Advanced Ultra Low-Power Deep Learning Applications with Neuromorphic Computing

Mark Barnell
Air Force Research Laboratory
Information Directorate
Rome, NY, USA
mark.barnell.1@us.af.mil

Courtney Raymond
Air Force Research Laboratory
Information Directorate
Rome, NY, USA
courtney.raymond.1@us.af.mil

Lisa Loomis
Air Force Research Laboratory
Information Directorate
Rome, NY, USA
lisa.loomis.3@us.af.mil

Darrek Isereau
SRC, Inc.
Defense Systems & Solutions
North Syracuse, NY, USA
isereau@srcinc.com

Daniel Brown
SRC, Inc.
Defense Systems & Solutions
North Syracuse, NY, USA
dtbrown@srcinc.com

Francesca Vidal
SRC, Inc.
Defense Systems & Solutions
North Syracuse, NY, USA
fvidal@srcinc.com

Steven Smiley
SRC, Inc.
Defense Systems & Solutions
North Syracuse, NY, USA
ssmiley@srcinc.com

***Abstract* — The latest Intel neuromorphic processor, Loihi 2, provides a breakthrough in Artificial Intelligence (AI) for computing at the edge, where sensor information is collected. The computing architecture does this by leveraging computations at the transistor level in a fashion analogous to the human brain's biological neural networks (vs. a Von Neumann compute architecture). The Loihi 2's high performance, small form factor, and low-power consumption makes it a unique capability that is well suited for use in devices. Our technical approach and findings support extreme computing needs for the internet of things (IoT) and various airborne platforms' applications. The recently released Loihi 2 and the novel research completed on this effort were combined to accelerate development and demonstration of a new concept of operation for machine learning at the edge. This research included the development of spiking neural networks (SNN) on sensor data representative of information sources from a small research platform. Our concept uses the representative sensor data to predict the platform mode through machine learning. Importantly, our technical approach allowed us to rapidly scale from IBM's TrueNorth Corelet framework to the *Lava* framework, which Intel's Loihi 2 neuromorphic processor utilizes. The use of the *Lava* framework demonstrates the art-of-the-possible in edge computing by demonstrating capabilities on small airborne platform sensor data and wide extensibility to other domains that can use this neuromorphic compute hardware. In summary, this research included the use of new compute frameworks, novel processing algorithms, and a unique concept of operation. This technical approach resulted in the classification of the platform mode given the sensor information with accuracies up to 97.6%.**

***Keywords* — Extreme Computing, Machine Learning, High Performance Embedded Computing, Neuromorphic Computing, Deep Learning, Intel Loihi 2, Autonomous Operation.**

Received an approval for public release by the Air Force Research Laboratory (AFRL) on 15 Sep 2023, case number AFRL-2023-4323. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of AFRL or its contractors. The data analyses were partially funded under AFRL's FA8750-21-C-1511 contract, which was started in 29 April 2021 and will continue through 29 January 2024

I. Introduction

This research advances extreme computing technologies (computing hardware, machine learning, algorithms) through the development and demonstration of new capabilities to support several use cases and applications. The research does this by using frameworks utilized by newly invented neuromorphic computing, machine learning algorithms, and the innovative use on platform sensor data with extensibility to other information sources, such as electro-optical, infrared and/or radar.

Background and insight into recent research, as well as demand signals that make this research appropriate and applicable are provided in Section II. The *Compute Hardware* used is introduced in Section III. Section IV describes the *Compute Software*. The *Neuromorphic Classification Algorithm* is described in Section V. The *Data Description* is introduced in Section VI. The *Processing Approach, Results Conclusions and Future Research* are described in Sections VII, VIII and IX respectively.

II. Background/Significance

The Air Force Research Laboratory, Information Directorate (AFRL), High Performance Systems Branch is developing and demonstrating new computing architectures that are providing unique high-performance embedded computing (HPEC) solutions meeting the most demanding operational and tactical processing requirements for emerging and future surveillance operations.

Sensor capabilities have become less expensive and more prevalent; this has created vast quantities of data which must be analyzed promptly to provide information in a timely manner. The data is either stored and/or downlinked for post processing delaying the time relevance of the information. This information can be utilized to better inform and support disaster relief efforts where reduced processing timelines can save lives. Further, more time saving can be obtained through upstream compute systems that operate autonomously, i.e., with very limited or no user interaction [1, 2, 3]. Therefore, sensors have rapidly increased in fidelity and are now able to collect vast quantities of data, which must be analyzed promptly to provide critical information [4].

III. Compute Hardware

In the late 2021, Intel released an advanced neuromorphic processor called *Loihi 2*. The Loihi 2 has over 2.3 billion transistors with over a million neurons per chip, which contain state variable allocation between 0 to 4096. This makes Loihi 2 outperform its predecessor by 10x [5]. The Loihi 2 supports low-power applications, below 1 Watt with a die area of only 31 mm^2. Loihi 2 is an advancement to its predecessor, which is a 60-mm^2 chip [6]. In addition, its intuitive Python-based API for specifying SNNs, a compiler and runtime library for building and executing SNNs make it a practical solution [7]. This makes it a favorable compute asset for extreme edge computing research and development. Additional details on this chip architecture are shown in Fig. 1.

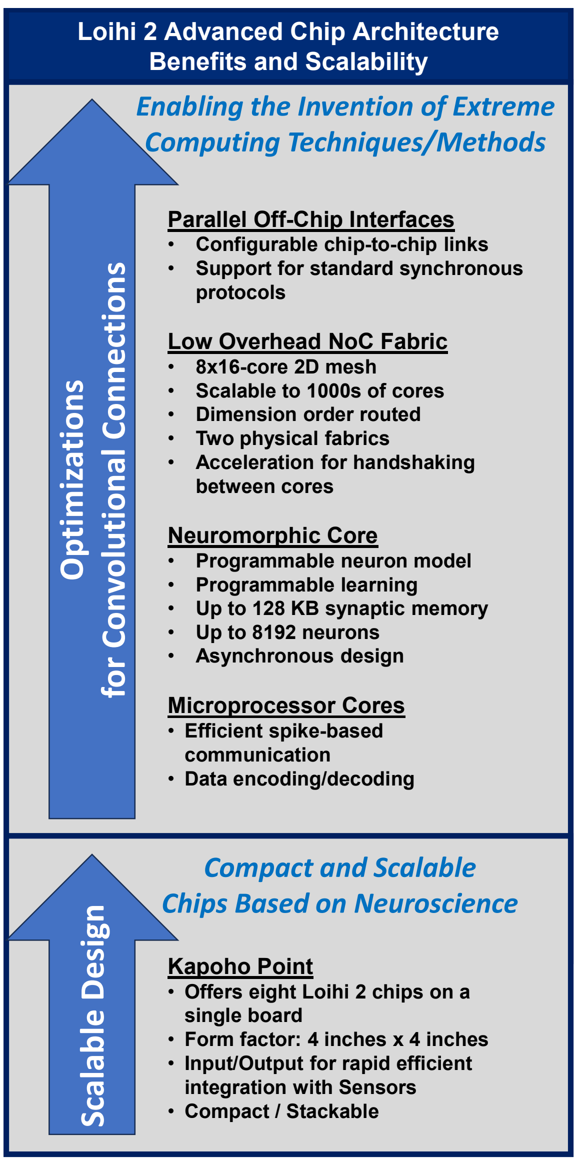

Fig. 1. Loihi 2 Chip Architecture, Benefits and Scalability

IV. Compute Software

With Loihi 2 came the release of *Lava,* a suite of libraries designed by Intel's Neuromorphic Research Community (INRC) to optimize models to run on neuromorphic hardware [8]. For our application, we used ***Lava-dl*** [9], which enables a process called *ANN-SNN Training,* Fig. 2.

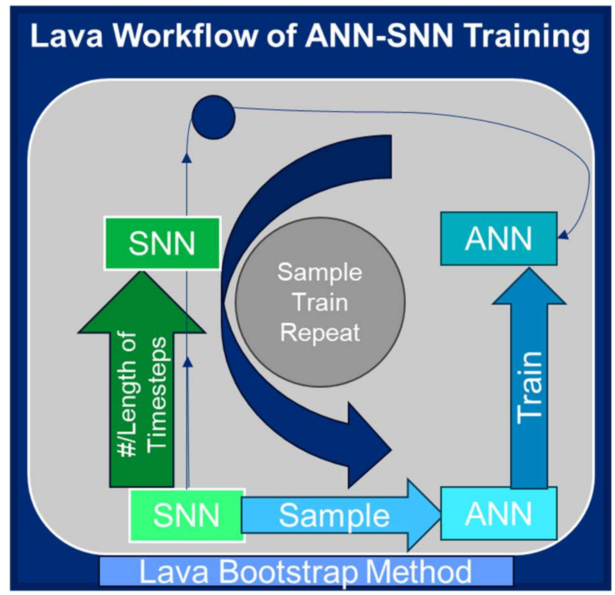

Fig. 2: Workflow Illustration of ANN-SNN Training

To get the best performance from a spiking neural network (SNN), direct training is often the preferred training method. However, directly training SNNs can take a long time and a large amount of compute resources, depending on model size and architecture. One known alternative to this is artificial neural network (ANN) to SNN conversion that takes advantage of rapidly training an ANN and then using a conversion tool to translate that model into an SNN. This process is based on the resemblance of a Leaky Integrate-and-Fire (LIF) neuron to a Rectified Linear Unit (ReLU) transfer function. This means that by using a LIF neuron and integrating over some number of timesteps, we can obtain a good representation of an ANN model trained with ReLU.

Lava-dl's ANN-SNN Bootstrap Training follows a similar workflow: (1) starting with an SNN model, we sample over a few iterations; (2) Bootstrap then dynamically creates a ReLU based ANN, which it trains on; (3) before converting back into an SNN for another round of sampling. Converting to an ANN enables quickly training the model and the sample/train cycle can also improve the gap in performance between an ANN and SNN.

V. Neuromorphic Classification Algorithm

Neuromorphic computing aims at a paradigm shift from Von Neumann-based architectures to distributed and co-integrated memory, the granularity at which this paradigm shift is achieved in digital implementations strongly varies between a distributed Von Neumann or full custom approaches [10, 11, 12]. These custom chip approaches enable the implementation of various algorithms/methods.

Neuromorphic systems hold a critical position in the investigation of novel architectures, as the brain exemplifies an exceptional model for accomplishing scalable, energy-efficient, and real-time embodied computation [13]. It promises to realize artificial intelligence while reducing the energy requirements of computing platforms [14].

The classification algorithm utilized for this research was a Multilayer Perceptron (MLP) network on the representative sensor data. Previously, we successfully completed research and development using a Gradient Boosted Machine (GBM) algorithm on the same dataset with the IBM TrueNorth neuromorphic processor. Different hyperparameters were adjusted on the MLP algorithm to find an optimal setting for maximizing accuracy on predicting the platform's mode from sensor data, e.g., using engine response speed (ERS). Multiple models were created for a given set of packets of sensor data. The packet size was less than 146 bytes total. In this paper, the engine response was grouped into an engine response speed, ERS and a custom data command, Custom_data_1, that included the engines response speed and various other sensors onboard the platform that appeared most correlated. These hyperparameters included the time steps, number of dense neurons, and the number of dense layers in the MLP network. Three models using ERS data and their hyperparameters are described in Table I. Three models using additional data and their hyperparameters are described in Table II.

TABLE I: HYPERPARAMETER SELECTION FOR ERS MODELS

Data Command: Engine Response Speed (ERS)				
Model #	# of Dense Layers	# of Dense Neurons	Timesteps	Accuracy
1	1	128	16	94.0%
2	2	256	16	92.8%
3	1	1024	32	93.8%

TABLE II: HYPERPARAMETER SELECTION FOR CUSTOM_DATA_1 MODELS

Data Command: Custom_data_1			
Model #	# of Dense Layers	# of Dense Neurons	Timesteps
1	1	512	16
2	1	512	32
3	1	1024	16

VI. DATA DESCRIPTION

The data used in this paper was from a representative small research-based platform equipped with various sensors. The sensor data included key sensory information from. Example data includes engines and their speed, sensor vector information (e.g., direction to the object of interest), global positioning system (GPS) data, magnetometer data, etc. The sensor data was grouped by information packets provided to the compute system.

VII. PROCESSING APPROACH

In order to quickly experiment and refine various models on the Loihi 2, the MLP architecture was chosen as our base model. Because *Lava-dl* follows the block format similar to PyTorch *ModuleList*, it is easy to configure the model architectures used for training and testing. Two methods of feeding the data to the algorithms were tested in this paper: One method, which was previously tested with PyTorch models trained on central processing unit (CPU), consisted of tiling the commands into 28 x 28 pixel squares with zero padding; the other method consisted of feeding the command data, which is an array of binary inputs of shape (320x1) directly to the models, without preprocessing or augmentation.

VIII. RESULTS

One interesting observation made when testing different models is that the *Lava-dl* models trained with augmented data, which meant each input was reshaped into a 28 x 28 chip and zero-padded, resulted in lower accuracies (an average of 77% across all models) than the baseline PyTorch CPU models (average of 90% across all models) trained on the same data.

When these *Lava-dl* models were trained directly on raw data, there was a significant increase in performance. Table III shows the comparison in accuracies between the *Lava-dl* models trained on two different inputs.

TABLE III: LAVA-DL MODEL ACCURACIES ON RAW VS. AUGMENTED DATA

Data Command: Custom_Data_1				
# of dense layers	# of dense neurons	Timesteps	Accuracy on raw data	Accuracy on augmented data
1	512	16	97.6%	78.8%
1	512	32	93.6%	78.86%
1	1024	16	92.4%	79.0%

This behavior could be due to the nature of data itself. In a conventional neural network classifier, the expected input is often an image, and the network extracts features from the pixels of a region in the image. Even when these images are converted into spiking data for SNN's, a scaling factor is used to convert them into graded spikes, giving the models adequate information for classification.

The data used in this paper is not an image but a binary sequence representing the state of a sensor. Augmenting this binary sequence could affect the way the model learns.

Tables IV and V show the accuracies of the fine-tuned models trained on raw data. In this case we see a significant improvement in results.

TABLE IV: ACCURACIES OF MODELS TRAINED ON ERS

Data Command: Engine Response Speed (ERS)				
Model #	# of Dense Layers	# of Dense Neurons	Timesteps	Accuracy
1	1	128	16	94.0%
2	2	256	16	92.8%
3	1	1024	32	93.8%

TABLE V: ACCURACIES ON MODELS TRAINED ON CUSTOM_DATA_1

Data Command: Custom_Data_1				
Model #	# of Dense Layers	# of Dense Neurons	Timesteps	Accuracy
1	1	512	16	97.6%
2	1	512	32	93.6%
3	1	1024	16	92.4%

IX. CONCLUSION AND FUTURE RESEARCH

More effective, efficient, and pervasive use of sensing systems such as sensors for smart buildings or cities, disaster surveying drones (where ultra-low power compute is needed) is more realizable than ever before. In addition, the Loihi 2 research chip can use the output of event-based sensors, such as DVS cameras, for further utilization [15]. Additionally, the system could provide command and control capabilities that enable various levels of autonomy and system management/utilization for building security. Applications include completing training operations with SNNs, where innovative new spike-based backpropagation [16], and/or complex spatial temporal timing is learned [17, 18].

Additional research illustrates the potential of neuromorphic sensing and processing for enabling smaller, more intelligent robots [19]. Neuromorphic systems ,such as this, can be used to calculate the short time Fourier transform of a signal and compute the optical flow of visual data with significant savings in computational cost compared to conventional approaches – applying such methods to various sensor and sensor data processing systems represents additional future work [20].

ACKNOWLEDGMENT

The authors would like to thank Dan Sherlock for his guidance, expertise, and technical contributions.

REFERENCES

[1] M. Barnell, C. Raymond, M. Wilson, D. Isereau, and and C. Cicotta, Target Classification in Synthetic Aperture Radar and Optical Imagery Using Loihi Neuromorphic Hardware, 2020, 1-6. 10.1109/HPEC43674.2020.9286246.

[2] M. Barnell, C. Raymond, C. Capraro, D. Isereau, C. Cicotta and N. Stokes, "High-Performance Computing (HPC) and Machine Learning Demonstrated in Flight Using Agile Condor®," 2018 IEEE High Performance extreme Computing Conference (HPEC), Waltham, MA, USA, 2018, pp. 1-4, doi: 10.1109/HPEC.2018.8547797.

[3] M. Barnell, C. Raymond, C. Capraro and D. Isereau, "Agile Condor: A scalable high performance embedded computing architecture," 2015 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 2015, pp. 1-5, doi: 10.1109/HPEC.2015.7322447.

[4] D. Isereau, C. Capraro, E. Cote, M. Barnell, and C. Raymond, Courtney, Utilizing High-Performance Embedded Computing, *Agile Condor*, for Intelligent Processing: An Artificial Intelligence Platform for Remotely Piloted Aircraft. 1155-1159. 10.1109/IntelliSys.2017.8324277, 2017.

[5] Intel Technology Brief, "Taking Neuromorphic Computing to the Next Level with Loihi 2", https://download.intel.com/newsroom/2021/new-technologies/neuromorphic-computing-loihi-2-brief.pdf

[6] M. Davies et al., "Loihi: A Neuromorphic Manycore Processor with On-Chip Learning," in IEEE Micro, vol. 38, no. 1, pp. 82-99, January/February 2018, doi: 10.1109/MM.2018.112130359.

[7] C. -K. Lin et al., "Programming Spiking Neural Networks on Intel's Loihi," in Computer, vol. 51, no. 3, pp. 52-61, March 2018, doi: 10.1109/MC.2018.157113521.

[8] Lava: A software framework for neuromorphic computing. https://github.com/lava-nc/lava, 2021.

[9] Lava DL: https://github.com/lava-nc/lava-dl, 2021.

[10] C. Frenkel, D. Bol, and G. Indiveri, Bottom-Up and Top-Down Neural Processing Systems Design: Neuromorphic Intelligence as the Convergence of Natural and Artificial Intelligence. CoRR abs/2106.01288, 2021.

[11] Cristina Silvano, Daniele Ielmini, Fabrizio Ferrandi, Leandro Fiorin, Serena Curzel, Luca Benini, et al. "A Survey on Deep Learning Hardware Accelerators for Heterogeneous HPC Platforms." arXiv preprint arXiv:2306.15552, 2023.

[12] Y. S. Yang and Y. Kim, "Recent Trend of Neuromorphic Computing Hardware: Intel's Neuromorphic System Perspective," 2020 International SoC Design Conference (ISOCC), Yeosu, Korea (South), 2020, pp. 218-219, doi: 10.1109/ISOCC50952.2020.9332961.

[13] J. Yik, et al. "NeuroBench: Advancing neuromorphic computing through collaborative, fair and representative benchmarking." arXiv preprint arXiv:2304.04640, 2023.

[14] K. Roy, A. Jaiswal and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing", *Nature*, vol. 575, pp. 607-617, 2019.

[15] R. Massa, A. Marchisio, M. Martina and M. Shafique, "An Efficient Spiking Neural Network for Recognizing Gestures with a DVS Camera on the Loihi Neuromorphic Processor," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-9, doi: 10.1109/IJCNN48605.2020.9207109.

[16] Amar Shrestha, Haowen Fang, Daniel Patrick Rider, Zaidao Mei and Qinru Qiu, "In-hardware learning of multilayer spiking neural networks on a neuromorphic processor." 2021 58th ACM/IEEE Design Automation Conference (DAC). IEEE, 2021.

[17] Haowen Fang, Brady Taylor, Ziru Li, Zaidao Mei, Hai (Helen) Li, Qinru Qiu, et al. "Neuromorphic algorithm-hardware codesign for temporal pattern learning." 2021 58th ACM/IEEE Design Automation Conference (DAC). IEEE, 2021.

[18] Fang, Haowen, Amar Shrestha, and Qinru Qiu. "Multivariate time series classification using spiking neural networks." 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020.

[19] Federico Paredes-Vallés, Jesse Hagenaars, Julien Dupeyroux, Stein Stroobants, Yingfu Xu, Guido de Croon, "Fully neuromorphic vision and control for autonomous drone flight." arXiv preprint arXiv:2303.08778, 2023.

[20] Garrick Orchard, E. Paxon Frady, Daniel Ben Dayan Rubin, Sophia Sanborn, Sumit Bam Shrestha, et al. "Efficient neuromorphic signal processing with loihi 2." 2021 IEEE Workshop on Signal Processing Systems (SiPS). IEEE, 2021.