



MPD Collaboration Meeting 8-10 November 2022



# Event reconstruction status

Alexander Zinchenko



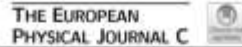
# Outline

- TPC fast digitizer
- TPC cluster finder
- Track reconstruction
- Application of machine learning for PID
- Hyperon reconstruction (background from event mixing)
- Machine learning at MPD
- Tracking in the end-cap region



# TPC fast digitizer

Eur. Phys. J. C (2021) 51:398  
https://doi.org/10.1146/epjc.2021.09366-4



Regular Article - Experimental Physics

## Simulating the time projection chamber responses at the MPD detector using generative adversarial networks

A. Maevskiy<sup>1,4</sup>, F. Ratnikov<sup>1,2,3</sup>, A. Zinchenko<sup>3,4</sup>, V. Riabov<sup>4,5</sup>

<sup>1</sup> HSE University, 20 Myasnikovskaya Ulitsa, Moscow, Russia  
<sup>2</sup> Yandex School of Data Analysis, 11-2 Timova Prizna Street, Moscow, Russia  
<sup>3</sup> Joint Institute for Nuclear Research, 6 Mich-Curie St, Dubna, Moscow Oblast, Russia  
<sup>4</sup> Petersburg Nuclear Physics Institute, 1, ul. Orlova-Zubovskaya, Gatchina, Leningradskaya Oblast, Russia

Received: 13 December 2020 / Accepted: 25 June 2021 / Published online: 10 July 2021  
© The Author(s) 2021

**Abstract** High energy physics experiments rely heavily on the detailed detector simulation models in many tasks. Running these detailed models typically requires a notable amount of the computing time available to the experiments. In this work, we demonstrate a new approach to speed up the simulation of the Time Projection Chamber tracker of the MPD experiment at the NICA accelerator complex. Our method is based on a Generative Adversarial Network – a deep learning technique allowing for implicit estimation of the population distribution for a given set of objects. This approach lets us learn and then sample from the distribution of raw detector responses, conditioned on the parameters of the charged particle tracks. To evaluate the quality of the proposed model, we integrate a prototype into the MPD software stack and demonstrate that it produces high-quality events similar to the detailed simulator, with a speed-up of at least an order of magnitude. The prototype is trained on the responses from the inner part of the detector and, once expanded to the full detector, should be ready for use in physics tasks.

### 1 Introduction

Computer simulations of high-energy physics experiments play a crucial role in a variety of relevant tasks, including detector geometry optimization [1,2], selecting best analysis strategies [3,4], and testing the Standard Model (SM) predictions and searching for new phenomena beyond the SM [5,6]. For a typical experimental data analysis, the number of simulated events usually translates directly to the uncertainty

of the final physics result. The amount of computational resources spent on the simulations usually takes a notable fraction of the total computing capabilities of an experiment and is comparable with that spent on the real data processing [7,8]. Therefore, faster approaches to event generation and simulation are in great demand for the existing and future high energy physics experiments.

The MPD detector is one of the two experiments at the NICA accelerator complex – a new heavy ion accelerator facility being constructed at the Joint Institute for Nuclear Research and located in Dubna, Russia [9,10]. The complex is designed to study the properties of dense baryonic matter. For the tracking, MPD utilizes a time projection chamber (TPC) in the central barrel [11]. TPC simulation is very CPU-intensive [12], and hence a fast simulation approach for TPC is highly desirable.

A typical approach to constructing models for fast simulation of particle physics detectors is to use a simplified detector geometry and a simplified model of the interaction of particles with matter [13]. This approach is justified for subsystems with a flat sensitive volume, such as silicon trackers, that measure the two-dimensional coordinate of a passing particle. For systems with a large volume, such as calorimeters or TPC-based trackers, this approach makes it difficult to achieve a reasonable compromise between accuracy and simulation speed.

Another fast simulation approach is an analytical parameterization of the detector responses, as can be seen in shower shape parameterizations for calorimeters [14]. This approach can significantly speed up the calorimeter simulation, but it makes it difficult to achieve high quality simulated data. A common solution for calorimeters is also to use the so-called “frozen showers” [13] when detailed simulated event responses are stored as a response library for subsequent reuse.

\* e-mail: artem.maevskiy@cern.ch (corresponding author)  
<sup>1</sup> e-mail: feder.ratnikov@cern.ch  
<sup>2</sup> e-mail: alexander.zinchenko@cern.ch  
<sup>3</sup> e-mail: v.riabov@gmail.com

## HSE team

## Generative Adversarial Networks for the fast simulation of the Time Projection Chamber responses at the MPD detector

A. Maevskiy<sup>1</sup>, F. Ratnikov<sup>1,2</sup>, A. Zinchenko<sup>3</sup>, V. Riabov<sup>4</sup>, A. Sukhorosov<sup>1</sup> and D. Evdokimov<sup>1</sup>

<sup>1</sup> HSE University, 20 Myasnikovskaya Ulitsa, Moscow, Russia  
<sup>2</sup> Yandex School of Data Analysis, 11-2 Timova Prizna Street, Moscow, Russia  
<sup>3</sup> Joint Institute for Nuclear Research, 6 Mich-Curie St, Dubna, Moscow Oblast, Russia  
<sup>4</sup> Petersburg Nuclear Physics Institute, 1, ul. Orlova-Zubovskaya, Gatchina, Leningradskaya Oblast, Russia  
E-mail: artem.maevskiy@cern.ch

**Abstract.** The detailed detector simulation models are vital for the successful operation of modern high-energy physics experiments. In most cases, such detailed models require a significant amount of computing resources to run. Often this may not be affordable and less resource-intensive approaches are desired. In this work, we demonstrate the applicability of Generative Adversarial Networks (GAN) as the basis for such fast-simulation models for the case of the Time Projection Chamber (TPC) at the MPD detector at the NICA accelerator complex. Our prototype GAN-based model of TPC works more than an order of magnitude faster compared to the detailed simulation without any noticeable drop in the quality of the high-level reconstruction characteristics for the generated data. Approaches with direct and indirect quality metrics optimization are compared.

### 1. Introduction

Simulation of particle detectors is inevitable in the High Energy Physics (HEP) experiments. For a typical HEP data analysis, the limited size of simulated data samples often contributes directly to the uncertainty in the final result. Since the number of simulated events that one can afford to produce is constrained by the computational efficiency of the simulation algorithms, faster algorithms are always desired [1].

Computational efficiency of the detailed simulation is often limited by the fine granularity of the physics simulation steps being performed. Therefore, a speed-up may be achieved by aggregating a sequence of such steps with a single estimate of the probability distribution for the last step output parameters, conditioned by the first step inputs. An important requirement for such a probability distribution estimate is that it should allow for efficient sampling. Generative Adversarial Networks (GAN) [2] are a good candidate for such a parametric estimate since they only require a forward pass through a neural network to generate new samples. In this work, we demonstrate an application of GANs for building a fast-simulation model of the Time Projection Chamber (TPC) detector at the MPD experiment at the NICA accelerator complex [3].

arXiv:2207.04340v1 [hep-ex] 9 Jul 2022

## Generative Surrogates for Fast Simulation: TPC Case

Fedor Ratnikov<sup>1,2\*</sup>, Artem Maevskiy<sup>1</sup>, Alexander Zinchenko<sup>3</sup>, Victor Riabov<sup>4</sup>, Alexey Sukhorosov<sup>1</sup>, Dmitrii Evdokimov<sup>1</sup>

<sup>1</sup> HSE University, 20 Myasnikovskaya Ulitsa, Moscow, Russia  
<sup>2</sup> Yandex School of Data Analysis, 11-2 Timova Prizna Street, Moscow, Russia  
<sup>3</sup> Joint Institute for Nuclear Research, 6 Mich-Curie St, Dubna, Moscow Oblast, Russia  
<sup>4</sup> Petersburg Nuclear Physics Institute, 1, ul. Orlova-Zubovskaya, Gatchina, Leningradskaya Oblast, Russia

### Abstract

Simulation of High Energy Physics experiments is widely used, necessary for both detector and physics studies. Detailed Monte-Carlo simulation algorithms are often limited due to the computational complexity of such methods, and therefore faster approaches are desired. Generative Adversarial Networks (GANs) are well suited for aggregating a number of detailed simulation steps into a surrogate probability density estimator readily available for fast sampling. In this work, we demonstrate the power of the GAN-based fast simulation model on the case of simulating the response for the Time Projection Chamber (TPC) in the MPD experiment at the NICA accelerator complex. We show that our model can generate high-fidelity TPC responses, while accelerating the TPC simulation by at least an order of magnitude. We describe alternative representation approaches for this problem and also outline the roadmap for the deployment of our method into the software stack of the experiment.

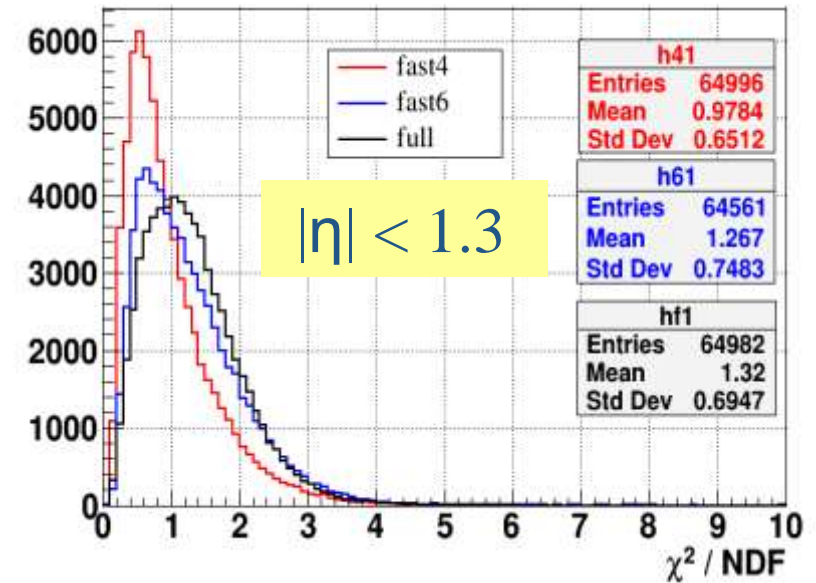
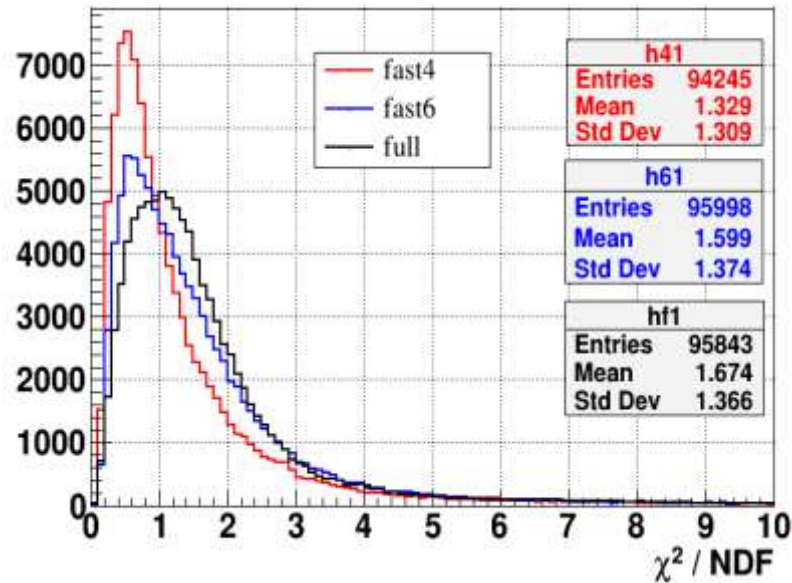
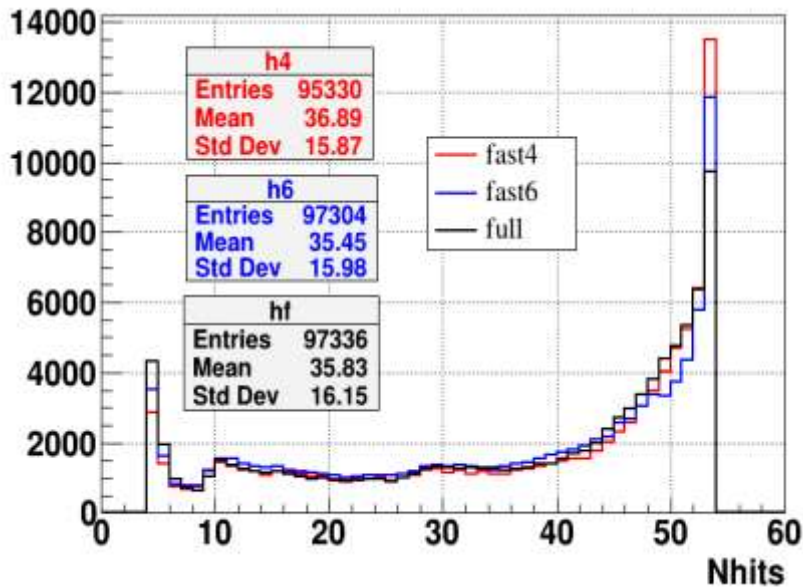
**Keywords:** fast simulation, time projection chamber, generative adversarial network

\*Corresponding author  
Email addresses: feder.ratnikov@gmail.com (Fedor Ratnikov), artem.maevskiy@cern.ch (Artem Maevskiy), alexander.zinchenko@cern.ch (Alexander Zinchenko), rriabov@gmail.com (Victor Riabov), asukhorosov@phs.jinr.ru (Alexey Sukhorosov), de-evd@mail.ru (Dmitrii Evdokimov)

Preprint submitted to Nuclear Instruments and Methods in Physics Research A July 11, 2022



# TPC fast digitizer



Future steps:

- ❑ Add the 7 input track parameter;
- ❑ Generate signals for 3 (2) neighbor padrows (to account for correlations)



# TPC cluster finder

**Tasks:** Two track separation (for femtoscopy) and cluster charge determination (for  $dE/dx$  identification)

Fresh ideas are welcome (Machine Learning?)



# Track reconstruction: reconstructed track data members

Int\_t fID; ///< trackID

Int\_t fNhits; ///< number of hits (just for convenience of tree browsing)

TrackDir fTrackDir; ///< track direction (inward/outward)

TrackType fTrackType; ///< track type (kBarrel/kEndcap)

Int\_t fLastLay; ///< the outermost layer number achieved

Int\_t fNofWrong; ///< number of wrong hits (different from the track ID)

TString fNode; ///< node path (for local track parameters)

TString fNodeNew; ///< node path (for local track parameters)

Double32\_t fPartID; ///< particle ID - dE/dx

Double32\_t fPos; ///< current track radial position

Double32\_t fPosNew; ///< propagated track radial position

Double32\_t fPosAtHit; ///< radial position of outermost hit

Double32\_t fChi2; ///< track Chi2

Double32\_t fChi2Vertex; ///< track Chi2-distance to primary vertex

Double32\_t fLength; ///< track length

Double32\_t fLengAtHit; ///< track length at last hit

TMatrixD \*fParam; ///< pointer to track parameter matrix (5x1)

TMatrixD \*fParamAtHit; ///< pointer to track parameter matrix at last hit

TMatrixDSym \*fCovar; ///< pointer to covariance matrix (15)

TMatrixDSym \*fWeightAtHit; ///< pointer to weight matrix at last hit

TVector3 fVertex; ///< track vertex – not used

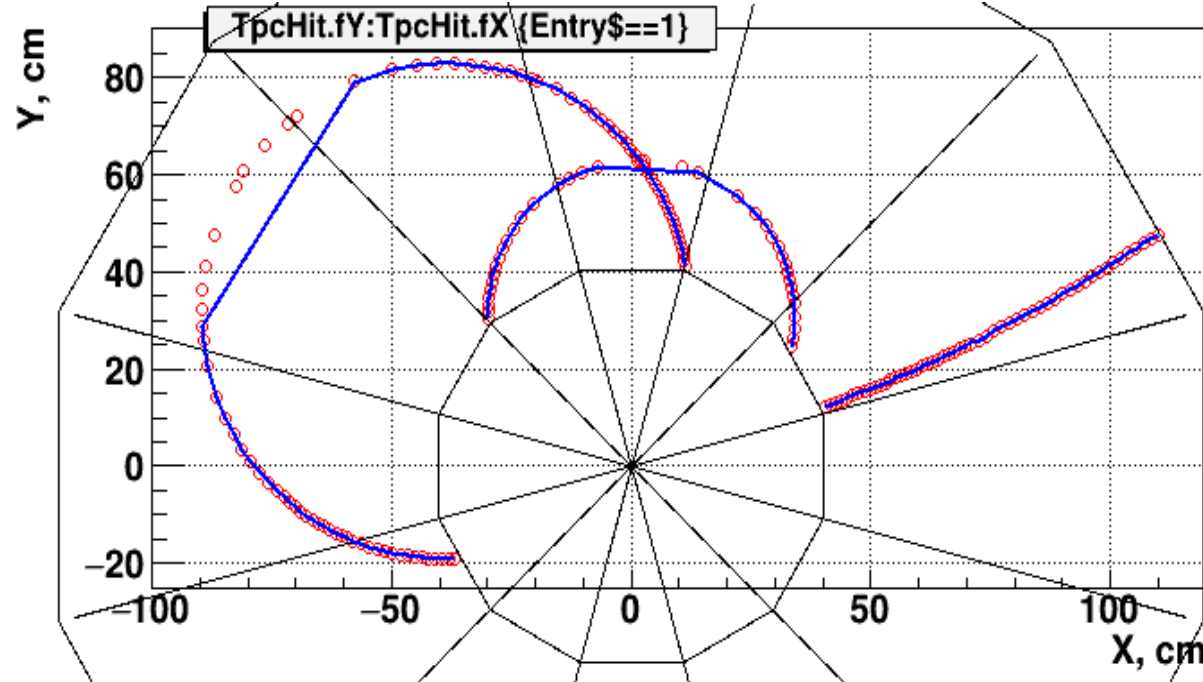
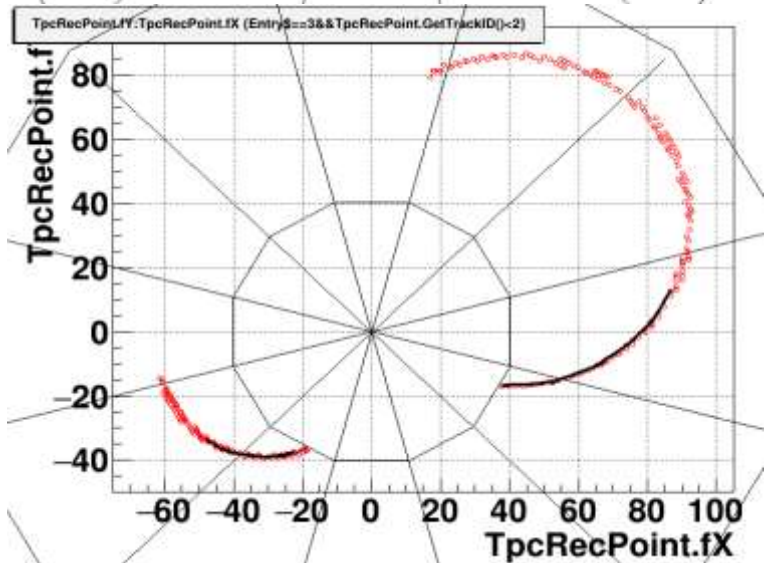
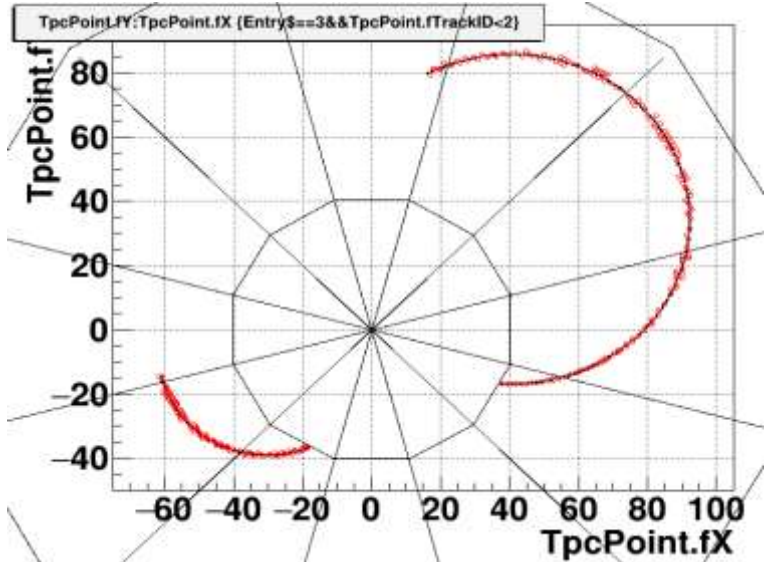
TrackFlag fFlag; ///< track flag

TClonesArray \*fTrHits; ///< track hits



# Track reconstruction: low- $p_T$ tracks

I. Tserruya,  
S. Rode



Required some changes of the Kalman filter main engine.

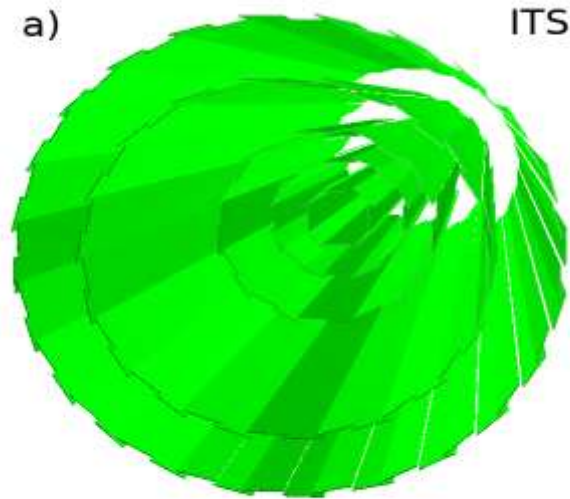
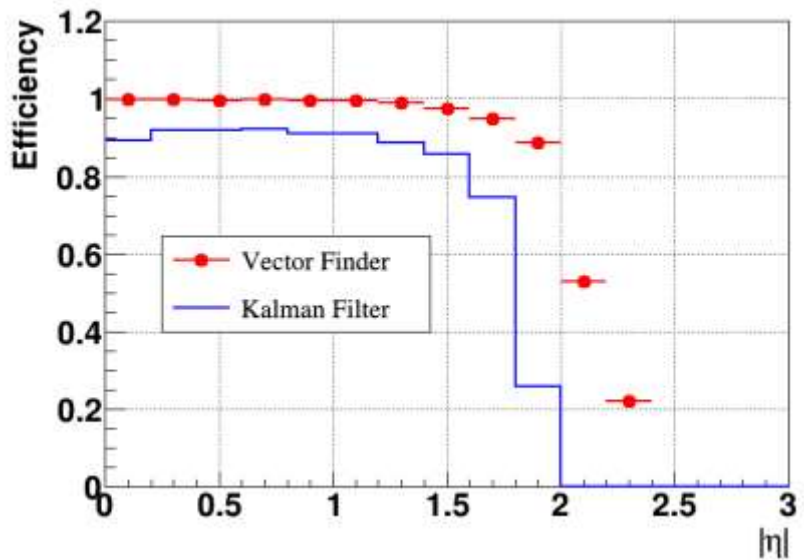
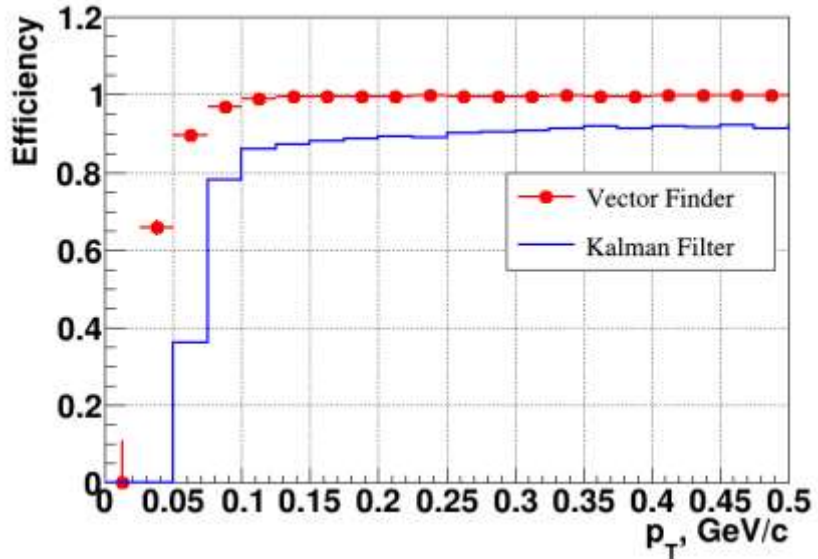
Some additional modifications might be necessary for realistic clusters.

Contingent on TPC response simulation for large crossing angles.

Not clear if it is possible on a short time scale.

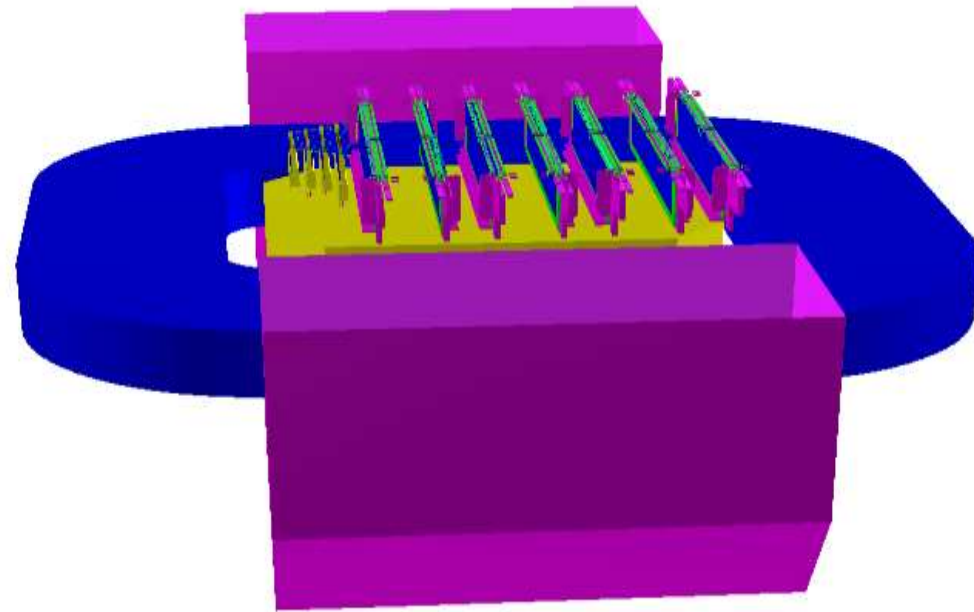


# Track reconstruction: Vector Finder toolkit for ITS



D. Zinchenko

BM@N inner tracker:  
VF with ML plugins

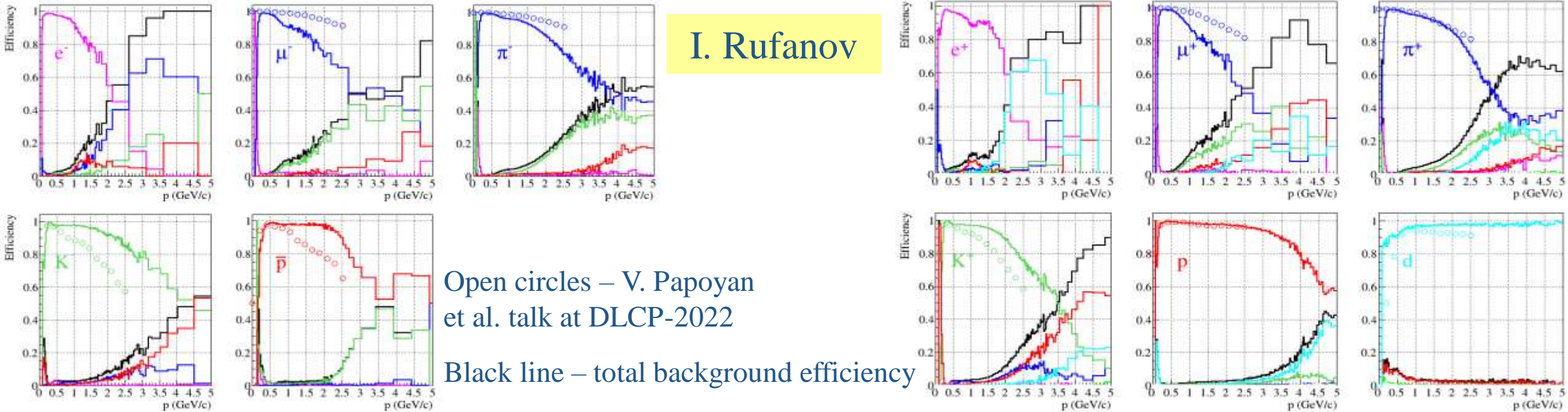






# TMVA PID for TOF Matched TPC Tracks

I. Rufanov



Open circles – V. Papoyan et al. talk at DLCP-2022

Black line – total background efficiency

- ❑ Input data: MPDROOT simulation of PHQMD min. bias BiBi collisions at 9.2 GeV
- ❑ TMVA method: Neural Networks with Multilayer Perceptrons (MLP)
- ❑ Track classification variables: Q, momentum, dE/dx, TOF, N of hits, pseudorapidity and DCA
- ❑ Currently, muons and pions are considered as a same classification case

**Plans:** find out the reason of low-momentum miss-classification; separate muon and pion classification cases; choose optimal set of variables for classification (include azimuthal angle)



# Hyperon reconstruction (background evaluation)

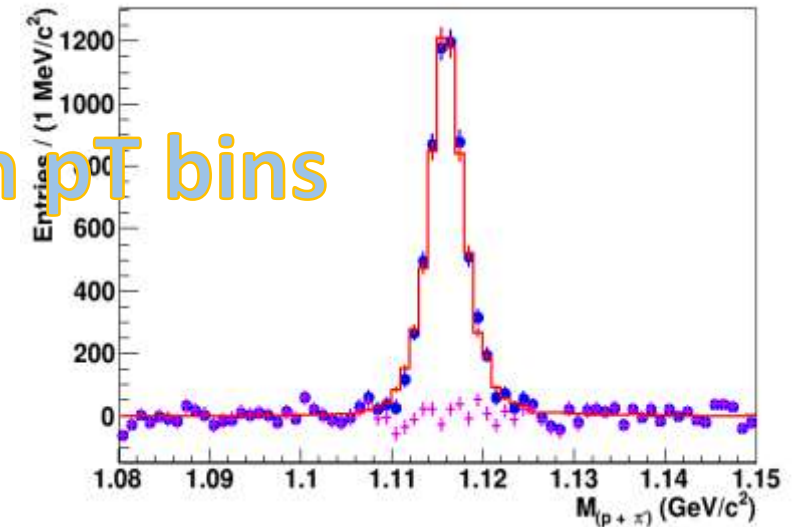
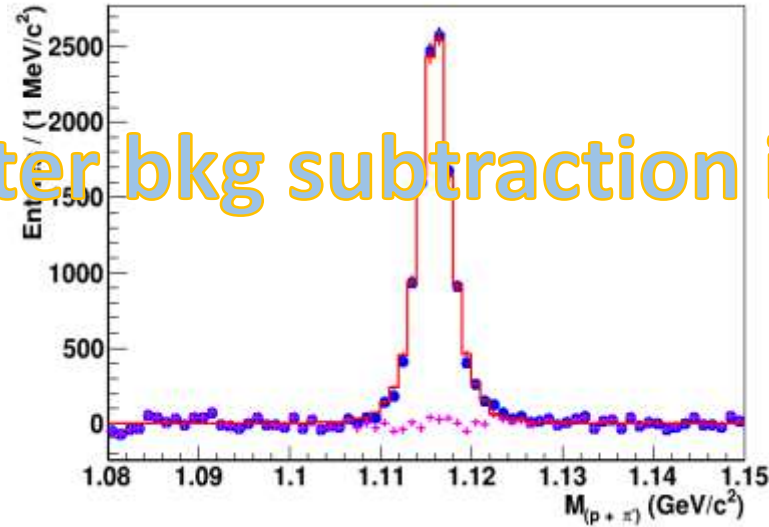
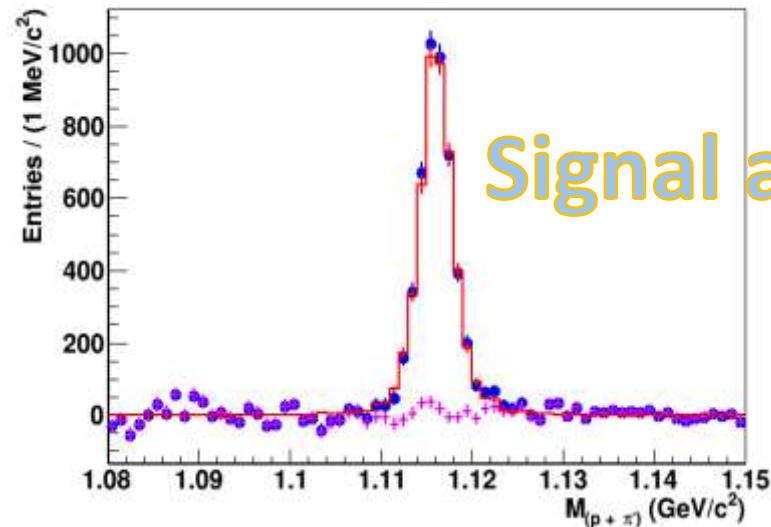
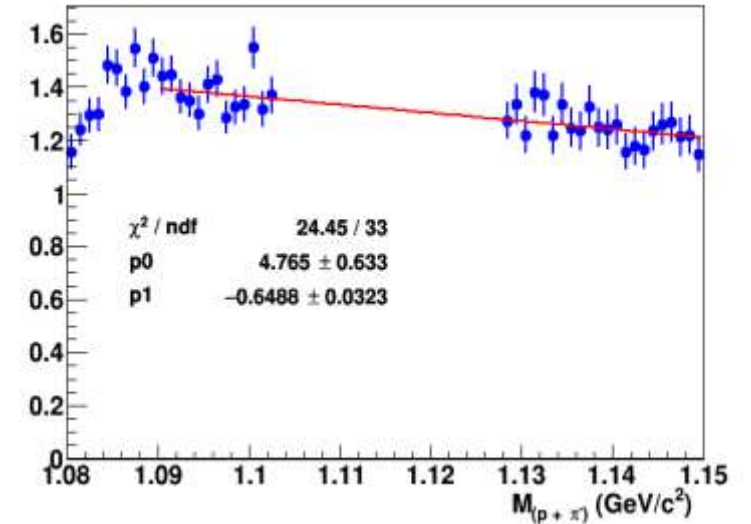
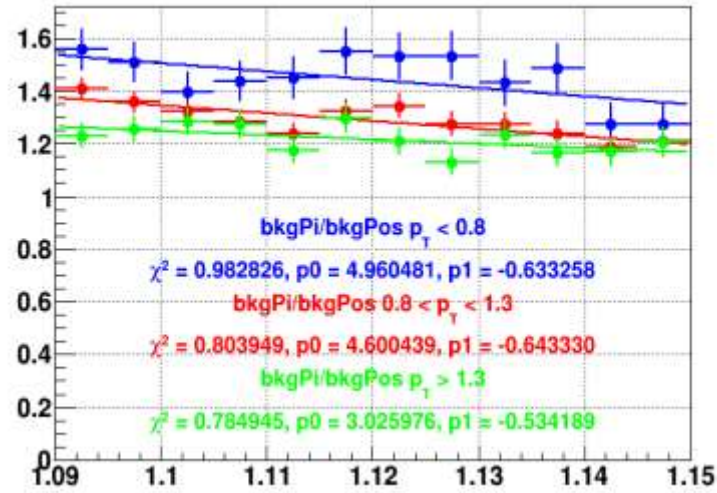
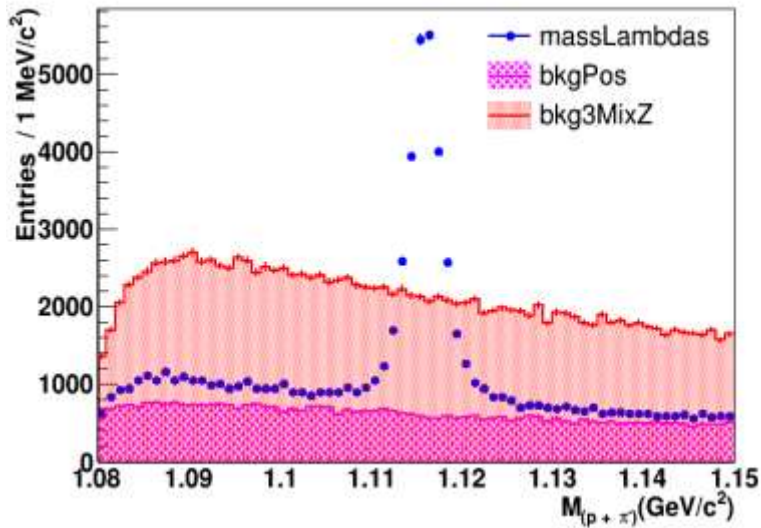
**Like-sign pairs:** pro – takes into account correlated effects,  
contra – different contamination from PID, limited statistics

**Event mixing:** pro – the same contamination from PID, high  
statistics



# $\Lambda$ reconstruction (bkg from like-sign pairs)

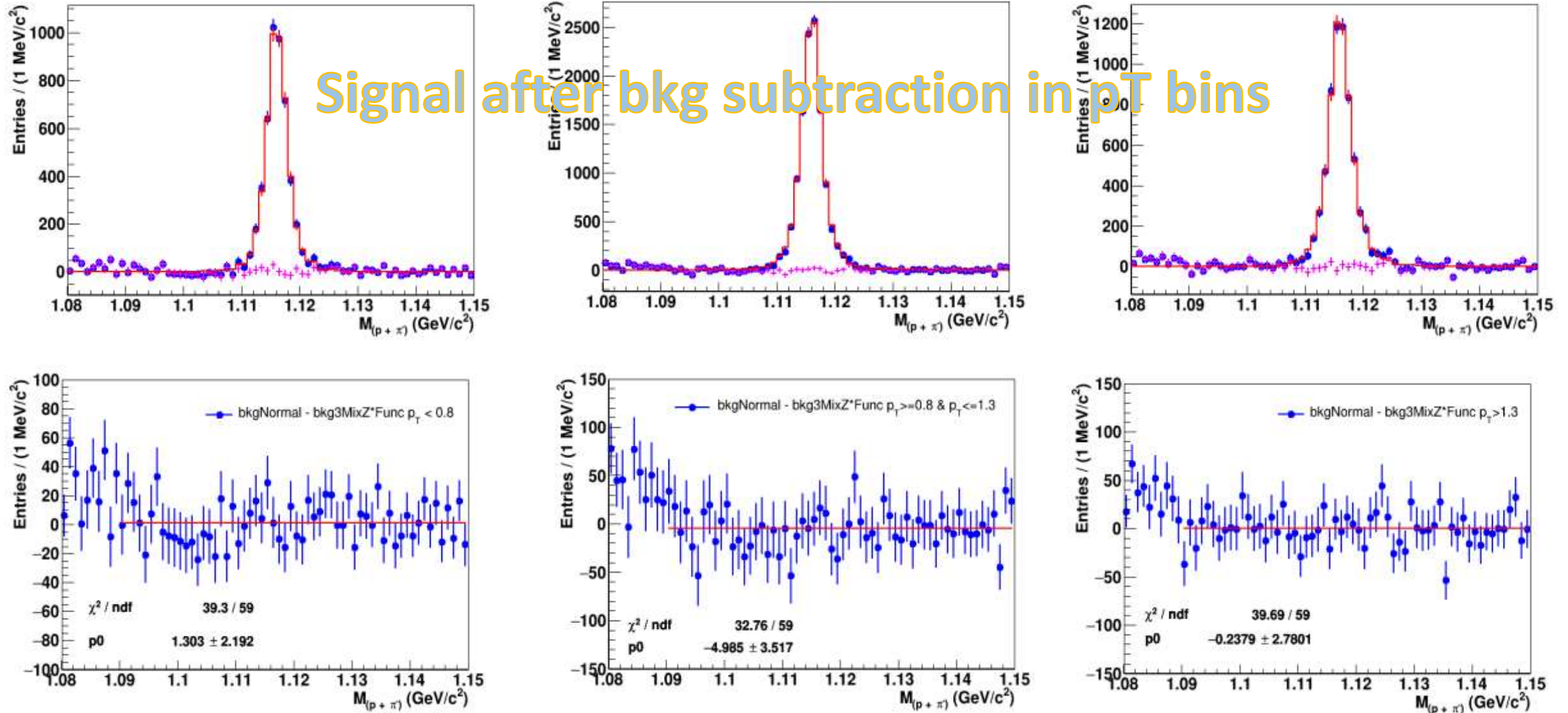
D. Suvarieva



Signal after bkg subtraction in  $p_T$  bins



# $\Lambda$ reconstruction (bkg from event mixing)



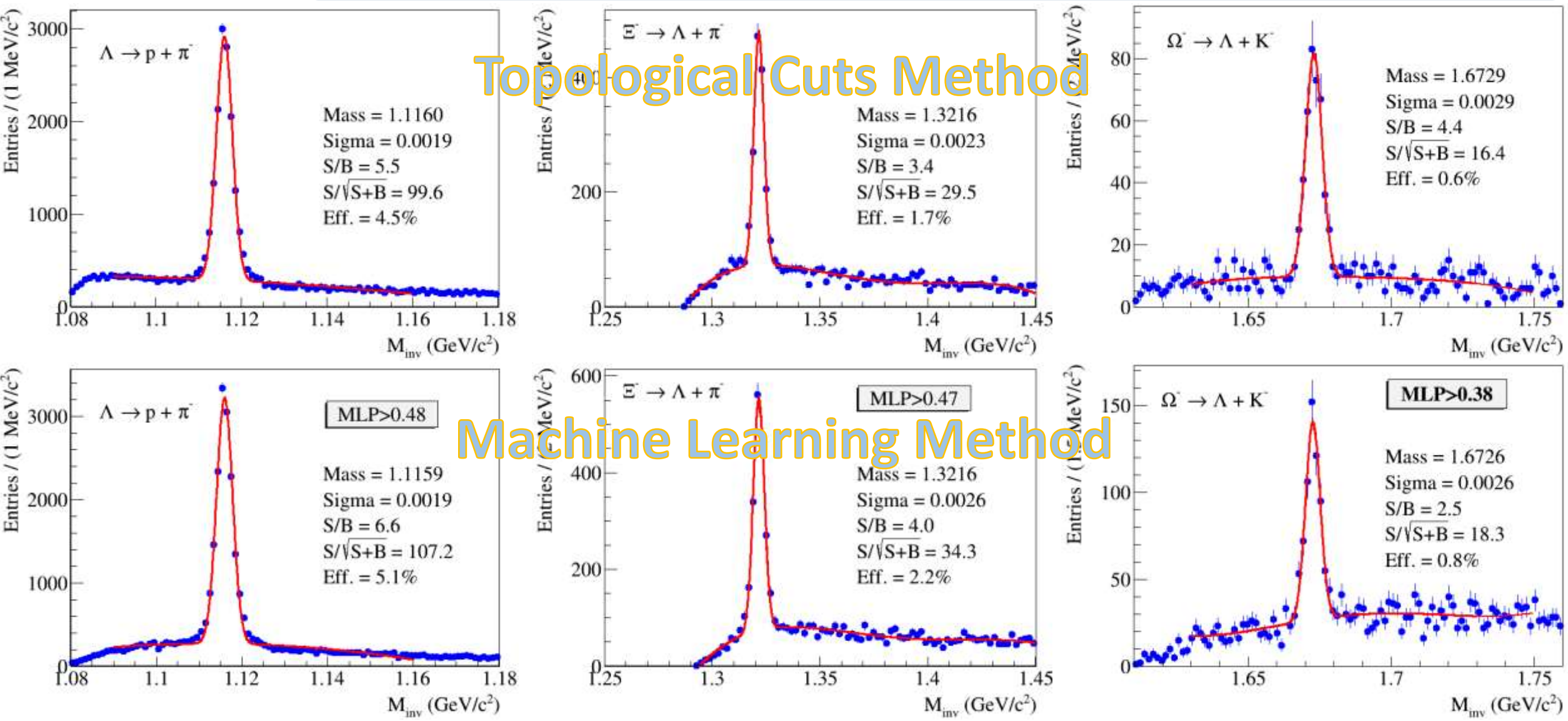


# Hyperon reconstruction: TC vs TMVA

V. Vasendina

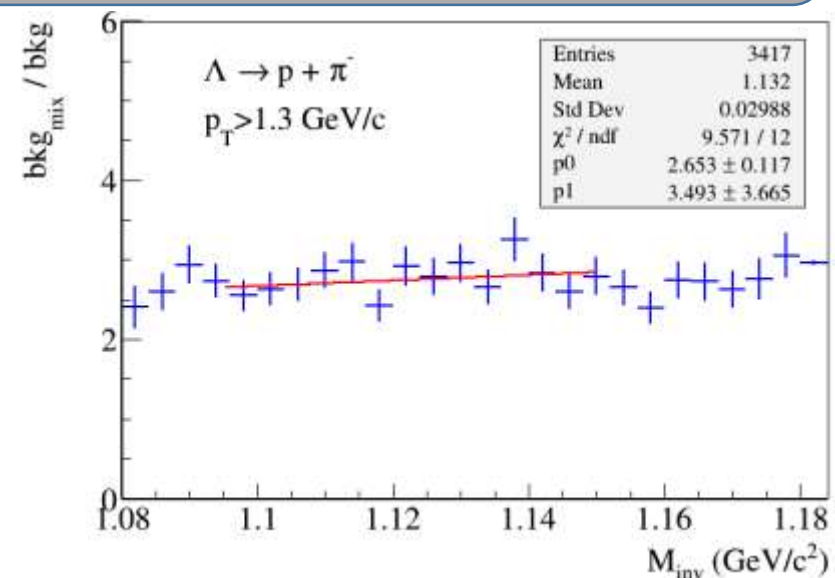
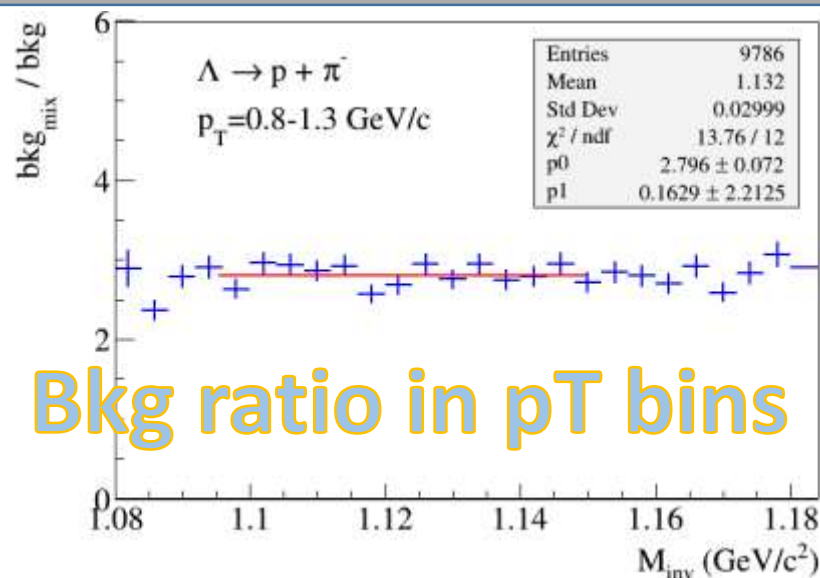
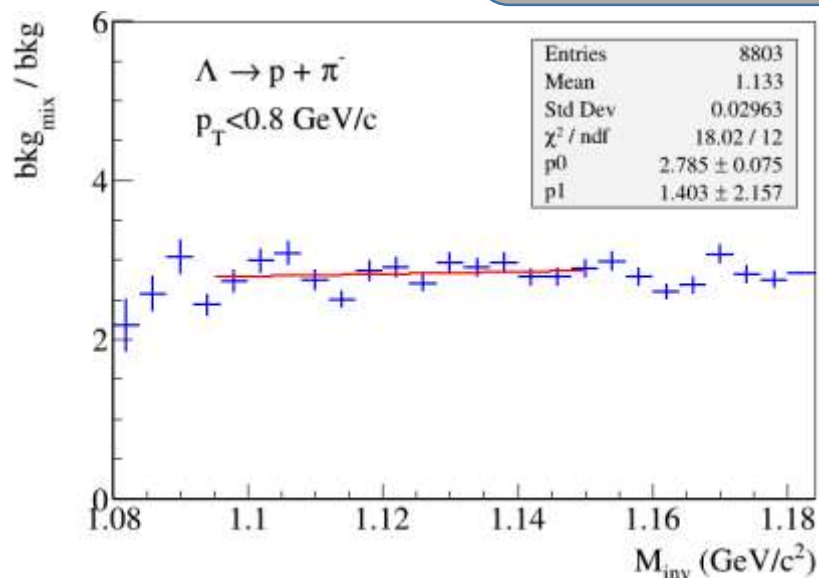
Topological Cuts Method

Machine Learning Method

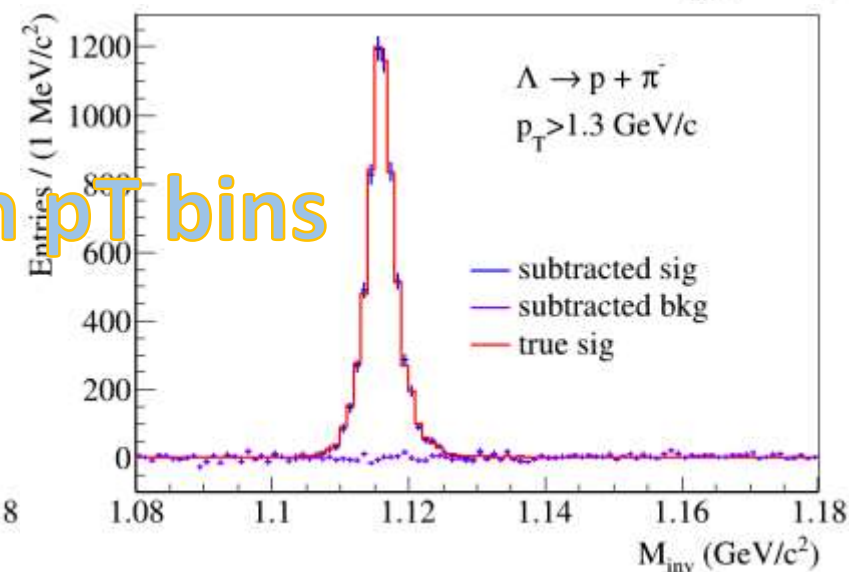
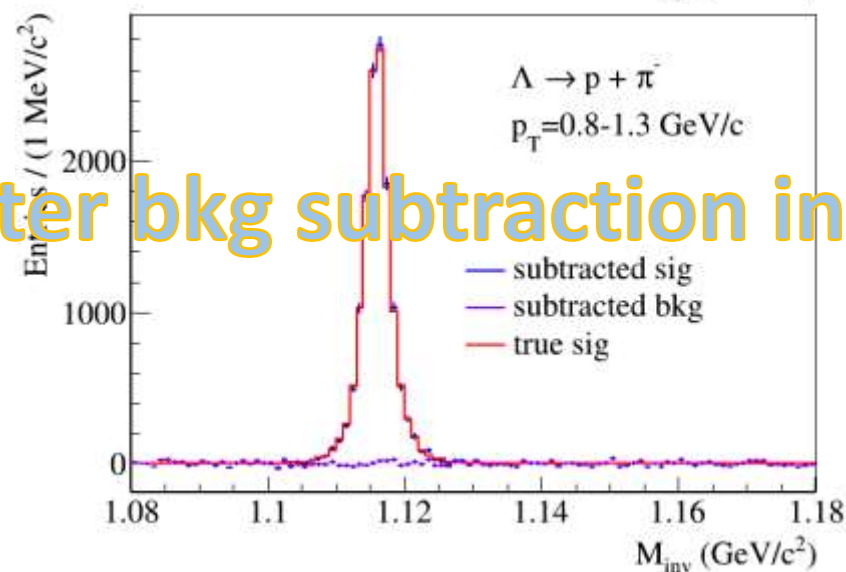
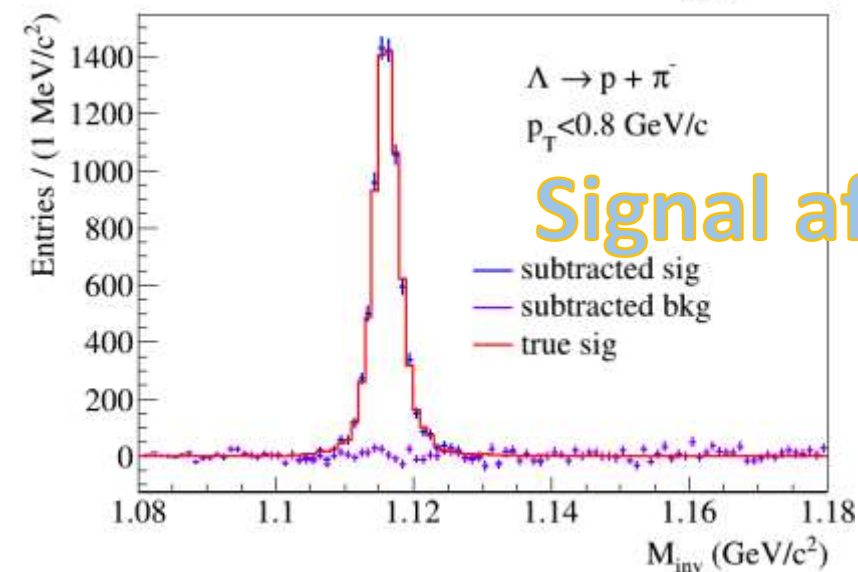




# $\Lambda$ reconstruction: TMVA & event mixing



Bkg ratio in pT bins



Signal after bkg subtraction in pT bins



# Application of Machine Learning at MPD

*Proceedings of the VIII International Conference "Distributed Computing and Grid-technologies in Science and Education" (GRID 2018), Dubna, Moscow region, Russia, September 10 - 14, 2018*

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

D.A. Zinchenko <sup>1,\*</sup>, E.G. Nikonov <sup>2</sup>, A.I. Zinchenko <sup>1</sup>

<sup>1</sup> VBLHEP, JINR, Dubna, Russia

<sup>2</sup> LIT, JINR, Dubna, Russia

E-mail: \*zinchenk1994@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 algorithm 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

Simulation of TPC and EMC response using generative models

Multivariate analysis for dilepton and open charm selection

+

Track reconstruction

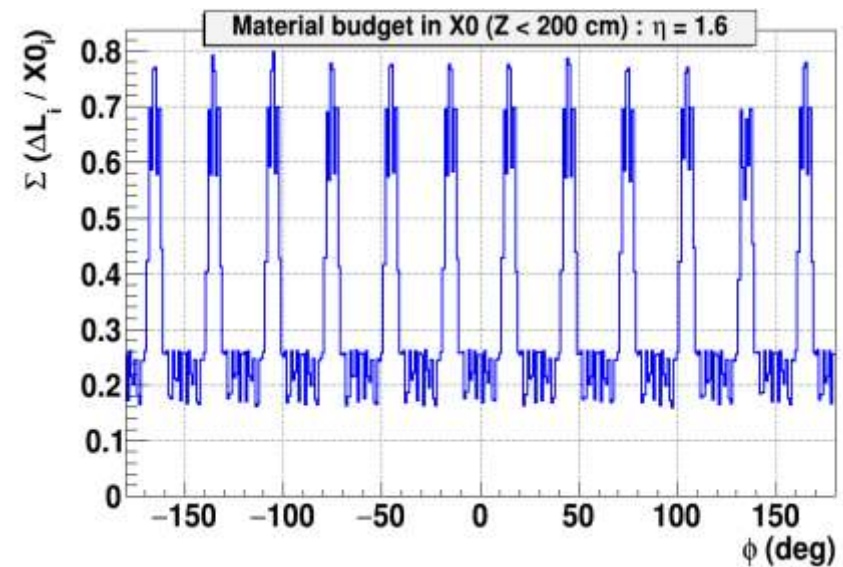
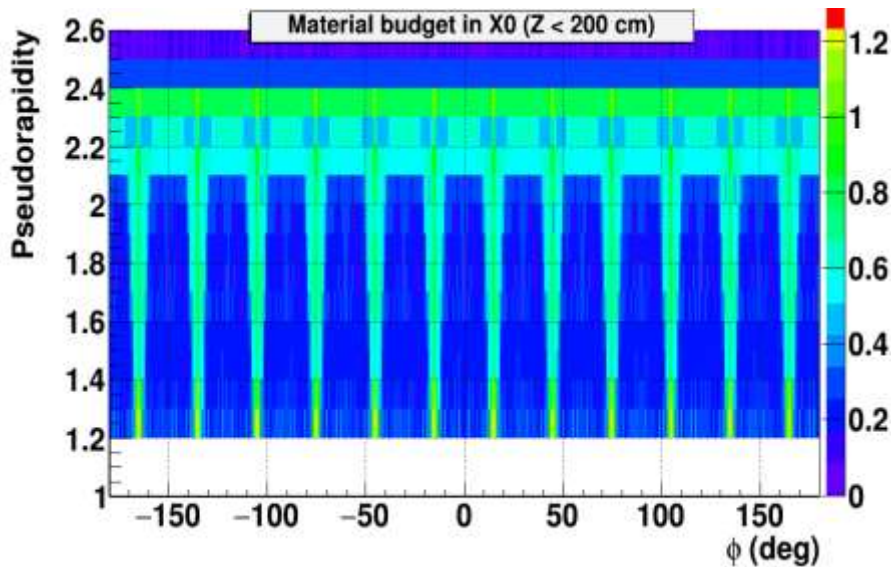
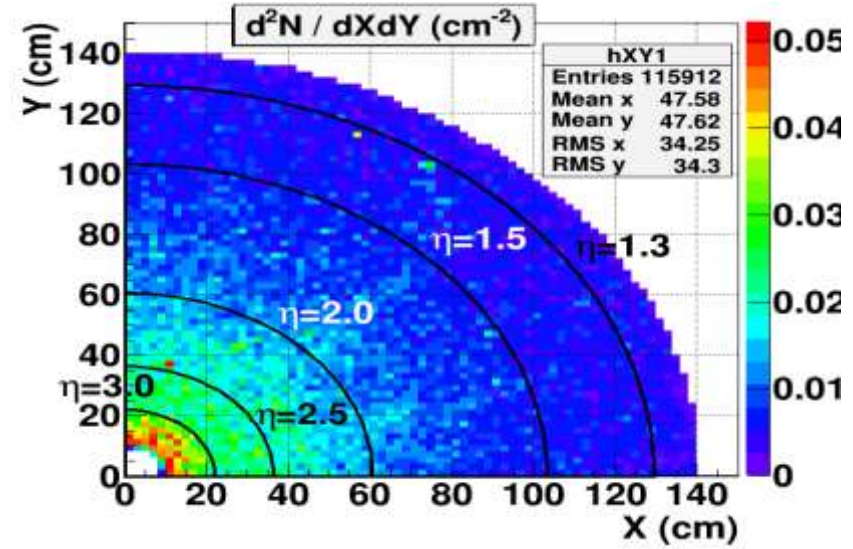
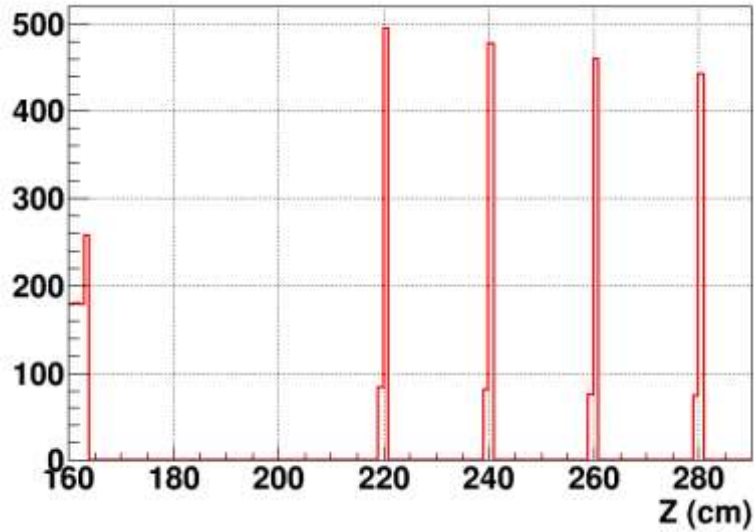
Particle identification

Hyperon selection

CEJIB-HEP-001/2018-2287/615-419-papers-119.pdf



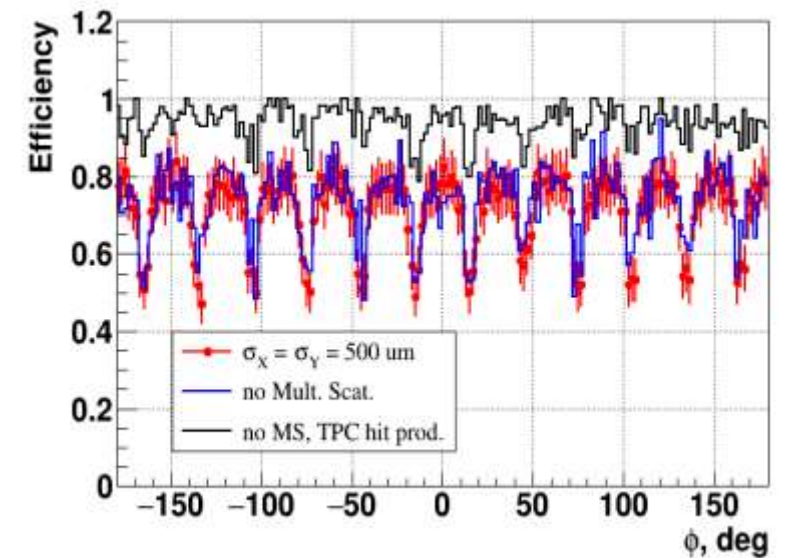
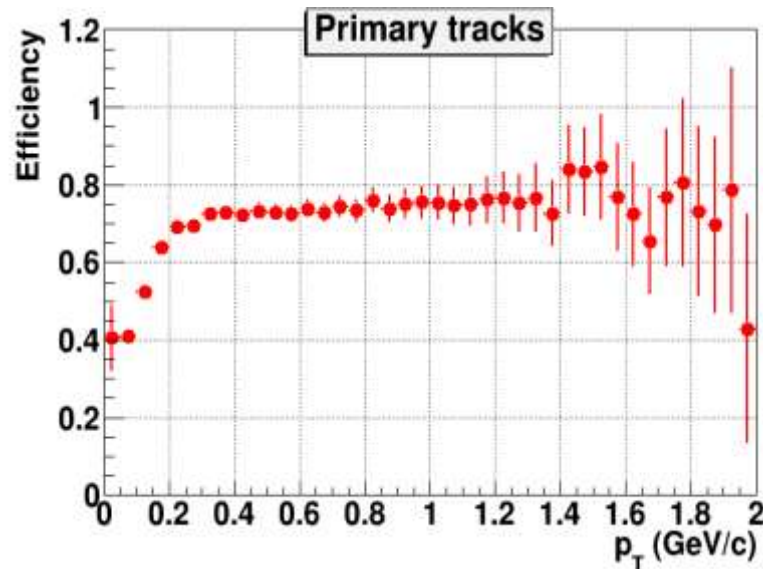
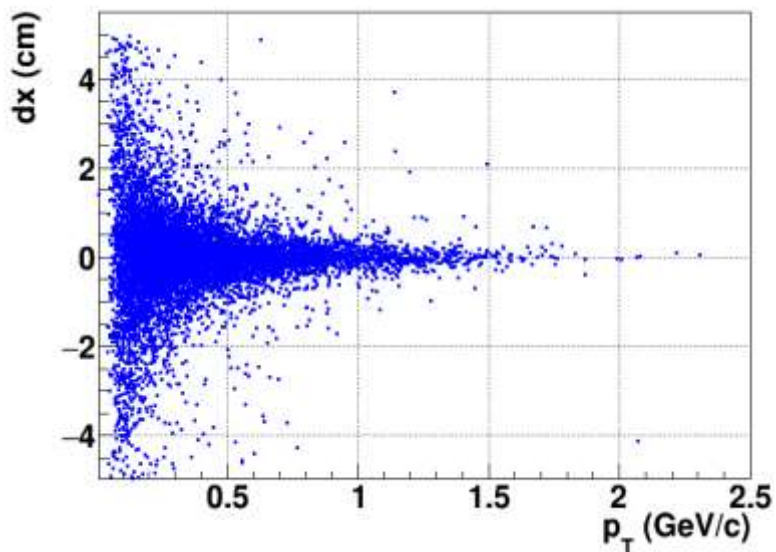
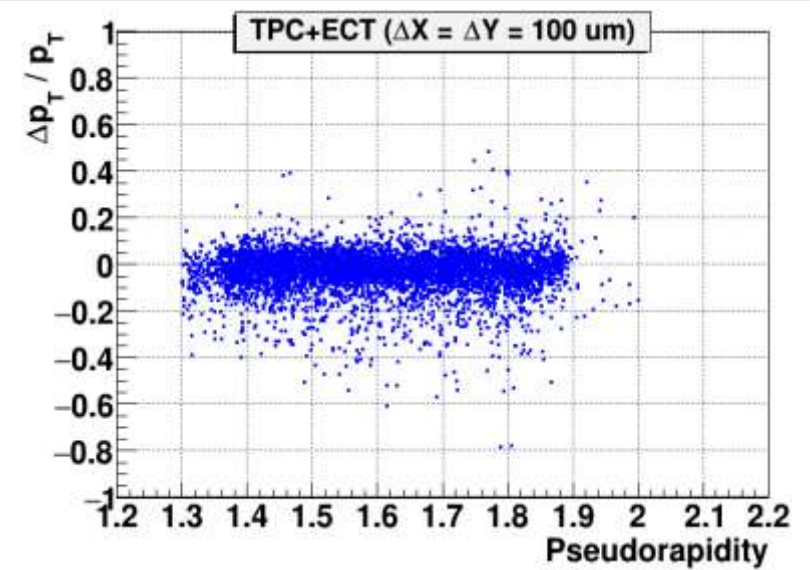
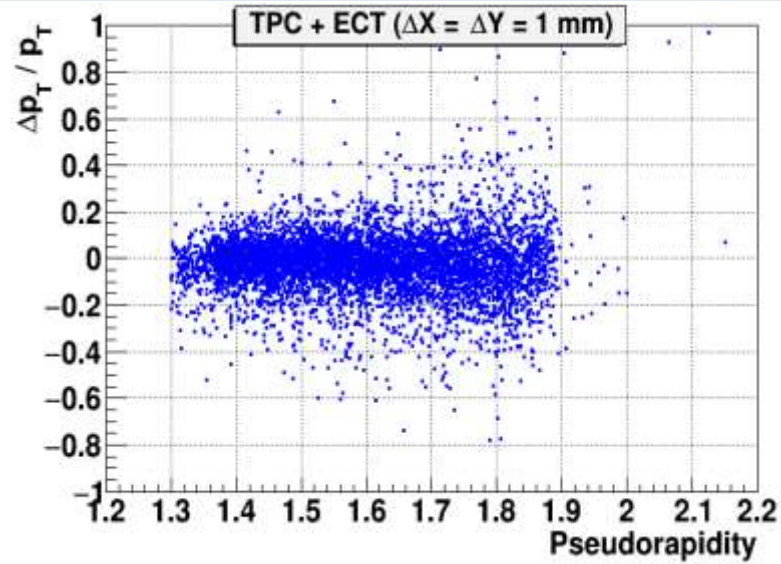
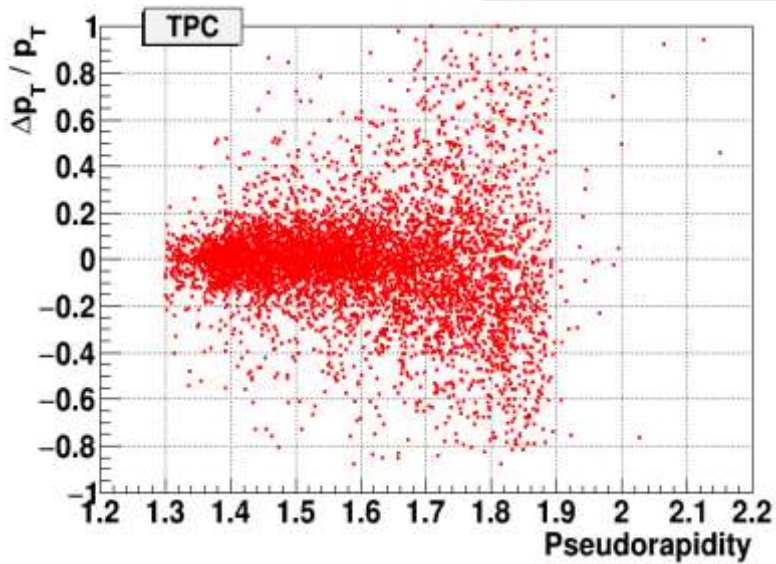
# Tracking in the end-cap region







# Tracking in the end-cap region





# Tracking in the end-cap region

- ❑ To continue with this, some technological solutions should be proposed – GEM detectors (estimated number of channels in double-sided GEMs with strip pitch of 1 mm exceeds that of the [BM@N](#) GEM tracker); scintillating fibers(?)
- ❑ Cluster finder for large dip angles might require some work
- ❑ Track fitting in the forward region with material budget map



# Summary / Outlook



- There are several active tasks on reconstruction