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

Sung Hak Lim

Theory Center, KEK

The 1st AEI Workshop for BSM and
the 9th KIAS Workshop on Particle Physics and Cosmology,

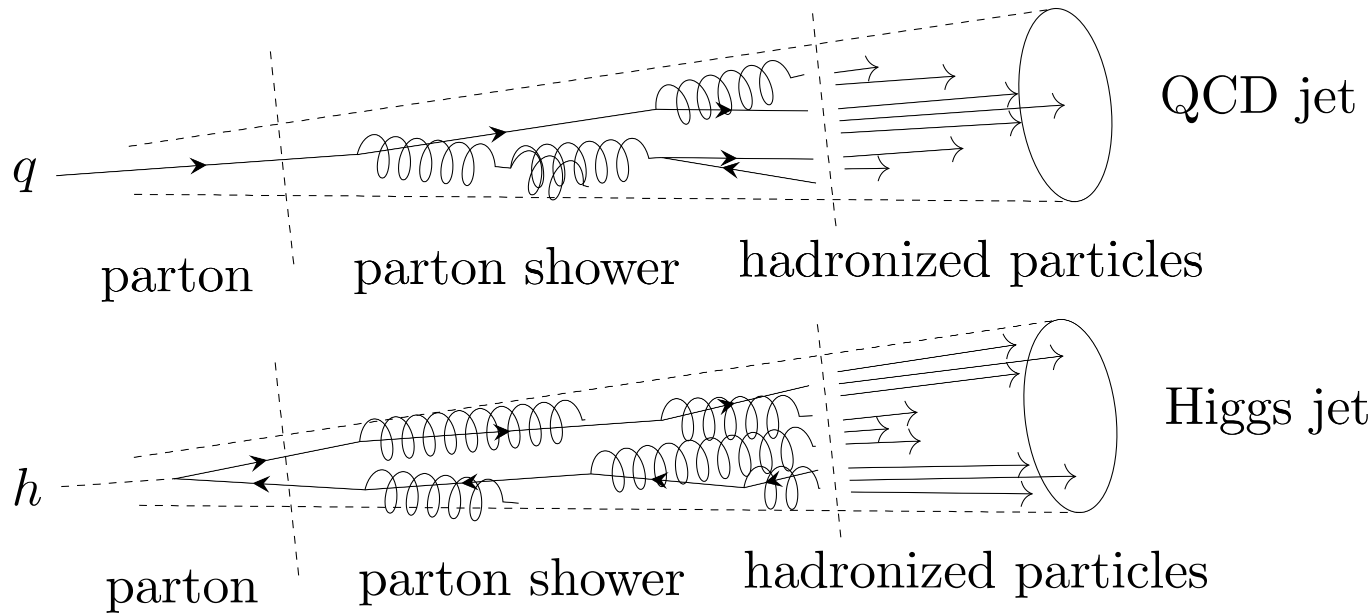
Jeju island, Korea

Nov. 2019

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, will appear in arXiv soon

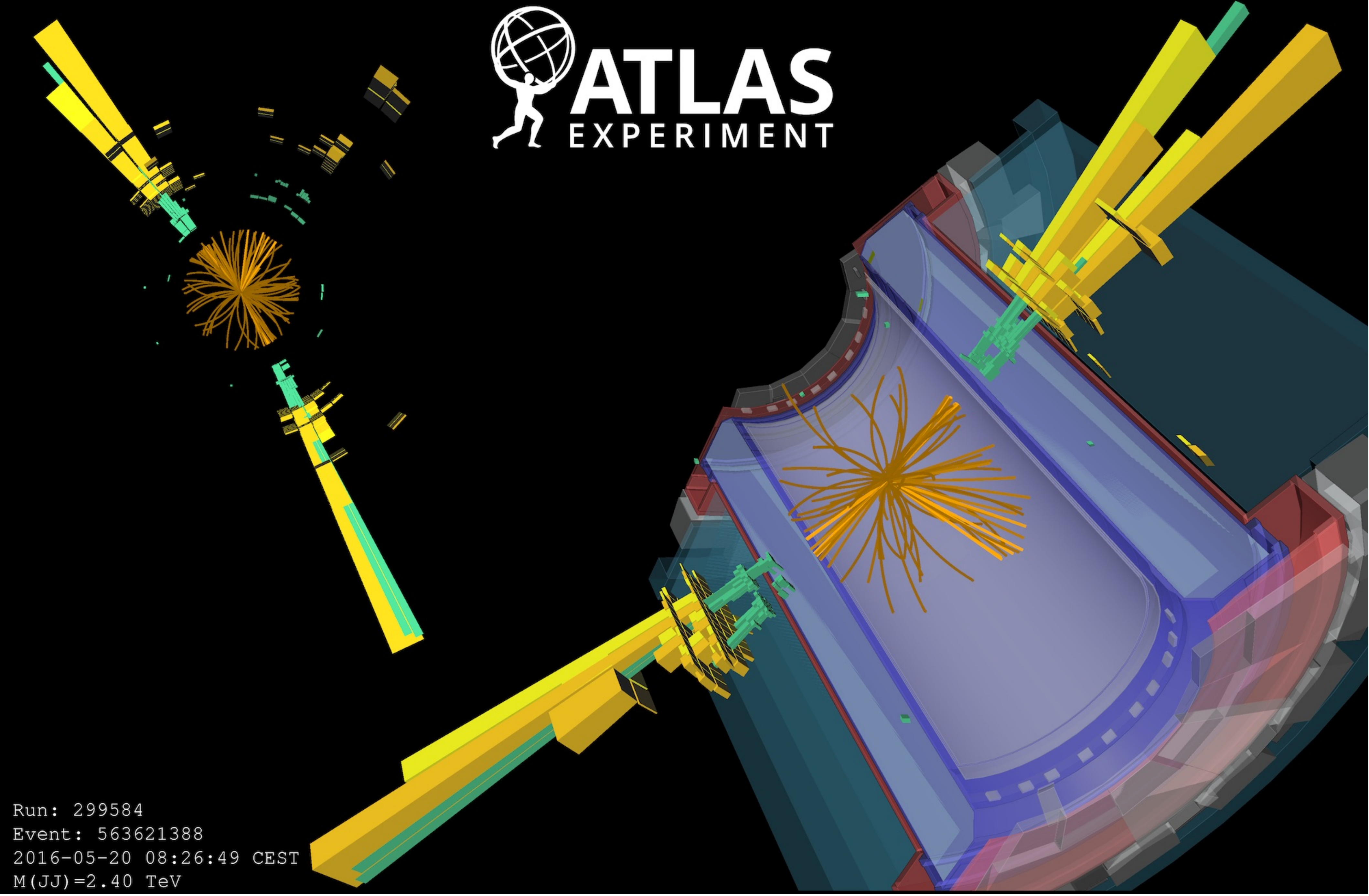
Boosted Jets:
Jets have substructure!

- As LHC stacking up multi TeV center-of-mass energy events, boosted heavy particles is produced and forms a single collimated cluster of particles similar to the QCD jets. ($m_{\mathrm{EW}}/\sqrt{\hat{s}} \approx \mathcal{O}[0.1]$)



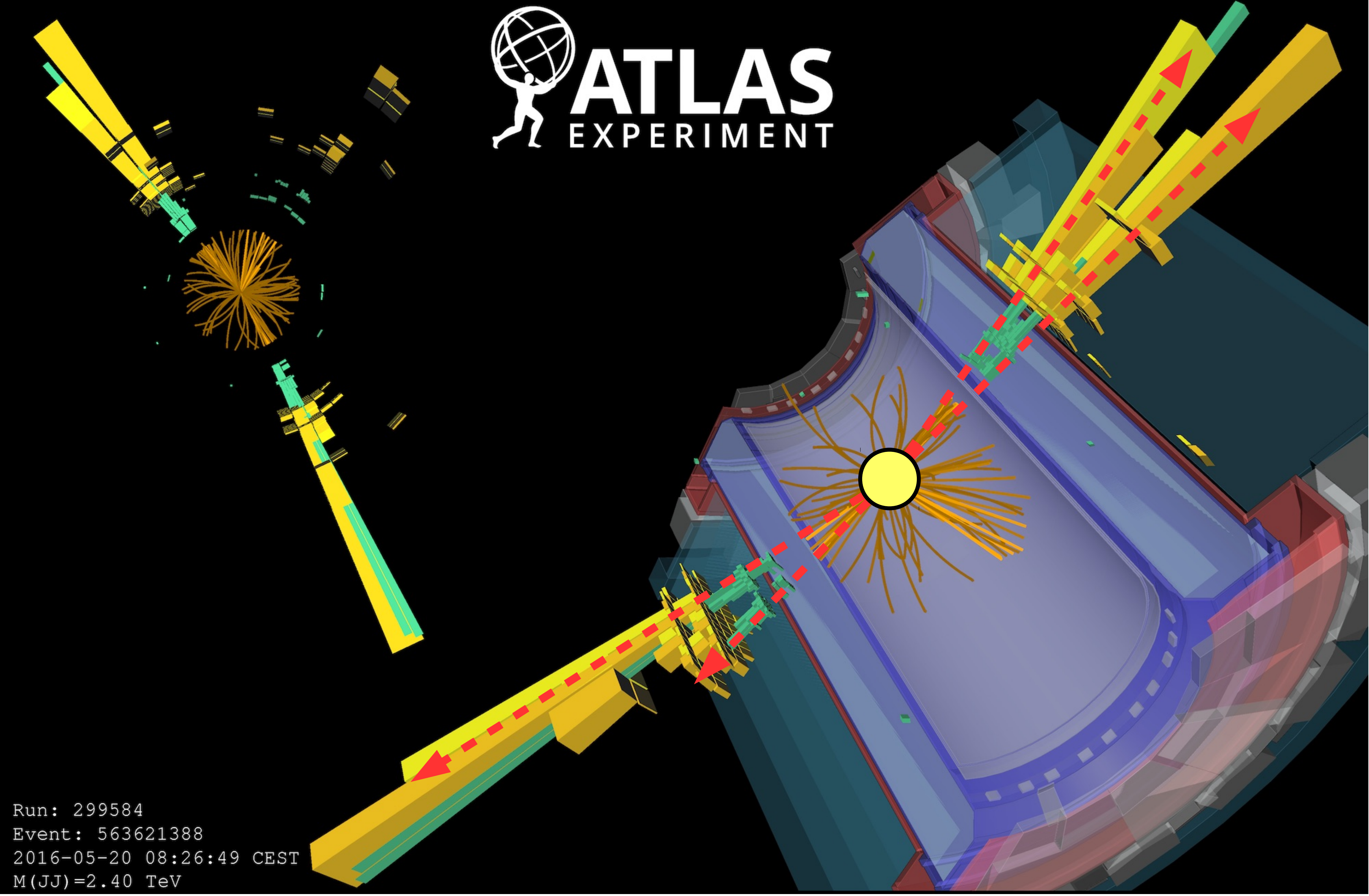
- We will see more and more these boosted jets!

LHC	**HE-LHC**	**FCC**
14 TeV →	27 TeV →	100 TeV



Two boosted jets from the old 2 TeV resonance searches...

https://atlas.cern/updates/physics-briefing/hunting-new-physics-boosted-bosons



We may require **deeper** understanding on these objects...

https://atlas.cern/updates/physics-briefing/hunting-new-physics-boosted-bosons

Deep Learning and Jet Physics

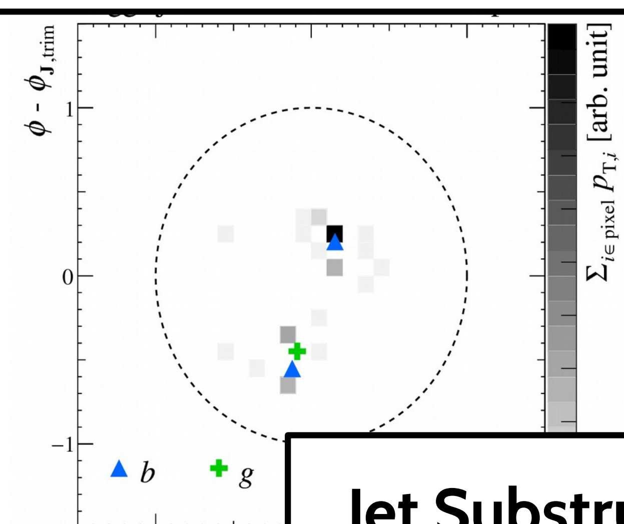
- We want a quick and reliable method for classifying those jets.
- Thanks to the development in physics and computer science…

Theoretical Understanding of QCD

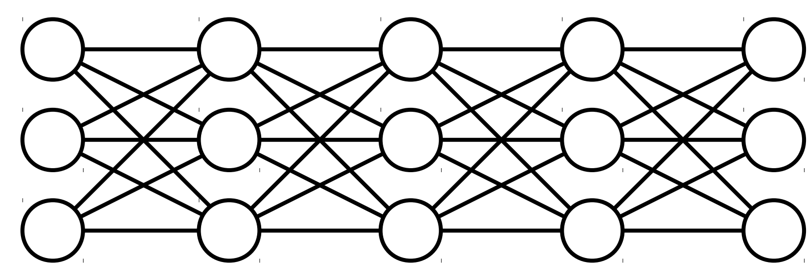
Development of Machine Learning

Huge Experimental Data

More Computational Power



Jet Substructure



Deep Learning

Classification Problem with Images

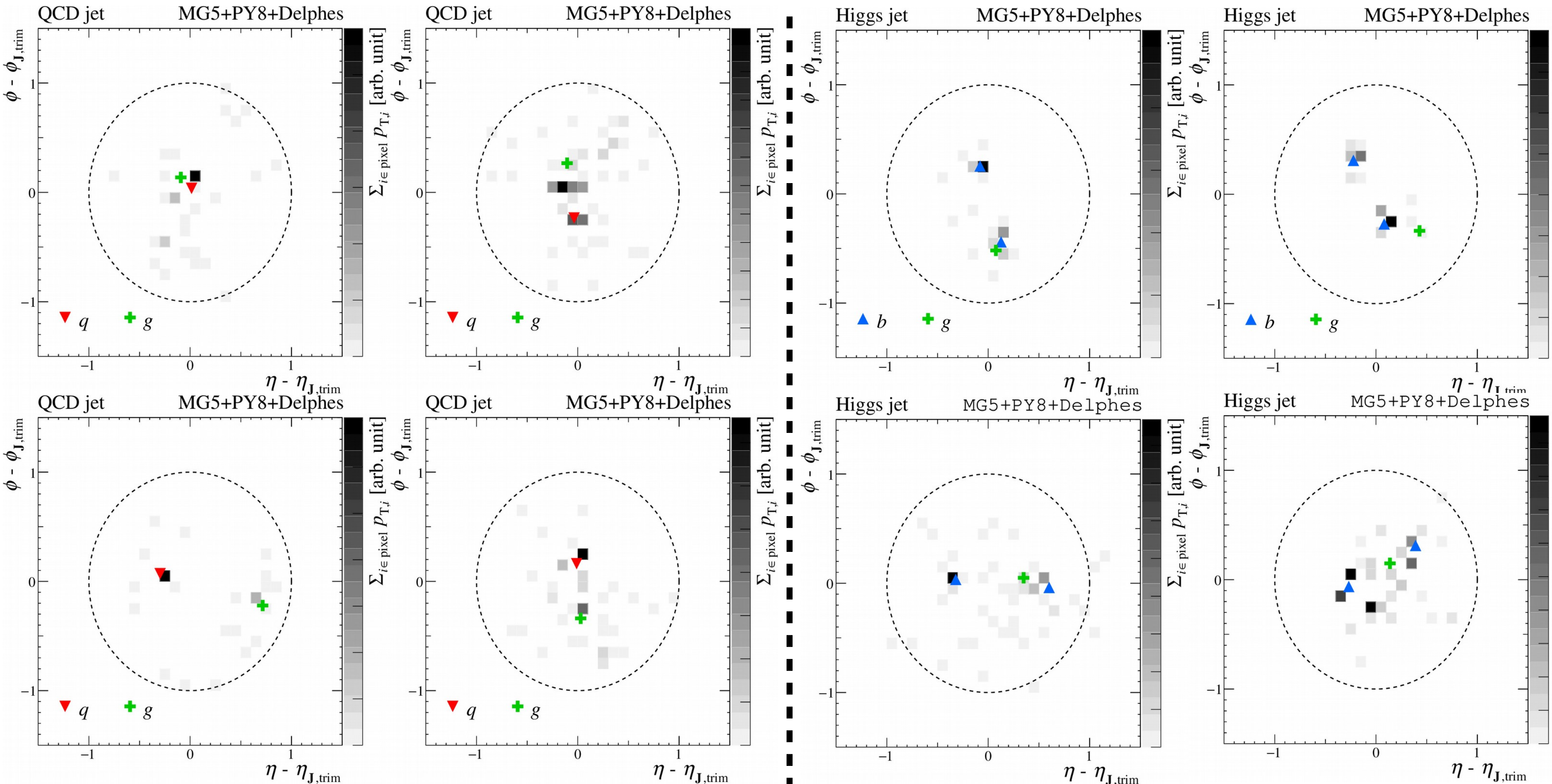
Can you distinguish cats and dogs from **pictures**?

Cats (?)

Dogs (?)

Classification Problem with Jet Images

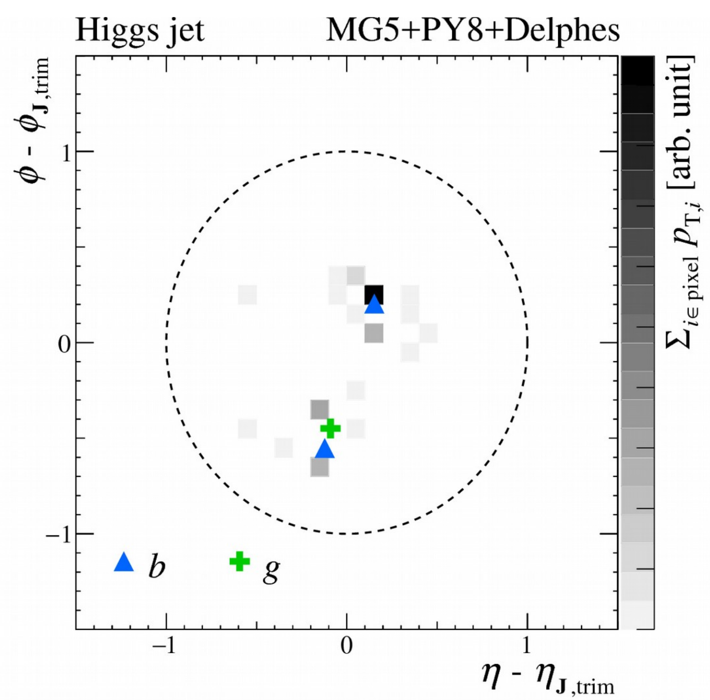
Can you distinguish <u>QCD jets</u> and <u>Higgs jets</u> from **reconstructed particles**?

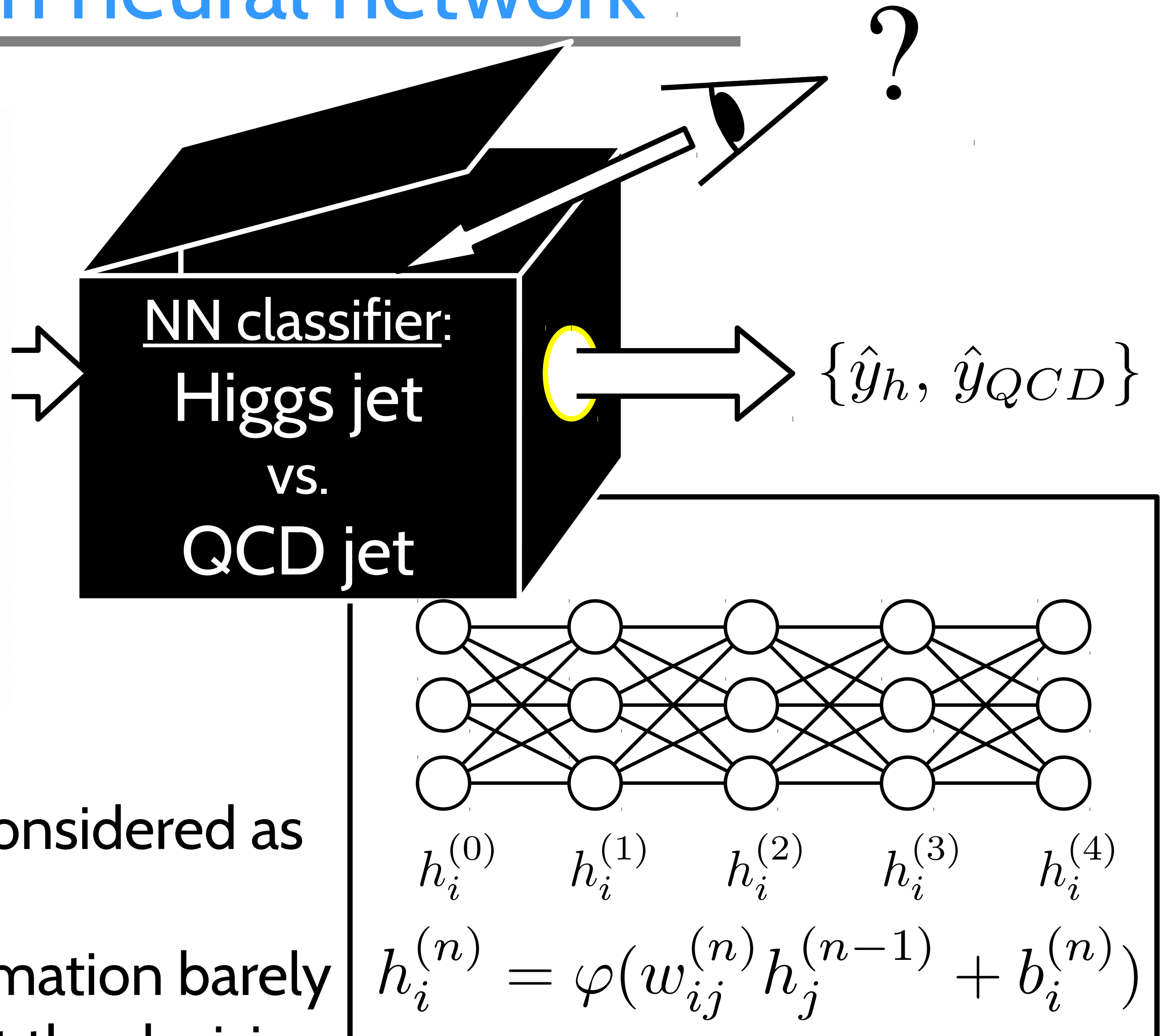


QCD jets

Higgs jets

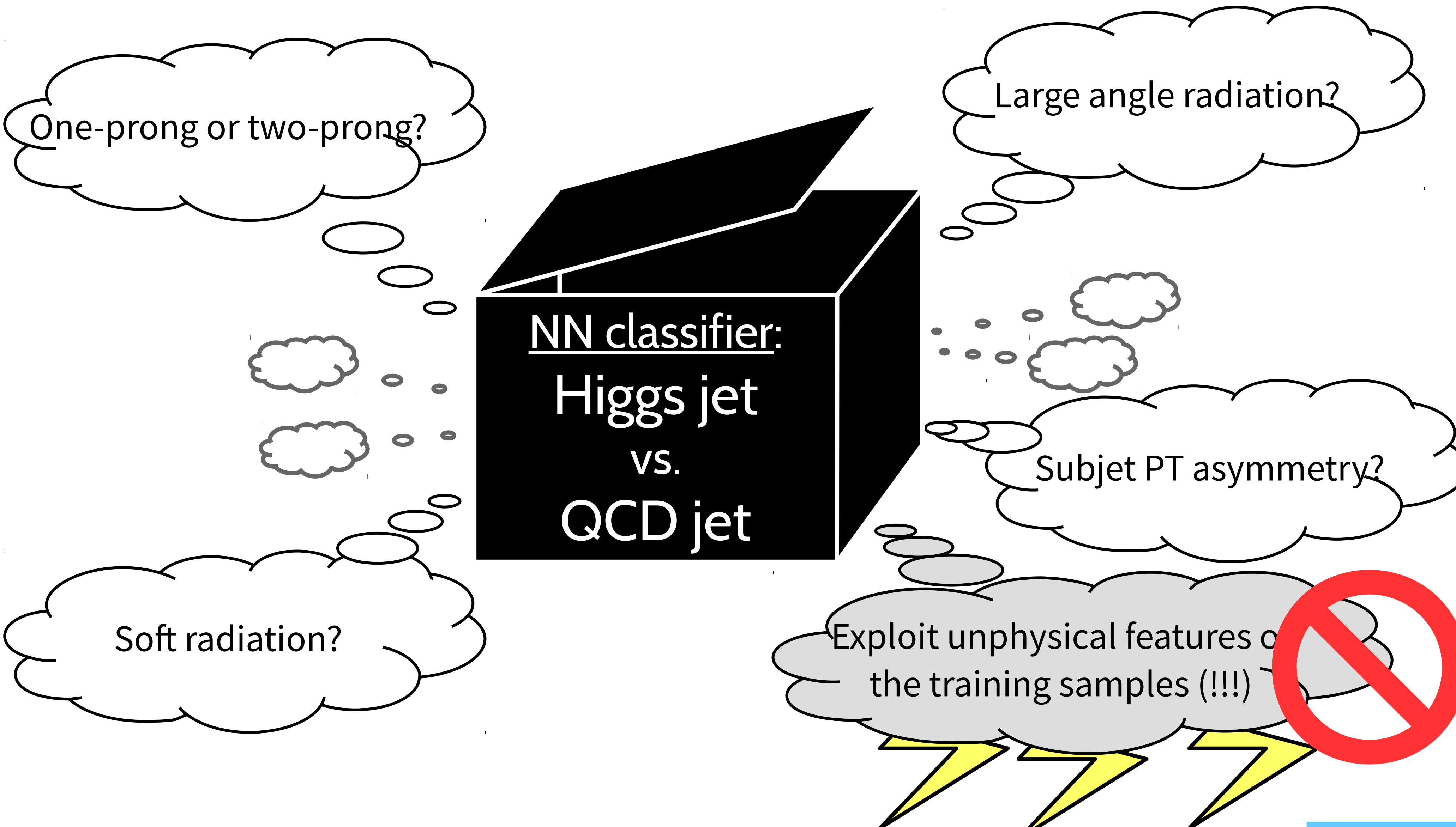
Difficulties on understanding the results from neural network





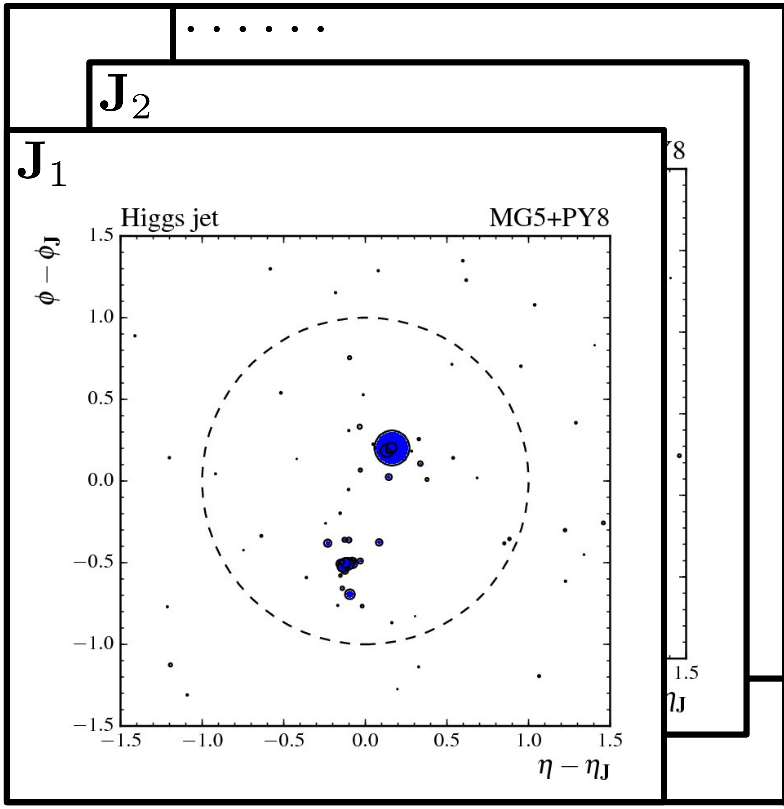
Neural network is often considered as a **black box** because studying its internal information barely gives you an insight about the decision...

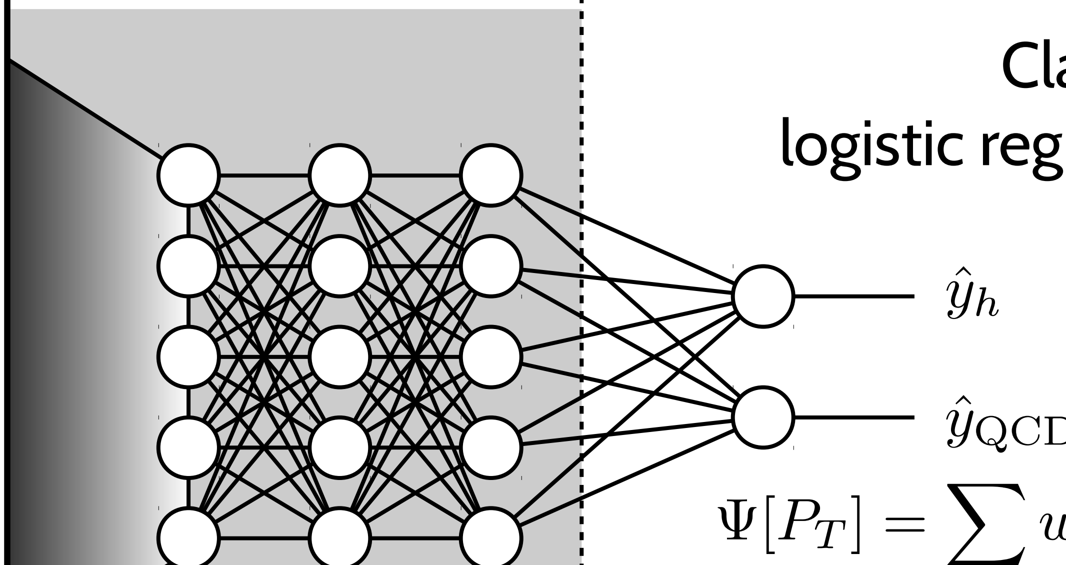
Difficulties on understanding the results from neural network



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

Basic Structure of a Neural Network Classifier





Classifier:
logistic regression

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

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

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

Inputs: energy flow

<u>Lots of inputs</u>

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

Feature map

<u>Too many parameters</u>

We simplify inputs and the feature map to **<u>linearize</u>** this classifier for the interpretability.

The decision boundary is <u>highly nonlinear</u>.
It is hard to <u>interpret</u> the NN itself...

Functional Taylor Expansion

See also
energy flow polyomials
arXiv:1712.07124

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

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

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

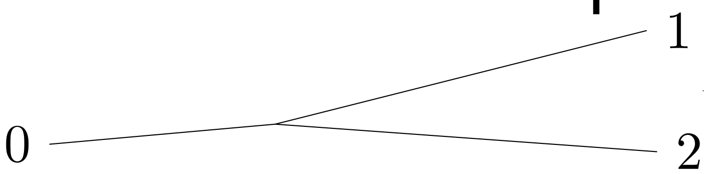
$$S_{2,ab}(R) = \int d\vec{R}_1\, d\vec{R}_2\; P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_2) \delta(R - R_{12})$$

Two-point correlation between constituents at distance *R*

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

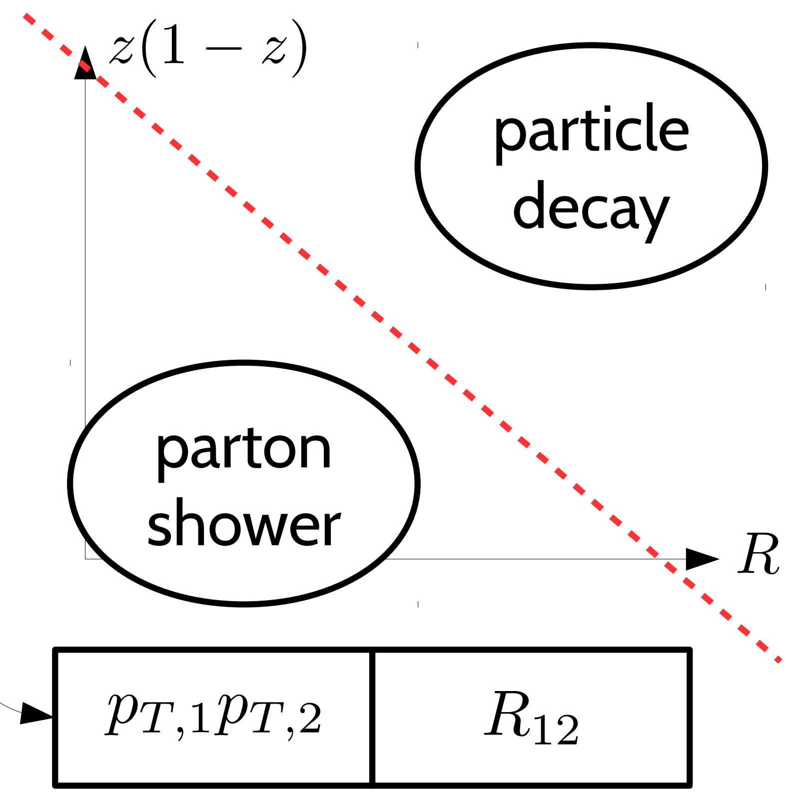
Kinematics inside Jet

The parameter set $(p_{T,0}, z, R)$ is a set of charateristic variables for the kinematics of parton evolution.

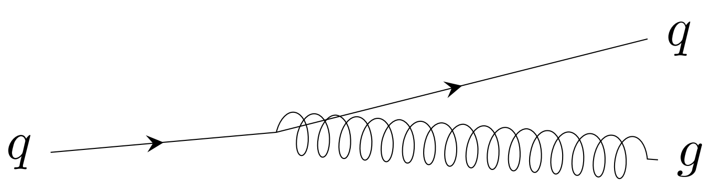


$$p_{T,1} = p_{T,0}z \qquad p_{T,2} = p_{T,0}(1-z)$$

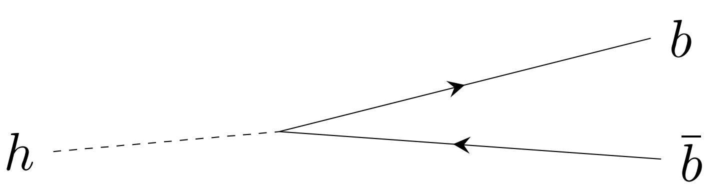
$$R^2 = (y_1 - y_2)^2 + (\phi_1 - \phi_2)^2$$



- parton shower



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



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

The second order term must be a dominant term for the classification.

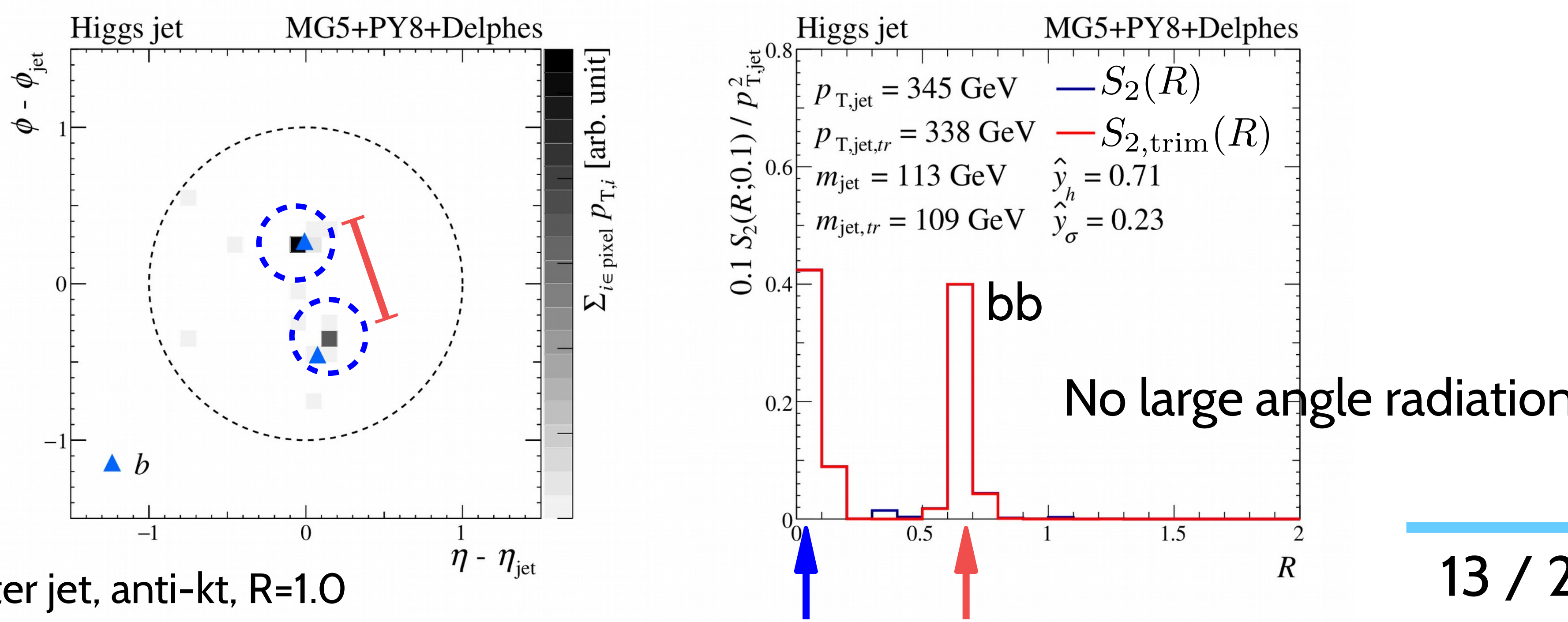
Two-Point Correlation Spectrum: Trimmed Spectrum

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

Two-Point Correlation Spectrum: Hard-Soft Correlation

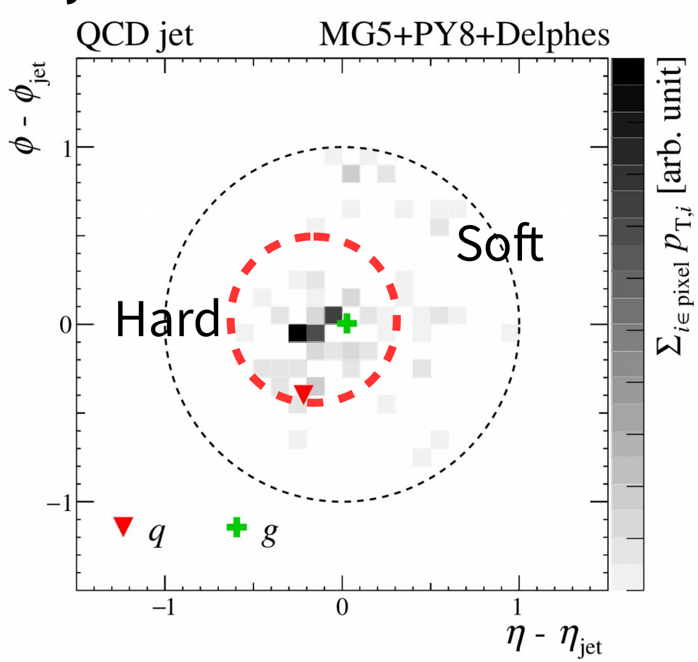
QCD jets have significant soft radiations. We may consider correlation between the soft parts and the hard parts.

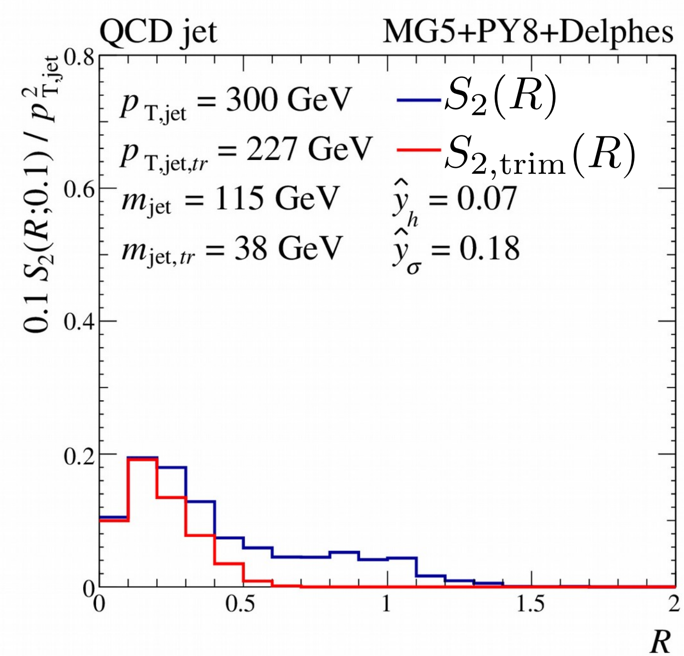
$$S_{2,\mathrm{soft}}(R) = S_2(R) - S_{2,\mathrm{trim}}(R)$$

sensitive to **hard-soft correlations**

subleading soft-soft correlations

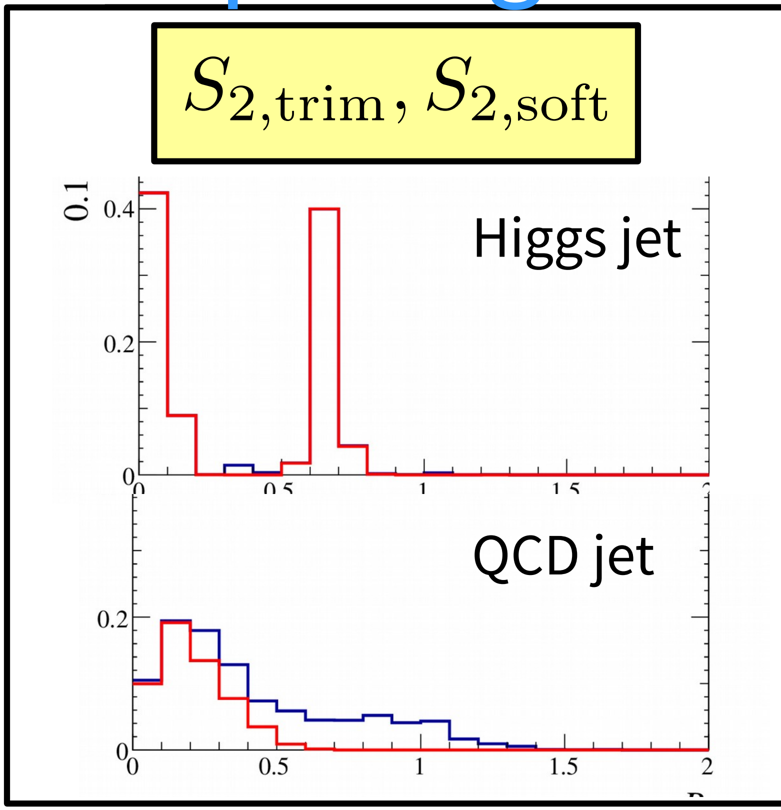
For QCD jet:

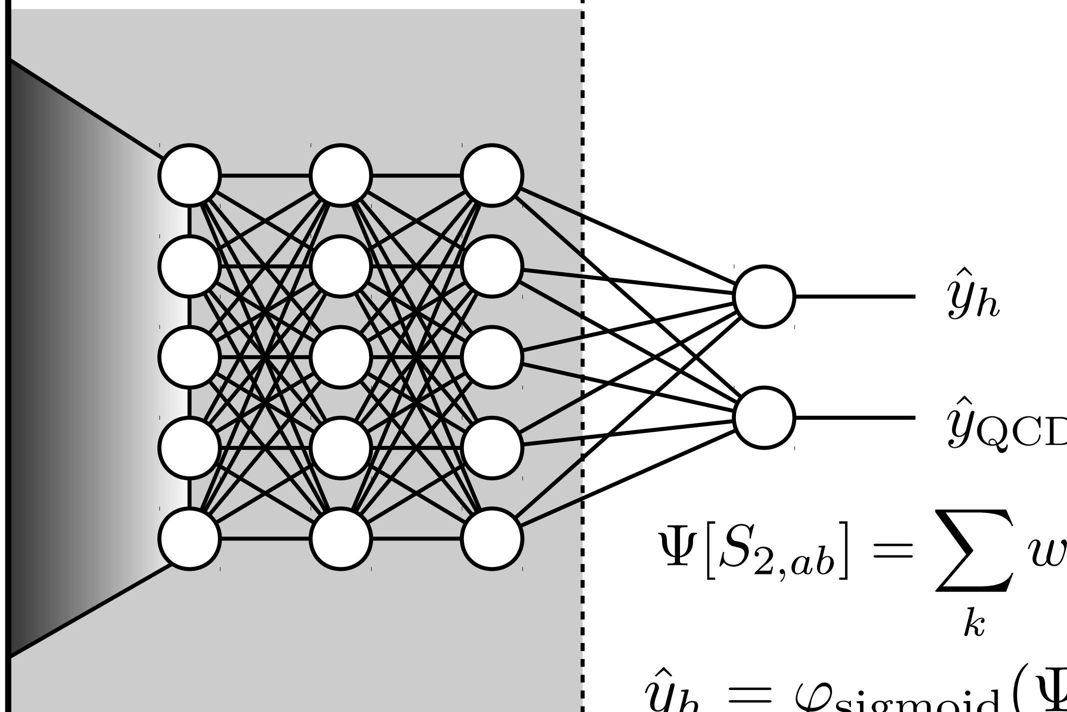




Calorimeter jet, anti-kt, R=1.0

Replacing CNN to MLP+S2





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

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

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

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

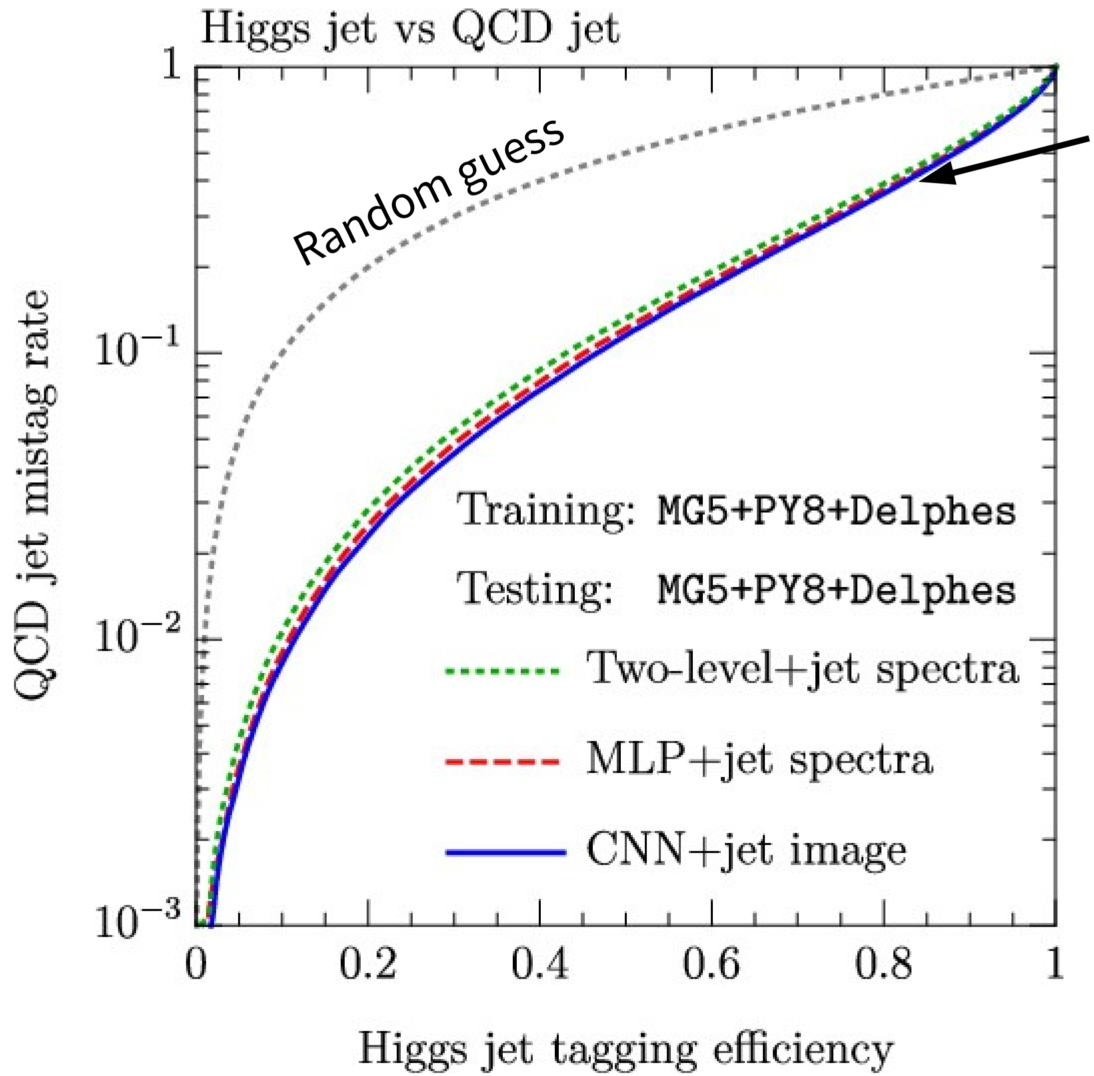
Feature map: MLP

Logistic regression

Although MLP is nonlinear, Higgs jets and QCD jets are separable with small number of inputs.

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

Equal Performance between CNN and MLP



similar performance

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

$$\mathrm{MLP} + \mathrm{S}_2 \sim \sum_{n=0}^{\infty} \mathcal{O}\left[P_T^{2n}\right]$$

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

Interpretable Setup:

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

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

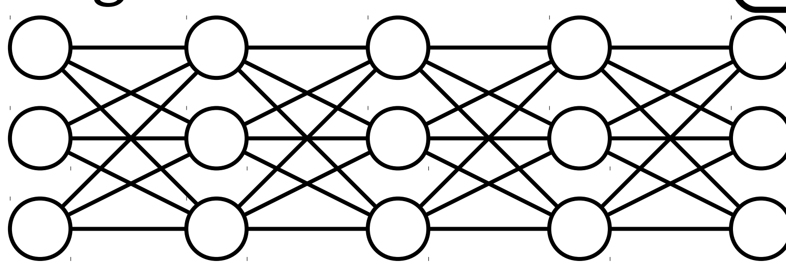
$$x_{\rm kin} = \{p_{T,{\rm jet}},\, m_{\rm jet}\}$$

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

Level2: kinematics analyzer

Use neural network to approximate the weight function.

$x_{\rm kin}$ $\Rightarrow$



$\Rightarrow$ $w_{ab}^{(2)}(R; x_{\rm kin})$

This architecture gives you two interpretable quantities:

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

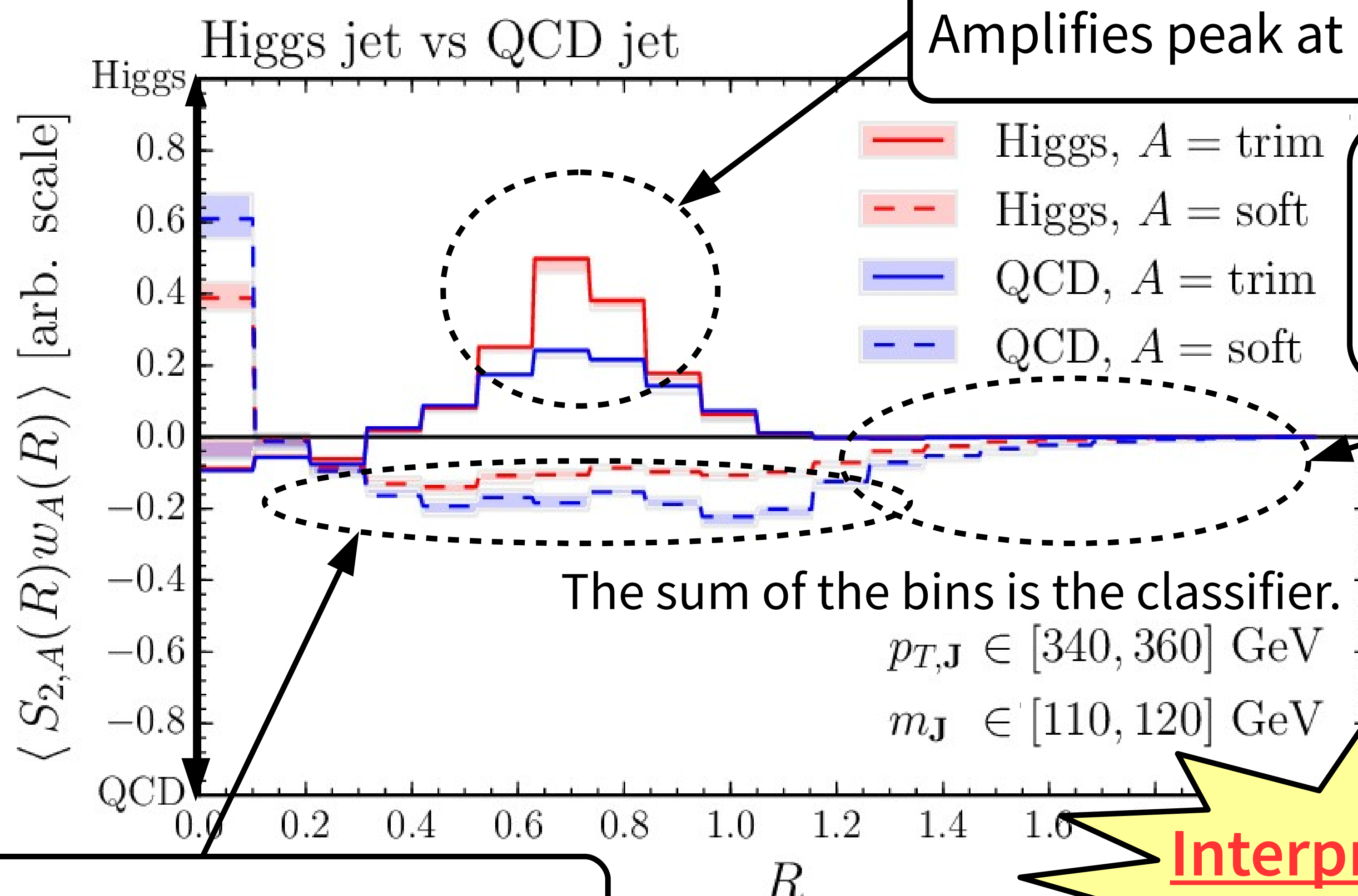
$S_{2,ab}(R) w_{ab}^{(2)}(R; x_{\rm kin})$ shows the **contribution** to the classifier.

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**

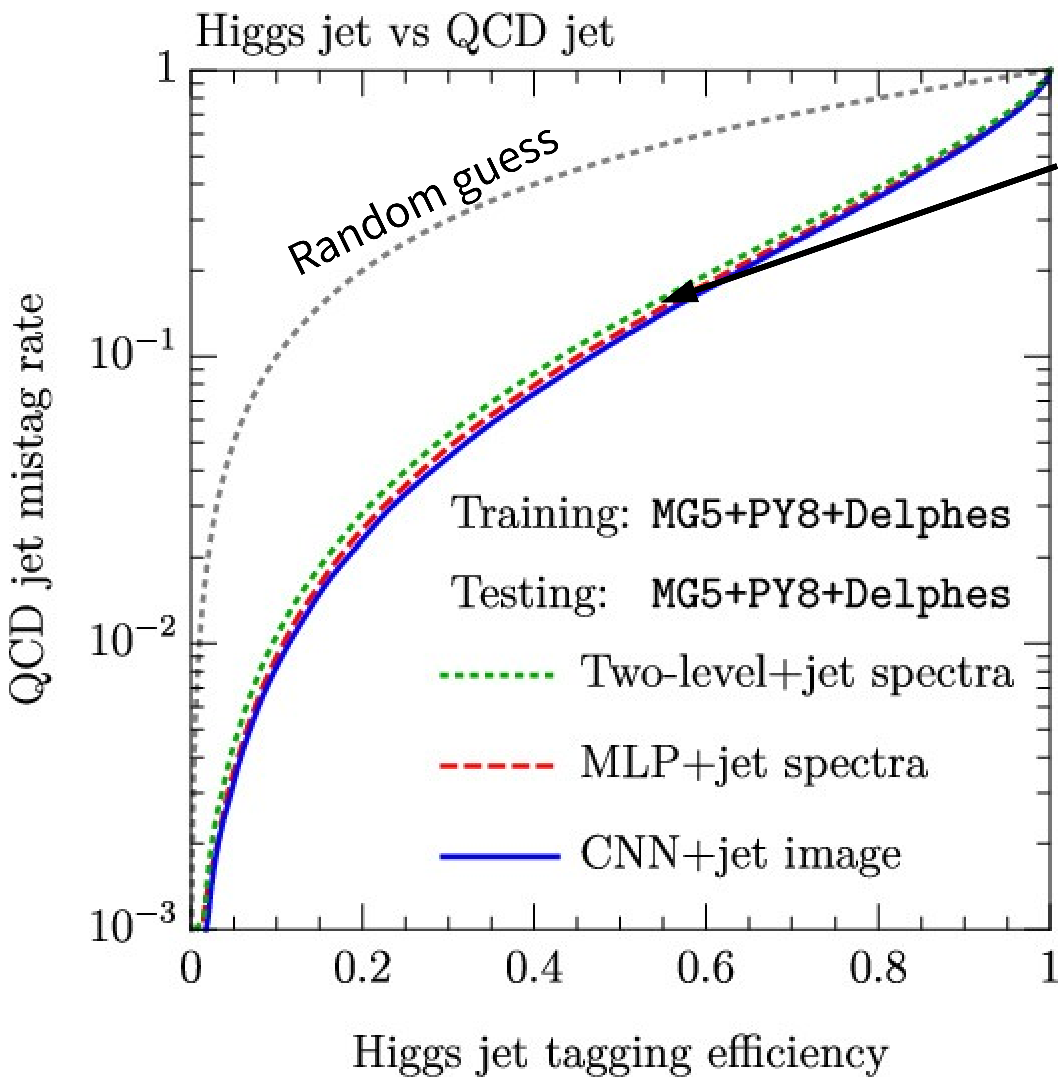


The sum of the bins is the classifier.

More soft activity: **QCD jet**

Interpretable

Similar Performance between CNN and the interpretable network



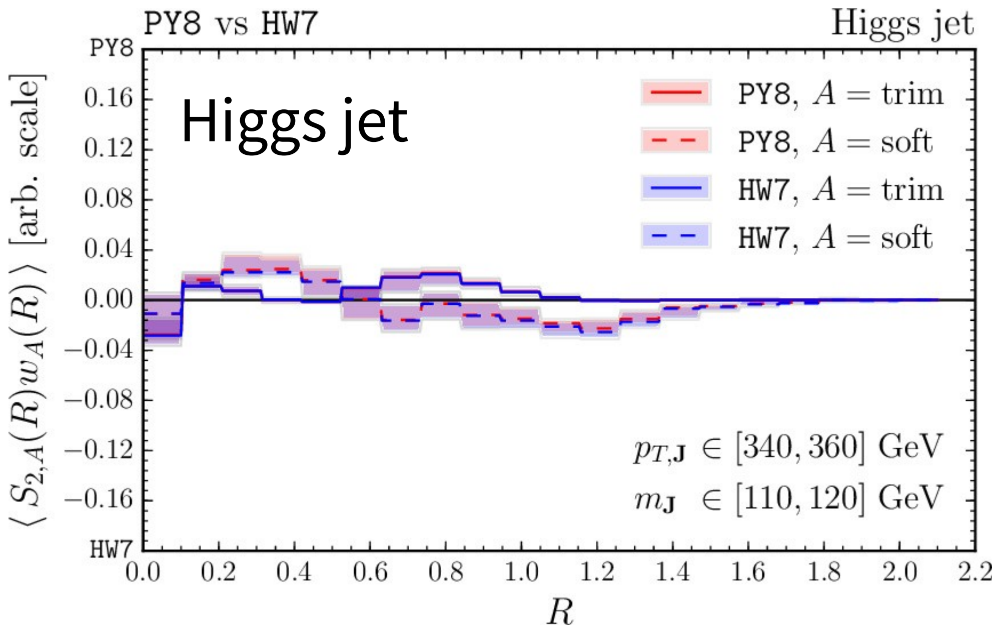
Still similar performance!

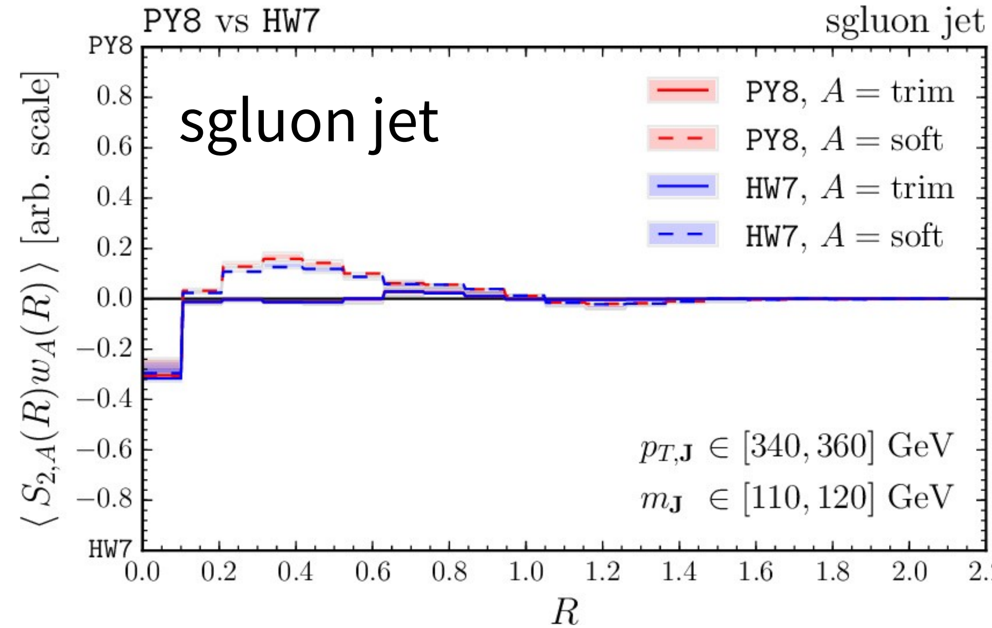
$$\mathrm{MLP} + \mathrm{S}_2 \sim \sum_{n=0}^{\infty} \mathcal{O}\left[P_T^{2n}\right]$$

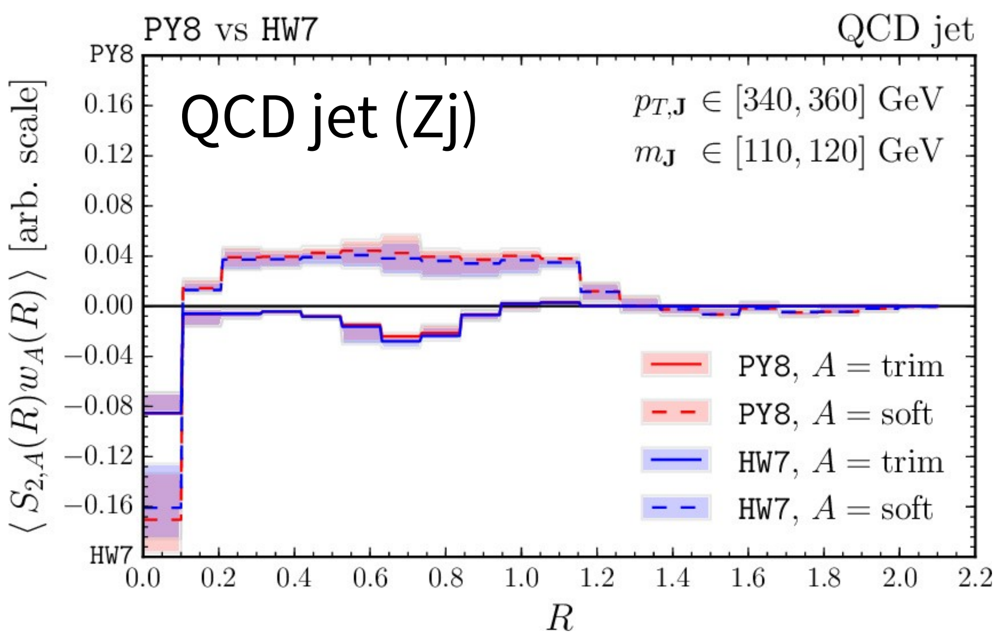
$$\text{Two-level} + S_2 \sim \mathcal{O}\left[P_T^2 \cdot w(p_{T,\mathbf{J}}, m_{\mathbf{J}})\right]$$

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

Comparing Pythia and Herwig







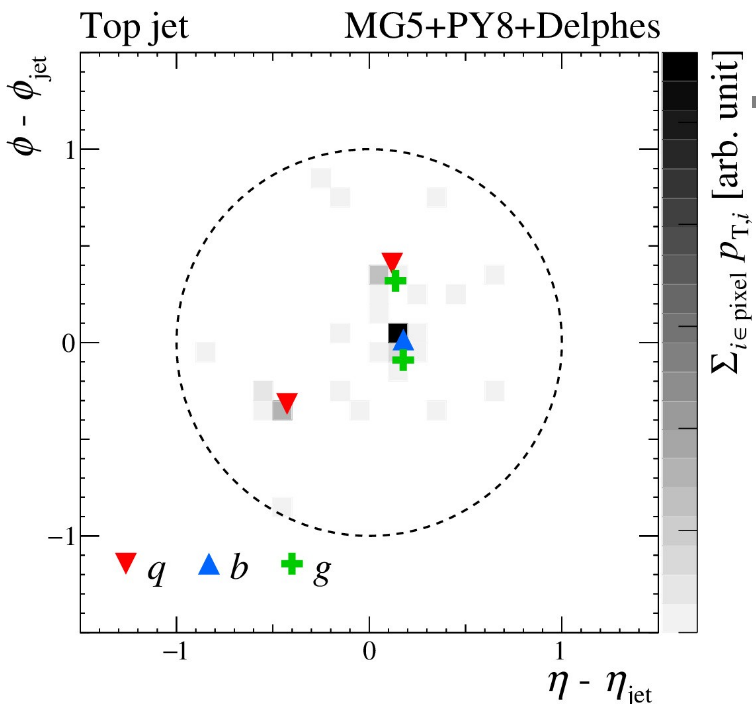
Most of the classifier contribution comes from soft activity

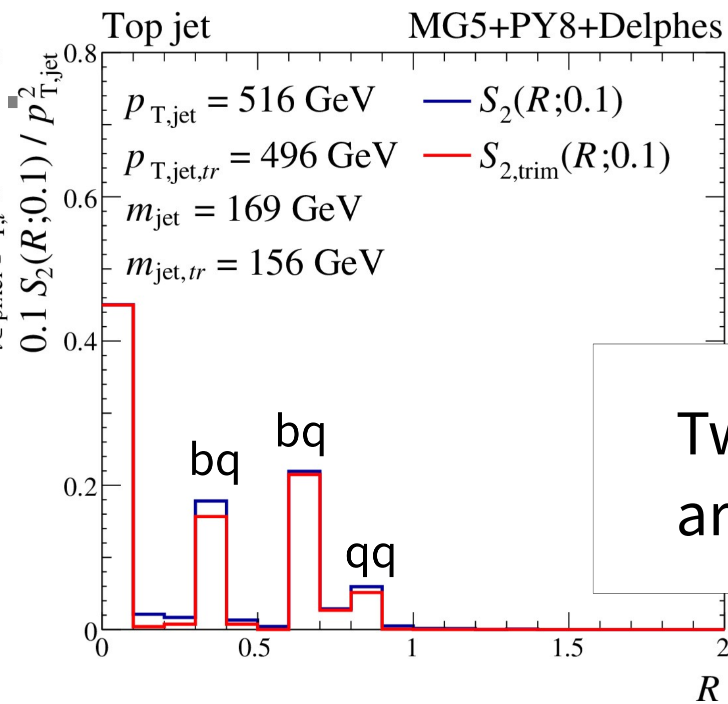
PY8 jets are more compact than HW7 jets

Singlet and octet have different behavior

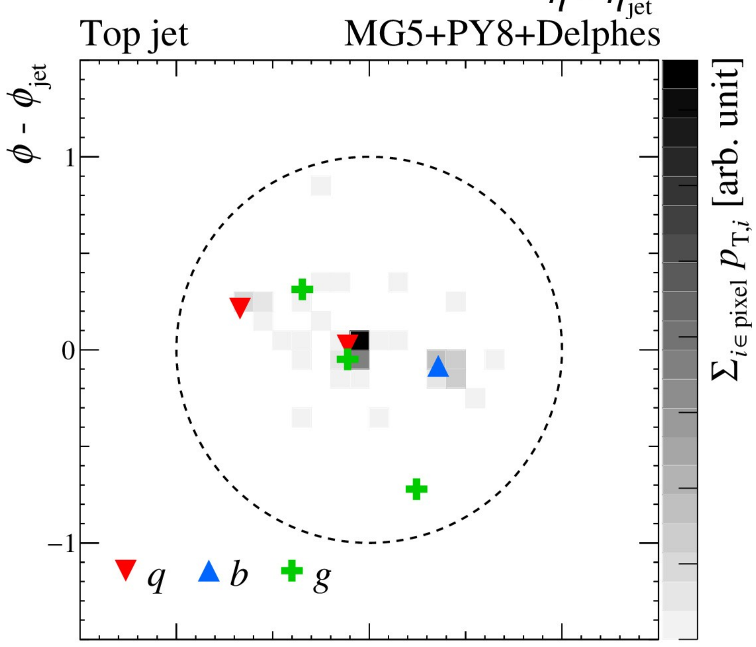
Preview: Top Jets

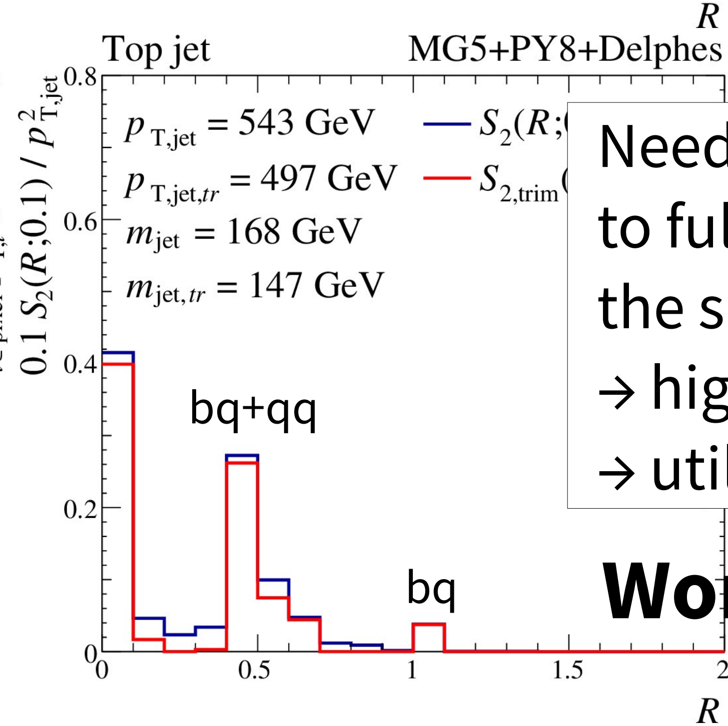
Preliminary





Two-point correlations are enough.





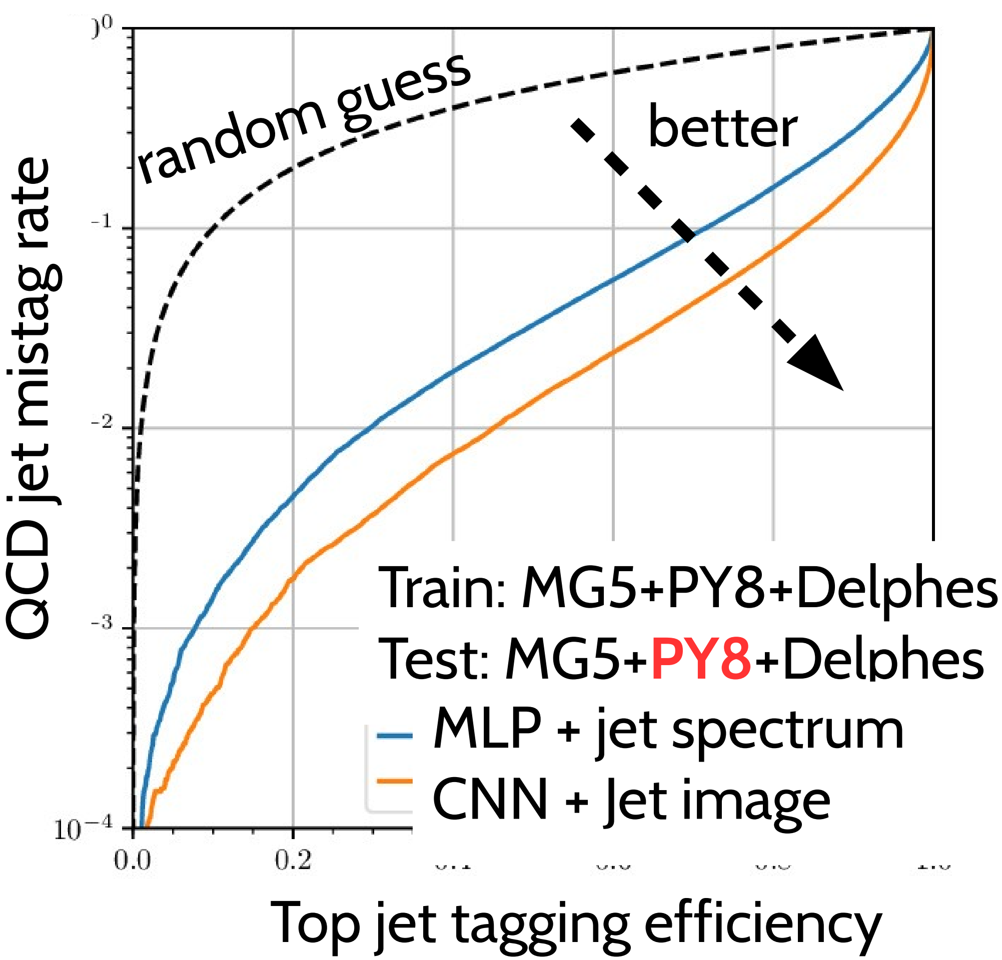
Calorimeter jet, anti-kt, R=1.0

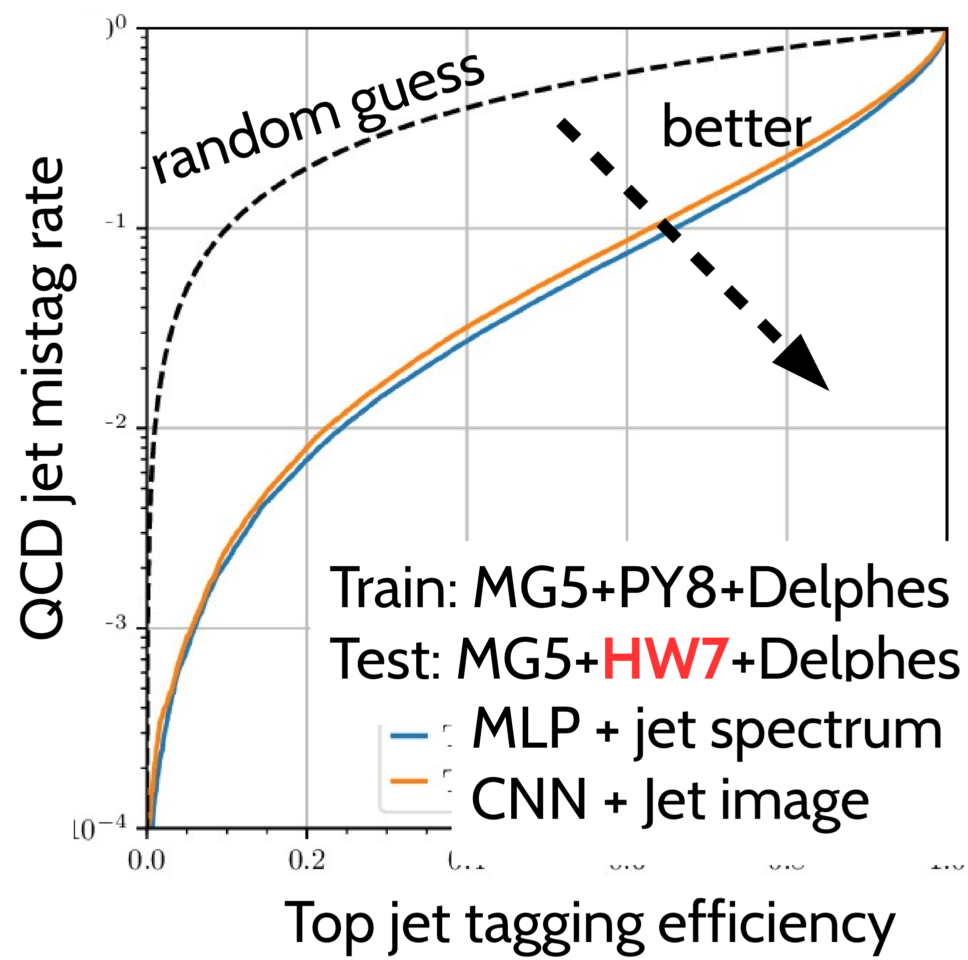
Need more information
to fully encode
the substructures.
→ higher order correlation?
→ utilize subjet information?

Work in progress...

Preview: Top Jets vs. QCD jets

Preliminary





The CNN and MLP+S2 setups are **comparable** within the uncertaintiy between PY8 and HW7.

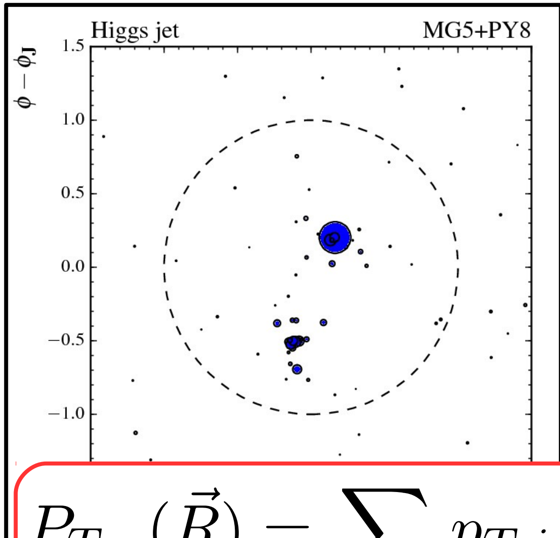
Work in progress...

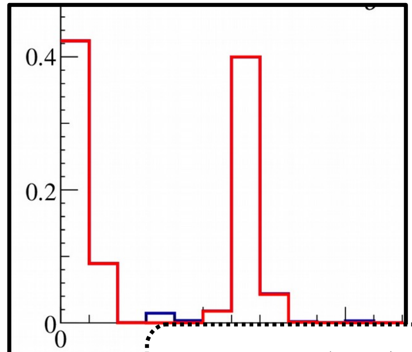
Summary

- 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 <u>**two-point correlation spectrum**</u> for analyzing jet substructures.
- The spectrum is derived from the jet image analysis and the corresponding two-level model is <u>**interpretable**</u>.
- 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





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

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

$$S_{2,ab}(R)$$

discretization

$$P_{T,a}(\vec{R}) = \sum_{i \in \mathbf{J}_a} p_{T,i}\, \delta(\vec{R} - \vec{R}_i) \rightarrow S^k_{2,ab}$$

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

discretization

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

sigmoid

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

x_{kin}

$$\begin{pmatrix} p_{T,\mathbf{J}} \\ m_{\mathbf{J}} \end{pmatrix} \rightarrow \text{MLP} \rightarrow w^k_{ab}(x_{\text{kin}})$$

MLP

Activation: ELU

discretization

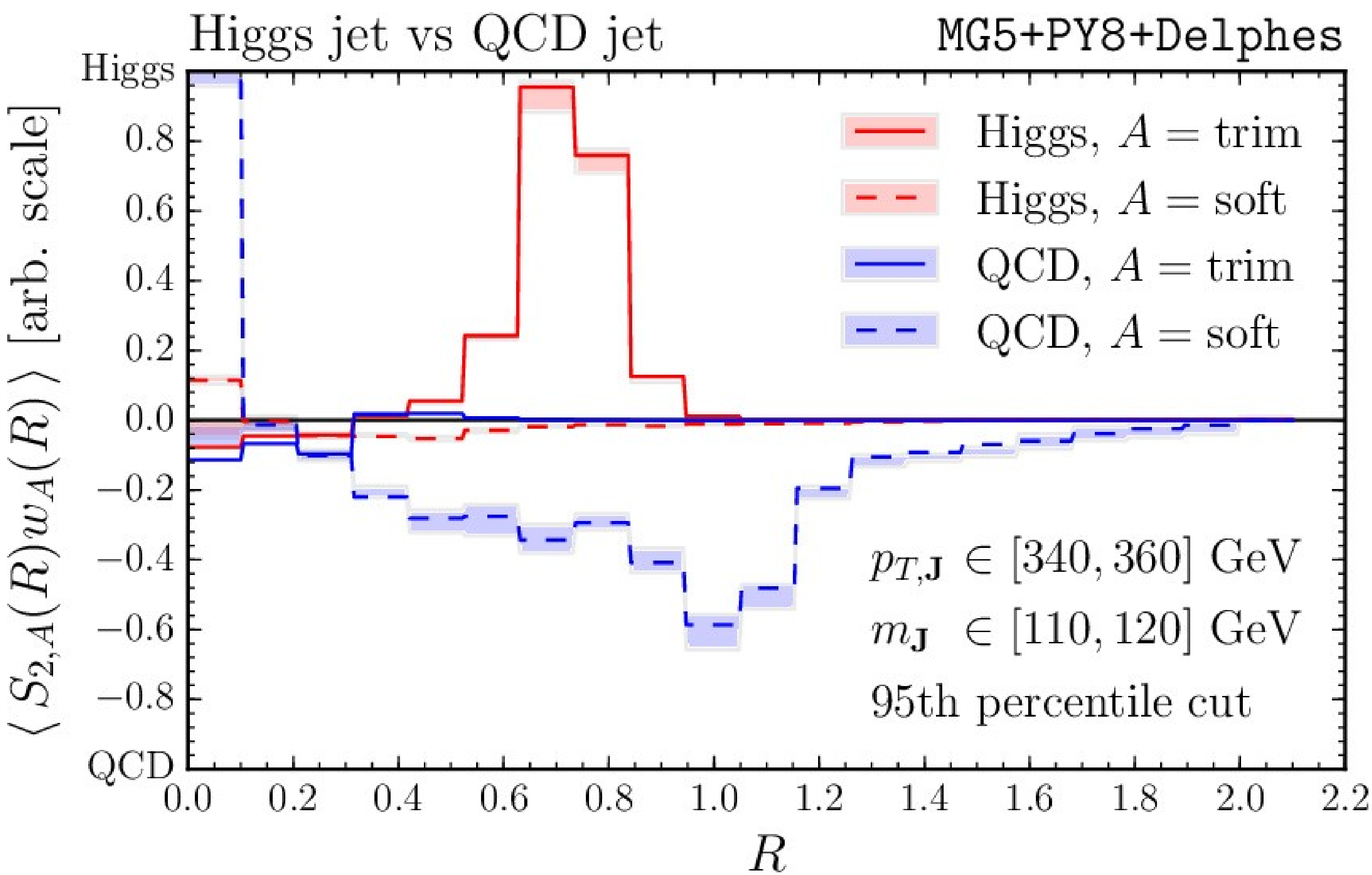
approximate

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

cf.

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

Average of the linear classifier outputs



Two-Prong jet: 1 vs 8

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