PWG2 Summary

V. Kireyeu* on behalf of the group

* vkireyeu@jinr.ru

MPD Collaboration meeting, JINR, Dubna, October 20-22, 2026

- 5 group analyses were included to the 3rd MPD Collaboration paper
- However, the work is going on:
- STAR-like PID and light hadrons results
- MAD PID (with ML) and light nuclei results
- Big update on resonances studies
- Byproduct bonus of ‘nuclei studies’
- Summary

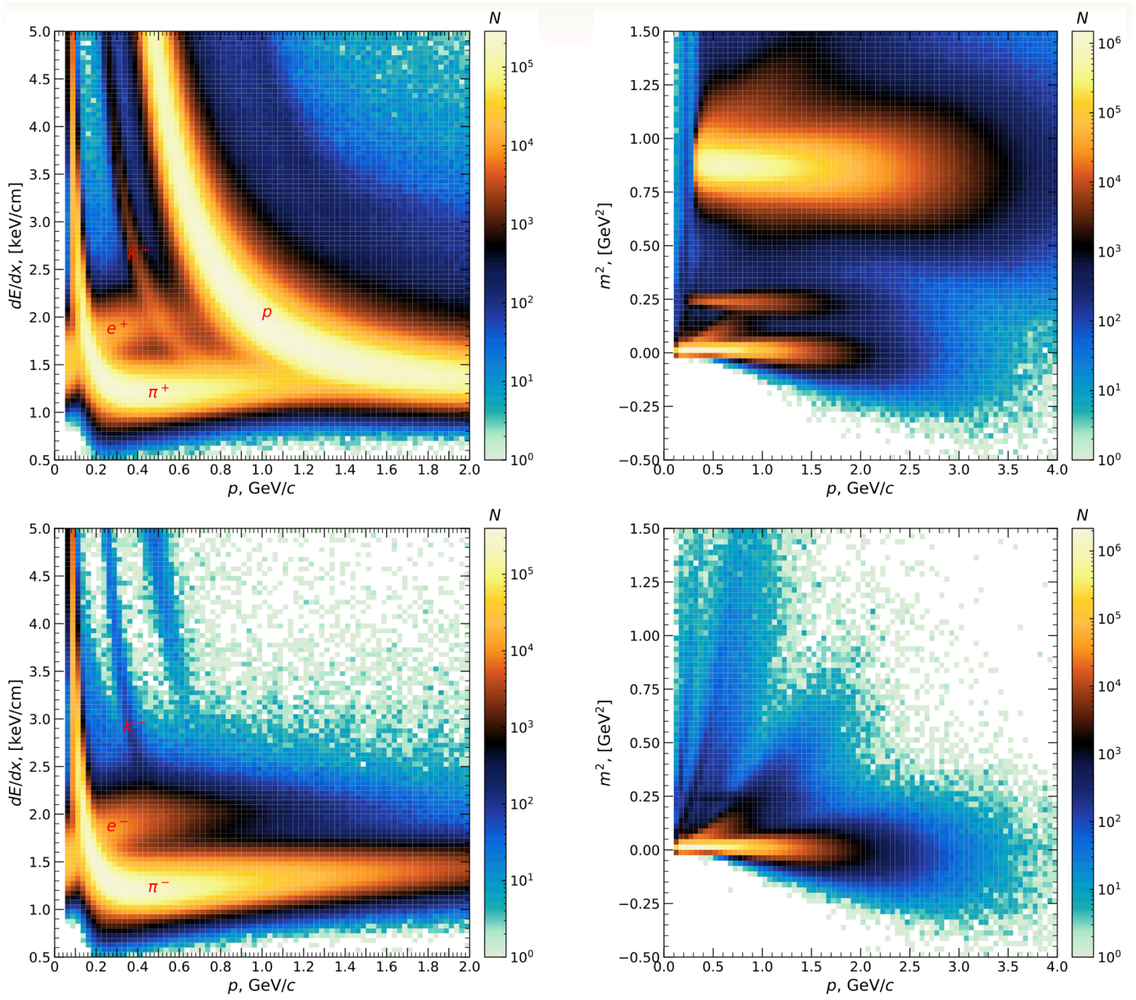
Light hadrons, realistic PID

Natalia Kolomoyets

Basic idea of the method: use 1D sections of dE/dx and m2 distributions to estimate the amount of particles of a given sort in every cell of the phase space.

General features:

- model-independent method;
- offline method;
- integral method (it doesn't work with separate tracks, but with the whole set ofthem)



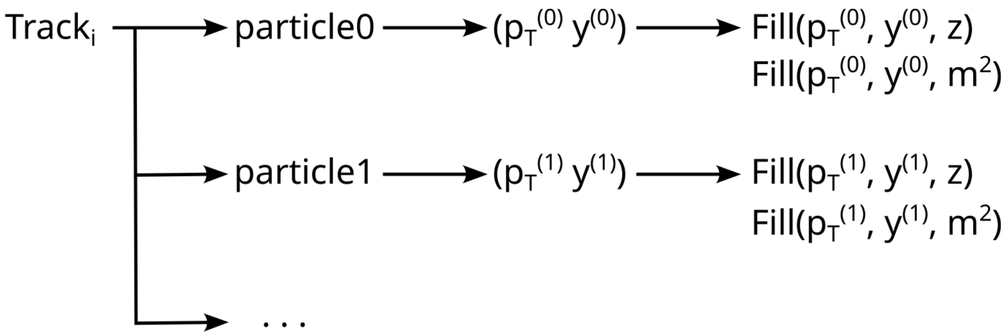
The implementation is based on: STAR, PRC 79 (2009), 034909, STAR, PRC 96 (2017), 044904, STAR, PRC 101 (2020), 024905

Cross-PWG presentation: https://indico.jinr.ru/event/6353/contributions/37667/attachments/26437/47257/Meeting-11.1.pdf

Light hadrons, realistic PID

Natalia Kolomoyets

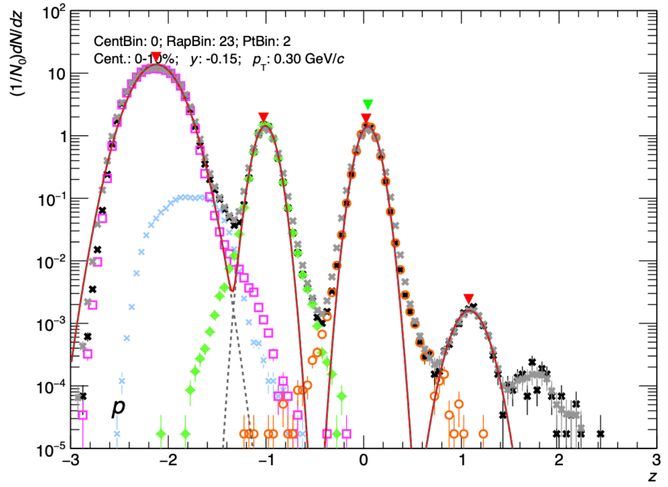
Each track contributes to the phase space of the each particle of interest:

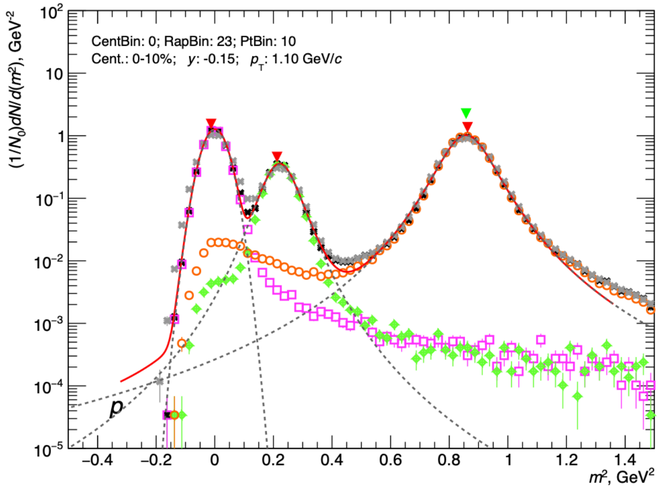


with $z = ln\dfrac{dE/dx}{(dE/dx)_{BB}}$ for $z = 0$ peak

Then for each phase space cell:

Result = number of particles in selected cell = area under corresponding peak





Light hadrons, realistic PID

Natalia Kolomoyets

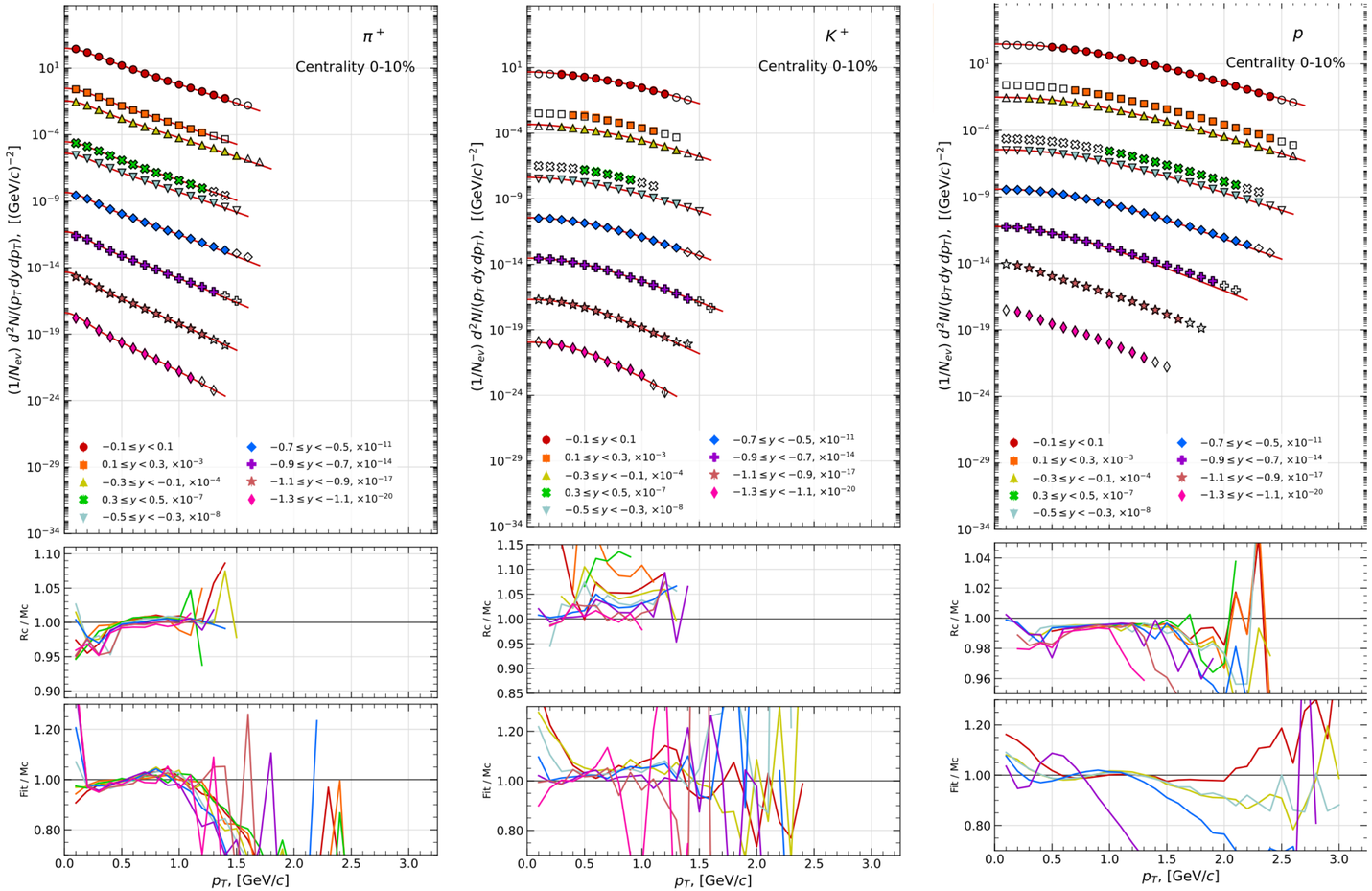
Model-independent method

The procedure does not depend on the colliding system and energy

Possible optimizations

Presented on Cross-PWG

Next stop: ICPPA-2026



Light nuclei, Data-driven PID (MAD PID)

Viktar Kireyeu

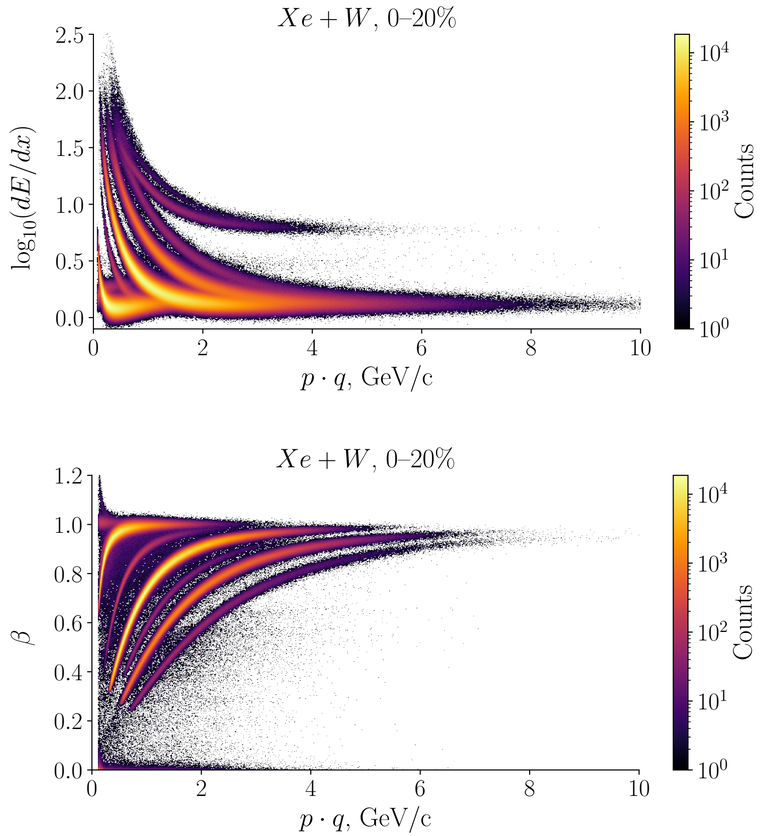
The method converts an approximately isolated response band into data-driven PID boundaries. The selected tracks are split into narrow pq bins, with at least 100 tracks per bin.

TPC: $x_k = \log_{10}(dE/dx)$ TOF: $x_k = \beta_k$

$$\mu_j = \mathrm{median}_k(x_k)$$

$$\sigma_j^{\mathrm{lower}} = 1.4826\ \mathrm{median}_{x_k<\mu_j}(\mu_j - x_k)$$

$$\sigma_j^{\mathrm{upper}} = 1.4826\ \mathrm{median}_{x_k>\mu_j}(x_k - \mu_j)$$

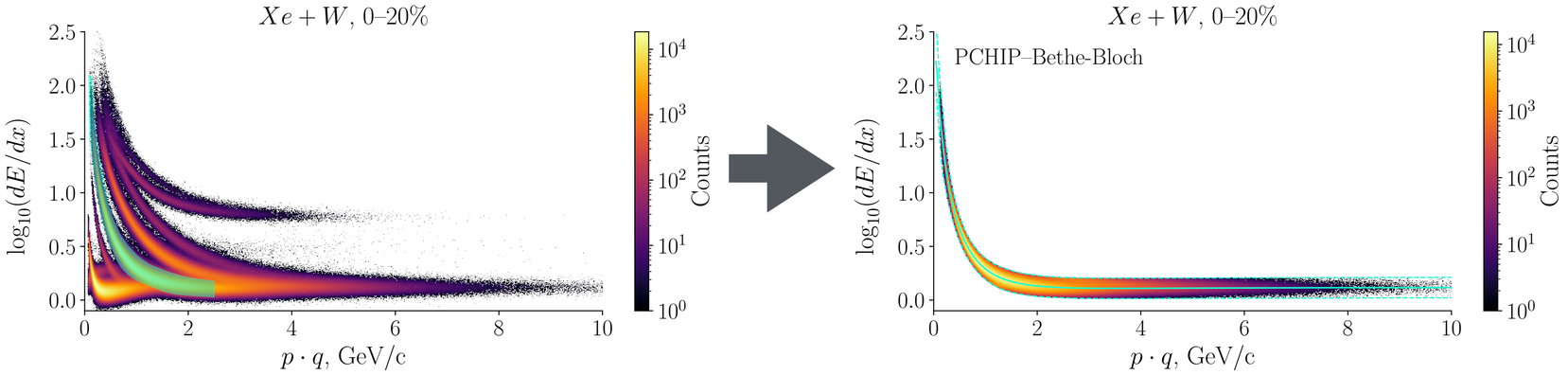


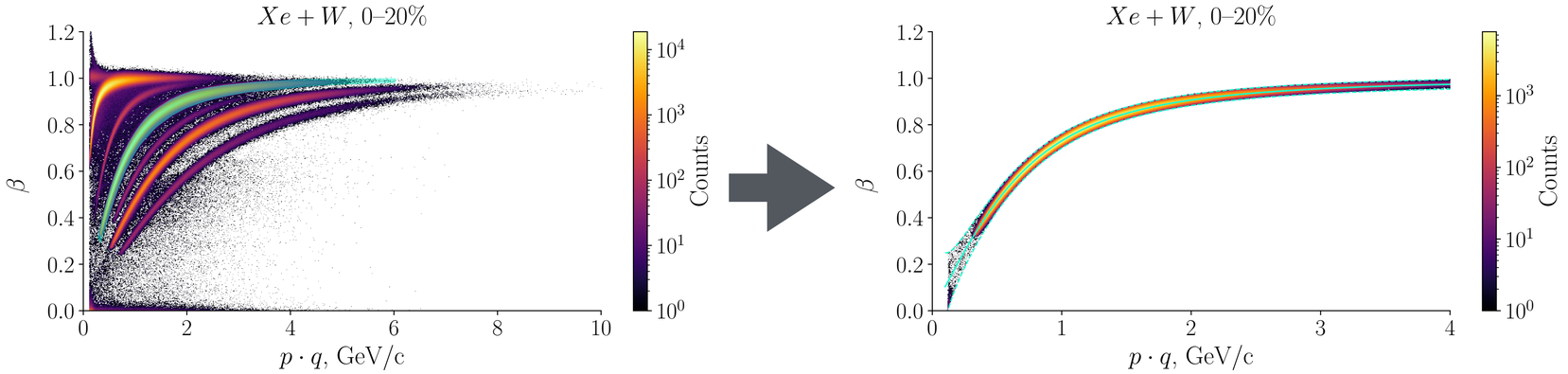
- Breakdown points: 50% for the median and 25% for the one-sided MAD widths; a single outlier can drive the mean and standard deviation without bound
- Separate lower and upper widths account for asymmetric response bands
- No response model, template or fitted signal component

Cross-PWG presentation: https://indico.jinr.ru/event/6302/contributions/37106/attachments/26247/46916/main.pdf

Light nuclei, MAD PID

Viktar Kireyeu





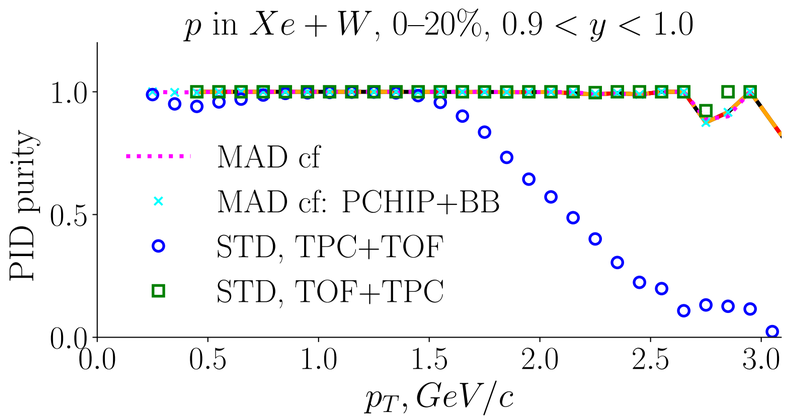
Truth-blind extraction reproduces the MC response

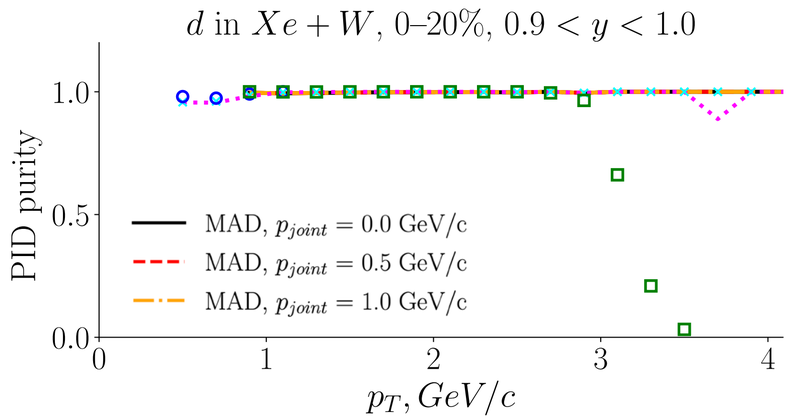
Species	Response	$\mathrm{med}_j(\lvert\Delta\mu_j\rvert/\sigma_{\mathrm{truth},j})$	$\mathrm{med}_j(C_{3,j})$
p	dE/dx	0.021	99.4%
p	β	0.129	98.7%
d	dE/dx	0.089	98.2%
d	β	0.048	98.2%
t	dE/dx	0.024	98.2%
t	β	0.033	97.9%
^{3}He	dE/dx	0.053	95.0%
^{3}He	β	0.079	98.2%

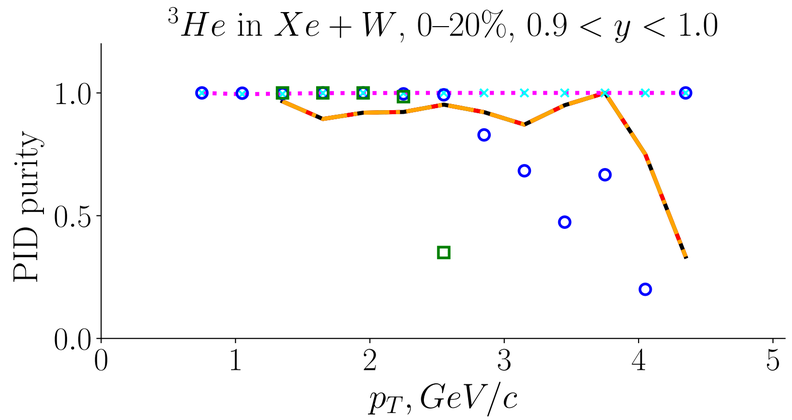
Light nuclei, MAD PID

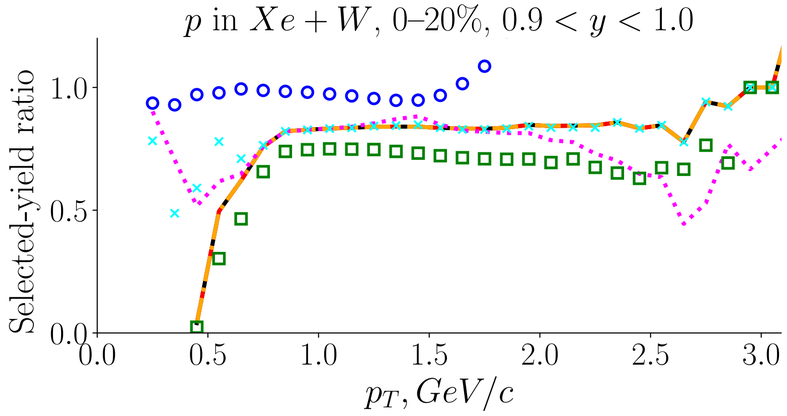
Viktar Kireyeu

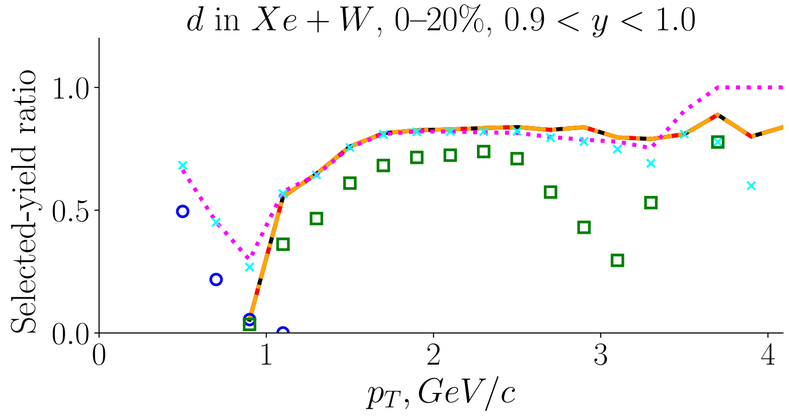
$$R_i^{\text{sel}} = \frac{N_i^{\text{selected}}}{N_i^{\text{reco,true}}};\ \text{Purity}_i = \frac{N_i^{\text{selected,true}}}{N_i^{\text{selected}}}$$

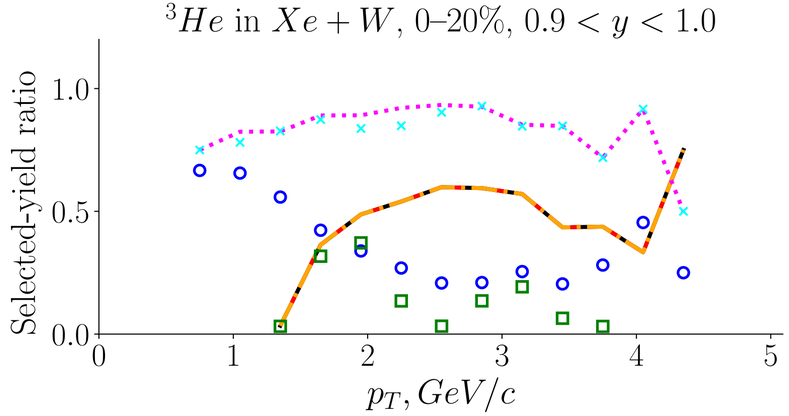










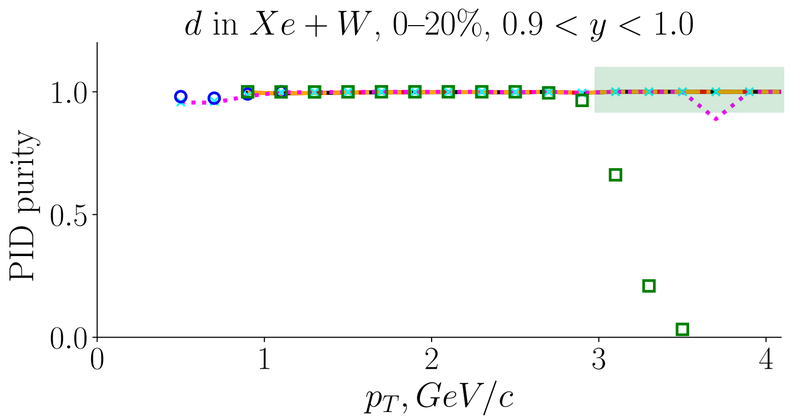


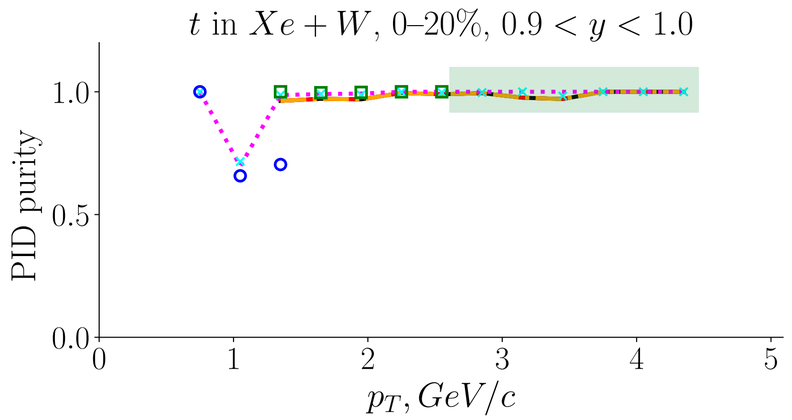
Method allows for the extension of measured region by ~2 GeV/c compared to standard procedures

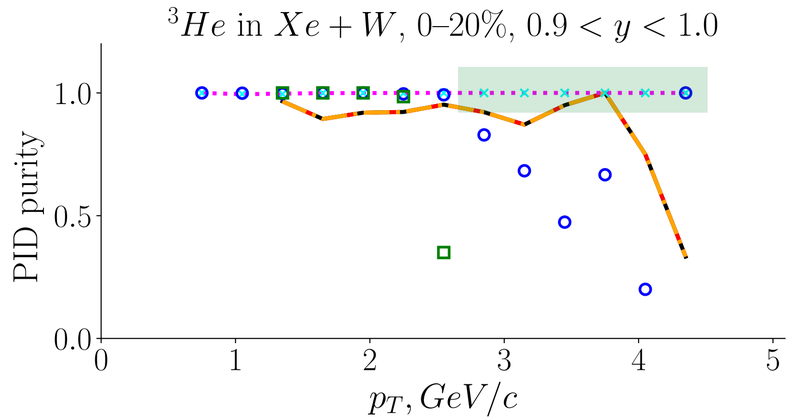
Light nuclei, MAD PID

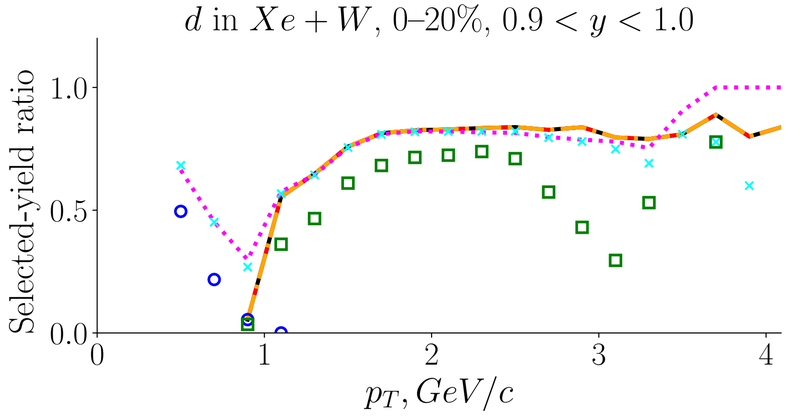
$$R_i^{\text{sel}} = \frac{N_i^{\text{selected}}}{N_i^{\text{reco,true}}};\ \text{Purity}_i = \frac{N_i^{\text{selected,true}}}{N_i^{\text{selected}}}$$

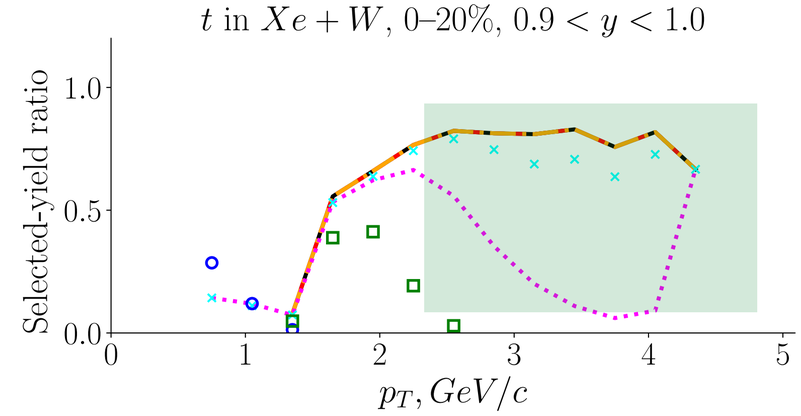
Viktar Kireyeu

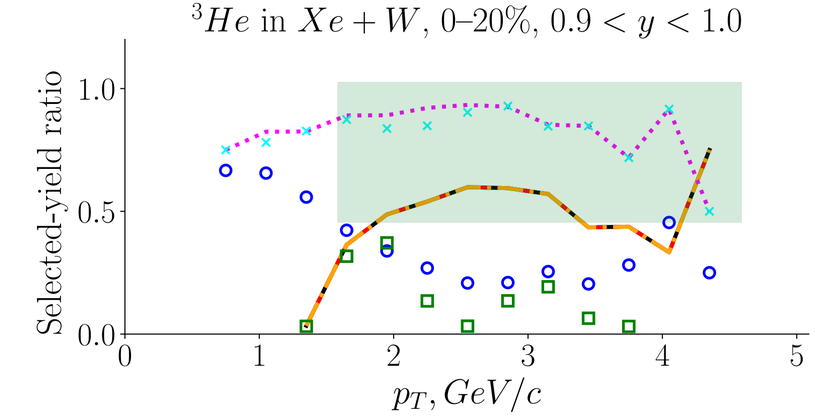












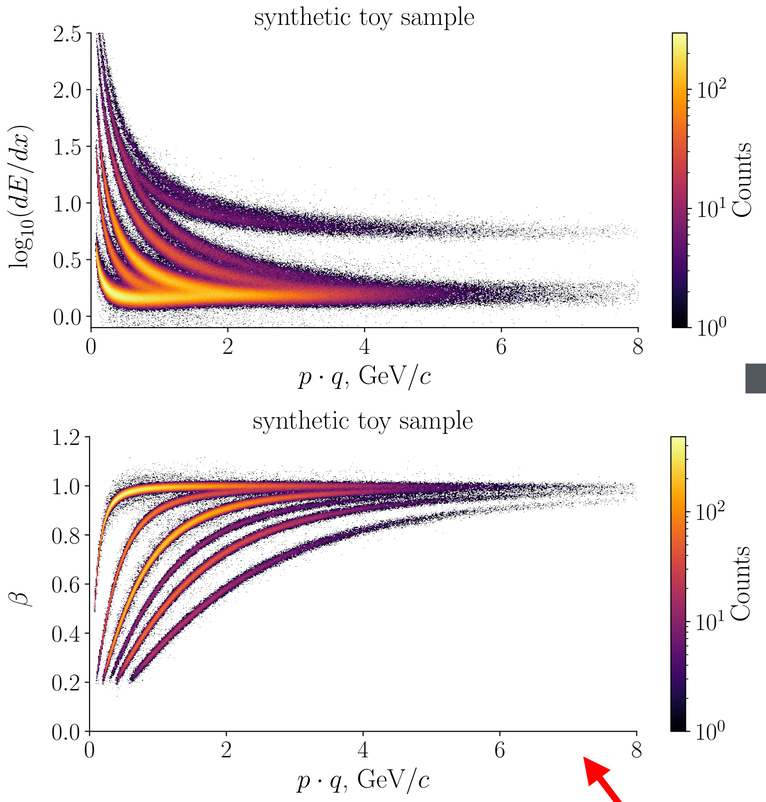
Method allows for the extension of measured region by ~2 GeV compared to standard procedures

Light nuclei, MAD PID

Viktar Kireyeu

Supervised Machine Learning:

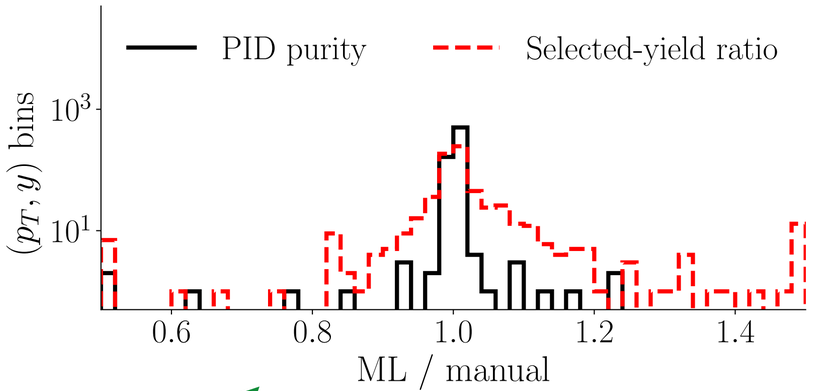
Do not make per-track ML — pre-select bands — run MAD PID



Two independent binary classifiers

for TPC and TOF:

HistGradientBoostingClassifier

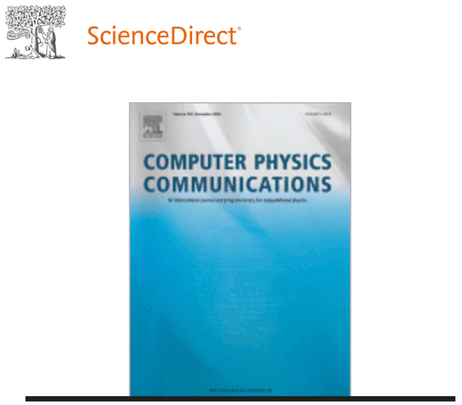


The model is trained on the synthetic toy sample and applied to the reconstructed data

Light nuclei, MAD PID

Viktar Kireyeu

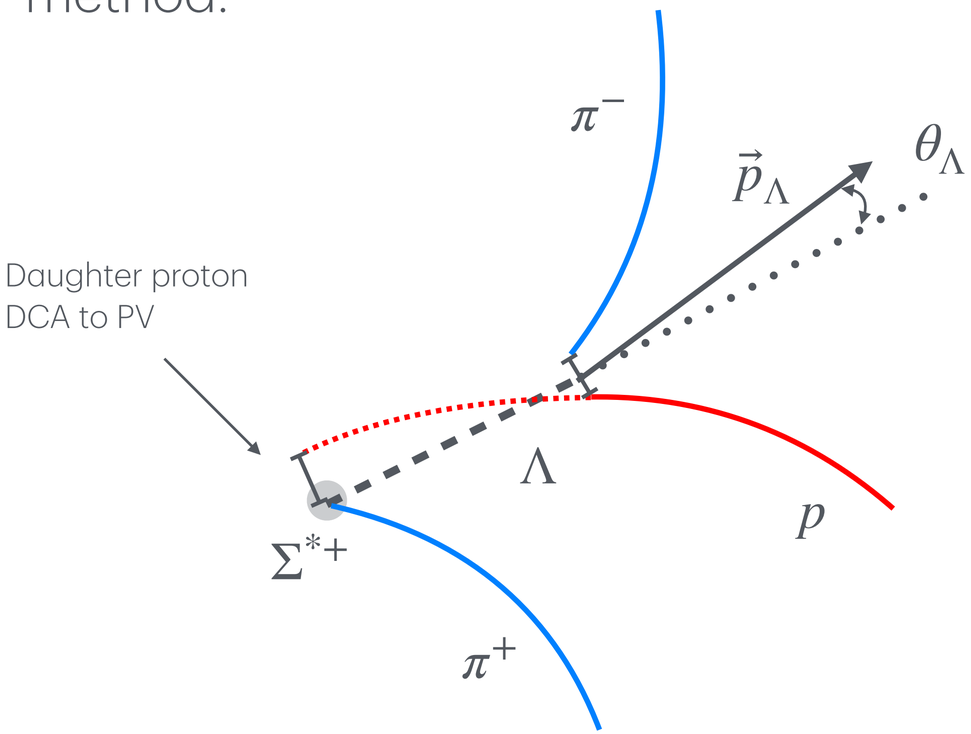
- Presented at Cross-PWG
- Published in CPC
- Fully reproducible chain available on GitHub under MIT license:
  - code, interface from ROOT to DataFrame,
    toy-sample, ML training macro, ML model



Update on resonances

Mikhail Malaev

Short-lived resonances are reconstructed via their hadronic decay channels using the invariant mass method.



$$\phi(1020) \rightarrow K^+K^- \qquad (c\tau \sim 45 \text{ fm}/c)$$
$$K^*(892)^0 \rightarrow \pi^\pm K^\pm \qquad (c\tau \sim 4.2 \text{ fm}/c)$$
$$\rho(770)^0 \rightarrow \pi^+\pi^- \qquad (c\tau \sim 1.3 \text{ fm}/c)$$

- Possible to measure due to large branching ratios in hadronic decay channels
- Probe of reaction dynamics: hadron chemistry, strangeness production, lifetime and properties of hadronic phase etc.

Cross-PWG presentation: https://indico.jinr.ru/event/6425/contributions/37952/attachments/26640/47628/Resonance_MMalaev.pdf

Update on resonances

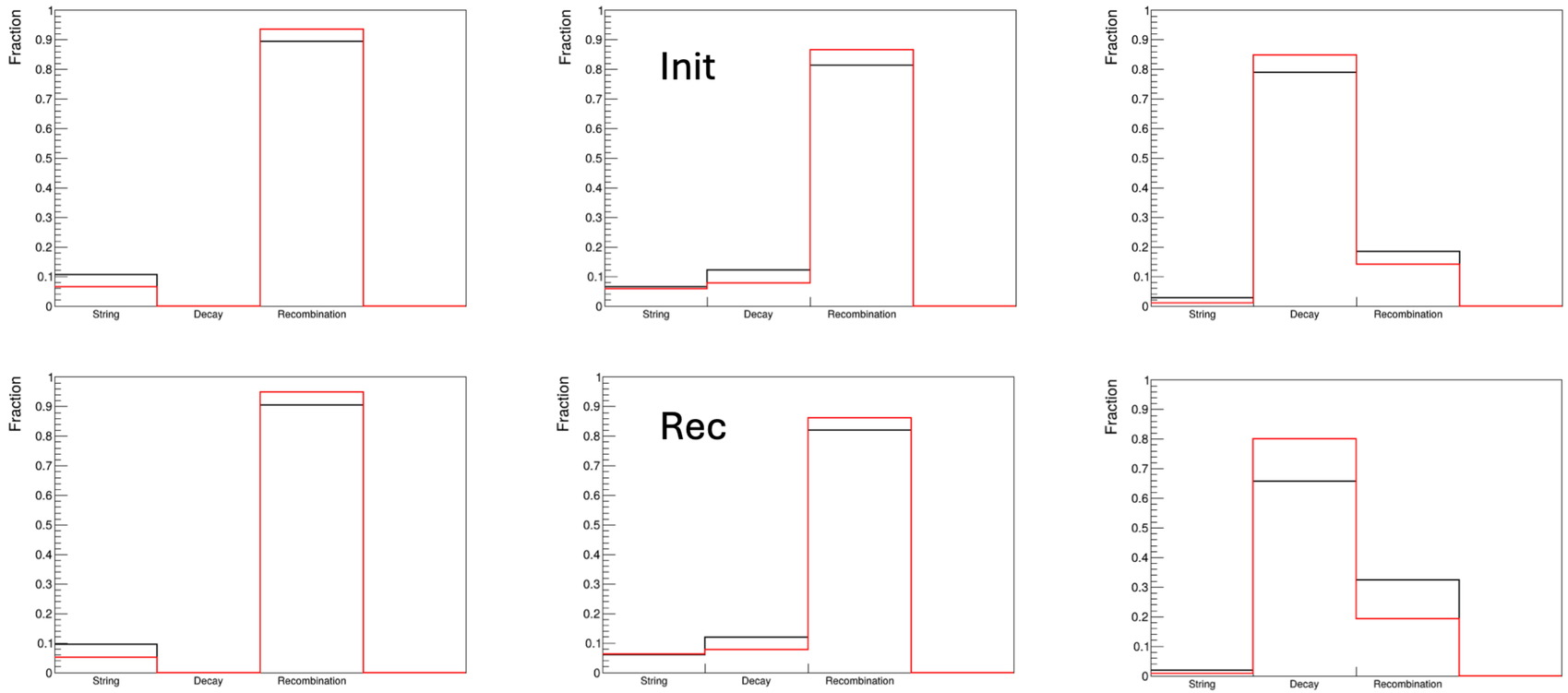
Mikhail Malaev

Init = reconstructed from daughter particles at decay moment
Rec = reconstructed at the end of evolution after all interactions

Production mechanisms
(UrQMD model used):

- String decay
- Decay of heavier particles
- Recombination of hadrons

ϕ(1020) K*(892)^0 ρ(770)

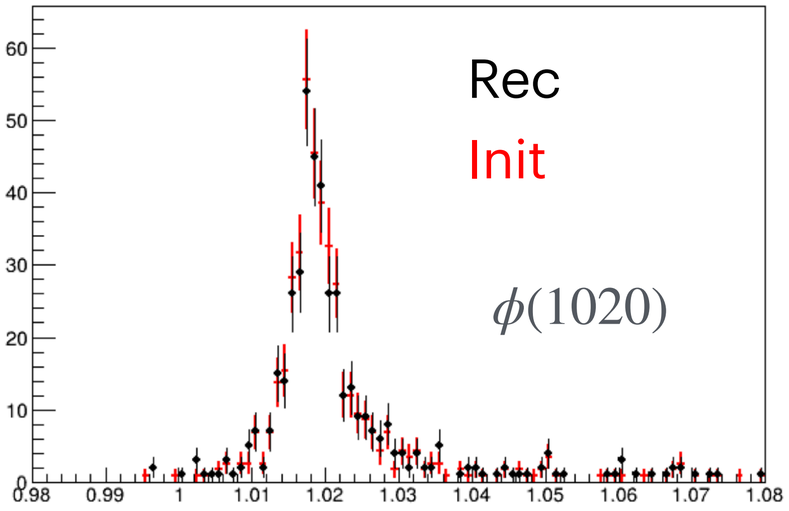


- Recombination is dominant $\phi(1020)$and $K^*(892)^0$
- Decays is dominant for $\rho(770)^0$

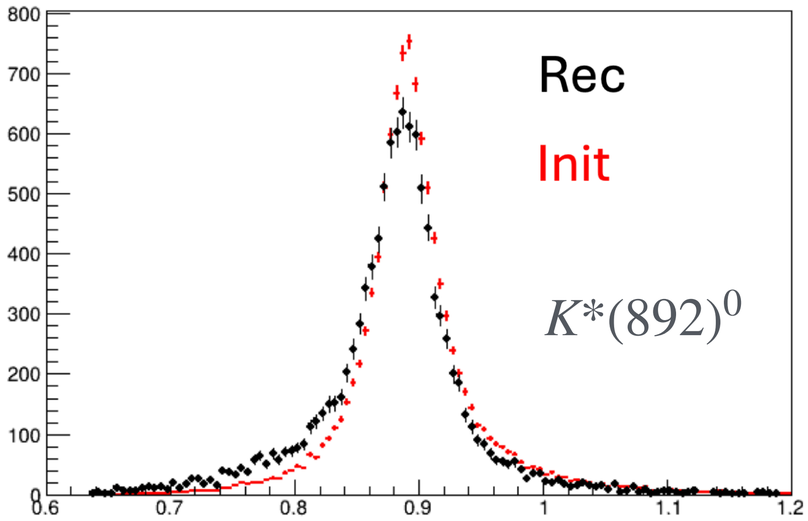
Red lines = low multiplicity events
Black lines = high multiplicity events

Update on resonances

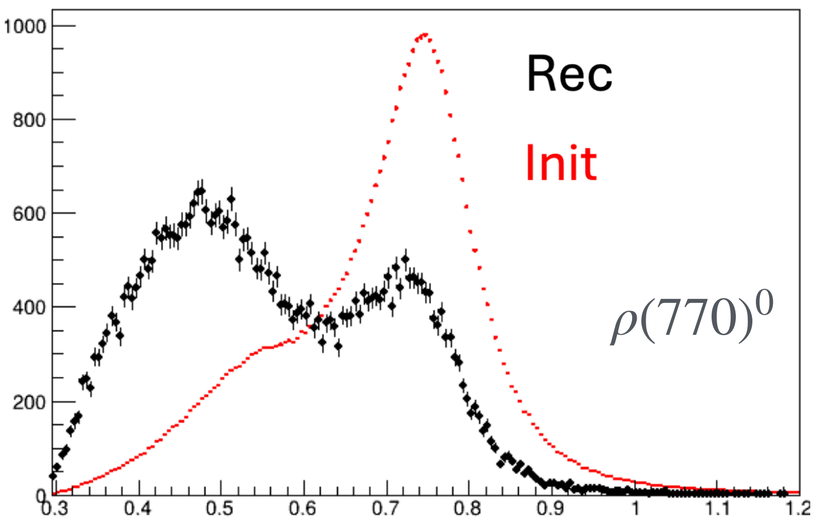
Mikhail Malaev



$0 < p_T < 0.5$ GeV/c

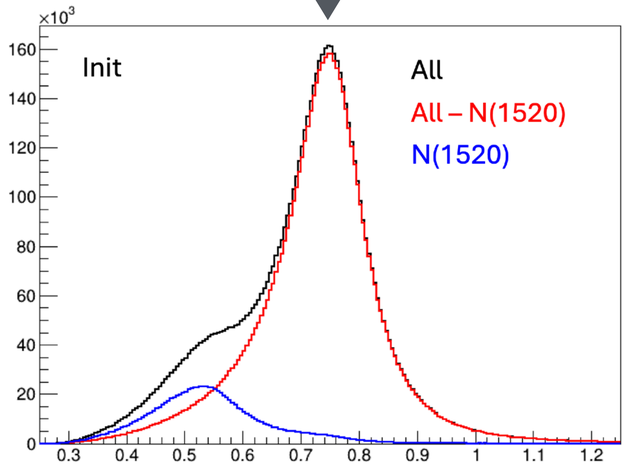


$0 < p_T < 0.5$ GeV/c



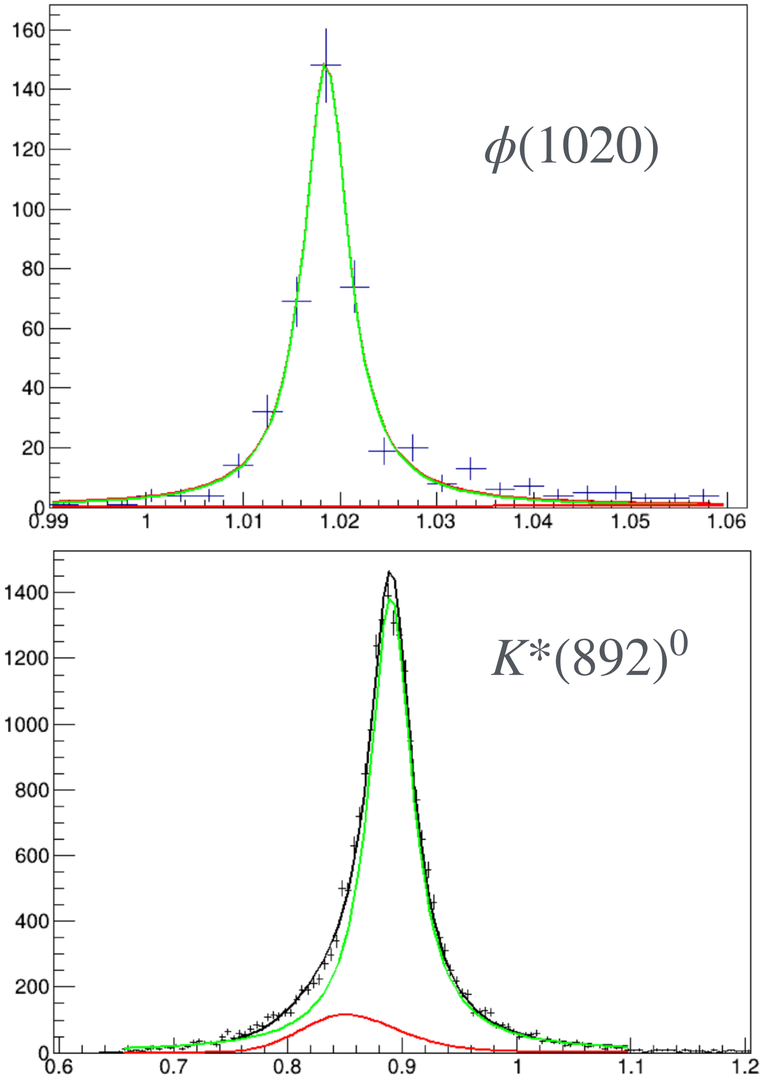
Peak shape modification:

- Modest to no for $K^*(892)^0$ and $\phi(1020)$
- Significant for $\rho(770)^0$ at lower p_T — can be explained by $N(1520)$ decay channel
- After exclusion of the $N(1520)$ — $\rho(770)^0$ peak shape form is back to expected



Update on resonances

Mikhail Malaev



Fitting procedure like in real data analysis for Rec distributions: signal + background:

- No peak shape modifications for $\phi(1020)$
- Modest peak shape modifications in high multiplicity events for $K^*(892)^0$

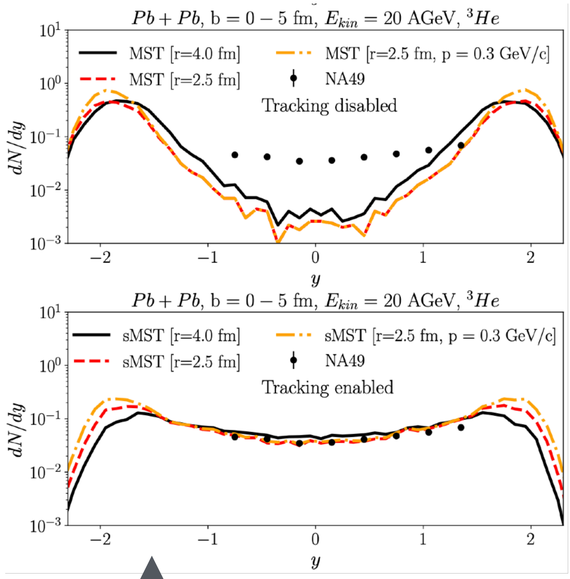
Baryonic resonances studies started with $\Delta(1232)^{++}$

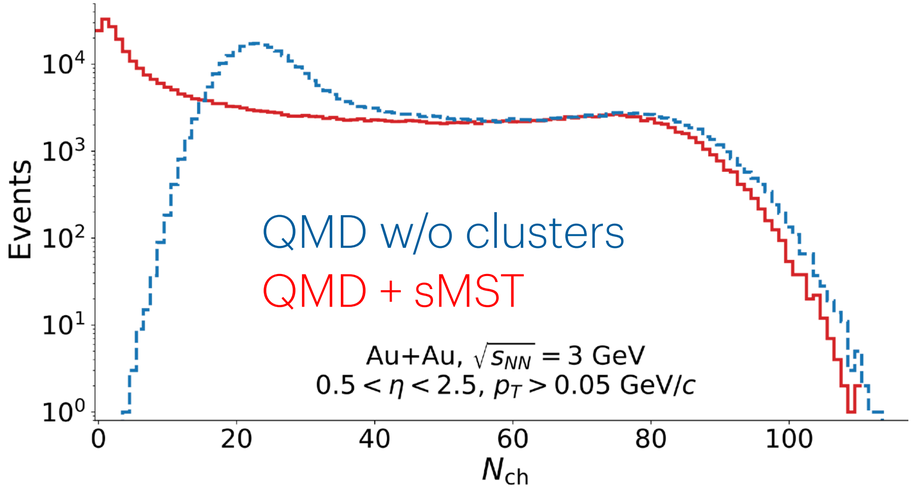
Plans to use PHSD transport approach for similar studies

Results presented on Cross-PWG

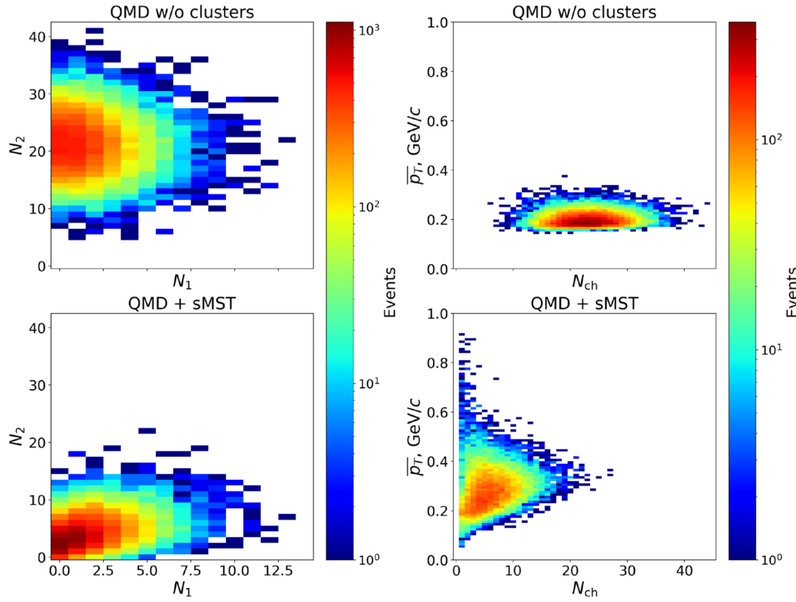
Next stops: ICPPA-2026, publication

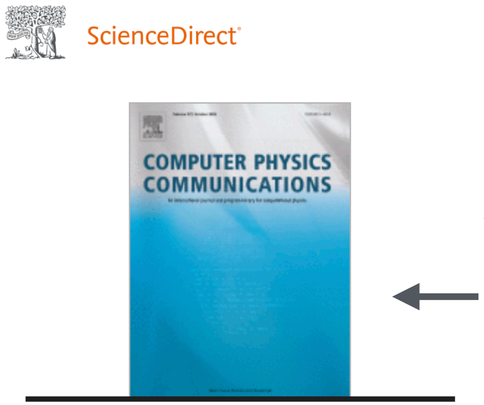
Byproduct bonus





12.5 < b < 13 fm





Computer Physics Communications

Date: October 2026
Article: 110278
Volume: Volume 327

View accessibility information

Published by

ELSEVIER

CCL sMST allows to add clusters to QMD-based models and thus make them available for different technical tasks and analyses **(if one is interested)**

Single baryon observables are sensitive to presence or absence of clusters in the models

Phys.Rev.C 113, 064909 (2026)

N_1: charged particles in $0.5 < \eta < 1.5$

N_2: charged particles in $1.5 \leq \eta < 2.5$

$N_{ch} = N_1 + N_2$

Summary

- 2 new PID methods
- One of them already has ML part, second is ML-ready (e.g. signal analysis for ‘worst’ regions)
- Nuclei spectra can be extended by extra 2 GeV/c (compared to standard procedures)
- Resonances analysis is actively updated

Backup

Cost of ML per-track classifier

- The training sample. Retrain on a different subsample and the same track gets a different probability. The spread is a genuine uncertainty and can only be measured by retraining an ensemble
- The training configuration. Random seed, train/test split, hyperparameters and the choice of input features change the result in the same way, and are rarely varied
- Trained on simulation, applied to data. Any imperfection in the detector description biases every track, silently: the classifier exploits fine correlations between many variables, including the badly modelled ones
- The composition. The classifier learns the particle mix of its training sample; applied where the mix differs, its probabilities are miscalibrated even with a perfect detector description
- Nothing to look at. There is no boundary to inspect and no polygon to correct, so a mismodelled region is not found by eye

The methods to quantify this were published years ago: ensembles, Poisson bootstrap, retraining spread. No PID analysis uses it.

ML MAD PID

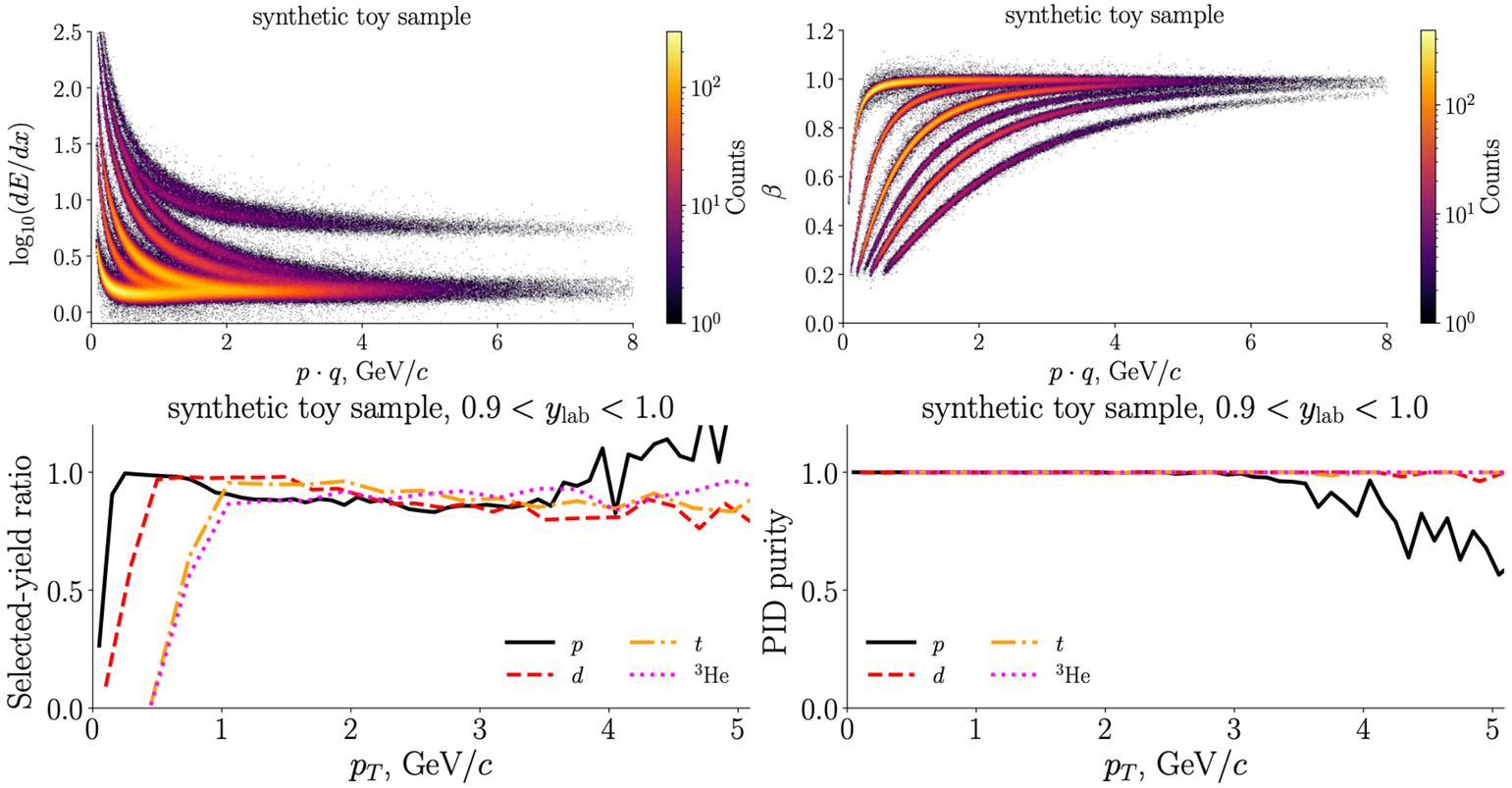


Figure C.1: Public synthetic example distributed with the code. Top: inclusive TPC and TOF response maps for positively charged tracks. Bottom: selected-yield ratio and purity obtained with the PCHIP+BB combined+fallback selection for protons, deuterons, tritons, and helium-3 in $0.9 < y_{\mathrm{lab}} < 1.0$.

Data

Same task: select protons with $\beta > 0$, fill a p_T histogram, write nothing. $\sim$ 15.5M of $\sim$ 54.7M tracks selected, identical in every backend. ROOT tree 905 MB, Parquet 1038 MB, Parquet (sorted) 911 MB.

Backend	Threads	Time, s	M tracks/s	Slower than best
TTree, C++, 3 branches	1	2.29	23.9	×22.0
TTree, C++, all branches	1	4.11	13.3	×39.5
RDataFrame, C++	1	4.05	13.5	×38.9
RDataFrame, C++	10	1.48	37.0	×14.2
RDataFrame, PyROOT	1	4.76	11.5	×45.8
RDataFrame, PyROOT	10	1.94	28.2	×18.7
Parquet, PyArrow	10	0.74	74.0	×7.1
Parquet, Pandas	10	0.67	81.7	×6.4
Parquet, Polars	10	0.16	350.9	×1.5
Parquet, Polars, sorted	10	0.10	526.3	×1.0

When compiled

Backend	Format	Threads	Time, s	M tracks/s	Interpreted
TTree, 3 branches, compiled	TTree	1	2.22	24.6	2.29
TTreeProcessorMT, compiled	TTree	10	0.47	116.2	0.49
RDataFrame, compiled	TTree	10	0.81	67.5	1.48
RNTupleReader, compiled	RNTuple	1	1.057	51.8	2.35
RNTupleReader MT, compiled	RNTuple	10	0.416	131.6	0.515
Polars, compiled	columnar	10	0.075	729.8	0.10

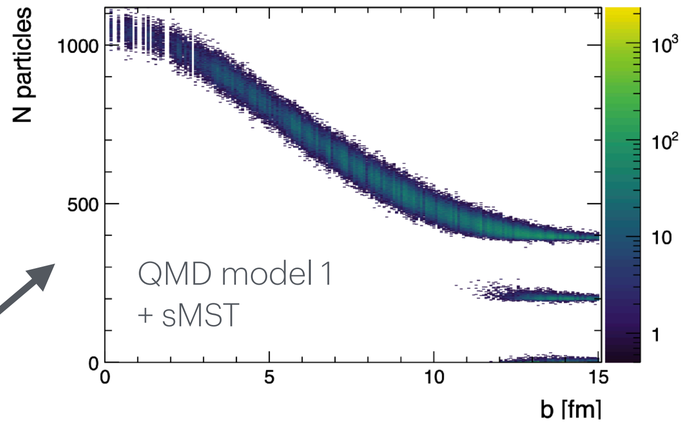
Same task on nested data: 1M events and 54.99M tracks. Select $|\eta| < 1$ protons with $\beta > 0$ at centrality $< 20\%$; fill p_T and event-multiplicity histograms.

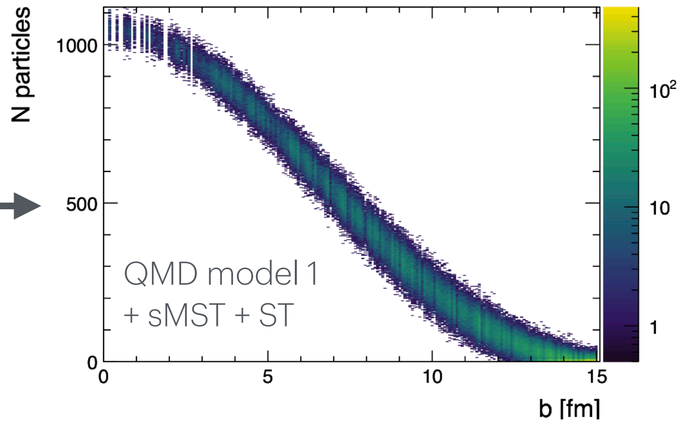
Backend	Format	Threads	Time, s	CPU, s	M tracks/s
TTreeReader, compiled	TTree	1	1.631	1.620	33.7
RDataFrame, compiled	TTree	10	0.793	2.360	69.4
TTreeProcessorMT, compiled	TTree	10	0.407	3.670	135.1
RNTupleReader, compiled	RNTuple	1	0.663	0.720	83.0
RNTupleReader MT, compiled	RNTuple	10	0.205	1.870	268.3
Polars, nested	Parquet	10	0.428	3.687	128.5
Polars, flat	Parquet	10	0.175	1.661	314.3
Polars, nested, sorted	Parquet	10	0.137	1.008	401.4
Polars, flat, sorted	Parquet	10	0.064	0.503	859.3

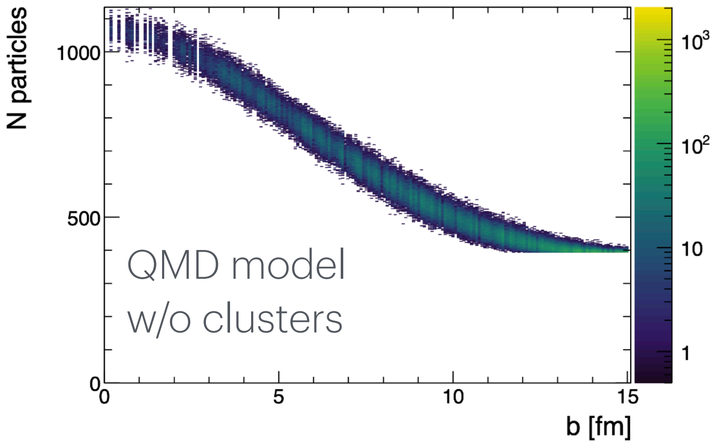
Spectators

ST = Spectator Treatment

In some models ‘splitting’ occurs when one or both spectator remnants fail the cluster selection and are counted as individual nucleons.







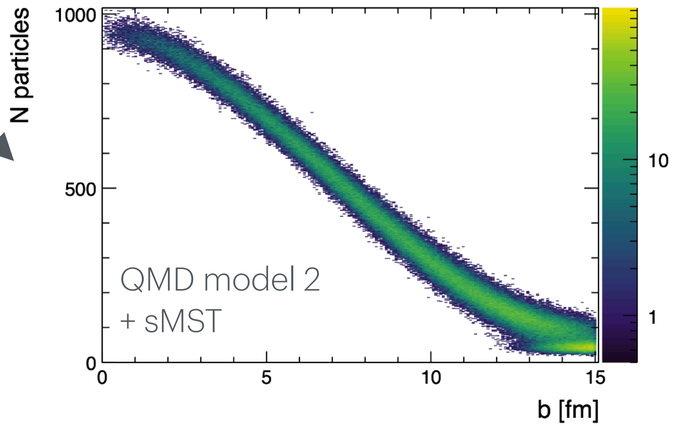
A Fermi-motion window around either beam defines spectator candidates:

$$\Delta y_F = \operatorname{arsinh}\left(\frac{p_F}{m_N c}\right) \simeq 0.315, \qquad p_F = 300\,\mathrm{MeV}/c$$

This gives the recoil-energy threshold: $\varepsilon_{\mathrm{cut}} = m_N c^2 \left[\cosh(\Delta y_F) - 1\right] \simeq 46.8\,\mathrm{MeV/baryon}$

The cluster kinetic energy per baryon in the nearer beam rest frame must satisfy:

$$\frac{K_{\mathrm{rel}}}{A} = \frac{\gamma_b \left(E_{\mathrm{cl}} - \beta_b c\,|p_{z,\mathrm{cl}}|\right) - M_{\mathrm{cl}} c^2}{A} < \varepsilon_{\mathrm{cut}}$$



‘Splitting’ is not universal across QMD-based models family

A is the cluster baryon number, M_{cl} its rest mass, and m_N the nucleon mass. E_{cl} and $p_{z,\mathrm{cl}}$ are measured in the NN CM frame. The beam speed is $\beta_b c$, with $\gamma_b = 1/\sqrt{1-\beta_b^2}$.

Protons

Phys.Rev.C 113, 064909 (2026)

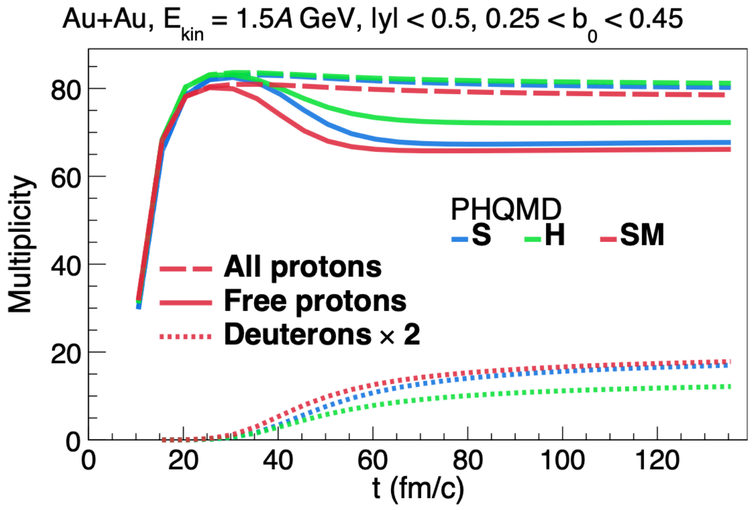


FIG. 5. Time evolution of the midrapidity multiplicities of the protons and deuterons for Au+Au collisions at $E_{kin} = 1.5$ A GeV for centrality $0.25 < b_0 < 0.45$. The solid lines are for the free unbound protons, the long dashed lines show all (free + bound) protons, the short dashed lines present deuterons. The deuteron lines are multiplied by a factor of 2 for better visibility. The colour code is the same as in Fig. 3.

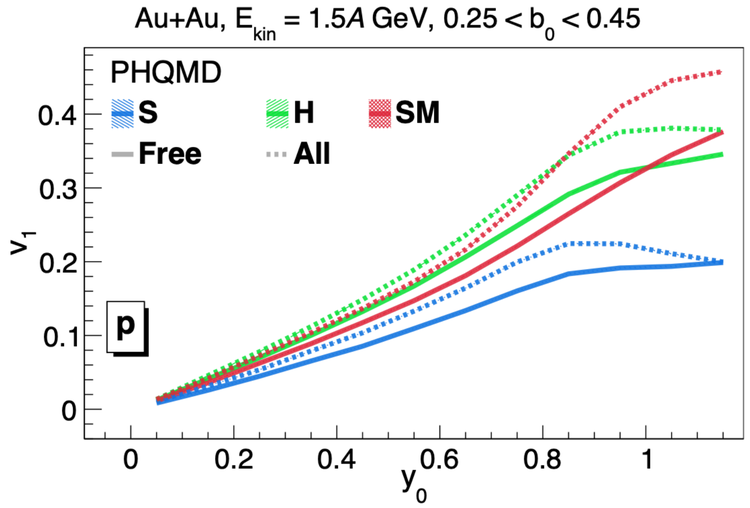


FIG. 6. v_1 of protons versus y_0 for Au+Au collisions at $E_{kin} = 1.5$ A GeV for the impact parameter range $0.25 < b_0 < 0.45$. The colour code is the same as in Fig. 3. The solid lines are for the free unbound protons, while the dashed lines show all (free + bound) protons.

The p_T integrated directed flow $v1$ of finally unbound protons is smaller than $v1$ of all (free + bound in clusters) protons