Gamma/Neutron Online Discrimination Based on Machine Learning With CLYC Detectors

Iván René Morales, Romina Soledad Molina, Mladen Bogovac, Nikola Jovalekic, Maria Liz Crespo, Kalliopi Kanaki, Giovanni Ramponi, *Life Senior Member, IEEE*, and Sergio Carrato

***Abstract*— An embedded system (ES) for gamma and neutron discrimination in mixed radiation environments is proposed, validated with an off-the-shelf detector consisting of a Cs2LiYCl6:Ce (CLYC) crystal coupled to a silicon photomultiplier (SiPM) cell array. This solution employs a machine learning classification model based on a multilayer perceptron (MLP) running on a commercial field-programmable gate array (FPGA), providing online single-event identification with 98.2% overall accuracy at rates higher than 200 kilocounts/s. Thermal neutrons and fast neutrons up to 5 MeV can be detected and discriminated from gamma events, even under pile-up scenarios with a dead-time lower than 2.5 μs. The system exhibits excellent size, weight, and power consumption (SWaP) characteristics, packed in a volume smaller than 0.6 l and weighing less than 0.5 kg, while ensuring continuous operation with only 1.5 W. These features render our proposal suitable for embedded applications where low SWaP is critical and radiation levels manifest large count rates variability, such as space exploration, portable dosimeters, radiation surveillance on uncrewed aerial vehicles (UAVs), and soil moisture monitoring.**

***Index Terms*— Cs2LiYCl6:Ce (CLYC), field-programmable gate array (FPGA), gamma/neutron discrimination, machine learning, pile-up.**

I. Introduction

THE use of Cs2LiYCl6:Ce (CLYC) scintillators in mixed radiation detectors has recently become widespread due to their remarkable gamma/neutron (γ/n) discrimination capability, enabling gamma and fast neutron spectroscopy as well as thermal neutron sensing [1], [2]. These features may be exploited in several applications where size, weight, and power consumption (SWaP) justify a design based on a single multipurpose detector, such as space and planetary exploration [3], [4], [5], [6], [7], [8], portable gamma/neutron dosimeters [9], [10], remote radiation surveillance on uncrewed aerial vehicles (UAVs) [11], [12], and soil moisture monitoring [13], [14]. Although traditional photomultiplier tubes (PMTs) coupled to CLYC crystals offer excellent pulse shape discrimination and notorious gamma spectroscopy resolution, silicon photomultipliers (SiPMs) exhibit better immunity to external magnetic fields, withstand higher mechanical stress, and require a less complex bias power supply. Moreover, SiPMs evince much lower SWaP characteristics than PMTs, making them excellent candidates for applications demanding portability and mechanical robustness [15]. However, the parasitic capacitance of SiPM arrays not only severely penalizes the inherent pulse shape discrimination feature of CLYC scintillators, but also increases the pulse decay times [16], [17]. Hence, a higher probability of pile-up and reduced γ/n separation harm the overall performance (OP) compared to PMT-based detectors, encouraging the development of advanced techniques for γ/n discrimination and pile-up recovery in large-pulse count-rate scenarios [18].

Machine learning (ML) algorithms have shown promising results in γ/n discrimination for CLYC [19], [20], [21] and organic scintillators [22], [23], [24]. Offline ML methods have also been demonstrated to recover information in high event rates where pile-up usually distorts individual pulse shapes [25], [26], [27]. Consequently, applications requiring γ/n discrimination at high count rates are very likely to need similar advanced processing techniques. If online processing is added to this already-complex endeavor, achieving these goals simultaneously turns into an ambitious task targeting high-performance digital signal processing [28].

Aiming at leveraging the potential of SiPM-based CLYC detectors under these challenging conditions, we propose an efficient solution for online γ/n discrimination, enabling high count rates and low dead-time, featuring pile-up recovery in a low-SWaP embedded system (ES). We achieved these outstanding characteristics by tackling the discrimination challenge with a field-programmable gate array (FPGA) for online digital signal processing, executing a pipeline with a fast feature extraction system, followed by an optimized ML discrimination model.

Received 18 July 2024; revised 23 September 2024 and 8 November 2024; accepted 12 November 2024. Date of publication 14 November 2024; date of current version 19 December 2024. *(Corresponding author: Iván René Morales.)*

Iván René Morales and Romina Soledad Molina are with the Dipartimento di Ingegneria e Architettura (DIA), Università degli Studi di Trieste (UNITS), 34127 Trieste, Italy, and also with the Multidisciplinary Laboratory (MLab), STI Unit, The Abdus Salam International Centre for Theoretical Physics (ICTP), 34151 Trieste, Italy (e-mail: ivanrene.moralesargueta@phd.units.it; rmolina@ictp.it).

Mladen Bogovac and Kalliopi Kanaki are with the Nuclear Science and Instrumentation Laboratory, Physics Section, Division of Physical and Chemical Sciences, Department of Nuclear Sciences and Applications, IAEA, Vienna International Center, 1400 Vienna, Austria (e-mail: M.Bogovac@iaea.org; K.Kanaki@iaea.org).

Nikola Jovalekic is with Teledyne Healthcare, X-Ray Solutions, 5611 GS Eindhoven, The Netherlands (e-mail: nikola.jovalekic@teledyne.com).

Maria Liz Crespo is with the Multidisciplinary Laboratory (MLab), STI Unit, Abdus Salam International Centre for Theoretical Physics (ICTP), 34151 Trieste, Italy (e-mail: mcrespo@ictp.it).

Giovanni Ramponi and Sergio Carrato are with the Dipartimento di Ingegneria e Architettura (DIA), Università degli Studi di Trieste (UNITS), 34127 Trieste, Italy (e-mail: ramponi@units.it; carrato@units.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNS.2024.3498321.

Digital Object Identifier 10.1109/TNS.2024.3498321

The remainder of this article is organized as follows. In Section II, an overview of the state-of-the-art (SOA) is provided, followed by Section III where the specific contributions of this work are described. In Section IV, the ES is summarized, detailing the hardware platform and the FPGA firmware design. Section V dives deep into the experimental data recording, offline data analysis, and tagging for the ML training datasets, highlighting the attributes that enhance the efficiency of the system compared to other ML-based implementations. The online feature extraction system and its implementation are detailed in Section VI. Section VII describes the ML model and the performed optimizations to provide outstanding γ/n discrimination performance combined with low-latency continuous operation. Section VIII explains the tests conducted to evaluate the performance and SWaP aspects, including comparisons with SOA. Final thoughts and motivation for future work are summarized in Section IX.

II. Related Work

Regarding online γ/n discrimination methods optimized for high count rates using CLYC scintillators, Johnson et al. [18] proposed a solution based on digital filters to recover piled-up events in high-event-rate scenarios. They computed a smoothed discrete-time derivative filter on the detector traces, obtaining a sharp pulse to signal the time of arrival of each event. These pulses were further reshaped by a passband filter to enhance the differences between the gamma and neutron events. Although their solution provided a straightforward online method for FPGA implementation, their proposal required a long buffering stage, causing a dead-time close to 100 ms. Unfortunately, the accuracy decreased at higher count rates, and the algorithm proof-of-concept was tested offline: no FPGA deployment was performed.

Wen and Enqvist [29] used a fine-tuned triangular filter matched to the CLYC detector's pulse shape applying Jordanov and Knoll et al. [30] recursive implementation. As in [18], they measured the timing between zero-crossings at the filter output to determine the event type under high rates, achieving a constant dead-time of 2.67 μs. Although the measured neutron discrimination sensitivity was fairly low (28.9%) at the maximum count rate (375 kcps), the detection rate increased to 100% at 25 kcps or lower. No hardware implementation was further developed, leaving the FPGA deployment as a future work.

In terms of online methods based on ML for γ/n discrimination, Simms et al. [31] designed a system to detect the presence of neutron sources using liquid scintillators, aiming at minimizing the contamination of false gamma counts. Their system relied on an unsupervised gamma mixture model executed in an FPGA. Although the reported latency was remarkable (3 μs), offline preprocessing was required. Moreover, no pile-up recovery nor rejection strategies were implemented, rendering the method unsuitable for high count rates.

Astrain et al. [32] implemented an online γ/n discrimination system based on a 1-D convolutional neural network. Their deployment achieved a maximum count rate of 79 kcps and a maximum dead-time of 50 μs with a steady overall classification accuracy of 98%. The system was capable of performing single-event γ/n identification and pile-up recovery of up to two simultaneous events. Of relevant interest is that they designed the neural network using experimental data from the tokamak fusion reactor at the Jet European Torus.

Michels et al. [28] recently developed a complete solution for online γ/n discrimination based on ML using a set of MLP models. Classification with an overall accuracy of 98.2% and pile-up recovery was tested at count rates as high as 1.11 Mcps with a maximum dead-time of 7.7 μs. As in [32], up to two piled-up events could be identified using a captured (triggered) trace within a time window. Triple and quadruple pile-ups could be detected but not discriminated, while more than four piled-up events were not tested. A methodology to guarantee the ground truth of the event types for the ML training was not conducted. These high-performance metrics were achieved using a top-tier benchtop FPGA system, unsuitable for portable applications.

Related to low-SWaP implementations, Huang et al. [9] proposed a portable neutron-gamma detector based on CLYC, enabled by a pile-up rejection stage followed by a partial charge-to-peak ratio algorithm. Their SWaP features were 4 L, 4.9 kg, and an average consumption of 8.8 W. The maximum neutron count rate was about 1 kcps.

Zhao et al. [33] documented a low-power and low-size γ/n discrimination ES targeting environmental radiation monitoring using a CLYC detector coupled to a PMT. Their design was compliant with the IP66 standard in a 3.34-L volume while operating with an average power draw of 3.5 W. Although no pile-up recovery was implemented, a maximum count rate of 270 cps was certified for events with energy higher than 1.25 MeV.

The latest commercial γ/n dose-rate meter from ThermoFischer Scientific (RadEye GN+) [10] is based on a CLYC scintillator and exhibits outstanding low-SWaP features: 0.18 L and 0.16 kg, while offering a battery lifetime of at least 300 h. However, the maximum neutron count rate is 1 kcps with a maximum gamma energy range of 1.3 MeV: reduced detection limits for typical mixed radiation field applications.

III. Contributions

By identifying the current trends in low-SWaP applications that require high count-rate γ/n discrimination, we propose an ES solution featuring the following contributions.

- Gamma/neutron discrimination with overall accuracy higher than 98% at count rates greater than 200 kcps.
- Continuous pile-up recovery for more than four successive events and dead-time as low as 2.5 μs.
- Low SWaP using an off-the-shelf integrated CLYC detector and a low-end FPGA on a custom hardware platform.
- Discrimination performed by a compressed machine learning model trained with a ground-truth dataset, validated by the expected CLYC scintillator behavior under mixed radiation fields.

The accuracy, count rate, and dead-time have been validated with a synthetic dataset of piled-up events, consisting of

continuous long traces from experimental pulses recorded with a CLYC detector in a γ/n mixed radiation field.

IV. ES Design

The ES is a self-contained autonomous device composed of an off-the-shelf CLYC detector and a custom hardware platform housed in a 3-D-printed case. The enclosure was printed using acrylonitrile butadiene styrene and was optimized to minimize the overall volume while providing a robust lightweight structure to safely pack the whole system for diverse use-case scenarios. The compact commercial SiPM-based (SENSL ArrayC-60035-4P) detector features a temperature-compensated power supply and an integrated preamplifier with low-SWaP characteristics [34].

The hardware platform includes an analog interface to connect the detector and sensors with the FPGA, responsible for the digital signal processing. A microcontroller unit (MCU) provides optional ports for external peripherals and commands the clocking and power management system. An external serial interface (UART) is used to assess the system performance through a direct connection to a computer using a USB port. Alternatively, data can also be stored in a nonvolatile memory for offline analysis. The system power supply can be sourced either from an external Li-ion battery pack or directly from the USB port. Fig. 1(a) shows the block diagram of the most relevant ES components, including the detector and the hardware platform. Fig. 1(b) shows the physical appearance of the ES as used in the experiments.

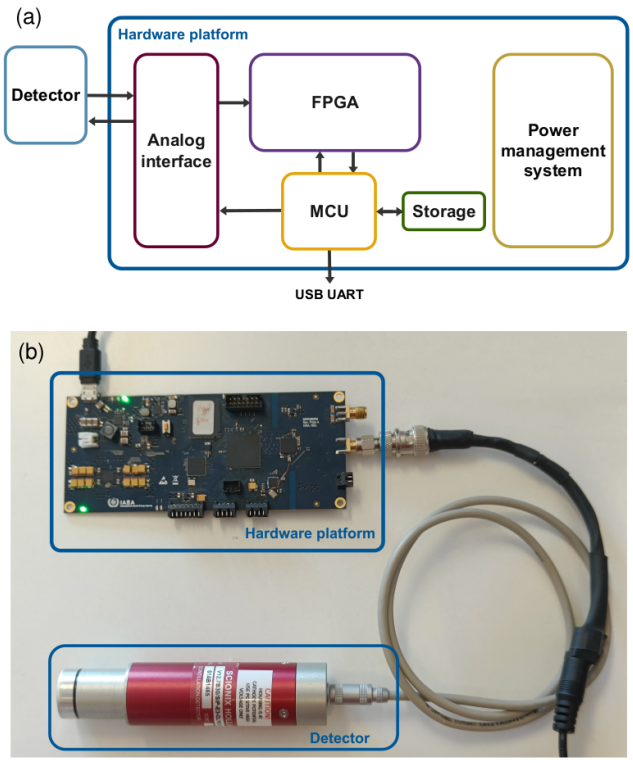

Fig. 1. ES design, featuring an off-the-shelf SiPM-based CLYC detector and a custom low-SWaP hardware platform. (a) Block diagram highlighting the most relevant elements of the platform. (b) Photograph of the system components as used in the experiments.

A. Hardware Design

The hardware platform was designed by the Nuclear Science and Instrumentation Laboratory (NSIL), International Atomic Energy Agency (IAEA). It has been devised with multifunctionality in mind, that is, its deployment in different field applications, targeting low-power operation [35]. It accommodates various subassemblies, including: i) input power monitoring and management subsystem; ii) digital interface controller; iii) digital signal-processing block; iv) data storage; and v) analog interface domain.

The backbone of the input power monitoring and management subsystem is based on a switched-mode Li-Poly battery charger with integrated system power path management. It performs several safeguard tasks such as over-voltage, over-current, and short-circuit protection.

Regarding low-power features, the hardware platform integrates two additional regulators and four load switches. These load switches can be used to shut down each subassembly providing a configurable system power consumption management.

The digital interface controller is based on a Cortex-M7 MCU, used to provide connectivity toward different peripherals through a USB port. In addition, tight control over different power domains is achieved via the input/output pins, managing the platform's low-power operation modes by enabling/disabling individual load switches. It is worth mentioning that the MCU lays in a separate power domain, automatically switched on when the external power supply is connected.

The platform offers three additional serial ports to connect other peripherals, such as an altimeter, a global navigation satellite system receiver, and a short-range radio for wireless communication. This modular architecture expands its potential use in different applications, such as deployment on UAV remote radiation monitors and portable dosimeters.

The main component in the digital signal-processing block is a low-end Artix-7 Series FPGA from AMD. This FPGA does not contain any hardcore processors, demanding significantly less power consumption compared to similar system-on-chip (SoC) solutions. In addition, this approach does not require a heat sink, relaxing the platform enclosure constraints. Therefore, the FPGA-based design is crucial for the SWaP reduction of the ES.

The platform also allows storing data onboard using an 8-GB eMMC memory. This increases the field deployment reliability when packet loss occurs in the wireless data link, allowing measurements to be recorded for further offline readout.

The analog interface domain encompasses the analog front end (AFE), a digitally controlled bias power supply for SiPMs, and an analog interface for the temperature sensor attached to the detector. The AFE enables direct coupling of the SiPM output using an impedance-matched input. These signals are further converted to the digital domain using a 14-bit 250 Msps analog-to-digital converter (ADC) with differential and integral nonlinearities of 1.7 and 4.5 LSB, respectively.

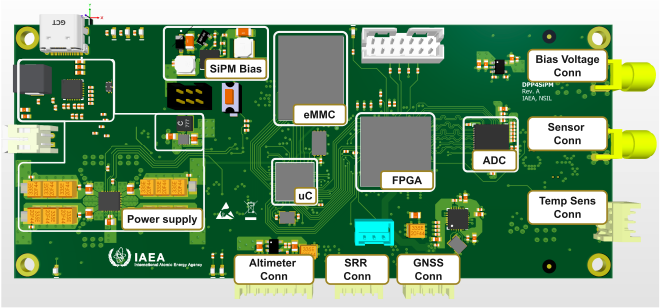

Fig. 2. Two-dimensional render of the platform with marked subassemblies.

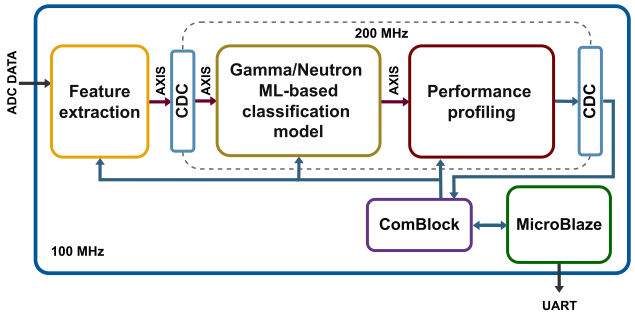

Fig. 3. FPGA firmware design integration. Feature extraction block captures a set of samples from the leading edges of events digitized by the system's ADC. The ML-based classification model performs the γ/n discrimination and feeds the performance profiling block for continuous operation benchmarking. ComBlock and MicroBlaze are used for parameter setup and results in visualization via a serial (UART) interface. The AXI stream protocol was used through the critical online processing chain: from the feature extraction block output to the performance profiling output.

The AFE features a resolution of 122 μV/LSB with a 2-V_{pk-pk} analog input range, and a measured baseline noise rms of 1.78 LSB (217 μV). As the integrated CLYC detector output saturates at 1.5-V_{pk-pk} with event energies of 6.1 MeV, the calibrated ES dynamic range is 67.8 dB with an expected energy resolution of 496 eV/LSB (4.07 keV/mV). Moreover, CLYC linearity is guaranteed below 8 MeV when coupled to an SiPM [36], resulting in a constant gain through all the detection ranges.

The 2-D render of the platform is shown in Fig. 2, emphasizing the most relevant subassemblies.

B. FPGA Firmware

The FPGA performs data acquisition and online digital signal processing in the ES. An overall view of the firmware design integration is shown in Fig. 3, on which the most relevant elements are presented. This implementation was developed using custom VHDL modules for the signal-processing stages. The main data flow is managed by standard AXI4-Stream interfaces [37], enabling automatic handshaking and flow control through the processing chain. Following this methodology, the output from the detector is continuously digitized by the ADC at 100 Msps and streamed into the feature extraction block, where the most relevant data are captured and transferred out. Subsequently, a clock-domain-crossing (CDC) stage is used to translate these samples from the low-power 100-MHz clock domain to a faster 200-MHz high-performance processing section. Here, the ML-based classification model computes the event type (γ/n) in less than 1 μs and feeds the performance profiling block with the inferred outcome. At this stage, the γ total count, n total count, γ pulse rate, and n pulse rate are continuously computed. Finally, another CDC transfers these measurements to the Communications Block (ComBlock) input registers [38], used to periodically gather the performance metrics and set the working parameters of the feature extraction, ML-based classification system, and profiling blocks. An AMD MicroBlaze soft-processor running at 100 MHz is used to connect the ComBlock with an external serial interface (UART), handled by a USB-UART bridge coded in the onboard MCU. This port is used to output the performance metrics once a second and might also serve as a drop-in solution for remote monitoring and diagnostics interface [39]. The CDC stages are implemented with AXI4-Stream FIFO blocks, simultaneously serving as a pipeline buffer for online processing and as a safe clock domain translator.

Power consumption optimization in the FPGA firmware design consists of accelerating only the ML inference and performance profiling blocks at 200 MHz while keeping the remaining components at 100 MHz. The ADC sampling frequency was also set to 100 Msps instead of the maximum 250 Msps available, which does not affect the γ/n discrimination capabilities and contributes to dynamic power reduction in both the ADC and FPGA. At 50 Msps, the feature extraction system was not able to accurately determine the time of arrival of the pulses, and the ML model complexity was not reduced proportionally, whereas, at 250 Msps, the ML model complexity increased significantly without providing relevant accuracy improvement, while increasing dead-time and power consumption. Thus, the feature extraction and the ML model benefited from the 100-Msps sampling rate.

V. Data Preparation

The experimental data for this work were recorded at the Neutron Science Facility (NSF), IAEA Laboratories, in Seibersdorf, Austria. Two neutron sources were used: an americium–beryllium (AmBe) passive source and a pulsed deuterium–deuterium (DD) active neutron generator [40]. A Cs-137 γ source was placed beside the CLYC detector during most of the measurements to provide valuable energy calibration data. Strong thermalization of the fast neutrons was caused by the DD operation housing module (high-density polyethylene shield) and the surrounding room objects. Fig. 4 shows a close-up view of the CLYC detector placed beside the beam port of the DD generator and the Cs-137 source, as used during the data recording session at the NSF. A lead block was positioned between the DD output and the detector to maximize the γ counts from the calibration source. Similarly, fast neutrons from the AmBe source suffered some scattering due to the surrounding matter in the room, even if no polyethylene shield was present during the experiment.

The detector's output signal port was coupled through a 50-Ω RG-316 cable to a CAEN DT5761 digitizer to obtain

Fig. 4. Integrated CLYC detector placed at the output window of the DD generator, separated by a lead block to maximize the Cs-137 γ counts for calibration.

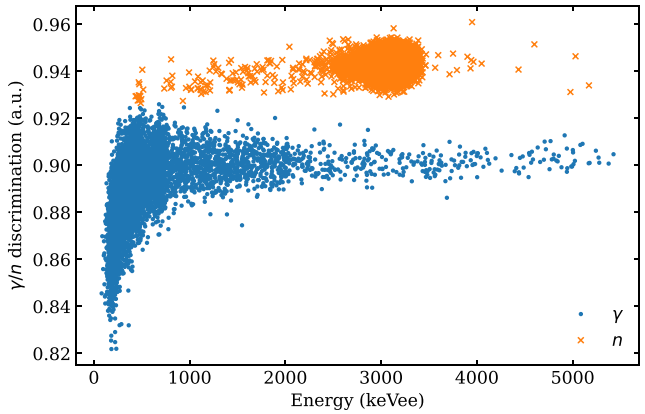

Fig. 5. Labeled γ/n discrimination matrix of the experimental data, including events from both the DD generator and the AmBe source. The fast neutron spectrum in gamma-equivalent energy units is merged with the thermal neutron events around 3.2 MeV, caused by the ^6Li capture reaction in the scintillator.

a dataset sampled at 4 Gsps with 10-bit amplitude resolution and virtually no dead-time [41]. More than 100 000 raw pulses were recorded for further offline preprocessing.

A. Data Wrangling

Data wrangling was performed using the Pandas library (version 1.5.2) to filter out undesirable events in the dataset, such as saturation or noise. Piled-up pulses were also removed, leaving only single clean traces to further synthesize piled-up events under controlled conditions. This cleanup process consisted of consecutive stages to deal with the different types of spurious pulses.

1) The dataset was downsampled from 4 Gsps to 100 Msps, matching the sampling rate of the low-SWaP ES.
2) Individual events from multiple data-recording sessions were mixed and indexed in a single data frame.
3) Baseline restoring was applied with the average of the initial 20 samples of each pulse/event. Amplitude polarity was flipped to obtain a set of positive pulses with their first samples located in an average of zero amplitude units.
4) High-energy amplitude-saturated events were removed.
5) Pile-up rejection was performed with pulse shape correlation using an efficient method to distinguish single events from distorted traces [42]: pulses with correlations lower than 98% were dropped. Because gamma and neutron pulses exhibit similar shape features in the SiPM-based detector, the same correlation template was used in both cases.
6) A second cleaning stage was performed using residuals analysis, on which the normalized average pulse shapes (γ/n) were subtracted from each amplitude-normalized trace. Once the sum of the residuals was computed, the outliers (most likely distorted pulses) were discarded. Less than 100 traces (<0.1%) were affected by this cleanup, promoted by low-energy events piled-up over high-energy pulses that passed undetected by the correlation stage.
7) Energy calibration was performed over the clean dataset, computed as the amplitude of the pulse, referenced to the well-known Cs-137 gamma 661.6-keV photopeak and the characteristic CLYC thermal neutron cluster close to 3.2 MeV in gamma energy-equivalent units (MeVee) [43].

Then, individual event tagging (gamma/neutron) was carried out using a frequency-based γ/n classification algorithm, based on a ground-truth labeling method supported by the interactions of the CLYC crystal under mixed radiation fields [43]. Fig. 5 depicts the γ/n discrimination matrix of the labeled events from both the DD generator and the AmBe source, shown as a 2-D histogram of discrimination index versus energy. The thermal neutron cluster is merged in the fast neutron spectrum, close to 3.2 MeV due to the Q-value of the ^6Li reaction channel [44]. Thus, fast and thermal neutrons have been seamlessly discriminated from gamma events, as conducted in [45].

The data wrangling process resulted in a clean dataset with γ/n tags containing original (raw) pulse traces, calibrated in keVee units. These events were used to synthesize piled-up pulses under controlled conditions: individual event type (γ/n), energy in gamma-equivalent units, and time of arrival. As from the nuclear reactions involving fast and thermal neutron interaction in CLYC crystals, it is well known that neutron events cannot be measured as pulses with energies less than 500 keV in the gamma energy scale (keVee) [46]. Thus, events with energy lower than this limit are disregarded in the pile-up dataset synthesis, as no γ/n discrimination is required and gamma pile-up recovery may be processed using an existing online FPGA-based method [47].

B. Pile-Up Synthesis

Although some piled-up traces were observed directly in the recorded data from the neutron sources, no ground-truth of the event type (γ/n) or time of arrival can be certainly defined. Therefore, a dataset of pile-up events was created using individual (tagged) events from the clean dataset obtained through the process presented in Section V-A, following the approach used in [22] and [48]. Several types of event combinations resulted from this process: $\gamma+\gamma$, $\gamma+n$, and $n+n$, with traces

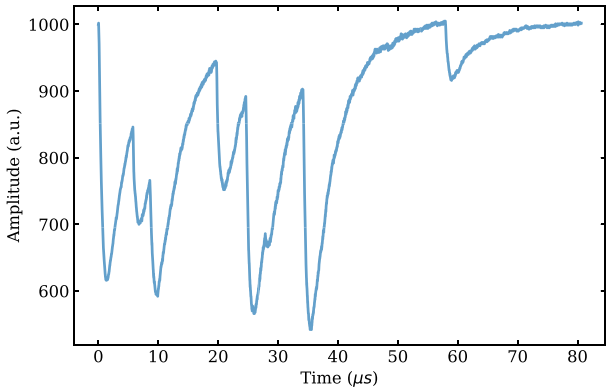

Fig. 6. Synthetic piled-up trace with eight $n + \gamma$ interleaved events from experimental data. Minimum time between pulses set to 2.5 μs and average event rate $\lambda = 200$ kHz.

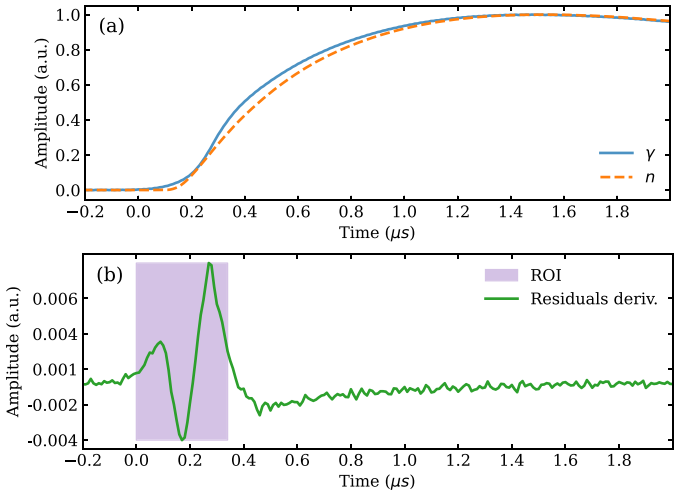

Fig. 7. (a) Gamma and neutron leading edges of normalized average pulse shapes. (b) Gamma/neutron pulse shape residual derivatives and ROI used to define the most relevant samples for the ML-based discrimination (350 ns).

containing up to eight consecutive pulses. The distribution of the time of arrival of piled-up events follows a Poisson process, modeled as an exponential distribution with rate parameter λ. Based on this, we created several combinations of traces at different average event rates: 10, 25, 100, and 200 kHz.

We also observed that a minimum separation between the time of arrival of each piled-up pulse was required to avoid losing a low-energy event in the trailing edge of a high-energy trace, caused by the long decay time of our CLYC detector. We set 2.5 μs as the minimum interpulse distance (MIPD), aiming at reducing the probability of missing an event with the aforementioned characteristics in a trace. MIPD defines the overall dead-time of the system, leaving the possibility of tuning it according to individual implementation requirements. Also, a minimum of 5% overlap between pulses was forced in the synthesis to accurately quantify the pile-up recovery of the online feature extraction stage (detailed in Section VI) in a worst case multiple pile-up scenario.

After applying these hard constraints, we assessed the Poisson behavior by fitting the time of arrival of the synthesized pulses with an exponential distribution. The fit of 60 000 piled-up traces resulted in a perfect coefficient of determination ($R^2 = 1$), with a measured pulse rate $\lambda_{\text{fit}} = (199.8 \pm 0.7)$ kHz that matches the value set for the synthesis $\lambda_{\text{set}} = 200.0$ kHz. Therefore, the established hard constraints did not influence the average rate of events nor the arrival times distribution.

Individual events were randomly picked from the clean dataset (detailed in Section V-A) and further piled-up as shown in Fig. 6, where an interleaved $n + g$ synthetic trace with MIPD $= 2.5$ μs and $\lambda = 200$ kHz is depicted.

C. Gamma/Neutron Pulse Shape Features

Traditional methods usually analyze the trailing edge of the detector traces to compute a γ/n discrimination factor. On the contrary, our solution analyzes the leading-edge features of gamma and neutron events to determine the most significant differences between both types, aiming at developing a reliable system under high-event-rate scenarios. This approach enables quick γ/n identification with fewer features, simultaneously decreasing the overall dead-time and significantly reducing the complexity of the machine learning model used for discrimination.

Offline analysis with the clean dataset from Section V-A was carried out, obtaining the amplitude-normalized average pulse representation of gamma and neutron events: the leading edges of these shapes are shown in Fig. 7(a). The residuals between the average neutron and gamma traces of Fig. 7(a) were computed bin by bin (sample by sample) to determine where the neutron and gamma traces show the highest variability. The derivative of the residual traces further highlights in a region of interest (ROI) the most significant differences in the build-up process of an event type, as shown in Fig. 7(b). The optimal ROI length was minimized to 350 ns, equivalent to the first 35 samples of an event digitized at 100 Msps. This value is the number of samples used to feed the ML γ/n classification model, as detailed in Sections VI and VII-A. Although the number of samples could be slightly reduced, a subsequent precise pulse alignment mechanism would be required, significantly impacting the overall dead-time by introducing a stochastic delay [28]. Minimizing the number of samples in the leading-edge ROI was crucial to notably improve the efficiency of the system compared to other offline and online ML-based solutions, which typically analyze the complete pulse shape or even multiple successive events [17], [20], [21], [22], [23], [28].

VI. Online Feature Extraction

An online feature extraction (FE) method has been devised aiming at sourcing the ML model with the right amount of data to accurately and efficiently perform the γ/n discrimination for each incoming event from the detector.

First, the input signal's discrete second derivative is used to determine the time of arrival of an event, which is then passed through a cross-level trigger (CLT) with a constant threshold th. Similar to [18] and [27], a smoothed discrete derivative is used to enable efficient event detection despite the pile-up scenario or variation in baseline, generally present in

high-event-rate conditions. We also mimic their bandpass filtering by adding consecutive derivative and smoothing stages, obtaining a smoothed second-order derivative (SSD).

Since the processing is carried out in the digital domain, discrete implementations are mandatory to model the online computations. The discrete derivative $w(n)$ was approximated by the central-difference operator detailed in the following equation:

$$w(n) = \frac{1}{2}[x(n) - x(n-2)]. \tag{1}$$

Its bandpass frequency response, recognizable from the transfer function (TF) in the $\mathcal{Z}$ domain shown in the following equation, avoids the amplification of undesired high-frequency components:

$$H_d(z) = \frac{1}{2}\left(1 - z^{-2}\right). \tag{2}$$

Smoothing is further performed by an N-point filter based on a moving-average implementation. We omitted the amplitude normalization $1/N$ to keep integer representation throughout the entire feature extraction process. This smoothing operator was chosen to avoid fine-tuning the filter for specific pulse shapes of different detectors, resulting in a reliable generic FE solution for multiple leading-edge timing parameters. Moreover, the discrete-time filter $y(n)$ is recursively defined in the computationally efficient implementation shown in (3) [49], whereas its $\mathcal{Z}$-domain TF is represented in (4)

$$y(n) = y(n-1) + w(n) - w(n-N) \tag{3}$$

$$H_s(z) = \frac{1 - z^{-N}}{1 - z^{-1}}. \tag{4}$$

The N averaging parameter was set to 64 after successive tuning iterations, providing a good noise level reduction while preserving the required features for leading-edge detection in pile-up scenarios [50], [51].

Being a linear time-invariant system, the TFs used to compute the two successive smoothed discrete derivatives (SSDs) can be combined into a single $\mathcal{Z}$-domain expression to define the feature extraction filter, shown in the following equation:

$$H(z) = H_d^2(z)H_s^2(z) = \frac{1}{4}\left[\frac{(1 - z^{-2})(1 - z^{-64})}{(1 - z^{-1})}\right]^2. \tag{5}$$

This FE filter evidences a stable and causal behavior for an infinite input trace. However, since the initial values of the delay lines z^{-2} and z^{-64} are reset at start-up in the FPGA deployment, a slight warm-up time close to 1.3 μs is required. After this short period, the SSD computation runs uninterruptedly with the data stream provided by the ADC.

Whenever the CLT is activated over the SSD signal, a time-over-threshold (ToT) counter of length M is immediately started. The ToT is used to ensure that the CLT was certainly triggered by an event instead of noise: only if the total number of samples surpassing the threshold is more than $M/2$, the event is registered. Also, the CLT functionality is disabled succeeding the end of the ToT M period, getting re-enabled only after a defined time-after-ToT (TAT), used to filter out events or noise that may incorrectly detect an event too close to each other. This safeguard introduces a deterministic dead-time that matches the MIPD from Section V-B.

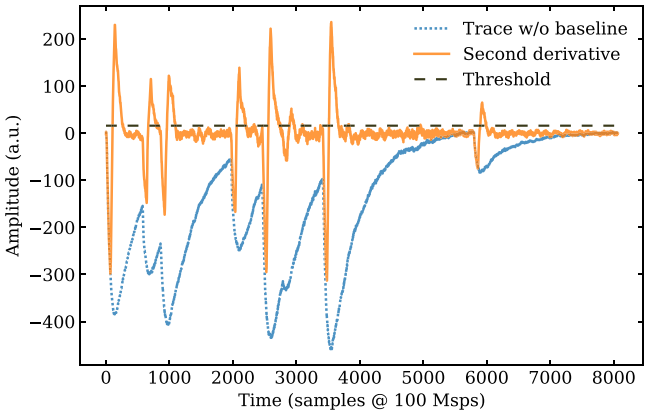

Fig. 8. Piled-up trace with eight successive events and smoothed second derivative (SSD) filter output. The baseline was removed on purpose from the original trace to facilitate the visual comparison of the plots. The SSD computation inherently removes any baseline component. The threshold for detection on SSD set to th = 16 ADC channels.

After the arrival of an event has been successfully identified by the FE system, the first R samples of its leading edge are captured and sent to the ML-based classification model. It was experimentally observed that the SSD pulses related to expected events in the original trace kept a ToT higher than 31 samples, with a threshold equal to four times the average SSD baseline (th = 16). Thus, the ToT parameter was set to $M = 64$ in the final implementation. A piled-up input trace along the computed SSD is shown in Fig. 8, depicting the actual threshold value where events may be detected only after the ToT criterion is met.

As for the residual analysis described in Section V-C, the chosen number of samples fed to the ML model is $R = 35$. In accordance, the interpulse safeguard parameter is set to TAT = 151, which summed with M and R equals to the expected overall dead-time: 250 samples at 100 MHz, or 2.5 μs. Although shorter dead-times are easily achievable by reducing TAT, the detection efficiency would not increase significantly owing to the long decay time of the CLYC.

The FE implementation was initially developed as an offline simulation in Python 3.10 and further translated to an online deployment for the FPGA using VHDL. Integer representation is used in both the Python simulation and the FPGA design, avoiding unnecessary complexity related to floating point in terms of resource utilization, while simultaneously improving throughput and latency [52].

The online feature extraction in the FPGA was verified to match with the Python simulation behavior using a VHDL test bench and further validated in hardware using AMD's Integrated Logic Analyzer (ILA), as shown in Fig. 9.

VII. Gamma/Neutron Discrimination Based on Machine Learning

From a computational perspective, a low-SWaP solution implies a reduction in power consumption during computation, storage, and memory access operations. As observed in SOA [53], [54], [55], ML-based models are computationally

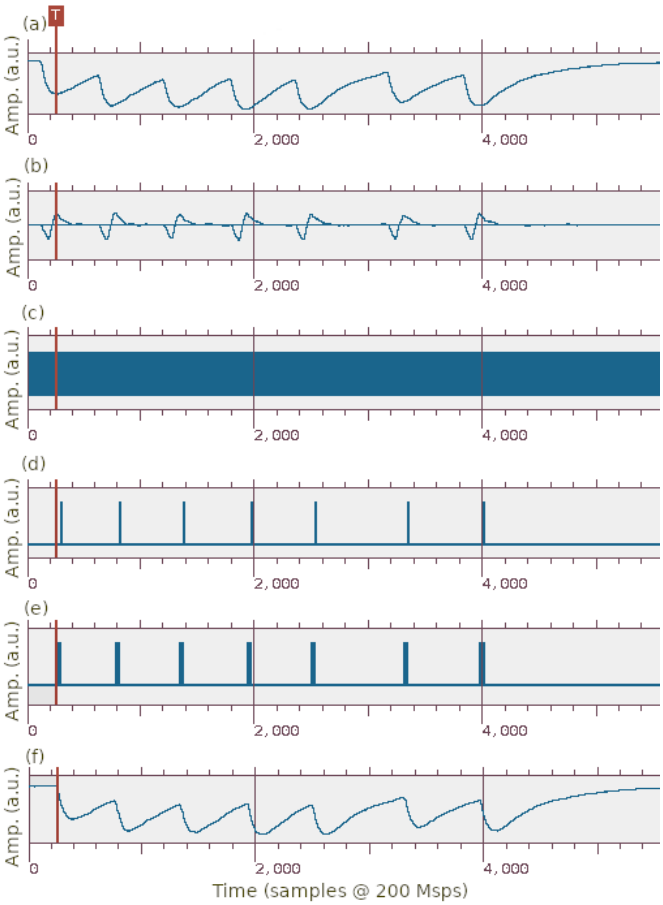

Fig. 9. Online feature extraction in the FPGA shown in the sliced view of a screenshot from AMD Vivado ILA, where the related signals are simultaneously captured using the same time base. (a) Digitized piled-up input trace from the ADC. (b) Smoothed second derivative of the input signal. (c) Valid data flag shows continuous single-clock cycle data consumption from the next processing stage at 200 MHz, sourced with the 100-MHz ADC signal. (d) End of leading-edge capture flag (AXI4-Stream TLAST). (e) Capture valid signal to fetch the first 35 leading-edge samples of the detected events (AXI4-Stream TVALID): the falling edge of this flag indicates the end of the buffering to the ML inference stage. (f) Delayed input signal, used to align the captured pulse with the deterministic latency of the feature extraction system: the leading-edge extraction starts exactly at the beginning of each event, piled-up or not.

costly, impacting the power efficiency of the overall system. Nevertheless, it has been demonstrated that these models can be implemented on edge devices without compromising the system performance [52], [56], [57]. Aiming at online γ/n discrimination on the edge, an optimized ML model was designed employing the end-to-end workflow for deploying deep neural networks on the SoC/FPGA we proposed in [58]. A good compromise between effectiveness, memory footprint, and inference time was achieved while improving power efficiency.

A. Model Training and Compression

The workflow employs three techniques to achieve efficient model compression: pruning (P), quantization (Q), and knowledge distillation (KD). P and Q are orthogonal to KD, aiding to achieve better performance and reducing the size of the model with minimum loss in accuracy [59]. P aims to

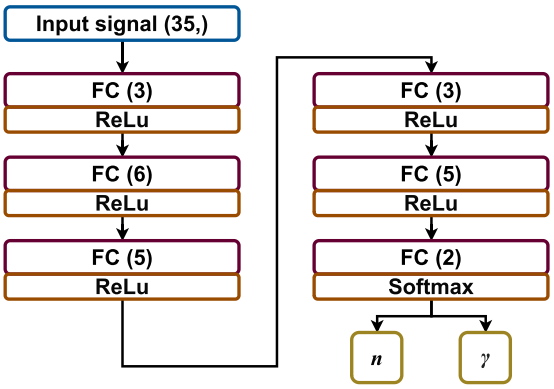

Fig. 10. Compressed architecture based on MLP for γ/n event discrimination.

decrease the number of parameters by removing neurons and connections. Q reduces the memory footprint by selecting the number of bits that represent the weights and biases. Finally, KD [60] is devoted to transferring knowledge from a teacher network to a smaller and faster target network (i.e., distilled or student), which can reproduce the teacher's behavior while being less computationally expensive. Furthermore, the workflow uses Bayesian optimization to heuristically define the hyperparameters.

To implement the workflow, the teacher and student architectures should be defined for the KD method. Furthermore, P and Q strategies are established within the student architecture because the KD learning process is performed through quantization-aware pruning (QAP) [61].

A multilayer perceptron (MLP) was trained as a teacher model to discriminate between neutron and gamma events. The MLP-based model was composed of six hidden layers chosen heuristically with 2623 parameters, exhibiting an overall accuracy of 99.0%. The input of the model consists of the first 35 amplitude samples of the leading edge of the pulse. Each hidden layer of the MLP is composed of fully connected (FC) neurons followed by rectified linear unit (ReLU) activation functions, whereas the output layer employs Softmax to provide a normalized discrimination result.

The training dataset was generated by processing the piled-up traces from Section V-B using the feature extraction method from Section VI. In total, 100 000 leading edges were generated, with a perfectly balanced distribution for each output class: 50 000 gamma and 50 000 neutron events. Events covering all the energy ranges under study were used to enable the ML model generalization capabilities to accurately infer the expected class (γ/n), no matter the average slope of the leading edges.

After the teacher model had undergone training, the student architecture was defined as an MLP with 8-bit fixed-point precision, target sparsity of 30.00%, and 217 hyperparameters distributed in six layers. After student architecture training through KD and QAP, the overall accuracy of the compressed model showed a slight decrease: from 99.0% to 98.2%. Specifically, the accuracy was 96.8% for gamma rays and 99.5% for neutrons. As can be observed, a reduction in the memory footprint was achieved through model compression while maintaining high accuracy. The compressed ML-based

architecture (or student) that performs γ/n event discrimination is illustrated in Fig. 10.

Once the compressed model was created, the hls4ml package [62] was employed to obtain a high-level synthesis (HLS) project to synthesize and export the model to the FPGA register transfer level (RTL). The resulting hardware block allows stream data transmission through dedicated interfaces optimized for low latency through the maximum parallelization of computations. Furthermore, the hardware generated for the inference process was fully implemented in the on-chip memory of the FPGA, significantly reducing the overall power consumption compared to implementations based on off-chip memory transactions mandatory for bigger ML models in low-end FPGA platforms [63], [64].

VIII. Results

The ES performance assessment and SWaP validation were conducted by quantifying experimental data of the most relevant indicators under diverse conditions. The measured performance metrics were count rate, dead-time, discrimination accuracy, and pile-up recovery capabilities. The FPGA resource utilization is also reported, highlighting the efficiency of the ML model and the overall firmware design. In terms of SWaP, the evaluated features were weight, volume, and power consumption. The experiment descriptions, their outcomes, and quantitative comparisons with the SOA are presented next.

A. Performance Assessment

Individual performance parameters are quantified and discussed, stressing the online capabilities of the system for high-event-rate scenarios. These parameters were experimentally measured with the performance profiling block in the FPGA, updating the results at 1 Hz through the UART port. Two tests have been devised with the CLYC detector aiming at providing a broad verification method as follows: evaluating maximum performance figures with synthesized piled-up events and using experimental raw data to validate the system behavior in real conditions. Besides, a complementary assessment was carried out with a fast organic detector pulse shape to showcase the potentiality of our system to work with other types of scintillators.

1) Verification With Synthetic Traces: Aiming at pushing the performance of the system to the design limits in terms of count rate and dead-time, a new set of piled-up traces was synthesized following the methodology described in Section V-B. This validation dataset was never used in the ML training or testing stages. Three types of pulse combinations were created on each trace, with up to 16 piled-up sequential events: all-gamma events (γ), all-neutron events (n), and alternated $\gamma + n$. These traces were stored and sequentially reproduced in three 24-h individual sessions using an arbitrary waveform generator (AWG) to stimulate the input of the system's AFE. To add variability to the traces, the AWG was set to sum 1% of white Gaussian noise atop the traces, reducing the probability of generating equal events during the long period of experimental validation in hardware. The following parameters were assessed at the intermediate stages and verified at the end of the individual tests: total gamma counts, total neutron counts, gamma count rates, and neutron count rates. These unique traces provided a solid reference for the total expected counts and rates after event discrimination.

The measurements revealed a worst case scenario for the feature extraction false-negative rate: 74 counts out of 1 million expected events were missed (74 ppm), equivalent to a sensitivity higher than 99.9%. Moreover, the ML experimental γ/n discrimination overall accuracy matched the expected model performance with three significant digits: 98.2%. Since the dataset used for the validation was generated with an average event rate of 200 kHz and a dead-time of 2.5 μs, the 74 ppm proportion of missed pulses is negligible, demonstrating the actual performance of the ES as a whole with peak count rates up to 400 kcps or $(2.5\ \mu\text{s})^{-1}$.

Moreover, aiming at measuring the maximum count rate of the digital signal-processing system and verifying the absolute minimum dead-time, the pulse shape of a gamma event from a fast organic scintillator was synthesized using the double-exponential model with a single 50-ns decay time [65]. This shape was continuously reproduced with the AWG at diverse pulse rates summed with 10% of white Gaussian noise, while plugged into the AFE input of the hardware platform.

The TAT safeguard parameter (from Section VI) was removed in the FPGA configuration for this test, unlocking the maximum event rate of the system for fast detectors. The highest measured figures yielded 1.01 Mcps with an overall dead-time lower than 995 ns and an average 129 ppm error in counting sensitivity: on pair with the event rate limit reported by [28] and the lowest dead-time among the reviewed SOA. Values higher than 1.01 Mcps resulted in a drastic number of missed events, establishing the maximum detection rate unless further optimization is performed.

These impressive metrics evidence the advantages of the FPGA reconfigurability and the flexibility of our ES design when paired with faster scintillators. Naturally, the ML model requires proper training for different pulse shapes to accurately discriminate γ/n on each type of detector.

2) Verification With Experimental Data: A verification stage with the experimental data was performed using raw traces recorded at the NSF with the CLYC detector. Three continuous traces were streamed sequentially into the ES AFE using the AWG at 100 Msps. Since no pulse cleanup was carried out, pile-up, saturation, and other perturbations were present in the traces. The data streams were selected as follows: i) a γ-only Cs-137 source placed beside the detector; ii) the AmBe neutron source along with the Cs-137 calibration source; and iii) the DD generator at 100% duty cycle with a lead shield in front of the detector, exhibiting higher neutron activity due to the thermalization caused by the surrounding high-density polyethylene.

Following the quantification approach from [66], the AmBe and DD γ/n relative abundances showed agreement with ii) and iii), respectively. Regarding i) the false-positive rate for neutrons remained below 2.54%, matching the expected gamma accuracy of the ML model. Similar to [9] and [28], Fig. 11 shows the online count rates during this verification

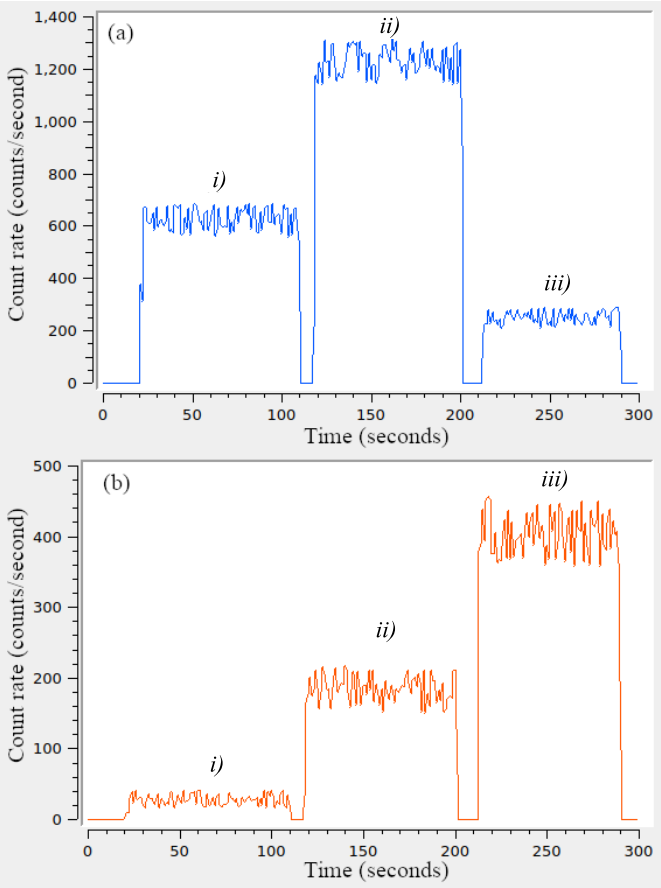

Fig. 11. Screenshot from SerialPlot showing the count rates computed by the performance profiling block in the ES. Event quantification rates are shown for (a) gamma and (b) neutrons, as discriminated online by our system. Three stimuli were tested using experimental traces recorded at NSF with the CLYC detector: i) γ-only Cs-137 source; ii) AmBe source plus Cs-137 calibration source; and iii) DD generator plus lead shield. Stimuli were reproduced as a continuous stream at 100 Msps with an AWG plugged directly into the ES AFE input.

stage, computed by the performance profiling block and transmitted via UART to the SerialPlot application [67].

B. FPGA Resource Utilization

Although a low-end FPGA part was chosen for the ES, the firmware design (from Section IV-B) demanded less than one-third of the total resources available. Table I summarizes the Vivado postimplementation report, where each relevant design element is described, including the overall occupation of the firmware. Some components such as the UART interface controller are not representative and were not included in the table.

C. SWaP Validation

The SWaP parameters were measured using common laboratory instruments with the reported digit significance, leading to the following results: the external case volume is 0.56 L, while the overall weight of the system is 0.41 kg. Power consumption was measured using an off-the-shelf USB power meter during a processing stress-test session, evidencing an average draw of 1.49 W when performing online discrimination on a stream of piled-up events at the maximum count rate.

TABLE I
FPGA RESOURCE UTILIZATION PER FIRMWARE DESIGN ELEMENT

Design element	LUTs	Registers	BRAMs	DSPs
Available resources	20800	41600	50	90
Overall utilization	30.4%	17.4%	24.0%	17.8%
Feature Ext. + CDC	8.6%	3.9%	8.0%	0.0%
MicroBlaze + memory	6.4%	2.5%	16.0%	3.3%
Performance profiling	0.9%	1.5%	0.0%	0.0%
ML classification	11.2%	7.0%	0.0%	14.4%
ComBlock	0.8%	0.7%	0.0%	0.0%

TABLE II
SWAP COMPARISON OF RECENT CLYC-BASED GAMMA/NEUTRON DISCRIMINATION SYSTEMS, INCLUDING SAW AND WAP SCORES

Work	**S** (L)	**W** (Kg)	**P** (W)	**SaW**	**WaP**
Ours	0.56	0.41	1.5	0.13	1.00
Huang [9]	4.08	4.9	8.8	0.00	0.01
Zhao [33]	3.34	-	3.5	-	-
[1]**Mesick** [7]	-	7.0	14	-	0.01
[1]**Hardgrove** [8]	-	3.4	9.6	-	0.02
Thermofischer [10]	0.18	0.16	-	1.00	-

D. Comparison With SOA

A broad comparison of the proposed system's SWaP and performance metrics with SOA is carried out, providing overall quantitative scores for each indicator.

In terms of low-SWaP measurables, Table II summarizes the parameters of recent CLYC-based γ/n discrimination implementations optimized for the same SWaP reduction goal. Size is measured in liters, weight in kilograms, and power consumption in Watts. An overall SWaP score cannot be computed due to the lack of information in one field for most of the reference works. In this regard, size and weight score (SaW) as well as weight and power (WaP) draw score were defined individually: SaW is the inverse of the product of weight and size (1/SW), while WaP is computed as the inverse product of WaP (1/WP). In both cases, the highest value serves as the reference for normalization and represents the best score in the range (0, 1).

Our proposal exhibits the highest WaP score and the lowest power consumption among the compared works. Although [10] evidences the smallest and lightest housing, the lack of performance may be a deciding factor in demanding applications: their maximum neutron count rate is 200× lower than ours, while the maximum detectable gamma energy is about one-fourth our limit.

Regarding performance metrics, our system was contrasted with recent developments targeting online γ/n discrimination in high-event-rate scenarios. A numerical score for pile-up rejection or recovery (PuP R/R) features is proposed due to

[1]These systems were optimized for reliability in outer-space environments, commonly requiring rugged (heavier and bigger) housings.

TABLE III
PERFORMANCE INDICATORS AND OVERALL PERFORMANCE SCORE

Work	CR (kcps)	1/DT (s^{-1})	Acc (%)	PuP R/R	OP
Ours	200	400×10^3	98.2	1.00	1.00
[2]**Michels** [28]	1100	130×10^3	98.2	0.50	0.89
[3]**Wen** [29]	25	375×10^3	100	1.00	0.12
[4]**Cruz** [68]	2000	83.3	100	0.50	0.01
[5]**Astrain** [32]	79	20×10^3	98.0	0.50	0.01

their relevance in high-count-rate applications, as detailed in the following categories. Score 1.00 is assigned to continuous R/R, where a system can continuously recover or reject events, no matter the total amount of successive piled-up pulses; score 0.50 relates to implementations on which up to four piled-up pulses can be recovered or rejected; score 0.00 is assigned to systems without R/R capabilities. Equal grading is designated to rejection and recovery on each category since the impact of rejection will be reflected in the OP of the system.

Table III and Fig. 12 summarize the quantitative performance features of each compared work: minimum sustained count rate (CR), inverse of maximum dead-time (1/DT), discrimination accuracy (Acc), and pile-up rejection or recovery capabilities (PuP R/R) rating. An OP score is also proposed to provide a single indicator for the systems, computed as the product of the evaluated parameters, further normalized in the range (0,1) with respect to the highest score. This metric provides a global insight into the total performance involving all the measured aspects. The parameter values in Fig. 12 are normalized to facilitate the visualization in a unit circle representation.

Notwithstanding the impressive low SWaP metrics achieved, our implementation also exhibited the best OP score and the lowest dead-time among the evaluated systems, evidencing the high efficiency of the proposed solution.

IX. DISCUSSION AND CONCLUSION

Our solution demonstrated to be a low-SWaP ES suitable for γ/n discrimination with 98.2% overall constant accuracy at event rates higher than 200 kcps. The feature extraction mechanism matched with an efficient ML model in a pipeline guaranteed online operation achieved uninterrupted pile-up recovery and ensured a total dead-time lower than 2.5 μs with a detection sensitivity higher than 99.9%. These characteristics, validated with a synthetic piled-up dataset from experimental pulses, position our proposal in an overall high-performance range similar to that of the benchtop solution from [28], with the added portability optimizations enabled by the low-SWaP parameters described in [9].

[2]Michels et al. [28] were able to discriminate two piled-up events only. Triple or quadruple pile-ups could not be individually distinguished.

[3]Wen and Enqvist [29] reported several count-rate and accuracy values: the highest accuracy/count-rate pair was chosen to maximize their OPS score. Their accuracy metric was defined as the system's sensitivity to neutrons, not the overall accuracy. Continuous pile-up rejection is used in this work.

[4]For Cruz et al. [68], discrimination accuracy is assumed as 100%, since no value was reported besides mentioning they obtained the expected γ/n spectra. Dead-time includes pulse processing and emissivity profile reconstruction.

[5]Astrain et al. [32] demonstrated pile-up recovery of up to two successive events.

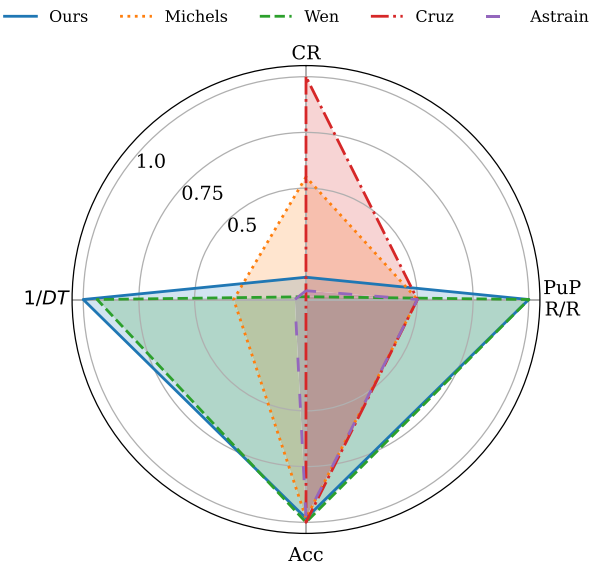

Fig. 12. Performance comparison of online γ/n discrimination systems, based on measurements for the highest value per category. Parameters normalized within the (0,1) range, being the highest value of the top performance indicator. CR defines the minimum γ/n discrimination sustained rate at which the reported overall accuracy is reached. 1/DT is the inverse of the maximum dead-time of each implementation. Acc represents the overall accuracy reported at the evaluated count rate. PuP R/R is a score based on the implemented pile-up rejection or pile-up recovery method. Our system is compared against Michels et al. [28], Wen and Enqvist [29], Cruz et al. [68], and Astrain et al. [32].

Our setup is also the first, among the most recent literature, to combine low-SWaP and online discrimination features into an ES capable of continuous pile-up recovery based on a CLYC detector. We also outperform in terms of γ/n count rate and maximum energy range the commercial portable gamma/neutron active dosimeter from [10]. SWaP reduction was achieved with a small and light, yet robust enclosure, housing an ES designed with power consumption optimizations at multiple layers: a hardware platform, an FPGA firmware, and an ML model.

This implementation is complementary to existing γ/n discrimination methods, relying on precise data tagging for the ML model training and validation. Employing a solid labeling method in combination with an optimized ML model was crucial to ensure an efficient online solution with high discrimination accuracy. We also demonstrated the flexibility of our system by adopting another detector, where keeping reduced SWaP features did not necessarily compel a performance compromise for pulses with much faster decay times. Thus, other scintillators with γ/n discrimination capabilities such as CLLB, NaI(Tl+Li), CLLBC, TLYC, and EJ276 could potentially benefit from this development.

Using this work as a starting point, γ/n spectroscopy under high count rate and pile-up scenarios can be performed by employing any known technique for γ as well as n methods [47], [69], further appending the label computed by our system. Also, proper gamma and neutron total ambient dose equivalent rates can be conducted using standard calibration methods [70], [71].

REFERENCES

[1] T. Quist and E. Cazalas, "Design, modeling, and initial testing of a multi-volume neutron spectrometer," *Radiat. Meas.*, vol. 166, Aug. 2023, Art. no. 106982.

[2] M. M. Bourne, C. Mussi, E. C. Miller, S. D. Clarke, S. A. Pozzi, and A. Gueorguiev, "Characterization of the CLYC detector for neutron and photon detection," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 736, pp. 124–127, Feb. 2014.

[3] G. Shen et al., "Using energy particle detection technology on the Tiangong's space station's wentian laboratory cabin module," *Aerospace*, vol. 10, no. 4, p. 373, Apr. 2023.

[4] D. D. S. Coupland, L. C. Stonehill, K. E. Mesick, and J. P. Dunn, "The SENSER CLYC experiment," in *Proc. IEEE Nucl. Sci. Symp., Med. Imag. Conf. Room-Temp. Semiconductor Detect. Workshop (NSS/MIC/RTSD)*, Oct. 2016, pp. 1–4.

[5] J. Köhler et al., "Measurements of the neutron spectrum in transit to Mars on the Mars science laboratory," *Life Sci. Space Res.*, vol. 5, pp. 6–12, Apr. 2015.

[6] D. Ölçek et al., "Development of a compact scintillator-based gamma-ray neutron detector for terrestrial and space applications," *J. Instrum.*, vol. 19, no. 1, Jan. 2024, Art. no. C01030.

[7] K. Mesick et al., "Development of a low-resource combined gamma-ray and neutron spectrometer for planetary science," in *Proc. Small Satell. Conf.*, 2020, pp. 1–8.

[8] C. Hardgrove et al., "The lunar polar hydrogen mapper CubeSat mission," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 35, no. 3, pp. 54–69, Mar. 2020.

[9] Y. Huang et al., "Development of a CLYC-based wide dose rate range portable neutron-gamma detector," *Social Sci. Res. Netw.*, pp. 1–17, Feb. 2024.

[10] Thermofischer. *RadEyeTM GN+ Gamma Neutron Pager*. Accessed: Mar. 7, 2024. [Online]. Available: https://www.thermofisher.com/order/catalog/product/4250631

[11] A. Barzilov and M. Kazemeini, "Unmanned aerial system integrated sensor for remote gamma and neutron monitoring," *Sensors*, vol. 20, no. 19, p. 5529, Sep. 2020.

[12] J. Hartman, A. Barzilov, and I. Novikov, "Remote sensing of neutron and gamma radiation using aerial unmanned autonomous system," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Oct. 2015, pp. 1–4.

[13] A. Cirillo and M. Caresana, "Calibration of a large-size wide-range neutron spectrometer," *Eur. Phys. J. Plus*, vol. 137, no. 7, p. 774, Jul. 2022.

[14] B. Fersch et al., "A dense network of cosmic-ray neutron sensors for soil moisture observation in a highly instrumented pre-alpine headwater catchment in Germany," *Earth Syst. Sci. Data*, vol. 12, no. 3, pp. 2289–2309, Sep. 2020.

[15] M. C. Recker, E. J. Cazalas, J. W. McClory, and J. E. Bevins, "Comparison of SiPM and PMT performance using a $Cs_2LiYCl_6 : Ce^{3+}$ (CLYC) scintillator with two optical windows," *IEEE Trans. Nucl. Sci.*, vol. 66, no. 8, pp. 1959–1965, Aug. 2019.

[16] N. Dinar, D. Celeste, M. Silari, V. Varoli, and A. Fazzi, "Pulse shape discrimination of CLYC scintillator coupled with a large SiPM array," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 935, pp. 35–39, Aug. 2019.

[17] J. Lu et al., "Pulse-shape discrimination of SiPM array-coupled CLYC detector using convolutional neural network," *Appl. Sci.*, vol. 12, no. 5, p. 2400, Feb. 2022.

[18] E. B. Johnson, C. Whitney, S. Vogel, J. F. Christian, K. Holbert, and P. Chandhran, "High event rate, pulse shape discrimination algorithm for CLYC," in *Proc. IEEE Int. Symp. Technol. Homeland Secur. (HST)*, Apr. 2015, pp. 1–7.

[19] Y. Ma et al., "A method for discriminating neutron and gamma waveforms based on a comparison of differences between pulse feature heights," *J. Radioanal. Nucl. Chem.*, vol. 333, no. 1, pp. 375–386, Jan. 2024.

[20] K. Zhao, C. Feng, S. Wang, Z. Shen, K. Zhang, and S. Liu, "N/γ discrimination for CLYC detector using a one-dimensional convolutional neural network," *J. Instrum.*, vol. 18, no. 1, Jan. 2023, Art. no. P01021.

[21] E. Doucet et al., "Machine learning n/γ discrimination in CLYC scintillators," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 954, Feb. 2020, Art. no. 161201.

[22] C. Fu, A. Di Fulvio, S. D. Clarke, D. Wentzloff, S. A. Pozzi, and H. S. Kim, "Artificial neural network algorithms for pulse shape discrimination and recovery of piled-up pulses in organic scintillators," *Ann. Nucl. Energy*, vol. 120, pp. 410–421, Oct. 2018.

[23] J. Griffiths, S. Kleinegesse, D. Saunders, R. Taylor, and A. Vacheret, "Pulse shape discrimination and exploration of scintillation signals using convolutional neural networks," *Mach. Learn., Sci. Technol.*, vol. 1, no. 4, Dec. 2020, Art. no. 045022.

[24] M. Gelfusa et al., "Advanced pulse shape discrimination via machine learning for applications in thermonuclear fusion," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 974, Sep. 2020, Art. no. 164198.

[25] I. Morad, M. Ghelman, D. Ginzburg, A. Osovizky, and N. Shlezinger, "Model-based deep learning algorithm for detection and classification at high event rates," *IEEE Trans. Nucl. Sci.*, vol. 71, no. 5, pp. 970–980, May 2024.

[26] C. H. Kim, S. Ahn, K. Y. Chae, J. Hooker, and G. V. Rogachev, "Restoring original signals from pile-up using deep learning," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 1055, Oct. 2023, Art. no. 168492.

[27] C. Yi et al., "Discrimination of piled-up neutron-gamma pulses using charge comparison method and neural network for CLYC detectors," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 1055, Oct. 2023, Art. no. 168561.

[28] N. M. Michels, A. J. Jinia, S. D. Clarke, H.-S. Kim, S. A. Pozzi, and D. D. Wentzloff, "Real-time classification of radiation pulses with piled-up recovery using an FPGA-based artificial neural network," *IEEE Access*, vol. 11, pp. 78074–78083, 2023.

[29] X. Wen and A. Enqvist, "Pulse shape discrimination of Cs_2LiYCl_6: Ce^{3+} detectors at high count rate based on triangular and trapezoidal filters," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 866, pp. 129–133, Sep. 2017.

[30] V. T. Jordanov and G. F. Knoll, "Digital synthesis of pulse shapes in real time for high resolution radiation spectroscopy," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 345, no. 2, pp. 337–345, Jun. 1994.

[31] L. M. Simms, B. Blair, J. Ruz, R. Wurtz, A. D. Kaplan, and A. Glenn, "Pulse discrimination with a Gaussian mixture model on an FPGA," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 900, pp. 1–7, Aug. 2018.

[32] M. Astrain, M. Ruiz, A. Stephen, R. Sarwar, A. Carpeo, and S. Esquembri, "Real-time implementation of the neutron/gamma discrimination in an FPGA-based DAQ MTCA platform using a convolutional neural network," *IEEE Trans. Nucl. Sci.*, vol. 68, no. 8, pp. 2173–2178, Aug. 2021.

[33] K. Zhao et al., "A low-power neutron and gamma-ray detector for environmental radiation monitoring using CLYC scintillator," *J. Instrum.*, vol. 18, no. 9, Sep. 2023, Art. no. P09043.

[34] Scionix. (Mar. 2017). *Thermal NEUTRON Detector—V12.7B30/SIP-E3-CLYC-X*. [Online]. Available: https://scionix.nl/wp-content/uploads/2017/03/V12.7B30_SIP-E3-CLYC-X.pdf

[35] *Nuclear Science and Instrumentation Newsletter*, Int. At. Energy Agency, Vienna, Austria, Feb. 1, 2021.

[36] S. West et al., "Compact readout of large CLYC scintillators with silicon photomultipler arrays," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 951, Jan. 2020, Art. no. 162928.

[37] H. Gao, H. Dai, and Y. Zeng, "High-speed image processing and data transmission based on Vivado HLS and AXI4-stream interface," in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Aug. 2018, pp. 575–579.

[38] M. L. Crespo et al., "Remote laboratory for E-learning of systems on chip and their applications to nuclear and scientific instrumentation," *Electronics*, vol. 10, no. 18, p. 2191, Sep. 2021.

[39] I. R. Morales, M. L. Crespo, and S. Carrato, "Open source remote diagnostics platform for custom instrumentation in nuclear applications," in *Applications in Electronics Pervading Industry, Environment and Society*, F. Bellotti et al., Eds., Cham, Switzerland: Springer, 2024, pp. 424–430.

[40] *Nuclear Science and Instrumentation Newsletter*, Int. At. Energy Agency, Vienna, Austria, Feb. 2022.

[41] CAEN S.P.A. (2012). *CAEN DT5761 Digitizer*. [Online]. Available: https://www.caen.it/products/dt5761/

[42] A. Cicuttin et al., "A simplified correlation index for fast real-time pulse shape recognition," *Sensors*, vol. 22, no. 20, p. 7697, 2022.

[43] I. R. Morales, M. L. Crespo, M. Bogovac, A. Cicuttin, K. Kanaki, and S. Carrato, "Gamma/neutron classification with SiPM CLYC detectors using frequency-domain analysis for embedded real-time applications," *Nucl. Eng. Technol.*, vol. 56, no. 2, pp. 745–752, Feb. 2024.

[44] N. Blasi et al., "Fast neutron detection efficiency of ^{6}Li and ^{7}Li enriched CLYC scintillators using an Am-Be source," *J. Instrum.*, vol. 13, no. 11, Nov. 2018, Art. no. P11010.

[45] I. R. M. Argueta, Jun. 2023, "Gamma and neutron tagged dataset from CLYC SiPM detector," doi: 10.5281/zenodo.8037239.

[46] M. C. Recker, "Enabling mobile neutron detection systems with CLYC," Ph.D. dissertation, Air Force Inst. Technol., Dec. 2019.

[47] V. H. Cao et al., "Flexible and low-cost FPGA-based multichannel analyzer for handheld measurement devices," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 1018, Dec. 2021, Art. no. 165808.

[48] A. Abdulaziz, J. Zhou, M. Fang, S. McLaughlin, A. Di Fulvio, and Y. Altmann, "A variational autoencoder for minimally-supervised pulse shape discrimination," *Ann. Nucl. Energy*, vol. 204, Sep. 2024, Art. no. 110496.

[49] R. G. Lyons, *Real-Time Computation of Signal Averages and Variances*. London, U.K.: Pearson, 2011, p. 799.

[50] X. Wang, Y. Chen, Q. Yin, S. Zeng, and Q. Xie, "Advantages of digitally sampling scintillation pulses in pileup processing in PET," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec. (NSS/MIC)*, Oct. 2009, pp. 144–147.

[51] Z. Gu, D. L. Prout, R. Taschereau, B. Bai, and A. F. Chatziioannou, "A new pulse pileup rejection method based on position shift identification," *IEEE Trans. Nucl. Sci.*, vol. 63, no. 1, pp. 22–29, Feb. 2016.

[52] R. S. Molina et al., "Compression of NN-based pulse-shape discriminators in front-end electronics for particle detection," in *Applications in Electronics Pervading Industry, Environment and Society*, S. Saponara and A. De Gloria, Eds., Cham, Switzerland: Springer, 2022, pp. 93–99.

[53] S. M. Mousavi, W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza, "Earthquake transformer—An attentive deep-learning model for simultaneous earthquake detection and phase picking," *Nature Commun.*, vol. 11, no. 1, p. 3952, Aug. 2020.

[54] Z. Peng, J. Yang, T.-H. Chen, and L. Ma, "A first look at the integration of machine learning models in complex autonomous driving systems: A case study on Apollo," in *Proc. 28th ACM Joint Meeting Eur. Softw. Eng. Conf. Symp. Found. Softw. Eng.*, Nov. 2020, pp. 1240–1250.

[55] I. Qureshi et al., "Medical image segmentation using deep semantic-based methods: A review of techniques, applications and emerging trends," *Inf. Fusion*, vol. 90, pp. 316–352, Feb. 2023.

[56] N. Ghielmetti et al., "Real-time semantic segmentation on FPGAs for autonomous vehicles with hls4ml," *Mach. Learn., Sci. Technol.*, vol. 3, no. 4, Dec. 2022, Art. no. 045011.

[57] L. G. Garcia et al., "Muon–electron pulse shape discrimination for water Cherenkov detectors based on FPGA/SoC," *Electronics*, vol. 10, no. 3, p. 224, Jan. 2021.

[58] R. S. Molina, I. R. Morales, M. L. Crespo, V. G. Costa, S. Carrato, and G. Ramponi, "An end-to-end workflow to efficiently compress and deploy DNN classifiers on SoC/FPGA," *IEEE Embedded Syst. Lett.*, vol. 16, no. 3, pp. 255–258, Sep. 2024.

[59] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A comprehensive survey on model compression and acceleration," *Artif. Intell. Rev.*, vol. 53, pp. 5113–5155, Feb. 2020.

[60] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[61] B. Hawks, J. Duarte, N. J. Fraser, A. Pappalardo, N. Tran, and Y. Umuroglu, "Ps and Qs: Quantization-aware pruning for efficient low latency neural network inference," *Frontiers Artif. Intell.*, vol. 4, Jul. 2021, Art. no. 676564.

[62] J. Duarte et al., "Fast inference of deep neural networks in FPGAs for particle physics," *J. Instrum.*, vol. 13, no. 7, Jul. 2018, Art. no. P07027.

[63] C. Zhang, G. Sun, Z. Fang, P. Zhou, P. Pan, and J. Cong, "Caffeine: Toward uniformed representation and acceleration for deep convolutional neural networks," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 38, no. 11, pp. 2072–2085, Nov. 2019.

[64] J. Haris, P. Gibson, J. Cano, N. B. Agostini, and D. Kaeli, "SECDA: Efficient hardware/software co-design of FPGA-based DNN accelerators for edge inference," in *Proc. IEEE 33rd Int. Symp. Comput. Archit. High Perform. Comput. (SBAC-PAD)*, Belo Horizonte, Brazil, Oct. 2021, pp. 33–43.

[65] S. Marrone et al., "Pulse shape analysis of liquid scintillators for neutron studies," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 490, nos. 1–2, pp. 299–307, 2002.

[66] R. C. Pereira et al., "Neutron/gamma discrimination code based on trapezoidal filter," *Fusion Eng. Des.*, vol. 134, pp. 118–122, Sep. 2018.

[67] H. Y. Özderya. (2023). *SerialPlot*. Accessed: Jun. 2, 2024. [Online]. Available: https://github.com/hyOzd/serialplot

[68] N. Cruz et al., "The design and performance of the real-time software architecture for the ITER radial neutron camera," *IEEE Trans. Nucl. Sci.*, vol. 66, no. 7, pp. 1310–1317, Jul. 2019.

[69] F. Ferrulli, M. Labalme, and M. Silari, "Investigation of CLYC-6 for thermal neutron detection and CLYC-7 for fast neutron spectrometry," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 1029, Apr. 2022, Art. no. 166460.

[70] I. Morales, A. Monterroso, and S. Urizar, "Design, assembly and calibration of a microcontroller-based Geiger–Müller doserate meter," in *Proc. IEEE Central Amer. Panama Student Conf. (CONESCAPAN)*, Sep. 2016, pp. 1–5.

[71] T. K. Nguyen et al., "Calibration of a neutron dose rate meter in various neutron standard fields," *Nucl. Sci. Techn.*, vol. 31, no. 6, p. 60, Jun. 2020.

Open Access funding provided by 'Università degli Studi di Trieste' within the CRUI CARE Agreement