CNN-BASED ARRIVAL TIME DETERMINATION IN PILED-UP PARTICLE COUNTER SIGNALS

R. Singh*, P. Boutachkov, GSI, Darmstadt, Germany
T. Habermann†, M. Kumm, Fulda University of Applied Sciences, Fulda, Germany

Abstract

Pulse pile-up limits the count rate, timing precision, and energy resolution of radiation detection systems in accelerator beam instrumentation and downstream experiments. We present the project status of a machine-learning based pile-up recovery system designed for real-time particle counting and time of arrival determination at pulse rates exceeding 10^7 particles/s. The convolutional neural network (CNN) architecture utilized in this project is trained on labelled scintillator data to identify pulses in the piled-up waveforms. This development is primarily aimed at beam spill characterization at GSI/FAIR using plastic scintillators, however the concept presented is general-purpose and could be applicable to any radiation detector. First model implementation is performed on the FPGA onboard a commercial digitizer and its performance is compared against the single threshold leading-edge discriminator.

INTRODUCTION

Radiation or particle detectors communicate the information of the species, charge, energy, arrival time of the detected particles via the electronic pulses produced at its output. Pulse pile-up is defined as the temporal overlap of these electronic pulses/signals when successive radiation events arrive within the resolving time of the detection system. A pulse pile-up arises whenever the mean interval between successive events approaches the characteristic response time of the detector-electronics chain. For radioactive sources, event inter-arrival times are exponentially distributed and intensity fluctuations scale as $\sqrt{N}$. The probability of k events within a resolving time τ at mean rate r is $P(k) = (r\tau)^k e^{-r\tau}/k!$. Poisson process discussed above is statistically the most favorable case for pile-ups, while in slowly extracted beam from synchrotrons, the situation can be significantly worse. The synchrotron beams typically exhibit intensity fluctuations far exceeding the Poisson limit, driven by power supply ripples and the nonlinear dynamics of the resonant extraction process; measured duty factors at GSI SIS-18 are typically 30–70% of the Poisson limit, meaning instantaneous rates regularly exceed several times the mean rate on microsecond timescales [1, 2]. The consequences of spill fluctuation-induced pile-up propagate directly into the physics experiments downstream. The HADES dilepton spectrometer [3], operating at GSI reported occupancy-related pile-up at its interaction rate of ∼20 kHz due to substantial aforementioned beam fluctuations on $\approx 10 - 100$ µs timescales that degrade tracking efficiency. The ability to precisely count and timestamp individual beam particles from piled-up signals with sub-nanosecond timing using any of the available counting detectors would simultaneously enable improved spill quality characterization, facilitating faster feedback for extraction optimization [4], and extended operating range in particle rates for the physics detectors downstream.

* r.singh@gsi.de
† The work was partially funded by the German Federal Ministry of Research, Technology and Space, Germany under Grant 05K25REA

Pile up Handling Approaches

Initial pile-up handling was focussed on identifying and rejecting piled-up pulses [5, 6], and this evolved into pile-up avoidance using either analog pulse shaping methods [7,8] or more trivially, by detector segmentation dividing the pulses into multiple channels. With the advent of digital processing, conceptual change from pile-up avoidance to pile-up recovery emerged [9]. Most of these "new" recovery methods posed the problem of pile-up as a linear superposition of independent pulses (modelled as analytical functions) whose amplitude and time of arrival had to be determined via various methods. Notable among these de-convolution type methods are template matching [10, 11], maximum likelihood estimation (MLE) [12, 13], and high-yield-pile-up-event-recovery (HYPER) [14]. HYPER is based on W-SUM method which deconvolves a single exponential decay into step function. A systematic comparison between these methods is shown in [13, 15]. Arguably the most general classical approach is MLE, however it requires iterative nonlinear optimization for each waveform segment resulting in variable and unbounded computation time, making it a challenge for real-time FPGA implementation at the throughputs required at our targeted rates.

More recently, deep learning approaches have been widely applied to the pile-up correction problem across multiple domains. Primary motivation is to overcome the assumptions in classical methods, e.g. pulses superimpose linearly and follow a known analytical shape. In addition, real detector pulses vary due to detector imperfections, gain drifts, saturation effects which are typically not captured by a pulse template. Deep neural networks trained on real labeled data can potentially learn all these effects often with a risk of model overfitting and improper generalization. A comparison of template matching and CNN in context of particle counting application can be found [16]. Although deep learning approaches appear very different from classical methods, it is helpful to understand them as a form of amortized inference [17], a trained neural network learns the approximate mapping from observed waveform to underlying parameters (peak times, amplitudes, count) that MLE would compute iteratively, but produces it in a single, fixed-time pass. The cost of the iterative optimization is during training, rather

than repeatedly at run time giving deterministic latency and resource usage independent of pulse rate. In this project, we aim to provide a framework with labelled piled-up data from various radiation detectors and deep learning models trained on this data. A summary of current and future data collection methods, first model implemented in the FPGA along with the inference from the trained model transferred to a new dataset are discussed in the next two sections followed by an outlook.

DATA COLLECTION AND LABELLING METHODS

Single-particle pulse data were collected at the HTP beamline at GSI. The beam consisted of $^{12}C^{6+}$ ions extracted from the SIS-18 synchrotron at a kinetic energy of 300 MeV/u.A plastic scintillator (BC400) coupled to a photomultiplier tube (PMT) served as the primary single-particle counter. All ions within the same spill share the same energy and charge state, producing PMT pulses with consistent characteristics, i.e. the area under each pulse is comparable and thus the total power in any given time window serve as an estimator of the number of constituent pulses [18]. Supervised training of the CNN requires precise ground truth labels for all the quantities to the determined, in our case that is the total particle count along with arrival times of individual particles within each waveform segment. The power indicator discussed above provided the first coarse classification of pulse counts and was used to verify the multiplicity in the labeling process. The time of arrival requires a more careful labelling. First labelling approach was with a combination of algorithmic peak detection followed by manual correction. This was a time consuming process but provided the first set of reliable data for training the network. Following that, synthetic piled-up pulses under linear superimposition assumption were generated using identified single pulses in previous step. These aspects were discussed in detail in the preceding contribution [18] and were used for training the CNN model reported in this contribution. The aforementioned labelling approaches are being extended by two more procedures discussed below.

Coincidence Between Slow and Fast Detectors Dedicated data was collected in recent experiments where coincidence measurements between slower detectors like plastic scintillator (SCI) with total pulse lengths of 25-40 ns against faster detectors like OTR counter [19], Diamond [20] and SiC (5-15 ns) were performed. This allows robust determination of time of arrival of extremely piled-up pulses under real operational conditions in the slower detectors using the pulses detected on fast detectors. An example is indicated in Fig. 1 where $^{238}U^{92+}$ beam is extracted at 700 MeV/u using knock out method (KO) with an excitation frequency of 2.36 MHz. Up to 5 ms waveforms are recorded with a sampling rate of 3.125 GSa/s at an average extraction rate of 20 MHz. It is well known that the KO excitation signal is imprinted on the spill [21] as evident from 2.36 MHz structure in Fig. 1. This data has not been used to update the model trained on data produced by previous labelling methods.

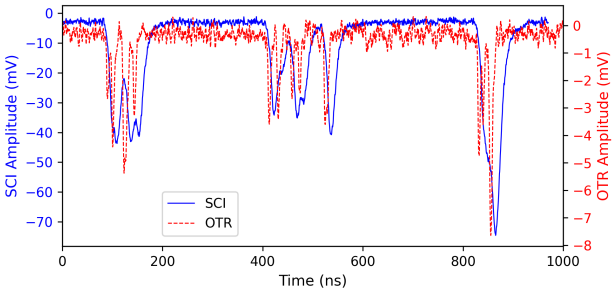

Figure 1: Scintillator and OTR pulses in coincidence.

Laser-based Labelling System Extensive ground truth data collection for PMTs is planned using two independently and precisely triggered UV diode lasers, each producing ≈ 50 ps laser pulses, at variable intervals. This scheme of irradiating the PMT at programmable rates and separations gives exact ground truth for both counting and timing, without the limitations of manual labeling, linear synthetic superposition or expensive beam time campaigns. This set-up will allow simulating piled-up signals from up to 4-5 particles within 10 ns. All the labelling data will be made available in an open repository [22] for scientific applications.

DAQ AND MACHINE LEARNING MODEL

The data acquisition platform chosen for this project is the Teledyne SP Devices ADQ35, a high-speed digitizer with on-board Xilinx Kintex UltraScale FPGA. The digitizer supports two operating modes: a 2-channel mode at 5 GSa/s per channel, and a single-channel mode at 10 GSa/s via interleaved ADC operation. The channel gain and offset can be individually and digitally programmed such that the input signal is scaled before the FPGA processing. The FPGA on the ADQ35 provides ample resources for custom firmware development: ~5,000 DSP slices, ~600k LUTs, and ~2,000 BRAM blocks. The ADQ35 optionally provides a firmware for Waveform and Pulse Detection (FWPD), which implements a leading-edge discriminator (LED) with programmable threshold. FWPD reports timestamp and pulse-height pairs, Time over Threshold (ToT) and area under the pulse for each detected pulse. Although LED is a weak baseline in comparison against our ML-based recovery system for piled-up signals, it is still the dominant method in actual operation for beam diagnostics and FWPD provides a modern commercial implementation of LED.

Baseline Architecture: PILONET

The baseline CNN architecture PIle-up LOcator Net (PILONet) developed in our previous work [18], inherits main ideas from YOLOv1 architecture [23] where the input samples are divided into equal-sized cells, and each cell predicts a confidence score together with a position offset of all peaks relative to start of the cell. Thus, each cell independently predicts up to $m = 4$ pulse peaks and outputs a relative position

or time $\hat{p}_x \in [0, 1)$ from the start of waveform window and a confidence score $\hat{p}_c \in [0, 1]$ for the predicted time. For the FPGA implementation, small modifications were made to the architecture presented in [18] to reduce the number of model parameters from 9,000 to around 2,000 parameters in total. The network consists of six convolutional layers where the first two layers downsample by a total stride of 32, which also defines the cell size, while the subsequent layers process each cell with receptive fields limited to few neighboring cells. The model architecture is shown in Table 1. At

Table 1: PILONET architecture for n cells as input, each cell containing 32 samples and detecting up to m pulses.

Layer	Kernel	Stride	Filters	Output shape
input	–	–	–	$(32n,\ 1)$
conv1	4	4	8	$(8n,\ 8)$
conv2	8	8	8	$(n,\ 8)$
conv3	3	1	16	$(n,\ 16)$
conv4	3	1	16	$(n,\ 16)$
conv5	1	1	16	$(n,\ 16)$
conv6	3	1	$2m$	$(n,\ 2m)$
reshape	–	–	–	$(n,\ m,\ 2)$

3.125 GSa/s, each cell spans 10.24 ns. The SCI followed by PMT produces pulses with ~8 ns rise time and ~30 ns decay (see Fig. 1), i.e. each pulse spans ~4 cells. For an individual pulse, the cell containing the peak must predict it, while the remaining cells must correctly identify that no pulse is present. This creates at least ~3:1 class imbalance between empty and occupied cells at single pulse level. In real spill data this baseline is misleading: the super-Poissonian microstructure of slow extraction shown in Fig. 1 produces dense bursts where nearly every cell contains a peak, separated by empty stretches. The local density of peak-bearing cells therefore can vary by orders of magnitude across the spill, motivating the evaluation of appropriate loss functions for training the PILONET [18, 24]. The receptive field of each cell at the output layer is 244 original samples (78 ns), covering roughly 2 complete pulse widths. The tool chain used to implement the trained CNN model on the FPGA is discussed in [25]. The utilized tool chain is data-rate aware and utilizes the FPGA resources optimally. Alternative FPGA porting methods of trained models are also available [26]. The CNN described above uses less than 15% of the available DSP and 6% of LUT resources, leaving substantial headroom for exploring further architectures. The inference latency is 291.2 ns.

Comparison with Leading Edge Discriminator

A subsection of recorded SCI waveforms during a non-optimized KO spill extracted with a 2.36 MHz extraction signal is shown in Fig. 1. This specific waveform has 512 cells corresponding to 16384 input samples and is played using a 500 MHz signal generator at 10 GSa/s and connected to the ADQ35 digitizer whose FPGA is running the CNN implementation (this dataset was not part of the training set)

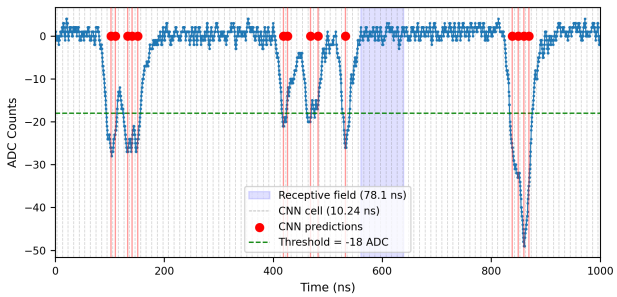

Figure 2: Comparison of detection performance at 20 MHz average particle rate. The cells are marked with dashed lines and receptive field with the shaded region. The optimal threshold setting for counting most pulses using FWPD is also shown.

discussed in the previous section. The PILONET predictions are marked with vertical lines in Fig. 2 where only ≈ 100 cells are shown for clarity. The first tunable user parameter in case of PILONET is the amplitude normalization to match with the training data which was adjusted using the digital input gain. Second parameter is the confidence threshold associated with each peak, which allows controlling the false positives. Amplitude normalization can also be found automatically by a low rate calibration run without piled-up pulses. All the peaks were correctly identified (in comparison to fast detector in Fig. 1) at the optimal normalization setting. Unsurprisingly LED performs poorly under pile-up conditions as observed, i.e. only 8 peaks instead of 15 could be identified for the best threshold setting marked in Fig. 2.

OUTLOOK

The first results from the current architecture are promising, but significant further work remains. The CNN parameters still need to be optimized, robustness to variations in the input signal needs to be characterized, and timing resolution needs to be quantified against an independent reference. A comparison with the TDC measurements at the 80 MHz cavity operation [27,28] is planned for this purpose. A two-head version of the network separating counting and timing into independently optimized outputs will be evaluated. Pulse energy determination will be added on top of the timing predictions using the W-SUM method [9]. Two extensions broaden the scope of the architecture. First, scaling to detectors with different pulse shapes by adjusting the digitizer sampling rate so that each pulse spans the same number of cells and the trained network transfers without retraining. Second, fundamentally different network designs will be explored: networks that compress each waveform window into a small set of summary numbers and reconstruct the underlying pulses from them, and networks designed to track the slowly varying state of the PMT itself — gain sag, recovery from saturation, baseline drift over long stretches of data, providing a continuous validity signal alongside the per-pulse predictions. The CNN count predictions will be integrated into the spill feedback system at GSI and tested with beam at the highest particle rates available.

REFERENCES

[1] R. Singh, P. Forck, and S. Sorge, "Reducing fluctuations in slow-extraction beam spill using transit-time-dependent tune modulation", *Phys. Rev. Applied*, vol. 13, p. 044076, 2020. doi:10.1103/PhysRevApplied.13.044076

[2] R. Singh, P. Boutachkov, P. Forck, S. Sorge, and H. Welker, "Slow Extraction Spill Characterization from Micro to Milli-Second Scale", in *Proc. IPAC'18*, Vancouver, Canada, Apr.-May 2018, pp. 2095–2098. doi:10.18429/JACoW-IPAC2018-WEPAK007

[3] G. Agakichiev *et al.*, "The high-acceptance dielectron spectrometer HADES", *Eur. Phys. J. A*, vol. 41, pp. 243–277, 2009. doi:10.1140/epja/i2009-10807-5

[4] P. Niedermayer, H. Bräuning, R. Singh, and T. Milosic, "Spill optimization system improving slow extraction at GSI", in *Proc. IPAC'25*, Taipei, Taiwan, Jun. 2025, pp. 956–959. doi:10.18429/JACoW-IPAC2025-TUPB008

[5] W. Feller, "On probability problems in the theory of counters", in *Studies and Essays, Courant Anniversary Volume*, pp. 105–115, 1948.

[6] G. Bertolini, V. Mandl, and G. Melandrone, "Pile-up detection circuit", *Nucl. Instrum. Meth.*, vol. 30, pp. 301–304, 1964. doi:10.1016/0029-554X(64)90053-5

[7] V. T. Jordanov and G. F. Knoll, "Digital synthesis of pulse shapes in real time for high resolution radiation spectroscopy", *Nucl. Instrum. Meth. A*, vol. 345, pp. 337–345, 1994. doi:10.1016/0168-9002(94)91011-1

[8] V. Drndarević, "Elimination of distortions of nuclear energy spectra caused by pulse pile-up", *Meas. Sci. Technol.*, vol. 5, pp. 1573–1575, 1994. doi:10.1088/0957-0233/5/12/024

[9] W. H. Wong and H. Li, "A scintillation detector signal processing technique with active pileup prevention for extending scintillation count rates", *IEEE Trans. Nucl. Sci.*, vol. 45, no. 3, pp. 838–842, 1998. doi:10.1109/23.682647

[10] B. Loeher *et al.*, "High count rate γ-ray spectroscopy with $LaBr_3$:Ce scintillation detectors", *Nucl. Instrum. Meth. A*, vol. 686, pp. 1–6, 2012. doi:10.1016/j.nima.2012.05.051

[11] J. Z. Zhang *et al.*, "Development of a pile-up pulse recovery algorithm for the $LaBr_3$ detector", *Nucl. Instrum. Meth. A*, vol. 1062, p. 169273, 2024. doi:10.1016/j.nima.2024.169273

[12] P. A. B. Scoullar, C. C. McLean, and R. J. Evans, "Real time pulse pile-up recovery in a high throughput digital pulse processor", *AIP Conf. Proc.*, vol. 1412, no. 1, pp. 270–277, Dec. 2011. doi:10.1063/1.3665324

[13] M. Bolić and V. Drndarević, "Pileup correction algorithms for very-high-count-rate gamma-ray spectrometry with NaI(Tl) detectors", *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2156–2161, 2009. doi:10.1109/TNS.2009.2021917

[14] W. H. Wong *et al.*, "Feasibility of a high-speed gamma-camera design using the high-yield-pileup-event-recovery method", *J. Nucl. Med.*, vol. 42, pp. 624–632, 2001.

[15] M.-R. Mohammadian-Behbahani *et al.*, "A comparison study of the pile-up correction algorithms", *Nucl. Instrum. Meth. A*, vol. 1047, p. 167724, 2023. doi:10.1016/j.nima.2022.167724

[16] S. E. Engel, P. Boutachkov, and R. Singh, "Application of Machine Learning towards Particle Counting and Identification", no. 11, pp. 508–511, Dec. 2022. doi:10.18429/JACoW-IBIC2022-WEP42

[17] K. Cranmer, J. Brehmer, and G. Louppe, "The frontier of simulation-based inference", *Proc. Natl. Acad. Sci.*, vol. 117, pp. 30055–30062, 2020. doi:10.1073/pnas.1912789117

[18] T. Habermann, M. Kumm, and R. Singh, "Application of convolutional neural networks for pile up correction in single particle counting", in *Proc. IBIC'25*, Liverpool, UK, Sep. 2025, pp. 152–155. doi:10.18429/JACoW-IBIC2025-MOPCO36

[19] R. Ghagi, R. Singh, P. Boutachkov, B. Walasek-Hoehne, C. Welsch, and H. Zhang, "UV–VIS emissions from high energy heavy ions: mechanisms and applications to single-ion detection", in *Proc. IBIC'25*, Liverpool, UK, Sep. 2025, pp. 249–253. doi:10.18429/JACoW-IBIC2025-MOPMO28

[20] M. Alfonsi *et al.*, "Validation of the diamond detectors for the Super Fragment Separator beam diagnostics", *J. Instrum.*, vol. 19, no. 05, p. C05009, 2024. doi:10.1088/1748-0221/19/05/C05009

[21] P. Niedermayer and R. Singh, "Excitation signal optimization for minimizing fluctuations in knock out slow extraction", *Sci. Rep.*, vol. 14, p. 10353, 2024. doi:10.1038/s41598-024-60966-y

[22] R. Singh and T. Habermann, "particle-counter-data", 2026. `https://git.gsi.de/r.singh/particle-counter-data`

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection", in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016. doi:10.1109/CVPR.2016.91

[24] M. Hamdan and T. Habermann, "Exploring convolutional neural network training strategies for pile-up correction in single particle counting", presented at IPAC'26, Deauville, France, May 2026, paper WEP6009, this conference,

[25] T. Habermann, M. Mecik, Z. Wang, C. D. Vera, M. Kumm, and M. Garrido, "Continuous-flow data-rate-aware cnn inference on fpga", *IEEE Trans. Circuits Syst. Artif. Intell.*, pp. 1–14, 2026. doi:10.1109/TCASAI.2026.3669843

[26] T. Aarrestad *et al.*, "Fast convolutional neural networks on fpgas with hls4ml", *Mach. Learn.: Sci. Technol.*, vol. 2, no. 4, p. 045015, Jul. 2021. doi:10.1088/2632-2153/ac0ea1

[27] S. Sorge, P. Forck, and R. Singh, "Spill ripple mitigation by bunched beam extraction with high frequency synchrotron motion", *Phys. Rev. Accel. Beams*, vol. 26, no. 1, p. 014402, Jan. 2023. doi:10.1103/PhysRevAccelBeams.26.014402

[28] T. Milosic, P. Forck, and R. Singh, "Sub-ns Single-Particle Spill Characterization for Slow Extraction", in *Proc. IBIC'21*, Pohang, Rep. of Korea, Oct. 2021, pp. 438–442. doi:10.18429/JACoW-IBIC2021-WEPP28