The information content of jet quenching and machine learning assisted observable design

We employ machine learning techniques to identify important features that distinguish jets produced in heavy-ion collisions from jets produced in proton-proton collisions [1]

[1] Yue Shi Lai, James Mulligan, Felix Ringer, MP [JHEP 10 (2022) 011]

See also https://arxiv.org/abs/2210.06450

Formulation of the problem

Use machine learning techniques to understand what observable are sensitive to jet quenching

- Prerequisites for a meaningful answer
  - **Completeness of data** <=> ML model reflects the training data
    - => Need clear understanding what questions we are trying to answer
    - => Select the input data appropriately <=> e.g. **IF model learns from input missing important elements of the problem it will produce meaningless (or potentially misleading) results**
  - **Model independent/agnostic methods**
  - **Human understandable/interpretable result**
  - **Theoretical understanding**

Formulation of the problem

Use machine learning techniques to understand what observable are sensitive to jet quenching

- Prerequisites for a meaningful answer
  - Completeness of data
  - Model independent/agnostic methods
  - Understandable/interpretable result
  - Theoretical understanding

- … our approach
  - include all relevant effects - e.g. background!
  - MC independent, use exp. data
  - symbolic regression; keep leading terms
  - formulate observables in connection with a theoretical formalism

What are the **maximally discriminating observables (understandable on theoretical level) of jets in AA from jets in pp?**

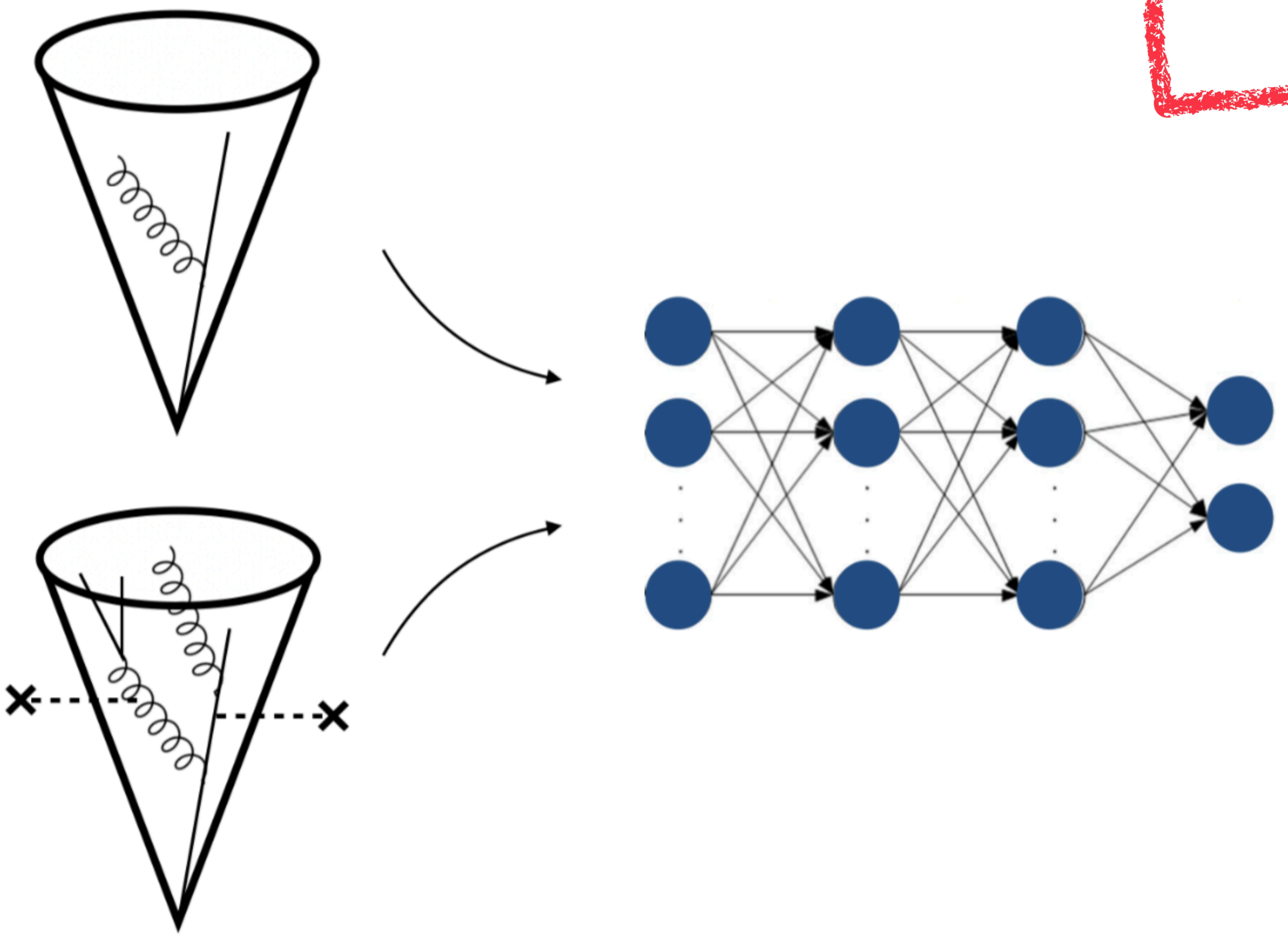
Formulation of the problem

“Simple” question:
Which jets are quenched and which are not?

A “trivial” answer:
Jets in pp are not quenched

A classification problem

CLEAR LABELS IN EXP. DATA!



pp → PYTHIA

vs.

AA → JEWEL

Eventually learn from exp data directly!

Figure 1. Schematic illustration of jets in pp (left) and heavy-ion AA (right) collisions. Interactions with the QGP can lead to a modification of the jet substructure. By training a binary classifier, the machine learns the relevant information that distinguishes jets in pp and AA collisions.

IRC-safe vs. IRC-unsafe architectures

Permutation-invariant neural networks based on deep sets

Zaheer et al. 1703.06114
Wagstaff et al. 1901.09006
Bloem-Reddy, Teh JMLR 21 90 (2020)

Unordered, variable-length sets of particles as input

Komiske, Metodiev, Thaler JHEP 01 (2019) 121

Particle Flow Network (PFN)

$$f(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} \Phi\left(p_i\right)\right)$$

latent space $d = 256$

Classifier

DNNs

Includes IRC-unsafe information

Energy Flow Network (EFN)

$$f(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} z_i \Phi\left(\hat{p}_i\right)\right)$$

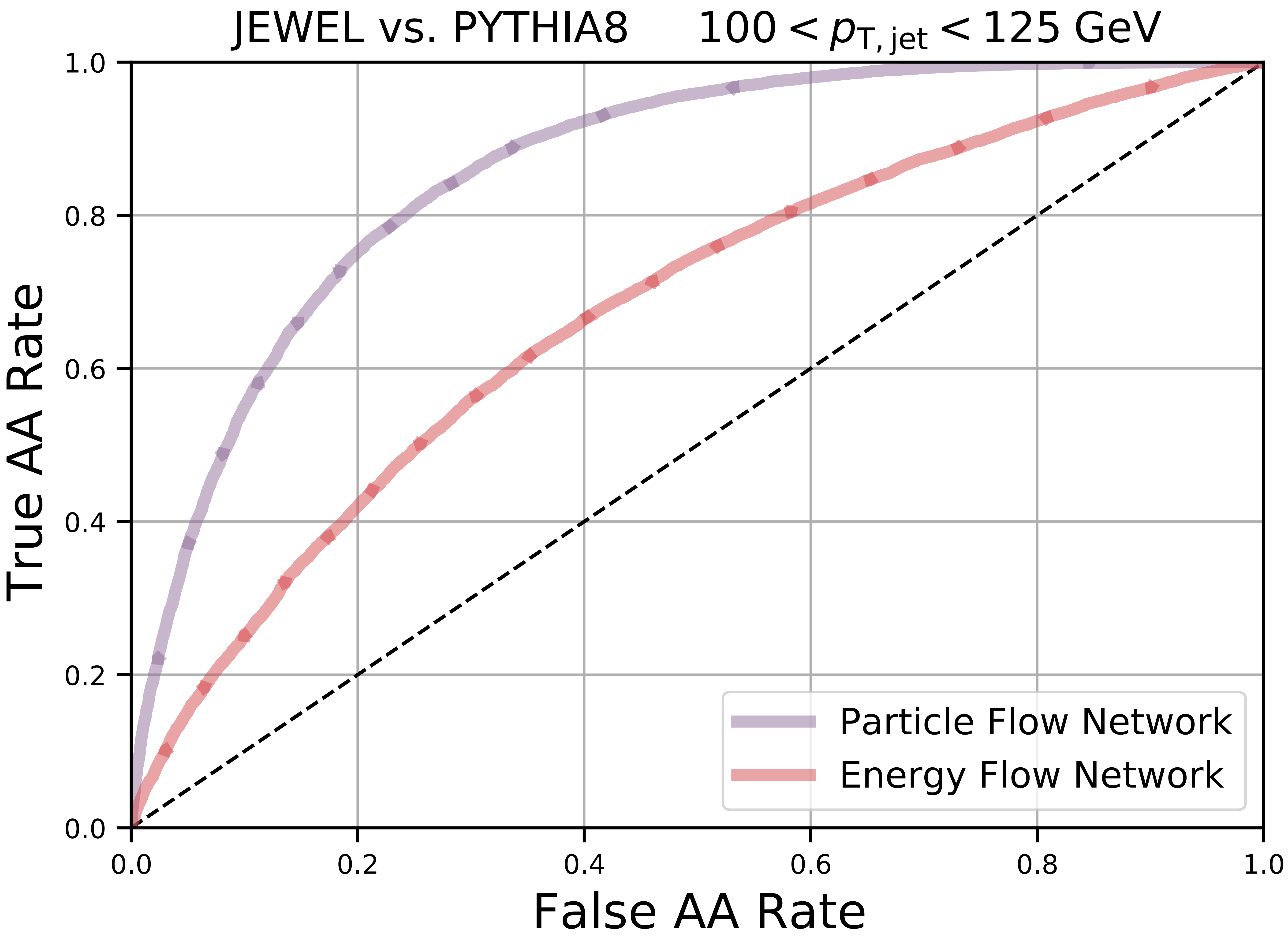
Classifier

DNNs

Includes only IRC-safe information

IRC-safe vs. IRC-unsafe physics

Lai, Mulligan, Płoskoń, Ringer JHEP 10 (2022) 011



We compare the IRC-unsafe network (PFN) to an IRC-safe network (EFN)

$$\underbrace{f(p_1, \ldots, p_M)}_{\text{Classifier}} = \underbrace{F}_{\text{DNNs}}\left(\sum_{i=1}^{M} z_i \underbrace{\Phi}_{\text{DNNs}} \left(\hat{p}_i \right) \right)$$

IRC-unsafe information contains significant discriminating power

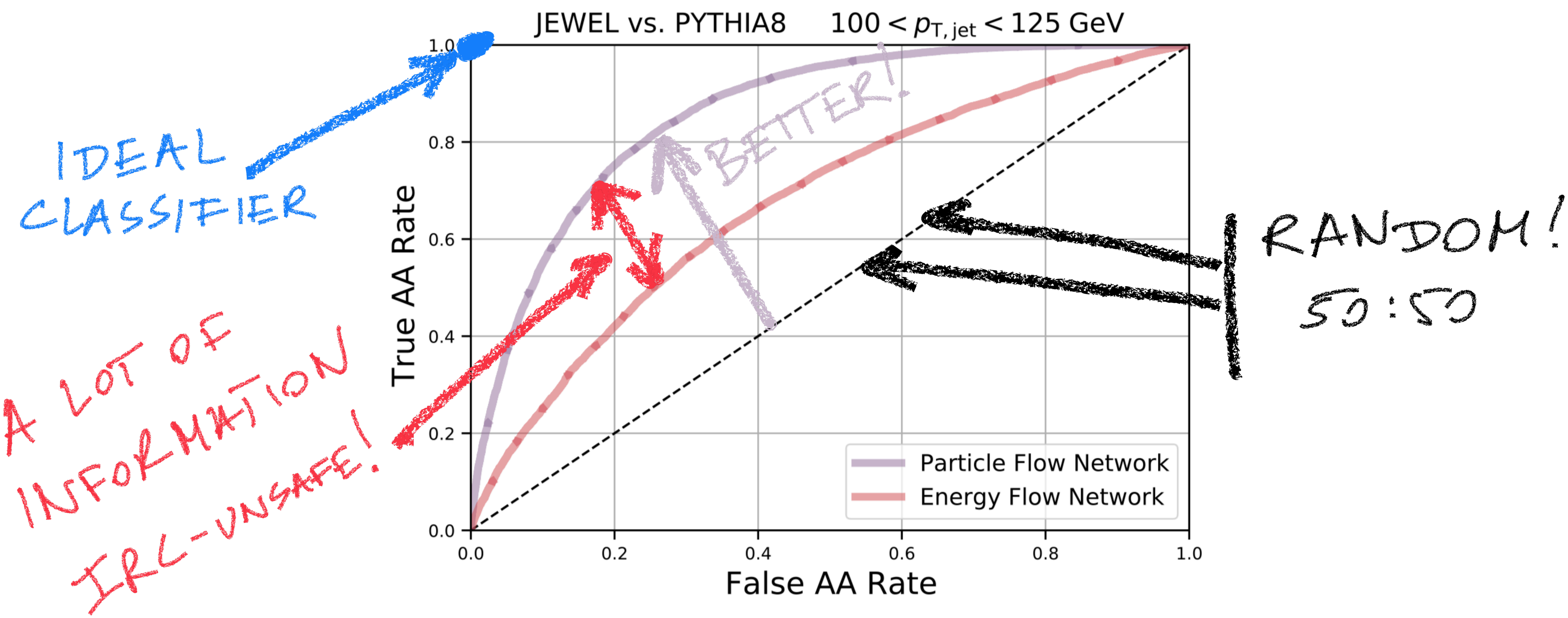


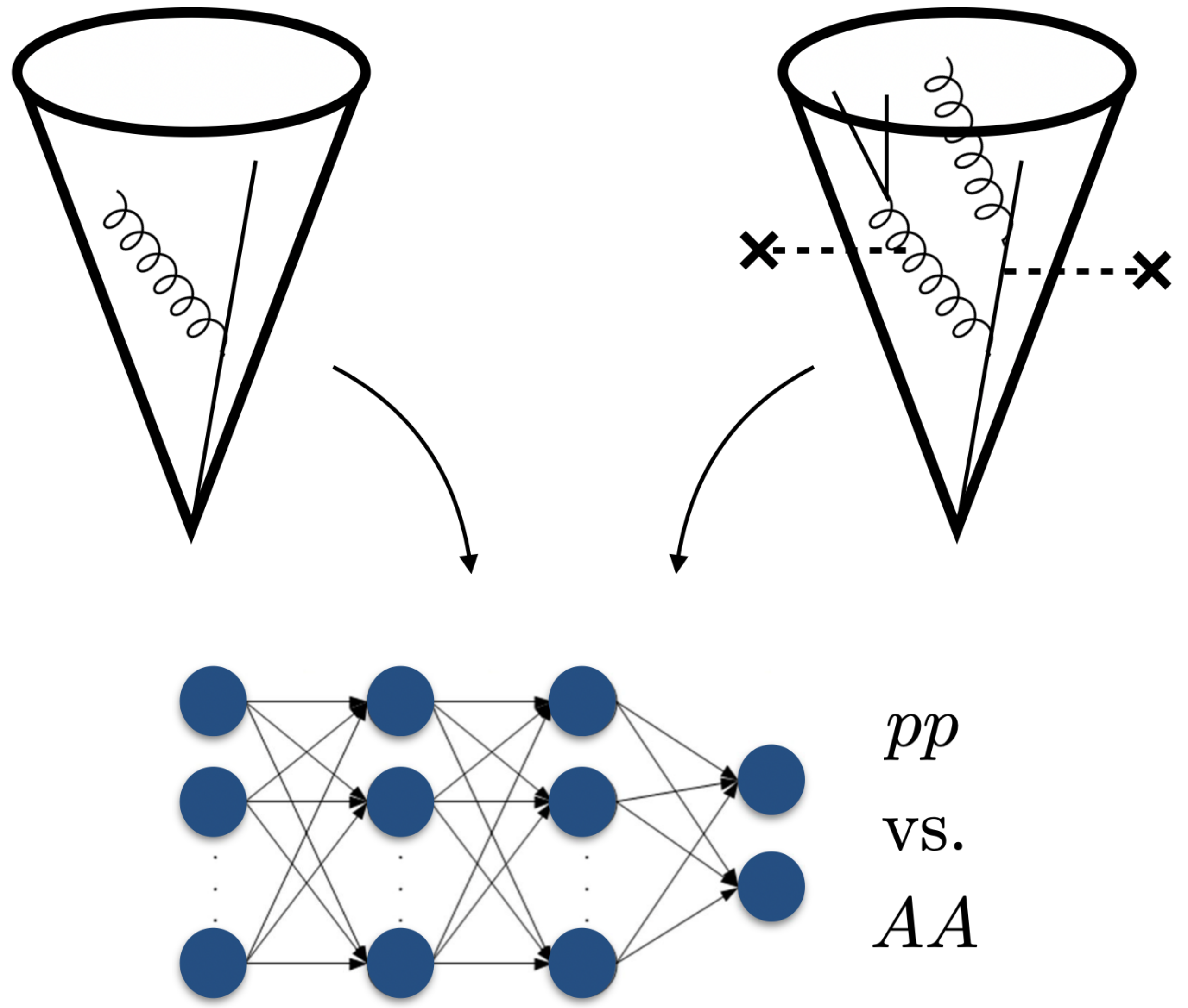
Figure 2. Classification performance of pp vs. AA jets quantified in terms of ROC curves using IRC-unsafe PFNs and IRC-safe EFNs. The jet samples in pp and AA collisions are obained from Pythia 8 [70] and Jewel [72, 73].

Guiding future measurements?
Observable design…

Observable design

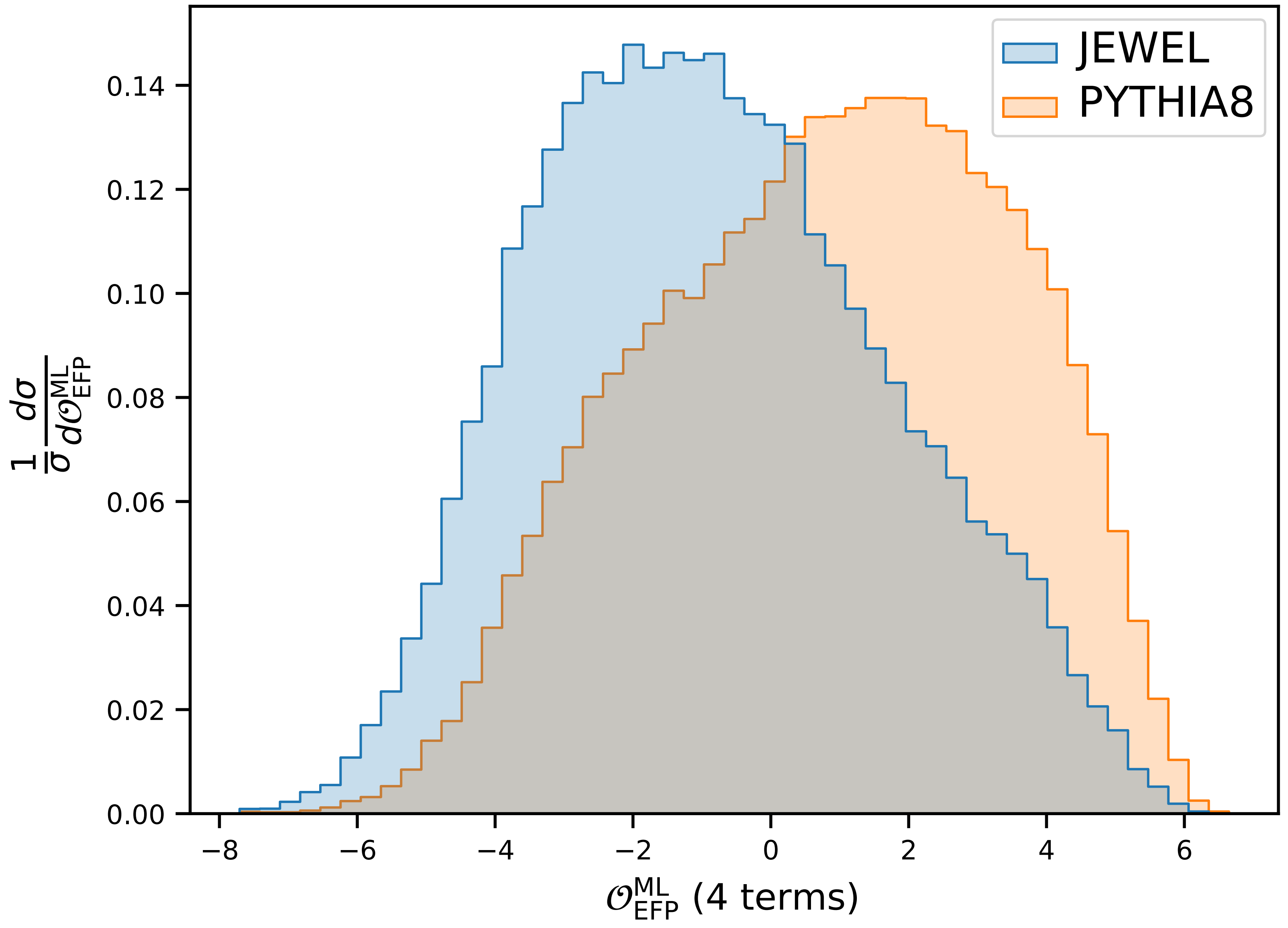
Lai, Mulligan, Płoskoń, Ringer JHEP 10, 011 (2022)

1. Design the **most strongly modified observable** that is **theoretically calculable**
2. Optimize discriminating power vs. complexity (trade d. power for simplicity)



$$\max_{c_G} \left| \frac{d\sigma_{AA}}{d\sigma_{pp}}(c_G) - 1 \right|$$

Symbolic regression



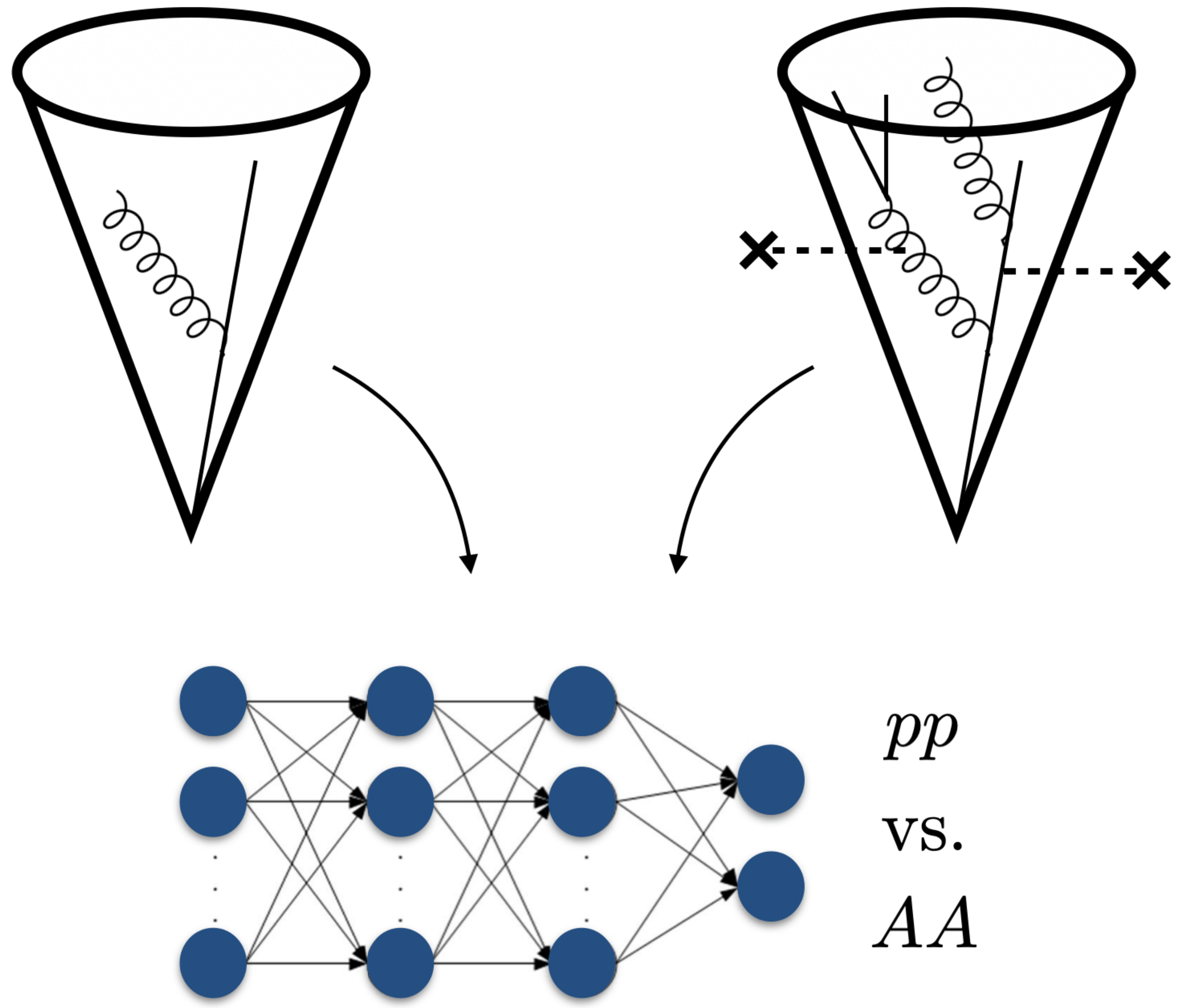
First step in a new paradigm: data-driven design of complete set of calculable observables

Complementary to Bayesian approach

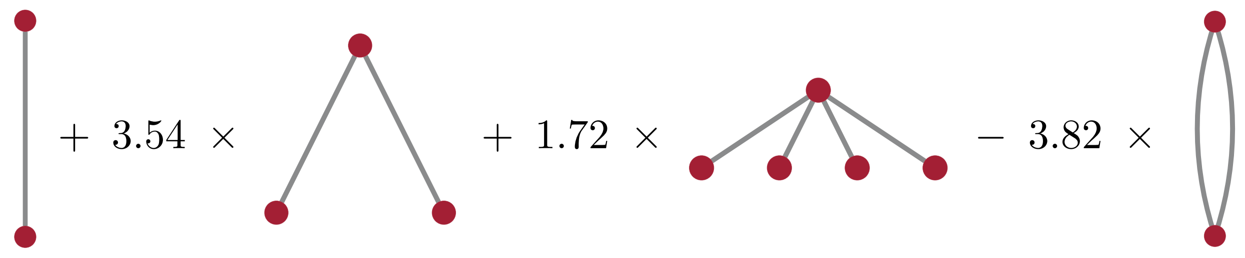
Observable design

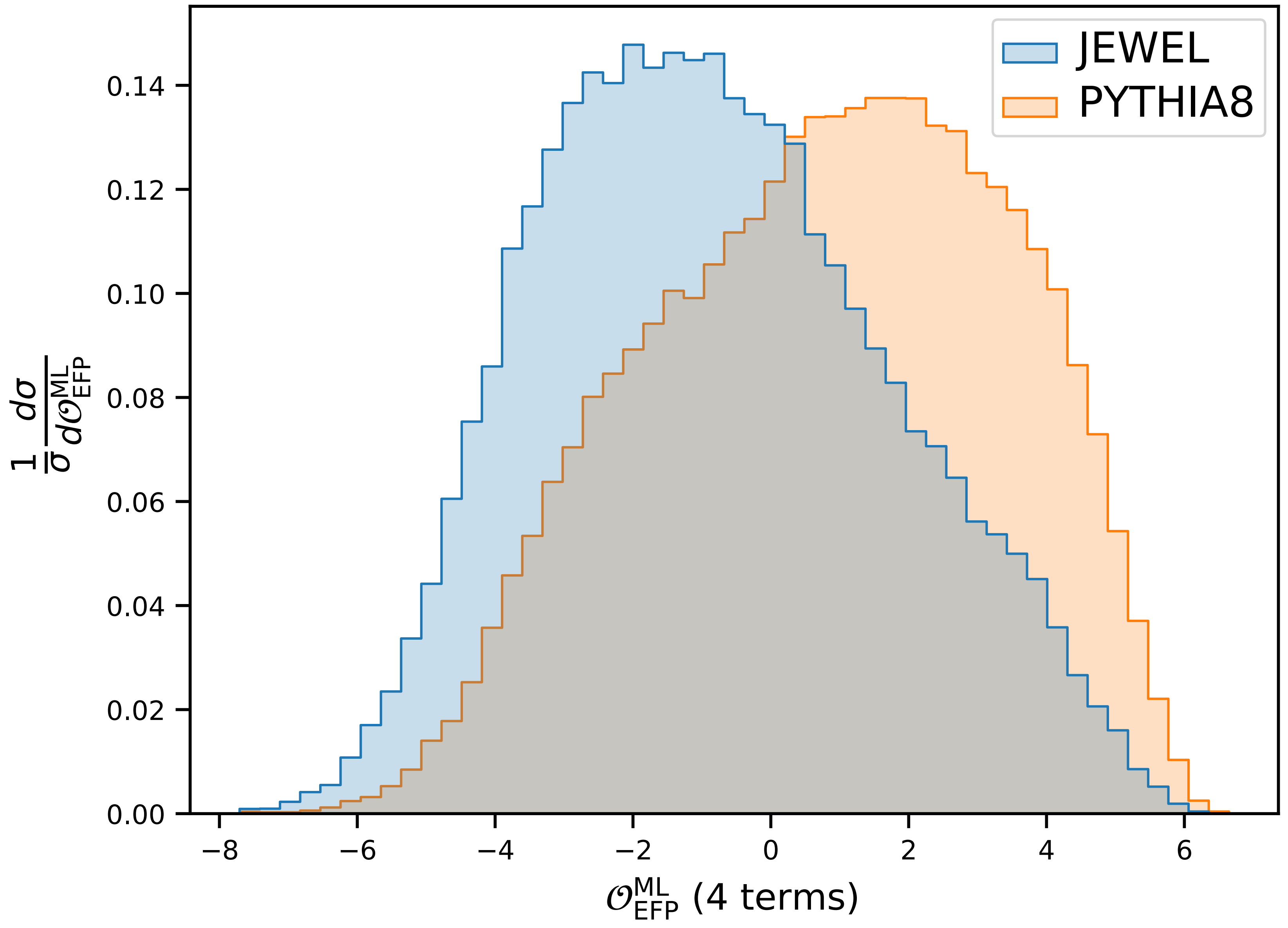
Lai, Mulligan, Płoskoń, Ringer JHEP 10, 011 (2022)

1. Design the **most strongly modified observable** that is **theoretically calculable**
2. Optimize discriminating power vs. complexity (trade d. power for simplicity)



$$\max_{c_G} \left| \frac{d\sigma_{AA}}{d\sigma_{pp}}(c_G) - 1 \right|$$



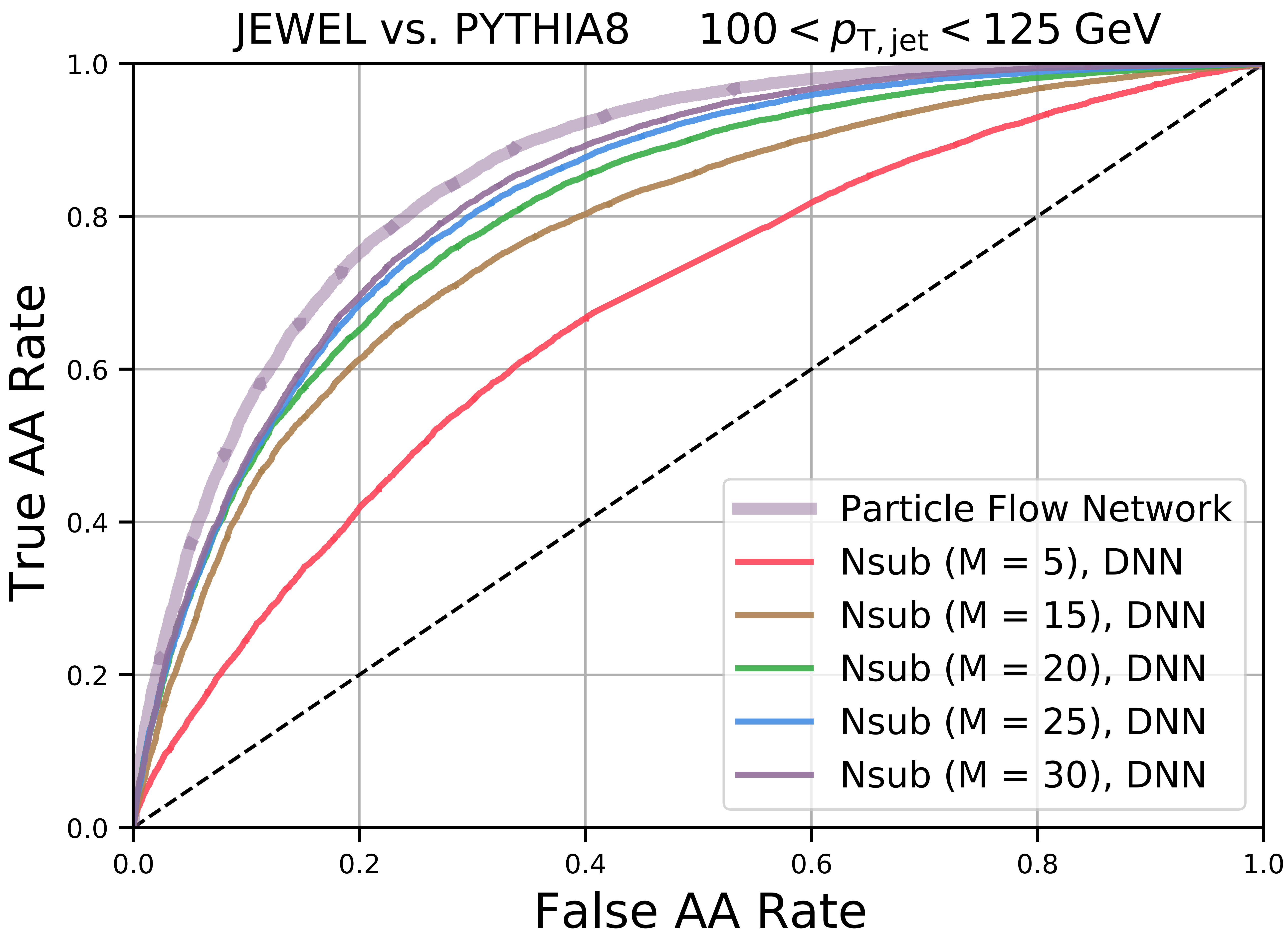


First step in a new paradigm: data-driven design of complete set of calculable observables

Complementary to Bayesian approach

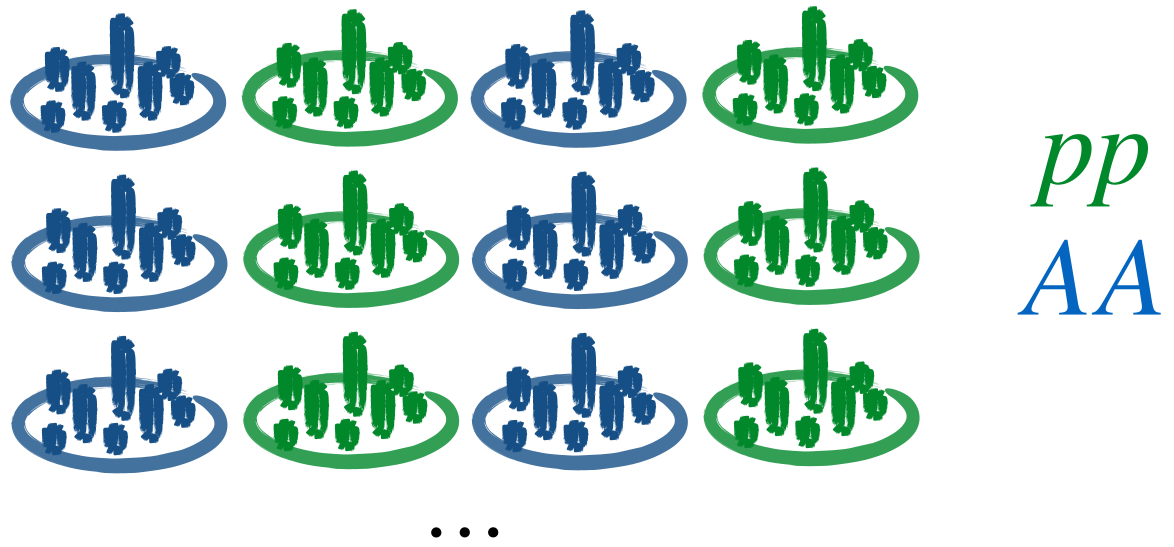
The information content of jet quenching

Lai, Mulligan, Płoskoń, Ringer JHEP 10, 011 (2022)



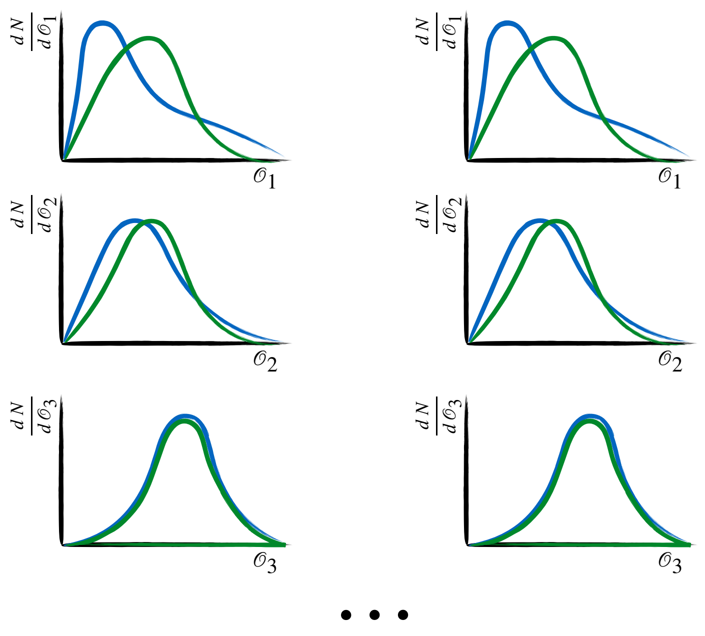
"Optimal" classifier

- Input: four-vectors of all jet particles



Jet substructure basis

- Input: $3M - 4$ observables



Systematic approach: how many observables does one need to measure?

Encoding information in terms of observables

N-subjetiness basis and Energy Flow Polynomials

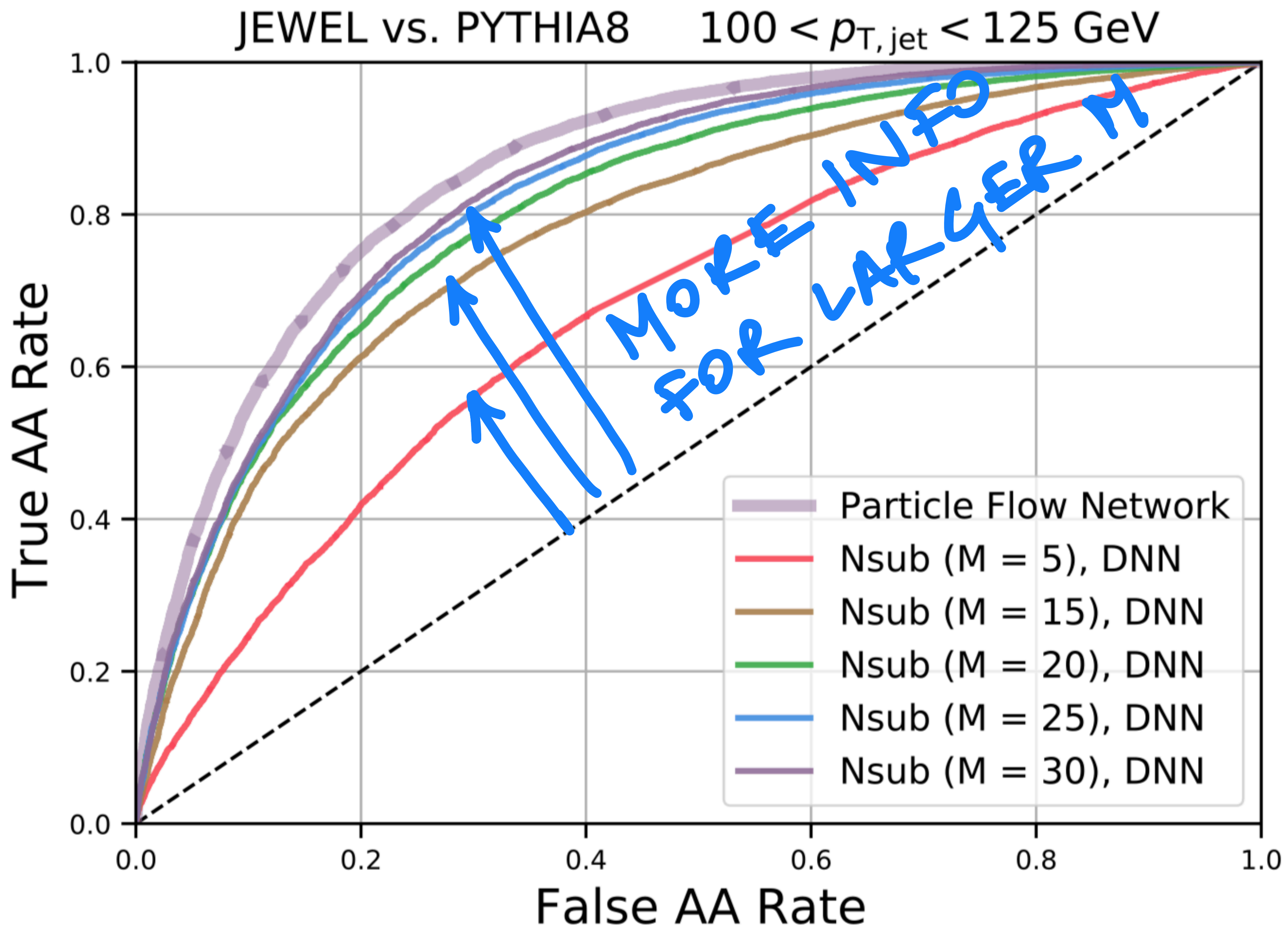


Figure 4. ROC curves for jets in pp vs. AA collisions using the N-subjettiness basis. For comparison we also show the result obtained using the classifier based on PFNs.

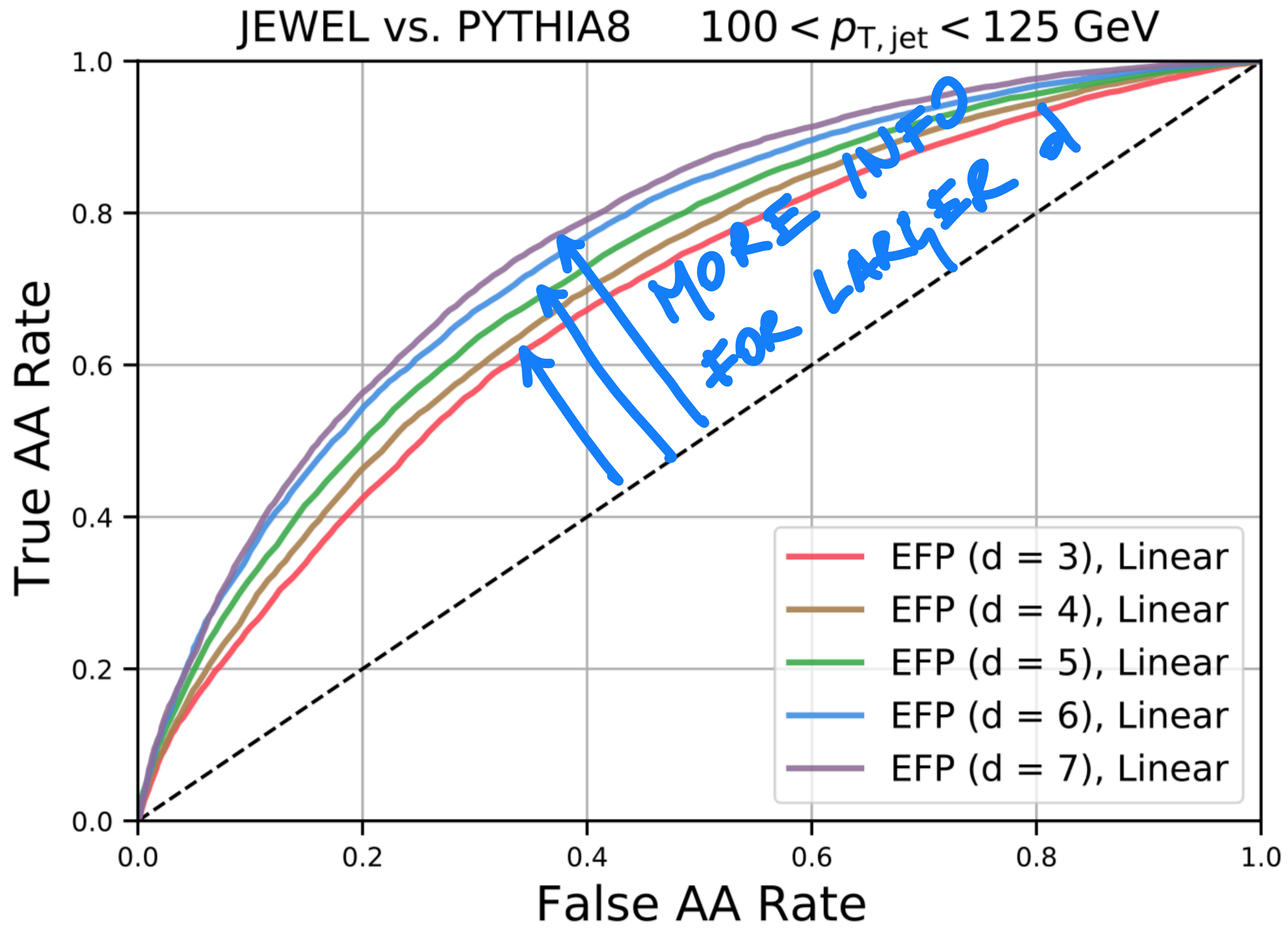


Figure 6. ROC curves for jets in pp vs. AA collisions using the EFP basis up to degree 7.

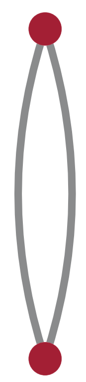
Observable design

Lai, Mulligan, Płoskoń, Ringer JHEP 10, 011 (2022)

Goal: design a *minimal* set of observables that are theoretically calculable

Example: design a **single observable** that is **maximally modified**

$$\mathcal{O}^{\mathrm{ML}} = \sum_{G \in \mathcal{G}} c_G \, \mathrm{EFP}_G$$ where EFP_G are energy flow polynomials

e.g. $\mathrm{EFP}_G =$  $= \sum_{i_1=1}^{M} \sum_{i_2=1}^{M} z_{i_1} z_{i_2} \theta^2_{i_1,i_2}$

We can determine the coefficients using **symbolic regression**

- Regularization allows us to choose complexity of allowed solution (e.g. how many nonzero coefficients)

$$\max_{c_G} \left| \frac{d\sigma_{AA}}{d\sigma_{pp}}(c_G) - 1 \right|$$

ML-assisted observable design

Lai, Mulligan, Płoskoń, Ringer JHEP 10, 011 (2022)

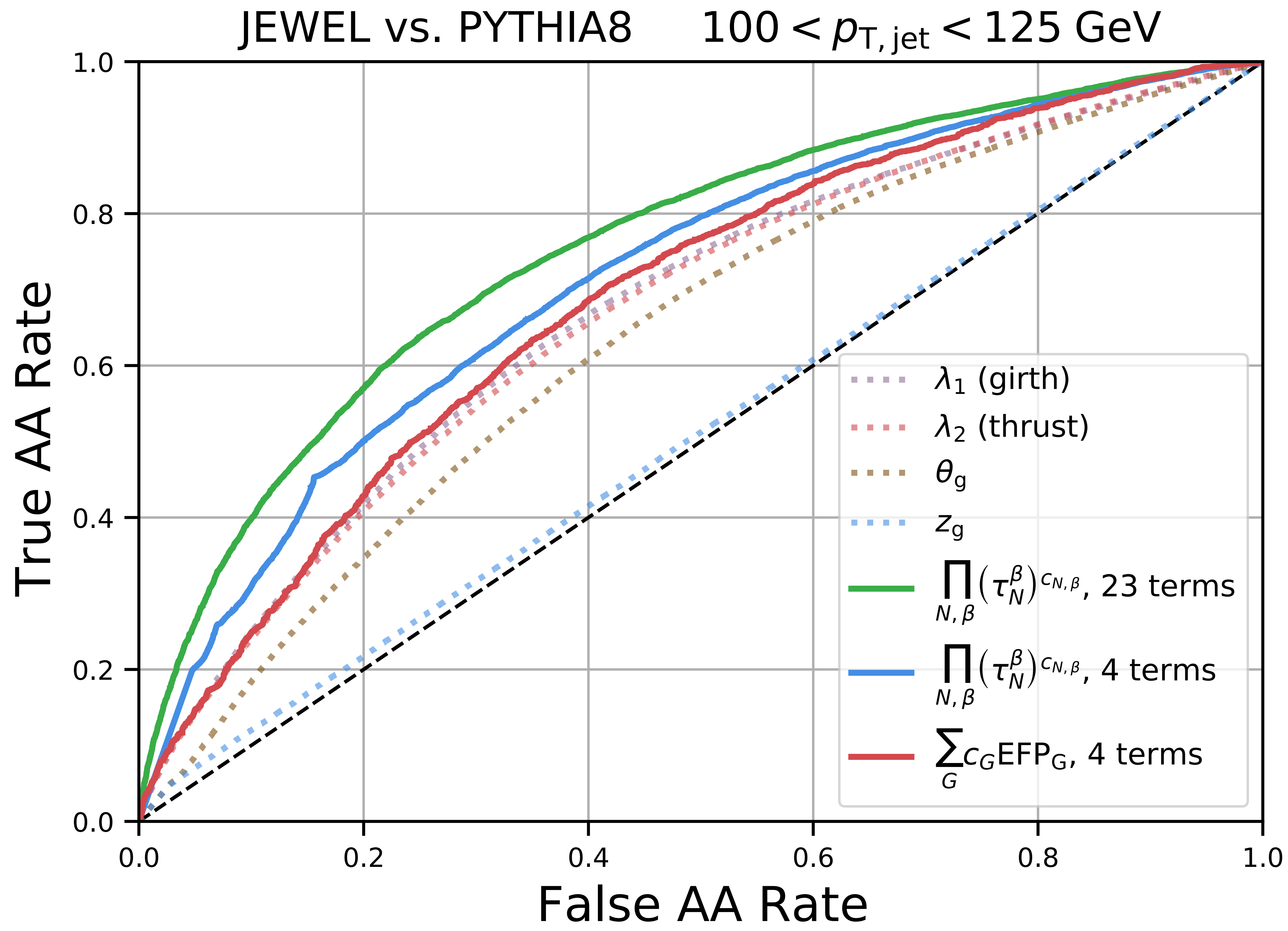
Lasso regression

$$O = \prod_{N<K,\ \beta\in\{0.5,1,2\}} \left(\tau_N^\beta\right)^{c_{N\beta}}$$

Stronger regularization drives $c_{N\beta}$ to zero

$\alpha = 0.01$ → 24 terms
$\alpha = 0.1$ → 4 terms
$\alpha = 0.5$ → 1 term

e.g. $(\tau_1^2)^{1.437}(\tau_5^2)^{0.068}(\tau_6^2)^{1.712} \times \ldots$



By training ML classifier and balancing the tradeoff of discriminating power and complexity, we can design the *most strongly modified* calculable observable

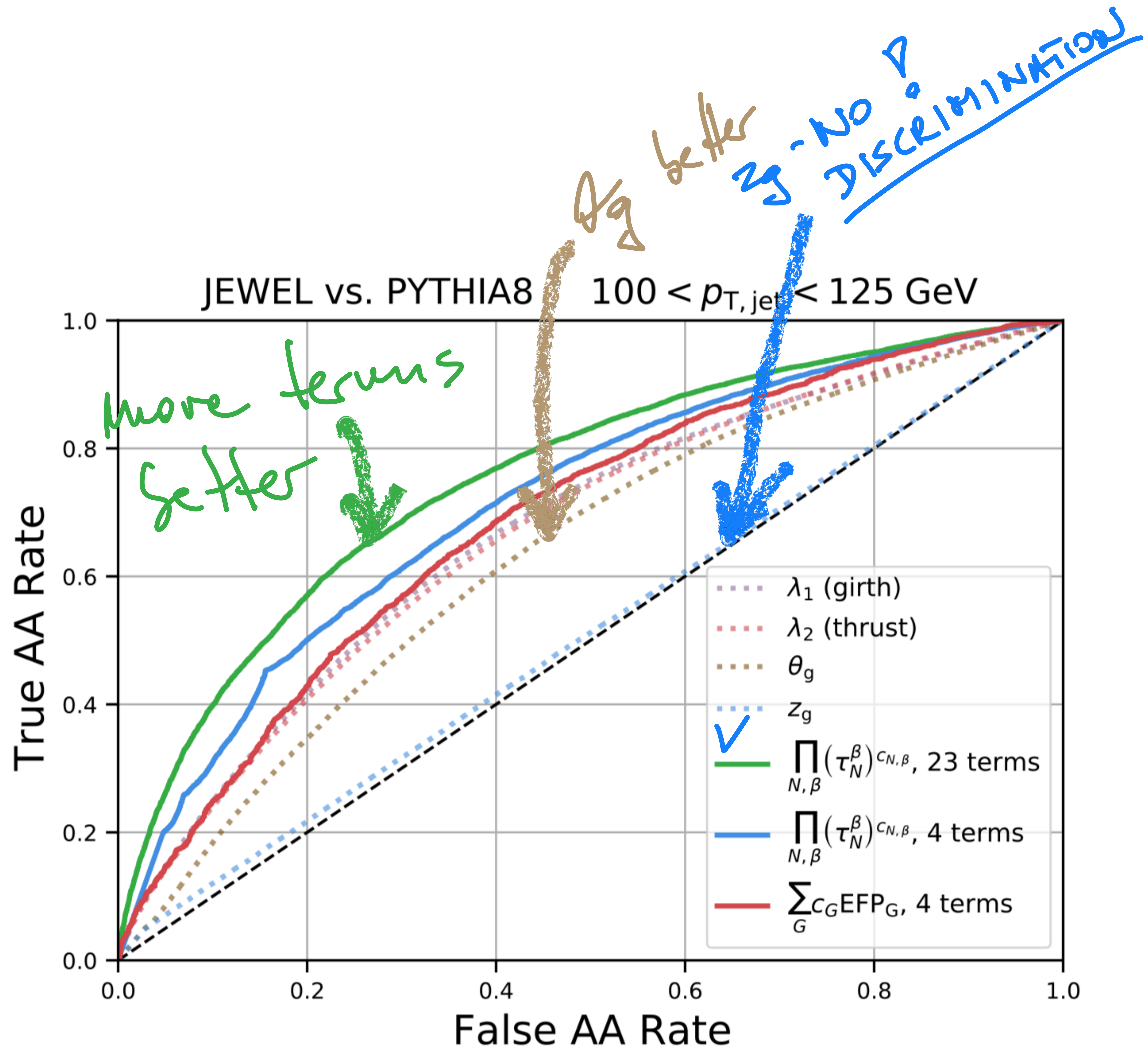


Figure 7. ROC curves for the Lasso regression using the N-subjettiness basis and EFPs. For comparison we also show the result for typical observables in heavy-ion collisions.

Realistic conditions include BACKRGOUND!
(aka heavy-ion underlying event)

1. Background <=> Noise

=> Loss of discrimination power

1. Background <=> Noise
   => Loss of discrimination power

2. Finite accuracy of background subtraction
   => Loss of signal/information

Jets within heavy-ion events

Adding realism to the problem - background and information loss

1. Background <=> Noise
   => Loss of discrimination power

2. Finite accuracy of background subtraction
   => Loss of signal/information

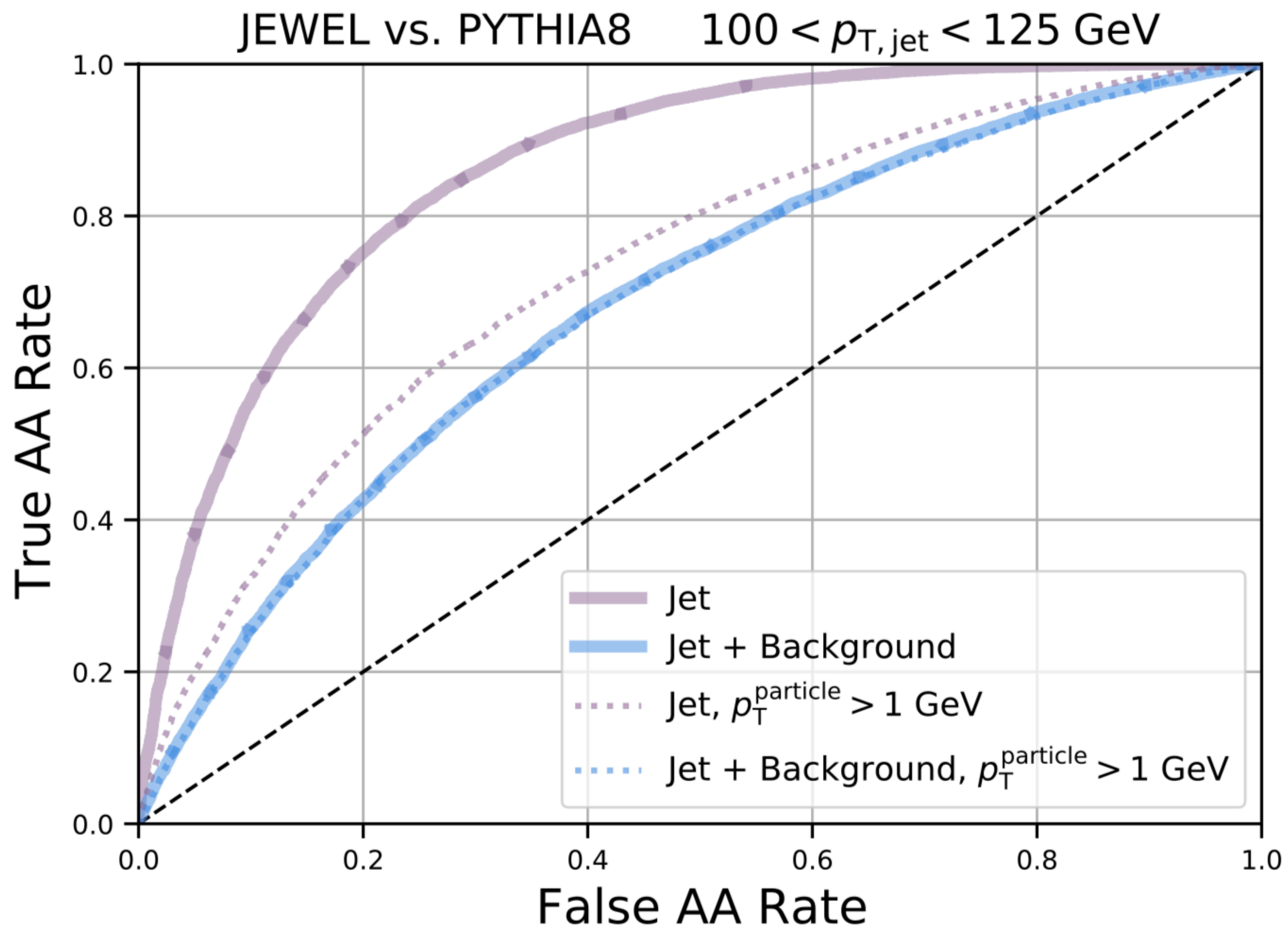


Figure 9. ROC curves for PFNs trained with (i) PYTHIA8/JEWEL jets, (ii) jets clustered from a combination of PYTHIA8/JEWEL events with a thermal background, with event-wide constituent subtraction applied ($R_{\mathrm{max}} = 0.25$), (iii) PYTHIA8/JEWEL jets only considering jet constituents with $p_T > 1$ GeV, and (iv) jets clustered from a combination of PYTHIA8/JEWEL events with a thermal background, only considering jet constituents with $p_T > 1$ GeV, with event-wide constituent subtraction applied ($R_{\mathrm{max}} = 0.25$).

Jets within heavy-ion events

Adding realism to the problem - background and information loss

1. Background <=> Noise
   => Loss of discrimination power

2. Finite accuracy of background subtraction
   => Loss of signal/information

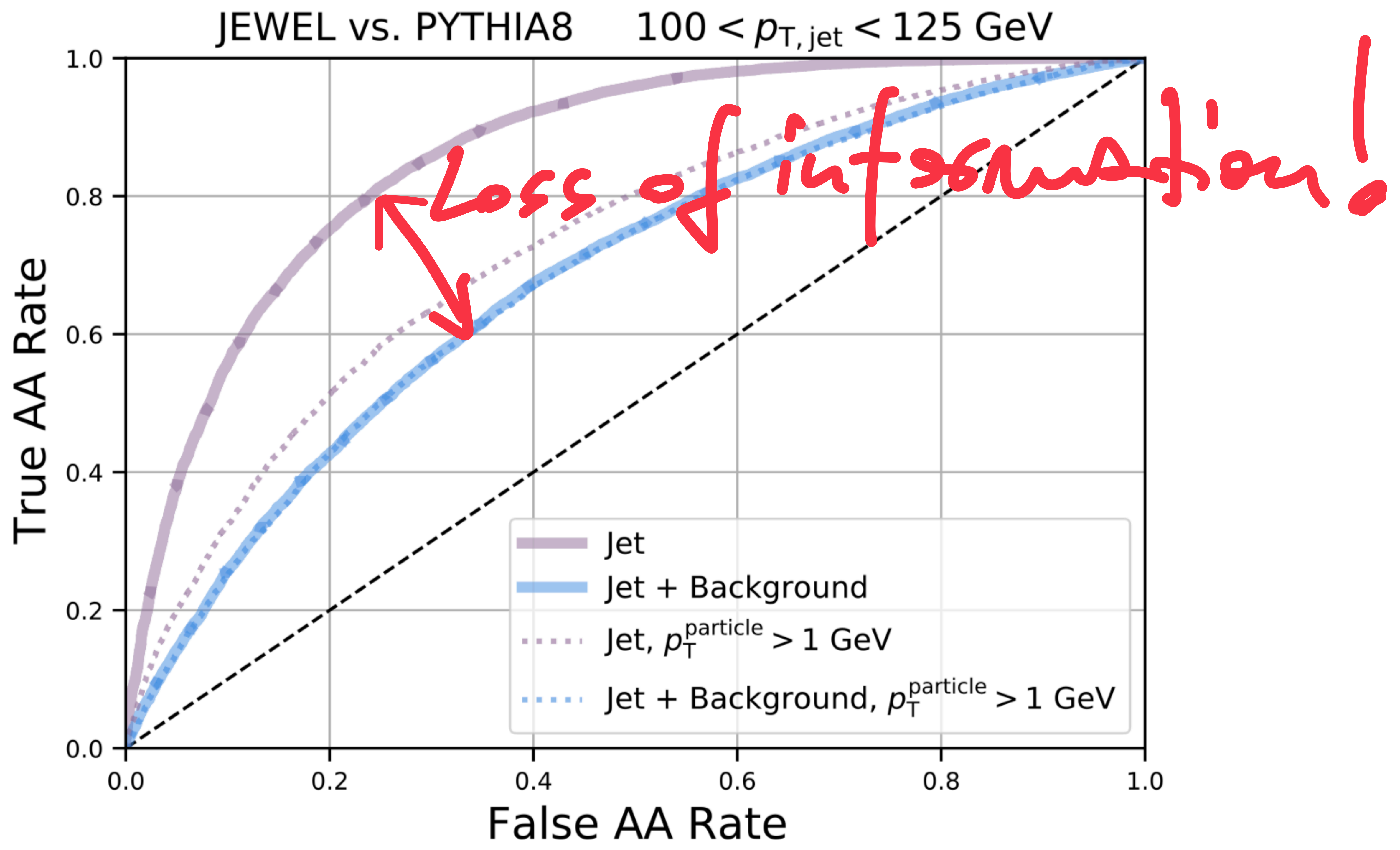


Figure 9. ROC curves for PFNs trained with (i) PYTHIA8/JEWEL jets, (ii) jets clustered from a combination of PYTHIA8/JEWEL events with a thermal background, with event-wide constituent subtraction applied ($R_{\mathrm{max}} = 0.25$), (iii) PYTHIA8/JEWEL jets only considering jet constituents with $p_T > 1$ GeV, and (iv) jets clustered from a combination of PYTHIA8/JEWEL events with a thermal background, only considering jet constituents with $p_T > 1$ GeV, with event-wide constituent subtraction applied ($R_{\mathrm{max}} = 0.25$).

Jets within heavy-ion events

Adding realism to the problem - background and information loss

1. Background <=> Noise
   => Loss of discrimination power

2. Finite accuracy of background subtraction
   => Loss of signal/information

Introduction of the realistic conditions (experimental backgrounds) critical applicability test!

ML w/o UE not OK!

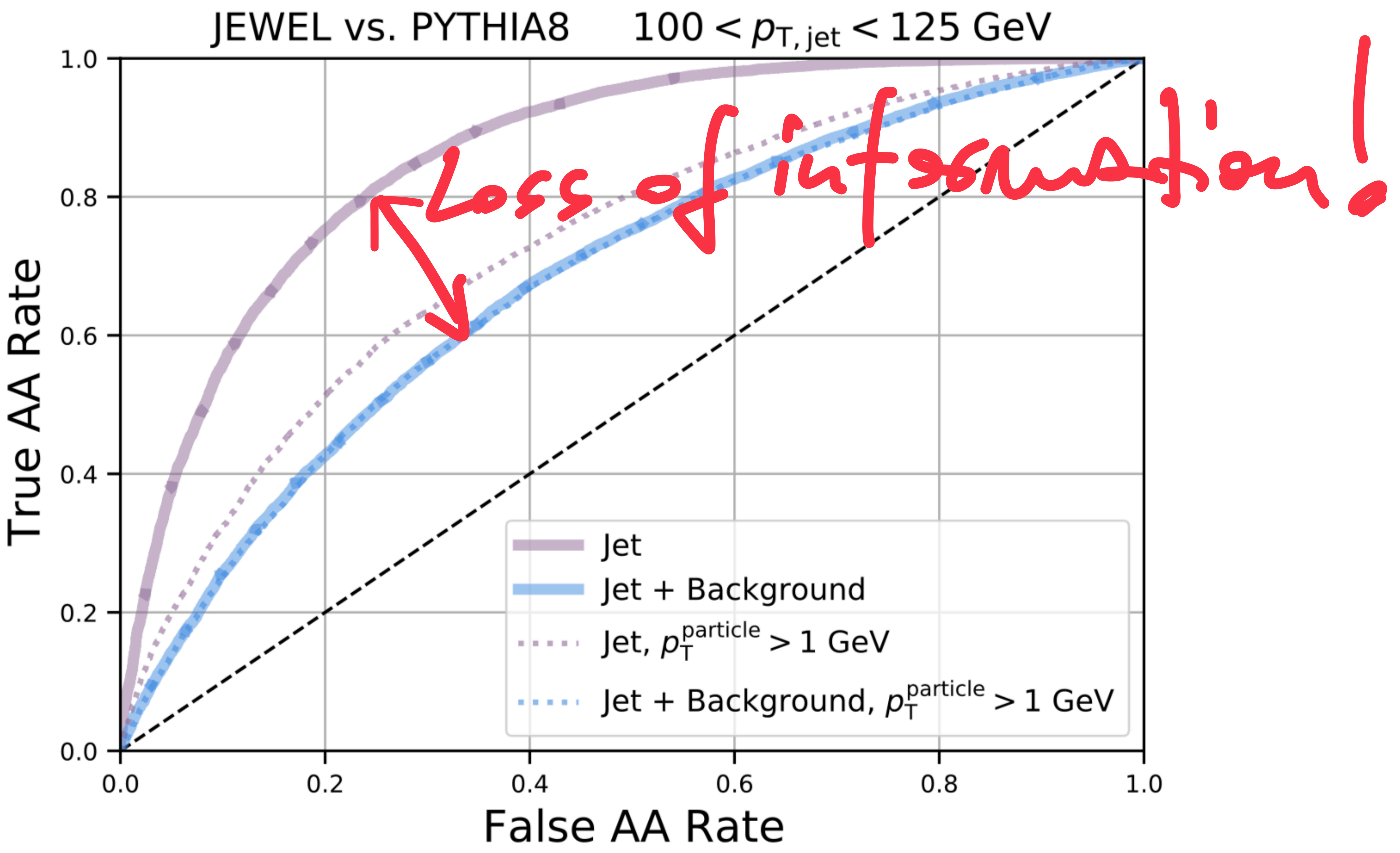
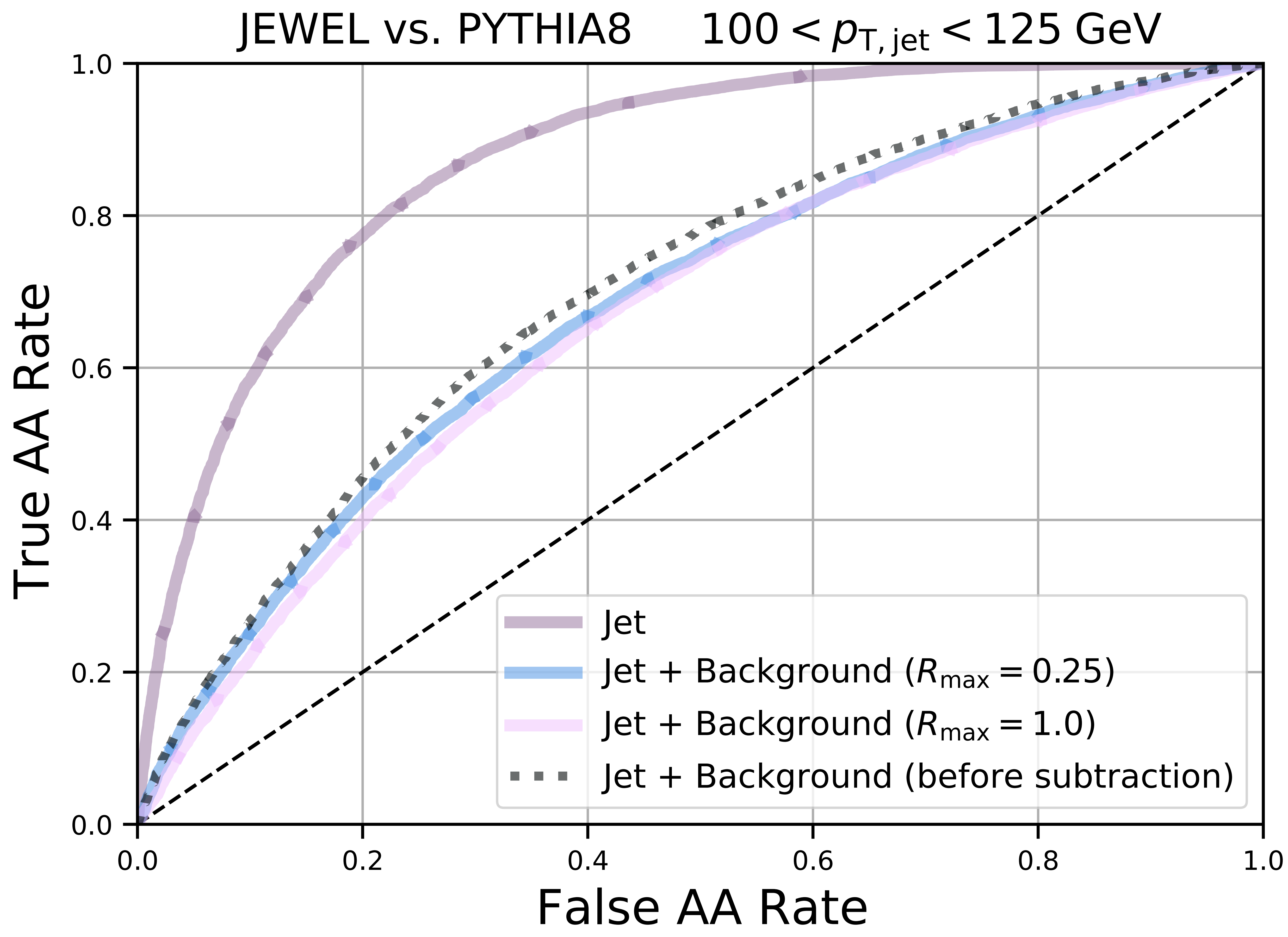


Figure 9. ROC curves for PFNs trained with (i) PYTHIA8/JEWEL jets, (ii) jets clustered from a combination of PYTHIA8/JEWEL events with a thermal background, with event-wide constituent subtraction applied ($R_{\max} = 0.25$), (iii) PYTHIA8/JEWEL jets only considering jet constituents with $p_T > 1$ GeV, and (iv) jets clustered from a combination of PYTHIA8/JEWEL events with a thermal background, only considering jet constituents with $p_T > 1$ GeV, with event-wide constituent subtraction applied ($R_{\max} = 0.25$).

Experimental guidance from ML

Lai, Mulligan, Płoskoń, Ringer JHEP 10, 011 (2022)

To what extent does the background destroy discriminating power?



Large, irrecoverable information loss

- Soft physics
- Background subtraction algorithm

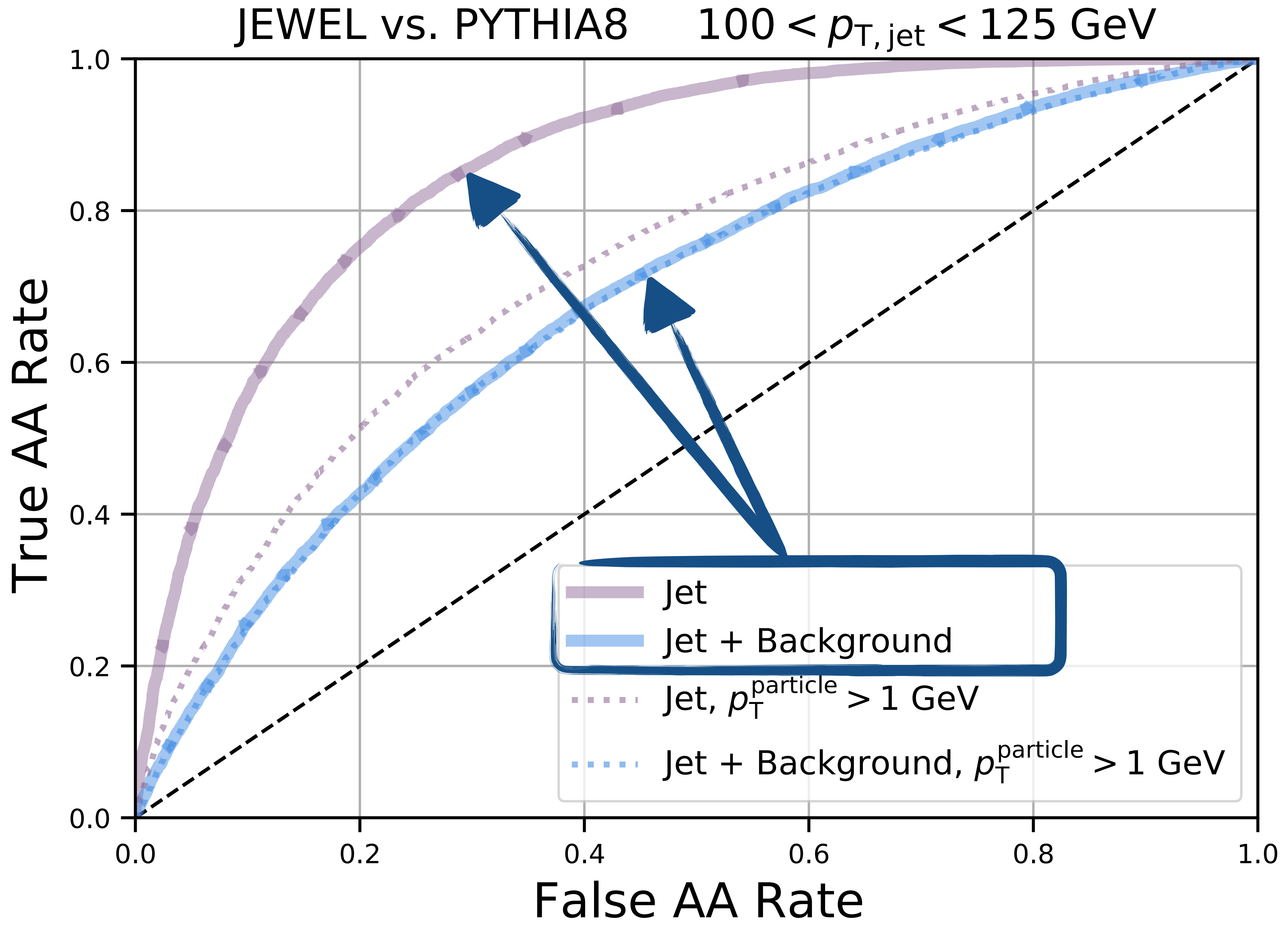
Delicate challenge: soft information is crucial to discriminate, yet UE fundamentally prevents much of this information from being accessed

First study quantifying the information loss of background subtraction algorithms

Information loss due to background

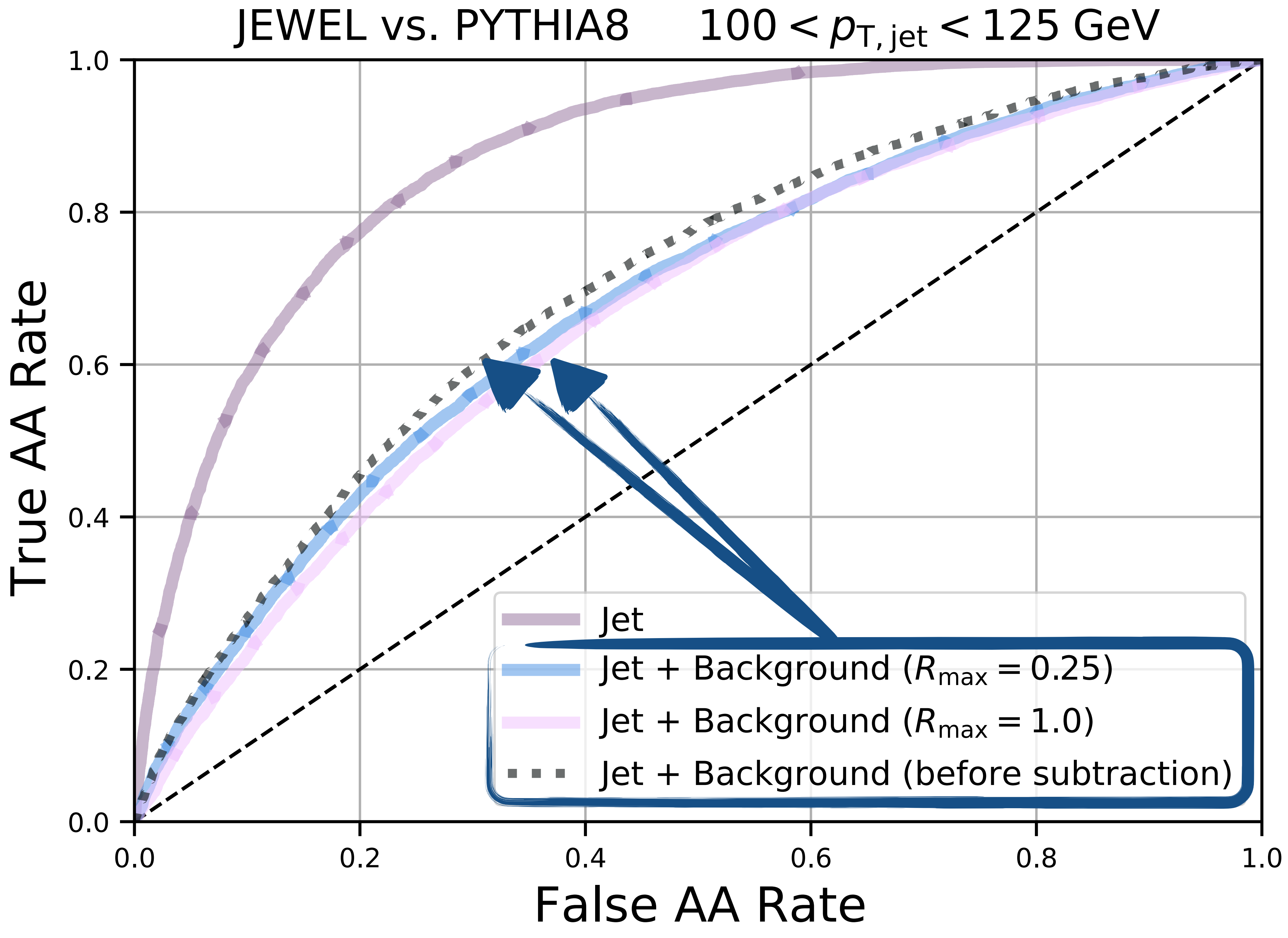
Lai, Mulligan, Płoskoń, Ringer JHEP 10 (2022) 011

Discriminating power is highly reduced by the fluctuating underlying event



Delicate challenge: soft information crucial, yet background prevents from being accessed

Background subtraction algorithms remove small but significant information



New metric to assess background subtraction algorithms

Conclusions

We propose that each of the three complementary studies in Sections 3-5 can be performed on experimental data:

- *Measuring the ROC curve.* The measured ROC curve can serve as an observable that can be compared to Monte Carlo event generators. Moreover, the distribution of information content with complete sets of jet substructure observables can provide a differential test of jet quenching models, to the extent that highly soft-sensitive observables, such as high-N N-subjettiness or high-dimension EFPs, can be reliably measured in the presence of the heavy-ion underlying event.

- *ML-assisted observable design.* Regardless of whether the classifier is trained on detector-level inputs or corrected inputs, symbolic regression can be used to identify approximate maximally discriminating observables. These identified observables can then be measured with traditional techniques: correcting for detector and background effects, and in principle comparing to jet quenching calculations.

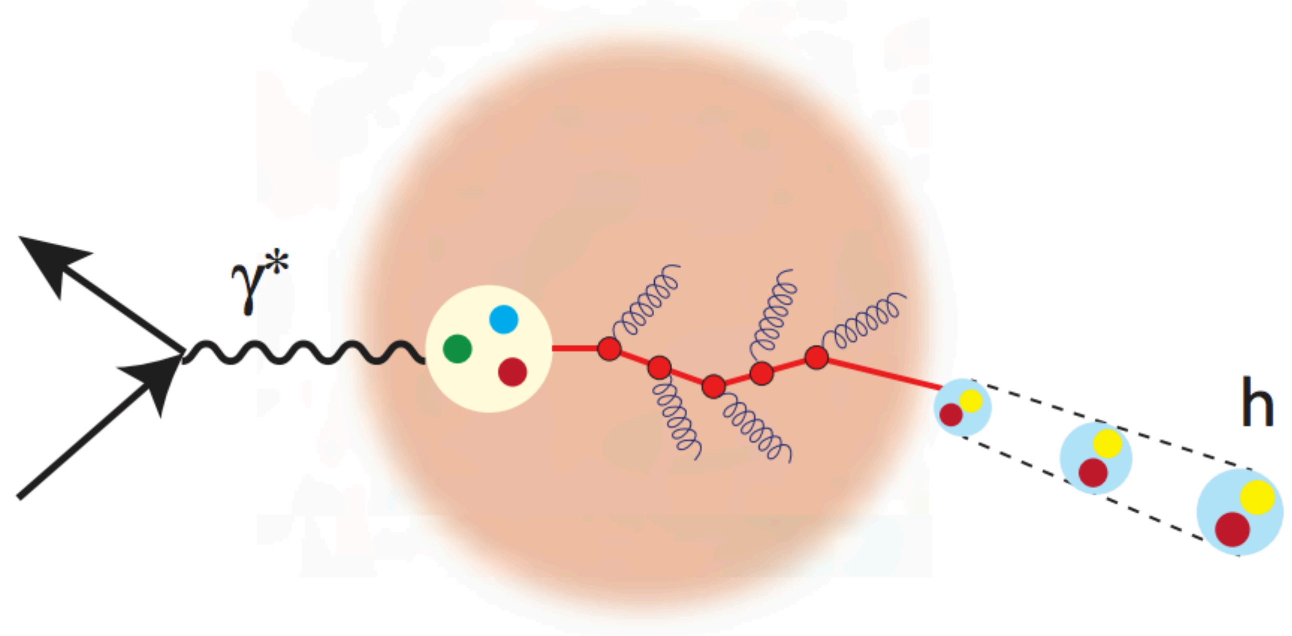
- *Information content and background subtraction techniques.* The information loss caused by various background subtraction algorithms can be quantified by comparing classification performance before and after subtraction, and can be used to select and tune subtraction algorithms to minimize information loss.

Maximizing cold nuclear matter effects

Lai, Mulligan, Płoskoń, Ringer JHEP 10 (2022) 011

Goal: extract transport properties of nuclear matter e.g. $\hat{q}$

Ru, Kang, Wang, Xing, Zhang, PRD 103, L031901 (2021)
Li, Liu, Vitev, PLB 816, 136261 (2021)

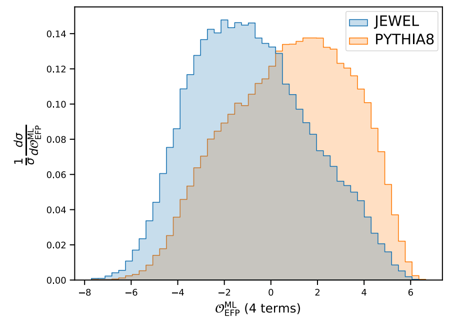


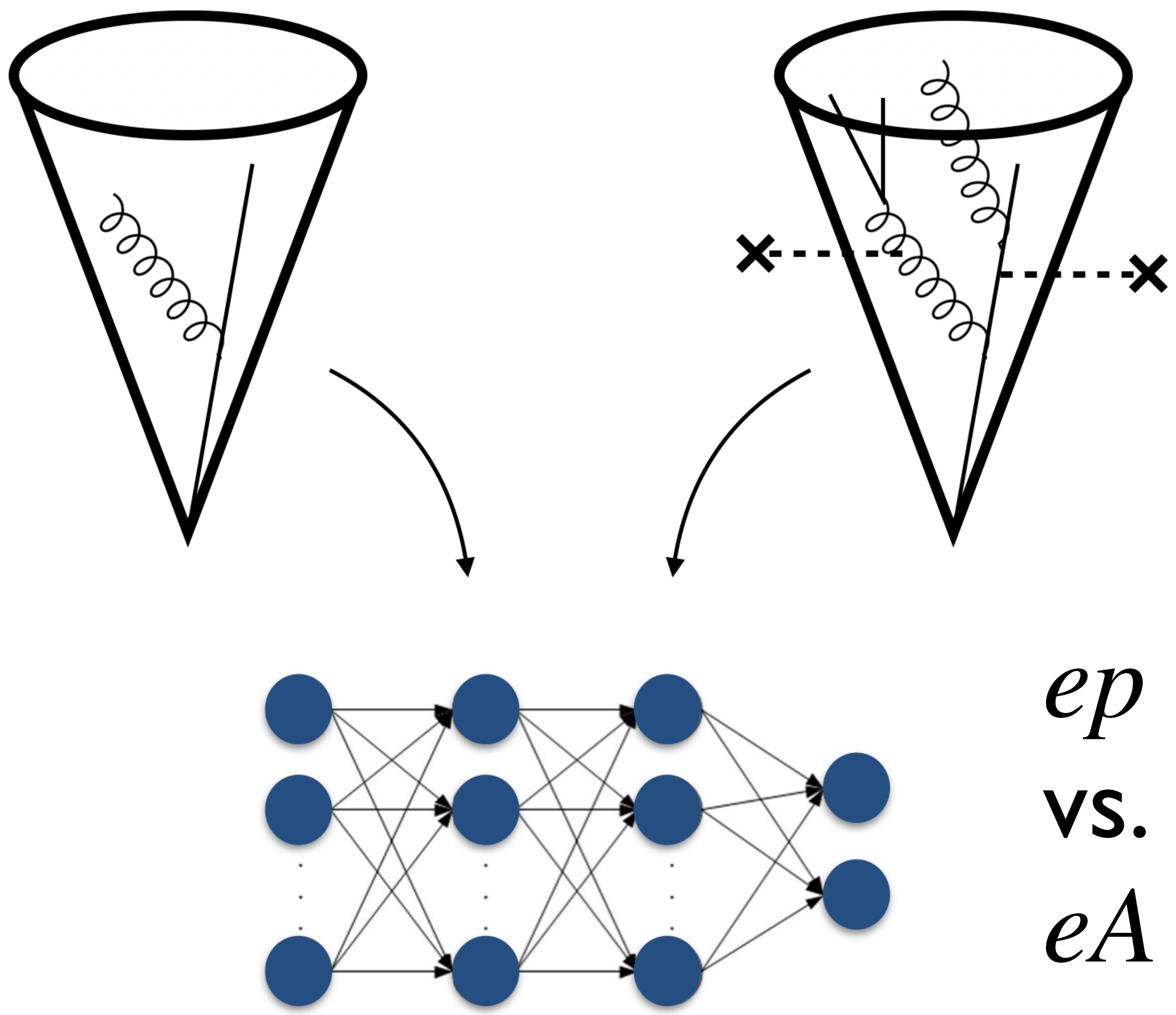
Train ML classifier to distinguish ep vs. eA jets

Can use *interpretable* ML:

- Gain insight about type of information responsible for differences: IRC-safe vs. IRC-unsafe, hard vs. soft
- Design maximally discriminating observables that are calculable in pQCD

$$\max_{\theta} \left| \frac{d\sigma_{eA}}{d\sigma_{ep}}(\theta) - 1 \right|$$





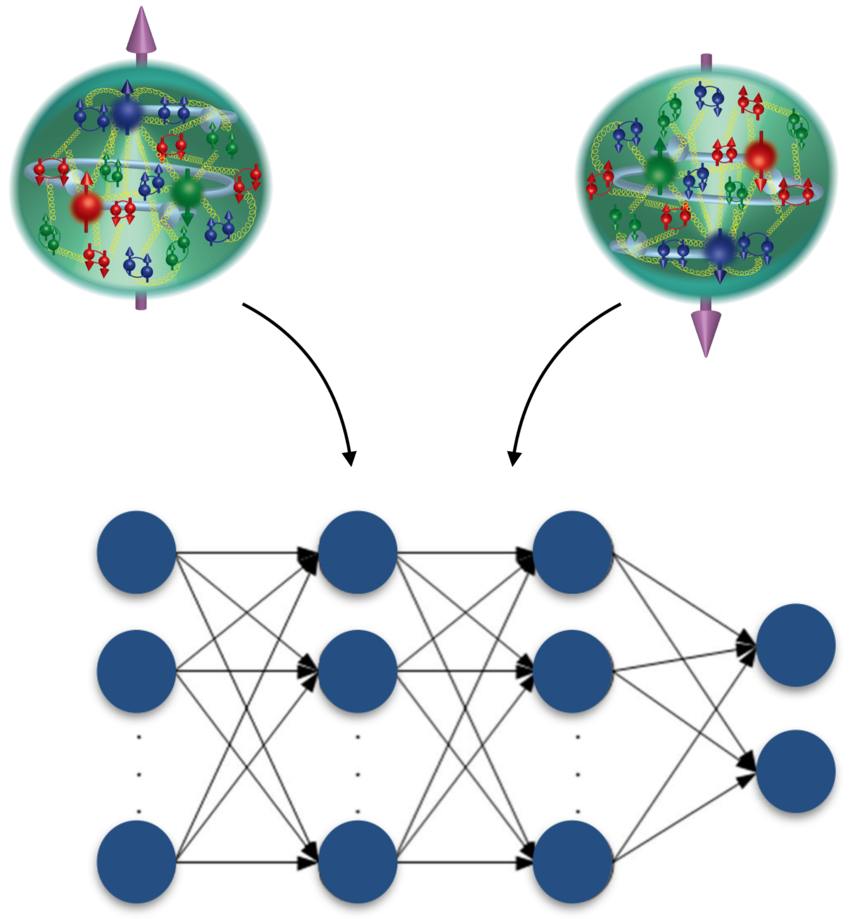
Can be applied directly on experimental data

Structure of QCD bound states

Spin physics

Lee, Mulligan, Płoskoń, Ringer, Yuan arXiv 2210.06450

Spin asymmetries encode information about internal structure of proton:



$$A = \frac{d\sigma^{\uparrow} - d\sigma^{\downarrow}}{d\sigma^{\uparrow} + d\sigma^{\downarrow}}$$

ML-assisted design of observables that maximize asymmetry

$$\max_{\theta} \left| A(\theta) \right|$$

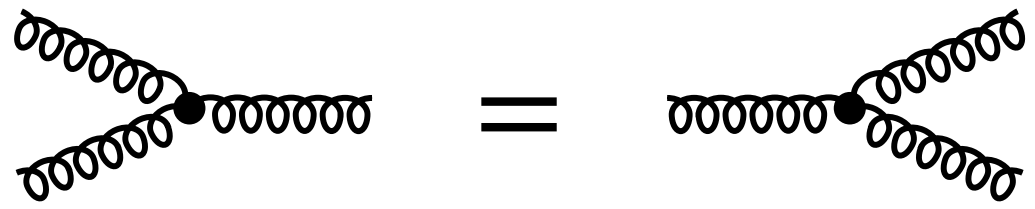
Train directly on experimental data: $\sigma^{\uparrow}$ vs. $\sigma^{\downarrow}$

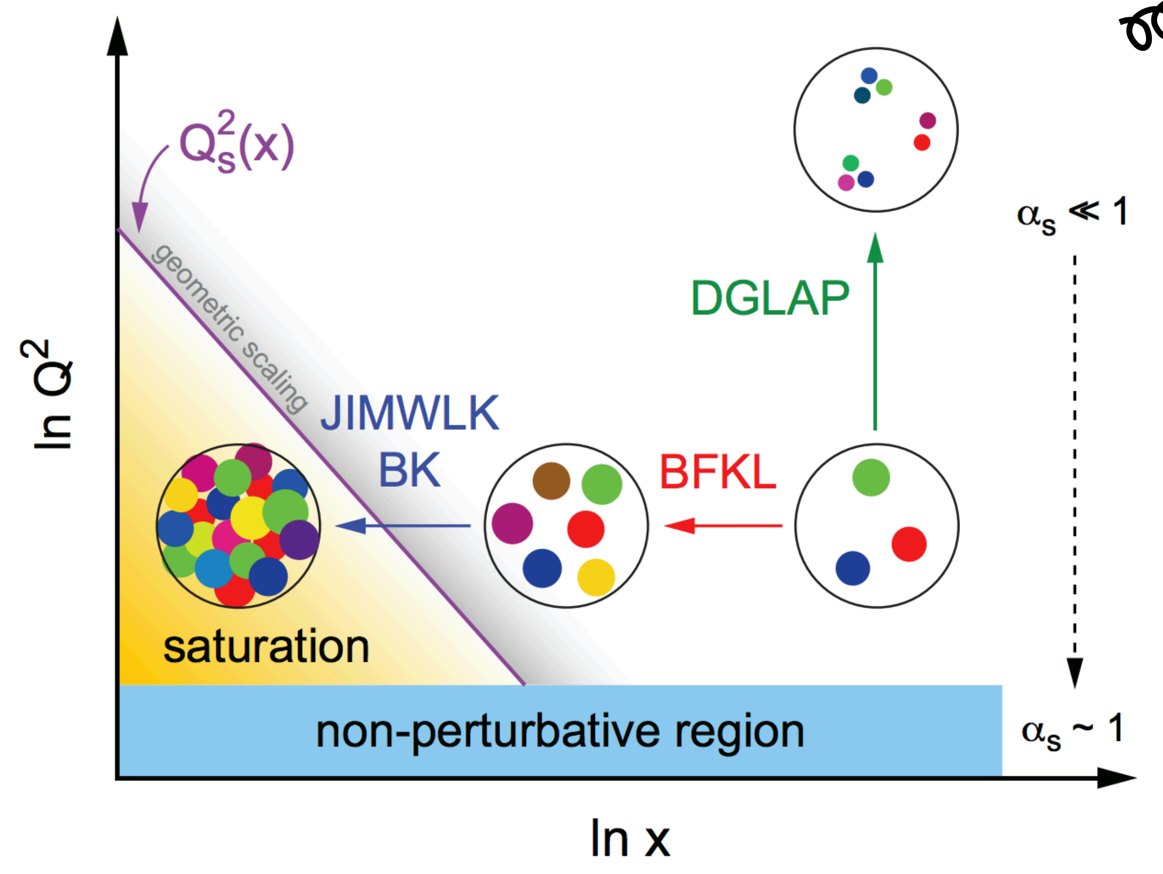
Can be applied at RHIC, EIC

Gluon saturation

EIC Yellow report, ALICE FoCal Proposal

Can we observe saturation of the gluon density in the nucleus?





Bayesian inference to test universal, non-linear scaling

Design sets of observables to test models

Can be applied at LHC, EIC

Thank you!

Encoding information in terms of observables

N-subjetiness basis and Energy Flow Polynomials

N-subjetiness

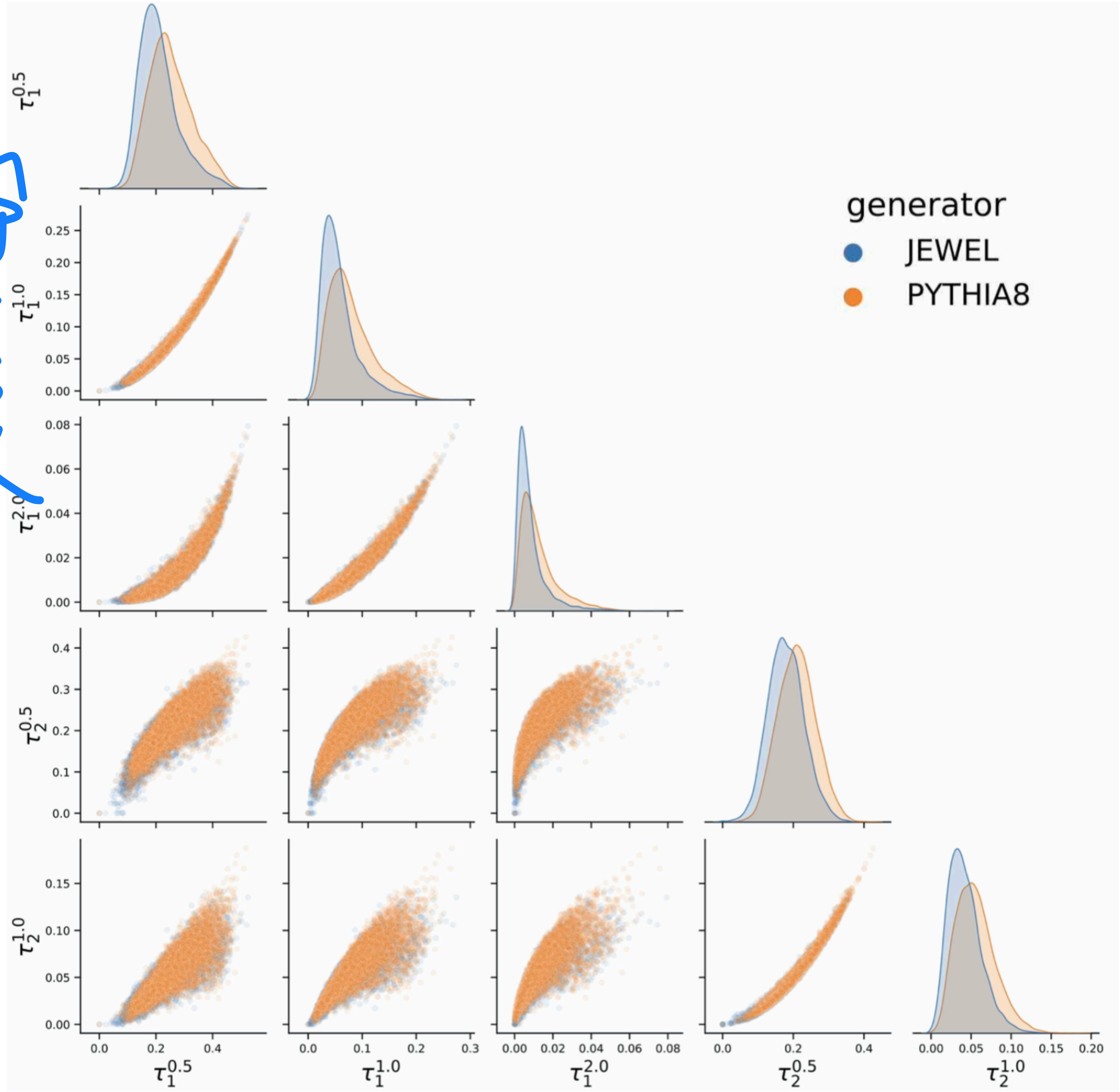


Figure 3. Scatter plot showing different N-subjettiness distributions (diagonal) and their pairwise correlations (off-diagonal panels) in pp and AA collisions without background. The pp and AA results shown here are obtained from Pythia 8 [70] and Jewel [72, 73], respectively.

EFP's

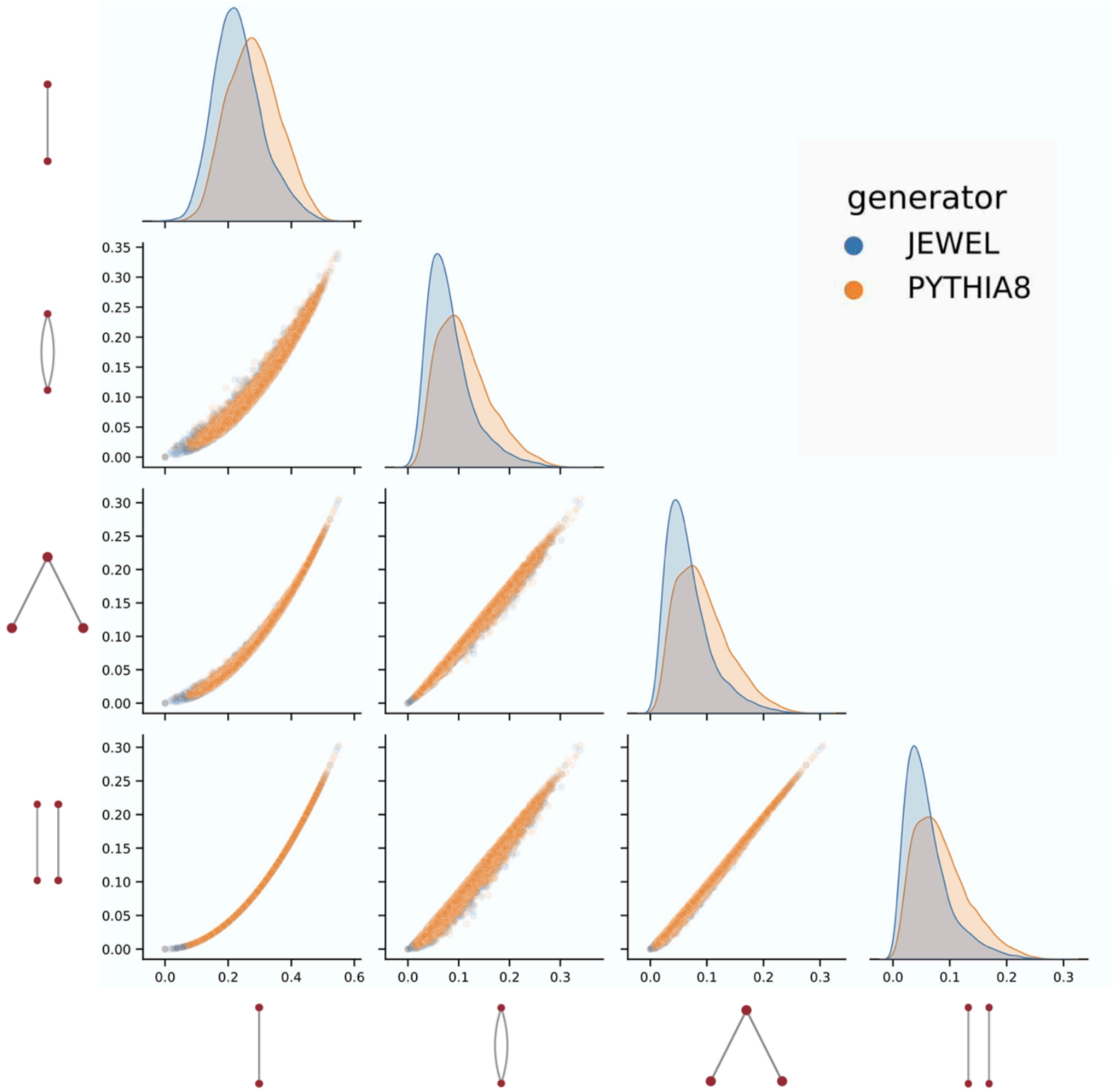
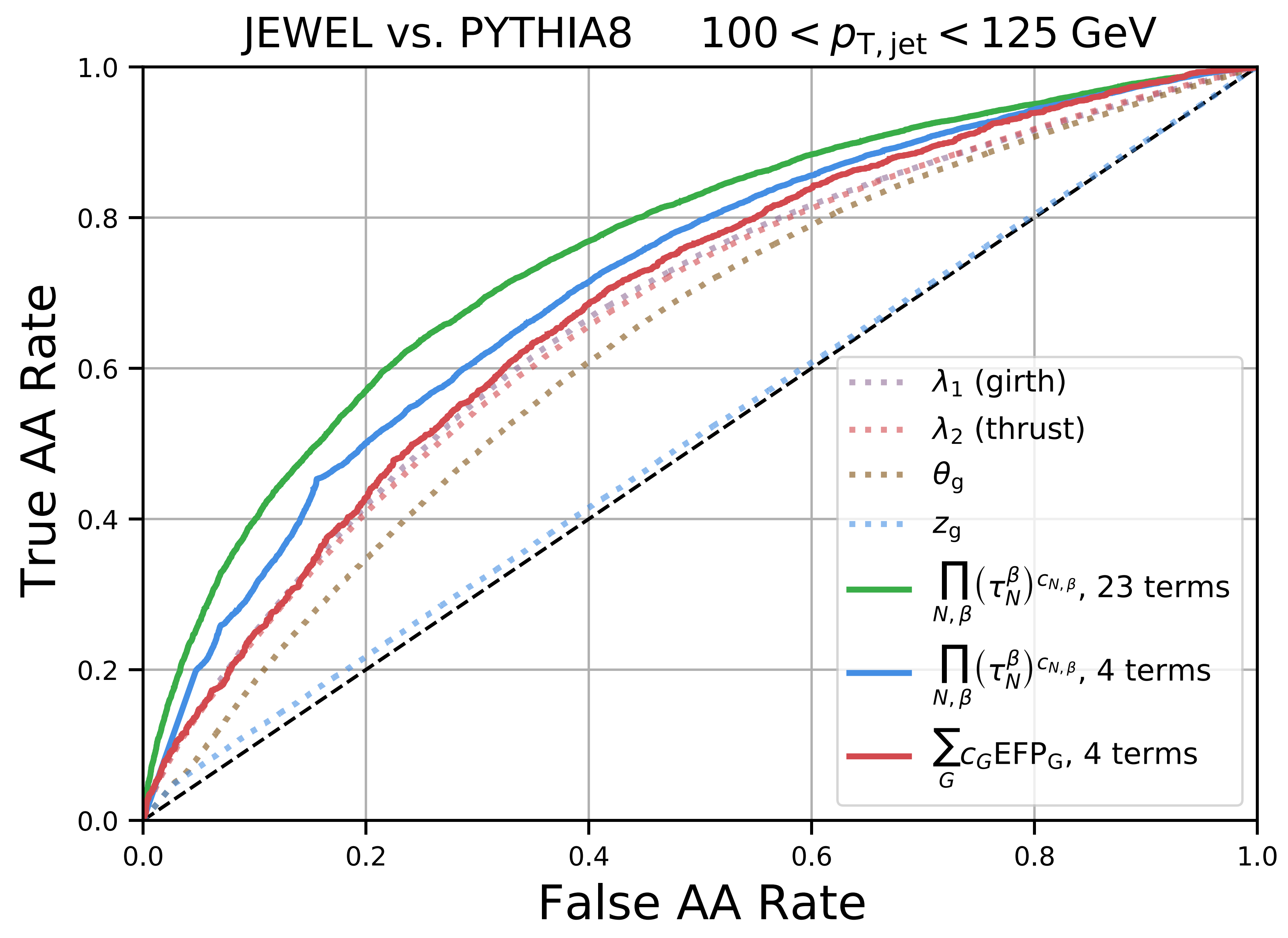


Figure 5. Scatter plot showing different EFP distributions (diagonal) and their pairwise correlations (off-diagonal panels) in pp and AA collisions without background. The pp and AA results shown here are obtained from Pythia 8 [70] and Jewel [72, 73], respectively.

Observable design

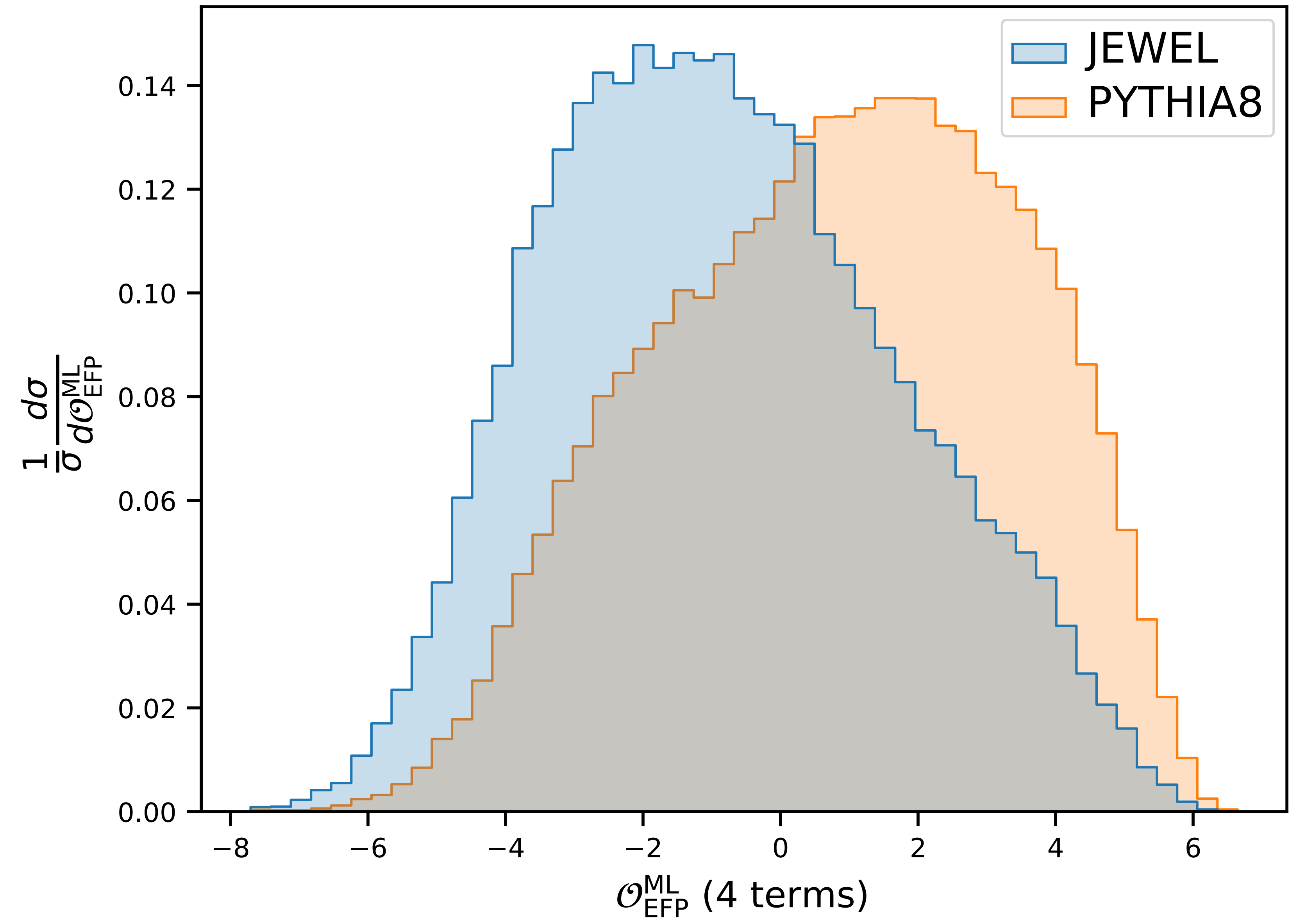
Lai, Mulligan, Płoskoń, Ringer JHEP 10 (2022) 011

By balancing the tradeoff of discriminating power and complexity, we can design the *most strongly modified* calculable observable



Approximate classifier with small number of features

"Symbolic regression" using Lasso



ML-assisted observable design provides guidance to experiments and theory — can then measure and calculate designed observables using traditional methods