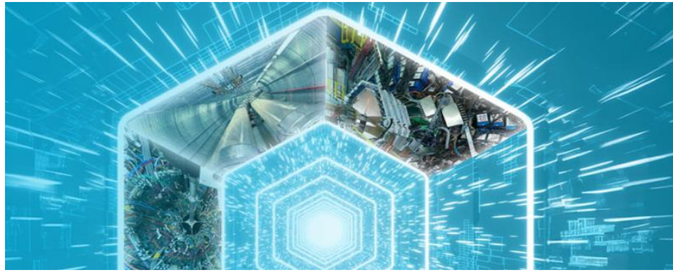


Interpretable Machine Learning applications to Jet Background Subtraction

Charles Hughes
Iowa State University

2024 RHIC & AGS Users' Meeting
Brookhaven National Laboratory, Upton, N.Y.
06/11/2024

U.S. DEPARTMENT OF ENERGY
Office of Science

Motivation

Looking at new methods for jet background subtraction in heavy ion collisions

Machine learning proving to be a useful tool – but wanted careful consideration

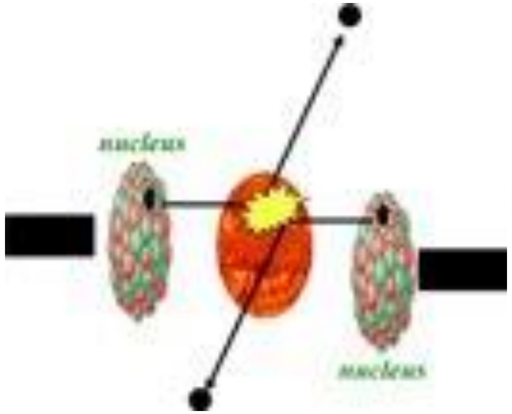
Today – presenting this careful consideration as "interpretable machine learning"

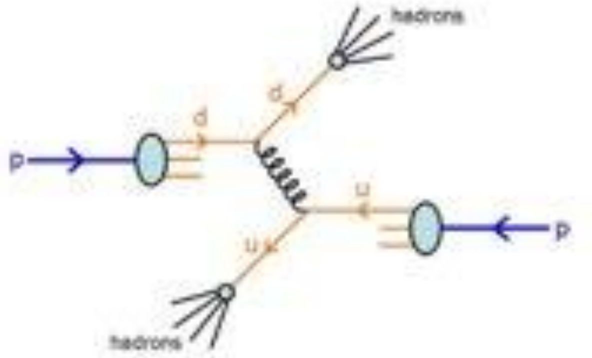
This application of a interpretable machine learning method provides a useful case study for current RHIC heavy ion experiments (and in general)

Jets in Heavy Ion Collisions

- Internally generated probe of QGP
- Jets lose energy in medium (AA) relative to vacuum (pp)
- Energy loss is path length dependent - giving us information about Quark Gluon Plasma (QGP) properties

Absolute Energy Loss (medium vs. vacuum)





Relative Energy Loss (more mediums vs. less medium)

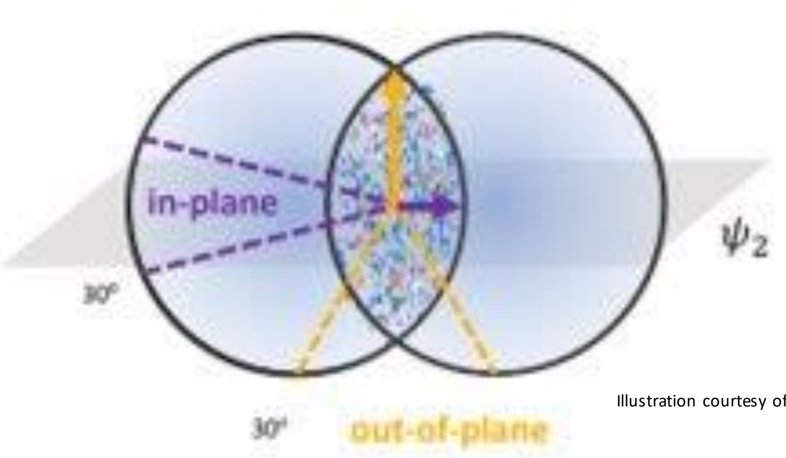
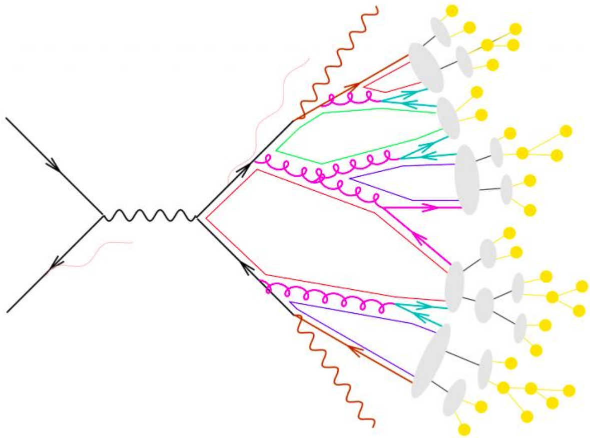


Illustration courtesy of Caitilin Beattie

Reconstructing Jets (hadron collisions)



- Jets are ambiguous objects – even at generator level

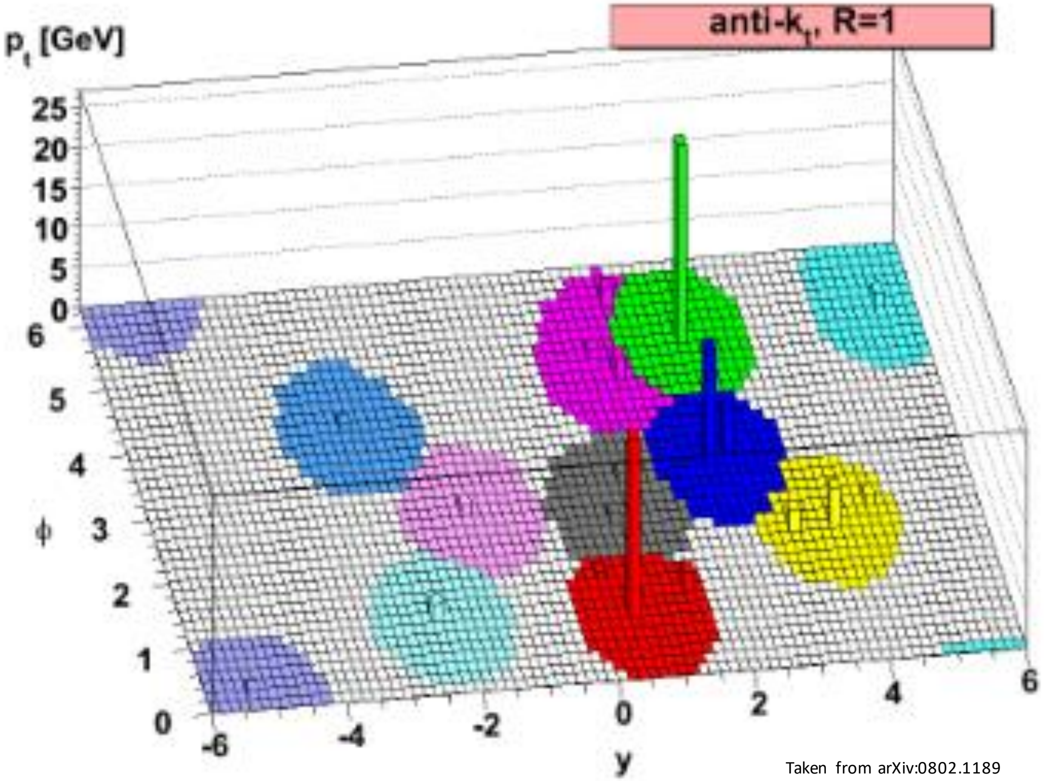
- Experiments rely on **reconstruction algorithms** (decide which particles go in the jet) and **recombination schemes** (decide how to calculate jet properties from particle properties)

- Example – anti-k_T algorithm w/ boost invariant p_T scheme

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} ,$$

$$d_{iB} = k_{ti}^{2p} ,$$

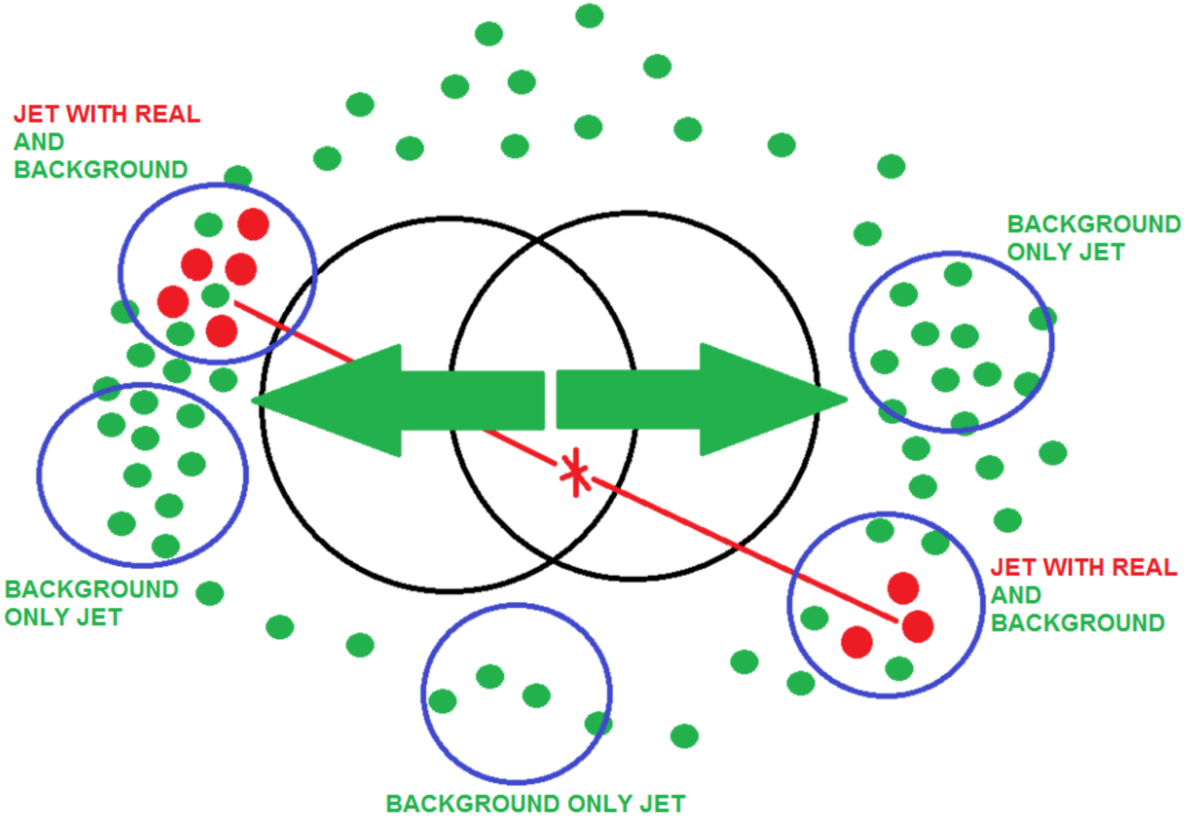
p= -1



Taken from arXiv:0802.1189

The Problem of Jet Background in Heavy Ion Collisions

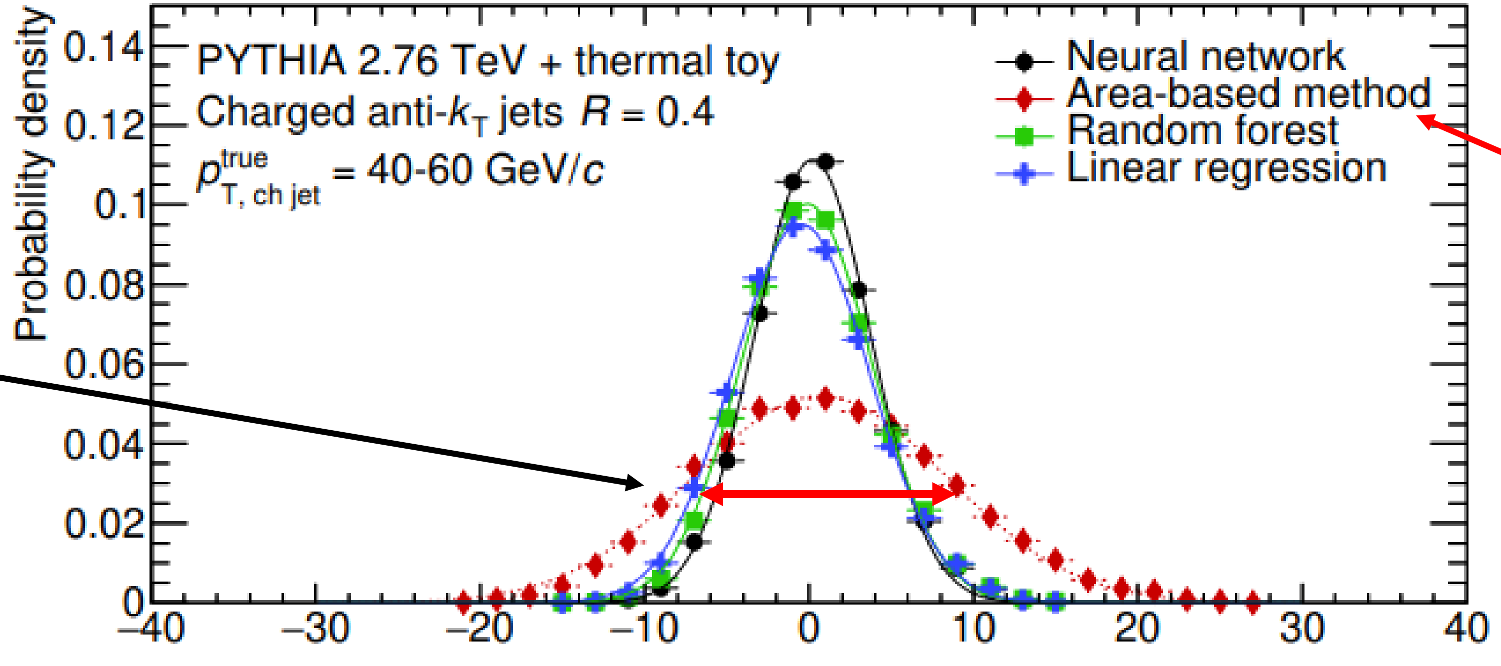
- Simplified picture
  - **Signal/Real particles** from hard scatterings
  - **Background particles** from soft processes
- **Background fluctuates in η, ϕ, event-to-event**
- **Jets** with **combinatorial background**
- **Jets composed of entirely combinatorial background**



Application of ML to Jet Background Subtraction

- Applications of ML to jet background subtraction achieve improved p_T resolution **at LHC energies**, particularly at low jet momentum.

- Can a neural network outperform traditional background subtraction methods **at RHIC energies?**



Width of δp_T determines momentum resolution

$p_T^{corr.} = p_T^{raw} - \rho A_{jet}$

Phys. Rev. C 99, 064904 (2019)

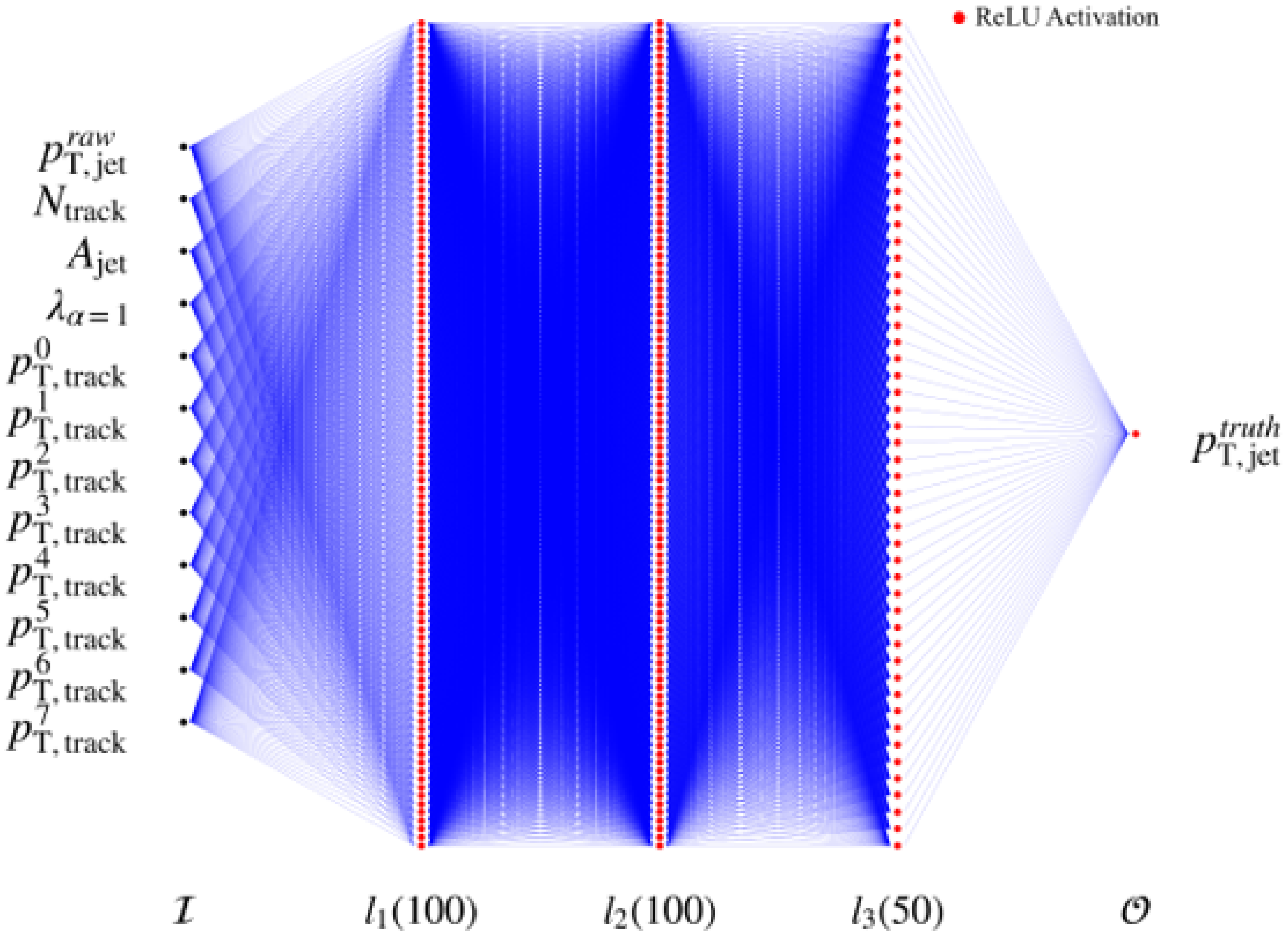
Jet Background Subtraction Study at RHIC Energies

- Analysis similar to Phys. Rev. C 99, 064904 (2019) (previous slide)
- **Signal Jets**: 200 GeV PYTHIA pp collisions
- **Heavy Ion Background**: TennGen tuned to 200 GeV AuAu
- Find charged anti-kT jets in Pythia + TennGen events and geometrically match to Pythia only jets
- Use matched PYTHIA jet momentum as ground truth: $\mathbf{p_{T,\,jet}^{truth} = p_{T,\,jet}^{pythia}}$

+ TennGen

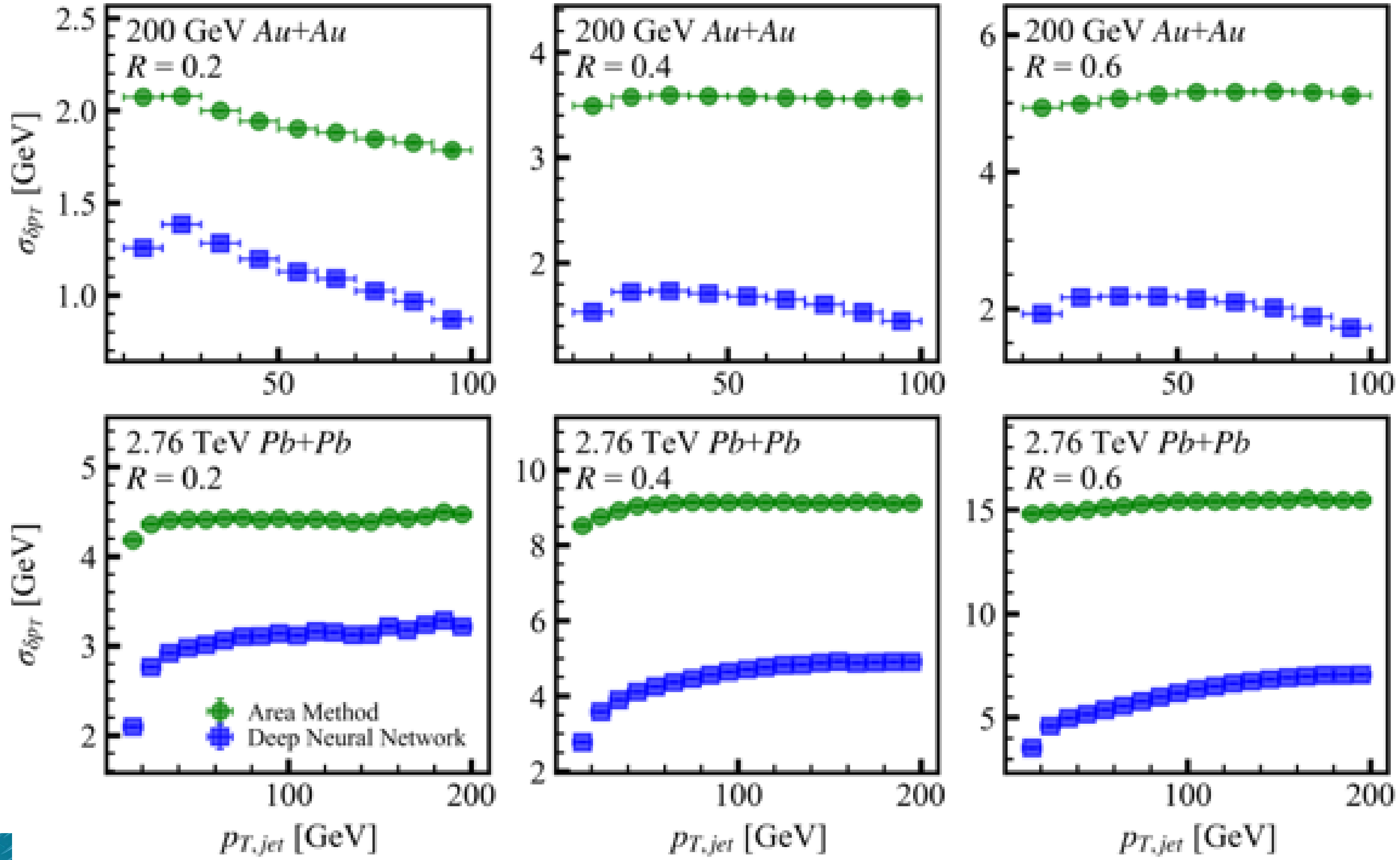
Neural Network Architecture

- Sequential dense network with 3 hidden layers
- Mean Squared Error Loss
- ADAM Optimizer
- ReLU activation functions
- 50/50 test/train split
- TensorFlow 2.10.0



Initial Jet Background Subtraction Study Results

- Width of δp_T from neural network is 2-3 times smaller for all jet p_T similar to the study at LHC energies



Initial Jet Background Subtraction Study Results

- Width of δp_T from neural network is 2-3 times smaller for all jet p_T similar to the study at LHC energies

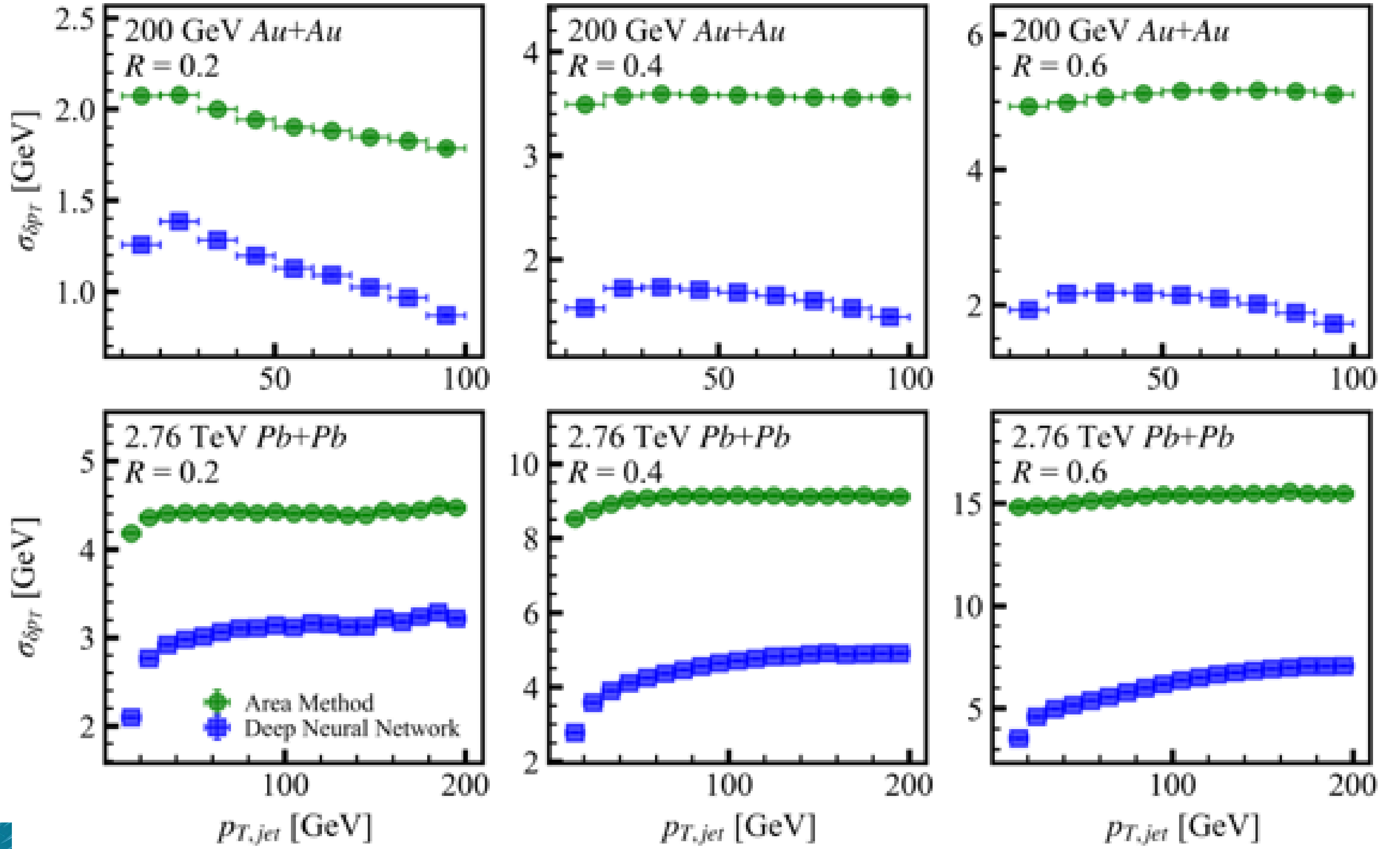
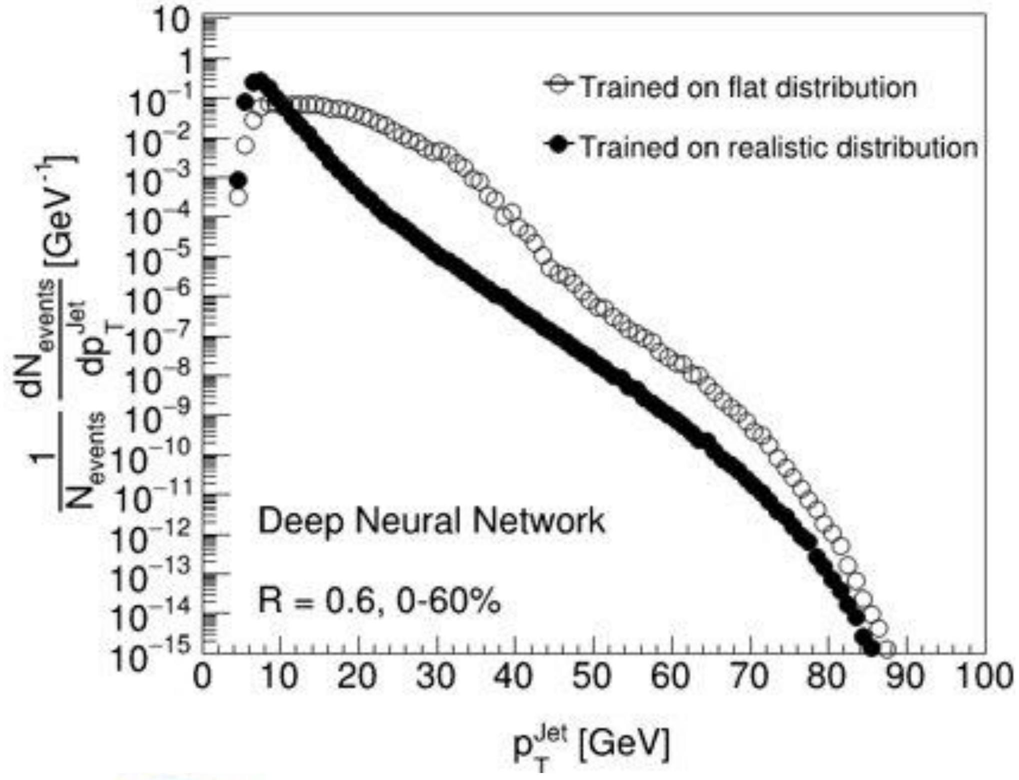
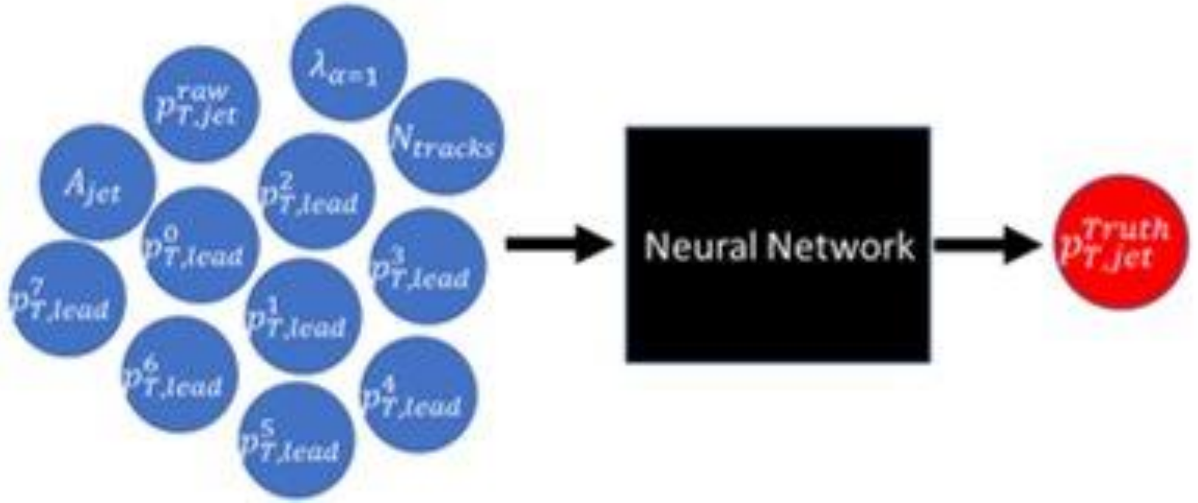


Image Credit: Ferdi Rizkiyanto

The Elephant in the Room

- Predictions biased by training
- Predictions only reliable within training phase space
- Offers little/no explanation for underlying physics of background subtraction

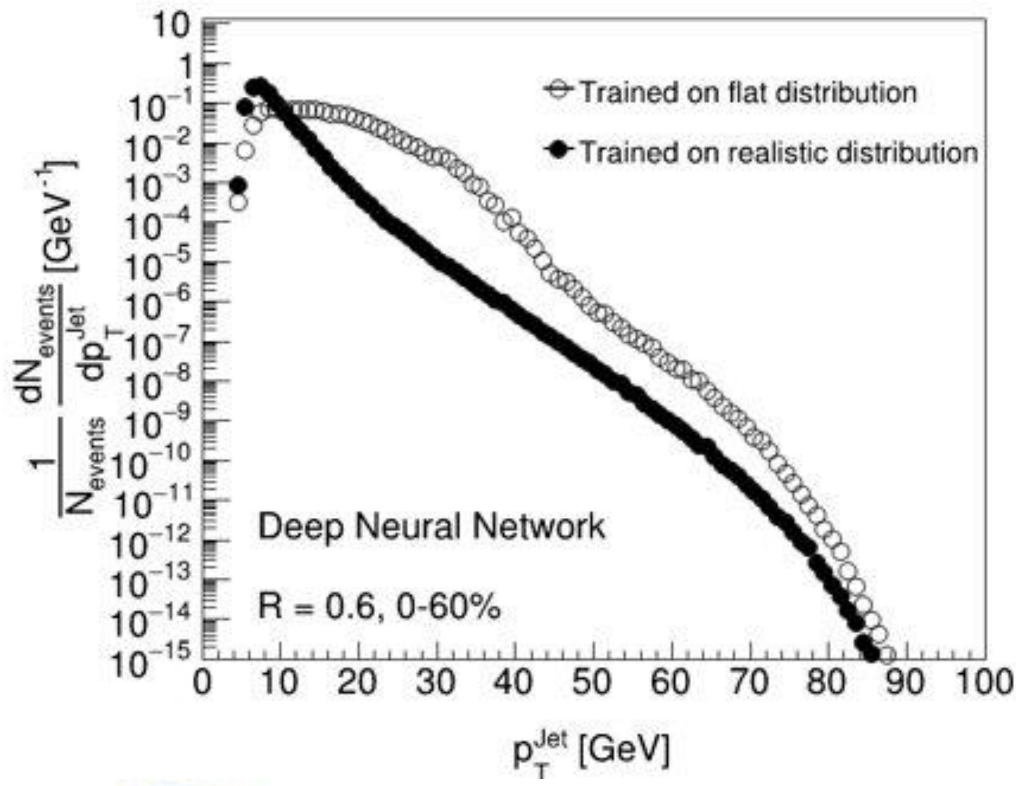


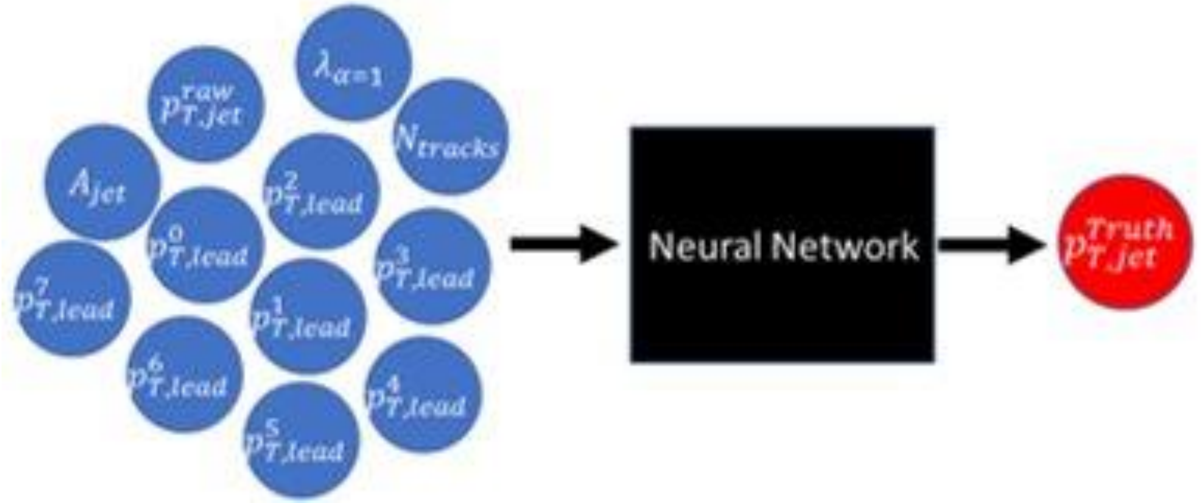


The Elephant in the Room

- Predictions biased by training
- Predictions only reliable within training phase space
- Offers little/no explanation for underlying physics of background subtraction
- By the way, these issues can be generally applicable...

(no free lunch theorem)





What we were trying to do, again?

Can a neural network outperform traditional jet background subtraction methods?

What we were trying to do, again?

Why does

~~Can~~ a neural network outperform traditional jet background subtraction methods?

Interpretable Machine Learning

1) Method must be equivalently applicable to data and simulation.

2) Predictions must be understood outside the range of training set.

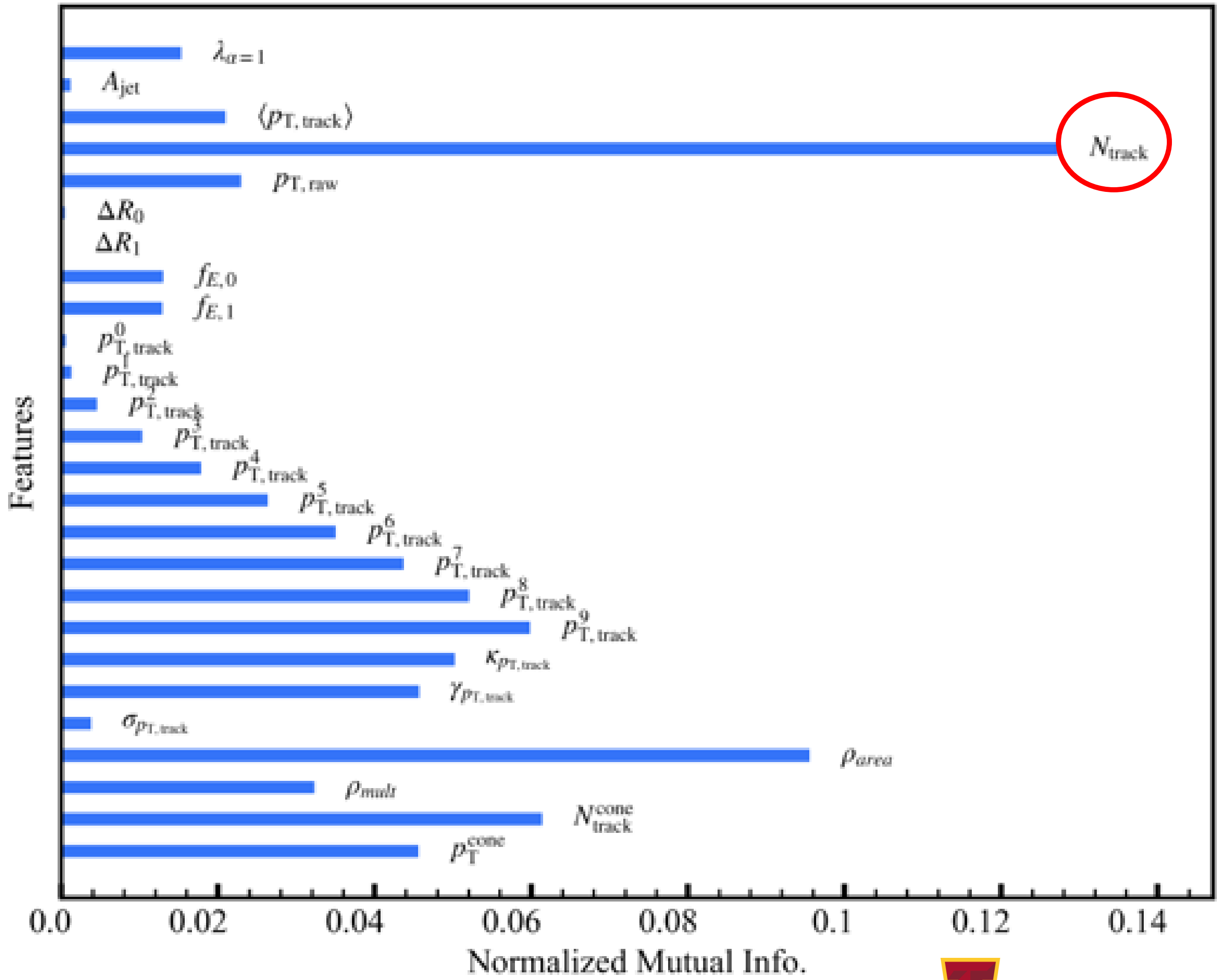
3) Systematic uncertainties can be assessed for predictions.

4) Learned relationships can be directly observed.

Principles listed in PhysRevC.108.L021901

2nd Look at All Those Input Features …

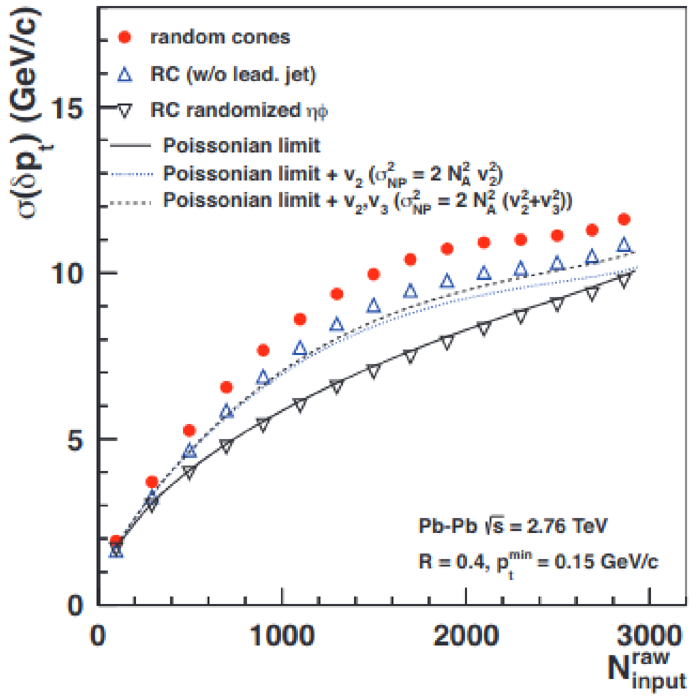
- N_{tracks} in the jet (multiplicity) has the largest mutual information to truth momentum
  - Mutual information: shared entropy between the joint prob P(X|Y) and the truth prob P(Y)
- Jet background fluctuations are driven by multiplicity
- JHEP 03 (2012) 053 (next slide)



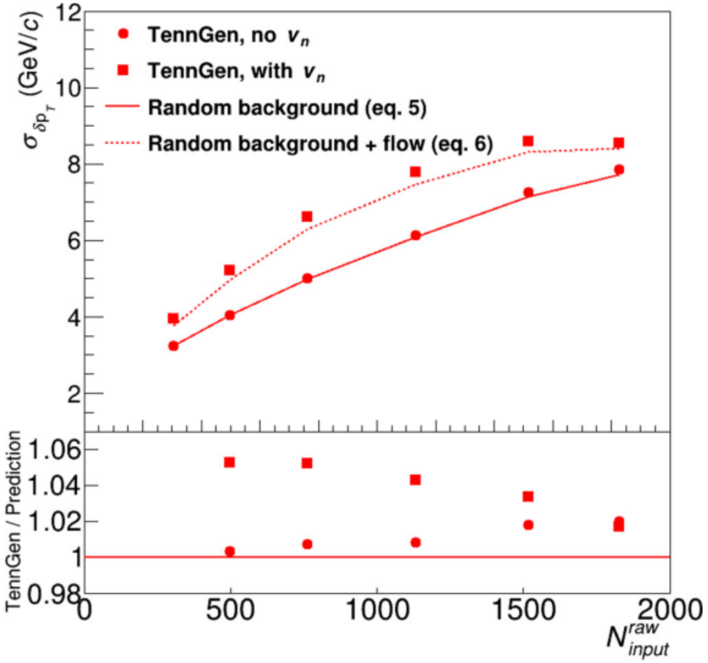
Multiplicity Method

- Jet background fluctuations well described in model
  - Assumes single particle pT spectra follows gamma distribution as in [Tannenbaum et. al.](#)

PbPb Data: JHEP 03 (2012) 053



TennGen PbPb Sim: PhysRevC.106.044915



N_A = avg. # of particles in cone

Assuming no flow (eq. 5)

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3 (eq. 6)

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$

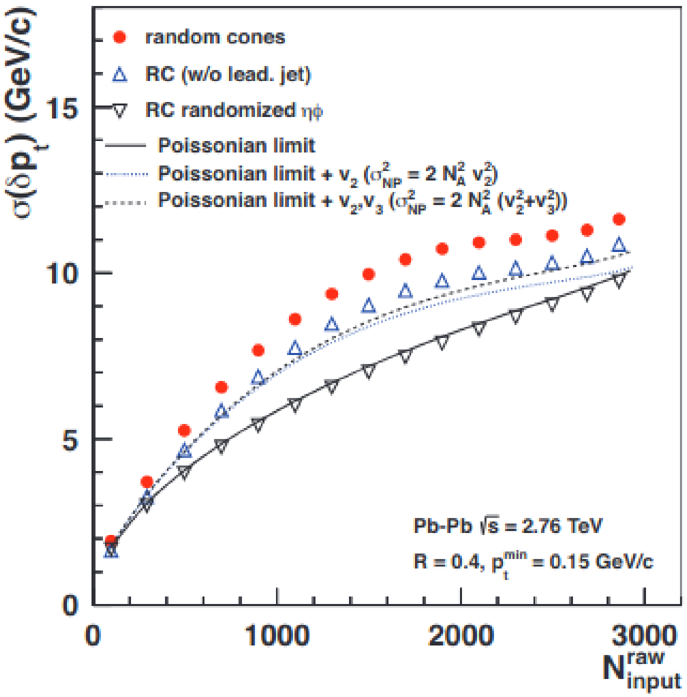
$$\delta p_t = \sum_i p_{t,i} - A \cdot \rho,$$

Measurement of jet background

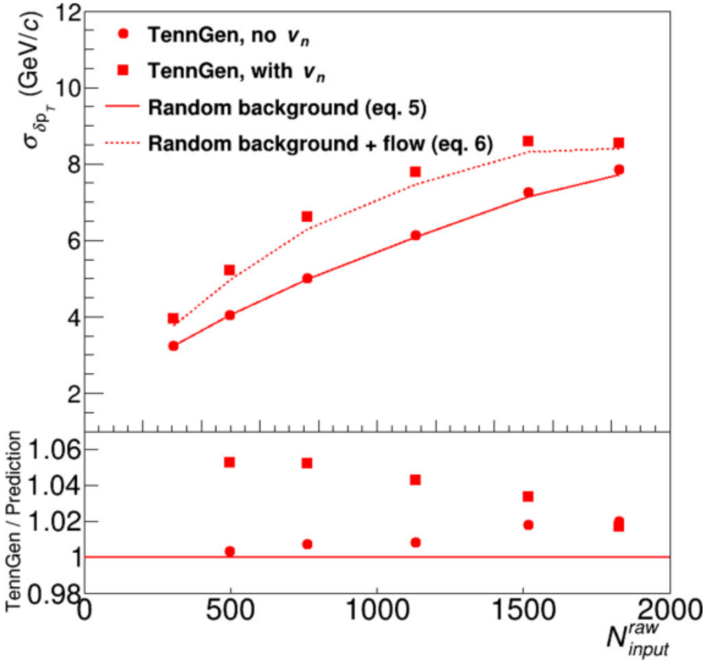
Multiplicity Method

- Jet background fluctuations well described in model
  - Assumes single particle pT spectra follows gamma distribution as in [Tannenbaum et. al.](#)

PbPb Data: JHEP 03 (2012) 053



TennGen PbPb Sim: PhysRevC.106.044915



N_A = avg. # of particles in cone

Assuming no flow (eq. 5)

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3 (eq. 6)

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$

$$\delta p_t = \sum_i p_{t,i} - A \cdot \rho,$$

Measurement of jet background

Multiplicity Method for background subtractions

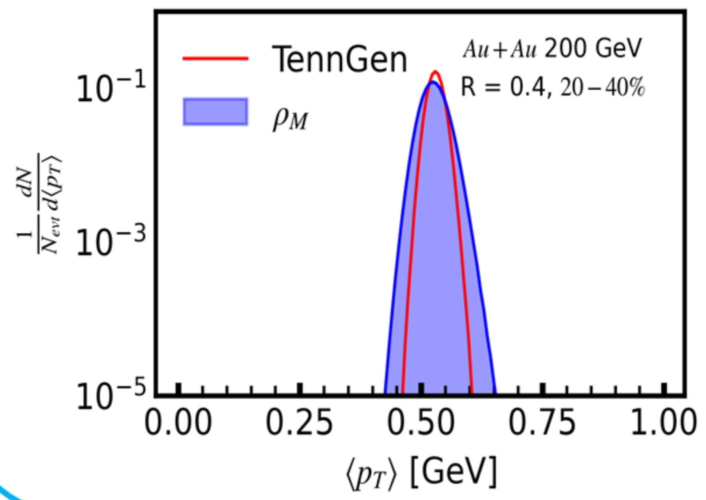
- Suggests that background subtraction technique could be:

$$p_T^{corr.} = p_T^{raw} - \langle p_T^{bkgd} \rangle \cdot N_{bkgd}$$

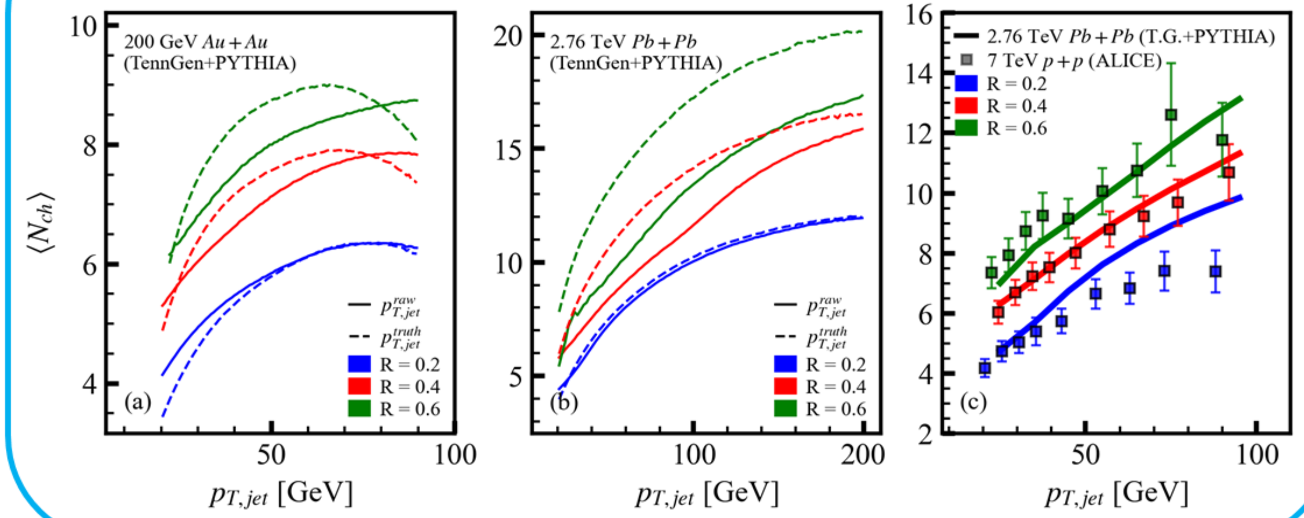
- Switch to:

$$p_T^{corr.} = p_T^{raw} - \rho_M \cdot \left(N - \langle N_{pythia} \rangle\right)$$

$$\rho_M \equiv median\left\{\langle p_{T,track} \rangle^{k_T\ jets}\right\} \approx \langle p_T^{bkdg} \rangle$$



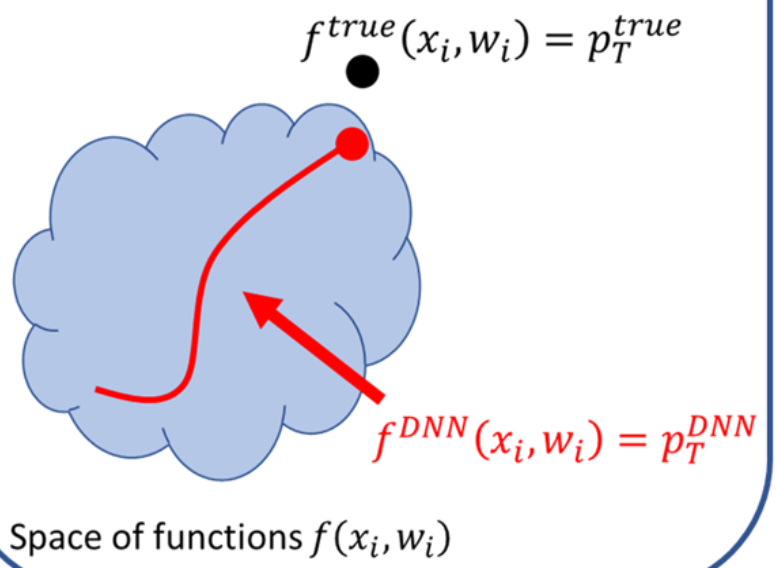
$$N_{bkgd} = N - N_{signal} \approx N - \langle N_{pthyia} \rangle$$



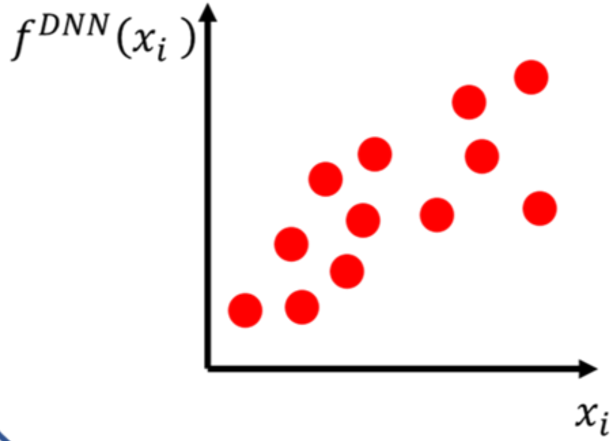
Connection Between Multiplicity Method and Neural Net

- Multiplicity Method physically well motivated and most important feature in DNN: can we learn more?

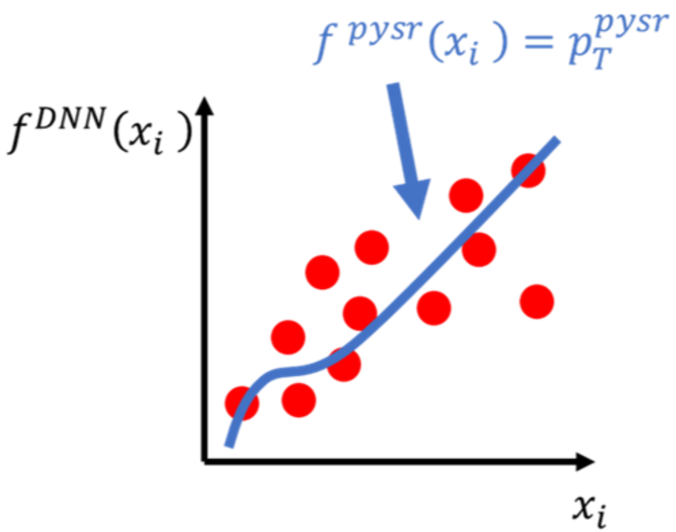
Use a deep neural network to map input jet features to the truth momentum.



Sample output of neural network across full range of input phase space.

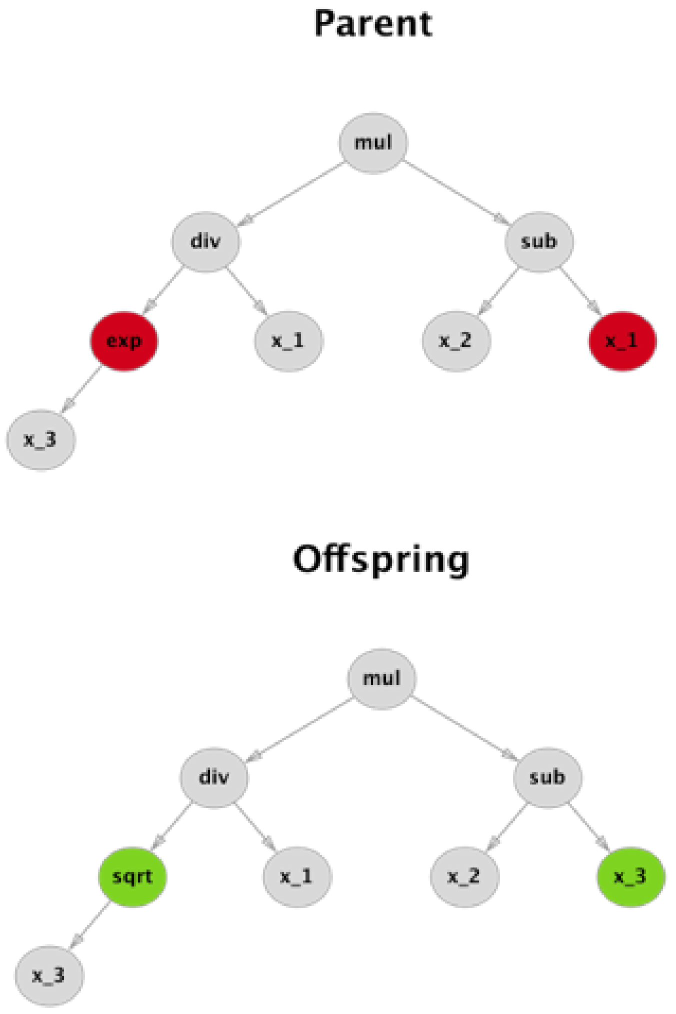


Fit jet features to neural network momentum prediction with symbolic regression.



PySR: Symbolic Regression

- PySR searches space of analytic expressions via multi-population evolutionary algorithm
- Genetic programming where 'traits' = operators.
- Each iteration creates new population with traits from each parent.
- Hyper-parameters:
  - # of generations: 50
  - # of populations: 22
  - # of individuals/population: 33
  - Max depth: 15
  - Loss function: MSE
  - Available genes: arithmetic/exponentiation/trig/sqrt/all input features
  - Pre-processing: 5 kNN input features
  - Output: 10 best equations

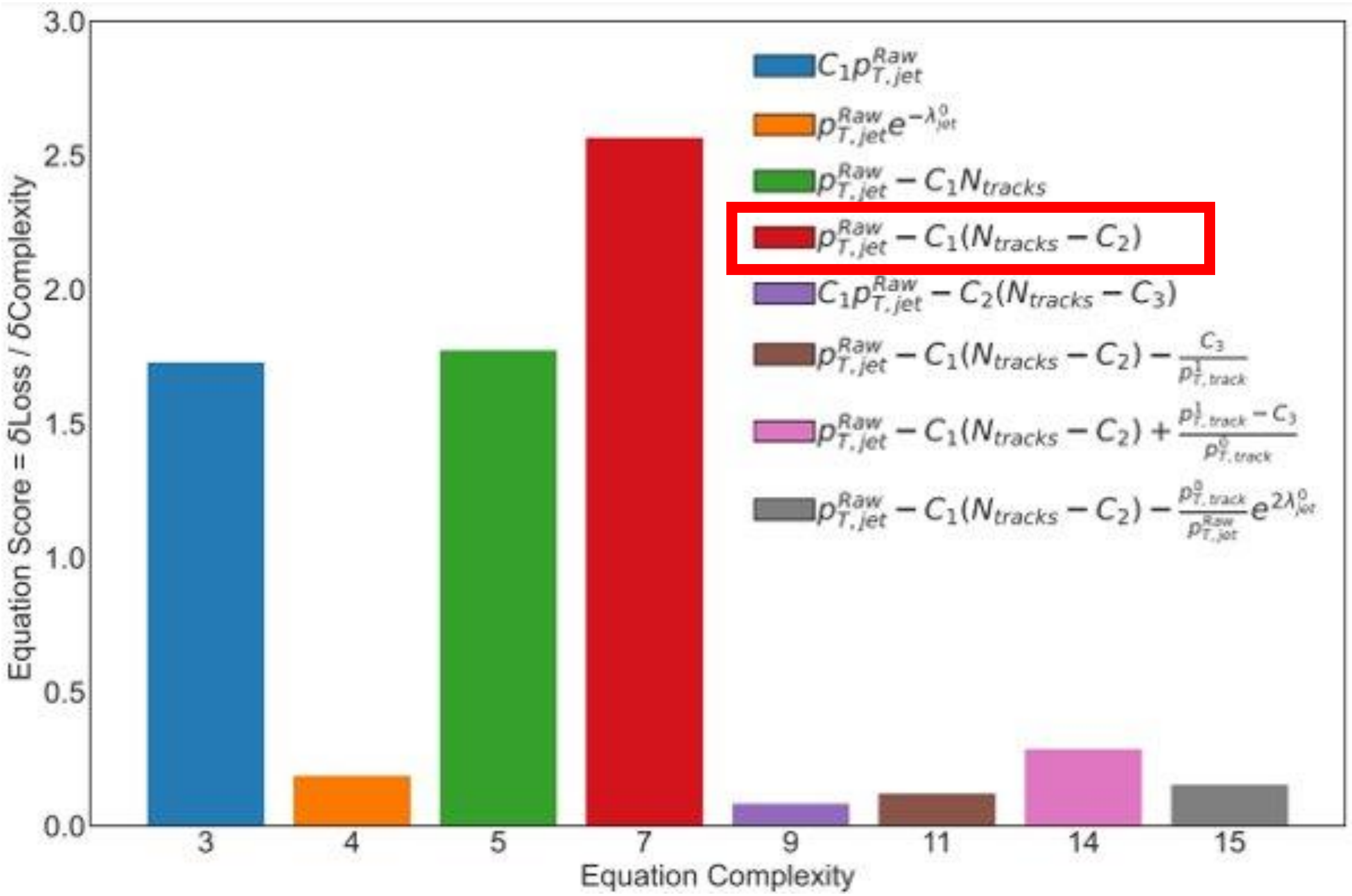


PySR Documentation:
https://arxiv.org/abs/2305.01582

https://arxiv.org/abs/2202.02306

PySR: Symbolic Regression - Results

- A linear form in N_{tracks} with 2 parameters is the least complex model that is the most accurate

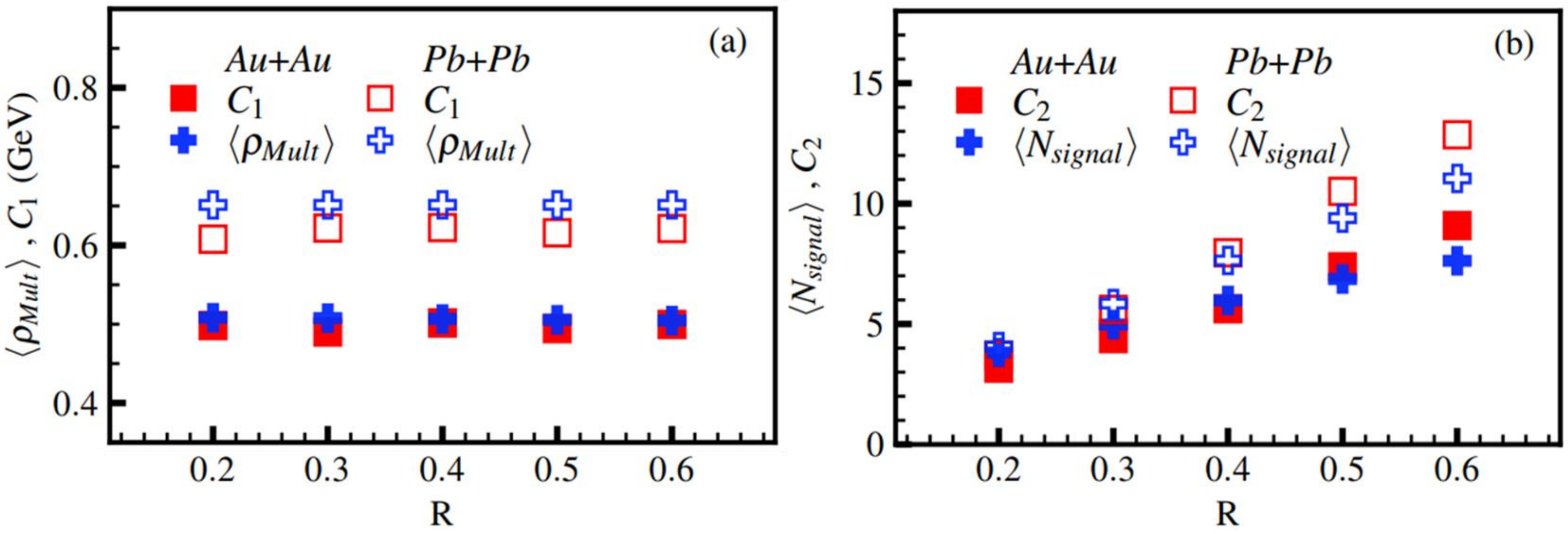


PySR: Symbolic Regression – Learned Parameters

- Constants learned by PySR are approximately the terms used in multiplicity background subtraction method.

$$p_T^{mult.} = p_T^{raw} - \rho_M * (N - \langle N_{pythia} \rangle)$$

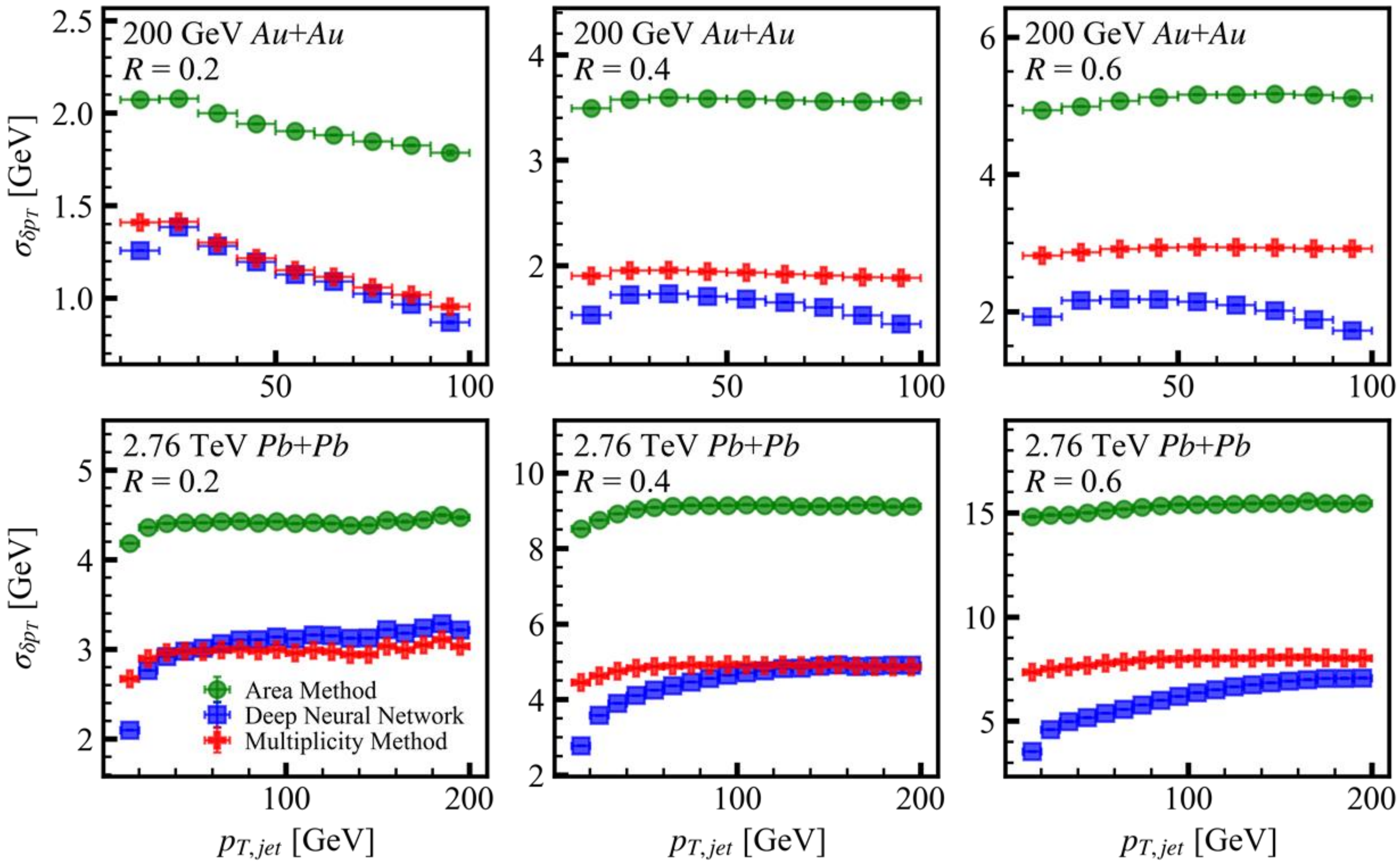
$$p_T^{psyr} = p_T^{raw} - C_1 * (N - C_2)$$



Results: Comparison of Jet Background Subtraction Methods

- Neural network picks up on multiplicity relationship

- Multiplicity method reproduces much of improvement achieved by neural network.



Conclusions/Discussions

- Using symbolic regression, the multiplicity method (2 free params) is shown to have a connection to the neural net (16K learned weights)
- The multiplicity method is physically well motivated, transparent, has quantifiable uncertainties, can be applied equally to simulation/data, and can be extrapolated
- However, this is not at all giving up on ML
- **Use ML as a tool to better understand** – pysr is a great example!

Authors/Thank Yous

Tanner Mengel

Patrick Steffanic

Charles Hughes

Antonio Da Silva

Christine Nattrass

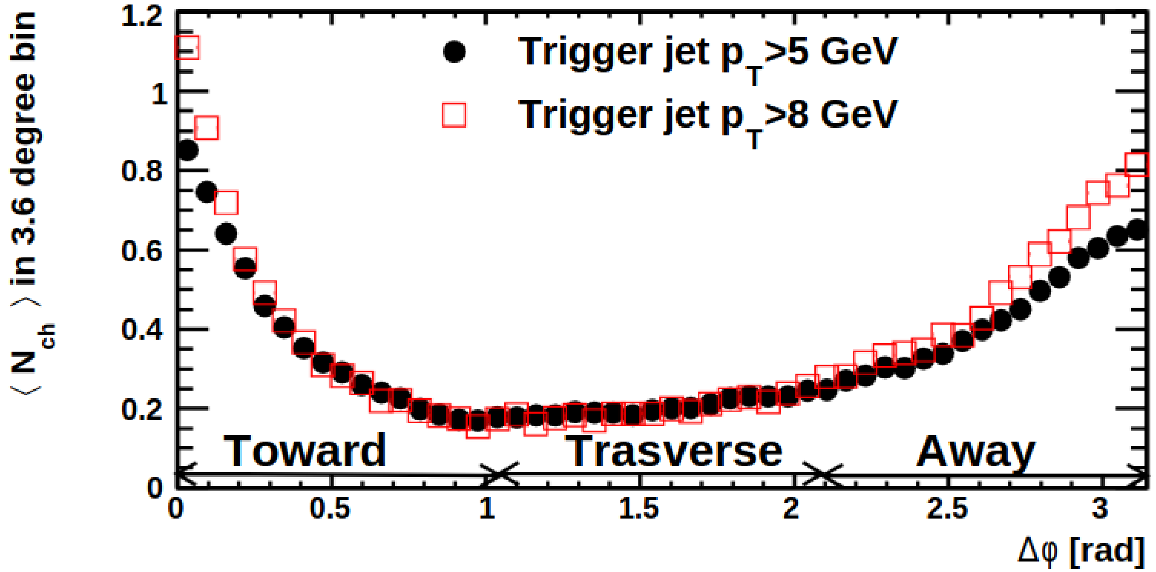
PhysRevC.108.L021901

Study also featured on PySR website!

Backup

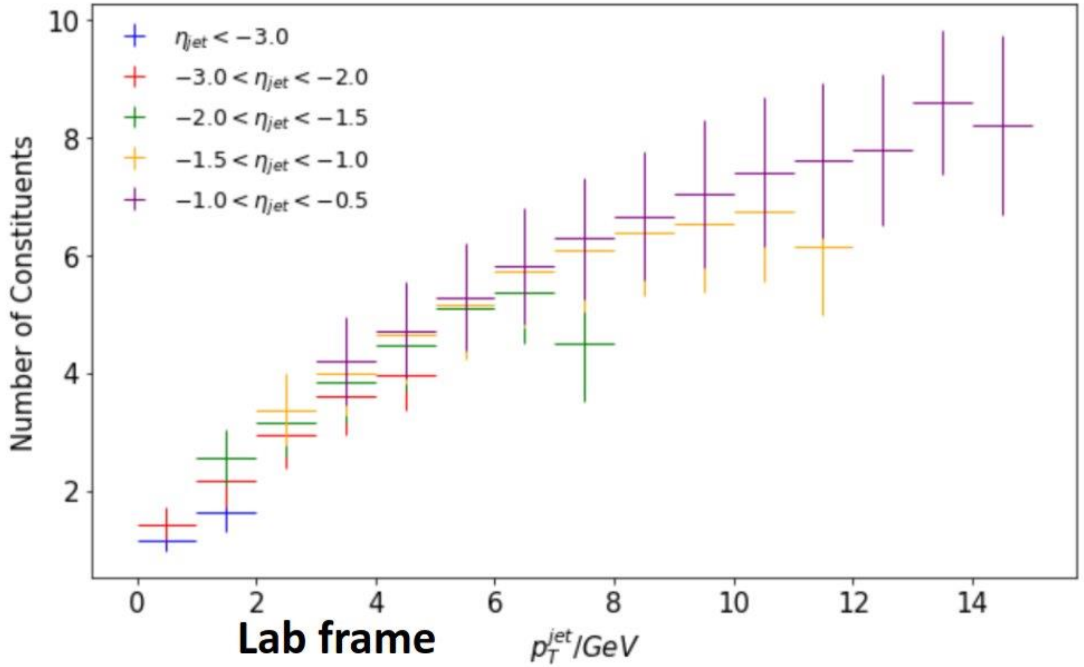
EIC jet multiplicity

- E+P jets – multiplicity < 1 particle



Plot from Brian Page

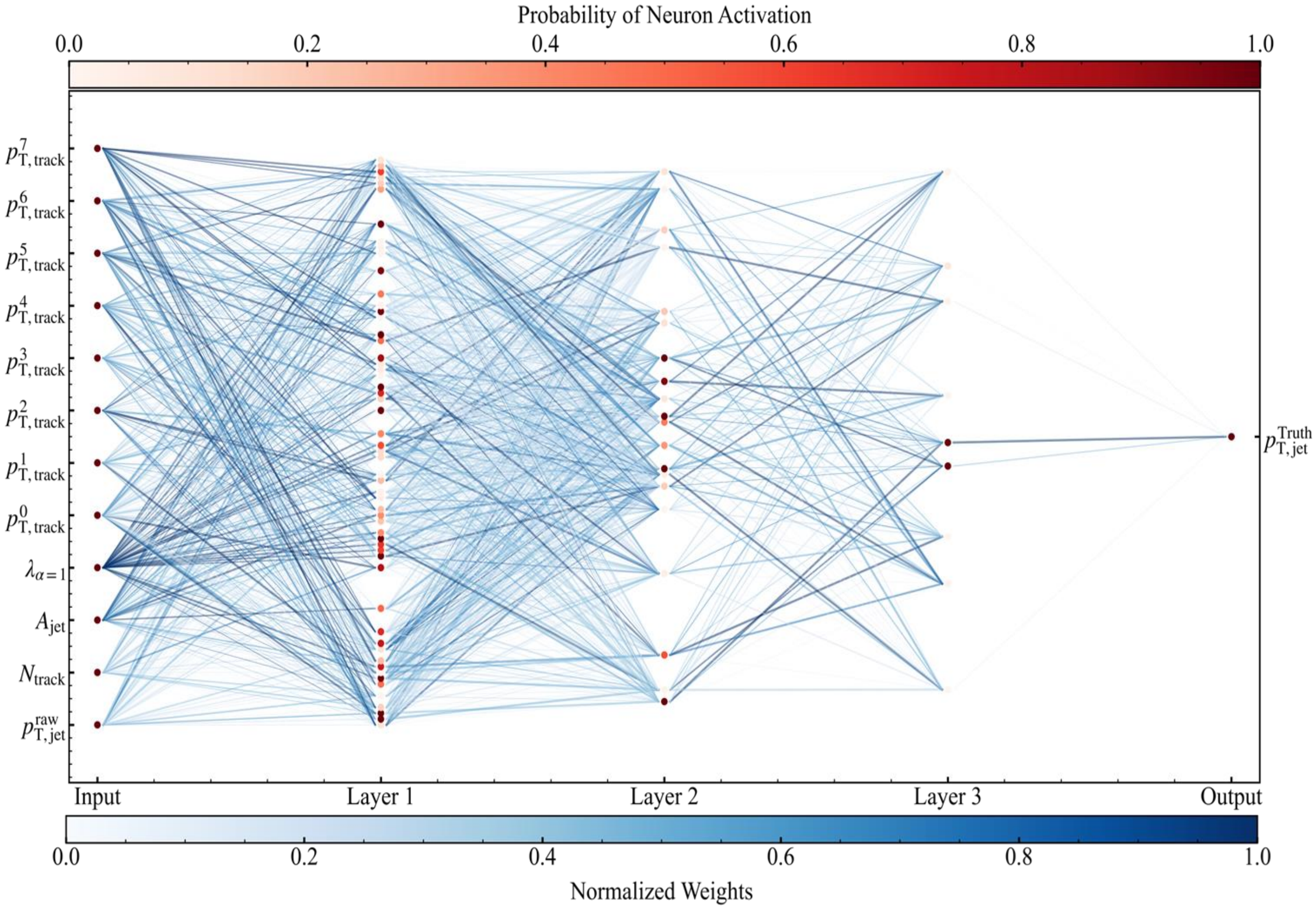
- E + A jets – multiplicity 5-10 particles



Plot from Miguel Arriata

Complexity vs. Accuracy

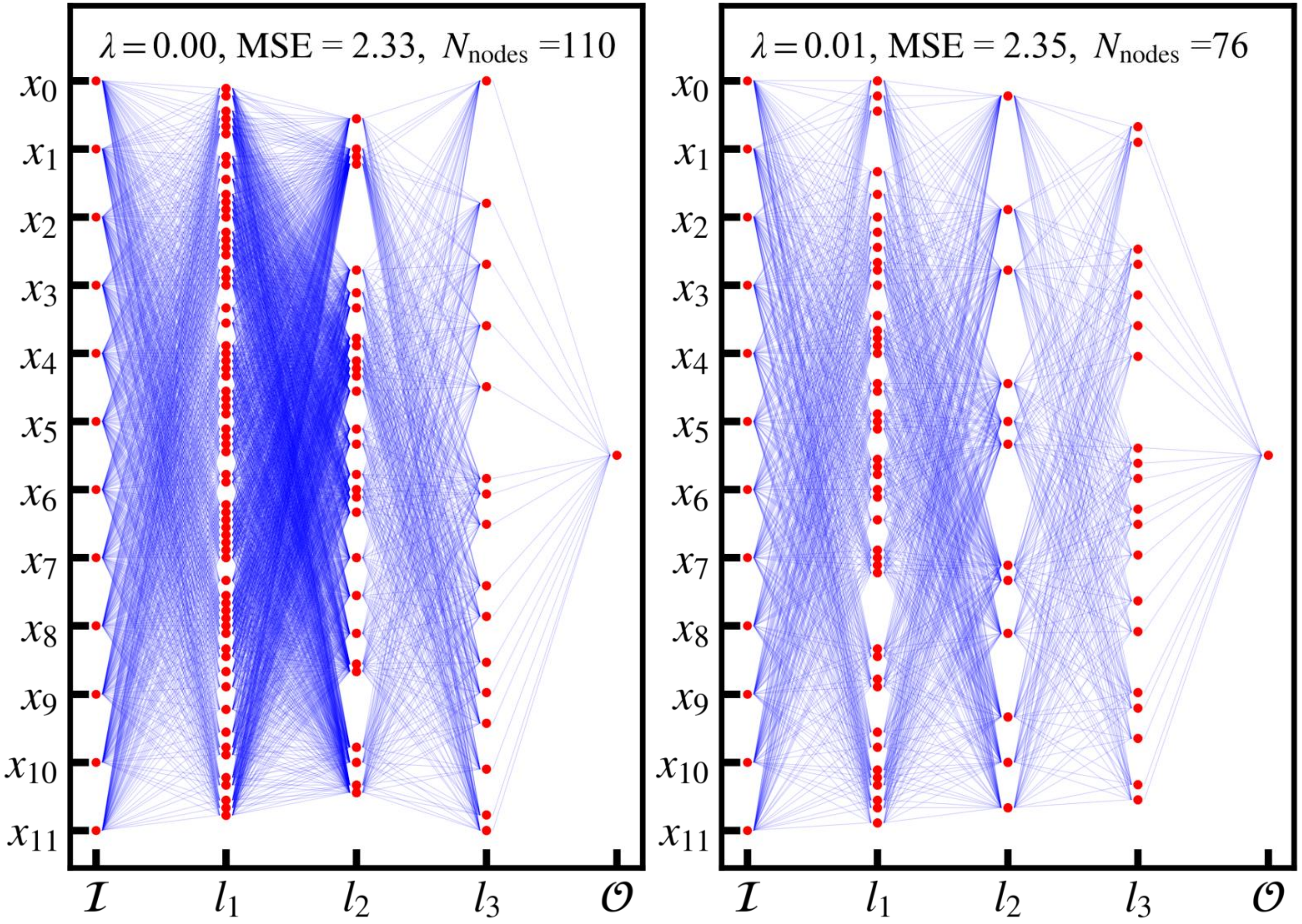
- Multiplicity method has 2 parameters
- Neural network has over 16,000 learned weights



Reducing Redundancy

- Introduce kernel regularization term to loss function.

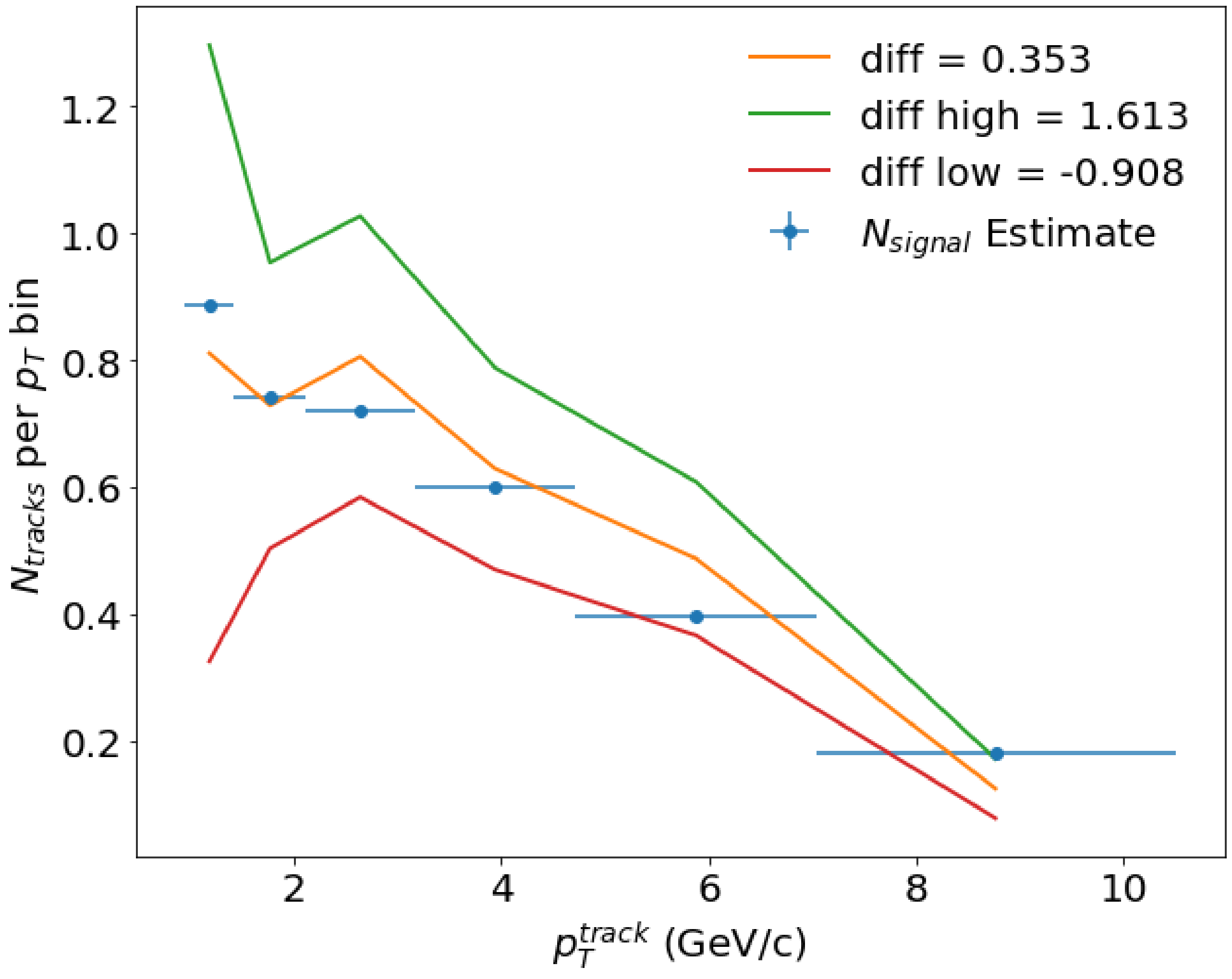
$$\mathcal{L} = \frac{1}{n}\sum_{i}^{n} |p_{T,jet}^{pred.} - p_{T,jet}^{truth}|^2 + \boxed{\lambda \|\boldsymbol{W}\|^2}$$



Uncertainty on <N_{Pythia}>

- Enhancement in jet multiplicity can be estimated via measured jet fragmentation functions

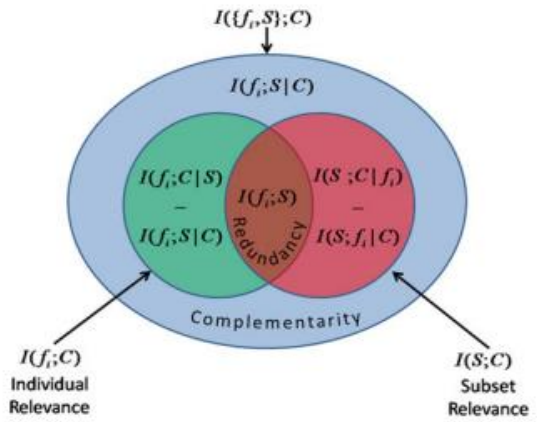
- At most 1 particle difference at low jet momentum

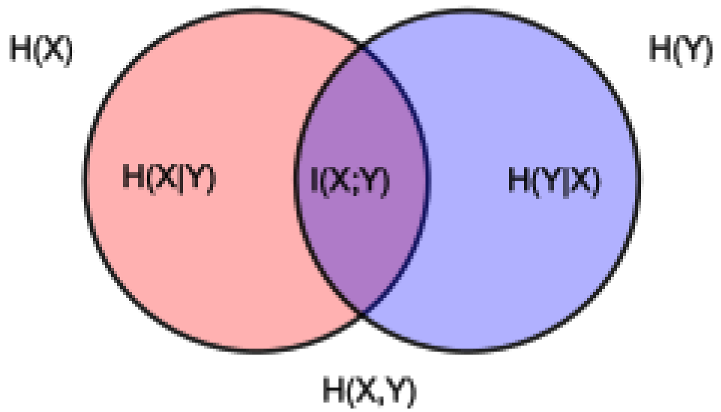


Mutual Information

- 'Distance' between joint probability and individual probability between two random variables
- The mutual information provides a measure of the relevance an input feature has in predicting the target variable

$$I(X;Y) \equiv H(X) - H(X|Y)$$



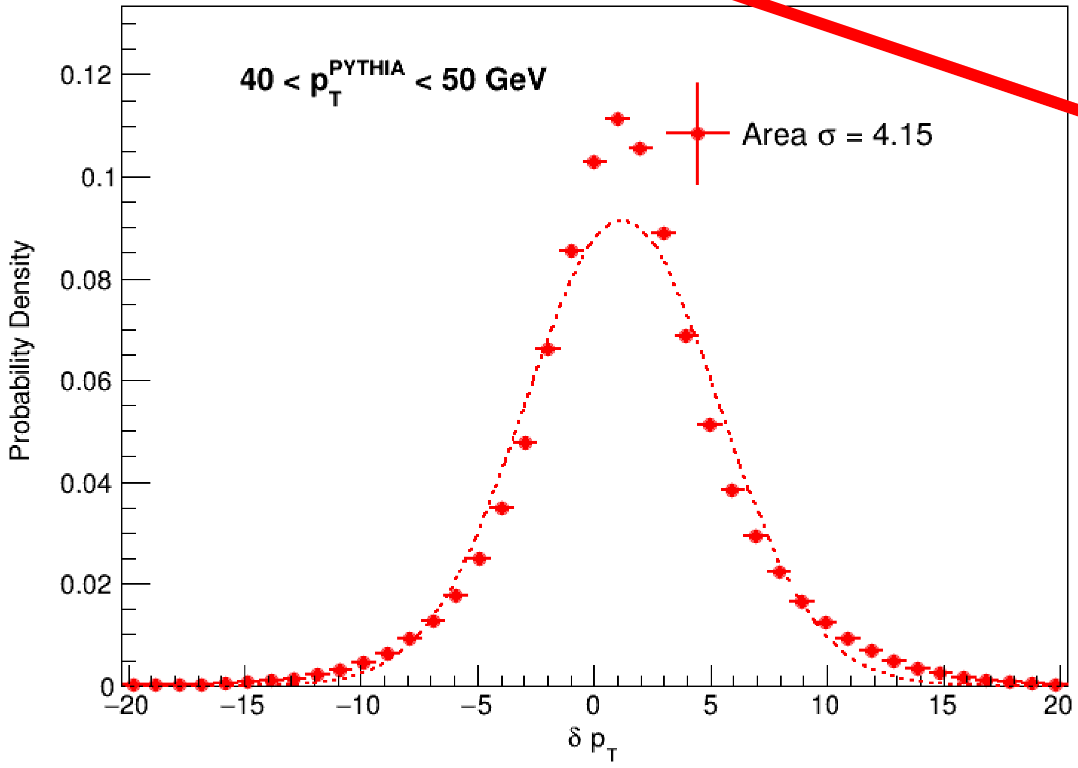


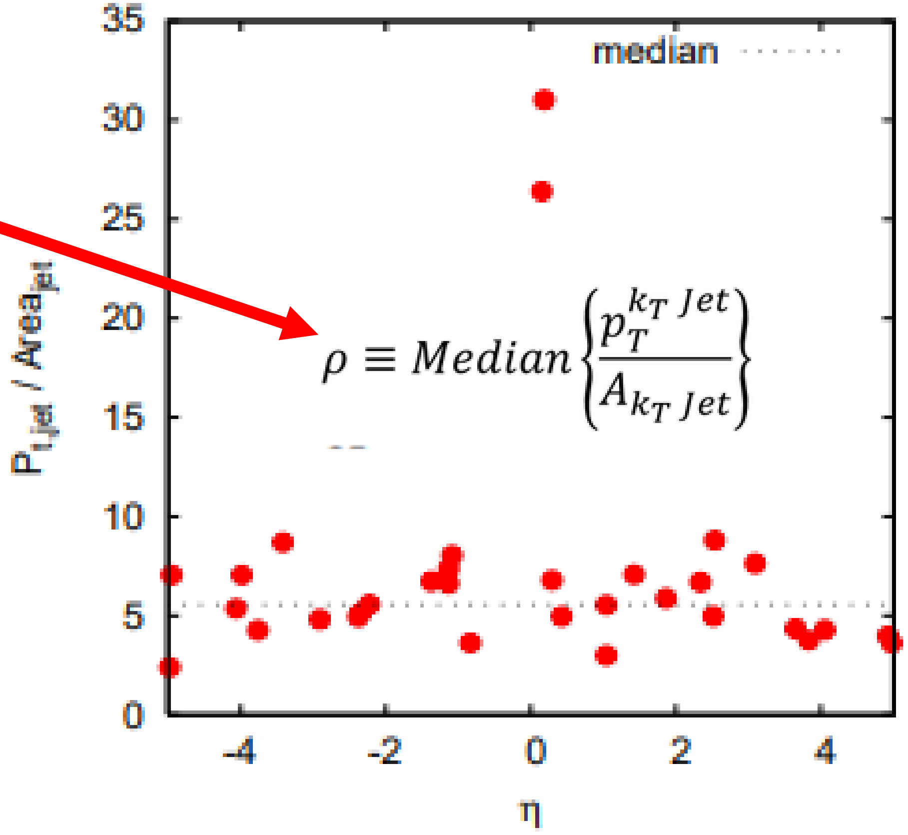
https://link.springer.com/article/10.1007/s00521-013-1368-0

Area- based subtraction method recap

- Area-based background subtraction method

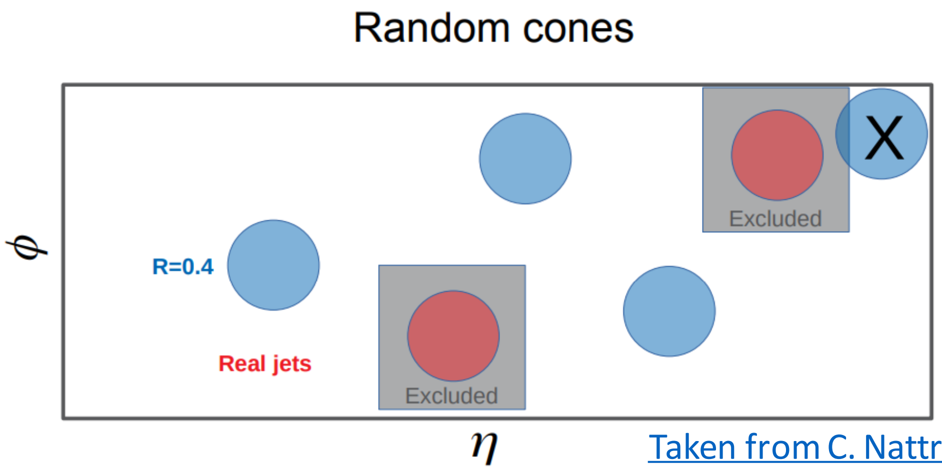
$p_T^{corr} = p_T^{raw} - rho * A$



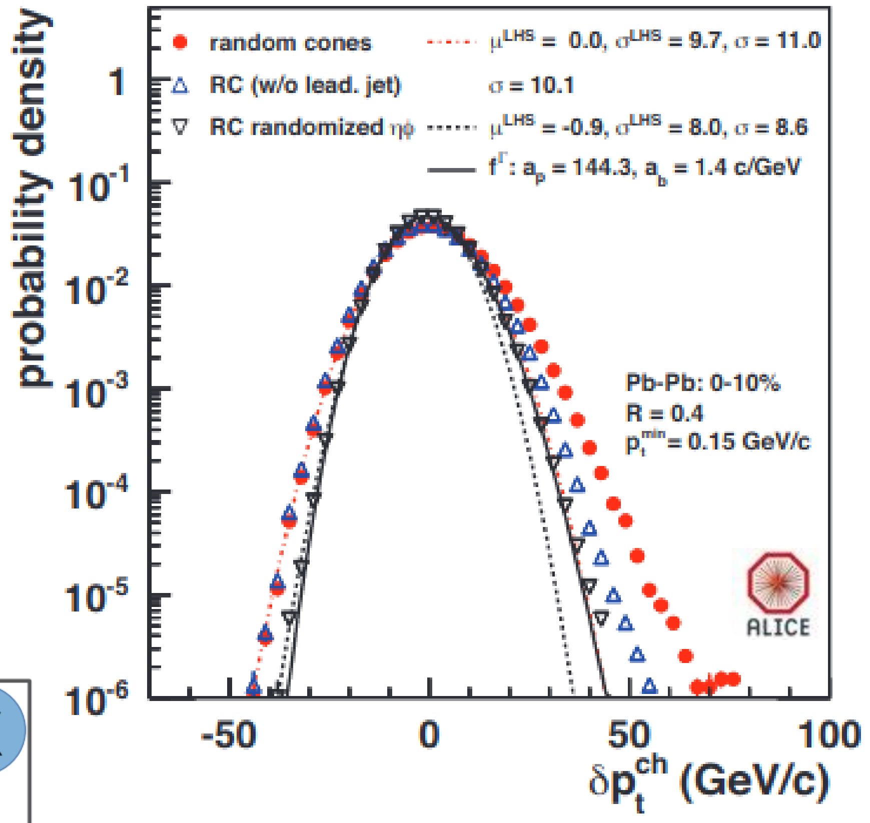


Measuring Background Fluctuations

- Following the analysis in *JHEP 03 (2012) 053*
  - Estimate rho
    - k_T jet finder
    - $\rho = \text{median}(p_T/A)$
  - Estimate jet background fluctuations
    - Draw random cone
    - $\delta p_T = p_T^{cone} - \rho * A^{cone}$
    - CalcuLate $\sigma(\delta p_T)$ - next slide



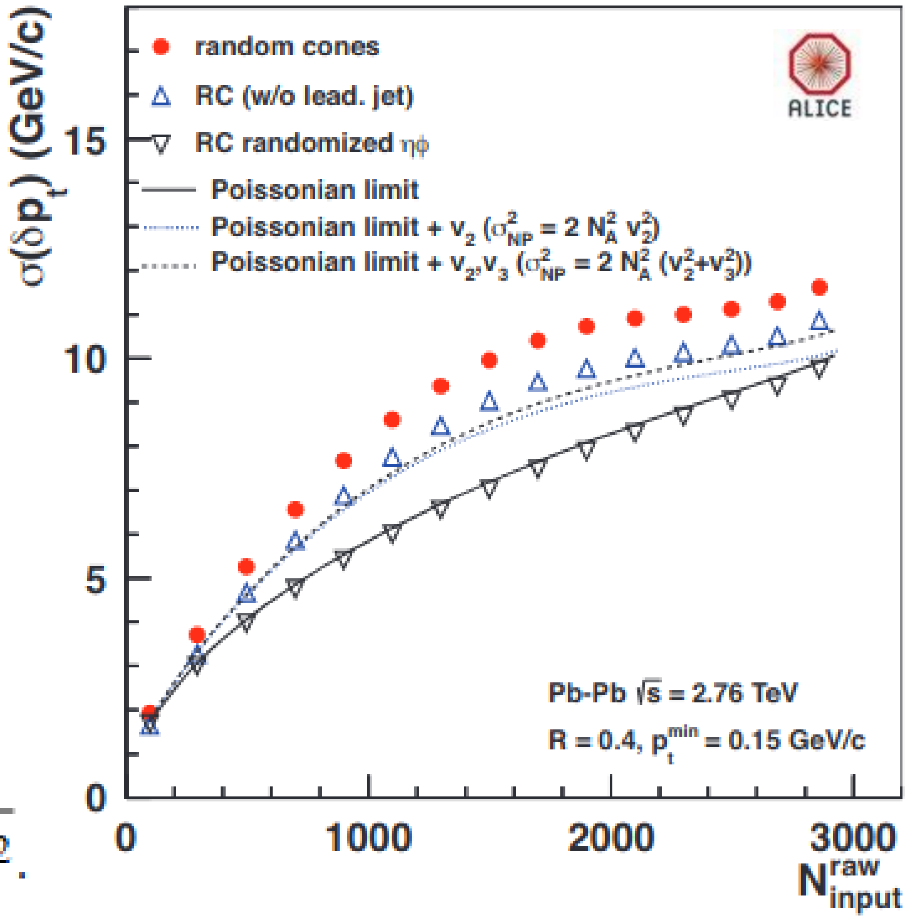
Taken from C. Nattrass



Taken from *JHEP 03 (2012) 053*

Characterizing Background Fluctuations

- Following the analysis in *JHEP 03 (2012) 053*
  - Calculate $\sigma(\delta p_T)$
  - Compare to model as in ([Tannenbaum et. al.]())
    - Assumes single particle p_T spectrum is gamma distribution, then:



N_A = # of particles in cone

Assuming no flow

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$

Characterizing Background Fluctuations

- Following the analysis in *JHEP 03 (2012) 053*
  - Calculate $\sigma(\delta pT)$
  - Compare to model as in (Phys.Lett.B 498 (2001) 29-34)
    - Assumes single particle pT spec

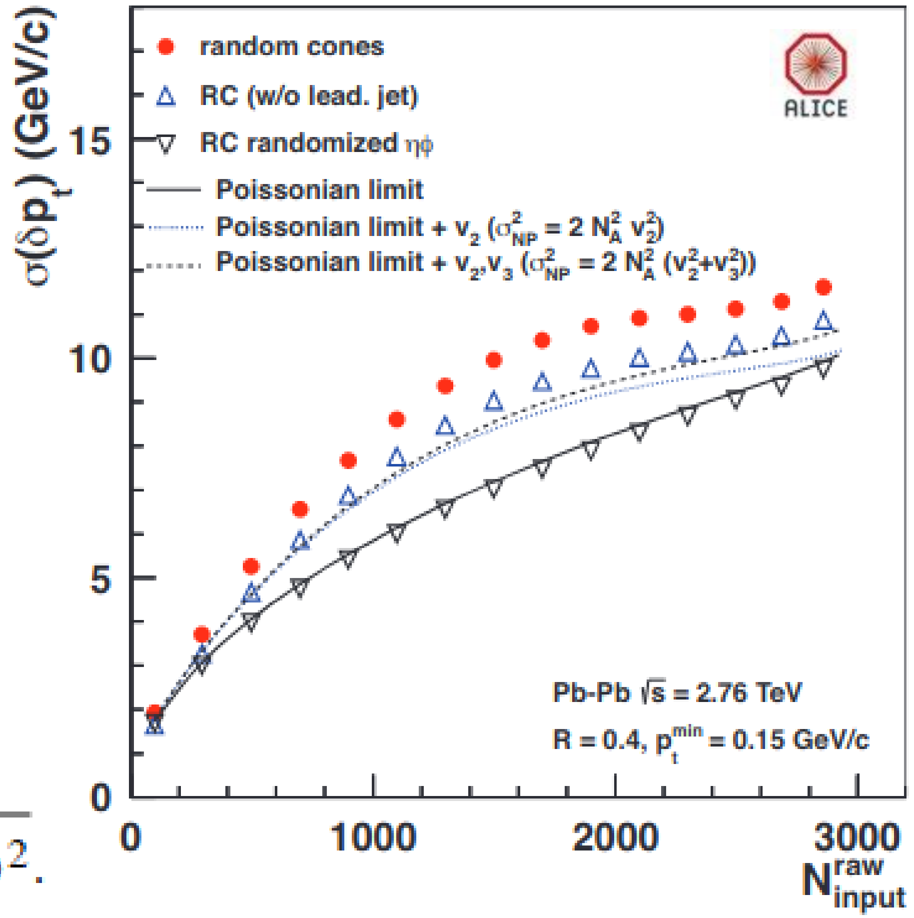
Width of N-fold convolution of gamma pT spectrum term

Poissonian fluctuations term

N_A = # of particles in cone

Assuming no flow $\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$

Accounts for v2/v3 $\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$



Characterizing Background Fluctuations

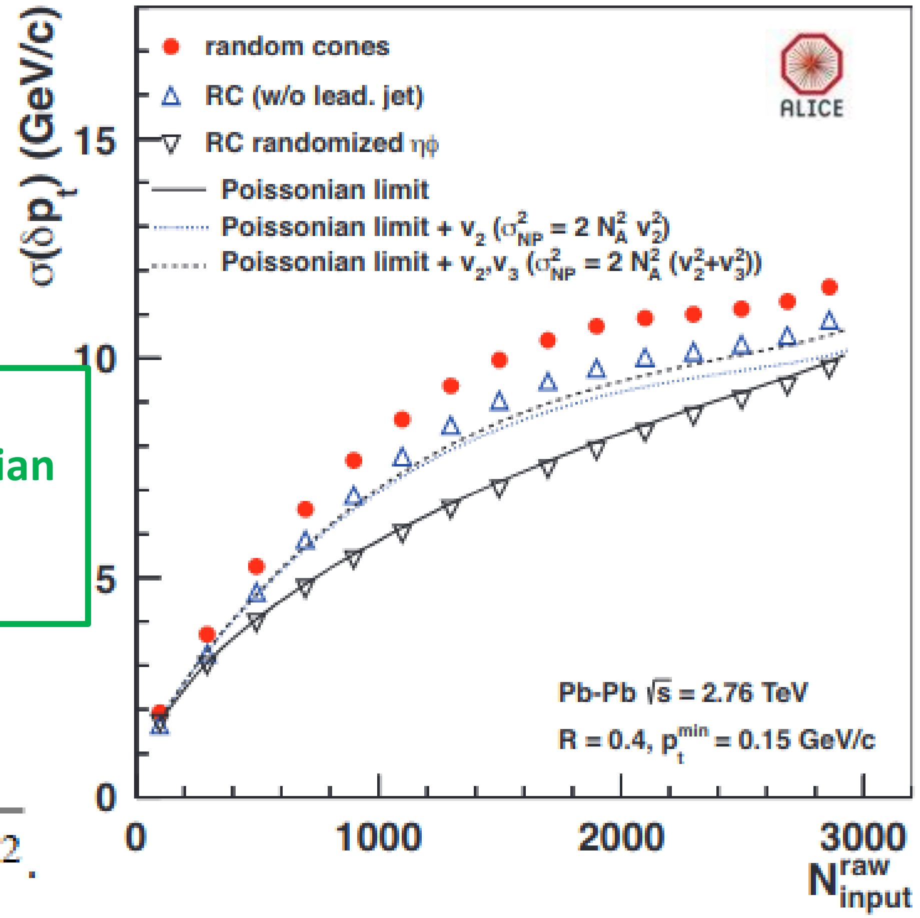
- Following the analysis in *JHEP 03 (2012) 053*
  - Calculate $\sigma(\delta pT)$
  - Compare to model as in (Phys.Lett.B 498 (2001) 29-34)
    - Assumes single particle pT spectrum is gamma distribution, then:

N_A = # of particles in cone

Non-Poissonian term (v2/v3)

Assuming no flow $$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3 $$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$



$$\sigma^2_{NP}(N_A) \approx 2v_2^2 N_A^2 \quad \text{OR} \quad (\sigma^2_{NP}(N_A) \approx 2N_A^2(v_2^2 + v_3^2))$$

Background Fluctuations - Model Studies

- Following the analysis in *JHEP 03 (2012) 053*

- Some questions to ask:

    - What can we learn from this background characterization in simple model studies ?

    - What implications does this have for background mitigation in jet observables in data ?

["Model studies of fluctuations in the background for jets in heavy ion collisions"]

Phys. Rev. C **106**, 044915 – Published 31 October 2022

Charles Hughes, Antonio Carlos Oliveira da Silva, and Christine Nattrass

Charles Hughes

Antonio Da Silva

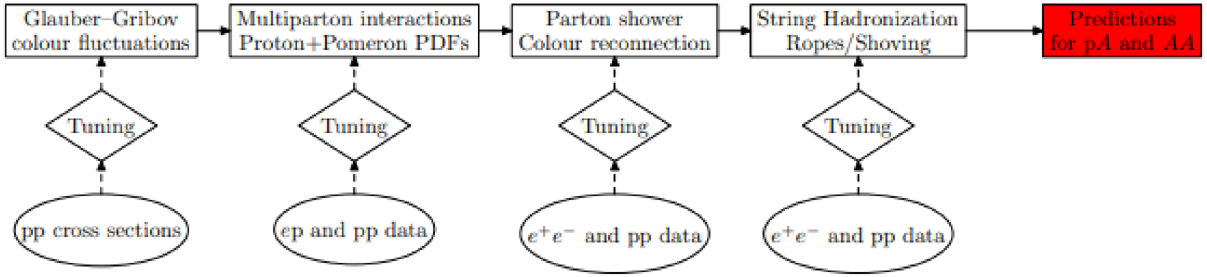
Christine Nattrass

Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]

- Looking at 2 models
  - Angantyr Pythia - [arXiv:1806.10820]
    - MPI/Diffractive Excitation
  - TennGen - [(github)]
    - Next Slide



Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

- TennGen:

What TennGen is for:

a) Computationally cheap way to generate particles with realist pT spectrum and flow as in heavy ion collisions (and NO OTHER correlations)

b) Understanding how a realistic heavy ion background affects jet finders/jet observables

c) Development of background subtraction/mitigation techniques

d) Seeing how analysis depends on background with/without v_1, v_2, v_3, etc...

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915](#)

- TennGen:

What TennGen is for:

a) Computationally cheap way to generate particles with realist pT spectrum and flow as in heavy ion collisions (and NO OTHER correlations)

b) Understanding how a realistic heavy ion background affects jet finders/jet observables

c) Development of background subtraction/mitigation techniques

d) Seeing how analysis depends on background with/without v_1, v_2, v_3, etc...

What TennGen is NOT for:

a) NOT a replacement for physics based generators (HIJING, AMPT, JEWEL, etc...)

b) NOT for jet/background physics interaction model studies (e.g. quenching/energy loss/back-reaction)

c) NOT for testing hydro/flow models

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

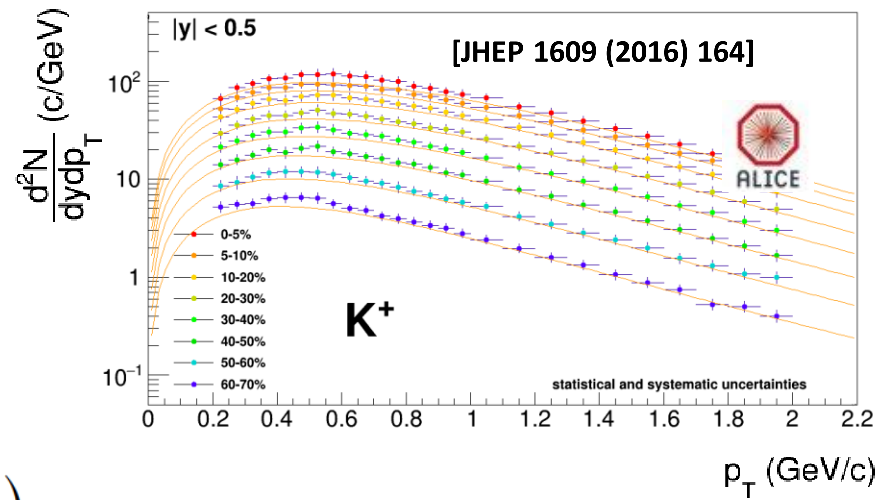
- TennGen:
  - Particle generator meant to simulate $\pi^{+/-/0}$, $K^{+/-}$, $p, \bar{p}$ in 2.76 TeV PbPb collisions (0-5% : 40-50 %)
  - Particle p_T according to fits of data to Boltzmann-Gibbs Blast Wave

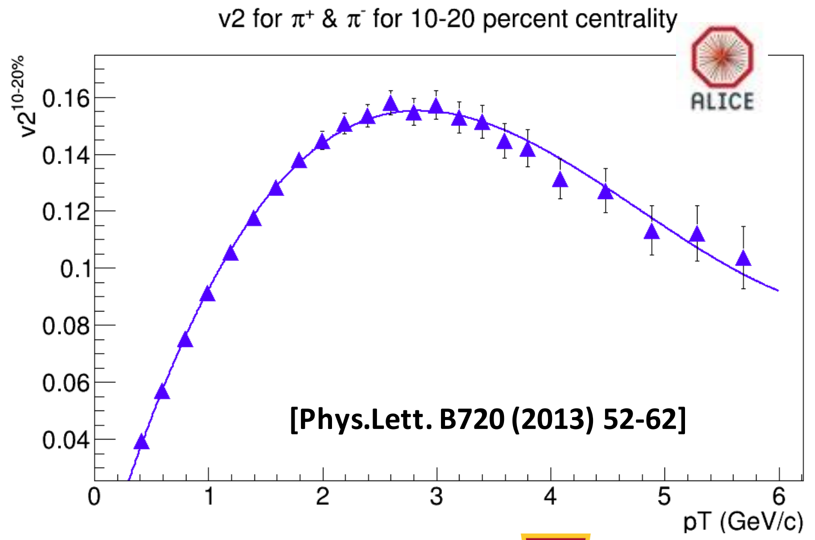
$$\frac{d^2N}{dp_T dy} = N p_T \int_0^1 r' dr' \left(\sqrt{m^2 + p_T^2}\right) \times I_0\left(\frac{p_T \sinh\left[\tanh^{-1}\left(\beta_s r'^n\right)\right]}{T_{\text{kin.}}}\right) \times K_1\left(\frac{\sqrt{m^2 + p_T^2}\cosh\left[\tanh^{-1}\left(\beta_s r'^n\right)\right]}{T_{\text{kin.}}}\right)$$

  - $v_n(p_T)$ from polynomial fits to data (v_1 : v_5)
  - Particle ϕ from Fourier Sum

$$\frac{dN}{d\phi} = \frac{N_0}{2\pi}\left(1 + \sum_{n=1}^{5} 2v_n \cos[n(\phi - \Psi_n)]\right)$$

  - Particle η Uniform (mid-rapidity approximation)

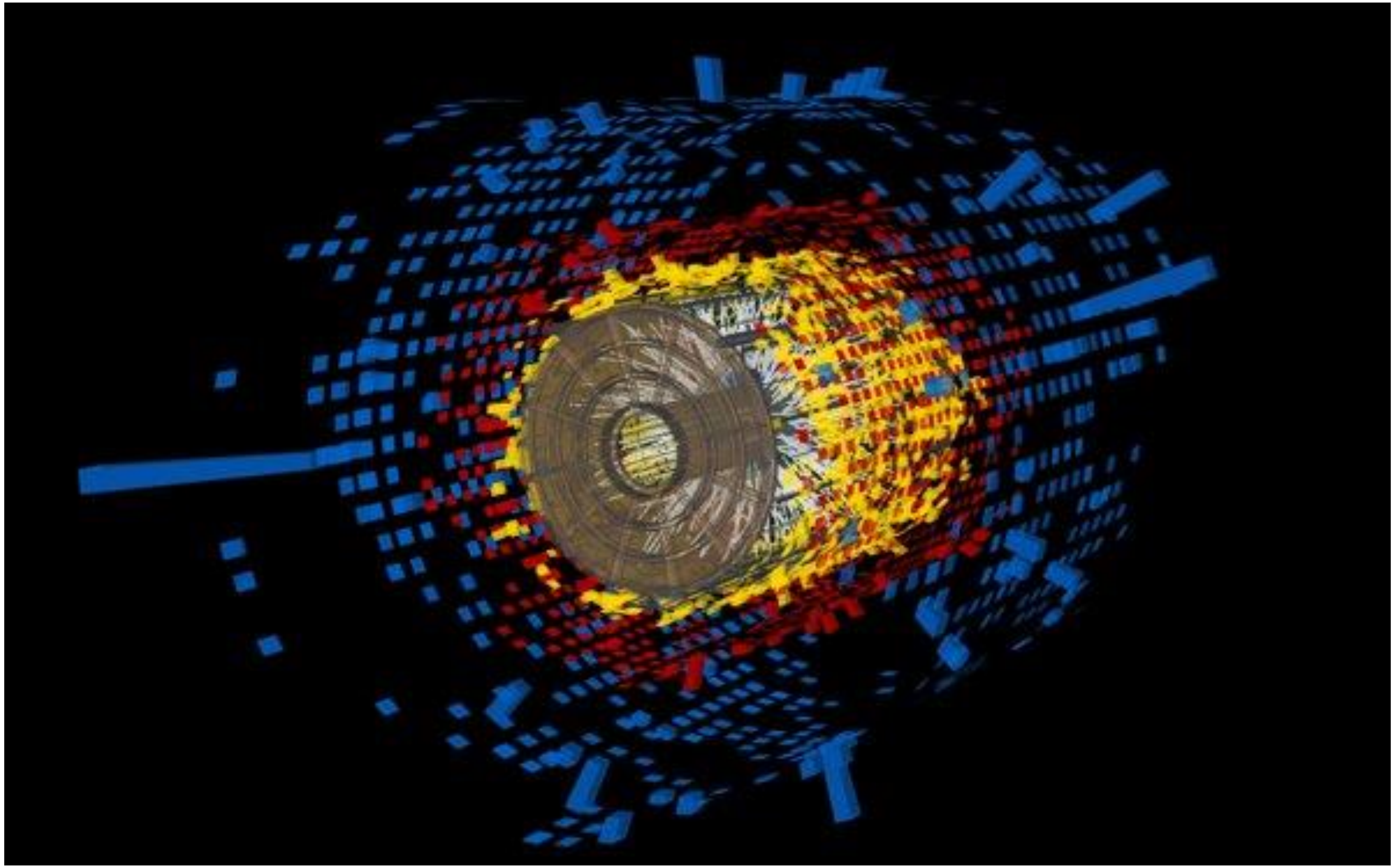




Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]

- TennGen: 40-50 % 2.76 TeV PbPb event display for sPHENIX

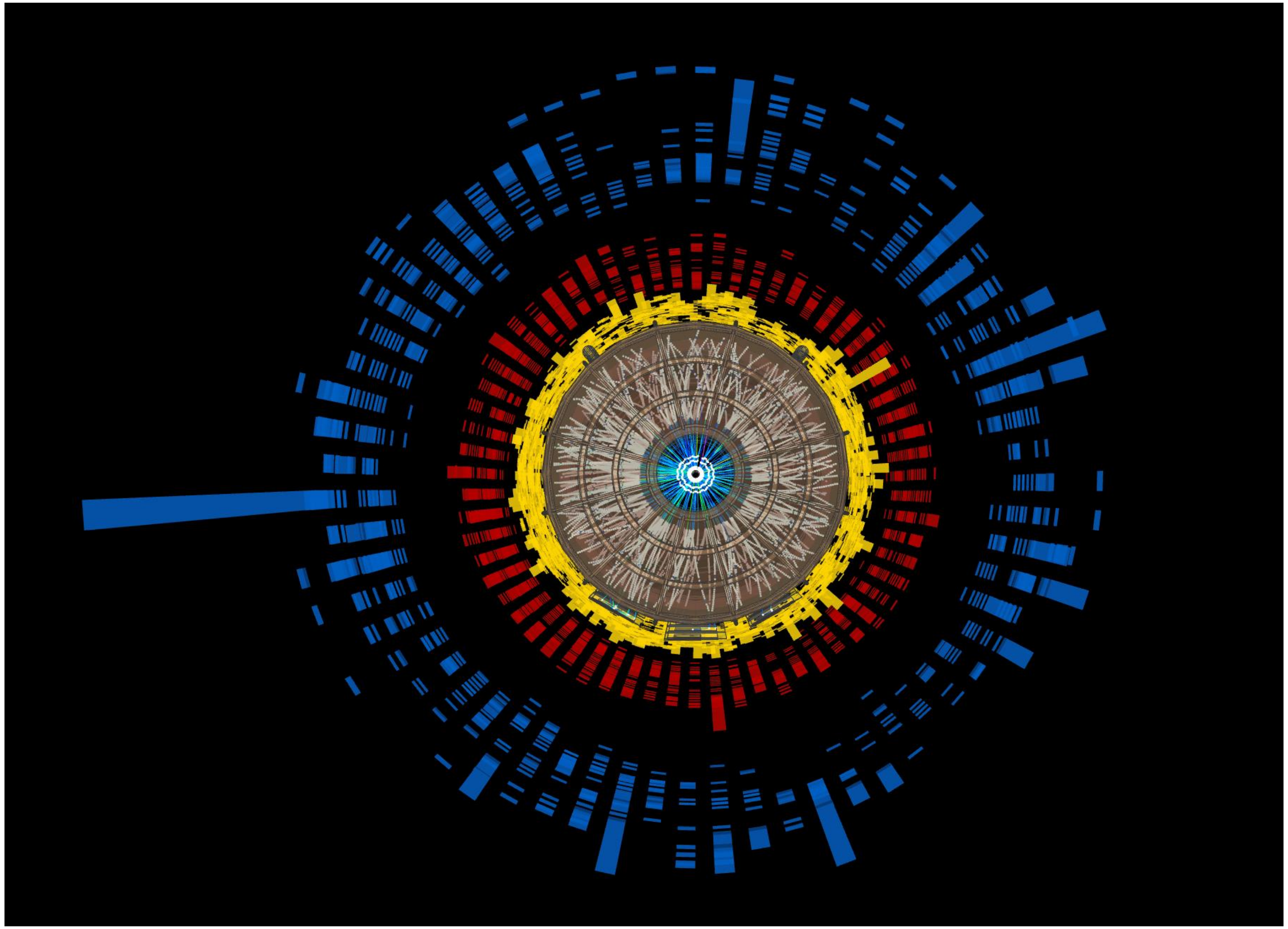


Event displays made by Ejiro Umaka

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

- TennGen: 40-50 % 2.76 TeV PbPb event display for sPHENIX



Event displays made by Ejiro Umaka

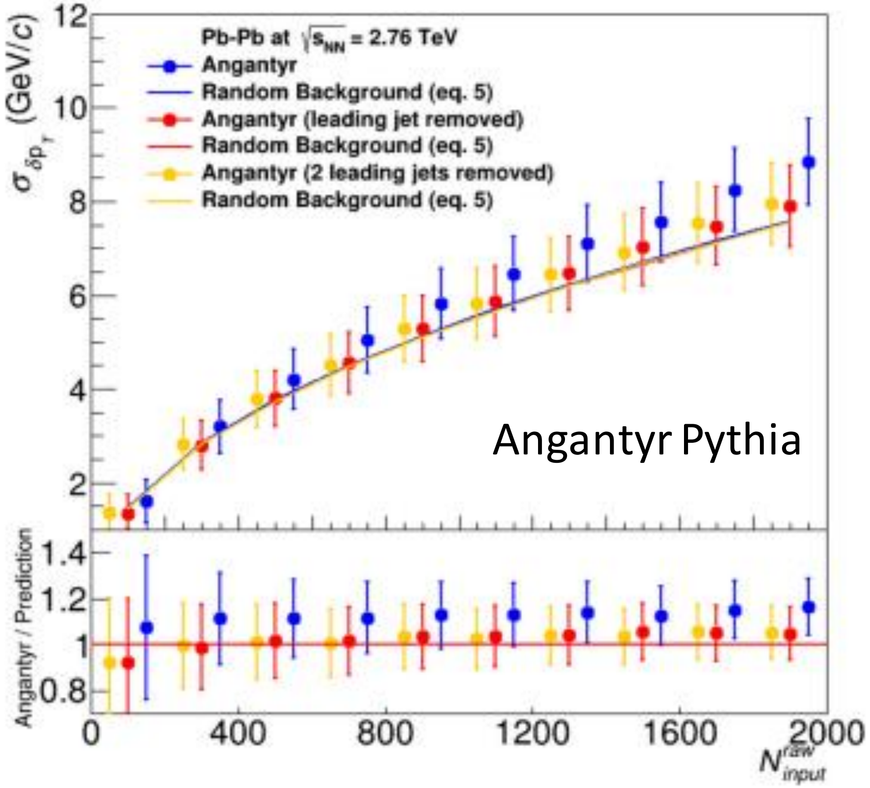
Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

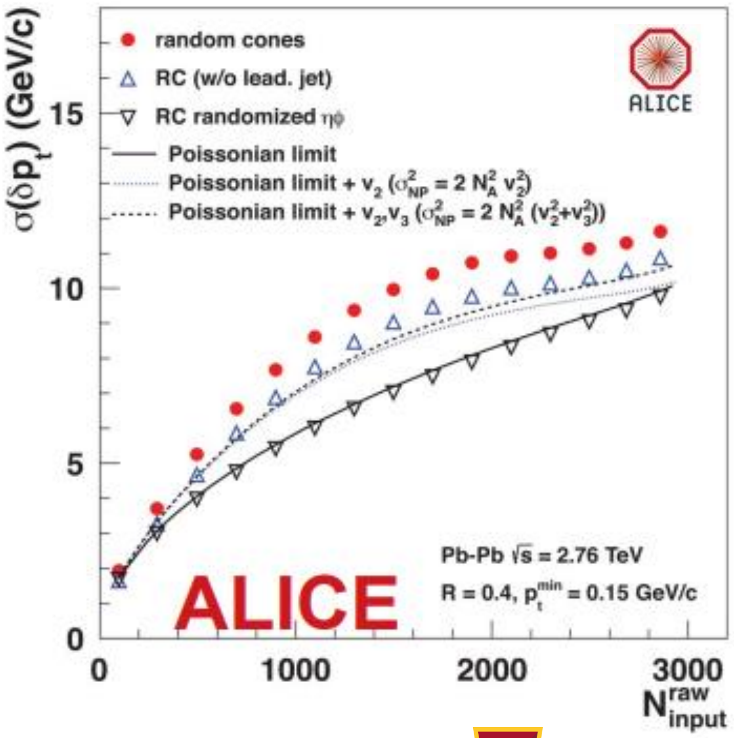
Eq. 5

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915



Abelev et. al. On behalf of ALICE
JHEP 03 (2012) 053



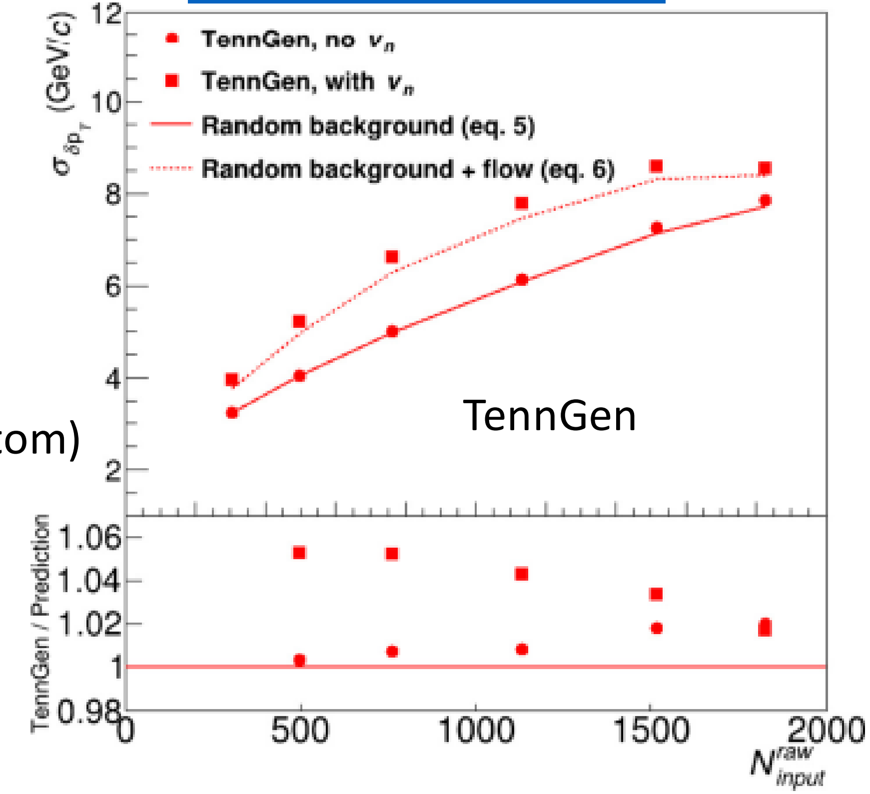
Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

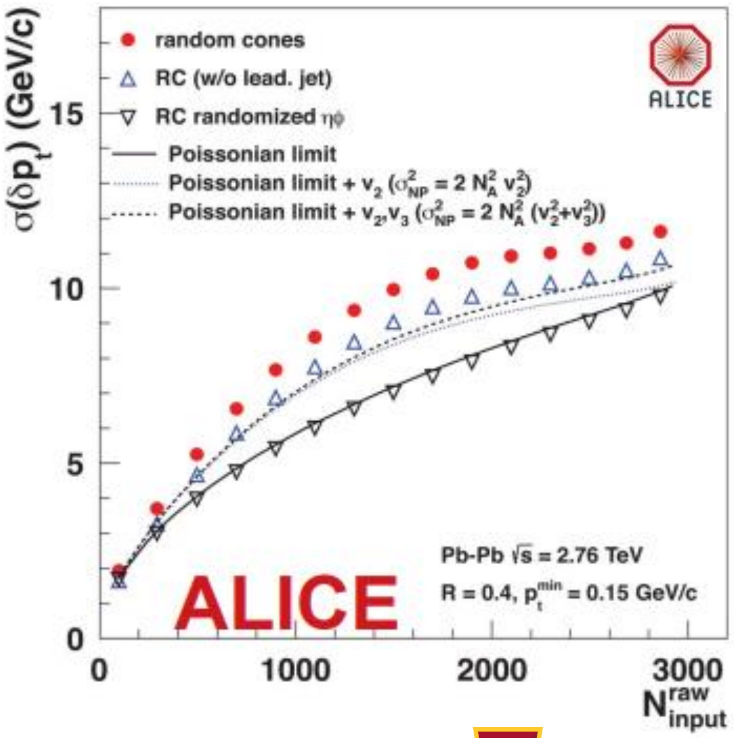
Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

(with flow -top)

(without flow – bottom)



Abelev et. al. On behalf of ALICE
JHEP 03 (2012) 053



Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

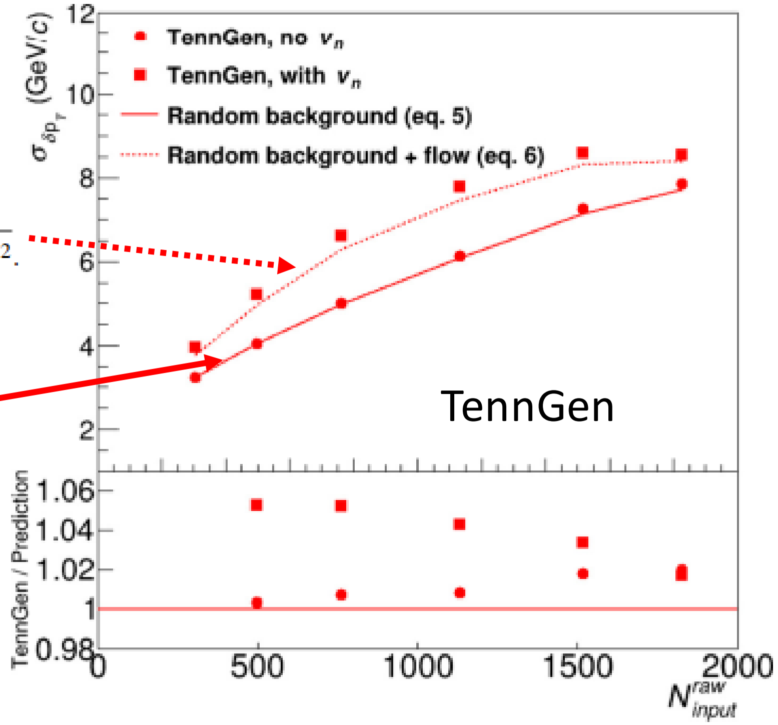
Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]

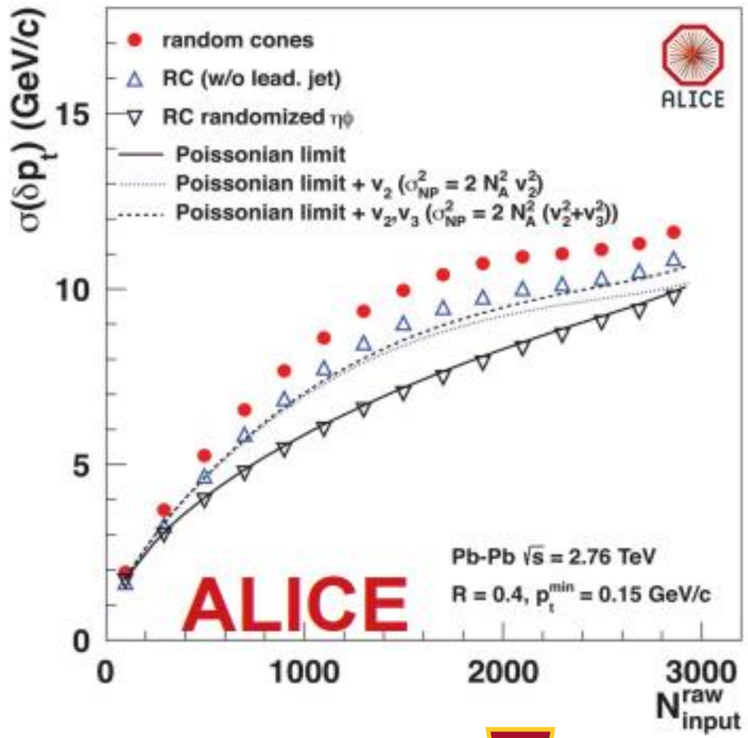
Abelev et. al. On behalf of ALICE
[JHEP 03 (2012) 053]

Eq. 6

$$\sigma(\delta p_{\mathrm{t}}) = \sqrt{N_{\mathrm{A}} \cdot \sigma^2(p_{\mathrm{t}}) + \left(N_{\mathrm{A}} + \sigma^2_{\mathrm{NP}}(N_{\mathrm{A}})\right) \cdot \langle p_{\mathrm{t}} \rangle^2}.$$

Eq. 5 $$\sigma(\delta p_{\mathrm{t}}) = \sqrt{N_{\mathrm{A}} \cdot \sigma^2(p_{\mathrm{t}}) + N_{\mathrm{A}} \cdot \langle p_{\mathrm{t}} \rangle^2}.$$





Background Fluctuations - Model Studies

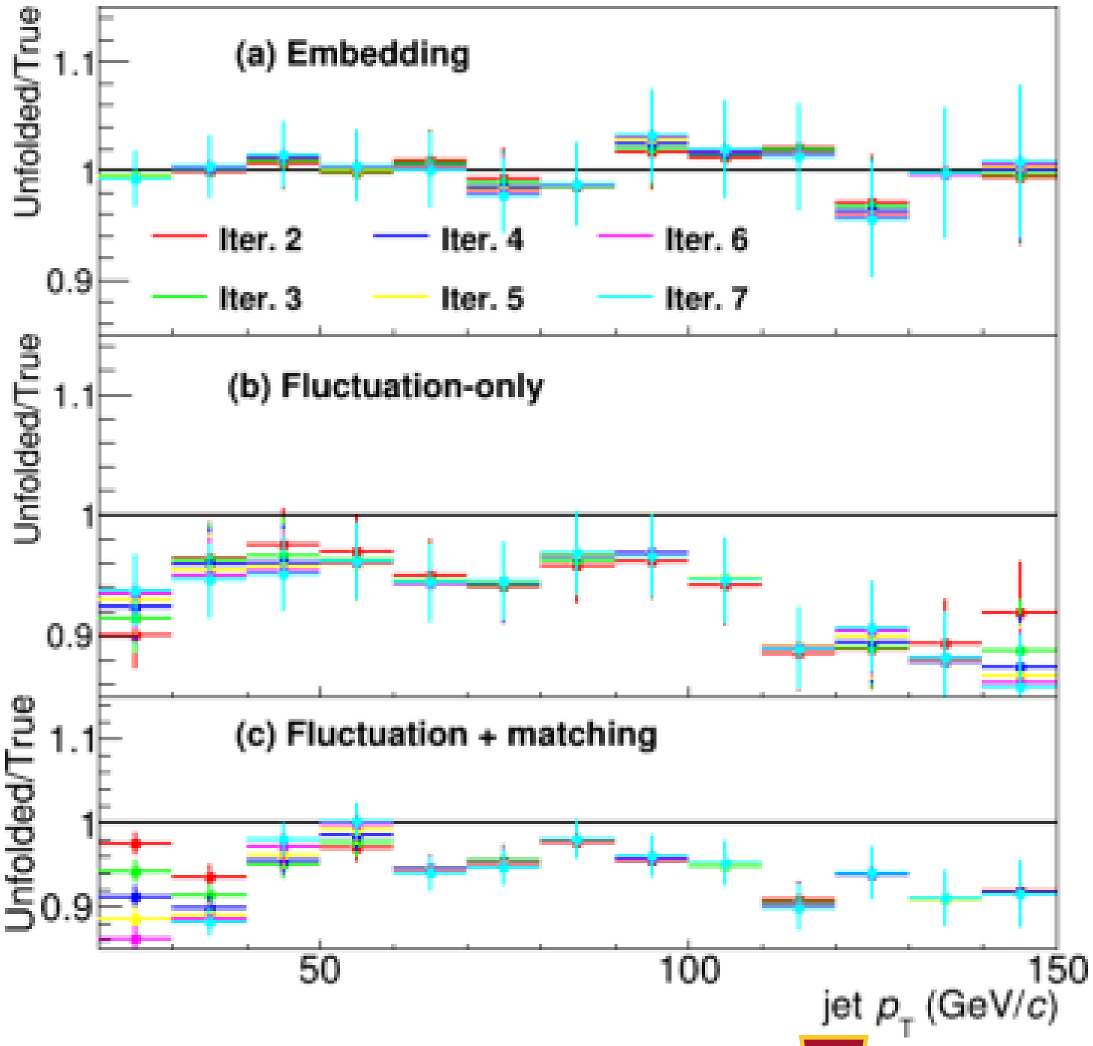
- What can we learn from this background characterization in simple model studies ?
  - Fluctuations widths dominated by Poissonian number fluctuations (expected as seen in data)
  - However, models such as Angantyr can differ by up to 13 % (no flow) and Tenngen up to 6 % (flow) RHIC & AGS Users' Meeting
  - The fluctuations in models do indeed depend on the choice of thermal spectrum etc... – details seem to be ~ 10 % effect
  - **What can be done ?**

Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]

Background Fluctuations - Model Studies - Unfolding

Hughes, da Silva, Nattrass [Phys. Rev. C 106, 044915]

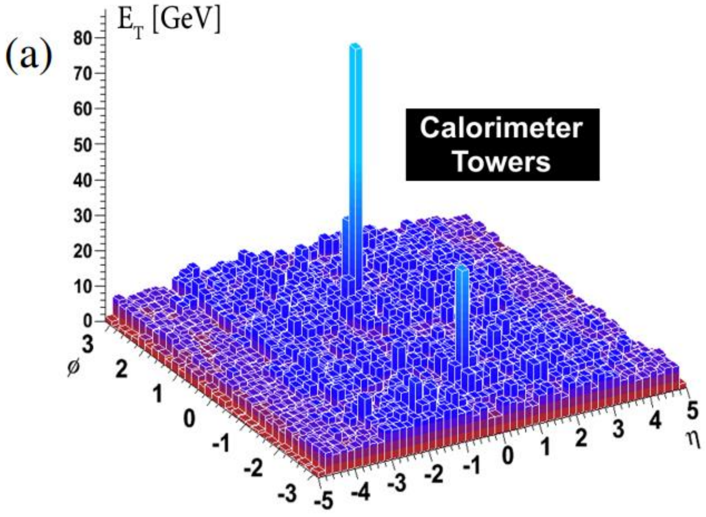
- Fluctuations in models are sensitive to details of model. What can be done ?
- Must unfold in a model just as done with data (because models have background !!!)
- Closure best when using an embedding technique (Pythia pp embedded in Pythia Angantyr PbPb)



Background Fluctuations - Subtraction

- Unfolding in a model to deal with background fluctuations
- What about subtracting the background in data ?
- Many techniques exist but one standard in the area-based subtraction method
    - $P_{T,\,jet}^{corr.} = p_{T,\,jet}^{raw} - \rho * A_{jet}$
- However, fluctuations in pT remain after
    - (std. Dev. ~ 20 GeV for R = 0.4)

Taken from arXiv:1702.07231



Background Fluctuations - Subtraction

- Fluctuations remain after area-based background subtraction technique
- But maybe we can devise a technique to correct for fluctuations ?

Background Fluctuations - Machine Learning

- Fluctuations remain after area-based background subtraction technique

- But maybe we can devise a technique to correct for fluctuations ?

  - Yep – It's Machine Learning

  - Technique from Haake and Loizides:

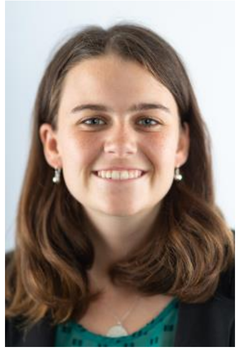
    "Machine-learning-based jet momentum reconstruction in heavy-ion collisions"

    Phys. Rev. C **99**, 064904

  - Improves upon Area Based method (reduces fluctuations)

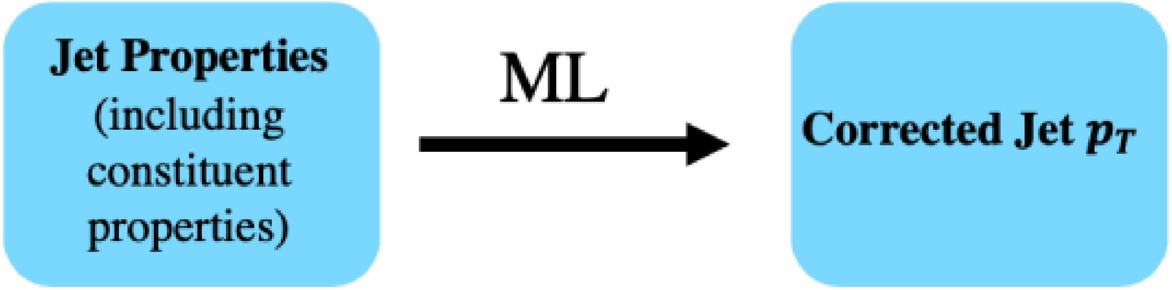
Background Fluctuations - Machine Learning

- Technique from Haake and Loizides: *"Machine-learning-based jet momentum reconstruction in heavy-ion collisions"* Phys. Rev. C **99**, 064904

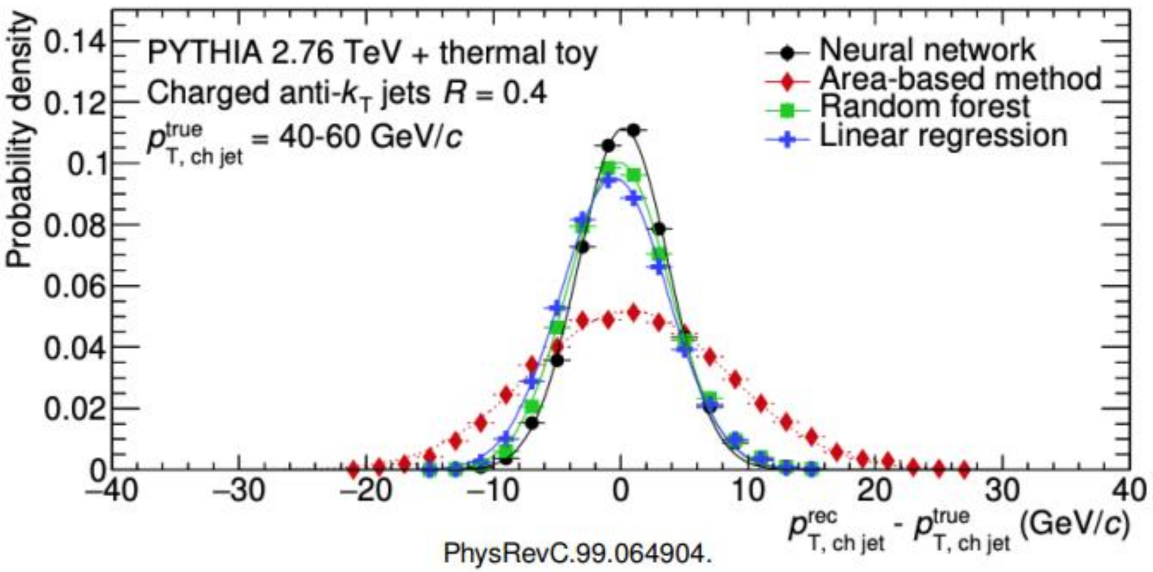


Hannah Bossi

- Further developed by Hannah Bossi and used in by her in ALICE Jet RAA measurements (arXiv:2208.14492)



Taken from H. Bossi



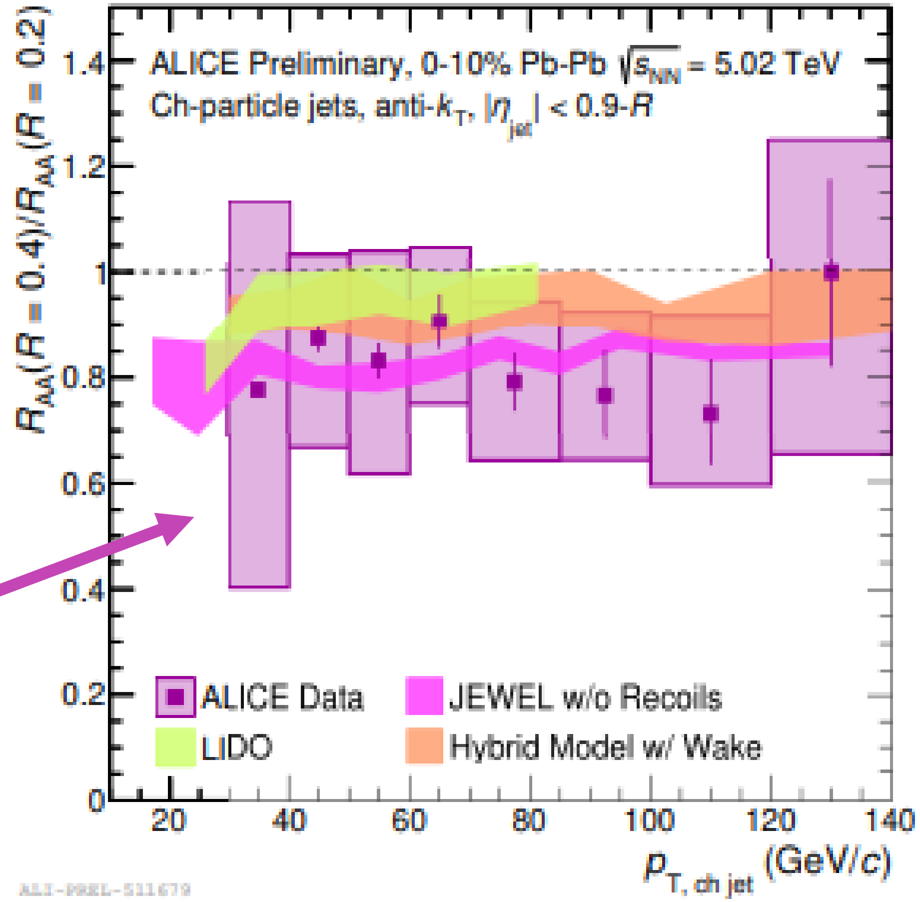
Background Fluctuations - Machine Learning

- Technique from Haake and Loizides: *"Machine-learning-based jet momentum reconstruction in heavy-ion collisions"* Phys. Rev. C **99**, 064904

- What does this technique buy you ?

- "*R-dependence of inclusive jet suppression and groomed jet splittings in heavy-ion collisions with ALICE*" [arXiv:2208.14492v1]
  - Unfolding still necessary
  - BUT
  - Reduced fluctuations
  - Lower momentum (down to p_T^{jet} = 30 GeV)

RHIC & AGS Users' Meeting



Taken from [H. Bossi]

Background Fluctuations - Machine Learning

- Following the analysis in *Phys. Rev. C 99, 064904*

- Some questions to ask:

    - Can we improve on these results ?

    - Can we interpret the ML methods to improve p_T resolution ?

["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]()

5th ML4Jets Workshop – 11/01/22 - 11/04/22

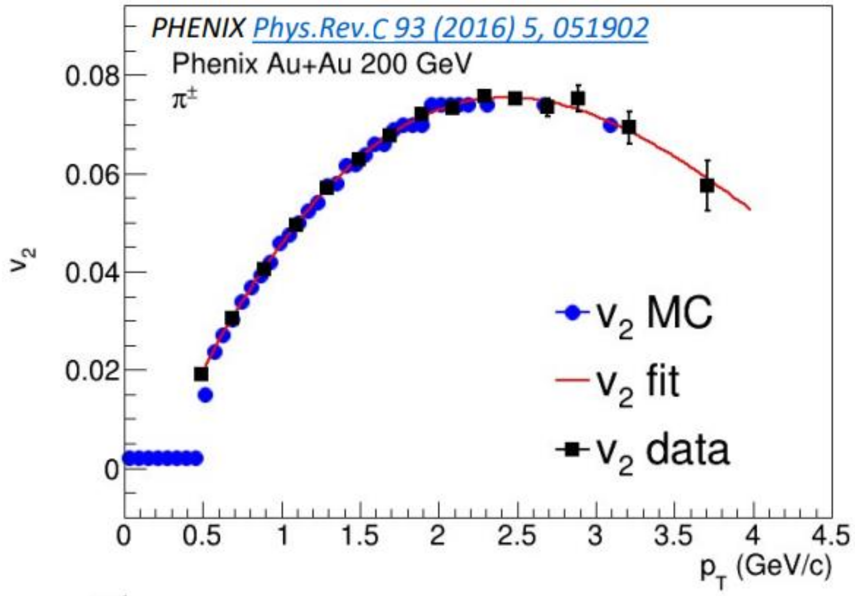
Tanner Mengel

Tanner Mengel

Background Fluctuations - Machine Learning

- Can we improve on results in *Phys. Rev. C 99, 064904* ?
- Add a more complex generator that includes flow (TennGen – fit to PHENIX and STAR data) for background
- Add in PYTHIA 8 for signal
  - Use p_T hard bins (1M events / bin)
- PYTHIA 8 is truth, we want to predict $p_{T,\,jet}^{pythia}$
- TennGen fit to RHIC data is background

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]()



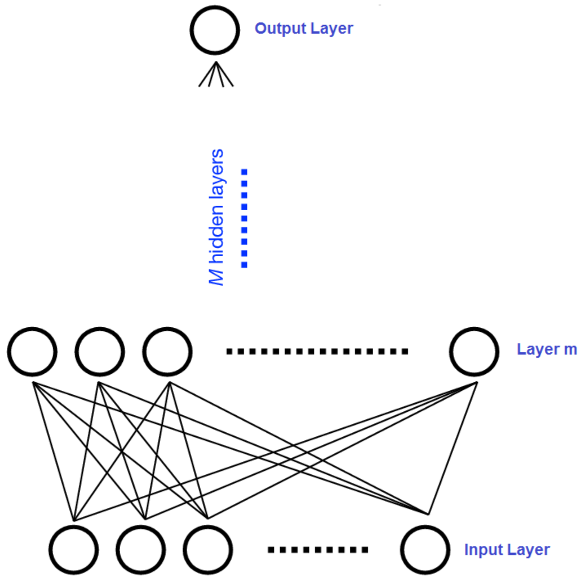
Taken from [T. Mengel]()

Background Fluctuations - Machine Learning

- Can we improve on results in *Phys. Rev. C 99, 064904* ? - 2 methods

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

Deep Neural Net (DNN)
Architecture: [N, 100, 100, 50, 1]
Features: [11]



Taken from ResearchGate

Physics Inspired (Multiplicity)

$P_T^{corr.} = p_T^{uncorr.} - rho(N^{jet}_{constit.} - \langle N^{jet}_{pythia\ constit.} \rangle)$

Inspired by:

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Assuming no flow

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + \left(N_A + \sigma^2_{NP}(N_A)\right) \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3

Background Fluctuations - Machine Learning

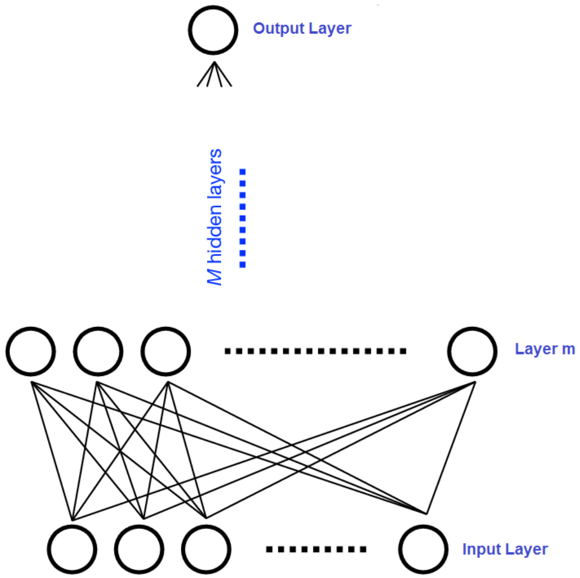
Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

- Can we improve on results in *Phys. Rev. C 99, 064904 ?* - 2 methods

Deep Neural Net (DNN)

Architecture: [N, 100, 100, 50, 1]

Features: [11]



Taken from ResearchGate

Physics Inspired (Multiplicity)

$P_T^{corr.} = p_T^{uncorr.} - rho(N^{jet}_{constit.} - \langle N^{jet}_{pythia} \rangle)$

Inspired by:

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Assuming no flow

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3

Background Fluctuations - Machine Learning

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

- Can we improve on results in *Phys. Rev. C 99, 064904 ?*
  - 2 methods

Deep Neural Net (DNN)
Architecture: [N, 100, 100, 50, 1]
Features: [1]

Physics Inspired (Multiplicity)
$P_T^{corr.} = p_T^{uncorr.} - rho(N^{jet}_{constit.} - \langle N^{jet}_{pythia} \rangle)$

- DNN is powerful but black box

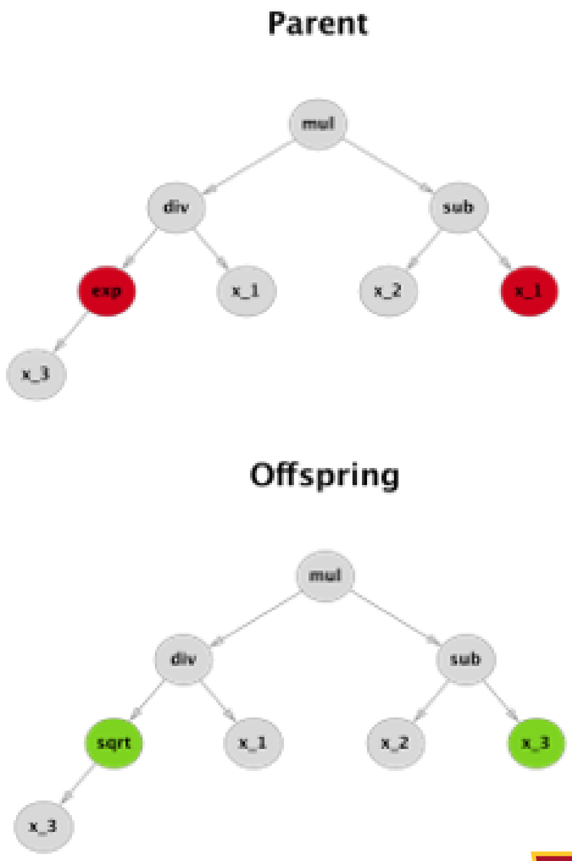
- Want to understand it better ?

- Symbolic regression

Background Fluctuations - Machine Learning

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

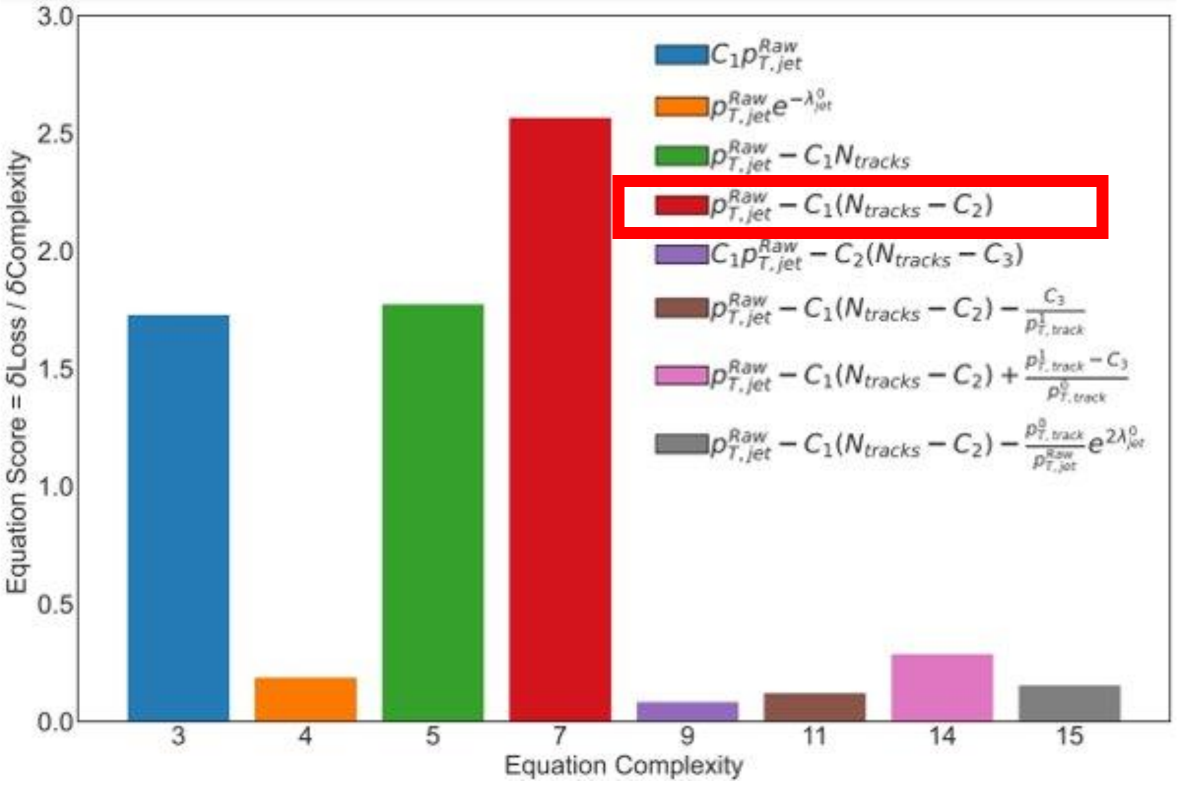
- Symbolic Regression
  - Genetic programming where "traits" = operators
  - Each iteration creates new population with traits from each parent.
  - Highest performing offspring selected
- Plan
  - Train DNN for jet p_T regression
  - Fit input space to DNN prediction using Symbolic Regression implementation in PySR



Background Fluctuations - Machine Learning

- Highest Scoring looks like multiplicity method !!!

Mengel
"Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"

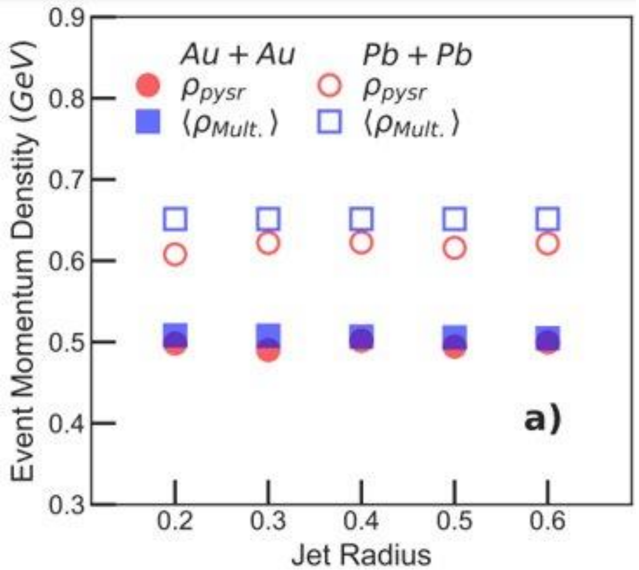


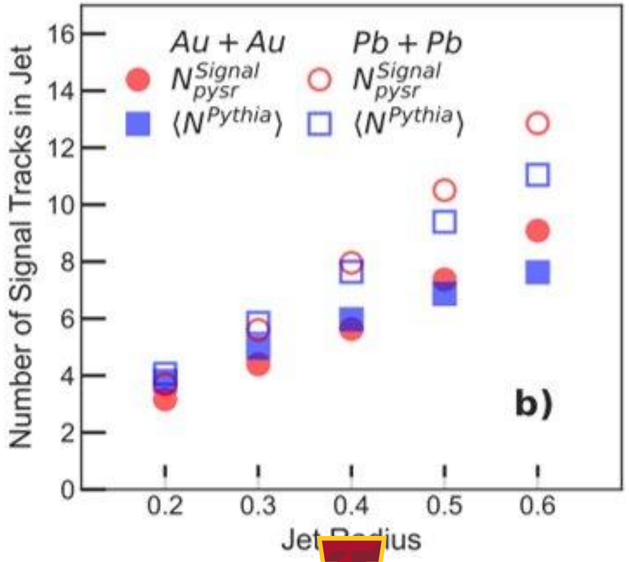
Physics Inspired (Multiplicity)

$P_T^{corr.} = p_T^{uncorr.} - rho(N^{jet}_{constit.} - \langle N^{jet}_{pythia\ constit.} \rangle)$

$C_1 \to \rho$

$C_2 \to \langle N_{pythia} \rangle$

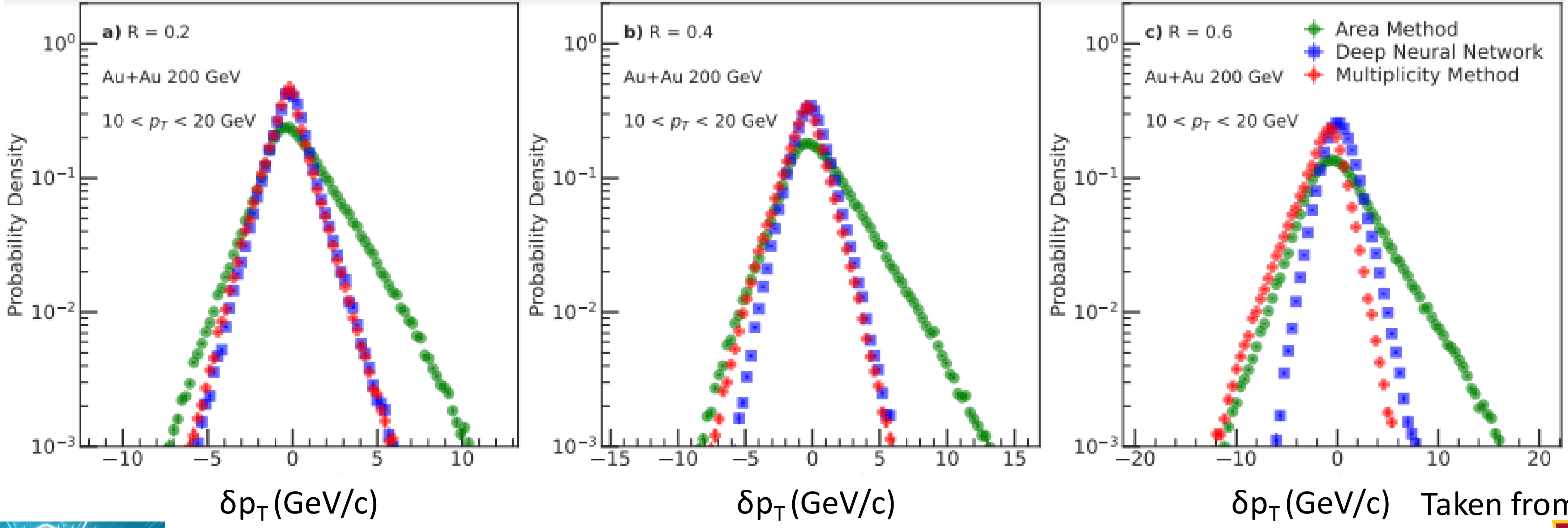




Background Fluctuations - Machine Learning - Results

- Compare performance across methods
- Measure $\delta p_T = p_T^{predicted} - p_T^{pythia}$
- Extract variance (next slide)

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

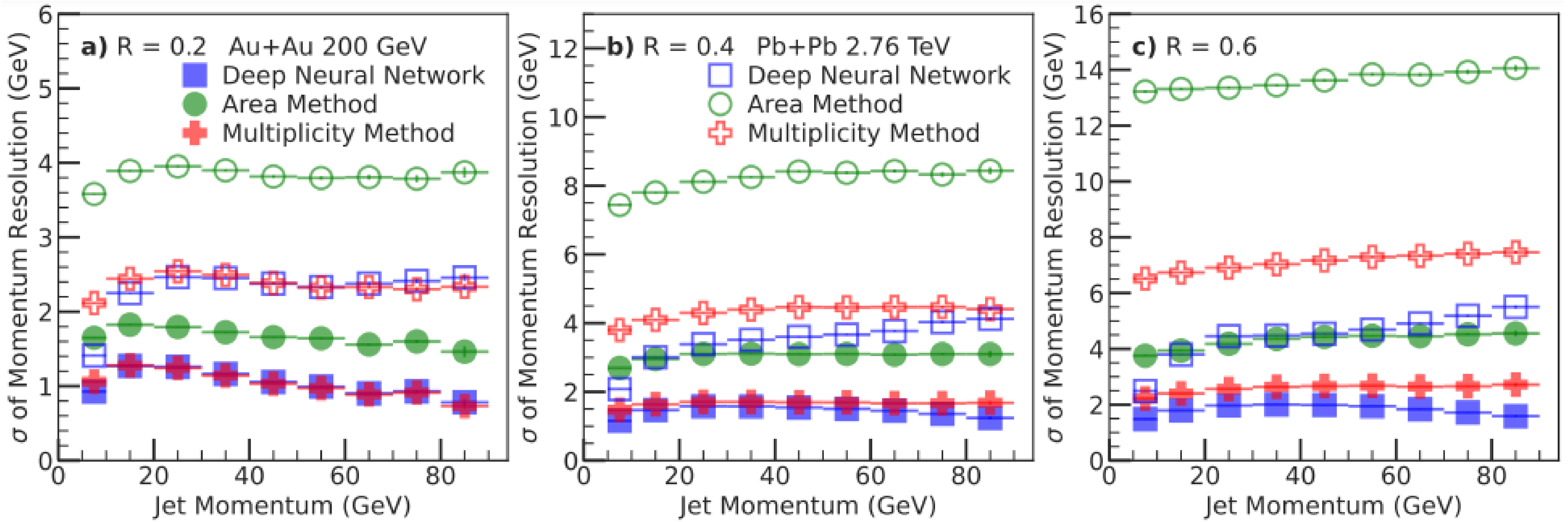


Taken from T. Mengel

Background Fluctuations - Machine Learning

- Extract variance of δpT distribution

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]



Taken from T. Mengel

Background Fluctuations - Machine Learning

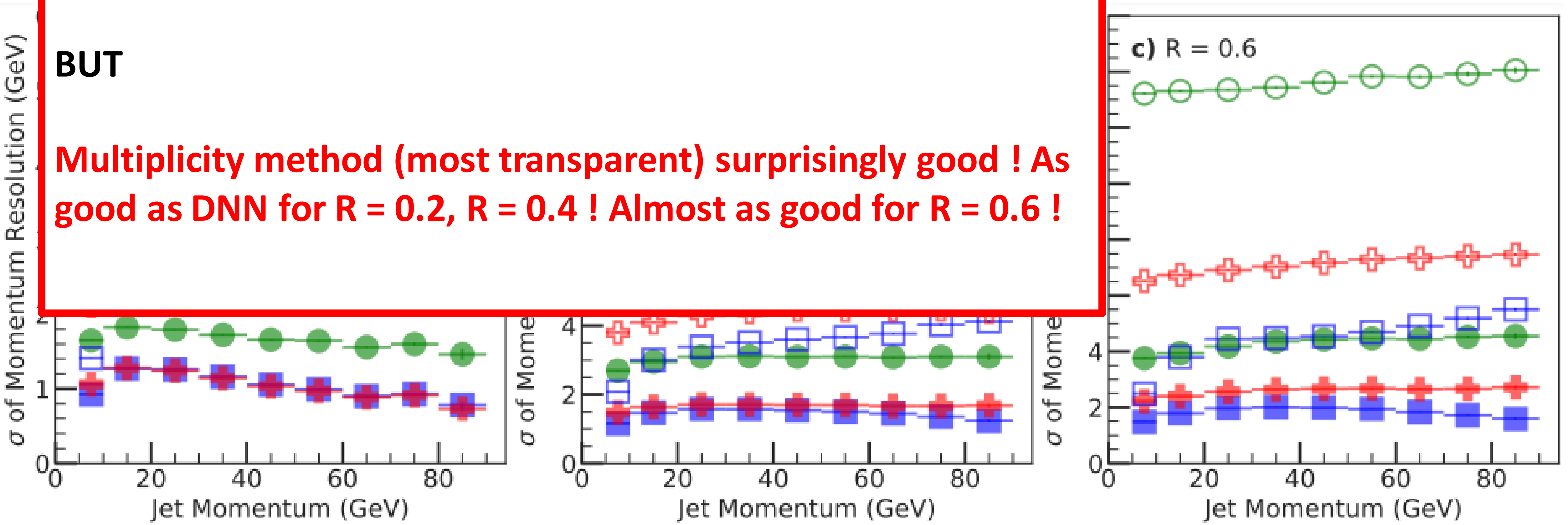
Area based method performs the worst (highest width)

DNN performs best in almost all cases

BUT

Multiplicity method (most transparent) surprisingly good ! As good as DNN for R = 0.2, R = 0.4 ! Almost as good for R = 0.6 !

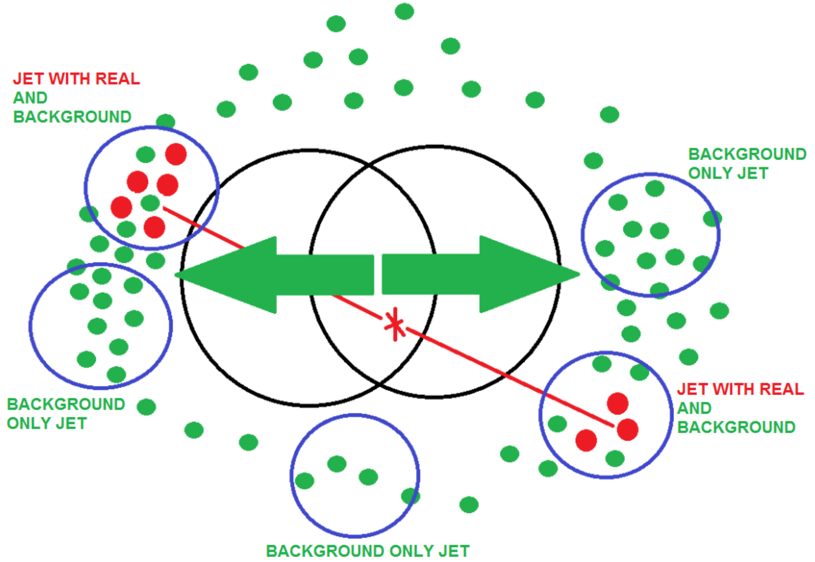
engel
[Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]



Taken from [T. Mengel]

Background Fluctuations - Mitigation

- We see that background subtraction can be heavily informed by machine learning methods where model studies play a crucial role
- We learned that these methods can often point to a simpler/more transparent background subtraction method
- What about the problem of combinatorial jets? Can we use model studies to look at mitigating those?



Background Fluctuations - Mitigation

- What about the problem of combinatorial jets? Can we use model studies to look at mitigating those?

- Look at the effect of cuts on removing combinatorial jets.

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

"Separating signal from combinatorial jets in a high background environment"

arXiv:2301.09148v2 – (also submitted to PRC)

Patrick Steffanic et. al.

Background Fluctuations - Mitigation

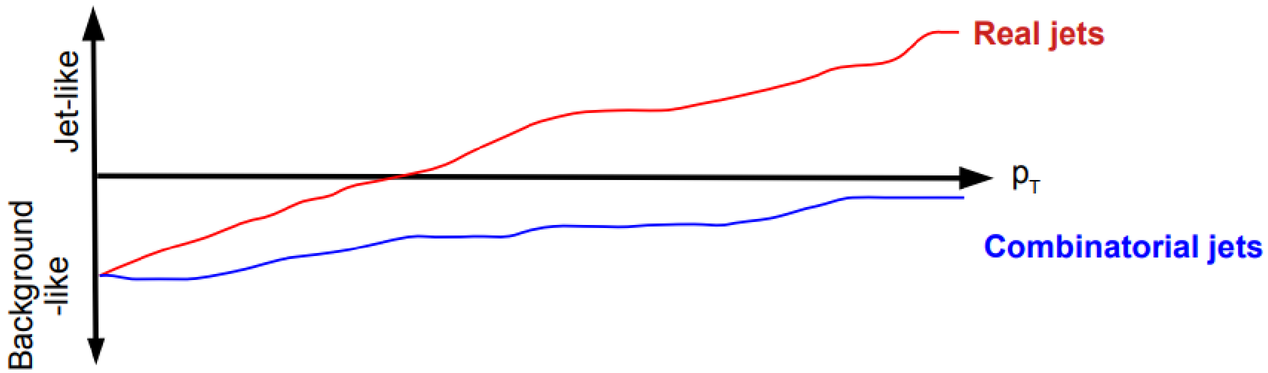
- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Machine Learning (Random Forests)

- Pythia 6 signal

- TennGen background

Steffanic et. al.
arXiv:2301.09148v2



Taken from C. Nattrass

Background Fluctuations - Mitigation

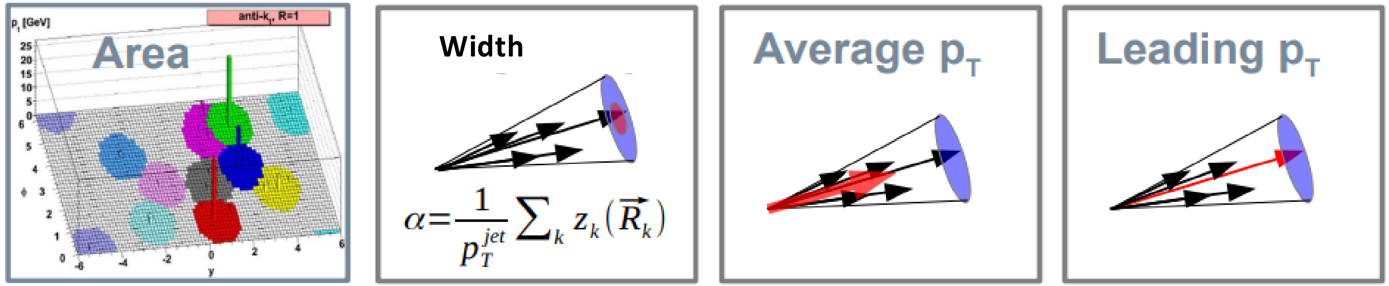
- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

Steffanic et. al.
arXiv:2301.09148v2

- Combinatorial jets: $p_T^{pythia} < 2\pi R^2$ GeV

- Signal jets: $p_T^{pythia} > 0.8 * p_T^{hard\ min.}$ GeV

- Observables:
  - Area: $N_g \langle A_g \rangle$
  - Jet Width: $\Sigma z_i(\Delta R_{i,\ jet})/p_T^{jet}$
  - Leading hadron p_T
  - Mean constituent pT: $\langle p_{T,\ constit.} \rangle$

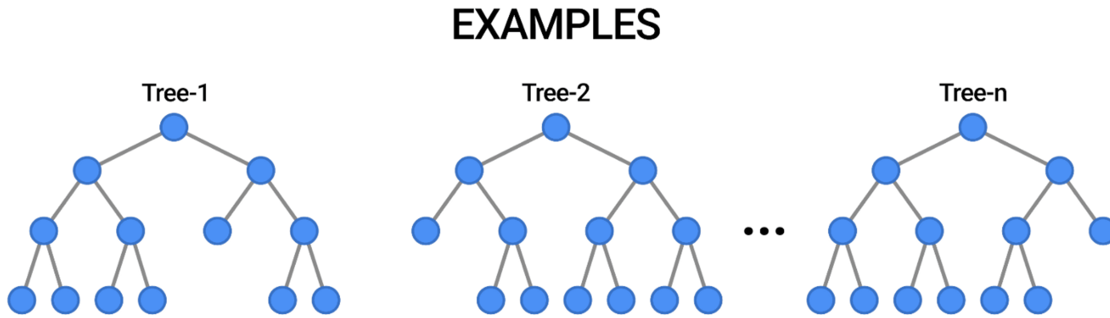


Taken from C. Nattrass

Background Fluctuations - Mitigation

Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?
- Random forest Ensemble Oracle Method
  - Train random forest
  - Apply single decision tree to predictions of random forest
  - Extract top level node of decision tree
    - This is the cut !



Taken from Tensorflow Blog

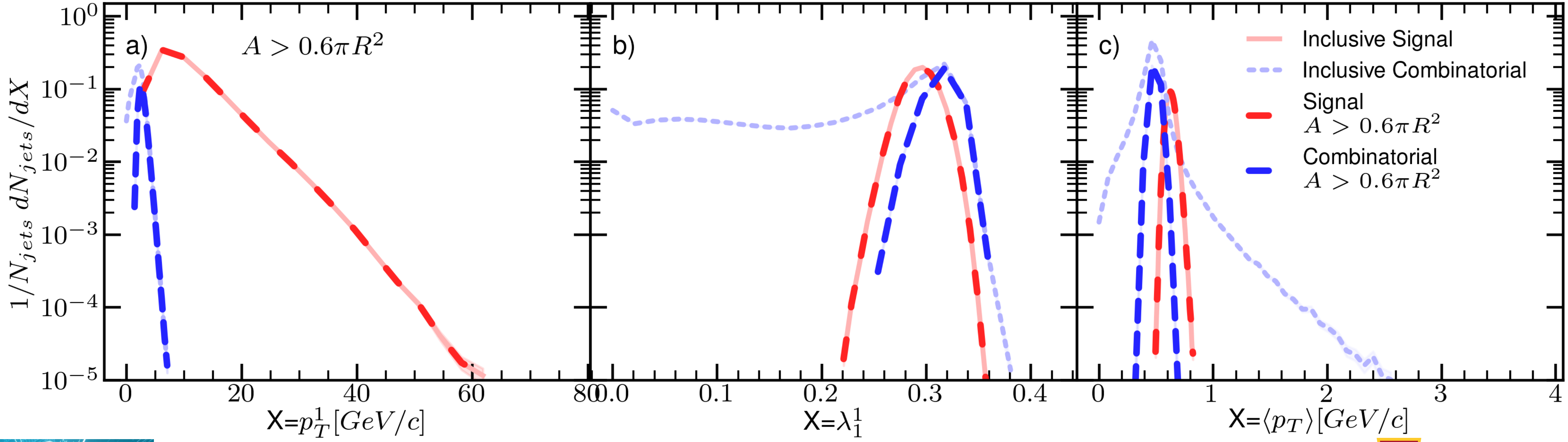
Parameter name	This study	Default
n_estimators	200	100
max_depth	3	None
min_samples_leaf	100	1
min_weight_fraction_leaf	0.1	0.0
max_samples	0.9	1.0
random_state	42	None

Background Fluctuations - Mitigation

Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Baseline "canonical cut" – $A > 0.6\pi R^2$

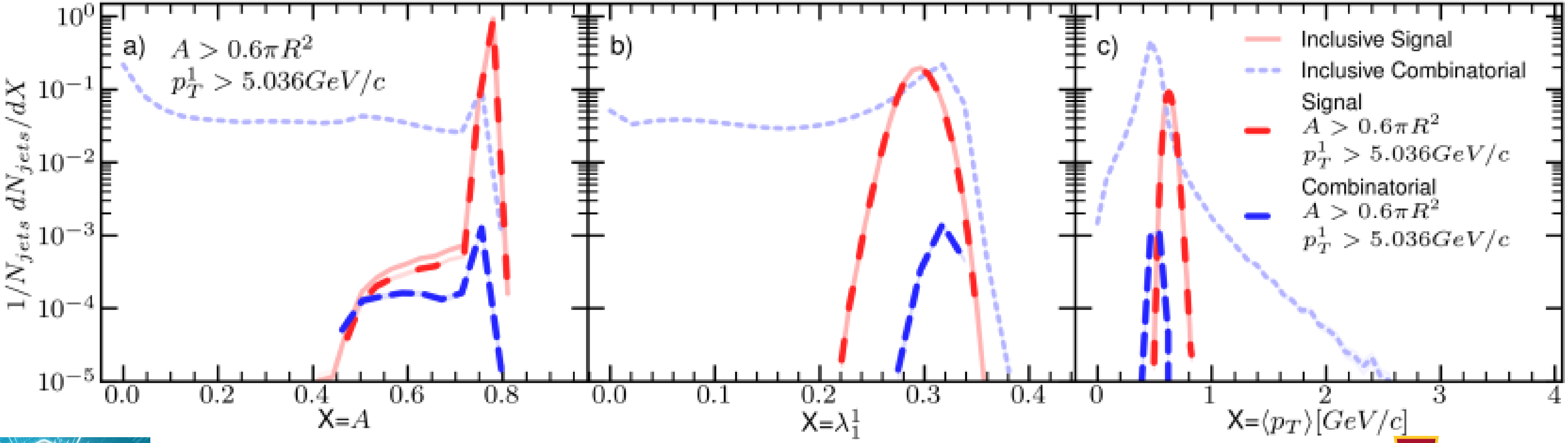


Background Fluctuations - Mitigation

Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- From Random forest – $p_T^{\text{lead. trk.}} > 5.036$ GeV/c (+ $A > 0.6\pi R^2$)



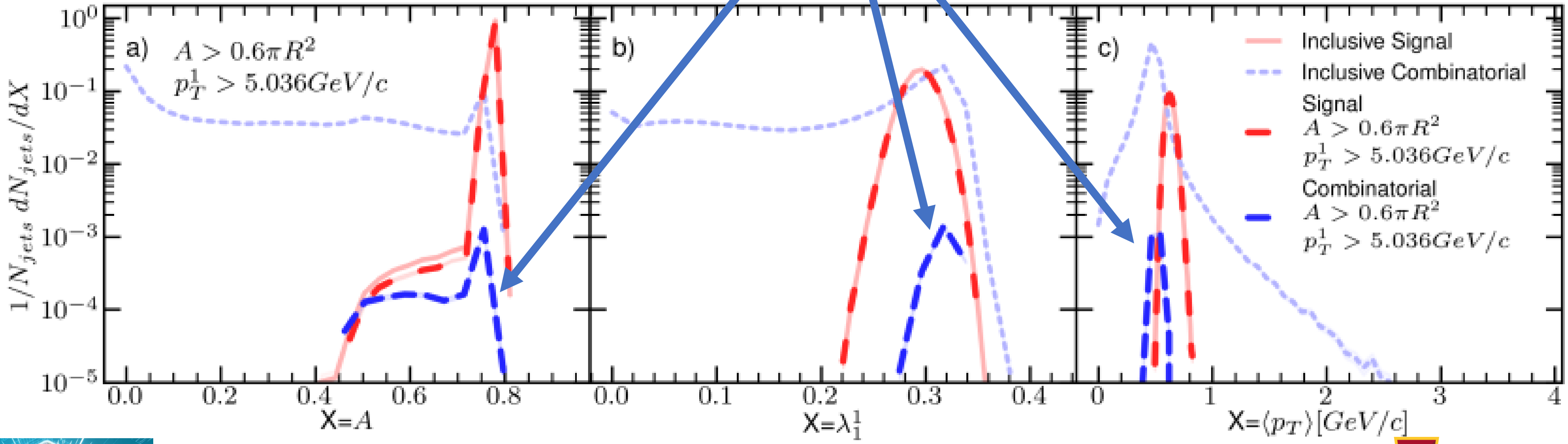
Background Fluctuations - Mitigation

Steffanic et. al. [arXiv:2301.09148v2]

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

These cuts work well but ALWAYS leave a population of combinatorial jets that look like signal jets. The addition of the leading hadron p_T cut removes a lot of combinatorial jets compared to area cut alone.

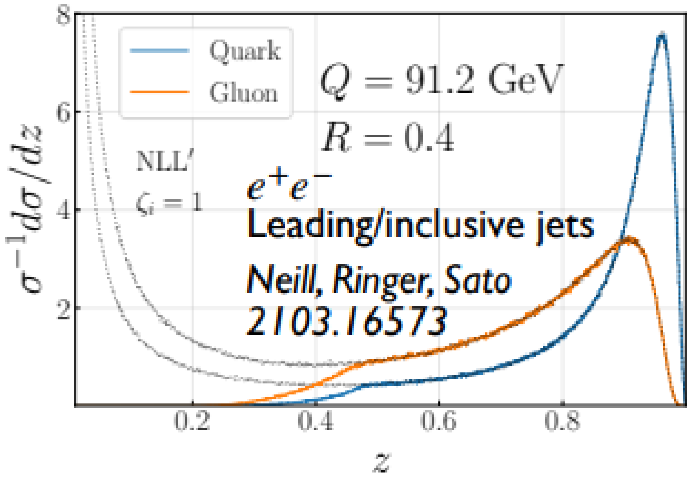
- From Random forest – $p_T^{\text{lead. trk.}} > 5.036$ GeV/c



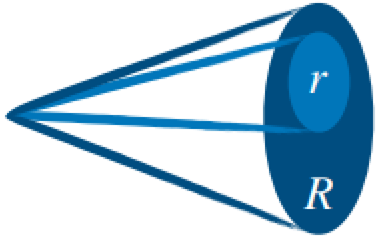
Background Fluctuations - Mitigation

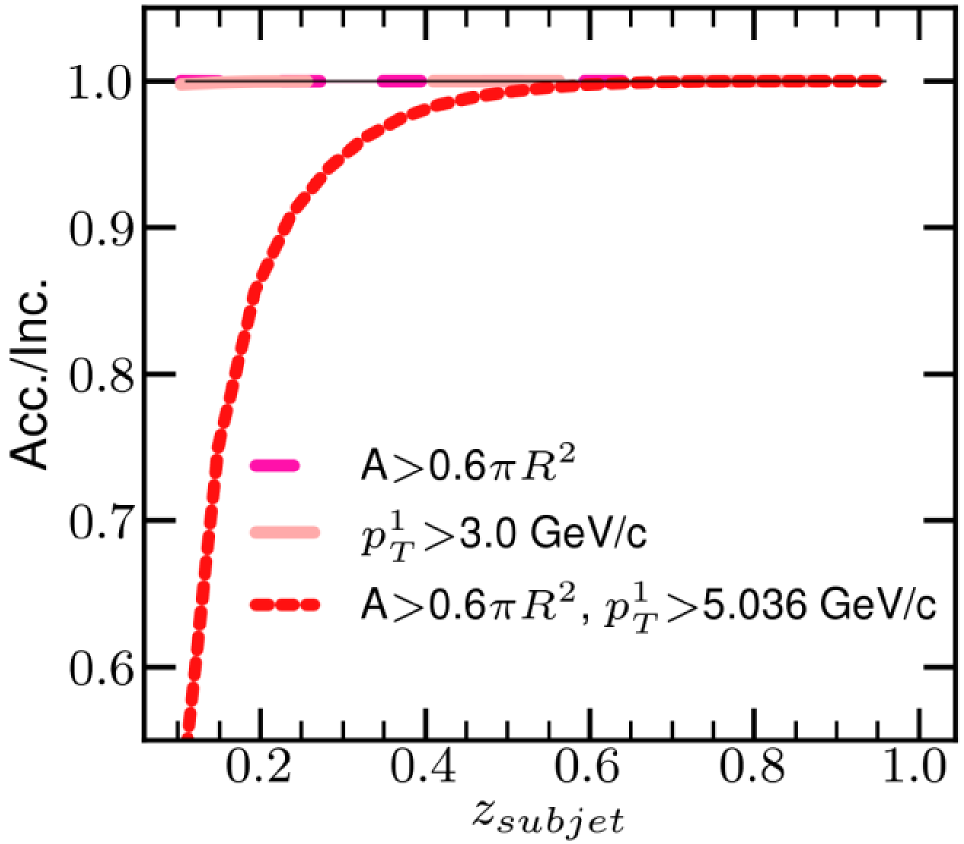
Steffanic et. al. arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?
- From Random forest – $p_T^{\text{lead. trk.}} > 5.036$ GeV/c
- This cut does induce a bias towards quark-like jets



$$z_{subjet} = \frac{p_T^{lead.\ subjet}}{p_T^{jet}}$$





Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*
- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)

Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*
- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)
- Machine learning background subtraction – powerful but obscure

Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*

- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)

- Machine learning background subtraction – powerful but obscure

- Multiplicity method (and similar methods) more transparent and nearly as effective in our studies (and can be informed by ML)

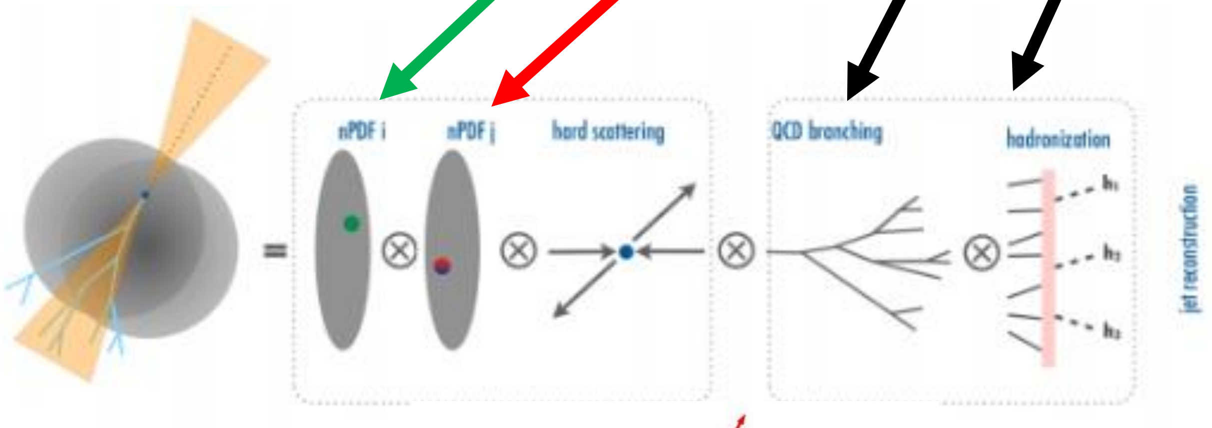
Background Fluctuations - Conclusions/Takeaways

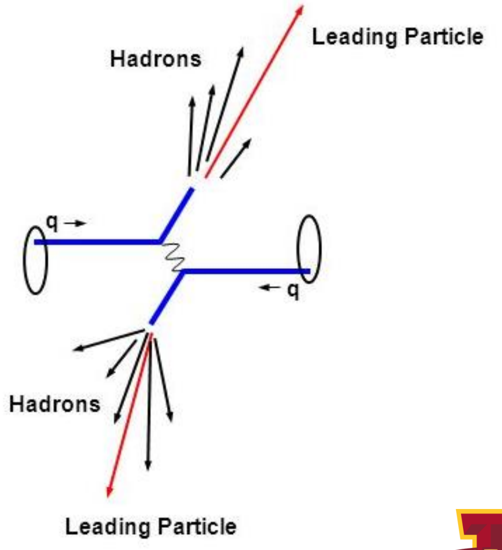
- The details of the background matter – *even in a model*
- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)
- Machine learning background subtraction – powerful but obscure
- Multiplicity method (and similar methods) more transparent and nearly as effective in our studies (and can be informed by ML)
- Combinatorial jets are a tough problem – no silver bullet
  - Cuts always leave some combinatorial jets that look like signal jets – may bias jets

What Is a Jet ?

- In a pp/AA collisions, 2 incoming partons scatter with large momentum transfer
- Scattered partons fragment and hadronize
- This forms back-to-back, collimated sprays of particles – jets

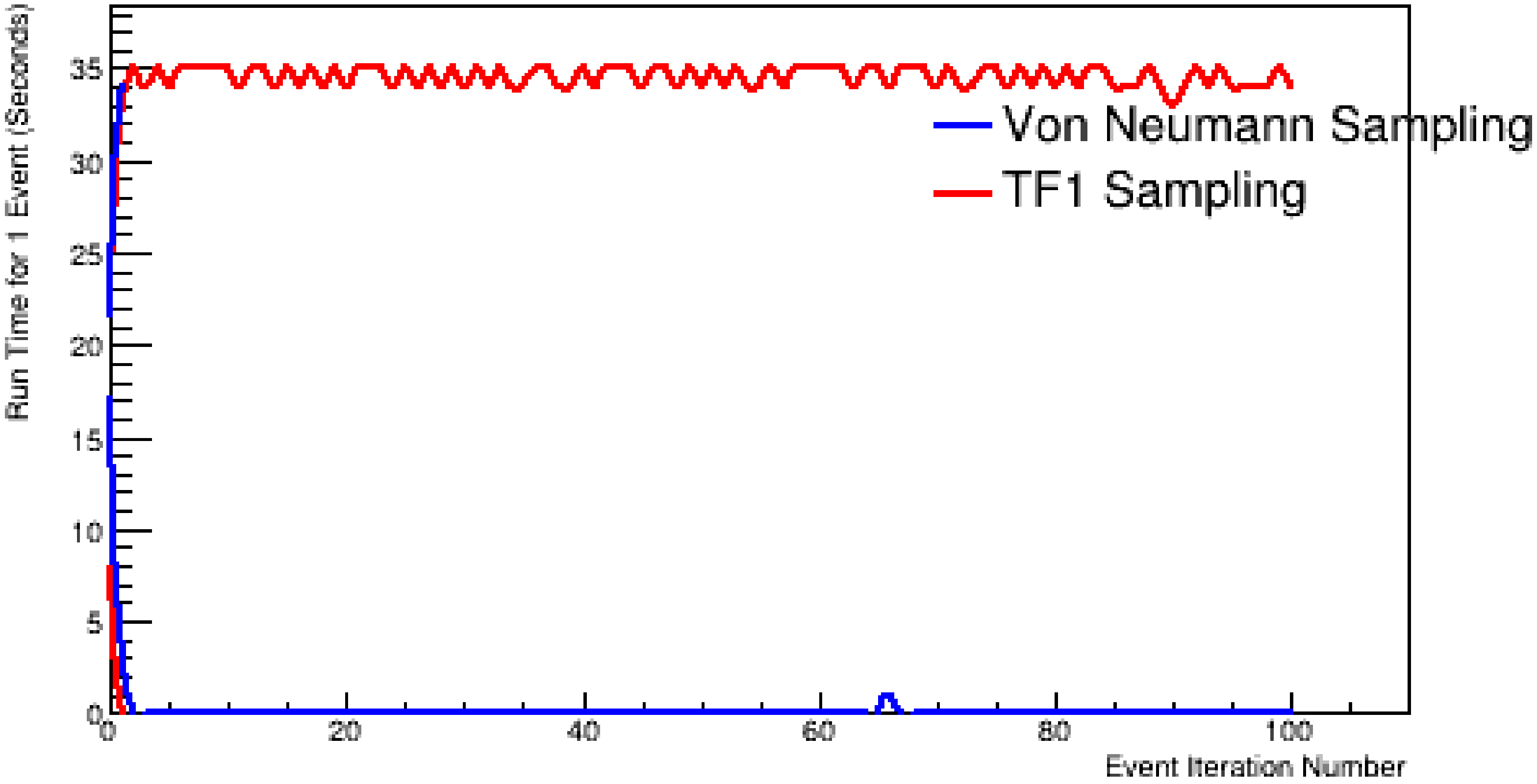
$$\frac{d^3\sigma^h}{dy d^2 p_T} = \frac{1}{\pi}\int dx_a \int dx_b f_a^A(x_a) f_b^B(x_b) \frac{d\sigma_{ab\to cX}}{d\hat{t}} \frac{D_c^h(z)}{z}$$





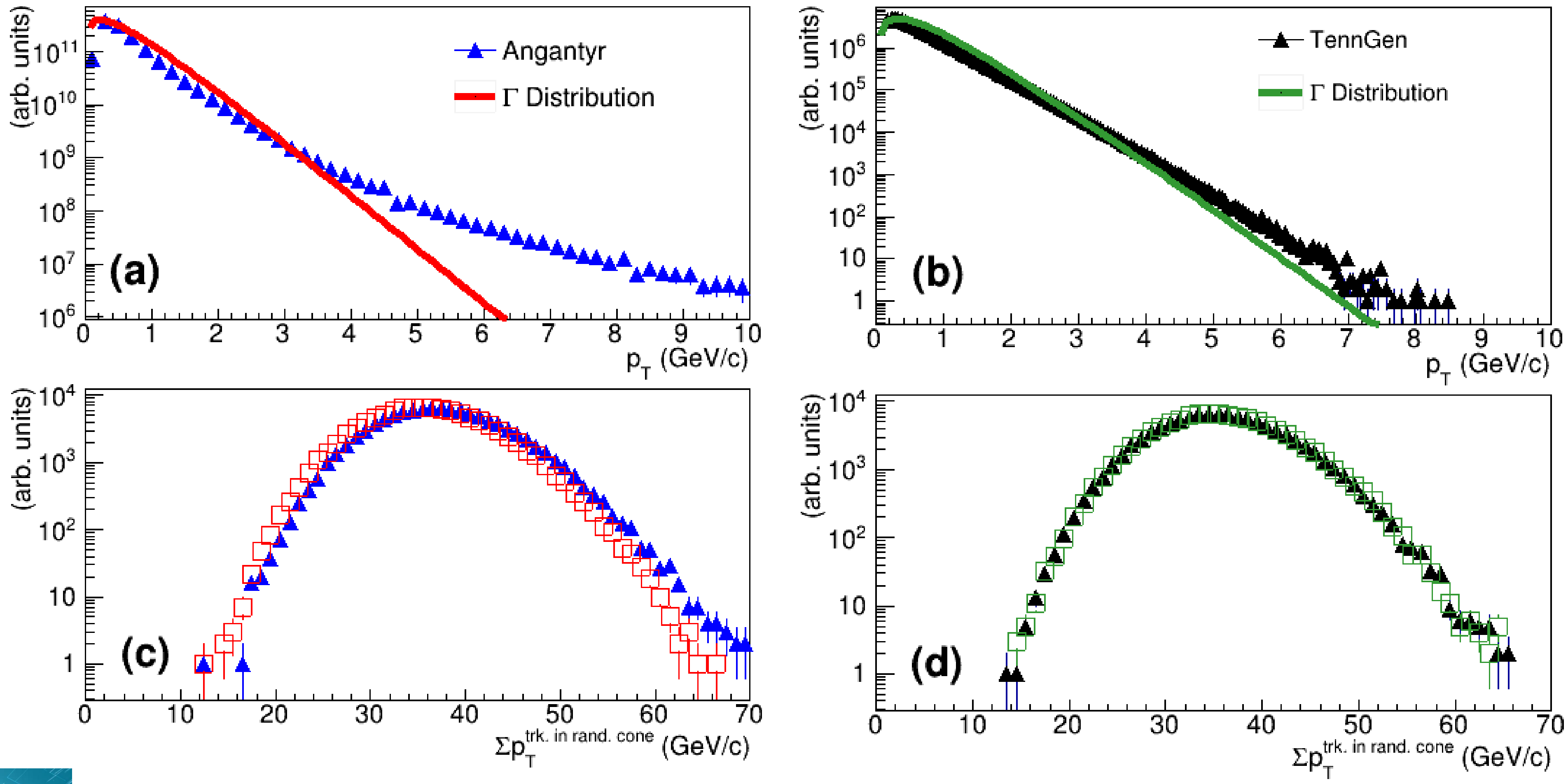
Backup

- TennGen v1 (PbPb only)
- TennGen v2 (AuAu)



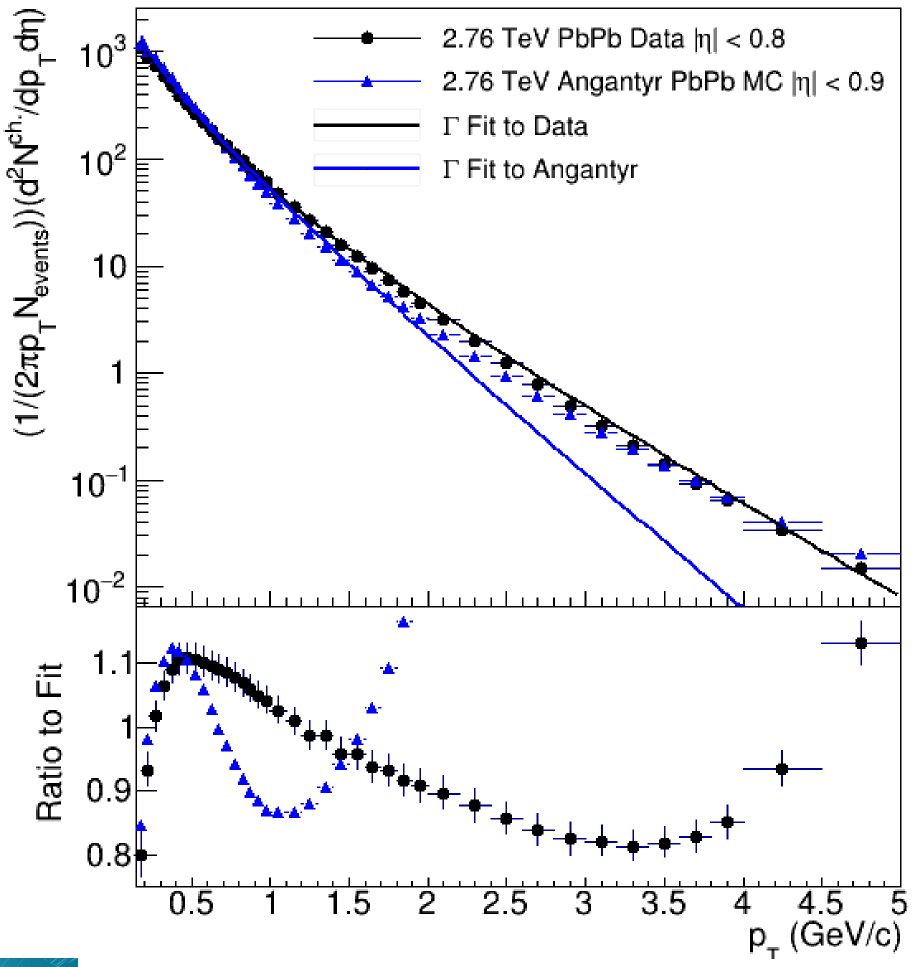
Backup

- Spectra shapes compared to Gamma Distribution



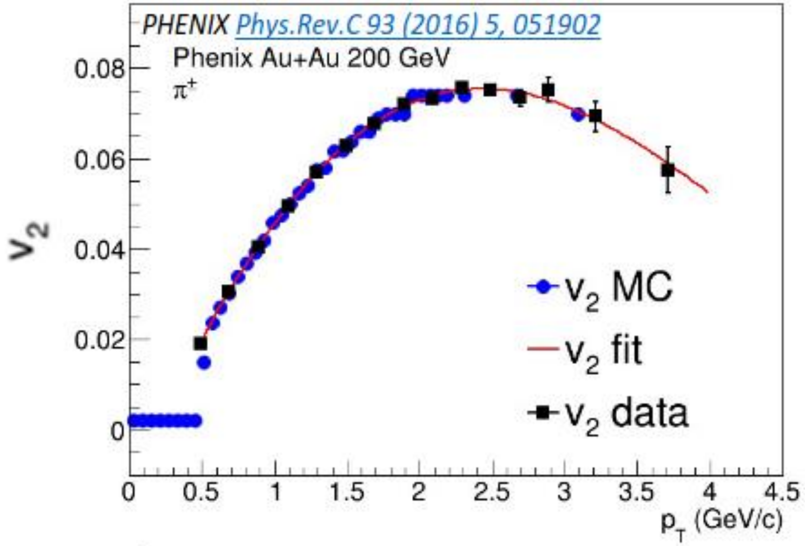
Backup

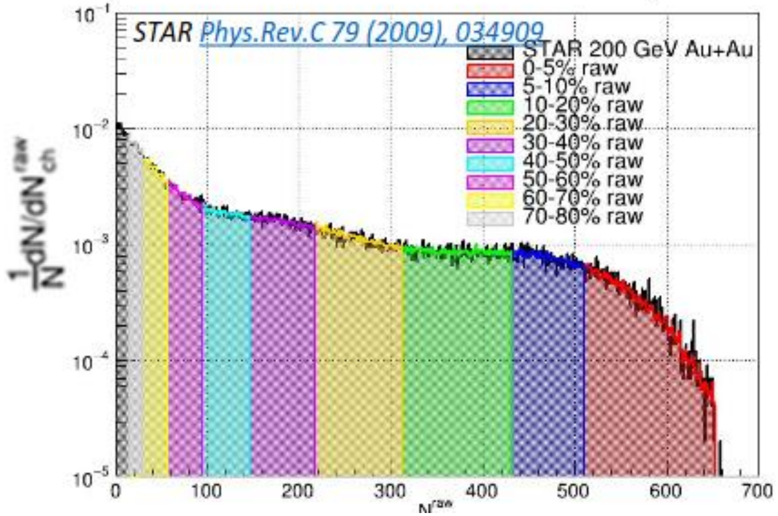
- Spectra shape compared to Gamma Distribution (Angantyr Only)



Backup

- PYTHIA8 (Signal):
  - 25 Million (1 million per p_T hard bin) p+p events at 200 GeV, Tune 14
- TennGen (Background): *C. Hughes et al* Phys. Rev. C 106 (2022), 044915
  - **Multiplicity**: Sampled from corrected N_{ch} distribution *STAR* Phys.Rev.C 79 (2009), 034909
  - $\boldsymbol{p_T}$: Identified particle p_T spectrum fit with Boltzmann-Gibbs Blast wave *PHENIX* Phys.Rev.C 88 (2013) 2, 024906
  - $\boldsymbol{\phi}$: Identified particle flow harmonics (v_2, v_3, v_4) *PHENIX* Phys.Rev.C 93 (2016) 5, 051902
  - $\boldsymbol{\eta}$: Uniform distribution |η| < 1.1
- Merge PYTHIA8 charged particles with TennGen Au+Au 200 GeV background
- Find anti-k_T jets
  - Only save jets with $p_T^{PYTHIA} > 5.0$ GeV
  - ~30 Million jets per dataset
- Take p_T^{PYTHIA} to be truth value
  - Train-Test split: 20/80%

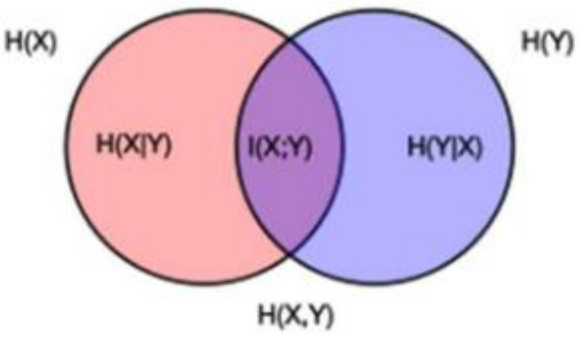


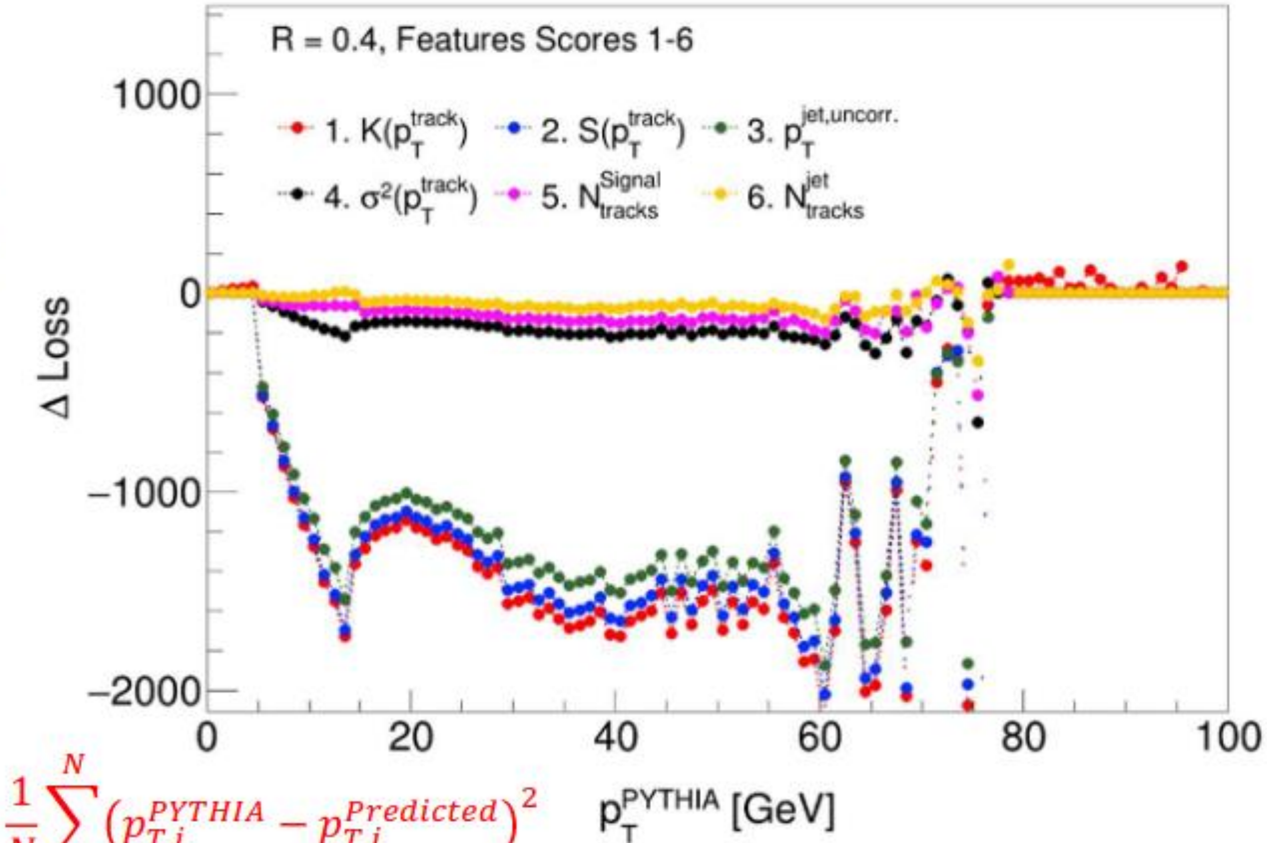


Backup

Feature space optimization

- Mutual Information: I(X;Y) = H(X,Y) – H(X|Y) –H(Y|X)
- Permutation Scoring: Randomly permutes feature to see change in Cost (mean squared error) evaluation





$$MSE = \frac{1}{N}\sum_{i}^{N}\left(p_{T,i}^{PYTHIA} - p_{T,i}^{Predicted}\right)^2$$

Backup

Genetic programming where ‘traits’ = operators.

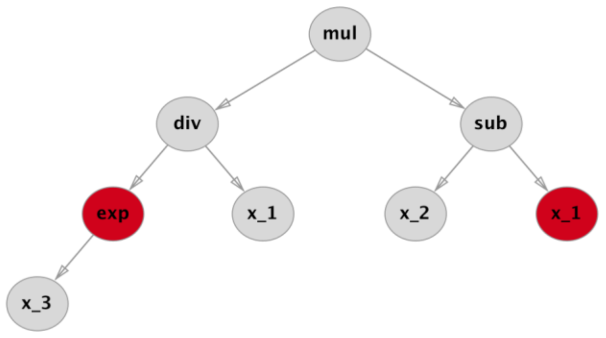
Each iteration creates new population with traits from each parent. Highest performing offspring are selected.

Proof of concept (Area correction):

Target = area_based_prediction

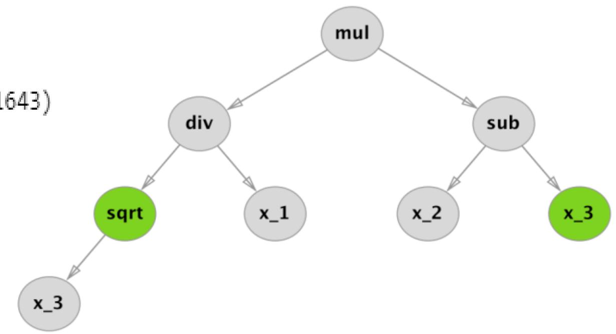
Inputs = jet_pt_raw , jet_area, median_pt_over_area

Parent



Complexity	Loss	Score	Equation
1	1.708e+02	4.292e-01	jet_pt_raw
3	5.111e+01	6.032e-01	(jet_pt_raw * 0.7568252)
4	8.534e+00	1.790e+00	(jet_pt_raw - (jet_area * median_pt_over_area))
5	8.412e+00	1.440e-02	(jet_pt_raw + (median_pt_over_area * -0.4965229))
9	8.237e+00	5.256e-03	(((jet_area - median_pt_over_area) * 0.47800952) - (jet_pt_raw * -0.9821865))
15	8.115e+00	2.473e-03	((((jet_pt_raw - ((median_pt_over_area * 0.4997403) + 1.1289326)) - -1.282973) * 0.96711856) - -0.95751643)

Offspring



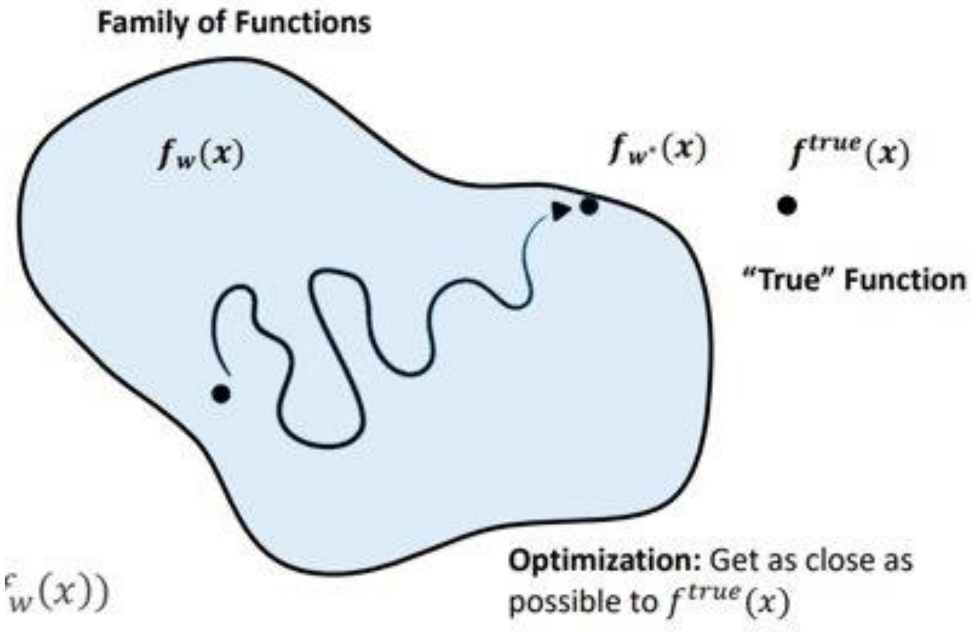

```
PySRRegressor.equations_ = [
        pick      score                                        equation  \
    0         0.000000                                       jet_pt_raw
    1         0.603197                         (jet_pt_raw * 0.7568252)
    2  >>>>   1.789895    (jet_pt_raw - (jet_area * median_pt_over_area))
```

Backup

- *SR models are good if defined arithmetic expression exists that maps inputs to outputs f(x) = y*

- *DNN models are good at factorizing and approximating multivariate mappings: f(x|theta) = yhat ~ y*

- *Plan*
  1) *Train DNN on jet pT regression*
  2) *Fit input space to DNN prediction using PySR*

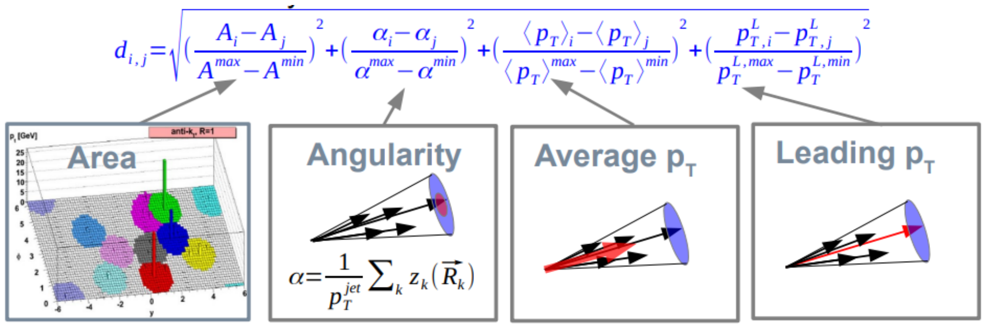


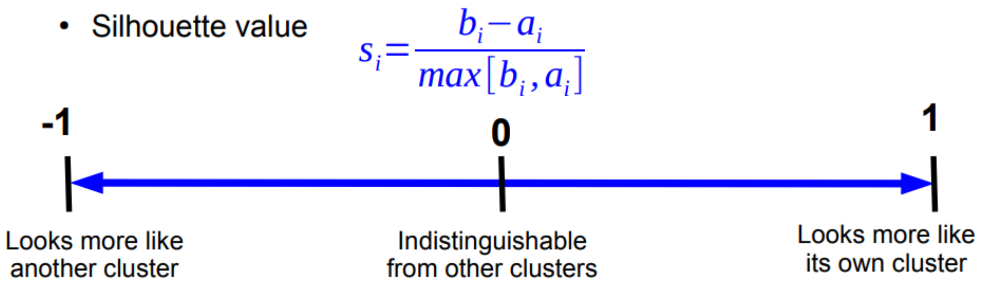
Backup

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Silhouette Measure:
  - $a_i = \langle d_{i,j} \rangle_{j \neq I}$
    - (avg. distance b/w jet candidate and others in **its own cluster**)
  - $b_i = \langle d_{i,j} \rangle$
    - (avg. Distance b/w jet candidate and others **in other clusters**)
  - $s_i = (b_i - a_i)/(\max[b_i, a_i])$

Steffanic
"Separating Signal from Combinatorial Jets in a High Background Environment" (arXiv entry)



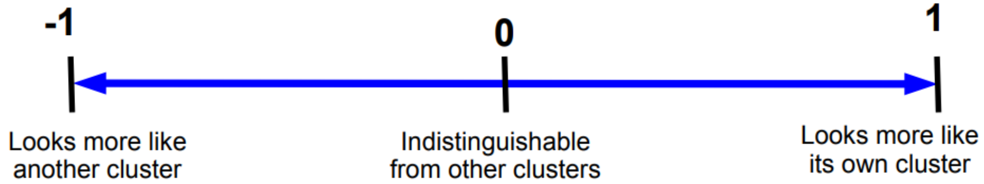


Taken from [C. Nattrass]

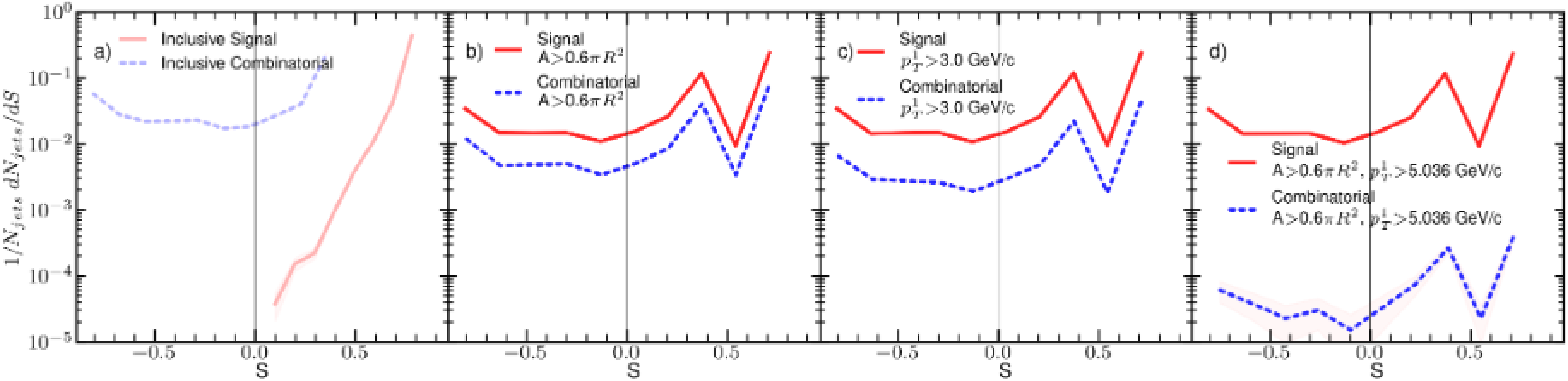
Backup

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Silhouette value $s_i = \frac{b_i - a_i}{max[b_i, a_i]}$



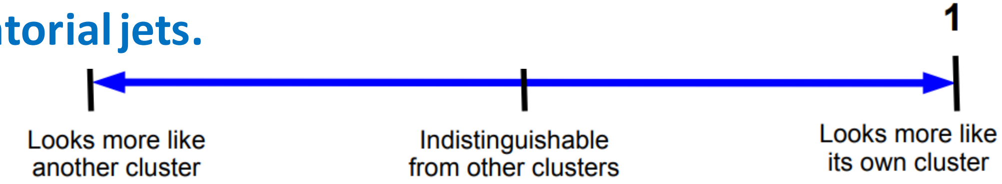
Taken from C. Nattrass



Backup

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

These cuts work well but ALWAYS leave a population of combinatorial jets that look like signal jets. The addition of the leading hadron pT cut removes a lot of combinatorial jets.



Taken from C. Nattrass