BindsNET: A machine learning-oriented spiking neural networks library in Python

Hananel Hazan, Daniel J. Saunders, Hassaan Khan, Darpan T. Sanghavi, Hava T. Siegelmann, and Robert Kozma

Biologically Inspired Neural and Dynamical Systems Laboratory, University of Massachusetts Amherst, College of Computer and Information Sciences, Amherst, MA, USA

December 12, 2018

Abstract

The development of spiking neural network simulation software is a critical component enabling the modeling of neural systems and the development of biologically inspired algorithms. Existing software frameworks support a wide range of neural functionality, software abstraction levels, and hardware devices, yet are typically not suitable for rapid prototyping or application to problems in the domain of machine learning. In this paper, we describe a new Python package for the simulation of spiking neural networks, specifically geared towards machine learning and reinforcement learning. Our software, called BindsNET, enables rapid building and simulation of spiking networks and features user-friendly, concise syntax. BindsNET is built on top of the PyTorch deep neural networks library, enabling fast CPU and GPU computation for large spiking networks. The BindsNET framework can be adjusted to meet the needs of other existing computing and hardware environments, e.g., TensorFlow. We also provide an interface into the OpenAI gym library, allowing for training and evaluation of spiking networks on reinforcement learning problems. We argue that this package facilitates the use of spiking networks for large-scale machine learning experimentation, and show some simple examples of how we envision BindsNET can be used in practice.

1 Introduction

The recent success of deep learning models in computer vision, natural language processing, and other domains [45] have led to a proliferation of machine learning software packages [33, 1, 9, 3, 25, 40]. GPU acceleration of deep learning primitives has been a major proponent of this success [10], as their massively parallel operation enables rapid processing of layers of
independent nodes. Since the biological plausibility of deep neural networks is often disputed [38], interest in integrating the algorithms of deep learning with long-studied ideas in neuroscience has been mounting [30], both as a means to increase machine learning performance and to better model learning and decision-making in biological brains [44].

Spiking neural networks (SNNs) [28, 29, 42] are sometimes referred to as the “third generation” of neural networks because of their potential to supersede deep learning methods in the fields of computational neuroscience [43] and biologically plausible machine learning (ML) [5]. SNNs are also thought to be more practical for data-processing tasks in which the data has a temporal component since the neurons which comprise SNNs naturally integrate their inputs over time. Moreover, their binary (spiking or no spiking) operation lends itself well to fast and energy efficient simulation on hardware devices.

Although spiking neural networks are not widely used as machine learning systems, recent work shows that they have the potential to be. SNNs are often trained with unsupervised learning rules to learn a useful representation of a dataset, which are then used as features for supervised learning methods [16, 15, 20, 35]. Trained deep neural networks may be converted to SNNs and implemented in hardware while maintaining good image recognition performance [15], demonstrating that SNNs can in principle compete with deep learning methods. In similar lines of work [27, 23, 32], the popular back-propagation algorithm has been applied to differentiable versions of SNNs to achieve competitive performance on standard image classification benchmarks, more evidence in support of the potential of spiking networks for ML problem solving.

The membrane potential (or voltage) of a spiking neuron is often described by ordinary differential equations. The membrane potential of the neuron is increased or decreased by presynaptic inputs, depending on their sign and strength. In the case of the leaky integrate-and-fire (LIF) model [42] and several other models, the neuron is always decaying to a rest potential $v_{rest}$. When a neuron reaches its threshold voltage $v_{thr}$, it emits a spike which travels to downstream neurons via synapses, its postsynaptic effect modulated by synaptic strengths, and its voltage is reset to some $v_{reset}$. Connections between neurons may also have their own dynamics, which can be modified by prescribed learning rules or external reward signals.

Several software packages for the discrete-time simulation of SNNs exist, with varying levels of biological realism and support for hardware platforms. Many such solutions, however, were not developed to target ML applications, and often feature abstruse syntax resulting in steep learning curves for new users. Moreover, packages with a large degree of biological realism may not be appropriate for problems in ML, since they are computationally expensive to simulate and may require a large degree of hyper-parameter tuning. Real-time hardware implementations of SNNs exist as well, but cannot support the rapid prototyping that some software solutions can.

Motivated by the foregoing shortcomings, we present the BindsNET spiking neural networks library, which is developed on top of the popular PyTorch deep learning library [33]. At its core, the software allows users to build, train, and evaluate SNNs composed of groups of neurons and their connections. The learning of connection weights is supported by various algorithms from the biological learning literature [21, 31]. A separate module provides an interface to the OpenAI gym [7] reinforcement learning (RL) environments library from BindsNET. A Pipeline
object is used to streamline the interaction between spiking networks and RL environments, removing many of the messy details from the purview of the experimenter. Still other modules provide functions such as loading of ML datasets, encoding of raw data into spike train network inputs, plotting of network state variables and outputs, and evaluation of SNN as ML models.

The paper is structured as follows: we begin in Section 2 with an assessment of the existing SNN simulation software and hardware implementations. In Section 3, the BindsNET library is described at a high level of detail, emphasizing the functional role of each software module, how they were motivated, and how they inter-operate. Code snippets and simple case studies are presented in Section 4 to demonstrate the breadth of possible BindsNET applications. Desirable future features are listed in 5 and potential research impacts are assessed in Section 6.

2 Review

In the last two decades, neural networks have become increasingly prominent in machine learning and artificial intelligence research, leading to a proliferation of efficient software packages for their training, evaluation, and deployment. On the other hand, the simulation of the “third generation” of neural networks has not been able to reach its full potential, due in part to their inherent complexity and computational requirements. However, spiking neurons excel at remembering a short-term history of their activation and feature efficient binary communication with other neurons, a useful feature in reducing energy requirements on neuromorphic hardware. Spiking neurons exhibit more properties from their biological counterpart than the computing units utilized by deep neural networks, which may constitute an important advantage in terms of practical computational power or ML performance.

Researchers that want to conduct experiments with networks of spiking neurons for ML purposes have a number of options for SNN simulation software. Many frameworks exist, but each is tailored toward specific application domains. In this section, we describe the existing relevant software libraries and the challenges associated with each, and contrast these with the strengths of our package.

We believe that the chosen simulation framework must be easy to develop in, debug, and run, and, most importantly, support the level of biological complexity desired by its users. We express a preference to maintain consistency in development by using a single programming language, and for it to be affordable or an open source project. We describe whether and how these aspects are realized in each competing solution.

The reviewed simulations are divided into two main categories: (1) software platforms targeting simulation on CPUs, GPUs, or both; (2) hardware implementations of spiking neural networks, as described below.

2.1 Software Simulation

Many spiking neural network frameworks exist, each with a unique set of use cases. Some focus on the biologically realistic simulation of neurons, while others on high-level spiking network functionality. To build a network to run even the simplest machine learning experiment, one will face multiple difficult design choices: Which biological properties should the neurons and
the network have? E.g., how many GABAergic neurons or NMDA/AMPA receptors to use, or what form of synapse dynamics? Many such options exist, some of which may or may not have a significant impact on the performance of an ML system.

For example, NEURON [8], Genesis [11], NEST [19], BRIAN [37], and ANNarchy [41] focus on accurate biological realistic simulation from sub-cellular components and biochemical reactions to complex models of single neurons up to the network level. These simulation platforms target the neuro-biophysics community and neuroscientists that wish to simulate multicompartment neurons, where each sub-compartment is a different part of the neuron with different functionalities, morphological details, and shape. These packages are able to simulate large SNNs on various types of systems, from laptops all the way up to HPC systems. However, each simulated component must be homogeneous, meaning that it must be built with a single type of neuron and a single type of synapse. If a researcher wants to simulate multiple types of neurons utilizing various synapse types, it may be difficult in these frameworks. For a more detailed comparison of development time, model performance, and varieties of models of neurons available in this biophysical software see [39].

A major benefit of the BRIAN, ANNarchy, and NEST packages is that, besides the built-in modules for neuron and connection objects, the programmer is able to specify the dynamics of neurons and connections using differential equations. This eliminates the need to manually specify the dynamic properties of each new neuron or connection object in code. The equations are compiled into fast C++ code in case of ANNarchy, vectorised and linear algebraic operations using NumPy and Basic Linear Algebra Subprograms (BLAS) in case of BRIAN and to a mix of Python and native C-like language (hoc) [22] which are responsible for the SNN simulation. In addition, in the NEST package, the programmer can combine pre-configured objects (which accepts arguments) to create SNNs. In all of these libraries, significant changes to the operation of the network components requires modification of the underlying code, a difficult task which stands in the way of fast network prototyping and break the continuity of the programming. At this time BindsNET does not support the solving of arbitrary differential equations describing neural dynamics, but rather, for simplicity, several popular neuron types are supported.

Frameworks such as NeuCube [26] and Nengo [4] focus on high-level behaviors of neural networks and may be used for machine learning experimentation. NeuCube supports rate coding-based spiking networks, and Nengo supports simulation at the level of spikes, firing rates, or high-level, abstract neural behavior. NeuCube attempts to map spatiotemporal input data into three-dimensional SNN architectures; however, it is not an open source project, and therefore is somewhat restricted in scope and usability. Nengo is often used to simulate high-level functionality of brains or brain regions, as a cognitive modeling toolbox implementing the Neural Engineering Framework [36] rather than a machine learning framework. Nengo is an open source project, written in Python, and supports a Tensorflow [1] backend to improve simulation speed and exploit some limited ML functionality. It also has options for deploying neural models on dedicated hardware platforms; e.g., SpiNNaker [34].

Other frameworks, e.g., CARLsim [6] and NeMo [17], focus on the high-level aspects of spiking neural networks, and are more suited to machine learning applications. Both packages allow simulation of large spiking networks built from Izhikevich neurons [24] with realistic synaptic
dynamics as their fundamental building block, and accelerate computation using GPU hardware. Like the frameworks before, low-level simulation code is written in C++ for efficiency, and programmers may interface with them with the simulator-independent PyNN Python library [13], or in MATLAB or Java.

The above packages are typically written in more than one programming language: the core functionality is implemented in a lower-level language (e.g., C++) to achieve good performance, and the code exposed to the user of the package is written in a higher-level language (e.g., Python or MATLAB) to enable fast prototyping. If such frameworks are not tailored to the needs of a user, lead to steep learning curves, and aren’t flexible enough to create a desired model, the user may have to program in both high- and low-level languages to make changes to the required internal components.

2.2 Neuromorphic Hardware Simulation

Spiking neural network frameworks based on hardware have several advantages over software simulation in terms of performance and power consumption. Different hardware solutions have different use cases. For example, SpiNNaker [34] combines cheap, generic, yet dedicated CPU boards together to create a powerful SNN simulation framework in hardware. Other platforms (e.g., TrueNorth [2], HRL, and Braindrop) involve the design of a new chip, used to emulate spiking neurons by making use of the physical properties of certain materials. A novel development is Intel’s Loihi platform for spike-based computation, outperforming all known conventional solutions [12]. Other solutions are based on programmable hardware, like FPGAs which transform neural equations to configurations of electronic gates in order to speed up computation. More specialized hardware such as ASIC and DSP can be used to parallelize and therefore accelerate the calculations required for the operation and interaction of neurons in a spiking network.

In order to conduct experiments in the hardware domain, one must usually learn a specific programming language targeted to the hardware platform, or carefully adapt an existing experiment to the unique hardware environment under the constraints as enforced by chip designers; e.g., the connectivity limitations of the TrueNorth chip, or the limited configuration options of the SpiNNaker platform. In either case, this is not an ideal situation for researchers who want to rapidly prototype and test the capabilities of SNNs in simulation, machine learning, or other domains.

3 Package structure

A summary of all the software modules of the BindsNET package is included in Figure 1.

Many BindsNET objects use the torch.Tensor data structure for computation; e.g., all objects supporting the Nodes interface use Tensors to store and update state variables such as spike occurrences or voltages. The Tensor object is a multi-dimensional matrix containing elements of a single data type; e.g., integers or floating points numbers with 8, 16, 32, or 64 bits of precision. They can be easily moved between devices with calls to
3.1 SNN simulation

BindsNET provides a Network object (in the network module) which is responsible for the coordination of one or many Nodes and Connections objects, and supports the use of Monitors for recording the state variables of these components. A time step parameter \( dt \) is the sole (optional) argument to the Network constructor, to which all other objects' time constants are scaled. The \( \text{run}(\text{inpts}, \text{time}) \) function implements synchronous updates (for a number of time steps \( \frac{\text{time}}{\text{dt}} \)) of all network components. This function calls \( \text{get_inputs}() \) to calculate presynaptic inputs to all Nodes instances (alongside user-defined inputs in \( \text{inpts} \)) as a subroutine. A \( \text{reset}() \)
method invokes resetting functionality of all network components, namely for resetting state variables back to default values. Saving and loading of networks to and from disk is implemented, permitting re-use of trained connection weights or other parameters.

The **Nodes** abstract base class in the **nodes** module specifies the abstract functions **step**(inpts, dt) and **_reset()**. The first is called by the **run()** function of a **Network** instance to carry out a single time step’s update, and the second resets spikes, voltages, and any other recorded state variables to default values. Implementations of the **Nodes** class include **Input** (neurons with user-specified or fixed spikes) **McCullochPittsNodes** (McCulloch-Pitts neurons), **IFNodes** (integrate-and-fire neurons), **LIFNodes** (leaky integrate-and-fire neurons), and **IzhikevichNodes** (Izhikevich neurons). Other neurons or neuron-like computing elements may be implemented by supporting the **Nodes** interface. Many **Nodes** object support optional arguments for customizing neural attributes such as threshold, reset, and resting potential, refractory period, membrane time constant, and more.

The **topology** module is used to specify interactions between **Nodes** instances, the most generic of which is implemented in the **Connection** object. The **Connection** is aware of source (presynaptic) and target (postsynaptic) **Nodes**, as well as a matrix of weights \( w \) of connections strengths. By default, connections do not implement any learning of connection weights, but do so through the inclusion of an **update_rule** argument. Several canonical learning rules from the biological learning literature are implemented in the **learning** module, including Hebbian learning (**hebbian()**), a variant of spike-timing-dependent plasticity (STDP) (**post.pre()**), and less well-known methods such as reward-modulated STDP (**mstdp()**). The optional argument **norm** to the **Connection** specified a desired sum of weights per target neuron, which is enforced by the parent **Network** during each call of **run()**. A **SparseConnection** object is available for specifying connections where certain weights are fixed to zero; however, this does not (yet) support learning functions. The **Conv2dConnection** object implements a two-dimensional convolution operation (using PyTorch’s **torch.nn.conv2d** function) and supports all update rules from the **learning** module.

### 3.2 Machine and reinforcement learning

**BindsNET** is being developed with machine and reinforcement learning applications in mind. At the core of these efforts is the **learning** module, which contains functions which can be attached to **Connection** objects to modify them during SNN simulation. By default, connections are instantiated with no learning rule. The **hebbian** rule (‘fire together, wire together’) symmetrically strengthens weights when pre- and postsynaptic spikes occur temporally close together, and the **post.pre** rule implements a simple form of STDP in which weights are increased or decreased according to the relative timing of pre- and postsynaptic spikes, with user-specified (possibly asymmetric) learning rates. The reward-modulated STDP (**mstdp**) and reward-modulated STDP with eligibility trace (**mstdp.**et) rules of Florian [18] are also implemented for use in basic reinforcement learning experiments.

The **datasets** module provides a means to download, pre-process, and iterate over machine learning datasets. For example, the **MNIST** object provides this functionality for the MNIST handwritten digits dataset. Several other datasets are supported besides, including CIFAR-10,
3.3 The Pipeline object

CIFAR-100, and Spoken MNIST. The samples from a dataset can be encoded into spike trains using the `encoding` module, currently supporting several functions for creating spike trains from non-negative data based on different statistical distributions and biologically inspired transformations of stimuli. Spikes may be used as input to SNNs, or even to other ML systems. A submodule `preprocess` of `datasets` allows the user to apply various pre-processing techniques to raw data; e.g., cropping, subsampling, binarizing, and more.

The `environment` module provides an interface into which a SNN, considered as a reinforcement learning agent, can take input from and enact actions in an reinforcement learning environment. The `GymEnvironments` object comprises of a generic wrapper for gym [7] RL environments and calls its `reset()`, `step(action)`, `close()`, and `render()` functionality, while providing a default pre-processing function `preprocess()` for observations from each environment. The `step(action)` function takes an action in the gym environment, which returns an observation, reward value, an indication of whether the episode has finished, and a dictionary of (name, value) pairs containing additional information. Another object, `DatasetEnvironment`, provides a generic wrapper around objects from the `datasets` module, allowing these to be used as a component in a `Pipeline` instance (see Section 3.3). The `environment.action` module provides methods for mapping one or more network layers’ spikes to actions in the environment; e.g., `select_multinomial()` treats a (normalized) vector of spikes as a probability distribution from which to sample an action for the environment’s similarly-sized action space.

Simple methods for the evaluation of SNNs as machine learning models are implemented in the `evaluation` module. In the context of unsupervised learning, the `assign_labels()` function assigns data labels to neurons corresponding to the class of data on which they spike most during network training [14]. These labels are to classify new data using methods like `all_activity()` and `proportion_weighting()` [20].

A collection of network architectures is defined in the `models` module. For example, the network structure of Diehl and Cook [14] is implemented by the `DiehlAndCook2015` object, which supports arguments such as `n_neurons`, `excite`, `inhib`, etc. with reasonable default values.

3.3 The Pipeline object

As an additional effort to ease prototyping of machine learning systems comprising spiking neural networks, we have provided the `Pipeline` object to compose an environment, network, an encoding of environment observations, and a mapping from network activity to the environment’s action space. The `Pipeline` also provides optional arguments for visualization of the environment and network state variables during network operation, skipping or recording observations on a regular basis, the length of the simulation per observation (defaults to 1 time step), and more. The main action of the pipeline can be explained as a four-step, recurring process, implemented in the pipeline `step()` function:

1. An action is selected based on the activity of one or more of the network’s layers during the last one or more time steps
2. This action is used as input to the environment’s `step()` function, which returns a new
observation, a scalar reward, whether the simulation has finished, and any additional information particular to the environment.

3. The observation returned from the environment is converted into spike trains according to the user-specified encoding function (either custom or from the encoding module) and request simulation time.

4. The spike train-encoded observation is used as input to the network.

Alongside the required arguments for the Pipeline object (network, environment, encoding, and action), there are a few keyword arguments that are supported, such as history and delta. The history_length argument indicates that a number of sequential observations are to be maintained in order to calculate differences between current observations and those stored in the history data structure. This implies that only new information in the environment’s observation space is delivered as input to the network on each time step. The delta argument (default 1) specifies an interval at which observations are stored in history. This may be useful if observations don’t change much between consecutive steps; then, we should wait some delta time steps between taking observations to expect significant differences. As an example, combining history_length = 4 and delta = 3 will store observations \{0, 3, 6, 9\}, \{3, 6, 9, 12\}, \{6, 9, 12, 15\}, etc. A few other keyword arguments for handling console output, plotting, and more exist and are detailed in the Pipeline object documentation.

A functional diagram of the Pipeline object is depicted in Figure 2.

![Pipeline Diagram](image)

Figure 2: A functional diagram of the Pipeline object. The four-step process involves an encoding function, network computation, converting network outputs into actions in an environment’s action space, and a simulation step of the environment. An encoding function converts non-spiking observations from the environment into spike inputs to the network, and a action function maps network spiking activity into a non-spiking quantity: an action, fed back into the environment, where the procedure begins anew. Other modules come into play in various supporting roles: the network may use a learning method to update connection weights, or the environment may simply be a thin wrapper around a dataset (in which case there is no feedback), and it may be desirable to plot network state variables during the reinforcement learning loop.

3.4 Visualization

BindsNET also contains useful visualization tools that provide information during network or environment simulation. Several generic plotting functions are implemented in the analysis.plotting module; e.g., plot_spikes() and plot_voltages() create and update plots dynamically instead of recreating figures at every time step. These functions are able to display spikes and voltages with a single call. Other functions include plot_weights() (displays connection weights),
plot_input() (displays raw input data), and plot_performance() (displays time series of performance metric).

The analysis.visualization module contains additional plotting functionality for network state variables after simulation has finished. These tools should allow experimenters to analyzed learned weights or spike outputs, or to summarize long-term behaviors of their SNN models. For example, the weights_movie() function creates an animation of a Connection’s weight matrix from a sequence of its values, enabling the visualization of the trajectory of connection weight updates.

4 Examples

We present some simple example scripts to give an impression of how BindsNET can be used to build spiking neural networks implementing machine learning functionality. BindsNET is built with the concept of encapsulation of functionality to make it faster and easier for generalization and prototyping. Note in the examples below the compactness of the scripts: fewer lines of code are needed to create a model, load a dataset, specify their interaction in terms of a pipeline, and run a training loop. Of course, these commands rely on many lines of underlying code, but the user no longer has to implement them for each experimental script. If changes in the available parameters are not enough, the experimenter can intervene by making changes in the underlying code in the model without changing language or environment, thus preserving the continuity of the coding environment.

4.1 Unsupervised learning

The DiehlAndCook2015 object in the models module implements a slightly simplified version of the network architecture discussed in Diehl and Cook [14]. A minimal working example of training a spiking neural network to learn, without labels, a representation of the MNIST digit dataset is given in Figure 4, and state variable-monitoring plots are depicted in Figure 3. The Pipeline object is used to hide the messy details of the coordination between the dataset, encoding function, and network instance. Code for additional plots or console output may be added to the training loop for monitoring purposes as needed.
from bindsnet.datasets import MNIST
from bindsnet.encoding import poisson
from bindsnet.pipeline import Pipeline
from bindsnet.models import DiehlAndCook2015
from bindsnet.environment import DatasetEnvironment

# Build Diehl & Cook 2015 network.
network = DiehlAndCook2015(n_inpt=784,
n_neurons=400,
exc=22.5,
inh=17.5,
dt=1.0,
norm=78.4)

# Specify dataset wrapper environment.
environment = DatasetEnvironment(dataset=MNIST(path='..../data/MNIST'),
train=True,
intensity=0.25)

# Build pipeline from components.
pipeline = Pipeline(network=network,
environment=environment,
encoding=poisson,
time=350,
plot_interval=1)

# Train the network.
for i in range(60000):
    pipeline.step()
    network.reset()

Figure 3: Unsupervised learning of the MNIST handwritten digits in BindsNET. The DiehlAndCook2015 model implements a simple spike timing-dependent plasticity rule between input and excitatory neuron populations as well as a competitive inhibition mechanism to learn prototypical digit filters from raw data. The DatasetEnvironment wraps the MNIST dataset object so it may be used as a component in the Pipeline. The network is trained on one pass through the 60K-example training data for 350ms each, with state variables (voltages and spikes) reset after each example.
4.2 Supervised learning

We can use a modified version of the DiehlAndCook2015 network to implement supervised learning of the CIFAR-10 natural images dataset. An minimal example of training a spiking
network to classify the CIFAR-10 dataset is given in Figure 5 with plotting outputs depicted in Figure 6. A layer of 100 excitatory neurons is split into 10 groups of size 10, one for each digit category. On each input example, we observe the label of the data and clamp a randomly selected excitatory neuron from its group to spike on every time step. This forces the neuron to adjust its filter weights towards the shape of current input example.

```python
from torch import Tensor
from numpy.random import choice
from bindsnet.datasets import CIFAR10
from bindsnet.encoding import poisson
from bindsnet.pipeline import Pipeline
from bindsnet.models import DiehlAndCook2015
from bindsnet.environment import DatasetEnvironment

# Build network.
network = DiehlAndCook2015(n_inpt=32*32*3,
n_neurons=100,
dt=1.0,
exc=22.5,
inh=17.5,
nu_pre=0,
nu_post=1e-2,
norm=78.4)

# Specify dataset wrapper environment.
environment = DatasetEnvironment(dataset=CIFAR10(path='../../../data/CIFAR10'),
                                train=True)

# Build pipeline from components.
pipeline = Pipeline(network=network,
environment=environment,
encoding=poisson,
time=50,
plot_interval=1)

# Train the network.
labels = environment.labels
for i in range(60000):
    # Choose an output neuron to clamp to spiking behavior.
c = choice(10, size=1, replace=False)
    clamp = {'Ae': 10 + labels[i].long() + Tensor(c).long()}

    # Run a step of the pipeline with clamped neuron.
pipeline.step(clamp=clamp)
network._reset()
```

Figure 5: The DiehlAndCook2015 model is instantiated with 100 neurons, and is otherwise identical to the model in the unsupervised learning example. The group of neurons named “Ae” refers to the excitatory layer of neurons in the Diehl and Cook [14] model. A random neuron from the subgroup corresponding to the data label is selected, and clamped to spike on each time step (using pipeline.step(clamp=clamp)).
4.3 Reinforcement learning

For a more complete view of the details involved in constructing an SNN and deploying a GymEnvironment instance, see the script depicted in Figure 7 and accompanying displays in Figure 6: Accompanying plots for supervised training of a DiehlAndCook2015 network with 400 excitatory and inhibitory neurons, trained by clamping random neurons from a particular subpopulation to spike when a particular input category is presented. The first twenty neurons represent the image category “plane”, the second twenty neurons represent the image category “car”, and so on. STDP is used to update the synapse weights.

Figure 6: Accompanying plots for supervised training of a DiehlAndCook2015 network with 400 excitatory and inhibitory neurons, trained by clamping random neurons from a particular subpopulation to spike when a particular input category is presented. The first twenty neurons represent the image category “plane”, the second twenty neurons represent the image category “car”, and so on. STDP is used to update the synapse weights.

4.3 Reinforcement learning

For a more complete view of the details involved in constructing an SNN and deploying a GymEnvironment instance, see the script depicted in Figure 7 and accompanying displays in Figure 6: Accompanying plots for supervised training of a DiehlAndCook2015 network with 400 excitatory and inhibitory neurons, trained by clamping random neurons from a particular subpopulation to spike when a particular input category is presented. The first twenty neurons represent the image category “plane”, the second twenty neurons represent the image category “car”, and so on. STDP is used to update the synapse weights.

4.3 Reinforcement learning

For a more complete view of the details involved in constructing an SNN and deploying a GymEnvironment instance, see the script depicted in Figure 7 and accompanying displays in
Figure 8. A three layer SNN is built to compute on spikes encoded from Space Invaders observations. The result of this computation is some spiking activity in the output layer, which are converted into actions in the game’s action space. The simulation of both the network and the environment are interleaved and appear to operate in parallel, and the network is able to adjust its parameters *online* using the reward-modulated STDP update rule $m_{stdp, st}$ taken from Florian [18].
4.3 Reinforcement learning

Figure 7: A spiking neural network that accept input from the SpaceInvaders-v0 gym Atari environment. The observations from the environment are downsampled and binarized before they are converted into Bernoulli-distributed vectors of spikes, one per time step. The history and delta keyword arguments are used to create a reasonable level of input spiking activity. The output layer of the network has 60 neurons in it, each subsequent 10 representing a different action in the Space Invaders game as a population code. An action is selected at each time step using the select_multinomial feedback function, which treats the averaged (normalized) activity over each output layer subpopulation as a probability distribution over actions.
4.4 Reservoir Computing

Reservoir computers are typically built from three parts: (1) an encoder that translates input from the environment that is fed to it, (2) a dynamical system based on randomly connected neurons (the reservoir), and (3) a readout mechanism. The readout is often trained via gradient descent to perform classification or regression on some target function. BindsNET can be used to build reservoir computers using spiking neurons with little effort, and machine learning functionality from PyTorch can be co-opted to learn a function from states of the high-dimensional reservoir to desired outputs. Code in for defining and simulating a simple reservoir computer is given in Figure 9 and plots to monitor simulation progress are shown in Figure 10. The outputs of the reservoir computer on the CIFAR-10 natural image dataset are used as transformed inputs to a logistic regression model. The logistic regression model is then trained to recognize the digits based on the features produced by the reservoir.

Figure 8: Accompanying plots for a custom spiking neural network’s which interacts with the SpaceInvaders-v0 reinforcement learning environment. Spikes and voltages of all neuron populations are plotted, and the Space Invaders game is rendered, as well as the downsampled, history- and delta-altered observation, which is presented to the network.
import torch
import torch.nn as nn
from bindnet import *

# Define logistic regression model using PyTorch.
class LogisticRegression(nn.Module):
    def __init__(self, input_size, num_classes):
        super(LogisticRegression, self).__init__()
        self.linear = nn.Linear(input_size, num_classes)
    def forward(self, x):
        return self.linear(x)

# Build a simple, two layer, "input-output" network.
network = Network(dt=1.0)
inpt = Input(3, 32, 32)
output = LIFNodes(625, threshold=52 + torch.randn(625))
network.add_layer(inpt, name='inpt')
num_classes = 10
network.add_layer(output, name='output')
for i in range(2):
    network.add_connection(n1, n2, weight=torch.randn(n1.n, n2.n), synapse=1.0)

# Create and train logistic regression model on reservoir outputs.
weights = torch.randn(625, num_classes, requires_grad=True)
model = LogisticRegression(625, num_classes)
model_weights = torch.nn.Parameter(weights)
optimizer = torch.optim.SGD(model.parameters(), lr=0.1)
for epoch in range(10):
    for i, (data, label) in enumerate(train_loader):
        network.run(inpts={'inpt': data}, time=250)
        train_pairs.append(network.monitors['output_spikes'].get('s').sum(-1))
        network.reset()
        if (i + 1) % 50 == 0:
            print('Train progress: %d / 500' % (i + 1))
        loss = criterion(output, label)
        optimizer.zero_grad(); output = model(s)
        loss.backward(); optimizer.step()

# Get MNIST test images and labels and create data loader.
images, labels = CIFAR10(path='./../data/CIFAR10').test() images *= 0.25
loader = zip(images, labels)

# Run test data on reservoir and store (spikes per neuron, label) pairs.
for i, (data, label) in enumerate(test_loader):
    network.run(inpts={'inpt': data}, time=250)
    test_pairs.append(network.monitors['output_spikes'].get('s').sum(-1))
    network.reset()
    if (i + 1) % 50 == 0:
        print('Test progress: %d / 500' % (i + 1))
    if (i + 1) == 500:
        break
    loss.backward(); optimizer.step()

# Test the logistic regression model on (spikes, label) pairs.
correct, total = 0, 0
for s, label in test_pairs:
    output = model(s); predicted = torch.max(output, 1)[1]
    total += 1; correct += (predicted == label).sum()

print('Accuracy of logistic regression on 500 test examples: %2f %%%' % (100 * correct / total))

Figure 9: A recurrent neural network built from 625 spiking neurons accepts inputs from the CIFAR-10 natural images dataset. An input population is connected all-to-all to an output population of LIF neurons with weights drawn from the standard normal distribution, which has voltage thresholds drawn from $\mathcal{N}(52, 1)$ and is recurrently connected to itself with weights drawn from $\mathcal{N}(0, \frac{1}{2})$. The reservoir is used to create a high-dimensional, temporal representation of the image data, which is used to train and test a logistic regression model created with PyTorch.
5 ONGOING DEVELOPMENT

BindsNET is still at an early stage of development, and there is much room for future work and improvement. Since it is an open source project and because there is considerable interest in the research community in using SNNs for machine learning purposes, we are optimistic that there will be numerous community contributions to the library. Indeed, we believe that public interest in the project, along with the strong support of the libraries on which it depends, will be an important driving factor in its maturation and proliferation of features. We mention some specific implementation goals:

- Additional neuron types, learning rules, datasets, encoding functions, etc. Added features should take priority based on the needs of the users of the library.

- Specialization of machine learning and reinforcement learning algorithms for spiking neural networks. These may take the form of additional learning rules, or some more complicated training methods that operate at the network level rather than on individual synapses.

- Tighter integration with PyTorch. Much of PyTorch’s neural network functionality may have interpretations in the spiking neural network context (e.g., as with Conv2dConnection).
• Conversion from deep neural network models implemented in PyTorch or specified in the ONNX format to near-equivalent spiking neural networks (as in Diehl et al. [15]).

• Performance optimization: improving the performance of library primitives will save time on all experiments with spiking neural networks.

• Automatic smoothing of SNNs: approximating spiking neurons as differentiable operations will permit the application of backpropagation to BindsNET SNNs. The torch.autograd automatic differentiation library [33] can then easily be applied to optimize the parameters spiking networks for ML problems.

6 Discussion

We have presented the BindsNET open source package for rapid biologically inspired prototyping of spiking neural networks with a machine learning-oriented approach. BindsNET is developed entirely in Python and is built on top of other mature Python libraries that lend their power to utilize multi-CPU or multi-GPU hardware configurations. Specifically, the ML tools and powerful data structures of PyTorch are a central part of BindsNET’s operation. BindsNET may also interface with the gym library to connect spiking neural networks to reinforcement learning environments. In sum, BindsNET represents an additional and attractive alternative for the research community for the purpose of developing faster and more flexible tools for SNN experimentation.

BindsNET comprises a spiking neural network simulation framework that is easy to use, flexible, and efficient. Our library is set apart from other solutions by its ML and RL focus; complex details of the biological neuron are eschewed in favor of high-level functionality. Computationally inclined researchers may be familiar with the underlying PyTorch functions and syntax, and excited by the potential of the third generation of neural networks for ML problems, driving adoption in both ML and computational neuroscience communities. This combination of ML programming tools and neuroscientific ideas may facilitate the further integration of biological neural networks and machine learning. To date, spiking neural networks have not been widely applied in ML and RL problems; having a library aimed at such is a promising step towards exciting new lines of research.

Researchers interested in developing spiking neural networks for use in ML or RL applications will find that BindsNET is a powerful and easy tool to develop their ideas. To that end, the biological complexity of neural components has been kept to a minimum, and high-level, qualitative functionality has been emphasized. However, the experimenter still has access to and control over groups of neurons at the level of membrane potentials and spikes, and connections at the level of synapse strengths, constituting a relatively low level of abstraction. Even with such details included, it is straightforward to build large and flexible network structures and apply them to real data. We believe that the ease with which our framework allows researchers to reason about spiking neural networks as ML models, or as RL agents, will enable advancements in biologically plausible machine learning, or further fusion of ML with neuroscientific concepts.

Although BindsNET is similar in spirit to the Nengo [4] neural and brain modeling software
in that both packages can utilize a deep learning library as a “backend” for computation, Nengo optionally uses TensorFlow in a limited fashion while BindsNET uses PyTorch by default, for all network simulation functionality (with the torch.Tensor object). Additionally, for users that prefer the flexibility and the imperative execution of PyTorch, BindsNET inherits these features and is developed with many of the same design principles in mind.

We see BindsNET as a simple yet attractive option for those looking to quickly build flexible SNN prototypes backed by an easy-to-use yet powerful deep learning library. It encourages the conception of spiking networks as machine learning models or reinforcement learning agents, and is perhaps the first of its kind to provide a seamless interface to RL environments. The library is supported by several mature and feature-full open source software projects, and experiences immediate benefits from their growth and efficiency improvements. Considered as an extension of the PyTorch library, BindsNET represents a natural progression from the second to third generation of neural network research and development.

Acknowledgements

This work has been supported in part by Defense Advanced Research Project Agency Grant, DARPA/MTO HR0011-16-I-0006 and by National Science Foundation Grant NSF-CRCNS-DMS-13-11165.

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[2] F. Akopyan, J. Sawada, A. S. Cassidy, R. Alvarez-Icaza, J. V. Arthur, P. Merolla, N. Imam, Y. Y. Nakamura, P. Datta, G.-J. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. L. Jackson, and D. S. Modha. Truenorth: Design and tool flow of a 65 mw 1 million neuron programmable neurosynaptic chip. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 34:1537–1557, 2015.

[3] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopesky, Y. Bengio, A. Bergeron, J. Bergstra, V. Bisson, J. Blecheer Snyder, N. Bouchard, N. Boulanger-Lewandowski, X. Bouthillier, A. de Brébisson, O. Breuleux, P.-L. Carrier, K. Cho, J. Chorowski, P. Christiano, T. Coolmans, M.-A. Côté, M. Côté, A. Courville, Y. N. Dauphin, O. Delalleau, J. Demouth, G. Desjardins, S. Dieleman, L. Dinh, M. Ducoffe, V. Dumoulin, S. Ebrahimi Kahou, D. Erhan, Z. Fan, O. Firat, M. Germain, X. Glorot, I. Goodfellow, M. Graham, C. Gulcehre, P. Hamel, I. Harlouchet,
J.-P. Heng, B. Hidasi, S. Honari, A. Jain, S. Jean, K. Jia, M. Korobov, V. Kulkarni, A. Lamb, P. Lamblin, E. Larsen, C. Laurent, S. Lee, S. Lefrancos, S. Lemieux, N. Léonard, Z. Lin, J. A. Livezey, C. Lorenz, J. Lowin, Q. Ma, P.-A. Manzagol, O. Mastrogiacomo, R. T. McGibbon, R. Memisevic, B. van Merriënboer, V. Michalski, M. Mirza, A. Orlandi, C. Pal, R. Pascanu, M. Pezeshki, C. Raffel, D. Renshaw, M. Rocklin, A. Romero, M. Roth, P. Sadowski, J. Salvatier, F. Savard, J. Schlüter, J. Schulman, G. Schwartz, I. V. Serban, D. Serdyuk, S. Shabanian, E. Simon, S. Spieckermann, S. R. Subramanyam, J. Szegedy, J. Tanguay, G. van Tulder, J. Turian, S. Urban, P. Vincent, F. Visin, H. de Vries, D. Ward-Farley, D. J. Webb, M. Willson, K. Xu, L. Xue, L. Yao, S. Zhang, and Y. Zhang. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016. URL http://arxiv.org/abs/1605.02688.

[4] T. Bekolay, J. Bergstra, E. Hunsberger, T. DeWolf, T. C. Stewart, D. Rasmussen, X. Choo, A. R. Voelker, and C. Eliasmith. Nengo: A python tool for building large-scale functional brain models. *Frontiers in Neuroinformatics*, 7(48), 2014. ISSN 1662-5196. doi: 10.3389/fninf.2013.00048. URL http://www.frontiersin.org/neuroinformatics/10.3389/fninf.2013.00048/abstract.

[5] Y. Bengio, D. Lee, J. Bornschein, and Z. Lin. Towards biologically plausible deep learning. *CoRR*, abs/1502.04156, 2015. URL http://arxiv.org/abs/1502.04156.

[6] M. Beyeler, K. D. Carlson, T.-S. Chou, N. D. Dutt, and J. L. Krichmar. Carlsim 3: A user-friendly and highly optimized library for the creation of neurobiologically detailed spiking neural networks. *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2015.

[7] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. *CoRR*, abs/1606.01540, 2016. URL http://arxiv.org/abs/1606.01540.

[8] N. T. Carnevale and M. L. Hines. *The NEURON Book*. Cambridge: University Press, 2006.

[9] T. Chen, M. Li, Y. Li, M. Lin, N. Wang, M. Wang, T. Xiao, B. Xu, C. Zhang, and Z. Zhang. Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems. *CoRR*, abs/1512.01274, 2015.

[10] S. Chetlur, C. Woolley, P. Vandermersch, J. Cohen, J. Tran, B. Catanzaro, and E. Shelhamer. cudnn: Efficient primitives for deep learning. *CoRR*, abs/1410.0759, 2014.

[11] H. Cornelis, A. L. Rodriguez, A. D. Coop, and J. M. Bower. Python as a federation tool for genesis 3.0. *PLOS ONE*, 7(1):1–11, 01 2012. doi: 10.1371/journal.pone.0029018. URL https://doi.org/10.1371/journal.pone.0029018.

[12] M. Davies, N. Srinivasa, T.-H. Lin, G. Chinya, Y. Cao, S. H. Choday, G. Dimou, P. Joshi, N. Imam, S. Jain, et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99, 2018.
[13] A. P. Davison, D. Brüderle, J. M. Eppler, J. Kremkow, E. Müller, D. Pecevski, L. U. Perrinet, and P. Yger. Pynn: A common interface for neuronal network simulators. *Frontiers in Neuroinformatics*, 2:3637 – 3642, 2008.

[14] P. U. Diehl and M. Cook. Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers in Computational Neuroscience*, Aug. 2015.

[15] P. U. Diehl, D. Neil, J. Binas, M. Cook, S. C. Liu, and M. Pfeiffer. Fast-classifying, high-accuracy spiking deep networks through weight and threshold balancing. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, July 2015. doi: 10.1109/IJCNN.2015.7280696.

[16] P. Ferr, F. Mamalet, and S. J. Thorpe. Unsupervised feature learning with winner-takes-all based stdp. *Frontiers in Computational Neuroscience*, 12:24, 2018. ISSN 1662-5188. doi: 10.3389/fncom.2018.00024. URL https://www.frontiersin.org/article/10.3389/fncom.2018.00024.

[17] A. Fidjeland, E. B. Roesch, M. Shanahan, and W. Luk. Nemo: A platform for neural modelling of spiking neurons using gpus. *2009 20th IEEE International Conference on Application-specific Systems, Architectures and Processors*, pages 137–144, 2009.

[18] R. V. Florian. Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural computation*, 19 6:1468–502, 2007.

[19] M.-O. Gewaltig and M. Diesmann. Nest (neural simulation tool). *Scholarpedia*, 2(4):1430, 2007.

[20] H. Hazan, D. J. Saunders, D. T. Sanghavi, H. T. Siegelmann, and R. Kozma. Unsupervised learning with self-organizing spiking neural networks. *IEEE/INNS IJCNN2018, July 15-19, 2018, Rio de Janeiro, Brazil (accepted)*, 2018.

[21] D. O. Hebb. *The organization of behavior: A neuropsychological theory*. Wiley, New York, 1949. ISBN 0-8058-4300-0.

[22] M. Hines, A. Davison, and E. Muller. Neuron and python. *Frontiers in Neuroinformatics*, 3: 1, 2009. ISSN 1662-5196. doi: 10.3389/neuro.11.001.2009. URL https://www.frontiersin.org/article/10.3389/neuro.11.001.2009.

[23] E. Hunsberger and C. Eliasmith. Spiking deep networks with lif neurons. *CoRR*, abs/1510.08829, 2015.

[24] E. M. Izhikevich. Simple model of spiking neurons. *IEEE transactions on neural networks*, 14 6:1569–72, 2003.

[25] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
[26] N. K. Kasabov. Neucube: A spiking neural network architecture for mapping, learning and understanding of spatio-temporal brain data. Neural Networks, 52:62 – 76, 2014. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2014.01.006. URL http://www.sciencedirect.com/science/article/pii/S0893608014000070.

[27] J. H. Lee, T. Delbruck, and M. Pfeiffer. Training deep spiking neural networks using backpropagation. Frontiers in Neuroscience, 10:508, 2016. ISSN 1662-453X. doi: 10.3389/fnins.2016.00508. URL https://www.frontiersin.org/article/10.3389/fnins.2016.00508.

[28] W. Maass. Lower Bounds for the Computational Power of Networks of Spiking Neurons. Neural Computation, 8(1):1–40, Jan. 1996. ISSN 0899-7667. doi: 10.1162/neco.1996.8.1.1. URL https://doi.org/10.1162/neco.1996.8.1.1.

[29] W. Maass. Networks of spiking neurons: The third generation of neural network models. Neural Networks, 10(9):1659–1671, Dec. 1997. ISSN 0893-6080. doi: 10.1016/S0893-6080(97)00011-7. URL http://www.sciencedirect.com/science/article/pii/S0893608097000117.

[30] A. H. Marblestone, G. Wayne, and K. P. Kording. Toward an integration of deep learning and neuroscience. Frontiers in Computational Neuroscience, 10:94, 2016. ISSN 1662-5188. doi: 10.3389/fncom.2016.00094. URL https://www.frontiersin.org/article/10.3389/fncom.2016.00094.

[31] H. Markram, J. Luebke, M. Frotscher, and B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. Science, 275 5297:213–5, 1997.

[32] P. O’Connor and M. Welling. Deep spiking networks. CoRR, abs/1602.08323, 2016.

[33] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch, 2017.

[34] L. A. Plana, D. M. Clark, S. Davidson, S. B. Furber, J. D. Garside, E. Painkras, J. Pepper, S. Temple, and J. Bainbridge. Spinnaker: Design and implementation of a gals multicore system-on-chip. JETC, 7:17:1–17:18, 2011.

[35] D. J. Saunders, H. T. Siegelmann, R. Kozma, and M. Ruszinkó. Stdp learning of image features with spiking neural networks. IEEE/INNS IJCNN2018, July 15-19, 2018, Rio de Janeiro, Brazil (accepted), 2018.

[36] T. C. Stewart. A technical overview of the neural engineering framework. Technical report, Centre for Theoretical Neuroscience, 2012.

[37] M. Stimberg, D. Goodman, V. Benichoux, and R. Brette. Equation-oriented specification of neural models for simulations. Frontiers in Neuroinformatics, 8:6, 2014. ISSN 1662-5196. doi: 10.3389/fninf.2014.00006. URL https://www.frontiersin.org/article/10.3389/fninf.2014.00006.
[38] D. G. Stork. Is backpropagation biologically plausible? In *International 1989 Joint Conference on Neural Networks*, pages 241–246 vol.2, 1989. doi: 10.1109/IJCNN.1989.118705.

[39] R. A. Tikidji-Hamburyan, V. Narayana, Z. Bozkus, and T. A. El-Ghazawi. Software for brain network simulations: A comparative study. *Frontiers in Neuroinformatics*, 11:46, 2017. ISSN 1662-5196. doi: 10.3389/fninf.2017.00046. URL https://www.frontiersin.org/article/10.3389/fninf.2017.00046.

[40] S. Tokui, K. Oono, S. Hido, and J. Clayton. Chainer: a next-generation open source framework for deep learning, 2015. URL http://learningsys.org/papers/LearningSys_2015_paper_33.pdf.

[41] J. Vitay, H. Dinkelbach, and F. Hamker. Annarchy: a code generation approach to neural simulations on parallel hardware. *Frontiers in Neuroinformatics*, 9:19, 2015. ISSN 1662-5196. doi: 10.3389/fninf.2015.00019. URL https://www.frontiersin.org/article/10.3389/fninf.2015.00019.

[42] W. M. K. W. Gerstner. *Spiking Neuron Models. Single Neurons, Populations, Plasticity.* Cambridge University Press, 2002.

[43] J. Wall and C. Glackin. Spiking neural network connectivity and its potential for temporal sensory processing and variable binding. *Frontiers in Computational Neuroscience*, 7:182, 2013. ISSN 1662-5188. doi: 10.3389/fncom.2013.00182. URL https://www.frontiersin.org/article/10.3389/fncom.2013.00182.

[44] J. X. Wang, Z. Kurth-Nelson, D. Kumaran, D. Tirumala, H. Soyer, J. Z. Leibo, D. Hassabis, and M. Botvinick. Prefrontal cortex as a meta-reinforcement learning system. bioRxiv, 2018. doi: 10.1101/295964. URL https://www.biorxiv.org/content/early/2018/04/13/295964.

[45] G. H. Y. LeCun, Y. Bengio. Deep learning. *Nature*, 521:436–444, May 2015.