

An introduction to neuromorphic computing and its potential impact for unattended ground sensors

Aaron J. Hill

Dept. 5974, Sensors & Embedded Systems
Sandia National Laboratories
Albuquerque, New Mexico
Email: ajhill@sandia.gov

Craig M. Vineyard

Dept. 1421, Cognitive & Emerging Computing
Sandia National Laboratories
Albuquerque, New Mexico
Email: cmviney@sandia.gov

Abstract—Neuromorphic computers are hardware systems that mimic the brain’s computational process phenomenology. This is in contrast to neural network accelerators, such as the Google TPU or the Intel Neural Compute Stick, which seek to accelerate the fundamental computation and data flows of neural network models used in the field of machine learning. Neuromorphic computers emulate the integrate and fire neuron dynamics of the brain to achieve a spiking communication architecture for computation. While neural networks are brain-inspired, they drastically oversimplify the brain’s computation model. Neuromorphic architectures are closer to the true computation model of the brain (albeit, still simplified). Neuromorphic computing models herald a 1000x power improvement over conventional CPU architectures. Sandia National Labs is a major contributor to the research community on neuromorphic systems by performing design analysis, evaluation, and algorithm development for neuromorphic computers. Space-based remote sensing development has been a focused target of funding for exploratory research into neuromorphic systems for their potential advantage in that program area; SNL has led some of these efforts. Recently, neuromorphic application evaluation has reached the NA-22 program area. This same exploratory research and algorithm development should penetrate the unattended ground sensor space for SNL’s mission partners and program areas. Neuromorphic computing paradigms offer a distinct advantage for the SWaP-constrained embedded systems of our diverse sponsor-driven program areas.

I. INTRODUCTION

The model of efficient computing is the human brain. At an average mass of 1.4-kg it packs approximately 86 billion neurons and 150 trillion synapses operating from a meager 20W of power with an energy cost per computation of on the order of 10 femto-joules [1]. The basic computation dynamics of the brain work as follows. A neuron is connected to other neurons via synapses¹. These synapses carry electrical signals² to down stream neurons. As a neuron receives these electrical signals, or spikes, a voltage potential accumulates in

the neuron’s cell body, or soma. The strength of the synaptic connection amplifies or depresses the electrical signal entering the soma. Each neuron has an inherent threshold for the voltage potential that it can withstand. Once this threshold is exceeded, the neuron fires by releasing a spike to all its downstream neurons via their synaptic connections. These dynamics all evolve over time. The spike has an associated delay (or travel time) for each synaptic connection and any built-up potential in the soma will slowly decay over time. It is often stated that the brain performs analog computation and digital communication. A more complex treatment of neural dynamics can be found in [2]–[6].

Researchers have long sought to develop an artificial computer that reaches the complexity scale of the mammalian brain [7], [8]. In the last 20 years there has been an explosion of research into the field of machine learning (ML). This field is predicated on the underlying model of the Artificial Neural Network (ANN). The ANN is a brain-inspired computational model that can be reduced to a *universal function approximator* [9]. The function is “programmed” through training on copious amounts of data utilizing highly computationally expensive back propagation algorithms [10]. The ANN borrows neural topology from the brain; the basic concept that neurons are connected by synapses which convey information. That is about as far as ANNs go in bio-inspiration. They remove the entire time domain complexity of the brain and replace it with non-linear differentiable transfer functions and vector-matrix multiplication operations to compute the synaptic input connections to the neuron.

Still, even with this extremely simplistic view of brain computational dynamics, ANNs have performed very well at specific human cognitive tasks, such as image recognition, natural language processing, playing games, and driving cars. Because of the increase of the computational density of these problems, their evaluation on conventional computer architectures requires an extreme amount of time and power. To this end strong efforts in industry were (and are being) undertaken to accelerate in hardware the implementation of ANNs. While much of this effort is aimed at solutions for large data centers, there is an additional effort to boost hardware implementation of ANNs for edge computing [11]. While these efforts are maturing and improving, they still

¹A synapse more directly refers to the connection between a neuron’s axon and the downstream neuron’s dendrites. A pre-synaptic neuron’s axon can “synapse” to the post-synaptic neuron’s soma, dendrite, or axon. The connection is not physical. There exists a small gap (the synaptic cleft) in which a neuro-chemical response allows the “connection” to manifest.

²The transfer of this signal is an electro-chemical process. As the electrical signal reaches the synaptic cleft it stimulates the release of neurotransmitters that are ejected and eventually land on chemical receptors that then stimulate a chemical to electrical conversion

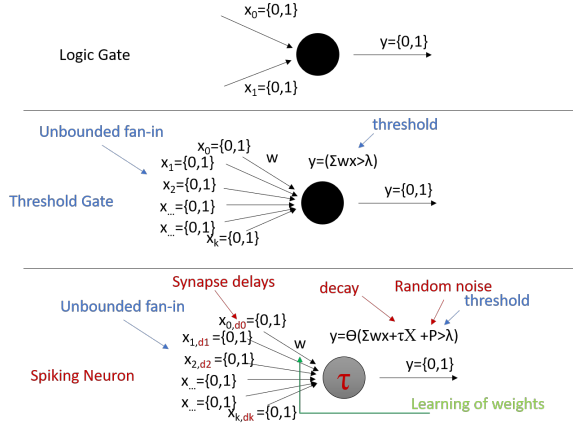


Fig. 1. The LIF neuron model is inspired by biological neurons and shares commonality between threshold gates which evolved from basic logic gates. Spiking neurons can be characterized by adding input delay and summation decay to a threshold gate model.

rely on conventional processing architectures and would be considered high-powered operations from the perspective of battery-operated embedded systems.

While the ubiquity of ML and Neural Network Accelerators has taken hold in the last 20 years, the Spiking Neural Network (SNN) model was being born in this same period. The SNN takes a different approach to brain-inspired computing than ANNs, one that is more brain inspired. The Leaky Integrate and Fire (LIF) neuron model (see Figure 1) is a common model used in SNN (in contrast to the ANN implementation of the non-linear transfer function). When a spike enters the neuron from a connected synapse, the synaptic weight is integrated into the membrane potential (the neuron’s accumulation register). At this time the neuron evaluates its potential to determine if it has exceeded its defined threshold. If the threshold has been exceeded the neuron will spike, or fire, transmitting a value of one³ to all its downstream neurons via the neuron’s synaptic connections. If the threshold was not exceeded, the membrane potential will be reduced, or leak, by some specified amount. After a spike occurs the membrane potential is reset to a defined value, typically zero. We refer to the hardware implementation of SNNs as Neuromorphic Computers (NMC).

Because of the vast complexity of the brain and the real resource limitations when engineering specific hardware solutions, there are many trade-offs made when designing an NMC. As a result there are many different implementations of SNN in hardware. But, all NMC systems implement the underlying principles of spike-based communication, asynchronous operation, and massively parallel processing elements with tightly integrated computation and memory. There is a debate in the field of whether NMC implementations should be digital, analog, or both. In a pursuit of understanding which

³Ultimately, this unit value is multiplied by the synaptic weight before accumulating into the neurons membrane potential. Thus, spikes have the end effect of increasing the neurons potential by the value of the synaptic weight—a measure of the “strength” of the synaptic connection.

approach is best, a number of large-scale implementations have emerged in the recent decade. On the analog front there is the Stanford Neurogrid [12] and the Heidelberg BrainScaleS [13] system. For digital systems there is IBM’s TrueNorth [14] and the Manchester SpiNNaker [15] system. Steve Furber provides a good treatment of these systems in [16]. A recent edition to the digital domain is Intel’s Loihi [17] system. These five systems are a huge step in the advancement of NMC research and development and SNL has varying degrees of experience with all of them.

II. WHY NEUROMORPHIC?

The scope of applications for ML is vast, but their practical realization in embedded applications is limited by the hardware resource requirements of the underlying systems. The types of AI applications we are familiar with on our mobile devices all require a connection to a data center in the “cloud” to evaluate the large-scale ML models required to complete the task. Such a communication and computation paradigm is a non-starter in the unattended ground sensor domain because these devices need to exist in the field for years on end from a single battery source. A much lower power system design for its communication and computation is required. We believe that SNN will easily enable low-power hardware evaluation of deeply complex AI tasks due to their event driven sparsity.

Two brain-inspired principles drive the innate low power nature of NMC, sparse connectivity, and event driven processing and communication. Each neuron in the brain connects to an average of 7,000 other neurons in an ionic soup of 86 billion neurons. Further, the entire brain is not spiking all the time during a task. By only requiring a small subset of the computational elements at any point in time, power consumption is greatly reduced by not powering unused elements. Second, neuron computation and firing dynamics are event driven. If a neuron does not receive a spike, it will not expend energy in processing it. The same is true for NMC designs. The event driven nature combined with the sparsity of connections is an evolutionary design phenomenon that reduces overall energy consumption. Whereas, ANN implementations will seek very densely connected layers of neurons which drive high computational cost and power.

The scientific community has accepted that Moore’s law has ended [18]. This is the self-fulfilling prophecy pronounced by Gordon Moore of Intel in 1965 that semiconductor complexity (transistor count) would double every two years. The statement has been misconstrued over time and the popular understanding is that semiconductor performance would double every 18 to 24 months. Unfortunately, the performance and transistor count progress is hitting physical reality. Performance has flat lined due to Dennard scaling about 15 years ago; as transistors get smaller, power density increases and clock frequency cannot be increased. Current process fabrication technology nodes are at 5-nm, which most in industry believe is the end; pushing any further is too cost prohibitive [18]. Regardless, its an inevitable reality that transistors can only become so small under current silicon CMOS technology.

A typical modern desktop CPU operates at 65W and a clock frequency of 2-3 GHz. If we assume that a handful (a number less than 10) of clock cycles are required on average for a single operation, then the math dictates that a typical CPU operation costs on the order of nano-joules of energy. In contrast, the basic operation of a NMC is a synaptic event, and the typical NMC system today costs on the order of pico-joules of energy per synaptic event [16]. This three orders of magnitude difference in base-operation energy cost is the loose basis of the 1000x power improvement. However, comparison of SNN to conventional computing is non-trivial and Vineyard *et al.* caution against such comparison because the fundamental compute paradigm between conventional and neuromorphic systems is very different. He argues that the computational objectives being optimized should serve as the basis for comparison [19]. His work [20] has shown, with existing hardware, at least a 100x power improvement of NMC systems over conventional CPU/GPU architectures on a Remote Sensing task. This is still an amazing improvement.

There are two notional reasons for these power gains with neuromorphic systems in addition to the sparsity and event driven nature of their architectures. The first is that there are improved communication protocols between massively parallel simple computing elements with co-located memory [1]. The other is that digital NMC systems interpret a spike as the communication of a single bit of information. Since, these systems only use energy to communicate the value of a 1, no energy is spent communicating the value of a 0, it is simply the absence of a spike. This temporal dynamic is one of the fascinating elements of NMC architectures. The integration of time in the information propagation and computation enables SNNs to efficiently extract temporal information from time-dependent data. Furthermore, the use of time in the computation model is an energy-free exchange of information; the precise time that a spike arrives conveys additional meaning. Intrinsically, NMCs are spatial-temporal systems providing different means of data representation and processing.

Neuromorphic computers will enable the future of computing and be more readily available near term to embedded systems. The National Security Program's (NSP) Intelligence Science for Proliferation (ISP) Investment Area (IA) should prepare for this future, today.

III. SANDIA'S LEADING ROLE IN NEUROMORPHIC COMPUTING RESEARCH

Neuromorphic Computing has grown dramatically at SNL over the past 10 years. The largest single investment in neural computing at SNL was the Hardware Acceleration of Adaptive Neural Algorithms (HAANA) Grand Challenge [21], a laboratory directed research & development (LDRD) effort that amounted to approximately \$15M over 3 years, and was funded from FY15-FY17 (PI Conrad James). HAANA was broad, covering three technical research areas (hardware, architectures, and algorithms) as well as a mission impact team. To convey the history of neural computing research at

SNL, it is convenient to discuss research efforts before, during, and after HAANA.

A. Pre-HAANA Funding

Before HAANA, research efforts were mostly conducted through small-scale LDRDs within different investment areas. Much of this work focused on technologies related to neuromorphic computing (i.e., computational neuroscience, memristor devices), as opposed to direct examination of neuromorphic computing. We specifically highlight two efforts (there are many others) which helped initiate the NMC research.

Episodic Memory and Neurogenesis (FY12-14; **PI Aimone**) was a CS&T LDRD focused on examining the biological process of adult neurogenesis, whereby the brain adds new neurons to facilitate learning. Insights regarding neurogenesis were used in adaptive algorithm research in HAANA and are motivating a project currently funded by DARPA and a submitted LDRD project.

Cognitive-inspired Cyber Protection Architecture (FY13-14; **PI Naegle**) was a Cyber Security LDRD focused on developing a streaming brain-inspired architecture for rapid pattern matching of streaming data. This effort involved evaluating novel algorithms and designing a simulator for the spiking neuromorphic architecture the effort developed.

B. HAANA-era Funding

Research in the community was dominated by HAANA, which encouraged significant growth in the program. This shifted the emphasis of research away from foundational neuroscience and hardware efforts to more algorithms and architecture research, especially in targeting applications such as image processing and cyber security. Over all, this effort produced 48 peer reviewed publications on neuromorphic computing research; only a handful are highlighted below with the full scope of research reported in [21].

Algorithms (**PI Aimone**): Focused primarily on theory behind algorithms for spiking neuromorphic hardware, developing adaptive deep learning algorithms, and formal analysis of different neuromorphic approaches.

Architecture (**PI Naegle**): Focused on developing a streaming neuromorphic architecture for cyber pattern matching and a prototype spiking neuromorphic architecture. Hill and Vineyard helped assess the trade-offs of algorithm and architecture interplay.

Hardware (**PI Marinella**): Focused on researching memristive devices for neuromorphic analog computing approaches and developing a mod-sim framework for assessing the value of different devices within analog architectures.

Cyber (**PI Doak**): Focused on identifying cyber security problems amenable to adaptive pattern matching and deep learning algorithms.

Neural Memory Management (FY16-17; **PI Vineyard**) was a CIS LDRD focused on identifying adaptive strategies for resource allocation inspired by neural dynamics.

Neural Adaptive Filtering (FY16-18; **PI Chance**) was a Global Security (GS) LDRD focused on adaptive filter

techniques inspired by the mammalian retinal structure. The algorithms developed out of this were later implemented to neuromorphic systems.

C. Recent Efforts (Post HAANA)

After HAANA, the main research pillars (algorithms, architecture, hardware) separated into efforts within specific disciplines. There has been an extraordinary boom in neuromorphic research efforts post-HAANA, all of which would be overwhelming to capture entirely. This summary will focus on neural algorithms and neuromorphic computing applications with a sampling to illustrate breadth and depth of the research. There are inevitably efforts of no lesser value left out and additionally other derivative projects with more of an applied ML or hardware focus that are not mentioned here.

Neuromorphic Remote Sensing (FY17-19; **PI Vineyard**) was a GS LDRD focused on evaluating the potential impact of neuromorphic processors on embedding image processing tasks on remote platforms, such as the focal plane array of a space system.

SPARR (FY17-FY19; **PI Hays**) was a GS LDRD focused on exploring brain-inspired sensing in the research and development of a novel spiking photonic focal plane. The effort showed the ability to perform wavelet transforms at the sensor.

Neural PDE (FY18-20; **PI Aimone**) was a CIS LDRD focused on developing spiking algorithms for solving partial differential equations (PDEs), with a particular focus on stochastic PDE solutions for diffusion based applications, such as radiation transport and molecular dynamics, using native stochastic capabilities inherent in spiking neuromorphic hardware platforms.

Neural Automatic Target Recognition (FY19-21; **PI Vineyard**) was an Autonomy for Hypersonics (A4H) LDRD focused on developing spiking algorithm implementations of target recognition algorithms such as MPM for synthetic aperture radar (SAR).

BrainSLAM (FY19; **PI Aimone**) was an exploratory A4H seedling LDRD focused on exploring the potential of brain-inspired navigation algorithms for GPS-deprived navigation in autonomous platforms. A follow on to this effort, extending the successful initial proof-of-concept is *NeuroGrid* (FY21-23; **PI Wang**).

Dragonfly Target Interception (FY19; **PI Chance**) was an exploratory A4H seedling LDRD that utilized target interception dynamics inspired by the Dragonfly's cortical networks; developing novel spike-based target interception algorithms. Extensions to this effort have pursued spiking implementations of the base continuous-valued computation as well as hardware demonstrations (FY20-present; **PI Chance**).

Low-Power Inference for Self-Driving Vehicles (FY20; **PI Severa**) was an EHS seedling effort demonstrating how low-precision spiking neural networks can help enable the onboard processing needs for autonomous vehicles.

ASC Beyond Moore's Law Neuromorphic Computing (**PI's Aimone & Vineyard**) is a multi-year, on-going effort exploring how neuromorphic computing technologies can enable

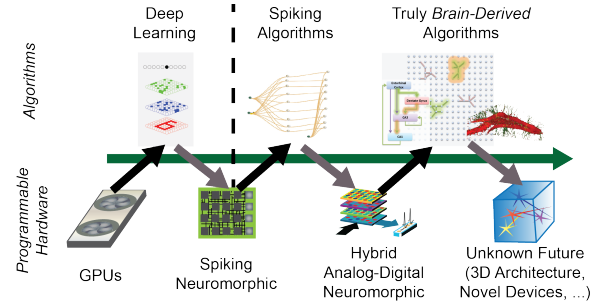


Fig. 2. Neuromorphic research has immediate benefits and long term potential in future systems. SNL is helping pave the way for the future growth of this research field with deep investments into algorithm, architecture, and hardware co-design to drive the future innovation of neuromorphic computing systems.

scientific computing workloads for next generation High Performance Computing (HPC) architectures.

Over the past decade, the “Cognitive and Emerging Computing” Department (1421) has built the Neural Exploration and Research Laboratory (NERL) to house a vast array of neural accelerators and neuromorphic computing systems to aid in their research efforts. Notably, they acquired an Intel Loihi-based system [22] with 50 million neurons, 50 billion synapses, and 384 Loihi chips in a 5U rack through a close research partnership between Intel and SNL⁴ [23].

Through the HAANA Grand Challenge and many other efforts, SNL has risen to become a major leader in all aspects of neuromorphic computing (Figure 2). This leadership is strengthened by SNL’s brain-inspired computing research collaboration efforts across multiple national labs such as LANL, LLNL, PNNL, and ORNL.

IV. NEUROMORPHIC RESEARCH IN REMOTE SENSING AND PROLIFERATION SENSING

Space-based Remote Sensing (RS) applications are tightly integrated into many facets of SNL’s mission domains. Remote Sensing systems are quickly evolving in three dimensions, hyper-spectrally (sensing more spectral bands), hyper-temporally (faster sampling rates) and hyper-spatially (increasing number of smaller pixels). This evolution creates the challenging problem of deploying expensive and computationally complex algorithms in SWaP constrained environments. There are three broad application areas which jointly capture the problem domain of Remote Sensing tasks: signal processing, signal classification, and signal understanding.

While RS has historically considered neural network approaches to improve on signal processing, classification and understanding, these approaches faltered due to high power consumption and a lack of strong theoretical understanding. Combined with SNL’s research excellence and recent advances in hardware acceleration of ANN and SNN, these neural inspired approaches are showing more promise. The newer

⁴In FY22 SNL will be receiving Intel’s 2nd generation architecture with 128 million neurons consuming under 40W, and in FY23 will be receiving a large-scale 2nd generation system using 3D scaling technology with 1 billion neurons consuming under 100W

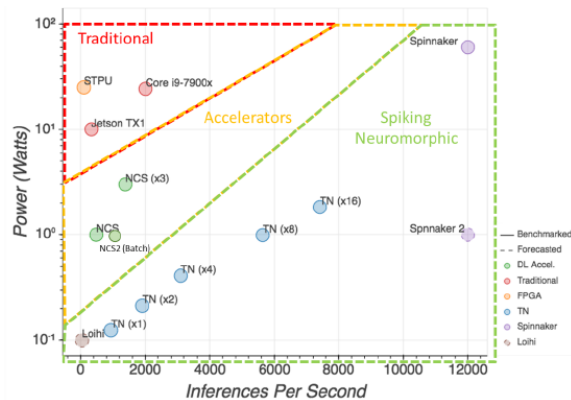


Fig. 3. A comparison of power vs performance for various traditional CPU/GPU, ANN Accelerator, and Spiking Neuromorphic systems. The inference task was a common SAR-based image processing task. The STPU is a SNL-developed neuromorphic architecture implemented on an FPGA and developed through the HAANA GC LDRD effort.

hardware implementations are often proving to be orders of magnitude faster computationally and cheaper in terms of power consumption (see Figure 3), making them well suited for resource constrained environments like Remote Sensing.

A 3-year LDRD effort that began in FY17 investigated the application of neuromorphic computing systems and SNN algorithms to the RS domain [24]. The research goal was to simulate several neural inspired architectures to assess and identify their computational implications for the RS domain, and to evaluate computationally intensive RS algorithms on these neural architectures. The result of this effort was the development of an algorithmic capability to produce spiking neuromorphic compatible networks, an understanding of how to benchmark and compare heterogeneous architectures, and a co-design capability to explore algorithm and architecture mapping. This work led to additional funding from SPP sources at \$2M for 2 years for developing SNN implementations of specific RS imaging algorithms.

Space-based RS shares a number of common parallels to terrestrial RS, specifically unattended ground sensors (UGS). Both mission areas operate in highly constrained SWaP environments, require sophisticated processing of sensor data, and demand more and more processing capability with each new mission. Remote Sensing domain problems have a natural segue into standalone battery-operated sensor applications that fall into the NA-22 program area.

Vineyard is currently leading an NA-22 funded effort known as the SEEK project. SEEK is a 2 year \$1.25M joint effort with LANL scoping the role of NMC for proliferation monitoring. SEEK hypothesizes that neuromorphic hardware, by virtue of its low-power brain-inspired communication structure and highly parallel design, offers a solution to the NA-22 mission area; enabling greater processing at the sensor. Complementing the development of NA-22 relevant neural algorithms and their characterization on emerging neuromorphic architectures, the effort will develop an understanding of which architectural features impact the joint algorithm and architecture perfor-

mance for NA-22 missions. A complementary NA-22 scoping study (FY21; PI Shank) considers how brain-inspired event-sensors may offer novel sensor capabilities for proliferation monitoring.

V. NEUROMORPHIC RESEARCH FOR UNATTENDED GROUND SENSORS

While application areas on the cutting edge of processing technology, such as supercomputers and data centers, are feeling the pinch of the death of Moore's law, the low-power embedded sensors space still has some time to take advantage of performance gains down the road. This is because the embedded processing ICs used for low-power applications lag behind the advancements of transistor scaling. But, the bemoaning of the HPC and data center universe is a warning call to the embedded computing world; it is a window into our inevitable future. Adaption of new computing paradigms and processing platforms takes years of research, development, testing, characterization, and trust building. The future of computing is heterogeneous specialization [25], [26], and it will also be the future of embedded systems. Embedded computing is intrinsically well suited for the paradigm shift NMC seeks to enable. While HPC has thrived in the pursuit of enabling max computations (marching towards exascale), the embedded world conversely seeks the minimal solution. Accordingly, we see opportunity for embracing this paradigm shift as a strategic advantage rather than out of necessity.

Many of the common ANN-based ML algorithms may offer value to many of our UGS sensing application needs, but the power requirements for implementing such algorithms limits their utility. A SNL developed tool (Whetstone) [27] exists for porting traditional ANN algorithms to SNN algorithms amenable for implementation to NMC systems. This reduces the burden of entry for NMC research into the low-power embedded systems domain and facilitates the exploration of more processing capability at the sensor node. Effectively, by enabling enhanced processing at the sensor, the amount of data transmitted for subsequent processing or analyst interpretation may be reduced. Additionally, more processing at the node allows the exploration of more feature rich classification of raw sensor data and better discrimination of events of interest.

There is no question that ML has been the buzzword in NSP ISP IA for the last 3-4 years. It is promising to see many mission relevant funding efforts exploring the use of ANN type algorithms to mission relevant problems. However, results seem mixed. For example, BALDR (investigation of radio isotope identification using ML algorithms) is an excellent example of successfully applying ML to a domain relevant problem. Other efforts have not been so successful due to the required degradation of the high-fidelity models, resulting in loss of accuracy, when moving to the embedded processing space. Further, many of the models are too large to implement or require too much computation time to execute, expending more power than the traditional algorithm for performing the same task. This is where NMC can help. Taking a deep look

into the respective ground base sensor algorithms and re-imagining them into SNN models for implementation on NMC hardware can provide computational and power gains where traditional ANN approaches are failing. This has been done in the RS domain, and is currently happening in the NA-22 domain.

In FY22, the NSP ISP LDRD IA funded a 2-year LDRD effort titled “Autonomous Reconfigurable Intelligence at the Edge”, as a first entry of this IA into neuromorphic computing research. This effort investigates brain-inspired computing algorithms applied to mission domain relevant problems in the ISP mission domain. Specifically, this effort seeks to leverage the notion of context switching in the brain with liquid state machines (LSM) [28] to develop an algorithm that can deploy multiple ML models on the same network. Additionally, many of the sensor phenomenologies for ground sensors are one-dimensional making it hard to perform accurate discrimination of the event data when looking at data from a single sensor. The LSM is a SNN-based algorithm that performs high dimensional feature space discrimination on streaming data; a type of data common in the UGS domain. This type of neural model can easily perform sensor data fusion while elevating the dimensionality of the data leading to better discrimination. Leveraging a natively spike-based algorithm for data sensor classification will make it easily amenable to neuromorphic hardware. This is a promising sign that neuromorphic research is starting to make headway into the NSP ISP IA. It is our hope that this paper has supported this continued path and provided evidence to the mission need.

ACKNOWLEDGMENT

This work was funded by Program Development funds in department 5974 under project and task number 275 / 30.09.08.01.01.05.06. Special thanks to David I. Toledo (5974) for supporting this effort, Nicole Woloshuk (106591) for managing the money, and department 1421 for outstanding cutting edge research in the field of neuromorphic computing.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-NA0003525.

REFERENCES

- [1] M. Bouvier, A. Valentian, T. Mesquida, F. Rummens, M. Reyboz, E. Vianello, and E. Beigne, “Spiking neural networks hardware implementations and challenges: A survey,” *ACM Journal on Emerging Technologies in Computing Systems (JETC)*, vol. 15, no. 2, pp. 1–35, 2019.
- [2] P. Sterling and S. Laughlin, *Principles of neural design*. MIT press, 2015.
- [3] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.
- [4] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.
- [5] E. M. Izhikevich, “Which model to use for cortical spiking neurons?” *IEEE transactions on neural networks*, vol. 15, no. 5, pp. 1063–1070, 2004.
- [6] W. Maass, “Networks of spiking neurons: the third generation of neural network models,” *Neural networks*, vol. 10, no. 9, pp. 1659–1671, 1997.
- [7] C. D. James *et al.*, “A historical survey of algorithms and hardware architectures for neural-inspired and neuromorphic computing applications,” *Biologically Inspired Cognitive Architectures*, vol. 19, pp. 49–64, 2017.
- [8] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, “A survey of neuromorphic computing and neural networks in hardware,” *arXiv preprint arXiv:1705.06963*, 2017.
- [9] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.
- [10] D. E. Rumelhart, R. Durbin, R. Golden, and Y. Chauvin, “Backpropagation: The basic theory,” *Backpropagation: Theory, architectures and applications*, pp. 1–34, 1995.
- [11] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, “Survey of machine learning accelerators,” in *2020 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE, 2020, pp. 1–12.
- [12] B. V. Benjamin *et al.*, “Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations,” *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014.
- [13] J. Schemmel, D. Bröderle, A. Grübl, M. Hock, K. Meier, and S. Millner, “A wafer-scale neuromorphic hardware system for large-scale neural modeling,” in *2010 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2010, pp. 1947–1950.
- [14] P. A. Merolla *et al.*, “A million spiking-neuron integrated circuit with a scalable communication network and interface,” *Science*, vol. 345, no. 6197, pp. 668–673, 2014.
- [15] S. B. Furber, F. Galluppi, S. Temple, and L. A. Plana, “The spinnaker project,” *Proceedings of the IEEE*, vol. 102, no. 5, pp. 652–665, 2014.
- [16] S. Furber, “Large-scale neuromorphic computing systems,” *Journal of neural engineering*, vol. 13, no. 5, p. 051001, 2016.
- [17] M. Davies *et al.*, “Loihi: A neuromorphic manycore processor with on-chip learning,” *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018.
- [18] R. Aitken. (2019) Moore’s law ending? no problem. [Online]. Available: <https://www.eetimes.com/moores-law-ending-no-problem/>
- [19] C. M. Vineyard, M. Plagge, and S. Green, “Comparing neural accelerators & neuromorphic architectures the false idol of operations,” in *Proceedings of the Neuro-inspired Computational Elements Workshop*, 2020, pp. 1–6.
- [20] C. Vineyard, W. Severa, M. Kagie, A. Scholand, and P. Hays, “A resurgence in neuromorphic architectures enabling remote sensing computation,” in *2019 IEEE Space Computing Conference (SCC)*. IEEE, 2019, pp. 33–40.
- [21] C. D. James, “Hardware acceleration of adaptive neural algorithms,” Sandia National Labs, Albuquerque, NM, Tech. Rep. SAND2017-12810, Nov. 2017.
- [22] Intel. (2020) Intel scales neuromorphic research system to 100 million neurons. [Online]. Available: <https://newsroom.intel.com/news/intel-scales-neuromorphic-research-system-100-million-neurons>
- [23] —. (2020) Intel and sandia national labs collaborate on neuromorphic computing. [Online]. Available: <https://www.intel.com/content/www/us/en/newsroom/news/sandia-national-labs-neuromorphic-computing.html>
- [24] C. M. Vineyard, “Neural inspired computation remote sensing platform,” Sandia National Labs, Albuquerque, NM, Tech. Rep. SAND2019-11291, Sep. 2019.
- [25] B. Aimone, B. Hwang, and S. Collis, “From bits to brilliance,” Sep. 2019, delivered to ES&T Priority 6 team of the Lab’s Strategic Planning effort.
- [26] J. Reinders. (2020) The future of computing is heterogeneous, cpu and friends. [Online]. Available: <https://www.hpcwire.com/2019/10/07/the-future-of-computing-is-heterogeneous-cpu-and-friends/>
- [27] W. Severa, C. M. Vineyard, R. Dellana, S. J. Verzi, and J. B. Aimone, “Training deep neural networks for binary communication with the whetstone method,” *Nature Machine Intelligence*, vol. 1, no. 2, pp. 86–94, 2019.
- [28] W. Maass, T. Natschlager, and H. Markram, “Real-time computing without stable states: A new framework for neural computation based on perturbations,” *Neural computation*, vol. 14, no. 11, pp. 2531–2560, 2002.