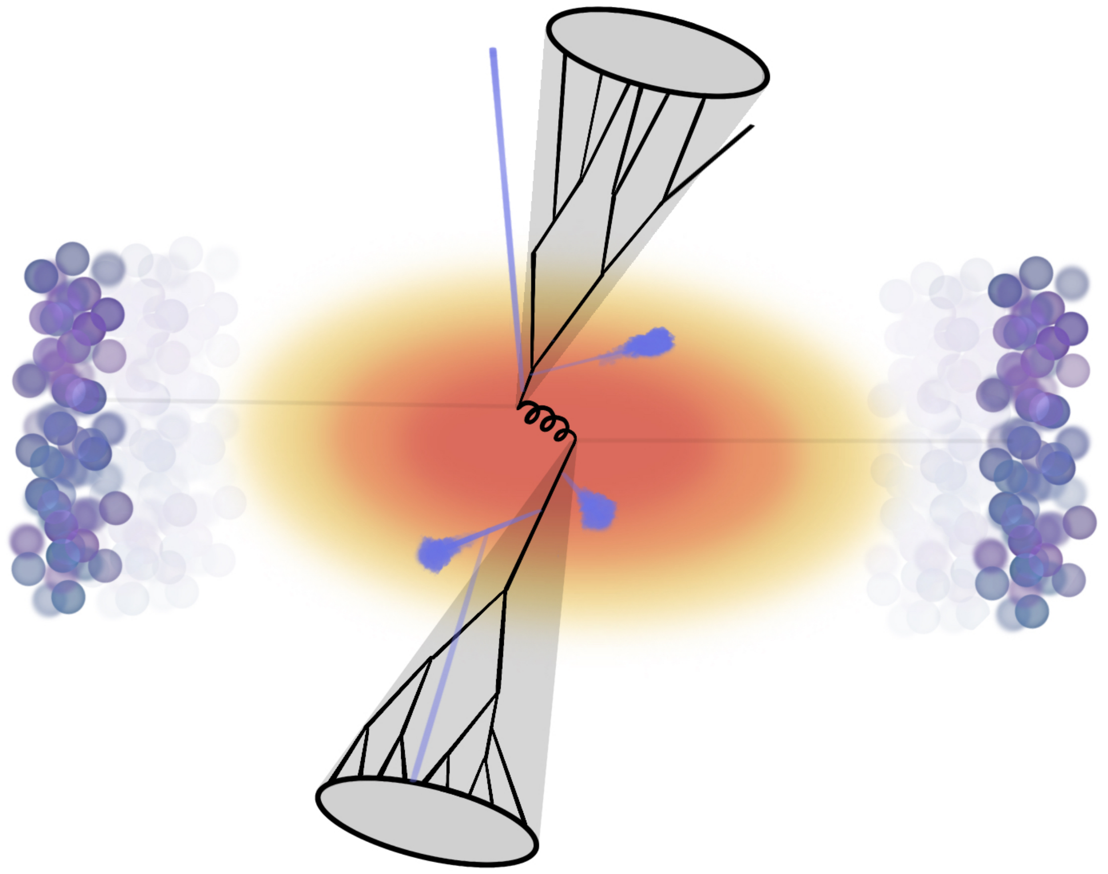
# The information content of jet quenching and machine learning assisted observable design

James Mulligan
in collaboration with
Yue Shi Lai, Mateusz Płoskoń, Felix Ringer

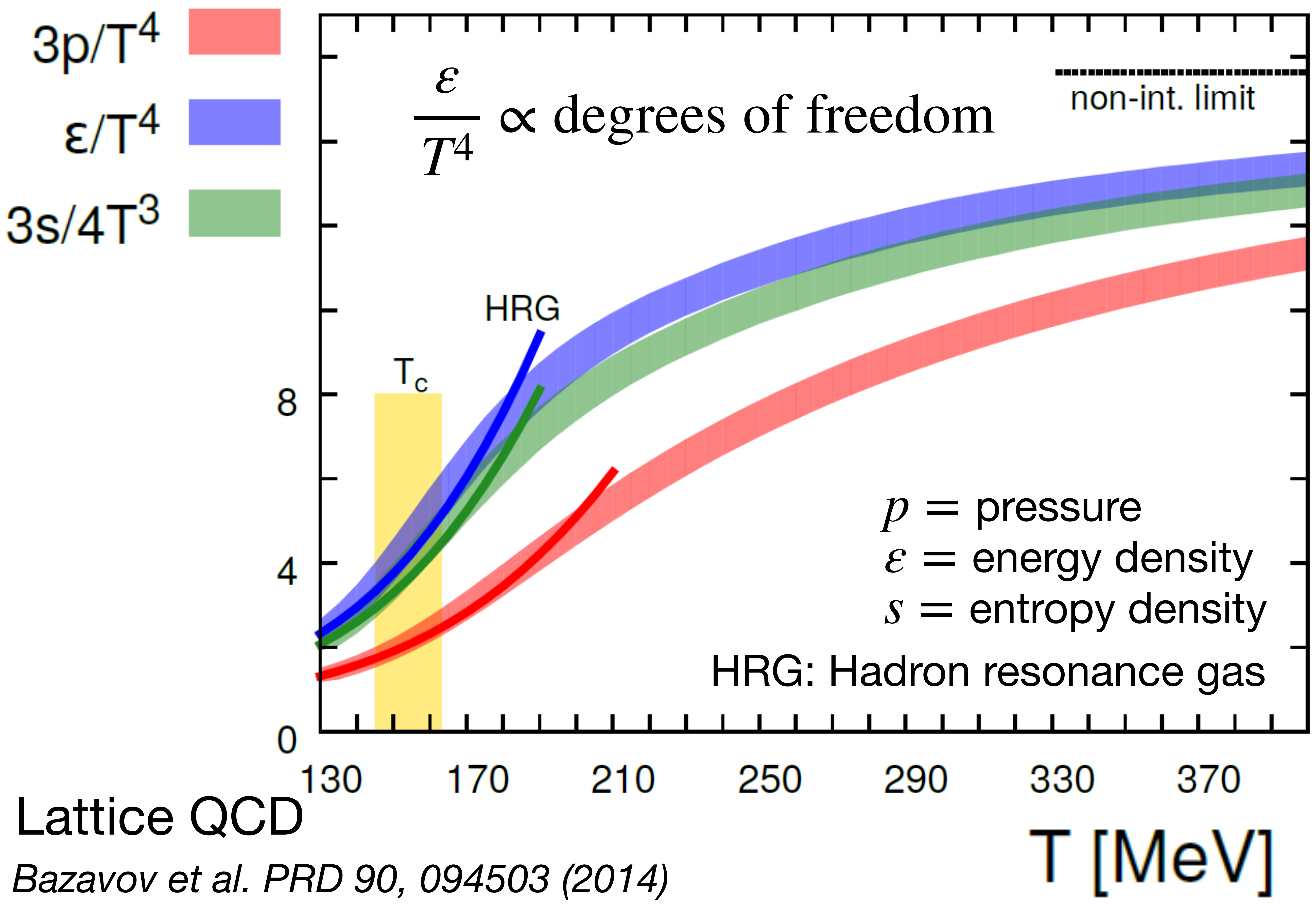


INT Workshop
Machine Learning for Nuclear Theory
March 31, 2022

Berkeley
UNIVERSITY OF CALIFORNIA

# The quark-gluon plasma

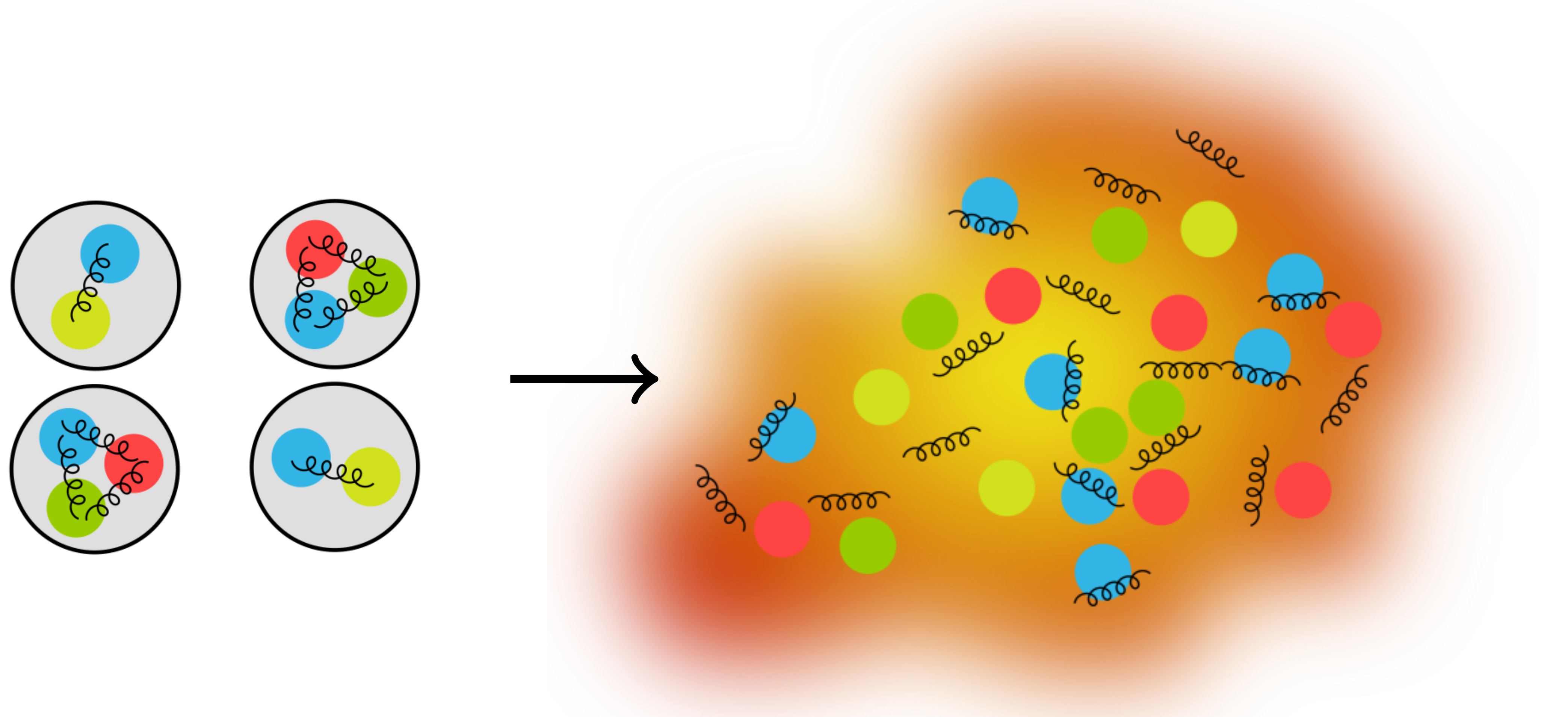
**If we heat nuclear matter to $T = \mathcal{O}(100\ \mathrm{MeV})$, thermodynamic quantities exhibit a rapid rise near a crossover temperature $T_c$**




Lattice QCD

*Bazavov et al. PRD 90, 094503 (2014)*

***Deconfinement of quarks and gluons***



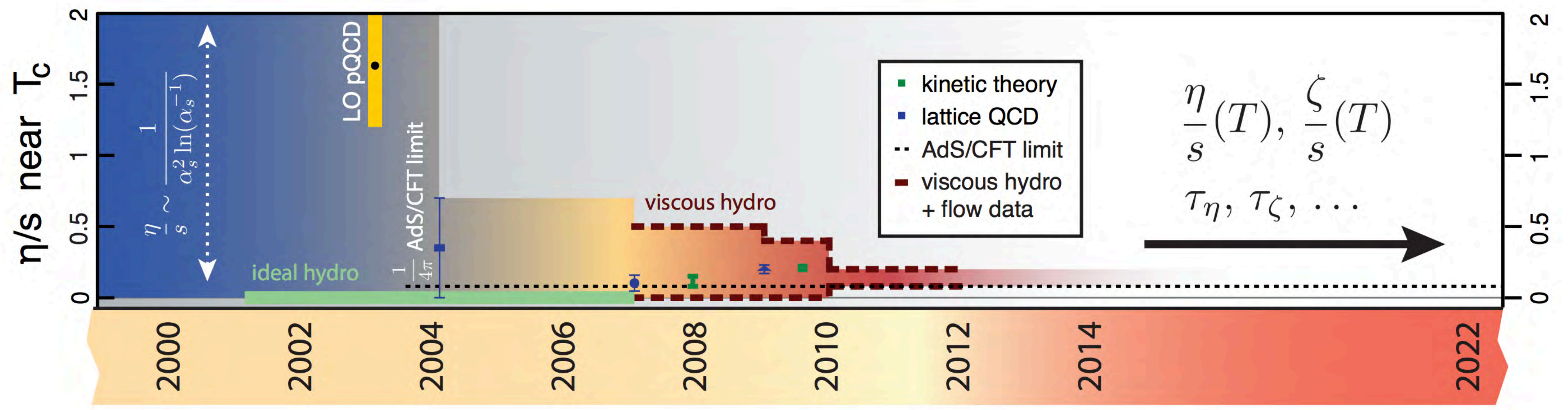
# The quark-gluon plasma

In the last two decades it has been established that hot QCD matter is:

- Deconfined
- Strongly-coupled

But much more to learn!

Low specific shear viscosity ($\eta/s$) → Strong coupling



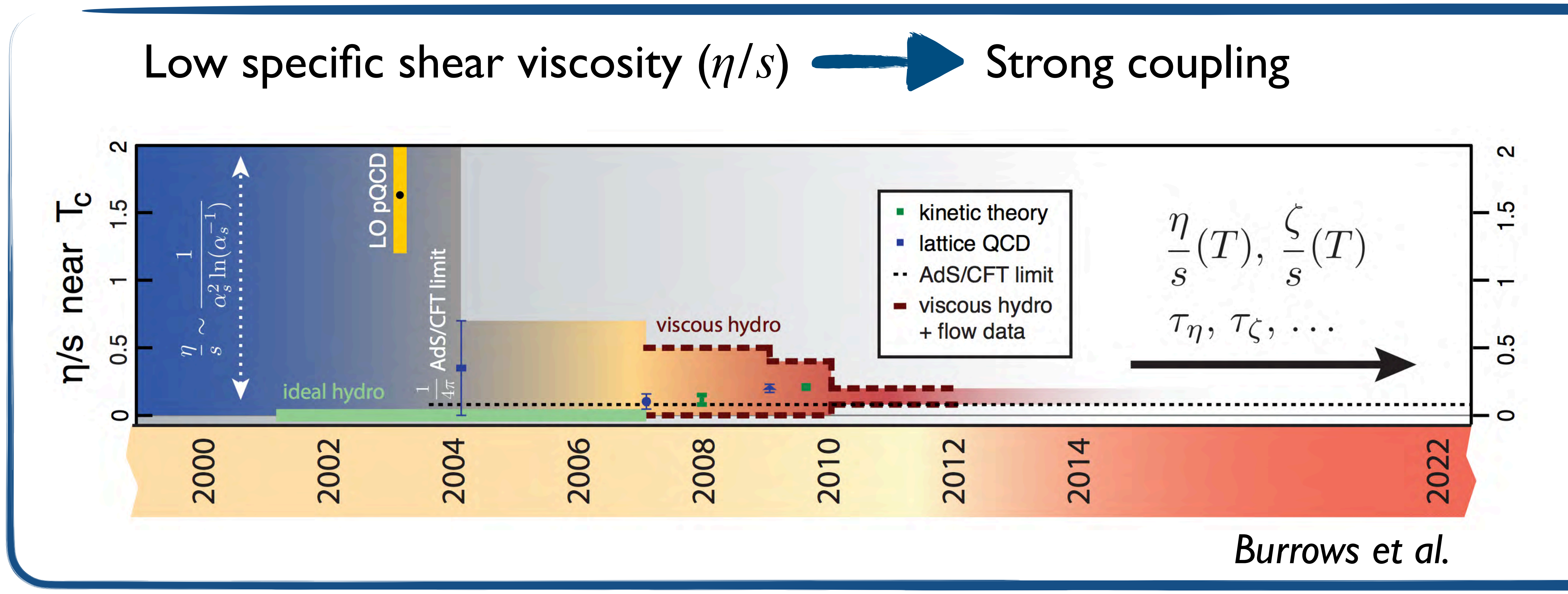

*Burrows et al.*

# The quark-gluon plasma

In the last two decades it has been established that hot QCD matter is:

- Deconfined
- Strongly-coupled

But much more to learn!




The quark-gluon plasma is a laboratory to study how complex properties emerge from the fundamental laws of quantum chromodynamics

*How does a strongly-coupled fluid arise from the Lagrangian of QCD?*

*What are the relevant degrees of freedom of the QGP as a function of resolution scale?*

*How does color confinement emerge?*

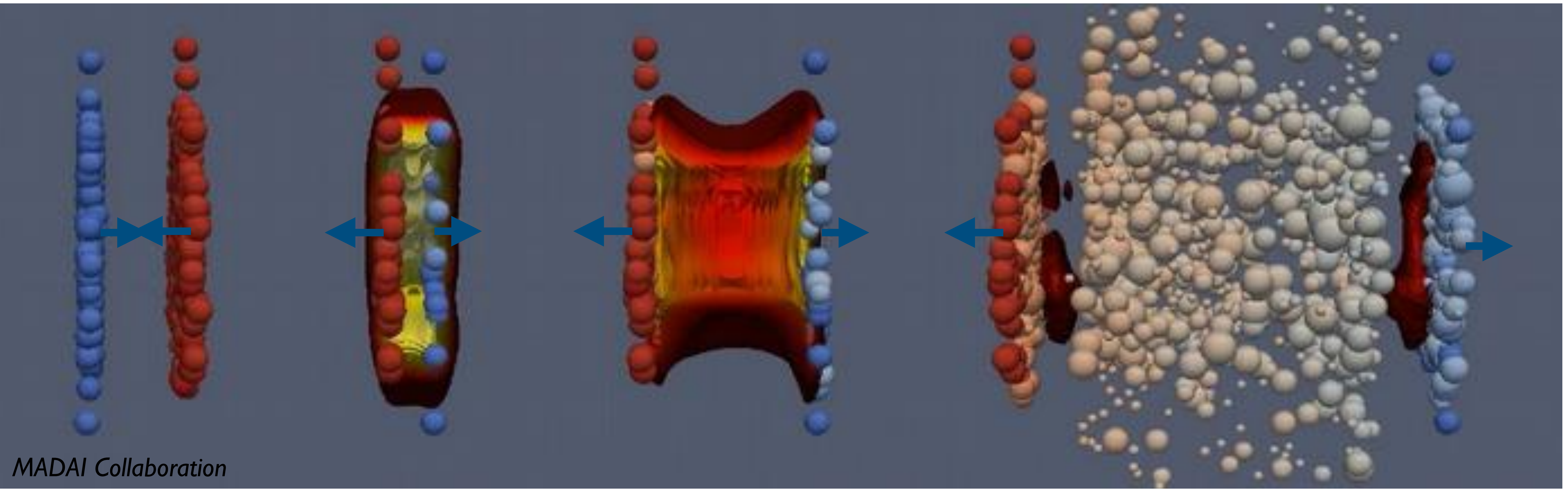
# Heavy-ion collisions


**We collide nuclei together at the**

*Large Hadron Collider (LHC)*

*Relativistic Heavy Ion Collider (RHIC)*

**to produce droplets of hot, dense quark-gluon plasma**

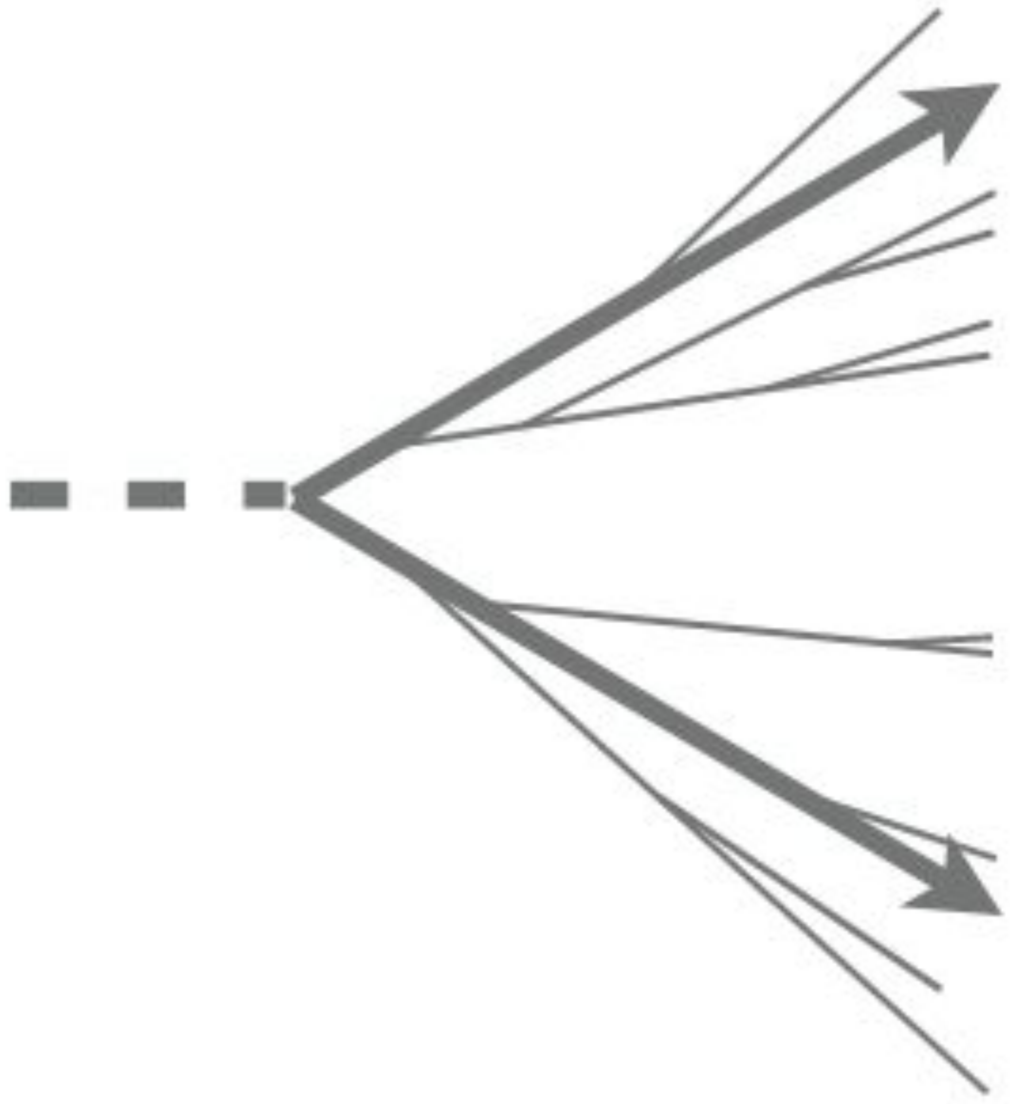



**Soft collisions transform kinetic energy of nuclei into region of large energy density**

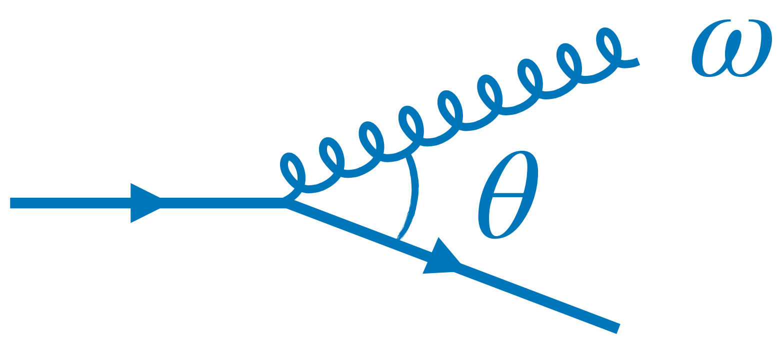
$T \approx 150\text{-}500$ MeV $\quad t \sim \mathcal{O}\left(10 \text{ fm}/c\right)$

# Jets

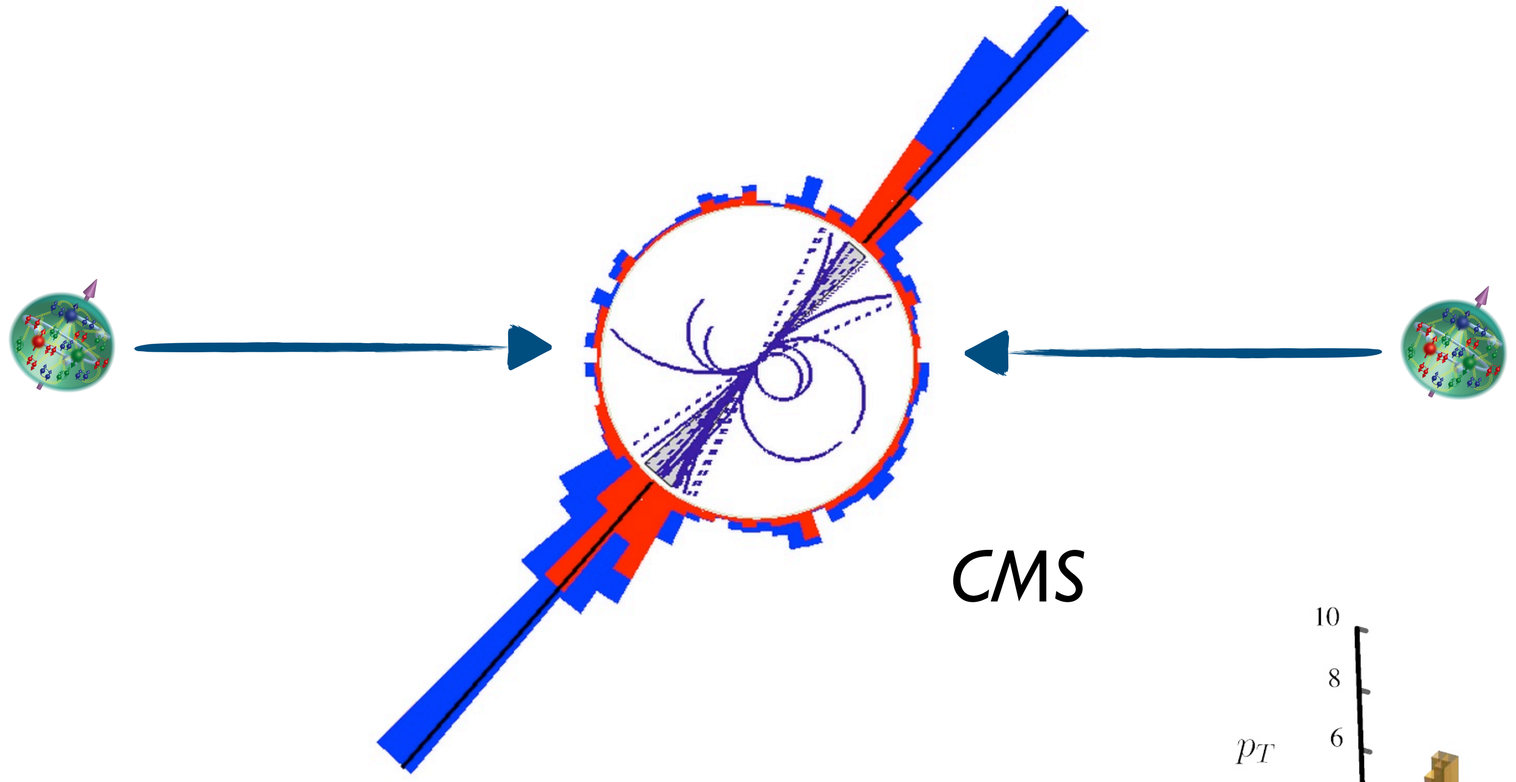
Collimated shower of particles arising from the iterative fragmentation of a **high energy quark or gluon**

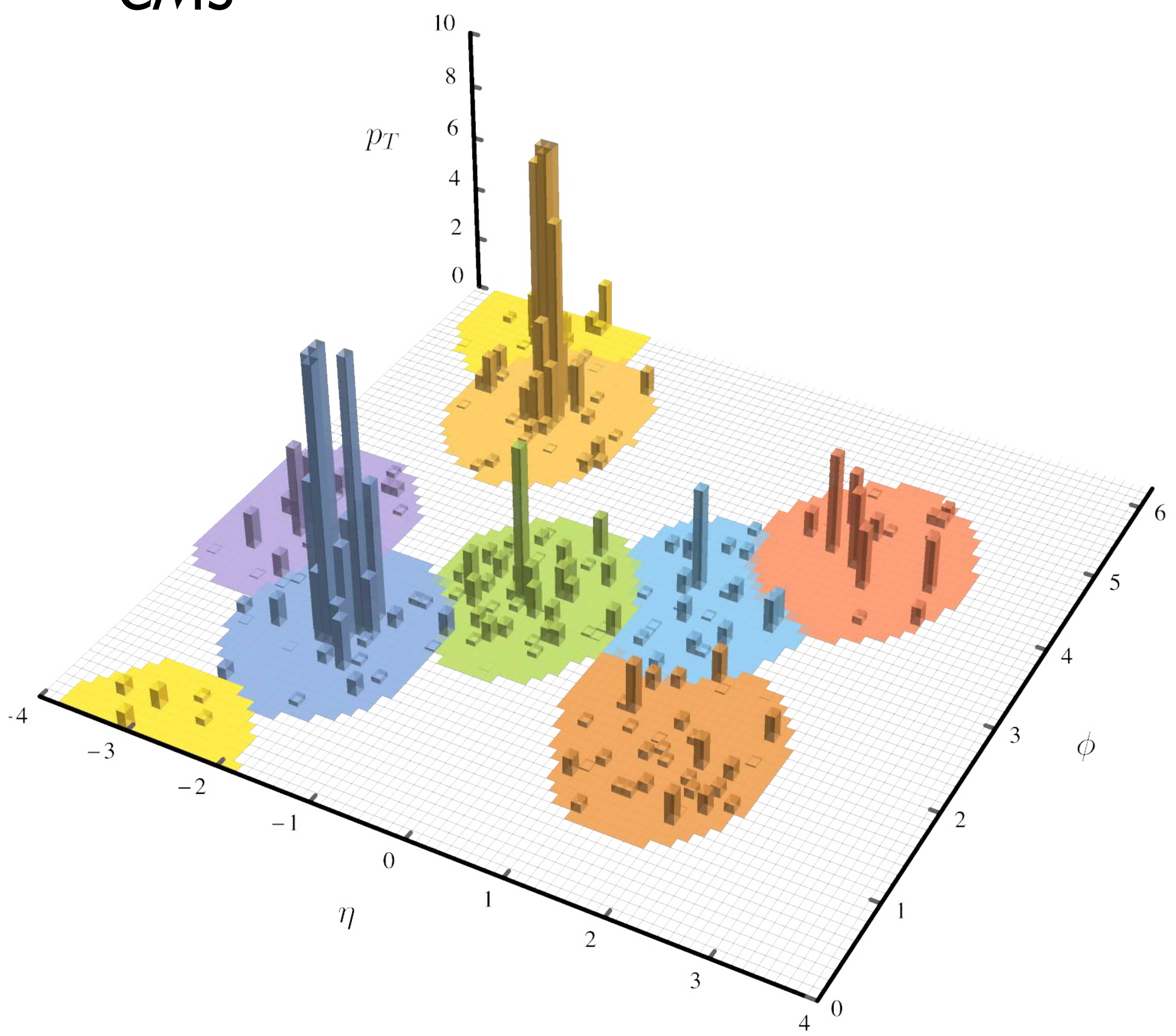


$$dP \sim \alpha_s C_R \frac{d\theta}{\theta}\frac{d\omega}{\omega}$$




Jets are produced abundantly in colliders

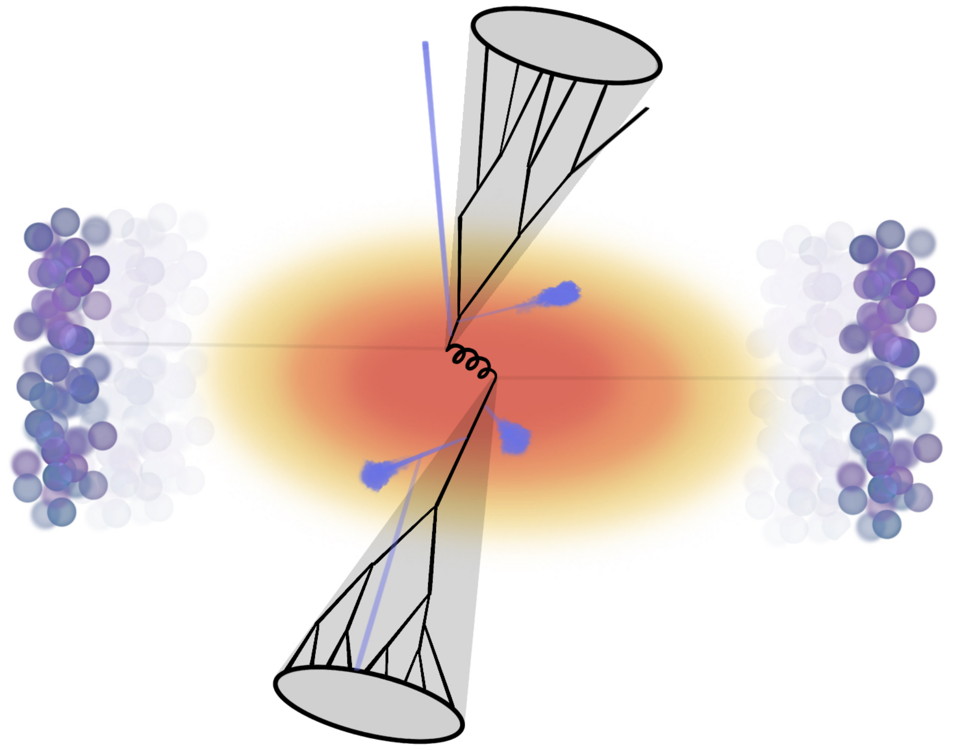


# Jet quenching in the quark-gluon plasma

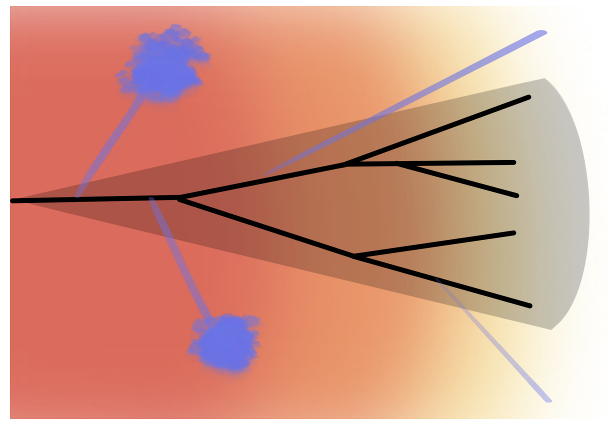
The QGP is too small and short-lived to be probed by traditional scattering beams

→ Use jets as probes

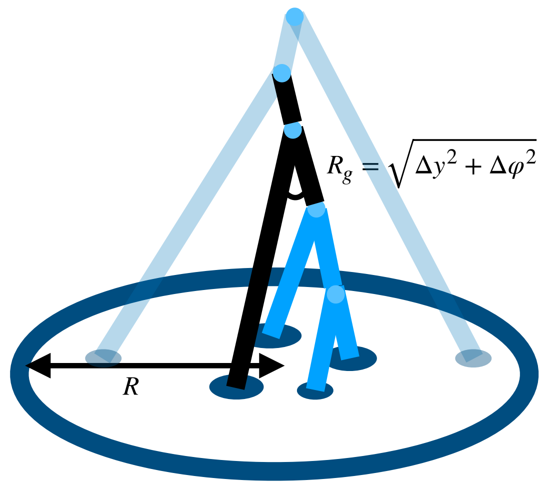


Jets interact with the quark-gluon plasma as they traverse it:

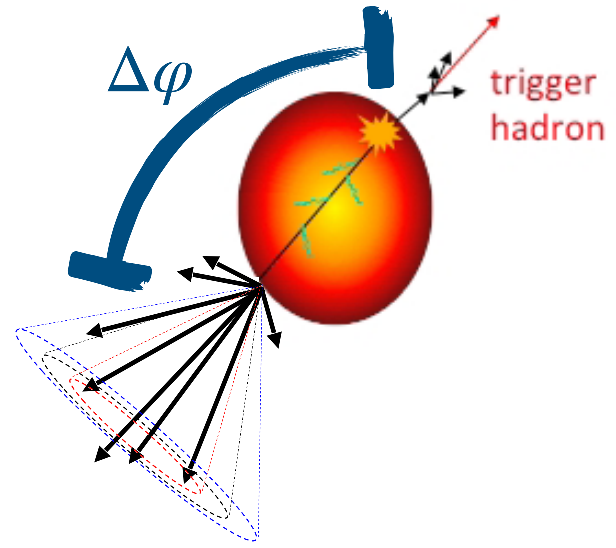
**"Energy loss"**



**Substructure modification**




**Deflection**

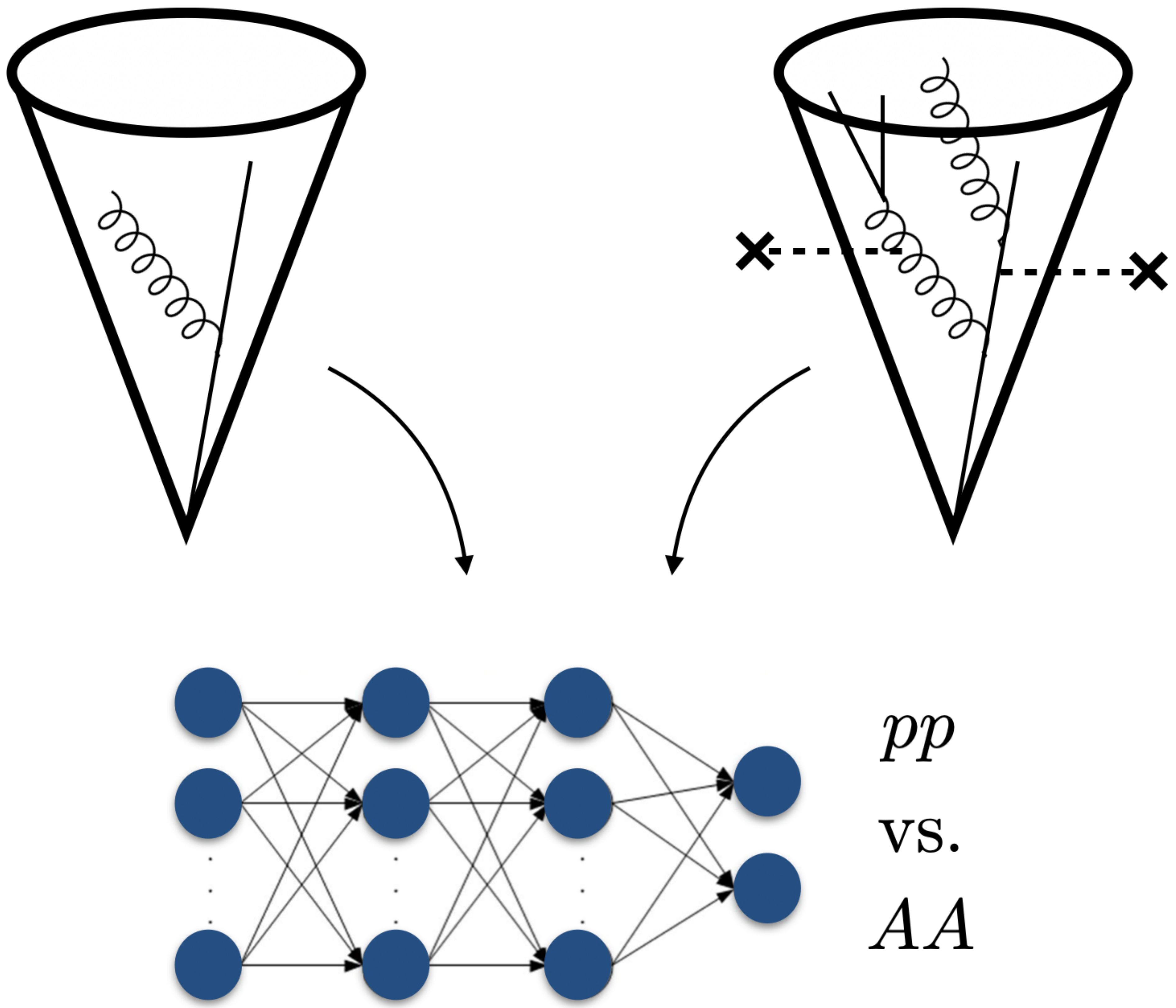



By modeling these interactions, we hope to determine the structure of the QGP

# Jet classification — pp vs. AA

We seek to understand how jets in heavy-ion collisions are different than jets in proton-proton collisions

→ Binary classification problem




**Goal:** Use ML to discriminate pp from AA jets in a way that is **theoretically interpretable**

# Outline

1. Identify the useful information content in the jet

2. Design optimal observables to maximize discrimination

3. Assess information loss due to backgrounds

# Outline

1. Identify the useful information content in the jet

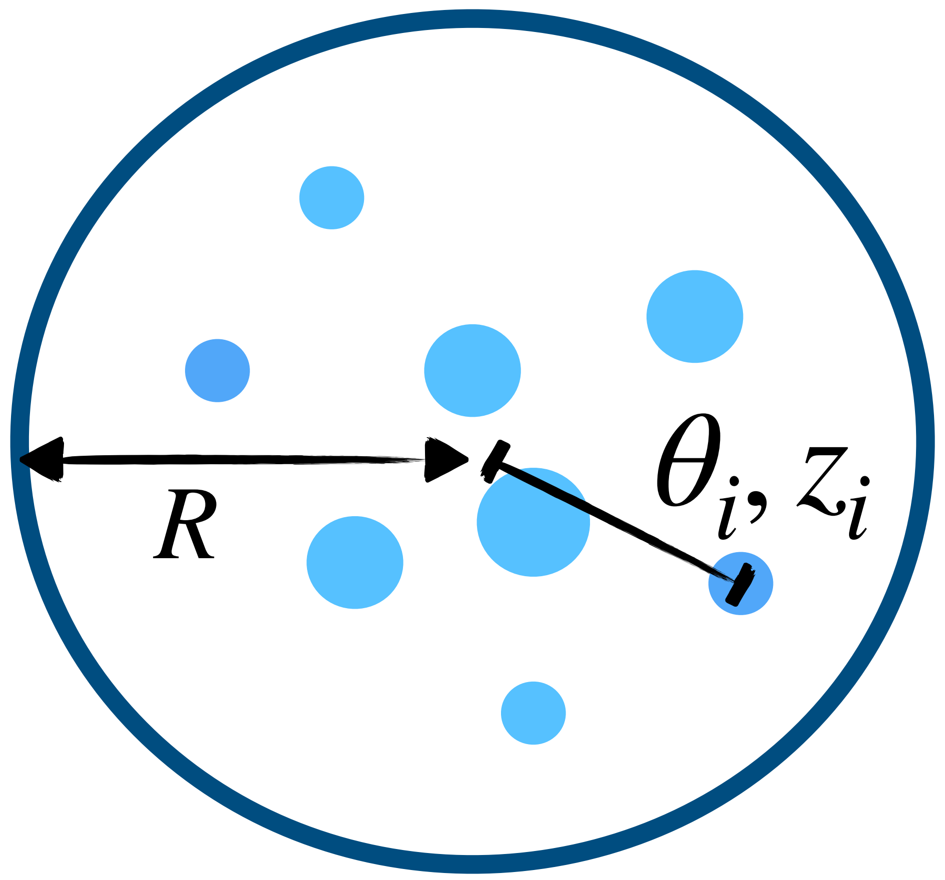
2. Design optimal observables to maximize discrimination

3. Assess information loss due to backgrounds

# Jet observables and IRC safety

We are free to construct any observable from the jet's constituents

e.g. $$\lambda_\alpha^\kappa = \sum_{i \in \text{jet}} z_i^\kappa \theta_i^\alpha$$

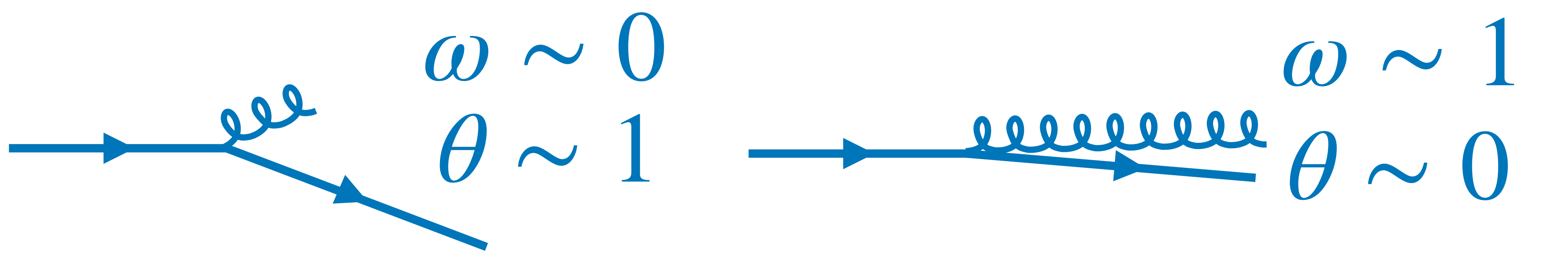



$$\theta_i = \frac{\sqrt{\Delta y^2 + \Delta \varphi^2}}{R}$$

$$z_i = \frac{p_{\mathrm{T},i}}{p_{\mathrm{T,jet}}}$$

However, usually only those combinations that obey **infrared-collinear (IRC) safety** are calculable in perturbative QCD

e.g. $$\lambda_{\alpha>0}^{\kappa=1} = \sum_{i \in \text{jet}} z_i \theta_i^\alpha$$




*Insensitive to soft/collinear emissions*

# IRC-safe vs. IRC-unsafe architectures

Permutation-invariant neural networks based on deep sets

Unordered, variable-length sets of particles as input

*Komiske, Metodiev, Thaler JHEP 01 (2019) 121*

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

**Particle Flow Network (PFN)**

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

Classifier

DNNs

latent space $d = 256$

Includes IRC-unsafe information

**Energy Flow Network (EFN)**

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

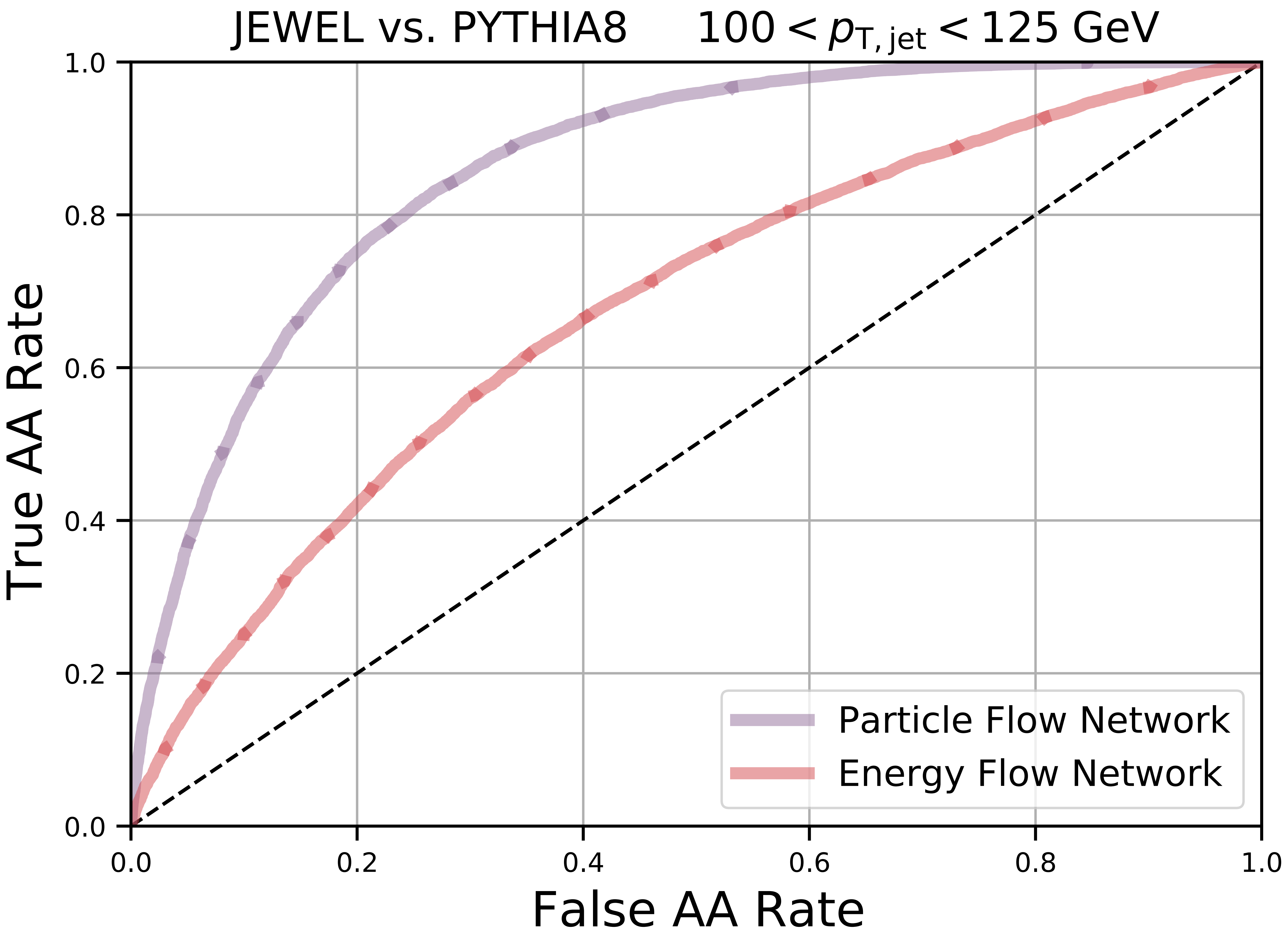
Classifier

DNNs

Includes only IRC-safe information

# IRC-safe vs. IRC-unsafe physics

*Lai, Mulligan, Płoskoń, Ringer arXiv 2111.14589*




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

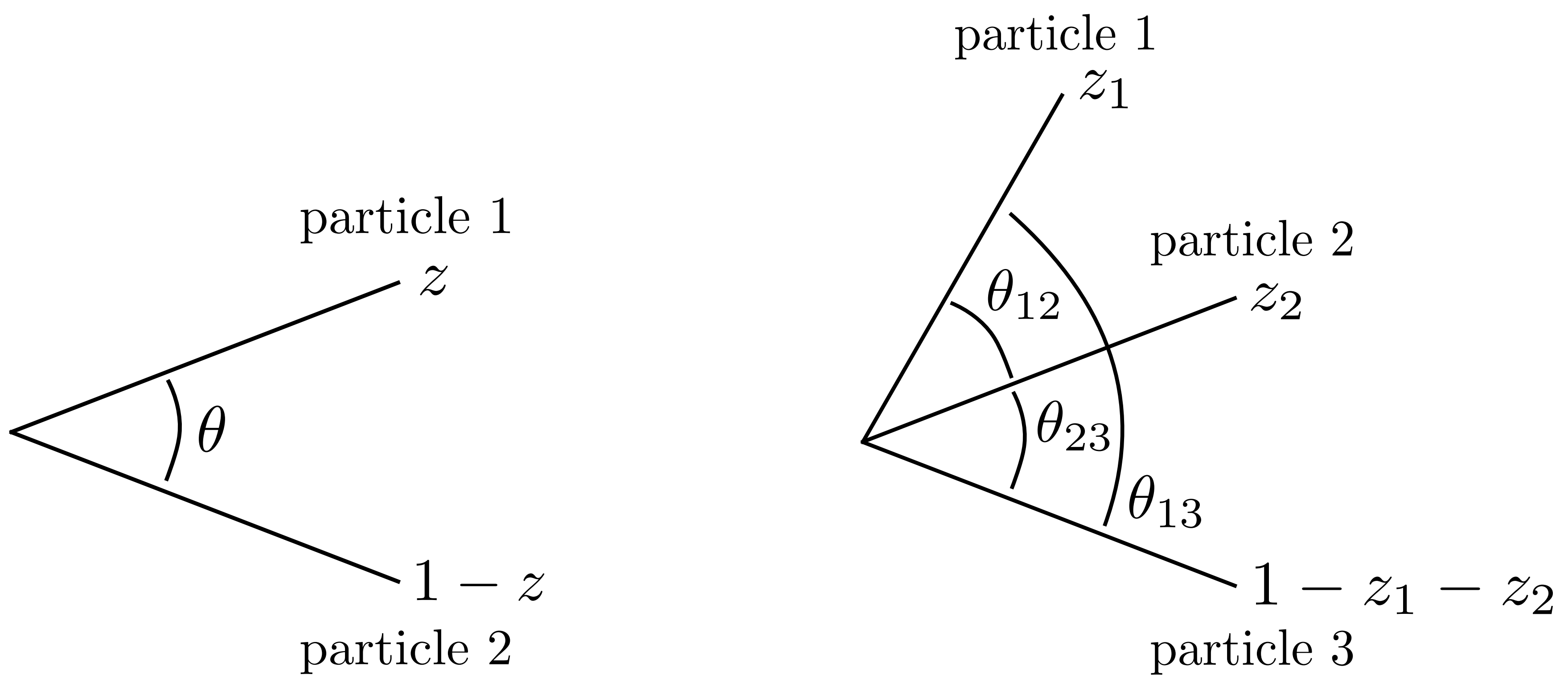
The PFN performs significantly better than the EFN

IRC-unsafe information contains significant discriminating power

# Jet classification in vacuum: Information content

*Datta, Larkoski JHEP 06 (2017) 073*

The substructure of a jet with $M$ final-state partons can be specified by $3M-4$ observables



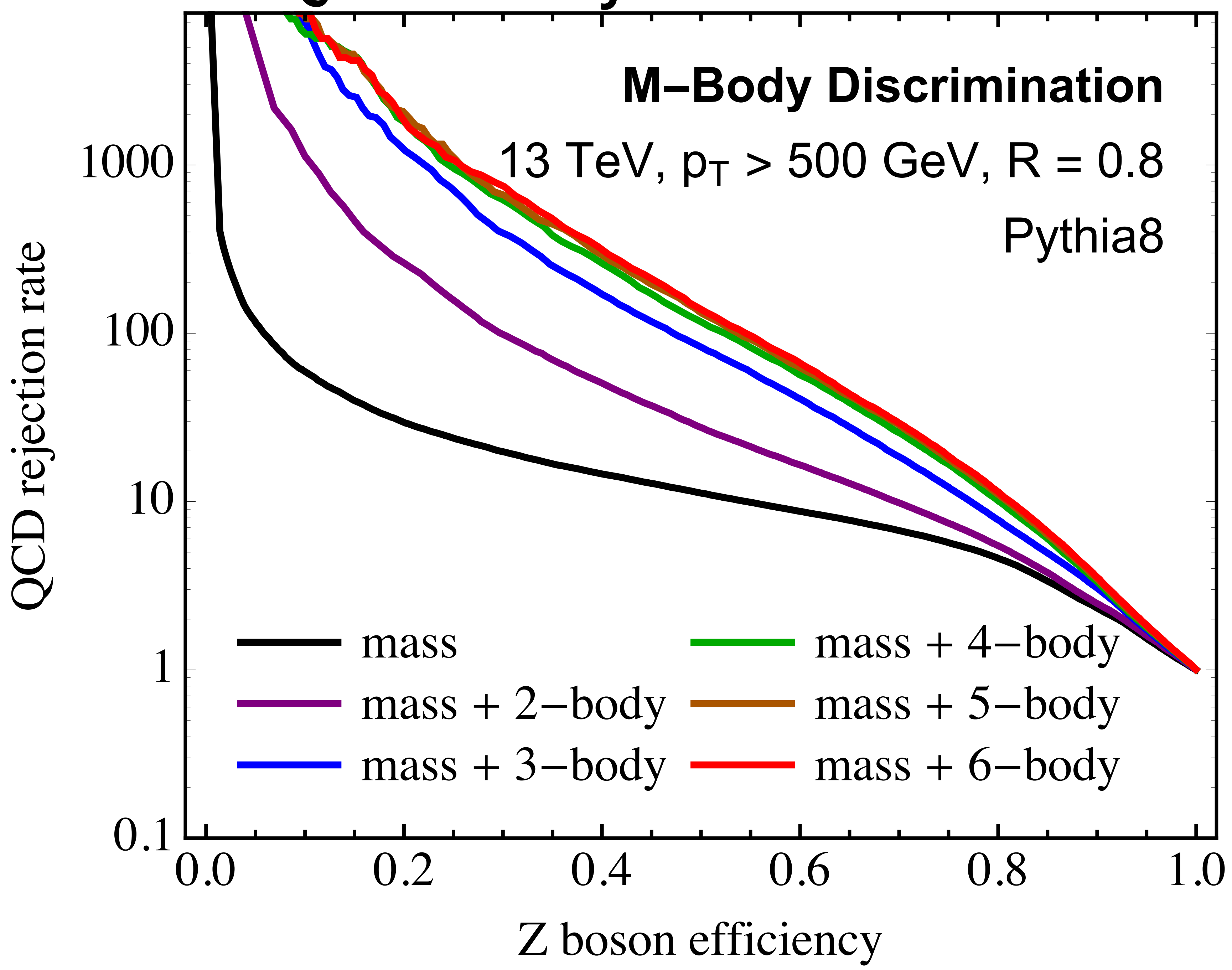

e.g. *N*-subjettiness basis:

$$\left\{\tau_1^{(0.5)},\tau_1^{(1)},\tau_1^{(2)},\tau_2^{(0.5)},\tau_2^{(1)},\tau_2^{(2)},\ldots,\tau_{M-2}^{(0.5)},\tau_{M-2}^{(1)},\tau_{M-2}^{(2)},\tau_{M-1}^{(1)},\tau_{M-1}^{(2)}\right\}$$

where $\tau_N^{(\beta)}=\frac{1}{p_T^{\text{jet}}}\sum_{i\in\text{Jet}}p_{Ti}\min\left\{R_{1i}^{\beta},R_{2i}^{\beta},\ldots,R_{Ni}^{\beta}\right\}$

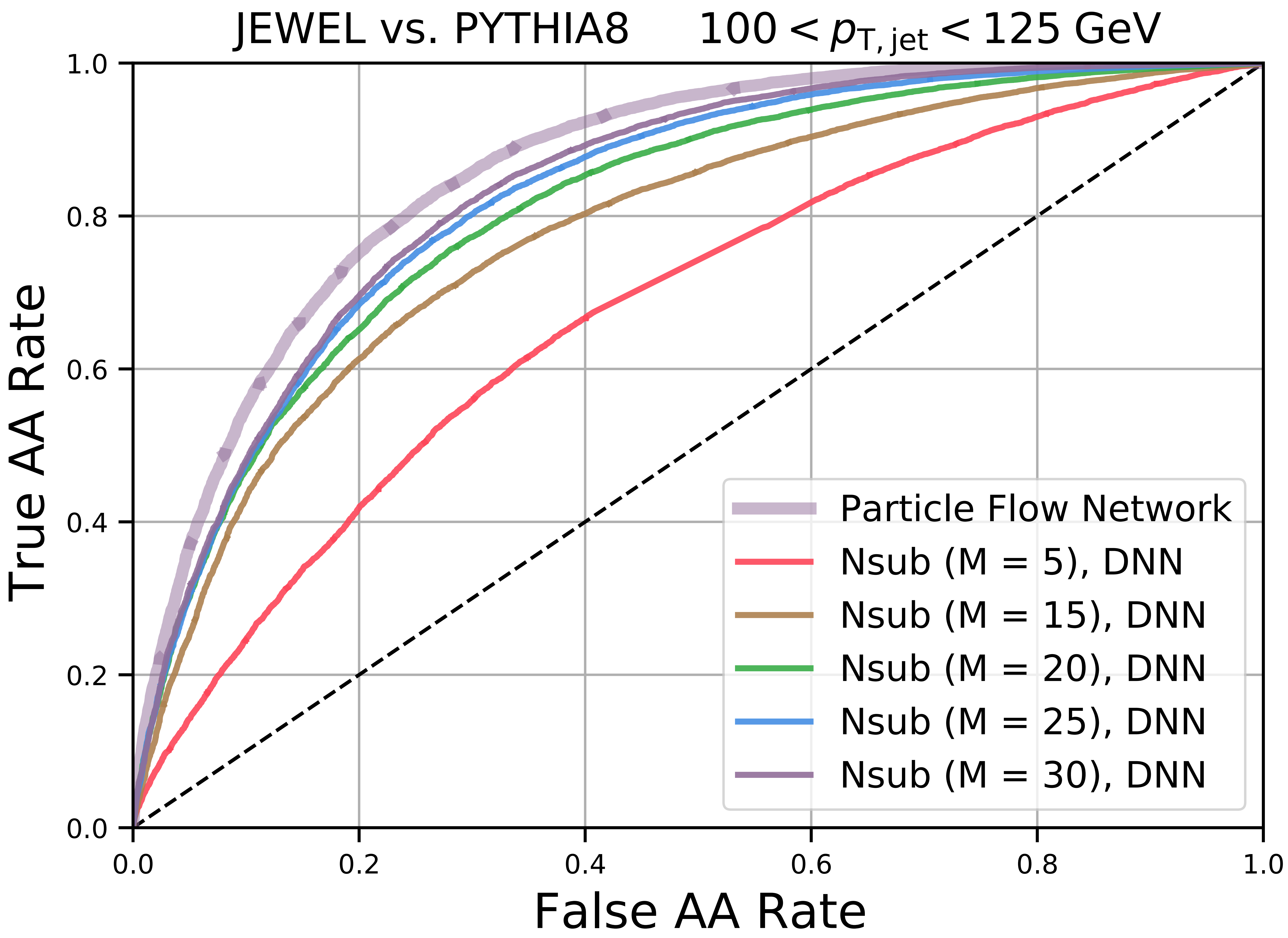
By constructing a complete set of observables encoding the jet's internal structure, can study at what point the information content saturates

*QCD vs. Z jet discrimination*



# Hard vs. soft physics

*Lai, Mulligan, Płoskoń, Ringer arXiv 2111.14589*



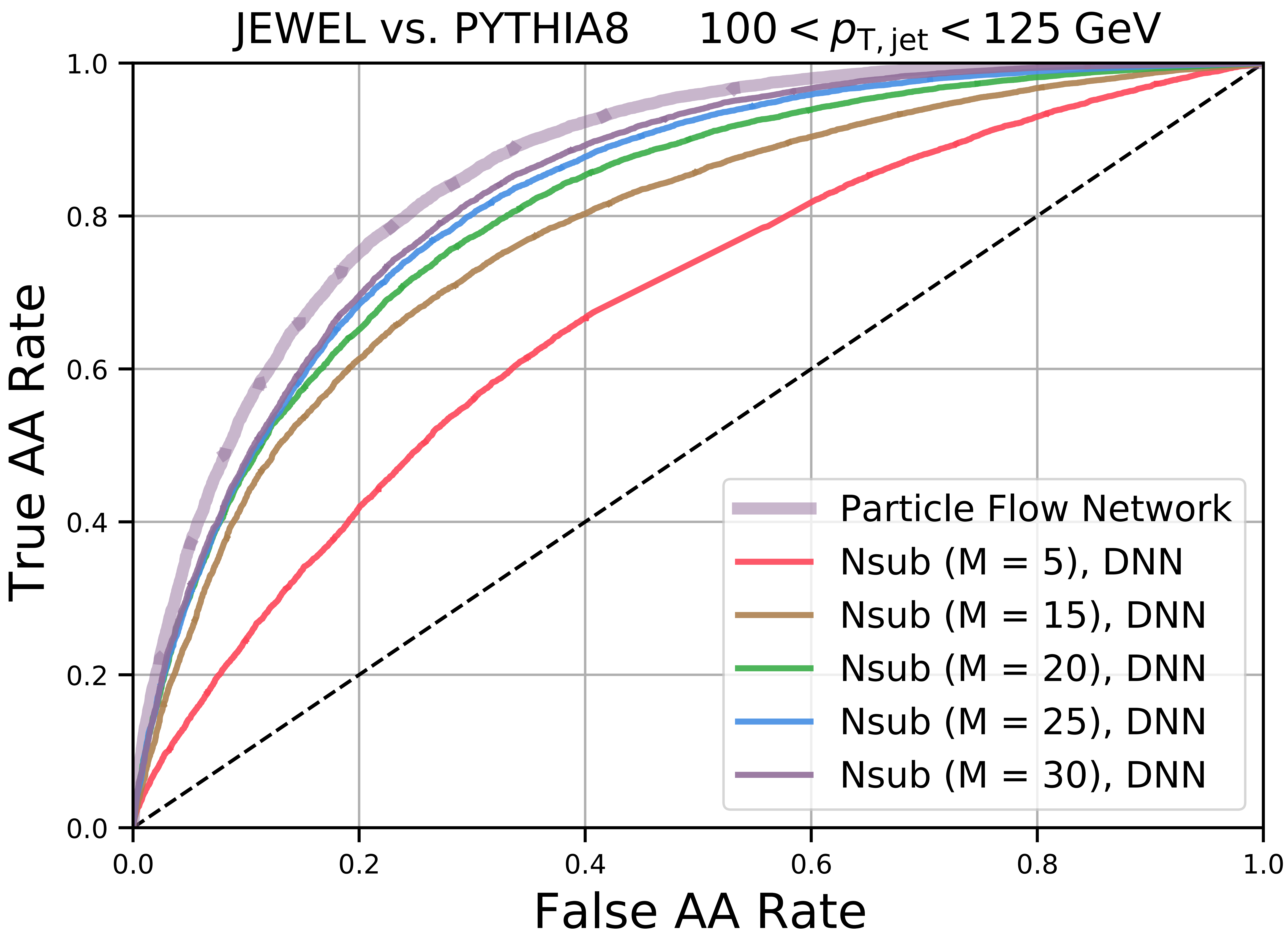

DNN with complete set of jet substructure observables as input

*N*-subjettiness basis:

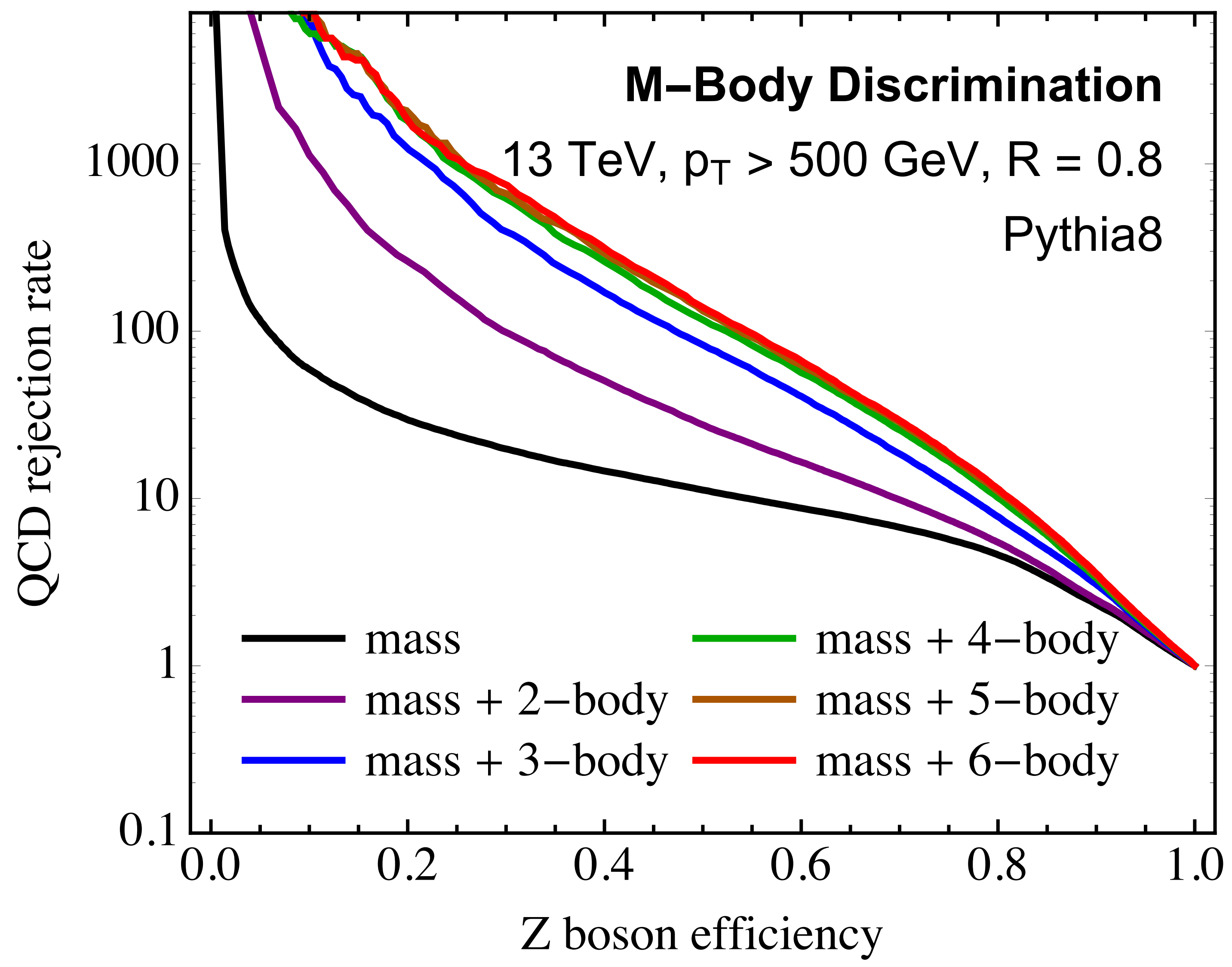
$$\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \ldots, \tau_{M-2}^{(0.5)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(0.5)}, \tau_{M-1}^{(1)}\right\}$$

Significant information in quenched jets up to $M \approx 25$

# Hard vs. soft physics

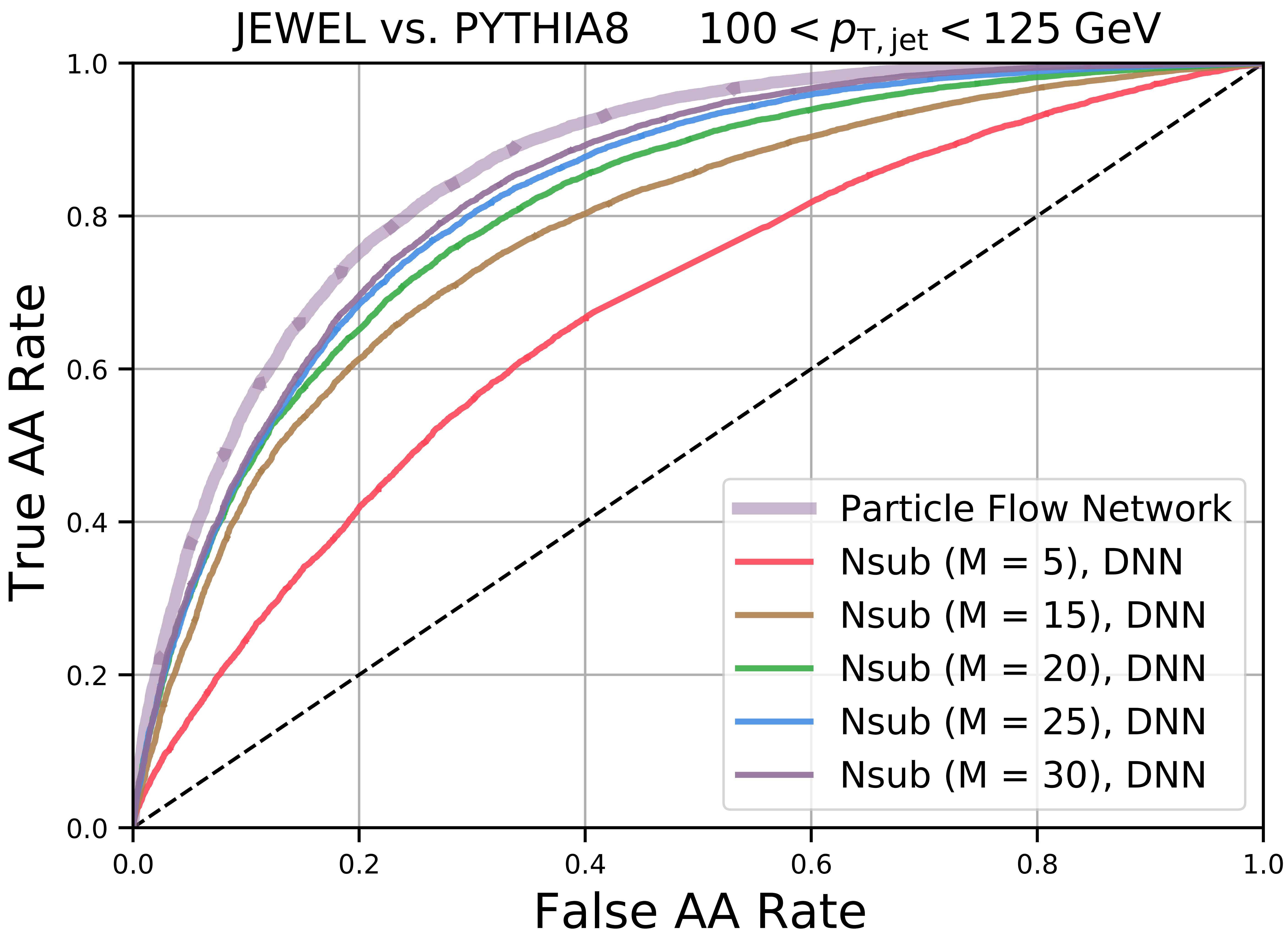



*Datta, Larkoski JHEP 06 (2017) 073*




→ Unlike QCD vs. Z jets (which saturate at $M = 4$), vacuum vs. quenched jets contain discriminating power in soft physics (high $M$-body phase space)

# Hard vs. soft physics




Deep set data representation (PFN) performs slightly better than substructure basis (DNN)

The difference can be due to:

- IRC-unsafe information in PFN
- Different data representations / training / hyperparameter performance

# Outline

1. Identify the useful information content in the jet

2. Design optimal observables to maximize discrimination

3. Assess information loss due to backgrounds

# ML-assisted observable design

Now that we have demonstrated an ML classifier, we can find observable(s) that can approximate the classifier

→ Theoretical calculability

Approximate the $3M-4$ N-subjettiness observables with e.g. product observables

Product observable: Sudakov safe

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

N-subjettiness exponents become weights in linear regression

$$\ln O = \sum_{N<K,\ \beta\in\{0.5,1,2\}} c_{N\beta} \ln \tau_N^\beta$$

# ML-assisted observable design

*Lai, Mulligan, Płoskoń, Ringer arXiv 2111.14589*

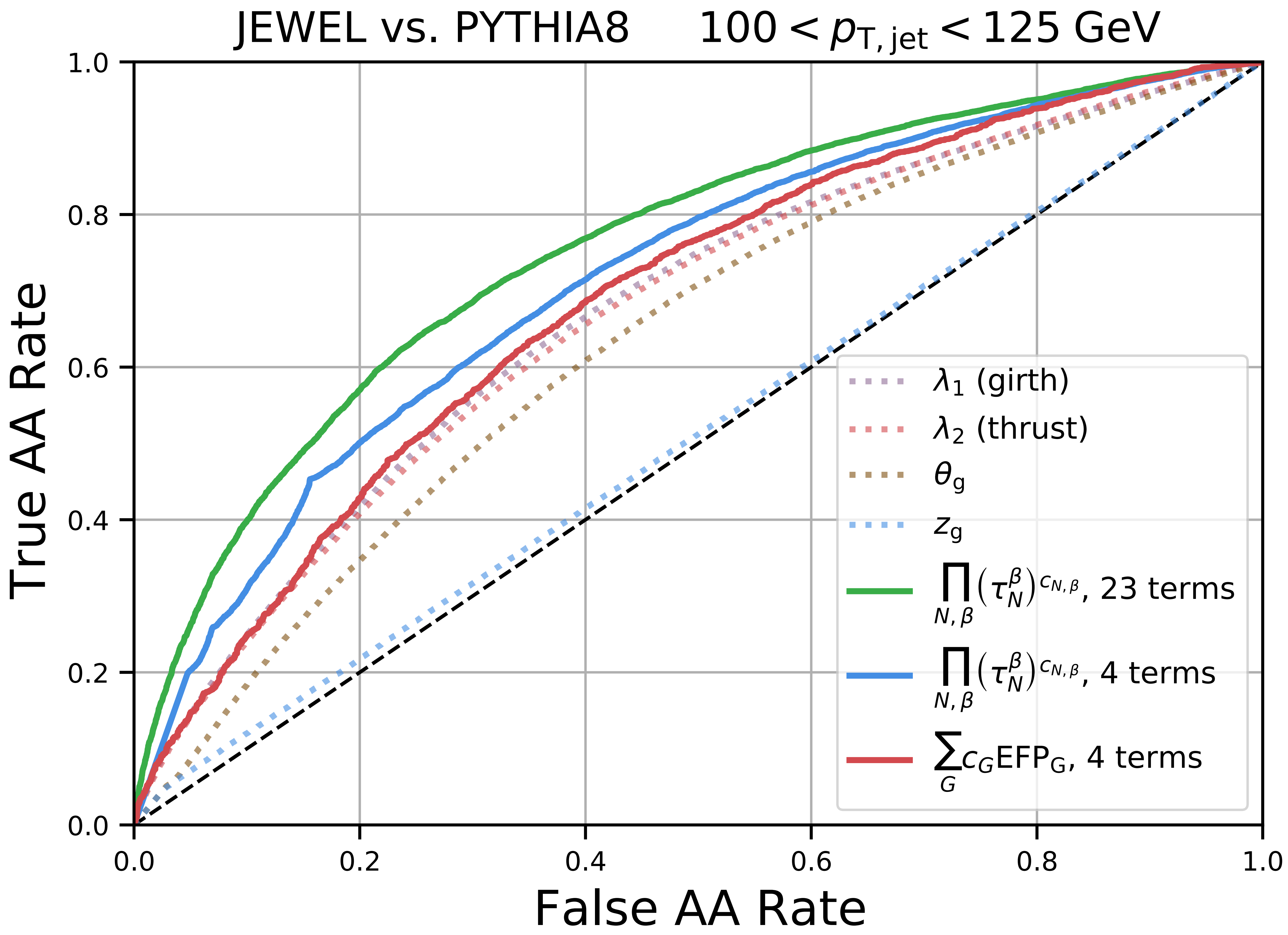
**Lasso regression**

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

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

# ML-assisted observable design

*Lai, Mulligan, Płoskoń, Ringer arXiv 2111.14589*

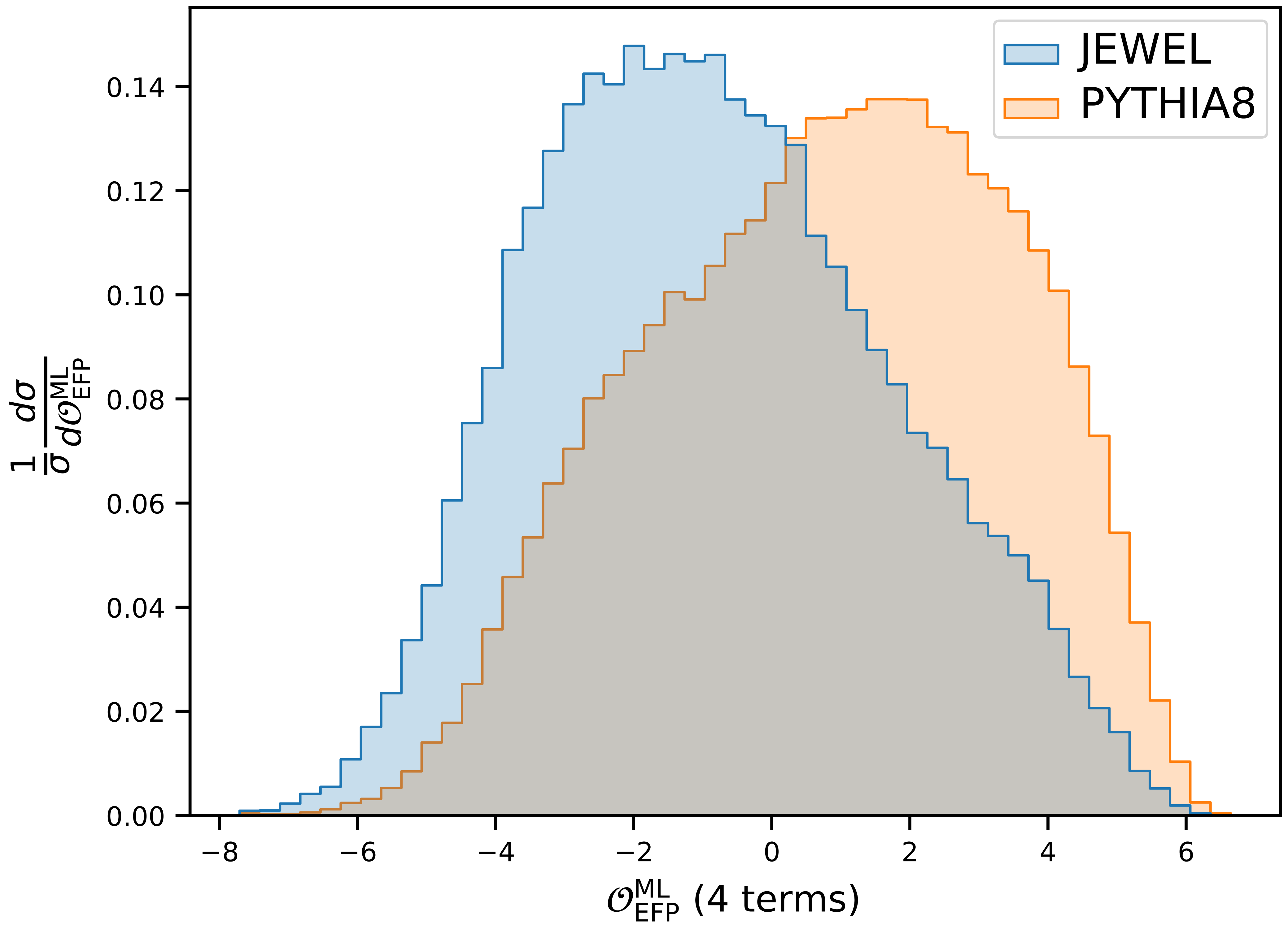
**Lasso regression**

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

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

$\alpha = 0.01 \longrightarrow$ 24 terms
$\alpha = 0.1 \longrightarrow$ 4 terms
$\alpha = 0.5 \longrightarrow$ 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

# Outline

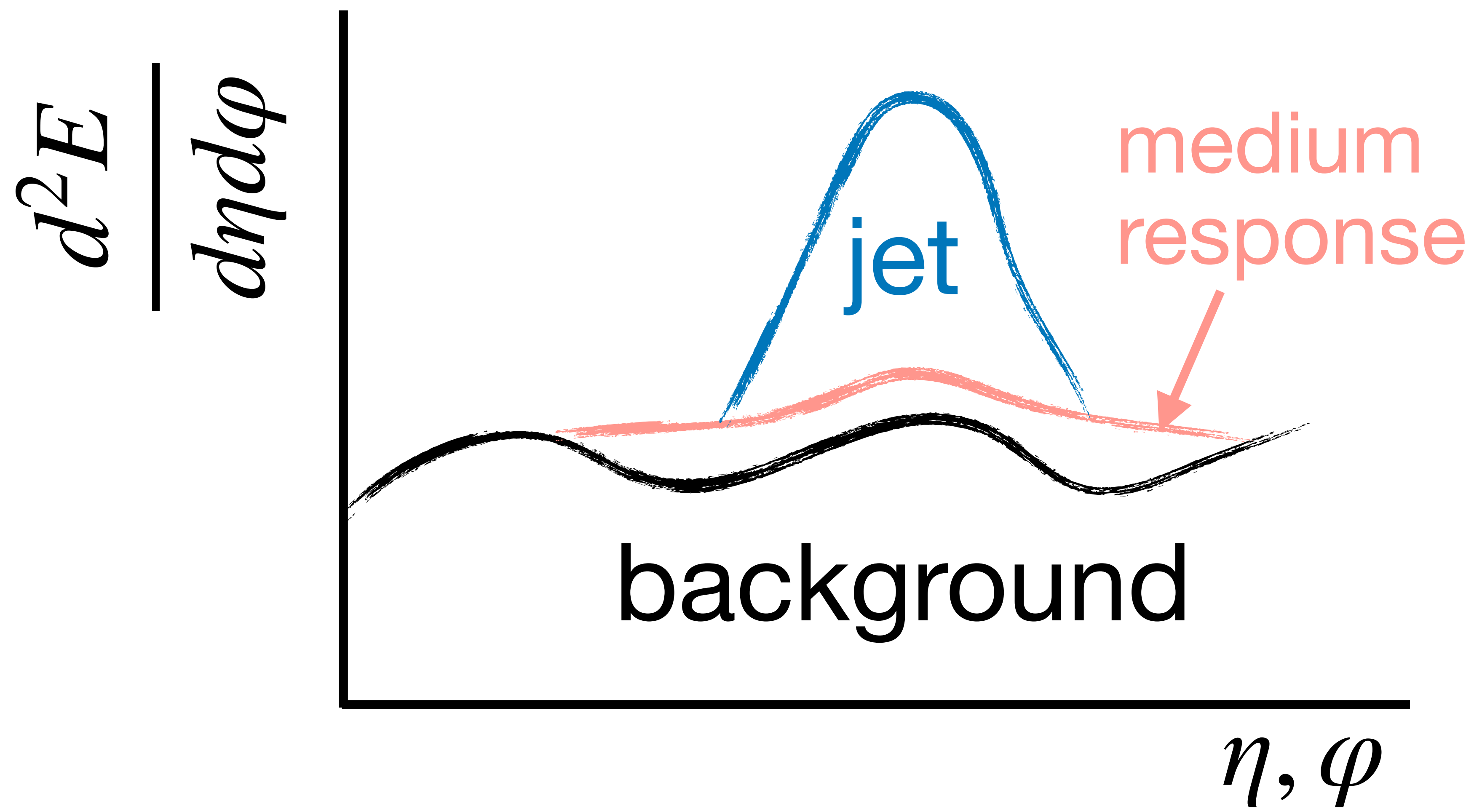
1. Identify the useful information content in the jet

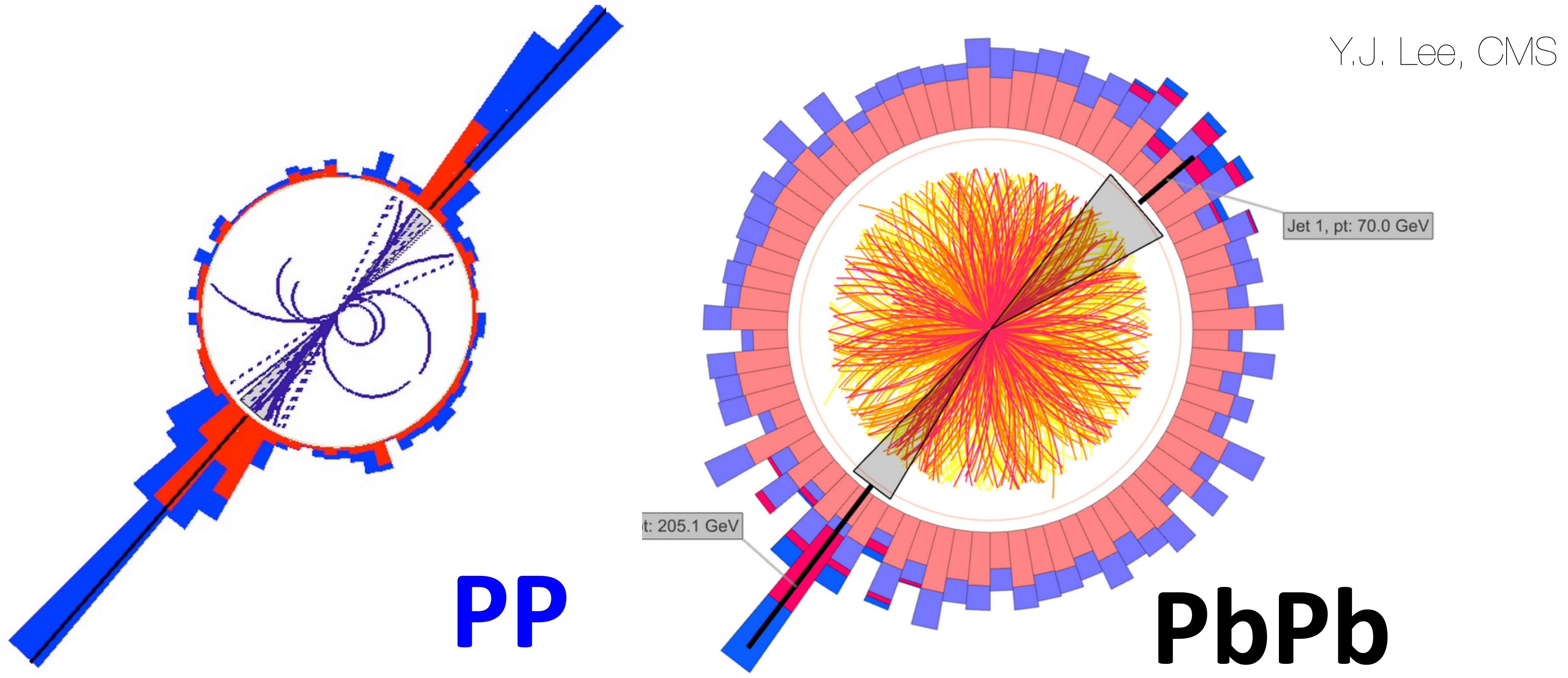
2. Design optimal observables to maximize discrimination

3. Assess information loss due to backgrounds

# Heavy-ion background

The soft collisions in a heavy-ion event produce a large, fluctuating underlying event



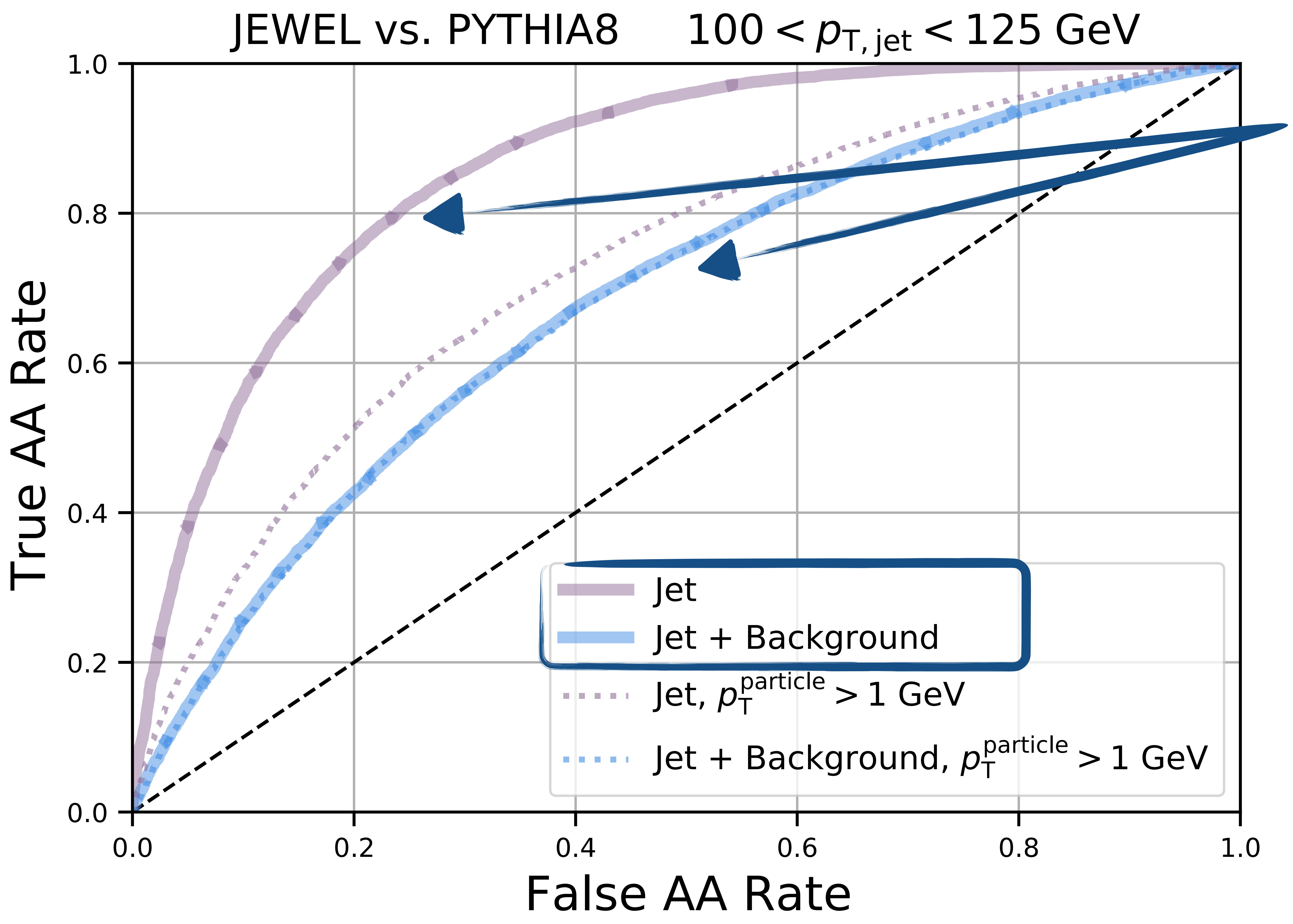




This is a major experimental and theoretical hurdle

**To what extent does the background destroy discriminating power?**

# Results — w/o vs. w/ background

*Lai, Mulligan, Płoskoń, Ringer arXiv 2111.14589*



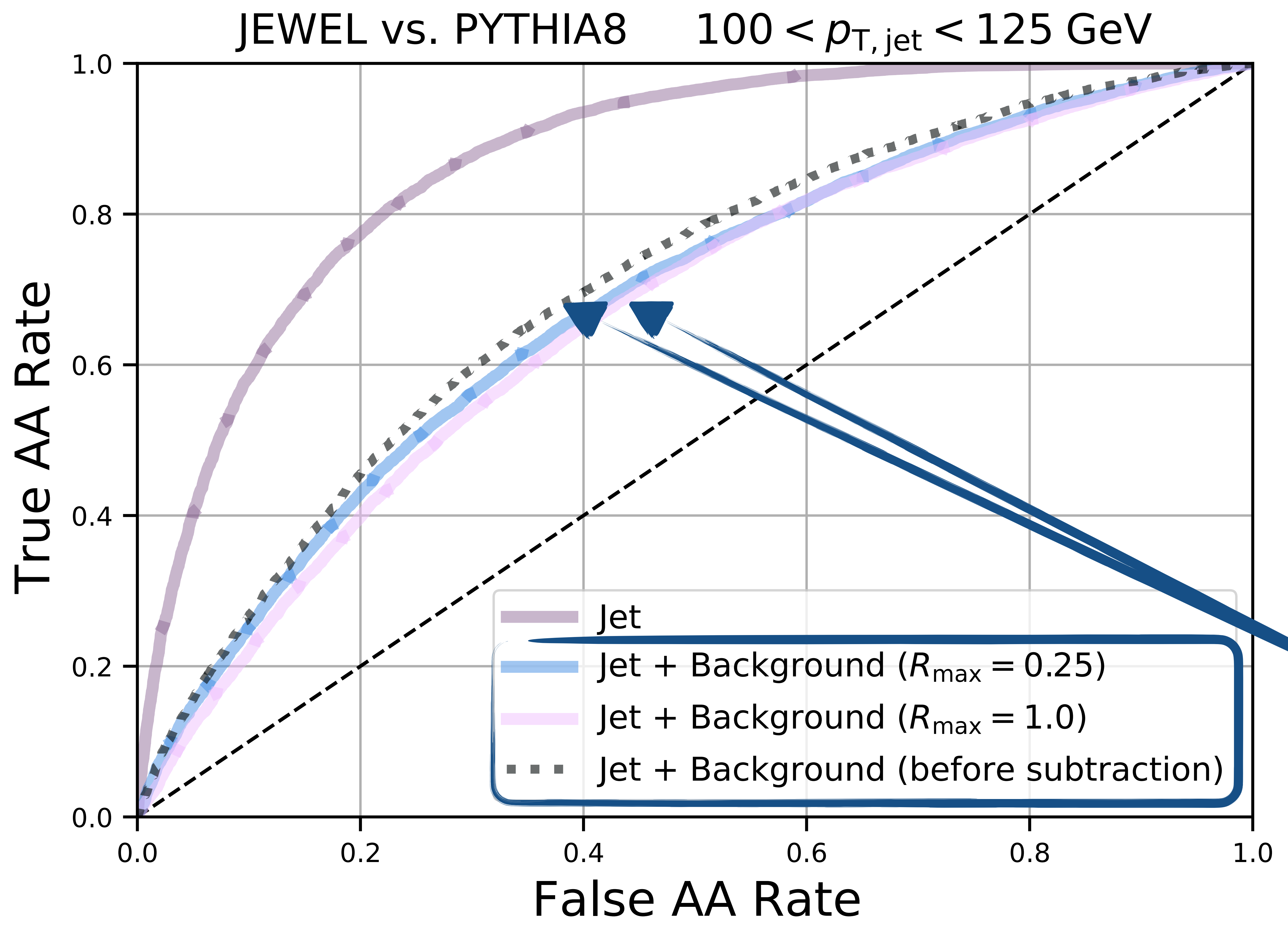

Discriminating power is highly reduced by the fluctuating underlying event

Large, irrecoverable information loss

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

# Assessing background subtraction algorithms

*Lai, Mulligan, Płoskoń, Ringer arXiv 2111.14589*




In practice, experiments must perform background subtraction procedures

This subtraction strictly removes information — use ML classifier performance as a metric to assess

Background subtraction does remove a small but significant amount of information

# Summary

## Theory-motivated binary classifiers to distinguish jets in heavy-ion collisions from those in proton-proton collisions

- New physics insights
  - Important information contained in soft emissions and IRC-unsafe physics
  - Large, irrecoverable information loss due to underlying event
- ML-assisted observable design of optimally discriminating (and calculable!) observables

## These methods can be applied directly to experimental data — labels are known

- Can also apply to full events: LHC, EIC