ORNL/SPR-2024/3423

OR22-Neuromorphic Rad Detector-PD3Ra Final Report

Approved for public release.
Distribution is unlimited.

James Ghawaly[1]
Andrew Nicholson[2]
Cathrine Schuman[3]
Mathew Swinney[2]
Brett Witherspoon[2]
Aaron Young[2]

September 2024

[1] Louisiana State University
[2] Oak Ridge National Laboratory
[3] University of Tennessee, Knoxville

OAK RIDGE
National Laboratory

ORNL IS MANAGED BY UT–BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

DOCUMENT AVAILABILITY

Online Access: US Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via https://www.osti.gov/.

The public may also search the National Technical Information Service's National Technical Reports Library (NTRL) for reports not available in digital format.

DOE and DOE contractors should contact DOE's Office of Scientific and Technical Information (OSTI) for reports not currently available in digital format:

US Department of Energy
Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831-0062
Telephone: (865) 576-8401
Fax: (865) 576-5728
Email: reports@osti.gov
Website: https://www.osti.gov/

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORNL/SPR-2024/3423

Physics Division

OR22-NEUROMORPHIC RAD DETECTOR-PD3RA FINAL REPORT

James Ghawaly[1]
Andrew Nicholson[2]
Cathrine Schuman[3]
Mathew Swinney[2]
Brett Witherspoon[2]
Aaron Young[2]

September 4, 2024

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADC	analog-to-digital converter
ANN	artificial neural network
ARAD	autoencoder radiation anomaly detection
COTS	commercial-off-the-shelf
CSA	charge-sensitive amplifier
Deffe	Data-Efficient Framework for Exploration
DNN	deep neural network
DSP	digital signal processing
EONS	Evolutionary Optimization of Neuromorphic Systems
FAR	false alarm rate
FDA	fully differential amplifier
FIFO	first in, first out
FP	false positive
FPGA	field-programmable gate array
HEU	highly enriched uranium
I^2C	Inter-Integrated Circuit
IC	integrated circuit
ICONS	International Conference on Neuromorphic Systems
IF	integrate-and-fire
IP	internet protocol
LBNL	Lawrence Berkeley National Laboratory
LIF	leaky integrate-and-fire
MCA	multichannel analyzer
MCU	microcontroller unit
MUSE	Modeling Urban Search Environments
NORM	Naturally Occurring Radioactive Material
OLED	organic light-emitting diode
ORNL	Oak Ridge National Laboratory
PCA	principal components analysis
PCB	printed circuit board
PD	probability of detection
PPV	positive predictive value
ROC	receiver operating characteristic
ROI	Region of Interest
RTOS	real-time operating system
SAD	spectral anomaly detector
SDK	software development kit
SiPM	silicon photomultiplier
SNN	spiking neural network
SNR	signal-to-noise ratio
SPI	serial peripheral interface
SPRT	sequential probability ratio test
STDP	spike-time dependent plasticity
TP	true positive
TPR	true positive rate
TTAC	Technical Testing and Analysis Center
UART	universal asynchronous receiver and transmitter
UTK	University of Tennessee, Knoxville
WGPu	weapons-grade plutonium

1. INTRODUCTION

In unattended monitoring scenarios, automated radiation detection algorithms must be able to detect low signal-to-noise ratio (SNR) anomalies in a potentially dynamic and noisy background and report these anomalies in a timely fashion. Dynamic and noisy backgrounds complicate the use of simple gross-counting algorithms because they can lead to either high false positive rates or low sensitivity. Algorithms that use the entire spectrum have been the most successful in this area; notable examples are the NSCRAD algorithm [1] developed at Pacific Northwest National Laboratory and recently the nonnegative matrix factorization approach [2] developed at Lawrence Berkeley National Laboratory (LBNL). These approaches use either spectral regions of interest or spectral decomposition to detect threat isotopes in the background.

Systems that are used for these applications are often composed of one or more radiation detectors that generally have spectroscopic capabilities, one or more computers that reads data from the sensor(s) and feeds data to a radiation anomaly detection algorithm (which may also perform isotope identification), and a rechargeable battery system for powering the device while it is deployed. Of these components, the computer(s) responsible for ingesting the radiation data and running it through an anomaly detection algorithm consumes the vast majority of the available power. As an example, the Bridgeport SiPM-3000 silicon photomultiplier array and multichannel analyzer designed for operating a scintillation detector typically consumes about 0.3 W of power [3], whereas the NVIDIA Jetson Xavier NX computing platform, which is aimed at edge computing applications, can consume between 7.5 and 15 W of power [4].

During the past decade, research has demonstrated tremendous success in using advanced machine learning algorithms for radiation anomaly detection, as was demonstrated in the deep neural network (DNN) R&D sponsored “Detecting Radiological Threats in Urban Areas” algorithm competition held on the Topcoder platform [5, 6, 7]. Unfortunately, these advanced algorithms generally come at the cost of increased computational demand, which in turn comes at the cost of increased energy consumption [8, 9]. This inverse relationship between algorithm performance and energy consumption limits the potential capabilities of these systems and forces their designers to make compromises in their design. By contrast, neuromorphic computing enables the implementation of these powerful machine learning methods without the extreme power consumption because units communicate using discrete rather than continuous signals [9, 10].

Neuromorphic technology is a cutting-edge field that promises to significantly revolutionize the ways in which advanced machine learning algorithms are designed and implemented across broad application spaces. In particular, neuromorphic computing enables traditionally power-hungry algorithms to be implemented in this new paradigm that significantly reduces their power consumption while reducing the form factor of the computing hardware. This project is the first formal investigation into the use of neuromorphic computing for improving current radiation anomaly detection methods in the proliferation detection space, rendering this effort both novel and potentially revolutionary in its impact on operations. Aside from the energy-saving advantages that neuromorphic computing provides, spiking neural networks (SNNs), which are the primary neuromorphic algorithm architecture, also have certain characteristics that enable them to be highly effective and efficient at processing spatiotemporal data streams [9, 11]. This capability may improve the neuromorphic algorithm’s performance relative to traditional algorithm methodologies that do not inherently process data temporally.

This report details the two-year rapid development, construction, and testing of neuromorphic SNN algorithms running on an embedded radiation detection system consisting of a NaI(Tl) gamma-ray detector and a combination of commercial-off-the-shelf (COTS) and custom electronics. Accomplishing the project goals involved leveraging the previous development of the TENNLab neuromorphic computing framework [12], the Caspian neuromorphic simulator, and the field-programmable gate array (FPGA)-based μCaspian neuromorphic hardware platform [13, 14] from Oak Ridge National Laboratory (ORNL) and University of Tennessee, Knoxville (UTK). The following sections describe the embedded hardware that was developed for this project and the data generated to train and evaluate algorithms, detail algorithm development, and evaluate the system with neuromorphic algorithms integrated with hardware in an experimental campaign.

2. HARDWARE AND SOFTWARE DEVELOPMENT

This section details the hardware design and power consumption of the embedded hardware system developed in this project, shown in Figure 1. This system was custom-designed to demonstrate the ultra low-power capabilities of neuromorphic computing for autonomous radiation monitoring. Two custom printed circuit boards (PCBs) were designed to integrate a low-power FPGA to host the neuromorphic processor, an embedded multichannel analyzer (MCA), a silicon photomultiplier (SiPM)-based NaI(Tl) scintillation detector, and an ultralow-power microcontroller. This hybrid design takes advantage of digital communications and traditional computing architectures when more appropriate while benefiting from the ultra low-power custom neuromorphic implementation for the primary processing component. The scope of this project was limited to optimizing the power consumption of the computing subsystem only. The SiPM and MCA were chosen for their low power use, comparable to the computing subsystem power target, but were not specifically optimized in this project.

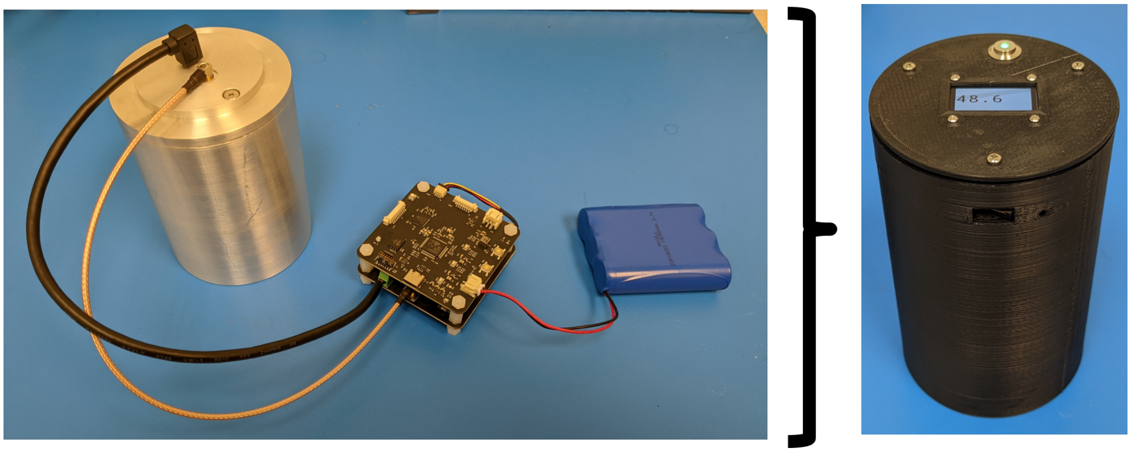


Figure 1. The embedded neuromorphic platform, detector, and casing.

2.1 DETECTOR SIGNAL PROCESSING

This project focused primarily on developing the hardware for the low-power neuromorphic computing subsystem and leveraged COTS components for the MCA and NaI(Tl) SiPM detector. However, the COTS components were selected to be suitable for an embedded system and a reasonable power budget. Most commercial MCAs are designed for communications and power over USB or Ethernet. An early power analysis showed that both USB and Ethernet required significantly more power than protocols such as a universal asynchronous receiver and transmitter (UART) or serial peripheral interface (SPI), which are more commonly found in low-power microcontroller-based designs. The GBS Elektronik GmbH MCA527 nano MCA was chosen for this project because it uses a lower power UART communications interface and has a smaller footprint than other options. The MCA527 nano uses the same digital signal processing (DSP) and signal processing firmware as the larger desktop variant but does not include the power supply electronics; therefore, it allows better integration with the high-efficiency power supply of the host system and lower overall power consumption.

Following is a description of the signal processing from radiation sensor to neuromorphic processing to provide background information for the hardware description sections. The NaI(Tl) scintillator emits photons as a result of incident gamma radiation. The SiPM then induces a photocurrent pulse that is integrated by a charge-sensitive amplifier (CSA) and then digitized with a sampling rate of 10 MHz.

The peak voltage observed in the integrated pulse is proportional to the energy of the incident radiation, and a spectroscopic filter is used to extract these peaks from the digitized waveform. The peaks are accumulated into a histogram of pulse heights to construct the gamma-ray spectrum observed by the detector.

Each bin of the spectra represents a range of energies determined by a calibration to account for variations in the detector and the resolution of the histogram. The microcontroller queries the MCA for the current gamma spectrum, inserts it into a moving window sum, applies a calibration, and encodes the result into spikes. This spike sequence is then transferred to the input first in, first out (FIFO) of the neuromorphic coprocessor for inference. After the neuromorphic coprocessor has processed the encoded spike train, the microcontroller reads output spikes from the FIFO and applies a threshold to the spike rate to detect an anomaly.

2.2 CUSTOM HARDWARE

The custom hardware consists of two printed circuit boards: (1) a mixed-signal carrier board that integrates COTS detector components and (2) the "Nuclide" board, which is the general-purpose computing and neuromorphic processing system. Figure 2 shows a block diagram of the system. The mixed-signal carrier board was designed to preprocess the analog signal from the NaI(Tl) SiPM detector and drive the analog-to-digital converter (ADC) of the digital MCA, which estimates and then communicates the gamma-ray spectra to the main processing board. A low-power fully differential amplifier (FDA) was developed to drive unbuffered input of the ADC and incorporates antialiasing and out-of-band filtering to increase the SNR of the signal of interest before digitization. The carrier board also contains power electronics required to supply the various voltages for the MCA's DSP and the SiPM detector from the single cell Li-ion battery design of system.

The "Nuclide" board utilizes a low-power STM32U5 based on the ARM Cortex-M33 core, a Lattice iCE40 UltraPlus FPGA, lithium-ion battery charger, and high-efficiency power supplies. This board was designed to also support other applications and projects related to neuromorphic computing and has been instrumented to enable power measurements of the various subsystems. Board-to-board style connectors are available to extend the functionality of the board with any peripheral that has a UART, SPI, or Inter-Integrated Circuit (I^2C) digital interface, or custom interface logic can be used on the FPGA. These connectors are currently being used for the organic light-emitting diode (OLED) display and a wireless communications module based on the Nordic Semiconductor nRF52840 transceiver. Figure 3a shows the printed MCA carrier board, and Figure 3b shows the custom microcontroller unit (MCU) and FPGA board.

An STM32U5 series microcontroller from STMicroelectronics functions as the host for the neuromorphic coprocessor and has a maximum core clock frequency of 160 MHz with many low-power states and other power-management features. The microcontroller manages numerous peripherals, including an external 512 MB NAND storage for data, a temperature and humidity sensor, various power supply controls, a multifunction USB interface, and several board-to-board connectors used for extensions such as a wireless communications module and display. In the system's lowest power mode, almost all of these peripherals are clock gated if they are integrated inside the microcontroller or power gated via switches if they are external.

The previously developed µCaspian neuromorphic core functions as a coprocessor and is implemented on the low-power Lattice iCE40 FPGA [13, 14]. The microcontroller loads the FPGA image via a SPI. After configuration, the same interface is used to configure and control the neuromorphic core. The µCaspian core is a digital event-driven processor operating on spikes occurring in the SNN. Thus, the processor's runtime is directly proportional to the number of input spikes and the resulting internal spikes that occur. Memory regions are only accessed in response to a spike event, which in turn causes information about the neuron's properties and synapse routing to be accessed.

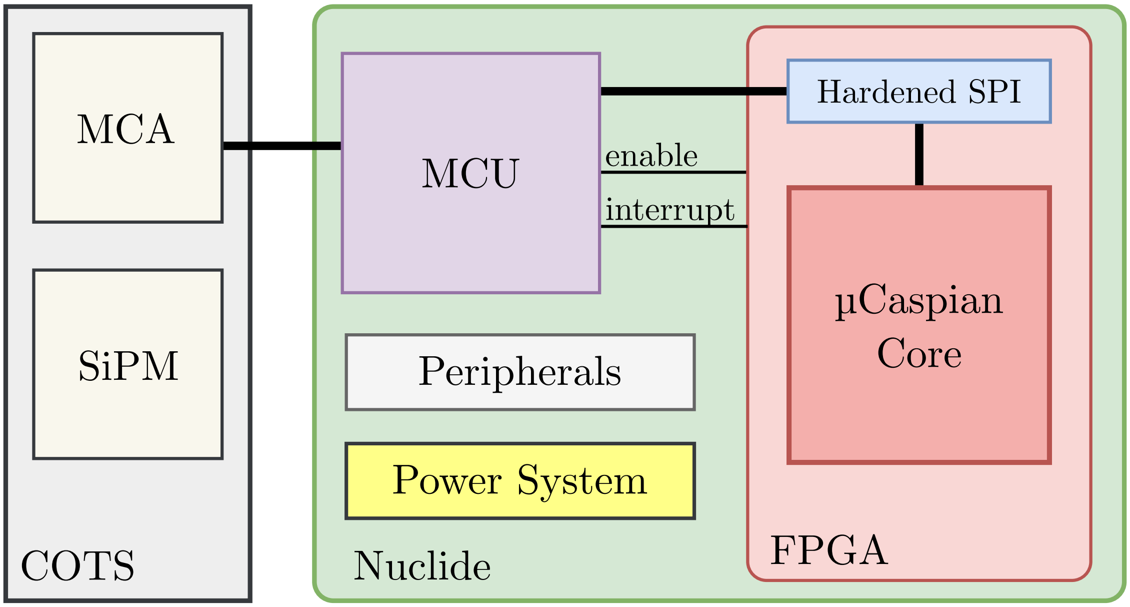
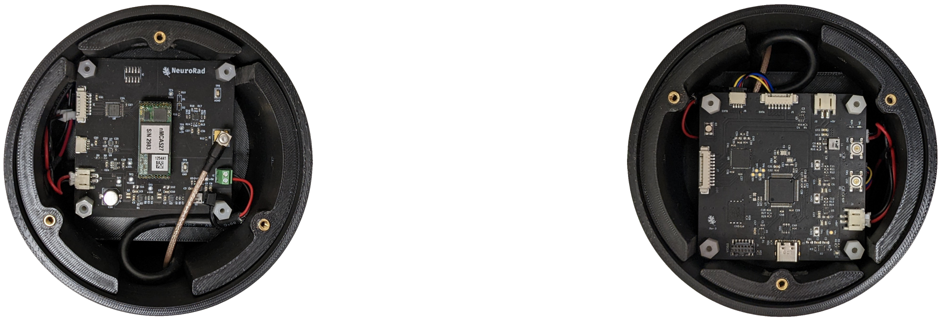


Figure 2. A block diagram illustrating the components of the detection system.



(a) Custom MCA carrier board with supporting power electronics.

(b) Custom PCB for the MCU and FPGA with lithium-ion battery power supply.

Figure 3. Custom printed circuit boards developed in this project.

Recent additions support lower-power operation when integrated with the microcontroller host. Most importantly, clock gating was added, significantly reducing power consumption when the system is not processing events. The design primarily achieves energy efficiency by quickly and efficiently evaluating the SNN, and then clock gating while idle. A hardened SPI core available on the FPGA is used for the SPI interface instead of programmable logic, saving both resources and power. Finally, an interrupt request signal and a FIFO buffer were added to allow the core to process all events without waiting and before waking the microcontroller from its low-power sleep state.

Every 500 ms, the microcontroller queries the MCA for the accumulated spectrum, inserts it into a moving window sum, applies a calibration, and encodes this calibrated input spectrum into spikes. This spike sequence is then transferred to the input FIFO of the neuromorphic coprocessor. The microcontroller then enters a low-power sleep state until the coprocessor finishes processing the events and makes an interrupt request to wake up the microcontroller. Once awake, the microcontroller reads the output spikes from the output FIFO and applies a threshold to the spike rate to check for an anomaly.

2.3 SOFTWARE AND FIRMWARE

An extensible software development kit (SDK) was built for developing firmware for the neuromorphic platforms. The SDK is based on the Zephyr real-time operating system (RTOS) and adds support for this custom platform. This support includes device drivers, application code, and host computer utilities for working with the platform. All project software is version controlled using Git and is hosted on ORNL's internal Gitlab instances. The SDK is hierarchical and uses a metatool to automatically clone various internal and external projects into a development workspace. A block diagram of some of the components and the hierarchical design is illustrated in Figure 4.

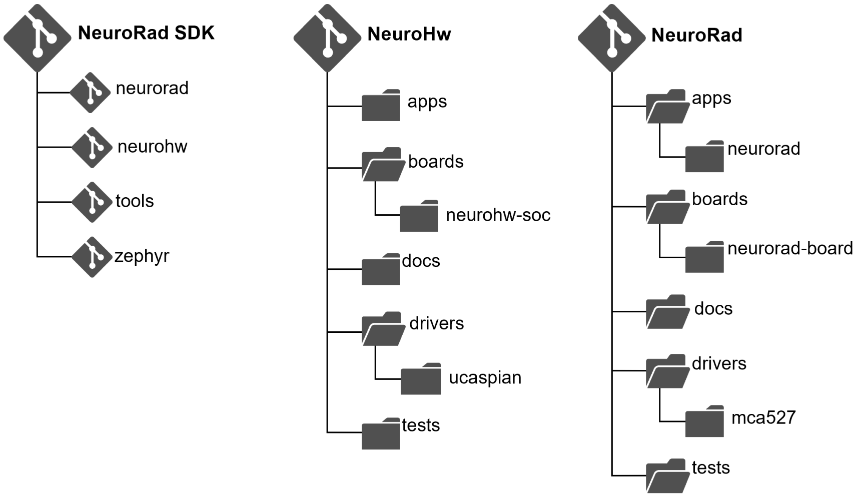


Figure 4. A depiction of the extensible software development kit.

To facilitate software development, a Visual Studio Code editor workspace automates the setup and build process, features a full-featured debugging configuration, and includes a curated list of additional plugins. Documentation is automatically generated from the source code repositories and includes a "getting started" guide.

The microcontroller firmware is based on the Zephyr[1] RTOS and leverages the power management features of the operating system whenever possible. An RTOS offers preemptive priority scheduling capabilities required for real-time computing. Moreover, the main advantages are the device driver abstractions

[1] https://www.zephyrproject.org/

and application-level interfaces that make software easy to port to different hardware platforms. The power supply system was designed to have high efficiency, even with very low output current, and is capable of recharging lithium-ion batteries via USB or solar cells. A full internet protocol (IP) networking stack is also available via USB-emulated Ethernet. The internal flash of the microcontroller has been partitioned into several regions for the bootloader, a primary and secondary firmware image, and various nonvolatile configuration options. A larger 512 MB flash integrated circuit (IC) is used for additional storage and is not removable, so it meets the security requirements of nuclear facilities. However, an optional configuration allows the storage to be easily accessed via the USB interface, which is similar to that of a typical USB storage device. Larger objects such as the FPGA images and neuromorphic network configurations are stored in the higher capacity storage and can be selected at runtime. This storage can also be used to record data that can be read later for further analysis or archival.

2.4 POWER ANALYSIS

Table 1 shows the average power consumption by each subsystem. The computing subsystem typically operates with nonessential peripherals disabled, resulting in the lowest power consumption. However, during laboratory testing, it is useful to enable many of the peripherals to simplify data collection and support debugging. Therefore, both power figures are given in the table. Power consumption is largely determined by the time each component is online; thus, optimizing sleep time for each component is important to reduce overall power consumption. A timing diagram and the runtime for each component for a typical 500 ms periodic interval is shown in Figure 5. The majority of this runtime is for communications between the components, and this time can vary depending on baud or clock rate settings. The total power consumption is largely determined by how long each component can remain in low-power sleep status during each interval.

Table 1. Power consumption by subsystem and components

Subsystem	Components	Typical power (maximum power)
Sensing	SiPM, CSA	70 mW
Sensing	MCA	215 mW
Computing	MCU, FPGA	6 mW (37 mW)
Total		291 mW (322 mW)

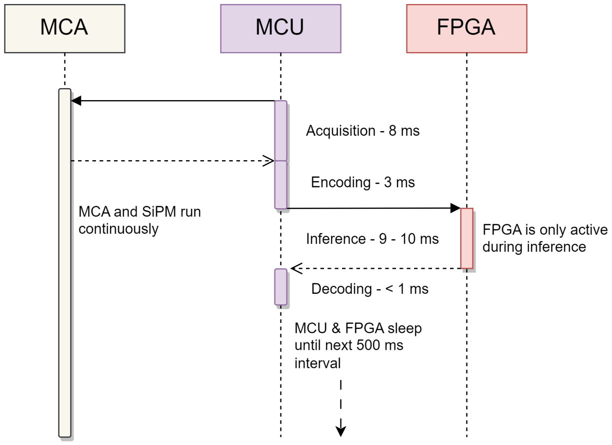

MCA
MCU
FPGA
Acquisition - 8 ms
Encoding - 3 ms
MCA and SiPM run continuously
Inference - 9 - 10 ms
FPGA is only active during inference
Decoding - < 1 ms
MCU & FPGA sleep until next 500 ms interval

Figure 5. A timing diagram of the interaction between the MCA, MCU, and FPGA.

3. TRAINING AND EVALUATION DATA DEVELOPMENT

Comprehensive and realistic labeled radiation datasets are needed in order to develop neuromorphic-capable radiation anomaly detection algorithms. Like other machine learning algorithms, neuromorphic-capable spiking neural networks benefit from large datasets that adequately span the statistical domain of the problem space. Initially, the project used the Top Coder dataset, which has already been developed and peer-reviewed through the DNN R&D Modeling Urban Search Environments (MUSE) program, and was used to hold two data competitions: one among government competitors and another held publicly [5, 6, 7]. This same dataset was used initially to train, test, and evaluate the neuromorphic-capable radiation anomaly detection algorithms.

However, additional data were needed to include a more diverse set of sources and shielding combinations. To this end, a new radiation dataset was created based on Monte Carlo simulations for the ORNL 7000 area. A detailed 3D Monte Carlo radiation transport model was used to generate detector response, for both background and threat sources, for a detector at several positions along a path near a static source. The background and threat source simulations were used to generate time-series synthetic datasets simulating the response to background, and they injected sources to a detector moving past threat sources at varying detector speeds, distance of closest approach, threat source type, threat source activity, and background composition. Synthetic datasets generated by detailed simulations are useful for radiation algorithm development because the ground truth for both background and threat source responses is well known, making labeling straightforward. This section details the radiation transport simulations and how they were used to create time-series datasets for 2 × 4 × 16 in. and 3 × 3 in. NaI(Tl) detectors.

3.1 ORNL 7000 AREA RADIATION TRANSPORT SIMULATIONS

For this project, a radiation transport model was developed for the Shift Monte Carlo code [15] based on the 7000 area of the ORNL campus. The model geometry is shown in Figure 6, compared with a Google Earth image of the environment, representing an area of approximately 1,200 × 600 ft. Nine separate volumetric background sources were attributed to the soil, concrete, and asphalt as appropriate in the model, each with concentrations of 1 Bq/kg for potassium, uranium, and thorium. This methodology enabled the creation of a realistic background using any concentration of Naturally Occurring Radioactive Material (NORM) chosen when generating the synthetic data set.

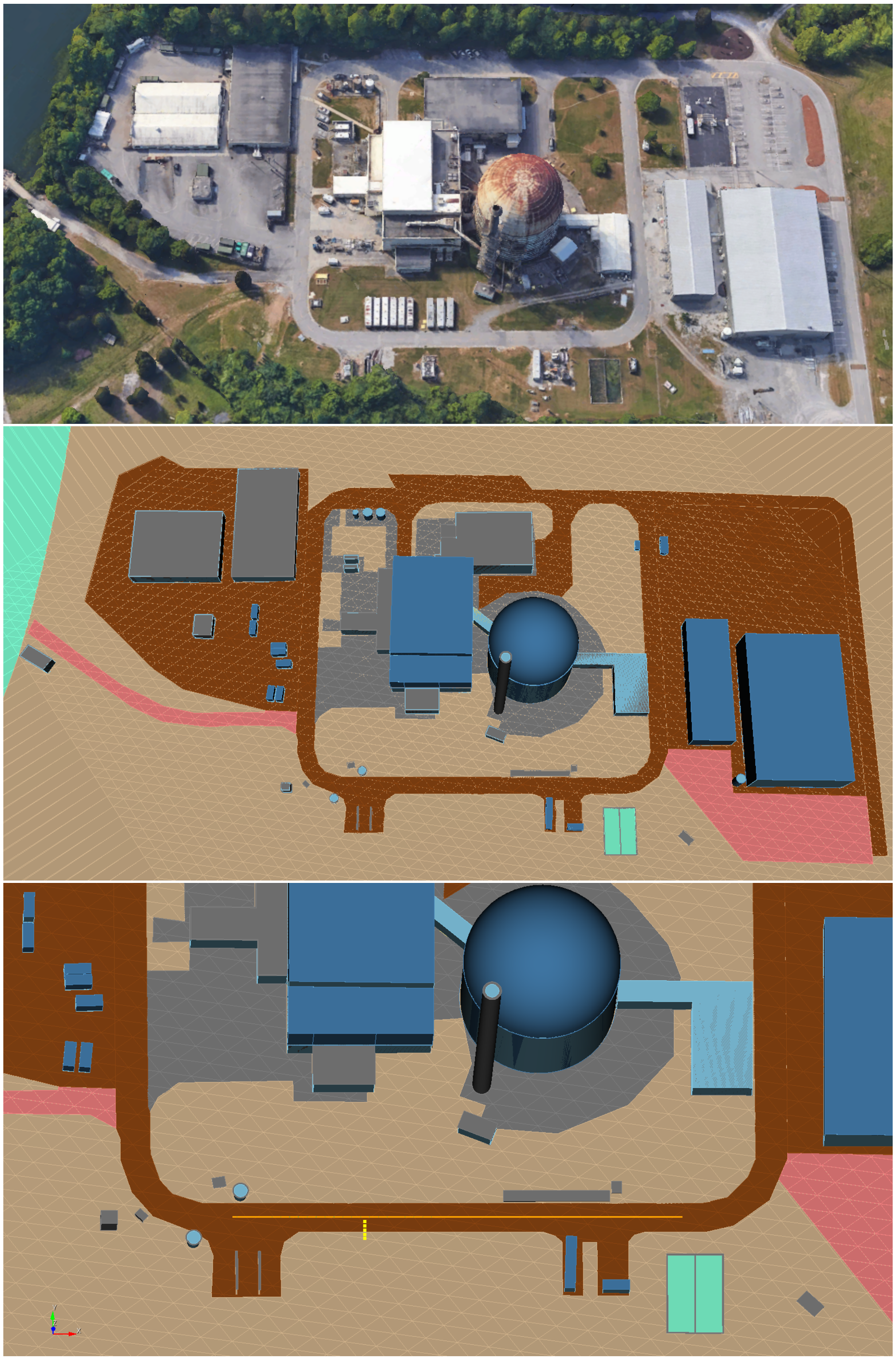


Figure 6. The 7000 area of the ORNL campus. (top) A Google Earth representation of the area and (middle) the corresponding model developed for the Shift Monte Carlo code. (bottom) Source locations (in yellow) and detector path (in orange) for the radiation transport calculations designed to create synthetic data in the Shift ORNL 7000 area model.

The threat sources placed in the Shift 7000 area simulations are shown in Table 2 and were located at standoff distances of 1, 2, 3, 4, and 5 m from the detector path, as shown in the bottom portion of Figure 6. The detector path traveled down the center of the road (as shown) as it approached and passed the source at the appropriate standoff distance. Although the end use case scenario for this system is envisioned as being a static detector–dynamic source system, the static source–dynamic detector scenario is much easier to simulate with the tools developed in previous programs [6, 16] because the dynamic source scenario would require many source simulations along the road. This configuration is expensive to compute compared with one source and one background simulation required for the static source–dynamic detector scenario.

Both threat and background simulations calculated gamma-ray flux (and flux moments) in a 1 m^3 voxelized mesh along the detector path. Calculated mesh flux moments were then processed into pulse-height mesh spectra for a specified detector type in a postprocessing step using a methodology described in Celik et al. [17]. This methodology started with the 2 × 4 × 16 in. detector response function developed in the TopCoder competition but ended with a 3 × 3 in. detector response function developed for this project using characterization data acquired in an experimental characterization.

Table 2. Source list for ORNL 7000 area simulations. Source shielding materials included stainless steel and polymethyl methacrylate.

Isotope	Activity (or mass)	Shielding
^{241}Am	1 µCi	Bare and 1 cm SS
^{133}Ba	1 µCi	Bare and 1 cm SS
^{57}Co	1 µCi	Bare and 1 cm SS
^{60}Co	1 µCi	Bare and 1 cm SS
^{137}Cs	1 µCi	Bare and 1 cm SS
^{177}Lu	1 µCi	Bare and 8 cm PMMA
WGPu	0.5, 1, 2, 4, 8 kg	Bare and 1 cm SS
HEU	1, 2.5, 5, 10, 25 kg	Bare and 1 cm SS

3.2 ORNL 7000 AREA SYNTHETIC DATASET GENERATION

Monte Carlo generated background and threat source pulse-height tallies (units of counts per second). These height tallies were used to generate time-series synthetic datasets that simulate a detector moving down the road by scaling background and threat source mesh voxel pulse-height tallies based on how much time the detector spent in each voxel given its speed and using Poisson sampling to generate realistic random noise. Data were generated in a series of runs, each a single detector passing down the road with a different set of parameters, including detector speed, detector offset from source, threat source type (including background only), threat source SNR, and potassium/uranium/thorium background activities. For these datasets, potassium/uranium/thorium background activities were informed from data measured at the Fort Indiantown Gap facility [18].

For a run with a predefined detector offset, detector speed, and threat source type, corresponding Monte Carlo–generated background and source pulse-height mesh tallies were loaded. Background pulse-height mesh tallies were scaled by chosen background activity concentrations, and source pulse-height mesh tallies were scaled by the chosen source activity. Threat source activities, α, were calculated using the summed threat background pulse-height spectrum at the distance of closest approach, S and B respectively, and the chosen SNR using the following equation:

$$\mathrm{SNR} = \frac{\alpha S}{\sqrt{\alpha S + B}} \rightarrow \alpha = \frac{(S \cdot (\mathrm{SNR})^2) \pm S \cdot \mathrm{SNR}\sqrt{4B + \mathrm{SNR}^2}}{2S^2}. \quad (1)$$

Finally, based on the detector speed and chosen integration time, the time the detector spent in each mesh voxel along the road was used to scale the pulse-height mesh tally to produce an average gamma-ray spectrum (in units of counts) for each voxel along the road. The background and source spectra were then sampled, using Poisson statistics, to generate the random noise a real detector would measure over a short period of time. Finally, the background and source spectra were added together to produce the final dataset. Two example high SNR runs are shown in Figure 7.

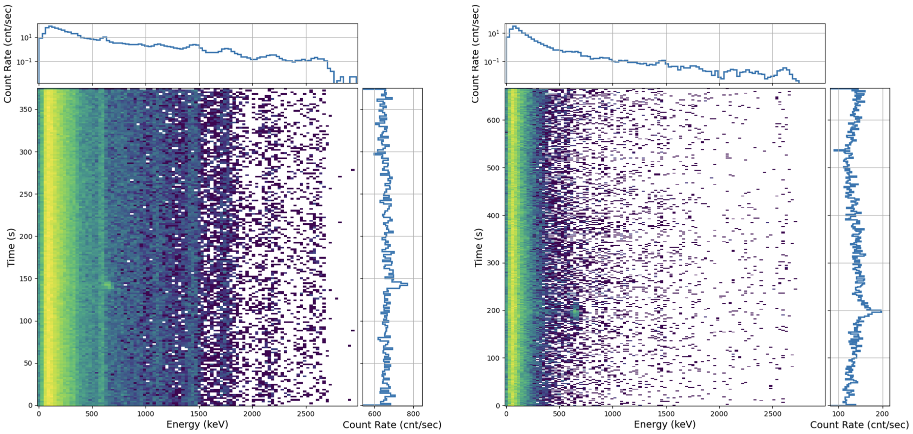


Figure 7. Two sample high-SNR runs from the (left) 2 × 4 × 16 in. NaI(Tl) and (right) 3 × 3 NaI(Tl) detectors.

These generated data were separated into two datasets—a training dataset and a test dataset—to create two independent datasets to train and evaluate networks. This procedure was performed twice in this project: once for a 2 × 4 × 16 in. detector response function developed in the TopCoder competition, and once for a 3 × 3 in. detector response function developed in an experimental characterization campaign using detectors purchased in this project. This project started with the 2 × 4 × 16 in. detector response function, even though this detector is not used in the final system design, because it was available to use while the system with the 3 × 3 in. system was being constructed. Data from the 2 × 4 × 16 in. dataset generally have higher count rates and less absolute variability in gamma-ray spectra because the detector is larger and has a higher efficiency. Parameters used in the 2 × 4 × 16 in. and the 3 × 3 in. training and test datasets are listed in Table 3.

Table 3. Parameters used for 2 × 4 × 16 in. and 3 × 3 in. training and test datasets

Parameter	2 × 4 × 16 in.		3 × 3 in.	
	Training	Testing	Training	Testing
Number of runs	13,350	13,350	29,400	17,460
Total time (days)	29.5	29.5	42.4	54.4
Source types	Table 2		Table 2	
SNR Range	0.5–9.0	0.5–9.0	1.0–10.0	1.0–10.0
Detector offsets (m)	1, 3, 5	1, 3, 5	1, 2, 3	1, 3, 4
Detector speeds (m/s)	0.3, 1.4, 2.5	0.3, 1.4, 2.5	0.5, 0.95, 1.4	0.15, 1.1, 2.0

4. NEURMOPRHIC ALGORITHM DEVELOPMENT

4.1 INTRODUCTION

The objective of the algorithm development is to develop an efficient and effective radiation anomaly detection algorithm that can be used as a trigger (tip and queue system) for activating higher-power/lower-efficiency algorithms and sensors. Neuromorphic algorithms take inspiration from the human brain and, more generally, biological neural networks to create artificial neural network (ANN) models with high computational and energy efficiency. The type of neural network methodology used in neuromorphic systems is called an SNN, named after the discrete "spiking" behavior of neurons in biological systems. In an SNN, each neuron is treated as a time-dependent dynamic system, better expressed as an ordinary differential equation where the neuron's response depends not only on the input to the neuron but also on the previous state of the neuron itself. Many models of spiking neurons exist, but this project focused on the two most efficient models, the integrate-and-fire (IF) neuron and the leaky integrate-and-fire (LIF) neuron. These two models are indeed highly efficient, and they are the only models supported by the Caspian neuromorphic hardware architecture.

Training SNNs is notoriously challenging, at least when compared with the simplicity and reliability of gradient descent via back propagation commonly used for traditional ANNs. Back propagation is not a viable training method for SNNs because spiking neuron models are nondifferentiable. Approximations of back propagation do exist, but they often do not leverage the temporal feature learning capability of SNNs because of the stateless nature of back propagation. As such, various other training algorithms have been developed based on Hebbian learning principles, spike-time dependent plasticity (STDP), and evolutionary adaptation. The Evolutionary Optimization of Neuromorphic Systems (EONS) learning methodology was leveraged because of its demonstrated success in real-world applications [19]. This methodology was originally created by one of the senior members of this project, Dr. Katie Schuman.

The TENNLab Framework [12], developed by researchers at UTK and ORNL, was used develop, train, evaluate, and deploy an SNN model for this project. The TENNLab Framework enables researchers to design and train SNNs using a variety of SNN training algorithms, including EONS. One of TENNLab's unique capabilities is that, by design, it enables hardware/software co-design. The SNN models are trained on accurate hardware simulators of the neuromorphic hardware system being used, with all of the computational and numerical limitations of the hardware. This strategy ensures that models trained via the framework are deployable on hardware after training without loss of accuracy or reliability caused by differences in hardware constraints and numerical precision.

4.2 ALGORITHM MOTIVATION

To guide the algorithm development process, the requirements, purpose, and desirable characteristics of the resulting model were first established. These guiding features listed and described as follows motivate the design decisions made throughout development.

- **High Anomaly Detection Rate**: The algorithm must have a high true positive rate to ensure reliability and sensitivity in the anomaly detection task.
- **Low False Alarm Rate**: False alarms (false positives) must be minimized to improve operational efficiency. In addition, because the algorithm is intended as a trigger for higher-power processes, false positives reduce the energy efficiency of the overall system.
- **Energy Efficiency**: As the primary component of the software, the algorithm is largely responsible for managing the overall energy of the system.

- **Compactness and Sparsity**: In order to run on the µCaspian neuromorphic hardware, the SNN must be quite small. In addition, because µCaspian is an event-driven architecture, activity sparsity in the network reduces latency and, thus, power consumption.
- **Real-Time Detection**: The algorithm must be capable of real-time detection to be useful for the intended operation as a top and queue system and to provide remote operators with anomaly alarms as soon as they are triggered.

An algorithm characterized by all of these traits is difficult to achieve. Prior work has demonstrated that, in scenarios characterized by an unknown and dynamic background radiation environment, machine learning–based approaches can yield significantly higher performance than traditional statistical or physics-based approaches [20]. Specifically for anomaly detection, deep learning models using unsupervised learning approaches, such as the autoencoder radiation anomaly detection (ARAD) model, outperform principal components analysis (PCA)-based models, sequential probability ratio test (SPRT) algorithms, and the ubiquitous $k\sigma$ algorithm in both simulated and real-world anomaly detection scenarios [21]. Nevertheless, machine learning models based on traditional deep learning technology are not without significant disadvantages. To operate in real time, the largest of these models often require consumer-grade GPUs, which are not feasible to deploy for more than a few hours on battery power. Existing hardware specifically designed for running deep learning models at the edge, such as the NVIDIA Jetson Orin nano, consumes more than 15 W of power when active, again limiting their utility for the desired application of long-term unattended deployment.

By contrast, neuromorphic computing seeks to develop hardware/software solutions that enable the computational performance of deep learning models while reducing the power to a level feasible for embedded and low-power edge computing devices. As stated in Section 4.1, neuromorphic SNNs are notoriously challenging to train effectively. As such, developing, training, and testing the SNN model for radiation anomaly detection was a significant part of the technical development cycle in this project.

4.3 INITIAL DEVELOPMENT

The first iteration of the SNN model was developed during the first year of the project. This iteration demonstrated the feasibility of using SNNs for radiation detection tasks using a known and well-understood dataset for development. This stage of development also set the foundation for the algorithm training and evaluation pipeline for the rest of the project's development cycle. A detailed account of the initial development of the algorithm model was presented and published in the 2022 International Conference on Neuromorphic Systems (ICONS) conference proceedings [22].

4.3.1 Dataset

We used the Urban Radiation Search Dataset, which was developed over the course of several years by an ORNL, Los Alamos National Laboratory, and LBNL team through the DNN R&D sponsored MUSE project. This dataset will be referred to as the Topcoder dataset throughout the rest of this report.

The Topcoder dataset is fully synthetic, simulating a vehicle-borne $2 \times 4 \times 16$ in. NaI(Tl) detector moving through an urban city street. The background radiation is highly dynamic, with materials in the simulated buildings containing naturally occurring ^{238}U (and daughters), ^{232}Th (and daughters), and ^{40}K in variable concentrations in line with experimentally validated concentrations [16]. The dataset contains six unique shielded and bare source configurations placed in various locations in the simulated city blocks. These sources are highly enriched uranium (HEU), weapons-grade plutonium (WGPu), ^{131}I, ^{60}Co, ^{99m}Tc, and a mixture source containing both HEU and ^{99m}Tc. The methods used to generate this dataset are outlined in detail in prior work [6]. Details on how the dataset was designed and distributed and how it can be accessed are also described in prior work [7].

Table 4. Parameter selection for the initial hyperparameter optimization experiment

Parameter name	Parameter values
Initial nodes	10, 50, 100
Initial edges	100, 300, 500
Encoding bins	1, 2, 4
Encoding method	spike, spikes, temporal, rate

4.3.2 Experimental Design

4.3.2.1 Algorithm Design and Operation

The initial version of the SNN algorithm operated in a stateless manner. The SNN processed each 1 s integrated gamma-ray spectrum sequentially and produced a binary value as the output: 0 indicating no anomaly and 1 indicating an anomaly. During each inference, the SNN is presented with a normalized spectrum containing 128 (initially) energy bins between 0 and 3 MeV, with each input neuron mapping to a single energy bin. In this case, the spectra are normalized to integral unity. The normalized count data in each bin are converted into a spike train according to the defined encoding method, and these spike trains are processed through the network for a defined period of time (50 ms). A spike rate decoder converts the spike train of the output neuron into a binary value, which is output as described above. After inference on a single spectrum, the SNN's state (neuronal activity and propagating speaks) is reset before the next inference. This workflow is illustrated in Figure 8.

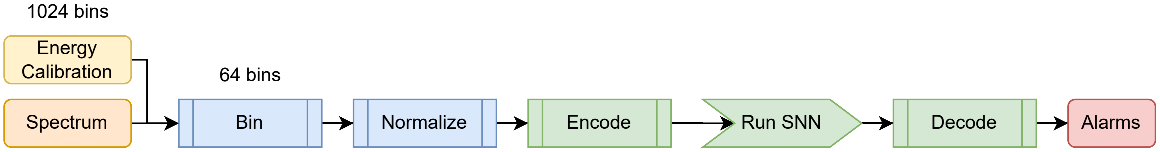


Figure 8. Inference workflow for NeuroRad SNN algorithms.

4.3.2.2 Hyperparameter Optimization

Without a precedent on the ideal set of initial design decisions and hyperparameter selections for SNNs on these types of tasks, a large hyperparameter optimization experiment was necessary. An ORNL-developed tool called the Data-Efficient Framework for Exploration (Deffe) was leveraged to explore the design space in coordination with EONS. Because of the many variables to consider, the initial experiment performed a search across a selection of parameters that strongly influence the SNN's behavior. These parameters are shown in Table 4. The "initial nodes" parameter refers to the number of neurons initially generated for each candidate SNN in the EONS population. The "initial edges" parameter refers to the number of connections (synapses) between neurons in each SNN. The "encoding bins" and "encoding method" parameters determine how input data are encoded into spikes.

4.3.2.3 Fitness Function

EONS optimizes the population of SNNs for a user-defined fitness function. The initial experiment used the F1-score as the fitness function. The F1-score is shown in Eq. (2), where the variables are defined as follows. The true positive rate (TPR), also known as recall, measures the ratio of the true positive (TP), or in this case, the anomalous source encounters that were correctly identified as an anomaly, to the total number of anomalies (P). The positive predictive value (PPV) measures the ratio of TP to the sum of TP and false positive (FP). EONS seeks to maximize the F1 score via the optimization process.

$$F1 = 2\frac{TPR \times PPV}{TPR + PPV} \tag{2}$$

4.3.2.4 Reference Algorithms

In order to benchmark the performance of the optimized SNN, its performance was evaluated against the performance of two reference algorithms evaluated on the same testing dataset from the Topcoder set. The two reference algorithms were the ubiquitous kσ algorithm [23] and a full-spectrum-based algorithm called the spectral anomaly detector (SAD) [24]. A simple statistical anomoly detection algorithm, kσ analyzes the windowed moving average of the background count rate and triggers an alarm when the foreground window count rate increases above *k* multiples of the background standard deviation, σ. The user sets the threshold value *k*, typically based on a receiver operating characteristic (ROC) curve to achieve the desired balance between false positive and true positive rates. The theory underlying the kσ algorithm is described in prior work [23]. Additional hyperparameters for the kσ algorithm are the widths of the background and foreground windows. A hyperparameter random search was performed in this study to optimize these windows for the highest performance, and the threshold *k* was set at a value that resulted in a false alarm rate (FAR) equal to that of the final SNN.

The second reference algorithm, SAD, is an algorithm that leverages the full gamma-ray energy spectrum to perform anomaly detection. The algorithm leverages PCA to determine a set of background spectrum components that can be combined to reconstruct any given spectrum in inference mode. Features in the spectrum arising from nonbackground sources cannot be reconstructed using the background feature components (SAD components are extracted from background-only data), and thus reconstruction error increases. A threshold can be set by the user, beyond which the algorithm triggers an alarm. Like for kσ, this threshold was set to match the FAR of the optimized SNN.

4.3.3 Results

Full details of the results are described elsewhere [22]. Following are some notable highlights from the initial results:

- The SNN achieved a low FAR on the US Department of Energy Urban Search Challenge dataset of only 0.5 per hour, despite the highly dynamic nature of the background radiation in the dataset. This FAR is well below the 1 per hour standard outlined in the ANSI N42.34 [25] and ANSI N42.53 standards [26].
- Compared with the baseline kσ algorithm, the SNN achieved a probability of detection (PD) over 10 times higher than kσ when evaluated at the same FAR of 0.5 per hour. Further developments and optimization in the latter half of 2022 resulted in an additional doubling of the SNN's TPR relative to the results published in prior work [22].
- A comparison was also made with a much more complex PCA-based spectral anomaly detection algorithm from the literature. This algorithm tends to perform on par with top-performing algorithms on this dataset. When evaluated at the same FAR of 0.6 per hour, the SNN achieved a PD about half that of the PCA-based algorithm, indicating room for further research and development of the SNN approach.
- The resulting SNN algorithm is small enough to fit on the smallest and lowest-power neuromorphic μCaspian hardware design.
- The evolved SNN architecture highlighted some interesting characteristics of SNNs when trained using EONS. Notably, the SNN heavily leverages the input neurons for computation as well as for

simple data input. This observation is in contrast to typical ANNs, where the input layer is typically responsible only for transmitting input data to deeper layers. Because EONS optimizes for both performance and computational efficiency, this result was not surprising but highlights the efficiency of our approach.

- The hyperparameter optimization study demonstrated that rate-based and spikes-based encoding methods, both with one encoding bin, performed best among all encoding methods. Temporal encoding performed the worst and failed to generate any useful networks. A starting node and edge size of 10 and 500, respectively, yielded the highest performing networks. As such, the model performed optimally with fewer initial neurons and more connections between the neurons.

4.4 SECOND DEVELOPMENT ITERATION

Establishing the workflow and initial design decisions for developing, training, and evaluating SNNs for radiation detection data allowed for further investigation of additional hyperparameters and evaluation of the model on more realistic datasets. This section describes this effort and the resulting performance. Additional details can be found in a paper that was presented and published at the 2023 ICONS conference proceedings [27].

4.4.1 7000 Dataset

Although the Topcoder dataset was an ideal starting point for developing an algorithm development and testing workflow, it did not accurately represent the intended use cases of the NeuroRad detection system. First, it models a mobile detection system traveling at vehicle speeds up to 13 m/s. The result is a highly dynamic background, especially when considering the high variation in the potassium/uranium/thorium concentrations of the simulated buildings. The intended use of this system is for static or mobile deployment at walking speeds. Training a model for such a use case on dynamic data will likely result in an algorithm with reduced sensitivity owing to the higher noise floor, which limits minimum detectable activity. Second, a more expansive set of anomaly source terms was desired to evaluate performance at different energy ranges. For these reasons, the dataset described in Section 3 was created based on prior work at ORNL. For this second iteration of development, the 7000 dataset used to train the models still used the original $2 \times 4 \times 16$ in. NaI(Tl) detector. Future experiments incorporated a new detector model for the 3×3 in. detector used in the final system.

4.4.2 Spiking Neural Network Features Developments

A new set of features to modify the behavior and capabilities of the SNN and methods of preprocessing the input data were developed. These features are described in detail in prior work [27], but are summarized as follows:

- **Temporality:** In the first iteration of the model, the SNN was essentially stateless between each inference. As such, it could not leverage the temporal pattern of a passing source, relying solely on spectral features for each spectrum, independent of features in past spectra. The team hypothesized that, by allowing the network to maintain its state between each inference, it could learn to leverage both the spectral and temporal features of an anomaly, thus improving performance. This feature was thus implemented with a simple way to enable/disable it for comparison purposes.
- **Spike Rate Thresholding:** Rather than training the model to directly output a binary value indicating the anomaly, the model was instead trained to output a high spike rate in the presence of an anomaly and a low spike rate otherwise. This strategy enables the user to set an alarm threshold on the spike rate, giving them the ability to tune the ratio between TPR and FAR to meet their specific use case requirements. This feature resulted in the need to redefine the fitness function, which is described in Section 4.4.3.

Table 5. Parameter selection for the second hyperparameter optimization experiment performed

Parameter name	Parameter values
Temporal	True, false
Leak	True, false
Energy binning scheme and count	*linear32*, *linear64*, *sqrt32*, *sqrt64*, *resnorm32*, *resnorm64*, *ROI*

- **Reduced Spectral Granularity:** For a small detector, such as the one used in this project's hardware, 128 bins may result in a noisy spectrum. Reducing the number of energy bins would increase the SNR at the expense of reduced peak resolution. To test the effect of different numbers of bins, a new energy bin count hyperparameter was created.
- **Nonlinear Energy Binning Schemes:** Because of the nonlinearity of the detector's energy resolution with respect to energy, a variety of energy binning schemes—including simple linear binning (*linear*), $\sqrt{E}$ binning (*sqrt*), and resolution normalized binning (*resnorm*)—were incorporated.
- **Region of Interest (ROI) Binning:** An additional binning scheme uses energy ROIs developed using the Censored Energy Window algorithm [28]. This algorithm develops a set of energy ROIs, which may overlap the maximized SNR for a predefined set of sources.

4.4.3 Fitness Function

The spike rate thresholding features necessitated the development of a new fitness function because the SNN no longer output an anomaly detection directly. The new fitness function is based on the ROC curve. After each step in the training procedure, the resulting networks are processed through the entire validation dataset, and a ROC curve is constructed by calculating TPR and FAR across thresholds. From this ROC curve, the TPR at an operationally relevant FAR of 1 per hour, called TPR_1, is extracted. One might consider using the TPR_1 directly as the fitness function; however, because the SNNs initially have FAR values much greater than 1 per hour, $\text{TPR}_1 = 0$ and any improvement in the value is left to random chance. To avoid this issue, the F1 score was calculated using the network's output spike decoder, just as was done for the initial experiments, and combined with TPR_1 extracted from the ROC curve. The resulting fitness function is shown in Eq. 3.

$$\text{F1} + \text{TPR}_1 \tag{3}$$

Initially, the F1 score dominated the fitness function, but eventually, the SNN's FAR increases above 1 and TPR_1 begins to dominate.

4.4.4 Hyperparameter Optimization

With new features implemented, a hyperparameter optimization study was necessary to quantitatively evaluate their performance. Table 5 shows the investigated parameters and their values. The parameters in Table 4 were set to the optimal values determined in prior experiments.

4.4.5 Results

The results of this study are outlined in detail in published work [27]; however, a summary of these results is listed here:

- Temporality increased detection performance by 55%, indicating that the networks could leverage temporal information to better detect sources. This significant performance improvement highlighted the importance of leveraging the anomaly's full spatiotemporal signature.

- Leakage had little to no effect on the performance of this application. This good news means that leakage hardware does not need to be implemented on the neuromorphic hardware design, resulting in better computational and energy efficiency. The *linear32* binning scheme had the lowest performance across binning schemes.
- The *sqrt* binning scheme, which nonlinearly scaled bin width as a function of energy, yielded the highest performance across all binning schemes. Additionally, using 32 bins in the spectrum as opposed to 64 did not decrease performance. Fewer bins decreases the size of the model and thus decreases inference latency and increases computational and power efficiency.
- The ROI input format yielded performance slightly higher than the *sqrt* method, however, the ROI method has its own limitations that must be considered. Notably, ROIs are only optimized for a specific set of source terms; thus, an algorithm based on a fixed set of ROIs risks transferability limitations to new sources on which the ROIs were not trained.
- Despite closing the gap between the SNN and the reference SAD algorithm, the best-performing SNN's TPR was about 20% lower than that of SAD, indicating an opportunity for additional optimization.

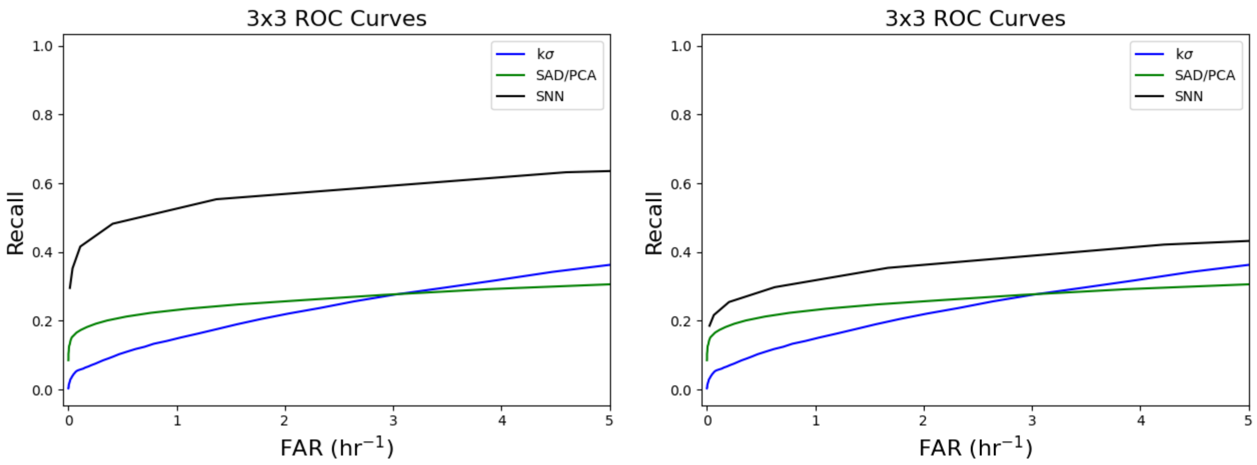
4.5 THIRD DEVELOPMENT ITERATION

The third iteration of development did not incorporate any major new features. Instead, it focused on performing a larger training operation comprising static hyperparameter configurations based on the findings of the first two development iterations. Networks with ROI binning schemes were trained in addition to networks with *sqrt32* binning. Modifications to the data and fitness function are listed here:

- **New detector model:** The detector in the datasets for both the first and second development iteration set measured 2 × 4 × 16 in., far larger than the 3 × 3 in. detector selected for the NeuroRad system, and with a different geometry. The result is a significant difference in detector response that leads to poor transferability of the algorithm to the real world. As such, a new detector response function was used for the NeuroRad detection systems to generate a new dataset based on the 7000 model described in Section 3.
- **Dataset size and characteristics:** The dataset was expanded to include more runs for each source term and background potassium/uranium/thorium configuration. In addition, the range of SNR values was expanded to incorporate more source encounters with low-integral SNR in an attempt to improve the model's low SNR performance.
- **Fitness function:** The fitness function was slightly modified, now being F1+2×TPR_1. This function more heavily weights the TPR_1 metric in the fitness evaluation, which is more operationally relevant than F1.

4.6 RESULTS

The third iteration of development resulted in two SNNs that significantly outperformed both the reference kσ and SAD algorithms on the testing dataset: one with ROI binning and the other with *sqrt32* binning. All of the resulting networks from the EONS training run were evaluated on the testing set, and the top-10 performing models for each binning scheme were saved. Interestingly, the networks that performed best in training did not always perform best in testing. One of the benefits of the EONS algorithm is that testing the many candidate models enables evaluation of the ones that transfer the best to new scenarios. The best-performing models on the test set were typically in the top 20% of models on the training set. The ROC curves for the best-performing SNN with the ROI binning scheme and with the *sqrt32* binning scheme are plotted in Figure 9. Although the ROI model has the highest performance among all networks, the *sqrt32* model may be a better choice in situations where the source of interest is not covered in the ROIs. In both cases, the ROC curves for the reference kσ and SAD algorithms are shown for comparison.



(a) SNN with ROI binning scheme.

(b) SNN with *sqrt32* binning scheme.

Figure 9. ROC curves for best performing SNN models on the 7000 dataset (third development iteration) with the 3 × 3 in. NaI(Tl) detector model, compared with the reference kσ and SAD algorithms.

5. EXPERIMENTAL CAMPAIGN AND RESULTS

To quantify overall system performance, a data collection campaign was conducted at ORNL's Technical Testing and Analysis Center (TTAC) facility from February 6–7, 2024. The purpose of this campaign was to understand the system performance when the top-performing neuromorphic algorithms were run on the new hardware outside of benchtop testing and virtual testbeds. This campaign allowed the project to collect valuable data, including new backgrounds and source configurations not available in the lab. It also allowed for an evaluation of the hardware performance.

TTAC is accredited by the National Voluntary Laboratory Accreditation Program to the ANSI/IEEE N42 Homeland Security Standards in the radiological, mechanical, environmental, and electromagnetic areas, as well as to the radiation protection instrumentation standards. The facilities used for this experiment included radiation sources and a track capable of moving sources past the (stationary) detector ranging from 0.5–2 m/s (or about 1–4.5 mph). The detector was placed on a table at a predetermined distance of closest approach to the track, and data were recorded. A photograph of the experimental setup is shown in Figure 10. The facility has lead-lined curtains to the left and right of the ends of the track to shield the source between passes and ensure the detector has time to return to background levels between passes.

The detector system was connected to a laptop for real-time data analysis to ensure the system was working as intended throughout the campaign. The system ran on battery power while collecting the overnight background, ensuring that this functionality worked as intended. Throughout the testing campaign, calibration tests were performed and background spectra were collected to ensure that the energy calibration was stable and that no additional sources were present in the data. In addition, the background data were used to set algorithm thresholds to a false alarm rate of 1 per hour.

For this experimental campaign, industrial radioactive sources were chosen that spanned from low energies (59.6 KeV ^{241}Am) to high energies (173 keV and 1,333 keV ^{60}Co) and in between (661 keV ^{137}Cs). In addition, special nuclear materials (HEU and WGPu) were used. All sources were bare (no shielding) except

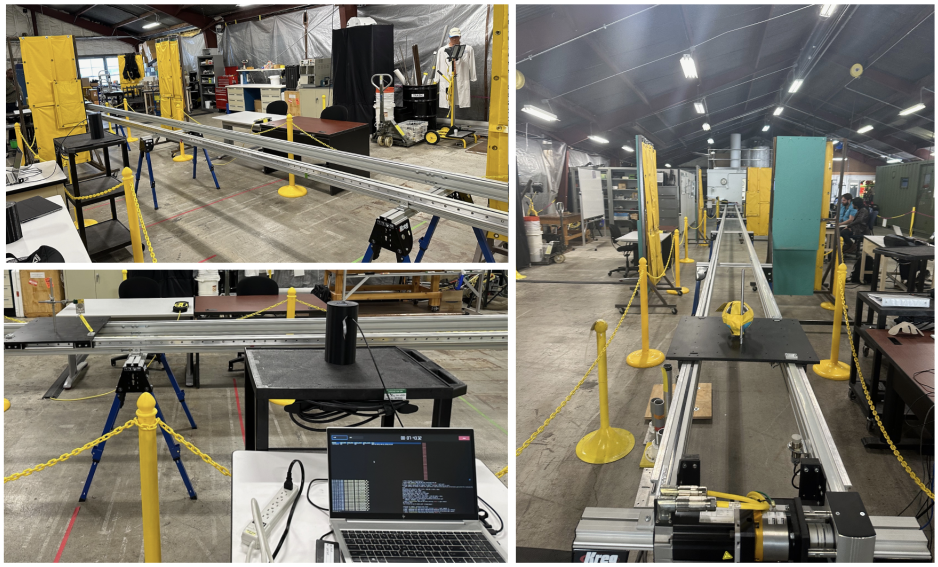


Figure 10. Experimental setup. (top left) The track used in the experiment with yellow lead-lined curtains. The detector (black cylinder) is located on a table 1-meter from the track. (right) Another view down the track. The HEU spherical source is loaded on the track. (bottom left) The detector connected to a laptop for real-time data analysis.

for the WGPu source, which used an approximately 5 mm thick copper shield to reduce signal from ^{241}Am buildup. The sources and experiment parameters are shown in Table 6. Before collecting dynamic data, static long-dwell measurements were collected with each of the sources except for ^{152}Eu at 1 m. These data were used to create measured source templates for later injection studies and to predict SNR for various offsets in the campaign.

Table 6. Source list for TTAC experiments

Isotope	Activity (µCi)	Shielding
^{241}Am	36.0	Bare
^{137}Cs	91.8	Bare
^{60}Co	40.0	Bare
^{152}Eu	39.2	Bare
HEU	2,950.0	Bare
WGPu	210,951.3	Copper shielding

Experiments were performed by moving each source past the detector system at various speeds (0.5, 1.0, and 2.0 m/s) and source–detector standoffs (1, 2, and 3 m) at the point of closest approach. For each source/speed/standoff, 10 source passes were completed to generate statistics for detector system performance characterization. The number of parameters (speed, offset, standoff, and number of passes) were chosen to complete the experiments in 2 days. Table 7 lists all of the data collected during the TTAC campaign.

Table 7. Source list for TTAC experiments

Run #	Source	Source Speed (m/s)	Standoff (m)	# Repetitions
1	Background	—	—	1

2	Calibration	—	—	1
3	Calibration	—	—	1
4	^{137}Cs	0.5	1	10
5	^{137}Cs	1.0	1	10
6	^{137}Cs	2.0	1	10
7	^{241}Am	0.5	1	10
8	Background	-	-	10
9	^{241}Am	1.0	1	10
10	^{241}Am	2.0	1	10
11	^{60}Co	0.5	1	10
12	^{60}Co	1.0	1	10
13	^{60}Co	2.0	1	10
14	^{152}Eu	0.5	1	10
15	^{152}Eu	1.0	1	10
16	^{152}Eu	2.0	1	10
17	WGPu	0.5	1	10
18	WGPu	1.0	1	10
19	WGPu	2.0	1	10
20	HEU	0.5	1	10
21	HEU	1.0	1	10
22	HEU	2.0	1	10
23	HEU	0.5	2	10
24	HEU	1.0	2	10
25	HEU	2.0	2	10
26	WGPu	0.5	2	10
27	WGPu	1.0	2	10
28	WGPu	2.0	2	10
29	^{152}Eu	0.5	2	10
30	^{152}Eu	1.0	2	10
31	^{152}Eu	2.0	2	10
32	^{60}Co	0.5	2	10
33	^{60}Co	1.0	2	10
34	^{60}Co	2.0	2	10
35	^{241}Am	0.5	2	10
36	^{241}Am	1.0	2	10
37	^{241}Am	2.0	2	10
38	^{137}Cs	0.5	2	10
39	^{137}Cs	1.0	2	10
40	^{137}Cs	2.0	2	10
41	Background	—	—	1
42	Calibration	—	—	1
43	^{137}Cs	0.5	3	10
44	^{137}Cs	1.0	3	10
45	^{137}Cs	2.0	3	10
46	^{241}Am	0.5	3	10
47	^{241}Am	1.0	3	10
48	^{241}Am	2.0	3	10
49	^{60}Co	0.5	3	10

50	^{60}Co	1.0	3	10
51	^{60}Co	2.0	3	10
52	^{152}Eu	0.5	3	10
53	^{152}Eu	1.0	3	10
54	^{152}Eu	2.0	3	10
55	WGPu	0.5	3	10
56	WGPu	1.0	3	10
57	WGPu	2.0	3	10
58	HEU	0.5	3	10
59	HEU	1.0	3	10
60	HEU	2.0	3	10

The measured count rate as a function of time for all source measurements for 1 m/s source speed is shown in Figure 11. As the source offset increases, the signal from the source becomes weaker and harder to detect. Most sources were easily detected for slow speeds and close distances of approach. As the sources increased in speed and distance of the closest approach, the detection probability decreased. Figure 12 shows results for all ^{241}Am measurements. The blue shaded areas indicate when the network spike count rate exceeded the detection threshold and an alarm was triggered. The alarm threshold was set to a false alarm rate of 1 per hour, using background collected at TTAC.

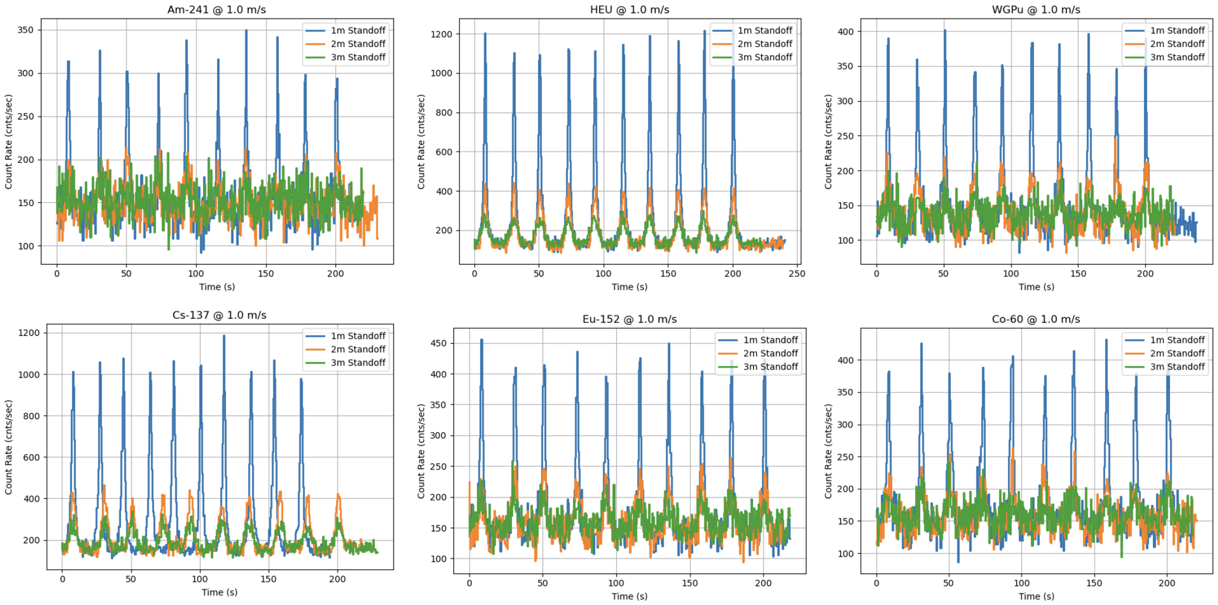


Figure 11. Measured total count rates for each source for variable standoff distances for the source moving at 1 m/s. For the ^{137}Cs subplot, the dwell time at the end of the track was increased after the 1 m standoff ^{137}Cs measurements.

The probability of detection for each parameter set was calculated by counting the number of alarms and dividing by the number of source encounters. Figure 13 shows algorithm results across all measured sources, speeds, and standoffs for the top-performing network (network-242). This network performed well for ^{241}Am, ^{60}Co, and ^{137}Cs but did not perform well for ^{152}Eu, HEU, or WGPu. This result was surprising

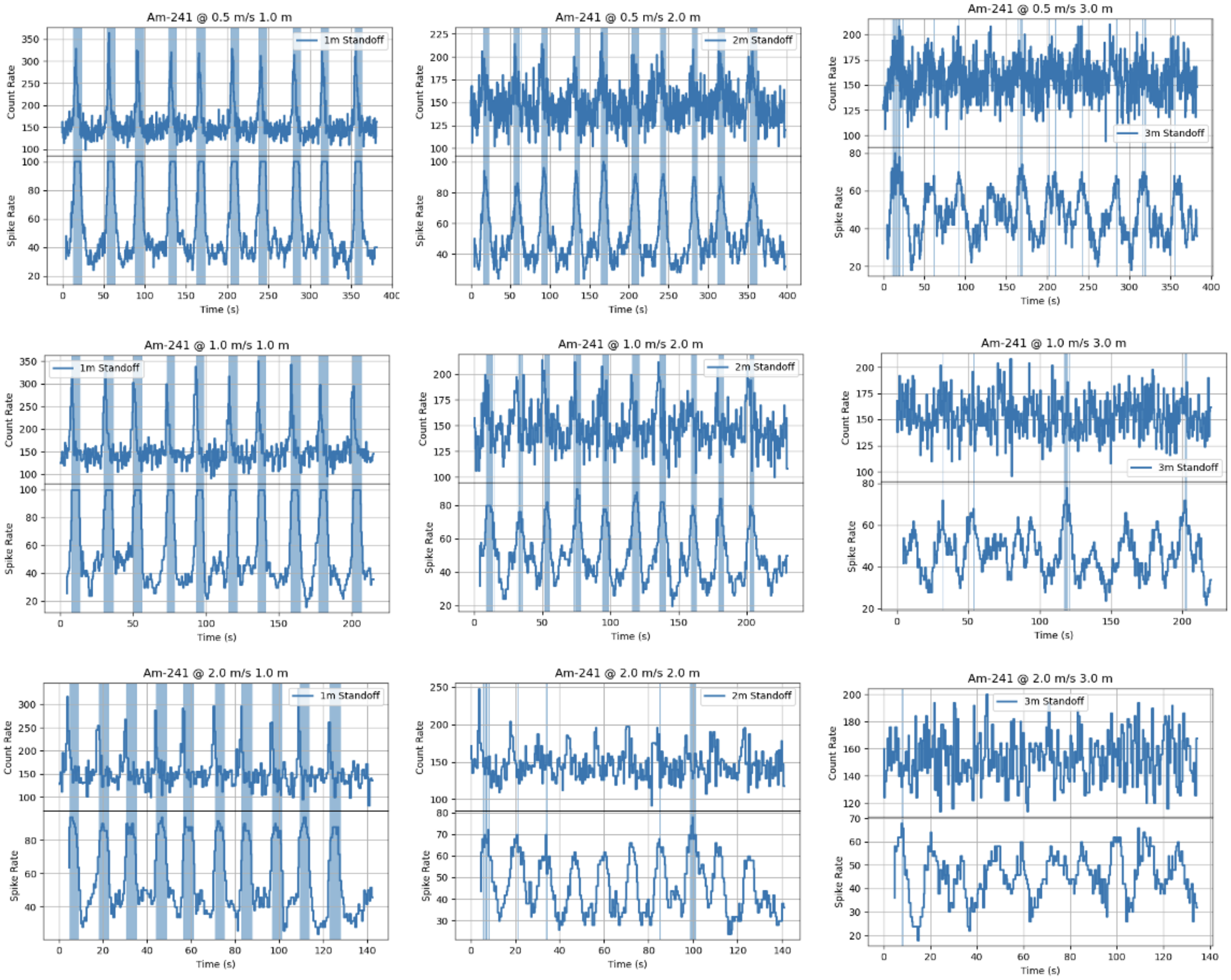


Figure 12. All data with alarms for ^{241}Am. The source–detector standoff increases from left to right, and the source speed increases from top to bottom.

because this network performed well in Monte Carlo–generated training and testing data. The next section discusses an investigation into the cause of this poor performance.

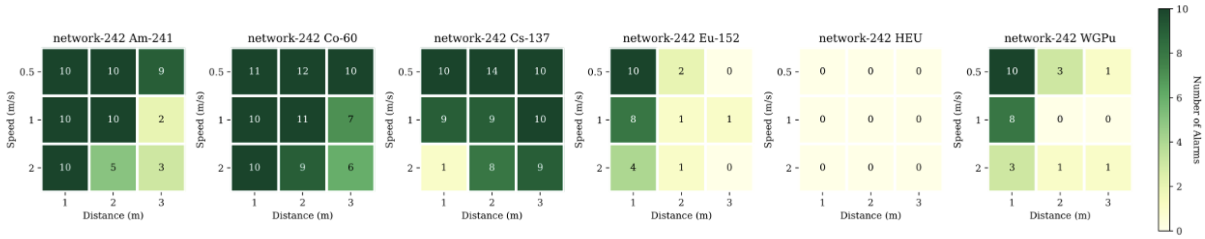


Figure 13. Algorithm results for the top performing network on the TTAC data. The maximum number of defections for each source/speed/standoff combination is 10.

The first major success of this campaign was demonstrating that the hardware worked as expected. No issues arose while running on battery power, and the data are high quality. The system collected overnight backgrounds on the battery without any data loss. The detector energy calibration of the system was stable and did not change during the measurement campaign. Although the algorithms performed well for the industrial sources, this campaign also revealed some discrepancies between how the SNN algorithms performed in the lab and how they performed on synthetic data used in training. Understanding why these discrepancies occurred will inform how we train these algorithms in the future.

6. TTAC CAMPAIGN DISCUSSION

After the TTAC campaign, an investigation was conducted to better understand algorithm results. During the training and testing on synthetic data, the SNN algorithms performed well and outperformed other anomaly detection algorithms such as kσ, SAD, and SPRT. One possible reason for the discrepancy between performance on synthetic data and measured data at TTAC is systematic differences in the data—both in the background and threat source responses. To understand whether these differences were a factor, this section compares measured and simulated background responses and studies algorithm performance when experimental and simulated response functions are mixed.

6.1 BACKGROUND RESPONSE STUDY

If the background response measured at TTAC is significantly different than the synthetic background used to train the SNN algorithms, then performance could be degraded. To better understand this issue, the background at TTAC was compared with the average Monte Carlo–generated backgrounds used in training and testing (Figure 14). Discrepancies in the number of counts from 1,460 keV (^{40}K) and 2,614 keV gamma-ray peaks(^{208}Tl—daughter from ^{232}Th decay) are to be expected because concentrations of potassium, uranium, and thorium vary in the environment and the Monte Carlo dataset was not generated for this specific facility. Potassium, uranium, and thorium were varied in the Monte Carlo test and training data to account for different environments, and Figure 14 represents only the average background in the testing set.

The spectral shape at low energies in the TTAC is significantly different from Monte Carlo backgrounds. This difference is likely owing to two main factors:

1. The low-energy scattering is much higher in the Monte Carlo simulations than the data measured at TTAC.
2. The low-energy cutoff is different in the detector system than in the Monte Carlo simulation.

Therefore, the training and testing data did not have enough variation at low energies. This result is important to the spectral encoding scheme, where the bounds of the incoming spike rate frequency are set to the expected minimum and maximum variation in each spectral energy bin. The reasons why the low-energy scattering is so much larger in the Monte Carlo simulations are under investigation, but if real-world backgrounds are included in the training dataset, then the effect of this systematic difference can be minimized.

To determine whether the difference in backgrounds has a measurable effect on algorithm performance, the long-dwell source data measured at TTAC was injected into the Monte Carlo backgrounds, and the SNN algorithm was run over the results. Long-dwell backgrounds were subtracted from the long-dwell static source measurements at TTAC to create background-subtracted source templates. These source templates were then scaled and Poisson sampled using a simple $^1/_{r^2}$ scaling based on source standoff and speed to create new datasets for algorithm evaluation. The ^{152}Eu source templates were not created because a long-dwell static source measurement for this source was not taken. A new algorithm threshold was programmatically determined using the Monte Carlo backgrounds so that the false alarm rate was 1 per hour, the same as used in the TTAC experiments. Algorithm results are shown in Figure 15. The results show similar performance to the TTAC experiments, although a bit improved for ^{241}Am and ^{137}Cs, and a high probability of detection for the industrial sources. The poor performance for HEU and WGPu continued. This investigation indicates that, although the difference in background does affect algorithm performance, it is not the primary culprit for poor performance.

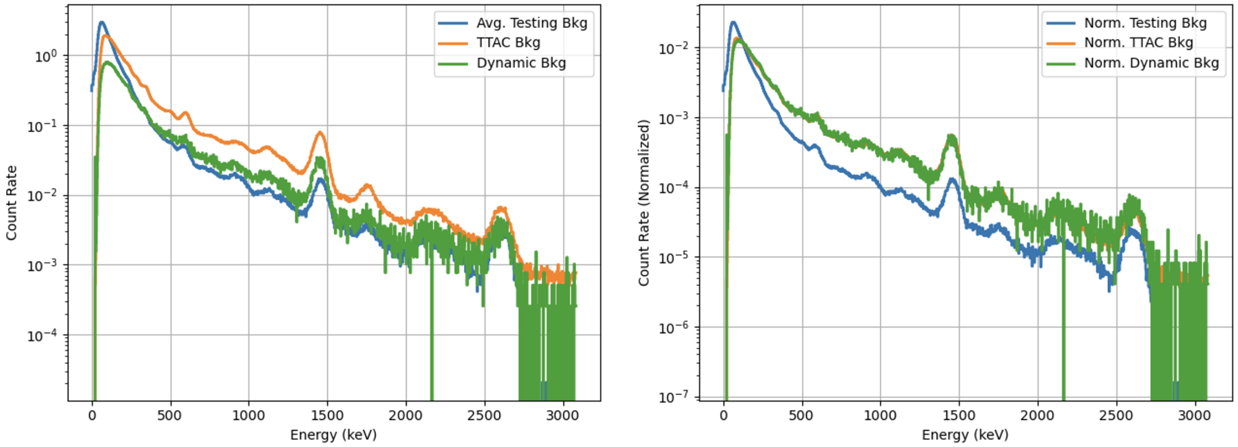


Figure 14. Comparison of various backgrounds: the average background in the Monte Carlo generated testing set used to evaluate neural network performance, the background measured at TTAC, and a dynamic background dataset collected via driving around the Knoxville, Tennessee, area. (left) The time-normalized count rate for each background. (right) Each spectrum normalized by the total number of counts.

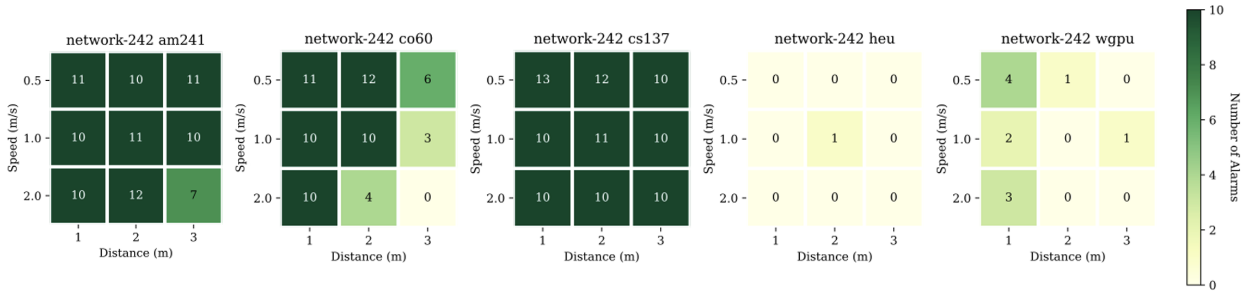


Figure 15. Algorithm results for the top performing network on TTAC source templates injected into virtual testbed backgrounds. The maximum number of defections for each source/speed/standoff combination is 10. The ^{152}Eu source templates were not created because a long-dwell static source measurement was not taken, so results for this isotope are absent.

6.2 THREAT SOURCE RESPONSE STUDY

Next, the team investigated whether the source definitions used in the synthetic data were significantly different from the sources measured in the TTAC campaign and whether this difference affects algorithm performance. The Monte Carlo–generated source's tallies are compared with long-dwell TTAC source spectra in Figure 16. At a high level, the measured TTAC and the Monte Carlo source spectra are similar, but some differences exist. The measured data seem to contain a slight energy calibration error at low energies, as shown by the ^{241}Am data. Furthermore, as indicated by the background comparison, the Monte Carlo data do not accurately reflect a low-energy cutoff in place in the measured data. The HEU data contain the largest source of discrepancy, which is likely due to differences in source geometry and composition from simulations.

A new dataset was generated by injecting Monte Carlo source spectra into measured TTAC backgrounds. The top-performing SNN algorithm processed the results, using the same threshold calculated for the

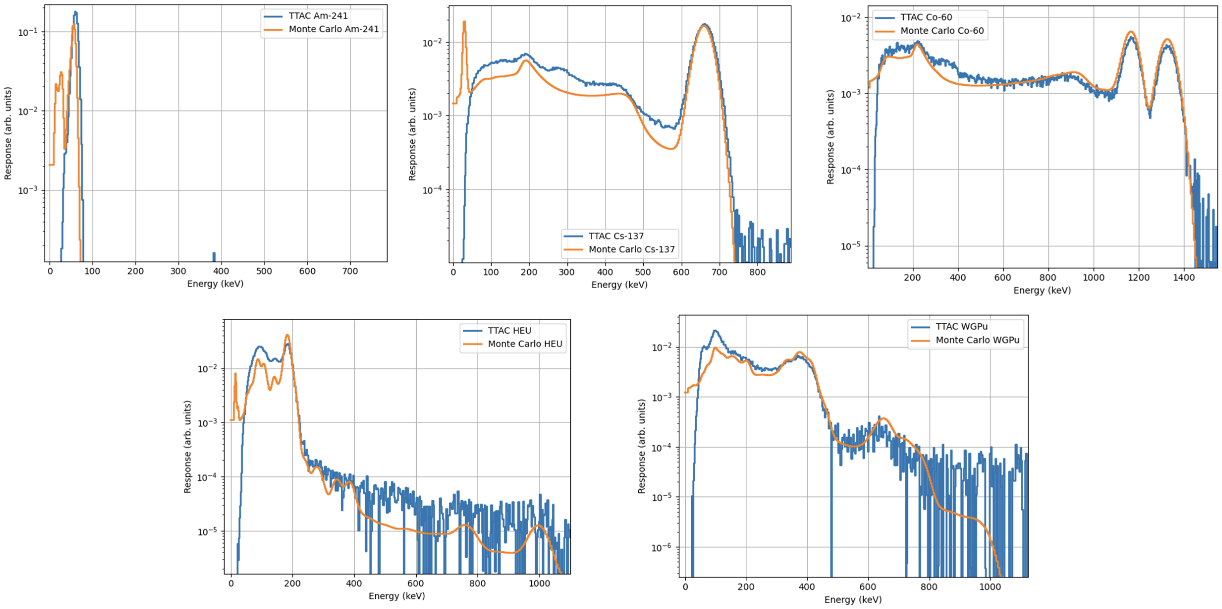


Figure 16. Comparison of the Monte Carlo generated source mesh tallies and measured long dwell source spectra measured at TTAC. Both spectra are scaled.

TTAC measurements (the background is the same in both) to generate the probability of detection results shown in Figure 17. The results exhibit improved performance over the TTAC experiments, with a high probability of detection for the industrial sources and WGPu, but still poor performance for HEU. This experiment indicates that the difference in source profiles is the primary driver of the discrepancy, although the missing detections in HEU are still puzzling.

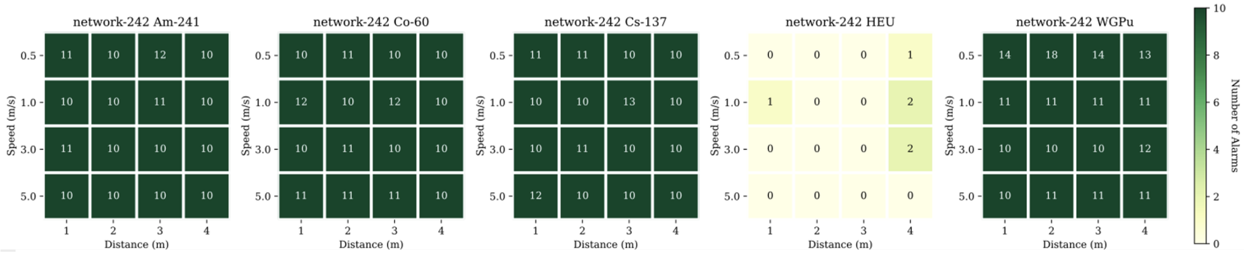


Figure 17. Algorithm results for the top performing network on TTAC source templates injected into virtual testbed backgrounds. The maximum number of defections for each source/speed/standoff combination is 10. The ^{152}Eu source templates were not created because a long-dwell static source measurement was not taken, so results for this isotope are absent.

6.3 DISCUSSION

This study indicates that the systematically different background at TTAC degraded algorithm performance but was not the primary cause of the poor performance for HEU and WGPu. Differences in the source responses between testing and training seem to have a larger effect, especially for WGPu. With all of the various source geometry and shielding configurations, the exact source response is impossible to predict for an unknown source. However, more source geometry and shielding configurations can be included to span the response space as much as possible, with the hope that this change will be enough to improve algorithm performance. The reason why HEU performance was so poor even when using synthetic source responses remains unknown. This issue will be investigated in future work.

7. REFERENCES

[1] David Michael Pfund et al. "Examination of Count-Starved Gamma Spectra Using the Method of Spectral Comparison Ratios". In: *IEEE Transactions on Nuclear Science* 54.4 (2007), pp. 1232–1238. DOI: 10.1109/TNS.2007.901202.

[2] K. J. Bilton et al. "Non-negative Matrix Factorization of Gamma-Ray Spectra for Background Modeling, Detection, and Source Identification". In: *IEEE Transactions on Nuclear Science* 66.5 (2019), pp. 827–837. DOI: 10.1109/TNS.2019.2907267.

[3] *SiPM-3000, Bridgeport Instruments*. https://www.bridgeportinstruments.com/products/sipm/sipm3k.html. Accessed April 2021.

[4] *Jetson Xavier NX*. https://developer.nvidia.com/embedded/jetson-xavier-nx. Accessed April 2021.

[5] *Detecting Radiological Threats in Urban Areas*. https://www.topcoder.com/lp/detect-radiation. Accessed March 2019.

[6] Andrew D. Nicholson et al. "Generation of Synthetic Data for a Radiation Detection Algorithm Competition". In: *IEEE Transactions on Nuclear Science* 67.8 (2020), pp. 1968–1975. DOI: 10.1109/TNS.2020.3001754.

[7] James M. Ghawaly et al. "Data for training and testing radiation detection algorithms in an urban environment". In: *Scientific Data* 7.1 (2020), p. 328. DOI: 10.1038/s41597-020-00672-2. URL: https://doi.org/10.1038/s41597-020-00672-2.

[8] Emma Strubell, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Modern Deep Learning Research". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.09 (Apr. 2020), pp. 13693–13696. DOI: 10.1609/aaai.v34i09.7123. URL: https://ojs.aaai.org/index.php/AAAI/article/view/7123.

[9] Catherine D. Schuman et al. "Neuromorphic computing for temporal scientific data classification". In: *Proceedings of the Neuromorphic Computing Symposium*. NCS '17. Knoxville, Tennessee: Association for Computing Machinery, 2017. ISBN: 9781450364423. DOI: 10.1145/3183584.3183612. URL: https://doi.org/10.1145/3183584.3183612.

[10] Bipin Rajendran et al. "Low-Power Neuromorphic Hardware for Signal Processing Applications: A Review of Architectural and System-Level Design Approaches". In: *IEEE Signal Processing Magazine* 36.6 (2019), pp. 97–110. DOI: 10.1109/MSP.2019.2933719.

[11] Alexander Kugele et al. "Efficient Processing of Spatio-Temporal Data Streams With Spiking Neural Networks". In: *Frontiers in Neuroscience* 14 (2020). ISSN: 1662-453X. DOI: 10.3389/fnins.2020.00439. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2020.00439.

[12] J. S. Plank et al. "The TENNLab Exploratory Neuromorphic Computing Framework". In: *IEEE Letters of the Computer Society* 1.2 (2018), pp. 17–20. DOI: 10.1109/LOCS.2018.2885976. URL: https://doi.ieeecomputersociety.org/10.1109/LOCS.2018.2885976.

[13] J Parker Mitchell et al. "Caspian: A neuromorphic development platform". In: *Proceedings of the 2020 Annual Neuro-Inspired Computational Elements Workshop*. 2020, pp. 1–6.

[14] J Parker Mitchell, Catherine D Schuman, and Thomas E Potok. "A small, low cost event-driven architecture for spiking neural networks on fpgas". In: *International Conference on Neuromorphic Systems 2020*. 2020, pp. 1–4.

[15] Tara M. Pandya et al. "Implementation, Capabilities, and Benchmarking of Shift, a Massively Parallel Monte Carlo Radiation Transport Code". In: *Journal of Computational Physics* 308 (Mar. 2016), pp. 239–272. ISSN: 00219991. DOI: 10.1016/j.jcp.2015.12.037. URL: http://linkinghub.elsevier.com/retrieve/pii/S0021999115008566.

[16] Daniel E. Archer et al. "Modeling Urban Scenarios & Experiments: Fort Indiantown Gap Data Collections Summary and Analysis". In: (Oct. 2017). DOI: 10.2172/1410932. URL: https://www.osti.gov/biblio/1410932.

[17] Cihangir Celik et al. "A directional detector response function for anisotropic detectors". In: *Nuclear Science and Engineering* 193.12 (2019), pp. 1355–1370.

[18] Mathew W. Swinney et al. "A Methodology for Determining the Concentration of Naturally Occurring Radioactive Materials in an Urban Environment". In: *Nuclear Technology* 203.3 (May 2018). ISSN: 0029-5450. DOI: 10.1080/00295450.2018.1458558. URL: https://www.osti.gov/biblio/1462899.

[19] C. D. Schuman et al. "Evolutionary Optimization for Neuromorphic Systems". In: *NICE: Neuro-Inspired Computational Elements Workshop*. 2020.

[20] Christine Michaela Anderson-Cook et al. *Radiation Detection Data Competition Report*. Tech. rep. Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2021.

[21] James M Ghawaly Jr et al. "Characterization of the autoencoder radiation anomaly detection (arad) model". In: *Engineering Applications of Artificial Intelligence* 111 (2022), p. 104761.

[22] James Ghawaly et al. "A neuromorphic algorithm for radiation anomaly detection". In: *Proceedings of the International Conference on Neuromorphic Systems 2022*. 2022, pp. 1–6.

[23] Lloyd A Currie. "Limits for qualitative detection and quantitative determination. Application to radiochemistry". In: *Analytical chemistry* 40.3 (1968), pp. 586–593.

[24] David Boardman, Mark Reinhard, and Alison Flynn. "Principal component analysis of gamma-ray spectra for radiation portal monitors". In: *IEEE Transactions on Nuclear Science* 59.1 (2012), pp. 154–160.

[25] "American National Standard Performance Criteria for Handheld Instruments for the Detection and Identification of Radionuclides". In: *ANSI N42.34-2015 (Revision of ANSI N42.34-2006)* (2016), pp. 1–60. DOI: 10.1109/IEEESTD.2016.7551091.

[26] "American National Standard Performance Criteria for Backpack-Based Radiation-Detection Systems Used for Homeland Security". In: *ANSI N42.53-2021 (Revision of ANSI N42.53-2013)* (2021), pp. 1–44. DOI: 10.1109/IEEESTD.2021.9502037.

[27] James Ghawaly et al. "Performance Optimization Study of the Neuromorphic Radiation Anomaly Detector". In: *Proceedings of the 2023 International Conference on Neuromorphic Systems*. 2023, pp. 1–7.

[28] Eric Lei et al. "Robust Detection of Radiation Threat Using Uncertain Censored Energy Windows". In: ().