

POSSIBLE APPLICATION AREAS OF MACHINE LEARNING TECHNIQUES AT MPD/NICA EXPERIMENT AND EVALUATION OF THEIR IMPLEMENTATION PROSPECTS IN DISTRIBUTED COMPUTING ENVIRONMENT

D.A. Zinchenko ^{1, a}, E.G. Nikonov ², A.I. Zinchenko ¹

¹ *VBLHEP, JINR, Dubna, Russia*

² *LIT, JINR, Dubna, Russia*

E-mail: ^azinchenk1994@gmail.com

At present, the accelerator complex NICA is being built at JINR (Dubna). It is intended for performing experiments to study interactions of relativistic nuclei and polarized particles (protons and deuterons). One of the experimental facilities MPD (MultiPurpose Detector) was designed to investigate nucleus-nucleus, proton-nucleus and proton-proton interactions.

During the preparation of the physics research program, the production of a large volume of simulated data is required, including high-multiplicity events of heavy-ion interactions with high energy. Realistic modelling of the detector response for such events can be significantly accelerated with a use of generative models.

A selection of rare physics processes traditionally uses machine learning based approaches.

For the high luminosity accelerator operation for the proton-proton interaction research program it will be necessary to develop high-level trigger algorithms and methods, based on machine learning techniques.

During the data taking, the tasks of the fast and efficient processing and storage of large amounts of experimental data will become more and more important, requiring involvement of distributed computing resources.

In this work these problems are considered in connection to the MPD/NICA experimental program preparation.

Keywords: machine learning, generative models, multivariate analysis, heavy-ion collisions

© 2018 Dmitry A. Zinchenko, Eduard G. Nikonov, Alexander I. Zinchenko

1. Detector geometry

The MPD (Figure 1) time projection chamber (TPC) is the main tracking detector of the central barrel and, together with the inner tracking system (IT), time of flight system (TOF) and electromagnetic calorimeter (ECal) has to provide charged particles momentum measurement with sufficient precision, particle identification and vertex reconstruction, two track separation and dE/dx measurement for hadronic and leptonic observables at pseudorapidities $|\eta| < 1.2$ and $p_T > 100$ MeV/c.

TPC is a well-known detector for 3-dimensional tracking and particle identification for high multiplicity events. In the conditions of the maximum charged particle multiplicity ~ 1000 in central Au+Au collisions and the event rate of about 7 kHz achieved at the NICA design luminosity, the TPC/MPD will provide:

- efficient tracking up to pseudorapidity region $|\eta| = 1.2$
- momentum resolution for charged particles under 3% in the transverse momentum range $0.1 < p_T < 1$ GeV/c
- two-track resolution of about 1cm
- hadron and lepton identification by dE/dx measurements with a resolution better than 8%

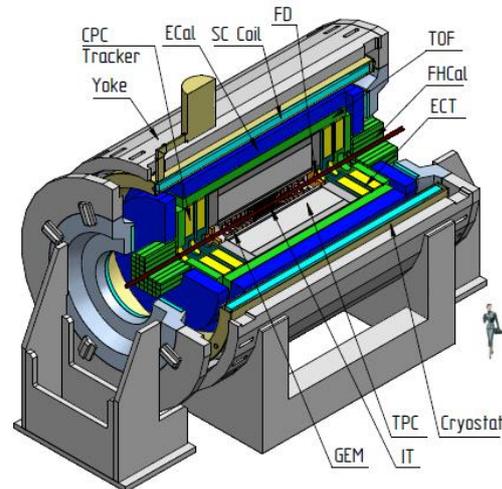


Figure 1. MPD detector

Ambitious physics goals of the MPD require excellent particle identification capability over as large as possible phase space volume. Identification of charged hadrons at intermediate momenta (0.1-2 GeV/c) is achieved by the time-of-flight (TOF) measurements which are complemented by the energy loss (dE/dx) information from the TPC.

The TOF system based on the Multigap Resistive Plate Counters (MRPC) will provide:

- large phase space coverage $|\eta| < 2$;
- high granularity to keep the overall system occupancy below 10-15% and minimize efficiency degradation due to double hits;
- good position resolution to provide efficient matching of TOF hits with TPC tracks;
- high combined geometrical and detection efficiency (better than 80%);
- identification of pions and kaons with $0.1 < p_T < 2$ GeV/c and (anti)protons with $0.3 < p_T < 3$ GeV/c;

The primary role of the electromagnetic calorimeter is to measure the spatial position and energy of electrons and photons produced in heavy ion collisions. It will also contribute to the particle identification due to its high time resolution.

The expected high multiplicity environment implies a high segmentation of the calorimeter. To have an adequate space resolution and good separation of overlapping showers, the transverse cell size should be small enough. Following these requirements, the "shashlyk"-type ECal is proposed to be developed using a tower as a basic building element. It has the transverse size of 4 cm^2 and the length of 40 cm and consists of 220 alternating tiles of Pb (0.3 mm) and plastic scintillator (1.5 mm). The whole ECal will contain ~ 43000 towers.

The silicon IT is planned to be installed at a later stage. It will be constructed from silicon pixel sensors based on MAPS technology. It will help to solve the following tasks. First, it will enhance track reconstruction for particles registered with all other subsystems. Namely, it will improve the tracking quality for low- p_T and/or large- η particles. Second, due to its excellent spatial resolution it will enhance MPD capabilities for rare probe studies, for example, multistrange hyperons. Moreover, it can bring the open charm physics sector within reach. In addition, due to its high processing speed it can be used for triggering on rear probes during the high-luminosity running for pp-collisions.

2. Simulation of the TPC and ECal response using generative models

In order to obtain more realistic estimates of the MPD performance, especially at forward pseudorapidities, a full realistic simulation of the detector response is needed. This is especially true for the TPC, where the simulation details affect cluster, hit and track reconstruction procedures. Due to its large size and high readout granularity, the respective simulation procedure is very time consuming. To better illustrate this, some relevant numbers can be presented. The TPC information is obtained in the form of the space distribution of the charge of ionization electrons produced by charged particles on their path through the active detector volume. Each of ~ 1000 particles will pass the gas thickness of ~ 79 cm and produce ~ 30 electrons per cm of track length (Figure 2 left). The charge distribution will be recorded using 12 readout chambers on each end of the detector. Each chamber contains 53 readout pad rows with the number of pads varying from 42 to 124 in the innermost and outermost pad rows, respectively (Figure 2 right). The pad charge distribution gives the transverse coordinate. In addition, for each pad the signal shape in time is digitized every 100 ns for a total of ~ 300 samples to obtain the longitudinal coordinate from the drift time. So, this digitization procedure is the most CPU-intensive part of the data processing chain (simulation and reconstruction). Current developments with Generative Adversarial Networks (GAN) used for the detector response description can potentially improve the situation [1] by producing single track TPC responses which can be combined in a superposition for a multi-track environment (Figure 3).

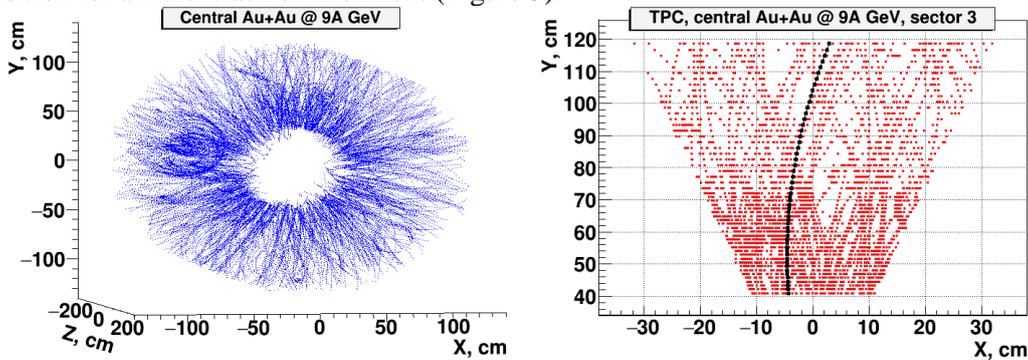


Figure 2: Left - 3-D visualization of trajectories of charged particles in the TPC, produced in Au+Au collision at $\sqrt{s_{NN}} = 9$ GeV. Right - transverse projection of particle trajectories passing through one readout chamber. Dots represent particle crossing points with median planes of pad rows. Black line shows a track of a 0.5 GeV/c pion.

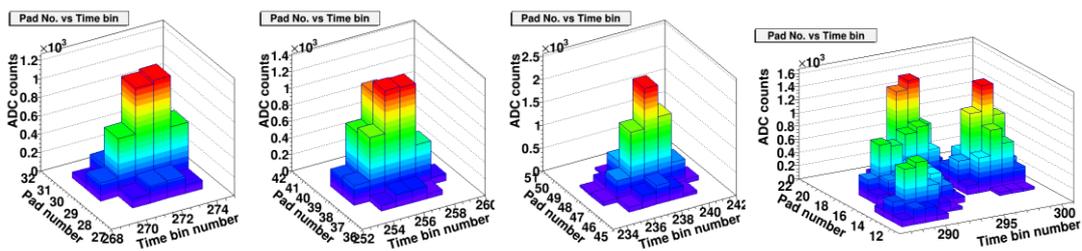


Figure 3: Left - charge distributions from 3 pad rows obtained for a 0.5 GeV/c pion from Figure 2 in the pad number versus time bin space. Vertical axis represent the signal amplitude. Right - charge distribution from several close tracks.

Another detector with a similar problem is the ECal where the simulation of electromagnetic shower development also takes quite some time because of the large number of alternate material layers along the radius and the number of readout elements. Here a similar approach with generative models is also applicable [2–5]. However such a fast simulation method would result in a simplified particle history evolution treatment and a less realistic response description for the TOF, located in front of the ECal at lower radius, because the return flux of backscattered particles originated in the shower will not be reproduced in the simulation.

3. Multivariate analysis for dilepton and open charm selection

The correlated e^+ or μ^+ μ^- pairs (dileptons), especially those from decays of vector mesons (ρ , ω , ϕ), are the best candidates to relate medium modifications of hadronic spectral function to the restoration of the chiral symmetry in A+A collisions because the vector meson decay products (i.e. electrons and positrons) interact only electromagnetically. Therefore they escape the interaction region unaffected by subsequent strong interactions in dense hadronic matter and carry to the detectors information about the conditions and properties of the medium at the time of their creation.

The experimental study of dileptons in heavy-ion collisions is a challenging task. The main difficulty is a huge combinatorial background of uncorrelated lepton pairs which mainly come from π^0 Dalitz decays and photon conversion in the detector material. A special attention should be paid to reduce this background as much as possible.

For dilepton studies, the electron and positron identification is based on a combination of measurements from three detector subsystems in order to achieve the best results: dE/dx in TPC, time-of-flight in TOF and ECal and E/p in ECal. The used identification method is as follows: for TPC tracks with a good match in TOF or ECal the time-of-flight measurement and momentum give an estimate of the particle velocity β . In addition, if the track reaches ECal, the calorimeter signal E for a measured momentum p provides another particle identification criterion E/p , which should be very close to 1 for electrons unlike for hadrons. Since there are several variables to select electrons, they can be combined to build a multivariate discriminator, which potentially could improve the selection quality [6,7] due to a better utilization of variable correlations and a possibility to include additional information such as, for example, the ECal shower topology (Figure 4).

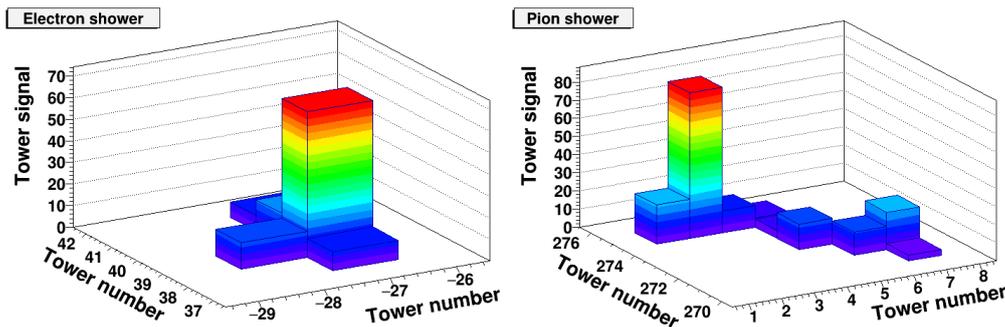


Figure 4: Left - ECal tower signal distribution obtained for a 0.5 GeV/c electron; right – the same for a 0.5 GeV/c pion. Horizontal axes represent the tower numbers in the longitudinal and tangential directions with respect to the beam line. Vertical axis shows the signal amplitude.

Another difficult physics topic, study of the charm production, could also benefit from multivariate analysis techniques [9] using information provided by the future silicon IT. Moreover, the new detector ability to select decays of short-lived particles (Figure 5 left) and high rate performance could potentially allow implementation of a high level trigger scheme for rare events produced in proton-proton collisions at high luminosity using extracted features of those events within machine learning approaches (Figure 5 right).

4. Distributed computing

Currently the problem of detector response simulation consists of two main steps, which require distributed computing application:

- 1) MC simulation with GEANT4:

for experiment purposes several million events need to be processed. Different events can be processed independently, so event-level data parallelism can be applied. Events might need to be processed more than once with software updates.

2) GAN training:

MC-generated events can be used as a training dataset for GAN model. While event-level data parallelism still can be applicable (some problems with model update according to gradients from different events to be solved), model-level parallelism needs to be considered, with different neurons/layers of the model trained independently.

Current generations of GPU/HPC are highly optimized for neural networks training, so JINR HybriLIT [10] cluster can be efficiently used for both of these problems.

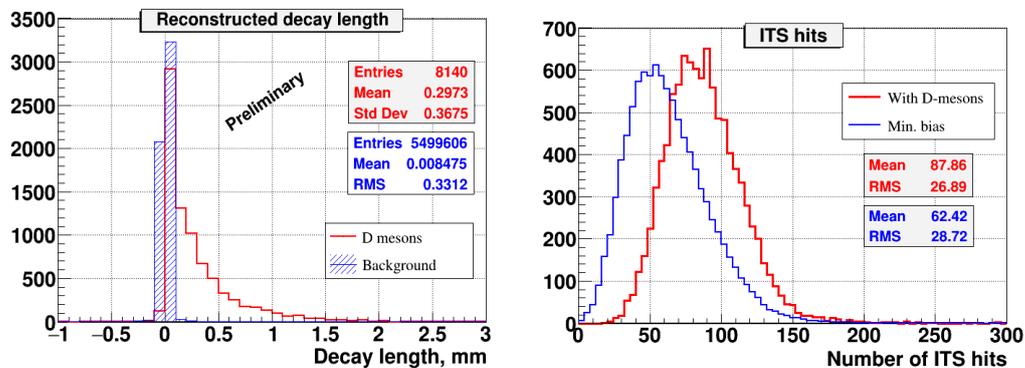


Figure 5: Left - reconstructed in IT decay length of $D^{+(\-)}$ -mesons from decays to $K^{-(+)}$ $\pi^{+(\-)}$ $\pi^{+(\-)}$ and background combinations; right - distributions of the number of IT hits in events with D-mesons and minimum bias events.

References

- [1] K. Deja et al., "Generative Models for Fast Cluster Simulations in the TPC for the ALICE Experiment", Available at: http://ii.pw.edu.pl/~ttrzcins/papers/ITSRCP_8.pdf
- [2] P. Musella, F. Pandolfi, "Fast and accurate simulation of particle detectors using generative adversarial neural networks", arXiv:1805.00850 [hep-ex]
- [3] S. Vallecorsa, "Generative models for fast simulation", Available at: https://indico.cern.ch/event/567550/papers/2656673/files/5841-SofiaVallecorsa_Plenary.pdf
- [4] Ch. Lester, "Machine Learning Based Simulation of Particle Physics Detectors", Available at: <https://www.hep.phy.cam.ac.uk/~lester/teaching/PartIIIProjects/2017-SeyonSivarijah-NeuralNet-JetSimulation.pdf>
- [5] L. deOliveira et al., "Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis", Available at: <http://inspirehep.net/record/1510258/files/s41781-017-0004-6.pdf>
- [6] O. Y. Derenovskaya, V. V. Ivanov, "Reconstruction and selection of $J/\psi \rightarrow e^+ e^-$ decays registered by CBM setup in 25 AGeV AuAu collisions", Phys. Part. Nucl. Lett. 11, 560 (2014).
- [7] S. Harabasz [HADES Collaboration], "Electron identification in Au+Au collisions at 1.23 GeV/u using multivariate analysis", J. Phys. Conf. Ser. 503, 012014 (2014).
- [8] J. Bouchet [STAR Collaboration], "Identification of charmed mesons using multivariate analysis in STAR experiment", J. Phys. Conf. Ser. 396, 022007 (2012).
- [9] A. Quintero, "Measurement of charm meson production in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV", Available at: https://drupal.star.bnl.gov/STAR/files/AQuintero_2016_final.pdf
- [10] JINR HybriLIT cluster: <http://hlit.jinr.ru/>