



OPEN ACCESS

EDITED BY

Malu Zhang,
National University of Singapore, Singapore

REVIEWED BY

Aili Wang,
International Campus, Zhejiang University,
China
Yufei Guo,
China Aerospace Science and Industry
Corporation, China

*CORRESPONDENCE

Samuel Schmidgall
✉ samuel.schmidgall@nrl.navy.mil

RECEIVED 09 March 2023

ACCEPTED 06 April 2023

PUBLISHED 12 May 2023

CITATION

Schmidgall S and Hays J (2023)
Meta-SpikePropamine: learning to learn with
synaptic plasticity in spiking neural networks.
Front. Neurosci. 17:1183321.
doi: 10.3389/fnins.2023.1183321

COPYRIGHT

© 2023 Schmidgall and Hays. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Meta-SpikePropamine: learning to learn with synaptic plasticity in spiking neural networks

Samuel Schmidgall^{1,2*} and Joe Hays¹

¹U.S. Naval Research Laboratory, Spacecraft Engineering Department, Washington, DC, United States,

²Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, United States

We propose that in order to harness our understanding of neuroscience toward machine learning, we must first have powerful tools for training brain-like models of learning. Although substantial progress has been made toward understanding the dynamics of learning in the brain, *neuroscience-derived* models of learning have yet to demonstrate the same performance capabilities as methods in deep learning such as gradient descent. Inspired by the successes of machine learning using gradient descent, we introduce a bi-level optimization framework that seeks to both solve online learning tasks and improve the ability to learn online using models of plasticity from neuroscience. We demonstrate that models of three-factor learning with synaptic plasticity taken from the neuroscience literature can be trained in Spiking Neural Networks (SNNs) with gradient descent via a framework of learning-to-learn to address challenging *online* learning problems. This framework opens a new path toward developing neuroscience inspired online learning algorithms.

KEYWORDS

online learning, synaptic plasticity, spiking neural networks, biologically inspired, meta-learning, learning to learn

Introduction

The ability to learn continually across vast time spans is a hallmark of the brain which is unrivaled by modern machine learning algorithms. Extensive research on learning in the brain has provided detailed models of synaptic plasticity—however, these models of learning have yet to produce the impressive capabilities demonstrated by deep neural networks trained with backpropagation. Despite our increasing understanding of biological learning, the most powerful methods for optimizing neural networks have remained backpropagation-based. However, backpropagation is not perfect. For example, when backpropagation is applied to a continuous stream of data, memory-stability issues arise since gradient descent approaches do not address the ability to update synapses continually without forgetting previously learned information (French, 1999). This is because backpropagation methods modify the weight of every synapse at every update, which causes task-specific information from previous updates to rapidly deteriorate. The tendency for backpropagation to overwrite previously learned tasks has made its use as an online learning algorithm impractical (Kirkpatrick et al., 2017; Parisi et al., 2019). The brain solves this problem by determining its own modification as a function of information that is locally available to neurons and synapses. This ability for self-modification is a process that has been fine-tuned through a long course of evolution, and is the basis of learning and memory in the brain (Hrvoj-Mihic et al., 2013). Can the success of gradient descent be combined with neuroscience models of learning in the brain?

Recent experimental evidence from neuroscience has provided valuable insight into the dynamics of learning in the brain (Frémaux and Gerstner, 2016; Gerstner et al., 2018). Two fundamental findings have led to recent successes in the development of online neuro-inspired learning algorithms (Lukasz Kusmierz et al., 2017; Kaiser et al., 2019; Bellec et al., 2020; Liu et al., 2021; Schmidgall et al., 2021). First, neurons and synapses in the brain maintain historical traces of activity. These traces, referred to as eligibility traces, are thought to accumulate the joint interaction between pre- and post-synaptic neuron factors. Eligibility traces do not automatically produce a synaptic change, but have been demonstrated to induce synaptic plasticity in the presence of top-down learning signal. Second, the brain has a significant quantity of top-down learning signals which are broadly projected by neurons from higher centers in the brain to plastic synapses to convey information such as novelty, reward, and surprise. These top-down signals often represent neuromodulator activity such as dopamine (Schultz et al., 1993; Seamans, 2007; Zhang et al., 2009; Steinberg et al., 2013; Speranza et al., 2021) or acetylcholine (Ranganath and Rainer, 2003; Hasselmo, 2006; Teles-Grilo Ruivo and Mellor, 2013; Brzosko et al., 2015; Zannone et al., 2018). The interaction between eligibility traces and top-down learning signals enables learning rules to connect interactions between long and short time scales (Frémaux and Gerstner, 2016; Gerstner et al., 2018).

Here, we demonstrate on two challenging online learning problems that models of neuromodulated synaptic plasticity from neuroscience can be trained in SNNs with the paradigm of learning to learn through gradient descent. These results demonstrate that neuromodulated synaptic plasticity rules can be optimized to solve temporal learning problems from a continuous stream of data, leading to dynamics that are optimized to address several fundamental *online* learning challenges. This new paradigm allows models of neuromodulated synaptic plasticity to realize the benefits from the success of gradient descent in machine learning while staying true to neuroscience. This opens the door for validating learning theories in neuroscience on challenging problems, as well as developing effective online learning algorithms which are compatible with existing neuromorphic hardware.

Learning in networks with plastic synapses

Learning how to learn online

The primary strategy for developing online learning systems has been to attempt discovering each piece of the system manually such that these pieces can one day be assembled to form an effective online learning system. Alternatively, the paradigm of meta-learning aims to learn the learning algorithm itself such that it ultimately discovers a solution that solves the inherent learning problems out of necessity (Clune, 2019). Meta-learning has been notoriously difficult to define, and is often used inconsistently across experiments—however, it is consistently understood to signify *learning how to learn*: improving the learning algorithm itself (Hospedales et al., 2020). More concisely, meta-learning is a learning paradigm that uses meta-knowledge from previous experience to improve its ability to learn in new contexts. This

differs from *multi-task* learning in that, multi-task learning aims to produce a model that performs well on multiple tasks that are explicitly encountered during the optimization period, whereas meta-learning primarily aims to produce a model that is able to learn novel tasks more efficiently.

Meta-learning consists of an inner (base) and outer (meta) loop learning paradigm (Hospedales et al., 2020). During base learning, an inner-loop learning algorithm solves a task, such as robotic locomotion or image classification, while optimizing a provided objective. During meta-learning, an outer-loop (*meta*) algorithm uses information collected from the base learning phase to improve the inner-loop (base learning) algorithm toward optimizing the outer-loop objective. It is proposed that there are three axes within the meta-learning paradigm: meta-representation (what?), meta-optimization (how?), and meta-objective (why?) (Hospedales et al., 2020). The meta-representation refers to the representation of meta-knowledge ω . This knowledge could be anything from initial model parameters (Finn et al., 2017; Rothfuss et al., 2018; Fakoor et al., 2019; Liu et al., 2019), the inner optimization process (Andrychowicz et al., 2016; Bello et al., 2017; Metz et al., 2018; Irie et al., 2022), or the model architecture (Zoph and Le, 2016; Liu et al., 2018; Lian et al., 2019; Real et al., 2019). The meta-optimizer refers to the choice of optimization for the outer-level in the meta-training phase which updates meta-knowledge ω . The meta-optimizer often takes the form of gradient-descent (Finn et al., 2017), evolutionary strategies (Houthoofd et al., 2018), or genetic algorithms (Co-Reyes et al., 2021). The meta-objective specifies the goal of the outer-loop learning process, which is characterized by an objective \mathcal{L}^{meta} and task distribution $\mathcal{D}_{source}^{test(i)}$.

To provide a more formal definition, the bilevel optimization perspective of meta-learning is presented as follows:

$$\omega^* = \arg \min_{\omega} \sum_{i=1}^M \mathcal{L}^{meta}(\theta^{*(i)}(\omega), \omega, \mathcal{D}_{source}^{test(i)}) \quad (1)$$

$$\text{s.t. } \theta^{*(i)}(\omega) = \arg \min_{\theta} \mathcal{L}^{task}(\theta, \omega, \mathcal{D}_{source}^{train(i)}) \quad (2)$$

Equation 1 represents the outer-loop optimization, which looks to find an optimal meta-representation ω^* defined by the selection of values ω such that the meta-objective loss \mathcal{L}^{meta} is minimized across a set of M tasks from the task testing distribution $\mathcal{D}_{source}^{test(i)}$. The minimization of \mathcal{L}^{meta} is dependent on finding $\theta^{*(i)}(\omega)$, which is the selection of values for θ that minimize the task loss \mathcal{L}^{task} using the meta-representation ω . In other words, $\theta^{*(i)}(\omega)$ looks to find the optimal θ for a given *training* distribution of data using ω and ω^* looks to find the optimal ω for a given *testing* distribution with $\theta^{*(i)}(\omega)$ that was optimized on the training distribution using a given ω .

Learning how to learn online with synaptic plasticity through gradient descent

In learning applications with networks of spiking neurons, synaptic plasticity rules have historically been optimized through black-box optimization techniques such as evolutionary strategies

(Bohnstingl et al., 2019; Schmidgall, 2020; Schmidgall and Hays, 2021), genetic algorithms (Elbrecht and Schuman, 2020; Jordan et al., 2021), or Bayesian optimization (Nessler et al., 2008; Kulkarni et al., 2021). This is because spiking dynamics are inherently non-differentiable, and non-differentiable computations prevent gradient descent from being harnessed for optimization. However, recent advances have developed methods for backpropagating through the non-differentiable part of the neuron with surrogate gradients (Shrestha and Orchard, 2018; Neftci et al., 2019; Li et al., 2021; Zenke and Vogels, 2021; Guo et al., 2022), which are continuous relaxations of the true gradient. These advances have also allowed gradient descent based approaches to be utilized for optimizing both the parameters defining plasticity rules and neuromodulatory learning rules in SNNs (Schmidgall et al., 2021). However, previous work optimizing these rules use neuromodulated plasticity as a dynamic which *compliments* the network on tasks which can be solved without it instead of using it as the learning algorithm itself (Schmidgall et al., 2021). State of the art methods which *do* use neuromodulated plasticity as a learning algorithm do not learn its dynamics from biological learning rules, but define rules which are derived from machine learning approaches (Bellec et al., 2020; Scherr et al., 2020). Instead, we desire to provide a paradigm of using learning rules from neuroscience that can be optimized to act as the learning algorithm through gradient descent.

An insight which enables this is the idea that synaptic plasticity in the presence of a neuromodulatory signal can be thought of as a meta-learning optimization process, with meta-knowledge ω being represented by the learned plasticity rule parameters and θ as the strength of synaptic weights representing the inner-level free parameters which change based on ω *online*. Since both the parameters governing the dynamics of the neuromodulatory signal and the plasticity rules in SNNs can be optimized through backpropagation through time (BPTT) (Schmidgall et al., 2021), the outer-loop training can be framed to optimize neuromodulated plasticity rules (Equation 1) which act as the inner-loop learning process (Equation 2). The optimization goal of outer-loop in Equation 1 is a selection of the neuromodulatory and plasticity parameters for ω which minimize the outer-loop loss \mathcal{L}^{meta} as a function of $\theta^{*(i)}(\omega)$, ω , and $\mathcal{D}_{source}^{test(i)}$. The optimization goal of the inner loop in Equation 2 is to find $\theta^{*(i)}(\omega)$ which is defined as a selection of the parameters for θ which minimize the inner-loop loss \mathcal{L}^{task} , such that θ is determined across time as a function of the plasticity equation and ω , which parameterizes the plasticity rules and the neuromodulatory dynamics, for a given task $\mathcal{D}_{source}^{train(i)}$. By optimizing the learning process, gradient descent, which acts on ω in Equation 1, is able to shape the dynamics of learning in Equation 2 such that it is able to solve problems that gradient descent is not able to solve on its own. To do this, learning problems are presented to emulate how biological organisms are trained to solve tasks in behavioral experiments—specifically with respect to the *online* nature of the task. The meta-learning process can then shape plasticity and neuromodulatory dynamics to address more fundamental challenges that are presented during the inner-loop task. Rather than manual design, these fundamental learning problems are addressed implicitly by the optimization process out of necessity for solving the meta-learning objective. As this

work will demonstrate, using neuromodulated plasticity as the meta-representation allows for the learning algorithm itself to be learned, making this optimization paradigm capable of *learning* to solving difficult temporal learning problems. This capability is demonstrated on an online one-shot continual learning problem and on a online one-shot image class recognition problem.

Experiments

One-shot continual learning: addressing credit assignment through one-shot cue association

Experience-dependent changes at the synapse serve as a fundamental mechanism for both short- and long-term memory in the brain. These changes must be capable of attributing the outcome of behaviors together with the necessary information contained in temporally-dependent sensory stimuli, all while ignoring irrelevant details; if the behavior produced by a particular stimuli led to a good outcome it should be reinforced and visa versa. The problem, however, is that the outcome of behavior is often not realized for a long and typically variable amount of time after the actions affecting that outcome are produced. Additionally, there are often many elements of sensory noise that could serve to distract the temporal-learner from proper credit assignment.

To examine these capabilities, a valuable learning experiment from neuroscience tests the cognitive capabilities of rodents in a T-maze. The T-maze can be described as an enclosed structure that takes the form of a horizontally-placed T (Lett, 1975; Wenk, 1998; Dudchenko, 2001; Deacon and Rawlins, 2006; Engelhard et al., 2019), with the maze beginning at the base of T and ending at either side of the arms, see Figure 1. The rodent moves down the base of the maze and chooses either side of the arms. In some experiments, a series of sensory cues are arranged along the left and right of the apparatus as the rodent makes progress toward the end of the maze. A decision as to which side of the maze will provide positive and negative reinforcement is based on the arrangement of these stimuli (Morcos and Harvey, 2016; Engelhard et al., 2019). The rodent is rewarded for choosing the side of the track with the majority of cues. This task is not trivial to solve since the rodent has to recognize that the outcome is not effected by the presentation ordering of the cues or which side the last cue was on. Rather, the cues must be counted independent of their ordering for each side and the sums must be compared to make a decision. Making learning more difficult, the reward for solving this problem is not presented until after a decision has been made, so the rodent must address credit assignment for its behavior across the time span of an entire cue experiment.

Previous experiments with SNNs in simulation have demonstrated that synaptic plasticity alone enables a network to solve this problem where it was not able to be solved with feedforward SNNs (Schmidgall et al., 2021) or Recurrent SNNs (RSNNs) (Bellec et al., 2020) using BPTT. However, previous work only considered learning in this environment in a setting where the neurons associated with a particular cue remained consistent across gradient updates and experiments. In this way,

there was no inner- and outer-level optimization. Rather, the synaptic plasticity served as a mechanism for memorization and cue-decision making, but not actually learning which cues and which decisions are associated with positive reward during the network time-horizon, and hence it does not qualify as a meta-learning problem. Additionally, during *in-vivo* rodent experiments, accurate cue-problem performance is demonstrated with only 7-12 sessions per mouse (Morcos and Harvey, 2016). This differs from the learning efficiency of Bellec et al. (2020) and Schmidgall et al. (2021), which take on the order of hundreds and thousands of training sessions respectively.

Many- and one-shot learning

Converting this experiment from neuroscience into simulation, sensory cues are emulated as probabilistic spike trains, with subgroups of neurons corresponding to particular sensory cues. Twenty sensory neurons are organized into four subgroups, five of which represent right-sided cues, five for left-sided cues, five of which display activity during the decision period, and five which purely produce spike noise (Figure 2B). To transform this problem into a meta-learning problem, the particular sensory neurons which are associated with cues, decision timings, and noise are randomly permuted (Figure 2C) making the temporal learner unable to know which neurons are associated with which stimuli at the beginning of each cue-association task. The network is then presented with a series of cue-trials (Figure 2A) and a reward signal at the end of each trial. The many-shot cue association experiment is as follows: (1) the neurons associated with particular cues in previous experiments are randomly permuted, (2) the network is placed at the beginning of the cue-maze, (3) a series of sensory inputs, noise, cues, and decision activity are input into the sensory neurons as the learner moves along the apparatus, (4) at the end of the maze the learner makes a decision (left or right) based on the sensory input and a reward signal is provided as input to the neuromodulatory network based on whether it was the correct decision, (5) the agent is placed at the beginning of the maze and starts from step 2 *without* resetting network parameters and traces for N trials for all K cues (left and right) acting as a *training* phase (i.e. inner loop solving Equation 2), (6) the performance of the network is *tested* based on the information that has been learned from the N -shot cue presentations, (7) plasticity parameters are updated through gradient descent based on evaluation performance for the final test trial (i.e. outer loop solving Equation 1), and the learning problem is repeated from step 1. The benefit of permuting the sensory neurons as a source of inner-loop learning is that it results in a large number of variations of the problem. With only 20 neurons there are $20! = 2.4 \cdot 10^{18}$ variations, which results in learning experiments which are unlikely to have repetitions in the problem domain.

One-shot learning is a particularly challenging variation of the N -shot learning paradigm, where N is set equal to one for each of the K classes. In this way, the learning model is only provided with *one* example of each class and must be capable of differentiating between classes based only on the given single example. One-shot learning is argued to be one of two important capabilities of the brain that is missing from models of learning in computational neuroscience (Brea and Gerstner, 2016).

Architecture

The DP-SNN in Figure 3 contains one-hidden layer of 48 Current-Based Leaky Integrate-and-Fire (CUBA) neurons (Methods, Equations 6, 7). Synaptic connections between the input neurons and the hidden layer neurons accumulate changes in an eligibility trace based on an additive pair-based STDP rule (Methods, Equation 12). The STDP paradigm typically describes two forms of change: Long-Term Potentiation (LTP) and Long-Term Depression (LTD). LTP describes an increase in synaptic strength while LTD describes a decrease in strength. For the pair-based STDP rule, an eligibility is maintained for the LTP dynamics of the pair-based rule and trace for the LTD dynamics. Pair-based STDP (Methods, Equation 11) represents plasticity based on the product of timing relationships between pairs of pre- and post-synaptic activity. Learning signals for the LTP trace and the LTD trace are produced by an independent neuromodulatory SNN (NM-SNN) using an input neuron specific modulatory signal (Methods, Equation 19) for both the LTP and LTD dynamics. Connections from the hidden neurons to the output activity are non-plastic synapses learned through gradient descent. During network initialization, only a fraction of neurons are connected, with a connection probability of 50%. Each initialized synapse is assigned to represent either an inhibitory synapse with 20% probability or an excitatory synapse with 80% probability, with inhibitory synapses producing negative currents in outgoing neurons and excitatory synapses producing positive ones. The neuromodulatory SNN contains two layers of 64 CUBA neurons (Methods, Equations 6, 7). The synapses are *non-plastic* and are fully-connected between layers. The NM-SNN receives the same sensory input as the DP-SNN in addition to the DP-SNN hidden neuron activity and a learning signal that occurs at the decision interval for the training cue sequences. Both the DP-SNN and the neuromodulatory SNN share the same meta-objective and are optimized jointly in an end-to-end manner with *BPTT* in the outer loop (i.e. Equation 1). Error for the one-shot learning task is calculated via binary cross entropy loss on the output neuron activity compared with the correct cue label (Figure 2H) during the testing data trajectory, with $\mathcal{L}^{meta} = -\sum_{i=1}^2 (\log(p(y_i)) + (1 - y_i)\log(1 - p(y_i)))$ and y_i equal to the weighted output neuron activity.

Experimental setup

The one-shot learning experiment in this work presents $M = 5$ cues (Figures 2A–C). During a cue presentation period the permuted cue neuron has a firing probability of 0.75. When the cue neuron is not active (during a cue presentation) the firing probability is 0.15. The noise neurons have a firing probability of 0.15 at each moment in time and the decision interval neurons have a firing probability of 0.75 during a decision period and 0.15 otherwise. The cue presentation period for each cue spans 25 ms which is followed by a 30 ms resting period between each cue. After the final cue there is a 50 ms resting period before the decision period which is 25 ms totalling 350 ms for each individual cue problem in the one-shot cue-association task. The simulation step size is set to 1 ms. An environment feedback signal arrives at the end of each cue-trial during the decision interval, requiring the

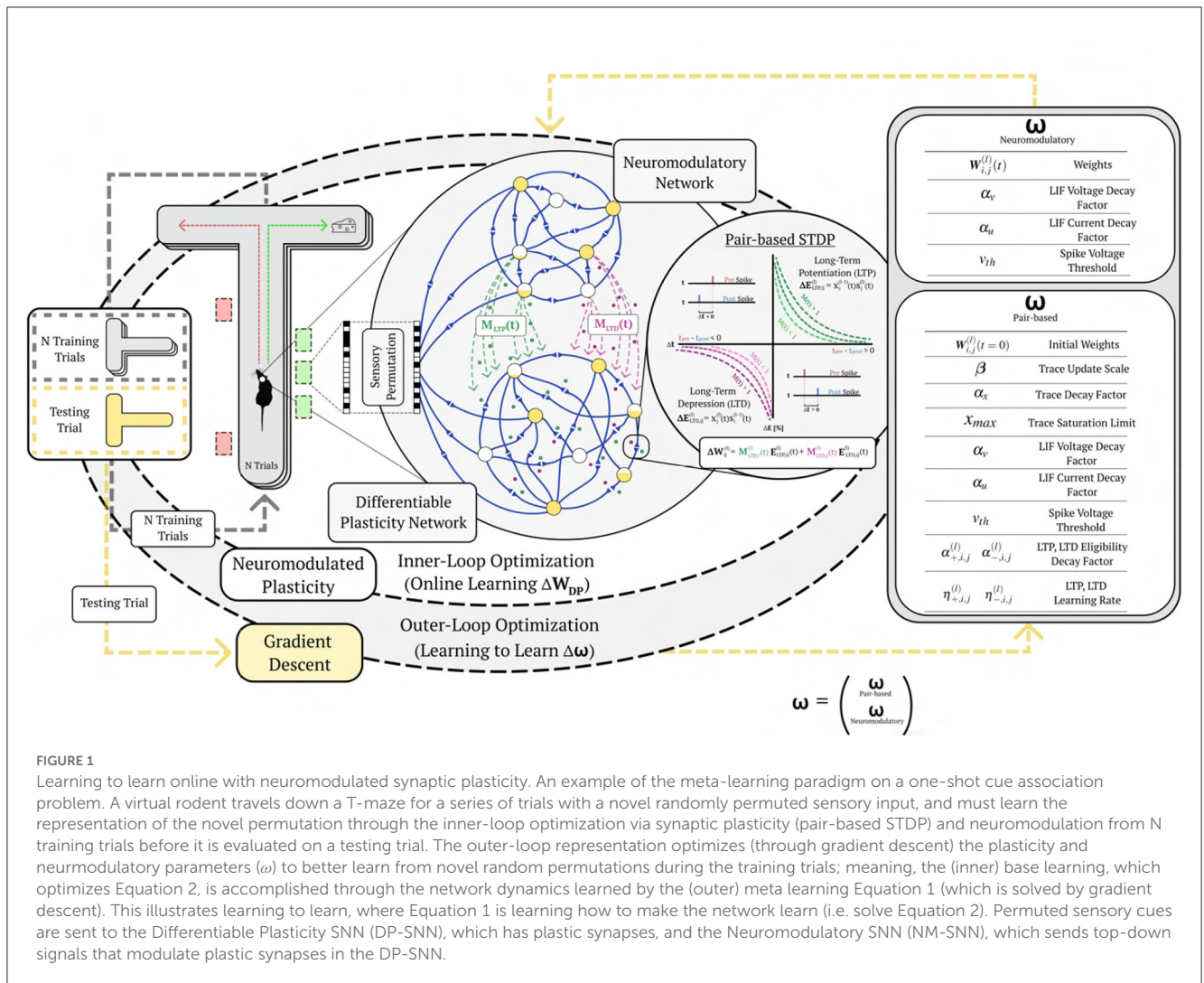


FIGURE 1 Learning to learn online with neuromodulated synaptic plasticity. An example of the meta-learning paradigm on a one-shot cue association problem. A virtual rodent travels down a T-maze for a series of trials with a novel randomly permuted sensory input, and must learn the representation of the novel permutation through the inner-loop optimization via synaptic plasticity (pair-based STDP) and neuromodulation from N training trials before it is evaluated on a testing trial. The outer-loop representation optimizes (through gradient descent) the plasticity and neuromodulatory parameters (ω) to better learn from novel random permutations during the training trials; meaning, the (inner) base learning, which optimizes Equation 2, is accomplished through the network dynamics learned by the (outer) meta learning Equation 1 (which is solved by gradient descent). This illustrates learning to learn, where Equation 1 is learning how to make the network learn (i.e. solve Equation 2). Permuted sensory cues are sent to the Differentiable Plasticity SNN (DP-SNN), which has plastic synapses, and the Neuromodulatory SNN (NM-SNN), which sends top-down signals that modulate plastic synapses in the DP-SNN.

synapses to store and process the necessary information relating the permuted input cues and the learning signal. This signal is only given to the neuromodulatory network during the training data phase (Figure 2A). The environment learning signal is two-dimensional binary vector, with the first element as one during a right-cue task, the second element as one during a left cue-task, and each element is otherwise zero.

Results

Synaptic plasticity occurs continuously at every moment in time rather than during select periods. Task specific knowledge is not able to be transferred between cue streams since cue frequency, cue ordering, and input permutations are randomly ordered. Rather information must be transferred between cue streams by improving online learning via the optimization of the meta-representation of plasticity—improving the learning algorithm itself (i.e. solving Equation 1). Recalling the definition of continual learning from Delange et al. (2021), information within a cue stream must be retained and improved upon across the two presented training trials without clear task divisions

being provided. Unimportant information in the form of noise neurons and random cue firings must be selectively recognized and forgotten. Critically, this requires the optimized learning algorithm to store learned information in synapses from the training cue trials without catastrophically forgetting in order to solve the testing cue trial.

A representative trial of the one-shot cue association problem is shown in Figure 2. Performance on the testing set of novel cue permutations yields 95.6% accuracy, which is averaged across 30 trainings with different randomly initialized parameters. Figure 4 demonstrates the performance accuracy of the network demonstrated in Figure 3 when the number of cues presented, M, are varied from 1-15. Interestingly, while the network plasticity rule was only optimized for $M = 5$ cues, the learning behavior exhibits the capability of accurately solving cue problems above and below the number of cues it was optimized for without additional training. Below $M = 5$, $M = 1$ obtains 98.1% accuracy and $M = 3$ obtains 96.7% accuracy. Above $M = 5$, there is a consistent loss in accuracy from $M = 7$ with 94.2% to $M = 15$ with 68.7%. These results demonstrate that the learned neuromodulated plasticity rule generalizes in the task solving domain with respect to the number of cues without additional training on the meta-representation.

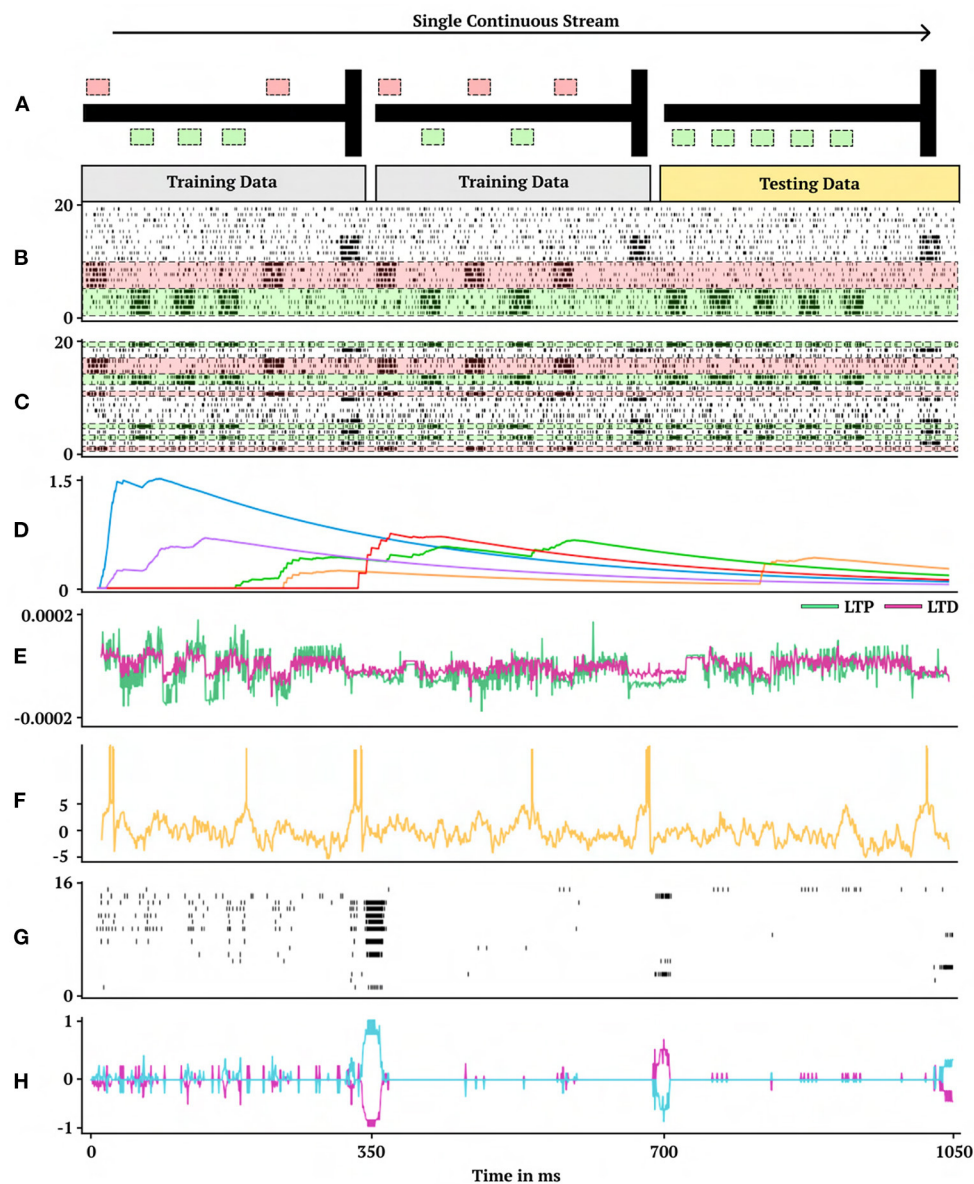


FIGURE 2

One-shot cue association. Visual demonstration of the continual one-shot learning paradigm for a trained neuromodulatory network. (A) Graphical interpretation of cue-association task. Two training data mazes are presented in a random order, one from each class of right (first maze) and left (second maze) cues followed by a testing data maze. Gray corresponds to training data which receives no reward to backpropagate and gold corresponds to testing data which does receive reward to backpropagate. (B) Non-permuted sensory information represented as spikes indexed from 0 to 20 (bottom to top). (C) Permuted form of sensory information presented in (B) indexed by which neuron receives spikes. (D) Eligibility trace dynamics (Methods, Equation 16) sampled from five random synapses. (E) LTP (green) and LTD (magenta) neuromodulatory dynamics from a random modulatory neuron. (F) Activity of a hidden neuron. (G) Sample of 16 hidden neuron spiking activity. (H) Action neuron activity.

Recognizing novel character classes from a single example

Several learning challenges are presented in Lake et al. (2017) with the aim of providing benchmarks that more closely demonstrate human-like intelligence in machines. The first among these challenges is the “characters challenge,” which aims at benchmarking a learning algorithm’s ability to recognize digits with few examples. The dataset for this challenge contains 1623 classes of handwritten characters across 50 unique alphabets, with

each character consisting of 20 samples (Lake et al., 2015). In this challenge, a learner is presented with a phase 1 image as well as a set of phase 2 images (Figure 5) where, one image presented is from the same phase 1 image class, and several other images presented are from other image classes. The phase 2 images are all presented simultaneously, and the learner must determine which image from phase 2 is in the phase 1 class. In the original design of this task, each image is able to be observed and compared simultaneously, and the image most closely matching the phase 1 image can be compared directly. A more challenging variation of this problem

which aligns more closely to biological learning is presented in Scherr et al. (2020), where each sample from phase 1 and phase 2 is presented sequentially instead of the learner being able to view and compare all samples simultaneously. The problem is considered solved correctly if the learner has the highest output activity for the image in phase 2 that matches the image class from phase 1. This variation of the characters challenge requires the learner to address the problem of holding information in memory across time and actively comparing that information with subsequently presented data, which even presents itself as a challenge for humans. Informal human testing from Scherr et al. (2020) demonstrates error rates around 15% based on 4 subjects and 100 trials.

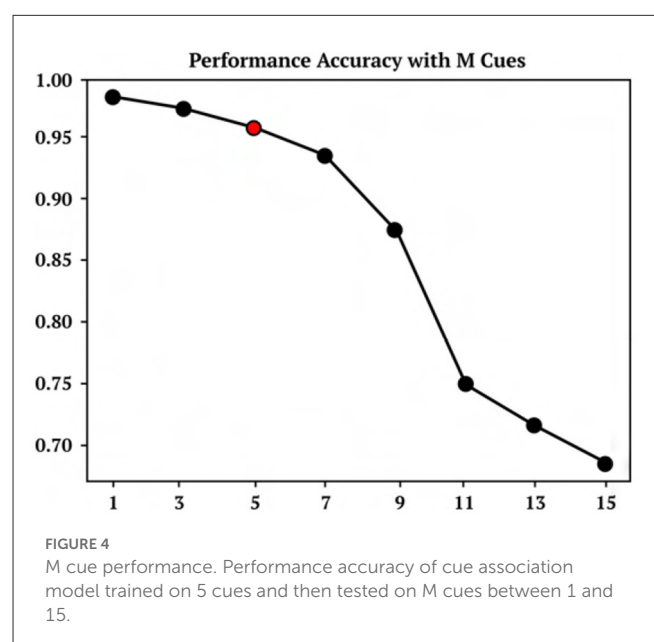
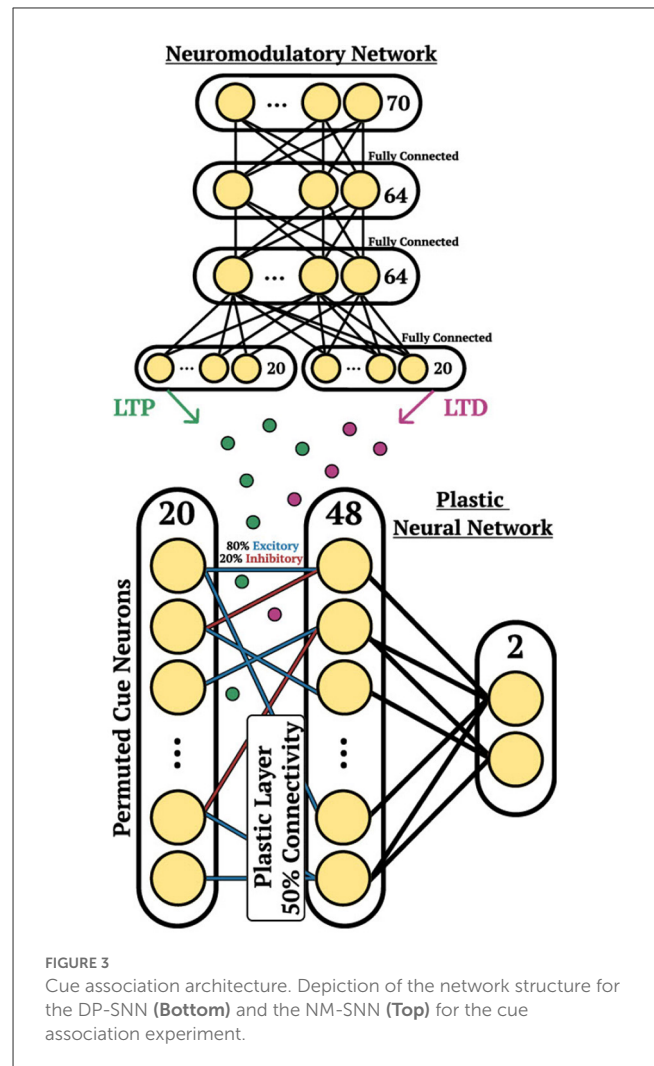
Experimental setup

Both phase 1 and phase 2 images are presented for 20 ms with a simulation step size of 1 ms. One image is presented in phase 1 and five images are presented in phase 2 for a total trajectory time of 120 ms. This causes the character challenge to be particularly difficult because the set of testing tasks is much larger than the set of training tasks. It is argued that the character presentation should be intentionally small such that the learner must carry out spike-based computation and learning vs. rate-based (Scherr et al., 2020). This time span is small compared to the average human visual reaction time which is around 331 ms (Jose and Gideon Praveen, 2010). The phase 1 and phase 2 character classes are selected uniformly from a categorical distribution and the phase 2 characters are organized with random ordering. Neuromodulatory signals are only sent by the NM-SNN to the DP-SNN during the 20 ms presentation of the phase 1 character. During this period, the synapses must be modified to recognize the phase 2 image that belongs to the same character class as the phase 1 image.

To increase the number of classes in the character dataset each character set is rotated by 90, 180, and 270 degrees, and are considered independent classes, increasing the number of character classes from 1623 to 6492. The character classes in the dataset are split into 20% testing and 80% training. There are $1.2 \cdot 10^{19}$ possible just on the ordering of character class arrangements in phase 2 in this problem, making it unlikely for the experiment to repeat any particular trial. Each gradient is computed across 256 cue trials and the model is updated for 2000 updates (Figure 5D).

Architecture

The character image is fed into a several layers of a CNN for pre-processing and is flattened at the output. The flattened output is used as current input to a layer of 196 spiking neurons, which represent the input of the DP-SNN and the NM-SNN, see Figure 6. The DP-SNN consists of one hidden layer with 48 CUBA neurons (Methods, Equations 6, 7). Synaptic connections between the 196 input neurons and the 48 hidden layer neurons store LTP and LTD dynamics in separate eligibility traces based on an additive triplet based STDP rule (Methods, Equation 15). The triplet-based STDP provides a more accurate representation of biological STDP dynamics compared with the pair-based rule through the use of a *slow* and *fast* post-synaptic trace which



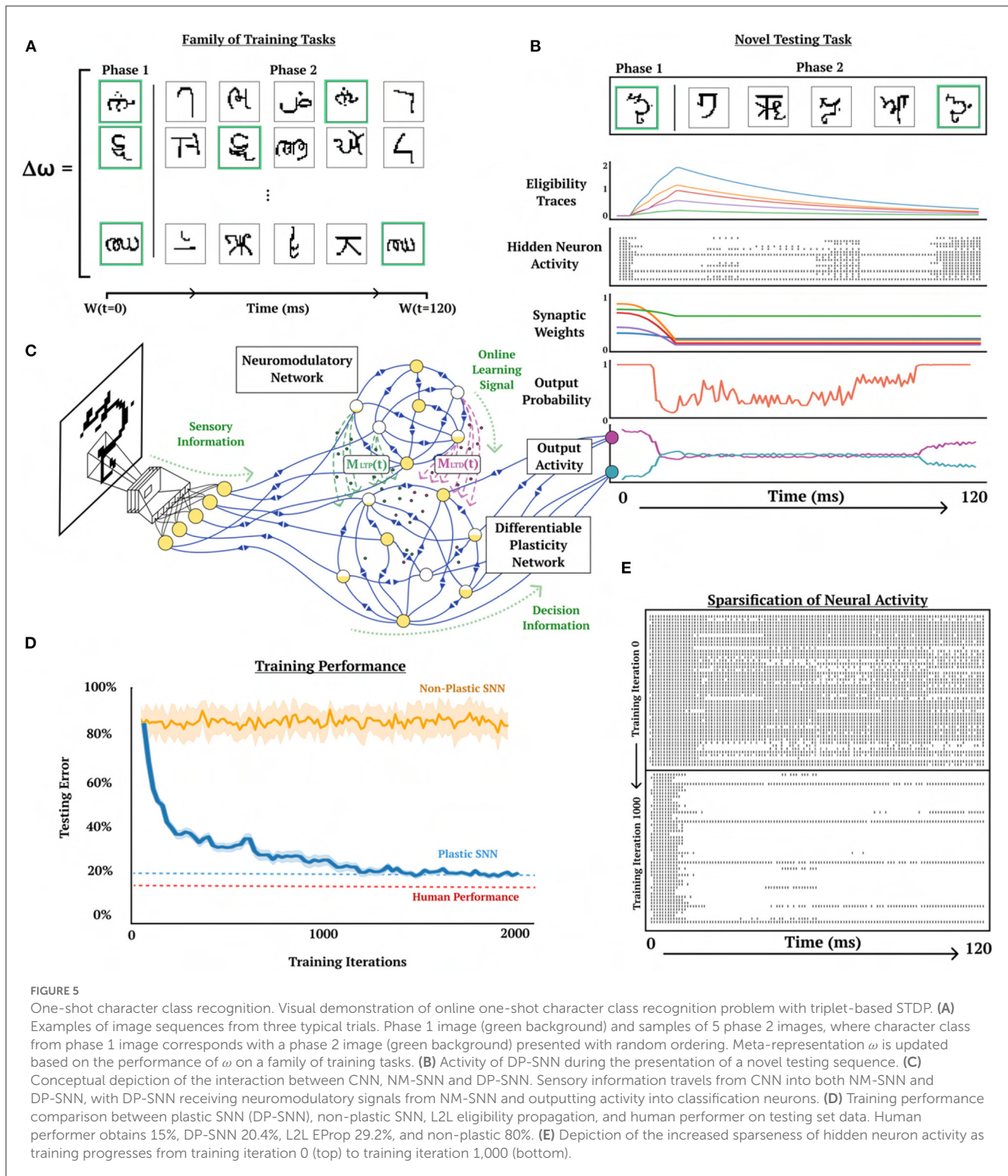


FIGURE 5

One-shot character class recognition. Visual demonstration of online one-shot character class recognition with triplet-based STDP. (A) Examples of image sequences from three typical trials. Phase 1 image (green background) and samples of 5 phase 2 images, where character class from phase 1 image corresponds with a phase 2 image (green background) presented with random ordering. Meta-representation ω is updated based on the performance of ω on a family of training tasks. (B) Activity of DP-SNN during the presentation of a novel testing sequence. (C) Conceptual depiction of the interaction between CNN, NM-SNN and DP-SNN. Sensory information travels from CNN into both NM-SNN and DP-SNN, with DP-SNN receiving neuromodulatory signals from NM-SNN and outputting activity into classification neurons. (D) Training performance comparison between plastic SNN (DP-SNN), non-plastic SNN, L2L eligibility propagation, and human performer on testing set data. Human performer obtains 15%, DP-SNN 20.4%, L2L EProp 29.2%, and non-plastic 80%. (E) Depiction of the increased sparseness of hidden neuron activity as training progresses from training iteration 0 (top) to training iteration 1,000 (bottom).

accumulate post-synaptic activity with varied trace decay factors (Methods, Equation 14). Connection probabilities between neurons are set to 50% during initialization, with 20% inhibitory synapses and 80% excitatory. Modulatory signals are produced by the NM-SNN using an input neuron specific modulatory signal (Methods, Equation 19) for both the LTP and LTD dynamics. The NM-SNN receives input from the image layer spiking neurons along

with the DP-SNN hidden neuron activity. However, to make the challenge more difficult, the NM-SNN does not receive any additional inputs and must generate neuromodulation from the same sensory information as the DP-SNN. The NM-SNN consists of two layers of 64 CUBA neurons with fully-connected non-plastic synapses. The pre-processing CNN consists of the following steps: (1) convolution from 1 to 4 channels with a kernel size

of 3, (2) batch norm, (3) ReLU operation, (4) max pooling with kernel and stride size of 2, (5) convolution from 4 to 4 channels with a kernel size of 3, (6) batch norm, (7) ReLU operation, and (8) a max pool with kernel and stride size of 2. This is then flattened and are used as current input to a 196 CUBA neurons which act as input for the DP-SNN and the NM-SNN.

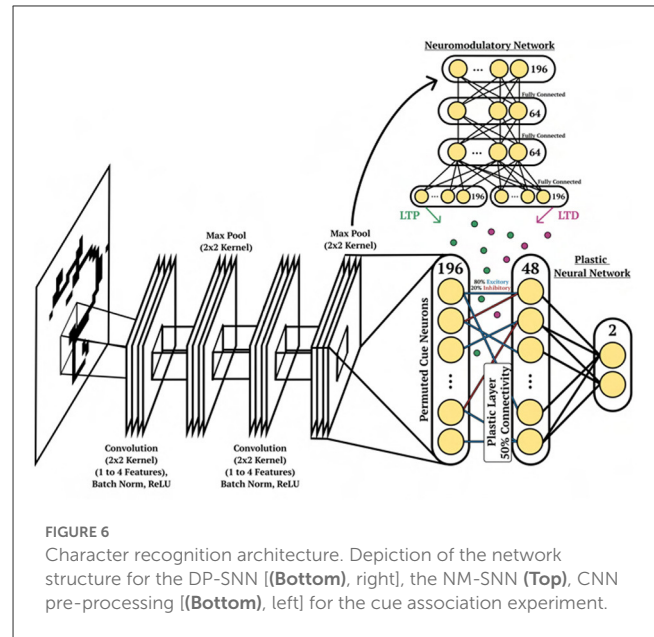
Results

The performance of the NM-SNN and DP-SNN is compared to a non-plastic SNN using the same connective structure. The non-plastic SNN is demonstrated to be unable to solve this task with a testing error average of around 80%, which is equivalent to random selection. On the other hand, the DP-SNN obtains a testing error of 20.4% after 2000 gradient steps on the outer-loop. This performance is comparable to informal human testing (Scherr et al., 2020) which is around 15%. Additionally, this performance is compared with its most similar counterpart, the L2L Eligibility Propagation method introduced in Scherr et al. (2020), which obtained a 29.2% accuracy. A surprising finding was that the DP-SNN obtains 64.1% accuracy on MNIST digits without any additional gradient steps on the plasticity parameters.

Discussion

In this paper, we introduce a method for learning to learn with neuroscience models of synaptic plasticity in networks of spiking neurons, where the neuromodulated plasticity dynamics are learned through gradient descent and online learning tasks are solved with the learned neuromodulated plasticity dynamics online. This framework was demonstrated on two challenging online learning tasks: a one-shot continual learning problem and a one-shot image class recognition problem. These challenges required neuromodulated plasticity to act as the mechanism of intra-lifetime learning, and presented a way for learning the parameters of plasticity with gradient descent such that it can address these problems.

Previous work on the development of online SNN learning algorithms includes the work of e-prop (Bellec et al., 2020), which is a plasticity rule that was mathematically derived from BPTT, where a learning signal defined by a given loss function over a task is projected to all neurons in the SNN using random feedback connections. This projected feedback interacts with an eligibility trace that accumulates the BPTT plasticity approximation to update synaptic weights. E-prop was demonstrated to be competitive with BPTT on several temporal learning benchmarks. In Scherr et al. (2020), a method called *natural e-prop* is introduced, which uses the plasticity dynamics of e-prop and learns a neuromodulatory signal toward solving several one-shot learning challenges. Another online learning algorithms for SNNs is Surrogate-gradient Online Error triggered Learning (SOEL) (Stewart et al., 2020). SOEL calculates a global error signal and uses surrogate gradient descent to create a plasticity-like rule for updating the network synapses online. Works like e-prop and SOEL are not competing algorithms, but rather are complimentary with respect to this framework.



The e.g. timing parameters, voltage parameters, and, the surrogate gradient parameters could be learned by gradient descent using these methods as the inner-loop optimization to produce an even more effective version of the existing algorithm. There have also been many previous contributions toward neuromodulated plasticity in non-spiking Artificial Neural Networks (ANNs) (Soltoggio et al., 2008; Risi and Stanley, 2012; Velez and Clune, 2017; Beaulieu et al., 2020; Miconi et al., 2020). However, plastic ANNs have been demonstrated to struggle maintaining functional stability across time due to their continuous nature which causes synapses to be in a constant state of change (Schmidgall and Hays, 2021). The effect of this instability was shown to not disturb the performance as significantly in plastic SNNs as it did on plastic ANNs. In addition, ANNs cannot be utilized together with neuromorphic hardware, impeding their applicability for low-power edge computing devices. While the learning to learn framework can produce powerful learning dynamics, one of the primary limitations of this work is the compute requirements. This is because to calculate the gradient, for each parameter of *each synapse* the error at the end of an episode must backpropagation all the way back through time which grows polynomial in complexity with respect to the episode duration. However, this is also a problem present in other learning to learn paradigms (Scherr et al., 2020).

To realize the full potential of this framework, described here are several topics for future research, including: incorporating cell-type specific neuromodulatory signals (Liu et al., 2021) into the learning process; exploring the addition of glial cell dynamics (Gordleeva et al., 2021; Ivanov and Michmizos, 2021); providing deeper insight into the learning capabilities of different plasticity rules in the neuroscience literature, such as the wide-range of existing voltage-dependent plasticity rules, rate-based plasticity rules, and spike-timing dependent plasticity rules; and exploring the use of this framework on robotic and reinforcement learning experiments. Another direction might explore learning the neural architecture in conjunction with the plasticity parameters,

since architecture is known to play a significant role in the function of neural dynamics (Gaier and Ha, 2019). Recent works have explored learning the plasticity rule equation in addition to the plasticity rule parameters (Jordan et al., 2021). A differentiable plasticity rule search constrained toward biological realism may provide more powerful learning applications of this framework.

Finally, addressing the problem of online learning has been a central focus of neuromorphic computing (Davies et al., 2021). The existing need for learning methods that can be used on these systems has impeded the use of neuromorphic systems in real-world applications. From a practical perspective, backpropagation on these systems is only envisioned as a utility for offline training since on-chip BPTT is expensive with respect to complexity, memory, and energy efficiency, and is not naturally suited for online learning. Instead, some neuromorphic systems have invested in *on-chip plasticity* in part to address online learning in hopes that an effective method for utilizing this capability is discovered. Neuromorphic processors implement *on-chip plasticity* by allowing the flexible reconfiguration of a set of local variables that interact to adapt synaptic weights (Jin et al., 2010; Davies et al., 2018, 2021; Rajendran et al., 2019; Pehle et al., 2022). The reconfiguration of these variables have historically modeled learning rules from the neuroscience literature. In spite of this, the goal of finding learning rules that can solve a wide variety of challenging problems (like backpropagation) while building off of the impressive capabilities of the brain remains open. We hope that this framework of learning to learn with backpropagation inspires the next generation of on-chip learning algorithms for the field of neuromorphic computing.

The framework of learning to learn with neuromodulated synaptic plasticity in this paper provides a method for combining the power of gradient descent with neuroscience models of plasticity, which opens the doors toward a better synthesis of machine learning and neuroscience.

Methods

Leaky Integrate-and-Fire

The Leaky Integrate-and-Fire (LIF) neuron model is a phenomenological model of neural firing-dynamics. Activity is integrated into the neuron and stored across time, and, once the stored activity surpasses a threshold value, a binary signal is emitted and the voltage is reset. The “leaky” part of the model name refers to an introduced time-dependent decay dynamic acting on the membrane potential. While the simplicity of the LIF dynamics deviates from the complexity of the biological neuron, the purpose of the model is to capture the essence of neuron dynamics while providing value from a computational perspective. The LIF neuron model requires among the fewest computational operations to implement compared with other neuron models.

To begin describing the LIF dynamics, we represent the continuous difference equation $\tau \frac{dv_j}{dt}$ for the voltage state $v_j(t) \in \mathbb{R}$ as a discrete time equation $v_j + \tau \frac{dv_j}{dt} = v_j(t + \Delta\tau)$ since computational models of spiking neurons typically operate across discrete update intervals.

$$v_j(t + \Delta\tau) = v_j(t) - \alpha_v[v_j(t) - v_{rest}] + RI_j(t), \quad (3)$$

The term $\alpha_v[v_j(t) - v_{rest}]$ represents the membrane potential leak, where $\alpha_v \in [0, 1]$ is the leak time-constant and $v_{rest} \in \mathbb{R}$ as the neuron resting potential, which is the value that the membrane potential returns to in the absence of external activity. $I_j(t) \in \mathbb{R}$ represents incoming current, which is the source of an increase in voltage $v_j(t)$ for the neuron j . This current is scaled by a resistance factor $R \in \mathbb{R}$.

$$s_j(t) = H(v_j(t)) = \begin{cases} 0 & v_j(t) \leq v_{th} \\ 1 & v_j(t) > v_{th} \end{cases}, \quad (4)$$

$H: \mathbb{R} \rightarrow \{0, 1\}$ is a piece-wise step function which, in the case of a spiking neuron, outputs 1 when a neuron’s membrane potential surpasses the defined firing threshold $v_{th} \in \mathbb{R}$ and otherwise outputs 0. In the LIF neuron model, once a neuron fires a spike, the membrane potential is reset to its resting potential $v_j(t) \leftarrow v_{rest}$.

In a spiking neural network, $I_j(t)$ from Equation 3 is defined as $I_j = \sum_i W_{ij}s_i(t)$, which represents the sum of weighted spikes from all pre-synaptic neurons i that are connected to post-synaptic neuron j . The weight of each spike is given by $W_{ij}(t) \in \mathbb{R}$, with $W_{ij}(t) < 0$ representing inhibitory connections, and $W_{ij}(t) > 0$ representing excitatory connections.

$$v_j(t + \Delta\tau) = v_j(t) - \alpha_v[v_j(t) - v_{rest}] + R \sum_i W_{ij}(t)s_i(t), \quad (5)$$

Consistent with Intel’s neuromorphic processor code named Loihi, our experiments use an adaptation of the LIF which incorporates current called the Current-based Leaky-integrate and fire (CUBA) neuron model (Davies et al., 2018).

$$u_i(t + \Delta\tau) = u_i(t) - \alpha_u[u_i(t) - u_{rest}] + \sum_j W_{ij}(t)s_j(t), \quad (6)$$

$$v_i(t + \Delta\tau) = v_i(t) - \alpha_v[v_i(t) - v_{rest}] + Ru_i(t). \quad (7)$$

In the CUBA neuron model, a decaying current trace $u_i(t)$ integrates incoming current $I_j = \sum_i W_{ij}s_i(t)$ from pre-synaptic neurons i into the post-synaptic current trace j in Equation 6. Then instead of current I_j directly modifying the neuron membrane potential $v_i(t)$ in Equation 7, the current trace $u_i(t)$ takes its place.

Backpropagation through spiking neurons

The role of $H(\cdot)$ in Equation 4 can be viewed analogously to the non-linear activation function used in artificial neural networks. However, unlike most utilized non-linearities, $H(\cdot)$ is non-differentiable, and hence backpropagating gradients becomes particularly challenging. To backpropagate through the non-differentiable function $H(\cdot)$, Spike Layer Error Reassignment in Time (SLAYER) is used. SLAYER represents the derivative of the spike function $H(\cdot)$ with a surrogate gradient, and backpropagates error through a temporal credit assignment policy (Shrestha and Orchard, 2018).

Spike-timing based plasticity rules

Spike-timing Dependent Plasticity rules, unlike rate-based models, are dependent on the relationship between precise spike-timing events in pre- and post-synaptic neurons (Gerstner et al., 1996; Markram et al., 1997; Bi and Poo, 1998; Sjöström et al., 2001). Equations for neuronal and synaptic plasticity dynamics are presented as discrete-time update equations as opposed to continuous-time equations to provide a closer correspondence to the computational implementation.

Synaptic traces

STDP can be defined as an iterative update rule through the use of synaptic activity traces.

$$\mathbf{x}_i^{(l)}(t + \Delta\tau) = \alpha_x \mathbf{x}_i^{(l)}(t) + f(\mathbf{x}_i^{(l)}(t)) \mathbf{s}_i^{(l)}(t). \quad (8)$$

The bio-physical meaning of the activity trace $\mathbf{x}_i^{(l)}(t) \in \mathbb{R} > 0$ is left abstract, as there are several candidates for the representation of this activity. For pre-synaptic events, this quantity could represent the amount of bound glutamate or the quantity of activated NMDA receptors, and for post-synaptic events the synaptic voltage by a backpropagating action potential or by calcium entry through a backpropagating action potential.

The variable $\alpha_x \in (0, 1)$ is traditionally represented as a quantity $(1 - \frac{1}{\tau})$, which decays the activity trace to zero at a rate inversely proportional to the magnitude of the time constant $\tau \in \mathbb{R} > 1$. The trace $\mathbf{x}_i^{(l)}(t)$ is updated by a quantity proportional to $f: \mathbb{R} \rightarrow \mathbb{R}$ in the presence of a spike $\mathbf{s}_i^{(l)}(t)$. This synaptic trace is referred to as an all-to-all synaptic trace scheme since each pre-synaptic spike is paired with every post-synaptic spike in time indirectly via the decaying trace.

In the linear case of this update rule $f(\mathbf{x}_i^{(l)}(t)) = \beta \in \mathbb{R} > 0$, the trace is updated by a constant factor β in the presence of a spike $\mathbf{s}_i^{(l)}(t)$.

$$\mathbf{x}_i^{(l)}(t + \Delta\tau) = \alpha_x \mathbf{x}_i^{(l)}(t) + \beta \mathbf{s}_i^{(l)}(t). \quad (9)$$

Another candidate for the function $f(\mathbf{x}_i^{(l)}(t))$ is $\beta[x_{max} - \mathbf{x}_i^{(l)}(t)]$, which updates the trace by a constant β together with a factor $[x_{max} - \mathbf{x}_i^{(l)}(t)]$ that scales the update depending on the relationship between $\mathbf{x}_i^{(l)}(t)$ and its proximity to the trace saturation point $x_{max} \in \mathbb{R} > 0$ (Morrison et al., 2008).

$$\mathbf{x}_i^{(l)}(t + \Delta\tau) = \alpha_x \mathbf{x}_i^{(l)}(t) + \beta [x_{max} - \mathbf{x}_i^{(l)}(t)] \mathbf{s}_i^{(l)}(t). \quad (10)$$

When $\beta < 1$, as $\mathbf{x}_i^{(l)}(t)$ approaches x_{max} , the update scale $[x_{max} - \mathbf{x}_i^{(l)}(t)]$ reduces the magnitude of the trace update, producing a soft-bounded range $0 \leq \mathbf{x}_i^{(l)}(t) \leq x_{max}$.

Pair-based STDP

The pair-based model of STDP describes a plasticity rule from which synapses are changed as a product of the timing relationship

between pairs of pre- and post-synaptic activity.

$$\begin{aligned} W_{ij}^{(l)}(t + \Delta\tau) = & W_{ij}^{(l)}(t) + A_{+,ij}(W_{ij}^{(l)}(t)) \mathbf{x}_i^{(l-1)}(t) \mathbf{s}_j^{(l)}(t) \\ & - A_{-,ij}(W_{ij}^{(l)}(t)) \mathbf{x}_j^{(l)}(t) \mathbf{s}_i^{(l-1)}(t). \end{aligned} \quad (11)$$

Weight potentiation is realized in the presence of a post-synaptic firing $\mathbf{s}_j^{(l)}(t) = 1$ by a quantity proportional to the pre-synaptic trace $\mathbf{x}_i^{(l-1)}(t)$. Likewise, weight depression is realized in the presence of a pre-synaptic $\mathbf{s}_i^{(l-1)}(t) = 1$ proportional to the post-synaptic trace $\mathbf{x}_j^{(l)}(t)$. Potentiation and depression are respectively scaled by $A_{+,ij}: \mathbb{R} \rightarrow \mathbb{R}$ and $A_{-,ij}: \mathbb{R} \rightarrow \mathbb{R}$, which are functions that characterize the update dependence on the current weight of the synapse $W_{ij}^{(l)}(t)$. Hebbian pair-based STDP models generally define $A_{+,ij}(W_{ij}^{(l)}(t)) > 0$ and $A_{-,ij}(W_{ij}^{(l)}(t)) > 0$, whereas anti-Hebbian models define $A_{+,ij}(W_{ij}^{(l)}(t)) < 0$ and $A_{-,ij}(W_{ij}^{(l)}(t)) < 0$.

Weight-dependence

An additive model of pair-based STDP defines $A_{+,ij}(W_{ij}^{(l)}(t)) = \eta_{+,ij}^{(l)}$, which scales LTP and LTD linearly by a factor $\eta_{+,ij}^{(l)} \in \mathbb{R}$ and $\eta_{-,ij}^{(l)} \in \mathbb{R}$ respectively.

$$\begin{aligned} W_{ij}^{(l)}(t + \Delta\tau) = & W_{ij}^{(l)}(t) + \eta_{+,ij}^{(l)} \mathbf{x}_i^{(l-1)}(t) \mathbf{s}_j^{(l)}(t) \\ & - \eta_{-,ij}^{(l)} \mathbf{x}_j^{(l)}(t) \mathbf{s}_i^{(l-1)}(t). \end{aligned} \quad (12)$$

Additive models of STDP demonstrate strong synaptic competition, and hence tend to produce clear synaptic specialization (Gilson and Fukai, 2011). However, without any dependence on the weight parameter for regulation, the weight dynamics may grow either without bound or, with hard bounds, bimodally (Rubin et al., 2001; Morrison et al., 2008; Gilson and Fukai, 2011).

A multiplicative, or weight dependent, model of pair-based STDP defines $A_{+,ij}(W_{ij}^{(l)}(t)) = \eta_{+,ij}(W_{max} - W_{ij}^{(l)}(t))$ for LTP, which scales the effect of potentiation based on the proximity of the weight $W_{ij}^{(l)}(t)$ to the defined weight soft upper-bound W_{max} . Similarly, LTD defines $A_{-,ij}(W_{ij}^{(l)}(t)) = \eta_{-,ij}(W_{ij}^{(l)}(t) - W_{min})$, which scales weight depression according to the defined soft-lower bound W_{min} .

$$\begin{aligned} W_{ij}^{(l)}(t + \Delta\tau) = & W_{ij}^{(l)}(t) + \eta_{+,ij}(W_{max} - W_{ij}^{(l)}(t)) \mathbf{x}_i^{(l-1)}(t) \mathbf{s}_j^{(l)}(t) \\ & - \eta_{-,ij}(W_{ij}^{(l)}(t) - W_{min}) \mathbf{x}_j^{(l)}(t) \mathbf{s}_i^{(l-1)}(t). \end{aligned} \quad (13)$$

LTP and LTD produce weight changes depending on their relationship to the upper- and lower-bound, with LTP more effective when weights are farther from the upper-bound and LTD more effective when weights are farther from the lower bound. The use of soft bounds in practice leads to LTD dominating over LTP (Kempster et al., 1999; Song et al., 2000; Rubin et al., 2001; van Rossum et al., 2001; Gilson and Fukai, 2011) and, opposite to additive pair-based STDP, fails to demonstrate clear synaptic specialization (Gilson and Fukai, 2011).

Additive and multiplicative models of STDP have been regarded as extremes among a range of representations, with

LTP as $A_{+,ij}(W_{ij}^{(l)}(t)) = \eta_{+,ij}(W_{max} - W_{ij}^{(l)}(t))^\mu$ and LTD as $A_{-,ij}(W_{ij}^{(l)}(t)) = \eta_{-,ij}(W_{ij}^{(l)}(t) - W_{min})^\mu$ (Gütig et al., 2003; Song et al., 2005). Here, the parameter μ acts as an exponential weight-dependence scale, with $\mu = 0$ producing an additive model, and $\mu = 1$ producing a multiplicative model. Values of $0 < \mu < 1$ result in rules with intermediate dependence on $W_{ij}^{(l)}(t)$.

Triplet-based STDP

Experimental data has demonstrated that pair-based STDP models cannot provide an accurate representation of biological STDP dynamics under certain conditions. Particularly, these rules cannot reproduce triplet and quadruplet experiments, and also cannot account for the frequency-dependence of plasticity demonstrated in STDP experiments (Senn et al., 2001; Sjöström et al., 2001).

$$\mathbf{x}_{i,\tau}^{(l)}(t + \Delta\tau) = \alpha_\tau \mathbf{x}_{i,\tau}^{(l)}(t) + f(\mathbf{x}_{i,\tau}^{(l)}(t)) \mathbf{s}_i^{(l)}(t) \quad (14)$$

To address the representation limitations of pair-based STDP, a plasticity rule based on a triplet interaction between one pre-synaptic spike and two post-synaptic spikes is proposed in Gjorgjieva et al. (2011). To implement this, a second *slow* synaptic trace is introduced for the post-synaptic neuron is introduced, with a time constant $\alpha_\tau \in \mathbb{R} > \alpha_x$, with α_x representing the decay rate of the *fast* synaptic trace from Equation (8). More specifically, the triplet model of STDP produces LTP dynamics that are dependent on the pre-synaptic trace $\mathbf{x}_i^{(l-1)}(t)$ [Equation (14)] and the *slow* post-synaptic trace $\mathbf{x}_{j,\tau}^{(l)}(t - \Delta\tau)$, which is evaluated at time $t - \Delta\tau$, one timestep prior to the evaluation of traces $\mathbf{x}_j^{(l)}(t)$ and $\mathbf{x}_i^{(l-1)}(t)$:

$$\begin{aligned} W_{ij}^{(l)}(t + \Delta\tau) = & W_{ij}^{(l)}(t) + A_{+,ij}(W_{ij}^{(l)}(t)) \mathbf{x}_i^{(l-1)}(t) \mathbf{x}_{j,\tau}^{(l)}(t - \Delta\tau) \\ & - A_{-,ij}(W_{ij}^{(l)}(t)) \mathbf{x}_j^{(l)}(t) \mathbf{s}_i^{(l)}(t). \end{aligned} \quad (15)$$

The triplet rule has demonstrated to explain several plasticity experiments more effectively than pair-based STDP (Sjöström et al., 2001; Wang et al., 2005; Gjorgjieva et al., 2011). Additionally, the triplet rule has been demonstrated to be capable of being mapped to the BCM rule under the assumption that (1) pre- and post-synaptic spiking behavior assumes independent stochastic spike trains, (2) LTD is produced in the presence of low post-synaptic firing rates, (3) LTP is produced in the presence of high post-synaptic firing rates, and (4) the triplet term is dependent on the average post-synaptic firing frequency (Pfister and Gerstner, 2006). If these requirements are matched, the presented triplet-based STDP rule demonstrates the properties of the BCM rule, such as synaptic competition which produces input selectivity, a requirement for receptive field development (Bienenstock et al., 1982; Pfister and Gerstner, 2006).

Neuromodulatory plasticity rules

Synaptic learning rules in the context of SNNs mathematically describe the change in synaptic strength between a pre-synaptic neuron i and post-synaptic neuron j . At the biological level, these

changes are products of complex dynamics between a diversity of molecules interacting at multiple time-scales. Many behaviors require the interplay of activity on the time-scale of seconds to minutes, such as exploring a maze, and on the time-scale of milliseconds, such as neuronal spiking. Learning rules must be capable of effectively integrating these two diverse time-scales. Thus far, the learning rules observed have been simplified to equations which modify the synaptic strength $W_{ij}^{(l)}(t)$ based on local synaptic activity without any motivating guidance and without the presence of external modulating factors.

Biological experiments have demonstrated that synaptic plasticity is often dependent on the presence of neuromodulators such as dopamine (Schultz et al., 1993; Seamans, 2007; Zhang et al., 2009; Steinberg et al., 2013; Speranza et al., 2021), noradrenaline (Ranganath and Rainer, 2003; Salgado et al., 2012), and acetylcholine (Ranganath and Rainer, 2003; Hasselmo, 2006; Teles-Grilo Ruivo and Mellor, 2013; Brzosko et al., 2015; Zannone et al., 2018). These modulators often act to regulate plasticity at the synapse by gating synaptic change, with recent evidence suggesting that interactions more complex than gating occur (Zhang et al., 2009; Frémaux and Gerstner, 2016; Gerstner et al., 2018). The interaction between neuromodulators and eligibility traces has served as an effective paradigm for many biologically-inspired learning algorithms (Seung, 2003; Frémaux et al., 2013; Bing et al., 2018; Bellec et al., 2019, 2020).

Eligibility traces

Rather than directly modifying the synaptic weight, local synaptic activity leaves an activity flag, or eligibility trace, at the synapse. The eligibility trace does not immediately produce a change, rather, weight change is realized in the presence of an additional signal. In the theoretical literature on three-factor learning, this signal has been theorized to be accounted for by *external*, or non-local, activity (Marder, 2012; Frémaux and Gerstner, 2016; Gerstner et al., 2018). For learning applications, this third signal could be a prediction error, or for reinforcement learning, an advantage prediction (Frémaux et al., 2013). In a Hebbian learning rule, the eligibility trace can be described by the following equation:

$$E_{ij}^{(l)}(t + \Delta\tau) = \gamma E_{ij}^{(l)}(t) + \alpha_{ij} f_i(x_i^{(l-1)}) g_j(x_j^{(l)}). \quad (16)$$

The constant $\gamma \in [0, 1]$ inversely determines the rate of decay for the trace, $\alpha_{ij} \in \mathbb{R}$ is a constant determining the rate at which activity trace information is introduced into the eligibility trace, f_i is a function of pre-synaptic activity $x_i^{(l-1)}$, and g_j a function of post-synaptic activity $x_j^{(l)}$. These functions are indexed by their corresponding pre- and post-synaptic neuron i and j since the synaptic activity eligibility dynamics may be dependent on neuron type or the region of the network.

Both rate- and spike-based models of plasticity can be represented with the eligibility trace dynamics described in Equation (16). Spike-based models of plasticity, such as the triplet-based (Equation 15) and pair-based model (Equation 11), often require two synaptic flags $E_{+,ij}^{(l)}$ and $E_{-,ij}^{(l)}$ for LTP and LTD respectively.

Modulatory eligibility traces

In the theoretical literature, eligibility traces alone are not sufficient to produce a change in synaptic efficacy (Frémaux and Gerstner, 2016; Gerstner et al., 2018). Instead, weight changes are realized in the presence of a third signal.

$$W_{ij}^{(l)}(t + \Delta\tau) = W_{ij}^{(l)}(t) + M(t)E_{ij}^{(l)}(t). \quad (17)$$

Here, $M(t) \in \mathbb{R}$ acts as a global third signal which is referred to as a neuromodulator. Weight changes no longer occur in the absence of the neuromodulatory signal, $M(t) = 0$. When the value $M(t)$ ranges from positive to negative values, the *magnitude* and *direction* of change is determined causing LTP and LTD to both scale and reverse in the presence of certain stimuli.

The interaction between individual neurons and the global neuromodulatory signal need not be entirely defined multiplicatively as in Equation 17, but can have neuron-specific responses defined by the following dynamics:

$$W_{ij}^{(l)}(t + \Delta\tau) = W_{ij}^{(l)}(t) + h_j(M(t))E_{ij}^{(l)}(t). \quad (18)$$

The function $h_j: \mathbb{R} \rightarrow \mathbb{R}$ is a neuron-specific response function which determines how the post-synaptic neuron j responds to the neuromodulatory signal $M(t)$. This form of neuromodulation accounts for random-feedback networks when $h_j(M(t)) = h(b_j M(t))$. However, this form of neuromodulation does not account for the general supervised learning paradigm through backpropagating error. Equation (18) must be extended to account for neuron-specific neuromodulatory signals:

$$W_{ij}^{(l)}(t + \Delta\tau) = W_{ij}^{(l)}(t) + M_j(t)E_{ij}^{(l)}(t). \quad (19)$$

In *layered* networks being optimized through backpropagation, the neuron-specific error is $M_j(t)$. In the case of backpropagation, $M_j(t)$ is calculated as a weighted sum from the errors in the neighboring layer closest to the output. The neuron-specific error in Equation 19 can also be computed with the dimensionality of the pre-synaptic neurons, $M_i(t)$, which was the form of neuromodulation used in both experiments from the Experiments section.

References

- Andrychowicz, M., Denil, M., Gomez, S., Hoffman, M. W., Pfau, D., Schaul, T., et al. (2016). Learning to learn by gradient descent by gradient descent. *Adv. Neural Inform. Process. Sys.* 29. doi: 10.48550/arXiv.1606.04474
- Beaulieu, S., Frati, L., Miconi, T., Lehman, J., Stanley, K. O., Clune, J., et al. (2020). Learning to continually learn. *arXiv preprint arXiv:2002.09571*.
- Bellec, G., Scherr, F., Hajek, E., Salaj, D., Legenstein, R., and Maass, W. (2019). Biologically inspired alternatives to backpropagation through time for learning in recurrent neural nets. *arXiv preprint arXiv:1901.09049*.
- Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., et al. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nat. Commun.* 11, 1–15. doi: 10.1038/s41467-020-17236-y
- Bello, I., Zoph, B., Vasudevan, V., and Le, Q. V. (2017). "Neural optimizer search with reinforcement learning," in *Proceedings of the 34th International Conference on Machine Learning* (Sydney: PMLR), 459–468.
- Bi, G.-q., and Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472. doi: 10.1523/JNEUROSCI.18-24-10464.1998
- Bienenstock, E. L., Cooper, L. N., and Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* 2, 32–48. doi: 10.1523/JNEUROSCI.02-01-00032.1982
- Bing, Z., Meschede, C., Huang, K., Chen, G., Rohrbein, F., Akl, M., et al. (2018). "End to end learning of spiking neural network based on r-stdp for a lane keeping vehicle," in *2018 IEEE International Conference on Robotics and Automation (ICRA)* (Brisbane, QLD: IEEE), 4725–4732. doi: 10.1109/ICRA.2018.8460482
- Bohnstingl, T., Scherr, F., Pehle, C., Meier, K., and Maass, W. (2019). Neuromorphic hardware learns to learn. *Front. Neurosci.* 13, 483. doi: 10.3389/fnins.2019.00483

Data availability statement

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

Author contributions

SS designed and performed the experiments as well as the analysis. SS wrote the paper with JH being active contributors toward editing and revising the paper as well as supervising the project. Both authors contributed to the article and approved the submitted version.

Funding

The program was funded by Office of the Under Secretary of Defense (OUSD) through the Applied Research for Advancement of S&T Priorities (ARAP) Program work unit IU64.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Brea, J., and Gerstner, W. (2016). Does computational neuroscience need new synaptic learning paradigms? *Curr. Opin. Behav. Sci.* 11, 61–66. doi: 10.1016/j.cobeha.2016.05.012
- Brzosko, Z., Schultz, W., and Paulsen, O. (2015). Retroactive modulation of spike timing-dependent plasticity by dopamine. *Elife* 4, e09685. doi: 10.7554/eLife.09685.017
- Clune, J. (2019). Ai-gas: Ai-generating algorithms, an alternate paradigm for producing general artificial intelligence. *arXiv preprint arXiv:1905.10985*.
- Co-Reyes, J. D., Miao, Y., Peng, D., Real, E., Levine, S., Le, Q. V., et al. (2021). Evolving reinforcement learning algorithms. *arXiv preprint arXiv:2101.03958*.
- Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., et al. (2018). Loihi: a neuromorphic manycore processor with on-chip learning. *Ieee Micro* 38, 82–99. doi: 10.1109/MM.2018.112130359
- Davies, M., Wild, A., Orchard, G., Sandamirskaya, Y., Guerra, G. A. F., Joshi, P., et al. (2021). Advancing neuromorphic computing with loihi: a survey of results and outlook. *Proc. IEEE* 109, 911–934. doi: 10.1109/JPROC.2021.3067593
- Deacon, R. M., and Rawlins, J. N. P. (2006). T-maze alternation in the rodent. *Nat. Protoc.* 1, 7–12. doi: 10.1038/nprot.2006.2
- Delange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., et al. (2021). A continual learning survey: defying forgetting in classification tasks. *IEEE Trans. Pattern Anal. Mach. Intell.* doi: 10.1109/TPAMI.2021.3057446
- Dudchenko, P. A. (2001). How do animals actually solve the t maze? *Behav. Neurosci.* 115, 850. doi: 10.1037/0735-7044.115.4.850
- Elbrecht, D., and Schuman, C. (2020). “Neuroevolution of spiking neural networks using compositional pattern producing networks,” in *International Conference on Neuromorphic Systems*, 1–5. doi: 10.1145/3407197.3407198
- Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., et al. (2019). Specialized coding of sensory, motor and cognitive variables in vta dopamine neurons. *Nature* 570, 509–513. doi: 10.1038/s41586-019-1261-9
- Fakoor, R., Chaudhari, P., Soatto, S., and Smola, A. J. (2019). Meta-q-learning. *arXiv preprint arXiv:1910.00125*.
- Finn, C., Abbeel, P., and Levine, S. (2017). “Model-agnostic meta-learning for fast adaptation of deep networks,” in *Proceedings of the 34th International Conference on Machine Learning* (Sydney: PMLR), 1126–1135.
- Frémaux, N., and Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Front. Neural Circuits* 9, 85. doi: 10.3389/fncir.2015.00085
- Frémaux, N., Sprekeler, H., and Gerstner, W. (2013). Reinforcement learning using a continuous time actor-critic framework with spiking neurons. *PLoS Comput. Biol.* 9, e1003024. doi: 10.1371/journal.pcbi.1003024
- French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends Cogn. Sci.* 3, 128–135. doi: 10.1016/S1364-6613(99)01294-2
- Gaier, A., and Ha, D. (2019). Weight agnostic neural networks. *Adv. Neural Inform. Process. Syst.* 32. doi: 10.48550/arXiv.1906.04358
- Gerstner, W., Kempter, R., Van Hemmen, J. L., and Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature* 383, 76–78. doi: 10.1038/383076a0
- Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., and Brea, J. (2018). Eligibility traces and plasticity on behavioral time scales: experimental support of neohobbian three-factor learning rules. *Front. Neural Circuits* 12, 53. doi: 10.3389/fncir.2018.00053
- Gilson, M., and Fukai, T. (2011). Stability versus neuronal specialization for stdp: long-tail weight distributions solve the dilemma. *PLoS ONE* 6, e25339. doi: 10.1371/journal.pone.0025339
- Gjorgjieva, J., Clopath, C., Audet, J., and Pfister, J.-P. (2011). A triplet spike-timing-dependent plasticity model generalizes the bienstock-cooper-munro rule to higher-order spatiotemporal correlations. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19383–19388. doi: 10.1073/pnas.1105933108
- Gordleeva, S. Y., Tsybina, Y. A., Krivososov, M. I., Ivanchenko, M. V., Zaikin, A. A., Kazantsev, V. B., et al. (2021). Modelling working memory in spiking neuron network accompanied by astrocytes. *Front. Cell. Neurosci.* 15, 86. doi: 10.3389/fncel.2021.631485
- Guo, Y., Chen, Y., Zhang, L., Liu, X., Wang, Y., Huang, X., et al. (2022). Im-loss: Information maximization loss for spiking neural networks. *Adv. Neural Inform. Process. Syst.* 35, 156–166.
- Gütig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric hebbian plasticity. *J. Neurosci.* 23, 3697–3714. doi: 10.1523/JNEUROSCI.23-09-03697.2003
- Hasselmo, M. E. (2006). The role of acetylcholine in learning and memory. *Curr. Opin. Neurobiol.* 16, 710–715. doi: 10.1016/j.conb.2006.09.002
- Hospedales, T., Antoniou, A., Micaelli, P., and Storkey, A. (2020). Meta-learning in neural networks: a survey. *arXiv preprint arXiv:2004.05439*. doi: 10.1109/TPAMI.2021.3079209
- Houthoofd, R., Chen, Y., Isola, P., Stadie, B., Wolski, F., Jonathan Ho, O., et al. (2018). Evolved policy gradients. *Adv. Neural Inform. Process. Syst.* 31.
- Hrvoj-Mihic, B., Bienvenu, T., Stefanacci, L., Muotri, A., and Semendeferi, K. (2013). Evolution, development, and plasticity of the human brain: from molecules to bones. *Front. Hum. Neurosci.* 7, 707. doi: 10.3389/fnhum.2013.00707
- Irie, K., Schlag, I., Csordás, R., and Schmidhuber, J. (2022). A modern self-referential weight matrix that learns to modify itself. *arXiv preprint arXiv:2202.05780*.
- Ivanov, V., and Michmizos, K. (2021). Increasing liquid state machine performance with edge-of-chaos dynamics organized by astrocyte-modulated plasticity. *Adv. Neural Inform. Process. Syst.* 34. doi: 10.48550/arXiv.2111.01760
- Jin, X., Rast, A., Galluppi, F., Davies, S., and Furber, S. (2010). “Implementing spike-timing-dependent plasticity on spinnaker neuromorphic hardware,” in *The 2010 International Joint Conference on Neural Networks (IJCNN)* (Barcelona: IEEE), 1–8. doi: 10.1109/IJCNN.2010.5596372
- Jordan, J., Schmidt, M., Senn, W., and Petrovici, M. A. (2021). Evolving interpretable plasticity for spiking networks. *Elife* 10, e66273. doi: 10.7554/eLife.66273
- Jose, S., and Gideon Praveen, K. (2010). Comparison between auditory and visual simple reaction times. *Neurosci. Med.* 2010.
- Kaiser, J., Hoff, M., Konle, A., Vasquez Tieck, J. C., Kappel, D., Reichard, D., et al. (2019). Embodied synaptic plasticity with online reinforcement learning. *Front. Neurobot.* 13, 81. doi: 10.3389/fnbot.2019.00081
- Kempter, R., Gerstner, W., and van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Phys. Rev. E* 59, 4498–4514. doi: 10.1103/PhysRevE.59.4498
- Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., et al. (2017). Overcoming catastrophic forgetting in neural networks. *Proc. Natl. Acad. Sci. U. S. A.* 114, 3521–3526. doi: 10.1073/pnas.1611835114
- Kulkarni, S., Parsa, M., Mitchell, J. P., and Schuman, C. (2021). “Training spiking neural networks with synaptic plasticity under integer representation,” in *International Conference on Neuromorphic Systems 2021*, 1–7. doi: 10.1145/3477145.3477152
- Lake, B. M., Salakhutdinov, R., and Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science* 350, 1332–1338. doi: 10.1126/science.aab3050
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behav. Brain Sci.* 40, e253. doi: 10.1017/S0140525X16001837
- Letz, B. T. (1975). Long delay learning in the t-maze. *Learn. Motiv.* 6, 80–90. doi: 10.1016/0023-9690(75)90036-3
- Li, Y., Guo, Y., Zhang, S., Deng, S., Hai, Y., and Gu, S. (2021). Differentiable spike: rethinking gradient-descent for training spiking neural networks. *Adv. Neural Inform. Process. Syst.* 34, 23426–23439.
- Lian, D., Zheng, Y., Xu, Y., Lu, Y., Lin, L., Zhao, P., et al. (2019). “Towards fast adaptation of neural architectures with meta learning,” in *International Conference on Learning Representations*.
- Liu, H., Simonyan, K., and Yang, Y. (2018). Darts: differentiable architecture search. *arXiv preprint arXiv:1806.09055*.
- Liu, H., Socher, R., and Xiong, C. (2019). “Taming maml: efficient unbiased meta-reinforcement learning,” in *Proceedings of the 36th International Conference on Machine Learning* (Long Beach, CA: PMLR), 4061–4071.
- Liu, Y. H., Smith, S., Mihalas, S., Shea-Brown, E., and Sümbül, U. (2021). Cell-type-specific neuromodulation guides synaptic credit assignment in a spiking neural network. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2111821118. doi: 10.1073/pnas.2111821118
- Lukasz Kusmierz, Isomura, T., and Toyozumi, T. (2017). Learning with three factors: modulating hebbian plasticity with errors. *Curr. Opin. Neurobiol.* 46, 170–177. doi: 10.1016/j.conb.2017.08.020
- Marder, E. (2012). Neuromodulation of neuronal circuits: back to the future. *Neuron* 76, 1–11. doi: 10.1016/j.neuron.2012.09.010
- Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic apss and epsps. *Science* 275, 213–215. doi: 10.1126/science.275.5297.213
- Metz, L., Maheswaranathan, N., Cheung, B., and Sohl-Dickstein, J. (2018). Meta-learning update rules for unsupervised representation learning. *arXiv preprint arXiv:1804.00222*.
- Miconi, T., Rawal, A., Clune, J., and Stanley, K. O. (2020). Backpropamine: training self-modifying neural networks with differentiable neuromodulated plasticity. *arXiv preprint arXiv:2002.10585*.
- Morcos, A. S., and Harvey, C. D. (2016). History-dependent variability in population dynamics during evidence accumulation in cortex. *Nat. Neurosci.* 19, 1672–1681. doi: 10.1038/nn.4403
- Morrison, A., Diesmann, M., and Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biol. Cybernet.* 98, 459–478. doi: 10.1007/s00422-008-0233-1
- Neftci, E. O., Mostafa, H., and Zenke, F. (2019). Surrogate gradient learning in spiking neural networks: bringing the power of gradient-based

- optimization to spiking neural networks. *IEEE Signal Process. Magaz.* 36, 51–63. doi: 10.1109/MSP.2019.2931595
- Nessler, B., Pfeiffer, M., and Maass, W. (2008). Hebbian learning of bayes optimal decisions. *Adv. Neural Inform. Process. Syst.* 21.
- Parisi, G. I., Kemker, R., Part, J. L., Kanan, C., and Wermter, S. (2019). Continual lifelong learning with neural networks: a review. *Neural Netw.* 113, 54–71. doi: 10.1016/j.neunet.2019.01.012
- Pehle, C., Billaudelle, S., Cramer, B., Kaiser, J., Schreiber, K., Stradmann, Y., et al. (2022). The brainscales-2 accelerated neuromorphic system with hybrid plasticity. *arXiv preprint arXiv:2201.11063*. doi: 10.3389/fnins.2022.795876
- Pfister, J.-P., and Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neurosci.* 26, 9673–9682. doi: 10.1523/JNEUROSCI.1425-06.2006
- Rajendran, B., Sebastian, A., Schmuker, M., Srinivasa, N., and Eleftheriou, E. (2019). Low-power neuromorphic hardware for signal processing applications: a review of architectural and system-level design approaches. *IEEE Signal Process. Magaz.* 36, 97–110. doi: 10.1109/MSP.2019.2933719
- Ranganath, C., and Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat. Rev. Neurosci.* 4, 193–202. doi: 10.1038/nrn1052
- Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. (2019). “Regularized evolution for image classifier architecture search,” in *Proceedings of the Aaai Conference on Artificial Intelligence, Volume 33*, 4780–4789. doi: 10.1609/aaai.v33i01.33014780
- Risi, S., and Stanley, K. O. (2012). “A unified approach to evolving plasticity and neural geometry,” in *The 2012 International Joint Conference on Neural Networks (IJCNN)* (Brisbane, QLD: IEEE), 1–8. doi: 10.1109/IJCNN.2012.6252826
- Rothfuss, J., Lee, D., Clavera, I., Asfour, T., and Abbeel, P. (2018). Promp: proximal meta-policy search. *arXiv preprint arXiv:1810.06784*.
- Rubin, J., Lee, D., and Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. *Phys. Rev. Lett.* 86, 364–367. doi: 10.1103/PhysRevLett.86.364
- Salgado, H., Köhr, G., and Trevino, M. (2012). Noradrenergic “tone” determines dichotomous model of cortical spike-timing-dependent plasticity. *Sci. Rep.* 2, 1–7. doi: 10.1038/srep00417
- Scherr, F., Stöckl, C., and Maass, W. (2020). One-shot learning with spiking neural networks. *BioRxiv*. doi: 10.1101/2020.06.17.156513
- Schmidgall, S. (2020). “Adaptive reinforcement learning through evolving self-modifying neural networks,” in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*. doi: 10.1145/3377929.3389901
- Schmidgall, S., Ashkanazy, J., Lawson, W. E., and Hays, J. (2021). Spikepropamine: differentiable plasticity in spiking neural networks. *Front. Neurobot.* 15, 629210. doi: 10.3389/fnbot.2021.629210
- Schmidgall, S., and Hays, J. (2021). Stable lifelong learning: spiking neurons as a solution to instability in plastic neural networks. doi: 10.1145/3517343.3517345
- Schultz, W., Apicella, P., and Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J. Neurosci.* 13, 900–913. doi: 10.1523/JNEUROSCI.13-03-00900.1993
- Seamans, J. (2007). Dopamine anatomy. *Scholarpedia* 2, 3737. doi: 10.4249/scholarpedia.3737
- Senn, W., Markram, H., and Tsodyks, M. (2001). An algorithm for modifying neurotransmitter release probability based on pre-and postsynaptic spike timing. *Neural Comput.* 13, 35–67. doi: 10.1162/089976601300014628
- Seung, H. S. (2003). Learning in spiking neural networks by reinforcement of stochastic synaptic transmission. *Neuron* 40, 1063–1073. doi: 10.1016/S0896-6273(03)00761-X
- Shrestha, S. B., and Orchard, G. (2018). Slayer: Spike layer error reassignment in time. *Adv. Neural Inform. Process. Syst.* 31. doi: 10.48550/arXiv.1810.08646
- Sjöström, P. J., Turrigiano, G. G., and Nelson, S. B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* 32, 1149–1164. doi: 10.1016/S0896-6273(01)00542-6
- Soltoggio, A., Bullinaria, J. A., Mattiussi, C., Dürr, P., and Floreano, D. (2008). “Evolutionary advantages of neuromodulated plasticity in dynamic, reward-based scenarios,” in *Proceedings of the 11th International Conference on Artificial Life (Alife XI), number CONF* (MIT Press), 569–576.
- Song, S., Miller, K. D., and Abbott, L. F. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* 3, 919–926. doi: 10.1038/78829
- Song, S., Sjöström, P. J., Reigl, M., Nelson, S., and Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* 3, e350. doi: 10.1371/journal.pbio.0030350
- Speranza, L., di Porzio, U., Viggiano, D., de Donato, A., and Volpicelli, F. (2021). Dopamine: the neuromodulator of long-term synaptic plasticity, reward and movement control. *Cells* 10, 735. doi: 10.3390/cells10040735
- Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., and Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nat. Neurosci.* 16, 966–973. doi: 10.1038/nn.3413
- Stewart, K., Orchard, G., Shrestha, S. B., and Neftci, E. (2020). Online few-shot gesture learning on a neuromorphic processor. *IEEE J. Emerg. Selected Top. Circuits Syst.* 10, 512–521. doi: 10.1109/JETCAS.2020.3032058
- Teles-Grilo Ruivo, L., and Mellor, J. (2013). Cholinergic modulation of hippocampal network function. *Front. Synapt. Neurosci.* 5, 2. doi: 10.3389/fnsyn.2013.00002
- van Rossum, M., Bi, G., and Turrigiano, G. (2001). Stable hebbian learning from spike timing-dependent plasticity. *J. Neurosci.* 20, 8812–8821. doi: 10.1523/JNEUROSCI.20-23-08812.2000
- Velez, R., and Clune, J. (2017). Diffusion-based neuromodulation can eliminate catastrophic forgetting in simple neural networks. *PLoS ONE* 12, e0187736. doi: 10.1371/journal.pone.0187736
- Wang, H.-X., Gerkin, R. C., Nauen, D. W., and Bi, G.-Q. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nat. Neurosci.* 8, 187–193. doi: 10.1038/nn1387
- Wenk, G. L. (1998). Assessment of spatial memory using the t maze. *Curr. Protoc. Neurosci.* 4, 8–5. doi: 10.1002/0471142301.ns0805bs04
- Zannone, S., Brzosko, Z., Paulsen, O., and Clopath, C. (2018). Acetylcholine-modulated plasticity in reward-driven navigation: a computational study. *Sci. Rep.* 8, 1–20. doi: 10.1038/s41598-018-27393-2
- Zenke, F., and Vogels, T. P. (2021). The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural Comput.* 33, 899–925. doi: 10.1162/neco_a_01367
- Zhang, J.-C., Lau, P.-M., and Bi, G.-Q. (2009). Gain in sensitivity and loss in temporal contrast of stdp by dopaminergic modulation at hippocampal synapses. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13028–13033. doi: 10.1073/pnas.0900546106
- Zoph, B., and Le, Q. V. (2016). Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.