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. To infer the state of the neural network, a first line of work has tried to model the neural activity with Hidden Markov Models [2, 3, 4]. Maximum entropy models have proved useful for the analysis of many complex systems (see for example [5,6]). Another line of research has used this approach to describe neural activity, focusing on spiking patterns lying within one time bin [7,8]. However, the latter is not prone to predict the temporal statistics of the neural activity [9]. In the following study, 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 spatio-temporal patterns significantly better than Ising models taking into account only pairwise 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.

\[ P(\{\sigma\}^{t+1}; \{\sigma'\}^t) = P(\{\sigma\}^{t+1}|\{\sigma'\}^t)P(\{\sigma'\}^t) \]

which maximizes the entropy

\[ H(\{\sigma\}^{t+1}; \{\sigma'\}^t) = -\sum_{\{\sigma\},\{\sigma'\}} P(\{\sigma\}^{t+1}; \{\sigma'\}^t) \ln P(\{\sigma\}^{t+1}; \{\sigma'\}^t) \]

with the constraints on the first- and second-order statistical moments of the activity \( m_{ij} = \langle \sigma_i \rangle >, C_{ij} = \langle \sigma_i \sigma_j(t) \rangle > \) and \( C_{ij}^t = \langle \sigma_i(t)\sigma_j(t+1) \rangle > \), the normalisation constraint, and the marginal distribution constraint:

\[ \sum_{\{\sigma\}} P(\{\sigma\}^{t+1}; \{\sigma'\}^t) = P(\{\sigma\}^{t+1}) \]

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

\[ P(\{\sigma\}^{t+1}|\{\sigma'\}^t) = \frac{1}{Z(\{\sigma\})} \exp \left( \sum_{i=1}^{N} h_i \sigma_i' + \sum_{i,j=1}^{N} J_{ij} \sigma_i' \sigma_j' \right) \]

\[ + \left( \sum_{i,j=1}^{N} J_{ij}^{t+1} \sigma_i' \sigma_j' \right) \]

\[ P(\{\sigma\}^{t+1}) \]

\[ Z(\{\sigma\}) \]

\[ \text{being the conditional partition function} \]

\[ \text{, and} \]

\[ P(\{\sigma\}^t|\{\sigma'\}^t)P(\text{stat})(\{\sigma\}) \]

so that:

\[ P(\{\sigma\}^t|\{\sigma'\}^t)P(\text{stat})(\{\sigma\}) = \frac{\exp \left( \sum_{i=1}^{N} h_i \sigma_i + \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j + \sum_{i,j=1}^{N} J_{ij}^{t+1} \sigma_i \sigma_j' \right)}{Z(\{\sigma'\})} P(\{\sigma'\}) \]
We then develop the extensive quantity \( \ln(Z(\{\sigma'\})) \) up to the second order:

\[
\ln(Z(\{\sigma'\})) = \ln(Z_{eff}) - \sum_{i=1}^{N} h'_i \sigma'_i - \sum_{i,j=1}^{N} J'_{ij} \sigma'_i \sigma'_j + o(\delta \sigma'^4)
\]

The k-th order terms are k products of \( J'_{ij} \). This approximation is thus valid in the weak temporal correlation limit.

Note that the coefficients of this development, \( \{h'_i, J'_{ij}\}_{i,j=1}^{N} \), can be obtained analytically from (3) by a straightforward computation. The final form for the transition function then becomes:

\[
P(\{\sigma\}|\{\sigma'\}) = \frac{1}{Z_{eff}} \exp \left( \sum_{i=1}^{N} h'_i \sigma_i + \sum_{i,j=1}^{N} J'_{ij} \sigma_i \sigma_j + \sum_{i=1}^{N} J'_i \sigma'_i + \sum_{i,j=1}^{N} J'_{ij} \sigma'_i \sigma'_j \right)
\]

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

\[
P_{stat}(\{\sigma\}) = \frac{\exp \left( \sum_{i=1}^{N} h'_{stat} \sigma_i + \sum_{i,j=1}^{N} J'_{ij} \sigma_i \sigma_j \right)}{\sum_{\{\sigma'\}} \exp \left( \sum_{i=1}^{N} h'_{stat} \sigma'_i + \sum_{i,j=1}^{N} J'_{ij} \sigma'_i \sigma'_j \right)}
\]

Since

\[
P_{stat}(\{\sigma\}) = \sum_{\{\sigma'\}} P(\{\sigma\}|\{\sigma'\}) P_{stat}(\{\sigma'\})
\]

the parameters \( \{h'_{stat}, J'_{ij}\}_{i,j=1}^{N} \) 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 \( P_{stat}(\{\sigma\}) \) and the time-invariant joint distribution \( P(\{\sigma\};\{\sigma'\}) \), without the marginalization condition. We obtain (6) for \( P_{stat}(\{\sigma\}) \), and:

\[
P(\{\sigma\};\{\sigma'\}) = \frac{1}{Z_{tr}} \exp \left( \sum_{i=1}^{N} h_i \sigma_i + \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j + \sum_{i=1}^{N} J'_i \sigma'_i + \sum_{i,j=1}^{N} J'_{ij} \sigma'_i \sigma'_j \right)
\]

The transition matrix is then determined by:

\[
P(\{\sigma\}|\{\sigma'\}) = \frac{P(\{\sigma\};\{\sigma'\})}{P_{stat}(\{\sigma'\})},
\]

which gives back (5) if we identify \( h'_i = h'_i - h_{stat}' \) and \( J'_{ij} = J'_{ij} - J'_{stat} \).

This model contains seven sets of parameters, \( \{h_i, h_{stat}'_i, J_{ij}, J_{stat}''_{ij}\}_{i,j=1}^{N} \). 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 \( \{h'_{stat}, J'_{stat}''_{ij}\}_{i,j=1}^{N} \) 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 \( Z_{eff} = \frac{1}{Z_{stat}} \). Therefore, the parameter set \( \{h'_i, J'_{ij}\}_{i,j=1}^{N} \) is also defined by \( \{h_i, J_{ij}, J_{stat}''_{ij}\}_{i,j=1}^{N} \) 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 [10], whose flip transition probability from one time step to the next is

\[
W(\sigma_i \rightarrow -\sigma_i) = \frac{1}{2\tau_0} \left( 1 - \sigma_i(t) \tanh \left( \sum_j J'_{ij}(t) + h'_{ij}(t) \right) \right)
\]

where \( \tau_0 \) is the effective time constant and \( J_{ij}, h_i \) are coupling constants of the neurons \( \sigma \) [11].

Fitting the model parameters to the corresponding \( m_i, C_{ij} \) and \( C'_{ij} \) values is a classical Boltzmann machine learning problem [12]. We started with an analytical approximation of the solution [13] followed by a gradient descent: at each time step, the \( m_i, C_{ij} \) and \( C'_{ij} \) 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 \( P_{stat}(\{\sigma\}) \), but assumed \( P(\{\sigma\};\{\sigma'\}) = P_{stat}(\{\sigma\}) P_{stat}(\{\sigma'\}) \) (this is equivalent to assume \( C_{ij} = m_im_j \)). The “independent model” assumed no second order interactions: all the previous parameters are null but the \( h'_{stat} \).

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 \cite{7} \cite{8}. Indeed, for a broad range of parameters in the Glauber model, the absolute correlation values are weak. However, their temporal extent controlled by \( \tau_0 \) (see Fig. 2D), 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: \( D_{JS}(P, Q) = H(0.5(P + Q)) - 0.5(H(P) + H(Q)) \) (where \( H(\cdot) \) is the Shannon entropy) measures the similarity between two distributions \( P \) and \( Q \) \cite{14}. Figure 2A 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 the temporal correlation functions 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. 2C-D). Above 1, the normalized difference \( \delta \log(D_{JS}) = (\log(D_{Markov}) - \log(D_{Ising}))/\log(D_{Ising}) \) quickly increases to reach a peak performance of 120% around 2.5, and then slowly decreases to a plateau of 46 % improvement from the Ising to the Markov model, for \( \tau_0 \geq 10 \). The Markov model thus performs better over a large range of \( \tau_0 \) values.
ues. From the experimental perspective, the Markov model prediction is at best when the ratio between the correlation time constant (Fig. 2D) 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, \( \frac{S_k}{S_1} \), where \( S_k \) is the entropy when taking into account the correlations up to the k-th order [7, 16]. 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. 2B), 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 spatio-temporal statistics.

Apart from describing the statistics of the activity, this model can also be used to generate surrogate rasters having the same statistics than 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. 3A). Although the generator only used the \( h_i \), \( J_{ij} \) and \( J_{ij}^{\text{stat}} \) 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 \( h_i^{\text{stat}} \) and \( J_{ij}^{\text{stat}} \). This result shows the consistency of the model: the transition matrix defined by the \( h_i \), \( J_{ij} \) and \( J_{ij}^{\text{stat}} \) parameters has indeed the stationary distribution defined by the \( h_i^{\text{stat}} \) and \( J_{ij}^{\text{stat}} \) coefficients in (6).

We then applied the same analysis to the surrogate data, to obtain a model of the surrogate statistics. Fig. 3B 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 biological data taken from [17], composed of 8 simultaneous multi-unit recordings in the cat parietal cortex in different sleep states (Slow Wave Sleep (SWS) and Rapid Eye Movement (REM)). For the activity recorded during SWS, the performance of the Markov model is significantly higher than for the Ising model. For a bin size of 10 ms, this was the case for different template sizes above 2 (Fig. 4A). The improvement was comparable to the difference between independent and Ising models. We estimated \( \delta \log(D_{JS}) \), the normalized log-difference between the Markov and Ising associated \( D_{JS} \), for different combinations of template and bin sizes. The result holds, with \( D_{JS} \) in the same order of magnitude, for larger bin sizes as long as the pattern length, defined as (template size) x (bin size), is below \( \sim 120 \) ms (Fig. 4C).

To see how the sleep state affects this result, we compared the \( \delta \log(D_{JS}) \) between the SWS and the REM activities (Fig. 4C). For pattern length below \( \sim 120 \) ms, while the Markov model outperforms the Ising model in describing the SWS activity, 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. 4B). This is indeed reminiscent of the case \( \tau_0 = 1 \) in the Glauber model (see Fig. 3C), and as a consequence, we observed no significant difference between the Ising and Markov models for intermediate pattern lengths. On the contrary, the SWS state exhibits larger correlation extent (similar to \( \tau_0 > 1 \) in Fig. 3C), and shows a persistent difference \( \delta \log(D_{JS}) \). To further emphasize this relation, we measured the correlation time constant \( \tau_0 \) 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)/τ₀. When rescaled, both states exhibit the same dependency with the pattern length (Fig. 4D). The Markov model is thus suited for the analysis of temporally correlated activity for different data sets and for pattern lengths up to 10 times their correlation time constant.

In conclusion, we have presented a probabilistic model which gives an account of the distributed spiking activity with relatively few parameters, and takes into account both spatial and temporal pairwise correlations. The model still predicts the occurrence probability of larger temporal 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 that have been used to show the failure of the ising model [9]. Beyond spiking assembly activity, other event-based data with long enough recordings might be interesting to analyze with this model (for example calcium transients [18]). 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 which used maximum entropy models [19].

We thank Michael Berry and Valérie Ego-Stengel for helpful discussions. Experimental data were obtained with Diego Contreras and Mircea Steriade, and were published previously [17]. Support by CNRS, ANR (Natstats, HR-cortex), and EU (Bio-I3: Facets FP6-2004-IST-FTPI 15879) grants. O.M. was supported by DGA and FRM fellowships.

![Image of graph](https://example.com/graph.png)

**FIG. 4:** (Color online) Test of the models on experimental data. A: Jensen-Shannon divergence $D_{JS}$ for the 3 models, estimated for the activity of 8 channels in cat parietal cortex, and for different template sizes. Bin width of 10 ms. B: Auto-correlation of the population averaged activity for the SWS and REM sleep states. The correlation time constants τ₀ 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, colour coded by the pattern length (template size x 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)/τ₀.
Roldan, J. Oliver, & H.E. Stanley. *Phys Rev E*, **65**:041905 (2002).

[16] E. Schneidman, S. Still, M.J. Berry, & W. Bialek. *Phys Rev Lett*, **91** (23):238701 (2003).

[17] A. Destexhe, D. Contreras, & M. Steriade. *J Neurosci*, **19**: 4595 (1999).

[18] C. Stosiek, O. Garaschuk, K. Holthoff, & A. Konnerth *Proc Natl Acad Sci U S A*, **100** (12):7319 (2003).

[19] The code of this model is available on ModelDB [http://senselab.med.yale.edu/ModelDB/](http://senselab.med.yale.edu/ModelDB/) (for more information, see [http://www.unic.cnrs-gif.fr](http://www.unic.cnrs-gif.fr)).