Acquisition and Representation of Spatio-Temporal Signals in Polychronizing Spiking Neural Networks

Felix Wang
Sandia National Laboratories
Albuquerque, New Mexico
felwang@sandia.gov

William M. Severa
Sandia National Laboratories
Albuquerque, New Mexico
wmsever@sandia.gov

Fred Rothganger
Sandia National Laboratories
Albuquerque, New Mexico
frothga@sandia.gov

ABSTRACT

The ability of an intelligent agent to process complex signals such as those found in audio or video depends heavily on the nature of the internal representation of the relevant information. This work explores the mechanisms underlying this process by investigating theories inspired by the function of the neocortex. In particular, we focus on the phenomenon of polychronization, which describes the self-organization in a spiking neural network resulting from the interplay between network structure, driven spiking activity, and synaptic plasticity. What emerges are groups of neurons that exhibit reproducible, time-locked patterns of spiking activity. We propose that this representation is well suited to spatio-temporal signal processing, as it naturally resembles patterns found in real-world signals. We explore the computational properties of this approach and demonstrate the ability of a simple polychronizing network to learn different spatio-temporal signals.

CCS CONCEPTS

• **Hardware → Neural systems**; • **Theory of computation →** *Self-organization*; • **Computer systems organization →** *Neural networks*; • **Computing methodologies →** Unsupervised learning; Learning latent representations.

KEYWORDS

Spiking neural networks, polychronization, signal processing, spatio-temporal coding, internal model

ACM Reference Format:

Felix Wang, William M. Severa, and Fred Rothganger. 2019. Acquisition and Representation of Spatio-Temporal Signals in Polychronizing Spiking Neural Networks. In *Neuro-inspired Computational Elements Workshop (NICE '19), March 26–28, 2019, Albany, NY, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3320288.3320291

1 INTRODUCTION

For a spiking neural network (SNN), the underlying mechanisms reside in the dynamics of the spiking neuron and their synaptic connections. As represented by spike trains, the flow of information within the SNN is governed by the synaptic strengths between neuronal units. Our focus is on how the neural system adjusts these strengths to better represent the signals from the environment such that the internal model may better determine desirable behavior in response to, or in anticipation of, incoming stimuli. With increasing fidelity of measurement, groups of neurons have been shown to exhibit complex spike-timing patterns with remarkable precision [18, 19]. Taking advantage of this precise temporal nature of spiking activity, we come to a phenomenon uniquely exhibited by SNNs: polychronization [11]. This is the self-organization of groups of neurons in the SNN that exhibit reproducible, time-locked patterns of spiking activity as a result of synaptic strengthening.

These patterns are called polychronous neural groups (PNGs). Due to their properties, we posit that the activation of a PNG may be considered to be an information-bearing signal at the level of the network, while simultaneously providing an integrative method for signal processing. Furthermore, due to their self-similarity to the spatio-temporal patterns received through the sensorium, we believe PNGs naturally lend themselves to composition. We propose that these properties make PNGs well suited to complex signal processing tasks such as language acquisition or object recognition.

To demonstrate this, we isolate and evaluate the polychronization learning mechanism in a toy network. This paper is organized as follows: we define polychronization and provide discussion of its computational properties in Section 2; we present some learning experiments in Section 3; we discuss algorithmic implications in the context of neuromorphic hardware in Section 4; and we conclude in Section 5.

NICE '19, March 26–28, 2019, Albany, NY, USA

ACM ISBN 978-1-4503-6123-1/19/03…$15.00
https://doi.org/10.1145/3320288.3320291

2 POLYCHRONOUS NEURAL GROUPS

At a high level, PNGs share similarity with the idea of a cell assembly, first proposed by Hebb, of a group of neurons that tends toward spiking as a collective due to their strong synaptic connections [8]. These cell assemblies were abstracted as an operational unit in the brain, and are extended and refined in the construction of PNGs.

2.1 Definition

Representationally, a PNG takes the form of a set of spatio-temporal pairs $G = \{(s_i, t_i)\}$, where the spatial component is indexed by *which* of the constituent neurons produces a spike, and the temporal component is indexed by *when*. Additionally, we may characterize the duration of a PNG as $\Delta_G = \max_{i \neq j}(t_j - t_i)$, and the activation sets as $A_G \subset G$. Here, activation sets are subsets of spatio-temporal pairs of G such that spiking activity coherent with the activation sets is sufficient to induce spiking activity of the full PNG.

Analogous to spiking at the level of the neural substrate, we consider the activation of a PNG to be the information bearing signal at the level of the network. Whereas individual neurons spike in response to local inputs, it follows from the characterization of PNGs that they are sensitive to distributed inputs.

2.2 Acquisition

The process by which PNGs are acquired follows the theory of neuronal group selection (TNGS) [5]. Here, the potential PNGs that are supported structurally by the SNN compose the *primary repertoire* of the network. As a result of the interaction with the spatio-temporal signals from the environment, certain PNGs out of this primary repertoire are strengthened, and others weakened. This self-organization through activity-dependent experiential selection provides the network with a *secondary repertoire*, composed of the PNGs that would subsequently participate in cognitive funtion. Coming from the bottom-up, the forward assembly of neurons results from the strengthening of individual synaptic connections; coming from the top-down, the backward selection of synaptic connections results from their mutual facilitation as a group.

Although the conditions in structural connectivity within an SNN that would support the existence of PNGs may at first seem to be an uncommon occurrence, just from the sheer number of convergent subgraphs that exist, it is typical that an SNN can combinatorially support many more PNGs than the number of neurons [12]. Consequently, it is typical that neurons will participate in the activation of multiple PNGs. Here, overlap in the constituent neurons between multiple PNGs may play an important role in facilitating the dynamic, functional connectivity that exists in co-active PNGs [20].

The capacity of an SNN in supporting a large number of coexisting PNGs is aided by a number of regulatory mechanisms. Primarily, the balance of excitatory connections with inhibitory ones maintains stability in the network by preventing runaway excitation as a result of the excitatory overlap between multiple PNGs. Additionally, metaplasticity mechanisms such as homeostatic synaptic scaling have been shown to increase the number of overlapping PNGs that may coexist within an SNN [7]. This leads to a greater degree of associativity in the network.

2.3 Computational Properties

In the brain, the construction of more abstract concepts from more primitive ones follows the process of "chunking", enabling the processing of increasingly complex signals while maintaining a limited usage of resources (e.g. working memory) [15]. As it relates to behavior, this enables an agent to manipulate and make estimates of increasingly generalizable hypotheses about the world.

Computationally, the main operation polychronization performs is to extract regularity from a spatio-temporal signal and represent it symbolically in the form of PNGs. This mirrors ideas from redundancy reduction, where the objective is to transform signals such that any hidden redundancies are made explicit [1]. As contrasted with dimensionality reduction, by explicitly capturing regularity in the signal more in line with a sparse coding approach, this form of redundancy reduction enables a system to more easily estimate probabilities and make predictions about statistically meaningful patterns, all the while using fewer computational resources.

The potential compositional ability of PNGs is particularly noteworthy. This is due to the self-similar representation of a PNG to the spatio-temporal signals that they represent. By remaining in the spatio-temporal domain, the composition of PNGs at higher levels of abstraction may follow the same self-organizing mechanism as that of PNGs closer to the sensorium. Furthermore, as a result of flexibility in network structure, these compositions are not restricted to a predefined hierarchy of representation, rather they may be able to integrate across multiple spatio-temporal scales [21].

3 LEARNING EXPERIMENTS

To evaluate our proposal that PNGs provide a suitable internal representation for processing spatio-temporal signals, we perform a number of learning experiments. Briefly, we train a simple network model on real-world datasets, and measure its PNG activations to perform classification. Here, we use the TIMIT speech dataset and the USPS handwritten digits dataset [6, 9]. Although high classification accuracy is preferable, the main objective of our experiments is to demonstrate capability.

3.1 Network Model

We simulate the network model using STACS, a simulation tool that enables interfacing an SNN with real-world signals [22]. For our experiments, we use a minimal SNN model that has been shown to exhibit polychronization [11]. The model consists of 1000 point neurons in an 80% to 20% balance between excitatory and inhibitory neurons, with sparse (10%) random connectivity between neurons. This balance was chosen to reflect the ratios found empirically in the neocortex [2].

Neurons are simulated according to the reduced-order dynamical systems model proposed by Izhikevich [10]. The system of equations describing the model dynamics is presented in equation 1, with state variables u, v, and parameters a, b, c, d. The parameters of the excitatory neurons are tuned to exhibit regular spiking behavior, corresponding to cortical pyramidal neurons, and the inhibitory neurons are tuned to exhibit fast spiking behavior, corresponding to cortical interneurons. The voltage threshold v_{thresh} for spiking is set to $30mV$ for both.

$$
\begin{aligned}
&\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I_{app} \\
&\frac{du}{dt} = a(bv - u) \\
&\text{if } v \geq v_{thresh}\,, \quad \text{then } \begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases}
\end{aligned}
\tag{1}
$$

Axonal delays from the excitatory neurons to their post-synaptic targets take on integer values between 1 and $20ms$, inclusive, and synaptic weights are calibrated such that the coincidence of at least two strongly connected pre-synaptic neurons will induce spiking activity in the post-synaptic neuron. These excitatory connections are subject to spike-timing-dependent plasticity (STDP) dynamics, with a synaptic weight range $w \in [0, 10]\mu A/cm^2$. For the inhibitory neurons, axonal delays are set to $1ms$, and the weights are fixed to $w = -5\mu A/cm^2$.

In contrast to the more distributed self-organization of a PNG, STDP is a local plasticity mechanism [14]. For our model, STDP is implemented through a trace variable, ω, that is set to 0.1 when a neuron spikes and decays exponentially with a time constant of $\tau_\omega = 20ms$. Corresponding to the positive and negative adjustments of the STDP curve, a filtered change in synaptic weight over a time

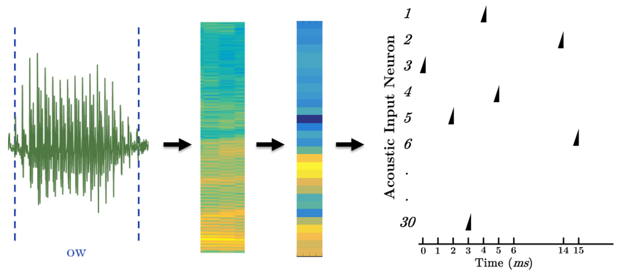

Figure 1: Transformation to spatio-temporal domain

period of 1s is updated as $\Delta w = \Delta w + \omega$ and $\Delta w = \Delta w - 1.2\omega$, respectively, for post-synaptic and pre-synaptic spikes, respectively. Subsequently, the synaptic weight is updated as $w = w + \Delta w + 0.01$, where the additional 0.01 term is used to prevent the synaptic connections from being prematurely pruned as to process potentially changing inputs during learning. Further details of this network may be found in [11].

Extending this network to accommodate signal processing, we implement an input interface as a layer of excitatory neurons whose spiking activity is driven by incoming signals. Although the axonal delays of this layer to the main network follow the same random distribution of integer values between 1 and 20ms, their synaptic weights are fixed to $w = 12\mu A/cm^2$ to induce spiking in the target neurons. This is to ensure that there is sufficient activity in the network for learning to occur.

3.2 Dataset

For a classic spatio-temporal signal, speech, we use the TIMIT speech dataset [6]. This dataset is noteworthy in that each speech utterance collected is phonetically time-aligned and tagged, and the sentences used were chosen to be phonetically balanced. Following the proposal by Lee and Hon which folds together a number of phonetic labels, we generate a phonetic inventory of 38 different phoneme classes [13]. We additionally fold these phonemes into the 5 broad phonetic categories (stops, fricatives, nasals, glides, and vowels) for a more general classification task.

To demonstrate the flexibility of polychronization to multiple signal domains, we also explored the visual input modality. In the spirit of the standard MNIST image classification task of handwritten digits, but due to the smaller size of the network, we used the USPS dataset, which provides 16x16 grayscale images of the handwritten digits 0 through 9 [9].

3.3 Signal Preprocessing

As a preprocessing step, we transform the raw data signals into the spatio-temporal domain. This is because the incoming stimuli to the network model are expected to be spike trains. Although the specific procedures for the speech and image data are different, the basic strategy is to convert magnitudes to a temporal delay (e.g. time-to-spike). An illustration of this transformation for speech (segmented into phonemes) is shown in Figure 1. The transformed signal is then interfaced through the set of excitatory input neurons.

For the speech signal, although we cannot capture the function of the auditory pathway in detail, we provide the network with a basic auditory system to obtain the input neural encoding. Moving from left to right in Figure 1, we initially convert the speech signal into the frequency domain using the short-time Fourier transform (STFT), corresponding to the function of the cochlea [16]. Subsequently, we transform the signal into a further reduced order basis borrowing the construction of the mel-scale by applying a psychoacoustically relevant filterbank over the power spectral density of the frequency content [4]. Finally, to translate the resulting signal into the spatio-temporal format suitable for processing by an SSN, we convert from densities to temporal delays by averaging their magnitudes over a given phoneme duration. The excitatory neuron corresponding to a given dimension of the reduced order basis then spikes according to this delay following the onset of a phoneme.

For the image signal, a far simpler approach was used, where each excitatory input neuron was assigned a pixel in the image. The transformation into the spatio-temporal domain was then simply achieved by mapping the magnitude of the pixel values to a temporal delay. This process is similar to how time-to-spike of light-sensitive receptors in the retina respond. For both the speech and image signals, the range of magnitudes to temporal delays was scaled to between 0 and 15ms.

3.4 Evaluation Metrics

We design a classification task that evaluates the predictive capacity of the network as a result of unsupervised learning via polychronization. For training, we present the network with the transformed spatio-temporal signals, balanced over the target classes. Here, synaptic plasticity is active, and we expect the network to self-organize through polychronization with respect to the statistical regularities in the signal. During testing, synaptic plasticity is inactive, and we are interested in measuring the network behavior with respect to the PNGs that formed during training. By considering only the spiking activity that follows the characterization of PNGs, we isolate the contribution of polychronization as the learning mechanism.

To identify and measure PNGs, we leverage information about the network structure to search through potential activation sets and follow axonal pathways. This is possible because appropriately timed spikes from an activating set of neurons of a PNG are sufficient to induce the consistent spike-timing pattern of the remaining spatio-temporal pairs due to their strengthened synaptic connectivity. Given an anchor neuron, we consider the various combinations of its pre-synaptic, or activation neurons that would result in spiking activity in an otherwise quiescent network. Consistent spike-timing patterns from these neurons are then used to compose the PNG by iteratively following strong synaptic connections starting from the activation set. Once identified, the set of spatio-temporal pairs of the PNG can be used as a template to detect percentage of activation within its duration.

In order to perform classification, we measure correlations between the different stimuli classes and their PNG activations. For each class, we compute $P(G_j|c_i)$ over the PNGs that formed.We may then estimate classes using a maximum likelihood approach,

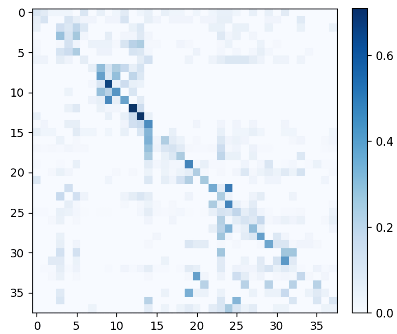

Figure 2: Confusion matrix for phoneme classes

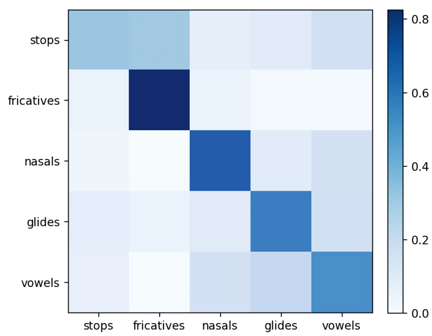

Figure 3: Confusion matrix for broad phonetic categories

following a naive Bayes assumption over the PNG activations as shown in equation 2.

$$
\begin{aligned}
c &= \arg\max_i P(c_i | \textstyle\prod_j G_j) \\
&= \arg\max_i \frac{\prod_j P(G_j|c_i)P(c_i)}{P(G_j)}
\end{aligned}
\tag{2}
$$

3.5 Classification Results

Using the same basic network model, we trained randomly initialized networks on 38 different phoneme classes, 5 broad phonetic categories, and 10 handwritten digits. Subsequently, we identified a total of 5095, 5192, and 8136 PNGs composing the secondary repertoire from the trained networks, respectively. Following the maximum likelihood approach described above, these networks achieved classification accuracy on the training data of 33.2%, 58.6%, and 48.1%, respectively. For the testing data, we achieved classification accuracy of 26.4%, 56.5%, and 41.9%, respectively. We provide the confusion matrices for the testing data in Figures 2, 3, and 4, respectively.

Although these classification accuracies do not achieve state-of-the-art performance, we note that they are significantly higher than chance (2.6%, 20%, and 10%, respectively). We note that the accuracies between training and testing are comparable, indicating that the acquisition of PNGs were generalizable, as opposed to overfitting on the training data. Most importantly, we find that in spite of any form of fine tuning with respect to the network

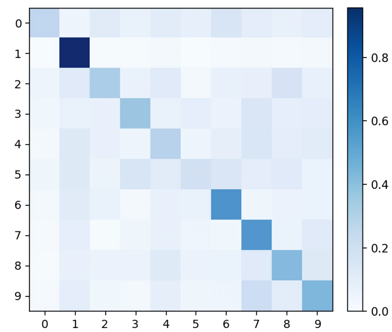

Figure 4: Confusion matrix for handwritten digits

model, and using a relatively simple classification method, we were able to demonstrate learning as a result of polychronization. Here, we can also be confident in that PNGs provide a suitable internal representation of information in an SNN as the only measurements made were the identification and activation of the PNGs. Finally, we observe that the mechanism of polychronization is flexible with respect to the input signal, provided it has been transformed into the spatio-temporal domain.

4 NEUROMORPHIC IMPLICATION

As it relates to implementation on neuromorphic hardware, aside from their representational capacity, a potential advantage of using PNGs for spatio-temporal signal processing is that they are distributed both spatially and temporally. That is, for a given PNG, there may be constituent neurons that have no direct connection, and the duration of the group activation may also extend beyond the maximum axonal delays between neurons. In this context, the self-organization of an SNN via polychronization may be considered to be a distributed method of learning that leverages local learning methods (e.g. STDP). Using a method that is inherently distributed like this should lend itself well to scalability both in terms of computation and communication.

4.1 Scalable Algorithms

In general, neural information processing at scale trades architectural complexity for algorithmic complexity. By forgoing more traditional global synchronization routines in favor of local coordination methods, large scale algorithms can achieve decreased communication latency, but interaction beyond local neighborhoods needs to pass through intermediate computational units. Algorithmically, a balance between signal propagation and attenuation is required to scale local interactions to global computation. And although there may be increased data locality per computational unit, application relevant data structures are increasingly distributed.

As a result, we believe that neural information processing at scale rests on emergent structures, such as exhibited by PNGs, where the complexity of the applications are managed only through the complexity supported by the algorithm [3]. To support the development of such algorithms on neuromorphic hardware, we find that a critical area to advance is the support of more collective operations on intermediate-scale spatio-temporal representations. Here, the focus is placed at middle ground between local computation at the level of the neural substrate and global computation at the level of a full network.

4.2 Computational Metrics

For benchmarking and comparing across different spatio-temporal learning approaches such as [17], we suggest a number of metrics for evaluation. Although the most prevalent metric is classification accuracy or task performance, this metric is not very informative on its own. While it may suggest that the network has learned useful information, it does not necessarily reveal how that information is structured or used. Additional metrics to supplement classification accuracy may be speed of learning, information capacity, and tolerance or robustness to noise. More relevant metrics for SNNs in particular may also include spike volume or efficiency, time to classification, distribution and size of PNGs, and the facilitation or interference of activation of related groups.

5 CONCLUSION

In this paper, we have presented an exploration of polychronization and their associated polychronous neural groups (PNGs), both computationally and in the context of neuromorphic hardware. We designed a set of experiments to isolate the learning capabilities of this learning mechanism in a toy network. We demonstrated that the self-organized spatio-temporal patterns that emerge from an unsupervised training process are predictive of the incoming signal and can be used for classification. Furthermore, we showed that the learning mechanism is flexible with respect to the specific input modality, as long as the signal is transformed into the spatio-temporal domain and its information is preserved.

Extending this work, we expect that improved methods for transforming signals into the spatio-temporal domain will yield more effective acquisition. Similarly, the introduction of additional regulatory mechanisms that support the formation of PNGs should also result in better representations. Network parameterization and signal encoding methods that result in greater or more efficient representational capacity should also be explored. We also suggest as a line of future work that focuses on biasing mechanisms, such as the introduction of supervisory or reinforcement learning signals. This direction would move learning away from a purely unsupervised regime and drive the formation of PNGs towards representations that are more relevant or useful with respect to a given task.

ACKNOWLEDGEMENT

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] Horace B. Barlow. 2001. Redundancy reduction revisited. *Network* 12 3 (2001), 241–53.
[2] Valentino Braitenberg and Almut Schuz. 1991. *Anatomy of the Cortex: Statistics and Geometry.* Berlin: Springer-Verlag.
[3] Roger C. Conant and W. Ross Ashby. 1970. Every good regulator of a system must be a model of that system. 1, 2 (1970), 89–97.
[4] S. Davis and P. Mermelstein. 1980. Comparision of parametric representations for monosyllabic word recognition in continuously spoken sentences. 28, 4 (1980), 357–66.
[5] Gerald M. Edelman. 1987. *Neural Darwinism: The Theory of Neuronal Group Selection.* New York: Basic Books.
[6] John Garofolo, Lori Lamel, William Fisher, Jonathan Fiscus, David Pallett, Nancy Dahlgren, and Victor Zue. 1993. TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1. (1993).
[7] Mira Guise, Alistair Knott, and Lubica Benuskova. 2015. Enhanced polychronization in a spiking network with metaplasticity. 9, 9 (2015).
[8] Donald O. Hebb. 1949. *Organization of Behavior.* New York: Wiley & Sons.
[9] Jonathan J. Hull. 1994. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16, 5 (May 1994), 550–554.
[10] Eugene M. Izhikevich. 2003. Simple Model of Spiking Neurons. 14, 6 (2003), 1569–72.
[11] Eugene M. Izhikevich. 2006. Polychronization: Computation with Spikes. 18 (2006), 245–82.
[12] Eugene M. Izhikevich, Joseph A. Gally, and Gerald M. Edelman. 2004. Spike-timing Dynamics of Neuronal Groups. *Cerebral Cortex* 14, 8 (2004), 933–944.
[13] K. F. Lee and H. W. Hon. 1989. Speaker-independent phone recognition using hidden Markov models. 37, 11 (1989), 1641–8.
[14] Henry Markram, Joachim Lubke, Michael Frostcher, and Bert Sakmann. 1997. Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. 275, 5297 (1997), 213–5.
[15] George Armitage Miller. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. 63, 2 (1956), 81–97.
[16] Brian C. J. Moore. 2004. *An Introduction to the Psychology of Hearing.* Elsevier Academic Press.
[17] Hélène Paugam-Moisy, Régis Martinez, and Samy Bengio. 2008. Delay learning and polychronization for reservoir computing. *Neurocomputing* 71, 7-9 (March 2008), 1143–1158.
[18] Yifat Prut, Eilon Vaadia, Hagai Bergman, Iris Haalman, Hamutal Slovin, and Moshe Abeles. 1998. Spatiotemporal Structure of Cortical Activity: Properties and Behavioral Relevance. *Journal of Neurophysiology* 79, 6 (1998), 2857–2874.
[19] John D. Rolston, Daniel A. Wagenaar, and Steve M. Potter. 2007. Precisely timed spatiotemporal patterns of neural activity in dissociated cortical cultures. *Neuroscience* 148, 1 (2007), 294 – 303.
[20] Botond Szatmáry and Eugene M. Izhikevich. 2010. Spike-Timing Theory of Working Memory. *PLOS Computational Biology* 6, 8 (2010), 1–11.
[21] Christian Tetzlaff, Christoph Kolodziejski, Irene Markelic, and Florentin Wörgötter. 2012. Time scales of memory, learning, and plasticity. *Biological Cybernetics* 106, 11 (2012), 715–726.
[22] Felix Wang. 2015. Simulation Tool for Asynchronous Cortical Streams (STACS): Interfacing with Spiking Neural Networks. In *Procedia Computer Science*, Vol. 61. 322–7.