Machine learning based jet and event identification at the Electron-Ion Collider

with applications to hadron structure and spin physics

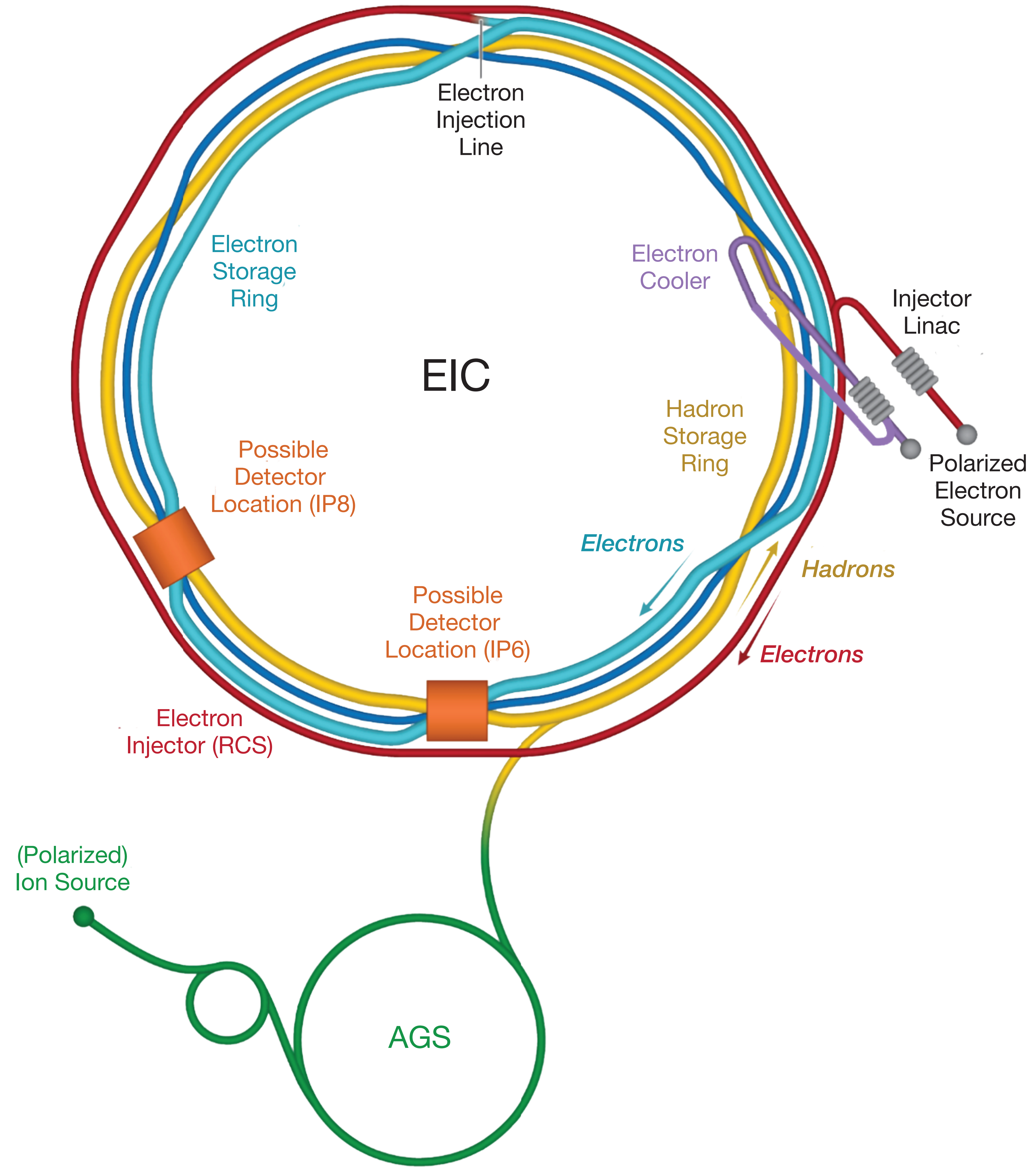
arXiv 2210.06450

K. Lee, J. Mulligan, M. Płoskoń, F. Ringer, F. Yuan

James Mulligan
UC Berkeley / LBNL

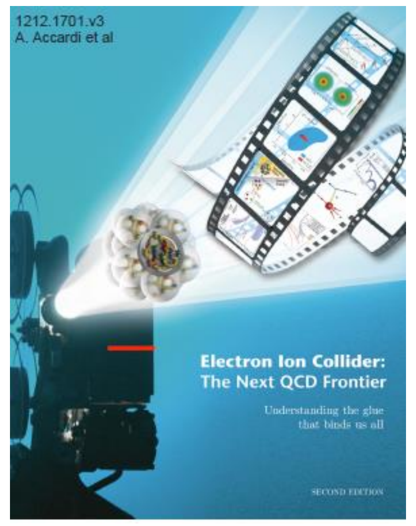
ML4Jets Workshop
Rutgers University
Nov 3, 2022

The Electron-Ion Collider

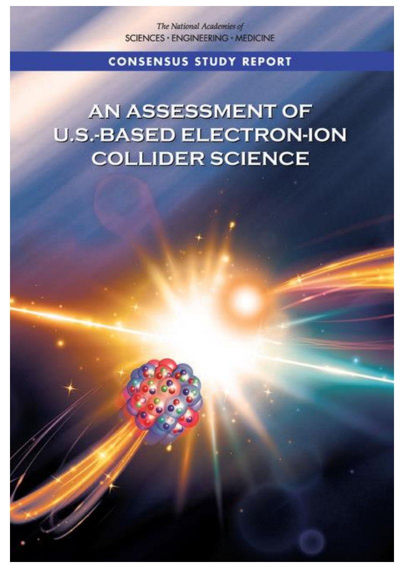


Precision QCD with ep and eA collisions in the 2030s

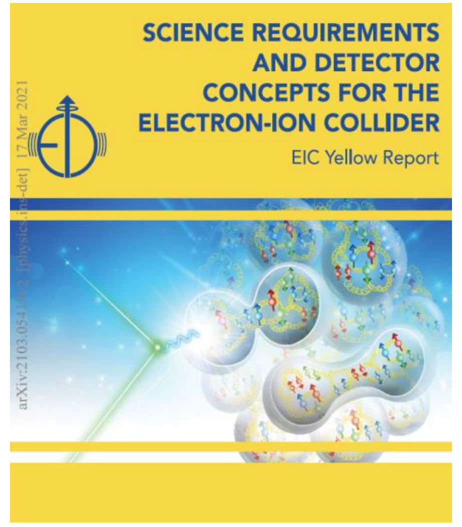
- Polarized electron and proton beams
- Variable ion species: Au, Pb, U
- Variable CM energy: $20 - 140$ GeV
- High luminosity: $10^{33} - 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$



White paper



NAS report

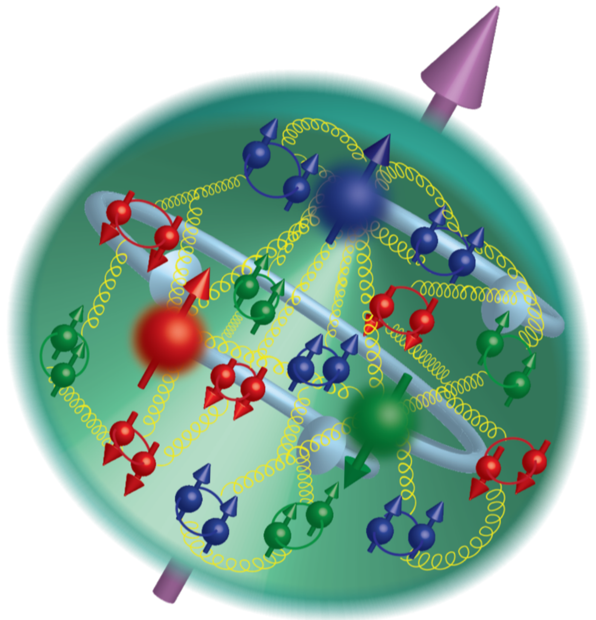


Yellow report

The EIC Science Program

Imaging the proton

How are quarks and gluons distributed in the proton?



Position and *momentum* space

$$W(x, \vec{b}_T, \vec{k}_T)$$

$f(x, \vec{b}_T)$

Generalized parton distributions (GPDs)

$f(x, \vec{k}_T)$

Transverse momentum dependent (TMDs)

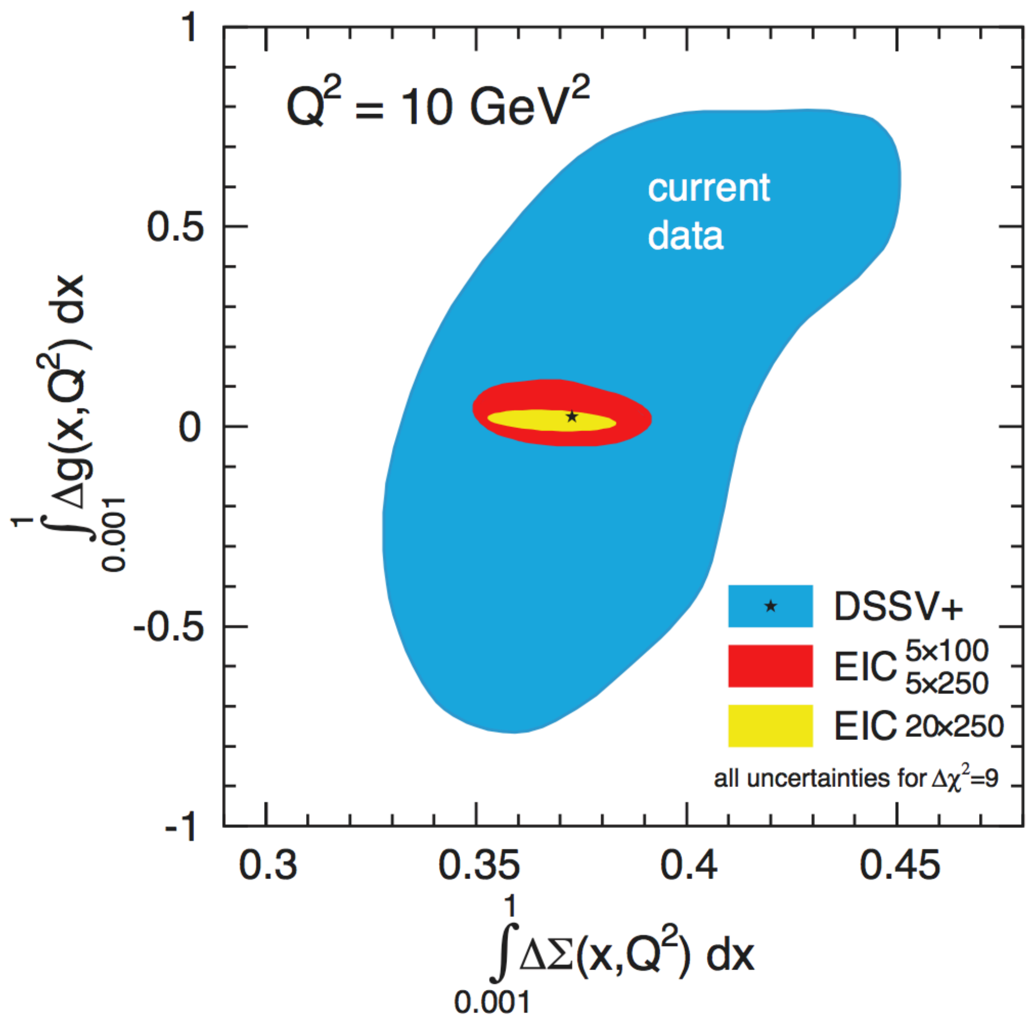
Proton spin puzzle

Where does the proton spin arise from?

$$\frac{1}{2} = \frac{1}{2}\Delta\Sigma + \Delta G + L_q + L_g$$

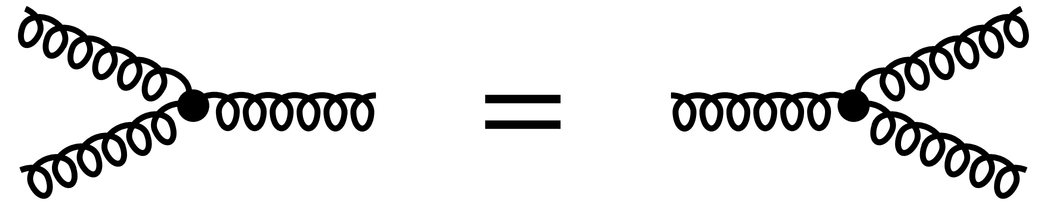
helicity

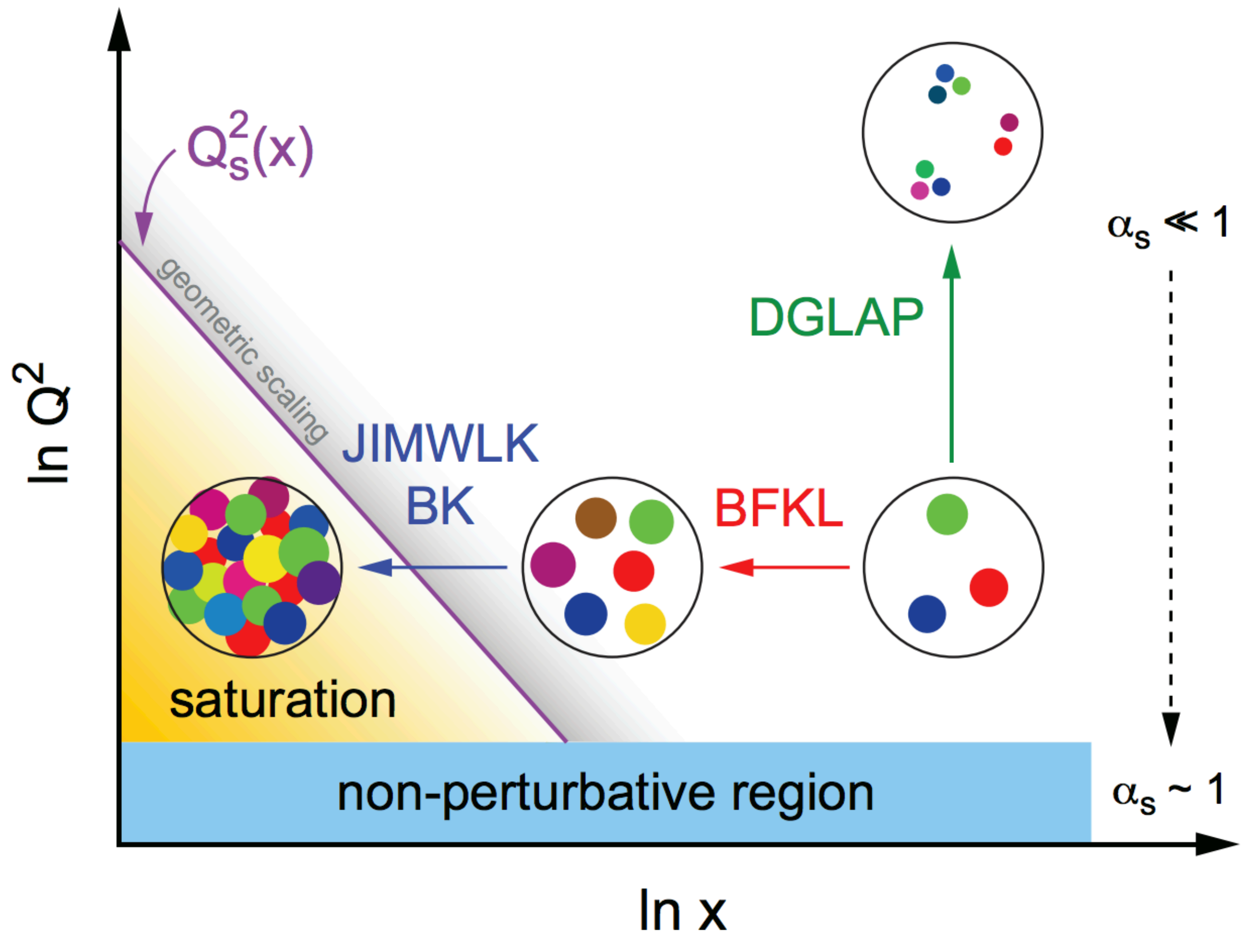
orbital angular momentum



Gluon saturation

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





…

This talk

1. Is machine learning based jet classification useful for the science program of the EIC?

2. How will machine learning based jet taggers perform at the relatively low EIC energies?

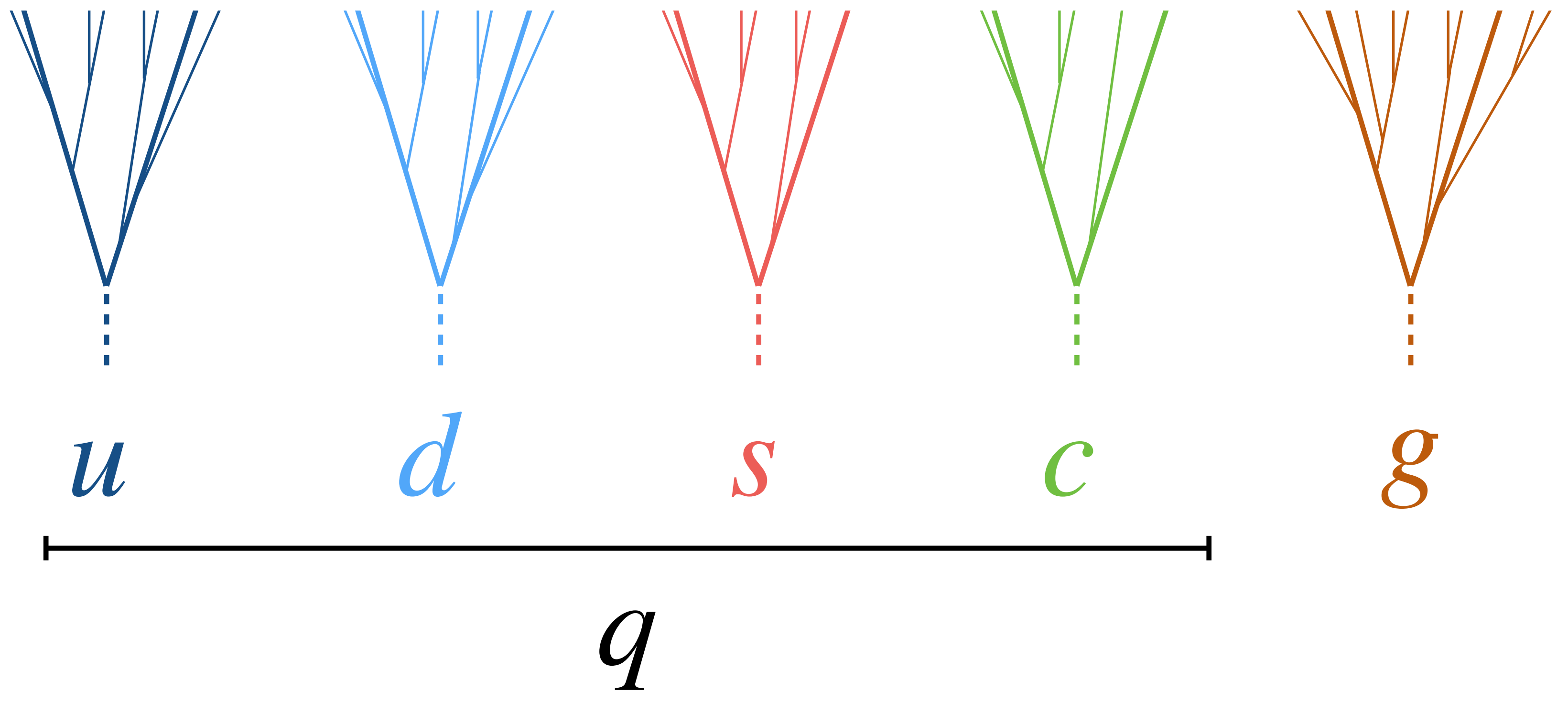
This talk

1. Is machine learning based jet classification useful for the science program of the EIC?

2. How will machine learning based jet taggers perform at the relatively low EIC energies?

Constraining TMDs with jet flavor tagging

Determining the flavor of a jet allows stronger constraints on TMDs by avoiding spin asymmetry cancellations of different flavors



Example: Sivers function (TMD PDF)

Burkhardt sum rule: $$\sum_{a=q,\bar{q},g} \int_0^1 \mathrm{d}x\, f_{1T}^{\perp(1)a}(x) = 0$$

If valence quarks dominate, then u, d Sivers functions have large cancellation

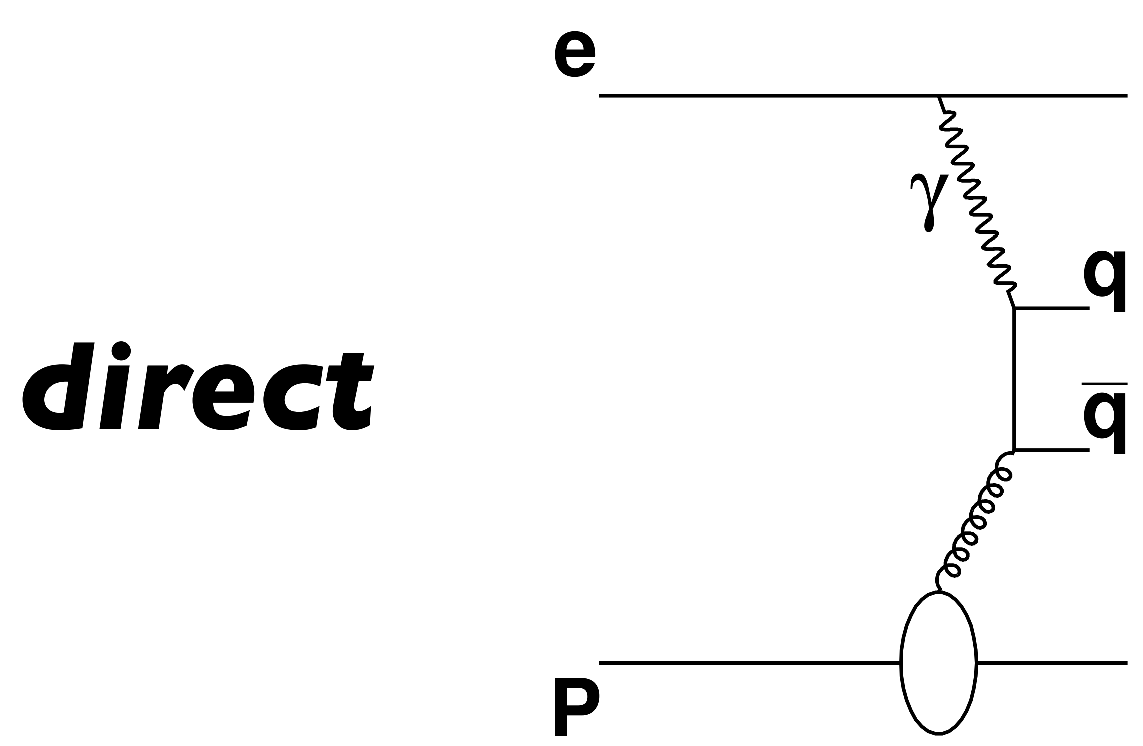
→ Tagging u, d jets separately will allow stronger constraints on Sivers function

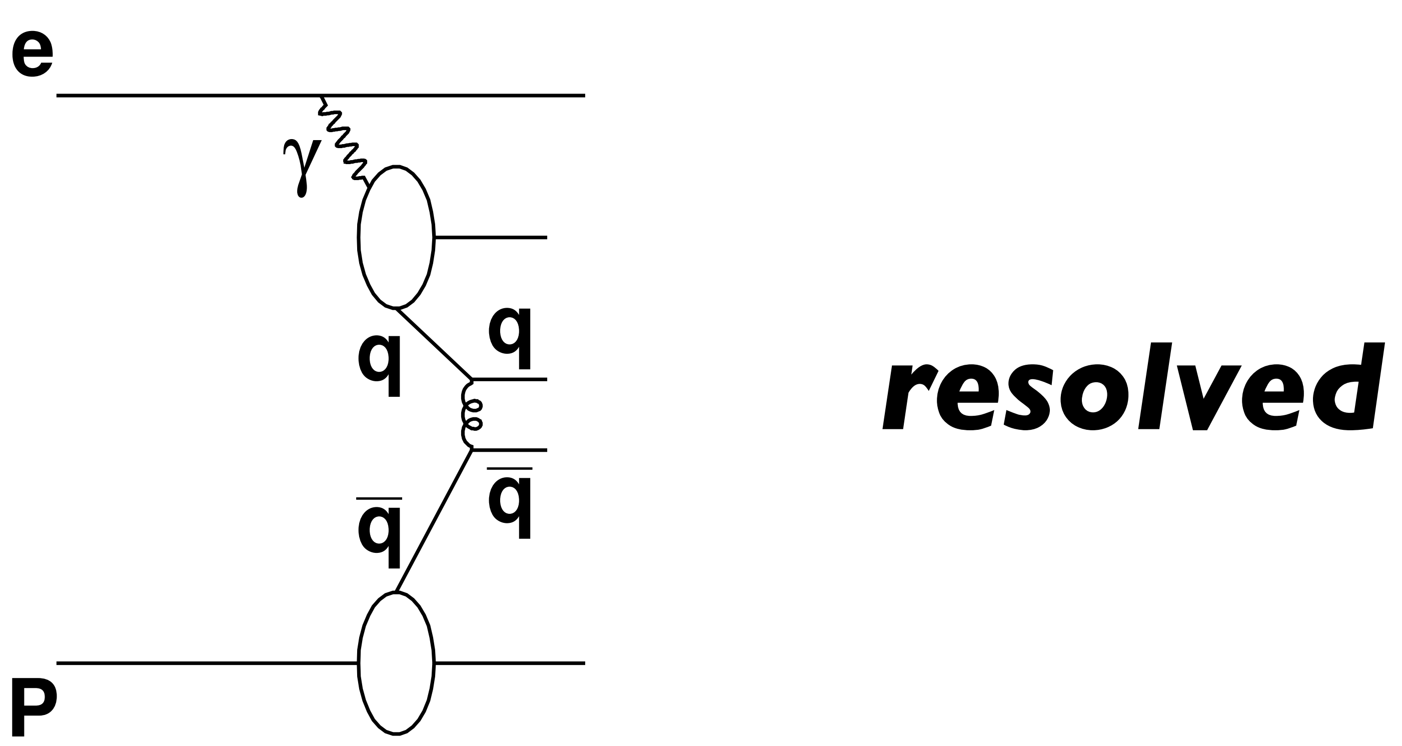
- Recent proposal: use jet charge → *Kang, Liu, Mantry, Shao PRL 125 242003 (2020)*
- Using ML can further boost separation

STAR, R. Fatemi EINN 2019

Constraining photon PDF with event tagging

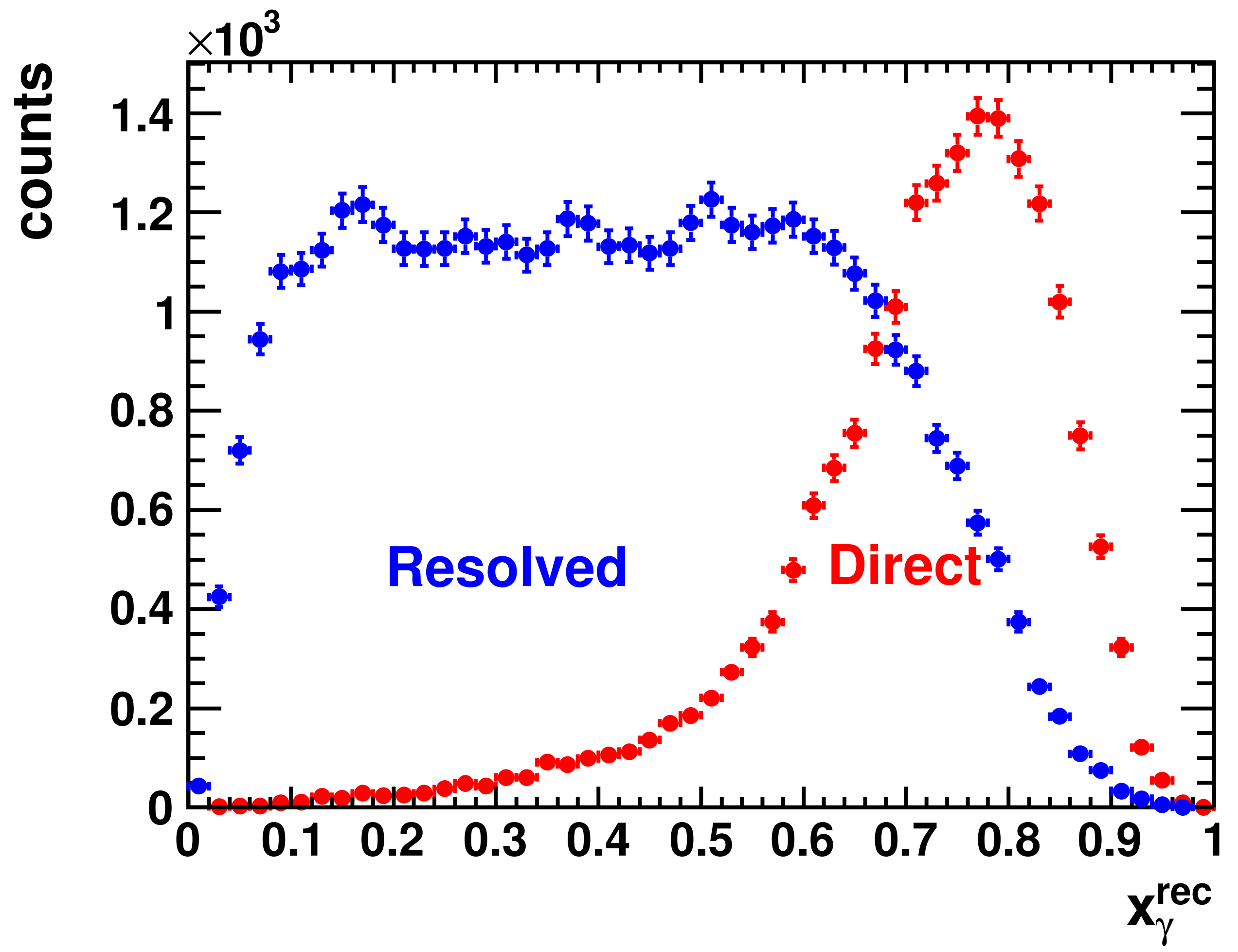
In photoproduction, the resolved processes probe the parton-in-photon PDF





Chu, Aschenauer, Lee, Zheng PRD 96 7, 074035 (2017)

By classifying direct vs. resolved photoproduction processes with ML, can enhance constraints on the photon PDF relative to traditional observables



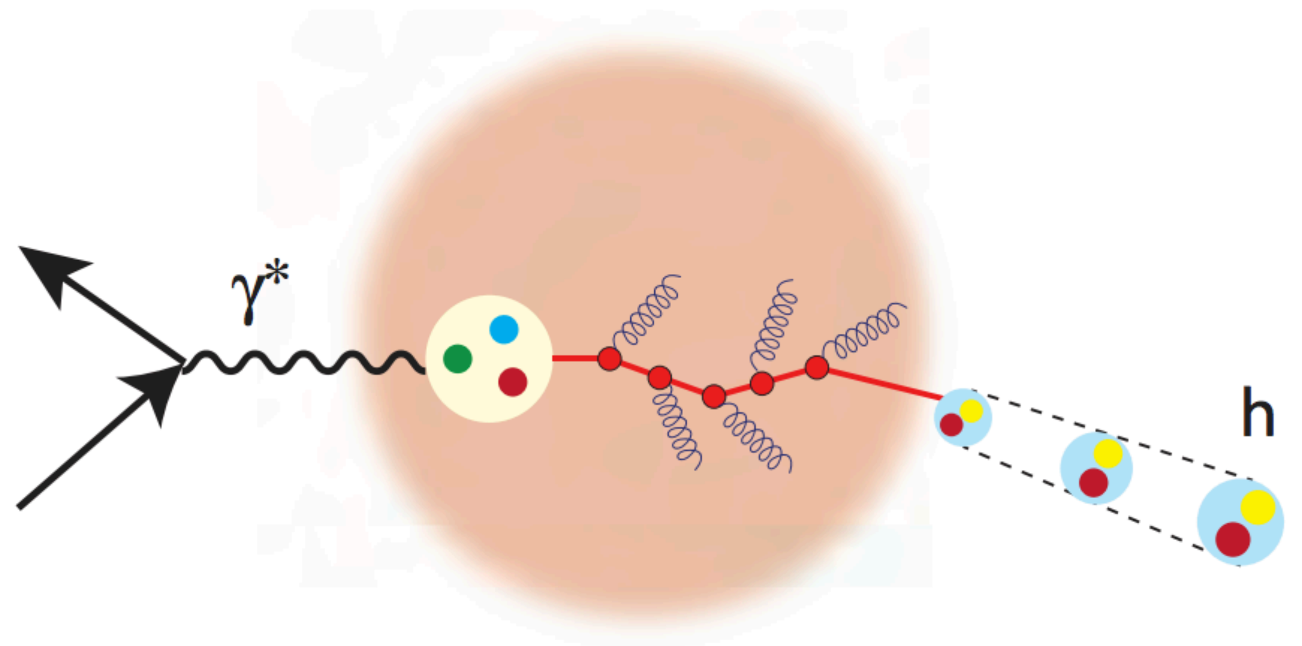
e.g. $$x_\gamma = \frac{1}{2E_e y}\left(p_{T,1}e^{-\eta_1} + p_{T,2}e^{-\eta_2}\right)$$

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)

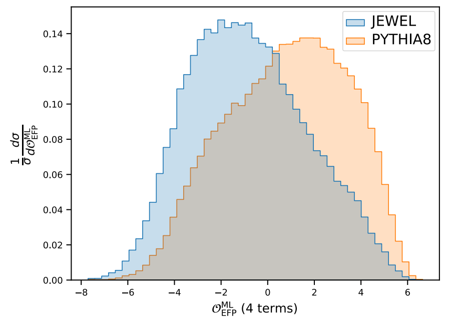


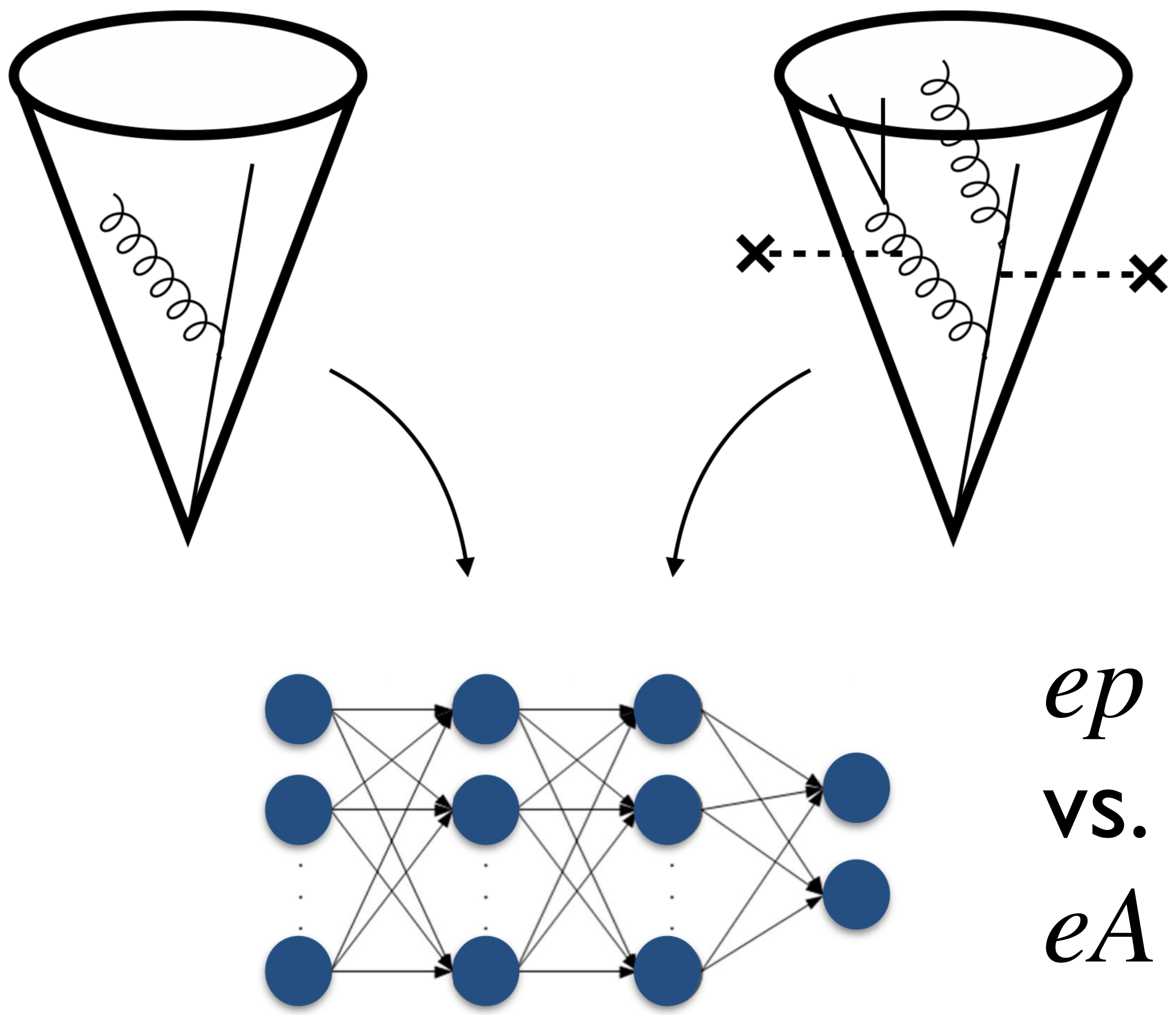
Train ML classifier to distinguish ep vs. eA jets

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

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





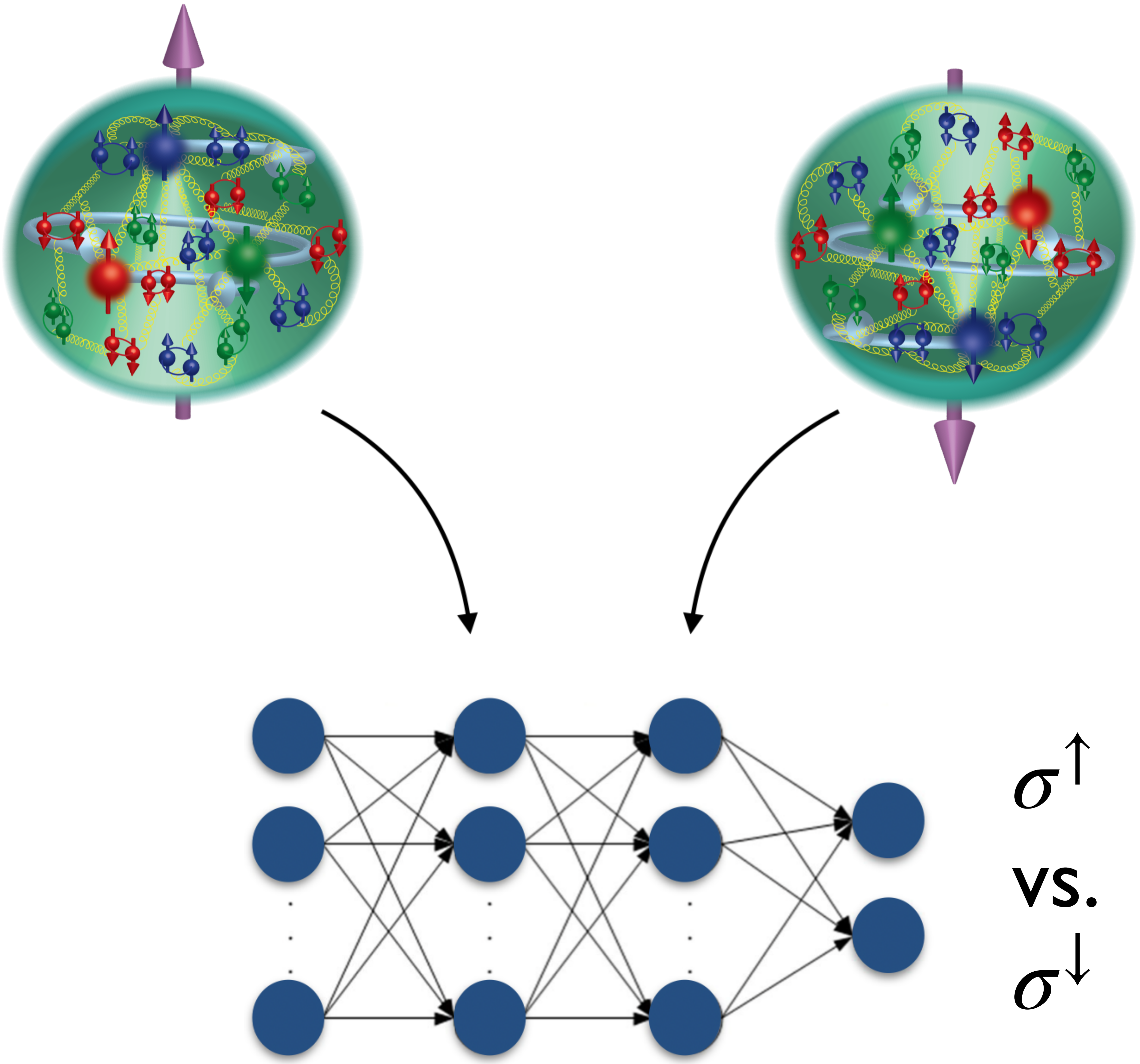
Can be applied directly on experimental data

Maximizing spin asymmetries

Goal: Measure non-zero TSSAs associated with jets:

$$A_{UT} = \frac{\mathrm{d}\sigma^{\uparrow} - \mathrm{d}\sigma^{\downarrow}}{\mathrm{d}\sigma^{\uparrow} + \mathrm{d}\sigma^{\downarrow}}$$

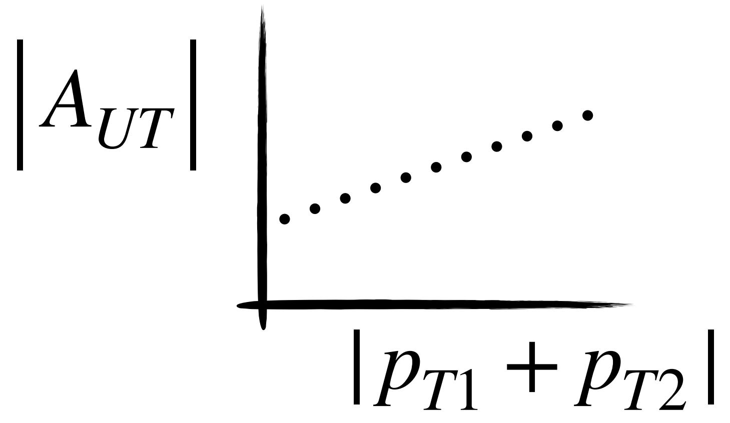
STAR PRL 99, 142003 (2007)
STAR arXiv 2205.11800



Train ML classifier to distinguish ↑ vs. ↓ jets

Can use *interpretable* ML to design maximally discriminating observables that are calculable in pQCD

$$\max_{\theta} |A_{UT}(\theta)| \longrightarrow$$



Can be applied directly on experimental data

Can be applied at RHIC now!

This talk

1. Is machine learning based jet classification useful for the science program of the EIC?

2. How will machine learning based jet taggers perform at the relatively low EIC energies?

Setup

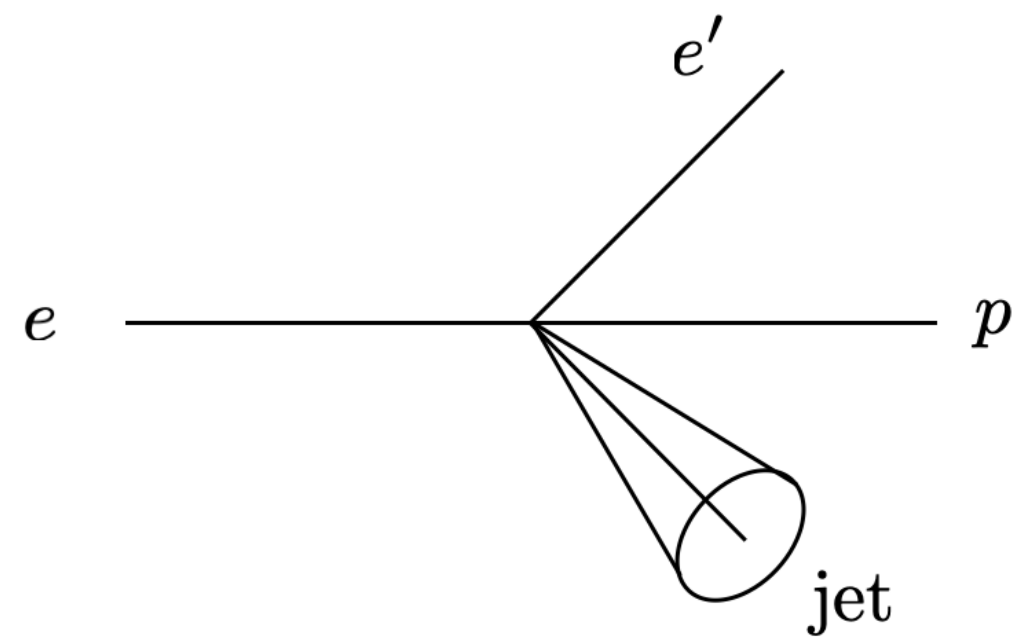
Event generation

PYTHIA6

- No detector simulation
- Vary minimum particle p_T, PID info

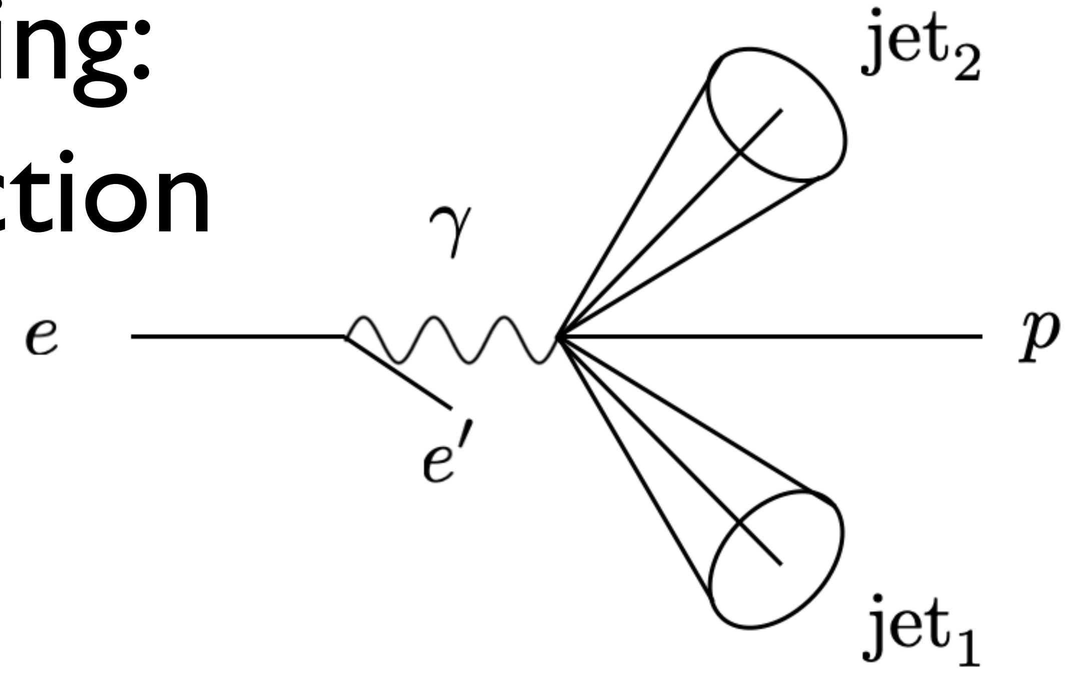
For u/d/s/c tagging:
LO DIS

$p_{T,\mathrm{jet}} > 10$ GeV



For quark/gluon tagging:
low-Q^2 photoproduction

$p_{T,\mathrm{jet1}} > 8$ GeV
$p_{T,\mathrm{jet2}} > 5$ GeV



Machine learning model

Binary classification: u vs. d, ud vs. s, …

Architecture: Particle Flow Networks

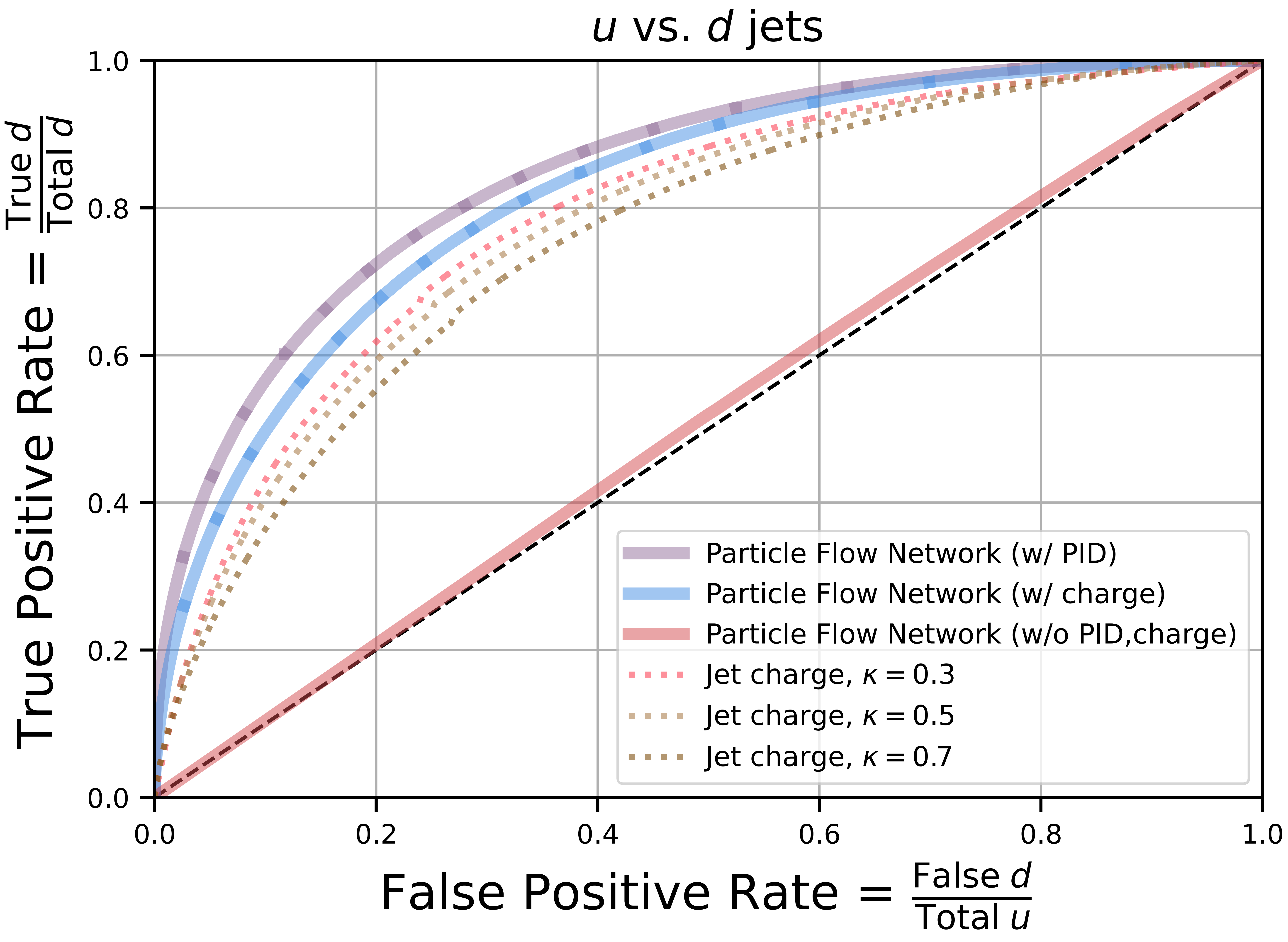
$$f(p_1, \ldots, p_M) = F\left(\sum_{i=1}^{M} \Phi\left(p_i\right)\right)$$

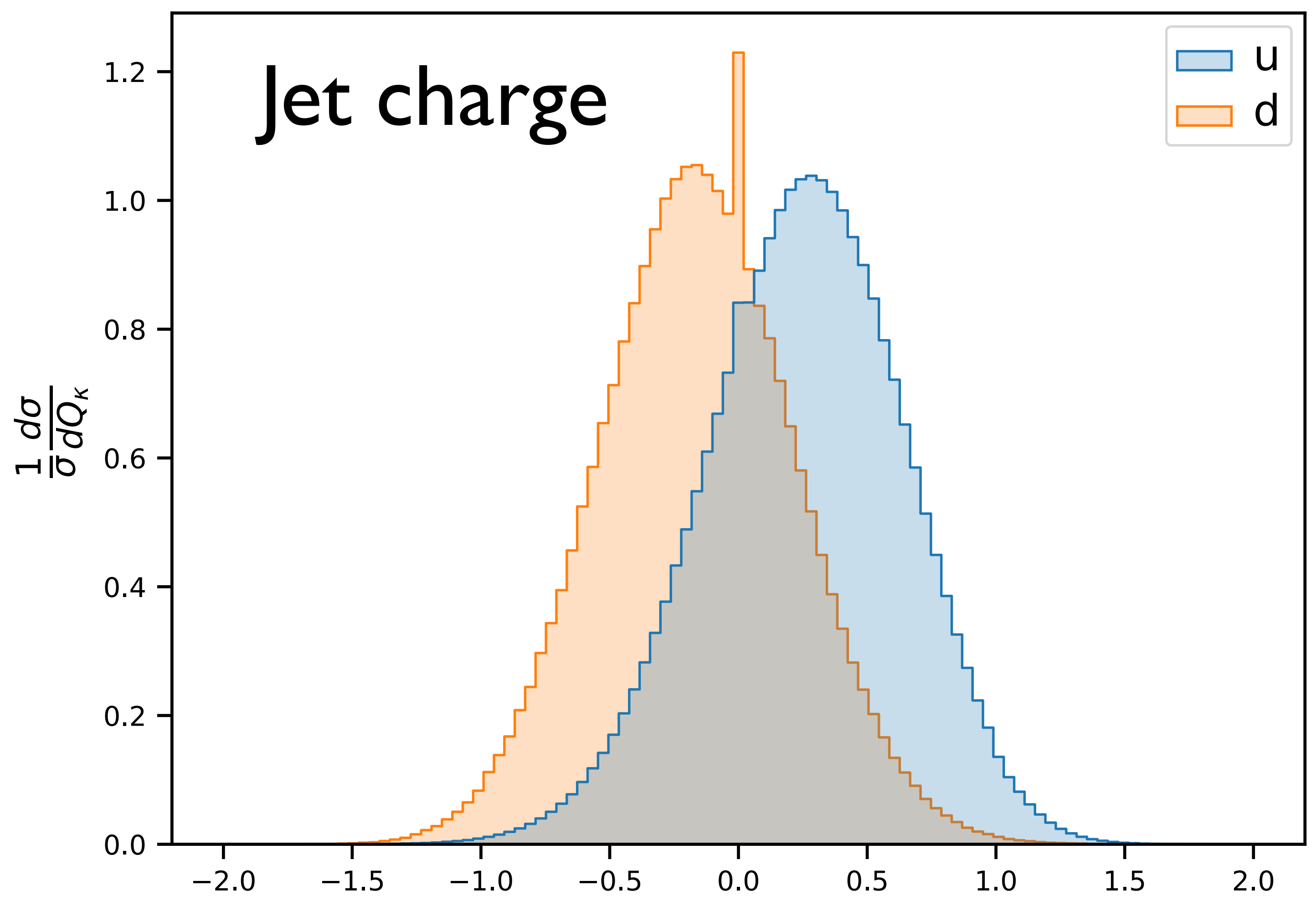
Classifier

DNNs

Komiske, Metodiev, Thaler JHEP 01 (2019) 121

Jet flavor tagging: u vs. d



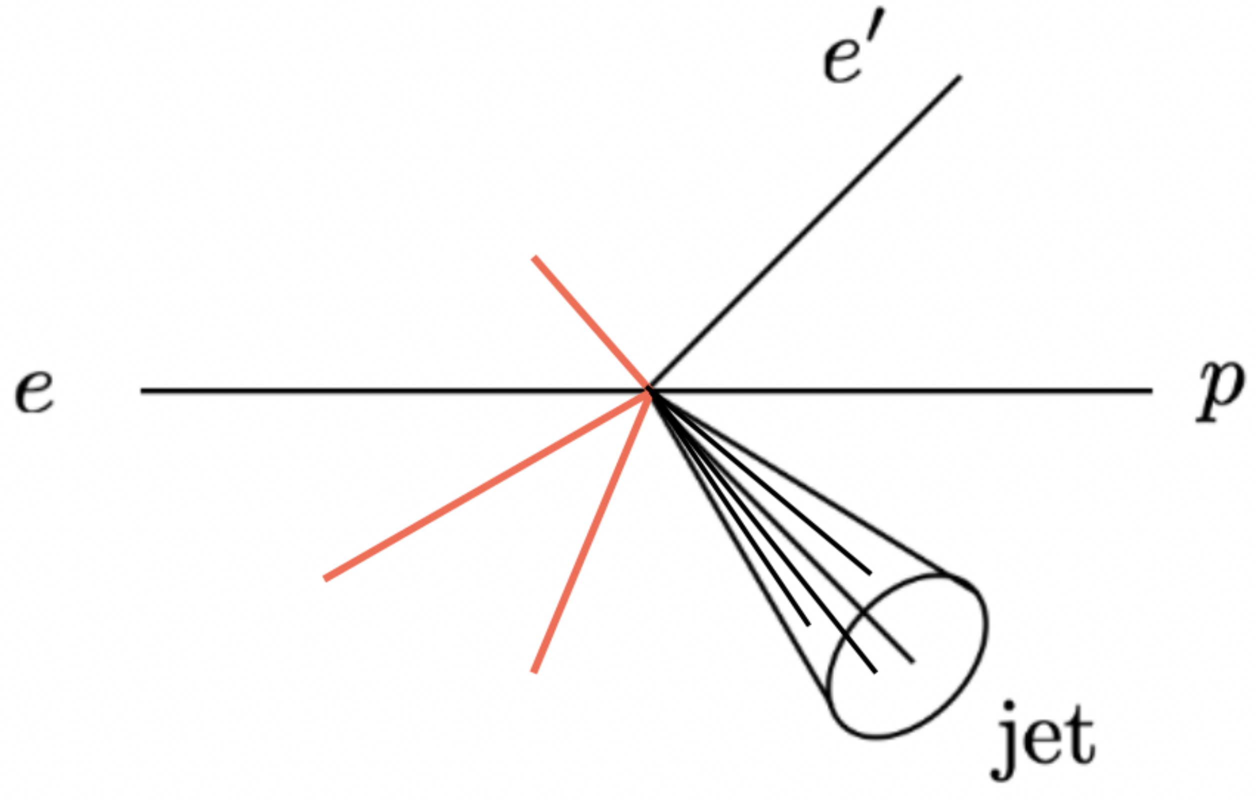


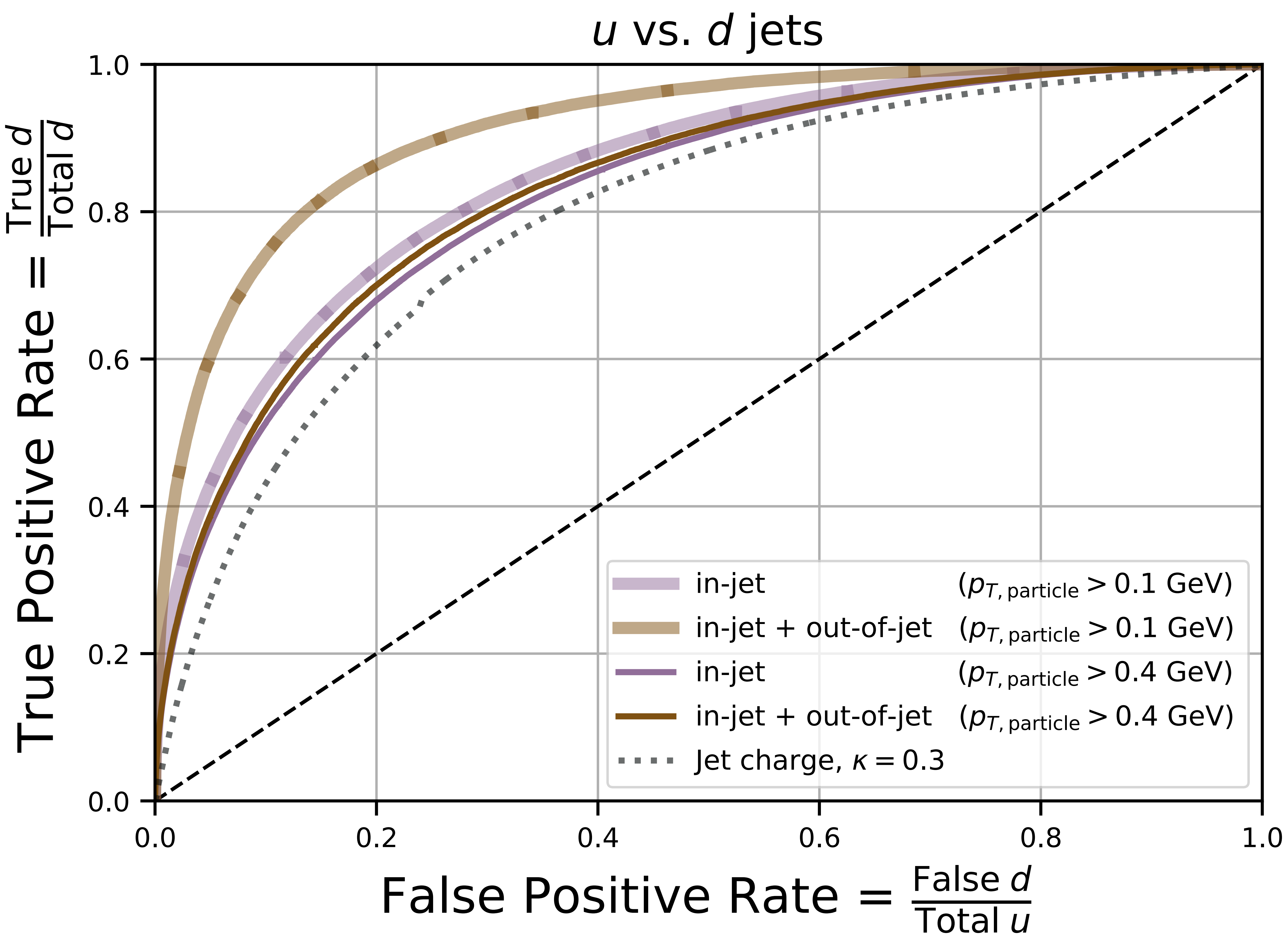
$$Q_\kappa = \sum_{i \in \text{jet}} z_i^\kappa Q_i$$

- ML outperforms jet charge
  - Charge information is crucial
  - Full PID does not gain much

Out-of-jet information

Does including out-of-jet information boost the jet flavor tagging performance?

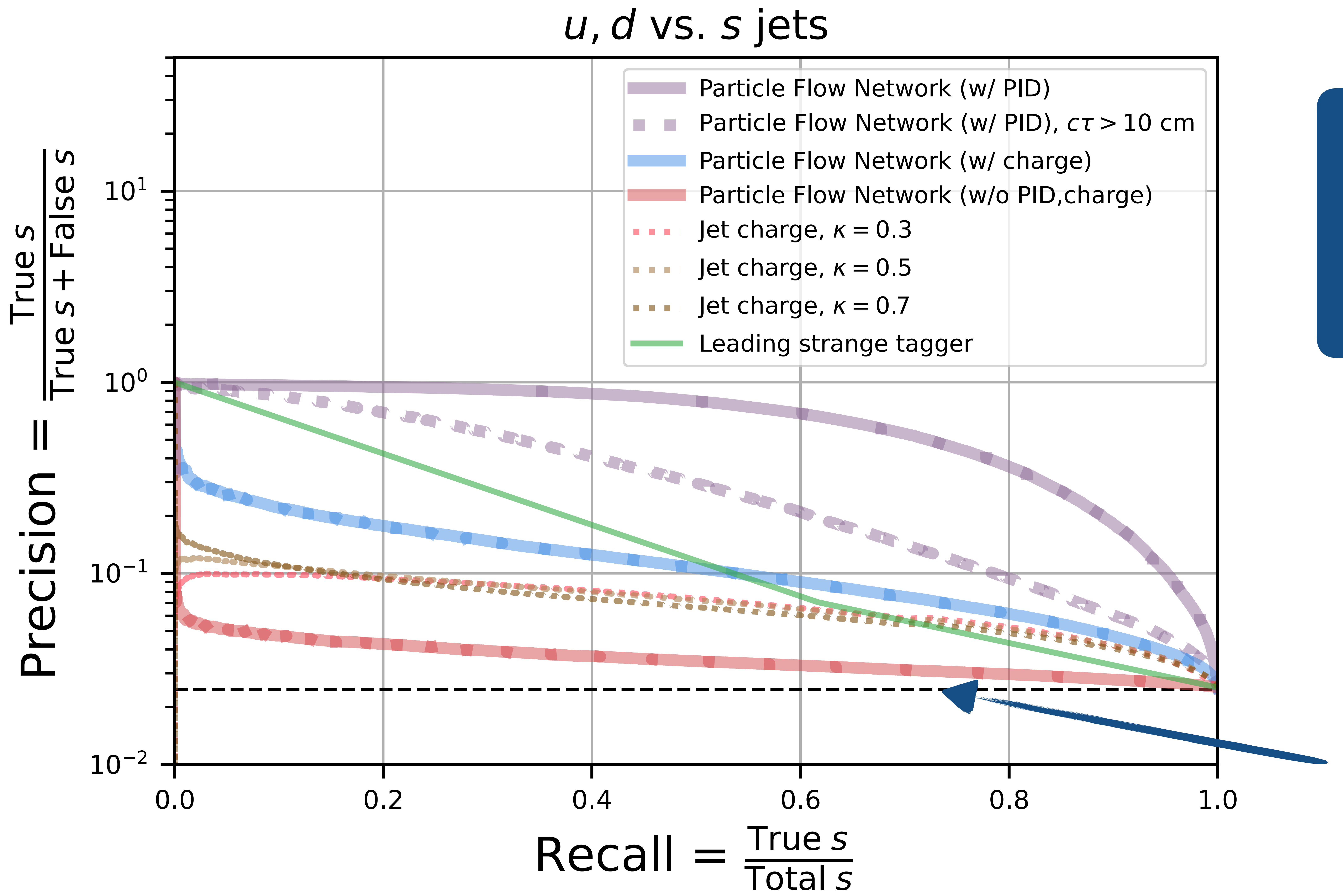




Significant gain from out-of-jet information

- Due to soft particles $0.1 < p_T < 0.4$ GeV

Jet flavor tagging: *ud* vs. *s*



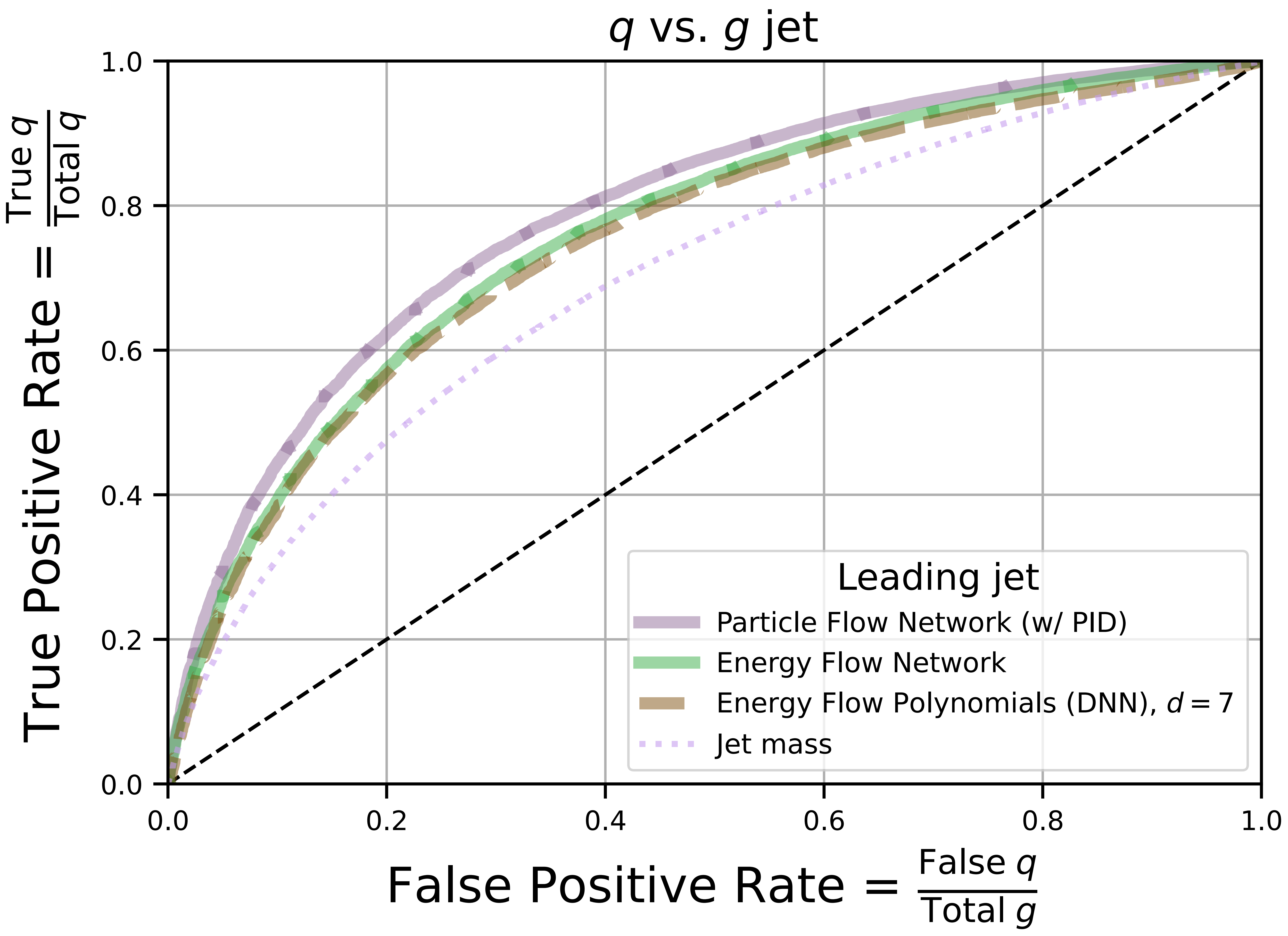
For strange: ML dramatically outperforms jet charge

- PID gives huge boost

We use precision-recall metric since there are ~40x more *ud* than *s*

- Precision ↔ Purity
- Recall ↔ Efficiency

Quark vs. gluon jet tagging



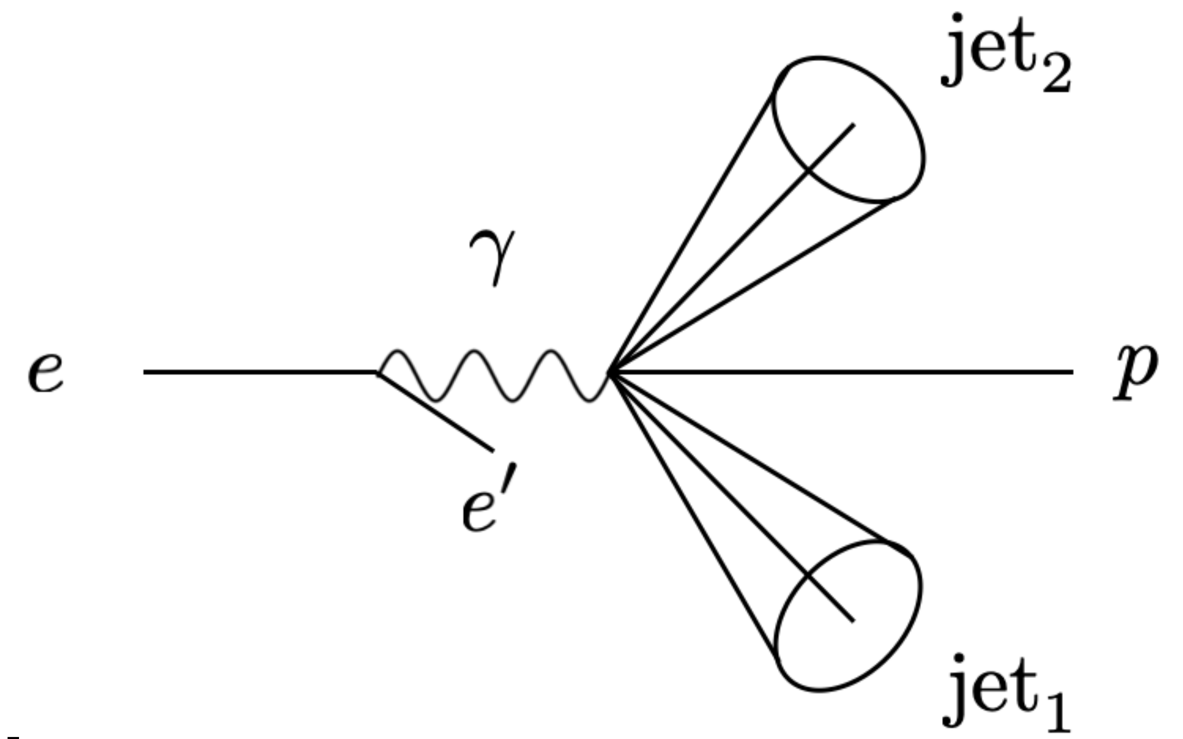
ML performance not as good as at LHC, but still reasonably good

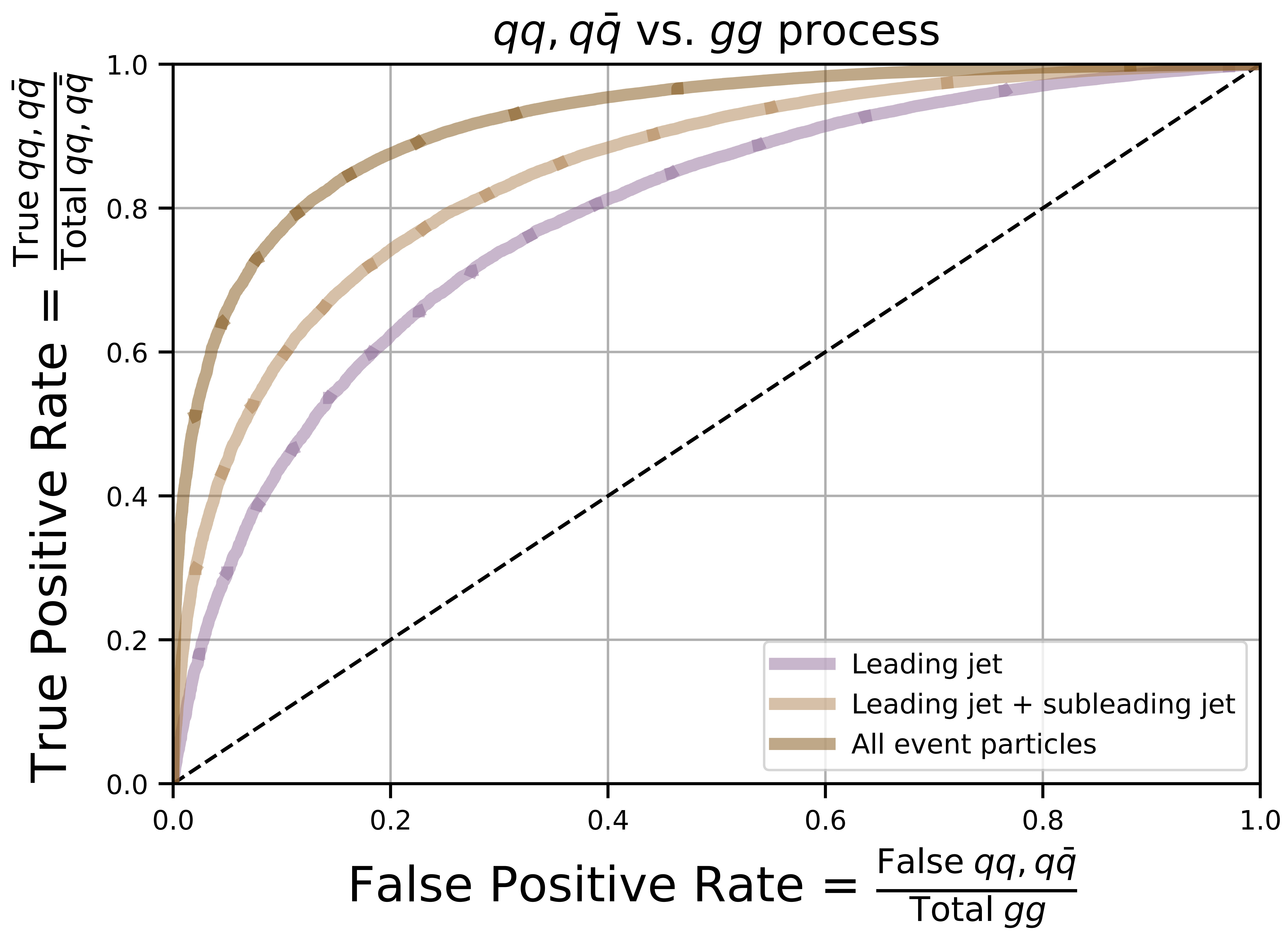
AUC	EIC	LHC
Particle Flow Network	0.79	0.91
Energy Flow Network	0.76	0.88
Energy Flow Polynomials	0.75	0.89

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

Komiske, Metodiev, Thaler, JHEP 04, 013 (2018)

Hard process tagging





We classify hard processes generating $qq/q\bar{q}$ vs. gg di-jets:

$$qq \to qq, q\bar{q} \to q\bar{q}, gg \to q\bar{q}, \gamma_T^* g \to q\bar{q}, \gamma_L^* g \to q\bar{q}$$

vs.

$$q\bar{q} \to gg, gg \to gg,$$

→ Can use this method to tag resolved photoproduction contributions

Significant improvement when adding subleading jet and out-of-jet particles

Summary

Machine learning can improve access to hadron structure and spin physics at the EIC

- Improve jet flavor tagging performance: constrain TMDs, photon PDF, …
- Maximize the size of spin asymmetries or cold nuclear matter effects — train directly on data

PYTHIA6 indicates that classification performance remains reasonably good at EIC

- Large performance boost from ML for strange and charm tagging when PID is included
- Large performance boost by including soft, out-of-jet particles

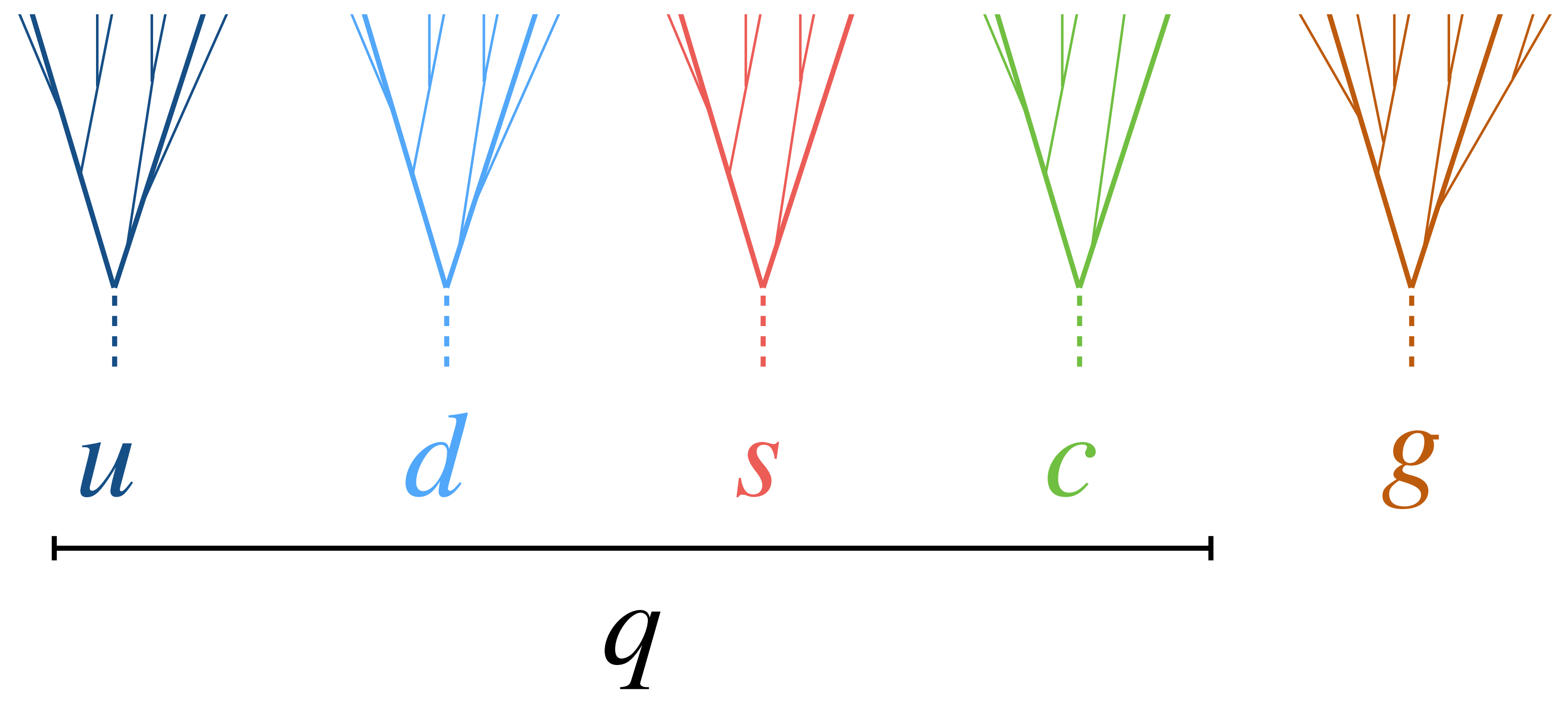
Outlook: Study model-dependence and connect ML results to theory

- Design analytically tractable observables and/or incorporate classifiers into global fits
- Explore ML architectures — data set to be made public soon

backup

Constraining TMDs with jet flavor tagging

Determining the flavor of a jet allows stronger constraints on TMDs by avoiding spin asymmetry cancellations of different flavors



Example: Collins fragmentation function

Schäfer-Teryaev sum rule: $$\sum_h \int_0^1 \mathrm{d}z\, H_{1,h/q}^{\perp(3)}(z) = 0$$

One usually measures identified hadrons to avoid e.g. π^+ cancellation with π^-

However the fragmentation functions still contain large parton flavor cancellations:

$$\int_0^1 \mathrm{d}z \left(H_{1,\pi^+/u}^{\perp(3)}(z) + H_{1,\pi^+/d}^{\perp(3)}(z) \right) \approx 0$$

→ Tagging jet flavor will allow stronger constraints on Collins fragmentation function

e.g. *Arratia, Kang, Produkin, Ringer PRD 201 7, 074015 (2020)*

Additional applications of jet flavor tagging

- Longitudinally polarized gluon distribution Δg — quark flavor and quark vs. gluon

  Zhou, Sato, Melnitchouk (JAM), PRD 105, 074022 (2022)

- Gluon Sivers function — quark vs. gluon

  Zheng, Aschenauer, Lee, Xiao, Yin, PRD 98, 034011 (2018)
  Liu, Ringer, Vogelsang, Yuan, PRL122, 192003 (2019)

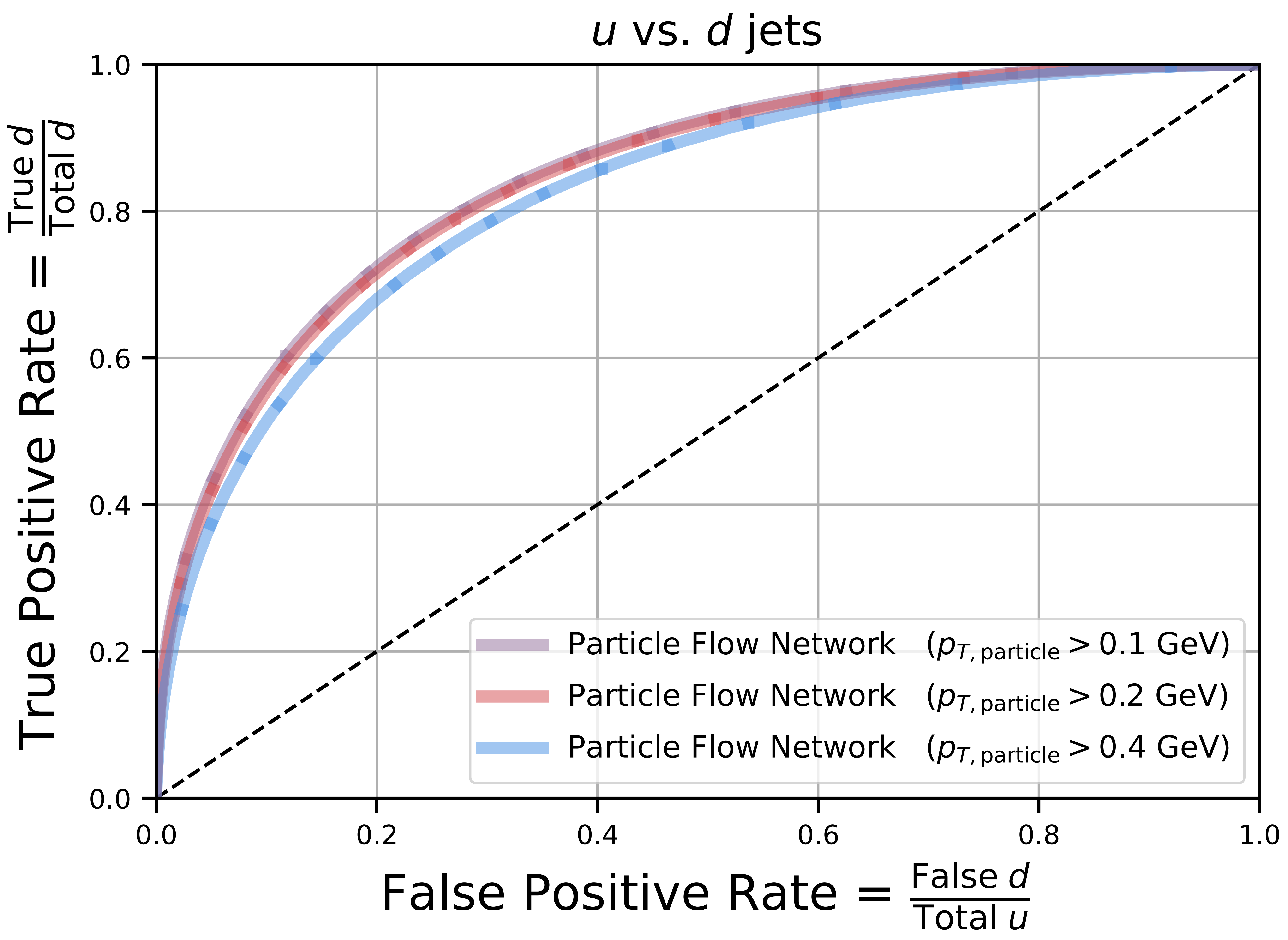
- Strange quark PDF — charm tagging

  Arratia, Furletova, Hobbs, Olness, Sekula, PRD 103, 074023 (2021)

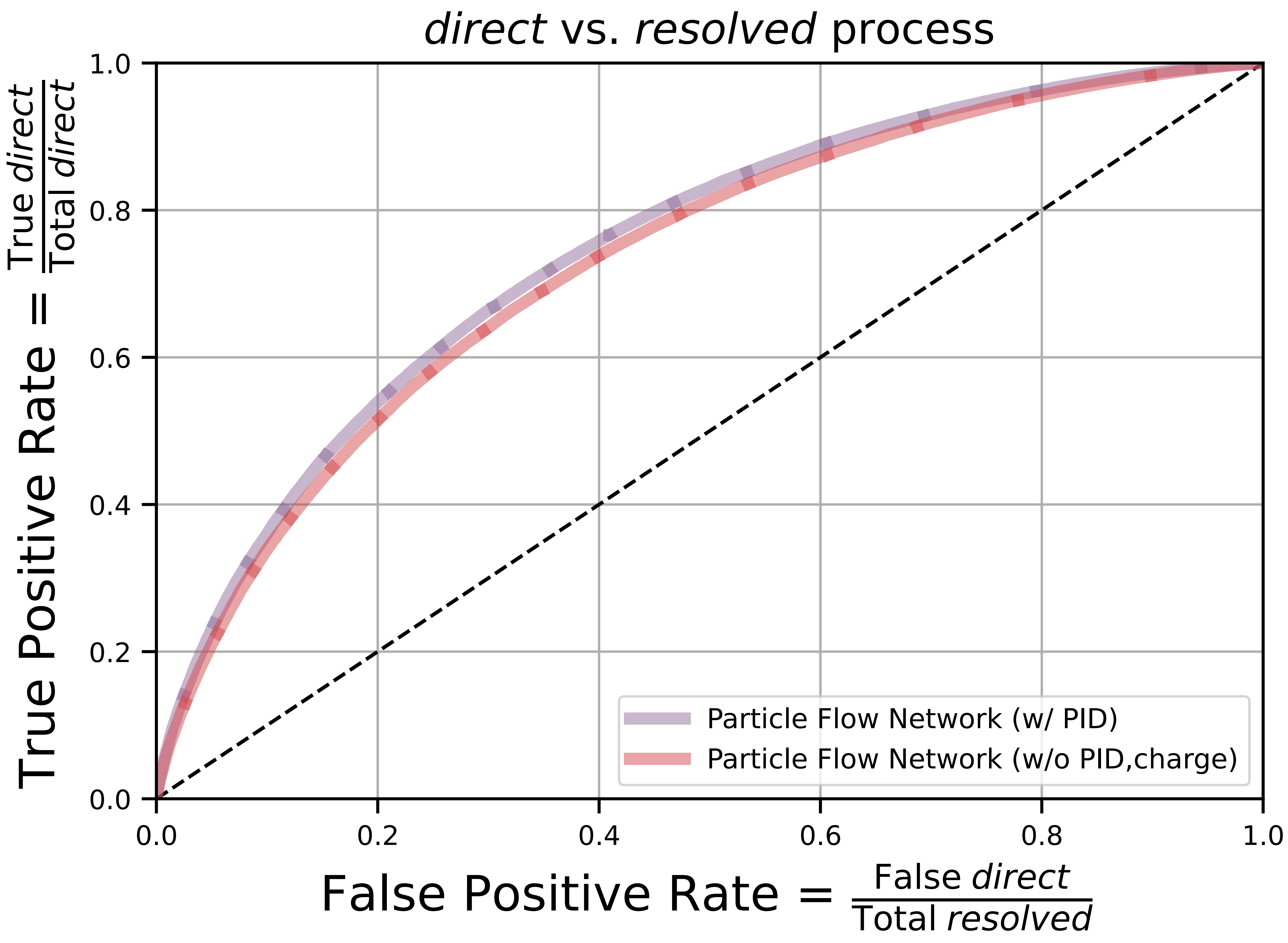
- BSM searches — quark flavor

  Li, Yan, Yuan, arXiv:2112.07747

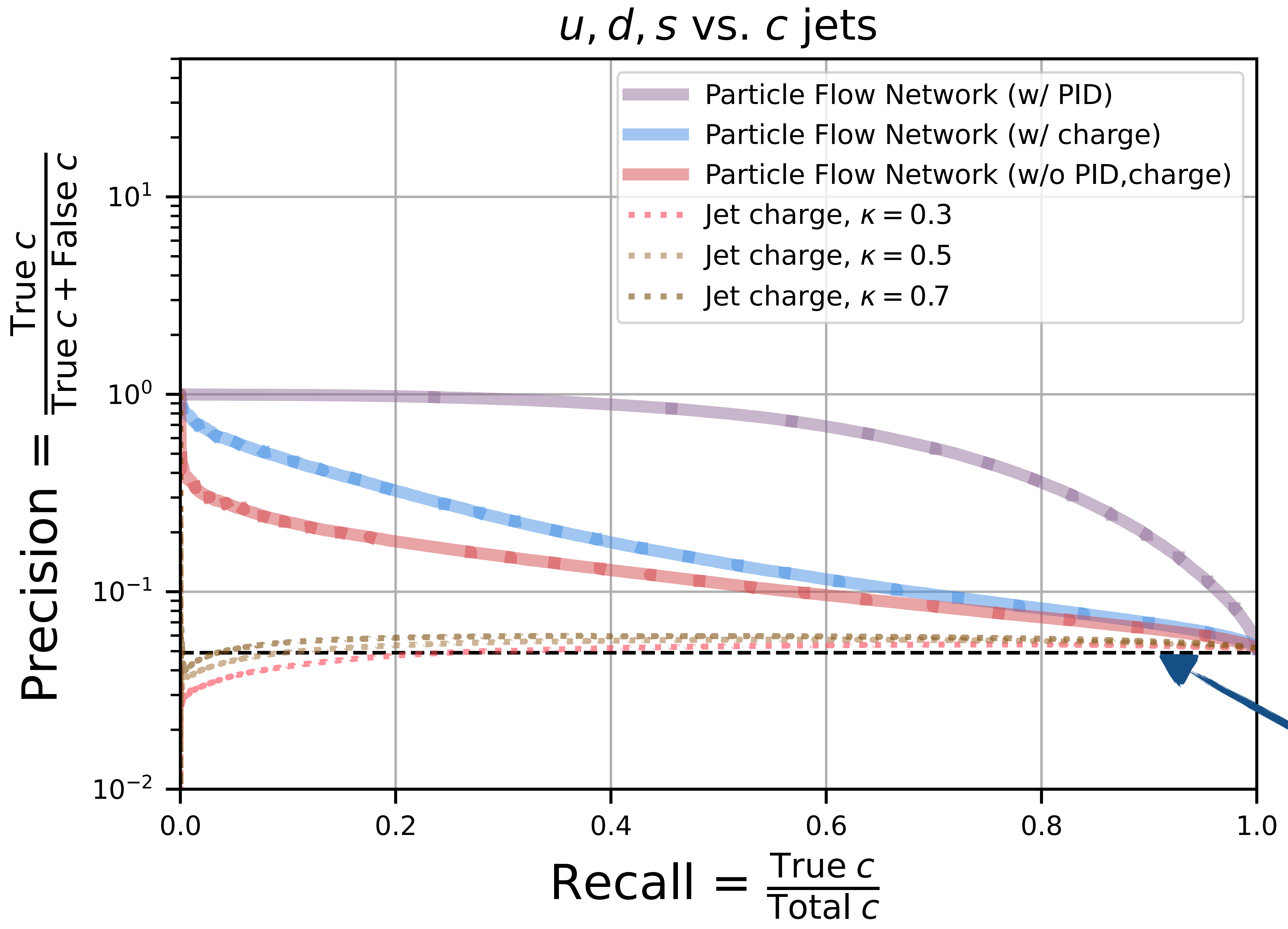
Dependence on minimum particle p_T



Direct vs. resolved photon tagging



Jet flavor tagging: *uds* vs. *c*



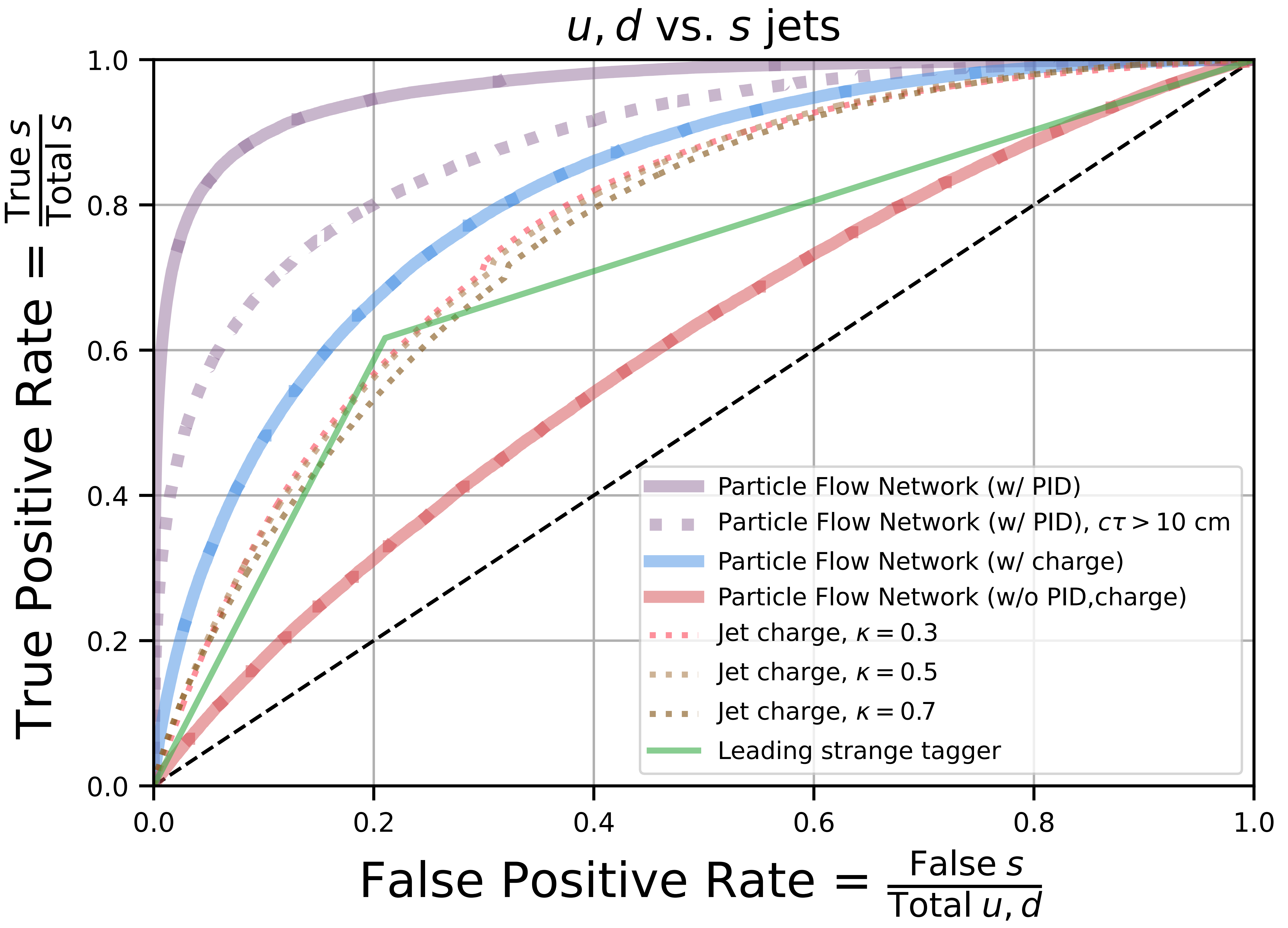
Random classifier

For charm: fragmentation pattern increasingly important, but PID is crucial

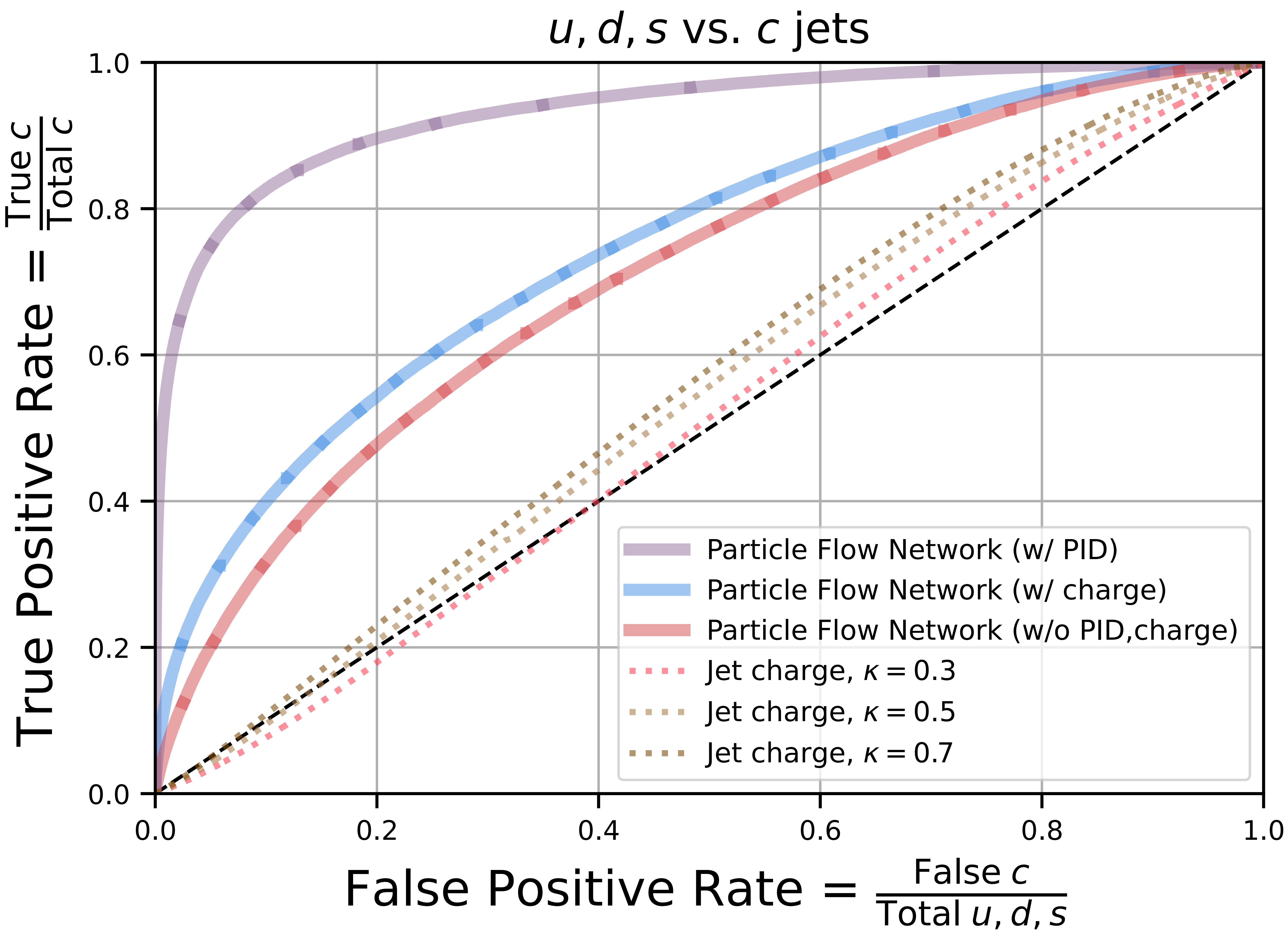
We use precision-recall metric since there are ~20x more *uds* than *c*

- Precision ↔ Purity
- Recall ↔ Efficiency

ud vs. s


u, d vs. s jets
True Positive Rate = $\frac{\text{True } s}{\text{Total } s}$
False Positive Rate = $\frac{\text{False } s}{\text{Total } u, d}$
0.0
0.2
0.4
0.6
0.8
1.0
Particle Flow Network (w/ PID)
Particle Flow Network (w/ PID), $c\tau > 10$ cm
Particle Flow Network (w/ charge)
Particle Flow Network (w/o PID,charge)
Jet charge, $\kappa = 0.3$
Jet charge, $\kappa = 0.5$
Jet charge, $\kappa = 0.7$
Leading strange tagger

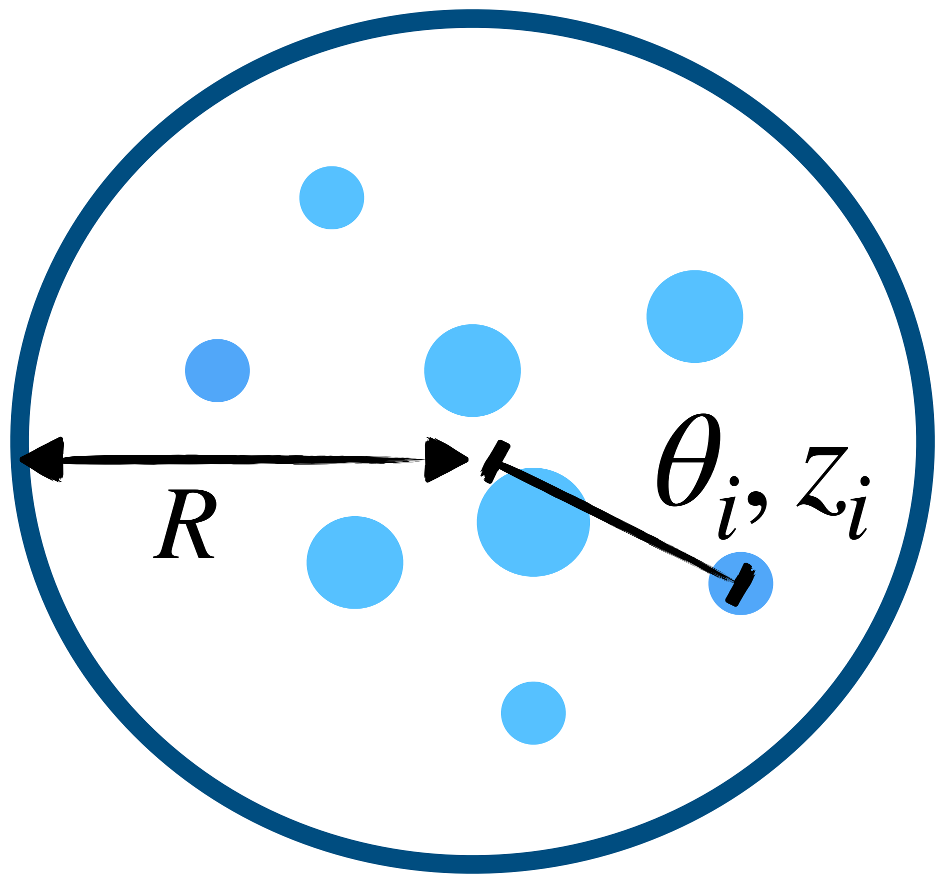
uds vs. c


u, d, s vs. c jets
True Positive Rate = $\frac{\text{True } c}{\text{Total } c}$
False Positive Rate = $\frac{\text{False } c}{\text{Total } u, d, s}$
0.0
0.2
0.4
0.6
0.8
1.0
Particle Flow Network (w/ PID)
Particle Flow Network (w/ charge)
Particle Flow Network (w/o PID,charge)
Jet charge, $\kappa = 0.3$
Jet charge, $\kappa = 0.5$
Jet charge, $\kappa = 0.7$

Jet observables and IRC safety

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

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

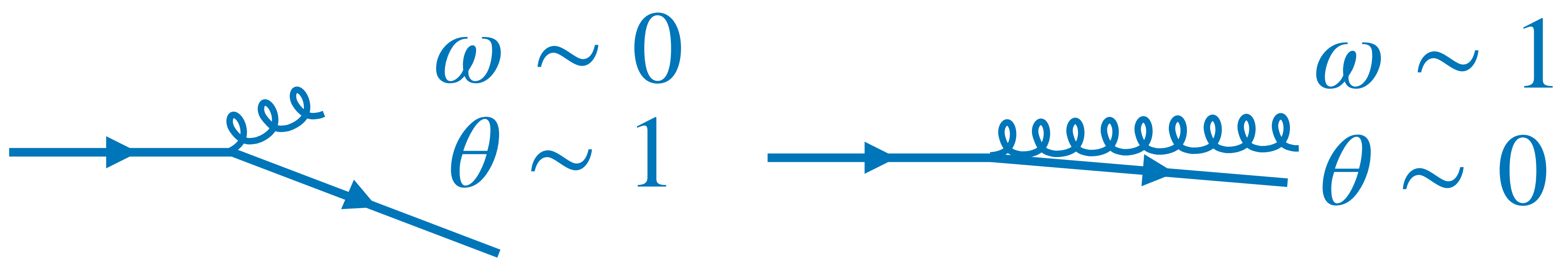


$$\theta_i = \frac{\sqrt{\Delta y^2 + \Delta\varphi^2}}{R}$$

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

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

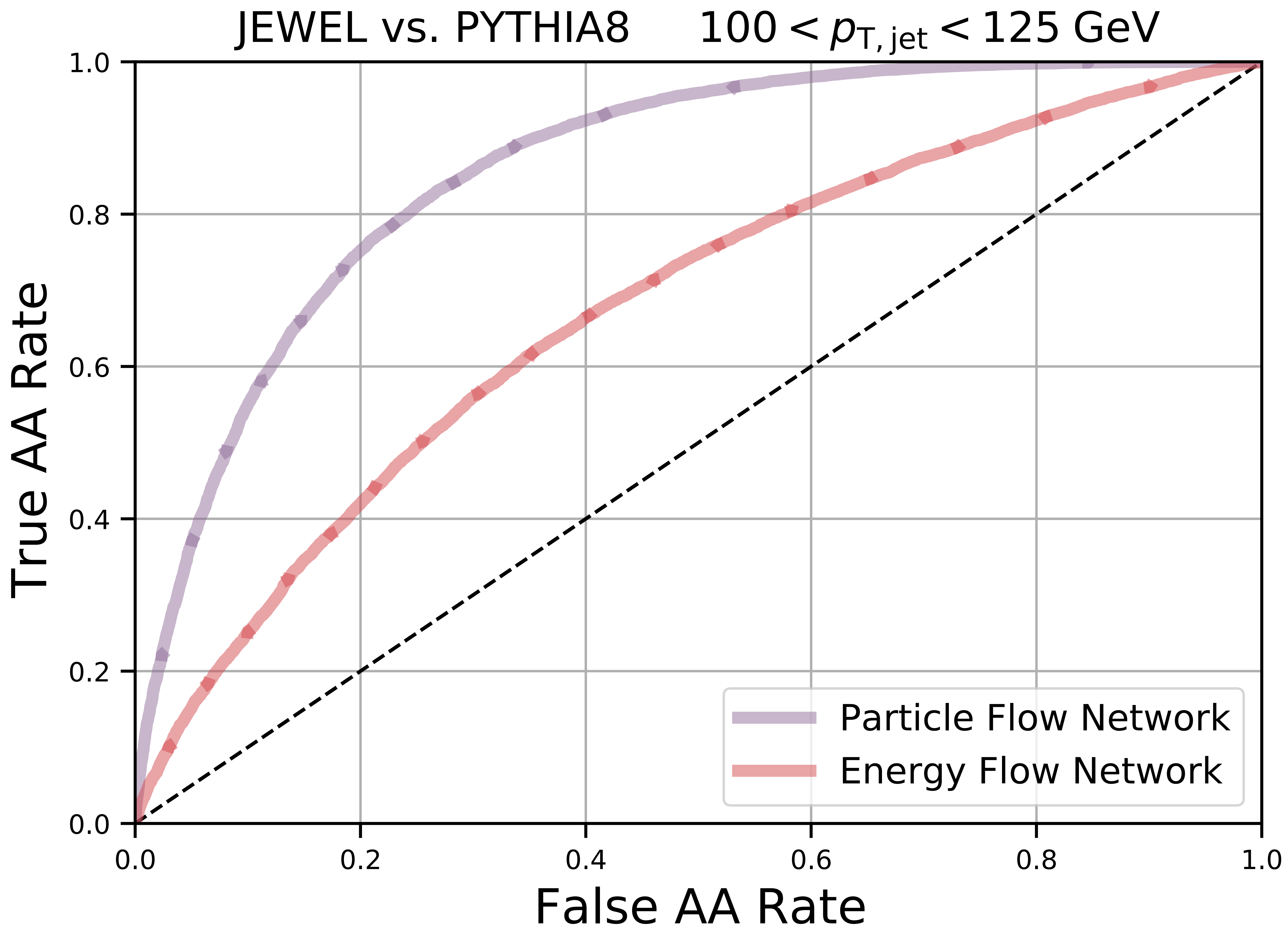
e.g. $$\lambda_{\alpha>0}^{\kappa=1} = \sum_{i \in \text{jet}} z_i \theta_i^\alpha$$



Insensitive to soft/collinear emissions

IRC-safe vs. IRC-unsafe physics

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



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

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

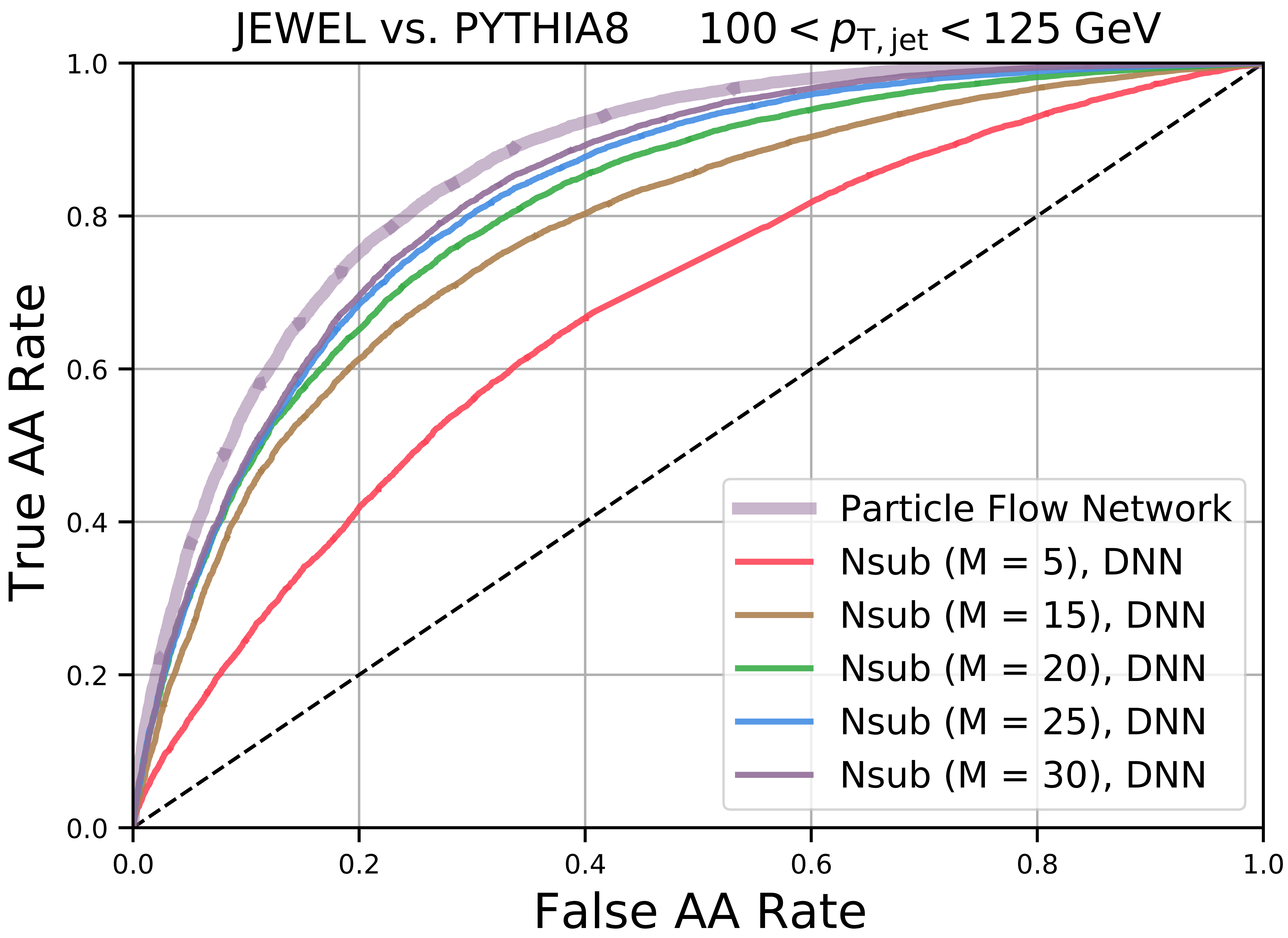
Classifier

DNNs

IRC-unsafe information contains significant discriminating power

Hard vs. soft physics

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



How many observables does one need to measure to saturate information?

DNN with $3M-4$ N-subjettiness basis observables as input:

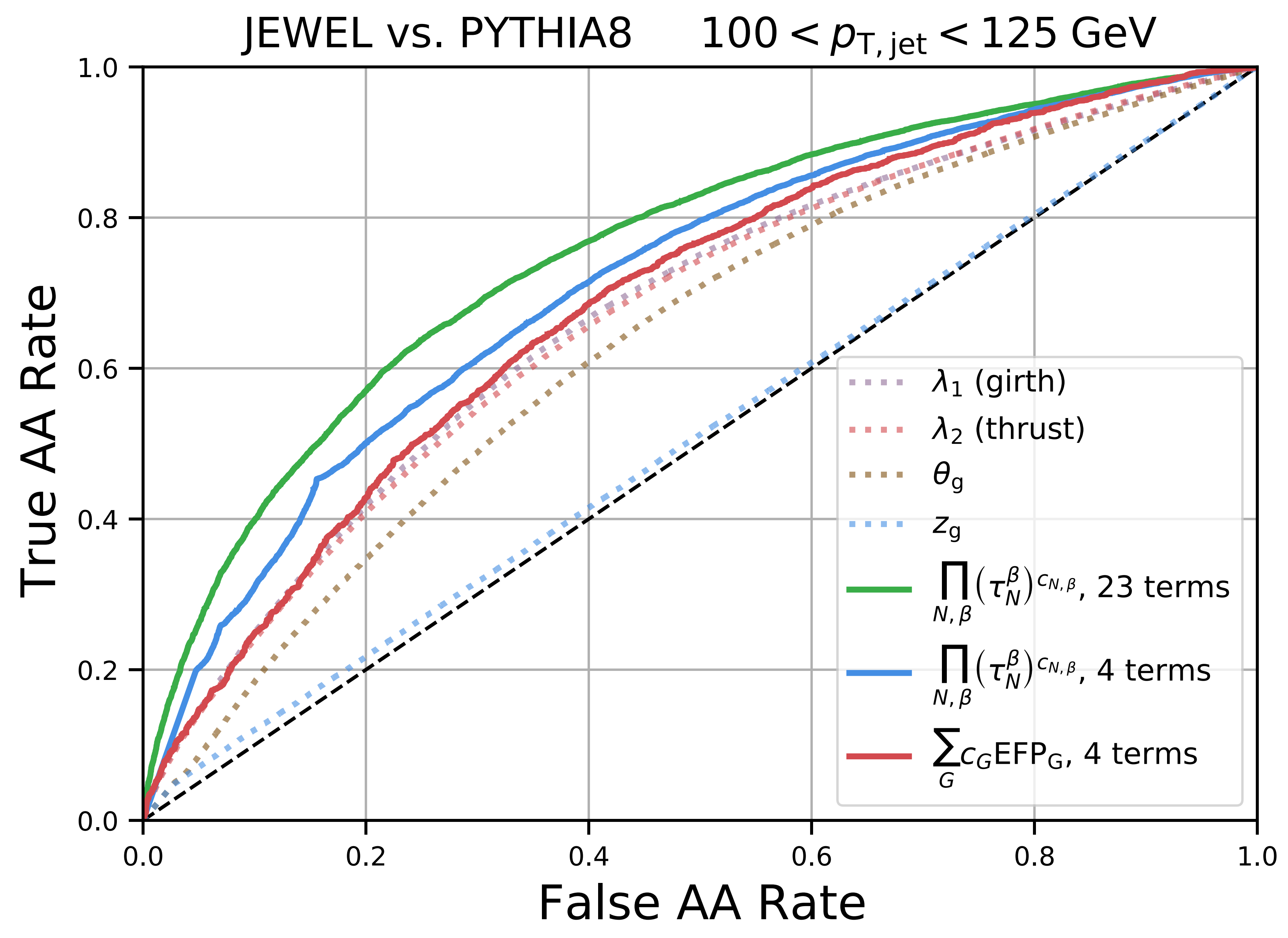
$$\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \ldots, \tau_{M-2}^{(0.5)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(0.5)}, \tau_{M-1}^{(1)}\right\}$$

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

Observable design

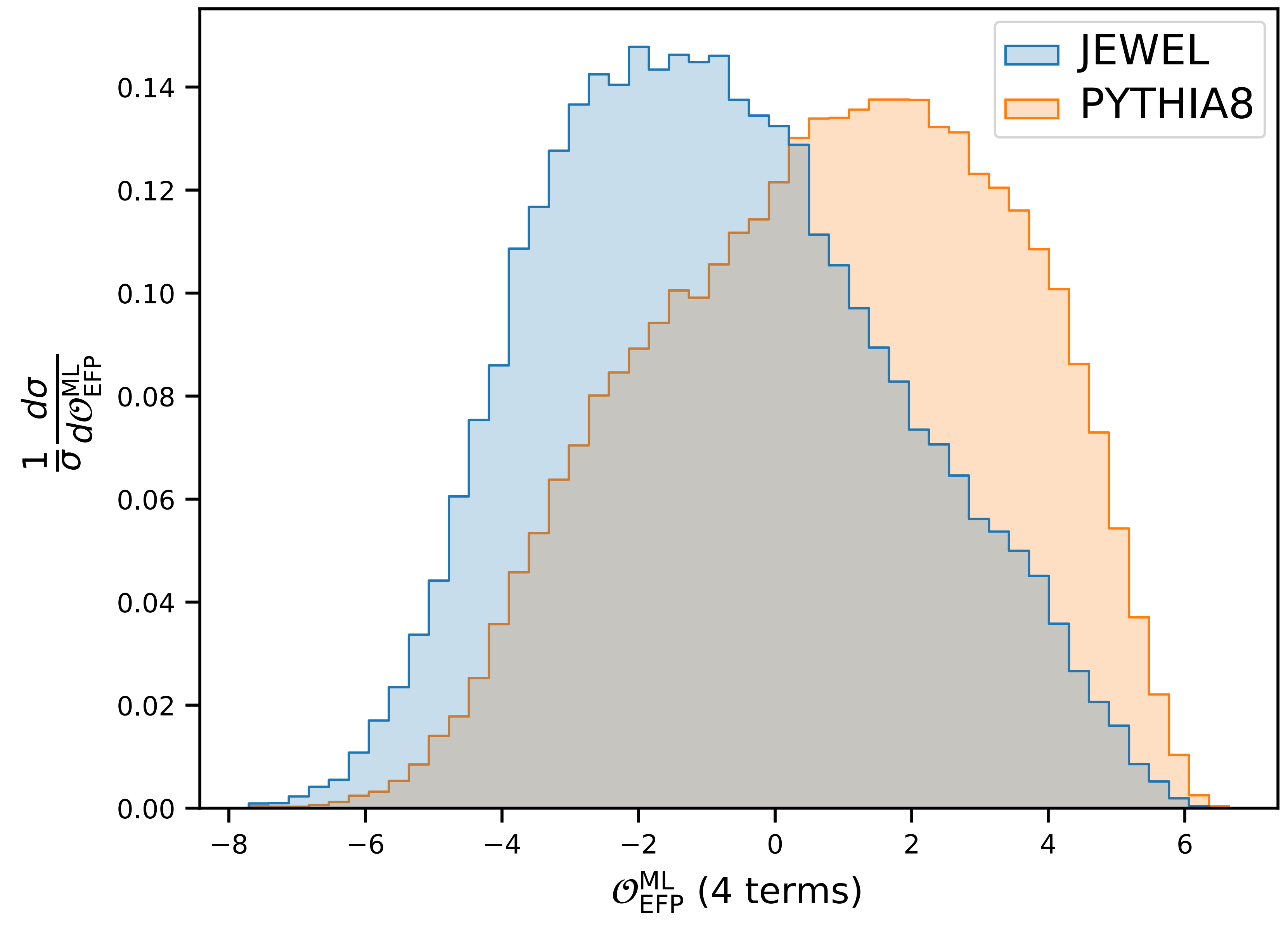
Lai, Mulligan, Płoskoń, Ringer JHEP 10 (2022) 011

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

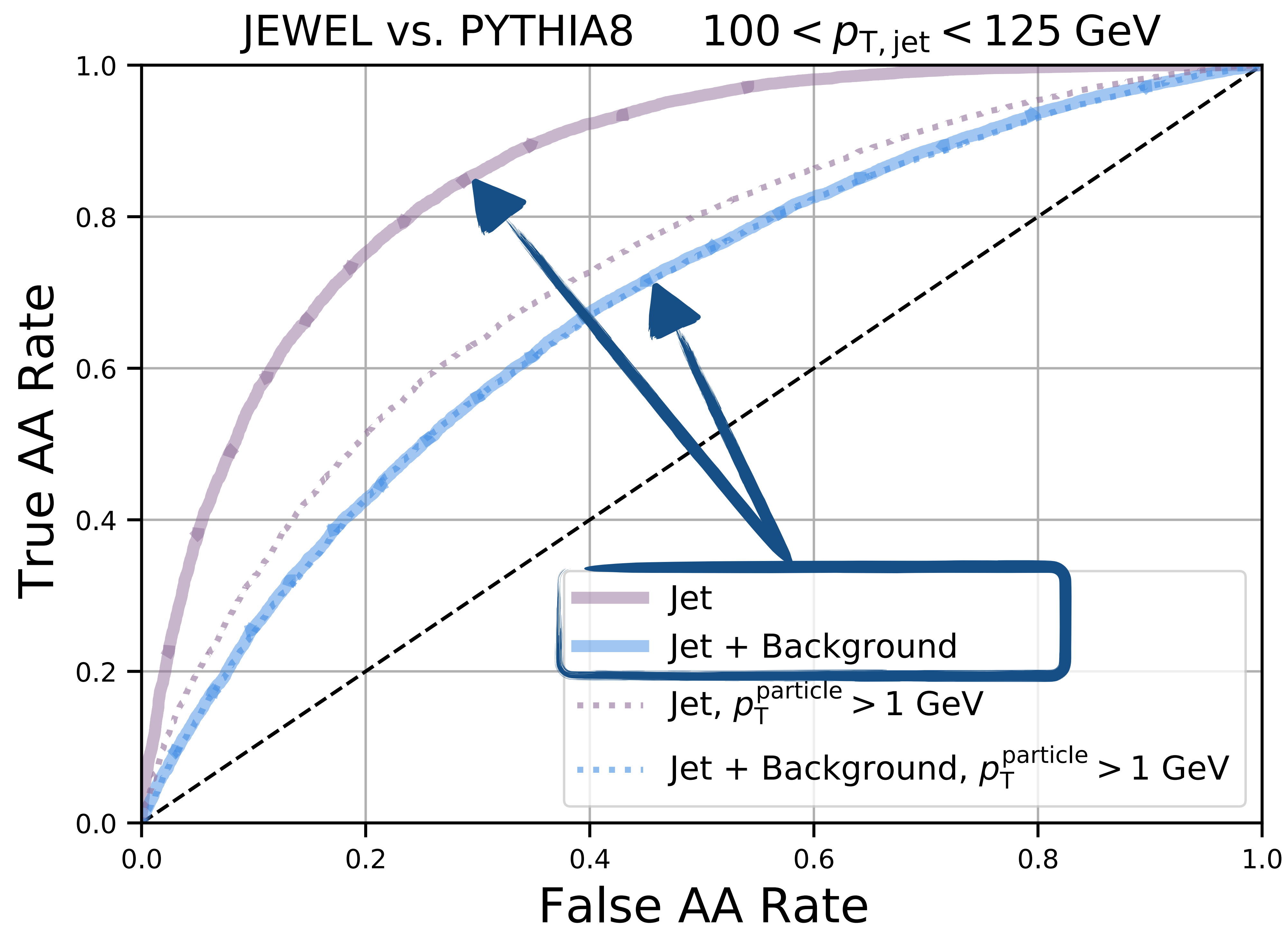


ML-assisted observable design provides guidance to experiments and theory — can then measure and calculate designed observables using traditional methods

Information loss due to background

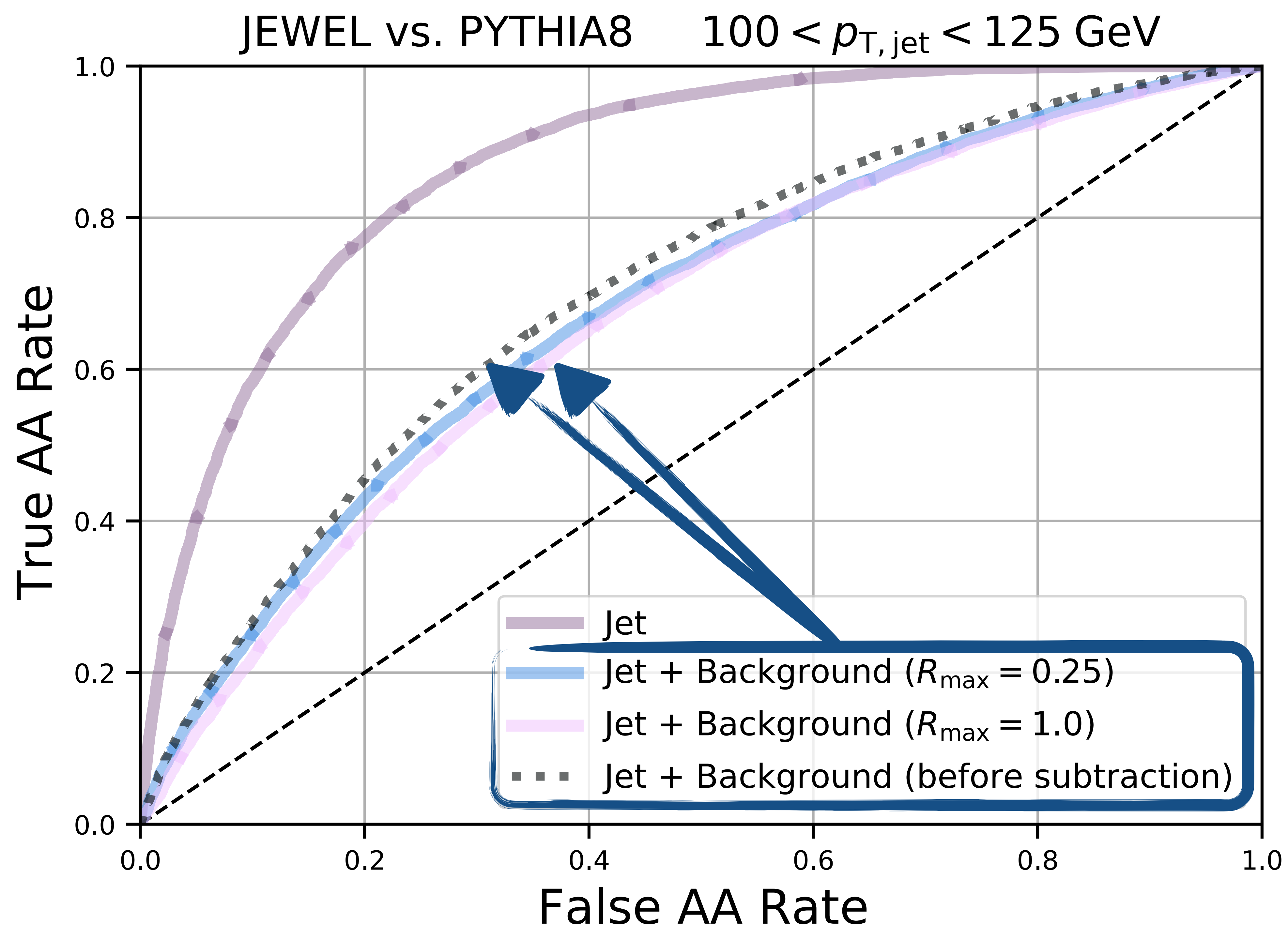
Lai, Mulligan, Płoskoń, Ringer JHEP 10 (2022) 011

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