PHYSICAL REVIEW LETTERS

Prediction of Spatiotemporal Patterns of Neural Activity from Pairwise Correlations

O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe*

Unité de Neurosciences Intégratives et Computationnelles (UNIC), UPR CNRS 2191, Gif-sur-Yvette, France
(Received 17 November 2008; published 2 April 2009)

We designed a model-based analysis to predict the occurrence of population patterns in distributed spiking activity. Using a maximum entropy principle with a Markovian assumption, we obtain a model that accounts for both spatial and temporal pairwise correlations among neurons. This model is tested on data generated with a Glauber spin-glass system and is shown to correctly predict the occurrence probabilities of spatiotemporal patterns significantly better than Ising models only based on spatial correlations. This increase of predictability was also observed on experimental data recorded in parietal cortex during slow-wave sleep. This approach can also be used to generate surrogates that reproduce the spatial and temporal correlations of a given data set.

The structure of the cortical activity and its relevance to sensory processing or motor planning are a long-standing debate [1]. There is a need to describe the structure of the spiking activity based on well-defined statistical models. One approach consists of inferring the state of the network based on Hidden Markov Models [2]. Another approach used maximum entropy models, which is common in the analysis of complex systems [3]. The latter focused on spike patterns lying within one time bin [4,5], but is not prone to predict the temporal statistics of the neural activity [6]. Here, we design a model inspired from both lines of research to better describe the neural dynamics. This model is a maximum entropy model-based on the correlation values, and respecting a Markovian assumption. Thus, it takes into account both spatial and temporal correlations. We show its ability to describe the spatiotemporal statistics of the activity on simple network models and recordings in the mammalian parietal cortex *in vivo*.

We consider N neurons whose spikes are recorded and binned, for a long time period, noted as $\{\sigma(t)\} := \{\sigma_i(t)\}_{i=1,\dots,N}$ where $\sigma_i \in \{-1; 1\}$. The purpose of a statistical model is to describe as closely as possible the probability distribution of the spatiotemporal patterns, $P(\{\sigma(t)\}, \{\sigma(t+1)\}, \dots)$ with a limited number of parameters. For that purpose, we make a Markovian hypothesis on this distribution, and aim at finding the joint distribution $P(\{\sigma\}^{\tau+1}; \{\sigma'\}^{\tau}) = P(\{\sigma\}^{\tau+1}|\{\sigma'\}^{\tau})P(\{\sigma'\}^{\tau})$ which maximizes the entropy $H(\{\sigma\}^{\tau+1}; \{\sigma\}^{\tau}) = -\sum_{\{\sigma\},\{\sigma'\}} P(\{\sigma\}^{\tau+1}; \{\sigma'\}^{\tau}) \ln[P(\{\sigma\}^{\tau+1}; \{\sigma'\}^{\tau})]$ with the constraints on the first- and second-order statistical moments of the activity $m_i = \langle \sigma_i \rangle$, $C_{ij} = \langle \sigma_i(t)\sigma_j(t) \rangle$ and $C_{ij}^1 = \langle \sigma_i(t)\sigma_j(t+1) \rangle$, the normalization constraint, and the marginal distribution constraint: $\sum_{\{\sigma'\}} P(\{\sigma\}^{\tau+1}; \{\sigma'\}^{\tau}) = P(\{\sigma\}^{\tau+1})$.

By using Lagrange multipliers, and then applying the marginal distribution constraint, we find

$$P(\{\sigma\}^{\tau+1}; \{\sigma'\}^{\tau}) = \frac{1}{Z(\{\sigma\})} \exp\left(\sum_{i=1}^{N} h_i^{\tau} \sigma_i' + \sum_{i,j=1}^{N} J_{ij}^{\tau} \sigma_i' \sigma_j' + \sum_{i,j=1}^{N} J_{ij}^{\tau+1,\tau} \sigma_i \sigma_j' \right) P(\{\sigma\}^{\tau+1}) \qquad (1)$$

$Z(\{\sigma\})$ being the conditional partition function, and $\{h_i, J_{ij}\}_{i,j=1}^{N}$ are the Lagrange multipliers corresponding to the constraints given by $\{m_i, C_{ij}\}_{i,j=1}^{N}$.

We assume that the detailed balance is satisfied for a stationary distribution $P_{\text{stat}}(\{\sigma\})$. Therefore, the Markovian matrix is also time-invariant and satisfies

$$P(\{\sigma'\}|\{\sigma\})P_{\text{stat}}(\{\sigma\}) = P(\{\sigma\}|\{\sigma'\})P_{\text{stat}}(\{\sigma'\}) \qquad (2)$$

so that

$$P(\{\sigma\}; \{\sigma'\}) = P(\{\sigma'\}|\{\sigma\})P_{\text{stat}}(\{\sigma\}) = \frac{\exp(\sum_{i=1}^{N} h_i \sigma_i + \sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j + \sum_{i,j=1}^{N} J_{ij}^1 \sigma_i'\sigma_j)}{Z(\{\sigma'\})} P_{\text{stat}}(\{\sigma'\}). \qquad (3)$$

We then develop the extensive quantity $\ln[Z(\{\sigma'\})]$ up to the second order:

$$\ln[Z(\{\sigma'\})] = \ln(Z_{\text{eff}}) - \sum_{i=1}^{N} h_i^r \sigma_i' - \sum_{i,j=1}^{N} J_{ij}^r \sigma_i' \sigma_j' + O(\delta\sigma'^3). \qquad (4)$$

The k-th order terms are k products of J_{ij}^1. This approximation is thus valid in the weak temporal correlation limit.

Note that the coefficients of this development, $\{h_i^r, J_{ij}^r\}_{i,j=1}^{N}$, can be obtained analytically from (3). The final form for the transition function then becomes

$$P(\{\sigma\}|\{\sigma'\}) = \frac{1}{Z_{\text{eff}}} \exp\left(\sum_{i=1}^{N} h_i \sigma_i + \sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j + \sum_{i,j=1}^{N} J_{ij}^1 \sigma_i'\sigma_j + \sum_{i=1}^{N} h_i^r \sigma_i' + \sum_{i,j=1}^{N} J_{ij}^r \sigma_i'\sigma_j' \right). \qquad (5)$$

Using the detailed balance, the stationary distribution is also restricted to the second order and has the generic form

$$P_{\text{stat}}(\{\sigma\}) = \frac{\exp(\sum_{i=1}^{N} h_i^{\text{stat}} \sigma_i + \sum_{i,j=1}^{N} J_{ij}^{\text{stat}} \sigma_i \sigma_j)}{\sum_{\{\sigma''\}} \exp(\sum_{i=1}^{N} h_i^{\text{stat}} \sigma_i'' + \sum_{i,j=1}^{N} J_{ij}^{\text{stat}} \sigma_i'' \sigma_j'')}. \tag{6}$$

Since

$$P_{\text{stat}}(\{\sigma\}) = \sum_{\{\sigma'\}} P(\{\sigma\}|\{\sigma'\}) P_{\text{stat}}(\{\sigma'\}), \tag{7}$$

the parameters $\{h_i^{\text{stat}}, J_{ij}^{\text{stat}}\}_{i,j=1}^{N}$ are fully determined by the m_i and C_{ij} values.

Numerically, we adopt a slightly different approach, which is shown to be equivalent to the approximation made above. We maximize separately the entropy of the stationary distribution $P_{\text{stat}}(\{\sigma\})$ and the time-invariant joint distribution $P(\{\sigma\}; \{\sigma'\})$, without the marginalization condition. We obtain (6) for $P_{\text{stat}}(\{\sigma\})$, and

$$P(\{\sigma\};\{\sigma'\}) = \frac{1}{Z_{\text{tr}}} \exp\Bigg(\sum_{i=1}^{N} h_i \sigma_i + \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j + \sum_{i,j=1}^{N} J_{ij}^1 \sigma_i' \sigma_j$$
$$+ \sum_{i=1}^{N} h_i' \sigma_i' + \sum_{i,j=1}^{N} J_{ij}' \sigma_i' \sigma_j'\Bigg). \tag{8}$$

The transition matrix is then determined by $P(\{\sigma\}|\{\sigma'\}) = \frac{P(\{\sigma\};\{\sigma'\})}{P_{\text{stat}}(\{\sigma'\})}$, which gives back (5) if we identify $h_i^r = h_i' - h_i^{\text{stat}}$ and $J_{ij}^r = J_{ij}' - J_{ij}^{\text{stat}}$.

This model contains seven sets of parameters, $\{h_i, h_i^{\text{stat}}, h_i^r, J_{ij}, J_{ij}^{\text{stat}}, J_{ij}^r, J_{ij}^1\}_{i,j=1}^{N}$. In order to be equivalent to the previous model, we must apply several constraints which will reduce the number of free parameters. The stationary parameters $\{h_i^{\text{stat}}, J_{ij}^{\text{stat}}\}_{i,j=1}^{N}$ are bound to the others by using the relation (7) as before. Then, we have to apply a normalization on the conditional probability distribution (5) to recover the marginalization condition, which is a special form of (4) with $Z_{\text{eff}} = \frac{Z_{\text{tr}}}{Z_{\text{stat}}}$. Therefore, the parameter set $\{h_i^r, J_{ij}^r\}_{i,j=1}^{N}$ is also defined by $\{h_i, J_{ij}, J_{ij}^1\}_{i,j=1}^{N}$ which are the only free parameters. This model is thus equivalent to the previous approximation and allows for more tractable numerical treatments.

To test the model, we first used a raster generated by a Glauber model [7], whose flip transition probability from one time step to the next is

$$W(\sigma_i \rightarrow -\sigma_i) = \frac{1}{2\tau_0}\Bigg[1 - \sigma_i(t)\tanh\Bigg(\sum_{j}[J_{ij}^g \sigma_j(t)$$
$$+ h_j^g \sigma_j(t)]\Bigg)\Bigg] \tag{9}$$

where τ_0 is the effective time constant and J_{ij}, h_i are coupling constants of the neurons σ [8].

To fit the model parameters to the corresponding m_i, C_{ij}, and C_{ij}^1 values, we started with an analytical approximation of the solution [9] followed by a gradient descent: at each time step, the m_i, C_{ij} and C_{ij}^1 predicted by the model were

estimated through a Monte Carlo algorithm, compared to the experimental ones, and the model parameters were updated according to the difference. The algorithm was stopped when the difference between the theoretical and experimental values was less than 0.005, of the order of the uncertainty on the m_i and C_{ij} estimations.

In the following, we compared this model to simpler versions already used in the literature. The "Ising model" has the same description of $P_{\text{stat}}(\{\sigma\})$, but assumed $P(\{\sigma\}, \{\sigma'\}) = P_{\text{stat}}(\{\sigma\})P_{\text{stat}}(\{\sigma'\})$ [4,6]. The "independent model" assumed no second-order interactions: all the previous parameters are null but the h_i^{stat}.

To estimate their performance in describing the statistics of the neural activity, we estimated the occurrence probability of several spiking patterns empirically and compared it to the ones predicted by each model. Figure 1 shows the prediction of the three models for the probability of patterns with, respectively, 1, 2, and 3 time bins. For 1-bin patterns, the Markov and the Ising model are equivalent, and showed a good prediction performance, with most of the points prediction being in the confidence interval of the estimated probability. For patterns with 2 and 3 time bins, the prediction remained satisfying for the Markov model, while it is strongly degraded for the Ising model. Note that the Ising and independent models give similar performances here, contrary to [4,5]. Indeed, for a broad range of parameters in the Glauber model, the absolute correlation values are weak. However, their temporal extent controlled by τ_0 [see Fig. 2(d)], is already sufficient to impair the Ising model performance.

We quantified the fit between the model prediction and the experimentally measured statistics by computing the Jensen-Shannon Divergence: $D_{\text{JS}}(P, Q) = H[0.5(P + Q)] - 0.5[H(P) + H(Q)]$ [where $H(\cdot)$ is the Shannon entropy] measures the similarity between two distributions P and Q. Figure 2(a) shows the value of D_{JS} for the three models, for different numbers of bins in the pattern. This confirmed our previous observation. For one bin, the Ising and the Markov model are equivalent, and performed better than the independent model. For two bins or more, the Markov model showed lower D_{JS} values than the Ising model and the independent model. This prediction performance does not vary significantly with the number of bins. The Markov model is thus able to predict the probability of a pattern even when it is composed of several bins. It thus describes with more accuracy the statistics of the neural activity over a large temporal extent.

The better performance of the Markov model compared to the Ising model has to be related with the shape of the correlation functions: if it can be reduced to a Dirac-like form, there should be no difference between the Markov and Ising models [case $\tau_0 = 1$ in Fig. 2(c) and 2(d)]. Above 1, the normalized difference $\delta \log(D_{\text{JS}}) = [\log(D_{\text{JS}}^{\text{Markov}}) - \log(D_{\text{JS}}^{\text{Ising}})]/\log(D_{\text{JS}}^{\text{Ising}})$ quickly increases to reach a peak performance of 120% around 2.5, and then slowly decreases to a plateau of 46% improvement from
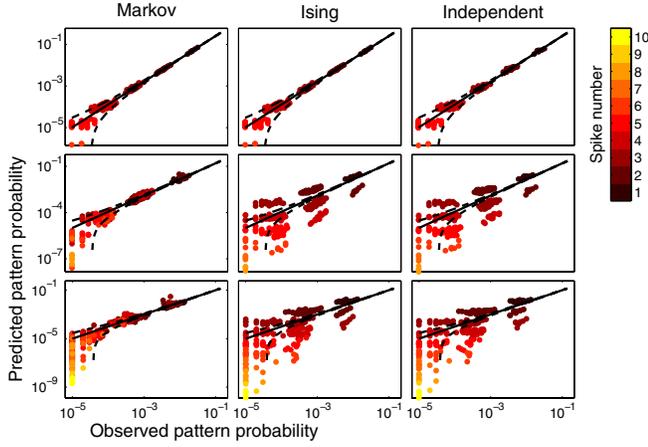
FIG. 1 (color online). Performance of the 3 statistical models to describe the statistics generated by the Glauber model ($\tau_0 = 2$). For each panel, we compared the probability of several patterns estimated empirically from the raster, and predicted by the corresponding model. Each point corresponds to a different pattern, picked up in the raster. The point color indicates the number of spikes in each pattern. The black line indicates equality, and the dashed curves the 95% confidence interval for the estimated probability. Each column corresponds to one of the three models described earlier. From left to right: the Markov, Ising and Independent models (see text). The different lines correspond to different pattern sizes (from top to bottom: 1, 2, and 3 temporal bins in the pattern).

the Ising to the Markov model, for $\tau_0 \gtrsim 10$. The Markov model thus performs better over a large range of τ_0 values. The prediction is at best when the ratio between the correlation time constant [Fig. 2(d)] and the bin size is around 2.5, but remains satisfying for larger ratios.

We also computed the fraction of the ensemble correlations that was captured by the Markov model, $\frac{I_2}{I_n} = \frac{S_1 - S_2}{S_1 - S_n}$, where S_k is the entropy when taking into account the correlations up to the k-th order [4]. This measures the improvement of the fit from the independent model to the Markov model. The value is maximal for two time bins, and then decreased [Fig. 2(b)], in line with the observed difference in D_{JS} between the independent and the Markov model. This Markov model is thus able to explain a major part of the higher order spatiotemporal statistics.

This model can also be used to generate surrogate rasters having the same statistics as the captured ones. For that purpose, starting from an initial random pattern, we generate at each time step a new pattern according to (5). We then compared the statistics of this new raster with the original prediction [Fig. 3(a)]. Although the generator only used the h_i, J_{ij}, and J_{ij}^1 coefficients of the model, the generated stationary probability is in very good agreement with the predicted stationary distribution estimated from the original data set, described by the h_i^{stat} and J_{ij}^{stat}. This result shows the consistency of the model: the transition matrix defined by the h_i, J_{ij}, and J_{ij}^1 parameters has indeed the stationary distribution defined by the h_i^{stat} and J_{ij}^{stat}
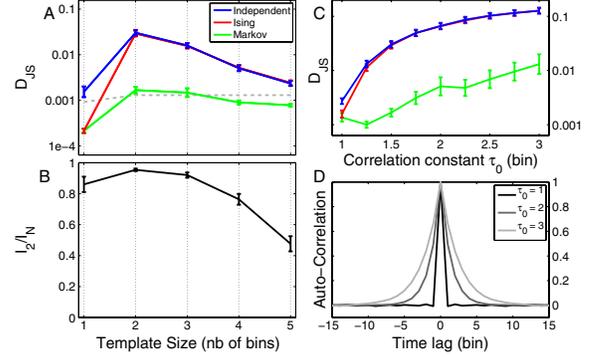

FIG. 2 (color online). Quantification of the models performance. (a) Jensen-Shannon Divergence D_{JS} between the prediction of the three statistical models, and the probabilities estimated empirically, for different pattern sizes. The raster has been generated by the Glauber numerical model with parameter $\tau_0 = 1.5$. The gray line indicates the value below which D_{JS} is not significantly different from zero ($p \leq 0.01$, [13]). (b) Quantification with the information ratio I_2/I_N. (c) Comparison for 2-bin pattern sizes, for different values of the τ_0 parameter in the Glauber model. (d) Autocorrelation of the population averaged activity for different τ_0.

coefficients in (6). We then applied the same analysis to the surrogate data, to obtain a model of the surrogate statistics. Figure 3(b) shows that we recover the same predictions than with the original analysis. The generator is thus producing a surrogate raster congruent with the statistical model.

We then tested the model on *in vivo* data, composed of 8 simultaneous multiunit recordings in the cat parietal cortex in different sleep states [Slow Wave Sleep (SWS) and Rapid Eye Movement (REM)] [10]. During SWS, the
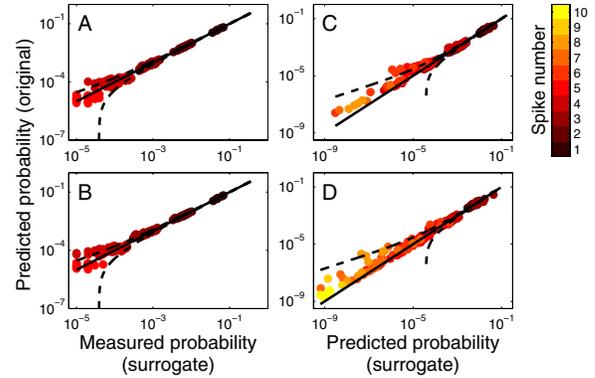

FIG. 3 (color online). Tests of the surrogate raster generator. (a) Comparison between the pattern probabilities in the surrogate raster, and the ones predicted by the model in the original analysis, for 1-bin patterns and a Glauber model with $\tau_0 = 1$ ($D_{JS} \simeq 0.0003$). Same representation as in Fig. 1. (b) Same comparison than (a) for a Glauber model with $\tau_0 = 1.5$ ($D_{JS} \simeq 0.0005$). (c) Comparison between the prediction of the model fitted on the original data ($\tau_0 = 2$), and the prediction fitted on the surrogate raster, for 2-bins pattern ($D_{JS} \simeq 0.0024$). (d) Same comparison than C for 3-bins patterns ($D_{JS} \simeq 0.0024$).
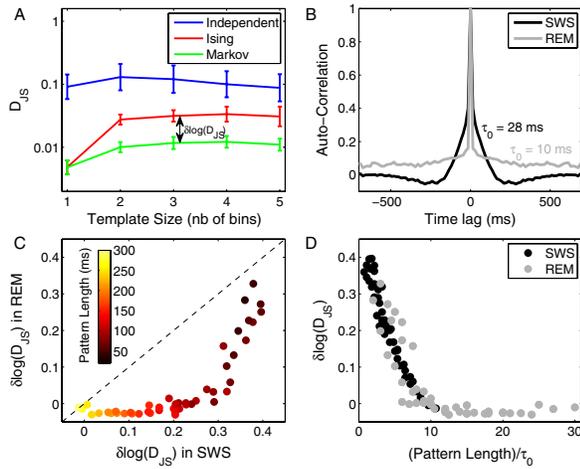
FIG. 4 (color online).   Test of the models on experimental data. (a) D_{JS} for the 3 models, estimated for the activity of 8 channels in cat parietal cortex, and for different template sizes. Bin width = 10 ms. (b) Autocorrelation of the population averaged activity for the SWS and REM sleep states. The correlation time constants τ_0 were estimated by fitting an exponential function. (c) Relative log-difference $\delta \log(D_{JS})$ between the Markov and Ising D_{JS}, compared for the SWS and the REM data. The dotted line indicates equality. The different points correspond to different combinations of template and bin sizes, color coded by the pattern length (template size $\times$ bin size). Points with black edge correspond to panel (b) values. (d) $\delta \log(D_{JS})$ for both states and for different pattern lengths, in unit of their respective correlation time constant (pattern length)$/\tau_0$.

performance of the Markov model is significantly higher than for the Ising model for different template sizes above 2 [Fig. 4(a)]. The improvement was comparable to the difference between independent and Ising models. We estimated $\delta \log(D_{JS})$ as above for different combinations of template and bin sizes. The result holds, with D_{JS} in the same order of magnitude, as long as the pattern length, defined as (template size) $\times$ (bin size), is below $\sim$120 ms [Fig. 4(c)]. To see how the sleep state affects this result, we compared the $\delta \log(D_{JS})$ between the SWS and the REM activities [Fig. 4(c)]. For pattern length below $\sim$120 ms, the improvement drops rapidly for the REM state. For very large pattern lengths ($\sim$300 ms), the Markov and Ising models perform equally well [$\delta \log(D_{JS}) = 0$] for both states. This faster drop of performance is related to the smaller correlation time constant in the REM state [Fig. 4(b)], and is reminiscent of the case $\tau_0 = 1$ in the Glauber model, while SWS seems more similar to the case $\tau_0 > 1$ [see Fig. 2(c)]. To further emphasize this relation, we measured the correlation time constant τ_0 for both states. We then computed $\delta \log(D_{JS})$ for different pattern lengths, expressed in unit numbers of their respective correlation time constant (pattern length)$/\tau_0$. When rescaled, both states exhibit the same dependency with the pattern length [Fig. 4(d)]. The Markov model is thus suited

for the analysis of different data sets and for pattern lengths up to 10 times their correlation time constant.

We have presented a probabilistic model which accounts for distributed spiking activity based on both spatial and temporal pairwise correlations. The model predicts the occurrence probability of spatiotemporal spike patterns, and can be used to generate surrogates which mimic the temporal and spatial correlation structure of the data. It would be interesting to test it on the specific data for which the Ising model fails [6]. Beyond spiking activity, other event-based data with long enough recordings might be interesting to analyze with this model [11]. This method of analysis will help to tackle fundamental issues about the structure of the neural activity, like the existence of higher order statistics or the Markovian nature of the temporal correlations. It could also impact on a broad range of areas of physics and biology [12].

*Corresponding Author: destexhe@unic.cnrs-gif.fr
[1] M. Abeles, *Local Cortical Circuits: An Electrophysiological Study* (Springer-Verlag, Berlin, 1982).
[2] B. M. Yu, A. Afshar, G. Santhanam, S. I. Ryu, K. V. Shenoy, and M. Sahani, Adv. Neural Inf. Process. Syst. **18**, 1545 (2006).
[3] J. N. Kapur, *Maximum Entropy Models in Science and Engineering* (Wiley, New York, 1990).
[4] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature (London) **440**, 1007 (2006).
[5] J. Shlens, G. D. Field, J. L. Gauthier, M. I. Grivich, D. Petrusca, A. Sher, A. M. Litke, and E. J. Chichilnisky, J. Neurosci. **26**, 8254 (2006).
[6] A. Tang, D. Jackson, J. Hobbs, W. Chen, J. L. Smith, H. Patel, A. Prieto, D. Petrusca, M. I. Grivich, A. Sher, P. Hottowy, W. Dabrowski, A. M. Litke, and J. M. Beggs, J. Neurosci. **28**, 505 (2008).
[7] K. H. Fischer and J. A. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge, England, 1991).
[8] We used a Glauber model of 8 units, with J_{ij}^g uniformly chosen in $[0.1, 0.1]$ and the h_j^g in $[-1.05, -1]$.
[9] T. Tanaka, Phys. Rev. E **58**, 2302 (1998).
[10] A. Destexhe, D. Contreras, and M. Steriade, J. Neurosci. **19**, 4595 (1999).
[11] C. Stosiek, O. Garaschuk, K. Holthoff, and A. Konnerth, Proc. Natl. Acad. Sci. U.S.A. **100**, 7319 (2003).
[12] Model available at ModelDB [http://senselab.med.yale.edu/ModelDB] (see also http://www.unic.cnrs-gif.fr).
[13] I. Grosse, P. Bernaola-Galvan, P. Carpena, R. Roman-Roldan, J. Oliver, and H. E. Stanley, Phys. Rev. E **65**, 041905 (2002).