Anomaly Detection with Spiking Neural Networks

Eric Moreno

IRIS-HEP Fellowship Proposal: July-August 2021

Mentors: Dr. Maurizio Pierini (CERN), Dr. Jean-Roch Vlimant (Caltech)

Project Keywords: Innovative Algorithms, Neuromorphic Computing, Anomaly Detection

1 Proposal

In recent years, there has been a revolution in Machine Learning (ML) techniques that can be attributed to the large increase in computational power, mainly harnessing the efficiency for matrix multiplication and convolution granted by Graphical Processing Units (GPUs). The most widespread of these techniques was the Artificial Neural Network (ANN), whose architectural principles were inspired by information processing and distributed communication in biological systems. Even though ANNs were created to mimic brain-like calculations, facilitating communication between artificial neurons and performing calculations inside these neurons, neuroscientists were quick to point out that biological systems perform completely differently from those proposed by any current type of ANN. As such, Artificial Intelligence (AI) has not reached the truly limitless potential that was once promised. Neuromorphic hardware and Spiking Neural Networks (SNNs) [1], which behave like the human brain, aim to fix this problem, bridging the gap between artificial and biological intelligence. The inherent strengths of these SNNs would be extremely useful at the Large Hadron Collider (LHC) with their need for fast inference and accurate data-processing of petabytes of time-series events.

SNNs use membrane potentials to mimic the behavior of organic information processing systems like the human brain. This architecture computes using asynchronous spikes that signal the occurrence of an event by temporally precise action potentials, as demonstrated in Fig. 1. As such, these individual neurons with action potentials and asynchronous spikes are difficult to simulate on traditional hardware, which led to the development of neuromorphic computing hardware like the Loihi research chip[1]. Neuromorphic hardware contains neurons that are optimized for SNNs and because of this, the most promising SNN results come from dedicated neuromorphic hardware. This combination of SNN algorithms and neuromorphic hardware results in favorable properties exhibited in real neural circuits like brains, such as analog computation, low power consumption, fast inference, event-driven processing, online learning, and massive parallelism. These properties could be the key to unlocking the intelligence that ANNs promised to give but failed to deliver on. Importantly, SNNs are inherently temporal architectures with the timing of neuron firings being directly responsible for the computation of the model. As such, they excel with time-series datasets, achieving state-of-the-art performance [2] on datasets like Permuted Sequential MNIST (PS-MNIST) [3] where data is fed in as a function of time. Naturally, the next application for these SNN architectures is at the frontiers of ML experimentation, at physics experiments like the LHC.

I propose the development and optimization of SNN models using neuromorphic training algorithms in real-physics use cases, like detection of anomalous events at the LHC, where protons are collided a billion times per second at extreme energies to create new fundamental particles. We will build on previous research completed by a team from Intel Labs, CERN, and myself, where we demonstrated the ability to detect gravitational-wave in simple single-detector time-series data at the Laser Interferometer Gravitational-Wave Observatory (LIGO) using recurrent autoencoders (paper submitted to Machine Learning: Science and Technology). The continuation of this project at the LHC involves more complex time-series data from multiple detector channels. As ML algorithms are the cutting edge of data processing at CERN, the development of this project will extend experimentation on the strengths of SNN models with respect to their classical ANN counterparts, enabling scientists at CERN to identify use cases for the powerful algorithms in their detectors. We will compare the SNN anomaly detection algorithms with proven algorithms, such as the Variational Autoencoder [4].

[1] `https://www.intel.com/content/www/us/en/research/neuromorphic-computing.html`

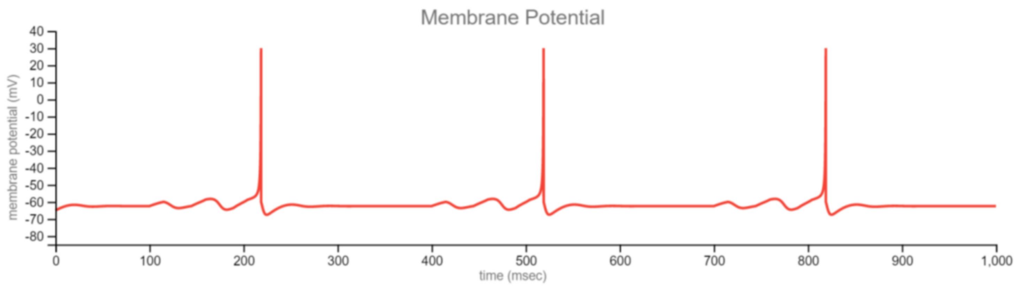

Figure 1: Membrane potential of a single spiking neuron. The neuron fires when the action potential achieves a specific threshold. After activation, the signal returns to the regular low potential.

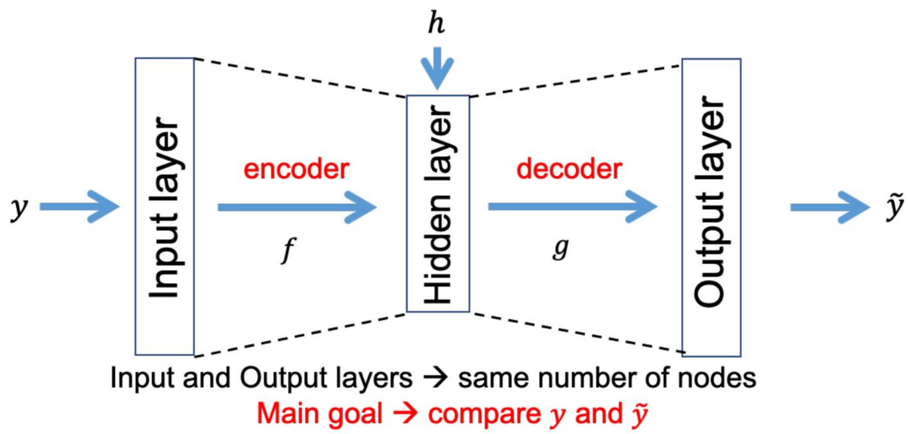

Figure 2: Illustration of an autoencoder architecture with generic nodes. For the proposed project, the encoder and decoder will be multi-layered Legendre Memory Units (LMU), a type of spiking recurrent cell. In comparing y and $\tilde{y}$, one can infer how much the given input differs from normal QCD noise, which can signal an anomalous detection.

This project will assist in improving real-time and offline anomaly detection algorithms at the LHC. At the LHC, data are collected at 40 MHz but only 1 kHz of data can be stored for physics studies. This is acceptable since the majority of collisions in the LHC are known physics which needs no further studies. Scientists search for needle-in-a-haystack events to study which contain anomalous physics. A typical LHC experiment operates a real-time selection L1 trigger system, that decides if an event should be stored or discarded. A L1 algorithm needs to operate within O(1 μsec) latency, which makes it an ideal selection for fast-inference neuromorphic hardware like Loihi. Additionally, offline searches (where accuracy is preferred) for anomalous events could also benefit from the inherent time-series nature of SNNs. This project will aim to prove that the LHC could benefit from an unsupervised SNN algorithms designed to identify outlier events, possibly highlighting the occurrence of new phenomena in LHC collisions. For this purpose, I will design a spiking recurrent autoencoder, similar to our previous LIGO gravitational-wave identification algorithm, which processes particle four momenta. Since the autoencoder training does not depend on any specific new physics signature, the proposed procedure does not make specific assumptions on the nature of new physics. The event selection based on this SNN algorithm would be complementary to classic LHC searches, delivering a list of previously unidentified anomalous events that the experimental collaborations could further scrutinize and even release as a catalog.

The SNN autoencoder would be trained to represent normal simulated QCD noise in a compressed latent space and decompress this encoded data as represented in Fig. 2. Specifically, the encoder and decoder cells will use the Legendre Memory Unit (LMU) [2], which is a spiking alternative to the Recurrent Neural Networks like Long-short memory networks (LSTMs) [5] and Gated Recurrent Unit (GRUs) [6]. As a part of the Intel Neuromorphic Research Community (INRC), I can access large-scale neuromorphic systems that can be used for testing and evaluation. To train the SNN autoencoder models on GPUs will require special backpropagation algorithms, SLAYER [7], which can overcome the problem of non-differentiability of the spike functions in the SNN. Additionally, evolutionary optimization (EO) algorithms for SNNs [8] can be used to further optimize models that are already trained.

2 Timeline

The timeline I propose is a total of three months, during which I will devote 100% of the time to the project. I will have no academic, teaching, or other research obligations during this time.

- Month 0-1: Design and implement LMU autoencoder toy model on LIGO GW data. This simple dataset will help to troubleshoot problems with the SNN autoencoder architecture and SNN training (back-propagation) before moving to more difficult LHC datasets. Deliverable: Functioning LMU autoencoder architecture.
- Month 1-1.5: Generate simulated dataset corresponding to LHC (specifically CMS) operating conditions with normal QCD events and beyond-standard-model (BSM) events which will be used to train and evaluate the LMU autoencoder. Deliverable: Simulated CMS dataset.
- Month 1.5-3: Modify and train LMU autoencoder to run on the simulated CMS dataset. Refine model using different spiking back-propagation algorithms. Optimize these models using evolutionary optimization algorithms. Deliverable: Full neuromorphic anomaly detection algorithm for CMS.

References

[1] M. Pfeiffer and T. Pfeil, *Deep learning with spiking neurons: Opportunities and challenges*, *Frontiers in Neuroscience* **12** (2018) 774.

[2] A. Voelker, I. Kajić, and C. Eliasmith, *Legendre memory units: Continuous-time representation in recurrent neural networks*, in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, et al., eds.), vol. 32, pp. 15570–15579, Curran Associates, Inc., 2019.

[3] Q. V. Le, N. Jaitly, and G. E. Hinton, *A simple way to initialize recurrent networks of rectified linear units*, *CoRR* **abs/1504.00941** (2015).

[4] O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant, *Variational Autoencoders for New Physics Mining at the Large Hadron Collider*, *JHEP* **05** (2019) 036, [`arXiv:1811.10276`].

[5] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural computation* **9** (12, 1997) 1735–80.

[6] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, *Empirical evaluation of gated recurrent neural networks on sequence modeling*, *CoRR* **abs/1412.3555** (2014) [`arXiv:1412.3555`].

[7] S. B. Shrestha and G. Orchard, *Slayer: Spike layer error reassignment in time*, in *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, et al., eds.), vol. 31, pp. 1412–1421, Curran Associates, Inc., 2018.

[8] C. D. Schuman, J. S. Plank, A. Disney, and J. Reynolds, *An evolutionary optimization framework for neural networks and neuromorphic architectures*, in *International Joint Conference on Neural Networks*, (Vancouver), July, 2016.