Machine Learning in Experiment & Analysis

(James) Daniel Brandenburg
Brookhaven National Laboratory
A STAR-perspective talk for the
MMXXII RHIC/AGS Users Meeting

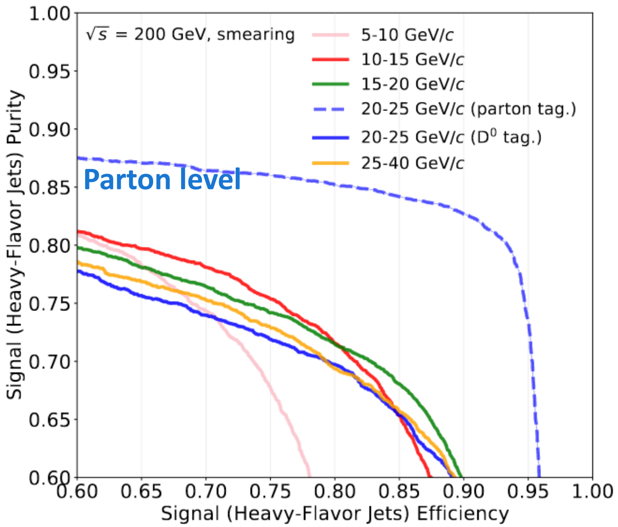
STAR ML Coordinators:
Raghav Kunnawalkam Elayavalli,
Jerome Lauret, JDB

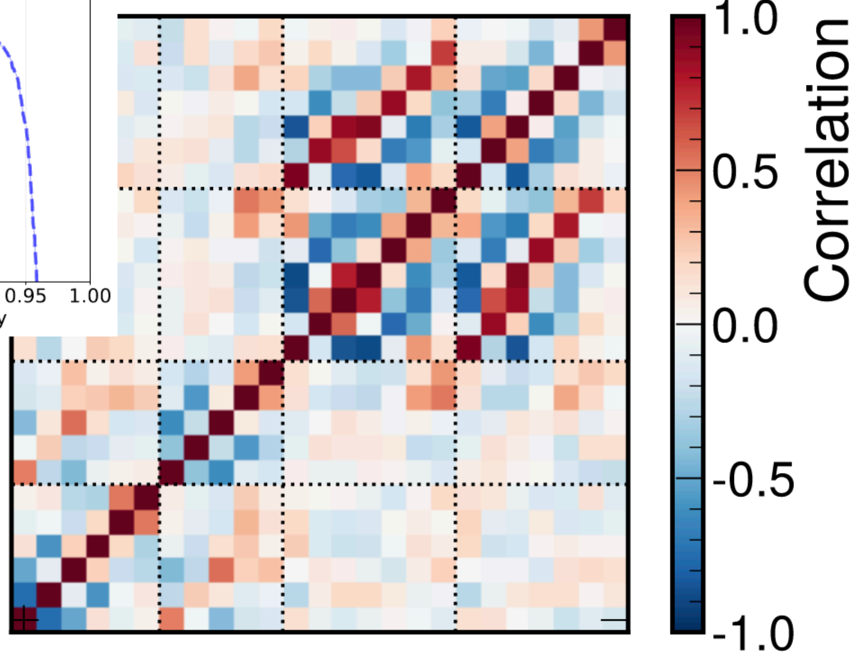
Teaching the Machine

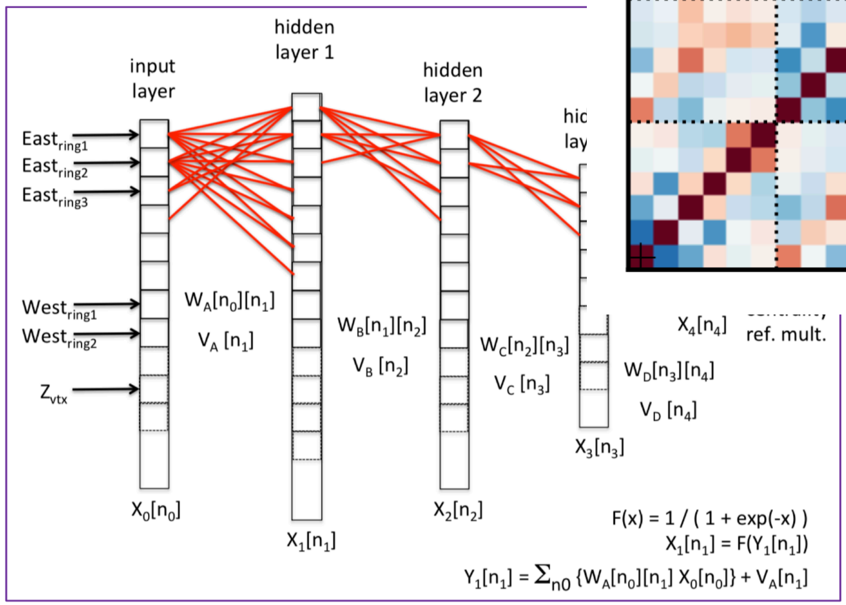
1. Identifying Heavy Flavor Jets

2. Multi-dimensional unfolding

3. A+A Event Characterization



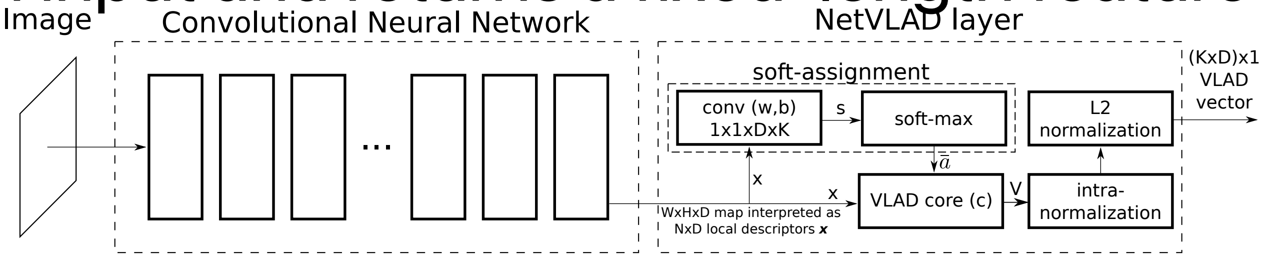




Identifying Jets with JetVLAD

Ponimatkin et. al JINST 17 P03017 (2021)

- Supervised learning model based on NetVLAD [Arandjelovic et al., arXiv:1511.07247]
- *NB: VLAD = Vector of Locally Aggregated Descriptors*
- NetVLAD takes a set of particles as an input and returns a fixed-length feature vector that characterizes it.
  - Similar principle as document2vec etc.
  - Document = set of tokens (characters)



Describe jets as a set of particles:

$$\mathfrak{J} = \{(p_{T,i}, \eta_i, \phi_i, \dots \dots)\}_{i=1}^{n}$$

- Characterize features in a natural coordinate space
- And, find similarity between ~~documents~~ jets in n-dimensional feature space

(a) Mobile phone query (b) Retrieved image of same place

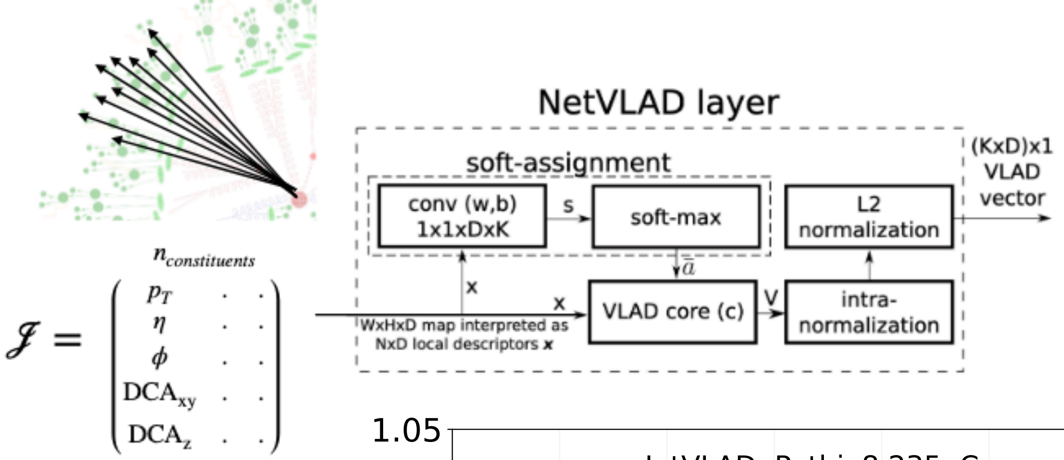
Figure 1. Our trained NetVLAD descriptor correctly recognizes the location (b) of the query photograph (a) despite the large amount of clutter (people, cars), changes in viewpoint and completely different illumination (night vs daytime). **Please see appendix C for more examples.**

Methodology

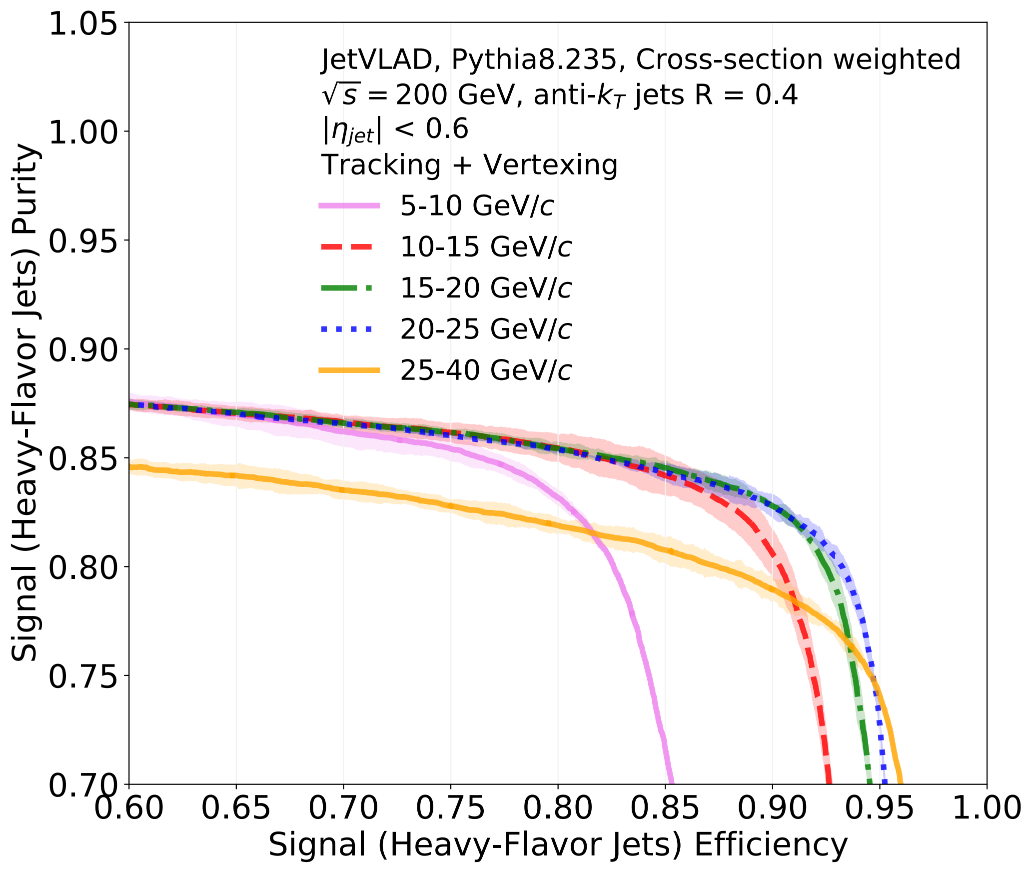
Jitka Mrazkova

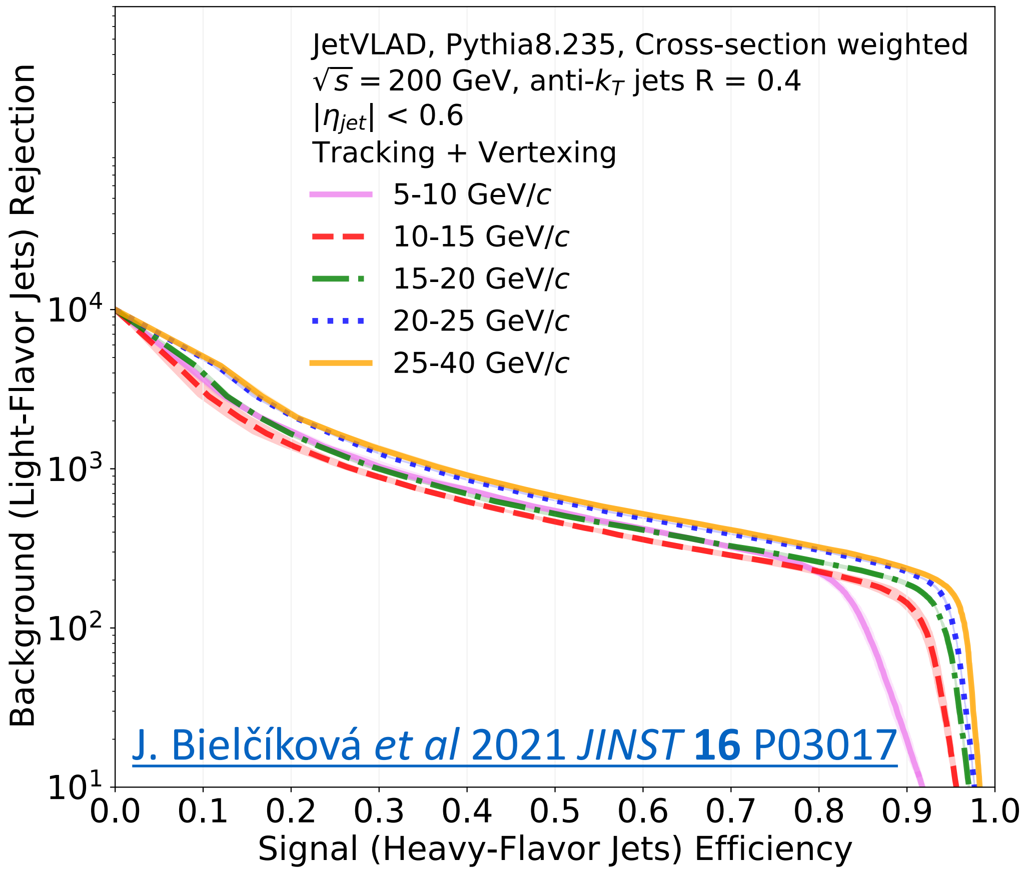
- PYTHIA 8.235 generator
  - $p + p$ collisions at $\sqrt{s} = 200$ GeV
  - Generate cross-section weighted samples
  - Identify jets from both:
    - Partons (c, b quarks)
    - Meson tagging ($D^0 = c\bar{u}$)
- PYTHIA events → STAR detector fast simulators
  - Perform simplified simulation of detector response, so that input is ~ real data
  - Gaussian smearing of p_T → account for tracking resolution
  - Event vertex information smeared by known STAR HFT resolution
- Network Input: $(p_T, \eta, \phi, DCA_{xy}, DCA_z)$
- Dataset split: 80% : 10% : 10% for training : testing : validation
- Grid search for hyper-parameter optimization

Tagging Results



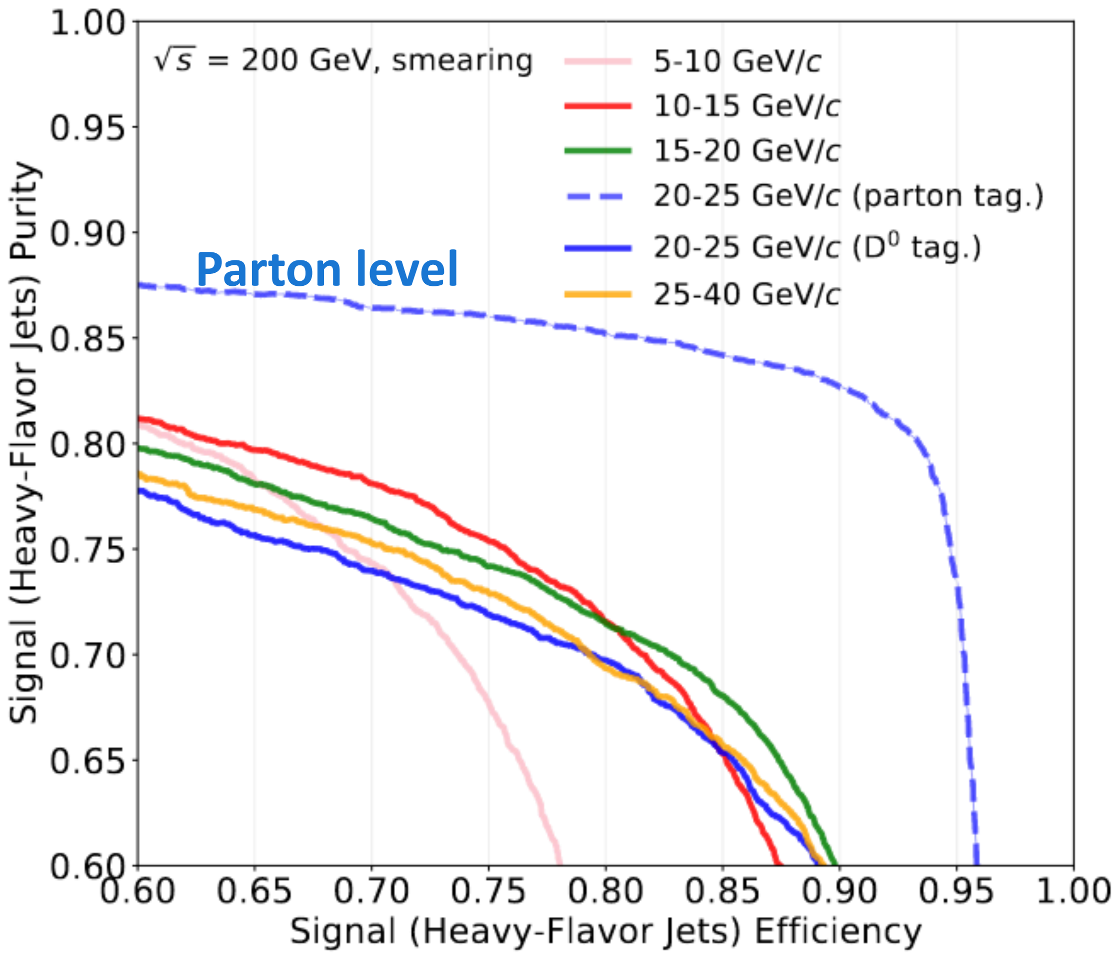
- Good tagging performance: $> 80\%$
- Background: Relatively high rejection at e.g. 80% efficiency
- Minimal dependence on jet p_T

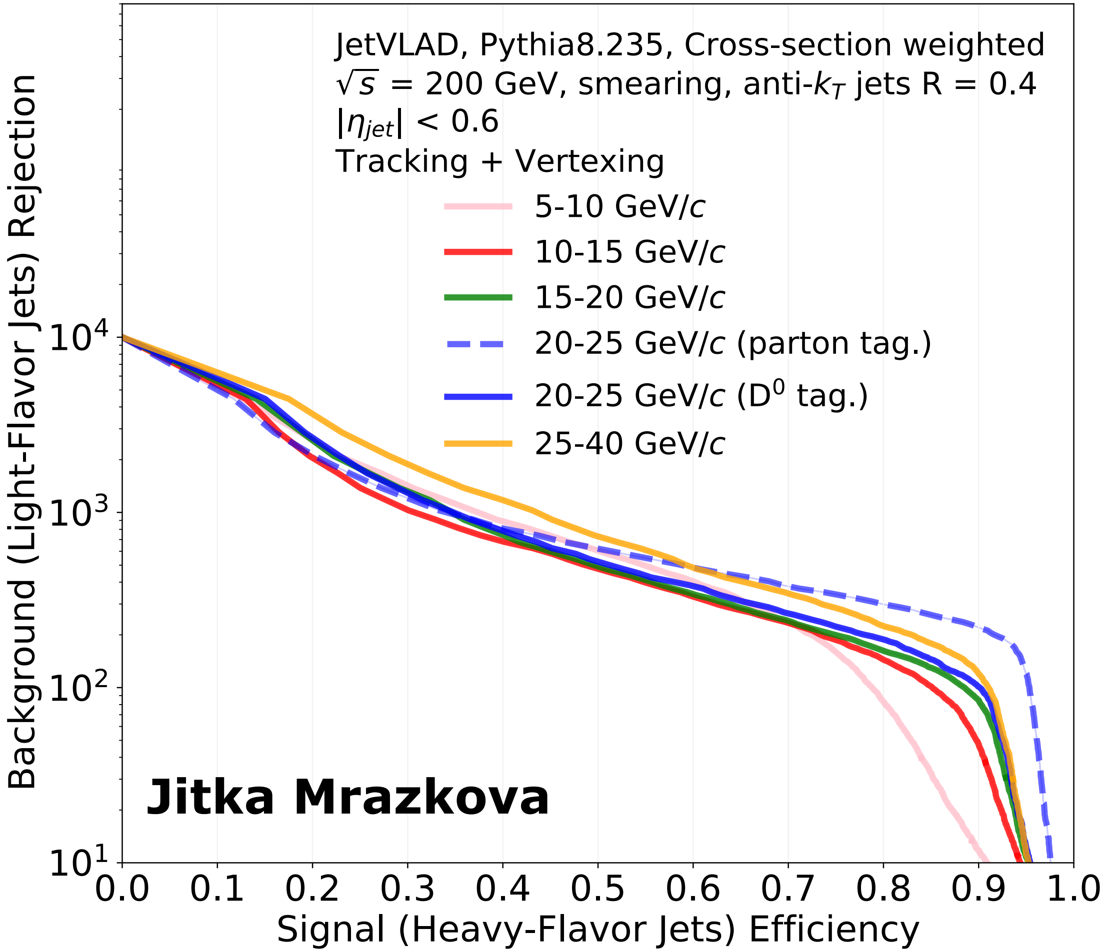




Jitka Mrazkova, et. al.

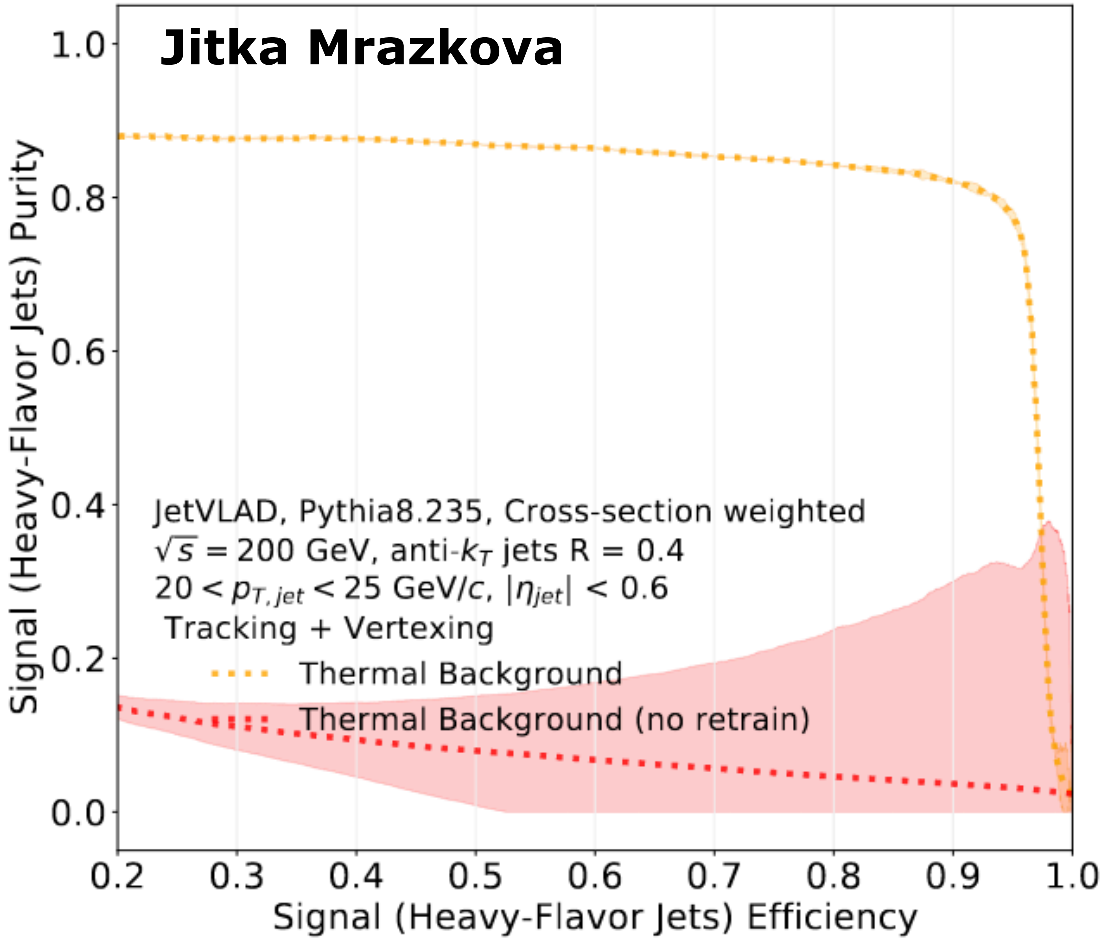
Robustness and effect of the jet id method?

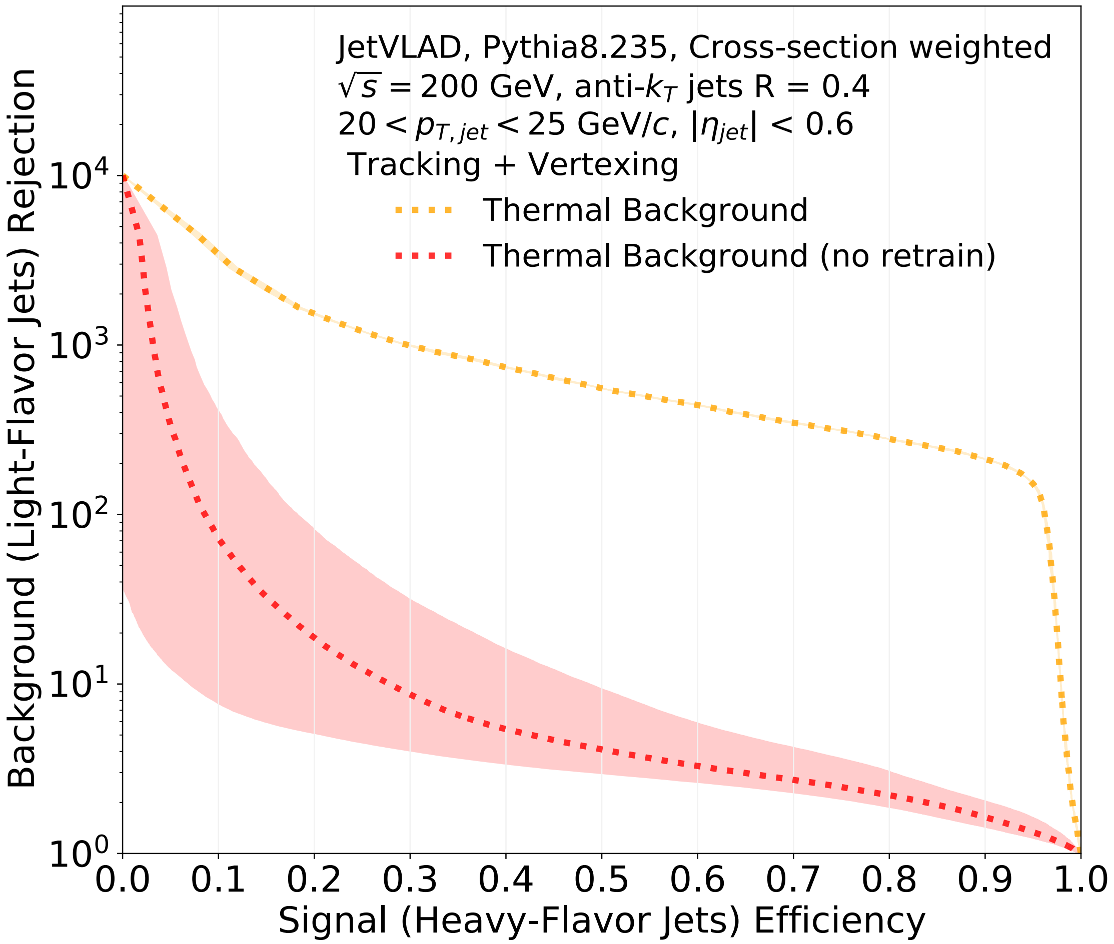




- Particle level (i.e. D^0 meson) tagging approach is more experimentally consistent

Robustness and effect of the jet id method?





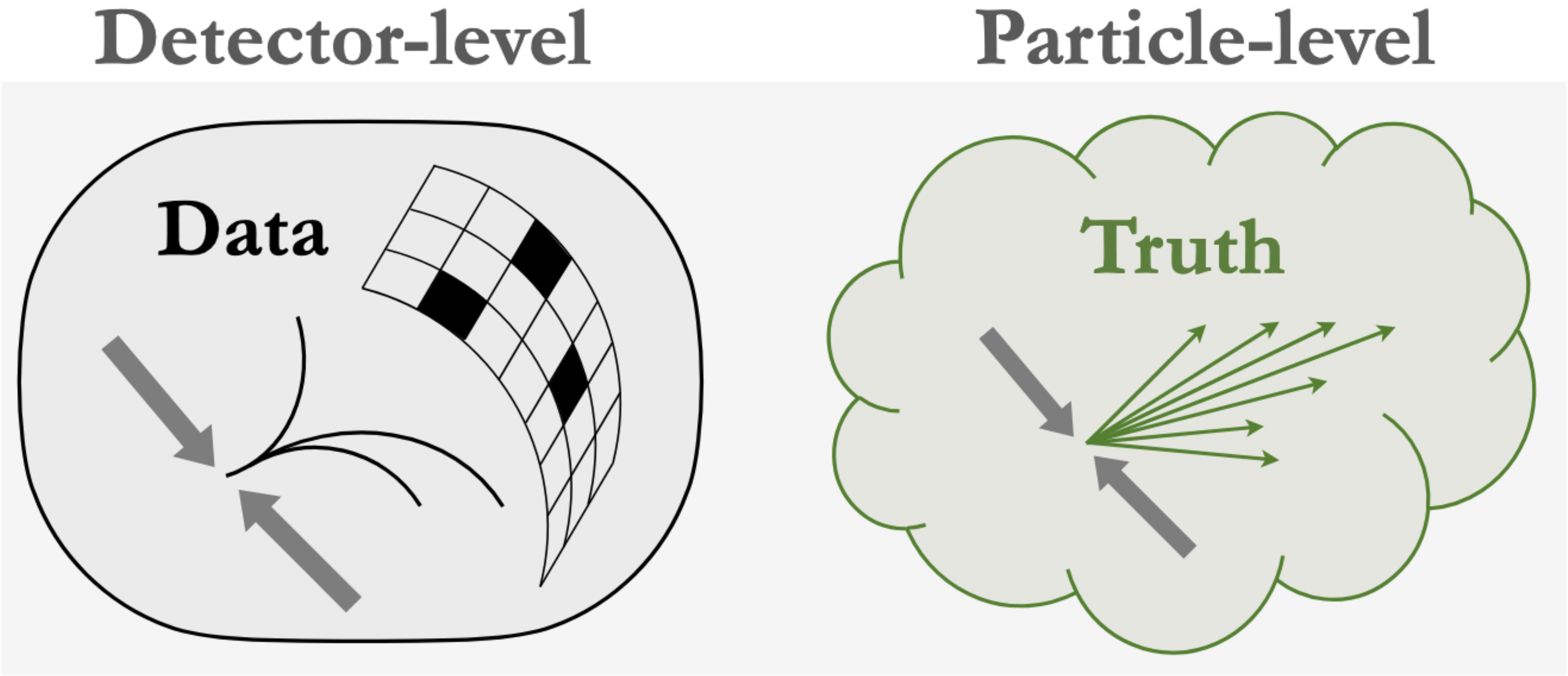
- Particle level (i.e. D^0 meson) tagging approach is more experimentally consistent
- Effect of thermal background from event?
  - Without retrain, performance suffers. After retrain, comparable performance as w/o

Next steps

- Study energy dependence: $\sqrt{s} = 510$ GeV, $\sqrt{s} = 7$ TeV …
- Utilize powerful JETSCAPE framework for additional studies
  - Jets in A+A events
  - Quenched jets
  - …
- Present methodology + initial results J. Bielčíková *et al* 2021 *JINST* **16** P03017
- + look out for new preprint with these new results **Mrazkova et. al in prep**

OK, now let's say we have our jets and do a measurement - How do we get back to the PHYSICS

Unfolding: Measurement → Physics



What are our options?!

Traditional Methods:

RooUnfold Methods

- iterative ("Bayesian");
- singular value decomposition
- simple inversion of the response matrix without regularization

PyUnfold

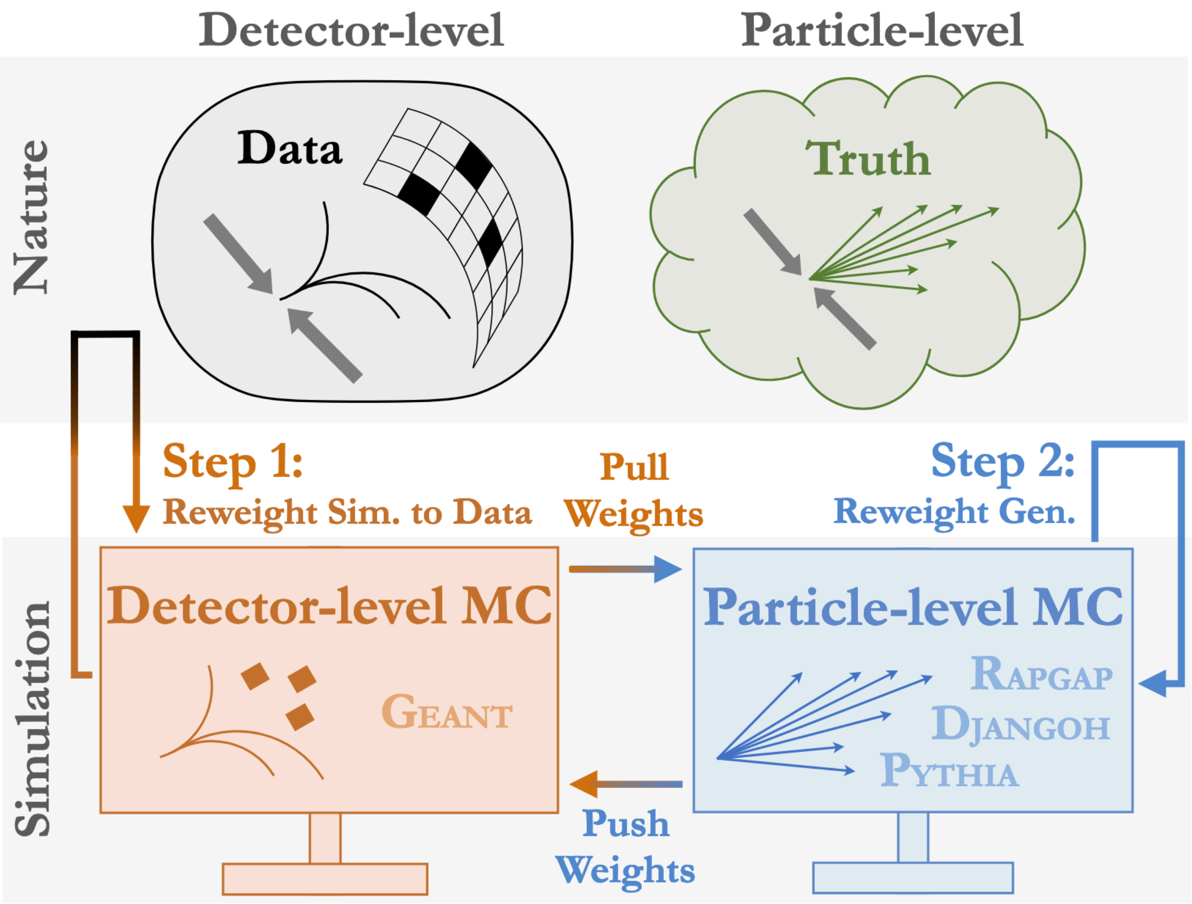
- Regularized inversion (Truncated Singular Value Decomposition)
- iterative ("Bayesian") = Richardson-Lucy

…

…

[Phys. Rev. Lett. **124**, 182001]

ML + Unfold = OmniFold



OMNIFOLD: Whats the big deal?

Special thanks to Ben Nachman

Three **challenges** in unfolding

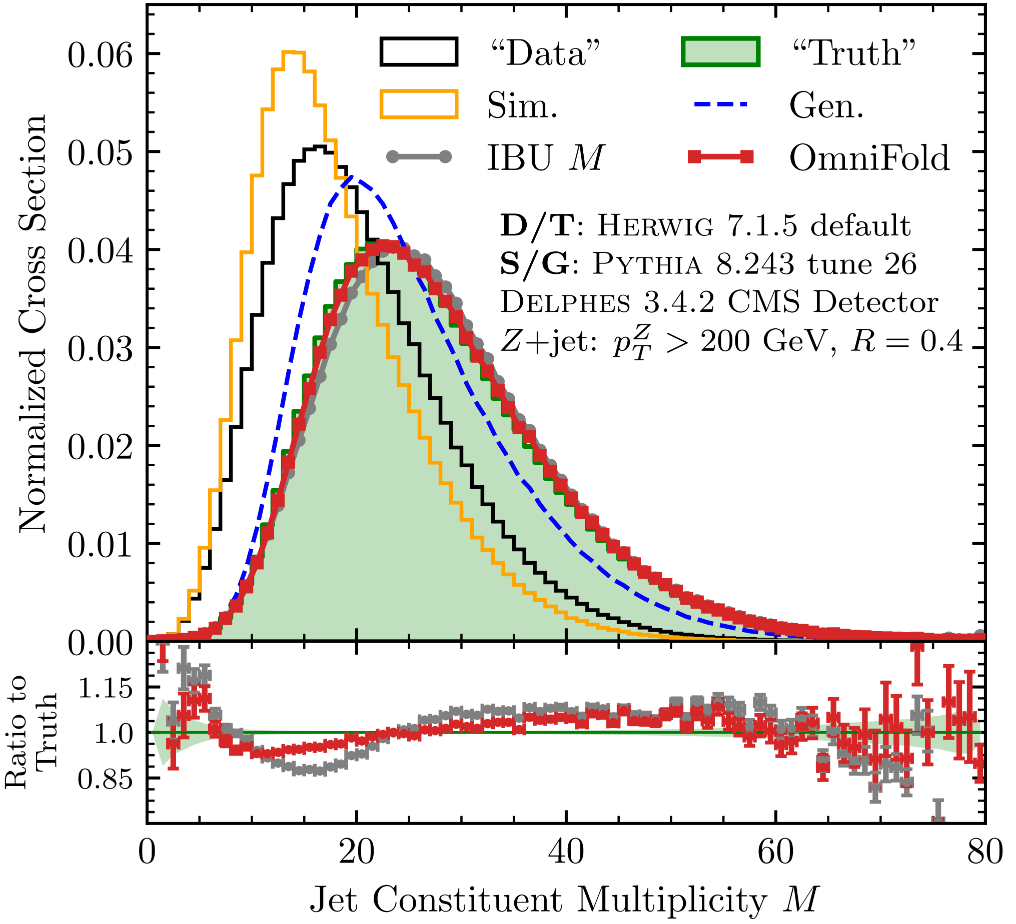
1. Binned observables
2. Limited dimensional simultaneous unfolding (partially consequence of #1)
3. Inability to encode effects of auxiliary features [see what you want to see]

- Closure on 3 flavors to demonstrate the effectiveness of OMNIFOLD's solutions
  - UNIFOLD : A single observable as input. This is an unbinned version of iterative Bayseian unfolding (IBU)
  - MULTIFOLD: Many observables as input (e.g. jet kinematics)
  - OMNIFOLD: Use full event/jet as input – no limit on information vector

$$\mathcal{I} = \langle y, p_T^e, q_T, \phi, \cdots \rangle$$

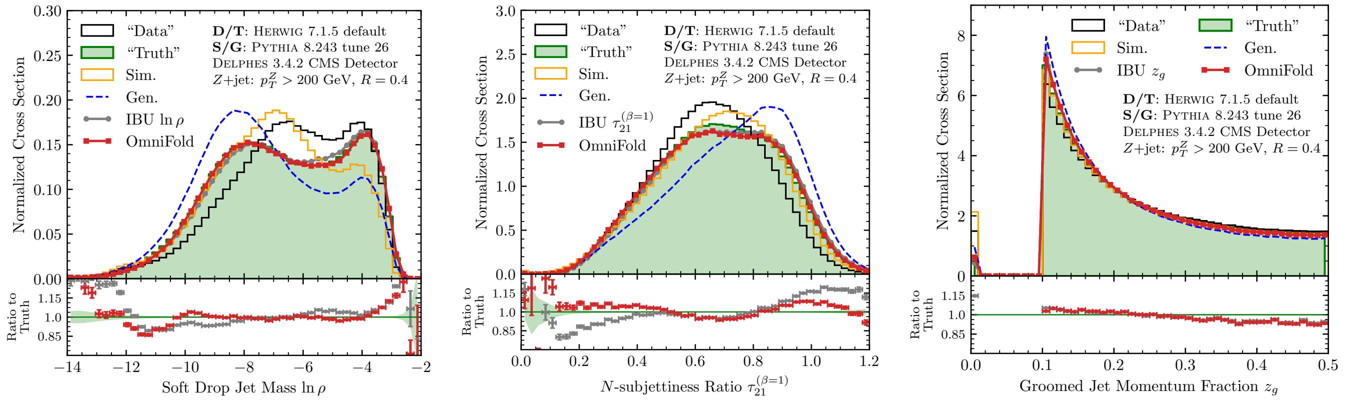
OMNIFOLD: Closure Example

Special thanks to Ben Nachman



- Many more examples in [Phys. Rev. Lett. **124**, 182001](Phys. Rev. Lett. 124, 182001)

- OMNIFOLD: Even in 1D cases, matches or exceed performance of e.g. IBU
- But what about more multidimensional cases?

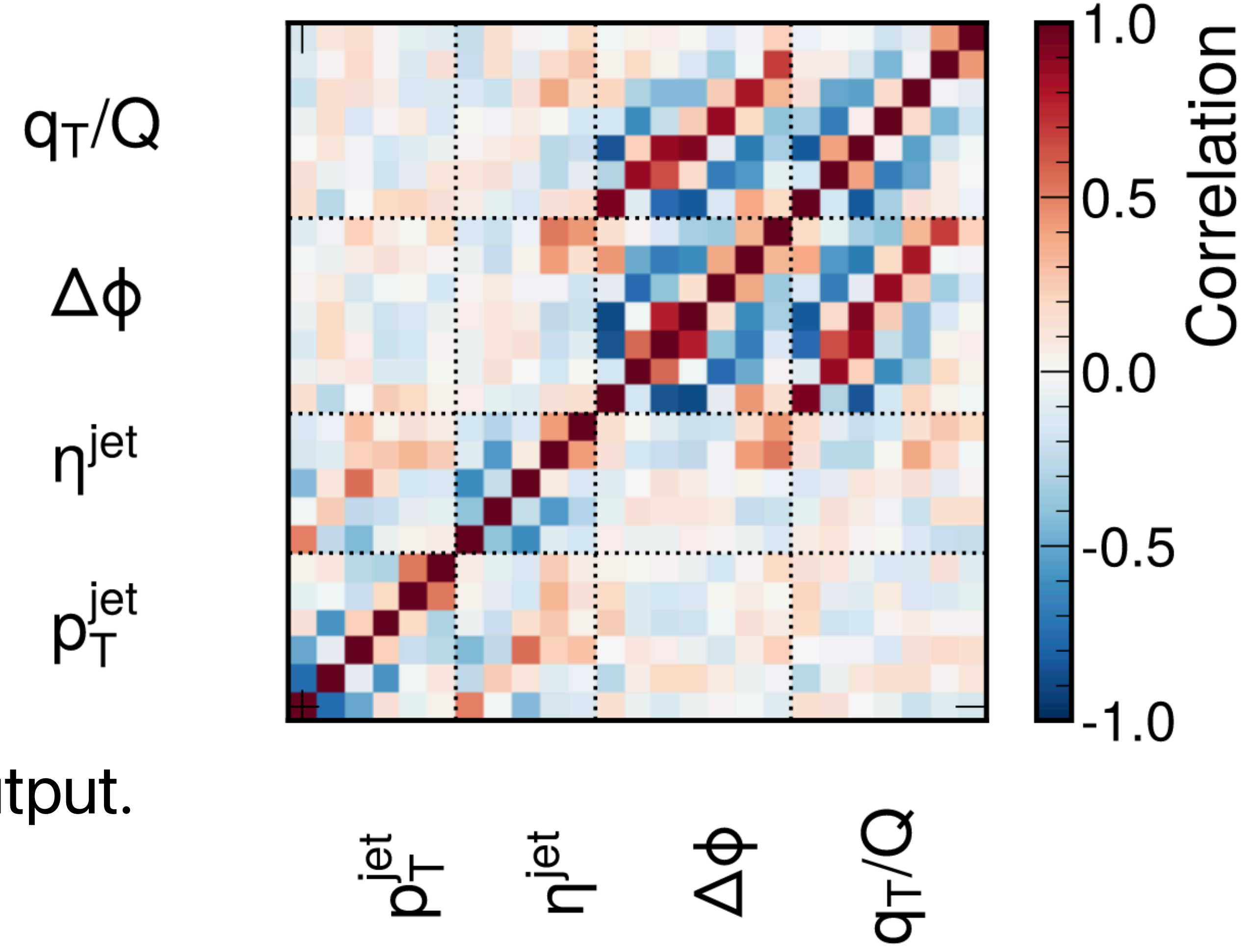


- Robust, consistent performance

Plus free lunch: Correlations *pro bono physico*

$$\mathscr{I} = \langle y, p_T^e, q_T, \phi, \cdots \rangle$$

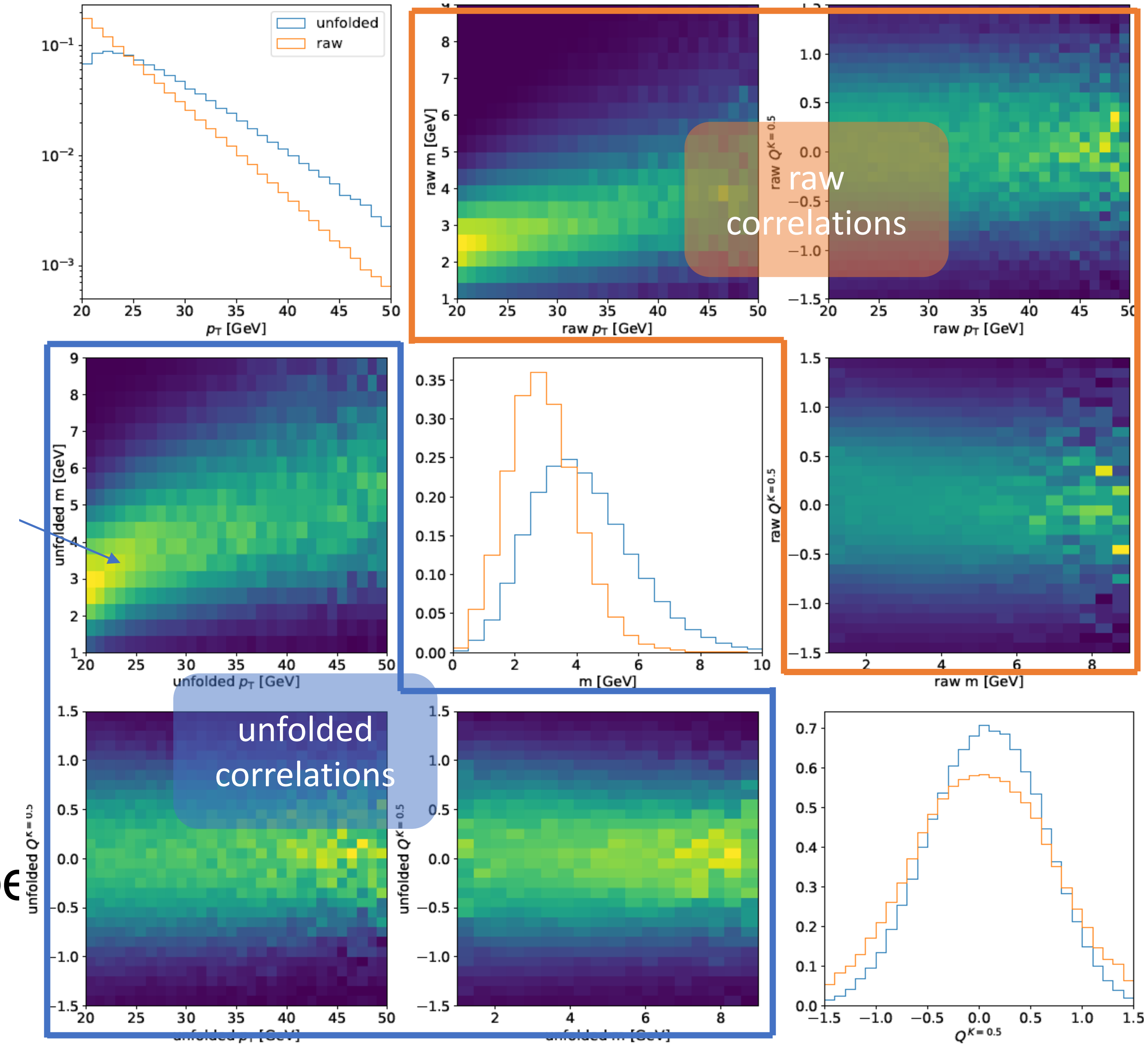
$$\color{green}{\mathscr{O} = \langle y, p_T^e, q_T, \phi, \cdots \rangle}$$



- ALL variable correlations as a natural output.
- Entirely unbinned!
  - Binning is only for visual demo

Applicability at RHIC

- PYTHIA 8.235 generator
  - Charged jets only for now
  - Jet algorithm anti-k_T $R = 0.4$
  - Jet selection:
  - pT > 10 GeV,
  - |eta| < 0.6,
  - number of constituents > 1,
  - passed SoftDrop.
  - SoftDrop grooming parameters:
    - beta = 0 and z_cut = 0.1.
- Correlations before and after unfolding
- Recovering ‘true’ correlations ope physics opportunities



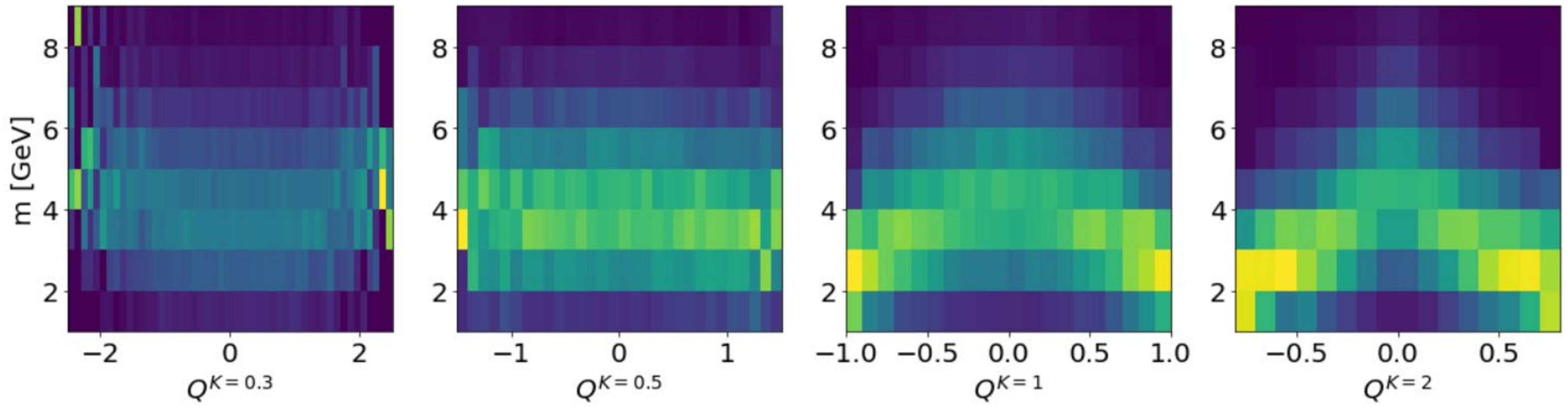
Work by Youqi Song (Yale)

Studies and Opportunities

Youqi Song (Yale)

- Explore correlations that encode fragmentation information
- E.g. Jet mass (m) vs. Jet charge (Q)

PYTHIA charged jets



More always better?

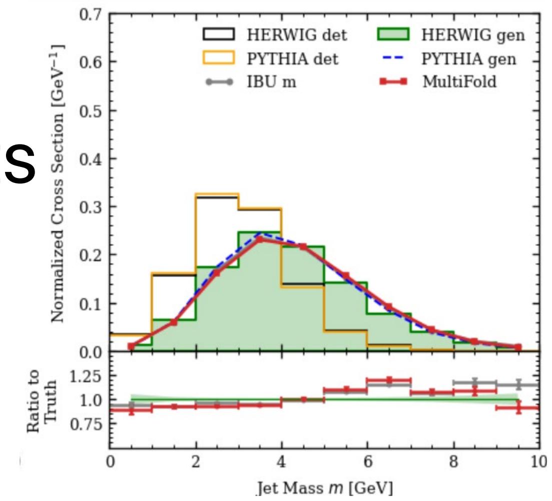
HERWIG & PYTHIA closure tests

6 inputs: $p_\perp, Q^{K=0.5}, M, M_g, R_g, z_g$

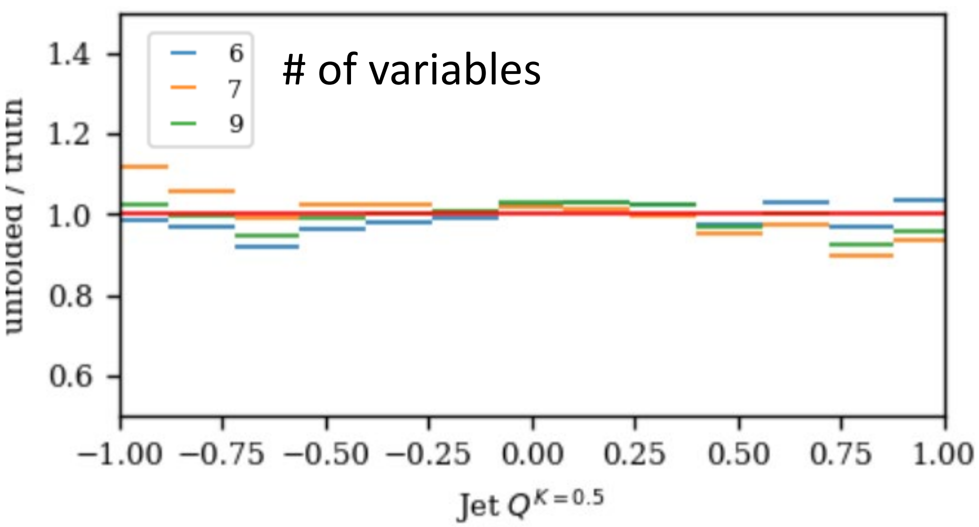
7 inputs: + $Q^{K=2}$

9 inputs: + underlying event mult & $p_\perp$

Or at least not worse?



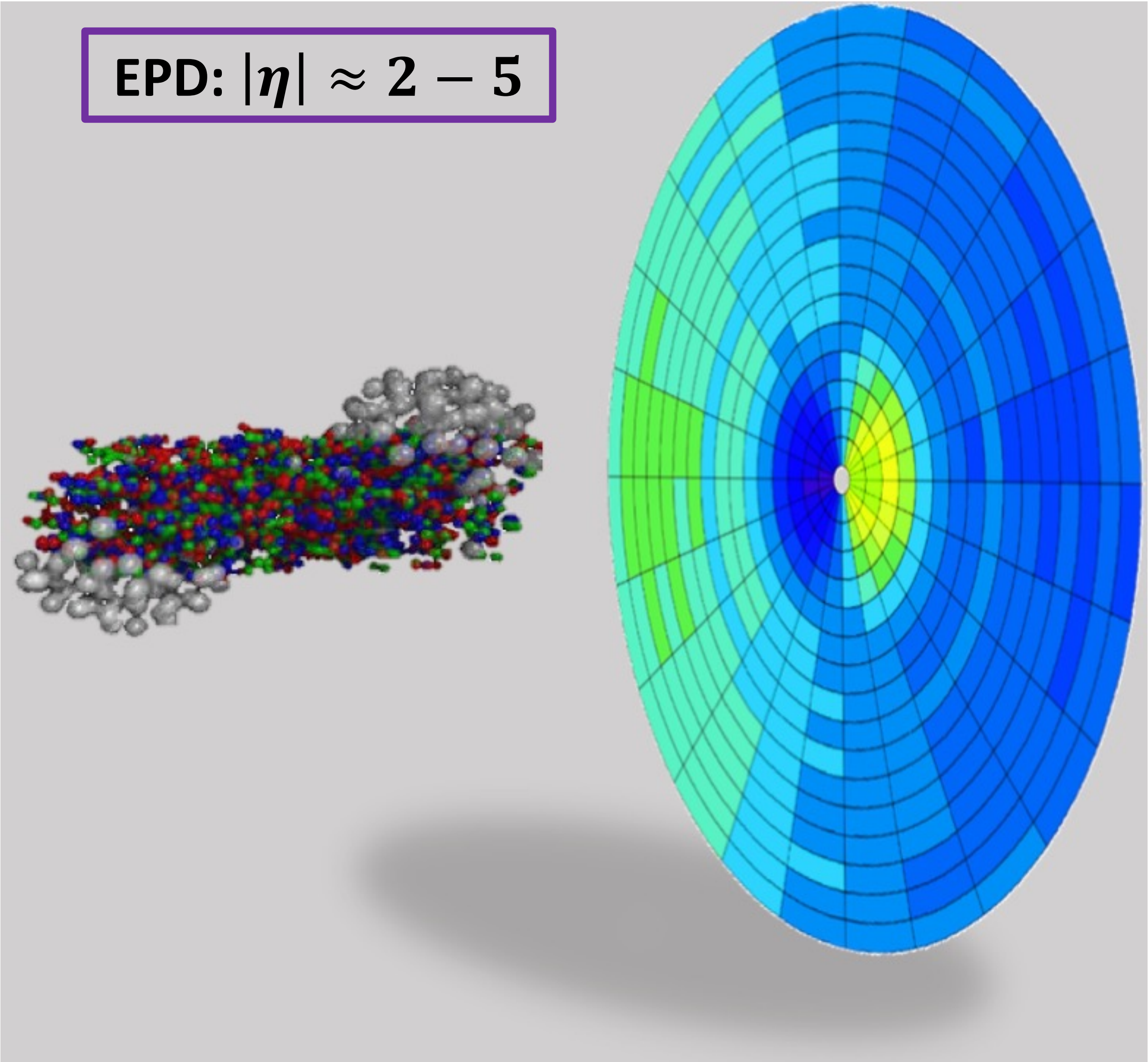
9 variables



STAR Event Plane Detector (EPD)

Goal:
Characterizing A+A events across a range of collision energies

Learn the event multiplicity (centrality) independent of the mid-rapidity tracking detector

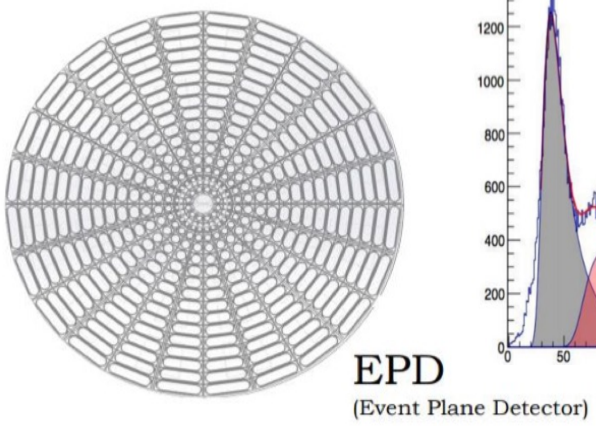


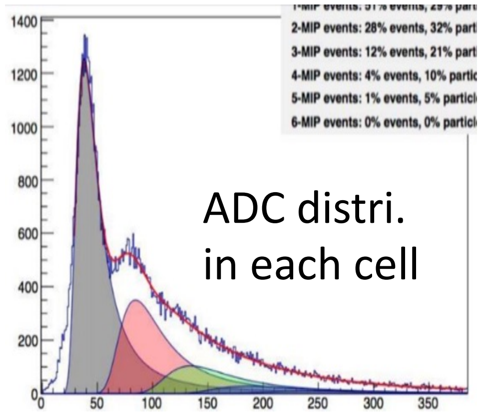
STAR Event Plane Detector (EPD)

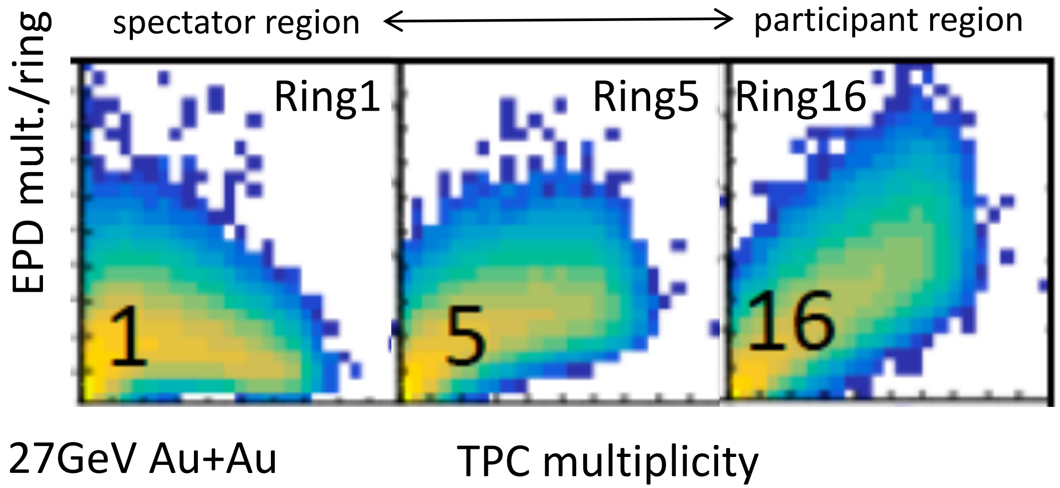
Shinichi Esumi, Yuri Sato

Challenges:

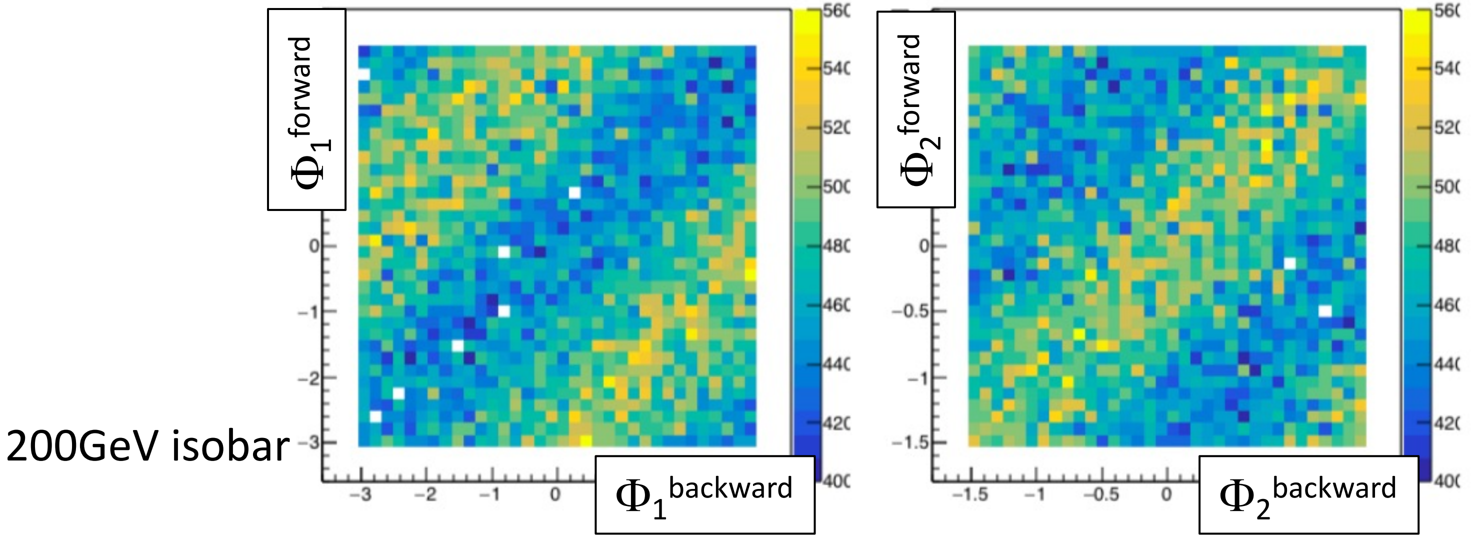
- Many different measurements
  - ADCs on 31 * 12 * 2 = 744 tiles
- ADC distribution depends on interaction location (z_{vtx})
- Energy dependence of mid-to-forward rapidity multiplicity
  - At various energies – strong y dependence of multiplicity spectrum
  - Changes in relative acceptance





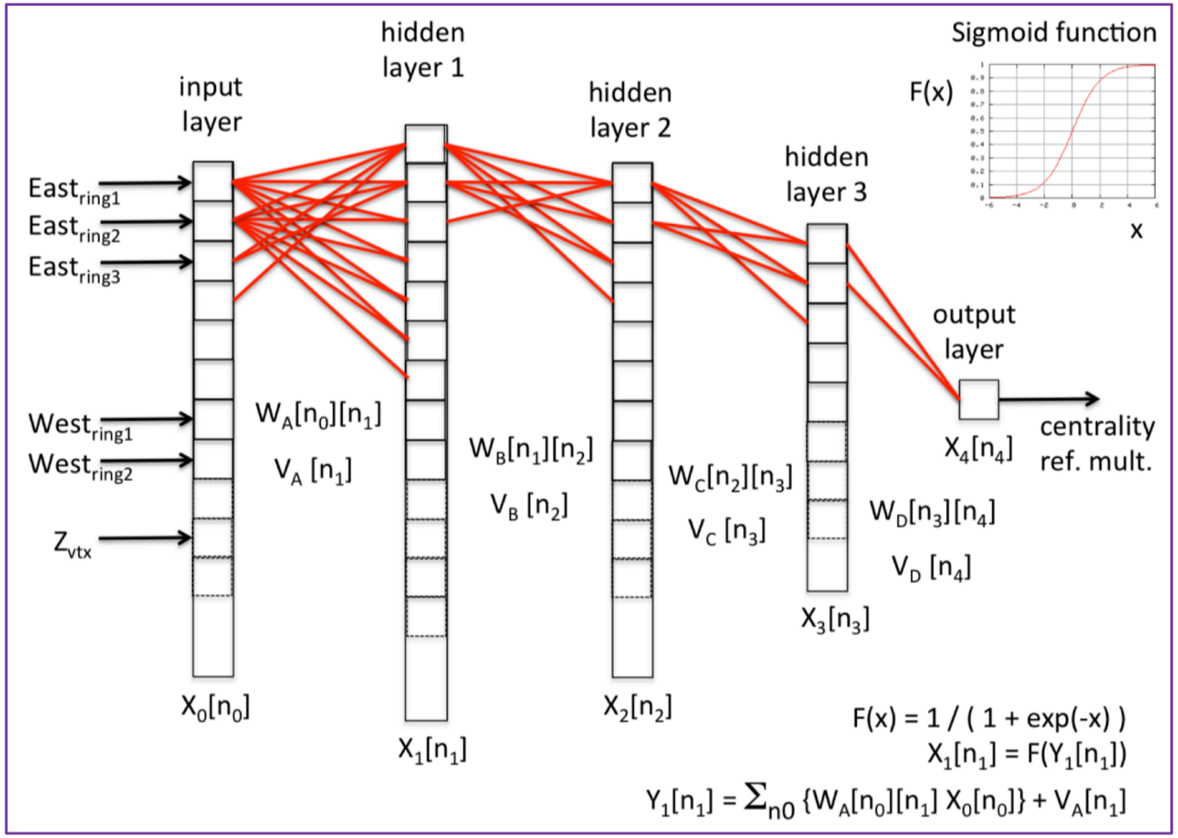


Forward-backward event-plane correlation



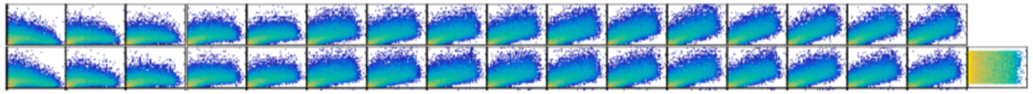
Methodology and network architecture

Shinichi Esumi, Yuri Sato



27GeV Au+Au 50k education events and 50k test events (EPD+TPC)

16 ADC sum values from 2 arms + zvertex = 33 input neurons



3 hidden layers (n1=66, n2=30, n3=10 neurons)

back-propagation to modify weight W and bias V

Error (target-output) : $E = 0.5\,(X_T[n_4] - X_4[n_4])^2$

$dE/dX_4 = X_T - X_4$, $dX_4/dY_4 = F'(X_4)$, $dY_4/dW_D = X_3$

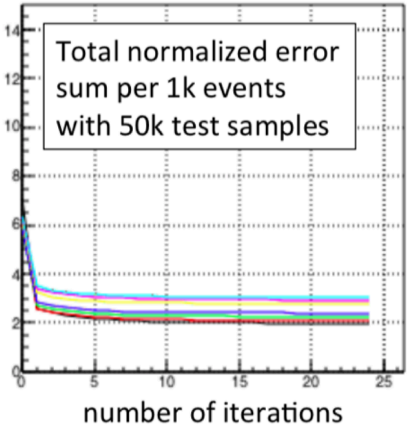
$dE/dW_D = (dE/dX_4)\,(dX_4/dY_4)\,(dY_4/dW_D)$

$W_D = W_D + \alpha\,(X_T - X_4)\,F'(X_4)\,X_3$

$V_D = V_D + \alpha\,(X_T - X_4)\,F'(X_4)$

$dE/dX_3 = (X_T - X_4)\,F'(X_4)\,W_D$

.....



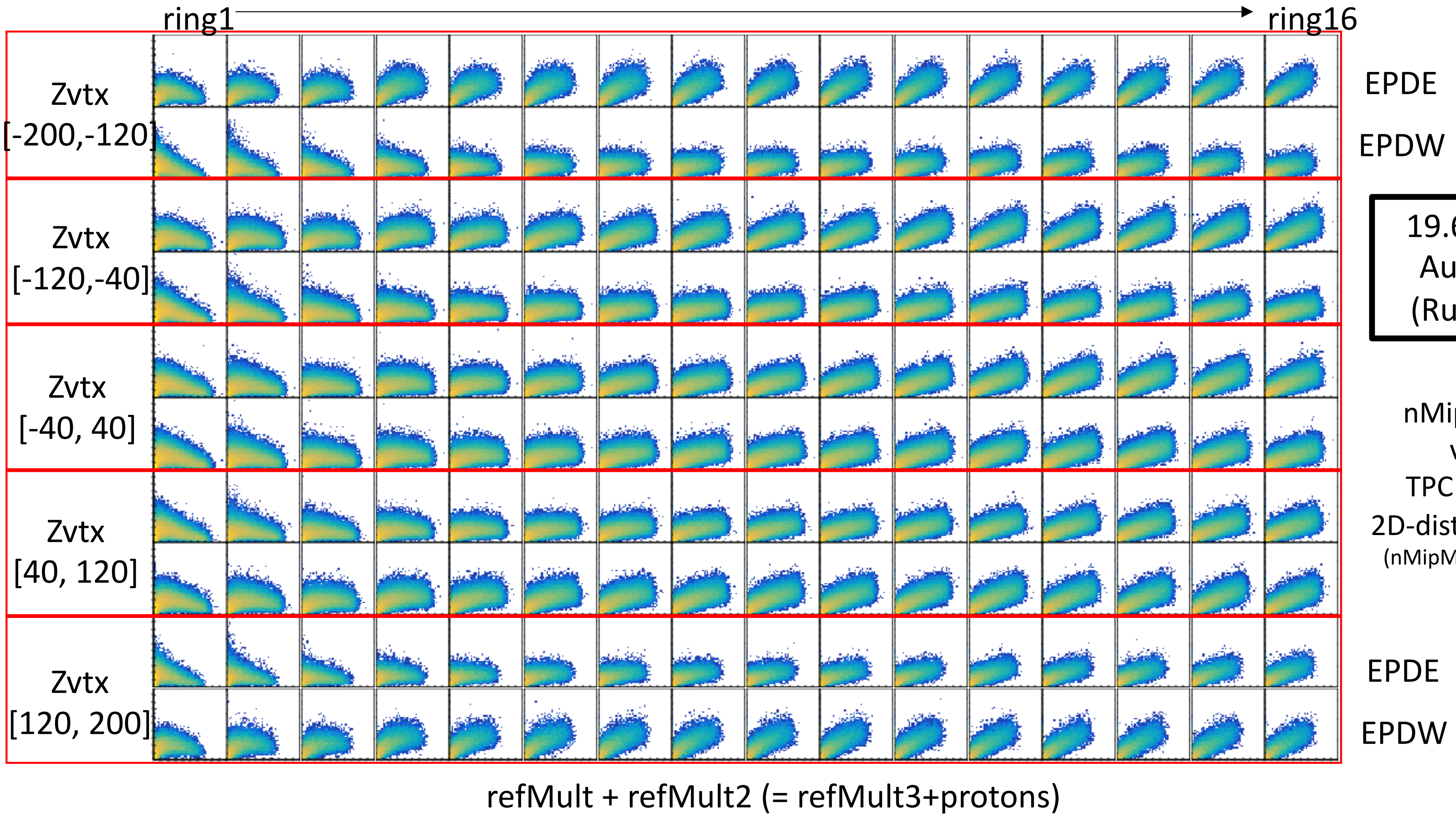
1 output layer with 1 neuron (as refmult)

ShinIchi Esumi (a few slides taken from Yuri Sato's Master thesis)
Inst. of Physics, Univ. of Tsukuba
Tomonaga Center for the History of the Universe (TCHoU)

Learn to predict the mid-rapidity multiplicity from:
Each set of EPD signals (2 x 16 ring summed ADCs)
+ primary interaction vertex location

(some) of the input data

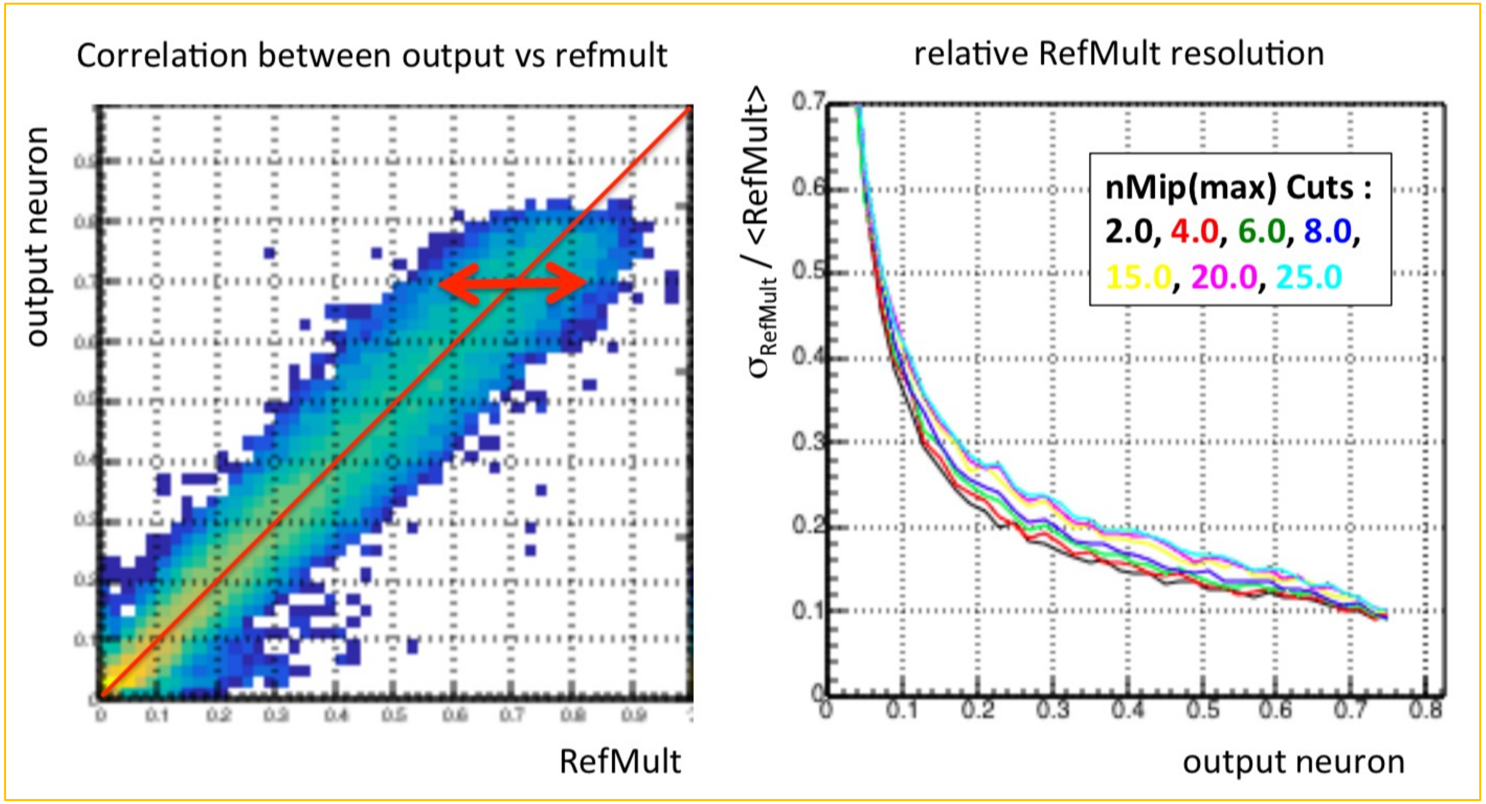
Shinichi Esumi, Yuri Sato

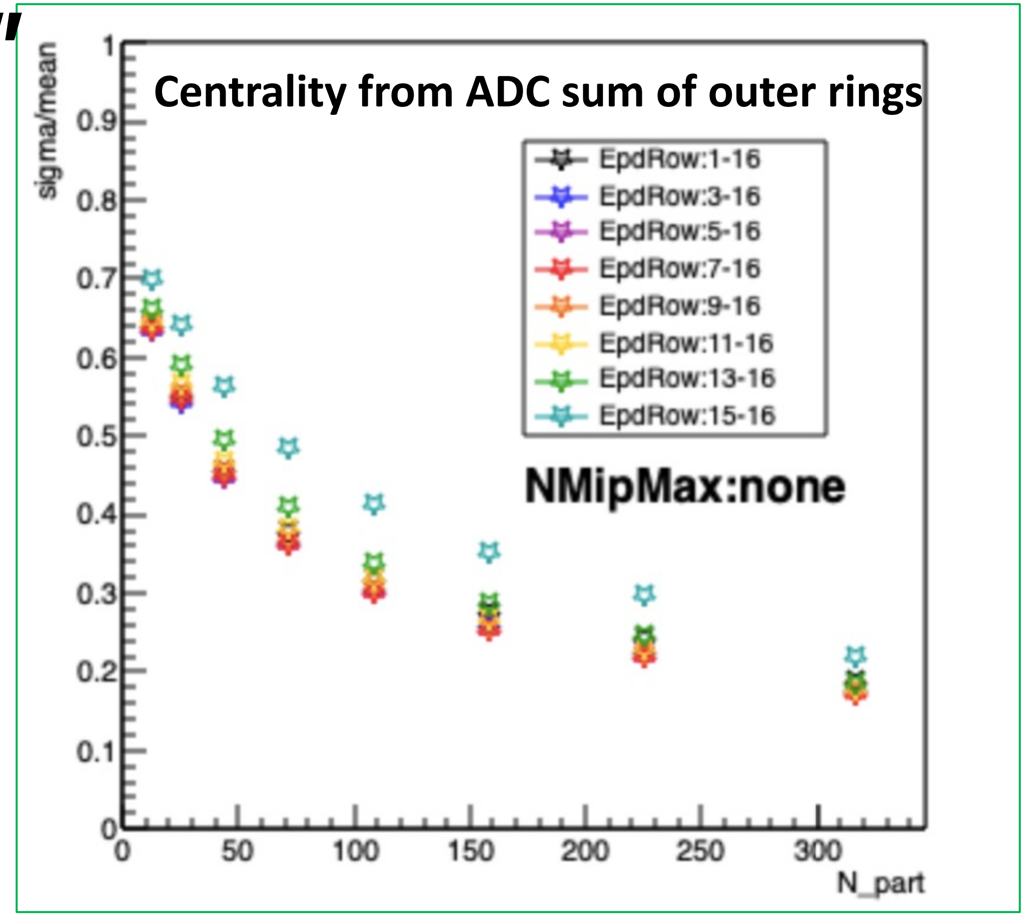


19.6GeV
Au+Au
(Run19)

nMip sum
vs
TPC mult.
2D-distribution
(nMipMaxCut=6)

Results & comparison to "basic"



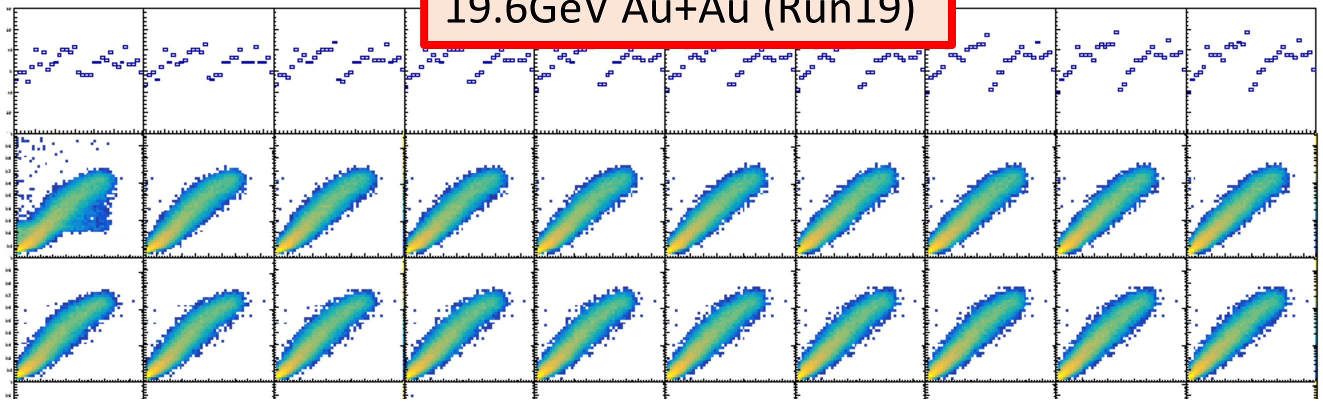


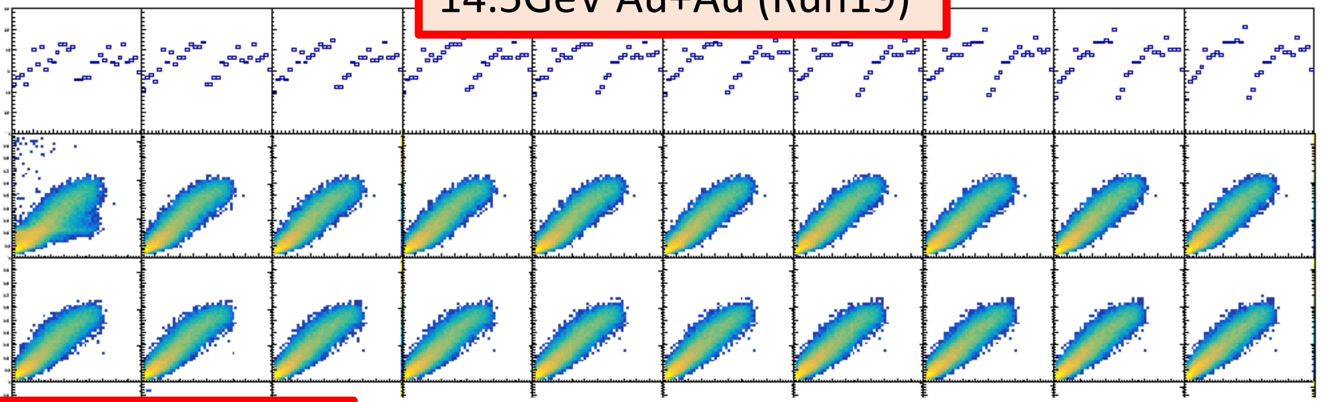
- Diagonal correlation – learning the multiplicity at mid-rapidity
- NN resolution is ~×2 better than "basic" single ring-sum
- Optimal resolution utilizing ADC spectrum up to 2 MIPs

Shinichi Esumi, Yuri Sato

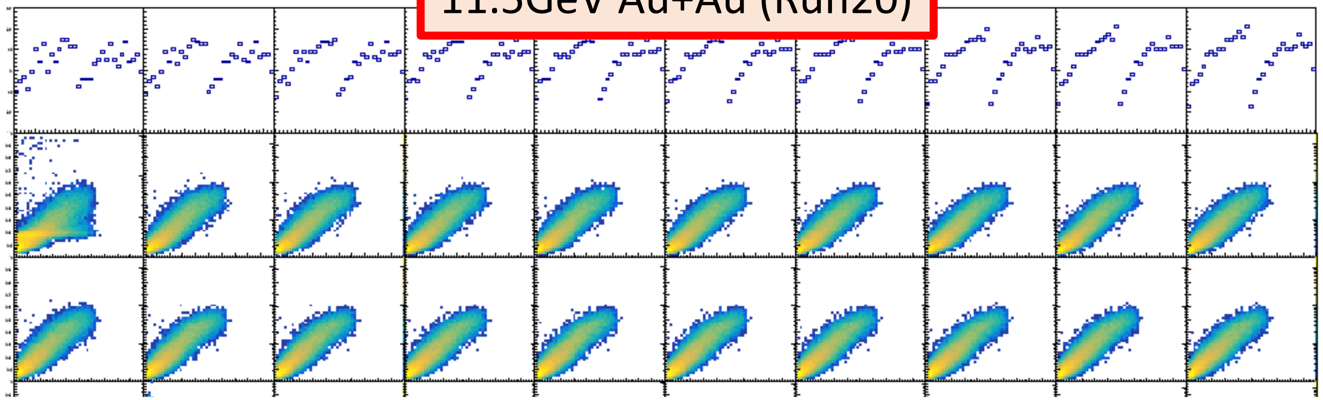
Energy dependence

- Strong correlation obtained at all energies
- Understand what was learned by looking at the weights





nMipMaxCut : 6
Zvtx [-100, 100]



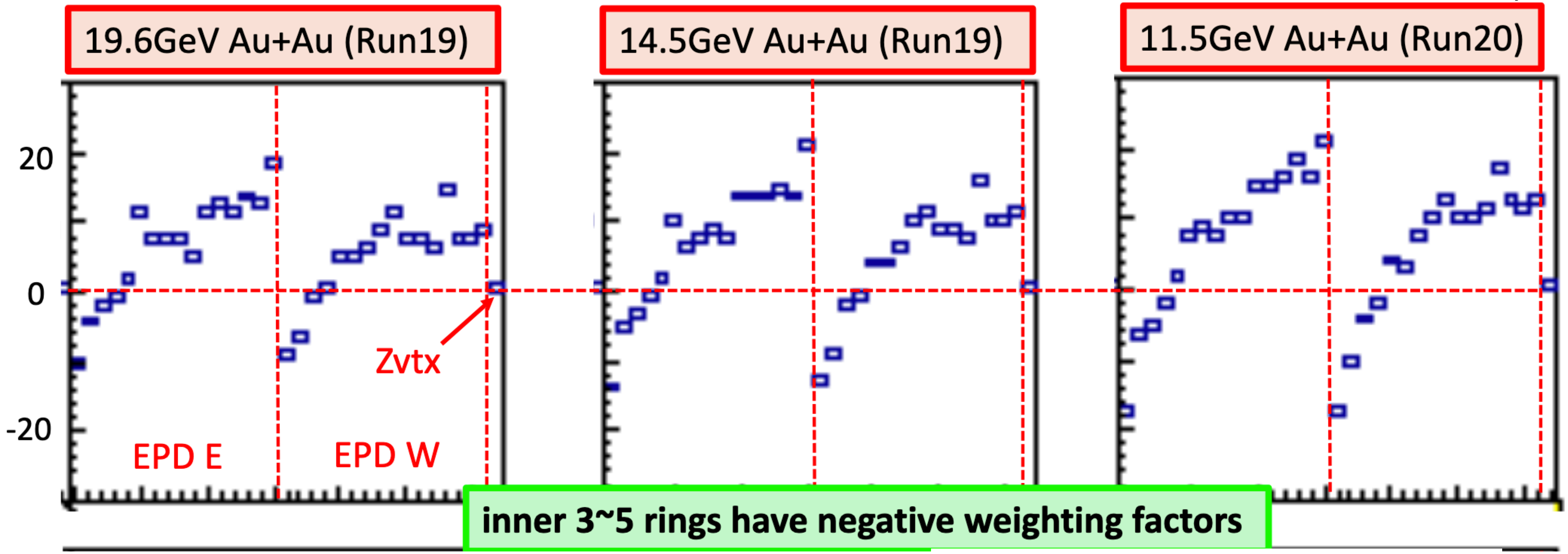
Beam energy dependence

(19.6 -> 14.5 -> 11.5 GeV)

Shinichi Esumi, Yuri Sato

Understanding the weights & energy dependence

Shinichi Esumi, Yuri Sato



- Sign change in weights, correspond to positive/negative correlation with mid-rapidity multiplicity
- Change in rapidity spectrum -> intuitive understanding of the changes in weights
- Event centrality characterization independent of mid-rapidity tracker
  - Important for studies susceptible to auto-correlations, e.g. fluctuation analyses

Summary

Exciting times in ML research

1. ML is a tool for improving analysis and extracting meaning from data
2. Fundamental ML research → HEP
   - **OmniFold:**
     - unites authors from various fields
     - Powerful tool for HEP, only just begun to dig into the possibilities
3. ML techniques can be understood, and made reliable for physics

Thanks to all whose work I showed