Interpretable Deep Learning for Two-Prong Jet Classification with Jet Spectra

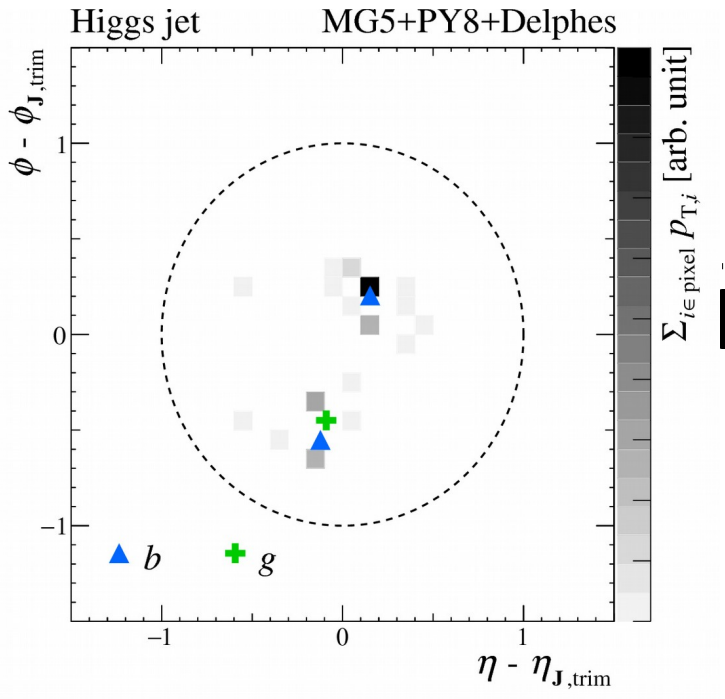
Sung Hak Lim
Theory Center, KEK

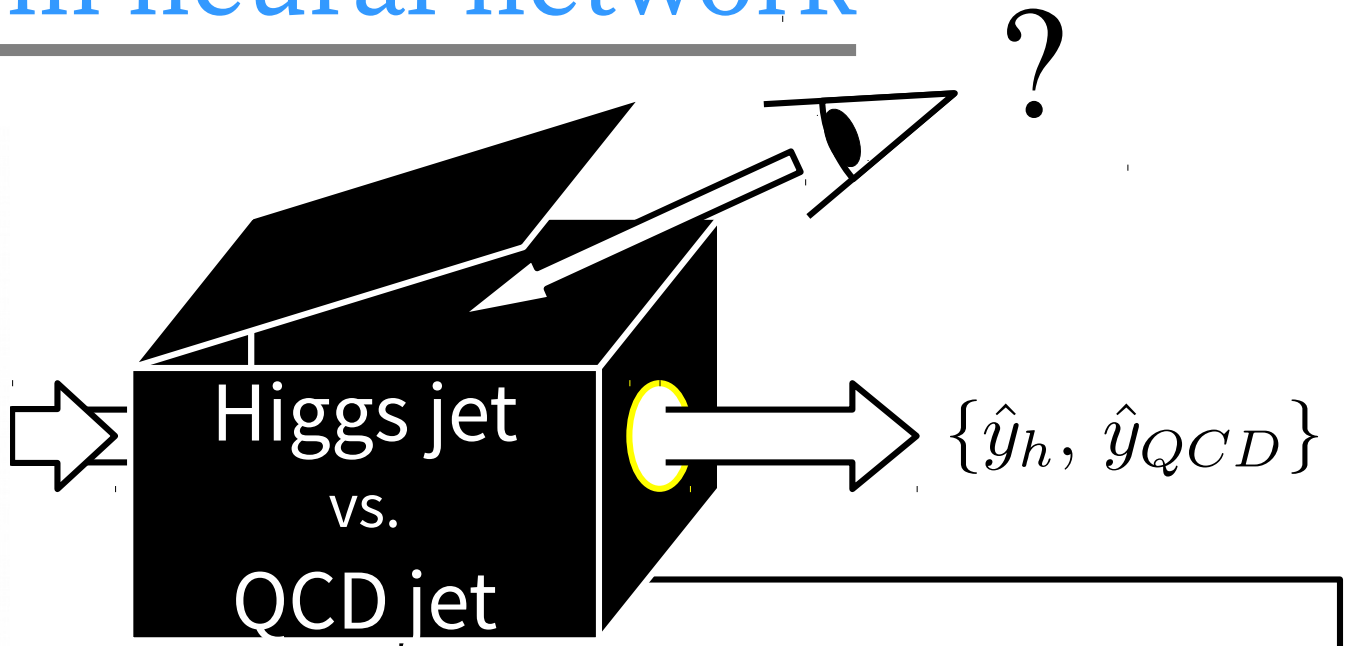
BOOST 2019, MIT
July 2019

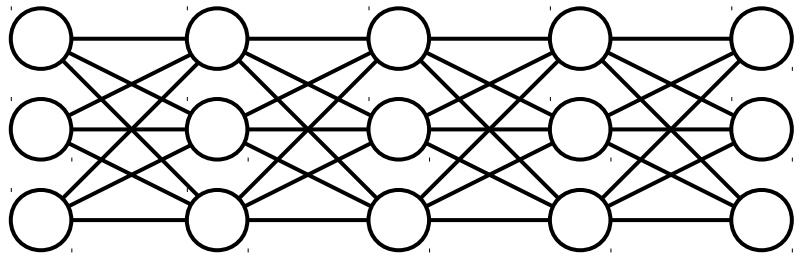
published
7 hours ago

S. H. Lim, M. M. Nojiri, arXiv:1807.03312, JHEP10(2018)181.
A. Chakraborty, **S. H. Lim**, M. M. Nojiri, arXiv:1904.02092, JHEP07(2019)135.
A. Chakraborty, **S. H. Lim**, M. M. Nojiri, M. Takeuchi, in preparation.

Difficulties on understanding the results from neural network



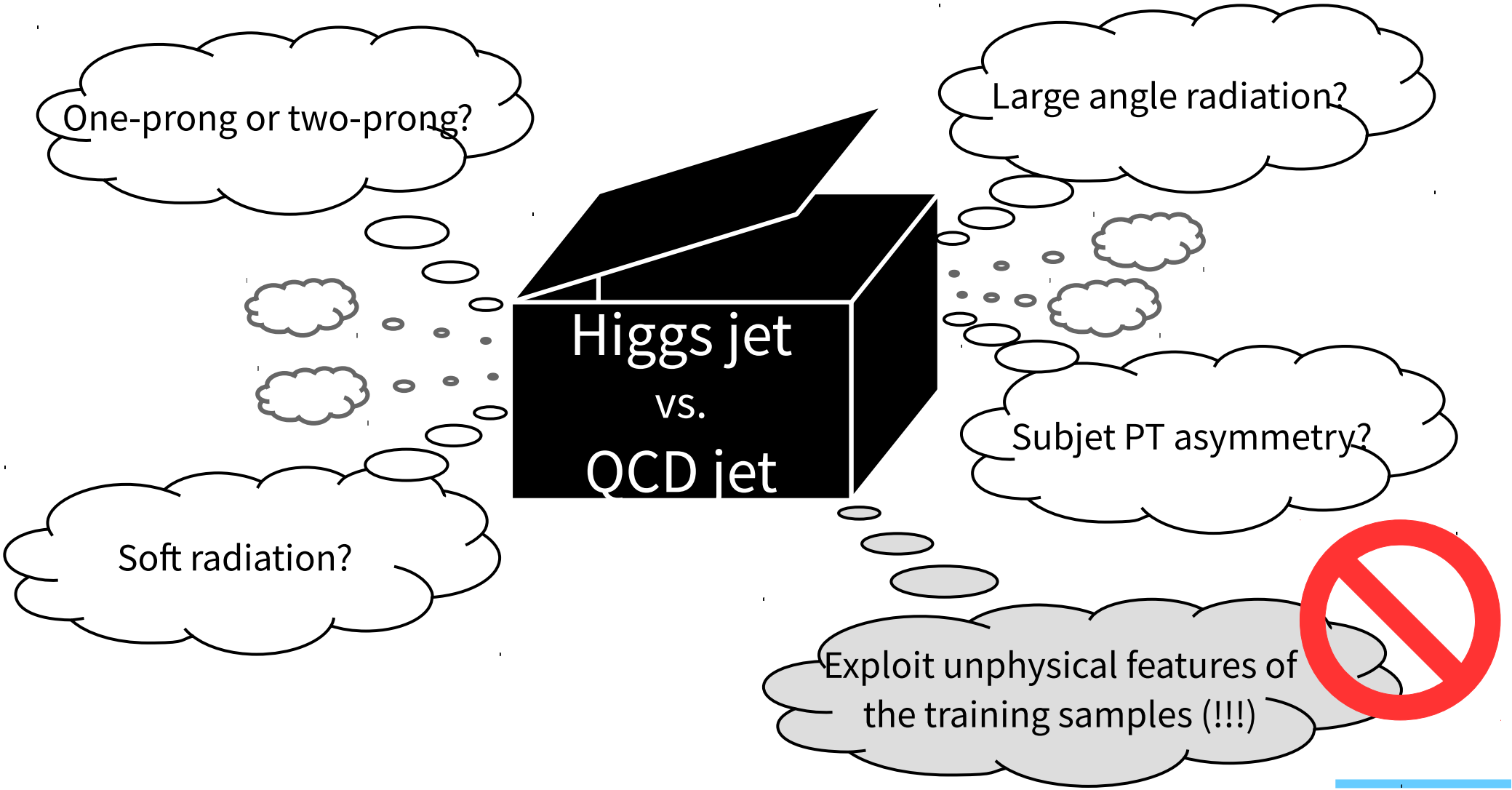




$$h_i^{(n)} = \varphi(w_{ij}^{(n)} h_j^{(n-1)} + b_i^{(n)})$$

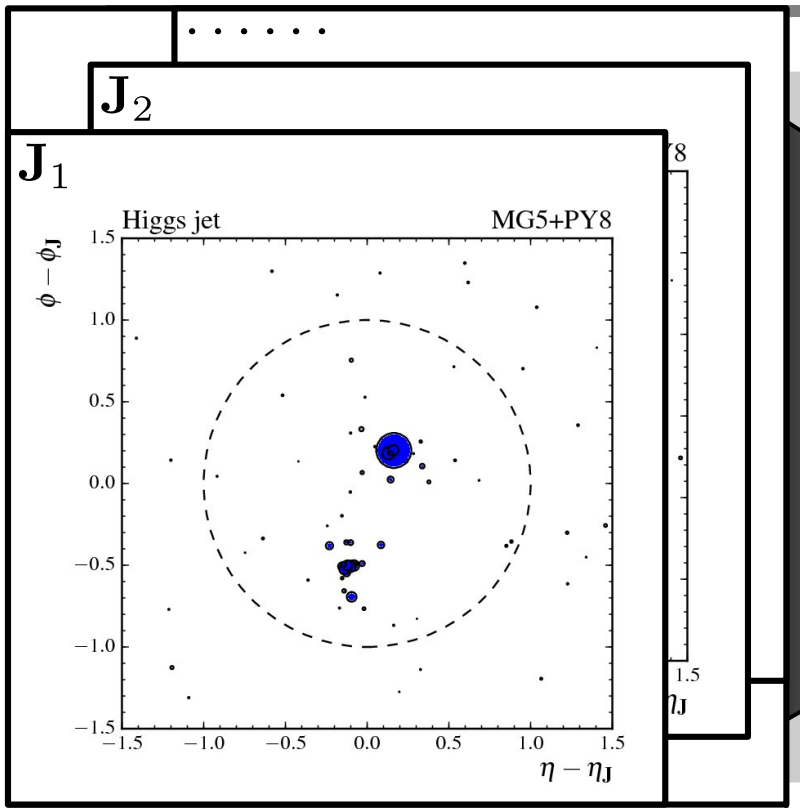
Neural network is often considered as a **black box** because studying its internal information barely gives you an insight about the decision...

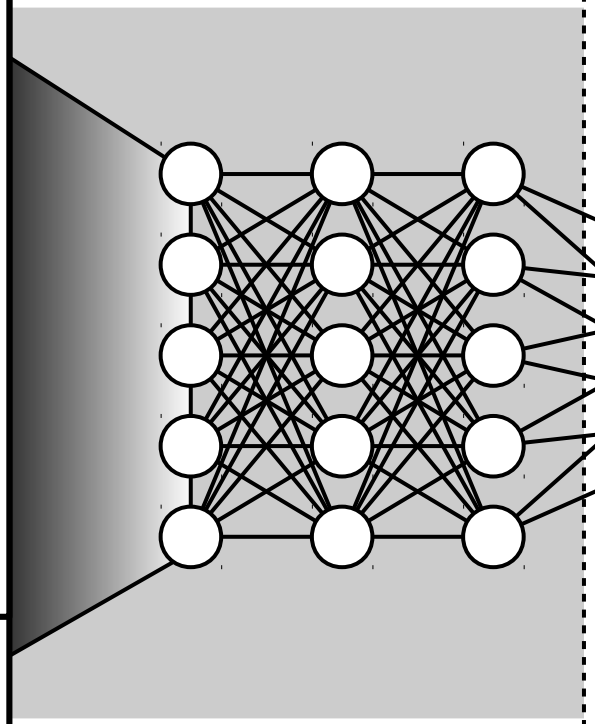
Difficulties on understanding the results from neural network



We also want to know **the reasoning** behind the decision!

Basic Structure of a Neural Network Classifier





$$\hat{y}_h$$

$$\hat{y}_{\mathrm{QCD}}$$

$$\Psi[P_T] = \sum_k w_k \phi_k[P_T]$$

$$\hat{y}_h = \varphi_{\mathrm{sigmoid}}(\Psi[P_T])$$

Logistic regression

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

Inputs: energy flow

Lots of inputs

$$\phi_k = \Phi_k[P_T]$$

Feature map

Too many parameters

Talk of Gregor and David for quick review

- N-subjettiness, ECFs…
- MLP, CNN, ParticleNet…
- EFPs, EFN, JUNIPR, …

The decision boundary is highly nonlinear.
It is hard to **interpret** the NN itself...

Functional Taylor Expansion

See also **energy flow polynomials**

Let us consider the "functional Taylor expansion" of the classifier.

$$\Phi[P_{T,a}] = w^{(0)} + \int d\vec{R}\; P_{T,a}(\vec{R}) w_a^{(1)}(\vec{R}) + \boxed{\frac{1}{2!}\int d\vec{R}_1 d\vec{R}_2\; P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_2) w_{ab}^{(2)}(\vec{R}_1, \vec{R}_2)} + \cdots$$

If we use only relative distance between constituents, the first nontrivial term is

$$w^{(0)} + p_{T,\mathbf{J}_a}\, w_a^{(1)}$$

$$\Phi[P_{T,a}] = \int dR\; S_{2,ab}(R) w_{ab}^{(2)}(R) + \cdots$$

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

Two-point correlation between constituents at distance R

Reduce the dimension of inputs
[Length/bin width]^2 → [Length/bin width]

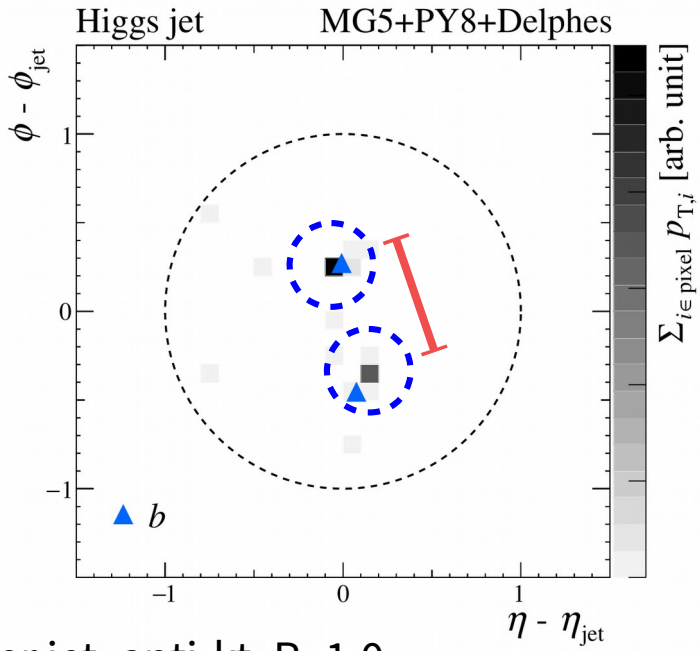
Two-Point Correlation Spectrum: Trimmed Spectrum

First, let us focus on correlation between hard constituents.
We may consider the two-point correlation spectrum of **trimmed jet**.

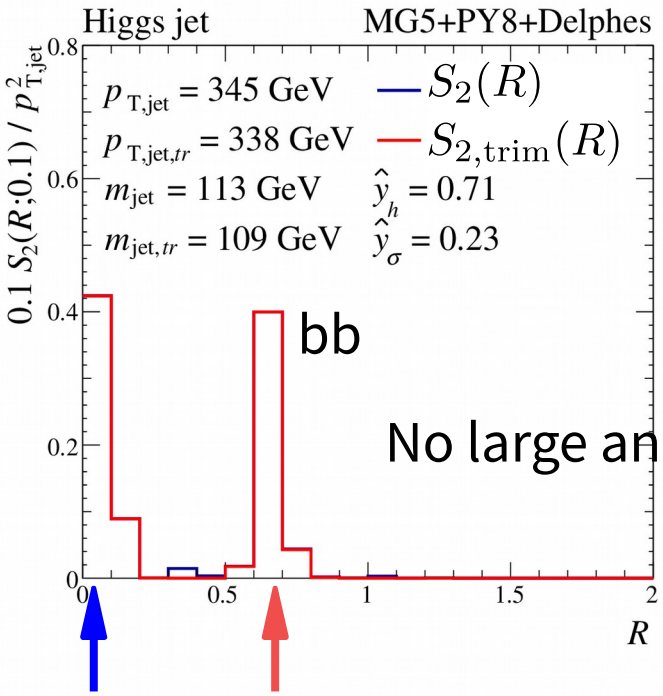
$$S_{2,\mathrm{trim}}(R) = \int d\vec{R}_1\, d\vec{R}_2\; P_{T,\mathbf{J}_{\mathrm{trim}}}(\vec{R}_1) P_{T,\mathbf{J}_{\mathrm{trim}}}(\vec{R}_2) \delta(R - R_{12})$$

only sensitive to **hard-hard correlations**

For Higgs jet:



Calorimeter jet, anti-kt, R=1.0



No large angle radiation

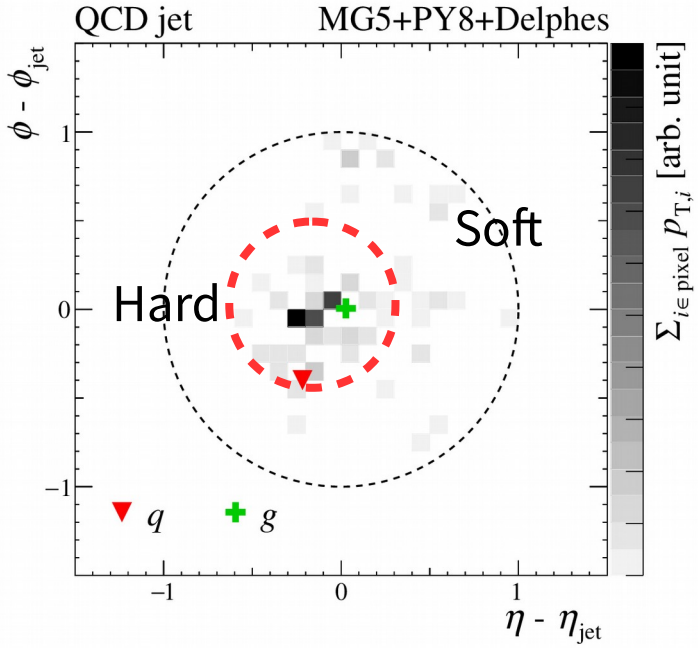
Two-Point Correlation Spectrum: Hard-Soft Correlation

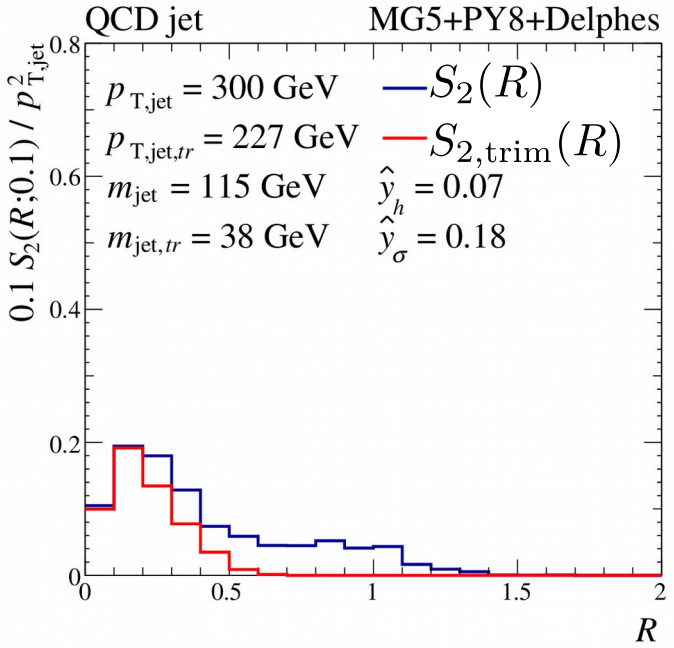
QCD jets have significant soft radiations. We may consider correlation between the soft parts and the hard parts.

$$S_{2,\mathrm{soft}}(R) = S_2(R) - S_{2,\mathrm{trim}}(R)$$

sensitive to **hard-soft correlations**
subleading soft-soft correlations

For QCD jet:

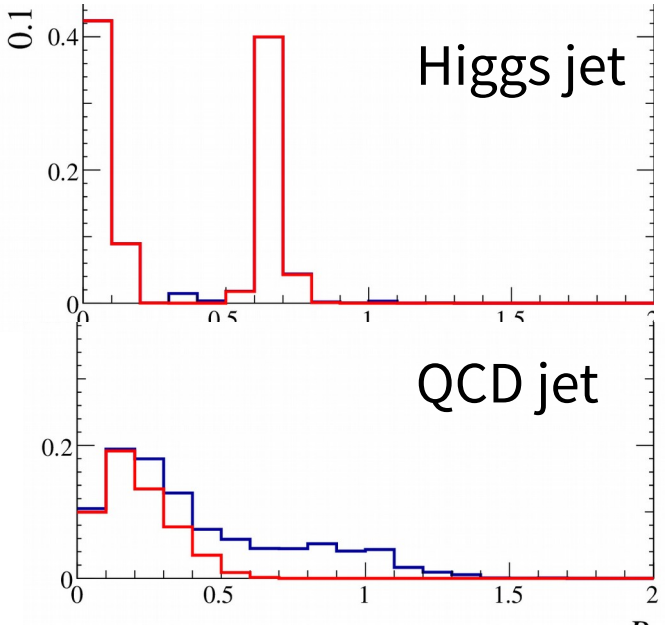


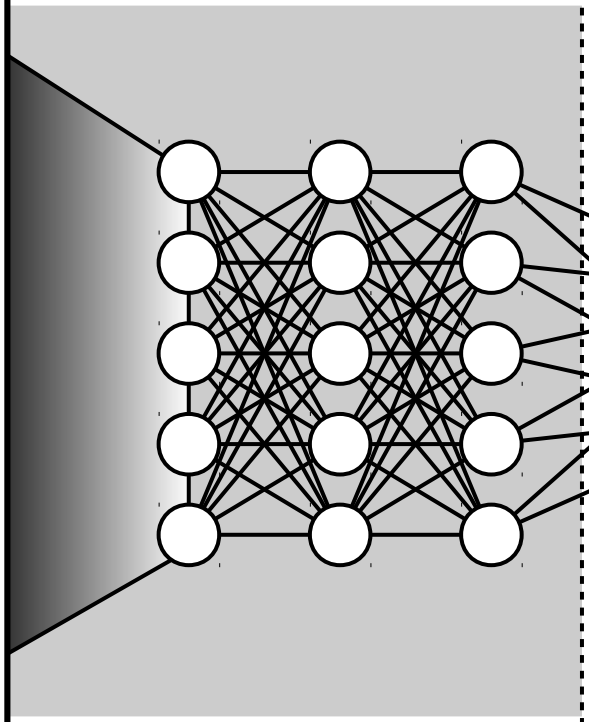


Calorimeter jet, anti-kt, R=1.0

Replacing CNN to MLP+S2

$S_{2,\mathrm{trim}}, S_{2,\mathrm{soft}}$





$\hat{y}_h$

$\hat{y}_{\mathrm{QCD}}$

$$\Psi[S_{2,ab}] = \sum_k w_k \phi_k[S_{2,ab}]$$

$$\hat{y}_h = \varphi_{\mathrm{sigmoid}}(\Psi[S_{2,ab}])$$

Inputs: two-point corr. Spectrum (+ jet mass, PT)

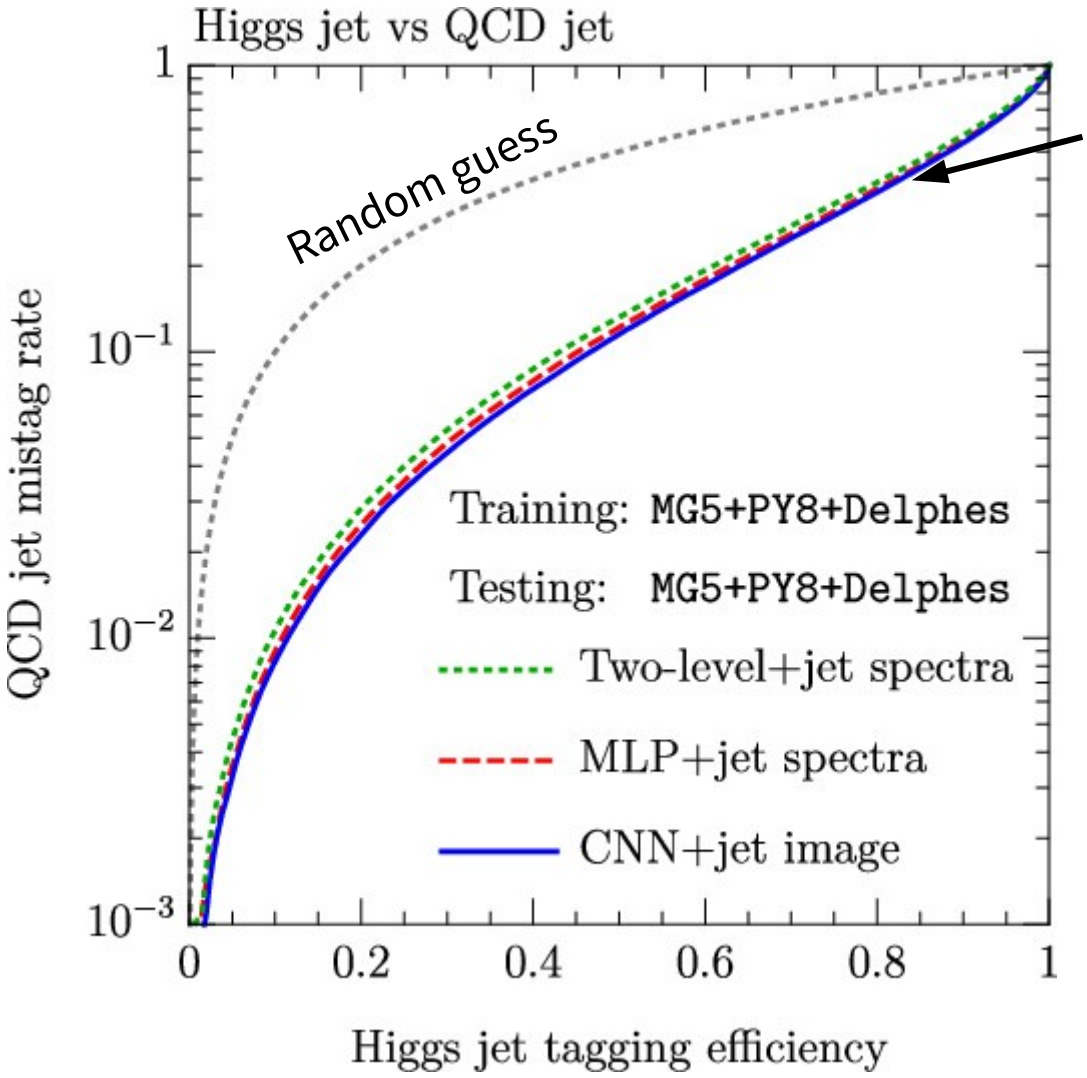
$$\phi_k = \Phi_k[S_{2,ab}]$$

Feature map: MLP

Logistic regression

This architecture represents: $\sum_{n=0}^{\infty} \mathcal{O}\left[P_T^{2n}\right]$

Equal Performance between CNN and MLP



Equal performance!!

$$\text{CNN} \sim \sum_{n=0}^{\infty} \mathcal{O}\left[P_T^n\right]$$

$$\text{MLP+S2} \sim \sum_{n=0}^{\infty} \mathcal{O}\left[P_T^{2n}\right]$$

For Higgs jet vs. QCD jet classification, **MLP+S2** is **sufficient.**

Interpretable Setup:

$$\sum_{n=0}^{\infty} \mathcal{O}\left[P_T^{2n}\right] \to \mathcal{O}\left[P_T^2\right] + \cdots$$

We may try the following two-level setup

Level1: substructure analyzer

$$\Phi[P_T; x_{\rm kin}] = \int dR\, S_{2,ab}(R) w_{ab}^{(2)}(R; x_{\rm kin})$$

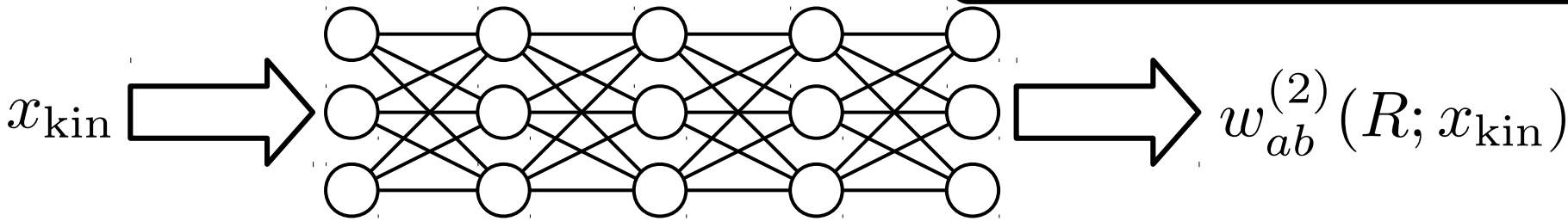
: IRC safe

$$x_{\rm kin} = \{p_{T,\rm jet}, m_{\rm jet}\}$$

$$\hat{R}_{b\bar{b}} = \frac{2m_h}{p_{T,h}}$$

Level2: kinematics analyzer

Use neural network to approximate the weight function.



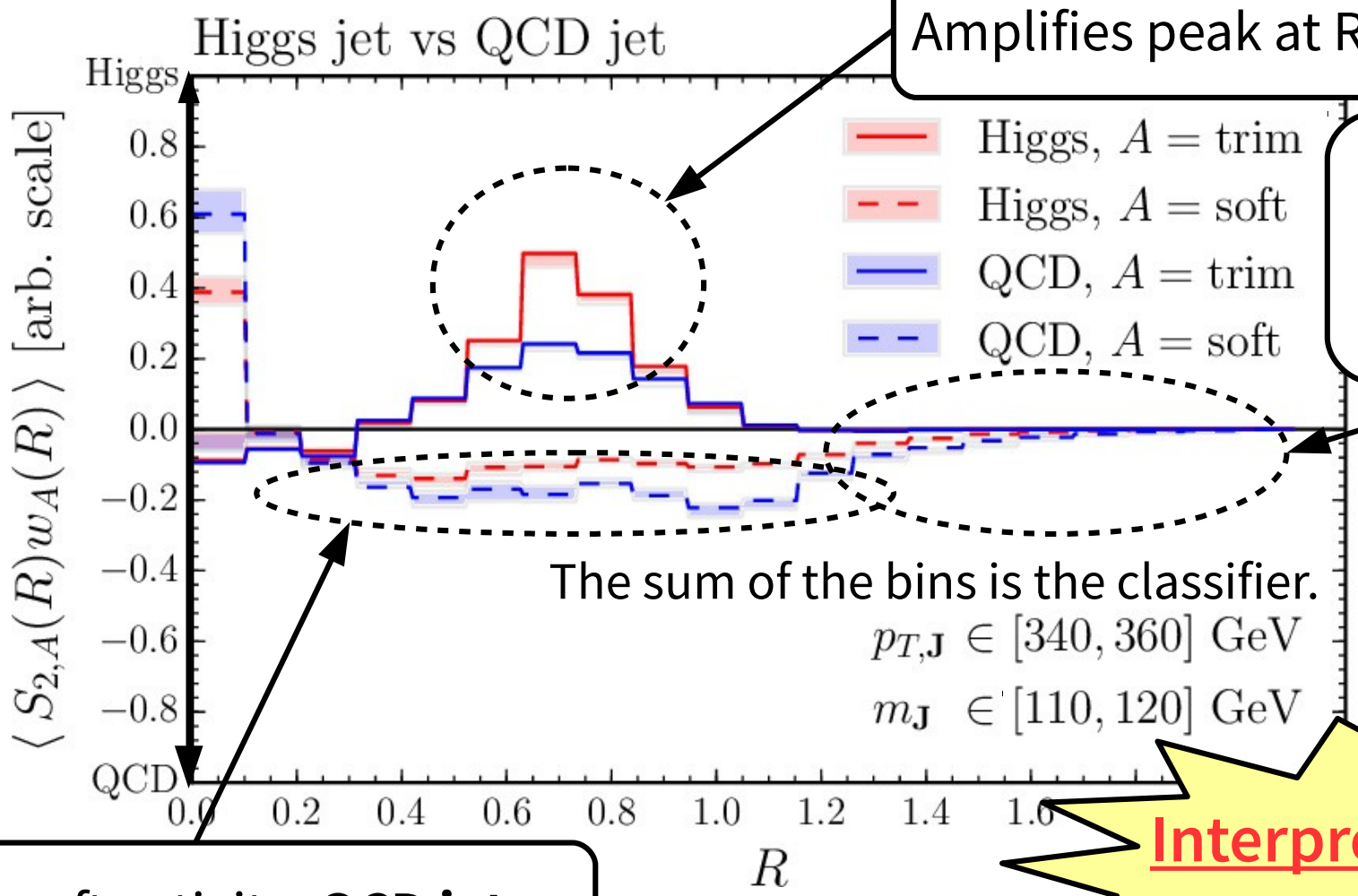
This architecture gives you two interpretable quantities:

$w_{ab}^{(2)}(R; x_{\rm kin})$ shows the **functional form** of the energy correlator.

$S_{2,ab}(R) w_{ab}^{(2)}(R; x_{\rm kin})$ shows the **contribution** to the classifier.

Average of the linear classifier outputs

$$\Phi[S_{2,ab}] = \int dR\, S_{2,\mathrm{trim}}(R) w^{(2)}_{\mathrm{trim}}(R) + \int dR\, S_{2,\mathrm{soft}}(R) w^{(2)}_{\mathrm{soft}}(R)$$



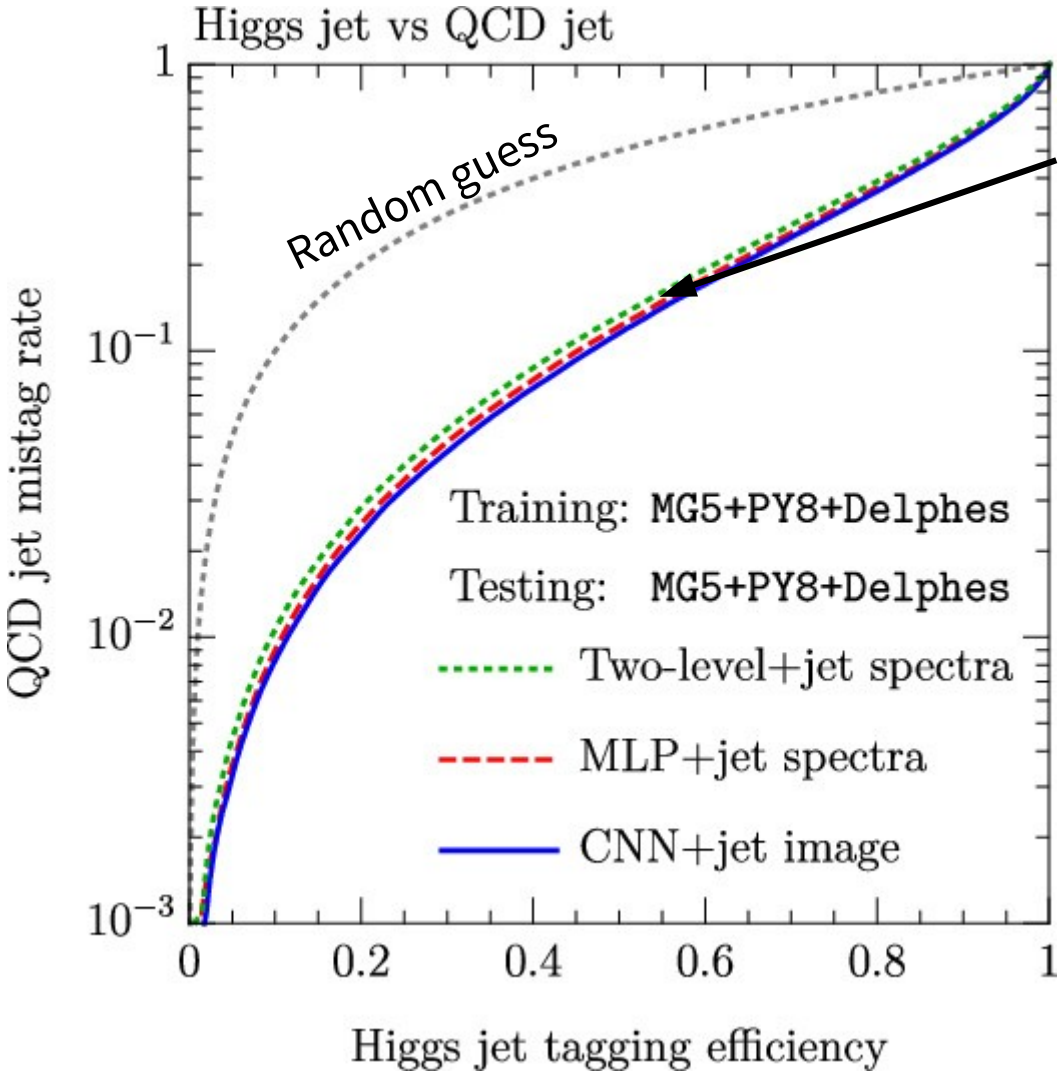
Amplifies peak at Rbb: **Higgs jet**

Large angle radiation: **QCD jet**

More soft activity: **QCD jet**

Interpretable

Similar Performance between CNN and the interpretable network



Still similar performance!

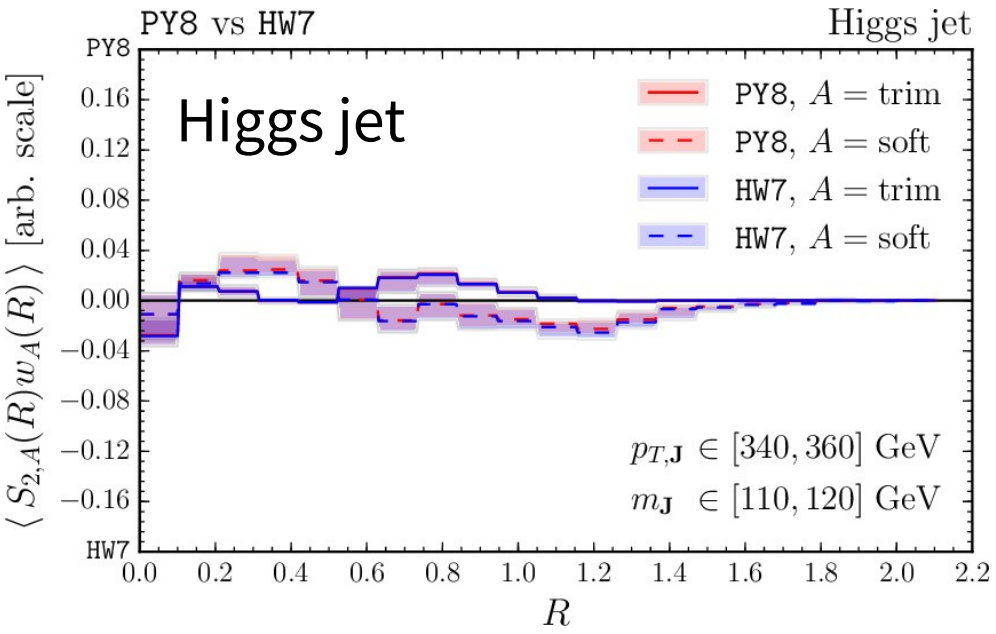
$$\text{MLP+S2} \sim \sum_{n=0}^{\infty} \mathcal{O}\left[P_T^{2n}\right]$$

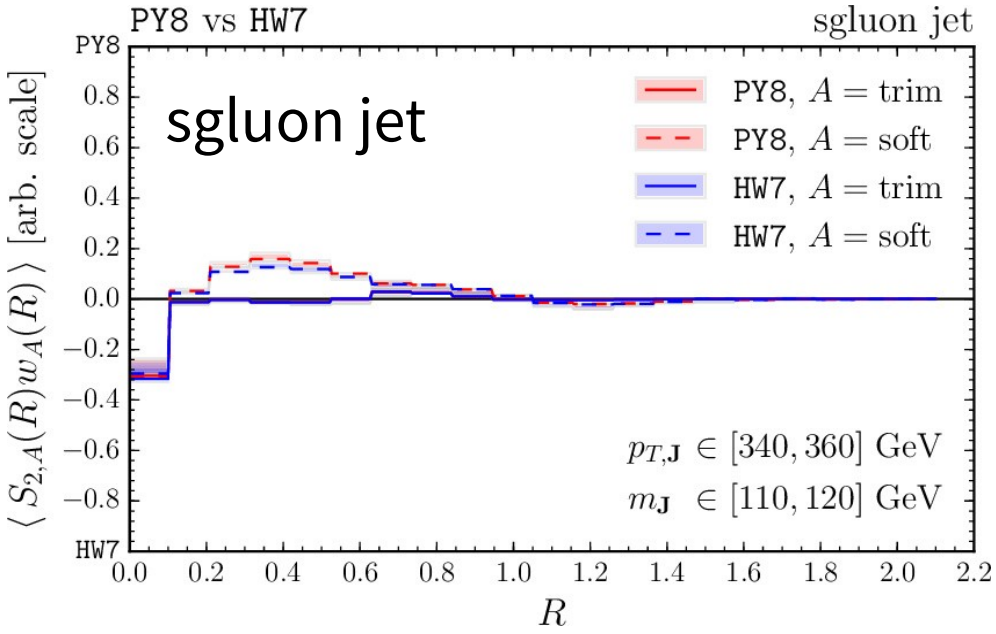
$$\text{Two-level+S2} \sim \mathcal{O}\left[P_T^2 \cdot w(p_{T,\mathbf{J}}, m_{\mathbf{J}})\right]$$

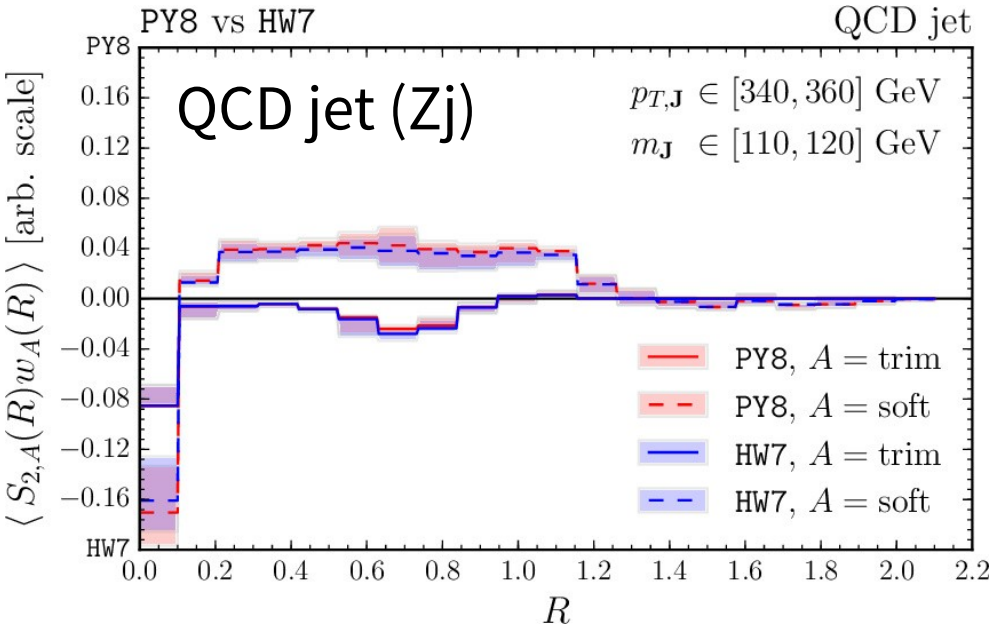
For Higgs jet vs. QCD jet classification, **Two-level+S2** is **sufficient.**

See importance of calibration on soft activity:
Petar's talk

Comparing Pythia and Herwig







Most of the classifier contribution comes from soft activity

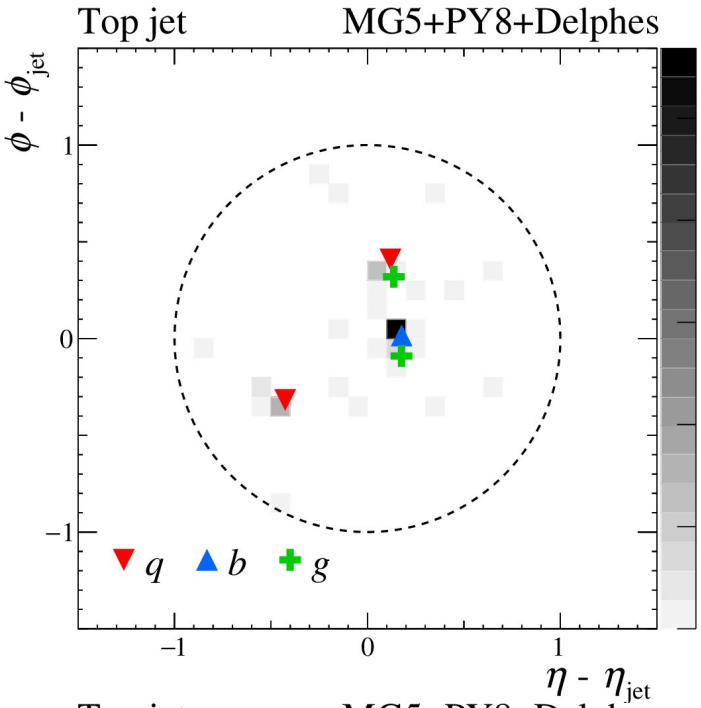
PY8 jets are more compact

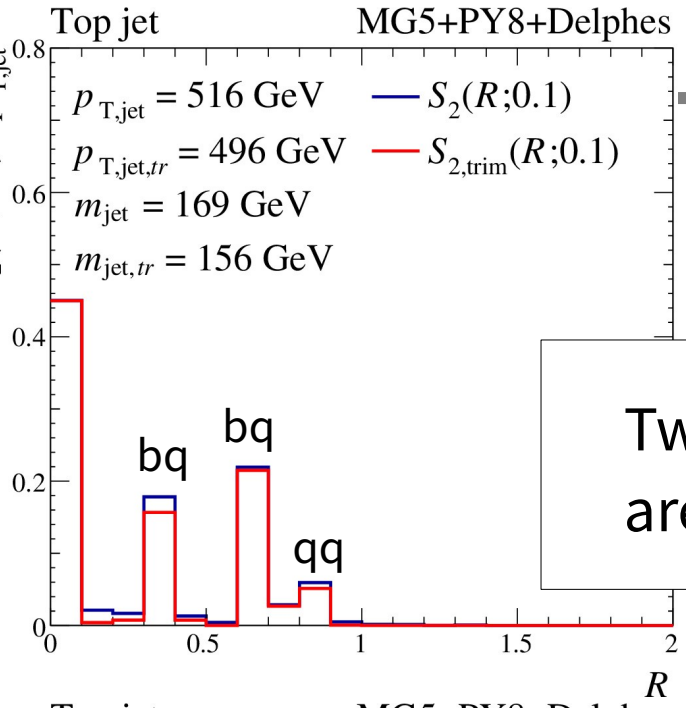
Singlet and octet have different behavior

Top Jets (work in progress...)

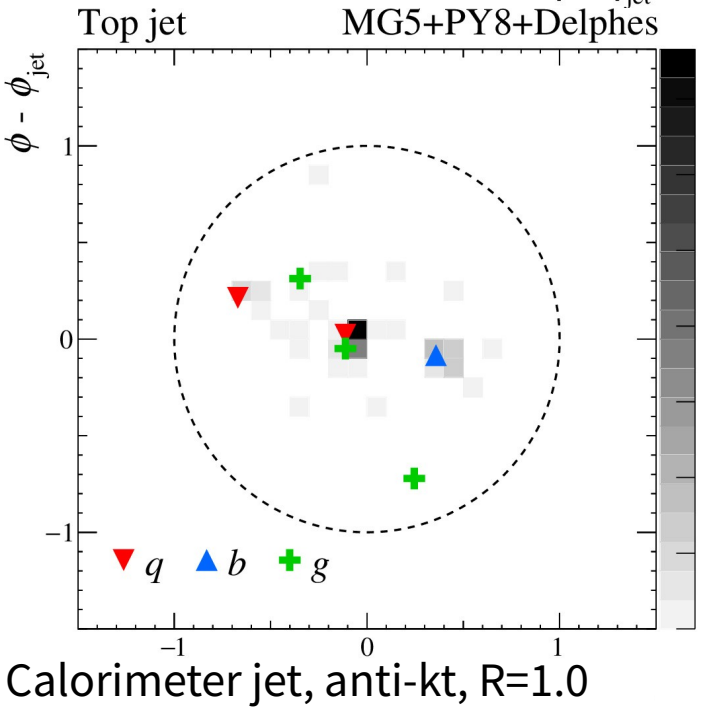
KEK

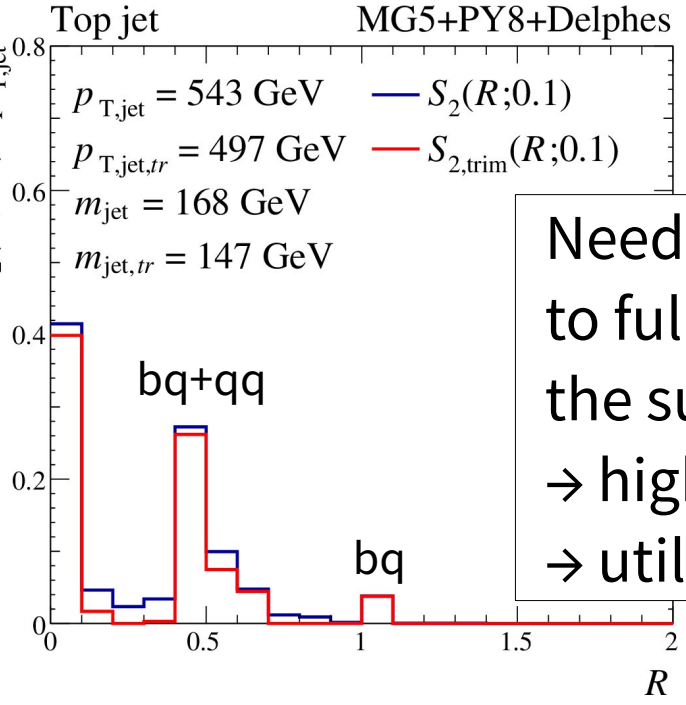
Preliminary





Two-point correlations are enough.





Need more information to fully encode the substructures.
→ higher order correlation?
→ utilize subjet information?

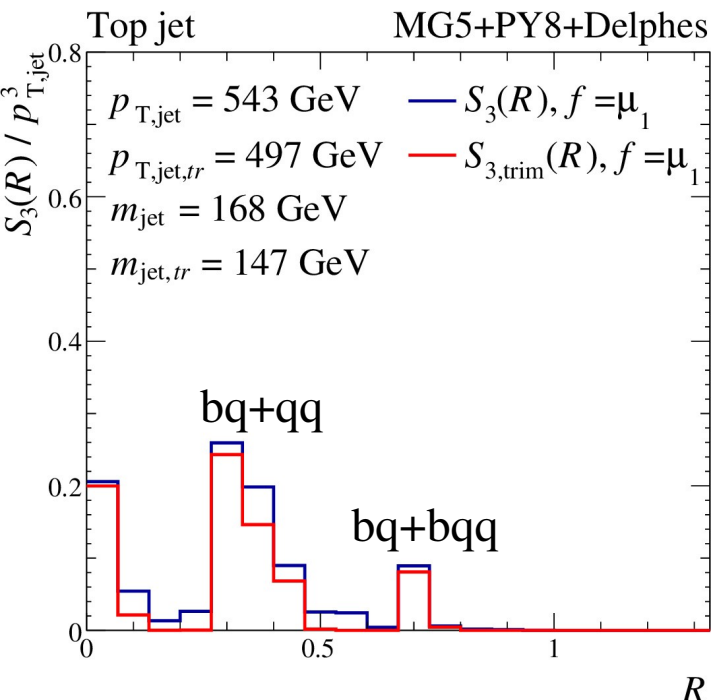
Calorimeter jet, anti-kt, R=1.0

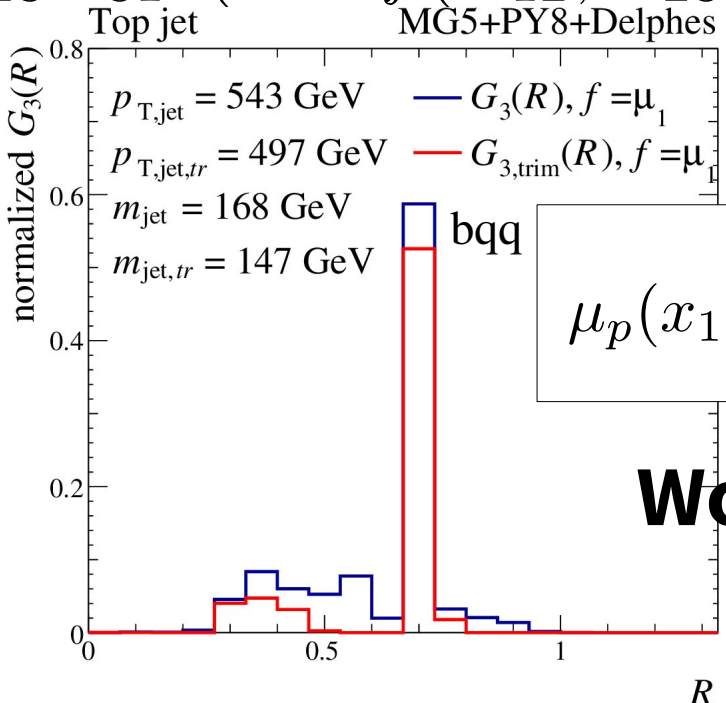
Three-point correlation spectrum

Preliminary

$$S_{3,abc}(R) = \int d\vec{R}_1 d\vec{R}_2 d\vec{R}_3 \; P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_2) P_{T,c}(\vec{R}_3) \times \delta(R - f(R_{12}, R_{23}, R_{31}))$$

$$G_{3,abc}(R) = \int d\vec{R}_1 d\vec{R}_2 d\vec{R}_3 \; P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_2) P_{T,c}(\vec{R}_3) \times R_{12} R_{23} R_{31} \delta(R - f(R_{12}, R_{23}, R_{31}))$$





$$\mu_p(x_1, \cdots, x_N) = \left(\frac{\sum_i x_i^p}{N} \right)^{\frac{1}{p}}$$

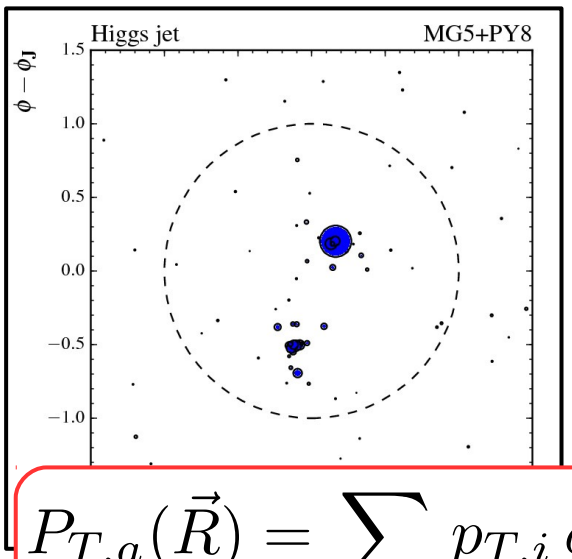
Work in progress...

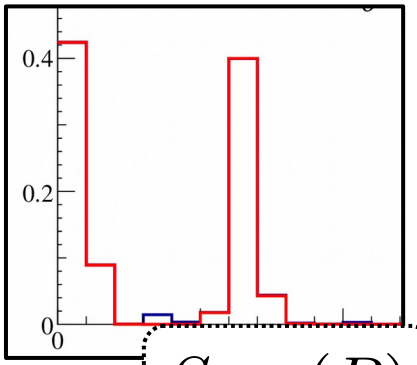
Summary

- For the next run of LHC and future colliders, we need a quick and reliable jet substructure analysis framework.
- We developed a machine learning framework using **two-point correlation spectrum** for analyzing jet substructures.
- The spectrum is derived from the jet image analysis and the corresponding two-level model is **interpretable**.
- This analysis strategy is not limited to Higgs jet vs. QCD jet classification, but we may use it for comparing Monte-Carlo simulations (and real data).
- Analyses with more complex objects (top jets…) are ongoing.

Please stay tuned!

BACKUP





$$S_{2,\mathrm{trim}} = S_{2,hh}$$

$$S_{2,\mathrm{soft}} = S_{2,hs} + S_{2,sh} + S_{2,ss}$$

$S_{2,ab}(R)$

discretization

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

$S^k_{2,ab}$

$$\int dR\; S_2(R) w(R; x_{\mathrm{kin}})$$

discretization

$$\sum_{k,a,b} S^k_{2,ab}\, w^k_{ab}(x_{\mathrm{kin}})$$

sigmoid

$\hat{y}_h,\ \hat{y}_{\mathrm{QCD}}$

x_{kin}

$\begin{pmatrix} p_{T,\mathbf{J}} \\ m_{\mathbf{J}} \end{pmatrix}$

MLP

Activation: ELU

$w^k_{ab}(x_{\mathrm{kin}})$

discretization

approximate

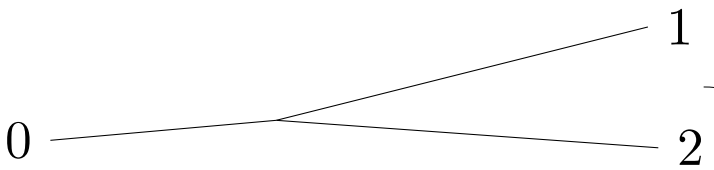
$w_{ab}(R; x_{\mathrm{kin}})$

cf.

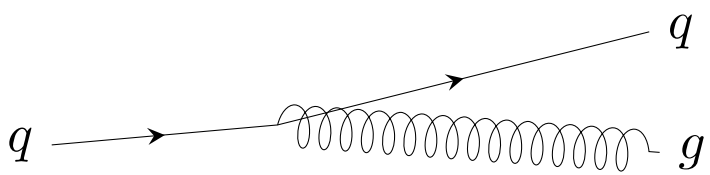
$$\hat{R}_{b\bar{b}} = \frac{2m_h}{p_{T,h}}$$

Why the linear classifier works?

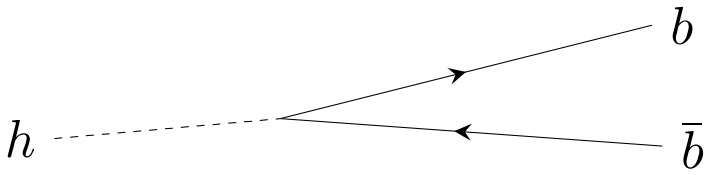
The parameter set $(p_{T,0}, z, R)$ determines the kinematics of parton evolution.



- parton shower



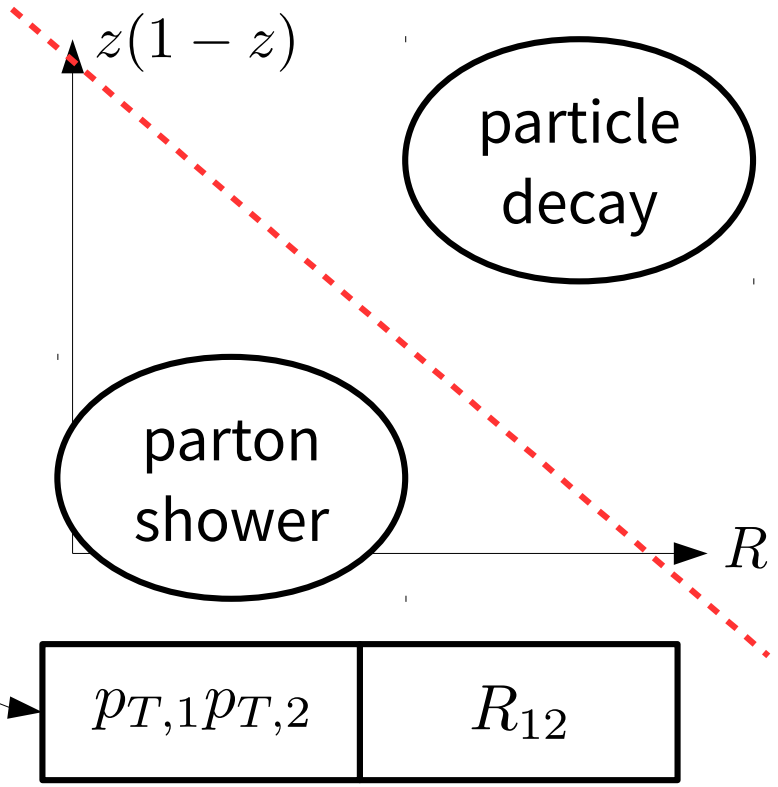
  - Collinear: small R
  - Soft: z ~ 0 or 1
- particle decay



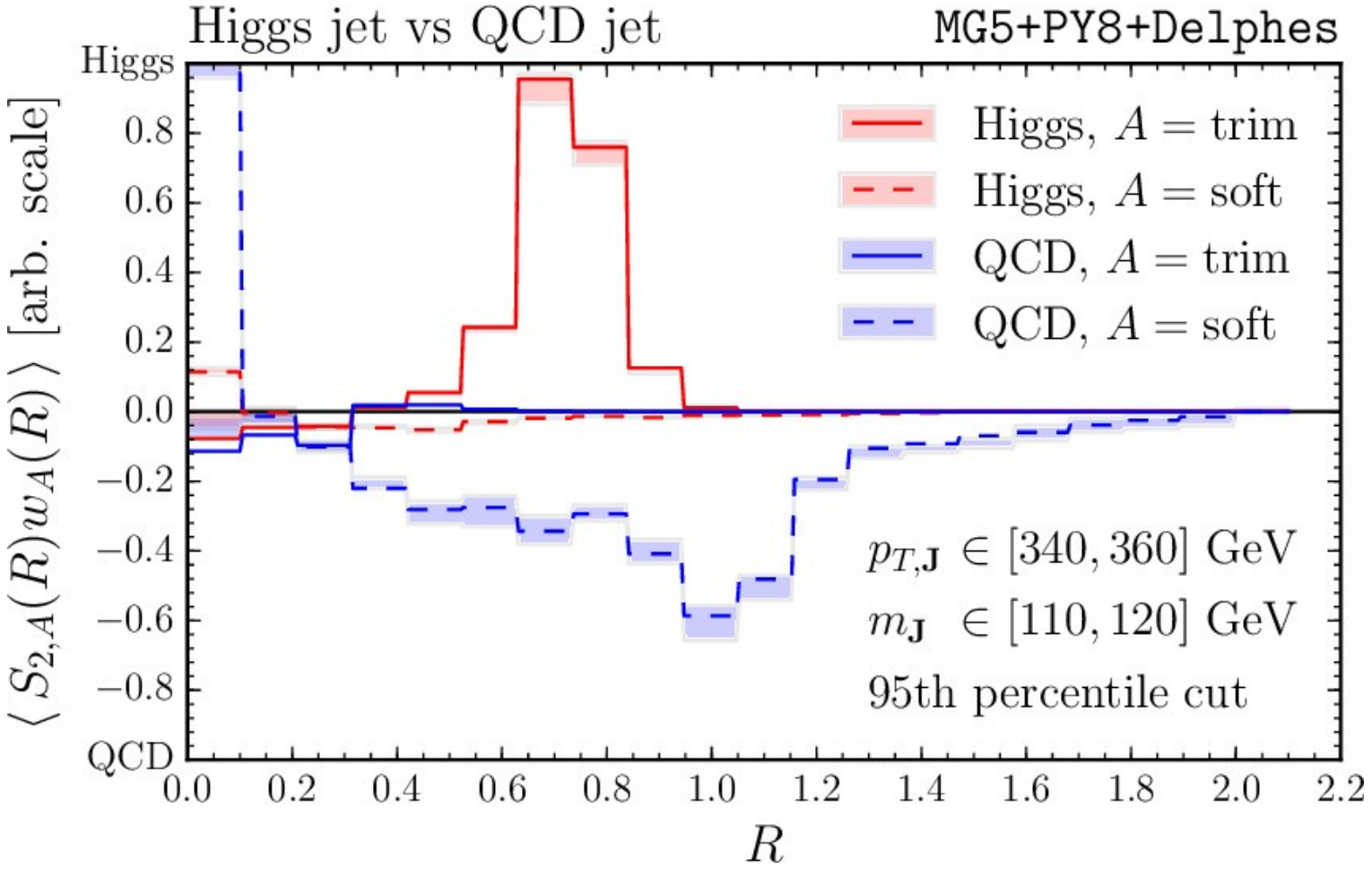
  - Transverse decay:
    R ~ 2m/pt and z ~ 0.5

$$p_{T,1} = p_{T,0} z \qquad p_{T,2} = p_{T,0}(1-z)$$

$$R^2 = (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2$$

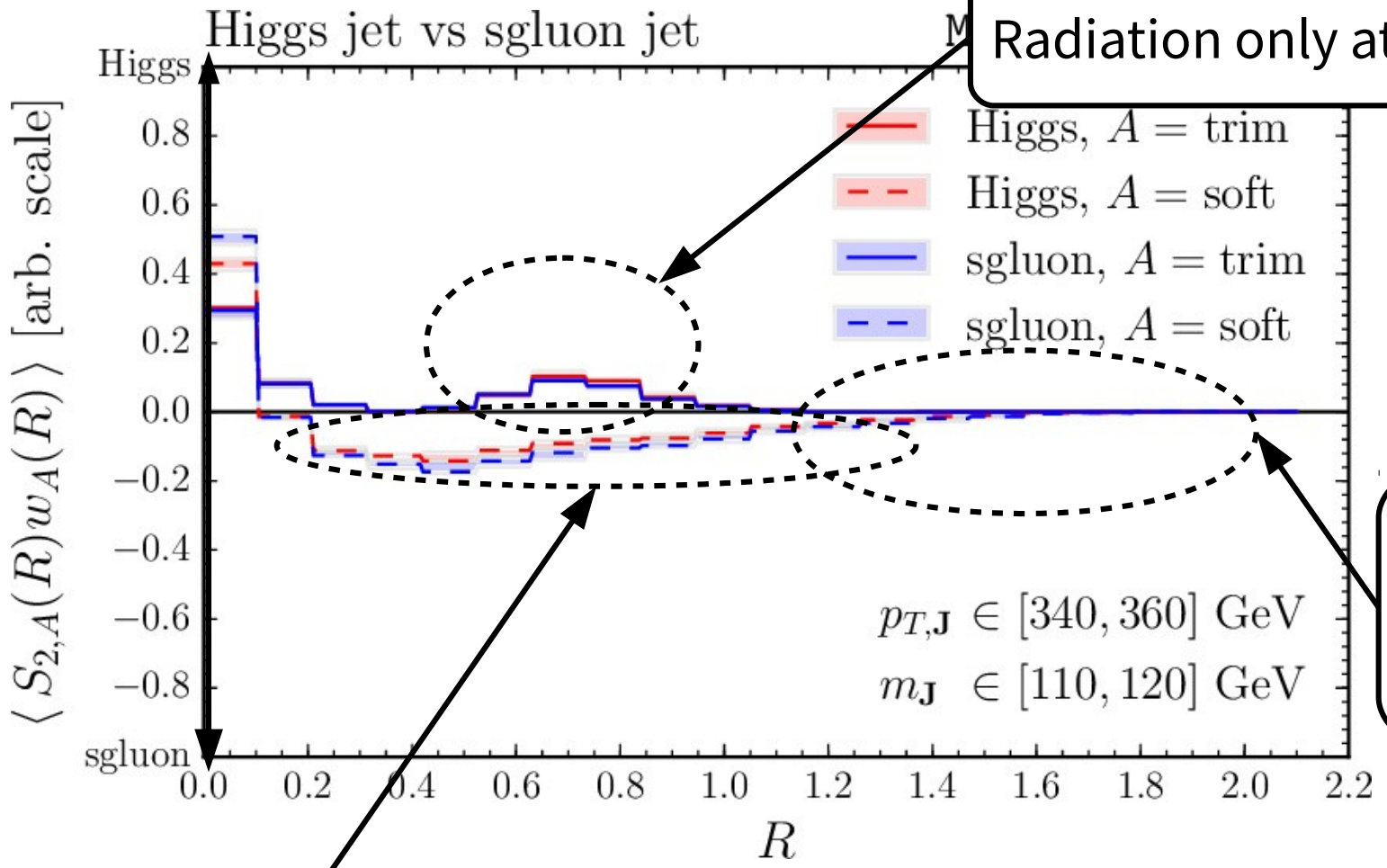


Average of the linear classifier outputs



Two-Prong jet: **1** vs **8**

Identification of color of originating parton: **1** vs **8**

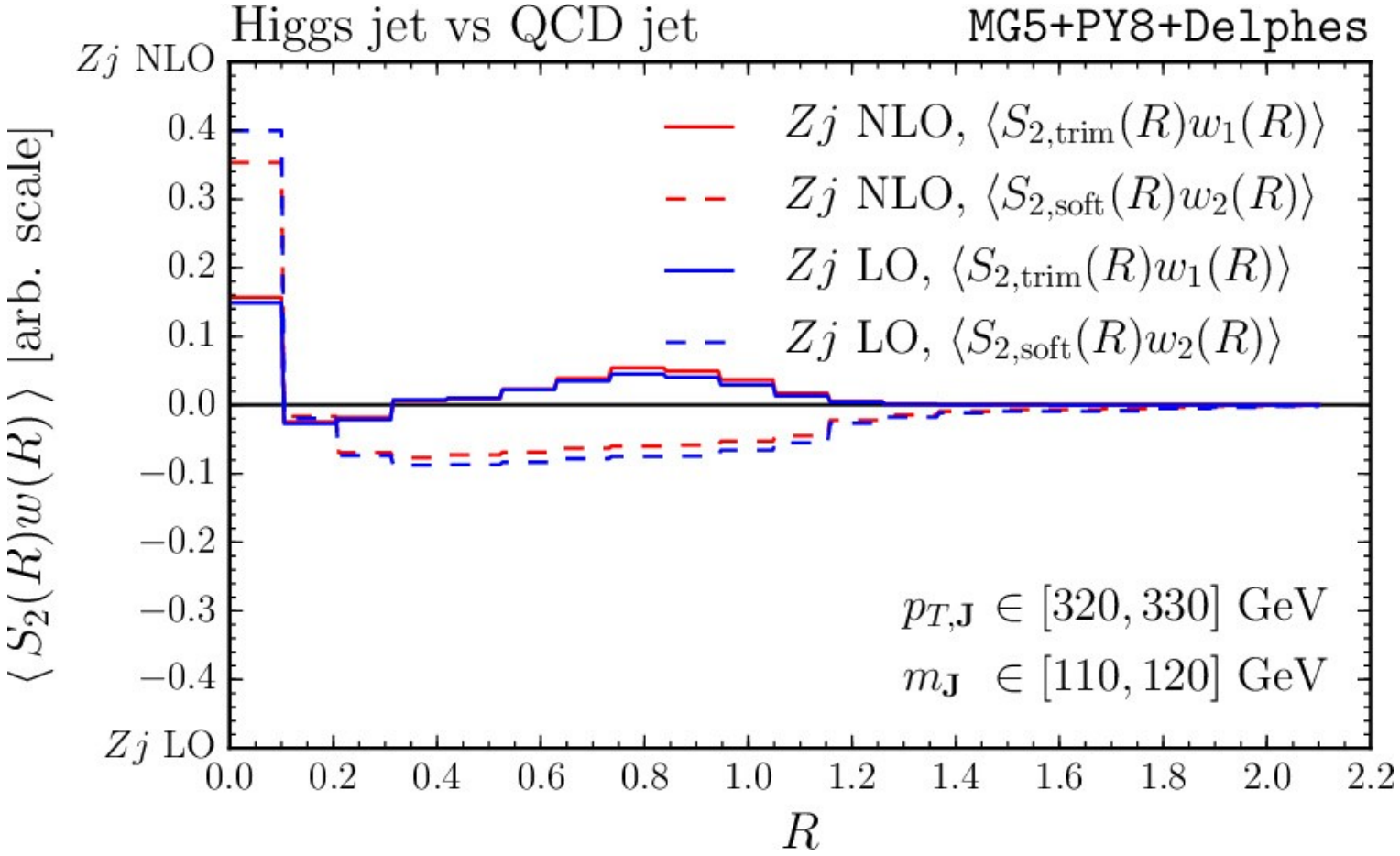


Radiation only at Rbb: **Higgs jet**

Large angle radiation: **sgluon jet**

More soft activity: **sgluon jet**

Training with MC@NLO events



We can check the sanity of training with negative weights.