IOP Conf. Series: Journal of Physics: Conf. Series **898** (2017) 032054 doi:10.1088/1742-6596/898/3/032054

Reconstruction of Micropattern Detector Signals using Convolutional Neural Networks

L. Flekova[1] and M. Schott[2]

[1]Department of Computer Science, Technische Universität Darmstadt, Germany,
[2] Institute for Physics, Johannes Gutenberg University, Mainz, Germany

E-mail: flekova@ukp.informatik.tu-darmstadt.de, mschott@cern.ch

Abstract. Micropattern gaseous detector (MPGD) technologies, such as GEMs or MicroMegas, are particularly suitable for precision tracking and triggering in high rate environments. Given their relatively low production costs, MPGDs are an exemplary candidate for the next generation of particle detectors. Having acknowledged these advantages, both the ATLAS and CMS collaborations at the LHC are exploiting these new technologies for their detector upgrade programs in the coming years. When MPGDs are utilized for triggering purposes, the measured signals need to be precisely reconstructed within less than 200 ns, which can be achieved by the usage of FPGAs.

In this work, we present a novel approach to identify reconstructed signals, their timing and the corresponding spatial position on the detector. In particular, we study the effect of noise and dead readout strips on the reconstruction performance. Our approach leverages the potential of convolutional neural network (CNNs), which have recently manifested an outstanding performance in a range of modeling tasks. The proposed neural network architecture of our CNN is designed simply enough, so that it can be modeled directly by an FPGA and thus provide precise information on reconstructed signals already in trigger level.

1. Introduction

In recent years, MicroMegas (Micro-Mesh gaseous structure) detectors [1, 2] received significant attention in the development of precision and cost-effective tracking detectors in nuclear and high energy physics experiments, e.g., they have been chosen as a baseline technology for the upgrade project of the ATLAS muon system [4]. In this paper, we provide a brief summary on the underlying concept. For a detailed discussion we refer the interested reader to [3].

MicroMegas detectors are gaseous parallel-plate detectors with two regions that are separated from each other by a thin metallic mesh, illustrated in Figure 1. In the drift region, typically of a height of several mm, the traversing charged particles ionise gas atoms, typically from a Ar/CO_2 gas-mixture. The resulting ionization electrons drift along the electric field towards the mesh. The amplification region, directly below the mesh, has a height of $\approx 100\,\mu$m. The electrical field in the amplification region is higher by two orders of magnitude, i.e., large enough to create electron avalanches. Those imply an enhancement of the primary electron signal by a factor of 10^4, enabling their measurement on the readout electrodes. MicroMegas detectors with a spatial resolution of up to $40\,\mu$m and a timing resolution of 5 ns have been successfully constructed and tested.

A typical readout scheme for MicroMegas detectors is based on APV25 (Analog Pipeline

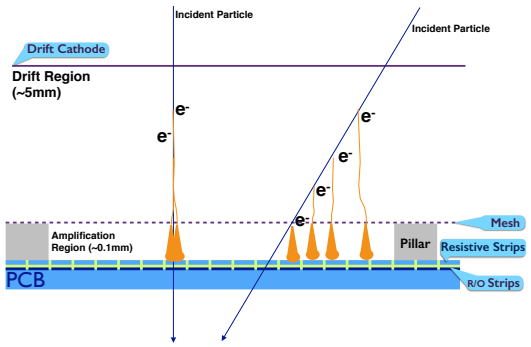

Figure 1. Illustration of the working principle of MicroMegas detectors with two incident particles.

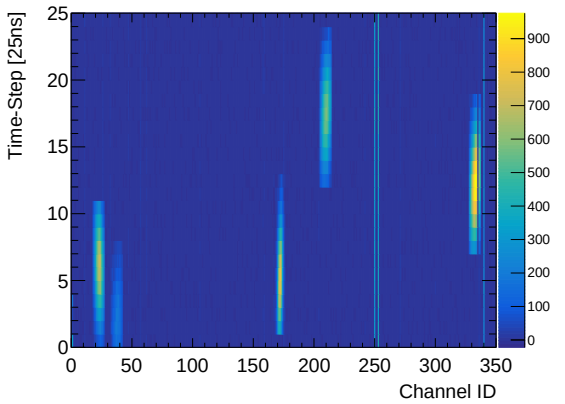

Figure 2. Example of a simulated event with 5 incident particles with a MicroMegas detector with 360 readout channels.

Voltage chips with 0.25 μm CMOS technology) hybrid cards [11] and the RD51 Scalable Readout System (SRS) [10]. The APV25 provides 128 readout channels with analog CR-RC shaped signals sampled at 40 MHz rate. This leads to a measurement of the integrated charge, measured in units of ADC counts, on each readout strip in 27 time bins of 25 ns.

A typical event of a five incident particles, recorded by a MicroMegas detector with APV25 readout scheme for 360 channels, is shown in Figure 2. The signal evaluation for each readout strip is shown for 25 time bins and a clear signal characteristic can be seen. In addition, several readout channels with significant noise, as well as dead channels are shown. The situation get complicated, when background events, e.g. stemming from high-rate environments, distort the signal characteristics.

While previous approaches for the signal identification rely on simple sequential algorithms, we present here a novel approach for the identification of MicroMegas signals in a highly parallel way, allowing for the implementation on an FPGA processor. This theoretically enables the reconstruction of events within less than 200 ns in complex and background rich environments. Our approach is based on convolutional neural networks [5], which became very popular in image classifications tasks [6, 7] and recently also in high energy physics [9], [8].

2. Neural Network Architecture

While the idea of neural-network learning has been known for decades [12, 13, 14], the rapid progress in computing infrastructure enabled deeper architectures, resulting into a deep learning boom for the past several years. Deep neural networks achieved an unprecedented performance in image classification [6] and speech recognition [15, 16, 17], and have been successfully used in high energy physics [8], [9].

Convolutional neural network (CNN) is a type of a feed-forward network network which contains one or more convolutional layers. These networks, developed mainly for image recognition problems [5], are useful for classification tasks in which we expect to find strong local clues regarding class membership, but these clues can appear in different places in the input. For example, the network can learn that a certain set of points and curves is a human face, regardless where on an image it appears [6].

The main idea behind a CNN is to learn a non-linear function (called a convolutional filter) to be applied on a sliding window over the image. This function transforms a window of $m \times n$ pixel into a vector that captures important properties of the shape in the window. A pooling

operation is then applied to combine the vectors resulting from the different windows into a single vector, usually by taking the maximum or the average value observed. The intention is to focus on the most important features, regardless of their location. The resulting vector is then fed further into a network that is used for prediction. Parameters of the filter function are updated during the training process to highlight the aspects of the data that are important for the task.

CNNs have demonstrated excellent performance on challenging visual classification tasks [6, 7]. Most notably, [6] show an unprecedented performance on the ImageNet 2012 classification benchmark, reducing an error rate from 26.1% to 16.4%.

In our experiments, we treat the input data of an event, recorded by a MicroMegas detector, as a 2D image, representing the readout channels and the time scale. Each event can be interpreted as gray-scale image with 360x25 pixels, each representing a value between 0 and 1000. We use a standard fully supervised CNN model, mapping the 2D input "image" through a set of convolutional filters and a sequence of hidden layers to a probability vector over the number of incident particles in the image (0-5), and a probability vector over the discrete approximate position features ("segments"). The combination of the shape learning through convolutional filters, and the position features fed into the lower layer, enables us to reconstruct the segments with a signal into the correct number of incident particles, and extrapolate an approximate center of each of those.

The convolutional filters are applied separately on small input regions (10x8 pixels) of the image in a parallel way as illustrated in Figure 3. Each operation of the subsequent layers is performed first for the small local subsets, before being combined at a later stage of the network. This approach enables a parallel training on the small regions, i.e. reducing significantly the complexity of the training algorithm. For each subset, we apply five filters for extracting the brightest region, the dead and noisy readout-strips, the 1D projection to the x- and y-axis, and the brightest point of the segment. First filter layer is optimized to be parallelized as much as possible.

Additionally, the information of the position of the local maximum (using a weighted mean used for position extraction) is routed directly to the second hidden layer (Fig. 3) layer, while the other filtered results are fed into separate feed forward networks and trained independently. Second filter layer includes also pooling layer, i.e., performing a downsampling operation. The output information gets therefore pooled and restructured, reducing the complexity from 360x25 to 90x4. Adjoint regions are identified and connected. The final output layer makes the decision of the presence of absence of up to five incident particles within the image, and returns the stored local maximum feature value of the segment containing the overall maximum of the present particles (note that one particle typically spreads over multiple segments and multiple particles may overlap).

3. Data Set and Training

We use simulated events of event signatures, in order to quantify the performance of spatial and time resolution, as well as signal efficiency and fake rate. These simulated event signatures contain between 0 and 5 signal events and include typical signal characteristics such as background noise, dead-channels and overlapping signals.

In total, our training set contains 20,000 events with different background and detector features. Our performance tests are conducted with an uncorrelated test-statistics of 1,000 events. In addition, the results, which are based on simulated samples, have been cross-checked with real detector data taken at the MAMI particle accelerator at the University of Mainz.

A backpropagation algorithm is used for network training. While the training itself can be performed on a normal PC based on floating point precision, the final application on an FPGA requires a reduced precision for the individual calculations. Hence, we reduced the precision of

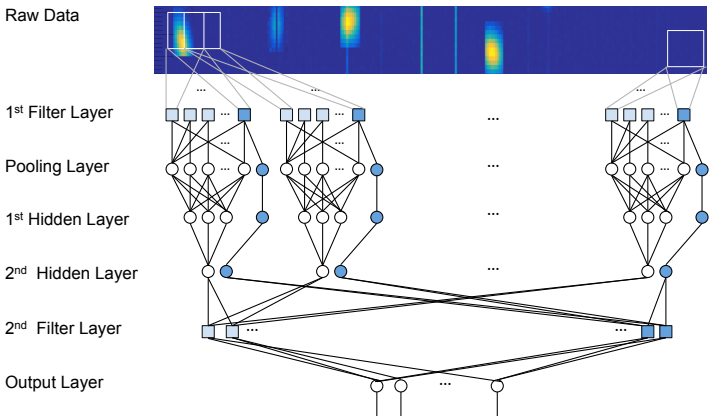

Figure 3. Overview of the network architecture used in the experiments. The filters are applied in parallel to different regions of the raw data. The pooling layer transforms the filtered information, selecting the most prominent information from adjoint regions. Subsequently, a feed-forward network architecture is employed.

input-patterns to 4-bit precision and the network-weights to a one byte precision. The sigmoidal output-function of each neuron in the network is reduced to a 4-bit precision using a look-up table.

The input data with a size 360x25 responds to a typical MicroMegas detector with a one dimensional readout length of 25-50 cm. Given the input bandwidth of 600-1000 bits per cycle for a typical FPGA[1], the input size is still too large. Therefore we also tested the performance of our network, when reducing the input information to 90x4 values, corresponding to 5 FPGA cycles for reading, before the actual filtering.

4. Results

In the following we compare the performance of the classifier, once based on the full input information using the full floating point precision during the classification (labelled as *full CNN*) and once based on the reduced input information with reduced number precision ((labelled as *reduced CNN*)).

The classification results for one simulated event with three indecent particles are illustrated in Figure 4 as example. The efficiency as well as the fake rate in dependency of the number of clusters, i.e. incident particles, is shown Figure 5. A particle is identified correctly if the reconstructed position is within 1mm of the expected position, otherwise it is considered as fake-reconstruction. The reconstruction efficiency is larger than 90% using the full network information and larger than 85% using the reduced network. The fake rate is always below 5%. It should be noted, that the fake rate for events with no incident particles in the event is 0, meaning that the observed fake rate for higher number of clusters is actually an artifact of the limited spatial resolution.

The spatial as well as the timing resolution is shown in Figure 6 for both network implementations. Similar results are observed using the 4 byte and 6 bit number precision. The spatial resolution is in the order of 250 μm, which is corresponding to the assumed width of the readout strips. A typical timing resolution of 36 ns is observed for both approaches.

The reduced CNN has been also tested with real data taken during testbeam measurements at MAMI electron accelerator at JGU Mainz, as well as the GIF++ facility at CERN for testing the stability under high background rates. In both cases we observe reconstruction efficiencies above 90%.

[1] A typical cycle rate in an FPGA can be assumed to be in the order of 500 MHz, i.e. 2 ns per cycle

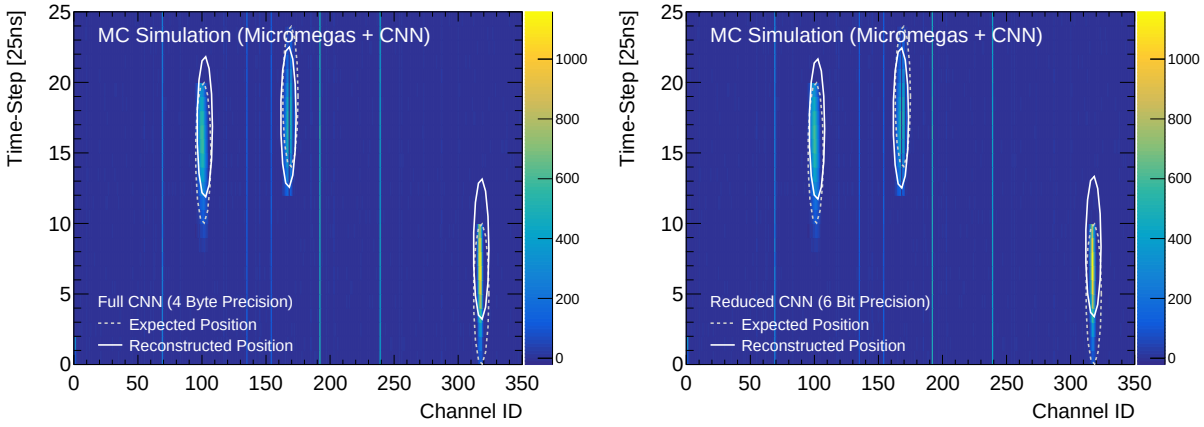

Figure 4. Example of simulated MicroMegas events with three incident particles. The expected position, as well as the reconstructed position using the full classifier (left) and the reduced classifier (right) are illustrated.

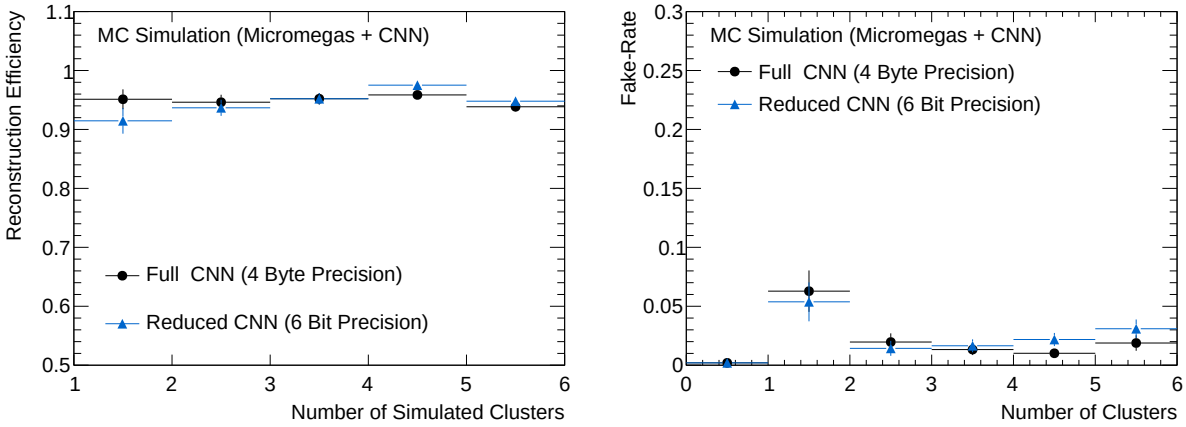

Figure 5. Expected efficiency (left) and the fake-rate (right) for the full and the reduced CNN.

5. Summary and Outlook

In this work, we presented studies on the fast signal reconstruction of MicroMegas detectors using convolutional neural networks. The network design has been optimized for the implementation on FPGAs, leading to a theoretical interpretation of the full recorded signal in less than 200 ns based on initial performance estimations. The expected spatial and timing resolution, which can be achieved using a typical MicroMegas detector design, is 250 μm and 35 ns, respectively. Even for high background rates, we expect an identification efficiency of more than 90% with a corresponding fake rate on the percent level. The chosen network architecture can be used as basis for several further applications, e.g. GEM detectors. The implementation on a Xilinx Virtex 6 FPGA (XC6VHX380T) is currently ongoing using OpenCL as FPGA design tool. The board used for this study has 382k logical units, 60k slices and 4 MB distributed RAM. Preliminary studies indicate that the process of locating the brightest regions in the first network layer takes the most time and should be optimized in the future. Once an initial and functional implementation of the proposed CNN is ready, we aim to optimize the design for less expensive FPGA devices and also compare to ARM CPU versions.

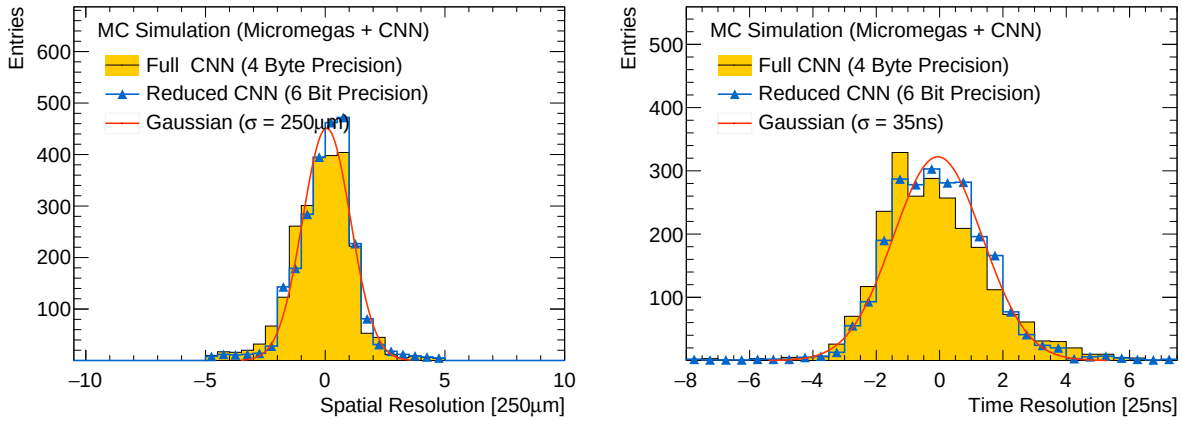

Figure 6. Expected spatial resolution (left) and the expected timing resolution (right) for the full and the reduced CNN.

References

[1] Y. Giomataris, P. Rebourgeard, J. Robert, and G. Charpak, MICROMEGAS: A High granularity position sensitive gaseous detector for high particle flux environments, Nucl. Instrum. Meth. A376 (1996) 2935.
[2] T. Alexopoulos, J. Burnens, R. de Oliveira, G. Glonti, O. Pizzirusso, et al., A spark-resistant bulk-micromegas chamber for high-rate applications, Nucl. Instrum. Meth. A640 (2011) 110118.
[3] Andriamonje, S. and others, Development and performance of Microbulk Micromegas detectors, Proceedings, 1st International Conference on MicroPattern Gaseous Detectors: Kolympari, Chania, Crete, Greece, 12-15 Jun 2009, JINST-5 (2010) P02001
[4] T. Kawamoto et al., New Small Wheel Technical Design Report, Tech. Rep. CERN-LHCC-2013-006. ATLAS-TDR-020, CERN, Geneva, Jun, 2013. https://cds.cern.ch/record/1552862.
[5] Y. LeCun and Y. Bengio (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.
[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097-1105. 2012.
[7] D. Ciresan, U. Meier and J. Schmidhuber, 2012. Multi-column deep neural networks for image classification. In IEEE Conference on Computer Vision and Pattern Recognition (2012), 3642-3649.
[8] P. Baldi, P. Sadowski, and D. Whiteson, Searching for Exotic Particles in High-Energy Physics with Deep Learning, Nature Commun., vol. 5, p. 4308, 2014.
[9] J. Searcy, L. Huang, M.-A. Pleier, and J. Zhu, Determination of the WW polarization fractions in pp W W j j using a deep machine learning technique, Phys. Rev., vol. D93, no. 9, p. 094033, 2016.
[10] S. Martoiu, H. Muller, and J. Toledo, Front-end electronics for the Scalable Readout System of RD51, IEEE Nucl. Sci. Symp. Conf. Rec. 2011 (2011) 20362038.
[11] L.Jones et al., The APV25 deep sub Micron readout chip for CMS channels detectors, Proceedings of 5th workshop on Chips electronics for LHC experiments CERN/LHCC/99-09 (1999) 162166.
[12] Rosenblatt, Frank. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological review 65, no. 6 (1958): 386.
[13] Rumelhart, D. E., G. E. Hinton, and R. J. Williams. "Learning internal representation by back propagation. Parallel distributed processing: exploration in the microstructure of cognition 1 (1986).
[14] G. E. Hinton. Learning distributed representations of concepts. In Proceedings of the eighth annual conference of the cognitive science society, vol. 1, p. 12. 1986.
[15] G. E. Dahl, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing 20.1 (2012): 30-42.
[16] G. E. Hinton, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups." IEEE Signal Processing Magazine 29.6 (2012): 82-97.
[17] F. Seide, G. Li, and D. Yu. Conversational Speech Transcription Using Context-Dependent Deep Neural Networks. In Interspeech (2011) 437-440.