Temporal Convolution in Spiking Neural Networks: a Bio-mimetic Paradigm

David Reid[1,2], Emanuele Lindo Secco[2]

[1]Artificial Intelligence Lab, School of Mathematics, Computer Science and Engineering, Liverpool Hope University, Hope Park, L16 9JD, UK

[2]Robotic Lab, School of Mathematics, Computer Science and Engineering, Liverpool Hope University, Hope Park, L16 9JD, UK

Abstract

Recent spectacular advances in *Artificial Intelligence* (AI), in large, be attributed to developments in *Deep Learning* (DL). In essence, DL is not a new concept. In many respects, DL shares characteristics of "traditional" types of *Neural Network* (NN). The main distinguishing feature is that it uses many more layers in order to learn increasingly complex features. Each layer convolutes into the previous by simplifying and applying a function upon a subsection of that layer. Deep Learning's fantastic success can be attributed to dedicated researchers experimenting with many different groundbreaking techniques, but also some of its triumph can also be attributed to fortune. It was the right technique at the right time.

To function effectively, DL mainly requires two things: (a) vast amounts of training data and (b) a very specific type of computational capacity. These two respective requirements have been amply met with the growth of the internet and the rapid development of GPUs. As such DL is an almost perfect fit for today's technologies.

However, DL is only a very rough approximation of how the brain works. More recently, *Spiking Neural Networks* (SNNs) have tried to simulate biological phenomena in a more realistic way. In SNNs information is transmitted as discreet spikes of data rather than a continuous weight or a differentiable activation function. In practical terms this means that far more nuanced interactions can occur between neurons and that the network can run far more efficiently (e.g. in terms of calculations needed and therefore overall power requirements). Nevertheless, the big problem with SNNs is that unlike DL it does not "fit" well with existing technologies. Worst still is that no one has yet come up with definitive way to make SNNs function at a "deep" level. The difficulty is that in essence *"deep"* and *"spiking"* refer to fundamentally different characteristics of a neural network: *"spiking"* focuses on the activation of individual neurons, whereas *"deep"* concerns itself to the network architecture itself [1].

However, these two methods are in fact not contradictory, but have so far been developed in isolation from each other due to the prevailing technology driving each technique and the fundamental conceptual distance between each of the two biological paradigms. If advances in AI are to continue at the present rate that new technologies are going to be developed and the contradictory aspects of DL and SNN are going to have to be reconciled.

Very recently, there have been a handful of attempts to amalgamate DL and SNN in a variety of ways [2]-one of the most exciting being the creation of a specific hierarchical learning paradigm in *Recurrent SNN* (RSNNs) called e-prop [3]. However, this paper posits that this has been made problematic because a fundamental agent in the way the biological brain functions has been missing from each paradigm, and that if this is included in a new model then the union between DL and RSNN can be made in a more *harmonious manner*. The missing piece to the jigsaw, in fact, is the glial cell and the unacknowledged function it plays in neural processing.

In this context, this paper examines how DL and SNN can be combined, and how glial dynamics cannot only address outstanding issues with the existing individual paradigms - for example the "weight transport" problem - but also act as the "glue" – e.g. pun intended - between these two paradigms. This idea has direct parallel with the idea of convolution in DL but has the added dimension of time. It is important not only *where* events happen but also *when* events occur in this new paradigm. The synergy between these two powerful paradigms give hints at the direction and potential of what could be an important part of the next wave of development in AI.

Deep Learning and Recurrent Spiking Neural Networks

An essential distinction between *Deep Learning* (DL) and *Recurrent Spiking Neural Networks* (RSNNs) is in the way the information flows from one neuron to another. DL can be seen as a continuous function approximators in which the information is *synchronously* processed whereas RSNN can be seen as transporting discrete packets of information that can be processed *asynchronously*. DL cares nothing about precisely when something happens, whereas a key characteristic of RSNN is the encoding of temporal information. At a fundamental level it would not be an unreasonable to consider the difference between DL and SNNs as being equivalent to the distinction between analog and digital neural networks.

Recently there has been growing interest in applying some of the techniques used in DL to the field of RSNNs [4]. This is undoubtedly because of DL popularity and its success in solving problems that where intractable even a few years ago. As noted in the abstract the dominance of DL can be attributed to be the right fit for the current technology, but it is also because it can utilize larger scale cognitive structures, whereas RSNN research has very often only focused on local interaction.

Early attempts to marry DL and RSNN involved transposing either the activation functions or weights in a Deep Learning system into a firing rate at an equivalent level in a RSNN. However, because RSNNs can use the timing between spikes as information [5], or the absolute timing of an event [6] they can model far more complex processes than Deep Learning. This allows for far more rapid decision making in the RSNN system as the synchronous calculation required by Deep Learning is not required. Taking this idea further RSNN research opens up the idea of "temporal codes" that may trigger different reactions in different subtle circumstances [7][8]. Using current technologies, this will come at the cost of longer training times, because fully spiking simulations are computationally more expensive than simulations of traditional Neural Networks. However, the number of spikes needed is typically far lower than in rate-coded simulations of SNNs, this opens up the opportunity for both far more efficient and simulation of more complex behavior than is possible with DL.

Despite this supervised training methods in SNN have not fundamentally changed since they were adapted from traditional neural networks. *Spike Time Dependent Plasticity* (STDP) is the most popular variant of the Hebbian rule for learning. STDP which rewards or punishes the firing of related neurons based on the time of the firing. If a neuron fires soon after another neuron fires it is heavily rewarded, if it preempts the firing pattern it is punished. In this way inputs that might be the cause of the post-synaptic neuron's excitation are made even more likely to contribute to the network, whereas inputs that are not the cause of the post-synaptic spike are made less likely to contribute to the network.

Systems such as *ReSuMe* [9] or the *Tempotron* [10] used these in single layer SNNs. Again, multi-layer systems typically use variants of backpropagation inherited from traditional NNs. Most multi-layer or deep SNNs depend on some parameters that can be differentiated and on which backpropagation can be performed. A famous early system, *SpikeProp* [11] derives a backpropagation rule for spike times in the output layer. However, SpikeProp, because of the vast computational expense of backpropagation, has not been applied to large scale problems.

Recently, Lee et al. [12] have used stochastic gradient descent on real-valued membrane potentials. Spike patterns are then subject to filtering before they are finally used for backpropagation. This method achieves state-of-the-art results for deep SNNs on tasks such as MNIST and N-MNIST. Recently Kulkarni and Rajendran [13] used a similar technique and achieved similar results. They minimized the distance between network output and a regular firing target spike train for the desired output neuron.

Wu et al. [14] use a spatio-temporal backpropagation rule for SNN by separating spatial input signals, which come from other neurons, from temporal dynamics arising from the spiking behavior of the neuron itself. Jin et al. [15], use a backpropagation rule for a rate-coded error signal on a longer "macro" time-scale, and updates on a shorter "micro" time-scale intended to capture individual spike effects.

However a potential major problem exists with all of these systems; the error is usually computed at the output layer. Since the information flow in biological axons is assumed to be unidirectional, it is unclear how error information derived from learning methods descended from Hebbian or back-propagation can reach earlier parts of the neural network at all.

Traditionally the answer to this problem is by utilizing bi-directional or random connections in order to provide recurrent feedback connections that modulate learning in earlier layers. This is the idea underlying Echo State Networks (ESN) or Liquid State Machines [16] and Reservoir computing [17]. Random backpropagation has been

postulated as one possible solution for training SNNs with backpropagation, by separating and synchronizing forward and backward passes, and by using symmetrical weights in both directions [18]. Research into energy-based networks and backpropagation [19] has also been performed. Recently Mozafari et al. [20] added multiple layers with reward-modulated Spike Time Dependent Placidity (|STDP) to a SNN to obtain fully spiking supervised training. In order to simultaneously train all layers with STDP within a deep network Thiele et al. [21] introduced neurons with two integrate-and-fire units decoupling learning with STDP and inference. Autoencoders be trained layer-by-layer with unsupervised STDP [22].

However the problem remains in that despite these modifications to the standard SNN model there is no biological justification form many of them. They are a pragmatic solution to a problem constrained by current technology and modus operandi. At its core are two questions:

- How to represent information in SNN, should weight changes constitute learning in such systems as they do in traditional neural networks?
- How should learning propagate in such a system?

The Weight Transport Problem and Credit Assignment

The weight transport problem has been around for almost as long as backpropagation itself [23].
It was recently reported that Geoffrey Hinton, the "father" of DL, said he was "deeply suspicious" of back-propagation, and said "my view is throw it all away and start again".
He outlined a number of problems; for example, in real biological systems:

1. There is no obvious source for a supervision signal
2. Cortical neurons do not communicate real-valued activities needed for back-prop (they send all or none spikes)
3. Neurons do not have symmetrical reciprocal connections. The feedback connections do not go back to the neurons that the feed forward connections came from
4. The neurons need to send two different types of signals (a forward pass: output = activity = y and a backward pass: output = error derivative w.r.t input = dE/dx)

However, Hinton mitigates the argument against backpropagation for each of the above problems by proposing how structures could be built that circumvent these neurological impediments.

Problems 1 and *2* can be easily addressed by using both self-regulatory systems and a different representation scheme for real values, such as used in spiking neural networks.

Problems 3 can be addressed by observing that the brain has many recurrent structures; these structures do not directly go back to the originating neuron but do often go back to the rough region of where the signal originated.

Problem 4 is a bit harder to address. At first this seems in direct contradiction of one interpretation of Dales law; sending two different signals simultaneously, in other words a single neuron representing two different quantities (error derivative and signal) depending on the different state of the system. However, Hinton, Bartunov, Bengio, and others, have recently shown that by modifying the standard interpretation of STDP even this problem can be addressed; spike represent signals and the difference in spiking rates over some interval represent the error derivatives that are backpropagated in the network [24].

However, even if these problems can be addressed as Lillicrap [25] pointed out backpropagation is on the extreme end of a type of learning that may occur in the brain, and that there are undoubtedly others, but it is the only one that is being used in AI research at present. At the other end of the spectrum learning could be done by global weight perturbation, but so far no learning mechanism has been found that is practicable has been found that can be used in even small scale networks (let alone deep ones). Backpropagation can be viewed as an intensely precise instrument that require significant time to repeatedly, synchronously, select neurons and update errors in the network, whereas global weight perturbation can be seen as a blunt instrument used to rapidly effect, asynchronously, global change on the whole network.

Between these two extremes there is evidence of learning algorithms that are yet unexplored. Hinton recently examined capsules or groups of neurons whose outputs represent different properties of the same entity with each capsule operating on a similarity voting scheme and implement [26]. As early as 1995 Hinton attempted to use a

"Wake- Sleep" algorithm, which used recognition weights in the wake phase and used generative weights in the sleep phase in order to explore different learning mechanisms on this continuum [27].

Recently however Bellec et al. [3] have come up with the idea of e-prop(agation) that enables network learning through gradient descent. This exciting development recognizes two key factors hitherto ignored in RSNN research:

1) The dynamics of neurons in the brain is enriched by continuously ongoing updates of traces of past activity on the molecular level, for example in the form of calcium ions.
2) In the brain there exists an abundance of top-down signals such as dopamine and acetylcholine, to name only a few, that inform local populations of neurons about sub-optimal performance of brain computations.

From this they prove that traces and learning signals can be combined to produce network learning through gradient descent – without back-propagation of signals through time or by recurrent synaptic connections.

In this model spikes are treated as binary variables (z_j^t) that is given a value of 1 if neuron j fires at time *t*. The goal of network learning is to find synaptic weights *W* that minimize a given loss function *E* where *E* depends on all or a subset of the spikes in the network. *E* measures in the case of regression or classification learning the deviation of the actual output (y_k^t) of each output neuron *k* at time *t* from its given target value ($y_k^{*,t}$). In other words *E* represents the effectiveness of a particular network of weights. In order to reduce *E* to a minimum then the gradient $\frac{dE}{dW_{ji}}$ for the weight $W_{ji}$of the synapse from neuron *i* to neuron *j* indicates how it should be reduced. This gradient is seen as a sum over the time steps *t* of the RSNN computation.

While *e-propagation* correctly identifies difficulties in the biological plausibility of *back-propagation* and also hints at calcium dynamics as being a possible solution it is conjectured here that the interaction of calcium dynamics add an extra level of interactivity that has the potential to dramatically increase the sophistication of the RSNN.

Representation of weights in SNNs

Most modern SNNs use membrane potential, threshold function or, more even sometimes, spike delay to represent weights. The calcium control hypothesis, says that a large calcium transient produces *Long Term Potentiation* (LTP) whereas a moderate increase in calcium results in *Long Term Depression* (LTD) [28]. This hypothesis has received significant experimental support in neuroscience.

To simulate learning in such a system it is necessary to define rules that translate calcium dynamics in the postsynaptic neuron to changes in synaptic strength.

Shouval [29] showed how this could be done:

$$\frac{dW_i}{dt} = \eta(\mathrm{Ca})(\Omega(\mathrm{Ca}) - \lambda \mathrm{Wi})$$

Equation 1

Here W_i is the synaptic efficacy of synapse *i*, Ω determines the sign magnitude of synaptic plasticity as a function of the Calcium (*Ca*) levels, η is a calcium dependent learning rate, which is usually a monotonically increasing function of calcium, and λ is decay constant which can be generically set to $\lambda=0$, no decay, or $\lambda=1$. If $\lambda=1$, for sustained calcium elevation, the synaptic efficacy converges to the value of Ω and the rate of convergence depends on η.

Learning takes place as neuron pairs become *"coincidence detectors"* by pre- and post-synaptic activity, in turn controlled by the blockage of the respective neuron's calcium receptors.

The Other Brain

However, none of the above considers lager scale structures that may be present in the brain or how these "macro" structures may be represented in a RSNN.

Whereas neurons provide fast point to point electro-chemical signaling that may occur at large distances, glial cells work at a chemical only and a fairly local level. Neurons signals are fast (up to 268 mph), glial calcium signals are slow (Ca^{2+} wavefront velocity typically being 1.05 ± 0.17 μm/s). neurons communicate directly, glial broadcasts; neurons.

Many neuroscientists are suggesting [30] that neurons and glia have a hitherto unsuspected complex interaction that forms the basis of cognition. While controversy currently exists about the extent and the exact mechanisms involved in calcium dynamics [31], there is a consensus that glia/neural interaction constitutes a complex electro chemical feedback mechanism. This is strongly suspected to be the principal agent in *Long Term Depression and Pontentiation* (LTD and LTP) as both require activation of NMDA-type glutamate receptors and the resulting entry of Ca^{2+} into the postsynaptic cell. As such the process of learning is intrinsically linked to this activity. *Yet only a handful of artificial neural models try to reproduce this interaction*.

A Simple Model

We attempt to model calcium dynamics using the Belousov-Zhabotinsky (B-Z) reaction. Calcium levels can influence neural activity in a number of ways, such as the regulation of metabolic activity, the regulation of cell growth, and the long-term modification of synaptic efficiency and is also implicated in the destruction of neurons. It has been shown that a decrease in extra cellular calcium levels brings about a reduction in the threshold for excitation of the nerve, and may even induce spontaneous electrical activity [32]. Conversely, an increase in calcium causes reduced excitability. More significantly, for this work, it has been discovered that calcium transients can have a direct influence on synaptic plasticity and is therefore linked to Hebbian like learning.

In our model we use elevated calcium levels as a way to directly influence learning (in an effort to address the weight transport problem) and as a way to stimulate/depress spike activity. The interaction of these traveling *wavefronts*, similar to interference patterns on the ripple of a pond, are an ideal mechanism to distribute weight changes across the surface of the neural substrate.

The interaction of these calcium wavefronts provide a biologically plausible mechanism to strengthen and weaken connection in the network, in other words *wavefront interaction* is used as a contrivance for the network to learn. Furthermore, the mathematical principles behind the B-Z reaction are well understood and are a good fit for inclusion into a more complex spiking neural network simulation [33]. Not only does this obviate the biological implausibility of backpropagation it has the potential to mimic, and expand upon, one of Deep Learning major breakthrough techniques, principally convolution.

Temporal convolution through the medium of calcium dynamics has the potential associate together groups of neurons activated at specific intervals of time. Different activation patterns would interfere with different groups of neurons. In this way, far more subtle patterns could be encoded onto the network, with each neuron playing a different role at different times. Such cooperation between the neurons in the network increases the potential to create a far more powerful classification network.

Proposed Systems

Diffusion model

With this model the calcium diffuses to the adjacent extra cellular medium. Either a partial differential equation such as used in the heat equation or in Fick's first or second laws of diffusion can be used or a more complex model using continuous time stochastic processes as used in Brownian motion can be used to model particle movement. In its simplest form the standard diffusion equation can be used to calculate the Ca^{2+} flux from one cell to another, for example, the flux J is proportional to the gradient in concentration of Ca^{2+}, namely

$$J = D\frac{dn}{dx}$$

where D is the diffusion coefficient of the material and $\frac{dn}{dx}$ is the concentration of calcium between adjacent cells.

Wave propagation model

The wave propagation model is more biologically relevant than the simple diffusion model presented above. Glial networks generate calcium waves spreading across relatively large areas of the neural substrate. These waves interact to produce complex interference patterns. It is these interference patterns and the Ca^{2+} levels encoded within them that form the basis of this model. The wave model, unlike the diffusion model, recognizes that Ca^{2+} stores are not unlimited and must be used sparingly.

It has very recently been discovered that in real biological systems a rapidly oscillating calcium wave spread quickly but is much more localized than with a calcium wave that is slowly oscillating. This is probably due to the amplitude of the initial calcium wave being strongly influenced by the initial concentration of calcium released. The larger the release, the larger the amplitude of the wave, and the longer it will last. The smaller the release of calcium the more quickly a new release can be made. Rapid oscillations will produce smaller amplitude and less longevity (in other words more suited to rapid localized interactions).

The proposed system uses the interactions between glia, synapse and neuron to provide macroscopic function. The interference patterns of raised levels of *Ca2+* waves modulate transmission properties of the surrounding synapses, which in turn has an effect on the functionality of the neurons they are connected to. This is illustrated in the Figure 1.

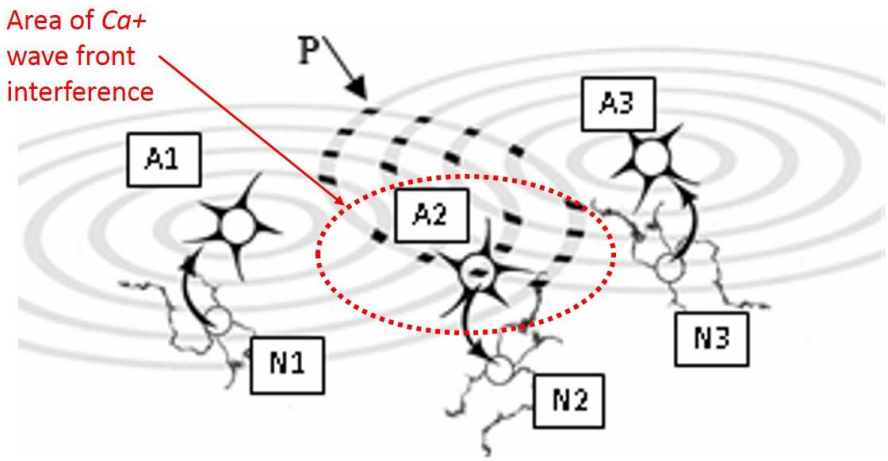

Figure 1 - *Interaction of neuron and glial network*

In Figure 1, at time t_1, the neurons *N1*, *N2*, *N3* are quiescent, as are astrocytes *A1*, *A2*, *A3*. At time t_2 ($t_2 > t_1$), neurons *N1* and *N3* are responding to an input stimulus, in turn stimulating astrocytes *A1* and *A2*. *A1* and *A2* emit calcium waves, which propagate through surrounding medium. At time t_3 ($t_3 > t_2$), points of constructive interference (*P*) will cause "hot spots" of glial activity (e.g. *A2*) stimulating nearby neurons (e.g. *N2*).

In this model raised calcium levels to have *three major effects* on synaptic functionality:

1. It has the potential to vary the *Post-Synaptic Potential* (PSP). For example, the model of the PSP of an excitatory or inhibitory synapse may, after raising calcium levels, have a much steeper initial climb and a much shallower tail. Moreover, the height and area under the curve is analogous to altering size of the synaptic gap, and by implication, the ability for neurotransmitters to cross this gap. If raised calcium levels reduces the size of this gap, or increases the ability of the synapse to send or receive neurotransmitters, then the curve used to model the PSP changes. In other word *synaptic plasticity*, or weight, changes. In this model, we use the equation below to simulate the effect of raised calcium levels. Where Q_{Sk} and V_{Sk} represent synaptic time constraints, the ratio between the two determines the initial steepness of the curve, the overall height of the curve, and the curve's decay:

$$E(S_k) = \frac{S_k}{\boldsymbol{Q_{Sk}}}\left(\mathbf{1} - \boldsymbol{exp}\left(-\frac{\boldsymbol{S_k}}{\boldsymbol{V_{sk}}}\right)\right)$$

Equation 2

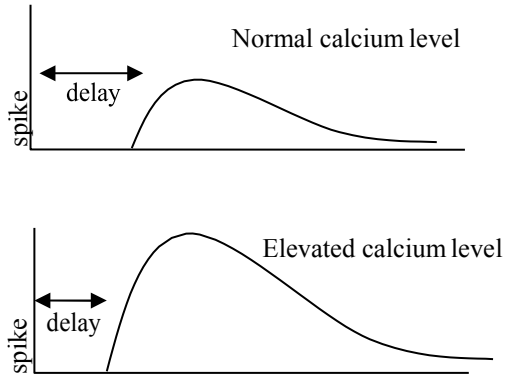

Figure 2 – *Time patterns of the Post-Synaptic Potential (PSP) with normal and elevated calcium level (top and bottom panels, respectively)*

2. New synapses may be formed if sufficient dendritic/axonal activity occurs between neurons. This direction (and motility) of dendritic development may be coordinated by the direction and strength of the calcium waves. Moreover, the interference between waves tells the system about the (i) *direction of the waves*, (ii) *the location of the initial wave* and (iii) *its immediate history*.
   The neuron itself is directly influenced by the synapses it is attached to, and therefore indirectly affected by the calcium wave. Furthermore, the refractory period of the neuron may be decreased if the calcium levels around the neuron are raised. Again, this makes biological sense; if a neuron becomes exhausted then the more stores of calcium it has to replenish itself the shorter its refractory periods will be (Figure 3).

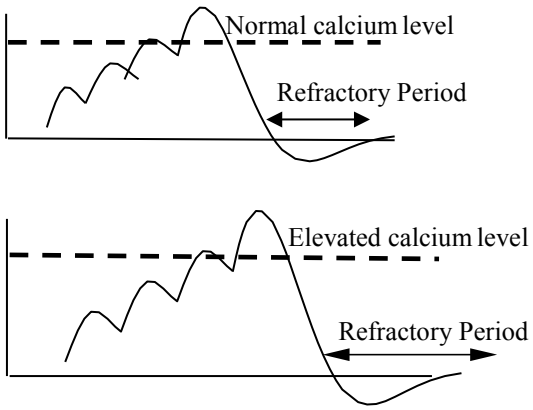

Figure 3 – *Patterns of the refractory period with normal and elevated calcium levels*

3. The glial cell responds to neuronal activity through the elevation of intracellular Ca^{2+} levels. This causes the release of neurotransmitters which in turn adapt the transmission of signals across the neural synapse, hence forming a feedback mechanism (which could be positive or negative depending on the type of neurotransmitter released).

Conclusion

This paper has shown how the glial network has biological relevance to RSNNs and proposes a proof-of-concept system. In particular, the paper has described how growing evidence for the central role of the glial network and calcium in neural stimulation/depression opens up the possibility of a step change in the computational power afforded by RSNN. More importantly it gives biological justification for new categories of learning (such as q-prop) as an alternative to back-propagation type learning.

We believe that such a biological-inspired approach may benefit the development of novel learning model as well as a better comprehension of the mechanism behind biological mechanism and, in particular, the learning process which can affect bio-mimetic and robotic applications.

References

[1] M. Pfeiffer, T. Pfeil, "Deep Learning with Spiking Neurons: Opportunities and Challenges", in *Frontiers in Neuroscience*: Frontiers Media, Volume 12, Article 774, October 2018.

[2] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, A. Maida, "Deep Learning in Spiking Neural Networks", in *Neural Networks*, Elsevier, Volume 111, March 2019, pp. 47-63.

[3] G. Bellec, F. Scherr, A. Subramoney, E. Hajek, D. Salaj, R. Legenstein, W. Maass, "A solution to the learning dilemma for recurrent networks of spiking neurons", in *bioRxiv*, 2019.

[4] V. Demin, D. Nekhaev, "Recurrent Spiking Neural Network Learning Based on a Competitive Maximization of Neuronal Activity", in *Frontiers in Neuroscience*, Frontiers Media, Volume 12, November 2018.

[5] R.Gütig, "To spike, or when to spike?", in *Current Opinion in Neurobiology*, Theoretical and Computational Neuroscience Special Issue, Elsevier, April 2014, pp. 134-139.

[6] M. A .Montemurro, M. J. Rasch, Y. Murayama, N.K. Logothetis, S. Panzeri, "Phase-of-firing coding of natural visual stimuli in primary visual cortex", in *Current Biology*, Elsevier, Volume 18, March 2008, pp. 375-380.

[7] W. Gerstner, R. Kempter, J. L. van Hemmen, H. Wagner, "A neuronal learning rule for sub-millisecond temporal coding", in *Nature*, Springer , Volume 363, September 1996, pp. 76-81.

[8] F. Rieke, D. Warland, R. Steveninck, W. Bialek, "Spikes: exploring the neural code", MIT Press, ISBN: 0-262-18174-6, 1999.

[9] F. Ponulak, A. Kasinski, "Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting", in *Neural Computation*, MIT Press, Volume 22, January 2010.

[10] R.Gütig, H. Sompolinsky, "The tempotron: a neuron that learns spike timing-based decisions", in *Nature Neuroscience,* Springer, Volume 9, March 2006.

[11] S. M. Bohte, J. N. Kok, H. La Poutre, "Spike-prop: error-backpropagation in multi-layer networks of spiking neurons", in Neurocomputing, Elsevier, Volume 48, 2002, pp. 17-37.

[12] J. H. Lee, T. Delbruck. M. Pfeiffer, "Training Deep Spiking Neural Networks Using Backpropagation", in *Frontiers in Neuroscience*, Frontiers Media, Volume 10, November 2016.

[13] S.R Kulkarni, B. Rajendran, "Spiking neural networks for handwritten digit recognition-Supervised learning and network optimization", in *Neural Networks*, Elsevier, Volume 10, July 2018, pp. 118-127.

[14] Y. Wu, L. Deng, G. Li, J. Zhu, L. Shi, "Spatio-Temporal Backpropagation for Training High-Performance Spiking Neural Networks", in *Frontiers in Neuroscience*, Frontiers Media, Volume 12, May 2018.

[15] Y. Jin, W. Zhang, P. Li, "Hybrid Macro/Micro Level Backpropagation for Training Deep Spiking Neural Networks", in *Neural and Evolutionary Computing*, IEEE, 2019.

[16] W. Maass, T. Natschläger, H. Markram, "Real-time computing without stable states: a new framework for neural computation based on perturbations", in *Neural Computation*, MIT Press, Volume 14, November 2002, pp. 2531–60.

[17] B. Schrauwen, D. Verstraeten, J. van Campenhout, "An overview of reservoir computing: theory, applications, and implementations", in *Proceedings of the European Symposium on Artificial Neural Networks ESANN*, 2007, pp. 471-482.

[18] Y. Bengio, D. Lee, J. Bornschein, T, Mesnard, Z. Lin, "Towards Biologically Plausible Deep Learning", arXiv:1502.04156, 2015.

[19] B. Scellier, Y. Bengio, "Equilibrium Propagation: Bridging the Gap between Energy-Based Models and Backpropagation", in *Frontiers in Computational Neuroscience*, Frontiers Media, Volume 11, May 2017.

[20] M. Mozafari, S. R. Kheradpisheh, T. Masquelier, A. Nowzari-Dalini, M. Ganjtabesh, "First-Spike-Based Visual Categorization Using Reward-Modulated STDP", in *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, Volume 12, December 2018.

[21] J. C. Thiele, O. Bichler, A. Dupret, "Event-based, timescale invariant unsupervised online deep learning with STDP", in *Frontiers in Computational Neuroscience*, Frontiers Media, Volume 12, June 2018.

[22] P. Panda, K. Roy, "Unsupervised Regenerative Learning of Hierarchical Features in Spiking Deep Networks for Object Recognition", in *Neural and Evolutionary Computing*, MIT Press, February 2016.

[23] S. Grossberg, “Competitive Learning: From Interactive Activation to Adaptive Resonance”, in *Cognitive Science*, Volume 11, Issue 1, 1987, pp. 23-63.

[24] S. Bartunov, A. Santoro, B. A. Richards, L. Morris, G.E. Hinton T. P. Lillicrap, “Assessing the Scalability of Biologically-Motivated Deep Learning Algorithms and Architectures”, in *32nd Conference on Neural Information Processing Systems (NIPS)*, arXiv:1807.04587v2, Montreal, Canada, 2018.

[25] T. P. Lillicrap, “Deep Learning and the Brain: Does the Brain Approximate Backpropagation?”, in *Carnegie Mellon University's BrainHub Victor Bearg Lecture*, https://www.youtube.com/watch?v=zQSNijL1fJg , 2018.

[26] G. E. Hinton, S. Sabour, N Frosst, “Matrix Capsules with EM Routing”, in *Proceedings of the International Conference on Learning Representations (ICLR)*, Vancouver, Canada, 2018.

[27] G. E. Hinton; P. Dayan; B. J. Frey, R. Neal, “The wake-sleep algorithm for unsupervised neural networks”, in *Science*, Amercan Association for the Advancement of Science, Volume 268, Issue 5214, May 1995, pp.1158–1161.

[28] J. Lisman, “A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory”, in *Proceedings of the National Academy of Science United States of America*, Volume 86, Issue 23, December 1989, pp. 9574-9578.

[29] H. Z. Shouval, G. C. Castellani, B. S. Blais, L. N Cooper, “Converging evidence for a simplified biophysical model of synaptic plasticity”, in *Proceedings of Biological Cybernetics*, Springer, January 2003.

[30] G. Perea, M. Sur, A Araque, “Neuron-glia networks: integral gear of brain function”, in *Frontiers of Cellular Neuroscience*, Frontiers Press, Volume 8, November 2014.

[31] D.S. Auld, R. Robitaille, “Glial Cells and Neurotransmission: An Inclusive View of Synaptic Function”, in *Neuron*, Elsevier, Volume 40, Issue2, October 2003, pp.389-400.

[32] B. Lu, Q. Zhang. H. Wang, Y. Wang, M. Nakayama, D. Ren, “Extracellular Calcium Controls Background Current and Neuronal Excitability via an UNC79-UNC80-NALCN Cation Channel Complex”, in *Neuron*, Elsevier, Volume 68, Issue 3, November 2010, pp.488-499.

[33] G. Chen, “A mathematical model for bifurcations in a Belousov–Zhabotinsky reaction”, in *Physica D: Nonlinear Phenomena*, Elsevier, Volume 145, Issue 3, November 2000, pp. 309-329.