

## Article

# Thermodynamic Formalism in Neuronal Dynamics and Spike Train Statistics

Rodrigo Cofré <sup>1,\*</sup>, Cesar Maldonado <sup>2</sup>, Bruno Cessac <sup>3</sup>

<sup>1</sup> CIMFAV-Ingemat, Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile; rodrigo.cofre@uv.cl

<sup>2</sup> IPICYT/División de Matemáticas Aplicadas, San Luis Potosí, Mexico; cesar.maldonado@ipicyt.edu.mx

<sup>3</sup> Universite Cote d'Azur, Inria, France, Inria Biovision team and Neuromod Institute

\* Correspondence: rodrigo.cofre@uv.cl

**Abstract:** The Thermodynamic Formalism provides a rigorous mathematical framework to study quantitative and qualitative aspects of dynamical systems. At its core there is a variational principle and corresponding, in its simplest form, to the Maximum Entropy principle, used as a statistical inference procedure to represent, by specific probability measures (Gibbs measures), the collective behaviour of complex systems. This framework has found applications in different domains of science, in particular, has been fruitful and influential in neurosciences. In this article, we review how the Thermodynamic Formalism can be exploited in the field of theoretical neuroscience, as a conceptual and operational tool, to link the dynamics of interacting neurons and the statistics of action potentials from either experimental data or mathematical models. We comment on perspectives and open problems in theoretical neuroscience that could be addressed within this formalism.

**Keywords:** Thermodynamic formalism; neuronal networks dynamics; maximum entropy principle; free energy and pressure; linear response; large deviations, ergodic theory

## 1. Introduction

Initiated by Boltzmann [1,2], the goal of statistical physics was to establish a link between the microscopic mechanical description of interacting particles in a gas or a fluid, and the macroscopic description provided by thermodynamics. Although, this program is, even nowadays, far from being completed [1,3], the work of Boltzmann and his successors opened new avenues of research, not only in physics but also in mathematics. Especially, the term "ergodic" coined by Boltzmann [1], inaugurated an important branch of mathematics that provides a rigorous link between the description of dynamical systems in terms of their trajectories and the description in terms of statistics of orbits, and, more generally, between dynamical systems theory and probability theory. At the core of the ergodic theory there is a set of "natural" dynamically invariant probability measures on the phase space, somewhat generalising the Liouville distribution for conservative systems with strong analogies with Gibbs distributions in statistical physics [4,5]. This strong connection gave, in particular, birth to the so-called Thermodynamic Formalism.

The introduction of the Thermodynamic Formalism occurred in the 1970s, and was primarily due to Yakov Sinai, David Ruelle, and Rufus Bowen [6–8]. The development of the Thermodynamic Formalism initially served to derive rigorous criteria characterising the existence and uniqueness of the Gibbs states in the infinite volume limit. Although Gibbs states and equilibrium states (see section 2.1.1) are naturally defined in finite volume systems, the extension to infinite volume ("thermodynamic limit") is far from straightforward. Indeed, it does not follow from the Carathéodory or Kolmogorov extension theorems [9,10], that the equilibrium states of the infinite volume define a measure, as there is no way to express the marginals associated to an infinite-volume Gibbs measure without making explicit reference to the measure itself [10]. Considering conditional probabilities rather than marginals Dobrushin, Lanford, and Ruelle, led to a different consistency condition that affords to build infinite volume Gibbs measures [5].

In the context of dynamical systems, Sinai, Ruelle, and Bowen were able to connect the theory of hyperbolic (Anosov) dynamical systems to results in statistical mechanics. Indeed, Sinai found an unexpected link between the equilibrium statistical mechanics of spin systems and the ergodic theory of Anosov systems by a codification using Markov partitions (see section 2.1.1 for details). This idea was later extended for a much more general class of hyperbolic systems [8,11,12]. While the Thermodynamic Formalism started as a branch of the rigorous statistical mechanics, nowadays, it is viewed from different communities as a branch of dynamical systems or stochastic processes.

There have been a few attempts to use the Thermodynamic Formalism in a different way other than as a natural mathematical foundation of statistical mechanics, for example, studying population dynamics [13,14], self-organised criticality [15], and statistics of spike trains in neural systems [16–22]. Neuronal networks are biological systems, whose components such as neurons, synapses, ionic channels, ..., are ruled by the laws of physics and written in terms of differential equations, hence, are dynamical systems. On the other hand, because of their large dimensionality, it is natural to attempt to characterise neuronal networks using methods inspired by statistical physics. For example, mean-field methods [23–26], density methods [27] or Fokker-Planck equations [28]. Most neurons produce short electrical pulses, called *action potentials* or *spikes*, and it is widely believed that the collective spike trains emitted by neuronal networks encode information about the underlying dynamics and response to stimuli [29–31]. Researchers have thus devoted a lot of effort to understand the correlations structure in the statistics of spike trains [32–39], (see [40] for a physicists-oriented review).

As spikes can be represented as binary variables, it is natural to attempt and adapt methods and concepts from statistical physics, and more specifically the statistical physics of spin systems. Unfortunately, the community of scientist addressing spike train statistics mostly stuck at the original form of a Ising spin-glass potential (pairwise interactions with random couplings), with some attempt of generalisations [34–36,41], but without making an explicit link between neuronal dynamics and spikes statistics. Indeed, in most of these approaches successive times are independent, thereby neglecting the time correlations and causality one may expect a network of neurons with interactions (exceptions can be found in the following articles [42–44]).

In this review, we illustrate how the Thermodynamic Formalism can be used to study the link between the neuronal dynamics and spike statistics, going far beyond Ising-like models, not only from experimental data, but also from mathematical models of neuronal networks, properly handling causal and memory dependent interactions between neurons and their spikes. We discuss extensions of this approach to non-stationary situations through a linear response formula. Additionally, the Thermodynamic Formalism provides a rigorous mathematical framework to study phase transitions that may illuminate the current discussions about signatures of criticality observed in many examples of neuroscience [45–53] and, especially, in Gibbs distributions inferred using the maximum entropy principle from experimental data [54–56]. The aim of this review is twofold. On the one hand, to bridge the gap between mathematicians working in the field of Thermodynamic Formalism and scientist interested in characterising spike statistics especially those applying the maximum entropy principle, which is placed here in a broader context. On the other hand, to show new perspectives in the field of mathematical neuroscience related to the Thermodynamic Formalism, including phase transitions.

The rest of the article is organized as follows. In section 2 we introduce the main tools and ideas of the Thermodynamic Formalism we use later. In section 3, 4 we present the uses of this formalism in neuroscience, for spike train statistics. In section 5 we end with a discussion and present perspectives for future works.

## 2. Mathematical setting of Thermodynamic Formalism

Our goal in this section is to present the basic tools and ideas of the Thermodynamic Formalism to the unfamiliar reader (detailed accounts of this subject can be found in [7,8,57–59])

## 2.1. General properties

To study a dynamical system from the perspective of the Thermodynamic Formalism, we need first of all, to describe the set of elements, that need to be either finite or countable. Thus, continuous particle characteristics such as position, speed, momentum do not enter in the setting we analyse here, unless one "coarse grain" the phase space with a finite or countable partition. Discrete particle characteristics like spin, symbols attached to specific features of a dynamical system, constitute the set of symbols denoted by  $A$  also referred as *alphabet*. Let  $A^M$  be the set of blocks of  $M$  symbols of the alphabet, that is, sequences of the form  $x_0x_1 \dots x_{M-1}$ , where  $x_i \in A, i = 0 \dots M-1$  and  $A^{\mathbb{N}}$ , is the set of right-infinite sequences of the form  $x = x_0x_1 \dots$ , with  $x_i \in A$  for all  $i \in \mathbb{N}$ . One may also consider the bi-infinite sequences  $A^{\mathbb{Z}}$  but we will mostly stick at  $A^{\mathbb{N}}$  in the sequel.

One can equip the space  $A^{\mathbb{N}}$  with a distance. We associate to  $\theta \in (0, 1)$  the distance  $d_\theta$  such that  $d_\theta(x, y) = \theta^m$ , where  $m$  is the largest non-negative integer such that  $x_i = y_i$  for every  $0 \leq i < m$ . In particular, if  $x = y$  then  $m = \infty$  and  $d_\theta(x, x) = 0$ , and, by convention, if  $x_0 \neq y_0$  then  $m = 0$ . Considering this distance, the metric space  $(A^{\mathbb{N}}, d_\theta)$  is compact [8].

To consider time order, we introduce the time evolution in the form of a left-shift,  $\sigma : A^{\mathbb{N}} \rightarrow A^{\mathbb{N}}$ , defined by  $\sigma(x)_i = x_{i+1}$  for all  $i \in \mathbb{N}$  and any  $x \in A^{\mathbb{N}}$ , where the index  $i$  refers to time.

Let us define continuous functions  $f : A^{\mathbb{N}} \rightarrow \mathbb{R}$ . We introduce the modulus of continuity of  $f$ :

$$\text{var}_k(f) := \sup\{|f(x) - f(y)| : x_i = y_i, \text{ for } i = 0, \dots, k-1\},$$

characterising the maximal variation of  $f$  on the set of infinite sequences agreeing on their first  $k$  symbols. The function  $f$  is continuous if  $\text{var}_k(f) \rightarrow 0$  when  $k \rightarrow \infty$ , and the continuity is exponential if  $\text{var}_k(f) \rightarrow 0$  exponentially fast. A function  $f : A^{\mathbb{N}} \rightarrow \mathbb{R}$  has range  $R$  if  $f(x) \equiv f(x_0, \dots, x_{R-1})$ , i.e.  $f$  only depend on the first  $R$  coordinates  $x_0, \dots, x_{R-1}$ . Functions with finite range are obviously continuous. The notion of continuity, and, especially, exponential continuity is essential when studying the thermodynamic formalism of functions with infinite range.

A useful concept for this formalism is that of the Markov partitions. Let  $\Omega$  be a general state space and a general dynamics  $\mathcal{G}$  on  $\Omega$ . Consider a finite covering of  $\Omega$ , made by sets called rectangles,  $\mathcal{R} = \{R_1, \dots, R_l\}$  with the property that for every pair  $i \neq j$ ,  $\text{int}(R_i) \cap \text{int}(R_j) = \emptyset$  and such that the closure of the image set  $G(R_i)$  equals a union of closures of sets in  $\mathcal{R}$ . There is a specific construction for different dynamical systems (for details we refer the reader to [8]). This construction allows one to make a natural coding for the state space, where the trajectories of the dynamical systems correspond to sequences of symbols of the labels identifying the sets  $R_i$ . Rectangles allow to partition a continuous phase space into a discrete finite partition. For hyperbolic dynamical systems, each point in the phase space admits local stable or unstable manifolds with non zero diameter. The edges of rectangles in this case, are made of these local stable and unstable manifolds and the corresponding partition is called "Markov partition" because the image of a rectangle under the dynamics can be represented by a Markov transition matrix [60].

### 2.1.1. Gibbs measures

We now define the measurable sets on  $A^{\mathbb{N}}$ . Let us denote the sequence  $x_0 \dots, x_{n-1}$  by  $x_{0,n-1}$ . The set  $[x_{0,n-1}] := \{y \in A^{\mathbb{N}} : y_0 = x_0, \dots, y_k = x_{k-1}\}$  is called cylinder. Cylinders define a Borel sigma algebra  $\mathcal{B}$  on  $A^{\mathbb{N}}$  (see for instance Chapter 1., in [61]). We are interested in probability distributions on  $(A^{\mathbb{N}}, \mathcal{B})$  also referred as (macro) states in physics. In the sequel we will skip the prefix "macro" and deal with "states" as probability measures on  $(A^{\mathbb{N}}, \mathcal{B})$ . There is a special class of states  $\nu$ , such that for any measurable set  $B \subset A^{\mathbb{N}}$ ,  $\nu(\sigma^{-1}(B)) = \nu(B)$ , where  $\sigma^{-1}(B)$  stands for the set of all the pre-images of  $\sigma$  of elements of  $B$ . These distributions are the set of shift-invariant probability measures denoted by  $\mathcal{M}_\sigma(A^{\mathbb{N}})$ . In general, it is possible to consider systems where there exists some forbidden transitions between symbols. In that case, we need to consider a subset of  $A$  invariant under the dynamics, i.e.  $X \subset A^{\mathbb{N}}$  such that  $\sigma(X) = X$ .

The analogy with the spin systems in statistical mechanics, at the root of the terminology "Thermodynamic Formalism", goes as follows. Let  $\phi : A^{\mathbb{N}} \rightarrow \mathbb{R}$  be a continuous function, also called "energy" or "potential". Two important examples are finite range potentials where  $\phi(x) \equiv \phi(x_{0,R-1})$ , or exponentially continuous potentials with infinite range. Then, the "energy" of a configuration of  $n$  sites based on the sequence  $x \in X$ , is given by the "Birkhoff sums":

$$S_n \phi(x) := \sum_{i=0}^{n-1} \phi \circ \sigma^i(x).$$

We define the measures that assign probability to the cylinder sets based on the potential function, with the so called "Gibbs property". There exists constants  $C > 1$  and  $F(\phi)$  such that for all  $x \in X$  and for all  $n \geq 1$

$$C^{-1} \leq \frac{\mu_\phi([x_{0,n-1}])}{\exp(\sum_{i=0}^{n-1} \phi \circ \sigma^i(x) - nF(\phi))} \leq C. \quad (1)$$

The measures satisfying the condition (1) are called "Gibbs measures". The quantity:

$$F(\phi) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \sum_{x_0 \cdots x_{n-1}} e^{S_n \phi(y)}, \quad (2)$$

for all  $y \in [x_{0,n-1}]$  [8], is called the "pressure" (or free energy) of the potential  $\phi$ . Observe that it does not depend on the measure  $\mu_\phi$  itself, only on the potential. Thus, two Gibbs measures associated to the same potential have the same pressure.

Given a continuous potential  $\phi$  such that  $\text{var}_k(\phi) \leq b\theta^k$  for some constants  $0 < \theta < 1$  and,  $b > 0$ , there is a unique shift-invariant probability measure satisfying the Gibbs property (1), [8]. Furthermore, under the same assumption for the potential, the associated Gibbs measure is mixing<sup>1</sup> and thus ergodic<sup>2</sup> (see [8,61] for definitions and details). Moreover, two continuous potentials  $\phi, \psi$  are called cohomologous with respect to the shift  $\sigma$  (and denoted by  $\phi \sim \psi$ ), if there is a continuous function  $u$  and a constant  $K$  such that:

$$\phi = \psi + u - u \circ \sigma - K.$$

Cohomologous potentials have the same Gibbs measure,  $\mu_\phi = \mu_\psi$ .

**Remark.** The word "range" used all over this review can be confusing for readers acquainted with statistical physics. As we are considering time evolution in dynamical systems, the "range" of a potential refers to a *time* range (not a spatial range). Thus, a range  $R$  potential corresponds to a function of trajectories  $x$  where the value of the function, at time  $n$ , depends only on  $R$  time coordinates  $n - R - 1, \dots, n$ . Especially, potentials of range 1 only depend on the present configuration, which means that the successive times are independent under the corresponding Gibbs distribution. This is the classical setting for the Maximum Entropy model used e.g. in neuroscience. We also call these specific Gibbs distribution memory-less.

### 2.1.2. Entropy and the variational principle

The Shannon entropy of a probability distribution quantifies the average level of "uncertainty" in its possible outcomes.

$$h(p) := - \sum_{x \in A} p(x) \log p(x). \quad (3)$$

<sup>1</sup> For any measurable set  $A, B$ ,  $\lim_{n \rightarrow \infty} \mu(A \cap \sigma^{-n}B) = \mu(A)\mu(B)$ .

<sup>2</sup> A measure  $\mu$  is said to be ergodic for the dynamics  $\mathcal{G}$  if for any measurable  $\mathcal{G}$ -invariant set  $A$ , ( $\mathcal{G}^{-1}(A) = A$ ), its measure is either  $\mu(A) = 0$  or  $\mu(A) = 1$ . See for instance Chapter 3 of [60] for a detailed introduction.

More generally, let  $\nu$  be a shift-invariant probability measure on  $X$ , we introduce the block entropy:

$$H_n(\nu) := - \sum_{x_{0,n-1} \in A^n} \nu([x_{0,n-1}]) \log \nu([x_{0,n-1}]).$$

For finite alphabets and because  $\nu$  is shift-invariant, the following limit exists,

$$h(\nu) := \lim_{n \rightarrow \infty} \frac{1}{n} H_n(\nu),$$

and is called the Kolmogorov-Sinai entropy or simply entropy of  $\nu$  [57,62]. The key is that the sequence  $H_n$  is sub-additive and the possible values lie on the compact set  $[0, \log |A|]$  (see [57]).

Another important quantity is the Kullback-Leibler (KL) divergence which quantifies the difference between probability distributions. Consider two probability distributions  $p, q$  on the same probability space  $X$ , the KL divergence is:

$$D(p||q) := \sum_{x \in X} p(x) \log \frac{p(x)}{q(x)}.$$

This divergence is finite whenever  $p$  is absolutely continuous with respect to  $q$ , and it is only zero if  $p = q$ .

Following the analogy with systems in statistical mechanics, an *equilibrium state* is defined as the measure that satisfies the so-called variational principle, namely:

$$\sup \left\{ h(\nu) + \int \phi d\nu : \nu \in \mathcal{M}_\sigma(X) \right\} = h(\mu_\phi) + \int \phi d\mu_\phi = F(\phi), \quad (4)$$

where  $\int \phi d\nu$  represents the expected value of  $\phi$  with respect to the measure  $\nu$  (would be noted  $\mathbb{E}_\nu(\phi)$  in a probabilistic context). Let us comment on this result. The first equality establishes that, among all shift-invariant probability measures, there is a unique one,  $\mu_\phi$ , which maximises  $h(\nu) + \int \phi d\nu$ . If  $\int \phi d\nu$  is fixed to some value, this corresponds to maximising the entropy under constraints. This is the Maximum Entropy Principle, but (4) is more general. The second equality states that the maximum,  $h(\mu_\phi) + \int \phi d\mu_\phi$  is exactly the pressure defined in eq. (2). The equation  $F(\phi) = h(\mu_\phi) + \int \phi d\mu_\phi$  establishes a link with thermodynamics as  $F(\phi)$  is equivalent to the free energy<sup>3</sup>. Thus, (4) coincides with the principle of free energy minimization in statistical mechanics but corresponds to a maximizing measure in our formalism because of a sign convention. Gibbs measures associated to potentials satisfying that  $\text{var}_k(\phi) \leq b\theta^k$ , as stated above, are equilibrium states. For more general potentials the notion of Gibbs states and equilibrium states may not coincide.

## 2.2. Observables and fluctuations of their time averages

An *observable*  $f$  is a function  $f : X \rightarrow \mathbb{R}$ , such that  $|f| < \infty$  ( $|\cdot|$  denote the absolute value), and which is time-translation invariant, i.e., for any time  $i$ ,  $f(x_{0,R-1}) = f(x_{i,i+R-1})$  whenever  $x_{0,R-1} = x_{i,i+R-1}$ .

An important observable, considered in the following sections, is the potential:

$$U_\lambda(x) = \sum_k \lambda_k f_k(x) \quad k \in \{1, \dots, K\}, \quad (5)$$

<sup>3</sup> In contrast to thermodynamics where free energy or pressure refer to different thermodynamic ensembles, we will not make such distinction here as it is irrelevant. Also, note that we do not keep the minus sign and take the Boltzmann constant equal to 1.

that is a linear combination of  $K$  observables  $f_k$  where the parameters  $\lambda_k$ 's  $\in \mathbb{R}$ . Here, we want to make a few important remarks. First, this decomposition is similar to the definition of potentials in thermodynamics, which are written in terms of a linear combination of intensive variables (e.g. temperature, chemical potential, and so on) and extensive functions of the configurations of the system (physical energy, number of particles). To emphasise this analogy we use the symbol  $U$  for the potential, instead of  $\phi$ . The function  $U_\lambda$  depends parametrically on the  $\lambda_k$ 's, hence we use the lower index  $\lambda$ , to denote  $U_\lambda$  as well as the corresponding Gibbs measure, denoted here  $\mu_\lambda$ .

Now, we denote  $A_n(f)$  the empirical average of the observable  $f$  in a sample of size  $n$ , that is,

$$A_n f(x) = \frac{1}{n} \sum_{i=0}^{n-1} f(\sigma^i x).$$

This quantity is important for empirical purposes. If  $\mu$  is ergodic, the convergence of the empirical average to the actual expected value is  $\mu$ -almost-sure, that is  $A_n(f) \xrightarrow{\text{a.s.}} \int f d\mu$  as  $n \rightarrow \infty$ .

### 2.3. Correlations

Let us consider a pair observables  $f, g \in L^2(\mu)$ , (square integrable functions with respect to  $\mu$ ). We define their correlation at time  $n$  by

$$C_{f,g}(n) := \int f \cdot g \circ \sigma^n d\mu - \int f d\mu \int g d\mu.$$

One also might be interested for the auto-correlation (or auto-covariance) of  $f$  at time  $n$ ,

$$C_f(n) := \int f \cdot f \circ \sigma^n d\mu - \left( \int f d\mu \right)^2.$$

Observe that these quantities might decay fast. The mixing property implies that the correlations go to zero with  $n \rightarrow +\infty$ .

#### 2.3.1. Properties of the pressure

From the pressure (2) important statistical information about the system can be obtained, in particular, correlations. A distinguished case corresponds to the potentials of the form (5). When the corresponding pressure is differentiable to any order, taking the successive derivatives of the pressure with respect to their conjugate parameters gives the average values of the observables, their correlations, and their high-order cumulants with respect to the Gibbs measure. That is, in general:

$$\frac{\partial^n F(U_\lambda)}{\partial \lambda_k^n} = \kappa_n \quad \text{for all } k \in \{1, \dots, K\}, \quad (6)$$

where  $\kappa_n$  is the cumulant of order  $n$  with respect to  $\mu_\lambda$ . In particular,  $\kappa_1$  is the mean of  $f_k$ ,  $\kappa_2$  is its variance,  $\kappa_3$  the skewness and  $\kappa_4$  the kurtosis. Partial derivatives with respect to pairs of parameters can also be considered [63]:

$$\begin{aligned} \frac{\partial^2 F(U_\lambda)}{\partial \lambda_k \partial \lambda_j} &= C_{f_k, f_j}(0) + \sum_{n=1}^{\infty} C_{f_k, f_j}(n) + \sum_{n=1}^{\infty} C_{f_j, f_k}(n) \\ &= \frac{\partial \mu[f_k]}{\partial \lambda_j} = \frac{\partial \mu[f_j]}{\partial \lambda_k}. \end{aligned} \quad (7)$$

Observe that we differentiate  $C_{f_k, f_j}$  from  $C_{f_j, f_k}$  as the dynamics may be irreversible in time.



**Remark.** The last two equations are fundamental. They establish a link between the variations in the average of the observable  $f_k$ , when slightly varying the parameter  $\lambda_j$  and the sum of time correlations between observables  $f_j, f_k$ . This result is known, in statistical physics as the fluctuation-dissipation theorem [64,65]. It relates, for example, in a ferromagnetic model, the variation of the magnetisation of the spin  $k$  to the variations of the local magnetic field  $h_k$ , via the magnetic susceptibility that is the second derivative of the free energy. This is also the context of linear response theory, that quantifies how a small perturbation of a parameter affects the average values of observables in terms of the unperturbed measure. As we discuss in section 4.5, equation (7) can also be used to extend results of the Thermodynamic Formalism to non-stationary situations.

Now, in the classical formulation in statistical physics and the Maximum Entropy models, only the correlation  $C_{f_k, f_j}(0)$  appears in (7), because successive times are independent (correlations  $C_{f_k, f_j}(n), C_{f_j, f_k}(n), n > 0$  vanish). When handling memory (thus, potentials with range  $R > 1$ ) there is an infinite sum (series) of correlations appearing in the linear response. This infinite sum converges whenever correlations are decaying sufficiently fast with time  $n$  (exponentially). In contrast, when they do not converge fast enough (e.g. power-law with a small exponent), the series diverge, leading to a divergence of the second derivative of the pressure, corresponding to a second-order phase transition<sup>4</sup>. Note that phase transitions in memory-less models can happen too if the instantaneous correlation function  $C_{f_k, f_j}(0)$  diverges, e.g. when the number of degrees of freedom in the system (number of spins, neurons) tends to infinity. Therefore, in the present setting second-order phase transitions can arise either when the number of degrees of freedom tends to infinity, or (not exclusive), when time correlations decay slowly.

Thus, eq. (7) connects the second derivative of the pressure, variations in static average of observables, and dynamical correlations. We discuss in the next sections theorems that relate the dynamical evolution to a criteria ensuring that time-correlations are exponentially decaying, preventing the possibility of second order phase transitions for systems with a finite number of degree of freedom.

### 2.3.2. Ruelle-Perron-Frobenius operator

Let  $C(X)$  be the set of continuous functions on  $X$ . Consider the potential  $\phi$ , and a continuous function  $f \in C(X)$ , so one can define a bounded linear operator associated to  $\phi$  (transfer operator), called the Ruelle-Perron-Frobenius (RPF) as follows:

$$\mathcal{L}_\phi f(x) = \sum_{y \in \sigma^{-1}x} e^{\phi(y)} f(y). \quad (8)$$

The spectral properties of (8) yields information to characterise the pressure and to study ergodic properties of the system, in particular, the rate of decay of their correlations functions [61]. For instance, if 1 is a simple eigenvalue and the modulus of each of the other eigenvalues is smaller than one, this is equivalent to be mixing [61]. When the potential considered is of finite range, then the transfer operator corresponds to a matrix and the whole formalism is equivalent to Markov chains defined on finite alphabets. A potential  $\phi$  is called *normalised* if  $\mathcal{L}_\phi(1) = 1$ . The log of a normalised potential of range  $R + 1$ , corresponds to the transition probabilities of a Markov chain with memory depth  $R$ . Moreover, in this case  $F(\phi) = 0$ . For Lipschitz observables in the finite dimensional case, the Perron-Frobenius theorem assures a unique eigenvector associated to the maximal eigenvalue, from which the unique invariant measure (Markov) is obtained. This measure has mixing properties, exponential decay of correlations,

<sup>4</sup> Here, follow the Ehrenfest classification of phase transitions [66]. There is a phase transition of order  $k$  if the pressure (free energy) is  $C^{k-1}$  but not  $C^k$ . Known examples are first-order, second-order, or infinite order (Kosterlitz-Thouless) phase transitions.

central limit theorem and a large deviations principle (see 2.3.4). When the operator acts on an infinite dimensional space (such as the space of continuous functions), then the spectrum of a bounded linear operator  $\mathcal{L}$  is given by the set  $\text{spec}(\mathcal{L}) = \{\lambda \in \mathbb{C} : \text{such that } (\lambda I - \mathcal{L}) \text{ has no bounded inverse}\}$ , this set may contain points that not necessarily are eigenvalues (see for instance [67]). In this case, the strategy is to find a proper subspace where the spectrum of  $\mathcal{L}$  has a finite number of such complex numbers whose norm is the spectral radius, say  $\rho$ , and the rest of the spectrum has norm strictly less than  $\rho$  (spectral gap). In this scenario, it is known that there is exponential decay of correlations for sufficiently regular class of observables (such as Lipschitz), and the central limit theorem holds. In the absence of the spectral gap, then one has sub-exponential decay of correlations, that breakdown the central limit theorem, and the phase transition phenomenon appear (for further details and precise definitions see [61], and the references therein).

To end up this section, note that given a potential  $\psi$  one can explicitly find a normalised potential  $\phi$  cohomologous to  $\psi$  as follows,

$$\phi := \psi + \log R - \log R \circ \sigma - \log \rho, \quad (9)$$

where  $R$  is the right eigenvector (real and positive) associated to the unique maximum eigenvalue  $\rho$  associated to  $\mathcal{L}_\psi$ .

**Remark.** Note that the normalisation of the potential  $\psi$  does not requires a partition function. In fact, as discussed below, the classical normalisation by a partition function is a particular case of (9), holding for memory-less potentials which *does not generalise* to range  $R > 1$  potentials.

### 2.3.3. Time averages and Central limit theorem

We have seen that if the measure  $\mu$  is ergodic the time averages  $A_n(f)$  converge almost surely to the expected value  $\int f d\mu$ . Now, we can ask about fluctuations around their expected value. The observable  $f$  satisfies the central limit theorem (CLT) with respect to  $(\sigma, \mu)$  if:

$$\frac{A_n(f) - n \int f d\mu}{\sqrt{n}} \xrightarrow{\text{law}} \mathcal{N}_{0, \sigma_f^2} \quad (10)$$

where  $\mathcal{N}_{0, \sigma_f^2}$  is the Gaussian distribution with mean zero and covariance  $\sigma_f^2$ , which is given by the following expression involving temporal correlations:

$$\sigma_{f,g}^2 = C_{f,g}(n)(0) + \sum_{n=1}^{\infty} C_{f,g}(n) + \sum_{n=1}^{\infty} C_{g,f}(n).$$

that is a particular case of (7).

Strong properties of convergence and exponential decay of correlations are ensured for Hölder continuous potentials in finite dimension. These properties are associated to the spectral gap property and does not hold (necessarily) for less regular potentials or in non-compact spaces [59,61].

### 2.3.4. Large deviations

The central limit theorem describes small fluctuations in the limit when  $n$  goes to infinity, rare events that are exponentially small is the matter of the large deviations theory.

An empirical average  $A_n(f)$  satisfies a large deviation principle (LDP) with rate function  $I_f$ , if the following limit exists:

$$I_f(s) := - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\{A_n(f) > s\}), \quad (11)$$

for  $s \in \mathbb{R}$ . The above condition for large  $n$  implies that  $\mathbb{P}(\{A_n(f) > s\}) \approx e^{-nI_f(s)}$ . In particular, if  $s > \int f d\mu$ , then  $\mathbb{P}(\{A_n(f) > s\})$  should tend to zero as  $n$  increases. The rate function tells us precisely



how fast this probability goes to zero. Computing the rate function from the equation (11), may be a laborious task. The Gärtner-Ellis theorem provides a way to compute  $I_f$  more easily [68]. To this end, let us introduce the *scaled cumulant generating function* (SCGF)<sup>5</sup> associated to the observable  $f$ , by

$$\Gamma_f(k) =: \lim_{n \rightarrow \infty} \frac{1}{n} \log \int e^{nkA_n(f)} d\mu \quad k \in \mathbb{R},$$

whenever the limit exists. If  $\Gamma_f$  is differentiable, then the Gärtner-Ellis theorem ensures that the average  $A_n(f)$  satisfies a LDP with rate function given by the Legendre transform of  $\Gamma_f$ , that is

$$I_f(s) = \max_{k \in \mathbb{R}} \{ks - \Gamma_f(k)\}.$$

Therefore, one can study the large deviations of empirical averages  $A_n(f)$  by first computing their SCGF, characterise its differentiability and then finding the Legendre transform.

The large deviation principle plays an important role in statistical mechanics. A large deviation principle can be used to relate entropy and free energy (here pressure) through a Legendre transform and to explain why variational principles arise in statistical mechanics [68,69].

#### 2.4. Potentials of range one

A specific case where the variational principle (4) holds, is when the potential has the form (5). Then, equilibrium states are probability distributions  $\mu_\lambda$ , that maximise the entropy (3), under the constraints of expected values of  $K$  observables  $\mathbb{E}_{\mu_\lambda}(f_k) := \sum_x f_k(x) \mu_\lambda(x) = C_k$  for  $k = 1, \dots, K$ . This problem can be solved introducing a Lagrange multipliers  $\lambda_k$  in the potential (5)

$$F(U_\lambda) := \max_p \{H(p) + \mathbb{E}_p(U_\lambda)\} = H(\mu_\lambda) + \mathbb{E}_{\mu_\lambda}(U_\lambda). \quad (12)$$

There exists a unique maximum entropy distribution  $\mu_\lambda$  (equilibrium state) satisfying the constraints. It turns out that the maximising distribution can be explicitly found for range one potentials and the distribution satisfies the Gibbs property (1), which in this particular case, reduces to,

$$\mu_\lambda(x) = \exp(-F(U_\lambda) + U_\lambda(x)) = \frac{\exp(U_\lambda(x))}{Z}, \quad (13)$$

for all  $x \in X$ . The equation (13) is obtained by considering  $F(U_\lambda) = \log Z$ , where  $Z$  is the "partition function". From equation (12), the expression for the entropy (3) and the Jensen inequality, one can obtain the formula for the pressure in this case:

$$F(U_\lambda) = \log \sum_{x \in X} e^{U_\lambda(x)}.$$

The constrained problem can be uniquely solved because the map  $\lambda \mapsto \mathbb{E}_{\mu_\lambda}(U)$  maps the real line monotonically onto the interval  $(\min U, \max U)$  [57].

For range one potentials, the measure of a block becomes a product distribution, given by:

$$\mu_\lambda([x_{0,n-1}]) = \prod_{i=0}^{n-1} \frac{\exp(U_\lambda(x_i))}{Z}. \quad (14)$$

<sup>5</sup> The name comes from the fact that the  $n$ -th cumulant of  $f$  can be obtained by  $t$  successive differentiation operations over of  $\Gamma_f(k)$  with respect to  $k$ , and then evaluating the result at  $k = 0$ .

As the index  $n$  corresponds to time, having an interaction depending on one single coordinate imply that configurations at distinct times are independent.

### 2.5. Finite range potentials

Equation (9) can be used to find the unique Markov measure associated to a finite range potential. As an example, consider a potential  $U$  of range two representing the pairwise interactions in a graph with incidence matrix  $I$ . The entries  $I(y, x) = 1$  represent the allowed transitions between symbols  $y \rightarrow x$  and  $I(y, x) = 0$  the forbidden. We introduce the finite  $|A| \times |A|$  transfer matrix  $\mathcal{L}_U$ , that corresponds to the RPF "operator" (8) restricted to a finite space.

$$\mathcal{L}_U(y, x) = I(y, x)e^{U(y)}, \quad y, x \in A, y \in \sigma^{-1}x \quad (15)$$

As anticipated in section 2.3.2, calling  $\rho$  the unique maximal positive eigenvalue of  $\mathcal{L}_U$  guaranteed by the Perron-Frobenius theorem, and  $R(x)$  and  $L(x)$  the  $x$ -th entry in the right and left eigenvectors associated to  $\rho$  respectively, we define a normalized potential  $\phi(y, x) = U(y) + \log R(y) - \log R(x) - \log \rho$  such that the matrix

$$P(y, x) = I(y, x)e^{\phi(y, x)} = \frac{I(y, x)R(y)e^{U(y)}}{\rho R(x)}$$

is stochastic, i.e.  $\sum_x P(y, x) = 1$ . It represents the transition probabilities of a Markov chain  $P(y \rightarrow x) = P(x | y)$ . The invariant measure  $p$  associated to the matrix  $P$  satisfying  $pP = p$  is

$$p(x) = \frac{R(x)L(x)}{\langle R, L \rangle}, \quad (16)$$

where  $\langle R, L \rangle = \sum_x R(x)L(x)$ . Note that normalisation is done without defining a partition function. The Markov measure  $\mu(p, P)$  of a block is given by  $\mu[x_{0,n}] = p(x_0)P(x_1, x_2) \cdots P(x_{n-1}, x_n)$  for  $x_k \in A, k = 0, \dots, n$ . Here, we have a nice way to show that this measure satisfies the Gibbs property using the Markov property  $\mu[x_{1,n} | x_0] = e^{\sum_{k=1}^n \log P(x_{k-1}, x_k)}$  where we see that the conditioning upon the first time is similar to left boundary conditions in statistical physics.

It follows from (9) and (16) that,  $\mu[x_{0,n}]$  obeys the variational principle and satisfy equation (1) where  $F(U) = \log \rho$ . The Gibbs measure  $\mu[x_{0,n}]$  give to each cylinder set, an exponential weight depending on the "energy" depending on smaller blocks.

### 2.6. Systems with infinite range potentials, Chains with infinite memory and Gibbs distributions

A general class of stochastic processes to deal with infinite memory are called *Chains with complete connections* [70,71]. These chains define non-markovian processes. However, Markovian approximations are possible and useful [72]. This section follows closely from [70].

**Definition** A system of transition probabilities is a family  $\{P_n\}_{n \in \mathbb{Z}}$  of functions with  $P_n(\cdot | \cdot) : A \times A_{-\infty, n-1} \rightarrow [0, 1]$  such that the following conditions hold for every  $n \in \mathbb{Z}$ :

(a) For every  $x_n \in A$  the function  $P_n(x_n | \cdot)$ , is measurable with respect to the filtration  $\mathcal{F}_{\leq n-1}$ .

(b) For every  $x_{-\infty, n-1} \in A_{-\infty, n-1}$ ,

$$\sum_{x_n \in A} P_n(x_n | x_{-\infty, n-1}) = 1.$$

**Definition** A probability measure  $\mu$  in  $\mathcal{P}(A_{-\infty,n}, \mathcal{F})$  is **consistent** with a system of transition probabilities  $\{P_n\}_{n \in \mathbb{Z}}$  if:

$$\int h(x_{-\infty,n}) \mu(dx) = \int \sum_{x_n \in A} h(x_{-\infty,n-1} x_n) P_n(x_n | x_{-\infty,n-1}) \mu(dx).$$

for all  $n \in \mathbb{Z}$  and all  $\mathcal{F}_{\leq n}$ -measurable functions  $h$ . The probability measure  $\mu$ , when it exists, is called a *chain with complete connections* consistent with the system of transition probabilities  $\{P_n\}_{n \in \mathbb{Z}}$ . It is possible that multiple measures are consistent with the same system of transition probabilities.

We now give conditions ensuring the existence and uniqueness of a probability measure consistent with the system of transition probabilities [70].

**Theorem** A system of continuous transition probabilities ( $\text{var}_m[P_n(x_n | \cdot)] \rightarrow 0$  as  $m \rightarrow +\infty$ ) on a compact space has at least one probability measure consistent with it.

**Definition** A system of transition probabilities is **non-null** if, for all  $n \in \mathbb{Z}$  and all  $x_{-\infty}^n \in A_{-\infty,n}$ :

$$P[x_n | x_{-\infty,n-1}] > 0$$

**Definition** A normalized potential has **bounded squared variations** if, for all  $n \in \mathbb{Z}$  and all  $x_{-\infty,n} \in A_{-\infty,n}$ :

$$\sum_{k \geq 0} \text{var}_k^2(\log P[x_n | x_{-\infty,n-1}]) < +\infty.$$

There exist a *unique* probability measure consistent with the system of transition probabilities if these are non-null and the associated normalised potential has bounded squared variations [70].

The link with the Gibbs distributions presented up to now is the following. Define  $\phi(n, x) : \mathbb{Z} \times A \rightarrow \mathbb{R}$  by:

$$\phi(n, x) \equiv \log P[x_n | x_{-\infty,n-1}], \quad (17)$$

and:

$$\Phi(m, n, x) = \sum_{r=m}^n \phi(r, x). \quad (18)$$

Then:

$$P[x_{m,n} | x_{-\infty,m-1}] = e^{\Phi(m,n,x)} = e^{\sum_{r=m}^n \phi(r,x)},$$

and:

$$\mu[x_{m,n}] = \int_{A_{-\infty,m-1}} e^{\Phi(m,n,x)} \mu(dx).$$

There is a mathematically well-founded correspondence between chains with complete connections and Gibbs distributions [5,70,73]. Let us discuss here some analogies. These last equations highlight the connection between chains with complete connections and Gibbs distributions in statistical physics where  $\phi$  acts as an “energy” [70]. The correspondence is obtained considering “time” as a 1-dimensional “lattice” and the “boundary conditions” as the past of the stochastic process. In contrast to statistical physics, there is no need to define a partition function (the potential is defined via transition probabilities, thus, is normalised).

While for chains with complete connections defined through transition probabilities the present is conditioned upon the past, Gibbs distributions in general allow conditioning “upon the future” as well. More generally, Gibbs distributions in statistical physics extend to probability distributions on  $\mathbb{Z}^d$  where the probability to observe a certain configuration of spins in a restricted region of space is constrained by the configuration at the boundaries of this region. They are therefore defined in terms of specifications [5,73], which determine finite-volume conditional probabilities when the exterior of

the volume is known. In one spatial dimension ( $d = 1$ ), identifying  $\mathbb{Z}$  with a time axis, this corresponds to conditioning both in the past and in the future. In contrast, families of transition probabilities with an exponential continuity rate define the so-called left-interval specifications (LIS) [70,74]. This leads to not equivalent notions of "Gibbsianness" [75].

In contrast to potentials studied up to now, the potential (18) is defined from transition probabilities (17) which *are not necessarily time-translation invariant*. This is the reason why the potential is noted  $\phi(n, x)$ , as it depends explicitly on time  $n$  and the configuration  $x$ . This case is closer to the setting where potential or energy is not necessarily invariant when moving along a lattice in statistical physics, therefore not constrained by the stationarity assumption made up to now. As we discuss in the next section, this is quite helpful to study neuronal network dynamics.

### 3. Thermodynamic Formalism in Neuroscience

In this section, we make the connection between the Thermodynamic Formalism and spiking neuronal dynamics. From the standpoint of mathematics, there are at least, two ways to consider spiking neuronal networks. First, as biological objects whose activity can be experimentally recorded using Multi-Electrode Arrays (MEA), often requiring sophisticated mathematical methods and algorithms for data analysis [76,77]. Second, neuronal networks are characterised by dynamical models, more or less derived from biophysics.

We start considering the statistical analysis of spike trains recorded from neuronal networks. For this case the Thermodynamic Formalism provides a powerful and insightful method to analyse the spatio-temporal statistics from experimental spike trains. We then focus on a specific question. When dealing with a model of spiking neurons how much the intrinsic dynamics of neurons, their interaction via synapses, the influence of stimuli, constraint the collective spatio-temporal spike statistics?

Neurons are (nonlinear) entities evolving in a concerted way (as they interact via synapses) and responding to external stimuli [30,31]. The theoretical analysis of this high dimensional systems can be made thanks to mathematical methods (dynamical systems, bifurcations theory, stochastic processes, partial differential equations) or theoretical physics (statistical physics, nonlinear physics). Here, one might be interested in what Thermodynamic Formalism can bring considering neuronal models dynamics. In this spirit we consider two models, the Integrate and Fire model and the Galves Löcherbach model. Most of our presentation focuses on stationary situation characterised by equilibrium states. We nevertheless consider the extension of Thermodynamic Formalism to non-stationary situations.

At the end of the section, we address a couple of open questions.

1. What is the natural alphabet for spiking neurons dynamics? As we shall see, although the binary representation of spikes is a good candidate, it is too naive, as the relevant alphabet can be constructed on time blocks of spikes. A subsidiary question is about the size (time depth) of these blocks.
2. Under which conditions can the Thermodynamic Formalism machinery be faithfully applied on a spiking neuronal network model?
3. What are the limits when the main theorems of the Thermodynamic Formalism can and cannot be applied and what are the consequences on neuronal dynamics and spike statistics?

#### 3.1. Statistics of spike trains and Gibbs distribution

The human brain is composed of about a hundred billion neurons that mostly communicate among themselves together using sequences of spikes<sup>6</sup>. Spikes are binary events. Although the action potentials can vary in duration, amplitude, and shape, depending e.g. on neurons' type they have a canonical form so that they can be considered as identical stereotyped events.

---

<sup>6</sup> Sub-threshold oscillations also play an important role [78] and in organs like the retina, most neurons do not spike.

The main physiological reason for spike occurrence is that they can propagate information on different scales in the nervous system (centimeters to one meter) essentially without attenuation (active conduction as opposed to passive, Ohmic, conduction). However, from a contemporary point of view, spikes are also considered as events constituting "bits" of information. In this paradigm, it is tempting to consider spike trains as objects containing a "neural code" [79], i.e., a language that neurons use to communicate and that one could decipher. This terminology should not be considered literally, as in contrast to computer codes, spike trains have a wide variability (e.g. the repetition of the same stimulus, even under controlled experimental conditions does not induce the same sequence of spikes as a response). In addition, nothing guarantees that there is only one code. Considered from the perspective of the Thermodynamic Formalism, the notion of neural codes can have several meanings. (1) Spike trains constitute a symbolic coding of voltage dynamics (which depends on neuronal interactions and stimuli); (2) The way how neuronal dynamical systems (especially spike trains) are affected by stimuli, provides a way for downstream networks to infer the stimulus (e.g. the retina encodes a visual scene in spike trains which are decoded by the visual cortex, capable to reconstruct a representation of the visual scene). Here, we essentially want to address the following questions: How to use Thermodynamic Formalism to fit experimental (or numerically generated) spike train data? and Which Gibbs distribution is produced by a network of neurons whose dynamics is known?

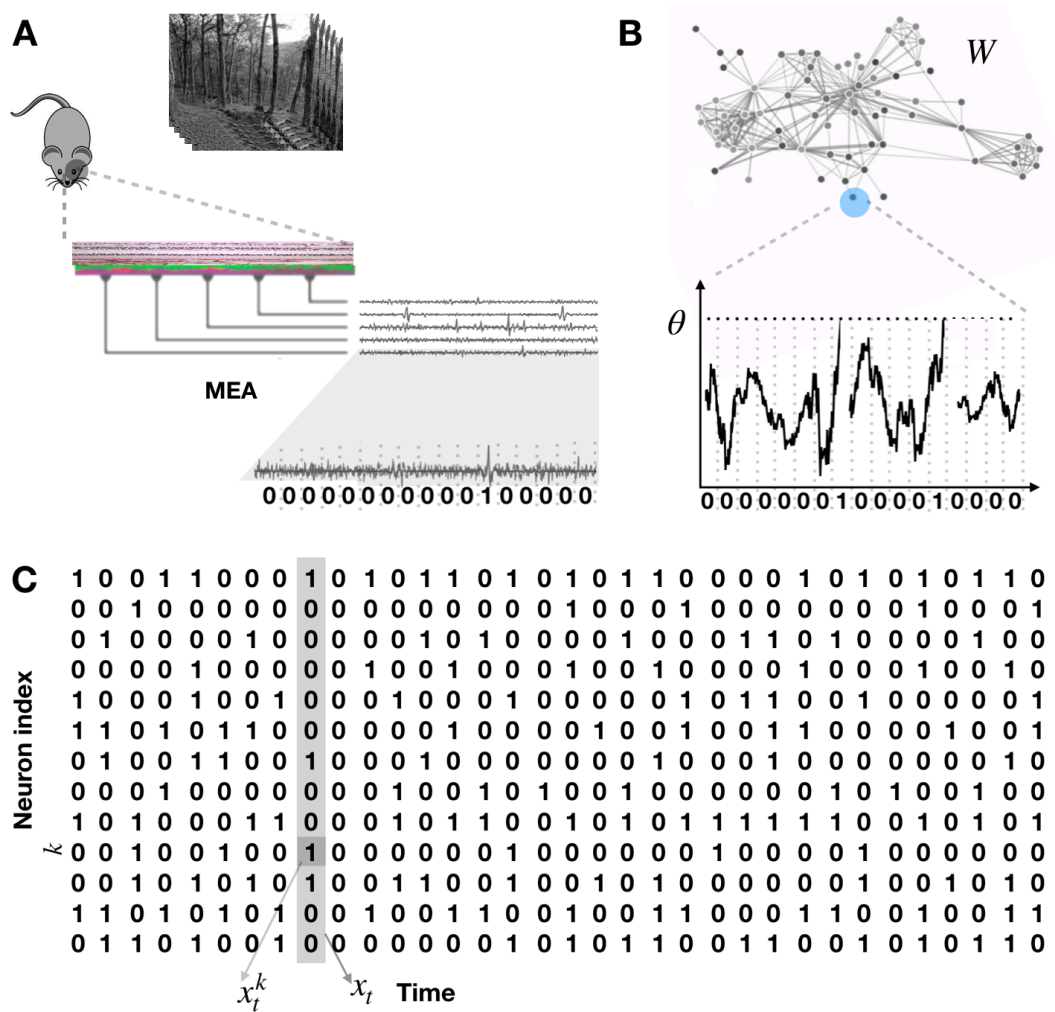
It is useful to consider spikes as instantaneous events (while the duration is about 1ms) and identify the maximum in the action potential course as the "time of the spike" [80]. This implicitly assumes that one considers dynamics on time scales larger than one millisecond. The binary representation is obtained by using a window of a constant "binning size" (of order 10 – 20 ms) over the continuous time course of membrane potentials and count how many spikes there are per neuron within each time bin. Two or more spikes may occur within the same time bin, in that case, the convention is to consider these events equivalent to just one spike. This, not harmless procedure [81], transforms experimental data into sequences of binary patterns (see figure 1) leading to the following symbolic description.

Denoting the discrete time index by  $n$ , the *spike-state* of neuron  $k$  is denoted by  $x_n^k \in \{0,1\}$  depending whether the  $k$ -th neuron emits a spike during the  $n$ -th time bin or not. A *spike pattern* is the spike-state of all the neurons in a network of  $N$  neurons at a given time bin, and is denoted by  $x_n := [x_n^k]_{k=1}^N$ . A *spike block* denoted by  $x_{n,r} := x_n x_{n+1} \cdots x_r$  is a sequence of spike patterns. The length of the spike block  $x_{n,r}$  is  $r - n + 1$ . A *spike train* denoted by  $x$  is the spike block representing the whole sequence of spike patterns. We consider spike trains of finite and infinite length. The set of all possible spike blocks of length  $R$  in a network of  $N$  neurons is denoted by  $A_R^N$ .

Thus, in comparison to the previous section, and especially, section 2.1, symbols are here spike blocks of length  $R$ , the alphabet, previously denoted  $A$  is denoted here  $A_R^N$ , making explicit the dependence in the number of neurons,  $N$ , and the block depth  $R$ . It is important to make this dependence explicit as we consider, later in this review, the effect of increasing  $N$  and  $R$ .

### 3.2. Conditional probabilities for spike trains

The probability that a biological neuron, embedded in a network, emits a spike in a given time bin depends on the history of all variables determining the evolution of the neural network (voltage, conductances, concentrations of ions, neurotransmitters, etc.). Most of these variables are not experimentally accessible. Even if they were, there would be no hope to predict, from this huge amount of information, the statistics of spikes. Dealing with neuronal models, the situation is simpler as there are fewer variables to control and their dynamics are known explicitly, but, even in this case, it is in general not possible to access spike statistics from dynamics. A simplification is to consider that the probability of a spike pattern depends only on the spikes emitted in the past by the network. This way, one can ignore the hidden dynamics of inaccessible variables and compute the statistics from



**Figure 1. From experimental spike trains to mathematical modelling.** **A) Experimental set-up.** MEA detect spikes from living neuronal tissue. In this illustration, the retina of a mammalian is put into the MEA and submitted to natural light stimuli. The membrane potential of retinal ganglion cells is recorded and analysed to extract the spikes using spike sorting algorithms [76,77]. **B) Mathematical models** of biophysically inspired spiking networks can be used to study spike trains. **Top.** Neurons, considered here as points in a lattice, interact via synaptic connections, on an oriented weighted graph, where the matrix of weights is denoted  $W$ . **Bottom.** A prominent mathematical class of models is the Integrate and Fire model where the membrane potential is modelled by a stochastic differential equation (black trajectory) with threshold condition  $\theta$ . The neuron is considered to spike whenever the membrane potential reaches the threshold. Then, it is reset to some constant value. Binning time using windows of a few ms length, one can associate to the continuous-time trajectory of the membrane potential a discrete-time sequence of 0s and 1s characterising the spike state of the neuron in each time window. **C) Spike trains.** Using the binary representation at the bottom of (B) for each neuron in a network one obtains sequences of binary spike patterns (spike trains) symbolically representing the underlying neuronal dynamics.

what can be measured. Still, characterise the probability of a spike pattern given the history of the system, is in general out of reach (with, at least, two exceptions described in the next section).

The idea is to characterise the spike trains statistics through a family of transition probabilities of the form:

$$P(x_{n+1} \mid x_{n-R,n}) \equiv e^{\phi(x_{n-R,n+1})} > 0. \quad (19)$$



where  $R$  is the memory of the spike sequence, i.e. the time horizon on which the present depends on the past. Having these transition probabilities and an initial condition (or initial distribution) one can define a Markov chain (or a chain with infinite memory if  $R \rightarrow +\infty$ ). It is possible, for some models, to write explicitly these probabilities. In eq. (19) we have assumed that all transition probabilities are strictly positive. This is necessary to ensure that the uniqueness of the corresponding Gibbs distribution. Then, one can associate to (19) a range  $R$  potential  $\phi(x_{n-R,n}) > -\infty$ . On experimental grounds the problem is to estimate these probabilities from data. Since there are  $2^{NR}$  possible spike blocks, for  $N$  and/or  $R$  big (e.g.  $NR > 20$ ) it becomes rapidly impossible to estimate these transition probabilities from experimental data using a frequentist approach, as most of these transitions do not even occur within the finite experimental sample.

However, one can try to guess the form of these transition probabilities. One possibility is to start with an ad-hoc form, capturing the main features of neuronal dynamics. A canonical example is the Generalised Linear Model (GLM) [82] where the transition probabilities take the form:

$$P(x_{n+1}^i | x_{n-R,n}) = f \left( b^i + \sum_{j \in \mathcal{B}_i} (H_{ij} * r^j)(n) \right) \quad (20)$$

where  $f$  is a non-linear function. The term  $b^i$  is a constant fixing the baseline firing rate of neuron  $i$ .  $K^i$  is a causal linear convolution kernel,  $*$  is the convolution product,  $H_{ij}$  is the memory kernel and  $r^j(n)$  is the spike train of neuron  $j$  before time  $n$ . In this case the memory kernel consider the spikes between  $n - R$  and  $n$ , but  $R$  can go arbitrarily to the past.

As we show in section 4.3 this form can be established for Integrate and Fire models. Eq. (20) gives an Ansatz for the marginal probability that neuron  $i$  spikes at time  $n + 1$ , given the history of the network. The general form (19), providing the join probability to have the spike pattern at time  $n + 1$  is obtained assuming that neurons are conditionally independent. This can be justified if one assumes that, due to synaptic transmission and delays, neurons have not the time to interact within one time bin. This means that time bins must not be too large and/or synapses must not be too fast (like gap junctions [83]).

**Remark.** The GLM, instead of describing conditional probabilities, characterises the spike rate or conditional intensity of an auto-regressive Poisson process.

A second approach is based on the variational principle (14), maximising entropy under constraints. Both approaches can be addressed from the perspective of the Thermodynamic Formalism.

### 3.3. The Hammersley-Clifford theorem

In our representation spikes take a binary value 0 or 1. Thus, any potential of range  $R$  is a function taking a finite set of values. A general theorem from Hammersley and Clifford [84,85] states that *any* range- $R$  observable, in particular, the potential  $\phi(x_{n-R,n})$ , can be written in the form

$$\phi(x_{n-R,n}) = \sum_l \phi_l m_l(x_{n-R,n}), \quad (21)$$

where the coefficients  $\phi_l$  corresponds to the decomposition of  $\phi$  in the space of finite range  $R$ -observables. This is analogous to the equation (5) with two main differences. First, the linear combination in (5) is used as an example making a link with Thermodynamics and the Maximum Entropy principle. Here, the decomposition (21) is a *systematic* expansion of any potential of range  $R$  defined over spikes sequences. Second, in contrast to (5), the observables, denoted  $f_k$  in (5) consider only finitely many values.

Equation (21) is a linear decomposition on a basis of eigenfunctions called, from now on, monomials [86]. They have the form:

$$m_l(x_{n-R,n}) = \prod_{k=1}^d x_{i_k}^{n_k}.$$

where  $n_k = 1, \dots, N$  is a neuron index, and  $i_k = n - R, \dots, n$  a time index. Thus,  $m_l(x_{n-R,n}) = 1$  if and only if, in the spike sequence  $x_{n-R,n}$  neuron  $n_k$  spikes at time  $i_k$  for all  $k = 1, \dots, d$ . Otherwise,  $m_l(x_{n-R,n}) = 0$ . The number  $d$  is the *degree* of the monomial; degree one monomials have the form  $x_{i_1}^{n_1}$ , taking the value 1 if and only if neuron  $n_1$  spikes at time  $i_1$ , degree two monomials have the form  $x_{i_1}^{n_1} x_{i_2}^{n_2}$ , taking the value 1 if and only if neuron  $n_1$  spikes at time  $i_1$  and neuron  $n_2$  spikes at time  $i_2$ , and so on. Thus, monomials provide a notion of spike interactions, similar to spins interactions in magnetic systems. For example, monomials of degree two correspond to pairwise interactions, like e.g. in Ising model. In contrast to the Ising model, the interactions considered here may involve time delay between spikes.

There are  $2^{NR}$  monomials for  $N$  neurons and a given range  $R$ . One can index them by an integer  $l$  in one-to-one correspondence with the set of pairs  $(i_k, n_k)$ . The advantage of the monomial representation is that focuses on spike events, which is natural for spiking neuronal dynamics. The Hammersley Clifford decomposition gives thus a canonical way to write any range  $R$  potentials as a linear combination of monomials of maximum degree  $R$ . This includes the GLM potential which can then be embedded in the same framework [83].

As emphasised above, the Hammersley Clifford decomposition is analogous to the expression of thermodynamics potentials as a sum of products of an intensive quantity (e.g. temperature) with an extensive one (e.g. the energy). Depending on the physical constraints of the problem, one defines a *thermodynamic ensemble* where the average value of extensive quantities (energy, number of particles, volume, magnetisation) is prescribed. Whereas first principles allow guessing the form of the potential in thermodynamics, there is no such recipe in neuroscience. Moreover, one cannot use the complete expansion (21) on practical grounds, simply because large degree monomials have a vanishing empirical probability. More precisely, the average value of a monomial of degree  $d$  decays exponentially fast with  $d$ . This leads to two problems. (i) How to determine (from data) the constraints which are necessary to correctly characterise the spike train statistics? (ii) Are there constraints that are equivalent? (i) can be addressed in the context of information geometry [87] while (ii) can be approached using cohomology [86]. We do not develop further these aspects here referring the reader to the quoted articles.

The variational principle (4) (or its finite version (12)) provides a systematic way for inferring Gibbs distributions from empirical average values of spikes interactions (monomials). We make the construction explicit in the next subsections.

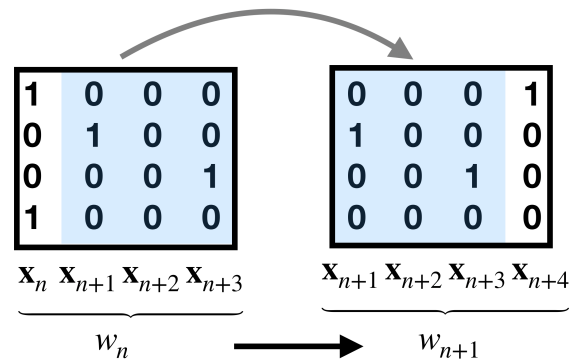
### 3.3.1. Finite memory, Markov chains and Gibbs distributions

We now show explicitly how to build a Gibbs measure from a finite set of experimental averages as constraints of the maximum entropy variational problem. We assume that these constraints involve events (monomials) over a memory depth  $R$ . We build the corresponding Markov chain using the material of section 2.3.2. We associate to each spike block  $x_{n,n+R-1}$  an integer  $w_n$ ,

$$w_n = \sum_{r=0}^{R-1} \sum_{k=1}^N 2^{k-1+Nr} x_{n+r}^k,$$

we write  $w_n \sim x_{n,n+R-1}$ . In this way a sequence of spike patterns (spike block) can be encoded as sequences of integers, that define the alphabet. We next define the incidence matrix ("grammar") between symbols of the alphabet. Not all transitions between symbols are legal or allowed. A transition

between the two symbols denoted by  $w_n \rightarrow w_{n+1}$  or  $w_n, w_{n+1}$  is legal if the corresponding blocks overlap according to this pattern, i.e. they have the block  $x_{n+1} \dots x_{n+R-1}$  in common, (see Fig. 2).



**Figure 2. Spike blocks transition** Example of legal transition  $w_n \rightarrow w_{n+1}$  between blocks of range four ( $R = 4$ ). The two blocks are  $w_n \sim x_{n,n+3}$  and  $w_{n+1} \sim x_{n+1,n+4}$  and have in common the block  $x_{n+1}x_{n+2}x_{n+3}$  in light blue.

This define an incidence matrix  $I(w', w) = 1$  if the transition between symbols  $w', w$  is legal and 0 otherwise. This incidence matrix defines the *grammar* of allowed and forbidden words or sequences of symbols. From this incidence matrix, we define the Perron-Frobenius transfer matrix  $\mathcal{L}_\psi$  in the same way as in (15). To obtain the unique Markov transition matrix of maximum entropy we follow the procedure explained in section 2.3.2.

Thus, for a given choice of monomials, we associate a potential of the form (21) where the  $\lambda_l$ s that do not correspond to a chosen monomial is set to 0. Then, one computes the empirical average of the chosen monomials from data. From the Perron-Frobenius theorem, there is a unique Gibbs measure, the Markov measure  $\mu_\lambda$  of the Markov chain, which solves the variational problem (14), giving a statistical model of data, minimizing the KL divergence between the empirical measure and  $\mu_\lambda$ . Here  $\lambda$  is the set of parameters  $\lambda_l$  that achieve the variational principle. These parameters can be computed numerically, either using the explicit form of the measure (16) [43] or using MonteCarlo methods [88]. The software PRANAS allows to handle spike trains statistics by computing numerically the Gibbs distribution solving (14) for up to 100 neurons [89].

### 3.3.2. Spectral gap and thermodynamic limit

For a potential of finite range  $R$  and a finite number of neurons provided  $\lambda_l > -\infty$ , for all  $l$  in (21), the Perron-Frobenius theorem guarantees the uniqueness of the Gibbs measure  $\mu_\lambda$  solving the variational principle (14). Moreover, the pressure being a real analytic function for finite range potentials, is infinitely differentiable with respect to parameters and there is an exponential decay of correlations with respect to time. This last aspect is due to the gap in the spectrum of the transfer matrix (15). These properties may not hold if either  $R \rightarrow +\infty$  or, if  $N \rightarrow \infty$  (corresponding to a thermodynamic limit) where the potential may lose regularity. Here, one has to consider the Thermodynamic Formalism in infinite dimension, on a functional space, the space of continuous functions. The case  $R \rightarrow +\infty$  corresponds to a potential with infinite range associated in our case to a spike statistics with infinite memory. This is discussed in the next section and in section 4 where we show that neuronal models can have such an unbounded memory. More generally, the limits  $N \rightarrow \infty$  or  $R \rightarrow +\infty$  can induce important effects, such as phase transitions, further commented in the discussion section.

## 4. Spiking Neuronal Network models: The Leaky Integrate-and-Fire and beyond

There are many models for neuronal dynamics both at the level of individual neurons and neuronal networks [30,31,90]. Here, we consider a canonical example of such a model, actually the first one proposed to the scientific community was introduced by Lapique in 1907 [91]. The main interest

is to make a nice link between dynamics, coding and spikes, paving the way to use Thermodynamic Formalism to analyse the spike train statistics.

#### 4.1. Dynamics and spikes

A fundamental equation in neuronal membrane potential dynamics is the conservation of electric charge, written, in its most canonical form, as follows:

$$C \frac{dV}{dt} = - \sum_X g_X (V - V_X) + I(t),$$

where  $C$  is the membrane capacitance of the neuron and  $V$  its membrane potential. The sum  $\sum_X$  holds on ionic currents of the form  $i_X = -g_X (V - V_X)$  involving specific ionic channels permeable to specific ions (e.g.  $Na^+$ ,  $K^+$ ,  $Cl^-$ ,  $Ca^{2+}$ ). Here, we include intrinsic neuron's current (e.g. sodium and potassium currents triggering a spike [92]) and synaptic currents [90]. The conductance,  $g_X$ , of channels of type  $X$  depends, in general, non linearly on activation variables, themselves dependent on the voltage.  $V_X$  is the Nernst reversal potential, i.e. the value of the membrane potential at which the current  $i_X$  reverses its direction. Finally,  $I(t)$  is an external current that can mimic, e.g., an injection by an electrode or an external stimulus.

In its simplest form, for a single isolated neuron, this equation takes the form:

$$C \frac{dV}{dt} = -\frac{1}{R} V + I(t), \quad (22)$$

where  $R$  the membrane resistance and the term  $g = \frac{1}{R}$  corresponds to a unique passive conductance. In this case, we consider only a leak current where the leak reversal potential is set to 0. This is the equation of an RC circuit which is quite simple, but quite far from a real neuron, as this equation does not even produce spikes. To circumvent this problem one introduces a threshold,  $\theta$ , such that equation (22) holds whenever  $V(t) < \theta$  (sub-threshold dynamics), corresponding to the "integrate" phase. In contrast, for all times  $t_{(r)}$  such that  $V(t_{(r)}) = \theta$ , two effects take place: (i) The membrane potential of the neuron is reset *instantaneously* to a rest value, here 0, without loss of generality; (ii) A spike is recorded at times  $t_{(r)}$  called "spike times". This is the "fire" phase (see Fig. 1).

While this is a simple artificial way to generate spikes, there is a huge price to pay on the mathematical ground because the threshold introduces a singularity set in the phase space where the dynamic is not differentiable. We develop this aspect below.

The generalisation of (22) to a network of  $N$  neurons is straightforward. Adding the contribution of synaptic currents,  $I_{(syn)}^k(t)$ , building the network interactions, we obtain:

$$C^k \frac{dV^k}{dt} + \frac{1}{R} V^k = I_{(syn)}^k(t) + S^k(t) + \sigma_B \zeta^k(t), \quad \text{if } V^k(t) < \theta, \quad (23)$$

where  $C^k$  is the membrane capacitance of neuron  $k$ ,  $V^k$ , its membrane potential. The resistance  $R$  is assumed to be the same for all neurons. In the synaptic current,

$$I_{(syn)}^k(t) = \sum_{j,r} W_{kj} \alpha(t - t_{(r)}^j), \quad (24)$$

the parameter  $W_{kj}$  represent the synaptic strength ("weight") from the pre-synaptic neuron  $j$  to the post-synaptic neuron  $k$  (see Fig. 1 B). Synaptic weights can be negative (inhibition) or positive (excitation). By convention,  $W_{kj} = 0$  if there is no connection from  $j$  to  $k$ . This way, the sum in (24) holds for  $j = 1 \dots N$ .

The function  $\alpha$  represent the time profile of the postsynaptic current induced by a pre-synaptic spike [93]. It has been experimentally observed that the tail of this function is *exponential*. On mathematical grounds this is essential. The sum in equation (24) consider *all* the spike times  $t_{(r)}^j$  emitted by all the pre-synaptic neurons  $j$  *before* time  $t$ . When considering the asymptotic regime  $t \rightarrow -\infty$  (to get rid of initial conditions) this sum may contain an infinite number of terms. Thus, to ensure the sumability of this series one needs that  $\alpha$  decays sufficiently fast (here exponentially fast).

Eq. (23) holds in the sub-threshold regime. The term  $S^k(t)$  represent an external stimulus, and  $\xi^k(t)$  is a white noise whose amplitude is modulated by  $\sigma_B$ . When the membrane potential of neuron  $k$  reaches the firing threshold at a firing time  $t_{(r)}^k$ , for some  $r$ , i.e.  $V^k(t_{(r)}^k) \geq \theta$ , the neuron  $k$  fires an action potential and its membrane potential is reset to a fixed reset value instantaneously (see figure 1 B).

While the equations (23), (24) looks rather simple, the right hand size of Eq. (23) depends, via (24) on a possibly *uncountable* set of events (the spike times) corresponding to a possible infinite history of the voltage dynamics of the network. In this sense, these equations do not represent a classical dynamical system, where the knowledge of the variables at a given time allows to compute the variables value at a future time by integrating the flow. Here, we require the knowledge of the history, back to the last time in the past where neuron was reset to make the integration. This history can go quite far in the past, with a dependence decaying like the tail of the function  $\alpha$ .

To circumvent these problems we need first to get rid of the fact that spike times belong *a priori* to an uncountable set. There are two alternatives. The first one, briefly explored in this section, consists of discretising time as done e.g. in [94]. This leads to important results in related with the Thermodynamic Formalism (see [95,96] for details). The second one, keeping time continuous, is commented in the discussion section.

#### 4.2. A discrete time version of the Leaky-Integrate and Fire model

The time discretisation of the model (23) reads:

$$\begin{cases} V_{n+1}^k = \gamma V_n^k + \sum_{j=1}^N W_{kj} x_n^j + S_n^k + \sigma_B \xi_{n+1}^k, & \text{if } V_n^k < \theta, \quad \text{Integrate phase;} \\ V_{n+1}^k = 0 \quad \text{and} \quad x_n^k = 1, & \text{if } V_n^k \geq \theta, \quad \text{Firing phase.} \end{cases} \quad (25)$$

For simplicity we have assumed that all neurons have the same capacitance  $C^k = C$ , and set  $\gamma = 1 - \frac{dt}{\tau}$ , where  $\tau = RC^7$ , with  $0 < \gamma < 1$ , and then taken  $dt = 1$ . If, at some, discrete-time  $n$ ,  $V_n^k$  exceeds the threshold  $\theta$ , the membrane potential is reset at time  $n + 1$  and a spike is recorded at  $n$  for neuron  $k$ , i.e.  $x_n^k = 1$ . Below the threshold, the random dynamical systems is ruled by (25). We have assumed that synapses are instantaneous. Then, the synaptic input is  $\sum_j W_{kj} x_n^j$ , that correspond to the pre-synaptic neuron  $j$  that acts on the post-synaptic neuron  $k$  whenever  $j$  spikes,  $x_n^j = 1$ .  $S_n^k$  is the time discretization of the external stimulus.  $\xi_n^k$  are independent standard Gaussian random variables.

It is easy to integrate equation (25) conditionally upon a fixed spike sequence  $x$ . A trajectory  $V = \{V_n^k, k = 1 \dots N, n \in \mathbb{Z}\}$  is *compatible* with this spike sequence if  $\chi(V_n^k > \theta) = x_n^k, \forall k = 1 \dots N, n \in \mathbb{Z}$ , where  $\chi(A)$  is the indicator function of the logical event  $A$ ,  $\chi(A) = 1$  if  $A$  is true,  $\chi(A) = 0$  otherwise. We discuss in more detail the compatibility condition in section 4.4, when dealing with symbolic coding. For the moment, assume that  $V$  and  $x$  are compatible. We note

<sup>7</sup>  $\tau$  is the characteristic time scale of the membrane response,  $dt$  is an integration time step which has to be much smaller than  $\tau$  to preserve the physical relevance, whereas it has to strictly positive to have a time-discretization scheme.

$\tau_k(x, n) = \max \{ l, l < n \mid x_l^k = 1 \}$  the last time before  $n$  where neuron  $k$  has spiked, thus whose voltage was reset to 0. Then:

$$V_{n+1}^k = \sum_{j=1}^N W_{kj} \eta_{kj}(n, x) + I_0 \frac{1 - \gamma^{n+1-\tau_k(x, n)}}{1 - \gamma} + \sum_{l=\tau_k(x, n)}^n \gamma^{n-l} S_l^k + \sigma_B \sum_{l=\tau_k(x, n)}^n \gamma^{n-l} \xi^k(l), \quad (26)$$

where:

$$\eta_{kj}(n, x) = \sum_{l=\tau_k(x, n)}^n \gamma^{n-l} x_l^j,$$

integrates the influence of pre-synaptic neuron  $j$  on the time interval  $[\tau_k(x, n) + 1, n]$ . Each spike emitted by this neuron, at times  $l$  in this time interval, contributes with the weight  $\gamma^{n-l}$  and there is no contribution at times where  $x_l^j = 0$ . The condition  $\gamma < 1$  implies an exponential decay in the spike history dependence with characteristic time:

$$\tau_\gamma = - \left[ \frac{1}{\log \gamma} \right].$$

Likewise,  $\sum_{l=\tau_k(x, n)}^n \gamma^{n-l} S_l^k$  integrates the stimulus influence on neuron  $k$  and  $\sigma_B \sum_{l=\tau_k(x, n)}^n \gamma^{n-l} \xi^k(l)$  is the integrated noise term. This is a Gaussian random variable, with mean zero and variance  $\sigma_B^2 \frac{1 - \gamma^{2(n+1-\tau_k(x, n))}}{1 - \gamma^2}$ . In (26) the dependence on the initial condition does not appear because we assume the initial time to be  $n_0 \rightarrow -\infty$ . So, either neuron  $k$  has spiked in the time interval  $]-\infty, n]$  and the voltage is reset to 0, or it has not spiked but the initial condition dependence decays like  $\gamma^{n-n_0}$  that vanishes when  $n_0 \rightarrow -\infty$ .

#### 4.3. Gibbs distribution of the discrete LIF model

Thanks to the integrated equation (26) and because the integrated noise is Gaussian it is now easy to compute the probability that neuron  $k$  spikes at time  $n + 1$  given the history  $x$ ,  $P(x_{n+1}^k = 1 \mid x_{-\infty, n}) = P(V_{n+1}^k \geq \theta \mid x_{-\infty, n})$ :

$$P(x_{n+1}^k = 1 \mid x_{-\infty, n}) = f \left( \frac{\theta - \sum_{j=1}^N W_{kj} \eta_{kj}(n, x) - \sum_{l=\tau_k(x, n)}^n \gamma^{n-l} S_l^k}{\sigma_B \sqrt{\frac{1 - \gamma^{2(n+1-\tau_k(x, n))}}{1 - \gamma^2}}} \right) \quad (27)$$

where  $f(z) = \int_z^{+\infty} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} dz$ . We used a small abuse of notation. The conditioning upon  $x_{-\infty, n}$  means, in fact, the conditioning upon the sequence  $x_{n-1, R(x)}$  where  $R(x) = \min \{ k = 1 \dots N \mid \tau_k(x, n) \}$ . We condition upon the spike history prior to  $n$  back to the last time in the past where all neurons have been reset at least once. While, for eq. (27) we just need to consider the history back to  $\tau_k(x, n)$ , the conditioning upon  $x_{n-1, R(x)}$  is necessary when considering the conditional joint probability of spiking patterns. The joint probability is conditionally independent given the past:

$$P(x_{n+1} \mid x_{-\infty, n}) = \prod_{k=1}^N \left( x_{n+1}^k P(x_{n+1}^k = 1 \mid x_{-\infty, n}) + (1 - x_{n+1}^k) (1 - P(x_{n+1}^k = 1 \mid x_{-\infty, n})) \right). \quad (28)$$

Let us comment this result. First, (28) is the transition probability, of the form (19) where the normalised Gibbs potential  $\phi$  can be explicitly written, in terms of the synaptic interactions and the parameter  $\sigma_B$  controlling the noise amplitude. It is a general fact, for three reasons. First, in contrast to (19) where the memory of the spike sequence was fixed independent on  $x$ , here the memory depends on  $x$ , providing a variable length Markov chain [97,98]. Second,  $R(x)$ , has a function of  $x$  is unbounded.



One can find for all  $r \in \{-\infty, n\}$  a sequence  $x$  such that  $R(x) = r$  (take the sequences where all  $x_n^k = 0$ ,  $k = 1 \dots N$ ,  $n > r$  and  $x_r^k = 1$  for at least one  $k$ ). We have thus to deal with the extension of Markov chains, to chains with unbounded memory introduced in section 2.6. The existence and uniqueness of a Gibbs distribution compatible with this chain is guaranteed by the exponential decay of the memory controlled by  $\gamma < 1$  [96]. In this case the potential fulfils the conditions of described in section 2.6. Finally, in (28), the transition probabilities explicitly depend on time, because of the stimulus dependent term. They are therefore not translation invariant. While the extension to Gibbs distribution to non time-translation invariant chains can be rigorously done (upon the exponential decay of memory [70]), we restrict now to the case without stimulus ( $S_n^k = 0$ ,  $\forall k = 1 \dots N$ ,  $n \in \mathbb{Z}$ ) to apply the thermodynamic formalism, until section 4.5 where we discuss linear response.

Note that (27) bares strong analogy with the GLM Ansatz (20).

#### 4.4. Markov partition and symbolic coding

In this section we consider the deterministic discrete-time neuronal network model obtained considering (23) with  $\sigma_B = 0$ , studied in detail in [95]. The threshold  $\theta$  of the voltage in a network of  $N$  neurons induces a natural partition of  $\mathbb{R}^N$ ,  $\mathcal{P} = \prod_{k=1}^N \mathcal{P}_{x^k}$ , where  $x^k \in \{0, 1\}$ ,  $\mathcal{P}_0 = ]-B, \theta[$ ,  $\mathcal{P}_1 = [\theta, B[$  where the bound  $B$  depends on synaptic weights [95,99]. If  $V_n^k \in \mathcal{P}_0$  it evolves according to the sub-threshold dynamics (23) and it does not emit a spike at time  $n$ . In contrast, if  $V_n^k \in \mathcal{P}_1$  it emits a spike at time  $n$  and its trajectory is set back to  $\mathcal{P}_0$  at time  $n + 1$ . Thus,  $\mathcal{P}$  is a natural partition in the sense that it informs about the spikes of each neuron. Therefore, to each trajectory  $V \equiv \{V_n^k, k = 1 \dots N, n \in \mathbb{Z}\}$  is associated an infinite spike sequence  $x$  such that  $x_n^k = 0 \iff V_n^k \in \mathcal{P}_0$  and  $x_n^k = 1 \iff V_n^k \in \mathcal{P}_1$ .

This partition can be used to generate a Markov partition [95], but, in general, the Markov partition is not  $\mathcal{P}$  but a finite refinement  $\mathcal{Q}$  of  $\mathcal{P}$ . The reason ensuring that a Markov partition exists is that (25) is contracting. More precisely, it contracts, in one step, at speed  $\gamma$  for directions (neurons) such that  $V^k < \theta$ , and it contracts, with an infinite speed, for directions such that  $V^k \geq \theta$  (reset). This generates however discontinuities in the mapping (25), and a singularity set  $\mathcal{S} = \{V \in \mathbb{R}^N \mid \exists k \in \{1 \dots N\}, V^k = \theta\}$  where the map associated to (25), hereafter denoted by  $\mathcal{G}$ , is discontinuous. Thus  $\mathcal{G}$  is piecewise continuous and piecewise contracting.

Now, recall that  $\mathcal{Q}$  is a Markov partition for the dynamics with contracting map  $\mathcal{G}$  if its elements satisfy  $\mathcal{G}(\mathcal{Q}_n) \cap \mathcal{Q}_{n'} \neq \emptyset \Rightarrow \mathcal{G}(\mathcal{Q}_n) \subset \mathcal{Q}_{n'}$ . In other words, the image of  $\mathcal{Q}_n$  is included in  $\mathcal{Q}_{n'}$  whenever the transition  $n \rightarrow n'$  is legal. Here, in general, the elements of  $\mathcal{P}$  do not satisfy this condition. This is because the image of a domain of  $\mathcal{P}$  usually intersects several domains (in this case, the image intersects the singularity set). From the neural network's perspective this means that, in general, it is not possible to know the spiking pattern at time  $n + 1$  knowing the spiking pattern at time  $n$ . There are several possibilities depending on the membrane potential values and not only on the firing state of the neurons. Of course, if, say  $\mathcal{P}_n$  is such that  $\mathcal{G}(\mathcal{P}_n)$  intersects several domains  $\mathcal{P}_{n_1}, \dots, \mathcal{P}_{n_l}$  one can take the preimages of these domains  $\mathcal{G}^{-1}(\mathcal{P}_{n_r})$  to construct a refinement of  $\mathcal{P}$  such that the Markov partition requirement is satisfied, in one iteration of the map. However, nothing guarantees that, at the second iteration, some elements of this new partition will not intersect the singularity set under  $\mathcal{G}^2$ .

Can we find a finite refinement of  $\mathcal{P}$  such that the trajectory of the partition elements never intersects several partition elements? It is shown in [95] that this property is satisfied for generic values of the synaptic weights  $W_{ij}$ . Essentially, it is based on the fact that the distance between the  $\Omega$ -limit set of (25) and the singularity set  $\mathcal{S}$ , is generically positive. In other words, each point in the partition elements  $\mathcal{Q}_n$ , has a local stable manifold with a finite diameter.

As a consequence, the deterministic discrete-time neuronal network model (25) admits a Markov partition, a refinement of the natural partition, providing a symbolic coding of the membrane potential trajectories in terms of spike sequences. In particular, once the initial condition dependence has been

removed, the evolution (26) (without noise) is only constrained by the stimulus. Thus, (26) provides a coding scheme of the stimulus in terms of spike sequences. The Markov partition is made of spike blocks, with finite memory depth  $R$ , that can be used to apply the thermodynamic formalism in the presence of noise. However,  $R$  - the memory depth of the corresponding Markov chain - depends on the parameters, in particular the synaptic weights.

In addition, note that the presence of a singularity set induces a weak form of initial conditions. Although the dynamics is contracting, a small perturbation of a trajectory can induce an evolution drastically different from the unperturbed trajectory, if the perturbation crosses the singularity set. In this case, e.g. there is a neuron,  $k$ , which does not spike, at time  $n$  in the unperturbed trajectory, and spikes at time  $n$  in the perturbed one, inducing a completely different evolution (cascade effect). The singularity set also induces the existence of ghost orbits,  $\exists k \in \{1 \dots N\}$ ,  $\forall n > 0$ ,  $V_n^k < \theta$  and  $\limsup_{n \rightarrow +\infty} V_n^k = \theta$ . Ghost orbits are however non-generic in a topological and a measure-theoretic sense. As a corollary, the  $\Omega$ -set is generically composed of finitely many periodic orbits with a finite period (whose length depends on parameters of the model, in particular, synaptic weights).

#### 4.5. Extensions

##### 4.5.1. Explicit form of the potential. GLM vs MaxEnt.

The model (25) makes a link between the dynamics of a neuronal network and the transition probabilities (19), where the dependence in the model parameters (in particular, synaptic weights and stimuli) is explicit. We have for this model an explicit potential, which, here, takes a GLM-like form (27), but is more general, as in contrast to the GLM, the effective interactions depend on time via powers of the leak term  $\gamma$ . This potential can also be written in terms of monomials using the Hammersley-Clifford decomposition (3.3), through a series expansion of the function  $f$ . This procedure generate a series of monomials with coefficients that can be explicitly computed (using the fact that, from the monomials definition (21)  $(x_{i_k}^{n_k})^m = x_{i_k}^{n_k m}$ , for any integer  $m > 0$ ). These coefficients are proportional to powers of  $\gamma < 1$ , so their strength decay exponentially fast allowing to truncate the potential to a finite number of terms, which produce canonical Markov approximations of different orders [72]. One obtain, to the lowest order, a Bernoulli potential, then pairwise terms, and so on.

##### 4.5.2. Linear response.

Another interesting consequence of the analysis of this model is that the potential may depend on a time dependent stimulus. Considering that the stimulus is of small amplitude and additive, one can take a Taylor expansion of the potential as powers of the stimulus, allowing to go beyond the stationarity assumption central in equilibrium statistical mechanics and in the Thermodynamic Formalism. In this case, it is possible to show that the variations in the spike statistics induced by the stimulus, can be described in terms of a linear response theory [100]. The main result is that the variation, in the average of an observable  $f$ , resulting from the application of a stimulus reads:

$$\delta\mu[f](n) \equiv \mu^S[f](n) - \mu^{(sp)}[f] = (K_f * S)(n),$$

where  $\mu^{(sp)}$  is the Gibbs distribution in spontaneous activity (without stimulus), and  $\mu^S$  is the Gibbs distribution in the evoked activity regime (with stimulus),  $\mu^S[f](n)$  means the average of  $f$  with respect to  $\mu^S$ , which depends on time (if the stimulus does), and  $\mu^{(sp)}[f]$  means the average of  $f$  with respect to  $\mu^{(sp)}$ , which does not depends on time. This variation in average is given by a convolution between a kernel  $K_f$ , depending on  $f$  (which can be expressed in terms of time correlations functions between monomials) and of the stimulus. The coefficients in the expansion of  $K_f$  depend on the parameters constraining dynamics (e.g. the synaptic weights in (25)). The correlations are computed with respect to the invariant Gibbs measure  $\mu^{(sp)}$ . In addition, the influence of monomials in the

expansion decreases with their order, so that one can obtain a reasonable approximation of the convolution kernel considering only averages of order two monomials (time-dependent pairwise correlations). This is therefore a result in the form of a fluctuation-dissipation "theorem" in statistical physics, with the difference that one considers time dependent correlations. This formula has proven to give astonishingly good results when computing the response to a time dependent stimulus in the model (25) [100].

#### 4.6. The Galves-Löcherbach model

We present a second example where the Gibbs potential can be computed. This model is known as the Galves-Löcherbach model introduced by Antonio Galves and Eva Löcherbach in [71] (see also [101,102]). This model is a generalization of [16], but considering an infinite (countable) network of neurons interacting in time with memory of variable length.

The model is build considering a stochastic chain  $(X_t)_{t \in \mathbb{Z}}$  taking values in  $\{0,1\}^I$ , where  $I$  is a countable set of neurons. The probability of a spike depends on the accumulated activity of the system since the last spike, thus, each spike depends on a variable length history, defining also a non-Markovian stochastic process. Extensions of this model has been made considering the hydrodynamic limit of the interacting neuronal system [103], classifying collective behavior according to parameter values [104], and the generalization to the continuous time [105,106].

For each neuron  $i \in I$  and each time  $t \in \mathbb{Z}$  let  $\tau_i(x, t)$  denote the last time before  $t$  at which neuron  $i$  fired a spike in the spike train  $x$ :

$$\tau_i(x, t) = \sup \left\{ s < t : x_s^i = 1 \right\},$$

and suppose that the synaptic weights  $W_{ij}$  have the uniform summability property:

$$\sup_{i \in I} \sum_j |W_{ij}| < \infty.$$

The joint probabilities are conditionally independent given the past:

$$P \left( x_t^i \mid x_{t-1} \right) = \prod_{i \in I} P \left( x_t^i \mid x_{t-1} \right), \quad (29)$$

where the probability of neuron  $i$  having a spike at time  $t$  is given by:

$$P \left( x_t^i = 1 \mid x_{t-1} \right) = h_i \left( \sum_j W_{ij} \sum_{s=\tau_i(x,t)}^{t-1} g_j(t-s) x_s^j \right). \quad (30)$$

where  $g_j(t-s)$  plays the role of the exponential  $\alpha$ -kernel in (24). The transition probabilities (30) have thus the form of a GLM model.

Under technical conditions of the functions  $h_i$  and  $g_j$  and  $W_{ij}$ , there exists a unique probability measure consistent with (29) and (30) (see Theorem 1 of [71]). To prove this claim they use a Kalikow-type decomposition of the infinite order transition probabilities. This type of decomposition has also been considered in Ref [71,107,108]. The setup considered in this work extends the chains infinite size and infinite memory.

## 5. Discussion and perspectives

In this review we introduce different ideas and tools from the Thermodynamical Formalism and show how they can be applied in theoretical neuroscience. As a summary, we grouped these approaches depending on two main characteristics. The first one is the number of neurons  $N$  that affects the cardinality of the alphabet considered and the range  $R$  of the potential (memory in transition

probabilities) associated to the equilibrium measure characterising the system. This is represented in table 1. The infinite number of neurons and infinite range cases are further discussed in section 5.2.

| Thermodynamic formalism and Gibbs measures |                           |                             |                                  |
|--|---------------------------|-----------------------------|----------------------------------|
| Number of neurons                          | Memory of the potential   |                             |                                  |
|  | Memoryless                | Finite                      | Infinite                         |
| Finite                                     | Boltzmann-Gibbs           | Gibbs in the sense of Bowen | Chains with complete connections |
| Infinite                                   | Countable state Bernoulli | Countable state Markov      | Chains with variable length      |

**Table 1.** Types of Gibbs measures potentially found in experimental data analysis or in the analysis of mathematical models of networks of interacting spiking neurons.

While there has been interesting applications of the Thermodynamic Formalism to neuroscience, there are interesting ideas and developments still to come. In this concluding section we raise questions, challenges and new avenues for the applications of the Thermodynamic Formalism to theoretical neurosciences.

5.1. Thermodynamic Formalism for more biologically plausible neuron models

In this review, we have considered rather academic models of neuronal networks where, especially, time is considered discrete. There are good reasons for that. As we remarked at the beginning of section 3, we were considering models, like the Integrate-and-Fire, where spikes arise instantaneously, in continuous time, thereby providing a *possibly uncountable* set of potential spike trains. The question is whether we are dealing here with a realistic property of biological neuronal networks, or, with an artifact created by the instantaneous reset. Real spikes have a duration (a few ms) and a refractory period (also a few ms), so for a fixed initial condition, spikes trains produced by a continuous-time neuronal network are countable. Now, it might be that the set of spike trains depend continuously on the initial condition so that we are still left with an uncountable set of spikes.

It is out of the scope of this review to discuss from a biological perspective, whether or not neuronal networks have the cardinality of the continuum (See [99,109,110] for a discussion on this topic). Instead, we have considered a strategy consisting of discretizing time, avoiding the problem of potentially uncountable spikes trains. Here, we briefly mention another strategy, allowing to associate to continuous time networks a countable set of spikes trains. Here, we briefly mention continuous-time networks with a countable set of spike trains. We first make the remark that the instantaneous reset of voltage is physical and biological nonsense, inducing pathologies in the dynamics [111]. On this basis, we use a convenient mathematical trick explained in the next paragraph, which can certainly be criticized on phenomenological grounds [110], especially when the dynamical system representing the neuronal activity is deterministic.

After spiking a biological neuron stays at rest a certain time (refractory period). So, the trick is the following. Fix  $\delta > 0$  and define a spiking variable  $x_n^k \in \{0, 1\}$ , where  $n$  is an integer, where  $x_n^k = 1$  if neuron  $k$  emits a spike in the time interval  $[n\delta, (n + 1)\delta[$  and  $x_n^k = 0$  otherwise. Recall that  $t_{(r)}^k$  denote the time at which neuron  $k$  emits its  $r$ -th spike . This reads:

$$x_n^k = \begin{cases} 1, & \text{if } \exists r, t_{(r)}^k \in [n\delta, (n + 1)\delta[; \\ 0, & \text{otherwise.} \end{cases}$$

Spiking variables are therefore time-discrete events with a time resolution  $\delta$ . When  $V^k$  reaches the threshold at time  $t_{(r)}^k$  it is reset to 0, and stays there until time  $(n + 1)\delta$ . After this, follows the sub-threshold evolution (23) until the next time where  $V^k$  reaches the threshold. Note that, in this modelling,  $\delta$  can be quite small compared to the time scales of the dynamics. In this way, the set of spike trains  $x$  becomes at most countable.

This trick can be used to generalise the Integrate-and-Fire model to a conductance-based Integrate-and-Fire model, introduced by Rudolph and Destexhe in [112] and mathematically studied in [17,18,99], where the synaptic conductance depends on the spike history. One can still show that there exists a unique Gibbs distribution with infinite range potential characterizing the spike train statistics in this case. The potential can be explicitly computed as a function of network parameters, even in the presence of a time-dependent stimulus.

Now, would the Thermodynamic Formalism apply to more realistic neuronal models like Hodgkin-Huxley [92], FitzHugh-Nagumo [113,114], Morris-Lecar models [115] ? (see [31,90,116] for a complete presentation of canonical neuronal models). In these models, closer to biology, the spike has a time course, thus are not considered as point events. However, we do not know any result establishing, e.g., the existence of Markov partitions and symbolic coding for these models and this seems to be out of reach for the moment. Still, one can bin the time and proceed as done in experiments where voltage is a time-continuous signal. Thus, one can still use the approach used in (19) to characterise the spike train statistics.

A natural question in this context is what is the link between spikes train statistics and the underlying dynamical model, with "hidden" dynamical variables such as membrane potential, but also, e.g. activation/inactivation variables? If we think in terms of spike coding, the alternative is the following. Either spike trains contain all the necessary information to characterise the dynamics, e.g. the spike response to a stimulus, and then characterizing (19) is sufficient. Or, there is additional information, not conveyed by spikes (e.g. sub-threshold oscillations [78,117,118]), and the "neural code" is not entirely contained in the spikes, somewhat ruining the hope of encoding neuronal messages purely in terms of spikes. This question is quite more general than the validity of Thermodynamic Formalism approach for these models.

What Thermodynamic Formalism brings to the analysis of these models is (1) a way to rigorously handle probabilistic representations of spikes (19); (2) To provide conceptual and mathematical tools to analyze spiking neuronal network models like (25), where a dynamical system formulation of biophysical variables can be mathematically related to spike coding and spike train statistics.

## 5.2. Phase transitions

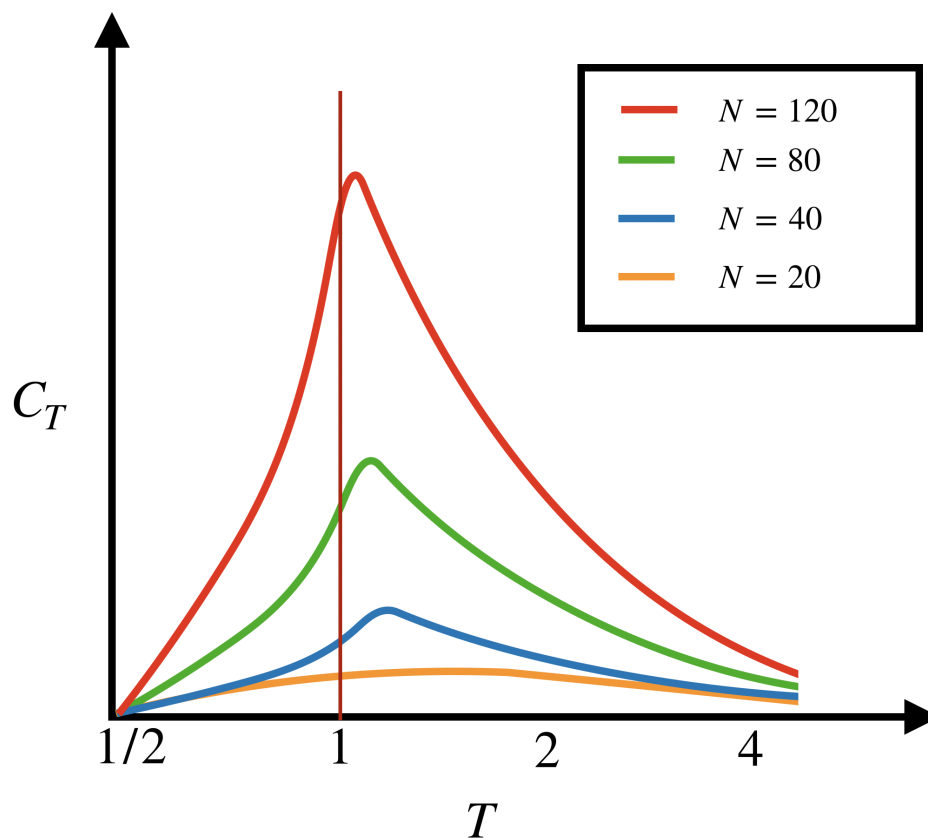
Several studies have shown that the population of vertebrate retinal ganglion cells responding to naturalistic stimulus is poised near a "critical state" [54,55]. From the maximum entropy joint distribution (13), a family of Gibbs distributions can be built introducing a parameter  $1/\beta$  (analogous to the inverse temperature), which scales all the Lagrange multipliers of the inferred Hamiltonian. When  $\beta \rightarrow 0$  (infinite temperature) the uniform distribution is obtained, and when  $\beta \rightarrow +\infty$  (zero temperature), the Dirac delta supported at the spike configuration(s) of minimal energy is obtained.  $1/\beta = 1$  corresponds to the inferred maximum entropy distribution from data. These studies have only analysed memoryless Gibbs distributions (13).

From this representation it is possible to compute the fluctuations (variance) of the energy  $U_\lambda$  as a function of the "temperature" parameter  $T$ . This quantity can be obtained as the second derivative of the pressure, eq. (6), which is, in thermodynamics, related to the heat capacity  $C_T$ . On numerical grounds, this quantity can be computed using MonteCarlo simulations and plotted as a function of the "temperature"  $T = \frac{1}{\beta}$ , for different network sizes, (see Fig 3).

The form of  $C_T$  versus  $T$  plot, for maximum entropy models of Ising type obtained from the recording of retinal ganglion cells responding to naturalistic stimuli are shown in Fig. 3 (redrawn from [55]). It can be observed that there is a clear, increasing, peak, at  $T = 1$ , which starts to manifest when larger and larger groups of neurons are considered. This presumable divergence of the heat capacity (or variance of  $U$ ) when  $N \rightarrow \infty$  (thermodynamic limit) is interpreted as a second order phase transition, (a so-called "critical regime" [119]).

The behavior of the specific heat observed in Fig. 3 suggests that the heat capacity of a maximum entropy distribution, fitted over an increasingly large group of neurons in the retina diverges. This

phenomenon has been considered a “signature of criticality” (details of this study and a discussion about whether criticality is functional for retinal ganglion cells can be found in Ref. [55]). Some criticism about this approach to diagnostic criticality has appeared arguing that the maximum entropy principle is likely to yield models that are close to singular values of parameters, akin to critical points in physics where phase transitions occur. Statistically distinguishable models tend to accumulate close to critical points, where the susceptibility (inverse Fisher Information) diverges in infinite systems [120]. These ideas have been also applied to numerical simulations of a canonical feed-forward population model showing that the specific heat diverges whenever the average correlation strength is independent of the population size [56], as in the random subsampling of correlations used in [55]. Also note that, for spike trains obtained from discrete Markov processes, binning generates a stochastic process with unbounded memory akin to induce spurious phase transitions [121].



**Figure 3. Signatures of criticality** Generic plot of heat capacity  $C_T$  versus temperature  $T$  for maximum entropy models built constraining firing rates and pairwise correlations of retinal ganglion cells responding to naturalistic stimuli [55]. A clear peak appear at  $T = 1$  when groups of increasingly large number of neurons are considered (thermodynamic limit).

This interesting approach leads nevertheless to several questions in the context of the Thermodynamic Formalism.

- **Does this signature of criticality** extends to Gibbs distributions with potentials of range  $R > 1$ , i.e. with memory? How does it depend on  $R$ ? We don't know about any experimental results addressing this issue. This question is related to the following:
- **What is this signature of criticality from the point of view of the Thermodynamic Formalism?** The occurrence of a second-order phase transition mathematically means that the pressure is  $C^1$  but not  $C^2$  when some limit is taken. Here, we have two possible limits: the range of the potential  $R$  tends to infinity or the number of neurons  $N$  tends to infinity. These two limits could also be addressed simultaneously and they do not necessarily commute. For finite  $R, N$  potentials



the Perron-Frobenius theorem guarantees the existence, uniqueness of the Gibbs measure and the analyticity of the pressure can be proved, preventing phase transitions. When  $R$  or  $N$  are infinite, the properties of the RPF operator (2.3.2) characterizes the presence or absence of phase transitions. Indeed, there are conditions ensuring a spectral gap for this operator, ensuring the exponential decay of correlations. Now, eq. (7) characterise the second derivative of the pressure as a time series of correlations which converges when the correlations decay exponentially. On the opposite, the non-summability of time correlations function implies the non-existence of the second derivative, thus, of a second-order phase transition. Therefore, a possibility to have a second-order phase transition is when the spectral gap property for the RPF operator when  $R \rightarrow +\infty$  or  $N \rightarrow +\infty$  is absent. In statistical mechanics, second-order phase transitions can be characterised by the way how the zeros of the partition function, written as a polynomial, pinch the real axis (Lee-Yang phenomenon) [122–124]. In our case, when  $R > 1$ , the object of interest is not the partition function but the largest eigenvalue of  $\mathcal{L}_\phi$  which has to stay analytic in the limit  $R$ , or  $N$ ,  $\rightarrow +\infty$ . The absence of the spectral gap property presents an analogy with the Lee-Yang phenomenon although we do not know about results establishing a deeper link.

- **Can we relate known examples of dynamical systems exhibiting phase transitions to models in neuroscience?** Another possible example to be interpreted in neuroscience is the Dyson model [125], in which there exists a phase transition in the sense of spontaneous magnetisation when the temperature goes to zero, due to an infinite range potential whose correlation does not decay exponentially fast. In our case, the range of the potential should be taken in time, keeping (possibly) the number of neurons finite. Other examples exist of rigorous characterizations of phase transitions in the thermodynamic description of Pomeau-Manneville intermittent maps, passing from an integrable density function associated with the measure to heavy-tailed densities [126]. An interesting result may hint the connection between the topological Markov map of the interval and stochastic chains of infinite order or chains with complete connections. In [127] is presented how to build a topological Markov map of the interval whose invariant probability measure is the stationary law of a given stochastic chain of infinite order. This is interesting in this context as we presented in 25 there are mathematical models of spiking neurons whose spike statistics is represented by chains of infinite order. This result or its inverse may hint to conditions in the parameters or conditions of the mathematical models of spiking neurons to exhibit second order phase transitions.
- **What could be the dynamical or mechanistic origins of a second-order phase transition in a spiking neuronal network model?** Handling experimental data is of course important, but for long experiments with living neuronal tissue, one cannot control the size of the sample, the stationarity of data, and so on. So, assume that we have been able to find an example of a Gibbs distribution exhibiting a second-order phase transition when  $R \rightarrow +$  or  $N \rightarrow +\infty$ . Can we build a spiking dynamical system, with finite  $R, N$ , which has this Gibbs distribution in the limit  $R$ , or  $N$ ,  $\rightarrow +\infty$  we observe a phase transition in the model? Then, what are the mechanistic origins (in the neuronal dynamics) of second-order phase transitions? It could be an interesting example to study the existence of a second-order phase transition in a simple neuronal model. Returning back to the discrete LIF model, the failure in the second-order differentiability of the pressure means the loss of exponential mixing, which, in the model (25) can arise in, at least, two cases. First, if  $\gamma = 1 - \epsilon$ ,  $\epsilon \rightarrow 0$ . This is a way to obtain a potential with increasing range as  $\epsilon \rightarrow 0$  with loss of summability of correlations. The corresponding orbits (reminiscent of the ghost orbits discussed in section 4.4) are such that it may take a long time for some neurons to be reset. Thereby, the memory to be considered is very long. This is however a case hardly interpretable from the neuroscience perspective. A second possibility is to analyse how the pressure depends on the spectrum of the synaptic weights matrix and to check whether there are cases (e.g. small world or scale-free lattices) where the spectral gap of the RPF vanishes.

### 5.3. What do Thermodynamic Formalism and Gibbs distribution may tell us about neuroscience?

One of the biggest challenges in science of the XXI century is to understand the brain functions within a conceptual framework capable to unify the multi-scale dynamics that take place in the brain. This framework should also make sense in the light of the overwhelming amount of experimental data and capable to predict macroscopic phenomena such as motor behaviour or visual experience from the activity of billions of neurons.

The relation between mathematics and physics has been historically symbiotic and the Thermodynamic Formalism is an interesting example of how ideas from physics may help to solve problems and introduce ideas into the field of mathematics. The history of the Thermodynamic Formalism also shows how purely mathematical results can be obtained as corollaries of physical laws, inverting the frequently assumed relation between physics and mathematics [128].

However, in the case considered in this review, the mathematical problem is motivated by biology, not by physics. While Eugene Wigner argues in favour of the “The unreasonable effectiveness of mathematics in the natural sciences” [129], Israel Gelfand, after spending several years working in mathematical problems related to biology replied with “The unreasonable ineffectiveness of mathematics in biology.” [130]. While there are reasons to argue that this is still the case, it is less clear if one can blame the field of mathematics or just the fact that we have not yet used the right tools or frameworks. There is a long tradition of using ideas from statistical physics to study neural networks, in particular, to represent the emergence of collective behaviour from the microscopic interactions, with the hope that statistical aspects of the collective behaviour will be independent of the details in these systems. This gave rise to major branches of theoretical neuroscience like dynamic mean-field methods [131–134] or Maximum Entropy approach [32,43,135,136]. There are considerably less articles using mathematical methods to analyse rigorously the collective behaviour of neuronal networks. Some methods used to the analysis are large deviations [137,138], Kalikow-type decomposition [71,139], stochastic processes [140,141,141–144], dynamical systems [116,145], etc.

Physicist have been able to make a deep connection between mechanics, statistical physics and thermodynamics. A similar quest is presumably guiding the research of (some) theoretical and experimental neuroscientists. While there still a long way from achieving this goal (as some argue we are still searching for principles [146]), during the last decades, mathematicians are playing a relevant role in the rigorous description of neural phenomena clarifying and rising conceptual problems in neuroscience.

From this perspective, the interest of Thermodynamic Formalism lays, certainly, in its main theorems and hypotheses. The analysis of these in the context of theoretical neuroscience may provide an original way to formulate conditions under which a spiking neuronal network could display a critical state, or when the way to handle data induce so severe bias in the spike statistics that they could induce spurious evidences of criticality.

We can hope that theoretical tools and ideas from the Thermodynamic Formalism and its current application in neuronal dynamics and spike train statistics should lead to a better and unified understanding of the neural phenomena. We also hope that the present review may serve as an encouragement for the mathematical community interested in applications of the thermodynamic formalism to study these interesting and important problems.

**Acknowledgments:** R.C. was supported by Fondecyt Proyecto 11181072 and CONICYT REDES Project No. 180151. CM acknowledges CONACYT (Mexico) for financial support through project No. A1-S-15528 and partially supported by the CONICYT REDES Project No. 180151. CM also wish to thank the CIMFAV for the wonderful hospitality and facilities during his stay. B.C. and R.C. acknowledge the INRIA program “Associated teams” for funding part of the project via the associated team MAGMA <https://team.inria.fr/biovision/associated-team-magma/>.

**Author Contributions:** The authors conceived the main ideas and concepts, wrote and revised the manuscript. All authors have read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                                     |
|------|-------------------------------------|
| KL   | Kullback-Leibler                    |
| MEP  | Maximum Entropy Principle           |
| RPF  | Ruelle-Perron-Frobenius             |
| LDP  | Large Deviation Principle           |
| SCGF | Scaled Cumulant Generating Function |
| GLM  | Generalized Linear Model            |
| LIF  | Leaky Integrate-and-Fire            |
| MEA  | Multi-Electrode Arrays              |

## Symbol list

|                     |   |
|---------------------|---|
| $x_n^k$             | Spike-state of neuron $k$ at time $n$                                     |
| $x_n$               | Spike pattern at time $n$   |
| $x_{n_1, n_2}$      | Spike block from time $n_1$ to $n_2$                                      |
| $A_{n, n+m}$        | Configuration space of spike blocks of $m$ spike patterns                 |
| $A_R^N$             | Configuration space of $N$ neurons and spike blocks of $R$ spike patterns |
| $\mathbb{E}_\nu(f)$ | Expected value of the observable $f$ w.r.t. the probability measure $\nu$ |
| $A_T(f)$            | Empirical average of the observable $f$ considering $T$ spike patterns    |
| $H[\mu]$            | Entropy of the probability measure $\mu$                                  |
| $\lambda_k$         | Lagrange multiplier parameter   |
| $U_\lambda$         | Potential or Energy function  |
| $F[U_\lambda]$      | Pressure associated to the potential $U_\lambda$                          |
| $\mu_\psi$          | Equilibrium measure associated to the potential $\psi$                    |
| $S_n\phi$           | Birkhoff sums associated to the potential $\phi$                          |
| $\Gamma_f$          | Scaled cumulant generating function of the observable $f$                 |
| $I_f$               | Rate function of the observable $f$                                       |
| $\mathcal{L}_\phi$  | Ruelle-Perron-Frobenius operator associated to the potential $\phi$       |
| $w_n$               | Integer associated to the spike block $x_{n, n+R-1}$                      |
| $C_{f,g}(n)$        | Correlation function between the observables $f$ and $g$ at time $n$      |
| $m_l$               | Monomial $l$  |
| $C_T$               | Heat capacity   |
| $V^k$               | Voltage of neuron $k$   |
| $\theta$            | Threshold   |

1. Gallavotti, G. Statistical Mechanics: A Short Treatise; Theoretical and Mathematical Physics, Springer Berlin Heidelberg, 1999.
2. Gallavotti, G. Nonequilibrium and Irreversibility; Springer Publishing Company, Incorporated, 2014.
3. Gaspard, P. Chaos, scattering and statistical mechanics; Vol. 9, Cambridge Non-Linear Science series, 1998.
4. Ruelle, D. Statistical Mechanics: Rigorous results; New York, 1969.
5. Georgii, H.O. Gibbs measures and phase transitions; De Gruyter Studies in Mathematics:9, Berlin; New York, 1988.
6. Sinai, Y.G. Gibbs measures in ergodic theory. Russian Mathematical Surveys **1972**. doi:10.1070/rm1972v027n04abeh001383.
7. Ruelle, D. Thermodynamic formalism; Addison-Wesley, Reading, Massachusetts, 1978.
8. Bowen, R. Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Second revised version.; Vol. 470, Lect. Notes in Math., Springer-Verlag, 2008.
9. Ash, R.; Doleans-Dade, C. Probability and Measure Theory, 2nd Edition; Academic Press, 1999.
10. Friedli, S., V.Y. Statistical Mechanics of Lattice Systems: a Concrete Mathematical Introduction; Cambridge University Press: Cambridge, 2017.
11. Young, L.S. Statistical properties of dynamical systems with some hyperbolicity. Ann. of Math. (2) **1998**, 147, 585–650.

12. Climenhaga, V.; Pesin, Y. Building thermodynamics for non-uniformly hyperbolic maps. Arnold Mathematical Journal **2017**, *3*, 37–82.
13. Dementrius, L. The Thermodynamic Formalism in Population Biology. Numerical Methods in the Study of Critical Phenomena; Della Dora, J.; Demongeot, J.; Lacolle, B., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1981; pp. 233–253.
14. Demetrius, L. Statistical mechanics and population biology. Journal of Statistical Physics **1983**, *30*, 709–753.
15. Cessac, B.; Blanchard, P.; Krüger, T.; Meunier, J.L. Self-Organized Criticality and thermodynamic formalism. Journal of Statistical Physics **2004**.
16. Cessac, B. Statistics of spike trains in conductance-based neural networks: Rigorous results. The Journal of Mathematical Neuroscience **2011**, *1*, 1–42. doi:10.1186/2190-8567-1-8.
17. Cessac, B. A discrete time neural network model with spiking neurons II. Dynamics with noise. Journal of Mathematical Biology **2011**, *62*, 863–900. doi:10.1007/s00285-010-0358-4.
18. Cofré, R.; Cessac, B. Dynamics and spike trains statistics in conductance-based Integrate-and-Fire neural networks with chemical and electric synapses. Chaos, Solitons and Fractals **2013**, *50*, 13–31.
19. Cofré, R.; Maldonado, C. Information Entropy Production of Maximum Entropy Markov Chains from Spike Trains. Entropy **2018**, *20*.
20. Galves, A.; Galves, C.; García, J.E.; Garcia, N.L.; Leonardi, F. Context tree selection and linguistic rhythm retrieval from written texts. The Annals of Applied Statistics **2012**, *6*, 186–209.
21. Cofré, R.; Maldonado, C.; Rosas, F. Large Deviations Properties of Maximum Entropy Markov Chains from Spike Trains. Entropy **2018**, *20*.
22. Cofré, R.; Videla, L.; Rosas, F. An Introduction to the Non-Equilibrium Steady States of Maximum Entropy Spike Trains. Entropy **2019**. doi:10.3390/e21090884.
23. Sompolinsky, H.; Crisanti, A.; Sommers, H. Chaos in random neural networks. Phys. Rev. Lett. **1988**, *61*, 259–262.
24. Buice, M.A.; Chow, C.C. Beyond mean field theory: Statistical field theory for neural networks. Journal of Statistical Mechanics: Theory and Experiment **2013**. doi:10.1088/1742-5468/2013/03/P03003.
25. Montbrió, E.; Pazó, D.; Roxin, A. Macroscopic Description for Networks of Spiking Neurons. Physical Review X **2015**, *5*, 021028.
26. Byrne, A.; Avitabile, D.; Coombes, S. Next-generation neural field model: The evolution of synchrony within patterns and waves. Phys. Rev. E **2019**, *99*, 012313. doi:10.1103/PhysRevE.99.012313.
27. Chizhov, A.V.; Graham, L.J. Population model of hippocampal pyramidal neurons, linking to refractory density approach to conductance-based neurons. Phys. rev. E **2007**, *75*, 114.
28. Brunel, N.; Hakim, V. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. Neural Computation **1999**, *11*, 1621–1671.
29. Rieke, F.; Warland, D.; de Ruyter van Steveninck, R.; Bialek, W. Spikes: Exploring the Neural Code; Bradford Books, 1997.
30. Abbott, L.; Dayan, P. The effect of correlated variability on the accuracy of a population code. Neural Comput **1999**, *11*, 91–101.
31. Gerstner, W.; Kistler, W. Spiking Neuron Models; Cambridge University Press, 2002.
32. Schneidman, E.; Berry II, M.; Segev, R.; Bialek, W. Weak pairwise correlations imply strong correlated network states in a neural population. Nature **2006**, *440*, 1007–1012.
33. Shlens, J.; Field, G.D.; Gauthier, J.L.; Grivich, M.I.; Petrusca, D.; Sher, A.; Litke, A.M.; Chichilnisky, E.J. The structure of multi-neuron firing patterns in primate retina. J Neurosci **2006**, *26*, 8254–66.
34. Tkačik, G.; Prentice, J.; Balasubramanian, V.; Schneidman, E. Optimal population coding by noisy spiking neurons. PNAS **2010**, *107*, 14419–14424.
35. Ganmor, E.; Segev, R.; Schneidman, E. The architecture of functional interaction networks in the retina. The journal of neuroscience **2011**, *31*, 3044–3054.
36. Ganmor, E.; Segev, R.; Schneidman, E. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. PNAS **2011**, *108*, 9679–9684.
37. Tkacik, G.; Marre, O.; Mora, T.; Amodei, D.; II, M.J.B.; Bialek, W. The simplest maximum entropy model for collective behavior in a neural network. Journal of Statistical Mechanics: Theory and Experiment **2013**, *2013*, P03011. doi:10.1088/1742-5468/2013/03/p03011.

38. Granot-Atedgi, E.; Tkačik, G.; Segev, R.; Schneidman, E. Stimulus-dependent Maximum Entropy Models of Neural Population Codes. *PLOS Computational Biology* **2013**, *9*, 1–14. doi:10.1371/journal.pcbi.1002922.
39. Nghiem, T.A.; Telenczuk, B.; Marre, O.; Destexhe, A.; Ferrari, U. Maximum-entropy models reveal the excitatory and inhibitory correlation structures in cortical neuronal activity. *Phys. Rev. E* **2018**, *98*, 012402. doi:10.1103/PhysRevE.98.012402.
40. Gardella, C.; Marre, O.; Mora, T. Modeling the correlated activity of neural populations: A review. *Neural computation* **2019**, *31*, 233–269.
41. Ferrari, U.; Deny, S.; Chalk, M.; Tkačik, G.; Marre, O.; Mora, T. Separating intrinsic interactions from extrinsic correlations in a network of sensory neurons. *Physical Review E* **2018**, *98*, 042410.
42. Marre, O.; El Boustani, S.; Frégnac, Y.; Destexhe, A. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Physical review letters* **2009**, *102*.
43. Vasquez, J.; Palacios, A.; Marre, O.; Berry II, M.; Cessac, B. Gibbs distribution analysis of temporal correlation structure on multicell spike trains from retina ganglion cells. *J. Physiol. Paris* **2012**, *106*, 120–127.
44. Gardella, C.; Marre, O.; Mora, T. Blindfold learning of an accurate neural metric. *Proceedings of the National Academy of Sciences of the United States of America* **2018**, [1710.05067]. doi:10.1073/pnas.1718710115.
45. Beggs, J.M.; Plenz, D. Neuronal Avalanches in Neocortical Circuits. *Journal of Neuroscience* **2003**, *23*, 11167–11177. doi:10.1523/JNEUROSCI.23-35-11167.2003.
46. Haldeman, C.; Beggs, J. Critical Branching Captures Activity in Living Neural Networks and Maximizes the Number of Metastable States. *Physical review letters* **2005**, *94*, 058101. doi:10.1103/PhysRevLett.94.058101.
47. Kinouchi, O.; Copelli, M. Optimal dynamical range of excitable networks at criticality. *Nature Physics* **2006**, *2*, 348–351.
48. Shew, W.L.; Yang, H.; Petermann, T.; Roy, R.; Plenz, D. Neuronal Avalanches Imply Maximum Dynamic Range in Cortical Networks at Criticality. *Journal of Neuroscience* **2009**, *29*, 15595–15600. doi:10.1523/JNEUROSCI.3864-09.2009.
49. Shew, W.L.; Plenz, D. The Functional Benefits of Criticality in the Cortex. *The Neuroscientist* **2013**, *19*, 88–100. doi:10.1177/1073858412445487.
50. Gautam, S.H.; Hoang, T.T.; McClanahan, K.; Grady, S.K.; Shew, W.L. Maximizing Sensory Dynamic Range by Tuning the Cortical State to Criticality. *PLOS Computational Biology* **2015**, *11*, 1–15. doi:10.1371/journal.pcbi.1004576.
51. Girardi-Schappo, M.; Bortolotto, G.S.; Gonsalves, J.J.; Pinto, L.T.; Tragtenberg, M.H.R. Griffiths phase and long-range correlations in a biologically motivated visual cortex model. *Scientific Reports* **2016**, *6*, 29561.
52. Touboul, J.; Destexhe, A. Can Power-Law Scaling and Neuronal Avalanches Arise from Stochastic Dynamics? *PLOS ONE* **2010**, *5*, 1–14. doi:10.1371/journal.pone.0008982.
53. Cocchi, L.; Gollo, L.L.; Zalesky, A.; Breakspear, M. Criticality in the brain: A synthesis of neurobiology, models and cognition. *Progress in Neurobiology* **2017**, *158*, 132 – 152. doi:https://doi.org/10.1016/j.pneurobio.2017.07.002.
54. Mora, T.; Bialek, W. Are biological systems poised at criticality? *J Stat Phys* **2011**, *144*.
55. Tkačik, G.; Mora, T.; Marre, O.; Amodei, D.; Berry II, M.; Bialek, W. Thermodynamics for a network of neurons: Signatures of criticality. *PNAS* **2015**, *112*.
56. Nonnenmacher, M.; Behrens, C.; Berens, P.; Bethge, M.; Macke, J.H. Signatures of criticality arise from random subsampling in simple population models. *PLoS Computational Biology* **2017**. doi:10.1371/journal.pcbi.1005718.
57. Chazottes, J.; Keller, G., *Mathematics of Complexity and Dynamical Systems*; Springer, 2011; chapter Pressure and Equilibrium States in Ergodic Theory, pp. 1422–1437.
58. Keller, G. *Equilibrium States in Ergodic Theory*; Cambridge University Press, 1998.
59. Sarig, O.M. Thermodynamic formalism for countable Markov shifts. *Ergodic Theory and Dynamical Systems* **1999**. doi:10.1017/S0143385799146820.
60. Katok, A.; Hasselblatt, B. *Introduction to the modern theory of dynamical systems*; Kluwer, 1998.
61. Baladi, V. *Positive Transfer Operators and Decay of Correlations*; Vol. 16, World Scientific, 2000.
62. Shields, P.C. *The ergodic theory of discrete sample paths*; Vol. 13, *Graduate Studies in Mathematics*, American Mathematical Society: Providence, RI, 1996; pp. xii+249.



63. Mayer, V.; Urbański, M. Thermodynamical formalism and multifractal analysis for meromorphic functions of finite order. *Memoirs of the American Mathematical Society* **2010**. doi:10.1090/s0065-9266-09-00577-8.
64. Kubo, R. Statistical-mechanical theory of irreversible processes. *J. Phys. Soc. (Japan)* **1957**, *12*, 570–586.
65. Kubo, R. The fluctuation-dissipation theorem. *Reports on Progress in Physics* **1966**, *29*, 255–284. doi:10.1088/0034-4885/29/1/306.
66. Jaeger, G. The Ehrenfest Classification of Phase Transitions: Introduction and Evolution. *Archive for History of Exact Sciences* **1998**, *53*, 51–81. doi:10.1007/s004070050021.
67. Dunford, N.; Schwartz, J. *Linear operators: Spectral operators*; Vol. 7, Wiley-Interscience, 1988.
68. Ellis, R. *Entropy, Large deviations and Statistical Mechanics*; Springer, Berlin, 1985.
69. Touchette, H. The large deviation approach to statistical mechanics. *Physics Reports* **2009**, *478*, 1–69.
70. Fernandez, R.; Maillard, G. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **2005**, *118*, 555–588.
71. Galves, A.; Löcherbach, E. Infinite Systems of Interacting Chains with Memory of Variable Length-A Stochastic Model for Biological Neural Nets. *Journal of Statistical Physics* **2013**. doi:10.1007/s10955-013-0733-9.
72. Fernández, R.; Galves, A. Markov approximations of chains of infinite order. *Bull. Braz. Math. Soc. (N.S.)* **2002**, *33*, 295–306. doi:10.1007/s005740200015.
73. Ruelle, D. Statistical mechanics of a one-dimensional lattice gas. *Commun. Math. Phys.* **1968**, *9*, 267–278.
74. Ny, A.L. Introduction to (Generalized) Gibbs Measures. *Ensaios Matematicos* **2008**, *15*.
75. Fernandez, R.; Gallo, S. Regular g-measures are not always Gibbsian. *Electron. Commun. Probab.* **2011**, *16*, 732–740.
76. Yger, P.; Spampinato, G.L.; Esposito, E.; Lefebvre, B.; Deny, S.; Gardella, C.; Stimberg, M.; Jetter, F.; Zeck, G.; Picaud, S.; Duebel, J.; Marre, O. A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *eLife* **2018**, *7*, e34518. doi:10.7554/eLife.34518.
77. Buccino, A.P.; Hurwitz, C.L.; Magland, J.; Garcia, S.; Siegle, J.H.; Hurwitz, R.; Hennig, M.H. SpikeInterface, a unified framework for spike sorting. *bioRxiv* **2019**. doi:10.1101/796599.
78. Stiefel, K.M.; Fellous, J.M.; Thomas, P.J.; Sejnowski, T.J. Intrinsic subthreshold oscillations extend the influence of inhibitory synaptic inputs on cortical pyramidal neurons. *European Journal of Neuroscience* **2010**. doi:10.1111/j.1460-9568.2010.07146.x.
79. Rieke, F.; Warland, D.; de Ruyter van Steveninck, R.; Bialek, W. *Spikes, Exploring the Neural Code*; M.I.T. Press, 1996.
80. Cessac, B.; Paugam-Moisy, H.; Viéville, T. Overview of facts and issues about neural coding by spikes. *J. Physiol. Paris* **2010**, *104*, 5–18.
81. Cessac, B.; Le Ny, A.; Löcherbach, E. On the mathematical consequences of binning spike trains. *Neural Computation* **2017**, *29*, 146–170.
82. Pillow, J.W.; Shlens, J.; Paninski, L.; Sher, A.; Litke, A.; Chichilnisky, E.J.; Simoncelli, E. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature* **2008**, *454*, 995–999.
83. Cessac, B.; Cofré, R. Spike train statistics and Gibbs distributions. *J. Physiol. Paris* **2013**, *107*, 368–368.
84. Hammersley, J.M.; Clifford, P. Markov fields on finite graphs and lattices. *unpublished* **1971**.
85. Moussouris, J. Gibbs and Markov Random Systems with Constraints. *J. Stat. Phys.* **1974**, *10*, 11–33.
86. Cofré, R.; Cessac, B. Exact computation of the maximum entropy potential of spiking neural networks models. *Physical Review E* **2014**, *107*, 368–368.
87. Herzog, R.; Escobar, M.J.; Cofre, R.; Palacios, A.G.; Cessac, B. Dimensionality Reduction on Spatio-Temporal Maximum Entropy Models of Spiking Networks. *bioRxiv* **2018**. doi:10.1101/278606.
88. Nasser, H.; Marre, O.; Cessac, B. Spatio-temporal spike trains analysis for large scale networks using maximum entropy principle and Monte-Carlo method. *Journal Of Statistical Mechanics* **2013**.
89. Cessac, B.; Kornprobst, P.; Kraria, S.; Nasser, H.; Pamplona, D.; Portelli, G.; Viéville, T. PRANAS: A New Platform for Retinal Analysis and Simulation. *Frontiers in Neuroinformatics* **2017**, *11*, 49. doi:10.3389/fninf.2017.00049.
90. Ermentrout, G.B.; Terman, D.H. *Mathematical Foundations of Neuroscience*, 1st edition. ed.; Springer, 2010.
91. Lapicque, L. Recherches quantitatives sur l'excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen.* **1907**, *9*, 620–635.



92. Hodgkin, A.; Huxley, A. A quantitative description of membrane current and its application to conduction and excitation in nerve cells. *Journal of Physiology* **1952**, *117*, 500–544.
93. Destexhe, A.; ZF, Z.M.; Sejnowski, T. *Kinetic models of synaptic transmission*; Cambridge, MA: MIT Press, 1998; pp. 1–26.
94. Soula, H.; Beslon, G.; Mazet, O. Spontaneous dynamics of asymmetric random recurrent spiking neural networks. *Neural Computation* **2006**, *18*.
95. Cessac, B. A discrete time neural network model with spiking neurons. Rigorous results on the spontaneous dynamics. *J. Math. Biol.* **2008**, *56*, 311–345.
96. Cessac, B. A discrete time neural network model with spiking neurons II. Dynamics with noise. *J. Math. Biol.* **2011**, *62*, 863–900.
97. Variable length Markov chains. *Annals of Statistics* **1999**. doi:10.1214/aos/1018031204.
98. Mächler, M.; Bühlmann, P. Variable length Markov chains: Methodology, computing, and software. *Journal of Computational and Graphical Statistics* **2004**. doi:10.1198/1061860043524.
99. Cessac, B.; Viéville, T. On Dynamics of Integrate-and-Fire Neural Networks with Adaptive Conductances. *Frontiers in neuroscience* **2008**, *2*.
100. Cessac, B.; Ampuero, I.; Cofre, R. Linear response for spiking neuronal networks with unbounded memory, 2020. under revision.
101. Galves, A.; Löcherbach, E.; Pouzat, C.; Presutti, E. A system of interacting neurons with short term plasticity. *J. Stat. Phys. (to appear)* **2019**.
102. Galves, A.; Löcherbach, E., Stochastic chains with memory of variable length. In *Festschrift in honour of the 75th birthday of Jorma Rissanen*; TICSP Series vol. 38, 2008; pp. 117–133.
103. De Masi, A.; Galves, A.; Löcherbach, E.; Presutti, E. Hydrodynamic Limit for Interacting Neurons. *Journal of Statistical Physics* **2014**, [1401.4264]. doi:10.1007/s10955-014-1145-1.
104. Fournier, N.; Löcherbach, E. On a toy model of interacting neurons. *Annales de l'institut Henri Poincaré (B) Probability and Statistics* **2016**, [1410.3263]. doi:10.1214/15-AIHP701.
105. Yaginuma, K. A Stochastic System with Infinite Interacting Components to Model the Time Evolution of the Membrane Potentials of a Population of Neurons. *Journal of Statistical Physics* **2016**. doi:10.1007/s10955-016-1490-3.
106. Hodara, P.; Löcherbach, E. Hawkes processes with variable length memory and an infinite number of components. *Advances in Applied Probability* **2017**. doi:10.1017/apr.2016.80.
107. Ferrari, P.A.; Maass, A.; Martínez, S.; Ney, P. Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory and Dynamical Systems* **2000**. doi:10.1017/S0143385700000924.
108. Comets, F.; Fernández, R.; Ferrari, P.A. Processes with long memory: Regenerative construction and perfect simulation. *Annals of Applied Probability* **2002**, [arXiv:math/0009204]. doi:10.1214/aoap/1031863175.
109. Cessac, B.; Paugam-Moisy, H.; Viéville, T. Overview of facts and issues about neural coding by spikes. *J. Physiol. Paris* **2010**, *104*, 5–18. doi:doi:10.1016/j.jphysparis.2009.11.002.
110. Kirst, C.; Timme, M. How precise is the timing of action potentials ? *Front. Neurosci.* **2009**, *3*, 2–3.
111. Cessac, B. A view of Neural Networks as dynamical systems. *International Journal of Bifurcations and Chaos* **2010**, *20*, 1585–1629. doi:10.1142/S0218127410026721.
112. Rudolph, M.; Destexhe, A. Analytical Integrate and Fire Neuron models with conductance-based dynamics for event driven simulation strategies. *Neural Computation* **2006**, *18*, 2146–2210.
113. FitzHugh, R. Impulses and physiological states in models of nerve membrane. *Biophys. J.* **1961**, *1*, 445–466.
114. Nagumo, J.; Arimoto, S.; Yoshizawa, S. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE* **1962**, *50*, 2061–2070.
115. Morris, C.; Lecar, H. Voltage oscillations in the barnacle giant muscle fiber. *Biophys J* **1981**, *35*, 193–213.
116. Izhikevich, E. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*; The MIT Press, Cambridge, MA, 2007.
117. Lampl, I.; Yarom, Y. Subthreshold oscillations of the membrane potential: A functional synchronizing and timing device. *Journal of Neurophysiology* **1993**. doi:10.1152/jn.1993.70.5.2181.
118. Engel, T.A.; Schimansky-Geier, L.; Herz, A.V.; Schreiber, S.; Erchova, I. Subthreshold membrane-potential resonances shape spike-train patterns in the entorhinal cortex. *Journal of Neurophysiology* **2008**. doi:10.1152/jn.01282.2007.

119. Ma, S. Modern Theory Of Critical Phenomena; M. New York: Routledge, 2001. doi:https://doi.org/10.4324/9780429498886.
120. Mastromatteo, I.; Marsili, M. On the criticality of inferred models. J. Stat. Mech. **2011**.
121. Cessac, B.; Ny, A.L.; Löcherbach, E. On the mathematical consequences of binning spike trains. Neural Computation **2017**, [1606.08567]. doi:10.1162/NECO\_a\_00898.
122. Yang, C.N.; Lee, T.D. Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation. Physical Review **1952**, *87*, 404–409. doi:10.1103/PhysRev.87.404.
123. Lee, T.D.; Yang, C.N. Statistical Theory of Equations of State and Phase Transitions. II. Lattice Gas and Ising Model. Physical Review **1952**, *87*, 410–419. doi:10.1103/PhysRev.87.410.
124. Privman, V.; Fisher, M.E. Universal critical amplitudes in finite-size scaling. **1984**, *30*, 322–327. doi:10.1103/PhysRevB.30.322.
125. Dyson, F.J. Existence of a phase-transition in a one-dimensional Ising ferromagnet. Comm. Math. Phys. **1969**, *12*, 91–107.
126. Venegeroles, R. Thermodynamic phase transitions for Pomeau-Manneville maps. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **2012**. doi:10.1103/PhysRevE.86.021114.
127. Collet, P.; Galves, A. Chains of Infinite Order, Chains with Memory of Variable Length, and Maps of the Interval. Journal of Statistical Physics **2012**. doi:10.1007/s10955-012-0579-6.
128. Ruelle, D. Is our mathematics natural? The case of equilibrium statistical mechanics. Bulletin of the American Mathematical Society **1988**. doi:10.1090/S0273-0979-1988-15634-0.
129. Wigner, E.P. The unreasonable effectiveness of mathematics in the natural sciences. Richard Courant lecture in mathematical sciences delivered at New York University, May 11, 1959. Communications on Pure and Applied Mathematics **1960**. doi:10.1002/cpa.3160130102.
130. Lesk, A.M. The unreasonable effectiveness of mathematics in molecular biology. Mathematical Intelligencer **2000**. doi:10.1007/BF03025372.
131. Sompolinsky, H.; Crisanti, A.; Sommers, H. Chaos in Random Neural Networks. Physical Review Letters **1988**, *61*, 259–262.
132. Faugeras, O.; Touboul, J.; Cessac, B. A constructive mean field analysis of multi population neural networks with random synaptic weights and stochastic inputs. Frontiers in Computational Neuroscience **2009**, *3*. doi:10.3389/neuro.10.001.2010.
133. Schuecker, J.; Goedeke, S.; Dahmen, D.; Helias, M. Functional methods for disordered neural networks. arXiv e-prints **2016**.
134. Helias, M.; Dahmen, D. Statistical field theory for neural networks; Vol. 970, Lecture Notes in Physics, Springer, 2020.
135. Tkacik, G.; Mora, T.; Marre, O.; Amodei, D.; Palmer, S.E.; Berry, M.J.; Bialek, W. Thermodynamics and signatures of criticality in a network of neurons. Proceedings of the National Academy of Sciences **2015**, *112*, 11508–11513.
136. Tkacik, G.; Marre, O.; Amodei, D.; Schneidman, E.; Bialek, W.; Berry, M.J. Searching for collective behavior in a large network of sensory neurons. PLoS Comput. Biol. **2014**, *10*, e1003408. doi:10.1371/journal.pcbi.1003408.
137. Faugeras, O.; MacLaurin, J. A large deviation principle for networks of rate neurons with correlated synaptic weights. Twenty Second Annual Computational Neuroscience Meeting: CNS\*2013; , 2013.
138. Faugeras, O.; MacLaurin, J. Asymptotic description of stochastic neural networks. I. Existence of a large deviation principle. Comptes rendus Mathématique **2014**, *352*, 841–846. doi:10.1016/j.crma.2014.08.018.
139. Ost, Guilherme; Reynaud-Bouret, P. Sparse space-time models: concentration inequalities and Lasso. Annales de l'IHP, Probabilités et Statistiques (accepté) **2019**.
140. Reynaud-Bouret, P.; Rivoirard, V.; Grammont, F.; Tuleau-Malot, C. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. Journal of Mathematical Neuroscience **2014**. doi:10.1186/2190-8567-4-3.
141. Delarue, F.; Inglis, J.; Rubenthaler, S.; Tanré, E. Global solvability of a networked integrate-and-fire model of McKean-Vlasov type. Annals of Applied Probability **2015**. doi:10.1214/14-AAP1044.
142. Cormier, Q.; Tanré, E.; Veltz, R. Hopf bifurcation in a Mean-Field model of spiking neurons. working paper or preprint.

143. Lambert, R.C.; Tuleau-Malot, C.; Bessaih, T.; Rivoirard, V.; Bouret, Y.; Leresche, N.; Reynaud-Bouret, P. Reconstructing the functional connectivity of multiple spike trains using Hawkes models. Journal of Neuroscience Methods **2018**. doi:10.1016/j.jneumeth.2017.12.026.
144. Albert, M.; Bouret, Y.; Fromont, M.; Reynaud-Bouret, P. Surrogate data methods based on a shuffling of the trials for synchrony detection: The centering issue. Neural Computation **2016**. doi:10.1162/NECO\_a\_00839.
145. Bressloff, P.C.; Coombes, S. Dynamics of strongly coupled spiking neurons. Neural Computation **2000**. doi:10.1162/089976600300015907.
146. Bialek, W. Biophysics: Searching for Principles; Princeton University Press, 2012.