Discovery of Salient Low-Dimensional Dynamical Structure in Neuronal Population Activity Using Hopfield Networks

Felix Effenberger[1](✉) and Christopher Hillar[2]

[1] Max-Planck-Institute for Mathematics in the Sciences,
Inselstr. 22, 04103 Leipzig, Germany
`felix.effenberger@mis.mpg.de`
[2] Redwood Center for Theoretical Neuroscience,
Berkeley, CA 94720, USA
`chillar@berkeley.edu`

Abstract. We present here a novel method for the classical task of finding and extracting recurring spatiotemporal patterns in recorded spiking activity of neuronal populations. In contrast to previously proposed methods it does not seek to classify exactly recurring patterns, but rather approximate versions possibly differing by a certain number of missed, shifted or excess spikes. We achieve this by fitting large Hopfield networks to windowed, binned spiking activity in an unsupervised way using minimum probability flow parameter estimation and then collect Hopfield memories over the raw data. This procedure results in a drastic reduction of pattern counts and can be exploited to identify prominently recurring spatiotemporal patterns. Modeling furthermore the sequence of occurring Hopfield memories over the original data as a Markov process, we are able to extract low-dimensional representations of neural population activity on longer time scales. We demonstrate the approach on a data set obtained in rat barrel cortex and show that it is able to extract a remarkably low-dimensional, yet accurate representation of population activity observed during the experiment.

Keywords: Neuronal population activity · Parallel spike train analysis · Spatiotemporal patterns · Hopfield network · Ising model

1 Introduction

Finding recurring spatiotemporal patterns (STP) in recorded spiking activity of neuronal populations is a classical problem in the data analysis of parallel spike trains, and quite a number of approaches to detect and classify recurring spatiotemporal patterns of neural population activity have been proposed [3,6]. Yet, most published methods so far either focus solely on synchrony detection [15,16,18] or assume a more or less noiseless scenario, seeking to classify exactly recurring STP in neuronal activity (apart from allowing some jitter in spike timing), see e.g. [5].

A. Feragen et al. (Eds.): SIMBAD 2015, LNCS 9370, pp. 199–208, 2015.
DOI: 10.1007/978-3-319-24261-3_16

Given the usually high variability of population responses to stimuli, the re-occurrence of such exactly repeating STP becomes more and more unlikely with increasing population size though. Despite this variability, there is strong experimental evidence that neural populations code information about stimuli in some form of STP, see e.g. [1,2]. Thus, a much more plausible situation is that some underlying STP appears in several "corrupted" variants, both expressing jitter in spike times and differing in a few missing or excess spikes. To find and classify recurring STP in parallel spike trains, we fit Hopfield networks (HN) to windowed, binned spiking activity of a population of cells using minimum probability flow [19] (MPF), a novel probabilistic learning rule for HN with many desirable properties [8,10]. We then use Hopfield network dynamics to classify the raw data and identify recurring STP. The presented method is robust to the aforementioned variability in the signal and able to extract the underlying recurring patterns, even for seldom occurring STP and large population sizes.

Modeling furthermore the sequence of occurring Hopfield memories as a Markov process, we are able to extract low-dimensional representations of neural population activity. We demonstrate the approach on a data set obtained from rat barrel cortex [14] and show that it is able to extract a remarkably low-dimensional, yet accurate representation of the average population response to whisker stimulation.

The paper is organized as follows. In Sect. 2 we give a short overview of the theoretical background, namely Hopfield networks and minimum probability flow parameter estimation. We then present our method in Sect. 3, followed by a demonstration of the method in Sect. 4. We conclude in Sect. 5.

2 Background

Hopfield networks [11] are a well-known model of memory and collective processing in networks of abstract McCulloch-Pitts [13] neurons.

The possible *states* of a HN are the same as those of a non-ferromagnetic Ising model, a classical model in statistical physics [12]. This discrete probability distribution has as states all binary vectors of length n, with the probability of a particular state $\mathbf{x} = (x_1, \ldots, x_n) \in \{0,1\}^n$ being

$$p_{\mathbf{x}} = \frac{1}{Z} \exp\left(\sum_{i<j} J_{ij} x_i x_j - \sum_i \theta_i x_i \right) = \frac{1}{Z} \exp\left(-E_{\mathbf{x}}\right), \tag{1}$$

in which $J \in \mathbb{R}^{n \times n}$ is a real symmetric matrix with zero diagonal (the *coupling matrix*), the vector $\theta \in \mathbb{R}^n$ is a bias or *threshold* term, and $Z = \sum_{\mathbf{x}} \exp(-E_{\mathbf{x}})$ is the *partition function* (which normalizes $\mathbf{p} = (p_{\mathbf{x}})_{\mathbf{x} \in \{0,1\}^n}$ to sum to 1). Typically, the expression inside the exponential of (1) is viewed as the negative of a quadratic energy function,

$$E_{\mathbf{x}} = -\frac{1}{2} \mathbf{x}^\top J \mathbf{x} + \theta^\top \mathbf{x}. \tag{2}$$

Thus, states $\mathbf{x}$ with low energy (2) appear most often under sampling from (1). It follows from basic theory (e.g. [4]) that the distribution defined by (1) is the maximum entropy distribution on binary vectors given its first and second order statistics (mean and covariance).

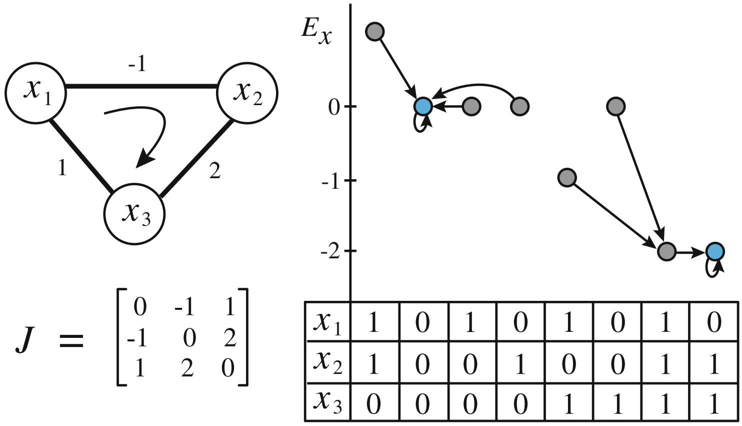
A HN is a recurrent network of binary nodes (representing spiking neurons) with deterministic dynamics. Formally, a HN on n nodes $\{1, \ldots, n\}$ consists of a symmetric coupling matrix $J \in \mathbb{R}^{n \times n}$ and a threshold vector $\theta \in \mathbb{R}^n$.

An *asynchronous dynamics update* of state $\mathbf{x}$ in a HN consists of iteratively replacing each x_i in $\mathbf{x}$ with a new value

$$x_i = \begin{cases} 1 & \text{if } \sum_{j \neq i} J_{ij} x_j > \theta_i, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The update given by Eq. (3) is inspired by computations exhibited in neurons [13] and a model neuron with such an update rule is often called a *McCulloch-Pitts neuron*. A fundamental property of HNs is that an asynchronous dynamics update given by Eq. (3) does not increase the energy given by Eq. (2). Thus, after a finite number of updates, each initial state $\mathbf{x}$ converges to a *fixed-point* $\mathbf{x}^*$ (also called *stable-point* or *memory*) of the dynamics. Intuitively, we may interpret the dynamics as an inference technique, producing the most probable nearby memory given a noisy version. See Fig. 1 for an example of a small HN and its energy landscape.

A basic problem is to construct Hopfield networks with a given set $\mathcal{D}$ of binary patterns as memories (i.e. local minima of $E_{\mathbf{x}}$). Such networks are useful



$$J = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$$

x_1	1	0	1	0	1	0	1	0
x_2	1	0	0	1	0	0	1	1
x_3	0	0	0	0	1	1	1	1

Fig. 1. Small Hopfield network. A 3-node Hopfield network with coupling matrix J and zero threshold vector θ. A state vector $\mathbf{x} = (x_1, x_2, x_3)$ has energy $E_{\mathbf{x}}$ as labeled on the y-axis of the diagram. Arrows represent one iteration of the network dynamics; i.e. x_1, x_2, and x_3 are updated by Eq. (3) in the order of the clockwise arrow. Resulting fixed-points $\mathbf{x}^*$ are indicated by blue circles (Color figure online).

for memory denoising and retrieval since corrupted versions of patterns in $\mathcal{D}$ will converge through the dynamics to the originals.

In contrast to traditional rules used for this task such as the *outer-product learning rule* [11] (OPR) and the *perceptron learning rule* [17] (PER) that face a number of limitations such as low memory capacity, bad generalization properties and high computational cost, we here use *minimum probability flow* [19] (MPF) to estimate the parameters of a Hopfield network. Applied to estimating the parameters in an Ising model/Hopfield network, Eq. (1), the minimum probability flow objective function [8,19] is:

$$K_{\mathcal{D}}(J,\theta) = \sum_{\mathbf{x}\in\mathcal{D}} \sum_{\mathbf{x}'\in\mathcal{N}(\mathbf{x})} \exp\left(\frac{E_{\mathbf{x}} - E_{\mathbf{x}'}}{2}\right). \tag{4}$$

Here, the *neighborhood* $\mathcal{N}(\mathbf{x})$ of $\mathbf{x}$ consists of those binary vectors which are Hamming distance 1 away from $\mathbf{x}$. The function in (4) is infinitely differentiable, jointly convex in the parameters, consists of only order $O(|\mathcal{D}|n)$ terms, and can be minimized using standard methods such as gradient descent. Notice also that when $K_{\mathcal{D}}$ is small, the energy differences $E_{\mathbf{x}} - E_{\mathbf{x}'}$ between points $\mathbf{x}$ in the dataset $\mathcal{D}$ and patterns $\mathbf{x}'$ in single-flip neighborhoods $\mathcal{N}(\mathbf{x})$ will be negative, making $\mathbf{x}$ a fixed-point of the Hopfield dynamics. Importantly to applications, much more is true: minimizing (4) given a storable set of patterns $\mathcal{D}$ will determine a Hopfield network storing those patterns as robust memories [8]. Moreover, the MPF objective function can naturally be turned into an online, neurologically plausible learning rule [9].

3 Our Method

The training data X are obtained by sliding a window of given length L over a binary matrix of dimension $N \times T$ representing the binned spiking activity of N cells over a time period of T bins, yielding $T - L$ binary vectors of length NL as training data, see Fig. 2.

After fitting a HN with NL nodes on the data using MPF, we converge each window of the raw, binned spiking data to its Hopfield memory.

We label the sequence of occurring memories by natural numbers in the order of their appearance so that we obtain a memory sequence $S = (s_1, \ldots, s_{T-L})$, with $s_i \in \{m_1, \ldots, m_k\} = M$, $k \leq T - L$, where M denotes the set of all distinct memories in S.

Note that usually $k \ll |S|$, as STP occurring in the raw data that have low Hamming distances are likely to converge to the same memory under the Hopfield dynamics as a result of fitting the network with MPF.

For each memory $m_i \in M$ we compute all pairwise one-step Markov transition probabilities to $m_j \in M$ $(1 \leq j \leq k)$ using data from S and the entropy over this probability distribution for each m_i, which we call the entropy of the memory m_i and denote by $H(m_i)$.

The entropy of a memory is a measure for how predictable the following network state is, according to the observed data. Memories with a more restricted

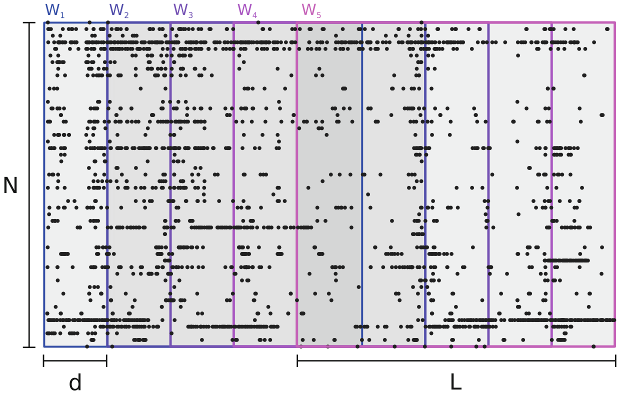


Fig. 2. Windowing of binned neural activity. To generate training data for the Hopfield network, spiking data of N cells are first binned and then training vectors are extracted using a sliding window. Windows of length L are shifted by d bins (here, we take d = 1) resulting in training vectors of dimension n = NL. The above illustration shows five overlapping windows.

set of following network states have lower entropy, ones with less predictable states have higher entropy. $H(m_i)$ can therefore be seen as a local measure (in time) for how deterministic the network dynamics evolve from that memory.

We then construct a directed graph with the elements of M as nodes. Two nodes m_i, m_j are connected by an edge (m_i, m_j) of weight w if their Markov transition probability $w = P(m_j|m_i)$ obtained from S is non-zero. We call this graph the *Markov graph* G_M of S. Paths and cycles (i.e. simple closed paths) in G_M along nodes with low entropy correspond to sequences of memory labels and thus sequences of STP of spiking activity that are prominently and reliably generated by the neuronal population.

4 Application to Data

We applied the proposed method to spiking data of recorded in the rat barrel cortex during repeated whisker stimulation [14] ($N = 16, T = 3.4 \cdot 10^4, L = 10$, 1 ms bins), see Fig. 3 for a raster plot of 50 trials of the experiment. For each trial, recorded spiking activity measured in 16 electrodes (multi unit activity, MUA) is shown. Whisker stimulation is performed at 1000 ms within each trial.

The number of different 160-dimensional patterns (corresponding to 10 ms of network activity) in the raw data is 161,171. After fitting a Hopfield network to the raw data and collecting the memories over the input data we obtain 577 distinct Hopfield memories, a 280-fold reduction in count, see also Fig. 4. The Markov transition probability matrix for the 577 memories is shown in Fig. 5, their probabilities and entropies in Fig. 6.

To ease data analysis, we further restrict the number of memories considered to the 50 memories of highest rank from this point on. The Markov Graph

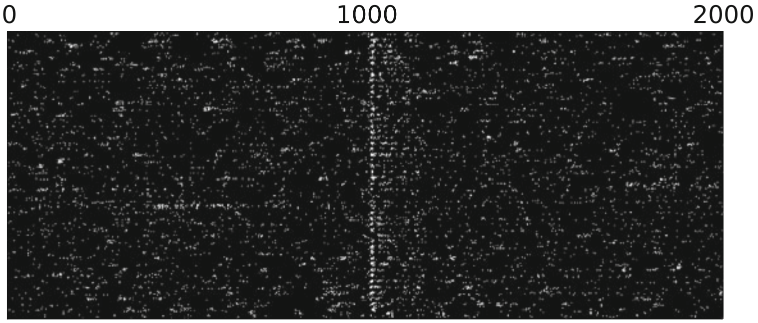


Fig. 3. Raster plots of 16 cells over 50 trials. Binned into 1 ms bins, stimulation of primary whisker at 1000 ms. White dots denote spiking activity. Horizontal axis shows time, vertical axis units/trials.

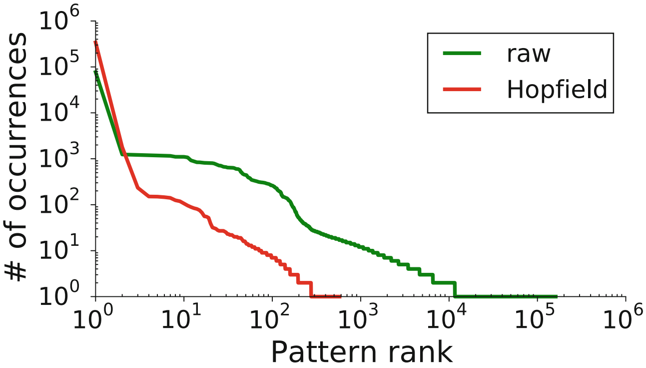


Fig. 4. Ranks of raw and Hopfield patterns of 10 ms length.

G_M is pruned accordingly, but node entropies are calculated on the full set of memories. To each Hopfield memory we associate memory triggered averages (MTAs, computed as the average of all raw patterns converging to the given Hopfield memory); these are shown in Fig. 7.

For this data set we find a 'central' node m_α (corresponding to memory label 1) in the Markov Graph G_M that has a high degree (sum of in- and out-degrees). This is characteristic for a situation in which the node is the termination (resp. starting) point of prominently occurring STP of network activity. Interestingly, the memory m_α occurs very frequently in the data ($p > 0.9$) and the node has low entropy. This we expect from a network's low-energy base state that it prominently rests in and repeatedly returns to. Using the information of the stimulus protocol, we indeed found that m_α corresponds to the resting state of the network (see top left MTA in Fig. 7).

We now look at cycles (i.e. closed, simple paths) in the Markov Graph G_M starting and terminating in some given node. We expect that cycles in G_M starting in m_α can give insight on how the network is driven out of its resting

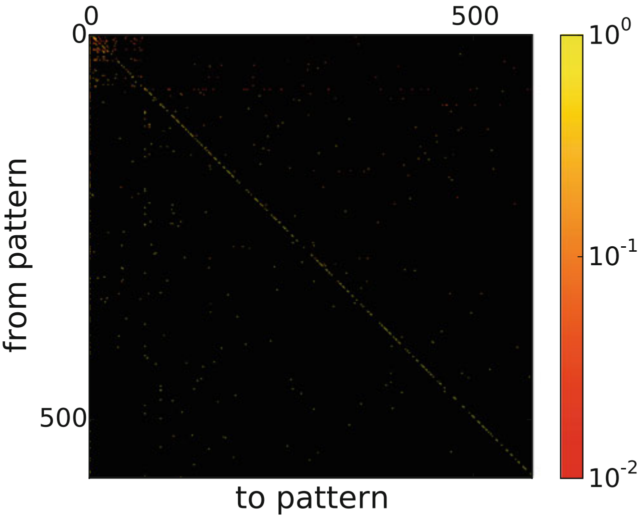


Fig. 5. Markov transition probabilities of all Hopfield memories (577) observed in the raw data.

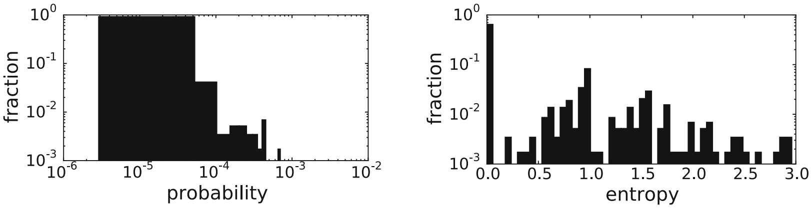


Fig. 6. Occurrence probabilities and entropies of Hopfield memories. Left: Occurrence probabilities of memories observed in raw data (cut at 10^{-2}). Right: Entropies of memories (calculated for each memory from its Markov transition probabilities) observed in raw data.

state by some stimulus and enters a transient sequence of excited states before falling back to the resting state. See Fig. 8 for the distribution of cycle lengths in the restricted Markov graph. Note that windows overlap by 9 ms in the present case, making the approximation of longer time-scale network dynamics via a first order Markov process (as paths in the Markov Graph) rather robust.

Tracing such cycles in G_M (and scoring them by their entropy, obtained as a weighted sum of the entropies of the cycle's nodes as a measure for how reliably that cycle is "visited" by the network dynamics), we find that the most STP associated with low entropy cycles indeed correspond closely to the average network response to whisker stimulation (that we computed from the raw data using knowledge of the simulation protocol), see Fig. 9. Note that our method was able to reconstruct the average network response without any knowledge of the stimulus protocol.

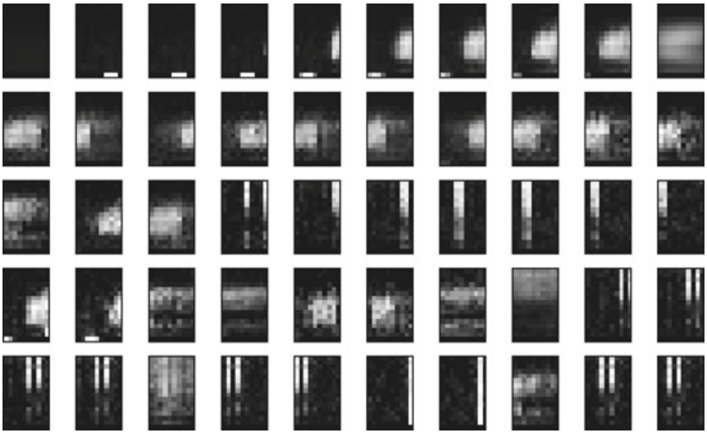


Fig. 7. Memory triggered averages. Memory triggered averages of 50 memories with highest rank observed in the raw data (ordered by decreasing rank, top left to bottom right). Each plot shows one MTA encoding a prominent STP of length 10 ms; a white pixel denotes high (1) spiking probability of a given neuron at a given time, a black pixel low spiking probability (0).

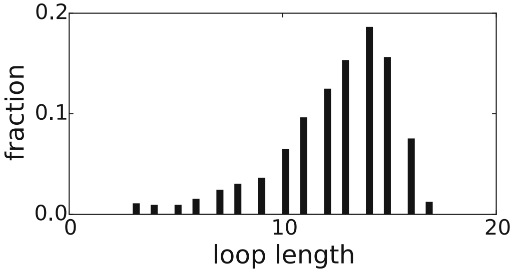


Fig. 8. Distribution of cycle lengths around base node 0 in reduced Markov graph.

Fig. 9. Reconstruction of stable network response. Left: Stimulus triggered average obtained from raw data obtained using knowledge of stimulus protocol. Right: Likely network response reconstructed by our method, corresponding to a low entropy path in the Markov graph, not using any knowledge of the stimulus protocol.

5 Conclusion

We presented here a method for identifying and classifying recurring spatiotemporal patterns in parallel neural spike trains based on Hopfield networks. In contrast to previously proposed methods [5,6,15,16,18], it does not solely focus on (partial) synchrony detection or finding exactly recurring patterns, nor does it face combinatorial explosion in the number of neurons or time steps considered.

The trained Hopfield networks denoise the data, grouping similar patterns together in a way that respects the underlying statistics of the data. They are thus able to identify prominent patterns reoccurring in the dataset, possibly corrupted by noise, and eliminate the large number of spurious patterns occurring rarely. In its memories the network encodes different structural aspects of the spiking data such as prominent temporal firing sequences that usually are very difficult to identify in the raw data or using standard methods. Modeling the sequence of occurring memories as a Markov chain, we have shown that the method is able to extract salient features of parallel spike trains in a fully unsupervised way.

We thus propose the method as a novel tool in mining parallel spike trains for possibly low-dimensional underlying network dynamics. An open source software in form of a Python package [7] allowing for the wider application of the method is currently in beta test and to be released soon.

Acknowledgements. The authors would like to thank Yuri Campbell for helpful comments on an earlier version of this manuscript.

References

1. Abeles, M., Bergman, H.: Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. J. Neurophysiol. **70**(4), 1629–1638 (1993)
2. Arieli, A., Shoham, D.: Coherent spatiotemporal patterns of ongoing activity revealed by real-time optical imaging coupled with single-unit recording in the cat visual cortex. J. Neurophysiol. **73**(5), 2072–2093 (1995)
3. Brown, E.N., Kass, R.E., Mitra, P.P.: Multiple neural spike train data analysis: state-of-the-art and future challenges. Nat. Neurosci. **7**(5), 456–461 (2004)
4. Cover, T., Thomas, J.: Elements of Information Theory, 2nd edn. Wiley, New Jersey (2006)
5. Gansel, K.S., Singer, W.: Detecting multineuronal temporal patterns in parallel spike trains. Front. Neuroinformatics **6**(May), 18 (2012)
6. Grün, S., Rotter, S.: Analysis of Parallel Spike Trains. Springer, Heidelberg (2010)
7. Hillar, C., Effenberger, F.: `hdnet` - hopfield denoising network. https://github.com/team-hdnet/hdnet (2015)
8. Hillar, C., Sohl-Dickstein, J., Koepsell, K.: Efficient and optimal little-hopfield auto-associative memory storage using minimum probability flow. In: NIPS Workshop on Discrete Optimization in Machine Learning (DISCML) (2012)
9. Hillar, C., Sohl-Dickstein, J., Koepsell, K.: Novel local learning rule for neural adaptation fits Hopfield memory networks efficiently and optimally. BMC Neurosci. **14**(Suppl 1), P215 (2013)

10. Hillar, C., Tran, N.: Robust exponential memory in Hopfield networks. arXiv e-prints (2014)
11. Hopfield, J.: Neural networks and physical systems with emergent collective computational abilities. Proc. Nat. Acad. Sci. **79**(8), 2554–2558 (1982)
12. Ising, E.: Beitrag zur Theorie des Ferromagnetismus. Zeitschrift fur Physik **31**, 253–258 (1925)
13. McCulloch, W., Pitts, W.: A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biol. **5**(4), 115–133 (1943)
14. Minlebaev, M., Colonnese, M., Tsintsadze, T., Sirota, A., Khazipov, R.: Early gamma oscillations synchronize developing thalamus and cortex. Science **334**(6053), 226–229 (2011)
15. Picado-Muiño, D., Borgelt, C., Berger, D., Gerstein, G., Grün, S.: Finding neural assemblies with frequent item set mining. Front. Neuroinformatics **7**(May), 9 (2013)
16. Pipa, G., Wheeler, D.W., Singer, W., Nikolić, D.: NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. J. Comput. Neurosci. **25**(1), 64–88 (2008)
17. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. **65**(6), 386 (1958)
18. Santos, L.V., Ribeiro, S., Tort, A.B.L.: Detecting cell assemblies in large neuronal populations. J. Neurosci. Methods **220**(2), 149–166 (2013)
19. Sohl-Dickstein, J., Battaglino, P., DeWeese, M.: New method for parameter estimation in probabilistic models: minimum probability flow. Phys. Rev. Lett. **107**(22), 220601 (2011)