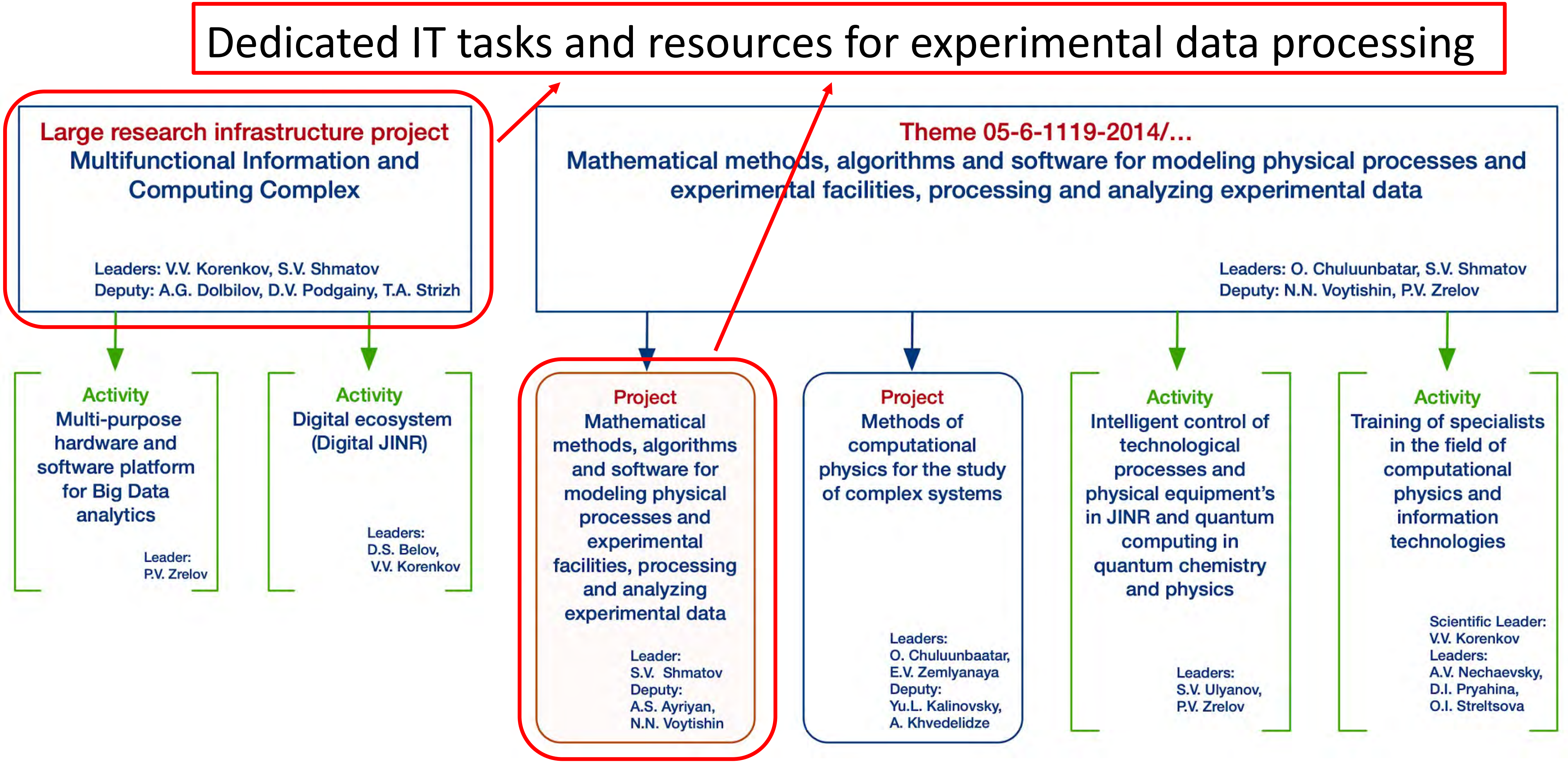
Methods and Algorithms of Experimental Data Processing @ JINR

Nikolay Voytishin
On behalf of
Meshcheryakov Laboratory of Information Technologies, JINR

The XVI International School-Conference
“The Actual Problems of Microworld Physics”

Information Technologies @ JINR



The Goals and Objectives of the Project

The project is aimed at

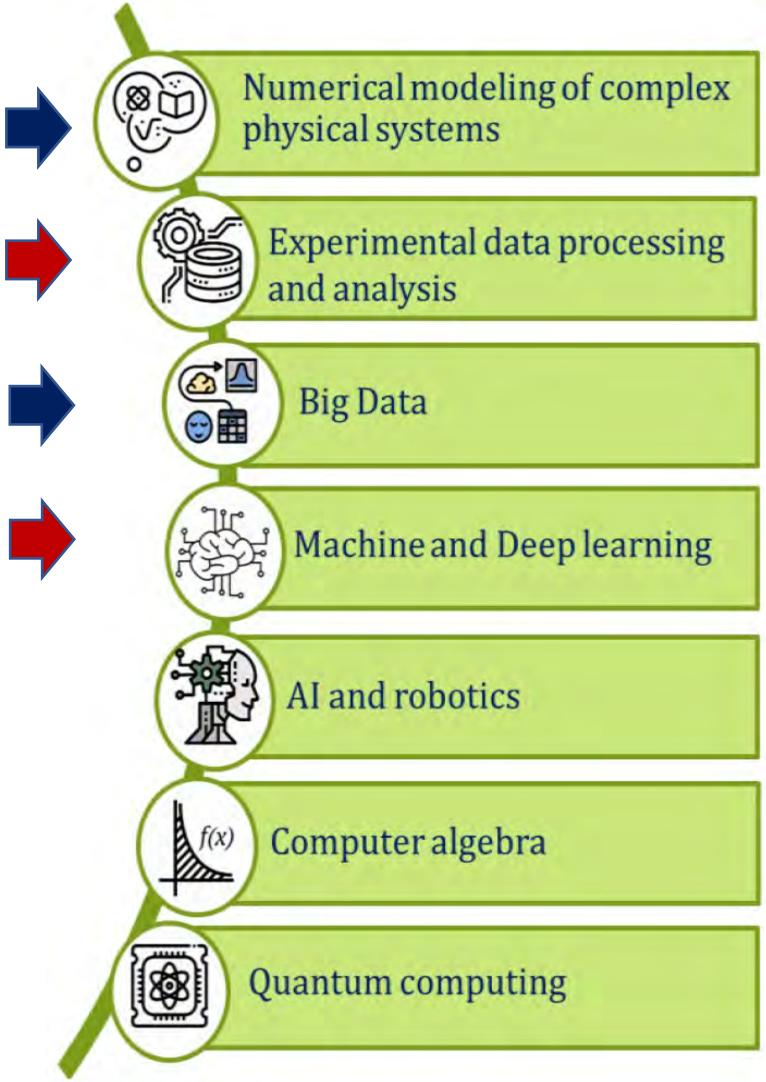
- organizing and providing computational support for the physics research programs implemented with the participation of JINR
- the development of mathematical methods and software for modeling physical processes and experimental facilities, processing and analyzing experimental data in the field of elementary particle physics, nuclear physics, neutrino physics, condensed matter, radiobiology, etc.

The particular attention is paid to the creation of systems for the distributed processing and analysis of experimental data as well as information and computing platforms to support research at JINR and other world centers.

The priorities are mathematical and computational physics to support the JINR large research infrastructure projects, and first of all the experiments at the NICA accelerator complex and the Baikal-GVD neutrino telescope.

Further cooperation will also be continued with the experiments at the largest world accelerator centers (CERN, BNL, etc.), experiments in the field of neutrino physics and astrophysics, radiobiological research programs.

The possibility of using the developed methods and algorithms within other fundamental science and applied projects is being considered.



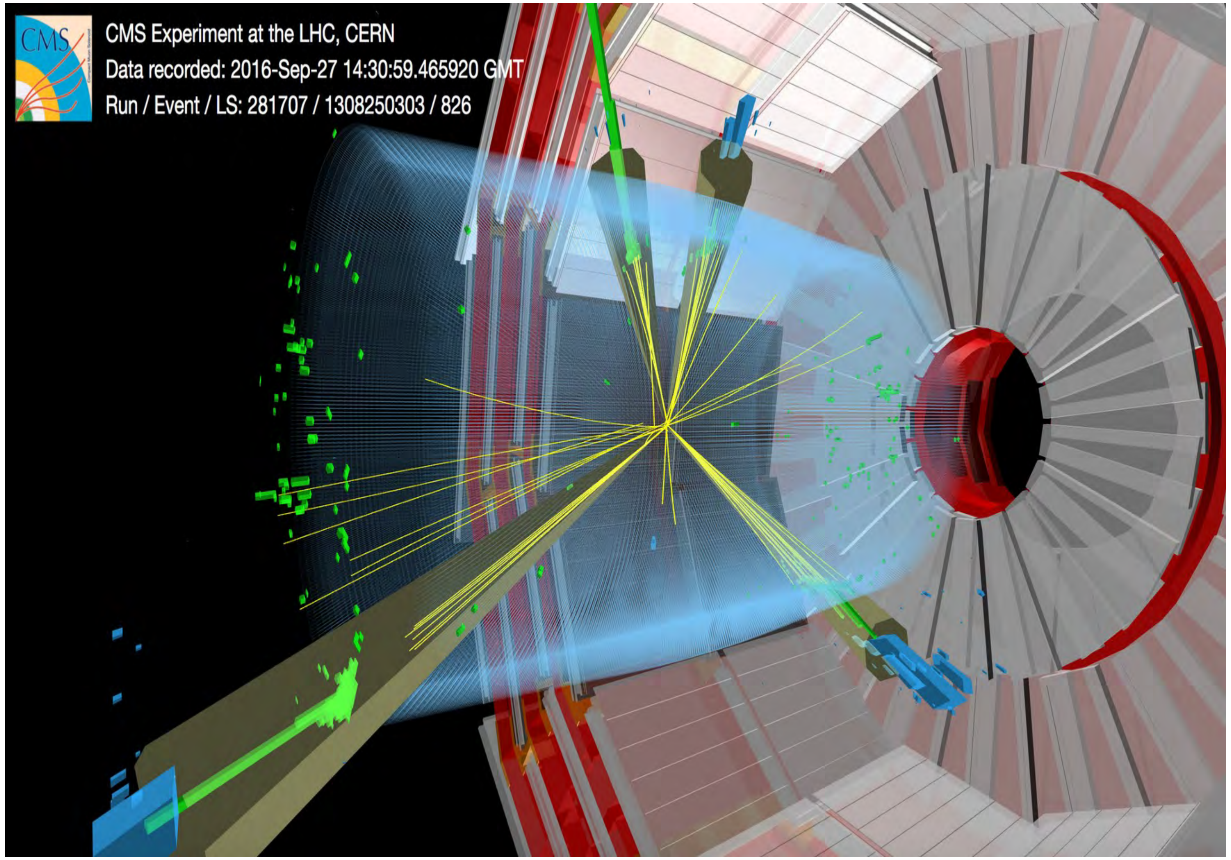
The Project Structure

Simulation of Physics Processes and Facilities	Reconstruction and Data Analysis	Software Environment for Experiments
Physics event simulation	Particle trajectory reconstruction	Data processing and analysis models
GEANT-simulation of experimental setups	Particle identification	Data models
	Reconstruction of physics processes	Software platforms and systems
	Experimental data analysis	Development and maintenance of DBs
		Event visualization

The main strategy is to use common solutions and methods for different experiments

Simulation of physical processes and experimental facilities

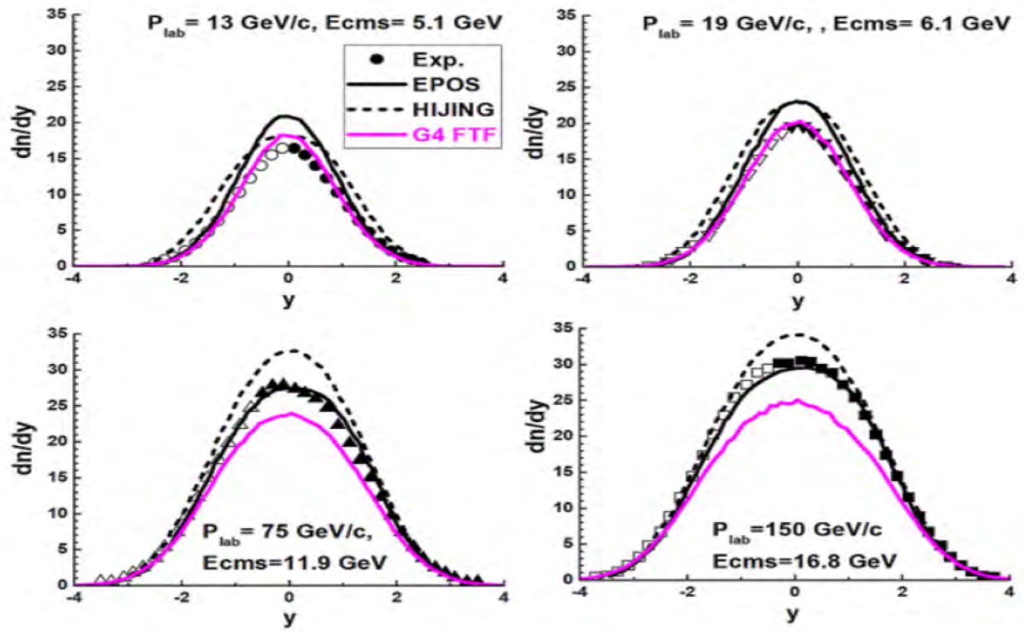
- analytical and numerical calculations of physical processes, software optimization, including tuning and adaptation of physics event generators;
- MC event production, development and support of information systems for event catalogues;
- participation in the creation of computer models of experimental facilities and simulation of elementary particles passing through them based on GEANT4 (and others) and fast simulation of the response of the detectors.



Modeling Experimental Facilities for the NICA, LHC, etc.

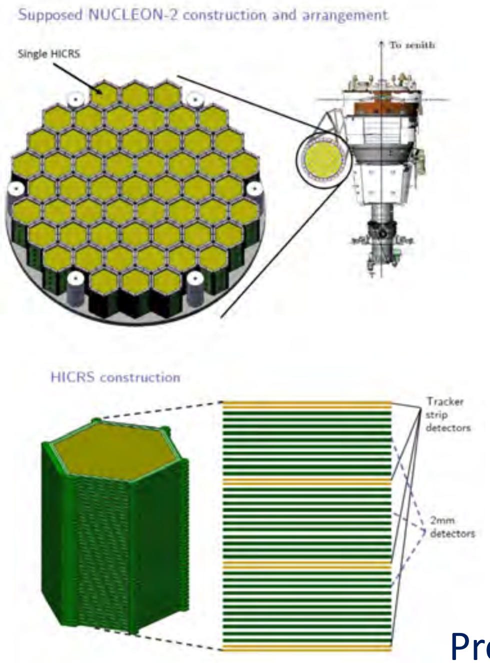
- Development, verification, validation and application of FTF (Fritiof) and QGSM (Quark-Gluon-String-Model) hadronic models

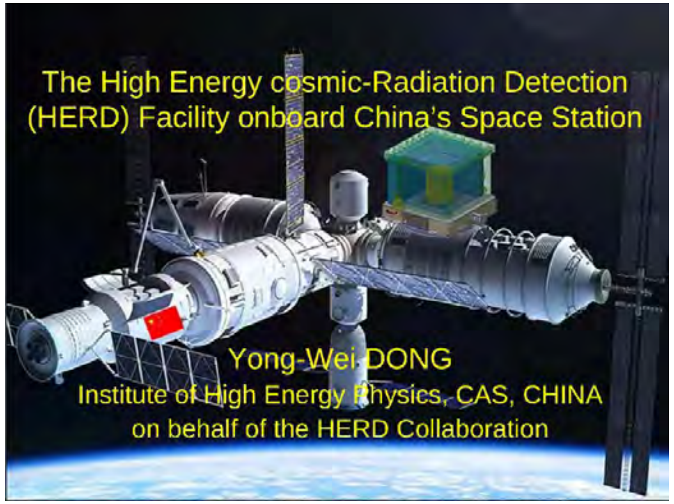
Rapidity distributions of π^- mesons in $^{40}Ar + {}^{45}Sc$ interactions (EPJ , C82 (2022)



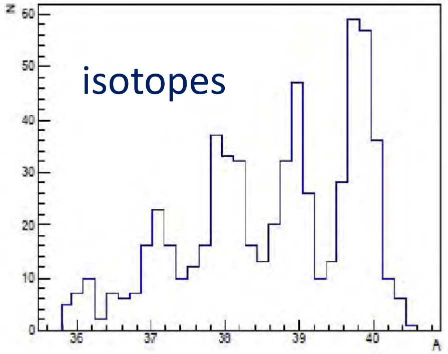
Exp. conclusion: "There is no model (EPOS, UrQMD, HIJING …) able to describe the data!" from NA61/SHINE Collab. on PP, $^{40}Ar + {}^{45}Sc$ and $^{7}Be + {}^{9}Be$

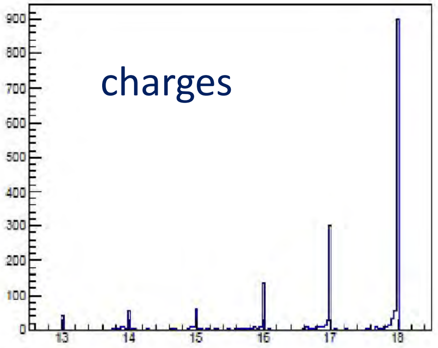
- Simulation and prototype testing for present and future orbital detectors: NUCLEON, NUCLEON-2, HERD





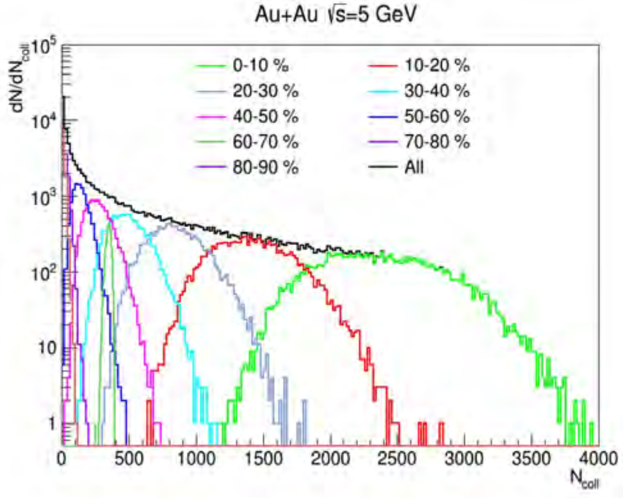
Prototypes tested @ Nuclotron JINR

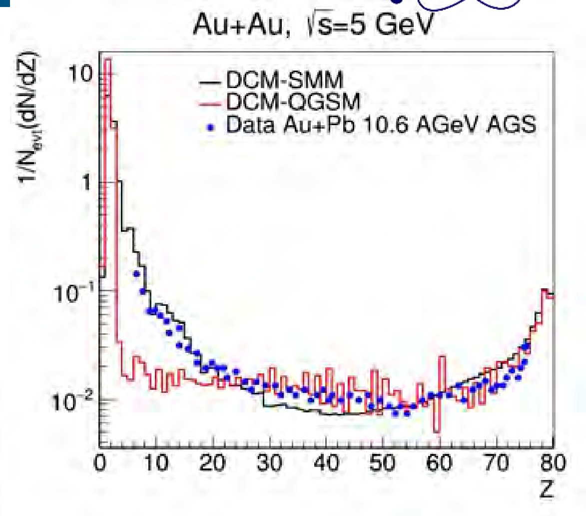




MC Generators for the NICA and LHC Experiments

- Development of the heavy ion collision generators
  - Dubna Cascade Model, Quark-Gluon-String Model, Statistical Multifragmentation Model for the NICA Experiments
  - tuning the HIJING generator with data of NA49 and NA61/SHINE @ CERN, STAR@RHIC (can be used in MPD and SPD experiments)
- Analytical and numerical methods for calculating neutron-proton systems under strong compression at the NICA



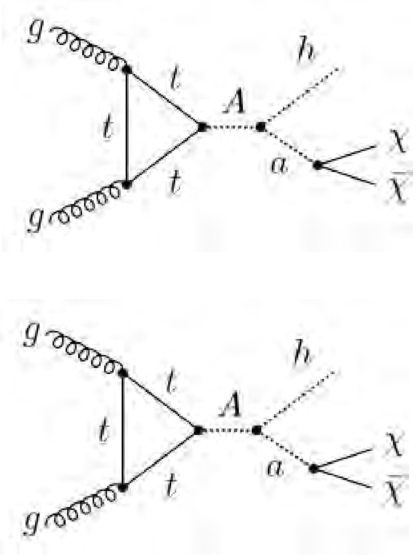


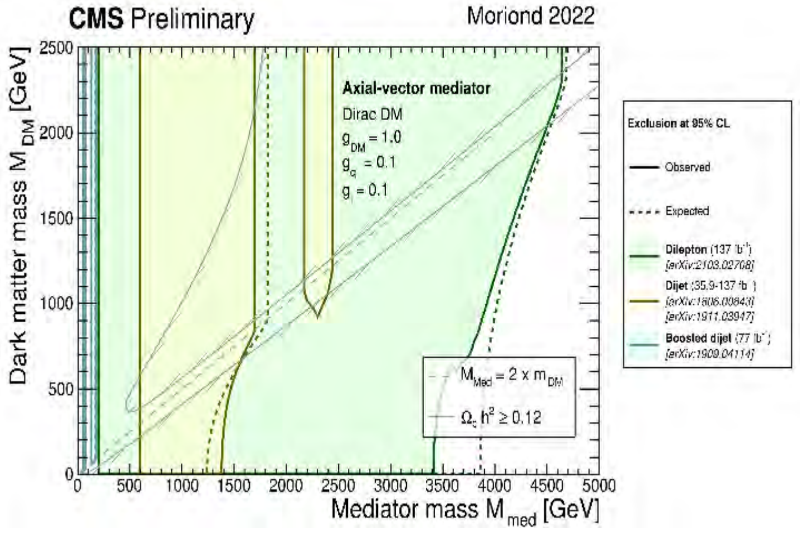
The priority tasks of the JINR Groups in LHC physics program (ATLAS and CMS) include searches for candidates for dark matter particles, tests of predictions of TeV-energy scenarios

- Fine tuning the generators for searches for new physics
  - revision of model parameters for 2HDM+a, 2HDM+s, etc.
  - simulation with Pythia8, QBH, MadGraph5_aMC@NLO + FeynRules (simplified DMM, HDM+a, 2HDM+s, etc.)
  - mass production + Geant4 response

Ex., Dark Matter can be probed with two fermions/two fermions + MET/higgs + MET/Z + MET in the final states

h ($\to$ bbar) + a ($\to \chi\chi$) = bbar + MET

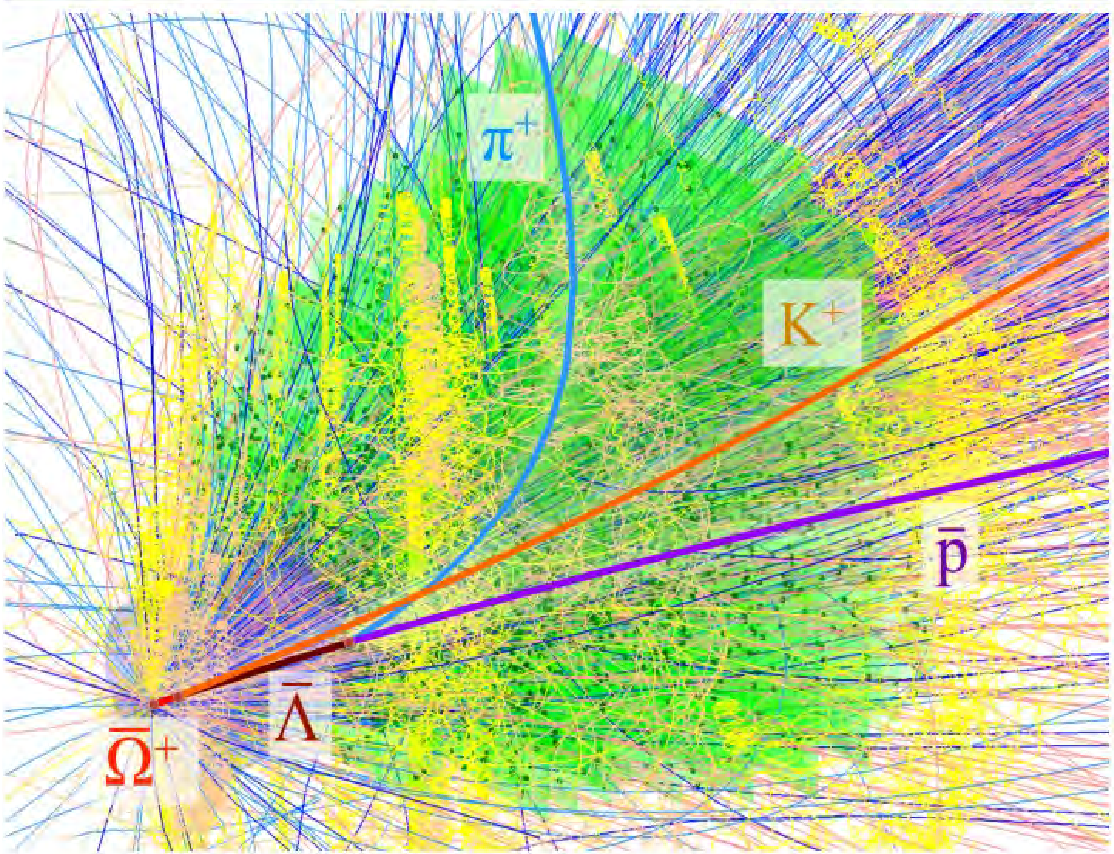




MLIT + BLTP + VBLHEP

Reconstruction of physical objects and analysis of experimental data

- development of algorithms, including those based on recurrent and convolutional neural networks for machine and deep learning tasks, and creation of corresponding software for the reconstruction of physical objects (tracks, particles, clusters, etc.) and physical processes;
- development of methods and algorithms for data analysis, including statistical analysis;
- adaptation of existing software for specific experiments, reconstruction and analysis of experimental data;
- analysis of Open Data of experiments, in particular, experiments at the LHC;
- conducting a global analysis of data from various experiments (in particular, a combined analysis of data from accelerator and astrophysical experiments in search for candidates for the role of dark matter).



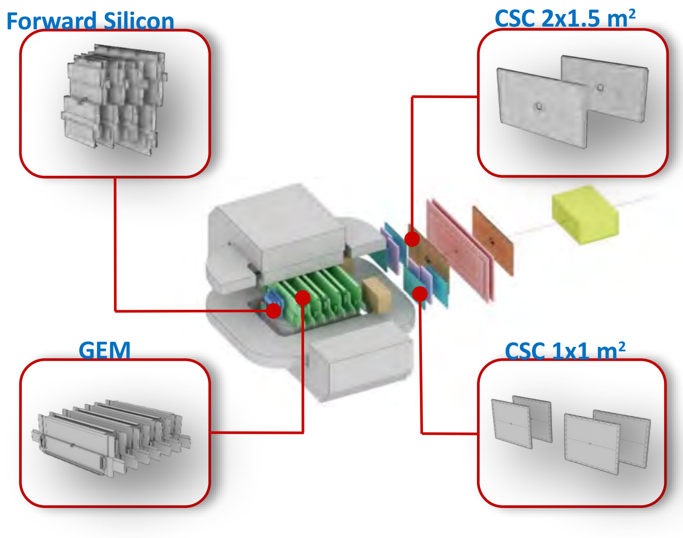
Моделирование и реконструкция в трекинговых детекторах эксперимента BM@N

Разработка, внедрение и развитие методов и алгоритмов с их последующей реализацией в виде комплекса проблемно-ориентированных программ для моделирования и реконструкции физических событий, а также обработки и анализа экспериментальных данных для координатных детекторов трековой системы эксперимента BM@N

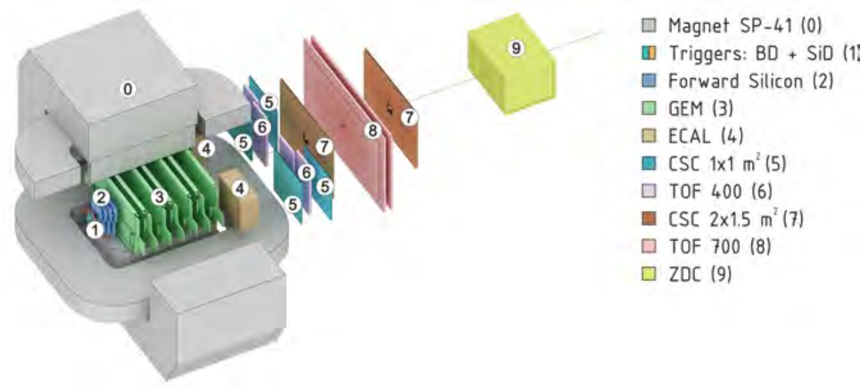
BM@N tracking system

Beam tracker		Inner tracker		Outer tracker	
Silicon Beam Tracker	Silicon Profilometers	Forward Silicon	GEM	CSC 1x1 m2	CSC 2x1.5 m2 (or DCH)

Иерархическая схема координатных детекторов трековой системы эксперимента BM@N в контексте последнего успешно проведенного восьмого сеанса в конце 2022 – начале 2023 годах

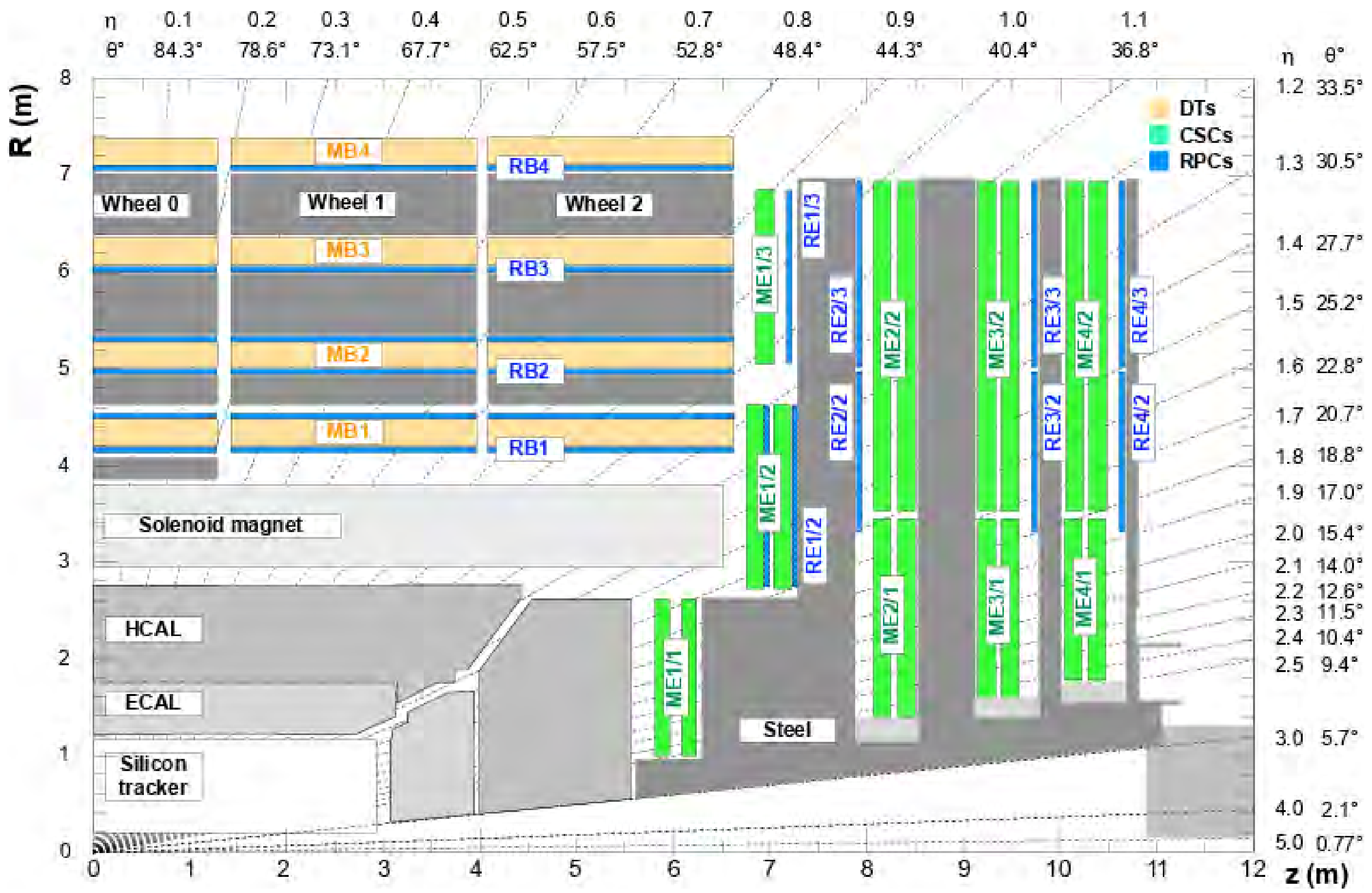


Координатные детекторы внутренней и внешней подсистем трековой системы эксперимента BM@N в контексте последнего проведенного восьмого сеанса в конце 2022 – начале 2023 годах

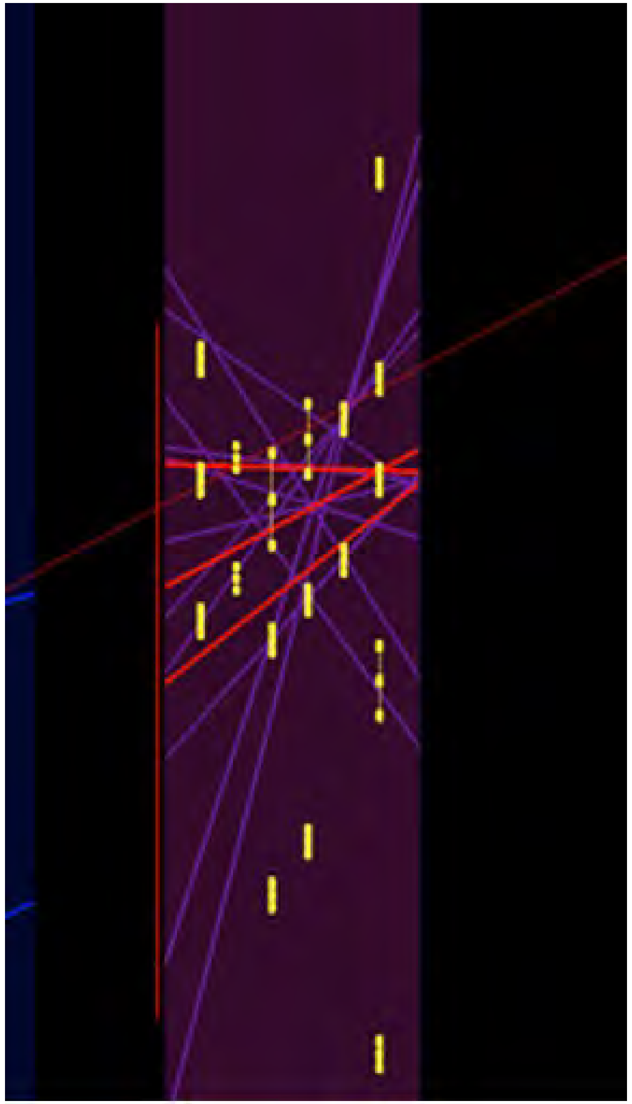


Конфигурация детекторов для RUN-8 BM@N эксперимента

CMS Endcap Muon System, Cathode strip chambers



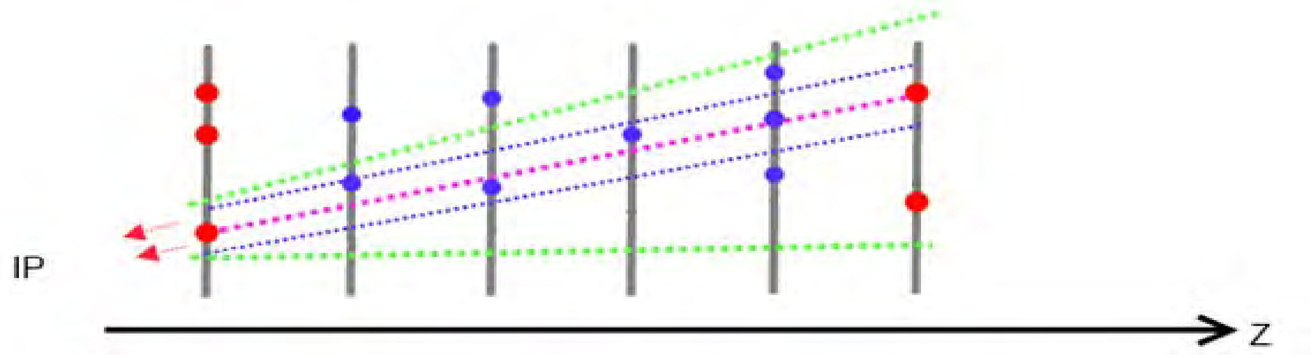
CSC placement in the CMS experimental setup



High pT muon passing through a CSC chamber

RU(RoadUsage) muon reconstruction algorithm

- Taking into account the IP while choosing the base hits in terms of coordinates for the future segment;
- Base Road for RecHit association:
  - A straight line is traced through the base hits
  - In the road formed along this line new hits from the inner layers are added to the segment

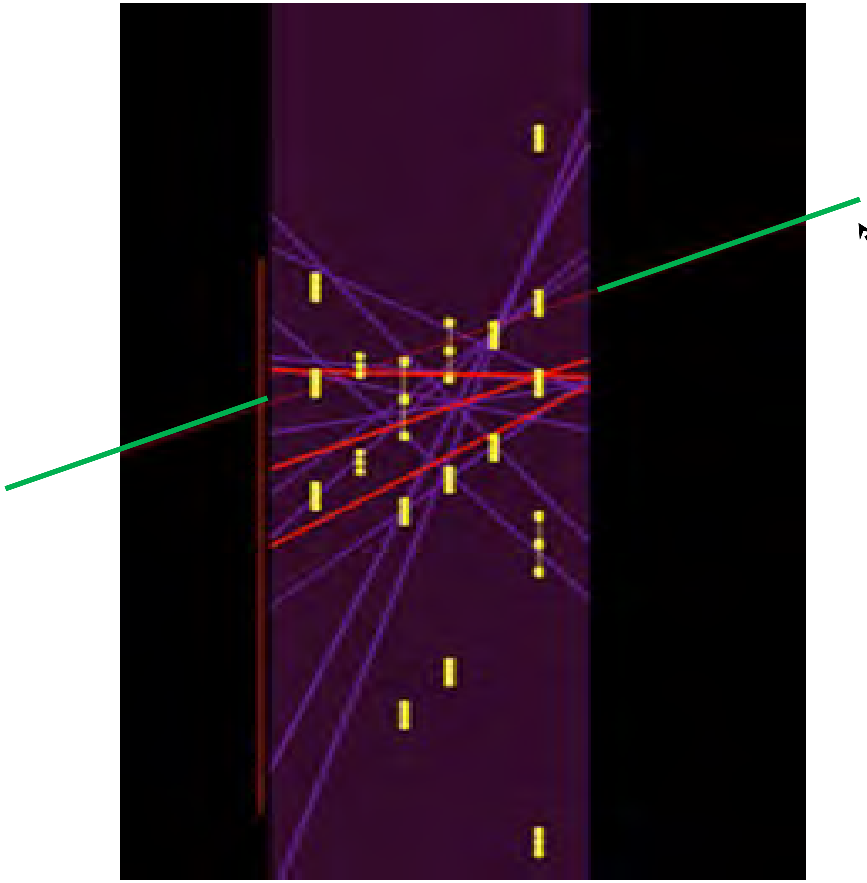


- Phi & tuned thresholds for each CSC station;
- Two iterations allowed:
  - First iteration is run with nominal thresholds;
  - Second iteration (if needed) is run with enlarged thresholds.
- If there are enough unused RecHits left the segment reconstruction is run again with the IP check turned off in order to reconstruct segments that correspond to displaced muons.

Implemented into the official CMS software package in July,2016

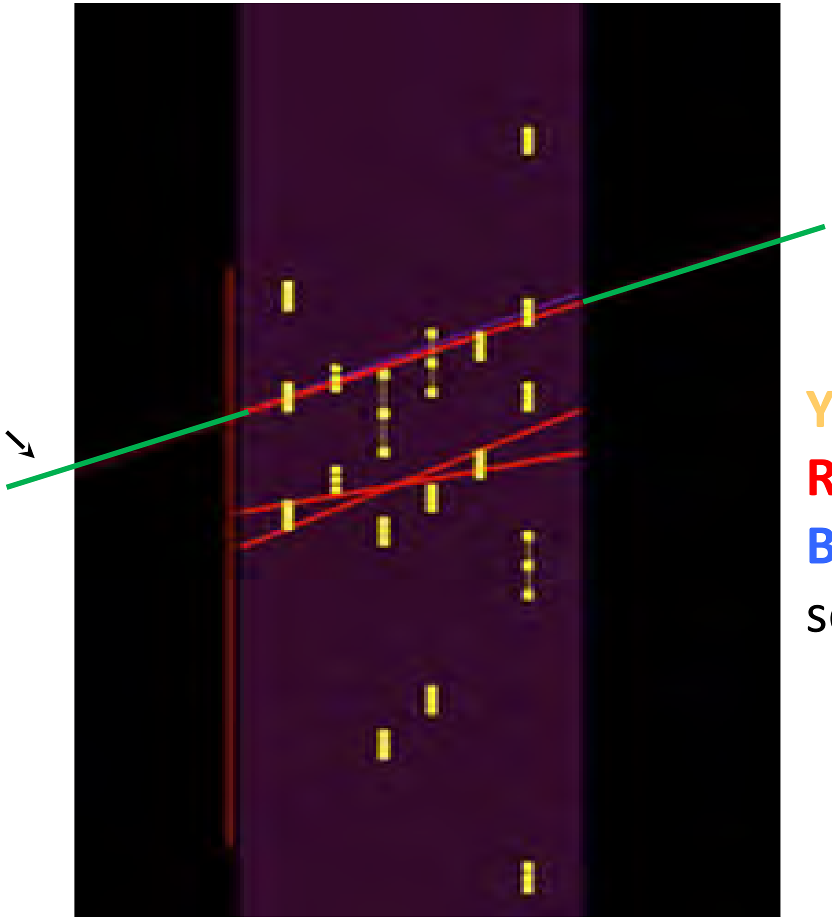
Starting 2017 it became the DEFAULT algorithm for reconstruction in CSC

High multiplicity event example - 84 RecHits in Chamber



Old
15 segments

Muon
trajectory



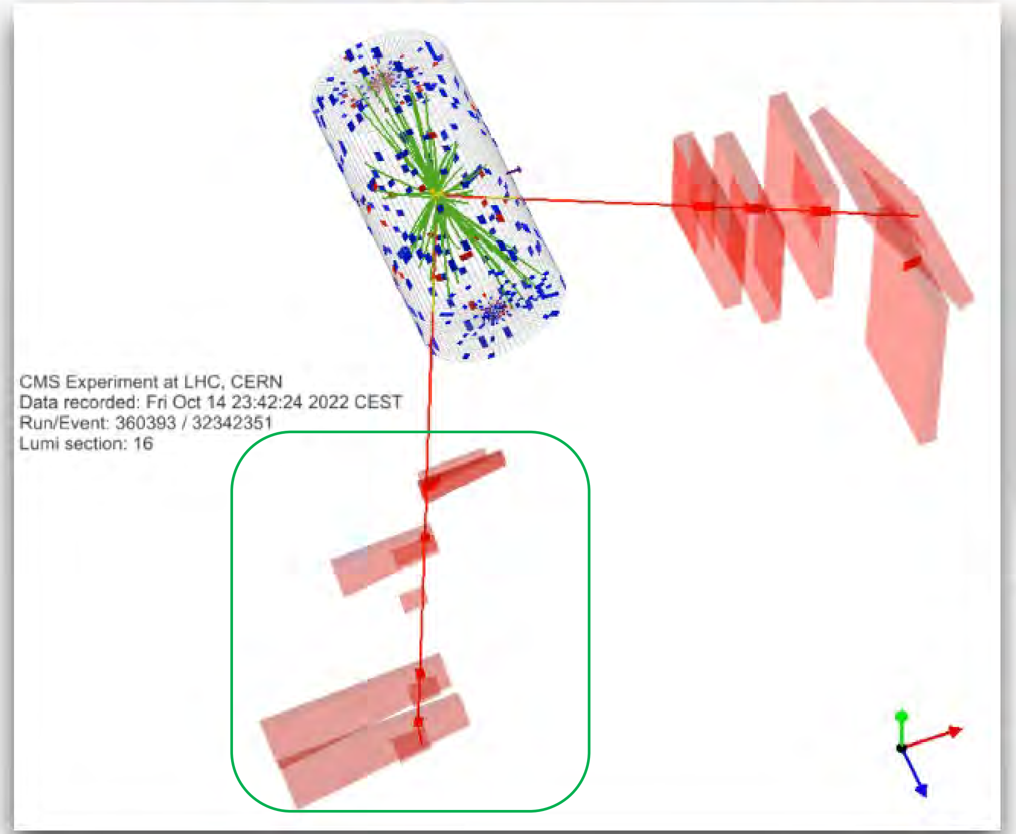
RU
4 segments

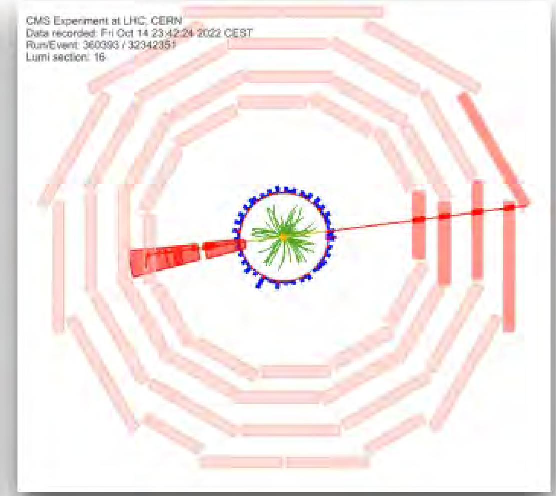
Muon p_T = 1000 GeV

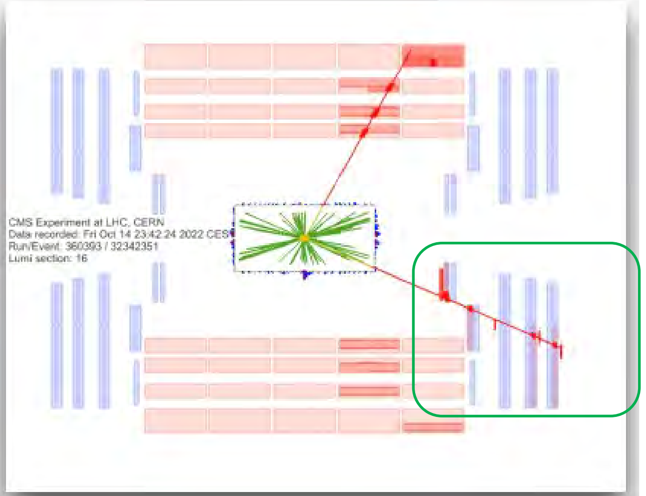
Yellow points – hits;
Red lines – muon segments;
Blue lines – secondary segments.

Highest di-muon mass event at CMS

Z′: Highest Mass Events







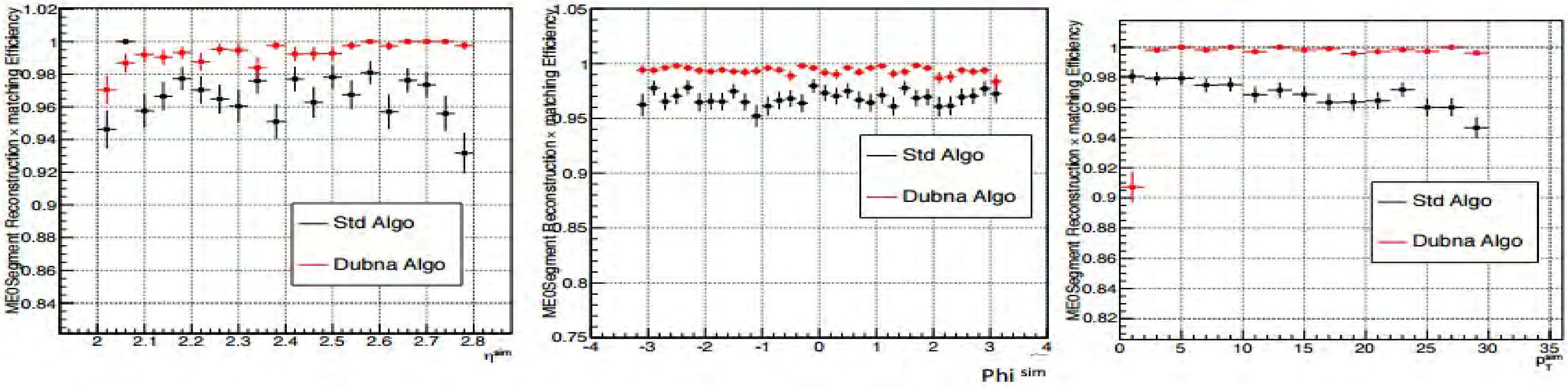
- Event ID: 360393:16:32342351
  - Dimuon invariant mass:
    - **TuneP + common vertex fit:** **<u>2407 ± 157 GeV</u>**
    - TuneP: 2378 GeV
    - Tracker track: 1868 GeV
    - Global track: 2530 GeV
  - mu1 (pT [GeV], η, φ) = (1161, 1.589, -3.014) ← CSC
  - mu2 (pT [GeV], η, φ) = (935, 0.538, 0.131)

A. Lanyov

Event displays for 2nd and 3rd highest mass events in backup

Usage of RU algorithm for tracking in GEMs @ CMS

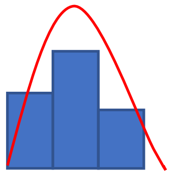
Example: Std vs Dubna Efficiency



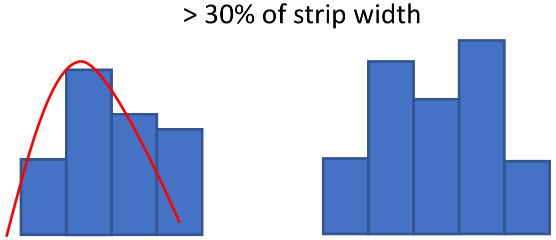
- ✓ Single Mu sample with PU=0, without noise, CMSSW_8_1_0_pre16
- ✓ Flat pT in 0-30 GeV, Flat eta in ME0 acceptance
- ✓ Rechits are Smeared points with perfect spatial resolution (thus, no realistic readout yet)

- Overall, the performance of Dubna algo are the better, as expected

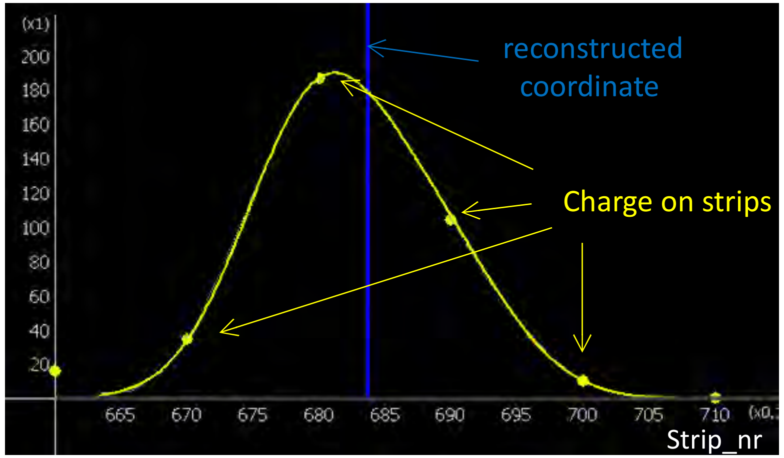
Overlapped signals delimitation in CSC of the CMS experiment



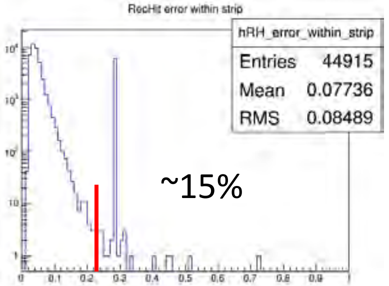
Regular strip charge distribution



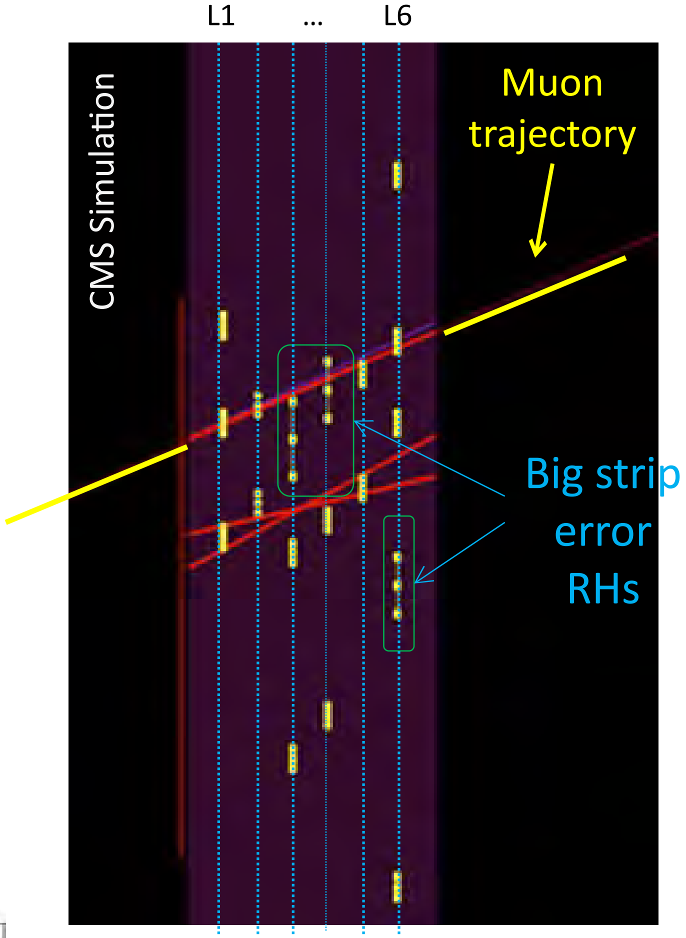
Abnormal strip charge distributions



If > =>
the coordinate is calculated using Center of Gravity (CoG) like algorithm



Hit Strip Errors



Event display in CSC
Hits with big strip errors fail to be add to the segment.

Introduction to wavelet analysis

The g2-WTS(Wavelet-TranSform method based on the usage of second degree wavelets ())[1] was chosen by us for overlapping signal recognition.

The main function in wavelet analysis is the following double gaussian function

(1)

While processing the signal it is divided in a set {}, where

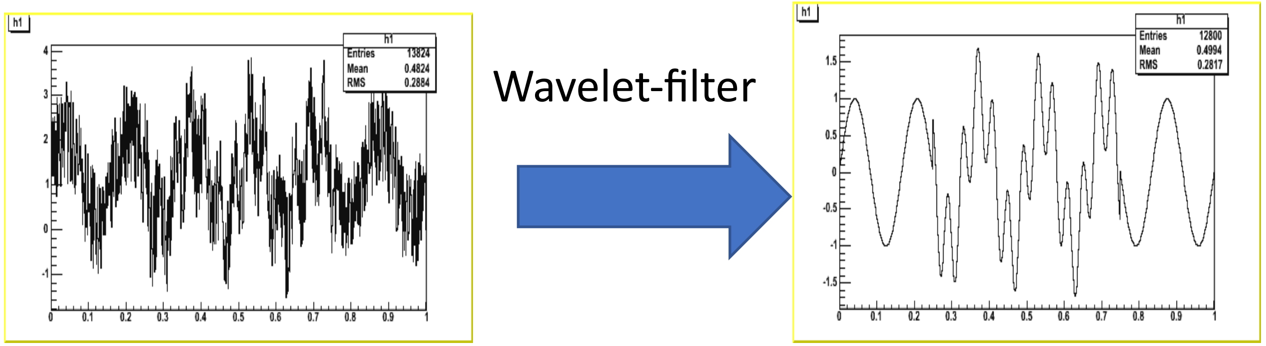
= (2)

Gauss wavelets W for the function G can be represented as follows:

(3)

where

.



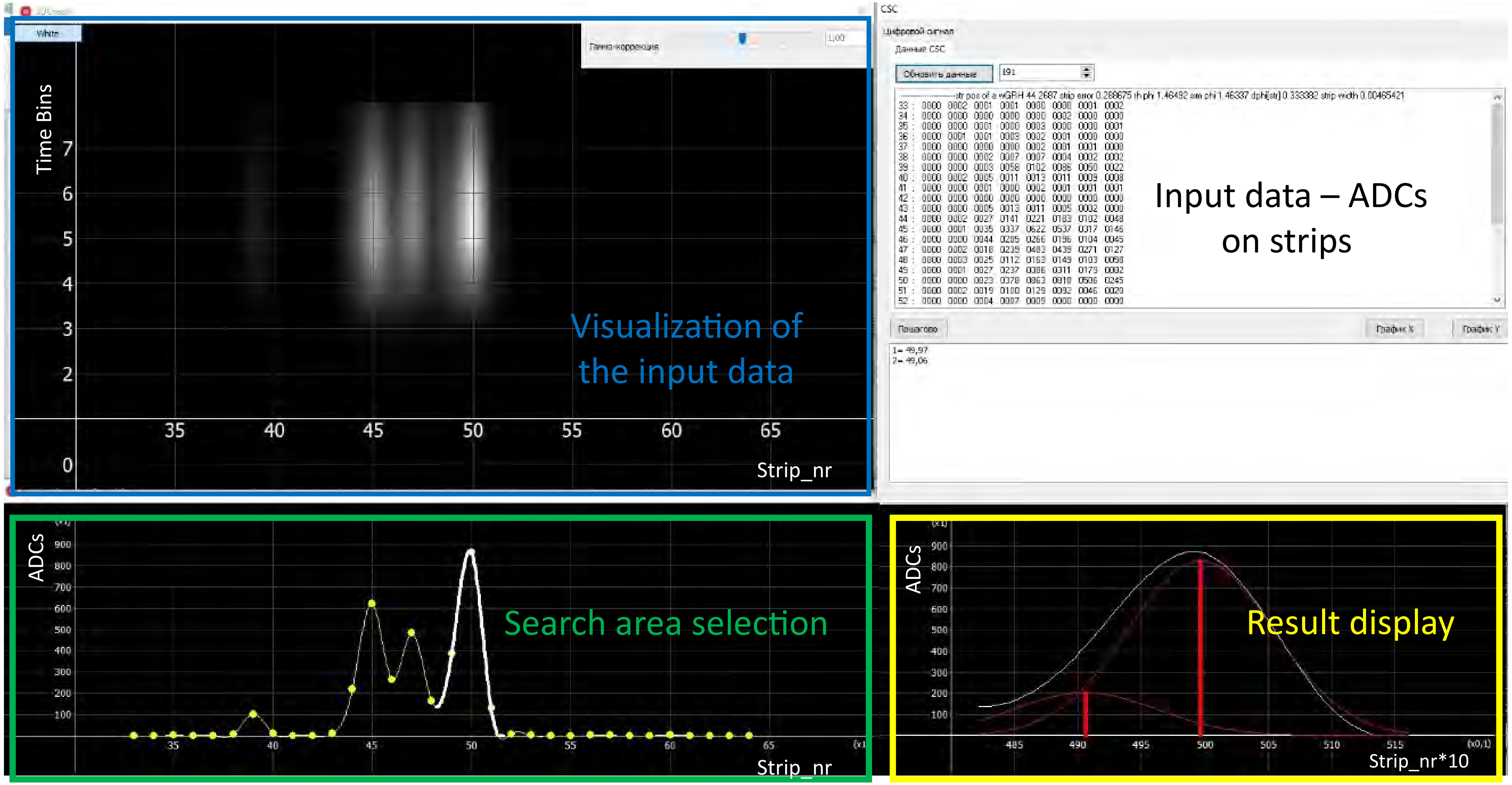
Here b is the shift, a is the scale and , – gaussian wavelet of the n-th degree, h- cluster of strip charges,- half-width of the wavelet.

To find the n-th wavelet coefficients of the set h we use:

(4)

G.Ososkov, A.Shitov Gaussian Wavelet Features and Their Applications for Analysis of Discretized Signals // Comp.Phys.Comm, v.126/1-2, 149-157, (2000)

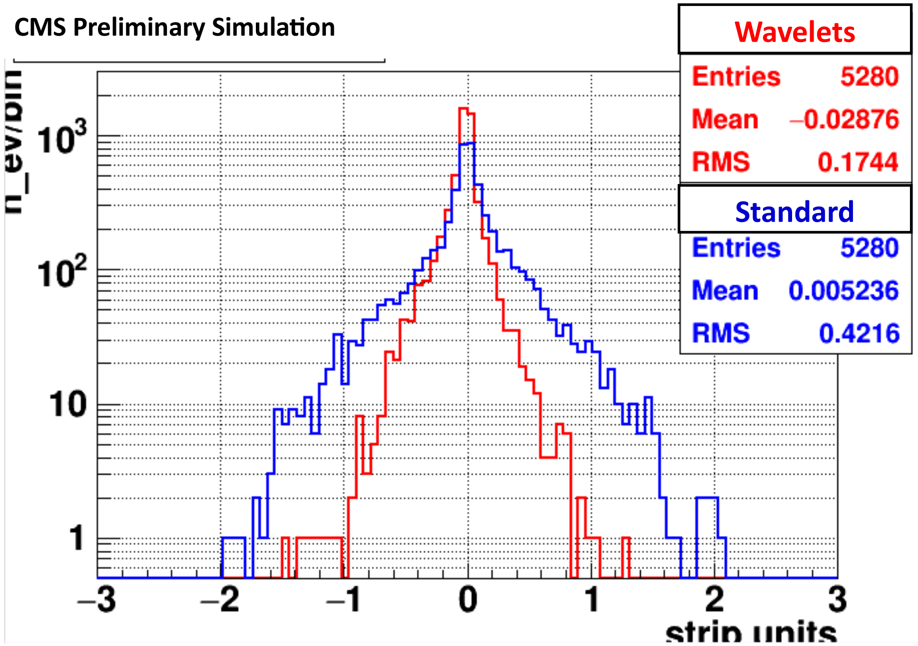
Universal tool for overlapped signals delimitation



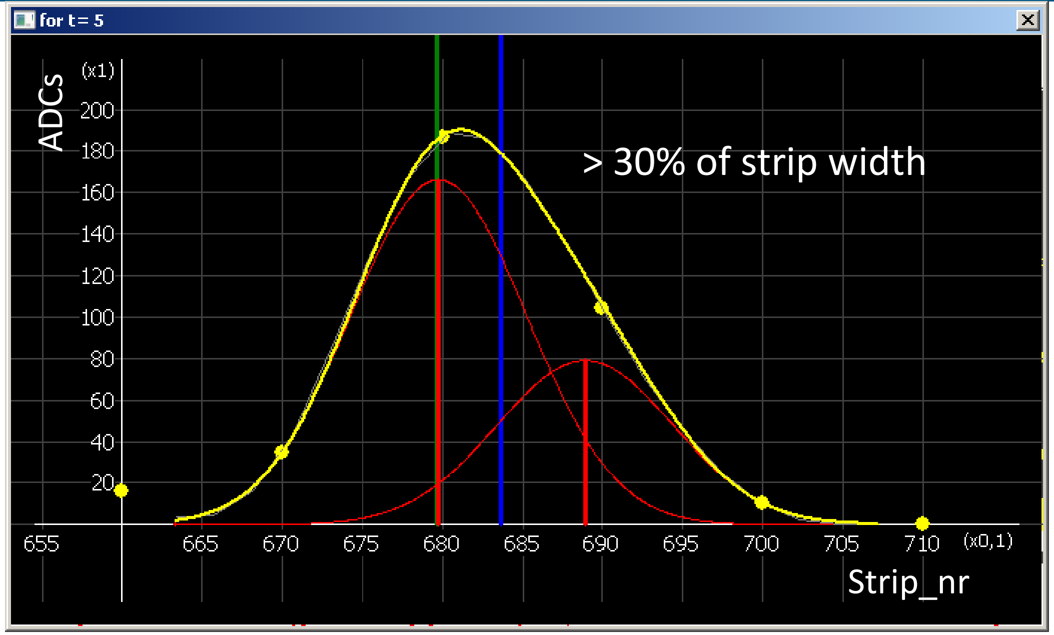
Universal tool for overlapping signal recognition

Overlapped signals – preliminary results

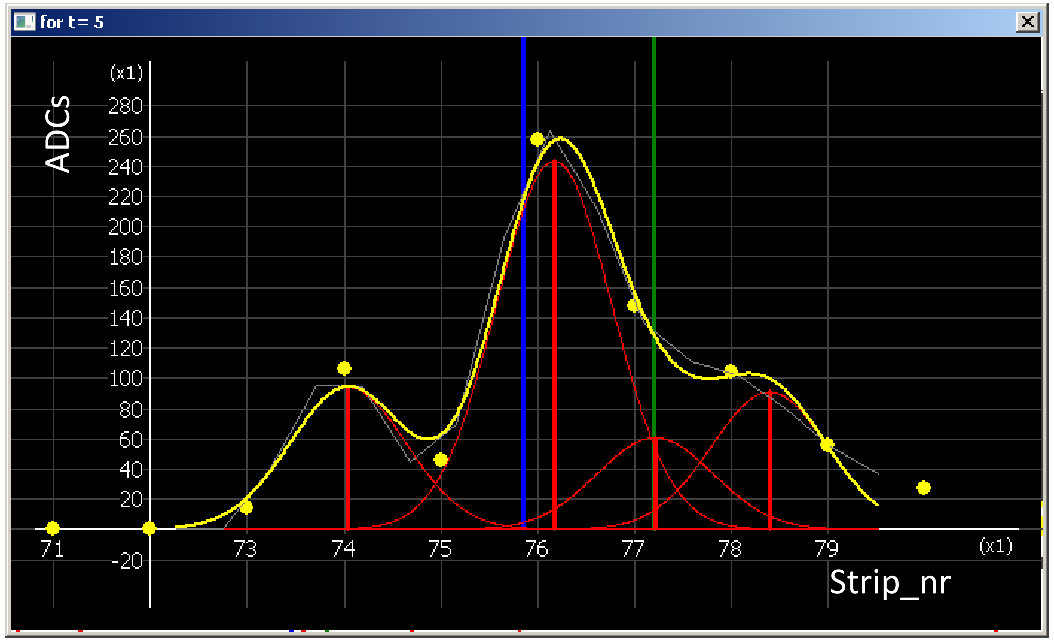
Yellow line – initial charge distribution;
Green line – simulated muon coordinate;
Red line – wavelet analysis;
Blue line – standard approach (CoG like).



Difference in terms of strip widths between simulated and reconstructed hit



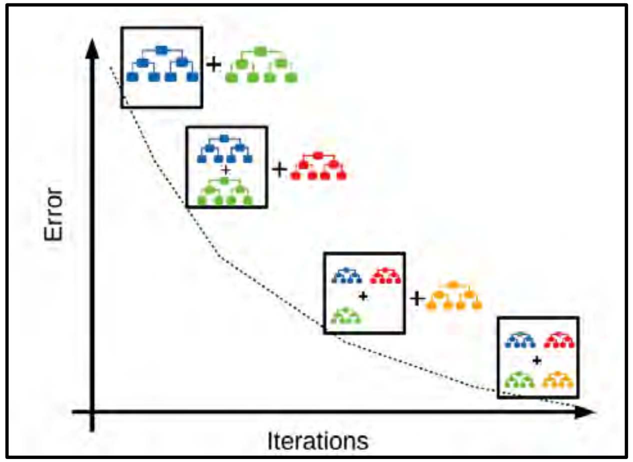
2 overlapped signals



4 overlapped signals

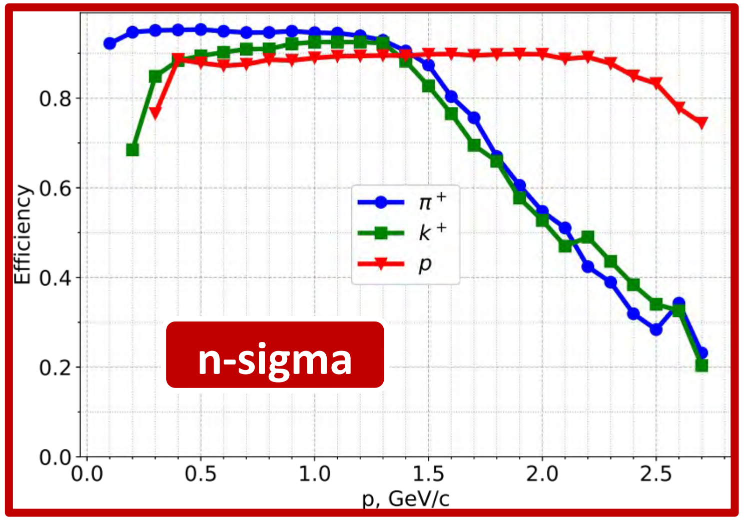
First approach of decision trees for PID in MPD experiment

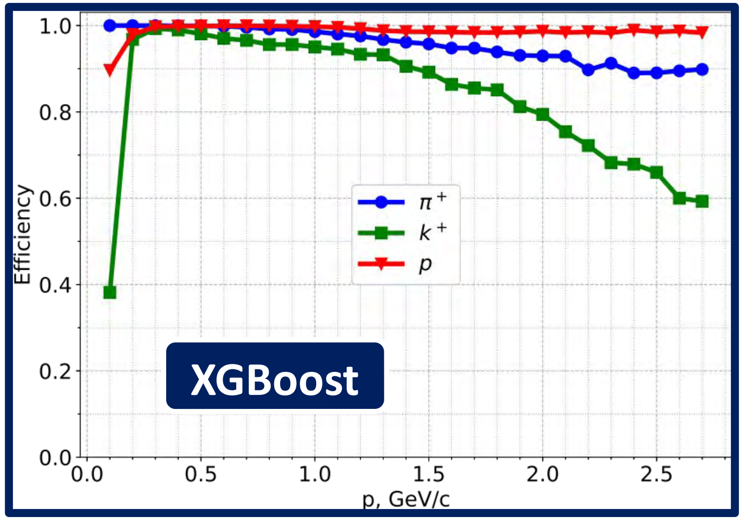
Particle identification is an important part of the physics analysis pipeline in MPD experiment and is provided by Time-Projection Chamber (TPC) and Time-of-Flight (TOF) detectors. We have applied gradient boosting approach for ensemble of decision trees with use of CatBoost library to PID on Monte-Carlo simulated data for TPC and TOF detectors.



Idea of the gradient boosting method

The proposed approach for PID problem in the MPD experiment shows promising results. In particular, it gives a multiple increase of an efficiency for small and large particle momentum ranges in comparison with so-called n-sigma method realized in mpdroot software.

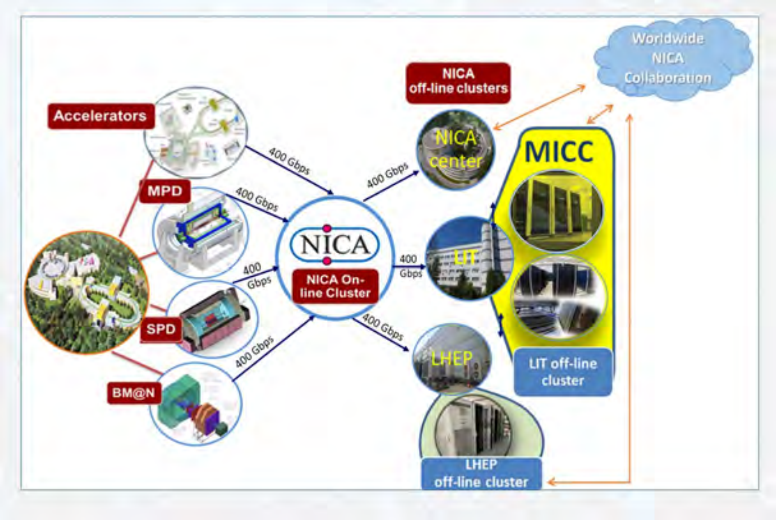


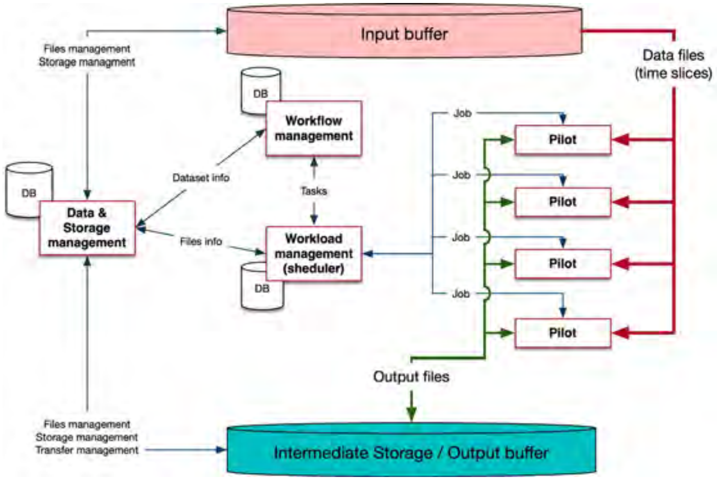


Efficiency of PID

Support and development of the software environment for JINR research

- Development of the software environment for processing and analyzing data from the NICA experiments
- Creation, implementation and development of an information and computing complex for processing, analyzing and storing data for the SPD experiment
- Creation of specialized databases and information systems for the Collaborations @ NICA and LHC (ATLAS, BM@N, MPD)

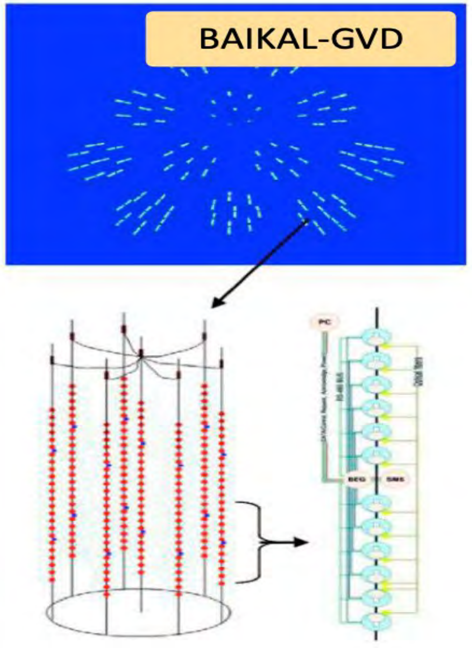




Развитие алгоритмов и системы обработки и анализа данных для экспериментов в области физики нейтрино

В рамках проекта производится развитие алгоритмов и системы обработки и анализа данных для реконструкции событий экспериментов в рамках реализации программы ОИЯИ в области физики нейтрино

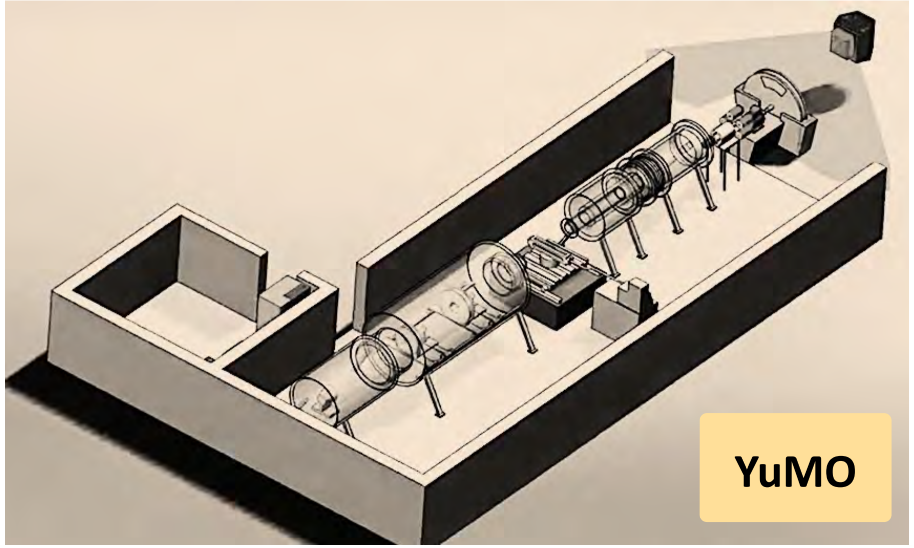
- Байкальский глубоководный нейтринный телескоп (BAIKAL-GVD) предназначенный для исследований потока космических нейтрино высоких энергий и поиск их источников для изучения темной материи, распадов сверхтяжелых частиц и др.
  - Отладка на большом объеме экспериментальных данных системы обработки данных и достижение её работоспособности в качестве основной системы в проекте BAIKAL-GVD
- Наземная гамма-обсерватория TAIGA предназначены для исследования космических и гамма-лучей в пределах пяти порядков по энергии и станет одним из основных детекторов в диапазоне энергий от ТэВ до сотен ПэВ с целью изучения галактических источников (певатронов).
  - Развитие программного обеспечения для реконструкции космических лучей и высокоэнергетических гамма-квантов на данных с детекторов TAIGA-HiSCORE и TAIGA-IACT



Исследовательские проекты ОИЯИ в области нейтронной физики

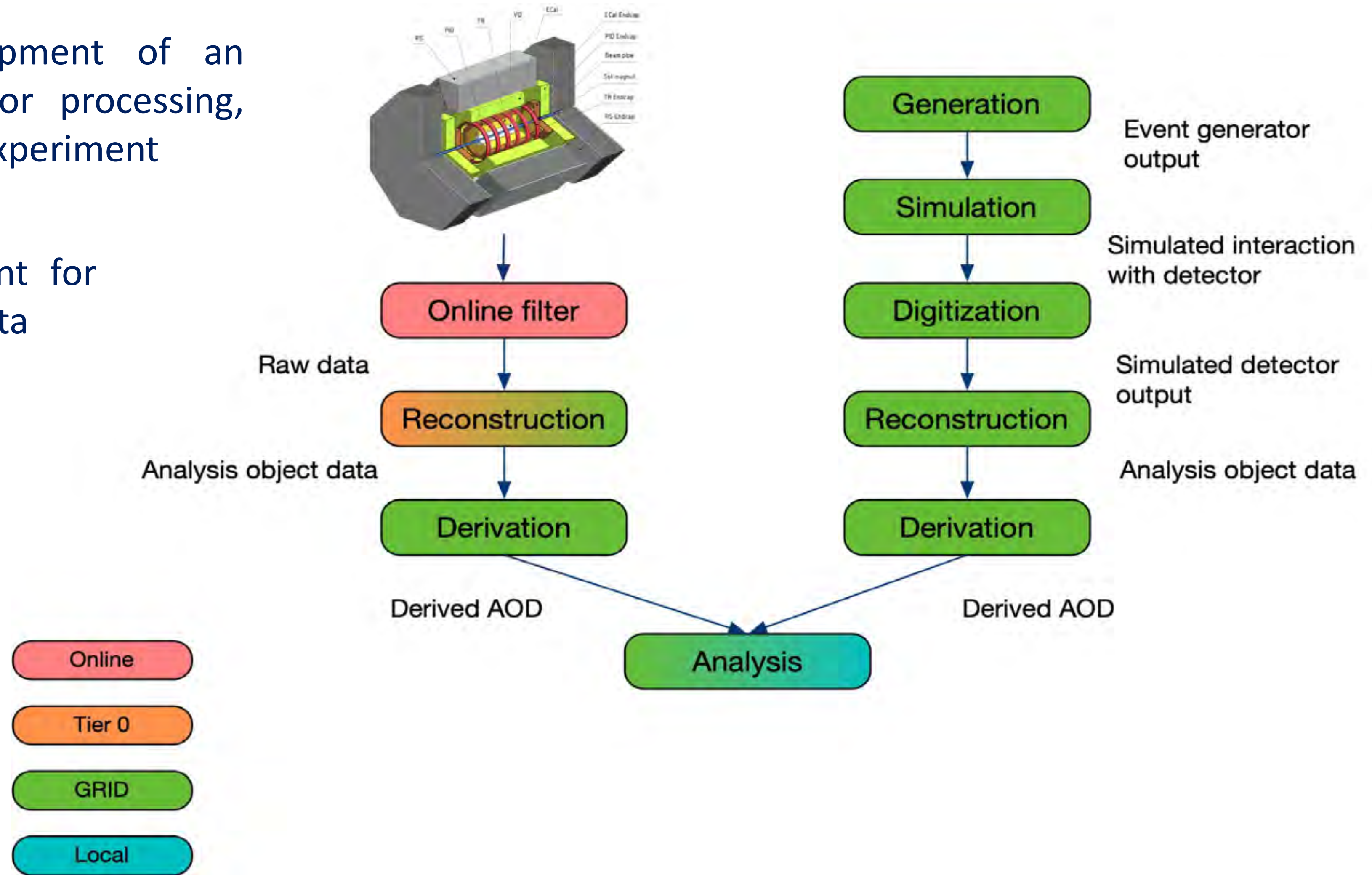
Ключевым элементом в применении метода малоуглового рассеяния нейтронов является обработка экспериментальных данных. Для решения целого спектра задач спектрометром ЮМО принципиальным и определяющим является пакета программ для быстрой и эффективной обработки экспериментальных данных. Целью предлагаемого проекта является развитие алгоритмического обеспечения и программного комплекса обработки данных спектрометра малоуглового рассеяния нейтронов ЮМО на реакторе ИБР-2, которая является крайне востребованной и актуальной.

В рамках проекта реализуется развитие пакета программ для первичной обработки малоугловых экспериментальных данных спектрометра ЮМО для многодетекторной системы с позиционно-чувствительным детектором с распределенными возможностями сочетания типов обработки, включая нормировки на потоки, адаптации к возможной смене частоты импульсов реактора ИБР-2, методов учета фоновых условий и адаптации к изменениям многодетекторной системы спектрометра ЮМО.



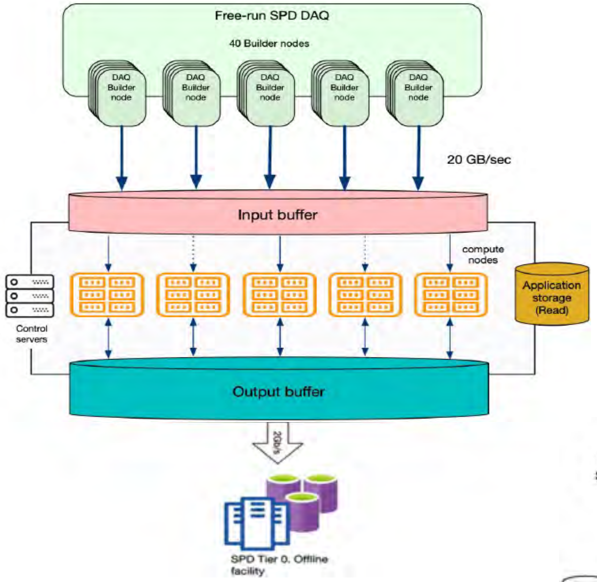
Computing for SPD Experiment

- Creation, implementation and development of an information and computing complex for processing, analysis and storage of data for the SPD experiment
- Development of a software environment for processing and analyzing experimental data



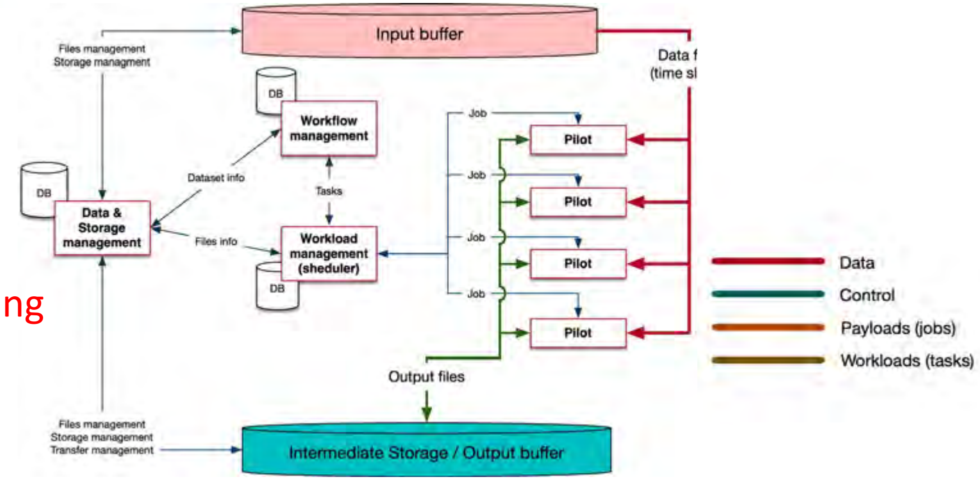
Computing for SPD Experiment: OnLine Filter and Off-line System

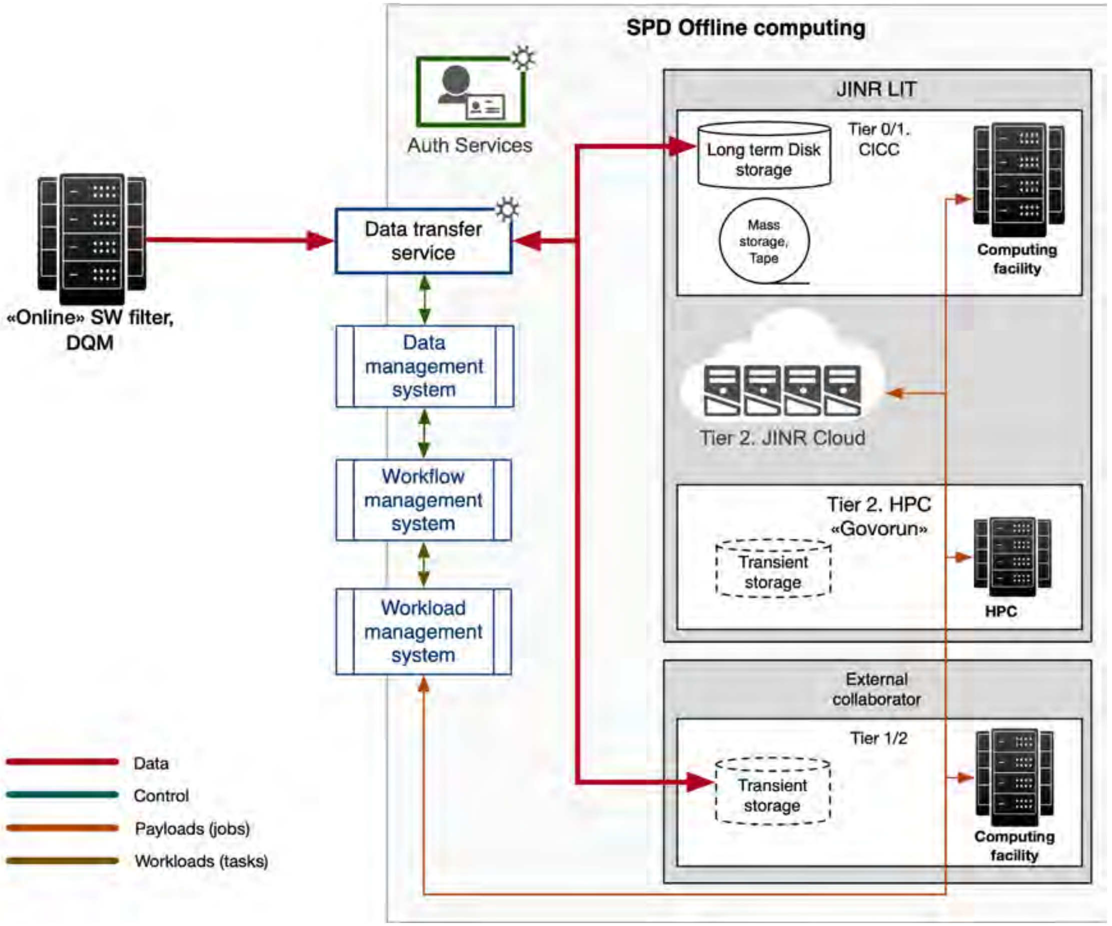
- design and creation of a software and hardware complex for organizing the high-throughput processing of data received from the SPD DAQ (SPD OnLine Filter)
- development of a model for processing and organizing data including distributed data processing of the SPD experiment on the resources of the collaboration participants.



✓ data flow ~3M ev./s. (~20 GB/s)
✓ no trigger

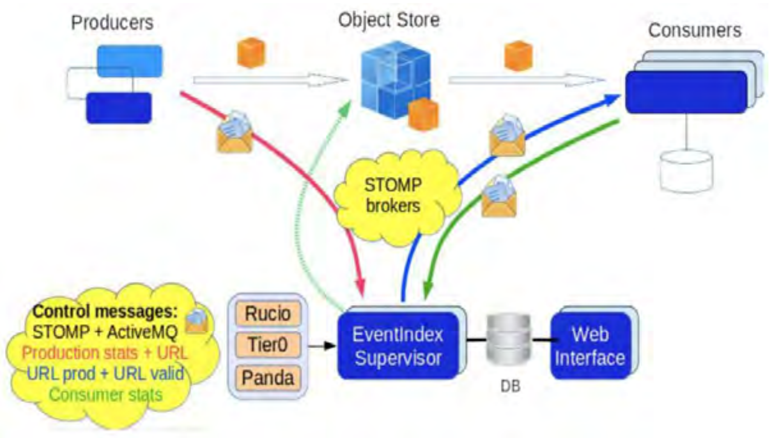
- dedicated computing farm
- middleware complex
- methods and approaches to filtering data in real time using artificial intelligence technologies;



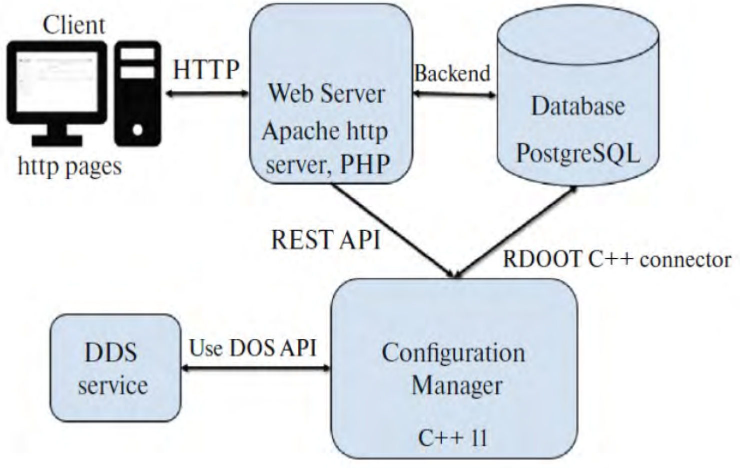


Data Bases for HEP Computing

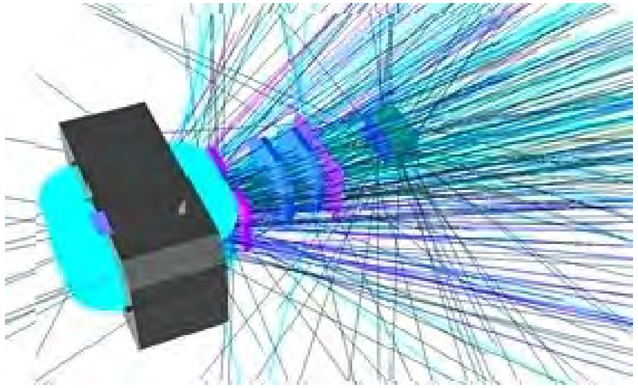
- ATLAS EventPickingService
  - the first version for automated event collection has been created.
  - the service was used for the second stage oa the «γγ→WW» analysis (136 K events)
  - further modernization of the service is ongoing according to the received results
- ATLAS CREST (Condition DB)
  - C++ API for CREST (implemented into Athena software package)
  - COOL2CREST converter (developed, but not yet implemented)
  - both parts require constant improvement for compatibility with the updated CREST server



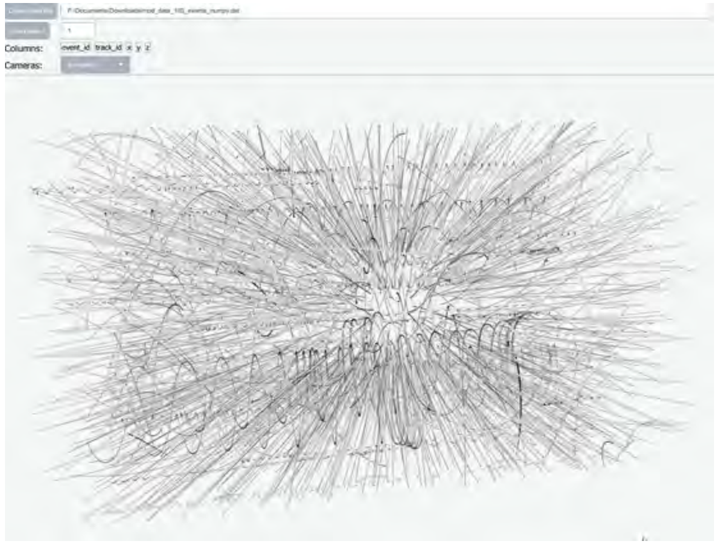
- Event metadata system for the experiments at NICA (BM@N, MPD and SPD)
  - has been designed to index events of the NICA experiments and store their metadata
  - quick search by required conditions and parameters used in various physics analyses for a set of physics events to use in further event data processing
- Geometry DB
  - The geometry database is the main element of the information system designed to store, process and manage information about the geometric models of detectors. It is implemented into BMNROOT software
- Configuration DB (Configuration Information System, CIS)
  - Configuration database for managing online applications for collecting and processing high energy physics events. It was created for process launching in BM@N environment.



ML методы для современных экспериментов ФВЭ



Эксперимент BM@N.



Трековый детектор TPC внутри магнита MPD. Иллюстрация смоделированного события столкновения пучков ионов золота, создающее тысячи треков.

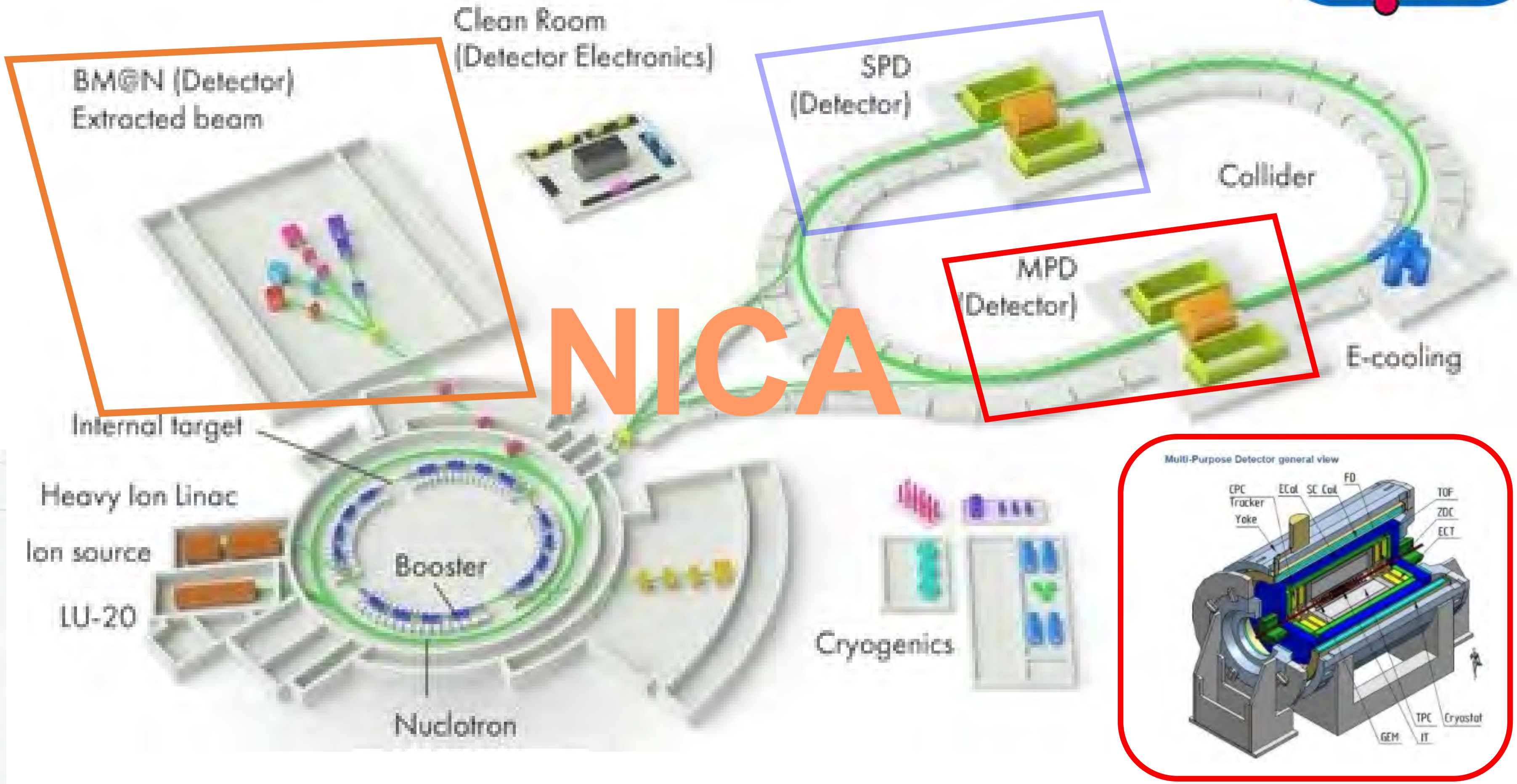


Схема ускорительного комплекса NICA с экспериментами MPD, SPD, BM@N

Задачи: реконструкция событий по данным измерений в трековых и других детекторах.

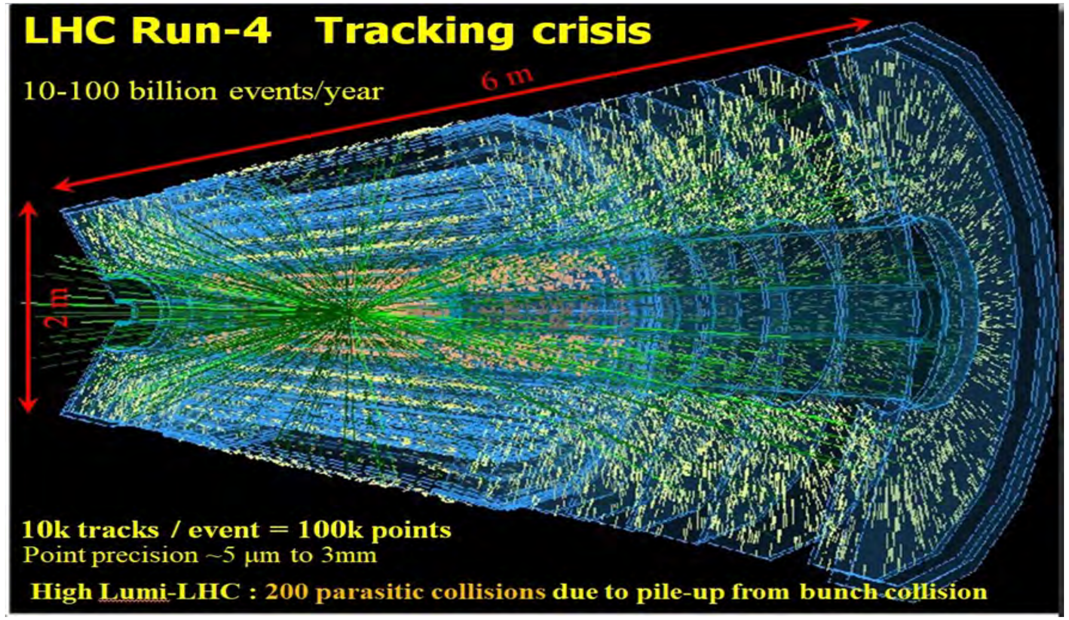
Problem Statement: The Need for Advanced Tracking Methods

Unprecedent scale of modern experiments:

- Up to 200 simultaneous proton-proton interactions is expected at High Luminosity Large Hadron Collider
- 200 particle tracks on average, 40K of tracks considering pile-up
- Traditional tracking methods cannot handle dense, overlapping particle tracks due to computational complexity and time constraints.

Deep Learning for Efficient Track Reconstruction:

- DL models can handle high-dimensional data and complex spatial correlations between tracks
- Multiple scattering and inhomogeneous magnetic field effects could be learned from training data
- Effective parallelization using GPUs out of the box
- TrackML Challenge was launched to explore new scalable approaches for particles tracking

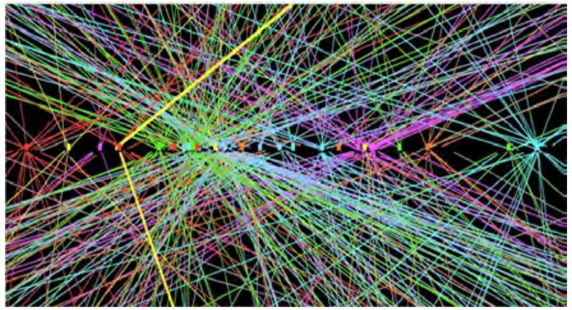


https://webific.ific.uv.es/web/en/content/taking-lhc-higher-luminosity

The same problems are expected also for the NICA megaproject experiments

Under high luminosity conditions, particles are accelerated not individually, but in groups – bunches

~15 cm



To cope with immense data volumes
a new high-throughput deep-learning based approach for tracking is needed

TrackML challenge 2018

In 2018, physicists from CERN and other physics centers around the world, including Russia, staged a competition - the TrackML challenge to solve a machine learning problem for particle tracking in high-energy physics at high luminosity

(DOI 10.1109/eScience.2018.00088)

For this purpose, a source code simulator program is made on the Kaggle platform, where a typical all-pixel LHC tracking detector of 10 layers generates physical events (Pythia ttbar) superimposed on 200 additional collisions. This yields typically 10000 tracks (100000 hits) in each event.

TrackML dataset

https://www.kaggle.com/c/trackml-particle-identification/overview

Detailed view of the short strip detector of the TrackML challenge with a simulated event with 200 pile-up interactions

Noticeable participants:

- 1st: top-quarks – Logistic regression for pairs and triplets, helix extrapolation (**8 min/event**).
- 2nd: outrunner – Dense NN for pair prediction, circle fitting (**3+ hrs/event**).
- 3rd: Sergey Gorbunov – Triplet seeds, helix fit with magnetic field estimation (**0.56 sec/event**).
- 9th: CPMP – DBSCAN clustering, filtered by module frequency (**10 hrs/event**, 30,000+ DBSCAN runs).
- 12th: Finnies – DBSCAN seeding, LSTM for predicting next 5 hits (**slow, no speed given**).

Most of the solutions repeat the classical pipeline for tracking – seeding followed by trajectory fitting.

TrackML challenge results

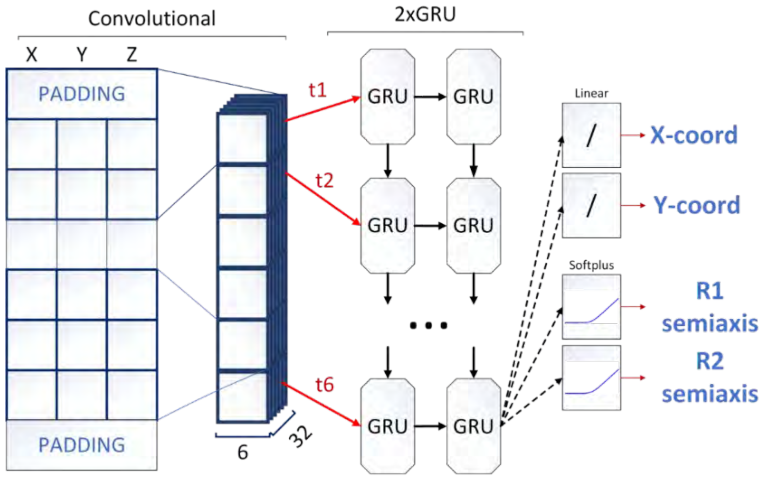
The TrackML competition has stimulated a lot of research where TrackML dataset has been used to train and verify different tracking neuromodels

- **Lots of graph neural network programs, e.g. https://arxiv.org/pdf/2003.11603**
- **There is also some interest in the application of Hopfield neural networks, but in a very different aspect, the slow evolution of the network is proposed to be dramatically accelerated by quantum annealing performed on a quantum D-Wave computer https://doi.org/10.1007/s42484-021-00054-w.**
- **Moreover, it is also proposed to apply quantum annealing to accelerate graph neural networks arXiv:2109.12636v1 [quant-ph] 26 Sep 2021**
- **These works have in many ways stimulated new and quite promising research on deep tracking carried out since 2018 at JINR MLIT for experiments of the NICA and BES-III projects**

Reports on real tracking tasks using LHC Run 2 and 3 data have already appeared outside the **TrackML challenge** (see, e.g., arXiv:2308.09471v1 [hep-ex] 18 Aug 2023)

Our achievements before the announcement of the TrackML challenge

1. Local tracking for the GEM detector of the BM@N experiment is particularly challenging due to the presence of a p number of fake hits, making it extremely difficult to find those hits at subsequent detector stations that are extensions of the processed track.



Scheme of the recurrent TrackNETv2 NN

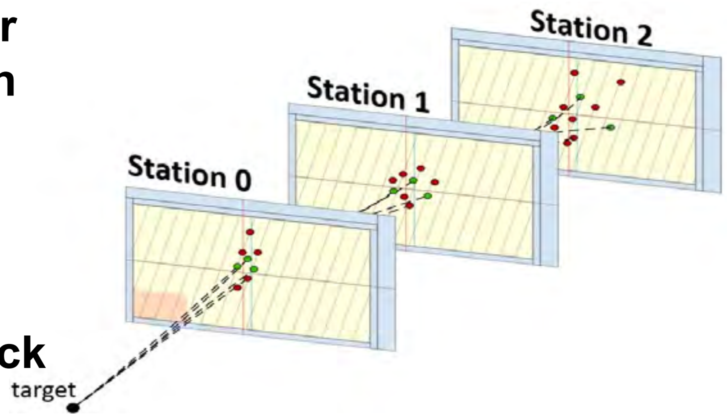
We used the flexibility of the GRU recurrent neural network architecture, which allowed us to overcome these difficulties and create a new end-to-end TrackNET neural network with a regression part of four neurons. Two neurons predict the ellipse center point on the next coordinate plane where the continuation of the candidate track should be searched for, and two more determine the semi-major axis of this ellipse. (See https://doi.org/10.1063/1.5130102)

Used Metrics:

Recall = ; Precision =

- - No. of correctly reconstructed true tracks.
- – No. of true tracks.
- – Total number of reconstructed tracks.

This gives us the opportunity to train our model using only the true tracks that can be extracted from the Monte Carlo simulation. Thus, we have obtained a neural network that performs track following similar to the Kalman filter, although without the final part where track fitting is performed

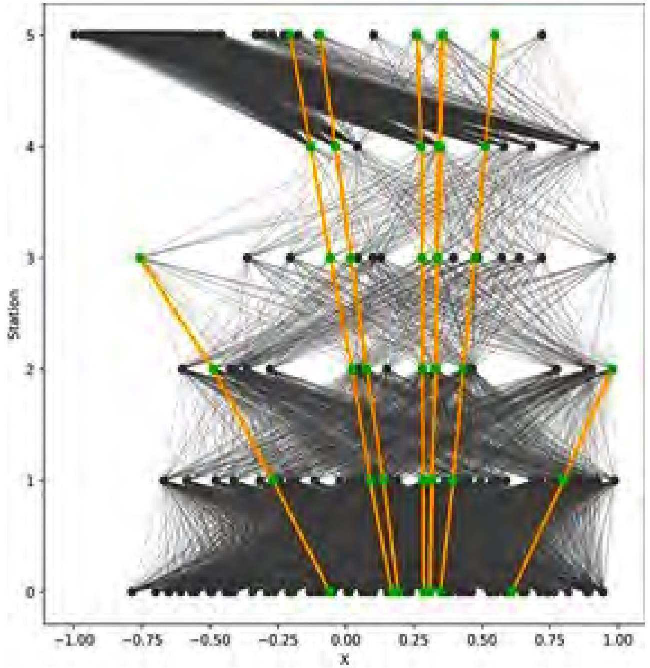


However, the beforementioned shortcomings of local tracking meant that applying TrackNET to the simulation data of BM@N run 7 gave a good recall of 97% but a precision >50%, which is unacceptable. The situation was later saved by a hybrid tracking approach using a graph neural network in the second run

Tracking for data from high luminosity experiments. GNN for BM@N

Application of graph neural networks (GNNs).

Consider an event as a graph in which the nodes are hits. Nodes between neighboring stations can be connected by edges, which are possible track segments. Nodes are not connected within the same detector layer. The tracking task for graph neural networks (GNNs) can be formulated as a graph edge classification problem - to determine which of the segments belong to real tracks and which ones should be discarded as false.



Graphical representation of the C + C, 4 GeV event of the BM@N experiment. Black nodes and edges correspond to the fakes, green nodes and yellow edges to the found tracks

This scheme is similar to the well-known global Denby-Peterson approach with a segmented Hopfield neural network, where the neural network takes a long time to self-train separately for each event. However GNNs, where we need to find edges that are segments of real tracks, can be trained on a sample of event graphs, where these edges are labeled with a binary vector indicating whether a particular edge is true (1) or not (0). This approach has been successfully implemented at CERN for model events from the pixel detector, but our attempts to adapt their GNN for BM@N events with a huge fake background failed due to the resulting memory space issues for loading the graph.

These problems disappeared when GNN was applied to the TrackNET output data in the second stage of tracking. By receiving as input an event represented as a graph of all candidate tracks generated in the first stage, GNN produced an acceptable tracking performance as a result

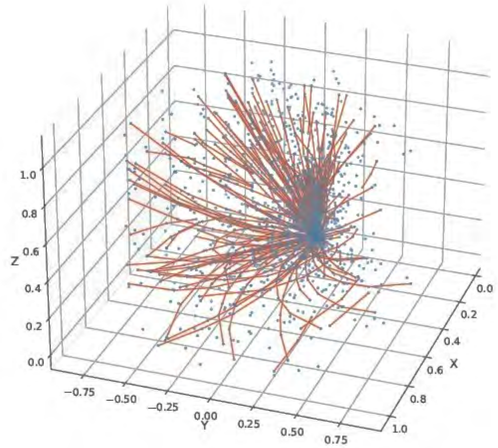
http://ceur-ws.org/Vol-2507/280-284-paper-50.pdf

TrackML Challenge. Our results

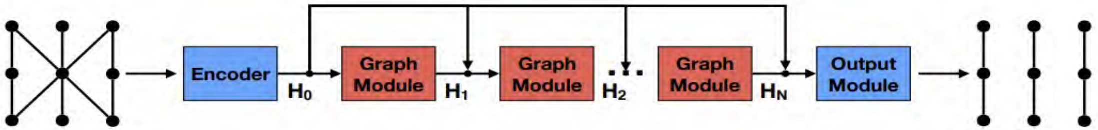
1. Improvement of HEPTrkX GNN model

https://arxiv.org/pdf/2003.11603

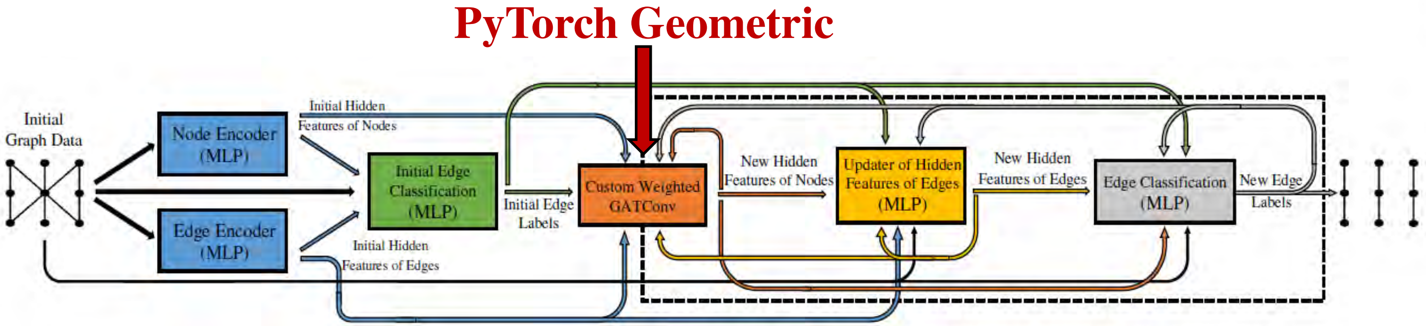
The HEPTrkX GNN model has been substantially redesigned.



In particular, the original structure



has been changed to

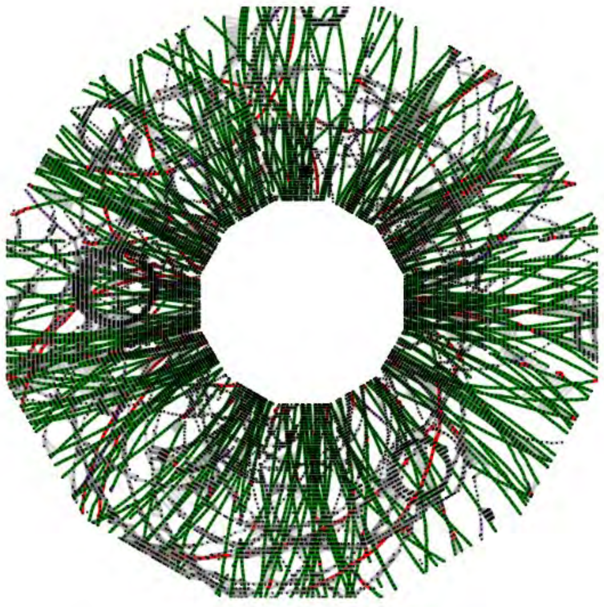


These improvements allowed to reduce the memory to accommodate the event graph CPU-RAM usage from 16 GB to 3GB and keep GPU-RAM usage up to 5GB at the batch size of 8.
Additional study how much the results deteriorate when uniformly distributed noise of 20% of the data is added.

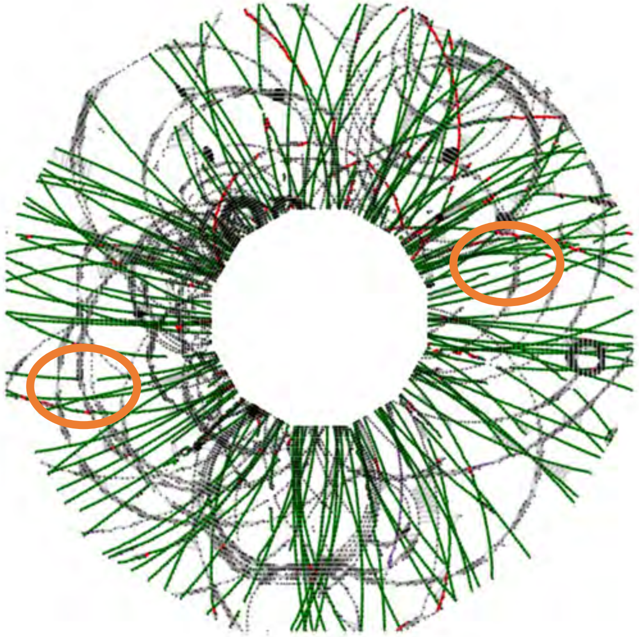
Metrics	w/o noise	20 % of noise
Accuracy	99 %	99 %
Purity & Efficiency	91 %	90 %
AUC	0.996	0.996

Adapting the GNN model to datasets obtained from the MPD experiment of the NICA project, taking into account their specifics, is the next step in the current study.

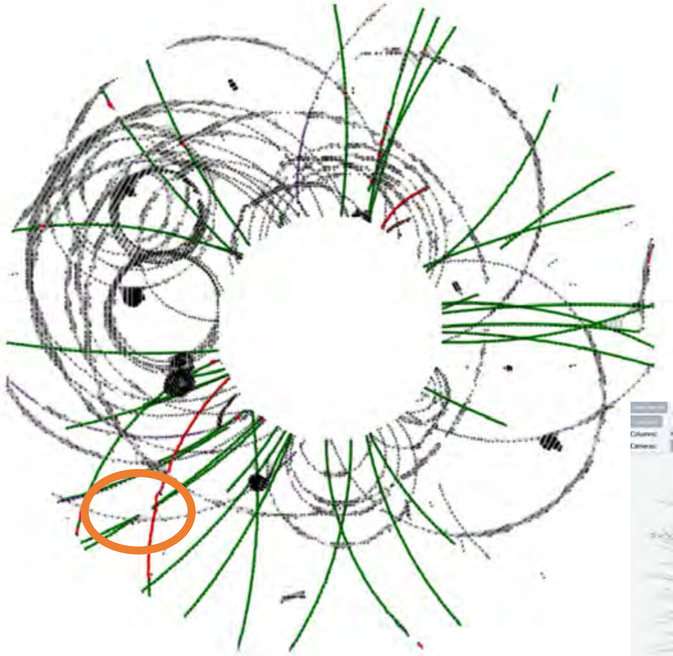
GNN для реконструкции треков в эксперименте MPD



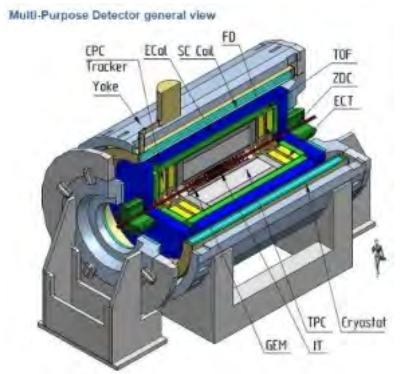
Большая множественность

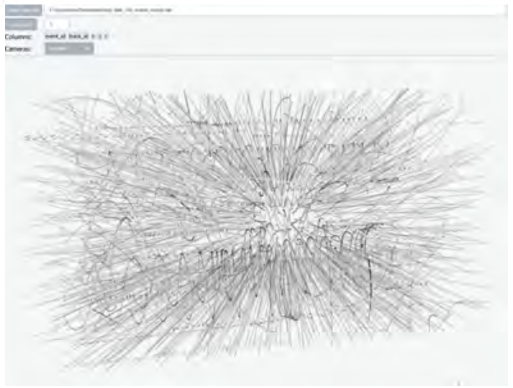


Средняя множественность



Малая множественность





Многоуровневая графическая нейронная сеть была разработана для трекинга в эксперименте MPD.

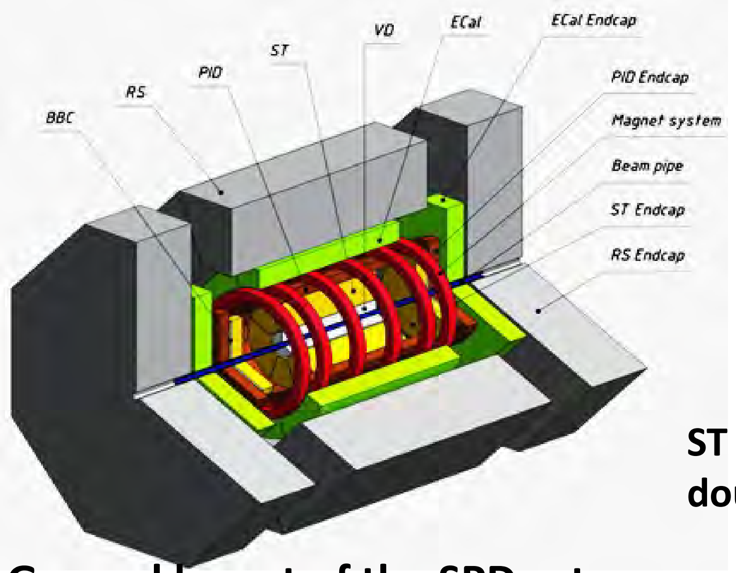
- Стандартная GNN (https://doi.org/10.48550/arXiv.2003.11603)
- Слой внимания
- Двух шаговая агрегация для учета кривизны трека

	Simple GNN	GNN with attention	**Full GNN**
Accuracy	95.9 %	96.1 %	**96.2 %**
Purity & Efficiency	91.8 %	92.2 %	**92.6 %**

1000 Au-Au событий
$p_t < 150$ MeV
80% - обучение
20% - тестирование

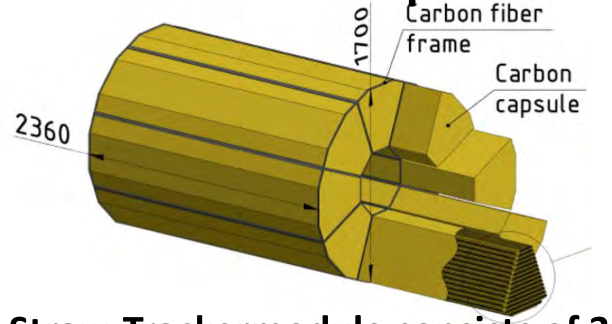
Tracking for data from high luminosity experiments. SPD NICA

SPD (Spin Physics Detector) is being developed to study the spin structure of proton, deuteron and other spin-related phenomena using polarized beams of protons and deuterons at collision energies up to 27 GeV and luminosities up to 10^{32} cm $^{-2}$ s $^{-1}$.



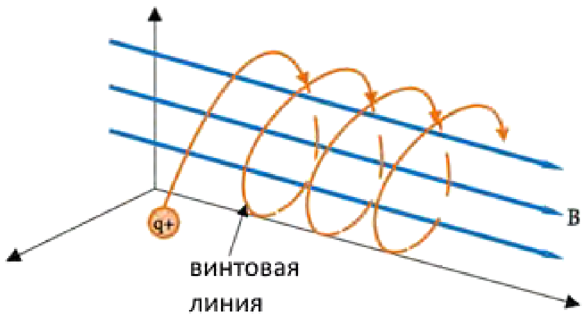
General layout of the SPD setup

Event data from the SPD will be received at 3 MHz as 10 ms time-slice data, with up to 40 events in each time-slice, i.e., one time-slice will contain up to 200 tracks and ~ 1100 hits per station



ST - Straw-Trackermodule consists of 35 double layers of straw-tubes

Reconstruction of events from a dataset of time-slices is required. For this purpose it is planned to develop an <u>algorithm for online filter</u> to process at least 100 time-slices per second



The calculations were performed according to <u>a simplified simulation scheme</u>:

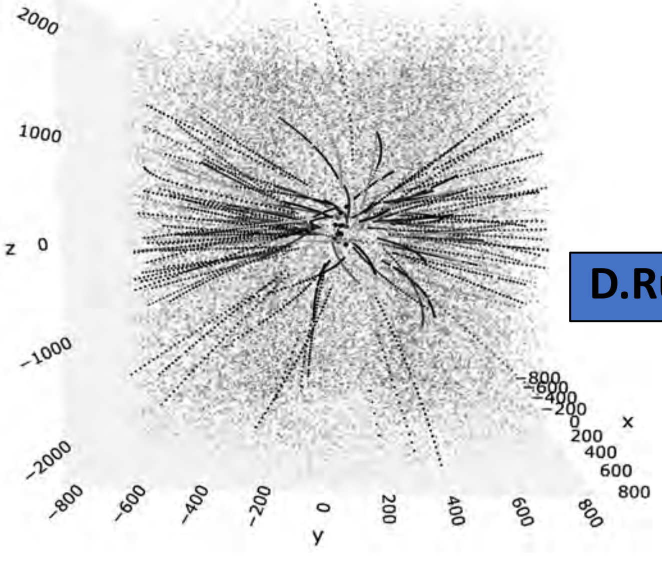
- Python script generates events with 1-10 random tracks.
- Transverse momentum: 100-1000 MeV/c (uniform).
- Random vertex coordinates within the collision area.
- Trajectories follow a helical path, defined by the pitch and radius equations.
- Simulated detector with 35 stations.
- Fake and noise hits simulated using randomly sampled points in detector space.

Deep tracking for SPD NICA timeslice data

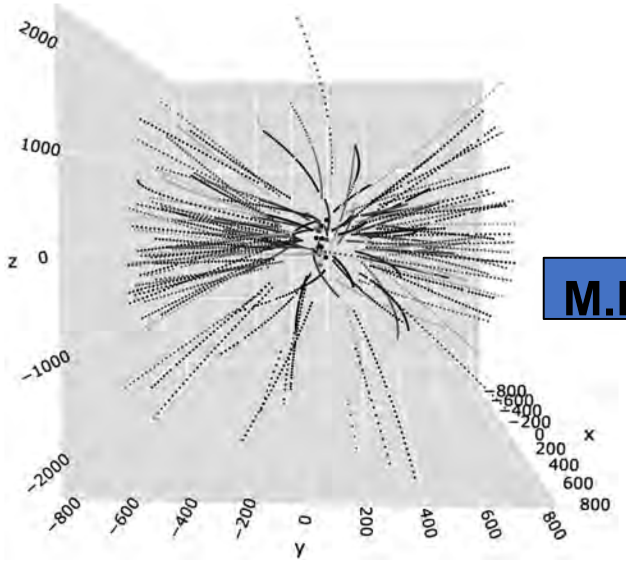
The main problems in SPD tracking are a huge number of fake signals, missed counts due to inefficiency of detectors and “left-right” ambiguity of straw-tubes Introducing appropriate complications in the TrackNET program inevitably slows down its work and reduces its efficiency.

Reconstruction of events from the time-slice dataset was performed in two stages

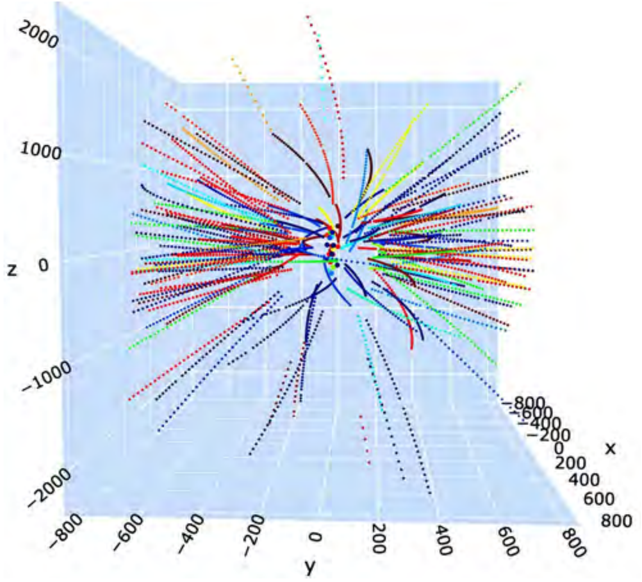
1. On-line tracking(TrackNET) → **2. Unraveling track array by events**



D.Rusov



M.Borisov



By fine-tuning TrackNET on the GOVORUN supercomputer, a processing speed of ~2000 model events per second with acceptable tracking efficiency was achieved

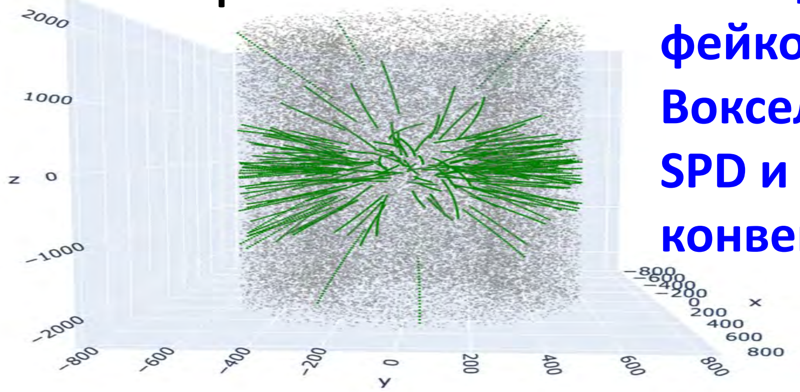
The event unraveling algorithm is based on clustering of feature vectors, obtained using Siamese neural network. The result is quite promising, but requires improvement due to insufficiently low efficiency.

https://doi.org/10.1134/S1063779624030237

Новые подходы с применением нейросети Point cloud transformer (PCT)

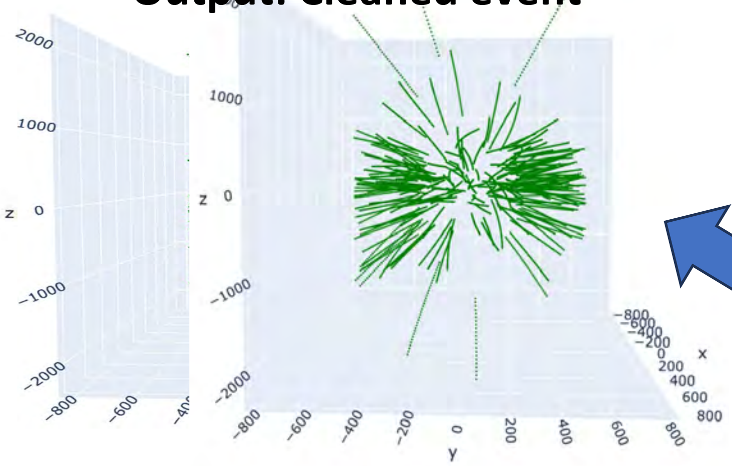
1. PCT для определения числа треков в событии (в планах).

Input: Raw event

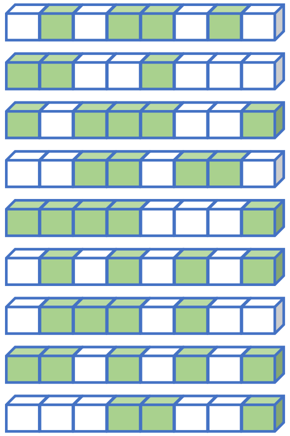


2.PCT для удаления фейков в тайм-слайсе. Вокселизация событий SPD и программный конвейер

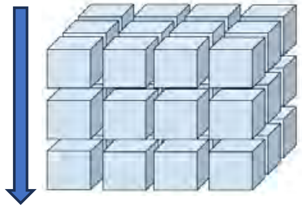
Output: Cleaned event



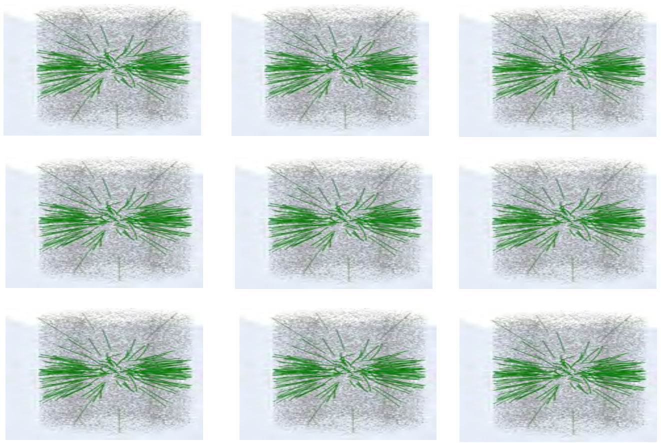
Объединяем все в одно событие

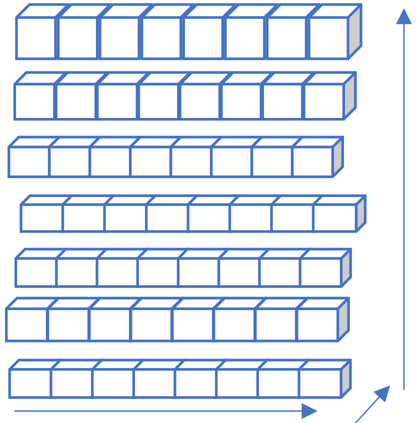


Делим пространство детектора на M вокселей, т.е. на более мелкие подпространства



Берем хиты из каждого вокселя и формируем батч из M x N x F подвыборок, считая, что каждая подвыборка - это мини-событие





M samples

N hits

F features

PCT= Point cloud transformer

PCT model

Классифицируем хиты на истинные и фейки

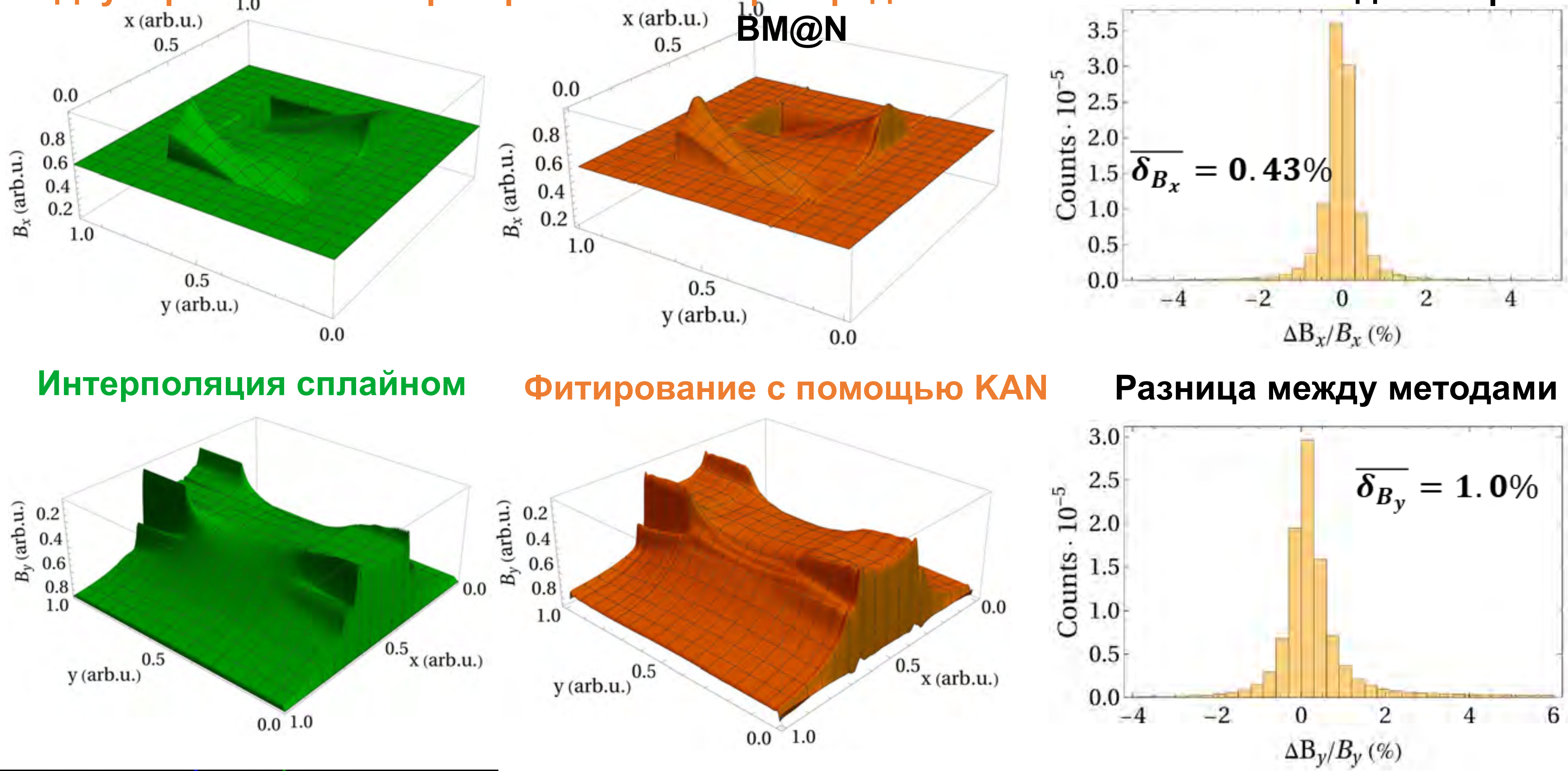
РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ЭСПЕРИМЕНТА

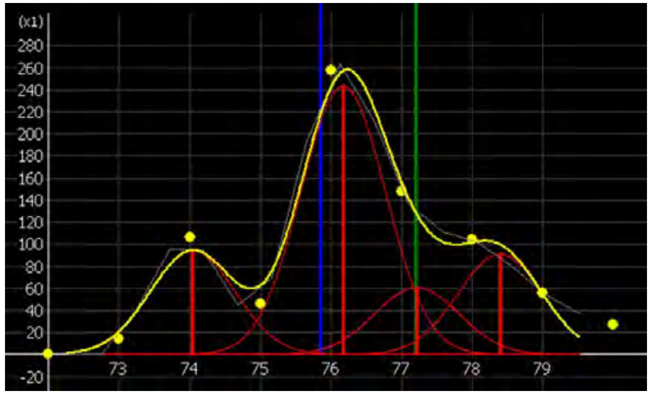
Число событий	Неэффек	Число вокселей	Precision	Recall
40	2%	512	0.98	0.98

https://doi.org/10.1134/S1063779624030638

ML/DL подходы для задач комплекса NICA

Двумерное сечение пространственного распределения магнитного поля в детекторе BM@N





Разработанный подход может быть применен к задаче **реконструкции свойств частиц и струй, получаемых на ускорителе NICA.** Например, разрешение перекрывающихся сигналов в трековых детекторах для обнаружения близко летящих частиц.

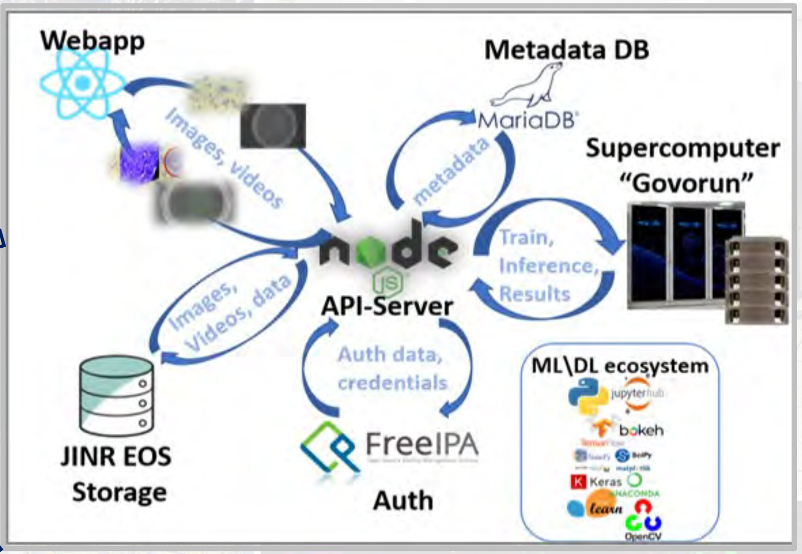
Предложена, улучшена и реализована **нейронная сеть Колмогорова-Арнольда (КАН),** совместимая с оптимизаторами семейства Adam.

KAN обладает **более высокой аппроксимационной способностью** по сравнению с классическим многослойным персептроном (MLP).

Для аппроксимации **магнитного поля в детекторе BM@N** был применен метод KAN. KAN намного быстрее, чем классический метод сплайнов. Значения поля аппроксимируются со средней ошибкой **0,43%** для компоненты поля $\mathbf{B_x}$ и **1%** для компонента $\mathbf{B_y}$.

Радиобиология и науки о жизни

Применение накопленного опыта к различным классам задач

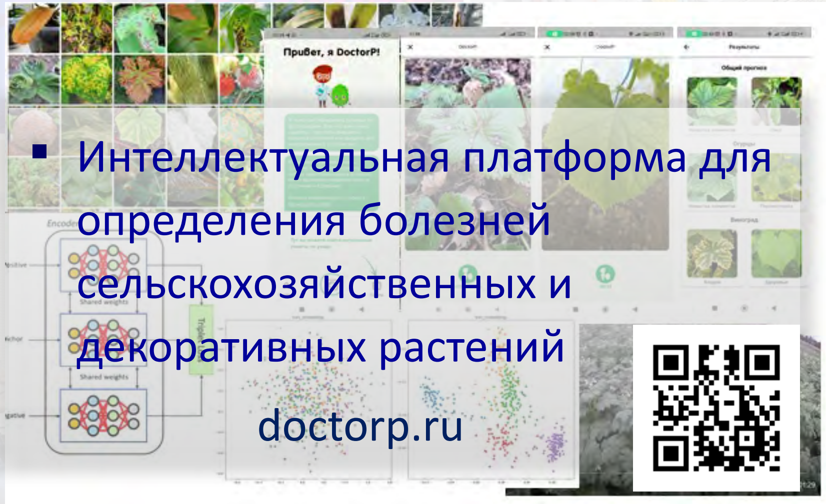


ML/DL/HPC экосистема

- ML/DL технологии
- Современные IT-решения для хранения, обработки и визуализации данных
- Статистический анализ

- Интеллектуальная платформа мониторинга окружающей среды

moss.jinr.ru



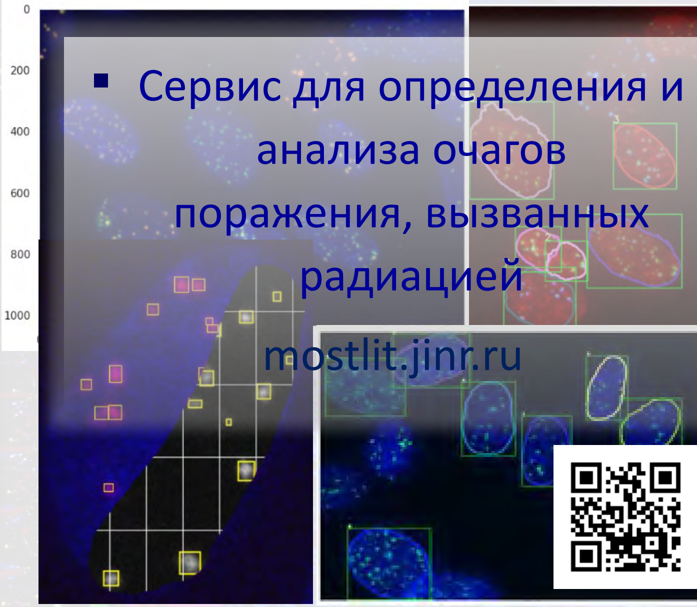
- Интеллектуальная платформа для определения болезней сельскохозяйственных и декоративных растений

doctorp.ru

- Информационная система для задач радиобиологии

bio.jinr.ru

хранение,быстрый доступ и анализ данных радиобиологических экспериментов

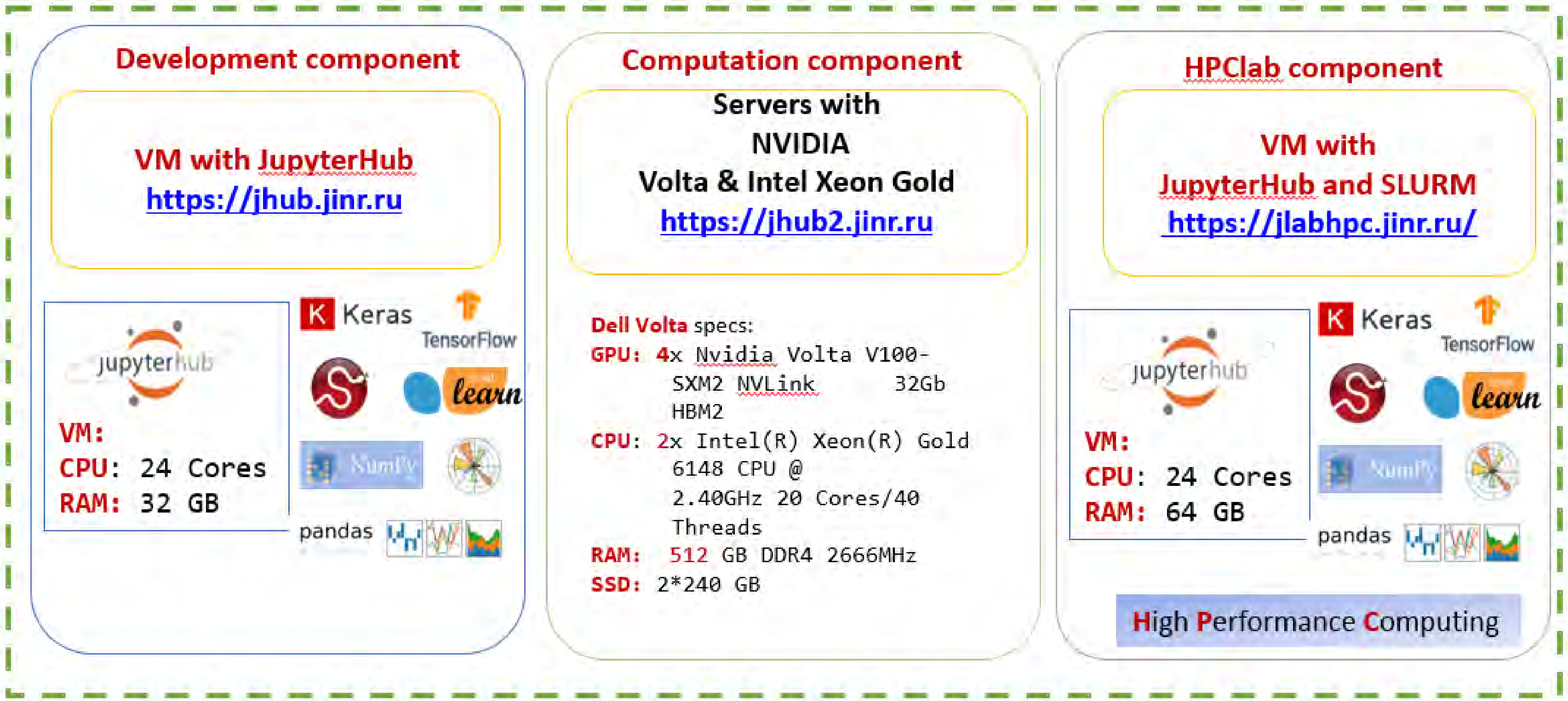


- Сервис для определения и анализа очагов поражения, вызванных радиацией

mostlit.jinr.ru

Экосистема для задач машинного обучения, глубокого обучения и анализа данных

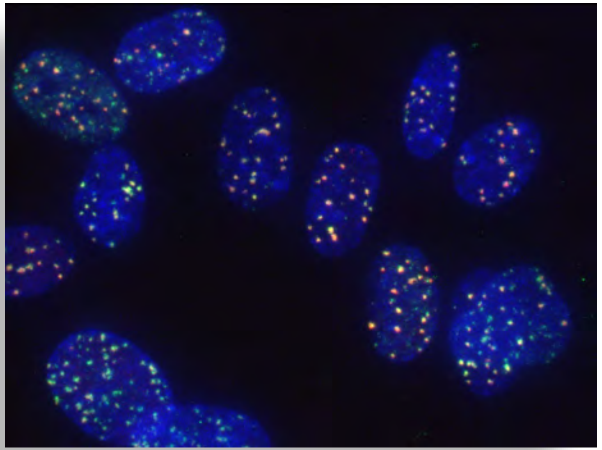
Обеспечение работ по разработке методов и алгоритмов машинного обучения и глубокого обучения, а также среда для анализа и визуализации данных. Экосистема реализована на базе JupyterHub –многопользовательской платформы по работе с Jupyter Notebook, включающая в себя набор библиотек и фреймворков:

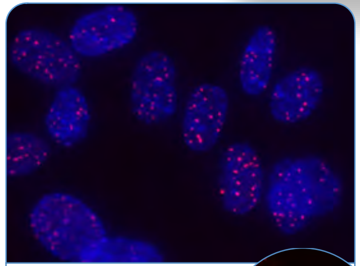


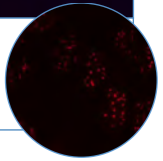
- вычислительная компонента предназначена для проведения ресурсоемких, массивно-параллельных задач обучения нейронных сетей с использованием графических ускорителей NVIDIA (https://jhub2.jinr.ru);
- компонента JLabHPC обеспечивающая проведение расчетов на вычислительных узлах платформы HybriLIT, в том числе на СК «Говорун» (https://jlabhpc.jinr.ru);
- компонента для разработки моделей и алгоритмов и анализа данных (https://jhub.jinr.ru).

Информационно-вычислительная системы для автоматизации обработки данных радиобиологических исследований

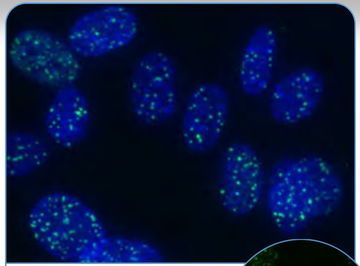
Радиационно-индуцированные Фокусы (РИФ)

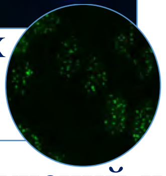




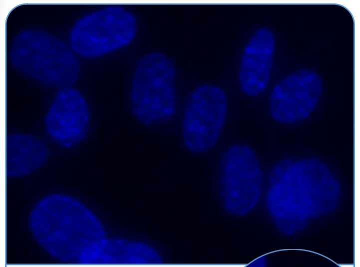


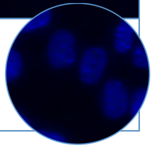
Клетки с 53BP1
DAPI+красные РИФ





Клетки с γH2AX
DAPI+зелёные РИФ

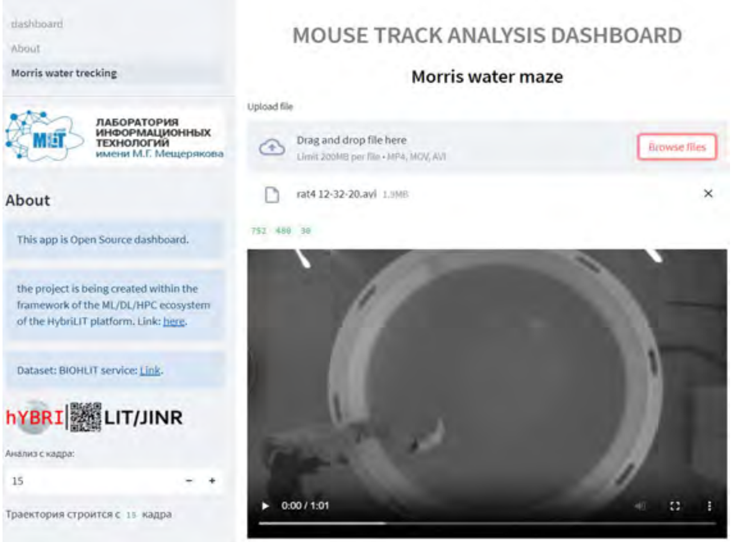


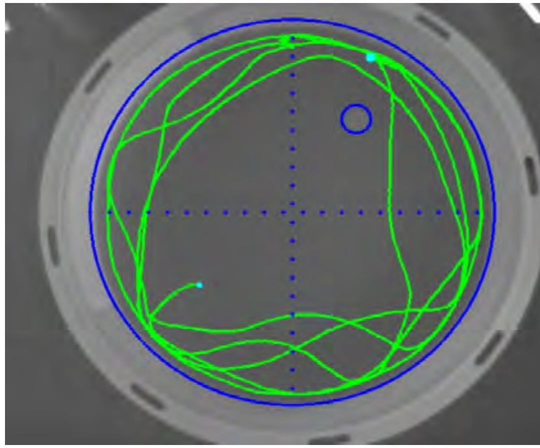


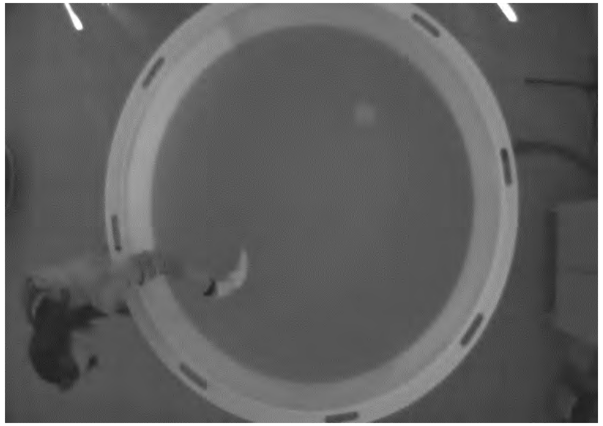
Ядра клеток
DAPI

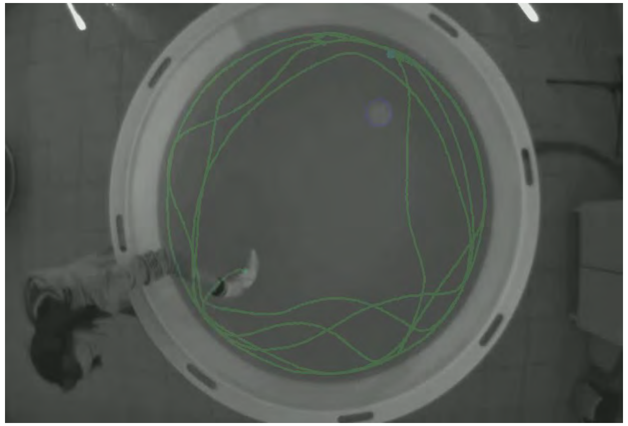
При действии ионизирующих излучений на живые клетки возникает целый спектр повреждений ДНК, в том числе двунитевые разрывы (ДР).

Для визуализации ДР ДНК используется метод иммуноцитохимического окрашивания: выявляются специфические белки-маркеры, которые накапливаются в сайтах возникновения ДР ДНК и формируют радиационно-индуцированные фокусы (РИФ).



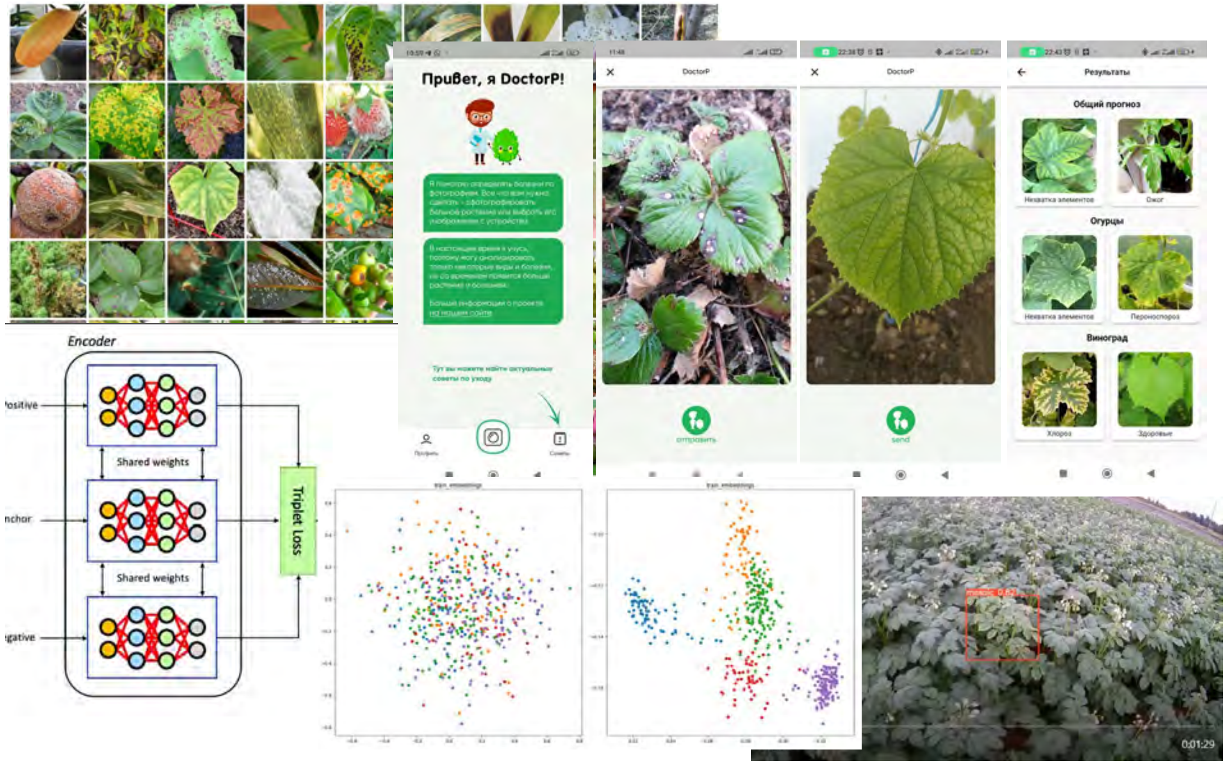






Построение траектории передвижения животного на основе алгоритмов компьютерного зрения.

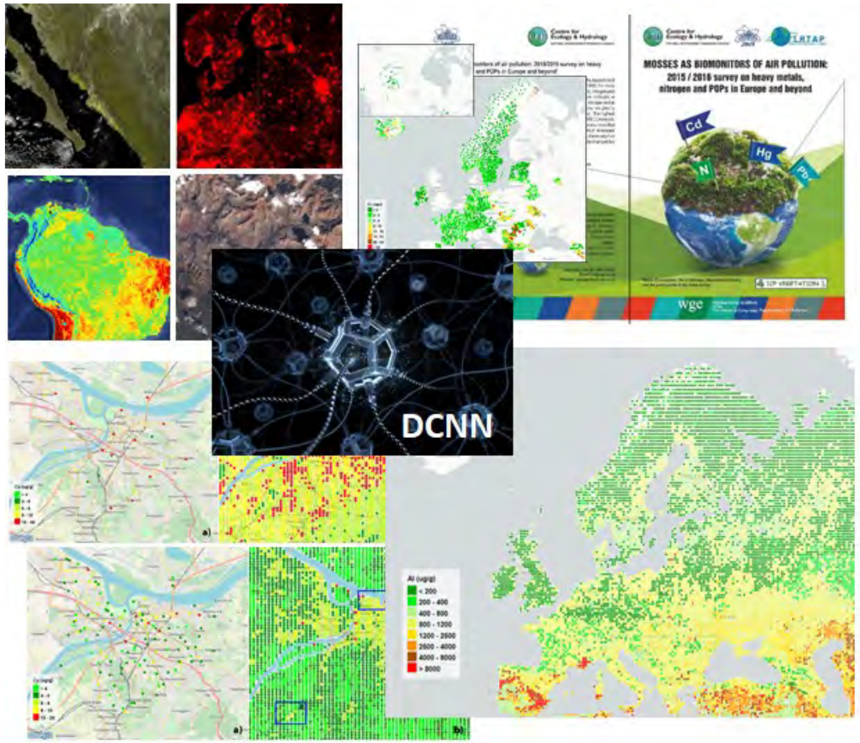
Развитие интеллектуальной платформы определения состояния сельскохозяйственных и декоративных растений



Результаты

Интеллектуальная платформа определения состояния сельскохозяйственных и декоративных растений, в которой представлен значительный объем определяемых болезней, подробные планы лечения и используются передовые технологии искусственного интеллекта.

Развитие платформы контроля и прогнозирования состояния окружающей среды



Результаты

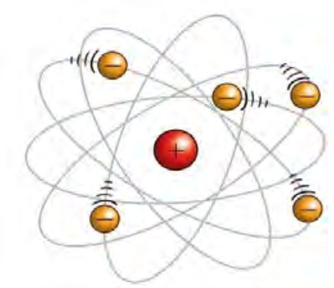
Платформа контроля и прогнозирования состояния окружающей среды, сочетающая в себе передовые технологии управления данными и искусственного интеллекта для решения задач экологического мониторинга.

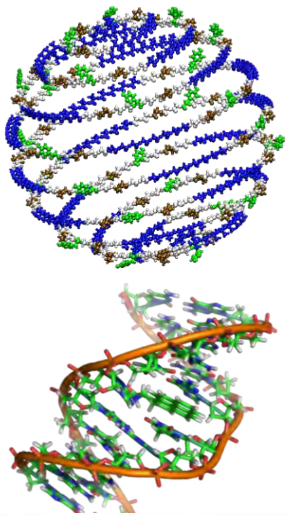
Квантовые симуляторе на СК «Говорун»

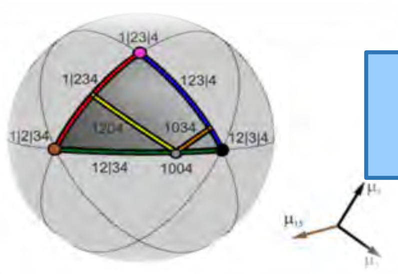
Создание квантового компьютера ⟷ Создание квантового симулятора

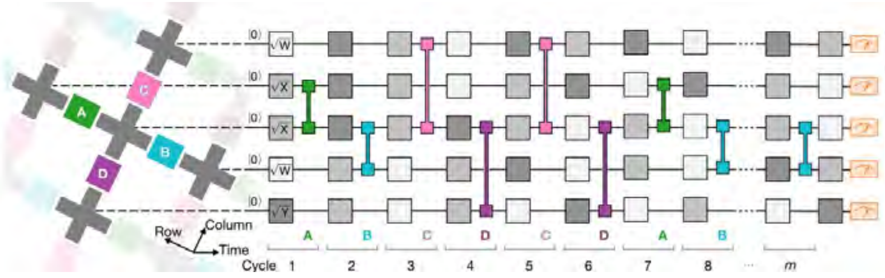
Разработка прикладных квантовых алгоритмов

Моделирование квантовых систем



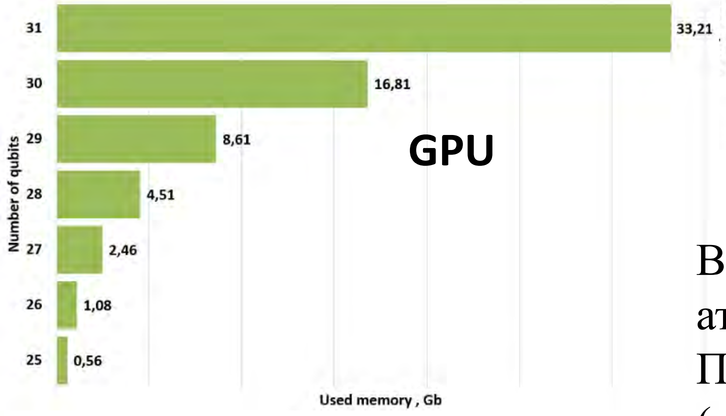


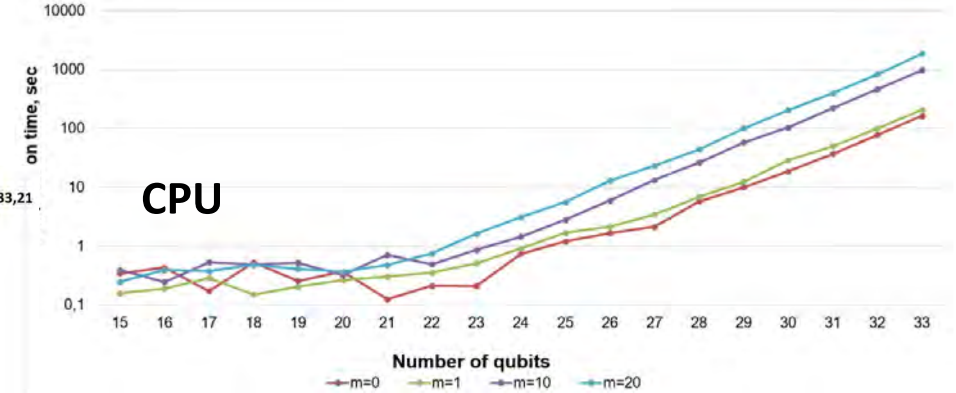




На СК «Говорун» была настроена программная среда, содержащая набор квантовых симуляторов, способных работать на различных вычислительных архитектурах.

Было показано, что на текущих ресурсах СК «Говорун» можно смоделировать до 38 кубитов на симуляторе QuEST. Однако этого числа недостаточно для полного моделирования электронных оболочек сверхтяжелых элементов, например, для моделирования 118-го элемента необходимо 118 кубитов.





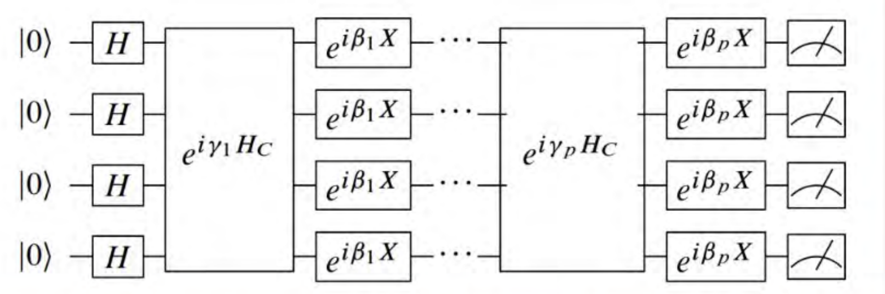
В рамках проекта «Сверхтяжелые ядра и атомы: пределы масс ядер и границы Периодической таблицы Д.В. Менделеева» (грант Министерства образования и науки РФ № 075-10-2020-117)

Quantum Computing Polygon

deployed on the ML/DL/HPC ecosystem of the HybriLIT platform.

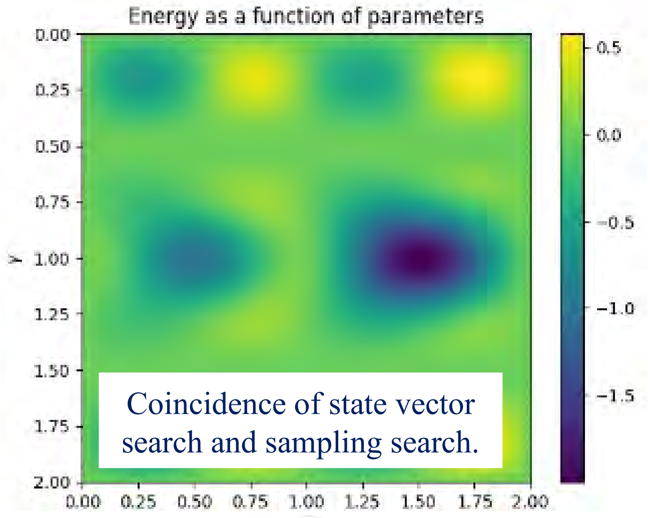
Cirq | dwave-neal | dwave-qbsolv | PennyLane | Qiskit | QuTiP_py36

As an example, we present a search for the ground state and its energy in the Ising model with a longitudinal magnetic field using the quantum approximation optimization algorithm (QAOA).



A quantum circuit to the variation ansatz of QAOA

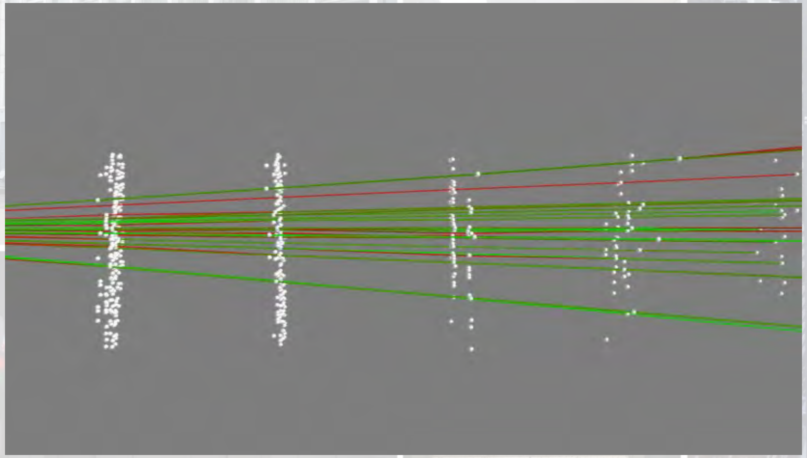
$$|\psi(\gamma,\beta)\rangle = \underbrace{U(\beta_p,B)U(\gamma_p,\mathcal{H})}_{p}\dots\underbrace{U(\beta_1,B)U(\gamma_1,\mathcal{H})}_{1}\, H^{\otimes n}|0\rangle^{\otimes n}$$

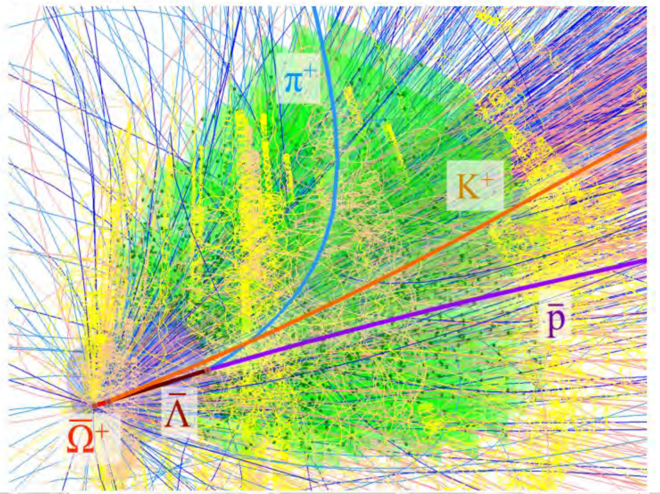


OpenMP for CPU computations and cuStateVec for GPU computations

Ising Model 3x3x3 lattice 27 qub	AMD EPYC 7763, 128 th.	Intel Xeon Platinum 8368Q, 128 threads	NVIDIA A100, cuStateVec
Comp. time	3 h 20 min	3 h 10 min	**14 min 35 sec**

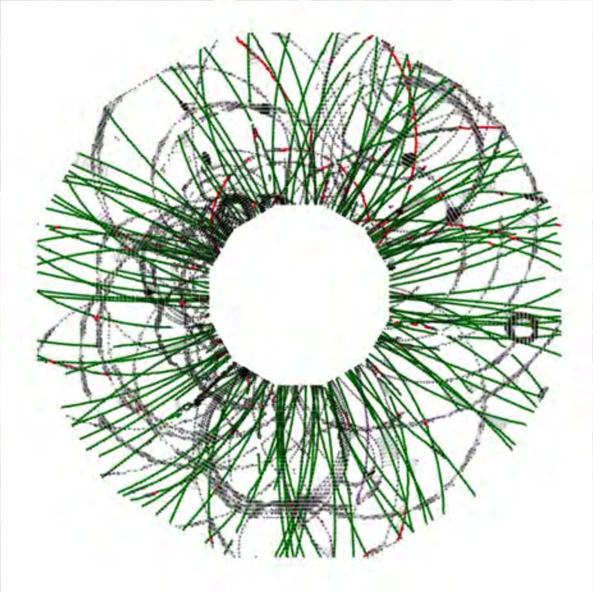
Yu. Palii, A. Bogolubskaya, and D. Yanovich: Quantum Approximation Optimization Algorithm for the Ising Model in an External Magnetic Field // PEPAN, V. 55, N. 3. Pp. 600-602, 2024.

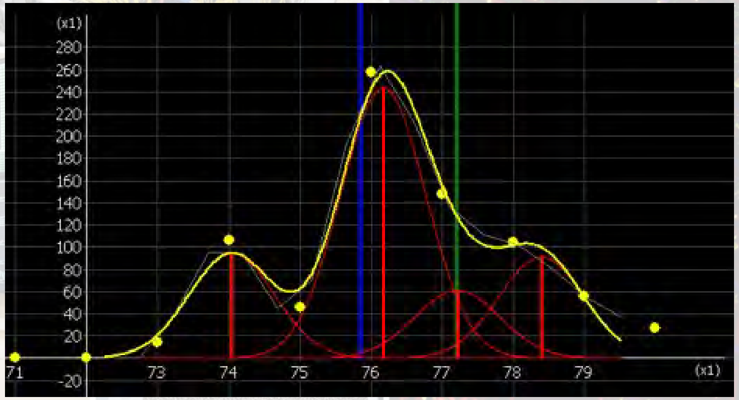


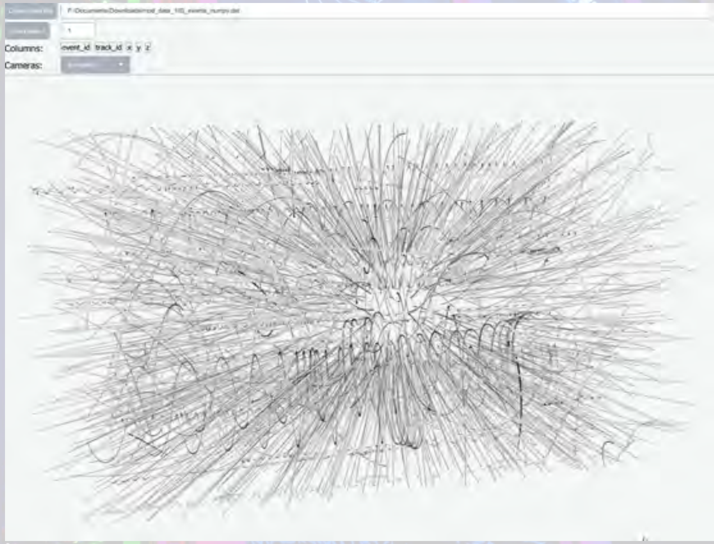


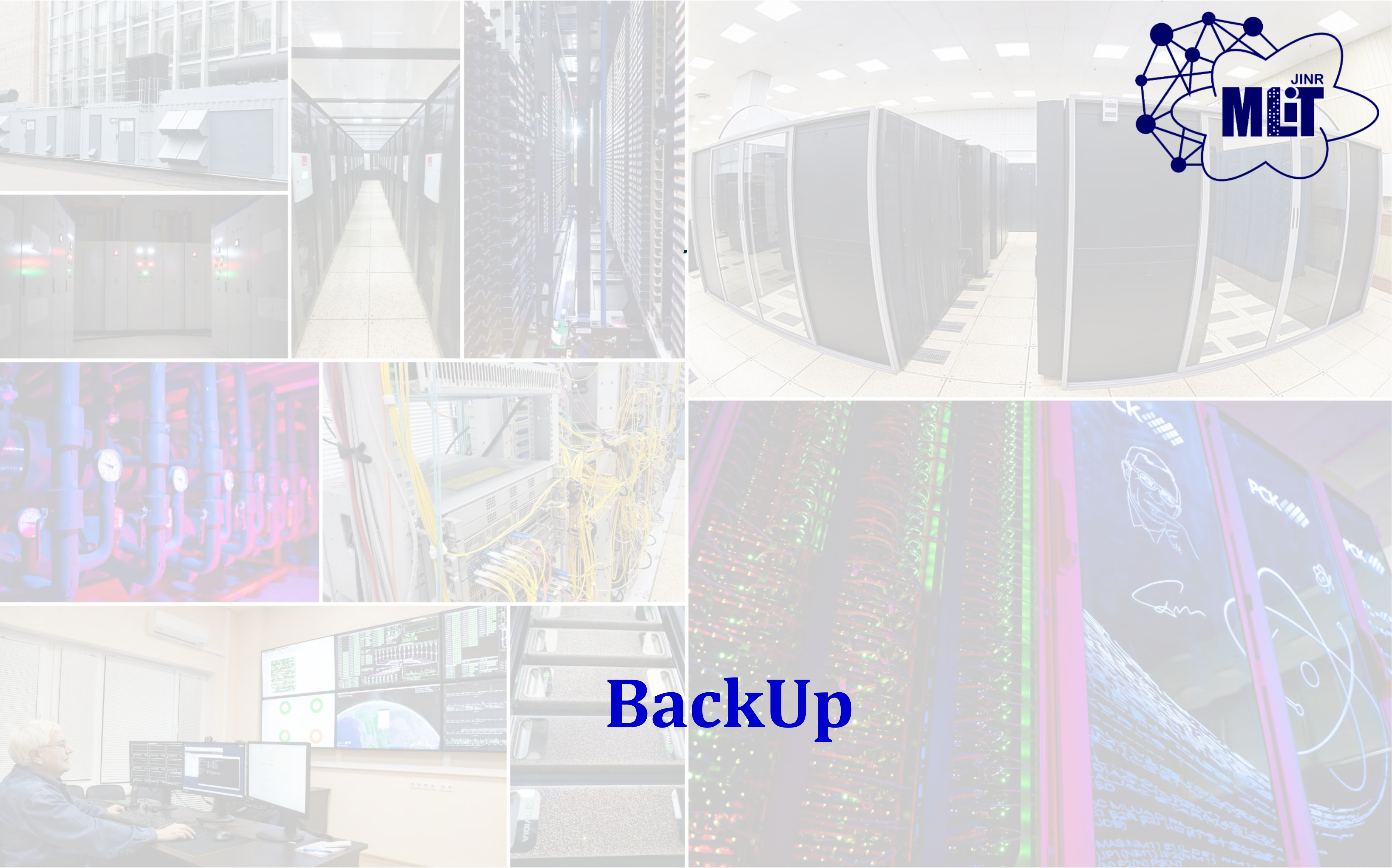
Thank you for your attention!

nvoytish@jinr.ru









BackUp

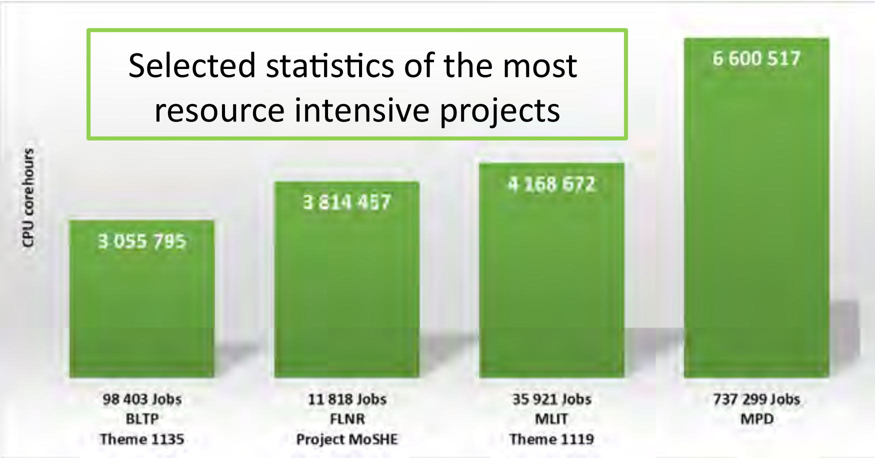
"Govorun" Supercomputer for JINR tasks

Projects that mostly intensively use the CPU resources of the "Govorun" supercomputer:

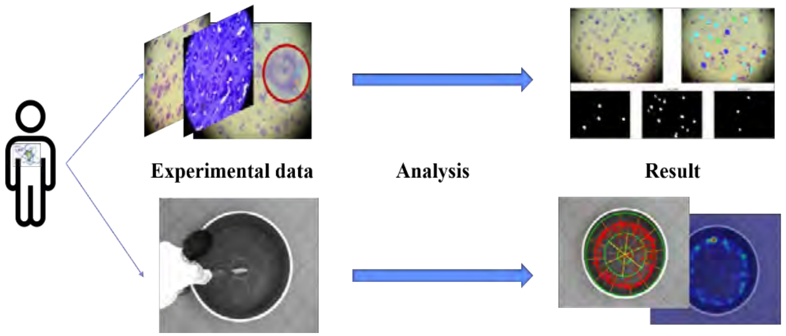
- NICA megaproject,
- simulation of complex physical systems,
- computations of the properties of atoms of superheavy elements,
- calculations of lattice quantum chromodynamics.

The GPU component is actively used for solving applied tasks by the neural network approach:

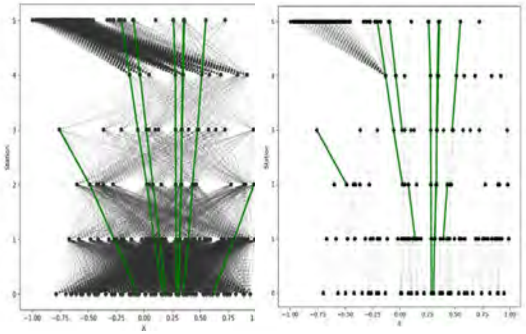
- processing of data from experiments at LRB,
- data processing and analysis at the NICA accelerator complex, etc.



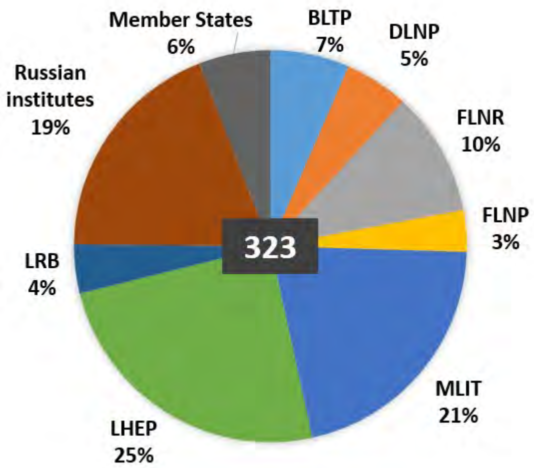
During 2022, **890 911** jobs were performed on the **CPU** component of the "Govorun" supercomputer, which corresponds to **18 543 076** core hours.



Information system for radiation biology tasks



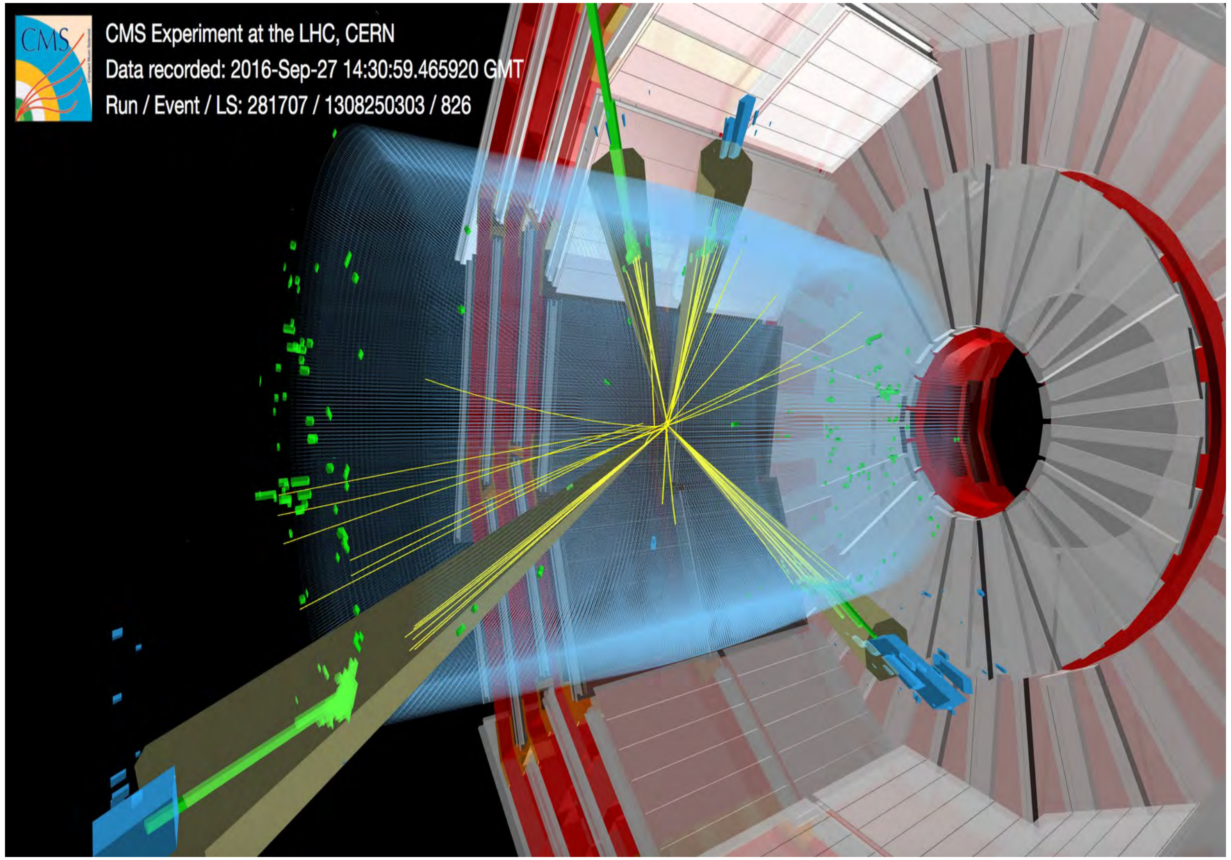
Neural network for HEP data reconstruction and analysis



The resources of the "Govorun" spercomputer are used by scientific groups from all the Laboratories of the Institute within **25 themes** of the JINR Topical Plan.

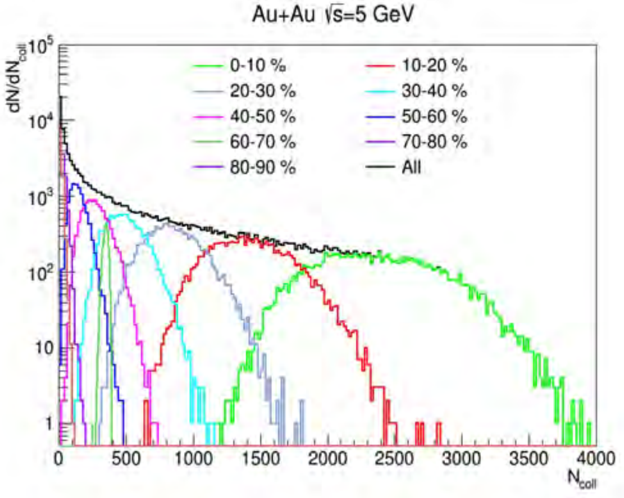
Simulation of physical processes and experimental facilities

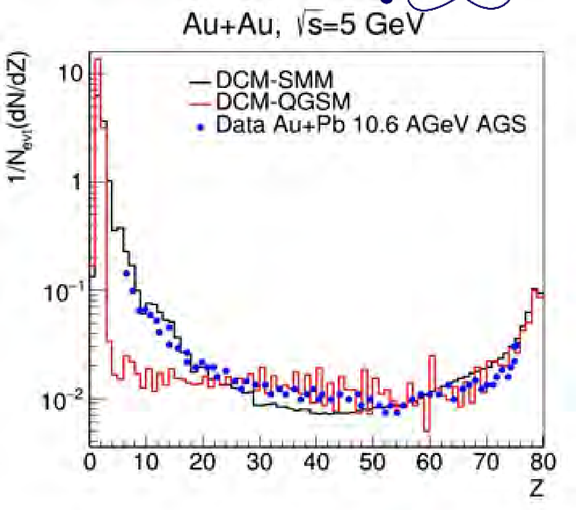
- analytical and numerical calculations of physical processes, software optimization, including tuning and adaptation of physics event generators;
- MC event production, development and support of information systems for event catalogues;
- participation in the creation of computer models of experimental facilities and simulation of elementary particles passing through them based on GEANT4 (and others) and fast simulation of the response of the detectors.



MC Generators for the NICA and LHC Experiments

- Development of the heavy ion collision generators
  - Dubna Cascade Model, Quark-Gluon-String Model, Statistical Multifragmentation Model for the NICA Experiments
  - tuning the HIJING generator with data of NA49 and NA61/SHINE @ CERN, STAR@RHIC (can be used in MPD and SPD experiments)
- Analytical and numerical methods for calculating neutron-proton systems under strong compression at the NICA



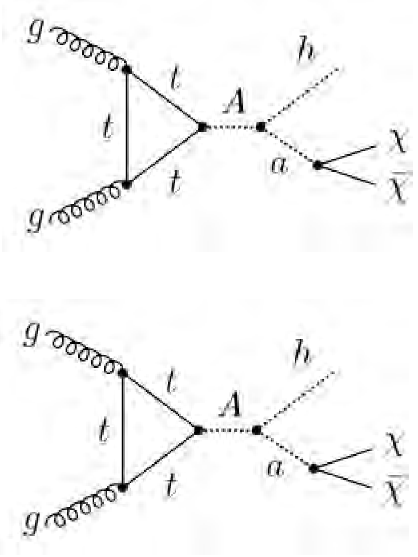


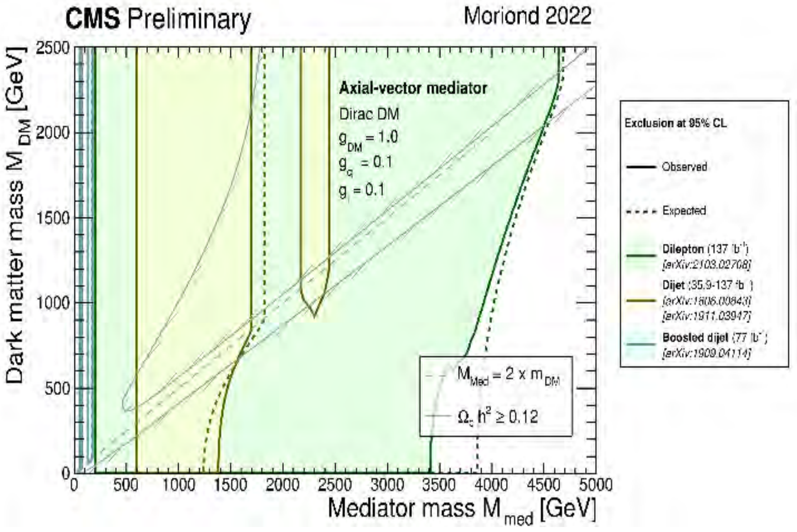
The priority tasks of the JINR Groups in LHC physics program (ATLAS and CMS) include searches for candidates for dark matter particles, tests of predictions of TeV-energy scenarios

- Fine tuning the generators for searches for new physics
  - revision of model parameters for 2HDM+a, 2HDM+s, etc.
  - simulation with Pythia8, QBH, MadGraph5_aMC@NLO + FeynRules (simplified DMM, HDM+a, 2HDM+s, etc.)
  - mass production + Geant4 response

Ex., Dark Matter can be probed with two fermions/two fermions + MET/higgs + MET/Z + MET in the final states

h (→ bbar) + a (→ χχ) = bbar + MET



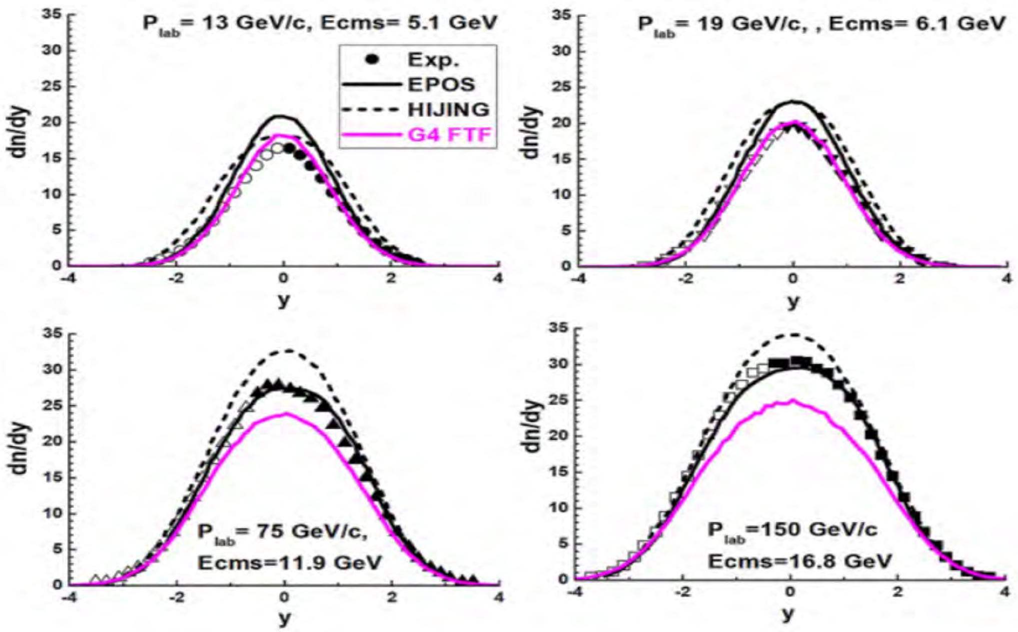


MLIT + BLTP + VBLHEP

Modeling Experimental Facilities for the NICA, LHC, etc.

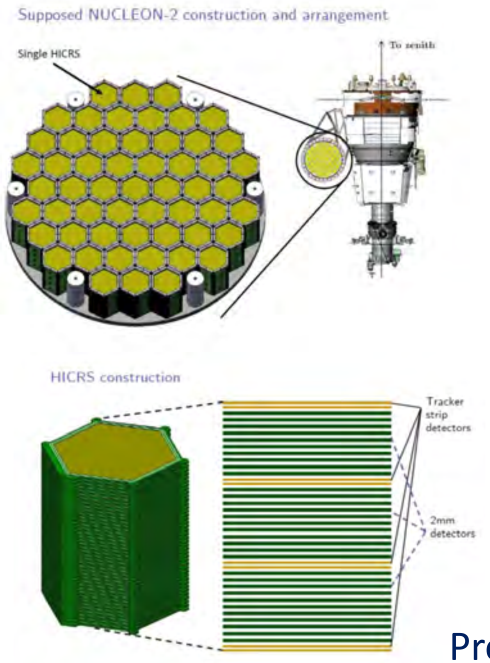
- Development, verification, validation and application of FTF (Fritiof) and QGSM (Quark-Gluon-String-Model) hadronic models

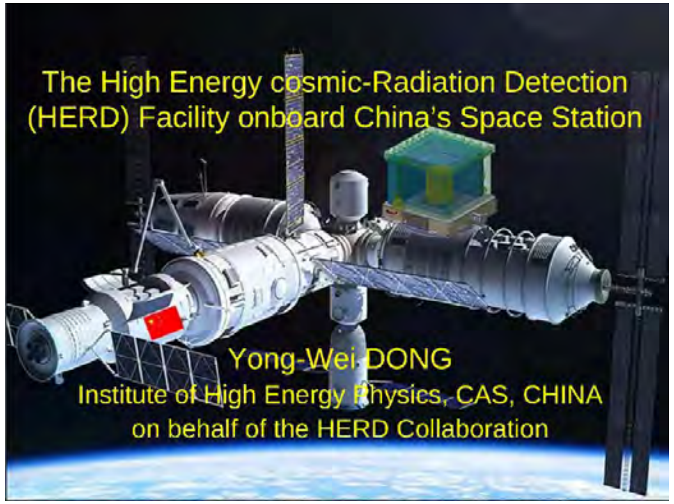
Rapidity distributions of π^- mesons in $^{40}Ar + ^{45}Sc$ interactions (EPJ , C82 (2022)



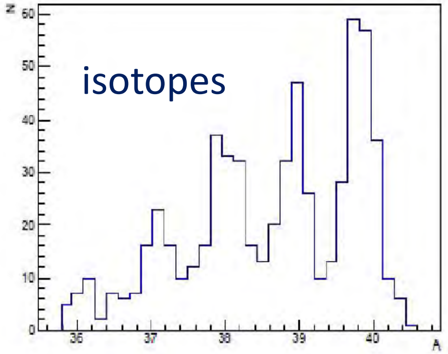
Exp. conclusion: “There is no model (EPOS, UrQMD, HIJING …) able to describe the data!” from NA61/SHINE Collab. on PP, $^{40}Ar + ^{45}Sc$ and $^{7}Be + ^{9}Be$

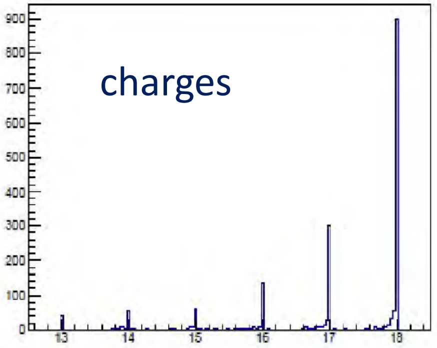
- Simulation and prototype testing for present and future orbital detectors: NUCLEON, NUCLEON-2, HERD





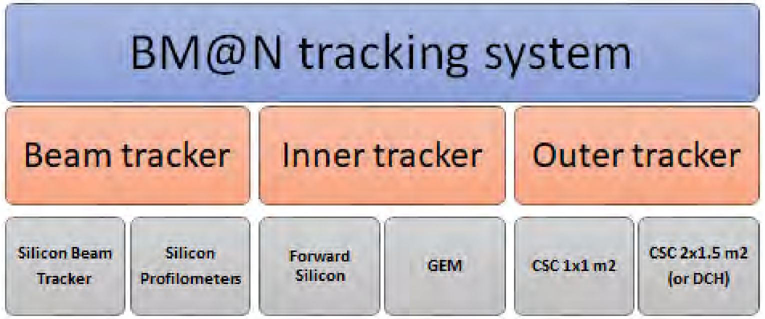
Prototypes tested @ Nuclotron JINR

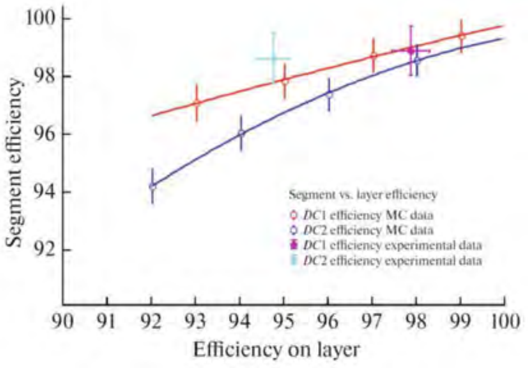




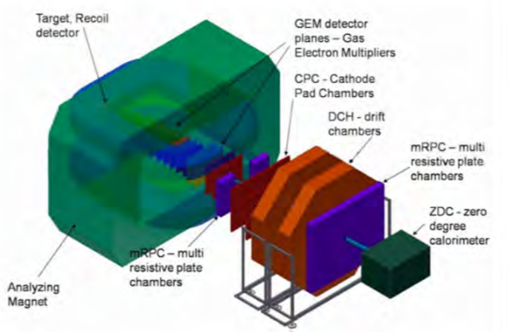
Tracking Algorithms for HEP Experiments

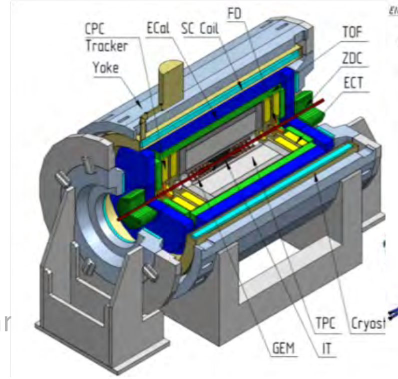
- Mathematical methods and software for processing and analyzing data from the experiments @ NICA
  - software for alignment and calibration of the BM@N STS (silicon chambers) and GEM (gas electron multipliers) track detectors

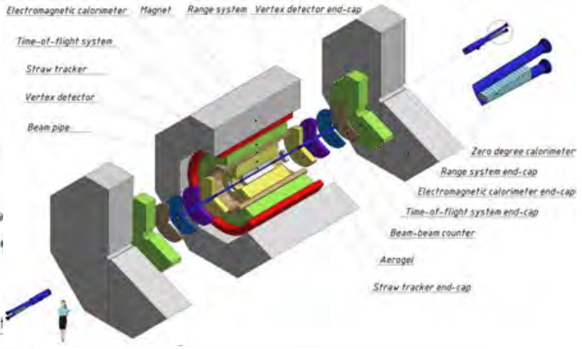




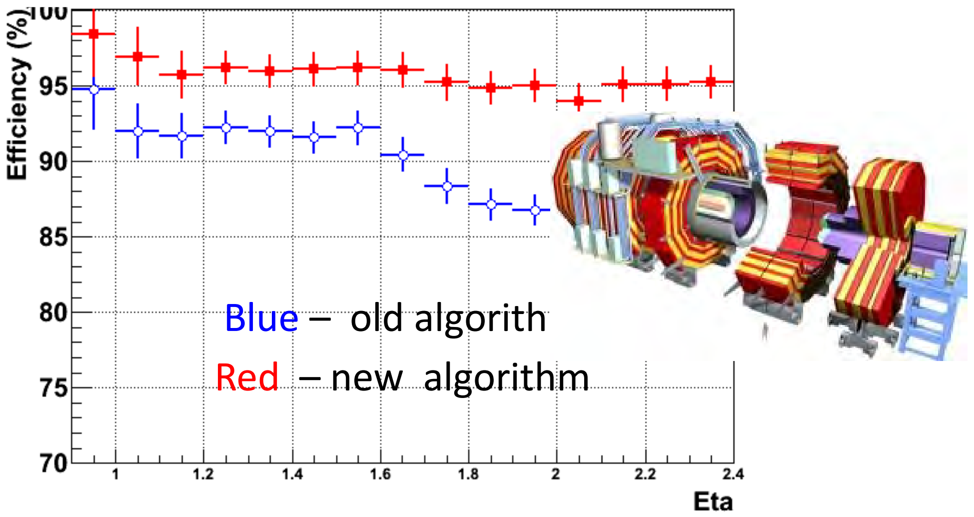
  - development and application of methods and algorithms for processing and analyzing experimental data for the coordinate detectors





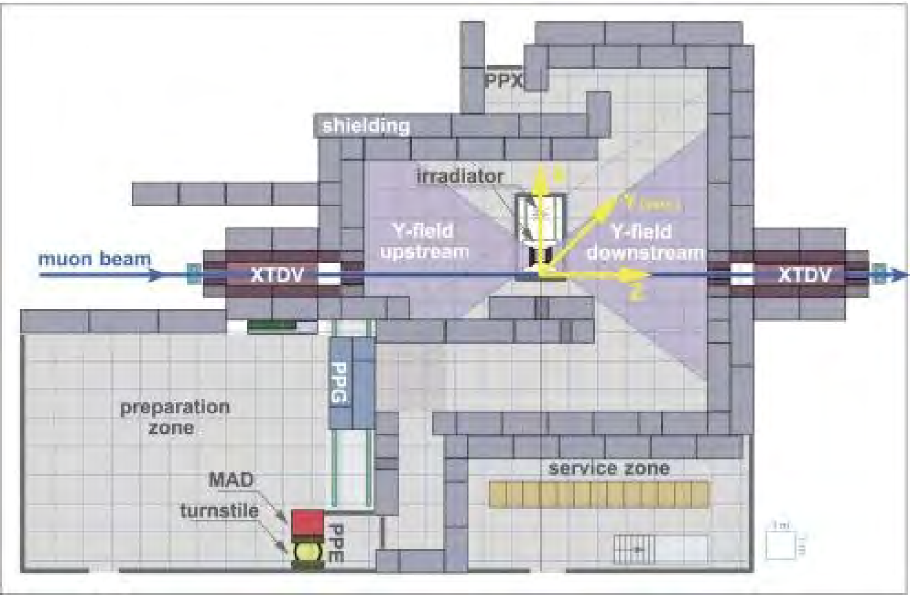


- Mathematical methods and software for muon reconstruction and the estimation of operation parameters of CMS detectors @ LHC
  - reconstruction of the cosmic muon trajectory in the setup for testing active elements of the CMS HGCal, as well as evaluation of the efficiency of HGCal modules;
  - usage of discrete wavelet analysis to recognize the coordinates of close-flying particles from over-lapping signals in the Cathode Strip Chambers (CSC). Evaluation of the operation parameters of CSC detectors and of the rate of background particles for different types of experimental data.

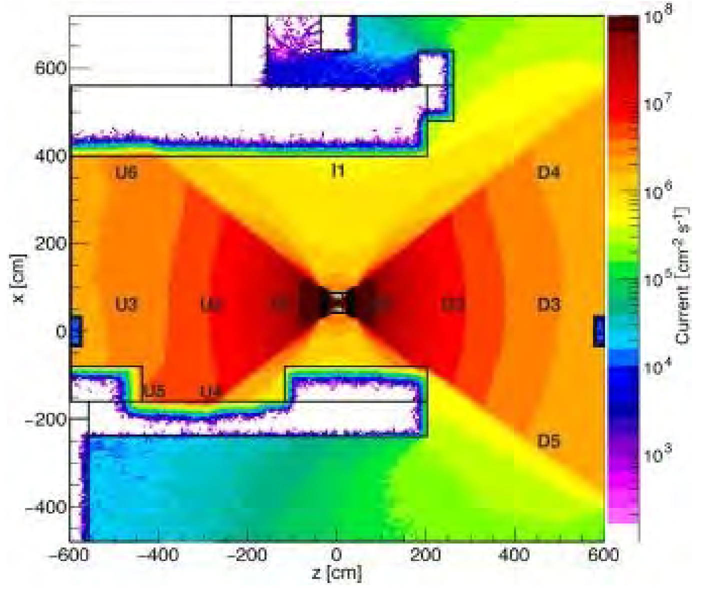


Testing RU algorithm with GIF++ data

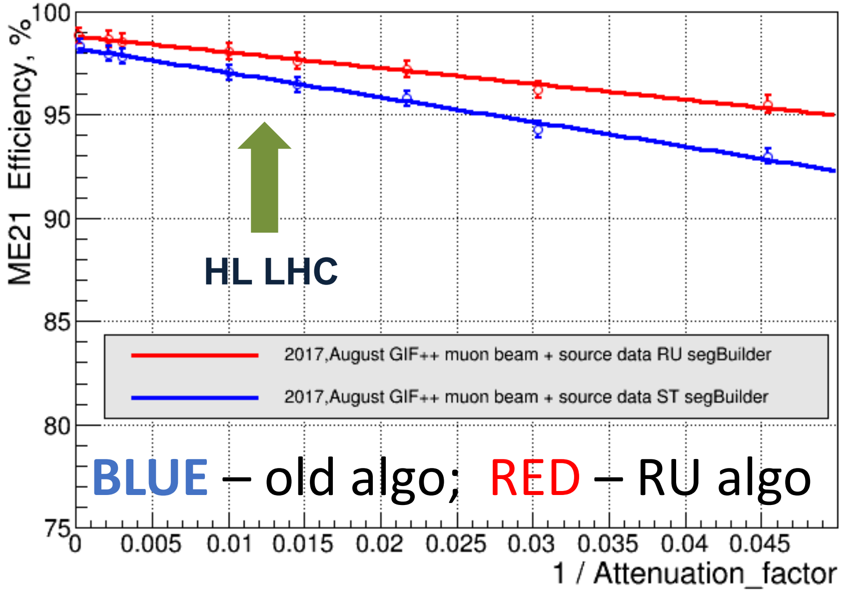
CERN Gamma Irradiation Facility **(GIF ++)**



Facility scheme



Radiation field map



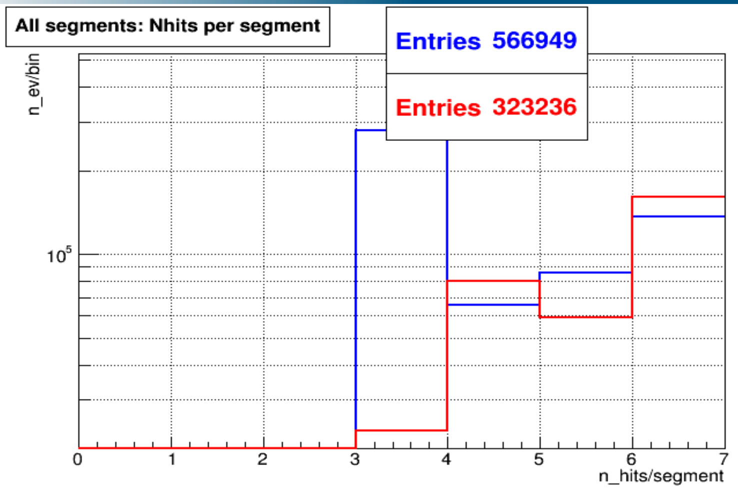
Reconstruction efficiency vs. background noise level

Goal: study of the performance and stability of the LHC detectors and future upgrades for the **HL-LHC at CERN**.

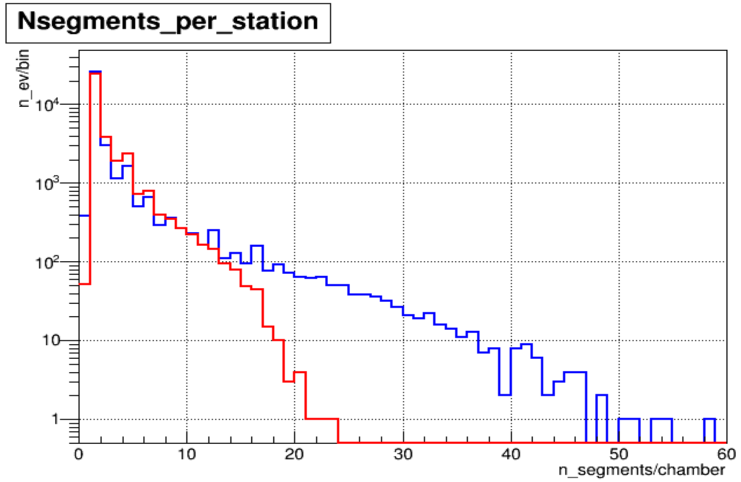
Methodology: high energy charged particle beams (mainly muon beam with momentum up to 100 GeV/c) are combined with a 14 TBq ^{137}Cesium source, allowing to accumulate doses equivalent to HL-LHC experimental conditions.

It was proven that CSCs can survive the HL-LHC conditions.

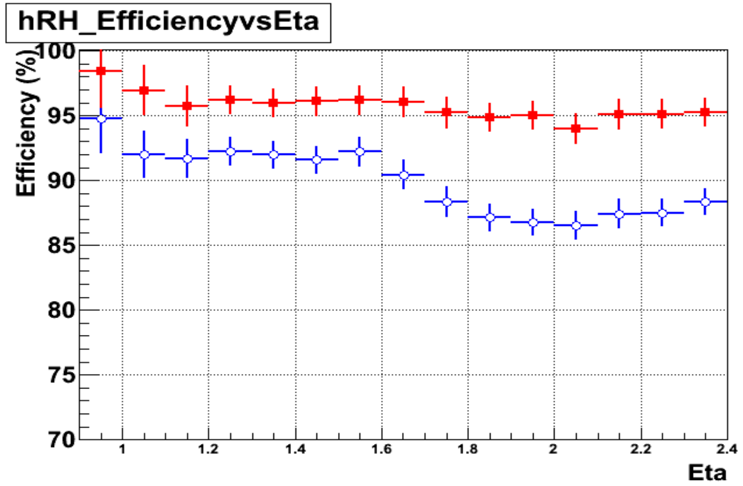
Two algorithms comparison



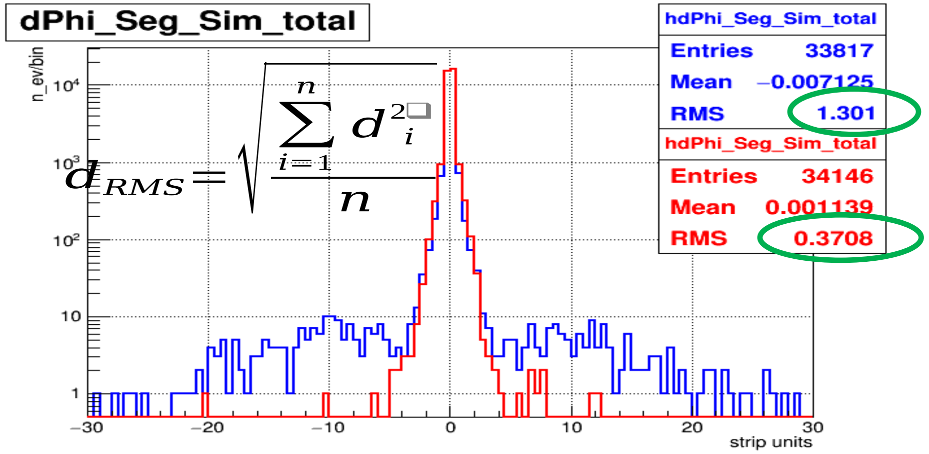
Segment length



Number of reconstructed segments in chamber



Reconstruction efficiency vs. pseudorapidity



$$d_{RMS} = \sqrt{\frac{\sum_{i=1}^{n} d_i^2}{n}}$$

Difference in precise coordinates between the reconstructed segment and simulated trajectory

- The number of short segments decreased by order of **12.**
- The number of reconstructed track segments has decreased, while the number of cases where no segments were reconstructed has decreased by **8 times**.
- The reconstruction efficiency has been improved and has become **more uniform** depending on the pseudorapidity.
- The reconstructed track segments became **3.5 times closer** to the simulated muon trajectory.

Simulation data results.

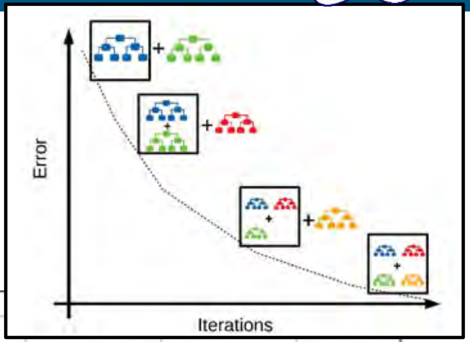
Muons with p_T = 1000 GeV

10K events.

Blue – previous algorithm; **red** – RU algorithm.

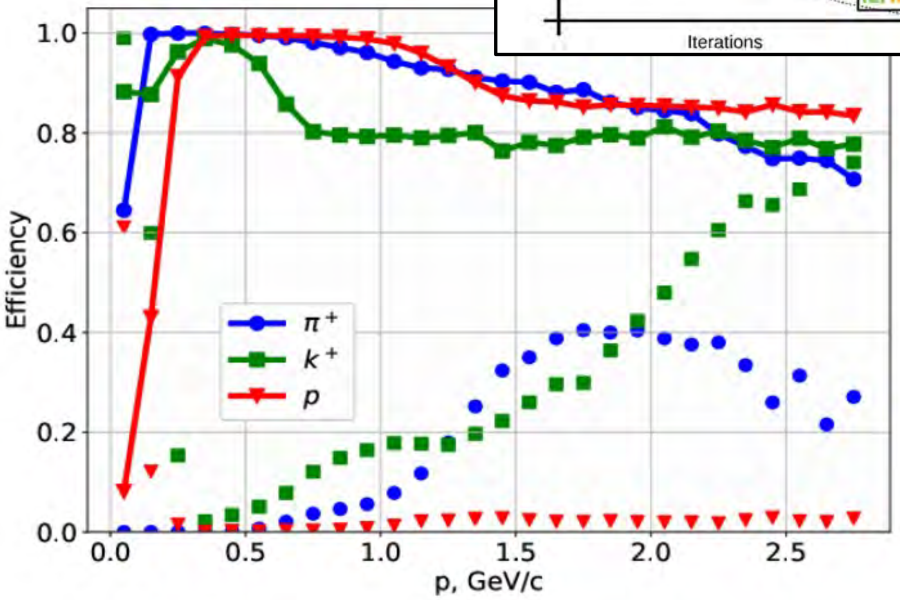
Machine Learning Methods for the Track Reconstruction and PID

- particle identification based on the gradient boosting of decision trees with MPD Time-Projection Chamber (TPC) and Time-of-Flight (TOF)

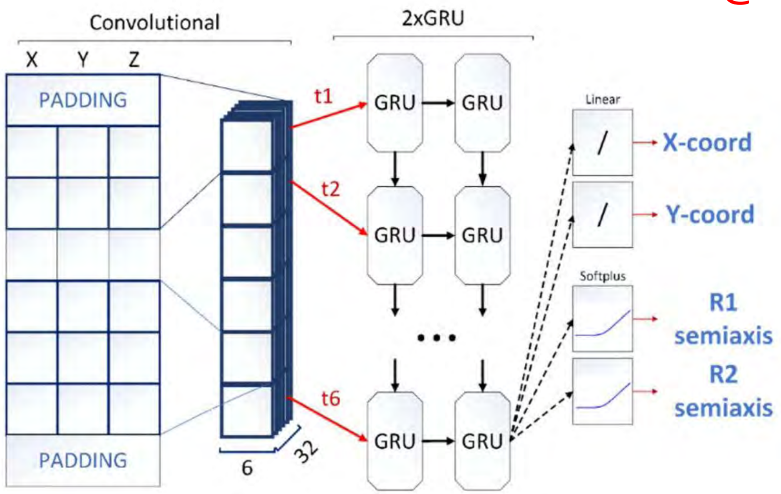


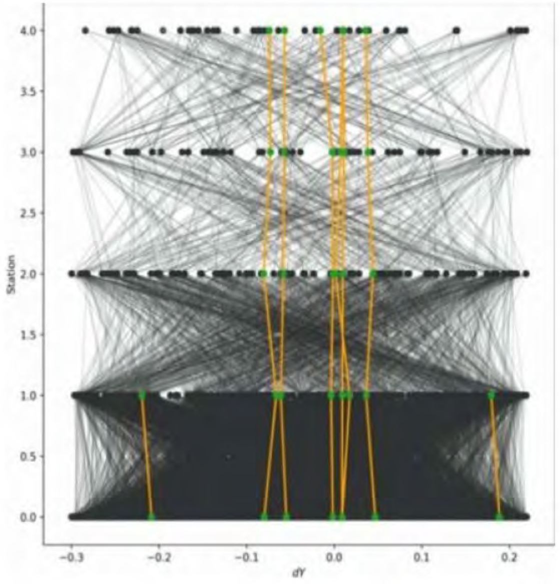
- new approaches to track recognition in SPD strip and pixel detectors based on a recurrent neural network and a graph network (already used for track recognition in the BM@N experiment at JINR and in the BESIII experiment in China)

MPD



BM@N





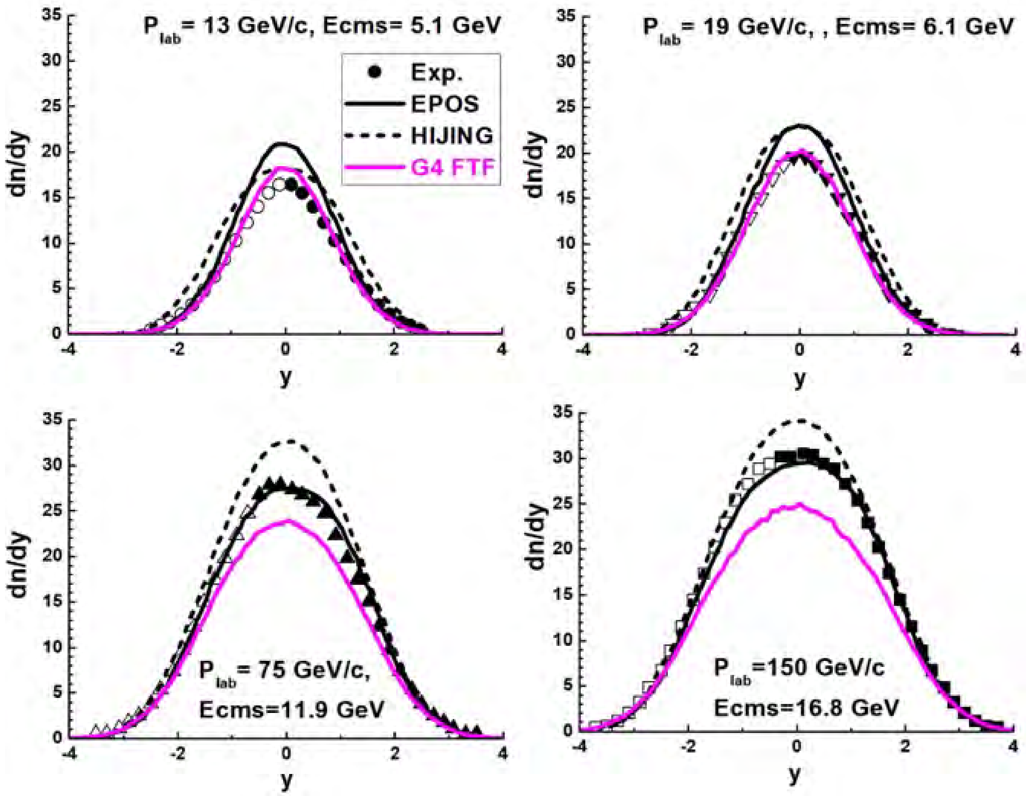
Preliminary results: accuracy of about 99% for testing data (18% of true segments are lost)

Efficiency and contamination of the identification of positively charged particles in the MPD model data, obtained by the method of decision trees with gradient boosting

Geant4 hadronic models (QGS and FTF) for HEP and cosmic ray experiments

GEANT4
A SIMULATION TOOLKIT

Rapidity distributions of π^- mesons in
Ar-40 + Sc-45 interactions (EPJ , C82 (2022)



Development, verification, validation and application

Key data: NA61/SHINE Collab. on PP, **Ar-40 + Sc-45** and Be-7 + Be-9 interactions

Exp. conclusion: "There is no model (EPOS, UrQMD, HIJING …) able to describe the data!"

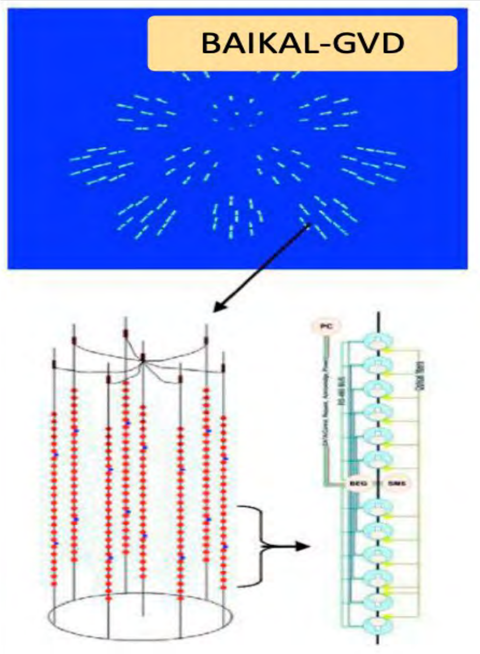
At the moment we have shown: Geant4 FTF model without QGP gives the best description of ^{40}Ar+^{45}Sc data at c.m.s. energy below 10 GeV

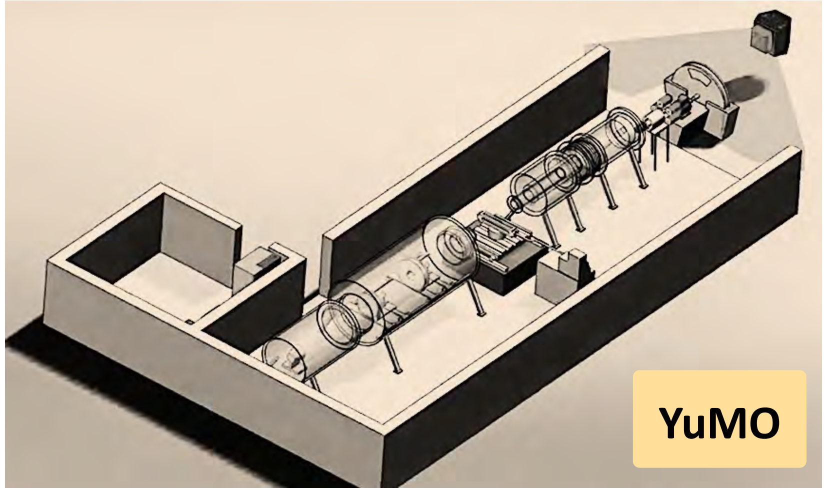
Tasks at hand:

1. Solve the problem with strange meson production;
2. Analyze BM@N data;
3. Simulate Au+Au interactions for NICA MPD experiment;
4. Simulate pp and dd interactions for NICA SPD experiment;
5. Estimate centrality detector responses;
6. Include spin in MC generators;
7. Include QGP a'la EPOS model;
8. Re-thinking of anti-nuclei interactions with nuclei;
9. Include structure of (anti) hyper-nuclei.
10. etc.

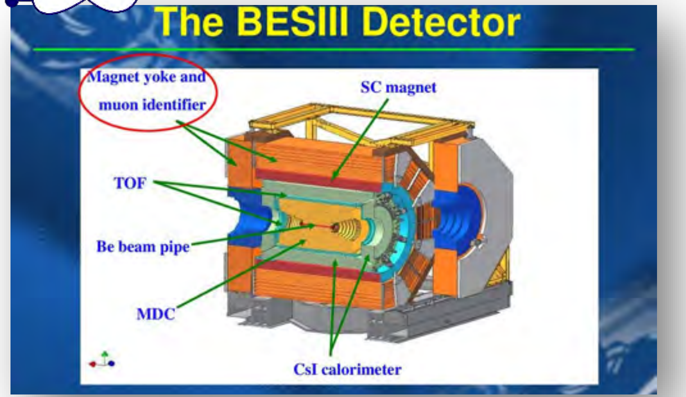
… and other fields

- data processing system for the BAIKAL-GVD project
- reconstruction of cosmic rays at the TAIGA gamma-ray observatory
  - several atmospheric Cherenkov telescopes (IACT) with a network of TAIGA-HiSCORE wide-aperture Cherenkov optical detectors
- track reconstruction in the proton digital calorimeter for proton therapy
- processing and analysis of neutron noise of the IBR-2M reactor
- software complex for data processing of the YuMO small-angle neutron scattering spectrometer

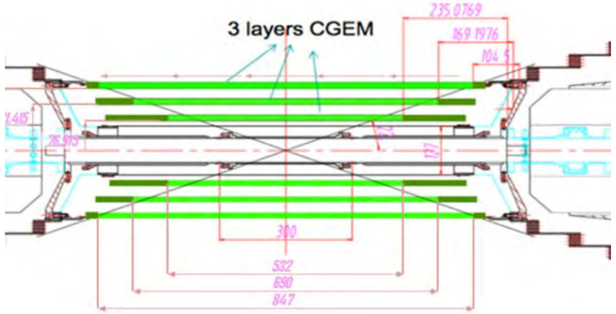




Tracking for data from high luminosity experiments. GNN for BES-III

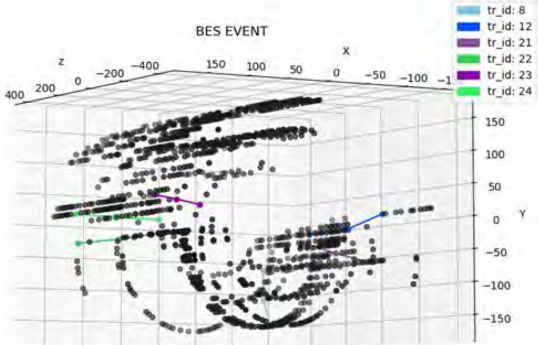


collider BESIII experiment



The CGEM-IT internal detector of the collider BESIII experiment, consisting of three detection cylinders

The presence of fakes and missed hits necessitated the use of a different type of GNN



All hits of a simulated event

The event graph is inverted into a linear digraph when edges are represented by nodes and nodes of the original graph are represented by edges. In this case, information about the curvature of track segments is embedded in the edges of the graph, making it easier to recognize tracks in a sea of fakes and noise. **In the process of training, the network receives as input an inverse digraph with labels of true edges - real track segments. The already trained GraphNet neural network as a result associates each edge with the value *x* ϵ *[0,1]* in the output. True path edges are those edges for which *x* is greater than some given threshold *(> 0.5)*. (http://ceur-ws.org/Vol-2507/280-284-paper-50.pdf)**

Tracking efficiency estimates. Evaluation of **accuracy** as a share of found tracks to the total number of candidate tracks is useless and even dangerous, because our sample is very unbalanced. It is accepted to use two metrics - **recall and precision. Recall** is the fraction of true tracks that the model was able to correctly reconstruct by finding all its hits. **Precision** is the fraction of true tracks among those that the model reconstructed

GraphNet	recall	precision
BES-III	96.23	90.64

Методы глубокого отслеживания для реконструкции событий в GEM детекторе эксперимента BM@N

Один из подходов представляет собой двухэтапный метод с предварительной обработкой данных путем направленного поиска в дереве k-d для поиска всех возможных кандидатов в треки, а затем с использованием глубокой рекуррентной нейронной сети для их классификации

Другой сквозной метод использует глубокую рекуррентную нейронную сеть для экстраполяции исходных треков, подобно фильтру Калмана, который извлекает необходимые параметры из данных.

Третий метод реализует нашу новую попытку адаптировать подход основанный на графовых нейронных сетях (GNN), разработанный в проекте HEP.TrkX в ЦЕРН, к данным, специфичным для GEM

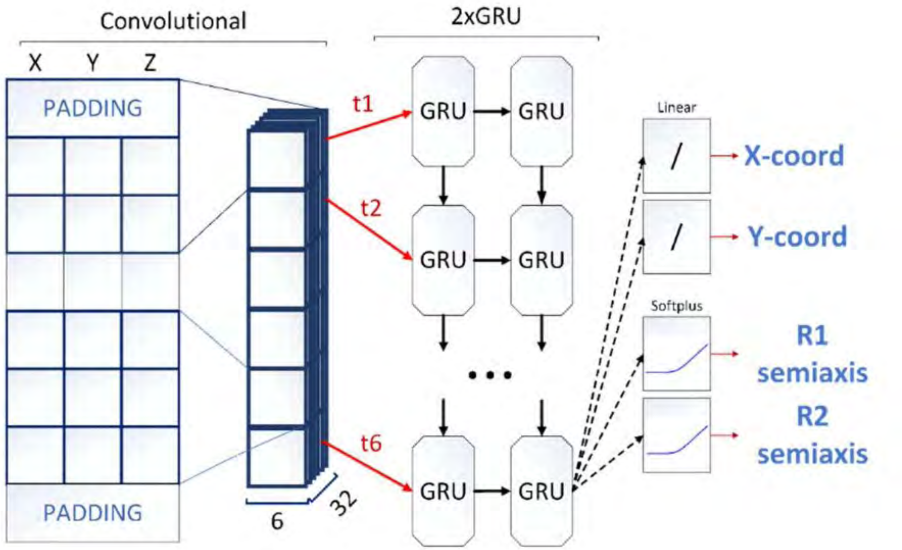
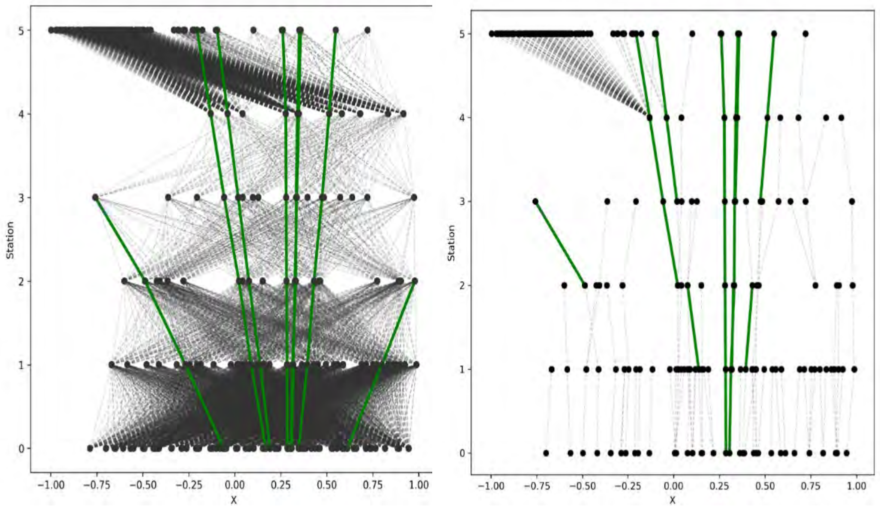


Схема сквозного обучения, основанная на глубоких рекуррентных сетях. Схема реализована в пакете TrackNETv2. Метод дает эффективность реконструкции порядка 99.2%.

Инициирование графа для GNN



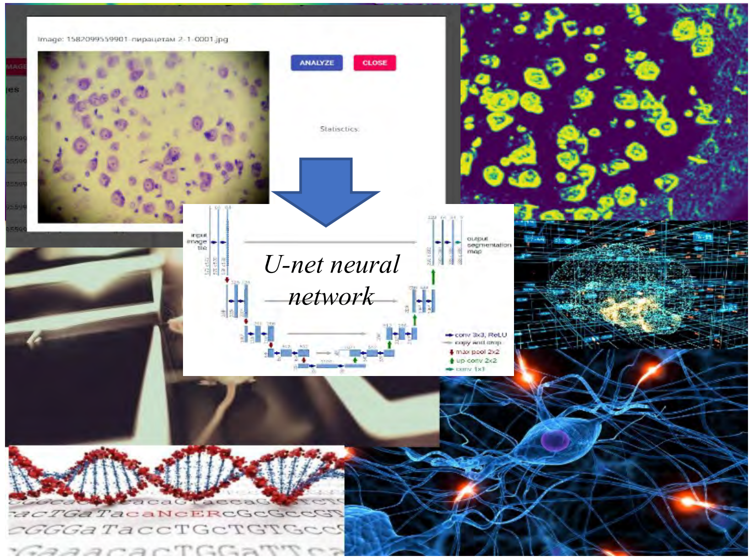
Предварительные результаты алгоритма минимального дерева ветвления:

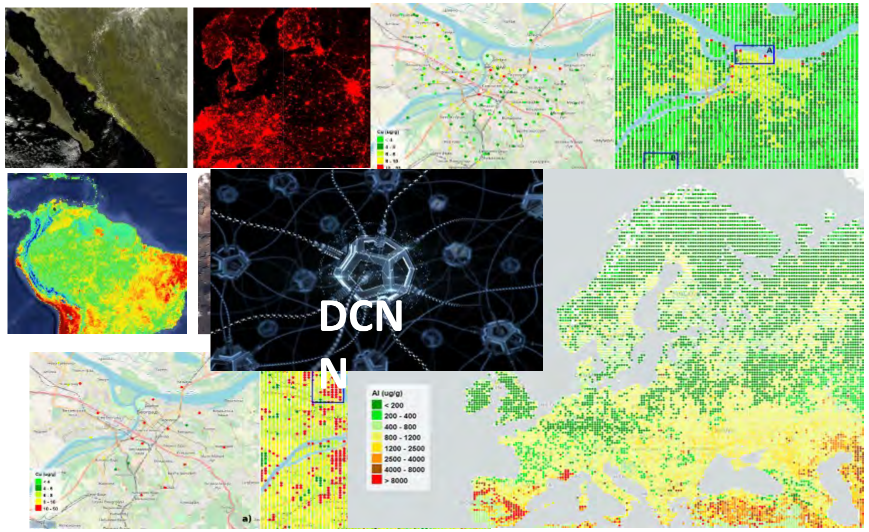
точность – 99% на тестовых данных (18% истинных сегментов были потеряны во всем наборе данных)

Goncharov et al. AIP Conf. Proc. 2163, 040003 (2019)
Goncharov et al. EPJ WoC 226, 03009 (2020)
Rusov et al. Studies in Comp. Intell. 1064, 32 (2022)

ИИ для задач ОИЯИ

I. Совместный проект ЛИТ и ЛРБ по созданию Информационной системы (ИС) как комплекса ИТ-решений, обеспечивающих хранение, анализ и визуализацию данных радиобиологических экспериментов. ИС основана на методах машинного и глубокого обучения и нейросетевых подходах.





II. В рамках сотрудничества ЛИТ и ЛНФ ведутся работы по прогнозированию загрязнения воздуха тяжелыми металлами с использованием данных биологического мониторинга, данных с космоснимков и различных технологий машинного и глубокого обучения.

III. Нейронные сети для решения задач реконструкции траекторий движения частиц в экспериментах ФВЭ (BM@N, BESIII, SPD et al.).