Parametric estimation of spike train statistics by Gibbs distributions : an application to bio-inspired and experimental data

Bruno Cessac, Juan Carlos Vasquez, Hassan Nasser, Horacio Rostro-Gonzalez, Thierry Viéville, Adrian Palacios

PARAMETRIC ESTIMATION OF SPIKE TRAIN STATISTICS BY GIBBS DISTRIBUTIONS : AN APPLICATION TO BIO-INSPIRED AND EXPERIMETAL DATA

Bruno Cessac,* Juan-Carlos Vasquez,† Hassan Nasser,‡ Horacio Rostro-González,§ Thierry Viéville¶ Adrian Palacios ‖

ABSTRACT We review here the basics of the formalism of Gibbs distributions and its numerical implementation, (its details published elsewhere [1], in order to characterizing the statistics of multi-unit spike trains. We present this here with the aim to analyze and modeling synthetic data, especially bio-inspired simulated data e.g. from Virtual Retina [2], but also experimental data Multi-Electrode-Array(MEA) recordings from retina obtained by Adrian Palacios. We remark that Gibbs distribution allow us to estimate the spike statistics, given a design choice, but also to compare different models, thus answering comparative questions about the neural code.

KEY WORDS Spike trains statistics, Gibbs measures, Numerical implementations, MEA recordings.

1 Context.

Processing and encoding of information in neural dynamics is a very active research field [3], although still much of the role of neural assemblies and their internal interactions remains unknown [4]. The simultaneously recording of the activity of groups of neurons (up to several hundreds) over a dense configuration, supply today a critical database to unravel the role of specific neural assemblies.

A recent popular approach has been applied by [5]. They have shown that a model taking into account pairwise synchronizations between neurons in a small assembly (10-40 retinal ganglion cells) describes most (90%) of the correlation structure and of the mutual information of the block activity, and performs much better than a non homogeneous Poissonian model. Analogous results were presented the same year by [6]. The model used by both teams is based on a probability distribution known as the Gibbs distribution of the Ising model and comes from statistical physics. We develop here a numerical method allowing to parametrically estimate and compare, models for the statistics of simulated multi-cell-spike trains. Our method is not limited to firing rates models, pairwise synchronizations as [5, 6, 7] or 1-step time pairwise correlations models as [8], but deals with general form of Gibbs distributions, with parametric potentials corresponding to a spike n-uplets expansion with multi-units and multi-times terms. The method is exact (in the sense that is does not involve heuristic minimization techniques) since it is based on spectral properties of a transition matrix associated with a Markov chain.

*Laboratoire JAD http://math.unice.fr& INRIA NeuroMathComp http://www-sop.inria.fr/neuromathcomp/

†INRIA NeuroMathComp

‡INRIA NeuroMathComp

§INRIA NeuroMathComp

¶INRIA Cortex http://cortex.loria.fr/

‖INRIA Cortex , UHP et Nancy 2 et CREA Ecole Polytechnique

2 Theory

Spike trains and Raster Plots

Lets denote by $i = 1 \dots N$ the neuron index, and assume we have discretized the time in bins of size Δt so t denotes the bin number, then to each membrane potential value corresponding to each bin $V_i(t)$ we associate a binary variable $\omega_i(t) = Z(V_i(t))$, where $Z(x) = 1$ whenever $x \geq \theta$ and $Z(x) = 0$ otherwise and θ is the firing threshold. The "spiking pattern" of the neural network at time t is the vector $\omega(t) = (\omega_i(t))_{i=1}^{N}$: it tells us which neurons are firing at time t, ($\omega_i(t) = 1$) and which neurons are not firing at time t ($\omega_i(t) = 0$). We denote by ω_s^t the sequence or *spike block* $\omega(s) \dots \omega(t)$. A bi-infinite sequence $\omega = \{\omega(t)\}_{t=-\infty}^{+\infty}$ of spiking patterns is called a "raster plot". In practice, raster plots are obviously finite sequences of spiking patterns In the next sections we give our main results whose details can be found in [1].

Empirical averaging

As a starting point, one computes experimental averages allowing us to estimate the average value of some prescribed *observable* $\phi(\omega)$, i.e. a function which associates to a raster plot a real number. Typical examples of observables are $\phi(\omega) = \omega_i(0)$ in which case the average of ϕ is the firing rate of neuron i; $\phi(\omega) = \omega_i(0)\omega_j(0)$ then the average of ϕ measures the probability of spike coincidence for neuron j and i; $\phi(\omega) = \omega_i(\tau)\omega_j(0)$ then the average of ϕ measures the probability of the event "neuron j fires and neuron i fires τ time step later" (or sooner according to $sign(\tau)$).

The average of ϕ is computed, from the experimental raster plots. Let us assume that it is sufficient to focus on time average for a single raster plot, via the time-empirical average:

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

Note that in (1) we have used the shift σ^t for the time evolution of the raster plot.

Inferring statistics with fixed observables
We are seeking a probability distribution ν which is compatible with these empirical average i.e. for all $l = 1 \dots L$,

$$\nu(\phi_l) = \pi_\omega^{(T)}(\phi_l) = C_l. \tag{2}$$

This is a minimal, but insufficient requirement, since one can construct infinitely many probability distributions satisfying these constraints. One has therefore to add more structure. We propose two additional requirements: Stationarity and using Gibbs measures as parametric statistical models.

Stationarity The first one corresponds to assuming that data are extracted from a stationary dynamics i.e. that statistics are invariant under time translation, during the observation window.On mathematical grounds, this corresponds to requiring that ν is invariant. Indeed, by definition, an invariant measure ν is ergodic if for any typical raster plot (i.e. ν-almost every raster plot) ω, and for any (measurable) observable ϕ, the empirical average $\pi_\omega^{(T)}(\phi)$ converges, as $T \to +\infty$ to the theoretical expectation of ϕ with respect to ν. Namely,

$$\lim_{T \to +\infty} \pi_\omega^{(T)}(\phi) = \nu(\phi),$$

for ν-almost every ω. This means that the empirical average better and better approximates the expected value as $T \to +\infty$

Gibbs distributions The second requirement states that one only wants to take into account the constraints (2) without adding additional assumptions. Statistical physics naturally proposes a canonical way to construct such a statistical model: "Maximizing the statistical entropy under the constraints (2)" [9]. In the present context, this amounts to solving a *variational principle*:

$$P[\boldsymbol{\psi}] = \sup_{\nu \in m^{(inv)}} h[\nu] + \nu[\boldsymbol{\psi}], \tag{3}$$

where $m^{(inv)}$ is the set of invariant measures and h the Kolmogorov-Sinai entropy or entropy rate. The term $\boldsymbol{\psi}$ is given by:

$$\boldsymbol{\psi} = \sum_{l=1}^{L} \lambda_l \phi_l, \tag{4}$$

where the λ_l's are adjustable Lagrange multipliers. It is called a "potential. Typically, in the case considered here, the observables ϕ_l and therefore the potential $\boldsymbol{\psi}$ depends on the raster plot over a finite time horizon R, i.e. $\boldsymbol{\psi}(\omega) = \boldsymbol{\psi}(\omega(0), \dots, \omega(R-1))$. One speaks of "range-$R$ potentials". They constitute Markovian approximations of more general (infinite range) potentials occurring e.g. in IF models with noise, where the degree of accuracy of the approximation can be mathematically controlled : the accuracy increases exponentially fast as R growths (see [10] for details).

A probability $\nu_{\boldsymbol{\psi}}$ which realizes the supremum (3), i.e.

$$P[\boldsymbol{\psi}] = h[\nu_{\boldsymbol{\psi}}] + \nu_{\boldsymbol{\psi}}[\boldsymbol{\psi}], \tag{5}$$

is called, in this context, an "equilibrium state". Typically, equilibrium states are also *Gibbs states* or *Gibbs distributions*. The Gibbs distribution property means that one can find some constants $P(\psi), c_1, c_2$ with $0 < c_1 \leq 1 \leq c_2$ such that for all $n \geq 0$ and for all $\omega \in X$:

$$c_1 \leq \frac{\mu_\psi\left(\omega_{-n}^0\right)}{\exp\left[-(n+1)P(\psi) + \sum_{k=-n}^{0} \psi(T^k\omega)\right]} \leq c_2.$$

where T is the *right* shift over the set of infinite sequences X i.e. $(T\omega)(t) = \omega(t-1)$, $t \leq 0$. This means that the probability of (large) sequences behaves like $\mu_\psi\left(\omega_{-n}^0\right) \sim \frac{\sum_{k=-n}^{0} \psi(T^k\omega)}{\exp{(n+1)P(\psi)}}$. The term $P(\psi)$, called the *topological pressure* in this context is the formal analog of a thermodynamic potential (free energy). It is a generating function for the cumulants of ψ.

3 Numerical Estimation of Gibbs Distributions

$P(\psi)$ and μ_ψ can be found by a spectral approach, being respectively the (unique) largest eigenvalue and related left eigenfunction of the Ruelle-Perron-Frobenius operator $\mathcal{L}_\psi f(\omega) = \sum_{\omega' : \sigma\omega' = \omega} \psi(\omega') f(\omega')$, acting on $C(X, \mathbb{R})$, the set of continuous real functions on X.

Ruelle-Perron-Frobenius operator
The Ruelle-Perron-Frobenius (RPF) operator was introduced by Ruelle in [11]. In the present case of range-R potentials it acts on vectors $\mathbf{v}$ in $\mathcal{H}$ as:

$$(L_{\boldsymbol{\psi}}\mathbf{v})_w = \sum_{w' \in 2^{NR}} e^{\boldsymbol{\psi}_{w'}} v_{w'} G_{w',w}, \tag{6}$$

where $w = \sigma w'$, σ denoting again the time evolution. Thus, here $L_{\boldsymbol{\psi}}$ is a positive $2^{NR} \times 2^{NR}$ *matrix* (while it acts on functional spaces in the infinite range case). We note $L_{\boldsymbol{\psi}}(w', w)$ the entries of $L_{\boldsymbol{\psi}}$.

The matrix G is called the *grammar*. It encodes the *essential fact* that the underlying dynamics is not able to produce all possible raster plots in its definition:

$$G_{w',w} = \begin{cases} 1, & \text{if the transition } w' \to w \text{ is admissible;} \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Note therefore that $L_{\boldsymbol{\psi}}$ is *sparse*.

In the present paper we make the assumption that the underlying (and hidden) dynamics is such that the $L_{\boldsymbol{\psi}}$ matrix is primitive, i.e. $\exists n > 0$, s.t. $\forall w, w'$ $L_{\boldsymbol{\psi}}^n(w', w) > 0$. This assumption holds for Integrate and Fire models with noise and is likely to hold for more general neural networks models where noise renders dynamics ergodic and mixing [10].

Then $L_{\boldsymbol{\psi}}$ obeys the Perron-Frobenius theorem[1]:

Theorem 1 *$L_{\boldsymbol{\psi}}$ has a unique maximal eigenvalue $s_{\boldsymbol{\psi}} = e^{P[\boldsymbol{\psi}]}$ associated with a right eigenvector $\boldsymbol{b}_{\boldsymbol{\psi}\rangle}$ and a left eigenvector $\boldsymbol{b}_{\langle\boldsymbol{\psi}}$ such that $L_{\boldsymbol{\psi}}\boldsymbol{b}_{\boldsymbol{\psi}\rangle} = s_{\boldsymbol{\psi}}\boldsymbol{b}_{\boldsymbol{\psi}\rangle}$, and $\boldsymbol{b}_{\langle\boldsymbol{\psi}}L_{\boldsymbol{\psi}} = s_{\boldsymbol{\psi}}\boldsymbol{b}_{\langle\boldsymbol{\psi}}$. Those vectors can be chosen such that $\boldsymbol{b}_{\langle\boldsymbol{\psi}}.\boldsymbol{b}_{\boldsymbol{\psi}\rangle} = 1$ where . is the scalar product in $\mathcal{H}$. The remaining part of the spectrum is located in a disk in the complex plane, of radius strictly lower than $s_{\boldsymbol{\psi}}$. Moreover, for all $\mathbf{v}$ in $\mathcal{H}$,*

$$\frac{1}{s_{\boldsymbol{\psi}}^n} L_{\boldsymbol{\psi}}^n \mathbf{v} \to \boldsymbol{b}_{\boldsymbol{\psi}\rangle}\boldsymbol{b}_{\langle\boldsymbol{\psi}} . \mathbf{v}, \tag{8}$$

The Gibbs distribution is

$$\nu_{\boldsymbol{\psi}} = \boldsymbol{b}_{\boldsymbol{\psi}\rangle}\boldsymbol{b}_{\langle\boldsymbol{\psi}}, \tag{9}$$

i.e. the probability of a spin block w of length R is

$$\nu_{\boldsymbol{\psi}}(w) = b_{\boldsymbol{\psi}\rangle,w} b_{\langle\boldsymbol{\psi},w},$$

where $b_{\boldsymbol{\psi}\rangle,w}$ is the w-th component of $\boldsymbol{b}_{\boldsymbol{\psi}\rangle}$.

Computing averages

Since $\nu_{\boldsymbol{\psi}}[\phi_l] = \sum_w \nu_{\boldsymbol{\psi}}[w]\phi_l(w)$ one obtains using (9):

$$\nu_{\boldsymbol{\psi}}[\phi_l] = \sum_{w \in \mathcal{H}} b_{\boldsymbol{\psi}\rangle,w}\phi_l(w) b_{\langle\boldsymbol{\psi},w}. \tag{10}$$

This provides a fast way to compute $\nu_{\boldsymbol{\psi}}[\phi_l]$ and will be used to tune the λ_l's so that the theoretical averages $\nu_{\boldsymbol{\psi}}[\phi_l]$ matches the empirical average C_l according to eq. (16).

Entropy

An exact expression for the Kolmogorov-Sinai entropy can be readily obtained from eq. (5) giving:

$$h[\nu_{\boldsymbol{\psi}}] = \log(s_{\boldsymbol{\psi}}) - \sum_l \lambda_l^* \nu_{\boldsymbol{\psi}}[\phi_l]. \tag{11}$$

Comparing several Gibbs statistical models

The choice of a potential (4), i.e. the choice of a set of observables, fixes a statistical model for the statistics of spike trains. Clearly, there are many choices of potentials and one needs to propose a criterion to compare them.

The Kullback-Leibler divergence,

$$d(\mu, \nu) = \limsup_{n \to \infty} \frac{1}{n} \sum_{[\omega]_0^{n-1}} \mu\left([\omega]_0^{n-1}\right) \log\left[\frac{\mu\left([\omega]_0^{n-1}\right)}{\nu\left([\omega]_0^{n-1}\right)}\right], \tag{12}$$

where μ and ν are two invariant probability measures, provides some notion of asymmetric "distance" between μ and ν. Minimizing this divergence, i.e. the conditional entropy $h[\mu|\nu]$, corresponds to minimizing "what is not explained in the empirical measure μ by the theoretical measure ν".

The computation of $d(\mu, \nu)$ is numerically delicate but, in the present context, the following holds. For μ an invariant measure and $\nu_{\boldsymbol{\psi}}$ a Gibbs measure with a potential $\boldsymbol{\psi}$, both defined on the same set of sequences Σ, one has [14, 15]:

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

This suggests us[2] to use this relation to compare different statistical models (corresponding to different potentials) by choosing the one which minimizes the quantity:

$$d(\pi_\omega^{(T)}, \nu_{\boldsymbol{\psi}}) = P[\boldsymbol{\psi}] - \pi_\omega^{(T)}(\boldsymbol{\psi}) - h(\pi_\omega^{(T)}). \tag{13}$$

The advantage is that this quantity can be numerically estimated, since the topological pressure is known from the Ruelle-Perron-Frobenius theorem, while $\pi_\omega^{(T)}(\boldsymbol{\psi})$ is directly computable. Since $\pi_\omega^{(T)}$ is fixed by the experimental raster plot, $h(\pi_\omega^{(T)})$ is fixed independently of the statistical model. Thus, comparing two statistical models corresponding to 2 potentials $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2$ amounts to comparing $P[\boldsymbol{\psi}_1] - \pi_\omega^{(T)}(\boldsymbol{\psi}_1)$ and $P[\boldsymbol{\psi}_2] - \pi_\omega^{(T)}(\boldsymbol{\psi}_2)$.

Introducing

$$\tilde{h}[\boldsymbol{\psi}] = P[\boldsymbol{\psi}] - \pi_\omega^{(T)}(\boldsymbol{\psi}) = P[\boldsymbol{\psi}] - \sum_l \lambda_l \pi_\omega^{(T)}(\phi_l), \tag{14}$$

the comparison of two statistical models, i.e. knowing if $\boldsymbol{\psi}_2$ is significantly "better" that $\boldsymbol{\psi}_1$, reduces to verify:

$$\tilde{h}[\boldsymbol{\psi}_2] \ll \tilde{h}[\boldsymbol{\psi}_1], \tag{15}$$

easily computable at the implementation level, as developed below. Note that $\tilde{h}$ has the dimension of entropy. Since we compare entropies, which units are bits of information, defined in base 2, the previous comparison units is well-defined.

Estimating the potential parameters

The final step of the estimation procedure is to find the parameters $\boldsymbol{\lambda}$ such that the Gibbs measure fits at best with the empirical measure. We have discussed why minimizing (13) seems the best choice in this context. Since $h(\pi_\omega^{(T)})$ is a constant with respect to λ, it is equivalent to minimize $\tilde{h}[\boldsymbol{\psi}_\lambda]$ eq. (14) while

$$\frac{\partial \tilde{h}}{\partial \lambda_l} = \frac{\partial P[\boldsymbol{\psi}_\lambda]}{\partial \lambda_l} - \pi_\omega^{(T)}(\phi_l) = \nu_{\boldsymbol{\psi}}(\phi_l) - \pi_\omega^{(T)}(\phi_l), \tag{16}$$

where $\nu_{\boldsymbol{\psi}}(\phi_l)$ is given by (10). Equivalently, we are looking for a Gibbs distribution $\nu_{\boldsymbol{\psi}}$ such that $\frac{\partial P[\boldsymbol{\psi}_\lambda]}{\partial \lambda_l} = \pi_\omega^{(T)}(\phi_l)$ which expresses that $\pi_\omega^{(T)}$ is tangent to P at $\boldsymbol{\psi}_\lambda$ [14].

4 Applications

This analysis opens up the possibility of developing efficients algorithms to estimate at best the statistic of spike

[1] This theorem has been generalized by Ruelle to infinite range potentials under some regularity conditions [12, 13].

[2] Although $\pi_\omega^{(T)}$ is not invariant. This approximation becomes exact as $T \to +\infty$.

trains from experimental data, using several guess potential and selecting the one which minimizes the KL divergence [1]. The idea is to start from a parametric form of potential (Ψ), of range R, and to compute the empirical average of all monomials ϕ_l from data (say, an experimental raster $\omega^{(exp)}$). Then, one adjust the parameters λ_l by minimizing the KL divergence between the Gibbs measure $\mu_{\psi_{\boldsymbol{\lambda}}^{(R)}}$ and the empirical measure attached to the experimental raster $\omega^{(exp)}$.

The complete computational strategy and the description of the implemented numerical algorithms is described in [1] and is freely available as a C++ source at http://enas.gforge.inria.fr.

Preliminary data analysis
As a first step, we observe the outlook of the correlation in these different data types. To do so, we estimate the correlation between a couple of neurons for several various raster length.
The correlation between two neuron is then:

$$C_{i,j} = |\mu(\omega_i(t)\omega_j(t)) - \mu(\omega_i(t))\mu(\omega_j(t))| \qquad (17)$$

Thus, $C_{i,j}$ is equal to how many time the two neurons i and j fired together minus the product of their firing rate.

The computed correlation between non-correlated data e.g between two neurons whose raster plots follow a Bernoulli distribution decreases to zero exponentially with a decay constant $\sim 1/\sqrt{T}$ as described by large deviations theory. Figure 1 using data obtained from virtual retina software ([2]) shows the same behavior as in randomly generated data (Bernoulli distribution, figure not shown) as independently of distance between cells.

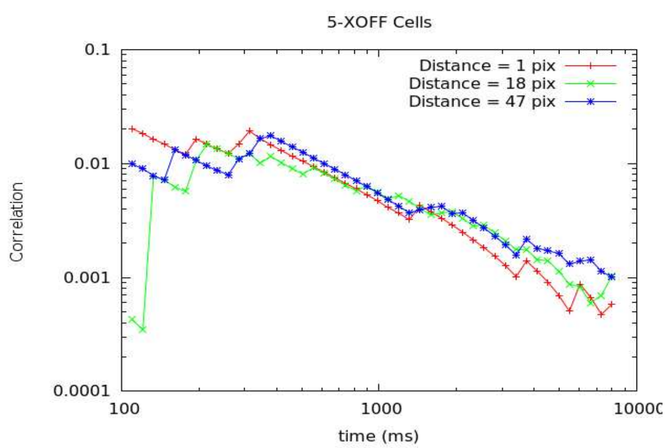

Figure 1. Correlation of spiking activity vs the sample size, computed from data obtained from Virtual Retina ([2]) for several couples of neurons situated at different distances. The distance is measured in pixel and the virtual retina is configured such with 1 neuron/pixel.

We perform the same analysis for real data acquired with and MEA chip (60 channels) on retinal ganglion cell with no stimuli (spontaneous regime). Figure 2 shows results for correlation between several couples of neurons, at different distances proving the presence of effective (i.e. non-zero asymptotically) correlations

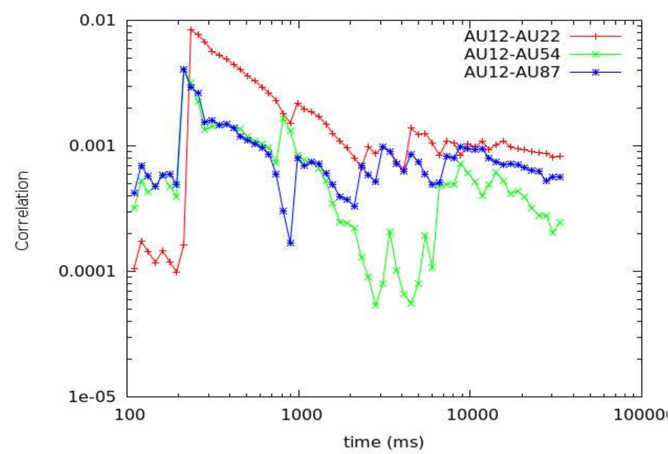

Figure 2. Correlation between ganglion cells in real MEA measurements. The figure shows the correlation for three couples of neurons, at several distances. The time scale corresponds to the first 30 sec. of the measured neurons .

Detailed Results for data
We expect to provide detailed results in time for the conference. In particular, MEA recordings obtained by Adrien Palacios are currently analyzed within this approach.

5 Discussion

In this paper we have addressed the question of characterizing the spike train statistics of a group of neurons in the stationary case, thanks to the framework of thermodynamic formalism in ergodic theory (see [16] for the related theory). We have shown that the Jaynes method, based on an a priori choice of a "guess" potential, with finite range, amounts to approximate the exact probability distribution by the Gibbs distribution of a Markov chain. The degree of approximation can be controlled by the Kullback-Leibler divergence. We will show detailed results on different type of data, in time for the conference

With respect to the state of the art, this method allows us to consider either more neurons with non-trivial statistics (e.g. beyond rate models and even models with correlation), thus targeting models with complex spike pattern. This method is in a sense the next step after Ising models, known as being able to represent a large but limited part of the encoded information (e.g. [5, 17]). Another very important difference with respect to other current methods is that we perform the explicit variational optimization of a well defined quantity, i.e., the KL-divergence between the observed and estimated distributions. The method proposed here does not rely on Monte Carlo Markov Chain methods but on a spectral computation based on the RPF operator, providing exact formula, while the spectral characteristics are easily obtained from standard numerical methods.

A step further, the non-trivial but very precious virtue of the method is that it allows us to efficiently compare models. We thus not only estimate the optimal parameters of a model, but can also determine among a set of models which model is the most pertinent. This means for instance that we can determine if either only rates, or rates and correlations matters, for a given piece of data. Another example is to detect if a given spike pattern is significant, with respect to a model not taking this pattern into account. These elements push the state of the art regarding statistical analysis of spike train a step further.

One weakness of the present work is that it only considers stationary dynamics, where e.g. the external current I_i is independent of time. However, real neural systems are submitted to non static stimuli, and transients play a crucial role, but considering non stationary requires to handle time dependent Gibbs measures. In the realm of ergodic theory applied to non equilibrium statistical physics, Ruelle has introduced the notion of time-dependent SRB measure [18]. A similar approach could be used here, at least formally.

Acknowledgment:.

References

[1] J.C. Vasquez, B. Cessac, and T. Vieville. Entropy-based parametric estimation of spike train statistics. *Journal of Computational Neuroscience*, 2010. submitted.

[2] Adrien Wohrer and Pierre Kornprobst. Virtual retina: a biological retina model and simulator, with contrast gain control. *Journal of Computational Neuroscience*, 26(2):219–249, 2009.

[3] F. Rieke, D. Warland, Rob de Ruyter van Steveninck, and William Bialek. *Spikes, Exploring the Neural Code*. The M.I.T. Press, 1996.

[4] Alexandre Pouget and Gregory C DeAngelis. Paying attention to correlated neural activity. *Nature Neuroscience*, 11(12):1371–1372, December 2008.

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

[6] Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Matthew I. Grivich, Dumitru Petrusca, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. The structure of multi-neuron firing patterns in primate retina. *The Journal of Neuroscience*, 26(32):8254–8266, August 2006.

[7] Jonathon Shlens, Greg D. Field, Jeffrey L. Gauthier, Martin Greschner, Alexander Sher, Alan 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.

[8] Olivier Marre, Sami El Boustani, Yves Frégnac, and Alain Destexhe. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Physical Review Letters*, 102(13):4, April 2009.

[9] E.T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106(620), 1957.

[10] B. Cessac. A discrete time neural network model with spiking neurons ii. dynamics with noise. *J.Math. Biol., submitted*, 2010.

[11] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Commun. Math. Phys.*, 9:267–278, 1968.

[12] D. Ruelle. *Statistical Mechanics: Rigorous results*. Benjamin, New York, 1969.

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

[14] G. Keller. *Equilibrium States in Ergodic Theory*. Cambridge University Press, 1998.

[15] J.R. Chazottes and G. Keller. *Pressure and Equilibrium States in Ergodic Theory*, chapter Ergodic Theory, page 26. Encyclopedia of Complexity and System Science, Springer, 2009.

[16] B. Cessac, H. Rostro-Gonzalez, J.C. Vasquez, and T. Viéville. How gibbs distribution may naturally arise from synaptic adaptation mechanisms: a model based argumentation. *J. Stat. Phys*, 136(3):565–602, August 2009.

[17] Marc Mezard and Thierry Mora. Constraint satisfaction problems and neural networks: A statistical physics perspective. *Journal of Physiology-Paris*, 103(1–2):107–113, January–March 2009.

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