ANATOMY OF "JET CLASSIFICATION USING DEEP LEARNING"

Mihoko Nojiri (KEK & KIPMU)

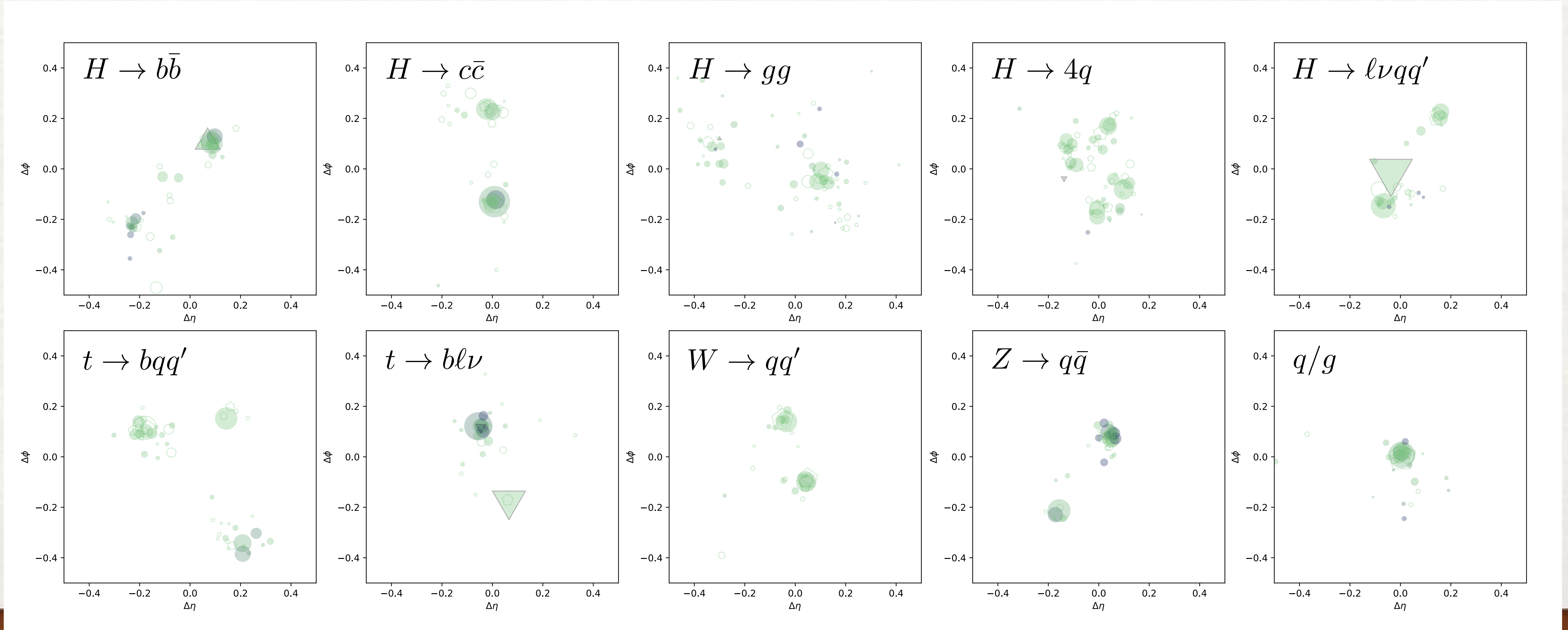
partly on work in collablation with Sung Hak Lim(Rutgers U)
Amon Furuichi(Nagoya U. and KEK) in preparation

This talk

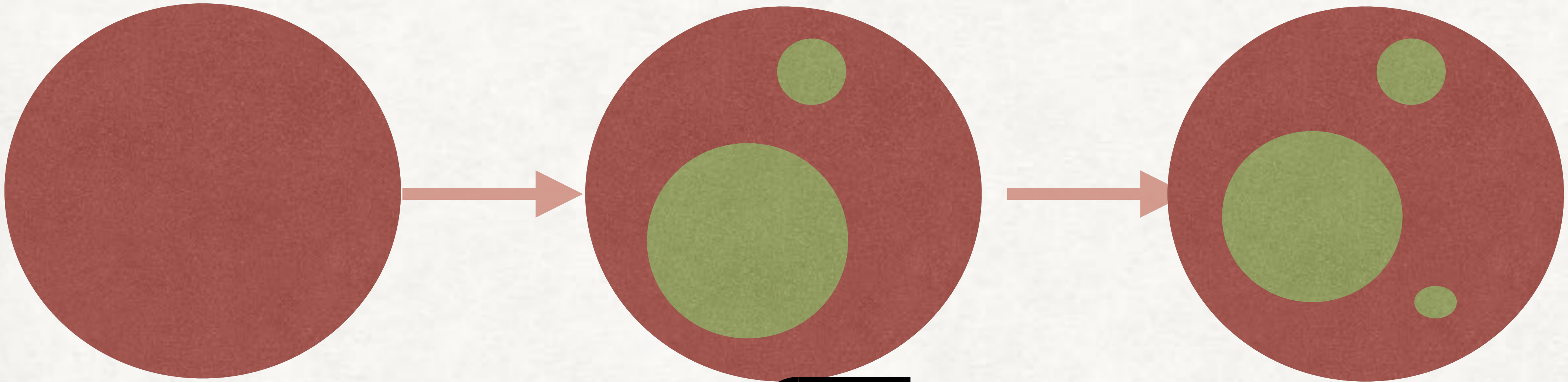
- Introduction History of jet physics (Triumph of QCD)
- Jet and Deep Learning (QCD theory → ML)
- (My) Skepticism and ANATOMY(ML meets soft physics)
- Toward Improved event simulations

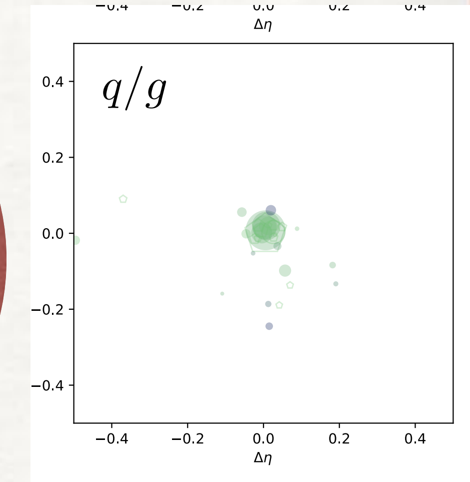
JET PHYSICS FOR BSM

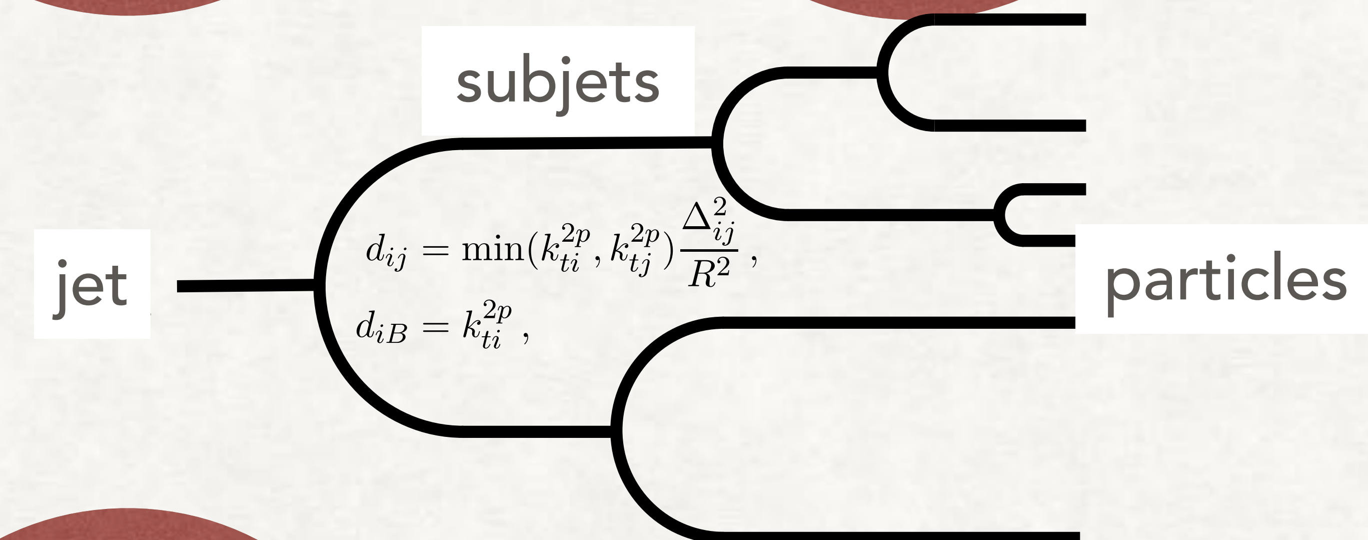
- LHC: boosted Higgs, boosted top for
  - heavy resonance search
  - SMEFT (high PT higgs boson, W, and Z distribution will be affected.)
- boosted objects look like a jet.   "jet substructure" is important to distinguish it from QCD jets

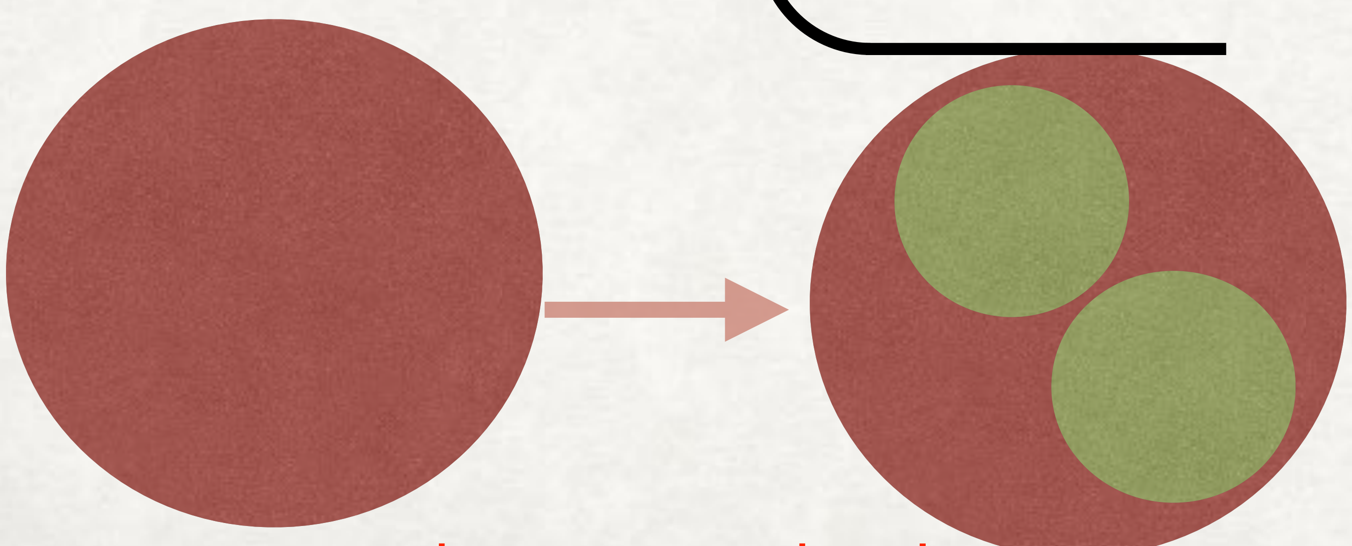


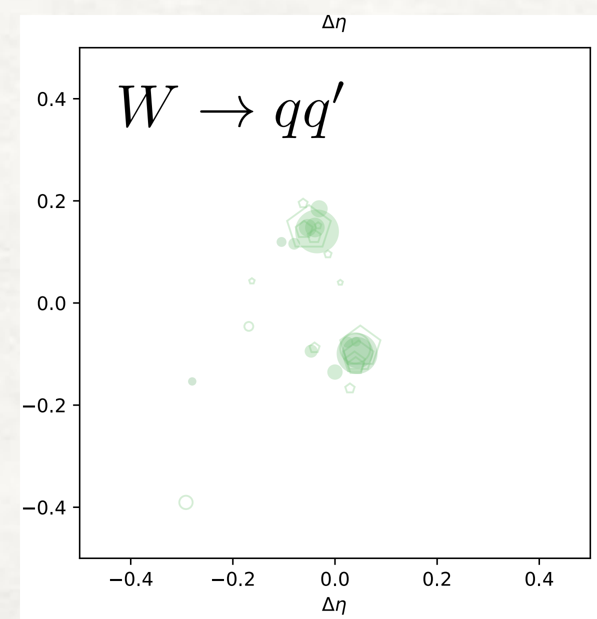
JET AND HEAVY OBJECT SEARCH
(BUTTERWORTH ET AL 2008)











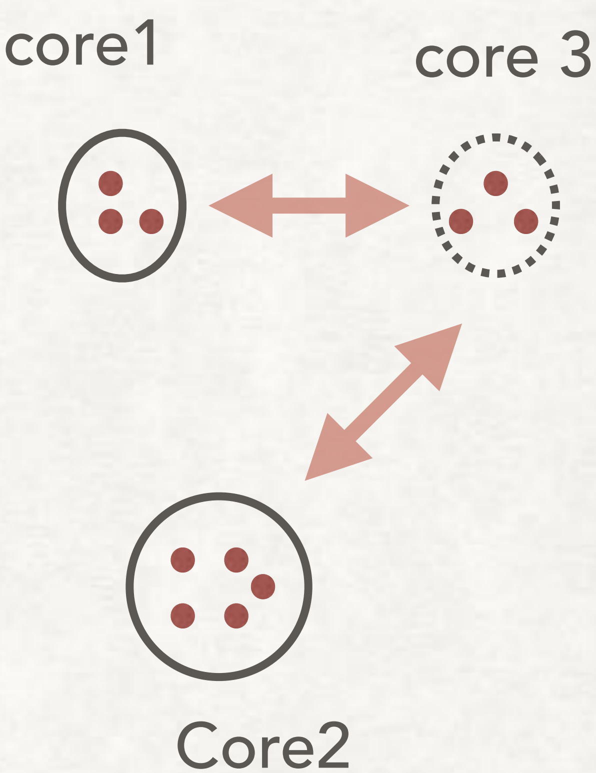
Make a cut on the cluster sequence backward (Butterworth et al 2008)

SEEDLESS IRC SAFE VARIABLES

- **n-subjettiness (2010 Thaler Tilburg)**

  minimize the distance to N axes

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \mathrm{Jet}} p_{Ti} \min\left\{R_{1i}^{\beta}, R_{2i}^{\beta}, \ldots, R_{Ni}^{\beta}\right\} .$$



- Energy Flow Polynomial(Komiske et al 1712.07124)

$$EFP_G = \sum_{i_1}^{M} \cdots \sum_{i_N}^{M} \ldots z_{i_1} \ldots z_{i_N} \prod_{k,l \in G} \theta_{i_k i_l}$$

ex $EFP_2^{\beta} = \sum_{i,j} z_i z_j \theta_{ij}$, $\theta_{ij} = [(y_i - y_j)^2 + (\phi_i - \phi_j)]^{\beta/2}$

  - linear in $z_i = E_i/E_J$ for all particle involved ← IRC safe (stable against soft and infrared divergence of QCD)

JET AND DEEP LEARNING

Deep learning and classifier $\Phi(x, \ldots)$

Multilayer perceptron(MLP)

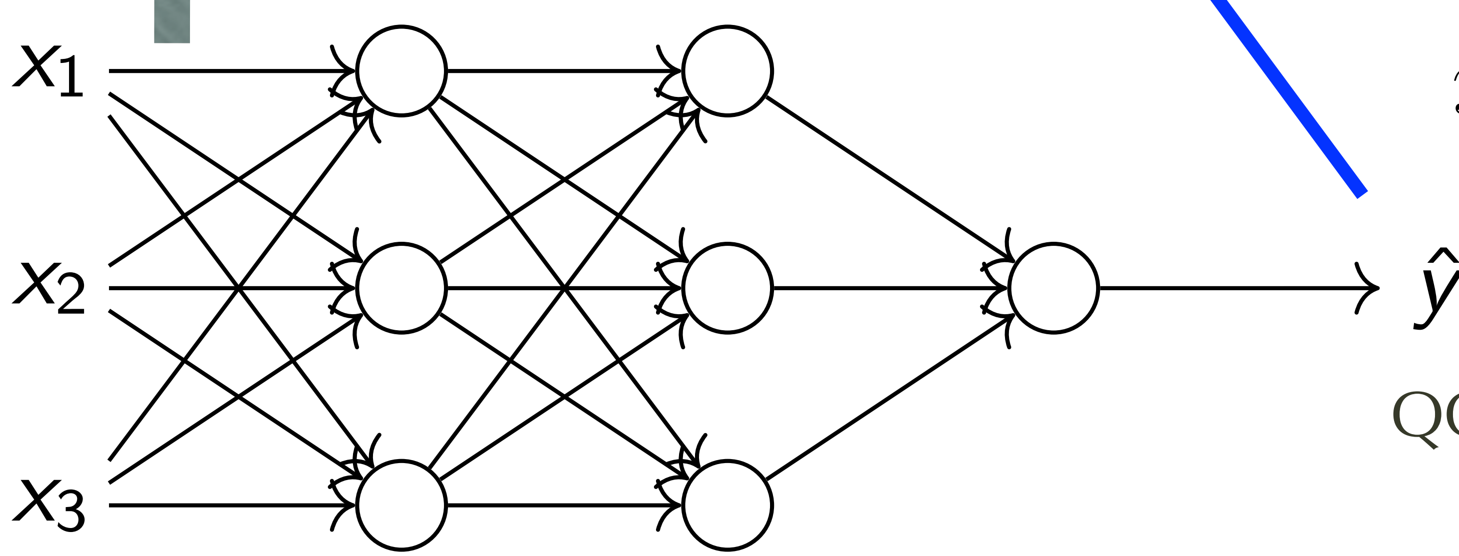
model parameter : $\mathbf{w_{ij}}, \mathbf{b_i}$

optimization $L(y, \hat{y})$

(gradient flow)

jet image or particles
+ further arch

High level
inputs
$j_i, \tau_i^{(\beta)}, EFP\ldots$



boosted jet
$y = (1, 0)$

QCD
$y = (0, 1)$

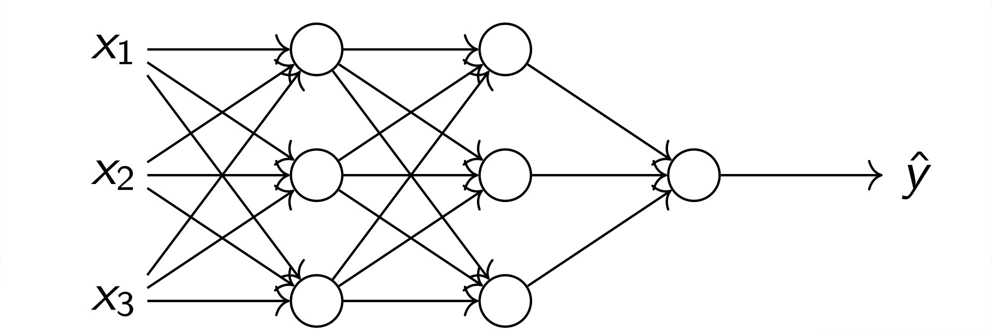
$$x_i^{(n+1)} = \varphi(w_{ij}^{(n)} x_j^{(n)} + b_i^{(n)})$$

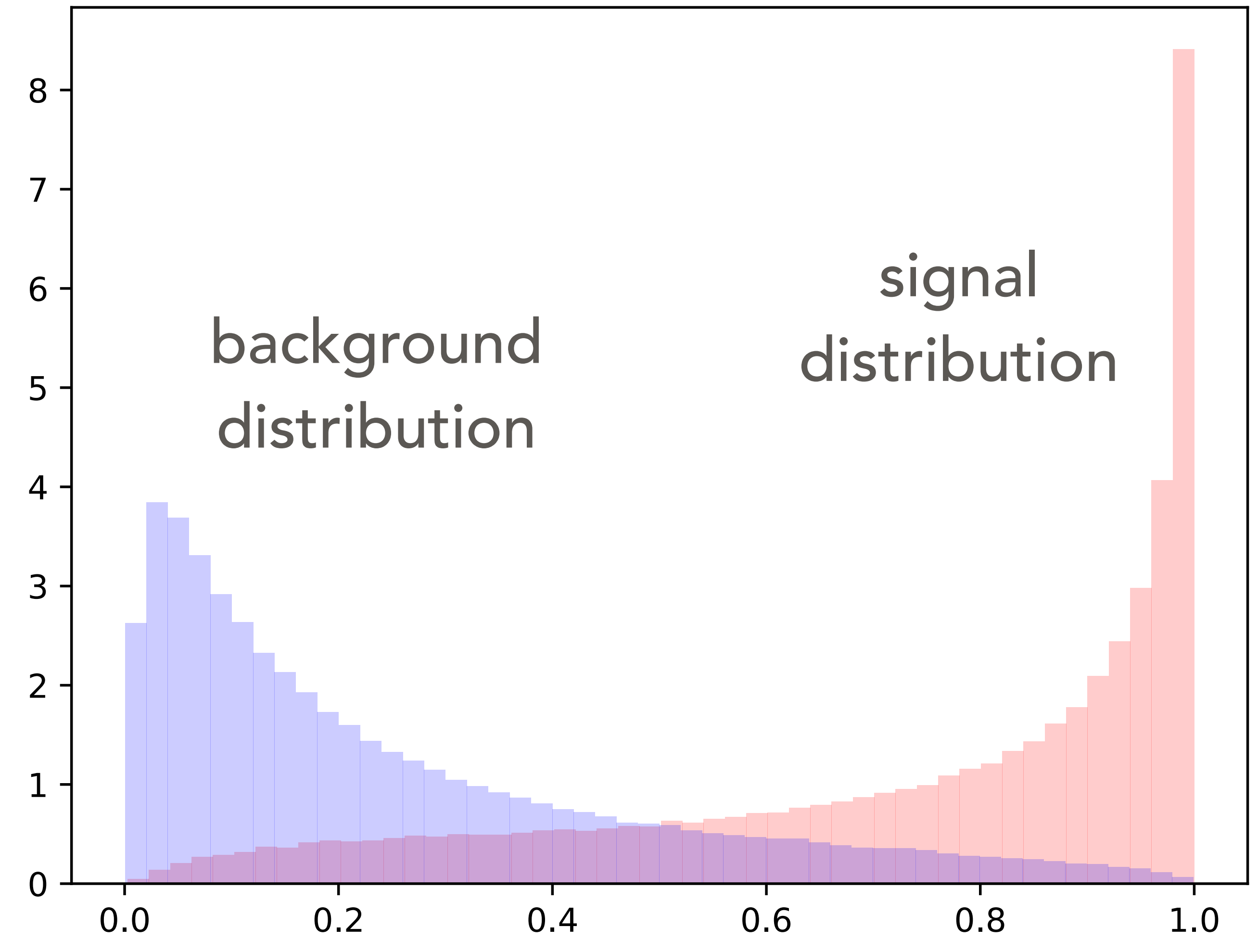
φ: activation function
source of non linearity

High "Representative power"

Event classifier $\Phi(x_i)$

x_i : event information





background like signal like

QUICK TOUR OF JET CLASSIFICATION MODELS

A. **Use Motivated** (safe) **input**

B. Use Most Fancy DL network at Time (many choices)

CNN(2017 Kasieczka et al)→**Particle Net** (Graph NN) (2019)

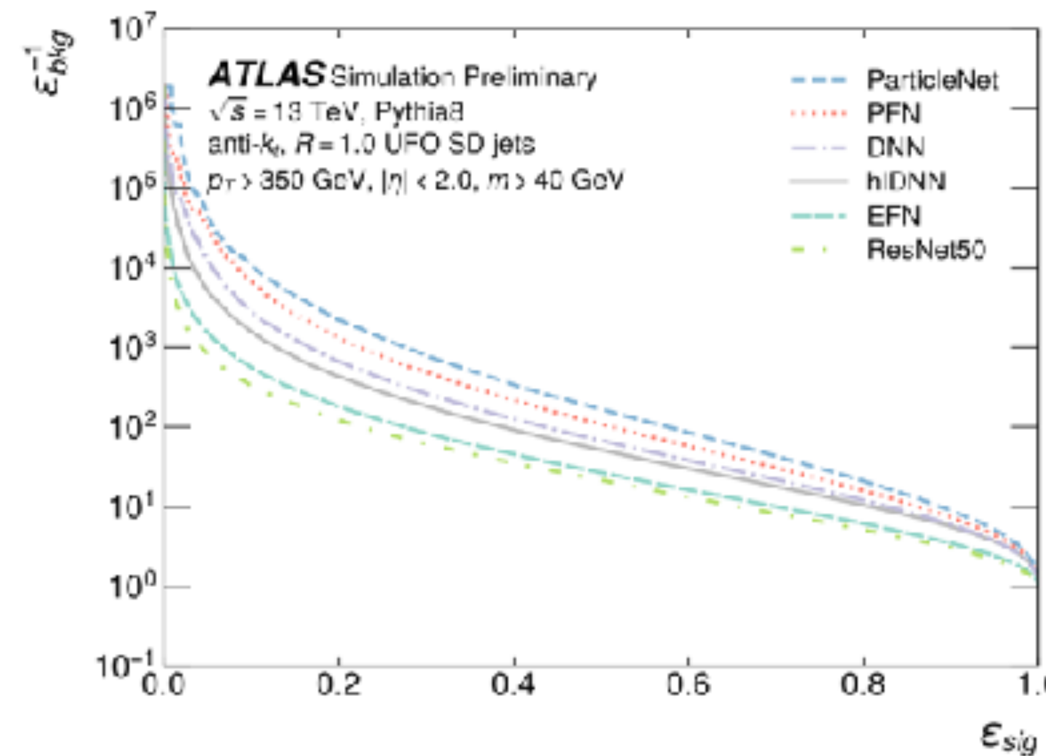
→**Particle Transformer**(Graph and Attention) (2022)

Large GAP between A & B →origin of the difference?

“low level" input improve classification

- e.g. calorimeter-cluster particle-flow objects observables “constituent-based tagging”
- Graph NN（ParticleNet）improve background rejection significantly
- **ML beats theory?**

rejection efficiency plot



	All classes		$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to 4q$	$H \to \ell\nu qq'$	$t \to bqq'$	t
	Accuracy	AUC	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{99\%}$	$\text{Rej}_{50\%}$	R
PFN	0.772	0.9714	2924	841	75	198	265	797	
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	
ParT	**0.861**	**0.9877**	**10638**	**4149**	**123**	**1864**	**5479**	**32787**	
ParT (plain)	0.849	0.9859	9569	2911	112	1185	3868	17699	

Qu et al 2022.03772

We cannot ignore such large gain.

TYPE A :ENERGY FLOW NETWORK (IRC SAFE)

(1810.05165 KOMISKE, METODIEV, THALER)

Deepset (permutation invariant, work for any number of constituents)

add over all constituents (work with any M)

Manifestly IRC safe set up

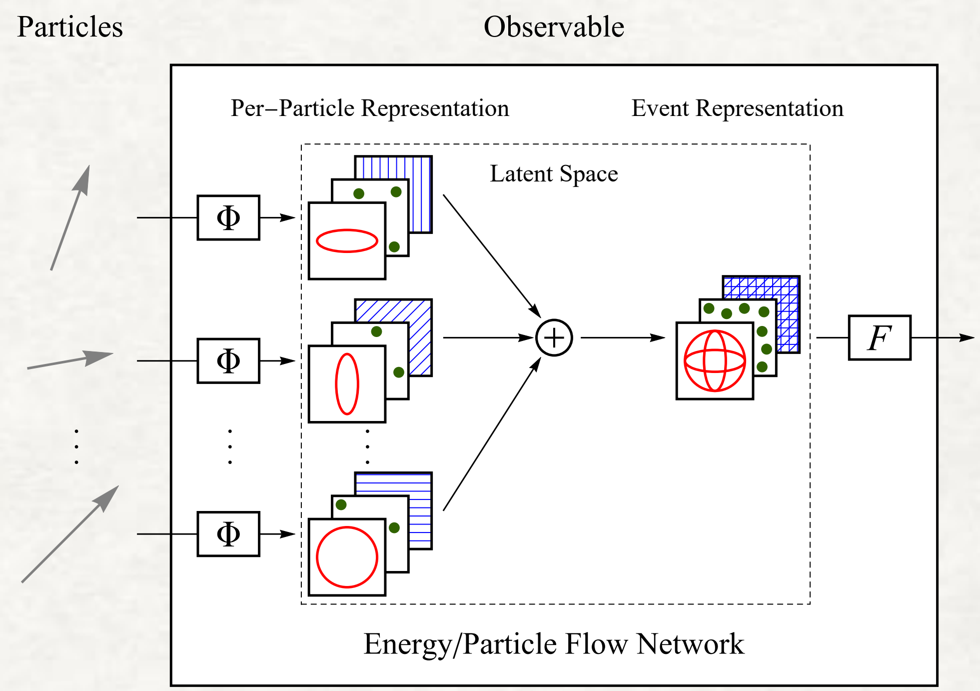
We learn this!

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

Weight by energy

$$z_i = E_i / \sum_j E_j \text{ or } z_i = p_{T,i} / \sum_j p_{T,j}$$

Limited to the correlation relative to leading jet

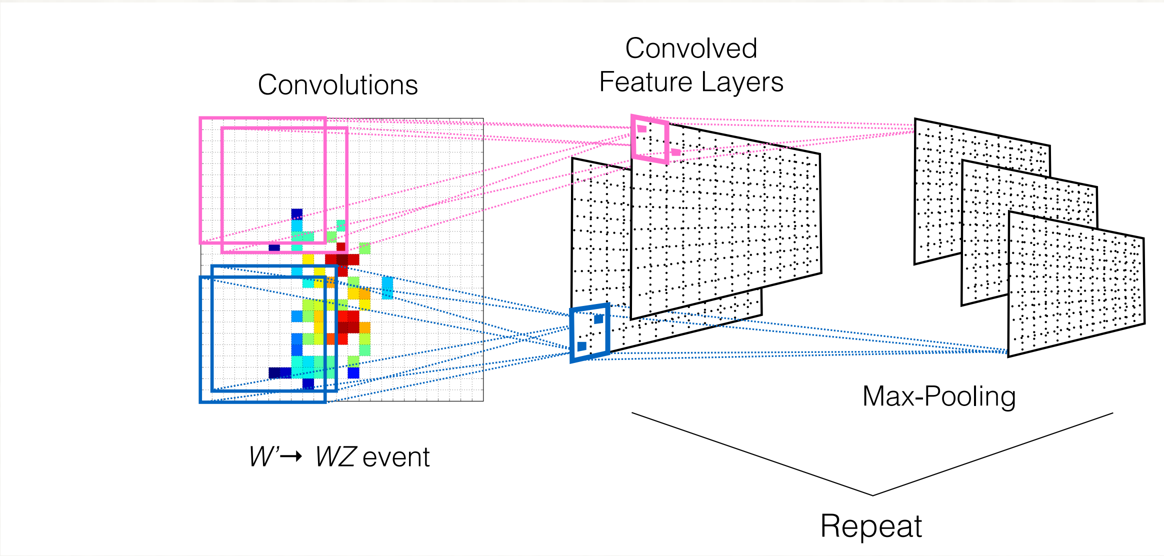


IRC unsafe

$$\text{PFN:} \quad F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$

TYPE B: CNN AND GRAPHES

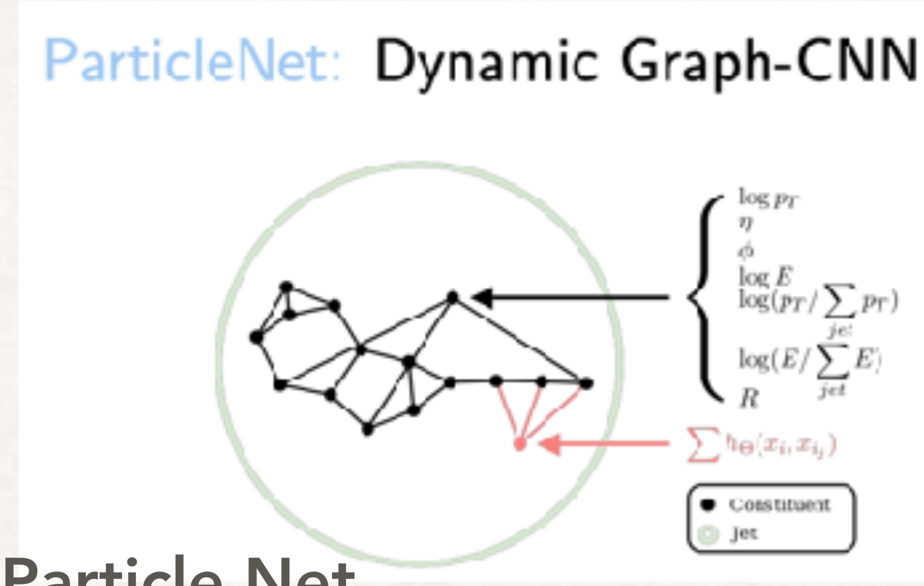
CNN



- Transfer image by KxK filter → cutoff (pooling) to find correlation.
- Performance: CNN> EFN
- Demerit "point by point" fluctuation near the boundary of the jets.

(Now it is independent to QCD etc)

GNN>> CNN



Particle Net

Input: vertex (particle information) N

edge (two point correlation) N^2

Calculate edge variables of nearby 7 vertex →update vertex and edge-> use this to select next 7 pairs

Particle Transformar

Attention Network, Aggregation

$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d_k} + \mathbf{U})V,$$

U: interaction N^2 but deep set like

SKEPTICISM AND ANATOMY

The Algorithm respect jet clustering

(2004.03540 PRL 124 222002)

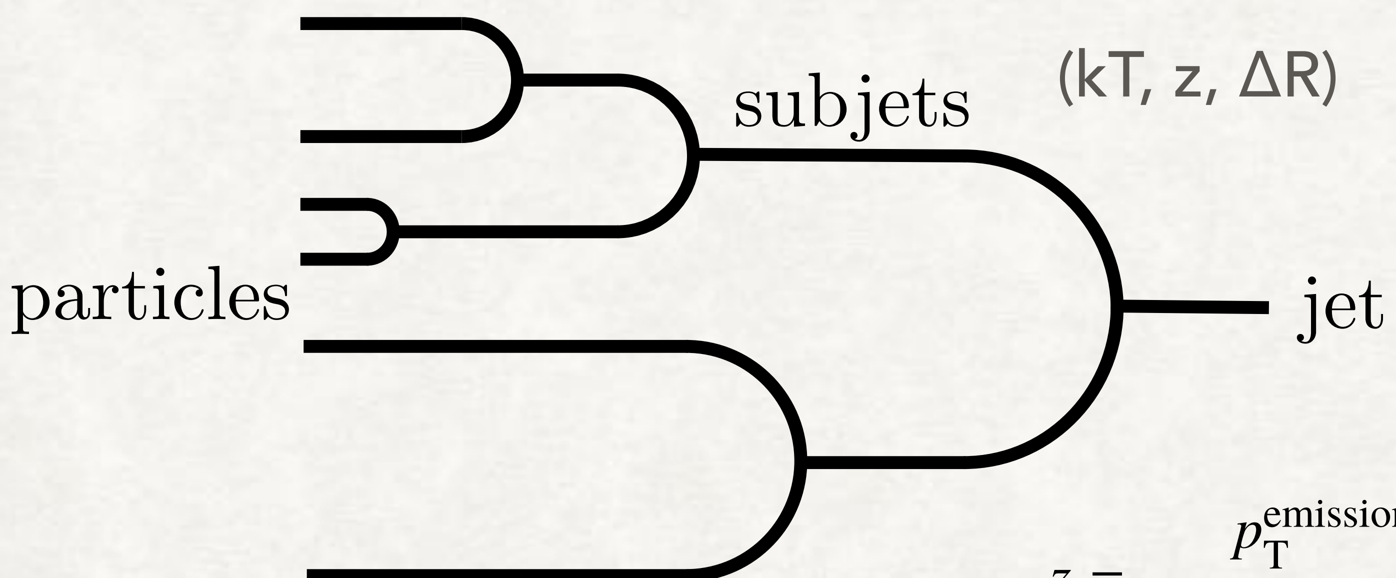
LUND JET PLANE
DREYER, SALAM, SOYEZ (1807.04758)

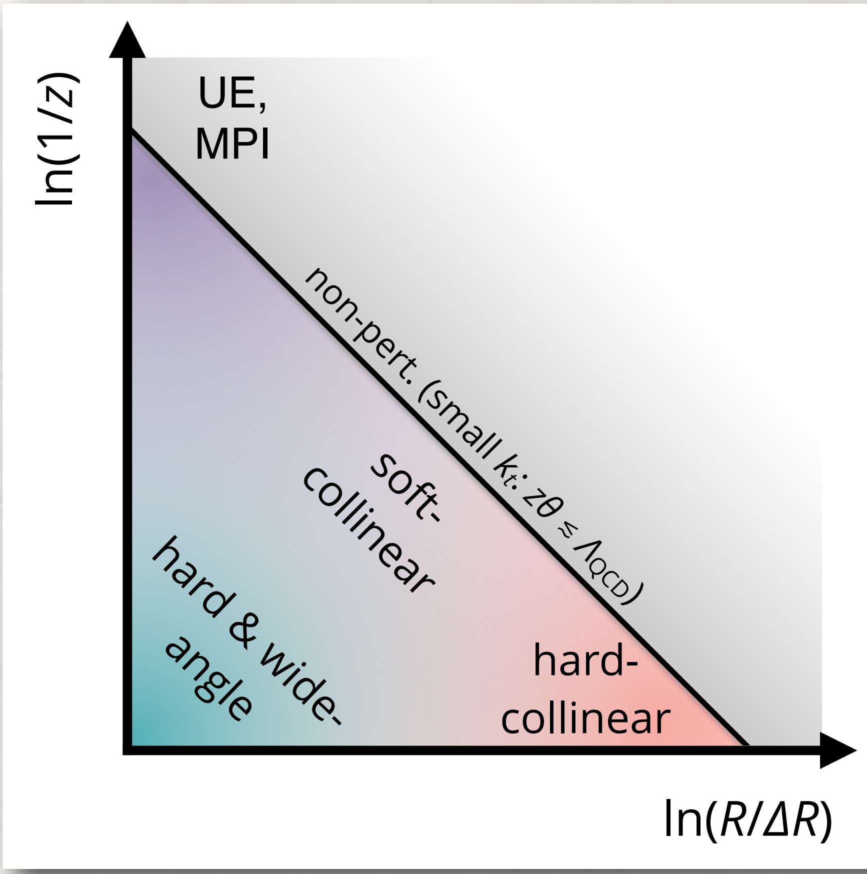
More reliable jet structure variable

Non-pert | soft collinear | hard collinear | hard and wide

(z', R')

(kT, z, ΔR)





(a) Schematic representation of the LJP.

$$z = \frac{p_T^{\text{emission}}}{p_T^{\text{emission}} + p_T^{\text{core}}}$$

LundNet(2012.08526 Dreyer and Qu)

The Algorithm respect jet clustering

(2004.03540 PRL 124 222002)

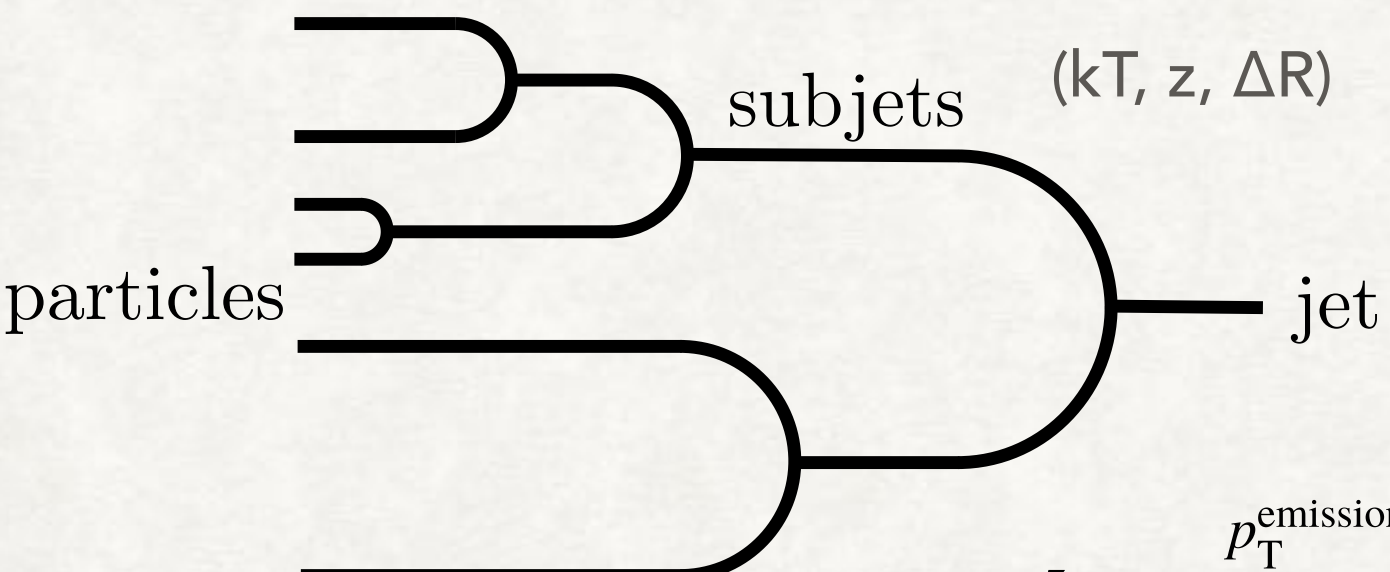
LUND JET PLANE
DREYER, SALAM, SOYEZ (1807.04758)

More reliable jet structure variable

Non-pert | soft collinear | hard collinear | hard and wide

(z', R')

(kT, z, ΔR)



$$z = \frac{p_{\mathrm{T}}^{\mathrm{emission}}}{p_{\mathrm{T}}^{\mathrm{emission}} + p_{\mathrm{T}}^{\mathrm{core}}}$$

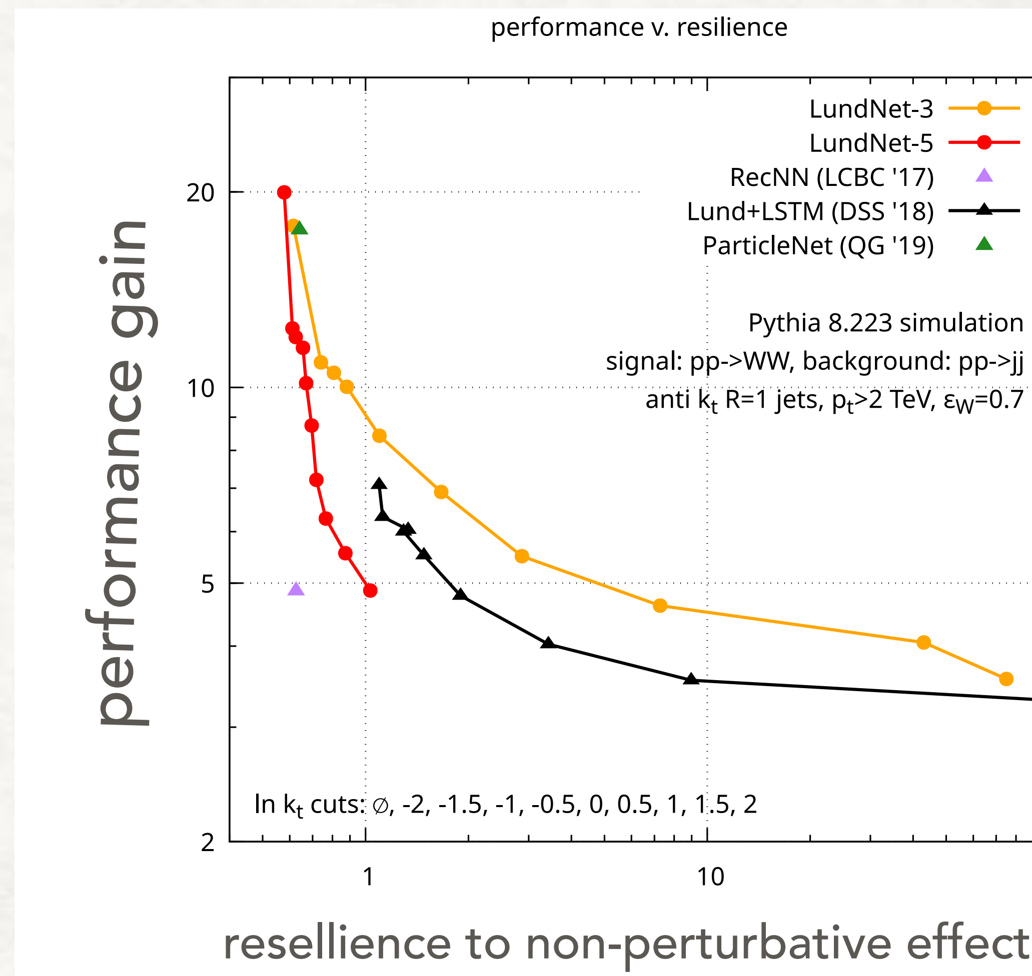


Figure 8. Performance $\frac{\epsilon_W}{\sqrt{\epsilon_{\mathrm{QCD}}}}$ versus resilience to non-perturbative effects.

LundNet(2012.08526 Dreyer and Qu)

EVENT GENERATOR DEPENDENCE IN PARTICLE TRANSFORMER

Furuichi Nojiri Lim

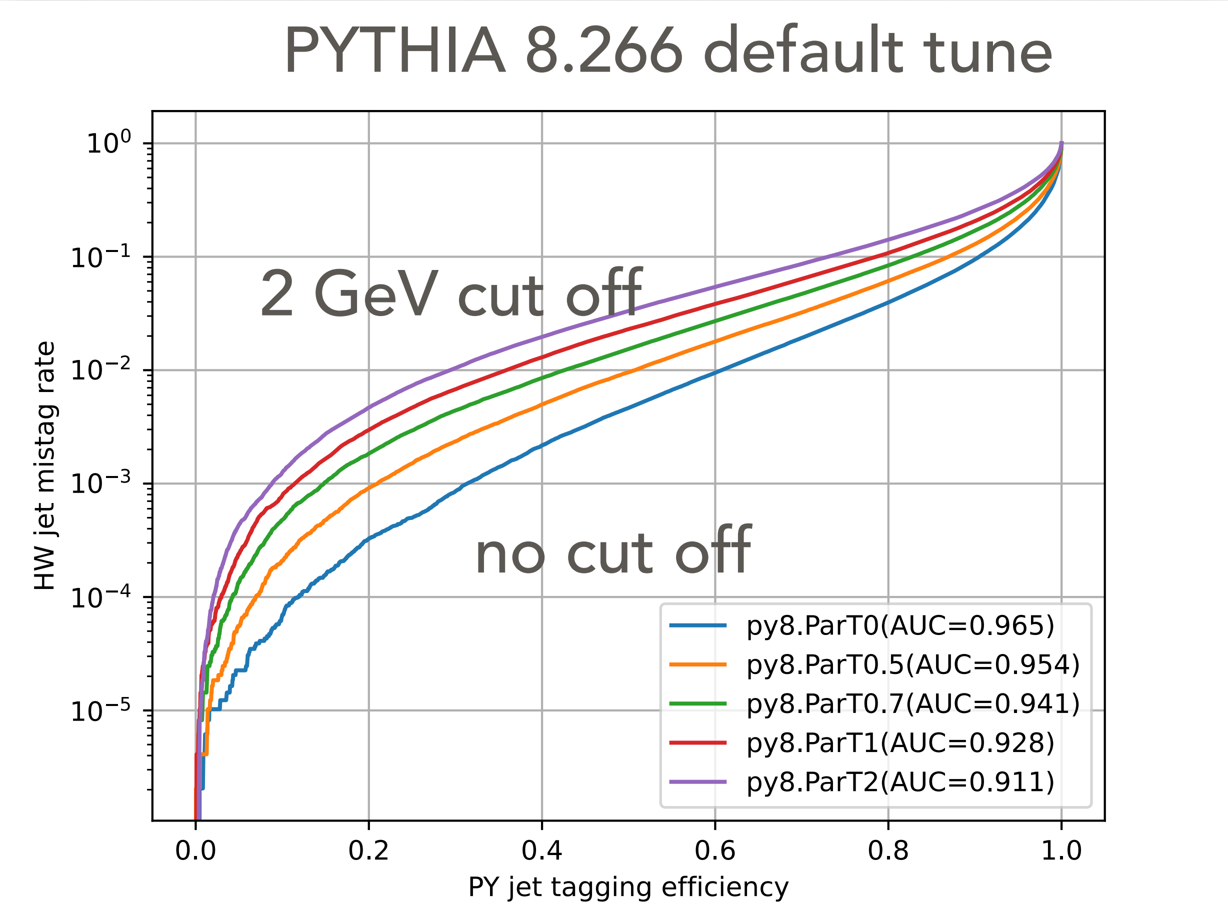
Sampel

500GeV <PT<600 GeV 150GeV <mJ<200GeV

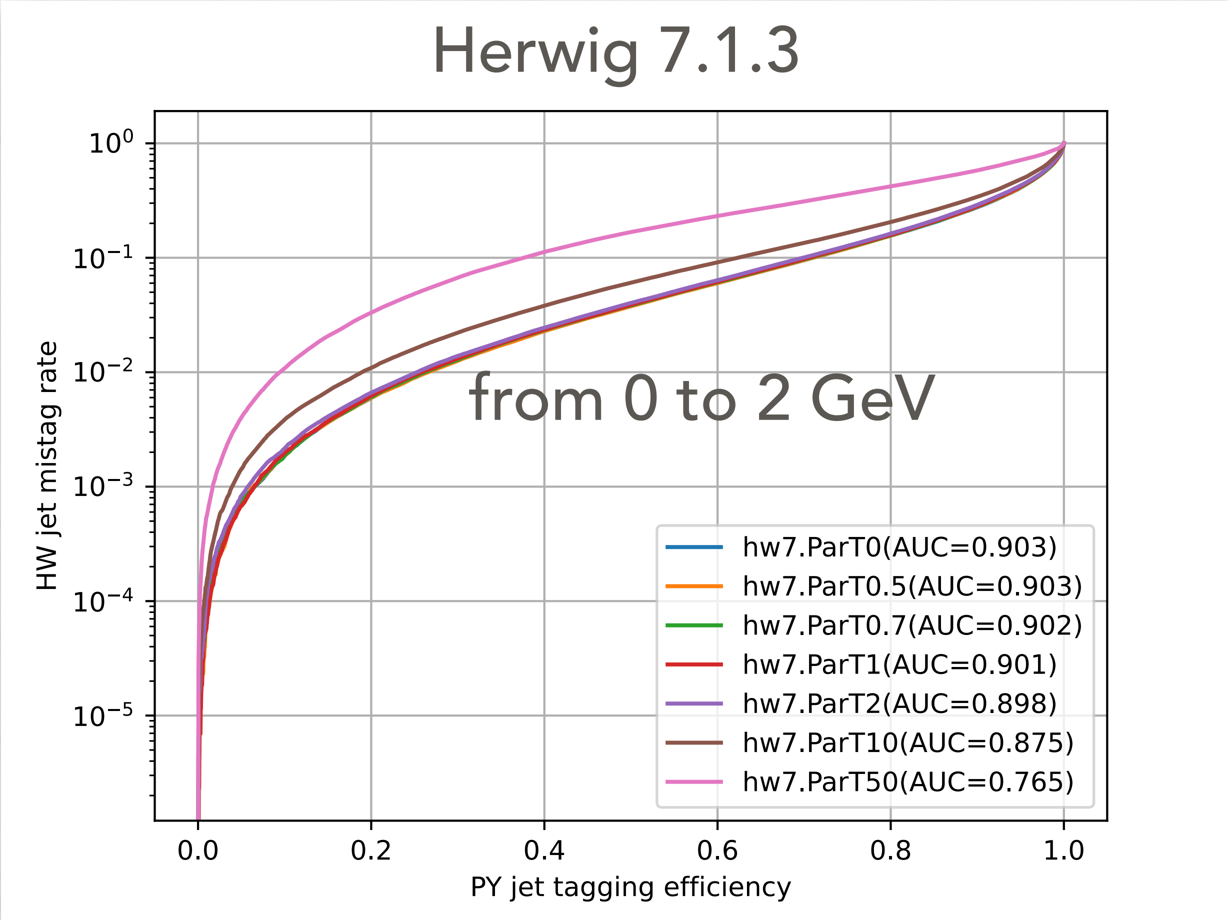
pixellated jet image $\Delta\eta, \Delta\phi = (0.1, 0.1)$ no track information

$$N_{QCD} = 0.35M \, , N_{top} = 1\mathrm{M}$$

PYTHIA event simulation



Herwig 7 simulation



(Probably the tune is a bit old)

EVENT GENERATOR DEPENDENCE IN PARTICLE TRANSFORMER

Furuichi Nojiri Lim

Sampel

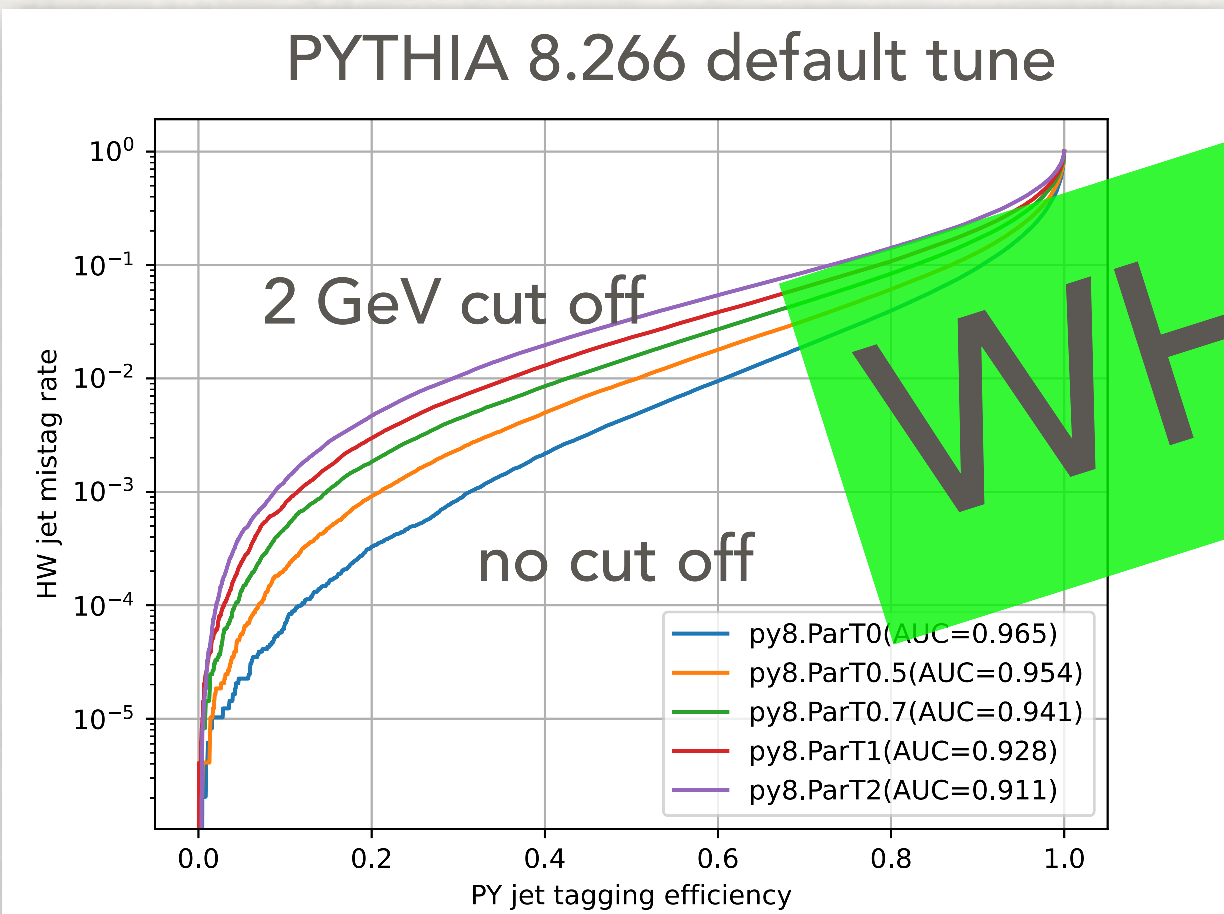
500GeV <PT<600 GeV 150GeV <mJ<200GeV

pixellated jet image $\Delta\eta, \Delta\phi = (0.1,0.1)$ no track information

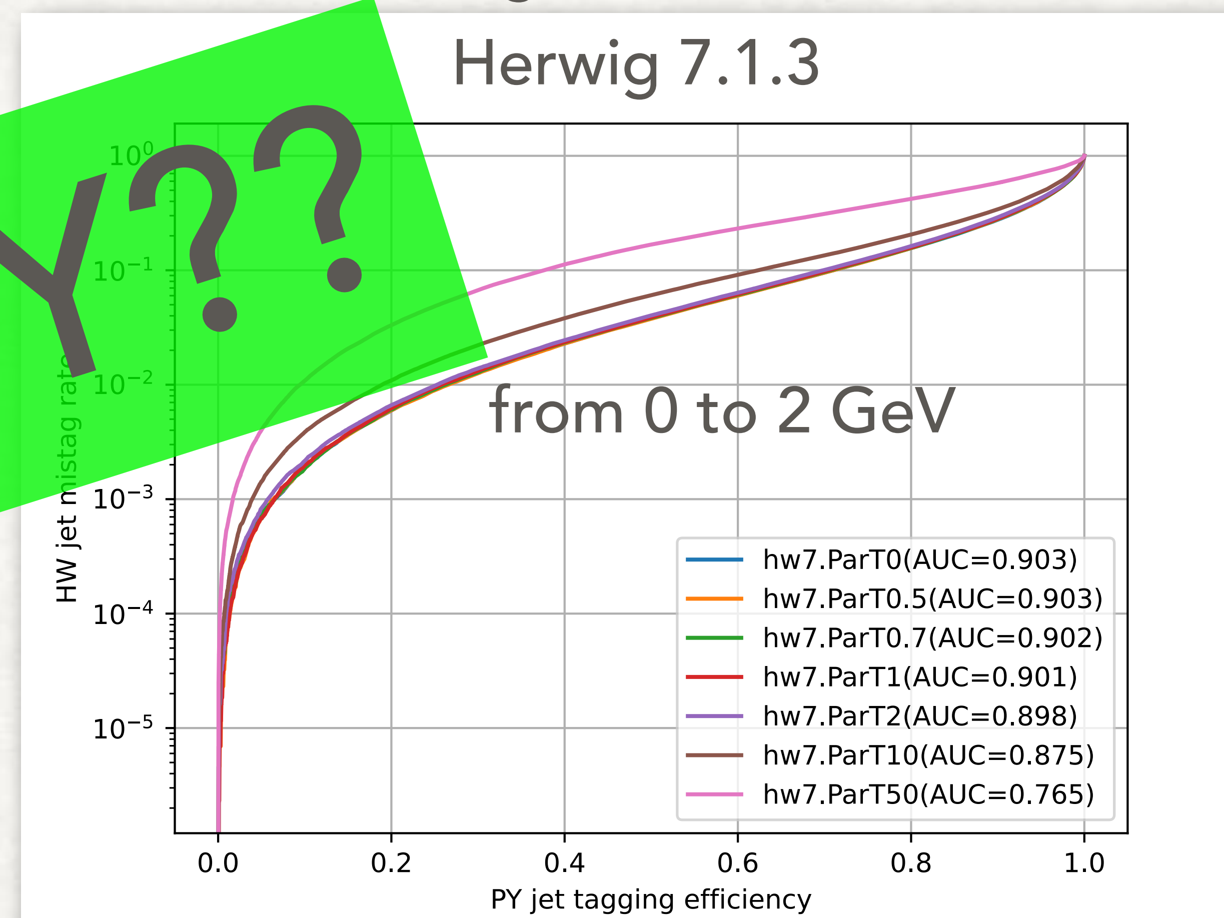
$$N_{QCD} = 0.35M \,,\, N_{top} = 1\text{M}$$

PYTHIA event simulation

Herwig 7 simulation

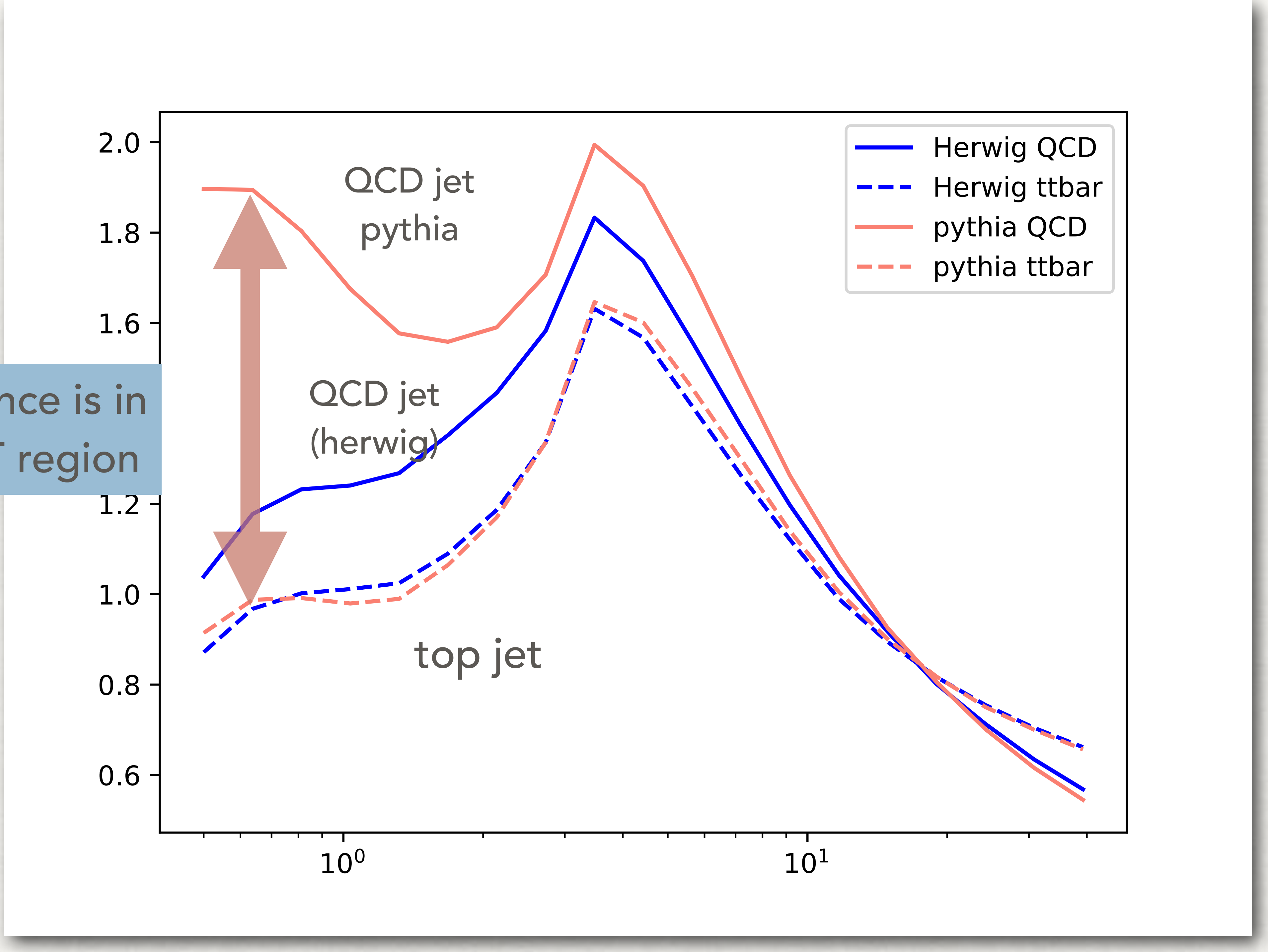


WHY??



(Probably the tune is a bit old)

PT DISTRIBUTION OF JET CONSTITUENTS



difference is in low pT region

Furuichi, Nojiri, Lim in preparation

SOFT PARTICLES SYSTEMATICS

Graph NNs are eager to learn the soft particle correlation if it is relevant to classification.

→Event Generator have to be modeled carefully toward low PT regions

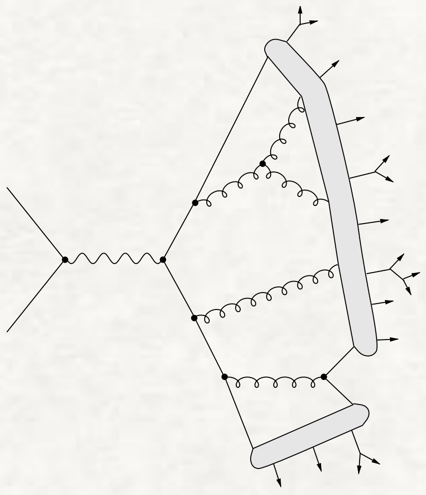
"SOFT PHYSICS MEETS ML "

- top vs QCD

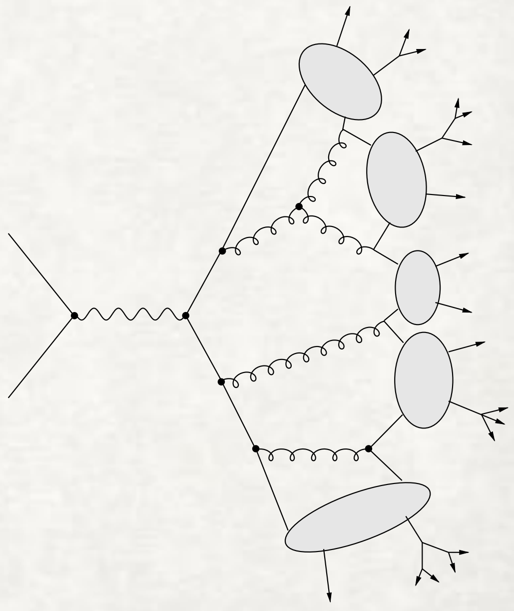
  Emission from top quark may be smaller

  Energy of decay product is small $E_q \ll E_t$

- hadronization modeling (ad hok model)
  - string model (PYTHIA)
  - cluster model (HERWIG Sherpa)



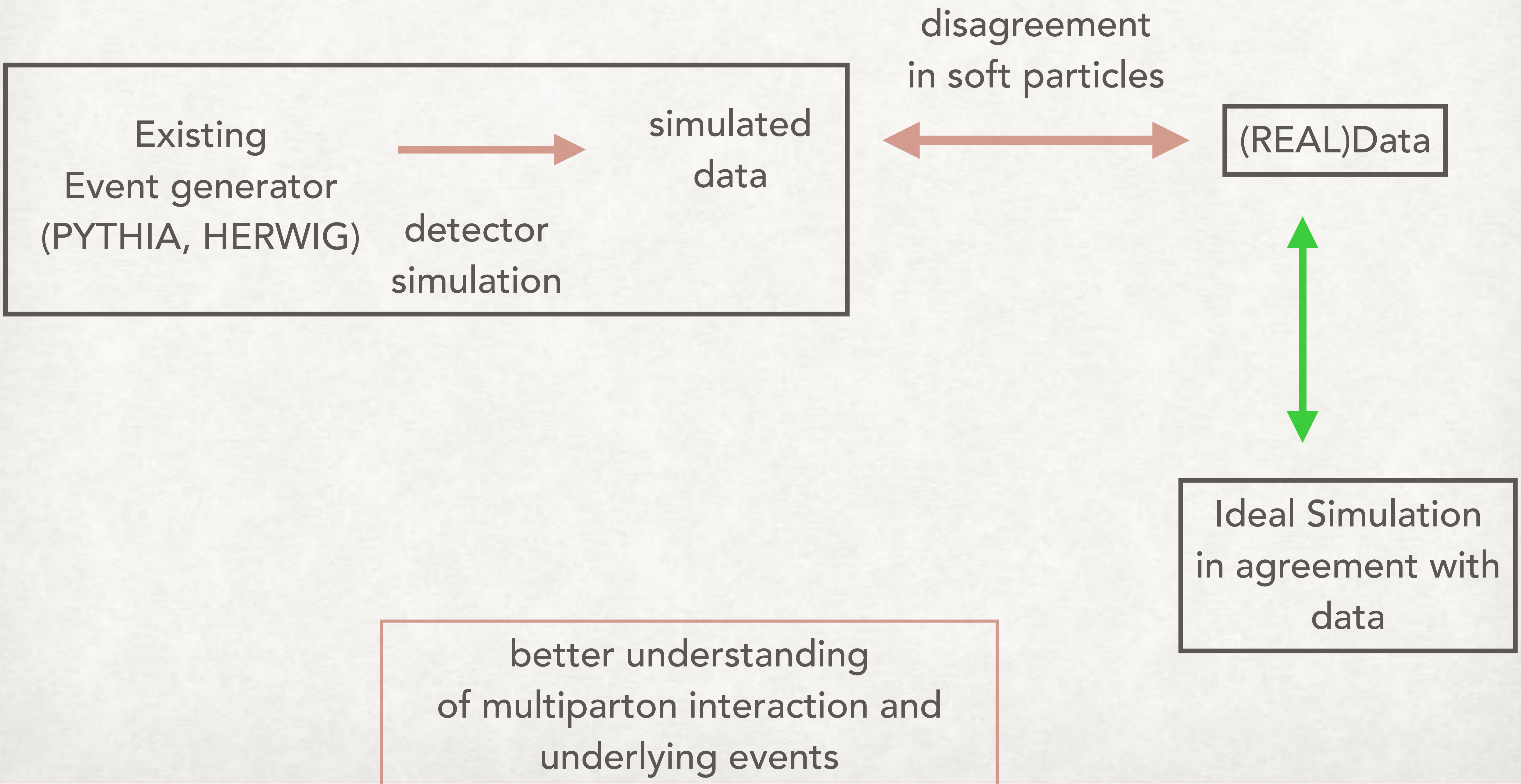
String model



Cluster model

HOW TO HANDLE DATA-MC DIFFERENCE

- Description of Low energy particle is empirical so we want to tune it by the data.



HOW TO HANDLE DATA-MC DIFFERENCE

- Description of Low energy particle is empirical so we want to tune it by the data.

Existing
Event generator
(PYTHIA, HERWIG)

detector
simulation

simulated
data

disagreement
in soft particles

classifier
$y = \Phi(x)$

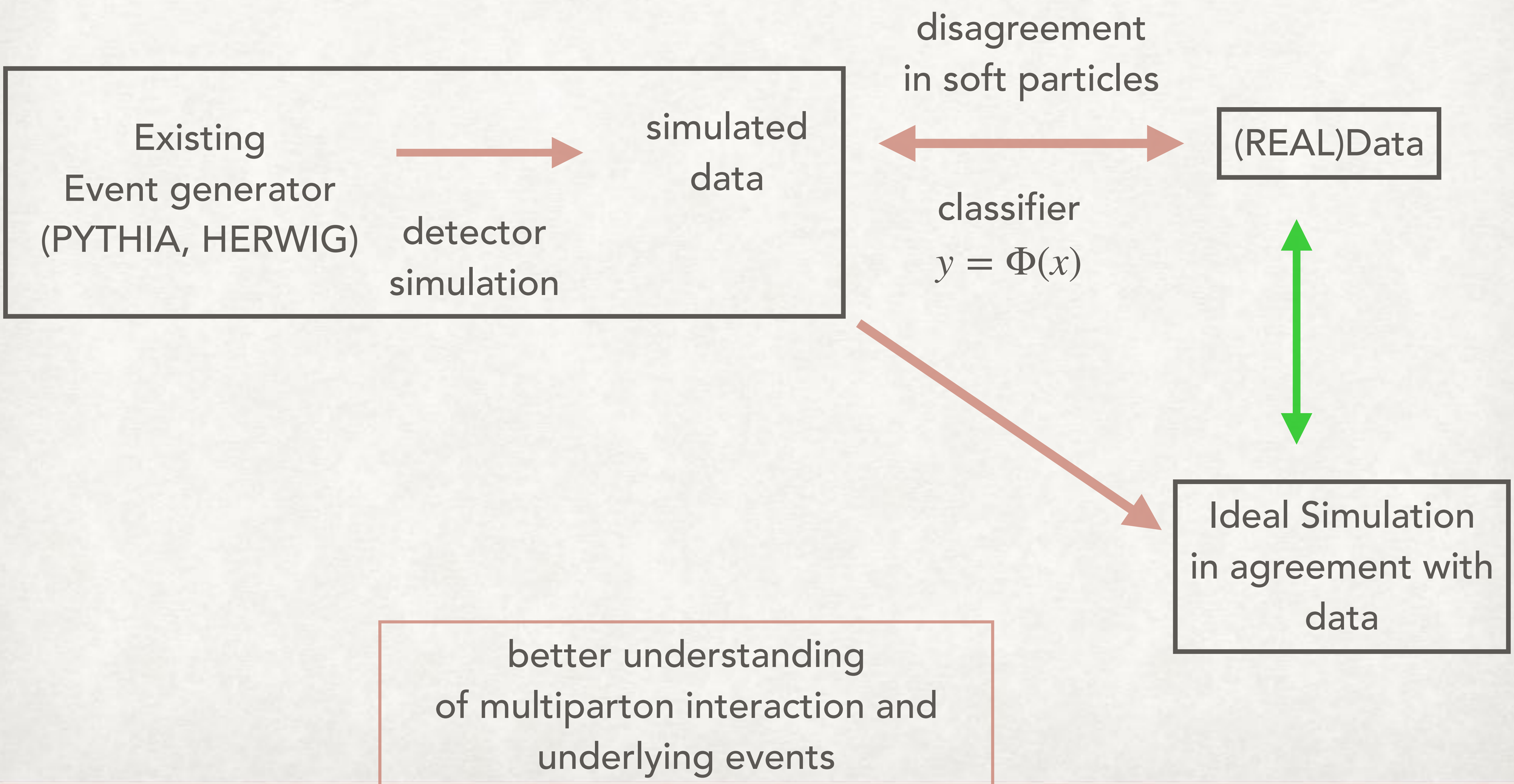
(REAL)Data

Ideal Simulation
in agreement with
data

better understanding
of multiparton interaction and
underlying events

HOW TO HANDLE DATA-MC DIFFERENCE

- Description of Low energy particle is empirical so we want to tune it by the data.



HOW TO HANDLE DATA-MC DIFFERENCE

- Description of Low energy particle is empirical so we want to tune it by the data.

Existing Event generator (PYTHIA, HERWIG) $\xrightarrow{\text{detector simulation}}$ simulated data

disagreement in soft particles

classifier $y = \Phi(x)$

(REAL)Data

Ideal Simulation in agreement with data

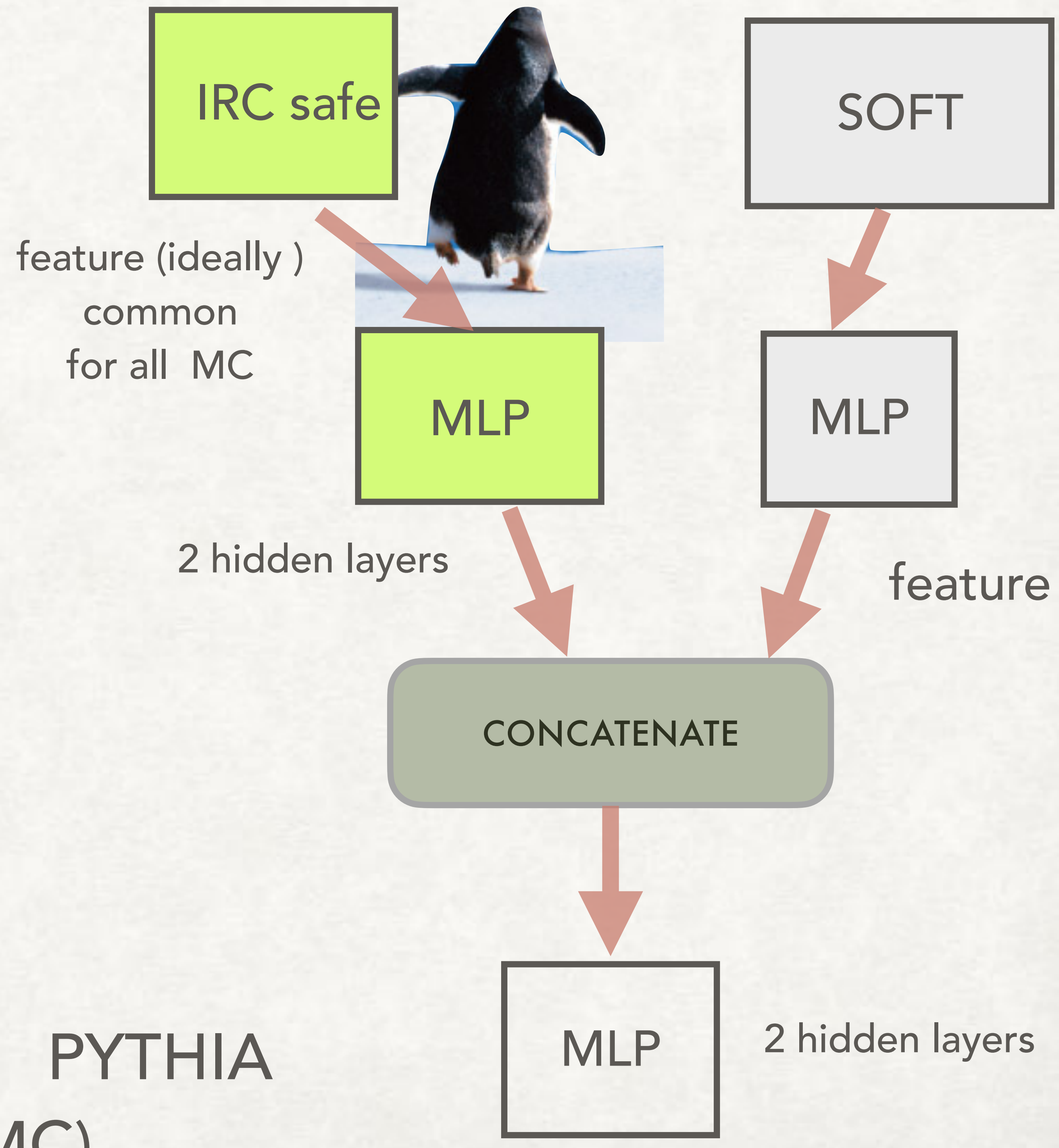
weight for i-th events $w_i = \dfrac{y_i}{1 - y_i}$

better understanding of multiparton interaction and underlying events

SOFT PARTICLES GEOMETRY IN JET CLASSIFICATION

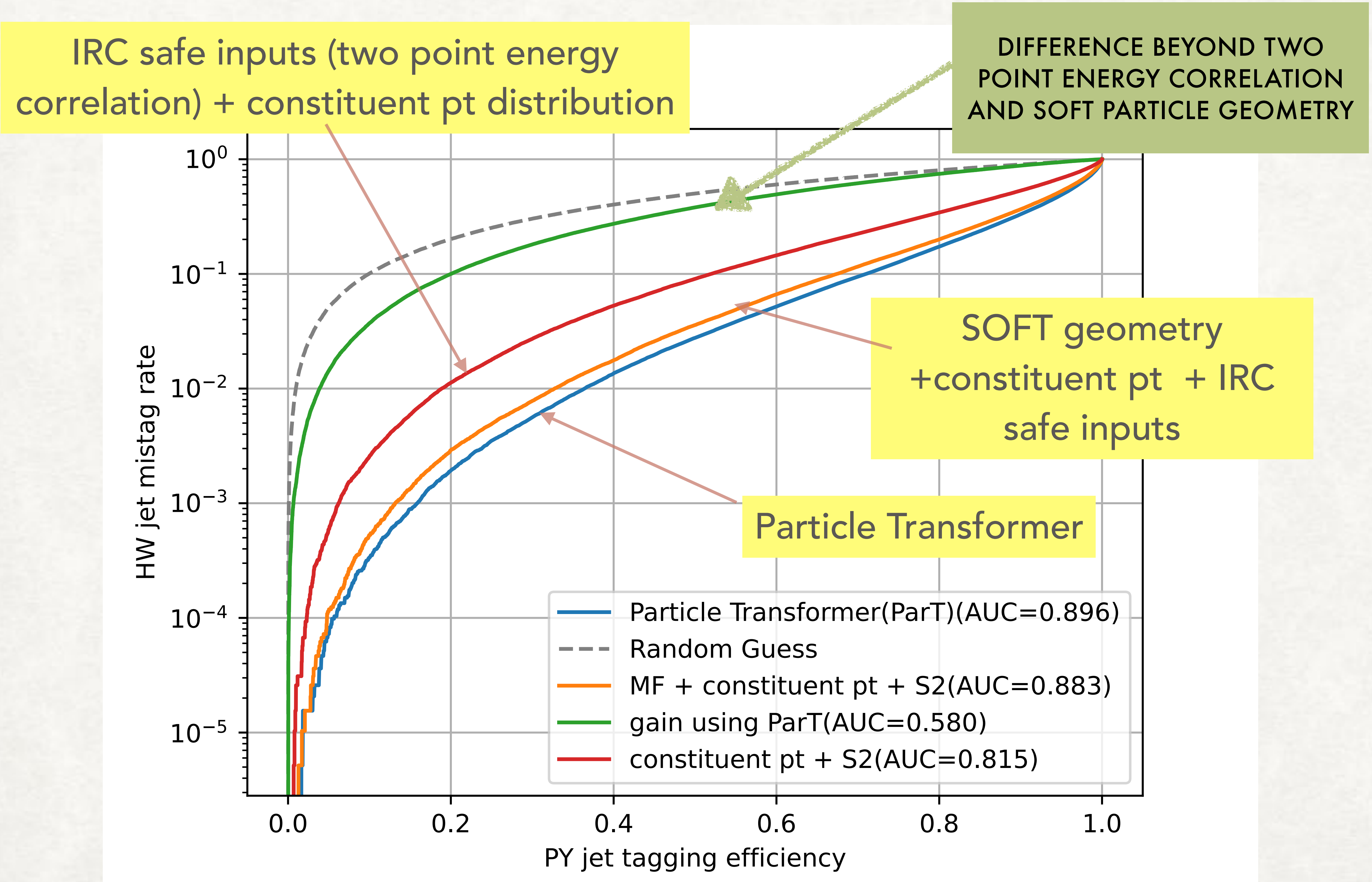
(FURUICHI, LIM, MN, IN PREPARATION)

- We have constructed a simple NN of following features represent GNN
  - "SOFT IMPUTS" and "IRC SAFE INPUTS" are separated before the first feature extraction.
  - A "complete" bases of aggregated input of #soft particles & geometry
  - No need of complicated network. just MLP



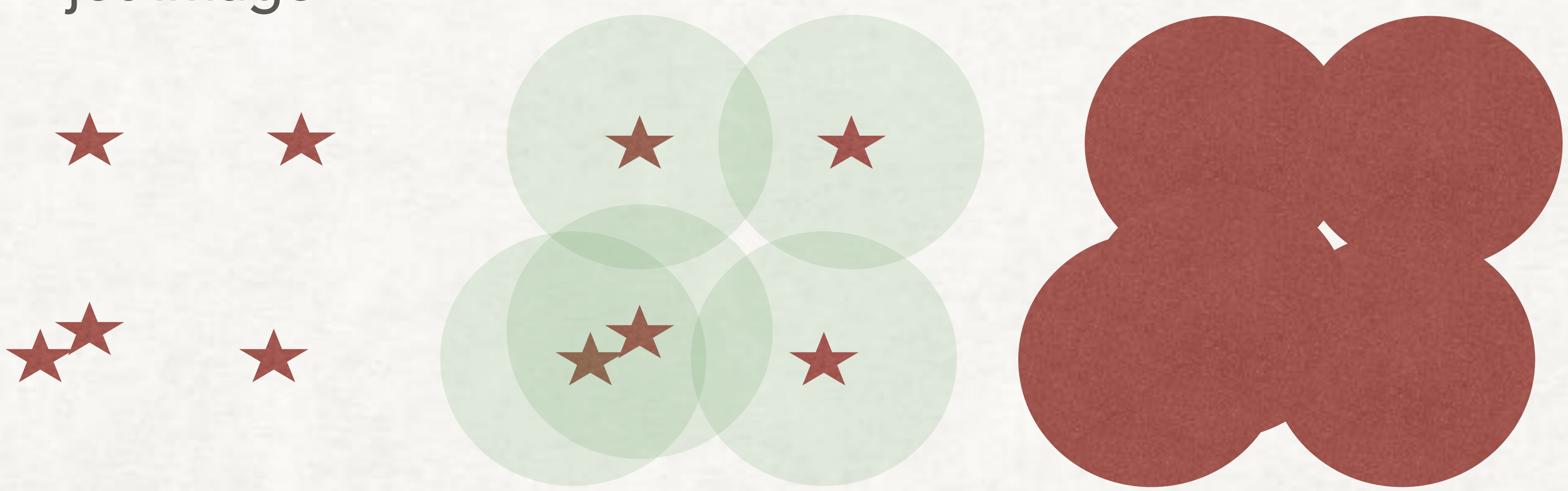
We test event reweighting   PYTHIA (DATA) vs HERWIG (MC)

QUANTIFYING QCD SIMULATION DIFFERENCES



SOFT INPUTS: MINKOWSKI FUNCTIONAL

Nojiri Lim



Minkowski Functionals

Area A(R)

Boundary Length L(R)

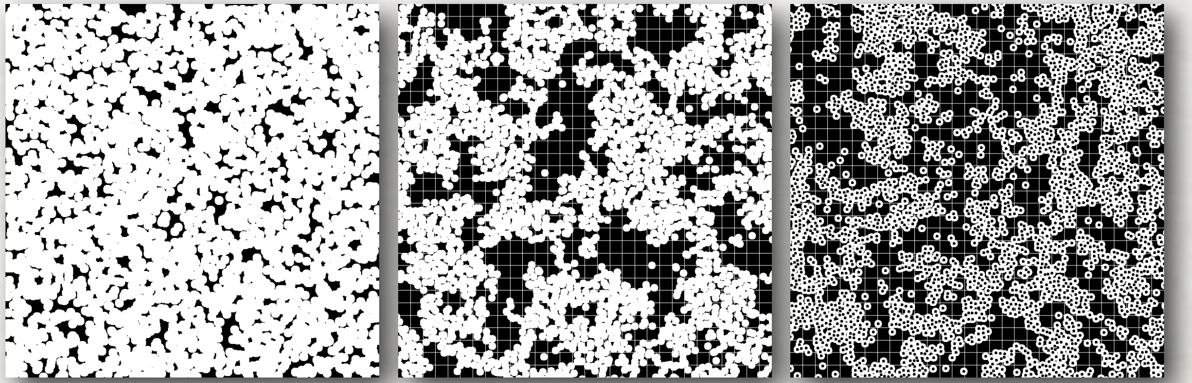
Euler Characteristic E(R)

A(R), L(R), E(R) is the base of all function F with $F(S \cup S') = F(S) + F(S') - F(S \cap S')$ and translation and rotation invariant

All point distribution information can be encoded here

Application in other field

- Statistical Physics (liquid crystal)
- Astrophysics



SOFT INPUTS 2. MINKOWSKI FUNCTIONAL

Nojiri Lim (2010.13469)

jet image   circle at every hit   **resulting image S**

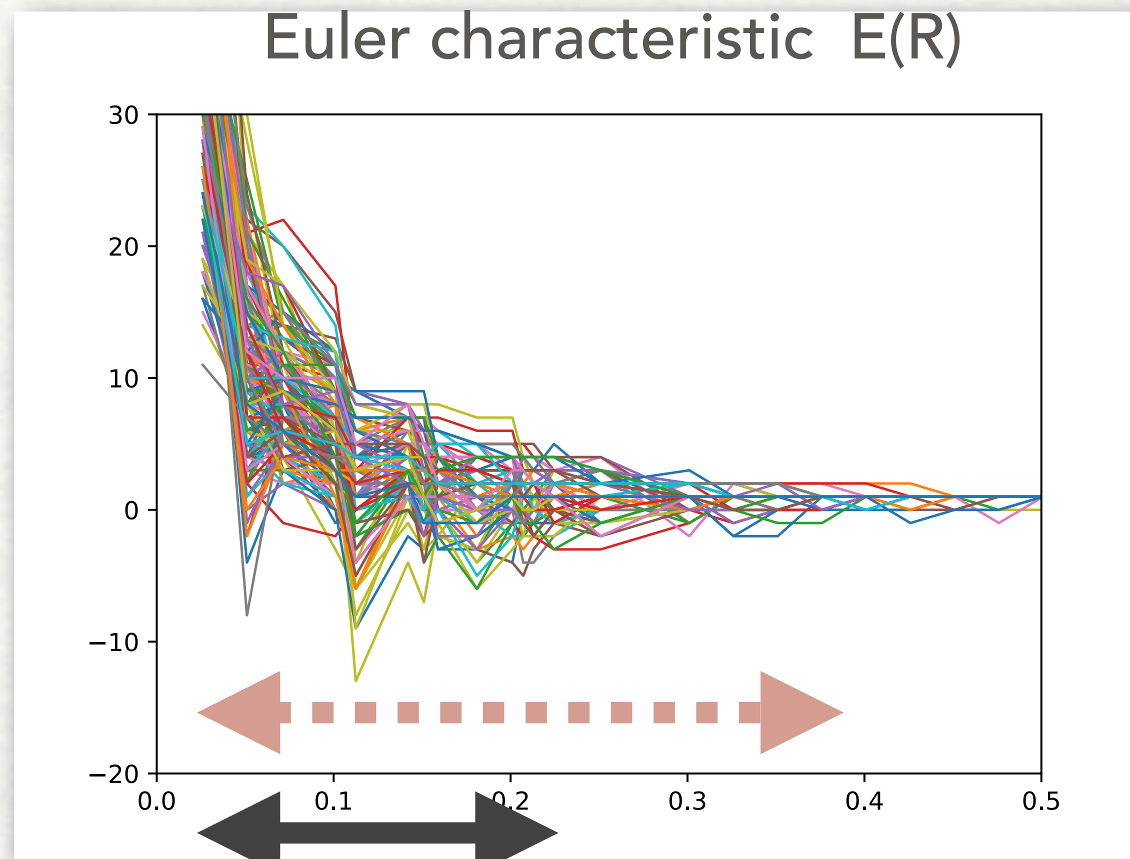
Minkowski Functionals

Area	A(R)
Boundary Length	L(R)
Euler Characteristic	E(R)

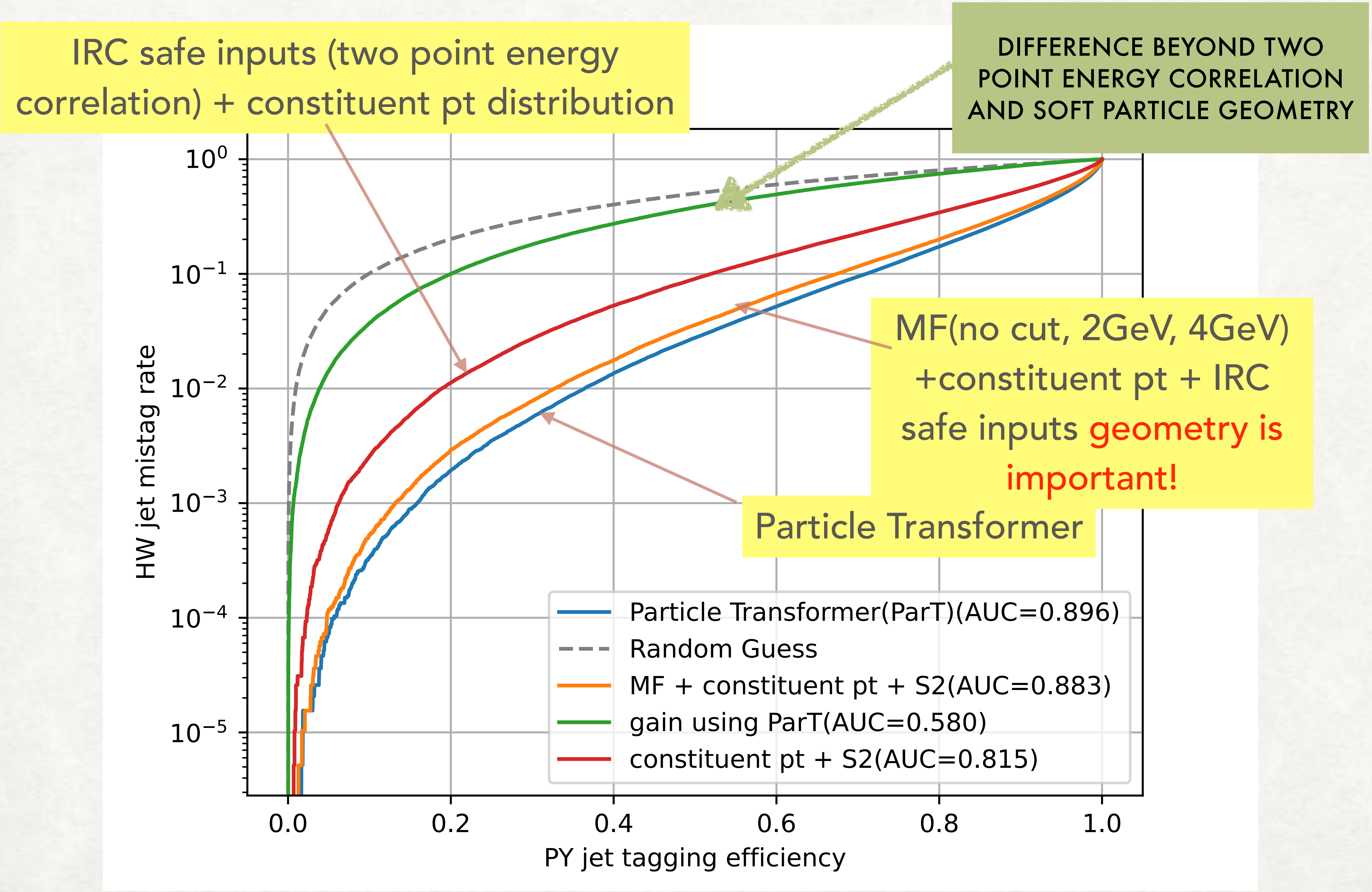
Several threshold (0.5GeV, 1 GeV, 2GeV, 4GeV)

aggregation of local information (Deepset)
Full geometry information up to rot/trans

number of particle, distance between particles, global effect such as color coherence…

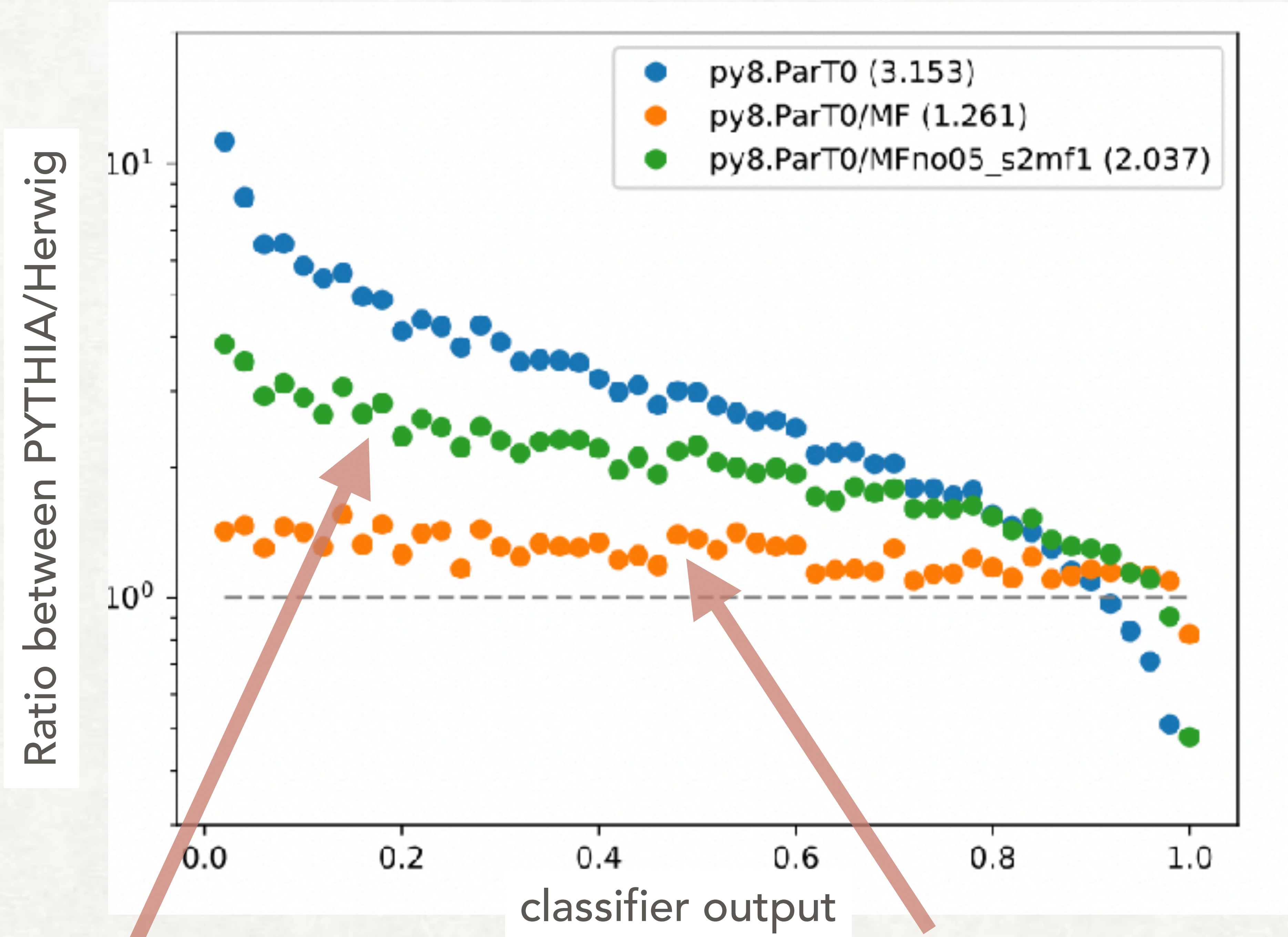


QUANTIFYING QCD SIMULATION DIFFERENCES



REWEIGHTING USING THE CLASSIFIERS

Output of classifier trained by Pythia events



IRC safe + constituent pT

reweight Herwig event
by (IRC safe+MF+constituent pT)

SUMMARY

- A part of improvement of DL comes from the soft particle information
    - Maybe there is more "unknown gods". GNN learns "everything"!
- We show the majority of soft particle effects can be parametrized by relatively simple aggregated inputs ("Minkowski Functionals") and simple MLP
- These inputs maybe used to
    - analyze experimental data,
    - reweight simulated events to improve the agreement between simulation and data in ML
    - GAN soft particles ?

to improve the "Event generators in the era of ML"

EXPERIMENTAL INTRO AT ML 4JET

BY PETAR MAKSIMOVIC (JOHNS HOPKINS)

QCD modeling for the future

- With a better QCD modeling, we could:
  - Train ML algorithms
    - → better data/MC agreement
    - → minimize signal efficiency systematics
  - Decorrelate taggers
    - → well-behaved background shapes → better bkg estimates
    - → if there's a BSM excesses, it would be "easier" to see
  - Estimate efficiencies of tagging jets with exotic substructure

    (see above)
- In general, experimentalist's life would become a lot easier

Certainly not easy --Maybe need "wish list" for soft QCD and ML

MLPhYs

Foundation of "Machine Learning Physics"

学習物理学の創成

Grant-in-Aid for Transformative Research Areas (A)

PD opening in KEK ML& particle, astro, cosmo.

(no ML publication history required.)

https://academicjobsonline.org/ajo/jobs/23019

Physics

Precise Prediction,
Mathematical description

Machine Learning

Innovation that can change society

Machine Learning Physics

discovery of new phenomena, new rule

Approach to fundamental Problem in Physics
by integrating machine learning and theoretical methods

STEALING

領域代表

Hashimoto

Kyoto

超弦理論
素粒子論
理論物理
機械学習

A01

Tomiya

Osaka Int' Tech

格子QCD
素粒子論
理論物理
機械学習

A02

Nojiri

KEK

素粒子論
理論物理
深層学習

A03

Otsuki

Sophia

物性理論
理論物理
機械学習

B01

Tanaka

RikenAIP

生成モデル
最適輸送 深層生成

B02

Kabashima

Tokyo

統計力学 情報理論
計算統計 機械学習

B03

Fukushima

Tokyo

理論物理 核理論
場の量子論機械学習

3 ATLAS+ 2 Belle
+ 2 Data scientists

BACK UP'S

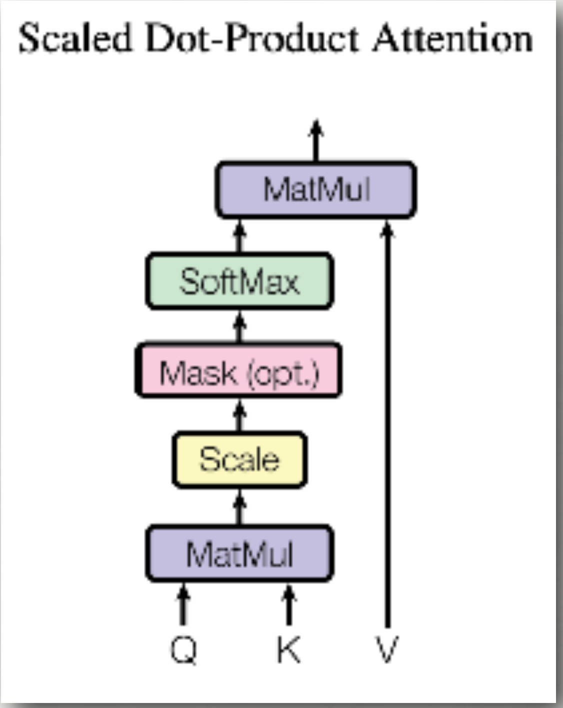
GRAPH NEURAL NETWORK (GNN)

- **ParticleNet,** treat nearby two point particle correlations directly

NN using
vertex (particle information)
edge (two point correlation)
as input.

Calculate edge variables of nearby 7 pairs
→update vertex and edge-> use this to select next 7 pairs

1902.08570 Qu and Goukos "Jet Tagging via Particle Clouds")



Particle Transformer

product of particles info

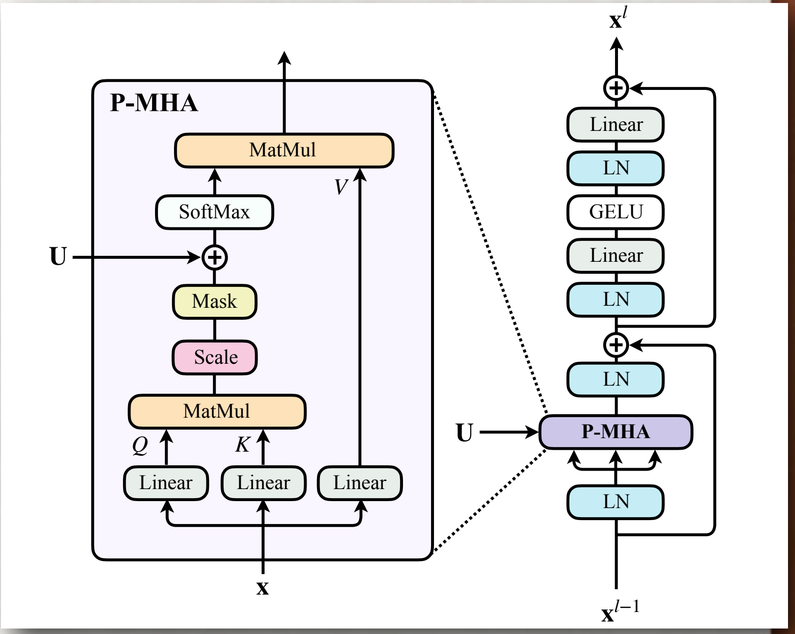
$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d_k} + \mathbf{U})V,$$

two point correlation of all particle

Edge information

$$\begin{aligned}
\Delta &= \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}, \\
k_{\mathrm{T}} &= \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})\Delta, \\
z &= \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})/(p_{\mathrm{T},a} + p_{\mathrm{T},b}), \\
m^2 &= (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,
\end{aligned}$$

pair



Now the idea of "distance" is controlled by the samples, not by theory.

GRAPH NEURAL NETWORK (GNN)

- **ParticleNet,** treat nearby two point particle correlations directly

NN using

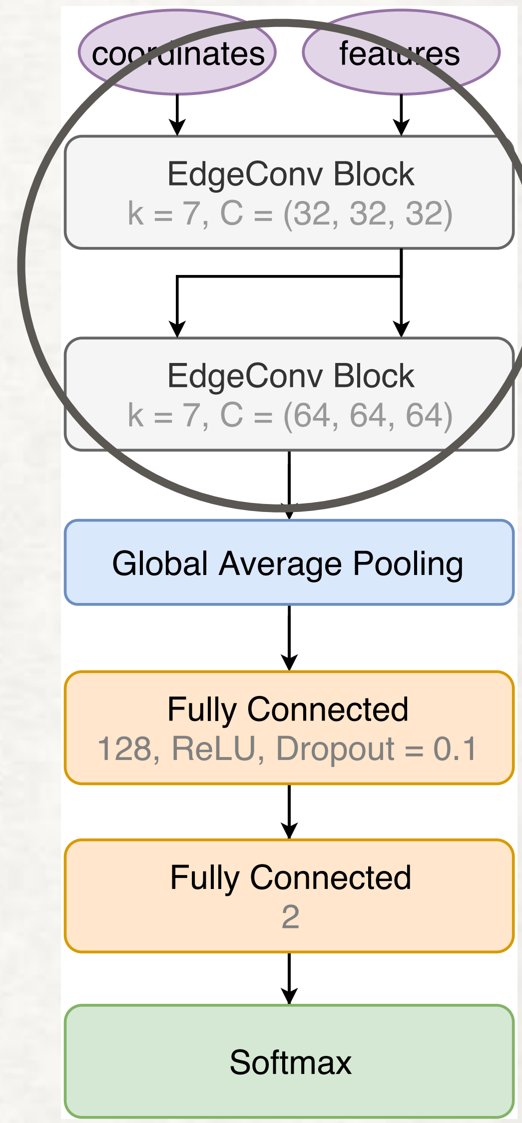
vertex (particle information)

edge (two point correlation)

as input.

Calculate edge variables of nearby 7 pairs

→update vertex and edge-> use this to select next 7 pairs



(b) ParticleNet-Lite

1902.08570 Qu and Goukos "Jet Tagging via Particle Clouds")

Particle Transformer

product of particles info

$$\text{P-MHA}(Q, K, V) = \text{SoftMax}(QK^T/\sqrt{d_k} + \mathbf{U})V,$$

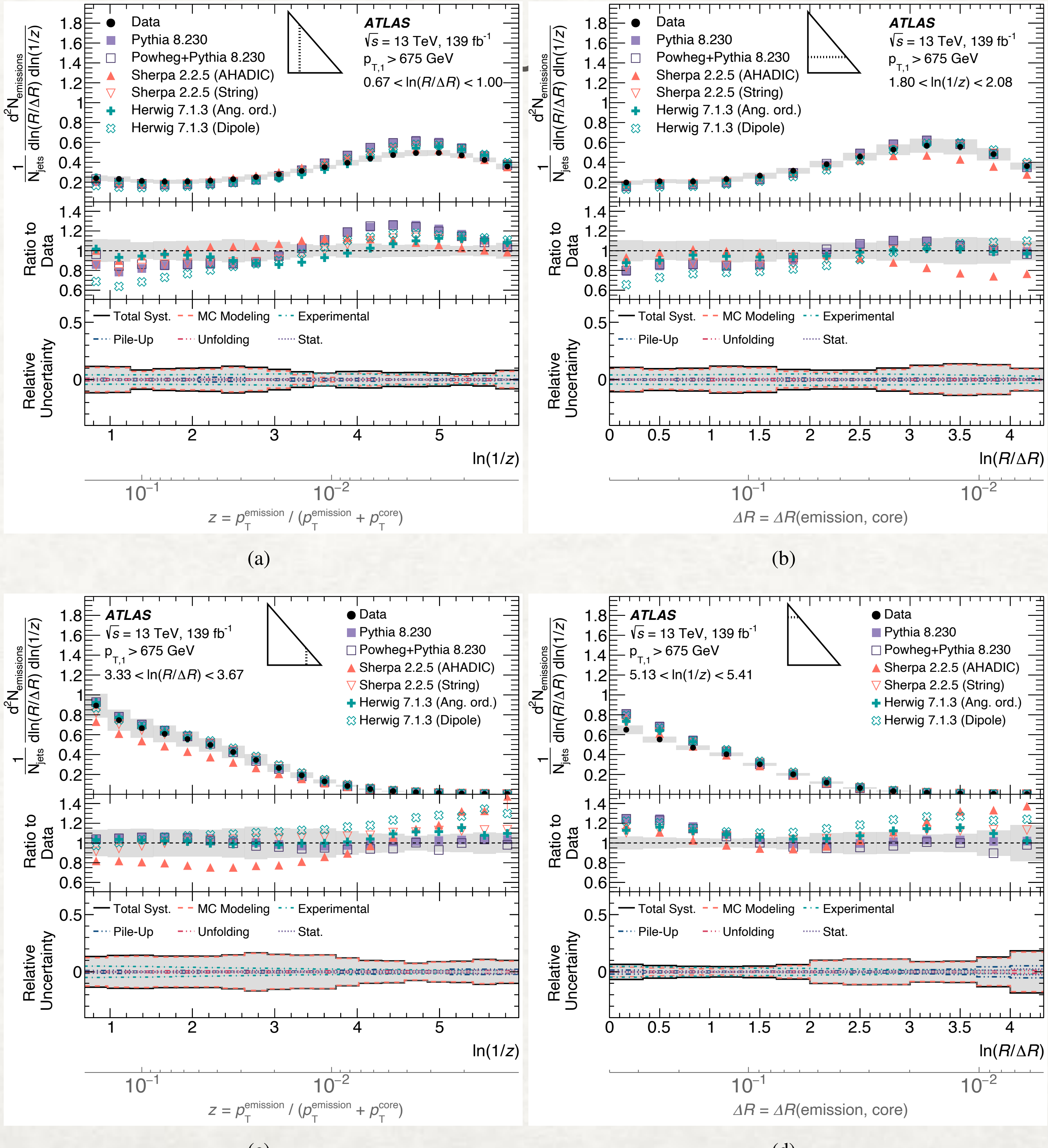
Edge information

pair

$$\begin{aligned}
\Delta &= \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}, \\
k_\mathrm{T} &= \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})\Delta, \\
z &= \min(p_{\mathrm{T},a}, p_{\mathrm{T},b})/(p_{\mathrm{T},a} + p_{\mathrm{T},b}), \\
m^2 &= (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,
\end{aligned}$$

Now the idea of "distance" is controlled by the samples, not by theory nor geometry.

Measurement of the Lund plane



(a) (b)

(c) (d)

Figure 3: Representative horizontal and vertical slices through the LJP. Unfolded data are compared with particle-level

IRC SAFE PART: TWO POINT ENERGY CORRELATION

Nojiri, Lim

EFN relay on jet direction (one point correlation) → two point correlation

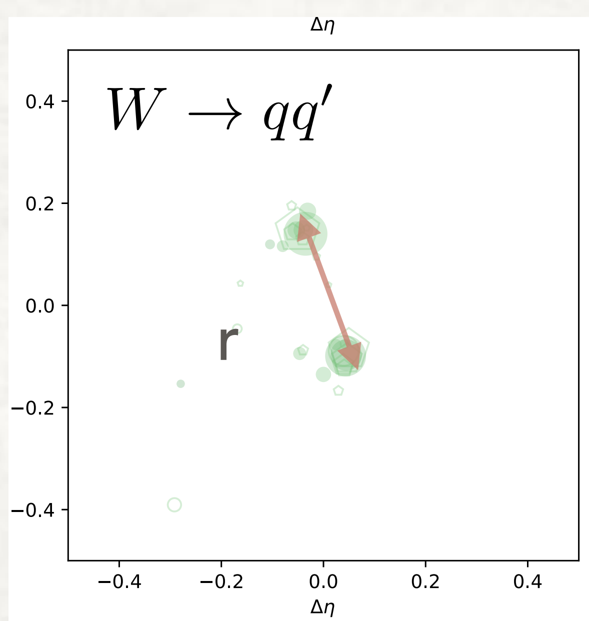
$$S_{2,ab}(R) = \sum_{i \in a, j \in b} p_{T,i} p_{T,j} \delta(R - R_{ij}).$$

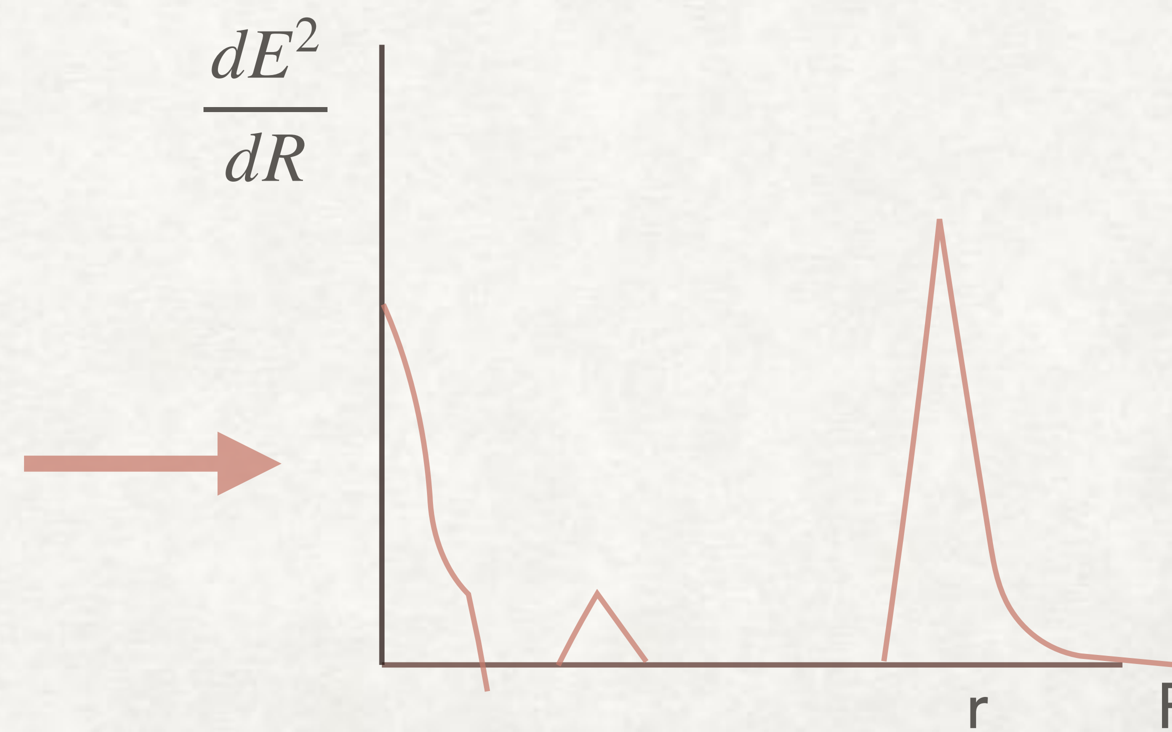
generating function

$$\mathrm{EFP}^n_{2,ab} = \int_0^\infty dR\, S_{2,ab}(R)\, R^n,$$

binning

$$S^{(k)}_{2,ab} = \int_{k\Delta R}^{(k+1)\Delta R} dR\, S_{2,ab}(R),$$

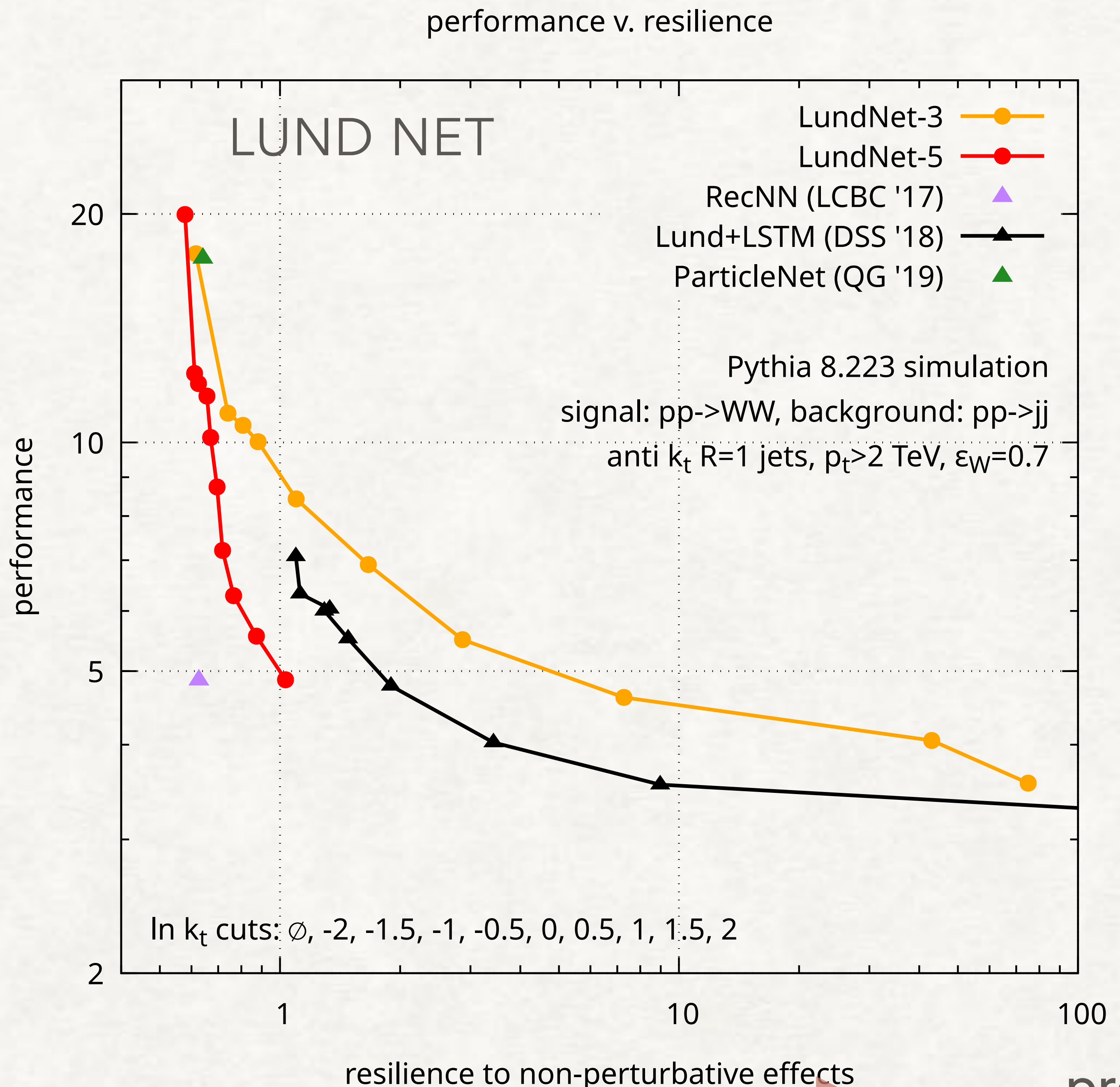




IS THIS GAIN REAL ?

better rejection

$$\frac{\epsilon(SG)}{\sqrt{\epsilon(BG)}}$$



predictability

Figure 8. Performance $\frac{\epsilon_W}{\sqrt{\epsilon_{\rm QCD}}}$ versus resilience to non-perturbative effects.

$$\zeta_{\rm NP} = \left(\frac{\Delta\epsilon_W^2}{\langle\epsilon\rangle_W^2} + \frac{\Delta\epsilon_{\rm QCD}^2}{\langle\epsilon\rangle_{\rm QCD}^2}\right)^{-1/2} \qquad \sim \frac{\epsilon}{\epsilon - \epsilon(\text{parton level})}$$