.05, 2.28$. The small cross is the inflexion point of the sigmoid curve. We remark that the position of the sigmoid curve is not fixed, in this case the value of $H^{-1}(m_-)$ has very small variation for the three values of a chosen.

From figure 2.4 we can see that if the synaptic reinforcement in the selective subpopulation is not strong enough to generate enough feedback to keep elevated activity, the selective subpopulation is active at low rate at about the same rate of the nonselective population; on the other hand, if excitatory feedback is too strong within the selective population the foreground neuron will be spontaneously active near saturation. For suitable choice of parameters, we can have three point of intersections between the curves, where the extreme ones represents dynamically stable solutions (black squares) and the central one (black circle) an unstable solution. The problem is that the higher rate solution occur at an extremely unrealistic, nearly saturated activity (see also simulation in fig.2.5).

2.3 Why attractor models have problems in reproducing realistic activity of memory states?

The problem of unrealistically high-rate solution is not restricted to the particular model we have considered but it's quite general. In fact, high rates memory states are found also with more realistic spiking neuron model both in finite size solutions [109] and in the balanced regime [20, 21].

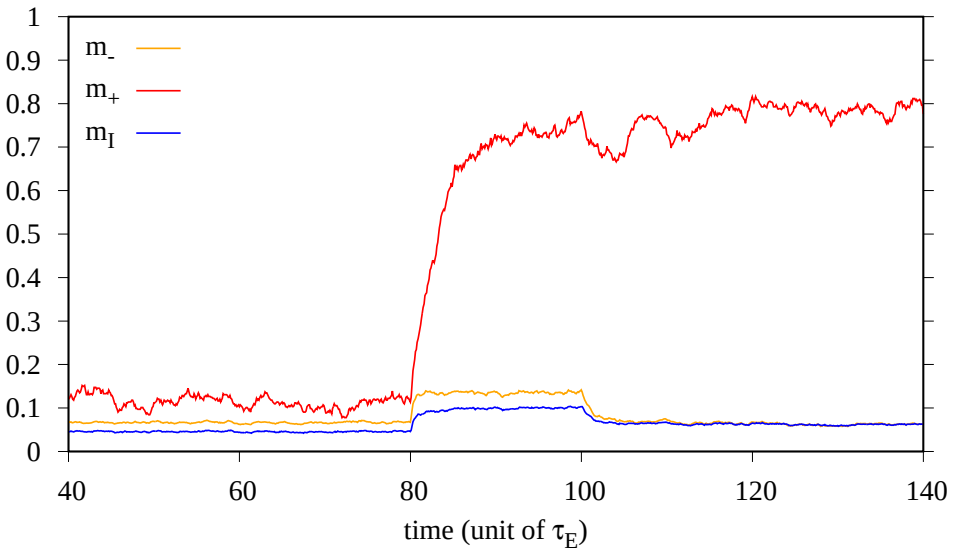

Figure 2.5: Simulation of a memory network model with $N_E = N_I = 10^4$ and $C = 1000$ storing $P = 3$ patterns with coding level $f = 0.05$. Connection parameters are from (1.70), $m0 = 0.01$ and synaptic reinforcement $a = 2.4$. A stimulus is delivered between $t = 80$ and 100 to the selective neurons of pattern 1 in the form of higher external input. After the end of the stimulation, the selective population keeps the activity elevated, although at a very high rate.

The reason of such an unsettling outcome of these models depends essentially on the interaction between two essential features:

- the input to a neuron is a *linear* function of the presynaptic activity
- regardless of the particular neuron model adopted, the single neuron input/output function has a stereotypical sigmoidal shape which derive from the response to a fluctuating input.

In fact, as the graphical analysis has shown, since the stable working memory solution corresponds to the intersection between a line and the concave part of the sigmoidal function, this results in a geometrical constrain which lead to an absolute lower bound on the solution: no-matter the parameter choice, the upper intersection occur always above the inflexion point of the neuron response curve; thus $m > 0.5$ for the binary neuron response function $H(x)$, for spiking neuron models this bound can be somewhat lower but still enough to yield memory activity rates consistently above spontaneous activity [109, 20, 21], contrary to experimental evidences, and in any case, to keep neurons from firing near saturation, the parameters must be chosen in a narrow range thus implying significant fine tuning and poor robustness. For instance, in the model studied in [21], weak potentiation (memory part of the connection scaling as $J \sim 1/C$) was considered in the balanced regime,

to get the lowest possible activity of the memory solution a fine tuning of coding level f is necessary, in particular it implied that f should be large. Another unpleasant feature of this model is that it produces a highly bimodal distribution of firing rates with the background neurons firing at a so much lower rate than the foreground that they form a distinct, and easily recognizable, population.

A direct consequence of the extremely high rate solution is that the activity in the memory state is not as temporally irregular as the spontaneous activity state[109, 21], contrary to the data which shows similar if not higher degree of irregularity.
The reason for this is that from spontaneous to memory solution the mean input to the foreground neurons increase, but the fluctuations σ_E remain unchanged (because is set by the activity of the background and inhibitory neurons); thus while the activity in the spontaneous state has subthreshold input and is fluctuation driven, in the memory state the mean input is fairly above threshold and relatively larger than fluctuations thus the activity is mean driven and much less temporally irregular. In fact, while balance is a necessary condition for irregular firing, it is not sufficient. That's because balance ensures only that the mean and fluctuations are independent of C, but does not rule out the possibility that the mean is much larger than the fluctuations, which would result in regular firing.
A model in which between spontaneous and memory states the mean does not change, but the variance does, was considered in [20] where both states are fluctuation driven, producing high variable activity.
Nevertheless, the memory solution occurs at unrealistically high rate also in this model and, even more problematically, multistability vanishes in the large C limit because the difference in the fluctuations between spontaneous and persistent activity vanishes, thus the range of multistability in the network with balanced persistent state is extremely small for realistic numbers of inputs per cell.

Short-term synaptic plasticity as a solution to unrealistically high-rate memory solution

The realization of a neural network model where multistability occur *robustly* between several states in which, as the experiments shows, neurons activity is *low* and displays *spatial and temporal irregularity* has proven difficult [110]. Low-rate activity and temporal irregularity are connected: if a model could generate a *balanced* memory solutions at low activity (slightly higher than spontaneous), the mean input would lie in the subthreshold part of the neuron transfer function yielding temporal irregularity automatically since the balanced regime ensures that input fluctuations and distance to threshold are of the same order and the activity is fluctuation driven.

In the first place, the object of study for this thesis has been an investigation of the possibility to achieve bistability not between a pair of balanced states, but between a low activity background state in the *balanced* regime and a slightly higher activity memory which is instead *partially balanced*, i.e. the foreground population is balanced at very low rate and the background population has a negatively unbalanced input. In this way both states would produce subthreshold inputs in both populations, yielding extremely low activity as some studies suggest [50]. This scenario turned out to be unfeasible because the condition for stability of the memory state automatically implied instability of the background state. The complete details of this investigations are given in appendix B.
Successively, we moved on to investigate whether we could obtain stable solution in the lower -convex- branch of the neuron transfer function by introducing a *nonlinear* dependence of the input on the presynaptic activity.

We assumed that such nonlinearity originates form *synaptic short-term plasticity*, i.e. activity-dependent modifications of the synaptic couplings[116, 117], drawing on numerous theoretical studies [118, 119, 120, 121, 122, 123, 124] which showed the considerable effect of synaptic short-term plasticity on the collective activity of the network. In particular, we considered a phenomenological model of synaptic *short-time depression* (STD) [125], a mechanism by which a synapse gradually decrease its strength when active, and recover to baseline value when inactive .

The originality of the present work consist in that STD effects have been examined within the balanced network framework for the first time. The advantage of balanced network is that fluctuation are generated autonomously by the network dynamics and can be calculated self-consistently; moreover, the high temporal variability of the activity seen in-vivo in both spontaneous and memory state would arise automatically as an effect of the balanced regime.

The result of the study is presented in two steps, corresponding to the subsequent two chapters: first, in chapter 3, the STD model is introduced and its effects analyzed in an unstructured random network through the derivation of a novel mean field theory and comparison with simulations, then STD dynamics is applied to the modified Willshaw model (presented at the beginning of the current chapter) and its effect on the bistability are examined.

Chapter 3

Balanced networks with depressing synaptic dynamics

Scientists are always assuming or hoping that things are simple,
and then discovering that they are not.
Gregory Bateson
Steps to an ecology of mind, 1972

In this chapter we introduce a model for synaptic dynamics in which the synaptic strength depends on the pre-synaptic neuron activity. Subsequently, we consider a model network where the synapses between excitatory neurons are endowed with this dynamics, thus rendering the input to any neuron a nonlinear function of the network activity. The corresponding mean field(MF) equations are derived and it will be shown that a closed form solution for this MF equations can be obtained under a simplifying assumption whose validity will be checked a-posteriori.

3.1 A model for synaptic dynamics

Synaptic transmission is a dynamic process: postsynaptic responses changes continuously according to the the presynaptic activity [24]. This prominent characteristic of chemical synaptic transmission is a crucial determinant of the response properties of synapses and, in turn, of the patterns of activity generated by neural networks.
In particular, short-time depression (STD) is an activity-dependent reduction in the efficacy of neuronal synapses. This phenomenon occurs because the synaptic transmission is mediated by chemical molecules, neurotransmitters, which are released when the synapses is activated by an action potential

from the presynaptic neuron, and are then re-absorbed and regenerated on a time scale of the corresponding metabolic process. Since a synapse has a finite amount of neurotransmitter available and the strength of the synaptic current that is generated depends on the amount of molecules released, if presynaptic stimulation occurs at a higher rate with respect to the recovery, the quantity of available neurotransmitters gradually decrease and the synaptic efficacy is reduced accordingly [126].

Inspired by the phenomenological model proposed in [125] for the dynamics of synapses in the context of spiking neuron models, we propose the following model for depressing synapse in a binary neuron model:

> The synaptic strength between two neurons is modulated by a continuous dynamical variable $x \in [0, 1]$ which evolves in function of the activity state of the presynaptic neuron $S(t)$ according to
>
> $$\frac{dx(t)}{dt} = \frac{1 - x(t)}{\tau_r} - U\, x(t)\, S(t) \tag{3.1}$$

The equation (3.1) can be written explicitly in the form

$$\frac{dx(t)}{dt} = \begin{cases} f_0(x) = \dfrac{1 - x(t)}{\tau_r} & \text{when} \quad S = 0 \\ f_1(x) = \dfrac{x_l - x(t)}{\tau_d} & \text{when} \quad S = 1 \end{cases} \tag{3.2}$$

where we have defined the quantities

$$x_l = \frac{1}{1 + U\tau_r} \qquad \tau_d = \frac{\tau_r}{1 + U\tau_r} = \tau_r\, x_l \tag{3.3}$$

Depending on the value of $S(t)$, the dynamics of the x variable is driven either by $f_0(x)$ or by $f_1(x)$, and the general solution for $x(t)$ in the two cases is, respectively

$$x_0(t) = a_0\, e^{-t/\tau_r} + 1 \tag{3.4a}$$

$$x_1(t) = a_1\, e^{-t/\tau_d} + x_l \tag{3.4b}$$

where the two constant a_0 and a_1 are determined by the initial conditions.

The qualitative picture that emerges from this model is the the following: assuming that J is the baseline strength of a synapse, when the presynaptic neuron is active ($S = 1$) the synaptic resources undergo a process of depletion and the efficacy decays exponentially toward Jx_l with a characteristic time τ_d; if the presynaptic neuron neuron is not active ($S = 0$), the synaptic resources are recovered and the efficacy grows exponentially toward the maximum J with a timescale τ_r.
If the time course of $S(t)$ is given, equation (3.1) is deterministic and the time-course of $x(t)$ is uniquely determined by the initial condition $x(0)$.

3.2 MF theory for a Balanced Network with STD

Mathematical descriptions of networks with dynamic synapses have been given in [119] by means of simplified rate-models in the style of [127].

Here we will derive an extension of the mean field theory for a randomly connected balanced network of binary neurons [25] to include depressing dynamics 3.1 on the synapses among excitatory neurons; this, as a preliminary step to the study of a balanced memory network with synaptic dynamics.

The input to a generic neuron of the excitatory or inhibitory population are, respectively

$$\begin{aligned} h_i^E(t) &= \frac{J_{EE}}{\sqrt{C}} \sum_{j=1}^{N_E} c_{ij}^{EE} x_j(t) S_j^E(t) - \frac{J_{EI}}{\sqrt{C}} \sum_{j=1}^{N_I} c_{ij}^{EI} S_j^I(t) + m_0 E \sqrt{C} \\ h_i^I(t) &= \frac{J_{IE}}{\sqrt{C}} \sum_{j=1}^{N_E} c_{ij}^{IE} S_j^E(t) - \frac{J_{II}}{\sqrt{C}} \sum_{j=1}^{N_I} c_{ij}^{II} S_j^I(t) + m_0 I \sqrt{C} \end{aligned} \tag{3.5}$$

where the amplitude of the excitatory synapses x_j follow the dynamics:

$$\frac{dx_j(t)}{dt} = \frac{1 - x_j(t)}{\tau_r} - U\, x_j(t)\, S_j(t)$$

For this model, a network configuration is specified by the state of the $2N$ binary spin variables S_i^A and the state of the N excitatory synaptic amplitudes x_i. It is therefore necessary to derive an evolution equation for the joint probability distribution

$$P(S_1^E, ..., S_N^E, S_1^I, ..., S_N^I, x_1, ..., x_N, t)$$

In the limit of highly diluted network $N \gg C$ the neurons activities become uncorrelated [88, 25] and we can write

$$P(S_1^E, ..., S_N^E, S_1^I, ..., S_N^I, x_1, ..., x_N, t) = \prod_{i=1}^{N} P(S_i^E, x_i, t) \prod_{i=1}^{N} P(S_i^I, t)$$

In the high connection limit, the net input to a neuron has a gaussian statistics which is fully specified by its mean and variance.

Mean input

Taking the population average of the fields in the two populations as usual we found the expression for the time dependent mean fields

$$h_E(t) = \sqrt{C}\{J_{EE}\, r(t) - J_{EI}\, m_I(t) + m_0\, E\} \tag{3.6a}$$

$$h_I(t) = \sqrt{C}\{J_{IE}\, m_E(t) - J_{II}\, m_I(t) + m_0\, I\} \tag{3.6b}$$

In this case, besides the usual population activities, synaptic plasticity defines a new macroscopic order parameter which can be interpreted as the fraction of available resources averaged over the active neurons:

$$r(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) S_i^E(t) \tag{3.7}$$

.

Input variance

In the limit $N \ll C \ll 1$ the fluctuation of the input to a neuron around its average h_A can be written as the sum of two gaussian random variable: a white noise $\eta(t)$ representing the temporal fluctuations, and x_i representing the spatial fluctuations due to the variability in the number of connections received by various neurons

$$h_i^A(t) = h_A + \sqrt{\sigma_A^2 - s_A^2}\,\eta(t) + s_A\, x_i$$

The calculation of the total fluctuation for the input (3.5) gives

$$\sigma_E^2(t) = J_{EE}^2\, v(t) + J_{EI}^2\, m_I(t) \tag{3.8a}$$

$$\sigma_I^2(t) = J_{IE}^2\, m_E(t) + J_{II}^2\, m_I(t) \tag{3.8b}$$

where we have defined

$$v(t) = \frac{1}{N}\sum_{i=1}^{N} x_i^2(t) S_i^E(t) \tag{3.9}$$

For the quenched fluctuations we have

$$s_E^2 = J_{EE}^2 p + J_{EI}^2 q_I \tag{3.10a}$$

$$s_I^2 = J_{IE}^2 q_E + J_{II}^2 q_I \tag{3.10b}$$

where we have defined the order parameter

$$p(t) = \frac{1}{N}\sum_{i=1}^{N} \Big\langle\, x_i(t) S_i^E(t) \,\Big\rangle^2 \tag{3.11}$$

The calculation of the order parameters r, v and p is not straightforward since it involves the determination of the probability distribution of the compound process (x, S) which depends on the history of S, thus a self-consistent calculation of the autocorrelation of the activities must be taken into account, which is quite complex. A more direct way to calculate the averages is by neglecting the self-correlations and approximating the flipping of $S(t)$ as a Markov Process.

Markov approximation for single neuron activity

The updating of a neuron in the excitatory population is a Poisson process with average rate $1/\tau_E$ and when an updating occurs, the state of the neuron is set to 1 or 0 depending on wether the input is above or below the threshold θ_E; the average activity $\langle S(t)\rangle = m(t)$ correspond to the probability that the neuron is in the active state at time t.
When the network is at equilibrium m correspond to the temporal average, i.e. the fraction of time that the neuron spend in the active state, but this is not sufficient to characterize the temporal structure of $S(t)$. In fact, nonzero self-correlation in the activity is present because after an update, the sooner comes the next switching time, the lesser the input field is likely to have changed, and so the neuron output: the smaller the interval between two updates the smaller the probability that the neuron flips its states at the second update.

If self correlations were absent, at every update the neuron can switch to (or remain in) the active state $S = 1$ with probability m (and switch to the silent state $S = 0$ with probability $1 - m$) regardless of the state at the previous update. A stochastic process with this characteristics is named *Dichotomous Markov Noise* (DMN): a two valued stochastic process with constant transition rates between the two states, K_{10} and K_{01} thus exponentially distributed waiting times in the two states [128]. A DMN model for $S(t)$ would be completely determined by the transition rates between the active and the silent state:

$$K_{10} = \frac{1-m}{\tau_E} \qquad K_{01} = \frac{m}{\tau_E} \tag{3.12}$$

Figure 3.1 shows an example of the time course of x with $S(t)$ given by the realization of a DMN. Assuming that the activities of neurons can be described by a DNM, we are effectively discarding the effect of self-correlation and we can solve the MF in closed form because the statistic of the x will be determined solely by the transition rates (3.12). At least from a qualitative point of view, we don't expect this approximation to yields predictions that are dramatically different form the proper MF theory in which autocorrelation are calculated self consistently. This will be checked A-posteriori, comparing the approximated MF theory with simulation (section 3.4).

Evolution equation for the p.d.f. of x

Under the hypotheses that $S(t)$ is a DMN, we can write and solve the evolution equation for the probability distribution function of the stochastic variable x which (3.1). The compound stochastic process [x(t),s(t)] can be described by the joint distributions

$$P_0(x,t) \equiv P(x, s = 0, t) \qquad P_1(x,t) \equiv P(x, s = 1, t)$$

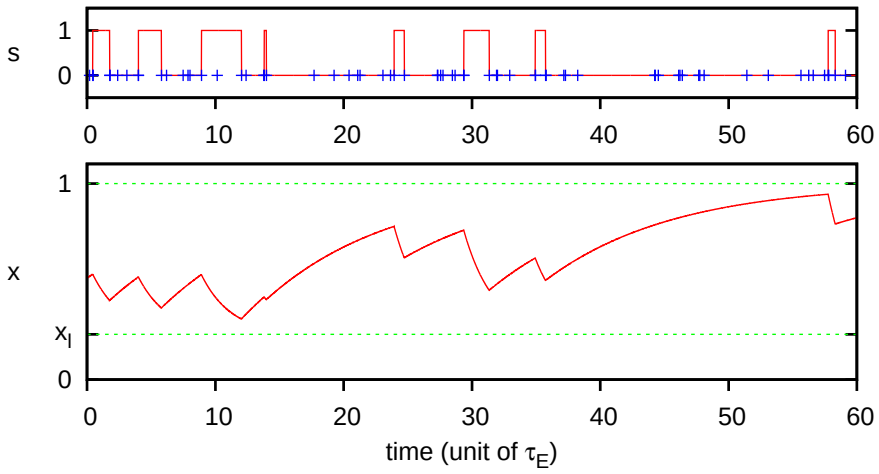

Figure 3.1: Time evolution of synaptic amplitude (lower panel) driven by a neuron with average activity $m = 0.15$ (upper panel) blue tics indicates the time of the neuron update. Time constants in unit of τ_E are: $\tau_r = 10$ and $U = 0.3$, yielding $\tau_d \simeq 2.3$ and $x_l \simeq 0.23$

and it can be shown that (see [129]), the temporal evolution of the joint distributions is governed by the master equation

$$\begin{cases} \partial_t P_0(x,t) = & -\partial_x[f_0(x)P_0(x,t)] - \Big(K_{01}P_0(x,t) - K_{10}P_1(x,t)\Big) \\ \partial_t P_1(x,t) = & -\partial_x[f_1(x)P_1(x,t)] + \Big(K_{01}P_0(x,t) - K_{10}P_1(x,t)\Big) \end{cases} \tag{3.13}$$

The marginal probability density of the stochastic variable $x(t)$ is given by

$$P(x,t) \;=\; P_0(x,t) + P_1(x,t)$$

Stationary distributions

The asymptotic steady-state distributions $P_0(x)$ and $P_1(x)$ are the solution of the system obtained by setting to zero the l.h.s. of the master equation (3.13), namely

$$\begin{cases} \partial_x[f_0(x)P_0(x)] \;=\; -\Big(K_{01}P_0(x) - K_{10}P_1(x)\Big) \\ \partial_x[f_1(x)P_1(x)] \;=\; +\Big(K_{01}P_0(x) - K_{10}P_1(x)\Big) \end{cases} \tag{3.14}$$

Summing the two equations above we get

$$\partial_x\Big[\; f_0(x)P_0(x,t) + f_1(x)P_1(x,t) \;\Big] = 0$$

which means that in the asymptotic steady state the total probability flux is a (spatial) constant

$$f_0(x)P_0(x) + f_1(x)P_1(x) = J$$

The support of $P(x)$ is the compact interval $[x_l; 1]$ therefore in the steady state the local probability flux must vanish at the edge of the interval, but since it is a constant we have that $J = 0$ everywhere in $[x_l; 1]$ which yield

$$P_1(x) = -\frac{f_0(x)}{f_1(x)} P_0(x) \tag{3.15}$$

we can substitute this relation in (3.14) to derive an ordinary differential equation for $P_0(x)$ alone

$$\frac{dP_0(x)}{dx} + P_0(x) \left[\frac{d}{dx} \log |f_0(x)| + \frac{K_{10}}{f_1(x)} + \frac{K_{01}}{f_0(x)} \right] = 0$$

solving the equation with the definitions given in (3.12, 3.3, 3.1), we obtain the general solution

$$P_0(x) = \mathcal{N} \left[1 - x\right]^{m\frac{\tau_r}{\tau_E} - 1} \left[x - x_l\right]^{(1-m)\frac{\tau_d}{\tau_E}} \tag{3.16}$$

and with the relation (3.15) we immediately find

$$P_1(x) = \mathcal{N} \left[1 - x\right]^{m\frac{\tau_r}{\tau_E}} \left[x - x_l\right]^{(1-m)\frac{\tau_d}{\tau_E} - 1} \tag{3.17}$$

The constant $\mathcal{N}$ is set by the normalization condition

$$\int_{xl}^{1} P(x)dx \; = \int_{xl}^{1} \; P_0(x) + P_1(x)dx = 1$$

explicitly:

$$\mathcal{N} = \left\{ \int_{xl}^{1} \left[\frac{1}{1-x} + \frac{1}{x - x_l} \right] \left[1 - x\right]^{m\frac{\tau_r}{\tau_E}} \left[x - x_l\right]^{(1-m)\frac{\tau_d}{\tau_E}} \; dx \right\}^{-1} \tag{3.18}$$

Figure 3.2 shows a comparison between the joint distributions $P_0(x)$ and $P_1(x)$ obtained from a simulation of the process and the theoretical stationary distribution given above.

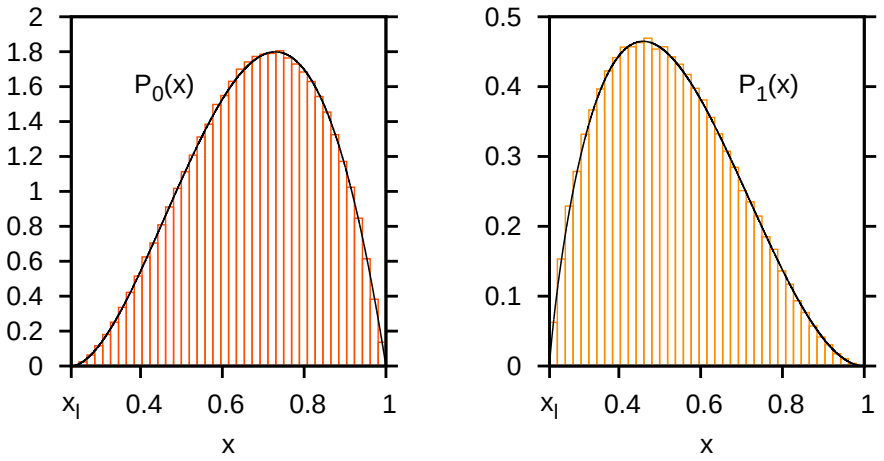

Figure 3.2: Bars: joint distributions $P_0(x)$ and $P_1(x)$ calculated from a simulation of the process with parameters $m = 0.15$, $\tau_r = 10$, $U = 0.3$ ($\tau_d \simeq 2.3$, $x_l \simeq 0.23$), (x, s) values were sampled every τ_E for a period of $10^6\tau_E$. Black lines: theoretical distributions $P_0(x)$ and $P_1(x)$ given in (3.16, 3.17) calculated for the same parameters.

We stress that $P_0(x)$ and $P_1(x)$ are joint distributions, taking the integral over x means marginalizing out the x variable so that we are left with the probability of having $S = 0$ and $S = 1$, then:

$$\int_{xl}^{1} P_0(x)dx = 1 - m \qquad \int_{xl}^{1} P_1(x)dx = m \tag{3.19}$$

Stationary moments of the distributions

We define the n^{th} moment of the joint distribution $P_a(x)$ as

$$\bar{x}_a^{(n)} \equiv \int_{x_l}^{1} x^n P_a(x)dx \qquad a = 0, 1$$

If we multiply the master equation (3.13) for x^n and integrate over x we get

$$\begin{aligned}
\frac{d\bar{x}_0^{(n)}}{dt} &= x^n f_0(x) P_0(x,t)\Big|_{xl}^{1} + \frac{n}{\tau_r}\left[\bar{x}_0^{(n-1)} - \bar{x}_0^{(n)}\right] - \left(K_{01}\bar{x}_0^{(n)} - K_{10}\bar{x}_1^{(n)}\right) \\
\frac{d\bar{x}_1^{(n)}}{dt} &= x^n f_1(x) P_1(x,t)\Big|_{xl}^{1} + \frac{n}{\tau_r}\left[\bar{x}_1^{(n-1)} - \frac{\bar{x}_1^{(n)}}{x_l}\right] + \left(K_{01}\bar{x}_0^{(n)} - K_{10}\bar{x}_1^{(n)}\right)
\end{aligned} \tag{3.20}$$

We are interested in the value of the moment of the stationary distributions $P_a(x)$ with $a = 0, 1$ namely the constant solution of (3.20).
We can calculate the first term on the r.h.s. of (3.20) using the explicit expression for the stationary distributions (3.16, 3.17) and we have that for any values of the parameters

$$x^n f_a(x) P_a(x)\Big|_{xl}^{1} = \mathcal{N}\, x^n \Big[1 - x\Big]^{m\frac{\tau_r}{\tau_E}} \Big[x - x_l\Big]^{(1-m)\frac{\tau_d}{\tau_E}} = 0$$

Dropping this boundary term and setting to zero the l.h.s. we get the following linear system

$$\begin{pmatrix} \frac{n}{\tau_r} + K_{01} & -K_{10} \\ -K_{01} & \frac{n}{\tau_r x_l} + K_{10} \end{pmatrix} \begin{pmatrix} \bar{x}_0^{(n)} \\ \bar{x}_1^{(n)} \end{pmatrix} = \frac{n}{\tau_r} \begin{pmatrix} \bar{x}_0^{(n-1)} \\ \bar{x}_1^{(n-1)} \end{pmatrix} \tag{3.21}$$

whose solutions give the stationary values of the n^{th} moments

$$\bar{x}_0^{(n)} = \frac{\left[n + U\tau_r + (1-m)\frac{\tau_r}{\tau_E}\right] \bar{x}_0^{(n-1)} + (1-m)\frac{\tau_r}{\tau_E}\bar{x}_1^{(n-1)}}{n + \frac{\tau_r}{\tau_E} + U\tau_r\left(1 + m\frac{\tau_r}{\tau_E}\right)} \tag{3.22a}$$

$$\bar{x}_1^{(n)} = \frac{\left[n + m\frac{\tau_r}{\tau_E}\right] \bar{x}_1^{(n-1)} + m\frac{\tau_r}{\tau_E}\bar{x}_0^{(n-1)}}{n + \frac{\tau_r}{\tau_E} + U\tau_r\left(1 + m\frac{\tau_r}{\tau_E}\right)} \tag{3.22b}$$

Since the moment of order n depends on the moment of order $(n-1)$, the moment of arbitrary order must be calculated recursively starting from the 0^{th} moments, which are the integral of the two distributions given in (3.19).

The calculation of the first moments of the stationary distributions gives

$$\bar{x}_0 = \frac{(1-m)\left(1 + U\tau_r + \frac{\tau_r}{\tau_E}\right)}{1 + \frac{\tau_r}{\tau_E} + U\tau_r\left(1 + m\frac{\tau_r}{\tau_E}\right)} \tag{3.23a}$$

$$\bar{x}_1 = \frac{m\,(1 + \frac{\tau_r}{\tau_E})}{1 + \frac{\tau_r}{\tau_E} + U\tau_r\left(1 + m\frac{\tau_r}{\tau_E}\right)} \tag{3.23b}$$

while for the second moment we have

$$\bar{x}_0^{(2)} = \frac{\left[2 + \frac{\tau_r}{\tau_d} + (1-m)\frac{\tau_r}{\tau_E}\right]\left[1 + \frac{\tau_r}{\tau_d} + \frac{\tau_r}{\tau_E}\right](1-m) + (1-m)\frac{\tau_r}{\tau_E} m \left[1 + \frac{\tau_r}{\tau_E}\right]}{\left[2 + \frac{\tau_r}{\tau_d} + \frac{\tau_r}{\tau_E}(1 + m\frac{\tau_r}{\tau_d})\right]\left[1 + \frac{\tau_r}{\tau_d} + \frac{\tau_r}{\tau_E}(1 + m\frac{\tau_r}{\tau_d})\right]} \tag{3.24a}$$

$$\bar{x}_1^{(2)} = \frac{\left[2 + m\frac{\tau_r}{\tau_E}\right] m \left[1 + \frac{\tau_r}{\tau_E}\right] + m\frac{\tau_r}{\tau_E}(1-m)\left[1 + \frac{\tau_r}{\tau_d} + \frac{\tau_r}{\tau_E}\right]}{\left[2 + \frac{\tau_r}{\tau_d} + \frac{\tau_r}{\tau_E}(1 + m\frac{\tau_r}{\tau_d})\right]\left[1 + \frac{\tau_r}{\tau_d} + \frac{\tau_r}{\tau_E}(1 + m\frac{\tau_r}{\tau_d})\right]} \tag{3.24b}$$

To simplify notation, from now on we set $\tau_E = 1$ so that all time constants are measured in units of τ_E.

Steady state MF equations in the DMN approximation

Under the DMN assumption all the moment of the distribution of x can be express in term of the average presynaptic activity m. In particular, the time average at equilibrium over a single synapse yields

$$\langle x_i(t) S_i^E(t) \rangle \equiv \int_{x_l}^1 x_i P_1(x) dx = \bar{x}_1(m_i) \tag{3.25}$$

$$\langle x_i^2(t) S_i^E(t) \rangle \equiv \int_{x_l}^1 x_i^2 P_1(x) dx = \bar{x}_1^{(2)}(m_i) \tag{3.26}$$

where $\bar{x}_1(m_i)$ and $\bar{x}_1^{(2)}$ are given in (3.23b) and (3.24b).

Therefore, the order parameters of the network, at equilibrium, can be calculated as population averages of the above functions, which for large N can be expressed as integrals over the distribution of the local activities:

$$r \equiv \langle r(t) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \bar{x}_1(m_i) = \int_0^1 \bar{x}_1(m) \rho_E(m) dm \tag{3.27a}$$

$$v \equiv \langle v(t) \rangle = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \bar{x}_1^{(2)}(m_i) = \int_0^1 \bar{x}_1^{(2)}(m) \rho_E(m) dm \tag{3.27b}$$

$$p = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} [\bar{x}_1(m_i)]^2 = \int_0^1 [\bar{x}_1(m)]^2 \rho_E(m) dm \tag{3.27c}$$

In the random network with static synapses (see section 1.7) the MF equations for the stationary average rates formed a closed system for mE, m_I because both h^A and σ^A depended exclusively on the population average rates, the quenched fluctuations could be determined separately as a closed system for q_E and q_I once the average rates were known.
In presence of synaptic dynamics, instead, the input mean h and the total variance σ depend on the new order parameters r and v that cannot be determined without knowing the local distribution of activity, thus the the quenched fluctuations q which depend also on the order parameter p

$$\begin{aligned} h_E &= \sqrt{C}(J_{EE}\, r - J_{EI}\, m_I + E m_0) \\ h_I &= \sqrt{C}(J_{IE}\, m_E - J_{II}\, m_I + I\, m_0) \end{aligned} \tag{3.28}$$

$$\begin{aligned} \sigma_E^2 &= J_{EE}^2\, v + J_{EI}^2 m_I & s_E^2 &= J_{EE}^2\, p + J_{EI}^2 q_I \\ \sigma_I^2 &= J_{IE}^2 m_E + J_{II}^2 m_I & s_I^2 &= J_{IE}^2 q_E + J_{II}^2 q_I \end{aligned}$$

therefore to find the stationary solution, a coupled system for the 7 order parameters of the system must be solved:

$$\begin{cases} m_E = H\left(\dfrac{\theta_E - h_E}{\sigma_E}\right) \\ m_I = H\left(\dfrac{\theta_I - h_I}{\sigma_I}\right) \\ q_E = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\dfrac{\theta_E - h_E + xs_E}{\sqrt{\sigma_E^2 - s_E^2}}\right)\right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ q_I = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\dfrac{\theta_I - h_I + xs_I}{\sqrt{\sigma_I^2 - s_I^2}}\right)\right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ r = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1 \left(H\left(\dfrac{\theta_E - h_E + xs_E}{\sqrt{\sigma_E^2 - s_E^2}}\right)\right) \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ v = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1^{(2)} \left(H\left(\dfrac{\theta_E - h_E + xs_E}{\sqrt{\sigma_E^2 - s_E^2}}\right)\right) \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ p = \displaystyle\int_{-\infty}^{+\infty} \left[\bar{x}_1 \left(H\left(\dfrac{\theta_E - h_E + xs_E}{\sqrt{\sigma_E^2 - s_E^2}}\right)\right)\right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \end{cases} \tag{3.29}$$

MF equations for a network with C connections per neuron

If we consider a simplified network architecture in which every neuron receives *exactly* C connections from each population, quenched fluctuations will be absent because and time averaged local rates are homogeneous across the network, i.e. the population distribution of activity of population A is

$$\rho_A(m) = \delta(m - m_A)$$

with m_A the average activity in that population.
With this simplification we have that $q_A = m_A^2$ and the order parameter of the network with STD (3.27) reduces to simple functions of the average excitatory population rate:

$$r = \int_0^1 \bar{x}_1(m)\delta(m - m_E)\, dm \quad \rightarrow \quad \bar{x}_1(m_E) \tag{3.30a}$$

$$v = \int_0^1 \bar{x}_1^{(2)}(m)\delta(m - m_E)\, dm \quad \rightarrow \quad \bar{x}_1^{(2)}(m_E) \tag{3.30b}$$

$$p = \int_0^1 [\bar{x}_1(m)]^2\delta(m - m_E)\, dm \quad \rightarrow \quad [\bar{x}_1(m_E)]^2 = r^2 \tag{3.30c}$$

with $\bar{x}_1(m)$ and $\bar{x}_1^{(2)}(m)$ given in (3.23b) and (3.24b).

And writing down the input mean and variance

$$\begin{aligned} h_E &= \sqrt{C}\,\left(J_{EE}\,\bar{x_1}(m_E) - J_{EI}\,m_I + m_0\,E\right) \\ h_I &= \sqrt{C}\,\left(J_{IE}\,m_E - J_{II}\,m_I + m_0\,I\right) \end{aligned} \tag{3.31}$$

$$\begin{aligned} \sigma_E^2 &= J_{EE}^2\left(\bar{x}_1^{(2)}(m_E) - [\bar{x}_1(m_E)]^2\right) + J_{EI}^2\left(m_I - m_I^2\right) \\ \sigma_I^2 &= J_{IE}^2\left(m_E - m_E^2\right) + J_{II}^2\left(m_I - m_I^2\right) \end{aligned} \tag{3.32}$$

we see that the input statistic, for this simplified network architecture, is completely determined by the population average rates, which are then the only independent parameters of the system and can be determined from the solution of the mean field system

$$\begin{cases} m_E = H\left(\dfrac{\theta_E - h_E}{\sigma_E}\right) \\ m_I = H\left(\dfrac{\theta_I - h_I}{\sigma_I}\right) \end{cases} \tag{3.33}$$

but in contrast with the network with static synapses, here we have that h_E and σ_E becomes *nonlinear* functions of m_E because of the synaptic dynamics. The effect of this nonlinearity on the balanced state are analyzed in the next section.

3.3 Balanced state with dynamical synapses

To get a qualitative insight of the balanced state in presence of depressing synaptic dynamics, as a first step it is convenient to consider the simpler MF equations (3.33) with (3.31,3.32) of the simplified architecture where all neurons receive C inputs for each population. After, we will get back to the complete model with variability in the connections, described by 3.29).

Balanced solution with no quenched fluctuations

The equilibrium rates in the $C \to \infty$ limit are found by the requirement that the average fields (3.31) must vanishes at leading order in $\sqrt{C}$. Whereas such "balancing" condition gave a linear system in the network with static synapses (1.38), in presence of STD we get a nonlinear system for the population average activities:

$$\begin{cases} J_{EE}\,\bar{x_1}(m_E) - J_{EI}\,m_I + m_0\,E = 0 \\ J_{IE}\,m_E - J_{II}\,m_I + m_0\,I = 0 \end{cases} \tag{3.34}$$

If we use the second equation to eliminate m_I in the first equation and substitute the explicit expression of $\bar{x_1}(m_E)$ we obtain

$$\frac{m_E(1+\tau_r)}{1+\tau_r+U\tau_r(1+m_E\tau_r)} = \frac{J_{EI}J_{IE}}{J_{EE}J_{II}}\,m_E + \frac{J_{EI}}{J_{EE}}\left(\frac{I}{J_{II}} - \frac{E}{J_{EI}}\right)m_0 \tag{3.35}$$

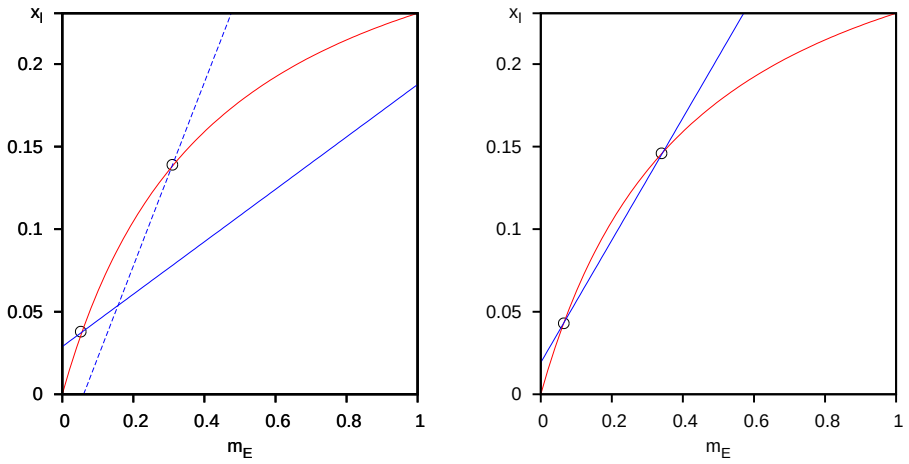

Figure 3.3: Graphic solution of equation (3.35) for excitatory rates. *Red line*: l.h.s. of equation (3.35) with $\tau_r = 10$, $\tau_d = 3$. *Blue line*: r.h.s. of equation (3.35) with a choice of parameter for which the intersection with the red curve are single (left panel) or double (right panel).

A graphical solution might be helpful to get the general picture: the equilibrium m_E is the abscissa of the intersection point(s) between the graph of the nonlinear function $\bar{x_1}(m_E)$ on the l.h.s. which includes all the effects of STD, and the straight line corresponding to the linear function on the r.h.s.; depending on the parameters a single or a double intersection can occur:

- The straight line crosses the negative part of y axis if

$$\frac{E}{I} > \frac{J_{EI}}{J_{II}} \tag{3.36}$$

  In this case a single intersection between the two curves in the interval $[0, 1]$ is possible (fig. 3.3 left, dashed line) .

- Otherwise, if

$$\frac{E}{I} < \frac{J_{EI}}{J_{II}} \tag{3.37}$$

  the straight line crosses the positive part of the y axis, and there can be a single (fig. 3.3 left, continuous line) or a double intersection (fig. 3.3 right, continuous line).

The solutions (3.34) can be found algebraically: rearranging the term in (3.35) we get the a quadratic equation for m_E

$$m_E^2 + bm_E + c = 0 \tag{3.38}$$

$$\begin{aligned} b &= \frac{1}{U\tau_r}\left[1+\frac{1}{\tau_r}+U\right]\frac{J_{EI}J_{IE}-J^d_{EE}J_{II}}{J_{EI}J_{IE}}+\frac{IJ_{EI}-EJ_{II}}{J_{EI}J_{IE}}\ m_0 \\ c &= \frac{1}{U\tau_r}\left[1+\frac{1}{\tau_r}+U\right]\frac{IJ_{EI}-EJ_{II}}{J_{EI}J_{IE}}\ m_0 \end{aligned} \tag{3.39}$$

where we have defined

$$J^d_{EE} = J_{EE}\,\frac{1}{1+\dfrac{U\tau_r}{1+\tau_r}} \tag{3.40}$$

To have meaningful values for m_E we require that the roots of (3.38) are real, positive, and less than 1.

When $c<0$ there are always two real roots ($b^2-4\,c>0$), one of which is positive and the other negative. In fact we have $c<0$ if (3.36) holds and the *single* equilibrium solution is given by

$$m_E = \frac{-b+\sqrt{b^2-4c}}{2} \tag{3.41a}$$

$$m_I = \frac{J_{IE}m_E+m_0 I}{J_{II}} \tag{3.41b}$$

While the network without static synapses displayed network activities that depended *linearly* on the external input (1.39), here we see that the nonlinear input summation determined by STD determines a nonlinear dependence of the population average activities on the external input.
If we take the limit $m_0\to 0$ in (3.38, 3.39) we have

$$\lim_{m_0\to 0^+} m_E = \begin{cases} 0 & \text{if} \quad \dfrac{J_{EI}}{J_{II}} > \dfrac{J^d_{EE}}{J_{IE}} \\ \dfrac{1}{U\tau_r}\left[1+\dfrac{1}{\tau_r}+U\right]\dfrac{J^d_{EE}J_{II}-J_{EI}J_{IE}}{J_{EI}J_{IE}} & \text{if} \quad \dfrac{J_{EI}}{J_{II}} < \dfrac{J^d_{EE}}{J_{IE}} \end{cases}$$

Therefore, the system has a single solution m_E which goes to zero when the external input vanish if the following inequalities hold

$$\frac{E}{I} > \frac{J_{EI}}{J_{II}} > \frac{J_{EE}}{J_{IE}}\frac{1}{1+U\frac{\tau_r}{1+\tau_r}} \tag{3.42}$$

If we consider the approximation of (3.41a) at leading order in m_0

$$m_E \;=\; m_0\,\frac{\partial}{\partial m_0}\left[\frac{-b+\sqrt{b^2-4c}}{2}\right]\Bigg|_{m_0=0} + O(m_0^2)$$

we have that the linear approximation of the population activities for very low external input m_0 gives

$$m_E \simeq \frac{EJ_{II}-IJ_{EI}}{J_{EI}J_{IE}-J_{II}J^d_{EE}}\ m_0 \qquad m_I \simeq \frac{EJ_{IE}-IJ^d_{EE}}{J_{EI}J_{IE}-J_{II}J^d_{EE}}\ m_0 \tag{3.43}$$

where the conditions (3.45) ensure that rates are positive. These expressions for the equilibrium rates are identical to those of the network without STD (1.39), with J^d_{EE} instead of J_{EE}. The effect of synaptic dynamics enters only through the coefficient (3.40) that reduces the absolute excitatory-to-excitatory strength.

Bistability ?

One striking difference introduced by dynamical synapses, it the possibility to have multiple balanced solutions. In fact equation (3.38) admit two solutions which are real, positive and smaller than 1, if the conditions[1]

$$\begin{cases} b^2 - 4c > 0 \\ b < 0 \\ 0 < c < 1 \end{cases} \quad \Rightarrow \quad \begin{cases} b < -2\sqrt{c} \\ 0 < c < 1 \end{cases} \tag{3.44}$$

are verified by the network parameters. But $c > 0$ corresponds to the condition (3.37) which is sufficient to allow the existence of the unbalanced solution in which the excitatory network is silent and the inhibitory network balance the external excitatory input ($h_E < 0$, $h_I \simeq 0$).
Thus, if (3.44) are fulfilled, the network has *three* equilibrium states: two balanced and a partially unbalanced one:

$$\begin{array}{lll} m_E^{(1)} = \dfrac{-b + \sqrt{b^2 - 4c}}{2} & m_E^{(2)} = \dfrac{-b - \sqrt{b^2 - 4c}}{2} & m_E^{(3)} = 0 \\ m_I^{(1)} = \dfrac{J_{IE} m_E + m_0 I}{J_{II}} & m_I^{(2)} = \dfrac{J_{IE} m_E + m_0 I}{J_{II}} & m_I^{(3)} = m_0 \dfrac{I}{J_{II}} \end{array}$$

While the partially unbalanced solution $m^{(3)}$ is certainly stable, the two balanced solutions appears at a saddle-node bifurcation therefore $m^{(2)}$ is unstable and we conclude that the system have *two stable equilibria.*
In the limit $m_0 \to 0$, we see that the excitatory activity

$$m_E^{(1)}(m_0 = 0) = \frac{1}{U\tau_r} \left[1 + \frac{1}{\tau_r} + U \right] \frac{J^d_{EE} J_{II} - J_{EI} J_{IE}}{J_{EI} J_{IE}}$$

would be positive provided that

$$\frac{E}{I} < \frac{J_{EI}}{J_{II}} < \frac{J_{EE}}{J_{IE}} \frac{1}{1 + U \frac{\tau_r}{1+\tau_r}} \tag{3.45}$$

Figure 3.4 shows an example of the bifurcation diagram when the system parameters allow bistabilty. We observe that the point in which the stable (straight line) and unstable (dotted line) solutions annihilate is where $b^2 - 4c = 0$; increasing m_0 we have that $b^2 - 4c < 0$ and the equation (3.38) has no real solution, thus the only stable solution is the unbalanced one.

[1] if x_1 and x_2 are the roots of $x^2 + bx + c$, we have that $b = -(x_1 + x_2)$ and $c = x_1 x_2$

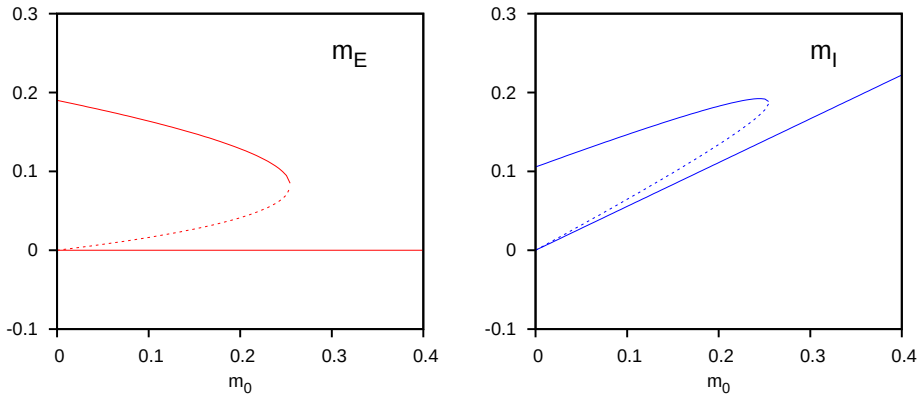

Figure 3.4: Excitatory (left panel) and inhibitory (right panel) average rates at equilibrium for a set of parameters that allows bistability. Continuous lines are the stable solutions, dashed lines the unstable ones. Parameters are $J_{EE} = 2$, $J_{EI} = 2$, $E = 1$, $J_{IE} = 1$, $J_{II} = 1.8$, $I = 1$, $\tau_r = 80$, $\tau_d = 20$.

Looking back to the graphical solution of (3.35) we can observe that the case in which (3.37) holds and there is a single intersection (figure 3.3 left, straight line) it corresponds necessarily to the unstable solution, because the stable solution would be above $m_E = 1$.

We point out that the occurrence of bistability in an unstructured network as a consequence of synaptic adaptation has been examined in [130] for a balanced network of spiking neurons with synaptic short-time facilitation.

Balanced solution with quenched fluctuations

After understanding the qualitative picture with the simplified model network without quenched fluctuations, we can go back to the original model where the neurons receive variable number of connections, C on average. The balancing condition for the average input in the $C \to \infty$ limit

$$\begin{cases} J_{EE}\, r \; - J_{EI}\, m_I + m_0\, E = 0 \\ J_{IE}\, m_E - J_{II}\, m_I + m_0\, I = 0 \end{cases} \tag{3.46}$$

yields the same nonlinear equation for the excitatory population activity

$$r \; = \frac{J_{EI} J_{IE}}{J_{EE} J_{II}}\, m_E + \frac{J_{EI}}{J_{EE}} \left(\frac{I}{J_{II}} - \frac{E}{J_{EI}} \right) m_0 \tag{3.47}$$

the only difference is that while in the simplified case where the homogeneity in local average rates allowed to express r as an algebraic function of the average excitatory rates $r = \bar{x}_1(m_E)$, this time the variability in the number of connections per neuron determines a variability in local average activities,

thus we can determine r as a function m_E by solving the complete statistics, i.e. taking m_E as a parameter of the system of equations:

$$\begin{cases} m_I = \dfrac{J_{IE} m_E + m_0 I}{J_{II}} \\ q_E = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\frac{\sigma_E H^{-1}(m_E) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ q_I = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\frac{\sigma_I H^{-1}(m_I) + x s_I}{\sqrt{\sigma_I^2 - s_I^2}} \right) \right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ r = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1 \left(H\left(\frac{\sigma_E H^{-1}(m_E) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ v = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1^{(2)} \left(H\left(\frac{\sigma_E H^{-1}(m_E) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ p = \displaystyle\int_{-\infty}^{+\infty} \left[\bar{x}_1 \left(H\left(\frac{\sigma_E H^{-1}(m_E) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \end{cases} \tag{3.48}$$

where the definitions for the input variances given in (3.28), and the residual inputs are found by inverting the transfer function $H(x)$ for m_E.
In figure 3.5 we can see that the graphical solution of (3.47) is completely analogous to the simplified case, and the effect of the inhomogeneities in the local rates does not change the qualitative picture.

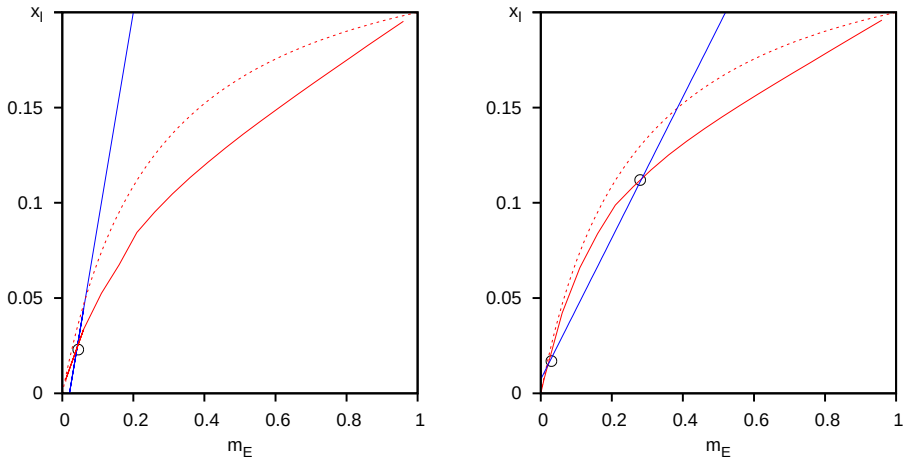

Figure 3.5: Graphic solution of equation (3.47) for excitatory rates: single (left panel) and double (right panel) intersections between the graph of $r(M_E)$, red continuous curve - obtained from a numerical solution of (3.48), and the r.h.s. of equation (3.47), blue line. *Red dashed line*: graph of $\bar{x}_1(m_E)$. ($\tau_r = 80$, $\tau_d = 20$).

The effect of quenched fluctuation is just to lower the population averaged synaptic efficacy of the active neurons, without altering dramatically the shape of the function; in particular, it can be demonstrated (using the argument outlined in appendix A.5) that when the external input is low m_0, at first order in m_0 we have $r(m_E) \simeq \bar{x}_1(m_E)$. Essentially, this is consequence of the fact that when activity levels are low, quenched fluctuations becomes negligible with respect to the temporal ones.
Depending on the parameters of the system, we can have, again, a single or a double intersection between the curves; similarly, one of the necessary condition for having a double intersection (the straight line must cross the positive part of the y axis) is sufficient to allow the existence of unbalanced solution (3.37). We can thus conclude that also the network with variable number of connection there can be bi-stability between the partially unbalanced solution and a balanced solution.

In the first chapter we have seen that for a network with static synapses, the balanced state yield a *linear* response in the $C \to \infty$ limit (1.39). Otherwise, when synapses are endowed with STD, the average input (3.6) involve a nonlinear dependence on the network activity through the order parameter r with the effect that the average activities in the balanced state have a *nonlinear* dependence on the external input m_0. Figure 3.6 shows the population average rates for three different values of the synaptic recovery time τ_r.

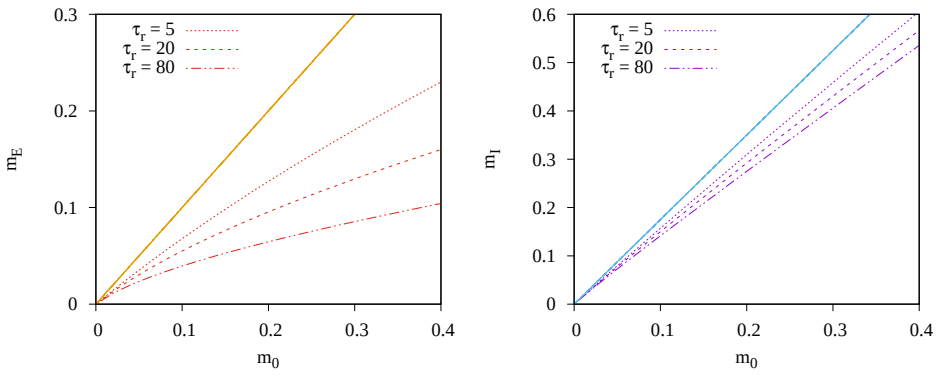

Figure 3.6: Average rates in the balanced state at $C \to \infty$ for a network with static synapses (continuous line) and for the same network with STD (dashed lines) with three different timescales for the synaptic recovery time (parameters (1.70), $U = 0.05$).

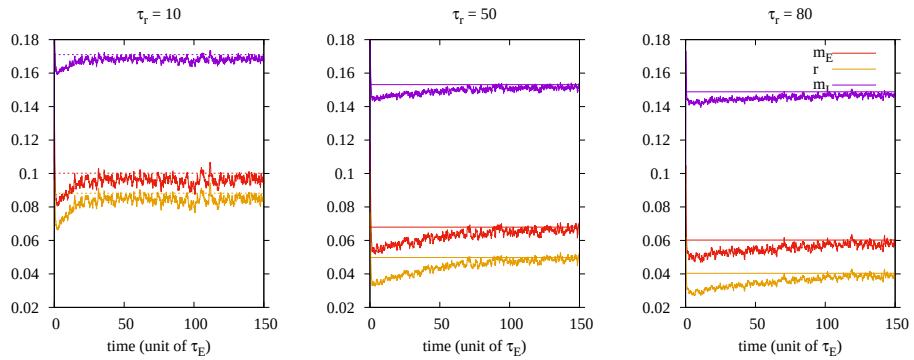

Figure 3.7: Time of convergence to equilibrium in presence of synaptic depression: comparison between the instantaneous average activities for three different values of τ_r with $U = 0.05$. Other parameters are (1.70), $m_0 = 0.1$

3.4 Comparing MF theory with simulations

To test the accuracy of the MF theory we calculated the solution of the MF equations (3.29) with $C = 10^3$ and compared it with the steady states order parameter measured from simulations of a network with $N_E = N_I = 10^4$.

First, we observe that while the network with static synapses converge to stationary state in a few network update cycle (see figure 1.14), in presence of synaptic depression, the stationary state is reached in a time-scale of the same order of magnitude of the synaptic recovery time-scale τ_r (figure 3.7). Despite the MF theory neglects the effect of autocorrelations, the predictions are generally in a very good agreement with the values of the order parameters measured from the simulation (fig 3.9), as it's confirmed by the distribution of the time averaged local rates (fig. 3.8).

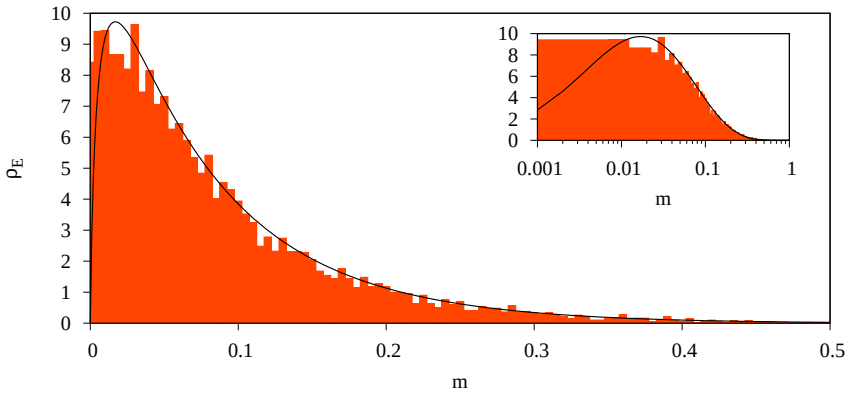

Figure 3.8: Distribution of excitatory neuron average activities with $\tau_r = 10$, $U = 0.05$, $m_0 = 0.08$ other parameters as usual. Comparison between theory and simulations.

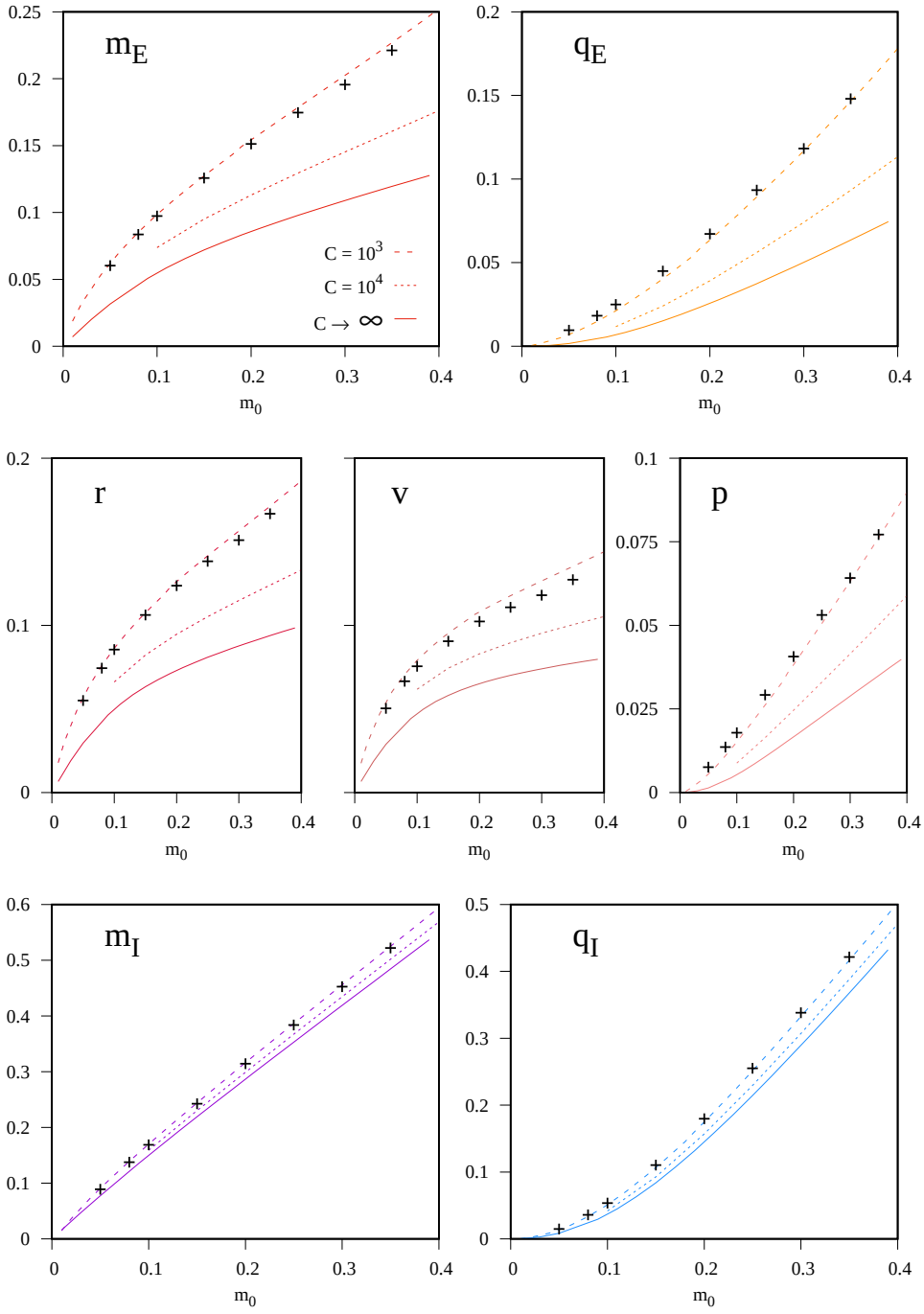

Figure 3.9: Example of MF solutions for the order parameter of the systems as a function of m_0 for $C = 10^3$, $C = 10^4$ and in the $C \to \infty$ limit. Crosses are the values measured form simulations of a network of $N = 10^4$ with $C = 10^3$. Parameters are (1.70) and $\tau_r = 10$, $U = 0.05$.

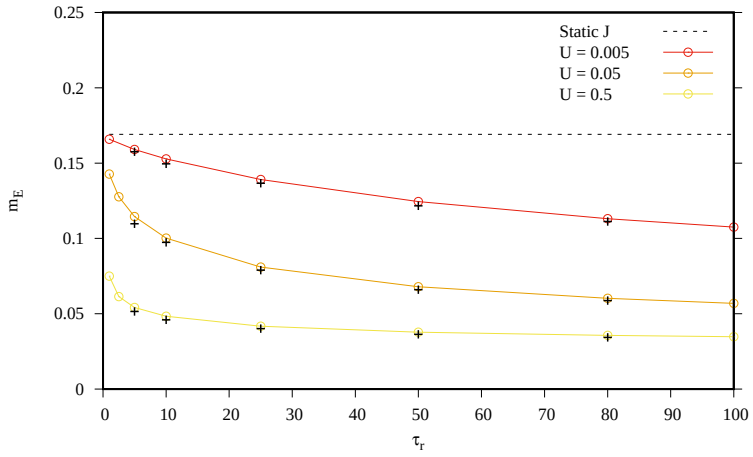

Figure 3.10: Excitatory activity as a function of the synaptic adaptation parameters, $m_0 = 0.1$

The values of the excitatory rates recorded from simulations are slightly smaller than the theoretical ones (fig. 3.9) and simulations for various values of the synaptic adaptation parameters τ_r and U shown in fig. 3.10 shows that the agreement are generally very good but the discrepancy are more pronounced for smaller values of τ_r. This is because the MF theory has been derived under the hypotheses of negligible self-correlations of neurons activities, and comparing the theoretical and simulated distributions of synaptic efficacy $P_1(x)$ (fig. 3.11) we see that when the timescale of the synaptic dynamics τ_r is much longer than the timescale of self-correlation decay, which is of the order of magnitude of the excitatory network update mean time τ_E (see section 1.8), the distribution of x driven by the activity $S(t)$ is better approximated by the distribution of the x driven by a dichotomous Markov noise with the same activity m.

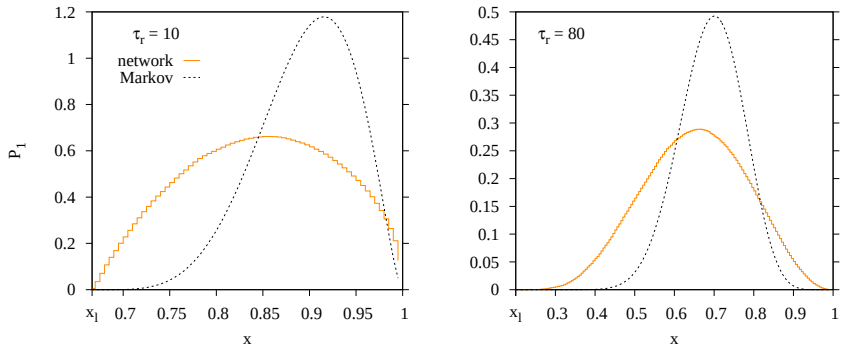

Figure 3.11: Verification of the Markov approximation, distribution of synaptic efficacies for a neuron in a network (orange) and distribution derived from a dichotomous Markov noise with the same activity level(black) for short τ_r (right) and long τ_r (right). $U = 0.05$

Chapter 4

A balanced memory network with depressing synapses

Nevertheless, the one and only thing of paramount interest to us in ourselves is, that we feel and think and perceive.

Erwin Schrödinger
What is Life?, 1944

In this chapter we extend the MF theory for a balanced network with short-term synaptic depression to the modified Willshaw memory model presented in section 2.2 to study how synaptic dynamics can modify the bistability in the sparse coding limit, and in particular if the current model can support the existence of low rate memory states.

4.1 MF theory for a memory model with STD

The modified Willshaw Model [19] gives a simple prescription to structure the connectivity between excitatory cells in order to reinforce the connection between neurons which are selective for a finite number P of *independent* memories patterns $\{\xi_i^\mu\}$ with coding level f

$$\begin{matrix}\forall i = 1, ..., N \\ \forall \mu = 1, ..., P\end{matrix} \qquad \xi_i^\mu = \begin{cases} 1 & \text{with probability } f \\ 0 & \text{with probability } 1-f \end{cases}$$

The excitatory-to-excitatory connections are built in the following way: if two neuron are simultaneously active in *at least* one of the patterns, the synapse between them is multiplied by a factor a, otherwise the synaptic strength is left at the baseline value $J_{EE}/\sqrt{C}$.

The symmetric synaptic structure is then diluted as usual yielding the following form for the synaptic matrix for the excitatory population:

$$J_{ij}^{EE} = c_{ij}^{EE} \frac{J_{EE}}{\sqrt{C}} \left\{ 1 + (a-1)\Theta \left[\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\nu \right] \right\} \tag{4.1}$$

The inputs to a generic excitatory and inhibitory neuron read

$$h_i^E(t) = \sum_{j=1}^{N_E} J_{ij}^{EE} x_j(t) S_j^E(t) + \frac{J_{EI}}{\sqrt{C}} \sum_{j=1}^{N_I} c_{ij}^{EI} S_j^I(t) + m_0 E \sqrt{C} \tag{4.2a}$$

$$h_i^I(t) = \frac{J_{IE}}{\sqrt{C}} \sum_{j=1}^{N_E} c_{ij}^{IE} S_j^E(t) + \frac{J_{II}}{\sqrt{C}} \sum_{j=1}^{N_I} c_{ij}^{II} S_j^I(t) + m_0 I \sqrt{C} \tag{4.2b}$$

where the recurrent excitatory synaptic matrix J_{ij}^{EE} is given in (4.1) and the x_j modulate the synaptic strength following a temporal dynamics determined by the state of S_j as specified by the model (3.1).

Input statistics

For each pattern ξ^ν the excitatory network is partitioned in two sub-populations: the selective (or foreground) neurons for which $\xi_i^\nu = 1$, and the nonselective (or background) neurons for which $\xi_i^\nu = 0$.
For a macroscopic characterization of the state of the system the patterns are all equivalent, therefore in the following we will drop the superscript ν and assume we are considering pattern ξ^1.

The population averaged inputs to the two excitatory sub-populations are

$$h_+(t) \equiv \frac{\sum_{i=1}^{N} \xi_i^1 h_i^E(t)}{\sum_{i=1}^{N} \xi_i^1} \qquad h_-(t) \equiv \frac{\sum_{i=1}^{N} (1-\xi_i^1) h_i^E(t)}{\sum_{i=1}^{N} (1-\xi_i^1)}$$

Calculation of the time averaged input for all the three sub-populations in the network proceeds in the same way as for the model with static synapses [19] presented in chapter 2 and gives

$$h_+ = \sqrt{C}\{ J_{EE}[afr_+ + (a\Pi + 1 - \Pi)(1-f)r_-] - J_{EI} m_I + E m_0\} \tag{4.3a}$$

$$h_- = \sqrt{C}\{ J_{EE}(a\Pi + 1 - \Pi)[fr_+ + (1-f)r_-] - J_{EI} m_I + E m_0\} \tag{4.3b}$$

$$h_I = \sqrt{C}\{ J_{IE}[fm_+ + (1-f)m_-] - J_{II} m_I + I m_0\} \tag{4.3c}$$

where $\Pi = 1 - (1-f^2)^{P-1}$ is the probability that a strong synapse exists between excitatory neurons in the background, or between a foreground and

a background cell. The difference is that with synaptic dynamics, besides the average activity in each population

$$m_+ \equiv \frac{\sum_{i=1}^{N} \xi_i^1 m_i^E}{\sum_{i=1}^{N} \xi_i^1} \qquad\qquad m_- \equiv \frac{\sum_{i=1}^{N} (1-\xi_i^1) m_i^E}{\sum_{i=1}^{N} (1-\xi_i^1)}$$

we have for each excitatory subpopulation the additional order parameters r representing the average synaptic efficacy for the active neurons

$$r_+ = \frac{\sum_{i=1}^{N} \xi_i^1 \langle x_i(t) S_i^E(t) \rangle}{\sum_{i=1}^{N} \xi_i^1} \qquad\qquad r_- = \frac{\sum_{i=1}^{N} (1-\xi_i^1) \langle x_i(t) S_i^E(t) \rangle}{\sum_{i=1}^{N} (1-\xi_i^1)}$$

As we have shown in section 1.6, in a sparsely connected network $C \ll N$ in the limit $N \to \infty$ and $C \to \infty$ the fluctuations of the input around the mean are Gaussian and can be separated into temporal and quenched(spatial) fluctuations:

$$h_i^A(t) = h_A + \eta_i^A(t) + \delta_i^A$$

The total variance of the input is given by

$$\sigma_+^2 = J_{EE}^2 [a^2 f v_+ + (a^2 \Pi + 1 - \Pi)(1-f) v_-] + J_{EI}^2 m_I \tag{4.4a}$$
$$\sigma_-^2 = J_{EE}^2 (a^2 \Pi + 1 - \Pi)[f v_+ + (1-f) v_-] + J_{EI}^2 m_I \tag{4.4b}$$
$$\sigma_I^2 = J_{IE}^2 [f m_+ + (1-f) m_-] + J_{II}^2 m_I \tag{4.4c}$$

with the definitions

$$v_+ = \frac{\sum_{i=1}^{N} \xi_i^1 \langle x_i^2(t) S_i^E(t) \rangle}{\sum_{i=1}^{N} \xi_i^1} \qquad\qquad v_- = \frac{\sum_{i=1}^{N} (1-\xi_i^1) \langle x_i^2(t) S_i^E(t) \rangle}{\sum_{i=1}^{N} (1-\xi_i^1)}$$

The variance of the time-independent fluctuating part of the input δ_i^A is:

$$s_+^2 = J_{EE}^2 [a^2 f p_+ + (a^2 \Pi + 1 - \Pi)(1-f) p_-] + J_{EI}^2 q_I \tag{4.5a}$$
$$s_-^2 = J_{EE}^2 (a^2 \Pi + 1 - \Pi)[f p_+ + (1-f) p_-] + J_{EI}^2 q_I \tag{4.5b}$$
$$s_I^2 = J_{IE}^2 [f q_+ + (1-f) q_-] + J_{II}^2 q_I \tag{4.5c}$$

with the definitions

$$v_+ = \frac{\sum_{i=1}^{N} \xi_i^1 \langle x_i(t) S_i^E(t) \rangle^2}{\sum_{i=1}^{N} \xi_i^1} \qquad\qquad v_- = \frac{\sum_{i=1}^{N} (1-\xi_i^1) \langle x_i(t) S_i^E(t) \rangle^2}{\sum_{i=1}^{N} (1-\xi_i^1)}$$

$$q_+(t) \equiv \frac{\sum_{i=1}^{N} \xi_i^1 [m_i^E(t)]^2}{\sum_{i=1}^{N} \xi_i^1} \qquad\qquad q_-(t) \equiv \frac{\sum_{i=1}^{N} (1-\xi_i^1) [m_i^E(t)]^2}{\sum_{i=1}^{N} (1-\xi_i^1)}$$

MF equations

A complete description of the network state is therefore given by 12 order parameters: 5 for each excitatory population and 2 for the inhibitory population. If self-correlations in the neurons' state are neglected (dichotomous Markov noise approximation), the stationary state value of the order parameters are a solutions of the fully coupled MF system

$$
\begin{cases}
m_+ = H\left(\dfrac{\theta_+ - h_+}{\sigma_+}\right) \\
m_- = H\left(\dfrac{\theta_- - h_-}{\sigma_-}\right) \\
m_I = H\left(\dfrac{\theta_I - h_I}{\sigma_I}\right) \\
q_+ = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\frac{\theta_E - h_+ + x s_+}{\sqrt{\sigma_+^2 - s_+^2}}\right)\right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
q_- = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\frac{\theta_E - h_- + x s_-}{\sqrt{\sigma_-^2 - s_-^2}}\right)\right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
q_I = \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\frac{\theta_I - h_I + x s_I}{\sqrt{\sigma_I^2 - s_I^2}}\right)\right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
r_+ = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1\left(H\left(\frac{\theta_E - h_+ + x s_E}{\sqrt{\sigma_+^2 - s_+^2}}\right)\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
r_- = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1\left(H\left(\frac{\theta_E - h_- + x s_-}{\sqrt{\sigma_-^2 - s_-^2}}\right)\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
v_+ = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1^{(2)}\left(H\left(\frac{\theta_E - h_+ + x s_+}{\sqrt{\sigma_+^2 - s_+^2}}\right)\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
v_- = \displaystyle\int_{-\infty}^{+\infty} \bar{x}_1^{(2)}\left(H\left(\frac{\theta_E - h_- + x s_-}{\sqrt{\sigma_-^2 - s_-^2}}\right)\right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
p_+ = \displaystyle\int_{-\infty}^{+\infty} \left[\bar{x}_1\left(H\left(\frac{\theta_E - h_+ + x s_+}{\sqrt{\sigma_+^2 - s_+^2}}\right)\right)\right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\
p_- = \displaystyle\int_{-\infty}^{+\infty} \left[\bar{x}_1\left(H\left(\frac{\theta_E - h_- + x s_-}{\sqrt{\sigma_-^2 - s_-^2}}\right)\right)\right]^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx
\end{cases}
\tag{4.6}
$$

Where $\bar{x_1}(m)$ and $\bar{x_1}^{(2)}(m)$ are given in (3.23b) and (3.24b) and are the first two moments of the distribution of synaptic efficacy driven by a memoryless neuron with average activity m.

MF for a memory network with C connections per neuron

A much more simple description can be obtained for a memory network in which there is no variability in the number of connections received by neurons.

As shown in section 3.2 for the network with unstructured connectivity, in a network where all neurons receive exactly C connections from the excitatory population (and exactly C from the inhibitory population) there is no cell-to-cell variability in the average activity, the distribution of the local average activity reduces to $\rho(m) = \delta(m - m_E)$ allowing to reduce the order parameters r and v to explicit functions of the average population activity:

$$r_\pm \to \bar{x}(m_\pm) \;=\; \frac{m_\pm(1+\tau_r)}{1+\tau_r+U\tau_r(1+m_\pm\tau_r)} \tag{4.7}$$

$$v_\pm \to\equiv \bar{x}^{(2)}(m_\pm) \;=\; \frac{m_\pm\left[2+\frac{\tau_r}{\tau_E}m_\pm\right]\left[1+\frac{\tau_r}{\tau_E}\right]+\frac{\tau_r}{\tau_E}m_\pm(1-m_\pm)\left[1+U\tau_r+\frac{\tau_r}{\tau_E}\right]}{\left[2+U\tau_r+\frac{\tau_r}{\tau_E}(1+U\tau_r\,m_\pm)\right]\left[1+U\tau_r+\frac{\tau_r}{\tau_E}(1+U\tau_r m_\pm)\right]}$$

The average input for each population are:

$$\begin{aligned}
h_+ =&\sqrt{C}\{\, J_{EE}[af\bar{x}_+ + (a\Pi+1-\Pi)(1-f)\bar{x}_-] - J_{EI}m_I + Em_0\}\\
h_- =&\sqrt{C}\{\, J_{EE}(a\Pi+1-\Pi)[f\bar{x}_+ + (1-f)\bar{x}_-] - J_{EI}m_I + Em_0\}\\
h_I =&\sqrt{C}\{\, J_{IE}[fm_+ + (1-f)m_-] - J_{II}m_I + Im_0\}
\end{aligned} \tag{4.8}$$

Quenched fluctuation absent, the input fluctuation is due exclusively to the temporal stochasticity in the presynaptic activity, and for the input variance we have

$$\begin{aligned}
\sigma_+^2 &= J_{EE}^2[a^2 f(\bar{x}_+^{(2)} - [\bar{x}_+]^2) + (a^2\Pi+1-\Pi)(1-f)(\bar{x}_-^{(2)} - [\bar{x}_-]^2)] + J_{EI}^2(m_I - m_I^2)\\
\sigma_-^2 &= J_{EE}^2(a^2\Pi+1-\Pi)[f(\bar{x}_+^{(2)} - [\bar{x}_+]^2) + (1-f)(\bar{x}_-^{(2)} - [\bar{x}_-]^2)] + J_{EI}^2(m_I - m_I^2)\\
\sigma_I^2 &= J_{IE}^2[f(m_+ - m_+^2) + (1-f)(m_- - m_-^2)] + J_{II}^2(m_I - m_I^2)
\end{aligned}$$

In this simpler model, the number of order parameters necessary to characterize completely the network state is reduced from 12 to the 3 population average activities, and their stationary value is found solving the system

$$\begin{cases}
m_+ = H\left(\dfrac{\theta_E - h_+}{\sigma_+}\right)\\
m_- = H\left(\dfrac{\theta_E - h_-}{\sigma_-}\right)\\
m_I = H\left(\dfrac{\theta_I - h_I}{\sigma_I}\right)
\end{cases} \tag{4.9}$$

4.2 Balanced regime with synaptic depression

To get a qualitative insight in the balanced regime we consider the model in which every neuron receives exactly C inputs from each population. Randomness in the number of connection will determine just a quantitative correction to the solutions found in the simplified architecture.

Balanced regime in the sparse coding limit $f \to 0$

As was shown in chapter 2, to have that in the $C \to \infty$ limit all three populations are in the balanced state (the mean input of all three subpopulation is of order 1 as the input variance) one possibility is to assume that the coding level of the memories scales with the connectivity as

$$f = \frac{F}{\sqrt{C}}$$

and the number of stored patterns $P = \alpha C$, with the loading parameter α.

In fact, with this scaling we get expressions for average inputs which are almost identical to (2.7) with the crucial difference that with dynamical synapses the average input becomes a *nonlinear* function of the recurrent excitatory activity through $\bar{x}$

$$\begin{aligned} h_I =& \sqrt{C}\{\, J_{IE}\, m_- - J_{II} m_I + I m_0\} \; + F J_{IE}\,(m_+ - m_-) \\ h_- =& \sqrt{C}\{\, J^*_{EE}\, \bar{x}(m_-) - J_{EI} m_I + E m_0\} + F J^*_{EE}(\bar{x}(m_+) - \bar{x}(m_-)) \\ h_+ =& \; h_- \; + (a-1)(1-\Pi) J_{EE} F\, \bar{x}(m_+) \end{aligned} \tag{4.10a}$$

where $J^*_{EE} = J_{EE}[a\Pi + 1 - \Pi]$. And analogously for the input variance.

Background an inhibitory population activities

When taking the limit $C \to \infty$ the variance are finite and for the two excitatory sub-populations becomes coincident

$$\begin{aligned} \sigma_E \equiv \sigma^2_+ \; =& \; \sigma^2_- = \; J^2_{EE}(a^2\Pi + 1 - \Pi)(\bar{x}^{(2)}_- - [\bar{x}_-]^2) + J^2_{EI}(m_I - m_I^2) \\ \sigma^2_I \; =& \; J^2_{IE}(m_- - m_-^2) + J^2_{II}(m_I - m_I^2) \end{aligned}$$

and since the difference between the input of the foreground and the background is finite, all three populations are automatically balanced as long as the large excitatory background and inhibitory population rates fulfil the balance condition

$$\begin{cases} J^*_{EE}\, \bar{x}(m_-) - J_{EI} m_I + E m_0 \; = 0 \\ \qquad J_{IE}\, m_- - J_{II} m_I + I m_0 \; = 0 \end{cases} \tag{4.11}$$

This condition is identical to that for an unstructured network with STD on excitatory synapses studied in chapter, the memory structure enters through the modified recurrent excitatory coupling $J^*_{EE} = a\Pi + 1 - \Pi$, and as was shown in chapter 3, provided that the parameters of the system satisfy the conditions

$$\frac{a\Pi + 1 - \Pi}{1 + \frac{U\tau_r}{1+\tau_r}} \frac{J_{EE}}{J_{IE}} < \frac{J_{EI}}{J_{II}} < \frac{E}{I}$$

there is a single stable equilibrium with average population activities given by

$$m_- = \frac{-b + \sqrt{b^2 - 4c}}{2} \qquad m_I = \frac{J_{IE} m_E + m_0 I}{J_{II}} \tag{4.12}$$

where

$$b = \frac{1}{U\tau_r}\left[1 + \frac{1}{\tau_r} + U\right] \frac{J_{EI}J_{IE} - (a\Pi + 1 - \Pi)J^d_{EE}J_{II}}{J_{EI}J_{IE}} + \frac{IJ_{EI} - EJ_{II}}{J_{EI}J_{IE}}\ m_0$$
$$c = \frac{1}{U\tau_r}\left[1 + \frac{1}{\tau_r} + U\right] \frac{IJ_{EI} - EJ_{II}}{J_{EI}J_{IE}}\ m_0$$

with $J^d_{EE} = J_{EE}/(1 + U\tau_r/(1 + \tau_r))$

Foreground population activity

At this point the equilibrium rate of the foreground population can be determined from the equation

$$m_+ = H\left(\frac{\theta_E - h_+}{\sigma_E}\right)$$

With the background an inhibitory average activities (4.12) determined by the balance condition, we can calculate the input variance σ_E and the net average residual input $\theta_E - h_- = \sigma_E H^{-1}(m_-)$, and with the input difference between foreground and background given by (4.10a) we get a nonlinear equation depending on m_+ alone

$$m_+ = H\left(H^{-1}(m_-) - \frac{J_{EE}(a-1)Fe^{-\alpha F^2}}{\sigma_E}\ \bar{x}(m_+)\right)$$

Since $\bar{x}(m_+)$ is a monotonically growing function in the interval $[0; 1]$, thus invertible, we can express m_+ as a function of $\bar{x}$ to get an equation for the variable $\bar{x}$:

$$\boxed{\frac{1 + \tau_r + U\tau_r}{1 + \tau_r - U\tau_r^2\, x}\ x\ =\ H\left(H^{-1}(m_-) - \frac{J_{EE}(a-1)Fe^{-\alpha F^2}}{\sigma_E}\ x\right)} \tag{4.13}$$

This is analogous to (2.9), the equation for a memory network with statical synapse (which is recovered setting $U = 0$), the essential difference is that

the l.h.s. of (4.13) is a nonlinear function of m_+ because of short-term depression of synapses, opening the possibility to make multiple intersections with the sigmoid function below the inflection point, thus allowing the low rate bistability sought for. We examine this possibility in the next section with a graphical analysis of the solutions of (4.13).

4.3 The effect of synaptic depression on bistability

The space of parameter of the system has 12 dimensions[1], but to study the graphical solution of (4.13) we can consider the problem in the general form

$$\underbrace{\frac{R-Q}{1-Qx}\,x}_{\phi(x)} \;=\; \underbrace{H(C-Dx)}_{\psi(x)} \tag{4.14}$$

of the occurrence of intersection between the functions $\phi(x)$ and $\psi(x)$ which depends on only 4 parameters, related to the parameters of the original problem by:

$$R = 1 + U\tau_r = \frac{1}{x_l} \qquad C = H^{-1}(m_-)$$

$$Q = U\tau_r \frac{1}{1+\frac{1}{\tau_r}} \qquad D = \frac{J_{EE}(a-1)}{\sigma_E} F e^{-\alpha F^2}$$

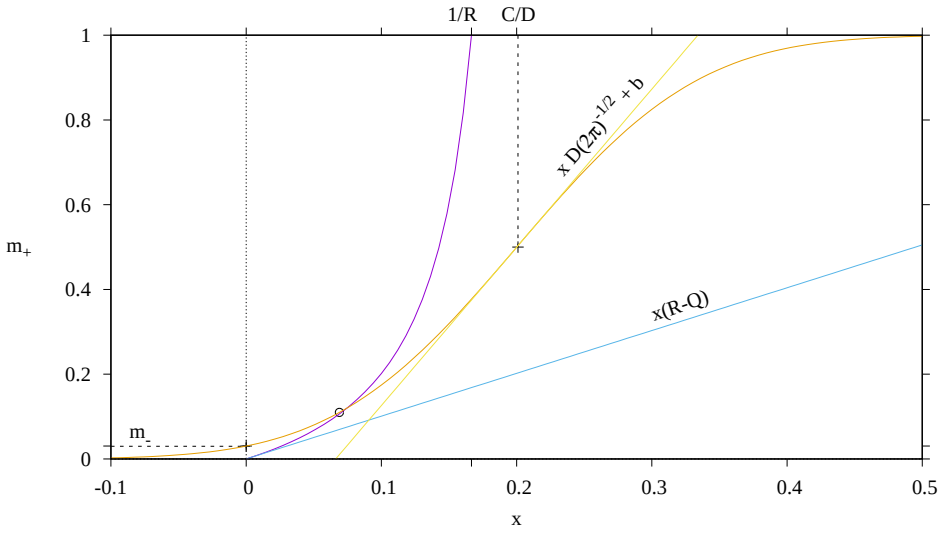

Figure 4.1: Geometrical significance of the parameters in $\phi(x)$ (red) and $\psi(x)$ (blue).

[1] 6 parameters for coupling strength J_{AB}, 3 for memory structure a, F, α, 2 for synaptic depression U, tau_r and 1 for external input m_0

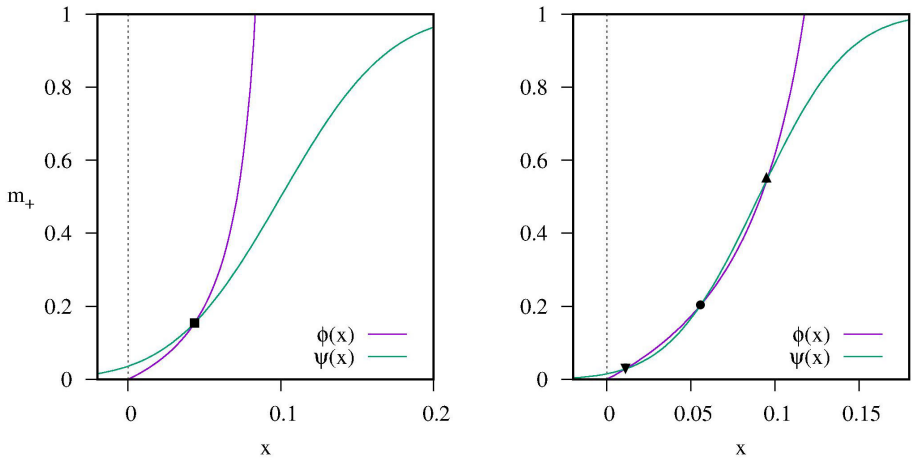

Figure 4.2: Equation (4.14) can have one (left) or three (right) solutions.

Figure (4.1) gives a summary of the relations between the four coefficients $\{R, Q, C, D\}$ and the geometrical properties of the curves, further details can be found in appendix C.
Of course the 4 parameters are not independent: while R and Q are related to U and τ_r only with an invertible variable transformaton, the coefficients C and D that determine the shape of function $\psi(x)$ depends on all the (12) parameters of the system because m_- and σ_E, appearing respectively in C and D, comes from the solution of the MF system (4.19).

It is easily understood that there are only two possible cases:

1. *the two curves meets in a single point* (fig. 4.2 left): the parameters allow a unique (stable) equilibrium solution m_+

2. *the two curves meets in three points* (fig. 4.2 right): the middle intersection corresponds to an un unstable solution while the upper and lower intersection correspond to stable solutions. Therefore the network is *bistable*: the lower activity solution m_+^{low} corresponds to background, unstructured, state of activity, the higher rate m_+^{hi} solution correspond to memory retrieval state.

In particular, we are interested to see if for certain values of the parameters there are two stable solutions for the foreground population fulfilling

$$m_+^{low} \simeq m_- \qquad\qquad m_+^{hi} \ll 1$$

as, for example in fig.4.3.
A preliminary exploration of the solutions of equation (4.14) as a function of the 4 parameters (see appendix C) showed that it is possible to have double intersection with the upper one occurring much lower than 1. This

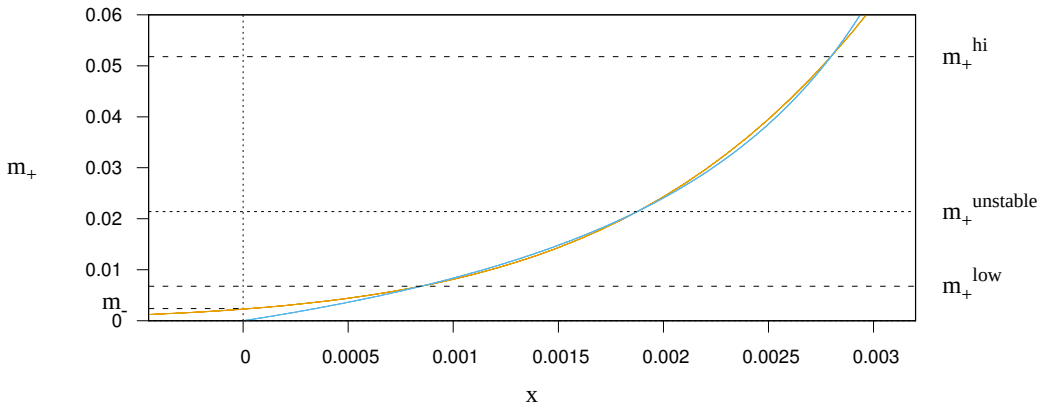

Figure 4.3: Example of two stable solution at low rate.

give a necessary condition for the existence of low rates solutions but it's not sufficient because the parameters C and D depends on R and Q. Indeed, scanning the full parameter space of (4.13) and calculating the solutions, showed that those region are accessible, proving our initial hypotheses:

> **for binary memory network model (modified Willshaw) with short-time depressing synapses, there is a finite region in the parameter space for which the selective population display bistability between two balanced states, both with average activities largely below saturation.**

As an example, a set of double solutions (m_+^{hi}, m_+^{low}) parameterized by U and τ_r is shown in figure 4.5, where we notice that the activity of the memory states m_+^{hi} lies all below 0.2.

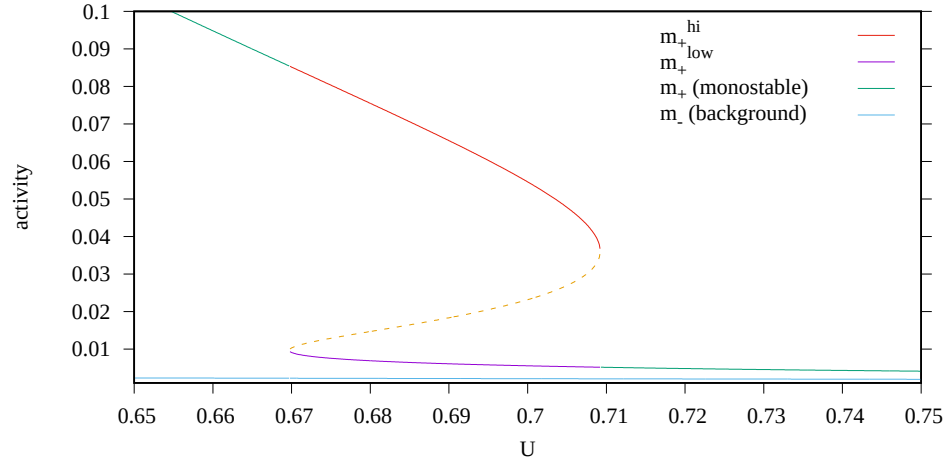

Figure 4.4: Bifurcation diagram. Solutions of (4.13) as a function of U, other parameters are fixed at $\tau_r = 80$, $a = 2, 2$, $F = 32$, $\alpha = 6 * 10^{-4}$, $m_0 = 0.005$

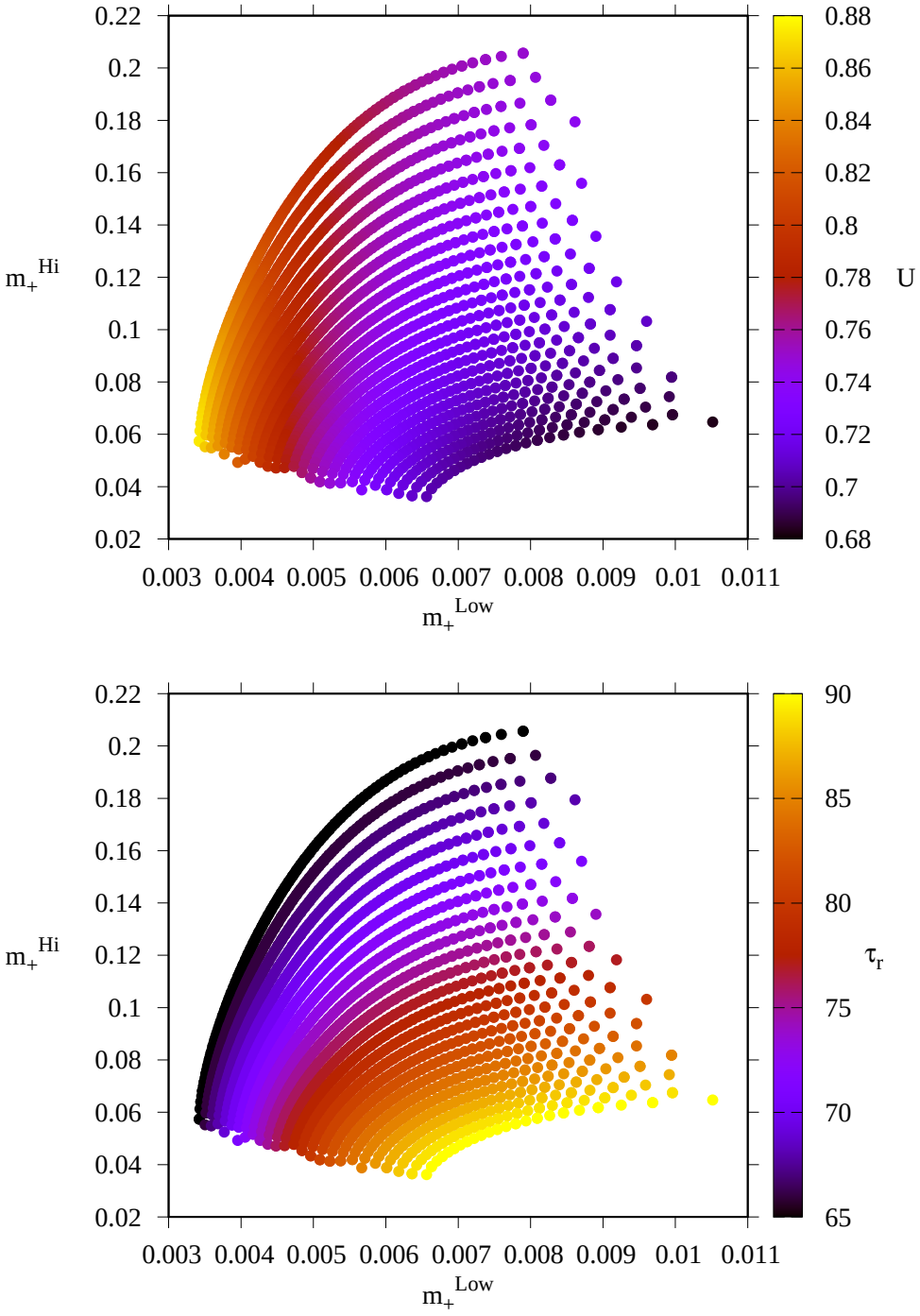

Figure 4.5: Bistability at low rate: example of a set of solutions for the foreground average activity as a function of the synaptic adaptation parameters U (top) and τ_r (bottom). Other parameters are fixed at $a = 2,2$, $F = 34$, $\alpha = 6 * 10^{-}4$, $m_0 = 0.005$

Model with quenched fluctuations

Now we consider the random architecture in which each neuron receives a random number of connections, C *on average* , which give rise to quenched fluctuations.

$$\begin{aligned} h_+ =&\sqrt{C}\{\ J^*_{EE}\, r_- - J_{EI} m_I + E m_0\} + F[a J_{EE} r_+ - J^*_{EE} r_-] \\ h_- =&\sqrt{C}\{\ J^*_{EE}\, r_- - J_{EI} m_I + E m_0\} + F J^*_{EE}[r_+ - r_-] \\ h_I =&\sqrt{C}\{\ J_{IE}\, m_- - J_{II} m_I + I m_0\}\ + F J_{IE}\,[r_+ - r_-] \end{aligned} \tag{4.15}$$

where we defined $J^*_{EE} \equiv J_{EE}[1 + (a-1)\Pi_\infty]$.
The total fluctuations becomes:

$$\begin{aligned} \sigma^2_+ \ = \sigma^2_- =&\ J^2_{EE}(a^2\Pi_\infty + 1 - \Pi_\infty) v_- + J^2_{EI} m_I \\ \sigma^2_I \ =&\ J^2_{IE} m_- + J^2_{II} m_I \end{aligned} \tag{4.16}$$

while for the quenched fluctuations we get

$$\begin{aligned} s^2_+ = s^2_- \ =&\ J^2_{EE}(a^2\Pi_\infty + 1 - \Pi_\infty) p_- + J^2_{EI} q_I \\ s^2_I \ =&\ J^2_{IE} q_- + J^2_{II} q_I \end{aligned} \tag{4.17}$$

and we have the notable fact that, in this limit, the fluctuation in the input to the foreground and the background population are equal. From now on we will indicate those quantities with the notation

$$\sigma_E \equiv \sigma_+ = \sigma_- \qquad\qquad s_E \equiv s_+ = s_-$$

In the balanced state, the term of order $\sqrt{C}$ in the average input must vanish for all populations (see section 2.2), but since the leading contribution in the excitatory part of the input comes solely form the background neurons (in the sparse limit the foreground are a vanishingly small fraction of the excitatory netowork), we only need to balance the background and the inhibitory population as in the case of the network without memory considered in previous chapter:

$$\begin{aligned} m_- =&\ \frac{J^*_{EE} J_{II}}{J_{EI} J_{IE}} \left[r_- + \frac{I}{J^*_{EE}} \left(\frac{E}{I} - \frac{J_{EI}}{J_{II}} \right) m_0 \right] \\ m_I =&\ \frac{J^*_{EE}}{J_{EI}} \left[r_- + \frac{E}{J^*_{EE}}\, m_0 \right] \end{aligned} \tag{4.18}$$

$$\begin{cases} q_- = \int_{-\infty}^{+\infty} \left[H\left(\dfrac{\sigma_E H^{-1}(m_-) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ q_I = \int_{-\infty}^{+\infty} \left[H\left(\dfrac{\sigma_I H^{-1}(m_I) + x s_I}{\sqrt{\sigma_I^2 - s_I^2}} \right) \right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ r_- = \int_{-\infty}^{+\infty} \bar{x}_1 \left(H\left(\dfrac{\sigma_E H^{-1}(m_-) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ v_- = \int_{-\infty}^{+\infty} \bar{x}_1^{(2)} \left(H\left(\dfrac{\sigma_E H^{-1}(m_-) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ p_- = \int_{-\infty}^{+\infty} \left[\bar{x}_1 \left(H\left(\dfrac{\sigma_E H^{-1}(m_-) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \end{cases} \tag{4.19}$$

As in the previous case, the foreground population average rates can then be determined as a solution of the following nonlinear equation for m_+

$$m_+ = H\left(H^{-1}(m_-) - \frac{J_{EE}(a-1)Fe^{-\alpha F^2}}{\sigma_+} \, r_+ \right) \tag{4.20}$$

where we have r_+ which, instead of being an algebraic function of m_+, results from the average over the local variablity of the input:

$$r_+ = \int_{-\infty}^{+\infty} \bar{x}_1 \left(H\left(\frac{\sigma_E H^{-1}(m_+) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}} \right) \right) \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \tag{4.21}$$

which is a function of m_+ alone since σ_E and s_E have been set by the balancing condition of the background and inhibitory population.
Since we are interested in regime of low activity, it is useful to observe that the low rate approximation $m_0 \ll 1$ (see section A.5) of (4.21) gives the expression obtained for a network with fixed number of connection

$$r_+ \longrightarrow \bar{x}_1(m_+) = \frac{(1+\tau_r)m_+}{1+\tau_r+\tau_r U(1+\tau_r m_+)} \tag{4.22}$$

which turn out to be the same function that we get for a network of a fixed number of connections per neurons considered at the beginning of the chapter, in fact at low rates the variability in the number of connections is less prominent because every neuron will have a small fraction of its afferent input which are *active*.

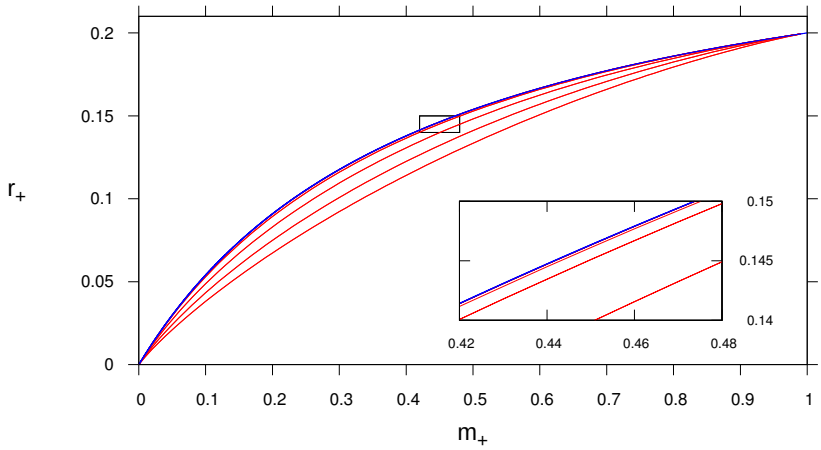

Figure 4.6: The red curves represent the frunction (4.21) calculated for different values of m_0 after solving the system (4.19) with parameters given in table (1.70) and $\tau_r = 7$, $U = 0.6$ for different values of the external input m_0. The curves, from the bottom, corresponds orderly to $m_0 = 0.4; 0.3; 0.2; 0.1; 0.05$. The blue curve is the approximated function (4.22). The inset show a magnification of the area indicated with a box in the main graph.

The graph in figure 4.6 gives a feel of the accuracy of the approximation. This allow us to state that the resulting picture will be essentially the same of the simplified case up to some quantitative correction, thus it's possible to achieve bistability at low rate, although it's likely that region in the parameter space where this occur might be different.

4.4 Numerical Simulations

We solved the MF equations at finite C (4.6) and find the parameters for which bistability at low rate occurred, but simulated networks failed to display bistability for such values. The discrepancies between theory and simulations, most likely, result from significant finite-size effects of the simulated network. In fact, we simulated networks of order $N \sim 10^4$ neurons and $C \sim 10^3$, and to consider the sparse coding limit ($f \ll 1$) we choose $f \sim 0.01 - 0.1$, obtaining selective populations with $Nf \sim 10^3 - 10^4$ neurons on average, each one receiving $fC \sim 10 - 100$ connections on average from other selective neurons. A proper comparison with the theoretical results would require the simulation of networks which are compatible with the assumptions of the mean field theory, namely $fN \gg C \gg 1$ and $f \ll 1$.

Although it wasn't possible to match quantitatively the theoretical predictions, we could find specific parameters for which the expected qualitative phenomenology is reproduced, i.e. with STD, the activity of the foreground neurons in the memory state is far below saturation (fig. 4.8) unlikely the case with static synapses (fig. 4.7).

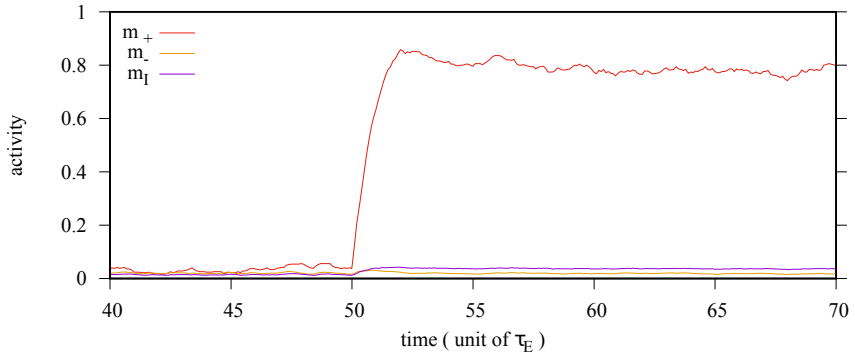

Figure 4.7: Example of a transition to a memory state upon stimulation in absence of synaptic adaptation.

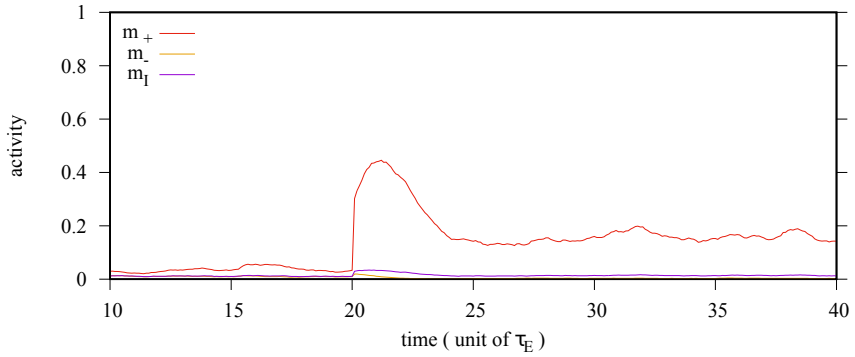

Figure 4.8: Example of transition to a memory state with synaptic depression.

The comparison of the distribution of local activities during the persistent activity states shows how the foreground neurons are all below saturation. Analogously to [21] the distribution is bimodal, and the selective population stands out from a background of much less active neurons.

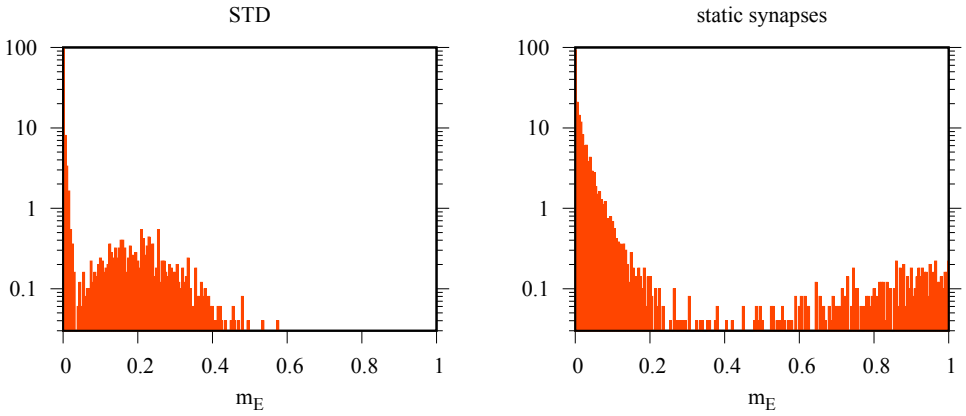

Figure 4.9: Comparison between the distribution of the local activities during memory state for a networ with static synapses (right) and a network with short time depression dynamics (left).

Moreover, we noted that for certain values of the parameters, the activity of the selective neurons displayed collective oscillations (fig. 4.10) with a period of the order of τ_r, suggesting that the stationary states can be not only fixed point attractors but can have a richer dynamics because of the additional timescales introduced by the synaptic adaptation.

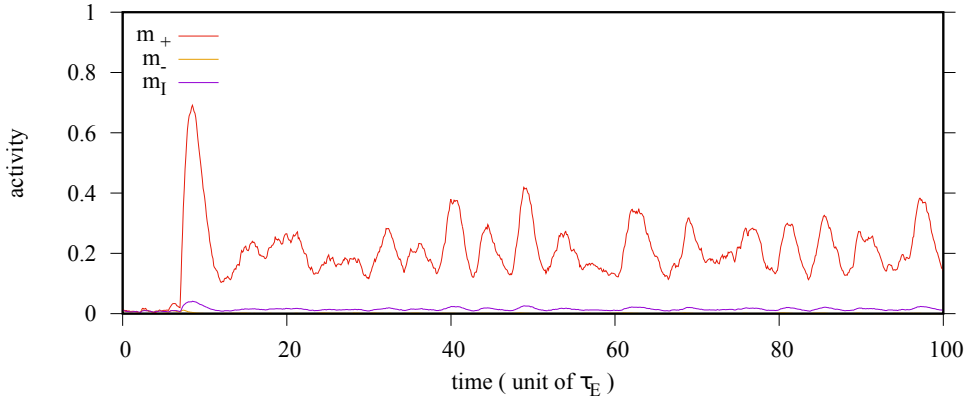

Figure 4.10: Global oscillations in the activity of the selective population of neurons $\tau_r = 5\tau_E$.

Conclusions

The nonlinearity of the single neuron response function allows standard attractor network models to exhibit bistability, provided that the excitatory recurrent feedback is sufficiently strong. This mechanism, as discussed, generically produces an unrealistic high level of activity (close to saturation) for the high rate -memory- state.

We have shown that for a memory network model operating in the balance state (where the distance of the input from threshold is of the same order of the input fluctuation), adding a dynamics to the synaptic couplings between the excitatory neurons indeed allow to have bistability with more plausible activity rates, far below saturation.
To do so, we have formulated a mean field theory for a network of binary neurons operating in the balanced regime with short-time depressing synapses under the assumption that neurons' self correlations were negligible. The analysis of the equilibrium solutions in the $C \to \infty$ limit showed that there are regions in the space of parameters where bistability at between two state at low activity is possible.
Simulations confirmed the qualitative picture, displaying memory state in which activity is low, temporally fluctuating and spatially inhomogeneous.

To achieve memory states at low activity rate, however, the parameters had to be fine tuned, highlighting that synaptic adaptation is an effective solution to achieve more realistic activity levels in the network, but it's a poorly robust one.
This suggest that despite the improvements to the current attractor framework like the one presented here, however successful in explaining the qualitative aspects of self-sustained persistent activity, this framework have some limitations that makes it hard to reconcile it with experimental observations, calling for a re-examination of the mechanism concerning the neural substrates of memory function.

Appendix A

Balanced Networks complementary material

A.1 Input statistics calculation

The expression for the mean and the variance of the input to a cell can also be calculated directly, in the thermodynamic limit.

Mean Input

The population average input is, by definition

$$h_A(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} h_i^A(t) \tag{A.1}$$

for population averaged quantities, the neuron index is dropped and the population index lowered.
Taking the expression of the input (1.5) we have

$$h_A(t) = \mathcal{J}_{AE} \sum_{j=1}^{N} \left[\left[c_{ij}^{AB}\right]\right] S_j^E(t) - \mathcal{J}_{AI} \sum_{j=1}^{N} \left[\left[c_{ij}^{AB}\right]\right] S_j^I(t) + C\mathcal{J}_{Aex} m_0 \tag{A.2}$$

where we use the double square bracket to indicate the population average

$$[[...]] \equiv \frac{1}{N} \sum_{i=1}^{N} ...$$

thus we have to take the average over the postsynaptic (output) connectivity for each neuron j. In general, this quantity might depend on the cell j, and this would imply that the activity S_j of some neuron in the sums appearing in A.3 are summed with different weights, corresponding to the fact that neuron i receives contact from the other neurons with different probabilities. In our case the network structuring is random (1.2) and in the large N limit:

$$\left[\left[c_{ij}^{AB}\right]\right] = \frac{C}{N}$$

then we can take this terms out of the summation over j and we get

$$h_A(t) = C\left\{ \mathcal{J}_{AE}\left[\frac{1}{N}\sum_{j=1}^{N} S_j^E(t)\right] - \mathcal{J}_{AI}\left[\frac{1}{N}\sum_{j=1}^{N} S_j^I(t)\right] + \mathcal{J}_{Aex} m_0 \right\} \quad \text{(A.3)}$$

where the term in square brackets are the population average activities

$$m_B(t) \equiv \frac{1}{N}\sum_{j=1}^{N} S_j^B(t) \quad \text{(A.4)}$$

and the average input in population A is

$$\boxed{h_A(t) = C\Big\{ \mathcal{J}_{AE} m_E(t) - \mathcal{J}_{AI} m_I(t) + \mathcal{J}_{Aex} m_0 \Big\}} \quad \text{(A.5)}$$

Input Variance

The variance of the input is defined as

$$\sigma_A^2(t) \equiv \mathrm{Var}[h_i^A(t)] = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \Big[\, \delta h_i^A(t) \,\Big]^2 \quad \text{(A.6)}$$

where we have that the input variation is

$$\begin{aligned} \delta h_i^A(t) =&\ h_i^A(t) - h_A(t) \\ =&\ \mathcal{J}_{AE}\sum_{j=1}^{N}\left(c_{ij}^{AE} - \frac{C}{N}\right) S_j^E(t) - \mathcal{J}_{AI}\sum_{j=1}^{N}\left(c_{ij}^{AI} - \frac{C}{N}\right) S_j^I(t) \end{aligned} \quad \text{(A.7)}$$

In general, if we have N uncorrelated random variables X_j, the variance of their sum is equal to the sum of their variance

$$\mathrm{Var}\left[\sum_{j=1}^{N} X_j\right] = \sum_{j=1}^{N} \mathrm{Var}\,[\, X_j \,] \quad \text{(A.8)}$$

In our case we can regard the recurrent part of the input to a cell at a certain time t as a sum of random variables. The external input is deterministic.

$$h_i^A(t) = \sum_{B}^{E,I}\sum_{j=1}^{N} X_{ij}^{AB} + \mathcal{J}_{Aex} m_0 \quad \text{(A.9)}$$

$$\text{where} \qquad X_{ij}^{AB} = (-1)^{\delta_{BI}} \mathcal{J}_{AB}\, c_{ij}^{AB} S_j^B(t)$$

the variables X_{ij} are certainly uncorrelated because for any neuron j, post-synaptic connections are independent random variables (1.2) and we can carry out the calculation of (A.6) using (A.8). We have that

$$\mathrm{Var}\left[\, X_{ij}^{AB} \,\right] = \left[\left[\left(X_{ij}^{AB}\right)^2\right]\right] - \left(\, \left[\left[X_{ij}^{AB}\right]\right] \right)^2 \quad \text{(A.10)}$$

The calculation of the mean of X_{ij}^{AB} in the large N limit gives

$$\left[\left[X_{ij}^{AB}\right]\right] = \frac{1}{N}\sum_{i=1}^{N} X_{ij}^{AB} = \frac{C}{N}(-1)^{\delta_{BI}}\mathcal{J}_{AB}S_j^B(t) \tag{A.11}$$

For the mean of the square we observe that both c_{ij} and S_j are binary random variables, therefore they remain unchanged when squared and we get

$$\left[\left[\left(X_{ij}^{AB}\right)^2\right]\right] = \frac{1}{N}\sum_{i=1}^{N}\left(X_{ij}^{AB}\right)^2 = \frac{C}{N}\mathcal{J}_{AB}^2 S_j^B(t) \tag{A.12}$$

Taking (A.12) and subtracting the square of (A.11) we find that the variance of the j^{th} term of the summation

$$\mathrm{Var}\left[\ X_{ij}^{AB}\ \right] = \frac{C}{N}\left(1-\frac{C}{N}\right)\mathcal{J}_{AB}^2 S_j^B(t) \simeq \frac{C}{N}\mathcal{J}_{AB}^2 S_j^B(t)$$

where the approximation holds in the limit of sparse connectivity $C \ll N$. For the full expression of the input variance we have

$$\sigma_A^2(t) = \sum_{B}^{E,I}\sum_{j=1}^{N}\mathrm{Var}\left[X_{ij}^{AB}\right] = \sum_{B}^{E,I} C\mathcal{J}_{AB}^2\left[\frac{1}{N}\sum_{j=1}^{N} S_j^B(t)\right] \tag{A.13}$$

where we recognize the average activities in the last term in square brackets. Therefore, the total variance for the input to population A reads

$$\boxed{\sigma_A^2(t) = C\{\mathcal{J}_{AE}^2 m_E(t) + \mathcal{J}_{AI}^2 m_I(t)\}} \tag{A.14}$$

A.2 Local stability of the Balanced State

The stability of the balanced state is studied. Local stability requires that a sufficiently small perturbation in the population will decay to zero. If (m_E, m_I) is the balanced solution of (1.26) we can write the perturbed solution as

$$m_A(t) = m_A + \delta m_A(t)$$

the perturbation in the rates affect the equilibrium field and the variance

$$\begin{aligned} h_A(t) &= h + \delta h(t) \\ \alpha_A(t) &= \alpha_A + \delta\alpha(t) \end{aligned} \quad \text{where} \quad \begin{aligned} \delta h_A(t) &= \sqrt{C}\left[J_{AE}\delta m_E(t) + J_{AI}\delta m_I(t)\right] \\ \delta\alpha_A(t) &= J_{AE}^2\delta m_E(t) + J_{AI}^2\delta m_I(t) \end{aligned}$$

The dynamical equations (1.26) can be linearized if the perturbation in the field is much smaller than the equilibrium field:

$$H\left(\frac{\theta_A - h_A - \delta h_A}{\sqrt{\alpha_A + \delta\alpha_A}}\right) \simeq H\left(\frac{\theta_A - h_A}{\sqrt{\alpha_A}}\right) + \delta m_E\frac{\partial H_A}{\partial m_E} + \delta m_I\frac{\partial H_A}{\partial m_I}$$

Since in the balanced state the equilibrium field is O(1), the linear approximation holds if

$$\delta h_A(t) \ll 1 \quad \Rightarrow \quad \delta m_A(t) \ll \frac{1}{\sqrt{C}}$$

Therefore, if we consider perturbations in the rates that are much smaller than $1/\sqrt{C}$ the linearized equation for the perturbed rates has the form

$$\frac{d}{dt}\begin{pmatrix}\delta m_E(t)\\ \delta m_I(t)\end{pmatrix} = \begin{pmatrix}\frac{1}{\tau_E}\frac{\partial H_E}{\partial m_E} - \frac{1}{\tau_E} & \frac{1}{\tau_E}\frac{\partial H_E}{\partial m_I}\\ \frac{1}{\tau_I}\frac{\partial H_I}{\partial m_E} & \frac{1}{\tau_I}\frac{\partial H_I}{\partial m_I} - \frac{1}{\tau_I}\end{pmatrix}\begin{pmatrix}\delta m_E(t)\\ \delta m_I(t)\end{pmatrix} \tag{A.15}$$

whose solutions are of the form:

$$\delta m_A(t) = c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$$

where λ_1, λ_2 are the eigenvalues of the matrix appearing on the r.h.s of A.15. The equilibrium solution are of locally stable if the perturbations drops to zero, therefore if the eigenvalues have a negative real part. Calculating the derivatives of the H_A function we get

$$\frac{\partial H_A}{\partial m_B} = \frac{1}{\sqrt{2\pi}}\ \exp\left[-\frac{(\theta_A - h_A)^2}{2\ \alpha_A}\right]\left(\sqrt{C}\frac{J_{AB}}{\sqrt{\alpha_A}} - \frac{J_{AB}^2 h_A}{2\ \alpha_A^{3/2}}\right)$$

In the large connectivity limit $C \gg 1$ the first term in round brackets is much larger and the second term can be neglected.

$$\frac{\partial H_A}{\partial m_B} \quad \underset{C\gg 1}{\simeq} \quad \sqrt{C}\ \frac{1}{\sqrt{2\pi\alpha_A}}\ \exp\left[-\frac{(\theta_A - h_A)^2}{2\ \alpha_A}\right] J_{AB} \tag{A.16}$$

If we find the roots of the characteristic polynomial and require that their real part be negative yields a condition on the timescales of the form

$$\frac{\tau_I}{\tau_E} < \tau^*(m_E, m_I, J)$$

In the $C \to \infty$ the timescale τ^* depends only on the external input m_0 and the coupling parameters J. Since the eigenvalues are both of order C, small perturbations will decay with an extremely short time constant, of order $1/\sqrt{C}$.

A.3 Partially frozen state

We are interested in understanding under which conditions and regimes of activity some of the neurons of the network remains "silent", i.e. the (fluctuating) input they receives *always* remain under threshold thus preventing the transition of the neuron to the active state. In other words: what should happen in order to have a *finite* subpopulation of quiescent neurons in the $N \to \infty$ limit?
The question can be answered if we look at the analytic form of the distribution of activity across a population (1.66), this function provides us with the following information: the fraction of neurons (in population A) whose time-averaged activity lies between 0 and m is given by

$$\int_0^m \rho_A(\mu) d\mu$$

The shape of the probability density function near 0 has to be studied. We recall that the population distribution of activity is a probability density function defined in the interval $(0:1)$

$$\rho_A(m) = \frac{1}{B} e^{(-x^2+(C_A+B_A x)^2)/2}\Big|_{x=x(m)}$$

$x(m)$ is the inverse function of $m(x) = H(C + Bx)$ where $H(z)$ is given by (1.60), C and B are defined in (1.67). To calculate $\rho_A(0)$, the value of $x(m = 0)$ has to be determined first: it is the x that fulfil

$$m(x) = H(C + xB) = \int_{C+xB}^{\infty} \frac{e^{x^2/2}}{\sqrt{2\pi}} = 0$$

The gaussian integral $H(C + Bx)$ is an always positive and monotonous function of x for every C, if $B > 0$ (always true by definition) is decreasing and the following limits holds:

$$\lim_{x\to-\infty} H(C + Bx) = 1 \qquad \lim_{x\to+\infty} H(C + Bx) = \lim_{x\to+\infty} m(x) = 0$$

therefore we can figure out the value of the density distribution function at the lower and upper extremes of its domain:

$$\rho_A(0) = \lim_{x\to+\infty} \rho_A(m(x)) = \lim_{x\to+\infty} \frac{1}{B_A} e^{[x^2(B_A^2-1)+x2B_A C_A+C_A^2]/2}$$
$$\rho_A(1) = \lim_{x\to-\infty} \rho_A(m(x)) = \lim_{x\to-\infty} \frac{1}{B_A} e^{[x^2(B_A^2-1)+x2B_A C_A+C_A^2]/2}$$

If $B_A \neq 1$ the exponent is a quadratic polynomial in x, therefore the two limits coincide and thus the values of $\rho(m)$ at the extremes of the interval are always equal. There are only two possibilities:

$$B_A^2 > 1 \quad \Rightarrow \quad \rho_A(0) = \rho_A(1) = \lim_{x\to\infty} \frac{1}{B_A} e^{[x^2(B_A^2-1)+x2B_A C_A+C_A^2]/2} = +\infty$$
$$B_A^2 < 1 \quad \Rightarrow \quad \rho_A(0) = \rho_A(1) = \lim_{x\to\infty} \frac{1}{B_A} e^{[x^2(B_A^2-1)+x2B_A C_A+C_A^2]/2} = 0$$

When $B = 1$ the exponent becomes linear and distribution reduces to:

$$\rho(x) = \frac{1}{B_A} e^{[x 2 B_A C_A + C_A^2]/2}$$

and the values at the extremes of the interval are no longer coincident, they can be:

$$\begin{array}{lll} \rho(0) = +\infty & \rho(1) = 0 & \text{if } A > 0 \\ \rho(0) = 0 & \rho(1) = +\infty & \text{if } A < 0 \end{array}$$

Thus, excluding the case $B = 1$ it has been demonstrated that depending on the value of B, the probability density function can either drop to zero at the extremum of its domain, or diverge[1].
To determine the conditions of occurrence of this transition the following quantity is defined

$$V_A = (B^2 - 1) = \frac{2\beta_A - \alpha_A}{\alpha_A - \beta_A} = \frac{q_E J_{AE}^2 + q_I J_{AI}^2}{(m_E - q_E) J_{AE}^2 + (m_I - q_I) J_{AI}^2} - 1 \tag{A.17}$$

When $V_A < 0$ the distribution goes to zero at the extremes, otherwise when $V_A > 0$ the divergences appear. The shape of the distribution, as well as V_A, depend only on m and q which are solutions of the MF equations for a given value of the external input m_0, therefore either the distribution and V_A depend -implicitly- on m_0 only (provided that all the structural parameters of the network are fixed).

Since $\alpha - \beta$ is always positive because of the constrain on q_A, we observe that the condition to have a bimodal distribution $V_A > 0$ can be stated as

$$\beta_A > \alpha_A - \beta_A \tag{A.18}$$

which means that the transition occurs when the amplitude of the quenched fluctuations of the input becomes greater than the amplitude of the temporal fluctuations.

Observation 1. *For any value of the parameters defining the network, the following limit holds:*

$$\lim_{m_0 \to 0^+} V_A = -1$$

Proof. Recalling the approximation of q for small m_0 given in (1.67) we can write V_A as a function of m_A only. Assuming that m_A goes to zero linearly with m_0, the fraction in the definition (A.17) vanishes. □

Figure A.1 shows how V_A depends on m_0 in both population, notice the "universal" limit $V_A(0) = -1$ and the point where V_A vanish which signal the transition to a "phase" in which the population distribution diverges at

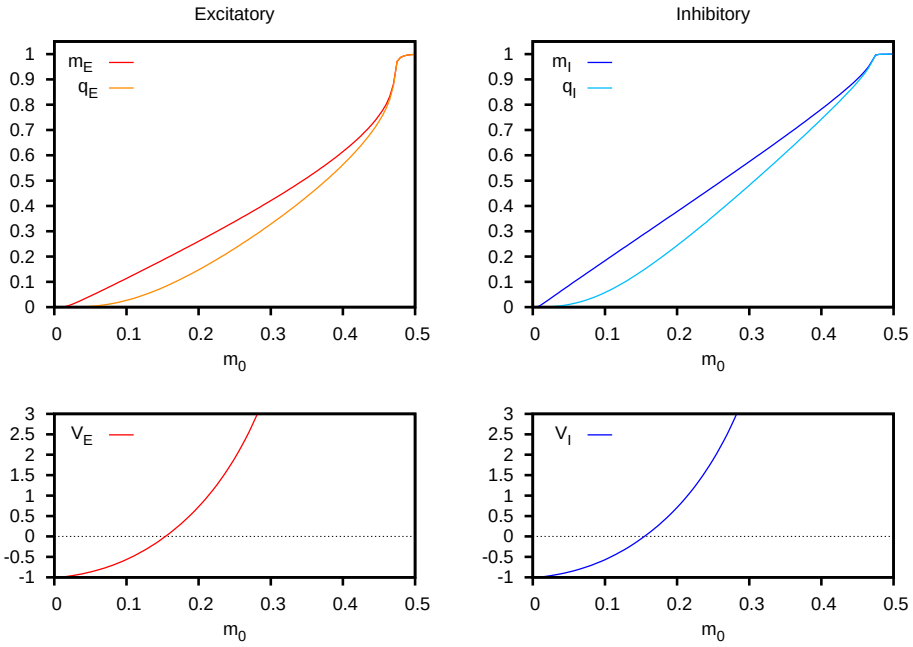

Figure A.1: From the definition (A.17) V depend explicitly on m and q which in turn are functions of m_0 (upper panels), therefore V is implicitly a function of m_0 (lower panels). Network parameters are those given in section **??**.

the extremes and macroscopic subpopulation of neurons at extremely low frequencies can be seen.

The existence of stationary regimes in which some of the neurons in the network remains silent ($m_i = 0$) and some others saturates ($m_i = 1$) has an interesting consequence on the dynamics of the system (see appendix **??**). Since those neurons does not change states along the dynamical evolution, they will be called *frozen neurons*.

Observation 2. *Above the m_0 for which the phase transition occurs, the system is non-ergodic.*

Proof. Let's assume that in the stationary state neuron $i = 1$ is silent, this mean that during the evolution the network will be confined in that subspace of the phase-space which correspond to the set of all the microstates for which $m_{1_A} = 0$, i.e. any transition to a states of the complementary subset with $m_{1_A} = 1$ is forbidden. □

[1]The fact that the integral of the function over the whole integral is always 1 ensure that the divergences are integrable and any improper integral with 0 or 1 as endpoint will converge to a finite value.

Therefore, the presence of silent (and saturated) neurons constrain the system to move in a subset of all the available states in the phase space; the higher the external input, the more frozen neurons appears, and the smaller the allowed phase-space gets. The set of non-forbidden microstates degenerate to a single state-point when the input becomes so high that all neurons are saturated.

In this situation, the dynamical scenario is the following: regardless the initial microstate, the system will evolve towards the region of the phase space which is compatible with the presence of frozen neuron and will be confined there; restricted to this subspace the system is ergodic, and the statistical distribution will vanish everywhere outside this region. Ergodicity breaking produces a shrinking of the set of available states [2].

A network with $N_E = N_I = 10000$ and $C = 1000$ and the usual parameters was simulated and *local* temporally averaged activities m_{i_A} have been calculated and represented on a 2-dimensional plot to give an explicit representation of a partially frozen phase. The duration of the simulation was set to $1000\tau_E$ as usual.
Two simulations were run with two different m_0, one below ($m_0 = 0.1$) the transition point and one below ($m_0 = 0.4$); the former corresponds to figures A.2 , the latter is represented in figures A.3.
We observe that the presence of a few silent (dark) neurons in the low-input simulation can be an effect of the simulation time length that imposes a cutoff on the lowest frequencies. Instead, for the high-input simulation, the appearance of clusters of frozen neurons is an intrinsic feature of the dynamics.

[2] instead of partitioning the phase-space into disconnected region corresponding to different stationary distribution as it happen, for instance, with the two available magnetization of a ferromagnetic Ising model.

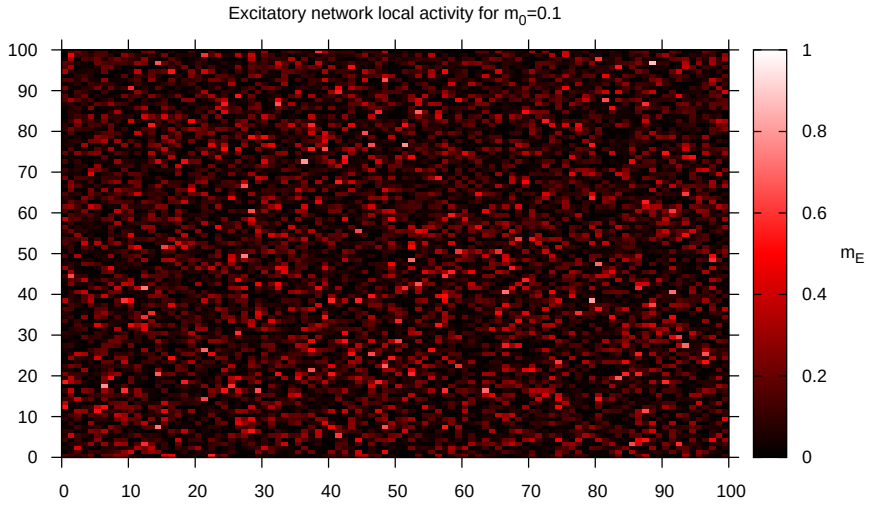

Figure A.2: 2D plot of the local activity of an excitatory neurons network with $N_E = 10000$ and external input $m_0 = 0.1$. There are 83 silent neurons (black pixels) and no saturated neurons (white pixels). Confront with the distribution in figure **??**.

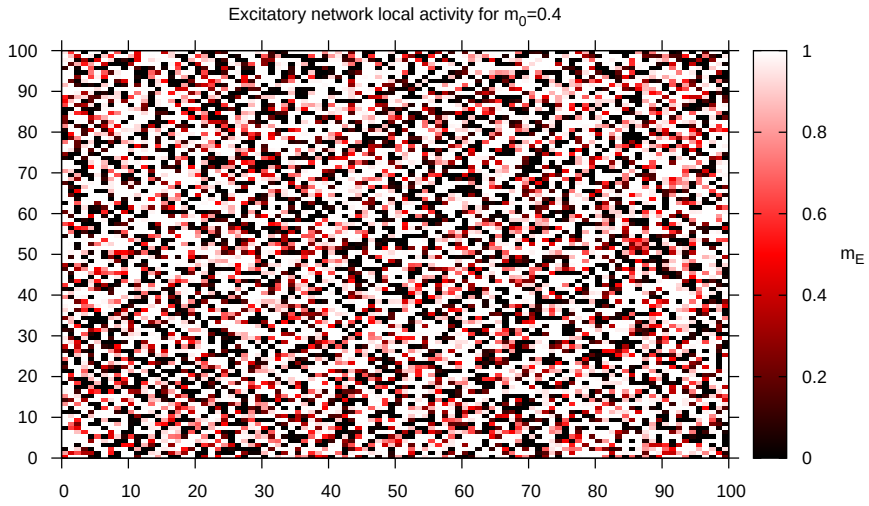

Figure A.3: 2D plot of the local activity of an excitatory neurons network with $N_E = 10000$ and external input $m_0 = 0.4$. There are 1547 silent neurons (black pixels) and 2957 saturated neurons (white pixels). Confront with the distribution in figure **??**.

A.4 Invariance transformations for the equilibrium solution

We define the following transformation for the synaptic couplings

$$\begin{aligned} J_{EA} &\rightarrow \gamma_E J_{EA} + \lambda_E J_{IA} & \qquad J_{IA} &\rightarrow \gamma_I J_{IA} + \lambda_I J_{EA} \\ J_{Eex} &\rightarrow \gamma_E J_{Eex} + \lambda_E J_{Iex} & \qquad J_{Iex} &\rightarrow \gamma_I J_{Iex} + \lambda_I J_{Eex} \end{aligned} \tag{A.19}$$

with $\gamma_A, \lambda_A > 0$. It is easy to see that if coupling constants are transformed according to (A.19), and the inequalities (1.47) are still respected, the population average rates in the balanced state (1.39) are unchanged.
Moreover, it can be proved that the q_A solution of (**??**) in the $C \to \infty$ are invariant under the transformation (A.19) with $\lambda_E = \lambda_I = 0$. In particular, this implies that in the $C \to \infty$ the population distribution of activity is invariant under all the transformation with $\lambda_E = \lambda_I = 0$.

Observation 3. *In the $C \to \infty$ limit, the solution of the mean field equation q_A are invariant under the scaling transformation (A.19) for $\lambda_E = \lambda_I = 0$ and $\gamma_{E,I} > 0$.*

Proof. It's very easy to see that the average activities m_A are invariant under the transformation: the solutions (**??**, **??**) do not change if all the coupling coefficients are multiplied for a positive number γ.
To demonstrate that q_A^∞ is invariant as well we should demonstrate that if equation (**??**) is transformed under scaling, its solution coincide with the unscaled one: $\tilde{q}_A^\infty = q_A^\infty$. First, we must figure out the transformation that the scaling induces on the quantities appearing in (**??**): for the input variance we get

$$\alpha_A^\infty \quad \longrightarrow \quad \tilde{\alpha}_A^\infty = \gamma^2 \alpha_A^\infty \tag{A.20}$$

To derive how the input h_A^∞ is transformed, we look for a solution of the transformed MF equation (1.42): recalling that m_A^∞ is unchanged under scaling we have

$$m_0 \Omega_A = H\left(\frac{\theta_A - \tilde{h}_A^\infty}{\sqrt{\tilde{\alpha}_A^\infty}}\right) = H\left(\frac{\theta_A - \tilde{h}_A^\infty}{\gamma\sqrt{\alpha_A^\infty}}\right) = H\left(\frac{\theta_A - [\theta_A + \frac{\tilde{h}_A^\infty - \theta_A}{\gamma}]}{\sqrt{\alpha_A^\infty}}\right)$$

the above identity is fulfilled if the unscaled value of the input is equal to the term in square bracket that can be inverted to find the sought transformation:

$$h_A^\infty \quad \longrightarrow \quad \tilde{h}_A^\infty = \theta_A + \gamma(h_A^\infty - \theta) \tag{A.21}$$

Last, the transformation induced on β is

$$\beta_A^\infty \quad \longrightarrow \quad \gamma^2 \left(J_{AE}^2 \tilde{q}_E^\infty + J_{AI}^2 \tilde{q_I}^\infty \right) \equiv \gamma^2 \tilde{\beta}_A^\infty \tag{A.22}$$

Using (A.20, A.21, A.22) we can write the transformed argument of the H function that appears the integrand of the equations of the system (**??**)

$$\frac{\theta_E - h_E^\infty + x\sqrt{\beta_E^\infty}}{\sqrt{\alpha_E^\infty - \beta_E^\infty}} \quad \longrightarrow \quad \frac{\gamma(h_A^\infty - \theta_A) + x\gamma\sqrt{\tilde{\beta}_A^\infty}}{\gamma\sqrt{\alpha_A^\infty - \tilde{\beta}_A^\infty}} \tag{A.23}$$

simplifying the γ we see that the argument of H remains unchanged as a function of the transformed variables, therefore the whole system of equations for the transformed variables $\tilde{q}_A^\infty$ keeps its form therefore, since it has a unique solution, for every $\gamma > 0$ we have $\tilde{q}_A^\infty = q_A^\infty$ □

Observation 4. *Invariance of the solutions of the MF equations under transformation (A.19) does not hold for finite C.*

Proof. It suffice to se that invariance changes the form of equations (1.61). In fact, the scaling induces the following transformation at the r.h.s. of the mean field equations:

$$\frac{\theta_A - h_A}{\sqrt{\alpha_A}} \quad \to \quad \frac{\theta_A - \gamma\, h_A}{\gamma\sqrt{\alpha_A}}$$

thus shifting the lower extreme of the gaussian integral and yielding a different r.h.s. term. Therefore, fixing m_0 and C finite, for each γ there is a solutions m_E and m_I □

Figure A.4 gives an example of how this invariance is lost for finite C: the curve $m_A(m_0)$ have been determined numerically for some values of the scaling factor γ_E. The degeneracy of the single solution for the $C \to \infty$ limit into a continuous spectrum at finite C is associated to the scaling symmetry breaking.

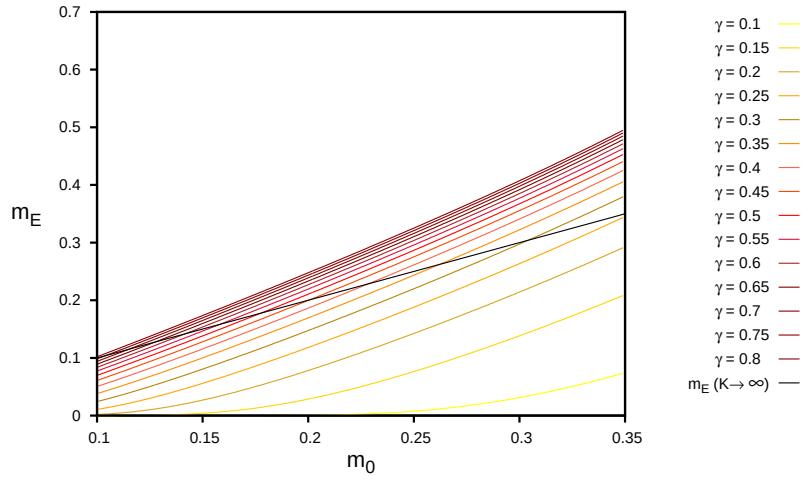

Figure A.4: Mean activity of the **excitatory** population as a function of the input m_0 for different values of the scaling parameter γ defined in (A.19). The unscaled couplings ($\gamma = 1$) are those given in (1.70) and $C = 1000$. *Black line*: $C \to \infty$ limit.

A.5 Low rate approximations

When $m_0 \ll 1$ the population average rates will also be very low $m \ll 1$ since the mean input is largely below threshold $(\theta - h)/\sigma \gg 1$. Recalling the definition (1.60) and considering the asymptotic expansion for $x \gg 1$

$$\mathrm{erfc}(x) = \frac{e^{-x^2}}{\sqrt{x\pi}} \left[1 + \sum_{n=1}^{\infty} (-1)^n \frac{(2n-1)!!}{(2x^2)^n} \right]$$

we can take the following approximation

$$m = H\left(\frac{\theta - h}{\sigma}\right) \simeq \frac{\sigma}{(\theta - h)\sqrt{2\pi}}\, e^{-\frac{(\theta-h)^2}{2\sigma^2}}$$

from which we can find an approximation for residual input:

$$\log(m) = -\log\left(\frac{\theta - h}{\sigma}\sqrt{2\pi}\right) - \frac{(\theta-h)^2}{2\sigma^2} \;\simeq\; \frac{(\theta-h)^2}{2\sigma^2} \quad \Rightarrow$$

$$\theta - h \;\simeq\; \sigma\sqrt{2|\log(m)|} \tag{A.24}$$

Quenched noise reduction at low m_0

When the external input m_0 is very low, the MF equations for the network with random connectivity are well approximated by the MF equations of the network in which the number of connections per neuron are fixed.
If $m_0 \ll 1$ the average rates in both populations will also be very low $m_{E,I} \ll 1$ and the quenched fluctuations will be much smaller than the temporal fluctuations

$$\frac{s_E}{\sqrt{\sigma_E^2 - s_E^2}} \ll 1 \qquad\qquad \frac{\sigma_E}{\sqrt{\sigma_E^2 - s_E^2}} \simeq 1$$

and we can we can make the approximation

$$H\left(\frac{\sigma_E H^{-1}(m_E) + x s_E}{\sqrt{\sigma_E^2 - s_E^2}}\right) \;\approx\; m_E \;+\; x\, O\left(\frac{s_E}{\sigma_E}\right)$$

The validity of this approximation reflect the fact that at very low rate, the population distribution of activity is very narrow and it's peaked around the average value. When m_0 increases, the population distribution of activity becomes narrower and increasingly skewed, making the approximation too crude.

Appendix B

Bistability between balanced and partially unbalanced states in a memory network

In the attempt to see if an attractor memory network model could support multistability between a background and memory state at realistically low level of activity, we studied whether a binary network model with memory structure, instead of having multiple *balanced* stable states, could have bistability between a spontaneous *balanced* state and a memory *partially balanced* state. It turned out that this scenario cannot be realized because the condition for stability of the partially unbalanced state excluded the stability of the balanced one.

B.1 The Model

The network is made up of two population of neurons: excitatory (E) and inhibitory (I) composed respectively by N_E and N_I units [1].
The pattern of connections among neurons is random and is realzed

Each neuron (independently on the population to which it belongs) receives an input from the generic neuron of population B with probability C/N_B and if a synapse $j_B \to i_A$ exists, its strength will be proportional to the populations coupling strength J_{AB}.
Formally, the elements of the connection matrix $J_{i_A j_B}$ are independent random variable identically distributed according to the density:

$$P(J_{i_A j_B}) = \frac{C}{N_A}\delta(J_{i_A j_B} - \frac{J_{AB}}{\sqrt{C}}) + (1 - \frac{C}{N_A})\delta(J_{i_A j_B})$$

For the excitatory network only, a finite number P of *independent* mem-

[1] In the following, every quantity Q relative to the i^{th} neuron of population A will be indicated O_i^A

ory patterns $\{\xi_i^\mu\}$ with coding level f are considered

$$\begin{matrix}\forall i = 1, ..., N \\ \forall \mu = 1, ..., P\end{matrix} \qquad \xi_i^\mu = \begin{cases} 1 & \text{with probability } f \\ 0 & \text{with probability } 1-f \end{cases}$$

The input to a generic excitatory and inhibitory neuron are respectively

$$h_i^E(t) = \sum_{j=1}^{N_E} c_{ij}^{EE} W_{ij} S_j^E(t) + \frac{J_{EI}}{\sqrt{C}} \sum_{j=1}^{N_I} c_{ij}^{EI} S_j^I(t) + m_0 \sqrt{C} J_{Eex} \tag{B.1a}$$

$$h_i^I(t) = \frac{J_{IE}}{\sqrt{C}} \sum_{j=1}^{N_E} c_{ij}^{IE} S_j^E(t) + \frac{J_{II}}{\sqrt{C}} \sum_{j=1}^{N_I} c_{ij}^{II} S_j^I(t) + m_0 \sqrt{C} J_{Iex} \tag{B.1b}$$

Where the recurrent excitatory synaptic matrix M_{ij} contains the information about the memories and is given by

$$W_{ij} = \left[\frac{J_{EE}}{\sqrt{C}} + \frac{\beta}{\sqrt{C} f(1-f)} \sum_{\mu=1}^{P} \xi_i^\mu (\xi_j^\mu - f) \right]$$

The strength of a single synapse in the excitatory subnetwork varies within

$$\frac{1}{\sqrt{C}} \left(J_{EE} - P \frac{\beta}{1-f} \right) \; < \; J_{ij}^{EE} \; < \; \frac{1}{\sqrt{C}} \left(J_{EE} + P \, \frac{\beta}{f} \right)$$

to make sure that recurrent excitatory synapse are positive we require that:

$$\boxed{P\beta \; < \; J_{EE}(1-f)} \tag{B.2}$$

B.2 Mean field

We consider only the excitatory recurrent part of the instantaneous input to the i^{th} excitatory neuron:

$$h_i^{EE}(t) = \sum_{j=1}^{N} c_{ij}^{EE} \left[\frac{J_{EE}}{\sqrt{C}} + \frac{\beta}{\sqrt{C} f(1-f)} \sum_{\mu=1}^{P} \xi_i^\mu (\xi_j^\mu - f) \right] S_j^E(t) \tag{B.3}$$

where the first term in the square brackets is the contribution of the unstructured part while the second term represent the synaptic structuring due to the memories.
Defining the time-averaged local activity as $m_j \equiv \langle S_i(t) \rangle$, the linearity of the time averaging operation enable us to find immediately the time average of the local input $h_i = \langle h_i(t) \rangle$ just replacing $S_j(t)$ with m_j.

We introduce the following quantities which will serve as order parameters:

Definition 1 (*Overlap with pattern ν*)**.**

$$m^{\nu}(t) \equiv \frac{1}{f(1-f)N} \sum_{i=1}^{N} (\xi_i^{\nu} - f) S_i(t)$$

Definition 2 (*Average Excitatory activity*)**.**

$$m_E(t) \equiv \frac{1}{N} \sum_{i=1}^{N} S_i(t)$$

Selective and nonselective populations

For each pattern ξ^{ν} the excitatory network can be partitioned in two populations:

- Selective (or foreground) neurons: cells for which $\xi_i^{\nu} = 1$
- Nonselective (or background) neurons: cells for which $\xi_i^{\nu} = 0$

It is naturally to define the following quantities:

Definition 3 (*Average activity of the pattern ν selective/nonselective population*)**.**

$$m_+^{\nu}(t) \equiv \frac{1}{fN} \sum_{i=1}^{N} \xi_i^{\nu} S_i(t) \qquad m_-^{\nu}(t) \equiv \frac{1}{(1-f)N} \sum_{i=1}^{N} (1 - \xi_i^{\nu}) S_i(t)$$

Those quantities are connected to the overlaps and to m_E by the following relations

$$m_E = f m_+^{\nu} + (1-f) m_-^{\nu} \qquad \forall\, \nu = 1...P \tag{B.4}$$

$$m^{\nu} = \frac{m_+^{\nu} - m_E}{1-f} = m_+^{\nu} - m_-^{\nu} \qquad \forall\, \nu = 1...P \tag{B.5}$$

Therefore for a macroscopic characterization of the state of the system, the set of the P overlaps and m_E is equivalent to a description in terms of all m_+ and one of the m_-. For our purpose it will turn out to be more convenient to use this last set of order parameters.

The order parameters have been defined in a time-dependent manner, the corresponding time-averaged ones are readily found simply substituting $S_i(t)$ with its time averaged value m_i.

Average inputs

If f is finite, in the $N \to \infty$ the selective and nonselective population are macroscopic population of neurons that can be studied with the mean field formalism. We define and calculate the following quantities:

Definition 4 (*Average input to pattern ν selective/nonselective population*)**.**

$$h_+^\nu(t) \equiv \frac{1}{fN}\sum_{i=1}^{N}\xi_i^\nu h_i(t) \qquad h_-^\nu(t) \equiv \frac{1}{(1-f)N}\sum_{i=1}^{N}(1-\xi_i^\nu)h_i(t)$$

If we consider only the structured part of the local field h_i (eq. B.3) and substitute it in the definition of h_+^ν given above we get:

$$\tilde{h}_+^\nu(t) = \frac{1}{fN}\sum_{i=1}^{N}\xi_i^\nu\sum_{j=1}^{N}c_{ij}\frac{\beta}{\sqrt{C}f(1-f)}\sum_{\mu=1}^{P}\xi_i^\mu(\xi_j^\mu - f)S_j(t)$$

Collecting together all the terms that depend on the index i we have:

$$\tilde{h}_+^\nu = \frac{\beta}{\sqrt{C}f^2(1-f)}\sum_{\mu=1}^{P}\sum_{j=1}^{N}(\xi_j^\mu - f)S_j(t)\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i^\nu\xi_i^\mu c_{ij}\right] \tag{B.6}$$

The term in square brackets is nothing but the definition of spatial average (over site i) of the random quantity $A_{ij}^{\mu\nu} \equiv \xi_i^\nu\xi_i^\mu c_{ij}$ for a given realization of the network.

> In general, if the $A_i(x)$ are N outcome of the same random variable $A(x)$ labeled by index i (and x indicates all the other parameters the random variable may depends on), the $N \to \infty$ limit of the spatial average over a particular realization corresponds to the expectation value of the random variable.
>
> $$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}A_i(x) = E[A(x)]$$

In our case $A_{ij}^{\mu\nu}$ is the product of three random variables, each of which are identically distributed for every i and j, therefore we can conclude that $A_{ij}^{\mu\nu}$ for $i, j = 1...N$ are the outcome of the random variable $A^{\mu\nu} \equiv \xi^\mu\xi^\nu c$ where the probability density of ξ^μ ($\mu = 1...P$) and c are respectively

$$P(\xi) = f\delta(1-\xi) + (1-f)\delta(\xi)$$
$$P(c) = \frac{C}{N}\delta(1-c) + (1-\frac{C}{N})\delta(c)$$

In our model there is clearly no correlation between the memory patterns ξ and the architecture of connections c, therefore recalling that if X and Y are uncorrelated random variables $E[XY] = E[X]E[Y]$ we have

$$E[\xi^\nu\xi^\mu c] = E[\xi^\nu\xi^\mu]E[c] = E[\xi^\nu\xi^\mu]\frac{C}{N}$$

If $\mu \neq \nu$ we have that ξ^μ and ξ^ν are independent and $E[\xi^\nu \xi^\mu] = E[\xi]^2 = f^2$ if otherwise $\mu = \nu$ we get $E[\xi^2] = E[\xi] = f$.
Putting together all the pieces we have

$$E[\xi^\nu \xi^\mu c] = f[f(1-\delta_{\mu\nu}) + \delta_{\mu\nu}]\frac{C}{N}$$

that can be substituted in the expression (B.6) yielding:

$$\tilde{h}_+^\nu(t) = \frac{\beta\sqrt{C}}{f(1-f)N}\sum_{\mu=1}^{P}[f(1-\delta_{\mu\nu}) + \delta_{\mu\nu}]\sum_{j=1}^{N}(\xi_j^\mu - f)S_j(t)$$

recognizing the expression that define the overlaps we can write

$$\tilde{h}_+^\nu(t) = \beta\sqrt{C}\sum_{\mu=1}^{P}[f(1-\delta_{\mu\nu}) + \delta_{\mu\nu}]m^\mu(t) \tag{B.7}$$

eventually, the expression we get after summing over the P patterns is

$$\tilde{h}_+^\nu(t) \;=\; \beta\sqrt{C}\left(m^\nu(t) + f\sum_{\mu\neq\nu}^{P} m^\mu(t)\right) \tag{B.8}$$

The calculation of the structured part of h_-^ν is carried out identically:

$$\tilde{h}_-^\nu(t) \;=\; \beta\sqrt{C}\left(f\sum_{\mu\neq\nu}^{P} m^\mu(t)\right) \;=\; \tilde{h}_+^\nu(t) - \beta\sqrt{C}m^\nu(t) \tag{B.9}$$

Using (B.5) we can write the full inputs in term of the foreground/background average activities m_+^ν and m_-^ν. For all pattern ξ^ν the average field on the selective and nonselective population read

$$\boxed{\begin{aligned} h_+^\nu = \sqrt{C}\Big[\; & (fJ_{EE} + \beta)m_+^\nu + (J_{EE}(1-f) - \beta)m_-^\nu \\ & + f\beta\sum_{\mu\neq\nu}^{P}(m_+^\mu - m_-^\mu) - J_{EI}m_I + Em_0\Big] \end{aligned}} \tag{B.10a}$$

$$\boxed{h_-^\nu = h_+^\nu \;-\; \sqrt{C}\;\beta(m_+^\nu - m_-^\nu)} \tag{B.10b}$$

and the average input to inhibitory neurons is

$$\boxed{h_I = \sqrt{C}\Big[J_{IE}(fm_+^\mu + (1-f)m_-^\mu) - J_{II}m_I + Im_0\Big]} \tag{B.11}$$

In general, the state of the excitatory network can be specified assigning the foreground and background activity for a particular pattern, let's say ξ^1, the average activities of the selective/non-selective populations for all others patterns cannot be arbitrary because of the overlaps and they must be consistent with the description of activity in term of pattern ξ^1. If the network state is specified by m_+^1 and m_-^1 we can determine m_+^μ and m_-^μ for $\mu \neq 1$ as follow:

- the selective population of any pattern ξ^μ is composed by fN neurons: Nf^2 of them are shared with the selective population of pattern ξ^1 and the remaining $Nf(1-f)$ belong to the nonselective population of pattern ξ^1; this enable us to write the weighted average:

$$m_+^\mu = \frac{m_+^1 N(f^2) + m_-^1 Nf(1-f)}{fN} \;=\; fm_+^1 + (1-f)m_-^1 = m_E$$

- the nonselective population of a pattern ξ^μ is composed by $N(1-f)$ neurons: $N(f-f^2)$ are shared with the selective population of pattern ξ^1 and the remaining $N(1-f)^2$ belongs to the nonselective population of pattern ξ^1, and we get:

$$m_-^\mu = \frac{m_+^1 Nf(1-f) + m_-^1 N(1-f)^2}{N(1-f)} = fm_+^1 + (1-f)m_-^1 = m_E$$

The symmetry of the system allow us to describe completely the system just in term of the selective and nonselective population for a generic pattern and the inhibitory population, we can drop the pattern index nu and write the significative inputs as:

$$h_+ = \sqrt{C}\Big[\,(fJ_{EE}+\beta)m_+ + (J_{EE}(1-f)-\beta)m_- - J_{EI}m_I + Em_0\Big] \quad \text{(B.12a)}$$

$$h_- = \sqrt{C}\Big[\,J_{EE}(fm_+ + (1-f)m_-) - J_{EI}m_I + Em_0\Big] \quad \text{(B.12b)}$$

$$h_I = \sqrt{C}\Big[\,J_{IE}(fm_+ + (1-f)m_-) - J_{II}m_I + Im_0\Big] \quad \text{(B.12c)}$$

B.3 MF solutions in the $C \to \infty$ limit

Average inputs (B.10, B.11) are $O(\sqrt{C})$, therefore in the $C \to \infty$ limit they will be divergent unless excitation and inhibition are *balanced* so that the leading term vanish and both the inputs remain finite.
If on the other hand the average excitatory and inhibitory inputs to a population will be *unbalanced* if the leading term dose not vanish, therefore in the in the $C \to \infty$ limit the neurons will be always active or always silent, depending on the sign of the leading term.

We examine the possibility of achieving bistability in the network, investigating de condition under which the MF system can admit these two type of solutions:

- a solution corresponding to the spontaneous, unselective activity state in which the average activity of the excitatory network is uniform and the average inputs of all populations are balanced
- a solution corresponding to the retrieval of a memory in which the selective neuron of a specific pattern are active and balanced with the inhibitory neurons while the nonselective neurons are silent because they are feeded with an inhibition-dominated unbalanced input

We observe that the second type of solution is meaningful because the nonselective neurons will be strictly silent only in the $C \to \infty$, for finite $C \gg 1$, instead, the activity will be extremely low but nonzero because the distance of the average input from threshold $(h - \theta)$ will be large and negative but still finite, thus large fluctuation can still set a neuron in the active state.

Spontaneous activity: full balancing

When a network is not engaged in the recall of a particular memory we say that it is in a state of *spontaneous activity* and the behavior is analogously to that of the unstructured network: excitatory and inhibitory populations are in a balanced state and the average fields vanish at first order in $\sqrt{C}$. We are then looking for a solution of the system

$$\begin{cases} h_+^\nu = 0 \\ h_-^\nu = 0 \\ h_I = 0 \end{cases} \qquad \text{for} \quad \nu = 1...P$$

The obvious ansatz is the following:

$$\begin{aligned} m_+^\mu \;=\; m_-^\mu \;&= m_E \in (0;1) \qquad \mu = 1...P \\ &\quad m_I \in (0;1) \end{aligned} \tag{B.13}$$

Since in this type of solution the activity of the selective/nonselective neurons is equal for all patterns we call it **symmetric state**.

Substituting (B.13) in the expressions of the average excitatory inputs (B.10) all the terms which are proportional to β disappear and $h_+^\mu = h_-^\mu \equiv h_E$ for all μ, so that the system of $2P+1$ equations collapse to a system of 2 equations which is identical to that of the memoryless network

$$\begin{cases} h_E = 0 \\ h_I = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} J_{EE} m_E - J_{EI} m_I + E m_0 = 0 \\ J_{IE} m_E - J_{II} m_I + I m_0 = 0 \end{cases}$$

whose solutions are

$$m_E = \left[\frac{E J_{II} - I J_{EI}}{J_{EI} J_{IE} - J_{EE} J_{II}} \right] m_0 \qquad m_I = \left[\frac{E J_{IE} - I J_{EE}}{J_{EI} J_{IE} - J_{EE} J_{II}} \right] m_0 \quad \text{(B.14)}$$

Again, the constraints to exclude unbalanced stationary solutions are (1.47).

Memory retrieval: partial balancing

The retrieval of a memory object is signaled by the enhancement of the activity in the sub-population of excitatory neurons which are selective for the pattern corresponding to that object. Our working hypotheses is that the activity of all other excitatory neurons (nonselective) drops to zero[2]. If we assume that pattern ξ^1 is the retrieved one, we look for a solution in which foreground neurons are active and nonsaturated while background neurons remains silent:

$$m_+^1 \in (0;1) \qquad m_-^1 = 0 \quad \text{(B.15)}$$

To obtain this state, the inhibitory population must balance the foreground yielding the cancelation of h_+^1 field at order $\sqrt{C}$ and, simultaneously, the field on nonselective cells should be $h_-^1 \sim -\sqrt{C}$, i.e. the network must settle in a state of *partial balancing*.

From the above consideration we can determine that the state of activity (B.15) yield the following ansatz for the form of the solution sought

$$\begin{aligned} & m_+^1 \in (0;1) && && m_-^1 = 0 \\ & m_+^\mu = m_-^\mu = f m_+^1 && \mu = 2 \ldots P \\ & m_I \in (0;1) \end{aligned} \quad \text{(B.16)}$$

we call it the **asymmetric state**.
Self-consistency requires that ξ^1 selective and inhibitory populations are balanced while ξ^1 nonselective input is unbalanced and negative. Populations that are either selective or nonselective for patterns $\mu \neq 1$ includes a finite fraction of silent neurons, thus the average inputs h_+^μ and h_-^μ will be dominated by the unbalanced inputs of order $-\sqrt{c}$ so we require that also the

[2]this can be strictly true only in the limit $C \to \infty$, for finite C, no matter how below threshold the input is, there is always a small but finite probability of firing.

fields h_+^μ and h_-^μ are unbalanced and negative. Therefore the retrieval state is the solution of the system

$$\begin{cases} h_+^1 = 0 \\ h_I = 0 \\ h_-^1 < 0 \\ h_+^{\mu\neq 1} = h_-^{\mu\neq 1} < 0 \end{cases} \Rightarrow \begin{cases} \left[(fJ_{EE}+\beta)m_+^1 - J_{EI}m_I + EX_0\right] = 0 \\ \left[fJ_{IE}m_+^1 - J_{II}m_I + IX_0\right] = 0 \\ h_+^1 - \sqrt{C}\beta m_+^1 < 0 \\ h_+^1 - \sqrt{C}\beta(1-f)m_+^1 < 0 \end{cases}$$

If $h_+^1 = 0$ the inequalities for the nonselective population inputs are automatically satisfied and we only need to solve the 2×2 system that represent the balancing between the foreground and inhibitory population:

$$\begin{cases} (fJ_{EE}+\beta)m_+^1 - J_{EI}m_I + Em_0 = 0 \\ fJ_{IE}m_+^1 - J_{II}m_I + Im_0 = 0 \end{cases}$$

whose solutions are

$$m_E = fm_+^1 = \left[\frac{EJ_{II} - IJ_{EI}}{J_{EI}J_{IE} - (J_{EE}+\beta/f)J_{II}}\right] m_0 \equiv \Omega_E(\beta/f)\, m_0 \quad \text{(B.17a)}$$

$$m_I = \left[\frac{EJ_{IE} - I(J_{EE}+\beta/f)}{J_{EI}J_{IE} - (J_{EE}+\beta/f)J_{II}}\right] m_0 \equiv \Omega_I(\beta/f)\, m_0 \quad \text{(B.17b)}$$

When $\beta = 0$ we recover the gain parameters of the unstructured network which yield the activity of the network in the symmetric solution.
Rates are positive and unbalanced solutions are excluded provided that

$$\frac{E}{I} > \frac{J_{EI}}{J_{II}} > \frac{J_{EE}+\beta/f}{J_{IE}} \quad \text{(B.18)}$$

keeping in mind that we must always have $\beta < J_{EE}(1-f)/P$ to preserve the positiveness of the excitatory synapses. It can be demonstrated that a fully unbalanced solution with $m_+^1 = m_I = 1$ is excluded for any m_0 if

$$f < \frac{EJ_{II} - IJ_{EI}}{EJ_{IE} - IJ_{EE}} \equiv f^* \quad \text{(B.19)}$$

In figure B.1 we plot the gain Ω_A appearing in (B.17) as a function of β/f for the following set of network parameters:

$J_{EE} = 1$	$J_{EI} = 2$	$E = 3$
$J_{IE} = 1$	$J_{II} = 1$	$I = 1$

(B.20)

The graph tells that the average network activity in the retrieval state is always higher with respect to the spontaneous activity.

The divergence of the gains occurs when β/f becomes too large, breaking down the constraint (B.18). We highlight that figure B.1 shows the gain of

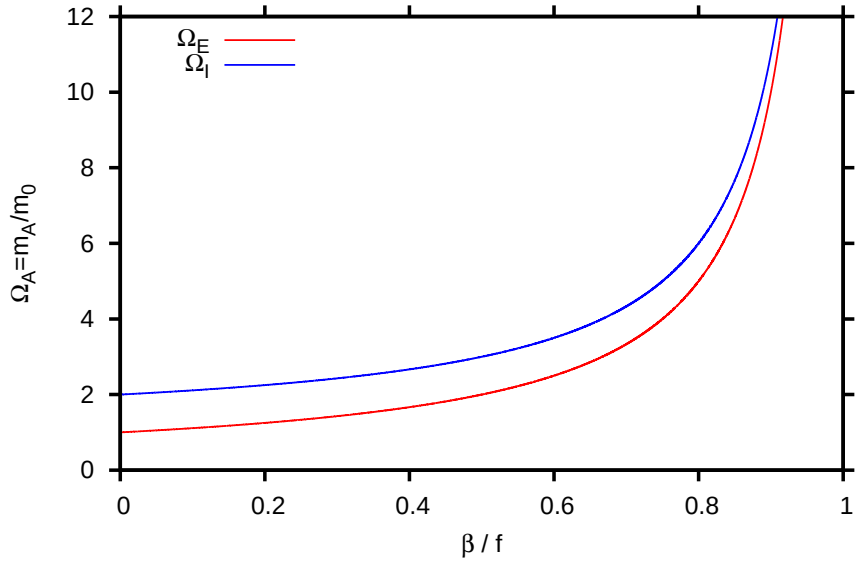

Figure B.1: Gain function for the excitatory and inhibitory population in the retrieval solution (B.1) as a function of β/f.

the excitatory population as a whole, but since the nonselective population is inactive, the gain for the selective population would be $\Omega_+ \equiv \Omega/f$. Since f shuold stay below f^* $(B.19)$, the activity of the selective neuron in the retrieval state cannot get lower than $m_0\Omega/f^*$. On the other hand, there is no lower limit for the coding level f, and Ω_+ can be made arbitrarily large. In general, for any P and f, we can take β arbitrarily low to fulfil $\beta < J_{EE}(1-f)/P$ and still have a retrieval solution. But its existence does not guarantee stability, therefore we must examine in what ranges of parameters this solution is stable.

B.4 Input variance

To determine the stability of the solution we need to know the variance of the inputs. Here we show how to calculate the variance of the field across the selective population. If we consider the recurrent excitatory part only, therefore we have to calculate the quantity :

$$(\sigma_+^\nu)^2 \equiv \frac{1}{Nf}\sum_{i=1}^{N}\xi_i^\nu[\tilde{h}_i(t)-\tilde{h}_+^\nu(t)]^2 = \frac{1}{Nf}\sum_{i=1}^{N}\xi_i^\nu\tilde{h}_i^2(t)-[\tilde{h}_+^\nu(t)]^2$$

We stress that this quantity will be the *total* input variance, i.e. the sum of the variance due to temporal and quenched fluctuations.
First we calculate the average of the squared input to the neurons selective

for pattern ν, writing explicitly the full expression we get:

$$\frac{1}{fN}\sum_{i=1}^{N}\xi_i^\nu[\tilde{h}_i(t)]^2 = \frac{1}{fN}\sum_{i=1}^{N}\xi_i^\nu\left\{\sum_{j=1}^{N}\frac{c_{ij}}{\sqrt{C}}\left[J_{EE}+\frac{\beta}{f(1-f)}\sum_{\mu=1}^{P}\xi_i^\mu(\xi_j^\mu-f)S_j(t)\right]\right\}^2 =$$

calculating the square of the input and collecting the terms that depends on i we get three terms

$$\begin{aligned}
&= \frac{J_{EE}^2}{Cf}\sum_{j,l=1}^{N}\left[\frac{1}{N}\sum_{i=1}^{N}c_{ij}c_{il}\xi_i^\nu\right]+\\
&\frac{2J_{EE}\beta}{Cf^2(1-f)}\sum_{j,l=1}^{N}\sum_{\alpha=1}^{P}\left[\frac{1}{N}\sum_{i=1}^{N}c_{ij}c_{il}\xi_i^\nu\xi_i^\alpha\right](\xi_j^\alpha-f)S_j(t)+\\
&\frac{\beta^2}{Cf^3(1-f)^2}\sum_{j,l=1}^{N}\sum_{\alpha,\beta=1}^{P}\left[\frac{1}{N}\sum_{i=1}^{N}c_{ij}c_{il}\xi_i^\nu\xi_i^\alpha\xi_i^\beta\right](\xi_j^\alpha-f)(\xi_l^\beta-f)S_j(t)S_l(t)
\end{aligned} \tag{B.21}$$

In the limit $N\to\infty$ the term in square bracket is the expectation value of the product of the random variables c and ξ, since the connectivity is independent from the patterns we can take the product of the expectation values

$$E[c_jc_l\,\xi] = E[c_jc_l]E[\xi] = \left[\delta_{jl}\frac{C}{N}\left(1-\frac{C}{N}\right)+\frac{C^2}{N^2}\right]E[\xi]$$

Putting back this expression in (B.21) and observing that $S_i^2 = S_i$ we get

$$\begin{aligned}
=&J_{EE}^2\left(1-\frac{C}{N}\right)+CJ_{EE}^2+\\
&\frac{\beta^2}{f^3(1-f)^2}\sum_{\alpha,\beta=1}^{P}E[\xi^\nu\xi^\alpha\xi^\beta]\left[\left(1-\frac{C}{N}\right)\frac{1}{N}\sum_{j=1}^{N}(\xi_j^\alpha-f)(\xi_j^\beta-f)S_j(t)\right]\\
&+\;C\,\frac{\beta^2}{f}\sum_{\alpha,\beta=1}^{P}E[\xi^\nu\xi^\alpha\xi^\beta]m^\alpha m^\beta
\end{aligned} \tag{B.22}$$

Recalling definitions (3, 2), the term in square brackets can be express in term of the order parameters, and if we take the the $N\to\infty$ limit for finite C we get

$$\left(1-\frac{C}{N}\right)\frac{1}{N}\sum_{j=1}^{N}\left(\xi_j^\alpha\xi_j^\beta-f(\xi_j^\alpha+\xi_j^\beta)+f^2\right)S_j(t) = f^2\left(m_+^{\alpha\beta}-m_+^\alpha-m_+^\beta+m_E\right)$$

The first term on the right hand side is a quantity that cannot be expressed in term of the order parameters we have defined so far, therefore it represent a new order parameter

Definition 5 (*Average activity in the population which is selective for both pattern α and β*)**.** For any $\alpha, \beta = 1...P$

$$m_+^{\alpha\beta}(t) \equiv \frac{1}{f^2 N} \sum_{i=1}^{N} \xi_i^\alpha \xi_i^\beta S_i(t)$$

In particular, if the indexes are equal:

$$m_+^{\alpha\alpha}(t) = \frac{1}{f} m_+^\alpha(t)$$

The average of the squared field (B.22) can be rewritten as

$$\underbrace{\frac{\beta^2}{f(1-f)^2} \sum_{\alpha,\beta=1}^{P} E[\xi^\nu \xi^\alpha \xi^\beta]\Big(m_+^{\alpha\beta} - m_+^\alpha - m_+^\beta + m_E\Big)}_{\text{(I)}} + \underbrace{C\frac{\beta^2}{f} \sum_{\alpha,\beta=1}^{P} E[\xi^\nu \xi^\alpha \xi^\beta] m^\alpha m^\beta}_{\text{(II)}}$$

Let's calculate the value of $E[\xi^\nu \xi^\alpha \xi^\beta]$

$$E[\xi^\nu \xi^\alpha \xi^\beta] = \begin{cases} \text{if } \alpha = \beta \qquad E[\xi^\nu \xi^\alpha] \;=\; \begin{cases} f & \text{if } (\alpha = \nu) \\ f^2 & \text{if } (\alpha \neq \nu) \end{cases} \\ \\ \text{if } \alpha \neq \beta \quad E[\xi^\nu \xi^\alpha \xi^\beta] \;=\; \begin{cases} f^2 & \text{if } (\alpha = \nu) \text{ or } (\beta = \nu) \\ f^3 & \text{if } (\alpha \neq \nu) \text{ and } (\beta \neq \nu) \end{cases} \end{cases}$$

Those results can be expressed formally in the following way:

$$\begin{aligned} E[\xi^\nu \xi^\alpha \xi^\beta] =\; & \delta_{\alpha\beta}\Big[f\delta_{\alpha\nu} + f^2(1-\delta_{\alpha\nu})\Big] \\ & + (1-\delta_{\alpha\beta})\Big[f^2(\delta_{\alpha\nu}+\delta_{\beta\nu}) + f^3(1-\delta_{\alpha\nu})(1-\delta_{\beta\nu})\Big] \\ =\; & f\Big\{(1-f)(1-2f)\delta_{\alpha\nu}\delta_{\beta\nu} + f(1-f)\big[\delta_{\alpha\beta}+\delta_{\alpha\nu}+\delta_{\beta\nu}\big] + f^2\Big\} \end{aligned}$$

PROBLEM: input variance is $O(C)$

If we go back to the definition of input variance we can write

$$(\sigma_+^\nu)^2; = \frac{1}{Nf} \sum_{i=1}^{N} \xi_i^\nu \tilde{h}_i^2(t) - [\tilde{h}_+^\nu(t)]^2 = \text{(I)} + \text{(II)} - [\tilde{h}_+^\nu(t)]^2$$

The last two terms are proportional to C and they should cancel if we want the variance to be $O(1)$. Let's compare (II) with the square of the average field (B.7)

$$\begin{aligned} [\tilde{h}_+^\nu(t)]^2 =& C\beta^2 \sum_{\alpha,\beta=1}^{P} m^\alpha m^\beta \Big\{(1-f)^2\delta_{\alpha\nu}\delta_{\beta\nu} + f(1-f)(\delta_{\alpha\nu}+\delta_{\beta\nu}) + f^2\Big\} \\ \text{(II)} =& C\beta^2 \sum_{\alpha,\beta=1}^{P} m^\alpha m^\beta \Big\{(1-f)(1-2f)\delta_{\alpha\nu}\delta_{\beta\nu} + f(1-f)\big[\delta_{\alpha\beta}+\delta_{\alpha\nu}+\delta_{\beta\nu}\big] + f^2\Big\} \end{aligned}$$

we see that they are not equal! Therefore taking the difference and carrying out the calculations we are left with a disturbing term of order C

$$(\sigma_+^\nu)^2 \;=\; (\mathrm{I}) + C\;\beta^2 f(1-f)\sum_{\mu\neq\nu}^{P}(m^\mu)^2$$

We can try to see what happen to this last term on the two standard solutions:

- **Symmetric solution.** The term of $O(C)$ vanishes for every ν because $m^\mu = 0$ for all $\mu = 1...P$)and we get $(\sigma_+^\nu)^2 = (I)$ which is $O(1)$.
- **Asymmetric solution.** In the single item retrieval solution of, say, pattern ξ^1 we have ($m_+^1 = m^1 < 1$, $m_-^1 = 0$) while $m^\mu = 0$ for any $\mu \neq 1$. In this case only the variance of the field for $\nu = 1$ are of $O(1)$ while all the others will be of $O(C)$, therefore in the large C limit $(\sigma_+^\mu)^2 \simeq C\beta^2 f(1-f)m^1$.
  This might seems pathologic, but if calculate the ratio of the fluctuation to the mean input for for non-retrived patterns ($\mu \neq 1$) in the large C limit we find

  $$\frac{\sigma_+^\mu}{|h_+^\mu|} = \frac{\sqrt{(\mathrm{I}) + C\beta^2 f(1-f)m^1}}{|h_+^1 - \sqrt{C}\beta(1-f)m^1|} \simeq \sqrt{\frac{f}{1-f}} \xrightarrow[f\ll 1]{} \sqrt{f}$$

  The field on the selective(and nonselective) neurons for the non-retrived pattern is largely below threshold because it is $O(-\sqrt{C})$, but if coding is low, its fluctuations will never bring it above threshold because they will be a factor $\sqrt{f}$ smaller, ensuring that background neurons stay silent, despite the fluctuations are $O(\sqrt{C})$.

The only thing left is the calculation of the expression (I):

$$(I) = \frac{\beta^2}{1-f}\sum_{\alpha,\beta=1}^{P}\left[(1-2f)\delta_{\alpha\nu}\delta_{\beta\nu} + f\big(\delta_{\alpha\beta}+\delta_{\alpha\nu}+\delta_{\beta\nu}\big) + \frac{f^2}{1-f}\right]\Big(m_+^{\alpha\beta} - m_+^\alpha - m_+^\beta + m_E\Big)$$

If we can characterize the state of the network in terms of selective/non-selective excitatory activity and inhibitory activity we can derive that the variances of the input to the three populations read

$$\begin{aligned}
\sigma_+ =& \left[J_{EE} + \frac{\beta}{f}\right]^2 f m_+ + \left[J_{EE} - \frac{\beta}{1-f}\right]^2 (1-f)m_- + \frac{\beta^2}{f}(P-1)m_E + J_{EI}^2 m_I \\
\sigma_- =& \left[J_{EE}^2 + \frac{\beta^2}{f}(P-1)\right] m_E + J_{EI}^2 m_I \\
\sigma_I =& \; J_{IE}^2 m_E + J_{II}^2 m_I
\end{aligned}$$

we recall that $m_E = f m_+ + (1-f)m_-$

B.5 Local stability of the solutions

Here we study the stability of the MF solutions with respect to small perturbations in the average activity of the neurons that are selective for a pattern.

The activity of populations that are selective and nonselective for the various patterns are not independent, we then consider a perturbation in the activity of a specific pattern ν and derive the perturbation that is induced on other patterns. Assuming

$$m_+^\nu(t) = m_E + \delta m_+^\nu(t) \qquad m_-^\nu(t) = m_E + \delta m_-^\nu(t)$$

the activity of all other patterns turns out to be

$$m_+^\mu(t) = m_-^\mu(t) = m_E + \underbrace{[f\delta m_+^\nu(t) + (1-f)\delta m_-^\nu(t)]}_{\delta m_+^\mu(t)=\delta m_-^\mu(t)} \tag{B.23}$$

thus the perturbation induced on other patterns activities result equal for both the selective and the nonselective population. This fact implies that the solution for the activity of any pattern $\mu \neq \nu$ is automatically determined once we discover the evolution of the perturbations on pattern ν alone, therefore it suffice studying the linearized dynamics for m_+^μ and m_-^μ alone[3]. Dropping the ν index for the perturbed pattern we have that the variation of the fields read

$$\begin{aligned}
\delta h_+ =& \sqrt{C}\left[(fJ_{EE}+\beta)\delta m_+(t) + (J_{EE}(1-f)-\beta)\delta m_-(t) - \delta m_I J_{EI}\right] \\
\delta h_- =& \sqrt{C}\left[fJ_{EE}\,\delta m_+(t) + J_{EE}(1-f)\,\delta m_-(t) - \delta m_I J_{EI}\right] \\
\delta h_I =& \sqrt{C}\left[(f\,\delta m_+(t) + (1-f)\,\delta m_-(t))J_{IE} - \delta m_I J_{II}\right]
\end{aligned}$$

we can linearize the dynamics provided that $\delta m \ll 1/\sqrt{C}$ for all perturbations and we get a 3×3 linear system of differential equations

$$\tau_E \frac{d}{dt}\begin{pmatrix}\delta m_+(t)\\ \delta m_-(t)\\ \delta m_I(t)\end{pmatrix} = \begin{pmatrix} \frac{\partial H_+}{\partial m_+} - 1 & \frac{\partial H_+}{\partial m_-} & \frac{\partial H_+}{\partial m_I} \\ \frac{\partial H_-}{\partial m_+} & \frac{\partial H_-}{\partial m_-} - 1 & \frac{\partial H_-}{\partial m_I} \\ \frac{\tau_e}{\tau_I}\frac{\partial H_I}{\partial m_+} & \frac{\tau_E}{\tau_I}\frac{\partial H_I}{\partial m_-} & \frac{\tau_E}{\tau_I}\frac{\partial H_I}{\partial m_I} - \frac{\tau_E}{\tau_I} \end{pmatrix}\begin{pmatrix}\delta m_+(t)\\ \delta m_-(t)\\ \delta m_I(t)\end{pmatrix} \tag{B.24}$$

Stability is ensured if the real part of all three eigenvalues is negative. The characteristic polynomial of a generic 3×3 matrix M reads[4]:

$$\mathcal{P}(\lambda) = -\lambda^3 + \lambda^2 \mathrm{tr}(M) + \lambda\frac{1}{2}[\mathrm{tr}(M^2) - \mathrm{tr}^2(M)] + \det(M)$$

[3] The same conclusion can be derived assuming independent perturbations on the foreground and background populations for all patterns and exploit the symmetries of the problem to reduce the $2P+1$ equations to just $2+1$.

[4] $\frac{1}{2}\left[\mathrm{tr}(M^2) - \mathrm{tr}^2(M)\right] = M_{12}M_{21} - M_{11}M_{22} + M_{32}M_{23} - M_{22}M_{33} + M_{31}M_{13} - M_{11}M_{33}$

The roots of a cubic polynomial $x^3 + ax^2 + bx + c$ can be all real or we can have a single real root and two complex conjugate. We can determine the nature of the roots by looking at the sign of the discriminant

$$\Delta = a^2b^2 - 4(b^3 + ca^3) + 9(\ 2abc - 3c^2) \tag{B.25}$$

If $\Delta \geq 0$ the roots are all real (if $\Delta = 0$ one of the roots has multiplicity 2) while if $\Delta < 0$ we have a single real root and two complex conjugate roots. It is easy to see that the coefficients of the polynomial can be expressed as a function of the roots, if we have three real roots

$$\begin{aligned} \mathrm{tr}(M) &= \lambda_1 + \lambda_2 + \lambda_3 \\ \frac{1}{2}\Big[\mathrm{tr}(M^2) - \mathrm{tr}^2(M)\Big] &= -\lambda_1\lambda_2 - \lambda_2\lambda_3 - \lambda_1\lambda_3 \\ \det(M) &= \lambda_1\lambda_2\lambda_3 \end{aligned}$$

if there is a single real root λ_1 and two complex conjiugate $\lambda_2 = \bar{\lambda_3} = ae^{i\alpha}$ the relations are:

$$\begin{aligned} \mathrm{tr}(M) &= \lambda_1 + 2\mathrm{Re}[\lambda_2] \\ \frac{1}{2}\Big[\mathrm{tr}(M^2) - \mathrm{tr}^2(M)\Big] &= -2\lambda_1\mathrm{Re}[\lambda_2] - a^2 \\ \det(M) &= \lambda_1 a^2 \end{aligned}$$

Therefore the stability of an equilibrium solution is guarantee in both cases if the following system of inequalities holds

$$\begin{cases} \mathrm{tr}(M) < 0 \\ \dfrac{1}{2}\Big[\mathrm{tr}(M^2) - \mathrm{tr}^2(M)\Big] < 0 \\ \det(M) < 0 \end{cases} \tag{B.26}$$

We can write the linearized system (B.24) in a compact notation as

$$\tau_E \frac{d}{dt}\delta m_i = M_{ij}\ \delta m_j$$

When evaluating the derivative of the functions H on the equilibrium solution, in the limit $C \gg 1$ we keep only the terms proportional to $\sqrt{C}$ as shown in (A.16). With this approximation we can write explicitly the matrix appearing in (B.24) as

$$M_{ij} = \sqrt{C}X_{ij}Y_{ij} - \delta_{ij} + \delta_{iI}\left(1 - \frac{\tau_E}{\tau_I}\right) \tag{B.27}$$

where $i, j = +, -, I$ and δ_{ij} is the usual Kronecker symbol. Matrix X depends on the input statistics at equilibrium, defining the average residual

field[5] as $u_A = \theta_A - h_A$ we have

$$X = \frac{1}{\sqrt{2\pi}} \begin{pmatrix} \frac{1}{\sigma_+} e^{-\frac{u_+^2}{2\sigma_+^2}} & 0 & 0 \\ 0 & \frac{1}{\sigma_-} e^{-\frac{u_-^2}{2\sigma_-^2}} & 0 \\ 0 & 0 & \frac{\tau_E}{\tau_I}\frac{1}{\sigma_I} e^{-\frac{u_I^2}{2\sigma_I^2}} \end{pmatrix} \tag{B.28}$$

Matrix Y depends on the network parameters read

$$Y = \begin{pmatrix} fJ_{EE} + \beta & J_{EE}(1-f) - \beta & -J_{EI} \\ fJ_{EE} & J_{EE}(1-f) & -J_{EI} \\ fJ_{IE} & J_{IE}(1-f) & -J_{II} \end{pmatrix} \tag{B.29}$$

We investigated the stability of the symmetric and asymmetric state for a network with the parameters listed in table (B.20) at different loading P and for different ratios of the neuron time constants τ_E/τ_I. Essentially, we determined numerically the set of parameters (m_0, β, f) for which the system (B.26) is fulfilled.

Stability of the asymmetric solution

In the retrieval asymmetric solution we have that $m_- = 0$ and $m_E = fm_+$. The variances of the field are

$$\begin{aligned} \sigma_+^2 &= \left[fJ_{EE}^2 + 2J_{EE}\beta + \beta^2 \left(\frac{1}{f} + P - 1 \right) \right] m_+ + J_{EI}^2 m_I \\ \sigma_-^2 &= \left[fJ_{EE}^2 + \beta(P-1) \right] m_+ + J_{EI}^2 m_I \\ \sigma_I^2 &= J_{IE}^2 f m_+ + J_{II}^2 m_I \end{aligned} \tag{B.30}$$

where the equilibrium rates are given in B.17. In this solution the nonselective neurons are off because the input is $O(-\sqrt{C})$ yielding $X_- \sim e^{-C}$. This imply that in the large C limit the approximation $M_{ij} \simeq \sqrt{C} X_{ij} Y_{ij}$ holds only for $i = +, I$, while for $i = -$ we have $M_{ij} \simeq -\delta_{ij}$. The matrix M for the asymmetric solution in the very large C reads:

$$M_a = \sqrt{C} \begin{pmatrix} X_+[fJ_{EE} + \beta] & X_+[J_{EE}(1-f) - \beta] & -X_+ J_{EI} \\ 0 & -1/\sqrt{C} & 0 \\ X_I f J_{IE} & X_I J_{IE}(1-f) & -X_I J_{II} \end{pmatrix} \tag{B.31}$$

Calculating the quantities appearing in the stability system of inequalities (B.26) we get

$$\mathrm{tr}(M_a) = \sqrt{C}\Big[X_+(fJ_{EE} + \beta) - X_I J_{II} \Big] - 1 \tag{B.32a}$$

[5] they are determined by inverting the fixed point equation (1.42) once the equilibrium rates are given.

$$\begin{aligned}\frac{1}{2}\Big[\mathrm{tr}(M_a^2) - \mathrm{tr}^2(M_a)\Big] = &- CX_+X_-\beta J_{EE} \\ &- CX_+X_I[J_{EI}J_{IE} - J_{II}(J_{EE} + \beta/f)]\, f\end{aligned} \tag{B.32b}$$

$$\det(M_a) \;= -CX_+X_I\left[J_{EI}J_{IE} - J_{II}(J_{EE} + \beta/f)\right] f \tag{B.32c}$$

This time the constraints (B.18) ensures that (B.32b) and (B.32c) are always negative, therefore the asymmetric solution is stable if the trace is negative:

$$\sqrt{C}\Big[X_+(fJ_{EE} + \beta) - X_I J_{II}\Big] - 1 < 0$$

Dividing by $\sqrt{C}$ and taking the $C \to \infty$ limit we find that the asymmetric state is stable when

$$\frac{fJ_{EE} + \beta}{J_{II}} \;<\; \frac{X_I}{X_+} \;=\; \frac{\tau_E}{\tau_I}\,\frac{\sigma_+}{\sigma_I}\,\exp\left[-\frac{u_I^2}{2\sigma_I^2} + \frac{u_+^2}{2\sigma_+^2}\right] \tag{B.33}$$

Stability of the symmetric solution

In the symmetric solution we have $m_+^\mu = m_-^\mu = m_E$ for all $\mu = 1...P$. The variance of the inputs to the populations are given by

$$\begin{aligned}\sigma_+^2 \;&=\; \left[J_{EE}^2 + \beta^2\left(\frac{f + (P-1)}{f(1-f)}\right)\right] m_E \;+\; J_{EI}^2 m_I \\ \sigma_-^2 \;&=\; \left[J_{EE}^2 + \beta^2\left(\frac{P-1}{f}\right)\right] m_E \;+\; J_{EI}^2 m_I \\ \sigma_I^2 \;&=\; J_{IE}^2 m_E \;+\; J_{II}^2 m_I\end{aligned} \tag{B.34}$$

with m_E and m_I given by B.13. The diagonal element of matrix X are always positive and, no matter how small they are, we can take C sufficiently large so that the approximation

$$M_{ij} \underset{C\gg 1}{\simeq} \sqrt{C}X_{ij}Y_{ij}$$

holds. Therefore, for large C, we can approximate the system (B.26) by

$$\begin{cases} \sqrt{C}\,\mathrm{tr}(XY) < 0 \\ C\dfrac{1}{2}\Big[\mathrm{tr}(X^2Y^2) - \mathrm{tr}^2(XY)\Big] < 0 \\ C^{\frac{3}{2}}\det(XY) < 0 \end{cases}$$

and dividing each inequality by the corresponding powers of C we get that the condition for for stability in the $C \to \infty$ limit are

$$\begin{cases} \mathrm{tr}(XY) < 0 \\ \dfrac{1}{2}\Big[\mathrm{tr}(X^2Y^2) - \mathrm{tr}^2(XY)\Big] < 0 \\ \det(XY) < 0 \end{cases} \tag{B.35}$$

If we calculate explicitly these quantities we get[6]

$$\mathrm{tr}(XY) = J_{EE}[fX_+ + (1-f)X_-] + X_+\beta - X_I J_{II} \tag{B.36a}$$

$$\begin{aligned}\frac{1}{2}\Big[\mathrm{tr}(X^2Y^2) - \mathrm{tr}^2(XY)\Big] = &- X_+X_-\beta J_{EE} \\ &- X_-X_I[J_{EI}J_{IE} - J_{II}J_{EE}]\,(1-f) \\ &- X_+X_I[J_{EI}J_{IE} - J_{II}(J_{EE}+\beta/f)]\,f\end{aligned} \tag{B.36b}$$

$$\det(XY) \;=\; X_+X_-X_I\,[J_{EI}J_{IE} - J_{EE}J_{II}]\,\beta \tag{B.36c}$$

If network parameters are subjects to the constraints (B.18) to exclude unbalanced solutions, it follows that (B.36c) will always be positive, therefore we conclude that with our hypotheses

> **The symmetric solution in the $C \to \infty$ limit is always unstable.**

[6] The make notation simpler we write X_{ii} as X_i

Appendix C

Region of bi-stability in the parameter space of a balanced memory network with STD

Study of $\phi(x)$

Function (**??**) depends only on the parameters that regulates the synaptic dynamics (τ_r, U) thus resulting independent from the statistical properties of the two neural population at equilibrium. Its graph is monotonously growing and passes through zero for any set of parameters, while $\phi(x_l) = 1$ where x_l was given in (3.3). It is useful to define the new parameters

$$R \equiv \frac{1}{x_l} = 1 + U\tau_r; \qquad Q \equiv \frac{U\tau_r^2}{\tau_r + 1} \tag{C.1}$$

so that we can write

$$\phi(x) = \frac{(R-Q)\,x}{1 - Q\,x} \tag{C.2}$$

This form allows a direct interpretation of the new parameters with respect to the geometrical characteristics of the graph of $\phi(x)$:

- The value of R sets the width of the interval in which the function grows from 0 to 1. Since the abscissa for which the function takes the value 1 is $1/R$, a large R means that the the function reach 1 in a short distance. This can be seen in the left panel of figure C.1 where the function (C.2) have been plot for some values of R keeping Q fixed at 0.5 .

- The derivative of function (C.2) in the origin is $R-Q$. Therefore Q set the steepness of the graph in 0. If $Q \lesssim R$ the graph will be almost flat in the origin then will rise very steeply as it approaches $x = 1/R$. On

the other extreme, if $Q \ll R$ the curve will rise almost in a straight line with steepness R. We can see an example of this in the right panel of figure C.1, where we plot some graphs of the function (C.2) for different Q while R is kept fixed at 5 .

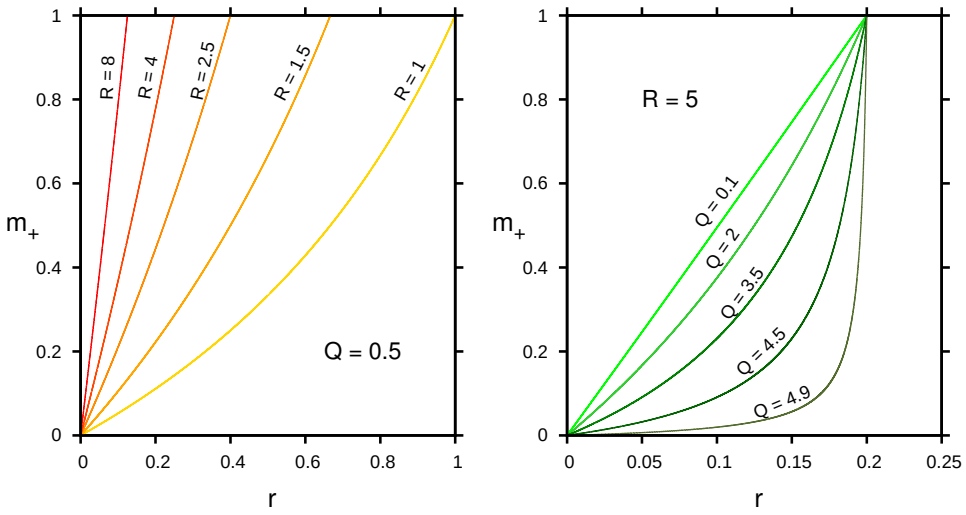

Figure C.1: see text

The above considerations are valid for arbitrary Q, R but from the definitions (C.1) and the fact that U and τ_r ar positive, it follows that we must respect the constraints:

$$R > 1 \qquad 0 < Q < (R-1) \tag{C.3}$$

Once R and Q have been set to get the desired shape for the graph[1], the corresponding values of τ_r and U can be determined through the inverse relations:

$$\tau_r = \tau_E \frac{Q}{R-1-Q} \qquad U = (R-1)\frac{R-1-Q}{\tau_E\, Q} = \frac{R-1}{\tau_r} \tag{C.4}$$

Study of $\psi(x)$

Now we study the function (**??**). Defining

$$C = H^{-1}(m_-) \qquad D = \frac{J_{EE}(a-1)}{\sigma_E} F e^{-\alpha F^2} \tag{C.5}$$

we can write

$$\psi(x) \equiv H(C - D\,x) \tag{C.6}$$

This function results simply form a linear transformation of the argument of the sigmoidal function $H(x)$ seen in (1.60), therefore its properties are determined exclusively by the value of the two parameters C and D:

[1]With the constraints (C.3) the derivative of the function in the origin $\phi'(0) = R - Q$ is bounded in the interval $(1; R)$, thus the graph can never start with a slope lower than 1

- C determine the shift of the function on the abscissae axis, i.e. $\phi(0) = H(C)$, therefore if $C > 0$, the function is shifted toward left, otherwise toward right. Some example of graph with different values of C can bee seen in the left panel of figure C.2. We notice that with definition (C.2) we have $H(C) = m_-$.

- H is a monotonically decreasing function of its argument, but since $D > 0$ (because $a > 1$) and there is a $-$ sign, ψ will be a monotonically increasing function. The magnitude of D is connected with the "stretching" of the function along the horizontal direction. This results clear if we calculate the derivative in the center of the function at $x = C/D$ (where $\psi(C/D) = 0.5$, we get $\psi'(C/D) = D/\sqrt{2\pi}$. If $0 < D \ll 1$ the function $\psi(x)$ will be very slowly increasing; conversely, if $D \gg 1$ the function will rise abruptly from 0 to 1 when crossing the central point $x = C/D$. See the right panel of figure C.2 for some examples.

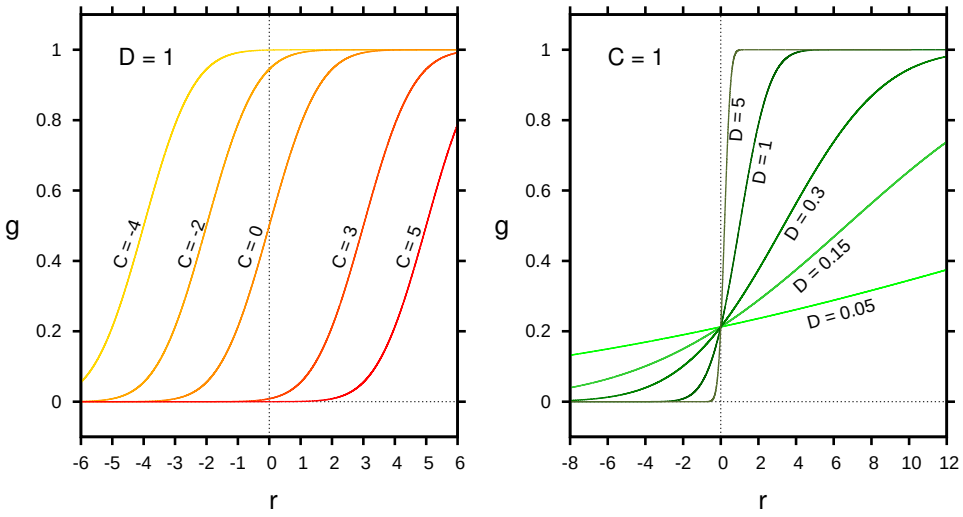

Figure C.2: see text

Parameter space

If $C < 0$ the center of function $\psi(x)$ has a negative abscissa and there can be only a single intersection, therefore we look for the regions of bi-stability in the domain

$$\Gamma = \{R > 1\ ;\ 0 < Q < (R-1)\ ;\ C > 0\ ;\ D > 0\ \} \tag{C.7}$$

We can think at the solution of equation (**??**) as a multi-branched function $\Lambda(R, Q, C, D)$ defined in Γ, and the corresponding value of the foreground activity are found calculating either ϕ or ψ:

$$\begin{aligned} m_+(R, Q, C, D) &= \phi(R, Q, \Lambda(R, Q, C, D)) \\ &= \psi(C, D, \Lambda(R, Q, C, D)) \end{aligned} \tag{C.8}$$

It is easy to verify that the equation is invariant for the parameter transformation $(R, Q, D) \to (\alpha R, \alpha Q, \alpha D)$ thus also its solutions:

$$m_+(R, Q, C, D) = m_+(\alpha R, \alpha Q, C, \alpha D) \tag{C.9}$$

In the 3-dimensional subspace

$$\gamma = \{R > 1; 0 < Q < (R-1)\ ;\ D > 0\} \subset \Gamma$$

the solutions are constant on the 'rays', thus the solutions can be parameterized on a 2-dimensional space.

It can be demonstrated that, in the space γ, there is a family of coordinate transformation $(R, Q, D) \to (\alpha, \tilde{R}, \tilde{D})$ defined for any $k > 1$ by

$$\begin{cases} \alpha = \dfrac{R - Q}{k} \\ \tilde{R} = \dfrac{kR}{R - Q} \\ \tilde{D} = \dfrac{kD}{R - Q} \end{cases} \qquad \begin{cases} R = \alpha \tilde{R} \\ Q = \alpha(\tilde{R} - k) \\ D = \alpha \tilde{D} \end{cases} \tag{C.10}$$

so that the new coordinate $\tilde{R}$ is constant over the rays of the space.
Writing the generic point $(R, Q, C, D) \in \Gamma$ in the new coordinates (C.10) and then applying the property (C.9) we have that

$$\begin{aligned} m_+(R, Q, C, D) &= m_+\left(\alpha\tilde{R}, \alpha(\tilde{R} - k), C, \alpha\tilde{D}\right) \\ &= m_+\left(\tilde{R}, (\tilde{R} - k), C, \tilde{D}\right) \end{aligned}$$

This means that the solutions depends on 3 independent parameters rather than 4, and we can explore all the possible values that the foreground activity can take by calculating m_+ on the 3-dimensional domain

$$\{\ \tilde{R} > k\ ;\ C > 0\ ;\ \tilde{D} > 0\ \}$$

Appendix D

Numerical methods

Here we present the technique employed to solve numerically the nonlinear system of equations obtained via MF theory and give details about the simulations of the networks. All the codes have been written in *C*.

D.1 Resolution of MF equations

The mean field system of equations at finite C are solved by interpreting the solutions as the stable fixed points of the fictitious dynamical system. For the standard balanced network, for example, it would be

$$\begin{cases} \tau_E \dfrac{dm_E}{dt} = -m_E + H\left(\dfrac{\theta_E - u_E}{\sqrt{\alpha_E}}\right) \\ \tau_I \dfrac{dm_I}{dt} = -m_I + H\left(\dfrac{\theta_I - u_I}{\sqrt{\alpha_I}}\right) \\ \tau_E \dfrac{dq_E}{dt} = -q_E + \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\dfrac{\theta_E - u_E + x\sqrt{\beta_E}}{\sqrt{\alpha_E - \beta_E}}\right)\right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \\ \tau_I \dfrac{dq_I}{dt} = -q_I + \displaystyle\int_{-\infty}^{+\infty} \left[H\left(\dfrac{\theta_I - u_I + x\sqrt{\beta_I}}{\sqrt{\alpha_I - \beta_I}}\right)\right]^2 \dfrac{e^{-x^2/2}}{\sqrt{2\pi}} dx \end{cases} \tag{D.1}$$

The idea is then to follow the time evolution of these quantities until they reach a stationary value.
In practice, the equations (D.1) are integrated using the Euler algorithm, at every time-step the absolute value of the increment of the quantities was checked - the stationary points were considered reached when all the increments became smaller than $\epsilon = 10^{-8}$.
Figure D.1 shows an example of trajectories of the macroscopic variables. The precise values of the time constants τ will determine the shape of the trajectories, of course, but do not affect the stationary points, therefore their

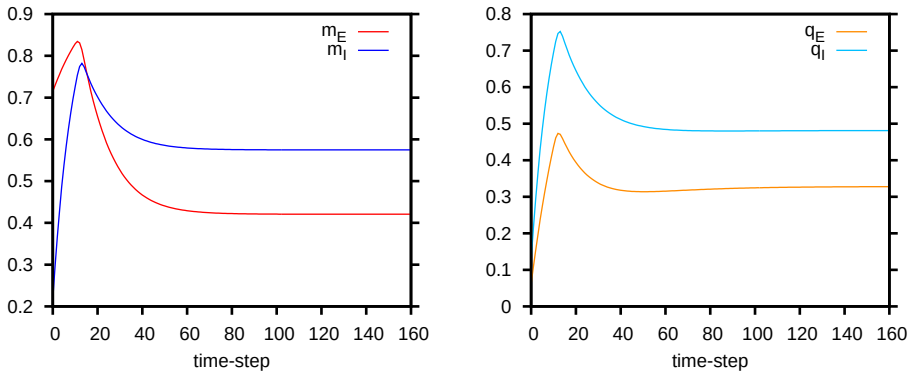

Figure D.1: Typical path of convergence to equilibrium values of the macroscopic variables . Parameters are $C = 1000$, $m_0 = 0.3$, table (D.7) and (**??**).

precise values are not important[1]. In the code they have been set to the following values: $\tau_E = 2$, $\tau_I = 1$. The evaluation of the transduction function $H(z)$ is done through an asymptotic expansion for the complementary error function (see *Numerical Recipes - the art of scientific computing*) which yields the correct value with a precision of 10^{-7}. Integrals are calculated with Simpson algorithm.

Population distribution of activity

Once that m_A and q_A are determined for population A, the distribution of activity is fully specified as seen in (1.66), one only need to find $x(m)$, inverting the transduction function $m(x) = H(A+Bx)$ with respect to x: this has been done finding the root of $f(x) = H(A + Bx) - m$ with a bisection algorithm.

The *Cumulative distribution function* was also calculated. In some parameters ranges the distribution of activity diverges in $m = 0$ or $m = 1$ (or both) making the direct numerical integration of (1.66) quite problematic. The easiest way to overcome this problem is to exploit the identity

$$C\rho(m) \equiv \int_0^m \rho(\mu)d\mu = \int_{x(m)}^{+\infty} \frac{e^{x^2/2}}{\sqrt{2\pi}}dx$$

where $x(m)$ is the inverse function of $m(x)$ as defined before.

[1]It has been seen that when $\tau_E = \tau_I$ trajectories could become highly irregular and do not converge to stationary points.

D.2 Network Simulations

Architecture generation

Network architectures are generated according to the probability distribution described in the previous section. Figure D.2 shows how the connections among populations are distributed in a random network generated with the simulation code.

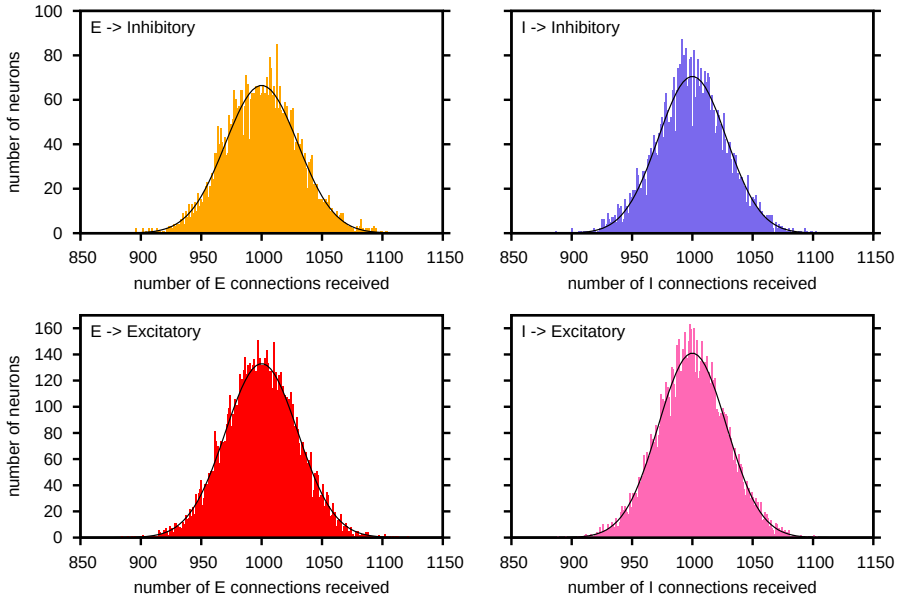

Figure D.2: The statistics of connectivity among populations in a network with $N_E = 10000$ and $N_I = 5000$ generated with the code employed for the simulations.

Network state initialization

The network is initialized assigning the state S_i of each neuron in population A the value 1 with probability $m_A(0)$ and calculating accordingly the initial field on each postsynaptic neuron.

Time evolution

The system has a single timescale determined by the ratio τ_E/τ_I, at every time-step a single neuron is updated and it is selected in the following manner: a random number x in $[0;1]$ is extracted, if $x < N_E/(N_E + N_I\tau_E/\tau_I)$ an excitatory neuron picked at random is updated, otherwise an inhibitory neuron.

The information on the net architecture was stored in the following way: each neuron is associated to a list where the index and the population of all the postsynaptic neurons is specified; in this way, whenever the state of the

neuron i_A is updated, the information on the new state is "feeded" directly to all (and only) its postsynaptic neurons just by browsing the list.

Observable quantities

During simulations the following time-dependent quantities are calculated at each timestep:

Population averaged instantaneous activity
it is calculated for both populations (A = E, I) as

$$m_A(t) = \frac{1}{N_A} \sum_{i_A=1}^{N_A} \sigma_{i_A}(t)$$

Recording these quantities at *every* time-step is unnecessary and quite memory consuming (especially if the network is large and the simulations is long). It seems more reasonable to calculate ad record it every given fraction of τ. In the simulations presented here the global activity was sampled every $\tau/10$.

Local time-averaged activity
This quantity represent the temporal average of the activity of th i^{th} neuron in population A from a certain time-step $\hat{t}$ to the end of the simulation $t = T$, formally:

$$m_{i_A} \equiv \langle \sigma_{i_A}(t) \rangle = \frac{1}{T - \hat{t}} \sum_{t=\hat{t}}^{T} \sigma_{i_A}(t)$$

In the simulation, at every time-step starting from $\hat{t}$, the "instantaneous time average" is calculated iteratively :

$$\overline{m}_{i_A}(t) = \frac{(t - 1 - \hat{t})\overline{m}_{i_A}(t-1) + \sigma_{i_A}(t)}{t - \hat{t}}$$

it is obvious that $\overline{m}_{i_A}(T)$ corresponds to the m_{i_A} defined above.

Final output

At the end of simulation, the following quantities are computed:

- $\mathbf{m_A}$. The population average of the local time-averaged activities, simply:

$$m_A = \frac{1}{N_A} \sum_{i_A=1}^{N_A} m_{i_A} \tag{D.2}$$

- $\mathbf{q_A}$. The population average of the squared local time-averaged activities:

$$q_A = \frac{1}{N_A} \sum_{i_A=1}^{N_A} (m_{i_A})^2 \tag{D.3}$$

- **Distribution of local time-averaged activity**
  The derivation of a population distribution was done in the following way: the domain of the possible values of m_{i_A} - the interval [0;1] - was partitioned in n_B equal "bins" of width $\Delta = 1/n_B$, then the number of local activities falling in each bin was assigned to the interval covered the bin thus yielding a step-function; each number of occurrence is then divided by $(N_E\Delta)$ so that the integral of the distribution results normalized to 1. Formally it can be written:

$$\begin{aligned} &\rho_A(m) = \frac{1}{N_E\Delta} \sum_{i=1}^{N_A} \Theta\Big[m_{i_A} - n\Delta\Big]\Theta\Big[\Delta(n+1) - m_{i_A}\Big] \\ &\text{for} \quad m \in [\, n\Delta;\ (n+1)\Delta\,] \end{aligned} \tag{D.4}$$

  where Θ is the step function and n runs from 0 to $n_B - 1$.

  The *cumulative* distribution was also calculated:

$$\begin{aligned} &C\rho_A(m) \equiv \int_0^m \rho(x)dx = \frac{1}{N_E\Delta} \sum_{i=1}^{N_A} \Theta\Big[\Delta(n+1) - m_{i_A}\Big] \\ &\text{for} \quad m \in [\, n\Delta;\ (n+1)\Delta\,] \end{aligned} \tag{D.5}$$

Network size and fluctuations

It is interesting to compare the results of the mean field theory with the simulation of a detailed network. For the purpose the MF equations have been integrated numerically () and three networks with sizes:

$$N_1 = 10000 \qquad N_2 = 20000 \qquad N_3 = 30000 \tag{D.6}$$

have been simulated each with the following protocol:

- **External input**. simulations were performed at four different values of the (constant) external input m_0: 0.1, 0.2, 0.3, 0.4
- **Initialization**. The initial network state have was set putting in the active state $0.2N$ excitatory and $0.3N$ inhibitory neurons picked at random; all the other where set in the inactive state.
- **Duration**. Overall simulation duration have been set to $1000\tau_E$; the averages of the observable quantities have been computed starting after $30\tau_E$.

The time constant of the two neuron was set to

$$\boxed{\tau_E = 2\tau_I}$$

, the value of all the other network parameters is listed in the following table:

$J_{EE} = 1$	$J_{EI} = 2$	$E = 2.5$	$\theta_E = 1$
$J_{IE} = 1$	$J_{II} = 1.8$	$I = 2.15$	$\theta_I = 0.7$

(D.7)

Average quantities: m **and** q

A total of 12 simulations have been performed, the quantities m and q (see (D.2) and (D.3) in D.2) have been calculated and reported in tables (D.8, D.9, D.10, D.11); the theoretical values derived with the MF theory for the correspondent values of m_0 are shown in the last column of each table.

m_E

m_0	N_1	N_2	N_3	MF theory
0.1	0.10952	0.11113	0.11088	0.11338
0.2	0.25741	0.25895	0.25826	0.26028
0.3	0.41821	0.41705	0.41754	0.42072
0.4	0.61220	0.61455	0.61032	0.61462

(D.8)

m_I

m_0	N_1	N_2	N_3	MF theory
0.1	0.18103	0.18235	0.18206	0.18347
0.2	0.37595	0.37722	0.37712	0.37785
0.3	0.57396	0.57355	0.57389	0.57476
0.4	0.78286	0.78298	0.78139	0.78258

(D.9)

q_E

m_0	N_1	N_2	N_3	MF theory
0.1	0.02579	0.02638	0.02644	0.02665
0.2	0.14585	0.14845	0.14624	0.14736
0.3	0.32687	0.32322	0.32471	0.32808
0.4	0.56417	0.56341	0.55891	0.56326

(D.10)

q_I

m_0	N_1	N_2	N_3	MF theory
0.1	0.05591	0.05653	0.05613	0.05765
0.2	0.24049	0.24449	0.24250	0.24370
0.3	0.47980	0.47600	0.47990	0.48144
0.4	0.74450	0.74425	0.74212	0.74283

(D.11)

Data are also represented in the graphics of figure D.3 together with the full curves found with the MF theory.

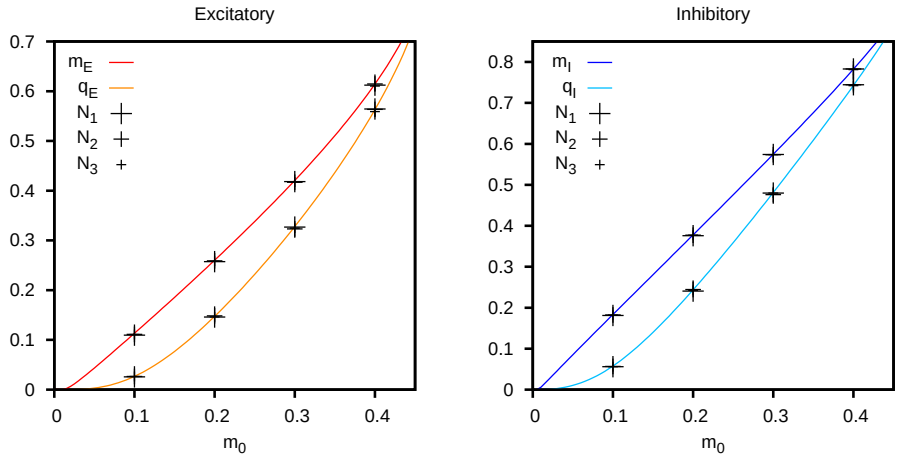

Figure D.3: *Continuous lines*: macroscopic variables found with mean field theory. *Crosses*: results of the simulations with the three networks (D.6)

It worth showing the time course of the population activities during a typical simulation to appreciate the amplitude of the temporal fluctuations for the three networks (figure D.4).

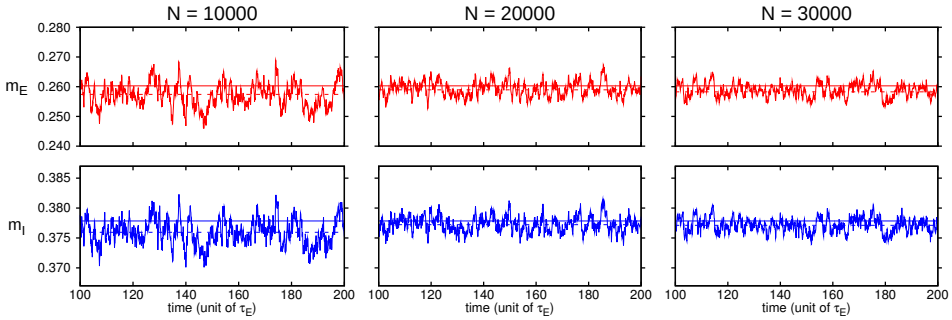

Figure D.4: Time course of the population averaged activity (see D.2) for both populations in the three networks considered with $m_0 = 0.2$. *Horizontal continuous lines*: time averages found with MF theory; *dashed lines*: time averages of the activities recorded in the simulations - numerical values of those averages can be read on the second line of tables (D.8) and (D.9)

D.3 Random Numbers generation

In stochastic processes simulation, a "good" random numbers generation algorithm is of fundamental importance. Obviously, the output numbers will never be truly random because they are the outcome of a set of computational operations.

A *pseudorandom number generator* (PRNG) is an algorithm for generating a sequence of numbers that approximates the properties of random numbers. The sequence is not truly random in that it is completely determined by a relatively small set of initial values, called the PRNG's state. A PRNG can be started from an arbitrary starting state using a *seed* state. It will always produce the same sequence thereafter when initialized with that state. The maximum length of the sequence before it begins to repeat is the *period* and it is determined by the size of the state, measured in bits. Although PRNGs will repeat their results after they reach the end of their period, a repeated result does not imply that the end of the period has been reached, since its internal state may be larger than its output.

Mersenne Twister PRNG Mersenne Twister (MT) is a widely-used fast pseudorandom number generator developed by Makoto Matsumoto and Takuji Nishimura in 1997 based on 32-bit operations. It has the period of $2^{19937} - 1$ iterations ($> 43 \times 10^{6,000}$), is proven to be equidistributed in (up to) 623 dimensions (for 32-bit values), and runs faster than other statistically reasonable generators. It is now increasingly becoming the random number generator of choice for statistical simulations and generative modeling. Recently, a 128-bit based PRNG has been presented, it is analogous to MT but making full use of Single Instruction Multiple Data (SIMD) operations of modern CPU (i.e., 128-bit) [131]. *SIMD-oriented Fast Mersenne Twister* (SFMT) is faster than earlier MT even if it's not compiled with SIMD support.

Uniformly distributed Random Numbers

A PRNG that produces uniformly distributed numbers in the interval $(0, 1)$ is a basic tool for several application, in particular for the generation of random numbers with distribution other than flat (es. Gaussian, Poissonian).

The choice for the simulations presented in this work has been the most recent version of the SFMT algorithm: dSFMT [2], it directly generates double precision floating point pseudorandom numbers which distributes uniformly in the range (0, 1). The algorithm support 10 different periods from $2^{521} - 1$ to $2^{216091} - 1$; in the simulations the parameters are set in order to get a period of $2^{19937} - 1$ for faster performance.

[2] version 2.1 was released on 4/18/2009 and is freely downloadable online at `http://www.math.sci.hiroshima-u.ac.jp/ m-mat/MT/SFMT/index.html`

Bibliography

[1] Joaquin M. Fuster and G. E. Alexander. Neuron Activity Related to Short-Term Memory. *Science*, 173(3997):652–654, aug 1971.

[2] Yasushi Miyashita and Han Soo Chang. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331(6151):68–70, jan 1988.

[3] S. Funahashi, C.J. Bruce, and Patricia S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331, 1989.

[4] Yuji Naya, Masatoshi Yoshida, and Yasushi Miyashita. Forward processing of long-term associative memory in monkey inferotemporal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(7):2861–2871, 2003.

[5] Patricia S. Goldman-Rakic. Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Comprehensive Physiology*, 1987.

[6] Daniel J. Amit. The Hebbian paradigm reintegrated: Local reverberations as internal representations. *Behavioral and Brain Sciences*, 18(04):617, feb 1995.

[7] Patricia S. Goldman-Rakic. Cellular basis of working memory. *Neuron*, 14(3):477–485, mar 1995.

[8] Xiao-Jing Wang. Synaptic reverberation underlying mnemonic persistent activity. *Trends in Neurosciences*, 24(8):455–463, aug 2001.

[9] Daniel Durstewitz, J K Seamans, and Terrence J. Sejnowski. Neurocomputational models of working memory. *Nature neuroscience*, 3 Suppl:1184–91, 2000.

[10] Donald Olding Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, 1949.

[11] S J Martin, P D Grimwood, and R G Morris. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience*, 23(Hebb 1949):649–711, 2000.

[12] Tomonori Takeuchi, Adrian J Duszkiewicz, and Richard G M Morris. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 369(1633):20130288, jan 2014.

[13] Shun Ichi Amari. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers*, C-21(11):1197–1206, 1972.

[14] W. A. Little. The existence of persistent states in the brain. *Mathematical Biosciences*, 19(1-2):101–120, 1974.

[15] J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558, 1982.

[16] Daniel J. Amit, Hanoch Gutfreund, and Haim Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32(2):1007–1018, 1985.

[17] Daniel J. Amit and Nicolas Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex*, 7(3):237–52, apr 1997.

[18] Nicolas Brunel and Xiao-Jing Wang. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *Journal of Computational Neuroscience*, 11(1):63–85, 2001.

[19] Carl van Vreeswijk and Haim Sompolinsky. Irregular activity in large networks of neurons. In C.C. Chow, B Gutkin, David Hansel, Claude Meunier, and J. Dalibard, editors, *Methods and Models in Neurophysics, Les Houches 2003 Session LXXX*, pages 345–406. Elsevier, 2005.

[20] Alfonso Renart, Rubén Moreno-Bote, Xiao-Jing Wang, and Néstor Parga. Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural computation*, 19(1):1–46, 2007.

[21] Yasser Roudi and Peter E Latham. A balanced memory network. *PLoS computational biology*, 3(9):1679–700, 2007.

[22] Yasushi Miyashita. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335(6193):817–20, oct 1988.

[23] Rishidev Chaudhuri and Ila Fiete. Computational principles of memory. *Nature Neuroscience*, 19(3):394–403, feb 2016.

[24] RS Zucker. Short-term synaptic plasticity. *Annual review of neuroscience*, 12(1):13–31, 1989.

[25] Carl van Vreeswijk and Haim Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–71, aug 1998.

[26] Eric R. Kandel, James Harris Schwartz, and Thomas M. Jessell. *Principles of neural science*. McGraw-Hill,, 4 edition, 2000.

[27] M Abeles. *Corticonics: Neural Circuits of the Cerebral Cortex*, volume 6. Cambridge University Press, 1991.

[28] Patric Hagmann, Leila Cammoun, Xavier Gigandet, Reto Meuli, Christopher J Honey, Van J Wedeen, and Olaf Sporns. Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159, jul 2008.

[29] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews. Neuroscience*, 10(3):186–98, mar 2009.

[30] Rodney J Douglas and Kevan a.C. Martin. Neuronal circuits of the neocortex. *Annual review of neuroscience*, 27:419–51, 2004.

[31] Valentino Braitenberg and Almut Schüz. *Cortex: Statistics and Geometry of Neuronal Connectivity*. Springer, 1998.

[32] Vernon B. Mountcastle. Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J Neurophysiol*, 20(4):408–434, 1957.

[33] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154.2, 1962.

[34] Vernon B. Mountcastle. An organizing principle for cerebral function: The unit model and the distributed system. In G. M. Edelman and V. B. Mountcastle, editors, *The mindful brain: Cortical organization and the group-selective theory of higher brain function.* MIT press, 1978.

[35] Valentino Braitenberg. Cortical architectonics: General and areal. In M.A.B. Brazier and H. Petsche, editors, *Architectonics of the cerebral cortex*, pages 443—-465. Raven Press, New York, 1978.

[36] Vernon B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120 (Pt 4:701–22, apr 1997.

[37] JS Griffith and G. Horn. An analysis of spontaneous impulse activity of units in the striate cortex of unrestrained cats. *The Journal of Physiology*, 186(3):516, 1966.

[38] H Noda and W Ross Adey. Firing variability in cat association cortex during sleep and wakefulness. *Brain research*, 18(3):513–26, mar 1970.

[39] B. D. Burns and A. C. Webb. The spontaneous activity of neurones in the cat's cerebral cortex. *Proceedings of the Royal Society of London. Series B, Biological sciences*, 194(1115):211–23, oct 1976.

[40] R Vogels. The response variability of striate cortical neurons in the behaving monkey. *Experimental Brain Research*, 1989.

[41] Rodney J Douglas, Kevan a.C. Martin, and David Whitteridge. An intracellular analysis of the visual responses of neurones in cat visual cortex. *Journal of Physiology*, 440:659–696, 1991.

[42] M.N. Shadlen and W.T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579, 1994.

[43] GR Holt and W R Softky. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of . . .*, 1996.

[44] W. Bair and C. Koch. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural computation*, 8(6):1185–1202, 1996.

[45] W R Softky and Christof Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1):334, 1993.

[46] M.N. Shadlen and W.T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(10):3870–96, may 1998.

[47] KW Koch and Joaquin M. Fuster. Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Experimental Brain Research*, 76(2):292–306, 1989.

[48] T. Hromádka, M.R. DeWeese, and Anthony M. Zador. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS biology*, 6(1):e16, 2008.

[49] DH O'Connor and SP Peron. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron*, 2010.

[50] S. Shoham, D.H. O'Connor, and Ronen Segev. How silent is the brain: is there a "dark matter" problem in neuroscience? *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 192(8):777–784, 2006.

[51] Alex Roxin, Nicolas Brunel, David Hansel, Gianluigi Mongillo, and Carl van Vreeswijk. On the Distribution of Firing Rates in Networks of Cortical Neurons. *Journal of Neuroscience*, 31(45):16217–16226, nov 2011.

[52] J.A. White, J.T. Rubinstein, and A.R. Kay. Channel noise in neurons. *J. Neurosci*, 16:3219–3235, 2000.

[53] Aldo A. Faisal, Luc P. J. Selen, and Daniel M. Wolpert. Noise in the nervous system. *Nature reviews. Neuroscience*, 9(4):292–303, 2008.

[54] W H Calvin and C.F. Stevens. Synaptic noise and other sources of randomness in motoneuron interspike intervals. *Journal of neurophysiology*, 31(4):574–87, jul 1968.

[55] ZF Mainen and Terrence J. Sejnowski. Reliability of spike timing in neocortical neurons. *Science*, 268(5216):1503, 1995.

[56] R J Sayer, M J Friedlander, and S J Redman. The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 10(3):826–836, 1990.

[57] M Matsumura, D Chen, T Sawaguchi, K Kubota, and E E Fetz. Synaptic interactions between primate precentral cortex neurons revealed by spike-triggered averaging of intracellular membrane potentials in vivo. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(23):7757–7767, 1996.

[58] D Lee, N L Port, W Kruse, and a P Georgopoulos. Variability and correlated noise in the discharge of neurons in motor and parietal areas of the primate cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 18(3):1161–1170, 1998.

[59] W. Bair, E Zohary, and W.T. Newsome. Correlated firing in macaque visual area MT: time scales and relationship to behavior. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 21(5):1676–1697, 2001.

[60] Christos Constantinidis and Patricia S. Goldman-Rakic. Correlated discharges among putative pyramidal neurons and interneurons in the primate prefrontal cortex. *Journal of neurophysiology*, 88(6):3487–3497, 2002.

[61] William R Softky and Christof Koch. Cortical Cells Should Fire Regularly, But Do Not. *Neural Computation*, 4(5):643–646, sep 1992.

[62] Edmund T. Rolls and M. J. Tovee. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of neurophysiology*, 73(2):713–26, feb 1995.

[63] P Foldiak and Peter Földiak. Sparse coding in the primate cortex. *The Handbook of Brain Theory and Neural Networks*, page 7, 2002.

[64] B.A. Olshausen and D.J. Field. Sparse coding of sensory inputs. *Current opinion in neurobiology*, 14(4):481–487, 2004.

[65] C.F. Stevens, Anthony M. Zador, and Others. Input synchrony and the irregular firing of cortical neurons. *Nature neuroscience*, 1(3):210–217, 1998.

[66] R Azouz and C M Gray. Cellular mechanisms contributing to response variability of cortical neurons in vivo. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 19(6):2209–23, mar 1999.

[67] Todd W. Troyer and Kenneth D. Miller. Physiological Gain Leads to High ISI Variability in a Simple Model of a Cortical Regular Spiking Cell. *Neural Computation*, 9(5):971–983, jul 1997.

[68] Emilio Salinas and Terrence J. Sejnowski. Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience*, 2(8):539–550, 2001.

[69] Charles M Gray, Peter König, Andreas K Engel, and Wolf Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–337, mar 1989.

[70] E. Vaadia, I Haalman, M Abeles, H Bergman, Y Prut, H Slovin, and A Aertsen. Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature*, 373(6514):515–8, feb 1995.

[71] Alfonso Renart, Jaime de la Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex D. Reyes, and Kenneth D Harris. The Asynchronous State in Cortical Circuits. *Science*, 327(5965):587–590, jan 2010.

[72] Misha V. Tsodyks and Terrence J. Sejnowski. Rapid state switching in balanced cortical network models. *Network: Computation in Neural Systems*, 6(2):111–124, 1995.

[73] Carl van Vreeswijk and Haim Sompolinsky. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science*, 274(5293):1724–1726, dec 1996.

[74] Y Shu. Turning on and off recurrent balanced cortical activity. *Nature*, 2003.

[75] Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, and David a McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(17):4535–45, apr 2006.

[76] M. Okun and I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature neuroscience*, 11(5):535–537, 2008.

[77] J.S. Isaacson and Massimo Scanziani. How Inhibition Shapes Cortical Activity. *Neuron*, 72(2):231–243, 2011.

[78] G.L. Gerstein and B. Mandelbrot. Random walk models for the spike activity of a single neuron. *Biophysical Journal*, 4(1):41–68, 1964.

[79] Alexander Lerchner, Cristina Ursta, John A. Hertz, Mandana Ahmadi, Pauline Ruffiot, and Soren Enemark. Response Variability in Balanced Cortical Networks. *Neural Computation*, 18(3):634–659, mar 2006.

[80] WS McCulloch and W Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biology*, 5(4):115–133, 1943.

[81] SI Amari. Homogeneous nets of neuron-like elements. *Biological Cybernetics*, 1975.

[82] J J Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 81(10):3088–92, may 1984.

[83] C.C. Chow and J.A. White. Spontaneous action potentials due to channel fluctuations. *Biophysical journal*, 71(6):3013–21, dec 1996.

[84] Elad Schneidman, B Freedman, and Idan Segev. Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural computation*, 10(7):1679–703, oct 1998.

[85] S Cash and Rafael Yuste. Linear summation of excitatory inputs by CA1 pyramidal neurons. *Neuron*, 22(2):383–394, 1999.

[86] Roberto Araya, Kenneth B Eisenthal, and Rafael Yuste. Dendritic spines linearize the summation of excitatory potentials. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49):18799–18804, 2006.

[87] Daniel J. Amit. *Modeling Brain Funtion*. Cambridge University Press, 1989.

[88] B. Derrida, E. Gardner, and A. Zippelius. An exactly solvable asymmetric neural network model. *EPL (Europhysics Letters)*, 4:167, 1987.

[89] a Sawatari and E M Callaway. Diversity and cell type specificity of local excitatory connections to neurons in layer 3B of monkey primary visual cortex. *Neuron*, 25(2):459–471, 2000.

[90] HJ Hilhorst, M. Nijmeijer, and Others. On the approach of the stationary state in Kauffman's random Boolean network. *Journal De Physique*, 48(2):185–191, 1987.

[91] B. Derrida, G. Weisbuch, and Others. Evolution of overlaps between configurations in random Boolean networks. *Journal de physique*, 47(8):1297–1303, 1986.

[92] Larry F. Abbott and Carl van Vreeswijk. Asynchronous states in networks of pulse-coupled oscillators. *Physical Review E*, 48(2):1483–1490, aug 1993.

[93] Wulfram Gerstner and J.L. van Hemmen. Coherence and incoherence in a globally coupled ensemble of pulse-emitting units. *Physical Review Letters*, 71(3):312–315, jul 1993.

[94] David Hansel, G Mato, and C Meunier. Synchrony in excitatory neural networks. *Neural Computation*, 1995.

[95] SJ Thorpe, Edmund T. Rolls, and S. Maddison. The orbitofrontal cortex: Neuronal activity in the behaving monkey. *Experimental Brain Research*, 49(1), jan 1983.

[96] GH Recanzone. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *Journal of . . .*, 2000.

[97] S. Funahashi. Neuronal activity related to saccadic eye movements in the monkey's dorsolateral prefrontal cortex. *Journal of . . .*, 1991.

[98] Albert Compte, Christos Constantinidis, J Tegner, S. Raghavachari, M.V. Chafee, Patricia S. Goldman-Rakic, and Xiao-Jing Wang. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of Neurophysiology*, 90(5):3441, 2003.

[99] M. Shafi, Y. Zhou, J. Quintana, C.C. Chow, Joaquin M. Fuster, and M. Bodner. Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience*, 146(3):1082–1108, 2007.

[100] AF Dean. The variability of discharge of simple cells in the cat striate cortex. *Experimental Brain Research*, 44(4):437–440, 1981.

[101] WE Vinje and JL Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 2000.

[102] John A. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the theory of neural computation*. Addison Wesley, 1991.

[103] Edmund T. Rolls and Alessandro Treves. *Neural networks and brain function*. Oxford University Press, 1998.

[104] Daniel J. Amit and Alessandro Treves. Associative memory neural network with low temporal spiking rates. *Proceedings of the National Academy of Sciences*, 86(20):7871, 1989.

[105] Nava Rubin and Haim Sompolinsky. Neural Networks with Low Local Firing Rates. *Europhysics Letters (EPL)*, 10(5):465–470, nov 1989.

[106] Wulfram Gerstner and J.L. van Hemmen. Associative memory in a network of spiking neurons, may 1992.

[107] Alfonso Renart, Nicolas Brunel, and Xiao-Jing Wang. Chapter 15 Mean-Field Theory of Irregularly Spiking Neuronal Populations and Working Memory in Recurrent Cortical Networks. In Jianfeng Feng, editor, *Computational Neuroscience A Comprehensive Approach*, chapter 15, pages 431–490. Chapman & Hall/CRC, 2004.

[108] Emanuele Curti, Gianluigi Mongillo, G.L. Camera, and Daniel J. Amit. Mean field and capacity in realistic networks of spiking neurons storing sparsely coded random memories. *Neural computation*, 16(12):2597–2637, 2004.

[109] Nicolas Brunel. Persistent activity and the single-cell frequency-current curve in a cortical network model. *Network: Computation in Neural Systems*, 11(4):261–280, nov 2000.

[110] Francesca Barbieri and Nicolas Brunel. Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Frontiers in Neuroscience*, 2(JUL):114–122, 2009.

[111] Earl K Miller, C a Erickson, and R Desimone. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience*, 16(16):5154–5167, 1996.

[112] Albert Compte. Temporally Irregular Mnemonic Persistent Activity in Prefrontal Neurons of Monkeys During a Delayed Response Task. *Journal of Neurophysiology*, 90(5):3441–3454, 2003.

[113] D J Willshaw, O P Buneman, and H C Longuet-Higgins. Non-holographic associative memory. *Nature*, 222(5197):960–2, jun 1969.

[114] D Golomb, Nava Rubin, and Haim Sompolinsky. Willshaw model: Associative memory with sparse coding and low firing rates. *Physical Review A*, 41(4):1843–1854, feb 1990.

[115] M P Young and S Yamane. Sparse population coding of faces in the inferotemporal cortex. *Science (New York, N.Y.)*, 256(5061):1327–31, may 1992.

[116] Robert S. Zucker and Wade G. Regehr. Short-term synaptic plasticity. *Annual review of physiology*, 64(1):355–405, mar 2002.

[117] Diasynou Fioravante and Wade G. Regehr. Short-term forms of presynaptic plasticity. *Current Opinion in Neurobiology*, 21(2):269–274, apr 2011.

[118] L. F. Abbott, J A Varela, K Sen, and S B Nelson. Synaptic depression and cortical gain control. *Science (New York, N.Y.)*, 275(5297):220–4, jan 1997.

[119] Misha V. Tsodyks, K Pawelzik, and Henry Markram. Neural networks with dynamic synapses. *Neural Computation*, 10(4):821–835, 1998.

[120] Galit Fuhrmann, Henry Markram, and Misha V. Tsodyks. Spike frequency adaptation and neocortical rhythms. *Journal of neurophysiology*, 88(2):761, 2002.

[121] Larry F. Abbott and Wade G. Regehr. Synaptic computation. *Nature*, 431(7010):796–803, oct 2004.

[122] Francesca Barbieri and Nicolas Brunel. Irregular persistent activity induced by synaptic excitatory feedback. *Frontiers in computational neuroscience*, 1(November):5, 2007.

[123] Gianluigi Mongillo, Omri Barak, and Misha V. Tsodyks. Synaptic theory of working memory. *Science*, 319(5869):1543–6, mar 2008.

[124] David Hansel and German Mato. Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks. *Journal of Neuroscience*, 33(1):133–149, jan 2013.

[125] Misha V. Tsodyks and Henry Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proceedings of the National Academy of Sciences*, 94(2):719–723, 1997.

[126] Ling-Gang Wu and J.Gerard G Borst. The Reduced Release Probability of Releasable Vesicles during Recovery from Short-Term Synaptic Depression. *Neuron*, 23(4):821–832, aug 1999.

[127] H R Wilson and J D Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal*, 12(1):1–24, 1972.

[128] I Bena. Dichotomous Markov Noise: exact results for out-of-equilibrium systems. *International Journal of Modern Physics B*, 2006.

[129] V Balakrishnan. On a simple derivation of master equations for diffusion processes driven by white noise and dichotomic Markov noise. *Pramana*, 40(4):259–265, apr 1993.

[130] Gianluigi Mongillo, David Hansel, and Carl Van Vreeswijk. Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission. *Physical Review Letters*, 108(15):13–17, 2012.

[131] Mutsuo Saito and M Matsumura. Monte Carlo and Quasi-Monte Carlo Methods 2006. In Alexander Keller, Stefan Heinrich, and Harald Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 607–622. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

Aknowledgments

My uttermost gratitude goes to Gianluigi Mongillo for having guided me throughout the long and rough path of this thesis with consideration, understanding and trust. A heartfelt thanks to Paolo Rossi for his kind attention and availability in supervising this work. Thank you also to David Hansel and Carl van Vreeswijk for their useful comments and indications.