Spiking neural networks for handwritten digit recognition—Supervised learning and network optimization

Article *in* Neural Networks · April 2018

DOI: 10.1016/j.neunet.2018.03.019

CITATIONS
7

READS
970

2 authors, including:

Shruti Kulkarni
New Jersey Institute of Technology
11 PUBLICATIONS **17** CITATIONS

SEE PROFILE

Neural Networks 103 (2018) 118–127

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Spiking neural networks for handwritten digit recognition—Supervised learning and network optimization

Shruti R. Kulkarni, Bipin Rajendran *

Department of Electrical and Computer Engineering, New Jersey Institute of Technology, NJ, 07102, USA

ARTICLE INFO

Article history:
Received 2 December 2017
Received in revised form 13 February 2018
Accepted 27 March 2018
Available online 6 April 2018

Keywords:
Neural networks
Spiking neurons
Supervised learning
Pattern recognition
Approximate computing
Neuromorphic computing

ABSTRACT

We demonstrate supervised learning in Spiking Neural Networks (SNNs) for the problem of handwritten digit recognition using the spike triggered Normalized Approximate Descent (NormAD) algorithm. Our network that employs neurons operating at sparse biological spike rates below 300 Hz achieves a classification accuracy of 98.17% on the MNIST test database with four times fewer parameters compared to the state-of-the-art. We present several insights from extensive numerical experiments regarding optimization of learning parameters and network configuration to improve its accuracy. We also describe a number of strategies to optimize the SNN for implementation in memory and energy constrained hardware, including approximations in computing the neuronal dynamics and reduced precision in storing the synaptic weights. Experiments reveal that even with 3-bit synaptic weights, the classification accuracy of the designed SNN does not degrade beyond 1% as compared to the floating-point baseline. Further, the proposed SNN, which is trained based on the precise spike timing information outperforms an equivalent non-spiking artificial neural network (ANN) trained using back propagation, especially at low bit precision. Thus, our study shows the potential for realizing efficient neuromorphic systems that use spike based information encoding and learning for real-world applications.

1. Introduction

The superior computational efficiency of biological systems has inspired the quest to reverse engineer the brain in order to develop intelligent computing platforms that can learn to execute a wide variety of data analytics and inference tasks (NAE, 2009). Artificial neural networks (ANNs), inspired by the network architecture of the brain, have emerged as the state-of-the-art for various machine learning applications. In particular, inspired by the Nobel prize winning work of Hubel and Weisel on elucidating the mechanisms of information representation in the visual cortex (Hubel & Wiesel, 1968), multi-layer convolutional neural networks have shown impressive performance for a wide variety of applications such as image recognition, natural language processing, speech recognition and video analytics (Ciregan, Meier, & Schmidhuber, 2012; Goldberg, 2016; Goodfellow, Bengio, & Courville, 2016; Goodfellow, Warde-Farley, Mirza, Courville, & Bengio, 2013; Hinton et al., 2012; Hinton, Osindero, & Teh, 2006; Karpathy et al., 2014; Krizhevsky, Sutskever, & Hinton, 2012; Lecun, Bottou, Bengio, & Haffner, 1998).

Nevertheless, the neurons in ANNs implement a memoryless nonlinear transformation of the input synaptic signals to create real-valued output signals. This is vastly different from the behavior of neurons in the brain, which encode information in the timing of binary signals, called action potentials or spikes based on the timing of incoming spike signals from upstream nodes. The third generation of artificial neural networks, also called spiking neural networks (SNNs), have been introduced to mimic this key aspect of information processing in the brain (Maass, 1997). There is growing evidence that SNNs have significant computational advantages as a result of their higher information representational capacity due to the incorporation of the temporal dimension (Brader, Senn, & Fusi, 2007; Crotty & Levy, 2005; Gutig & Sompolinsky, 2006; Hopfield & Brody, 2004). Furthermore, SNNs issue spikes sparsely – the observed spike rate in biological networks is in the range of 0.1 to 300 Hz – and they operate in an event-driven manner (Gabbiani & Metzner, 1999; Roxin, Brunel, Hansel, Mongillo, & van Vreeswijk, 2011; Shoham, O'Connor, & Segev, 2006; Wang et al., 2016). Therefore, highly energy efficient neuromorphic systems can be realized in hardware based on SNNs, as is evidenced by recent demonstrations (Benjamin et al., 2014; Furber, Galluppi, Temple, & Plana, 2014; Gehlhaar, 2014; Merolla et al., 2014; Qiao et al., 2015).

Earlier efforts to build learning algorithms for SNNs were inspired by recent discoveries from neuroscience that shed light on the synaptic (neuronal interconnections) mechanisms of adaptation based on the difference in the times of issue of pre- and

* Corresponding author.
E-mail address: bipin@njit.edu (B. Rajendran).

https://doi.org/10.1016/j.neunet.2018.03.019
0893-6080/

post-synaptic spikes. The most prominent among them is the Remote Supervised Method (ReSuMe) (Ponulak & Kasinski, 2010), that adjusts the synaptic weights based on the precise timing differences of the input and output neurons, inspired by the spike timing dependent plasticity (STDP) rule. Other spike based learning algorithms that have been proposed include the SpikeProp algorithm (though it was restricted to single spike learning) (Bohte, Kok, & La Poutre, 2002), SPAN and PSD, which converted spikes to smoothened analog signals and defined a continuous time cost function for training (Mohemmed, Schliebs, Matsuda, & Kasabov, 2012; Yu, Tang, Tan, & Li, 2013). Another important spike based supervised learning rule was the Chronotron rule which used piece-wise gradient descent and was demonstrated to be efficient in identifying different classes of random spike trains (Florian, 2012). Recently, the reward modulated STDP or R-STDP learning has shown superior performance on several benchmark problems compared to STDP SNNs and even traditional CNNs, even though training was limited to a single layer in the network (Mozafari, Kheradpisheh, Masquelier, Nowzari-Dalini, & Ganjtabesh, 2017). A variant of ReSuMe algorithm, called the Delay Learning (DL)-ReSuMe, in addition to the synaptic weights, made use of the transmission delays of synapses interconnecting the neurons as parameters to train the network (Taherkhani, Belatreche, Li, & Maguire, 2015). This algorithm has been shown to be superior in terms of accuracy and speed of convergence compared to the basic ReSuMe algorithm. The accurate synaptic efficiency adjustment method is another spike-error triggered supervised learning rule based on STDP, which optimizes a cost function defined in terms of membrane potential differences (Xie, Qu, Yi, & Kurths, 2017). This method has been used to demonstrate excellent performance in several UCI datasets with few training parameters. The Synaptic Kernel Inverse Method (SKIM) (Tapson et al., 2013) evaluates the weights analytically rather than learning them iteratively and has been applied to the problem of speech based digit recognition in a small network with 50 neurons. Based on the SKIM method, the convex optimized synaptic efficiencies (CONE) algorithm was developed (Lee, Kukreja, & Thakor, 2017) and was used for the problem of gait detection. The generalization capability of this algorithm and the noise tolerance of a variation of the algorithm called CONE-R has also been demonstrated.

Our work focuses on applying a precise spike based supervised learning algorithm to the MNIST (Modified National Institute of Standards and Technology database) handwritten digit classification problem and optimizing the network in terms of the number of learning parameters for implementation in energy and memory constrained hardware.

In addition to the above mentioned learning methods, unsupervised learning algorithms for SNNs have also been explored, based on the biological spike timing dependent plasticity (STDP) rule (Allred & Roy, 2016; Diehl & Cook, 2015; Kheradpisheh, Ganjtabesh, Thorpe, & Masquelier, 2017; Masquelier & Thorpe, 2007; Panda & Roy, 2016; Roy & Basu, 2017; Tavanaei & Maida, 2017). While these networks use multi-layered convolution architectures with more than one million parameters and have achieved over 98% accuracy on the MNIST dataset (Kheradpisheh et al., 2017; Tavanaei & Maida, 2017), we demonstrate similar accuracy with 13× fewer parameters.

There are also several efforts directed towards developing architectures with adaptive and evolving network structures (Kasabov, 2014; Kasabov et al., 2016; Takuya, Haruhiko, Hiroharu, & Shinji, 2016; Wang, Belatreche, Maguire, & McGinnity, 2015, 2017). SpikeTemp and SpikeComp are algorithms where neurons are progressively added in the classifier layer as the training algorithm approaches the optimal point (Wang et al., 2015, 2017). The recently developed evolving architecture called NeuCube, directly inspired by the brain Kasabov (2014), incorporates weight adjustments based on supervised and unsupervised rules and additionally, adds new network neurons as per training requirements.

Besides the above-mentioned approaches for designing learning algorithms for SNNs that operate directly in the spike domain, several authors have proposed to convert ANNs trained with the well-established backpropagation algorithm to SNNs so that the latter can be used as inference engines (Cao, Chen, & Khosla, 2015; Diehl et al., 2015; Hunsberger & Eliasmith, 2016; Hunsberger, Eric, 2018; Rueckauer, Lungu, Hu, & Pfeiffer, 2016; Rueckauer, Lungu, Hu, Pfeiffer, & Liu, 2017). ANN-to-SNN conversion imposes that the firing rate of a spiking neuron in the SNN be proportional to the activation output of a non-spiking neuron in the ANN. Various techniques such as approximating the response of a spiking neuron with a smooth differentiable ReLU-like function, weight normalization, noise addition, lateral inhibition or spiking rate based pooling masks, which is similar to max pooling operation, have been employed to this end. Using these approaches, state-of-the-art inference accuracies have been demonstrated in spike domain equivalent of deep learning networks such as VGG-16 and Inception-V13 for ImageNet classification problem, and close to 2× reduction in the number of operations needed compared to CNNs for smaller problems such as MNIST and CIFAR-10 (Rueckauer et al., 2017). Recently, a more biologically plausible algorithm called the Feedback Alignment (FA) has been proposed, which unlike the standard backpropagation uses two different sets of weights in the feed-forward and feedback paths (Lillicrap, Cownden, Tweed, & Akerman, 2016). This method has also been demonstrated in SNNs, using approximate differentiable functions of leaky integrate and fire (LIF) spiking neurons to train them in an online manner. However, the FA rule has lower performance compared to the standard backpropagation rule (Hunsberger, Eric, 2018).

Towards the goal of demonstrating a learning SNN capable of high accuracy and efficiency, we use the recently proposed Normalized Approximate Descent (NormAD) algorithm to train the output layer weights of a three-layered network with fixed convolutional kernel weights in the hidden layer. This spike-triggered weight update rule frames the learning task as a supervised optimization problem aimed at tuning the membrane potential to create spikes at desired time instants. Compared to other deterministic learning algorithms in the spike domain such as ReSuMe, at least 10× faster convergence characteristics have been demonstrated using this algorithm for generating arbitrarily desired spike streams (Anwani & Rajendran, 2015).

Prior SNN based demonstration of handwritten digit recognition using spiking versions of backpropagation of errors has achieved 98.7% based on a fully connected 4−layer network and 99.31% with convolutional spiking networks, but also with more than 4× higher number of trainable synapses compared to our network (Lee, Delbruck, & Pfeiffer, 2016). The training algorithm employed in that work has a cost function that is continuous in time defined in terms of the low pass filtered spike trains (both input and output). Compared to the state-of-the-art networks which have shown over 99% accuracy, our SNN trained with NormAD shows an accuracy of 98.17% on the test set of the MNIST database, with 4× fewer synaptic learning parameters (Ciregan et al., 2012; Goodfellow et al., 2013; Lecun et al., 1998; Lee et al., 2016). Furthermore, if the network architecture and number of synaptic parameters are kept the same, we show that the accuracy and performance of the NormAD trained SNN is slightly better than that of an equivalent ANN trained using backpropagation.

This paper is organized as follows. We introduce the basic units of SNNs in Section 2. Section 3 describes the architecture of our network, the spike encoding at the input and output of the network, and the training algorithm used for weight updates. Section 4 describes several hyper-parameter tuning experiments and the

results achieved on the MNIST database. Section 5 discusses the optimization of the network for implementation in energy and memory constrained hardware platforms by approximating the neuronal dynamics and using low-precision bits for storing the synaptic weights. Finally, Section 6 summarizes the key conclusions of our work.

2. Spiking neural networks

SNNs are the third generation of neural networks employing neuron models that are inspired by the biological mechanisms of neuronal signaling. While the mechanism of spike process in biological neurons depends on complex interactions of ion-channels on the cell membrane, a computationally simpler leaky integrate and fire (LIF) model is typically used for simulation of spiking neural networks (Abbott, 1999). This model represents the potential of a neuron as the voltage across a capacitor connected in parallel with a leaky conductance path, and is charged by incoming input currents. Accordingly, the membrane potential $V(t)$ evolves according to the differential equation:

$$C\frac{dV(t)}{dt} = -g_L(V(t) - E_L) + I_{syn}(t). \tag{1}$$

When $V(t) \geq V_T$, a spike is issued and transmitted to the downstream synapses; the membrane potential is reset to its resting value E_L after the spike. We use $E_L = -70$ mV and $V_T = 20$ mV in our simulations. $C = 300$ pF and $g_L = 30$ nS model the membrane's capacitance and leak conductance, respectively. Biological neurons enter a refractory period immediately after a spike is issued during which another spike cannot be issued. We implement this by holding the membrane potential at $V(t) = E_L$ for a short period $t_{ref} = 3$ ms after the issue of a spike. We also limit the membrane potential to the range $[E_L, V_T]$ through clipping.

The spikes arriving at a synapse having a strength (weight) w, will generate a post-synaptic current ($I_{syn}(t)$) in its downstream neuron, given by the following expressions:

$$c(t) = \sum_i \delta(t - t^i) * \left(e^{-t/\tau_1} - e^{-t/\tau_2}\right) \tag{2}$$

$$I_{syn}(t) = w \times c(t) \tag{3}$$

where t^i denotes the time of issue of the ith incoming spike and $*$ is the convolution operator. The variables $\tau_1 = 5$ ms and $\tau_2 = 1.25$ ms model the shape of the synaptic current kernel $c(t)$ and denote its falling and rising time constants, respectively. Note that the time of issue of spikes of a LIF neuron depends on the incoming spike times and synaptic strength in a strong nonlinear fashion, due to the weighted summation, integration and reset.

3. Network architecture

As illustrated in Fig. 1, we designed a simple 3-layer SNN for classification of handwritten digits from the MNIST database. Since MNIST images are 28×28 pixels, our network's input layer has 784 neurons and the output layer has 10 neurons, each corresponding to a particular digit. The input layer neurons connect to 8112 hidden layer neurons through twelve *a priori* fixed 3×3 sized convolutional kernels. The synapses connecting this hidden layer to the output layer are trained using the NormAD algorithm.

3.1. Input encoding

Biological sensory neurons employ complex transformations such as rate coding, time-of-spike coding, population coding and phase coding to encode real-world information in the spike domain (Panzeri, Brunel, Logothetis, and Kayser 2010). Time-encoding machines that convert band-limited input signals to the spike domain such that their perfect reconstruction is possible have been proposed in Lazar, Simonyi, and Tóth (2005). There are also some recent works that use Gaussian receptive fields or Poisson encoding to directly translate real-valued inputs to spike times (Diehl & Cook, 2015; Wang, Belatreche, Maguire, & McGinnity, 2010). As we are dealing with static images, we translate each gray-scale pixel value, in the range [0, 255], to currents that can be applied as inputs to the spiking neurons. Accordingly, each pixel value k is converted into a constant input current for the LIF neuron as

$$i(k) = I_0 + (k \times I_p) \tag{4}$$

where $I_p = 101.2$ pA is a scaling factor and $I_0 = 2700$ pA is the maximum constant amplitude current that does not generate a spike in the LIF neuron in Eq. (1). As a result, a LIF neuron in the input layer issues spikes that are uniformly spaced in time, with a frequency that is sub-linearly proportional to the magnitude of its input current (Khan, Ghani, & Khurram, 2017).

3.2. Convolutional feature extraction

The convolution layer of our network uses *a priori* determined fixed weights for the different feature maps and serves to detect the key features of the image. The filter kernels are continuous curves as shown in Fig. 2(left), and incorporate both excitatory and inhibitory connections. Our kernels are only 3×3 pixels and were inspired by biological studies that suggest that the first few layers of the visual cortex consist of small-sized visual receptive fields (Hubel & Wiesel, 1968).

The filter kernels are spatially convolved with 28×28 spike trains arriving from the input layer neurons, over a simulation period T, with a stride of 1, resulting in feature maps of size 26×26. The weight kernels have an overall net higher inhibition than excitation, as it helped to better suppress the spikes from unwanted edges of the input digit image in the corresponding feature map. Fixed weights based on Gabor filters have been used before as the first layer in a deep convolution neural network, and have shown an improvement in the accuracy for the MNIST dataset compared to the original LeNet-5 network (Calderón, Roa, & Victorino, 2003; Lecun et al., 1998). We use relatively simpler edge detection filters in the hidden layer of our network.

The spikes from the input layer neurons pass through these synaptic weight kernels to generate currents to the hidden layer neurons. The magnitude of the current entering the hidden layer neurons is scaled such that on an average their output spike rate is limited to 10 Hz. Fig. 2(right) shows the 2D feature maps depicting the number of spikes generated by the neurons in the hidden layer when an exemplary image of digit 9 from the MNIST data-set is presented to the network for $T = 100$ ms. The different kernels are able to effectively encode the edges and features of the input image in spike domain.

3.3. Learning layer

The synaptic weights connecting the hidden layer to the output classifier layer are trained using the NormAD algorithm (Anwani & Rajendran, 2015). The weights are initialized to zero at the beginning of training. Weights of all the 8112×10 synapses in this fully-connected layer of the network are updated at the end of presentation of each image, which lasts for the interval T, as

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \Delta\mathbf{w}. \tag{5}$$

The weight update, $\Delta\mathbf{w}$, is calculated only when there is a discrepancy between the spike times in the desired ($S^d(t)$) and

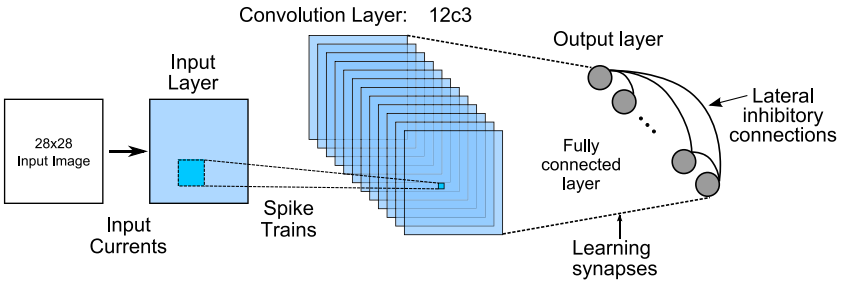

Fig. 1. The proposed spiking neural network architecture for handwritten digit classification. The spike trains from the input layer with 28 × 28 neurons are spatially convolved with twelve filters (or convolution kernels) of size 3 × 3, resulting in the twelve feature maps of size 26 × 26. The synapses connecting the 8112 convolution layer neurons and the 10 output layer neurons are tuned during training. There is a fixed winner-take-all (WTA) lateral inhibition between the neurons in the output layer.

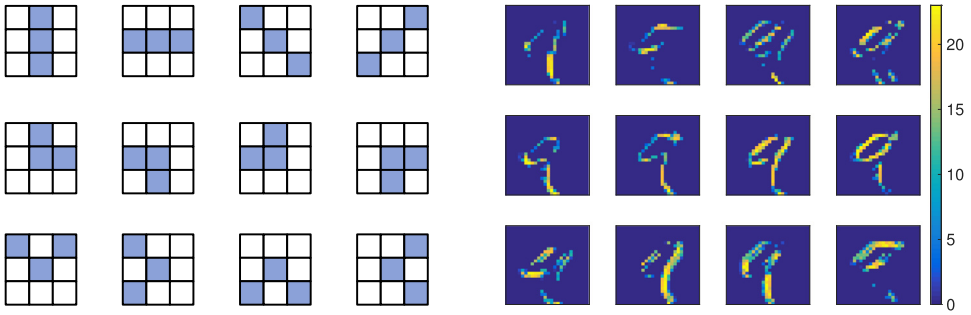

Fig. 2. (left) Convolution filters used in our SNN are of size 3 × 3 pixels. The blue pixels are the excitatory weights, while white pixels are inhibitory values. The magnitude of the excitatory weight is 1.6 times that of the inhibitory weight. (right) The twelve spike count feature maps corresponding to these filters obtained when an exemplary image of digit '9' was presented to the network. The color intensities in the 2D map depict the number of spikes generated by the neurons of the hidden layer when the input was presented for $T = 100$ ms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

observed ($S^o(t)$) spike trains, $e(t) = S^d(t) - S^o(t)$. As described in Anwani and Rajendran (2015), this is achieved by defining a cost function in terms of the error between the desired ($V_{des}(t)$) and observed ($V(\mathbf{w}, t)$) neuron membrane potentials as

$$J(\mathbf{w}) = \frac{1}{2}\int_0^T |e(t)|(V_{des}(t) - V(\mathbf{w}, t))^2 dt. \tag{6}$$

Using gradient descent on the instantaneous cost $J(\mathbf{w}, t)$ obtained by restricting the limits of integral in Eq. (6) to an infinitesimally small interval around time t, the instantaneous weight update term can be written as

$$\Delta\mathbf{w}(t) = \eta(t)\nabla_{\mathbf{w}}J(\mathbf{w}, t) \tag{7}$$

with

$$\nabla_{\mathbf{w}}J(\mathbf{w}, t) = |e(t)|(V_{des}(t) - V(\mathbf{w}, t))\nabla_{\mathbf{w}}V(\mathbf{w}, t) \tag{8}$$

$\eta(t)$ is a time dependent proportionality constant in Eq. (7). By *normalizing* and *approximating* the dependence of membrane potential on the weights, it is possible to obtain a closed form relationship for the weight update as

$$\Delta\mathbf{w} = r\int_0^T e(t)\frac{\hat{\mathbf{d}}(t)}{\|\hat{\mathbf{d}}(t)\|}dt \tag{9}$$

where

$$\hat{\mathbf{d}}(t) = \mathbf{c}(t) * \hat{h}(t), \text{ with } \hat{h}(t) = \exp(-t/\tau_L)u(t). \tag{10}$$

Here, $c(t)$ is the synaptic kernel as described in Eq. (2) and $u(t)$ is the Heaviside step function. The constant $\tau_L = 1$ ms represents the approximation for the neuronal time constant, during training phase. Normalization helps in eliminating the dependency on $V_{des}(t)$, which is an unknown term. The weight update depends only on the output spike error $e(t)$ and the incoming spike trains, captured in $\hat{d}(t)$. The constant r, having the dimensions of synaptic conductance, is a function of the number of input neurons, and is set to 200 pS for our network with 8112 incoming synapses per output neuron.

In our network, the desired signal $S^d(t)$ for the label neuron is a uniform spike train with a frequency of 285 Hz, corresponding to a spike every 3.5 ms, which is slightly higher than the LIF refractory period of 3 ms. There are no spikes in the $S^d(t)$ for all the other neurons.

3.4. Lateral inhibition at the output layer

In addition to the feed-forward inputs from the convolution layer neurons, each output layer neuron also receives lateral inhibitory inputs from the remaining 9 output neurons, implement winner-take-all (WTA) dynamics, similar to Lee et al. (2016). When a neuron spikes, its outgoing WTA synapses inject a negative current to other neurons, thereby suppressing their spikes, as illustrated in Fig. 3.

3.5. Training methodology

During training, each image is presented to the network for a duration T and all the output layer weights are updated after every image, similar to a stochastic gradient descent (SGD) rule. We divide the MNIST training set into two parts: 50,000 for training and remaining 10,000 for validation. In each training epoch, all the 50,000 images are presented once to the network. All the neurons' membrane potentials are initialized to their resting value of $E_L = -70$ mV and the synaptic current variables are cleared at the beginning of each simulation. The dynamics of the SNN is evaluated by numerical integration with a time-step of $\Delta t = 0.1$ ms which is 10 times smaller than the learning time constant, $\tau_L = 1$ ms used in the NormAD algorithm (see Section 3.3). The validation set is used to tune the hyper-parameters of the network such as the variation in the learning rate, optimal number of convolution kernels and the

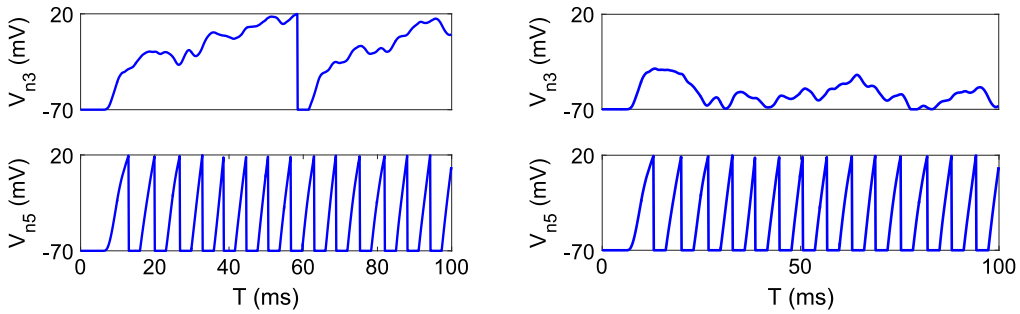

Fig. 3. Membrane potential of two output layer neurons '3' and '5', when an input image of digit '5' was presented to the network. (left) Membrane potential without lateral inhibition and (right) with lateral inhibition. It can be seen that lateral inhibition has suppressed the incorrect neuron '3' from issuing a spike.

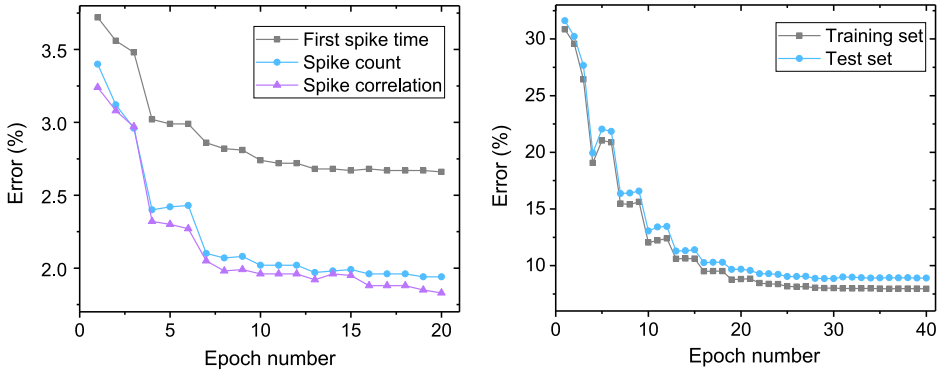

Fig. 4. (left) The 3-layer SNN error on the MNIST test data-set based on the count, correlation and first-spike-time metrics. It can be seen that the network classification error in terms of first neuron to spike (in gray) during the presentation interval T, is worse by almost 1% compared to either count (blue) or the correlation metric (magenta). (right) For a 2-layer SNN without the hidden layer, the error saturates to about 8%, even at 40 epochs of training, illustrating the importance of the hidden layer.

presentation duration as discussed in the following subsections. The network accuracy was determined on the MNIST test set consisting of 10,000 images. The details of the GPU implementation of the algorithm are available in the supplementary material.

4. Results

We now discuss the results of various experiments that we conducted in our study to optimize the performance of our network. We start with the baseline experiments that were conducted to analyze network performance, and then discuss the sensitivity of the network to signal encoding parameters such as image presentation duration, learning rate schedules and the network size.

4.1. Accuracy metrics in spike domain

We primarily used two metrics to measure the accuracy of our network — the first based on the spike count and the second based on the correlation C, of the observed spike trains with respect to a reference spike train. In the count metric, the network's output is decided based on the neuron having the highest spike count. The spike correlation measure (Schreiber, Fellous, Whitmer, Tiesinga, & Sejnowski, 2003) between the output spike train $S_i^o(t)$ for each neuron i in the output layer and a reference spike train $S^r(t)$ is defined as follows:

$$C_i = \frac{\langle L[S_i^o(t)], L[S^r(t)] \rangle}{\|L[S_i^o(t)]\| \, \|L[S^r(t)]\|} \tag{11}$$

where

$$L[S(t)] = S(t) * \exp(-t/\tau)u(t). \tag{12}$$

Here $\langle \mathbf{x}, \mathbf{y} \rangle$ represents the dot product of vectors $\mathbf{x}$ and $\mathbf{y}$. The training signal with a frequency $f_{out} = 285$ Hz is also used as the reference signal during inference. The neuron with the highest value of C is declared the winner of the classification.

The SNN is trained on the MNIST training set for 20 epochs. It can be seen from Fig. 4 (left) that precise timing of spikes measured using the correlation metric gives a slightly higher accuracy for classification, though the spike count metric is a simpler metric to evaluate. The classification accuracy of the network is reported using the correlation metric for the succeeding sections in this paper, with explicit mention of the count metric whenever it is used. We also considered the classification accuracy based on the output neuron that spiked first during the input presentation. However, the accuracy based on this metric at the end of 20 epochs was only about 97.34%. While there is a significant drop in accuracy compared to the correlation and spike count metrics, the prediction can be made within 20 ms of image presentation in 99% of input samples using the first-to-spike metric. This trade-off between latency and accuracy may be especially attractive for low-power approximate computing applications.

We also note the crucial role the convolutional hidden layer plays in improving the network accuracy — in a 2-layer network with the 784 input neurons connected directly to the 10 output layers, the network's error saturates around 8% (Fig. 4(right)).

4.2. Learning rate schedule optimization

As discussed in Anwani and Rajendran (2015), the optimal learning rate for the NormAD algorithm depends on the number of input neurons, N_{inp} and scales according to a $N_{inp}^{-1/2}$ rule. We studied several protocols (learning rate schedules) to decrease the learning rate during training (Table 1), which resulted in lowering the network error by nearly 0.5% (Fig. 5).

Epoch dependent learning rate schedules have shown accuracy improvement in previous works for ANN training (Gokmen &

Table 1
Learning rate schedules.

Scheme	Learning rate (pS)
Schedule 1	$r_0 = 200$, constant over all epochs, n
Schedule 2	$(1/n)$ decrease: $r(n) = \frac{r_0}{(1+k\times n)}$
Schedule 3	Exponential decrease: $r(n) = r_0 \exp(-k \times n)$
Schedule 4	Step decrease by half every 5 epochs
Schedule 5	Step decrease by half every 3 epochs

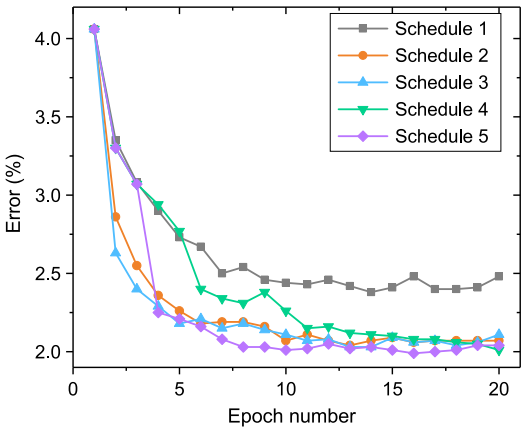

Fig. 5. Network error on the validation set for five different rate schedules listed in Table 1.

Vlasov, 2016; Lecun et al., 1998; Lee et al., 2016); in our study, we experimented with these and several other schedules, shown in Table 1. We use schedule 5 which gave the best validation error after convergence, for the rest of experiments in the paper.

4.3. Network parameter optimization

We also optimized the design parameters of the network such as the number of the convolution kernels used in the hidden layer and the time period T used for presenting each input image to the network. Increasing T results in longer integration time to learn the features of each image, as more spikes (or error points) are produced, resulting in a larger magnitude for the weight update. However, from the perspective of improving the throughput for network performance and preventing over-fitting, smaller values of T are more desirable. Fig. 6 shows the network performance as a function of the number of convolution kernels and the presentation duration T for the images. The network accuracy is optimized with 12 kernels and a presentation duration of $T = 100$ ms. We used a constant inhibitory WTA synaptic strength of 1 nS for all connections in the output layer.

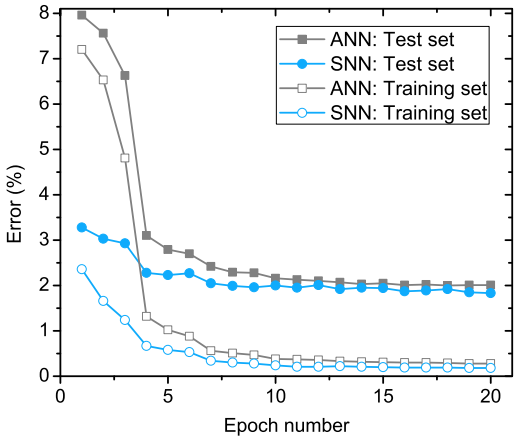

Fig. 7. Comparison of the MNIST error for the 3-layer SNN and an equivalent ANN with the same network structure during 20 epochs of training. The SNN performance (0.18% error for training set and 1.83% error for test set at convergence) is slightly better than that of the ANN (0.28% error for training set and 2.0% for test set at convergence).

4.4. MNIST accuracy results

Having optimized the network hyper-parameters, we trained our SNN with the complete MNIST dataset (60,000 images) for 20 epochs. The SNN achieved an accuracy of 99.82% on the MNIST training set and 98.17% on the test set.

We also trained an equivalent ANN with the same architecture, i.e., the same number of neurons and connectivity patterns (but without the lateral WTA connection) as the SNN in Fig. 1. We used the rectified linear unit (ReLU) as the activation function of the neurons in this network. The weights of the fully-connected layer were adjusted by the standard gradient descent rule by back-propagating the network error. After fine-tuning the learning rate schedule, this ANN achieved an accuracy of 98.0% on the MNIST test set, which is close to the best case accuracy of around 98.50% reported on an equivalently sized three-layered ANN (Tapson, De Chazal, & van Schaik, 2015). The performance for training and test sets for the SNN and ANN networks for 20 epochs of training is shown in Fig. 7. This comparison shows that SNNs trained using the NormAD algorithm can obtain performance similar to equivalent ANNs in benchmark classification problems.

Fig. 8 shows the average of the trained weights of the synapses from the 12 feature maps to each of the 10 output neurons of SNN. When the network is trained on the first 100 images, the weight maps closely resemble the images of the training set digits, though the test set accuracy using these weights was only about 65.8%.

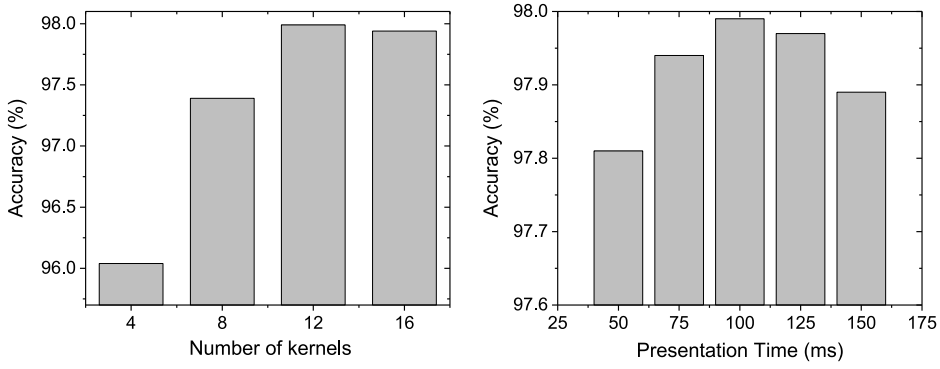

Fig. 6. (left) Classification accuracy on the MNIST test set as a function of the number of convolutional kernels; (right) the presentation duration, T. The network accuracy is optimized with 12 kernels and a presentation duration of $T = 100$ ms.

Table 2
MNIST classification accuracy comparison — our network architecture achieves over 98% accuracy with at least four times fewer parameters than the state-of-the-art networks.

Network and learning algorithm (BP stands for back-propagation)	Number of learning synapses	Test set Accuracy
ANN (LeNet-5) (Lecun et al., 1998)	331,984	99.05%
GCNN (LeNet-5 + Gabor filters) (Calderón et al., 2003)	331,984	99.32%
MCDNN (Multi-column Deep NN) (Ciregan et al., 2012)	1,574,600	99.77%
DNN with DropConnect (Wan, Zeiler, Zhang, Cun, & Fergus, 2013)	2,508,470	99.79%
SNN, with STDP (Diehl & Cook, 2015)	5,017,600	95.0%
Deep SNN with STDP (Kheradpisheh et al., 2017)	5,875,456	98.40%
Fully connected SNN, with BP (Lee et al., 2016)	328,984	98.77%
Convolution SNN with BP (Lee et al., 2016)	581,520	99.31%
Spiking ConvNet (Diehl et al., 2015)	1,422,848	99.11%
SNN, with NormAD (this work)	81,120	98.17%
ANN, with BP (this work)	81,120	98.0%

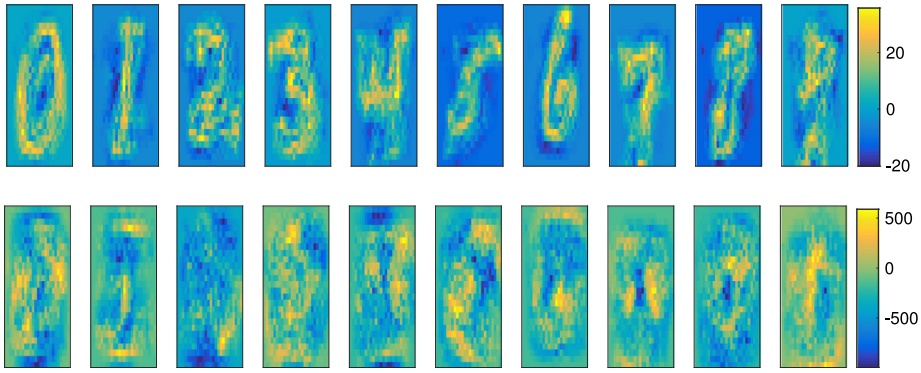

Fig. 8. The average of the trained weights (in pS) from the 12 kernels in the hidden layer to the 10 neurons in the output layer is the effective internal representation of the digits learned by the network. (Top) The average weights in the output layer of the SNN after 100 images presented once for training (when the test set accuracy was only 65.8%) and (Bottom) average weights after training (i.e., with 98.17% accuracy).

When the network is trained with all the 60,000 images in the training set, the test set accuracy rises to 98.17%, thanks to a more complex representation of the images that are captured by the synaptic weights in the network.

To benchmark the classification performance of our network, we compare the accuracy and number of learning synapses in other state-of-the-art approaches for MNIST handwritten digit classification (Table 2). We note that while the accuracy of our approach is about 1.6% worse than the best in class approach, our network achieves this accuracy with four to twenty times fewer number of trainable synaptic weights.

Table 3 presents the confusion matrix for the SNN based classification of the MNIST test data-sets into 10 classes. It can be seen that for all the digits, the true positive rate is 97% and above, demonstrating the high selectivity of the classifier layer, even though this is not easily discernible from the weight maps (Fig. 8). Only five images failed to elicit any spikes in the output neurons.

5. Network optimization

We now discuss the network optimization studies to translate the software design for energy and memory constrained hardware platforms.

5.1. Low precision weight encoding

The ability of a network to maintain its accuracy even when the precision for storing the network parameters is limited, is crucial for efficient hardware implementations. It has been observed that accuracy degrades significantly when low-precision weights are used for network emulation. For instance, a 5% drop in accuracy (with the MNIST data-set) was observed even with 5-bits of fixed-point precision for the synaptic weights in Stromatias et al. (2015).

We test the ability of our SNN and ANN for inference as a function of the precision of trained weights. We train the weights of both these networks in double-precision and then measure the inference accuracy by quantizing these weights, similar to the approach taken in Kulkarni and Rajendran (2015) for designing a scalable hardware solution. The histograms of the weights of our SNN and ANN after training with NormAD and gradient descent, respectively, are observed to be log-normally distributed. Our quantization studies showed that dividing the range of weights into linear bins, rather than log-linear bins gives lesser degradation in performance. Fig. 9 shows the drop in accuracy for our networks as the number of levels for representing the trained weights are reduced. It can be seen that even at 3-bit quantization, the degradation in SNN accuracy is within 1.0% for $T = 100$ ms compared to the floating point baseline. Further, across all quantization values, the degradation in accuracy of the ANN is slightly worse than that of the spiking network. It is also worth pointing out that compared to previous reports such as Stromatias et al. (2015), where the input spike rate was as high as 1500 Hz, the firing rate in our SNN is in the range of 10 to 300 Hz, which is closer to the observed biological spike rates. These results hence demonstrate the robustness of the SNN architecture and its suitability for memory constrained hardware platforms.

5.2. Approximating neuronal dynamics

We also study the SNN's performance when the dynamics of the neurons is evaluated with lower precision. As mentioned in Section 3.5, the time step for numerical integration was chosen to be 0.1 ms for learning. Even though there will be some error in the precise time of spike issue, a larger time step can be used when the network is used for inference.

Table 3
Confusion matrix for the SNN's predicted output shows high selectivity of the NormAD trained classifier layer for each digit.

Actual predicted	0	1	2	3	4	5	6	7	8	9
0	973	0	3	0	2	2	9	1	4	4
1	0	1126	1	0	0	0	2	4	0	4
2	2	3	1015	4	1	1	0	9	1	1
3	0	2	0	996	0	7	1	1	6	4
4	0	1	2	0	964	0	1	1	5	7
5	0	1	0	6	0	876	3	0	1	3
6	2	1	1	0	5	3	940	0	1	0
7	1	1	6	2	0	1	0	1005	3	7
8	1	0	1	1	1	2	1	3	947	3
9	0	0	2	1	9	0	0	3	6	975
No spike	1	0	1	0	0	0	1	1	0	1
Total	980	1135	1032	1010	982	892	958	1028	974	1009

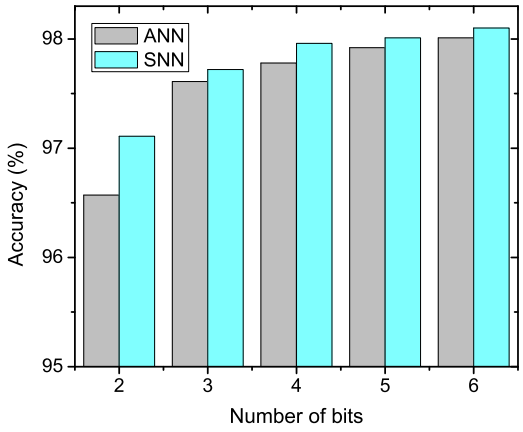

Fig. 9. Test accuracy as a function of the precision of the trained weights in the SNN and ANN. Even at 2-bit precision, the SNN accuracy is only about 1% worse than the floating point baseline. Further, the SNN accuracy is better than the corresponding ANN especially at low bit-precision.

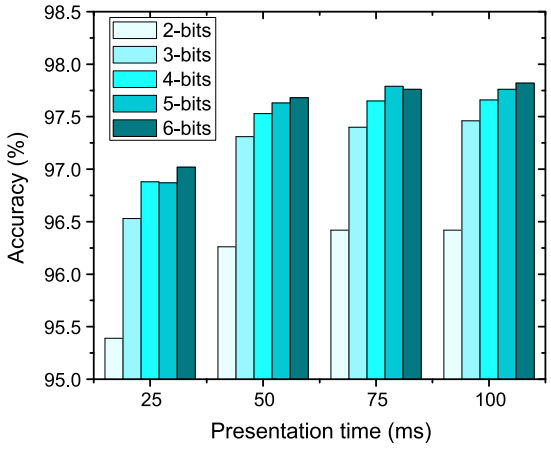

Fig. 10. MNIST test accuracy (count metric) as a function of bit-precision of weights and the presentation time T, when the neuronal dynamics is approximated with a larger integration time step of 1 ms. Even at 3-bits of precision and with $T = 50$ ms, the drop in accuracy is within 1% of the baseline.

With $\Delta t = 1$ ms, the neuronal response can be calculated 10× faster; Fig. 10 shows the test accuracy as a function of bit-precision and presentation times for the 3-layer SNN. Here, we used the count metric to determine the test accuracy to simplify the computation further. At a bit-precision of 3-bits, the digit identification can be completed in just 50 ms or with 50 points of neuronal integration with an accuracy of 97.31%. Hence, close to base-line accuracies can be maintained in approximate network evaluation that permits higher throughput for classification.

6. Conclusion and future work

We presented a highly compact and efficient 3-layer spiking neural network for identifying handwritten digits, that achieved an accuracy of 98.17% on the MNIST dataset using the NormAD learning algorithm. All information in the network is encoded and processed in the spike domain at sparse biological spike rates. Our studies show that using the precise time of spike issue for classification gives slightly better accuracy compared to the simpler rate coding method. We have also presented two techniques to co-optimize the network for hardware implementation, by reducing the bit-precision of weights and approximating the neuronal dynamics with higher integration time-step size.

The best convolution networks in both spiking and non-spiking versions that have achieved over 99% accuracy on the MNIST database use at least over 300,000 adjustable synapses. The NormAD-trained SNN, on the other hand, has 4× fewer learning parameters, making it amenable for implementation on custom neuromorphic hardware with on-chip learning. Our studies also show that as low as 3-bits of weight precision is sufficient to maintain close to baseline accuracies in the SNN when used for inference. Compared to an equivalent ANN with similar network architecture, the spike based training approach also shows better accuracy, especially at lower precision for synaptic weight storage.

The NormAD weight update rule as used in this study can be applied only for tuning the strength of synapses connected to the output layer of a network. However, the methodology used to derive this rule can be extended to adjust the weights of networks with hidden layers in a spike-triggered manner, based on the chain rule of derivatives. Such weight update rules could be then used to pre-train autoencoders which could be stacked and trained to develop deep spiking networks, following the approaches used in deep learning (Hinton et al., 2006). Quantifying the performance of such deep spiking networks and determining their accuracy–efficiency trade-offs for large benchmark classification problems is identified as a topic for future exploration.

Acknowledgments

The authors acknowledge valuable discussions with Prof. Osvaldo Simeone at NJIT and Navin Anwani and Bharath Sastha from IIT Bombay. We also gratefully acknowledge the comments of the anonymous reviewers which helped improve the paper.

This research was supported in part by the CAMPUSENSE project grant from CISCO Systems Inc, the Semiconductor Research Corporation (2016-SD-2717), and the National Science Foundation grant 1710009. Resources of the High-Performance Computing facility at NJIT was used in this work.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neunet.2018.03.019.

References

Abbott, L. F. (1999). Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, *50*, 303–304. http://dx.doi.org/10.1016/S0361-9230(99)00161-6.

Allred, J. M., & Roy, K. (2016). Unsupervised incremental STDP learning using forced firing of dormant or idle neurons. In *2016 international joint conference on neural networks* http://dx.doi.org/10.1109/IJCNN.2016.7727509.

Anwani, N., & Rajendran, B. (2015). NormAD - Normalized Approximate Descent based supervised learning rule for spiking neurons. In *International joint conference on neural networks* (pp. 1–8), http://dx.doi.org/10.1109/IJCNN.2015.7280618.

Benjamin, B. V., Gao, P., McQuinn, E., Choudhary, S., Chandrasekaran, A. R., Bussat, J. M., et al. (2014). Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proceedings of the IEEE*, *102*(5), 699–716. http://dx.doi.org/10.1109/JPROC.2014.2313565.

Bohte, S. M., Kok, J. N., & La Poutre, H. (2002). Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing*, *48*(1), 17–37. http://dx.doi.org/10.1016/S0925-2312(01)00658-0.

Brader, J. M., Senn, W., & Fusi, S. (2007). Learning real-world stimuli in a neural network with spike-driven synaptic dynamics. *Neural Computation*, *19*(11), 2881–2912. http://dx.doi.org/10.1162/neco.2007.19.11.2881.

Calderón, A., Roa, S., & Victorino, J. (2003). Handwritten digit recognition using convolutional neural networks and gabor filters. In *Proc. Int. Congr. Comput. Intell, 2003*.

Cao, Y., Chen, Y., & Khosla, D. (2015). Spiking deep convolutional neural networks for energy-efficient object recognition. *International Journal of Computer Vision*, *113*(1), 54–66. http://dx.doi.org/10.1007/s11263-014-0788-3.

Ciregan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 3642–3649), http://dx.doi.org/10.1109/CVPR.2012.6248110.

Crotty, P., & Levy, W. B. (2005). Energy-efficient interspike interval codes. *Neurocomputing*, *65*, 371–378. http://dx.doi.org/10.1016/j.neucom.2004.10.031. Computational Neuroscience: Trends in Research 2005.

Diehl, P. U., & Cook, M. (2015). Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers in Computational Neuroscience*, *9*, 99. http://dx.doi.org/10.3389/fncom.2015.00099.

Diehl, P. U., Neil, D., Binas, J., Cook, M., Liu, S. C., & Pfeiffer, M. (2015). Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In *2015 international joint conference on neural networks* (pp. 1–8), http://dx.doi.org/10.1109/IJCNN.2015.7280696.

Florian, R. V. (2012). The chronotron: A neuron that learns to fire temporally precise spike patterns. *PloS One*, *7*(8), e40233. http://dx.doi.org/10.1371/journal.pone.0040233.

Furber, S. B., Galluppi, F., Temple, S., & Plana, L. A. (2014). The SpiNNaker Project. *Proceedings of the IEEE*, *102*(5), 652–665. http://dx.doi.org/10.1109/JPROC.2014.2304638.

Gabbiani, F., & Metzner, W. (1999). Encoding and processing of sensory information in neuronal spike trains. *Journal of Fish Biology*, *202*(10), 1267–1279 arXiv:http://jeb.biologists.org/content/202/10/1267.full.pdf.

Gehlhaar, J. (2014). Neuromorphic processing: A new frontier in scaling computer architecture. In *ASPLOS '14. Proceedings of the 19th international conference on architectural support for programming languages and operating systems* (pp. 317–318). New York, NY, USA: ACM. http://dx.doi.org/10.1145/2541940.2564710.

Gokmen, T., & Vlasov, Y. (2016). Acceleration of deep neural network training with resistive cross-point devices: Design considerations. *Frontiers in Neuroscience*, *10*, 333. http://dx.doi.org/10.3389/fnins.2016.00333.

Goldberg, Y. (2016). A Primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research (JAIR)*, *57*, 345–420.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. URL http://www.deeplearningbook.org.

Goodfellow, I., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout networks. In S. Dasgupta, & D. McAllester (Eds.), *Proceedings of machine learning research: vol. 28. Proceedings of the 30th international conference on machine learning* (pp. 1319–1327). Atlanta, Georgia, USA: PMLR, URL http://proceedings.mlr.press/v28/goodfellow13.html.

Gutig, R., & Sompolinsky, H. (2006). The tempotron: a neuron that learns spike timing-based decisions. *Nature Neuroscience*, *9*(3), 420–428. http://dx.doi.org/10.1038/nn1643.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., et al. (2012). Deep Neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97. http://dx.doi.org/10.1109/MSP.2012.2205597.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554.

Hopfield, J. J., & Brody, C. D. (2004). Learning rules and network repair in spike-timing-based computation networks. *Proceedings of the National Academy of Sciences*, *101*(1), 337–342. http://dx.doi.org/10.1073/pnas.2536316100.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243.

Hunsberger, E., & Eliasmith, C. (2016). Training spiking deep networks for neuromorphic hardware. arXiv preprint arXiv:1611.05141.

Hunsberger, Eric (2018). *Spiking deep neural networks: Engineered and biological approaches to object recognition* (Ph.D. thesis), UWSpace, URL http://hdl.handle.net/10012/12819.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *2014 IEEE conference on computer vision and pattern recognition* (pp. 1725–1732), http://dx.doi.org/10.1109/CVPR.2014.223.

Kasabov, N. K. (2014). NeuCube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. *Neural Networks*, *52*, 62–76. http://dx.doi.org/10.1016/j.neunet.2014.01.006.

Kasabov, N., Scott, N. M., Tu, E., Marks, S., Sengupta, N., Capecci, E., et al. (2016). Evolving spatio-temporal data machines based on the NeuCube neuromorphic framework: design methodology and selected applications. *Neural Networks*, *78*, 1–14. http://dx.doi.org/10.1016/j.neunet.2015.09.011.

Khan, S. Q., Ghani, A., & Khurram, M. (2017). Population coding for neuromorphic hardware. *Neurocomputing*, *239*, 153–164.

Kheradpisheh, S. R., Ganjtabesh, M., Thorpe, S. J., & Masquelier, T. (2017). STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks*. http://dx.doi.org/10.1016/j.neunet.2017.12.005.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 25* (pp. 1097–1105). Curran Associates, Inc..

Kulkarni, S. R., & Rajendran, B. (2015). Scalable digital CMOS Architecture for Spike based Supervised Learning. In L. Iliadis, & C. Jayne (Eds.), *Engineering applications of neural networks: 16th international conference, EANN 2015, Rhodes, Greece, September 25-28 2015. Proceedings* (pp. 149–158). Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-23983-5_15.

Lazar, A. A., Simonyi, E. K., & Tóth, L. T. (2005). Time encoding of bandlimited signals, an overview. In *Proceedings of conference on telecommunication systems, modeling and analysis*. Citeseer.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. http://dx.doi.org/10.1109/5.726791.

Lee, J. H., Delbruck, T., & Pfeiffer, M. (2016). Training deep spiking neural networks using backpropagation. *Frontiers in Neuroscience*, *10*, 508. http://dx.doi.org/10.3389/fnins.2016.00508.

Lee, W. W., Kukreja, S. L., & Thakor, N. V. (2017). CONE: Convex-optimized-synaptic efficacies for temporally precise spike mapping. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(4), 849–861. http://dx.doi.org/10.1109/TNNLS.2015.2509479.

Lillicrap, T. P., Cownden, D., Tweed, D. B., & Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature Communications*. http://dx.doi.org/10.1038/ncomms13276.

Maass, W. (1997). Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, *10*(9), 1659–1671. http://dx.doi.org/10.1016/S0893-6080(97)00011-7. = http://www.sciencedirect.com/science/article/pii/S0893608097000117.

Masquelier, T., & Thorpe, S. J. (2007). Unsupervised learning of visual features through spike timing dependent plasticity. *PLOS Computational Biology*, *3*(2), 1–11. http://dx.doi.org/10.1371/journal.pcbi.0030031.

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., et al. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, *345*(6197), 668–673. http://dx.doi.org/10.1126/science.1254642.

Mohemmed, A., Schliebs, S., Matsuda, S., & Kasabov, N. (2012). SPAN: Spike pattern association neuron for learning spatio-temporal spike patterns. *International Journal of Neural Systems*, *22*(04). http://dx.doi.org/10.1142/S0129065712500128. PMID: 22830962.

Mozafari, M., Kheradpisheh, S. R., Masquelier, T., Nowzari-Dalini, A., & Ganjtabesh, M. (2017). First-spike based visual categorization using reward-modulated stdp. arxiv preprint arXiv:1705.09132 [q-bio.NC].

NAE (2009). National academy of engineering - Reverse-engineer the brain. Available at http://bit.ly/1PmsLiX.

Panda, P., & Roy, K. (2016). Unsupervised regenerative learning of hierarchical features in Spiking Deep Networks for object recognition. In *2016 international joint conference on neural networks* (pp. 299–306), http://dx.doi.org/10.1109/IJCNN.2016.7727212.

Panzeri, S., Brunel, N., Logothetis, N. K., & Kayser, C. (2010). Sensory neural codes using multiplexed temporal scales. *Trends in Neurosciences*, *33*(3), 111–120. http://dx.doi.org/10.1016/j.tins.2009.12.001.

Ponulak, F., & Kasinski, A. (2010). Supervised learning in spiking neural networks with ReSuMe: Sequence learning, classification, and spike shifting. *Neural Computation*, *22*(2), 467–510. http://dx.doi.org/10.1162/neco.2009.11-08-901.

Qiao, N., Mostafa, H., Corradi, F., Osswald, M., Stefanini, F., Sumislawska, D., et al. (2015). A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128k synapses. *Frontiers in Neuroscience*, *9*, 141. . http://dx.doi.org/10.3389/fnins.2015.00141.

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., & van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience*, *31*(45), 16217–16226. http://dx.doi.org/10.1523/JNEUROSCI.1677-11.2011. URL http://www.jneurosci.org/content/31/45/16217. arXiv:http://www.jneurosci.org/content/31/45/16217.full.pdf.

Roy, S., & Basu, A. (2017). An online unsupervised structural plasticity algorithm for spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(4), 900–910. http://dx.doi.org/10.1109/TNNLS.2016.2582517.

Rueckauer, B., Lungu, I.-A., Hu, Y., & Pfeiffer, M. (2016). Theory and tools for the conversion of analog to spiking convolutional neural networks. arXiv preprint arXiv:1612.04052.

Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., & Liu, S.-C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Frontiers in Neuroscience*, *11*, 682. http://dx.doi.org/10.3389/fnins.2017.00682. URL https://www.frontiersin.org/article/10.3389/fnins.2017.00682.

Schreiber, S., Fellous, J.-M., Whitmer, D., Tiesinga, P., & Sejnowski, T. J. (2003). A new correlation-based measure of spike timing reliability. *Neurocomputing*, *52*, 925–931. http://dx.doi.org/10.1016/S0925-2312(02)00838-X. Computational Neuroscience: Trends in Research 2003.

Shoham, S., O'Connor, D. H., & Segev, R. (2006). How silent is the brain: is there a "dark matter" problem in neuroscience? *Journal of Comparative Physiology A*, *192*(8), 777–784. http://dx.doi.org/10.1007/s00359-006-0117-6.

Stromatias, E., Neil, D., Pfeiffer, M., Galluppi, F., Furber, S. B., & Liu, S.-C. (2015). Robustness of spiking Deep Belief Networks to noise and reduced bit precision of neuro-inspired hardware platforms. *Frontiers in Neuroscience*, *9*, 222. http://dx.doi.org/10.3389/fnins.2015.00222.

Taherkhani, A., Belatreche, A., Li, Y., & Maguire, L. P. (2015). DL-ReSuMe: A delay learning-based remote supervised method for spiking neurons. *IEEE Transactions on Neural Networks and Learning Systems*, *26*(12), 3137–3149. http://dx.doi.org/10.1109/TNNLS.2015.2404938.

Takuya, T., Haruhiko, T., Hiroharu, K., & Shinji, T. (2016). A training algorithm for spike sequence in spiking neural networks –A discussion on growing network for stable training performance. In *2016 12th international conference on natural computation, fuzzy systems and knowledge discovery* (pp. 1773–1777) http://dx.doi.org/10.1109/FSKD.2016.7603446.

Tapson, J., Cohen, G., Afshar, S., Stiefel, K., Buskila, Y., & Wang, R. et al. (2013). Synthesis of neural networks for spatio-temporal spike pattern recognition and processing. arXiv preprint arXiv:1304.7118.

Tapson, J., De Chazal, P., & van Schaik, A. (2015). Explicit computation of input weights in extreme learning machines. In J. Cao, K. Mao, E. Cambria, Z. Man, & K.-A. Toh (Eds.), *Proceedings of ELM-2014 Volume 1:: Algorithms and theories* (pp. 41–49). Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-14063-6_4.

Tavanaei, A., & Maida, A. S. (2017). Multi-layer unsupervised learning in a spiking convolutional neural network. In *2017 International joint conference on neural networks* (pp. 2023–2030), http://dx.doi.org/10.1109/IJCNN.2017.7966099.

Wan, L., Zeiler, M., Zhang, S., Cun, Y. L., & Fergus, R. (2013). Regularization of Neural Networks using DropConnect. In *Proceedings of machine learning research*: *vol. 28. Proceedings of the 30th international conference on machine learning* (pp. 1058–1066). Atlanta, Georgia, USA: PMLR, URL http://proceedings.mlr.press/v28/wan13.html.

Wang, J., Belatreche, A., Maguire, L., & McGinnity, M. (2010). Online versus offline learning for spiking neural networks: A review and new strategies. In *2010 IEEE 9th international conference on cyberntic intelligent systems* (pp. 1–6), http://dx.doi.org/10.1109/UKRICIS.2010.5898113.

Wang, J., Belatreche, A., Maguire, L. P., & McGinnity, T. M. (2015). SpikeComp: An evolving spiking neural network with adaptive compact structure for pattern classification. In S. Arik, T. Huang, W. K. Lai, & Q. Liu (Eds.), *Neural information processing: 22nd international conference, ICONIP 2015, Istanbul, Turkey, November 9-12, 2015, Proceedings, Part II* (pp. 259–267). Cham: Springer International Publishing. http://dx.doi.org/10.1007/978-3-319-26535-3_30.

Wang, J., Belatreche, A., Maguire, L. P., & McGinnity, T. M. (2017). Spiketemp: An enhanced rank-order-based learning approach for spiking neural networks with adaptive structure. *IEEE Transactions on Neural Networks and Learning Systems*, *28*, 30–43. http://dx.doi.org/10.1109/TNNLS.2015.2501322.

Wang, B., Ke, W., Guang, J., Chen, G., Yin, L., Deng, S., et al. (2016). Firing frequency maxima of fast-spiking neurons in human, monkey, and mouse neocortex. *Frontiers in Cellular Neuroscience*, *10*, 239. http://dx.doi.org/10.3389/fncel.2016.00239. 27803650[pmid].

Xie, X., Qu, H., Yi, Z., & Kurths, J. (2017). Efficient training of supervised spiking neural network via accurate synaptic-efficiency adjustment method. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(6), 1411–1424. http://dx.doi.org/10.1109/TNNLS.2016.2541339.

Yu, Q., Tang, H., Tan, K. C., & Li, H. (2013). Precise-spike-driven synaptic plasticity: learning hetero-association of spatiotemporal spike patterns. *Plos One*, *8*(11), 1–16. http://dx.doi.org/10.1371/journal.pone.0078318.