The effects of sequence learning on the correlated activity in the Hierarchical Temporal Memory

Victor Hernandez-Urbina
vhernandezurbina@numenta.com

November 4, 2016

Abstract

Correlated activity in groups of neurons has been reported in many studies and in different experimental protocols in practically all areas of neuroscience. Although its use has proven to be useful to reveal functional connectivity among different brain areas, its role in neural information processing is still under much debate. In the present study, we use a biologically plausible neural network model to investigate the nature of correlated activity in the context of sequential learning. We are able to replicate some observations from experimental neuroscience, and to present some predictions regarding the effect of learning on the structure of neural correlations.

1 Introduction

Correlated activity of population of neurons provides relevant information regarding the functional connectivity of brain networks. The observation of correlated activity is the first step towards revealing a purported causal relationship between two different brain areas. For correlated activity to emerge in a single region, there has to be correlated activity within the region. The study of correlated activity at the level of single cells poses enormous experimental challenges particularly when trying to correlate the observed correlation patterns to the stimulus being presented to a sensory organ due to the noisy nature of neurons.

Moreover, studying brain correlated activity imposes computational challenges as it requires the analysis of large amounts of data from simultaneously recorded neurons. This is the main reason why most studies of correlations at the level of individual neurons are primarily focused on pairwise cell activity.

The correlated activity within a group of neurons might provide some insights regarding the encoding mechanisms of the neural substrate. Information might be encoded both in the individual activity of neurons, as well as in the correlations among their activities [6]. Thus, the study of neural correlations is relevant not only to reveal functional connectivity in the brain without recurring to invasive methods but also to decipher the neural code.

Neural correlations have been measured in many cortical regions in different experimental conditions over the past two decades. Observations vary across studies. However, most of them agree that pairwise correlations are typically small and positive. Nevertheless, small and perhaps undetectable correlation in pairs of neurons can have a large effect at the population level [1]. That is, even if we observe low correlated activity when considering the spiking activity of only two cells, the network might actually exhibit highly correlated states in the form of synchronous activation or quiescence of groups of neurons which carry vital information to the system.

Cohen and Kohn present a summary of relevant studies regarding neuronal correlations in different brain areas, and under different behavioural settings [3]. These authors describe some experimental factors that contribute to the variability in observed measurements. These include measurements performed on few spikes per trial, and the length of the time window in which spikes are counted, among others. Such factors might lead to incorrect interpretations of the correlated, or decorrelated activity, of neuronal populations.

It has been observed that correlations are usually higher for neuron pairs that are closer to each other, and also possess similar tuning properties As well, it has been reported without a satisfactory explanation that neural correlations are weak in the input layers of primary sensory cortex, and that correlations in motor areas seem to be consistently lower than those in sensory cortex.

As mentioned above, most researchers agree that neural correlations between pairs of cells are low. Schneidmann and colleagues show in a study based on neuronal activity in the retina of salamanders exposed to natural video clips that weak pairwise correlations coexist with strong collective behaviour in the response of groups of neurons [13]. Figure 1 shows the distribution of pairwise correlation coefficients obtained from such study. As well, these authors suggest that correlated neural activity provides a neural code with error-correcting properties.

In their study, Schneidmann and colleagues estimate the probability of observing particular binary words which are obtained from raster plots of spike trains from retinal activity compared to the probability of observing

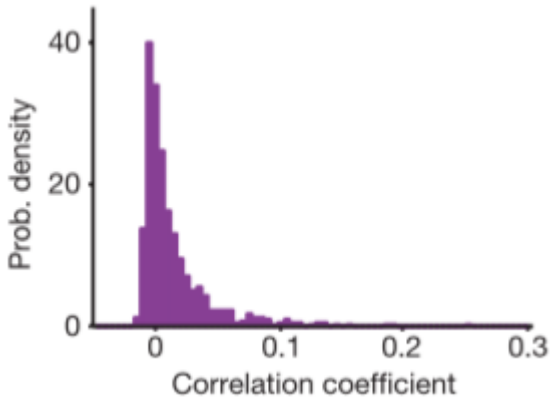
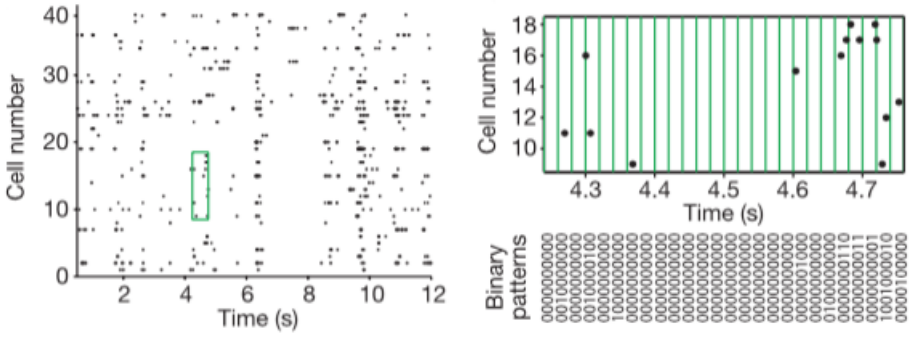


Figure 1: Pairwise correlation coefficients obtained from the activity of retinal ganglion cells of salamander. Correlations between neuron pairs are weak. (Image taken from Ref. [13])

such words in a random model based on Poisson spiking. This process is shown in Fig. 2. A subsample comprising 10 neurons from the spiking activity of retinal ganglion cells is considered along with a time window Δt in which the occurrence of action potentials is measured for all cells in the sample. This would yield binary patterns or words of length 10, if for a given time t the i-th cell fires, then a 1 is placed in the i-th position of the binary pattern, and 0 otherwise.

The researchers reported that although the pairwise neural correlations are low, the probability of observing particular binary words is larger than predicted by the random model. Thus, implying that the collective behaviour of these cells exhibit strong correlated states. This is shown in Fig. 3. The rate of occurrence for some binary patterns is larger than what is predicted by the independence assumption, that is, assuming that the firing of each cell is independent from one another. This implies that although the activity of neuron pairs is weakly correlated, the population exhibits states of activation that occur more often than predicted by chance. Such population states might encode information that is relevant to the system. As a corollary, this study also implies that it is not required to study higher-order correlations. Such results are also confirmed when fitting a maximum entropy model to their data [13, 14].

What are the origins of correlated activity? Authors differ regarding the alleged origin of neural correlations. Some claim that it is the inevitable



(a) Raster plot obtained from spike trains of 40 retinal ganglion cells.

(b) Generation of binary patterns from neural activity.

Figure 2: Binary patterns from the activity of retinal ganglion cells are obtained by subsampling the spike trains from 40 cells and considering a time window Δt in which the occurrence of action potentials is measured. The green box in (a) corresponds to the upper raster plot in (b) If cell i emits a spike within Δt, a 1 is placed at the i-th position in the binary pattern, and 0 otherwise. In this example, the binary patterns consist of 10 bits. (Image taken from Ref. [13])

consequence of shared input in groups of neurons, while others have shown in models how shared input is not a necessary factor for correlations to emerge [12, 4]. Moreover, it has been suggested that correlated activity might in fact be detrimental for neural processing as it reduces the accuracy of neural encoding [12].

However, Ecker and co-workers suggest that such is the case for homogeneous populations of neurons, in which all cells possess the same linear tuning properties [5]. In this scenario, reduced neural correlations induced by attention or by learning and adaptation have been interpreted as evidence for a more efficient population code. In their model, Ecker and colleagues show that increasing the amount of correlations can substantially improve encoding accuracy, when the neural population comprises heterogeneous tuning properties, a feature that is more akin to real brain networks. Such improvement results from a drop in noise entropy which is generally associated with increasing correlated activity. Thus, in contrast to current belief, these authors found that decreasing correlations does not necessarily lead to increased information. If correlations are large enough, increasing them can substantially increase the encoding accuracy [5].

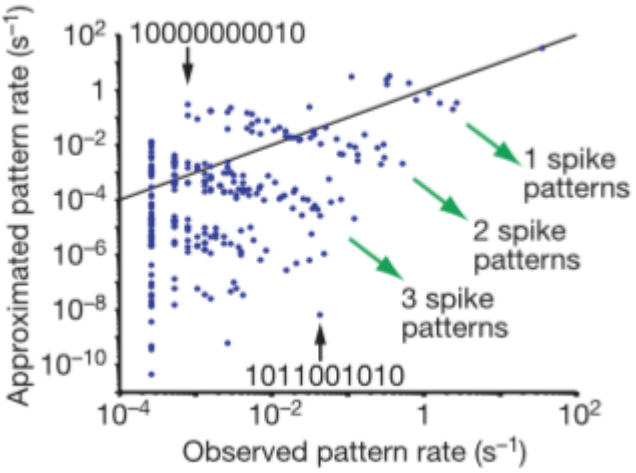


Figure 3: Rate of occurrence of each pattern predicted if all cells are independent (Poisson model) versus the observed measured rate from retinal ganglion cells. Black solid line shows equality. Two examples of extreme misestimation of the predicted pattern rate by the independent model vs. the observed rate are highlighted. (Image taken from Ref. [13])

One question that remains open is what is the role of neural correlations on learning? Or better said, how the process of learning affects the quality of the structure of correlations? Researchers have found that the correlation structure is a target for learning-dependent enhancement of sensory encoding [10]. The study focuses on the differences between the correlations in the tuning curves of two cells, and their correlated activity when presented with a fixed stimulus. The goal of the study is to reveal how these differences impact an organism's behaviour. The authors observed the behavioural responses of birds to particular stimulus motifs which can be thought of elemental auditory sequences. Authors claim that learning selectively alters the structure of correlations, both in the correlation between tuning curves, and in the correlated response of two different neurons for a particular stimulus. Such observation holds for stimuli that are behaviourally relevant Thus, this study concludes that the structure of correlations in a population of neurons carries biologically significant information that is independent of single neurons, and that is crucial to the transformation of purely sensory codes into neural signals that ultimately drive learned behaviour [10].

Neural correlations are significantly smaller in trained versus naive animals, which is thought to result in an improvement on coding efficiency [8] confirming the suggestions made by Ecker and colleagues [5]. In a study based on measuring the responses of neuron pairs in the dorsal medial superior temporal area (MST) of awake monkeys subject to training, authors found that training reduced neural correlations uniformly regardless of the tuning similarity between pairs of neurons. Their study concludes that the impact of correlated activity on population coding depends on the structure of the neural noise and its dependence on the correlations between tuning curves.

Other authors, however, have observed that learning actually increases the correlated activity in a population of neurons. In an experiment set in the context of Bayesian models of perception, Berkes and colleagues measured the population activity within the visual cortex of awake ferrets in response to natural and artificial scenes [2]. In their protocol, spontaneous neural activity and evoked neural activity of the visual cortex are associated to the prior and posterior distributions respectively, in a Bayesian framework. These authors reported an improved match between mean evoked activity and spontaneous activity as the visual cortex matures. This implies that the animal's internal model of visual perception adapts appropriately to statistical features of natural scenes as the organism ages. This observation is also associated with an increase in correlated neural activity opposite to current belief in which neural coding becomes sparse and uncorrelated after learning or stimulus adaptation. Thus, this study reports that the activity of neurons in both evoked and spontaneous activity becomes increasingly correlated and non-sparse with age. In fact, these authors claim that such increased correlations are important for the match between evoked and spontaneous activity [2].

In summary, the presence, origins, and role of correlated activity within a population of neurons is still under debate with authors claiming that its origin should be found in shared neural input, and others opposing such idea. Others claim that correlated activity might be detrimental to neural encoding, while others affirm that in fact correlations are required for encoding precision given the heterogeneous nature of neural cells. When it comes to its relationship with learning and behaviour, some authors have claimed that the process of learning impacts the structure of correlations by decreasing them, which results also in sparser activity within the population. However, other authors have reported the opposite.

The present study is focused on studying the structure of the correlated activity of the temporal memory (TM) of the Hierarchical Temporal Mem-

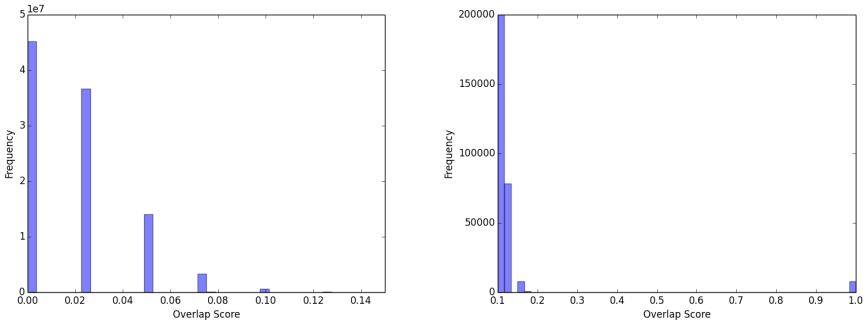
ory (HTM) theory as a result of sequence learning. The TM is an algorithm for on-line learning of time-series and sequential data based on principles of neocortical computation. Its learning mechanisms resemble elemental Hebbian plasticity mechanisms. Moreover, it simulates NMDA spiking to implement prediction of future input incoming to the network in the form of depolarized neuron states. Our analysis gives us the opportunity to find associations between results from experimental neuroscience and HTM theory, as well as to make specific predictions to be tested in experimental setups. Also, our study is motivated by the current debate surrounding the existence of correlated activity in neural networks, and by the possibility of deriving some predictions from computational modelling to experimental neuroscience.

2 Methods

We carry all our experiments in a temporal memory (TM) with 2048 columns, 8 cells per column, and 2% sparsity, which means that at any time only 40 columns become active in response to input. We define four learning scenarios that result from the nature of the sequences used for learning in the TM. In all of these scenarios we keep track of the spike trains generated by each cell in the TM, which will be used to compute the pairwise correlations after learning. We define a very simple measure of accuracy for the TM which works as follows: after presenting a symbol to the TM we keep track of the predicted columns at current time t_i. Then, at time t_{i+1} we consider the active columns as a result of current input. The accuracy is the fraction of columns active at time t_{i+1} that were predicted at time t_i.

The learning scenarios considered are:

1. *Random sequences.* We generated 1000 sequences, each sequence consists of 10 symbols taken at random, that is, by choosing 40 columns at random out of a total of 2048 columns. This results in a sparse distributed representation (SDR) which is required as input for the TM. The process yields 10000 random symbols with very low overlap among them. In this scenario at the beginning of each training epoch, we shuffle the set of sequences and then present each sequence to the TM until all have been processed. We repeat this for N number of epochs ($N = 0$, 5, 10, and 15). A time-step is defined as a presentation of a single symbol from a sequence to the TM. Thus, by the end of simulation time we end up with a total number of time-steps equal to the number of epochs times the number of sequences times the number



(a) Random scenario.

(b) High-order scenario.

Figure 4: Distribution of overlap scores among SDRs in two scenarios: *(a)* random, and *(b)* high-order. Overlap between two SDRs in the random scenario is very low, which results in a distribution that decays around 10%. On the contrary, overlap among SDRs in the high-order scenario is larger, and in fact, given the way in which the dataset was constructed, some SDRs have perfect overlap.

of symbols per sequence. This number is the temporal dimension of the tracked spike trains. In Fig. 4a we show the distribution of overlap scores obtained from estimating the overlap among all SDRs in this dataset. This distribution shows very low overlap among SDRs, which is expected given the random nature of the generation of these structures.

2. *High-order random sequences.* In the previous scenario we observed that the TM was able to learn the sequences very fast due to the nature of the sequences used. There was no ambiguity nor high-order sequences that would have made necessary extensive periods of learning. To introduce ambiguity in the sequences we created a set of high-order sequences in the following way: we generate a random sequence s consisting of 10 SDRs, then we construct high-order sequence s' by copying the contents of s into s' but substituting the first and last symbols by random SDRs. We repeat this process for 500 times which results in a total of 1000 sequences. The introduced ambiguity affected the accuracy of the TM as well as the structure of the pairwise correlations (see below). Figure 4b shows the distribution of overlap scores among the SDRs generated in this particular scenario.

As can be seen, the overlap scores are larger than in the random case, and in fact, there are SDRs with 100% overlap. This results from the particular way in which sequences are constructed in this scenario.

3. *Real-world data.* In this scenario we use the NYC taxi dateset which consists of passenger count measured at different times of the day for a year period. In this scenario each time-step is defined as the presentation of a single measurement of the dataset, that is, a single passenger-count number. This yields 17520 time-steps. Also, in this scenario there is no notion of epoch, that is, we present a single measurement to the TM only once. Figure 5a shows the first 1000 entries of this dataset. As can be seen, the data exhibits a recurring pattern of behaviour. As well, in this scenario we use a scalar encoder and spatial pooler in order to generate the appropriate set of SDRs to serve as input for the TM. In Fig. 5c we show the distribution of overlap scores for SDRs generated from the NYC taxi dataset.

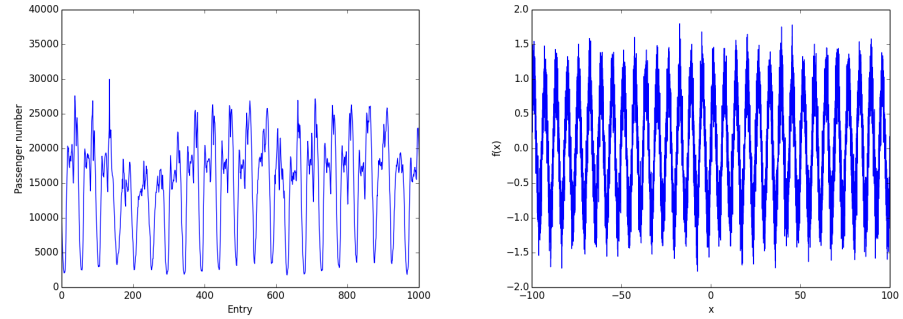
4. *Periodic artificial data.* In this scenario we generate a sinusoidal function and we add random Gaussian noise. In Fig. 5b we show an example of the sinusoidal function with noise added. We consider 10000 records sampled from this function. As with the previous scenario, we use a scalar encoder and a spatial pooler to generate appropriate SDRs to be fed to the TM. As well, here there is no notion of epoch as each record sampled from the sinusoidal function is presented once to the TM. Figure 5d shows the distribution of overlap scores. As with the NYC taxi dataset, the overlapping among SDRs is large and decays as it reaches unity. Large overlap scores are expected in periodic data as similar SDRs might occur in the data as time goes by.

Will the particular shape of the distribution of overlap scores from our datasets have an impact on the shape of neural correlations? We will explore this question in the following sections.

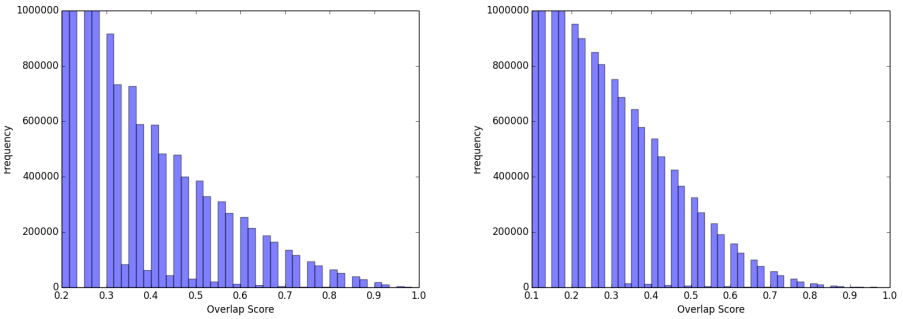
3 Results

3.1 Pairwise correlations are small

Pairwise correlations were estimated by subsampling the spike trains obtained from tracking activations in all cells. The number of cells considered in the sample was 1000; and we considered time windows of 1000 time steps of duration, that is, every 1000 time steps of simulation time we sampled



(a) NYC taxi dataset (b) Noisy sinusoid function.



(c) NYC taxi dataset (d) Noisy sinusoid function.

Figure 5: Raw data and overlap score distribution for two scenarios: *(a)*, *(c)* NYC taxi dataset, and *(b)*, *(d)* noisy sinusoid function. Both datasets are periodic, and this fact is captured by the distribution of overlap scores among SDRs.

1000 cells from the TM and considered its spike trains in the last 1000 time steps. Correlations were computed by estimating the Pearson correlation coefficient (PCC) between all neuron pairs in the sample. Next, we considered the distribution of such pairwise correlation coefficients.

In all scenarios considered the distribution of pairwise correlations exhibit weak correlations among neuron pairs. For instance, in Fig. 6 we show the distribution of pairwise correlations for the random scenario. Here, we can see how most pairwise correlations are small: no greater than 0.10, and also there are positive as well as negative correlations. There are correlations larger than those but occur so few times that they might be considered

outliers. This behaviour was also observed in the high-order random scenario, in which sequences possess a large overlap score among each other. In Fig. 7 we present the distribution of pairwise correlations for the high-order scenario.

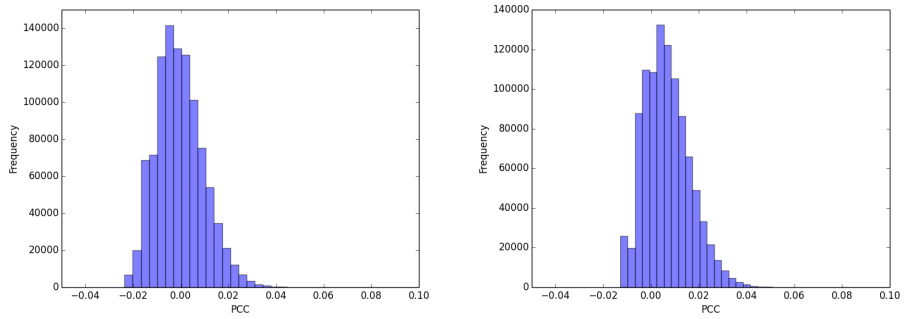
What is the effect of learning over the nature of pairwise correlations? In the case of the random and high-order scenarios, we observed that the effects of presenting the TM with repeated presentations of the same sequences for the purpose of learning (i.e. $epochs > 0$) had an impact over the quality of the distribution of pairwise correlations. As training times become larger the distribution shrinks in the sense that positive pairwise correlations decrease (see Fig. 14).

For the other scenarios, NYC taxi dataset and the noisy sinusoid, the shape of the distribution of pairwise correlations looks different. In both cases although there are no training periods larger than one for the TM, we do have them for the spatial pooler (SP) which will refine the process of generating SDRs from numerical values. Increasing training times in the SP stage has the effect of making SDRs more resilient to noise, which has the following consequence: similar input values will be mapped to similar SDRs. This process indeed results in changes in the shape of the distribution of pairwise correlations, as well as in the distribution of columns that become active during simulation time: the *column usage*.

For instance, in Fig. 8 we show how the number of records used for training the SP affects the distribution of column usage in the TM. As more items are used to train the SP, the less number of columns remain idle during simulation time.

The same behaviour is observed in the noisy sinusoid scenario (see Fig. 9). Naturally, the column usage is related to the correlated activity within the TM; thus, the shape of the distribution of column usage is also correlated with the shape of the distribution of pairwise correlations.

In Figure 10 we show the distribution of pairwise correlations from the TM for the NYC taxi dataset and for different number of training items in the SP. As can be seen in this image, the mode of the distribution is around a negative value, whose frequency becomes larger as more records are included in the training set for the SP. Unlike the random and high-order cases, the distribution exhibits a larger left-tail which implies the non-negligible existence of negative correlations. As well, this distribution exhibits a slowly decaying right-tail which is larger than in the random and high-order cases. In any case, pairwise correlations are weak, and in most cases negative. For instance, compare Fig. 6d with 10c. In the former, negative correlations barely exist and the distribution does not exhibit any



(a) $epochs = 1$ (b) $epochs = 5$

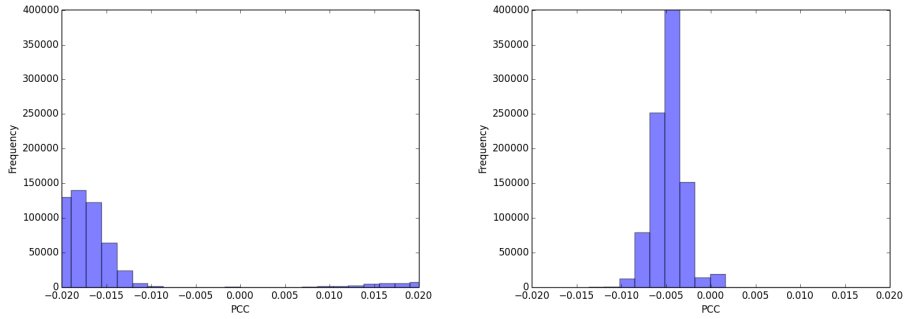
(c) $epochs = 10$ (d) $epochs = 15$

Figure 6: Distribution of pairwise correlations as estimated by the Pearson correlation coefficient of spike trains of neuron pairs for the random scenario. Here we also show the shape of the distribution as we increase the training periods (epochs) of the TM when presented with this particular dataset.

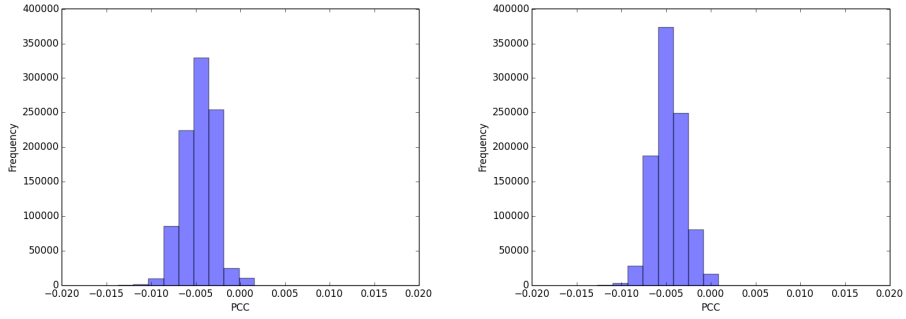
slowly-decaying right-tail.

A similar behaviour is observed in the correlated activity of the TM when using the noisy sinusoid dataset as input. However in this case a sharp transition between positive and negative correlations can be noticed (see Fig. 11). Negative correlations abound as in the case of the NYC taxi dataset; also, as in the latter case, left and right tails can be identified in the distribution.

We hypothesize that the particular behaviour of these periodic scenarios, namely the NYC taxi and the noisy sinusoid datasets, is related to the particular nature of the input stimulus. The periodic nature of the dataset



(a) $epochs = 1$ (b) $epochs = 5$



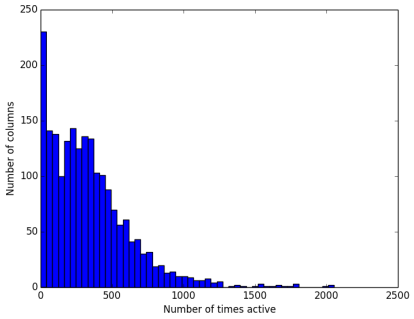
(c) $epochs = 10$ (d) $epochs = 15$

Figure 7: Same as Fig. 6 for the high-order scenario.

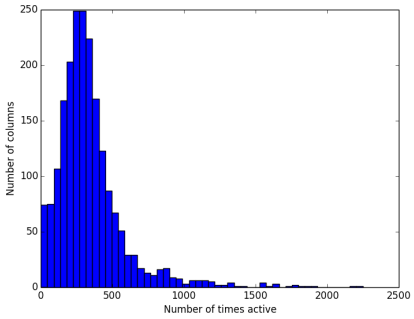
affects the overlap scores among SDRs in these two datasets (see Fig. 5). This ultimately affect the nature of the neural correlations. For instance, we would expect that as the TM learns, negative correlations start to emerge as the process of learning in the TM implies that cells within columns will be silent while one other cell from the same column is active in order to represent the current input. This means that cells within columns will become negatively correlated as the TM learns a sequence.

What are the effects of increasing the training periods of the TM for the NYC taxi and the noisy sinusoid datasets? For this purpose, we present the dataset in repeated occasions to the TM, and keep track of the evolution of the distribution of pairwise neural correlations.

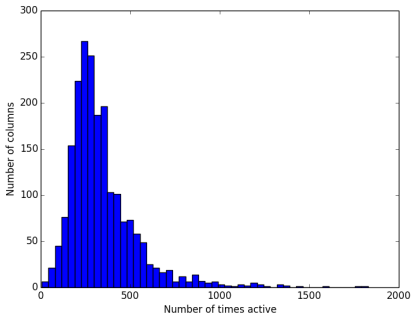
In Fig. 12 we present the distribution of pairwise correlations for the NYC taxi dataset at two different stages of learning: one at 1000 time-



(a) 5,000 training items



(b) 10000 training items



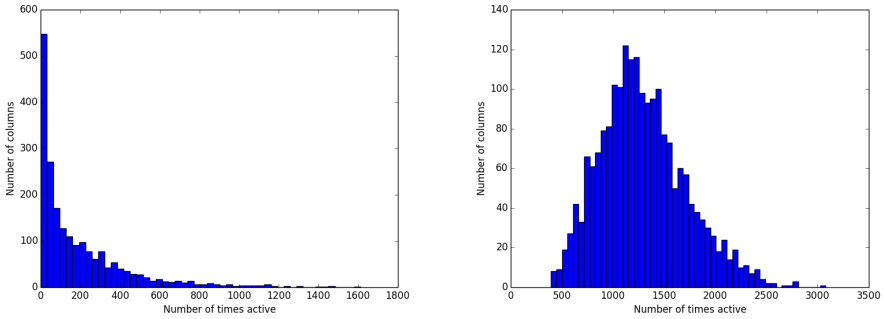
(c) 15,000 training items

Figure 8: Distribution of active columns or column usage in the TM. The x axis represent the number of times that a column has been active during simulation time. As we varied the number of training items for the SP, the number of idle columns decrease.

steps when the TM is at the early stages of learning, and another at 51,000 time-steps when the TM is completing 3 epochs of learning.

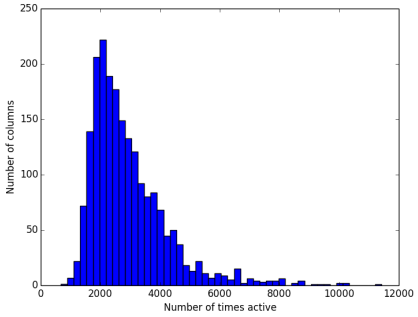
Unlike the random and high-order scenarios, the negative correlations do not vanish, nor the distribution shrinks. On the contrary, negative correlations abound possibly as a result of the periodic nature of the dataset.

A similar behaviour is observed in the noisy sinusoid dataset. As training periods in the TM increase, the right tail becomes more evident. Negative correlations do not vanish as occurs with the random and high-order scenarios, which again can be related to the periodic nature of the dataset.



(a) 1000 training items

(b) $5,000$ training items

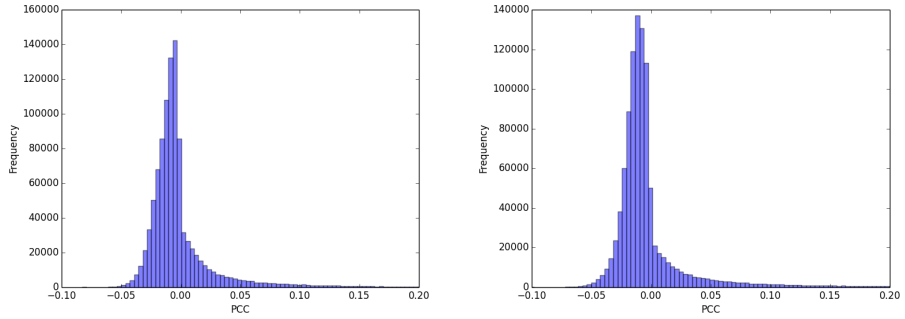


(c) 10000 training items

Figure 9: Same as Fig. 8 for the noisy sinusoid scenario. The number of items used to train the SP result in different shapes in the distribution of column usage in the TM.

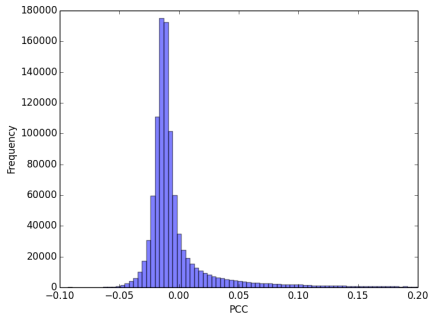
3.2 Network states are strongly correlated

Following the study of Schneidman and colleagues [13] we enquired about the frequency that binary patterns encoded by synchronous activation of a group of N neurons occur in the network as a result of stimulus presentation. The motivation behind this kind of analysis is to reveal whether weak pairwise correlations might have an overall effect on population coding when considering higher-order correlations or anticorrelations, and comparing them against a random model. In other words, this analysis serves as a proxy to study higher-order correlations within a population as well as a first step to study how a population of cells encode information. For instance, a binary pattern x of length N occurring more often than predicted



(a) 5,000 training items

(b) 10000 training items



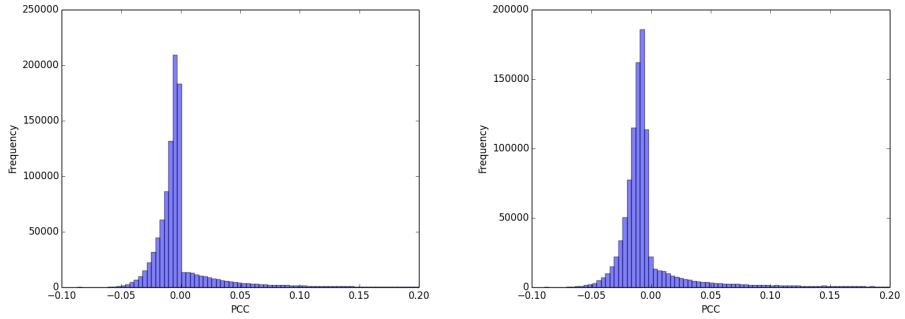
(c) 15,000 training items

Figure 10: Distribution of pairwise neural correlations for the NYC taxi dataset and different numbers of records used as training set for the SP. Unlike previous cases, this distribution exhibits slow decaying tails.

by a random model might carry neural information relevant to the organism. This is what is meant by network states in the work of Schneidman and co-workers [7, 13].

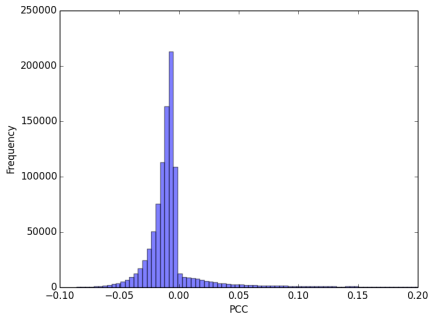
Fig. 15 shows an example of a raster plot obtained by subsampling $N = 10$ cells from the spike trains of the TM. Because in our experiments we are working with discrete time there is no need to consider a Δt window of time to count spikes as in Refs. [13, 2]. Therefore, we can build a binary pattern from spike trains by consider the activity of N subsampled cells at time-step t. Ten cells make a binary pattern of length 10 in which the i-th entry is 1 if the i-th cell is active at time t, and zero otherwise.

For the random and high-order scenarios, the "spiking" activity of a TM



(a) 1000 training items

(b) 4,000 training items



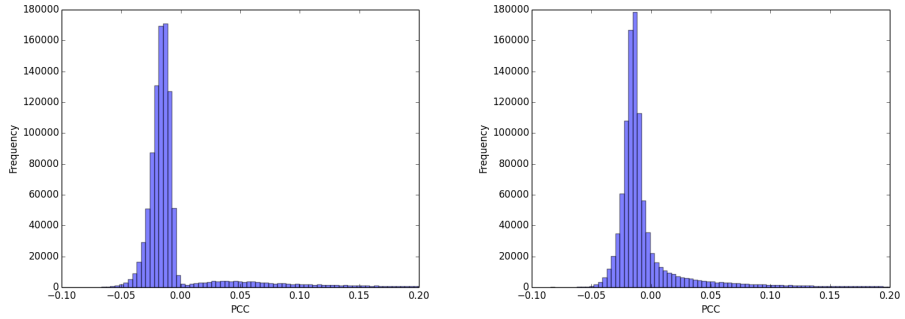
(c) 10000 training items

Figure 11: Same as Fig. 10 for the noisy sinusoid dataset.

in the early stages of learning resembles the activity of a Poisson spiking model with mean firing rate of 18 spikes per second[1]. Figure 16 shows raster plots of sampled cells from the TM and from the Poisson spiking model. The activity from both models is practically indistinguishable, suggesting that the TM behaves like a random spiking model as the TM starts to learn sequences.

We estimated the distribution of waiting times also known as inter-spike intervals (ISI) by sampling the activity of 1000 random cells in the TM, and then estimating the distribution of the time-steps in-between two spikes in such sample. The same process was repeated with the Poisson model, although in this case the spiking of 1000 cells was generated *ex profeso*. In

[1]In our simulations, a discrete time-step denotes a millisecond. Therefore 1000 time-steps result in a second of simulation.



(a) TM at time-step 1000 (b) TM at time-step $51,000$

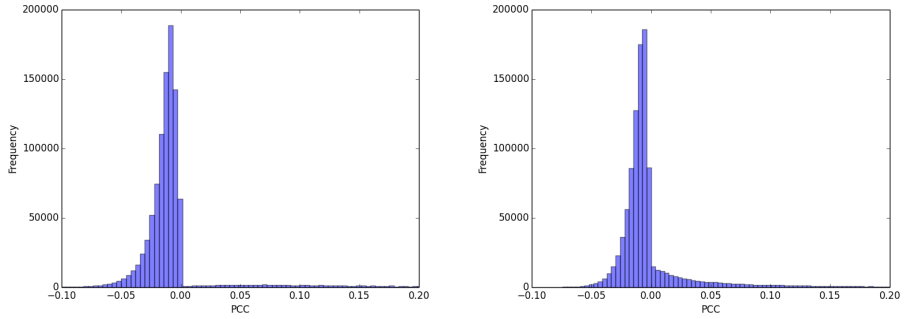
Figure 12: Distribution of pairwise correlations for the NYC taxi dataset. Here we increased the training periods of the TM by presenting 3 times the whole dataset. In *(a)* we show the distribution at the first stages of learning, whereas in *(b)* we present it at the last stages. As can be seen, negative correlations do not vanish as learning periods increase.

Fig. 17 we present the distribution of inter-spike intervals of both models. As can be seen, these distributions are nearly identical.

In order to study how frequent binary patterns occur in the TM, and how does this frequency compares to the random Poisson model we sampled $N = 10$ cells from both models by considering their complete spike history throughout the simulation. Then, following Ref. [13], we generated all 2^N binary patterns possible, and estimated the probability of observing each of them in the samples obtained from the TM and the random model. We repeated this process for 100 trials.

In Fig. 18 we show the probability of observing binary words of length $N = 10$ in the TM in the random and high-order scenarios compared to the probability of observing them in the Poisson spiking model. Each blue dot represents a binary word, e.g. [1011001101], from the total $2^{10} = 1,024$ possible words. The solid black line represents the identity function; a point lying in this line represent binary words that occur as frequently as predicted, whereas points lying below (above) the line denote binary words that occurred more (less) frequent than predicted by the random model.

In Fig. 18a dots close to the right top corner, which happen to occur very frequently and with the same probability in the TM and the Poisson model, correspond to binary patterns with only one bit "on". Similarly, the dots in the middle of the figure, which happen to occur more frequent in the TM



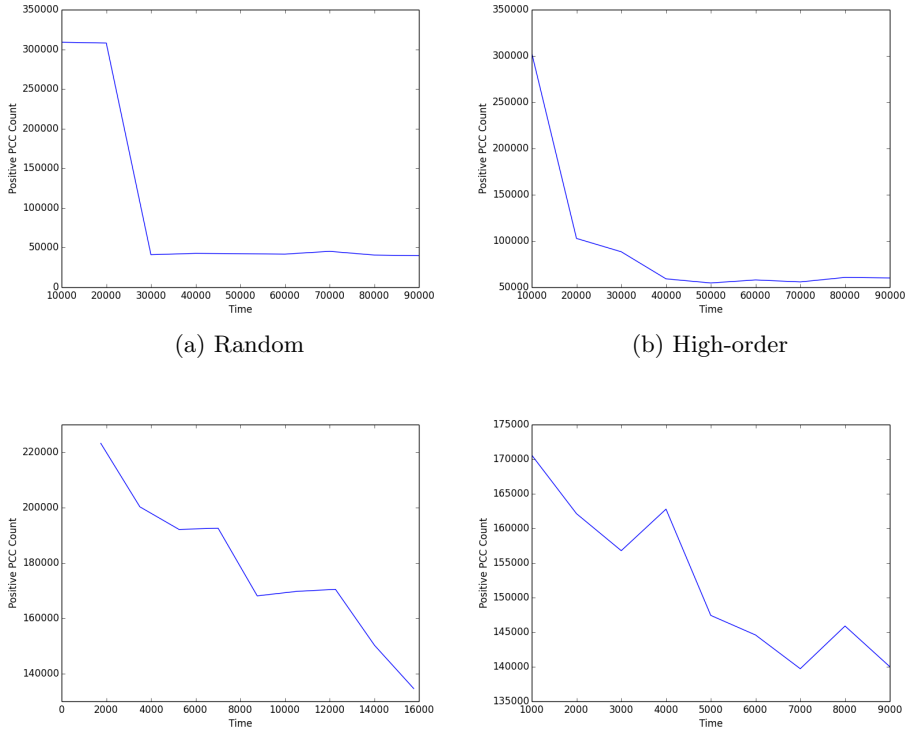
(a) TM at time-step 1000

(b) TM at time-step 29,000

Figure 13: Similar to Fig. 12 for the noisy sinusoid dataset. In *(a)* we show the distribution at the first stages of learning, whereas in *(b)* we present it at the last stages. As can be seen, negative correlations do not vanish as learning periods increase.

than predicted by the Poisson model, correspond to binary patterns with only two bits "on". The dots close to the upper left corner correspond to binary patterns with 3 or more bits "on". Here, there are patterns that occur more frequently than predicted by the random model, as well as patterns that occur less often than predicted by chance. Likewise, in Fig. 18b dots in the upper-right corner correspond to binary patterns with only one bit "on", whereas the dots in the bottom-left corner correspond to patterns with more than one bit "on".

Does the TM behave like a Poisson spiking model? We began this section presenting results that suggest that the TM behaves like a Poisson spiking model during the early stages of learning for the random and high-order scenarios. We showed how the spiking activity of the TM resembles that of a Poisson spiking model (see Figs. 16 and 17). However, when it comes to considering population of neurons representing binary patterns, and comparing their occurrence to a random model, the situation looks different. If the TM behaved like a Poisson spiking model, then when looking at particular patterns of N number of bits, we would expect them to occur as frequent as predicted by the model. However, this is not the case, and some binary patterns occur more often than predicted, while others are in the opposite situation. This suggests that binary patterns do not occur in a random fashion, and that in fact the activation of cells in the TM is not completely independent from one another although the distribution of



(a) Random

(b) High-order

(c) NYC taxi dataset

(d) Noisy sinusoid

Figure 14: Trace of amount of positive pairwise correlations for all experiments. The number of positive correlations decrease as learning increases in the TM. This implies that correlations are becoming weaker as learning progresses.

pairwise correlations points to the opposite idea.

The activation of cell populations (in this case of size $N = 10$) carries meaningful information for the whole network. In summary, we observed that the probability of observing particular binary patterns in the TM is larger than predicted from the random model. This implies, as in the study by Schneidman et al., and others [13, 11, 2], that although pairwise correlations are weak, the network exhibits strong correlated states in the form of synchronous activity of neurons that represent useful information to the population.

These observations were obtained in the random and high-order scenarios

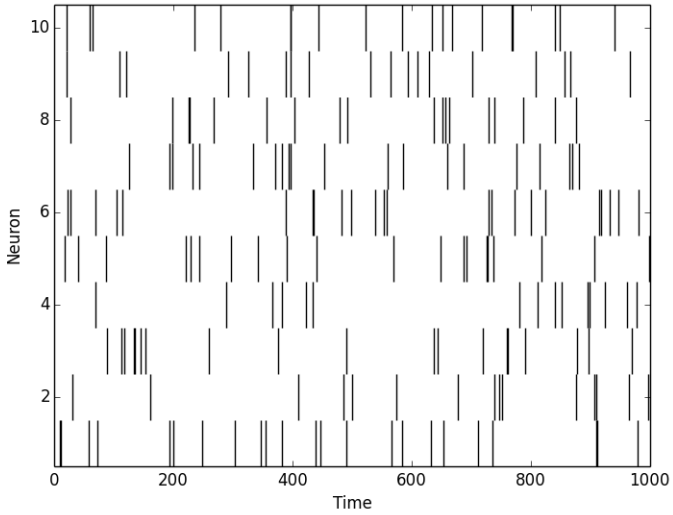
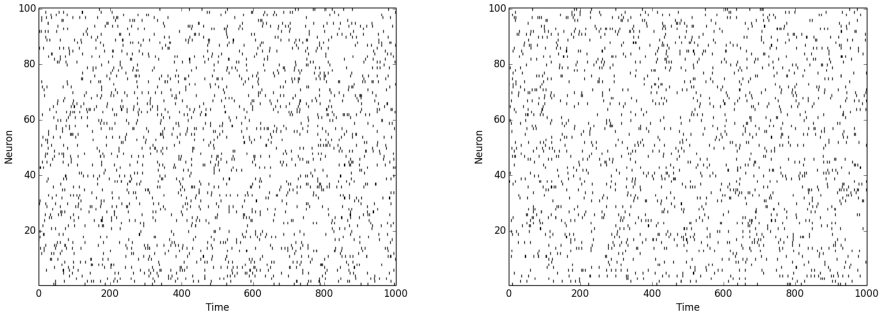


Figure 15: Raster plot of 10 cells in the TM for 1000 time-steps. A binary pattern is formed by considering the synchronous activation of cells in a discrete time-step t.

when the TM was presented with the dataset only once, that is, the number of epochs was one. What happens when we increase the number of training periods in the TM? As we increase the learning periods in the TM, the activity of the network resembles that of a Poisson spiking model even less. For instance, the distribution of waiting times (ISI) exhibits a long right-tail, which implies the presence of many cells with very short inter-spike intervals (high firing rate) coexisting with cells whose inter-spike intervals are very large (low firing rate). A Poisson spiking model cannot account for these observations.

Figure 19 shows the ISI distribution from the activity of a sampled group of cells in the TM when it was exposed to 5 training periods in the random scenario. It can be seen that there is a large number of cells with very short waiting times, that is, cells that are firing very often. Then, the distribution decays abruptly leaving a slowly-decaying right tail, which implies the presence of few cells whose waiting times are very large but it is not negligible.

We measured the *coefficient of variation*, which characterizes the vari-



(a) Temporal Memory (b) Poisson model

Figure 16: Raster plots of 100 cells sampled from *(a)* the TM, and *(b)* a Poisson spiking model. The activity from both models is practically indistinguishable.

ability in the inter-spike intervals [9]. This coefficient is given by:

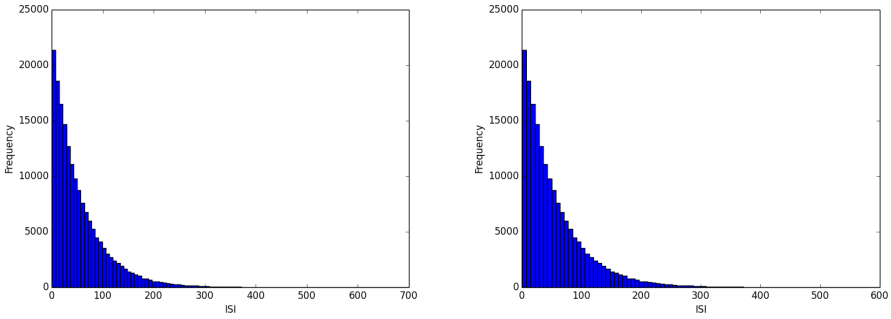
$$C_V = \frac{\sigma_{isi}}{\langle isi \rangle} \tag{1}$$

where $\langle isi \rangle$ refers to the mean of the inter-spike interval, whereas σ_{isi} refers to its standard deviation. It is known that a coefficient of variation of unity is a property of a Poisson process [9]. When measuring this coefficient on the ISI distribution of the TM after 5 periods of training, we obtained $C_V > 1$, whereas when there is no training periods (epochs = 0) the estimation of the coefficient yielded $C_V \approx 1$.

This particular shape of the ISI distribution was observed in the other scenarios, namely, high-order, NYC taxi dataset, and noisy sinusoid dataset. In these cases, the ISI distribution exhibits a slowly-decaying right tail even in the absence of training periods. Figure 20 shows the ISI distribution of the remaining scenarios (high-order, NYC taxi dataset and the noisy sinusoid dataset). All of these cases show a slowly-decaying right tail, which marks a divergence from the random Poisson spiking model.

Thus, in summary when it comes to the shape of the ISI distribution we observe two types of behaviour:

- *Exponential.* The distribution resembles an exponential distribution, which is known to reflect the presence of a Poisson spiking model [9]. This behaviour is observed in the random and high-order scenarios in



(a) ISI distribution of TM

(b) ISI distribution of Poisson model

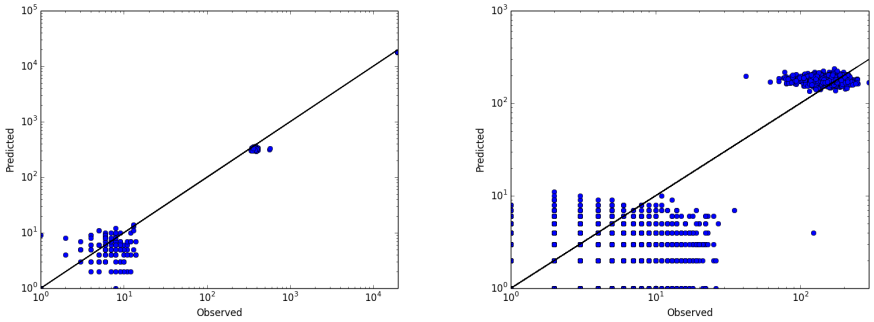
Figure 17: Distribution of ISI from the activity of 1000 cells sampled from *(a)* the TM, and *(b)* a Poisson spiking model. The activity from both models is practically indistinguishable.

the absence of training periods, which means that the TM observes every sequence only once.

- *Long tail.* The distribution looks like a power-law distribution with a slow-decaying right tail. This implies the existence of many cells whose firing rate is large coexisting with a few cells with low firing rate, and whose existence is non-negligible. This behaviour was observed in the random and high-order scenarios when the TM is exposed to training periods (epochs > 1), as well as all other scenarios even in the absence of training periods for the TM.

What is the origin of these two different behaviours of the ISI distribution? We hypothesize that the origin of this behaviour is related to the amount of novelty and familiarity that the TM has for the incoming input. A sequence of SDRs is fed to the TM. If the SDR is novel to the TM, then a set of columns burst. This increases the spiking activity in the TM, making it denser. This in turn, results in activity that can be approximated by a Poisson spiking model with certain accuracy. However, as the TM is presented with familiar input, complete columns stop bursting and the activity becomes sparser with few cells firing constantly and a majority of cells firing sporadically.

Figure 21 shows the distribution of spike counts from samples of 1000 cells in all scenarios considered and no training periods. The random and high-order cases yield a more symmetrical distribution than the other cases.



(a) Random (b) High-order

Figure 18: Frequency of occurrence of binary patterns in the TM (*observed*) for the random and high-order scenarios vs. the frequency predicted by the random model (*predicted*). Solid black line represents identity. (Read text.)

Moreover, these are the only case that can be related to a Poisson spiking model. The other distributions exhibit a tendency to develop a right tail which implies the existence of a few cells that spike frequently.

As mentioned above, when the TM receives familiar input, instead of bursting, all cells within a column will remain silent but one. A familiar input does not refer only to the same SDR occurring in different times during an experiment, but also to SDRs that have high overlap among each other.

The cumulative effect of column bursting increases the activity within the network resulting in an exponential shape of the ISI distribution, and its subsequent approximation to a Poisson spiking model. On the opposite, when columns stop bursting as a result of familiar input to the TM, the activity becomes sparser which results in the long-tail distribution of inter-spike intervals.

Power-law distributions of inter-spike intervals have been related to the conditional maximization of firing-rate entropy (CMFE) [15]. This hypothesis states that *in vivo* cortical neurons try to maximize the amount of information represented by their spike trains while keeping metabolic costs and uncertainty on output spike trains at minimum. It has been shown that a consequence of neurons operating under such constraints is the emergence of long and slow-decaying tails in ISI distributions [15].

Although a more rigorous examination is still underway, the ISI distributions in our experiments resemble a power-law distribution with slowly-

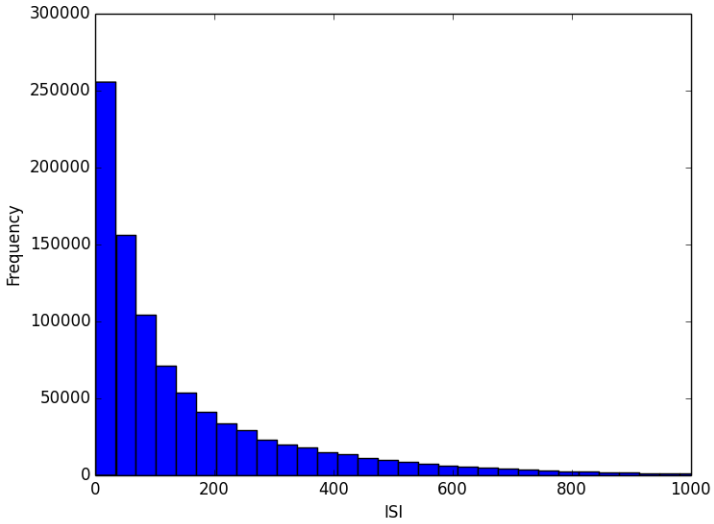


Figure 19: ISI distribution of the TM for the random scenario. Here, we exposed the TM with 15 training periods. Unlike Fig. 17a, the distribution exhibits an abrupt decay and a slow decaying right tail.

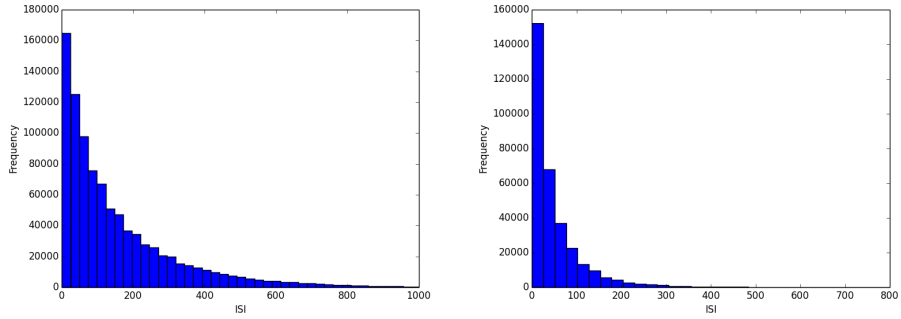
decaying right tails that could relate the activity of the TM with large training periods to the CMFE hypothesis. This will be a direction of future research.

3.3 Neural activity becomes sparse as learning increases

The particular shape of the ISI distribution (Fig. 19) is correlated with the activity within the TM becoming sparser. In other words, as the training epoch increases the activity in the TM becomes sparser. The phenomenon of sparse neural activity correlated with learning has been reported in experimental settings elsewhere [1, 3].

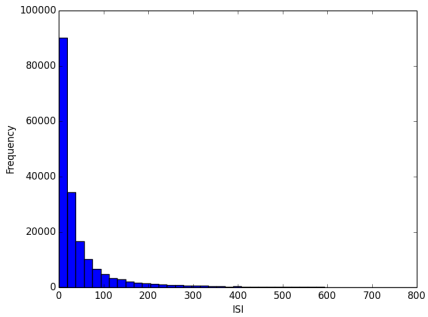
In Fig. 22 we present the evolution of the spike count for all cases considered. In most cases the trace exhibits a decreasing trend that implies that the activity is decreasing within the network as the TM learns from its input. In the NYC taxi and noisy sinusoid datasets we observe that although the trace is non-monotic, it exhibits a decreasing trend.

In Fig. 23 we present some examples of raster plots for two scenarios: random and high-order, at two stages of the learning process. Activity within the network is dense at the early stages of learning, but it becomes



(a) High-order

(b) NYC taxi dataset



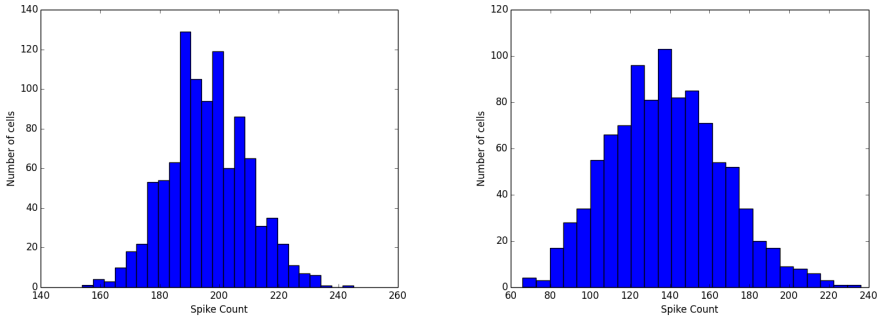
(c) Noisy sinusoid

Figure 20: ISI distributions for the three remaining scenarios considered in our experiments. For the high-order scenario (a), training periods were greater than one. In all these cases, the distribution exhibits a slowly-decaying right tail.

sparse as learning continues.

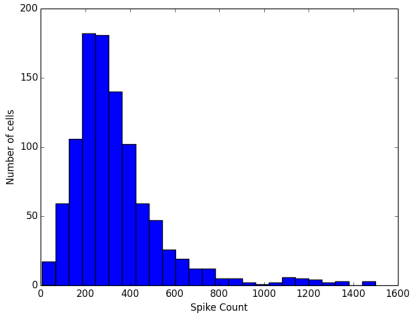
In Fig. 24 we present examples of raster plots for the other two scenarios: NYC taxi and noisy sinusoid datasets. In such plots it is less evident that the activity within the network is becoming sparser as learning continues. However a look at Fig. 25 gives a clearer idea of how sparsity evolves during learning.

Figure 25 shows the sparsity trace of all scenarios considered. The y axis of this plot represents the percentage of cells active in function of time. From a total of $16,384$ cells in the network, 2% activity corresponds to roughly 320 cells, which correspond to 40 columns bursting as a result of novel input

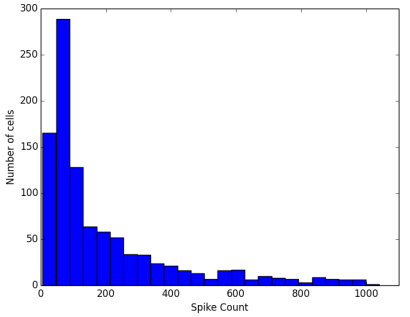


(a) Random

(b) High-order



(c) NYC taxi dataset

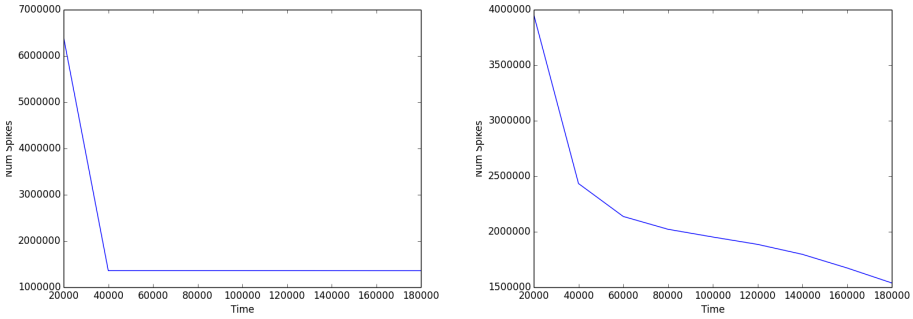


(d) Noisy sinusoid

Figure 21: Distribution of spike counts from the TM for all different scenarios and no training periods.

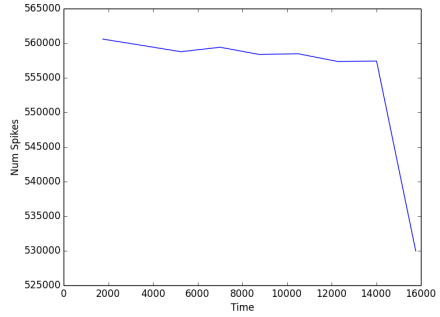
being presented to the TM. In a situation in which the TM learns accurately its input, whenever the TM is presented with a correctly predicted SDR only 40 cells will become active, this corresponds to roughly 0.2% of cells active at a given time.

In all cases considered the network converges to a sparse state of activity in which only a few cells are active. This behaviour is expected. As the TM learns from its input, less cells become active up to a point in which only 40 cells are active when there is perfect accuracy. The shape and speed of the decrease of activity within the network depends on the nature of the dataset.

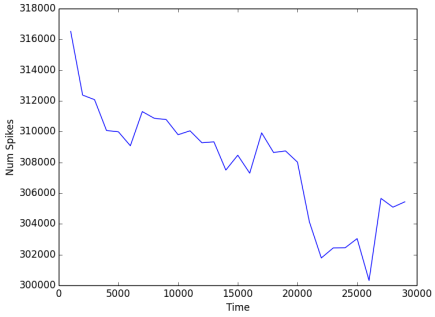


(a) Random

(b) High-order



(c) NYC taxi

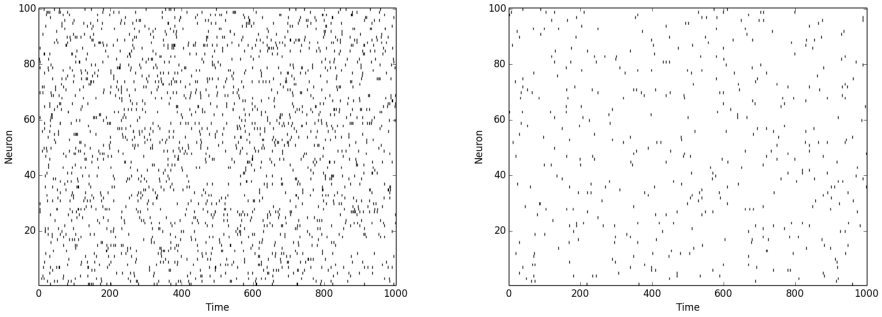


(d) Noisy sinusoid

Figure 22: Evolution of spike count in all scenearios considered. In most cases the trace exhibits a decreasing trend implying that the activity within the network is becoming sparser.

3.4 Learning increases the amount of negative correlations

Next we investigated on the effect of learning over the structure of pairwise neural correlations. We expected the amount of negative correlations increase as the TM learns the input. The reason behind this hypothesis is the following: novel input causes columns within the TM to burst, this leads to positively correlated activity of the cells within the bursting columns. As the TM learns the sequence, cells within columns stop bursting and on the contrary, we would observe only one cell per column becoming active. This leads to negative correlations among cells within the same column. This in consequence would lead to an increase in the number of negative correlations, once the TM learns its input better.



(a) Random scenario: early stages

(b) Random scenario: later stages

(c) High-order scenario: early stages

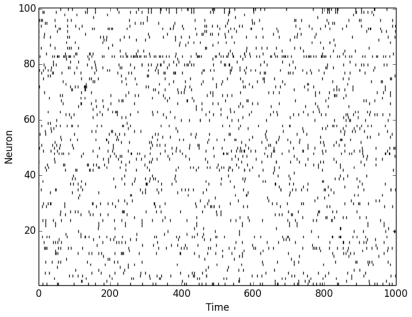
(d) High-order scenario: later stages

Figure 23: Raster plots of two scenarios: *(a)* and *(b)* Random, and *(c)* and *(d)* High-order. The activity in the early stages of learning is denser (a and c) than in its later stages (b and d).

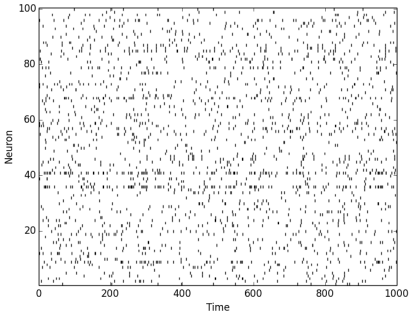
For scenarios such as the random and high-order cases, we were required to inspect the number negative correlations while the TM learns from the input, whereas for the NYC taxi and noisy sinusoid scenarios we required to assess the amount of negative correlations also as an effect of learning by the SP.

We report that negative pairwise correlations increase when increasing the training periods of the TM in all cases considered. Figure 26 shows the trace of the number of negative pairwise correlations. Here, the TM was learning while data was being presented to it. The amount of negative neural correlations increase with time as a result of training.

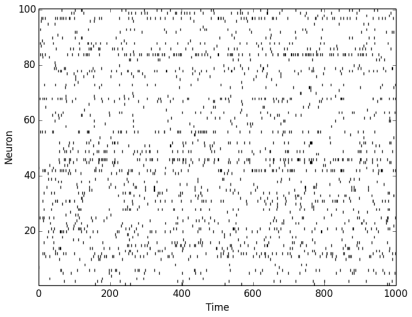
Moreover, we observed that the training periods in the SP have an effect



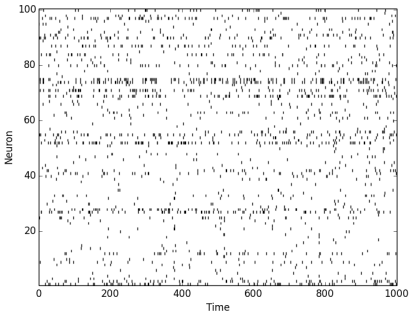
(a) NYC taxi dataset: early stages



(b) NYC taxi dataset: later stages



(c) Noisy sinusoid scenario: early stages

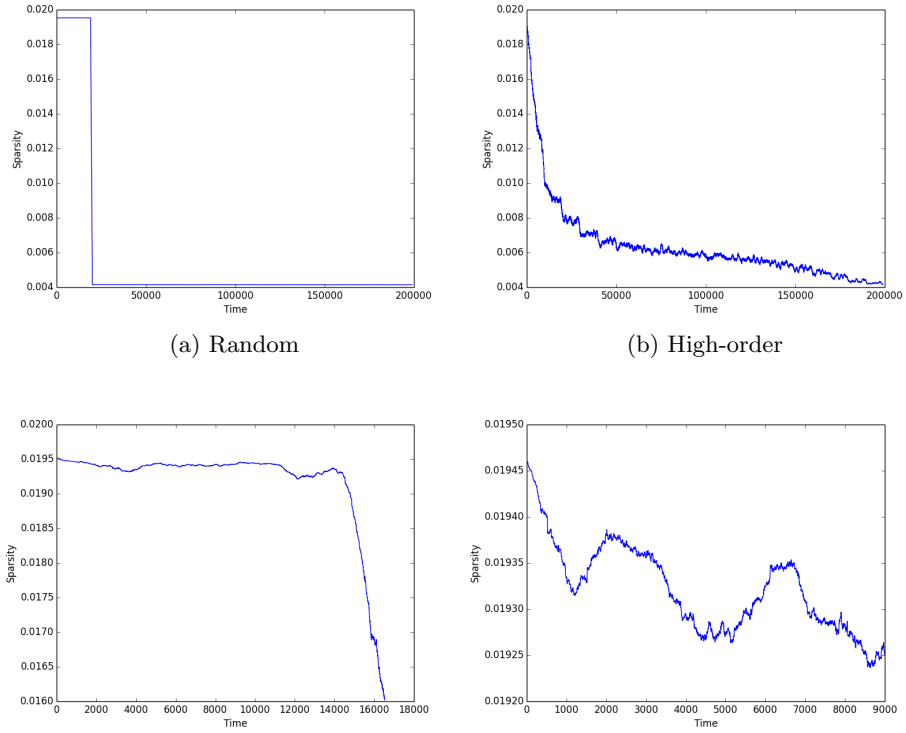


(d) Noisy sinusoid scenario: later stages

Figure 24: Raster plots of two scenarios: *(a)* and *(b)* NYC taxi dataset, and *(c)* and *(d)* noisy sinusoid. Activity within the network is denser in the early stages of training than in later stages. However, unlike Fig. 23, the decrease in activity is not too pronounced.

on the number of negative pairwise correlations in the TM. Here, we varied the number of training items for the SP. This not only has an impact on the amount of negative correlations emerging in the TM, but also in the frequency in which columns become active. As mentioned above, we refer to this as *column usage* and we showed two examples of these distributions in Figs. 8 and 9. These plots present the distribution of the number of times that a column has been active during simulation time. A tail in this distribution implies the existence of columns that are active very frequently.

In Fig. 27 we present the trace of the number of negative pairwise correlations in function of the training items in the SP. Here we fixed the number



(a) Random

(b) High-order

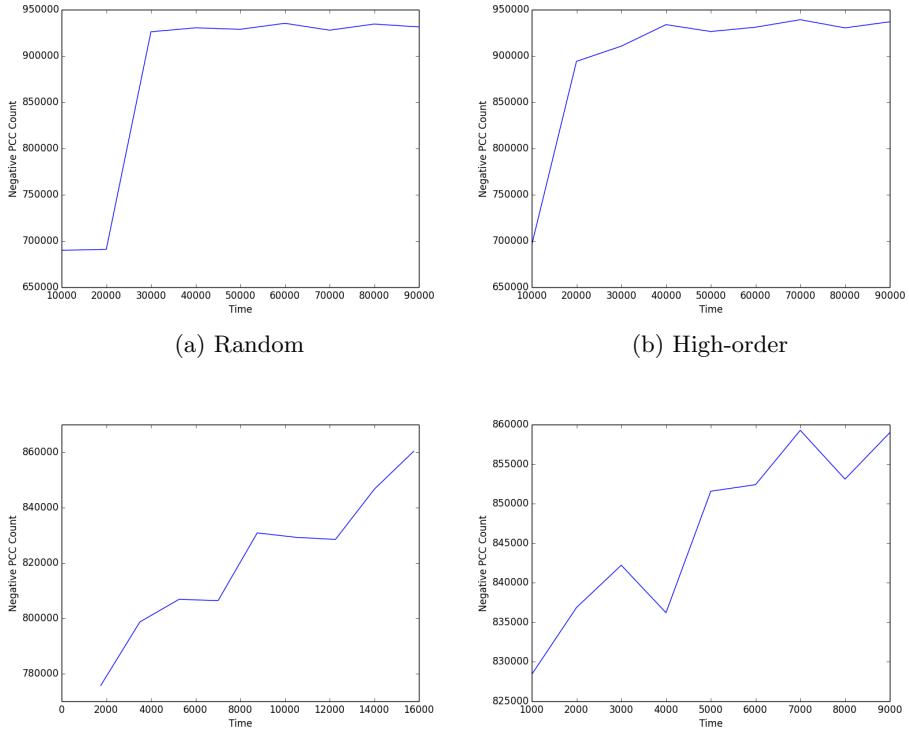
(c) NYC taxi dataset

(d) Noisy sinusoid

Figure 25: Sparsity trace for all scenarios considered. In all cases considered activity within the network tends to become sparser. The speed in which this occurs depends on the nature of the dataset.

of training periods in the SP but varied the number of items in the training set of the SP. In the case of the NYC taxi dataset (Fig. 27a) the number of negative pairwise correlations exhibit an increasing trend, whereas for the noisy sinusoid dataset (Fig. 27b) the trend is not clear.

In sum, we conclude that the amount of negative pairwise correlations increases with training in the experiments considered. However we were not able to verify that this phenomenon is clearer when sampling from cells within same columns as the process of choosing cells might be choosing cells from columns that are never active, or that become active sparsely as a result of the particular input being presented to the TM.



(a) Random

(b) High-order

(c) NYC taxi dataset

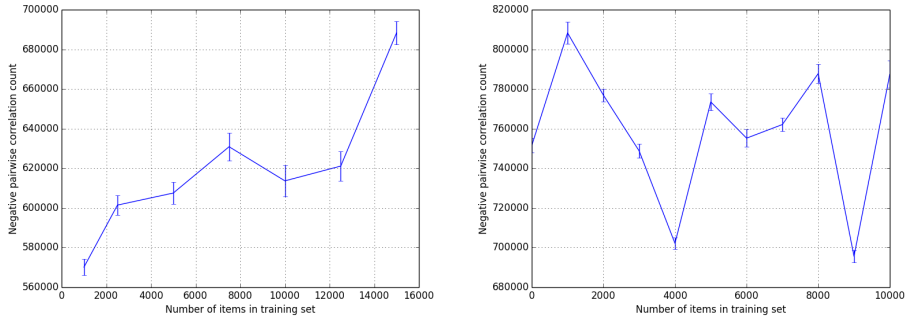
(d) Noisy sinusoid

Figure 26: Trace of amount of pairwise negative correlations all experiments. The number of negative correlations increases with training until saturation (a and b).

4 Conclusions

We have shown that for the random and high-order scenarios, the TM behaves almost like a Poisson spiking model in the absence of training. However, when measuring the rate in which binary patterns occur we observe that its behaviour differs from that of a random spiking model, and some patterns occur with more or less frequency than predicted. This analysis is a proxy for studying higher order correlations, that is, correlations that go beyond simple neuron pairs.

However, in the presence of learning (epochs > 1) the TM not longer behaves like a Poisson spiking model. This can be observed both in the shape of the distribution of inter-spike intervals as well as in the coefficient



(a) NYC taxi dataset (b) Noisy sinusoid

Figure 27: Amount of negative pairwise correlations in function of training items in the SP for the NYC taxi (a) and noisy sinusoid (b) datasets.

of variation, which becomes larger than unity with longer training periods. This observation is valid for the random and high-order scenarios when training periods are larger than unity, as well as for the NYC taxi and noisy sinusoid scenarios. In the last two cases the observation was valid even when the TM was exposed to the data only once. We believe that this is due to the nature of the data, which exhibits large overlapping scores, implying that some SDRs might resemble each other. This in consequence is interpreted by the TM as already-seen input. We showed how the distribution of the inter-spike intervals exhibits a slow-decaying right tail which has several implications on neural computation. Nevertheless, such implications remain to be explored in the current model.

We have shown as well that in all scenarios considered pairwise correlations are weak, and learning has the effect of reducing the amount of positive correlations. Moreover, we have shown how learning increases the amount of anti-correlated activity captured by the presence of negative correlations. This phenomenon is also related to the training parameters in the spatial-pooler for cases in which we use real numbers as input to the TM.

Thus, based in our observations we present the following predictions which might be testable in an experimental setting involving learning in neocortical tissue:

- Sequence learning in the neocortex has the effect of weakening pairwise neural correlations.

- Learning might increase the amount of negative correlations. This

observation will be related to the nature of the stimuli being fed to the network. Random input could be thought of artificial stimuli, whereas periodic input could be thought of natural stimuli.

- As well, this results in sparse activity within a neural network. This has also been hypothesized elsewhere in experimental neuroscience.
- The distribution of inter-spike intervals will exhibit a slowly-decaying right tail that will resemble a power-law distribution as a result of learning.

Further direction of work includes fitting a model to the TM at later stages of learning in order to repeat the Schneidman analysis comprising binary words and comparing their rate of occurrence between random model and TM. As well, further work includes the design of experiments that are more akin to computational neuroscience such as the presentation of bars to the TM. Lastly, we should as well investigate more about the implications of observing heavy-tailed distributions of inter-spike intervals in theoretical and experimental neuroscience.

References

[1] Bruno B Averbeck, Peter E Latham, and Alexandre Pouget. Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366, 2006.

[2] Pietro Berkes, Gergő Orbán, Máté Lengyel, and József Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013):83–87, 2011.

[3] Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature neuroscience*, 14(7):811–819, 2011.

[4] Alexander S Ecker, Philipp Berens, Georgios A Keliris, Matthias Bethge, Nikos K Logothetis, and Andreas S Tolias. Decorrelated neuronal firing in cortical microcircuits. *science*, 327(5965):584–587, 2010.

[5] Alexander S Ecker, Philipp Berens, Andreas S Tolias, and Matthias Bethge. The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience*, 31(40):14272–14283, 2011.

[6] Hugo Gabriel Eyherabide and Inés Samengo. When and why noise correlations are important in neural decoding. *The Journal of Neuroscience*, 33(45):17921–17936, 2013.

[7] Elad Ganmor, Ronen Segev, and Elad Schneidman. A thesaurus for a neural population code. *Elife*, 4:e06134, 2015.

[8] Yong Gu, Sheng Liu, Christopher R Fetsch, Yun Yang, Sam Fok, Adhira Sunkara, Gregory C DeAngelis, and Dora E Angelaki. Perceptual learning reduces interneuronal correlations in macaque visual cortex. *Neuron*, 71(4):750–761, 2011.

[9] David Heeger. Poisson model of spike generation. *Handout, University of Standford*, 5:1–13, 2000.

[10] James M Jeanne, Tatyana O Sharpee, and Timothy Q Gentner. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, 78(2):352–363, 2013.

[11] Urs Köster, Jascha Sohl-Dickstein, Charles M Gray, and Bruno A Olshausen. Modeling higher-order correlations within cortical microcolumns. *PLoS Comput Biol*, 10(7):e1003684, 2014.

[12] Alfonso Renart, Jaime De La Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D Harris. The asynchronous state in cortical circuits. *science*, 327(5965):587–590, 2010.

[13] Elad Schneidman, Michael J Berry, Ronen Segev, and William Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[14] Elad Schneidman, Susanne Still, Michael J Berry, William Bialek, et al. Network information and connected correlations. *Physical review letters*, 91(23):238701, 2003.

[15] Yasuhiro Tsubo, Yoshikazu Isomura, and Tomoki Fukai. Power-law inter-spike interval distributions infer a conditional maximization of entropy in cortical neurons. *PLoS Comput Biol*, 8(4):e1002461, 2012.