J Stat Phys (2009) 136: 565–602
DOI 10.1007/s10955-009-9786-1

How Gibbs Distributions May Naturally Arise from Synaptic Adaptation Mechanisms. A Model-Based Argumentation

B. Cessac · H. Rostro · J.C. Vasquez · T. Viéville

Received: 19 December 2008 / Accepted: 30 June 2009 / Published online: 11 July 2009

Abstract This paper addresses two questions in the context of neuronal networks dynamics, using methods from dynamical systems theory and statistical physics: (i) How to characterize the statistical properties of sequences of action potentials ("spike trains") produced by neuronal networks? and; (ii) what are the effects of synaptic plasticity on these statistics? We introduce a framework in which spike trains are associated to a coding of membrane potential trajectories, and actually, constitute a symbolic coding in important explicit examples (the so-called gIF models). On this basis, we use the thermodynamic formalism from ergodic theory to show how Gibbs distributions are natural probability measures to describe the statistics of spike trains, given the empirical averages of prescribed quantities. As a second result, we show that Gibbs distributions naturally arise when considering "slow" synaptic plasticity rules where the characteristic time for synapse adaptation is quite longer than the characteristic time for neurons dynamics.

Keywords Neurons dynamics · Spike coding · Statistical physics · Gibbs distributions · Thermodynamic formalism

1 Introduction

Spike Trains as a "Neural Code" Neurons activity results from complex and nonlinear mechanisms [28, 30, 39, 46], leading to a wide variety of dynamical behaviours [28, 43]. This activity is revealed by the emission of action potentials or "spikes". While the shape of an action potential is essentially always the same for a given neuron, the succession of spikes emitted by this neuron can have a wide variety of patterns (isolated spikes, periodic spiking,

B. Cessac (✉)
Laboratoire J.A. Dieudonné, U.M.R. C.N.R.S. N° 6621, Université de Nice Sophia-Antipolis, Sophia-Antipolis, France
e-mail: bruno.cessac@inria.fr

B. Cessac · H. Rostro · J.C. Vasquez · T. Viéville
INRIA, 2004 Route des Lucioles, 06902 Sophia-Antipolis, France

bursting, tonic spiking, tonic bursting, etc. . . .) [14, 48, 87], depending on physiological parameters, but also on excitations coming either from other neurons or from external inputs. Thus, it seems natural to consider spikes as "information quanta" or "bits" and to seek the information exchanged by neurons in the structure of spike trains. Doing this, one switches from the description of neurons in terms of membrane potential dynamics, to a description in terms of spike trains. This point of view is used, in experiments, by the analysis of *raster plots*, i.e. the activity of a neuron is represented by a mere vertical bar each time this neuron emits a spike. Though this change of description raises many questions, it is commonly admitted in the computational neuroscience community that spike trains constitute a "neural code".

This raises however other questions. How is "information" encoded in a spike train: rate coding [1], temporal coding [85], rank coding [32, 69], correlation coding [52]? How to measure the information content of a spike train? There is a wide literature dealing with these questions [3, 5, 36, 53, 65–67, 79], which are inherently related to the notion of *statistical characterizations* of spike trains, see [30, 39, 72] and references therein for a review. As a matter of fact, a prior to handle "information" in a spike train is the definition of a suitable probability distribution that matches the empirical averages obtained from measures. Thus, in some sense that we make precise in this paper, the choice of a set of quantities to measure (observables) constrains the form of the probability characterizing the statistics of spike trains.

As a consequence, there is currently a wide debate on the canonical form of these probabilities. While Poisson statistics, based on the mere knowledge of frequency rates, are commonly used with some success in biological experiments [37, 38], other investigation evidenced the role of spikes coincidence or correlations [41, 42] and some people have proposed non Poisson probabilities, such as Ising-like Gibbs distributions, to interpret their data [78, 86]. It is important to note here that beyond the determination of the "right" statistical model there is the aim of identifying which type of information is used by the brain to interpret the spike trains that it receives, coming from different sources. As an example choosing a model where only frequency rates matters amounts to assuming that spikes coming from different neurons are essentially treated independently.

However, determining the form of the probability characterizing the statistics of spike trains is extremely difficult in real experiments. In the present paper, we focus on neural networks *models* considered as dynamical systems, with a good mathematical and numerical control on dynamics. In this context we argue that Gibbs distributions are indeed natural candidates whenever a set of quantities to measure (observables) has been prescribed. As a matter of fact Poisson distributions and Ising-like distributions are specific examples, but, certainly, do not constitute the general case.

Synaptic Plasticity The notion of neural code and information cannot be separated from the capacity of neuronal networks to evolve and adapt by *plasticity* mechanisms, and especially *synaptic plasticity*. The latter occurs at many levels of organization and time scales in the nervous system [8]. It is of course involved in memory and learning mechanisms, but it also alters excitability of brain area and regulates behavioural states (e.g. transition between sleep and wakeful activity). Therefore, understanding the effects of synaptic plasticity on neurons dynamics is a crucial challenge. On experimental grounds, different synaptic plasticity mechanisms have been exhibited from the Hebbian's ones [44] to Long Term Potentiation (LTP) and Long Term Depression (LTD), and more recently to Spike Time Dependent Plasticity (STDP) [7, 61] (see [27, 30, 40] for a review). Modeling these mechanisms requires both a bottom-up and top-down approach.

This issue is tackled, on theoretical grounds, by inferring "synaptic updates rules" or "learning rules" from biological observations [8, 60, 63] and extrapolating, by theoretical or numerical investigations, the effects of such synaptic rule on such neural network *model*. This bottom-up approach relies on the belief that there are "canonical neural models" and "canonical plasticity rules" capturing the most essential features of biology. Unfortunately, this results in a plethora of canonical "candidates" and a huge number of papers and controversies. In an attempt to clarify and unify the overall vision, some researchers have proposed to associate learning rules and their dynamical effects to general principles, and especially to"variational" or "optimality" principles, where some functional has to be maximised or minimised [11, 24, 31, 70, 71, 88, 89]. Therefore, in these "top-down" approaches, plasticity rules "emerge" from first principles. Unfortunately, in most examples, the validations of these theories has been restricted to considering isolated neurons submitted to input spike trains with ad hoc statistics (typically, Poisson distributed with independent spikes [88, 89]).

However, addressing the effect of synaptic plasticity in neuronal networks where dynamics is *emerging* from collective effects and where spikes statistics are *constrained* by this dynamics seems to be of central importance. This is the point of view raised in the present paper, where, again, we focus on models, which are simplifications of real neurons. Even in this case, this issue is subject to two main difficulties. On one hand, one must identify the generic collective dynamical regimes displayed by the model for different choices of parameters (including synaptic weights). On the other hand, one must analyse the effects of varying synaptic weights when applying plasticity rules. This requires to handle a complex interwoven evolution where neurons dynamics depends on synapses and synapses evolution depends on neuron dynamics. The first aspect has been addressed by several authors using mean-field approaches (see e.g. [77] and references therein), "Markovian approaches" [84], or dynamical system theory (see [17] and references therein). The second aspect has, up to our knowledge, been investigated theoretically in only a few examples with Hebbian learning [29, 80, 81] or discrete time Integrate and Fire models with an STDP like rule [82, 83] and is further addressed here.

What the Paper is About To summarize the previous discussion the study of neuronal networks at the current stage is submitted to two central questions:

- How to characterize the statistics of spike trains in a network of neurons?
- How to characterize the effects of synaptic plasticity on this network dynamics and especially on spike trains statistics?

In this paper we suggest that these two questions are closely entangled and must both be addressed in the same setting. In order to have a good control on the mathematics we mainly consider the case of neural networks models where one has a full characterization of the generic dynamics [16, 18]. Thus, our aim is not to provide general statements about biological neural networks. We simply want to have a good mathematical control of what is going on in specific models, with the hope that this analysis should shed some light on what happens (or *does not* happen) in "real world" neural systems. However, though part of the results used here are rigorous, this work relies also on "working assumptions" that we have not been able to check rigorously. These assumptions, which are essentially used to apply the standard theorems in ergodic theory and thermodynamic formalism, provide a logical chain which drives us to important conclusions that could be checked in experimental data.

The paper is organised as follows. In Sect. 2 we introduce a framework in which spike trains are associated to a coding of membrane potential trajectories, and actually, constitute a symbolic coding in explicit examples. On this basis, we show how Gibbs distributions are

natural probability measures to describe the statistics of spike trains, given the data of known empirical averages. Several authors have discussed the relevance of Gibbs distribution in the context of Hopfield model (see [2] and references therein) and more recently to spike trains [55, 90]. A breakthrough has been made in [78, 86]. Actually, our approach has been greatly influenced by these two papers, though our results hold in a wider context. In Sect. 3, we discuss the effect of synaptic adaptation when considering "slow" synaptic plasticity rules where the characteristic time for synapse adaptation is quite a bit longer that the characteristic time for neurons dynamics. These rules are formulated in the context of thermodynamic formalism, where we introduce a functional, closely related to thermodynamic potentials like free energy in statistical physics, and called "topological pressure" in ergodic theory. In this setting we show that the variations of synaptic weights leads to variation of the topological pressure that can be smooth ("regular periods"), or singular ("phase transitions"). Phase transitions are in particular associated to a change of "grammar" inducing modifications in the set of spike trains that the dynamics is able to produce. We exhibit a functional, closely related to the topological pressure, that decreases during regular periods. As a consequence, when the synaptic weights converge to a fixed value, this functional reaches a minimum. This minimum corresponds to a situation where spike trains statistics are characterized by a Gibbs distribution, whose potential can be explicitly written. An example with numerical computations is presented in Sect. 4.

2 Neuron Dynamics

2.1 Neural State

We consider a set of N neurons. Each neuron i is characterized by its state, X_i, which belongs to some compact set $\mathcal{I} \in \mathbb{R}^M$. M is the number of variables characterizing the state of one neuron (we assume that all neurons are described by the same number of variables). A typical example is an integrate and fire model where $M = 1$ and $X_i = V_i$ is the membrane potential of neuron i and $\mathcal{I} = [V_{min}, V_{max}]$ (see Sect. 2.5). Other examples are provided by conductances based models of Hodgkin-Huxley type[1] [46]. Then $X_i = (V_i, m_i, n_i, h_i)$ where m_i, n_i are respectively the activation variable for Sodium and Potassium channels and h_i is the inactivation variable for the Sodium channel.

The evolution of these N neurons is given by a deterministic dynamical system of form:

$$\mathbf{X}(t+1) = \mathbf{F}_{\boldsymbol{\gamma}}[\mathbf{X}(t)] \tag{1}$$

where $\mathbf{X}(t) = \{X_i(t)\}_{i=1}^N$ represents the dynamical state of a network of N neurons at time t, while time is discrete (for a discussion on time discretisation in spiking neural networks see [18]). Thus $\mathbf{X} \in \mathcal{M} = \mathcal{I}^N$ where $\mathcal{M}$ is the phase space of (1), and $\mathbf{F}_{\boldsymbol{\gamma}}(\mathcal{M}) \subset \mathcal{M}$. The map $\mathbf{F}_{\boldsymbol{\gamma}} : \mathcal{M} \to \mathcal{M}$ depends on a set of parameters $\boldsymbol{\gamma} \in \mathbb{R}^P$. The typical case considered here is $\boldsymbol{\gamma} = (\mathcal{W}, \mathbf{I}^{(ext)})$ where $\mathcal{W}$ is the matrix of synaptic weights and $\mathbf{I}^{(ext)}$ is some external current, assumed to be independent of time in the present paper (see Sect. 2.5 for two explicit examples). Thus $\boldsymbol{\gamma}$ is a point in a $P = N^2 + N$ dimensional space of control parameters.

[1] Note that Hodgkin-Huxley equations are differential equations, while (1) corresponds to a discrete time evolution. We assume that we have discretized time with a time scale that can be arbitrary small. A detailed discussion on these aspects for conductance based IF models has been presented in [18]. Some further comments are given below.

2.2 Natural Partition

Neurons are excitable systems. Namely, neuron i "fires" (emits a spike or action potential), whenever its state X_i belongs to some connected region $\mathcal{P}_1$ of its phase space. Otherwise, it is quiescent ($X \in \mathcal{P}_0 = \mathcal{I} \setminus \mathcal{P}_1$). In Integrate and Fire models neuron i fires whenever its membrane potential V_i exceeds some threshold θ. In this case, the corresponding region is $\mathcal{P}_1 = [\theta, V_{max}]$. In Fitzhugh-Nagumo [34, 35, 64] or Hodgkin-Huxley model [46], the firing corresponds to the crossing of a manifold called the threshold separatrix [28] which separates the phase space into two connected regions $\mathcal{P}_0$ and $\mathcal{P}_1$. For N identical neurons this leads to a "natural partition" $\mathcal{P}$ of the product phase space $\mathcal{M}$ (see Fig. 1). Call $\Lambda = \{0, 1\}^N$, $\boldsymbol{\omega} = (\omega_i)_{i=1}^N \in \Lambda$. Then, $\mathcal{P} = \{\mathcal{P}_{\boldsymbol{\omega}}\}_{\omega \in \Lambda}$, where $\mathcal{P}_{\boldsymbol{\omega}} = \mathcal{P}_{\omega_1} \times \mathcal{P}_{\omega_2} \times \cdots \times \mathcal{P}_{\omega_N}$. Equivalently, if $\mathbf{X} \in \mathcal{P}_{\boldsymbol{\omega}}$, all neurons such that $\omega_i = 1$ are firing while neurons such that $\omega_k = 0$ are quiescent. We call therefore $\boldsymbol{\omega}$ a "spiking pattern".

2.3 Raster Plots

To each initial condition $\mathbf{X} \in \mathcal{M}$ we associate a "raster plot" $\omega = \{\boldsymbol{\omega}(t)\}_{t=0}^{+\infty}$ such that $\mathbf{X}(t) \in \mathcal{P}_{\boldsymbol{\omega}(t)}, \forall t \geq 0$. We write $\mathbf{X} \rightharpoonup \omega$. Thus, ω is the sequence of spiking patterns displayed by the neural network when prepared with the initial condition $\mathbf{X}$. On the other way round, we say that an infinite sequence $\omega = \{\boldsymbol{\omega}(t)\}_{t=0}^{+\infty}$ is *an admissible raster plot* if there exists $\mathbf{X} \in \mathcal{M}$ such that $\mathbf{X} \rightharpoonup \omega$. We call $\Sigma_{\boldsymbol{\gamma}}$ the set of admissible raster plots for the set of parameters $\boldsymbol{\gamma}$.

The dynamics (1) induce a dynamics on the set of raster plot in the following way:

$$\begin{array}{ccc} \mathbf{X} & \xrightarrow{\mathbf{F}_{\boldsymbol{\gamma}}} & \mathbf{F}_{\boldsymbol{\gamma}}(\mathbf{X}) \\ \downarrow & & \downarrow \\ \omega & \xrightarrow{\sigma_{\boldsymbol{\gamma}}} & \sigma_{\boldsymbol{\gamma}}(\omega) \end{array}$$

In this construction $\sigma_{\boldsymbol{\gamma}}$, called the "left shift", shifts the raster plot left-wise at each time step of the dynamics. Thus, in some sense, raster plots provide a code for the orbits of (1). But, the correspondence may not be one-to-one. That is why we use the notation $\rightharpoonup$ instead of $\rightarrow$.

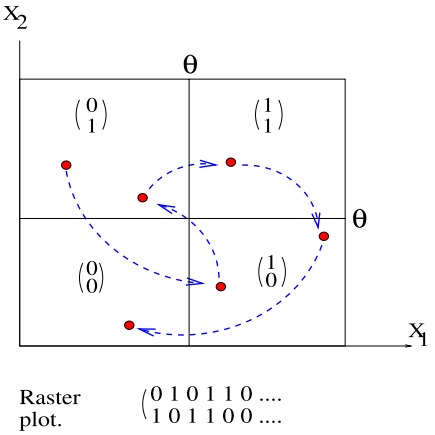

Fig. 1 The phase space $\mathcal{M}$ (here represented for $N = 2$ neurons with constant threshold) is partitioned in 2^N parts $\mathcal{P}_{\boldsymbol{\omega}} = \prod_{i=1}^N \mathcal{P}_{\omega_i}$, $\boldsymbol{\omega}$ being a spiking pattern. In this way, one associates naturally to an orbit of (1) a raster plot

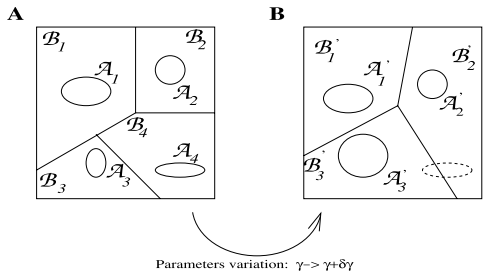

Fig. 2 Schematic illustration of the attractor landscape for neural network models. (**A**) The phase space is partitioned into bounded domains $\mathcal{B}_l$ and for each initial condition in $\mathcal{B}_l$ the initial trajectory is attracted toward an attractor (e.g. a fixed point, a periodic orbit or a more complex attractor) $\mathcal{A}_l$. (**B**) If the parameters (external current, weights) change, the landscape is modified and several phenomena can occur: change in the basins shape, number of attractors, modification of the attractor as for $\mathcal{A}_3$ in this example; A point belonging to $\mathcal{A}_4$ in (**A**), can, after modification of the parameters, converge either to attractor $\mathcal{A}'_2$ or $\mathcal{A}'_3$

Let us introduce the following notation. If we are interested in a prescribed sequence of spiking patterns $\boldsymbol{\omega}(s), \ldots, \boldsymbol{\omega}(t)$, from time s to time t, we denote by $[\omega]_{s,t}$, the set of raster plots whose spiking patterns from time s to time t match this sequence (cylinder set).

2.4 Asymptotic Dynamics

We are here mainly interested in the asymptotic behavior of (1) and set a few notations and notions. The ω-limit set, Ω, is the set of accumulation points of $\mathbf{F}^t_{\boldsymbol{\gamma}}(\mathcal{M})$. Since $\mathcal{M}$ is closed and invariant, we have $\Omega = \bigcap_{t=0}^{\infty} \mathbf{F}^t_{\boldsymbol{\gamma}}(\mathcal{M})$. In dissipative systems (i.e. a volume element in the phase space is dynamically contracted), the ω-limit set typically contains the attractors of the system. A compact set $\mathcal{A} \in \mathcal{M}$ is called an *attractor* for $\mathbf{F}_{\boldsymbol{\gamma}}$ if there exists a neighborhood $\mathcal{U}$ of $\mathcal{A}$ and a time $n > 0$ such that $\mathbf{F}^n_{\boldsymbol{\gamma}}(\mathcal{U}) \subset \mathcal{U}$ and $\mathcal{A} = \bigcap_{t=0}^{\infty} \mathbf{F}^t_{\boldsymbol{\gamma}}(\mathcal{U})$. In all examples considered in the present paper $\mathbf{F}_{\boldsymbol{\gamma}}$ is dissipative and the phase space is divided into finitely many attraction basins each of them containing an attractor (see Fig. 2). Simple examples of attractors are stable fixed points, or stable periodic orbits. More complex attractors such as chaotic attractors can be encountered as well. The attraction basins and attractors change when the parameters $\boldsymbol{\gamma}$ vary. These changes can be smooth (structural stability) or sharp (typically this arises at bifurcations points).

2.5 A Generic Example: Generalized Integrate and Fire Models

2.5.1 Quiescent Stage of the Neuron

Among the various models of neural dynamics, generalized integrate and fire models play a central role, due to their (relative) simplicity, while the biological plausibility is well accepted [40, 74]. A representative class of such models is of form:

$$\frac{dV_k}{dt} = -\frac{1}{\tau_L}(V_k - E_L) + i_k^{(ext)} - i_k^{(syn)}(V_k, t, \{t_j^{(n)}\}_t), \tag{2}$$

defining the evolution of the membrane potential V_k of neuron k. Here $\tau_L = RC \simeq 10$–20 ms is the membrane time-constant related to the membrane resistance and its electric capacity, while $E_L \simeq -80$ mV is the related reversal potential. The term $i_k^{(ext)}$ is an external

"current"[2] assumed to be time constant in this paper. In the end of this section we shall however consider the case where some noise is superimposed upon the constant external current. $t_j^{(n)}$ is the n-th firing time[3] of neuron j and $\{t_j^{(n)}\}_t$ is the list of firing times of all neurons up to time t.

The synaptic currents reads:

$$i_k^{(syn)}(V_k, t, \{t_j^{(n)}\}_t) = (V_k - E^+) \sum_{j\in\mathcal{E}} g_{kj}(t, \{t_j^{(n)}\}_t) + (V_k - E^-) \sum_{j\in\mathcal{I}} g_{kj}(t, \{t_j^{(n)}\}_t),$$

where $E^\pm$ are reversal potential (typically $E^+ \simeq 0$ mV and $E^- \simeq -75$ mV). $\mathcal{E}$ and $\mathcal{I}$ refers respectively to excitatory and inhibitory neurons and the $+$ $(-)$ sign is relative to excitatory (inhibitory) synapses. Note that conductances are always positive thus the sign of the post-synaptic potential (PSP) is determined by the reversal potentials $E^\pm$. At rest ($V_k \sim -70$ mV) the $+$ term leads to a positive PSP while $-$ leads to a negative PSP.

Conductances depend on past spikes via the relation:

$$g_{kj}(t, \{t_j^{(n)}\}_t) = G_{kj} \sum_{n=1}^{M_j(t,\mathbf{V})} \alpha_j(t - t_j^{(n)}).$$

In this equation, $M_j(t, \mathbf{V})$ is the number of times neuron j has fired at time t and $\{t_j^{(n)}\}_t$ is the list of firing times of all neurons up to time t. G_{kj} are positive constants, proportional to the *synaptic efficacy*:

$$\begin{cases} W_{kj} = E^+ G_{kj} & \text{if } j \in \mathcal{E}, \\ W_{kj} = E^- G_{kj} & \text{if } j \in \mathcal{I}. \end{cases} \tag{3}$$

We use the convention $W_{kj} = 0$ if there is no synapse from j to k. Finally, α represents the unweighted shape (called a α-shape) of the post-synaptic potentials. Known examples of α-shapes are $\alpha(t) = Ke^{-t/\tau}H(t)$ or $\alpha(t) = Kte^{-t/\tau}H(t)$, where H is the Heaviside function.

Then, we may write (2) in the form (see [18] for more details):

$$\frac{dV_k}{dt} + g_k V_k = i_k,$$

with:

$$g_k(t, \{t_j^{(n)}\}_t) = \frac{1}{\tau_L} + \sum_{j=1}^{N} g_{kj}(t, \{t_j^{(n)}\}_t),$$

and:

$$i_k(t, \{t_j^{(n)}\}_t) = \frac{E_L}{\tau_L} + \sum_{j\in\mathcal{E}} W_{kj} \sum_{n=1}^{M_j(t,\mathbf{V})} \alpha_j(t - t_j^{(n)}) + \sum_{j\in\mathcal{I}} W_{kj} \sum_{n=1}^{M_j(t,\mathbf{V})} \alpha_j(t - t_j^{(n)}) + i_k^{(ext)}.$$

[2] This is a slight abuse of language since we have divided (2) by the membrane capacity.

[3] For continuous time systems, the firing times of neuron k, for the trajectory $\mathbf{V}$, is defined by:

$$t_k^{(n)}(\mathbf{V}) = \inf\left\{t \mid t > t_k^{(n-1)}(\mathbf{V}),\ V_k(t) \geq \theta\right\},$$

where $t_k^{(0)} = -\infty$.

2.5.2 Fire Regime

The previous equations hold whenever the neuron is quiescent, i.e. whenever membrane potential is smaller than a threshold $\theta > 0$, usually depending on time (to account for characteristics such as refractory period of the neuron) and on the neuronal state. Here, however, we assume that θ is a constant (in fact without loss of generality [40]). When the membrane potential exceeds the threshold value, the neuron "fires" (emission of an action potential or "spike"). The spike shape depends on the model. In the present case, the membrane potential is reset instantaneously to a fixed value $V_{reset} \simeq E_L$, corresponding to the value of the membrane potential when the neuron is at rest. For simplicity we set $V_{reset} = 0$ without loss of generality. This modeling of the spike introduces a natural notion of spike time, but the price to pay is to introduce a discontinuity in dynamics. Moreover, since spike is instantaneous the set of possible spike times occurring within time interval is uncountable in the continuous case. Introducing a minimal time discretisation at scale δ removes this pathology.

2.5.3 Time Discretisation

Assuming that spike times are only known within a precision $\delta > 0$ [18] one shows that the dynamics of membrane potential, discretized at the time scale δ writes:

$$V_k(t+\delta) = \rho_k(t, t+\delta, [\omega]_{0,t}) V_k(t) + J_k(t, [\omega]_{0,t}) \tag{4}$$

where:

$$J_k(t, [\omega]_{0,t}) = \int_t^{t+\delta} i_k(s, [\omega]_{0,t}) \rho_k(s, t+\delta, [\omega]_{0,t}) ds,$$

$$\rho_k(s, t+\delta, [\omega]_{0,t}) = e^{-\int_s^{t+\delta} g_k(s', [\omega]_{0,t}) ds'},$$

$$g_k(t, [\omega]_{0,t}) = \frac{1}{\tau_L} + \sum_{j=1}^{N} G_{kj} \sum_{n=1}^{M_j(t,\omega)} \alpha^{\pm}(s - t_j^{(n)}),$$

where $M_j(t, \omega)$ is the number of spikes emitted by neuron j up to time t, in the raster plot ω. In the sequel we assume that $\alpha^{\pm}(u) = 0$ whenever $|u| \geq \tau_M$, i.e. $g_k(t, [\omega]_{0,t})$ depends on past firing times over a finite time horizon τ_M. We also set $\delta = 1$.

2.5.4 Model I and Model II

A step further, two additional simplifications can be introduced:

1. Assuming that $\rho_k(t, t+\delta, [\omega]_{0,t}) = \rho$ is almost constant, which is equivalent to considering "current" synapses instead of "conductance" synapses, i.e. neglect the dependency of $i_k^{(syn)}$ with respect to V_k;
2. Considering a much simpler form for $J_k(t, [\omega]_{0,t})$, where each connected neuron simply increments the membrane potential during its firing state. This is equivalent to considering the post-synaptic profiles in the previous equations as instantaneous step-wise profiles.

Considering these assumptions leads the following dynamics of membrane potential:

$$\mathbf{F}_{\boldsymbol{\gamma},i}(\mathbf{V}) = \rho V_i\,(1 - Z[V_i]) + \sum_{j=1}^{N} W_{ij} Z[V_j] + I_i^{ext}; \quad i = 1 \ldots N, \tag{5}$$

where $\mathbf{F}_{\boldsymbol{\gamma},i}$ is the i-th component of $\mathbf{F}_{\boldsymbol{\gamma}}$, $Z(x) = 1$ if $x \geq \theta$ and 0 otherwise, both the integrate and firing regime being integrated in this equation. It turns out that this corresponds exactly to the time discretisation of the standard integrate and fire neuron model, which as discussed in e.g., [47] provides a rough but realistic approximation of biological neurons behaviors. This model is called *model I* in the sequel whereas the model based on (4) is called *model II*.

2.5.5 *Genericity results for Models I and II*

The map $\mathbf{F}_{\boldsymbol{\gamma}}$ in models I and II is locally contracting [16, 18], but it is not smooth due to the sharp threshold in neurons firing definition. The *singularity set* where $\mathbf{F}_{\boldsymbol{\gamma}}$ is not continuous is:

$$\mathcal{S} = \{\mathbf{V} \in \mathcal{M} \mid \exists i = 1 \ldots N, \text{ such that } V_i = \theta\}.$$

This is the set of membrane potential vectors such that at least one of the neurons has a membrane potential exactly equal to the threshold. Because of this, dynamics exhibit sensitivity to perturbations for orbits which approach too close the singularity set.

Now, call

$$d(\Omega, \mathcal{S}) = \inf_{\mathbf{V} \in \Omega} \inf_{t \geq 0} \min_{i=1 \ldots N} |V_i(t) - \theta|. \tag{6}$$

Then, the following theorem holds [16, 18].

Theorem 1 *For a generic (in a metric and topological sense) set of parameters* $\boldsymbol{\gamma}$, $d(\Omega, \mathcal{S}) > 0$. *Consequently,*

1. Ω *is composed of finitely many periodic orbits with a finite period, Thus, attractors are generically stable period orbits.*
2. *There is a one-to-one correspondence between membrane potential trajectories and raster plots, except for a negligible set of* $\mathcal{M}$.
3. *There is a finite Markov partition.*

Remarks

1. Note however that, depending on parameters (synaptic weights, external current), some periods can be quite large (*well beyond any accessible computational time*). Also, the number of stable periodic orbits can grow exponentially with the number of neurons.
2. Result 2 means that spike trains provide a symbolic coding for the dynamics.
3. See Sect. 2.6.1 for the implications of result 3.

These models constitute therefore nice examples where one has a good control on dynamics and where one can apply the machinery of thermodynamic formalism (see Appendix). Note that since dynamics is eventually periodic, one may figure out that there is nothing non trivial to say from the respect of spike train statistics. The key point is that period are practically so large that there is any hope to see the periodicity in real experiments. Thus,

one may do "as if" the system were "chaotic". This is a fortiori the case if one superimposes upon the external current a small amount of noise to the external current $\mathbf{I}^{(ext)}$.

Also, though dynamics of model I and II may look rather trivial compared to what is expected from biological neural networks, it might be that such models are able to approximate trajectories of a real neural network, for suitable values of $\boldsymbol{\gamma}$ (e.g. after a suitable adaptation of the synaptic weights) and provided N, the number of neurons, is sufficiently large. This type of properties are currently sought by computational neuroscientist with interesting indications that "IF models are good enough" to approximate biological neurons spike trains [54]. See [73] for a recent illustration of this.

2.6 Spikes Dynamics and Statistics

2.6.1 Grammar

Though dynamics (1) produces raster plots, it is important to remark that it is not able to produce *any* possible sequence of spiking patterns. This fundamental fact is most often neglected in computational neuroscience literature and leads, e.g. to severe overestimation of the system's entropy. Also, it plays a central role in determining spike train statistics.

There are therefore *allowed* and *forbidden* sequences depending on conditions involving the detailed form of $\mathbf{F}_{\boldsymbol{\gamma}}$ and on the parameters $\boldsymbol{\gamma}$ (synaptic weights and external current). Let $\boldsymbol{\omega}, \boldsymbol{\omega}'$ be two spiking patterns. The transition $\boldsymbol{\omega} \to \boldsymbol{\omega}'$ is *legal* or *admissible* if there exists a neural state $\mathbf{X}$ such that neurons fire according to the firing pattern $\boldsymbol{\omega}$ and, at the next time, according to the firing pattern $\boldsymbol{\omega}'$. Equivalently, $\mathbf{F}_{\boldsymbol{\gamma}}(\mathcal{P}_{\boldsymbol{\omega}}) \cap \mathcal{P}_{\boldsymbol{\omega}'} \neq \emptyset$. An admissible transition must satisfy *compatibility conditions* depending on synaptic weights and currents.

As an example, for model I, compatibility conditions write, for all $i = 1 \dots N$:

(a) i is such that $\omega_i = 1$ and $\omega'_i = 1 \Leftrightarrow \sum_{j=1}^{N} W_{ij}\omega_j + I_i^{ext} \geq \theta$,
(b) i is such that $\omega_i = 1$ and $\omega'_i = 0 \Leftrightarrow \sum_{j=1}^{N} W_{ij}\omega_j + I_i^{ext} < \theta$,
(c) i is such that $\omega_i = 0$ and $\omega'_i = 1 \Leftrightarrow \exists \mathbf{X} \in \mathcal{P}_{\boldsymbol{\omega}}$ such that $\rho X_i + \sum_{j=1}^{N} W_{ij}\omega_j + I_i^{ext} \geq \theta$,
(d) i, is such that $\omega_i = 0$ and $\omega'_i = 0 \Leftrightarrow \exists \mathbf{X} \in \mathcal{P}_{\boldsymbol{\omega}}$ such that $\rho X_i + \sum_{j=1}^{N} W_{ij}\omega_j + I_i^{ext} < \theta$.

Note that, while the two first conditions only depend on the partition element $\mathcal{P}_{\boldsymbol{\omega}}$, the two last ones depend on the point $\mathbf{X} \in \mathcal{P}_{\boldsymbol{\omega}}$. Basically, this means that the natural partition is not a Markov partition. The idea is then to find a refinement of the natural partition such that allowed transitions depend only on the partition-elements.

A nice situation occurs when the refined partition is *finite*. Equivalently, compatibility conditions can be obtained by a *finite* set of rules or a finite *grammar*. This is possible if there exists some finite integer r with the following property: Construct blocks of spiking patterns $\boldsymbol{\omega}(0) \cdots \boldsymbol{\omega}(r-1)$ of width r. There are at most 2^{Nr} such possible blocks (and in general quite a bit less *admissible* blocks). Label each of these blocks with a symbol $\alpha \in A$, where A is called an *alphabet*. Each block corresponds to a connected domain $\bigcap_{s=0}^{r-1} \mathbf{F}_{\boldsymbol{\gamma}}^{-s}[\mathcal{P}_{\boldsymbol{\omega}(s)}] \subset \mathcal{M}$ and to a cylinder set $[\omega]_{0,r-1} \subset \Sigma_{\boldsymbol{\gamma}}$. Define a *transition matrix* $G_{\boldsymbol{\gamma}} : A \times A \to \{0,1\}$, depending on $\boldsymbol{\gamma}$, with entries $g_{\alpha\beta}$, such that $g_{\alpha\beta} = 1$ if the transition $\beta \to \alpha$ is admissible, and 0 otherwise. To alleviate the notations in the next equation write $\alpha(t) = [\omega]_{t,t+r-1}$, $\forall t \geq 0$. If

$$\Sigma_{\boldsymbol{\gamma}} = \big\{\omega \mid g_{\alpha(t+1)\alpha(t)} = 1, \forall t \geq 0\big\},$$

then all admissible raster plots are obtained via the finite transition matrix $G_{\boldsymbol{\gamma}}$ which provides the grammar of spiking patterns sequences.

In models I, II such a finite grammar exists for the (generic) set of parameters $\boldsymbol{\gamma}$ such that $d(\Omega, \mathcal{S}) > 0$. Moreover, there are open domains in the space of parameters where the grammar is fixed and is not affected by small variations of synaptic weights or current.

For more general models, it might be that there is no finite grammar to describe all admissible raster plots. When dealing with experiments, one has anyway only access to finite raster plots and the grammar extracted from empirical observations has a finite number of rules (see [21, 22, 25] for nice applications of this idea in the field of turbulence). This empirical grammar is compatible with the dynamics at least for the time of observation.

The grammar depends on $\boldsymbol{\gamma}$. Assume now that we are continuously varying $\boldsymbol{\gamma}$. This corresponds to following some path in the space of control parameters. Then, this path generically crosses domains where grammar is fixed, while at the boundary of these domains, there is a grammar modification, and thus a change in the set of admissible raster plots that dynamics is able to produce. Synaptic adaptation, as described in Sect. 3, corresponds precisely to this situation. Thus, we expect that the neural network exhibits "regular" periods of adaptation where spike trains before and after synaptic changes, correspond to the same grammar. Between these regular periods, sharp changes in raster plots structure are expected, corresponding somehow to displaying new transitions and new rules in the spike code. Actually, the ability of neural networks to display, after adaptation, a *specific* subset of admissible sequences, provided by a specific grammar, is an important issue of this paper.

2.6.2 Spike Responses of Neurons

Neurons respond to excitations or stimuli by finite sequences of spikes. In model I and II a stimulus is typically the external current but more general forms (spike trains coming from external neurons) could be considered as well. Thus, the dynamical response R of a neuronal network to a stimuli S (which can be applied to several neurons in the network), is a sequence $\boldsymbol{\omega}(t) \dots \boldsymbol{\omega}(t+n)$ of spiking patterns. "Reading the neural code" means that one seeks a correspondence between responses and stimuli. However, the spike response does not only depend on the stimulus, but also on the network dynamics and therefore fluctuates randomly (note that this apparent randomness is provided by the dynamical evolution and does not require the invocation of an exogenous noise). Thus, the spike response is sought as a conditional probability $P(R|S)$ [72]. Reading the code consists in inferring $P(S|R)$ e.g. via Bayesian approaches, providing a loose dictionary where the observation of a fixed spikes sequences R does not provide a unique possible stimulus, but a set of stimuli, with different probabilities. Having models for conditional probabilities $P(R|S)$ is therefore of central importance. For this, we need a good notion of statistics.

2.6.3 Performing Statistics

These statistics can be obtained in two different ways. Either one repeats a large number of experiments, submitting the system to the same stimulus S, and computes $P(R|S)$ by an average over these experiments. This approach relies on the assumption that the system has the same statistical properties during the whole set of experiments (i.e. the system has not evolved, adapted or undergone bifurcations meanwhile). In our setting this amounts to assuming that the parameters $\boldsymbol{\gamma}$ have not changed.

Or, one performs a time average. For example, to compute $P(R|S)$, one counts the number of times $n(R, T, \omega)$ when the finite sequence of spiking patterns R, appears in a spike train ω of length T, when the network is submitted to a stimulus S. Then, the probability

$P(R|S)$ is estimated by:

$$P(R|S) = \lim_{T\to\infty} \frac{n(R,T,\omega)}{T}.$$

This approach implicitly assumes that the system is in a stationary state.

The empirical approach is often "in-between". One fixes a time window of length T to compute the time average and then performs an average over a finite number $\mathcal{N}$ of experiments corresponding to selecting different initial conditions. In any case the implicit assumptions are essentially impossible to control in real (biological) experiments, and difficult to prove in models. So, they are basically used as "working" assumptions.

To summarize, one observes, from $\mathcal{N}$ repetitions of the same experiment, $\mathcal{N}$ raster plots $\omega_m, m = 1\ldots\mathcal{N}$ on a finite time horizon of length T. From this, one computes experimental averages allowing to estimate $P(R|S)$ or, more generally, to estimate the average value, $\langle\phi\rangle$, of some prescribed observable $\phi(\omega)$. These averages are estimated by:

$$\bar{\phi}^{(\mathcal{N},T)} = \frac{1}{\mathcal{N}T}\sum_{m=1}^{\mathcal{N}}\sum_{t=1}^{T}\phi(\sigma_{\boldsymbol{\gamma}}^t\omega_m). \tag{7}$$

Typical examples of such observables are $\phi(\omega) = \omega_i(0)$ in which case $\langle\phi\rangle$ is the firing rate of neuron i; $\phi(\omega) = \omega_i(0)\omega_j(0)$ then $\langle\phi\rangle$ measures the probability of spike coincidence for neuron j and i; $\phi(\omega) = \omega_i(\tau)\omega_j(0)$ then $\langle\phi\rangle$ measures the probability of the event "neuron j fires and neuron i fires τ time step later" (or sooner according to the sign of τ). In the same way $P(R|S)$ is the average of the indicatrix function $\chi_R(\omega) = 1$ if $\omega \in R$ and 0 otherwise. Note that in (7) we have used the shift $\sigma_{\boldsymbol{\gamma}}^t$ for the time evolution of the raster plot. This notation is more compact and more adapted to the next developments than the classical formula, reading, e.g., for firing rates $\frac{1}{\mathcal{N}T}\sum_{m=1}^{\mathcal{N}}\sum_{t=1}^{T}\phi(\boldsymbol{\omega}_m(t))$.

This estimation depends on T and $\mathcal{N}$. However, one expects that, as $\mathcal{N}, T \to \infty$, the empirical average $\bar{\phi}^{(\mathcal{N},T)}$ converges to the theoretical average $\langle\phi\rangle$, as stated e.g. from the law of large numbers. Unfortunately, one usually does not have access to these limits, and one is lead to extrapolate theoretical averages from empirical estimations. The main difficulty is that these observed raster plots are produced by an underlying dynamics which is usually not explicitly known (as it is the case in experiments) or impossible to fully characterize (as it is the case in most large dimensional neural networks models). Even for models I and II, where one has a full characterization of generic orbits, a numerical generation of raster plots can produce periodic orbits whose period is out of reach. Thus, one is constrained to propose ad hoc statistical models. This amounts to assuming that the underlying dynamics satisfies specific constraints.

2.6.4 Inferring Statistics from an Hidden Dynamics

We follow this track, proposing a generic method to construct statistical models from a prescribed set of observables. For this, we shall assume that the observed raster plots are generated by a *uniformly hyperbolic* dynamical system, where spike trains constitute a symbolic coding for dynamics (see the Appendix for more details). This a technical assumption allowing us to develop the thermodynamic formalism on a safe mathematical ground. Basically, we assume that the observed raster plots can be generated by a dynamical system which is chaotic with exponential correlation decay, and that a finite Markov partition exists, obtained by a refinement of the natural partition.

This is obviously a questionable assumption, but let us give a few arguments in its favor. First, it has been argued and numerically checked in [73] that finite sequence of spiking

patterns, produced by an hidden neural network, including data coming from biological experiments, can be exactly reproduced by a gIF model of type II, by adding hidden neurons and noise to the dynamics. This suggests that gIF models are somehow "dense" in the space of spiking neural networks dynamical systems, meaning that any finite piece of trajectory from a spiking neural network can be approached by a gIF model trajectory. Second, adding noise to model I, II makes them uniformly hyperbolic,[4] though not continuous. So, we expect the theory developed below to apply to model I, II. Now, dealing with spike train coming from a neural network whose mathematical properties are unknown (this is especially the case with biological neural networks), and whose statistical properties are sought, the idea is to do as if these data were generated by a dynamical system like model I or II.

As a matter of fact, the choice of a statistical model always relies on assumptions. Here we make an attempt to formulate these assumptions in a compact way with the widest range of application. These assumptions are compatible with the statistical models commonly used in the literature like Poisson models or Ising like models à la Schneidman and collaborators [78], but lead also us to propose more general forms of statistics. Moreover, our approach incorporates additional elements such as the consideration of neurons dynamics, and the grammar. This last issue is, according to us, fundamental, and, to the best of our knowledge, has never been considered before in this field. Finally, this postulate allows us to propose algorithms that can be applied to real data with an posteriori check of the initial hypotheses [20].

On this basis we propose the following definition. Fix a set ϕ_l, $l = 1 \dots K$, of observables, i.e. functions $\Sigma_{\boldsymbol{\gamma}} \to \mathbb{R}$ which associate real numbers to sequences of spiking patterns. Assume that the empirical average (7) of these functions has been computed, for a finite T and $\mathcal{N}$, and that $\bar{\phi}_l^{(T,\mathcal{N})} = C_l$.

A *statistical model* is a probability distribution ν on the set of raster plots such that:

1. $\nu(\Sigma_{\boldsymbol{\gamma}}) = 1$, i.e. the set of non admissible raster plots has a zero ν-probability.
2. ν is ergodic for the left-shift $\sigma_{\boldsymbol{\gamma}}$ (see Sect. A.1 in the Appendix for a definition).
3. For all $l = 1 \dots K$, $\nu(\phi_l) = C_l$, i.e., ν is compatible with the empirical averages.

Note that item 2 amounts to assuming that statistics are invariant under time translation. On practical grounds, this hypothesis can be relaxed using sliding time windows. This issue is discussed in more details in [20]. Note also that ν depends on the parameters $\boldsymbol{\gamma}$.

Assuming that ν is ergodic has the advantage that one does not have to average *both* over experiments *and* time. It is sufficient to focus on time average for a single raster plot, via the time-empirical average:

$$\pi_\omega^{(T)}(\phi) = \frac{1}{T} \sum_{t=1}^{T} \phi(\sigma_{\boldsymbol{\gamma}}^t \omega)$$

defined within more details in the Appendix, Sect. A.2.

[4] Contraction is uniform in model I. In model II it can be bounded by a constant strictly lower that 1. Introducing noise amounts to adding, for each neuron, an unstable direction where randomness is generated by a chaotic one-dimensional system. One forms in this way a skew product where the unstable fiber is the random generator and the stable one corresponds to the contracting dynamics of membrane potentials (see (4), or (5)). Because of the discontinuity of the map, the main difficulty is to show that, in this extended system, generic points have a local stable manifold of sufficiently large diameter (Cessac and Fernandez, in preparation). See [9, 19] for an application of this strategy in the framework of Self-Organized Criticality.

Item 3 can be generalized as follows.

3′. $d(\pi_\omega^{(T)}, \nu) \to 0$, as $T \to \infty$,

where $d(\mu, \nu)$ is the relative entropy or Kullback-Leibler divergence between two measures μ, ν (see (54) in the Appendix).

Dealing with real or numerical data, it is obviously not possible to extrapolate to $T \to \infty$, but the main idea here is to minimize the "distance" $d(\pi_\omega^{(T)}, \nu)$ between the empirical measure, coming from the data, and the statistical model. Especially, if several models can be proposed then, the "best" one minimizes the Kullback-Leibler divergence (see Theorem 2 in the Appendix). The main advantage is that $d(\pi_\omega^{(T)}, \nu)$ can be numerically estimated using the thermodynamic formalism. Note however that this approach may fail at phase transition points where $d(\mu, \nu) = 0$ does not necessarily imply $\mu = \nu$ [21]. Phase transition can be numerically detected from empirical data [26].

2.6.5 *Gibbs Measures as Statistical Models*

Variational Principle Statistical physics naturally proposes a canonical way to construct a statistical model: "Maximizing the entropy under the constraints $\nu(\phi_l) = C_l$, $l = 1 \dots K$", (see [50] for a beautiful and deep presentation of what became, since then, a "folklore" result). In the context of thermodynamic formalism this amounts to solving the following variational principle (see (43) in the Appendix).

$$P[\boldsymbol{\psi}] = \sup_{\nu \in m^{(inv)}} (h[\nu] + \nu[\boldsymbol{\psi}]),$$

where $m^{(inv)}$ is the set of invariant measures for σ and h the Kolomogorov-Sinai entropy or entropy rate (see Appendix for more details). The "potential" $\boldsymbol{\psi}$ is given by $\boldsymbol{\psi} = \sum_{l=1}^{K} \lambda_l \phi_l$ where the λ_l's are adjustable Lagrange multipliers. A measure $\nu_{\boldsymbol{\psi}}$ which realizes the supremum, i.e.

$$P[\boldsymbol{\psi}] = h\left[\nu_{\boldsymbol{\psi}}\right] + \nu_{\boldsymbol{\psi}}\left[\boldsymbol{\psi}\right],$$

is called, in this context, an "equilibrium state". The function $P[\boldsymbol{\psi}]$ is called the "topological pressure".

Gibbs States Uniformly hyperbolic dynamical systems have equilibrium states. Moreover, in this case, equilibrium states are *Gibbs states* (and vice-versa). A Gibbs state, or Gibbs measure, is a probability measure such that, one can find some constants c_1, c_2 with $0 < c_1 \leq 1 \leq c_2$ such that for all $n \geq 1$ and for all $\omega \in \Sigma$:

$$c_1 \leq \frac{\nu_{\boldsymbol{\psi}}(\omega \in [\omega]_{0,n-1})}{\exp(-nP[\boldsymbol{\psi}] + S^{(n)}\boldsymbol{\psi}(\omega))} \leq c_2$$

(cf. (51) in the Appendix), where $S^{(n)}\boldsymbol{\psi}(\omega) = \sum_{t=0}^{n-1} \boldsymbol{\psi}(\sigma_{\boldsymbol{\gamma}}^t \omega)$. Basically, this means that the probability that a raster plot starts with the bloc $[\omega]_{0,n-1}$ behaves like $\frac{\exp(S^{(n)}\boldsymbol{\psi}(\omega))}{Z_n}$. One recognizes the classical Gibbs form where space translation in lattice system is replaced by time translation (shift $\sigma_{\boldsymbol{\gamma}}^t$) and where the normalization factor Z_n is the partition function. Note that $P[\boldsymbol{\psi}] = \limsup_{n\to\infty} \frac{1}{n} \log Z_n$, so that $P[\boldsymbol{\psi}]$ is the formal analog of a thermodynamic potential (like free energy).

Topological Pressure The topological pressure is a convex function. Moreover, this is the generating function for the cumulants of the Gibbs distribution. Especially, the Lagrange multipliers λ_l can be tuned to a value λ_l^* such that $\nu_{\boldsymbol{\psi}}[\phi_l] = C_l$, (item 3) using:

$$\left.\frac{\partial P[\boldsymbol{\psi}]}{\partial \lambda_l}\right|_{\lambda_l=\lambda_l^*} = C_l. \tag{8}$$

It expresses that the Gibbs state is given the tangent of the pressure at λ_l^* [57]. Note that we have here assumed that the topological pressure is differentiable, namely that we are away from a phase transition. Higher order moments are obtained in the same way, especially second order moments related to the central limit theorem obeyed by Gibbs distributions [12, 13, 57]. It is also possible to obtain averages of more complex functions than moments [51]. The topological pressure is obtained via the spectrum of the Ruelle-Perron-Frobenius operator and can be calculated numerically when $\boldsymbol{\psi}$ has a finite (and small enough) range (see Appendix for more details).

Validating a Gibbs Statistical Model The Kullback-Leibler divergence provides some notion of "distance" between two measures. For μ an invariant measure and $\nu_{\boldsymbol{\psi}}$ a Gibbs measure with a potential $\boldsymbol{\psi}$, both defined on the same set of sequences Σ, one has (see (55) in the Appendix):

$$d\left(\mu, \nu_{\boldsymbol{\psi}}\right) = P[\boldsymbol{\psi}] - \int \boldsymbol{\psi} d\mu - h(\mu).$$

Though $\pi_\omega^{(T)}$ is not invariant, this suggests to use this relation to compare different statistical models (i.e. when several choices of observables ϕ_l are possible) by choosing the one which minimizes the quantity:

$$d(\pi_\omega^{(T)}, \nu_{\boldsymbol{\psi}}) = P[\boldsymbol{\psi}] - \pi_\omega^{(T)}(\boldsymbol{\psi}) - h(\pi_\omega^{(T)}) \tag{9}$$

(see Theorem 2 in the Appendix). The advantage is that this quantity can be numerically estimated [20].

Probability of Spiking Patterns Blocs In this context, the probability of a spiking pattern block $R = [\boldsymbol{\omega}]_{0,n-1}$ of length n corresponding to the response R to a stimuli S "behaves like" (in the sense of (51)):

$$P\left[R|S\right] = \nu\left[\omega \in R|S\right] \sim \frac{1}{Z_n[\boldsymbol{\lambda}^*(S)]} \exp\left[\sum_{l=1}^{K} \lambda_l^*(S) \sum_{t=0}^{n-1} \phi_l(\sigma_{\boldsymbol{\gamma}}^t \omega)\right], \tag{10}$$

where the conditioning is made explicit in the dependence of $\boldsymbol{\lambda}^*$ in the stimulus S. (Typically, referring to model I and II, S is an external current and is incorporated in $\boldsymbol{\gamma}$.) This equation holds as well when $S = 0$ (no stimulus). Obviously, for two different stimuli the probability $P(R|S)$ may drastically change. Indeed, different stimuli imply different parameters $\boldsymbol{\gamma}$ thus a different dynamics and different raster plots, with different statistical weights on a specific bloc R. The grammar itself can also change, in which case R may be forbidden for a stimulus and allowed for another one.

2.7 Examples

2.7.1 Canonical Examples

Firing Rates If $\phi_l(\omega) = \omega_l(0)$, then $\pi_\omega^{(T)}(\phi_l) = r_l$ is the average firing rate of neuron l within the time period T. Then, the corresponding statistical model is a Bernoulli distribution where neuron l has a probability r_l to fire at a given time. The probability that neuron l fires k times within a time delay n is a binomial distribution and the inter-spike interval is Poisson distributed [40].

Spikes Coincidence If $\phi_l(\omega) \equiv \phi_{(i,j)}(\omega) = \omega_i(0)\omega_j(0)$ where, here, the index l is an enumeration for all (non-ordered) pairs (i, j), then the corresponding statistical models has the form of an Ising model, as discussed by Schneidman and collaborators in [78, 86]. As shown by these authors in experiments on the salamander retina, the probability of spike blocs estimated from the "Ising" statistical model fits quite better to empirical date than the classical Poisson model.

Enlarged Spikes Coincidence As a generalization one may consider the probability of co-occurrence of spikes from neuron i and j within some time interval τ. The corresponding functions are $\phi_l(\omega) = \omega_i(0)\omega_j(\tau)$ and the probability of a spike bloc R writes:

$$P[R|S] = \frac{1}{Z_n[\boldsymbol{\lambda}^*(S)]} \exp\left[\sum_{i,j} \lambda^*_{i,j}(S) \sum_{t=0}^{n-1} \omega_i(t)\omega_j(t+\tau)\right].$$

Further generalizations are considered below.

2.8 Conclusion

The main conclusion of this section is that Gibbs measures constitute *optimal* (in the sense of entropy maximization) statistical models whenever a set of observable has been prescribed. It also allows to select a model between several choices by minimizing the Kullback-Leibler divergence between the empirical measure and the statistical models.

In all the examples presented above the statistical model is determined by the choice of an *a priori* form for the potential. In the next section, we explicitly compute the potential in neural networks with synaptic adaptation.

3 Synaptic Plasticity

3.1 Synapses Update as an Integration over Spikes Trains

Synaptic plasticity corresponds to the evolution of synaptic efficacy (synaptic weights). More precisely, in our notations (3), W_{ij} essentially provides the maximal amplitude of the post-synaptic potential induced, at the synapse linking j to i, when neuron j fires a spike. Synaptic weights evolve in time according to the spikes emitted by the pre- and post-synaptic neuron. In other words, the variation of W_{ij} at time t is a function of the spiking sequences of neurons i and j from time $t - T_s$ to time t, where T_s is time scale characterizing

the width of the spike trains influencing the synaptic change. In most examples the synapse update writes:

$$\delta W_{ij}(t+1) = g\left(W_{ij}(t), [\omega_i]_{t-T_s,t}, [\omega_j]_{t-T_s,t}\right), \quad t > T_s, \tag{11}$$

with $[\omega_i]_{t-T_s,t} = [\omega_i(t-T_s)\ldots\omega_i(t)]$. Thus, typically, synaptic adaptation results from an integration of spikes over the time scale T_s.

3.2 Spikes Responses of Synapses

The synaptic variation δW_{ij} is the integrated response of the synapse from neuron j to neuron i when neuron j sends a spike sequence $[\omega_j]_{t-T_s,t}$ and neuron i fires according to $[\omega_i]_{t-T_s,t}$. This response is not a deterministic function, while g is. Thus (11) is an approximation. As a matter of fact, the explicit form of g is usually derived from phenomenological considerations as well as experimental results where synaptic changes can be induced by *specific* simulations conditions, defined through the firing frequency of pre- and post-synaptic neurons [10, 33], the membrane potential of the post-synaptic neuron [4], or spike timing [7, 58, 61] (see [59] for a review). Thus, these results are usually based on a repetition of experiments involving the excitation of pre- and post-synaptic neurons by specific spike trains. The phenomenological plasticity rules derived from these experiments are therefore of *statistical* nature. Namely, they do not tell us what will be the exact changes induced on synapses when this or this spike train is applied to pre- and post-synaptic neuron. Instead, they provide us the *average* synaptic change. Thus, the function $g(W_{ij}, [\omega_i]_{t-T_s,t}, [\omega_j]_{t-T_s,t})$ in (11) is typically a statistical average of the synaptic response when the spike train of neuron j (resp. i) is $[\omega_j]_{t-T_s,t}$ (resp. $[\omega_i]_{t-T_s,t}$), and the actual synaptic weight value is W_{ij}.

In this paper we investigate the effects of those synaptic plasticity rules when the characteristic time scale T_s is quite a bit larger than the time scale of evolution of the neurons. Namely, we consider *slow* adaptation rules. In this situation the synaptic weights update can be written as a function of the empirical average (42). We therefore consider adaptation rules of form:

$$g\left(W_{ij}, [\omega_i]_{t-T_s,t}, [\omega_j]_{t-T_s,t}\right) = \epsilon \pi_\omega^{(T)}\left[\phi_{ij}(W_{ij}, .)\right], \tag{12}$$

where ϵ is a parameter that will be typically small. $\phi_{ij}(W_{ij}, \omega)$ is a function that we now make explicit.

Following [40] canonical form of adaptation rules use an expansion in singlet, pairs, triplets etc of spikes. Formally, one may write a generic form for these rules. Fix $T_s < \infty$ a positive integer. Let $\mathcal{L}$ be the finite set of ordered lists L in $\{-T_s, -T_s+1, \ldots, 0\}$ with the form $L = (t_1, t_2, \ldots, t_L)$ with $-T_s \le t_1 < t_2 < \cdots < t_L \le 0$. For $L \in \mathcal{L}$, $1 \le i \le N$, we call an *i-monomial* a function $m_{i,L}(\omega) = \omega_i(t_1)\ldots\omega_i(t_L)$. We call an (i, j) *polynomial* a function:

$$\phi_{ij}(W_{ij}, \omega) = \sum_{L_1, L_2 \in \mathcal{L}} h_{ijL_1L_2}(W_{ij}) m_{i,L_1}(\omega) m_{j,L_2}(\omega), \tag{13}$$

where $h_{ijL_1L_2}(W_{ij})$ are smooth functions of W_{ij}. They can be constant, linear functions of W_{ij} ("multiplicative" rules) or nonlinear functions allowing for example to constrain W_{ij} within bounded values (e.g. "hard bound" or "soft bounds" rules). The form (13) is the most general form of synaptic adaptation rules considered in this paper, while g has the explicit

form:

$$g\left(W_{ij},[\omega_i]_{t-T_s,t},[\omega_j]_{t-T_s,t}\right)$$
$$=\epsilon\sum_{k,l=0}^{K}\sum_{\substack{-T_s\le t_1<\cdots<t_k\le 0\\ -T_s\le s_1<\cdots<s_l\le 0}} h_{ij;t_1,\ldots,t_k,s_1,\ldots,s_l}(W_{ij})$$
$$\times\,\pi_\omega^{(T)}\left[\omega_i(t+t_1)\ldots\omega_i(t+t_k)\omega_j(t+s_1)\ldots\omega_j(t+s_l)\right]. \tag{14}$$

Though explicit, this formulation is heavy, and we use instead the form (12), (13).

3.3 Examples of Adaptation Rule Instantiation

Hebbian Learning uses firing rates. A typical example corresponds to $g_{ij}(W_{ij},[\omega_i]_{t-T_s,t},[\omega_j]_{t-T_s,t})=\epsilon\frac{1}{T_s}\sum_{s_1,s_2=t-T_s}^{t}(\omega_i(s_1)-r_i(s_1))(\omega_j(s_2)-r_j(s_2))$ (correlation rule) where $r_i(t)=\frac{1}{T_s}\sum_{s=t-T_s}^{t}\omega_i(s)$ is the frequency rate of neuron i in the raster plot ω, computed in the time windows $[t-T_s,t]$.

Spike-Time Dependent Plasticity as derived from Bi and Poo [7] provides the average amount of synaptic variation given the delay between the pre- and post-synaptic spike. Thus, "classical" STDP writes [40, 49]:

$$g\left(W_{ij},[\omega_i]_{t-T_s,t},[\omega_j]_{t-T_s,t}\right)=\frac{\epsilon}{T_s}\sum_{s_1,s_2=t-T_s}^{t}f(s_1-s_2)\omega_i(s_1)\omega_j(s_2) \tag{15}$$

with:

$$f(x)=\begin{cases}A_-e^{\frac{x}{\tau_-}}, & x<0,\ A_-<0;\\ A_+e^{-\frac{x}{\tau_+}}, & x>0,\ A_+>0;\\ 0, & x=0;\end{cases} \tag{16}$$

where the shape of f has been obtained from statistical extrapolations of experimental data. Hence STDP is based on a second order statistics (spikes correlations). There is, in this case, an evident time scale $T_s=\max(\tau_-,\tau_+)$, beyond which f is essentially zero.

"Nearest neighbors" STDP (according to the terminology of [49]) writes:

$$g\left(W_{ij},[\omega_i]_{t-T_s,t},[\omega_j]_{t-T_s,t}\right)=\frac{\epsilon}{T_s}\sum_{s=t-T_s}^{t}f(\tau_j(s)-s)\omega_i(\tau_j(s))\omega_j(s), \tag{17}$$

with $\tau_j(s)=\min_{t,\omega_j(t)=1}|t-s|$.

Generalized STDP As a last example, [40] propose a rule which corresponds to:

$$g(W_{ij},[\omega_i]_{t-T_s,t},[\omega_j]_{t-T_s,t})$$
$$=\epsilon\left[a_1^{pre}\sum_{s=t-T_s}^{t}\omega_j(s)+a_1^{post}\sum_{s=t-T_s}^{t}\omega_i(s)+\sum_{s_1,s_2=t-T_s}^{t}f(s_1-s_2)\omega_i(s_1)\omega_j(s_2)\right]. \tag{18}$$

We capture thus the main standard synaptic adaptation rules with (13).

3.4 Coupled Dynamics

We consider now the following coupled dynamics. Neurons are evolving according to (1). We focus here on *slow* synapses dynamics. Namely, synaptic weights are constant for $T \geq T_s$ consecutive dynamics steps, where T is large. This defines an "adaptation epoch". At the end of the adaptation epoch, synaptic weights are updated according to (11). This has the consequence of modifying neurons dynamics and possibly spike trains. The weights are then updated and a new adaptation epoch begins. We denote by t the update index of neuron states (neuron dynamics) inside an adaptation epoch, while τ indicates the update index of synaptic weights (synaptic plasticity). Call $\mathbf{X}^{(\tau)}(t)$ the state of the neurons at time t within the adaptation epoch τ. Let $W_{ij}^{(\tau)}$ be the synaptic weights from neuron j to neuron at i in the τ-th adaptation epoch. At the end of each adaptation epoch, the neuron dynamics time indexes are reset, i.e. $x_i^{(\tau+1)}(0) = x_i^{(\tau)}(T), i = 1 \dots N$. The coupled dynamics writes:

$$\begin{cases} \mathbf{X}^{(\tau)}(t+1) = \mathbf{F}_{\boldsymbol{\gamma}^{(\tau)}}(\mathbf{X}^{(\tau)}(t)), \\ \delta W_{ij}^{(\tau)} \stackrel{\text{def}}{=} W_{ij}^{(\tau+1)} - W_{ij}^{(\tau)} = g(W_{ij}^{(\tau)}, [\omega_i]_{t-T_s,t}, [\omega_j]_{t-T_s,t}). \end{cases} \tag{19}$$

Recall that $\boldsymbol{\gamma} = (\mathcal{W}, \mathbf{I}^{(ext)})$ (see Sect. 2.1) and $\boldsymbol{\gamma}^{(\tau)}$ is the set of parameters at adaptation epoch τ. In the present setting the external current $\mathbf{I}^{(ext)}$ is kept fixed and only synaptic weights are evolving. Basically, $\mathbf{I}^{(ext)}$ is used as an *external stimulus*.

3.5 Statistical Effects of Synaptic Plasticity

Synaptic plasticity has several prominent effects. First, modifying the synaptic weights has an action on the dynamics resulting, either in smooth changes, or in sharp changes. In the first case, corresponding to parameters variations in a domain where the dynamical system is structurally stable, small variations of the synaptic weights induce smooth variations on the ω-limit set structure (e.g. points are slightly moved) and in the statistics of orbits. The grammar is not changed. On the opposite, when crossing bifurcations manifolds, the slightest change in one synaptic weight value results typically in a drastic reorganization of the ω-limit set structure where attractors and their attraction basin are modified (see Fig. 2 and [29, 80] for an illustration of this in the case of frequency rates neural networks with Hebbian learning). But this can also change the grammar and the set of admissible raster plots. Some forbidden transitions become allowed, some allowed transitions become forbidden. Finally, the spikes train statistics are also modified. Typically, the time-empirical average of the raster plot changes ($\pi^{(T)}_{\omega^{(\tau)}} \to \pi^{(T)}_{\omega^{(\tau+1)}}$) and the corresponding statistical model also evolves.

But synaptic adaptation is not a mere variation of synaptic weights. Indeed, synaptic weights are associated with the coupled dynamics (19), meaning that synaptic changes depends on neurons dynamics, itself depending on parameters. Thus, a synaptic adaptation process corresponds to following a path in the space of parameters $\boldsymbol{\gamma}$; this path is not determined a priori but evolves according to neurons dynamics. Though this path can belong to a unique domain of structural stability, this is not the case in general. At some point during the adaptation, this path crosses a bifurcation manifold inducing sharp changes in the dynamics and in the grammar.[5] Hence, after several changes of this type one can end up

[5]There is here an analogy with phase transitions in statistical physics [6]. Detecting and characterizing these phase transitions from empirical data is possible provided that there are only a few control parameters [26].

with a system displaying raster plots with a structure rather different from the initial situation. These changes depend obviously upon the detailed form of neuron dynamics (1) and upon the synaptic update mechanism (11); they are also conditioned by parameters such as stimuli. We now analise these effects in the light of thermodynamic formalism and Gibbs distributions.

3.5.1 Static Synaptic Weights

Let us first consider the situation where $\delta\mathcal{W} = 0$, corresponding to synaptic weights that do not evolve. Typically, this is the case if synaptic weights matrix converges to an asymptotic value $\mathcal{W}^*$. From (12) this corresponds to:

$$\pi_\omega^{(T)}\left[\phi_{ij}\right] = 0, \quad \forall i, j \in \{1, \dots N\}. \tag{20}$$

This imposes a condition on the average value of ϕ_{ij}. Therefore, from Sect. 2.6.5 *this imposes that the statistical model is a Gibbs measure ν with a potential of form*:

$$\boldsymbol{\psi}^* = \Phi + \boldsymbol{\lambda}^*.\boldsymbol{\phi}, \tag{21}$$

where $\boldsymbol{\psi}^* = (\psi_{ij}^*)_{i,j=1}^N$, $\boldsymbol{\phi} = (\phi_{ij})_{i,j=1}^N$, $\boldsymbol{\lambda}^* = (\lambda_{ij}^*)_{i,j=1}^N$ and $\boldsymbol{\lambda}^*.\boldsymbol{\phi} = \sum_{i,j=1}^N \lambda_{ij}^*\phi_{ij}$. The potential Φ in (21) is such that $\Phi(\omega) = 0$ if ω is admissible and $\Phi(\omega) = -\infty$ if it is forbidden, so that forbidden raster plots have zero probability. This is a way to include the grammar in the potential (see Appendix). The statistical parameters λ_{ij}^*, are given by (8) in Sect. 2.6.5, and, making ϕ_{ij} explicit (see (13)):

$$\left.\frac{\partial P[\psi]}{\partial\lambda_{ij}}\right|_{\boldsymbol{\lambda}=\boldsymbol{\lambda}^*} = \nu_{\boldsymbol{\psi}^*}(\phi_{ij}) = \sum_{L_1,L_2\in\mathcal{L}} h_{ijL_1L_2}(W_{ij}^*)\nu_{\boldsymbol{\psi}^*}\left[m_{i,L_1}m_{j,L_2}\right] = 0. \tag{22}$$

Since $m_{i,L_1}m_{j,L_2}$ are monomials, this equation thus imposes *constraints* on the probability

$$\nu_{\boldsymbol{\psi}^*}\left[\omega_i(t_1)\dots\omega_i(t_{L_1})\omega_j(s_1)\dots\omega_j(s_{L_2})\right],$$

of spikes n-uplets $\omega_i(t_1)\dots\omega_i(t_L)\omega_j(s_1)\dots\omega_j(s_L)$. Note that condition (22) corresponds to an *extremum* for the topological pressure as a function of $\boldsymbol{\lambda}$. Note also that the variational formulation of Gibbs (equilibrium) state writes, in this case $P[\boldsymbol{\psi}^*] = \sup_{\nu\in m^{(inv)}} h[\nu]$ since $\nu(\psi) = 0$. Hence, the corresponding Gibbs measure has *maximal entropy*.

Let us emphasize what we have obtained. The statistical model that fits with the condition (20) in Sect. 2.6.5 is a Gibbs distribution such that the probability of a spin block R of length n is given by:

$$P\left[R|S\right] = \frac{1}{Z_n[\boldsymbol{\lambda}^*(S)]}\exp\left[\sum_{t=1}^n \boldsymbol{\psi}^*(\sigma_\gamma^t\omega)\right],$$

where $\boldsymbol{\psi}^*$ depends on S via the statistical parameters $\boldsymbol{\lambda}^*$ and via Φ (grammar).

When the situation $\delta\mathcal{W} = 0$ corresponds to the asymptotic state for a synaptic adaptation process, this potential provides us the form of the statistical model *after adaptation*, and *integrates all past changes in the synaptic weights*. We now discuss this process within details.

3.6 Variational Formulation of Synaptic Plasticity

3.6.1 Synaptic Weights Update and Related Potentials

Let us now formalize the coupled evolution (19) in the context of thermodynamic formalism. The main idea is to make the assumption that at each adaptation step, $\pi^{(T)}_{\omega^{(\tau)}}$ can be approximated by a Gibbs measure $\nu_{\boldsymbol{\psi}^{(\tau)}}$ with potential $\boldsymbol{\psi}^{(\tau)}$ and topological pressure $P[\boldsymbol{\psi}^{(\tau)}]$. In this case, when T is large, synaptic adaptation writes:

$$\delta W_{ij}^{(\tau)} = \epsilon \nu_{\boldsymbol{\psi}^{(\tau)}}\left[\phi_{ij}(W_{ij}^{(\tau)}, .)\right]. \tag{23}$$

The synaptic update results in a change of parameters $\boldsymbol{\gamma}$, $\boldsymbol{\gamma}^{(\tau+1)} = \boldsymbol{\gamma}^{(\tau)} + \delta\boldsymbol{\gamma}^{(\tau)}$. This induces a variation of the potential $\boldsymbol{\psi}^{(\tau+1)} = \boldsymbol{\psi}^{(\tau)} + \delta\boldsymbol{\psi}^{(\tau)}$ and of the pressure $P[\boldsymbol{\psi}^{(\tau+1)}] = P[\boldsymbol{\psi}^{(\tau)}] + \delta P^{(\tau)}$. We now distinguish two situations both arising when synaptic weights evolve.

3.6.2 Smooth Variations

Let us first assume that these variations are smooth, i.e. one stays inside a domain where dynamics is structurally stable and the grammar $G_{\boldsymbol{\gamma}^{(\tau)}}$ is not modified. We assume moreover that the topological pressure is differentiable with respect to the variation $\delta\boldsymbol{\gamma}^{(\tau)}$.

Let us define:

$$\mathcal{F}^{(\tau)}_{\boldsymbol{\phi}}(\mathcal{W}) = P\left[\boldsymbol{\psi}^{(\tau)} + (\mathcal{W} - \mathcal{W}^{(\tau)}).\boldsymbol{\phi}(\mathcal{W}^{(\tau)})\right] - P\left[\boldsymbol{\psi}^{(\tau)}\right], \tag{24}$$

so that $\mathcal{F}^{(\tau)}_{\boldsymbol{\phi}}(\mathcal{W}^{(\tau)}) = 0$. Note that $\mathcal{F}^{(\tau)}_{\boldsymbol{\phi}}(\mathcal{W})$ is *convex*, due to the convexity of the topological pressure (see (48) in Appendix).

Then, using (49) in Appendix, the adaptation rule (23) can be written in the form:

$$\delta\mathcal{W}^{(\tau)} = \epsilon \nabla_{\mathcal{W}=\mathcal{W}^{(\tau)}} \mathcal{F}^{(\tau)}_{\boldsymbol{\phi}}(\mathcal{W}). \tag{25}$$

Since, in this section, pressure P is assumed to be smooth, one has, using (49), (50):

$$\begin{aligned}&\epsilon\left(P\left[\boldsymbol{\psi}^{(\tau)} + \delta\mathcal{W}^{(\tau)}.\boldsymbol{\phi}(\mathcal{W}^{(\tau)})\right] - P\left[\boldsymbol{\psi}^{(\tau)}\right]\right)\\ &\quad = \delta\mathcal{W}^{(\tau)}.\left[I + \frac{\epsilon}{2}\kappa^{(\tau)}\right].\delta\mathcal{W}^{(\tau)} + O(\delta\mathcal{W}^{(\tau)^3}),\end{aligned} \tag{26}$$

where $\kappa^{(\tau)}$ is the tensor with entries:

$$\kappa^{(\tau)}_{i_1,j_1,i_2,j_2} = C^{(\tau)}_{\phi_{i_1,j_1}\phi_{i_2,j_2}}(0) + 2\sum_{t=1}^{+\infty} C^{(\tau)}_{\phi_{i_1,j_1}\phi_{i_2,j_2}}(t); \quad i_1, j_1, i_2, j_2 = 1 \ldots N, \tag{27}$$

and $C^{(\tau)}_{\phi_{i_1,j_1}\phi_{i_2,j_2}}(t) = \nu_{\boldsymbol{\psi}^{(\tau)}}[\phi_{i_1,j_1} \circ \sigma^t_{\boldsymbol{\gamma}^{(\tau)}}\phi_{i_2,j_2}] - \nu_{\boldsymbol{\psi}^{(\tau)}}[\phi_{i_1,j_1}]\nu_{\boldsymbol{\psi}^{(\tau)}}[\phi_{i_2,j_2}]$ is the correlation function of $\phi_{i_1,j_1}, \phi_{i_2,j_2}$ for the measure $\nu_{\boldsymbol{\psi}^{(\tau)}}$. Using the explicit form (13) of ϕ_{ij} one can see that $\kappa^{(\tau)}$ is a sum of time correlations between uplets of spikes. This is a version of the fluctuation-dissipation theorem where the response to a smooth variation of the potential $\boldsymbol{\psi}^{(\tau)}$ is given in terms of a series involving the time correlations of the perturbation [75, 76]. This series converges provided that dynamics is uniformly hyperbolic.

For sufficiently small ϵ, the matrix $I + \frac{\epsilon}{2}\kappa^{(\tau)}$ is positive and:

$$P\left[\boldsymbol{\psi}^{(\tau)} + \delta\mathcal{W}^{(\tau)}.\boldsymbol{\phi}(\mathcal{W}^{(\tau)})\right] \geq P\left[\boldsymbol{\psi}^{(\tau)}\right].$$

It follows that the variation $\delta^{(\tau)}\mathcal{F}_\phi = \mathcal{F}_{\boldsymbol{\phi}}^{(\tau+1)}(\mathcal{W}^{(\tau+1)}) - \mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}(\mathcal{W}^{(\tau+1)})$ is given by:

$$\begin{aligned}\delta^{(\tau)}\mathcal{F}_{\boldsymbol{\phi}} &= P\left[\boldsymbol{\psi}^{(\tau)}\right] - P\left[\boldsymbol{\psi}^{(\tau)} + \delta\mathcal{W}^{(\tau)}.\boldsymbol{\phi}(\mathcal{W}^{(\tau)})\right] \\ &= -\frac{1}{\epsilon}\delta\mathcal{W}^{(\tau)}.\left[I + \frac{\epsilon}{2}\kappa^{(\tau)}\right].\delta\mathcal{W}^{(\tau)} - O(\delta\mathcal{W}^{(\tau)^3}), \qquad (28)\\ \delta^{(\tau)}\mathcal{F}_{\boldsymbol{\phi}} &= -\epsilon\nu_{\boldsymbol{\psi}^{(\tau)}}\left[\boldsymbol{\phi}\right]\left[I + \frac{\epsilon}{2}\kappa^{(\tau)}\right].\nu_{\boldsymbol{\psi}^{(\tau)}}\left[\boldsymbol{\phi}\right] - O(\delta\mathcal{W}^{(\tau)^3}).\end{aligned}$$

This variation is therefore *negative* when ϵ is sufficiently small.

We come therefore to the following important conclusion. The adaptation rule (12) is a *gradient* system where the function $\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}$ *decreases* when iterating synaptic adaptation rules. Were the transition $\tau \to \tau + 1$ to be smooth for all τ, would $\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}$ reach a minimum[6] at some $\mathcal{W}^*$ as $\tau \to \infty$. Such a minimum corresponds to $\nabla_{\mathcal{W}^*}\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)} = 0$, thus to $\delta\mathcal{W} = 0$ according to (25). Hence, this minimum corresponds to a *static distribution* for the synaptic weights. Therefore, according to Sect. 3.5.1 the potential $\boldsymbol{\psi}^{(\tau)}$ would converge to the potential (21) as $\tau \to +\infty$. However, we cannot expect these transitions to be smooth for all τ.

3.6.3 Singular Variations

As we saw, during the synaptic adaptation process, the corresponding path in the space of parameters usually crosses bifurcations manifold inducing sharp changes in the dynamics. At those points pressure may not be smooth corresponding to a phase transition. These changes are not necessarily easy to detect numerically, though algorithms allowing to detect phase transitions from finite samples exist, based on rigorous results [26]. Indeed, searching "at blind" for all possible kind of phase transitions in so large dimensional dynamical systems, without any idea of what can happen, seems to be desperate, even for model I and II, especially when thinking of what can already happen in one dimensional systems [6]. So we shall not discuss within more details this aspect, focusing on grammar changes.[7]

Indeed, from the point of view of neural network analysis, grammar changes implies changes in the type of raster plots that the network is able to produce. An interesting situation occurs when the set of admissible raster plots obtained after adaptation belongs to $\Sigma_{\boldsymbol{\gamma}^{(\tau)}} \cap \Sigma_{\boldsymbol{\gamma}^{(\tau+1)}}$. In this case, adaptation plays the role of a *selective mechanism* where the set of admissible raster plots, viewed as a neural code, is gradually reducing, producing after n steps of adaptation a set $\bigcap_{m=1}^{n}\Sigma_{\boldsymbol{\gamma}^{(m)}}$ which can be rather small. This has been observed by Soula et al. in [83] for model I with Spike Time Dependent Plasticity, though not analyzed in those terms. If we consider the situation where (1) is a neural network submitted to some

[6] Additional constraints are required ensuring that $\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}$ does not tend to $-\infty$. Typically, such constraints amount to bound the synaptic weights variation (soft or hard-bounds rules) by a suitable choice of functions $h_{ijL_1L_2}$ in (13).

[7] We actually conjecture that the only possible phase transitions in model I and II are grammar changes occurring when crossing the set where $d(\Omega, \mathcal{S}) = 0$ (see (6)).

stimulus, where a raster plot ω encodes the spike response to the stimulus, then $\Sigma_{\boldsymbol{\gamma}}$ is the set of all possible raster plots encoding this stimulus. Adaptation results in a reduction of the possible coding, thus reducing the variability in the possible responses.

3.6.4 Evolution of the Potential

Iterating the adaptation process, one expects to have periods of smooth variations, punctuated by sharp transitions where potential and grammar change. We qualify them under the generic name of "phase transitions" without further specifications. We write $\tau \equiv (l, n)$ where l indexes the phase transition and n the number of epoch since the last phase transition. The adaptation process corresponds now to a sequence $\boldsymbol{\psi}^{(\tau)} \equiv \boldsymbol{\psi}^{(l,n)}$ of Gibbs potentials. Thus, $\boldsymbol{\psi}^{(l,0)}$ is the potential arising just after the l-th transition, with the convention that $\boldsymbol{\psi}^{(l,0)}(\omega) = -\infty$ if ω is forbidden, such that $\boldsymbol{\psi}^{(l,0)}$ characterizes the grammar after the l-th phase transition. We call a *regular period* the succession of adaptation epochs between two phase transitions.

By definition, during a regular period the synaptic update $\mathcal{W}^{(\tau+1)} = \mathcal{W}^{(\tau)} + \delta\mathcal{W}^{(\tau)}$ induces a smooth variation of the potential $\boldsymbol{\psi}^{(\tau+1)} = \boldsymbol{\psi}^{(\tau)} + \delta\boldsymbol{\psi}^{(\tau)}$, and the variational principle (43) selects a new Gibbs measure $\nu_{\boldsymbol{\psi}^{(\tau+1)}}$ with a potential:

$$\boldsymbol{\psi}^{(\tau+1)} = \boldsymbol{\psi}^{(\tau)} + \delta\boldsymbol{\psi}^{(\tau)}. \tag{29}$$

It follows that:

$$\boldsymbol{\psi}^{(l,n)} = \boldsymbol{\psi}^{(l,0)} + \sum_{m=0}^{n} \delta\psi^{(\tau_l+m)}, \tag{30}$$

where τ_l is the epoch where the l-th phase transition aroused. The explicit form of $\delta\boldsymbol{\psi}^{(\tau)}$ (as well as $\boldsymbol{\psi}^{(\tau)}$) depends on the detailed form of the dynamics, and there is little hope to determine it in general, except when the adaptation process converges to a solution with $\delta\mathcal{W} = 0$ (see Sect. 3.5.1). According to Sect. 3.6.2, the function $\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}$ (see (24)) decreases during regular periods.

When a phase transition occurs, the relation (29) does not hold anymore. We now write:

$$\boldsymbol{\psi}^{(l+1,0)} = \boldsymbol{\psi}^{(l,n_l)} + \delta\boldsymbol{\psi}_{reg}^{(l+1,0)} + \delta\boldsymbol{\psi}_{sing}^{(l+1,0)} \tag{31}$$

where n_l is the last epoch before the phase transition $l+1$. $\delta\boldsymbol{\psi}_{reg}^{(l+1,0)}$ contains the regular variations of the potential while $\delta\boldsymbol{\psi}_{sing}^{(l+1,0)}$ contains the singular variations corresponding to a change of grammar.

We end up with the following picture. During regular periods the grammar does not evolve, the potential changes according to (30), (31), and the function $\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}$ decreases. Then, there is a sharp change in the grammar and the potential. The function $\mathcal{F}_{\boldsymbol{\phi}}^{(\tau)}$ may also have singularities and may sharply increase. If the adaptation rule converges, the potential (30) converges to $\boldsymbol{\psi}^* = \Phi + \boldsymbol{\lambda}^*.\boldsymbol{\phi}$ (see (21)). This potential Φ characterizes the grammar and the statistical weight of allowed raster plots is given by $\boldsymbol{\lambda}^*.\boldsymbol{\phi}$. The potential $\boldsymbol{\psi}^*$ contains all changes in the grammar and statistics arising during the evolution (30, 31). Thus it contains the history of the system and the evolution of the grammar.

4 A Numerical Example

4.1 Model

4.1.1 Adaptation Rule

As an example we consider an adaptation rule inspired from (15) with an additional term $r_d W_{ij}^{(\tau)}$, $-1 < r_d < 0$, corresponding to passive LTD.

$$\delta W_{ij}^{(\tau)} = \epsilon \left[r_d W_{ij}^{(\tau)} + \frac{1}{T} \sum_{t=T_s}^{T+T_s} \omega_j^{(\tau)}(t) \sum_{u=-T_s}^{T_s} f(u) \omega_i^{(\tau)}(t+u) \right], \tag{32}$$

where $f(x)$ is given by (16) and with:

$$T_s \stackrel{\text{def}}{=} 2\max(\tau_+, \tau_-).$$

Set:

$$\begin{aligned} S_i(\omega) &= \sum_{u=-T_s}^{T_s} f(u)\omega_i(u), \\ H_{ij}(\omega) &= \omega_j(0) S_i(\omega), \end{aligned} \tag{33}$$

$\mathbf{H} = \{H_{ij}\}_{i,j=1}^N$, and:

$$\phi_{ij}(W_{ij}, \omega) = r_d W_{ij} + H_{ij}(\omega),$$

with $\boldsymbol{\phi} = \{\phi_{ij}\}_{i,j=1}^N$, where ϕ_{ij} is a finite range potential with range $2T_s$. Then (32) has the form (12), $\delta W_{ij}^{(\tau)} = \epsilon \pi_{\omega^{(\tau)}}^{(T)} \phi_{ij}(W_{ij}, .)]$.

4.1.2 Effects of the Adaptation Rule on the Synaptic Weights Distribution

The term $S_i(\omega)$ (see (33)) can be either negative, inducing Long Term Depression, or positive inducing Long Term Potentiation. In particular, its average with respect to the empirical measure $\pi_{\omega^{(\tau)}}^{(T)}$ reads:

$$\pi_{\omega^{(\tau)}}^{(T)}(S_i) = \eta r_i(\tau) \tag{34}$$

where:

$$\eta = \left[A_- e^{-\frac{1}{\tau_-}} \frac{1 - e^{-\frac{T_s}{\tau_-}}}{1 - e^{-\frac{1}{\tau_-}}} + A_+ e^{-\frac{1}{\tau_+}} \frac{1 - e^{-\frac{T_s}{\tau_+}}}{1 - e^{-\frac{1}{\tau_+}}} \right] \tag{35}$$

and where $r_i(\tau) = \pi_{\omega^{(\tau)}}^{(T)}(\omega_i)$ is the frequency rate of neuron i in the τ-th adaptation epoch.

The term η neither depend on ω nor on τ, but only on the adaptation rule parameters $A_-, A_+, \tau_-, \tau_+, T_s$. Equation (34) makes explicit 3 regimes.

- *Cooperative regime.* If $\eta > 0$ then $\pi_{\omega^{(\tau)}}^{(T)}(S_i) > 0$. Then synaptic weights have a tendency to become more positive. This corresponds to a cooperative system [45]. When iterating adaptation, dynamics become trivial with neurons firing at each time step or remaining quiescent forever.

- *Competitive regime.* On the opposite if $\eta < 0$ synaptic weights become negative. This corresponds to a competitive system [45].
- *Intermediate regime.* The intermediate regime corresponds to $\eta \sim 0$. Here no clear cut tendency can be distinguished from the average value of S_i and spikes correlations have to be considered as well.

4.1.3 Static Weights

Thanks to the soft bound term $r_d W_{ij}$ the synaptic adaptation rule admits a static solution given by:

$$W_{ij} = -\frac{\pi^{(T)}_{\omega^{(\tau)}}[\omega_j(0)S_i(\omega)]}{r_d}. \tag{36}$$

Note that this equation can have several solutions.

Using the same decomposition as the one leading to (34) we obtain:

$$\pi^{(T)}_{\omega^{(\tau)}}\left[\omega_j(0)S_i(\omega)\right] = A_- \sum_{u=-T_s}^{-1} e^{\frac{u}{\tau_-}} \pi^{(T)}_{\omega^{(\tau)}}\left[\omega_j(0)\omega_i(u)\right] + A_+ \sum_{u=1}^{T_s} e^{-\frac{u}{\tau_+}} \pi^{(T)}_{\omega^{(\tau)}}\left[\omega_j(0)\omega_i(u)\right].$$

Note that $\omega_j(0)\omega_i(u) \geq 0$, thus the first term is negative and the second one is positive (see (16)). The sign of W_{ij} depend on the parameters A_-, A_+, T_s, but also on the relative strength of the terms $\pi^{(T)}_{\omega^{(\tau)}}[\omega_j(0)\omega_i(u)]$.

4.1.4 Convergence to the Static Weights Solution

The synaptic adaptation rule (32) defines a mapping $\mathcal{Q}$ on the set of synaptic weights:

$$W_{ij}^{(\tau+1)} = W_{ij}^{(\tau)}(1+\epsilon r_d) + \epsilon \pi^{(T)}_{\omega^{(\tau)}}\left[H_{ij}\right] = \mathcal{Q}_{ij}(\mathcal{W}^{(\tau)})$$

with $|1+\epsilon r_d| < 1$, where $\pi^{(T)}_{\omega^{(\tau)}}$ depends on $\mathcal{W}^{(\tau)}$. Thus,

$$\frac{\partial \mathcal{Q}_{ij}}{\partial W_{kl}} = (1+\epsilon r_d)\delta_{ij,kl} + \epsilon \frac{\partial \pi^{(T)}_{\omega^{(\tau)}}[H_{ij}]}{\partial W_{kl}},$$

where $\delta_{ij,kl} = 1$ if $i = k$, $j = l$ and 0 otherwise. The second term is a linear response characterizing the variation of $\pi^{(T)}_{\omega^{(\tau)}}[H_{ij}]$ with respect to small variations of W_{kl}. Approximating $\pi^{(T)}_{\omega^{(\tau)}}$ by a Gibbs distribution with a potential $\boldsymbol{\psi}^{(\tau)}$, it writes $C_{\boldsymbol{\psi}^{(\tau)},H_{ij}}(0) + 2\sum_{t=0}^{+\infty} C_{\boldsymbol{\psi}^{(\tau)},H_{ij}}(t)$ where $C_{\boldsymbol{\psi}^{(\tau)},H_{ij}}(t)$ is the time-t correlation function between $\boldsymbol{\psi}^{(\tau)}$ and H_{ij}. When $\nu_{\boldsymbol{\psi}^{(\tau)}}$ is smooth with respect to $\mathcal{W}$, the derivative is dominated by the first term and $\mathcal{Q}$ is contracting for sufficiently small ϵ. This is however not a sufficient condition for the convergence of (32) because $\mathcal{Q}$ is not continuous everywhere. It has jumps whenever $\boldsymbol{\gamma}$ crosses the boundary of a structural stability domain.

If the static solution is contained in a open ball where $\mathcal{Q}$ is continuous the contraction property ensures the convergence to a static solution for any initial condition in this ball (Brouwer theorem).

In this paper we postulate the convergence and check it numerically.

4.1.5 Spike Train Statistics in a Static Weights Regime

As emphasized in Sects. 2.6.5 and 3.5.1, when the synaptic adaptation rule converges to a fixed point, the corresponding statistical model is a Gibbs measure with a potential:

$$\boldsymbol{\psi}^* = \Phi + \boldsymbol{\lambda}^*.\boldsymbol{\phi},$$

where Φ contains the grammar and $\boldsymbol{\lambda}$ are free statistical parameters. The value $\boldsymbol{\lambda}^*$ of these parameters in the potential $\boldsymbol{\psi}^*$ is determined by the relation:

$$\left.\frac{\partial P[\boldsymbol{\psi}]}{\partial \lambda_{ij}}\right|_{\boldsymbol{\lambda}^*} = r_d W_{ij}^* + \nu_{\boldsymbol{\psi}^*}[H_{ij}] = 0, \quad \forall i, j, \tag{37}$$

where the pressure is given by:

$$P\left[\boldsymbol{\psi}\right] = r_d \boldsymbol{\lambda}.\mathcal{W} + \lim_{T\to\infty} \frac{1}{T} \log \sum_{\omega\in\Sigma^{(T)}} e^{\boldsymbol{\lambda}.S_T\mathbf{H}(\omega)}.$$

This procedure provides us the explicit form of the raster plot probability distribution when the adaptation rule converges. But the price to pay is that we have to determine *simultaneously* the N^2 parameters λ_{ij}^* on which the Gibbs measure depends. Focusing on the joint probability of a small set of neurons (pairs, triplets) this constraint can be relaxed in order to be numerically tractable.

4.2 Numerical Checks

The main goal of Sect. 4 is to provide an example of the theoretical concepts developed in this paper. The emphasis is however not put on numerical results which will be the main topic of a forthcoming paper. Therefore, we focus here on numerical simulations for model I only, since the simulations for model II are quite more computer-time consuming, and we consider only one case of η value corresponding to the intermediate regime defined above.

4.2.1 Implementation

Neurons Dynamics Previous numerical explorations have shown that a Model I- network of N neurons, with synapses taken randomly from a distribution $\mathcal{N}(0, \frac{C^2}{N})$, where C is a control parameter, exhibits a dynamics with very large periods in specific regions of values of the space (ρ, C) [16, 18]. On this basis, we choose $N = 100$, $\rho = 0.95$, $C = 4.0$, $N = 100$. The external current $\mathbf{I}^{(ext)}$ in (5) is given by $I_i^{ext} = 0.06 + 0.01\mathcal{N}(0, 1)$. Note that fixing a sufficiently large average value for this current avoids a situation where neurons stops firing after a certain time ("neural death").

STDP Implementation An efficient implementation of a STDP rule does not only depend on the analytic form of the rule but also on the respective time scales characterizing neurons and synapses evolutions. In "on-line-protocols", where the synaptic changes occur at the same time scale as neurons dynamics the so-called recursive approach appears to be the more appropriated (see [91] and references therein). In our case, we use instead an offline protocol, where we register the dynamics of the system on a long time windows, then compute the STDP modification, then let the system evolve again to its attractor (see Sect. 3). The offline protocols are very expensive in machine-time, especially when using long spike trains to get

a reliable time average $(\pi^{(T)}_{\omega^{(\tau)}}(\phi_{ij}))$. For this reason we need to add some words about our implementation.

We register spike trains in a binary code. Indeed, this is the cheapest way in memory requirements, though it might be expensive for accessing specific spike times. Also, bit-wise operations are faster than their equivalents on other types of data. Finally, there exist very fast methods for computing the number of bits on any specific variable of type of integer. (The faster for large number of iterations is a look-up table of precomputed values, but in-line methods—using parallel methods based on masks—are not so far in terms of performances. For details and speed comparison see the G. Manku website http://infolab.stanford.edu/~manku/bitcount/bitcount.html.) Numerical comparison of this method with the direct one that records the dynamics in Boolean arrays and compute STDP spike by spike shows enormous performance difference growing exponentially as the length of trains increases.

Computation of $\delta^{(\tau)}\mathcal{F}_\phi$ To check the result in Sect. 3.6.2, stating that $\mathcal{F}^{(\tau)}_{\boldsymbol{\phi}}(\mathcal{W})$ is a decreasing function of τ (i.e. $\delta^{(\tau)}\mathcal{F}_{\boldsymbol{\phi}}$ is negative) during regular periods, we use the following method [21].

Fix a potential $\boldsymbol{\psi}$. If T is the length of the experimental raster plot, one divides it into k blocs of length n, such that $T = k \times n$. Call:

$$P_{n,k}(\omega) = \frac{1}{n}\log\left(\frac{1}{k}\sum_{j=0}^{k-1}\exp(S^{(n)}_j(\omega))\right), \tag{38}$$

with $S^{(n)}_j(\omega) = \sum_{t=jn}^{jn+n-1}\boldsymbol{\psi}(\sigma^t\omega)$. Then, the following result can be proved [21]:

$$P[\psi] = \lim_{n\to+\infty}\lim_{k\to+\infty} P_{n,k}(\omega), \tag{39}$$

for $\nu_{\boldsymbol{\psi}}$ almost-every ω. This computation requires the knowledge of $\boldsymbol{\psi}$.

The result in [21] can be extended straightforwardly to the following case. If $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2$ are two potentials for the same grammar, and $\delta\boldsymbol{\psi} = \boldsymbol{\psi}_2 - \boldsymbol{\psi}_1$ then:

$$P[\boldsymbol{\psi}_2 - P[\boldsymbol{\psi}_1] = \lim_{n\to+\infty}\lim_{k\to+\infty}\frac{1}{n}\log\left(\frac{1}{k}\sum_{j=0}^{k-1}\exp\left(\sum_{t=jn}^{jn+n-1}\delta\boldsymbol{\psi}(\sigma^t\omega)\right)\right). \tag{40}$$

Since $\delta^{(\tau)}\mathcal{F}_\phi = P[\boldsymbol{\psi}^{(\tau)}] - P[\boldsymbol{\psi}^{(\tau)} + \delta\mathcal{W}^{(\tau)}.\boldsymbol{\phi}(\mathcal{W}^{(\tau)})]$,

$$\delta^{(\tau)}\mathcal{F}_\phi = -\lim_{n\to+\infty}\lim_{k\to+\infty}\frac{1}{n}\log\left(\frac{1}{k}\sum_{j=0}^{k-1}\exp\left(\sum_{t=jn}^{jn+n-1}\delta\mathcal{W}^{(\tau)}.\boldsymbol{\phi}(\mathcal{W}^{(\tau)}, \sigma^t(\omega^{(\tau)}))\right)\right). \tag{41}$$

Expanding this equation in power series of ϵ one recovers the expansion (26). In particular, subtracting to (41) the leading term (of order $\delta\mathcal{W}^{(\tau)}.\delta\mathcal{W}^{(\tau)}$) one gets an estimation of the "linear response" term $\kappa^{(\tau)}_{i_1,j_1,i_2,j_2}$ in (27). However, on practical grounds, the computation of (41) requires very long time series to reduce significantly the finite fluctuations of the empirical average. Similar problems are encountered in the computation of linear response from other methods [15].

This result holds for any adaptation rule (any polynomial $\boldsymbol{\phi}$). In the case of the adaptation rule (32), this reduces to computing:

$$\delta^{(\tau)}\mathcal{F}_\phi = -r_d \delta\mathcal{W}^{(\tau)}.\mathcal{W}^{(\tau)} - \lim_{n\to+\infty}\lim_{k\to+\infty}\frac{1}{n}\log\left(\frac{1}{k}\sum_{j=0}^{k-1}\exp\left(\sum_{t=jn}^{jn+n-1}\epsilon\delta\mathcal{W}^{(\tau)}.\mathbf{H}(\mathcal{W}^{(\tau)},\sigma^t(\omega^{(\tau)}))\right)\right).$$

4.2.2 Simulation Results

We have run simulations for STDP parameters values: $r_d = 0.99$, $\epsilon = 0.001$, $\tau_+ = 16$, $\tau_- = 32$, $A_+ = 1.0$, and $T_s = 64$, corresponding to standard values [49]. We fix $\eta = -0.01$ corresponding to the intermediate regime discussed in Sect. 4.1.2. This fixes the value of A_- via (35). The length of spike trains is $T = 3941 = 4096 - 128 = 128 \times 32 - 2 \times T_s$, where 32 is the number of bits in long integer, in our (machine-dependent) implementation. An extended description will be published elsewhere. The main results are summarized in Fig. 3.

In this figure (top) we have represented the evolution of the distribution of synaptic weights $\mathcal{W}^{(\tau)}$ (see captions). For this value of η, the histogram evolves to a unimodal distribution with some nodes having strong synaptic weights. In Fig. 3 (center) we depict the raster plots at the beginning and at the end of synaptic adaptation process. In Fig. 3 (bottom-left) we have shown the evolution of the Frobenius norm for $\delta\mathcal{W}^{(\tau)}$ (i.e., $\sum_{i,j}^N |\delta\mathcal{W}_{ij}^{(\tau)}|^2$), which converges to 0. This shows the convergence of the rule in this case. We have also plotted the average weights modification $\frac{1}{N^2}\sum_{i,j}\delta W_{ij}$. Finally in Fig. 3 (bottom-right) is represented $\delta^{(\tau)}\mathcal{F}_\phi$ computed with 512 block of size 512. We see clearly that this quantity is negative as expected from (28) and tends to 0. We have also represented the term $-\frac{1}{\epsilon}\delta\mathcal{W}^{(\tau)}.\delta\mathcal{W}^{(\tau)}$ corresponding to the first order expansion in (28).

5 Conclusion

In this paper, we have considered the questions of characterizing the statistics of spike trains generated by neuronal networks, and how these statistics evolve during synaptic adaptation, from the angle of dynamical systems theory and ergodic theory. We have introduced a framework where notions such as raster plots, empirical averages of observables, and synaptic plasticity rules, coming from neuroscience, can be formalized in the language of dynamical systems. In this context, we propose a strategy to produce optimal statistical models of spike trains, based on the maximization of statistical entropy where expectation of observables must be compatible with their empirical averages. These models are Gibbs measures. Considering adaptation rules with slow dynamics, we also suggest that such Gibbs measures may result from related adaptation rules whenever synaptic weights converge to a fix value. In this case, synaptic weights resume the whole history of the neuronal network contained in the structure of the generating function of cumulants (the topological pressure) and in the structure of allowed/forbidden raster plots (grammar).

To our opinion this theoretical paper is a beginning. Our hope is now to apply this strategy to data coming from biological neuronal networks, while only formal neural networks were considered here. At the present stage there is indeed a crucial issue to be able to characterize spike train statistics and to discriminate statistical models. For example, it is still an open problem to decide, for a given spike train in a given experiment, whether

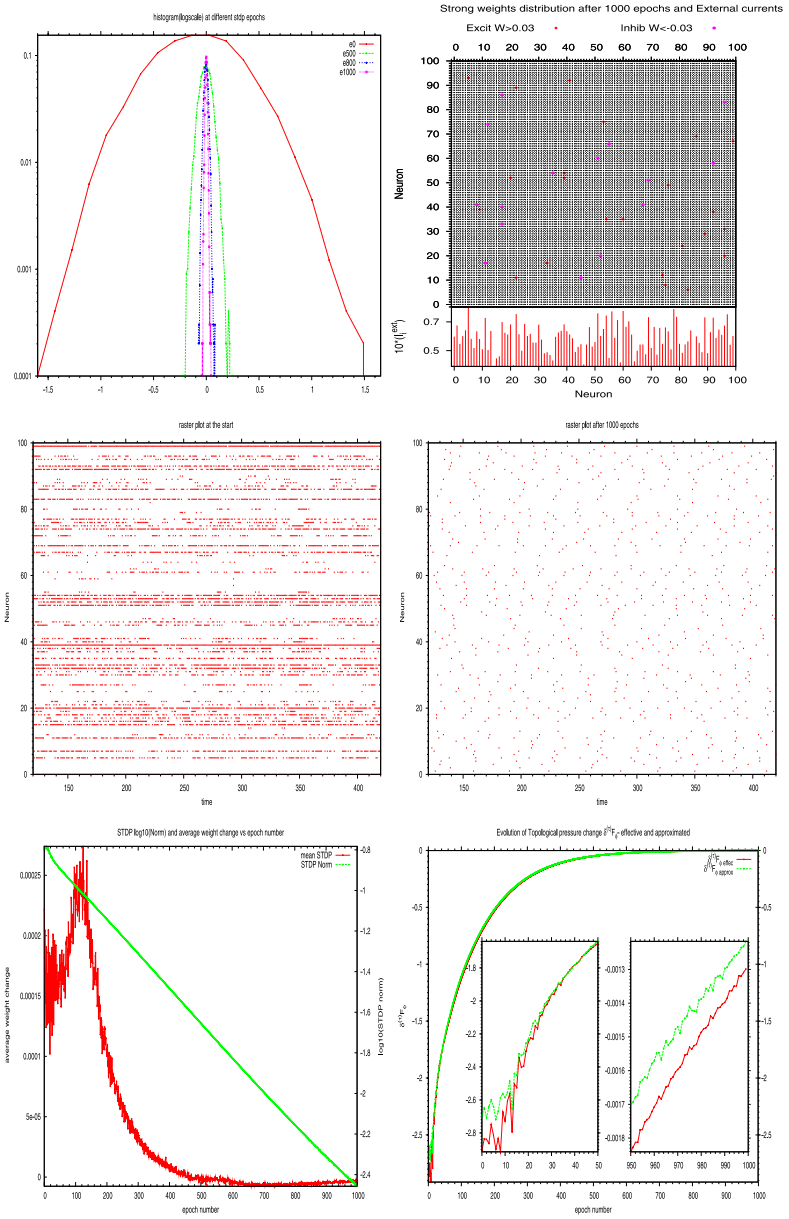

Fig. 3 (Color online) The parameters are $N = 100$, $r_d = -0.4$, $\eta = -0.01$, $\epsilon = 0.01$. *Top*: (*left*) Histogram in log scale of synaptic connectivity in the network after 1000 epochs of STDP; (*right*) map of strongest connections, either inhibitory (*pink dots*) or excitatory (*red diamonds*) (i.e., $|W_{ij}| > 0.03$). We also include at the bottom of this map the external current of each neuron (multiplied by 10). *Middle*: (*left*) A comparison of raster plots at the beginning and (*right*) the end of the synapses evolution. *Bottom*: (*left*) Evolution of the mean synaptic change $\frac{1}{N^2}\sum_{i,j}\delta\mathcal{W}_{ij}^{(\tau)}$ (*red*) and the Frobenius norm for STDP matrix on log scale i.e., $\log_{10}(\sum_{i,j}|\delta\mathcal{W}_{ij}^{(\tau)}|^2)$ (*green*); (*right*) Variations of Topological Pressure (with 2 zoom graphics showing sharp changes) after each epoch as calculated explicitly in (41) with 512 trains of size 512 (*red*), and first order term approximated by $\frac{1}{\epsilon}\delta\mathcal{W}^{(\tau)}.\delta\mathcal{W}^{(\tau)}$ (*green*)

firing rate are sufficient for the statistical characterization (leading to uncorrelated Poisson models) or if higher order orders statistics such as correlations are also necessary. This is more than an academic question. Beyond the choice of such or such statistical model, there are hypotheses on how neuronal network convert stimuli from the external word into an action. Discriminating statistical models is the way to select the best hypothesis.

Our approach opens up the possibility of producing numerical methods for such a discrimination, e.g. based on Kullback-Leibler divergence minimization obtained from (9). Some of these methods are already available as open source code on http://enas.gforge.inria.fr/. They are based on the theory developed in the present paper, relying itself on ergodic theory and thermodynamic formalism. At the present stage, the overwhelming richness of thermodynamic formalism has not really been considered in neuroscience. We hope that this paper will attract the attention of the neuroscience community on this point.

Acknowledgements This work has been supported by the INRIA ARC MACACC and the Doeblin CNRS federation. We thank the Reviewer for helpful remarks, useful references and constructive criticism.

Appendix A

In this appendix we present the thermodynamic formalism material used in the paper. Though all quantities defined below depend on $\boldsymbol{\gamma}$, we have dropped this dependence, which is not central here, in order to alleviate notations. We follow [21–23, 57, 68] in the presentation of Gibbs measures in ergodic theory. In accordance with the body of the text Σ is the set of admissible raster plots and we denote by $\mathcal{B}$ the related Borel sigma-field.

A.1 Ergodic Measures

Fix ϕ a continuous function. The time average of ϕ along the orbit of $\mathbf{X}$ is given by

$$\pi_{\mathbf{X}}(\phi) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \phi(\mathbf{X}(t)).$$

If μ is an invariant measure,[8] then, according to Birkhoff theorem, this limit exists for μ-almost every $\mathbf{X}$. Standard theorems in ergodic theory ensure that (1) has at least one invariant measure [56] but there are typically many. Among them, ergodic measures play a distinguished role. An invariant measure μ is ergodic if any real measurable invariant function is μ-almost surely constant. As a corollary of Birkhoff's theorem $\pi_{\mathbf{X}}(\phi) = \mu(\phi)$, for μ-almost every $\mathbf{X}$, where $\mu(\phi) = \int \phi d\mu$ is the expectation of ϕ with respect to μ. Namely, the empirical average $\pi_{\mathbf{X}}(\phi)$ is equal to the "ensemble" average $\mu(\phi)$ whatever the initial condition, provided it is chosen in the support of μ. Any invariant measure can be written as a convex decomposition of ergodic measures.

A.2 Raster Plots Statistics

Fix $n > 0$, a set of times $t_1, \ldots t_n$, and a set of prescribed spiking patterns $\boldsymbol{\omega}(t_1) \ldots \boldsymbol{\omega}(t_n)$. The set of raster plots $\omega' \in \Sigma$ such that $\boldsymbol{\omega}'(t_k) = \boldsymbol{\omega}(t_k), k = 1 \ldots n$ contains therefore raster

[8] Namely $\mu(\mathbf{F}_{\boldsymbol{\gamma}}{}^{-1} A) = \mathbf{F}_{\boldsymbol{\gamma}}(A)$ for any measurable set $A \subset \mathcal{B}$.

plots where the spiking patterns at prescribed times $t_1 \dots t_n$ are imposed (cylinder set). By (countable) intersection and union of cylinder sets one can generate all possible *events* in Σ, such as "neuron i is firing at time t_1", or "neuron i_1 is firing at time t_1, and neuron i_2 is firing at time t_2, ... and neuron i_n is firing at time t_n", etc. ...

The statistical properties of raster plots are inherited from the statistical properties of orbits of (1), via the correspondence $\nu[C] = \mu[\{\mathbf{X} \rightharpoonup \omega, \omega \in C\}]$, where C is a cylinder set. On practical grounds, these statistical properties are obtained by empirical average. The probability of a cylinder set C is given by:

$$\pi_\omega^{(T)}(C) = \frac{1}{T} \sum_{t=1}^{T} \chi(\sigma^t \in C) \tag{42}$$

where χ is the indicatrix function, while the time average of some function $\phi : \Sigma \to \mathbb{R}$ is given by:

$$\pi_\omega^{(T)}(\phi) = \frac{1}{T} \sum_{t=1}^{T} \phi(\sigma^t \omega).$$

When ν is ergodic the following holds.

$$\nu = \pi_\omega \stackrel{\text{def}}{=} \lim_{T \to \infty} \pi_\omega^{(T)}(.),$$

for ν-almost every ω, where $\pi_\omega(.)$ is called the *empirical measure* for the raster plot ω.

A.3 Thermodynamic Formalism

A.3.1 Working Assumptions

In this paper, we have adopted a pragmatic approach based on the idea that statistical models obtained via the applications of principles in thermodynamic formalism (which are basically the same as those used in statistical physics) could constitute good prototypes for the analysis of real spike trains statistics. In this way, our goal is not to prove theorems but instead to use already proved theorems. However, theorems have hypotheses that we now explicit and whose plausible validity is discussed.

Basically, to apply the machinery described below we need first to assume that raster plots constitute a symbolic coding for the orbits of (1), i.e. there is a one-to-one correspondence between membrane potential trajectories and raster plots, except for a negligible subset of initial conditions. We also assume that symbolic trajectories (raster plots) are generated by a finite grammar. This hypothesis is discussed in Sect. 2.6.1, and it can actually be rigorously established for models I and II without noise. The case with noise is under current investigations. The corresponding transition matrix G defines a sub-shift of finite type or topological Markov chain. Actually, part of the results presented below are extensions of standard results on Markov chains. We also assumed that G is *primitive* i.e. $\exists n_0 > 0$ such that $\forall i, j, \forall n > n_0$ $(G^n)_{ij} > 0$. This amounts to assuming that the dynamics on each attractor is topologically mixing.[9]

[9] A dynamical system $(\mathbf{F}, \mathcal{M})$ is topologically mixing if for any pair of open sets $U, V \subset \mathcal{B}$, there exists $n_0 \geq 0$ such that, $\forall n > n_0$, $\mathbf{F}^n U \cap V \neq \emptyset$.

All these hypotheses hold if the dynamical system generating the raster plots is continuous and uniformly hyperbolic. Following the definition of [68], a map $\mathbf{F} : \mathcal{M} \to \mathcal{M}$ is uniformly hyperbolic on a set Λ if:

1. There exists a splitting of the tangent space in unstable (E^u) and stable (E^s) subspaces such that there exists $C, \mu > 0$ with $\|D\mathbf{F}^t\|_{E^s}, \|D\mathbf{F}^{-t}\|_{E^u} \leq C\mu^t$, $t > 0$.
2. Λ contains a dense orbit.
3. The periodic orbits in Λ are dense (and Λ consists of more than a single closed orbit).
4. There exists an open set $U \supset \Lambda$ such that $\Lambda = \bigcap_{t=-\infty}^{\infty} F^t U$.

Adding noise to model I, II renders them uniformly hyperbolic (see footnote 4 in the text) but they are not continuous due to the singularity of the firing threshold. Hence, additional work is necessary to prove that there is a finite Markov partition (Cessac and Fernandez, in preparation). Throughout this Appendix we assume that all these hypotheses hold.

A.3.2 Potentials

Fix $0 < \Theta < 1$. We define a metric on Σ, the set of raster plots generated by the grammar G, by $d_\Theta(\omega, \omega') = \Theta^p$, where p is the largest integer such that $\boldsymbol{\omega}(t) = \boldsymbol{\omega}(t')$, $0 \leq t \leq p-1$. (The related topology thus structures the fact that raster plots coincide until $t = p-1$).

For a function $\boldsymbol{\psi} : \Sigma \to \mathbb{R}$ and $n \geq 1$ define

$$var_n \boldsymbol{\psi} = \sup \left\{ |\boldsymbol{\psi}(\omega) - \boldsymbol{\psi}(\omega')| : \boldsymbol{\omega}(t) = \boldsymbol{\omega}'(t), 0 \leq i < n \right\}.$$

We denote by $C(\Sigma)$ the space of functions $\boldsymbol{\psi}$ such that $var_n \boldsymbol{\psi} \to 0$ as $n \to \infty$. This is the set of continuous real functions on Σ.

A function $\boldsymbol{\psi} \in C(\Sigma)$ is called a *potential*. A potential is *regular*[10] if $\sum_{n=0}^{+\infty} var_n(\boldsymbol{\psi}) < \infty$. Equivalently, $\boldsymbol{\psi}$ is Hölder. For a positive integer r a *range-r* potential is a potential such that $\boldsymbol{\psi}(\omega) = \boldsymbol{\psi}(\omega')$, if $\boldsymbol{\omega}(t) = \boldsymbol{\omega}'(t)$, $0 \leq t < r$. That is, $\boldsymbol{\psi}$ depends only on the r first spiking patterns $\boldsymbol{\omega}(0), \ldots, \boldsymbol{\omega}(r-1)$. Thus, a finite range potential is necessarily regular. Examples are $\omega_{i_1}(0)$ (range-1 potential), $\omega_{i_1}(0)\omega_{i_2}(k)$ (range-k potential), $\omega_{i_1}(0)\omega_{i_2}(k)\omega_{i_3}(l)$ (range-$\max(k,l)$ potential), etc.... The potential ϕ_{ij} introduced in Sect. 3.2 is a range-T_s potential. More generally, all potential introduced in the paper have finite range.

A specific example of potential, used in this paper, corresponds to integrating the grammar into a potential Φ such that $\Phi(\omega) = 0$ is ω is allowed, and $\Phi(\omega) = -\infty$ otherwise. In this case, however, the potential is not continuous anymore but only upper semi-continuous.

A.3.3 Equilibrium States

Let ν be a σ-invariant measure and $\Sigma^{(n)}$ the set of admissible cylinders of lenght n. Call:

$$h^{(n)}[\nu] = -\sum_{\omega \in \Sigma^{(n)}} \nu\left([\omega]_{0,n-1}\right) \log \nu\left([\omega]_{0,n-1}\right).$$

Then,

$$h[\nu] = \lim_{n \to \infty} \frac{h^{(n)}[\nu]}{n}$$

[10] This condition is analogous to the potential decay ensuring the existence of a thermodynamic limit in statistical mechanics [62, 75]. In the present context it ensures that the Gibbs measure related to the potential $\boldsymbol{\psi}$ is unique and that it is also the unique equilibrium state.

is the entropy of ν. Let $m^{(inv)}$ be the set of invariant measures for σ, then the *pressure* of a potential $\boldsymbol{\psi}$ is defined by:

$$P[\boldsymbol{\psi}] = \sup_{\nu \in m^{(inv)}} \left(h[\nu] + \nu[\boldsymbol{\psi}]\right). \tag{43}$$

This supremum is attained—not necessarily at a unique measure—where:

$$P[\boldsymbol{\psi}] = h\left[\nu_{\boldsymbol{\psi}}\right] + \nu_{\boldsymbol{\psi}}\left[\boldsymbol{\psi}\right].$$

$\nu_{\boldsymbol{\psi}}$ is called an *equilibrium state*, as it maximises some version of the (neg) free energy. Equivalently, a measure ν is an equilibrium state for a potential $\boldsymbol{\psi} \in C(\Sigma)$ if and only if, for all potential $\boldsymbol{\eta} \in C(\Sigma)$:

$$P\left[\boldsymbol{\psi} + \boldsymbol{\eta}\right] - P\left[\boldsymbol{\psi}\right] \geq \nu\left[\boldsymbol{\psi}\right]. \tag{44}$$

This means that ν is a tangent functional for P at $\boldsymbol{\psi}$. In particular $\boldsymbol{\psi}$ has a unique equilibrium state if and only if P is differentiable at $\boldsymbol{\psi}$, i.e. $\lim_{t\to 0} \frac{1}{t}(P[\boldsymbol{\psi} + t\boldsymbol{\phi}] - P[\boldsymbol{\psi}] = \nu(\boldsymbol{\psi}))$, for all $\boldsymbol{\phi} \in C(\Sigma)$.

A.3.4 Properties of the Topological Pressure

We review some important properties of the topological pressure used in this paper. Basically, the topological pressure has the same properties of thermodynamic potentials like free energy. First,

$$P[\boldsymbol{\psi}] = \limsup_{n\to\infty} \frac{1}{n} \log(Z_n(\boldsymbol{\psi})), \tag{45}$$

where $Z_n(\psi)$ is a partition function:

$$Z_n(\boldsymbol{\psi}) = \sum_{\omega \in \Sigma^{(n)}} e^{S^{(n)}\boldsymbol{\psi}(\omega)}, \tag{46}$$

and

$$S^{(n)}\boldsymbol{\psi}(\omega) = \sum_{t=0}^{n-1} \boldsymbol{\psi}(\sigma^t \omega). \tag{47}$$

Also, the topological pressure is a convex functional of the potentials i.e.

$$P\left[\alpha\boldsymbol{\psi}_1 + (1-\alpha)\boldsymbol{\psi}_2\right] \leq \alpha P\left[\boldsymbol{\psi}_1\right] + (1-\alpha)P\left[\boldsymbol{\psi}_2\right]; \quad \alpha \in [0,1]. \tag{48}$$

Like the free energy, the topological pressure is a generating functional (when it is differentiable). For example,

$$\left.\frac{\partial P[\boldsymbol{\psi} + \alpha\boldsymbol{\phi}]}{\partial\alpha}\right|_{\alpha=0} = \nu_{\boldsymbol{\psi}}[\boldsymbol{\phi}], \tag{49}$$

and:

$$\left.\frac{\partial^2 P[\boldsymbol{\psi} + \alpha_1\boldsymbol{\phi}_1 + \alpha_2\boldsymbol{\phi}_2]}{\partial\alpha_1\partial\alpha_2}\right|_{\alpha_1=\alpha_2=0} = C_{\boldsymbol{\phi}_1\boldsymbol{\phi}_2}(0) + 2\sum_{t=0}^{+\infty} C_{\boldsymbol{\phi}_1\boldsymbol{\phi}_2}(t), \tag{50}$$

where:

$$C_{\phi_1\phi_2}(t) = \nu_{\psi}\left(\phi_1 \circ \sigma^t \phi_2\right) - \nu_{\psi}(\phi_1)\nu_{\psi}(\phi_2),$$

is the correlation function of ϕ_1, ϕ_2 at time t.

A.3.5 Gibbs Measure

Assume that the transition matrix of the shift σ is primitive and that ψ is regular. Then, there is a unique ergodic measure ν_{ψ}, called a Gibbs measure, for which one can find some constants c_1, c_2 with $0 < c_1 \leq 1 \leq c_2$ such that for all $n \geq 1$ and for all $\omega \in \Sigma$:

$$c_1 \leq \frac{\nu_{\psi}\left(\omega \in [\omega]_{0,n-1}\right)}{\exp(-nP[\psi] + S^{(n)}\psi(\omega))} \leq c_2, \tag{51}$$

where $S^{(n)}\psi(\omega)$ is given by (47). ν_{ψ} is also the unique equilibrium state.

A.3.6 Ruelle-Perron-Frobenius Operator

The Ruelle-Perron-Frobenius (RPF) operator for the potential ψ, denoted by L_{ψ}, acts on functions $g \in C(\Sigma)$ as $L_{\psi}g(\omega) = \sum_{\omega', \omega = \sigma\omega'} e^{\psi(\omega')}g(\omega')$. This operator has a unique maximal eigenvalue $s = e^{P[\psi]}$ associated to a right eigenfunction b_{ψ} and a left eigenfunction ρ_{ψ} (probability measure) such that $L_{\psi}b_{\psi}(\omega) = sb_{\psi}(\omega)$ and $\int L_{\psi}vd\rho_{\psi} = s\int vd\rho_{\psi}$, for all $v \in C(\Sigma)$. The remaining part of the spectrum is located in a disk in the complex plane, of radius strictly lower than s. Finally, for all $v \in C(\Sigma)$

$$\frac{1}{s^n}L_{\psi}^n v \to b_{\psi}\int vd\rho_{\psi}.$$

The Gibbs measure is $\nu_{\psi} = b_{\psi}\rho_{\psi}$.

Given ψ and knowing its right eigenfunction b_{ψ} and pressure $P[\psi]$ one can construct a new potential:

$$\Psi(\omega) = \psi(\omega) - \log(b_{\psi}(\sigma\omega)) + \log(b_{\psi}(\omega)) - P\left[\psi\right], \tag{52}$$

such that $L_{\Psi}\mathbf{1} = 1$, where $\mathbf{1}$ is the constant function $\mathbf{1}(\omega) = 1$. Such a potential is called "normalised". Its pressure is zero.

If ψ has a finite range-r then the RPF operator reduces to the transition matrix of a $(r+1)$-step Markov chain. Thus, b_{ψ} and ρ_{ψ} can be easily determined. The Gibbs measure is none other than the invariant measure of this chain. This can be used to generate a raster plot distributed according to a Gibbs distribution with a potential ψ. Moreover, for ν_{ψ}-almost-every raster plot ω:

$$\lim_{T\to+\infty} \frac{\pi_{\omega}^{(T)}[j\omega(1)\ldots\omega(r-1)]}{\pi_{\omega}^{(T)}[\omega(1)\ldots\omega(r)]} = e^{\psi(j\omega(1)\ldots\omega(r-1))}. \tag{53}$$

This allows to obtain of ψ numerically [22].

A.3.7 Kullack-Leibler Divergence

There is a last important property. Let μ be an invariant measure and $\nu_{\boldsymbol{\psi}}$ a Gibbs measure with a potential $\boldsymbol{\psi}$, both defined on the same set of sequences Σ. Let

$$d(\mu, \nu_{\boldsymbol{\psi}}) = \limsup_{n\to\infty} \frac{1}{n} \sum_{[\omega]_{0,n-1}} \mu\left([\omega]_{0,n-1}\right) \log\left[\frac{\mu([\omega]_{0,n-1})}{\nu_{\boldsymbol{\psi}}([\omega]_{0,n-1})}\right] \tag{54}$$

be the relative entropy (or Kullack-Leibler divergence) between μ and ν. Then,

$$d\left(\mu, \nu_{\boldsymbol{\psi}}\right) = P[\boldsymbol{\psi}] - \mu(\boldsymbol{\psi}) - h(\mu). \tag{55}$$

If $\mu = \nu$, $d(\mu, \nu) = 0$. The converse may not be true if the potential is not regular.

If a raster ω is typical for the Gibbs measure $\nu_{\boldsymbol{\psi}}$ then one expects that $\pi_\omega^{(T)}$ becomes closer to $\nu_{\boldsymbol{\psi}}$ than any other Gibbs measure (for another potential $\boldsymbol{\psi}'$) as T growths. This provides a criterion to compare two Gibbs measures (and to discrimate between several statistical models). Indeed, the following theorem holds [22].

Theorem 2 *For any pair of distinct*[11] *regular potentials* $\boldsymbol{\phi}, \boldsymbol{\psi}$, *there exists an integer* $N \equiv N(\boldsymbol{\phi}, \boldsymbol{\psi}, \omega)$ *such that, for all* $T \geq N$,

$$d\left(\pi_\omega^{(T)}, \nu_{\boldsymbol{\psi}}\right) < d\left(\pi_\omega^{(T)}, \nu_{\boldsymbol{\phi}}\right) \tag{56}$$

for $\nu_{\boldsymbol{\psi}}$*-almost every* ω.

This says that $d(\pi_\omega^{(T)}, \nu_{\boldsymbol{\psi}})$ becomes, for sufficiently large T, smaller than the Kullback-Leibler divergence between $\pi_\omega^{(T)}$ and any other Gibbs measure.

References

1. Adrian, E., Zotterman, Y.: The impulses produced by sensory nerve endings: Part II: The response of a single end organ. J. Physiol. (Lond.) **61**, 151–71 (1926)
2. Amit, D.J.: Modeling Brain Function—the World of Attractor Neural Networks. Cambridge University Press, Cambridge (1989)
3. Arabzadeh, E., Panzeri, S., Diamond, M.: Deciphering the spike train of a sensory neuron: Counts and temporal patterns in the rat whisker pathway. J. Neurosci. **26**(36), 9216–9226 (2006)
4. Artola, A., Bröcher, S., Singer, W.: Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. Nature **347**(6288), 69–72 (1990)
5. Barbieri, R., Frank, L.M., Nguyen, D.P., Quirk, M.C., Wilson, M.A., Brown, E.N.: Dynamic analyses of information encoding in neural ensembles. Neural Comput. **16**, 277–307 (2004)
6. Beck, C., Schloegl, F.: Thermodynamics of Chaotic Systems: An Introduction. Cambridge University Press, Cambridge (1995)
7. Bi, G., Poo, M.: Synaptic modification by correlated activity: Hebb's postulate revisited. Ann. Rev. Neurosci. **24**, 139–166 (2001)
8. Bienenstock, E.L., Cooper, L., Munroe, P.: Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. J. Neurosci. **2**(1), 32–48 (1982)
9. Blanchard, P., Cessac, B., Krueger, T.: What can one learn about self-organized criticality from dynamical system theory? J. Stat. Phys. **98**, 375–404 (2000)
10. Bliss, T., Gardner-Medwin, A.: Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetised rabbit following stimulation of the perforant path. J. Physiol. **232**, 357–374 (1973)

[11] Non cohomologous.

11. Bohte, S.M., Mozer, M.C.: Reducing the variability of neural responses: a computational theory of spike-timing-dependent plasticity. Neural Comput. **19**(2), 371–403 (2007)
12. Bowen, R.: Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms, vol. 470. Springer, New York (1975)
13. Bowen, R.: Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms, Revised Version. Springer, New York (2008)
14. Brette, R., Gerstner, W.: Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. J. Neurophysiol. **94**, 3637–3642 (2005)
15. Cessac, B.: Does the complex susceptibility of the Hénon map have a pole in the upper-half plane? A numerical investigation. Nonlinearity **20**, 2883–2895 (2007)
16. Cessac, B.: A discrete time neural network model with spiking neurons. Rigorous results on the spontaneous dynamics. J. Math. Biol. **56**(3), 311–345 (2008)
17. Cessac, B., Samuelides, M.: From neuron to neural networks dynamics. EPJ Spec. Top. Top. Dyn. Neural Netw. **142**(1), 7–88 (2007)
18. Cessac, B., Viéville, T.: On dynamics of integrate-and-fire neural networks with adaptive conductances. Front. Neurosci. **2**(2) (2008)
19. Cessac, B., Blanchard, P., Krueger, T., Meunier, J.: Self-organized criticality and thermodynamic formalism. J. Stat. Phys. **115**(516), 1283–1326 (2004)
20. Cessac, B., Vasquez, J., Viéville, T.: Parametric estimation of spike train statistics (2009, submitted)
21. Chazottes, J.: Entropie relative, dynamique symbolique et turbulence. Unpublished doctoral dissertation, Université de Provence—Aix Marseille I (1999)
22. Chazottes, J., Floriani, E., Lima, R.: Relative entropy and identification of Gibbs measures in dynamical systems. J. Stat. Phys. **90**(3–4), 697–725 (1998)
23. Chazottes, J., Keller, G.: Pressure and equilibrium states in ergodic theory. In: Ergodic Theory. Springer, New York (2009)
24. Chechik, G.: Spike-timing-dependent plasticity and relevant mutual information maximization. Neural Comput. **15**(7), 1481–1510 (2003)
25. Collet, P., Galves, A., Lopez, A.: Maximum likelihood and minimum entropy identification of grammars. Random Comput. Dyn. **3**(3/4), 241–250 (1995)
26. Comets, F.: Detecting phase transition for Gibbs measures. Ann. Appl. Probab. **7**(2), 545–563 (1997)
27. Cooper, L., Intrator, N., Blais, B., Shouval, H.: Theory of Cortical Plasticity. World Scientific, Singapore (2004)
28. Cronin, J.: Mathematical Aspects of Hodgkin-Huxley Theory. Cambridge University Press, Cambridge (1987)
29. Daucé, E., Quoy, M., Cessac, B., Doyon, B., Samuelides, M.: Self-organization and dynamics reduction in recurrent networks: stimulus presentation and learning. Neural Netw. **11**, 521–33 (1998)
30. Dayan, P., Abbott, L.F.: Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. MIT Press, Cambridge (2001)
31. Dayan, P., Hausser, M.: In: Thrun, S., Saul, L., Schoelkopf, B. (eds.) Plasticity Kernels and Temporal Statistics, vol. 16. MIT Press, Cambridge (2004)
32. Delorme, A., Perrinet, L., Thorpe, S.: Networks of integrate-and-fire neurons using rank order coding b: spike timing dependent plasticity and emergence of orientation selectivity. Neurocomputing **38**(40), 539–45 (2001)
33. Dudek, S., Bear, M.F.: Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. J. Neurosci. **13**(7), 2910–2918 (1993)
34. FitzHugh, R.: Mathematical models of threshold phenomena in the nerve membrane. Bull. Math. Biophys. **17**, 257–278 (1955)
35. FitzHugh, R.: Impulses and physiological states in models of nerve membrane. Biophys. J. **1**, 445–466 (1961)
36. Gao, Y., Kontoyiannis, I., Bienenstock, E.: Estimating the entropy of binary time series: methodology, some theory and a simulation study. Entropy **10**(2), 71–99 (2008)
37. Georgeopoulos, A.P., Merchant, H., Naselaris, T., Amirikian, B.: Mapping of the preferred direction in the motor cortex. PNAS **104**(26), 11068–11072 (2007)
38. Georgopoulos, A., Kalaska, J., Caminiti, R., Massey, J.: On the relations between the direction of two-dimensional arm movements and cell discharge in primary motor cortex. J. Neurosci. **2**, 1527–1537 (1982)
39. Gerstner, W., Kistler, W.: Spiking Neuron Models. Cambridge University Press, Cambridge (2002)
40. Gerstner, W., Kistler, W.M.: Mathematical formulations of Hebbian learning. Biol. Cybern. **87**, 404–415 (2002)
41. Grammont, F., Riehle, A.: Precise spike synchronization in monkey motor cortex involved in preparation for movement. Exp. Brain Res. **128**, 118–122 (1999)

42. Grammont, F., Riehle, A.: Spike synchronization and firing rate in a population of motor cortical neurons in relation to movement direction and reaction time. Biol. Cybern. **88**, 360–373 (2003)
43. Guckenheimer, J., Labouriau, I.S.: Bifurcation of the Hodgkin-Huxley equations: a new twist. Bull. Math. Biol. **55**, 937–952 (1993)
44. Hebb, D.: The Organization of Behavior: A Neuropsychological Theory. Wiley, New York (1949)
45. Hirsch, M.: Convergent activation dynamics in continuous time networks. Neural Netw. **2**, 331–349 (1989)
46. Hodgkin, A., Huxley, A.: A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. **117**, 500–544 (1952)
47. Izhikevich, E.: Simple model of spiking neurons. IEEE Trans. Neural Netw. **14**(6), 1569–1572 (2003)
48. Izhikevich, E.: Which model to use for cortical spiking neurons? IEEE Trans. Neural Netw. **15**(5), 1063–1070 (2004)
49. Izhikevich, E., Desai, N.: Relating stdp to bcm. Neural Comput. **15**, 1511–1523 (2003)
50. Jaynes, E.: Information theory and statistical mechanics. Phys. Rev. **106**, 620 (1957)
51. Ji, C.: Estimating functionals of one-dimensional Gibbs states. Probab. Theory Relat. Fields **82**(2), 155–175 (1989)
52. Johnson, D.: Sensory discrimination: neural processes preceding discrimination decision. J. Neurophysiol. **4**(6), 1793–1815 (1980)
53. Johnson, D.: Neural population structure and consequences for neural coding. J. Comput. Neurosci. **16**(1), 69–80 (2004)
54. Jolivet, R., Rauch, A., Lescher, H.R., Gerstner, W.: Integrate-and-Fire Models with Adaptation Are Good Enough. MIT Press, Cambridge (2006)
55. Kang, K., Amari, S.I.: Discrimination with spike times and isi distributions. Neural Comput. **20**, 1411–1426 (2008)
56. Katok, A., Hasselblatt, B.: Introduction to the Modern Theory of Dynamical Systems. Kluwer Academic, Dordrecht (1998)
57. Keller, G.: Equilibrium States in Ergodic Theory. Cambridge University Press, Cambridge (1998)
58. Levy, W., Stewart, D.: Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience **8**(4), 791–797 (1983)
59. Malenka, R.C., Nicoll, R.A.: Long-term potentiation—a decade of progress? Science **285**(5435), 1870–1874 (1999)
60. von-der Malsburg, C.: Self-organisation of orientation sensitive cells in the striate cortex. Kybernetik **14**, 85–100 (1973)
61. Markram, H., Lübke, J., Frotscher, M., Sakmann, B.: Regulation of synaptic efficacy by coincidence of postsynaptic ap and epsp. Science **275**, 213 (1997)
62. Meyer, D.: The Ruelle-Araki Transfer Operator in Classical Statistical Mechanics. Lecture Notes in Physics, vol. 123. Springer, Berlin (1980)
63. Miller, K., Keller, J., Stryker, M.: Ocular dominance column development: analysis and simulation. Science **245**(4918), 605–615 (1989)
64. Nagumo, J., Arimoto, S., Yoshizawa, S.: An active pulse transmission line simulating nerve axon. Proc. IRE **50**, 2061–2070 (1962)
65. Nemenman, I., Lewen, G., Bialek, W., de Ruyter van Steveninck, R.: Neural coding of a natural stimulus ensemble: Information at sub-millisecond resolution. PLoS Comput. Biol. **4**, e1000025 (2006)
66. Nirenberg, S., Latham, P.: Decoding neuronal spike trains: how important are correlations. Proc. Nat. Acad. Sci. **100**(12), 7348–7353 (2003)
67. Osbone, L., Palmer, S., Lisberger, S., Bialek, W.: Combinatorial coding in neural populations. arXiv:0803.3837 (2008)
68. Parry, W., Pollicott, M.: Zeta Functions and the Periodic Orbit Structure of Hyperbolic Dynamics, vols. 187–188. Asterisque (1990)
69. Perrinet, L., Delorme, A., Samuelides, M., Thorpe, S.: Networks of integrate-and-fire neuron using rank order coding a: how to implement spike time dependent Hebbian plasticity. Neurocomputing **38** (2001)
70. Rao, R., Sejnowski, T.J.: In: Solla, S., Leen, T., Muller, K. (eds.) Predictive Sequence Learning in Recurrent Neocortical Circuits, vol. 12. MIT Press, Cambridge (1991)
71. Rao, R., Sejnowski, T.J.: Spike-timing-dependent Hebbian plasticity as temporal difference learning. Neural Comput. **13**(10), 2221–2237 (2001)
72. Rieke, F., Warland, D., de Ruyter van Steveninck, R., Bialek, W.: Spikes, Exploring the Neural Code. MIT Press, Cambridge (1996)
73. Rostro-Gonzalez, H., Cessac, B., Vasquez, J., Viéville, T.: Back-engineering of spiking neural networks parameters. J. Comput. Neurosci. (2009, submitted)
74. Rudolph, M., Destexhe, A.: Analytical integrate and fire neuron models with conductance-based dynamics for event driven simulation strategies. Neural Comput. **18**, 2146–2210 (2006)

75. Ruelle, D.: Statistical Mechanics: Rigorous Results. Benjamin, New York (1969)
76. Ruelle, D.: Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. J. Stat. Phys. **95**, 393–468 (1999)
77. Samuelides, M., Cessac, B.: Random recurrent neural networks. Eur. Phys. J., Spec. Top. **142**, 7–88 (2007)
78. Schneidman, E., Berry, M., Segev, R., Bialek, W.: Weak pairwise correlations imply string correlated network states in a neural population. Nature **440**, 1007–1012 (2006)
79. Sinanović, A., Johnson, D.: Toward a theory of information processing. Signal Process. **87**(6), 1326–1344 (2007)
80. Siri, B., Berry, H., Cessac, B., Delord, B., Quoy, M.: Effects of Hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. J. Physiol. Paris **101**(1–3), 138–150 (2007). arXiv:0706.2602
81. Siri, B., Berry, H., Cessac, B., Delord, B., Quoy, M.: A mathematical analysis of the effects of Hebbian learning rules on the dynamics and structure of discrete-time random recurrent neural networks. Neural Comput. **20**(12), 12 (2008). arXiv:0705.3690v1
82. Soula, H.: Dynamique et plasticité dans les réseaux de neurones à impulsions. Unpublished doctoral dissertation, INSA Lyon (2005)
83. Soula, H., Beslon, G., Mazet, O.: Spontaneous dynamics of asymmetric random recurrent spiking neural networks. Neural Comput. **18**, 1 (2006)
84. Soula, H., Chow, C.C.: Stochastic dynamics of a finite-size spiking neural networks. Neural Comput. **19**, 3262–3292 (2007)
85. Theunissen, F., Miller, J.: Temporal encoding in nervous systems: a rigorous definition. J. Comput. Neurosci. **2**, 149–162 (1995)
86. Tkacik, G., Schneidman, E., Berry, M., Bialek, W.: Ising models for networks of real neurons. arXiv:q-bio/0611072 (2006)
87. Touboul, J.: Bifurcation analysis of a general class of nonlinear integrate-and-fire neurons. SIAM J. Appl. Math. **68**(4), 1045–1079 (2008)
88. Toyoizumi, T., Pfister, J.-P., Aihara, K., Gerstner, W.: Generalized Bienenstock-Cooper-Munroe rule for spiking neurons that maximizes information transmission. Proc. Nat. Acad. Sci. **102**, 5239–5244 (2005)
89. Toyoizumi, T., Pfister, J.-P., Aihara, K., Gerstner, W.: Optimality model of unsupervised spike-timing dependent plasticity: synaptic memory and weight distribution. Neural Comput. **19**, 639–671 (2007)
90. Wood, F., Roth, S., Black, M.: Modeling neural population spiking activity with Gibbs distributions. In: Weiss, Y., Schölkopf, B., Platt, J. (eds.) Advances in Neural Information Processing Systems, vol. 18, pp. 1537–1544. MIT Press, Cambridge (2006)
91. Zou, Q.: Modèles computationnels de la plasticité impulsionnelle: synapses, neurones et circuits. Unpublished doctoral dissertation, Université Paris VI (2006)