Accounting Machine Learning in Jet classification

Sung Hak Lim, Mihoko Nojiri (JHEP, 2018)
Amit Chakraborty, Sung Hak Lim, Mihoko Nojiri (arXiv 1904.02092)

Machine learning in Jet Physics

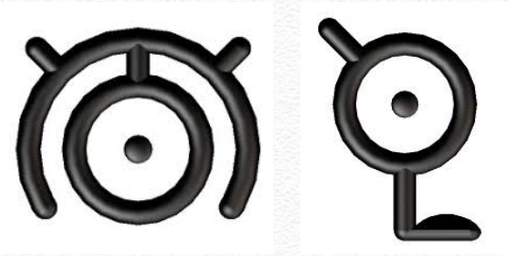
Run III→HL-LHC→FCC?

Huge data

*Need to push our understanding further
** And we want to do this quickly

hadronization,
PDF, parton shower
modeling, …

high pT objects (Events in Tail)
soft object, mono somthing
or something unkown

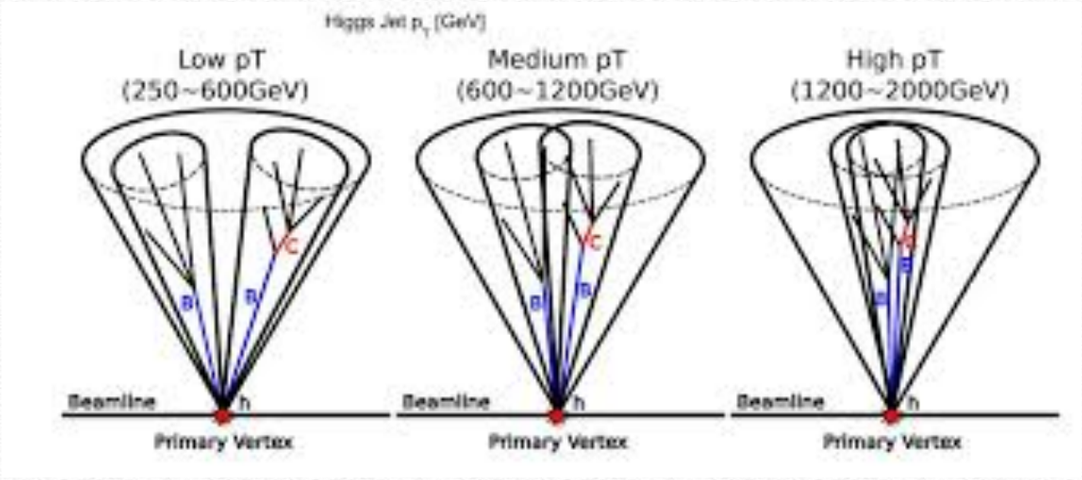


Physics outputs
effective operator,
top partner
dark matter..

Jets and Deep Learning

- Jet physics :One of the most successful area of QCD **(Theoretical understanding + computation)**
  - Jet reconstruction algorithm (kT, CA, antikT)
  - QCD jet (huge background) vs boosted Higgs →Jet substructure (mass drop), Top reconstruction (BSM search)
  - Various jet quantities and minimal Validation Analysis (leaving optimization to algorithm)

theory and computational developments

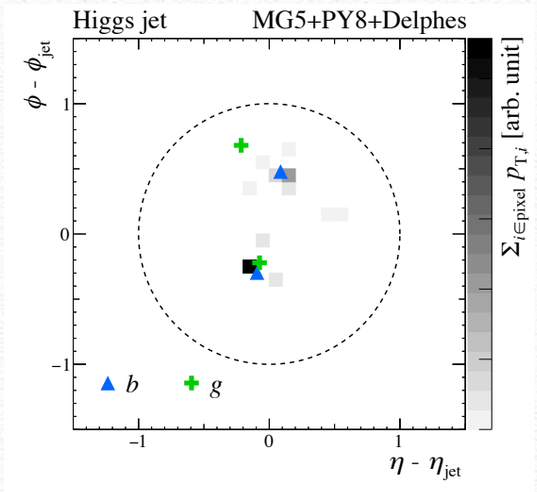


New wave —— machine learning

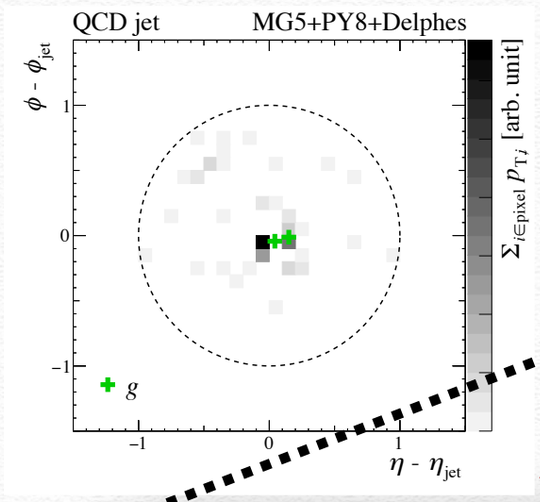
ML:basic units

Input:Jet images

Higgs



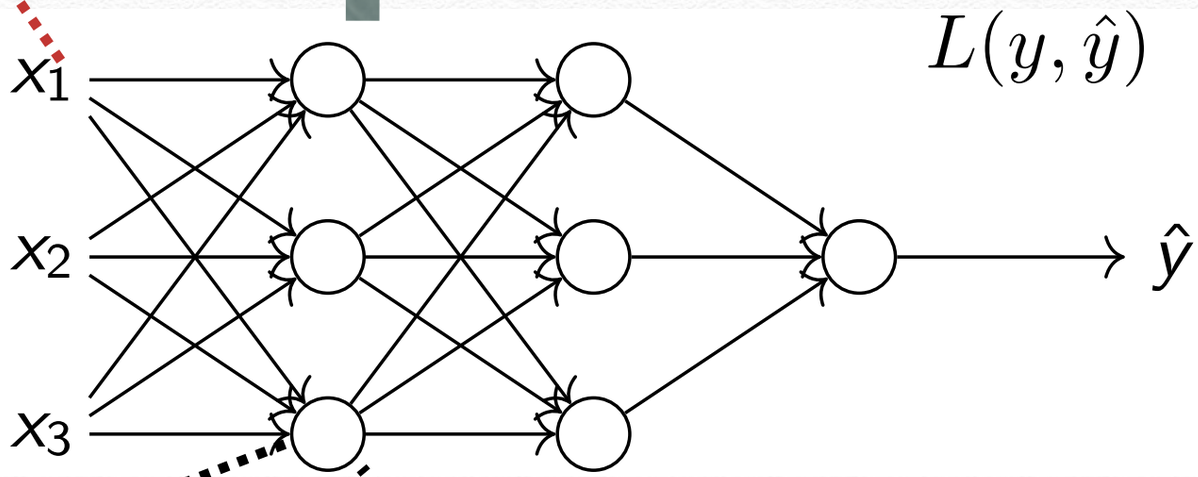
QCD



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

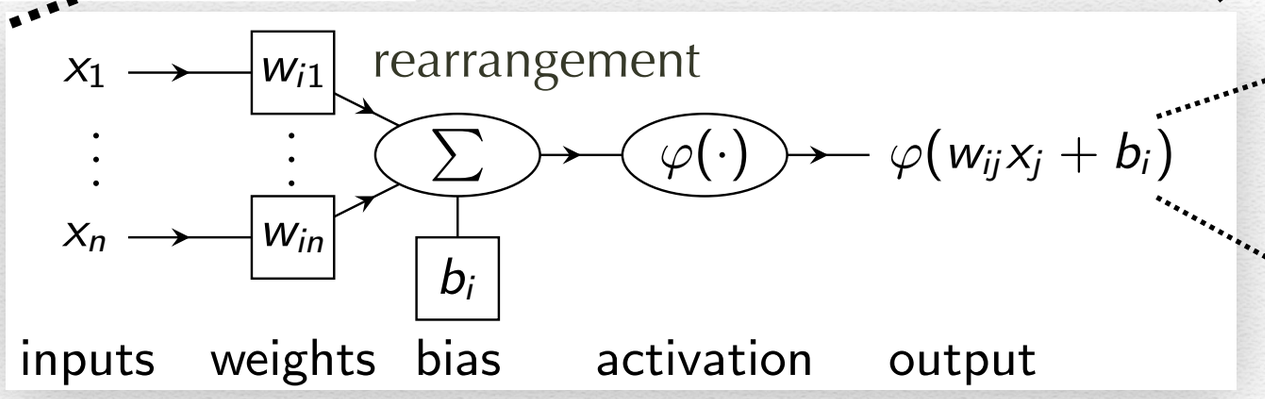
optimization

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

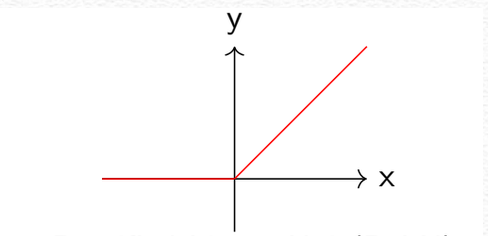


Higgs
$y = (1, 0)$

QCD
$y = (0, 1)$



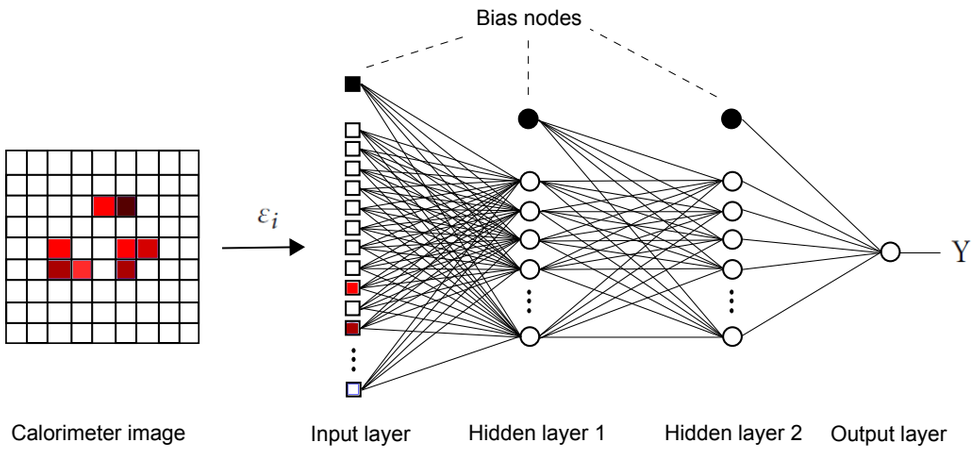
φ: source of non linearity

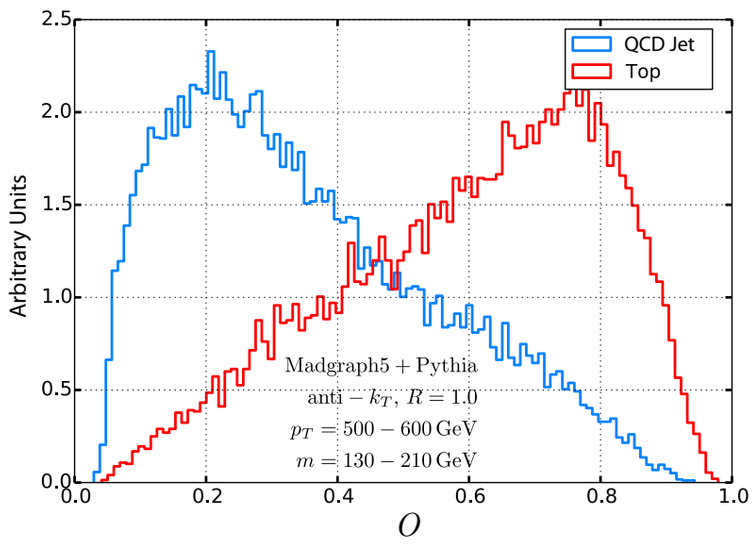


Effort to have more “structured the input”

Almeida et al 1501.05968

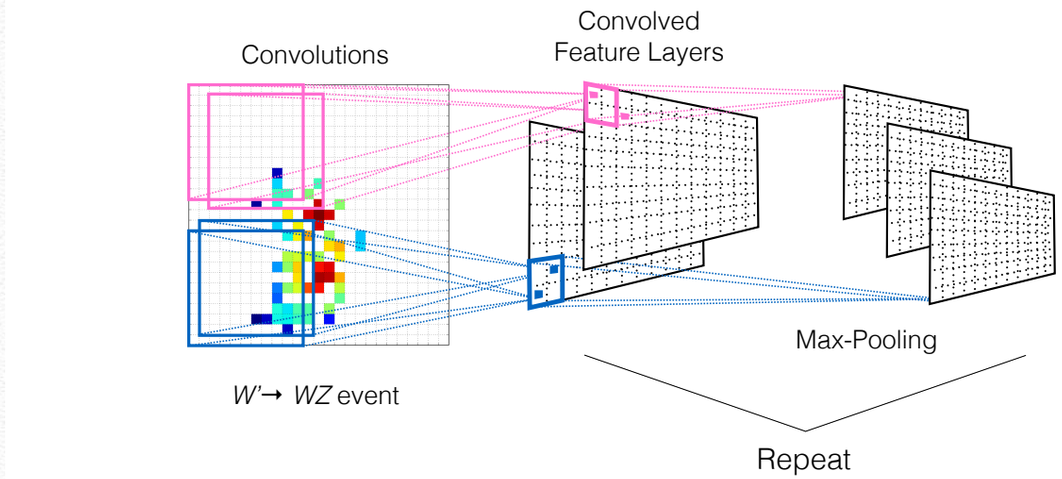
DNN (all bins in a line)



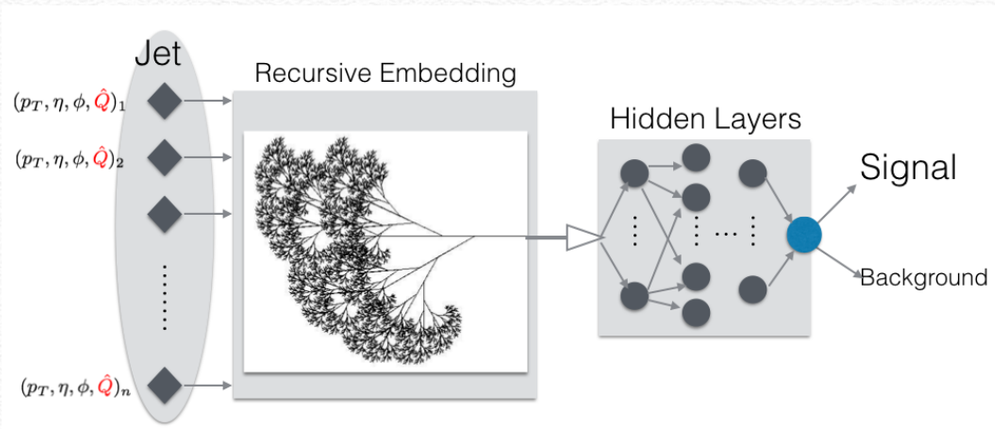


CNN

correct nearby image first 1511.05190



Recursive(Taoli Cheng 1711.02633)



“ Maybe we do not know physics behind, architecture do the job for you”??

We want to know what ML doing

- Why ML give us better result in classification?
- What is happening in Hidden layer?
- The result is consistent with out current QCD knowledge or anything beyond that?

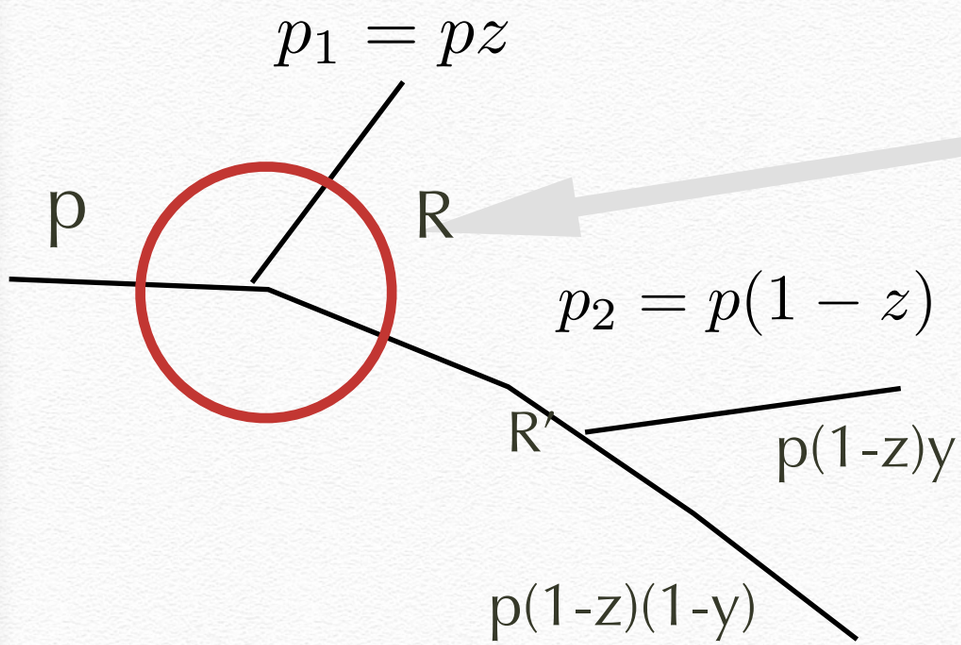
summary of our strategy

1. Instead using **jet image(400 inputs) CNN**, give more motivated and **minimum (~40) inputs** (**jet spectrum**) to DNN
2. Compare the CNN classifier and the DNN
3. Construct an "Interpretable model" using jet spectrum

They are Equivalent!

Minimum input (Jet spectrum)

- Monte Carlo →Parton splitting + hadronization



(p, R, z) describe
parton shower spritting

$$p_1 p_2 = p^2 z(1-z)$$

$$p_1^2 + p_2^2 = p^2[(1-z)^2 + z^2$$

Our minimum input (Jet Spectrum)

$S_2(R, \Delta R) = \sum_{ij} p_{Ti} p_{Tj}$ for $R < R_{ij} < R + \Delta R$ for R = 0, 0.1, 0.2 …

sum over pairs of jet constituents
proportional to momenta of particles to be IRC safe (C-correlator

Relation to the jet image classifiers

jet = energy flow (+ ...) $P_T(\vec{R}) = \sum_{i \in \mathbf{J}} p_{T,i}\, \delta(\vec{R} - \vec{R}_i),$

calorimeter hit position

*classifier using energy flow $h_i = \hat{\Psi}_i[P_T]$

$$h_i = w_i^{(0)} + \int d\vec{R}\, P_{T,a}(\vec{R}) w_{i,a}^{(1)}(\vec{R}) + \frac{1}{2!}\int d\vec{R}_1 d\vec{R}_2\, P_{T,a}(\vec{R}_1) P_{T,b}(\vec{R}_2) w_{i,ab}^{(2)}(\vec{R}_1, \vec{R}_2) + \cdots$$

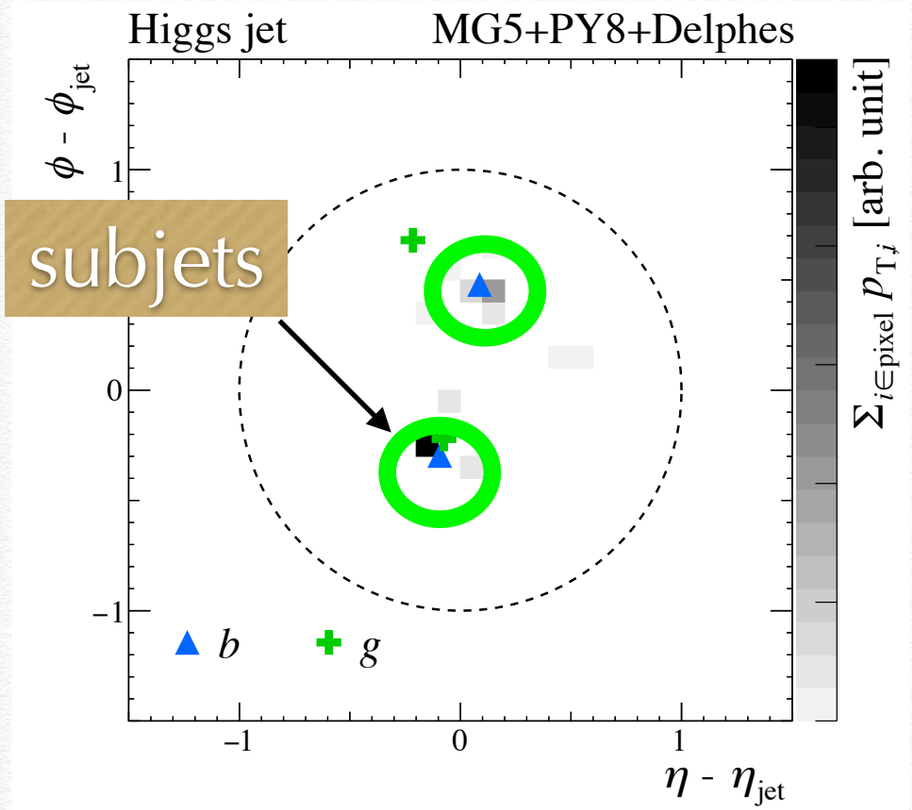
If w only depends on R_{12}

$$\frac{1}{2!}\int dR\, S_{2,ab}(R)\, w_{i,ab}^{(2)}(R) + \cdots$$

*classifier using Jet spectrum

$$h_i = w_i^{(0)}(\vec{x}_{\rm kin}) + \int dR\, S_{2,A}(R)\, \frac{w_{i,A}^{(2)}(R; \vec{x}_{\rm kin})}{2} + \frac{1}{2}\int dR_1 dR_2\, S_{2,A}(R_1) S_{2,B}(R_2)\, \frac{w_{i,AB}^{(4)}(R_1, R_2; \vec{x}_{\rm kin})}{12} + \cdots .$$

Soft and hard activity in the jet



momentum density plot

jet~ subjets(IRC safe)

subjets above certain pt cut

=>trimmed jet

$$\mathbf{J}_{\text{trim}} = \bigcup_{\substack{a \\ \frac{p_{T,\mathbf{J}_a}}{p_{T,\mathbf{J}}} \geq f_{\text{trim}}}} \mathbf{J}_a \ .$$

$S_{2\text{trim}}$

$$S_{2,\text{trim}}(R;\Delta R) = \frac{1}{\Delta R} \sum_{i,j \in \mathbf{J}_{\text{trim}}} p_{T,i}\ p_{T,j} \cdot I_{[R,R+\Delta R)}(R_{ij}),$$

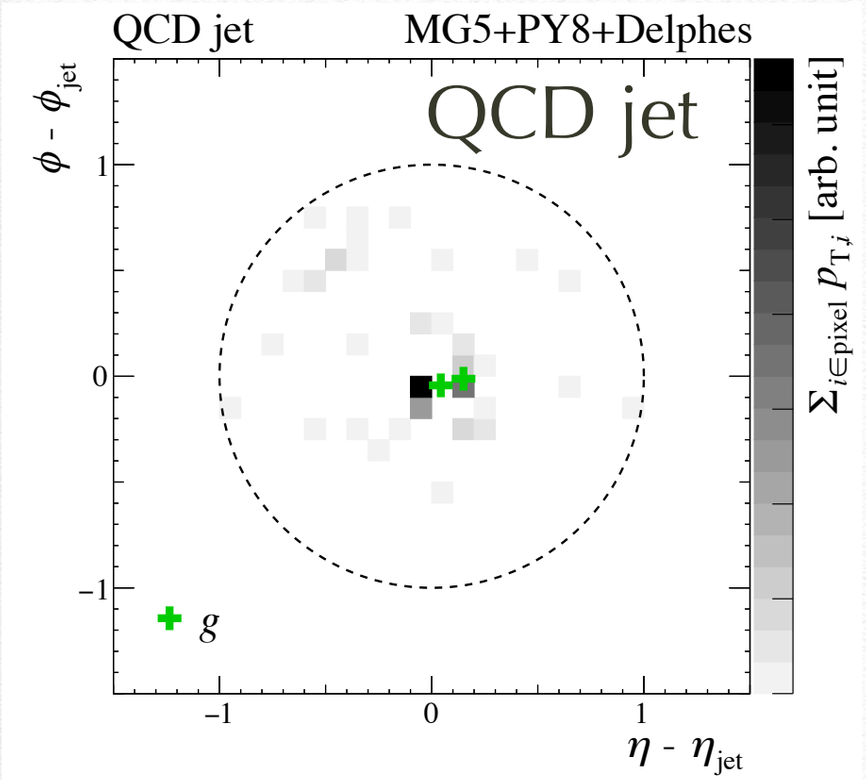
="hard-hard" correlation

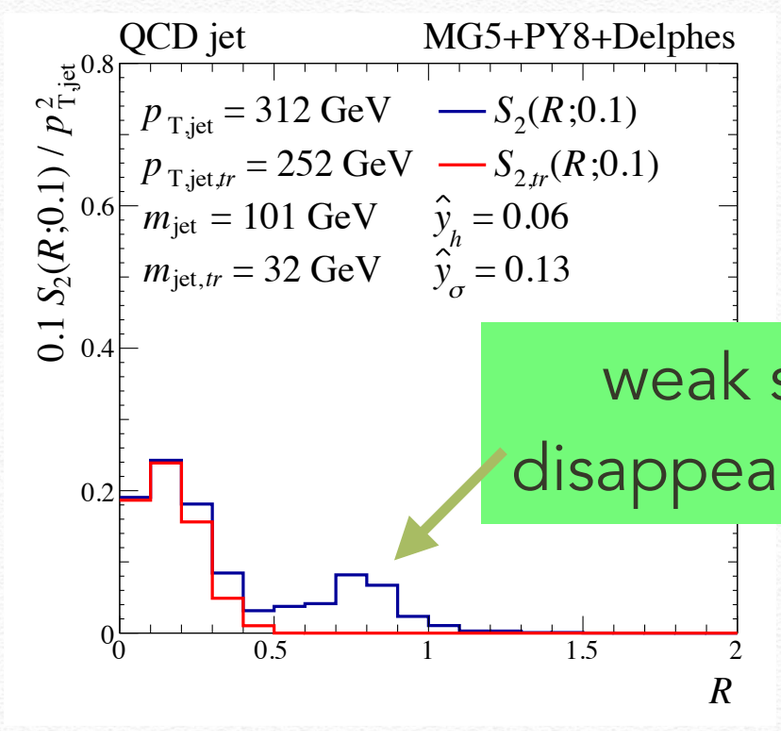
$S_{2\text{soft}}$

$$S_{2,\text{soft}}(R;\Delta R) = S_2(R;\Delta R) - S_{2,\text{trim}}(R;\Delta R).$$

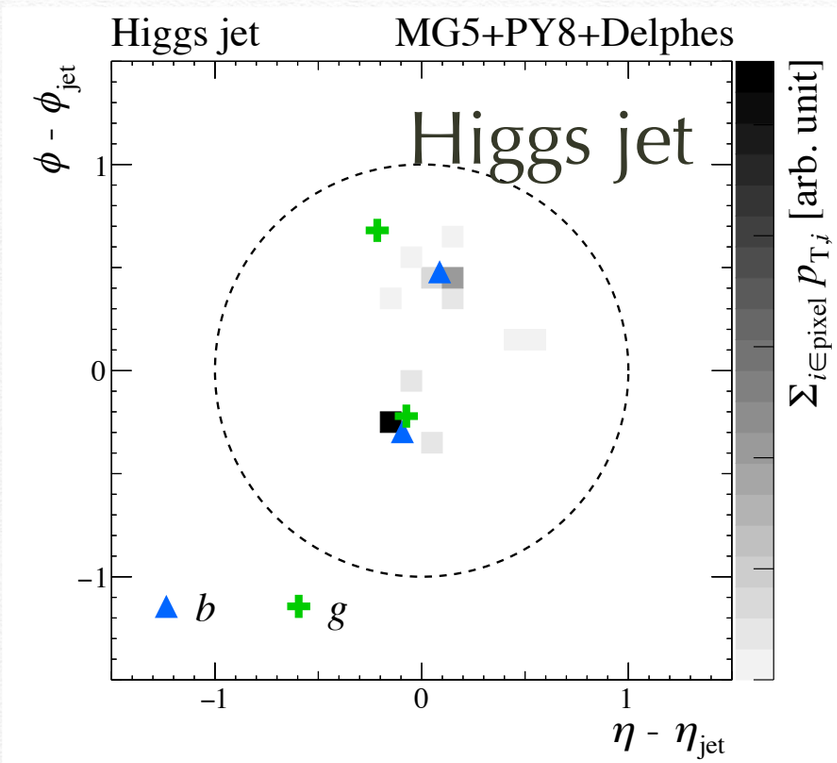
= "soft-hard correlation" + "soft-soft" correlation

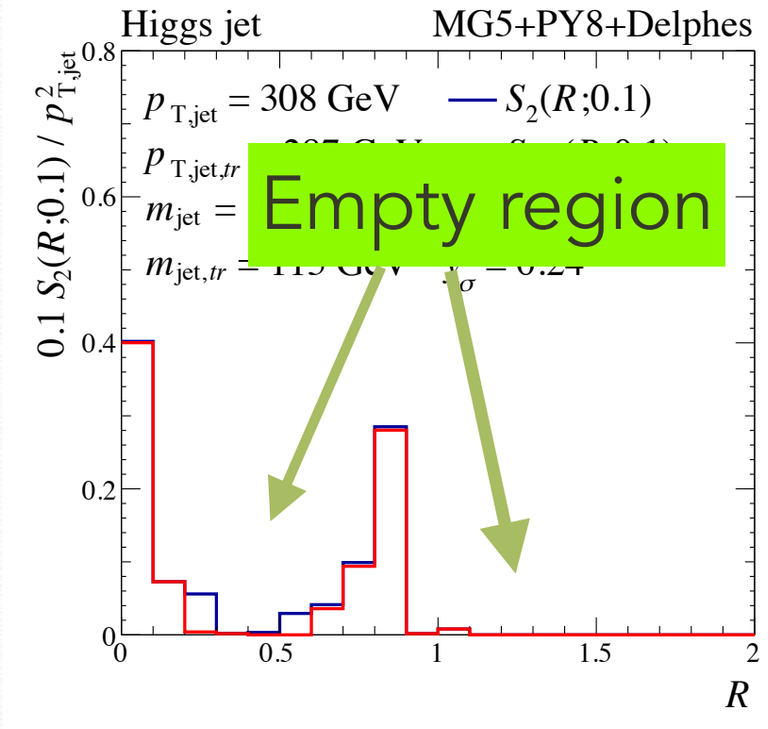
Typical $S_2(R)$ distribution





weak second peak
disappear after trimming

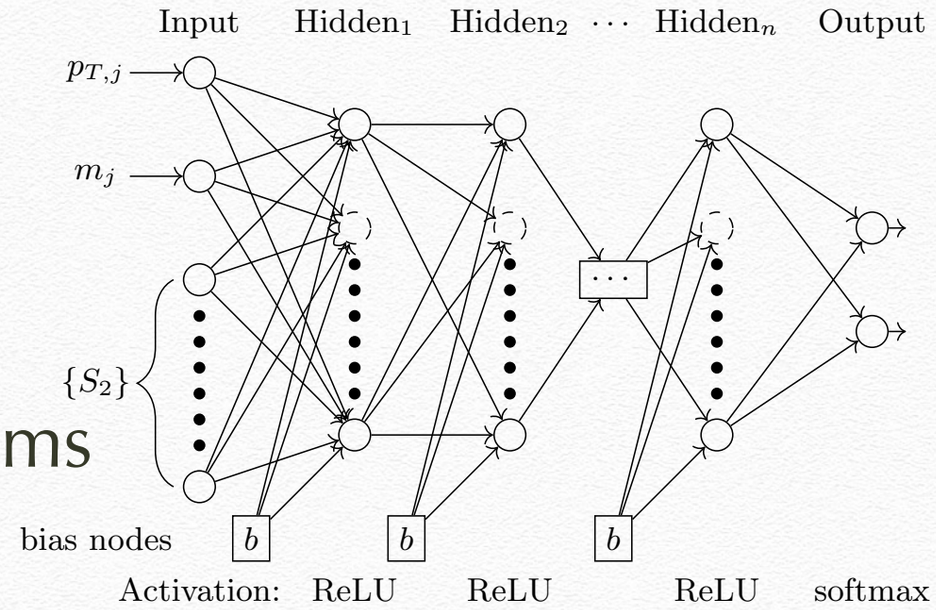


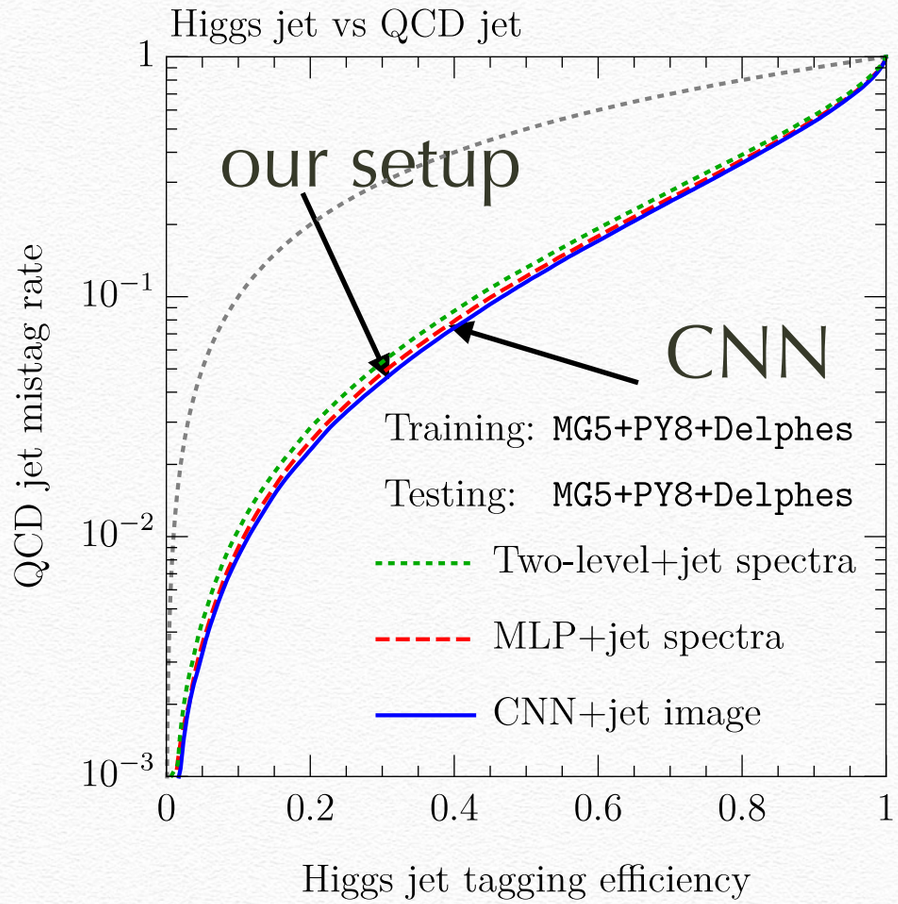


Empty region

CNN vs DNN with S_{2trim} (R) and S_{2soft} (R)

Jet spectrums





for 300GeV<pT<400GeV and 100GeV<mj<150GeV

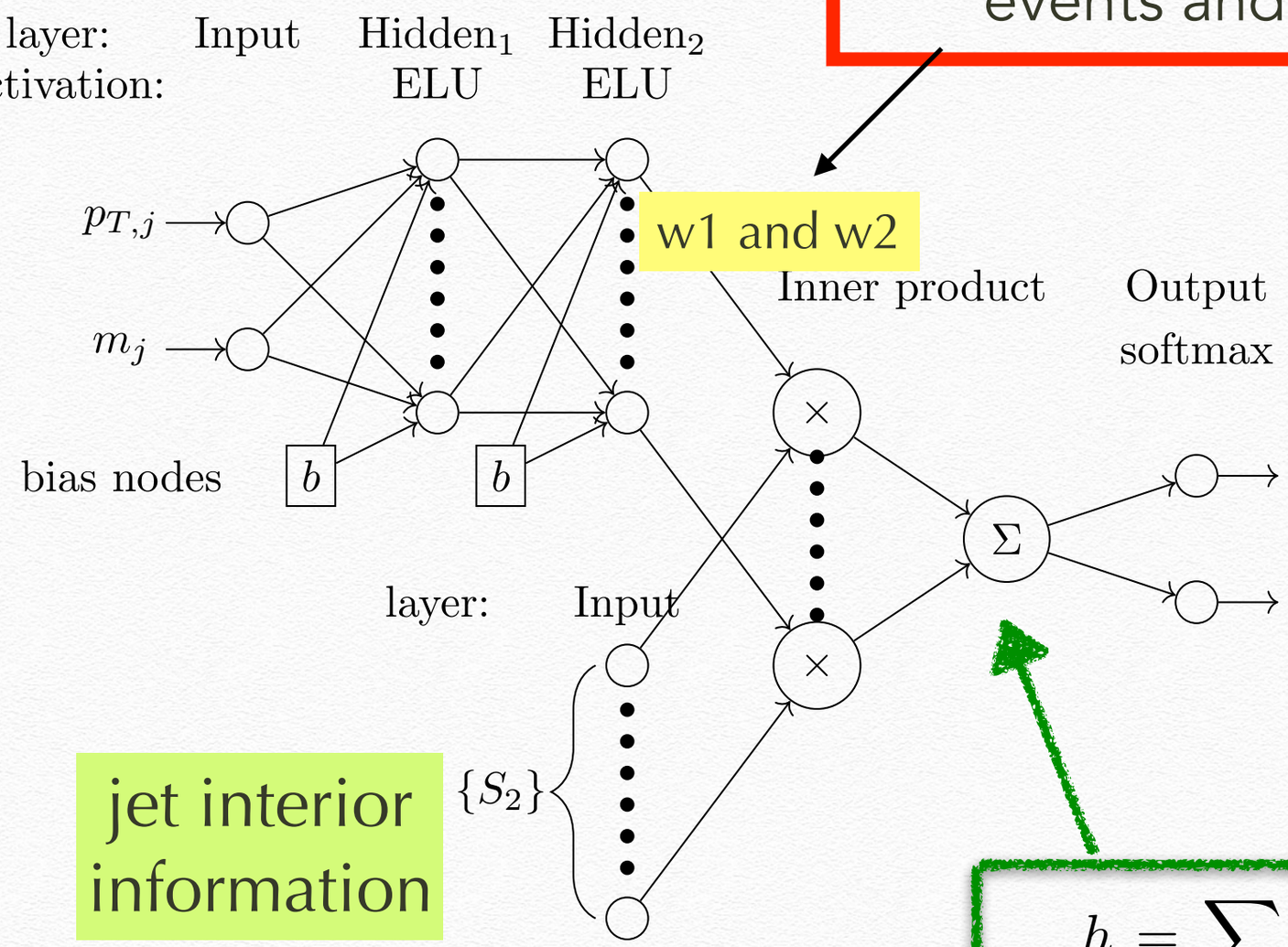
The results only slightly worse than CNN

- **Two point correlation pick up most important effects. (Or NN with jet image "find" two-point correlation by itself (not proven)**
- Chance to understand **the distribution contributing to the decision?**

[kinetic module] · [**radiation module**]

= [**Interpretable model**]

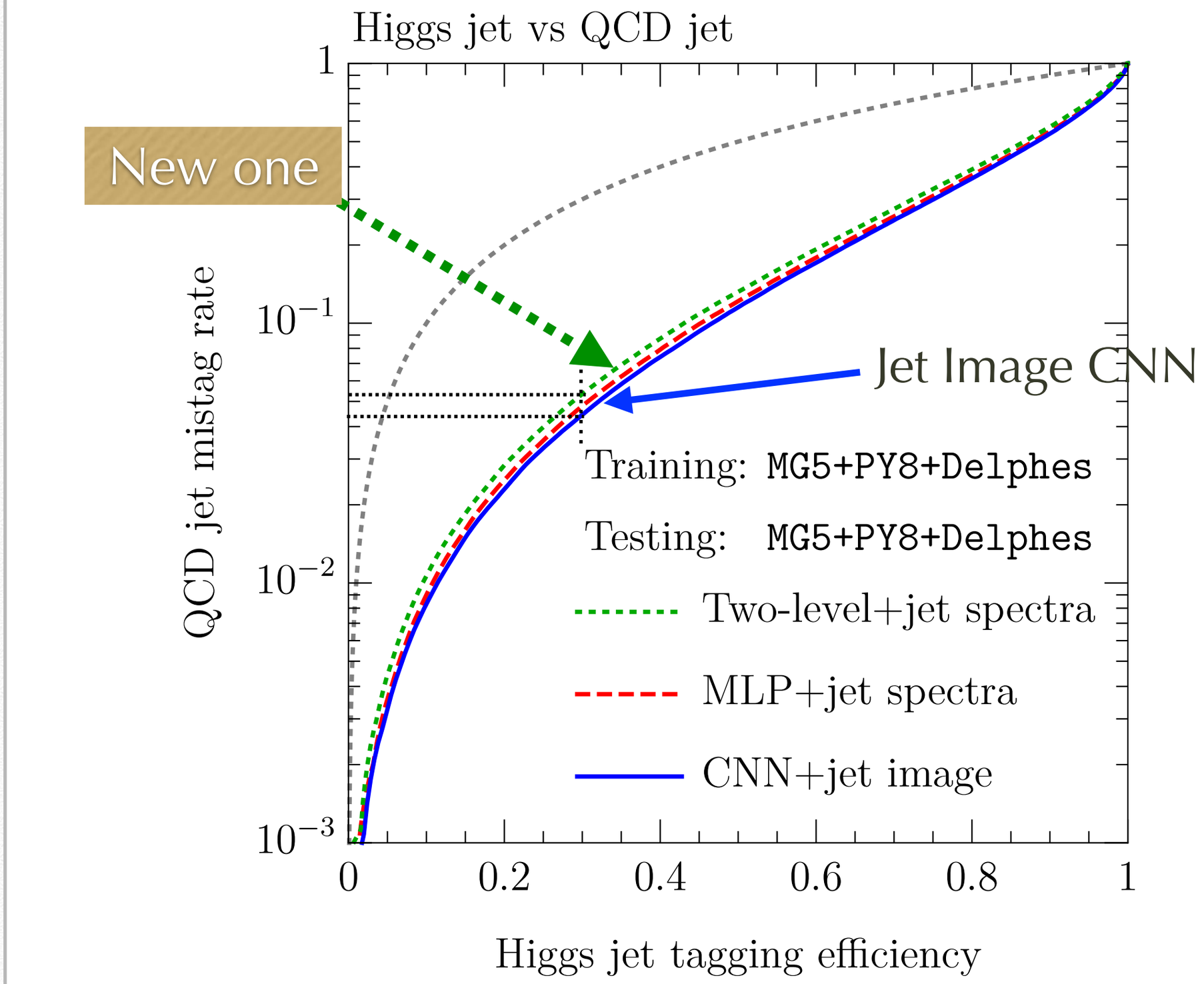
w1 and w2 are trained to be MAX for Higgs events and MIN for QCD jet events



jet interior information

$$h = \sum_k S^k_{2,\mathrm{trim}}\, w^k_1 + \sum_k S^k_{2,\mathrm{soft}}\, w^k_2,$$

This simple classifier performs nearly as good as previous ones


Higgs jet vs QCD jet
New one
Jet Image CNN
Training: MG5+PY8+Delphes
Testing: MG5+PY8+Delphes
Two-level+jet spectra
MLP+jet spectra
CNN+jet image
QCD jet mistag rate
Higgs jet tagging efficiency
1
10^{-1}
10^{-2}
10^{-3}
0
0.2
0.4
0.6
0.8
1

kinetic module vs radiation module with readout

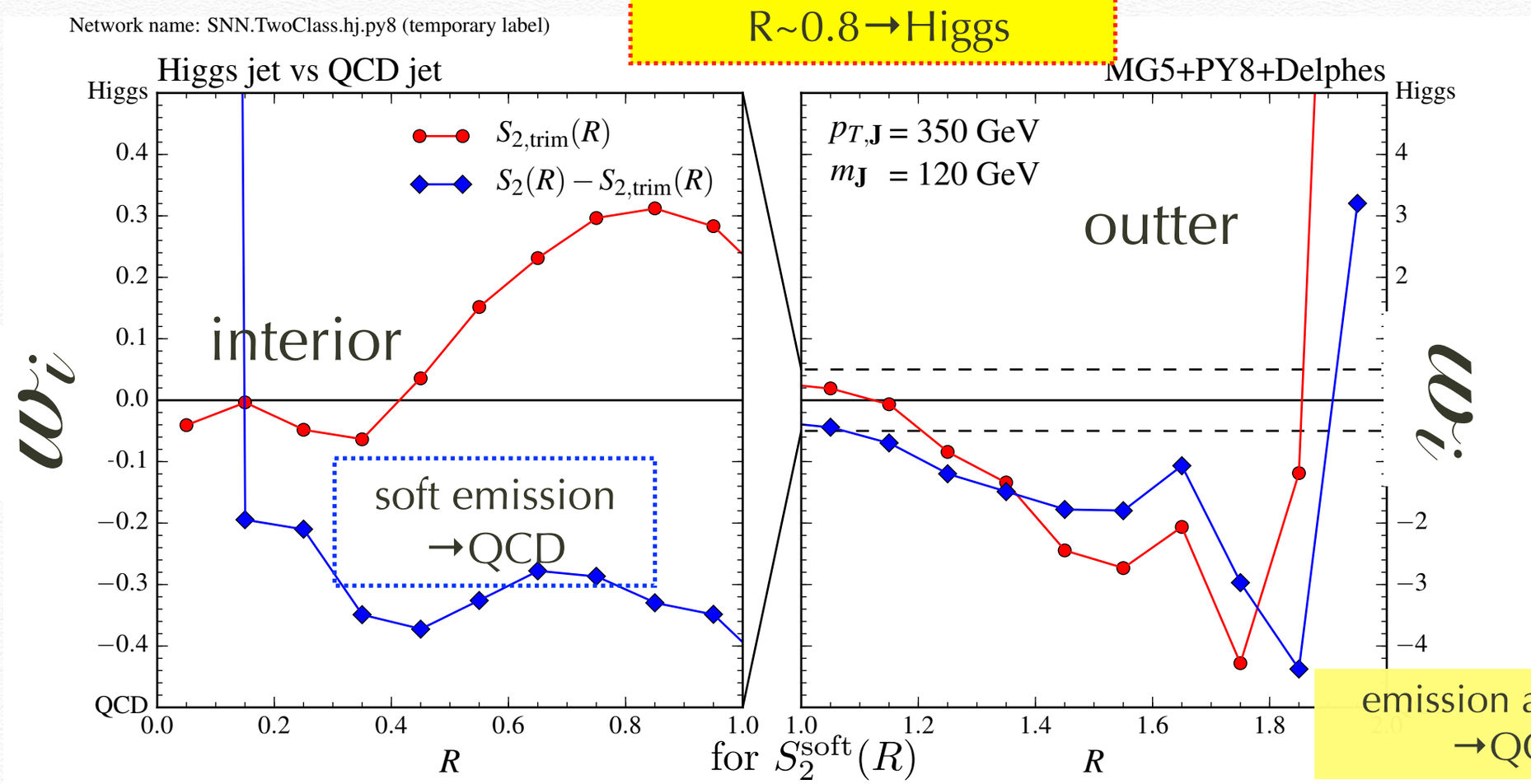
- inner product

$$h = \sum_k S^k_{2,\mathrm{trim}}\, w^k_1 + \sum_k S^k_{2,\mathrm{soft}}\, w^k_2, \quad \text{training}$$

coefficient w1 and w2 is instructed to depend on mass and momentum but not jet spectrum

QCD jet vs Higgs jet

trimmed jet emission at R~0.8→Higgs



emission at R> 1.0
→QCD

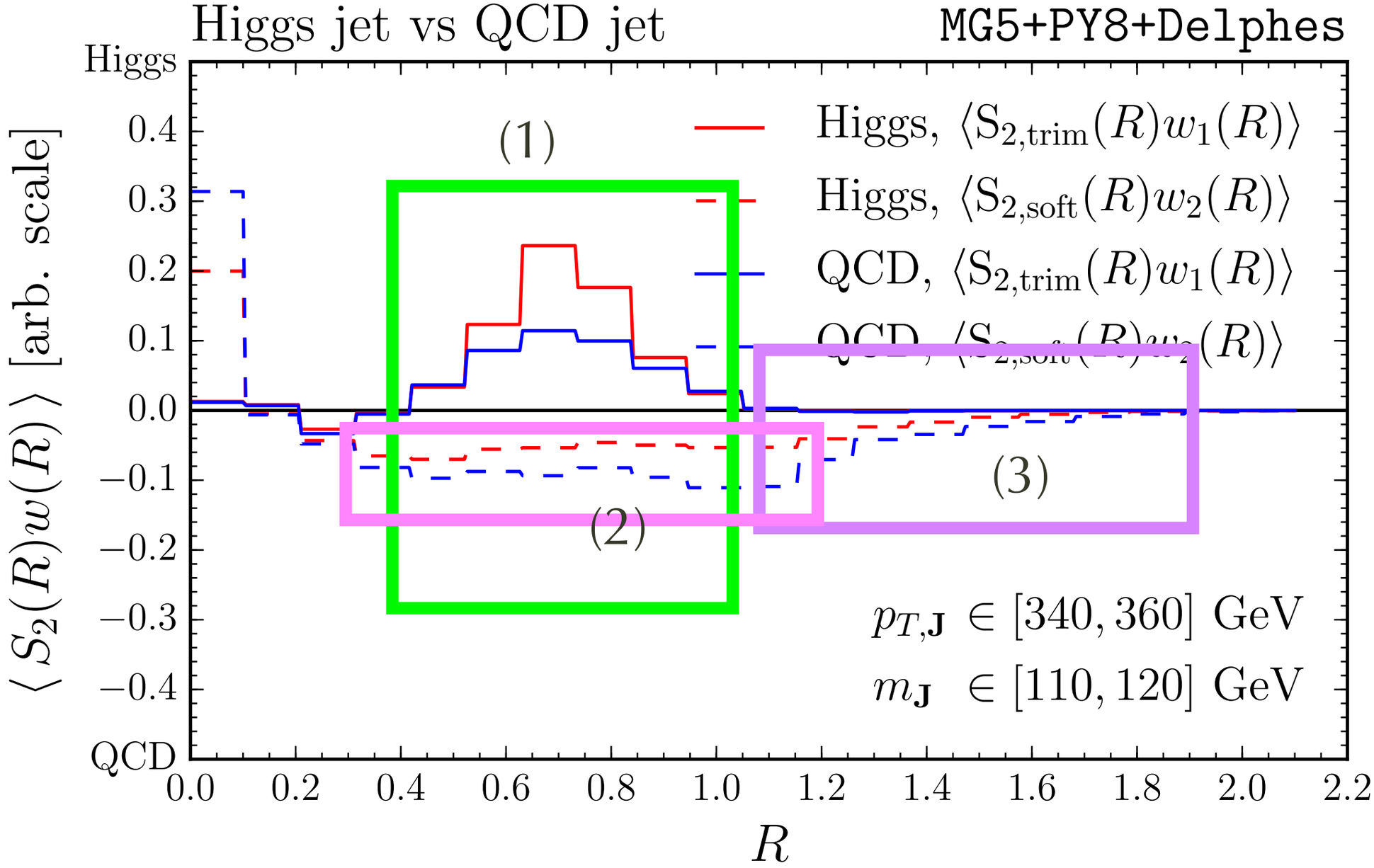
"Reasons" behind jet classification

two hard substructure making peaks in S_{2trim}

Large Hard-and-soft correlation

soft emission near jet boundary

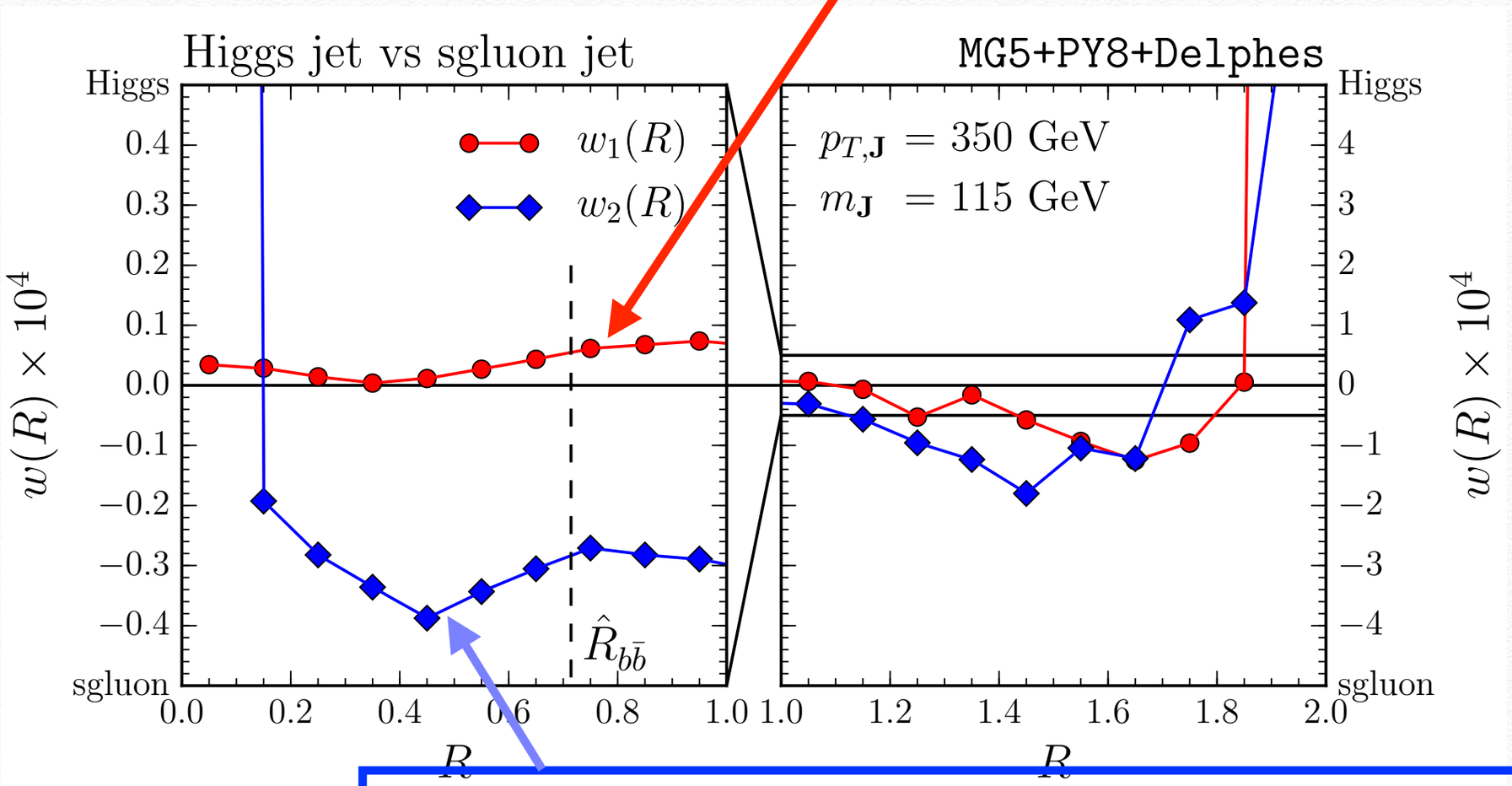
contribution to $\langle h \rangle = \langle S_2 w \rangle$



Color of heavy particles

- Higgs (siglet scalar) vs color octet particle decaying into 2b (same mass)

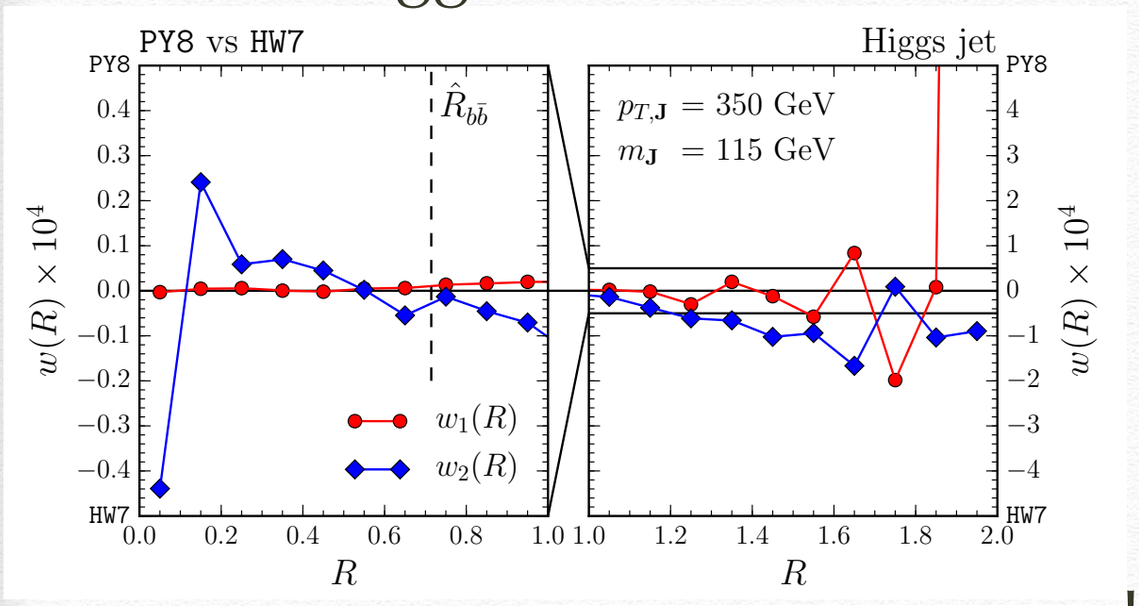
no mass peak for S_{2trim}
because the two masses are same.



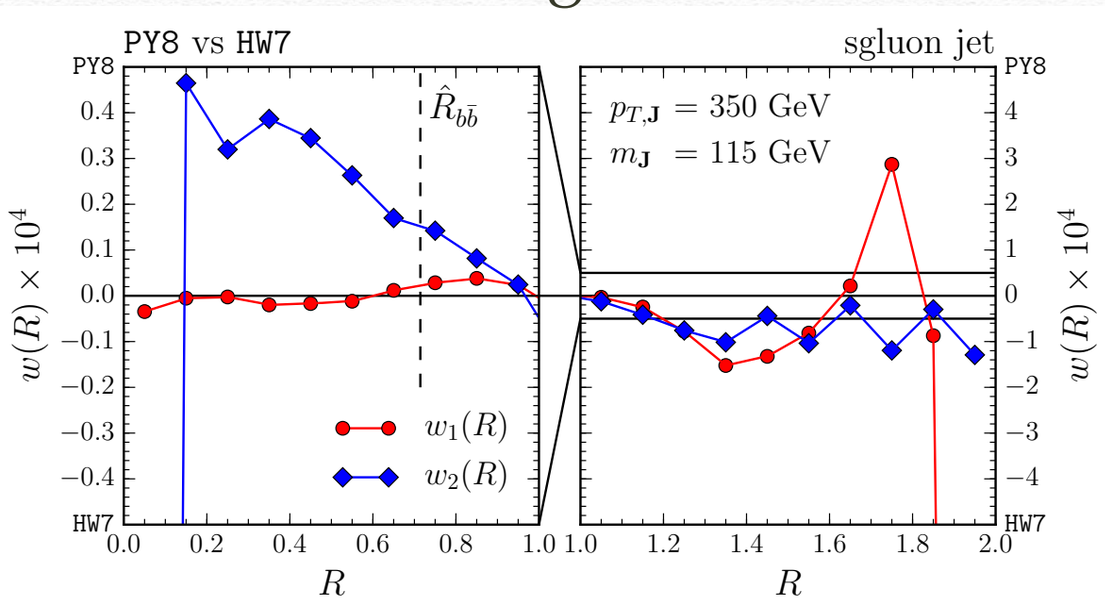
any soft activity outside
trimmed jet trigger octet interpretation

PY8 vs HW7 classification

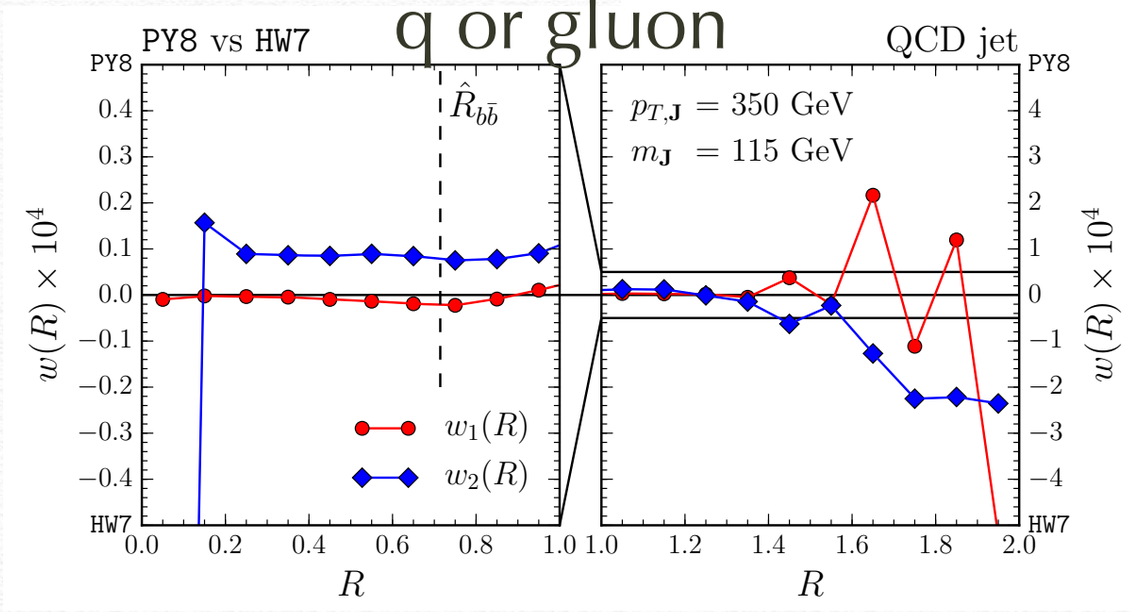
Higgs boson



Scalar gluon



q or gluon



* Herwig radiates more in large angle (angular ordering vs pT ordering ?)

* **Trimmed distribution is same (Hard(er) Physics is common)**

Summary

- We have extracted the core distribution relevant to ML H-j classification.
- **Input: reduced from NxN to N** (N~20 for our case.) We can use remaining numerical resource to understand whole events…
- **Modeling of soft emissions** turns out to be important.
- CNN using jet image is probably spending most of the time to find jet spectrum. (But not exact proof.)
- Nature of jets may be more interesting than parton shower, in that case something beyond jet spectrums is needed.
- Application to top process is on the way.