On the Self-Repair Role of Astrocytes in STDP Enabled Unsupervised SNNs

Mehul Rastogi, Sen Lu, Nafiul Islam, Abhronil Sengupta

arXiv:2009.03473v2 [cs.NE] 16 Nov 2020

***Abstract*—Neuromorphic computing is emerging to be a disruptive computational paradigm that attempts to emulate various facets of the underlying structure and functionalities of the brain in the algorithm and hardware design of next-generation machine learning platforms. This work goes beyond the focus of current neuromorphic computing architectures on computational models for neuron and synapse to examine other computational units of the biological brain that might contribute to cognition and especially self-repair. We draw inspiration and insights from computational neuroscience regarding functionalities of glial cells and explore their role in the fault-tolerant capacity of Spiking Neural Networks (SNNs) trained in an unsupervised fashion using Spike-Timing Dependent Plasticity (STDP). We characterize the degree of self-repair that can be enabled in such networks with varying degree of faults ranging from 50% - 90% and evaluate our proposal on the MNIST and Fashion-MNIST datasets.**

***Index Terms*—Spiking Neural Networks, Astrocytes, Spike-Timing Dependent Plasticity, Unsupervised learning**

I. Introduction

Neuromorphic computing has made significant strides over the past few years - both from hardware [1]–[4] and algorithmic perspective [5]–[8]. However, the quest to decode the operation of the brain have mainly focused on spike based information processing in the neurons and plasticity in the synapses. Over the past few years, there has been increasing evidence that glial cells, and in particular astrocytes, play a crucial role in a multitude of brain functions [9]. As a matter of fact, astrocytes represent a large proportion of the cell population in the human brain [9]. There have been also suggestions that complexity of astrocyte functionality can significantly contribute to the computational power of the human brain. Astrocytes are strategically positioned to ensheath tens of thousands of synapses, axons and dendrites among others, thereby having the capability to serve as a communication channel between multiple components and behave as a sensing medium for ongoing brain activity [10]. This has led neuroscientists to conclude that astrocytes play a major role in higher order brain functions like learning and memory, in addition to neurons and synapses. Over the past few years, there have been multiple studies to revise the neuron-circuit model for describing higher order brain functions to incorporate astrocytes as part of the neuron-glia network model [9], [11]. These investigations clearly indicate and quantify that incorporating astrocyte functionality in network models influence neuron excitability, synaptic strengthening and, in turn, plasticity mechanisms like Short-Term Plasticity and Long-Term Potentiation, which are important learning tools used by neuromorphic engineers.

The key distinguishing factors of our work against prior efforts can be summarized as follows:
(i) While recent literature reports astrocyte computational models and their impact on fault-tolerance and synaptic learning [9], [11]–[14], the studies have been mostly confined to small scale networks. This work is a first attempt to explore the self-repair role of astrocytes at scale in unsupervised SNNs in standard visual recognition tasks.
(ii) In parallel, there is a long history of implementing astrocyte functionality in analog and digital CMOS implementations [15]–[21]. More recently, emerging physics in post-CMOS technologies like spintronics are also being leveraged to mimic glia functionalities at a one-to-one level [22]. However, the primary focus has been on a brain-emulation perspective, i.e. implementing astrocyte computational models with high degree of detail in the underlying hardware. We explore the aspects of astrocyte functionality that would be relevant to self-repair in the context of SNN based machine learning platforms and evaluate the degree of bio-fidelity required.
(iii) While Refs. [23], [24] discusses impact of faults in unsupervised STDP enabled SNNs, self-repair functionality in such networks have not been studied previously.

While neuromorphic hardware based on emerging post-CMOS technologies [3], [25]–[28] have made significant advancements to reduce the area and power efficiency gap of Artificial Intelligence (AI) systems, such emerging hardware are characterized by a host of non-idealities which has greatly limited its scalability. Our work provides motivation toward autonomous self-repair of such faulty neuromorphic hardware platforms. The efficacy of our proposed astrocyte enabled self-repair process is measured by the following steps: **(i)** Training SNNs using unsupervised STDP learning rules in networks equipped with lateral inhibition and homeostasis, **(ii)** Introducing "faults"[1] in the trained weight maps by setting a randomly chosen subset of the weights to zero and **(iii)** Implementing self-repair by re-training the faulty network with astrocyte functionality augmented STDP learning rules. We

Manuscript dated September, 2020.

The authors are with the School of Electrical Engineering and Computer Science, The Pennsylvania State University, University Park, PA 16802, USA. M. Rastogi is also affiliated with Department of of Computer Science and Information Systems, Birla Institute of Technology and Science Pilani, Goa Campus, Goa 403726, India. E-mail: sengupta@psu.edu.

[1]Note that "faults" are disjoint from the concept of "dropout" [29] used in neural network training. In dropout, neurons are randomly deleted (along with their connections) only during training to avoid overfitting. In contrast, faults in our work refer to static non-ideal stuck at zero synaptic connections present during both the training and inference stages.

also compare our proposal with sole STDP based re-training strategy and substantiate our results on the MNIST and F-MNIST datasets.

II. Materials and Methods

A. Astrocyte Preliminaries

In addition to astrocyte mediated meta-plasticity for learning and memory [12], [30]–[32], there has been indication that retrograde signalling via astrocytes probably underlie self-repair in the brain. Computational models demonstrate that when faults occur in synapses corresponding to a particular neuron, indirect feedback signal (mediated through retrograde signalling by the astrocyte via endocannabinoids, a type of retrograde messenger) from other neurons in the network implements repair functionality by increasing the transmission probability across all healthy synapses for the affected neuron, thereby restoring the original operation [12]. Astrocytes modulate this synaptic transmission probability (PR) through two feedback signalling pathways: direct and indirect, responsible for synaptic depression (DSE) and potentiation (e-SP) respectively. Multiple astrocyte computational models [12], [30]–[32] describe the interaction of astrocytes and neurons via the tripartite synapse where the astrocyte's sensitivity to 2-arachidonyl glycerol (2-AG), a type of endocannabinoid, is considered. Each time a post synaptic neuron fires, 2-AG is released from the post synaptic dendrite and can be described as:

$$\frac{d(\mathrm{AG})}{dt} = \frac{-\mathrm{AG}}{\tau_{\mathrm{AG}}} + r_{\mathrm{AG}}\delta(t - t_{\mathrm{sp}}) \tag{1}$$

where, AG is the quantity of 2-AG, τ_{AG} is the decay rate of 2-AG, r_{AG} is the 2-AG production rate and t_{sp} is the time of the post-synaptic spike.

The 2-AG binds to receptors (CB1Rs) on the astrocyte process and instigates the generation of IP_3, which subsequently binds to IP_3 receptors on the Endoplasmic Reticulum (ER) to open channels that allow the release of Ca^{2+}. It is this increase in cystolic Ca^{2+} that causes the release of gliotransmitters into the synaptic cleft that is ultimately responsible for the indirect signaling. The Li-Rinzel model [33] uses three channels to describe the Ca^{2+} dynamics within the astrocyte: J_{pump} models how Ca^{2+} is pumped into the ER from the cytoplasm via the Sarco-Endoplasmic-Reticulum Ca^{2+}-ATPase (SERCA) pumps, J_{leak} describes Ca^{2+} leakage into the cytoplasm and J_{chan} models the opening of Ca^{2+} channels by the mutual gating of Ca^{2+} and IP_3 concentrations. The Ca^{2+} dynamics is thus given by:

$$\frac{d\mathrm{Ca}^{2+}}{dt} = J_{\mathrm{chan}} + J_{\mathrm{leak}} - J_{\mathrm{pump}} \tag{2}$$

The details of the equations and their derivations can be obtained from Refs. [12] and [34].

The intracellular astrocytic calcium dynamics control the glutamate release from the astrocyte which drives e-SP. This release can be modelled by:

$$\frac{d(\mathrm{Glu})}{dt} = \frac{-\mathrm{Glu}}{\tau_{\mathrm{Glu}}} + r_{\mathrm{Glu}}\delta(t - t_{\mathrm{Ca}}) \tag{3}$$

where, Glu is the quantity of glutamate, τ_{Glu} is the glutamate decay rate, r_{Glu} is the glutamate production rate and t_{Ca} is the time of the Ca^{2+} threshold crossing. To model e-SP:

$$\tau_{\mathrm{eSP}}\frac{d(\mathrm{eSP})}{dt} = -\mathrm{eSP} + m_{\mathrm{eSP}}\mathrm{Glu}(t) \tag{4}$$

where, τ_{eSP} is the decay rate of e-SP and m_{eSP} is a scaling factor. Eq. (4) substantiates that the level of e-SP is dependent on the quantity of glutamate released by the astrocyte.

The released 2-AG also binds directly to pre-synpatic CB1Rs (direct signaling). A linear relationship is assumed between DSE and the level of 2-AG released by the post-synaptic neuron as:

$$\mathrm{DSE} = -\mathrm{AG} \times K_{\mathrm{AG}} \tag{5}$$

where, AG is the amount of 2-AG released by the post-synaptic neuron and is found from Eq. (1) and K_{AG} is a scaling factor. The PR associated with each synapse is given by the following equation:

$$\mathrm{PR}(t) = \mathrm{PR}(t_0) + \mathrm{PR}(t_0) \times \left(\frac{\mathrm{DSE}(t) + \mathrm{eSP}(t)}{100}\right) \tag{6}$$

where, $\mathrm{PR}(t_0)$ is the initial PR of the synapses, e-SP and DSE are given by Eq. (4) and (5) respectively. In the computational models, the effect of DSE is local to the synapses connected to a particular neuron whereas all the tripartite synapses connected to the same astrocyte receive the same e-SP. Under no-fault condition, the DSE and e-SP reach a dynamic equilibrium where the PR is unchanged over time, resulting in a fixed firing rate for the neurons. When a fault occurs, this balance subsides and the PR changes according to Eq. (6) to restore the firing rate to its previous value. To showcase this effect consider for instance, Fig. 1 where a simple SNN with two post-synaptic neurons is depicted. Let us assume that each post-neuron receives input spikes from 10 pre-neurons. The initial PR of the synapses were set to 0.5. Fig. 1(a) is the case with no faults, while in Fig. 1(b), faults have occurred after some time in 70% of the synapses associated with post-neuron N2 (Fig. 2). Note, here "faults" imply that the synapses do not take part in transmission of the input spikes i.e. have a PR of

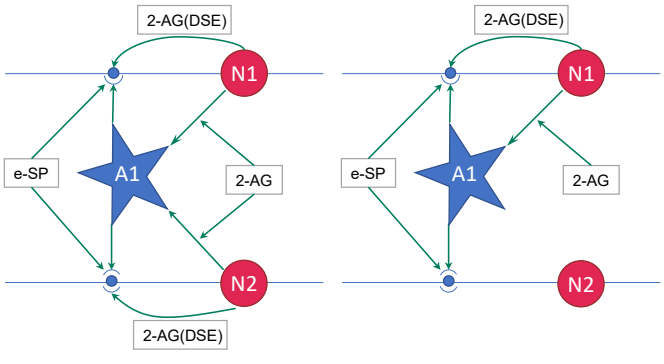
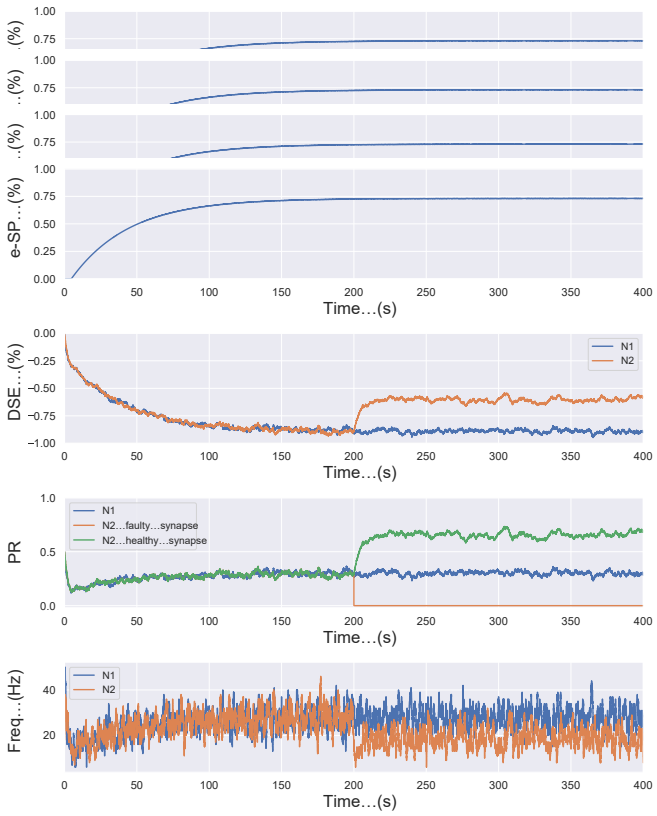


Fig. 1. (a) Network with no faults, (b) Network with fault occurring in synapse associated with neuron N2 [12]. 2-AG is local signal associated with each synapse while e-SP is a global signal. A1 is the astrocyte.



(d)

Fig. 2. Simulation results of the network in Fig. 1 using the computational model of astrocyte mediated self-repair from [12]. Total simulation time is 400s. At 200s, faults are introduced in 70% of the synapses connected to N2. All the synapses have PR(t_0)=0.5. (a) e-SP of N1 and N2. It is the same for both N1 and N2 since e-SP is a global function, (b) DSE of N1 and N2. It is different for each neuron as it is dependent upon the neuron output. At 200s, after the introduction of the faults in N2, only DSE of N2 changes, (c) PR of different types of synapses connected to N1 and N2, and (d) Firing rate of neurons N1 and N2.

0. This results in a drop of the firing frequency associated with N2 while operation of N1 is not impacted. Thus, the amount of 2-AG released by N2 decreases, which increases DSE and in turn increases the PR of the associated synapses of N2 where no faults have occurred. Hence, we observe in Fig. 2(d) that the increased PR recovers the firing rate and approaches the ideal firing frequency. Note that the degree of self-recovery, i.e. the difference between the recovered and ideal frequency is a function of the fault probability. The simulation conditions and parameters for the modelling are based on Ref. [12]. Interested readers are directed to Ref. [12] for an extensive discussion on the astrocyte computational model and the underlying processes governing the retrograde signalling.

A key question that we have attempted to address in this work is the computational complexity at which we require to model the feedback mechanism to implement autonomous repair in such self-learning networks. Simplifying the feedback modelling would enable us to implement such functionalities by efficient hardware primitives. For instance, the core functionality of astrocyte self-repair occurs in conjunction with STDP based learning in synapses. Fig. 3 shows a typical STDP learning rule where the change in synaptic weight varies exponentially with the spike time difference between the pre- and post-neuron [35], according to measurements performed in rat glutamatergic synapses [36]. Typically, the height of the STDP weight update for potentiation/depression is constant (A_+/A_-). However, astrocyte mediated self-repair suggests that the weight update should be a function of the firing rate of the post-neuron [35]. Assuming the fault-less expected firing rate of the post-neuron to be f_{ideal} and the non-ideal firing rate to be f, the synaptic weight update window height should be a function of $\Delta f = f_{ideal} - f$. The concept has been explained further in Fig. 3 and is also in accordance with Fig. 2 where the PR increase after fault introduction varies in a non-linear fashion over time and eventually stabilizes once the self-repaired firing frequency approaches the ideal value. The functional dependence is assumed to be that of a sigmoid function – indicating that as the magnitude of the fault, i.e. deviation in the ideal firing frequency of the neuron increases, the height of the learning window increases in proportion to compensate for the fault [35]. Note that the term "fault" for the machine learning workloads, described herein, refers to synaptic weights (symbolizing PR) stuck at zero. Therefore, with increasing amount of synaptic faults, $f << f_{ideal}$, thereby increasing the STDP learning window height significantly. During the self-healing process, the frequency deviation gradually reduces and thereby the re-learning rate also becomes less pronounced and finally saturates once the ideal frequency is reached. While our proposal is based on Ref. [35], prior work has been explored in the context of a prototype artificial neural network with only 4 input neurons and 4 output neurons. Extending the framework to an unsupervised SNN based machine learning framework therefore requires significant explorations, highlighted next.

B. Neuron Model and Synaptic Plasticity

We utilized the Leaky Integrate and Fire (LIF) spiking neuron model in our work. The temporal LIF neuron dynamics are described as,

$$\tau_{mem}\frac{\partial v(t)}{\partial t} = -v(t) + v_{rest} + I(t) \tag{7}$$

where, $v(t)$ is the membrane potential, τ_{mem} is the membrane time constant, v_{rest} is the resting potential and $I(t)$ denotes the total input to the neuron at time t. The weighted summation of synaptic inputs is represented by $I(t)$. When the neuron's

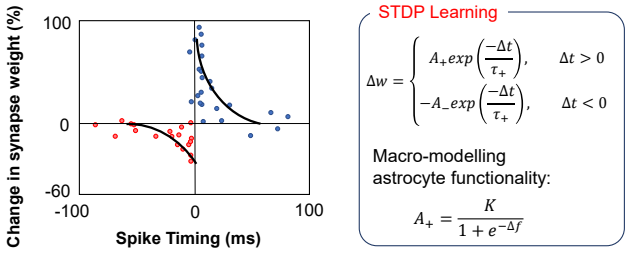


Fig. 3. In the above equations, the STDP learning window height is a non-linear increasing function of the deviation Δf from the ideal firing frequency of the post-neuron.

membrane potential crosses a threshold value, $v_{th}(t)$, it fires an output spike and the membrane potential is reset to v_{reset}. The neuron's membrane voltage is fixed at the reset potential for a refractory period, δ_{ref}, after it spikes during which it does not receive any inputs.

In order to ensure that single neurons do not dominate the firing pattern, homeostasis [6] is also implemented through an adaptive thresholding scheme. The membrane threshold of each neuron is given by the following temporal dynamics,

$$\begin{aligned} v_{th}(t) &= \theta_0 + \theta(t) \\ \tau_{theta}\frac{\partial\theta(t)}{\partial t} &= -\theta(t) \end{aligned} \tag{8}$$

where, $\theta_0 > v_{rest}, v_{reset}$ and is a constant. τ_{theta} is the adaptive threshold time constant. The adaptive threshold, $\theta(t)$ is increased by a constant quantity θ_+, each time the neuron fires, and decays exponentially according to the dynamics in Equation 8.

A trace [37] based synaptic weight update rule was used for the online learning process [6], [23]. The pre and post-synaptic traces are given by x_{pre} and x_{post} respectively. Whenever the pre (post) - synaptic neuron fires, the variable x_{pre} (x_{post}) is set to 1, otherwise it decays exponentially to 0 with spike trace decay time constant, τ_{trace}. The STDP weight update rule is characterized by the following dynamics,

$$\Delta w = \begin{cases} \eta_{post} * x_{pre} & \text{on post-synaptic spike} \\ -\eta_{pre} * x_{post} & \text{on pre-synaptic spike} \end{cases} \tag{9}$$

where, η_{pre}/η_{post} denote the learning rates for pre-synaptic / post-synaptic updates respectively. The weights of the neurons are bounded in the range of $[0, w_{max}]$. It is worth mentioning here that the sum of the weights associated with all post-synaptic neurons is normalized to a constant factor, w_{norm} [23].

C. Network Architecture

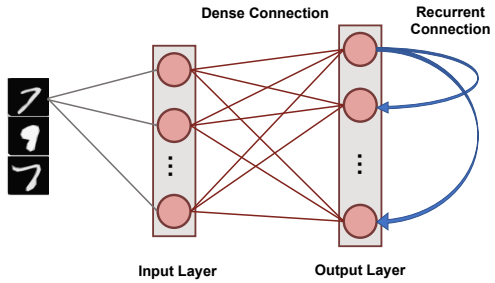


Fig. 4. The single layer SNN architecture with lateral inhibition and homeostasis used for unsupervised learning.

Our SNN based unsupervised machine learning framework is based on single layer architectures inspired from cortical microcircuits [6]. Fig. 4 shows the network connectivity of spiking neurons utilized for pattern-recognition problems. Such a network topology has been shown to be efficient in several pattern-recognition problems, such as digit recognition [6] and sparse encoding [38]. The SNN, under consideration, has an Input Layer with the number of neurons equivalent to the dimensionality of the input data. Input neurons generate spikes by converting each pixel in the input image to a Poisson spike train whose average firing frequency is proportional to the pixel intensity. This layer connects in an all-to-all fashion to the Output Layer through excitatory synapses. The Output layer has $n_{neurons}$ LIF neurons characterized by homeostasis functionality. It also has static (constant weights) recurrent inhibitory synapses with weight values, $w_{recurrent}$, for lateral inhibition to achieve soft Winner-Take-All (WTA) condition. Each neuron in the Output Layer has an inhibitory connection to all the neurons in that layer except itself. Trace-based STDP mechanism is used to learn the weights of all synapses between the Input and Output Layers. The neurons in the Output Layer are assigned classes based on their highest response (spike frequency) to input training patterns [6].

D. Challenges and Astrocyte Augmented STDP (A-STDP) Learning Rule Formulation

One of the major challenges in extending the astrocyte based macro-modelling in such self-learning networks lies in the fact that the ideal neuron firing frequency is a function of the specific input class the neuron responds to. This is substantiated by Fig. 5 which depicts the histogram distribution of the ideal firing rate of the wining neuron in the fault-less network. Further, due to sparse neural firing, the total number of output spikes of the winning neurons over the inference window is also small, thereby limiting the amount of information (number of discrete levels) that can be encoded in the frequency deviation, Δf. This leads to the question: *Can we utilize another surrogate signal that gives us information about the degree of self-repair occurring in the network over time while being independent of the class of the input data?*

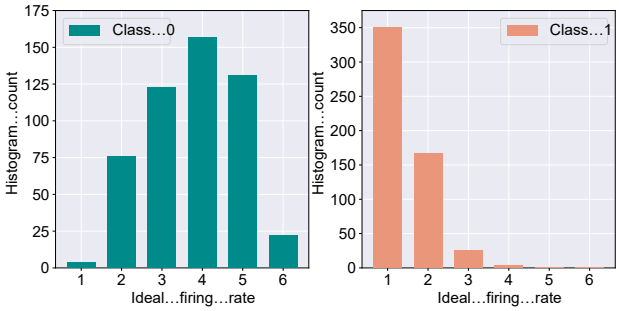


Fig. 5. Histogram count of the ideal firing rate of neurons responding to digit '0' versus digit '1' (measured from 5000 test examples of the MNIST dataset).

While the above challenge is related to the process of reducing the STDP learning window over time, we observed that using sole STDP learning or with a constant enhanced learning rate consistently reduced the network accuracy over time (Fig. 7). Fig. 8 also depicts that normal STDP retraining with faulty synapses slowly loses their learnt representations over time. Re-learning all the healthy synaptic weights uniformly using STDP with an enhanced learning rate should at least result in some accuracy improvement for the initial epochs of re-training, even if the modulation of learning window height over time is not incorporated in the self-repair

framework. The degradation of network accuracy starting from the commencement of the retraining process signified that some additional factors may have been absent in the astrocyte functionality macro-modelling process, which is independent from the above challenge of modulating the temporal behavior of the STDP learning window.

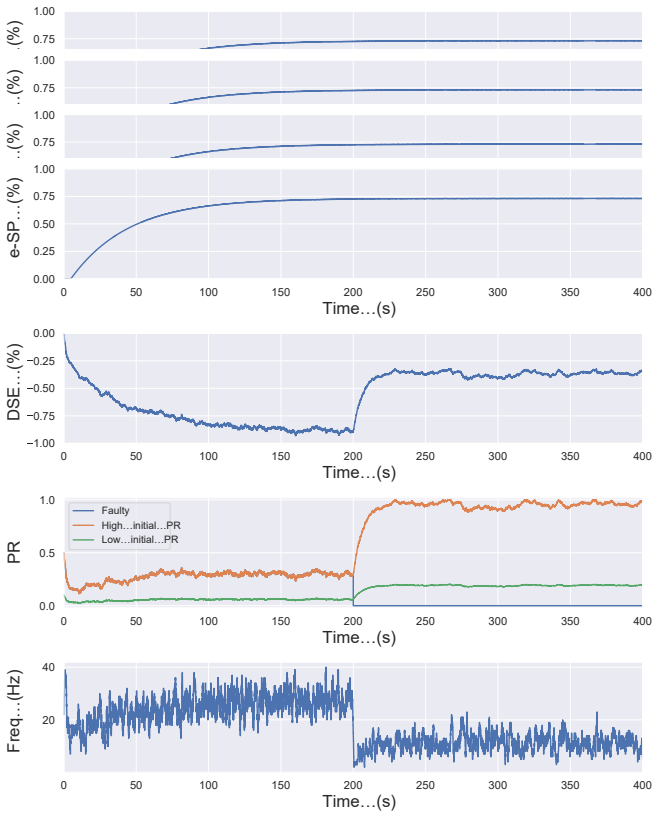
In that regard, we draw inspiration from Eq. 6, where we observe that the initial fault-free value of the PR acts as a scaling factor for the self-repair feedback terms DSE and e-SP. We perform a similar simulation for the network shown in Fig. 1, with each neuron receiving input from 10 synapses. However in this case, we set the initial PR of all of the synapses to 0.5, except one connected to N2; for which the initial PR was set to 0.1. In other words, 9 of the synapses connected to N2 have a PR(t_0)=0.5, while for one PR(t_0)=0.1. The lower initial PR value symbolizes a weaker connection. The network is simulated for 400s and at 200s, the associated PR of 8 of the synapses with higher initial PR are reduced to 0 to signify faulty condition (Fig. 6). We observe that after the introduction of the faults, the PR of the synapses with the higher initial PR value is enhanced greatly compared to the one with the lower initial PR. This leads us to the conclusion that synapses that play a greater role in postsynaptic firing also play a greater role in the self-repair process compared to other synapses.

Since our unsupervised SNN is characterized by analog synaptic weights in the range of $[0, w_{max}]$, we hypothesized that this characteristic might underlie the reason for the accuracy degradation and designed a preferential self-repair learning rule for healthier synapses. This was found to result in significant accuracy improvement during the retraining process (discussed in next section). Our formulated A-STDP learning rule formulation is therefore also guided by the following question: *Can we aggressively increase the healthy synaptic weights during the initial learning epochs which preserves the original representations learnt by the network?*

Driven by the above observations, we formulated our Astrocyte Augmented STDP (A-STDP) learning rule during the self-repair process as,

$$\Delta w = \begin{cases} \eta_{post} * x_{pre} * (\boldsymbol{w}/\boldsymbol{w_\alpha})^{\boldsymbol{\sigma}} & \text{on post-synaptic spike} \\ -\eta_{pre} * x_{post} & \text{on pre-synaptic spike} \end{cases} \tag{10}$$

where, w_α represents the the weight value at the α-th percentile of the network and serves as the surrogate signal to guide the retraining process. Fig. 9 depicts the temporal behavior of w_α for the 98-th percentile of the weight distribution. After faults are introduced, w_α is significantly reduced and slowly increases over time during the re-learning process. It finally saturates off at the bounded value w_{max}. The term w/w_α ensures that the effective learning rate for healthier synapses ($w > w_\alpha$) is much higher than the learning rate for weaker connections ($w < w_\alpha$) while σ dictates the degree of non-linearity. Since w_α increases over time, the enhanced learning process also reduces and finally stops once w_α saturates. It is worth mentioning here that w_α, σ and w_{max} are hyperparameters for the A-STDP learning rule. All



(d)

Fig. 6. Simulation results of the network in Fig. 1 using the computational model of [12] with synapses having different initial PR values. Total simulation time is 400s. At 200s, faults are introduced in 8 synapses with high initial PR connected to N2. (a) e-SP of N1 and N2, (b) DSE of N2, (c) PR of the 3 types of synapses connected to N2 (orange: healthy synapse with PR(t_0)=0.5, green: healthy synapse with PR(t_0)=0.1 and blue: faulty synapse with PR(t_0)=0.5 till 200s and PR(t_0)=0 afterwards) and (d) Firing rate of neuron N2

.

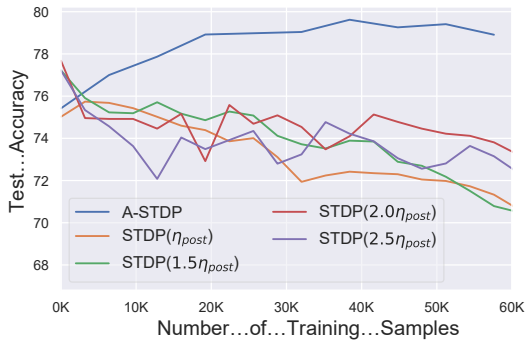
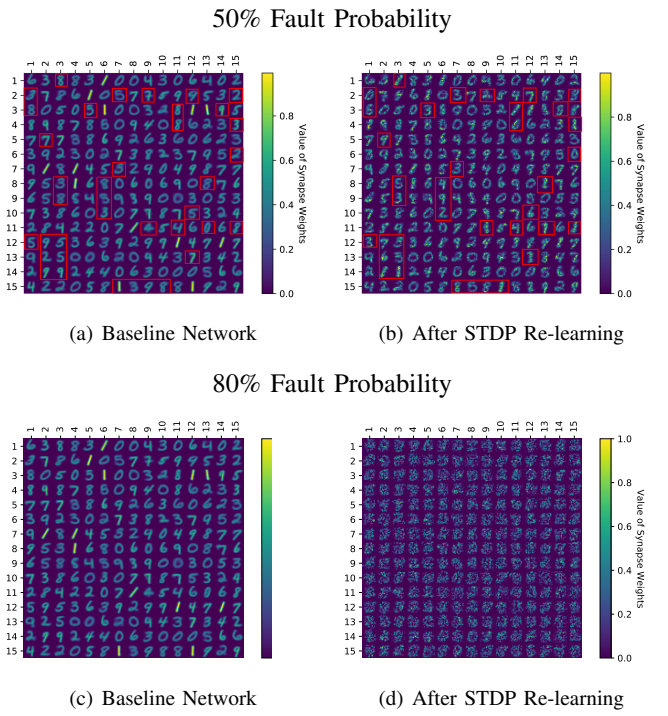


Fig. 7. Test accuracy of a 225 neuron network on the MNIST dataset with 70% faulty connections with normal and enhanced learning rates during STDP re-training process. Re-training with A-STDP rule is also depicted.

hyperparameter settings and simulation details are presented in the next section.

III. Results

We evaluated our proposal in the context of unsupervised SNN training on standard image recognition benchmarks under two settings: scaling in network size and scaling in network



(a) Baseline Network (b) After STDP Re-learning

(c) Baseline Network (d) After STDP Re-learning

Fig. 8. (a)-(d) Learnt weight patterns for 225 neuron network on the MNIST dataset are shown. Re-training the network with sole STDP learning causes distortion of the weight maps (50% and 80% fault cases are plotted). The red boxes in (a) and (b) highlight how the the neurons can change association toward a particular class during re-learning thereby forgetting their original learnt representations. Receptive fields of all neurons undergo distortion for the 80% fault case.

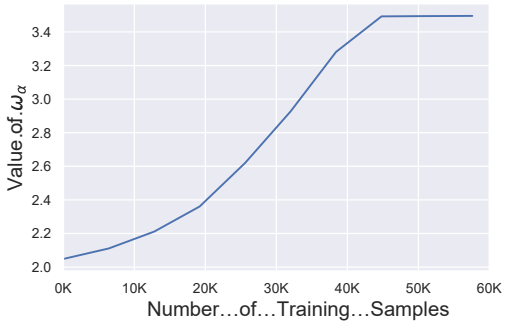
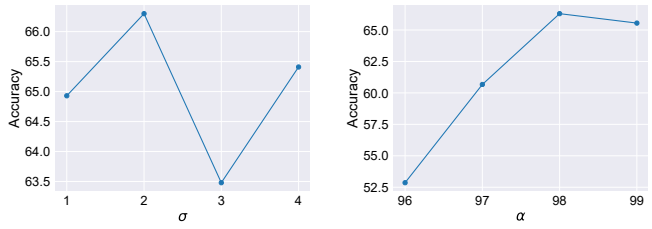


Fig. 9. Value of w_α (98-th percentile from weight distribution of the entire network) during the self-repair process using A-STDP learning rule for a 225 neuron network on the MNIST Dataset with 80% faulty connections.

complexity. We used MNIST [39] and Fashion-MNIST [40] datasets for our analysis. Both datasets contain 28 $\times$ 28 grayscale images of handwritten digits / fashion products (belonging to one of 10 categories) with 60,000 training examples and 10,000 testing examples. All experiments are run in PyTorch framework using a single GPU with a batchsize of 16 images. In addition to standard input pre-processing for generating the Poisson spike train, the images in F-MNIST dataset also undergo Sobel filtering for edge detection before being converted to spike trains. The SNN implementation is done using a modified version of the mini-batch processing enabled SNN simulation framework [41] in BindsNET [42], a PyTorch based package (Link). In addition to dataset complexity scaling, we also evaluated two networks with increasing size (225 and 400 neurons) on the MNIST dataset. For the MNIST dataset, the baseline test accuracy of the ideal network was 89.53% and 92.02% respectively. A 400-neuron network was used for the F-MNIST dataset with 77.35% accuracy. The baseline test accuracies are superior/comparable to prior reported accuracies for unsupervised learning on both datasets. For instance, Ref. [6] reports 87% accuracy for an STDP trained network with 400 neurons while Ref. [43] reports the best accuracy of 73.1% for state-of-the-art clustering methods on the F-MNIST dataset. Table I lists the network simulation parameters used in this work. It is worth mentioning here that all hyperparameters were kept unchanged (from their initial values during training) in the self-repair process. We also kept the hyperparameters, w_α and σ for the A-STDP rule unchanged for all fault simulations. Fig. 10 shows a typical ablation study of the hyperparameters α and σ. For this study, we trained a 225-neuron network with 90% faults. We divided the training set into training and validation subsets in the ratio of 5:1 respectively through random sampling. The two accuracy plots shown in Fig. 10 are models retrained on the training subset and then evaluated on the new validation set. Further hyperparameter optimizations for different fault conditions can potentially improve the accuracy improvement even further.



(a) (b)

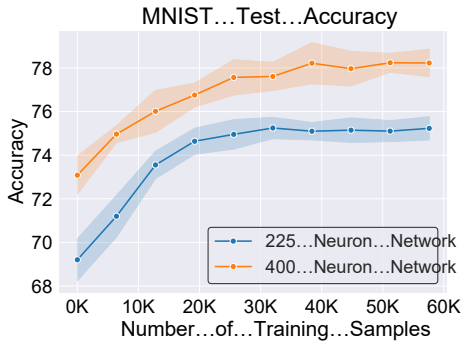
Fig. 10. Ablation studies for the hyperparameters (a) σ (with fixed $\alpha = 98$) and (b) α (with fixed $\sigma = 2$) in A-STDP learning rule.

The network is first trained with sole STDP learning rule for 2 epochs and the maximum test accuracy network is chosen as the baseline model. Subsequently, faults are introduced by randomly deleting synapses (from the Input to the Output Layer) post-training. Each synaptic connection was assigned a deletion probability, p_{del}, to decide whether the connection would be retained in the faulty network. For this work, p_{del} was varied between 0.5 - 0.9 to analyze the network and re-train after introducing faults. Note that A-STDP learning rule is only used during this self-repair phase. It is worth mentioning here, that weight normalization by factor w_{norm} (mentioned in Section III-B) is used before starting the re-training process. This helps to adjust the relative magnitude of firing threshold relative to the weights of the neurons (since the resultant magnitude diminishes due to fault injection).

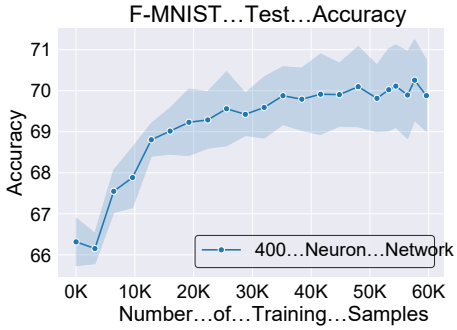
Fig. 11 shows the test classification accuracy as a function of re-learning epochs for a 225 / 400 neuron network with 80% probability for faulty synapses. After the faults are

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Membrane Time Constant, τ_{mem}	100ms
Spike Trace Decay Time Constant, τ_{trace}	20ms
Resting Potential, v_{rest}	-65mV
Threshold Voltage Constant, θ_0	-52mV
Membrane Reset Potential, v_{reset}	-60mV
Refractory Period, δ_{ref}	5ms
Adaptive Threshold Time Constant, τ_{theta}	10^7ms
Adaptive Threshold Voltage Increment, θ_+	0.05
Post-Synaptic Learning Rate, η_{post}	10^{-2} (MNIST) 4×10^{-3} (F-MNIST)
Pre-Synaptic Learning Rate, η_{pre}	10^{-4} (MNIST) 4×10^{-5} (F-MNIST)
Normalization Factor, w_{norm}	78.4
No. of Excitatory Neurons, $n_{neurons}$	225 / 400 (MNIST) 400 (F-MNIST)
Static Inhibitory Synaptic Weight, $w_{recurrent}$	-120 (MNIST) -250 (F-MNIST)
A-STDP Weight-Percentile Hyperparameter, α	98
A-STDP Non-Linearity Hyperparameter, σ	2



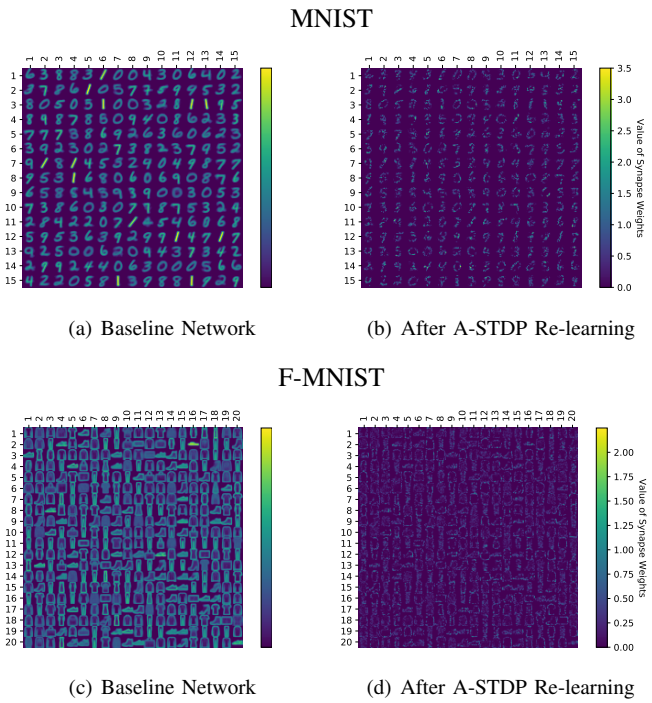
(a)



(b)

Fig. 11. Improvement of test accuracy during re-learning is depicted as a function of the training samples using A-STDP learning rule on the (a) MNIST (225 and 400 neuron networks) and (b) F-MNIST datasets (400 neuron network). Mean and standard deviation of the accuracy is plotted for 80% fault simulation in the networks.

introduced, the network accuracy improves over time during the self-repair process. The mean and standard deviation of test accuracy from 5 independent runs are plotted in Fig. 11. Fig. 12 depicts the initial and self-repaired weight maps of the 225 (MNIST) and 400 (F-MNIST) neuron networks, substantiating that original learnt representations are preserved



(a) Baseline Network (b) After A-STDP Re-learning

(c) Baseline Network (d) After A-STDP Re-learning

Fig. 12. (a-d) Initial and self-repaired weight maps of the 225 (400) neuron network trained on MNIST (F-MNIST) dataset corresponding to 80% fault simulations.

during the re-learning process. Table II summarizes our results for all networks with varying degrees of faults. The numbers in parentheses denote the standard deviation in accuracy from the 5 independent runs. Since sole STDP learning resulted in accuracy degradation for most of the runs, the accuracy is reported after 1 re-learning epoch. For some cases, some accuracy improvement through normal STDP was also observed. The maximum accuracy is reported for the A-STDP re-training process. After repair through A-STDP, the network is able to achieve accuracy improvement across all level of faults, ranging from 50% - 90%. Interestingly, A-STDP is able to repair faults even in a 90% faulty network and improve the testing accuracy by almost 9% (5%) for the MNIST (F-MNIST) dataset. Further, the accuracy improvement due to A-STDP scales up with increasing degree of faults. Note that the standard deviation of the final accuracy over 5 independent runs is much smaller for A-STDP than normal STDP re-training, signifying that the astrocyte enabled self-repair is consistently stable, irrespective of the initial fault locations.

IV. DISCUSSION

The work provides proof-of-concept results toward the development of a new generation of neuromorphic computing platforms that are able to autonomously self-repair faulty non-ideal hardware operation. Extending beyond just unsupervised STDP learning, augmenting astrocyte feedback in supervised gradient descent based training of SNNs needs to be explored along with their implementation on neuromorphic datasets [44]. In this work, we also focused on aspects of astrocyte operation that would be relevant from a macro-modelling perspective for self-repair. Further investigations on under-

TABLE II
SELF-REPAIR RESULTS FOR A-STDP ENABLED SNNS

Network Description	Fault Probability	Accuracy after fault injection (%)	Accuracy after weight normalization (%)	Accuracy after STDP re-training (%)	Accuracy after A-STDP re-training (%)	Accuracy Gain from A-STDP
MNIST Dataset						
225 Excitatory Neurons Baseline Accuracy = 89.53%	50%	75.41 ± 2.28	83.04 ± 0.49	76.69 ± 0.98	84.06 ± 0.70	1.02
	60%	70.72 ± 1.41	80.25 ± 0.70	73.85 ± 1.18	82.19 ± 0.28	1.95
	70%	61.86 ± 2.03	76.06 ± 1.22	70.57 ± 0.48	79.39 ± 1.13	3.33
	80%	30.42 ± 3.26	69.13 ± 0.85	67.42 ± 1.37	75.42 ± 0.48	6.29
	90%	9.92 ± 0.11	56.69 ± 1.45	61.40 ± 1.38	65.57 ± 1.28	8.89
400 Excitatory Neurons Baseline Accuracy = 92.02%	50%	79.32 ± 2.57	85.56 ± 0.24	80.96 ± 1.24	87.16 ± 0.12	1.59
	60%	73.01 ± 2.15	82.61 ± 0.22	79.12 ± 1.28	85.17 ± 0.33	2.56
	70%	61.20 ± 1.10	79.77 ± 0.61	77.51 ± 0.62	83.00 ± 0.40	3.22
	80%	30.18 ± 2.60	73.08 ± 0.87	73.26 ± 1.16	78.68 ± 0.73	5.58
	90%	9.80 ± 0.27	59.90 ± 1.16	67.80 ± 0.77	68.85 ± 0.48	8.95
Fashion-MNIST Dataset						
400 Excitatory Neurons Baseline Accuracy = 77.35%	50%	58.62 ± 1.24	73.85 ± 0.50	73.51 ± 0.30	75.88 ± 0.38	2.02
	60%	39.12 ± 1.19	71.85 ± 1.36	72.23 ± 0.60	75.16 ± 0.49	3.31
	70%	16.61 ± 0.69	70.21 ± 0.44	70.63 ± 0.70	73.14 ± 0.44	2.93
	80%	10.00 ± 0.22	66.32 ± 0.58	68.80 ± 0.47	70.82 ± 0.57	4.51
	90%	10.04 ± 0.28	60.24 ± 0.86	63.92 ± 0.77	65.49 ± 0.40	5.25

standing the role of neuroglia in neuromorphic computing can potentially forge new directions related to synaptic learning, temporal binding, among others.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AS developed the main concepts. MR, SL and NI performed all the simulations. All authors assisted in the writing of the paper and developing the concepts.

ACKNOWLEDGMENTS

The work was supported in part by the National Science Foundation grants BCS #2031632, ECCS #2028213 and CCF #1955815.

REFERENCES

[1] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, "A million spiking-neuron integrated circuit with a scalable communication network and interface," *Science*, vol. 345, no. 6197, pp. 668–673, 2014.

[2] M. Davies, N. Srinivasa, T. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, Y. Liao, C. Lin, A. Lines, R. Liu, D. Mathaikutty, S. McCoy, A. Paul, J. Tse, G. Venkataramanan, Y. Weng, A. Wild, Y. Yang, and H. Wang, "Loihi: A neuromorphic manycore processor with on-chip learning," *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.

[3] A. Sengupta and K. Roy, "Encoding neural and synaptic functionalities in electron spin: A pathway to efficient neuromorphic computing," *Applied Physics Reviews*, vol. 4, no. 4, p. 041105, 2017.

[4] S. Singh, A. Sarma, N. Jao, A. Pattnaik, S. Lu, K. Yang, A. Sengupta, V. Narayanan, and C. R. Das, "Nebula: A neuromorphic spin-based ultra-low power architecture for snns and anns," in *2020 ACM/IEEE 47th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 2020, pp. 363–376.

[5] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks," *IEEE Signal Processing Magazine*, vol. 36, pp. 61–63, 2019.

[6] P. U. Diehl and M. Cook, "Unsupervised learning of digit recognition using spike-timing-dependent plasticity," *Frontiers in computational neuroscience*, vol. 9, p. 99, 2015.

[7] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, "Going deeper in spiking neural networks: VGG and residual architectures," *Frontiers in neuroscience*, vol. 13, 2019.

[8] S. Lu and A. Sengupta, "Exploring the connection between binary and spiking neural networks," *arXiv preprint arXiv:2002.10064*, 2020.

[9] S. L. Allam, V. S. Ghaderi, J.-M. C. Bouteiller, A. Legendre, A. Nicolas, R. Greget, S. Bischoff, M. Baudry, and T. W. Berger, "A computational model to investigate astrocytic glutamate uptake influence on synaptic transmission and neuronal spiking," *Frontiers in computational neuroscience*, vol. 6, p. 70, 2012.

[10] W.-S. Chung, N. J. Allen, and C. Eroglu, "Astrocytes control synapse formation, function, and elimination," *Cold Spring Harbor perspectives in biology*, vol. 7, no. 9, p. a020370, 2015.

[11] R. Min, M. Santello, and T. Nevian, "The computational power of astrocyte mediated synaptic plasticity," *Frontiers in computational neuroscience*, vol. 6, p. 93, 2012.

[12] J. Wade, L. J. McDaid, J. Harkin, V. Crunelli, and S. Kelso, "Self-repair in a bidirectionally coupled astrocyte-neuron (AN) system based on retrograde signaling," *Frontiers in computational neuroscience*, vol. 6, p. 76, 2012.

[13] S. Y. Gordleeva, S. V. Stasenko, A. V. Semyanov, A. E. Dityatev, and V. B. Kazantsev, "Bi-directional astrocytic regulation of neuronal activity within a network," *Frontiers in computational neuroscience*, vol. 6, p. 92, 2012.

[14] M. De Pittà, V. Volman, H. Berry, V. Parpura, A. Volterra, and E. Ben-Jacob, "Computational quest for understanding the role of astrocyte signaling in synaptic transmission and plasticity," *Frontiers in computational neuroscience*, vol. 6, p. 98, 2012.

[15] M. Ranjbar and M. Amiri, "An analog astrocyte–neuron interaction circuit for neuromorphic applications," *Journal of Computational Electronics*, vol. 14, no. 3, pp. 694–706, 2015.

[16] G. Karimi, M. Ranjbar, M. Amirian, and A. Shahim-Aeen, "A neuromorphic real-time vlsi design of ca2+ dynamic in an astrocyte," *Neurocomputing*, vol. 272, pp. 197–203, 2018.

[17] R. K. Lee and A. C. Parker, "A CMOS circuit implementation of retrograde signaling in astrocyte-neuron networks," in *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. IEEE, 2016, pp. 588–591.

[18] M. Amiri, S. Nazari, and M. Janahmadi, "Digital configuration of astrocyte stimulation as a new technique to strengthen the impaired astrocytes in the tripartite synapse network," *Journal of Computational Electronics*, vol. 17, no. 3, pp. 1382–1398, 2018.

[19] S. Nazari, K. Faez, M. Amiri, and E. Karami, "A digital implementation of neuron–astrocyte interaction for neuromorphic applications," *Neural Networks*, vol. 66, pp. 79–90, 2015.

[20] J. Liu, J. Harkin, L. P. Maguire, L. J. McDaid, and J. J. Wade, "SPANNER: a self-repairing spiking neural network hardware architecture," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 4, pp. 1287–1300, 2017.

[21] Y. Irizarry-Valle, A. C. Parker, and J. Joshi, "A cmos neuromorphic approach to emulate neuro-astrocyte interactions," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2013, pp. 1–7.

[22] U. Garg, K. Yang, and A. Sengupta, "Emulation of astrocyte induced neural phase synchrony in spin-orbit torque oscillator neurons," *arXiv preprint arXiv:2007.00776*, 2020.

[23] D. J. Saunders, D. Patel, H. Hazan, H. T. Siegelmann, and R. Kozma, "Locally connected spiking neural networks for unsupervised feature learning," *Neural Networks*, vol. 119, pp. 332 – 340, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608019302333

[24] H. Hazan, D. Saunders, D. Sanghavi, H. Siegelmann, and R. Kozma, "Lattice map spiking neural networks (lm-snns) for clustering and classifying image data," *Annals of Mathematics and Artificial Intelligence*, 09 2019.

[25] B. L. Jackson, B. Rajendran, G. S. Corrado, M. Breitwisch, G. W. Burr, R. Cheek, K. Gopalakrishnan, S. Raoux, C. T. Rettner, A. Padilla, A. G. Schrott, R. S. Shenoy, B. N. Kurdi, C. H. Lam, and D. S. Modha, "Nanoscale electronic synapses using phase change devices," *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 9, no. 2, p. 12, 2013.

[26] D. Kuzum, R. G. Jeyasingh, B. Lee, and H.-S. P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," *Nano letters*, vol. 12, no. 5, pp. 2179–2186, 2011.

[27] S. H. Jo, T. Chang, I. Ebong, B. B. Bhadviya, P. Mazumder, and W. Lu, "Nanoscale memristor device as synapse in neuromorphic systems," *Nano letters*, vol. 10, no. 4, pp. 1297–1301, 2010.

[28] S. Ramakrishnan, P. E. Hasler, and C. Gordon, "Floating gate synapses with spike-time-dependent plasticity," *Biomedical Circuits and Systems, IEEE Transactions on*, vol. 5, no. 3, pp. 244–252, 2011.

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[30] V. Volman, E. Ben-Jacob, and H. Levine, "The astrocyte as a gatekeeper of synaptic information transfer," *Neural computation*, vol. 19, no. 2, pp. 303–326, 2007.

[31] S. Nadkarni and P. Jung, "Dressed neurons: modeling neural–glial interactions," *Physical biology*, vol. 1, no. 1, p. 35, 2004.

[32] ——, "Modeling synaptic transmission of the tripartite synapse," *Physical biology*, vol. 4, no. 1, p. 1, 2007.

[33] Y.-X. Li and J. Rinzel, "Equations for InsP3 receptor-mediated Ca^{2+} i oscillations derived from a detailed kinetic model: a Hodgkin-Huxley like formalism," *Journal of theoretical Biology*, vol. 166, no. 4, pp. 461–473, 1994.

[34] M. De Pittà, M. Goldberg, V. Volman, H. Berry, and E. Ben-Jacob, "Glutamate regulation of calcium and ip 3 oscillating and pulsating dynamics in astrocytes," *Journal of biological physics*, vol. 35, no. 4, pp. 383–411, 2009.

[35] J. Liu, L. J. McDaid, J. Harkin, S. Karim, A. P. Johnson, A. G. Millard, J. Hilder, D. M. Halliday, A. M. Tyrrell, and J. Timmis, "Exploring self-repair in a coupled spiking astrocyte neural network," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 3, pp. 865–875, 2018.

[36] G.-q. Bi and M.-m. Poo, "Synaptic modification by correlated activity: Hebb's postulate revisited," *Annual review of neuroscience*, vol. 24, no. 1, pp. 139–166, 2001.

[37] A. Morrison, M. Diesmann, and W. Gerstner, "Phenomenological models of synaptic plasticity based on spike timing," *Biological cybernetics*, vol. 98, pp. 459–78, 07 2008.

[38] P. Knag, J. K. Kim, T. Chen, and Z. Zhang, "A sparse coding neural network ASIC with on-chip learning for feature extraction and encoding," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 4, pp. 1070–1079, 2015.

[39] Y. LeCun and C. Cortes, "MNIST handwritten digit database," 2010. [Online]. Available: http://yann.lecun.com/exdb/mnist/

[40] H. Xiao, K. Rasul, and R. Vollgraf. (2017) Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms.

[41] D. J. Saunders, C. Sigrist, K. Chaney, R. Kozma, and H. T. Siegelmann, "Minibatch processing in spiking neural networks," *arXiv preprint arXiv:1909.02549*, 2019.

[42] H. Hazan, D. J. Saunders, H. Khan, D. Patel, D. T. Sanghavi, H. T. Siegelmann, and R. Kozma, "Bindsnet: A machine learning-oriented spiking neural networks library in python," *Frontiers in Neuroinformatics*, vol. 12, Dec 2018. [Online]. Available: http://dx.doi.org/10.3389/fninf.2018.00089

[43] Q. Zhu and Z. Wang, *Neural Processing Letters*, 06 2019.

[44] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," *Frontiers in Neuroscience*, vol. 9, p. 437, 2015. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2015.00437