Maximum entropy models reveal the correlation structure in cortical neural activity during wakefulness and sleep

Trang-Anh Nghiem,[1] Bartosz Teleńczuk,[1] Olivier Marre,[2] Alain Destexhe,[1,*] and Ulisse Ferrari[3,4,*]

[1]*Laboratory of Computational Neuroscience, Unit de Neurosciences, Information et Complexit, CNRS, Gif-Sur-Yvette, France.*
[2]*Institut de la Vision, INSERM and UMPC, 17 rue Moreau, 75012 Paris, France*
[3]*Sorbonne Université, INSERM, CNRS, Institut de la Vision, 17 rue Moreau, 75012 Paris, France*
[4]*Correspondence should be sent to* `ulisse.ferrari@gmail.com`.

Maximum Entropy models can be inferred from large data-sets to uncover how local interactions can generate collective dynamics. Here, we employ such methods to investigate the characteristics of neurons recorded by multielectrode arrays in the cortex of human and monkey throughout states of wakefulness and sleep. Taking advantage of the separation of cells into excitatory and inhibitory types, we construct a model including this biological feature. By comparing the performances of Maximum Entropy models at predicting neural activity in wakefulness and deep sleep, we identify the dominant interactions between neurons in each brain state. We find that during wakefulness, dominant functional interactions are pairwise while during sleep, interactions are population-wide. In particular, inhibitory cells are shown to be strongly tuned to the inhibitory population. This shows that Maximum Entropy models can be useful to analyze data-sets with excitatory and inhibitory cells, and can reveal the role of inhibitory cells in organizing coherent dynamics in cerebral cortex.

Keywords: Maximum Entropy models, human cortex, monkey cortex, brain states, wakefulness, Slow-Wave Sleep

Introduction

To analyze a complex system, one is interested in finding a model able to explain the most about empirical data, with the least forms of interactions involved. Such a model should reproduce the statistics observed in the data, while making the least possible assumptions on the structure and parameters of the system. In other terms, one needs the simplest, most generic model that generates statistics matching the empirical values - this implies *maximisation of entropy* (MaxEnt) in the system, with constraints imposed by the empirical statistics [7].

In a seminal paper [14], a framework equivalent to the Ising model in statistical physics was used to analyze the collective behavior of neurons. This approach was based on the assumption that pairwise interactions between neurons can account for the collective activity of the neural population. Indeed, it was shown for experimental data, from the retina and cerebral cortex, that this approach can predict higher order statistics, including the probability distribution of the whole population's spiking activity. Even though the empirical pairwise correlations were very weak, the model performed significantly better than a model reproducing only the firing rates without considering correlations. The Ising model was subsequently demonstrated to efficiently reproduce the data better than models with smaller entropy [4], as well as to analyse neural recordings in a variety of brain regions in different animals, ranging from the salamander retina [14, 18] to the cerebral cortex of mice [6], rats [16], and cats [9].

A complementary approach was recently introduced [11], aiming at reproducing the dependencies between single neuron activity and whole-population dynamics in the mouse and monkey visual cortex. Later work [5] generalised this approach to model the full profile of coupling strength variation with increasing population activity, and applied the model to the salamander retina.

Recent advances in experimental methods have allowed the recording of the spiking activity of up to a hundred neurons throughout hours of wakefulness and sleep, for instance using multi-electrode arrays, also known as Utah arrays. Inspection of neurons' spike waveforms and their functional interaction with other neurons made the

---

*These authors contributed equally.

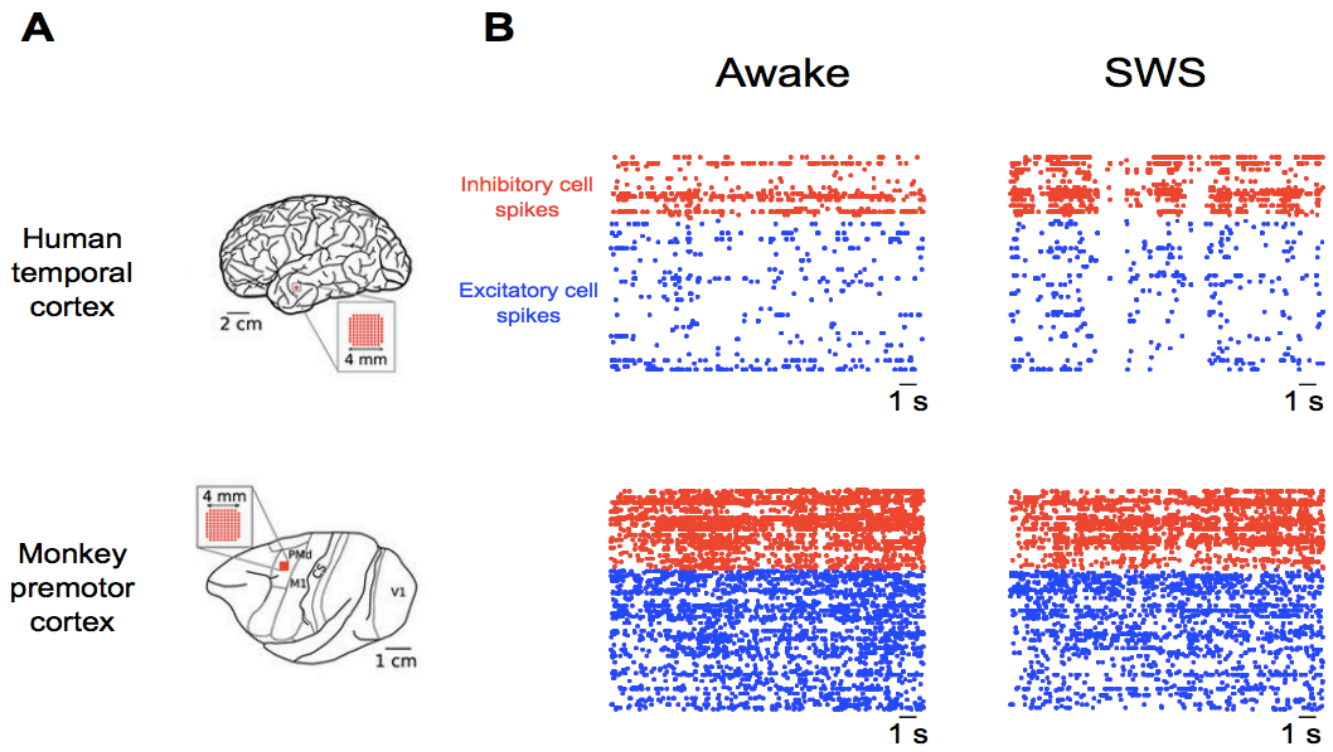

FIG. 1: **Multi-electrode array recordings. A**) Utah array position in human temporal cortex (top) and monkey prefrontal cortex (bottom) [17]. **B**) Awake (left) and SWS (right) raster plots for human (top) and monkey (bottom) data. Excitatory (E) cells are represented in blue, while inhibitory (I) cells are in red.

discrimination of excitatory (E) and inhibitory (I) cell types possible [2, 12]. Such data-sets therefore provide a further step in the probing of the system, due to the unprecedented availability of the simultaneously recorded dynamics of E and I neurons.

In the present paper, we would like to apply Maximum Entropy models to analyze human and monkey Utah-array recordings, which is done here for the first time. We would also like to investigate in which way such models may treat the availability of two recorded (E, I) populations, and what sort of supplementary insight they provide on the system's global dynamics.

Results

We study Utah-array recordings of spiking activity in the temporal cortex of a human patient and in the premotor cortex of a macaque monkey (see methods), in wakefulness and SWS, as shown in Fig. 1. Spike times are binned into time-bins of 50 ms (human data) and 25 ms (monkey data) to produce the population's spiking patterns (see Methods). From there, we can compute the correlations in the spiking patterns and fit models to these empirical correlations.

A. Pairwise Ising model

Pairwise correlations have been found to exhibit invariance with distance [13], even across brain regions [8]. Here, we wonder what this intriguing observation implies for functional interactions between neurons, and the information conveyed by pairwise correlations on such interactions. Therefore, we investigate whether pairwise covariances are sufficient to capture the main features of neural activity, for E and I cells during wakefulness and SWS.

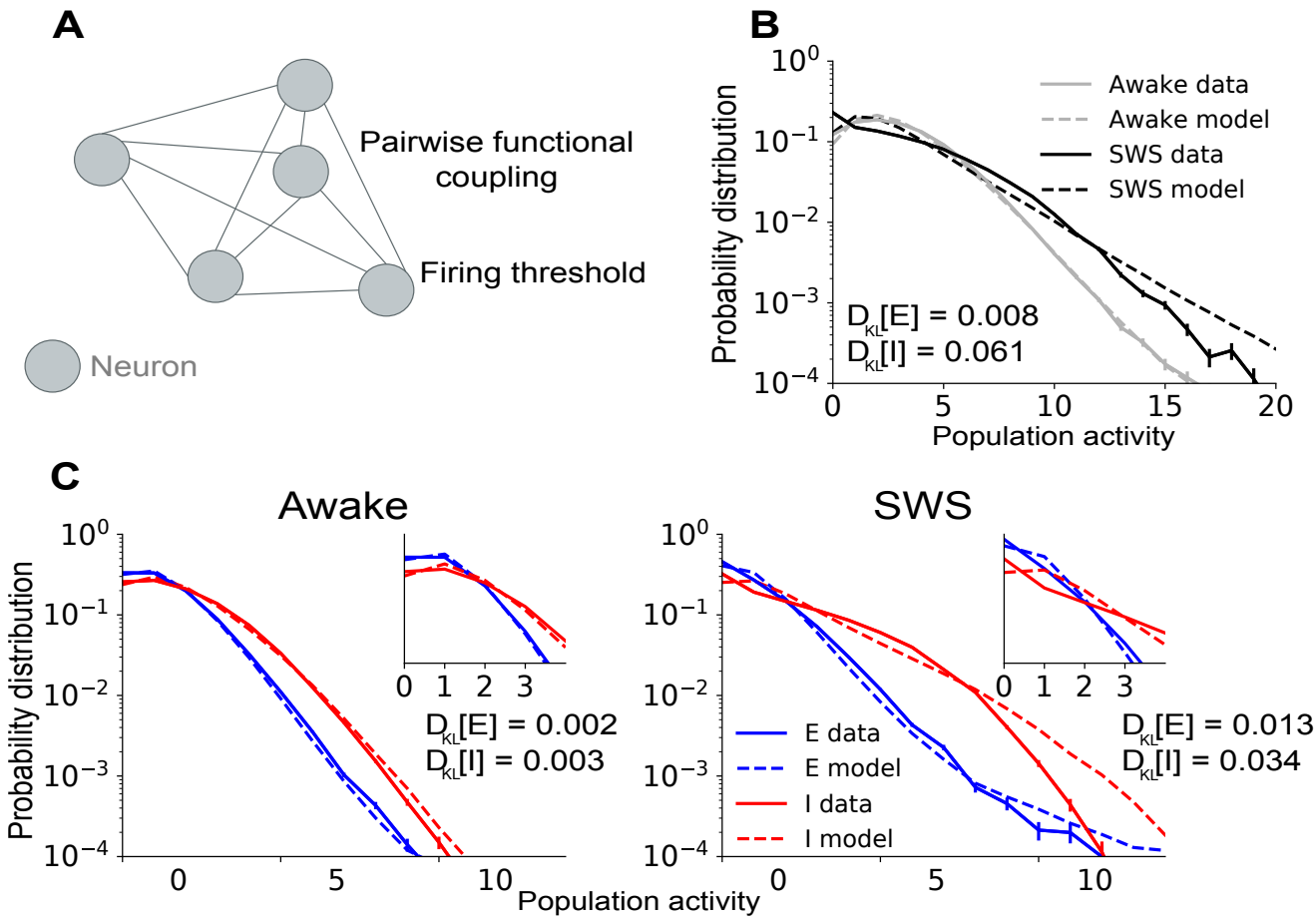

FIG. 2: **Pairwise Ising model fails to predict SWS synchronous activity, especially for inhibitory cells. A**) Model schematic diagram. Model parameters are each neuron's bias toward firing, and symmetric functional couplings between each pair of neurons. **B**) Empirical and predicted probability distributions of the population activity $K = \sum_i \sigma_i$ for the population of excitatory (E) cells, and that of inhibitory (I) cells. The Ising model more successfully captures the population statistics during Awake than SWS, especially for medium and large K values. **C**) Empirical and predicted population activities for E and I cells. The model particularly fails at reproducing the statistics of inhibitory population activity. These results are consistent with the presence of transients of high activity and strong synchrony between I cells during SWS.

To test this, we use a maximum entropy model, the most generic and least constrained model, to reproduce only and exactly the single neurons' spiking probability, and the pairwise covariances observed in the data.
As has already been shown [1, 14], this model takes the form of a disordered Ising model (see picture in Fig. 2A):

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\left(\sum_i b_i \sigma_i + \sum_{i \neq j} J_{ij} \sigma_i \sigma_j \right) \tag{1}$$

where we denote σ_i activity of neuron i given time bin (1: spike, 0: silence), b_i the bias of neuron i, controlling how heavily it is biased toward firing or silence, and J_{ij} the (symmetric) coupling between neurons i and j, tuning the resulting pairwise covariance between the neurons.

We use the algorithm introduced by [3] to infer the model's parameters b_i and J_{ij} on Awake and SWS data separately, for each Utah array recording.
Then we test how well the model describes neural activity during wakefulness as compared to during SWS. In particular, synchronous events involving many neurons may not be well accounted by the pairwise nature of the Ising model interactions. To test this, as introduced in [18], we quantify the empirical probability of having K neurons active in the same time window: $K = \sum_i \sigma_i$. Fig. 2B compare the empirical probability distributions with model predictions. During the Awake state, the Ising model is able to account for the empirical statistics, while during SWS it partially fails.

This is confirmed by the measures of the Kullback-Leibler divergence, $D_{\mathrm{KL}} \equiv \sum_K P^{\mathrm{Data}}(K) \log P^{\mathrm{Data}}(K)/P^{\mathrm{Model}}(K)$, between empirical and model-predicted distributions (Fig. 2B). This difference can be ascribed to the presence of high activity transients, known to modulate neurons activity during SWS [15] and responsible for the larger covariances, as seen in [12]. In order to investigate the Ising model's failure during SWS, in Fig. 2C we compare the predictions for $P(K)$, separating excitatory and inhibitory cell populations. For Awake periods, the model is able to reproduce both cell types' behaviors. However, during periods, the model largely fails at predicting the empirical statistics, in particular for the I population. This is confirmed by estimates of the Kullback-Leibler divergences (see Fig. 2). Fig. S1 shows similar results for the analysis on monkey recording.

These results highlight the relevance of the pairwise Ising model to reproduce $P(K)$ for all neurons, E and I, during wakefulness. Neural dynamics during wakefulness can therefore be described as predominantly driven by pairwise interactions. Conversely, the model's notable over-estimation of $P(K^I)$ for inhibitory neurons during SWS suggests that other dynamics are at work during SWS, especially on I neurons, that pairwise couplings alone may not account for.

Higher-order, perhaps even population-wide interactions are needed to accurately depict neural activity during SWS. This is consistent with the synchronous firing of the neural population during transients of high activity in SWS, correlating neurons across long distances, most notably pairs of I cells[8].

So far, our findings from inferring a pairwise Ising model on our datasets have highlighted that while pairwise interactions were sufficient to depict neural activity during wakefulness, higher-order, population-wide interactions may be at work during SWS, especially on I cells.

B. Single-population model

While the pairwise Ising model appeared to be sufficient to describe neural network dynamics during wakefulness, it failed to account for SWS activity, especially for I cells. This breakdown is likely due to transients of high activity and strong population-wide synchrony in SWS [12]. To account for such dynamics, one could consider the interaction between each cell and the whole population: indeed, such approaches have proven to be able to successfully describe cortical neural activity [11]. We investigate whether cell-to-population interactions exist in our data-set by studying the neurons' tuning curves to the population. Neuron-to-population tuning curves (see Methods)indicate how much a neuron's activity is modulated by the total activity of the rest of the network [5]. In Fig. 3A we present tuning curves for ten example E or I cell during both Awake and SWS states. These examples provide strong evidence for cell-to-population tuning. In order to quantify population tuning, we estimate how much a cell, either E or I, is *sensitive* to the activity of the rest of the population, i.e. how much its activity fluctuates depending on the population activity (see Methods). As can be observed in Fig. 3B, and consistently with our previous results, we find that cells are sensitive to the population especially during SWS. Similar results are valid for the monkey recording as well (Fig. S2A). Since we have established cell-to-population interactions do take place during SWS, we wish to determine to what extent they are sufficient in capturing the characteristics of neural activity during sleep.

To this purpose, we use a null model [5] for the dependencies between neuron firing and population activity $P(\sigma_i = 1, k = K(\boldsymbol{\sigma})$ where we denote $K(\boldsymbol{\sigma})$ the number of neurons spiking in any time bin. In this model (Fig. 4A), the probability of cell firing is described by the strength of its coupling to the population:

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\Big(\sum_i h_{ik} \delta_k^{K(\boldsymbol{\sigma})} \sigma_i\Big), \tag{2}$$

where h_{iK} is the coupling between neuron i and the whole population when K neurons are active. δ_k^K is the Kronecker delta, taking value one when the number K of active neurons is equal to a given value k and zero otherwise. h_{iK} is the coupling from cell i to the population when K neurons are active. For example, a chorister neuron, that fires most often when many others are firing, would have h_{iK} increasing with K. Conversely, a soloist neuron, that fires more frequently when others are silent, would have h_{iK} decreasing with K. Z is the normalisation constant, that can be computed by summing over all possible activity configurations $Z = \sum_{\boldsymbol{\sigma}} \exp\Big(\sum_{i=1}^N h_{ik} \delta_k^K \sigma_i\Big)$. Importantly, Z and its derivative allow us to determine the statistics of the model, such as the mean firing rate and the pairwise covariances. As an analytical expression exists for Z, the statistics may be derived analytically from the values of the couplings, making this model solvable (see Methods).

To evaluate to what extent the model describes the data well and hence captures empirical statistics it was not designed to reproduce, we study the predicted pairwise correlations as compared to the empirical ones.

In order to test this model, in Fig. 4B, we compare the empirical pairwise covariances with their model predictions.

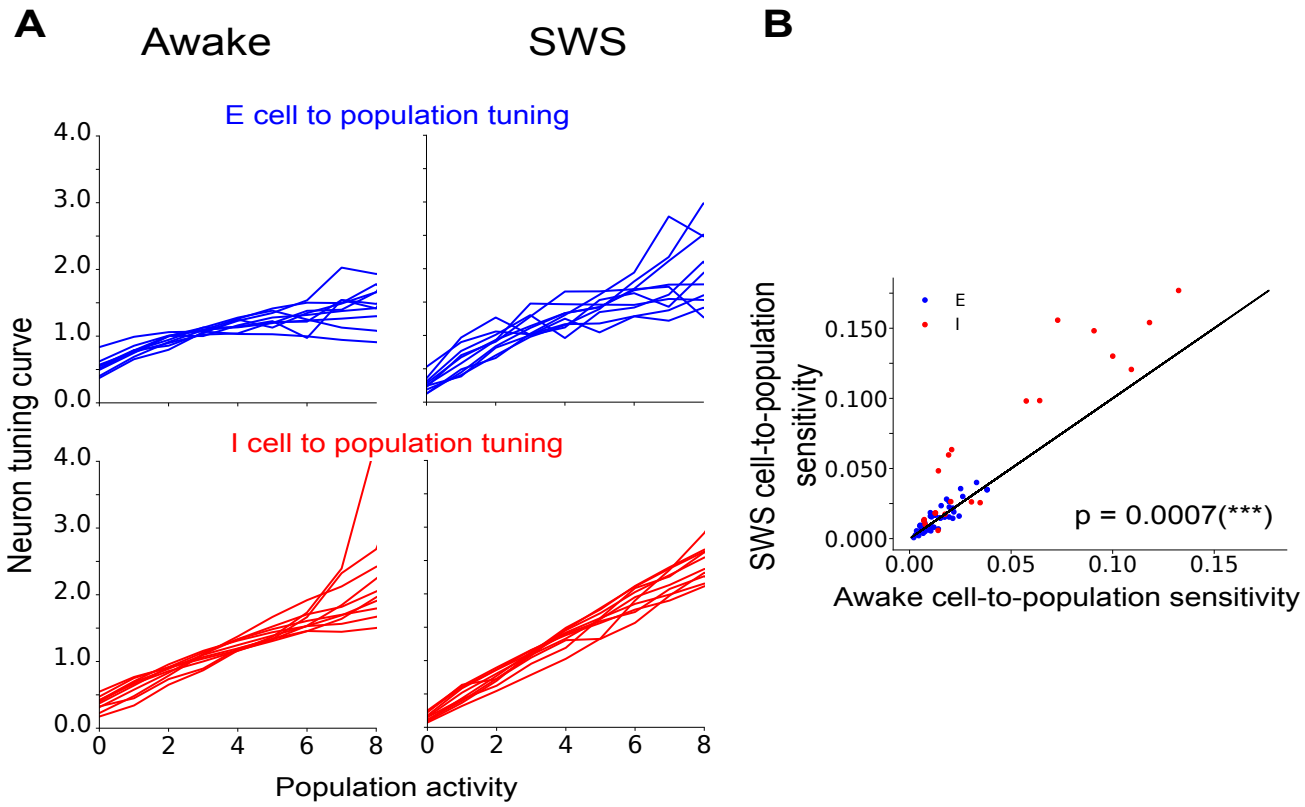

FIG. 3: **Neural firing is tuned to the neural population's activity, particularly during SWS. A**) Tuning curves of ten example cells (see text and methods) showing that cells are tuned to the population activity. **B**) Scatterplot of the cell sensitivity to the population activity (see Methods). Cells are very consistently more sensitive during SWS (p-value < 0.001, Wilcoxon sing-ranked test).

Pearson correlations confirm that the population statistics are better reproduced by the model during SWS than during awake states (Fig. 4). For monkey recording, the effect is even larger since the model entirely fails to account for Awake pairwise statistics (Fig. S2B). While the effect may be amplified by the fact that the Pearson correlations are relatively larger during SWS, this is the opposite of what was observed for the pairwise Ising model: a model reproducing only empirical neuron-to-population interactions seems adequate at depicting neural dynamics during SWS but not during wakefulness.

In particular, the model appears to perform best at reproducing the empirical statistics during SWS cell types for I neuron pairs. By contrast, E-E pairwise covariances are the most poorly reproduced during wakefulness. Consistently with our previous findings, this supports our hypothesis according to which inhibitory activity, and to a certain extent excitatory activity, is dominated by population-wide interactions rather than local pairwise mechanisms, such that a MaxEnt 'population model' is mostly sufficient at capturing the key dynamics. Nevertheless, this model still under-estimates the higher I-I pairwise covariances, suggesting something more may still be needed to capture the interactions involving I cells during SWS. Since they are strongly synchronised even across distances [2, 12], we hypothesise I cells could be tuned to the I population only, rather than the whole population.

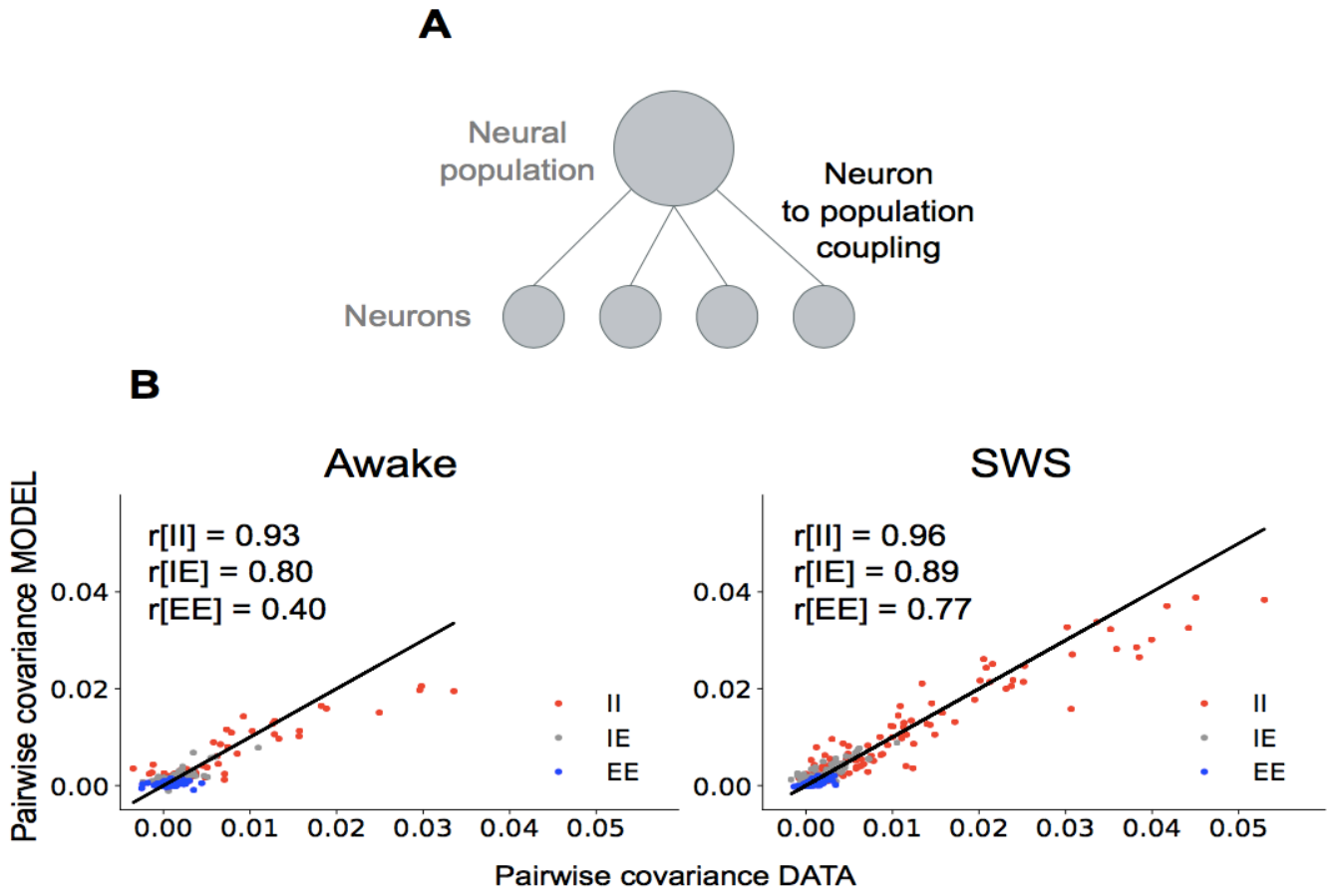

FIG. 4: **Single-population model shows better performance during SWS than wakefulness.** **A**) model schematic diagram. Parameters h_{iK} are the couplings between each cell i and the whole population activity $K = \sum_i \sigma_i$. **B**) Pairwise covariances, empirical against predicted, for Awake (left) and SWS (right) states. Consistently with Fig. 3B, the success for SWS, most noticeably for I-I pairs, suggests these cells are most responsive to whole-population activity.

C. Two-population model

We have shown that neural interactions during SWS were more population-wide than during wakefulness. However, interactions between each cell and the whole population could not account for the high covariances observed particularly amongst pairs of I cells in SWS. We therefore wonder if I cells are tuned to the I population only. Indeed, as shown in Fig. 5A, examination of the tuning curves of each cell to the E and the I populations separately revealed homogeneous and strong tuning for I cells to the I population, compared to I cells to the whole population or the E population. In order to quantify this effect, we estimated the cell sensitivity to both populations separately (see Methods). The comparison in Fig. 5B suggests I cells are significantly more sensitively tuned to the I population than to the E population. The effect is even larger for monkey recordings (Fig. S3A).

To study tuning to the two populations separately, we now refine the previous model to take into account the couplings between each cell and the E population and each cell and the I population separately. Because of the results of Fig. 5B, we expect this model to perform better at reproducing the main features of the data during SWS. We want the model to only and exactly reproduce the empirical $P(\sigma_i = 1, k^E = K^E(\boldsymbol{\sigma}))$ and $P(\sigma_i = 1, k^I = K^I(\boldsymbol{\sigma}))$ for all neurons i and all values empirically taken by K^E and K^I.
The probability of obtaining any firing pattern $\boldsymbol{\sigma}$ is given by (see picture in Fig. 6)

$$P(\boldsymbol{\sigma}) = \frac{1}{Z} \exp\Big(\sum_i (h^E_{ik^E} \delta^{K^E}_{k^E} + h^I_{ik^I} \delta^{K^I}_{k^I}) \sigma_i \Big), \tag{3}$$

where K^E is the number of E neurons spiking and K^I the number of I neurons spiking in any time bin, and $h^E_{iK^E}$ the

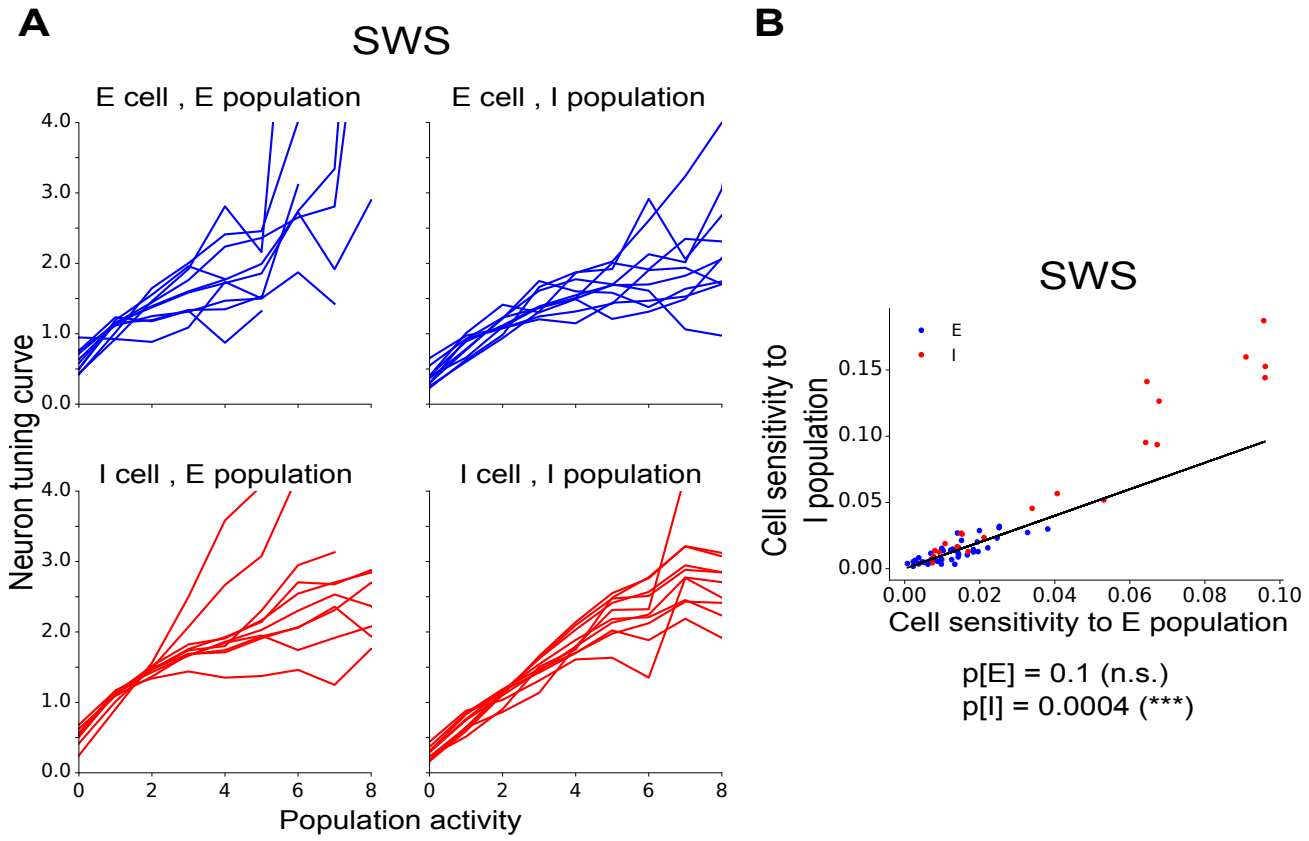

FIG. 5: **I cells are more specifically tuned to the I population during SWS. A**) Example tuning curves from ten cells of each type to each type of population during SWS, and similarly for the I population. **B**) Scatter-plot of cell sensitivity to E versus I population, during SWS. I cell are more tuned to I population than the E one (p-value $< 10^{-3}$, Wilcoxon sign-ranked test). E cells, instead, are weakly sensitive to both populations.

coupling between neuron i and the whole E population when k neurons are active, resp. $h^I_{iK^I}$ to the I population. Z the normalisation, $\delta^{K^E}_{k^E}$ is the Kronecker delta, equal to one when the number K^E of active E neurons is equal to a given value k^E, and respectively for inhibitory cells. It can be shown (see methods), following an analogous reasoning to that employed in [5], that this model is also analytically solvable in that the normalisation function Z may be derived analytically. Using the expression for Z, as described in the methods, allows us to analytically predict the model statistics for any given set of couplings.
As for the previous models, we want to assess whether this model is sufficient to describe the data, that is if it can accurately predict a data statistic it was not specifically designed to reproduce. To this purpose we test pairwise covariances. We also aim to evaluate how prediction performance compares with the single-population model on the whole population (Fig. 4) described previously.

For both human (Fig. 6B) and monkey (Fig. S3B) the two-population model provides better predictions for pairwise covariances than the single-population model in SWS. All the Pearson correlation are improved and large I-I covariance are no longer systematically under-estimated. To verify the improvement in model performance was not solely due to this model possessing more parameters, we repeat the inference on the same data with the cell types (E or I) shuffled, and find that the prediction deteriorates significantly, as highlighted in Fig. 6C. This demonstrates that taking into consideration each cell's couplings with the E population and the I population separately is more relevant than taking into account its couplings with any sub-populations of the same size.

As this confirms the two-population model is a good model for SWS, cell-to-population tuning can be considered

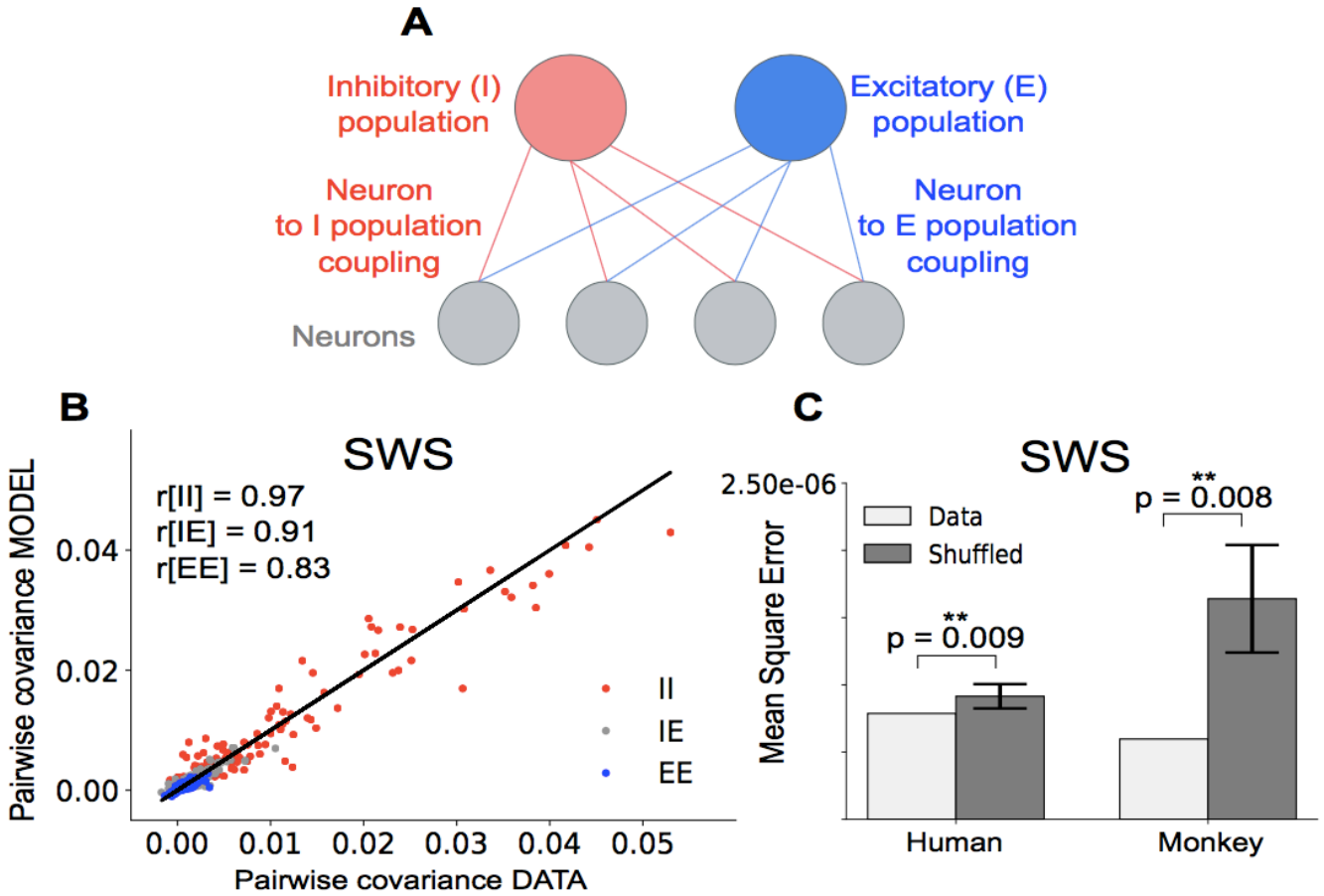

FIG. 6: **Two-population model shows significant improvement in prediction for all types of cells. A**) Schematic diagram of the two population model. Parameters $h^E_{iK^E}, h^I_{iK^I}$ are the couplings between each cell i and the E population activity $K^E = \sum_{i\in E} \sigma_i$ and the I population activity $K^I = \sum_{i\in I} \sigma_i$. **B**) Pairwise covariances, empirical against predicted, for the two-population model, during SWS. Improvement compared to the whole-population model is confirmed by the Pearson correlations. **C**) deterioration of prediction by shuffling cell types for the human and monkey data-sets. This effect demonstrates that knowledge of cell types significantly contributes to improving model prediction. This is confirmed by the Mann-Whitney U test p-values.

as the most relevant statistics to depict neural behaviour during SWS. This highlights the importance of I cells' tuning to the I population in understanding neural activity in SWS, as shown previously (Fig. 5).
Additionally, with the two-population model, E-I correlations are also reproduced with increased accuracy as compared to the single-population model. This result confirms the two-population model is an improvement over any of the previously used one-population models. Improved E-I covariance prediction also suggests that the two-population model successfully captures some of the cross-type interactions between the E and I populations, a non-trivial result since the two populations are not directly coupled to one another by design of the model.

Discussion

In this paper, we tested MaxEnt models on human and monkey Utah-array recordings where excitatory and inhibitory populations were discriminated, during the states of wakefulness and SWS. The models were designed to only reproduce one key feature of the data, to assess how well they captured the key mechanisms and hence reproduced empirical statistics for each brain state. We found that the pairwise Ising model's performance highlighted pairwise interactions as dominant in cortical activity during wakefulness, but insufficient to describe the inhibitory population activity during SWS. By contrast, MaxEnt models reproducing neuron-to-population dependencies

showed I neurons' activity was strongly modulated by the I population activity.

Comparison of MaxEnt models.
While pairwise Ising models had previously been shown to accurately predict statistical features of neural interaction in a wide array of data-sets [14, 16, 18] and theoretical scenarios [4], we observed that for the same neural system, activity can be well-reproduced in one brain state (wakefulness) and not the other (SWS)(Fig. 2). We identify I cells as responsible for breaking pairwise sufficiency during SWS, suggesting instead that I cells' interactions are long-range in space and population-wide, which explains recent empirical observations [8, 12].

Conversely, we found that models based on cell-to-whole population interactions, as introduced by [11], to only be relevant to SWS, entirely failing to replicate the empirical pairwise correlations in the monkey premotor cortex (Fig. reffigSupp:one-pop). Even for SWS, I cells' strong pairwise correlations were consistently underestimated.

Eventually, the two-population model provides a good trade-off for modelling neural interactions in SWS, and in particular the strongly correlated behaviour of I cells. As has become experimentally possible, discriminating E and I cell types within a recorded population greatly improves the capacity of a model to capture empirical neural dynamics. This also provides evidence for the existence of neural behaviour intermediate between 'soloist cell' and 'chorister cell'.

Model implications on internal dynamics of E and I populations, and E-I balance.
Our results also shed light on the workings of the E/I balance and its lessened robustness during SWS. Indeed, it has been shown that while I cells are less numerous, they are in equal parts more active, such that the number of spikes from E cells and from I cells at any point in time was similar, a balance that only remained robust at finer time scales during wakefulness [2].

Now, the two-population model's better reproduction of E-I pairwise covariances is compatible with the existence of such a balance: even though the E and I populations were not coupled by design of the model, this result is evidence that the model does capture the interaction between the two populations, suggesting synchrony between E and I populations to some degree.

Moreover, even with the same transients of high activity for the whole population, where the E/I balance is conserved, the way E cells interact with the E population is qualitatively different from that I cells interact with the I population. The different tuning curve shapes for E cells to the E population and I cells to the I populations support this: the strong and constant slopes for all I cell's tuning curves to the I population, for both high and low K^I seem to show that throughout both high-activity and low-activity transients, neuron-to-population interactions are predominant. This may suggest the role of I cells in synchronizing slow oscillations in the neural population, perhaps even across longer distances in the cortex.

Cell-to-cell and cell-to-population couplings as the competition between internal network dynamics and common external inputs.
Mechanisms underlying the interactions we noticed may occur at multiple scales, as well as involve the interplay between scales. Indeed, population-wide interactions winning over short-range pairwise interactions may be due to a different network connectivity, with reinforced structural couplings between I cells over long distances. However, this could also be explained by I cells presenting very similar responses to received inputs (as highlighted by similar tuning curves). Finally, larger common inputs to the I population, that may be different to common inputs to the E population, may also be a plausible mechanism driving the observed activity.

Using biophysical models of spiking neurons could allow the independent exploration of each of these mechanisms and their effects on the resulting cell-to-cell and cell-to-population interactions. Conversely, MaxEnt models could be used to constrain biophysical models in providing a quantitative way of comparing their statistical features to empirical neural statistics.

Acknowledgments

We thank M. Chalk, C. Gardella, G. Tkačik, M. di Volo and C. Capone for useful discussion. Research funded by European Community (Human Brain Project, H2020-720270), ANR TRAJECTORY, ANR OPTIMA, French State

program Investissements dAvenir managed by the Agence Nationale de la Recherche [LIFESENSES: ANR-10-LABX-65], NIH grant U01NS09050 and a AVIESAN-UNADEV grant.

Methods

A. Data-set

We work with an intra-cranial multi-electrode array recording of 92 neurons in the temporal cortex of an epileptic patient, the same data-set used by [12] and [2]. The record of interest spans across approximately 12 hours, including periods of wakefulness as well as several stages of sleep. Recordings were performed in layer II/III of the middle temporal gyrus, in an epileptic patient (even though far from the epileptic focus and therefore not recording epileptic activity outside of generalised seizures). Data acquisition in that region was enabled by implanting a multi-electrode array, of dimensions 1 mm in thickness and 4x4 mm in area, with 96 micro-electrodes separated by 400 μm spacings. The array was originally implanted for medical purposes. A 30 kHz sampling frequency was employed for recording. Switches in brain state (Awake, SWS, REM, seizure, ...) throughout the recording were noted from the patient's behavioural and physiological parameters. Using spike sorting methods on the obtained data, 92 neurons have been identified. Analysis of the spike waveforms for each of these cells allowed their classification as RS and FS cells. Using the spike times of each neurons, cross-correlograms for all pairs of neurons were also computed to determine whether each neuron's spikes had an excitatory (positive correlation) or an inhibitory (negative correlation) effect on other neurons through putative monosynaptic connections. It should be noted that cells found to be excitatory exactly corresponded to those classified as RS, while all inhibitory cells were also FS. We only retained cells spiking all throughout the recording for our analyses, yielding 71 cells of which 21 were I cells.
Similarly, spiking activity in layer III/IV of the premotor cortex of a macaque monkey was recorded by Utah-array, throughout a night, and an hour of target pursuit task on the following day. A 10 kHz sampling frequency was employed for recording. Classification of brain states was performed by examination of the Local Field Potential (LFP). Spike-sorting yielded 152 cells, of which 141 spiked throughout the whole recording. Clustering on features of the spike waveform has allowed for the sorting of cells as RS and FS. Excluding cells for which clustering was uncertain within a 30-percent margin yielded 81 cells, of which 38 were FS, over which all subsequent analyses were performed as presented in [2].
Time bin size was chosen in order to have one to few spikes from each neuron per time bin, while still having sufficient spikes from the whole population per time bin to compute statistics such as the pairwise covariances and the cell-to-population dependencies. Since I cells were consistently more active, this was equivalent to balancing a sufficient number of spikes from E neurons with sufficiently few spikes from any I neuron per time bin. In the human temporal cortex, where activity was considerably sparse, the chosen time bin size was 50 ms. In the interest of having comparable numbers of spikes per time bin and pairwise covariances, a time bin size of 25 ms was chosen for the monkey motor cortex.

B. Inference methods

Inferring the parameters from a MaxEnt model may be understood as a Lagrange multiplier problem, where one maximises the entropy while taking as constraints that the desired model-predicted statistics match their empirical values. Then each model parameters is the Lagrange multipliers for one constraint on one observable to reproduce. Taking, for example, the pairwise Ising model, the statistics we want to reproduce are the neuron mean firing rates and the pairwise covariances. The corresponding model parameters are the firing thresholds b_i and the pairwise couplings J_{ij} respectively. We therefore want to maximise

$$S_{\text{MaxEnt}} = \max_{P} \min_{b,J} \Bigg[-\sum_{\boldsymbol{\sigma}} P(\boldsymbol{\sigma}) \log P(\boldsymbol{\sigma}) + \sum_i b_i \Big(\sum_{\boldsymbol{\sigma'}} \sigma'_i P(\boldsymbol{\sigma'}) - < \sigma_i >_{\text{data}} \Big) \tag{4}$$

$$+ \sum_{j<i} J_{ij} \Big(\sum_{\boldsymbol{\sigma'}} \sigma'_i \sigma'_j P(\boldsymbol{\sigma'}) - < \sigma_i \sigma_j >_{\text{data}} \Big) \Bigg] . \tag{5}$$

One can verify that maximizing S_{MaxEnt} with respect to P and with the chosen constraints [1, 5], gives the form of each of the models given previously in Eq. 1, Eq. 2, and Eq. 3. A Hessian analysis may prove that the problem is well-posed, as the solution exists and it is unique.

The key challenge thus resides in finding a method that quickly converges to this solution. Thanks of the explicit form of Eq. (5), the gradient of the log-likelihood ($\ell = -S_{\text{MaxEnt}}$) with respect to the model parameters can be computed as differences between empirical and model-predicted averages of the conjugated observables. For example, the gradient with respect to the bias b_i of the Ising model can be estimated as $\langle \sigma_i \rangle_{\text{data}} - \langle \sigma_i \rangle_{\text{model}}$. The inference can thus be performed by an ascendant dynamics that requires to estimate model averages of observables. For the Ising model we applied the Markov-Chain Monte-Carlo method introduced in [3, 10]. For the one population model, we applied the Newton dynamics proposed in [5] For the two-population model, we modify the algorithm of [5] to take into account two populations. We found that a simpler steepest descent dynamics, that does not take into account the Hessian, was fast enough for our data-sets.

C. Tuning curves and sensitivity to population

In order to quantify the dependence of each cell on the rest of the population's activity, we used tuning curves [5] and sensitivity to the population.

Tuning curves characterize the dependence of the average activity of a cell conditioned to the activity of either the population, either the E or I sub-population. The tuning curve of neuron i is defined as $m_i(k)/\langle\sigma_i\rangle$, where $\langle\sigma_i\rangle$ gives the cell's mean activity across all time bins. $m_i(k)$ is defined as:

$$m_i(k) \equiv P(\,\sigma_i = 1\,|\,K_{\setminus i}(\sigma) = k\,) = \frac{P(\,\sigma_i = 1\,,\,K_{\setminus i}(\sigma) = k\,)}{P(\,K_{\setminus i} = k\,)} \tag{6}$$

where $K_{\setminus i}(\sigma)$ is the number of neurons activity in the configuration σ, when cell i has been excluded.

Sensitivity to population. A tuning curve shows the whole profile of the dependence of a cell activity on the rest of the population. In order to quantify this effect, we introduced the cell sensitivity to the population, depicting the cell's fluctuation in activity across states of population activity:

$$\text{Sensitivity}_i \equiv \sqrt{\sum_k \Big(\, m_i^2(k)\, P(\,K_{\setminus i} = k\,)\,\Big) - \sum_k \Big(\, m_i(k)\, P(\,K_{\setminus i} = k\,)\,\Big)^2} = \sqrt{\sum_k \Big(\, m_i^2(k)\, P(\,K_{\setminus i} = k\,)\,\Big) - \langle\sigma_i\rangle^2}\,. \tag{7}$$

D. Two-population model

In this section, we generalize the analysis of the one-population model introduced in [5] to the case of two populations. From our model introduced in Eq. 3, we can define the couplings $h^E_{iK^E} \equiv h^E_{ik^E}\delta^{K^E}_{k^E}$ for E cells to the E and the I populations, and respectively for I cells, such that the probability of a firing pattern occurring is

$$P(\boldsymbol{\sigma}) = \frac{1}{Z}\exp\Big(\sum_{i=1}^{N}(h^E_{iK^E} + h^I_{iK^I})\sigma_i\Big), \tag{8}$$

The model is said *solvable* as the normalisation Z can be expressed analytically. Note that the model is invariant under several gauge transformations as a number of linear combinations of its parameters $h^E_{iK^E}$ and $h^I_{iK^I}$ do not affect the probability distribution. Z and its derivative allow us to determine the statistics of the model, such as the mean firing rate and pairwise covariances.

Normalisation. From Eq. (8) the normalisation is defined as

$$Z = \sum_{\boldsymbol{\sigma}} \exp\Big(\sum_{i=1}^{N}(h^E_{iK^E} + h^I_{iK^I})\sigma_i\Big) \tag{9}$$

where we sum over all possible firing patterns $\boldsymbol{\sigma}$. We may decompose this sum into terms Z_{k^E,k^I} with given E and I population activities, such that

$$Z = \sum_{k^E=0}^{N^E}\sum_{k^I=0}^{N^I} Z_{k^E,k^I} \tag{10}$$

Then, we have

$$Z_{k^E,k^I} = \sum_{\substack{\boldsymbol{\sigma}\\ K^E=k^E,K^I=k^I}} \exp\Big(\sum_{i=1}^{N}(h^E_{iK^E} + h^I_{iK^I})\sigma_i\Big) \tag{11}$$

where we sum over all possible firing patterns for all neurons for which K^E excitatory neurons active and K^I inhibitory neurons active. This is equivalent to summing over all possible patterns of E and I neurons independently, i.e.:

$$\begin{aligned} Z_{k^E,k^I} &= \sum_{i^E_1<\ldots<i^E_{k^E}}\ \sum_{i^I_1<\ldots<i^I_{k^I}} \exp\Big(\sum_{b=1}^{k^E} h^E_{i^E_b k^E} + h^I_{i^E_b k^I}\Big)\exp\Big(\sum_{c=1}^{k^I} h^E_{j^I_c k^E} + h^I_{j^I_c k^I}\Big) \\ &= \Big[\sum_{i^E_1<\ldots<i^E_{k^E}} \exp\Big(\sum_{b=1}^{k^E} h^E_{i^E_b k^E} + h^I_{i^E_b k^I}\Big)\Big]\Big[\sum_{i^I_1<\ldots<i^I_{k^I}} \exp\Big(\sum_{c=1}^{k^I} h^E_{i^I_c k^E} + h^I_{i^I_c k^I}\Big)\Big], \end{aligned} \tag{12}$$

where the i^E_b spans over all the active E neurons for a given E activation pattern, and respectively for the j^I_c for active I neurons. The result may be written as a product of two terms as these terms share no parameters in common.

Here, the first term is summed over all possible firing patterns for E neurons that yield $K^E = k^E$, and similarly for I neurons in the second term. Now, analogously to [5] let Q be a polynomial such that the products over all i

$$Q_E(X) = \prod_{\substack{i=1\\ i\in E}}^{N^E} 1 + X\exp(h^E_{ik^E} + h^I_{ik^I}), \tag{13}$$

where we take the product over all i excitatory neurons, and similarly for $Q_I(X)$ multiplying over all inhibitory neurons. Now, the coefficient of Q_E of order X^{k^E}, denoted $\text{Coeff}[Q_E, X^{k^E}]$, corresponds to the sum over all the products of k^E terms of the form $\exp(h^E_{ik^E})$,

or in other words, the sum over all products of combinations of E neurons $i_1^E < ... < i_{k(E)}^E$, which is exactly equivalent to the first term of equation 12. Since the same obviously applies for I neurons, we have

$$Z_{k^E,k^I} = \text{Coeff}[Q_E, X^{k^E}] \times \text{Coeff}[Q_I, X^{k^I}], \tag{14}$$

As the Q coefficients can be recursively computed, Z is analytically computable, thus the model is solvable. Next we derive the statistics of the model from Z.

Model statistics The statistics predicted by the model is given by differentiating Z. We use this to predict the mean firing rates and pairwise covariances from the population couplings $h^E_{ik^E}$ and $h^I_{ik^I}$.

Firing rates. As we defined in Eq.s (9) and (10),

$$Z = \sum_{k^E=0}^{N^E} \sum_{k^I=0}^{N^I} \exp\Big(\sum_{i=1}^{N} (h^E_{ik^E}\delta^{K^E}_{k^E} + h^I_{ik^I}\delta^{K^I}_{k^I})\sigma_i\Big), \tag{15}$$

Thus, the joint probability of a given neuron spiking and a given number of E neurons spiking in any time bin is as follows:

$$P(\sigma_i = 1, k^E = K^E) =< \sigma_i \delta^{K^E}_{k^E} >= \frac{\partial lnZ}{\partial h^E_{iK^E}} \tag{16}$$

Recalling our expression for Z_{k^E,k^I} in Eq. (14), this yields

$$P(\sigma_i = 1, k^E = K^E) = \frac{1}{Z}\sum_{k^I}\Big(\text{Coeff}[Q'_E, X^{K^E}] \times \text{Coeff}[Q_I, X^{k^I}]\Big) \tag{17}$$

where

$$Q'_E \equiv \frac{\partial lnZ}{\partial h^E_{iK^E}}\text{Coeff}[Q_E, X^{k^E}] = Xe^{h^E_{iK^E}+h^I_{ik^I}} \prod_{j\neq i}(1 + Xe^{h^E_{jK^E}+h^I_{jk^I}}) \tag{18}$$

and *i is an E neuron.* For an I neuron's firing probability one can swap around E and I in Eq. (17). This allows the straightforward derivation of the firing rate by summing over all values of K^E (resp. K^I for I neurons).

Pairwise correlations. Likewise, the pairwise correlations may be computed from

$$< \sigma_{i_E}\sigma_{j_I} >= \frac{1}{Z}\sum_{k^E}\sum_{k^I}\Big(\text{Coeff}[Q'_E, X^{k^E}] \times \text{Coeff}[Q'_I, X^{k^I}]\Big) \tag{19}$$

for two neurons of different types, and

$$< \sigma_{i_E}\sigma_{j_E} >= \frac{1}{Z}\sum_{k^E}\sum_{k^I}\Big(\text{Coeff}[Q''_E, X^{k^E}] \times \text{Coeff}[Q_I, X^{k^I}]\Big) \tag{20}$$

where

$$Q''_E = X^2 e^{h^E_{ik^E}+h^I_{ik^I}+h^E_{jk^E}+h^I_{jk^I}} \prod_{l\neq i,j}(1 + Xe^{h^E_{lk^E}+h^I_{lk^I}}) \tag{21}$$

and i is an E neuron (extending to I neurons is very straightforward).

Shuffle tests. To verify whether information on cell types significantly improves the model's prediction performance, for each species, we perform a series of ten inferences on the same SWS data-set. Each time, the cell labels are independently shuffled, while the number of E cells and the number of I cells remains the same. The Mean Square Error (MSE) on the predicted pairwise covariances is computed every time. We found it to be consistently larger for the shuffled trials compared to that where empirical cell types are known. This is quantified by the Mann-Whitney U test on the samples of $\text{MSE}_{\text{shuffled}}[n] - \text{MSE}_{\text{data}}$ with n ranging from one to ten.

---

[1] S. Cocco and R. Monasson. *Adaptive cluster expansion for the inverse Ising problem: convergence, algorithm and tests.* J. Stat. Phys. , **147**:252, 2012.

[2] N. Dehghani, A. Peyrache, B. Telenczuk, M. Le Van Quyen, E. Halgren, S.S. Cash, N.G. Hatsopoulos, and A. Destexhe. Dynamic balance of excitation and inhibition in human and monkey neocortex. *Scientific Reports*, 6(23176), 2016.

[3] U. Ferrari. Learning maximum entropy models from finite-size data sets: A fast data-driven algorithm allows sampling from the posterior distribution. *Phys. Rev. E*, 94:023301, 2016.

[4] U. Ferrari, T. Obuchi, and T. Mora. Random versus maximum entropy models of neural population activity. *Physical Review E*, 95(4):042321, 2017.

[5] C. Gardella, O. Marre, and T. Mora. A tractable method for describing complex couplings between neurons and population rate. *eneuro*, 3(4):0160, 2016.

[6] L. S. Hamilton, J. Sohl-Dickstein, A. G. Huth, V. M. Carels, K. Deisseroth, and S. Bao. *Optogenetic Activation of an Inhibitory Network Enhances Feedforward Functional Connectivity in Auditory Cortex.* Neuron , **80**:1066–76, 2013.

[7] E. T. Jaynes. *On The Rationale of Maximum-Entropy Method.* Proc. IEEE , **70**:939, 1982.

[8] M. Le Van Quyen, L. E. Muller, B. Telenczuk, E. Halgren, S. Cash, N. G. Hatsopoulos, N. Dehghani, and A. Destexhe. High-frequency oscillations in human and monkey neocortex during the wake–sleep cycle. *Proceedings of the National Academy of Sciences*, 113(33):9363–9368, 2016.

[9] O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Phys. Rev. Lett.*, 102:138101, 2009.

[10] T.-A. Nghiem, O. Marre, A. Destexhe, and U. Ferrari. Pairwise ising model analysis of human cortical neuron recordings. In *International Conference on Geometric Science of Information*, pages 257–264. Springer, 2017.

[11] M. Okun, N. A. Steinmetz, L. Cossell, M. F. Iacaruso, H. Ko, P. Barthó, T. Moore, S. B. Hofer, T. D. Mrsic-Flogel, M. Carandini, et al. Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515, 2015.

[12] A. Peyrache, N. Dehghani, E.N. Eskandar, J.R. Madsen, W.S. Anderson, J.A. Donoghue, L.R. Hochberg, E. Halgren, S.S. Cash, and A. Destexhe. Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. *Proceedings of the National Academy of Sciences*, 109(5):1731–1736, 2012.

[13] A. Peyrache, M. Khamassi, K. Benchenane, S.I. Wiener, and F.P. Battaglia. *Replay of rule-learning related neural patterns in the prefrontal cortex during sleep* . Nat. Neurosci. , **12** :919–26, 2009.

[14] E. Schneidman, M. Berry, R. Segev, and W. Bialek. *Weak pairwise correlations imply strongly correlated network states in a population* . Nature , **440**:1007, 2006.

[15] M Steriade, D Contreras, R Curró Dossi, and A Nunez. The slow (¡ 1 hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *Journal of Neuroscience*, 13(8):3284–3299, 1993.

[16] G. Tavoni, U. Ferrari, F. P. Battaglia, S. Cocco, and R. Monasson. Functional coupling networks inferred from prefrontal cortex activity show experience-related effective plasticity. *Network Neuroscience*, 1(3):275–301, 2017.

[17] B. Teleńczuk, N. Dehghani, M. Le Van Quyen, S. S. Cash, E. Halgren, N. G. Hatsopoulos, and A. Destexhe. Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *Scientific reports*, 7:40211, 2017.

[18] G. Tkacik, O. Marre, D. Amodei, E. Schneidman, W Bialek, and M.J. Berry. *Searching for collective behaviour in a network of real neurons* . PloS Comput. Biol., **10(1)**:e1003408, 2014.

Supplementary figures

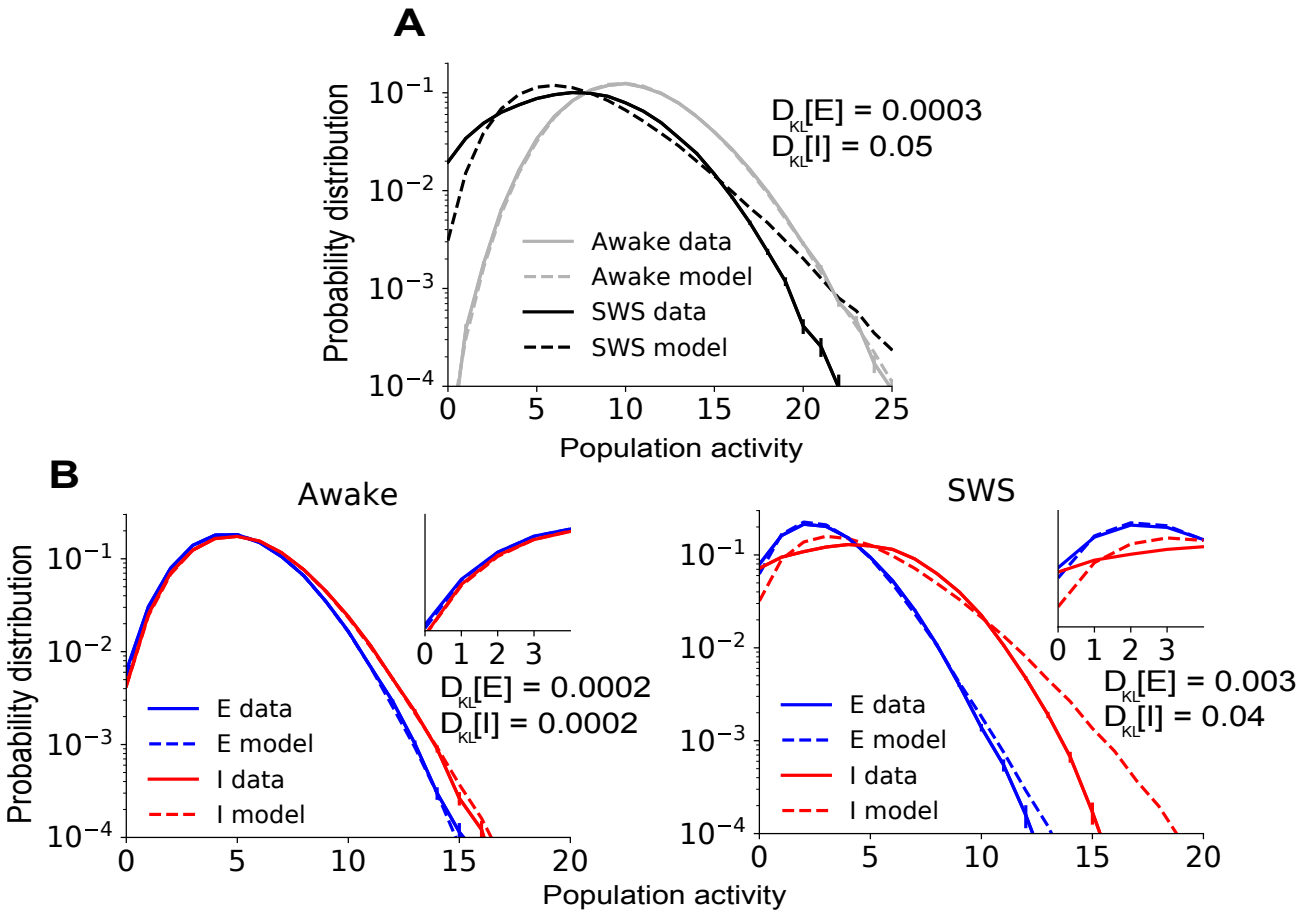

FIG. S1: **Pairwise Ising model analysis on monkey recording.** **A**) Empirical and predicted distributions of the population activity $K = \sum_i \sigma_i$ for the population of excitatory (E) cells, and that of inhibitory (I) cells. On the monkey data, the Ising also performs better at capturing the population statistics during Awake than SWS. **B**) Empirical and predicted population activities for E and I cells. The model fails at reproducing the statistics of inhibitory population activity during SWS, similarly to in the human data.

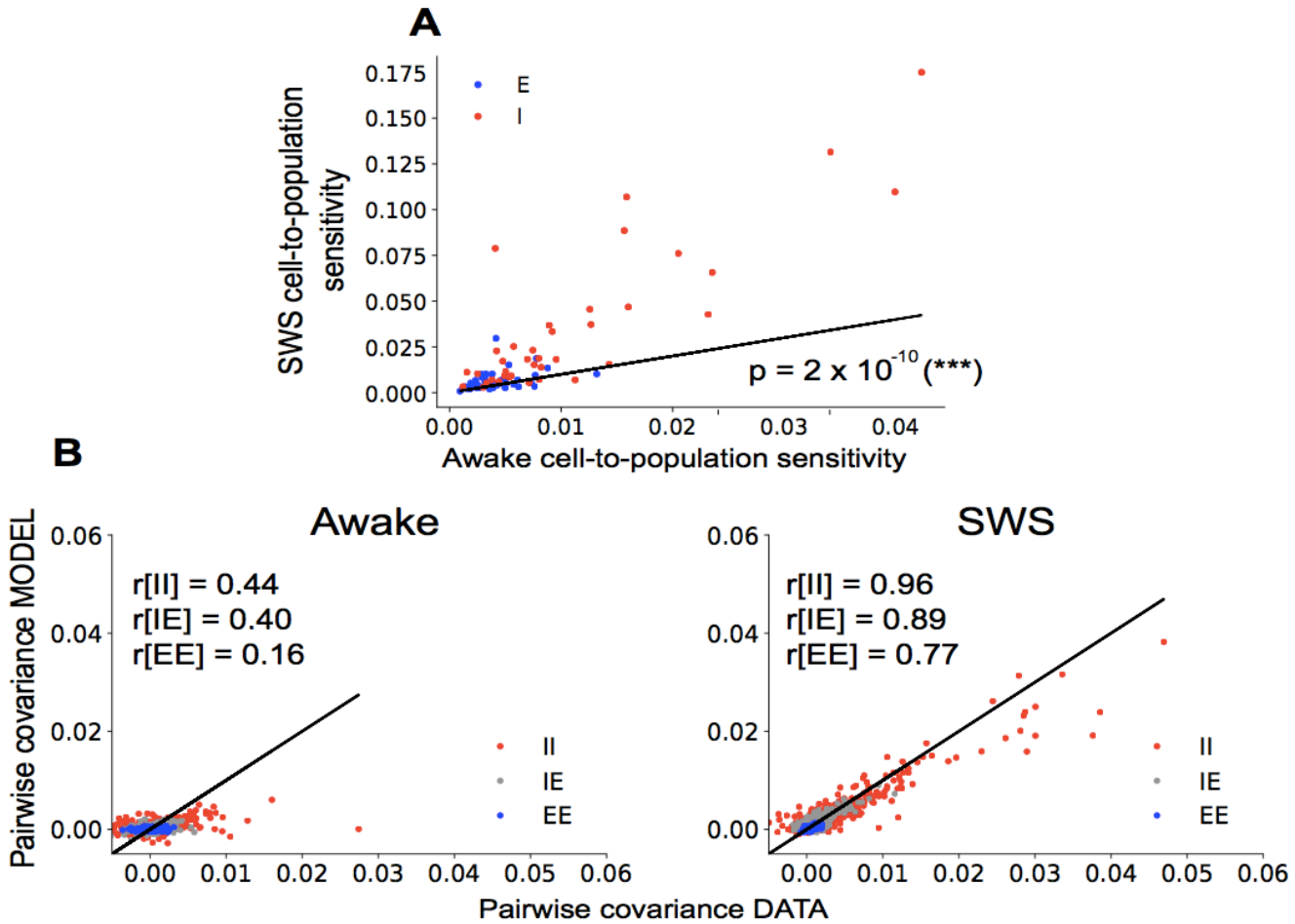

FIG. S2: **Single-population model analysis on monkey recording.** **A**) Scatter-plot of the cell sensitivity to the population activity. Cells are consistently more sensitive during SWS (p-value < 0.001). **B**) Pairwise covariances, empirical against predicted, for awake (left) and SWS (right) states. Relative success for SWS, especially I-I pairs suggests these cells are most responsive to whole-population activity, even though the model tends to under-estimate the larger pairwise covariances. The model completely fails to account for pairwise covariances during Awake.

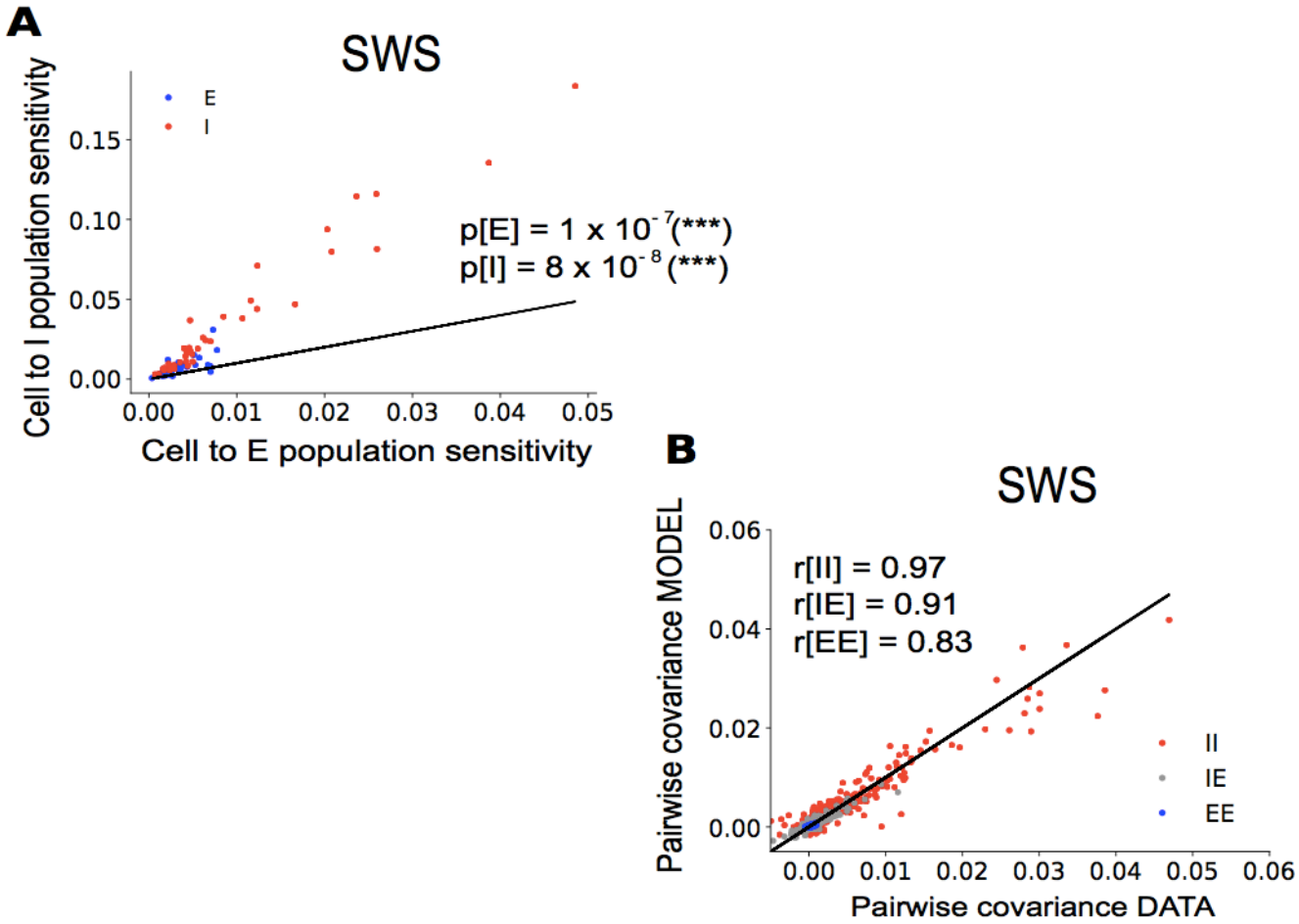

FIG. S3: **Two-population model shows significant improvement in prediction for all types of cells. A**) Scatter-plot of cell sensitivity to E versus I population, during SWS. Both I and E cells are more tuned to I population than the E one (p-value $< 10^{-3}$, Wilcoxon sign-ranked test). **B**) Pairwise covariances, empirical against predicted, for the two-population model, during SWS. Improvement compared to the whole-population model is confirmed by the Pearson correlations.