

A Neuron Model Independent Path Integral Explored via Binary Assemblies

Master thesis

by Christian Keup

under supervision of

Tobias Kühn &
Prof. Dr. Moritz Helias

Institute of Computational and Systems Neuroscience (INM-6) &
Institute for Advanced Simulation (IAS-6) &
JARA Brain Institute I, Jülich Research Centre
and Department of Physics, Faculty 1, RWTH Aachen University

30th of November, 2017

CONTENTS

I	Introduction	3
I.	Summary	3
II.	Basics of Biological Neuronal Networks and their Theoretical Description	3
A.	Introduction	3
B.	Local network activity	5
C.	Neuron models	5
III.	The Assembly Hypothesis	6
IV.	Advance Considerations on the Theoretical Approach	7
II	Mean Field Treatment and Simulation of Assembly Model	8
V.	Mean Field Formulation	8
A.	Model description	8
B.	Derivation of mean field equations	9
C.	Mean field equations for assembly model	10
VI.	Parameter Choices	10
VII.	Emergence of the Second Attractor	11
A.	Potential landscape	12
B.	The second attractor cannot emerge at activities $m \leq 0.5$	13
VIII.	Spontaneous Transitions	15
A.	Flatness of the potential	16
IX.	Pairwise Correlations	18
A.	Numerical solution of equations for linearized correlations	18
B.	Separation of equations by neglecting feedback of assembly to bath	19
III	Path Integral Formulation	19
X.	A Neuron Model Independent Path Integral Formulation	20
XI.	Derivation of MFT as Tree Level Approximation for one Population	21
A.	Time dependent mean activity	22
B.	Time lagged covariance for the stationary state	24
C.	Fixed vs. free indegree and the connection to standard MFT	26
XII.	Propagators	27
XIII.	1-Loop Correction to the Mean Activity	31
XIV.	Generalization to Multiple Populations	33
A.	Solution in tree level approximation	34
XV.	Clustering by connection probability vs. synapse strength	35

	2
XVI. Other Neuron Models	36
IV Discussion	38
XVII. Relation to other Path Integral Formalisms	39
A. Overview of approaches in the literature	39
B. Relationship to the presented approach	39
XVIII. Strengths and Weaknesses of the new Formalism	40
XIX. General Discussion of the Assembly Model	40
XX. Neuroscientific Implications for Experiment and Modeling of Cell Assemblies	41
XXI. Acknowledgments	41
V Appendix	41
XXII. Simulation and Data Analysis	42
A. Simulation	42
B. Bifurcation diagram	42
C. Extraction of single neuron correlations from population sums	43
XXIII. Supplementary Calculations Mean Field Treatment	43
A. Numerical solution of equations for linearized correlations	43
B. Analytical solution by neglecting feedback of assembly to bath	44
Covariances of the homogeneous EI-network	45
Covariances of the assembly subnetwork	46
XXIV. Supplementary Calculations Path Integral Formulation	47
A. Path integral with discretized time	47
B. Exact equation for $Q_1^*(t, s)$	48
C. Finding $S_{12}^{(RR)}(t, s)$ by explicit calculation of the expectation value	51
D. Disorder average for fixed indegree	52
E. Some further calculations in greater detail	54
XXV. References	55
References	55

Part I

Introduction

I. SUMMARY

We present the basic exploration of a novel path integral formulation for models of biological neuronal networks that allows to keep the neuron model unspecified until quantities are explicitly calculated. This is done at the example of a binary neuron network containing an assembly, which is suited to discuss the limits of standard mean field theory and the feasibility of a path integral approach, while also being of some neuroscientific interest.

Advanced theoretical approaches to the description of neuronal network activity are still in their infancy, but much needed, due to its nonlinearities, statistical nature, and nonequilibrium dynamics. There is thus now renewed interest in the transfer of mathematical tools developed in statistical physics to the theory of neural networks. We model a Hebbian cell assembly as a group of $O(100)$ excitatory binary neurons with increased coupling embedded in a larger balanced random network. Using standard mean-field theory and simulation, the system properties and parameter dependencies are analysed, especially emergence of the high activity state, spontaneous transitions, and pairwise correlations. We then introduce the path integral formulation, applying it to a single population of binary neurons. We show the relation of the tree level approximation to mean field theory, calculate propagators and a 1-loop diagram, and generalize to multiple populations. The formulation specifically uses generic properties of neuronal networks, which allows the formal description of the systems properties before an effective neuron model is fixed. It is analytically feasible for rate, binary and, possibly, spiking neurons. The implications of the results for the assembly model and the relation of our formulation to other path integral approaches are discussed. We stress that a functional form unlocks tools to treat critical phenomena, large fluctuations, disorder and may lead eventually to effective coarse grained theories by using renormalization group methods.

II. BASICS OF BIOLOGICAL NEURONAL NETWORKS AND THEIR THEORETICAL DESCRIPTION

A. Introduction

The human brain shows no obvious qualitative difference to the brain of other mammals¹. The main novelty of the mammalian brain compared to reptiles is the cerebral cortex, a sheet of neurons covering the subcortical structures and nuclei (Figure 1, Left). While most subcortical structures have a specialized function, various parts ('areas') of the neocortex (forming the well known folded surface of the human brain) specialize on such diverse functions as sensory perception, motor control, language comprehension and generation, planning and logical reasoning, while its architecture is nearly the same. This supports the view that a patch of neocortex forms a generic computing unit that specializes to a task according to its external inputs. How this so called microcircuit works, however, is unknown. It is organized in six visually distinguishable layers that differ in the neuron types contained and their connection architecture. To better understand the difficulties of determining the exact structure and operational principles of the cortex, consider the structure and function of a single cortical neuron (see Figure 1, Right):

There are three main parts; the dendrite, where inputs arrive via synapses from other neurons and are led to the cell body (soma), which holds the cell nucleus and where all inputs are gathered and transformed into an output signal that then travels along the axon, which branches out again and forms synapses on the target neurons dendrites. The output signals, called spikes or action potentials, are short (~ 1 ms) electrical pulses that actively travel along the axon as a potential difference between intra- and extracellular medium. This is facilitated by ion channels in the cell membrane that open and close depending on the voltage, allowing the spike to propagate like a wave without dispersion or attenuation. When this all-or-nothing signal arrives at the synapses, a chemical neurotransmitter is released, which quickly dissipates over the synaptic cleft and binds to receptors on the target side. The effects depend on the neurotransmitter and receptor; but all synapses made by a neurons axon always release only one transmitter (which is known as Dale's law), allowing a classification. Most commonly, ion channels are opened, resulting either in

¹ any factual information in Part I without explicit citation can be found in the textbooks by Abeles (1991), Kandel (2013), Rolls (2016)

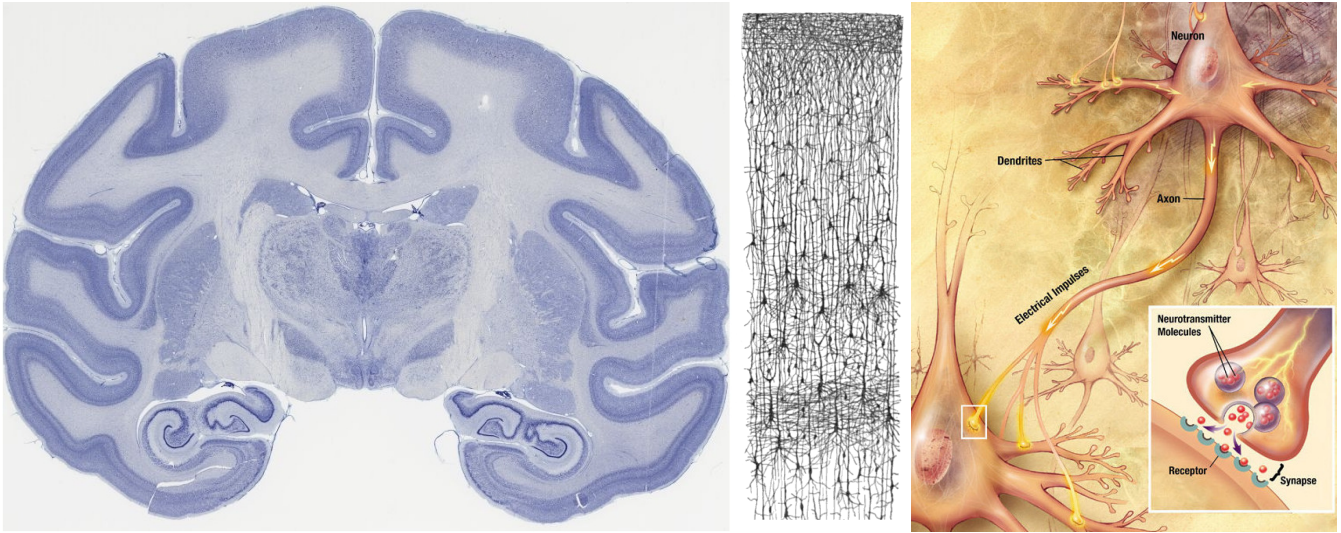


Figure 1. Left: Nissl stained coronal section of macaque monkey brain. The dark blue folded outer layer is the cerebral cortex. In the center, subcortical nuclei of Thalamus and Basal Ganglia can be seen. The high contrast folded 'extension' of the cortex in the lower part is the hippocampal formation. Image retrieved from <http://brainmaps.org/HBP-JPG/20/n19.jpg> (Mikula et al. 2007)

Middle: Historical drawing of Golgi stained neocortical slice by Santiago de Ramon y Cajal published in 1899. Only a subset of neurons is stained. (Source: Wikimedia Commons, Public Domain)

Right: Schematic drawing of neuron anatomy. (Source: Wikimedia Commons, Public Domain)

a local de- or hyperpolarization of the targeted dendrite, the strength of which depends on the specific synapse, but some receptors also have modulatory effects. The membrane potential changes combine with those elicited by other active synapses on the local dendrite and are passively conducted along the dendritic arborizations (contrary to the active propagation of spikes in the axon). If the input of many synapses combined results in the depolarization of the axon initial segment (or axon hillock) beyond a threshold (typically -55 mV , while the resting potential is -70 mV), a new spike is emitted. There are some exceptions from this description, such as electrical coupling by gap junctions (Fukuda et al. 2006), axo-axonic synapses (Kubota et al. 2016) and dendritic spikes (Manita et al. 2017).

The dendrite can extend on the order of a millimetre away from the soma (which is a few μm in size), while the axon, being typically below one μm in diameter, branches, depending on the cell type, not only in the local area of $\sim 1\text{ mm}$, but may reach to any other area of the brain, more than 10 cm away. Each neocortical neuron receives on the order of 10,000 synapses, roughly half of which come from cells in the local network, while the other half originates from other areas (Abeles 1991). The neuron density in neocortex is approximately $30,000\text{ mm}^{-3}$. Due to this multi-scale architecture, full reconstruction of only a local network by structural imaging would require to trace many thousands of tightly packed, intertwining 3D objects at nm resolution for a distance in the millimetre range. This has until now only been done for a $40 \times 40 \times 50\text{ }\mu\text{m}$ volume (Kasthuri et al. 2015). Larger scale methods are used, but cannot give the full connectivity.

Also the dynamics of a single neuron are significantly influenced by processes on several levels: short and long term plasticity on the level of synapses, nonlinear and morphology dependent integration of inputs by the dendrite (Stuart & Spruston 2015), local and global modulatory neurotransmitters, local network activity, global network activity, and structural plasticity. These effects can be modeled and experimentally accessed on their respective spatio-temporal scale, but their combination results in a tremendous multi-scale problem. However, we believe that the basic operation of cortex should follow a relatively small set of principles, which, if understood, would provide a much simplified effective picture of the observable processes, see eg. Rolls (2016). In the meantime, several levels of description exist in neuroscience, each with specific effective models of the described processes. These are: The level of individual receptors, gene expressions and molecular processes, individual synapses, dendritic dynamics, networks of single neurons, and networks of neuron populations or brain areas. The present work is concerned with networks of single neurons.

B. Local network activity

Roughly 80 % of the neocortical neurons have an excitatory (depolarizing) effect on their postsynaptic targets, while the remaining 20 % are inhibitory (hyperpolarizing), however, inhibition dominates in terms of total strength. Many details and open questions are available about different cell types and variations between layers, see for an overview Harris & Mrsic-Flogel (2013), Hattori et al. (2017), and there are attempts to simulate the microcircuit at the single neuron level with increasing complexity (Markram et al. 2015, Potjans & Diesmann 2014). However, we are here interested in a simplified model of the local network in a single neocortical layer, where we only distinguish between an excitatory and an inhibitory population. The network activity of pyramidal neurons (being the most common excitatory neuron in layer 2/3 and 5) in vivo is characterized by seemingly random firing (termed asynchronous-irregular) with rates of 1 to 30 Hz (in layer 2/3 there is also a large percentage of nearly silent neurons (Barth & Poulet 2012)). In the presence of sleep or anesthesia large oscillations are present, and, depending on the cortical area and layer, information processing may result in strong transient activity changes. In the quiet, wakeful state however, the network settles into a stationary state in which mean activities are stable and the correlation between neuron pairs is very low. The membrane potential at the soma has been found to fluctuate strongly around a value below threshold, crossing it as a rare, seemingly noise driven event. This behavior is due to a dynamic balancing of excitatory and inhibitory inputs on a millisecond scale, so that the neuron is not driven by the mean input, but by the input variance. It can be reproduced by sparse random networks of simplified neurons that neglect synaptic and dendritic dynamics, and has been termed the balanced state (Brunel 2000, Renart et al. 2010, van Vreeswijk & Sompolinsky 1998). The effect arises if inhibition dominates excitation, which causes a decorrelation of responses and a reduced effective time constant (Renart et al. 2010, Tetzlaff et al. 2012). The network model we will use in this work is of this type. There is some discussion whether a distinction must be made between 'loose' and 'tight' balance, see Denève & Machens (2016), Helias et al. (2014).

C. Neuron models

A large zoo of neuron models with varying amounts of simplifications exists in the literature. They can be roughly categorized by descending complexity: multi-compartment neurons, allowing to simulate the effects of dendritic morphology, point neurons, which neglect dendrite morphology but effectively describe single neuron behavior in varying complexity, and population or neural field models, describing large numbers of neurons as a single entity. It should be noted that the various models are as yet only rarely linked by rigorous mathematical approximations (but see Grytskyy et al. (2013)), making a comparison across models difficult. In the following, the point neuron models of most relevance to the present work are shortly described.

A very successful and widely used model is the Leaky-Integrate-and-Fire neuron (LIF), that approximates the spiking behavior of biological neurons surprisingly well (Burkitt 2006). There are several extensions and variations, but we describe the simple current-based LIF neuron: The somatic membrane potential is modeled as a leaking capacitor that receives currents for each synaptic activation according to the synaptic weight (the inputs are thus summed linearly). If the membrane potential crosses threshold, it is reset to the resting potential and a delta-distribution representing the spike is sent to all postsynaptic targets. Also a refractory period can be added in which no further spike can be emitted. The equation governing the membrane potential is

$$\tau_m \frac{dV_i}{dt}(t) = -V_i(t) + \tau_m \sum_{j=1}^N J_{ij} s_j(t-d) \quad (1)$$

where τ_m is the membrane time constant, d the (optional) conduction delay, J_{ij} the synaptic weight and $s_j(t) = \sum_k \delta(t - t_k^j)$ the incoming spike trains from other neurons. Reset and spike emission are handled 'manually'. The activity in a random network of these neurons can be described in mean field approximation by a Fokker-Plank equation. This model was used for the to date largest simulation of cortical activity in numbers of neurons by Schmidt et al. (2016).

Mathematically, it is expensive to model membrane potential and spike times as two variables with complicated dependence for each neuron. It is thus attractive to describe only the neuron activity without explicitly modeling the membrane potential. This is achieved by the binary neuron with asynchronous update : Each neuron is represented

by a binary variable as either active, $x_i = 1$, or silent, $x_i = 0$. The state of a neuron is decided at update by an activation function describing the probability of the neuron being active for the given input:

$$x_i(t_k) = F\left(\sum_j J_{ij}x_j(t_k - d)\right). \quad (2)$$

The leaky membrane is effectively modeled by updating each neuron at random times, given by a Poisson process with mean frequency $\frac{1}{\tau_m}$ (asynchronous update). This means that for low mean activities, when the probability of a neuron to be active in two consecutive updates is small, each activation ('spike') influences its postsynaptic targets only on a timescale corresponding to the membrane time constant τ_m . The delay d is usually set to zero, in good approximation for a local network, so that the system is Markovian. Commonly used activation functions are sigmoidal shapes or the Heaviside function. Especially in computer science, binary neurons are also frequently used with synchronous update (all neurons are updated at each time step), where the correspondence to a LIF neuron is less clear.

The last type of neuron model we describe is the rate neuron, ubiquitously used in theoretical studies due to its mathematical simplicity: Instead of describing single 'spikes', the rate neurons activity is a continuous variable representing the firing rate. This allows to describe network as a system of coupled differential equations. An example equivalent to the model by Sompolinsky et al. (1988):

$$\partial_t x_i = -x_i + \Phi\left(\sum_{j=1}^N J_{ij}x_j\right). \quad (3)$$

Again the input is mapped to an output rate by a sigmoidal function, here $\Phi(h_j) = \tanh(gh_j)$, modeling that a neuron does not fire for small inputs, and for very large inputs the firing rate saturates. Noise can be added to model the stochastic nature of spike counts. If the activation function is linear, or linearized around the working point of a stationary solution, the system reduces to an Ornstein-Uhlenbeck process (Grytskyy et al. 2013).

III. THE ASSEMBLY HYPOTHESIS

Until now, only networks with random connectivity were discussed. But there must be some structure in the biological network that allows the brain the processing of information and learning. Several hypotheses exist about the possible cortical principles of operation and neuronal encoding of information, see i.e. associative networks (Rolls 2016), synfire chains (Abeles 1991), or predictive coding (Hohwy 2013). It is agreed upon, however, that synaptic and structural plasticity, that is the activity dependent rewiring of the network, are largely responsible for learning. Synaptic plasticity refers to the strengthening or weakening of existing synapses, while structural plasticity refers to the growth or degeneration of dendrites, axons and synapses. While exact plasticity rules are diverse and not conclusively established (Sjöström et al. 2008), an important type is spike timing dependent plasticity (STDP) between excitatory cells (Sjöström & Gerstner 2010): Suppose a connection $A \rightarrow B$; if B frequently fires shortly after receiving a spike from A , the connection is strengthened. If however B fires just before receiving a spike from A , the connection is weakened. The potentiating effect seems to be larger than the depressing one, resulting in a net increase of weights in the population. Then a homeostatic mechanism is required to normalize total synaptic strength, for which the altered network state during sleep may be a candidate (Tononi & Cirelli 2014). STDP represents a modern view of an observation made already by Hebb (1949) and others, leading to the formulation of the assembly hypothesis of sensory processing:

If neurons respond to the same stimulus with increased activity, connections between them will be formed and strengthened, resulting in a neuronal cluster (assembly). Then if only a part of the assembly members should be activated by a sensory input, they will likely cause the remaining members to spike as well due to the strong connectivity, making the assembly a much more reliable indicator of the stimulus being present than a collection of unconnected neurons. This corresponds to subjective experience: Consider seeing a cat that is partially occluded by a fence; it will nonetheless be clearly recognized as a cat, and the occluded features mentally 'added'. Similarly, a known type of object will be recognized, even if there are some alterations to it. Also, thinking of a particular subject will remind one of related topics. These effects can be understood as activation spreading in a network of many assemblies, where neurons can be members in several ones, resulting in the assemblies to be connected according

to their relatedness. The assembly hypothesis therefore provides a theory of associative processing and memory, see Palm (1982), Rolls (2016) for more sophisticated descriptions. Furthermore, an assembly may sustain itself after activation in a high activity state even in the absence of the original stimulus. This may be involved in the integration of sequential stimuli and working memory.

Experimental evidence for the existence of neuronal assemblies has long been difficult to obtain because the two main options to experimentally observe them, to test the cortical local networks for clustering and to find groups of neurons that reliably activate together in relation to a specific stimulus, are severely limited by the technological difficulty to reconstruct the connections or to record the activity of many neurons at once: Given a method to record from 10 neurons and assembly sizes of a few dozen neurons, it would be unlikely to hit several assembly members by chance when sampling from a local network of several thousand neurons. In recent years however, imaging and recording technology have evolved considerably, for example allowing to record from hundreds of neurons in animals and partly reconstruct their connectivity, see Lee et al. (2016). And indeed, starting by the observation that local connectivity between pyramidal cells in layer 5 is nonrandom, bidirectional connections are overrepresented and of higher synaptic weight than average (Song et al. 2005), it was found that the connection probability and weight between two neurons rises for increasing number of common synaptic partners (Perin et al. 2011). Furthermore, pyramidal neurons in mouse V1 layer 2/3 that show similar orientation tuning have increased connection probability, larger than average synapses and share common input from layer 4 and 2/3 (Cossell et al. 2015, Ko et al. 2013, 2011, Lee et al. 2016, Yoshimura et al. 2005). Also, groups that repeatedly activate together were found in vivo (Miller et al. 2014) and in vitro (Cossart et al. 2003, Mao et al. 2001). In mouse hippocampus, temporally ordered replay of activated groups of place cells was reported (Malvache et al. 2016). Nonetheless, the existence and functional role of Hebbian assemblies is not conclusively established. See also the reviews on attractor up states (Miller 2016), formation of assemblies (Holtmaat & Caroni 2016), their detection (Wallace & Kerr 2010), various concepts connected to the term assembly (Abeles 2011, Buzsáki 2010), physiological evidence (Huyck & Passmore 2013) and especially the general one by Palm et al. (2014).

Theoretically, it is known that an assembly can form a high activity point attractor. The problem of overlapping assemblies was investigated by studying the number of possible attractors (that is, the memory capacity) in simplified networks, see Palm (2013) for a review,. An extensively studied example is the Hopfield network (Hopfield 1982, 2007) which uses a symmetric connectivity matrix and binary or rate neurons, allowing to define an energy on the state space. The computational capabilities of interconnected associative networks were considered by Palm (1982) and have been demonstrated by an autonomous robot that is capable of simple visual perception, language recognition and action (Markert et al. 2009). In simulations, it was shown that including clustering in a random network introduces rich dynamics on slow timescales and leads to decreased variability in response to external stimuli (Litwin-Kumar & Doiron 2012). A similar model yields activity resembling the spontaneous and evoked activity in gustatory cortex (Mazzucato et al. 2015). The dynamical properties of such a network were investigated in more depth by Setareh et al. (2017), see Part IV for further discussion.

IV. ADVANCE CONSIDERATIONS ON THE THEORETICAL APPROACH

In the mathematical treatment of neuronal networks, severe simplifications are often used. Most commonly, homogeneous, randomly connected populations are assumed, pairwise correlations neglected and the input to a neuron approximated as a gaussian random variable. This we refer to as mean field theory (MFT). It should be noted, that the term is not clearly defined in neuroscience. The mean field equations are derived differently for each neuron model that allows analytical treatment, and usually not by truncating a systematic expansion of exact equations, but by rather ad-hoc assumptions, such as the self averaging property, negligible correlations and identically distributed gaussian input variables. While the validity of these approximations may often be undisputed, a problem arises when corrections or extensions to MFT are calculated, because there is no obvious systematic way to include effects up to a specific order. This problem can be completely alleviated by a path integral formulation of the system, which allows not only well known expansions, but also the use of nonperturbative methods. It could be expected, that known MF equations are recovered in tree level approximation. Furthermore, a path integral form also makes available many advanced tools about which extensive knowledge has been developed in quantum field theory and statistical physics: Feynman diagrams, equations of motion, treatment of critical phenomena, disorder, (broken-) symmetries, and the functional renormalization group. All these could be of great value to the theory of neuronal networks, where multiple

scales interact and critical points, disorder and fluctuation driven dynamics are present. Thus there is renewed interest in functional methods, building on two formalisms for stochastic systems developed about 30 years ago, the MSR DJ (De Dominicis & Peliti 1978, Janssen 1976, Martin et al. 1973) for stochastic differential equations and the Doi-Peliti (Peliti 1985) for binary variables. An overview is given in Section XVII A.

In this work, we want to gain analytical insight into assembly dynamics, and test an idea about a path integral formulation that is not specific to the neuron model. A single assembly in an otherwise unclustered network is an interesting system for this purpose, for the following reasons: First, due to its mesoscopic size of $\mathcal{O}(100)$ neurons, significant fluctuations of the mean assembly activity are expected and the assumption of gaussianity underlying MFT might be violated, also because of increased correlations due to the positive feedback loop. Second, since a strongly connected assembly can maintain a low and a high activity state which is not the case for the inhibition dominated unclustered network, there must be a phase transition from mono- to bistability depending on the clustering strength. Third, as we do not assume a (biologically unrealistic) symmetric connectivity as in a Hopfield network, the system is not in equilibrium, making it difficult to define an energy function. This is interesting for the path integral, since our formulation does not require the construction of an explicit action. Fourth and lastly, the parameter dependencies of the dynamical properties of an assembly are, to our knowledge, not treated explicitly in the literature. Now it remains to decide on a neuron model: For rate neurons, a MSR DJ path integral is already well established, and we are interested in more realistic neuron models. However, a true spiking model like the LIF neuron might pose serious challenges to analytical treatment. Therefore, since we would like to first test the formalism and compare to known results in order to gain an understanding of the formalism, we use binary neurons with asynchronous update, which can be regarded as one step of approximation below a true spiking model.

We will now commence in Part II first with the description and mean field analysis of the model, compare the results to simulations which we can treat as ground truth, and then introduce the path integral formulation in Part III, first treating only a single population for simplicity, then generalize to multiple populations, and shortly comment on other neuron models. In Part IV the path integral is discussed from a theoretical viewpoint and compared to other approaches in the literature, then the same is done for the assembly results by neuroscientific considerations.

Part II

Mean Field Treatment and Simulation of Assembly Model

V. MEAN FIELD FORMULATION

A. Model description

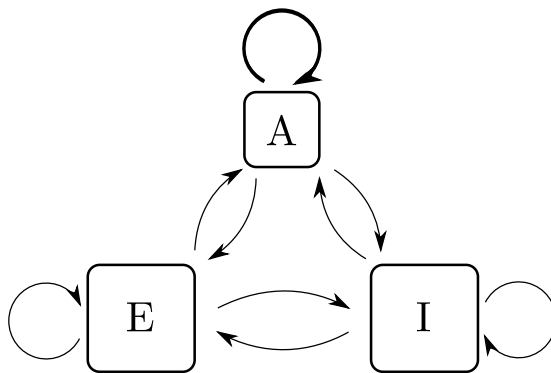


Figure 2. Schematic of network model

The model is composed of three populations of binary neurons: excitatory bath, inhibitory bath and assembly (Figure 2). They will be referenced by the indices $\alpha \in \{E, I, A\}$. Binary neurons with asynchronous random update and Heaviside activation function are used, allowing a relatively tractable analytic treatment while still displaying a simplified form of the characteristic spiking of neurons for low activities. The populations should have no further substructure, therefore connections are assigned randomly with a connection probability p . To simplify mean field treatment, this is done in such a way that the number of inputs to each neuron is the the same ('fixed indegree'). The total number of neurons is $2N$, with N excitatory and N inhibitory neurons, thus each receives Np exc. and the same amount of inh. inputs. All synapses coming from excitatory neurons (exc. bath and assembly) have positive weight $J_e = \frac{J_0}{\sqrt{N}}$, while inhibitory neurons have negative weight $J_i = -g \frac{J_0}{\sqrt{N}}$. Assembly and excitatory bath differ in that the connection probability between assembly neurons is increased by factor a , and the assembly size N_A is small compared to the bath populations. The assembly has negligible influence on the background activity. To keep the indegree fixed for all neurons, the network can be constructed as follows: Start with N excitatory and N inhibitory neurons, each receiving Np randomly drawn exc. and as many inh. inputs. Now single out N_A exc. neurons and replace E->A by A->A connections until the desired assembly connectivity is reached. This results in the connectivity matrix (where $K_{\alpha\beta}$ denotes the number of connections from population β to each neuron in population α)

$$K_{\alpha\beta} = \begin{pmatrix} (N - N_A)p & Np & N_Ap \\ (N - N_A)p & Np & N_Ap \\ (N - aN_A)p & Np & aN_Ap \end{pmatrix}. \quad (4)$$

Note that each row sums correctly to the indegree $2Np$. The analog matrix of the synapse weights is then (with the connection going from β to α)

$$J_{\alpha\beta} = \begin{pmatrix} J_e & J_i & J_e \\ J_e & J_i & J_e \\ J_e & J_i & J_e \end{pmatrix}. \quad (5)$$

B. Derivation of mean field equations

For the following derivation of the mean field equations, we use the notation and approach of Helias et al. (2014) as orientation.

Neurons are chosen randomly and their state $x_i(t) \in \{0, 1\}$ is updated according to the rule

$$x_i(t) = H\left(\underbrace{\sum_j J_{ij} x_j(t)}_{h_i(t)} - \theta\right).$$

The threshold theta models not only the input needed to cause a biological neuron to spike, but can incorporate also a constant external input to the network. The mean interval between updates of the same neuron is τ , which is therefore the characteristic time for which the neurons state will influence its neighbors, thus mimicking the membrane time constant of the LIF neuron model.

We are now interested in the mean population activities $\bar{x}_\alpha(t) = \sum_{i_\alpha} x_{i_\alpha}(t)$. In the absence of any connections and thus input, the mean activity would die out exponentially with time constant τ due to the update process. This would correspond to the simple equations of motion $\tau \partial_t \bar{x}_\alpha(t) + \bar{x}_\alpha(t) = 0$. Then the average activation by input can be added as an additional source

$$\tau \partial_t \bar{x}_\alpha(t) + \bar{x}_\alpha(t) = \langle H(h_i(t) - \theta) \rangle_{i \in \alpha}.$$

This equation can also be derived more strictly from the master equation on the network state space, see eg. Helias et al. (2014), which is omitted here for brevity. As $h_{i_\alpha}(t)$ is a sum of many binary distributed variables x_j that are weakly correlated (Tetzlaff et al. 2012), it can be approximated as gaussian distributed, with mean and variance

$$\mu_\alpha(t) = \sum_\beta K_{\alpha\beta} J_{\alpha\beta} \bar{x}_\beta(t) \quad (6)$$

$$\sigma_\alpha^2(t) = \sum_\beta K_{\alpha\beta} J_{\alpha\beta}^2 \bar{x}_\beta(t) (1 - \bar{x}_\beta(t)). \quad (7)$$

This closes the equation, with the average input becoming

$$\begin{aligned}\langle H(h_i(t) - \theta) \rangle_{i \in \alpha} &= \int dh H(h - \theta) \exp\left(-\frac{1}{2} \frac{(h - \mu_\alpha)^2}{\sigma_\alpha^2}\right) \\ &= \frac{1}{2} \operatorname{erfc}\left(\frac{\theta - \mu_\alpha}{\sqrt{2}\sigma_\alpha}\right),\end{aligned}$$

using the complementary error function $\operatorname{erfc}(\theta) = \int_\theta^\infty \frac{2}{\sqrt{\pi}} \exp(-x^2) dx$.

Therefore the mean field equations are:

$$\tau \partial_t \bar{x}_\alpha(t) + \bar{x}_\alpha(t) = \frac{1}{2} \operatorname{erfc}\left(\frac{\theta - \mu_\alpha(\{\bar{x}_\beta\})}{\sqrt{2}\sigma_\alpha(\{\bar{x}_\beta\})}\right) = \Phi(\mu_\alpha(\{\bar{x}_\beta\}), \sigma_\alpha(\{\bar{x}_\beta\})). \quad (8)$$

C. Mean field equations for assembly model

We can now consider the assembly model as specified above in Section V A. The mean inputs to the three populations are given by

$$\begin{aligned}\mu_E &= ((N - N_a)\bar{x}_E - Ng\bar{x}_I + N_a\bar{x}_A) p \frac{J_0}{\sqrt{N}} \\ \mu_I &= \mu_E\end{aligned} \quad (9)$$

$$\mu_A = ((N - aN_a)\bar{x}_E - gN\bar{x}_I + aN_a\bar{x}_A) p \frac{J_0}{\sqrt{N}} \quad (10)$$

and the input variances

$$\begin{aligned}\sigma_E^2 = \sigma_I^2 &= ((N - N_a)\bar{x}_E(1 - \bar{x}_E) + Ng^2\bar{x}_I(1 - \bar{x}_I) + N_a\bar{x}_A(1 - \bar{x}_A)) p \frac{J_0^2}{N} \\ \sigma_A^2 &= ((N - aN_a)\bar{x}_E(1 - \bar{x}_E) + Ng^2\bar{x}_I(1 - \bar{x}_I) + aN_a\bar{x}_A(1 - \bar{x}_A)) p \frac{J_0^2}{N}.\end{aligned} \quad (11)$$

We can immediately conclude that the mean activities of the excitatory and inhibitory bath populations, $\bar{x}_E$ and $\bar{x}_I$, are equal in this model, since $\mu_E = \mu_I$ and $\sigma_E^2 = \sigma_I^2$. We will now first consider the case of $a = 1$, where the assembly is just a regular part of the excitatory bath without special clustering. Then we have

$$\begin{aligned}\mu = \mu_E = \mu_I = \mu_A &= \bar{x}(1 - g) N p \frac{J_0}{\sqrt{N}} \\ \sigma^2 = \sigma_E^2 = \sigma_I^2 = \sigma_A^2 &= \bar{x}(1 - \bar{x})(1 - g) N p \frac{J_0^2}{N},\end{aligned}$$

so the system reduces to a single mean field equation

$$\tau \partial_t \bar{x}(t) + \bar{x}(t) = \frac{1}{2} \operatorname{erfc}\left(\frac{\theta - \mu(\bar{x}(t))}{\sqrt{2}\sigma(\bar{x}(t))}\right). \quad (12)$$

The stationary solutions of the equation can be graphically visualized by setting $\partial_t \bar{x} = 0$, then plotting the right and left hand sides of the equation and looking for intersections of the two functions (see Section VII).

VI. PARAMETER CHOICES

We now shortly discuss the chosen parameter values. The model cannot capture the biological reality quantitatively due to its level of abstraction. However, it should do so qualitatively for the modeled type of dynamics, and at least

operate with parameters in the same order of magnitude as in a corresponding more realistic network. The main feature is the network size and synaptic structure. A human cortical neuron has $\mathcal{O}(10^4)$ synapses and a connection probability to other neurons in its local ($\sim 1\text{mm}$) environment of 10 – 20% (Rolls 2016). Since this would imply an $N \approx \mathcal{O}(10^5)$, which would be computationally expensive, it is advantageous to scale down the network while keeping the activity features constant. This is done by scaling the synapse weight with $\frac{1}{\sqrt{N}}$ (Barral & Reyes 2016, van Vreeswijk & Sompolinsky 1998), thus keeping the input variance σ^2 of a neuron independent of N , while the mean input will increase as $\mu \sim \sqrt{N}$, since there are $2Np$ inputs per neuron. The mean input (and indirectly, the mean activity) can however be controlled by varying the threshold θ , for which higher values represent less external input, and lower values more. Therefore we absorb the scaling of the mean input into θ to keep also the base activity independent of N .

This is done by considering that as the bath activity $\bar{x}_E$ must be a solution of (12) independently of N , the argument of the error function complementary must be independent of N . Thus

$$\frac{\theta - \bar{x}(1-g)Np\frac{J_0}{\sqrt{N}}}{\sqrt{2}\sigma(\bar{x})} = \text{erfc}^{-1}(2\bar{x})$$

$$\theta = \sqrt{2}\sigma(\bar{x})\text{erfc}^{-1}(2\bar{x}) + \sqrt{N}\bar{x}(1-g)pJ_0. \quad (13)$$

Note that this construction models only the local network, where the spatial connectivity profile is assumed to be constant, treating everything else as external input. The mean activity of cortical pyramidal neurons in layer 2/3 is (translated into activations per membrane time constant, see Kühn & Helias (2017)) approximately 0.01 – 0.1, while those in layer 5 have somewhat higher activity (O’Connor et al. 2010). Therefore we consider a range of mean bath activities, but $\bar{x}_E = 0.1$ as the rough standard. The network size used for the example plots in the following sections is usually $N = 2000$, as this allowed an easily manageable simulation runtime. The value of the base synaptic weight is $J_0 = 5$ and $g = 2$. The absolute scale of J_0 is irrelevant, as it only leads to a redefinition of θ , and the value of g reflects the important point that the network is inhibition dominated. The membrane time constant of the neurons, modeled by the mean update intervals, is chosen as $\tau = 10\text{ms}$. Biological time constants vary roughly between 5 – 60 ms, depending on cell types, areas and layers, see the NeuroElectro.org database (Tripathy et al. 2015). The biological synaptic delay and transmission times in axon and dendrite, which are of the order of a few ms (Lefort et al. 2009), where neglected in theory and simulation, as the effect should be small. The assembly size N_A is varied mostly between 50 and 150, with $N_A = 70$ being the standard example. The lower bound is effected by the chosen clustering method (increasing the number of connections), while the upper bound is theoretically open. No qualitatively different dynamics arise from changes in assembly size, see Section VII. In the literature, assumed local assembly sizes range from tens to hundreds of neurons, e.g. Buzsáki (2010), Mazzucato et al. (2015), Perin et al. (2013).

VII. EMERGENCE OF THE SECOND ATTRACTOR

We now discuss the mean field solutions of the assembly activity for varying clustering strengths a and compare the theoretical results to simulations of the system.

Considering the assembly with arbitrary a , we start by noticing that when setting $\bar{x}_A = \bar{x}_E$ in the equations

$$\mu_A = ((N - aN_a)\bar{x}_E - gN\bar{x}_I + aN_a\bar{x}_A)p\frac{J_0}{\sqrt{N}} \quad (14)$$

$$\sigma_A^2 = ((N - aN_a)\bar{x}_E(1 - \bar{x}_E) + Ng^2\bar{x}_I(1 - \bar{x}_I) + aN_a\bar{x}_A(1 - \bar{x}_A))p\frac{J_0^2}{N}, \quad (15)$$

the assembly input mean and variance are still identical to those of the bath populations. Therefore the stationary solution of the system without assembly, henceforth called ground state, is also always a stationary point of the assembly activity (note that it must not always be stable, as will be discussed shortly). This is a property of the particular method to implement an assembly by increased connection probability only, when considering increased synaptic weights instead, the assembly ground state activity increases with a , see Section XV. There is another very interesting structure in the equations: Note that the variables a and N_a always appear in conjunction as aN_a . Therefore,

in mean field approximation, any change in the cluster size N_a can be absorbed by an antiproportional adjustment in the clustering factor a , leaving the eqs. invariant. Thus any result for a certain cluster size, like the critical points discussed below, can be translated to (in MF approx.) equivalent systems with the scaling $a \sim \frac{1}{N_a}$.

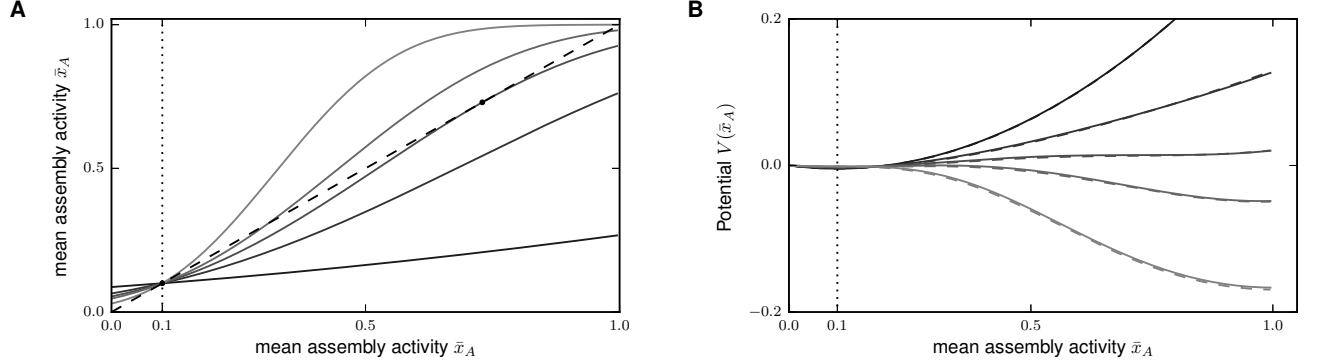


Figure 3. (A) Graphical representation of stationary MF equations. Parameters are $N = 2000$, $N_A = 70$ and $\bar{x}_E = 0.1$. $\Phi(\bar{x}_A)$ is plotted for $a = \{1(\text{black}), 3, a_{crit.1}, 5, a_{crit.2}(\text{light gray})\}$, dashed black line gives the angle bisector $\bar{x}_A$. (B) Corresponding potentials. Solid lines by numerical evaluation of (16), dashed lines are the analytical approximation (17).

To study the assembly mean field equation for varying a , it is most instructive to look at the graphical representation of the stationary equation. In Panel A of Figure 3 $\Phi(\mu_A, \sigma_A)$ is plotted for some values of a showing the whole range of qualitatively different regimes. Before starting the discussion, we note that if $\partial_{\bar{x}}\Phi < 1$ at an intersection (as for the ground state) a slight deviation from the fixpoint in any direction will decay due to $\tau\partial_t\bar{x}_{+\epsilon} = \Phi(\{\bar{x}_{+\epsilon}\}) - \bar{x}_{+\epsilon} < 0$ and $\tau\partial_t\bar{x}_{-\epsilon} > 0$, the fixpoint is therefore stable. If on the other hand $\partial_{\bar{x}}\Phi > 1$, the time derivatives of the mean activity have the same sign as the deviation and the fixpoint is an unstable one. Therefore, starting at $a = 1$, there is only the stable ground state. Then with increasing clustering, at some value $a_{crit.1}$ the curve touches the angle bisector, spawning one unstable and one stable fixedpoint at higher activity. This corresponds to a first order phase transition. The system is now bistable, displaying a ground and an upstate, which can be reached by fluctuations or external stimuli not contained in the mean field approximation. After further increase of the clustering to $a_{crit.2}$, the ground state loses stability, constituting a second order phase transition. From now only the upstate is stable, at very high activities, and the ground state remains as an unstable fixpoint. This three regimes are independent of the model parameters, as long as $\bar{x}_E < 0.5$, $g > 1$ and $N \gg N_A$, which are biologically and theoretically clear constraints.

The numeric values of the stationary points were extracted by bisectioning and compared to a simulation of the same system with NEST (the NEural Simulation Tool (Bos et al. 2015)), shown in Figure 4. For a description of the simulation and analysis methods, see Section XXII.

The mean field theory predictions are inaccurate in the vicinity of the critical points, but show qualitatively the correct structure. The differences between MFT and simulation can be understood by the fluctuation dominated dynamics in these regimes.

A. Potential landscape

For this purpose, let us define a potential for the system: Consider a particle moving in a potential with overdamped dynamics driven by noise. The equation of motion is given by

$$\tau\partial_t x(t) + \partial_x V(x) = \xi(t),$$

where, in the simplest case, ξ is gaussian white noise with amplitude $\langle \xi(t)\xi(t') \rangle = 2\tau T\delta(t-t')$ determined by the 'temperature' T . We can identify the potential as

$$V(\bar{x}_A) = \int_{\text{const.}}^{\bar{x}_A} d\bar{x}'_A \left[\bar{x}'_A - \frac{1}{2} \text{erfc} \left(\frac{\theta - \mu_A(\bar{x}_E, \bar{x}'_A)}{\sqrt{2}\sigma_A(\bar{x}_E, \bar{x}'_A)} \right) \right]. \quad (16)$$

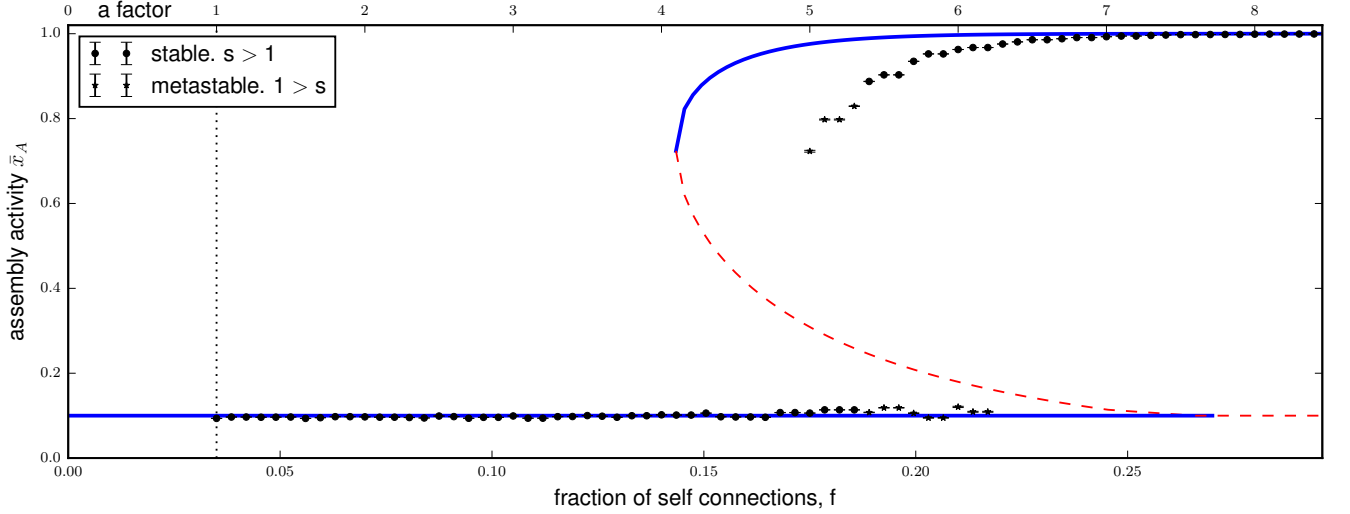


Figure 4. Bifurcation diagram showing the comparison of mean field theory to simulation. Parameters are $N = 2000$, $N_A = 70$ and $\bar{x}_E = 0.1$. Blue lines are stable fixpoints in MFT, the red dashed separatrix corresponds to the unstable fixpoint, simulation results are shown in black. Full dots indicate the dominating state (global minimum), stars the less frequently visited local minimum. Two adjacent dots with the same activity result from float to int rounding artifacts of K_{AA} (#of intra assembly connections).

This expression can be integrated analytically under the approximation that the $\bar{x}_A$ dependence of $\sigma_A(\bar{x}_E, \bar{x}_A)$ can be neglected (Figure 3 Panel B shows a comparison to the numerically evaluated Potential). Using integration by parts, $\int_0^x dy \operatorname{erfc}(y) = y \operatorname{erfc}(y)|_0^x + \int_0^x dx y \frac{2}{\sqrt{\pi}} \exp(-y^2) = x \operatorname{erfc}(x) - \frac{1}{\sqrt{\pi}} (\exp(-x^2) - 1)$, so that with the substitution $y = -\frac{\theta - \mu_A(\bar{x}_A)}{\sqrt{2}\sigma_A}$ leading to $dy = \frac{a N_a p J_0}{\sqrt{2N}\sigma_A} d\bar{x}'_A$ the potential is approximated as

$$V(\bar{x}_A) \approx \frac{1}{2} \bar{x}_A^2 + \frac{\sqrt{N}}{a N_a p J_0} [(\theta - \mu_A(\bar{x}_A)) \Phi(\mu_A(\bar{x}_A), \sigma_A) - (\theta - \mu_A(0)) \Phi(\mu_A(0), \sigma_A)] - \frac{\sqrt{N}\sigma_A}{\sqrt{2\pi} a N_a p J_0} \left[\exp\left(-\frac{(\theta - \mu_A(\bar{x}_A))^2}{2\sigma_A^2}\right) - \exp\left(-\frac{(\theta - \mu_A(0))^2}{2\sigma_A^2}\right) \right]. \quad (17)$$

Figure 3 Panel B shows how the potential changes with varying a . There are two main observations: First and trivially, if the minimum depth gets smaller than the noise amplitude, it can no longer constitute an effective (meta)stable state, which explains why the up-state appears in the simulation for higher a than the minimum in MFT, and why the ground state loses stability for lower values than predicted by MFT. Second, the difference between the simulated activity in the upstate and the MFT prediction could be explained by an asymmetric minimum which is rather flat on the left side. However, while this asymmetric form can be seen right after the critical point where the minimum appears, it seems no longer the case for values of a where the simulation starts to show a stable up-state. This inconsistency could be expectedly solved by including fluctuation corrections to the calculation, in which case the effective potential could show the critical point to emerge later, and the asymmetric form then still present in the correct range to explain the low upstate activities. This calculation is difficult to perform in the present framework, but can be done systematically in the path integral formulation as a loopwise expansion, see Part III.

B. The second attractor cannot emerge at activities $m \leq 0.5$

An interesting question is whether the up-state must always be clearly distinguishable from the down-state by a big leap in the activity. This is of relevance for the experimental detection of assembly activity in neural tissue (Latham & Nirenberg 2004). For our model system, the question can be mathematically addressed:

In a stationary point of assembly activity $\bar{x}_A$, the self-consistency equation (8) is given by

$$\bar{x}_A = \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{pN}{2}} \frac{\frac{\theta}{\sqrt{N}pJ_0} - (1-f)\bar{x}_E + g\bar{x}_I - f\bar{x}_A}{\sqrt{((1-f)\bar{x}_E(1-\bar{x}_E) + g^2\bar{x}_I(1-\bar{x}_I) + f\bar{x}_A(1-\bar{x}_A))}} \right) \quad (18)$$

$$= \frac{1}{2} \operatorname{erfc} \left(\sqrt{\frac{pN}{2}} \frac{b - f\bar{x}_A}{\underbrace{\sqrt{a_c + f\bar{x}_A(1-\bar{x}_A)}}_{c(\bar{x}_A)}} \right) \quad (19)$$

where we used the fraction of inputs to an assembly neuron that come from other assembly members $f = a \frac{N_a}{N}$ to shorten the expression and then collected constant terms. The discussion simplifies considerably if we can neglect the dependence of $c(\bar{x}_A)$ on $\bar{x}_A$, which is a good approximation if f is small compared to $a_c \sim \mathcal{O}(1)$. To address this question we consider the scaling of f with N : Use as reference point the emergence of the up-state at $a_{crit.1}$ and keep the argument of the erfc fixed while varying N . Using the scaling of θ obtained in (13)

$$\begin{aligned} \operatorname{erfc}^{-1}(2\bar{x}_A) &= \frac{\sigma(\bar{x}_E) \operatorname{erfc}^{-1}(2\bar{x}_E) - fpJ_0 \sqrt{\frac{N}{2}} (\bar{x}_A - \bar{x}_E)}{\sigma_A(\bar{x}_A)} \\ \Rightarrow f &= a \frac{N_a}{N} = \frac{1}{\sqrt{N}} \frac{\sqrt{2}}{pJ_0 (\bar{x}_A - \bar{x}_E)} [\sigma(\bar{x}_E) \operatorname{erfc}^{-1}(2\bar{x}_E) - \sigma_A(\bar{x}_A) \operatorname{erfc}^{-1}(2\bar{x}_A)] \\ &\Rightarrow a \sim \mathcal{O}(\sqrt{N}). \end{aligned} \quad (20)$$

Note that $\sigma_A(\bar{x}_A)$ is not independent of N if $f \sim \mathcal{O}(\frac{1}{\sqrt{N}})$, but $\lim_{N \rightarrow \infty} \sigma_A(\bar{x}_A) = \sigma(\bar{x}_E)$ so that the dependence can be neglected for large N , confirming the validity of the approximate scaling $f \sim \mathcal{O}(\frac{1}{\sqrt{N}})$.

We can thus neglect the dependence of $c(\bar{x}_A)$ on $\bar{x}_A$ for large network sizes and proceed with the argument concerning the up-state activity. After renaming constants, equation (19) then has the form

$$\bar{x}_A = \frac{1}{2} \operatorname{erfc}(\tilde{b} - \tilde{c}\bar{x}_A) \equiv \frac{1}{2} F(\bar{x}_A)$$

which makes explicit that $F(x)$ is just a error complementary function shifted and stretched in x direction. This transformations conserve the shape of the function, especially the point where the derivative gets maximal (the inflexion point) will always have the function value $\frac{1}{2} F(x_{inflex.}) = 0.5$. This immediately implies that if the equation has two solutions, one must be above and one below 0.5, and more than two solutions are not possible (see Figure 3). Therefore, the upstate must always be at $\bar{x}_A \geq 0.5$ and since realistic base activities for cortical pyramidal neurons are $\bar{x}_E \ll 0.5$, the two states are well separated. Expected generalization to more realistic architectures is discussed in Part IV.

VIII. SPONTANEOUS TRANSITIONS

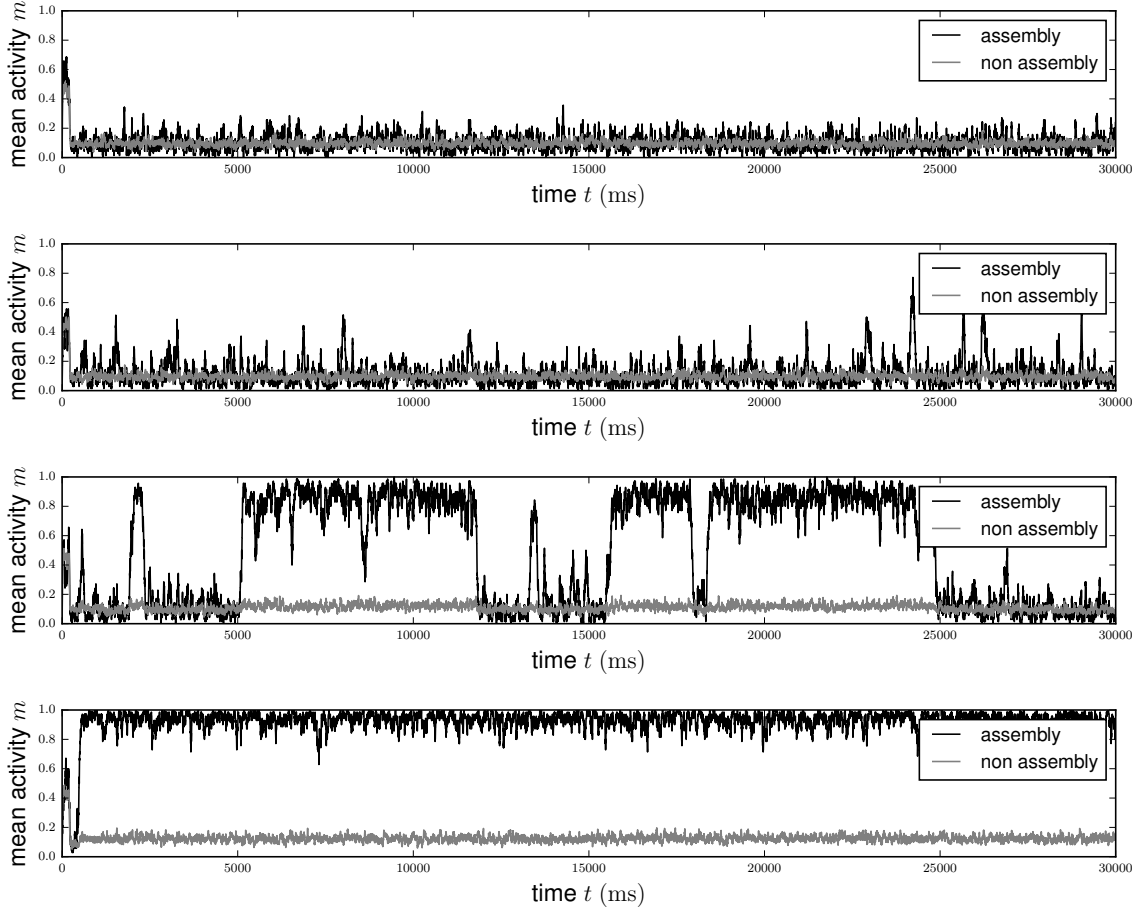


Figure 5. Spontaneous assembly activity for varying a , showing the 3 different regimes. Parameters are $N = 2000$, $N_A = 70$, $\bar{x}_E = 0.1$ and values of a (from top to bottom): 3.5, 4.8, 5.3, and 5.7. Black line shows population average of assembly activity, gray line the bath activity. The activity bump at $t = 0$ is a short external drive to start the activity in the network.

We have seen that the assembly activity can be approximately described as the noise driven dynamics of an overdamped particle in a potential. In the bistable regime we expect spontaneous transitions between the stable states depending on the depth of the potential wells compared to the fluctuation sizes. Figure 5 shows these dynamics of the simulated system for the different regimes. The fluctuations grow as the first critical point is approached, then spontaneous transitions are observed as approximately determined by the relative depths of the potential minima, and lastly the fluctuations decrease as the up-state gains dominance and the ground state loses stability. The evolution of the mean activity and transition rates in the simulation is shown in Figure 6.

In Section VII we observed an invariance of the MF eqs. under variation of the cluster size if $a \sim \frac{1}{N_A}$ is scaled accordingly. We expect however, that smaller clusters display larger fluctuations since the activation of one member neuron should have a much larger impact. This can indeed be seen in the simulation, where the standard error of the assembly activity in the ground state is 0.08 for $N_A = 50$, $a = 6$ and 0.045 for $N_A = 100$, $a = 3$ (with $N = 2000$, $\bar{x}_E = 0.1$), but is inaccessible to mean field theory, which neglects crosscorrelations between neurons. Also the bifurcation diagrams of these two systems, which are equivalent in MFT, look quite different, see Figure 7. Interestingly, the system with less assembly members is closer to the MF prediction. To understand this effect, fluctuation corrections to the effective potential need to be considered. Correlations will be investigated in an extension to MFT in Section IX, and in Part III we calculate a 1-loop correction term to the MF equation that is proportional to the second derivative of the input gain function, which is negative in the upstate and grows with N_A , thus providing the kind of effect seen

in Figure 7.

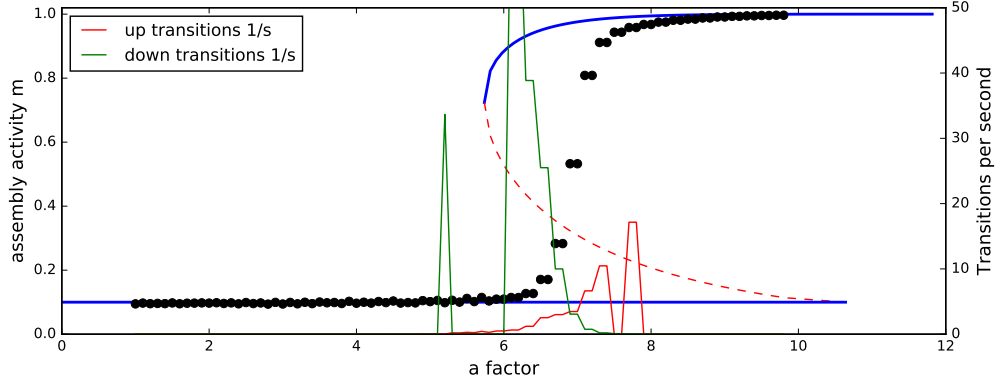


Figure 6. Evolution of spontaneous mean activity and transition rates from simulation on the background of the MFT bifurcation diagram. Parameters are $N = 2000$, $N_A = 50$ and $\bar{x}_E = 0.1$. Again, two dots with the same activity are a float to int rounding artifact.

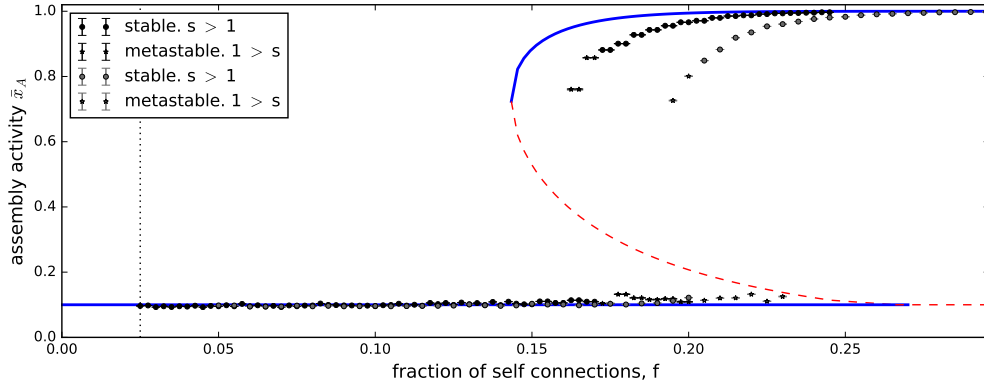


Figure 7. Comparison of bifurcation diagrams of two systems with identical MF equations. Black dots: $N_A = 50$, gray dots: $N_A = 100$. Other parameters: $N = 2000$ and $x_E = 0.1$.

A. Flatness of the potential

We consider the question how small the potential wall between minima can be depending on the model parameters. Consider Figure 8 panel A: The configuration where both minima have the same depth, with a minimal hill in between is given when the separatrix is at $x_A = 0.5$, so the MF eq. fulfills $\frac{1}{2} = \Phi(\bar{x}_A = 0.5)$. We can always choose a such that this condition is met. Then the height of the potential wall is controlled by the derivative at this point $\partial_{\bar{x}_A} \Phi|_{\bar{x}_A=0.5}$, giving a relation between the flatness of the potential and the remaining parameters.

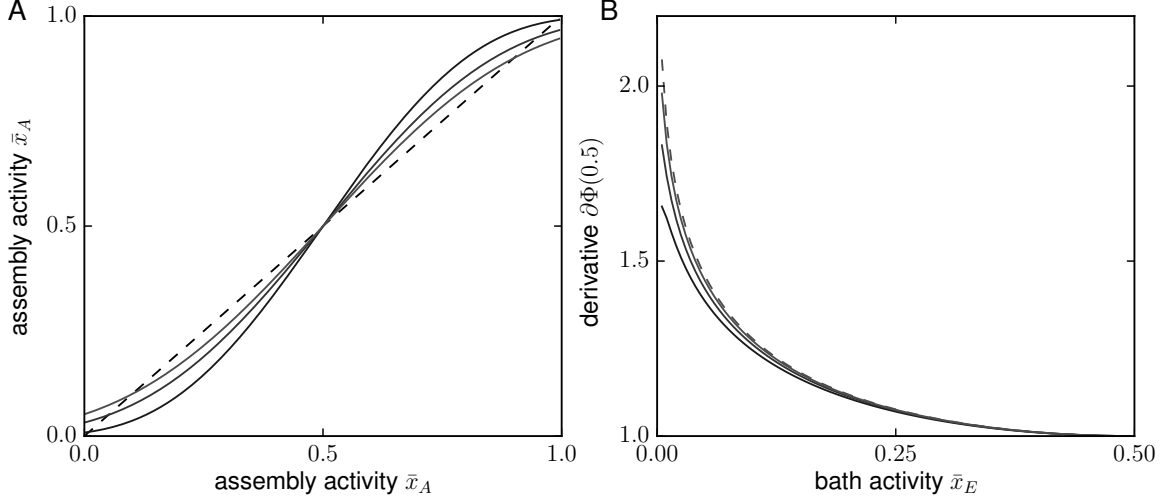


Figure 8. (A) Ansatz for determining flatness of the potential. $\Phi(\bar{x}_A)$ for $\bar{x}_E = \{0.01, 0.05, 0.1\}$ with f chosen according to (21), $N = 2000$. (B) Derivative at the separatrix as a function of $\bar{x}_E$. Shows (22) for $N = \{2000 \text{ (black)}, 8000, 64000 \text{ (light gray)}\}$, and the approximation neglecting the term in brackets (dashed line).

We start with the condition fixing the separatrix, using the relation for the threshold θ obtained in (13) to keep the base activity fixed to a $\bar{x}_E$:

$$\begin{aligned}
 \frac{1}{2} &= \Phi(0.5) = \frac{1}{2} \operatorname{erfc} \left(\frac{\theta - \mu(0.5)}{\sqrt{2}\sigma(0.5)} \right) \\
 \Rightarrow \operatorname{erfc}^{-1}(1) &= \frac{1}{\sigma(0.5)} \left(\sigma(\bar{x}_E) \operatorname{erfc}^{-1}(2\bar{x}_E) - fpJ_0 \sqrt{\frac{N}{2}} (0.5 - \bar{x}_E) \right) \\
 \Rightarrow f &= \sqrt{\frac{2}{N}} \frac{\sigma(\bar{x}_E) \operatorname{erfc}^{-1}(2\bar{x}_E)}{pJ_0 (0.5 - \bar{x}_E)}, \tag{21}
 \end{aligned}$$

giving the condition for f which we then insert into the expression for the derivative. At $\bar{x}_A = 0.5$ the outer derivative gives just $\frac{1}{\sqrt{\pi}}$ and the inner derivative of $\sigma(\bar{x}_A)$ vanishes, thus

$$\begin{aligned}
 \Phi'(0.5) &= \sqrt{\frac{N}{2\pi}} fpJ_0 \frac{1}{\sigma(0.5)} \\
 &= \frac{1}{\sqrt{\pi}} \frac{\operatorname{erfc}^{-1}(2\bar{x}_E)}{0.5 - \bar{x}_E} \frac{\sigma(\bar{x}_E)}{\sigma(0.5)} \\
 &= \dots \\
 &= \frac{1}{\sqrt{\pi}} \frac{\operatorname{erfc}^{-1}(2\bar{x}_E)}{0.5 - \bar{x}_E} \left[1 + \sqrt{\frac{2\bar{x}_E(1 - \bar{x}_E)}{Np(1 + g^2)}} \left(\frac{1}{4\bar{x}_E(1 - \bar{x}_E)} - 1 \right) \frac{\operatorname{erfc}^{-1}(2\bar{x}_E)}{0.5 - \bar{x}_E} \right]^{-\frac{1}{2}}. \tag{22}
 \end{aligned}$$

The Potential is flat if $\Phi(0.5) = 1$. A plot of the result is shown in panel B of Figure 8, showing that the term $\frac{1}{\sqrt{\pi}} \frac{\operatorname{erfc}^{-1}(2\bar{x}_E)}{0.5 - \bar{x}_E}$ is dominating, especially for larger system sizes $N > 10^5$. Thus the derivative depends mostly on the baseactivity, while all other parameters, appearing as the product $Np(1 + g^2)$ have only little influence.

IX. PAIRWISE CORRELATIONS

Understanding the correlation structure of the assembly model is necessary to develop insight into the fluctuation dynamics of the system. Since the assembly is subject to strong internal recurrent excitation, the pairwise crosscorrelations between members must contribute significantly to the fluctuation strength, which is not captured in the mean field treatment. We use an extension of MFT (Buice et al. 2010, Ginzburg & Sompolinsky 1994, Helias et al. 2014) that considers the linearized effect of one neuron on its neighbor, leading to equations for the population averaged pairwise crosscorrelations depending on the working point $\{\mu, \sigma\}$ (which is determined by the mean field solution):

$$2c_{\alpha\beta} = \sum_{\gamma} (\omega_{\alpha\gamma} c_{\gamma\beta} + \omega_{\beta\gamma} c_{\gamma\alpha}) + \omega_{\alpha\beta} \frac{a_{\beta}}{N_{\beta}} + \omega_{\beta\alpha} \frac{a_{\alpha}}{N_{\alpha}} \quad (23)$$

with the effective coupling matrix $\omega_{\alpha\beta} = S(\mu_{\alpha}, \sigma_{\alpha}) J_{\alpha\beta} K_{\alpha\beta}$. The autocovariance a_{α} can be expressed as $\bar{x}_{\alpha} - \bar{x}_{\alpha}^2$, due to the neurons being binary variables. The susceptibility is given by

$$S(\mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\theta - \mu_k)^2}{2\sigma_k^2}}. \quad (24)$$

A. Numerical solution of equations for linearized correlations

Looking at (23), we see that $c_{\alpha\beta}$ is symmetric in its indices, as should be the case for a covariance matrix. Thus we have a system of 6 linear equations, which can be written matrix form

$$\mathbb{M} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ea} \\ c_{ii} \\ c_{ia} \\ c_{aa} \end{pmatrix} = \vec{b}. \quad (25)$$

The explicit forms of $\mathbb{M}$, $\vec{b}$ and further details this section are given in appendix Section XXIII A. The linear system of equations could be solved analytically, but would yield very uninformative long expressions. Using a numeric solver, we also iterated the equation by using the resulting crosscorrelations as additional contribution to the fluctuation strength $\sigma(\{\bar{x}_{\alpha}\})$ and calculating the updated stationary solution $\{\mu, \sigma\}$ of the MF eqs., which was then used for the next iteration of (25), until convergence. Figure 9 shows a comparison between the numerical solution, the analytical approximation of the next section and the correlations extracted from simulation. The linearized correlations are only a rough approximation, and not suited to describe the spontaneously switching regime, since they are effectively assuming a quadratic potential around the working point, thus neglecting the asymmetric form of the potential minima discussed in Section VII A and the second potential well. It is not clear how to systematically extend the framework further. However, systematic corrections follow naturally in the path integral formulation, which is subject of the next Part.

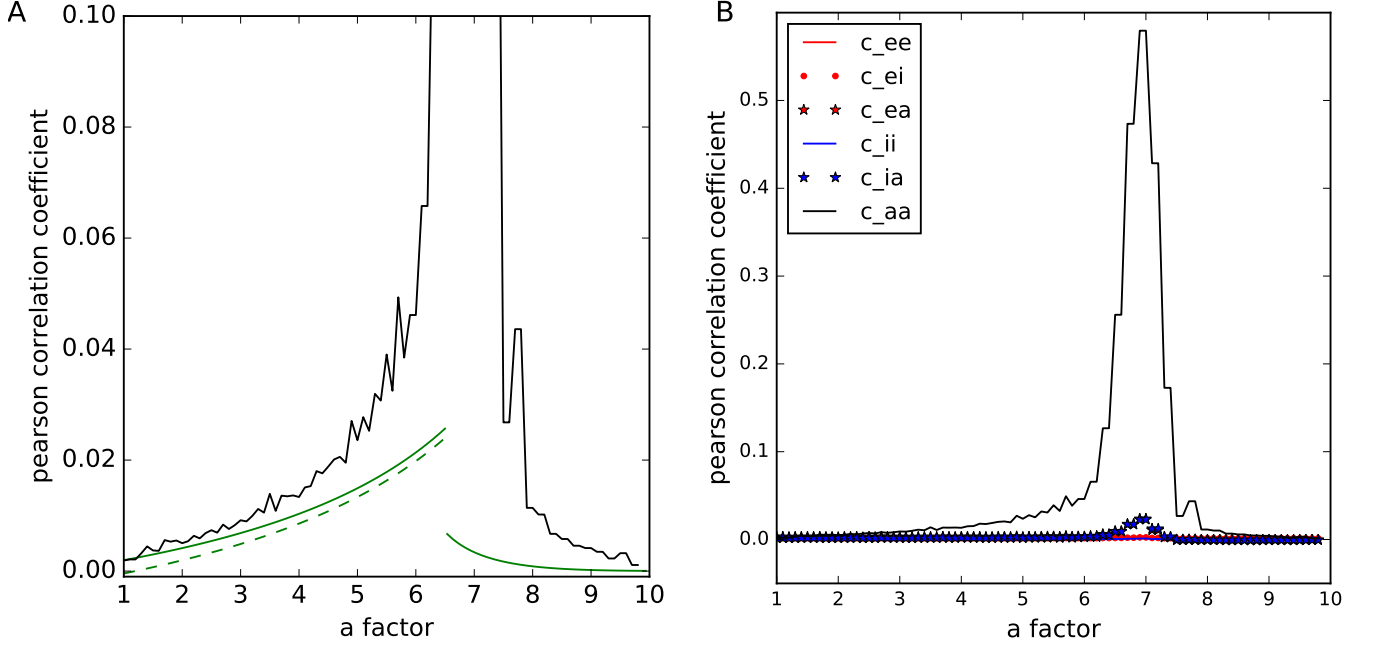


Figure 9. Population averaged pairwise crosscorrelations. (A) comparison of c_{aa} from simulation (black), numerical gaussian theory (green) and analytic approximation (green dashed). The jump in the gaussian theory values results from manually changing the working point from the downstate to the upstate. (B) simulation results for the other crosscorrelations. Parameters are $N = 2000$, $N_A = 70$ and $\bar{x}_E = 0.1$.

B. Separation of equations by neglecting feedback of assembly to bath

The justified approximation that the assembly has no significant influence on the bath activity can be expressed by setting ω_{ae} and ω_{ia} to zero. Then the system of equations should decouple into the homogeneous EI-network, and the purely excitatory assembly network receiving feedforward input by the former. Indeed, reordering the the entries of the correlations vector such that the assembly correlations are on one side, $\vec{c}^T = (c_{ee}, c_{ei}, c_{ii}, c_{ea}, c_{ie}, c_{aa})$, the matrix decomposes into four 3x3 blocks

$$\mathbb{M} = \begin{pmatrix} \mathbb{M}_{EI} & \mathbb{M}_{fback} \\ \mathbb{M}_{ffwd} & \mathbb{M}_A \end{pmatrix}. \quad (26)$$

Setting $\omega_{ae} = \omega_{ia} = 0$ results in $\mathbb{M}_{fback} = 0$, so the system of equations decomposes into

$$\mathbb{M}_{EI} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ii} \end{pmatrix} = \vec{b}_{hom}. \quad (27)$$

$$\mathbb{M}_{ffwd} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ii} \end{pmatrix} + \mathbb{M}_A \cdot \begin{pmatrix} c_{ea} \\ c_{ia} \\ c_{aa} \end{pmatrix} = \vec{b}_a, \quad (28)$$

allowing us to solve the equation for the homogeneous EI-network first and then use the resulting covariances to obtain also the solution of the assembly system. This is easily done by gaussian elimination, giving the known results for the homogeneous EI-net (Helias et al. 2014) and a still lengthy expression for c_{aa} (result shown in Figure 9). The explicit expressions can be found in Appendix Section XXIII B.

Part III

Path Integral Formulation

We now introduce a novel path integral formulation that allows to leave the specification of the neuron model implicit. The idea is that while the activity (output) of neurons is very dynamic and variable, the input of neurons is, independent of the neuron model, always a sum of a large number of activities, and therefore approximately gaussian distributed, thus yielding a path integral that can be expanded. The neuron model then only appears as the distribution describing the input to activity relation, and the equations are closed by the connectivity linking the activities to the inputs.

X. A NEURON MODEL INDEPENDENT PATH INTEGRAL FORMULATION

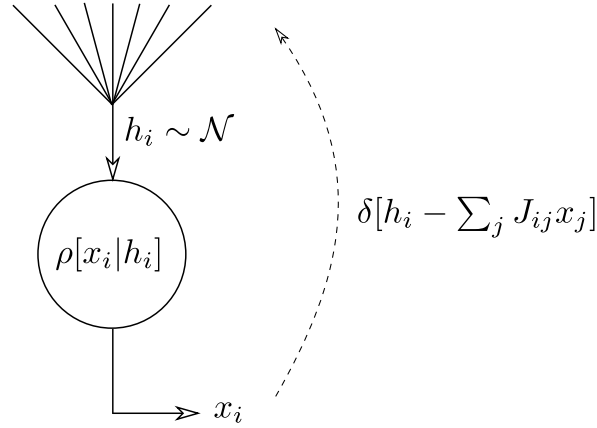


Figure 10. Schematic showing generic features of neuronal networks and the notation used.

We are interested in the disorder averaged expectation value of some observable $O[\mathbf{h}]$ where $h_i(t)$ is the input field of neuron i

$$\langle O[\mathbf{h}] \rangle_{\mathbf{J}, \mathbf{h}} := \int \mathcal{D}\mathbf{h} \langle \rho[\mathbf{h}](\mathbf{J}) \rangle_{\mathbf{J}} O[\mathbf{h}] \quad (29)$$

because the observable itself does not depend on the realization of the connectivity $\mathbf{J}$ the disorder average can be executed on the probability distribution only, and does not need to be repeated for different $O[\mathbf{h}]$. We can express $\rho[\mathbf{h}](\mathbf{J})$ by its relation to the the neuron activities x_i , leading to

$$\begin{aligned} \langle \rho[\mathbf{h}](\mathbf{J}) \rangle_{\mathbf{J}} &:= \left\langle \int \mathcal{D}\mathbf{x} \Pi_i \left\{ \delta[h_i - \sum_j J_{ij} x_j] \right\} \Pi_j \{ \rho_j[x_j | h_j] \} \right\rangle_{\mathbf{J}} \\ &= \int \mathcal{D}\hat{\mathbf{h}} \int \mathcal{D}\mathbf{x} \langle \exp(\hat{\mathbf{h}}^T \mathbf{h} - \hat{\mathbf{h}}^T \mathbf{J} \mathbf{x}) \rho[\mathbf{x} | \mathbf{h}] \rangle_{\mathbf{J}} \\ &= \int \mathcal{D}\hat{\mathbf{h}} \exp(\hat{\mathbf{h}}^T \mathbf{h}) \int \mathcal{D}\mathbf{x} \langle \exp(-\hat{\mathbf{h}}^T \mathbf{J} \mathbf{x}) \rangle_{\mathbf{J}} \rho[\mathbf{x} | \mathbf{h}], \end{aligned} \quad (30)$$

where the delta distribution was represented by its Fourier transform and the imaginary unit absorbed into $\hat{\mathbf{h}}$. Because the path integral goes over all time points each matrix multiplication means also an integration over time, leading to the short hand $\hat{\mathbf{h}}^T \mathbf{J} \mathbf{x} = \sum_{i,j} \int dt ds \hat{h}_i(t) \mathbf{J}_{ij}(t, s) x_j(s)$.

We identify the coupling term as the characteristic function $\varphi_{\mathbf{J}}(\mathbf{y}) = \langle \exp(\sum_{i,j} y_{ij} J_{ij}) \rangle_{\mathbf{J}}$ of the distribution of the couplings $\mathbf{J}$ evaluated at the point $y_{ij} = -\int \hat{h}_i(t) x_j(t) dt$. As the standard example, we assume the network has a

connectivity given by $J_{ij} = \frac{J_0}{\sqrt{N}} B(p)$, where $B(p)$ is the Bernoulli distribution. Thus the synaptic weights are all the same, but the indegree is approximately gaussian distributed and not fixed. This is a difference to the networks with fixed indegree considered in the simulations, the impact of which will be pointed out in Section XI C. Using the connectivity statistics assumed, we can execute the disorder average approximately by noting that high moments are suppressed

$$\varphi_{\mathbf{J}}(\mathbf{y}) \simeq \exp \left(\frac{m}{\sqrt{N}} \sum_{ij} y_{ij} + \frac{g^2}{2N} \sum_{ij} y_{ij}^2 + O(N^{-\frac{3}{2}}) \right) \quad (31)$$

with $m = J_0 p$ and $g^2 = J_0^2 p(1-p)$.

Note that this step did not require the connectivity statistics that were assumed here as an instructive, simple case. The generalization to multiple populations is done in Section XIV and a calculation for fixed indegree can be found in Section XXIV D. Interestingly, also a distribution for the synaptic weights could be considered, i.e. $J_{ij} \sim \frac{J_0}{\sqrt{N}} \text{Lognormal}(\mu, \sigma^2) B(p)$ as suggested by experiments (Song et al. 2005), where one may also consider higher moments than the first two.

After inserting this back into (30) a Hubbard-Stratonovich transform can be performed by introducing variables R_1 and Q_1 , enforcing their definition by a delta constraint, and again using the Fourier representation which gives two additional auxiliary variables R_2 and Q_2 :

$$\begin{aligned} \langle \rho[\mathbf{h}](\mathbf{J}) \rangle_{\mathbf{J}, \mathbf{h}} &\simeq \int \mathcal{D}\hat{\mathbf{h}} \mathcal{D}\mathbf{h} \int \mathcal{D}\mathbf{x} \Pi_i \exp \left(\hat{h}_i h_i - \underbrace{\int \hat{h}_i(t) \frac{m}{\sqrt{N}} \sum_j x_j(t) dt}_{=: R_1(t)} \right. \\ &\quad \left. + \underbrace{\iint \hat{h}_i(t) \hat{h}_i(s) \frac{g^2}{2N} \sum_j x_j(t) x_j(s) dt ds}_{=: \frac{1}{2} Q_1(t, s)} + O(N^{-\frac{3}{2}}) \right) \rho[x_i | h_i] \\ &= \iint \mathcal{D}R_1 \mathcal{D}R_2 \iint \mathcal{D}Q_1 \mathcal{D}Q_2 \exp(Q_2^T Q_1 + R_2^T R_1) \\ &\quad \times \underbrace{\int \mathcal{D}\hat{\mathbf{h}} \mathcal{D}\mathbf{h} \int \mathcal{D}\mathbf{x} \exp \left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1) - R_2^T \frac{m}{\sqrt{N}} \sum_j x_j + \frac{1}{2} \hat{\mathbf{h}}^T Q_1 \hat{\mathbf{h}} - \frac{g^2}{N} \sum_j x_j^T Q_2 x_j \right)}_{=: \Omega(R_1, R_2, Q_1, Q_2)} \rho[\mathbf{x} | \mathbf{h}]. \end{aligned} \quad (32)$$

$$\quad (33)$$

We have thus expressed the disorder averaged probability distribution of the inputs in the generic form $\exp(S(R_1, R_2, Q_1, Q_2))$, with the action

$$S(R_1, R_2, Q_1, Q_2) = Q_2^T Q_1 + R_2^T R_1 + \ln(\Omega(R_1, R_2, Q_1, Q_2)). \quad (34)$$

It is therefore straight forward to define the moment, cumulant and vertex generating functions. Note that $R_1(t) = \frac{m}{\sqrt{N}} \sum_j x_j(t)$ is the mean input and $Q_1(t, s) = \frac{g^2}{N} \sum_j x_j(t) x_j(s)$ the effective input variance generated by the disordered connectivity.

XI. DERIVATION OF MFT AS TREE LEVEL APPROXIMATION FOR ONE POPULATION

We know that generally, to zeroth order, the vertex generating function is given by $\Gamma_0 = -S_0$. As expressed by the equation of state, the minimum of Γ gives the mean value of its argument. At zeroth order, this corresponds to the saddle-point approximation of the probability distribution, or the neglect of fluctuations (note that here Q_1 represents fluctuations in the system, but is just a variable in $S(R_1, R_2, Q_1, Q_2)$, so that its mean is taken into

account). The solution of the equation of state for vanishing external input $\frac{d\Gamma_0}{dq}\Big|_{q^*} = 0$ gives then

$$R_1^*(t) = -\frac{\delta \ln \Omega(R_1, R_2, Q_1, Q_2)}{\delta R_2(t)} \quad (35)$$

$$= \frac{1}{\mathcal{Z}(0)} \frac{m}{\sqrt{N}} \sum_j \langle x_j \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$$

$$R_2^*(t) = -\frac{\delta \ln \Omega(R_1, R_2, Q_1, Q_2)}{\delta R_1(t)}$$

$$= \frac{1}{\mathcal{Z}(0)} \sum_i \langle \hat{h}_i \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$$

$$Q_1^*(t, s) = -\frac{\delta \ln \Omega(R_1, R_2, Q_1, Q_2)}{\delta Q_2(t, s)} \quad (36)$$

$$= \frac{1}{\mathcal{Z}(0)} \frac{g^2}{N} \sum_j \langle x_j(t) x_j(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}.$$

$$Q_2^*(t, s) = -\frac{\delta \ln \Omega(R_1, R_2, Q_1, Q_2)}{\delta Q_1(t, s)}$$

$$= -\frac{1}{\mathcal{Z}(0)} \frac{1}{2} \sum_i \langle \hat{h}_i(t) \hat{h}_i(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}.$$

The R^*, Q^* at the average means here that these variables are set fixed to their mean value, eg. by considering them as delta distributed around this value. Also note that the averages here are automatically normalized by $\mathcal{Z}(0) = \langle 1 \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$ due to the derivative of the logarithm. As a consequence of the normalization the moments of the response field $\hat{\mathbf{h}}$ should be zero (Coolen 2000b), so one can see that also $R_2(t) = Q_2(t, s) = 0$.

A. Time dependent mean activity

We now solve (35) explicitly for binary neurons.

$$\begin{aligned} R_1^*(t) &= \frac{1}{\mathcal{Z}(0)} \sqrt{Nm} \int \mathcal{D}\mathbf{h} \underbrace{\mathcal{D}\hat{\mathbf{h}} \exp\left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1^*) + \hat{\mathbf{h}}^T Q_1^* \hat{\mathbf{h}}\right)}_{\rho^*[\mathbf{h}]} \frac{1}{N} \sum_i \int \mathcal{D}\mathbf{x} x_i \rho[\mathbf{x}|\mathbf{h}]. \\ &= \frac{1}{\mathcal{Z}(0)} \sqrt{Nm} \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] \frac{1}{N} \sum_i \int \mathcal{D}x_i x_i \rho[x_i|h_i] \end{aligned} \quad (37)$$

where we used that $\int \mathcal{D}x_j x_i \rho[x_j|h_j] = x_i$ for $i \neq j$ due to the normalization. The explicit form of $\rho[x_i|h_i]$ for binary neurons with a Heaviside activation function is given by

$$\begin{aligned} \rho[x_i|h_i](t) &= x_i(t) p[h_i](t) + (1 - x_i(t)) (1 - p[h_i](t)) \\ &= (2x_i(t) - 1) p[h_i](t) + 1 - x_i(t) \end{aligned} \quad (38)$$

$$p[h_i](t) = \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} H(h_i(t') - \theta),$$

leading to

$$\begin{aligned}
R_1^*(t) &= \frac{1}{\mathcal{Z}(0)} \sqrt{Nm} \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] \frac{1}{N} \sum_i \int \mathcal{D}x_i \left((2x_i^2 - x_i) p_i[h_i](t) + x_i - x_i^2 \right) \\
&= \frac{1}{\mathcal{Z}(0)} \sqrt{Nm} \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] \frac{1}{N} \sum_i \sum_{x_i=0}^1 x_i p_i[h_i](t) \\
&= \frac{1}{\mathcal{Z}(0)} \sqrt{Nm} \frac{1}{N} \sum_i \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] H(h_i(t') - \theta),
\end{aligned} \tag{39}$$

where we used $x_i^2 = x_i$ and that the path integral amounts to a sum over the states 0 and 1. Before inserting $\rho^*[\mathbf{h}]$ into this equation, we perform the path integral over $\hat{\mathbf{h}}$ by noting that it is the Fourier transform of a Gaussian

$$\begin{aligned}
\rho^*[\mathbf{h}] &= \int_{-i\infty}^{i\infty} \mathcal{D}\hat{\mathbf{h}} \exp \left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1^*) + \frac{1}{2} \hat{\mathbf{h}}^T Q_1^* \hat{\mathbf{h}} \right) \\
&= \mathcal{F} \left[\exp \left(-\frac{1}{2} x^T Q_1^* x \right) \right] (\mathbf{h} - R_1^*(t)) \\
&= \mathcal{Z}_{\hat{\mathbf{h}}}(0) \exp \left(-\frac{1}{2} (\mathbf{h} - R_1^*)^T Q_1^{*-1} (\mathbf{h} - R_1^*) \right) \\
&= \prod_i \rho^*[h_i]
\end{aligned} \tag{40}$$

with Q_1^{*-1} being the inverse function in the sense

$$\int ds Q_1^*(t, s) Q_1^{*-1}(s, u) = \delta(t - u). \tag{41}$$

The mathematical validity of these steps can be confirmed by defining the continuous time path integral as the limit of a discrete version of M time steps leading from $t_0 = -\frac{T}{2}$ to $t_0 + M\Delta t = t_M = \frac{T}{2}$:

$$\int \mathcal{D}x := \lim_{T \rightarrow \infty} \lim_{\Delta t \rightarrow 0} \left(\int \prod_t dx_t \right). \tag{42}$$

An integral of the form $\mathbf{x}^T \mathbf{A} \mathbf{x}$ then becomes a matrix product $\sum_{s,u} -\frac{(\Delta t)^2}{2} x_s A_{su} x_u$. This allows a rigorous treatment of the gaussian integral, see Appendix Section XXIV A.

This is also a good point to calculate the full normalization factor:

$$\begin{aligned}
\mathcal{Z}(0) &= \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] \int \mathcal{D}x_i \rho[x_i|h_i] \\
&= \underbrace{\mathcal{Z}_{\hat{\mathbf{h}}}(0) \int \mathcal{D}\mathbf{h} \exp \left(-\frac{1}{2} (\mathbf{h} - R_1^*)^T Q_1^{*-1} (\mathbf{h} - R_1^*) \right)}_{=:\mathcal{Z}_{\mathbf{h}}(0)} \underbrace{\sum_{x_i=0}^1 ((2x_i - 1)) p_i[h_i](t) + 1 - x_i}_{=1} \\
&= \mathcal{Z}_{\hat{\mathbf{h}}}(0) \mathcal{Z}_{\mathbf{h}}(0).
\end{aligned} \tag{43}$$

Now, inserting $\rho^*[\mathbf{h}]$ into (39)

$$\begin{aligned}
R_1^*(t) &= \frac{\mathcal{Z}_{\hat{\mathbf{h}}}(0)}{\mathcal{Z}(0)} \sqrt{Nm} \frac{1}{N} \sum_i \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \int \mathcal{D}\mathbf{h} \exp \left(-\frac{1}{2} (\mathbf{h} - R_1^*)^T Q_1^{*-1} (\mathbf{h} - R_1^*) \right) H(h_i(t') - \theta) \\
&= \frac{\mathcal{Z}_{\hat{\mathbf{h}}}(0) \mathcal{Z}_{\mathbf{h}}(0)^{\frac{N-1}{N}}}{\mathcal{Z}(0)} \sqrt{Nm} \frac{1}{N} \sum_i \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \int \mathcal{D}h_i \exp \left(-\frac{1}{2} (h_i - R_1^*)^T Q_1^{*-1} (h_i - R_1^*) \right) H(h_i(t') - \theta).
\end{aligned} \tag{44}$$

To solve the $\int \mathcal{D}h_i$ integral, we again consider the limit of the discretized time version. $Q_{1,ts}^{*-1}$ is now a positive definite, symmetric matrix. In order to calculate the interesting integral

$$\int \prod_u (\mathcal{D}h_{i,u}) \underbrace{\exp\left(-\frac{1}{2} \sum_{t,s} (h_{i,t} - R_{1,t}^*)^T Q_1^{*-1} (h_{i,s} - R_{1,s}^*)\right)}_{\rho^*[h_i]} H(h_{i,t'} - \theta)$$

we need the marginalized form $\rho^*[h_{i,t'}]$, where all times except t' have been integrated over. To obtain this form, we use the moment generating function, which must be the gaussian with inverted variance

$$Z[j_i] = \exp\left(-\frac{1}{2} \sum_{t,s} j_{i,t} Q_{i,ts}^* j_{i,s}\right).$$

Then the moment generating function of the marginalized distribution is given by setting all $j_{i,t}$ with $t \neq t'$ to zero. This can be seen by the definition of Z as the Fourier transform of the probability distribution, where setting a $j=0$ just amounts to integrating over the corresponding variable without the exponential, thus marginalizing it. Therefore,

$$Z[j_{i,t'}] = \exp\left(-\frac{1}{2} j_{i,t'} Q_{i,t't'}^* j_{i,t'}\right)$$

implying that also in the continuous limit

$$\rho^*[h_i(t')] = \exp\left(-\frac{1}{2} \frac{(h_i(t') - R_1^*)^2}{Q_1^*(t', t')}\right). \quad (45)$$

Combined in (44) with the Heaviside function this gives a complementary errorfunction (defined as $\int_{\theta}^{\infty} \frac{1}{\sqrt{\pi}} \exp(-x^2) = \frac{1}{2} \text{erfc}(\theta)$) with the same prefactor as the gaussian integral without the $H(h_{i,t'} - \theta)$ would produce, so the normalization factors combine seamlessly and the final implicit equation is

$$\begin{aligned} R_1^*(t) &= \frac{\mathcal{Z}_{\mathbf{h}}(0) \mathcal{Z}_{\mathbf{h}}(0)}{\mathcal{Z}(0)} \sqrt{Nm} \frac{1}{N} \sum_i \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \frac{1}{2} \text{erfc}\left(\frac{\theta - R_1^*(t')}{\sqrt{2Q_1^*(t', t')}}\right) \\ &= \sqrt{Nm} \frac{1}{2} \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \text{erfc}\left(\frac{\theta - R_1^*(t')}{\sqrt{2Q_1^*(t', t')}}\right) \end{aligned} \quad (46)$$

$$= \sqrt{Nm} \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \Phi(R_1^*(t'), Q_1^*(t', t')). \quad (47)$$

Using the relation $R_1^*(t) = \sqrt{Nm} \bar{x}(t)$ and taking the derivative on both sides, this becomes (note that the t dependence inside the integral necessitates the use of the product rule)

$$\tau \partial_t \bar{x}(t) + \bar{x}(t) = \frac{1}{2} \text{erfc}\left(\frac{\theta - R_1^*(t)}{\sqrt{2Q_1^*(t, t)}}\right), \quad (48)$$

which is of the same form as the result from dynamic mean field theory (8). We can identify $R_1^*(t)$ with $\mu(t)$, but as we will see in Section XI C, this is not immediately clear for $Q_1^*(t, t)$ and $\sigma(t)$. First, we derive an expression for Q_1^* .

B. Time lagged covariance for the stationary state

We try to solve (36) for the binary neurons:

$$\begin{aligned} Q_1^*(t, s) &= \frac{1}{\mathcal{Z}(0)} \frac{g^2}{N} \sum_i \langle x_i(t) x_i(s) \rangle_{J, \mathbf{h}}^{R^*, Q^*} \\ &= \frac{1}{\mathcal{Z}(0)} \frac{g^2}{N} \sum_i \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] \int \mathcal{D}x_i x_i(t) x_i(s) \rho[x_i(s)|x_i(t)] \rho[x_i(t)|h_i(t)]. \end{aligned} \quad (49)$$

To write down the x_i path integral, we have to introduce a new probability distribution $\rho[x_i(s)|x_i(t)]$. Assume, w.l.o.g. due to $Q_1^*(t, s) = Q_1^*(s, t)$, that $s > t$. A derivation of the general expression made by Tobias Kühn is given in Section XXIV B, but leads to the complicated integro-differential equation (136). Thus we here approximate the distribution by assuming the stationary state and that the effect of a state change in one unit has negligible influence on the mean activity. Then the probabilities of the four different configurations can be expressed in simple terms by the mean activity $\bar{x} = \frac{R_1^*}{\sqrt{Nm}}$ and probability of (no) update $e^{-\frac{t-s}{\tau}}$, leading to

$$\begin{aligned}
\rho^{stat.}[x_i(s)|x_i(t)] &= x_i(s) \left(\frac{R_1^*}{\sqrt{Nm}} + \left(1 - \frac{R_1^*}{\sqrt{Nm}} \right) e^{-\frac{s-t}{\tau}} \right) x_i(t) \\
&\quad + (1 - x_i(s)) \left(1 - \frac{R_1^*}{\sqrt{Nm}} - \left(1 - \frac{R_1^*}{\sqrt{Nm}} \right) e^{-\frac{s-t}{\tau}} \right) x_i(t) \\
&\quad + x_i(s) \left(\frac{R_1^*}{\sqrt{Nm}} - \frac{R_1^*}{\sqrt{Nm}} e^{-\frac{s-t}{\tau}} \right) (1 - x_i(t)) \\
&\quad + (1 - x_i(s)) \left(1 - \frac{R_1^*}{\sqrt{Nm}} + \frac{R_1^*}{\sqrt{Nm}} e^{-\frac{s-t}{\tau}} \right) (1 - x_i(t)) \\
&= 2x_i(s)x_i(t)e^{-\frac{s-t}{\tau}} + x_i(s) \left(-1 + 2\frac{R_1^*}{\sqrt{Nm}} \left(1 - e^{-\frac{s-t}{\tau}} \right) \right) \\
&\quad - x_i(t)e^{-\frac{s-t}{\tau}} + 1 - \frac{R_1^*}{\sqrt{Nm}} + \frac{R_1^*}{\sqrt{Nm}} e^{-\frac{s-t}{\tau}}.
\end{aligned} \tag{50}$$

The normalization is given by summing over the $x_i(s)$ state space only:

$$\begin{aligned}
\sum_{x_i(s)=0}^1 \rho^{stat.}[x_i(s)|x_i(t)] &= 2x_i(t)e^{-\frac{s-t}{\tau}} + \left(-1 + 2\frac{R_1^*}{\sqrt{Nm}} \left(1 - e^{-\frac{s-t}{\tau}} \right) \right) - 2x_i(t)e^{-\frac{s-t}{\tau}} + 2 \left(1 - \frac{R_1^*}{\sqrt{Nm}} + \frac{R_1^*}{\sqrt{Nm}} e^{-\frac{s-t}{\tau}} \right) \\
&= 1.
\end{aligned}$$

When inserted into (49), the expression for $\rho[x_i(s)|x_i(t)]$ is multiplied by $x_i(t)x_i(s)$, so that when executing the sum over the binary state space, only the $x_i(t) = x_i(s) = 1$ state survives, giving

$$\begin{aligned}
Q_1^{*stat.}(t-s) &= \frac{1}{\mathcal{Z}(0)} g^2 \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \left(e^{-\frac{s-t}{\tau}} + \frac{R_1^*}{\sqrt{Nm}} \left(1 - e^{-\frac{s-t}{\tau}} \right) \right) \frac{1}{N} \sum_i \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] H(h_i(t') - \theta) \\
&= g^2 \left(e^{-\frac{s-t}{\tau}} + \frac{R_1^*}{\sqrt{Nm}} \left(1 - e^{-\frac{s-t}{\tau}} \right) \right) \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \frac{1}{2} \text{erfc} \left(\frac{\theta - R_1^*}{\sqrt{2Q_1^*(0)}} \right)
\end{aligned} \tag{51}$$

$$= g^2 \left(e^{-\frac{s-t}{\tau}} + \frac{R_1^*}{\sqrt{Nm}} \left(1 - e^{-\frac{s-t}{\tau}} \right) \right) \frac{R_1^*}{\sqrt{Nm}}. \tag{52}$$

Here we made explicit that Q_1^* is only a function of the time lag $t-s$ for the stationary state and used the result (46) on the integrals. By using $R_1^*(t) = \sqrt{Nm}\bar{x}(t)$ and $s > t$, the result can be written in the more compact form

$$Q_1^*(\Delta t) = g^2 \bar{x} \left(e^{-\frac{|\Delta t|}{\tau}} + \bar{x} \left(1 - e^{-\frac{|\Delta t|}{\tau}} \right) \right). \tag{53}$$

For zero time lag, as needed in (48), we obtain

$$Q_1^*(t, t) = g^2 \bar{x}.$$

Other than the expression for $Q_1^*(\Delta t)$, this result is exact, as $\rho[x_i(t)|x_i(t)] = 1$. It can also be understood by trivially noting that due to the binary variables, $x_i^2 = x_i$, so $Q_1^*(t, t) = \frac{g^2}{N} \sum_i \langle x_i^2(t) \rangle_{R^*Q^*} = g^2 \bar{x}(t)$.

C. Fixed vs. free indegree and the connection to standard MFT

We find that the difference between fixed and free indegree is that Q_1^* contains the autocovariances $\langle \delta x_j \delta x_j \rangle_{\mathbf{J}, \mathbf{h}}$ in the first, and the autocorrelations $\langle x_j x_j \rangle_{\mathbf{J}, \mathbf{h}}$ in the latter case (R_1^* is the same). Then we proceed to compare our results with the ad-hoc MF equations used in Part II.

To obtain $Q_1^{*fixed}(t, s)$, we need to carry out the disorder average $\langle \exp(\hat{\mathbf{h}} \mathbf{J} \mathbf{x}) \rangle_{\mathbf{J}}$ with the additional constraint that $\sum_i J_{ij} = Np$. The full calculation can be found in the appendix Section XXIV D, resulting in the slightly changed definition of the Hubbard-Stratonovich field

$$Q_1^{fixed}(t, s) = \frac{g^2}{N-1} \sum_i x_i(t) \delta x_i(s).$$

Then the saddlepoint approximation gives

$$Q_1^{*fixed}(t, s) = Q_1^{*distr.}(t, s) - \frac{g^2}{N(N-1)} \sum_{i \neq j} \bar{x}(t) \bar{x}(s) - \frac{g^2}{N(N-1)} \sum_{i \neq j} \underbrace{\langle \delta x_i(t) \delta x_j(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}}_{=0}, \quad (54)$$

where the last term vanishes in zero order approximation, because as shown in (40) there is no coupling between units as $\rho^*[\mathbf{h}] = \prod_i \rho^*[h_i]$ factorizes. This is because of $R_2^* = Q_2^* = 0$, which removes the couplings seen in the general expression (33). Thus

$$Q_1^{*fixed}(t, s) = Q_1^{*distr.}(t, s) - g^2 \bar{x}(t) \bar{x}(s) = g^2 \bar{x}(1 - \bar{x}), \quad (55)$$

Now we can compare the obtained tree level approximation (48) to the 'ad-hoc' MF equation (8). As mentioned at the end of Section XI A, the equations have the same structure and $\mu(t)$ can be identified with $R_1^*(t)$. Only between $\sigma^2(t)$ and $Q_1^*(t, t)$ a difference emerges:

$$\begin{aligned} Q_1^{*fixed} &= J_0^2 p(1-p) \bar{x}(1-\bar{x}) \\ \sigma^2 &= J_0^2 p \bar{x}(1-\bar{x}). \end{aligned} \quad (56)$$

This difference, which is small for $p \ll 1$, is rooted in the way we construct the two approximations. Consider the (biologically unrealistic) limiting case of a fully connected network: Then, $p = 1$ which means that $Q_1^{*fixed} = 0$. This is because $Q_1(t, s)$ represents an effective noise generated by the structural disorder in the system, and there is no such disorder in the fully connected network. Standard mean field theory skips over this issue by assuming that the neuron activities are independently and identically distributed, and can thus incorporate also the activity fluctuation that is not due to disorder but due to the autocovariance. This corresponds to including gaussian fluctuations of R_1 , but neglecting crosscovariances, as seen by calculating $\langle (R_1 - R_1^*)^2 \rangle_{\Omega(R, Q)} = \frac{m^2}{N} \sum_{i,j} \langle \delta x_i \delta x_j \rangle_{\Omega(R, Q)} \stackrel{i.i.d.}{=} J_0^2 p^2 \bar{x}(1-\bar{x})$, which exactly cancels the $(1-p)$ in $\tilde{Q}_1^*$ after redefining $\tilde{Q}_1 = Q_1 + \delta R_1^2$. Notice that assuming i.i.d. x_i in the fully connected network (where $p = 1$) is only justified because of the random update process (as all neurons see the same input), which is information specific to the neuron model and therefore not used a priori in our path integral expansion. If MFT is defined as a gaussian approximation of the input h with the additional assumption of i.i.d. x_i , then

$$\langle h_i \rangle_{\mathbf{J}, \mathbf{x}} = \sum_j \langle J_{ij} x_j \rangle_{\mathbf{J}, \mathbf{x}} \stackrel{i.i.d. x_j}{=} \mu \quad \text{and} \quad \langle \delta h_i^2 \rangle_{\mathbf{J}, \mathbf{x}} = \sum_{j,k} \langle J_{ij} J_{ik} x_j x_k \rangle_{\mathbf{J}, \mathbf{x}} - \sum_{j,k} \langle J_{ij} x_j \rangle_{\mathbf{J}, \mathbf{x}} \langle J_{ik} x_k \rangle_{\mathbf{J}, \mathbf{x}} \stackrel{i.i.d. x_j}{=} \sigma^2. \quad (57)$$

In comparison, the variables R_1^* and Q_1^* are given by (if we assume $\langle J_{ij} J_{mn} \rangle = \delta_{im} \langle J_{ij} J_{in} \rangle$, that is row-wise i.i.d. J_{ij})

$$R_1^* = \sum_j \langle J_{ij} \rangle_{\mathbf{J}} \langle x_j \rangle_{\mathbf{x}} \quad \text{and} \quad Q_1^* = \sum_{j,k} \langle J_{ij} J_{ik} \rangle_{\mathbf{J}} \langle x_j x_k \rangle_{\mathbf{x}}. \quad (58)$$

In general, if the distributions of J_{ij} and x_j are independent, we can show the relation between the approaches:

$$\begin{aligned}
\langle \delta h_i^2 \rangle_{\mathbf{J}, \mathbf{x}} &= \sum_{j,k} \langle J_{ij} J_{ik} x_j x_k \rangle_{\mathbf{J}, \mathbf{x}} - \sum_{j,k} \langle J_{ij} x_j \rangle_{\mathbf{J}, \mathbf{x}} \langle J_{ik} x_k \rangle_{\mathbf{J}, \mathbf{x}} \\
&= \sum_{j,k} \langle J_{ij} J_{ik} \rangle_{\mathbf{J}} \langle x_j x_k \rangle_{\mathbf{x}} - \sum_{j,k} \langle J_{ij} \rangle_{\mathbf{J}}^2 \langle x_j \rangle_{\mathbf{x}}^2 \\
&= \sum_{j,k} (\langle J_{ij} J_{ik} \rangle_{\mathbf{J}} + \langle J_{ij} \rangle_{\mathbf{J}} \langle J_{ik} \rangle_{\mathbf{J}}) \langle x_j x_k \rangle_{\mathbf{x}} - \sum_j \langle J_{ij} \rangle_{\mathbf{J}}^2 \langle x_j \rangle_{\mathbf{x}}^2 \\
&= \sum_{j,k} \langle J_{ij} J_{ik} \rangle_{\mathbf{J}} \langle x_j x_k \rangle_{\mathbf{x}} + \sum_{j,k} \langle J_{ij} \rangle_{\mathbf{J}}^2 \langle x_j x_k \rangle_{\mathbf{x}} \\
&= Q_1^* + \langle R_1 \rangle_{\Omega(R,Q)} \stackrel{i.i.d. x_j}{=} \sigma^2.
\end{aligned} \tag{59}$$

Looking at (57) and (58), R_1^* and Q_1^* have the advantages, that the two averages are separated without any assumptions on the x_i , the contributions of the disorder are made explicit and that the tools provided by the path integral form are available. However, the expansion differs from an 'intuitive' cumulant hierarchy in h , as constructed by Buice et al. (2010). Also, if the J_{ij} statistics readily allow to calculate μ and σ and no insight beyond the MF eqs. is needed, standard mean field theory is simple to use. It could be interesting to think of alternative expansions, e.g. exactly obtaining MFT at zeroth order, or a double expansion making explicit both orders of $\frac{1}{N}$ and network-generated fluctuations.

XII. PROPAGATORS

While we used only the maximum values of R and Q in the saddle point approximation of Section XI, that is, approximated the probability distributions of R_1, R_2, Q_1, Q_2 by delta distributions at their peak, we will incorporate fluctuations around this value. The distribution is now approximated as a gaussian, the maximum of which corresponds to the maximum of the full distribution. Because of

$$\rho[R_1, R_2, Q_1, Q_2] = \exp(S[R_1, R_2, Q_1, Q_2]),$$

we are interested in the second derivatives of the action (34):

$$S_{11}^{(RR)}(t, s) := \frac{\delta^2 S}{\delta R_1(t) \delta R_1(s)} \Big|_{R^*, Q^*} = \left\langle \left(\sum_i \hat{h}_i(t) \right) \left(\sum_i \hat{h}_i(s) \right) \right\rangle_{Q^*, R^*} = 0$$

$$S_{12}^{(RR)}(t, s) \equiv S_{21}^{(RR)}(s, t) := \frac{\delta^2 S}{\delta R_1(t) \delta R_2(s)} \Big|_{R^*, Q^*} = \delta(t-s) + \frac{m}{\sqrt{N}} \left\langle \sum_{i,j} \hat{h}_i(t) x_j(s) \right\rangle_{Q^*, R^*} \tag{60}$$

$$S_{22}^{(RR)}(t, s) := \frac{\delta^2 S}{\delta R_2(t) \delta R_2(s)} \Big|_{R^*, Q^*} = \frac{m^2}{N} \left\langle \sum_{i,j} x_i(t) x_j(s) \right\rangle_{Q^*, R^*} - \frac{m^2}{N} \left\langle \sum_i x_i(t) \right\rangle_{Q^*, R^*} \left\langle \sum_j x_j(s) \right\rangle_{Q^*, R^*} \tag{61}$$

and

$$S_{11}^{(QQ)}(t, s, u, v) := \frac{\delta^2 S}{\delta Q_1(t, s) \delta Q_1(u, v)} \Big|_{R^*, Q^*} := \frac{1}{4} \left\langle \sum_{i,j} \hat{h}_i(t) \hat{h}_i(s) \hat{h}_j(u) \hat{h}_j(v) \right\rangle_{Q^*, R^*} = 0$$

$$S_{12}^{(QQ)}(t, s, u, v) \equiv S_{21}^{(QQ)}(u, v, t, s) := \delta(t-u) \delta(s-v) + \frac{g^2}{2N} \left\langle \sum_{i,j} \hat{h}_i(t) \hat{h}_i(s) x_j(u) x_j(v) \right\rangle_{Q^*, R^*} \tag{62}$$

$$S_{22}^{(QQ)}(t, s, u, v) := \frac{g^4}{N^2} \left\langle \sum_{i,j} x_i(t) x_i(s) x_j(u) x_j(v) \right\rangle_{Q^*, R^*} - Q_1^*(t, s) Q_1^*(u, v) \tag{63}$$

as well as

$$S_{11}^{(RQ)}(t, u, v) \equiv S_{11}^{(QR)}(u, v, t) := \frac{1}{2} \left\langle \sum_{i,j} \hat{h}_i(t) \hat{h}_j(u) \hat{h}_j(v) \right\rangle_{Q^*, R^*} = 0$$

$$S_{12}^{(RQ)}(t, u, v) \equiv S_{21}^{(QR)}(u, v, t) := \frac{g^2}{N} \left\langle \sum_{i,j} \hat{h}_i(t) x_j(u) x_j(v) \right\rangle_{Q^*, R^*} \quad (64)$$

$$S_{21}^{(RQ)}(t, u, v) \equiv S_{12}^{(QR)}(u, v, t) := \frac{m}{2\sqrt{N}} \left\langle \sum_{i,j} x_i(t) \hat{h}_j(u) \hat{h}_j(v) \right\rangle_{Q^*, R^*} \quad (65)$$

$$S_{22}^{(RQ)}(t, u, v) \equiv S_{22}^{(QR)}(u, v, t) := \frac{m g^2}{N^{\frac{3}{2}}} \left\langle \sum_{i,j} x_i(t) x_j(u) x_j(v) \right\rangle_{Q^*, R^*} - R_1^*(t) Q_1^*(u, v). \quad (66)$$

For explicit treatment however, we will only consider those terms with two derivatives in R , because the other terms are suppressed by at least $\frac{1}{\sqrt{N}}$. $S_{12}^{(RR)}(t, s)$ can be calculated by explicitly solving the gaussian integrals, which is longish and care must be taken with the time arguments, see the attempt in Section XXIV C. Instead, we note that it is given as the derivative of the right hand side of (35), which was already calculated up to (46) without using the self consistency equation itself, as this must be done after taking the derivatives. This gives a much simpler method, that is also applicable to higher derivatives.

$$S_{12}^{(RR)}(t, s) = \frac{\partial}{\partial R_1(t)} \frac{\partial S}{\partial R_2(s)} \Big|_{R^*, Q^*}$$

$$= \delta(t-s) + \sqrt{N} m \frac{1}{2} \int_{-\infty}^s \frac{ds'}{\tau} \frac{\delta}{\delta R_1(t)} \Big|_{R^*, Q^*} e^{-\frac{s-s'}{\tau}} \operatorname{erfc} \left(\frac{\theta - R_1^*(s')}{\sqrt{2Q_1^*(s', s')}} \right)$$

We see here, that the $S_{12}^{(RR)}(t, s)$ is causal, ie. is only nonzero for $t < s$. The nonzero part is then, due to $\partial_r \frac{1}{2} \operatorname{erfc}(\theta(r)) = \frac{\partial \theta}{\partial r} \frac{\partial}{\partial \theta} \int_{\theta}^{\infty} dx \frac{1}{\sqrt{\pi}} \exp(-x^2) = -\frac{1}{\sqrt{\pi}} \exp(-\theta^2) \frac{\partial \theta}{\partial r}$:

$$S_{12}^{(RR)}(t, s) = \delta(t-s) + \sqrt{N} m H(s-t) \int_{-\infty}^s \frac{ds'}{\tau} e^{-\frac{s-s'}{\tau}} \delta(t-s') \frac{1}{\sqrt{2\pi Q_1^*(s', s')}} \exp \left(-\frac{1}{2} \frac{(\theta - R_1^*(s'))^2}{Q_1^*(s', s')} \right)$$

$$= \delta(t-s) + \sqrt{N} m H(s-t) \frac{1}{\tau} e^{-\frac{s-t}{\tau}} \frac{1}{\sqrt{2\pi Q_1^*(t, t)}} \exp \left(-\frac{1}{2} \frac{(\theta - R_1^*(t))^2}{Q_1^*(t, t)} \right). \quad (67)$$

$$= \delta(t-s) + \frac{m}{\sqrt{N}} K(t, s). \quad (68)$$

Therefore, in the stationary case:

$$K(\Delta t) = H(\Delta t) \frac{1}{\tau} e^{-\frac{\Delta t}{\tau}} \underbrace{N \left(\frac{\sqrt{N} m}{2\pi g^2 R_1^*} \right)^{\frac{1}{2}} \exp \left(-\frac{1}{2} \sqrt{N} m \frac{(\theta - R_1^*)^2}{g^2 R_1^*} \right)}_{\equiv k(R_1^*)}.$$

Note that from now on we use as a convention the notation $f(t, s) \rightarrow f(\Delta t)$ with $\Delta t = s - t$ for stationary cases. The Fourier transformed $K(\Delta t)$ is

$$K(\omega) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} d\Delta t \frac{k(R_1^*)}{\tau} \exp \left(-\Delta t \left(\frac{1}{\tau} + i\omega \right) \right)$$

$$= \frac{k(R_1^*)}{\sqrt{2\pi}} \frac{-1}{1 + i\tau\omega}. \quad (69)$$

Note that we use the convention $\mathcal{F}[f](\omega) = \frac{1}{\sqrt{2\pi}} \int dt f(t) e^{-i\omega t}$. Next we consider $S_{22}^{(RR)}(t, s)$, and splitting (61) into auto- and crosscorrelations, we use that the crosscovariances vanish in saddle point approximation,

$$\begin{aligned} S_{22}^{(RR)}(t, s) &= \frac{m^2}{N} \left\langle \sum_i x_i(t) x_i(s) \right\rangle_{Q^*, R^*} + \frac{m^2}{N} \left\langle \sum_{i \neq j} x_i(t) x_j(s) \right\rangle_{Q^*, R^*} - R_1^*(t) R_1^*(s) \\ &= \frac{m^2}{g^2} Q_1^*(t, s) + \frac{m^2}{N} N(N-1) \frac{R_1^*(t) R_1^*(s)}{Nm^2} - R_1^*(t) R_1^*(s) \\ &= \frac{m^2}{g^2} Q_1^*(t, s) - \frac{1}{N} R_1^*(t) R_1^*(s) \end{aligned}$$

so only the autocovariances remain. Note that we use 'correlation' to signify the second moment and 'covariance' for the second cumulant. In the stationary case, $Q_1^*(\Delta t)$ is given by (53), so that

$$S_{22}^{(RR)}(\Delta t) = m^2 \bar{x}(1 - \bar{x}) e^{-\frac{|\Delta t|}{\tau}}. \quad (70)$$

To determine the propagators, we need to invert the hessian matrix $\mathbf{S}^{(2)}(t, s)$ in the sense (see (135))

$$\int \mathbf{S}^{(2)}(t, s) \mathbf{\Delta}(s, r) ds = \text{diag}(\delta(t - r)). \quad (71)$$

Neglecting all components with derivatives in Q and using $S_{11}^{(RR)}(t, s) = 0$, this leads to

$$\int S_{12}^{(RR)}(t, s) \Delta_{21}^{(RR)}(s, u) ds = \delta(t - u) \quad (72)$$

$$\int S_{12}^{(RR)}(t, s) \Delta_{22}^{(RR)}(s, u) ds = 0 \quad (73)$$

$$\int S_{21}^{(RR)}(t, s) \Delta_{11}^{(RR)}(s, u) + S_{22}^{(RR)}(t, s) \Delta_{21}^{(RR)}(s, u) ds = 0 \quad (74)$$

$$\int S_{21}^{(RR)}(t, s) \Delta_{12}^{(RR)}(s, u) + S_{22}^{(RR)}(t, s) \Delta_{22}^{(RR)}(s, u) ds = \delta(t - u). \quad (75)$$

From (73) we see that $\Delta_{22}^{(RR)}(s, u) = 0$, and therefore the second term in (75) vanishes. $\Delta_{22}^{(RR)} = 0$ also means that not only the mean of R_2^* vanishes (35), but also its fluctuations; thus $R_2 = 0$.

(72) and (75) allow to write $\Delta_{21}^{(RR)} = (S_{12}^{(RR)})^{-1}$ and $\Delta_{12}^{(RR)} = (S_{21}^{(RR)})^{-1}$ so that because of (60)

$$\Delta_{21}^{(RR)}(t, s) = \Delta_{12}^{(RR)}(s, t). \quad (76)$$

We can solve (74) by multiplying from the left with $\Delta_{12}^{(RR)}$, which yields

$$\begin{aligned} 0 &= \int \underbrace{\int \Delta_{12}^{(RR)}(r, t) S_{21}^{(RR)}(t, s) dt}_{\delta(r-s)} \Delta_{11}^{(RR)}(s, u) ds + \int \int \Delta_{12}^{(RR)}(r, t) S_{22}^{(RR)}(t, s) \Delta_{21}^{(RR)}(s, u) ds dt \\ &\Rightarrow \Delta_{11}^{(RR)}(r, u) = - \int \int \Delta_{12}^{(RR)}(r, t) S_{22}^{(RR)}(t, s) \Delta_{21}^{(RR)}(s, u) ds dt. \end{aligned} \quad (77)$$

To evaluate this expression, we first need the explicit forms of $\Delta_{21}^{(RR)}$ and $\Delta_{12}^{(RR)}$. Starting from (72) we use (68):

$$\begin{aligned} \delta(u - t) &= \int S_{12}^{(RR)}(t, s) \Delta_{21}^{(RR)}(s, u) ds \\ &= \Delta_{21}^{(RR)}(t, u) + \frac{m}{\sqrt{N}} \int K(t, s) \Delta_{21}^{(RR)}(s, u) ds. \end{aligned} \quad (78)$$

For the stationary state, we can solve the Dyson equation in Fourier domain by the convolution theorem:

$$\begin{aligned}
\delta(u-t) &= \Delta_{21}(u-t) + \frac{m}{\sqrt{N}} \int \Delta_{21}(u-s) K(s-t) ds \\
&= \Delta_{21}(u-t) + \frac{m}{\sqrt{N}} \int \Delta_{21}(u-t-\bar{s}) K(\bar{s}) d\bar{s} \\
&= \Delta_{21}(u-t) + \frac{m}{\sqrt{N}} (\Delta_{21} * K)(u-t)
\end{aligned}$$

doing a FT in $u-t$, the convolution becomes a product

$$\frac{1}{\sqrt{2\pi}} = \Delta_{21}(\omega) + \frac{m}{\sqrt{N}} \sqrt{2\pi} \Delta_{21}(\omega) K(\omega)$$

and we can solve for the propagator in Fourier domain, giving

$$\Delta_{21}(\omega) = \frac{1}{\sqrt{2\pi}} \frac{1}{1 + \frac{m}{\sqrt{N}} \sqrt{2\pi} K(\omega)}. \quad (79)$$

Thus, using the result (69), the propagator is given by

$$\begin{aligned}
\Delta_{21}^{(RR)}(\omega) &= \frac{1}{\sqrt{2\pi}} \frac{1}{1 + \frac{mk(R_1^*)}{\sqrt{N}} \frac{-1}{1+i\tau\omega}} \\
&= \frac{1}{\sqrt{2\pi}} \frac{1+i\tau\omega}{1-\tilde{k}+i\tau\omega} \quad (80)
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \left(1 + \frac{\tilde{k}}{1-\tilde{k}+i\tau\omega} \right). \quad (81)$$

The interpretation of $\tilde{k}$ is the derivative of the input gain function. The form is $\tilde{k} = \sqrt{N}m\partial_{R_1}\Phi(R_1, Q_1)$, that is the derivative of the right part of the stationary MF eq. $R_1^* = \sqrt{N}m\Phi(R_1^*, Q_1^*)$, thus we know that $\tilde{k} \leq 1$. Therefore $\Delta_{21}^{(RR)}(\omega)$ has no pole on the real axis, except at a critical point, when $\tilde{k} = 1$.

Next we would like to calculate Δ_{11} . For this purpose we notice that for the stationary state, (77) has a convolution structure and simplifies in Fourier space. We again use that the second time argument should be larger than the first, defining this as positive time lag.

$$\begin{aligned}
\Delta_{11}^{(RR)}(u-r) &= - \int dt \Delta_{12}^{(RR)}(t-r) \int ds S_{22}^{(RR)}(s-t) \Delta_{21}^{(RR)}(u-s) \\
&= \dots = - \left(\left(\Delta_{21}^{(RR)} * S_{22}^{(RR)} \right) * \Delta_{12}^{(RR)} \right)(u-r) \quad (82)
\end{aligned}$$

(see (152) for details) so that in Fourier space

$$\begin{aligned}
\Delta_{11}^{(RR)}(\omega) &= -2\pi \Delta_{21}^{(RR)}(\omega) S_{22}^{(RR)}(\omega) \Delta_{12}^{(RR)}(\omega) \\
&= -2\pi \Delta_{21}^{(RR)}(\omega) S_{22}^{(RR)}(\omega) \Delta_{21}^{(RR)}(-\omega),
\end{aligned}$$

where we used $\Delta_{21}^{(RR)}(t, s) = \Delta_{12}^{(RR)}(s, t)$ and that $\int dt f(-t) e^{-i\omega t} = \int dt' f(t') e^{-i(-\omega)t'} = \mathcal{F}[f](-\omega)$.

The Fourier transform of $S_{22}(\Delta t)$ is needed: (for details see (153))

$$\begin{aligned}
S_{22}(\omega) &= m^2 \bar{x}(1-\bar{x}) \mathcal{F}[e^{-\frac{|\Delta t|}{\tau}}](\omega) \\
&= \dots = -\frac{m^2 \tau}{\sqrt{2\pi}} \bar{x}(1-\bar{x}) \frac{2}{1+(\tau\omega)^2}. \quad (83)
\end{aligned}$$

We can thus calculate $\Delta_{11}^{(RR)}$:

$$\begin{aligned}\Delta_{11}^{(RR)}(\omega) &= -2\pi\Delta_{21}^{(RR)}(\omega)S_{22}^{(RR)}(\omega)\Delta_{21}^{(RR)}(-\omega) \\ &= \frac{m^2\tau}{\sqrt{2\pi}}\bar{x}(1-\bar{x})\frac{2}{1+(\tau\omega)^2}\frac{1+i\tau\omega}{1-\tilde{k}+i\tau\omega}\cdot\frac{1-i\tau\omega}{1-\tilde{k}-i\tau\omega} \\ &= \frac{m^2\tau}{\sqrt{2\pi}}\bar{x}(1-\bar{x})\frac{2}{(1-\tilde{k})^2+\tau^2\omega^2}.\end{aligned}\tag{84}$$

A good crosscheck is provided by noting that the terms appearing in (84) have exactly the same structure as those appearing in the expression for the cross spectrum of linearized rate models, see eq. (8) in Grytskyy et al. (2013).

By integrating out ω , we obtain the zero time lag propagator (due to $\int d\omega \hat{f}(\omega) = \int d\omega \int dt f(t) \exp(-\omega t) = \int dt f(t)\delta(t) = f(0)$)

$$\Delta_{11}^{(RR)}(0) = \frac{m^2\tau}{\sqrt{2\pi}}\bar{x}(1-\bar{x}) \int d\omega \frac{2}{(1-\tilde{k})^2+\tau^2\omega^2}$$

where the integral can be conveniently computed using the residue theorem. For this we close the integration curve in a semicircle on the half plane of positive i . The added curve does not contribute to the integral, since after noticing $|\omega^2| \geq |\omega|^2$ we can estimate

$$\lim_{R \rightarrow \infty} \left| \int_{\text{arc}(R)} \frac{1}{(1-\tilde{k})^2+\tau^2\omega^2} \right| \leq \lim_{R \rightarrow \infty} \int_{\text{arc}(R)} \frac{1}{1+(\tau|\omega|)^2} = \lim_{R \rightarrow \infty} \frac{\pi R}{1+(\tau R)^2} = 0.$$

Thus we replace the integral by the residue at $\omega = i\frac{1-\tilde{k}}{\tau}$, which can be read off from the partial fraction decomposition

$$\frac{2}{(1-\tilde{k})^2+\tau^2\omega^2} = \frac{1}{(1-\tilde{k})} \left(\frac{1}{1-\tilde{k}+i\tau\omega} + \frac{1}{1-\tilde{k}-i\tau\omega} \right) = \frac{1}{i\tau(1-\tilde{k})} \left(\frac{1}{\omega - i\frac{1-\tilde{k}}{\tau}} - \frac{1}{\omega + i\frac{1-\tilde{k}}{\tau}} \right),$$

leading to the result

$$\begin{aligned}\Delta_{11}^{(RR)}(0) &= \frac{m^2\tau}{\sqrt{2\pi}}\bar{x}(1-\bar{x})2\pi i \frac{1}{i\tau(1-\tilde{k})} \\ &= \sqrt{2\pi} \frac{m^2}{1-\tilde{k}} \bar{x}(1-\bar{x}).\end{aligned}\tag{85}$$

As with $\Delta_{21}^{(RR)}$, we can see the divergence for $\tilde{k} \rightarrow 1$ that corresponds to critical points.

XIII. 1-LOOP CORRECTION TO THE MEAN ACTIVITY

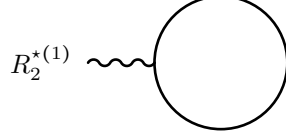
The effective action is given by

$$\Gamma(R, Q) = -S(R, Q) + \Gamma_{\text{fl}}(R, Q).\tag{86}$$

Γ_{fl} can be expressed diagrammatically, for which the propagators $\Delta_{11}^{(RR)}$, $\Delta_{12}^{(RR)}$, $\Delta_{22}^{(RR)}$ and vertices $V_n = \frac{\partial^n S}{\partial R^n} \Big|_{R=R^*}$ given by the higher derivatives of the action are needed. However, if we are interested in the one-loop correction to R_1^* only,

$$\frac{\partial \Gamma^{(1)}}{\partial R_2} \Big|_{\star(1)} = -R_1^{\star(1)} - \frac{\partial \ln \Omega}{\partial R_2} \Big|_{\star(1)} + \frac{\partial \Gamma_{fl}^{(1)}}{\partial R_2} \Big|_{\star(1)} = 0,\tag{87}$$

and there is just one type of possible diagrams:



We therefore need only the three point vertices $V_{112}^{(RRR)}$, $V_{122}^{(RRR)}$ and $V_{222}^{(RRR)}$. Furthermore, the diagram containing $V_{222}^{(RRR)}$ necessarily has its loop closed by $\Delta_{22}^{(RR)} = 0$ and can therefore be discarded. Also loops closed by the response function $\Delta_{21}^{(RR)}$ should vanish. However, we were not able to show this immediately. We calculate here only the $V_{112}^{(RRR)} \Delta_{22}^{(RR)}$ diagram, and leave the response loop to future work.

The vertex is

$$\begin{aligned}
 V_{112}^{(RRR)}(t, s, u) &= \frac{1}{3!} \left. \frac{\partial^3 S}{\partial R_1(t) \partial R_1(s) \partial R_2(u)} \right|_{R, Q} = \frac{1}{3!} \left. \frac{\partial}{\partial R_1(t)} S_{12}^{(RR)}(s, u) \right|_{R, Q} \\
 &= \frac{1}{3!} \frac{\partial}{\partial R_1(t)} H(u-s) \sqrt{N} m \frac{1}{\tau} e^{-\frac{u-s}{\tau}} \frac{1}{\sqrt{2Q_1^*(s, s)}} \exp\left(-\frac{1}{2} \frac{(\theta - R_1(s))^2}{Q_1^*(s, s)}\right) \\
 &= -\frac{1}{3!} \delta(t-s) H(u-s) \sqrt{N} m \frac{1}{\tau} e^{-\frac{u-s}{\tau}} \frac{\theta - R_1(s)}{\sqrt{2}(Q_1(s, s))^{\frac{3}{2}}} \exp\left(-\frac{1}{2} \frac{(\theta - R_1(s))^2}{Q_1(s, s)}\right), \\
 &= \frac{1}{3!} \delta(t-s) H(u-s) \sqrt{N} m \frac{1}{\tau} e^{-\frac{u-s}{\tau}} \frac{\partial^2}{\partial R_1(s)} \Phi(R_1(s), Q_1(s, s))
 \end{aligned} \tag{88}$$

so the diagram consisting of $V_{112}^{(RRR)}$ and $\Delta_{11}^{(RR)}$, having symmetry factor 3, gives in the stationary case (for which we know the form of $\Delta_{11}^{(RR)}$):

$$\begin{aligned}
 \left. \frac{\partial \Gamma_{fl}^{(1)}}{\partial R_2} \right|_{\star(1)}^{V_{112} \Delta_{11}} &= 3 \int dt ds V_{112}^{(RRR)}(t, s, u) \Delta_{11}^{(RR)}(t, s) \Big|_{\star(1)} \\
 &= \frac{1}{2} \int dt ds \Delta_{11}^{(RR)}(s-t) \Big|_{\star(1)} \delta(t-s) H(u-s) \sqrt{N} m \frac{1}{\tau} e^{-\frac{u-s}{\tau}} \frac{\partial^2}{\partial R_1^2} \Phi(R_1(s), Q_1(s, s)) \Big|_{\star(1)} \\
 &= \frac{1}{2} \sqrt{N} m \Delta_{11}^{(RR)}(0) \Big|_{\star(1)} \int_{-\infty}^u \frac{ds}{\tau} e^{-\frac{u-s}{\tau}} \frac{\partial^2}{\partial R_1^2} \Phi(R_1(s), Q_1(s, s)) \Big|_{\star(1)}
 \end{aligned} \tag{89}$$

Note that evaluating at $\star(1)$ corresponds to the self consistent one-loop approximation. We have thus derived a term of (87):

$$R_1^{*(1)} = \sqrt{N} m \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \left[\Phi(R_1^*(t'), Q_1^*(t', t')) + \frac{1}{2} \Delta_{11}^{(RR)}(0) \Big|_{\star(1)} \frac{\partial^2}{\partial R_1^2} \Phi(R_1(s), Q_1(s, s)) \Big|_{\star(1)} \right] + \left. \frac{\partial \Gamma_{fl}^{(1)}}{\partial R_2} \right|_{\star(1)}^{V_{122} \Delta_{21}},$$

so taking the time derivative (analogous to (48)), gives the corrected mean field equation:

$$\partial_t R_1^{*(1)} + R_1^{*(1)} = \Phi(R_1^*(t'), Q_1^*(t', t')) + \frac{1}{2} \Delta_{11}^{(RR)}(0) \Big|_{\star(1)} \frac{\partial^2}{\partial R_1^2} \Phi(R_1(s), Q_1(s, s)) \Big|_{\star(1)} + \underbrace{\partial_t \left. \frac{\partial \Gamma_{fl}^{(1)}}{\partial R_2} \right|_{\star(1)}^{V_{122} \Delta_{21}}}_{=0?}. \tag{90}$$

The correction corresponds to what can be expected from linear response theory, see Schuecker et al. (2016).

XIV. GENERALIZATION TO MULTIPLE POPULATIONS

To apply the method used on the toy model to the three population assembly model (assembly, excitatory and inhibitory bath), we note that the synaptic matrix $\mathbf{J}$ is now subdivided into 9 blocks with different statistics:

$$\begin{pmatrix} J_{ee} & J_{ei} & J_{ea} \\ J_{ie} & J_{ii} & J_{ia} \\ J_{ae} & J_{ai} & J_{aa} \end{pmatrix} \quad (91)$$

It is thus useful to introduce population indices $\alpha \in \{e, i, a\}$ for the relevant quantities. We want to execute the disorder average in (30)

$$\langle \rho[\mathbf{h}](\mathbf{J}) \rangle_{\mathbf{J}} := \int \mathcal{D}\hat{\mathbf{h}} \exp(\hat{\mathbf{h}}_{\alpha}^T \mathbf{h}_{\alpha}) \int \mathcal{D}\mathbf{x} \langle \exp(-\hat{\mathbf{h}}_{\alpha}^T \mathbf{J}_{\alpha\beta} \mathbf{x}_{\beta}) \rangle_{\mathbf{J}} \rho[\mathbf{x}|\mathbf{h}]. \quad (92)$$

Proceeding in analogy to (31), the moments of the Bernoulli distributed connectivity become dependent on the receiving and sending population, leading to

$$\varphi_{\mathbf{J}}(\mathbf{y}) \simeq \exp \left(\sum_{\alpha, \beta} \frac{m_{\alpha\beta}}{\sqrt{N}} \sum_{i_{\alpha}, j_{\beta}} y_{ij} + \sum_{\alpha, \beta} \frac{g_{\alpha\beta}^2}{2N} \sum_{i_{\alpha}, j_{\beta}} y_{ij}^2 + O(N^{-\frac{3}{2}}) \right) \quad (93)$$

(note that the terms go still in orders of N because this is the general scaling of the synapses with the whole system size) and the Hubbard-Stratonovich transform

$$\langle \rho[\mathbf{h}](\mathbf{J}) \rangle_{\mathbf{J}, \mathbf{h}}$$

$$\begin{aligned} & \simeq \int \mathcal{D}\hat{\mathbf{h}} \mathcal{D}\mathbf{h} \int \mathcal{D}\mathbf{x} \exp \left(\sum_{\alpha} \sum_{i_{\alpha}} \hat{h}_{i_{\alpha}} h_{i_{\alpha}} - \sum_{\alpha} \sum_{i_{\alpha}} \int \hat{h}_{i_{\alpha}}(t) \underbrace{\sum_{\beta} \frac{m_{\alpha\beta}}{\sqrt{N}} \sum_{j_{\beta}} x_{j_{\beta}}(t) dt}_{=: R_{1, \alpha\beta}(t)} \right. \\ & \quad \left. + \sum_{\alpha} \sum_{i_{\alpha}} \iint \hat{h}_{i_{\alpha}}(t) \hat{h}_{i_{\alpha}}(s) \underbrace{\frac{1}{2} \sum_{\beta} \frac{g_{\alpha\beta}^2}{N} \sum_{j_{\beta}} x_{j_{\beta}}(t) x_{j_{\beta}}(s) dt ds}_{=: Q_{1, \alpha\beta}(t, s)} + O(N^{-\frac{3}{2}}) \right) \rho[x_i|h_i] \\ & = \iint \prod_{\alpha, \beta} (\mathcal{D}R_{1, \alpha\beta} \mathcal{D}R_{2, \alpha\beta}) \iint \prod_{\alpha, \beta} (\mathcal{D}Q_{1, \alpha\beta} \mathcal{D}Q_{2, \alpha\beta}) \prod_{\alpha, \beta} \exp(Q_{1, \alpha\beta}^T R_{1, \alpha\beta} + R_{2, \alpha\beta}^T R_{1, \alpha\beta}) \\ & \quad \times \underbrace{\prod_{\alpha} \int \mathcal{D}\hat{\mathbf{h}}, \mathbf{h}, \mathbf{x} \exp \left(\hat{\mathbf{h}}_{\alpha}^T \left(\mathbf{h}_{\alpha} - \sum_{\beta} R_{1, \alpha\beta} \right) - \sum_{\beta} R_{2, \alpha\beta}^T \frac{m_{\alpha\beta}}{\sqrt{N}} \sum_{j_{\beta}} x_{j_{\beta}} + \frac{1}{2} \hat{\mathbf{h}}_{\alpha}^T \hat{\mathbf{h}}_{\alpha} \sum_{\beta} Q_{1, \alpha\beta} - \sum_{\beta} \frac{g_{\alpha\beta}^2}{N} \sum_{j_{\beta}} x_{j_{\beta}}^T Q_{2, \alpha\beta} x_{j_{\beta}} \right)}_{=: \Omega_{\alpha}(R, Q)} \rho[\mathbf{x}|\mathbf{h}] \quad (94) \end{aligned}$$

We notice that the interacting part of the theory factorizes on the receiving side but is coupled by the connections being sent from other populations, which is physically logical. Note also, that we defined $R_{1, \alpha\beta}$ and $Q_{1, \alpha\beta}$, but these only appear as $\sum_{\beta} R_{1, \alpha\beta} =: R_{1, \alpha}$ and $\sum_{\beta} Q_{1, \alpha\beta} =: Q_{1, \alpha}$, which could also be defined as the Hubbard-Stratonovich fields. We keep the dependence on the sending population however, in order to be able to tear apart the different contributions later.

We now again arrived at an action

$$S(R_1, R_2, Q_1, Q_2) = \sum_{\alpha, \beta} (Q_{2, \alpha\beta}^T Q_{1, \alpha\beta} + R_{2, \alpha\beta}^T R_{1, \alpha\beta}) + \sum_{\alpha} \ln(\Omega_{\alpha}(R, Q)). \quad (95)$$

A. Solution in tree level approximation

Going again to zero order in the effective potential, $\Gamma_0 = -S_0$, the equation of state gives

$$R_{1,\alpha\beta}^*(t) = -\frac{\delta \ln \Omega_\alpha(R, Q)}{\delta R_{2,\alpha\beta}(t)} \quad (96)$$

$$= \frac{1}{\mathcal{Z}(0)} \frac{m_{\alpha\beta}}{\sqrt{N}} \sum_{j_\beta} \langle x_{j_\beta} \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$$

$$R_{2,\alpha\beta}^*(t) = -\frac{\delta \ln \Omega(R, Q)}{\delta R_{1,\alpha\beta}(t)}$$

$$= \frac{1}{\mathcal{Z}(0)} \sum_{i_\alpha} \langle \hat{h}_{i_\alpha} \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$$

$$Q_{1,\alpha\beta}^*(t, s) = -\frac{\delta \ln \Omega(R, Q)}{\delta Q_{2,\alpha\beta}(t, s)} \quad (97)$$

$$= \frac{1}{\mathcal{Z}(0)} \frac{g_{\alpha\beta}^2}{2N} \sum_{j_\beta} \langle x_{j_\beta}(t) x_{j_\beta}(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$$

$$Q_{2,\alpha\beta}^*(t, s) = -\frac{\delta \ln \Omega(R, Q)}{\delta Q_{1,\alpha\beta}(t, s)}$$

$$= -\frac{1}{\mathcal{Z}(0)} \sum_{i_\alpha} \frac{1}{2} \langle \hat{h}_{i_\alpha}(t) \hat{h}_{i_\alpha}(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}.$$

As in the one population case, $R_{2,\alpha\beta}^*(t) = Q_{2,\alpha\beta}^*(t, s) = 0$ because the moments of the response fields vanish.

We now solve (96) explicitly, in analogy to Section XI A.

$$\begin{aligned} R_{1,\alpha\beta}^*(t) &= \frac{1}{\mathcal{Z}(0)} \frac{N_\beta}{\sqrt{N}} m_{\alpha\beta} \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] \frac{1}{N_\beta} \sum_{i_\beta} \int \mathcal{D}x_{i_\beta} x_{i_\beta} \rho[x_{i_\beta}|h_{i_\beta}] \\ &= \frac{1}{\mathcal{Z}(0)} \frac{N_\beta}{\sqrt{N}} m_{\alpha\beta} \frac{1}{N_\beta} \sum_{i_\beta} \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \int \mathcal{D}\mathbf{h} \rho^*[\mathbf{h}] H(h_{i_\beta}(t') - \theta). \end{aligned} \quad (98)$$

where we used that $\int \mathcal{D}x_j x_j \rho[x_j|h_j] = x_i$ for $i \neq j$ due to the normalization and the explicit form of $\rho[x_i|h_i]$ for binary neurons with a Heaviside activation function given by (38). Also the gaussian integration (40) carries over:

$$\begin{aligned} \rho^*[\mathbf{h}] &= \prod_{\alpha} \int_{-\infty}^{\infty} \mathcal{D}\hat{\mathbf{h}} \exp \left(\hat{\mathbf{h}}_{\alpha}^T \left(\mathbf{h}_{\alpha} - \sum_{\gamma} R_{1,\alpha\gamma}^* \right) + \frac{1}{2} \hat{\mathbf{h}}_{\alpha}^T \hat{\mathbf{h}}_{\alpha} \sum_{\gamma} Q_{1,\alpha\gamma}^* \right) \\ &= \prod_{\alpha} \mathcal{F} \left[\exp \left(-\frac{1}{2} x^T \sum_{\gamma} Q_{1,\alpha\gamma}^* x \right) \right] \left(\mathbf{h}_{\alpha} - \sum_{\gamma} R_{1,\alpha\gamma}^* \right) \\ &= \mathcal{Z}_{\hat{\mathbf{h}}}(0) \prod_{\alpha} \exp \left(-\frac{1}{2} \left(\mathbf{h}_{\alpha} - \sum_{\gamma} R_{1,\alpha\gamma}^* \right)^T \sum_{\gamma} Q_{1,\alpha\gamma}^{*-1} \left(\mathbf{h}_{\alpha} - \sum_{\gamma} R_{1,\alpha\gamma}^* \right) \right) \\ &= \prod_{\alpha} \prod_{i_\alpha} \rho^*[h_{i_\alpha}]. \end{aligned} \quad (99)$$

We here use the full probability distribution $\rho^*[\mathbf{h}]$ instead of $\rho^*[\mathbf{h}_\alpha]$ because the Heaviside term in (98) contains $h_{i_\beta} \notin \mathbf{h}_\alpha$ if $\beta \neq \alpha$. Therefore the full distribution is needed to perform the average (the same was already implicitly done for x_{i_β} in (98)). When plugging the form of the distribution into (98), the marginalized version, where all integrals except that over $h_{i_\beta}(t')$ have been executed, reads

$$\rho^*[h_{i_\beta}(t')] = \exp\left(-\frac{1}{2} \frac{(h_{i_\beta}(t') - \sum_\gamma R_{1,\beta\gamma}^*(t'))^2}{\sum_\gamma Q_{1,\beta\gamma}^*(t', t')}\right). \quad (100)$$

Combined in (98) with the Heaviside function this gives again the complementary error function, so that the result is

$$R_{1,\alpha\beta}^*(t) = m_{\alpha\beta} \frac{N_\beta}{\sqrt{N}} \frac{1}{2} \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} \operatorname{erfc}\left(\frac{\theta - \sum_\gamma R_{1,\beta\gamma}^*(t')}{\sqrt{2 \sum_\gamma Q_{1,\beta\gamma}^*(t', t')}}\right). \quad (101)$$

Thus, using the relation $R_{1,\alpha\beta}^*(t) = \frac{N_\beta}{\sqrt{N}} m_{\alpha\beta} \bar{x}_\beta(t)$ and taking the derivative, the system of coupled differential equations describing the dynamics of the networks populations mean activities reads

$$\tau \partial_t \bar{x}_\beta(t) + \bar{x}_\beta(t) = \frac{1}{2} \operatorname{erfc}\left(\frac{\theta - \sum_\gamma R_{1,\beta\gamma}^*(t')}{\sqrt{2 \sum_\gamma Q_{1,\beta\gamma}^*(t', t')}}\right). \quad (102)$$

The equal time input autocorrelation follows as

$$\begin{aligned} Q_{1,\alpha\beta}^*(t, t) &= \frac{1}{Z(0)} \frac{g_{\alpha\beta}^2}{N} \sum_{j_\beta} \left\langle \overbrace{x_{j_\beta}(t) x_{j_\beta}(t)}^{x^2=x} \right\rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \\ &= g_{\alpha\beta}^2 \bar{x}_\beta(t). \end{aligned} \quad (103)$$

This reproduces the result from mean field theory for arbitrary numbers of populations, only with the known difference in Q_1^* , as discussed in Section XI C.

XV. CLUSTERING BY CONNECTION PROBABILITY VS. SYNAPSE STRENGTH

So far, assembly formation was controlled by the parameter a , acting exclusively on the connection probability. However, it is evident that an assembly could also be formed by increasing the synaptic weights between members, or a combination of both methods. To understand in how far this would lead to different results, consider (93). The moments of the connectivity matrix blocks are $m_{\alpha\beta} = J_{\alpha\beta} p_{\alpha\beta}$ and $g_{\alpha\beta}^2 = J_{\alpha\beta}^2 p_{\alpha\beta} (1 - p_{\alpha\beta})$, therefore increasing the connection probability between two populations $p_{\alpha\beta}$ by factor a or doing the same with $J_{\alpha\beta}$ instead would make no difference in the value of the mean input strength $m_{\alpha\beta}$. For the input variance however,

$$\begin{aligned} g_{\alpha\beta, \text{conn.}}^2 &= J_{\alpha\beta}^2 a p_{\alpha\beta} (1 - a p_{\alpha\beta}) \\ &= a^2 J_{\alpha\beta}^2 p_{\alpha\beta} (1 - p_{\alpha\beta}) - J_{\alpha\beta}^2 p_{\alpha\beta} a (a - 1) \\ &= g_{\alpha\beta, \text{syn.}}^2 - J_{\alpha\beta}^2 p_{\alpha\beta} a (a - 1), \end{aligned} \quad (104)$$

which holds for both fixed and distributed indegree. Thus, when replacing connectivity clustering $p \rightarrow ap$ by synapse weight clustering $J \rightarrow aJ$, the mean input weight from that population is conserved but its variance increases significantly for large a . This is of course logical, since we have few strong synapses instead of many weak ones. However, for the case of a close to 1 the difference is small, due to the factor $a - 1$ in the correction term.

What are then the effects on the assembly dynamics? The way we introduced connection clustering in (4) was by $p_{aa} \rightarrow ap_{aa}$ and $p_{ae} \rightarrow \frac{1-a}{1-\frac{N_a}{N}} p_{ae} = a_{ae} p_{ae}$, where the second adjustment keeps the total excitatory input to any neuron

constant. Therefore, synapse weight clustering just amounts to doing $J_{aa} \rightarrow aJ_{aa}$ and $J_{ae} \rightarrow a_{ae}J_{ae}$ instead, so all calculations can be adopted by simply changing g_{aa}^2 and g_{ae}^2 according to (104). We especially obtain the same mean field equation (102) only with changed $Q_{1,aa}^*$ and $Q_{1,ae}^*$. For large N , or more precisely small $f = \frac{aN_a}{N} \sim \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$ (see (20)), the decrease of $Q_{1,ae}^*$ is negligible, as $a_{ae} \approx 1$. Therefore, we find

$$Q_{1,syn}^*(t, t) \approx Q_{1,conn.}^*(t, t) + J_0^2 pa(a-1)\bar{x}_A(1-\bar{x}_A). \quad (105)$$

The effects of this additional term are: First and trivially, there are larger fluctuations. Second, because $Q_{1,syn}^*$ is no longer linear in a , the ground state activity is not constant, but increases for bigger a due to the growing denominator of the error function complementary, constituting a stretch in x-direction. Third, this stretch also effects a smaller derivative of the gain function for the considerations of Section VIII A, resulting in more shallow minima and a smaller bistable regime. In a simulation, all these effects can be observed, see Figure 11.

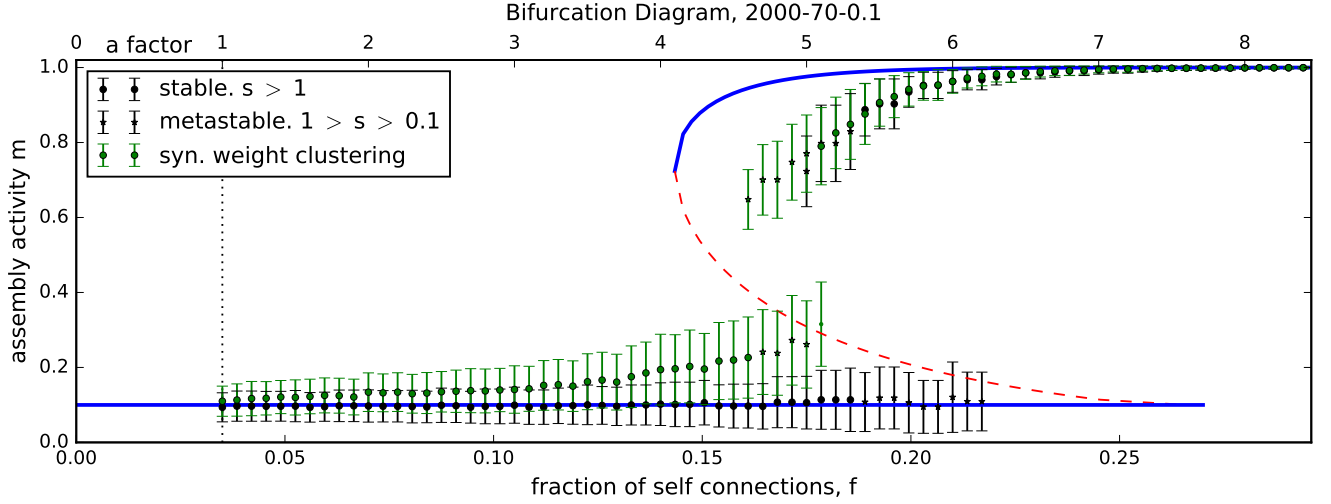


Figure 11. Comparison of clustering by synaptic weight (green) and connection probability (black). Blue curves are the MF-prediction for connection clustering. Error bars indicate the standard deviation of the spontaneous mean activity in the given minimum. The values in the bistable regime of the synapse weight clustered assembly are likely inaccurate, as the extraction procedure is not completely stable against very large fluctuations (see Section XXII).

Note that if very large a factors are required, eg. for very small assemblies, connection clustering is limited by full connectivity, and synapse strength clustering would need to rely on very strong synapses, both of which may become biologically implausible. However it would be expected that biological assemblies display increased connection probability as well as synapse strengths. The interesting, albeit of course logically clear, effect of this is that the mean input strength

$$m_{aa} = a_{syn}.J_{aa}a_{conn}.p_{aa}$$

scales with an effective a factor that is given by the product of the two a_{syn} . and a_{conn} . Therefore it is not only feasible to reach very large effective a factors in biologically plausible settings, making small assemblies displaying bistability possible, but also medium sized assemblies that do not show very obvious morphological statistics, as doubled synapse strength combined with doubled connection probability can easily be enough to reach the bistable regime.

XVI. OTHER NEURON MODELS

While the comprehensive treatment of other neuron models is beyond the scope of this work, we give a short exposition of possible future directions.

The way in which the neuron model enters into the formulation is via the conditional probability functional $\rho[x_i|h_i]$. For binary neurons with asynchronous update this took the form

$$\begin{aligned}\rho[x_i|h_i](t) &= x_i(t) p[h_i](t) + (1 - x_i(t)) (1 - p[h_i](t)) \\ &= (2x_i(t) - 1) p[h_i](t) + 1 - x_i(t)\end{aligned}\tag{106}$$

$$p[h_i](t) = \int_{-\infty}^t \frac{dt'}{\tau} e^{-\frac{t-t'}{\tau}} H(h_i(t') - \theta).$$

For rate neurons, consider the Langevin equation $(\frac{d}{dt} + 1) y_i = \sum_{j=1}^N J_{ij} \phi(y_j) + \xi_i(t)$ with Gaussian white noise $\langle \xi_i(t) \xi_j(s) \rangle = D \delta_{ij} \delta(t - s)$, which corresponds to the SCS model (Sompolinsky et al. 1988) with additional noise. Note that here, y_i is not the activity of neuron i . Instead, we define $x_i = \phi(y_i)$ so that $h_i = \sum_{j=1}^N J_{ij} \phi(y_j) = \sum_{j=1}^N J_{ij} x_j$ as used in the derivation in Section X. Also, it is here simpler to not construct $\rho[x_i|h_i]$, but its characteristic function $Z[j_i|h_i] = \int dx_i \rho[x_i|h_i] \exp(-j_i x_i)$, which amounts to the same for the averages $\langle \dots \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}$, since for $R_2^* = Q_2^* = 0$ the action (33) does not depend on x_i other than through $\rho[x_i|h_i]$. To do so, consider that

$$\begin{aligned}Z[j_i|h_i] &= \int \mathcal{D}x_i \rho[x_i|h_i] \exp(-j_i x_i) \\ &= \int \mathcal{D}x_i \int \mathcal{D}y_i \delta[x_i - \phi(y_i)] \rho[y_i|h_i] \exp(-j_i x_i)\end{aligned}\tag{107}$$

$$\begin{aligned}&= \int \mathcal{D}y_i \rho[y_i|h_i] \exp(-j_i \phi(y_i)) \\ &= \int \mathcal{D}y_i \langle \delta[(\partial_t + 1)y_i - h_i - \xi_i] \rangle_{\xi_i} \exp(-j_i \phi(y_i)) \\ &= \int \mathcal{D}y_i \mathcal{D}\hat{y}_i \exp\left(\hat{y}_i^T (\partial_t + 1)y_i - \hat{y}_i^T h_i + \frac{D}{2} \hat{y}_i^T \hat{y}_i - j_i \phi(y_i)\right),\end{aligned}\tag{108}$$

where we used the characteristic function of the noise $\langle \exp(-\hat{y}_i \xi_i) \rangle_{\xi_i} = \exp\left(\frac{D}{2} \hat{y}_i^T \hat{y}_i\right)$. If we now want to calculate, for example (35), we find that $\mathbf{h}$ and $\hat{\mathbf{h}}$ can be integrated out:

$$\begin{aligned}R_1^*(t) &= \frac{1}{Z(0)} \frac{m}{\sqrt{N}} \sum_i \langle x_i \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \\ &= \frac{1}{Z(0)} \frac{m}{\sqrt{N}} \sum_i \int \mathcal{D}\mathbf{y} \mathcal{D}\hat{\mathbf{y}} \int \mathcal{D}\hat{\mathbf{h}} \mathcal{D}\mathbf{h} \exp\left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1^*) + \frac{1}{2} \hat{\mathbf{h}}^T Q_1^* \hat{\mathbf{h}} - \hat{\mathbf{y}}^T \mathbf{h}\right) \partial_{j_i} \exp\left(\hat{\mathbf{y}}^T (\partial_t + 1)\mathbf{y} + \frac{D}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} - j \phi(\mathbf{y})\right) \\ &= \frac{1}{Z(0)} \frac{m}{\sqrt{N}} \sum_i \int \mathcal{D}\mathbf{y} \mathcal{D}\hat{\mathbf{y}} \partial_{j_i} \exp\left(+\hat{\mathbf{y}}^T (\partial_t + 1)\mathbf{y} + \frac{1}{2} (\hat{\mathbf{y}} - R_1^*)^T Q_1^* (\hat{\mathbf{y}} - R_1^*) + \frac{D}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} - j \phi(\mathbf{y})\right).\end{aligned}\tag{109}$$

This expression is also the one obtained in the MSRDJ formalism, see eqs. (30) and (33) in Schuecker et al. (2016) (only that we use $R_1^* \neq 0$). The equivalence between the two approaches can also be shown generally by looking at the full action and noting that integrating the delta constraint in (107) just replaces $x_i \rightarrow \phi(y_i)$ also in the R_2 and

Q_2 terms:

$$\Omega(R_1, R_2, Q_1, Q_2)$$

$$\begin{aligned}
&= \int \mathcal{D}\hat{\mathbf{h}}, \mathbf{h}, \mathbf{y}, \mathbf{x} \exp \left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1) + \frac{1}{2} \hat{\mathbf{h}}^T Q_1 \hat{\mathbf{h}} - \frac{m}{\sqrt{N}} R_2^T \mathbf{x} - \frac{g^2}{N} \mathbf{x}^T Q_2 \mathbf{x} \right) \delta(\mathbf{x} - \phi(\mathbf{y})) \rho[\mathbf{y}|\mathbf{h}] \exp(-\mathbf{j}\mathbf{x}) \\
&= \int \mathcal{D}\hat{\mathbf{h}}, \mathbf{h}, \hat{\mathbf{y}}, \mathbf{y} \exp \left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1) + \frac{1}{2} \hat{\mathbf{h}}^T Q_1 \hat{\mathbf{h}} - \hat{\mathbf{y}}^T \mathbf{h} - \frac{m}{\sqrt{N}} R_2^T \phi(\mathbf{y}) - \frac{g^2}{N} \phi(\mathbf{y})^T Q_2 \phi(\mathbf{y}) + \hat{\mathbf{y}}^T (\partial_t + 1) \mathbf{y} + \frac{D}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} - \mathbf{j} \phi(\mathbf{y}) \right) \\
&= \int \mathcal{D}\hat{\mathbf{y}}, \mathbf{y} \exp \left(-\frac{m}{\sqrt{N}} R_2^T \phi(\mathbf{y}) - \frac{g^2}{N} \phi(\mathbf{y})^T Q_2 \phi(\mathbf{y}) + \hat{\mathbf{y}}^T (\partial_t + 1) \mathbf{y} + \frac{1}{2} (\hat{\mathbf{y}} - R_1^*)^T Q_1^* (\hat{\mathbf{y}} - R_1^*) + \frac{D}{2} \hat{\mathbf{y}}^T \hat{\mathbf{y}} - \mathbf{j} \phi(\mathbf{y}) \right).
\end{aligned} \tag{110}$$

$$(111)$$

Thus, by integrating out $\mathbf{h}$ and $\hat{\mathbf{h}}$, we obtained the disorder averaged MSRDJ action for the SCS model, see eq. (30) in Schuecker et al. (2016) (only that we use $R_1, R_2 \neq 0$).

A very interesting question is the application to the spiking LIF model. An initial idea to describe the output spike train could be

$$x_i(t) = \begin{cases} 1 & \text{for } \int_{-\infty}^t e^{\frac{t'-t}{\tau}} (h_i(t') - \theta \delta(1 - x_i(t'))) dt' = \theta \\ 0 & \text{else} \end{cases}, \tag{112}$$

so that

$$\begin{aligned}
\rho[x_i|h_i](t) &= \int_0^\epsilon d\epsilon' x_i(t) \delta \left[\epsilon' + \theta - \int_{-\infty}^{t_k} dt' e^{\frac{t'-t}{\tau}} (h_i(t') - \theta \delta(1 - x_i(t'))) \right] \\
&\quad + \int_0^\epsilon d\epsilon' (1 - x_i(t)) \left(1 - \delta \left[\epsilon' + \theta - \int_{-\infty}^{t_k} dt' e^{\frac{t'-t}{\tau}} (h_i(t') - \theta \delta(1 - x_i(t'))) \right] \right) \\
&= \int_0^\epsilon d\epsilon' 2x_i(t) \delta \left[\epsilon' + \theta - \int_{-\infty}^{t_k} dt' e^{\frac{t'-t}{\tau}} (h_i(t') - \theta \delta(1 - x_i(t'))) \right] + (1 - x_i(t)).
\end{aligned} \tag{113}$$

The exploration of this, or other ideas, remains to further work.

Part IV

Discussion

The mathematical description of recurrent networks beyond large, homogeneous populations is a field of active research. We have here looked at the limits of available standard tools for a model that features a mesoscopic population size, phase transitions and large fluctuation dynamics, after which we explored a novel functional formulation that could in theory not only allow the use of advanced tools from statistical physics, but also the direct comparison of different neuron models in a single framework. The parameter dependence of dynamical assembly properties were investigated, especially assembly size and clustering strength. Also pairwise correlations and fluctuation driven state transitions, effecting the breakdown of MFT were considered. The path integral formulation was developed for binary neurons, up to expressions for the propagators allowing the calculation of 1-loop corrections. In tree level approximation, the relation to standard mean field theory was shown. For the SCS model of rate neurons, the formulation is equivalent to the MSRDJ form.

XVII. RELATION TO OTHER PATH INTEGRAL FORMALISMS

A. Overview of approaches in the literature

Some path integral approaches to stochastic disordered systems have recently been reviewed in Hertz et al. (2017), Chow & Buice (2015) and Schuecker et al. (2016). A short and possibly subjective overview focusing on the application to neuronal networks is as follows: For rate neurons, whose dynamics are described by stochastic differential equations, the MSRDJ (De Dominicis & Peliti 1978, Janssen 1976, Martin et al. 1973) and Onsager-Machlup (Onsager & Machlup 1953) formulations are available. The main conceptual steps of the MSRDJ formalism are: The path of the system is parameterized by delta constraints enforcing the SDE at all time points. Writing the delta distribution in its Fourier representation introduces a response field acting as a source term, and the components of the SDE now appear in the exponent, yielding an action. In the case of gaussian white noise, the unphysical response field can be integrated out, resulting in the Onsager-Machlup form of the action. For binary variables, symmetric couplings are most commonly used, which allows the direct definition of an energy for the configuration state, the familiar Ising spin Hamiltonian, see Coolen (2000a,b). Allowing asymmetric coupling results in the loss of detailed balance, so that non-equilibrium techniques are necessary. Independent of the update type and connection structure, if a network of coupled binary variables is Markovian and the transition rates between states are known, it can be exactly described by a master equation. Then the Doi-Peliti formalism (Doi 1976, Peliti 1985) allows a second quantization procedure, by introducing creation and annihilation operators which bring the master eq. into a form corresponding to a Schrödinger eq. with imaginary time, and a coherent state representation then yields the field theory. The action allows to read off an effective Langevin equation, but the obtained fields are not directly corresponding to physical quantities and curious effects such as 'imaginary' noise appear (Howard & Täuber 1997). This route was taken by Buice & Cowan (2007). Alternatively to the field theoretic formulation, a moment hierarchy can be constructed (Buice et al. 2010). Lefevre & Biroli (2007) started from the probability distribution of the 'jump process' $J(t_i) = n(t_{i+1}) - n(t_i)$ that connects two time points of a Markov process, used the MSRDJ approach on this discretized SDE, and showed that the resulting action is related to that of the Doi-Peliti form by a Hopf-Cole transformation of the fields. While there is no path integral formulation for leaky-integrate-and-fire neurons as yet, Ocker et al. (2017) recently formulated a path integral and diagrammatic expansion for Hawkes processes as an effective model for spiking networks.

B. Relationship to the presented approach

Our approach has methodological similarity to the one of Lefevre & Biroli (2007) in that we start from a probability distribution and then perform a MSRDJ-like derivation on the closing equation, represented as a delta constraint. This, and also the use of the input field $\mathbf{h}$ instead of the activities $\mathbf{x}$, is also treated similarly in the section on Ising spins of Hertz et al. (2017). However, we use three further ingredients: 1. use of generic properties of neuronal networks, 2. no specification of the neuron or connectivity model, 3. a disorder average over realizations of the connectivity. Using the generic properties of a neuronal network, that is large divergence of the neuron output, large convergence on the input side, small correlations in the balanced state, and a neuron model linking input to output, allows us to formally derive a 'general' path integral form for neuronal networks. Therefore the connectivity and network model only need to be added to gain concrete expressions, while the abstract structure is the same. The disorder average, also discussed by Coolen (2000b) and Schuecker et al. (2016), allows the natural access to a formulation in the population variables R and Q . It would be of interest to investigate the correspondence of our method to known formulations for specific neuron models, e.g. for the binary neurons of (Buice & Cowan 2007) (they use the Doi-Peliti formulation and a slightly different binary model) or Hawkes processes (Ocker et al. 2017), but this remains to further work. The relation to MFT was discussed in Section XI C. For the SCS model (Sompolinsky et al. 1988) of rate neurons, we have shown how to obtain the MSRDJ action from our formulation. A major difference to other formalisms is that we do not construct an explicit action in that not the full multivariate probability distribution is needed. Instead, increasingly complicated conditional probabilities need to be added only when higher cumulants are calculated, so that we needed only expressions for $\rho[x_i|h_i](t)$ and $\rho[x_i(s)|x_i(t)]$ in our derivations in Part III. This allows the extension to neuron models for which no action is known, such as the LIF neuron, the attempt of which would be an interesting next step.

XVIII. STRENGTHS AND WEAKNESSES OF THE NEW FORMALISM

Because the presented formalism is theoretically suited to describe a large class of models in the same framework, it may be useful to gain insight into the relations between networks of different neuron models and connectivity structures. Propagators, vertices and effective equations of motion could be directly compared, allowing to better understand which effects can or cannot be modeled by the various neurons. No explicit action needs to be constructed, so that more realistic spiking models with a membrane potential may be describable. This however has the downside, that a closed form for the possibly very complicated conditional probability $\rho[x_i(s)|x_i(t)]$ is needed in order to calculate $Q_1^*(t, s)$, and successively higher order conditional probabilities if cumulants beyond the second are calculated. The disorder average on the one hand allows the natural introduction of the population variables R and Q by the Hubbard-Stratonovic transform and the use of a large class of connectivity statistics, but also introduces higher order interaction terms which complicate the theory. It should also be noted the disorder average only provides a useful description of a single realization of the network, if the system has the self-averaging property. This can be assessed by a replica calculation, for which the path integral is the natural language. In this work we only explored a few basic aspects of the formalism and concentrated on binary neurons, therefore many questions remain open, such as in how far analytic calculations remain possible. Also the use of symmetries, renormalization group methods, or a supersymmetric formulation as done for SDE's (Hertz et al. 2017), could be attempted. Because of the ubiquitous time integrals in the formulation, the expressions can be involved, so if one is interested in stationary properties only, networks that have the Markov property might be described more easily using another method, as only equal-time correlators can appear. Concerning the range of connectivity models that can be described, note that for the disorder average, we used only the first two moments of a populations connectivity statistics, arguing that higher moments are suppressed for large populations due to the synapse weights scaling as $\frac{1}{\sqrt{N}}$. An interesting question is also, whether one could go beyond point neuron models and include dendritic branches, as then the probability functional $\rho[x_i|h_i]$ may not exist. One idea how to do handle the nonlinear summation of inputs would be to introduce an 'input neuron' for each dendritic branch h_{ik} , for which inputs are still summed linearly, and then collect these in the input to output mapping $\rho[x_i|\sum_k h_{ik}]$. Then however, J_{ij} would no longer be a square matrix, and due to the smaller number of inputs per dendritic branch, the assumption on the connectivity statistics is more easily violated. Also it is to be expected, that for increasingly complicated neuron models $\rho[x_i|h_i]$ needs to be determined numerically or by simulation.

XIX. GENERAL DISCUSSION OF THE ASSEMBLY MODEL

The generalizability of the model to biologically realistic assemblies can be considered as follows. The use of binary neurons with Heaviside activation function and asynchronous updates combined with dominant inhibition allows to qualitatively model a spiking network in the balanced state for low activities. Since the upstate activity comes close to 1, an assembly of spiking neurons would behave slightly differently. While the spike frequency of LIF neurons is limited by the refractory period, the assembly upstate will stabilize below this maximal frequency since input during the refractory period is lost. This would not qualitatively alter the dynamics, only replace the hard activity cap in the present model by a soft ceiling. This could be modeled by using binary neurons with a sigmoidal activation function. Mechanisms that are not modeled and could give rise to qualitatively different effects are: 1. Nonlinear summation and propagation of inputs by the dendrite. One effect expected here could be an increased coincidence detection due to clustering of assembly inputs on the same dendritic branch (Holtmaat & Caroni 2016). 2. Further excitatory and inhibitory neuron populations with specific connection patterns. Note that as long as these populations are not strongly influenced by the assembly, they may be treated as external input. 3. Spike frequency adaptation and synaptic short term plasticity. Adaptation, if strong enough, results in the destabilization of upstates after some time as the firing rates of the assembly members gradually decrease. This can be pictured as a flattening movement of the sigmoidal shape in Figure 3, see the equivalent Figure 7 of Setareh et al. (2017), which was obtained by simulating a clustered network using spiking neurons which include adaptation.

Concerning the generalizability of the proof in Section VII B that the upstate emerges at significantly higher activities than the ground state, we expect that it is not effected by points 1. and 3. However, Latham & Nirenberg (2004) show that an assembly significantly coupled to local inhibition can be dynamically stabilized to a relatively low firing rate upstate, as also commented on by Miller (2016). In this case the gap can be small, but still exists as long as the gain function has the sigmoidal shape.

XX. NEUROSCIENTIFIC IMPLICATIONS FOR EXPERIMENT AND MODELING OF CELL ASSEMBLIES

Of most interest for experiment and modeling would be predictions on: 1. Assembly sizes, 2. connection probabilities and synapse strengths, 3. a relation between assembly size and connection prob. / synapse strength, 4. the relation of activity patterns and structural network parameters and 5. understanding of assembly dynamics.

In Section VII we have seen that, to mean field approximation in our model, if assembly size and clustering strength are scaled antipropotionally as $a \sim \frac{1}{N_a}$, the system properties are kept invariant. Therefore, in want of clear measurements of one or the other, predictions on 1. or 2. are not possible from this approach. (While for synapse clustering the ground state activity changes as well, this should be difficult to detect if one assumes that the cortical neurons of interest can each be part of one or several assemblies). We can confirm that the general range of a few dozen to a few hundred neurons per assembly could be consistent with biologically realistic parameters.

Concerning 3., the question is raised in which activity regime the assemblies are expected to be. While this may depend on the brain structure or cortical layer, expected are either the bistable regime, the first critical point or a subcritical configuration. Since external input to the assembly or also the whole network can manipulate the regime, such that increased input may push a subcritical assembly to bistability, and the cortex operates under changing global or local state modifications such as attention, neurotransmitters like dopamine, or even sleep, assemblies could be expected to operate in a slightly or largely subcritical state, that is only tuned towards criticality or bistability by the modulatory influences. Therefore, to give more than rough guidelines on 3., a more detailed and realistic model is necessary, which could be based on the present work.

Concerning 4. and 5., insight into the activity dynamics of assemblies may be useful to experiment for their detection from observed network activity (for examples, see Buzsáki (2010), Miller (2016)), or by active probing of the network dynamics by stimuli (Beltramo et al. 2013, Setareh et al. 2017). The present work could be extended to study the assembly response to fluctuating input or short stimuli. To capture dynamics in the bistable regime, fluctuation corrections to the effective potential are necessary due to the mesoscopic assembly size. While quantitative predictions would again need a more realistic model, the qualitative behavior of a simplified (or even analytically treatable) model is useful to gain an understanding of complicate phenomena. For example, an assembly response function could be used to study the interactions of multiple assemblies, a topic that is of paramount importance to sensory processing, memory, or other functions, if the brain relies on assemblies for computation. Studies in this direction have until now relied on computer simulations of networks on the neuron level, e.g. Litwin-Kumar & Doiron (2012), Mazzucato et al. (2015). Furthermore, the spontaneous state transitions (Section VIII) may be a crucial ingredient to the resting state (see Luczak et al. (2009)), involved in the random aspects of mind wandering or dreaming, or a source of the neuronal avalanches widely reported by *in vitro* studies (Friedman et al. 2012, Hernandez-Urbina & Michael Herrmann 2016). While these where future directions, we think that the basic understanding of assembly dynamics and parameter dependencies developed in this work is valuable for modelers of attractor networks, since it allows more systematic interpretation of the dynamics observed in simulations, and may inform theoretical deliberations.

XXI. ACKNOWLEDGMENTS

The author is very much indebted to Tobias Kühn and Moritz Helias, for their excellent and very friendly supervision. He would like to thank the people of the INM-6 institute, who constitute an interesting and valuable environment. Furthermore, the author is thankful to Phil McHardy, for sparking his interest in neuroscience, his flatmates and family for many reasons, and Mia.

Part V

Appendix

XXII. SIMULATION AND DATA ANALYSIS

All simulations were done using NEST v2.11.4 (Bos et al. 2015). The assembly model was implemented as a python script according to the description in Section V A, using the `mc_culloch_pitts_neuron()` model, which is equivalent to the binary neuron with asynchronous update. Unlike the mathematical model, the synapses in the simulation have a delay of 0.1 ms, which is not expected to lead to any notable differences. All numerical calculations, data analysis and plotting was carried out by custom python scripts utilizing numpy, scipy and matplotlib. The code is freely available upon request. Most implementations are very straight forward and annotated by comments, but we will now shortly describe some nontrivial steps or design decisions. The main structure is very simple: `assembly_arraysim.py`, `attractor_datamining_functions.py` and `attractor_theory_functions.py` are resource files, containing functions which are called by the various other scripts. Data is stored to and read from `../data/` and plots saved to `../figures/`. The naming scheme for files is usually of the form 'description'_'suffix'.dat, where 'description' gives the content or purpose and 'suffix' the important parameters in the form $N_N_A_a_{factor}_m_E$.

A. Simulation

Since the mean bath activity m_E is usually given to the simulation routine as a parameter, the threshold that results in this mean activity needs to be computed. This is done numerically according to the mean field formula (12). As this is an approximation, the actual simulated mean activity will deviate from the intended one. The deviation was quantified in Figure 12, showing that it is small for $N > 2000 \wedge m_E > 0.05$. The death of activity for very small values of m_E is hypothesized to result from the gaussian approximation of the input strength: The actual input has a finite cutoff given by its binomial distribution, but this tail is overestimated by the gaussian approximations which ignores this cutoff, resulting in a pathological estimation of the threshold.

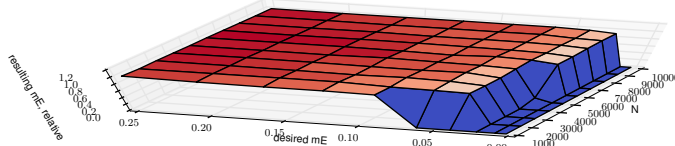


Figure 12. Simulated m_E relative to desired m_E for different N and m_E

B. Bifurcation diagram

In order to generate the bifurcation diagram, the function `stationarity_scanner()` is called in the simulations of `sim_bifurcation_diagram.py`. This scanner samples the activity space of the assembly by repeatedly driving it to various mean activities and then letting it relax again. This data is then used by `sim_bifurcation_diagram.py` as follows: The algorithm tries to skip transients by waiting 80 ms (arbitrary, determined by trial and error) after the net is released at some activity, and then looks whether the activity is above or below 0.5 and for how long. This list is then used to determine whether the up- or downstate exists and how stable it is. Of course, minima that are very small compared to the fluctuations are likely to be not detected, which is, however, because the system does not stay in these minima.

C. Extraction of single neuron correlations from population sums

assume it was measured for each time step of a simulation the sum of the activities of one half of a population of binary neurons, and the sum of the other half, also. Then

$$\begin{aligned} \left\langle \left\langle \sum_{k=0}^{\frac{N}{2}} n_k \sum_{l=\frac{N}{2}+1}^N n_l \right\rangle_{sim} \right\rangle_{trial} &= \sum_{k=0}^{\frac{N}{2}} \sum_{l=\frac{N}{2}+1}^N \langle \langle n_k n_l \rangle_{sim} \rangle_{trial} \\ &= \frac{N^2}{4} \cdot c \end{aligned} \quad (114)$$

where c is the average cross correlation of a single neuron pair.

Of course, the trial average needs not always be done by several simulations; instead, if the simulation is long enough compared to its intrinsic timescales, the sim average alone will approach the combination of sim and trial averaging. This is because the network state does not depend on its history of states for more than its average memory lifetime. As one simulation means only one realization of the connectivity however, it is a good estimator of the trial average only for sufficiently large numbers of connections.

To obtain the autocorrelation in the same way (now using two times the same half of the population), the cross-correlation needs to be known and subtracted:

$$\begin{aligned} \left\langle \left\langle \sum_{k=0}^{\frac{N}{2}} n_k \sum_{l=0}^{\frac{N}{2}} n_l \right\rangle_{sim} \right\rangle_{trial} &= \sum_{k,l=0}^{\frac{N}{2}} \langle \langle n_k n_l \rangle_{sim} \rangle_{trial} \\ &= \frac{N}{2} \cdot a + \frac{N}{2} \left(\frac{N}{2} - 1 \right) \cdot c \end{aligned}$$

However, this may introduce an increased statistical error, as the (large) N^2 contribution is subtracted, leaving a small quantity with the error of a larger one, thus heightening the relative error.

Luckily, the autocorrelation of a binary variable can be determined more easily, as due to $n^2 = n$

$$\langle n^2 \rangle - \langle n \rangle^2 = m - m^2 \quad (115)$$

XXIII. SUPPLEMENTARY CALCULATIONS MEAN FIELD TREATMENT

A. Numerical solution of equations for linearized correlations

starting from (6.4.1) in tn_corr (some steps after (33) of Grytskyy et al. (2013)):

$$2c_{\alpha\beta} = \sum_{\gamma} (\omega_{\alpha\gamma} c_{\gamma\beta} + \omega_{\beta\gamma} c_{\gamma\alpha}) + \omega_{\alpha\beta} \frac{a_{\beta}}{N_{\beta}} + \omega_{\beta\alpha} \frac{a_{\alpha}}{N_{\alpha}} \quad (116)$$

with $\omega_{\alpha\beta} = S(\mu_{\alpha}, \sigma_{\alpha}) J_{\alpha\beta} K_{\alpha\beta}$. The autocovariance a_k can be expressed as $m_k - m_k^2$, due to the neurons being binary variables (115). The susceptibility is given by

$$S(\mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\theta - \mu_k)^2}{2\sigma_k^2}} \quad (117)$$

Where $\mu_k = \sum_i K_{ki} J_{ki} m_i$ and $\sigma_k^2 = \sum_i K_{ki} J_{ki}^2 a_i$ if correlations are assumed to be small and can be neglected. The connections and weights in our case are, using $\alpha, \beta \in \{E, I, A\}$, see (4) and (5):

$$K_{\alpha\beta} = \begin{pmatrix} (N - N_a)p & Np & N_a p \\ (N - N_a)p & Np & N_a p \\ Np \cdot (1 - f) & Np & Np \cdot f \end{pmatrix} \quad (118)$$

and

$$J_{\alpha\beta} = \begin{pmatrix} J_e & J_i & J_e \\ J_e & J_i & J_e \\ J_e & J_i & J_e \end{pmatrix} \quad (119)$$

so that, using $S_k = S(\mu_k, \sigma_k)$

$$\omega_{\alpha\beta} = p \cdot \begin{pmatrix} S_e(N - N_a)J_e & S_e N J_i & S_e N_a J_e \\ S_i(N - N_a)J_e & S_i N J_i & S_i N_a J_e \\ S_a N J_e(1 - f) & S_a N J_i & S_a N J_e f \end{pmatrix} \quad (120)$$

Looking at (116) we see that $c_{\alpha\beta}$ is symmetric in its indices, as should be the case for a covariance matrix. Thus we have a system of 6 linear equations:

$$\mathbb{M} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ea} \\ c_{ii} \\ c_{ia} \\ c_{aa} \end{pmatrix} = \begin{pmatrix} -2\omega_{ee} \frac{a_e}{N_e} \\ -\omega_{ei} \frac{a_i}{N_i} - \omega_{ie} \frac{a_e}{N_e} \\ -\omega_{ea} \frac{a_a}{N_a} - \omega_{ae} \frac{a_e}{N_e} \\ -2\omega_{ii} \frac{a_i}{N_i} \\ -\omega_{ia} \frac{a_a}{N_a} - \omega_{ai} \frac{a_i}{N_i} \\ -2\omega_{aa} \frac{a_a}{N_a} \end{pmatrix}, \quad (121)$$

where

$$\mathbb{M} = \begin{pmatrix} 2\omega_{ee} - 2 & 2\omega_{ei} & 2\omega_{ea} & 0 & 0 & 0 \\ \omega_{ie} & \omega_{ee} + \omega_{ii} - 2 & \omega_{ia} & \omega_{ei} & \omega_{ea} & 0 \\ \omega_{ae} & \omega_{ai} & \omega_{ee} + \omega_{aa} - 2 & 0 & \omega_{ei} & \omega_{ea} \\ 0 & 2\omega_{ie} & 0 & 2\omega_{ii} - 2 & 2\omega_{ia} & 0 \\ 0 & \omega_{ae} & \omega_{ie} & \omega_{ai} & \omega_{ii} + \omega_{aa} - 2 & \omega_{ia} \\ 0 & 0 & 2\omega_{ae} & 0 & 2\omega_{ai} & 2\omega_{aa} - 2 \end{pmatrix}. \quad (122)$$

This equation can be solved analytically, but more conveniently numerically for the specific case. After solving (121) numerically using Numpy, the resulting $\vec{c}$ was used to update the population fluctuations as $\sigma_\alpha^2 = \sum_\beta K_{\alpha\beta} J_{\alpha\beta}^2 a_\beta + \sum c...$. Then the new stationary point of the MF eq. was used to update $S_\alpha(\mu_\alpha, \sigma_\alpha)$, leading to the next iteration, until convergence.

B. Analytical solution by neglecting feedback of assembly to bath

The justified approximation that the assembly has no significant influence on the bath activity can be expressed by setting ω_{ae} and ω_{ia} to zero. Then the system of equations should decouple into the homogeneous EI-network, and the purely excitatory assembly network receiving feedforward input by the former.

We start by setting ω_{ae} and ω_{ia} to zero:

$$\mathbb{M} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ea} \\ c_{ii} \\ c_{ia} \\ c_{aa} \end{pmatrix} = \begin{pmatrix} -2\omega_{ee} \frac{a_e}{N_e} \\ -\omega_{ei} \frac{a_i}{N_i} - \omega_{ie} \frac{a_e}{N_e} \\ -\omega_{ae} \frac{a_e}{N_e} \\ -2\omega_{ii} \frac{a_i}{N_i} \\ -\omega_{ai} \frac{a_i}{N_i} \\ -2\omega_{aa} \frac{a_a}{N_a} \end{pmatrix}, \quad (123)$$

where

$$\mathbb{M} = \begin{pmatrix} 2\omega_{ee} - 2 & 2\omega_{ei} & 0 & 0 & 0 & 0 \\ \omega_{ie} & \omega_{ee} + \omega_{ii} - 2 & 0 & \omega_{ei} & 0 & 0 \\ \omega_{ae} & \omega_{ai} & \omega_{ee} + \omega_{aa} - 2 & 0 & \omega_{ei} & 0 \\ 0 & 2\omega_{ie} & 0 & 2\omega_{ii} - 2 & 0 & 0 \\ 0 & \omega_{ae} & \omega_{ie} & \omega_{ai} & \omega_{ii} + \omega_{aa} - 2 & 0 \\ 0 & 0 & 2\omega_{ae} & 0 & 2\omega_{ai} & 2\omega_{aa} - 2 \end{pmatrix}. \quad (124)$$

the order of the covariances in the vector is a bit unfortunate in this situation. By exchanging the places of c_{ea} and c_{ii} the decoupled structure of $\mathbb{M}$ gets visible

$$\mathbb{M} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ii} \\ c_{ea} \\ c_{ia} \\ c_{aa} \end{pmatrix} = \begin{pmatrix} -2\omega_{ee} \frac{a_e}{N_e} \\ -\omega_{ei} \frac{a_i}{N_i} - \omega_{ie} \frac{a_e}{N_e} \\ -2\omega_{ii} \frac{a_i}{N_i} \\ -\omega_{ae} \frac{a_e}{N_e} \\ -\omega_{ai} \frac{a_i}{N_i} \\ -2\omega_{aa} \frac{a_a}{N_a} \end{pmatrix}, \quad (125)$$

where

$$\mathbb{M} = \begin{pmatrix} 2\omega_{ee} - 2 & 2\omega_{ei} & 0 & 0 & 0 & 0 \\ \omega_{ie} & \omega_{ee} + \omega_{ii} - 2 & \omega_{ei} & 0 & 0 & 0 \\ 0 & 2\omega_{ie} & 2\omega_{ii} - 2 & 0 & 0 & 0 \\ \omega_{ae} & \omega_{ai} & 0 & \omega_{ee} + \omega_{aa} - 2 & \omega_{ei} & 0 \\ 0 & \omega_{ae} & \omega_{ai} & \omega_{ie} & \omega_{ii} + \omega_{aa} - 2 & 0 \\ 0 & 0 & 0 & 2\omega_{ae} & 2\omega_{ai} & 2\omega_{aa} - 2 \end{pmatrix} \quad (126)$$

The matrix decomposes into four 3x3 blocks

$$\mathbb{M} = \begin{pmatrix} \mathbb{M}_{EI} & \mathbb{M}_{fback} \\ \mathbb{M}_{fwd} & \mathbb{M}_A \end{pmatrix} \quad (127)$$

and because $\mathbb{M}_{fback} = 0$ the system of equations becomes

$$\mathbb{M}_{EI} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ii} \end{pmatrix} = \begin{pmatrix} -2\omega_{ee} \frac{a_e}{N_e} \\ -\omega_{ei} \frac{a_i}{N_i} - \omega_{ie} \frac{a_e}{N_e} \\ -2\omega_{ii} \frac{a_i}{N_i} \end{pmatrix} \quad (128)$$

$$\mathbb{M}_{fwd} \cdot \begin{pmatrix} c_{ee} \\ c_{ei} \\ c_{ii} \end{pmatrix} + \mathbb{M}_A \cdot \begin{pmatrix} c_{ea} \\ c_{ia} \\ c_{aa} \end{pmatrix} = \begin{pmatrix} -\omega_{ae} \frac{a_e}{N_e} \\ -\omega_{ai} \frac{a_i}{N_i} \\ -2\omega_{aa} \frac{a_a}{N_a} \end{pmatrix} \quad (129)$$

so that we can indeed first solve the equation for the homogeneous EI-network first and then use the resulting covariances to obtain also the solution of the assembly system.

Covariances of the homogeneous EI-network

The solution of (128) can be obtained by Gaussian elimination yielding

$$c_{ee} = \frac{a_e}{N_e} A - \frac{\omega_{ei}}{\omega_{ee}} C$$

$$c_{ei} = \frac{C}{A}$$

$$c_{ii} = \frac{\omega_{ii} \frac{a_i}{N_i} A + \omega_{ie} B}{\omega_{ie} \omega_{ei} - (\omega_{ii} - 1) A}$$

with the definitions

$$A = \omega_{ee} + \omega_{ii} - 2 - \frac{\omega_{ie} \omega_{ei}}{\omega_{ee} - 1}$$

$$B = -\omega_{ei} \frac{a_i}{N_i} + \frac{a_e}{N_e} \left(\frac{\omega_{ee} \omega_{ie}}{\omega_{ee} - 1} - \omega_{ie} \right)$$

$$C = (2\omega_{ii} - 2)AB - 4\omega_{ie} \omega_{ei} B + 2\omega_{ii} \omega_{ei} \frac{a_i}{N_i} A$$

This general result can be simplified if the generic case is assumed, ie. a network of N excitatory and γN inhibitory neurons, where all excitatory synapses have weight J , all inhibitory synapses have weight $-gJ$ and each neuron (be it excitatory or inhibitory) gets K randomly drawn excitatory inputs, as well as γK randomly drawn inhibitory inputs. Then, $S_e = S_i = S$ allowing the definition $\omega = KJS$.

Because here $\omega_{ie} = \omega_{ee} (= \omega)$ and $\omega_{ei} = \omega_{ii} (= -\gamma g \omega)$, it is possible by close examination of (128) to read off $c_{ei} = \frac{1}{2} (c_{ee} + c_{ii})$ allowing to reduce the 3x3 matrix to 2x2:

$$\begin{pmatrix} 2\omega - \gamma g \omega - 2 & -\gamma g \omega \\ \omega & \omega - 2\gamma g \omega - 2 \end{pmatrix} \begin{pmatrix} c_{ee} \\ c_{ii} \end{pmatrix} = \begin{pmatrix} -2\omega \frac{a}{N} \\ -2g\omega \frac{a}{N} \end{pmatrix}$$

where also $a_i = a_e = m(1 - m)$. This equation can be reordered to

$$\left[\mathbf{1} - \frac{\omega}{2} \begin{pmatrix} 2 - \gamma g & -\gamma g \\ 1 & 1 - 2\gamma g \end{pmatrix} \right] \cdot \begin{pmatrix} c_{ee} \\ c_{ii} \end{pmatrix} = \frac{\omega a}{N} \begin{pmatrix} 1 \\ -g \end{pmatrix}$$

which is the same as in tn_corr sect.6.4 . The solution is then, analogously

$$\begin{pmatrix} c_{ee} \\ c_{ii} \end{pmatrix} = \left[\begin{pmatrix} 1 \\ -g \end{pmatrix} + \frac{\omega}{2} (g^2 \gamma - 1) \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \omega g \begin{pmatrix} \gamma \\ 1 \end{pmatrix} \right] \frac{\omega a}{N (1 - \frac{1}{2} L) (1 - L)}$$

$$c_{ei} = \frac{1}{2} (c_{ee} + c_{ii}) \tag{130}$$

$$L = \omega(1 - \gamma g)$$

Covariances of the assembly subnetwork

Using these results, we can now solve (129). Here, no opportunity to find a further reduction of the matrix dimension is evident. In order to reduce the number of parameters and thus the length of the result while keeping most of the plausible parameter range, we again go to the generic case:

This means $N_e = N$, $N_i = \gamma N$, $J_e = J_a = J$, $J_i = -gJ$ and $a_e = a_i = a$, $S_e = S_i = \frac{S_a}{\kappa} = S$, where we kept the possible deviation of S_a from S (this only occurs when $m_a \neq m$, therefore, $\kappa = 1$ in the subcritical regime). Now we can define again $\omega = pNJS$, yielding the connectivity matrix

$$\omega_{\alpha\beta} = \begin{pmatrix} \omega & -g\gamma\omega & 0 \\ \omega & -g\gamma\omega & 0 \\ (1-f)\kappa\omega & -g\gamma\kappa\omega & f\kappa\omega \end{pmatrix} \tag{131}$$

Using the abbreviations

$$\begin{aligned}
A &= -(1-f)\kappa\left(\frac{a}{N} + c_{ee}\right) + g\gamma\kappa c_{ei} \\
B &= g\gamma\kappa\left(\frac{a}{N} + c_{ii}\right) - (1-f)\kappa c_{ei} \\
C &= F\left(f\kappa - g\gamma - \frac{2}{\omega}\right) - g^2\gamma^2 \\
D &= -2g\gamma\kappa F + 2g\gamma\kappa(1-f) \\
E &= -2f\kappa\frac{a_a}{N_a}F - 2(1-f)\kappa A \\
F &= 1 + f\kappa - \frac{2}{\omega}
\end{aligned}$$

the solution of the linear system obtained by gaussian elimination is then:

$$\begin{aligned}
c_{ea} &= \frac{A}{F} + g\gamma\frac{B}{C} + g^2\gamma^2\frac{A}{FC} \\
c_{ia} &= F\frac{B}{C} + g\gamma\frac{A}{FC} \\
c_{aa} &= \frac{1}{2\left(f\kappa - \frac{1}{\omega}\right)}\left(\frac{E}{F} - \frac{DB}{C} - g\gamma\frac{DA}{FC}\right)
\end{aligned} \tag{132}$$

Unfortunately these expressions are still very involved. The result is shown in Figure 9.

XXIV. SUPPLEMENTARY CALCULATIONS PATH INTEGRAL FORMULATION

A. Path integral with discretized time

Consider the continuous time path integral measure to be defined by the limes of a discrete version of M time steps leading from $t_0 = -\frac{T}{2}$ to $t_0 + M\Delta t = t_M = \frac{T}{2}$:

$$\int \mathcal{D}x := \lim_{T \rightarrow \infty} \lim_{\Delta t \rightarrow 0} \left(\int \prod_t dx_t \right). \tag{133}$$

Here it can be seen why the path integral only exists as a ratio with it's normalization, as the infinite product will usually not converge. Then the integral in the exponent can be written as a matrix multiplication

$$\begin{aligned}
\int \mathcal{D}x \exp\left(-\frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + i\mathbf{j}^T \mathbf{x}\right) &= \lim_{T \rightarrow \infty} \lim_{\Delta t \rightarrow 0} \left(\int \prod_t dx_t \exp\left(\sum_{s,u} -\frac{(\Delta t)^2}{2} x_s A_{su} x_u + i \sum_s \Delta t j_s x_s \right) \right)^{\Delta t} \\
&= \lim_{T \rightarrow \infty} \lim_{\Delta t \rightarrow 0} \sqrt{\frac{2\pi}{\det(\mathbf{A}(\Delta t)^2)}} \exp\left(-\frac{1}{2} \sum_{s,u} (\Delta t)^2 j_s \frac{A_{su}^{-1}}{(\Delta t)^2} j_u \right) \\
&= \sqrt{\frac{2\pi}{\det[A(t,s)]}} \exp\left(-\frac{1}{2} \mathbf{j}^T \mathbf{A}^{-1} \mathbf{j}\right)
\end{aligned} \tag{134}$$

. The gaussian integral can only be performed if $\mathbf{A}$ is symmetric and positive definite. For Q_1^* , symmetry follows from $Q_1^*(t,s) = Q_1^*(s,t)$ and positive definiteness is given if the function is dominated by the necessarily positive zero

time lag autocorrelations $Q_1^*(t, t)$. Finally, one can now see that $A^{-1}(s, u) := \frac{A_{su}^{-1}}{(\Delta t)^2}$ indeed fulfills the condition (41)

$$\begin{aligned}
\int dt \delta(t - u) &= \iint dt ds Q_1^*(t, s) Q_1^{*-1}(s, u) \\
&= \lim_{\Delta t \rightarrow 0} \sum_{t, s} (\Delta t)^2 Q_{1, ts}^* \frac{Q_{1, su}^{*-1}}{(\Delta t)^2} \\
&= \lim_{\Delta t \rightarrow 0} \sum_{t, s} Q_{1, ts}^* Q_{1, su}^{*-1} \\
&= \lim_{\Delta t \rightarrow 0} \sum_t \text{diag}_{tu} = 1.
\end{aligned} \tag{135}$$

B. Exact equation for $Q_1^*(t, s)$

The following derivation until (138) was done by Tobias Kühn.

We need the quantity

$$\rho[x(t), x(s) | \mathbf{h}],$$

i.e. the probability that the neuron was in the state $x(t)$ at the time t and the state $x(s)$ at the time $s < t$ given the input history $\mathbf{h}$. We can express this using the conditioned probability of $x(t)$ given the state $x(s)$:

$$\rho[x(t), x(s) | \mathbf{h}] = \underbrace{\rho[x(t) | x(s), \mathbf{h}]}_{\xrightarrow{t \rightarrow s} \delta_{x(t), x(s)}} \rho[x(s) | \mathbf{h}].$$

Clearly, the conditioned probability has to go to 1 if t approaches s . This will give us a check for our result that we will calculate in the following separately for all four possibilities for the states $x(t), x(s) = 0, 1$:

$$\rho[x(t) = 1 | x(s) = 1, \mathbf{h}] = P(\text{no updates in } (s, t)) + P(\text{last update in } (s, t) \text{ to up-state})$$

$$\begin{aligned}
&= \exp\left(-\frac{t-s}{\tau}\right) + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} \\
&= 1 + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (T(h(t')) - 1) \frac{dt'}{\tau}
\end{aligned}$$

$$\rho[x(t) = 0 | x(s) = 0, \mathbf{h}] = 1 - \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}$$

$$\rho[x(t) = 1 | x(s) = 0, \mathbf{h}] = \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}$$

$$\rho[x(t) = 0 | x(s) = 1, \mathbf{h}] = \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (1 - T(h(t')))) \frac{dt'}{\tau}.$$

Summarized, this gives

$$\begin{aligned}
& \rho[x(t)|x(s), \mathbf{h}] \\
&= x(t)x(s) \left(1 + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (T(h(t')) - 1) \frac{dt'}{\tau}\right) \\
&+ (1-x(t))(1-x(s)) \left(1 - \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}\right) \\
&+ x(t)(1-x(s)) \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} \\
&+ (1-x(t))x(s) \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (1-T(h(t')))) \frac{dt'}{\tau} \\
&= x(t) \left[x(s) \left(1 - \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) \frac{dt'}{\tau}\right) + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} \right] \\
&+ (1-x(t)) \left[1 - \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} + x(s) \left(\int_s^t \exp\left(-\frac{t-t'}{\tau}\right) \frac{dt'}{\tau} - 1 \right) \right] \\
&= x(s) \left[x(t) \left(1 + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (T(h(t')) - 1) \frac{dt'}{\tau}\right) + (1-x(t)) \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (1-T(h(t')))) \frac{dt'}{\tau} \right] \\
&+ (1-x(s)) \left[x(t) \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} + (1-x(t)) \left(1 - \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}\right) \right] \\
&\xrightarrow{t \searrow s} x(t)x(s) + (1-x(t))(1-x(s)) = \delta_{x(t), x(s)}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
& P(x(t), x(s) | \mathbf{h}) \\
&= P(x(t) | x(s), \mathbf{s}) P(x(s) | \mathbf{h}) \\
&= x(s) \int_{-\infty}^s \exp\left(-\frac{s-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} \\
&\cdot \left[x(t) \left(1 + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (T(h(t')) - 1) \frac{dt'}{\tau}\right) + (1-x(t)) \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (1-T(h(t')))) \frac{dt'}{\tau} \right] \\
&+ (1-x(s)) \left(1 - \int_{-\infty}^s \exp\left(-\frac{s-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}\right) \\
&\cdot \left[x(t) \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) dt' + (1-x(t)) \left(1 - \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}\right) \right].
\end{aligned}$$

and we deduce

$$\begin{aligned}
& \sum_{x(s)=0,1} \sum_{x(s)=0,1} x(t) x(s) P(x(t), x(s) | \mathbf{h}) \\
&= \int_{-\infty}^s \exp\left(-\frac{s-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} \left(1 + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) (T(h(t')) - 1) \frac{dt'}{\tau}\right) \\
&= \int_{-\infty}^s \exp\left(-\frac{s-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau} \left(\exp\left(-\frac{t-s}{\tau}\right) + \int_s^t \exp\left(-\frac{t-t'}{\tau}\right) T(h(t')) \frac{dt'}{\tau}\right),
\end{aligned}$$

which leads to

$$\begin{aligned}
& Q^*(t, s) \\
&= \frac{g^2}{2} \left[\exp\left(-\frac{t-s}{\tau}\right) \int_{-\infty}^s \exp\left(-\frac{s-t'}{\tau}\right) \langle T(x) \rangle_{x \sim \mathcal{N}(R(t'), Q(t', t'))} \right. \\
&\quad \left. + \int_{-\infty}^s \frac{dt'}{\tau} \int_s^t \frac{dt''}{\tau} \exp\left(-\frac{(s+t)+t'+t''}{\tau}\right) \langle T(x) T(x') \rangle \left(\begin{array}{c} x \\ x' \end{array} \right)_{\sim \mathcal{N}} \left(\begin{pmatrix} R(t') \\ R(t'') \end{pmatrix}, \begin{pmatrix} Q(t', t') & Q(t', t'') \\ Q(t'', t') & Q(t'', t'') \end{pmatrix} \right) \right].
\end{aligned}$$

Taking the derivative with respect to t , we get

$$\begin{aligned}
& \tau \frac{d}{dt} Q^*(t, s) \\
&= -Q^*(t, s) + \int_{-\infty}^s \exp\left(-\frac{s-t'}{\tau}\right) \langle T(x) T(x') \rangle \left(\begin{array}{c} x \\ x' \end{array} \right)_{\sim \mathcal{N}} \left(\begin{pmatrix} R(t') \\ R(t) \end{pmatrix}, \begin{pmatrix} Q^*(t', t') & Q^*(t', t) \\ Q^*(t, t') & Q^*(t, t) \end{pmatrix} \right) \frac{dt'}{\tau} \\
&\stackrel{t''=s-t'}{=} -Q^*(t, s) + \int_0^\infty \exp\left(-\frac{t''}{\tau}\right) \langle T(x) T(x') \rangle \left(\begin{array}{c} x \\ x' \end{array} \right)_{\sim \mathcal{N}} \left(\begin{pmatrix} R(s-t'') \\ R(t) \end{pmatrix}, \begin{pmatrix} Q^*(s-t'', s-t'') & Q^*(s-t'', t) \\ Q^*(t, s-t'') & Q^*(t, t) \end{pmatrix} \right) \frac{dt''}{\tau}
\end{aligned} \tag{136}$$

In the stationary state, R is constant and Q^* just depends on the time lag $\Delta t := t - s$ and we have (compare also van Vreeswijk & Sompolinsky (1998))

$$\tau \frac{d}{d\Delta t} Q^*(\Delta t) = -Q^*(\Delta t) + \int_0^\infty e^{-\frac{t''}{\tau}} \langle T(x) T(x') \rangle \left(\begin{array}{c} x \\ x' \end{array} \right)_{\sim \mathcal{N}} \left(\begin{pmatrix} R \\ R \end{pmatrix}, \begin{pmatrix} Q^*(0) & Q^*(-(\Delta t + t'')) \\ Q(\Delta t + t'') & Q^*(0) \end{pmatrix} \right) \frac{dt''}{\tau} \tag{137}$$

$$\begin{aligned}
& \stackrel{\text{i.b.p.}}{=} -Q^*(\Delta t) + \langle T(x) T(x') \rangle \left(\begin{array}{c} x \\ x' \end{array} \right)_{\sim \mathcal{N}} \left(\begin{pmatrix} R \\ R \end{pmatrix}, \begin{pmatrix} Q^*(0) & Q^*(-\Delta t) \\ Q(\Delta t) & Q^*(0) \end{pmatrix} \right) \\
&+ \sum_{n=1}^\infty \tau^n \frac{d^n}{d\Delta t^n} \langle T(x) T(x') \rangle \left(\begin{array}{c} x \\ x' \end{array} \right)_{\sim \mathcal{N}} \left(\begin{pmatrix} R \\ R \end{pmatrix}, \begin{pmatrix} Q^*(0) & Q^*(-(\Delta t)) \\ Q(\Delta t) & Q^*(0) \end{pmatrix} \right)
\end{aligned} \tag{138}$$

Note that assuming the entries $Q^*(\Delta t + t'')$ on the off diagonal in (137) to be independent of t'' results in a trivial execution of the time integral $\int_0^\infty e^{-\frac{t''}{\tau}} \frac{dt''}{\tau} = 1$, so that the equation is solved by an exponential

$$\begin{aligned}
Q^*(\Delta t) &= Q^*(0) e^{-\frac{\Delta t}{\tau}} + \langle T(x) T(x') \rangle \left(1 - e^{-\frac{\Delta t}{\tau}}\right), \\
&\approx g^2 \bar{x} e^{-\frac{\Delta t}{\tau}} + g^2 \bar{x}^2 \left(1 - e^{-\frac{\Delta t}{\tau}}\right)
\end{aligned} \tag{139}$$

which corresponds to the approximate solution (53) used in the main text.

C. Finding $S_{12}^{(RR)}(t, s)$ by explicit calculation of the expectation value

$$\begin{aligned}
S_{12}^{(RR)}(t, s) &= S_{21}^{(RR)}(s, t) = \delta(t - s) + \frac{m}{\sqrt{N}} \left\langle \sum_{i,j} \hat{h}_i(t) x_j(s) \right\rangle_{Q^*, R^*} \\
&= \delta(t - s) + \frac{1}{Z(0)} \frac{m}{\sqrt{N}} \sum_{i,j} \int \mathcal{D}\mathbf{h} \mathcal{D}\hat{\mathbf{h}} \hat{h}_i(t) \exp\left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1^*) + \frac{1}{2} \hat{\mathbf{h}}^T Q_1^* \hat{\mathbf{h}}\right) \int \mathcal{D}\mathbf{x} x_j(s) \rho[\mathbf{x}|\mathbf{h}].
\end{aligned} \tag{140}$$

First, we execute the $\hat{\mathbf{h}}$ integration. Using the substitution $\tilde{h}_i(t) = \hat{h}_i(t) + \int du Q_1^{*-1}(t, u) (h_i(u) - R_1(u))$ gives

$$\begin{aligned}
&\int \mathcal{D}\hat{\mathbf{h}} \hat{h}_i(t) \exp\left(\hat{\mathbf{h}}^T (\mathbf{h} - R_1^*) + \frac{1}{2} \hat{\mathbf{h}}^T Q_1^* \hat{\mathbf{h}}\right) \\
&= \int \mathcal{D}\tilde{\mathbf{h}} \left(\tilde{h}_i(t) - \frac{h_i - R_1^*}{Q_1^*}(t)\right) \exp\left(\tilde{\mathbf{h}}^T (\mathbf{h} - R_1^*) - (\mathbf{h} - R_1^*)^T \frac{1}{Q_1^*} (\mathbf{h} - R_1^*) + \frac{1}{2} \left(\tilde{\mathbf{h}} - \frac{\mathbf{h} - R_1^*}{Q_1^*}\right)^T Q_1^* \left(\tilde{\mathbf{h}} - \frac{\mathbf{h} - R_1^*}{Q_1^*}\right)\right) \\
&= \int \mathcal{D}\tilde{\mathbf{h}} \left(\tilde{h}_i(t) - \frac{h_i - R_1^*}{Q_1^*}(t)\right) \exp\left(\frac{1}{2} \tilde{\mathbf{h}}^T Q_1^* \tilde{\mathbf{h}} - \frac{1}{2} (\mathbf{h} - R_1^*)^T \frac{1}{Q_1^*} (\mathbf{h} - R_1^*)\right) \\
&= -Z_{\tilde{\mathbf{h}}}(0) \frac{h_i - R_1^*}{Q_1^*}(t) \exp\left(-\frac{1}{2} (\mathbf{h} - R_1^*)^T \frac{1}{Q_1^*} (\mathbf{h} - R_1^*)\right),
\end{aligned} \tag{141}$$

where we used in the last step that $\int \mathcal{D}\tilde{\mathbf{h}} \tilde{h}_i \exp\left(\frac{1}{2} \tilde{\mathbf{h}}^T Q_1^* \tilde{\mathbf{h}}\right) = 0$ by symmetry, so that only the second term contributes. Plugging the expression back into (140) gives

$$\begin{aligned}
S_{12}^{(RR)}(t, s) &= \delta(t - s) - \frac{Z_{\tilde{\mathbf{h}}}(0)}{Z(0)} \frac{m}{\sqrt{N}} \sum_{i,j} \int \mathcal{D}\mathbf{h} \frac{h_i - R_1^*}{Q_1^*}(t) \exp\left(-\frac{1}{2} (\mathbf{h} - R_1^*)^T \frac{1}{Q_1^*} (\mathbf{h} - R_1^*)\right) \int \mathcal{D}\mathbf{x} x_j(s) \rho[\mathbf{x}|\mathbf{h}] \\
&= \delta(t - s) - \frac{Z_{\tilde{\mathbf{h}}}(0)}{Z(0)} \frac{m}{\sqrt{N}} \sum_{i,j} \int_{-\infty}^s \frac{ds'}{\tau} e^{-\frac{s-s'}{\tau}} \\
&\quad \times \int \mathcal{D}\mathbf{h} \int du \frac{h_i(u) - R_1^*(u)}{Q_1^*(t, u)} \exp\left(-\frac{1}{2} (\mathbf{h} - R_1^*)^T \frac{1}{Q_1^*} (\mathbf{h} - R_1^*)\right) H(h_j(s') - \theta).
\end{aligned} \tag{142}$$

Now the $\mathbf{h}$ integration can be considered. In the discretized version all the $d\mathbf{h}(v)$ integrations with $v \neq u, s'$ are gaussian and can be collapsed to the normalization. Then, for the case $i \neq j$, the integral over $h_i(u)$ takes the again form $\int d\mathbf{h} \mathbf{h} \exp(-\frac{\mathbf{h}^2}{2Q_1^*}) = 0$, and the same happens for $u \neq s'$. Thus only the term with $i = j$ and $u = s'$ survives, which

gives

$$\begin{aligned}
& \int \mathcal{D}\mathbf{h} \frac{h_i(s') - R_1^*(s')}{Q_1^*(t, s')} \exp\left(-\frac{1}{2}(\mathbf{h} - R_1^*) \frac{1}{Q_1^*}(\mathbf{h} - R_1^*)\right) H(h_i(s') - \theta) \\
&= \mathcal{Z}_{\mathbf{h} \setminus s'} \int_{\theta - R_1^*(s')}^{\infty} d(h_i(s') - R_1^*(s')) \frac{h_i(s') - R_1^*(s')}{Q_1^*(t, s')} \exp\left(-\frac{1}{2}(\mathbf{h}(s') - R_1^*(s')) \frac{1}{Q_1^*(s', s')}(\mathbf{h}(s') - R_1^*(s'))\right) \\
&= \mathcal{Z}_{\mathbf{h} \setminus s'} \frac{Q_1^*(s', s')}{Q_1^*(t, s')} \int_{\theta - R_1^*(s')}^{\infty} dy \frac{y}{Q_1^*(s', s')} \exp\left(-\frac{1}{2} \frac{y^2}{Q_1^*(s', s')}\right) \\
&= -\mathcal{Z}_{\mathbf{h} \setminus s'} \frac{Q_1^*(s', s')}{Q_1^*(t, s')} \int_{\theta - R_1^*(s')}^{\infty} dy \frac{\partial}{\partial y} \exp\left(-\frac{1}{2} \frac{y^2}{Q_1^*(s', s')}\right) \\
&= -\mathcal{Z}_{\mathbf{h} \setminus s'} \frac{Q_1^*(s', s')}{Q_1^*(t, s')} \exp\left(-\frac{1}{2} \frac{(\theta - R_1^*(s'))^2}{Q_1^*(s', s')}\right). \tag{143}
\end{aligned}$$

$$= -\mathcal{Z}_{\mathbf{h}} \frac{1}{\sqrt{2\pi} Q_1^*(t, s')} \exp\left(-\frac{1}{2} \frac{(\theta - R_1^*(s'))^2}{Q_1^*(s', s')}\right) \tag{144}$$

Plugging this back into (142), we obtain the final result:

$$\begin{aligned}
S_{12}^{(RR)}(t, s) &= S_{21}^{(RR)}(s, t) = \delta(t - s) + \frac{m}{\sqrt{N}} \left\langle \sum_{i,j} \hat{h}_i(t) x_j(s) \right\rangle_{Q^*, R^*} \\
&= \delta(t - s) + \frac{m}{\sqrt{N}} \sum_i \int_{-\infty}^s \frac{ds'}{\tau} e^{-\frac{s-s'}{\tau}} \frac{1}{\sqrt{2\pi} Q_1^*(t, s')} \exp\left(-\frac{1}{2} \frac{(\theta - R_1^*(s'))^2}{Q_1^*(s', s')}\right) \tag{145}
\end{aligned}$$

$$= \delta(t - s) + \frac{m}{\sqrt{N}} \tilde{K}(t, s). \tag{146}$$

However, this differs slightly from the result in the main text (68), in that $Q_1^*(t, s')$ appears in the denominator instead of $\sqrt{Q_1^*(s', s')}$. We think that the above calculation must be erroneous, since the result (68) leads to very logical further expressions, especially fulfilling the cross check to covariances in linear networks below (84). We nonetheless show this calculation because the error was not immediately found.

D. Disorder average for fixed indegree

The following derivation up to (151) was done by Tobias Kühn.

We want to calculate $\langle \exp(-\hat{h}^T J x) \rangle_J$ under the constraint of fixed in-degrees, i.e. we demand that

$$\sum_j J_{ij} = K \frac{j_0}{\sqrt{N}} \quad \forall i.$$

As the only information about the distribution of the synapses is captured in the first two cumulants, we have to calculate those.

We first observe that there are only $\binom{N}{K}$ possibilities to draw K out of N possible synapses for every postsynaptic neuron. We obtain the moments of single synapse coming from neuron i and terminating at an arbitrary neuron j by considering that for a binary variable n , all moments are identical to p ($n = 1$):

$$\kappa_{ij} = \frac{j_0}{\sqrt{N}} \frac{p \binom{N-1}{K-1} p^{K-1}}{\binom{N}{K} p^K} = \frac{j_0}{\sqrt{N}} \frac{K}{N} = \frac{j_0}{\sqrt{N}} p$$

As in the unrestricted case, the second cumulant for two identical connections is given by $\langle\langle J_{ij}^2 \rangle\rangle \frac{j_0^2}{N} p(1-p)$. For different presynaptic neurons, however, we have only $\binom{N-2}{K-2}$ possibilities to draw the missing $K-2$ synapses, which gives for the second moment

$$\langle J_{ij} J_{ik} \rangle = \frac{j_0^2}{N} \frac{p^2 \binom{N-2}{K-2} p^{K-2}}{\binom{N}{K} p^K} = \frac{j_0^2}{N} \frac{K}{N} \frac{K-1}{N-1} = \frac{j_0^2}{N} p \frac{K-1}{N-1}.$$

Lumping this together in one expression, we get

$$\kappa_{ij}^{(1)} = \frac{m}{\sqrt{N}} = \frac{j_0 p}{\sqrt{N}} \quad \forall i, j \quad (147)$$

$$\kappa_{ij,kl}^{(2)} = \frac{j_0^2 p(1-p)}{N} \left(\delta_{ik} \left(\delta_{jl} - \frac{1 - \delta_{jl}}{N-1} \right) \right) = \frac{g^2}{N} \left(\delta_{ik} \left(\delta_{jl} - \frac{1 - \delta_{jl}}{N-1} \right) \right) \quad (148)$$

$$= \frac{g^2}{N} \delta_{ik} \left(\delta_{jl} \frac{N}{N-1} - \frac{1}{N-1} \right) = \frac{g^2}{N-1} \delta_{ik} \left(\delta_{jl} - \frac{1}{N} \right). \quad (149)$$

Note that different synapses targeting the same postsynaptic neuron are anticorrelated because the number of incoming synapses is limited by K . With that, we can rewrite the disorder average as

$$\exp \left(\frac{g^2}{2(N-1)} \left(\sum_{i,k} \hat{h}_i \hat{h}_i x_j x_j - \sum_{i,k} \hat{h}_i \hat{h}_i x_j \frac{1}{N} \sum_l x_l \right) - \sum_{i,j} \kappa_{ij}^{(1)} \hat{h}_i x_j \right).$$

$$\begin{aligned} & \langle \exp(-\hat{h}^T J x) \rangle_J \\ & \approx \Pi_i \exp \left(- \int \hat{h}_i(t) \frac{m}{\sqrt{N}} \sum_j x_j(t) + \frac{g^2}{2(N-1)} \iint \hat{h}_i(t) \hat{h}_i(s) \sum_j x_j(t) x_j(s) - \frac{g^2}{2(N-1)} \iint dt ds \hat{h}_i(t) \hat{h}_i(s) \frac{1}{N} \sum_{k,l} x_k(t) x_l(s) \right) \\ & = \Pi_i \exp \left(- \int \hat{h}_i(t) \frac{m}{\sqrt{N}} \sum_j x_j(t) + \frac{g^2}{2(N-1)} \iint \hat{h}_i(t) \hat{h}_i(s) \sum_j x_j(t) \underbrace{\left(x_j(s) - \frac{1}{N} \sum_l x_l(s) \right)}_{=: \delta x_j(s)} \right). \end{aligned} \quad (150)$$

Note that the definition of δx involves an average over the population, not over realizations. Assuming that the system is self-averaging, however, this does not make any difference. A natural choice for the auxiliary variable Q_1 now would be

$$Q_1(t, s) = \frac{g^2}{2(N-1)} \sum_j x_j(t) \delta x_j(s). \quad (151)$$

Then we can find the relation to the expression for fixed indegree:

$$\begin{aligned}
Q_1^{\star fixed}(t, s) &= \frac{g^2}{N-1} \sum_i \langle x_i(t) \delta x_i(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \\
&= \frac{g^2}{N-1} \sum_i \left\langle x_i(t) \left(x_i(s) - \frac{1}{N} \sum_j x_j(s) \right) \right\rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \\
&= \frac{N}{N-1} Q_1^{\star distr.}(t, s) - \frac{g^2}{N(N-1)} \sum_{i,j} \langle x_i(t) x_j(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \\
&= \frac{N}{N-1} Q_1^{\star distr.}(t, s) - \frac{g^2}{N(N-1)} \left(\sum_i \langle x_i(t) x_i(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} + \sum_{i \neq j} \langle x_i(t) x_j(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \right) \\
&= Q_1^{\star distr.}(t, s) - \frac{g^2}{N(N-1)} \sum_{i \neq j} \langle x_i(t) x_j(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*} \\
&= Q_1^{\star distr.}(t, s) - \frac{g^2}{N(N-1)} \sum_{i \neq j} \bar{x}(t) \bar{x}(s) - \frac{g^2}{N(N-1)} \underbrace{\sum_{i \neq j} \langle \delta x_i(t) \delta x_j(s) \rangle_{\mathbf{J}, \mathbf{h}}^{R^*, Q^*}}_{=0}
\end{aligned}$$

The last term vanishes in zero order approximation, because as shown in (40) there is no coupling between units as $\rho^*[\mathbf{h}] = \prod_i \rho^*[h_i]$ factorizes. This is because of $R_2^* = Q_2^* = 0$, which removes the couplings seen in the general expression (33). Thus

$$Q_1^{\star fixed}(t, s) = Q_1^{\star distr.}(t, s) - g^2 \bar{x}(t) \bar{x}(s).$$

E. Some further calculations in greater detail

The convolution structure of the propagator Δ_{11} used in (82) is seen as follows:

$$\begin{aligned}
\Delta_{11}^{(RR)}(u-r) &= - \int dt \Delta_{12}^{(RR)}(t-r) \int ds S_{22}^{(RR)}(s-t) \Delta_{21}^{(RR)}(u-s) \\
&\stackrel{\bar{s}=s-t}{=} - \int dt \Delta_{12}^{(RR)}(t-r) \int d\bar{s} S_{22}^{(RR)}(\bar{s}) \Delta_{21}^{(RR)}(u-\bar{s}-t) \\
&= - \int dt \Delta_{12}^{(RR)}(t-r) \left(\Delta_{21}^{(RR)} * S_{22}^{(RR)} \right)(u-t) \\
&\stackrel{\bar{t}=t-r}{=} - \int dt \Delta_{12}^{(RR)}(\bar{t}) \left(\Delta_{21}^{(RR)} * S_{22}^{(RR)} \right)(u-\bar{t}-r) \\
&= - \left(\left(\Delta_{21}^{(RR)} * S_{22}^{(RR)} \right) * \Delta_{12}^{(RR)} \right)(u-r).
\end{aligned} \tag{152}$$

The steps leading to the Fourier transform of S_{22} in (83) are

$$\begin{aligned}
 S_{22}(\omega) &= m^2 \bar{x}(1 - \bar{x}) \mathcal{F}\left[e^{-\frac{|\Delta t|}{\tau}}\right](\omega) \\
 &= \frac{m^2 \tau}{\sqrt{2\pi}} \bar{x}(1 - \bar{x}) \left[\int_0^\infty dt e^{-t} \exp(-i\tau\omega t) + \int_{-\infty}^0 dt e^t \exp(-i\tau\omega t) \right] \\
 &= \frac{m^2 \tau}{\sqrt{2\pi}} \bar{x}(1 - \bar{x}) \left[\frac{-1}{1 + i\tau\omega} + \frac{-1}{1 - i\tau\omega} \right] \\
 &= -\frac{m^2 \tau}{\sqrt{2\pi}} \bar{x}(1 - \bar{x}) \frac{2}{1 + (\tau\omega)^2}.
 \end{aligned} \tag{153}$$

XXV. REFERENCES

-
- Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex*. (Cambridge University Press).
- Abeles, M. (2011). Cell assemblies. *Scholarpedia* 6, 1505.
- Barral, J. & Reyes, A. D. (2016). Synaptic scaling rule preserves excitatory-inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience* 19, 1690–1696.
- Barth, A. L. & Poulet, J. F. A. (2012). Experimental evidence for sparse firing in the neocortex. *Trends in Neurosciences* 35, 345–355.
- Beltramo, R., D’Urso, G., Maschio, M. D., Farisello, P., Bovetti, S., Clovis, Y., Lassi, G., Tucci, V., Tonelli, D. D. P., & Fellin, T. (2013). Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. *Nature Neuroscience* 16, nn.3306.
- Bos, H., Morrison, A., Peyser, A., Hahne, J., Helias, M., Kunkel, S., Ippen, T., Eppler, J. M., Schmidt, M., Seeholzer, A., et al. (2015). NEST 2.10.0. Tech. rep., Zenodo. DOI: 10.5281/zenodo.44222.
- Brunel, N. (2000). Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons. *Journal of Computational Neuroscience* 8, 183–208.
- Buice, M. A. & Cowan, J. D. (2007). Field-theoretic approach to fluctuation effects in neural networks. *Physical Review E* 75, 051919.
- Buice, M. A., Cowan, J. D., & Chow, C. C. (2010). Systematic fluctuation expansion for neural network activity equations. *Neural computation* 22, 377–426.
- Burkitt, A. N. (2006). A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input. *Biological Cybernetics* 95, 1–19.
- Buzsáki, G. (2010). Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron* 68, 362–385.
- Chow, C. C. & Buice, M. A. (2015). Path Integral Methods for Stochastic Differential Equations. *The Journal of Mathematical Neuroscience (JMN)* 5, 8.
- Coolen, A. C. C. (2000a). Statistical Mechanics of Recurrent Neural Networks I. Statics. arXiv:cond-mat/0006010 arXiv:cond-mat/0006010.
- Coolen, A. C. C. (2000b). Statistical Mechanics of Recurrent Neural Networks II. Dynamics. arXiv:cond-mat/0006011 arXiv:cond-mat/0006011.
- Cossart, R., Aronov, D., & Yuste, R. (2003). Attractor dynamics of network up states in the neocortex. *Nature* 423, 283–288.
- Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* 518, 399–403.

- De Dominicis, C. & Peliti, L. (1978). Field-theory renormalization and critical dynamics above T_c : Helium, antiferromagnets, and liquid-gas systems. *Physical Review B* *18*, 353–376.
- Denève, S. & Machens, C. K. (2016). Efficient codes and balanced networks. *Nature Neuroscience* *19*, nn.4243.
- Doi, M. (1976). Second quantization representation for classical many-particle system. *Journal of Physics A: Mathematical and General* *9*, 1465.
- Friedman, N., Ito, S., Brinkman, B. A. W., Shimono, M., DeVille, R. E. L., Dahmen, K. A., Beggs, J. M., & Butler, T. C. (2012). Universal critical dynamics in high resolution neuronal avalanche data. *Physical Review Letters* *108*, 208102.
- Fukuda, T., Kosaka, T., Singer, W., & Galuske, R. A. W. (2006). Gap Junctions among Dendrites of Cortical GABAergic Neurons Establish a Dense and Widespread Intercolumnar Network. *Journal of Neuroscience* *26*, 3434–3443.
- Ginzburg, I. & Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Physical Review E* *50*, 3171–3191.
- Grytskyy, D., Tetzlaff, T., Diesmann, M., & Helias, M. (2013). A unified view on weakly correlated recurrent networks. *Frontiers in Computational Neuroscience* *7*.
- Harris, K. D. & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature* *503*, 51–58.
- Hattori, R., Kuchibhotla, K. V., Froemke, R. C., & Komiyama, T. (2017). Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nature Neuroscience* *20*, 1199–1208.
- Hebb, D. O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. (Wiley).
- Helias, M., Tetzlaff, T., & Diesmann, M. (2014). The Correlation Structure of Local Neuronal Networks Intrinsically Results from Recurrent Dynamics. *PLOS Computational Biology* *10*, e1003428.
- Hernandez-Urbina, V. & Michael Herrmann, J. (2016). Neuronal avalanches in complex networks. *Cogent Physics* *3*, 1–13.
- Hertz, J. A., Roudi, Y., & Sollich, P. (2017). Path integral methods for the dynamics of stochastic and disordered systems. *Journal of Physics A: Mathematical and Theoretical* *50*, 033001. arXiv: 1604.05775.
- Hohwy, J. (2013). *The Predictive Mind*. (Oxford, New York: Oxford University Press).
- Holtmaat, A. & Caroni, P. (2016). Functional and structural underpinnings of neuronal assembly formation in learning. *Nature Neuroscience* *19*, 1553–1562.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* *79*, 2554–2558.
- Hopfield, J. J. (2007). Hopfield network. *Scholarpedia* *2*, 1977.
- Howard, M. J. & Tauber, U. C. (1997). ‘Real’ versus ‘imaginary’ noise in diffusion-limited reactions. *Journal of Physics A: Mathematical and General* *30*, 7721.
- Huyck, C. R. & Passmore, P. J. (2013). A review of cell assemblies. *Biological Cybernetics* *107*, 263–288.
- Janssen, H.-K. (1976). On a Lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. *Zeitschrift für Physik B Condensed Matter* *23*, 377–380.
- Kandel, E. R. (2013). *Principles of Neural Science, Fifth Edition*. (McGraw Hill Professional).
- Kasthuri, N., Hayworth, K., Berger, D., Schalek, R., Conchello, J., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T., et al. (2015). Saturated Reconstruction of a Volume of Neocortex. *Cell* *162*, 648–661.
- Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S. B., & Mrsic-Flogel, T. D. (2013). The emergence of functional microcircuits in visual cortex. *Nature* *496*, 96–100.
- Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature* *473*, 87–91.
- Kubota, Y., Karube, F., Nomura, M., & Kawaguchi, Y. (2016). The Diversity of Cortical Inhibitory Synapses. *Frontiers in Neural Circuits* *10*.
- Kühn, T. & Helias, M. (2017). Locking of correlated neural activity to ongoing oscillations. *PLOS Computational Biology* *13*, e1005534.
- Latham, P. E. & Nirenberg, S. (2004). Computing and Stability in Cortical Networks. *Neural Computation* *16*, 1385–1412.
- Lee, W.-C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glatfelter, K., & Reid, R. C. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature* *532*, 370–374.

- Lefevre, A. & Biroli, G. (2007). Dynamics of interacting particle systems: stochastic process and field theory. *Journal of Statistical Mechanics: Theory and Experiment* 2007, P07024–P07024. arXiv: 0709.1325.
- Lefort, S., Tómm, C., Floyd Sarria, J. C., & Petersen, C. C. H. (2009). The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex. *Neuron* 61, 301–316.
- Litwin-Kumar, A. & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience* 15, 1498–1505.
- Luczak, A., Barthó, P., & Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* 62, 413–425.
- Malvache, A., Reichinnek, S., Villette, V., Haimerl, C., & Cossart, R. (2016). Awake hippocampal reactivations project onto orthogonal neuronal assemblies. *Science (New York, N.Y.)* 353, 1280–1283.
- Manita, S., Miyakawa, H., Kitamura, K., & Murayama, M. (2017). Dendritic Spikes in Sensory Perception. *Frontiers in Cellular Neuroscience* 11.
- Mao, B.-Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C., & Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices. *Neuron* 32, 883–898.
- Markert, H., Kaufmann, U., Kara Kayikci, Z., & Palm, G. (2009). Neural associative memories for the integration of language, vision and action in an autonomous agent. *Neural Networks* 22, 134–143.
- Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and Simulation of Neocortical Microcircuitry. *Cell* 163, 456–492.
- Martin, P. C., Siggia, E. D., & Rose, H. A. (1973). Statistical Dynamics of Classical Systems. *Physical Review A* 8, 423–437.
- Mazzucato, L., Fontanini, A., & La Camera, G. (2015). Dynamics of multistable states during ongoing and evoked cortical activity. *Journal of Neuroscience* 35, 8214–8231.
- Mikula, S., Trotts, I., Stone, J. M., & Jones, E. G. (2007). Internet-Enabled High-Resolution Brain Mapping and Virtual Microscopy. *NeuroImage* 35, 9–15.
- Miller, J.-e. K., Ayzenshtat, I., Carrillo-Reid, L., & Yuste, R. (2014). Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences* 111, E4053–E4061.
- Miller, P. (2016). Itinerancy between attractor states in neural systems. *Current Opinion in Neurobiology* 40, 14–22.
- Ocker, G. K., Josic, K., Shea-Brown, E., & Buice, M. A. (2017). Linking structure and activity in nonlinear spiking networks. *PLOS Computational Biology* 13, e1005583.
- O’Connor, D. H., Peron, S. P., Huber, D., & Svoboda, K. (2010). Neural Activity in Barrel Cortex Underlying Vibrissa-Based Object Localization in Mice. *Neuron* 67, 1048–1061.
- Onsager, L. & Machlup, S. (1953). Fluctuations and Irreversible Processes. *Physical Review* 91, 1505–1512.
- Palm, G. (1982). *Neural Assemblies - An Alternative Approach to Artificial Intelligence*. (Springer).
- Palm, G. (2013). Neural associative memories and sparse coding. *Neural Networks* 37, 165–171.
- Palm, G., Knoblauch, A., Hauser, F., & Schüz, A. (2014). Cell assemblies in the cerebral cortex. *Biological Cybernetics* 108, 559–572.
- Peliti, L. (1985). Path integral approach to birth-death processes on a lattice. *Journal de Physique* 46, 1469–1483.
- Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences* 108, 5419–5424.
- Perin, R., Telefont, M., & Markram, H. (2013). Computing the size and number of neuronal clusters in local circuits. *Frontiers in Neuroanatomy* 7, 1–10.
- Potjans, T. C. & Diesmann, M. (2014). The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model. *Cerebral Cortex* 24, 785–806.
- Renart, A., Rocha, J. d. l., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The Asynchronous State in Cortical Circuits. *Science* 327, 587–590.
- Rolls, E. T. (2016). *Cerebral Cortex: Principles of Operation*. (Oxford, New York: Oxford University Press).
- Schmidt, M., Bakker, R., Shen, K., Bezgin, G., Hilgetag, C.-C., Diesmann, M., & van Albada, S. J. (2016). Full-density multi-scale account of structure and dynamics of macaque visual cortex. arXiv:1511.09364v4 [q-bio] arXiv: 1511.09364v4.

- Schuecker, J., Goedeke, S., Dahmen, D., & Helias, M. (2016). Functional methods for disordered neural networks pp. 1–24. arXiv: 1605.06758.
- Setareh, H., Deger, M., Petersen, C. C. H., & Gerstner, W. (2017). Cortical dynamics in presence of assemblies of densely connected weight-hub neurons. *Frontiers in Computational Neuroscience* *11*, 1–22.
- Sjöström, J. & Gerstner, W. (2010). Spike-timing dependent plasticity. *Scholarpedia* *5*, 1362.
- Sjöström, P. J., Rancz, E. A., Roth, A., & Häusser, M. (2008). Dendritic Excitability and Synaptic Plasticity. *Physiological Reviews* *88*, 769–840.
- Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in Random Neural Networks. *Physical Review Letters* *61*, 259–262.
- Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology* *3*, e68.
- Stuart, G. J. & Spruston, N. (2015). Dendritic integration: 60 years of progress. *Nature Neuroscience* *18*, nn.4157.
- Tetzlaff, T., Helias, M., Einevoll, G. T., & Diesmann, M. (2012). Decorrelation of Neural-Network Activity by Inhibitory Feedback. *PLOS Computational Biology* *8*, e1002596.
- Tononi, G. & Cirelli, C. (2014). Sleep and the Price of Plasticity: From Synaptic and Cellular Homeostasis to Memory Consolidation and Integration. *Neuron* *81*, 12–34.
- Tripathy, S. J., Burton, S. D., Geramita, M., Gerkin, R. C., & Urban, N. N. (2015). Brain-wide analysis of electrophysiological diversity yields novel categorization of mammalian neuron types. *Journal of Neurophysiology* *113*, 3474–3489.
- van Vreeswijk, C. & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Computation* *10*, 1321–1371.
- Wallace, D. J. & Kerr, J. N. (2010). Chasing the cell assembly. *Current Opinion in Neurobiology* *20*, 296–305.
- Yoshimura, Y., Dantzker, J. L. M., & Callaway, E. M. (2005). Excitatory cortical neurons form fine-scale functional networks. *Nature* *433*, 868–873.