

# Top Tagging and Machine Learning

Sung Hak Lim  
Rutgers University



RUTGERS

Top 2023,  
Traverse City, MI, USA

Sep. 2023

Based on 1807.03312, 1904.02092, 2003.11787, 2010.13469  
+ work in progress with A. Furuichi and M. M. Nojiri

# Boosted Top Jet Tagging and Machine Learning

Sung Hak Lim  
Rutgers University



RUTGERS

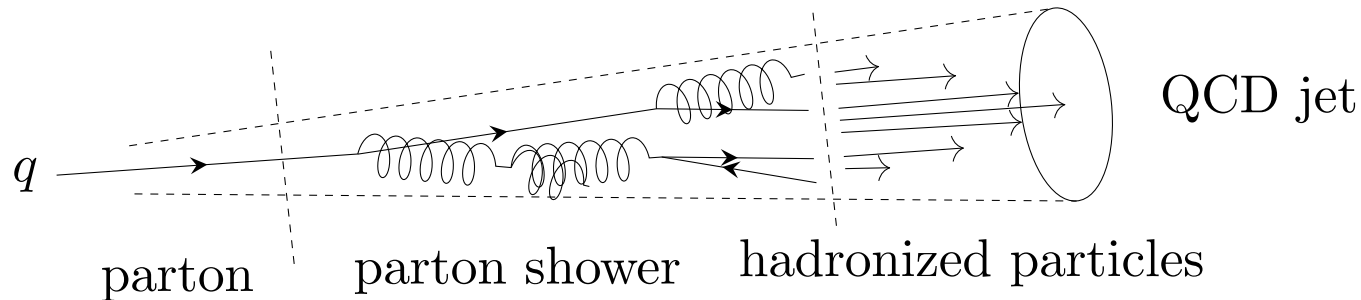
Top 2023,  
Traverse City, MI, USA

Sep. 2023

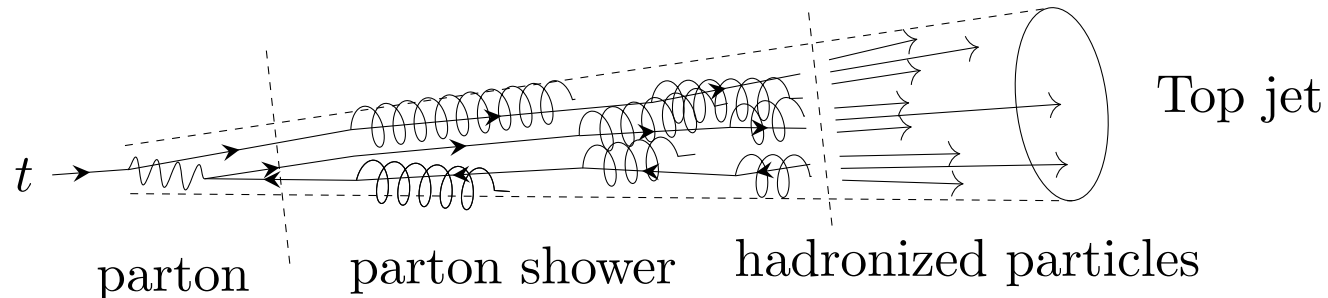
Based on 1807.03312, 1904.02092, 2003.11787, 2010.13469  
+ work in progress with A. Furuichi and M. M. Nojiri

# What is a Jet?

At a hadron-hadron collider, such as LHC, a collimated particle cluster called a jet appears when a colored parton is produced.

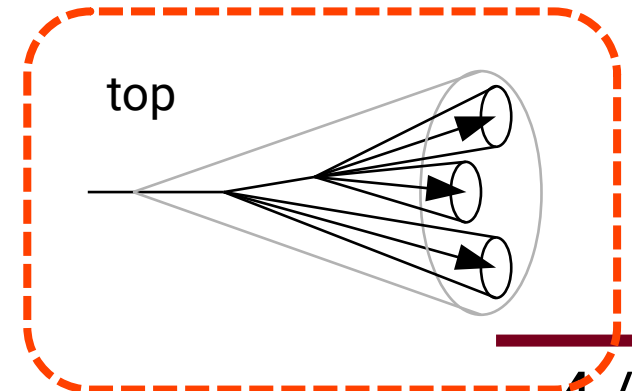
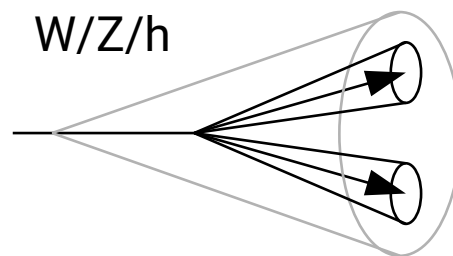
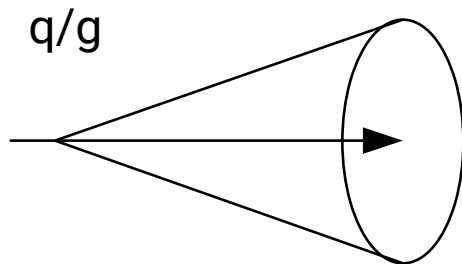
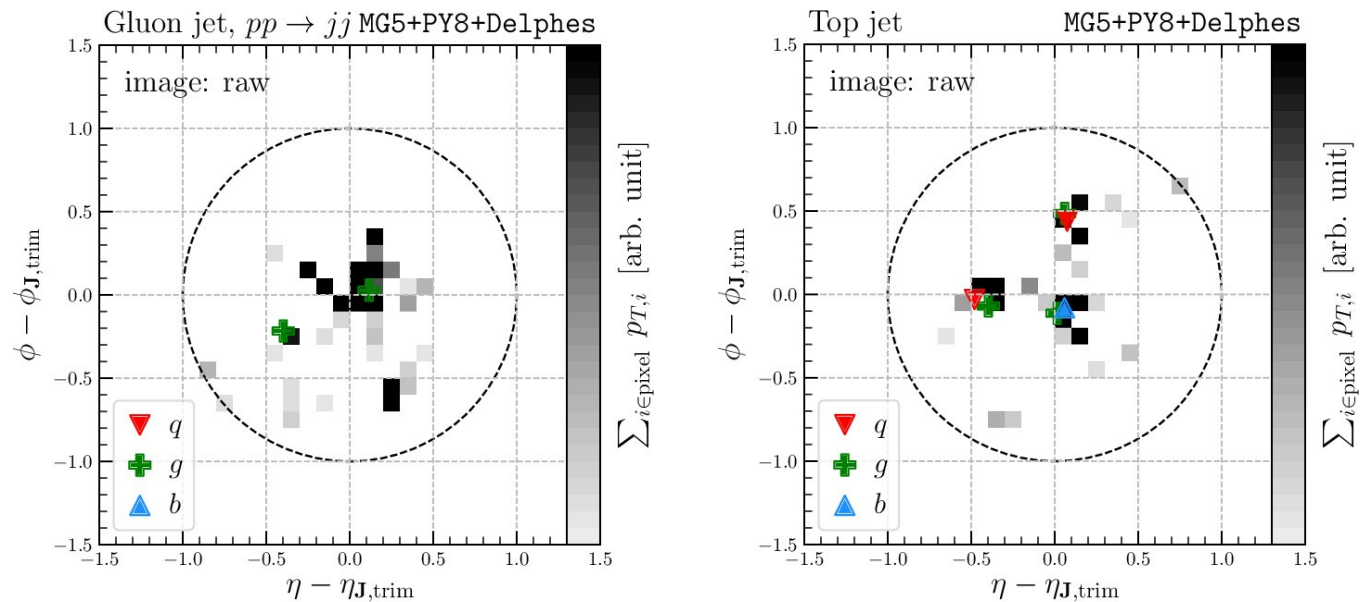


Since we see more and more multi TeV center-of-mass energy events at colliders, boosted heavy particles are produced. They can form a single collimated cluster of particles similar to the QCD jets.



# Jets have substructure!

In order to distinguish non-trivial jets from the QCD jets, we need to check features of jets called substructure:

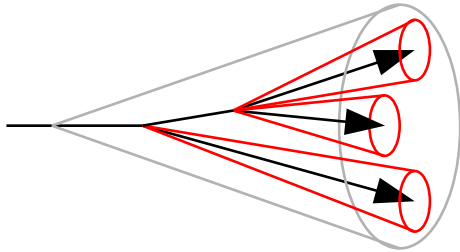


# Top Jet Substructures

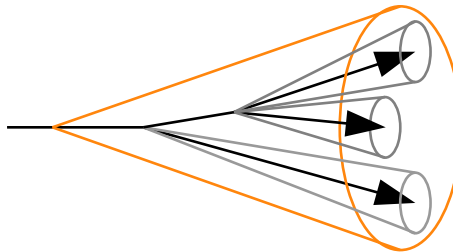
---

Boosted Top jets is very rich in characteristic features. Good example for tagger study and benchmarking various ML architectures!

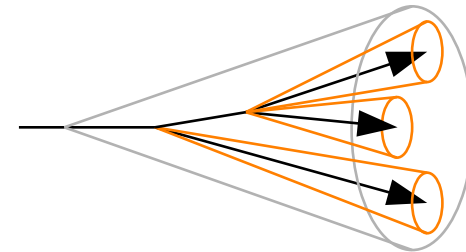
Three-prong



Color triplet

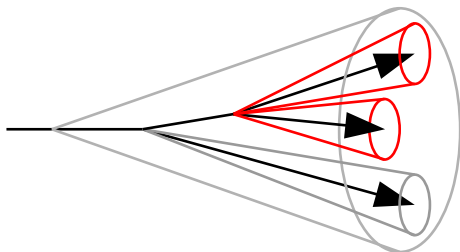


Color triplet subjets

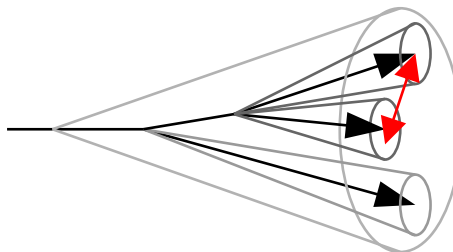


Top jet also have **W boson jet** inside.

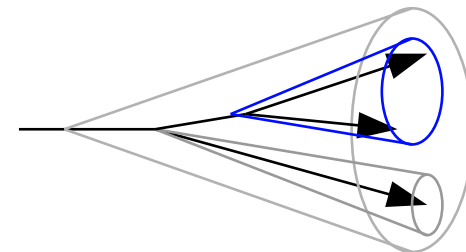
Two-prong subjet inside



Color connection

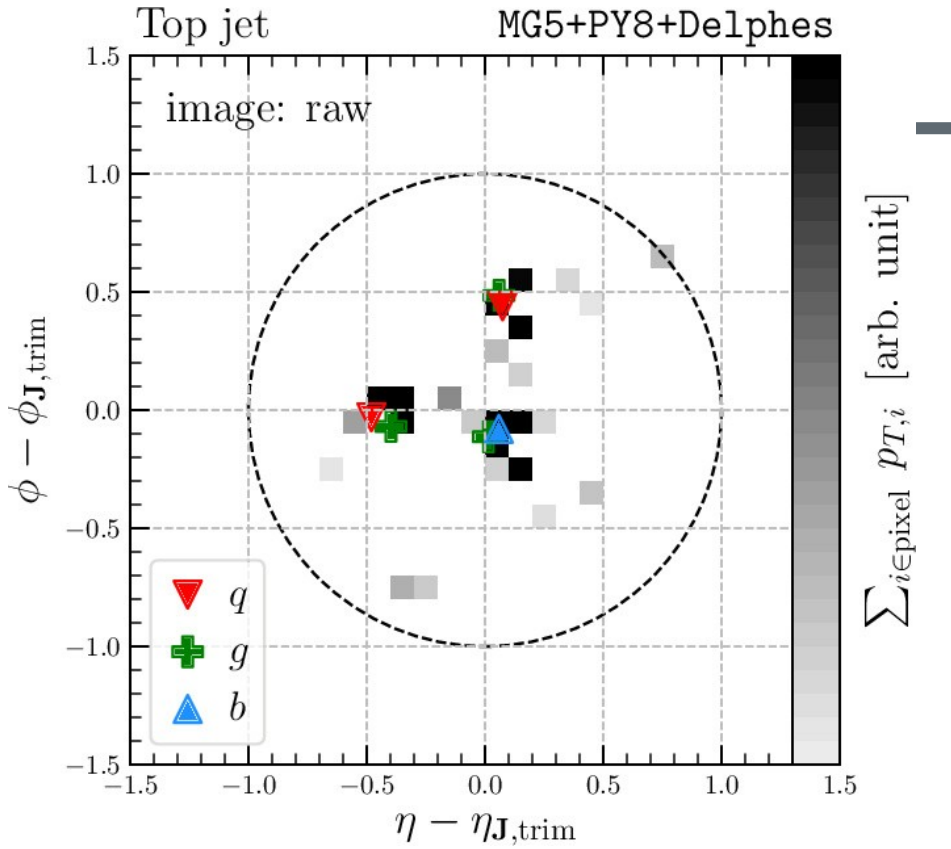


Color singlet

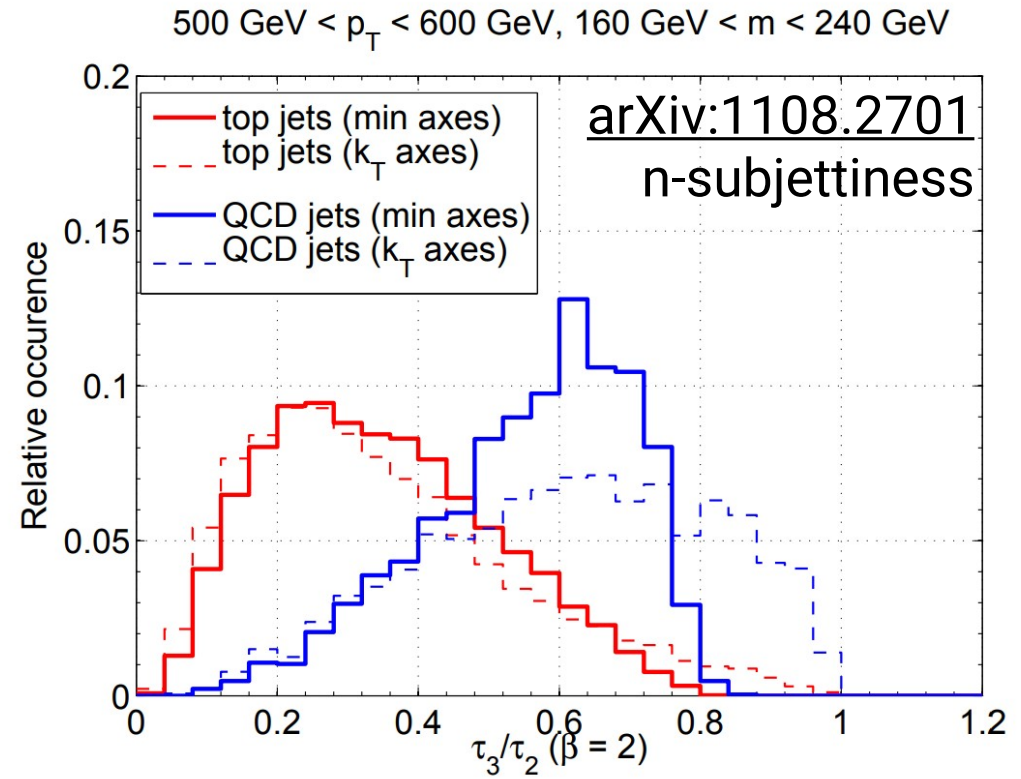


Jet tagger requires understanding of these patterns either implicitly or explicitly. Machine learning is very useful tool for analyzing these patterns!

There are **two** approaches for building ML based jet taggers:



Directly analyze  
**Low-Level Feature (LLF)**



Construct physics-motivated  
**High-Level Features (HLF)**

Use CNN/GNN/Transformers to analyze LLF

- ParticleNet
- ParticleTransformer
- LorentzNet, PELICAN (equivariant GNN/Transformer)

- Jet PT, mass, (basic kinematics)
- N-subjettiness
- Energy Flow Polynomials
- Constituent Multiplicities...

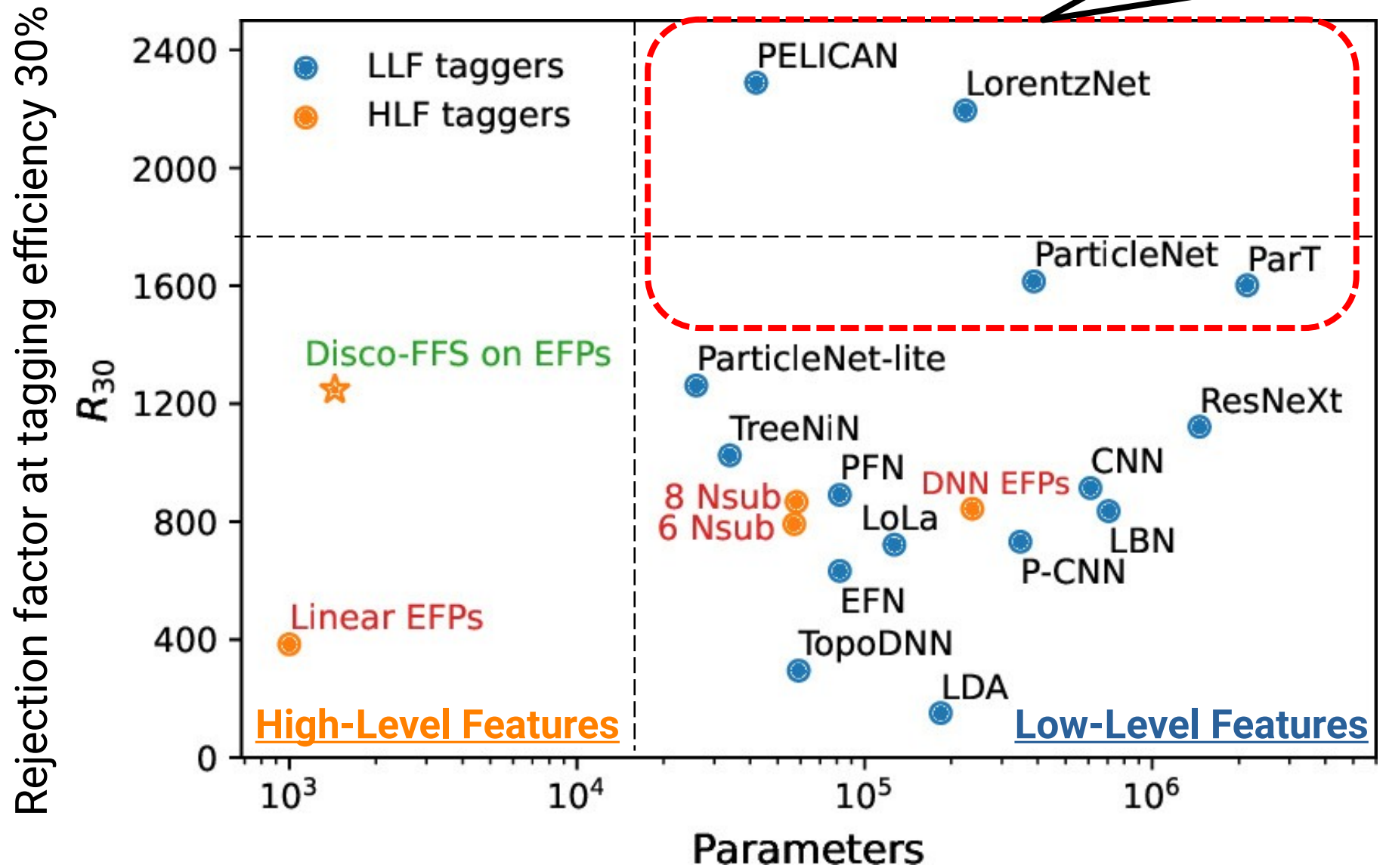
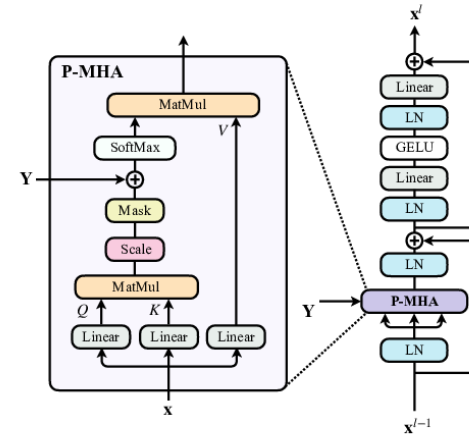
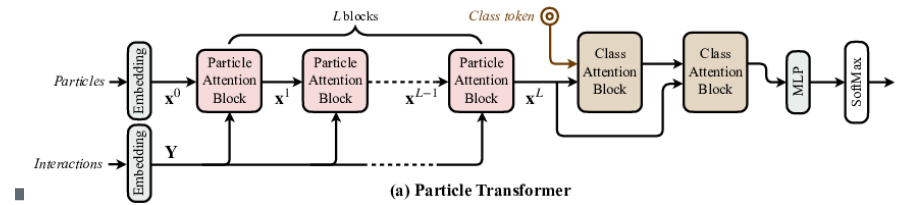
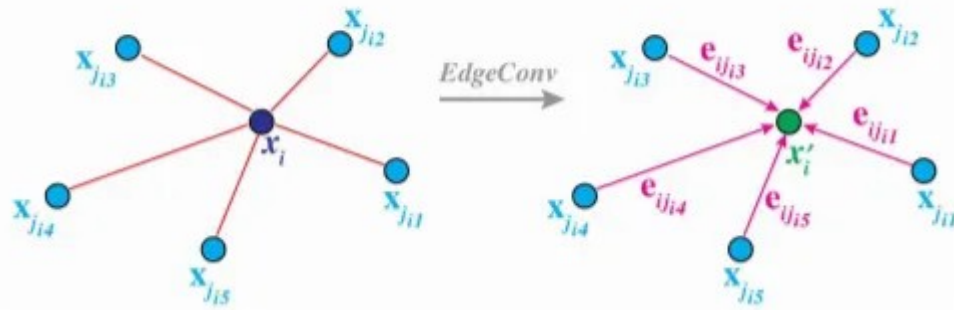
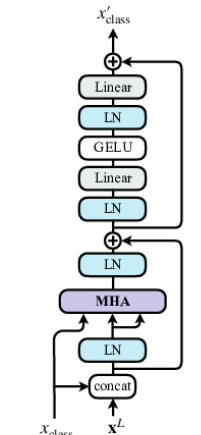


Figure from R. Das, G. Kasieczka, D. Shih, 2212.00046



(b) Particle Attention Block



(c) Class Attention Block

GNN / Transformers are working great. But because they are general purpose low-level feature analysis tools, it is hard to understand outcome other than the fact that they estimated the classifier output (likelihood ratio) more precisely.

Can we build up a high-level feature based classifier equally performing well? YES!

Advantages:

- simpler network: less training uncertainty (at a cost of expressivity)
- interpretable (by understanding HLF inputs)

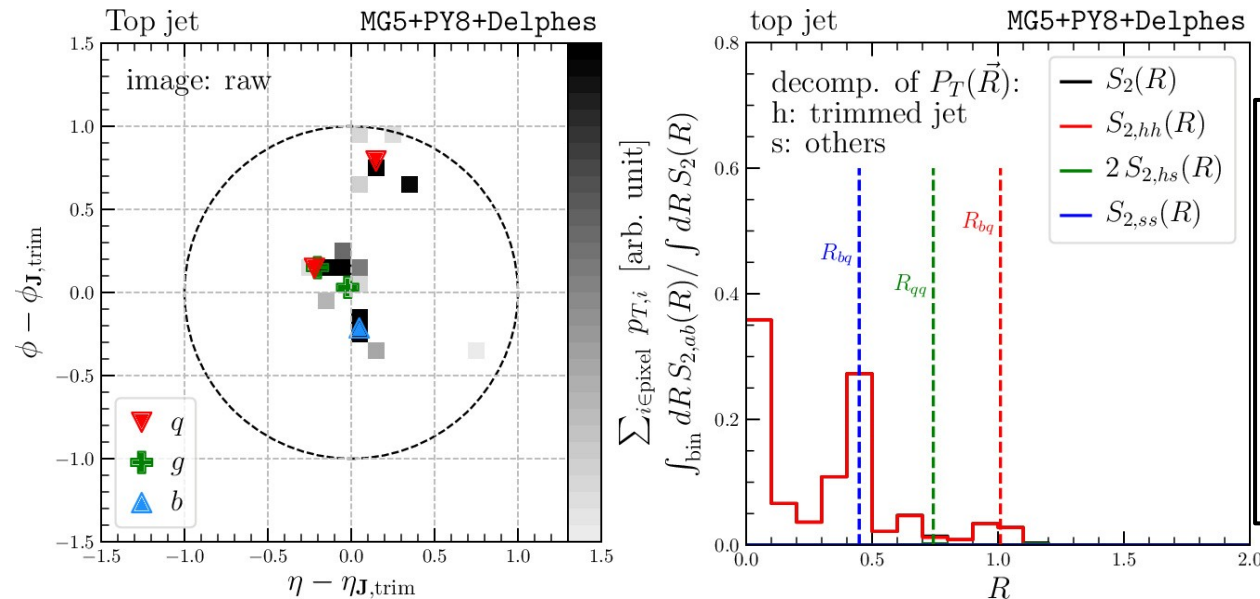


# Two-point energy correlation spectrum

Two-point energy correlation is an aggregated energy correlation between two constituents at a distance  $R$ .

$$S_{2,ab}(R) = \int d\vec{R}_1 d\vec{R}_2 P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_2) \delta(R - R_{12})$$

$$P_{T,a}(\vec{R}) = \sum_{i \in \mathbf{J}_a} p_{T,i} \delta(\vec{R} - \vec{R}_i)$$



Two-point energy correlation captures three characteristic angular scales of three prong substructures.

# IRC-safe energy correlator based Neural Networks

A. Chakraborty, **SHL**, M. M. Nojiri, M. Takeuchi,  
2003.11787

We use the two point correlation  $S_2$  as inputs to MLP.  
The resulting network is called Relation Network,  
a type of GNN using only edge features.

$$\text{First linear layer: } \int dR S_2(R) \phi^e(R) = \sum_{i,j \in J} p_{T,i} p_{T,j} \phi^e(R_{ij})$$

Graph Networks

Relation Network

Utilizes edge features

$$F\left[\sum_{i,j \in J} \phi^e(p_i, p_j)\right]$$

Raposo, et al. (1702.05068),  
Santoro, et al. (1706.01427)

IRC-safe energy correlator  
based Networks

Relation Network

Utilizes two-point energy correlation

$$F\left[\sum_{i,j \in J} p_{T,i} p_{T,j} \phi^e(R_{ij})\right]$$

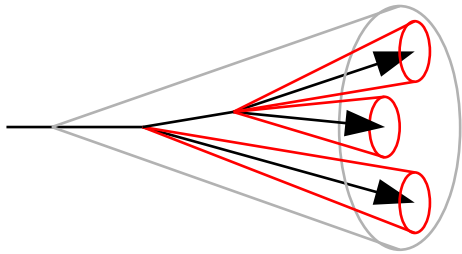
Chakraborty, **SHL**, Nojiri, and Takeuchi (2003.11787)

This network is able to analyze most of prong substructures  
and their correlations.

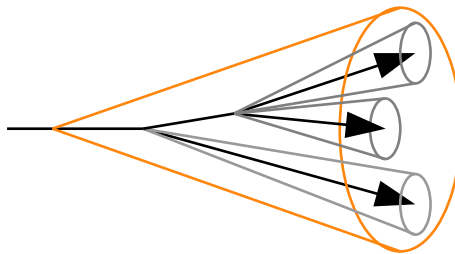
# What else do we need?

Boosted Top jets is very rich in characteristic features

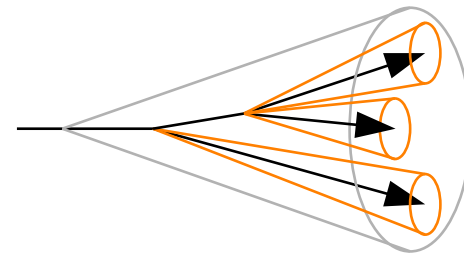
Three-prong



Color triplet

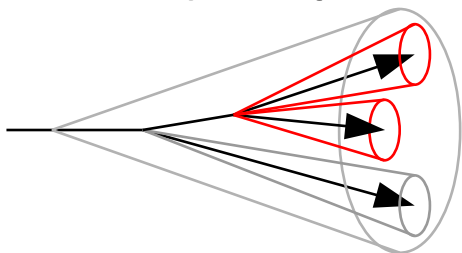


Color triplet subjects

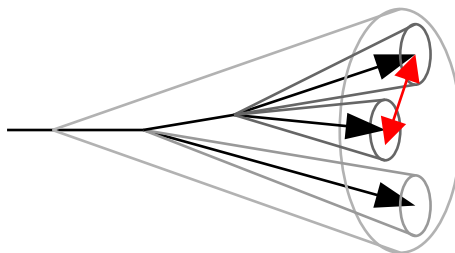


Top jet also have W boson jet inside.

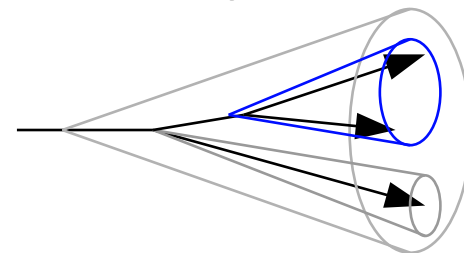
Two-prong subjet inside



Color connection



Color singlet



Color charge sensitive variable → constituent multiplicity.

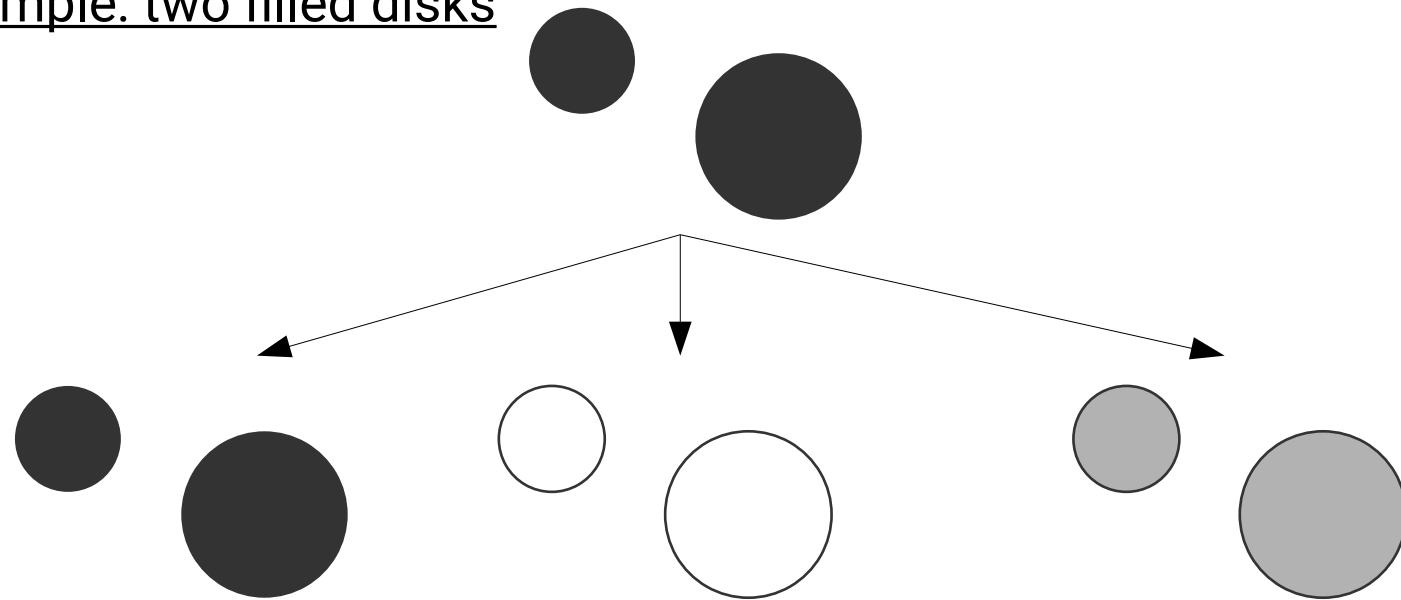
We will consider its generalization: Minkowski Functionals

# Minkowski Functionals

**Minkowski functionals** (MFs) are the basis of geometric measure (called valuation) of a given set.

For 2D object analysis, there are three MFs:

Example: two filled disks



Area

$$MF_0 = \int_A d^2\vec{r}$$

Boundary length

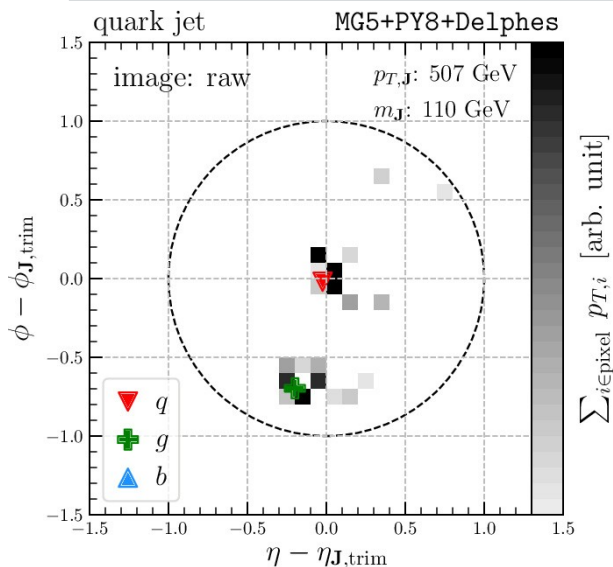
$$MF_1 = \frac{1}{2\pi} \int_{\partial A} d\vec{r}$$

Euler characteristic

$$MF_2 = \frac{1}{2\pi^2} \int_{\partial A} \frac{1}{R} d\vec{r} \quad (\text{Gauss-Bonnet})$$

With these three numbers, we can describe all the geometric measures related to this 2D objects. (Hadwiger's theorem)

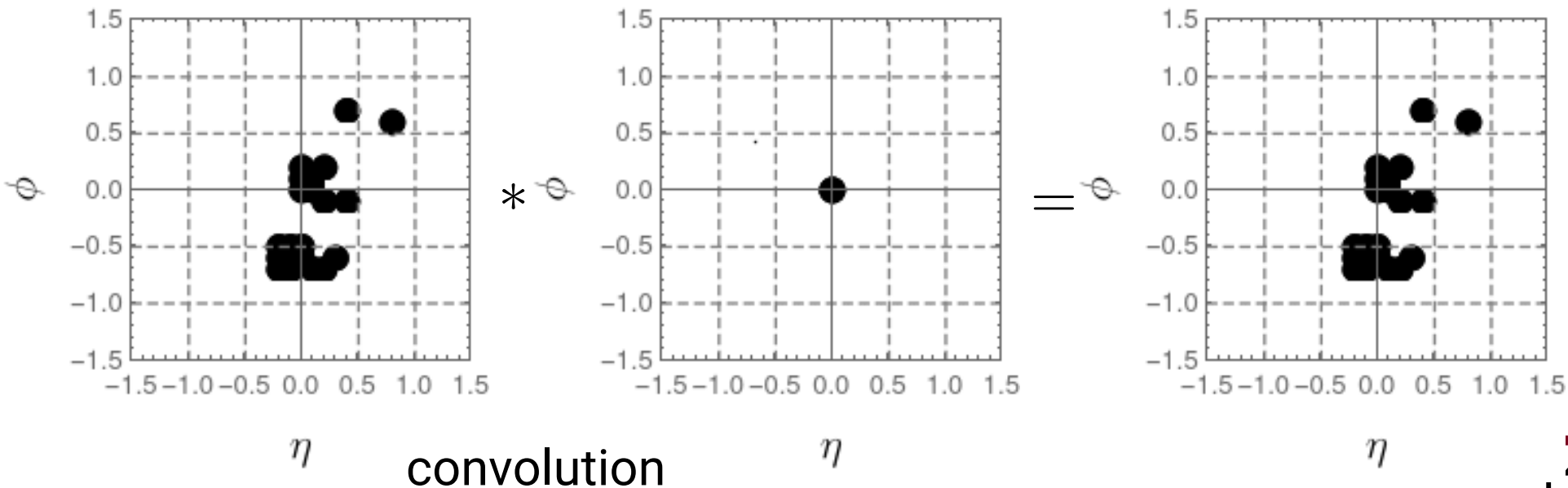
# Mathematical Morphology: Minkowski Functionals and Dilation



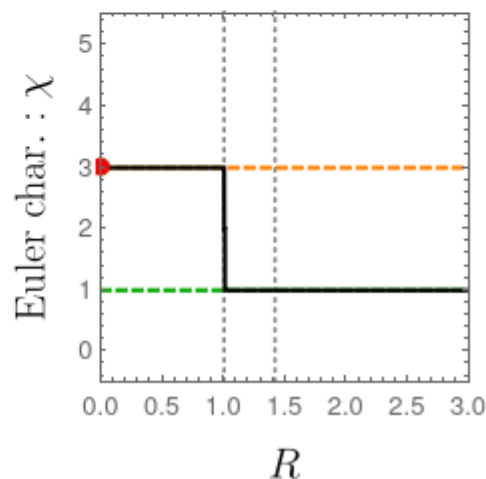
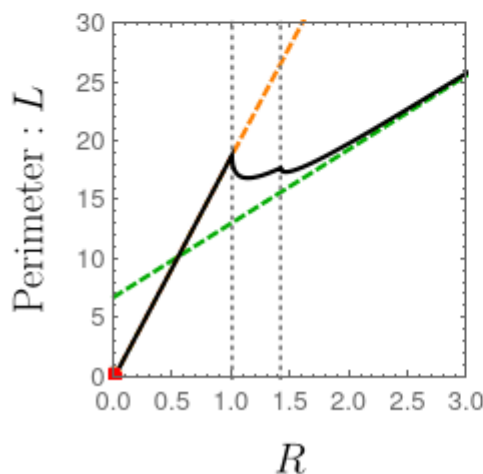
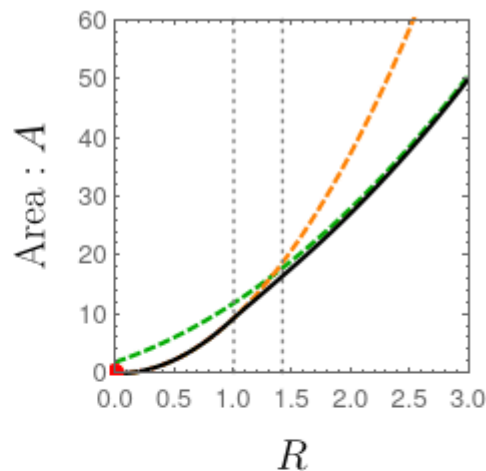
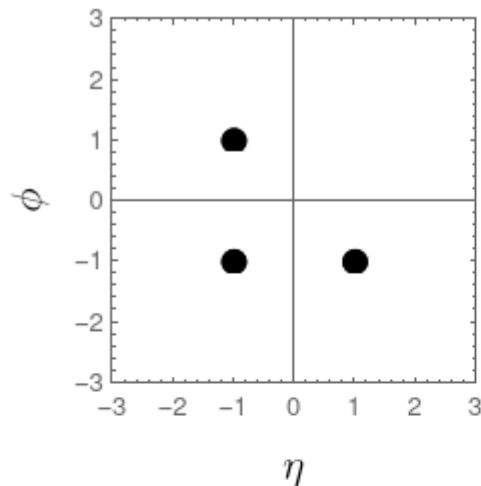
↓ Binarize

For the jet constituents analysis, we **binarize** the points using energy cutoffs and apply **dilation** on the binary image.

This morphological operation may be regarded as coarse-graining, and it will allow us to systematically analyze the **geometry of jet constituents** when we use this together with the MFs.



# Mathematical Morphology and Minkowski Functionals



If we have more constituents, such behavior changes may happen multiple times.

Start:

Cech complex:  
three points

First change happens when the nearest-neighbor meets

$$R = 1$$

Cech complex:  
an L-shaped line

Second change happens when the next nearest-neighbor meets

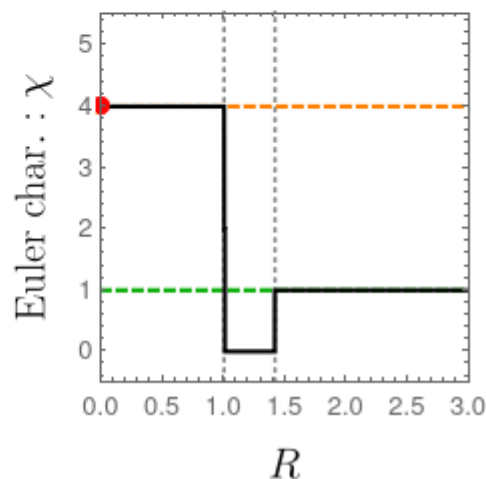
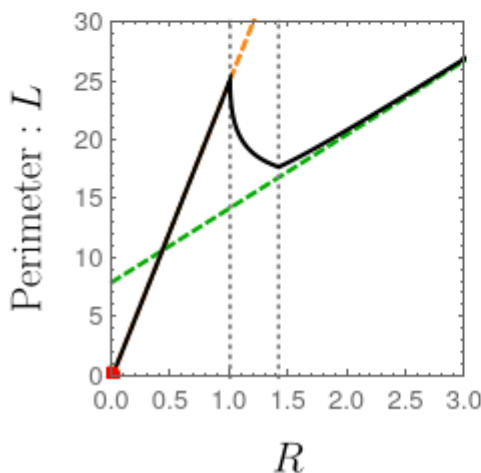
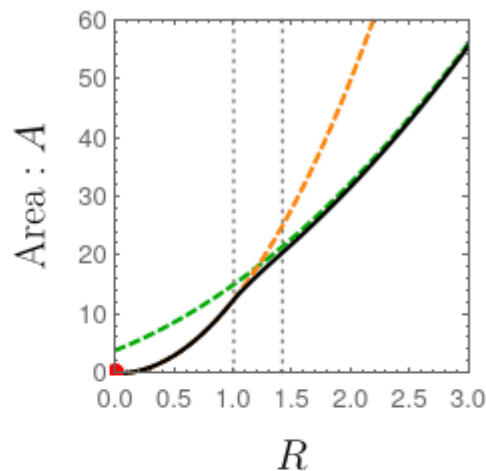
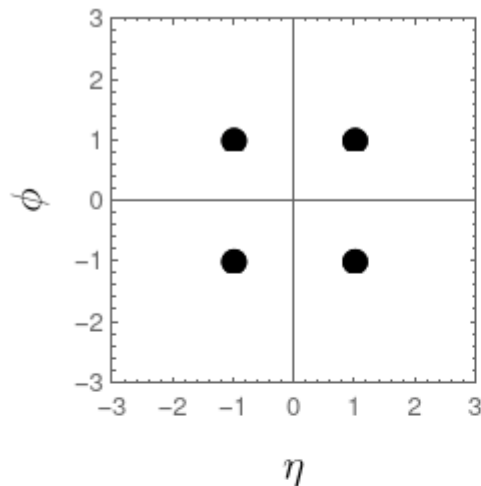
$$R = \sqrt{2}$$

Cech complex:  
a right triangle

**Orange:** asymptote as  $r \rightarrow 0$

**Green:** asymptote as  $r \rightarrow \text{infinity}$

# Mathematical Morphology and Minkowski Functionals



During the dilation, some peculiar topological structures may appear. For example, when a **hole** appears, the Euler characteristic can record that clearly.

Start: four constituents

$$\chi = 4$$

Cech complex:  
four dots

Hole appears:

cancels Euler characteristic by 1.

$$\chi = 1 - 1 = 0$$

Cech complex:  
a square

Hole disappears:

$$\chi = 1$$

Cech complex:  
a filled square

The topology of the jet constituents can be analyzed.

**Orange**: asymptote as  $r \rightarrow 0$

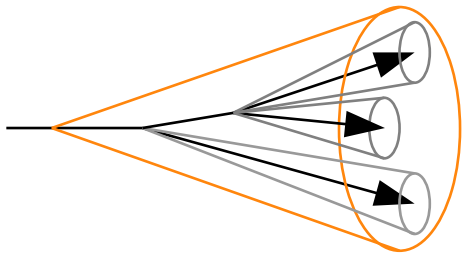
**Green**: asymptote as  $r \rightarrow \infty$

# Subjet color charges

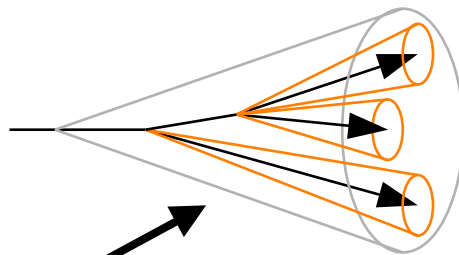
---

Minkowski functionals will deal with jet topology and color charges.  
How about subjet color charges?

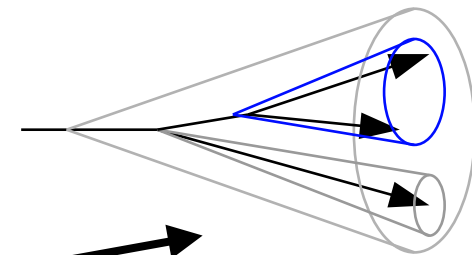
Color triplet



Color triplet subjets



Color singlet



Subjet  
Constituent Multiplicity

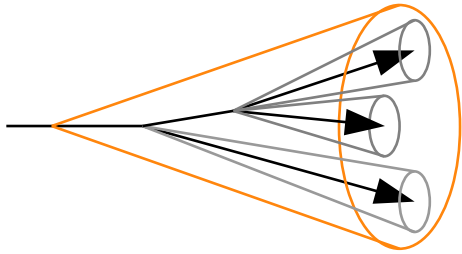




# Subjet color charges

Minkowski functionals will deal with jet topology and color charges.  
How about subjet color charges?

Color triplet



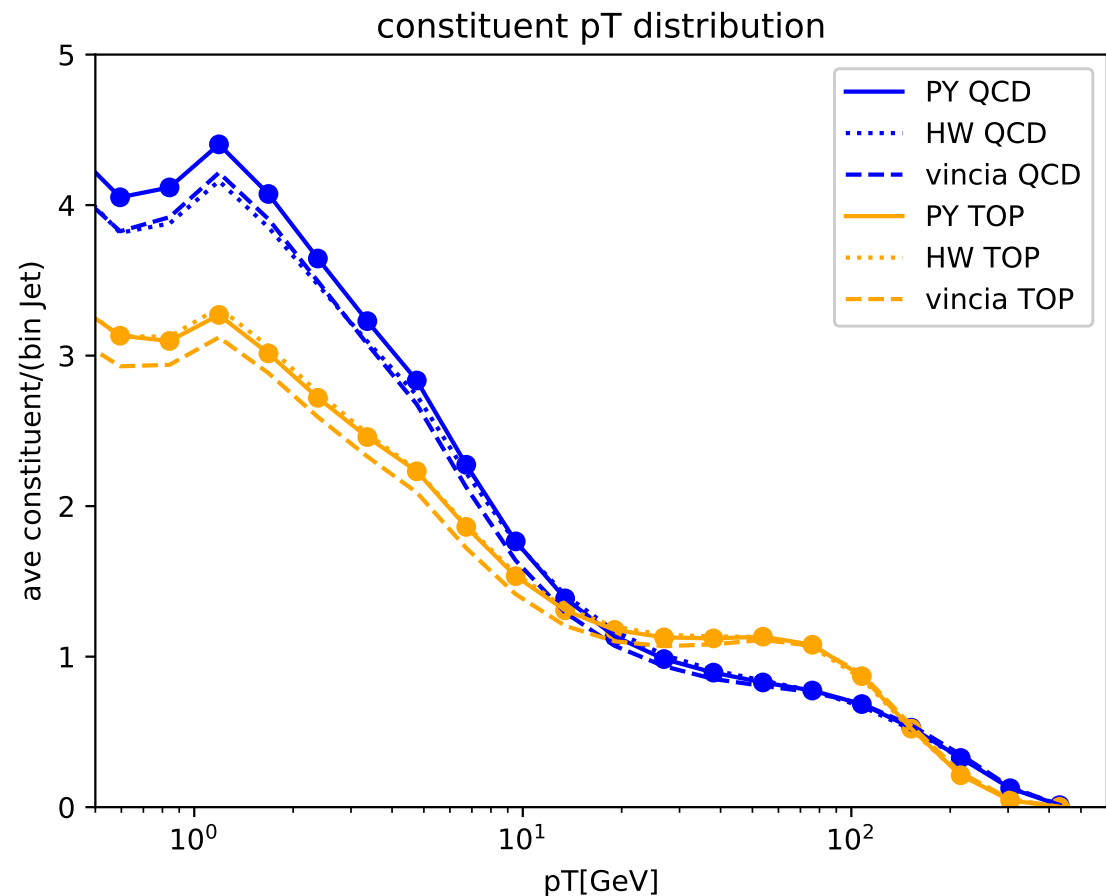
Subjet  
Constituent Multiplicity

Constituent PT  
histogram:  
capturing multiplicity  
conditioned on pt

Each subjet has different  
pt so this also provides  
extra information.

Color triplet subjets

Color singlet



# (Combined) Analysis Model

Jet Kinematics  
(PT, mass, ...)

Generalization of  
Constituent Multiplicity:  
Minkowski Functionals  
(Euler Char., Length, Area)

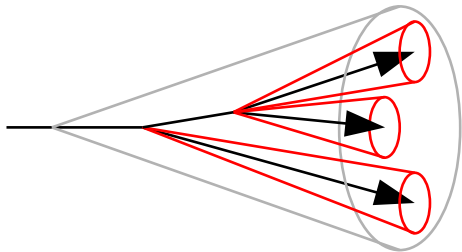
We will consider a NN  
analysing all these  
features.

Two-Point  
Energy Correlations S2  
(Relation Network)

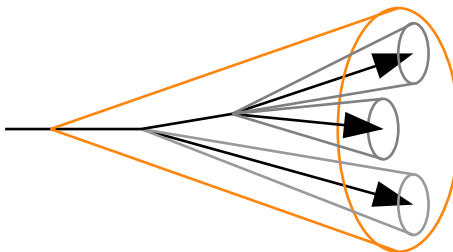
Subjet  
Constituent Multiplicity  
+ constituent PT histogram

All Top jet features  
below are covered by  
these inputs!

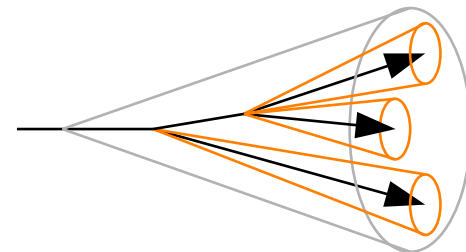
Three-prong



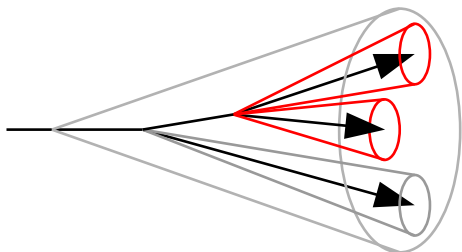
Color triplet



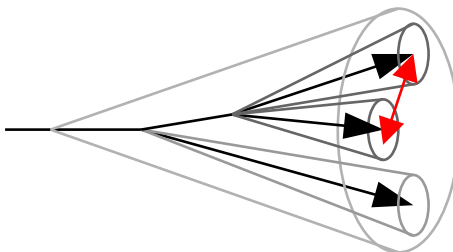
Color triplet subjets



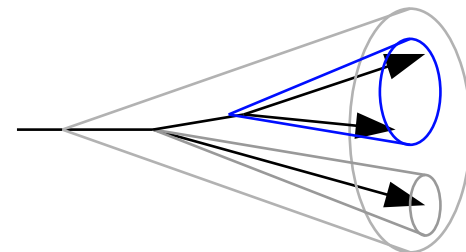
Two-prong subjet inside



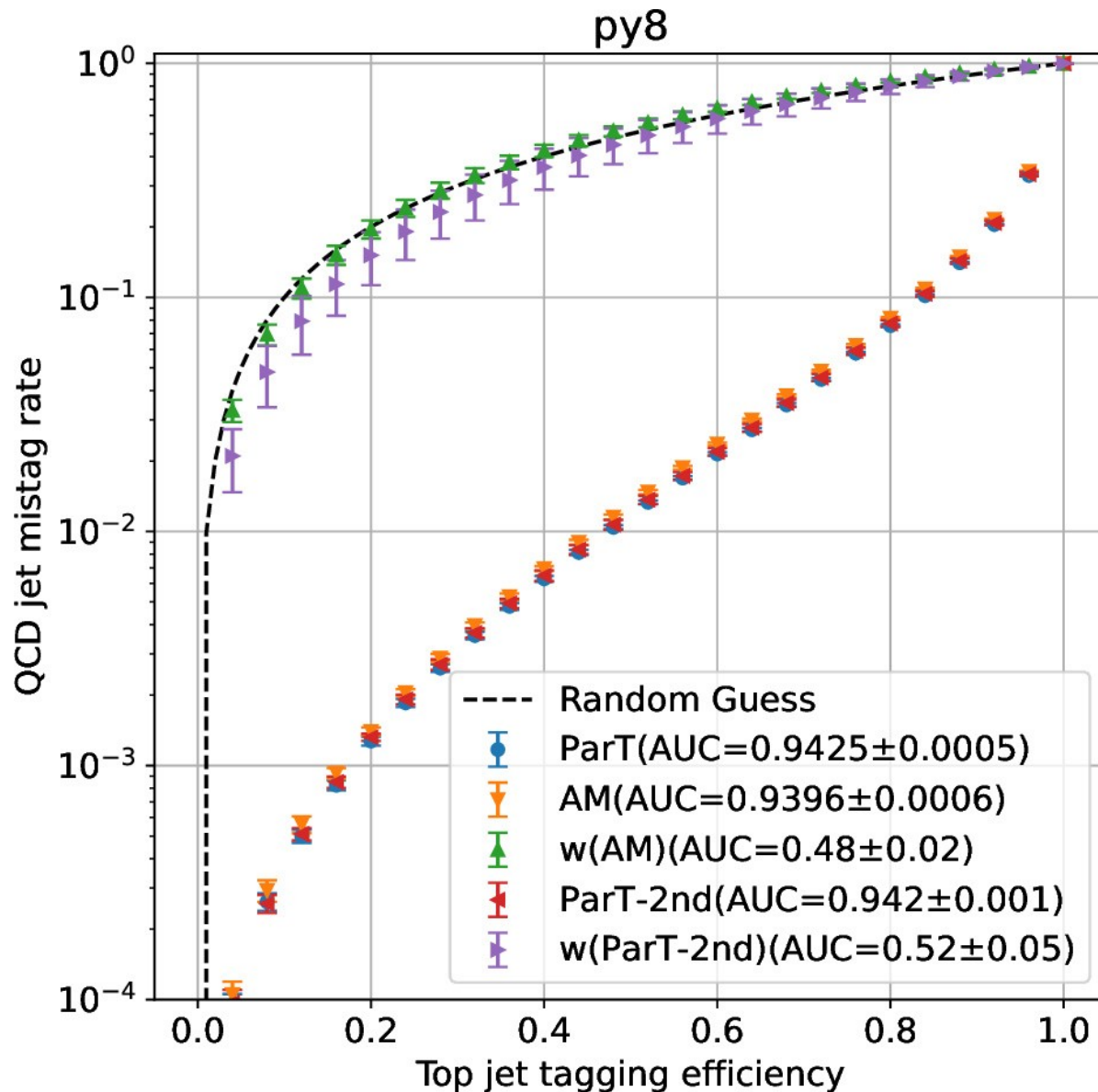
Color connection



Color singlet



# ROC curve



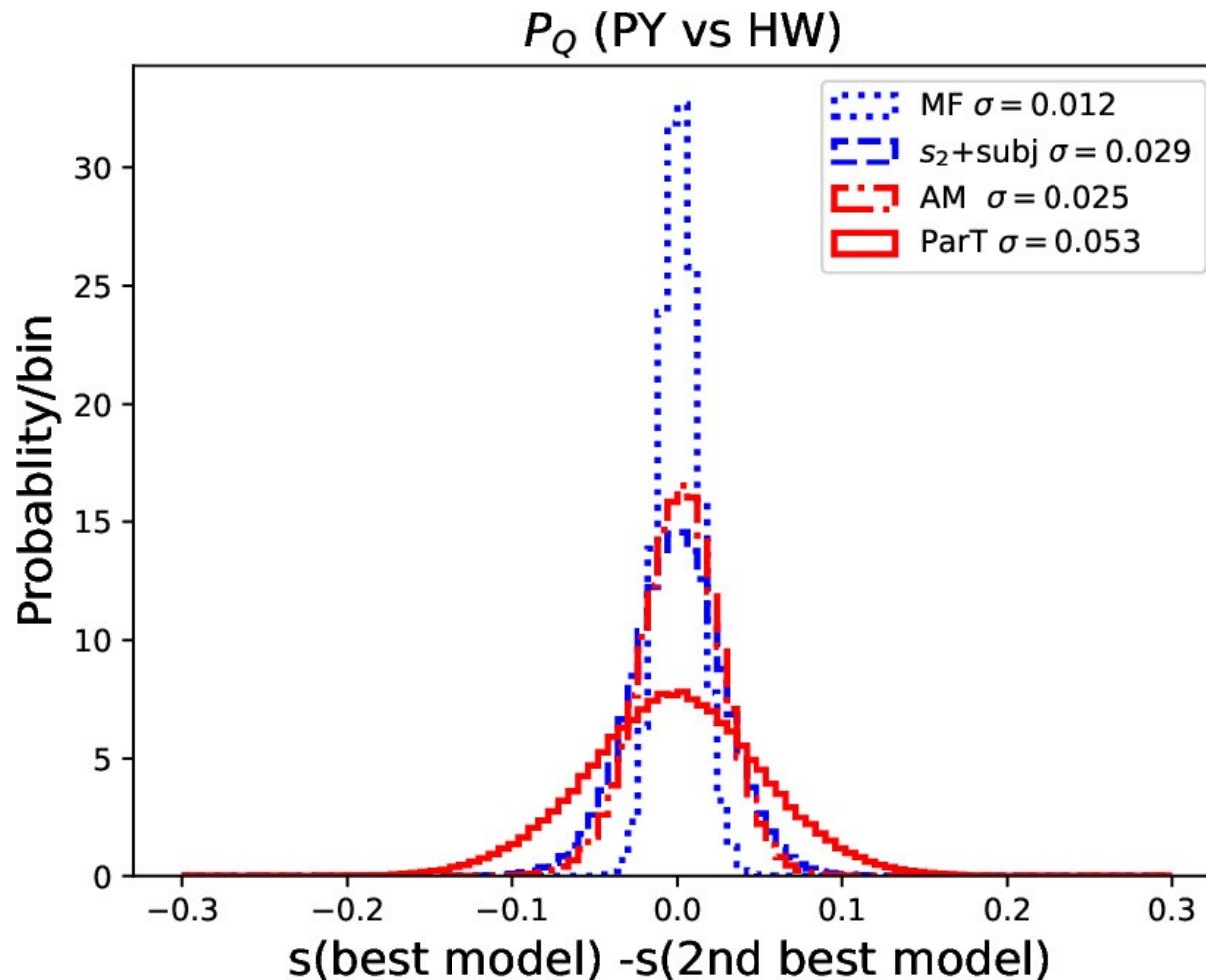
We compare the tagging performance of our analysis model to Particle Transformer working on pixellated jet constituents with HCAL resolution scale (0.1)

ROC curves are almost the same!

# Low variance!

---

One Advantage of using high-level feature based networks is low variance of training compared to low-level feature based networks.



Less training uncertainty in classifier training.

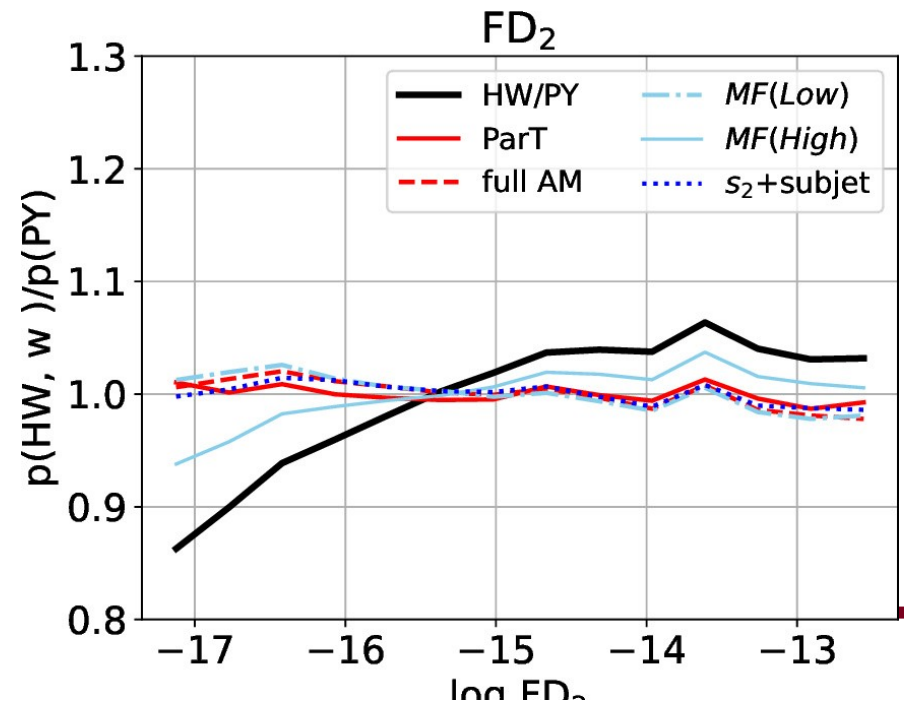
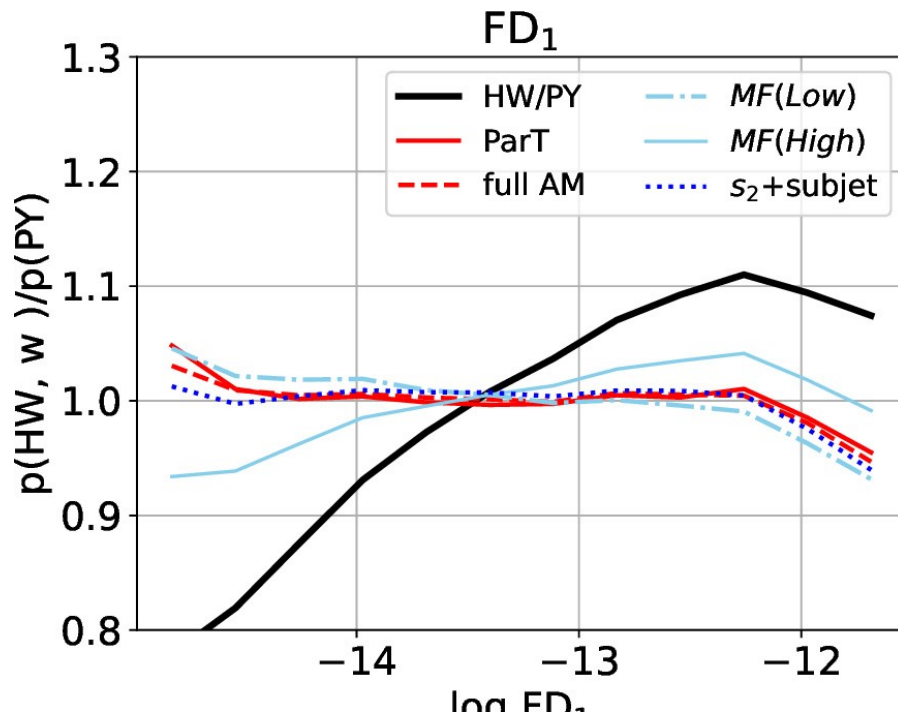
# Classifier-based sample reweighting

Low variance - high performance classifier is especially useful when we use classifier for reweighting MC samples for the calibration.

$$\hat{y}(x) = \frac{1}{1 + \frac{p_{\text{Data}}(x)}{p_{\text{MC}}(x)}}$$

Less training uncertainty on likelihood ratio estimation  
→ more accurate reweighting! Work in progress...

Reweight energy flow polynomial distributions important in top tagging  
(found by DisCo method: 2212.00046)



# Conclusion

---

- We introduced an analysis model for jet classification using two-point energy correlations and Minkowski functionals.
- We showed that this High-Level Feature based classifier shows competitive performance compared to the state-of-the-art classifiers such as ParticleNet and ParticleTransformers. at HCAL resolution scale,
- Our method is more constrained setup than those SotA methods without losing tagging performance much, we have less training uncertainty.
- Less training uncertainty is valuable especially when using classifier as density-ratio estimator, and using it for re-weighting Monte Carlo generated samples.

---

# Backups

# IRC-safe energy correlator based Neural Networks

Graph Networks

Relation Network

Utilizes edge features

$$F\left[\sum_{i,j \in J} \phi^e(p_i, p_j)\right]$$

Raposo, et al. (1702.05068),  
Santoro, et al. (1706.01427)

IRC safety

IRC-safe energy correlator based Networks

Relation Network

Utilizes two-point energy correlation

$$F\left[\sum_{i,j \in J} p_{T,i} p_{T,j} \phi^e(R_{ij})\right]$$

Chakraborty, **SHL**, Nojiri, and Takeuchi (2003.11787)

Deep Sets  
(Particle Flow Network)

Utilizes vertex features

$$F\left[\sum_{i \in J} \phi^v(p_i)\right]$$

Zaheer, et al. (1703.06114)

IRC safety

Energy Flow Network

Utilizes one-point energy correlation:  
permutation invariant  
energy-weighted linear sum of angular map

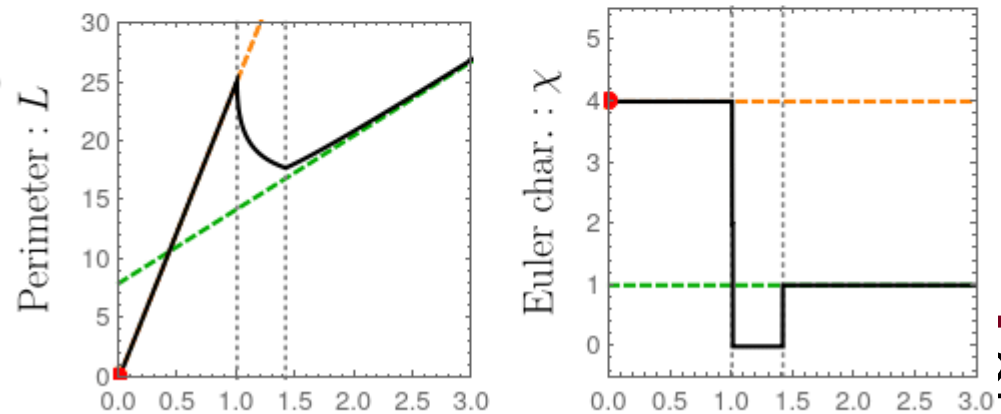
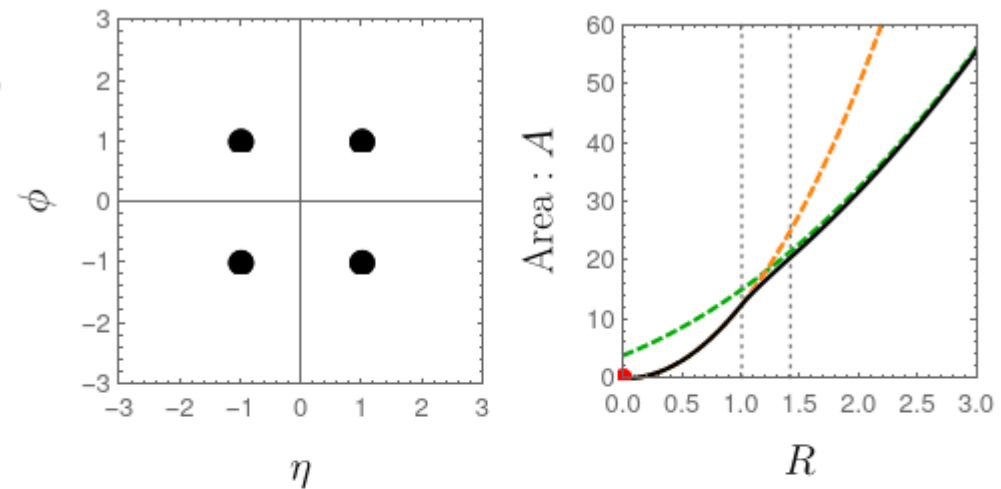
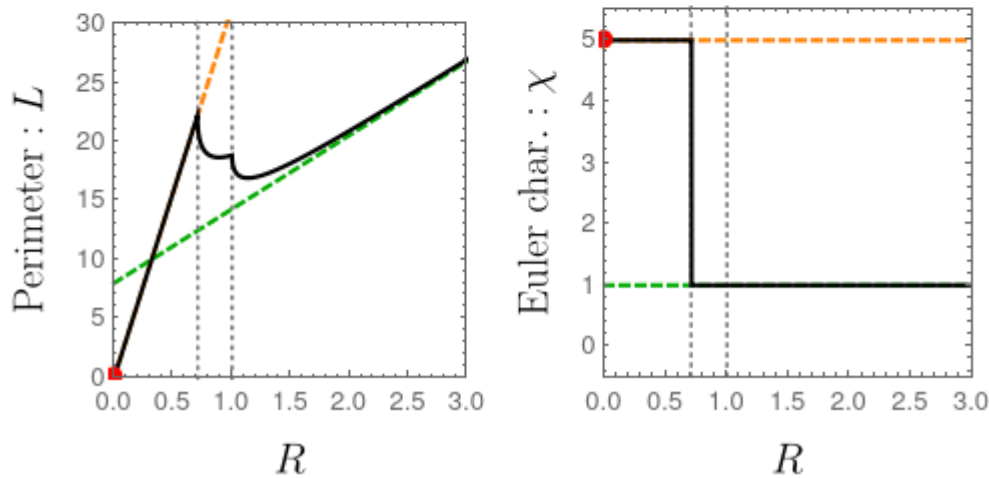
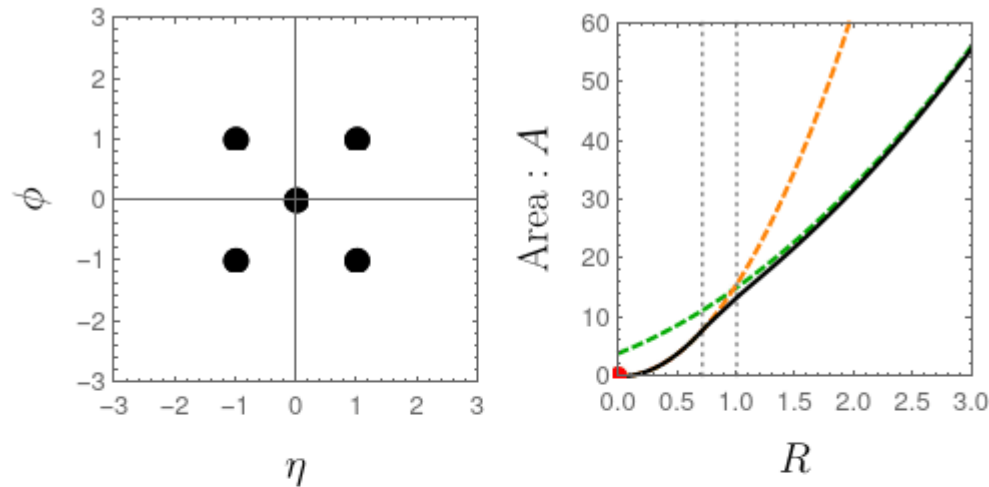
$$F\left[\sum_{i \in J} p_{T,i} \phi^v(\vec{R}_i)\right]$$

Komiske, Metodiev, and Thaler (1810.05165)



# Mathematical Morphology and Minkowski Functionals

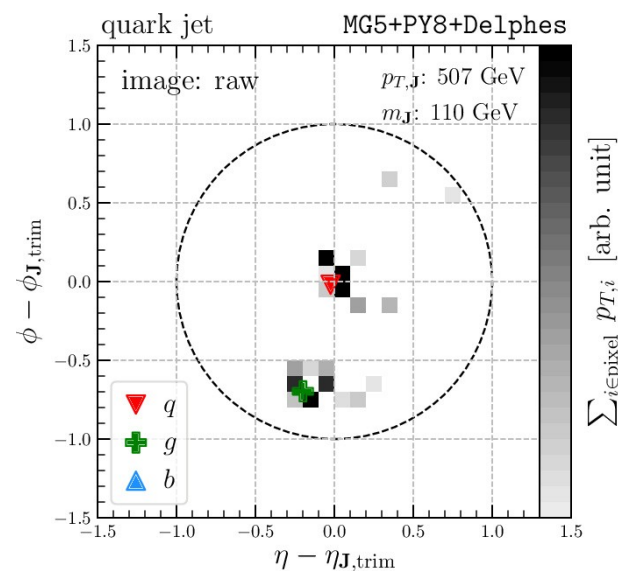
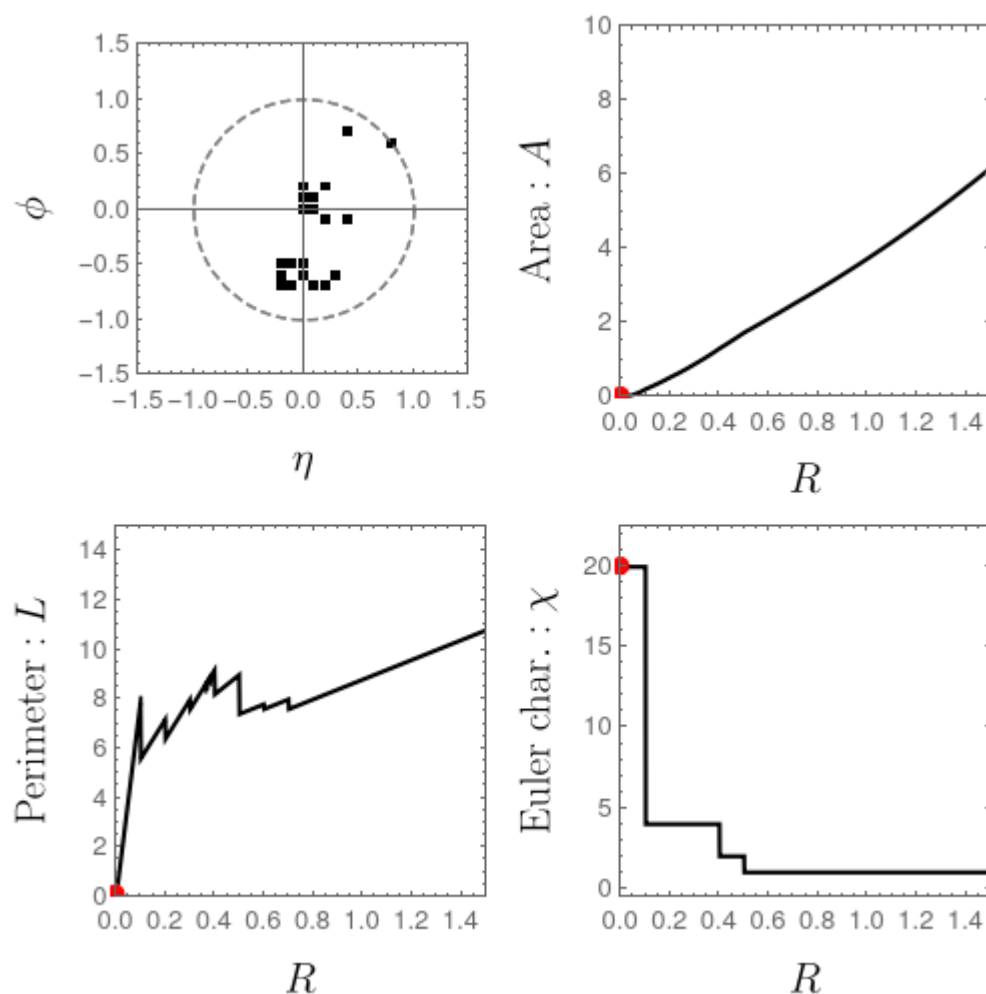
A similar point distribution with a different topology may exist, but the MFs could capture the fine difference.



**Orange:** asymptote as  $r \rightarrow 0$   
**Green:** asymptote as  $r \rightarrow \infty$

# Morphological Analysis on (pixellated) Jet Image

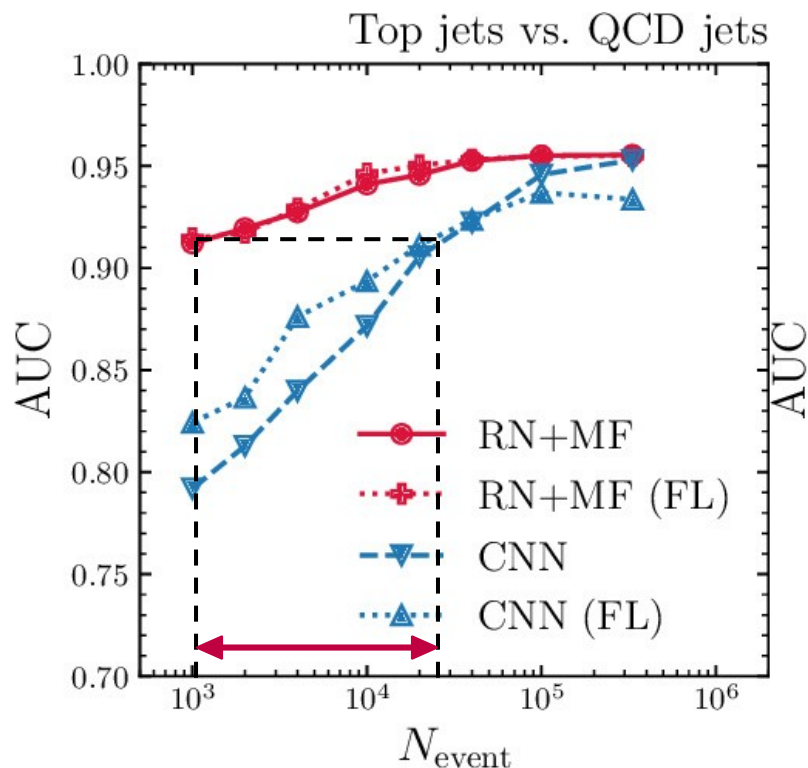
In the case of the analysis on jet images, we use squares for the dilation in order to preserve underlying geometry of the data.



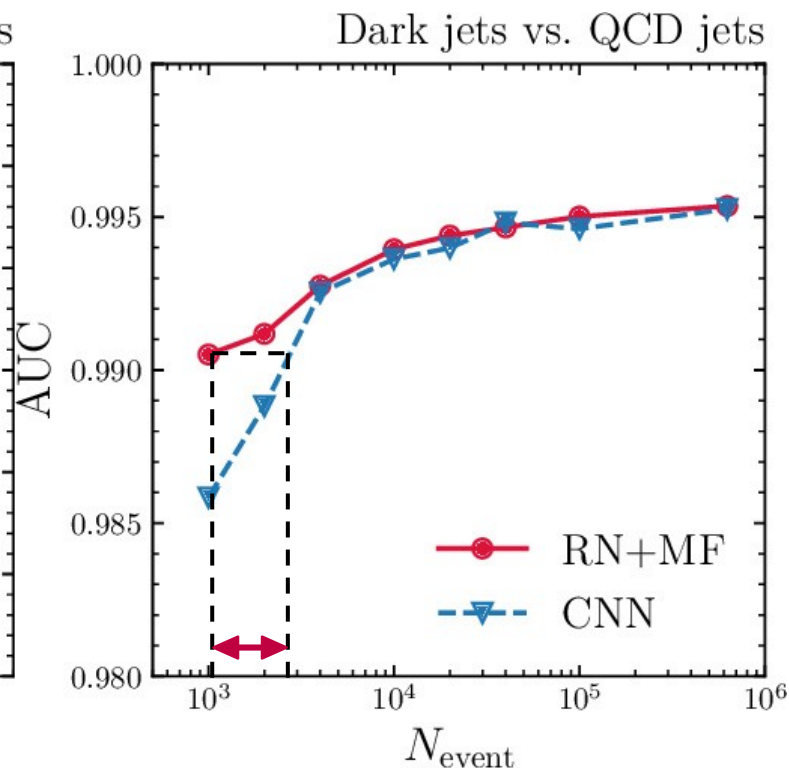
One interesting property of this setup is that all the calculation steps of the MFs can be written in terms of **discrete convolutions**.

# Constrained Architectures and Low-Shot Learning

We showed that our RN+MF has comparable performance to the CNN. Moreover, it has advantages when the dataset is small, because RN+MF is more constrained architecture than the CNN.



RN+MF is much less sample-demanding thanks to its constraints.



A smaller factor 3 gap is here, but this example has an exclusive phase space region parameterized by MFs.

# Sample description

---

- All the SM jets are simulated by MG5+pythia8.3
- Dark jets are simulated by pythia8.3
- Top jet vs. QCD jet
  - Jet constituents: Delphes EFlows
  - $PT \in [500, 600]$  GeV
  - $Mass \in [150, 200]$  GeV
  - Leading pt anti-kt jets with radius 1.0
  - For top jets, all the originating b-quarks and quarks must be within jet radius 1.0 from the jet center.