RECENT PROGRESS ON JET SUBSTRUCUTRE FOR BSM

Mihoko Nojiri (KEK & KIPMU)
partly work in collabration with Sung Hak Lim(Rutgers U)
Amon Furuichi(Nagoya U. and KEK)

JET PHYSICS FOR BSM

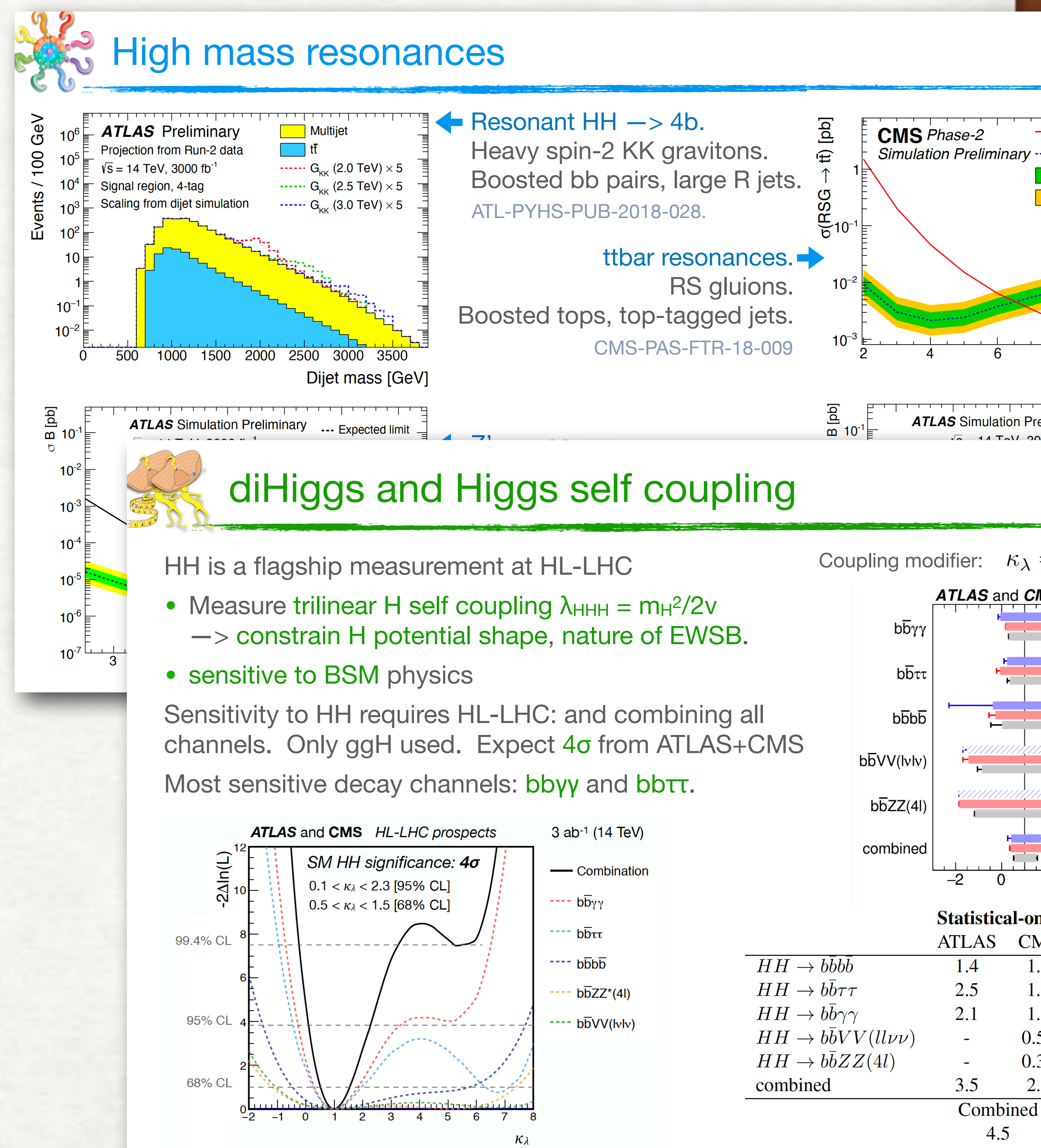
- boosted Higgs, boosted top for

Sezen Sekmen Aug 24 SUSY2021

  - heavy resonance search
  - SMEFT (high PT higgs boson, W, and Z distribution will be affected.)

- boosted objects look like a jet(narrow). "jet substructure" is important to distinguish it from QCD.

- systematic understanding of quark and gluon jets is important to estimate backgrounds.



JET SUBSTRUCTURE

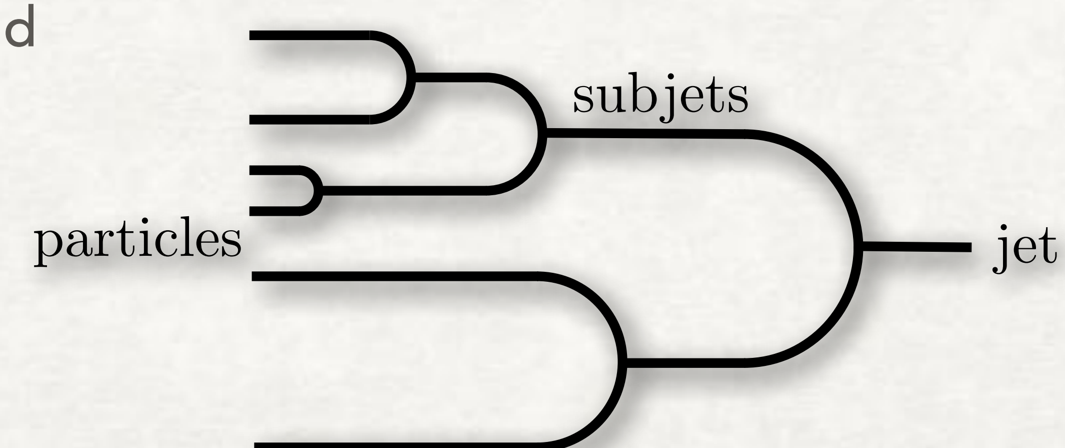
- Jet : Originally defined from highest PT object in the cluster (80's)
- KT algorithm (91) → general seedless jet algorithms (Cacciari Salam Soyez "fast jet" (2006))
  - merging the pairs with the smallest distance d_{ij} one by one nd "update" jet constituents

$$\begin{aligned} d_{ij} &= \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}\,, \\ d_{iB} &= k_{ti}^{2p}\,, \end{aligned}$$

angle between the pair p=1, 0, -1 for k_T, CA, anti-k_T

- Jet substructure for heavy object search (2008 Butterworth et al) going backward the cluster sequence. Systematically finding cores expected in heavy particle decays.
- mass drop→ soft drop

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$



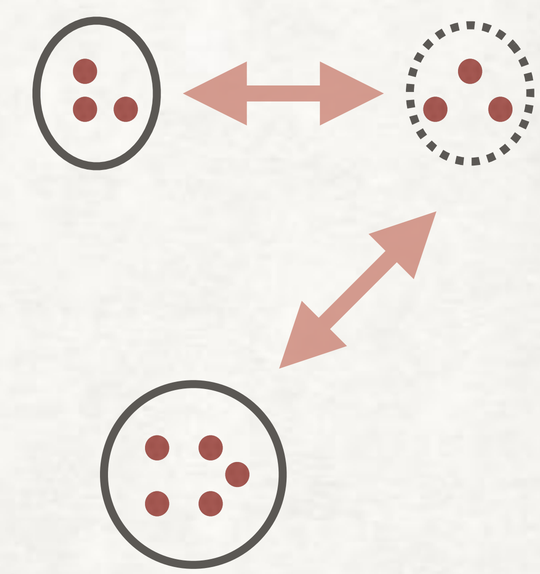
SEEDLESS IRC SAFE VARIABLES

- Cluster jet with large size R by anti-KT and see inside.

- n-subjettiness (2010 Thaler Tilburg) (minimize the distance to N axes.) +grooming

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

core1



Core2

- Energy Flow Polynomial(Komiske et al 1712.07124)

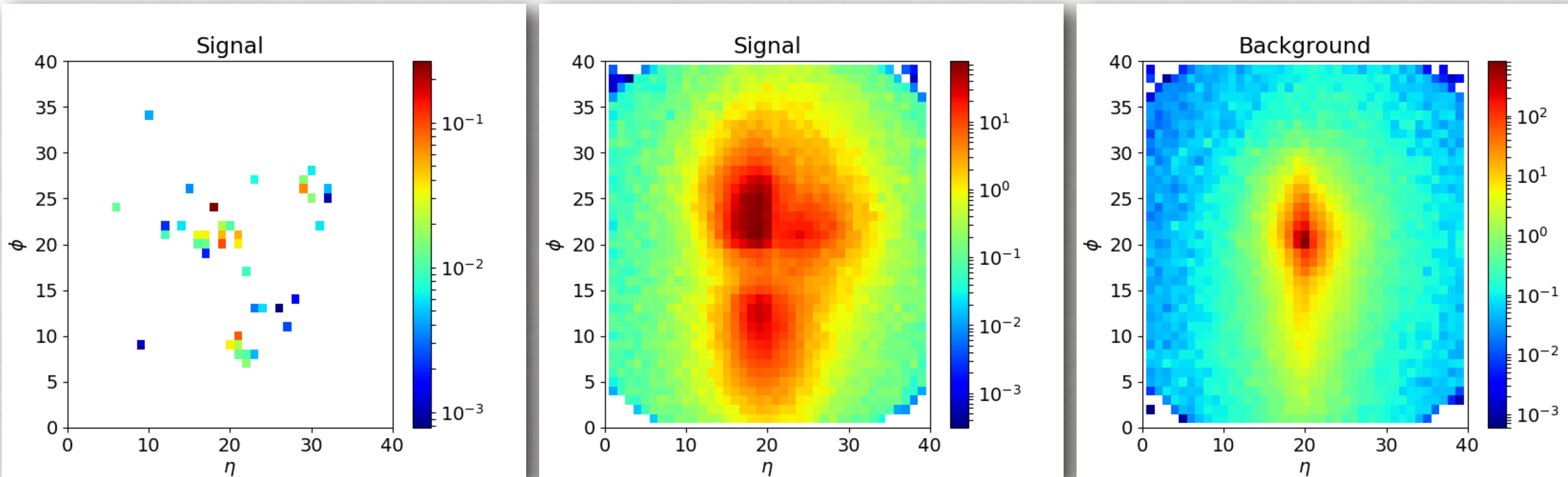
$$EFP_G = \sum_{i_1}^{M} \cdots \sum_{i_N}^{M} \ldots z_{i_1} \ldots z_{i_N} \prod_{k,l \in G} \theta_{i_k i_l}$$

- 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 E_i for the particle involved ← IRC safe

BILDING CLASSIFIER

single event　　　top 10,000　　　qcd 10,000



jet with pT=500~ 650GeV from 1902.09914

<u>work flow</u>

jet → (High level variable z_i) → classifier $\Phi(z_i)$ cut based, BDT, DL

or jet constituent itself

(classifier : increase the prob. of signal and reduce the prob. of background)

Deep Learning and classifier $\Phi(x_i)$

output: $\mathbf{w_{ij}}, \mathbf{b_i}$

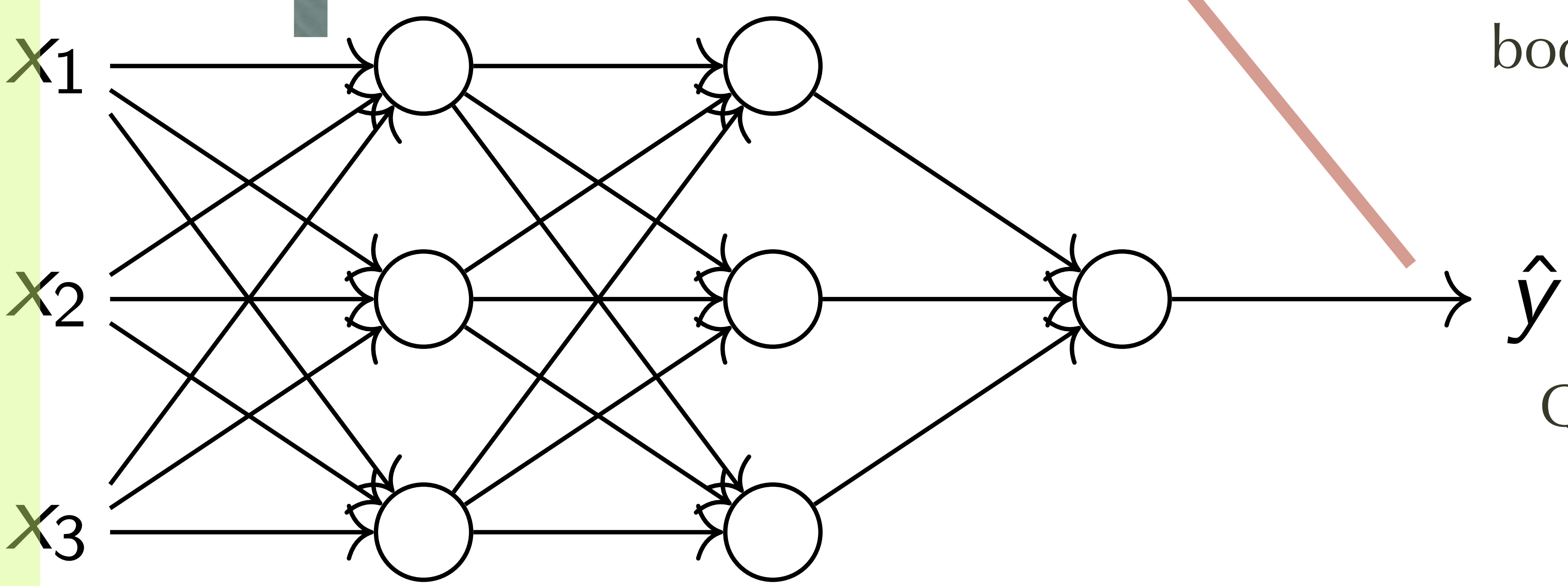
optimization $L(y, \hat{y})$

(gradient flow)

jet image
+ further arch

High level
inputs

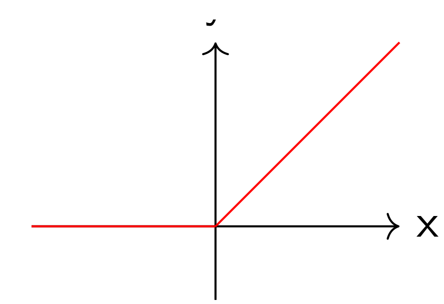
$j_i, \tau_i^{(\beta)}, EFP...$



boosted jet
$y = (1, 0)$

QCD
$y = (0, 1)$

$\varphi(w_{ij}x_i + b_i)$



Rectified Linear Unit(ReLU)

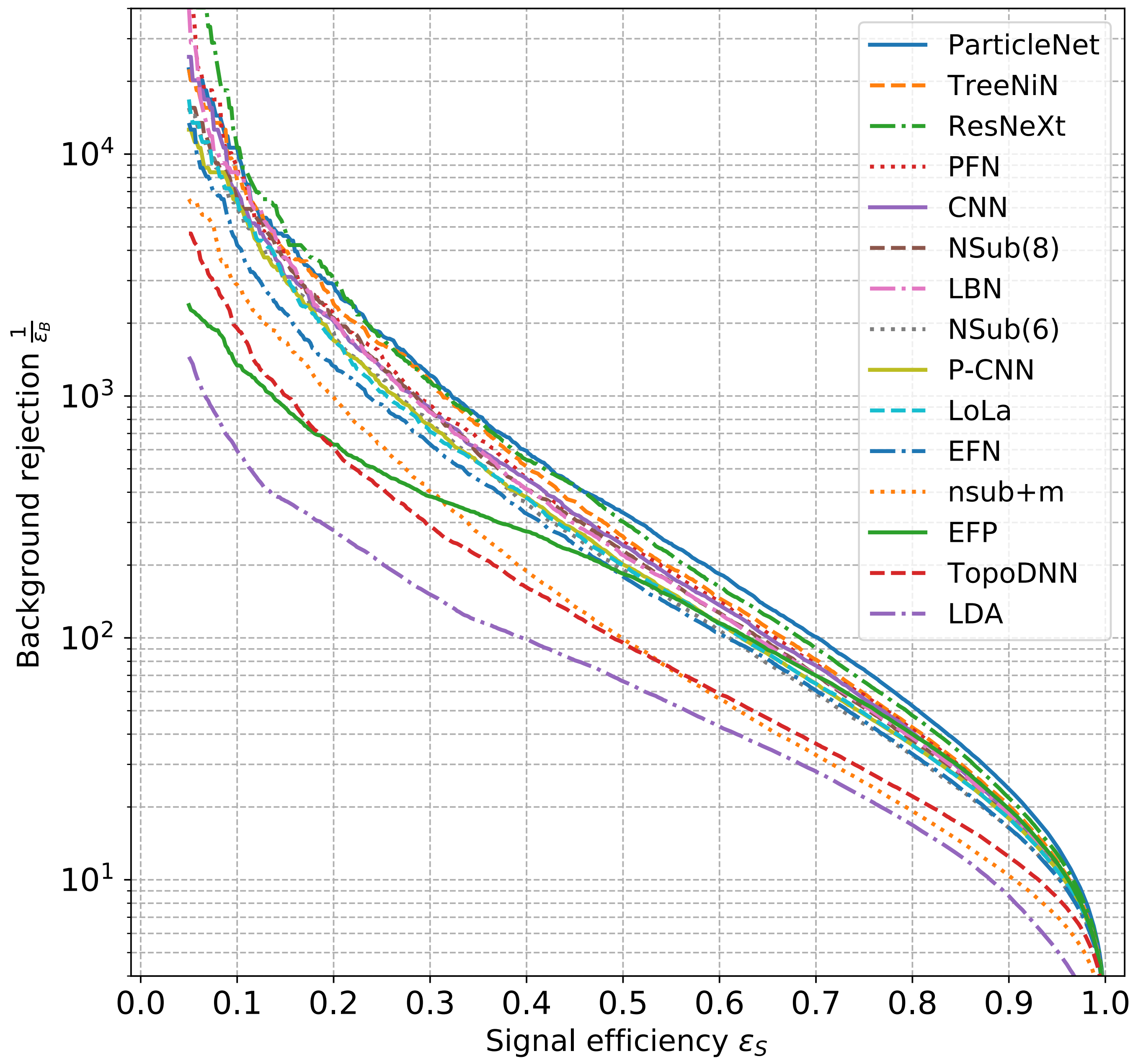
$\varphi_{\mathrm{ReLU}}(x) = x\theta(x)$

High "Representative power" if
#node #layers are large

φ: activation function
source of non linearity

Top taggers

MACHINE LEARNING LANDSCAPE OF TOP TAGGERS (1902.09914)



WARNING this plot is made with

1. detector simulation

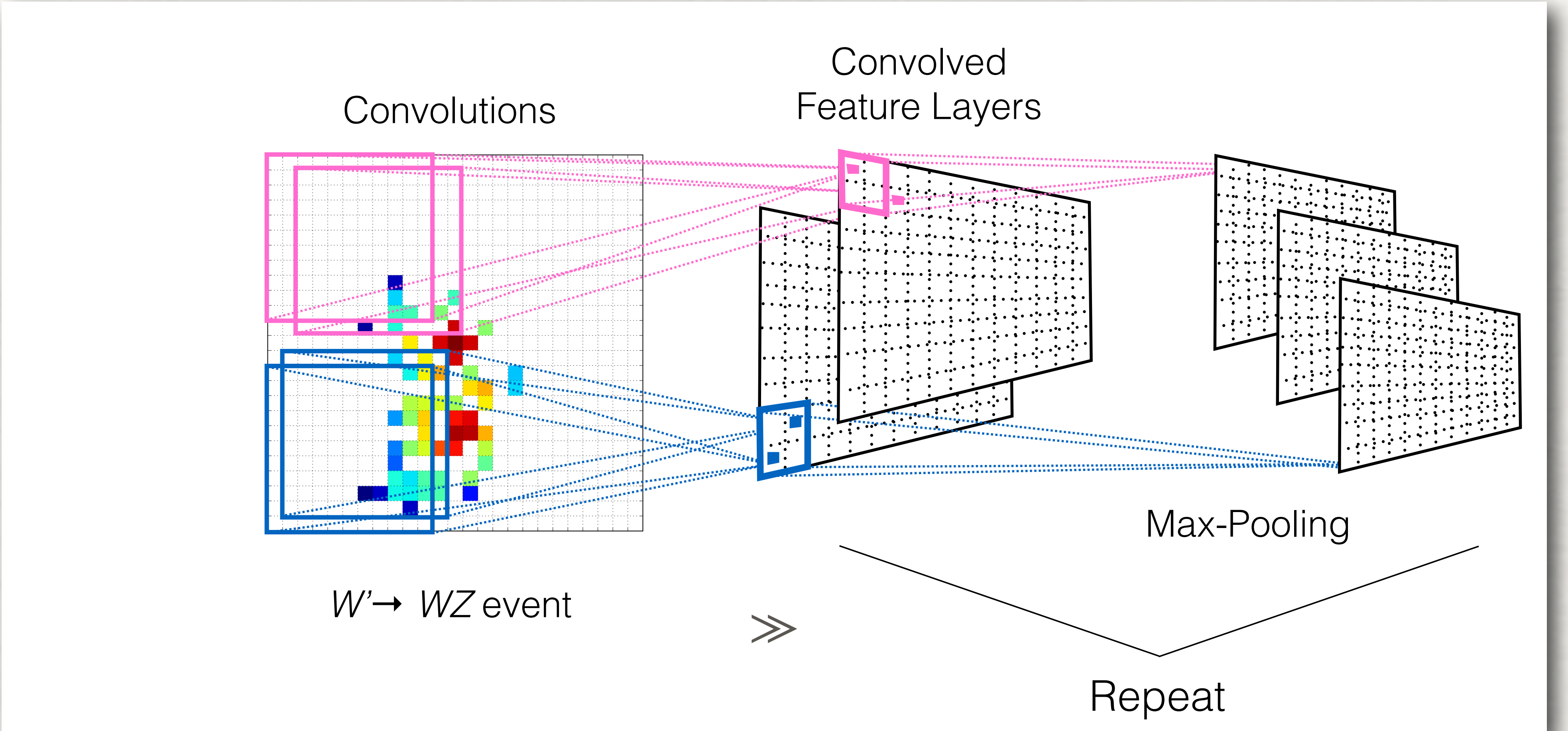
2. 550GeV<p_T<650GeV

3. no p_T marginalization
 (ML learn p_T difference)

4. no jet mass cut
(while mass cut is important)

talks in SUSY 2021
Aug 24 Jinmian Li

JET AS IMAGE (CNN)



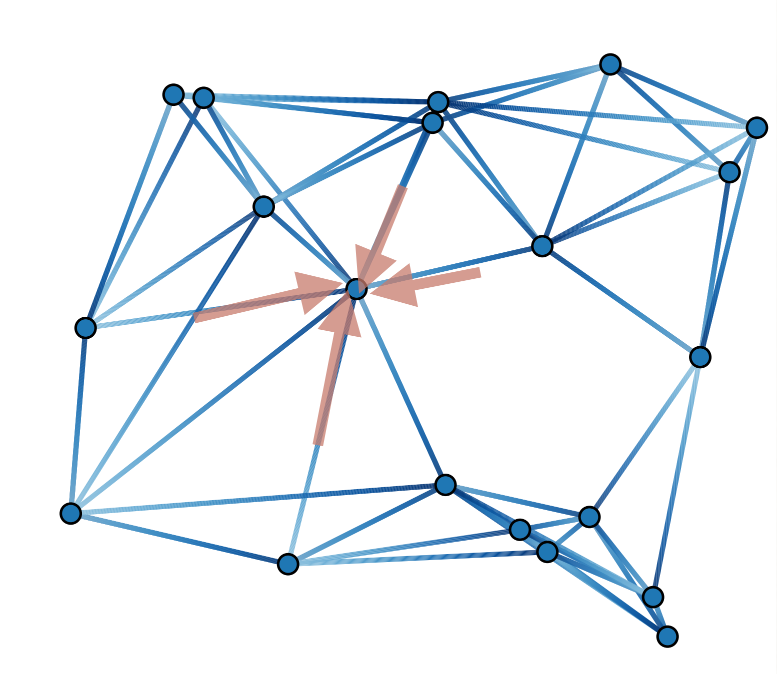
- Transfer image by NxN filter → some cutoff (pooling) →… to find correlation.
- Performance: jet Image (Ecal hit) CNN, ResNet ≫BDT based on human made variables
- **Why a NN is better than the other? What kind of event is excluded additionally? What is the key feature?**

GRAPH NEURAL NETWORK (GNN)

- ParticleNet, treat nearby two point particle correlations directly

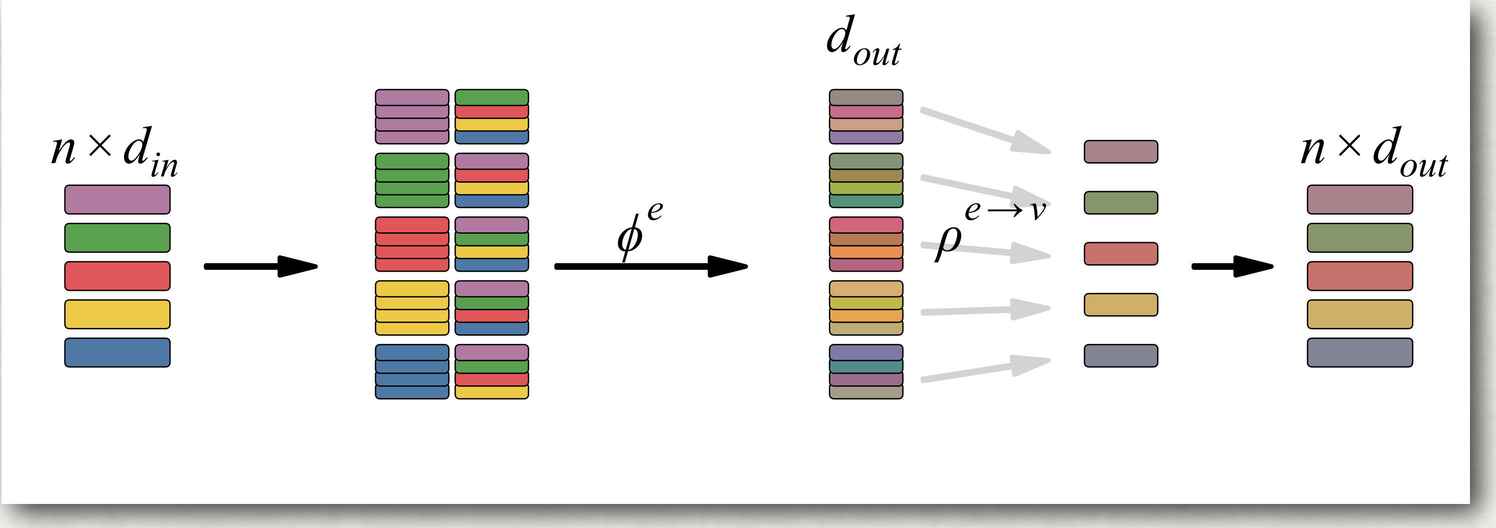
NN using
vertex (particle information)
edge (two point correlation)
as input.

Calculate edge variables and update vertex



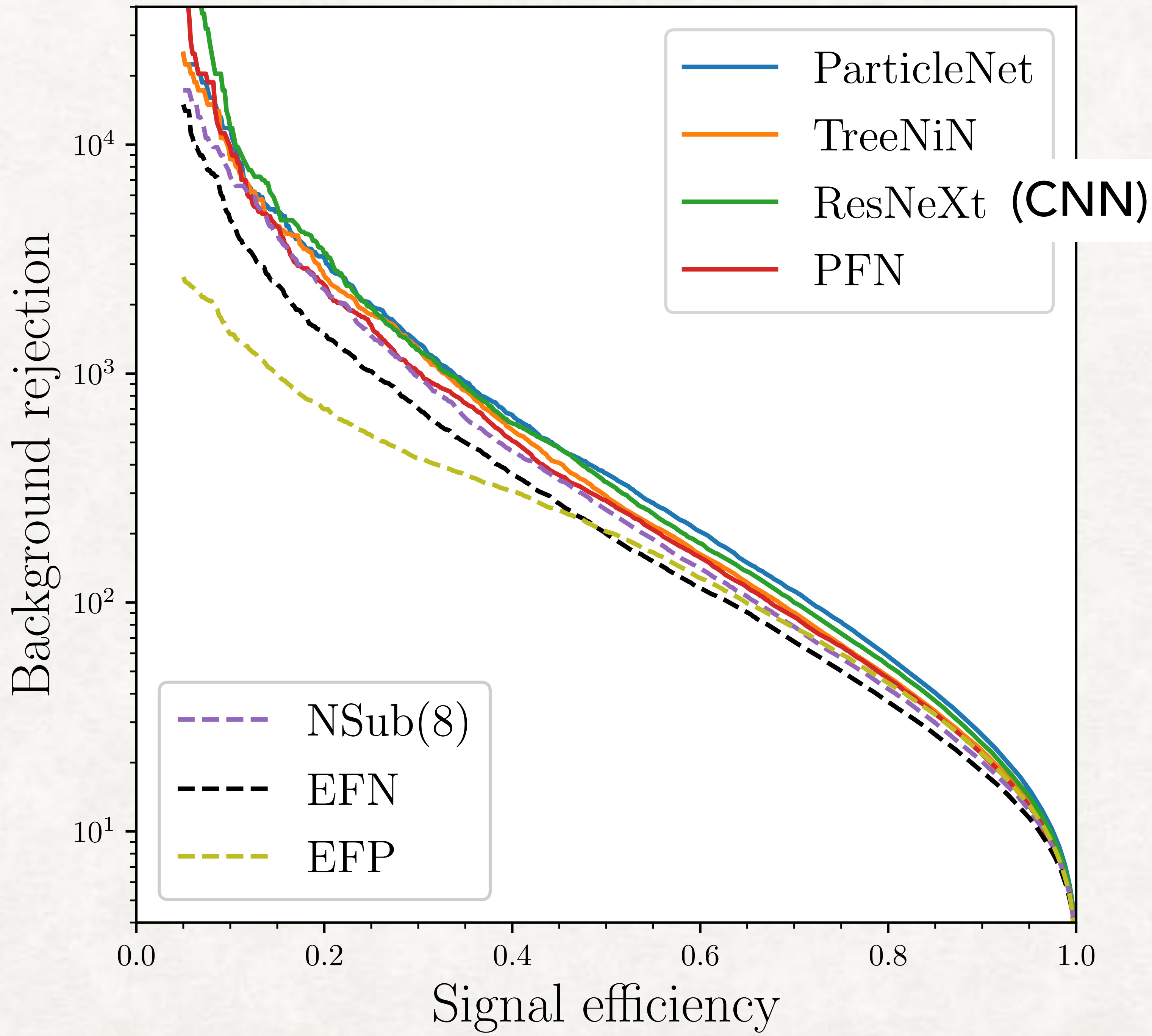
1902.08570 Qu and Goukos "Jet Tagging via Particle Clouds")

from 2007.13681 Shlomi, Battaglia Vilmant

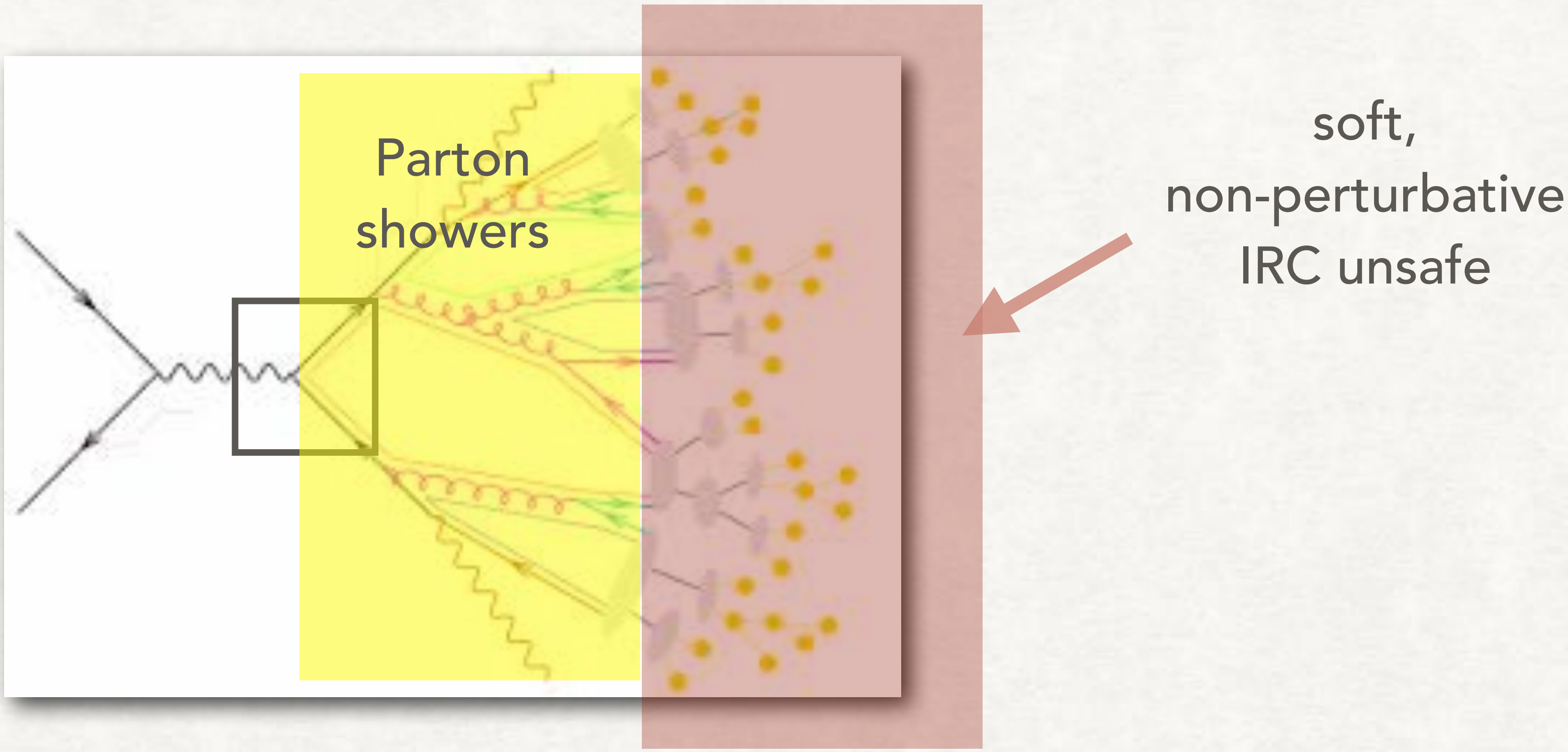


TOP TAGGERS (SELECTED)

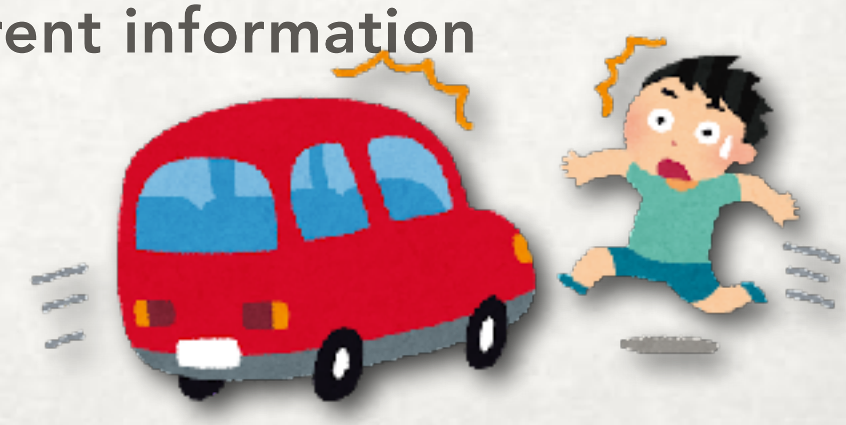
Solid, Image or particle
(general, may not be IRC safe)



WHAT IS THE ROLE OF OTHERS



- **IRC safe object:** subjet, energy correlation(C-correlator) , theoretical prediction
- **IRC "sensitive" Objects:** number of tracks, particles, soft emissions. Theoretically difficult. **MC modeling is bad** (Pythia vs Herwig , Shepa… vs real data ） **Color coherence etc.. Soft particle distribution also has parent information**
- Jet image contains both IRC safe and IRC unsafe obs. and DL may use it without prejudges



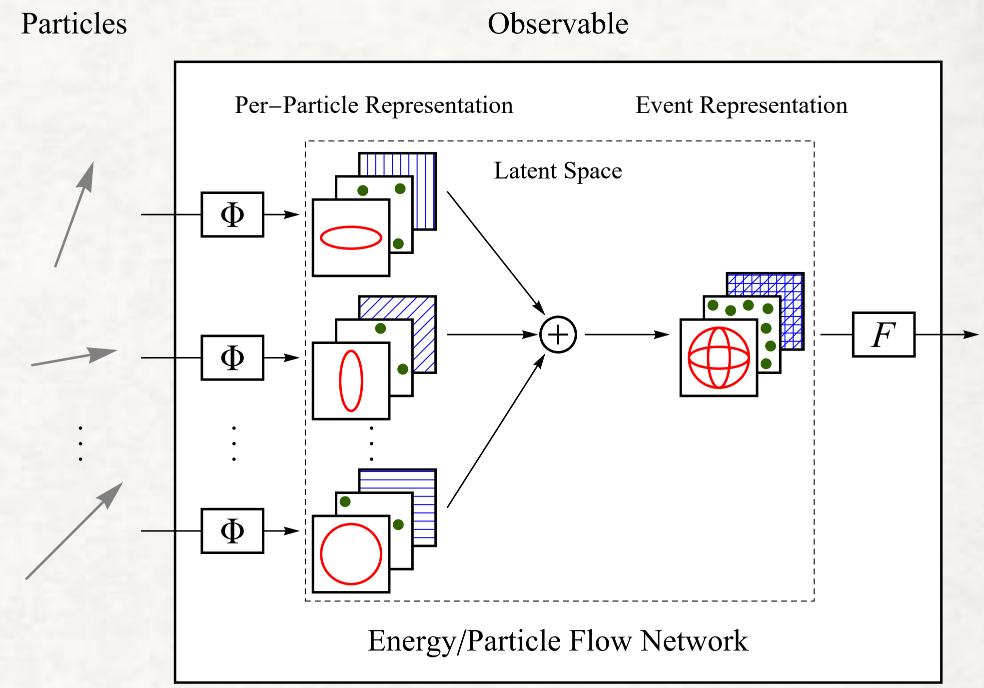
ENERGY FLOW NETWORK (IRC SAFE)

(1810.05165 KOMISKE, METODIEV, THALER)

Manifestly IRC safe set up

computationally cheaper because 1 point correlation

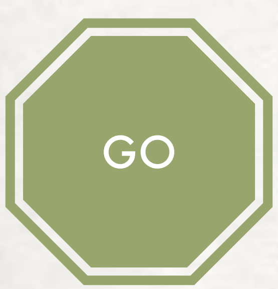
$$\mathcal{O}(\{p_1,\ldots,p_M\}) = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right),$$



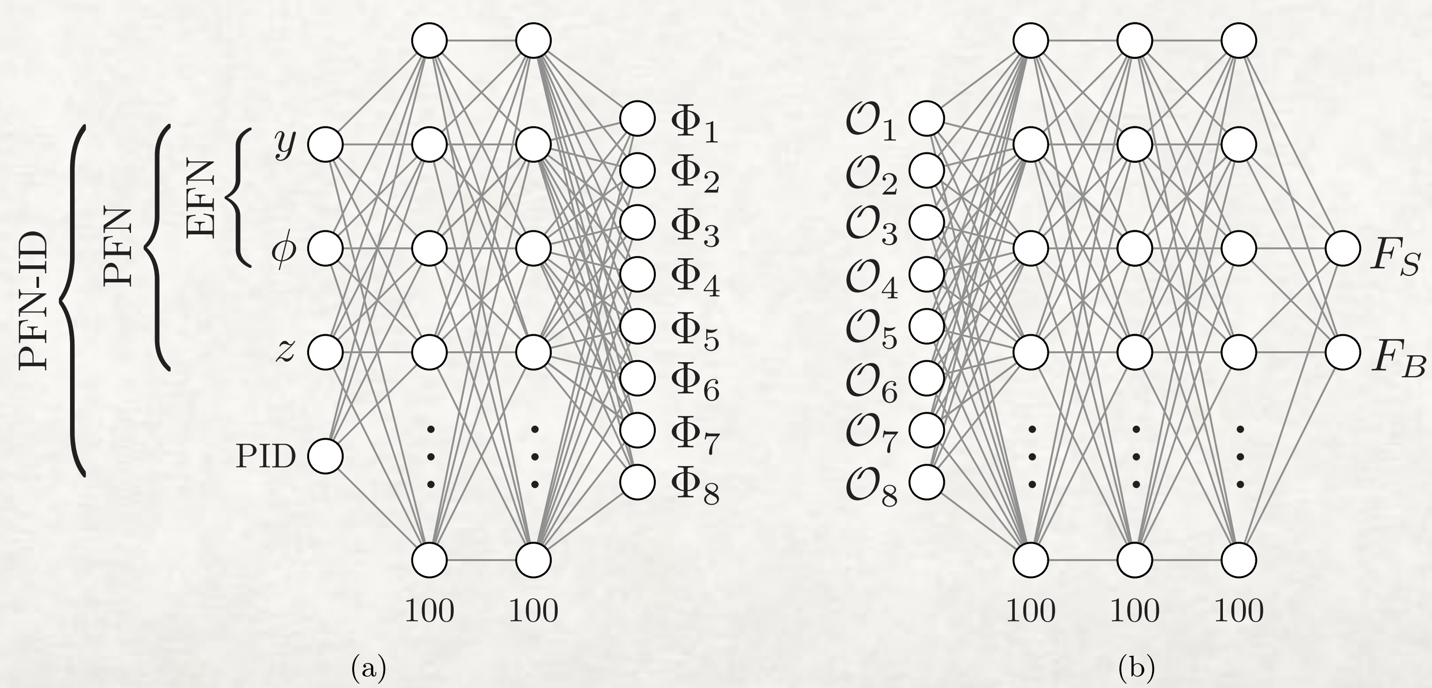
Deep set (permutation invariant)

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

IRC unsafe



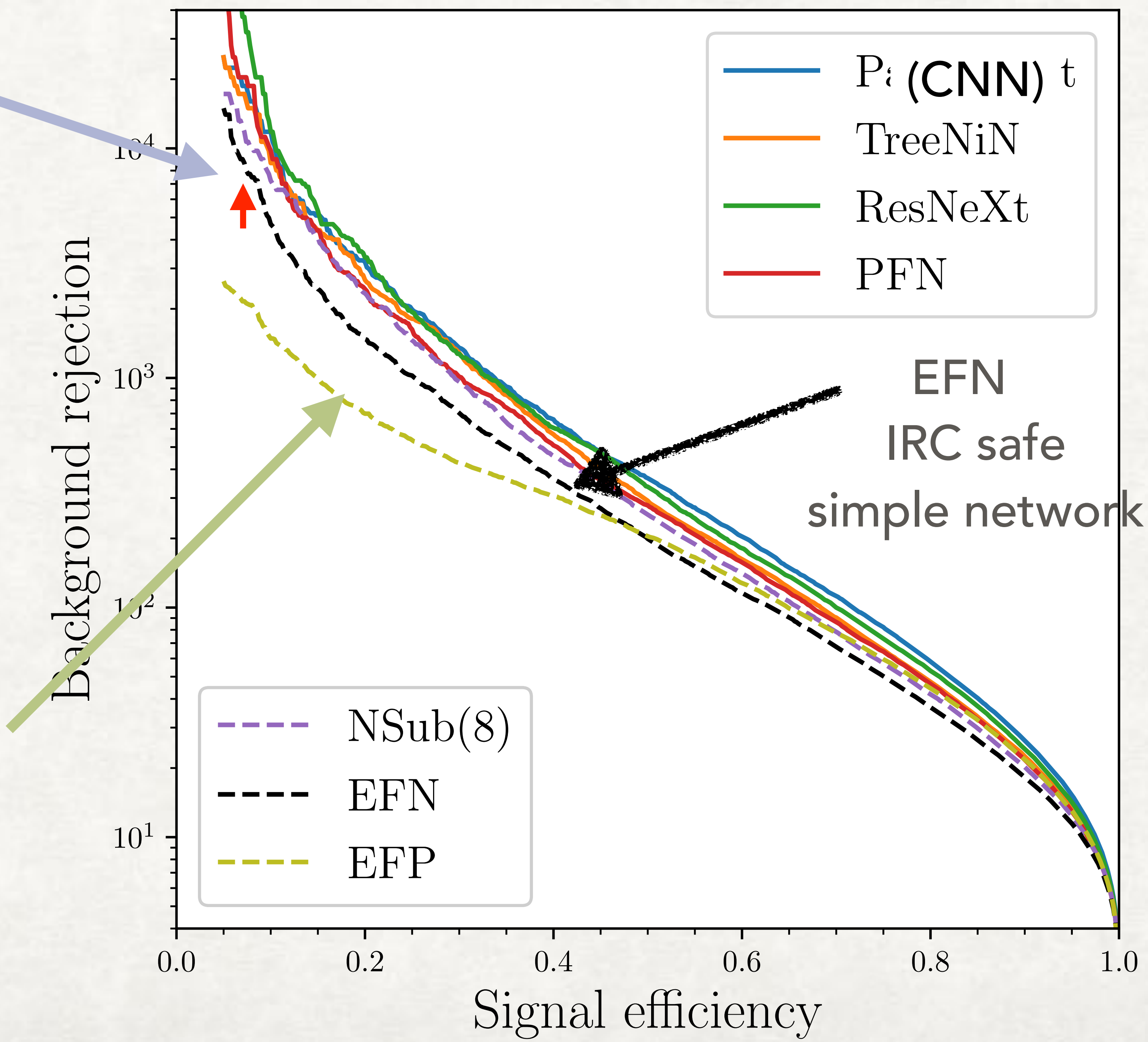
$$\text{PFN:} \quad F\left(\sum_{i=1}^{M} \Phi(p_i)\right)$$



MACHINE LEARNING LANDSCAPE OF TOP TAGGERS (1902.09914)

NSub(8): τ's up to 8 $\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \ldots, \tau_{m-2}^{(0.5)}, \tau_{m-2}^{(1)}, \tau_{m-2}^{(2)}, \tau_{m-1}^{(1)}, \tau_{m-1}^{(2)}\right\}$

(correlate with IRC unsafe quantity)



EFN
IRC safe
simple network

EFP: IRC safe and
hand made inputs

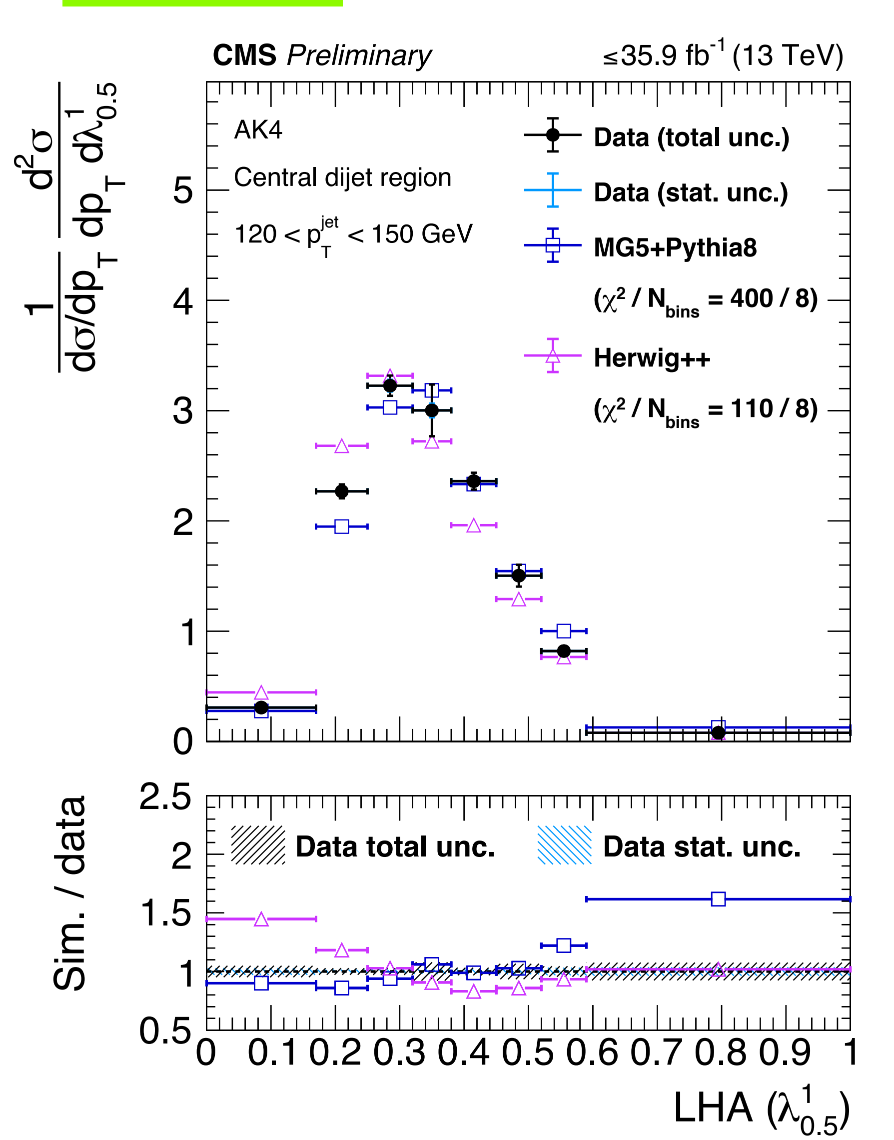
DO WE UNDERSTANDING JET PARAMETERS?

CMS PAS SMP-20-010

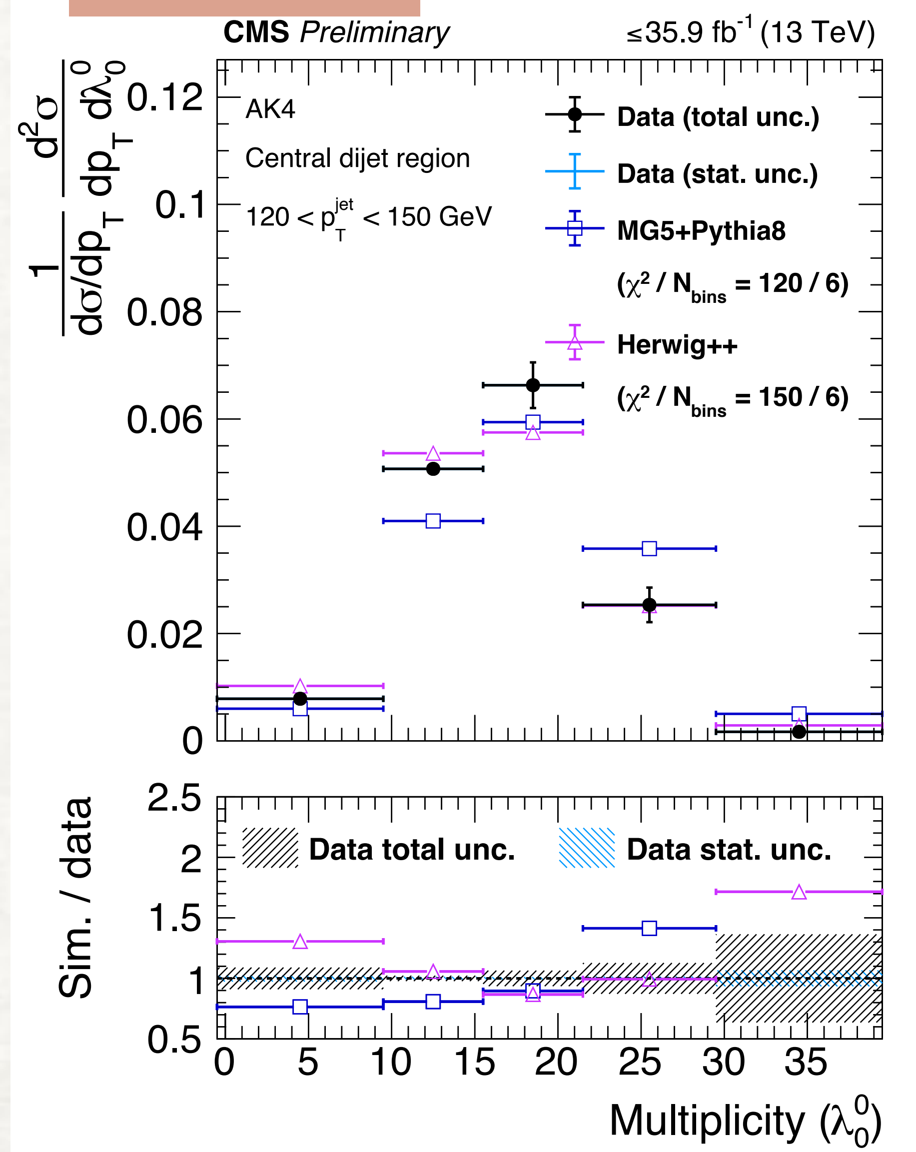
See also ATLAS JHEP 08 2019 033

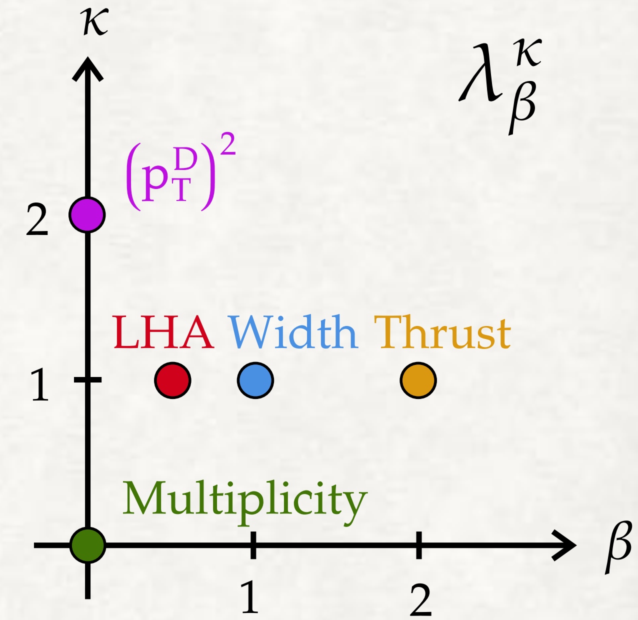
$$\lambda_\beta^\kappa = \sum_{i \in \mathrm{jet}} z_i^\kappa \left(\frac{\Delta R_i}{R} \right)^\beta ,$$

IRC safe



IRC unsafe



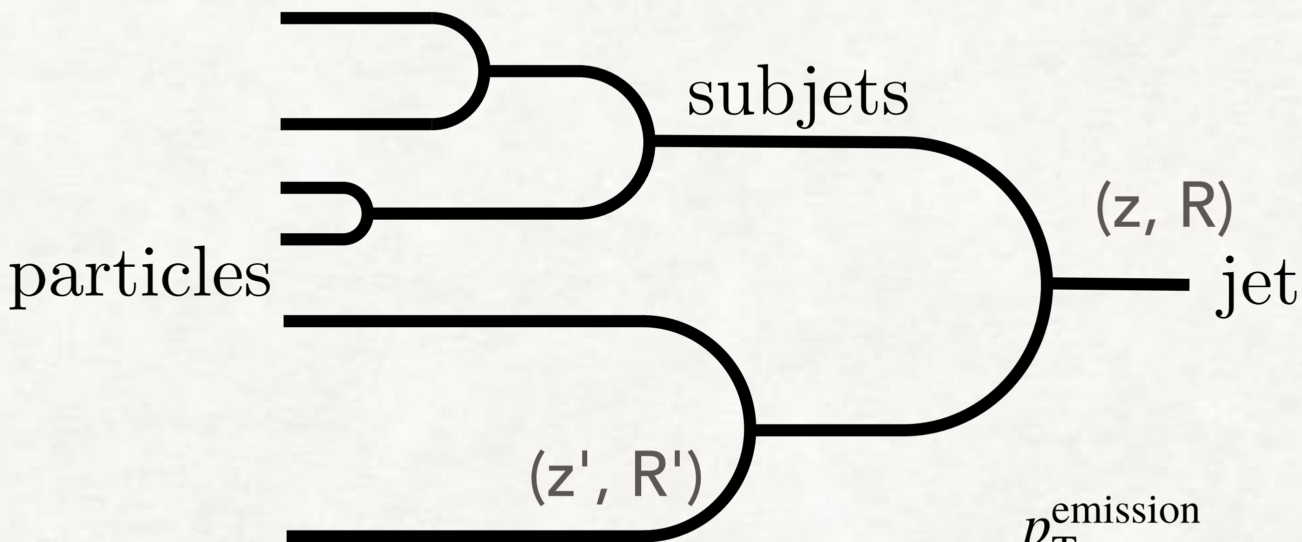


everybody has problem

LUND JET PLANE
DREYER, SALAM, SOYEZ (1807.04758)

(2004.03540 PRL 124 222002)

More reliable jet structure variable



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

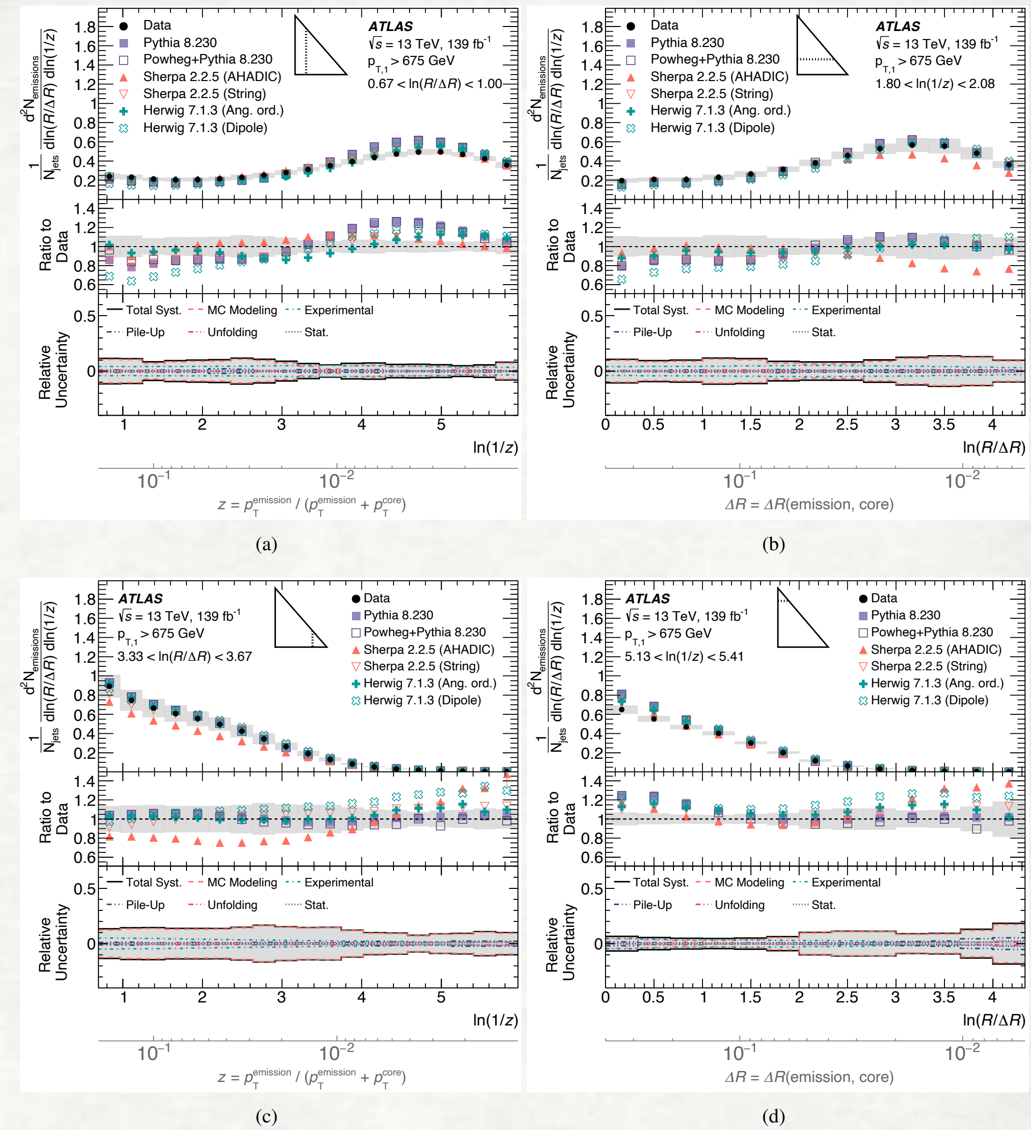
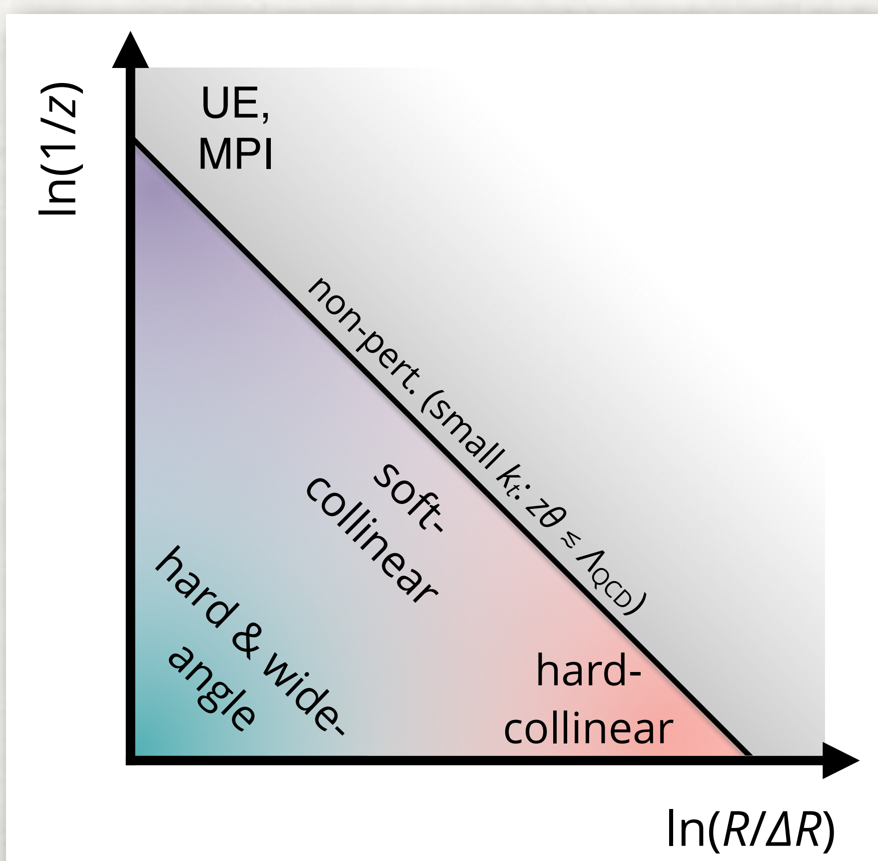


Figure 3: Representative horizontal and vertical slices through the LJP. Unfolded data are compared with particle-level simulation from several MC generators. The uncertainty band includes all sources of systematic and statistical uncertainty. The inset triangle illustrates which slice of the plane is depicted: (a) $0.67 < \ln(R/\Delta R) < 1.00$, (b) $1.80 < \ln(1/z) < 2.08$, (c) $3.33 < \ln(R/\Delta R) < 3.67$, and (d) $5.13 < \ln(1/z) < 5.41$.



(a) Schematic representation of the LJP.

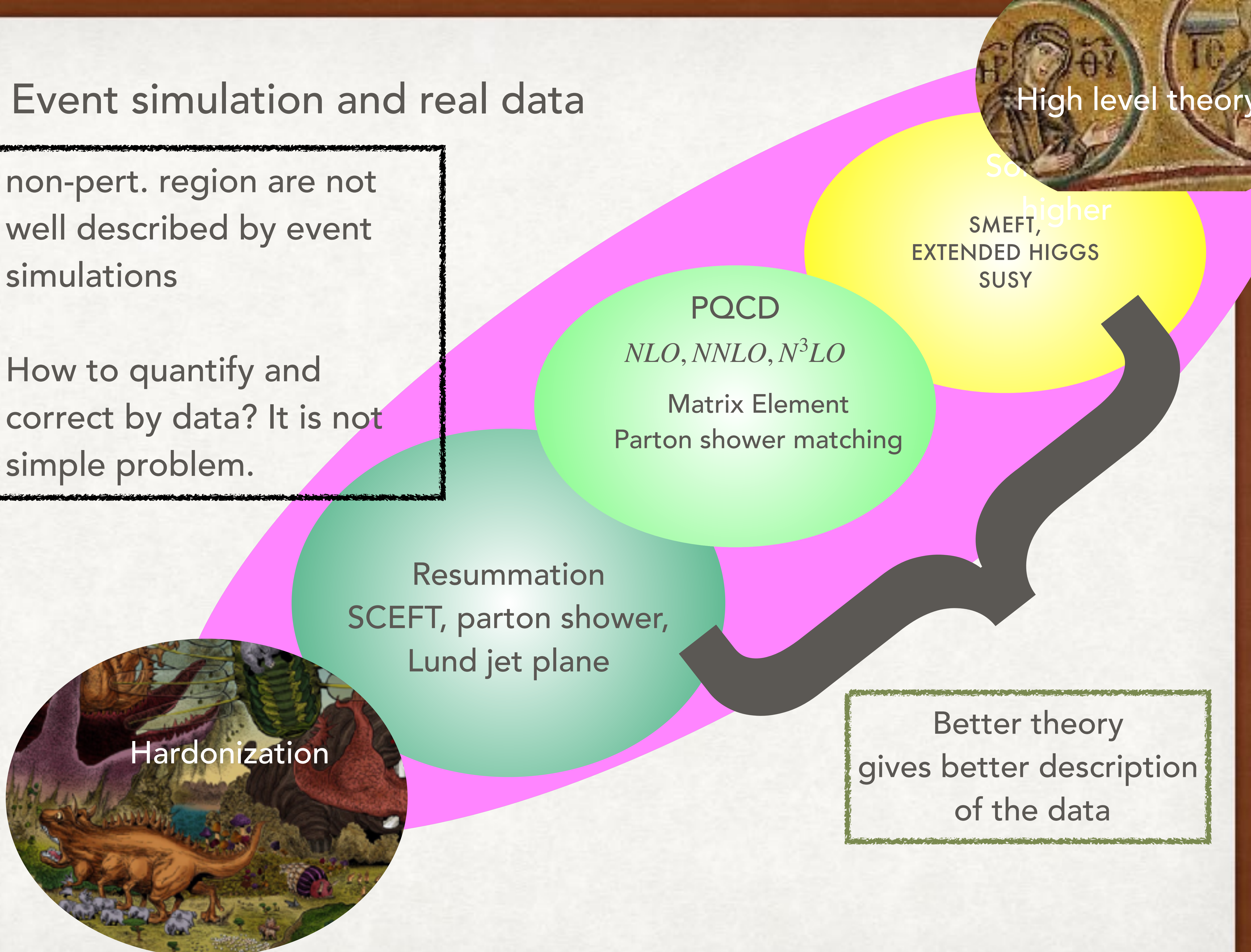
LundNet(2012.08526 Dreyer and Qu) use Lunt jet plane information and see **large improvement** from ParticleNet with limited (or low) resilience to QCD and detector effect.

(by particle level simulation for the sample with p_T>500GeV for top. no jet mass cut and no p_T marginalization)

Event simulation and real data

non-pert. region are not well described by event simulations

How to quantify and correct by data? It is not simple problem.



SOFT PARTICLE MORPHOLOGY

Sung Hak Lim and M.N. (2020, 2021)

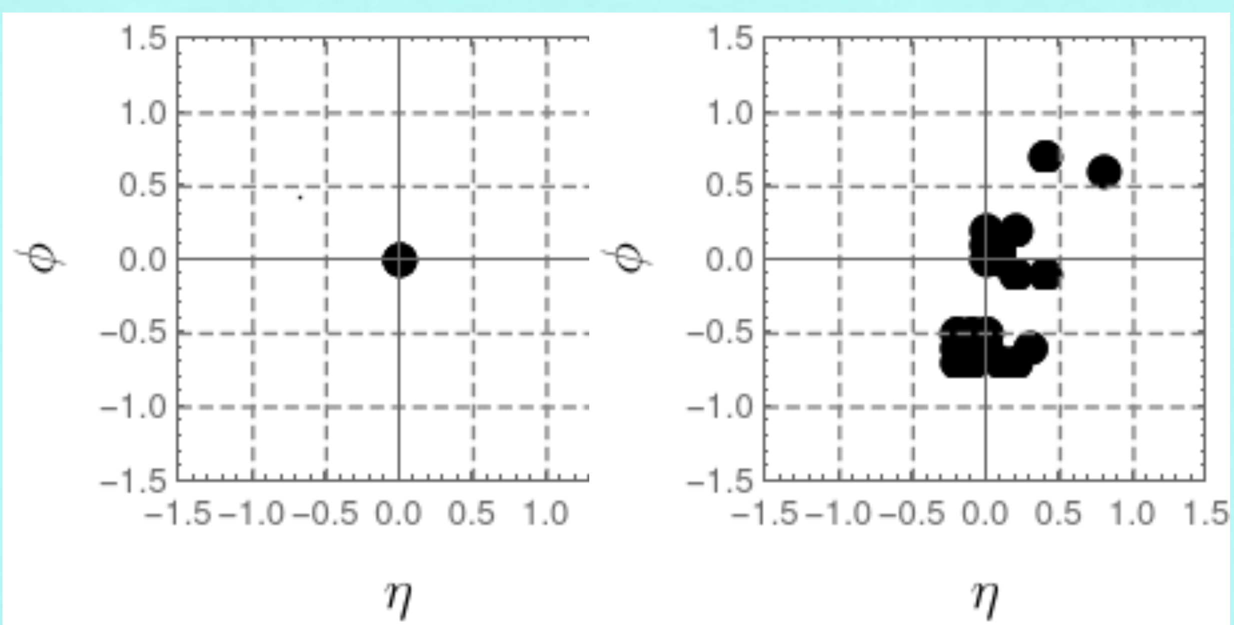
- counting soft particle & coherence. The increase number of constituents as pT cut reduces. We are not interested in the individual hits but counts and spatial distribution matters.
- geometrical information →persistent analysis & Minkowski Functional

Persistent analysis for jet

points → union of circle with size R

common in CMB analysis, material physics

String theory



Minkowski Functional

valuation

$$F(B_i \cup B_j) = F(B_i) + F(B_j) - F(B_i \cap B_j)$$

invariance in translation and rotation g

$$F(gB_i) = F(B_i)$$

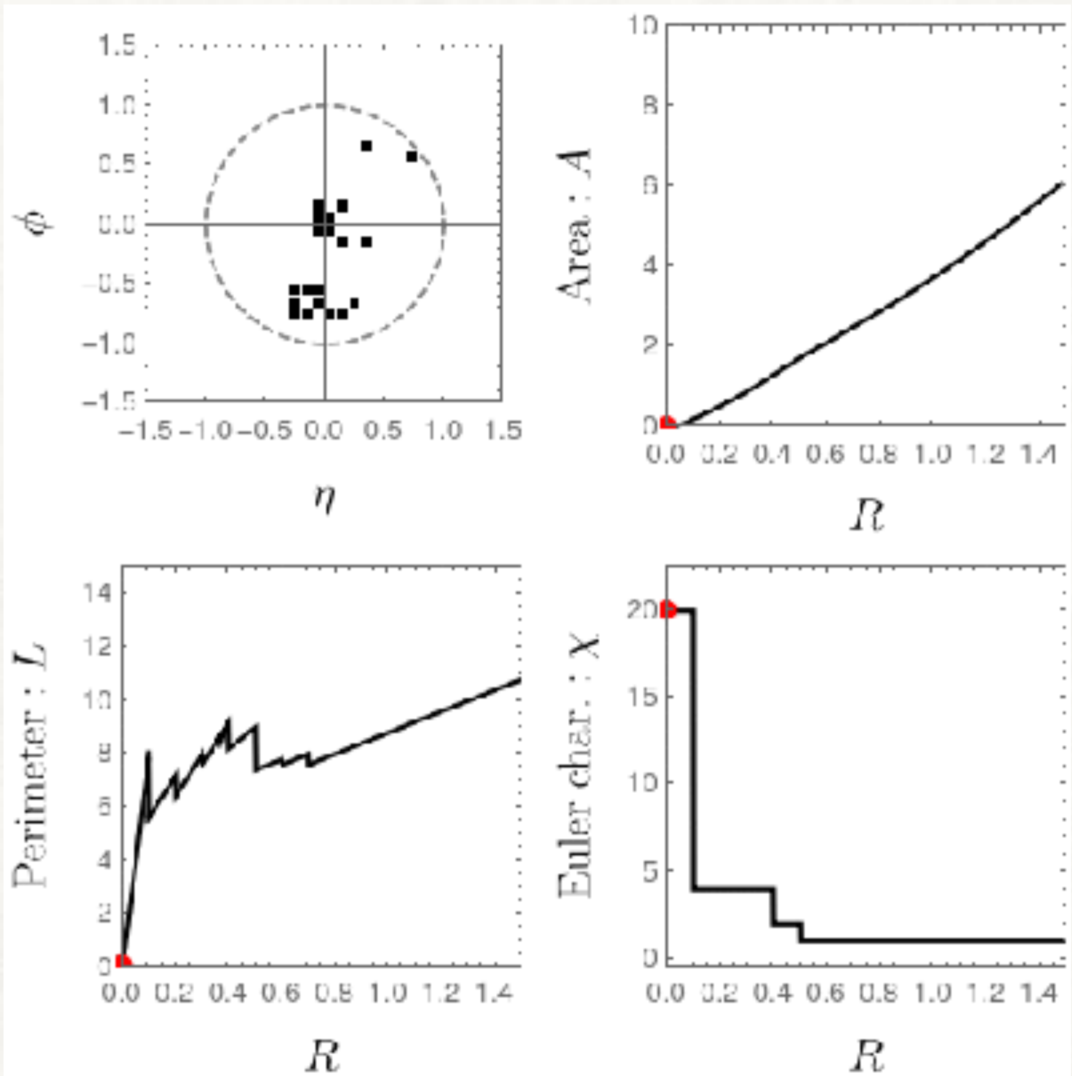
F(B)can be area A, perimeter L and euler ch χ of the image

Hadwiger's theorem in 2dim

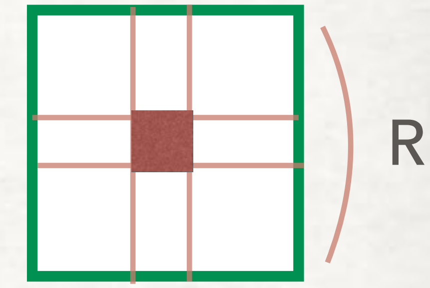
$$^{\forall}F(B_i) = c_1 A(B_i) + c_2 L(B_i) + c_3 \chi(B_i)$$

IMAGE PROCESSING

Map from Nx N bits information to N integers



$A(R), L(R), \chi(R)$



Number of pixels covered by dilated image (integers)

$A^{(0)}, A^{(1)} \dots$

Length L(0), L(1)..

up to 7

*relative distance information are encoded (counting and geometry) *

*Active bins treated equally. Reduced statistical fluctuation (order 1 $\rightarrow 1/\sqrt{N}$)

*MF has "Convolutional representation": save CNN minimization cost

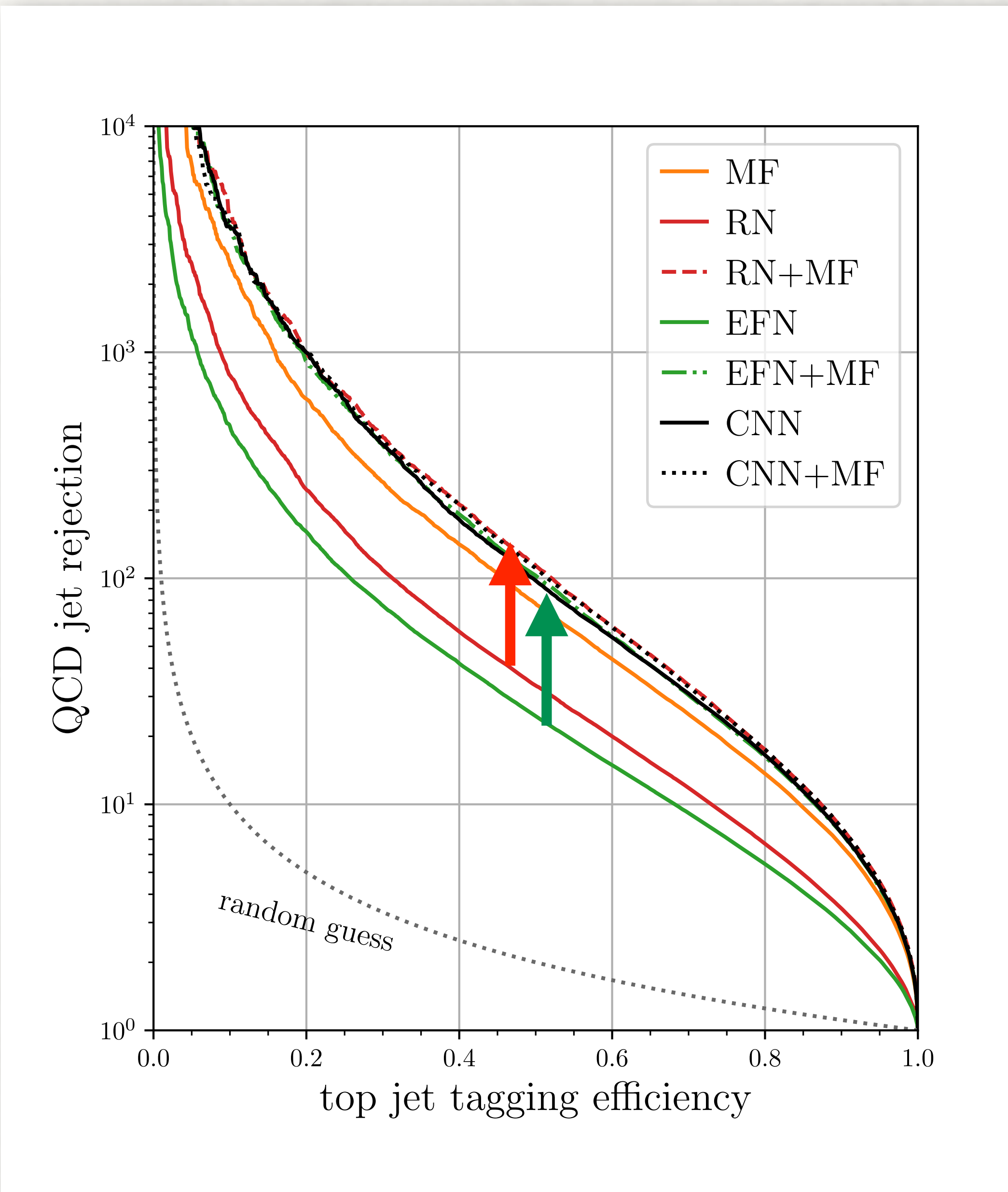
Other approaches "Jet Topology (Lingfeng Li et al 2006. 12446)"

Fractal based observables.. Davighi et al 1703.00914

Adding MF to standard networks

Higher performance(top)

$p_T^{\rm cut}$ = default,2,4,8



Faster training (> CNN)

	AUC	$t_{\rm train}/N_{\rm epoch}$ $N_{\rm batch}=20$	$t_{\rm train}/N_{\rm epoch}$ $N_{\rm batch}=200$
MF	0.9467	793 s / 564 epochs	954 s / 363 epochs
RN	0.9038	288 s / 186 epochs	619 s / 214 epochs
RN+MF	0.9552	418 s / 255 epochs	1057 s / 288 epochs
CNN	0.9529		31020 s / 1483 epochs
CNN+MF	0.9547		12319 s / 530 epochs
EFN	0.8900	535 s / 120 epochs	723 s / 108 epochs
EFN+MF	0.9521	725 s / 149 epochs	813 s / 111 epochs

Warning

detector simulation

$$500{\rm GeV} < p_{Tj} < 650{\rm GeV},$$

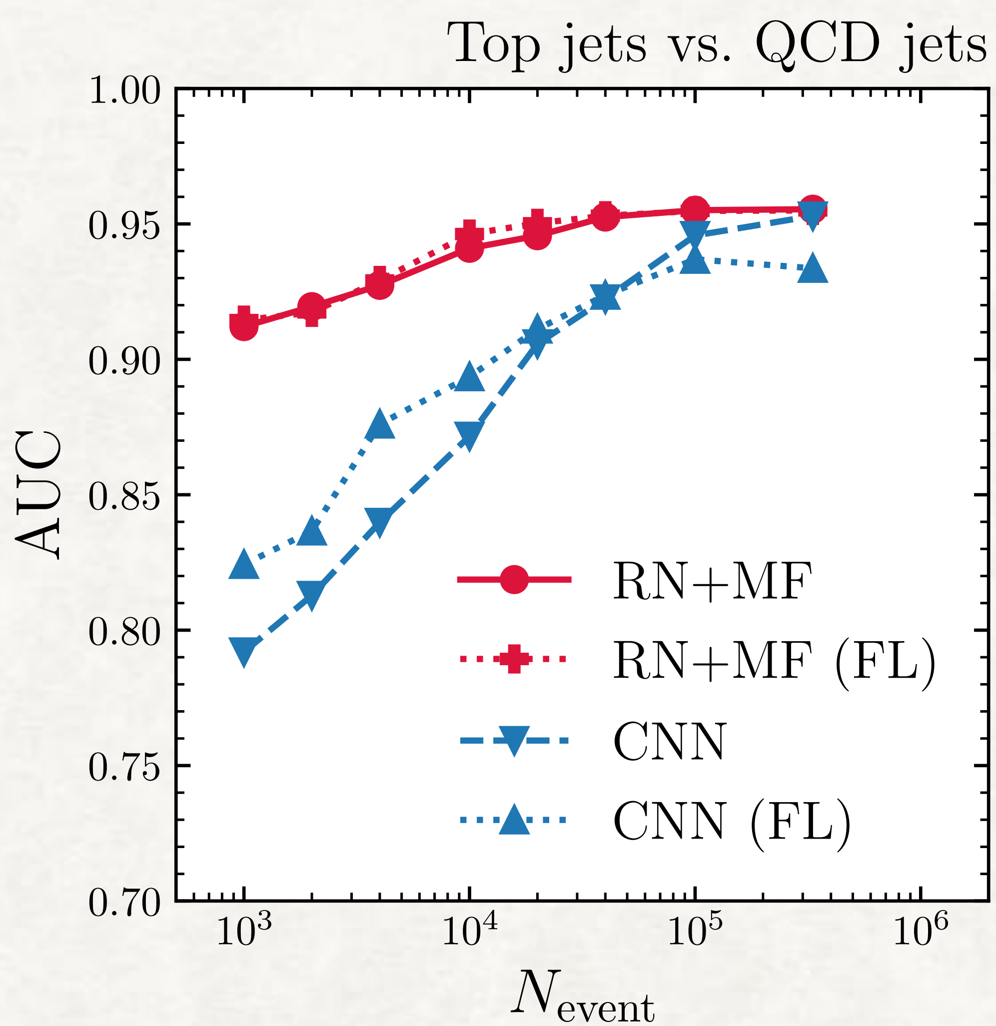
$$150{\rm GeV} < m_j < 200{\rm GeV}$$

p_T marginarized 1.5×10^5 samples

TRAINING STABILITY

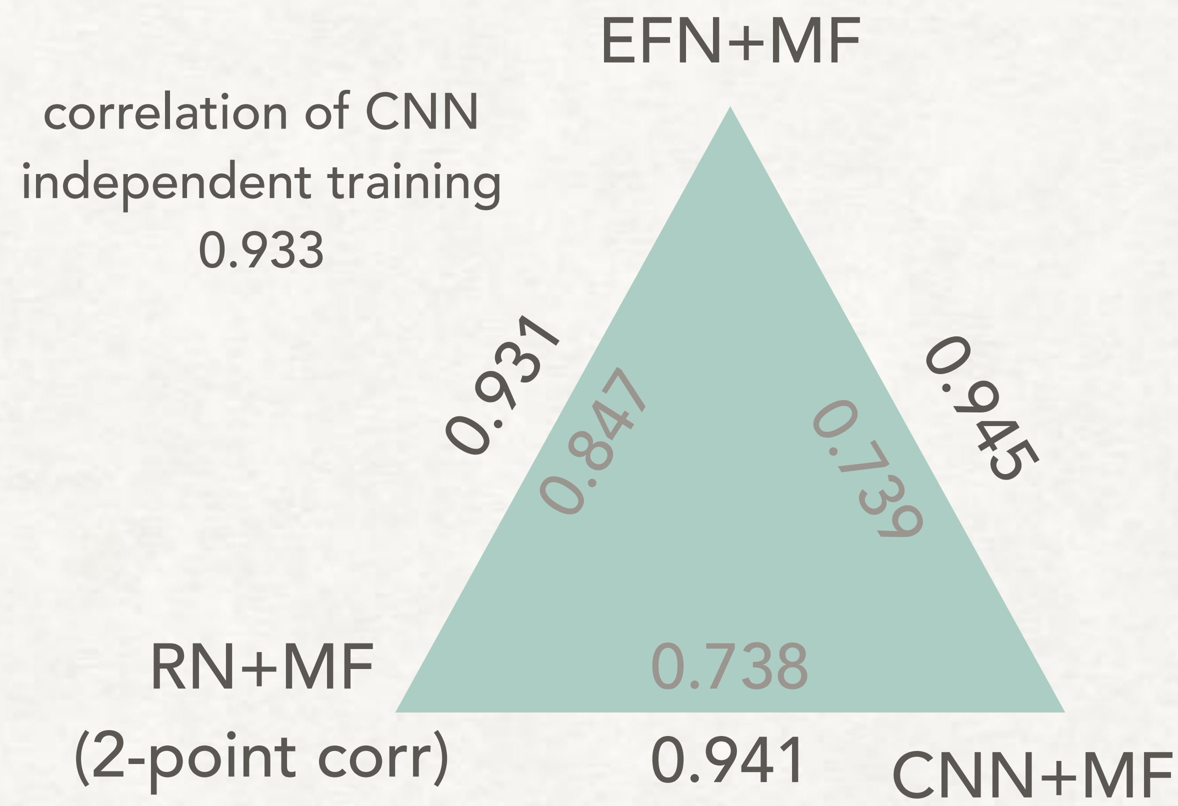
same performance with smaller samples

RN(relation network) utilize aggegated 2-point energy correlation.



no cost on fining feature variables and the performance is same

Correlation of the output of the model with MF



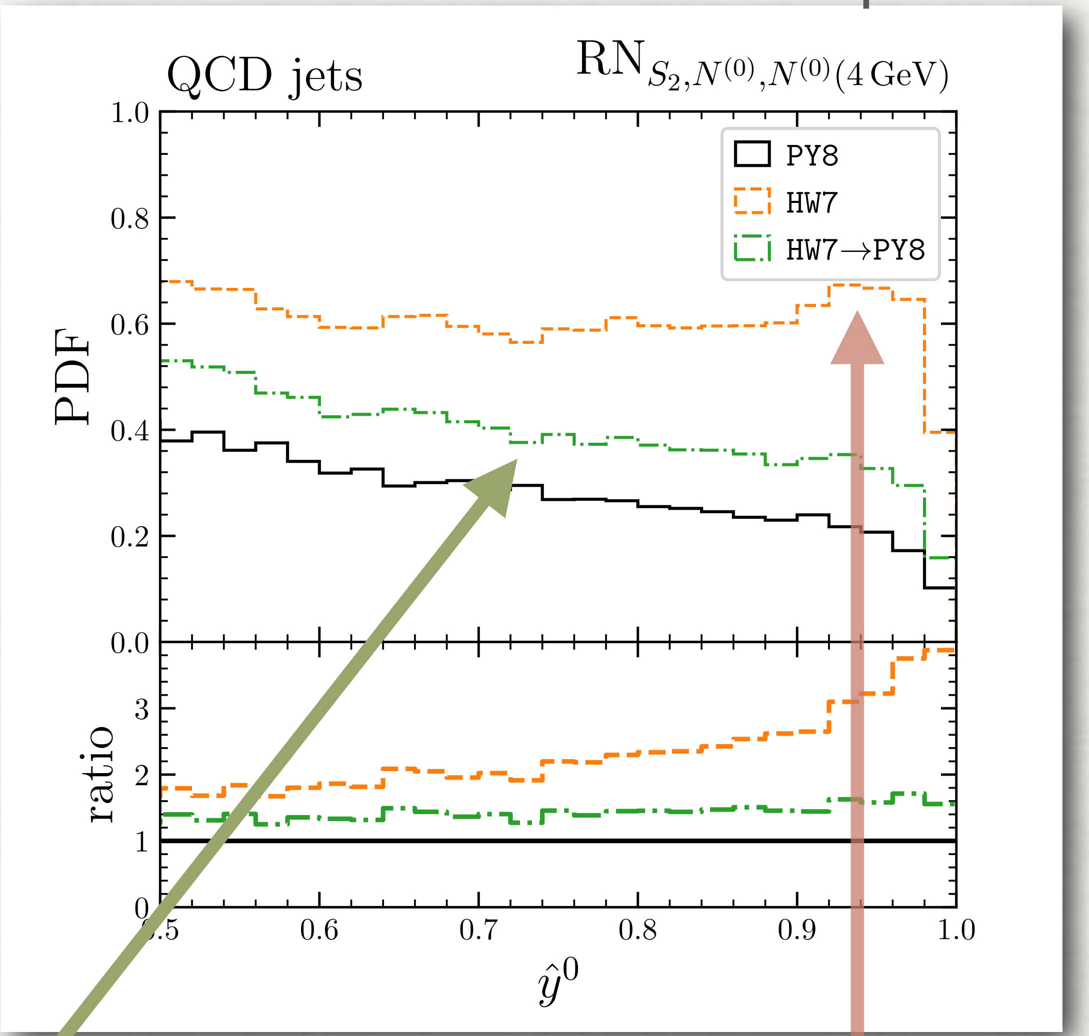
MF is efficient way to express soft particle distribution without using bin by bin information

CALIBRATION

Classifier is sensitive to IRC unsafe quantity,
→MC distribution have to be
tuned by the data

- ex Classifier trained by simulated data find fake signal in real data
- Event reweighing via MF
- Multi-dim reweighting using MaskedAutoagressiveFlow in progress (with Furuichi, Lim)

Top like



Reweighting to reproduce
"correct" number of pixel

Pythia trained classifier regard
some of Herwig QCD jet as "top"

GAN based approach abyBaldi et al
ArXiv 2012.11944
"How to GAN Higher jet resolution"

SUMMARY

- LHC →HL-LHC: access to high energy tail region. boosted object is more important.
- DL allows us to utilize multi-correlation among jet variables. Event reconstruction using QCD features (jet, rapidity gap etc)
- Current success of LHC is based on deeper QCD understanding. good interplay between QCD-EventSimulation-Experiment would be important for jet physics
- More and more interesting applications of DL => see Tilman's talk.

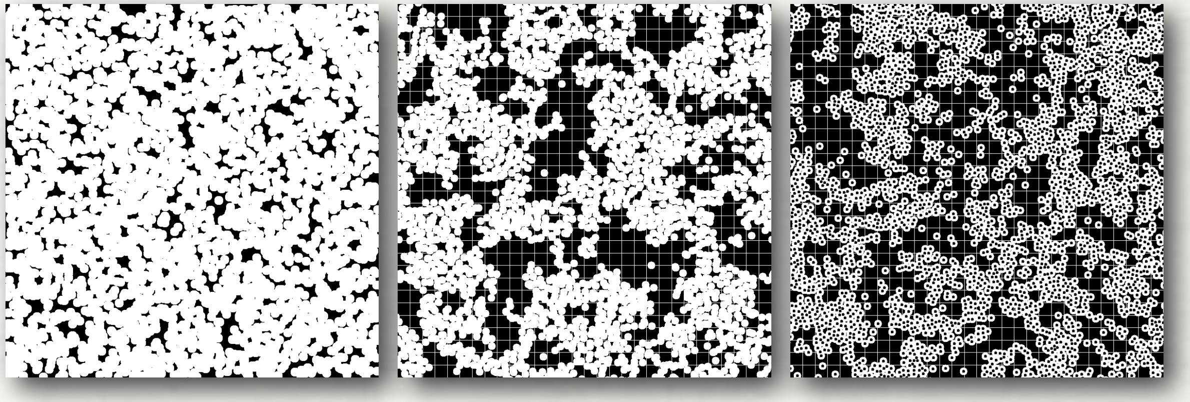
BACKUPS

APPLICATION OF MF IN THE OTHER FIELD

Statistical Physics

Occupation V, Surface(S) → nature of material

Mecke and Stoyan (2000)



Porous micro emulsion colloid

Mecke and Stoyan (2000)

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

Powerful to quantitatively describe point distribution

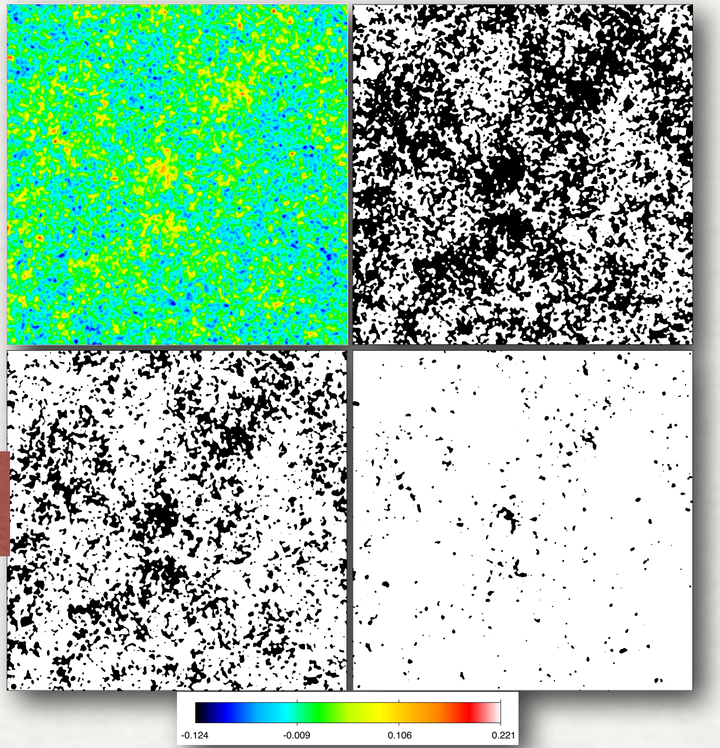


FIG. 1: *Top left panel: example of a simulated 12-square-degree convergence map in the fiducial cosmology, with intrinsic ellipticity noise from source galaxies and $\theta_G = 1$ arcmin Gaussian smoothing. A source galaxy density of $n_{gal} = 15/\text{arcmin}^2$ at redshift $z_s = 2$ was assumed. 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 Functinals

MF: PYTHIA VS HERWIG DIJET

$$\mathcal{D}(i) = \frac{f_P(i) - f_H(i)}{f_P(i) + f_H(i)},$$

$$f_A(i) = \frac{N_A(i)}{\sum_i N_A(i)}$$

N_H: number of herwig sample in a bin

N_P : number of pythia sample in a bin

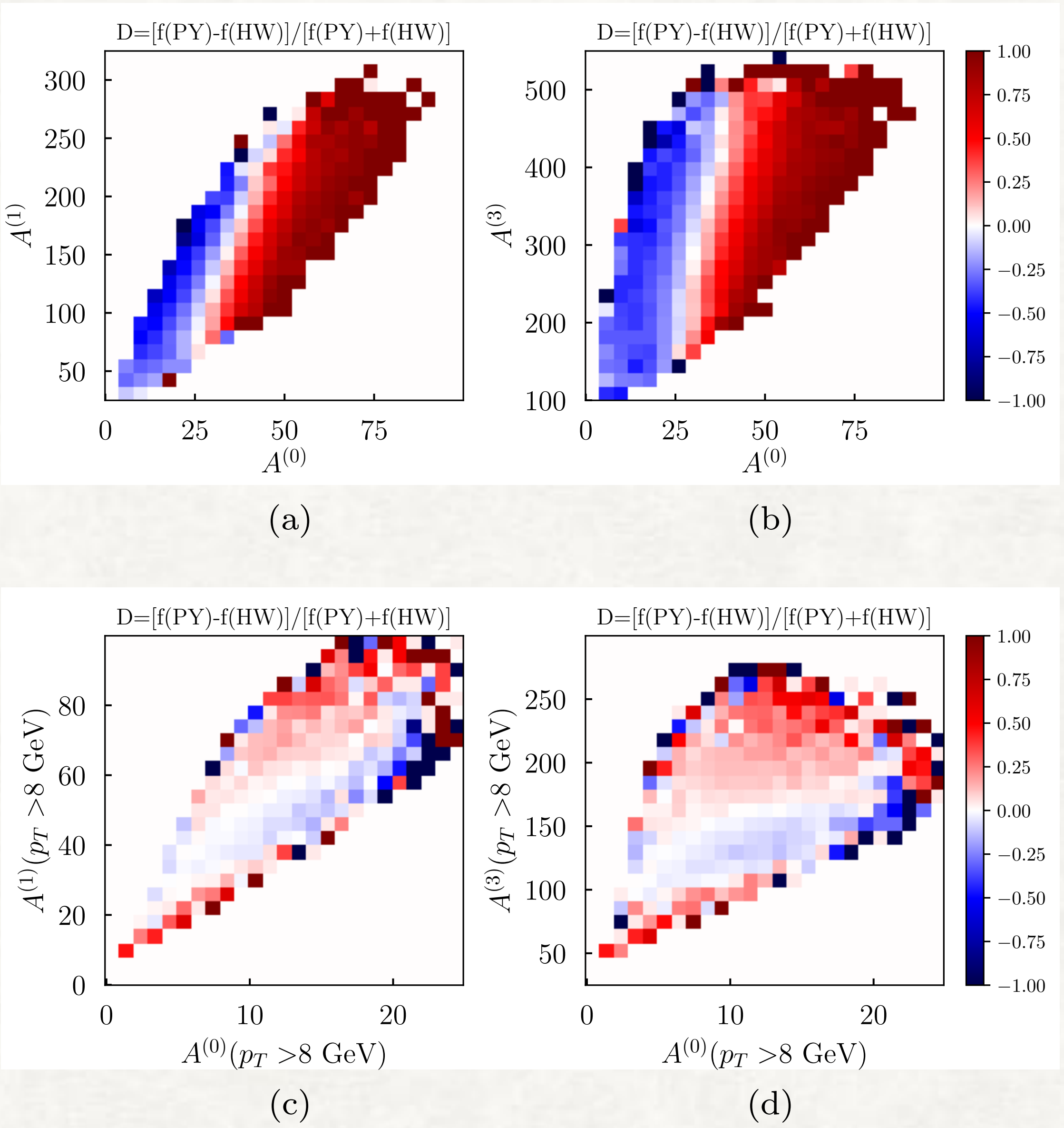


FIG. 8. The asymmetry $\mathcal{D}$ of the $(A^{(0)}, A^{(k)})$ distributions simulated by PYTHIA8 and HERWIG7. Figures (a) and (c) show the asymmetry of $(A^{(0)}, A^{(1)})$ distributions. Figures (b) and (d) show the asymmetry of $(A^{(0)}, A^{(3)})$ distributions. No p_T filter is applied to (a) and (b), while $p_T > 8$ GeV filter is applied for (c) and (d).

RELATION NETWORK

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

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