Spike-Timing Dependent Plasticity model for the digit recognition task

Internship report

Sofiya Garkot, Maayan Levy, Tim P. Vogels

Institute of Science and Technology Austria
Computational Neuroscience and Neurotheory group

Abstract

Understanding and modeling a brain's performance enabled recent advances in Machine Learning. Spiking neural networks (SNNs) with an unsupervised Spike-Timing Dependent plasticity (STDP) rule present the most realistic approximation of neuronal dynamics. In this study, an SNN was investigated on its ability to form memory consolidations in response to ten classes of stimuli - digits of the MNIST dataset. The network consisted of 2500 Leaky Integrate-and-Fire (LIF) neurons with conductance-based synapses. Unlike other studies, the model neither received any information on the labels nor used teaching nodes to correct its behavior. However, the results showed no formation of neuronal assemblies in response to different sets of inputs.

Keywords: Spike-Timing Dependent Plasticity, Spiking Neural Networks, digit recognition

1. Introduction

Deep neural networks (DNNs) show impressive results in predicting the input data with high accuracy. However, their performance is bounded by the underlying data distribution as well as a number of adjustable parameters, raising the need for alternative models. A more resource-efficient and realistic model of brain information processing is a spiking neural network.

Spiking neural networks

Like biological neurons, the nodes of an SNN communicate via discrete asynchronous signals - spikes. Based on the inputs a brain infers the reality of the environment. The information about it is transferred to the central processing unit by trains of action potentials. Each of the spike trains represents a particular aspect of a stimulus and, depending on the ability of a system to match the current firing patterns to the existing ones, a certain combination can be recognized as a concept.

A representation of the environment is not stored in the activity of one neuron but is a result of the firing of many sub-networks each representing certain part. For instance, the authors of the study by Quiroga et al [15] found individual neurons coding the presence of a particular mental concept. Rather than assuming the existence of a "grandmother cell" corresponding to a memory, they assume that a studied neuron belonged an assembly representing it.

Firstly introduced by Hebb in 1949 [4], a neuronal assembly is a sub-network of strongly connected neurons characterized by spation-temporal properties. They are formed as a result of repeated exposure to a stimuli, which is later interpreted as a memory. Neurons belonging to an assembly can be located in different brain areas, making the experimental attempts to measure them extremely hard. On average each cortical neuron has 10.000 synapses, which leads to an enormous number of possible formations of patterns [5].

Plasticity

As opposed to a neuron in a DNN, a biological neuron does not explicitly and immediately receive information on how a whole network performs on a particular task. The feedback on the network performance does not affect the synaptic strengths of neurons in other layers. Due to memory consolidation in form of synaptic strengths, it is important to find a best approximation of the rule governing their change.

Synaptic plasticity is a fundamental mechanism of learning and memory consolidation because it rules the formation of the assemblies. STDP is the closest approximation to biological synaptic modification mechanism [6]. According to the STDP rule, a synapse is potentiated when a presynaptic action potential precedes a postsynaptic spike and is depressed otherwise. However, the computational models following this rule does not show a result in line with biological observations. In a network with this STDP rule stronger synaptic connections tend to grow while most of the weaker connections fall into the depression region. As reported in the study by Babadi and Abbott [1], the introduction of a time shift $\Delta t > 0$ is able to stabilize the synapses.

From the biophysical perspective, the shift is a result of the slow kinetics of NMDA receptors. The sharp transition from Long-Term-Depression (LTD) to Long-Term-Potentiation (LTP) occurs as a result of Ca2+ influx through NMDA receptors, which are initially blocked by Mg2+. After the arrival of the presynaptic spike, the Mg2+ block is removed almost immediately as a result of depolarization, but NMDA activation occurs a few milliseconds after the arrival of the presynaptic spike. Thus, these few milliseconds are needed to initiate the LTP [5].

Problem Description

The aim of the project is to test whether a biologically plausible model of a neural network can lead to the formation of neuronal assemblies in response to ten classes of handwritten digits from the MNIST (Modified National Institute of Standards and Technology database) dataset [8].

Spiking neural networks were firstly introduced by Maass in 1997 [10] as the third generation of neural networks. Among different variations of SNNs, the unsupervised networks with the STDP learning rule are the best approximation to biological neuronal dynamics.

One of the first attempts to use this network configuration was done by Brader et al. in 2007 [2]. The network consisted of 784 excitatory neurons densely connected by plastic synapses with the readout layer of size 10. The authors used a population of inhibitory neurons that balance the incoming excitation, and teacher neurons that impose selectivity of output neurons, reaching 96.5% of classification accuracy. However, this configuration of a network is of questionable biological plausibility.

In the study by Diehl et al. [3] the recognition accuracy of 95% was reached by a network of 6400 Leaky-Integrate-and-Fire (LIF) neurons with synapses governed by the STDP rule. However, the authors use a specific connectivity pattern: all excitatory neurons were connected to inhibitory ones in a one-to-one fashion, which does not match the 4:1 ratio observed in vivo.

There are applications of reward-modulated SNNs with STDP rule [13], as well as deep convolutional SNNs [7] reaching up to 98.17% accuracy.

The motivation to study an SNN's classification in form of activity patterns is the result of the study by Litwin-Kumar [9]. The authors showed the ability of a network to form interconnected neuronal assemblies with realistic plasticity rules. The learned patterns were later noticed as spontaneous dynamics reflecting past stimuli. To test the network's ability to complete a pattern the authors presented stimuli to half of the neurons of an assembly. The difference in pre- and post-training activity of the unstimulated part of an assembly led to the conclusion that the presence of a stimuli class can be read out from the activity of neurons, that did not receive the input.

2. Methods

Network

The studied SNN consists of 2500 LIF neurons with initial random connectivity. The ratio of excitatory to inhibitory population is 4:1. The probability of synaptic connection within the excitatory population is 12%, excitatory to inhibitory is 20%, inhibitory to excitatory - 22%, and within inhibitory is 24%. The model of synaptic connections is conductance-based [19] as seen on Eq. 1.

$$\tau \frac{dV}{dt} = (V_{rest} - V) + g_{ex}(E_{ex} - V) + g_{inh}(E_{inh} - V) \tag{1}$$

Supported by the evidence of a study by Song et al. [17], the initial values of a synaptic conductance were sampled from a log-normal distribution with parameters $\mu = -0.6$ and $\sigma = 0.5$.

The strength of a connection corresponds to an increase in conductance between pre- and post-synaptic neurons. The choice of the shift in the STDP rule is 4 ms (Fig. 1). The amplitude of potentiation is 0.008 nS and the amplitude of depression is -0.004 ms.

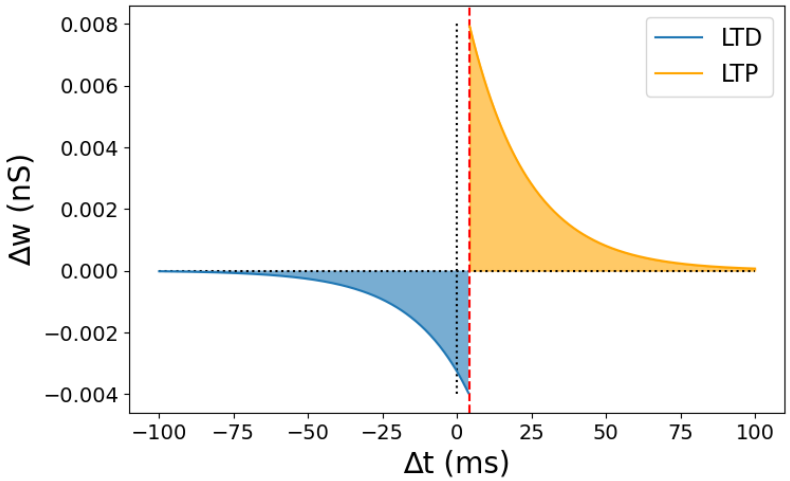


Figure 1: Spike-Timing-Dependent Plasticity rule
The x axis corresponds to time difference between post- and pre-synaptic spike times. The y axis corresponds to a change in conductance between the neurons in response to a spike. The regions of potentiation and depression are colored in yellow and blue correspondingly.

Pattern representation

A sample from the MNIST dataset is presented to the network pixel-wise. Each pixel is transformed into a Poisson spike train with firing rates proportional to its luminescence value. The firing rates were normalized to the range of 0 to 40 Hz. Each encoded spike train is shown randomly to 8% of the excitatory population of the network during 300 ms.

The input-to-network connectivity is motivated by the activity of biological neurons in response to natural stimuli. For instance, the range of 3%-15% of activated neurons has been recorded for the olfactory system in the neurons in the piriform cortex by Stettler and Axel [18].

The neurons receiving the input are chosen randomly, but the pixel-to-neuron mapping is the same across the samples. Every part of a picture is shown to the same random part of the network such that a network is able to differentiate between the stimuli by activating different neurons in response to an area of a picture.

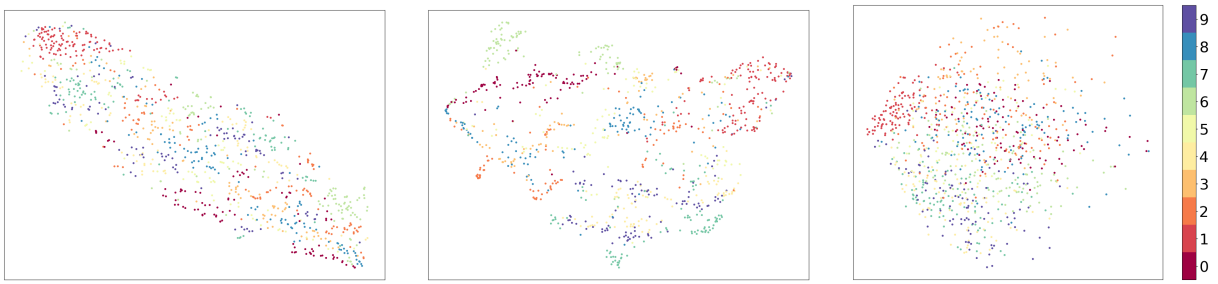
In order to estimate network performance during a short time period, only 10% of the dataset was used resulting in 6000 train samples and 1000 test samples. The synaptic connections of the excitatory population were plastic during the training phase, and during the test phase, the plasticity was turned off.

3. Results

Assemblies

To observe the response to different classes of digits, the total number of spikes of 2500 neurons in response to every sample of the test set is projected onto 2D space using the dimensionality reduction algorithm t-distributed stochastic neighbor embedding (t-SNE) [11]. The results did not show any formation of clusters based on the firing activity of the network. All of the tested dimensionality reduction

algorithms (t-SNE, Uniform Manifold Approximation and Projection (UMAP) [12], and principal component analysis (PCA) [14]) did not show the formation of the clusters in response to the digit classes (Fig. 2).



(a) The spiking activity of the neurons projected onto 2D using t-SNE. The x- and y-axis are expressed in arbitrary units.

(b) The spiking activity of the neurons projected onto 2D using UMAP. The x- and y-axis are expressed in arbitrary units.

(c) The spiking activity of the neurons projected onto 2D using PCA. The x- and y-axis are expressed in arbitrary units.

Figure 2: Results of dimensionality reduction on spiking activity of the network. The number of spikes of 2500 neurons was projected onto 2D using corresponding dimensionality reduction algorithms. The network was stimulated by 1000 samples of the test set. Each sample is color-coded according to the class. No formation of the clusters using the spiking response of the network can be found.

To detect the formed assemblies, the most active neurons of each class of digit were recorded and analyzed. The top 1% of neurons by firing activity intersected across the classes on 45%, which led to the conclusion that mostly class-specific assemblies were either weak or did not form. The number of spikes in response to each class of digits on the test set is plotted on the heatmap (Fig. 3). The results suggest that a network was rather linearly responding to the input density of the signal by output firing rate with some amount of variation between the classes.

To investigate the reason for the overlapped most spiking neurons, the synaptic weights were recorded during the training, and the dynamics of weights distribution can be seen in Fig. 4. The results suggest, that the network was not balanced by the inhibitory population during training, causing the shift to an unbalanced bimodal distribution during the first 300 training epochs. Afterward, the number of stronger synapses grows till the end of training.

STDP

To change the trend of distribution change, a variety of experiments were done on the learning rule. None of the tested shifts in the range from 0 to 10 ms led to lognormal weight distribution after training. The left extreme of 0 ms (Appendix A.5) led to even earlier unbalancing of the weights, and the right extreme of 10 ms (Appendix A.6) led to the silencing of all the synapses during early training epochs.

Another set of experiments was made on the amplitude of depression and potentiation and their impact on balancing the distribution of the synaptic weights. As expected, higher values of potentiation amplitude led to an increase in the number of synapses with strong connections, while a wider span of the amplitude of depression led to increasing in the number of weak synapses. However, both strategies led to unbalanced bimodal synaptic strength distribution after the training.

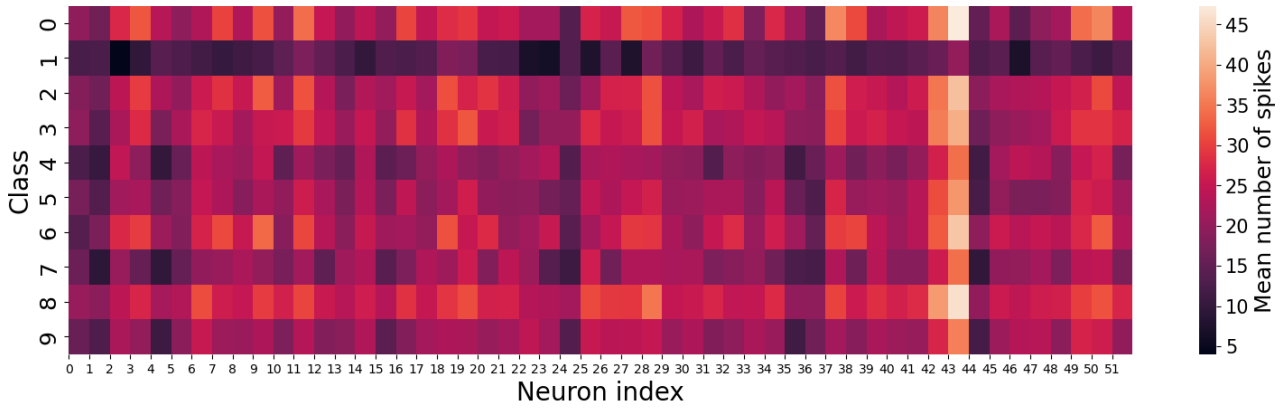


Figure 3: The average number of spikes of a neuron in response to a sample from a digit class. The y-axis corresponds to a class of digits. The x-axis corresponds to indices of most spiking neurons across ten classes (arbitrary units). The heatmap shows the mean number of spikes of a neuron in response to a subset from the test set. The neurons do not form class-dependent assemblies.

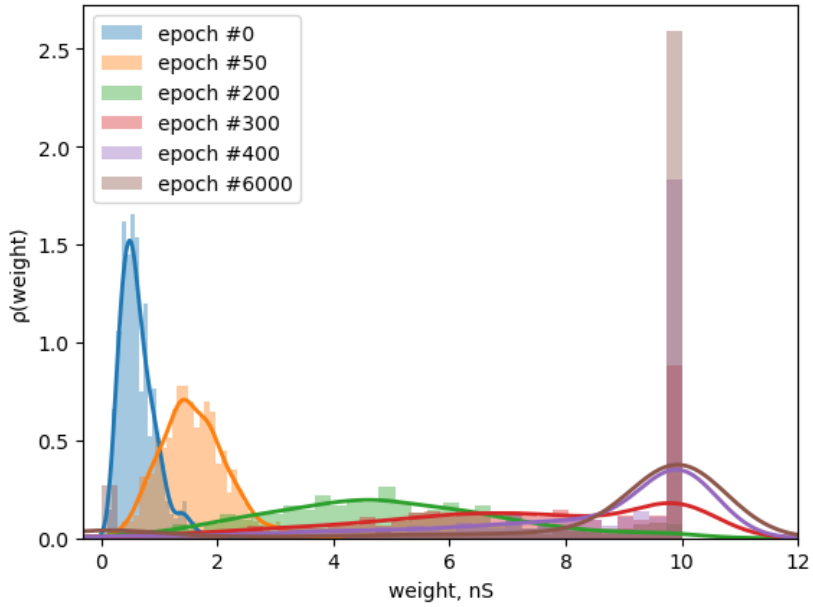


Figure 4: Synaptic weights distribution of excitatory population during training. The x-axis corresponds to a synaptic strength in nanosiemens. The y-axis shows the density across all the synapses from excitatory neurons. The weights were recorded after corresponding epochs, and their density is color-coded and approximated using kernel density estimation. The network shifts the excitatory synaptic weights during the first 300 epochs, staying in the unbalanced bimodal state till the end of the training.

4. Discussion

The results showed that despite the introduction of a shift, the network was unable to preserve the balanced state of synaptic weights. After 300 epochs of training, the weights of synapses were shifted to two extremes, causing the following trend. The subset of neurons, that have the highest spiking activity in response to a digit class, overlaps by 45% across all the classes. Thus, it is not able to form different assemblies in response to ten classes of inputs.

The current implementation of an SNN did not show the desired formation of synaptic connectivity patterns. A potential cause is the dense input-to-network mapping. Thus, further experiments with a larger number of neurons in the network should be made to evaluate it. Further experiments on the STDP rule are needed to detect the possible formation of the assemblies, as reported in previous studies.

Another path of investigation of the current setup is the result of Song et al. [17] questioning the plausibility of random connectivity. The results of the study showed that the strengths of synaptic connections with common pre- or postsynaptic neurons were correlated. Thus, strong synapses are more clustered than weak ones, which could be realized by creating a subnetwork of excitatory neurons with strong connections. Another line of improvement is the order and duration of training. Inspired by the results of a study by Smolen et al. [16], the proposed approach would be to use a self-sustained model and present a stimulus sequentially: 5 periods of 200 ms of input followed by 200 ms without it. Thus the total duration of learning of one sample would increase to 2 s.

Acknowledgements

Many thanks for the supportive supervision of Dr. Maayan Levy, who initiated and is leading this project now. I am grateful to meet and learn from the members of Vogels lab, who gave helpful answers to my questions on neuronal dynamics and shared their life journeys with me during the last 3 months.

References

[1] B. Babadi and L. Abbott, "Intrinsic stability of temporally shifted spike-timing dependent plasticity", PLoS computational biology **6**, e1000961 (2010) 10.1371/journal.pcbi.1000961.

[2] J. Brader, W. Senn, and S. Fusi, "Learning real-world stimuli in a neural network with spike-driven synaptic dynamics", Neural computation **19**, 2881–912 (2007) 10.1162/neco.2007.19.11.2881.

[3] P. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity", Frontiers in Computational Neuroscience **9**, 10.3389/fncom.2015.00099, ISSN: 1662-5188 (2015) 10.3389/fncom.2015.00099.

[4] D. O. Hebb, *The organization of behavior: A neuropsychological theory* (Wiley, New York, 1949).

[5] E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of neural science, fifth edition* (Jan. 2013), ISBN: 9780071390118.

[6] R. Kempter, W. Gerstner, and L. van Hemmen, "Hebbian learning and spiking neurons", Phys. Rev. E **59**, 10.1103/PhysRevE.59.4498 (1999) 10.1103/PhysRevE.59.4498.

[7] S. R. Kulkarni and B. Rajendran, "Spiking neural networks for handwritten digit recognition—supervised learning and network optimization", Neural Networks **103**, 118–127, ISSN: 0893-6080 (2018) https://doi.org/10.1016/j.neunet.2018.03.019.

[8]Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE **86**, 2278–2324 (1998) 10.1109/5.726791.

[9]A. Litwin-Kumar and B. Doiron, "Formation and maintenance of neuronal assemblies through synaptic plasticity", Nature communications **5**, 5319 (2014) 10.1038/ncomms6319.

[10]W. Maass, "Networks of spiking neurons: the third generation of neural network models", Neural Networks **10**, 1659–1671, ISSN: 0893-6080 (1997) https://doi.org/10.1016/S0893-6080(97)00011-7.

[11]L. van der Maaten and G. Hinton, "Visualizing data using t-SNE", Journal of Machine Learning Research **9**, 2579–2605 (2008).

[12]L. McInnes, J. Healy, and J. Melville, *Umap: uniform manifold approximation and projection for dimension reduction*, 2018, 10.48550/ARXIV.1802.03426.

[13]M. Mozafari, M. Ganjtabesh, A. Nowzari-Dalini, S. J. Thorpe, and T. Masquelier, "Bio-inspired digit recognition using reward-modulated spike-timing-dependent plasticity in deep convolutional networks", Pattern Recognition **94**, 87–95, ISSN: 0031-3203 (2019) https://doi.org/10.1016/j.patcog.2019.05.015.

[14]K. Pearson, "On lines and planes of closest fit to systems of points in space", The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **2**, 559–572 (1901) 10.1080/14786440109462720.

[15]R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain", Nature **435**, 1102–1107 (2005).

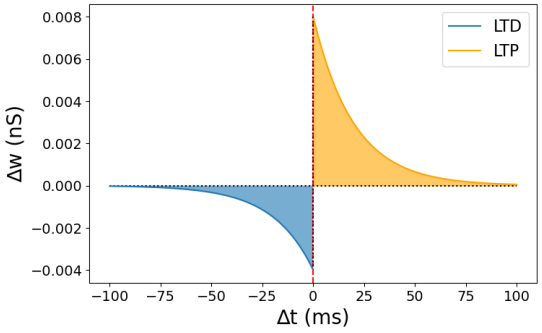
[16]P. Smolen, Y. Zhang, and J. Byrne, "The right time to learn: mechanisms and optimization of spaced learning", Nature Reviews Neuroscience **17**, 77–88 (2016) 10.1038/nrn.2015.18.

[17]S. Song, P. Sjöström, M. Reigl, S. Nelson, and D. Chklovskii, "Highly nonrandom features of synaptic connectivity in local cortical circuits", PLoS biology **3**, e68 (2005) 10.1371/journal.pbio.0030068.

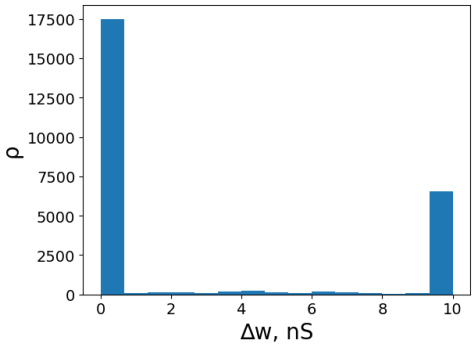
[18]D. D. Stettler and R. Axel, "Representations of odor in the piriform cortex", Neuron **63**, 854–864, ISSN: 0896-6273 (2009) https://doi.org/10.1016/j.neuron.2009.09.005.

[19]T. P. Vogels and L. Abbott, "Signal propagation and logic gating in networks of integrate-and-fire neurons", The Journal of neuroscience : the official journal of the Society for Neuroscience **25**, 10786–95 (2005) 10.1523/JNEUROSCI.3508-05.2005.

Appendix A. STDP

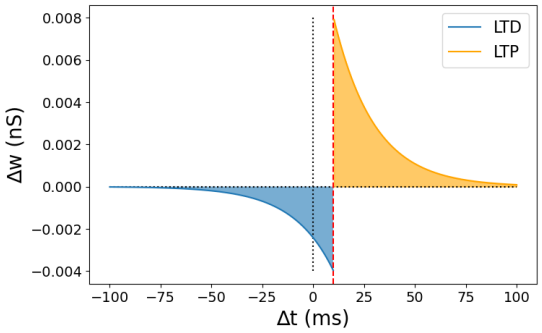


(a) **STDP rule with 0 ms shift.** The x-axis corresponds to the time difference between post- and presynaptic spikes. The y-axis corresponds to a change in the conductance of a synapse. Potentiation and depression regions are color-coded in yellow and blue correspondingly.

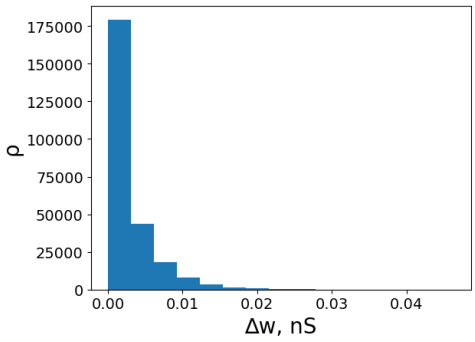


(b) **The resulting distribution of weights in excitatory population.** The x-axis corresponds to the conductance between excitatory neurons. The y-axis depicts the density.

Figure A.5: STDP rule without a time shift and resulting weights distribution. The results show that synaptic weights follow bimodal distribution after the training.



(a) **STDP rule with 10 ms shift.** The x-axis corresponds to the time difference between post- and presynaptic spikes. The y-axis corresponds to a change in the conductance of a synapse. Potentiation and depression regions are color-coded in yellow and blue correspondingly.



(b) **The resulting distribution of weights in excitatory population.** The x-axis corresponds to the conductance between excitatory neurons. The y-axis depicts the density.

Figure A.6: STDP learning rule with 10 ms shift and resulting weights distribution. The results show that all of the synapses were silenced during the training resulting in a weakly connected excitatory population.