Morphology for Jet Classification

Understanding and Improving jet classification
using "Minkowski Functional"

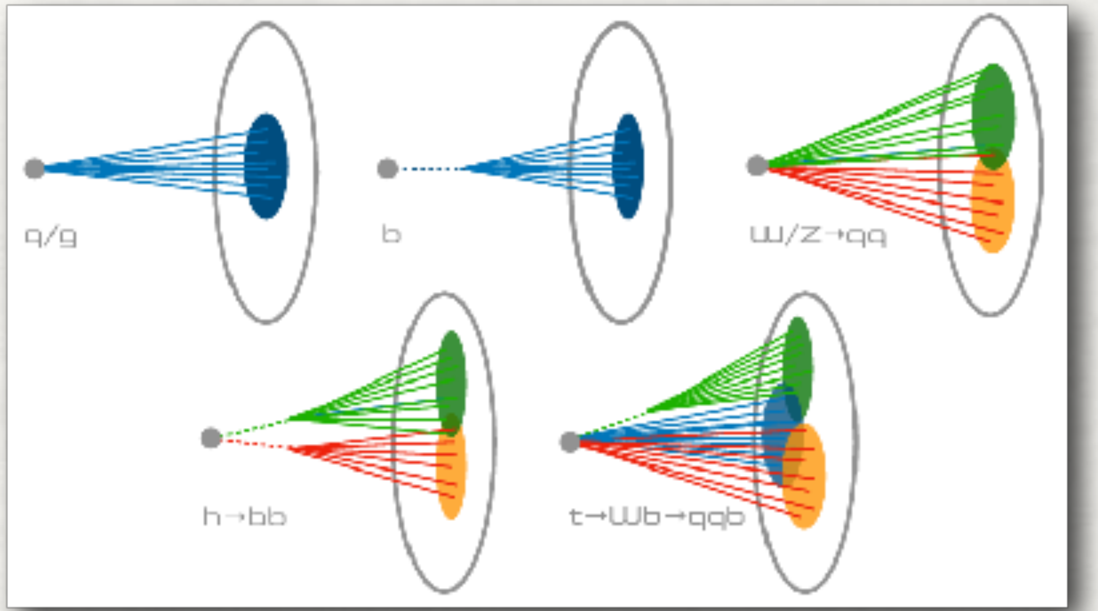
Mihoko Nojiri & Sung Hak Lim

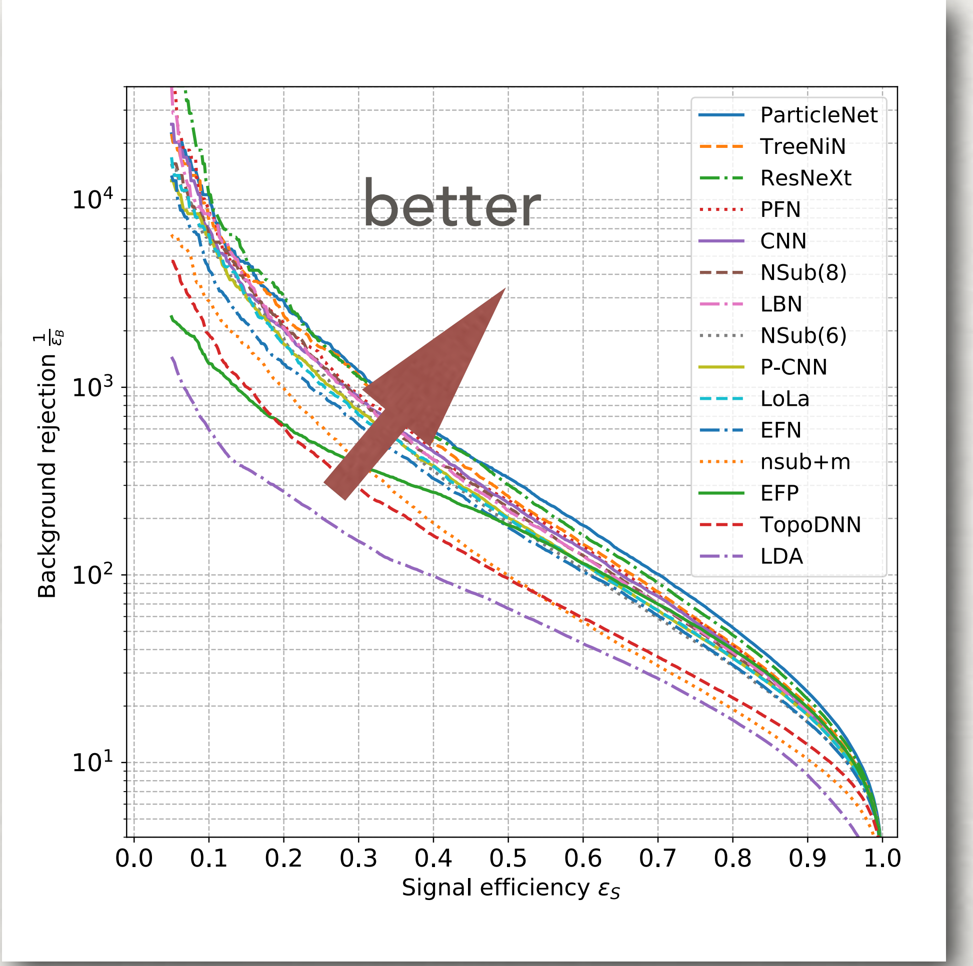
2010.13469, 2003.11787(JHEP), 1904.02092(JHEP)

JET CLASSIFICATION AND DEEP LEARNING

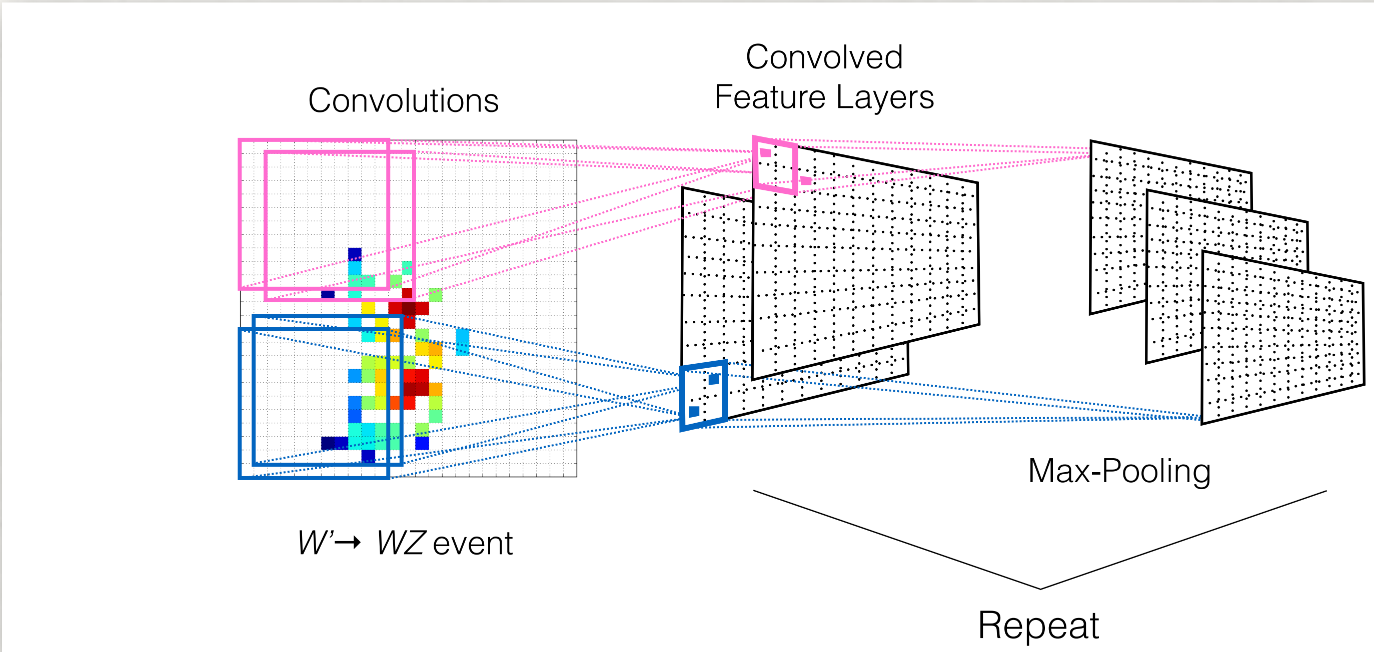
JET CLASSIFICATION

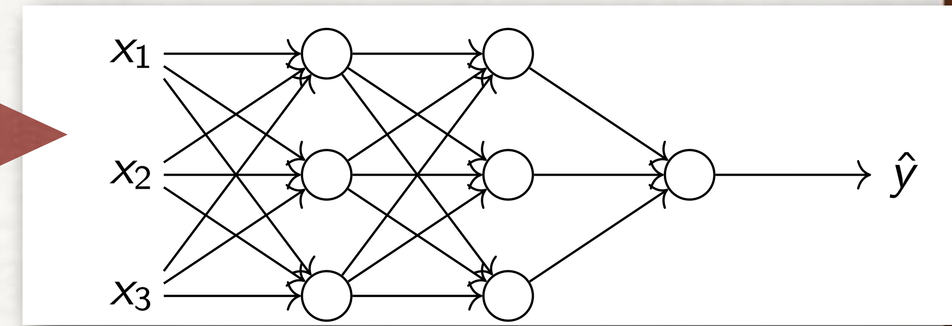
- LHC →HL-LHC → FCC-hh (far future)
- Better experimental sensitivity using ML → Deep Learning (Many architecture in market.)
  - QCD jet vs top, Higgs, W, new physics
  - Anomaly
  - Moving High level inputs to low level inputs becomes possible, and it gives better performance.
  - For jet physics, Jet Image→ CNN, Recurrent neural network(RNN), particle net,,,

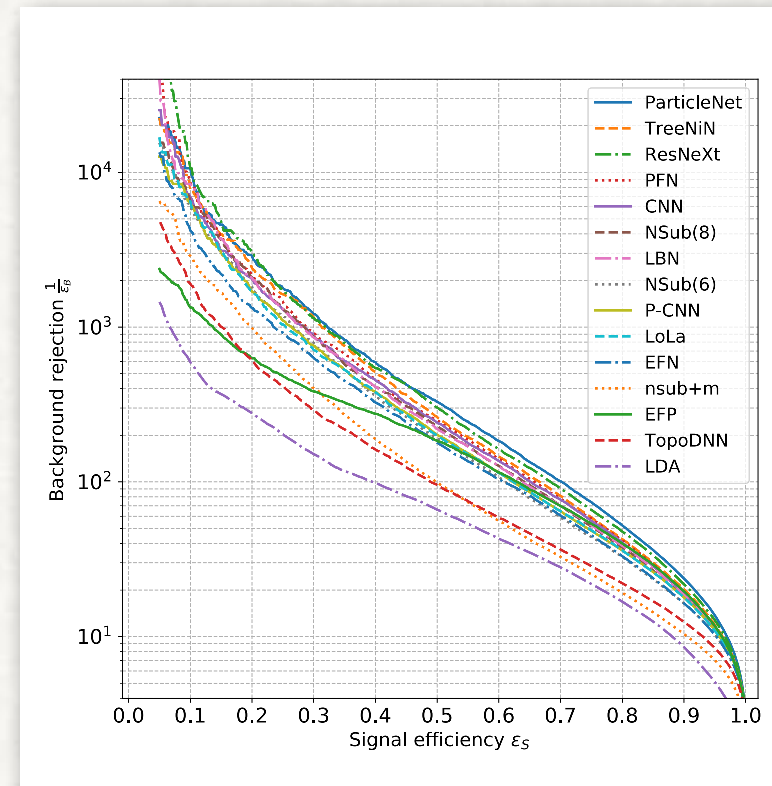




CNN(CONVOLUTIONAL NEURAL NETWORK) AND TOP TAGGING

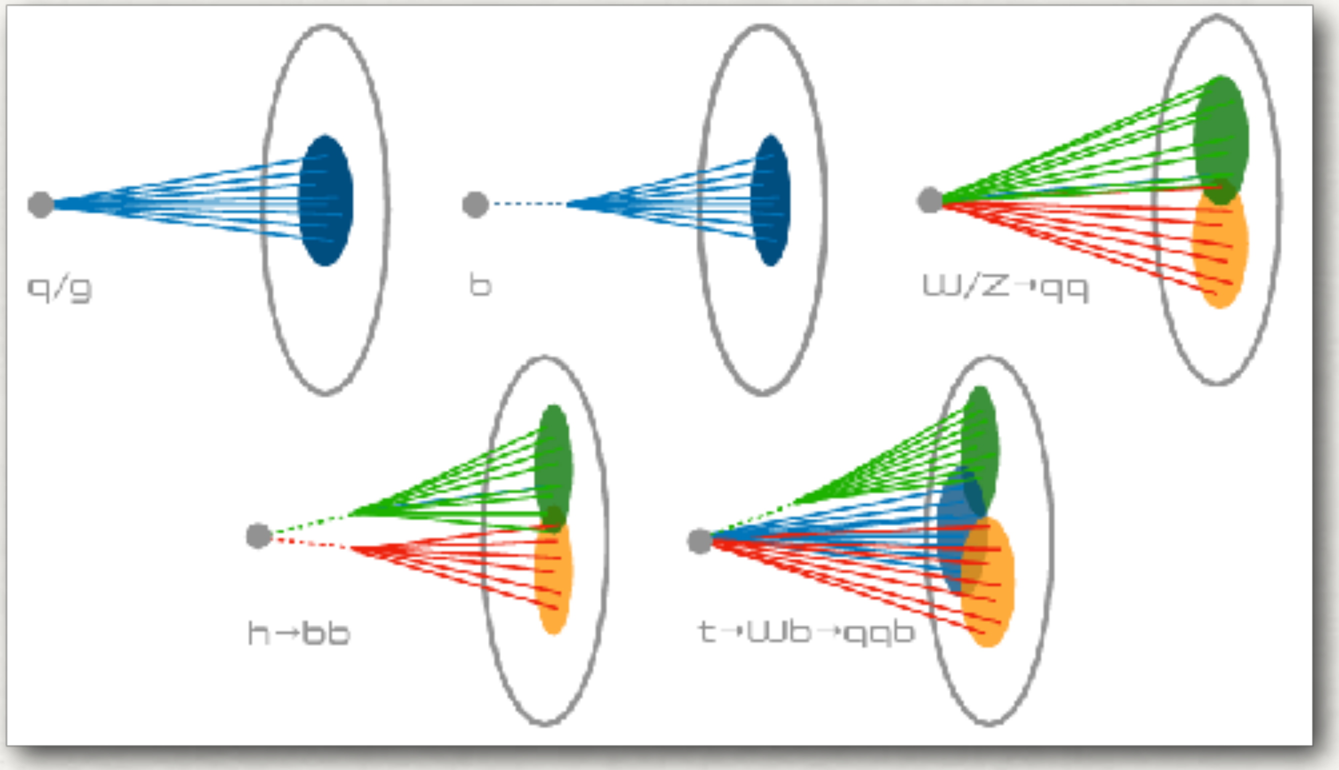


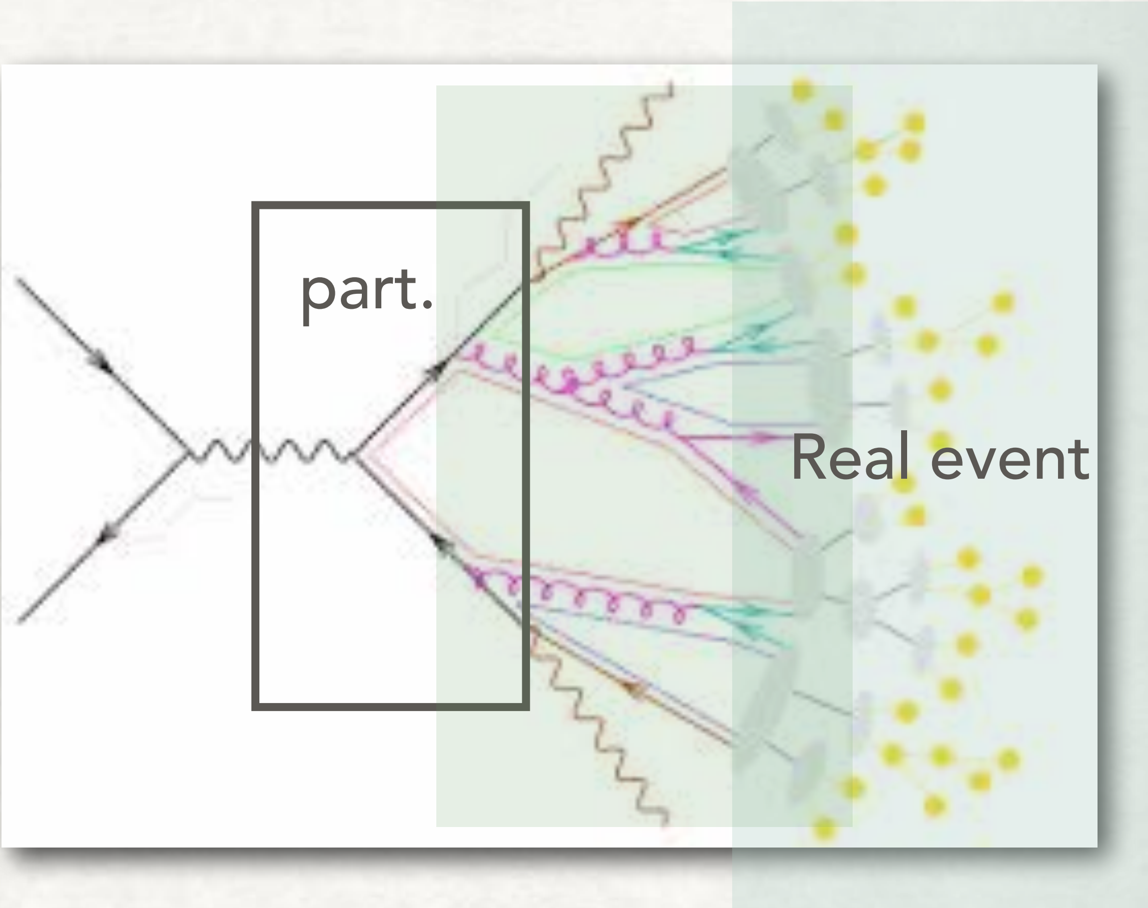




- Transfer image by NxN filter -> some cutoff (pooling) -> to find correlation.
- CNN, ResNeXT… use correlation of particles and show similar performance.
- Particle net, treat nearby particle correlation direction
- **Why one NN is better than the other? What kind of event is excluded additionally? What is the key feature?**

TWO QUANTITY FOR JET CLASSIFICATION.





- **IRC safe object:** :subjet, Energy correlation(C-correlator)
- **IRC sensitive Objects:** number of tracks, particles MC modeling Is bad (Pythia vs Hewig vs real data)
    - Color coherence etc.. Soft particle distribution → parent information
- Jet image contains both IRC safe and IRC unsafe object and jet classifier use it without prejudges

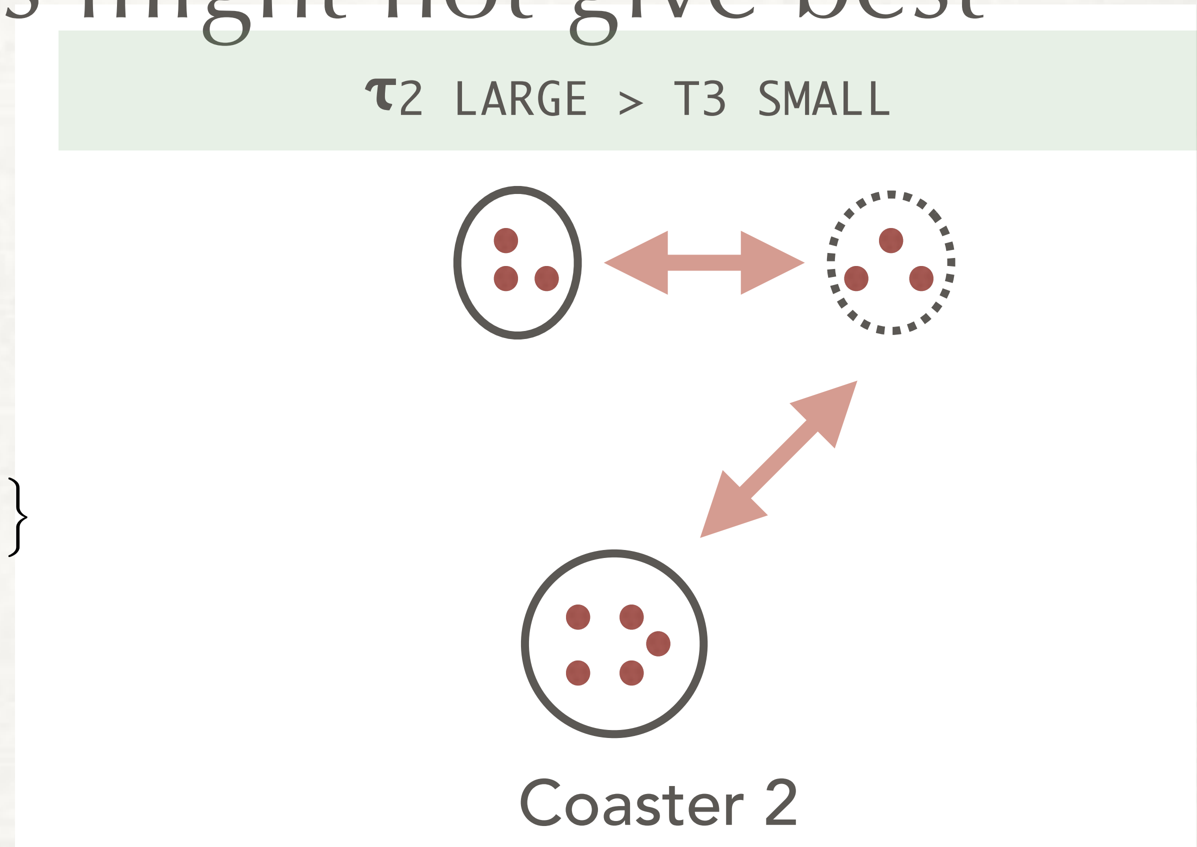
DL using IRC safe objects might not give best

N-subjettiness MLP classification

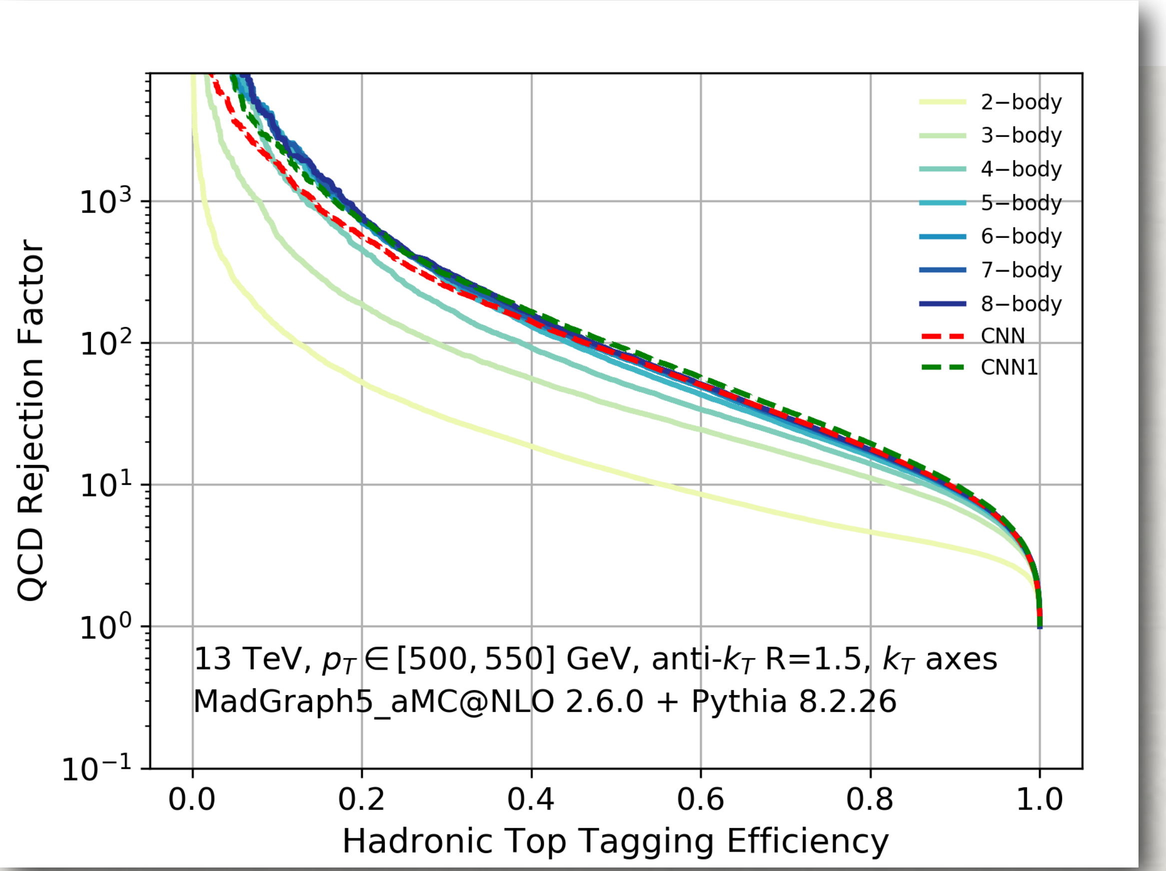
$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i\in \mathrm{Jet}} p_{Ti} \min\left\{R_{1i}^{\beta}, R_{2i}^{\beta}, \dots, R_{Ni}^{\beta}\right\}.$$

$$\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_5^{(1)}, \tau_5^{(2)}\right\}$$

arXiv 1704.08249 Datta Larkoski



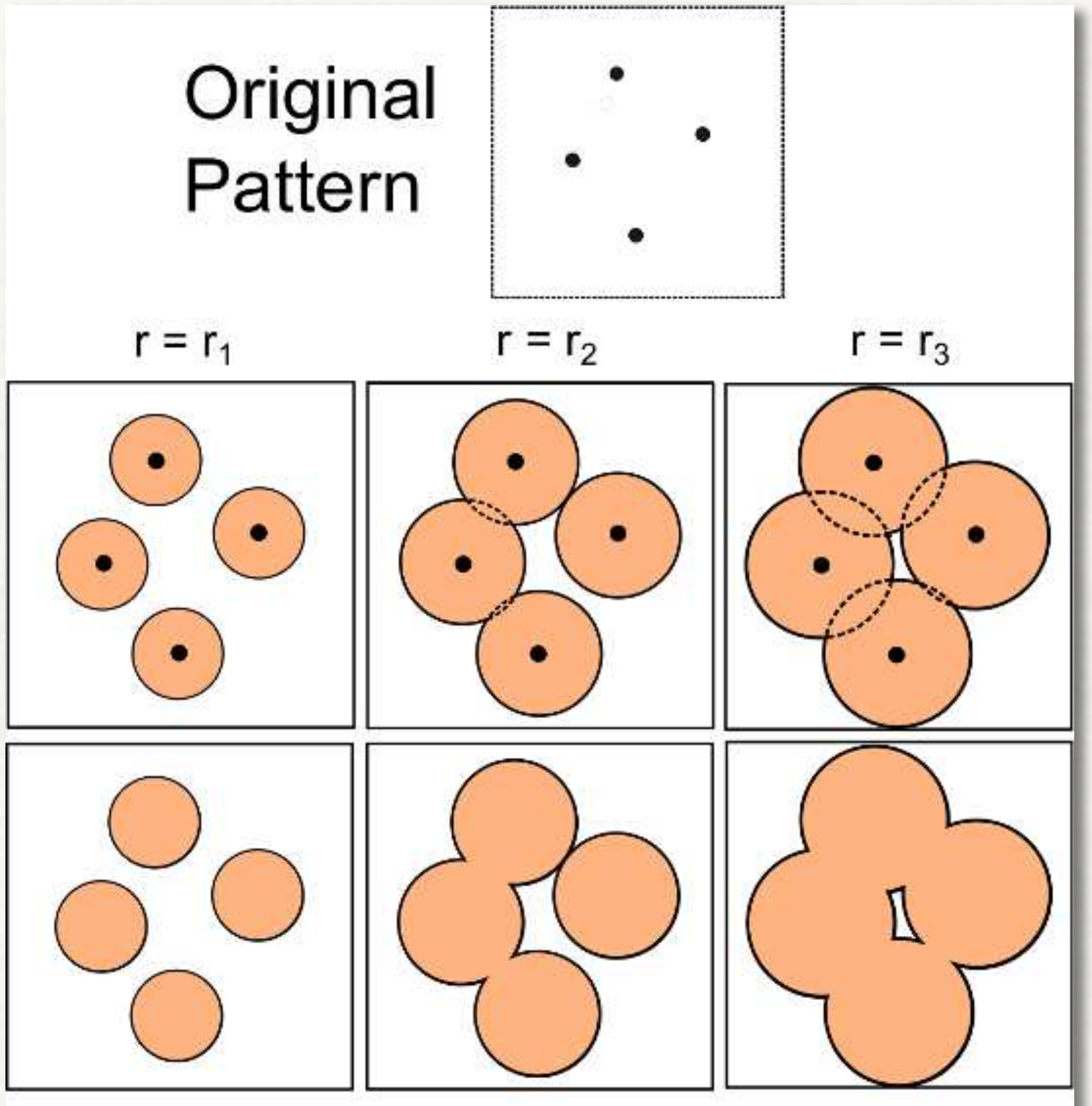
CNN vs N-subjettiness
Liam Moore et al 1807.04769



Need very higher order τ(5 body)

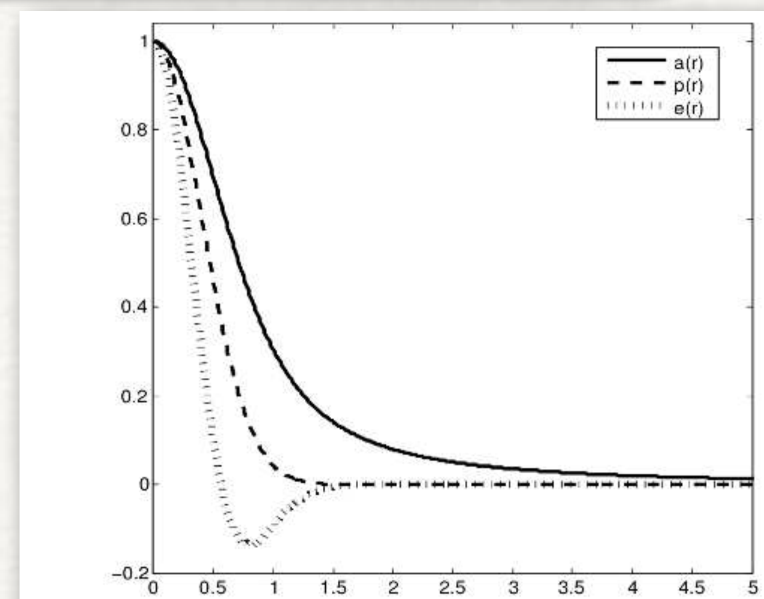
MINKOWSKI FUNCTIONAL
TO QUANTIFY SOFT PARTICLE DISTRIBUTION

“MINKOWSKI FUNCTIONAL ” MAP POINT DISTRIBUTION TO REAL FUNCTION WITH OUT LOSING INFORMATION



(a)

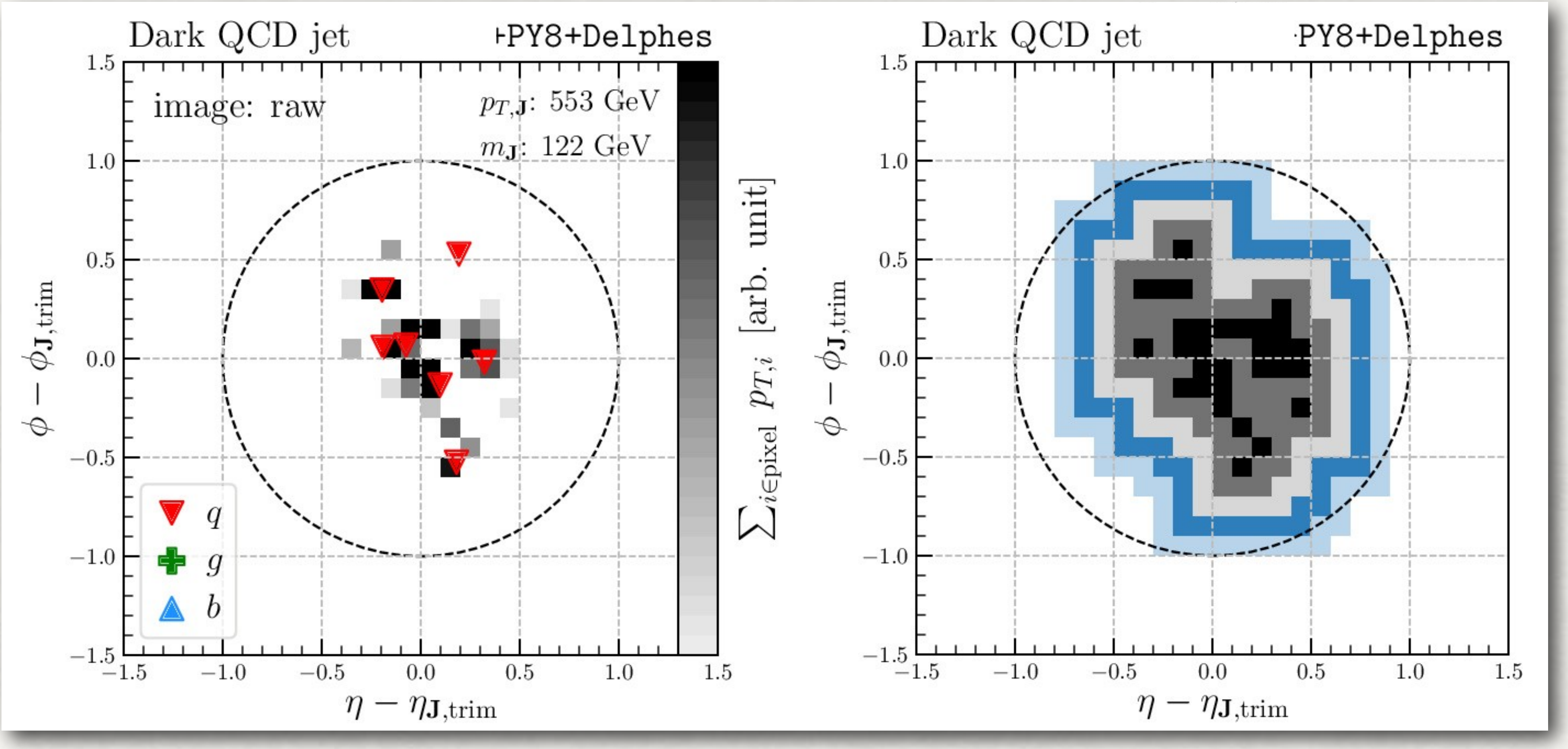
Area	$$a(r) = \frac{A(r)}{\pi N r^2}$$
Perimeter	$$p(r) = \frac{P(r)}{2\pi N r}$$
Euler characteristic	$$e(r) = \frac{\chi(r)}{N}$$

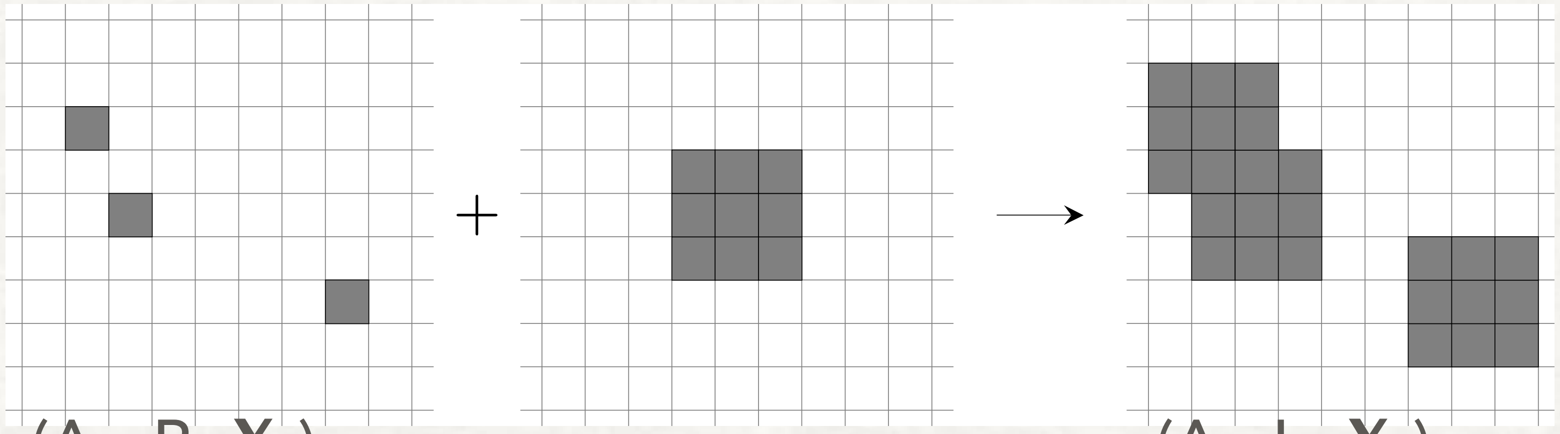


(b)

These three function represent point distribution in 2 dim up to rotation and translation
Hadwiger’s theorem. (Integral geometry)

JET IMAGE PROCESSING





→ repeat

$(A_0, P_0, \mathbf{X}_0)$ $(A_1, L_1, \mathbf{X}_1)$

We use A(n), P(n), X(n) of the jet image to quantify IRC unsafe activity

HOW TO QUANTIFY PARTICLE DISTRIBUTION

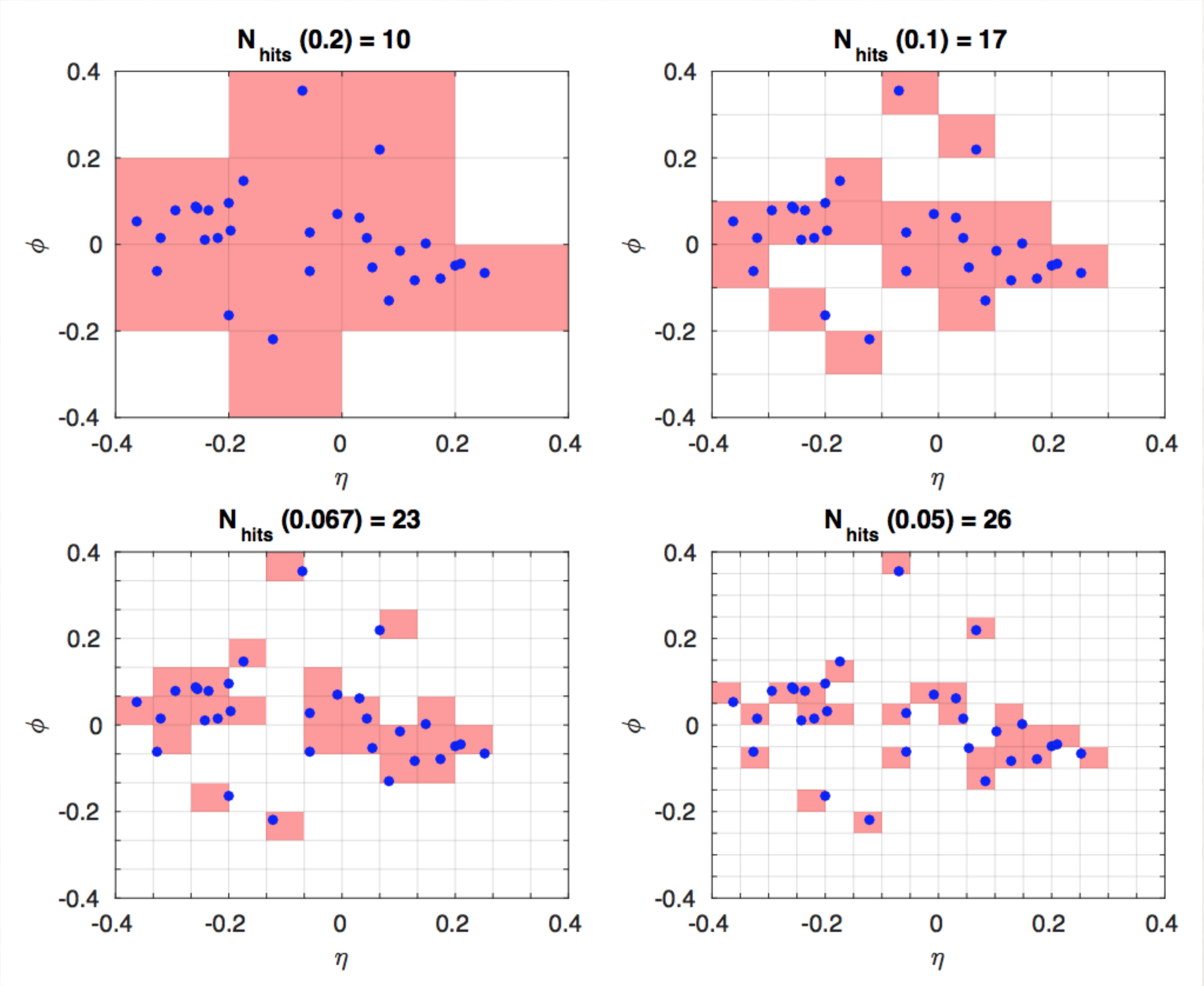
- **IRC safe** (quantity proportional to Energy) : subjects variables , energy flow polynomial,

- **IRC sensitive :** Jet Morphology particle distribution charge distribution,

    - Minkowski function (This work)

    - “Jet Toplogy” (Lingfeng Li, Tao Liu, Si-Jun Xu) arXiv 2006.12446]

    - “Fractal Based Observable” (Joe Davighi and Philip Harris) arXiv 1703.00914 Eur.Phys.J. C78 (2018) no.4, 334



WHAT IS THE GAIN

- Small fluctuation of the input → reduce the
  - A_0, A_1, A_2… A_0 =N(particle) number of particle in the jet ~ O(1/√N) fluctuation
  - Number of particle in jet image (ηij, φij, 0 or 1) ~ O(1) fluctuation.
- (A_0,P_0,χ_0), (A_1,P_1,χ_1) .. recovers particle distribution in the jet up to “rotation” and “translation”
- Easy to calculate :Realization of MF b v: lookup table from 2x2 partition to 3 vectors that sum up to MF

$$v(f_i) \rightarrow (A_i, P_i, \chi_i) \text{ 、}$$

$$(A, P, \chi) = \Sigma_i v(f_i)$$

Table 1. Look-up table for Minkowski functionals.

Conf.		A	P	χ	Conf.		A	P	χ
1		0	0	0	9		1/4	1	1/4
2		1/4	1	1/4	10		1/2	2	−1/2
3		1/4	1	1/4	11		1/2	1	0
4		1/2	1	0	12		3/4	1	−1/4
5		1/4	1	1/4	13		1/2	1	0
6		1/2	1	0	14		3/4	1	−1/4
7		1/2	2	−1/2	15		3/4	1	−1/4
8		3/4	1	−1/4	16		1	0	0

APPLICATION OF MF

Statistical Physics

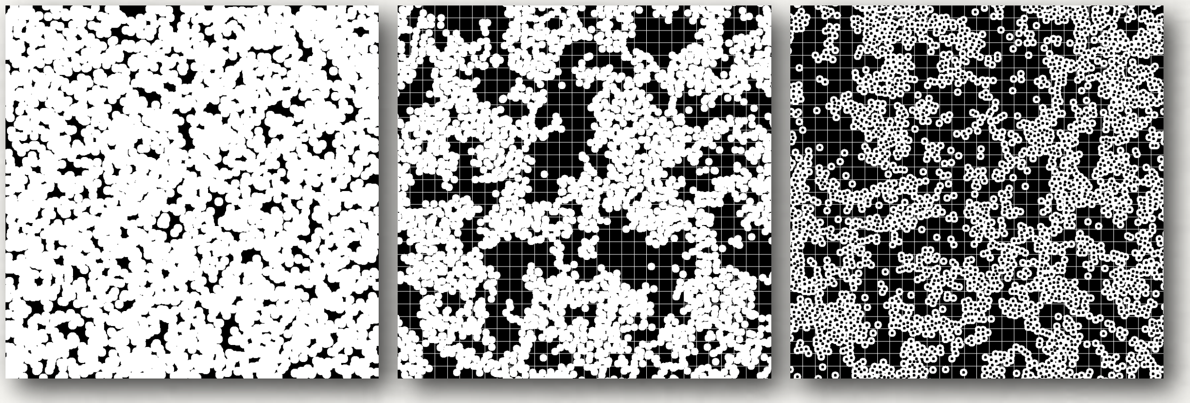
Left Porous medium

Middle: microemulsion

Right : Colloid

Occupation V, Surface(S) → material physics

Mecke and Stoyan (2000)



- Astrophysics : star and galaxy distribution, simulation study, non-Gaussinaity of CMB, weak lensing..

Powerful to quantitatively describe point distribution

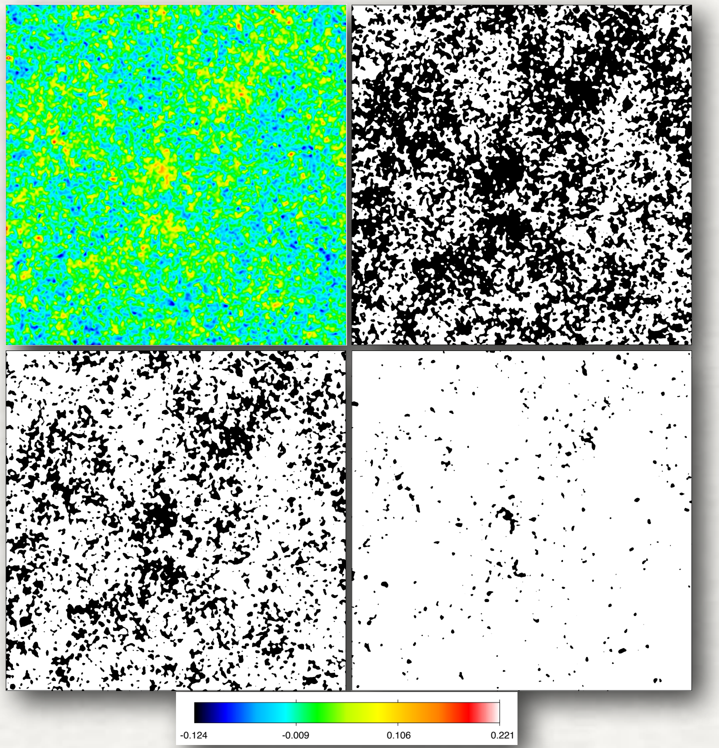


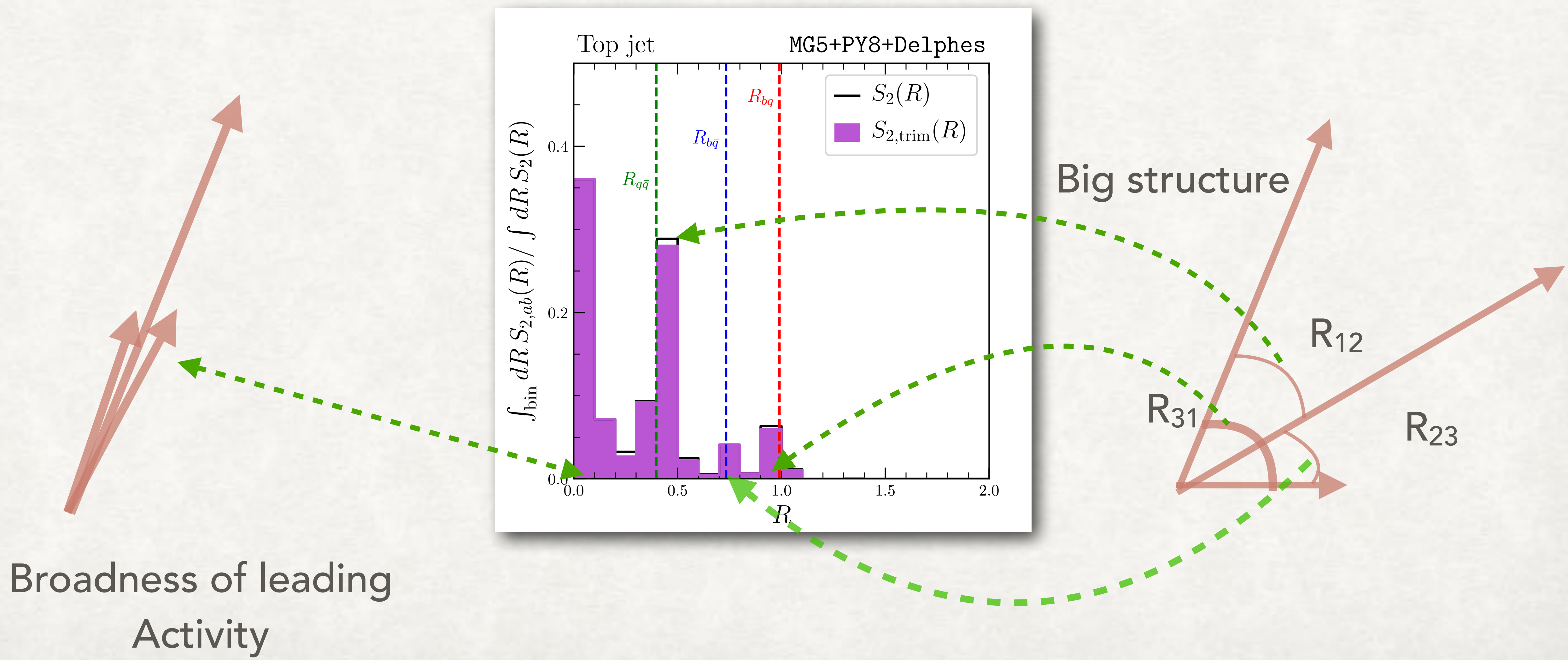
FIG. 1: *Top left panel: example of a simulated 12-square-degree convergence map in the fiducial cosmology, with intrinsic ellipticity noise from source galaxies and $\theta_G = 1$ arcmin Gaussian smoothing. A source galaxy density of $n_{gal} = 15/\text{arcmin}^2$ at redshift $z_s = 2$ was assumed. Other three panels: the excursion sets above three different convergence thresholds κ, i.e. all pixels with values above (below) the threshold are black (white). The threshold values are $\kappa = 0.0$ (top right), $\kappa = 0.02$ (bottom left), and $\kappa = 0.07$ (bottom right). The Minkowski Functionals V_0, V_1, and V_2 measure the area, boundary length, and Euler characteristic (or genus), respectively, of the black regions as a function of threshold.*

Kratochvil 1109.6334 Proving Cosmology with Weak Lensing Minkowski Functional s

OUR STRATEGY (IRC SAFE PART)

- **Incorporate soft distribution** by Minkowski Function
- **Incorporate hard distribution** by aggregated two point correlation S_2 and $S_2(\theta)$ after removing leading subjet

$$S_{2,\mathbf{J}_a\mathbf{J}_b}(R) = S_{2,ab}(R) = \int d\vec{R}_1 d\vec{R}_2 \, P_{T,\mathbf{J}_a}(\vec{R}_1) P_{T,\mathbf{J}_b}(\vec{R}_2) \, \delta(R - R_{12}),$$



WHAT IS S_2

IRC safe quantity: Theoretical estimate

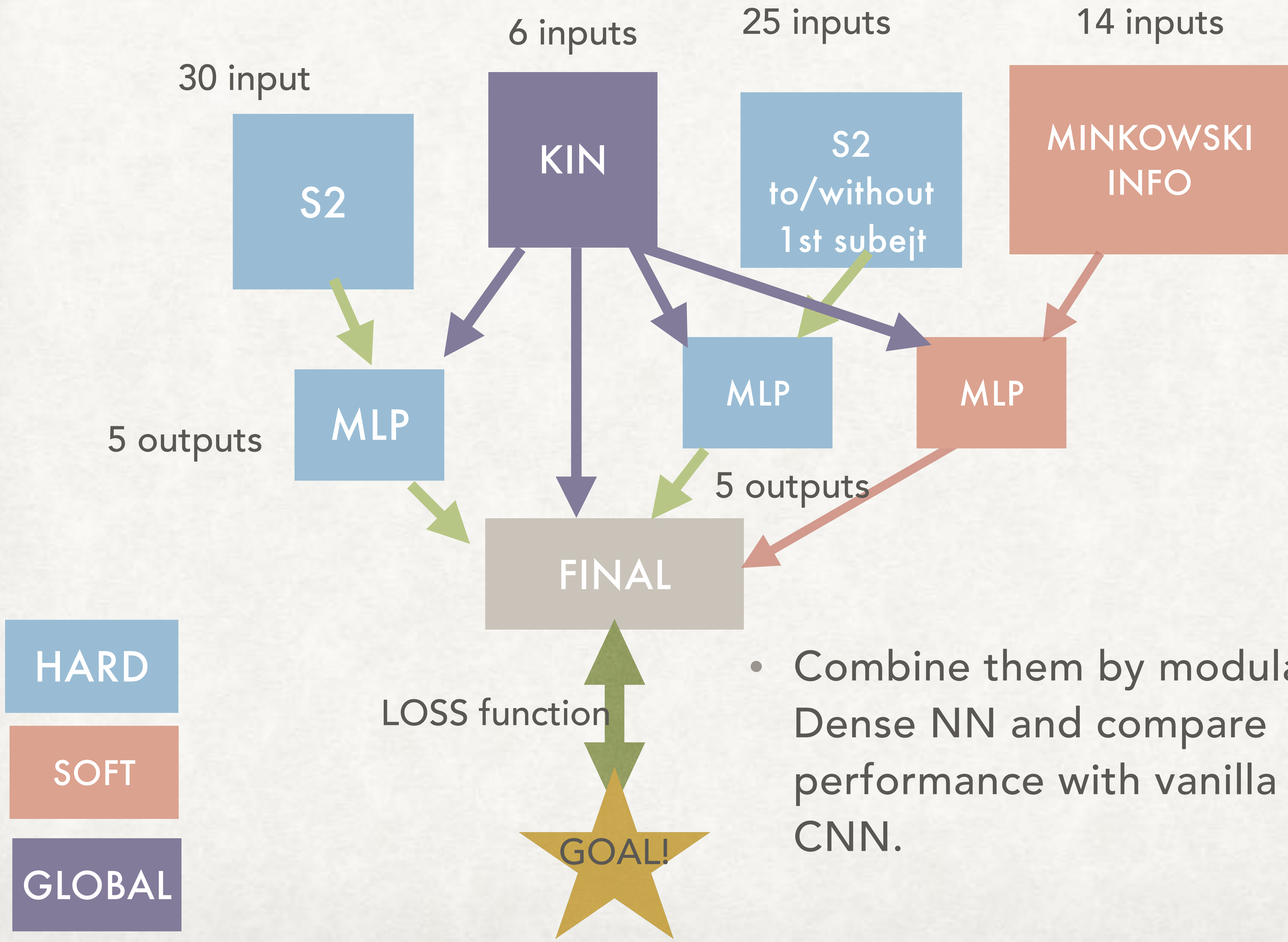
"Energy Flow Polynomials: A complete linear basis for jet substructure"
Komiske, Metoviev and Thaler JHEP 04 (2018)

$$\mathrm{EFP}^{(n)}_{2,ab} = \int d\vec{R}_1 d\vec{R}_2 \, P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_1) \, R^n_{12} = \sum_{i \in \mathbf{J}_a} \sum_{j \in \mathbf{J}_b} p_{T,i} p_{T,j} R^n_{ij}.$$

$S2 \leftrightarrows \mathrm{EFP}_2^{(0)}, \mathrm{EFP}_2^{(1)}, \mathrm{EFP}_2^{(2)}, \mathrm{EFP}_2^{(3)} \ldots$

$$\mathrm{EFP}^{(n-1)}_{2,ab} = \int_0^\infty dR \, R^{n-1} \cdot S_{2,ab}(R).$$

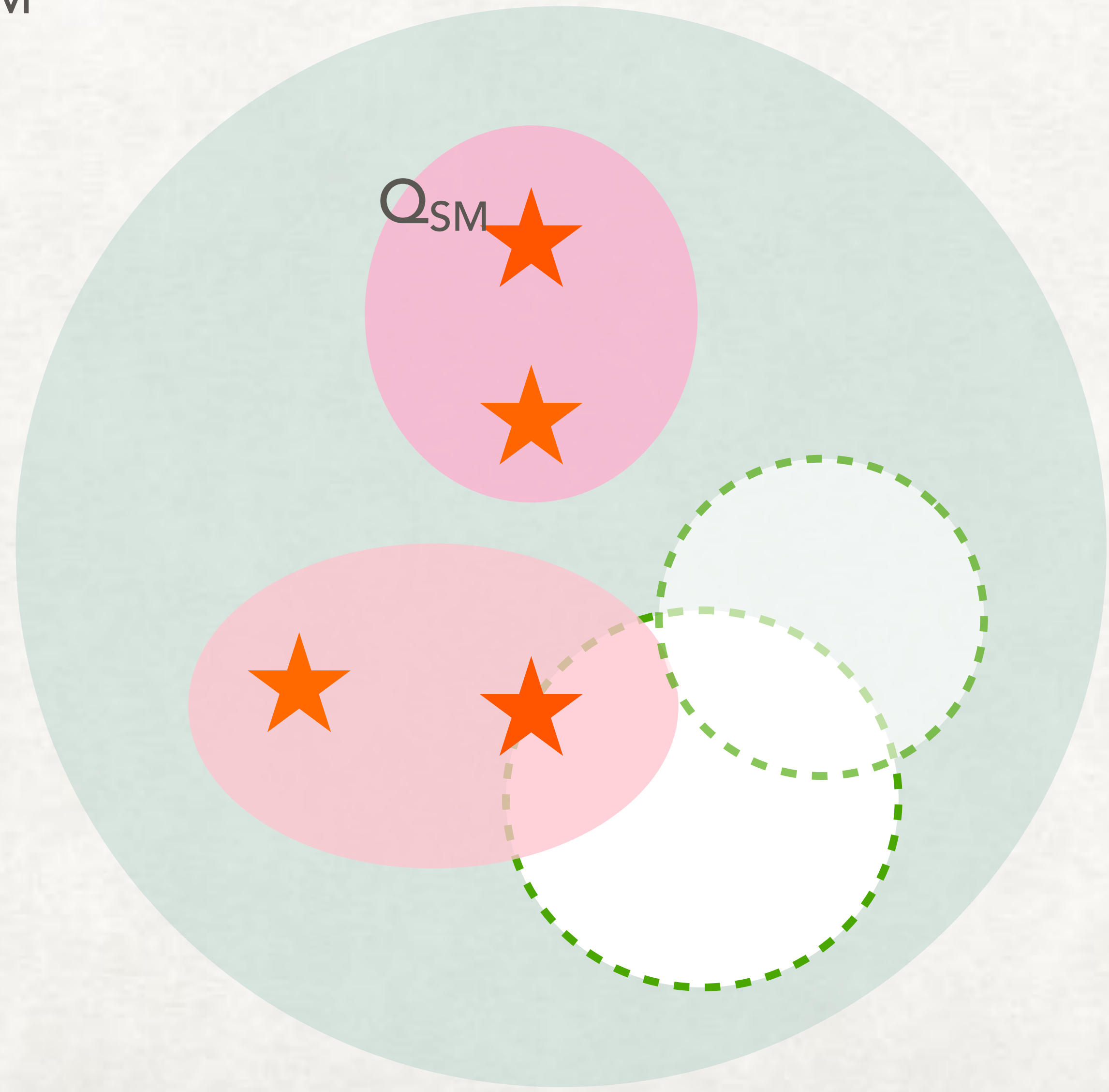
BACK UP : OUR RELATION NETWORK(RN)



- Combine them by modulate Dense NN and compare performance with vanilla CNN.

MODEL STUDY

- Dark jet $p\ p \rightarrow Z' \rightarrow Q_{Dark}\ \ Q_{Dark}$
- Dark parton shower → stable and unstable hadrons
- $\rho_{Diagonal} \rightarrow\ \ Q_{SM}\ Q_{SM}$

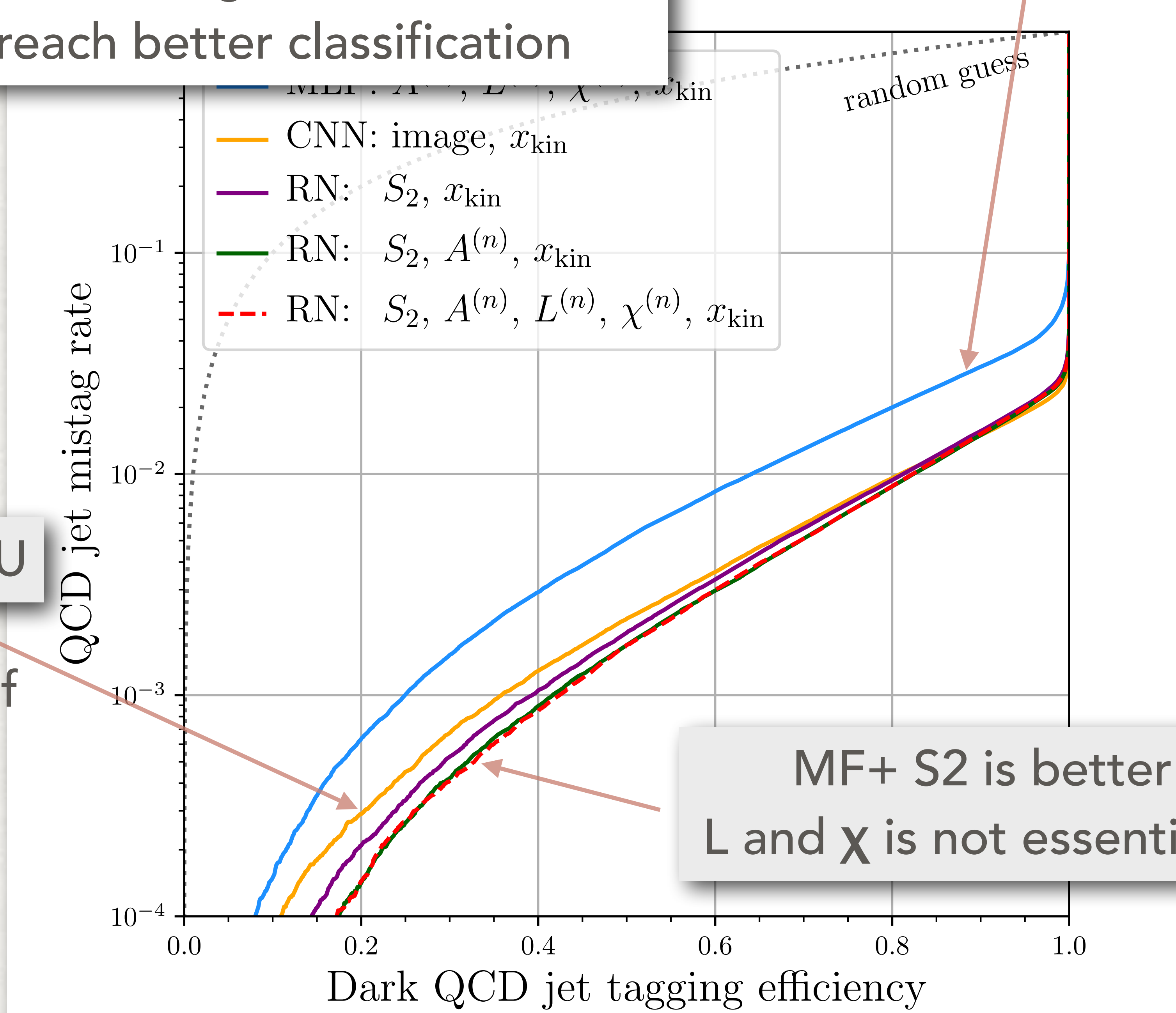


RESULTS FOR DARK JET VS QCD

Jet PT 150~300GeV Jet mass 30~ 70GeV

MF only reject lots of QCD jet

Even though CNN using more information
RN+ MF reach better classification



CNN using ReLU

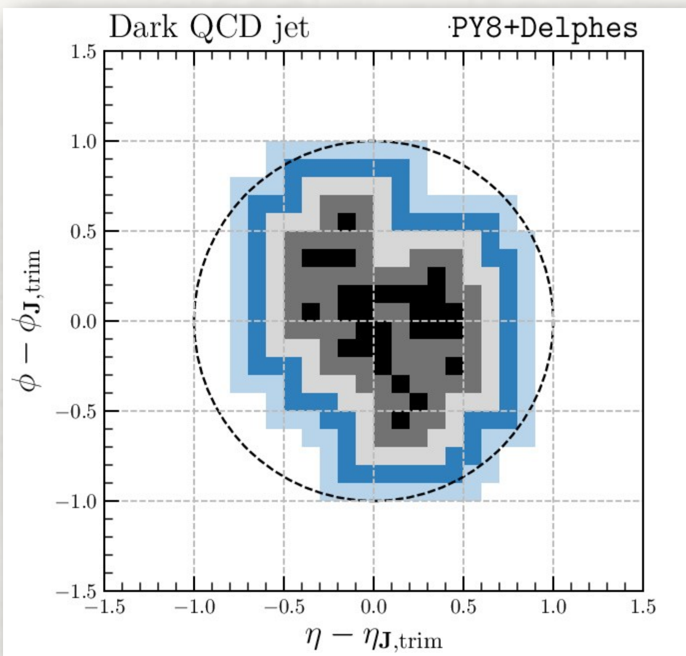
(Performance of
ELU is low)

MF+ S2 is better than CNN
L and χ is not essential for this case

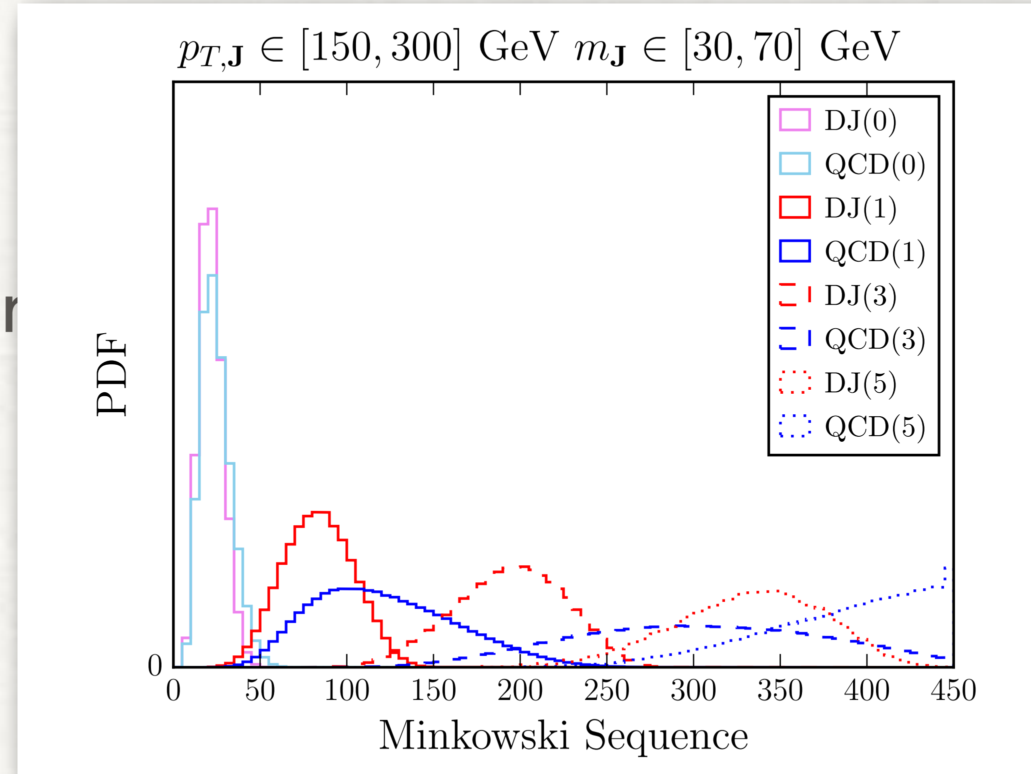
RN + MF also improve Top vs QCD (but improvement is smaller)

WHAT CNN TRAINING DOES?

Smaller MF endpoint suggests
Compact soft activities

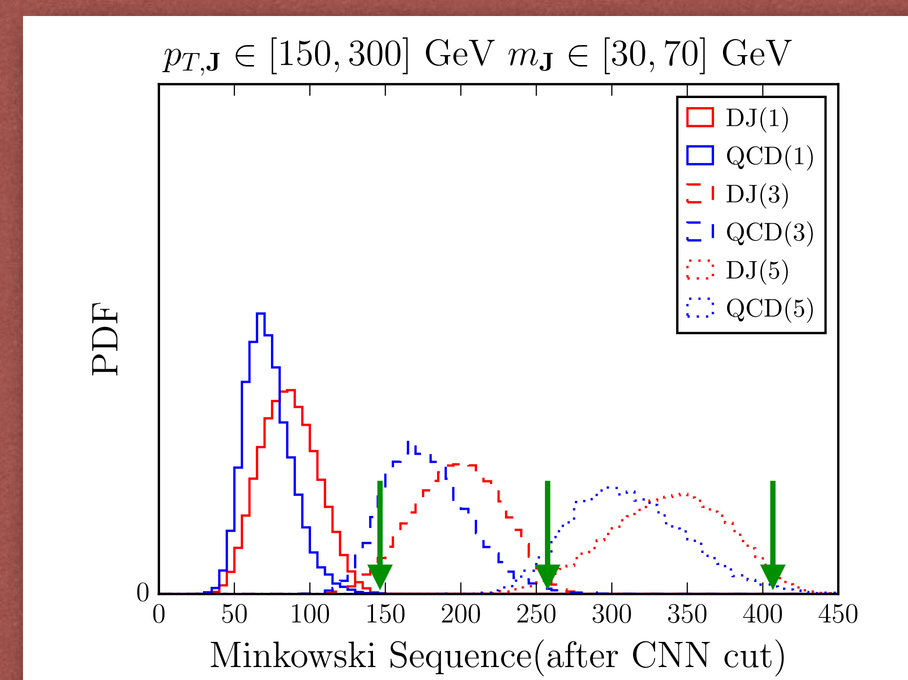


Original distribution



Cut using CNN

Providing MF is more efficient!
1/10 reduction of computation time and Memory usage.

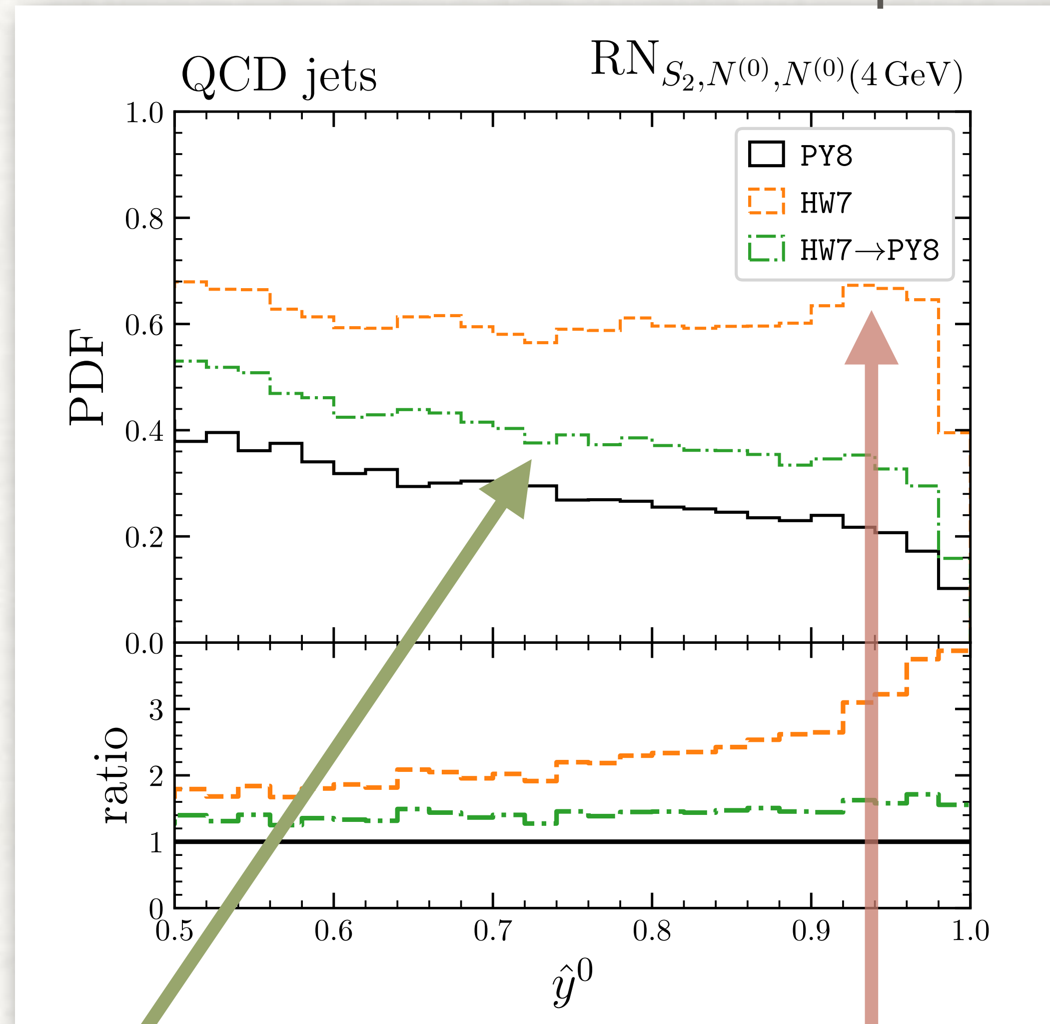


CNN allowing 10% rejection of signal

MERIT TO SEPARATE SOFT ACTIVITY (CALIBRATION)

- Herwig, Pythia, Sherpa: MC does not agree on soft particle distributions, especially for QCD jets.
- Improving the model using data takes time.
- Event reweighing via MF rather than jet image would be more feasible.
- Exercise of top event shows IRC safe distribution is not affected by reweighting

Top like



Pythia trained classifier classifying Herwig QCD jet as "top"

Reweighting to reproduce $A^{(0)}$

SHORT SUMMARY

- Minkowski functional is good tool to describe the n dim distributions of featureless points, and it works in Jet Physics too.
- Dark Jet vs QCD: CNN discover MF (without being told)
- MF + RN is better than CNN although(and because) it uses only part of the jet image. Reducing fluctuation by aggregation is the key.
- Future
    - Application to other HEP physics? (Such as displaced vertex, other detector such as water tank, cosmic shower) …
    - Connection to QCD?