Effects Of Real-time Synaptic Plasticity Using Spiking Neural Network Architecture

Matthew Frazier Nishant Shukla Worthy Martin

***Abstract*—Artificial Neural Networks is a promising approach to study human brain computation in hopes of achieving similar learning by artificial agents. Recent architecture design of a low-power supercomputer by the University of Manchester, the SpiNNaker, has made it easier to design highly parallel brain-inspired algorithms. We used the SpiNNaker machine to implement a neural network capable of rewiring its connection in real-time while trying to minimize information loss and maximize the decrease in statistical dependence. We believe this is the first use of synaptogenesis on a spiking neural network architecture, laying the framework for future efficient brain-like neural networks. Additionally, we explored scalability issues and unintended pitfalls with this approach.**

***Keywords*—Neural Networks, Synaptogenesis, SpiNNaker**

I. Introduction

ARTIFICIAL neural networks often require vast amounts of computation memory and space [1]. Some of the most effective algorithms such as back propagation do not scale well for a network consisting of a billion or more neurons [4]. Furthermore, most neural network models do not fully utilize the machine on which they run. The ones that do fully utilize their machine are typically implemented on Graphical Processing Units (GPUs) which disproportionately drain energy.

Finding a solution to revive neural networks to the same computational caliber as other paradigms, such as Support Vector Machines (SVM), would better adhere to biologically based design. Studying biologically inspired structures as opposed to statistically based ones allows us to grasp a better understanding of the human brain. By designing a neural network experiment on a new kind of computer architecture, this paper aims to demonstrate the high performing nature of neural networks.

Recently, deep feedforward neural networks have risen in popularity [22]. Multiple deep-learning approaches have proven to be practical. However, a major issue with deep-learning is the computation time for large networks. Deep-learning does not scale well due to the innate use of back-propagation in the algorithm.

Our approach to neural network learning is through an unsupervised model on a spiking neural network architecture. By utilizing almost every core of the 72-core energy-efficient embedded system, we are capable of reproducing complex neural network patterns in less than half the energy-cost. The network creates and removes synapses in real-time, so not only are weights learned and adjusted simultaneously, but so are the actual connections between the neurons. We demonstrate that introducing synaptic modification on a neural network leads to a biologically accurate model that outperforms previous methods in both performance and energy costs [19].

II. Background

THE subfield of AI known as neural network computation has recently received a large and growing amount of attention. Put simply; neural networks (NN) are computational models inspired by biological brains that are capable of machine learning. A NN usually consists of interconnected "neurons" (in quotation marks for they are artificially simulated to varying degrees of biological accuracy – but that topic is outside the scope of this paper) which compute from their inputs and produce various outputs, depending on the model.

This paper focuses on types of spiking neural network (SNN) models, in which the neuronal communication (and therefore, the computation of the system as a whole) is achieved via message spikes, or action potentials, from one neuron that is synapsed onto another. In this introduction, we first describe how a SNN is a good computational representation of the brain. We then describe various different spike-dependent spike-time learning methods used on a SNN, followed by an exploration of various types of SNN. We discuss why the neuromimetic hardware we use is a good platform for experimentation with SNN [11], and then describe our approach to the problem of simulating synaptogenesis on such a device and propose a solution. Our design and implementation sections explore this solution in detail, and our results section describe the outcome of our experiment. Furthermore, we discuss future work that remains to be conducted, including interesting implications such as computability beyond Von Neuman architecture and Turing machines.

A. The Brain And Other Spiking Neural Networks

The brain is a spiking neural network. That is to say; biological neurons pass information around the brain via action potentials which travel from a neurons' soma (cell body) down the axon and generate synaptic events which are received by the dendrites of all other connected neurons. Figure 1 diagrams the relevant biological prerequisites.

The process has been simplified greatly as neurobiology is not the focus of this paper, nor are the biological processes yet completely understood. What is important are the overarching properties exhibited by brains. Namely; that brains are fast, power-efficient, and capable of obtaining, representing, and integrating complex multi-dimensional information sets into useful knowledge (from unreliable sensory inputs, at that). Moreover, the property of being able to fairly reliably solve highly complex problems is something we want computers

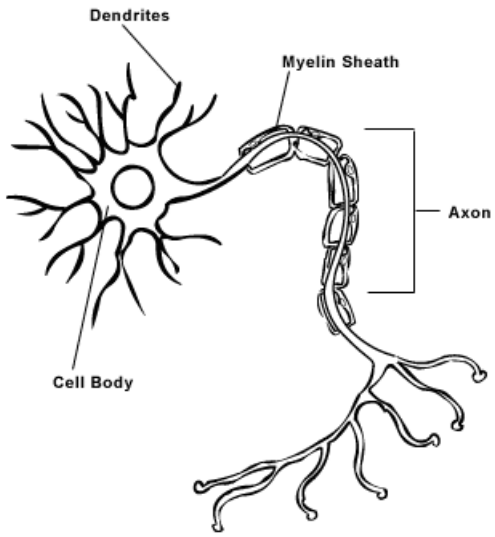

Fig. 1. Visual representation of a biological neuron identifying the cell body, dendrites, axon, and myelin sheath. Source: [27]

to have. Progress is however being made towards the goal of making machines more like minds; indeed, the bus in a computers' hardware may operate at tens of MHz, while an axon may carry only tens to hundreds of action potentials per second, about five orders of magnitude slower than a machine [10]. The power efficiency of the brain is still far ahead of even the best neuromimetic hardware but the gap is closing quickly, considering that biological evolution had several hundred million years of a head start.

One very important property of the brain that has received little attention by computer scientists, which is also believed to be quintessential in mankind's ability to adapt to and solve new and ever-more difficult problems, is neural plasticity involving the dynamic creation or removal of synapses. It would be erroneous to say that no degree of plasticity has been implemented in any type of SNN, but to the best of our knowledge there is yet no model which allows networks to add or remove synapses at runtime. Given that we do not yet fully understand the mechanisms governing neural plasticity in our own heads, it suffices to say that there is yet much work to be done in this field of computational neuroscience in which artificial neural networks dynamically alter their own topology [24].

B. Overview of Spike-time coding and learning methods

Before discussing different types of artificial spiking neural networks, it is necessary to describe various spike-time coding and learning methods that are relevant to understanding how the different types of spiking neural networks function.

Firstly, the distinction between rate- and rank-order coding must be addressed. Rate order coding is, as it sounds, a way to describe neural responses to stimulus in terms of firing rate, paying attention to the timing between spikes of a single input. Rank order, on the other hand, focuses on the relative timing of spikes from all inputs [12]. Rate order has a number of drawbacks, though the primary one is its inability to transmit and process information in a short period of time; with n neurons being able to transmit over the course of 10 ms only $log_2(nC1)$ bits of information [26]. Rank order, under the same constraints, is capable of achieving $log_2(n!)$ bits of information [16].

As detailed in [26], rate order can be considered an $analog-to-frequency$ converter, while rank order would be an $analog-to-delay$ converter, as the time an integrate-and-fire neuron takes to reach threshold is dependent upon input strength. During an initial learning phase, in which a new output neuron i is created for each N-dimensional training input pattern, connection weights $w_j, i(j = 1, 2, ..., N)$ are calculated based on the rank order learning rule:

$$w_j, i(t) = \alpha \cdot mod^{order(j,i)} \tag{1}$$

where w_j, i is the connection weight between post-synaptic neuron i and pre-synaptic neuron j, α is a learning parameter, mod is a modulation factor (determining the importance of the order of the first spikes), with $order(j, i)$ representing the order of spikes arriving at neuron i from all synapses to the neuron i (has value 0 for the first spike to neuron i, increases incrementally for each subsequent spike from other pre-synaptic neurons) [16]. Some models using rank order coding will also merge output neurons with similar weight vectors based on the Euclidean distance between them, thus reducing the number of redundant neurons that must be simulated. In short, rank order (RO) coding is assumed to be a better basis for model construction as it allows networks to process more information in a shorter amount of time with fewer redundant neurons [26]. One of the only downsides to rank order over rate order is a slightly increased computational overhead.

Moving on to learning methods, spike time dependent plasticity (STDP) implements plasticity through the use of long-term potentiation (LTP) and depression (LTD) [5], [6], [13]. Put simply, the connection weight between two neurons increases if the pre-synaptic neuron spikes before the post-synaptic neuron, and the weight decreases in the reverse case. Spike driven synaptic plasticity (SDSP) is a semi-supervised variant of STDP in which a threshold $Vmth$ is given to the post-synaptic neurons' membrane potential [16]. If the membrane potential is above $Vmth$ when an input spike arrives, potentiation occurs. Otherwise, the synapse experiences depression. These two cases are typically either shortly before or shortly after a post-synaptic spike is emitted, respectively.

STDP is the most commonly discussed paradigm for learning in neural networks. Its benefits include its unsupervised nature, and, significantly, the ability for the LTP and LTD weight modifications to be calculated by the post-synaptic neuron upon receipt of a spike. This allows the updates to be performed at a single point in the simulation: at the spike arrival event, which can further increase the efficiency of a simulation [5], [6].

We will later describe a new learning method referred to as synaptogenetic spike time dependent plasticity (SSTDP),

utilizing STDP and governed by biologically inspired synaptogenetic formulas taken from [3], and influenced by Information Theory from Shannon.

C. Types of Spiking Neural Networks

(a) **A spiking neural network (SNN)** is the most basic form, in which synaptic weights are static and neurons communicate via message spikes. While they have some drawbacks, an SNN is still a fairly powerful biologically-inspired model.

(b) **An evolving spiking neural network (eSNN)** evolves its functionality and structure based on incoming information using the RO rule, but only during a learning phase [16]. Once a neurons' weight has been set, however, this model does not allow for any further tuning of synaptic weights to reflect on other incoming spikes at the same synapse. That is to say; these synapses can capture some long-term memory during the learning phase, but have little ability to capture any short term memory, which is detrimental to the computational capabilities of the network as a whole [16].

(c) **A dynamically evolving spiking neural network (deSNN)** represents an improvement over eSNN in that it implements not only RO learning during an initial learning phase, but also a dynamic synaptic plasticity mechanism so that it will continue to learn and improve performance during a recall phase; the deSNN proposed in [16] uses the SDSP mechanism. This fixes the primary shortcoming of the eSNN model; allowing for short term memory to be stored easily, and to be potentiated as long as it remains useful [16].

D. SpiNNaker

The Spiking Neural Network Architecture (SpiNNaker) board is a neuromimetic hardware system designed by the University of Manchester with the eventual intent of simulating the human brain. Each SpiNNaker processing node is a multi-core system-on-chip (SoC) with 1GB SDRAM, containing 18 ARM968 processor cores, connected in a toroidal triangular mesh with passed messages being managed by an on-chip router for an asynchronous packet-switched network-on-chip (NoC) [20]. It is the brainchild of Steve Furber, and has a number of properties that make it useful for neural computation:

(a) **Fast:** The SpiNNaker system has a bisection bandwidth of over 5 billion packets/s [8]. It is, by design, meant to model spiking neural networks in real time. However, it can be said to outperform biology in terms of raw speed of neural activity [7], and must be slowed down to run real-time simulations [8], [21].

(b) **Power efficient:** A simulation run on the SpiNNaker system has been documented taking 100 nJ per neuron per millisecond and 43 nJ per postsynaptic potential, a smaller power consumption than any other recorded digital simulation [23].

(c) **Massively parallelized:** The brain is a massively parallel system populated with many low-performance asynchronous components known as neurons [9]. It follows that a digital system meant to effectively and efficiently emulate neural activity would also be massively parallelized [18], though each ARM core is substantially more complex than a single neuron, and can efficiently model many neurons in real time [8]. Each NoC is extended seamlessly to surrounding chips, creating a power-efficient system-wide interprocessor communication network [20].

(d) **Globally asynchronous, locally synchronous (GALS)** In a GALS system, individual processing cores have their own clock signals, potentially with different frequencies, and require no interprocessor phase-alignment [20]. Each SpiNNaker chip is its own GALS system, with an independently-clocked router managing intercore and interchip communications, leading to the SpiNNaker system as a whole also being a GALS system [9]. Major benefits of a GALS approach include that it eliminates any top-level system constraints (i.e. it is decentralized), and any timing closure issues [20]. It also offers increased flexibility regarding process variability [9].

(e) **Fault Tolerant:** In any large-scale systems fault tolerance is a major concern, and is increases in importance with the number of fallible subsystems [8]. Fault tolerance was paid a great deal of attention by the SpiNNaker development team, and listing all the built-in mechanisms at various levels of abstraction is outside the scope of this paper. To summarize, fault tolerance mechanisms exist at nearly every level of the system, including those related to processors, the interrupt controller, timers, and packet communications, allowing subsystems to generally fail gracefully without greatly affecting the performance of the system as a whole [8].

(f) **Scalable:** Due to the nature inherent to its GALS design, it is arbitrarily scalable. The only cost associated with scaling up the system is power consumption. The full million-core machine has an expected power budget of around 90 kW [8].

To summarize, the SpiNNaker specifically disregards three significant axioms of conventional supercomputing (memory coherence, synchronization, determinism) in a way that makes it well-suited for a wide range of biological and non-biological applications [8].

E. Synaptic Plasticity

A difficult problem in designing neural networks is configuring the initial parameters to optimize the network. For example, defining how the neurons are connected together greatly affects the efficiency and performance in learning [3]. Perceptrons are often designed as bipartite graphs, where every neuron in the pre-synaptic layer is connected to every neuron in the post-synaptic layer. This brute-force approach to constructing a network topology may achieve satisfactory results if the

weights converge to the optimal values within an acceptable time. However, a bipartite graph scales exponentially to the number of nodes in each layer. In our solution, we let the neural network modify itself, providing a "hands-off" approach to designing a network.

Other than updating synaptic weights to achieve learning, we introduce two additional forms of synaptic plasticity. In our study, we allow new synapses to be formed (synaptogenesis), or existing synapses to be entirely disconnected. The three types of plasticity undergo the following principles [3], [17]:

(a) **Synaptogenesis** - Unlike most networks, we allow ours to form new connections between neurons while learning occurs. The creation of a new synapse depends on whether the post-synaptic neuron is at an optimal activation. If not, the network considers forming a new synapse. Additionally, synaptogenesis only occurs between neighboring neurons, ensuring a constraint on locality. Synapses are formed by taking into account both receptivity and avidity [3], [17].

Avidity is the measure of ability for each neuron to participate in a new synaptic connection, defined as

$$A_i(t) = \frac{a}{(a + \sum_j \sum_k W_{i_k j}(t))}$$

$$a = \begin{cases} 1.0 * 10^{33} & \text{for unlimited avidity} \\ 1.0 & \text{for moderate avidity} \\ 1.0 * 10^{-3} & \text{for limiting avidity} \end{cases}$$

The receptivity of new synaptic connections is inversely proportional to the running average of the neuron's activation. We use the following equation for receptivity,

$$R_j(t) = \frac{r_1}{r_1 + \bar{y}_j(t)^{r_2}}$$

$$\bar{y}(t) = 0.99\bar{y}_j(t-1) + 0.01Y_j(t)$$

(r_1 and r_2 are experimental constants)

Finally, the probability of forming a new synapse between two neurons i and j depends on both avidity and receptivity.

$$\text{Prob(of new synapse } ij) \propto A_i(t) * R_j(t)$$

(b) **Weight Modification** - Similar to most unsupervised weight modification rules, we use a deterministic process depending on the pre- and post-synaptic activities. This type of weight adjustment is based off the Hebbian rule, where links between nodes that fire together strengthens. Specifically, weights are adjusted using the following formula,

$$\triangle w(t+1) \propto f(\text{post}_j(t)) * g(\text{pre}_j(t), w_{ij}(t))$$

(c) **Synapse Removal** - Lastly, we allow our network to undergo synaptic removal, a stochastic process where the probability of removal becomes non-zero if excitatory synapses fall bellow a specified threshold. Given some constant $\delta > 0$, probability of synaptic removal is defined as

$$\text{Prob(removal } ij) = \begin{cases} 0 & \text{if } w_{ij}(t) + \triangle w_{ij}(t+1) > \delta \\ > 0 & \text{otherwise} \end{cases}$$

Synaptic connections are slowly modified through additions or removals until optimal output firing levels are obtained.

Levy asserts that "information theory has gained popularity in recent years as a tool for understanding brain recordings."

III. Design

NEURAL networks consist of biologically inspired relationships between individual nodes (neurons). Learning in such a network occurs by intelligently adjusting weights on the links between the nodes. We define the relationship between nodes into the following three categories:

(a) **Static** - where the network topology is fixed, and can only change by manually adjusting the relationship between nodes. Most neural networks fall into this category because regardless of the number of nodes n, only one fixed relationship forms between them:

$$\text{NumberOfTopologies}(n) = 1$$

(b) **Semi-Dynamic** - in which the network is not static, and there is some freedom for nodes to rewire with other nodes in real time. The complexity grows exponentially. Each node has non-zero probability to be rewired with a constant c number of other nodes. Given n such nodes, the number of possible topologies can be up to

$$\text{NumberOfTopologies}(n) = c^n$$

(c) **Dynamic** - where each node has non-zero probability to be wired with any other neuron in the entire network. Given n nodes, the number of possible topologies becomes

$$\text{NumberOfTopologies}(n) = n^{n-1}$$

A biological neural network such as the human brain does not follow the static network topology. Connections between neurons in the brain are regularly formed and removed over time [2], [3]. Moreover, such a neural network is not fully dynamic either, since locality is a physical constraint. For example, a neuron on the far end of the left hemisphere of a brain might never directly connect with a neuron on the right hemisphere.

We designed a programming framework on SpiNNaker's interface to enable a semi-dynamic network topology. Each neuron has the flexibility to rewire with none, all, or some of the fixed number of other neurons, following our synaptic plasticity rules. To test performance, the network analyzed the MNIST database of handwritten digits. We compared information loss (see appendix A) from a network following our synaptic plasticity model, to that without synaptogensis and synapse removal.

In order to implement a spiking neural network, we choose an existing neural model. There exists multiple neural models for implementing a spiking neural network. Some of the most popular models are briefly covered below.

(a) **Integrate-and-fire (IF)** is one such model which increments the voltage while a current is present, and finally resets the voltage once a threshold V_{th} is reached. Every voltage reset produces a spike for that neuron. While computationally efficient, the IF model does not exhibit properties of the cortical spiking neurons.

(b) **Leaky integrate-and-fire (LIF)** models are a biologically revised version of the previous model. This model allows for "forgetting" isolated input currents. However, it also is insufficient for modeling cortical spiking neurons.

(c) **Izhikevich** model is an adaptation of the Integrate-and-fire model that uses a quadratic non-linearity when adjusting the voltage. This model is known for its applicability to large-scale simulations of cortical neural networks [14], [15].

We use the Izhikevich model since it our best candidate for a biologically inspired neuron that is simple enough to implement on a new computer architecture. By running an ensemble of neurons on the SpiNNaker, we wish to see both a more energy and time efficient approach for emulating neural networks. In direct comparison to Izhikevich's paper on polychronization, we present a spiking neural network on the SpiNNaker that can also exhibit reproducible time-locked but not synchronous firing patterns. This time-lock pattern is referred to as polychronization [15].

These polygroups are possible due to introducing delays in the network. The SpiNNaker machine produces natural delays as packets are sent to and from chips. We take advantages of the delays to discover these polygroups. A network of five neurons with embedded delays can produce fourteen polygroups. Generally, the number of polygroups is exponentially greater than the number of neurons in a network, as long as the delays are properly synchronized. The human brain consists of over 10^{11} neurons, which have the capability of producing even more polygroups. With our implementation, we aim to uncover the polygroups from the machine's innate delays.

IV. Implementation

BY taking advantage of the multiple independent cores on the SpiNNaker, we were able to emulate efficient real-time synaptogenesis. The neural network consisted of 39 excitatory input neurons, 9 inhibitory input neurons, and 16 excitatory output neurons.

A. Input Spikes

The SpiNNaker neural network was given a set of letters as input. Each letter was represented by a 48-dimensional binary vector of $1s$ or $0s$. Figures 2 and 3 are examples of letters with fairly similar input vectors, for illustration.

Intuitively, a 1 represents a colored pixel, and a 0 represents a blank pixel in a visual 8 x 6 image representation of the letter. Each coordinate of the 48-dimensional input spikes its corresponding input neuron. Figure 4 shows a letter with its input vector.

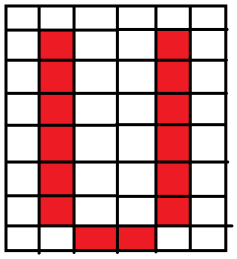

Fig. 2. The letter U represented as a series of colored and blank pixels.

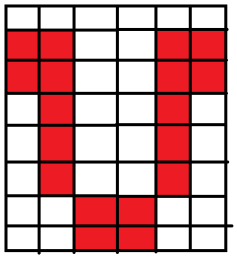

Fig. 3. The letter V represented as a series of colored and blank pixels.

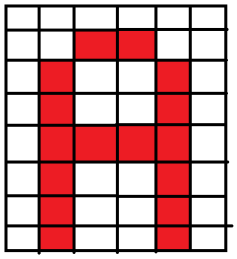

Fig. 4. The letter A is represented by the pixel sequence 000000001100010010010010011110010010010010010010

o The letters were randomly generated by a Python script which sends a spike to the SpiNNaker machine serially. Three chips receive the input generated by the Python script, and feed their generated spikes to the fourth as an output chip, illustrated in figure 5.

B. Floating Point Arithmetic

The SpiNNaker machine has no hardware support for floating point arithmetic [8]. However, the Izhikevich neural model adjusts voltage variables based off the following equations.

$\triangle v = 0.04v^2 + 5v + 140 - u + I$

$\triangle u = a(bv - u)$

In order to alleviate the rounding errors from only using integer arithmetic on the SpiNNaker machine, we scale the equations up to remove all floating points values. Then we use integer division in the final step to obtain the actual result.

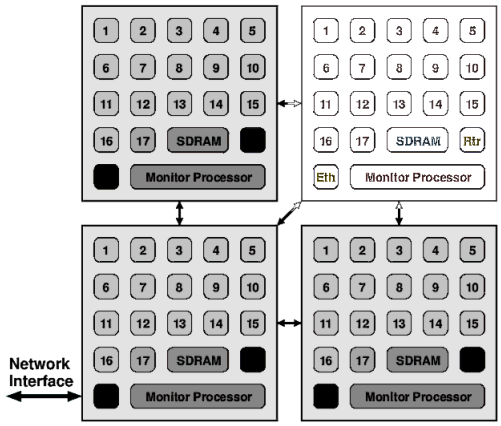

Fig. 5. Illustration of three input and one output chips on the SpiNNaker machine. The input chips consist of 9 total inhibitory neurons and 39 total excitatory neurons. Each of the output neurons are excitatory.

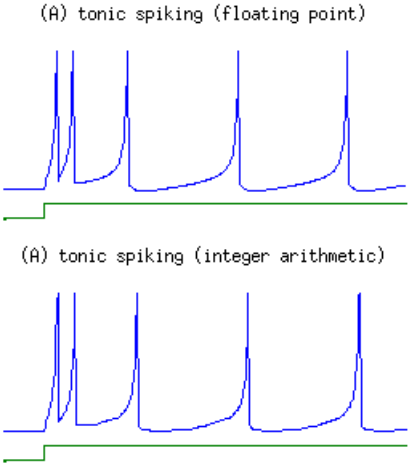

Fig. 6. Tonic spiking of a neuron.

The rounding errors of implementing the Izhikevich model are typically minute if apparent at all, shown in the figures 6 through 8 and 10 through 17. This is discussed in greater detail in appendix B.

C. Parallel Design

The algorithm written on the multi-core parallel SpiNNaker system is a translation of sequential MATLAB code used in [15]. Instead of dealing with lists of neurons, we were able to simple use a variable per each neuron core. Likewise, instead of using a two-dimensional array, where one dimension was the index of a neuron, we were able to simply use a one-dimensional array per each neuron core. Arguably, the parallel implementation on the SpiNNaker was easier to design due to the code only ever acting locally on one core.

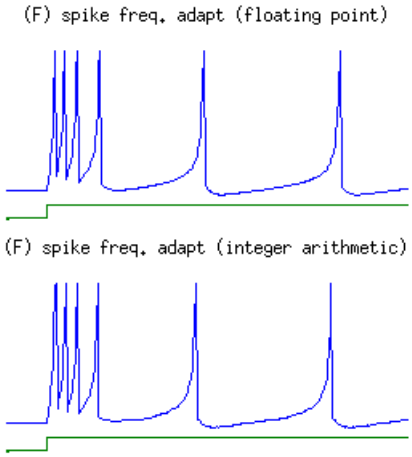

Fig. 7. Spike frequency adaptation of a neuron.

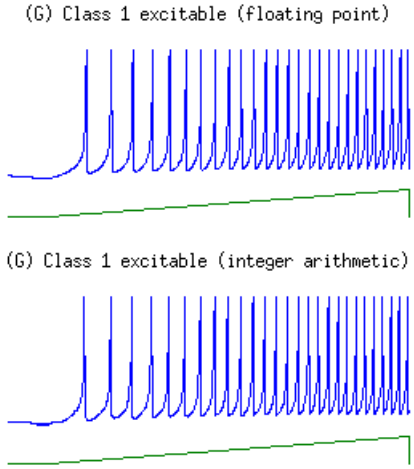

Fig. 8. Class 1 excitable neuron.

We used the weight modification rule proposed by Levy to categorize an input set of letters [3], [17].

We implemented a single-layer perception as a proof of concept to demonstrate synaptogenesis on the SpiNNaker board. In our network, every input node broadcasts a data packet to six of the output nodes. To simulate the creation and rewiring of synapses, not all messages are registered by the output neurons.

Each packet received by an output node is looked up in the output node's local list of incoming connections. If the address from the packet is not listed in the incoming connections list, it will be ignored, and no further computation will be done.

Synaptic connectivity is broken when weights fall below some negative threshold. More specifically, when an average running weight dropped below $-\delta$ for some $\delta > 0$, the link is considered disconnected.

New connections are established between nodes when an average rate of activity drops below 25%. In this case, the next

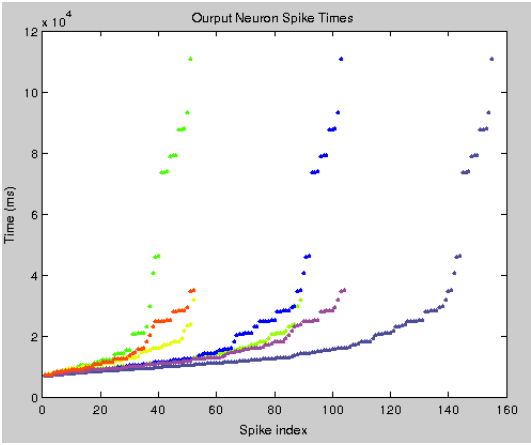

Fig. 9. The spike times of output neurons are shown in the diagram above. Each neuron is represented by a different shade of color. Some neurons share the same spiking patterns, so there are not 16 distinct curves. The curves represent the timings when each neuron fires.

data packet received from a muted neuron becomes unmuted, in hopes to raise the average rate of activity.

We combined the typical STDP mechanisms [6] with synaptogenetic formulas from [3], [17] to created a new synaptic plasticity mechanism, which we refer to as synaptogenetic spike time dependent plasticity (SSTDP). This mechanism, implemented locally on each postsynaptic neuron, handles both synaptic weight and network-level topological modifications.

V. Outcomes

After training the network on 1000 letters pixel-by-pixel, we were able to observe the oscillatory property described by Izhikevich [15]. The output neurons initially fire at a frequency of $0.476Hz$, which increases gradually. This agrees with our hypothesis that polygroup behavior is possible on a spiking neural network architecture. Refer to Figure 9 for a plot of the output spike times.

A. Results

The SpiNNaker machine used for our experiment was powered solely by USB. The total power consumption of the simulation was at most 4 Watts, which is less than 75% of the corresponding simulation on a ThinkPad T530 Laptop.

Spikes were delivered from the host machine to the SpiNNaker board in serial. Due to the clocking differences between the host machine and SpiNNaker board, we slowed down the simulation by a factor of 1000. Moreover, the internal SpiNNaker messages are prone to packet loss. Regardless of these shortcomings, the SpiNNaker simulation produced expected results.

B. Conclusion

Through empirical evidence, we have shown that by taking advantage of a spiking neural network architecture such as the SpiNNaker, we can implement a high performance energy efficient neural network system. Without the computation drawbacks of back-propagation, or the growing space complexity of global variables, we are able to train a network on a machine that uses less than 5 Watts of energy [8]. Furthermore, we have demonstrated that a neural network can produce outstanding results by having its input be fed in series as opposed to in parallel.

VI. Pitfalls

Some disadvantages are present when considering our approach for synaptogenesis on the SpiNNaker. Programming parallel applications is a challenge in its own right, but a lack of floating point support on the SpiNNaker system [8] compounds the difficulty, especially when the neural dynamics parameters and coefficients in the differential equations of Izhikevich's simple neural model are floating point numbers [14]. To get around this, we rationalized all coefficients in the equations and multiplied the entire set of equations by the lowest common multiple of the denominators of the rationalized coefficients. This incurred a slight loss of precision, but not so much that the models we used exhibited different neural activity behavior than their original counterparts with the same input. Source code illustrating this is available upon request.

Also, due to inherent limitations to Ethernet speed, our simulation had to be slowed by a factor of 1000 to allow for reliable I/O communications with the host machine. There are plans in motion to accommodate for faster communication between SpiNNaker boards and host machines, so this will no longer be an issue once higher speed I/O is achieved.

VII. Further Study

This paper reveals multiple questions that still need to be examined further. A key feature of the SpiNNaker machine is its ability to connect to other SpiNNaker chips for scalable performance. Future papers can examine the performance of a network of multiple SpiNNaker chips together. It's been shown that with "quarter of a million neurons, tens of millions of synapses and dynamic activity of over a billion synaptic events per second can be delivered within a 30 W power envelope. [25]"

APPENDIX A
SHANNON'S ENTROPY AND MUTUAL INFORMATION

Entropy is the measure of information in a system. We use it to measure the unpredictability of information content. Mutual information is a measure of the mutual dependence between two random variables. It can be expressed in terms of entropy.

A. Entropy

The entropy of a series of data is the expected value of the information.

$$H(X) = E[I(X)]$$

Information has the units of bits when the logarithm in the equation below is in base 2.

$$I(X) = -log(P(X))$$

B. Mutual Information

The mutual information between two random variables X and Y is expressed in the equation below.

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) log(\frac{p(x,y)}{p(x)p(y)})$$

An equivalent, but simpler equation is in terms of only entropy of the X and Y random variables.

$$I(X;Y) = H(X) - H(X|Y)$$

APPENDIX B
FLOATING POINT VS. INTEGER ARITHMETIC

Figures 6 through 8 and 10 through 17 illustrate the comparison between using the floating point arithmetic found in Izhikevich's original neural models [14] and the integer arithmetic we formulated for our experiment. In most cases the behavior of each is indistinguishable from one another. However, note the visually apparent rounding errors figures 10 and 12. Since our experiment used exclusively the tonic spiking integer arithmetic (with slight modifications to distinguish between excitatory and inhibitory neurons) from figure 6, the rounding errors were not a problem.

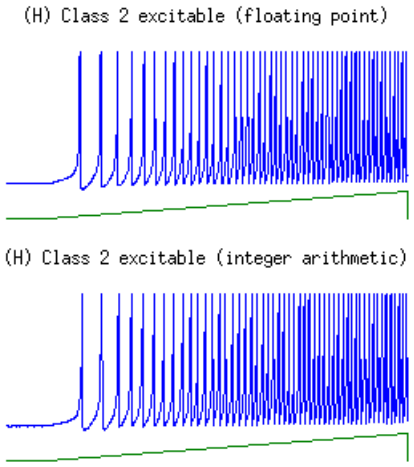

Fig. 10. Class 2 excitable neuron.

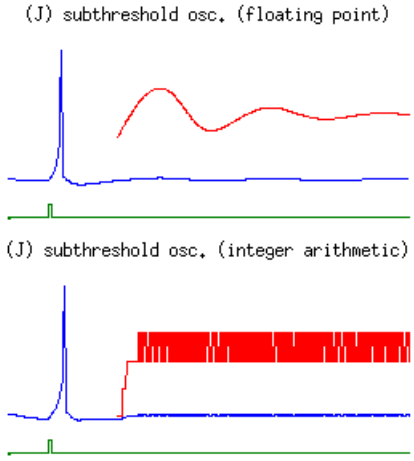

Fig. 11. Subthreshold oscillatory neuron.

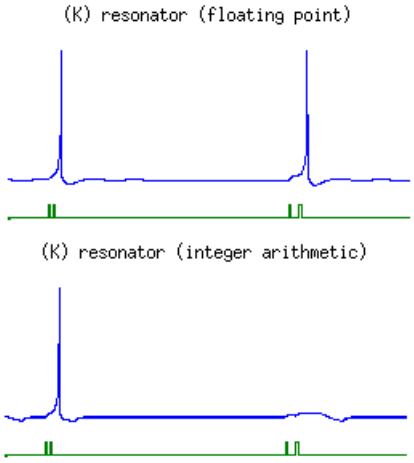

Fig. 12. Spiking resonator neuron.

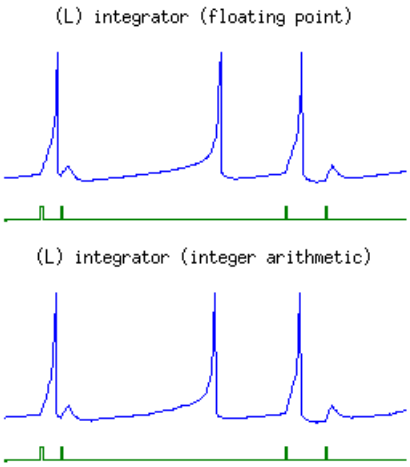

Fig. 13. Integrator neuron.

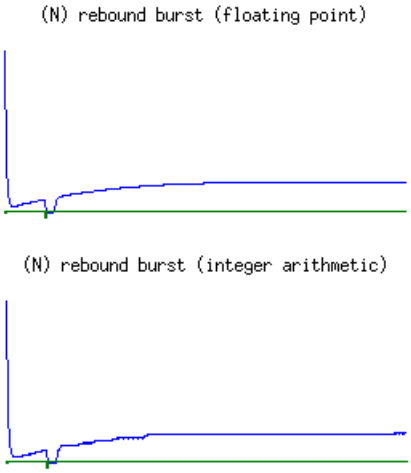

Fig. 14. Rebound burst neuron.

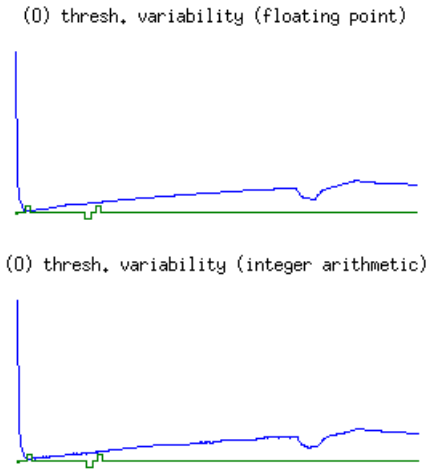

Fig. 15. Threshold variability neuron.

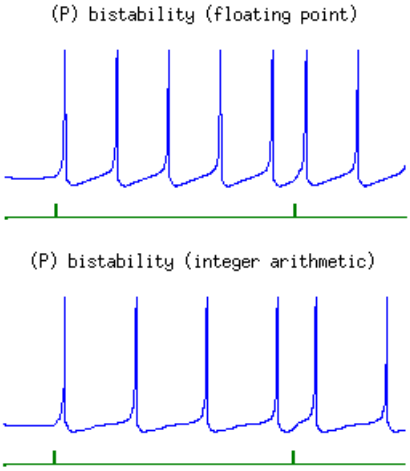

Fig. 16. Bistability neuron.

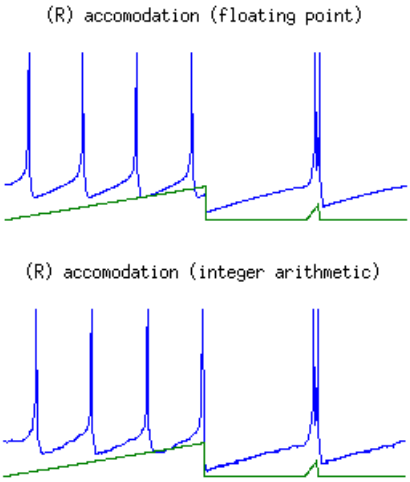

Fig. 17. Accommodation neuron.

ACKNOWLEDGMENT

The authors would like to thank Professor Worthy Martin, Associate Professor of Computer Science at the University of Virginia. The authors also greatly appreciate the generosity of the SpiNNaker team from the APT Research Group at the University of Manchester, who were gracious enough to not only loan the neuro-computing device, but also to help us debug issues whenever possible.

REFERENCES

[1] D. Adelsberger-Mangan and W. Levy in *International Joint Conference on Neural Networks 4*, 1992

[2] D. Adelsberger-Mangan and W. Levy in *Biological Cybernetics 67*, 1992, pp. 469-477

[3] D. Adelsberger-Mangan and W. Levy in *Biological Cybernetics 70*, 1993, pp. 81-87

[4] Q. Dai and N. Liu in *Neurocomputing Vol. 94*, 2012, pp.152-158

[5] S. Davies et al. in *Proceedings of International Joint Conference on Neural Networks*, San Jose, CA, 2011

[6] S. Davies et al. in *Neural Networks 32* 2012, pp. 314

[7] S. Davies et al. in *Proceedings of the 8th ACM International Conference on Computing Frontiers*, New York, NY, 2011, Article No. 15

[8] S. Furber et al. in *IEEE Transactions on Computers, Vol. 62, No. 12*, 2013, pp. 2454-2467

[9] S. Furber and A. Brown in *Ninth International Conference on Application of Concurrency to System Design*, 2009, pp. 3-12

[10] S. Furber and S. Temple in *Journal of the Royal Society Interface 4*, 2007, pp. 193-206

[11] F. Galluppi et al. in *Proceedings of the 9th conference on Computing Frontiers*, New York, NY, 2012, pp. 183-192

[12] F. Galluppi and S. Furber in *Internation Joint Conference on Neural Networks*, San Jose, CA, 2011, pp. 943-950

[13] J. Humble et al. in *From Brains to Systems*, Springer New York, 2011, pp. 19-31

[14] E. Izhikevich in *IEEE Transactions on Neural Networks 14*, 2003, pp. 1569-1572

[15] E. Izhikevich in *Neural Computation 18*, 2006, pp. 245-282

[16] N. Kasabov et al. in *Neural Networks 41*, 2013, pp. 188-201

[17] W. Levy in *Neurocomputing 58-60*, 2004, pp. 343-350

[18] J. Navaridas et al. in *Proceedings of the 23rd international conference on Supercomputing*, 2009, pp. 286-295

[19] L. Plana et al. in *IEEE Design and Test archive, Vol. 24 No. 5*, 2007, pp. 454-463

[20] L. Plana et al. in *ACM Journal on Emerging Technologies in Computing Systems, Vol. 7, No. 4*, 2011

[21] A. Rast et al. in *Neural Networks 24*, 2011, pp. 961-978

[22] S. Siniscalchi et al. in *Neurocomputing Vol 106*, 2013, pp. 148-157

[23] T. Sharp et al. in *Journal of Neuroscience Methods 210*, 2012, pp. 110-118

[24] K. Stanley and R. Miikkulainen in *Evolutionary Computation, Vol. 10 No. 2*, 2002, pp. 99-127

[25] E. Stromatias et al. in *The International Joint Conference on Neural Networks*, 2013, pp. 1-8

[26] S. Thorpe and J. Gautrais in *Computational Neuroscience; trends in research*, 1998, pp. 113-118

[27] Image, unedited, taken from http://ehumanbiofield.wikispaces.com/neurons+WG under the Creative Commons license (http://creativecommons.org/licenses/by-sa/3.0/)