

# ジェット構造と 解釈可能な機械学習

Sung Hak Lim, Mihoko Nojiri (JHEP, 2018)

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

# 今後のLHC

統計は今後も増える（発見の物理から精密測定 of 物理へ）

標準模型のプロセスのずれを探す。

より厳しいカットをかけて、面白いイベントを探す

過去の成功体験（LEP） vs QCD 特有の難しさ

## 馬鹿でかい統計

hadronization,  
PDF, parton shower  
modeling, ...

high  $p_T$  objects (Events in Tail)  
soft object, mono something  
or something unknown



\* システムティクスへの理解が重要

\*\* 解析のスピードアップの必要

Physics outputs  
effective operator,  
top partner  
dark matter..

# 機械学習(ML)への期待

- ❖ 機械学習を使って

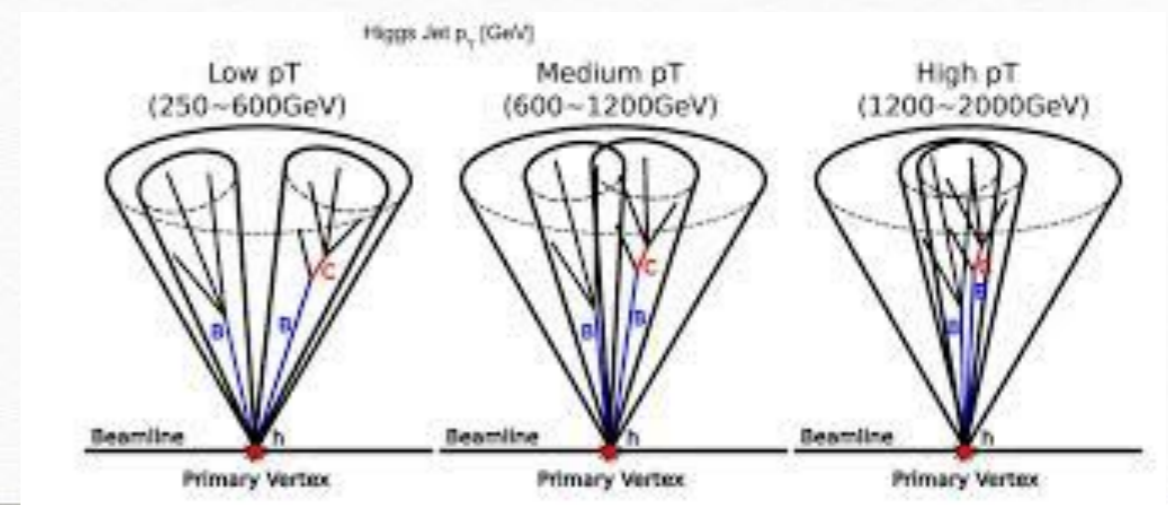
- ❖ イベントの再構成の効率化、高速化

- ❖ ジェットやイベントの特性を学習させて、top W H などのを取り出す

- ❖ 目的 ルミノシティに見合ったアウトプット QCD の不定性を実データを使って迂回（理解しないで済みます）あるいはソフトはプロセスの理解の促進

# Jets と機械学習

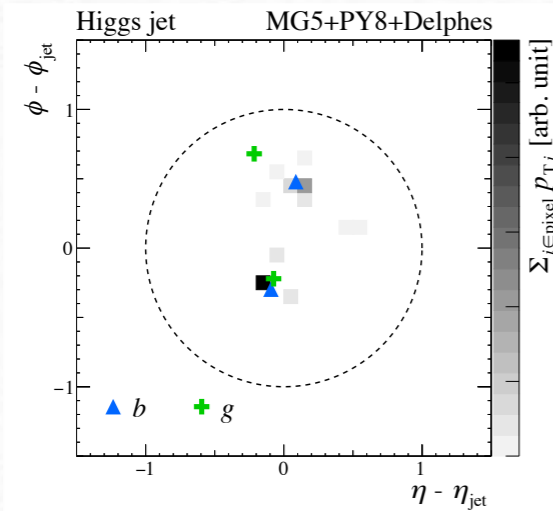
- ❖ Jet physics : QCDの中ではソフトな物理を含むにも変わらず比較的成功的している (Theoretical understanding + computation)
- ❖ 理論と現象論パートンシャワー→ハドロン
- ❖ 実験データとの比較 **theory and computational developments**
- ❖ Jet reconstruction algorithm ( kT, CA, antikT) IRC safe アルゴリズム
- ❖ numerical progress (**fastjet**  $N^3 \rightarrow N \log N$ )
- ❖ Jet substructure (mass drop),  $H_{i,s}$ , W, Top reconstruction (BSM search)
- ❖ minimal Validation Analysis (leaving optimization to algorithm) → **機械学習**



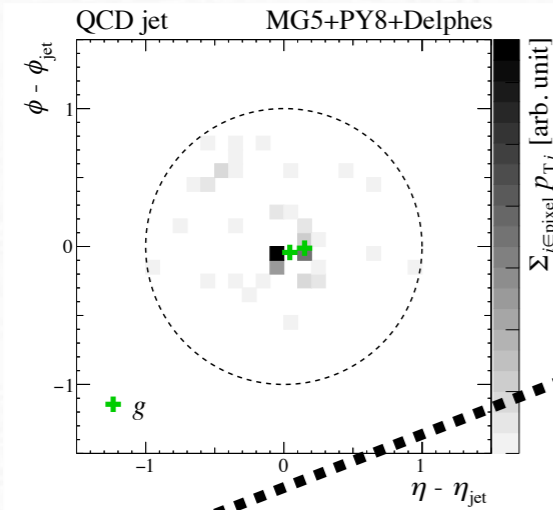
# 機械学習の基本要素

Input: Jet images

Higgs



QCD



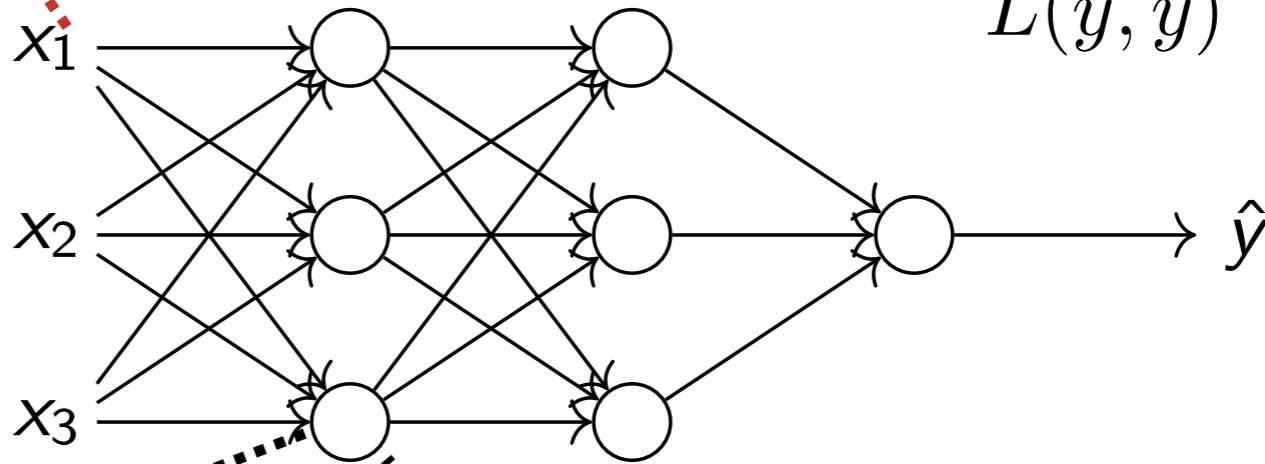
output:  $w_{ij}, b_i$

optimization

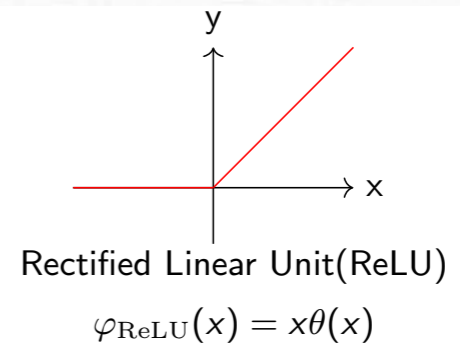
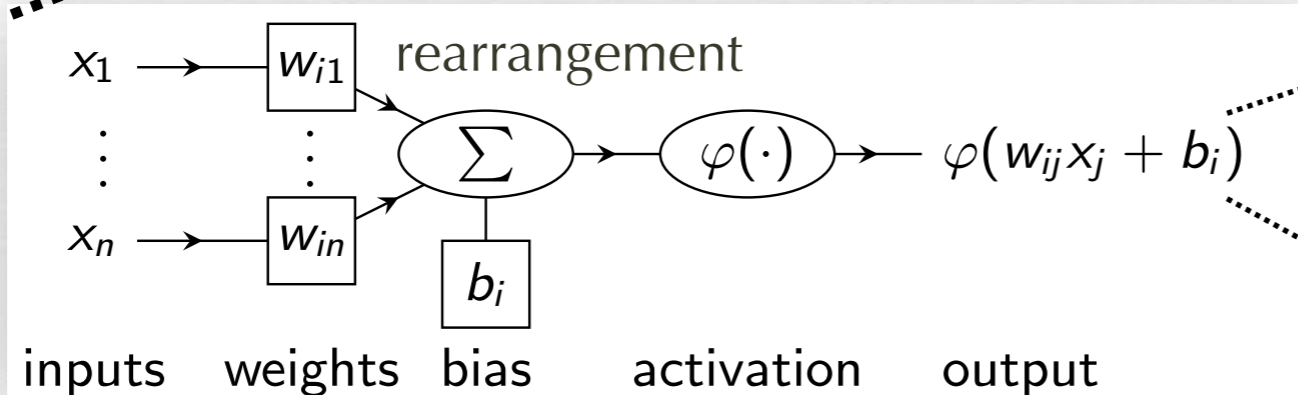
$$L(y, \hat{y})$$

Higgs  
 $y = (1, 0)$

QCD  
 $y = (0, 1)$



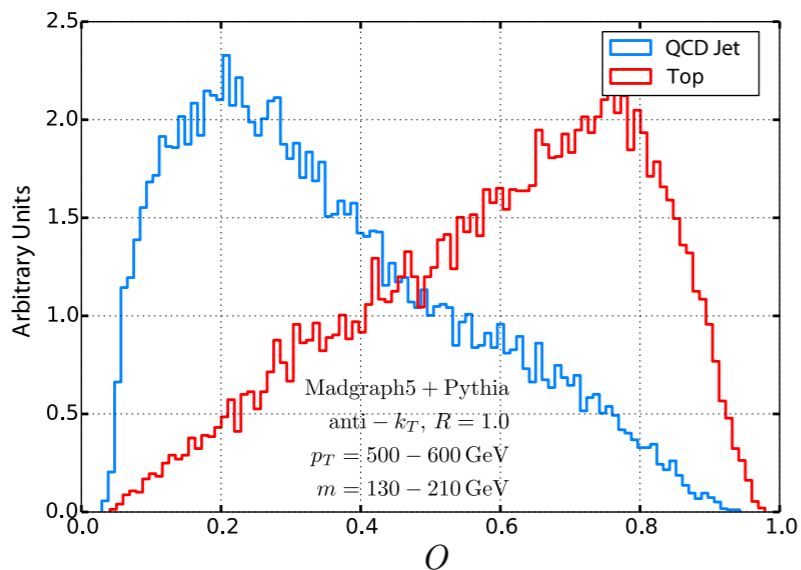
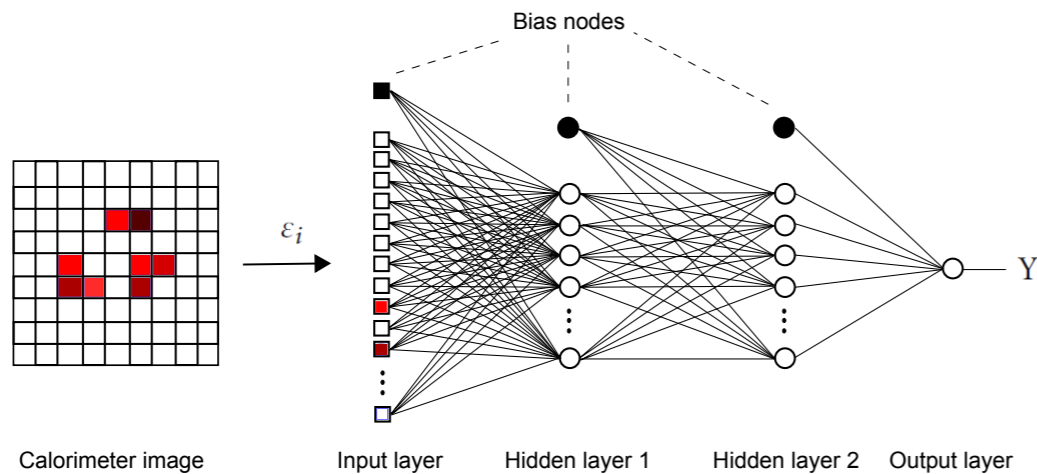
$\varphi$ : source of non linearity



# より“構造を持った入力”へ

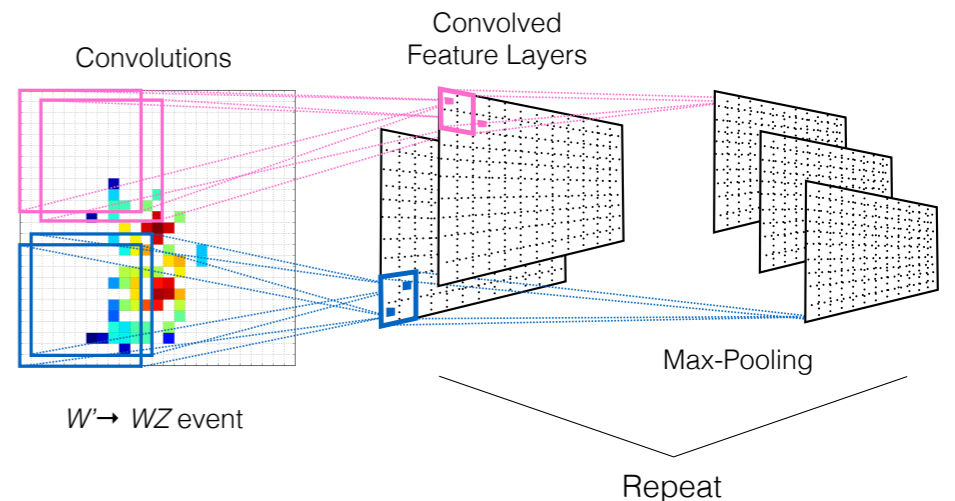
Almeida et al 1501.05968

DNN (all bins in a line)

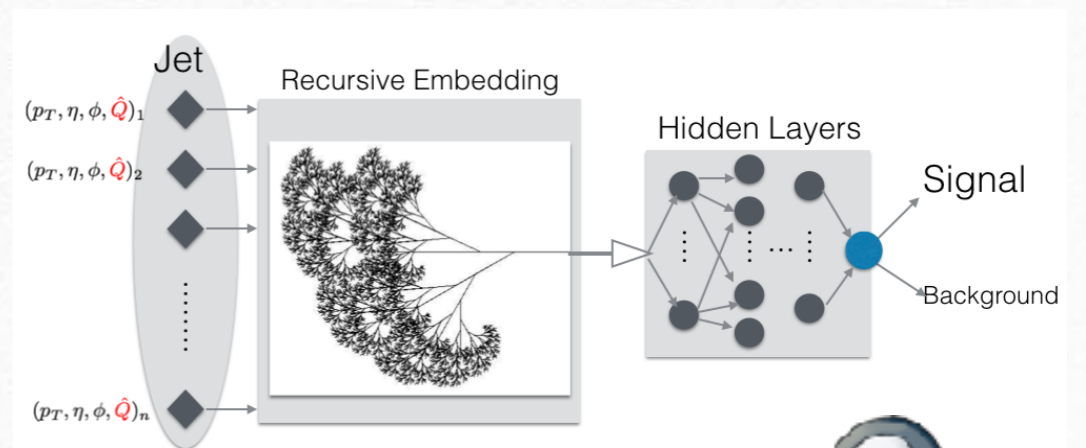


CNN

correct nearby image first 1511.05190



Recursive(Taoli Cheng 1711.02633)



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

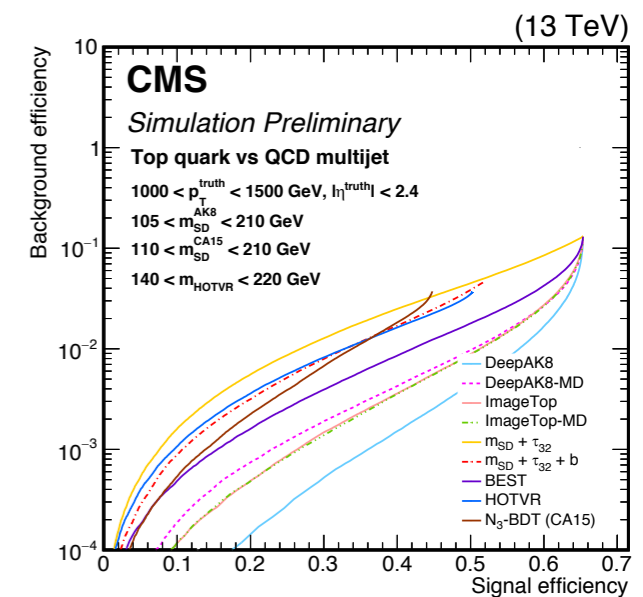
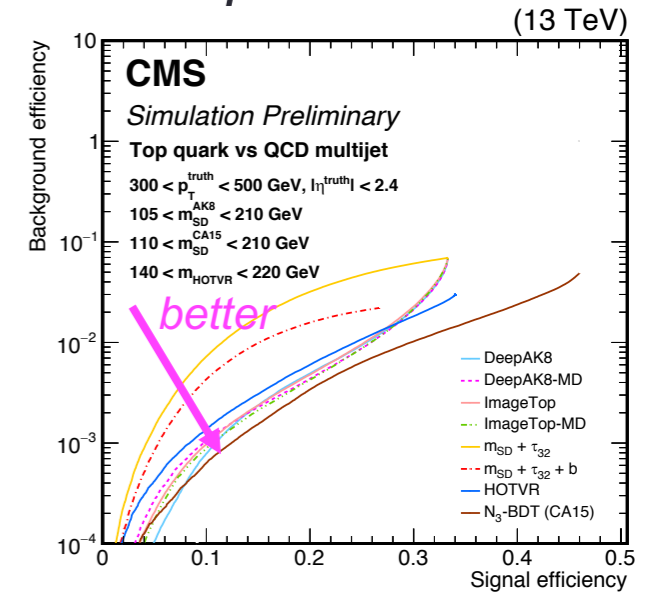




# Top Tagger performance in MC

- Performance comparisons in two different ranges of  $p_T(\text{truth})$
- $m_{SD} + \tau_{32}$ : addition of the subjet b tagging in the reduces the misidentification probability for t quarks by up to  $\sim 50\%$  (depends on  $p_T$ ).
- **HOTVR**: performance lies between  $m_{SD} + \tau_{32}$  and  $m_{SD} + \tau_{32} + b$ .
- **$N_3 - \text{BDT (CA15)}$** : shows improved performance compared to the aforementioned algorithms, particularly in the low  $p_T$  range.
  - The improved performance stems from the usage of the ECFs, which provide complementary information to  $\tau_{32}$ . In the low- $p_T$  region, the gain is mainly due to the use of larger-cone jets (i.e. jets clustered with  $R = 1.5$ ).
- **BEST**: targets the high- $p_T$  regime and shows similar performance to the ECF algorithm in this regime.
- The best discrimination is achieved with algorithms based on lower level information: **ImageTop and DeepAK8** algorithms.
  - ImageTop & DeepAK8-MD yield comparable performance in low & high  $p_T$  regions
- The optimal performance in terms of ROC curves is achieved with the nominal version of DeepAK8 over the entire  $p_T$  region.

## top vs QCD



Significant improvement from the new developments 5

# 「機械学習が何をやってるか知りたい」

- ❖ ML だとなぜ良い結果が得られるか。
- ❖ Hidden layerは何をしているか
- ❖ QCD との関係

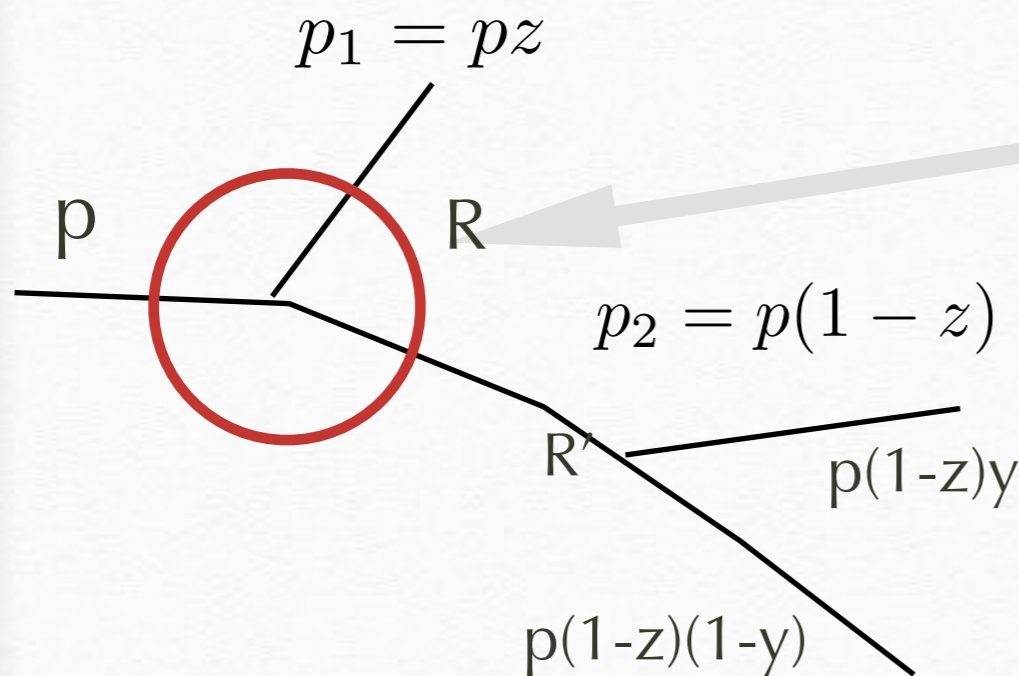
## やったことのまとめ

1. jet image(400 inputs) CNNはやめにして、もう少し**意味のある量** (~40) (jet spectrum) の DNNをやった
2. CNN classifier vs DNN **They are Equivalent!**
3. さらに“Interpretable model”を作って、何が使われているかみる。



# Input を絞る (Jet spectrum)

❖ Monte Carlo → Parton splitting + hadronization



$(p, R, z)$  describe  
parton shower splitting

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

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

Our minimum input (Jet Spectrum)

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

sum over pairs of jet constituents

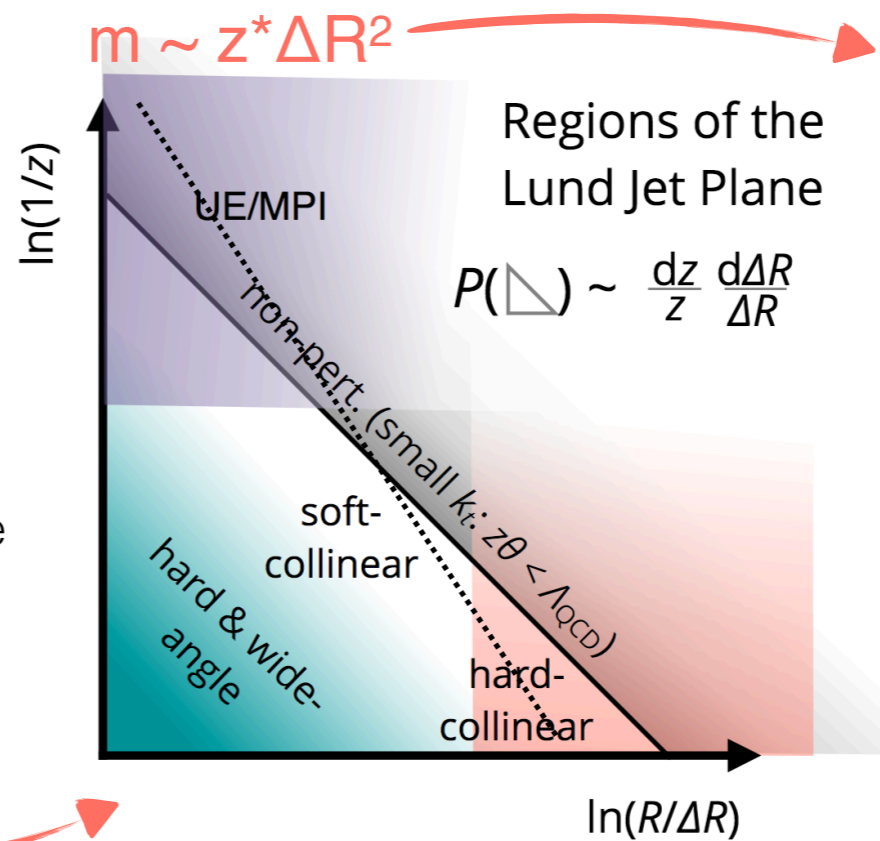
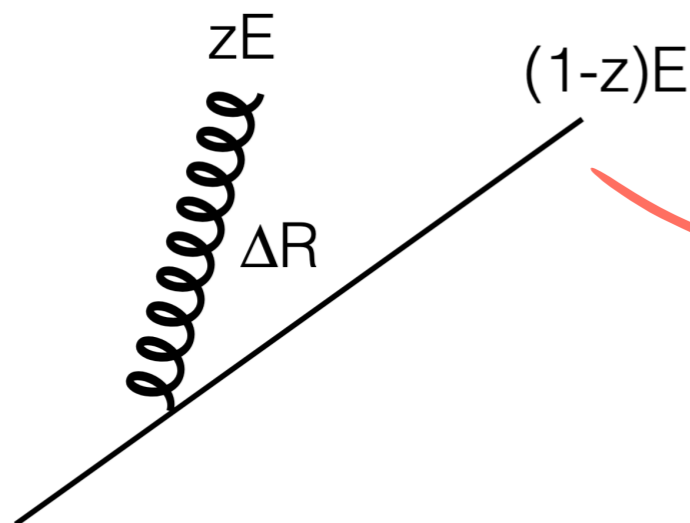
proportional to momenta of particles to be IRC safe (C-correlator)

Jet Spectrum ~ Energy Energy correlation, C-correlator (1996 Tkachov)  
 (IRC safe な要件で作った量 比較的起源の古い概念)

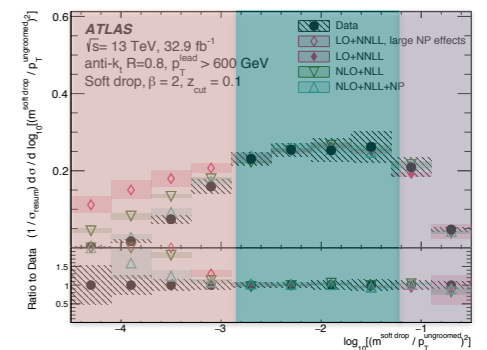
最近の Lund Plane とも近い (Dreyer, Salam, Soyez, JHEP 12(2018)064)  
 ( CA jet clustering のヒストリーを 2次元面に射影)

# The Lund Plane

- ▶ A jet may be approximated as soft emissions around a hard core which represents the originating quark or gluon
- ▶ Emissions may be characterized by
  - ▶  $z$  = relative momentum of emission wrt jet core
  - ▶  $\Delta R$  = angle of emission relative to the jet core



The jet mass is just one diagonal line in this space



So what if we could measure the whole thing?

The Lund Plane is the phase space of these emissions: it naturally factorises perturbative and non-perturbative effects, UE/MPI, etc.

# jet image を使った分類との関係

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

\*classifier using energy flow

$$h_i = \hat{\Psi}_i[P_T]$$

calorimeter hit position

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

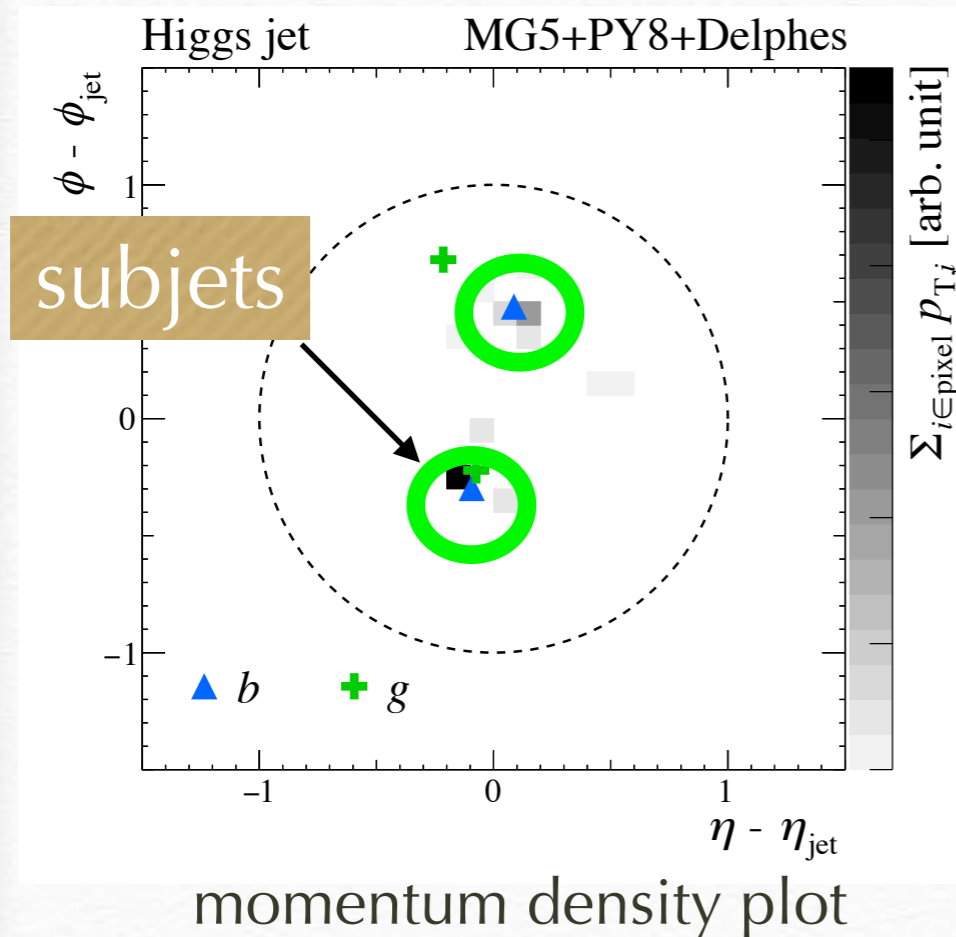
If w only depends on  $R_{12}$

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

\*classifier using Jet spectrum

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

# Soft と hard の activity を区別する



jet ~ subjects (IRC safe)  
 subjects above certain pt cut  
 => trimmed jet

$$\mathbf{J}_{\text{trim}} = \bigcup_a \mathbf{J}_a \cdot \frac{p_{T, \mathbf{J}_a}}{p_{T, \mathbf{J}}} \geq f_{\text{trim}}$$

ほんとは R に寄った  
 pt カットにした方が良かった

$S_{2\text{trim}}$

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

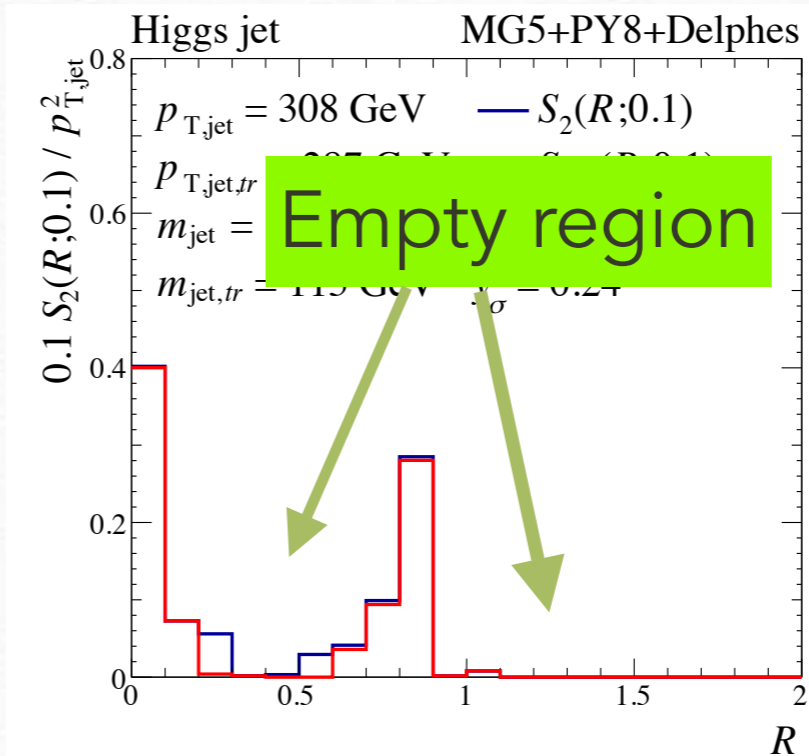
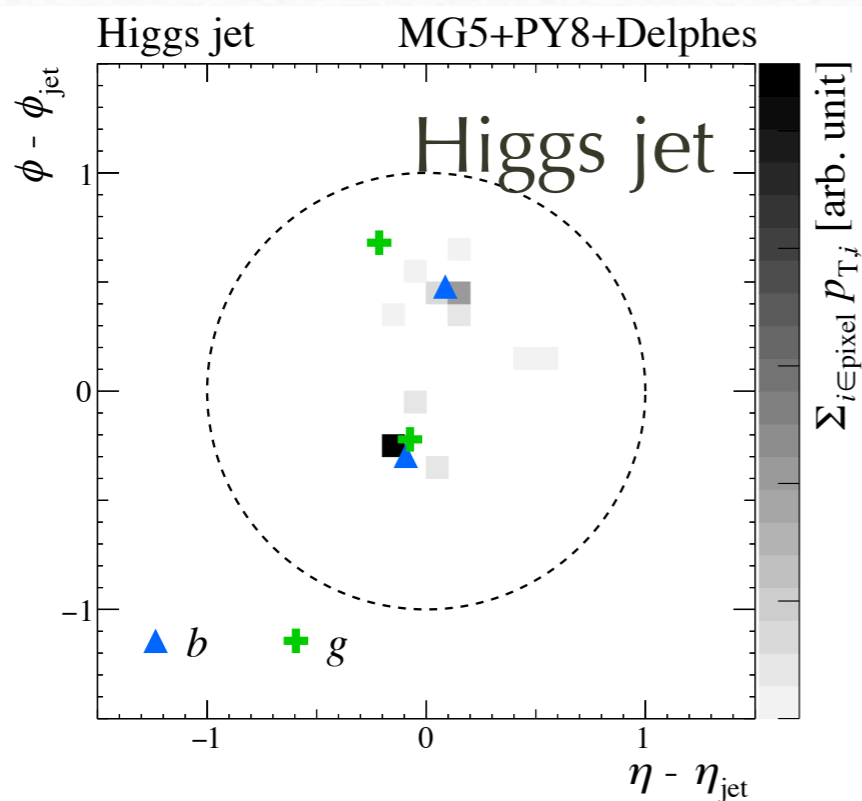
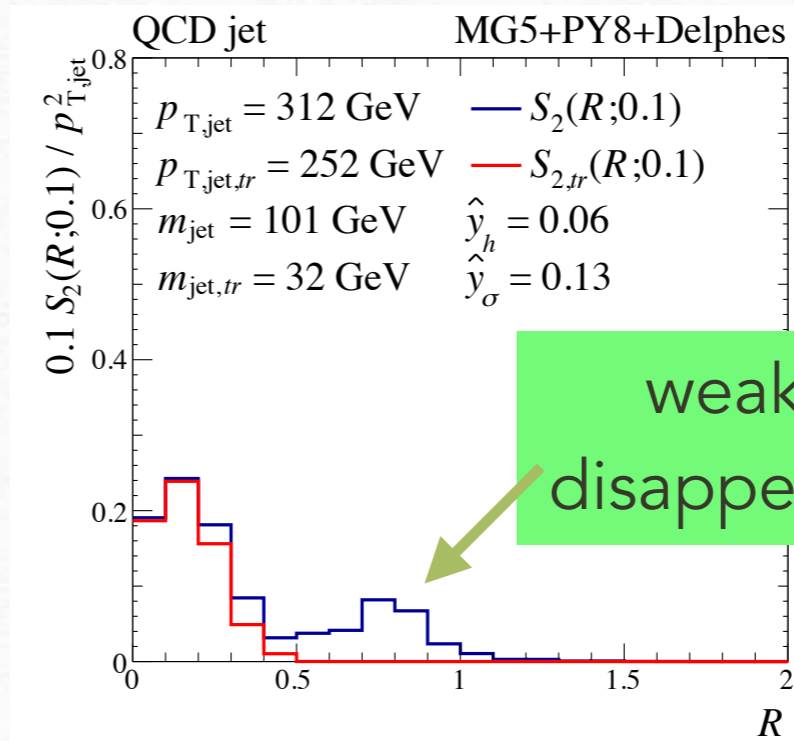
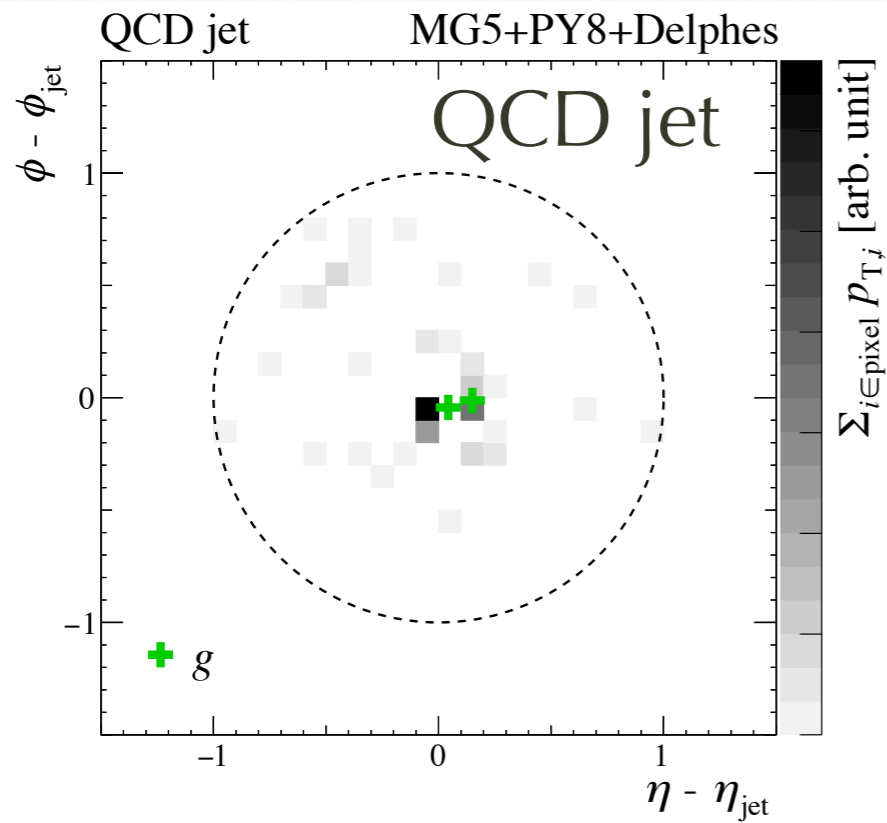
= "hard-hard" correlation

$S_{2\text{soft}}$

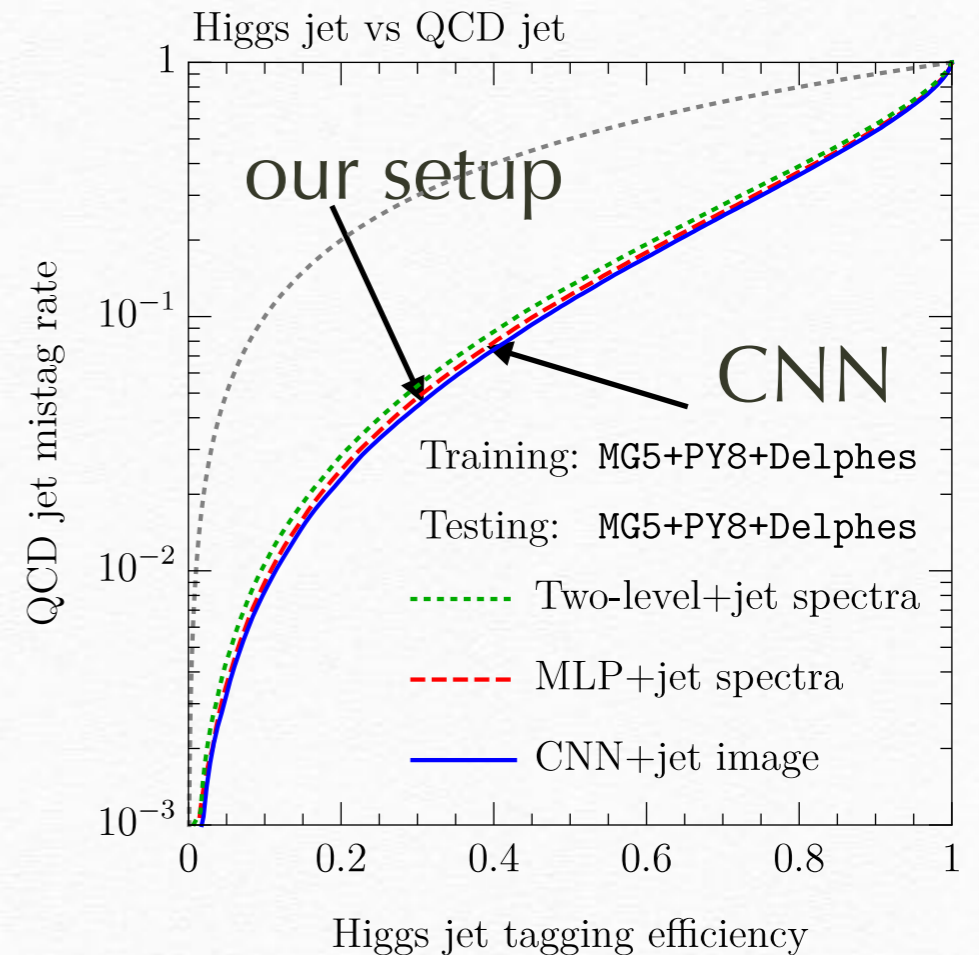
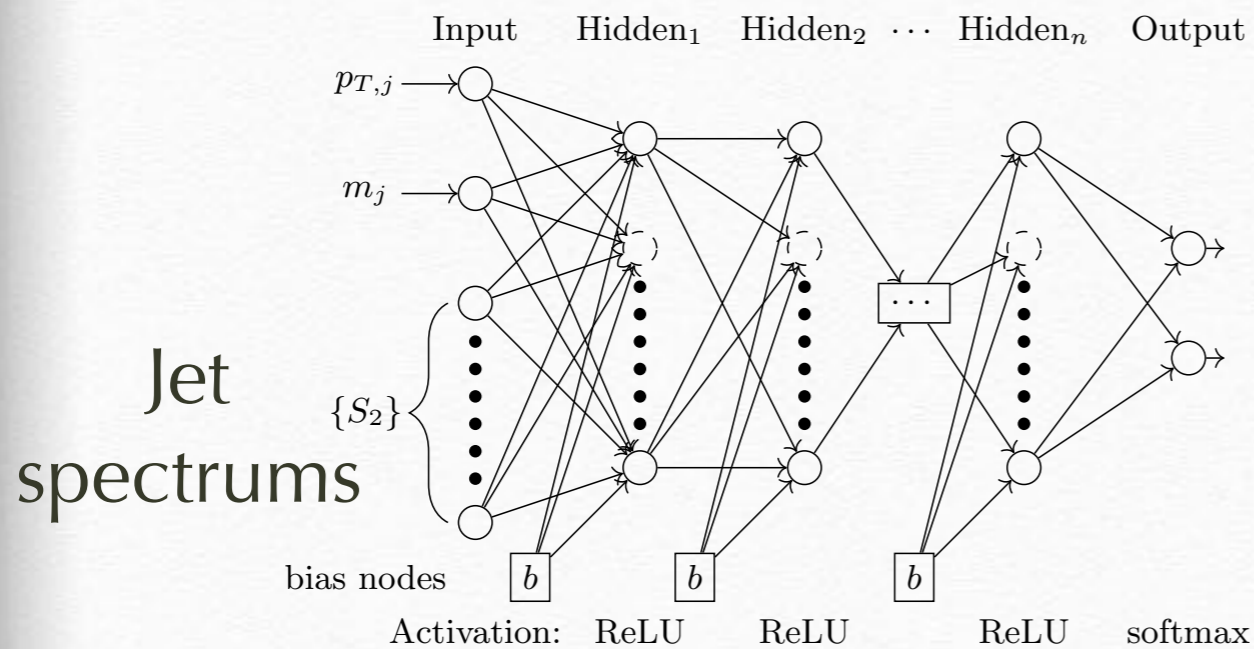
$$S_{2, \text{soft}}(R; \Delta R) = S_2(R; \Delta R) - S_{2, \text{trim}}(R; \Delta R).$$

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

# Typical $S_2(R)$ distribution



# CNN vs DNN with $S_{2\text{trim}}(R)$ and $S_{2\text{soft}}(R)$



for  $300\text{GeV} < p_T < 400\text{GeV}$  and  $100\text{GeV} < m_j < 150\text{GeV}$

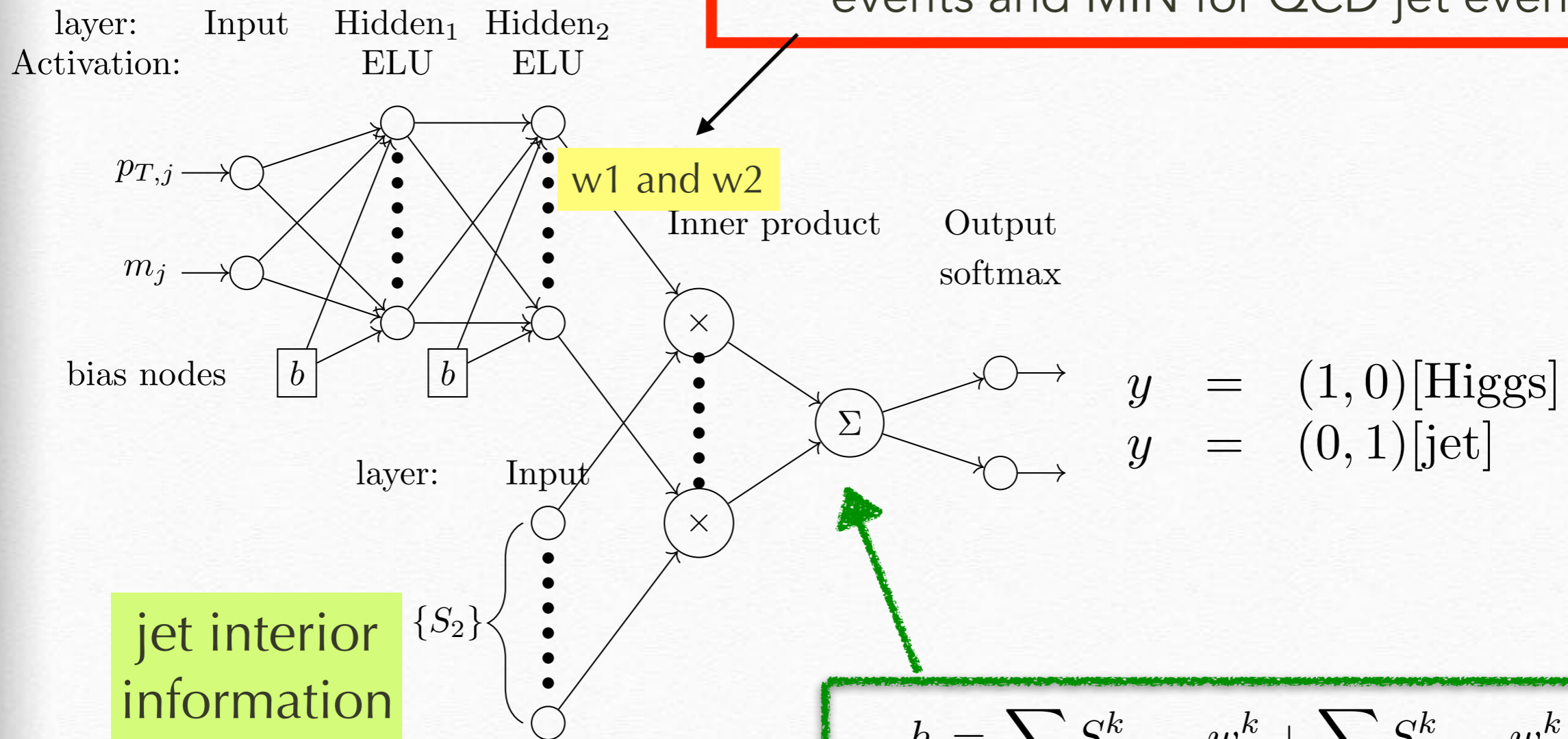
The results only slightly worse than CNN

- ❖ 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]

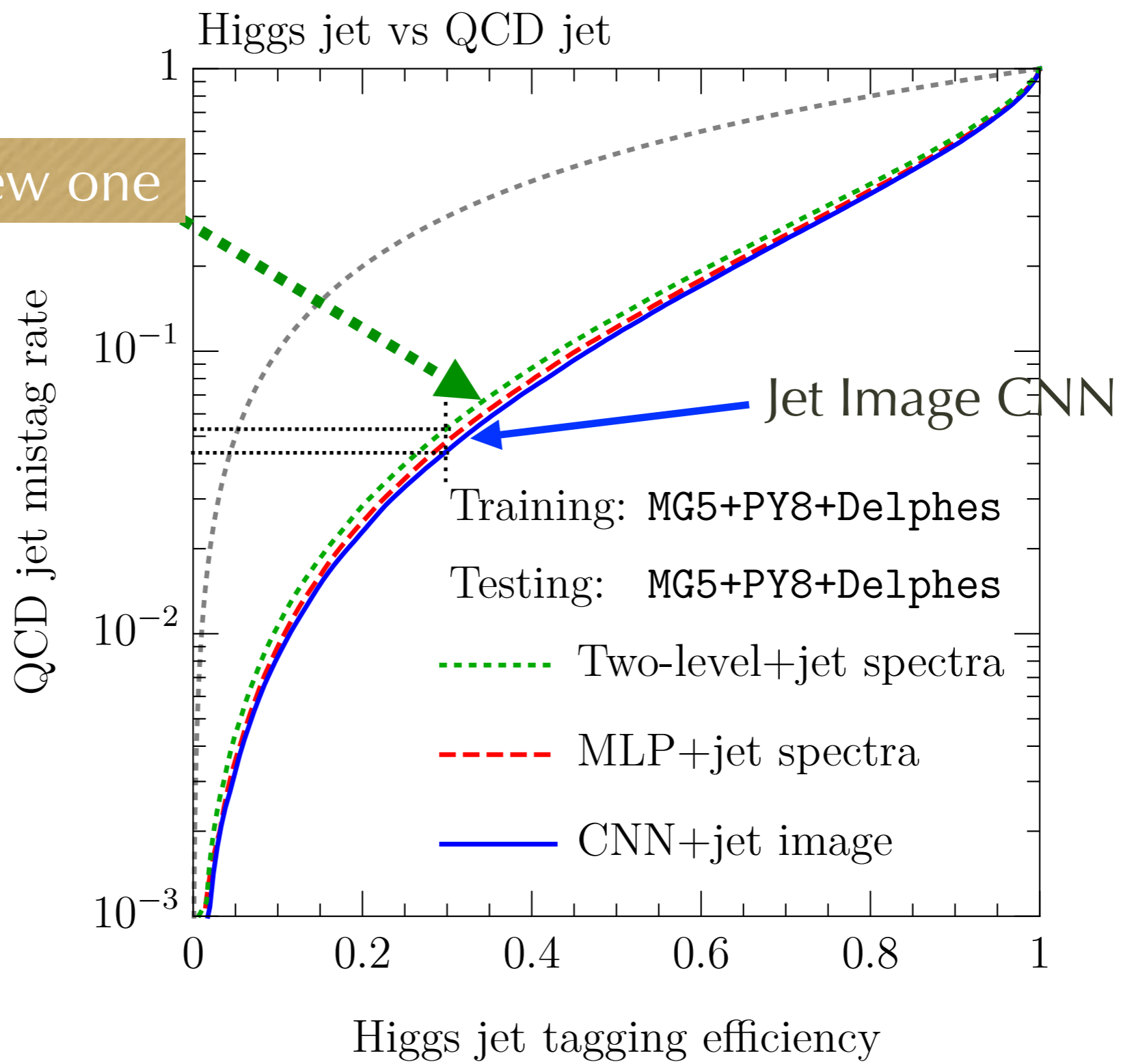
w1 and w2 are trained to be MAX for Higgs events and MIN for QCD jet events



$$h = \sum_k S_{2,\text{trim}}^k w_1^k + \sum_k S_{2,\text{soft}}^k w_2^k,$$

This simple classifier performs nearly as good as previous ones 15

New one





# kinetic module vs radiation module with readout

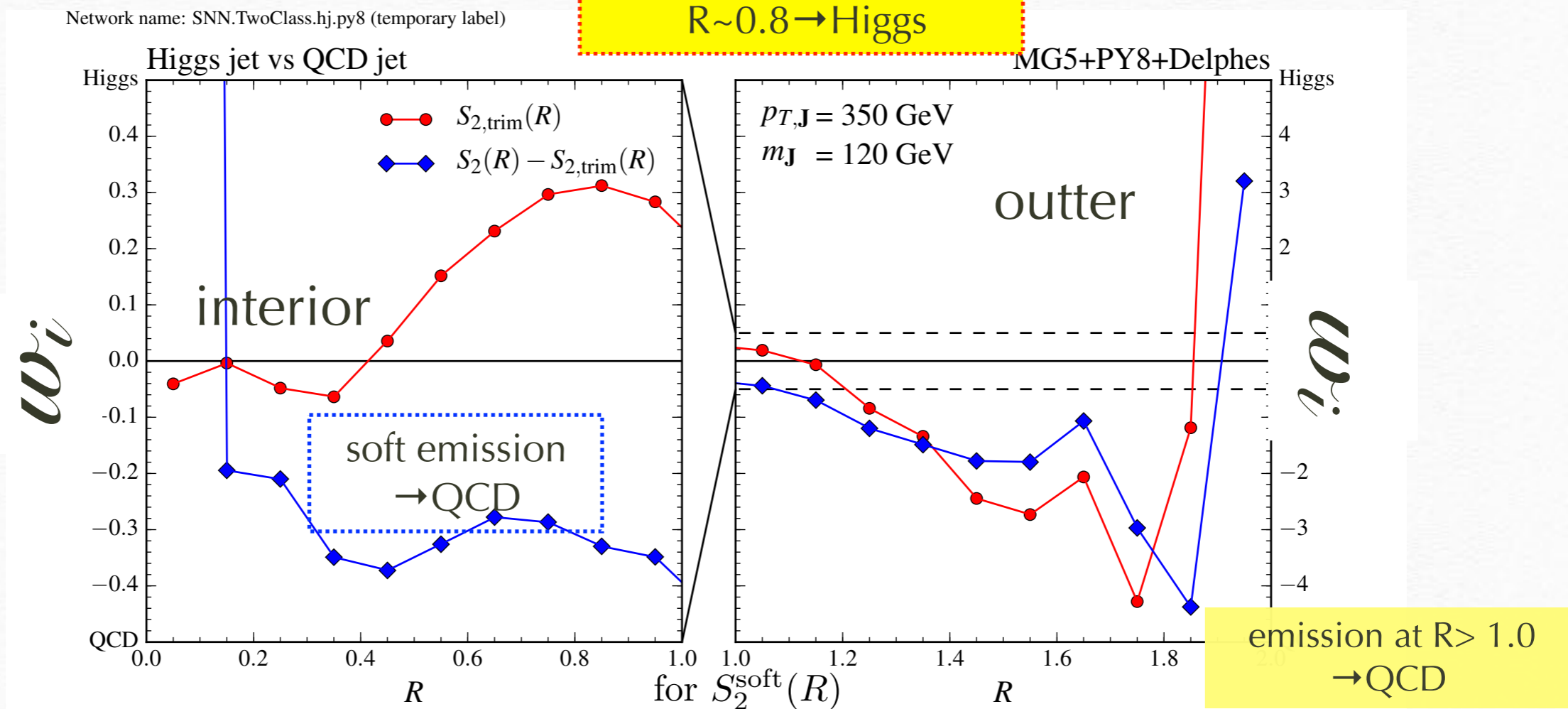
❖ inner product

$$h = \sum_k S_{2,\text{trim}}^k w_1^k + \sum_k S_{2,\text{soft}}^k w_2^k, \text{ training} \rightarrow$$

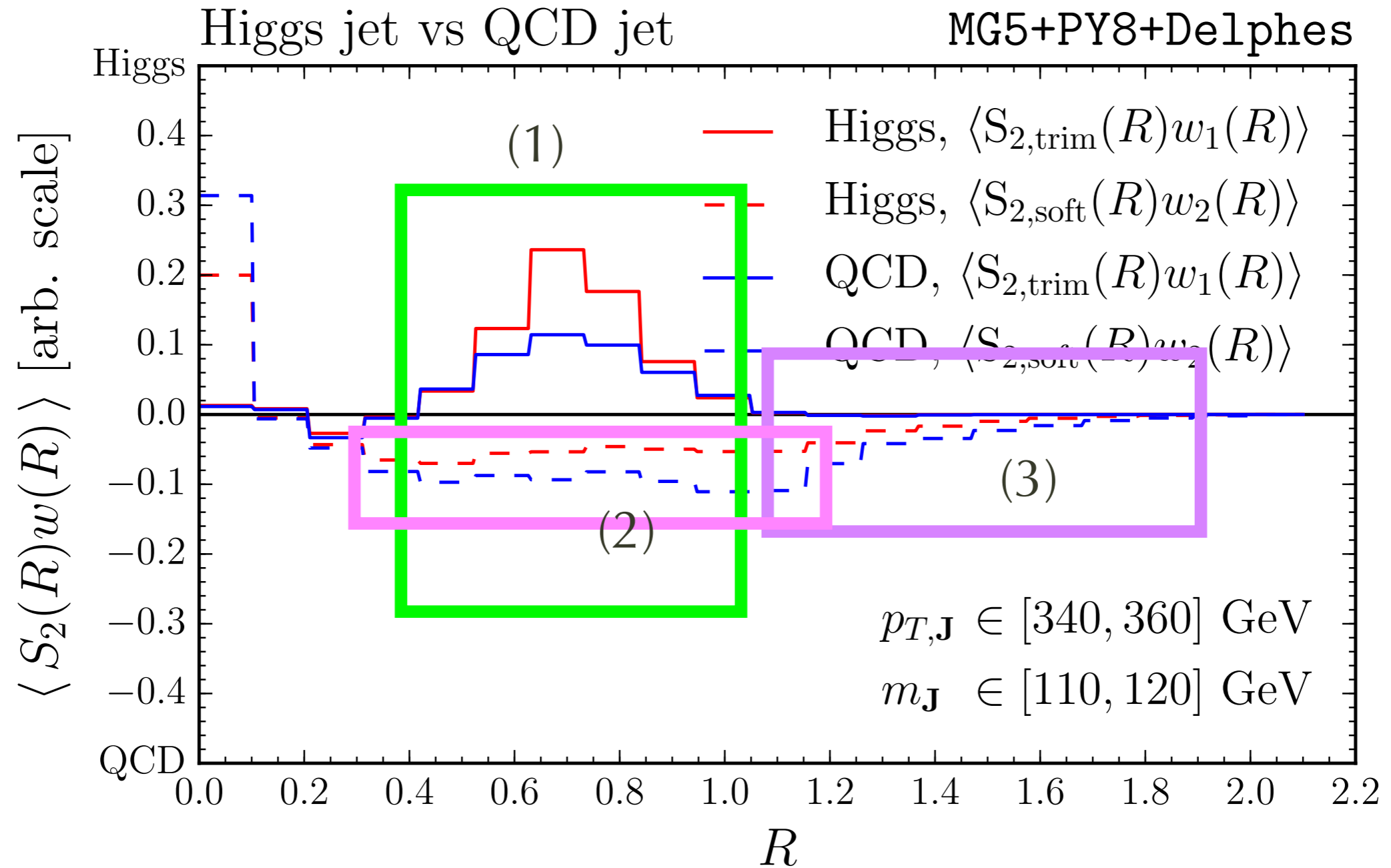
coefficient  $w_1$  and  $w_2$  is instructed to depend on mass and momentum but not jet spectrum

QCD jet vs Higgs jet

trimmed jet emission at  $R \sim 0.8 \rightarrow$  Higgs



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



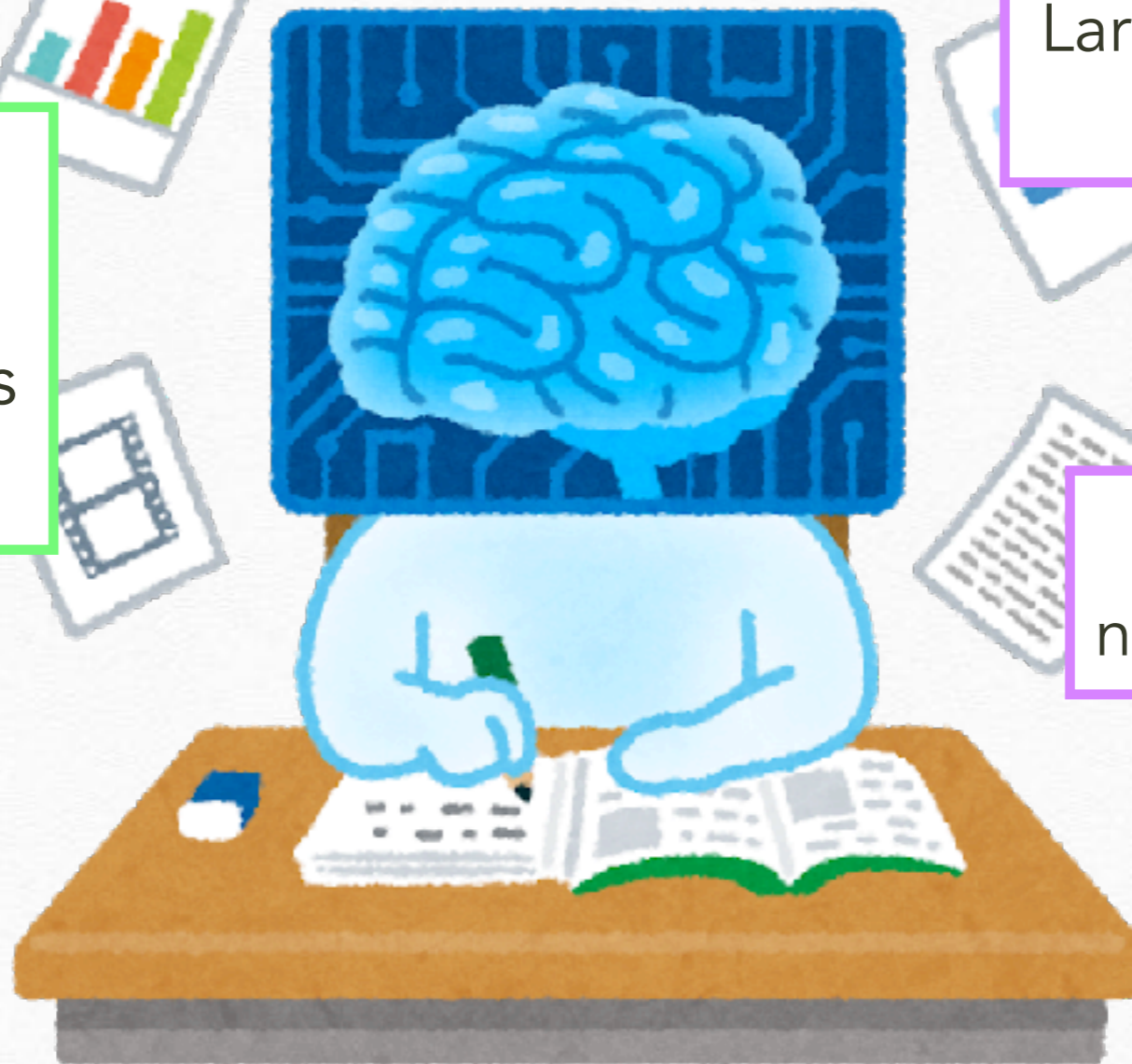
# "Reasons" behind jet classification



two hard  
substructure  
making peaks  
in  $S_{2\text{trim}}$

Large Hard-and-soft  
correlation

soft emission  
near jet boundary

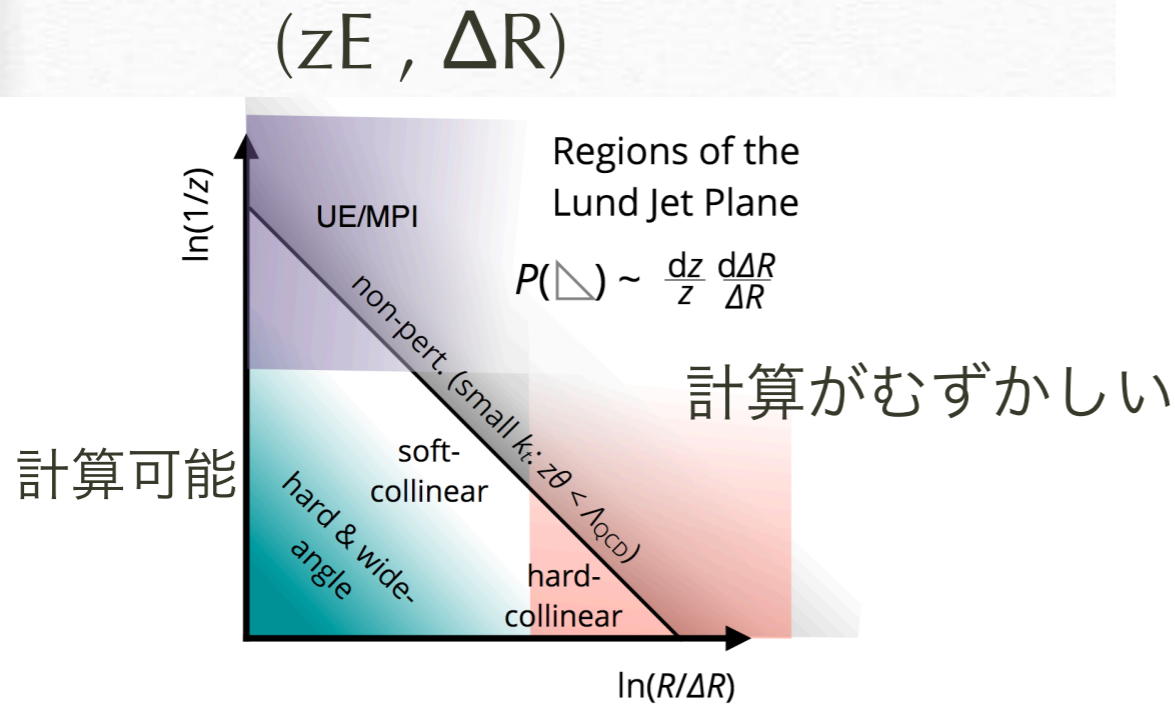


# MCは「正しい」か

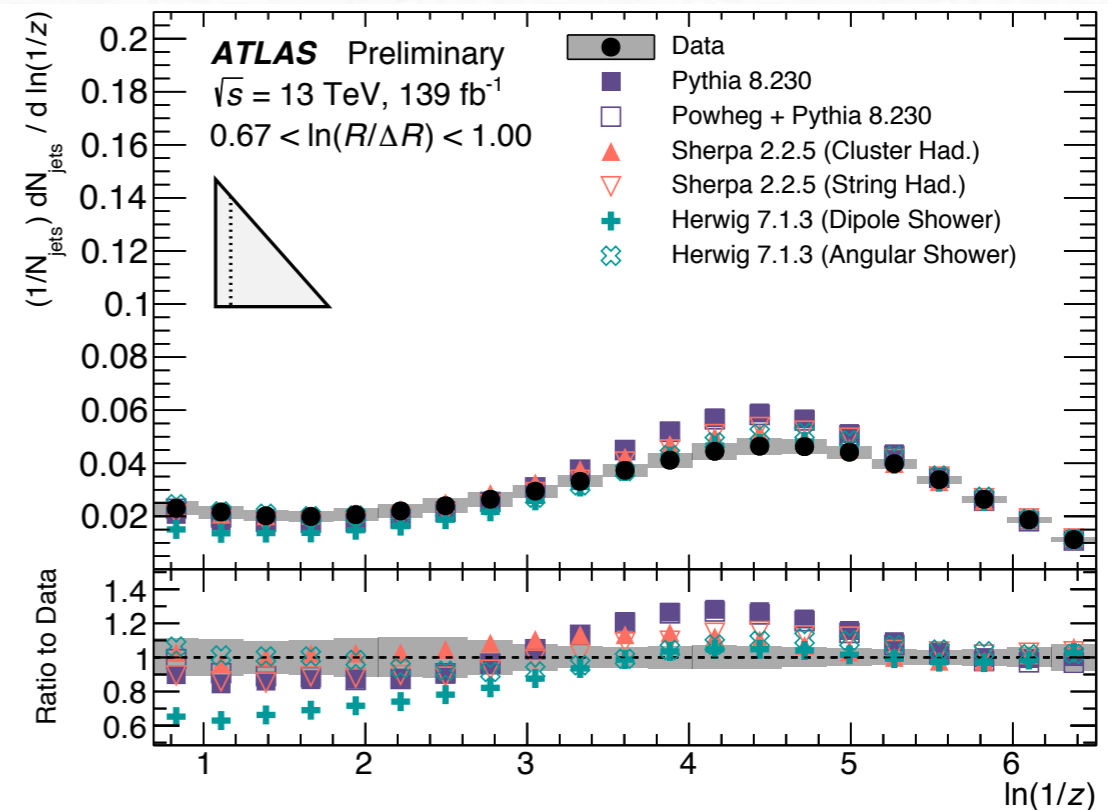
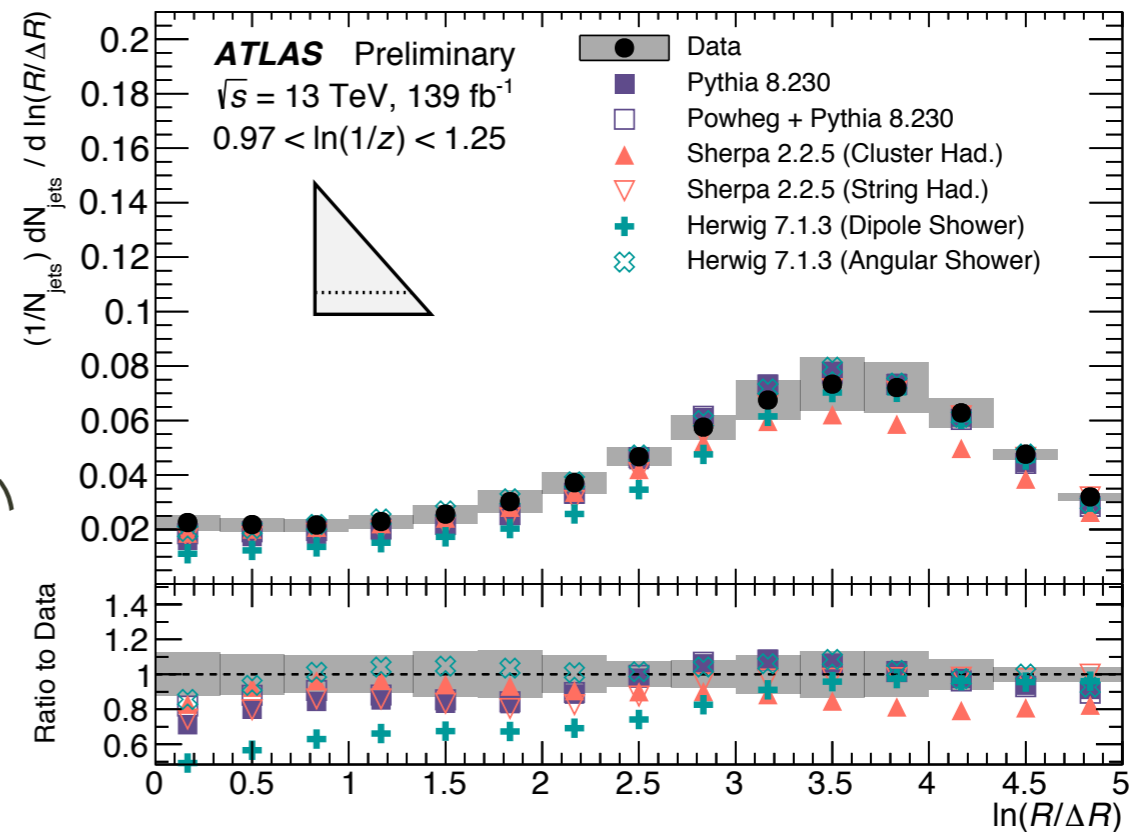
- ❖ quark → **パートンシャワー** → たくさんの quark gluon ( $Q_{\text{start}}, Q_{\text{hadron}}, \alpha_s$ )
  - ❖ **Herwig** angular ordering (より大きい角度から小さい角度に向かって放出が起こる。メリット angular ordering が保証されている)
  - ❖ **Pythia, Sherpa** pT ordering (よりkT の大きい放出から小さい放出が順次起こる)
  - ❖ options Dipole VINCIA (より複雑なカーネル)
- ❖ **パートンシャワーからハドロンへ ( $Q_{\text{hadron}}$ )**
- ❖ クラスタモデル (Herwig と Sherpa) parton shower 中の qqbar から近いものを見つけてペアを作り崩壊させる。
  - ❖ String 模型 (Pythia) color connection がちぎれて Meson ができる。
- ❖ **parton shower + hadronization** のパラメーターがデータで調整されたものが、使われている現象論的な量。 (Matrix element といった精度が評価できる量と異なる。
- ❖ できるだけ、resumして評価できるところは抑えて、その範疇でjetの物理をやりたいというのが今の流れ

# MC は現象をうまく記述しているか?

Lund jet Plane (Jet core からの分布)



- ▶ Non-trivial differences between different generators and unfolded data
- ▶ Region dominated by hard and wide-angle splitting is affected by parton shower
- ▶ Only small effects seen from UE/MPI (as expected)

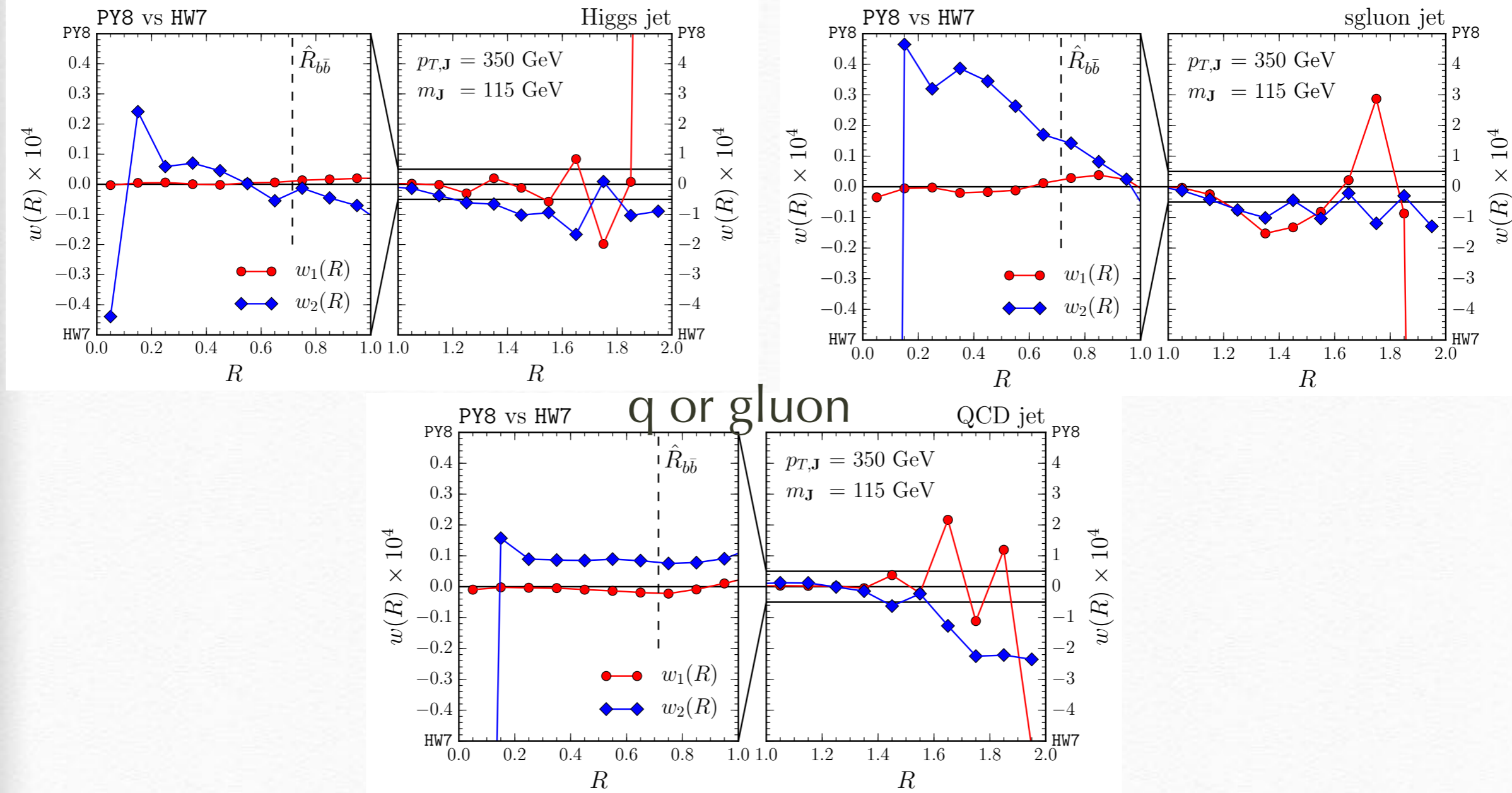


from Boost 2019 talk by Jennifer Roloff

# PY8 vs HW7 はどこが違うか

Higgs boson

Scalar gluon

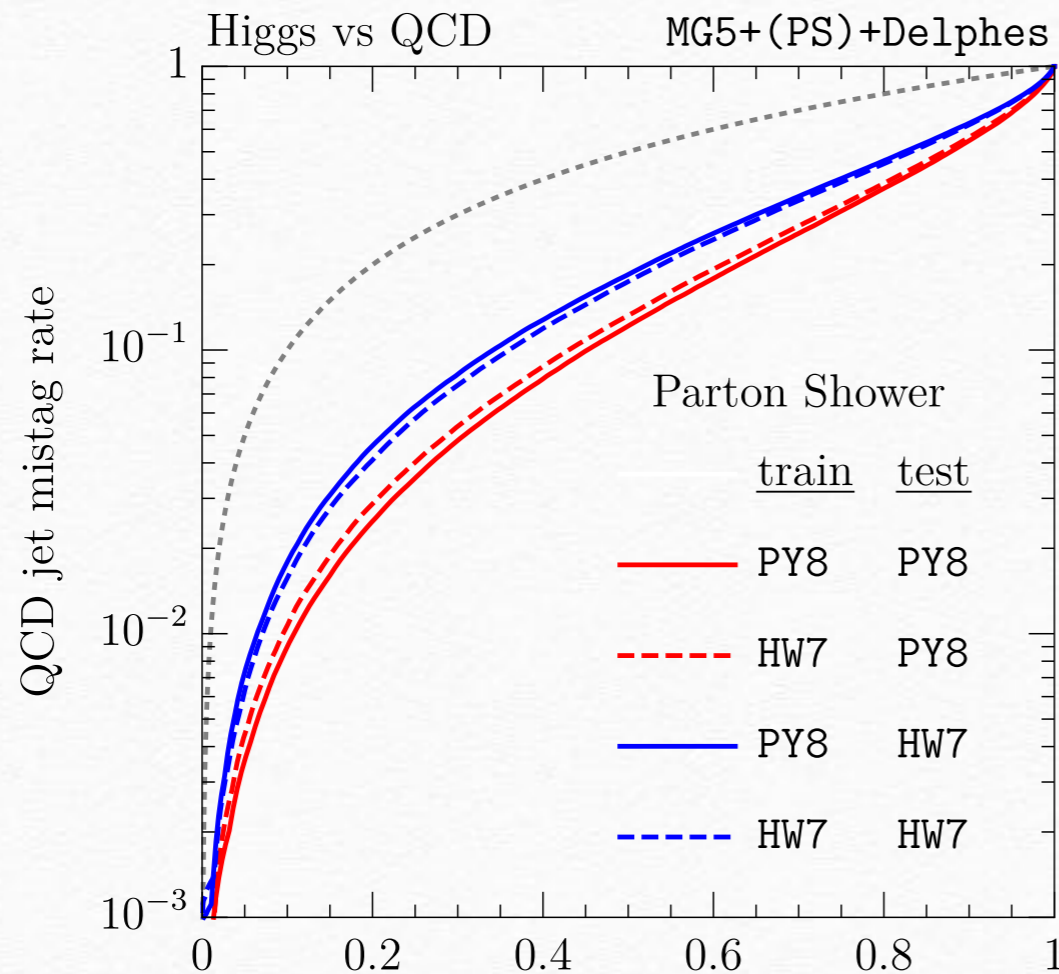


- \* 違うMCで同じプロセスを作って、機械学習で分類させてみる。
- \* ハードなサブジェットの分布は同じ。
- \* ソフト部分は違う。(そしてそれにかかなり依存して分類をやっている。)

# モンテカルロ依存性

color singlet or octet vs QCD

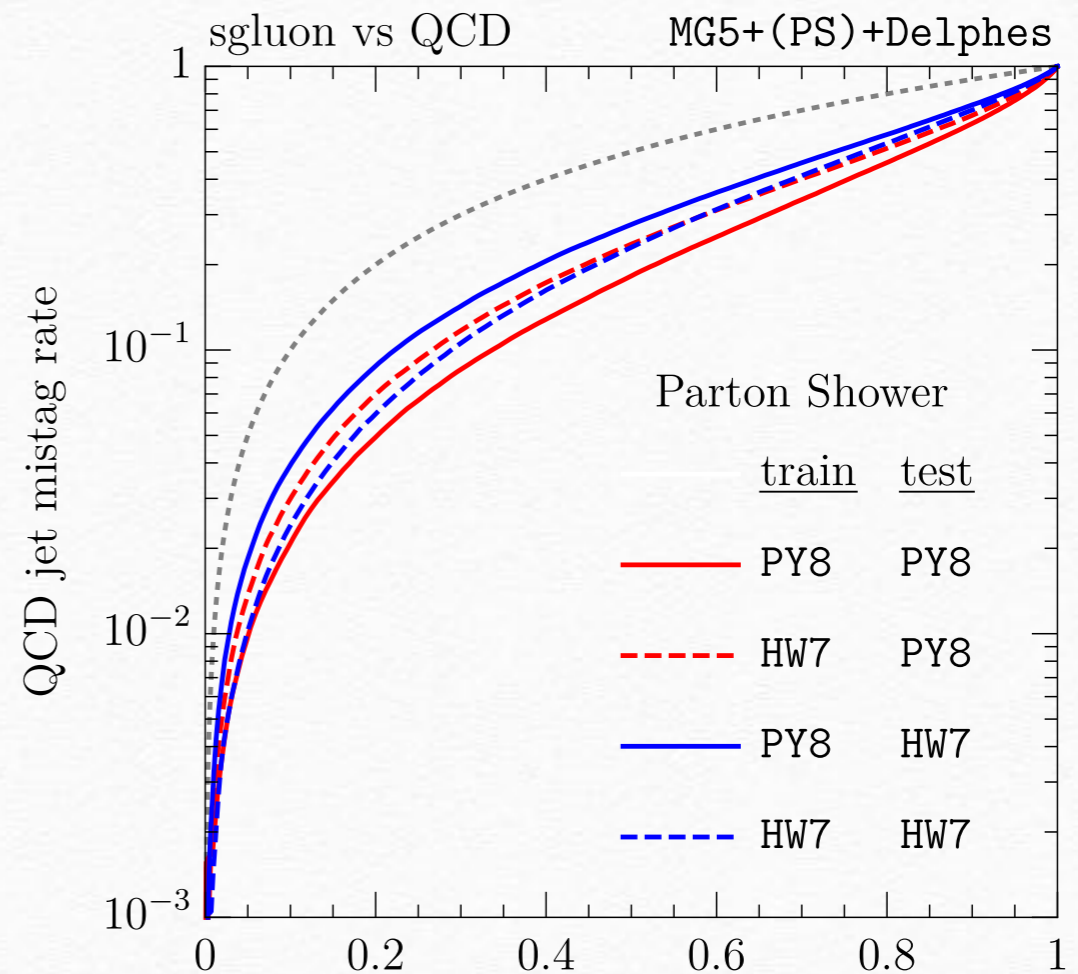
## singlet (Higgs)



Higgs jet tagging efficiency

MCによらず機械学習が注目している分布が同じ

## octet(scalar gluon)



sgluon jet tagging efficiency

機械学習が注目する特徴すら同じでない場合

# まとめ

- ❖ ML H-QCD ジェット分類で[実際に参照されている分布]を同定した。
- ❖ **ジェットイメージ  $N \times N$  to スペクトル  $N$  ( $N \sim 20$  for our case.)**  
とても簡単に収束するし、結果も安定。**ジェットイメージを使ってるCNNはかなりの場合ジェットスペクトルを発見するために時間を無駄にしてると思う。**
- ❖ **機械学習はソフトな分布も使っている。MC と実データが違うことからくる不定性がある。**
- ❖ 自然がパートンシャワーで記述できるとは限らない。
- ❖ トップはもっと複雑で計算中