Anatomy of Jet Classification using Deep Learning

Sung Hak Lim
Rutgers University

ML4Jets 2023,
DESY, Hamburg, Germany

Nov. 2023

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

Current Status of Jet Taggers

In the yesterday's experiment overview talk by Kevin, we saw that the current state-of-art jet taggers are **neural networks** analyzing **low level inputs** (jet constituent features) directly.

Classification

arXiv:2202.03772

Analysis

	All classes		$H \to b\bar{b}$	$H \to c\bar{c}$	$H \to gg$	$H \to 4q$	$H \to \ell\nu qq'$	$t \to bqq'$	$t \to b\ell\nu$	$W \to qq'$	$Z \to q\bar{q}$
	Accuracy	AUC	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{99\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{99.5\%}$	$\text{Rej}_{50\%}$	$\text{Rej}_{50\%}$
PFN	0.772	0.9714	2924	841	75	198	265	797	721	189	159
P-CNN	0.809	0.9789	4890	1276	88	474	947	2907	2304	241	204
ParticleNet	0.844	0.9849	7634	2475	104	954	3339	10526	11173	347	283
ParT	**0.861**	**0.9877**	**10638**	**4149**	**123**	**1864**	**5479**	**32787**	**15873**	**543**	**402**
ParT (plain)	0.849	0.9859	9569	2911	112	1185	3868	17699	12987	384	311

+ PELICAN

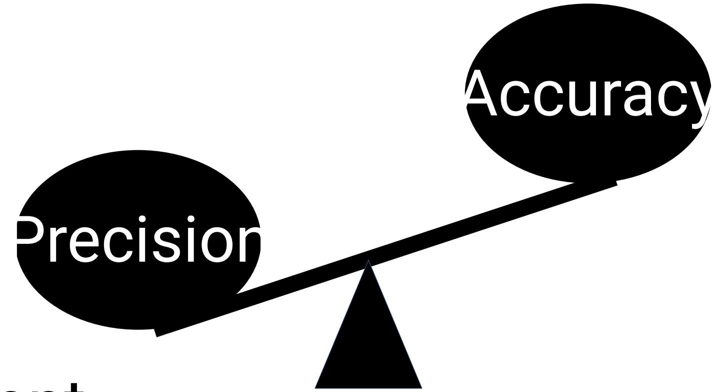
As a HEP theorist, one problem that I want to discuss:

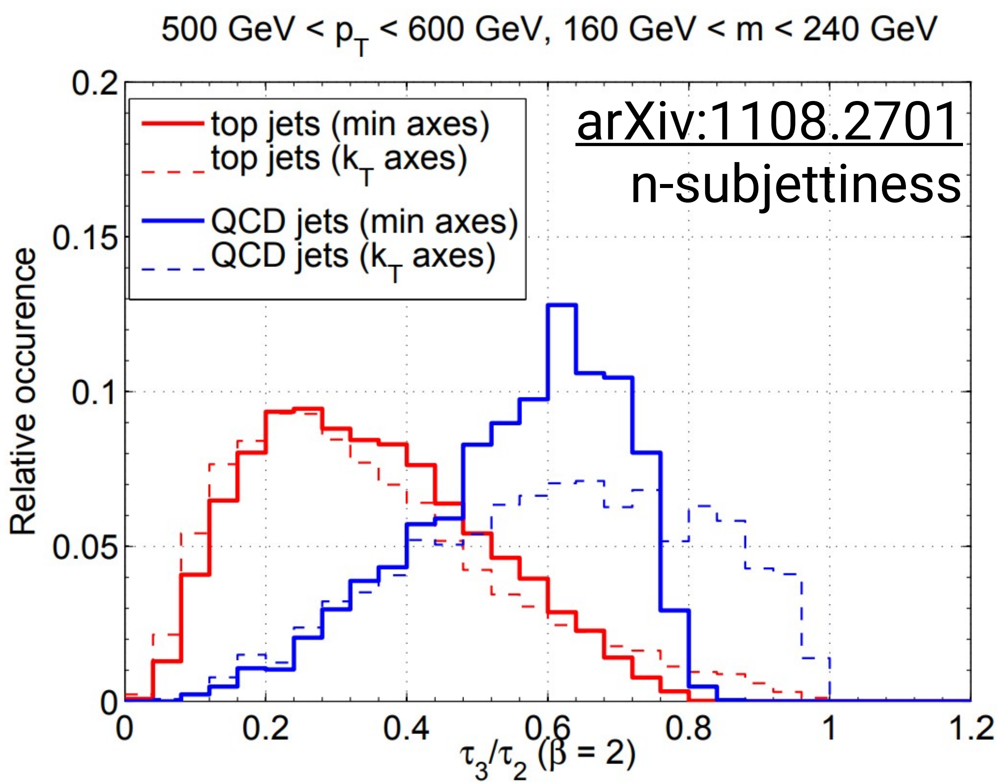
Can we build up a **high-level feature based jet tagger** equally performing well?

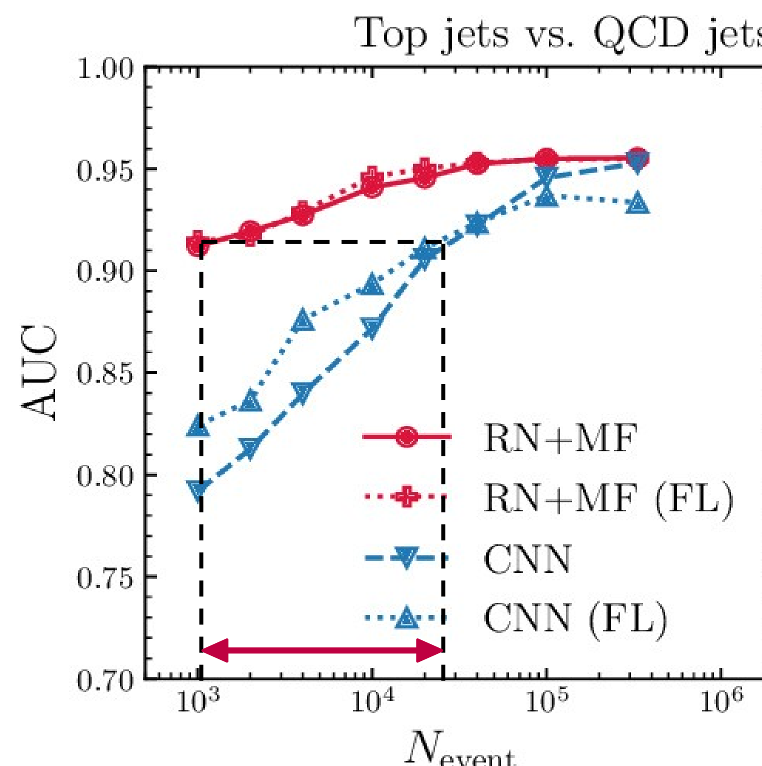
If we could do that, what are the advantages of such HLF based tagger?

Advantages of High Level Feature based Jet Taggers

- Interpretable (by understanding HLF inputs)
- Advantages from Bias-Variance tradeoff
  - Less training uncertainty
  - Less sample demanding
- Faster in evaluation, memory efficient
  - Simple networks (such as MLP) are sufficient.



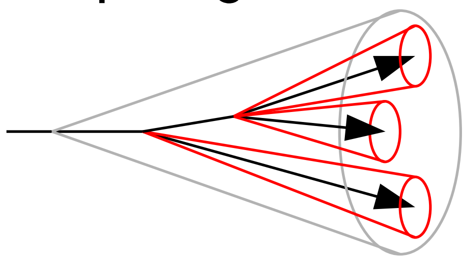




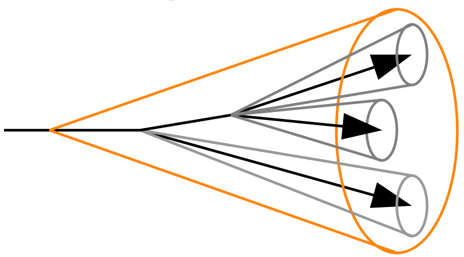
Anatomy of Top Jets

In order to build an **<u>high performing HLF based top jet tagger</u>**, we have to build up HLFs capturing the all <u>features of **top jets** completely</u>.
What are the features of top jets?

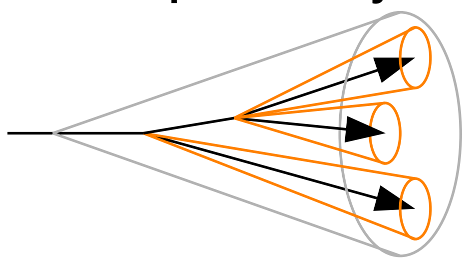
Three-prong



Color triplet

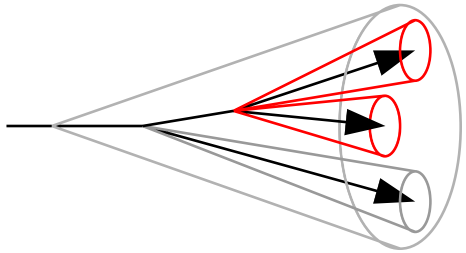


Color triplet subjets

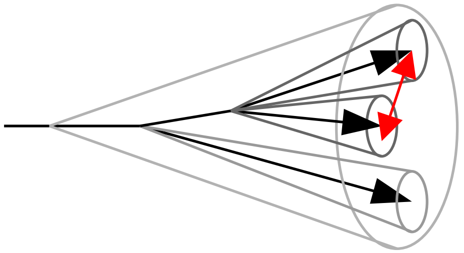


Top jet also have **<u>W boson jet</u>** inside.

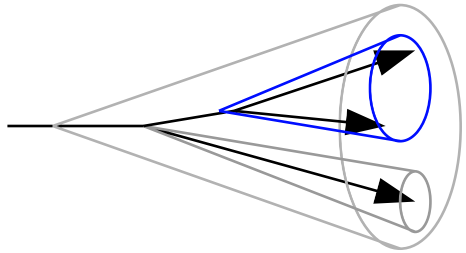
Two-prong subjet inside



Color connection



Color singlet



We will introduce an **<u>analysis model</u>** combining HLF analyzing architectures specialized for analyzing the above features.

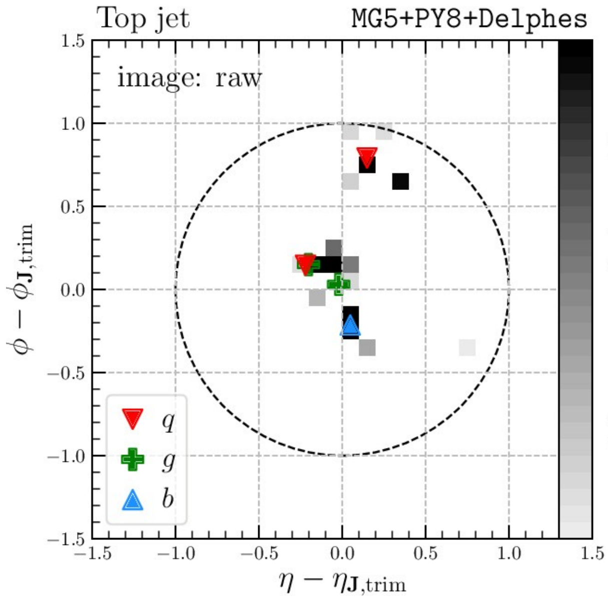
Two-point energy correlation spectrum

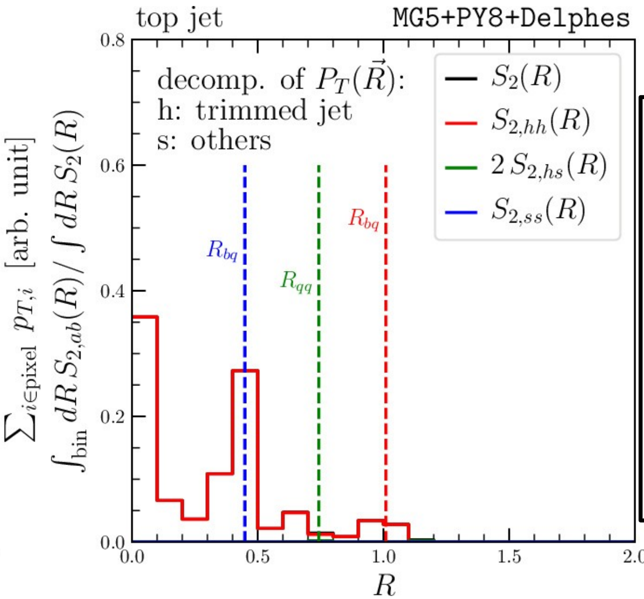
SHL, M. M. Nojiri, 1807.03312
A. Chakraborty, **SHL**, M. M. Nojiri,1904.02092
A. Chakraborty, **SHL**, M. M. Nojiri, M. Takeuchi, 2003.11787

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 S2 as inputs to MLP.
The resulting network is called Relation Network,
a type of GNN using only edge features.

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

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

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

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

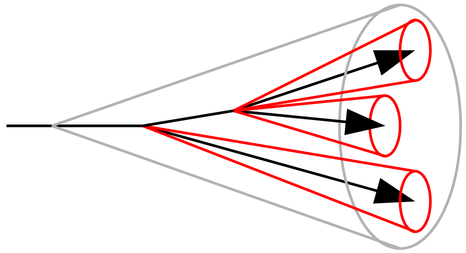
Chakraborty, **SHL**, Nojiri, and Takeuchi (2003.11787)

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

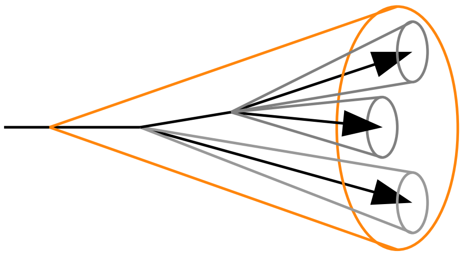
What else do we need?

Boosted Top jets is very rich in characteristic features

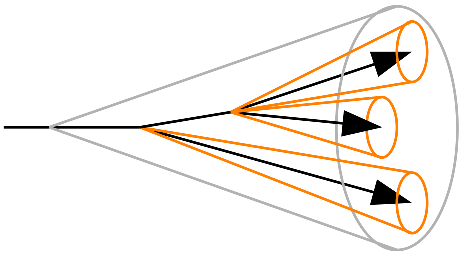
Three-prong



Color triplet

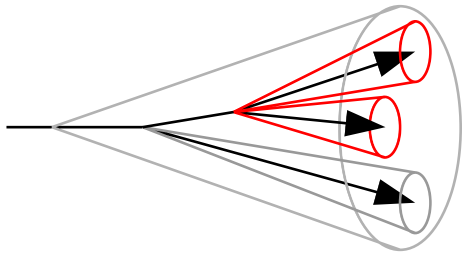


Color triplet subjets

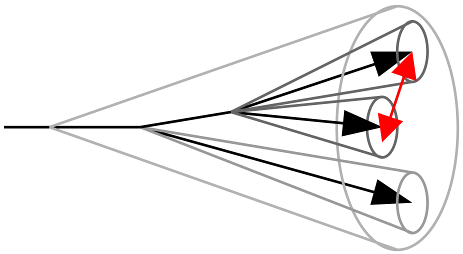


Top jet also have <u>**W boson jet**</u> inside.

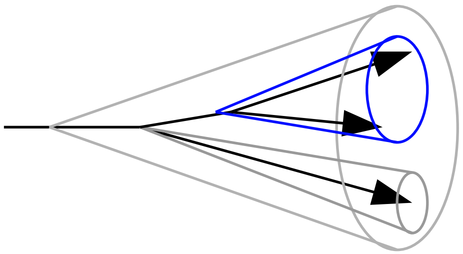
Two-prong subjet inside



Color connection



Color singlet

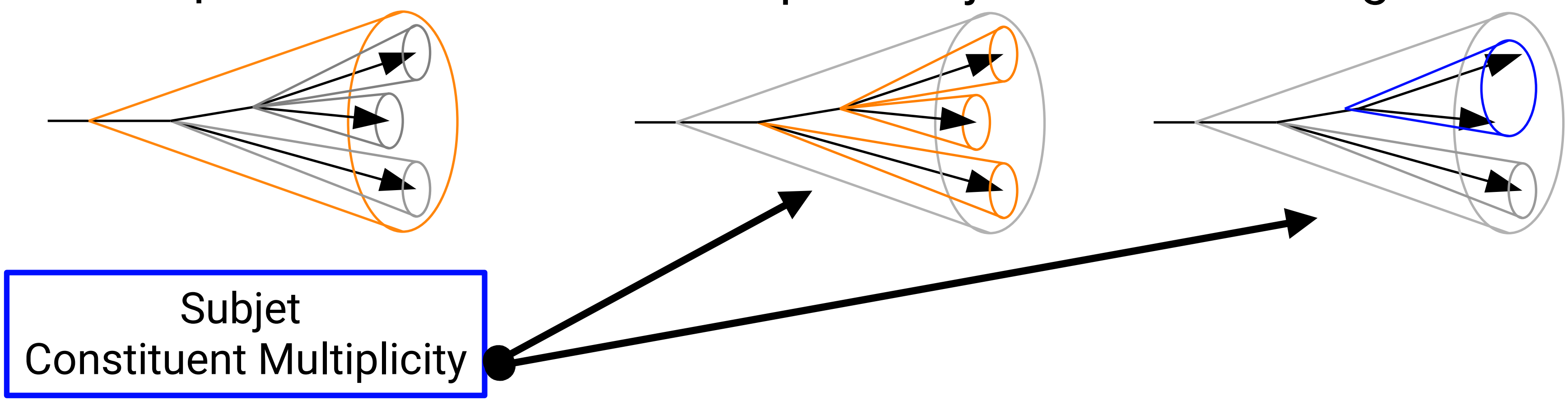


<u>**Color charge senstive variable**</u> → constituent multiplicity.

Subjet color charges

Constituent multiplicity is sensitive to the color charge of originating parton of jet. (IRC unsafe)

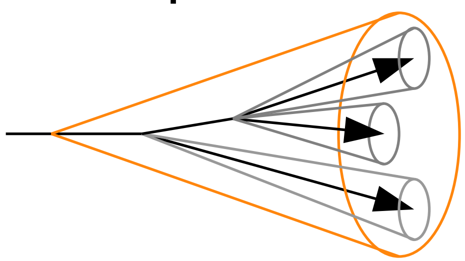
Color triplet | Color triplet subjets | Color singlet



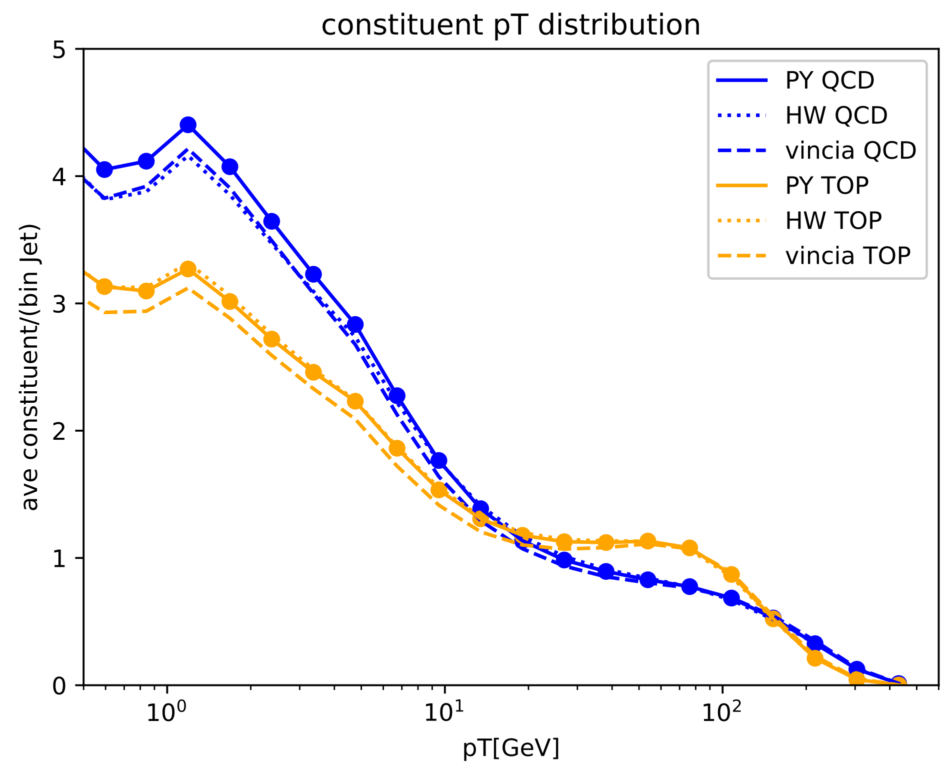
Subjet color charges

Constituent multiplicity is sensitive to the color charge of originating parton of jet. (IRC unsafe)

Color triplet



Colo



Subjet
Constituent Multiplicity

Constituent PT
histogram:
capturing multiplicity
conditioned on pt

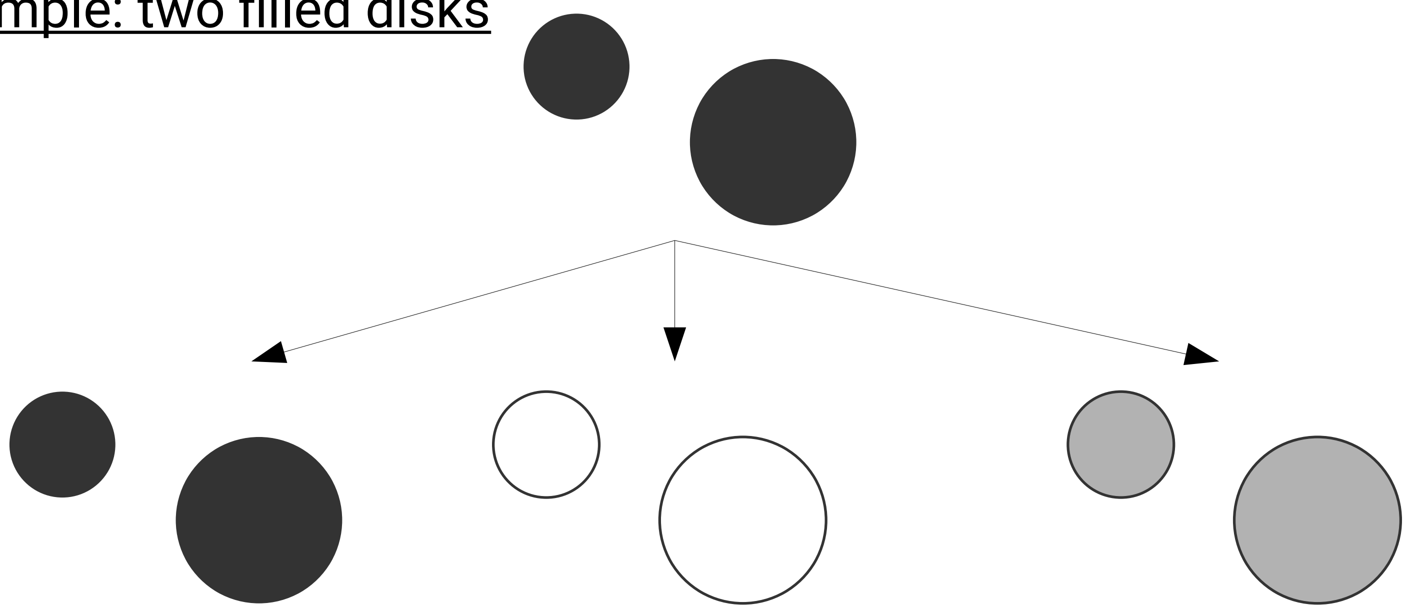
This counting variable analysis seems good, but it can be further extended

Minkowkski Functionals: a generalization of counting observables

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

$$\mathrm{MF}_0 = \int_A d^2\vec{r}$$

Boundary length

$$\mathrm{MF}_1 = \frac{1}{2\pi}\int_{\partial A} d\vec{r}$$

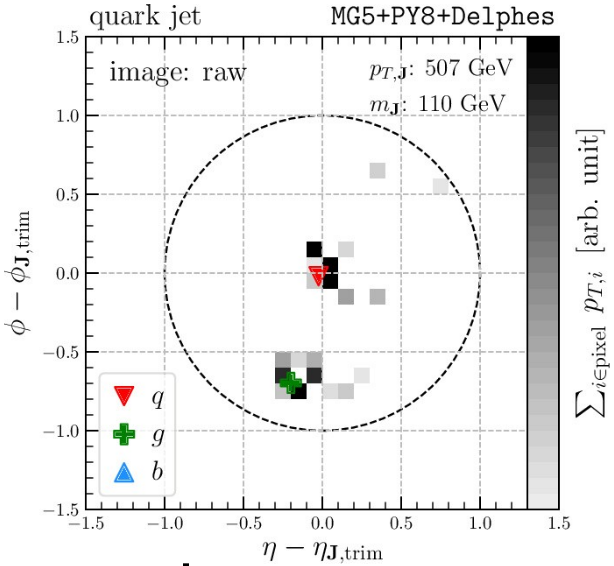
Euler characteristic

(Gauss-Bonnet)

$$\mathrm{MF}_2 = \frac{1}{2\pi^2}\int_{\partial A} \frac{1}{R} d\vec{r}$$

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

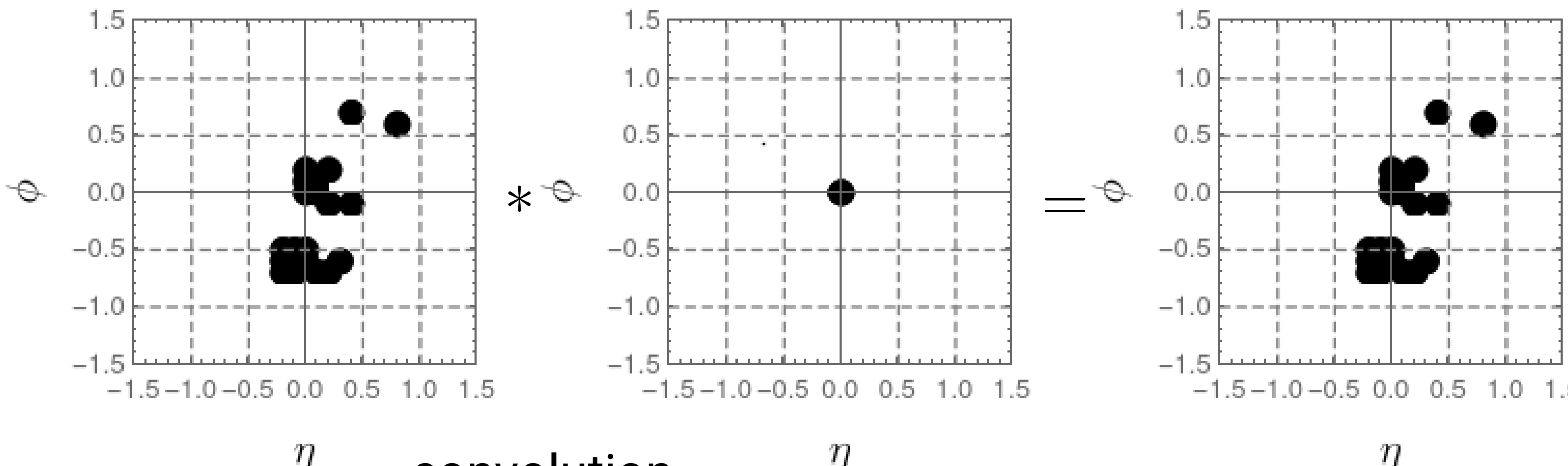
Mathematical Morphology: Minkowski Functionals and Dilation



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

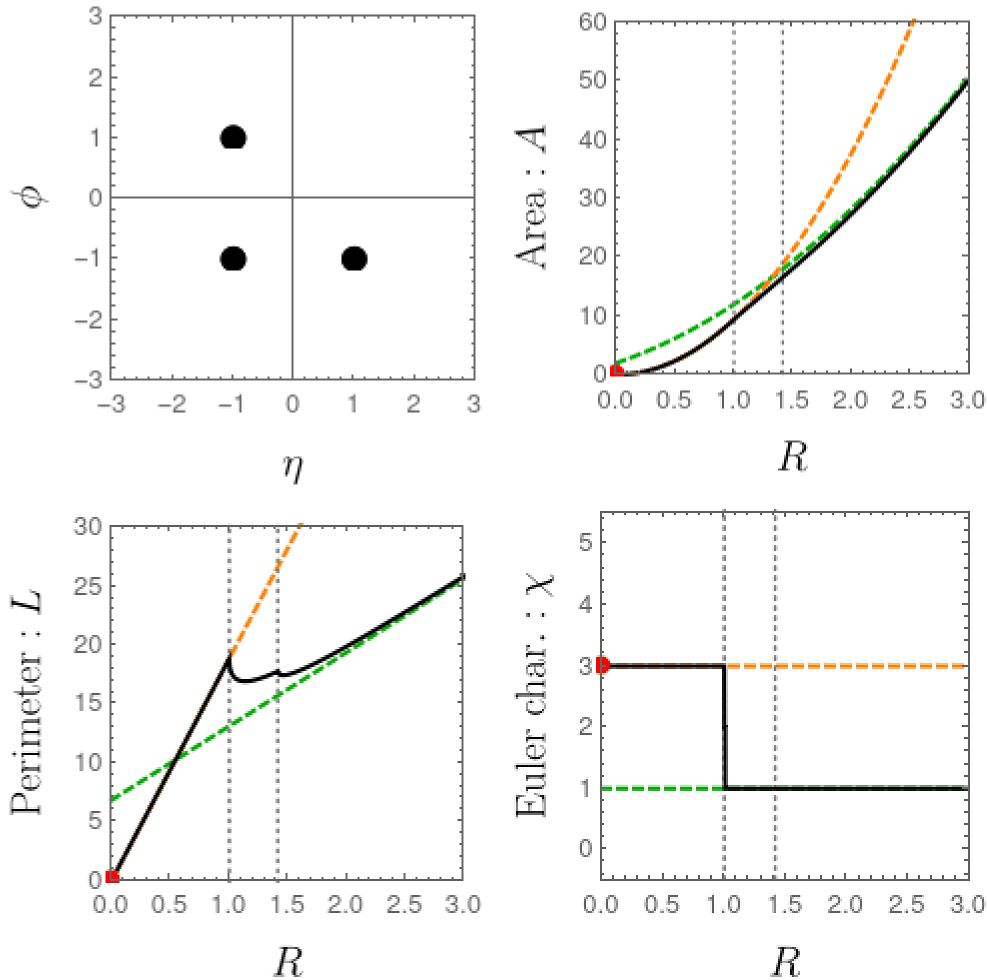
This morphological operation may be regarded as coarse-graining,
and it will allow us to systematically analyze the **geometry of jet constituents** when we use this together with the MFs.

Binarize



convolution

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: asymtote as r → 0
Green: asymtote as r → infinity

(Combined) Analysis Model

Jet Kinematics
(PT, mass, ...)

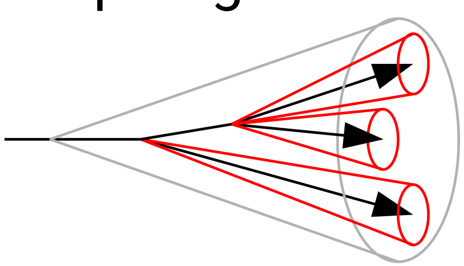
Generalization of
Constituent Multiplicity:
Minkowski Functionals
(Euler Char., Length, Area)

Two-Point
Energy Correlations S2
(Relation Network)

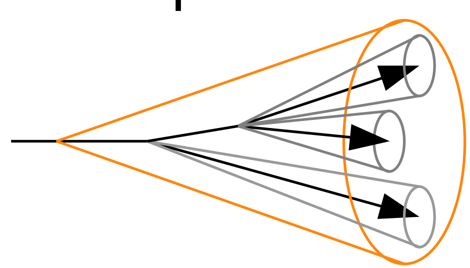
Subjet
Constituent Multiplicity
+ constituent PT histogram

We will consider a NN analysing all these features.
All Top jet features below are covered by these inputs!

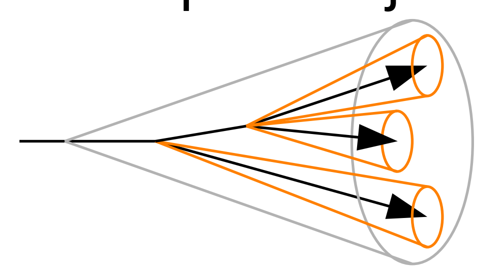
Three-prong



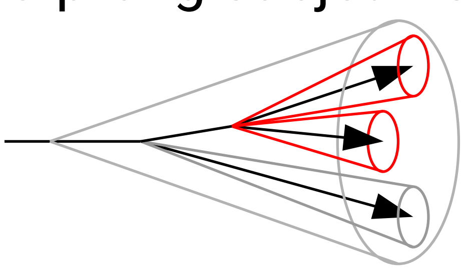
Color triplet



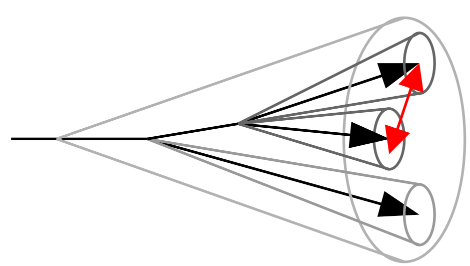
Color triplet subjets



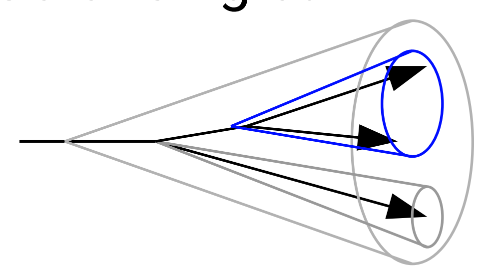
Two-prong subjet inside



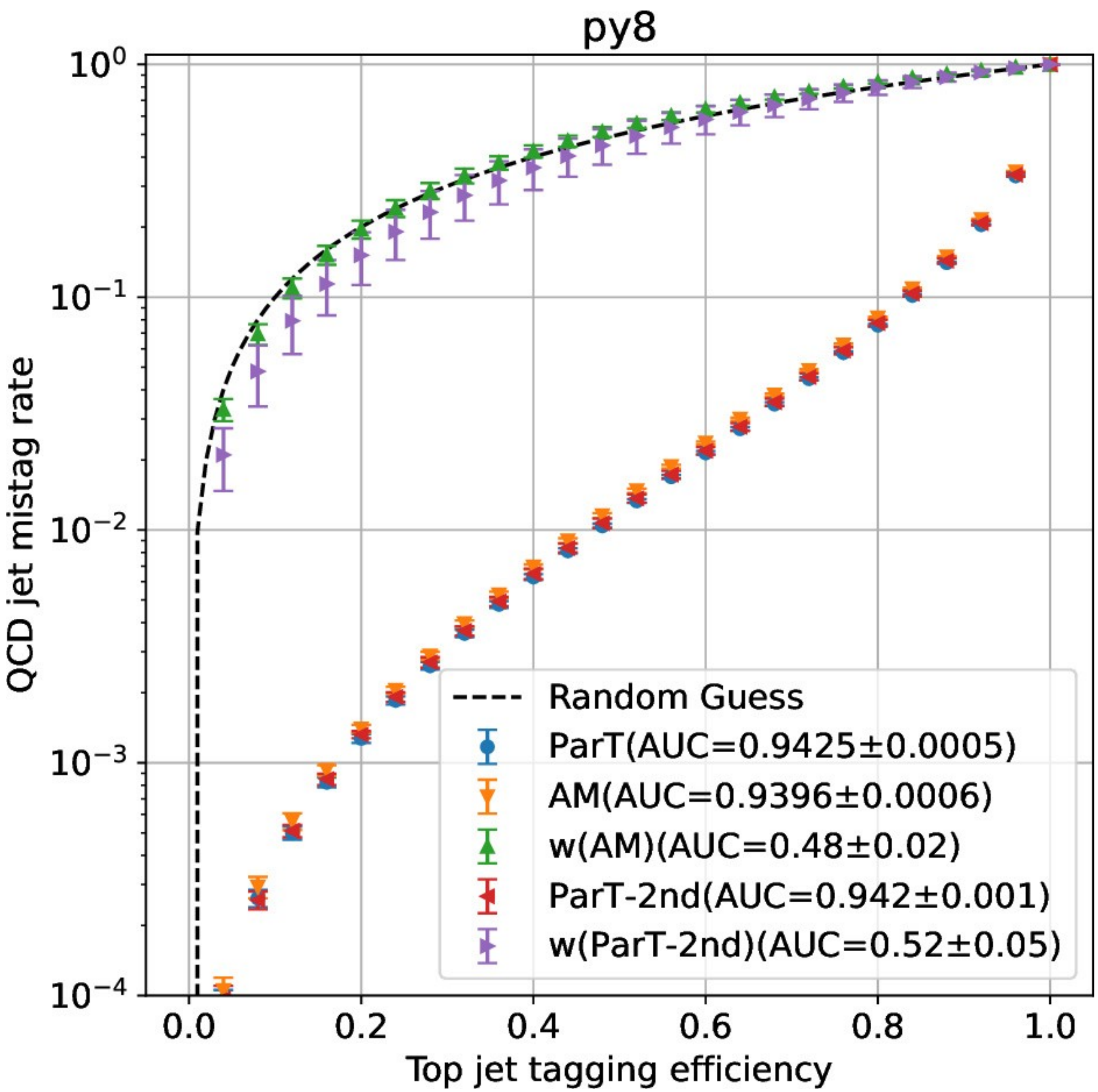
Color connection



Color singlet



ROC curve

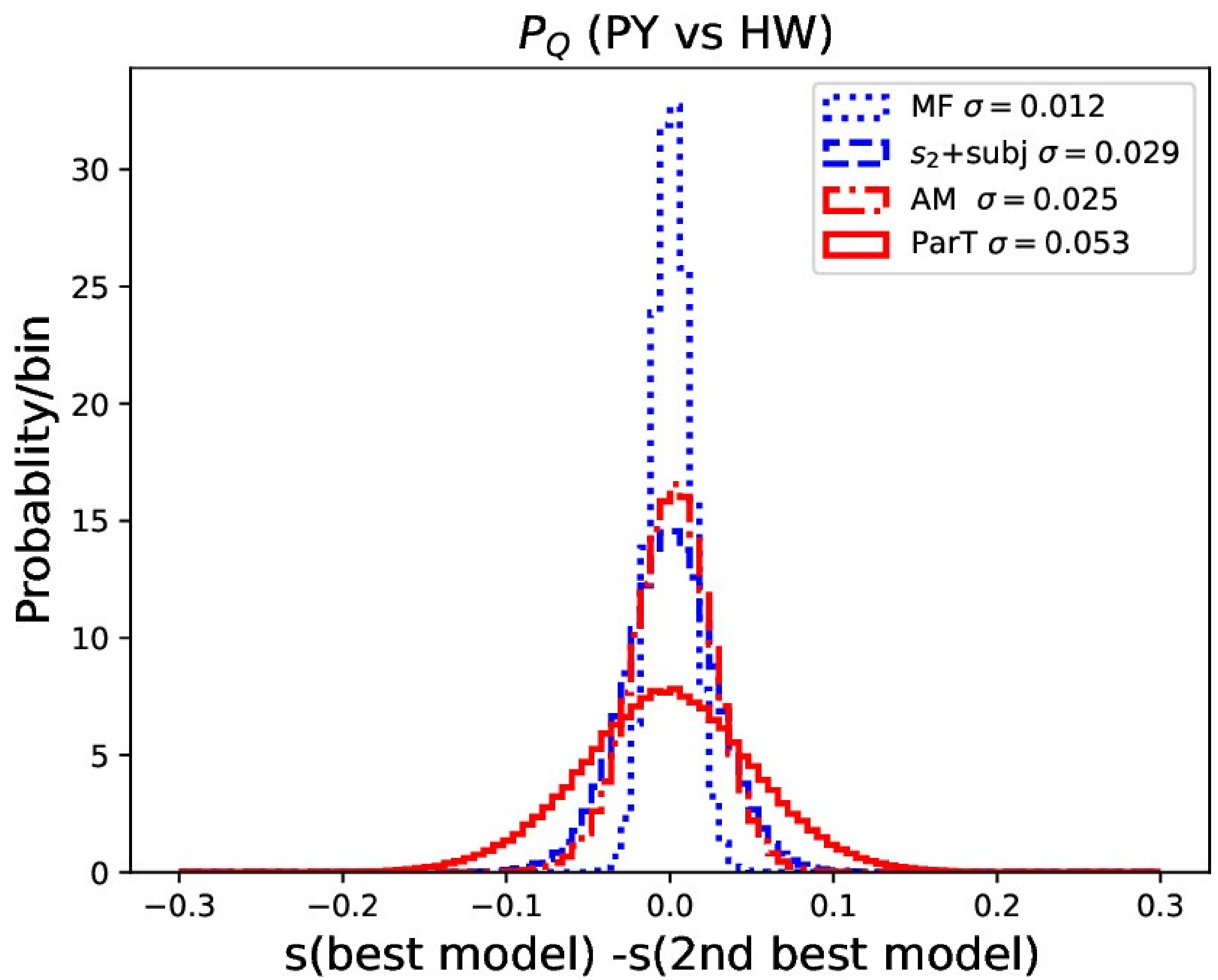


We compare the tagging performance of our analysis model to Particle Transformer working on pixellated jet constituents with HCAL resolution scale (0.1)

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.

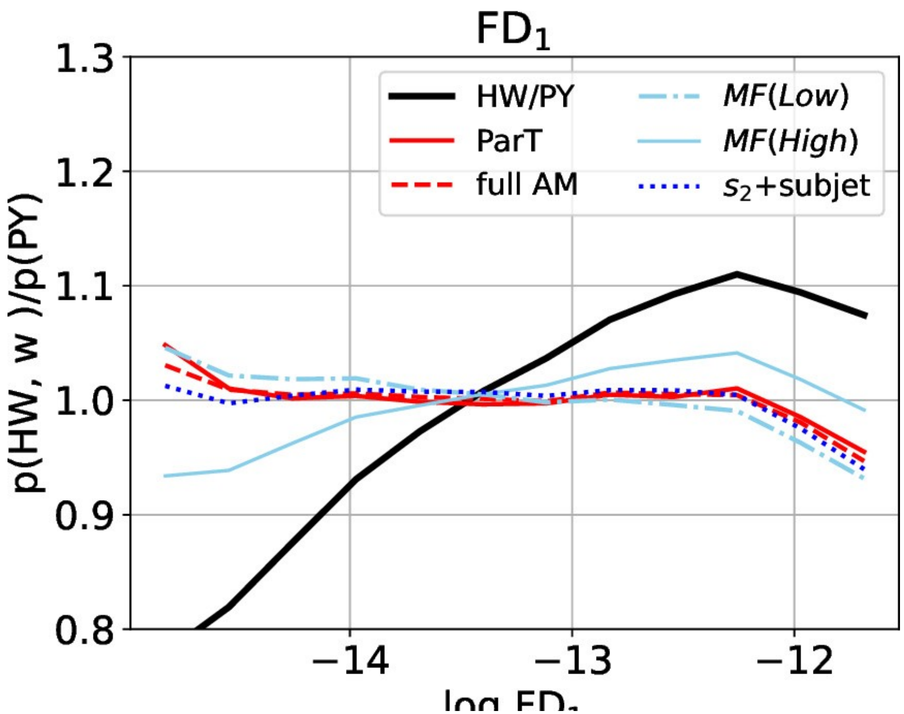
Classifier-based sample reweighting

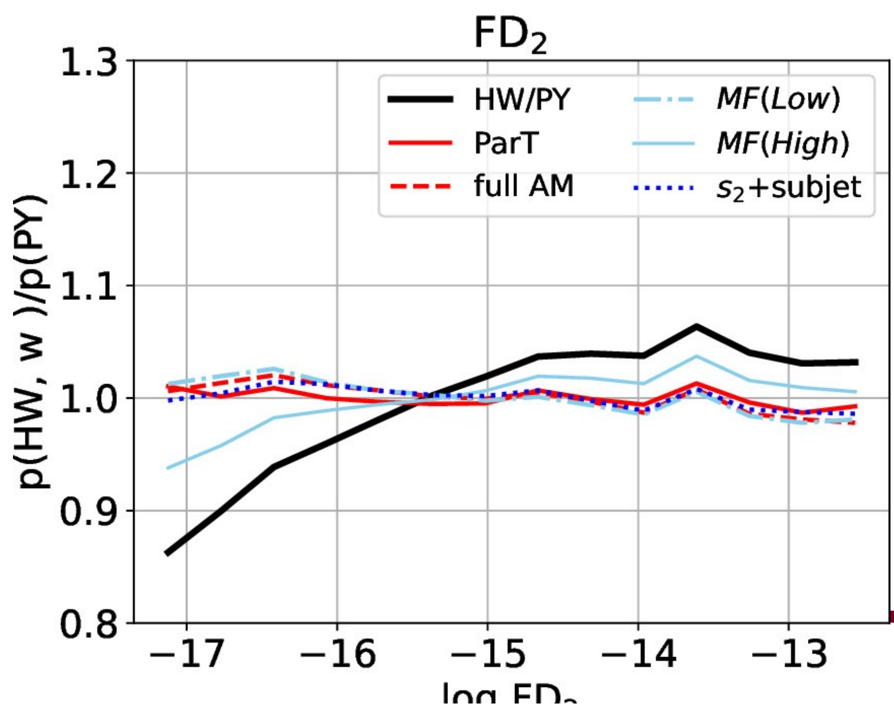
Low variance - high performance classifier is especially useful when we use classifier for reweighting MC samples for the calibration.

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

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

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





Conclusion

- We introduced an analysis model for jet classfication using two-point energy correlations and Minkowski functionals.
- We showed than this High-Level Feature based classifier shows competetive 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

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

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

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

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

Deep Sets (Particle Flow Network)

Utilizes vertex features

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

Zaheer, et al. (1703.06114)

IRC safety

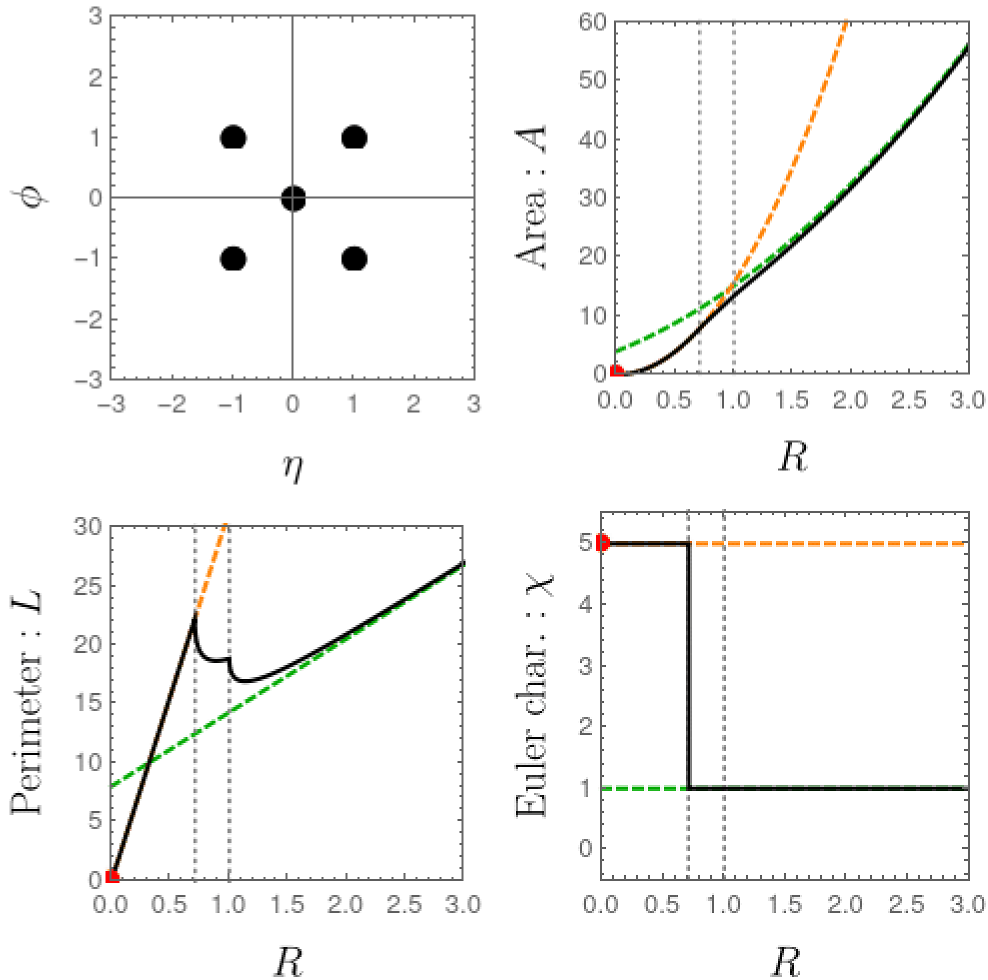
Energy Flow Network

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

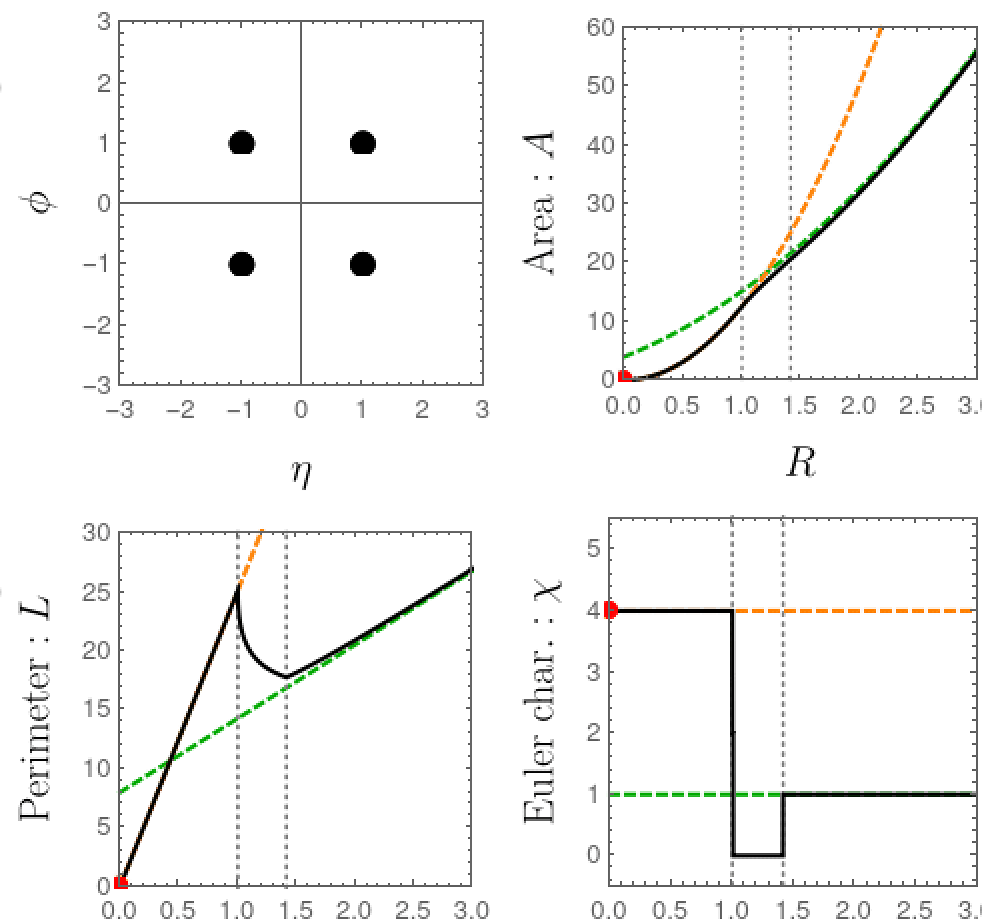
$$F[\sum_{i\in J} p_{T,i}\phi^v(\vec{R}_i)]$$

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.

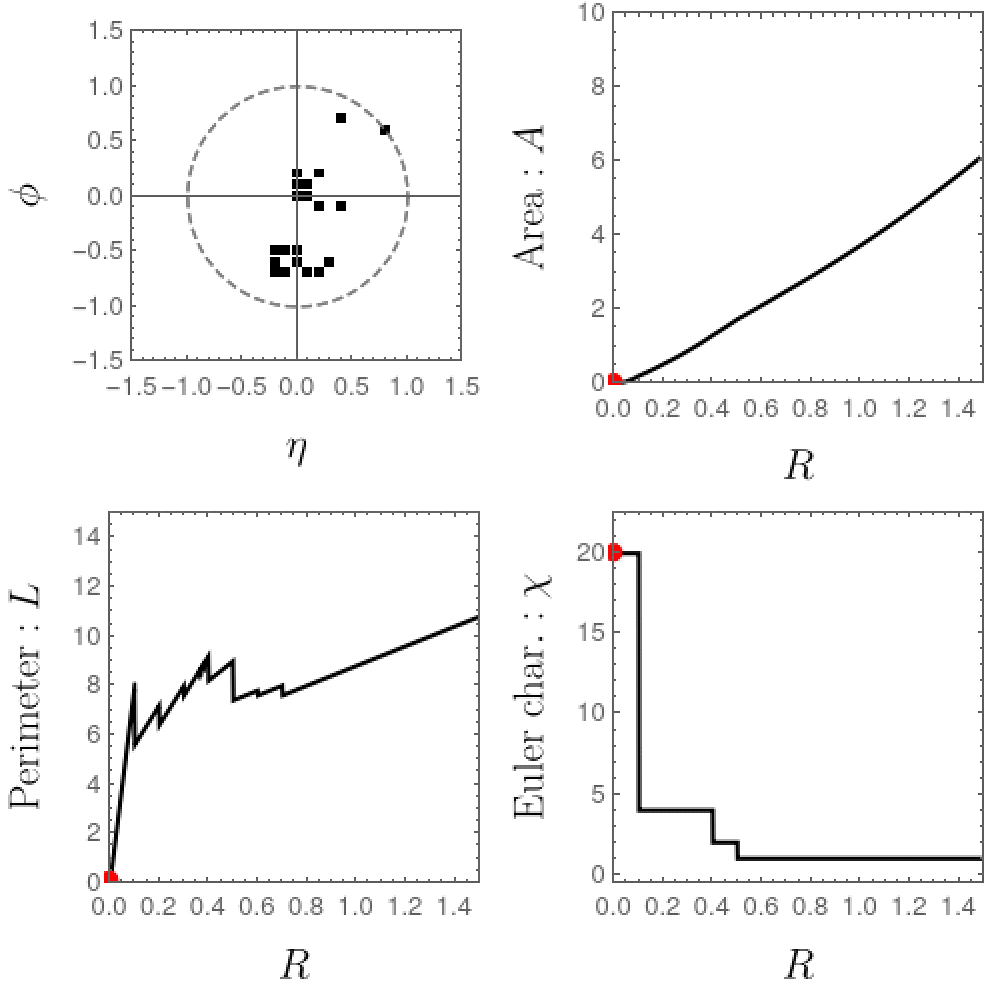


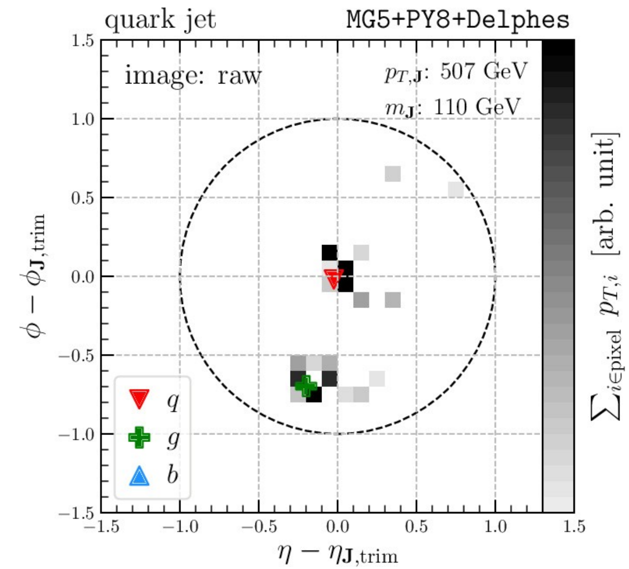
Orange: asymtote as r → 0

Green: asymtote as r → infinity

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.

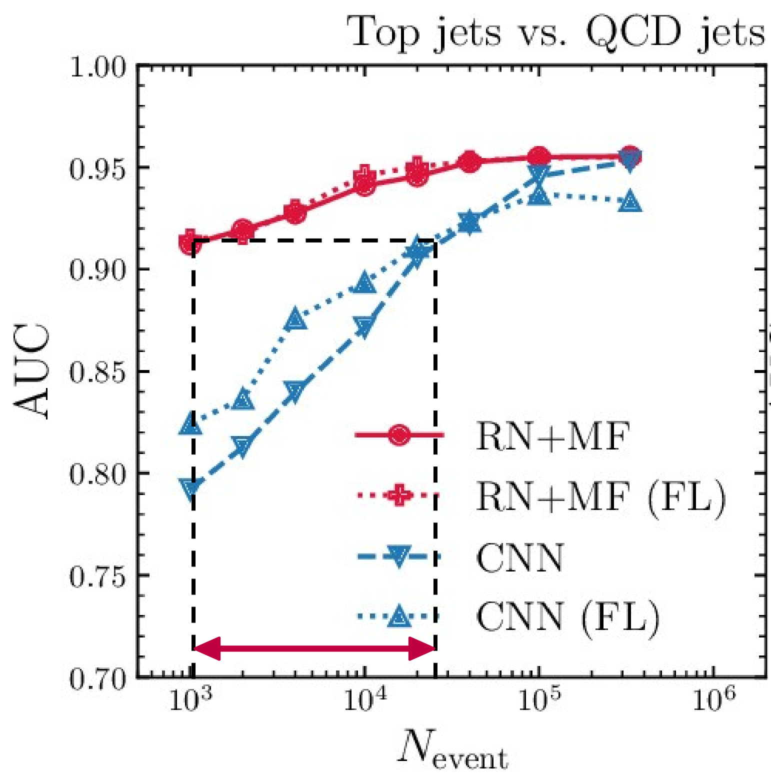


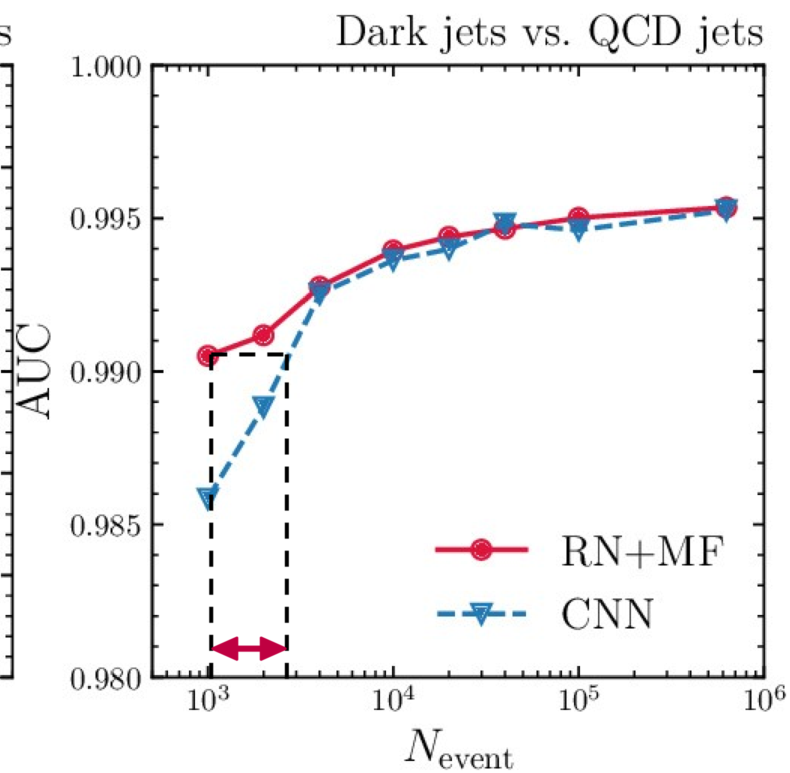


One interesting property of
this setup is that
all the calculation steps of the MFs
can be written in terms of
discrete convolutions.

Constrained Architectures and Low-Shot Learning

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





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

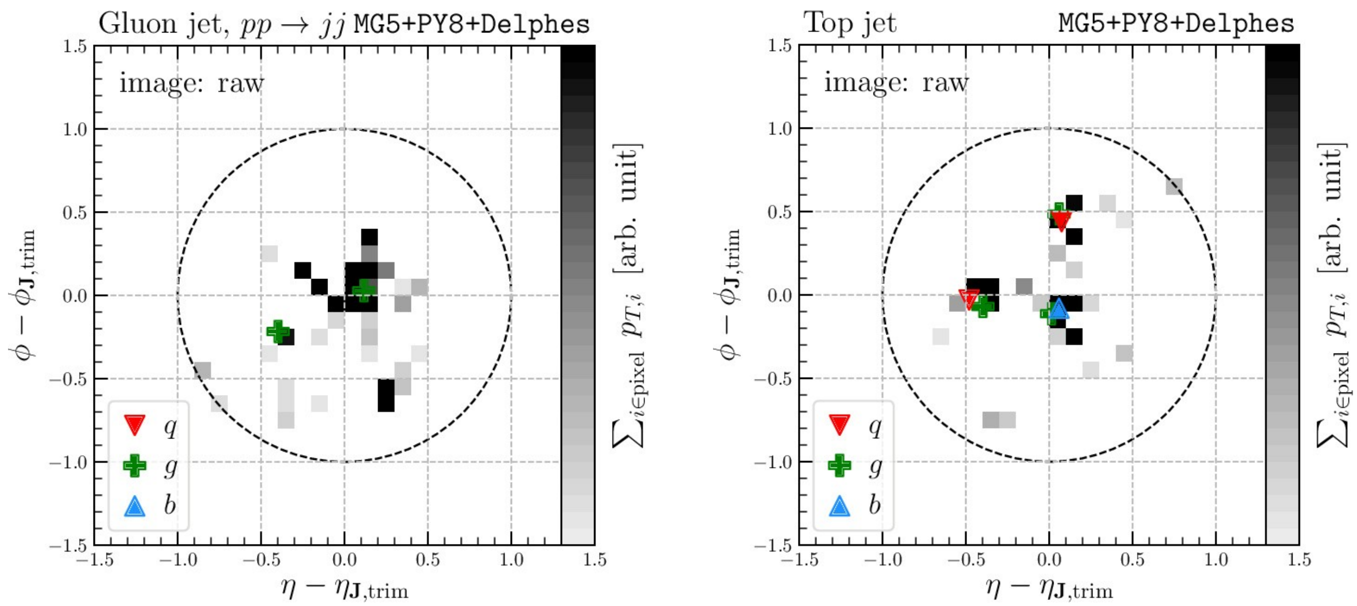
A smaller factor 3 gap is here, but this example has an exclusive phase space region parameterized by MFs.

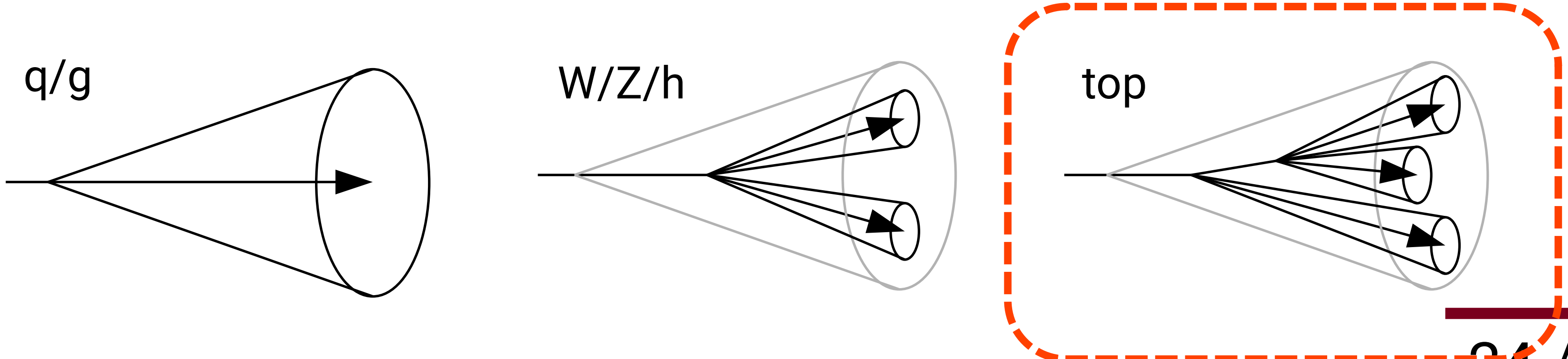
Sample description

- 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.

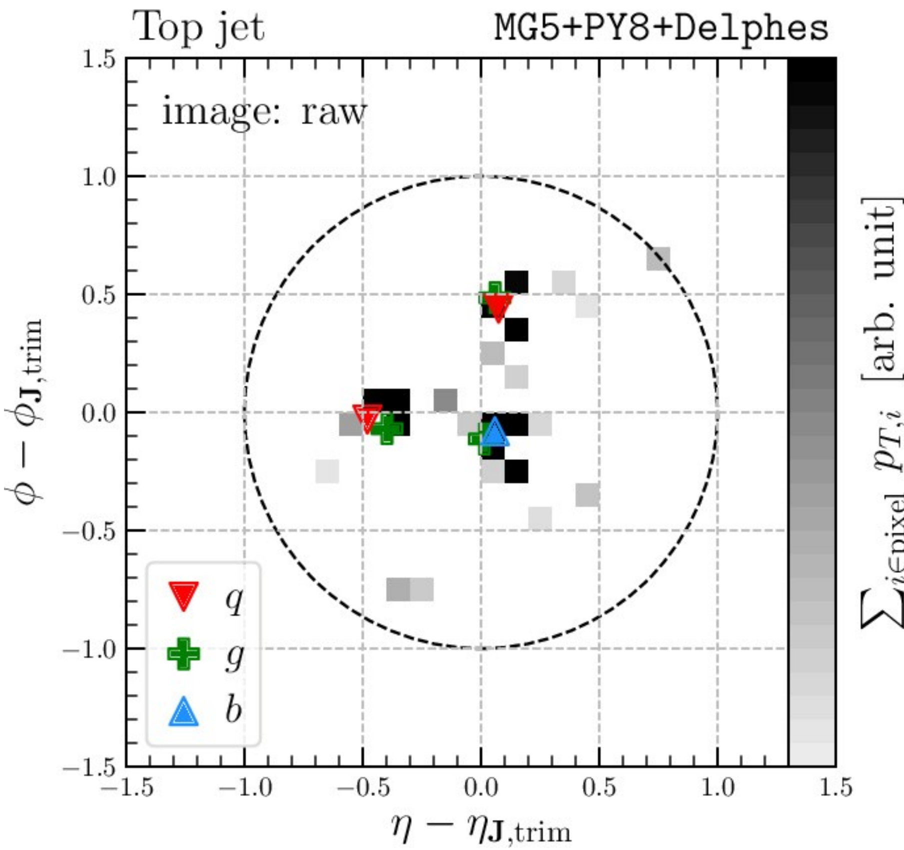
Jets have substructure!

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

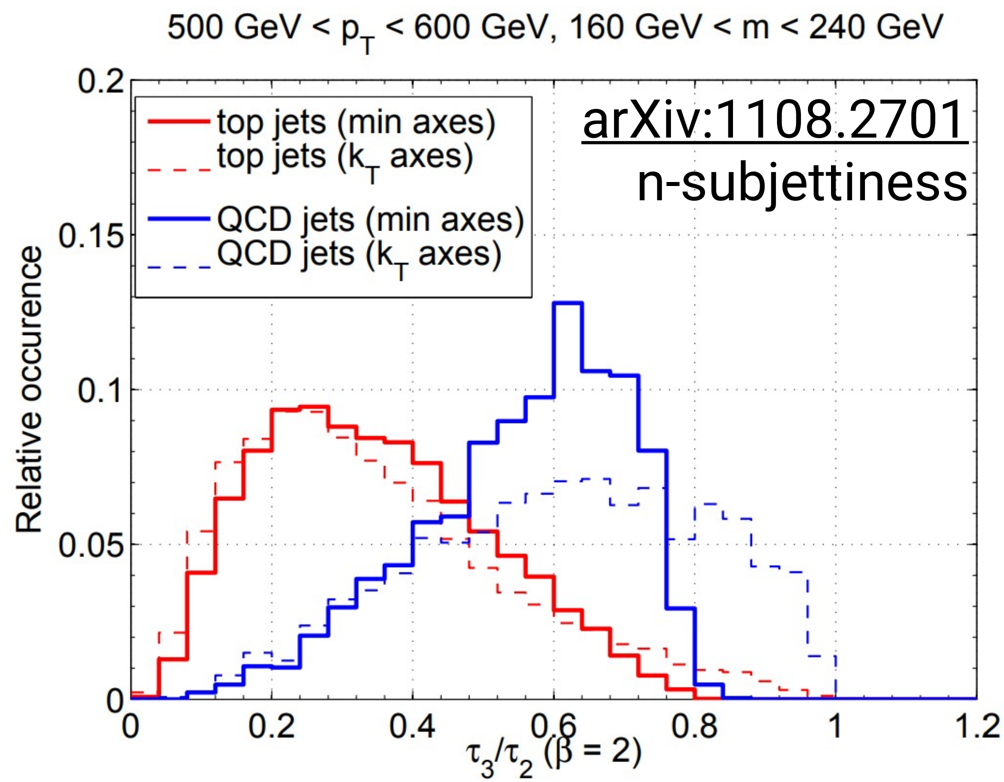




There are **two** approaches for building ML based jet taggers:



Directly analyze
Low-Level Feature (LLF)



Construct physics-motivated
High-Level Features (HLF)

Use CNN/GNN/Transfomers to analyze LLF
- ParticleNet
- ParticleTransformer
- LorentzNet, PELICAN (equivariant GNN/Transformer)

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

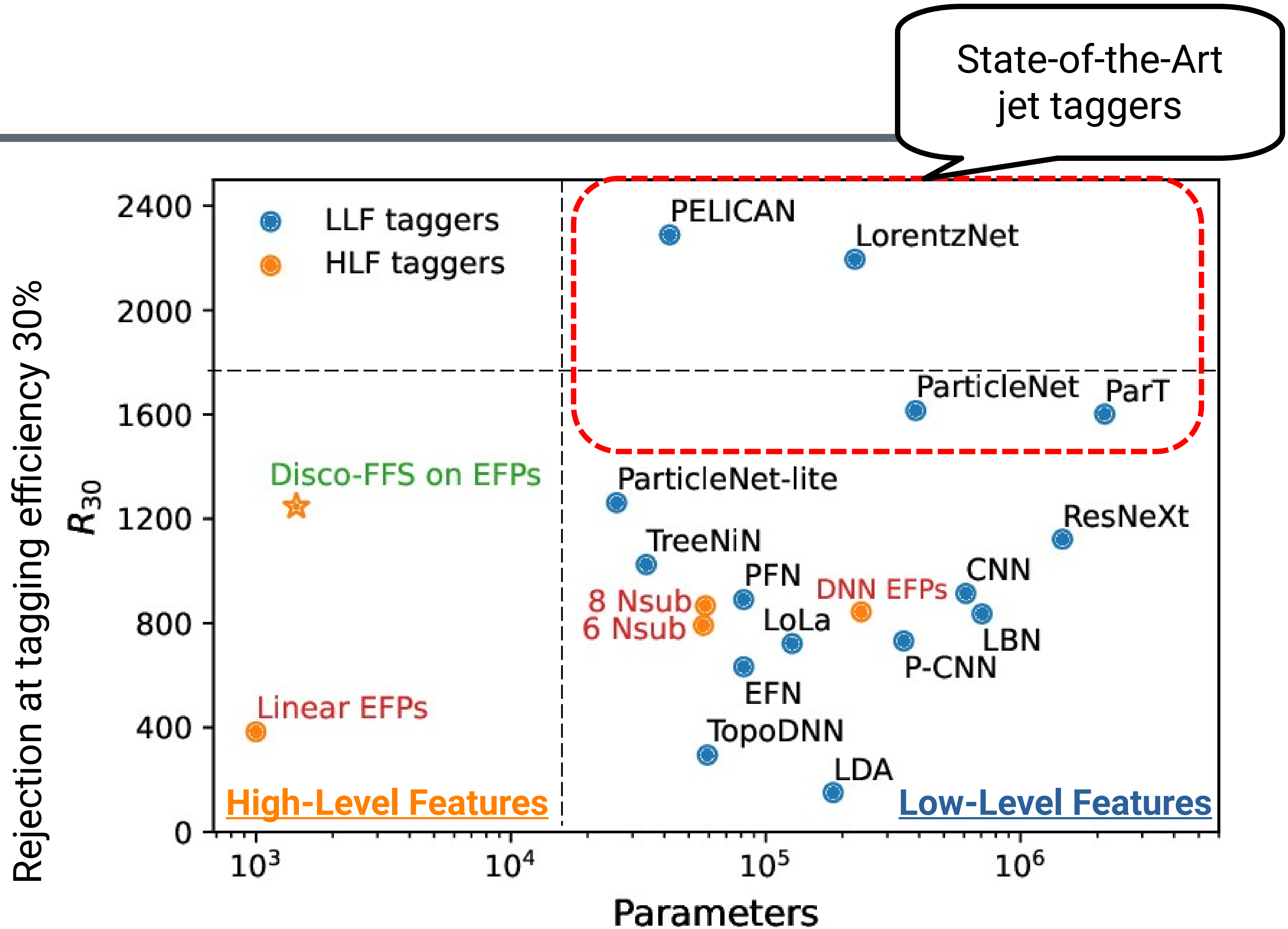
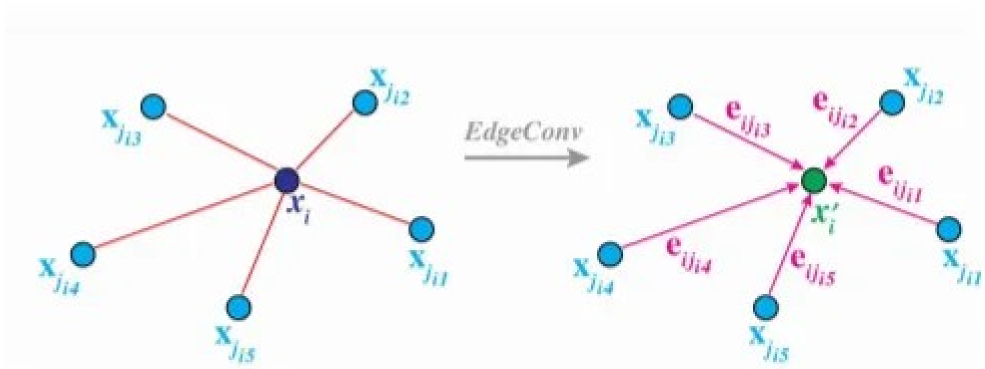
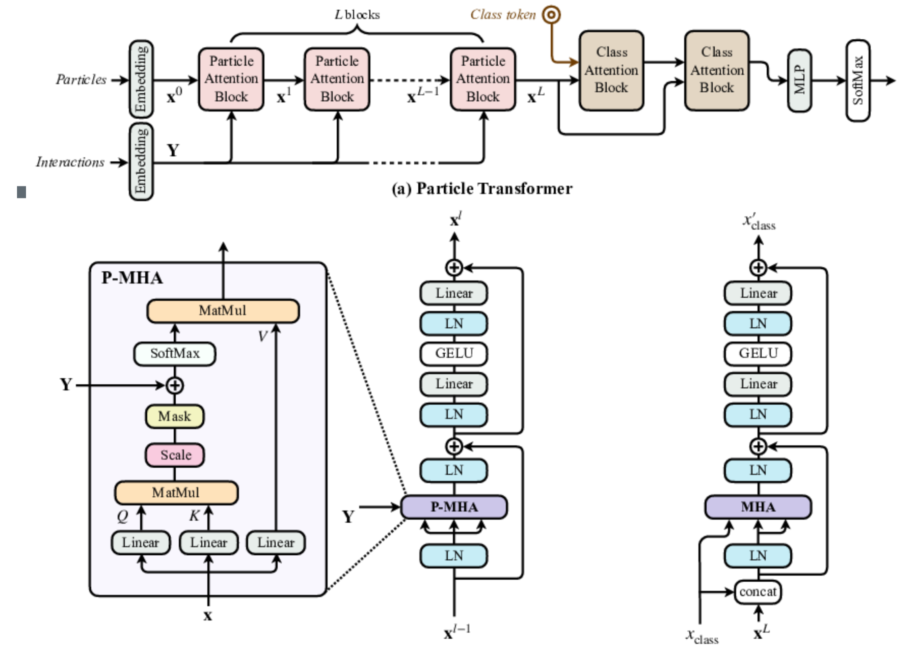


Figure from R. Das, G. Kasieczka, D. Shih, 2212.00046





(a) Particle Transformer

(b) Particle Attention Block

(c) Class Attention Block

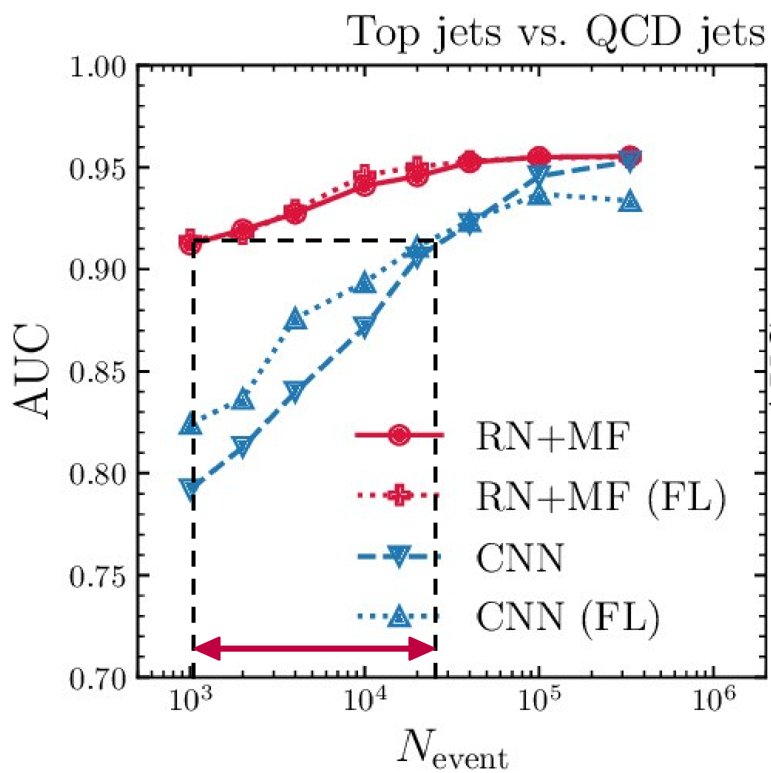
GNN / Transformers are working great. But because they are general purpose low-level feature analysis tools, it is hard to understand outcome other than the fact that they estimated the classifier output (likelihood ratio) more precisely.

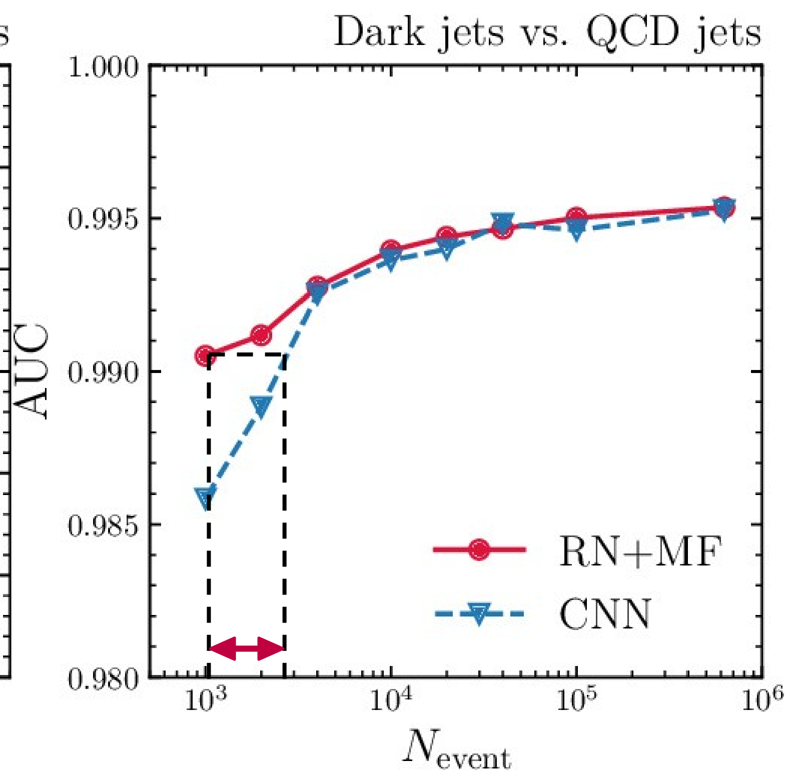
Can we build up a high-level feature based classifier equally performing well? YES!

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

Constrained Architectures and Low-Shot Learning

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

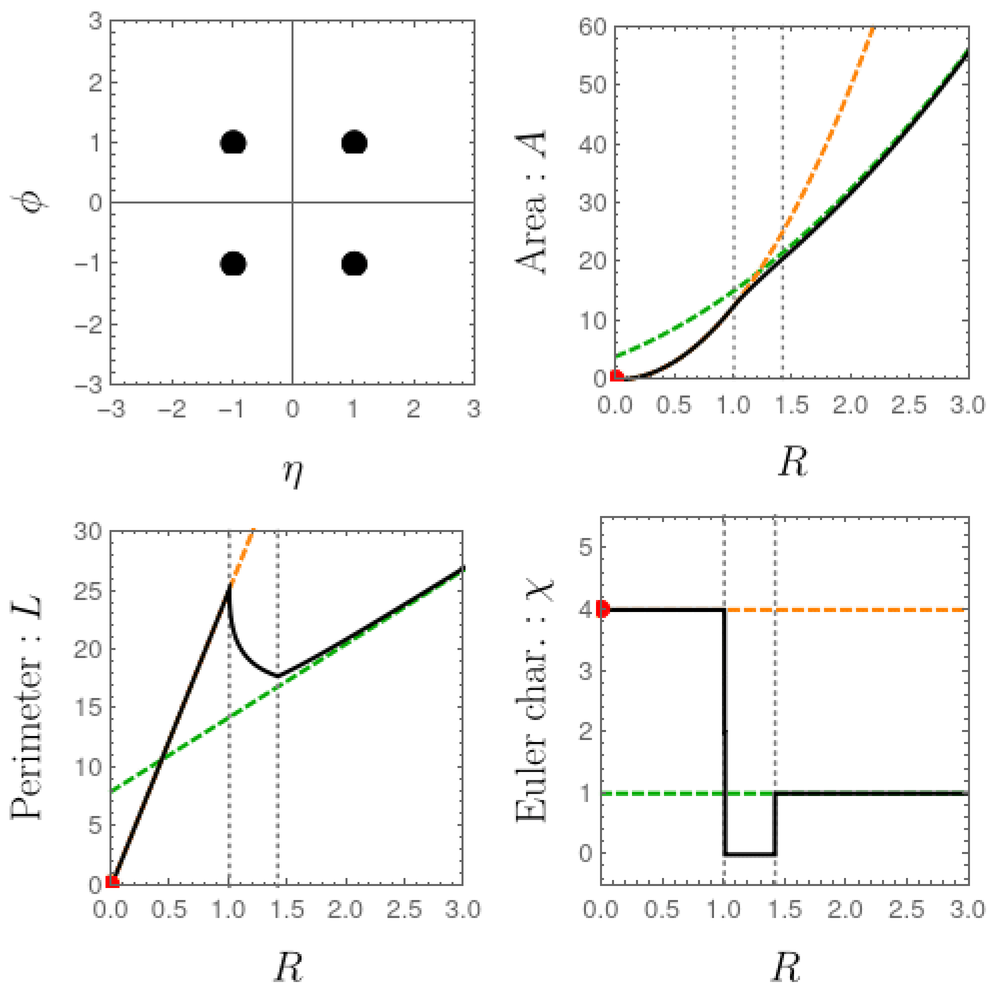




RN+MF is much less sample-demanding thanks to its constraints.

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

Mathematical Morphology and Minkowski Functionals



Orange: asymtote as r → 0
Green: asymtote as r → infinity

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.