HAL
open science

Linear response in neuronal networks: from neurons dynamics to collective response

Bruno Cessac

To cite this version:

Bruno Cessac. Linear response in neuronal networks: from neurons dynamics to collective response. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2019, 29 (103105), 10.1063/1.5111803 . hal-02280089

HAL Id: hal-02280089

https://inria.hal.science/hal-02280089v1

Submitted on 6 Sep 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Linear response in neuronal networks: from neurons dynamics to collective response

B. Cessac[a)]

(Dated: 6 September 2019)

We review two examples where the linear response of a neuronal network submitted to an external stimulus can be derived explicitly, including network parameters dependence. This is done in a statistical physics-like approach where one associates to the spontaneous dynamics of the model a natural notion of Gibbs distribution inherited from ergodic theory or stochastic processes. These two examples are the Amari-Wilson-Cowan model[3,131] and a conductance based Integrate and Fire model[101,102].

One of the contemporary challenges in neuroscience is to understand how our brain processes external world information. For example, our retina receives the light coming from a visual scene and efficiently converts it into trains of impulses (action potentials) sent to the brain via the optic nerve. The visual cortex is then able to decode this flow of information in a fast and efficient way. How does a neuronal network, like the retina, adapts its internal dynamics to stimuli, yet providing a response that can be successfully deciphered by another neuronal network ? Even if this question is far from being resolved, there exist successful methods and strategies providing partial answers. To some extent, as developed in this paper, this question can be addressed from the point of non equilibrium statistical physics and linear response theory. Although neuronal networks are outside the classical scope of non equilibrium statistical physics - interactions (synapses) are not symmetric, equilibrium evolution is not time-reversible, there is no known conserved quantity, no Lyapunov function - an extended notion of Gibbs distribution can be proposed, directly constructed from the dynamics where the linear response can be derived explicitly, including network parameters dependence.

I. INTRODUCTION

Our nervous system has the capacity to display adapted responses to its environment, in a fast way, e.g. compatible with survival, with low energy consumption so as to maintain the body temperature in a narrow range. The main cells involved in this process are the neurons, although other non neuronal cells, like glia, play a central role too, (see the website `http://www.networkglia.eu/en/classicpapers`). Our brain is made of neuronal networks with specific functions. A central question in neuroscience is to understand how these neuronal networks contribute to encode and decode the external world information so as to be able to respond in a fast and adapted way. Behind this question is the hope to be able, one day, to decipher the "neural code"[100], so as, for example, to stimulate the impaired retina of a patient in a proper way (e.g. with a retinal prosthesis) to partially restore his vision (see e.g.[126] and references therein).

Although we are yet far from this, it is already possible to characterise the response of sensory neurons to stimuli (e.g. in the retina or in the visual cortex) . The first step of this characterisation uses a linear response approach, where the response R to a stimuli is approximated by a spatio-temporal response kernel K, convolved with the stimulus S, $R \sim K * S$. This kernel can be empirically computed from the cell's activity, thereby allowing the deconvolution of the signal to find back the stimulus[26,95]. In the field of vision, the spatio-temporal kernel K characterises the receptive field of a sensory neuron and is a fundamental building element in Hubel and Wiesel's theory of visual perception[65]. It is known that the receptive field K does not only results from the intrinsic properties of the cell, but from its connections with other neurons. It involves therefore network effects. It is also known that the linear response approach is in general not sufficient to fully characterise the neurons' response property. It is only a first, essential step, but non linear corrections are usually necessary.

From a more fundamental point of view, it is not established yet how the non linear neurons' and synapses' dynamics, the synaptic architecture, the influence of noise, ... shape neuronal networks' response to external world stimulations and what makes a linear response theory relevant in this context. Understanding this is one contemporary task of theoretical neuroscience. The problem can be stated as follows. Assume that a neuronal network receives a time-dependent stimulus $S(t)$ from time t_0 to time t_1. Even if the stimulus is applied to a subset of neurons in the network, its influence will eventually propagate to other neurons, directly or indirectly connected. This will result in a complex process where the effect of the stimulus is interwoven with neurons' dynamics.Yet, is there a way to disentangle, deconvolve, the effect of the stimulus from the "background" neuronal activity ?

A classical way to tackle this question is to consider the stimulation as a perturbation of a state of "spontaneous" activity. The response to the stimulation is then written as an expansion involving correlations, of higher and higher orders, computed with respect to the spontaneous activity. In the field of neuroscience this expansion is often called a Volterra expansion[100]. The first term, where the response is proportional to the stimulus, is the linear response.

Linear response theory is also at the core of non equilibrium statistical physics[31,51]. The response of a system, originally at equilibrium, to a time-dependent perturbation is proportional

[a)]Université Côte d'Azur, Inria, Biovision team, France

to the stimulus, with proportionality coefficients obtained via correlations functions of induced currents, where the correlations are computed with respect to the equilibrium distribution. These are Green-Kubo relations[62,75]. The equilibrium state - which plays the role of the spontaneous state in the previous paragraph - is characterised, in statistical physics, by a Gibbs distribution.

It is therefore interesting to try and connect neuronal networks models dynamics and non equilibrium statistical physics, although the link is not straightforward. Statistical physics certainly applies to, e.g., characterise ionic transfer at the level of neurons and synapses, but here we are addressing the question at another level. When modelling neuronal networks one uses simplified models where the microscopic activity (at the molecular level) has been averaged out to produce a dynamical system with reduced mesoscopic variables (local voltage of a neuron, conductances, concentration of neurotransmitters, ...). It is not evident *a priori* that the formalism of equilibrium and non equilibrium statistical physics can be used at this mesoscopic level, so as to characterise the neural response to stimuli. In particular, dynamics is clearly non Hamiltonian, not time reversible, dissipative. Therefore, in such an analogy, what should be the form of the "energy" in the Gibbs distribution? Do their exist the analogue of currents ? For which quantity ? What is "transported" ? Although, there exist many approaches in neuroscience with a statistical physics flavour - the use of maximum entropy principle to analyse spike trains[111,114,130], the free energy principle from K. Friston et al[44–46], among many others - they are based on a formal analogy with statistical physics/thermodynamics principles, instead of being derived from the collective neuronal dynamics, in a kinetic-like theory.

Thus, the main question we want to address in this paper is:

- *Are there model-examples where the spontaneous spatio-temporal activity of a neuronal network can be characterised by a suitable form of Gibbs distribution, constructed from the neurons' dynamics, thereby allowing to derive a linear response theory where the linear response kernel is explicitly written in terms of the parameters governing the dynamics ?*

The resolution of this problem would also give hints to some pending questions in neuroscience:

- *How does a stimulation applied to a subgroup of neurons in a population affect the dynamics of the whole network ?*
- *How to measure the influence of a neuron's stimulation on another neuron, especially if they are not synaptically connected ?* This question is related to the notion of "effective" or "functional" connectivity between neurons (or groups of neurons) at the core of many researches in neuroscience[47]. Several definitions of this connectivity can be given, not necessarily equivalent (e.g. based on pairwise correlations, causality, or mutual information[9,15]).
- *How does this connectivity relate to synaptic connectivity and non linear dynamics ? Is it related to a transported quantity, typically "information" ?*

In this paper, we address these points on the basis of two paradigmatic neuronal network models, the Amari-Wilson-Cowan model[3,131] and the conductance based Integrate and Fire model proposed by M. Rudolph and A. Destexhe in[101,102]. The main idea here is to associate, to the spontaneous dynamics of the model, a natural notion of Gibbs distribution inherited from ergodic theory or stochastic processes and to derive the linear response from this.

This is a review paper, based on a lecture given in the LACONEU 2019 summer school in Valparaiso, `http://laconeu.cl/`. The main material has been published elsewhere although the presentation and discussion are original. It addresses both to the community of neuroscientists - interested in the applications of ideas and techniques coming from statistical physics and dynamical systems to neuroscience - and to experts in non equilibrium statistical physics interested in questions in neuroscience where their knowledge could be helpful. Although based on a well established mathematical framework the results presented here, from a physicist point of view, are non rigorous as we are most of the time at the border of theorems or far from these borders.

The paper is organised as follows. In section II we give a brief introduction to neuronal modelling and non equilibrium statistical physics for non expert readers. Then, we develop two examples where a linear response theory can be derived for a neuronal network, with explicit relations between the parameters and equations defining the neurons dynamics and the linear response convolution kernel. This way, we give partial answers to the questions raised in this introduction. In section III, we present a former work with J.A. Sepulchre where linear response can be lead relatively far using the fact that the network dynamics is chaotic[18–20]. In section IV, we present an ongoing work with R. Cofré[24] where we study the derivation of a linear response in a spiking neural network model using the formalism of Markov chain and chains with complete connections[90]. The last section is devoted to discussion.

II. BASES

A. Neuronal networks

We give here a brief summary of neuronal dynamics for the non familiar reader, so as to help him/her better understand the models presented later. For more details see[30,36,58].

1. *The dynamics of neuronal voltage*

Neurons are cells able to use ionic transfers to change locally their membrane potential (voltage), that is, the difference between the local electric potential inside the cell and the local electric potential outside the cell. The fundamental equation

controlling the time evolution of the voltage V_k of neuron k is based on local charge conservation:

$$C_k \frac{dV_k}{dt} = i_{ion,k} + i_{syn,k} + i_{ext,k} + i_{noise,k}, \tag{1}$$

where C_k is the neuron's membrane capacity. It relates the neuron's voltage variations to currents flowing through the neuron's membrane. The main currents are:

a. Ionic currents in the body of the neuron (soma, axon). They are due to fluxes of ions going through ionic channels in the membrane. These channels can open or close depending on the membrane voltage as well as other variables depending on the channel's type. These currents are summarised by the term:

$$i_{ion,k} = -\sum_X g_{k,X}(\,.\,)\,(V_k - E_X), \tag{2}$$

where the sum holds on ionic channels types, selective to ionic species (sodium, potassium, chloride, ...). The quantity $g_{k,X}$ is the conductance of the channel of type X. It is roughly proportional to the density of open channels of type X in the membrane. It depends on hidden variables (activation, inactivation, ...) summarised here as a dot (.) without further indication. E_X is the reversal or Nernst potential of X corresponding to the voltage V_k at which the ionic current of type X changes its direction.

Typical ionic currents are those generating action potentials, also called spikes. These are fast (of order milliseconds), large increase in the membrane voltage (depolarisation) due the influx of positive ions (typically, sodium), followed by a fast decrease (re-polarisation) due to the efflux of potassium, then by a refractory period during which the neuron cannot emit a spike any more. If the shape of the spike is essentially invariant for a given neuron, the sequence of spikes it emits may vary significantly in timing. The term $i_{ion,k}$ can contain many others ionic current types, regulating the neuron's activity, in particular, spikes timing (see[36] for a nice mathematical and biophysical presentation).

b. Synaptic currents. An increase (depolarisation) in the voltage of the pre-synaptic neuron j induces a release of neurotransmitters which diffuse to the post-synaptic neuron k and bind to specific receptors with different possible mechanisms. This triggers the opening of ionic channels generating a local (at the level of the synapse) synaptic current, $i_{syn,k}$ that takes a similar form as (2):

$$i_{syn,k} = -\sum_j g_{kj}(\,.\,)\,(V_k - E_{kj}), \tag{3}$$

although the mechanisms regulating the synaptic conductance g_{kj} from j to k are of a different nature. (See[33] for a clear and synthetic presentation of neurotransmitter-receptors modelling). The sum holds on pre-synaptic neurons connected to k. E_{kj} is the reversal potential of the ionic species triggering the current from j to k.

c. External current. Neurons can be excited by external stimuli (typically an external current imposed by an electrode). This corresponds to the current $i_{ext,k}$ in (1). In general, it depends explicitly on time. In the paper, $i_{ext,k}$ will design a stimulus exogenous to the network, without specifying how it is injected to the neurons (it could be e.g. a synaptic current coming from neurons of another network).

d. "Noise" current. Finally, $i_{noise,k}$ is a stochastic term that mimics "noise" in neurons' dynamics (thermal noise inducing a probabilistic ionic channel activation, diffusion of neurotransmitters, ...). In general, $i_{noise,k}$ is modelled as a white noise.

Note that, on biophysical grounds equation (1) holds for a small piece of the neuron's membrane. In this paper, though, we will consider neurons as points (no spatial structure), so that eq. (1) holds for the whole neuron.

2. Collective response

The collective dynamics of neurons (1) in the absence of stimulation ($i_{ext,k}(t) = 0$) is called *spontaneous*. It results from the intrinsic non linear dynamics of neurons (the term $i_{ion,k}$) as well as from neurons' interactions (the term $i_{syn,k}$). It can therefore be quite complex (bursting[67] , chaotic[2,74,97], generating waves[13], ...). In addition, the noise term introduces some degree of stochastic. Thus, in general, one is not attempting to analyse the individual trajectories of the dynamical system (1), but instead, one is studying statistical properties (e.g. spike rates or spikes correlations). It is a reasonable assumption, used in all the model we know, to consider that statistics of the spontaneous activity is stationary (time-translation invariant). This means that neurons' spike rates in spontaneous activity are constant, or that the pairwise spike correlation between 2 neurons only depends on the time interval between the spikes emitted by these neurons.

When the neuronal network is submitted to an external influence (the term $i_{ext,k}$ in (1)), the collective dynamics is called *stimulus evoked* or *stimulus dependent*. As the stimulation usually depends explicitly on time, the stationarity assumption, *stricto-sensu* does not hold. Yet, many theoretical tools are grounded on a stationarity assumption: especially all methods based on entropy (maximum entropy, mutual information). Other approaches, as the one presented in this paper, are not mathematically constrained by this hypothesis. The alternative strategy used here considers that the stimulus has a small amplitude, so that the neuronal network responds proportionally to the stimulus. In other words, the difference between spontaneous activity statistics and evoked statistics is proportional to the stimulus amplitude. In the context of linear response theory the proportionality coefficient is computed from correlations functions in spontaneous activity.

3. Simplified models

Equation (1) hides a large number of additional differential equations ruling conductances, calcium dynamics, synaptic activity and so on, which are just impossible to study mathematically in full generality. They are also hard to simulate,

not only because of the large number of variables and equations, but also, and mainly, because of the large number of parameters entering in the biophysics, which have to be determined from experiments. Modellers are therefore using simplified versions of (1) focusing on some specific aspects and questions, as we will do in this paper.

As a salient feature of most neurons in the nervous system (although not all of them) is to produce spikes, a standard modelling approach is to characterise the activity of a neuron or of a neuronal network by focusing on spikes timing. Neural activity is then represented by spike trains, or firing rates - the number of spikes emitted per second by a neuron. A firing rate model is presented in section III. A spiking model is presented in section IV.

B. Linear response in statistical physics

We give a brief summary of non equilibrium statistical physics (with a physicists point of view) for non familiar readers. We adopt a classical description (see[77] for a didactic, yet wide, introduction to the subject) disregarding, for the moment, more general and synthetic recent approaches[5,8] (see discussion section).

We consider a system characterised by a microstate ω (a point in the phase space, a spin configuration, ...). For simplicity we consider that ω takes a countable number of values.

When the system is at equilibrium the probability to observe the microstate ω is:

$$P[\omega] = \frac{1}{Z} e^{-\frac{H(\omega)}{k_B T}} \tag{4}$$

where $Z = \sum_\omega e^{-\frac{H(\omega)}{k_B T}}$ is called partition function, with k_B, the Boltzmann constant and T the temperature in Kelvin. The function of ω:

$$H(\omega) = \sum_\alpha \lambda_\alpha X_\alpha(\omega), \tag{5}$$

is called the energy of the microstate ω. The functions X_α are extensive quantities (proportional to the number of particles) such as energy, electric charge, volume, number of particles, magnetic field, ... The conjugated parameters λ_α correspond to intensive quantities (not proportional to the number of particles), like temperature, electric potential, pressure, chemical potential, ... In general they depend on the location in the physical space (e.g. the temperature depends on the position in a fluid). At equilibrium they are uniform in space though.

The form of H, i.e. the choice of the λ_α and X_α is constrained by the physical properties of the system. It is also constrained by boundary conditions. In standard statistical physics courses, the Gibbs distribution form (4) is obtained as a consequence of a principle, the Maximum Entropy Principle[68]: maximising the statistical entropy under the constraint that the average value of H is fixed. The statistical entropy is proportional to the Shannon entropy (up the, fundamental, Boltzmann constant), making a deep link between thermodynamics and information theory. More general definitions of Gibbs distributions exist though, not constrained by entropy, and constructed from dynamics. We see two examples in this paper. In this setting, maximising entropy is a consequence of large deviations theory[32].

Equilibrium statistical physics allows to establish macroscopic laws from first principles, These laws summarise a complex, non linear dynamics, with a large number of particles, in a few macroscopic variables related by a few equations[77]. A well known example is the law of ideal gas, $PV = nRT$. A natural question is whether spontaneous neuronal dynamics could obey similar laws.

A non equilibrium situation arises when the λ_αs are not uniform in space, generating gradients ("thermodynamic forces"), $\vec{\nabla}\lambda_\alpha$ (temperature gradient, electric potential gradient ...). These gradients result in currents density $\vec{j}_\alpha$ of X_α (e.g. a temperature gradient induces a heat current). In Onsager theory, currents density are functions of gradients:

$$\vec{j}_\alpha = \vec{\mathscr{F}}_\alpha(\vec{\nabla}\lambda_1, \ldots, \vec{\nabla}\lambda_\beta, \ldots).$$

If gradients are small and if $\vec{\mathscr{F}}_\alpha$ is differentiable:

$$\vec{j}_\alpha \sim \sum_\beta L_{\alpha\beta} \vec{\nabla}\lambda_\beta + \ldots,$$

where the coefficients $L_{\alpha\beta}$ are called Onsager coefficients[91].

Typical examples are the Ohm's law where the electric current density $\vec{j}_{el} = -\sigma_E \vec{\nabla} V$, is proportional to the gradient of electric potential with a factor σ_E, the electric conductivity, or the Fourier's law $\vec{j}_Q = -\lambda \vec{\nabla} T$ where the heat flux is proportional to the temperature gradient.

Assuming the gradients are small enough so that the system can be divided into mesoscopic cells at equilibrium (quasi static approximation) the Onsager coefficients can be derived as correlation functions computed with respect to the equilibrium distribution. This constitutes the Green-Kubo relations[62,75]:

$$L_{\alpha\beta} \propto \int_0^{+\infty} \langle \vec{j}_\alpha(0) . \vec{j}_\beta(s) \rangle_{eq} \, ds, \tag{6}$$

where $\langle\,\rangle_{eq}$ denotes the average with respect to the Gibbs equilibrium probability (4). We assume here that $\langle \vec{j}_\alpha(0) \rangle_{eq} = \langle \vec{j}_\beta(s) \rangle_{eq} = 0$ so that (6) is the time integral of correlation of currents, where correlations are computed at equilibrium.

An important relation, which allows to derive Onsager coefficients from dynamics in several examples[50] is the entropy production, $\sigma \equiv \frac{dS}{dt}$. It is given, to the first order[80], in terms of the Onsager coefficients by:

$$\sigma = \sum_\alpha \vec{\nabla}\lambda_\alpha . \vec{j}_\alpha = \sum_{\alpha,\beta} \vec{\nabla}\lambda_\alpha L_{\alpha\beta} \vec{\nabla}\lambda_\beta. \tag{7}$$

The interesting point is that the Green Kubo relations can be obtained, in some cases, from the microscopic dynamics ruling the evolution of the microstate, using different techniques: this has been done in dynamical systems and ergodic theory[48–50,106], or stochastic processes and Markov chains

(See[70] and references therein). A recent method, synthetising previous approaches has also been proposed by C. Maes and co-workers[5,8]. The application of these formalisms allows to construct a linear response theory in neuronal models, from the equations ruling neurons' dynamics, as we now show.

III. FROM FIRING RATE NEURONS DYNAMICS TO LINEAR RESPONSE

A. The Amari-Wilson-Cowan model

As a first example of linear response in neural network we consider a canonical model of neuronal dynamics, the Amari-Wilson-Cowan model[3,131,132]. It consists of a set of N neurons, $i = 1 \dots N$, with membrane voltage V_i, whose dynamics is given by the dynamical system:

$$\frac{dV_i}{dt} = -V_i + \sum_{j=1}^{N} J_{ij} f(V_j(t)) + \varepsilon S_i(t); \quad i = 1 \dots N. \tag{8}$$

This equation can be derived from equation (1) up to several approximations explained e.g. in[35,38]. Note that the decay (leak) term $-V_i$ has a more general form $-\frac{1}{\tau}(V_i - V_L)$ where V_L is a reversal potential and τ is a characteristic time. It is easy to get to the form (8) by rescaling time and voltages.

We will also consider the discrete time version of (8):

$$V_i(t+1) = \sum_{j=1}^{N} J_{ij} f(V_j(t)) + \varepsilon S_i(t); \quad i = 1 \dots N. \tag{9}$$

This form is more convenient to understand the mechanisms that shape the linear response, namely the interplay between neurons interactions and non linear dynamics, because one can follow the effect of a stimulus time-step by time-step. For this reason we will mainly stick at the discrete time dynamics, except in the next subsection (Contractive regime), where the derivation of the linear response and its consequences are more straightforward and familiar to readers in the continuous time case.

Neurons are coupled via synaptic weights J_{ij} characterising the strength of interaction from the pre-synaptic neuron j to the post-synaptic neuron i. This defines an oriented graph, i.e. $J_{ij} \neq J_{ji}$ in general, in contrast to physics where interactions are symmetric. This graph is also signed: when $J_{ij} > 0$ the interaction (synapse) is excitatory, when $J_{ij} < 0$, it is inhibitory.

In this model, the pre-synaptic neuron j influences the post synaptic neuron i via its firing rate (probability to emit a spike in a small time interval) which is a function $f(V_j)$ of the pre-synaptic neuron voltage. Here, f is a non linear, sigmoid function as depicted in Fig. 1. A typical form for f is:

$$f(x) = \frac{1}{2}\left(1 + \tanh(gx)\right). \tag{10}$$

The sigmoidal shape has a deep biological importance. Indeed, one can distinguish 3 rough regions (Fig. 1). In region I (low voltage), the neuron does not emit spikes. In region II, $f(V)$ is roughly linear. In region III (high voltage), the firing rate reaches a plateau, fixed by the refractory period.

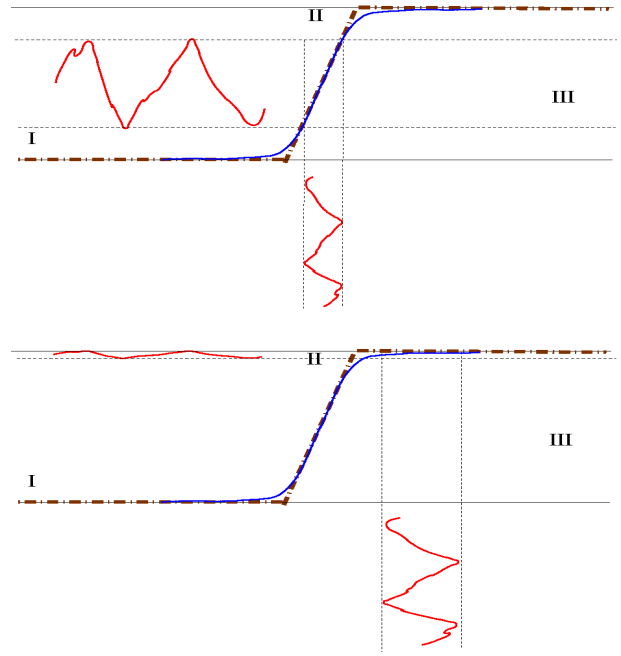

FIG. 1. The sigmoidal shape of the function f and its effects on voltage fluctuations. Top. When the neuron's voltage fluctuates around the inflection point of the sigmoid, and if the gain g is large enough fluctuations are amplified. Bottom. When the neuron's voltage fluctuates in the flat parts of the sigmoid (here, the saturated region) fluctuations are damped. Dashed red lines correspond to a piecewise linear approximation of the sigmoid, allowing to delimit regions I, II, III.

The parameter g in (10), either called "gain" or "non linearity", is of up most importance. On biophysical grounds, it characterises the sensitivity of the neuron's firing rate to fluctuations of its voltage. Consider indeed Fig. 1, top. When g is larger than 1 the fluctuations are amplified by f in region II. In contrast, in region I and III they are damped. This remark, made at the level of single neuron, has a deep importance when interpreting the linear response of a network governed by eq. (8) or (9). On dynamical grounds this effect controls the local expansion / contraction in the phase space, as developed below.

Finally, in eq. (8), (9), $S_i(t)$ is an "external stimulus". Here, it depends only on time but the analysis made below affords a situation where S_i depends also on the neuron's voltage.

We introduce the state vector $\vec{V} = \left(V_i \right)_{i=1}^{N}$, the stimulus vector $\vec{S} = \left(S_i \right)_{i=1}^{N}$ and the matrix of synaptic weights $\mathscr{J} = \left(J_{ij} \right)_{i,j=1}^{N}$. With a slight abuse of notations we write $f(\vec{V})$ for the vector $\left(f(V_i) \right)_{i=1}^{N}$. Then, we may rewrite (8) in vector form:

$$\frac{d\vec{V}}{dt} = \underbrace{-\vec{V} + \mathscr{J}.f(\vec{V})}_{\vec{F}(\vec{V})} + \varepsilon \vec{S}(t), \tag{11}$$

whereas (9) becomes:

$$\vec{V}(t+1) = \underbrace{\mathscr{J}.f(\vec{V}(t))}_{\vec{G}(\vec{V}(t))} + \varepsilon\vec{S}(t). \tag{12}$$

Here, we don't make any hypothesis on the matrix of weights $\mathscr{J}$, except that it's entries are bounded in absolute value. This implies that the spontaneous dynamics ($\varepsilon = 0$) can be restricted to a compact set $\mathscr{M}$.

We remark that the Jacobian matrix of (11), $D\vec{F}_{\vec{V}}$ have the form:

$$D\vec{F}_{\vec{V}} = -\mathscr{I} + \mathscr{J}.\mathscr{D}(\vec{V}), \tag{13}$$

where $\mathscr{I}$ is the $N \times N$ identity matrix and $\mathscr{D}(\vec{V})$ is the diagonal matrix:

$$\mathscr{D}(\vec{V}) = diag\left(f'(V_i), i = 1 \ldots N\right). \tag{14}$$

Likewise, the Jacobian matrix of (12), $D\vec{G}_{\vec{V}}$, reads:

$$D\vec{G}_{\vec{V}} = \mathscr{J}.\mathscr{D}(\vec{V}). \tag{15}$$

B. Contractive regime

1. Dynamics

In order to illustrate the main ideas of this section we start by considering a specific regime of the Amari-Wilson-Cowan model, the contractive regime[21,82]. This regime occurs when neurons have a very low gain g. In this case, the dynamical systems (11) and (12) have a unique fixed point attracting all trajectories (absolutely stability). On mathematical grounds, this is the starting point for the analysis of the bifurcations cascade occurring in the model when increasing g. On neuronal grounds, this is a low reactivity regime where, after a sufficiently long time and whatever the initial condition, neurons fire with a constant rate in spontaneous activity. Although this regime is somewhat trivial it brings nevertheless several interesting hints of linear response in more complex situations.

The existence of a contractive regime is deduced from the following argument. If $g = 0$, $\mathscr{D}(\vec{V}) = 0$ and all eigenvalues of $D\vec{F}_{\vec{V}}$ are equal to -1 (resp. all eigenvalues of $D\vec{G}_{\vec{V}}$ vanish). By continuity, for g sufficiently small, all the eigenvalues of $D\vec{F}_{\vec{V}}$ have a strictly negative real part, $\forall\vec{V} \in \mathscr{M}$, (resp. all the eigenvalues of $D\vec{G}_{\vec{V}}$ have a modulus strictly smaller than 1). From this, one can establish that there is a g value, $g_{as}(\mathscr{J})$, depending on $\mathscr{J}$, such that, for $g < g_{as}(\mathscr{J})$, the spontaneous dynamical system ($\varepsilon = 0$) has a unique fixed point $\vec{V}^*$ attracting all trajectories in $\mathscr{M}$. When g further increases, one gets out of this regime by a cascade of bifurcations leading to a more and more complex dynamics, discussed in section III C.

We consider now the perturbed neural network ($\varepsilon > 0$) in this contractive regime. We first make the computation for the continuous time system as it is straightforward and familiar to most readers. We look for solutions of the form $\vec{V} = \vec{V}^* + \vec{\xi}$. This is a standard linear stability analysis. We have:

$$\frac{d\vec{\xi}}{dt} = D\vec{F}_{\vec{V}^*}.\vec{\xi} + \varepsilon\vec{S}(t) + O\left(\varepsilon^2\right),$$

with solution:

$$\vec{\xi}(t) = e^{D\vec{F}_{\vec{V}^*}(t-t_0)}.\vec{\xi}(t_0) + \varepsilon\int_{t_0}^{t} e^{D\vec{F}_{\vec{V}^*}(t-s)}.\vec{S}(s)ds + O\left(\varepsilon^2\right),$$

where t_0 is the initial time where we start to apply the stimulus. This solution contains a transient term, dependent on the initial condition, and a stimulus dependent term. We now consider the steady state regime corresponding to $t_0 \to -\infty$: the stimulus was applied in a very distant time in the past, quite longer than the longest characteristic time of the dynamics. In this limit:

$$\vec{\xi}(t) \sim \varepsilon\int_{-\infty}^{t} e^{D\vec{F}_{\vec{V}^*}(t-s)}.\vec{S}(s)ds = \varepsilon\left[\chi * S\right](t), \tag{16}$$

where the integral converges since all eigenvalues have negative real part. We have introduced the matrix:

$$\chi(t) = e^{D\vec{F}_{\vec{V}^*}t}. \tag{17}$$

Equation (16) is a first example of a linear response formula, where the deviation ξ from the solution of spontaneous dynamics (here, the fixed point $\vec{V}^*$), is proportional to the stimulus, and expressed by a convolution with the linear response kernel (17). Thus, the linear response kernel K discussed in the introduction reads $K = \varepsilon\chi$.

The same result holds *mutatis mutandis* for the discrete time dynamical system (12) with a discrete time convolution:

$$\xi(t) = \varepsilon\sum_{\tau=-\infty}^{t-1} DG_{\vec{V}^*}^{t-\tau-1}.\vec{S}(\tau) + O(\varepsilon^2) \tag{18}$$

where :

$$\chi(t) = D\vec{G}_{\vec{V}^*}^{t}, \tag{19}$$

the t-th iterate of Jacobian matrix $D\vec{G}$ at $\vec{V}^*$. We come back to the derivation of (18), for a more general case, in section III C. The series (18) converges because, in the contractive regime, the eigenvalues of $D\vec{G}_{\vec{V}^*}$ have a modulus strictly lower than 1.

2. Susceptibility and resonances

The Fourier transform of (16) is:

$$\hat{\xi}(\omega) = \varepsilon\hat{\chi}(\omega).\hat{S}(\omega), \tag{20}$$

where ω is a real frequency. $\hat{\chi}(\omega)$, called the complex susceptibility, characterises the response to a harmonic stimulus with frequency ω, with an amplification factor (the modulus $|\hat{\chi}(\omega)|$) and a phase factor (the argument of $\hat{\chi}(\omega)$). $\hat{\chi}(\omega)$ can be analytically continued in the complex plane (complex frequencies, $\omega = \omega_r + i\omega_i$). Its explicit computation

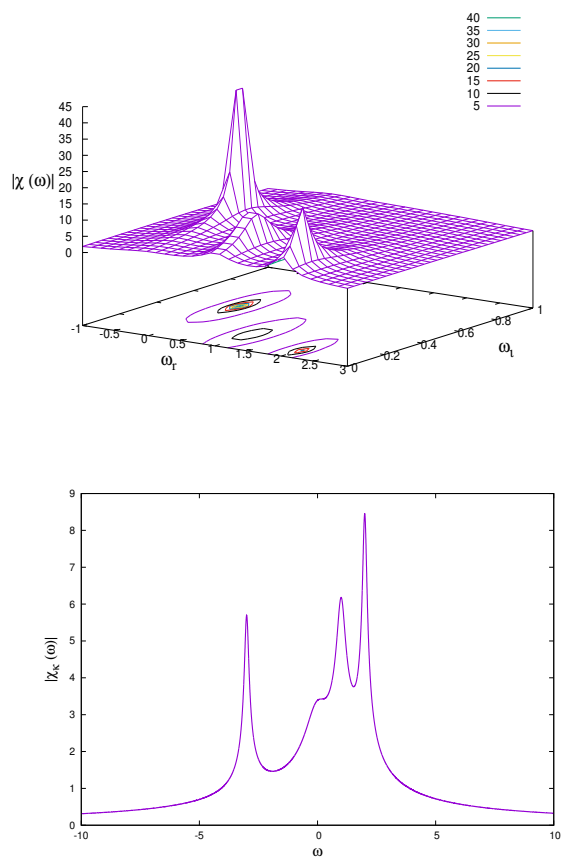

FIG. 2. Schematic plot of the resonances. Up. The modulus $|\hat{\chi}(\omega)|$ for complex frequencies. Bottom. $|\hat{\chi}(\omega)|$ projected on the real axis.

is straightforward. We give it first in the continuous time case. We note $\lambda_k = \lambda_{k,r} + i\lambda_{k,i}$ the eigenvalues of $D\vec{F}_{\vec{V}^*}$. Note that, as the matrix $\mathscr{J}$ is not symmetric, the eigenvalues λ_k are complex in general. Denoting P the transformation matrix diagonalising $D\vec{F}_{\vec{V}^*}$ we have $\hat{\chi}(\omega) = P.\hat{\chi}'(\omega).P^{-1}$ where $\hat{\chi}'_k(\omega) = \hat{\chi}'_{k,r}(\omega) + i\,\hat{\chi}'_{k,i}(\omega)$ is the diagonal matrix with entries:

$$\begin{aligned} \hat{\chi}'_{k,r}(\omega) &= -\frac{\lambda_{k,r}+\omega_i}{\left(\lambda_{k,r}+\omega_i\right)^2+\left(\lambda_{k,i}-\omega_r\right)^2}, \\ \hat{\chi}'_{k,r}(\omega) &= \frac{\lambda_{k,i}-\omega_r}{\left(\lambda_{k,r}+\omega_i\right)^2+\left(\lambda_{k,i}-\omega_r\right)^2}. \end{aligned} \tag{21}$$

The salient remark is that this expression has poles in the complex plane, located at $\omega_r = \lambda_{k,i}$, $\omega_i = -\lambda_{k,r}$, $k = 1\ldots N$. On the real axis (real frequencies) the trace of these poles gives peaks (resonances) where the response to harmonic perturbation is maximal. The peaks are located at $\omega_r = \lambda_{k,i}$ while their width depends on $\omega_i = -\lambda_{k,r}$. The situation is sketched in Fig. 2, where we have plotted $|\hat{\chi}(\omega)|$. We have only shown 4 poles for the legibility of the figure (2 poles are close to each other, as better seen on the real axis projection, bottom figure).

A similar computation can be done in the discrete time case. If λ_k denote now the eigenvalues of $D\vec{G}_{\vec{V}^*}$ the diagonal susceptibility $\hat{\chi}'(\omega)$ has entries $\hat{\chi}'_k(\omega) = \frac{1}{1-\lambda_k e^{i\omega}}$. Resonances are therefore now given by $\lambda_k = e^{-i\omega}$.

3. *The neural interpretation of resonances*

The existence of these resonances means the following. Exciting a neuron, or a group of neurons, with a harmonic perturbation, there are frequencies where the response to the perturbation is maximal. Especially, groups of neurons synchronise for a given resonance frequency. These groups depend on the corresponding eigenvector of the Jacobian matrix.

Let us understand now how these resonances relate to neurons' interactions and non linear dynamics. This is actually simpler for the discrete time dynamical system (9), mainly because we are going to study the propagation, step by step, of a signal along the network edges.

Assume therefore that we are injecting a periodic stimulus S, with frequency ω, at only one neuron, say j. How does this stimulation affect the other neurons in the network ? In the contractive regime the answer is relatively simple because neurons, in spontaneous activity, have a constant voltage (fixed point $\vec{V}^*$). When the stimulus is injected at neuron j, its voltage V_j oscillates periodically around this equilibrium value V_j^*. These oscillations are then synaptically transmitted to its post synaptic neighbours. If ε is small enough, the synaptic action of j to neuron k_1 is $\varepsilon J_{k_1 j} f'(V_j^*) S_j(t)$ to order 1 in ε. Thus, it is proportional to the synaptic weight, and to the derivative of f at V_j^*.

We find the effect illustrated in fig. 1: depending on which region of the sigmoid neuron j's potential is, the fluctuation induced by the stimulus are either non linearly contracted (in region I, III) or linearly multiplied by g in region II (where g is smaller than 1 in the contractive regime, in contrast to fig. 1 top).

More generally, we see that the first-order effect of the stimulus applied at j, on a neuron i, connected to j via a synaptic path $j \to k_1 \to k_2 \to k_3 \to k_4 \to i$, as in fig. 3, is proportional to the stimulus $\varepsilon S_j(t)$ with a proportionality coefficient $\prod_{l=1}^{5} J_{k_l k_{l-1}} f'(V^*_{k_{l-1}})$, where we set $k_0 = j$ and $k_5 = i$.

Thus, the effect of a stimulus applied to j at time 0, on neuron i, σ time step later, is proportional to:

$$\chi_{i,j}(\sigma) = \sum_{\gamma_{ij}(\sigma)} \prod_{l=1}^{\sigma} J_{k_l k_{l-1}} \prod_{l=1}^{\sigma} f'(V^*_{k_{l-1}}), \tag{22}$$

where the sum holds on all paths $\gamma_{ij}(\sigma)$ connecting j to i in σ time steps. This coefficient is nothing but the entry ij of $D\vec{G}^{\sigma}_{\vec{V}^*}$, the Jacobian matrix of the σ-th iterate of the mapping $\vec{G}_{\vec{V}^*}$. This gives an interesting, network oriented, interpretation of equation (19) where one can explicitly compute the linear response convolution kernel with an explicit dependence in the parameters determining the network dynamics.

When applying a periodic stimulus to neuron j, the effect propagates through the network edges, with a weight proportional to the synaptic weight and to the derivative f' at the corresponding node (neuron), as in Fig. 3. The effects of all these paths sum up at neuron i, generating contributions that

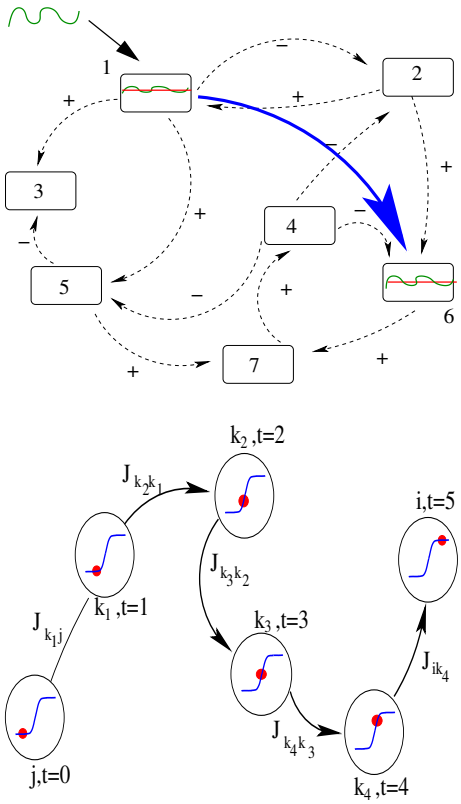

FIG. 3. Propagation of a stimulation throught the network. Top. A stimulus (green trace) is applied to neuron 1, superimposed upon the rest activity (red trace). This stimulation propagates, through the network, up to neuron 6 inducing a response (green trace in the box 6). Thus, the stimulation of neuron 1 influences neuron 6 even if there is no direct pathway between them. This is summarised by the blue arrow. This effect results from the summation of the stimulus influences propagating through network pathways, depending on neurons' voltage along those pathways. This is represented in the bottom figure. The state of the neuron (fixed point) is represented in red on the graph of the sigmoid function. The stimulus response is proportional to the derivative of the sigmoid at the red point.

can add up or interfere. This depends on the matrix $\mathscr{J}$ and on the neurons' rest state $\vec{V}^*$. It also depends on the frequency ω. At resonance frequencies the effects of the stimulus cumulate in an optimal way, generating a maximal response.

We have seen that these resonances are given by the eigenvalues of the Jacobian matrix $D\vec{G}_{\vec{V}^*}$. Interestingly, eigenvalues of a matrix can be expressed in cyclic expansions (using trace formula and ζ functions[57,93,103], see also `http://chaosbook.org/`), in terms of closed loops in the connectivity circuit defined by the matrix $D\vec{G}_{\vec{V}^*}$. Resonances correspond therefore to constructive interferences along these closed loops. This establishes a nice correspondence between the dynamical response to a stimulus, the network topology and the non linear dynamics.

C. Chaotic dynamics

1. *Beyond the contractive regime*

We now consider the Amari-Wilson-Cowan model outside the contractive regime, when g is larger than $g_{as}(\mathscr{J})$. There, strong non linear effects take place. Increasing g beyond the contractive regime results in general in codimension 1 local or global bifurcations[25,34], the most common being the destabilisation of the fixed point, $\vec{V}^*$, by a Hopf bifurcation, giving rise to periodic oscillations; or the appearance of other fixed points by saddle node bifurcation (or pitchfork when $f(x)$ has the symmetry $f(x) = -f(-x)$).

Here, we will consider a situation where dynamics become chaotic when g is large enough. On neuronal grounds, chaotic network have deserved a lot of attention since the 80s. Especially, the neurobiologist W. Freeman[43] argued that chaos in the brain may provide the brain vital flexibility while allowing it to visit many states in quick succession. The underlying idea is that chaotic dynamics constitute an infinite reservoir of information, that a neuronal network can access e.g. via a proper training (typically Hebbian learning or spike time dependent plasticity). This idea is at the core of many applications of neural networks such as echo state machines, liquid state machines, deep learning. The Amari Wilson Cowan model is a paradigmatic example exhibiting chaos where one can illustrate the computational properties of neuronal networks at the edge of chaos[10,86,118,119].

In view of the present paper, the interest of chaos is that a linear response theory can be formally obtained in this case, despite, and actually, thanks to, the fact that dynamics is chaotic. This apparent paradox is discussed below. We will consider a situation where chaos occurs via a transition by quasi-periodicity[110]. When increasing g out of the contractive regime the fixed point $\vec{V}^*$ destabilises by a Hopf bifurcation. As g increases a second Hopf bifurcation generates a 2 torus densely covered by trajectory. Then, frequency locking arises when crossing Arnold tongues[4]. In the region where Arnold tongues overlap (edge of chaos) one can see succession of periodic and quasi-periodic windows until the appearance of a chaotic regime (strange attractor), in general by a period doubling cascade although other scenarios as possible (see[52,79], for a description of the possible scenarios). Numerical measurement of the Lyapunov spectrum exhibit one or more positive Lyapunov exponents in this regime[16,119], corresponding to initial conditions sensitivity. The Lyapunov spectrum allows compute the dimension of the attractor via Kaplan-Yorke formula[71]. This dimension is not integer corresponding to a fractal attractor. Finally, from the Newhouse-Ruelle-Takens theorem[87] the attractor obtained at the end of this cascade of bifurcations is arbitrary close to an Axiom A attractor. We come back to this last point below.

The dynamical regimes resulting from the increase in the gain parameter obviously depend on the form of the matrix $\mathscr{J}$. A typical situation where this transition to chaos occurs is when the synaptic weights J_{ij} are random, independent, with Gaussian entries $\mathscr{N}(0, \frac{J^2}{N})$. The parameter J con-

trols therefore the variance of the distribution. This case has been proposed by H. Sompolinsky and co-workers in a seminal paper dating back to 1988[120]. They considered the case $f(x) = \tanh(gx)$ for the continuous time model (8), so that the destabilisation of the stable fixed point $\vec{V}^*$ arises when $-1 + g\Re(s_1) = 0$ where s_1 is the (random) eigenvalue of $\mathscr{J}$ with the largest real part. Now, the asymptotic distribution of the spectrum ($N \to +\infty$) of this family of random matrices is known from a theorem due to Girko[59,60]. The spectral density converges to a uniform distribution on the disk of center 0 and radius J in the complex plane. Thus, in the limit $N \to +\infty$ ("thermodynamic" limit), the fixed point destabilises for $gJ = 1$. The same holds for the discrete time model (9). It holds as well when breaking the symmetry $f(x) = -f(-x)$ although the bifurcations map is more complex[25].

The spectral behaviour gives actually a qualitative way to explain why chaos is occurring in this case[25,34]. When N is finite, while increasing g, eigenvalues of the Jacobian matrix cross the boundary of dynamical stability by complex conjugate pairs, leading to the transition to chaos by quasi-periodicity. As N increases the range of g values where this cascade of transitions occurs shrinks to 0. In the thermodynamic limit, infinitely many eigen-modes simultaneously destabilise when $gJ = 1$, leading to infinite dimension chaos[16]. Note that this scenario extends a priori to more complex forms of synaptic weights matrices, typically, having correlations, although this has not yet been really explored.

The choice for the mean and variance scaling of the synaptic weights is essential here: the variance scaling ensures that the sum of synaptic inputs have bounded variations as N growths, ensuring a proper, non trivial, thermodynamic limit $N \to +\infty$; having a zero mean ensures that dynamics is fluctuations-driven (neurons are mostly in region II of figure 1) leading to the so-called "balanced state"[128].

Random neural networks with random independent entries have attracted a lot of activity since the work[120]. Especially, H. Sompolinsky and A. Zippelius[121,122] developed an efficient dynamic mean-field approach for spin-glasses, which was later used to analyze the model (8). Although this theory is one of the most beautiful I know in the field of theoretical neuroscience (see[112] for a recent review) it requires a thermodynamic limit and deals with the average behaviour (weak convergence) of networks in this limit (although almost-sure convergence results now exist[37]) rendering difficult the interpretation of the dynamic mean-field equations and their solutions. Additionally, it relies heavily on the independence assumption of J_{ij} (although large deviations techniques now allow to access correlated weights[37]).

In contrast, the study proposed here deals with a given network with a finite size. Although the numerical examples presented in this paper were generated by a random model with independent $\mathscr{N}(0, \frac{J^2}{N})$ synaptic weights, the derivation of the linear response does not rely on this assumption. We do not average on the distribution of synaptic weigths, and we don't take the thermodynamic limit. As we see below, a nice resonances structure is produced by finite size models that is washed out by the mean-field approach (see[84] for a recent result on resonances in mean-field theory of chaotic network, including plasticity. The resonance structure is quite different from what is obtained in finite networks). Beyond neuronal networks, this resonance structure has important consequences in the field of dynamical systems and statistical physics of chaotic systems, as discussed below.

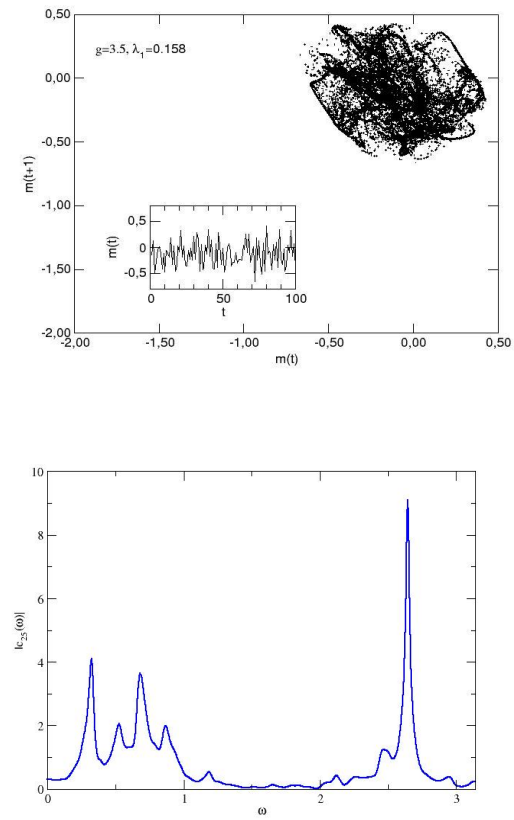

FIG. 4. Top. Representation of a strange attractor generated by (9). This is obtained by plotting $m(t+1)$ versus $m(t)$ where $m(t) = \frac{1}{N}\sum_{j=1}^{N} V_j(t)$. Bottom. Power spectrum of $m(t)$, ω is the frequency (between $[0, \pi]$ as time is discrete).

In the next steps we are going to consider the discrete time version (9). The main reason for this is that it affords an easy computation of the linear response in the chaotic regime, with a simple interpretation. An example of chaotic attractor obtained with (9) is given in Fig. 4 (top). In this regime the power spectrum of voltages (bottom) is continuous, but not flat, in contrast to white noise. There are peaks in the spectrum, corresponding to resonances in dynamics, called Ruelle-Pollicott resonances[57,93,96,105], as discussed below. These resonances are independent of the neuron. They are reminiscent of fig. 2 although dynamics here is quite more complex. The peaks are related to the succession of bifurcations (two Hopf bifurcations and frequency locking) leading to chaos.

D. Linear response in the chaotic regime

1. ε-expansion.

We now consider two versions of the dynamical system (12). The spontaneous dynamics version:

$$\vec{V}(t+1) = \vec{G}\left(\vec{V}(t)\right),$$

and the perturbed version:

$$\vec{V}'(t+1) = \vec{G}\left(\vec{V}'(t)\right) + \varepsilon\vec{S}(t). \tag{23}$$

We assume that the stimulus is switched on at time t_0 so that $\vec{V}(t_0) = \vec{V}'(t_0)$.

We define $\vec{\delta}V(t) = \vec{V}'(t) - \vec{V}(t)$ the difference between the trajectories of the two systems. We have thus $\vec{\delta}V(t_0+1) = \vec{V}'(t_0+1) - \vec{V}(t_0+1) = \varepsilon\vec{S}(t_0)$. At time t_0+2:

$$\vec{\delta}V(t_0+2) = \vec{G}\left(\vec{V}(t_0+1) + \varepsilon\vec{S}(t_0)\right) + \varepsilon\vec{S}(t_0+1) - \vec{G}(\vec{V}(t_0+1))$$

We now make a Taylor expansion of $\vec{G}\left(\vec{V}(t_0+1) + \varepsilon\vec{S}(t_0)\right)$ in powers of ε:

$$\vec{\delta}V(t_0+2) = \varepsilon\left[DG_{\vec{V}(t_0+1)}.\vec{S}(t_0) + \vec{S}(t_0+1)\right] + \varepsilon^2\vec{\eta}(t_0+1),$$

where $\vec{\eta}(t_0+1)$ contains terms of degree higher than ε. Note that we don't assume that $\varepsilon^2\vec{\eta}(t_0+1)$ is negligible so the equation is exact.

Iterating this procedure for larger times we obtain:

$$\vec{\delta}V(t) = \varepsilon\sum_{\tau=t_0}^{t-1} DG^{t-\tau-1}_{\vec{V}(\tau+1)}.\vec{S}(\tau) + \varepsilon^2\vec{R}(t) \tag{24}$$

where, again, we do not assume that $\varepsilon^2\vec{R}(t)$ is small and negligible.

This formula, which generalises (19), looks a bit useless. Indeed, a linear response theory would neglect the term $\varepsilon^2\vec{R}(t)$. But, in contrast to the contractive regime where the eigenvalues of $DG^{t-\tau-1}_{\vec{V}*}$ had a modulus < 1, ensuring the convergence of the series, here, the higher orders cannot be neglected, precisely because the system is chaotic. An initial perturbation, as tiny as it is, is locally amplified by dynamics. More precisely, dynamics is expansive in directions tangent to the attractor (positive Lyapunov exponents) and contractive in directions transverse to the attractor (negative Lyapunov exponents). This is illustrated in Fig. 5. The sum of all Lyapunov exponents is negative expressing that volume is contracted in the phase space. Asymptotic spontaneous dynamics lives on the attractor, so that a small perturbation - in our case, the stimulus $\varepsilon\vec{S}(t)$ - has generically a component tangent to the attractor and a component transverse to the attractor. The transverse component is exponentially damped, the tangent component is exponentially amplified. Thus, the net effect is an amplification of the stimulus effect, ruining any hope to neglect the "residual" term $\varepsilon^2\vec{R}(t)$ in (24). Therefore, it seems impossible to obtain a linear response on long times unless taking ridiculously small perturbations. This is the essence of the Van Kampen objection[127].

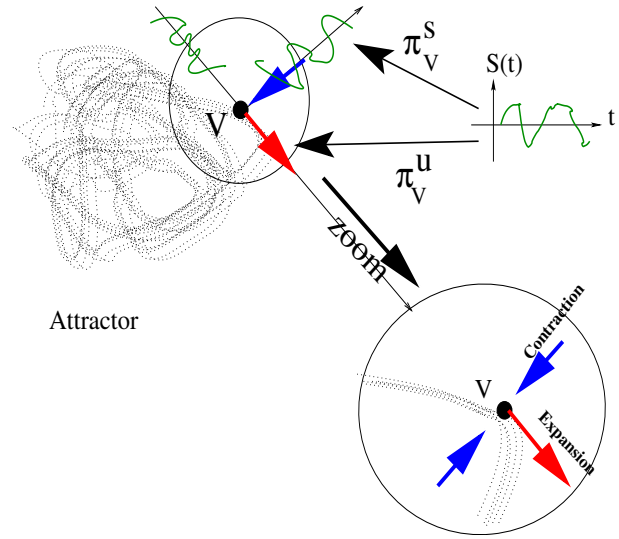

FIG. 5. The trajectory of (12) following the attractor of the spontaneous dynamics is perturbed by the time dependent stimulus $S(t)$ (green). The stimulation projects on the local stable manifold (projection $\pi^s_{\vec{V}}$) and is contracted; it also projects on the local unstable manifold (projection $\pi^u_{\vec{V}}$) and is expanded. The directions of projections depend on the point $\vec{V}$ on the attractor.

2. Expansive dynamics and ergodic average

To make one step further, let us now consider in more detail the effect of a small perturbation in a celebrated example, the Lorenz attractor[78]. In fig. 6 we have represented a trajectory (in red) and a small perturbation of the red trajectory (in green). One clearly sees the initial condition sensitivity: the two trajectories initially diverge exponentially fast. However, because the phase space is compact, non linear folding takes place and the two trajectories get closer (they can get arbitrary close from Poincaré's recurrence theorem), without crossing though (from Cauchy's theorem on uniqueness of solutions). After a sufficiently long time it becomes impossible to distinguish the 2 trajectories; the "green" attractor looks very much like the "red" one.

This example, illustrated here with the famous continuous time Lorenz model, illustrates a deep property that we are going to use now, ergodicity. This property holds, *mutatis mutandis* for our discrete time chaotic system. Consider, in our model, an initial condition $\vec{V}$ and its trajectory $\vec{G}^t(\vec{V})$, $t \geq 0$; consider a function $\Phi : \mathbb{R}^N \to \mathbb{R}^K$ (observable). Then, the time average of Φ on the trajectory, defined by the limit $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Phi(\vec{G}^t(\vec{V}))$, exists for typical initial conditions. We define now what we mean by "typical".

There exist a probability measure μ with support on the strange attractor Ω, such that:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Phi\left[\vec{G}^t(\vec{V})\right] \overset{\mu\ a.s.}{=} \int_\Omega \Phi(\vec{V})\mu(d\vec{V}) \tag{25}$$

for μ-almost every initial condition $\vec{V}$. The time average of Φ along orbits is equal to the average of Φ with respect to μ

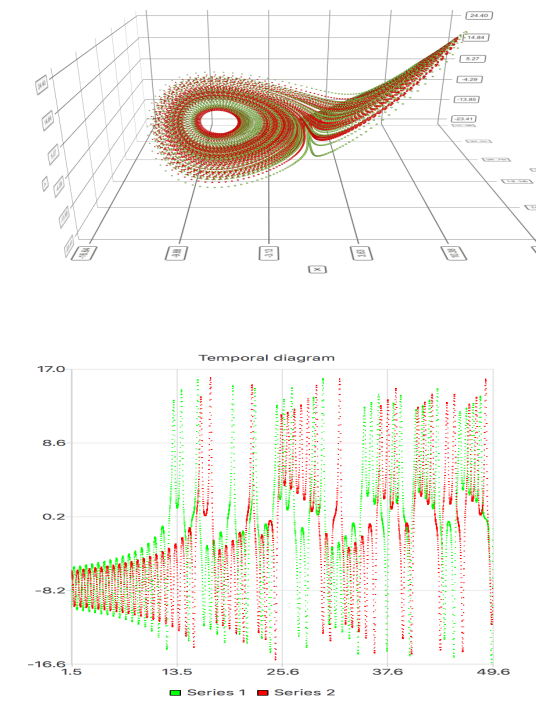

FIG. 6. Illustration of ergodicity on Lorenz attractor (top). We plot a trajectory (red) and a small perturbation of it (green). The green and red attractor look similar. Bottom. Temporal evolution of the 2 trajectories. Although initial condition sensitivity initially separates the two trajectories, they mix (without crossing in the 3 dimensional space) after a certain time.

(average on the attractor Ω).

There exist a class of dynamical systems, called uniformly hyperbolic or axiom A (discussed in more detail below), for which the measure μ is obtained by the weak limit:

$$\mu \overset{w}{=} \lim_{t\to+\infty} \vec{G}^t \mu_L, \tag{26}$$

where μ_L is the Lebesgue measure on the phase space $\mathscr{M}$, and $\vec{G}^t\mu_L$ is the image of μ_L by $\vec{G}^t$. Such a measure is called the Sinai-Ruelle-Bowen measure[11,104,107,117].

For the SRB measure the following holds:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \Phi\left[\vec{G}^t(\vec{V})\right] \overset{\mu_L\ a.s.}{=} \int_{\mathscr{M}} \Phi(\vec{V})\mu(d\vec{V}) \equiv \langle\Phi\rangle_{eq}. \tag{27}$$

This relation expresses that the time average of the trajectory of a "typical" initial condition, namely, selected with a uniform probability (Lebesgue measure) in $\mathscr{M}$, or any probability having a density with respect to μ_L, e.g. Gaussian, is equal to the average with respect to the SRB measure carried by the strange attractor.

3. Equilibrium versus non equilibrium

This has very deep physical meaning, reflecting an intuitive notion, somewhat initiated by Boltzmann who invented the word "ergodic"[51]. Starting from a "typical" microstate $\vec{V}$, i.e. selected with a natural probability, e.g. uniform or Gaussian, the time average of an observable Φ along the dynamical evolution of $\vec{V}$ is equal to the ensemble average of Φ with respect to the probability measure (macrostate) μ. The SRB measure μ plays therefore the role of the Boltzmann-Gibbs distribution in statistical physics, but it is constructed from the dynamical evolution without need to invoke the maximum entropy principle. We refer to the unperturbed situation as an equilibrium situation. This is the reason why we used the subscript eq in equation (27).

We want however to make here a small remark on the terminology. The theory used here, developed by Gallavotti, Cohen, Ruelle among others, was initially intended to characterise, from a dynamical system perspective, thermodynamic systems initially at equilibrium and perturbed by external forces, where the excess of energy is dissipated with a thermostat, ensuring a non equilibrium steady state. In this context, the equilibrium state has a density with the phase volume measure (the Liouville measure), while the non equilibrium state is characterised by a time dependent SRB measure which is not absolutely continuous. In contrast, our equilibrium state is a SRB measure, not absolutely continuous with respect to the volume element (it is only absolutely continuous along the unstable manifold). The theory developed by these authors extends to our case though[106].

Being ergodic SRB is stationary (time-translation invariant). In addition, the SRB measure is a Gibbs distribution whose "energy' is known and has the form:

$$H(\vec{V}) = -\log\det \pi^u_{\vec{V}} D\vec{G}_{\vec{V}}; \tag{28}$$

where $\pi^u_{\vec{V}}$ is the local projection on the unstable manifold. The average energy with respect to the SRB measure, $\langle H\rangle_{eq}$, is the sum of positive Lyapunov exponents. The SRB measure obeys a maximum entropy principle and its entropy is the sum of positive Lyapunov exponents[12]. This is Pesin formula[94].

An immediate consequence of ergodicity, somewhat appearing visually in Fig. 6, is that the time average of Φ on a perturbed trajectory is equal to the time average of the unperturbed trajectory:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Phi(\vec{G}^t(\vec{V})) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\Phi(\vec{G}^t(\vec{V}+\vec{\delta}V)). \tag{29}$$

This essentially expresses that time average smooths out the initial condition sensitivity and suggests to define a linear response theory via a proper averaging. Thinking of non equilibrium statistical physics this is exactly what we need. When considering particles in a fluid submitted to gradient of temperature, it is clear that molecular chaos and initial conditions sensitivity holds at the level of particles. But, at the level of a population, ensemble average, an order emerges, expressed by Fourier law, where the transport coefficient is expressed via

correlations of flux computed at equilibrium from the Gibbs distribution.

However, to define linear response in this context, one needs to extend the stationary situation exposed in this section, to a non stationary situation where the map defining the dynamics depends on time.

In this context, it is possible to define a time-dependent SRB measure by[106] (using our notations):

$$\mu_t = \lim_{n\to+\infty} \vec{G}'_t \dots \vec{G}'_{t-n}\mu_L, \tag{30}$$

where $\vec{G}'_t$ is the time-dependent map defined in (23). Equivalently, for an observable Φ the quantity:

$$\langle \Phi \rangle_t = \int_{\mathscr{M}} \Phi(\vec{V})\mu_t(d\vec{V}) \tag{31}$$

is the average value of the observable Φ at t, in the perturbed time-dependent evolution (23).

4. Linear response theory

D. Ruelle has developed a linear response theory for uniformly hyperbolic dynamical systems that we are going to use here[106]. That is, we assume for the moment that our chaotic attractor is uniformly hyperbolic. We come back to this point in section III E 4.

We note $\delta_t\mu[\Phi] = \langle \Phi \rangle_t - \langle \Phi \rangle_{eq}$ the difference between the average of Φ, at time t, for the perturbed time-dependent system and the average of Φ in the unperturbed system. This is the response of the perturbed system at time t, for the observable Φ and the stimulus $\vec{S}$. Ruelle formula reads, in our case[18–20]:

$$\delta_t\mu[\Phi] = \varepsilon \sum_{\tau=-\infty}^{t-1} \int \mu(d\vec{V}) D\vec{G}^{t-\tau-1}_{\vec{V}} \vec{S}(\tau).\vec{\nabla}_{\vec{V}(t-\tau-1)}\Phi + O(\varepsilon^2) \tag{32}$$

(actually, Ruelle formula extends to cases where the stimulus depends on the state $\vec{V}$).

Let us comment this formula in the simple case where $\Phi(\vec{V}) = \vec{V}$ so that $\vec{\nabla}\Phi = \vec{1}$. Then, $\delta_t\mu[\Phi] = \delta_t\mu\left[\vec{V}\right]$ is the difference between the average voltage at time t for the perturbed system and the unperturbed average. This gives;

$$\delta_t\mu\left[\vec{V}\right] = \varepsilon \sum_{\tau=-\infty}^{t-1} \chi(t-\tau-1).\vec{S}(\tau) + O(\varepsilon^2) \tag{33}$$

where :

$$\chi(t-\tau-1) = \int \mu(d\vec{V}) D\vec{G}^{t-\tau-1}_{\vec{V}} = \langle D\vec{G}^{t-\tau-1} \rangle_{eq} \tag{34}$$

is the linear response matrix. It is defined as the average of the Jacobian matrix $D\vec{G}^{t-\tau-1}$ with respect to the equilibrium SRB state μ. Note that (32) is a discrete time convolution, so that we can rewrite it in the form:

$$\delta_t\mu\left[\vec{V}\right] = \varepsilon\,[\chi * S](t), \tag{35}$$

similar to (18). Thus, eq. (34) provides an explicit form for the linear response kernel $K = \varepsilon\chi$ in the chaotic regime of the discrete time Amari-Wilson-Cowan model.

Let us now compare equation (32) to equation (24) obtained by a naive Taylor expansion where we had no hope to neglect the residual term $\vec{R}(t)$, which actually increases in time due to the positive Lyapunov exponents. In contrast, here the residual term $O(\varepsilon^2)$ remains under control and tends to zero like ε^2 when $\varepsilon \to 0$. Why is it so ?

This is sketched in Fig. 5. The stimulus perturbation locally projects on stable and unstable directions. In the stable direction, dynamics is contracting so perturbation is damped exponentially fast. In the unstable direction dynamics is expanding leading to amplification of a perturbation at the level of trajectories. However, considering averaging, as done in (34), the situation is different. It results indeed that the projection of the linear response on the unstable foliation is a correlation function between the observable Φ and a current $j_X = -\langle div^u \vec{X} \rangle_{eq}$ where div^u is the divergence computed along the attractor and $\vec{X}$ a smooth perturbation. More precisely, one can define a local Riemmanian metric $\mathscr{G}$ on the attractor so that the current reads[108]:

$$j_X = -\frac{1}{\sqrt{\det\mathscr{G}}} \sum_i \frac{\partial\left(\sqrt{\det\mathscr{G}}\,X^i\right)}{\partial x^i},$$

where X^i are the coordinates of the projection of $\vec{X}$ and x^i the local coordinates on the attractor (expressed e.g. in the Lyapunov basis). In this context, Green-Kubo relation and Onsager coefficients can be computed from the entropy production (7), not only at the lowest order, but also to higher orders[5,8,50,108].

Correlation functions decay exponentially fast in chaotic (uniformly hyperbolic) systems (exponential mixing) ensuring the convergence of the series (32). The decay rates are controlled by the eigenvalues of an evolution operator, acting on probabilities measures, the Ruelle-Perron-Frobenius operator, whose eigenvalues are the Ruelle-Pollicott resonances. The projection of these complex poles on the real axis gives the peaks appearing in the power spectrum Fig. 4.

Thus, the cumulated effect of the stimulus along a trajectory, obtained via time averaging, does not diverge. It converges, on the stable foliation, because of volume contraction, and on the unstable direction because of exponential mixing.

To sum up, the main difference between (32) and (19) is averaging with respect to the equilibrium measure. This makes physical sense. As pointed above, there is no hope to characterise the response of a fluid to a temperature gradient at the level of a particles trajectory, but it is possible at the level of densities.

The link with non equilibrium statistical physics can be formally pursued further as discussed in the conclusion of this section.

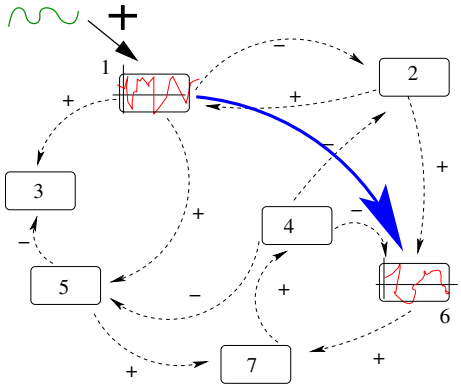

FIG. 7. Propagation of a stimulation throught the network in the chaotic case. In contrast to fig. 3 the dynamics of the stimulated neuron (1) is now evolving chaotically (red trajectory) and the stimulus (green trace) is superimposed upon it. This affects the dynamics of neuron 6 (blue arrow) but one has to disentangle the effect of the stimulus from the spontaneous activity in this neuron dynamics (red trace).

E. Linear response in the neural network model

1. Explicit form of the susceptibility

In the discrete time model (9) the Jacobian matrix $D\vec{G}_{\vec{V}} = \mathscr{J}.\mathscr{D}(\vec{V})$, where $\mathscr{D}$ is defined in eq. (14). It is then easy to compute χ whose entry χ_{ij} reads:

$$\chi_{ij}(\sigma) = \sum_{\gamma_{ij}(\sigma)} \prod_{l=1}^{\sigma} J_{k_l k_{l-1}} \langle \prod_{l=1}^{\sigma} f'(V_{k_{l-1}}(l-1)) \rangle_{eq}, \qquad (36)$$

where the sum holds on all synaptic paths connecting neuron j to neuron i in σ steps, with $k_0 = j$ and $k_\sigma = i$.

It is interesting to compare this equation to eq. (22) obtained in the contractive regime, where the attractor of dynamics was a fixed point. The straightforward difference is that we have now to average over the SRB measure the product of $f'(V_{k_{l-1}}(l-1))$. Actually, one obtains (22) by replacing the SRB measure by the Dirac measure on the attracting fixed point $\vec{V}^*$. This formula could be extended as well in the presence of noise.

Let us now interpret the meaning of the product $\langle \prod_{l=1}^{\sigma} f'(V_{k_{l-1}}(l-1)) \rangle_{eq}$ weighting each path $j \overset{\sigma}{\to} i$. As we have seen in section III B the response of the post synaptic neuron k_1 to a small variation of the pre synaptic voltage V_j which is proportional to $J_{k_1 j} f'(V_j)$. Now, the main differences with the contractive regime are: (i) V_j evolves in time; (ii) the gain $g > 1$ can be quite high. This second aspect is essential because, near the inflexion point of the sigmoid, f is expansive. It amplifies a small perturbation; in contrast it is contractive in the saturated parts (see Fig. 1). This is actually precisely this interplay between expansion and contraction which is essential to render dynamics chaotic for sufficiently large g (combined with the asymmetry of synaptic weights J_{ij}, as the model (9) with symmetric synapses has a Lyapunov function).

For clarity, let us consider, as in section III B the case when the signal is injected only at neuron j and let us study the response of neuron i. When the signal is injected at j, at time $t-\tau-1$, its propagation to i via the network pathways is weighted by the products of terms $J_{k_l k_{l-1}} f'(V_{k_{l-1}}(l-1))$. The contributions of all these paths sum up, with positive or negative weight (depending on the product of $J_{k_l k_{l-1}}$ along the path), with a small or large amplitude. The convolution form (35) expresses that the response of neuron i at time t integrates the past influence of the stimulus injected at j, propagating via many paths with different lengths and summing up at i.

Now, the bracket $\langle\,\rangle_{eq}$ expresses that this influences is given by the ergodic average of the products $\prod_{l=1}^{\sigma} f'(V_{k_{l-1}}(l-1))$. Remarkably, the averaging is done with respect to the equilibrium measure. We have here a strong analogy with non equilibrium statistical physics where transport coefficients are computed with respect to correlations functions of flux computed *at equilibrium*, as exposed in section II B.

2. Numerics

The main problem with eq. (36) is that it is numerically intractable. However, its Fourier transform is computable as we now explain. The (discrete time) Fourier transform of $\chi_{ij}(t)$ is:

$$\hat{\chi}_{ij}(\omega) = \sum_{t=-\infty}^{+\infty} e^{i\omega t} \chi_{ij}(t).$$

Consider now two perturbations, $\varepsilon \cos(\omega t)\vec{e}_j$ and $-\varepsilon \sin(\omega t)\vec{e}_j$ where $\vec{e}_j$, is the canonical basis vector in direction j and denote $\vec{V}^{(1)}, \vec{V}^{(2)}$ the corresponding perturbed dynamics. Then, one can show that:[18]

$$\hat{\chi}_{ij}(\omega) = \lim_{T\to+\infty} \frac{1}{T\varepsilon} \sum_{t=0}^{T-1} e^{i\omega(t-1)} \left[V_i^{(1)}(t) + i V_i^{(2)}(t) \right]. \qquad (37)$$

This allows the numerical computation of the complex susceptibility by time-averaging the trajectories of the two perturbed dynamics. Other numerical methods have been proposed in[1,5] that could be more efficient.

3. Resonances

The computation of $\hat{\chi}$ allows to exhibit resonances, in a similar way as in section III B, but with a very different structure. First, these resonances are not given in terms of eigenvalues of the Jacobian matrix $D\vec{G}_{\vec{V}}$ because, in contrast to section III B, $D\vec{G}_{\vec{V}}$ is now evolving dynamically along the strange attractor, as well as its eigenvalues. Rather than eigenvalues, Lyapunov exponents express the average expansion/contraction rates, but I don't know about any result relating the Lyapunov spectrum to resonances. Here, resonances are obtained numerically using the form (37). We now interpret these resonances, first, in the context of dynamical systems theory and statistical physics, and then in

the context of neuronal networks.

A classical wisdom coming from fluctuation-dissipation theorem in statistical physics is that the complex susceptibility is the Fourier transform of the corresponding correlation function, so that the resonances are peaks in the power spectrum. However, the implicit assumption underlying this result is that the equilibrium distribution has a density with respect to the Lebesgue (or Liouville) measure.

The situation is more complex in the case of strange attractors because the SRB measure is absolutely continuous only along the unstable manifold (parallel to the attractor) and is singular (fractal) transverse to the attractor. As a consequence, Ruelle's theory asserts that the linear response operator is the sum of two contributions. There is a regular term, corresponding to the response to perturbations "parallel" to the attractor (locally projected along the unstable manifold). This term is a correlation function and, consequently obeys the Fluctuation-dissipation theorem. The poles of its Fourier transform are the Ruelle-Pollicott resonances. They give the rate of mixing of the chaotic system, or, equivalently, the relaxation rate to equilibrium for a perturbation "on" the attractor. These poles are independent of the observable. Thus, in our case, they are independent on the pre-synaptic, post-synaptic neuron pair. These resonances are observed in Fig. 4.

There is a second term in the linear response operator, corresponding to the response to perturbations locally projected along stable manifolds, namely transverse to the attractor. Therefore, this term exists only in the dissipative case. It does not obey fluctuation-dissipation theorem and its resonances have a different structure. These exotic resonances, theoretically predicted by Ruelle[106], were, to my best knowledge exhibited for the first time by J.A. Sepulchre and myself in[18]. That was in the model (9). An example is shown in fig. 8 where, in blue is plotted the power spectrum (with peaks corresponding to Ruelle-Pollicott resonances) and in red are plotted the resonances in the complex susceptibility.

From the point of view of neuronal dynamics resonances have the following interpretation. In the chaotic regime neurons have firing rates evolving in an unpredictable and erratic way. Yet, there is an hidden dynamical structure revealed by the resonances in Fig. 8. First, upon exciting neuron j with a weak amplitude, harmonic signal, with resonant frequency ω superimposed upon the chaotic activity of neuron j, the effects of this stimulation will propagate through the network, and can actually be measured. Indeed, some neurons in the network will respond with a maximal amplitude (given by the modulus of $\hat{\chi}$). This effect can be exploited to transmit a signal with slow modulation through the network - despite chaos - using a resonance frequency as a carrier frequency. This signal, scrambled in the chaotic dynamics, can nevertheless be recovered on specific nodes, using an appropriate average procedure. See[19] for further details.

This leads to a define an effective connectivity (sketched by the blue arrow in Fig. 7) depending on the frequency. It does not coincide with the synaptic graph. It does not coincide either with a graph built on correlation functions precisely because the susceptibility has a contribution which is not a correlation function. This suggests that groups of neurons can synchronise upon excitation of the network with a suitable resonance frequency. These clusters of neurons would depend on the excitation frequency, supporting the idea that neurons can dynamically constitute far more clusters than expected from the study of the synaptic graph.

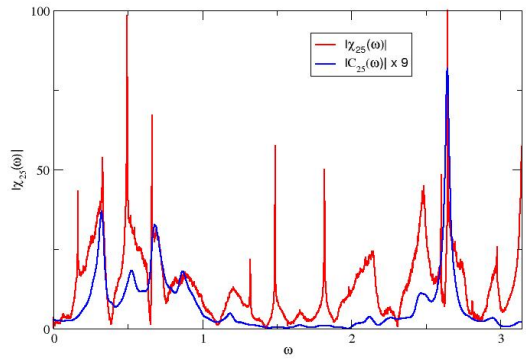

FIG. 8. Resonances in the response of a neuron (here 2) when exciting another neuron (here neuron 5) with an harmonic stimulus. In red we have plotted the modulus of $\hat{\chi}_{25}$ (eq. (36)) as a function of the real frequency ω. In blue is plotted the power spectrum of neuron j (this is the same as in fig. 4). The resonances which are not in the power spectrum are the stable resonances predicted by D. Ruelle.

4. *Uniform hyperbolicity*

Ruelle's theory mathematically requires the dynamical system (at least the dynamics on the attractor) be uniformly hyperbolic. Uniform hyperbolicity means that for for every $\vec{V} \in \Omega$ there is a splitting of the tangent space $T_{\vec{V}}\Omega = E_s(\vec{V}) \oplus E_u(\vec{V})$, and there are constants $C > 0$ and $0 < \lambda < 1$, such that, for every $n \in \mathbb{N}$, one has $\|D\vec{G}^n_{\vec{V}}.\vec{u}\| \leq C\lambda^n\|\vec{u}\|, \quad \vec{u} \in E_s(\vec{V})$ (exponential contraction on the stable space $E_s(\vec{V})$), and $\|D\vec{G}^{-n}_{\vec{V}}.\vec{u}\| \leq C\lambda^n\|\vec{u}\|, \quad \vec{u} \in E_u(\vec{V})$, (exponential expansion on the unstable space $E_s(\vec{V})$)[72]. In a nutshell this means that there are no neutral eigendirections (eigenvalues with modulus 1) for the derivative $D\vec{G}$.

Does the dynamical system (9) obey this condition ? This is a tricky question for which I only have partial answers. First, remark that we are dealing here with *random* attractors as each realisation of the J_{ij} generates a different attractor. So the answer to this question would require a probabilistic treatment. For a given realisation of J_{ij}s one can measure the Lyapunov spectrum and check that there are positive Lyapunov exponents. But having positive Lyapunov exponents is not enough to guarantee uniform hyperbolicity because neutral points may exist, as e.g. in the Henon model. What is known from the Newhouse-Ruelle-Takens theorem[87] is that the attractor resulting from a transition to chaos by quasi-periodicity can be perturbed by a smooth perturbation to give

rise to an axiom A (uniformly hyperbolic) attractor. So our attractor is, in this sense, "close" to uniform hyperbolicity, no more. From the shape of the Jacobian matrix and the form of the sigmoid, it is clear that there is a subset of the phase space where $D\vec{G}_{\vec{V}}$ has some eigenvalue with modulus 1. The question is whether the strange attractor Ω intersects this region. The answer depends on the matrix $\mathscr{J}$ and I don't know about any methods allowing to solve this question.

Therefore, the derivation made for the Amari Wilson Cowan model in the chaotic regime are made "as if" this system was uniformely hyperbolic. This assumption is called "chaotic hypothesis". It has been proposed by Gallavotti and Cohen[48,49].This conjecture agrees with the fact that many chaotic time evolutions behave as if they corresponded to uniformly hyperbolic dynamics.

Note that there are known example of chaotic attractors with neutral points (like the Henon's attractor) where linear response is violated[6,7,69,109]. The nice point is that this violation can be detected numerically[17]. We have not observed such effects in the examples of model (9) that we have studied.

F. Conclusion of section III

In this section, we have been able to answer the questions raised in the introduction using a linear response theory developed for chaotic dynamical systems using a form of Gibbs distribution, the SRB measure. We have been able to compute the convolution kernel in terms of the parameters and function ruling the non linear neurons dynamics. We have characterised how a stimulation of weak amplitude, applied to a group of neurons, influences the whole network. The functional connectivity obtained this way, clumsy sketched by the blue arrow in Fig. 7, is quite different from the synaptic connectivity as it results from a complex interplay between the network structure, the non linear dynamics and the statistics of orbits characterised by the SRB measure.

We want to address a few pending questions.

a. Information transport. Can we relate the effective connectivity to some notion of information transport that could be relevant for neuronal networks ? We stay here at a formal level and the following discussion would require further developments. A good candidate to quantify the effect of gently varying the voltage V_j of neuron j is the derivative $f'(V_j)$, or better, its log, $\log f'(V_j)$. Why the log ? Because the effect of f' is multiplicative along trajectories and because it controls the exponential expansion/contraction rate along the orbits of the dynamics. Actually, the average volume contraction rate in the discrete time Amari-Wilson-Cowan model is $\langle \log \left| \det D\vec{G} \right| \rangle_{eq} = \log |\det \mathscr{J}| + \sum_{j=1}^{N} \langle \log f'(V_j) \rangle_{eq}$. This average volume contraction, corresponds to an entropy production rate, as used in[50,106]. It has a static, synaptic network dependent term, and a dynamical term. From the analogy with (7), one may define a formal current form neuron j to neuron i by:

$$\vec{j}_{ij} = \frac{\partial \langle \log f'(V_i) \rangle_{eq}}{\partial V_j}. \tag{38}$$

This involves the derivative of the SRB state which can be computed by (32). This could be a formal way to define the blue arrow in Fig. 7. It remains however to check whether this formal current can be measured and how it relates to more classical indicators like information flow or Granger causality.

b. How large can be ε ? Linear response theory assumes that higher order terms are negligible. This assumption holds true for a range of ε values where the second order term is smaller than the first order. Thus, answering this question requires a priori compute the second order term. Although there is a general formulation of higher order terms[108], their computation has not been done in the present context. In the presented work, the validity of the linear response was numerically done, checking that e.g. dividing ε by 2 gives a response with an amplitude divided by 2. The typical values for ε were of order $10^{-2} - 10^{-3}$, thus small compared to the amplitude of the voltage. However, as pointed out in the introduction, linear response theory is only a first step toward characterising neural responses to stimuli. We further develop this point in the conclusion section.

IV. FROM SPIKING NEURONS DYNAMICS TO LINEAR RESPONSE

In this section, we address again the questions raised in the introduction, in a different context, for a spiking neuronal network. As most neurons communicate by spikes there are many models of spiking neurons that we can roughly divide in two categories: (1) (Non) linear smooth dynamical systems, based on (1), like Hodgkin-Huxley[64], FitzHugh-Nagumo[42,85] or Morris-Lecar[83] where the spike is a continuous function of time, shaped by the ionic currents involved in (1). These models are close to biophysics but remain quite hard to study at a theoretical level. For this reason researchers have developed a second category: (2) Non linear, non smooth dynamical systems, based on (1), like the Integrate and Fire model[76] and its generalisations[14,66,101] where the spike is a discontinuous function of time. The voltage V obeys equation (1) (subthreshold dynamics), with simplified ionic current terms, until V reaches a threshold. Then it is instantaneously reset to a rest value and a spike is emitted. This class of model is simpler to study and simulate although the introduction of a discontinuity at the threshold has a price to pay (non smoothness). In the case of the linear response theory developed here, this requires to use rather different tools than the previous section.

Here, I present a summary of work done in collaboration with Rodrigo Cofré where we developed a linear response theory in a conductance-based integrate and fire model[23,24,28]. As in the previous section, we are able to derive a linear response formula where the response kernel is written in terms of spike correlations, with an explicit dependence in the networks parameters. This work is in line with other approaches

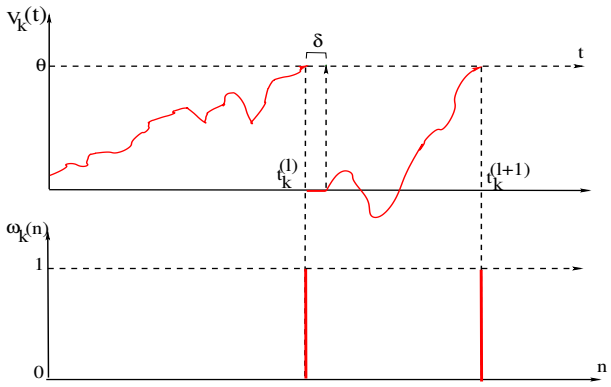

FIG. 9. Integrate and Fire dynamics. When the voltage V_k reaches the threshold θ, at time $t_k^{(l)}$, it is reset to 0 and a spike is recorded (Fire). Voltage stays at this value during a time interval δ. Then evolution of V_k starts again (Integrate).

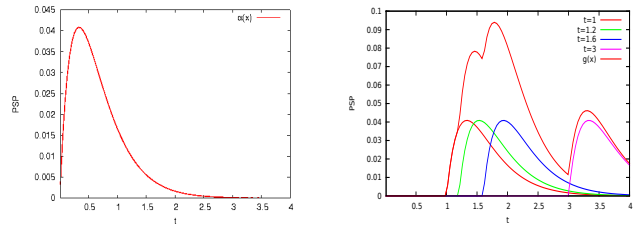

FIG. 10. Conductance of the model. Left. The function α of eq. (40). Right. Spikes are emitted at times $1,3,12,16$, generating an increase in the conductance g_{kj} controlled by the α function (40). The total conductance is the sum of the α profiles.

attempting to understand how the correlated spiking activity of a neuronal network reflects the neurons interactions as well as their collective response to stimuli. Modelers have proposed methods based on statistical physics (Maximum Entropy Models[41,111,114,115], stochastic processes (Markov chain, Hawkes processes[98,99]) or phenomenological models (Linear-non linear, Generalized Linear Models[26,92,116]) to achieve this purpose. Yet, there are rather few modelling studies trying to relate the collective dynamics of neurons with the spike statistics of the network, in spontaneous activity as well as in the presence of a stimulus.

A. Model

1. Spike emission

As in the previous section we consider N neurons with voltage V_k, $k = 1 \dots N$. But dynamics is of a different nature because we focus here on spike statistics. Spike emission is a complex process[30]. In Integrate and Fire models this process is quite simplified. The principle is illustrated in fig. 9.

We fix a voltage threshold θ. Below threshold, V_k follows and evolution based on eq. (1) (see eq. (45) below). If the neuron k's voltage reaches the threshold at time t this neuron emits a spike. Then, its voltage is reset to a rest value (taken here to be 0 without loss of generality) and the neuron stays at this value (quiescent) during a time interval $\delta > 0$.

We note $t_k^{(l)}$ the time of occurrence of the l-th spike emitted by neuron k. We define a spiking variable $\omega_k(n) \in \{0,1\}$ where n is an integer. We set $\omega_k(n) = 1$ if neuron k emits a spike in the time interval $[n\delta, (n+1)\delta[$ and $\omega_k(n) = 0$ otherwise. This reads:

$$\omega_k(n) = \begin{cases} 1, & \text{if} \quad \exists l, t_k^{(l)} \in [n\delta, (n+1)\delta[\\ 0, & \text{otherwise.} \end{cases} \tag{39}$$

Spiking variables are therefore time-discrete events with a time resolution δ. We define the spike pattern of the network at time n by the vector $\omega(n) = \big(\ \omega_k(n) \ \big)_{k=1}^N$. A spike block ω_m^n, $m < n$, is the matrix of spike patterns $\big(\omega(m), \dots, \omega(n)\big)$. A spike train is a bi-infinite spike block $\omega_{-\infty}^{+\infty}$. We note it ω for simplicity.

In this section we are going to consider a mixed dynamics with continuous time variables and discrete time variables. Especially, we will consider functions of the type $f(t,\omega)$ where t is the continuous time variable and ω the discrete time spike train. In this notation, however, $f(t,\omega)$ signifies $f(t,\omega_{-\infty}^{[t]})$ where $[t]$ is largest integer smaller or equal than t. This condition expresses causality: the function $f(t,\omega)$ depends on the spikes emitted *before* time t.

2. Sub-threshold dynamics

The subthreshold dynamics of neuron k is based on a model proposed by M. Rudolph and A. Destexhe in[101]. It starts from the charge conservation equation (1), where the ionic current is a simple, passive leak term $-g_L(V_k - E_L)$. The external current (stimulus) takes the form $\varepsilon S_k(t)$ and the noise term reads $\sigma_B \xi_k(t)$ where $\xi_k(t)$ is a white noise. σ_B controls the amplitude of the noise.

The synaptic current from pre-synaptic neuron j to post-synaptic neuron k reads $-g_{kj}(t,\omega)(V_k - E_{kj})$ where E_{kj} is the reversal potential associated with the synapse $j \to k$. In this model, the synaptic conductance g_{kj} depends on time and on the previous spiking history of the pre-synaptic neuron j. Whenever neuron j emits a spike (at time $t_j^{(n)}$) the conductance increases by an amount $G_{kj}\alpha(t - t_j^{(n)})$, where G_{kj} is the maximal conductance, and:

$$\alpha(t) = \frac{t}{\tau} e^{-\frac{t}{\tau}} H(t), \tag{40}$$

mimics the time profile in the synaptic increase upon emission of a pre-synaptic spike (Fig. 10). Here, $H(t)$ is the Heaviside function.

The total conductance is $g_{kj}(t) = G_{kj} \sum_{n \geq 0} \alpha(t - t_j^{(n)})$. It depends therefore on the spike history, represented by the spike times $t_j^{(n)}$ preceding t. The set of all possible such times is uncountable. In order to have a dependence in a countable set of events we use the spike time discretisation (39) to ob-

tain:

$$g_{kj}(t,\omega) = G_{kj} \sum_{n\geq 0} \alpha(t-n\delta)\omega_j(n) \tag{41}$$

Setting:

$$W_{kj} = G_{kj}E_{kj}, \tag{42}$$

$$\alpha_j(t,\omega) = \sum_{n\geq 0} \alpha(t-n\delta)\omega_j(n), \tag{43}$$

$$i_k(t,\omega) = g_L E_L + \sum_j W_{kj}\alpha_j(t,\omega) + \varepsilon S_k(t) + \sigma_B \xi_k(t), \tag{44}$$

we arrive at the final equation for the sub-threshold dynamics of neuron k:

$$C_k \frac{dV_k}{dt} + g_k(t,\omega)V_k = i_k(t,\omega), \quad \text{if} \quad V_k < \theta. \tag{45}$$

Let us summarise. Spike time is discretised in time bins $n\delta$. Voltage, current and conductance time t is continuous. When the voltage of neuron k, V_k, reaches the threshold, $V_k(t_k^{(l)}) = \theta$, with $t_k^{(l)} \in [n\delta,(n+1)\delta[$ for some n it is reset to 0, and the l-th spike of neuron k is recorded at discrete time n, $\omega_k(n) = 1$. Voltage stays at 0 until time $(n+1)\delta$ where it follows the sub-threshold evolution (45) until the next time where V_k reaches the threshold. Note that, in this modelling, δ can be quite small compared to the time scales of the dynamics.

Remark. Although eq. (45) looks quite simple (it is a differential equation, linear in V_k, with time dependent coefficients) it hides a real complexity, the dependence in the history ω. The coefficients $g_k(t,\omega), i_k(t,\omega)$ depends on the trajectory of V_k which itself depends on the history of spikes anterior to t. This involves compatibility conditions between the trajectory and the spike train ω which constitutes a symbolic coding of trajectories. These aspects are further discussed in[22].

3. Solutions

For a time t, a spike train ω and a neuron k we note $\tau_k(t,\omega)$ the last time anterior to t where the neuron membrane potential was reset.

It is easy to integrate the linear system (45) from time $\tau_k(t,\omega)$ to time t. The corresponding flow is:

$$\Gamma_k(t_1,t,\omega) = \begin{cases} e^{-\frac{1}{C_k}\int_{t_1}^t g_k(u,\omega)\,du}, & \text{if} \quad t \geq t_1 \geq \tau_k(t,\omega); \\ 0, & \text{otherwise}. \end{cases}$$

We obtain:

$$V_k(t,\omega) = \underbrace{V_k^{(eq)}(t,\omega) + V_k^{(S)}(t,\omega)}_{V_k^{(det)}(t,\omega)} + V_k^{(noise)}(t,\omega), \tag{46}$$

where:

$$V_k^{(eq)}(t,\omega) = V_k^{(syn)}(t,\omega) + V_k^{(L)}(t,\omega),$$

is the spontaneous contribution with:

$$V_k^{(syn)}(t,\omega) = \frac{1}{C_k}\sum_{j=1}^N W_{kj} \int_{\tau_k(t,\omega)}^t \Gamma_k(t_1,t,\omega)\,\alpha_j(t_1,\omega)\,dt_1,$$

the synaptic interaction term, and,

$$V_k^{(L)}(t,\omega) = \frac{E_L}{\tau_{L,k}} \int_{\tau_k(t,\omega)}^t \Gamma_k(t_1,t,\omega)\,dt_1,$$

where:

$$\tau_{L,k} \stackrel{\text{def}}{=} \frac{C_k}{g_L}.$$

The second term in (46) corresponds to the contribution of the external stimulus:

$$V_k^{(S)}(t,\omega) = \varepsilon \frac{1}{C_k} \int_{\tau_k(t,\omega)}^t S_k(t_1)\Gamma_k(t_1,t,\omega)\,dt_1. \tag{47}$$

The last one is the stochastic part of the membrane potential:

$$V_k^{(noise)}(t,\omega) = \frac{\sigma_B}{C_k} \int_{\tau_k(t,\omega)}^t \Gamma_k(t_1,t,\omega)\,dB_k(t_1),$$

where $B_k(t_1)$ is a Brownian process, thus Gaussian.

B. Gibbs distribution

We are interested in the statistic of spikes generated by this model, in spontaneous activity and in the presence of the stimulus. We want to propose a linear response theory in this context. For this we first show that spike statistics is associated to a form of Gibbs distribution.

1. Transition probabilities

Here, we want to compute the probability to have a spiking pattern $\omega(n)$ given the history, something informaly reading like $\mathbb{P}_n[\,\omega(n)\,|H_{<n}\,]$, where $H_{<n}$ is the history anterior to n, depending on voltage and spikes history. To simplify this dependence, we are going to make the approximation that the spike pattern $\omega(n)$ depends only on the spike history. This makes sense if one wants to use this theoretical approach to analyze experimental data on spike trains recordings with Multi-Electrode Arrays for example. Here, indeed, one has only access to spikes history, not to voltage[88].

Under this assumption $H_{<n}$ identifies with $\omega_{-\infty}^{n-1}$, where the memory extends in principle to the far past (here $-\infty$). This is an important point. In Integrate and Fire models voltage memory is reset when the neuron spikes, so memory goes back up to the last time anterior to n where the neuron has spiked. However, this time can be quite far in the past, giving rise to variable length Markov chain, as illustrated in Fig.

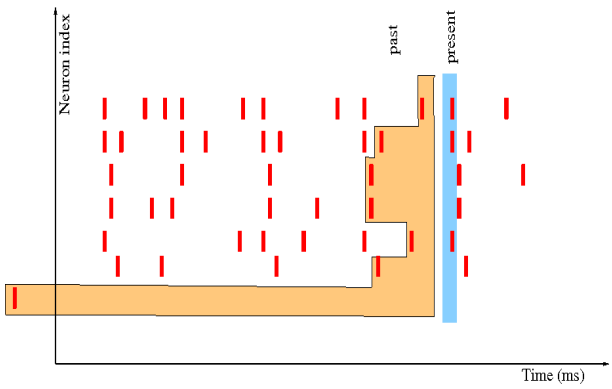

FIG. 11. Conditional probability of a spiking pattern (in blue) given the history. In Integrate and Fire models voltage memory (in orange) is reset when the neuron spikes (in red), so memory goes back up to the last time anterior to n where the neuron has spiked (red spike in the orange area). This time is variable and can extend quite far in the past, giving rise to variable length Markov chain.

11. In addition, the spike memory does not only depend on voltage, it is also depends on conductances, as expressed in (41), so memory is, in principle, infinite. We are therefore seeking probabilities of the form $P_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]$, where the subscript n expresses that these probabilities depend on general on the discrete time n.

One can show[22] that these transition probabilities are well approximated by:

$$\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]=\prod_{k=1}^{N}\mathbb{P}_n\left[\,\omega_k(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right], \tag{48}$$

where:

$$\mathbb{P}_n\left[\,\omega_k(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]=\omega_k(n)\,\Pi\left(\frac{\theta-V_k^{(det)}(n-1,\omega)}{\sigma_k(n-1,\omega)}\right)+\left(1-\omega_k(n)\right)\left(1-\Pi\left(\frac{\theta-V_k^{(det)}(n-1,\omega)}{\sigma_k(n-1,\omega)}\right)\right), \tag{49}$$

where $V_k^{(det)}(n-1,\omega)$ is the deterministic part of the voltage, given by (46) and:

$$\sigma_k^2(n-1,\omega)=\left(\frac{\sigma_B}{C_k}\right)^2\int_{\tau_k(n-1,\omega)}^{n-1}\Gamma_k^2(t_1,n-1,\omega)\,dt_1,$$

corresponds to the variance of the noise integrated along the flow up to time $n-1$. Finally

$$\Pi(x)=\frac{1}{\sqrt{2\pi}}\int_x^{+\infty}e^{-\frac{u^2}{2}}\,du.$$

2. Gibbs distribution

Probabilities of this form define a generalisation of Markov chain called *chains with complete connections*, a notion introduced by Onicescu and Mihoc in 1935[90]. The terminology chains with infinite memory can be also found. Under suitable conditions[40,81] these transition probabilities define a probability μ on the set of spike trains which is a generalisation of the probability consistent with a Markov chain:

$$\begin{aligned}&\int h\left(\omega_{-\infty}^{n}\right)\mu(d\omega)\\&=\int\sum_{\omega(n)\in\mathscr{A}}h\left(\omega_{-\infty}^{n-1}\omega(n)\right)\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]\mu(d\omega).\end{aligned} \tag{50}$$

with $\mathscr{A}=\{0,1\}^N$. This equality must hold for all $n\in\mathbb{Z}$ and measurable functions h. Obviously, the fact of conditioning by an infinite past requires some caution and conditions on the transition probabilities to ensure the existence of μ.

μ has actually strong analogies with Gibbs distributions in rigorous statistical mechanics, where the set of spike trains can be viewed as a one dimensional spin chain labeled by the time index. An important difference, though, is that Gibbs distributions are constructed by conditioning upon left and right boundary conditions. In contrast, in the present context we only condition upon the past (left specification). For this reason, μ is rather called a Left Interval Specification (LIS)[40,89]. There are mathematical examples showing that LIS have different properties than left-right conditionned Gibbs distribution[39]. However, to the best of my knowledge there is no effective, operational way, to distinguish these two notions from a finite sample obtained e.g. from numerical simulations, as used most of the time in the field of computational neuroscience. As a consequence, I will not distinguish these two notions and call μ a Gibbs distribution.

There are standard theorems ensuring the existence and uniqueness of a Gibbs distribution in this sense. In our case, these theorems apply because of the exponential decay of conductances (41) which induces an exponential decay of memory, the continuity of the family of transition probabilities and the summability of their variations (see[22,28] for details).

A consequence of the definition (50) is that the probability of a spike block ω_m^n, given the past reads:

$$\mathbb{P}_n\left[\,\omega_m^n\,\middle|\,\omega_{-\infty}^{m-1}\,\right]=e^{\Phi(m,n,\omega)}=e^{\sum_{r=m}^{n}\phi(r,\omega)} \tag{51}$$

where:

$$\Phi(m,n,\omega)=\sum_{r=m}^{n}\phi\left(r,\omega\right), \tag{52}$$

and

$$\phi\left(n,\omega\right)\equiv\log\mathbb{P}_n\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right], \tag{53}$$

so that $\phi\,(n,\omega)$ has the form of a Gibbs energy with infinite range. This "energy" being the log of a probability, there is no need to normalise with a partition function. We call ϕ a normalised energy (although it does not have the physical dimension of an energy, it is rather an information - bit rate). It depends explicitly on the model parameters via eq. (46). It contains, as well, the stimulus influence, via the term (47). We note $\phi^{(eq)}\,(\,\omega\,)$ the energy in spontaneous activity, $\varepsilon = 0$ and $\mu^{(eq)}$ the corresponding Gibbs distribution. It does not depend on n as dynamics and transition probabilities are stationary in this case. $\phi^{(eq)}\,(\,\omega\,)$ plays therefore the role of energy in the equilibrium Gibbs distribution (4). Yet, it does not have the form (5). We come to this point in section IV C 2 (eq. (57)).

C. Linear response

1. General form

Assume now that the spiking neuronal network receives a time-dependent stimulus $S(t)$ from time t_0 to time t_1. Even if the stimulus is applied to a subset of neurons in the network, its influence will eventually propagate to other neurons, directly or indirectly connected. The stimulus will act on spikes timing, modifying spike correlations. We note $n_0 = [t_0]$, the integer part of t_0. For times anterior to n_0, μ identifies with $\mu^{(eq)}$, that is, for any $m < n \leq n_0$, for any block ω_m^n, $\mu\,[\,\omega_m^n\,] = \mu^{(eq)}\,[\,\omega_m^n\,]$. In contrast, for $n > n_0$ spike statistics is modified.

We consider a function (observable) $f(t,\omega)$. How is its average modified by the application of the stimulus ? We set $\mathbb{E}_\mu\,[\,f(t,.)\,] \stackrel{\text{def}}{=} \mathbb{E}_{\mu^{(eq)}}\,[\,f(t,.)\,] + \delta\,[\,f(t,.)\,]$ where $\delta\,[\,f(t,.)\,] = 0$ for $t < t_0$ and $\delta\,[\,f(t,.)\,] \neq 0$ for $t \geq t_0$. We want to compute $\delta\,[\,f(t,.)\,]$ when ε, the stimulus amplitude, is weak enough. Using a formal expansion of the energy (53) one obtains $\delta\mu\,[\,f(t)\,]$ in the classical convolution form[24]:

$$\delta\mu\,[\,f(t)\,] = \varepsilon\,\left[\,\kappa_f * S\,\right](\,t\,), \tag{54}$$

where the convolution kernel takes the form:

$$\kappa_{k,f}\,(\,t-t_1\,) = \frac{1}{C_k}\sum_{r=-\infty}^{[t-t_1]} C^{(sp)}\left[\,f(t-t_1,.),\frac{\mathscr{H}_k^{(1)}(r,.)}{\sigma_k(r-1,.)}\,\Gamma_k(0,r-1,.)\right]. \tag{55}$$

Here the function $\mathscr{H}_k^{(1)}$ is obtained via a first order expansion of (53). It contains therefore in particular the synaptic weights (42). The notation $C^{(sp)}\,[\,A(t,.)B(s,.)\,]$ signifies the correlation of the functions $A(t,\omega)$ and $B(s,\omega)$ where the averaging is done with respect to the spike train distribution, in spontaneous dynamics, $\mu^{(eq)}\cdot$ Thus, the linear response kernel is here as well obtained as a sum of correlation functions computed with respect to the spontaneous dynamics. Note that the kernel depends on the flow of the dynamics $\Gamma_k(0,r-1,.)$ and on noise (term $\sigma_k()$). As in the equation (36) computed in section III D, the linear response kernel is obtained by a summation on the whole spike history.

2. Approximations

a. Markovian approximation. The exponential decay of the synaptic response α in (43) entails the existence of a time scale, τ, after which the dependence in the past is essentially lost. This suggests to use a Markovian approximation where the transition probabilities $P\left[\,\omega(n)\,\middle|\,\omega_{-\infty}^{n-1}\,\right]$ with infinite memory are replaced by $P\left[\,\omega(n)\,\middle|\,\omega_{n-D}^{n-1}\,\right] > 0$, with a fixed memory depth $D > 0$. Dynamics is then given by a Markov chain and $\mu^{(eq)}$ is the invariant probability of this chain. The memory depth D of the chain is constrained by the characteristic times:

$$\tau_{d,k} = \frac{C_k}{g_L + \tau\sum_{j=1}^N G_{kj}\nu_j}, \tag{56}$$

where ν_j is the firing rate of neuron j[24].

b. Hammersley-Clifford decomposition. In the Markovian approximation the energy (52) becomes a real function of spike blocks ω_0^D. A general theorem from Hammersley and Clifford[63] states that the normalised energy $\phi^{(eq)}\,(\,\omega\,)$ can then decomposed on a basis of interaction functions:

$$\phi^{(eq)}\,(\,\omega\,) = \sum_l \phi_l\, m_l\,(\,\omega\,) \tag{57}$$

where:

$$m_l(\omega) = \prod_{k=1}^n \omega_{i_k}(t_k). \tag{58}$$

where $i_k = 1\ldots N$ is a neuron index, and $t_k = 0\ldots D$. Thus, $m_l(\omega) = 1$ if and only if, in the spike train ω, neuron i_1 spikes at time $t_1,\ldots,$ neuron i_k spikes at time t_k. Otherwise, $m_l(\omega) = 0$. The number n is the degree of the interaction; degree one

interactions have the form $\omega_{i_1}(t_1)$, degree 2 interactions have the form $\omega_{i_1}(t_1)\omega_{i_2}(t_2)$, and so on. These interactions involve in general a time delay between spikes.

The form (57) bares analogy with the energy form (5) where the interactions m_l play the role of the X_αs. More generally, an energy of the form:

$$H(\omega) = \sum_\alpha \lambda_\alpha X_\alpha(\omega) \tag{59}$$

where the $X_\alpha(\omega)$s are functions of blocks ω_0^D is associated to a Markov chain via the Perron-Frobenius theorem[55,113]. Under moderate assumptions (λ_α are bounded from below), this chain has a Gibbs invariant probability[29].

In the simplest case, ($D = 0$, pairwise interactions) this probability is the Gibbs distribution of an Ising model[29]. This has attracted much interest in the computational neuroscience community although the Ising form is just the simplest non trivial potential existing in this context. We come back to this point below.

c. Further approximations. It is possible to simplify the form (55) with the following approximations:

(i) Replace $\tau_k(r-1,\omega)$ by $-\infty$;

(ii) Replace $\Gamma_k(t_1, r-1, \omega) = e^{-\frac{1}{C_k}\int_{t_1}^{r-1} g_k(u,\omega)\,du}$ by $e^{-\frac{(r-1-t_1)}{\tau_{d,k}}}$.

Then, linear response reads:

$$\delta^{(1)}\mu[f(t)] = -\frac{2}{\sigma_B}\sum_{k=1}^{N}\frac{1}{\sqrt{\tau_{d,k}}}\sum_{r=-\infty}^{n=[t]}\left[\quad\right]\left(S_k * e_{d,k}\right)(r-1) \tag{60}$$

where:

$$e_{d,k}(u) = e^{-\frac{u}{\tau_{d,k}}}$$

and where the term $\left[\quad\right]$ is an expansion of correlation functions between the observable f and spikes interactions. In the case of a memory depth $D = 1$ it reads :

$$\left[\quad\right] = \gamma_k^{(1)}\mathscr{C}^{(eq)}[f(t,\cdot),\omega_k(r)] + \sum_{i=1}^{N}\gamma_{k;i}^{(2)}\mathscr{C}^{(eq)}[f(t,\cdot),\omega_k(r)\,\omega_i(r-1)] + \sum_{i,j=1}^{N}\gamma_{k;ij}^{(3)}\mathscr{C}^{(eq)}[f(t,\cdot),\omega_k(r)\,\omega_i(r-1)\,\omega_j(r-1)] + \cdots \tag{61}$$

where correlations are computed with respect to the equilibrium probability. For example, the variation induced by the stimulus, in the firing rate of neuron m, at time t is given by:

$$\delta^{(1)}\mu[\omega_m(t)] = -\frac{2}{\sigma_B}\sum_{k=1}^{N}\frac{1}{\sqrt{\tau_{d,k}}}\sum_{r=-\infty}^{n=[t]}\left[\begin{array}{ll} & \gamma_k^{(1)}\mathscr{C}^{(eq)}[\omega_m(t),\omega_k(r)] \\ +\sum_{i=1}^{N} & \gamma_{k;i}^{(2)}\mathscr{C}^{(eq)}[\omega_m(t),\omega_k(r)\,\omega_i(r-1)] \\ +\sum_{i,j=1}^{N} & \gamma_{k;ij}^{(3)}\mathscr{C}^{(eq)}[\omega_m(t),\omega_k(r)\,\omega_i(r-1)\,\omega_j(r-1)] \\ + & \cdots \end{array}\right]\left(S_k * e_{d,k}\right)(r-1) \tag{62}$$

More generally, it involves correlations to pairwise, triplets, etc spike time correlations. The coefficients $\gamma_k^{(l)}$ depends on the synaptic weights W_{kj} and are therefore constrained by neurons interactions.

D. Conclusions of section IV

In this section, we have derived a linear response for a spiking neural network. As for the Amari Wilson Cowan model we end up with a convolution kernel depending on synaptic

graph and equilibrium correlations. Yet, the form obtained is quite more complex than the form (36), and quite harder to interpret. The interesting point is that the terms of this expansion, which are spike-time correlations computed at equilibrium, can be obtained by an ergodic, time average, in spontaneous activity. The numerical computation of high order terms is, however, cumbersome and will be the subject of a forthcoming paper.

As in section III the correlations decay exponentially fast, so that it is possible to truncate the expansion (60) to low orders, e.g. pairwise interactions for neurons, and small memory depth, as e.g. eq. (62). Yet, many terms remain. At the lowest non trivial order the spontaneous energy contains only synchronous pairwise interactions of the form $\omega_i(0)\omega_j(0)$ and the energy (59) corresponds to a Ising model, where successive times are independent. An expansion similar to (60) has be done by S. Cocco et al[27]. The Ising model has been used by many authors to analyze spike trains statistics, especially in retina data[56,111,114,124,125]. It neglects, however, higher order correlations which have been shown to play a role in spike response to stimuli[53,54,129], e.g. in the retina.

This work is at a less advanced stage than the previous one and there remain some pending questions:

a. Resonances. Submitting this spiking network to an harmonic stimulus, one may observe resonances, like in the Amari-Wilson-Cowan model. In the Markovian case (finite memory with depth D) these resonances are the eigenvalues of the transition matrix defined through the transition probabilities $P\left[\,\omega(n)\,\middle|\,\omega_{n-D}^{n-1}\,\right]$. They are difficult to compute directly though, even numerically, because the dimension of the matrix growths exponentially fast with the number of neurons and the memory depth D. Analytic computations seem out of reach. Therefore, resonances would have to be computed the same way as in section III E 2, eq. (37). We expect the same observations as for the Amari-Wilson-Cowan model: frequency dependent effective connectivity and frequency dependent clusters of synchronising neurons.

b. Structure of spike correlations. One of the goal guiding the work presented in this section is to understand how the spatio-temporal correlations present in the trajectory of a moving object are reflected in a spiking neuronal network when this moving object is the stimulus. Cells are excited by the moving object, but, as they are connected, this excitation triggers a wave of activity in the network which can for example enhance the effect, inducing anticipation[123]. We are essentially interested in applying this formalism to understand how the retina encodes motion, and how can one decode this motion for the spatio-temporal correlations observed in the output retinal cells (Ganglion cells).

V. CONCLUSION

In this paper we have considered, at a modelling level, the effects induced on a neuronal network by the weak stimulation of a sub group of neurons, using linear response theory. The goal was actually twofold: (i) analyse this effect starting from the equations ruling the neurons dynamics; (ii) make a link to linear response in non equilibrium statistical physics . For this we borrowed existing results, either coming from ergodic theory and chaotic dynamical systems, or from Markov chain and chains with complete connections. In the two proposed examples we ended up in a relation expressing the linear response to the stimulus as a convolution of this stimulus with a kernel. The form of the kernel is complex: as for statistical physics (e.g. kinetic theory) it depends on the microscopic dynamics, on the interactions (in our case, the synapses) and it is obtained via a suitable averaging that smooths the microscopic trajectory and establish a description at a mesoscopic level. The averaging, performed with respect to the equilibrium distribution, corresponds to an ergodic time-average, with no necessity to consider a thermodynamic limit where the number of neurons tends to infinity.

The analogy with statistical physics can go relatively far allowing us to define susceptibility, currents and Green-Kubo like relations. However, one of the main difficulty here is to obtain *a priori* the correct form for the "energy" (5). In physics this form is guided by first principles, mechanics or thermodynamics. What could be the equivalent principles in neuroscience, if any ? At the moment we are faced to two possible strategies: Either use formal analogies with statistical physics, e.g. proposing Ising model as a canonical or lowest order normal form for the energy, or try and extract the energy form from microscopic dynamics. The first approach raises the question of the role played by higher order terms and the way how to characterise their effects. The second approach resembles the program initiated by Boltzman to fund thermodynamics from mechanics, a program far from being completed yet, despite deep progresses[50]. Finally, the pending question is: why should statistical physics give fruitful insight in the understanding of neuronal dynamics and, more generally, neuroscience ? Beyond the fact that this conceptual analogy has lead to interesting new concepts in neuroscience, like in Friston's theory[44], the answer lies maybe above statistical physics, in large deviations and Markov chains[70].

This is actually the hidden link between section III and section IV. Indeed, a way to define the SRB state and to derive a linear response theory for chaotic systems is to use a salient property of chaotic (uniformely hyperbolic-like) dynamical systems. They have Markov partitions and can be encoded by Markov chains. Hyperbolicity allows indeed to split the phase space into a finite partition, constructed from local pieces of stable and unstable manifolds, and to encode the dynamics by a Markov transition matrix between the elements of this partition. The transition probabilities of this chain are weighted by the exponential of the potential (28). The SRB state is the invariant probability of this chain, while the induced large deviations of the entropy production allows to define currents and Onsager coefficients[48–50].

An interesting alternative approach, that could be fruitful when applied to neuronal models, has been developed by C. Maes in co-workers in[5,8]. It is based on dynamical operator (also called Koopman operator[57]) acting on observables, whose adjoin is the Ruelle-Perron-Frobenius operator. This

approach somewhat synthesises and generalises previous approaches giving a fast and compact way to obtain the linear response and higher order terms. It would be interesting to apply this method to neuronal network although it might be difficult to obtain a tractable form of the evolution operator.

More generally, the main obstacle to apply methods coming from non equilibrium statistical physics is that they have mainly be developed for physical problems and most examples are guided by physics and thermodynamics. In contrast, we don't have such guidelines here. We have no a priori idea of what is the form of the energy, and thermodynamics principles are useless to understand dynamics; for example there is no notion of "heat" or "work" that would allow to understand (8) or (45). The "thermodynamics of the brain" is still under investigation[73].

Returning to linear response, we would like to address several questions left aside in the paper.

a. Beyond linear response. In this paper, the linear response has been derived from an expansion in ε, the amplitude of the perturbation. What about the higher order terms that we have neglected ? In the context of chaotic systems higher order terms have been computed by Ruelle[108] and are therefore accessible mathematically in the model (9). Their form is quite hard to interpret in our context though, in contrast to the simple first order term of eq. (36). They are also very hard to estimate numerically or experimentally. It is known from Volterra expansion theory that higher order terms are given by high order correlations, which are quite difficult to estimate experimentally, requiring very large samples. We are confronted here to a similar problem.

If we stick at the questions addressed in this paper, the ε-expansion approach does not seem reasonable beyond the first order. In the field of neuroscience researchers prefer to correct the linear response convolution by a static non linearity[100]. In my opinion, the main interest of linear response theory in the context of neuronal modelling is to give us a notion of derivative of a statistical quantity (the average of an observable) with respect to a time dependent perturbation, in terms of the dynamics ruling the evolution of neurons. This leads, in the case studied here, to an explicit form for the convolution kernel, especially how it depends on dynamics and synaptic connections. In this setting, one sees explicitly that the "response" of a cell is not intrinsic to that cell, but depends on its dynamical surrounding. This approach could be useful to infer the receptive field of sensory cells from the structure of the neuronal circuits they are involved in. For example, it is known, in the retina, that the center-surround receptive field structure of bipolar and ganglion cells is due to the lateral inhibitory connectivity of horizontal cells. More complex circuits, processing motion, exist as well in the retina, involving amacrine cells whose role is far from being understood yet[61]. The present approach could afford to anticipate the role played by these cells in shaping the receptive field of cells responding to motion features.

b. Effective interactions. Linear response lead us to define a notion of effective interactions from the underlying non linear dynamics. As we argued, these interactions are not the synaptic connections, and they do not reduce to correlations. It is commonly accepted in the neuroscience community that "information" is transported by neurons. This is characterised by mutual information, relative entropy, Granger causality, ... In this paper we came out with the proposal (38), still on a shaky ground, based on the non linear effects induced by the sigmoid and a notion of entropy production. A next step is to investigate how to compute this quantity and how it compares with standard indicators.

c. Closeness to bifurcations. Neurons can exhibit drastic changes (bifurcations) in their behaviour when they are stimulated with a stimulus of increasing amplitude. The same holds as well in a neuronal network. If some neurons are close to a bifurcation point, a tiny stimulation of a single neuron can drastically change its activity, which, in turn, can induce bifurcations of other neurons in an avalanche like or wave activity. Thus, a small perturbation leads to a macroscopic change (diverging susceptibility). Such mechanism play an important role in neuronal dynamics. It is a current belief, observed in models, that learning and plasticity drive neuronal networks close to bifurcations points where they respond fast and optimally to learned stimuli[10,86,118,119]. On mathematical grounds, near bifurcations point, there is a loss of structural stability that can ruin any hope to have a linear response theory although structural stability is not necessary to obtain linear response and can be extended near bifurcations point under some conditions[7].

Linear response can be useful though to characterise the approach to such critical points. For example, the divergence of susceptibility could correspond to resonances converging to the real axis, in a Lee-Yang like phenomenon[133], providing a strong analogy between the behaviour of neuronal networks near bifurcations and phase transitions.

ACKNOWLEDGMENTS

I would like to acknowledge Jacques-Alexandre Sepulchre and Rodrigo Cofré from the good time we spend together thinking and working on the linear response results presented in this paper. I am grateful to Maria-José Escobar, Patricio Orio, Rodrigo Cofré and Wael El-Deredy to give me the opportunity to teach this lecture in the LACONEU summer school 2019. I would like to acknowledge the students of this school for their questions and their enthusiasm.

I thank the Reviewers for their insightful comments and constructive criticism.

[1] R. V. Abramov and A. J. Majda. New approximations and tests of linear fluctuation-response for chaotic non-linear forced-dissipative dynamical systems. *Journal of Nonlinear Science*, 18(3):303–341, 2008.

[2] K. Aihara, T. Takabe, and M. Toyoda. Chaotic neural networks. *Physics Letters A*, 144(6):333 – 340, 1990.

[3] S. Amari. Characteristics of random nets of analog neuron-like elements. *Syst. Man Cybernet. SMC-2*, 1972.

[4] V. I. Arnold. *Geometrical Methods in the Theory of Ordinary Differential Equations*. Springer–Verlag New York Inc., 1983.

[5] M. Baiesi and C. Maes. An update on the nonequilibrium linear response. *New Journal of Physics*, 15(1):013004, 2013.

[6] V. Baladi. On the susceptibility function of piecewise expanding interval maps. *Communications in Mathematical Physics*, 275(3):839–859, 2007.

[7] V. Baladi, M. Benedicks, and D. Schnellmann. Linear response, or else. ICM Seoul 2014 talk, August 2014.

[8] U. Basu and C. Maes. Nonequilibrium response and frenesy. *Journal of Physics: Conference Series*, 638:012001, 2015.

[9] D. Battaglia, I. Guyon, V. Lemaire, J. Orlandi, B. Ray, and J. Soriano. *Neural Connectomics Challenge*. Springer, 2017.

[10] N. Bertschinger and T. Natschläger. Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, 16(7):1413–1436, 2004.

[11] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lect. Notes.in Math.* Springer-Verlag, New York, 1975.

[12] R Bowen and D Ruelle. The ergodic theory of axiom A flows. *Inven. Math.*, 29:181–202, 1975.

[13] P. Bressloff. *Waves in Neural Media*. Lecture Notes on Mathematical Modelling in the Applied Sciences. Springer, 2014.

[14] R. Brette and W. Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, 94:3637–3642, 2005.

[15] J. Cabral, M. L. Kringelbach, and G. Deco. Functional connectivity dynamically evolves on multiple time-scales over a static structural connectome: Models and mechanisms. *NeuroImage*, 160:84 – 96, 2017. Functional Architecture of the Brain.

[16] B. Cessac. Increase in complexity in random neural networks. *Journal de Physique I (France)*, 5:409–432, 1995.

[17] B. Cessac. Does the complex susceptibility of the hénon map have a pole in the upper-half plane ? a numerical investigation. *Nonlinearity*, 20:2883–2895.

[18] B. Cessac and J.A. Sepulchre. Stable resonances and signal propagation in a chaotic network of coupled units. *Phys. Rev. E*, 70(056111), 2004.

[19] B. Cessac and J.A. Sepulchre. Transmitting a signal by amplitude modulation in a chaotic network. *Chaos*, 16(013104), 2006.

[20] B. Cessac and J.A. Sepulchre. Linear response in a class of simple systems far from equilibrium. *Physica D*, 225(1):13–28, 2007.

[21] B. Cessac. Absolute Stability criteria for random asymmetric neural networks. *J. of Physics A*, 27:L927-L930, 1994.

[22] B. Cessac. Statistics of spike trains in conductance-based neural networks: Rigorous results. *The Journal of Mathematical Neuroscience*, 1(8):1–42, 2011.

[23] B. Cessac and R. Cofré. Spike train statistics and Gibbs distributions. *Journal of Physiology-Paris*, 107(5):360–368, November 2013. Special issue: Neural Coding and Natural Image Statistics.

[24] B. Cessac and R. Cofré. Linear response for spiking neuronal networks with unbounded memory. submitted, arxiv preprint 1905.13424 , October 2019.

[25] B. Cessac, B. Doyon, M. Quoy, and M. Samuelides. Mean-field equations, bifurcation map, and route to chaos in discrete time neural networks. *Physica 74 D*, pages 24–44, 1994.

[26] E. J. Chichilnisky. A simple white noise analysis of neuronal light responses. *Network: Comput. Neural Syst.*, 12:199–213, 2001.

[27] S. Cocco, S. Leibler, and R. Monasson. Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *PNAS*, 106(33):14058–14062, 2009.

[28] R. Cofré and B. Cessac. Dynamics and spike trains statistics in conductance-based integrate-and-fire neural networks with chemical and electric synapses. *Chaos, Solitons & Fractals*, 50(13):3, 2013.

[29] R. Cofré and B. Cessac. Exact computation of the maximum-entropy potential of spiking neural-network models. *Phys. Rev. E*, 89(052117), 2014.

[30] P. Dayan and L.F. Abbott. *Theoretical Neuroscience : Computational and Mathematical Modeling of Neural Systems*. MIT Press, 2001.

[31] S.R. de Groot and P. Mazur. *Non-equilibrium Thermodynamics*. Dover Books on Physics. Dover Publications, 1984.

[32] A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38. Springer, 1998.

[33] A. Destexhe, Z. F. Mainen, and T.J. Sejnowski. Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of Computational Neuroscience*, 1(3):195—230, 1994.

[34] B. Doyon, B. Cessac, M. Quoy, and M. Samuelides. Control of the transition to chaos in neural networks with random connectivity. *International Journal of Bifurcation and Chaos*, 03(02):279–291, 1993.

[35] B. Ermentrout. Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics*, 61:353–430, 1998.

[36] B. Ermentrout and D. Terman. *Foundations of Mathematical Neuroscience*. Interdisciplinary Applied Mathematics. Springer, 2010.

[37] O. Faugeras and J. MacLaurin. Asymptotic description of neural networks with correlated synaptic weights. *Entropy*, 17(7):4701, 2015.

[38] O. Faugeras, J. Touboul, and B. Cessac. A constructive mean field analysis of multi population neural networks with random synaptic weights and stochastic inputs. *Frontiers in Computational Neuroscience*, 3(1), 2009.

[39] R. Fernandez and S. Gallo. Regular g-measures are not always gibbsian. *Electron. Commun. Probab.*, 16:732–740, 2011.

[40] R. Fernandez and G. Maillard. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.*, 118(3-4):555–588, 2005.

[41] U. Ferrari, S. Deny, M. Chalk, G. Tkačik, O. Marre, and T. Mora. Separating intrinsic interactions from extrinsic correlations in a network of sensory neurons. *Physical Review E*, 98(4):042410, 2018.

[42] R. FitzHugh. Mathematical models of threshold phenomena in the nerve membrane. *Bulletin of Mathematical Biology*, 17(4):257–278, 1955.

[43] W J. Freeman, Yong Yao, and Brian Burke. Central pattern generating and recognizing in olfactory bulb: A correlation learning rule. *Neural Networks*, 1(4):277 – 288, 1988.

[44] K. Friston, J. Kilner, and L. Harrison. A free energy principle for the brain. *Journal of Physiology-Paris*, 100:70–87, 2006.

[45] K. Friston. The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, 13(7):293–301, 2009. 00822.

[46] K. Friston. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2):127–138, 2010. 02703.

[47] K. Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011.

[48] G. Gallavotti and E. G. D. Cohen. Dynamical ensembles in stationary states. *Journal of Statistical Physics*, 80(5):931–970, 1995.

[49] G Gallavotti and E.G.D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, 74:2694–2697, 1995.

[50] G. Gallavotti. Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem. *Journal of Statistical Physics*, 84(5):899–925, 1996.

[51] G. Gallavotti. *Nonequilibrium and Irreversibility*. Springer Publishing Company, Incorporated, 2014.

[52] J.M. Gambaudo and C. Tresser. *Le chaos, théorie et expériences*. Collection CEA, 1988.

[53] E. Ganmor, R. Segev, and E. Schneidman. The architecture of functional interaction networks in the retina. *The journal of neuroscience*, 31(8):3044–3054, 2011.

[54] E. Ganmor, R. Segev, and E. Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *PNAS*, 108(23):9679–9684, 2011.

[55] F.R. Gantmacher. *The theory of matrices*. Chelsea Publishing Company, 1959. Two volumes.

[56] C. Gardella, O. Marre, and T. Mora. Modeling the correlated activity of neural populations: A review. *Neural computation*, 31(2):233–269, 2019.

[57] P. Gaspard. *Chaos, Scattering and Statistical Mechanics*. Cambridge Nonlinear Science Series. Cambridge University Press, 2005.

[58] W. Gerstner and W. M. Kistler. Spiking Neuron Models. Cambridge University Press, 2002.

[59] V.L. Girko. Circular law. *Theor. Prob. Appl*, 29:694–706, 1984.

[60] V.L. Girko. *Theory of Random Determinants*. Mathematics and its Applications. Springer Netherlands, 2012.

[61] T. Gollisch and M. Meister. Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, 65(2):150–164, January 2010.

[62] M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. ii. irreversible processes in fluids. *The Jour-*

nal of Chemical Physics, 22(3):398–413, 1954.

[63] J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. *unpublished*, 1971.

[64] A.L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, 117:500–544, 1952.

[65] D.H. Hubel and T.N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat visual cortex. *J Physiol*, 160:106–154, 1962.

[66] E. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572, 2003.

[67] E. Izhikevich. Neural excitability, spiking, and bursting. *International Journal of Bifurcation and Chaos*, 10:1171–1266, 2000.

[68] E.T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620, 1957.

[69] Y. Jiang and D. Ruelle. Analyticity of the susceptibility function for unimodal markovian maps of the interval. *Nonlinearity*, 18, 2005.

[70] M. Kaiser, R.L. Jack, and J. Zimmer. Canonical structure and orthogonality of forces and currents in irreversible markov chains. *Journal of Statistical Physics*, 170(6):1019–1050, 2018.

[71] J. L. Kaplan and J. A. Yorke. Chaotic behavior of multidimensional difference equations. In Heinz-Otto Peitgen and Hans-Otto Walther, editors, *Functional Differential Equations and Approximation of Fixed Points*, pages 204–227, Berlin, Heidelberg, 1979. Springer Berlin Heidelberg.

[72] A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*. Kluwer, 1998.

[73] J.S. Kirkaldy. Thermodynamics of the human brain. *Biophysical Journal*, 5(6):981 – 986, 1965.

[74] H. Korn and P. Faure. Is there chaos in the brain? ii. experimental evidence and related models. *C.R. Biologies*, 326:787–840, 2003.

[75] R. Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[76] L. Lapicque. Recherches quantitatifs sur l'excitation des nerfs traitée comme une polarisation. *J. Physiol. Paris*, 9:620–635, 1907.

[77] M. Le Bellac, F. Mortessagne, and G. G. Batrouni. *Equilibrium and Non-Equilibrium Statistical Thermodynamics*. Cambridge University Press, 2004.

[78] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141, 1963.

[79] R.S. Mackay and C. Tresser. Transition to topological chaos for circle maps. *Physica D*, 19(2):206–273, 1986.

[80] C. Maes and M. H. van Wieren. Time-symmetric fluctuations in nonequilibrium systems. *Physical review letters*, 96:240601, 2006.

[81] G. Maillard. *Introduction to chains with complete connections*. Ecole Federale Polytechnique de Lausanne, winter 2007.

[82] K. Matsuoka. Stability conditions for nonlinear continuous neural networks with asymmetric connection weights. *Neural Networks*, 5:495–500, 1992.

[83] C Morris and H Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophys J*, 35(1):193–213, 1981.

[84] S. P. Muscinelli, W. Gerstner, and T. Schwalger. How single neuron properties shape chaotic dynamics and signal transmission in random neural networks. *arXiv e-prints*, December 2018.

[85] J. Nagumo, S. Arimoto, and S. Yoshizawa. An active pulse transmission line simulating nerve axon. *Proceedings of the IRE*, 50:2061–2070, 1962.

[86] J. Naudé, B. Cessac, H. Berry, and B. Delord. Effects of Cellular Homeostatic Intrinsic Plasticity on Dynamical and Computational Properties of Biological Recurrent Neural Networks. *Journal of Neuroscience*, 2013.

[87] S. Newhouse, D. Ruelle, and F. Takens. Occurrence of strange axiom a attractors near quasiperiodic flows on $t^m, m \geq 3$. *Comm. Math. Phys.*, 64(1):35–40, 1978.

[88] MEA records actually a voltage activity on recording electrodes. Then, a spike sorting algorithm is used to attribute recorded voltage traces to neurons, and then, construct the spike train. Therefore, MEA could in principle allow to obtain the probability of having a spike pattern given the voltage history.

[89] A. Le Ny. Introduction to (generalized) gibbs measures. *Ensaios Matematicos*, 15:1–126, 2008.

[90] O. Onicescu and G. Mihoc. Sur les chaînes statistiques. *C. R. Acad. Sci. Paris*, 200:511–512, 1935.

[91] L. Onsager. Reciprocal relations in irreversible processes. i. *Phys. Rev.*, 37:405–426, 1931.

[92] S. Ostojic and N. Brunel. From spiking neuron models to linear-nonlinear models. *PLOS Computational Biology*, 7(1):1–16, 2011.

[93] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque*, 187(188):1–268, 1990.

[94] Y. B. Pesin. Characteristic lyapunov exponents and smooth ergodic theory. *Uspekhi Matematicheskikh Nauk*, 32(4):55–112, 1977.

[95] J.W. Pillow, L. Paninski, V.J. Uzzell, E.P. Simoncelli, and E.J. Chichilnisky. Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience*, 25(47):11003–11013, 2005.

[96] M. Pollicott. On the rate of mixing of axiom a flows. *Inventiones mathematicae*, 81(3):413–426, 1985.

[97] H. Preissl, W. Lutzenberger, and F. Pulvermüller. Is there chaos in the brain? *Behavioral and Brain Sciences*, 19(2):307–308, 1996.

[98] P. Reynaud-Bouret, V. Rivoirard, F. Grammont, and C. Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. *Journal of Mathematical Neuroscience*, page 4:3, April 2014.

[99] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Inference of functional connectivity in Neurosciences via Hawkes processes. In *1st IEEE Global Conference on Signal and Information Processing*, Austin, United States, December 2013.

[100] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the Neural Code*. Bradford Books, 1997.

[101] M. Rudolph and A. Destexhe. Analytical integrate and fire neuron models with conductance-based dynamics for event driven simulation strategies. *Neural Computation*, 18:2146–2210, 2006.

[102] M. Rudolph and A. Destexhe. How much can we trust neural simulation strategies? *Neurocomputing*. 70(10–12):1966–1969, 2007

[103] D. Ruelle. Generalized zeta-functions for axiom A basic sets. *Bull. Amer. Math. Soc*, pages 153–156, 1976.

[104] D. Ruelle. *Thermodynamic formalism*. Addison-Wesley,Reading, Massachusetts, 1978.

[105] D. Ruelle. Resonances for axiom A flows. *J. Differential Geom.*, 25(1):99–116, 1987.

[106] D. Ruelle. Smooth dynamics and new theoretical ideas in nonequilibrium statistical mechanics. *J. Statist. Phys.*, 95:393–468, 1999.

[107] D. Ruelle. A measure associated with axiom A attractors. *American Journal of Mathematics*, 98(3):619–654, 1976.

[108] D. Ruelle. Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms. *Nonlinearity*, 11(1):5–18, 1998.

[109] D. Ruelle. Differentiating the absolutely continuous invariant measure of an interval map f with respect to f. *Communications in Mathematical Physics*, 258:445–453, 2005. 10 pages, plain TeX.

[110] D. Ruelle and F. Takens. On the nature of turbulence. *Comm. Math. Phys.*, 20:167–192, 1971.

[111] E. Schneidman, M.J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[112] J. Schuecker, S. Goedeke, D. Dahmen, and M. Helias. Functional methods for disordered neural networks. *arXiv e-prints*, May 2016.

[113] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, 2006.

[114] J. Shlens, G.D. Field, J.L. Gauthier, M.I. Grivich, D. Petrusca, A. Sher, A.M. Litke, and E.J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, 26(32):8254, 2006.

[115] J. Shlens, G. D. Field, J. L. Gauthier, M. Greschner, A. Sher, A. M. . Litke, and E. J. Chichilnisky. The structure of large-scale synchronized firing in primate retina. *The Journal of Neuroscience*, 29(15):5022–5031, April 2009.

[116] E P Simoncelli, L Paninski, J Pillow, and O Schwartz. Characterization of neural responses with stochastic stimuli. In M Gazzaniga, editor, *The Cognitive Neurosciences, III*, chapter 23, pages 327–338. MIT Press, 2004.

[117] Ya Sinai. Gibbs measures in ergodic theory. *Russ. Math. Surveys*, 27(4):21–69, 1972.

[118] B. Siri, H. Berry, B. Cessac, B. Delord, and M. Quoy. Effects of hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. *Journal of Physiology-Paris*, 2007.

[119] B. Siri, H. Berry, B. Cessac, B. Delord, and M. Quoy. A mathemati-

cal analysis of the effects of hebbian learning rules on the dynamics and structure of discrete-time random recurrent neural networks. *Neural Computation*, 20(12):12, December 2008.

[120] H. Sompolinsky, A. Crisanti, and HJ Sommers. Chaos in Random Neural Networks. *Physical Review Letters*, 61(3):259–262, 1988.

[121] H. Sompolinsky and A. Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47(5):359–362, 1981.

[122] H. Sompolinsky and A. Zippelius. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25(11):6860–6875, 1982.

[123] S. Souihel and B. Cessac. Anticipation in the retina and the primary visual cortex : towards an integrated retino-cortical model for motion processing. In *ICMNS 2019 - The 5th International Conference on Mathematical NeuroScience*, Copenhague, Denmark, June 2019.

[124] G. Tkacik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, Michael J. Berry, and William Bialek. Thermodynamics and signatures of criticality in a network of neurons. *Proceedings of the National Academy of Sciences*, 112(37):11508–11513, 2015.

[125] G. Tkačik, Elad Schneidman, Michael J. Berry II, and William Bialek. Spin glass models for a network of real neurons. *arXiv preprint arXiv:0912.5409*, 2009.

[126] T. K. Tran. Large scale retinal modeling for the design of new generation retinal prostheses, PhD thesis, U. Nottingham, 2015.

[127] NG van Kampen. A discussion on linear response theory. *Physica Norvegica*, 5:279, 1971.

[128] C. VanVreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10:1321–1372, 1998.

[129] J.C Vasquez, O. Marre, A. Palacios, M.J Berry, and B. Cessac. Gibbs distribution analysis of temporal correlation structure on multicell spike trains from retina ganglion cells. *J. Physiol. Paris*, 106(3-4):120–127, May 2012.

[130] J. C Vasquez, B. Cessac, H. Rostro-Gonzalez, and T. Viéville. How Gibbs Distributions May Naturally Arise from Synaptic Adaptation Mechanisms. A Model-Based Argumentation Journal of Statistical Physics, 136(3), 565–602, 2009.

[131] H.R. Wilson and J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, 12:1–24, 1972.

[132] H.R. Wilson and J.D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biological Cybernetics*, 13(2):55–80, September 1973.

[133] C. N. Yang and T. D. Lee. Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation. *Physical Review*, 87:404–409, August 1952.