JHEP 1507 (2015) 086

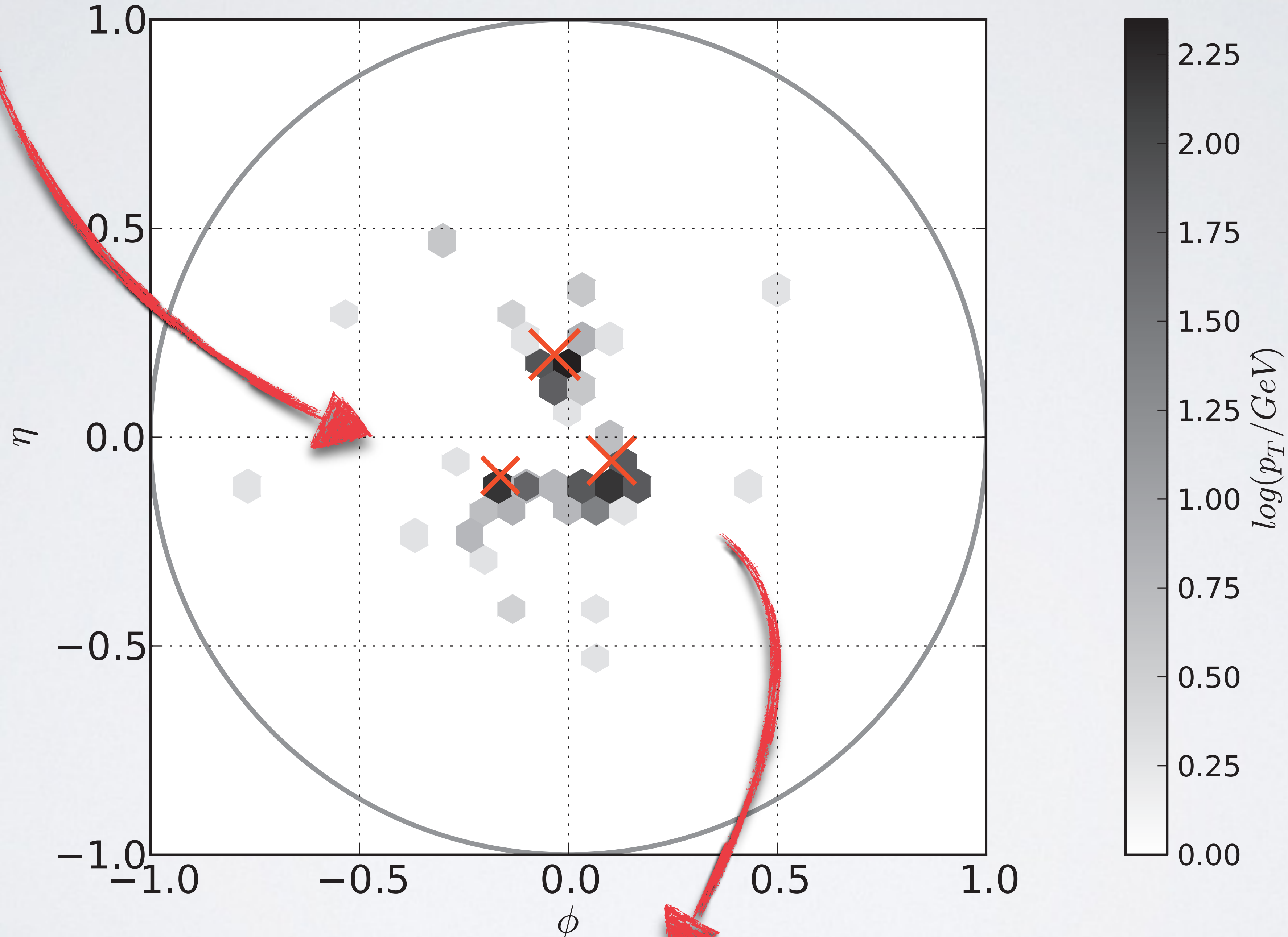
Boosted top identification with pattern recognition

— jet substructure with neural networks —

Leandro Almeida (Ecole Normal), Mihailo Backovic (CP3-UCL),
Seung J. Lee (KAIST), Maxim Perelstein (Cornell)

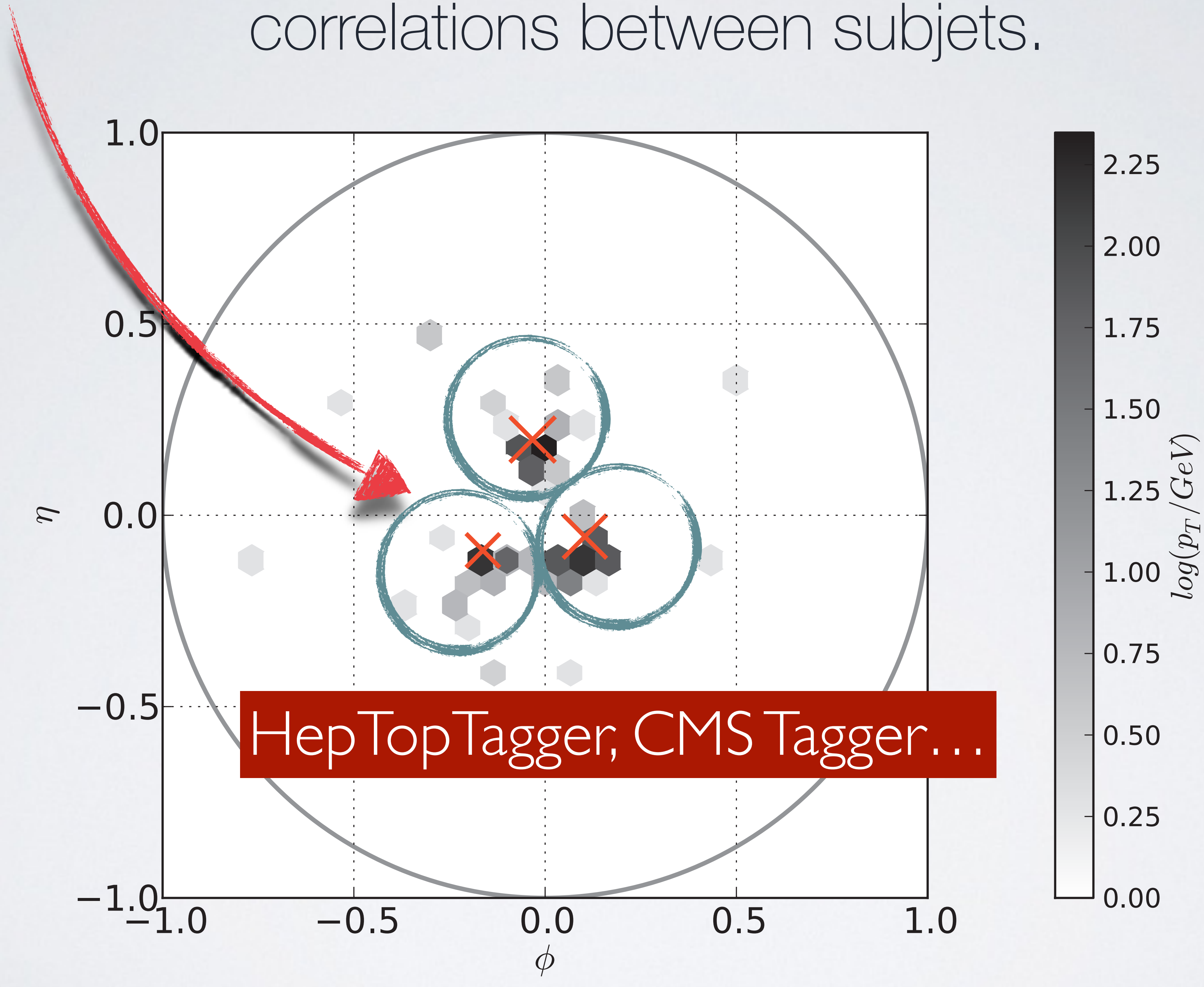
Let's begin with a **picture**…

This is a high p_T top jet (a very good one indeed!)…

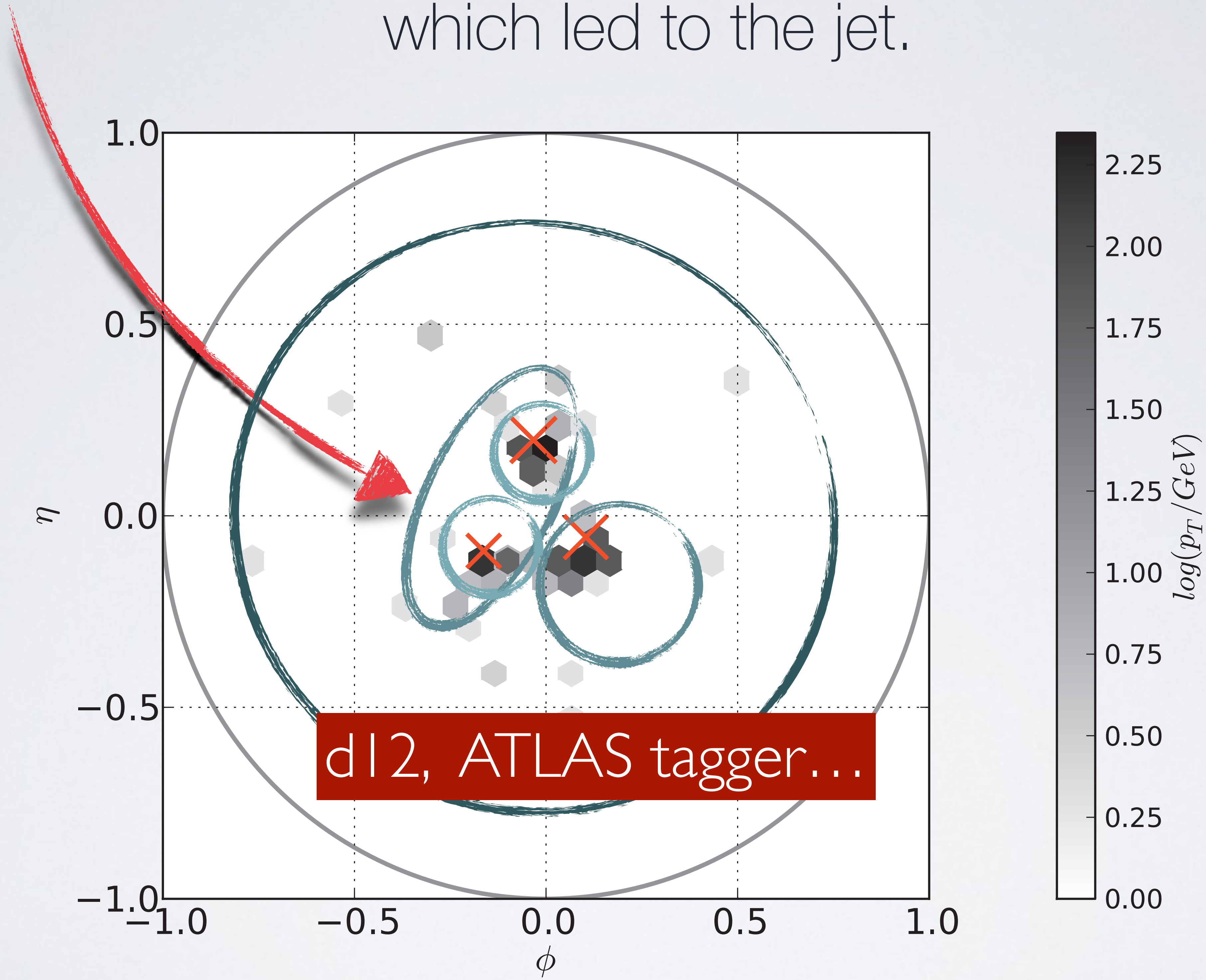


… and there are **many ways to interpret** this picture

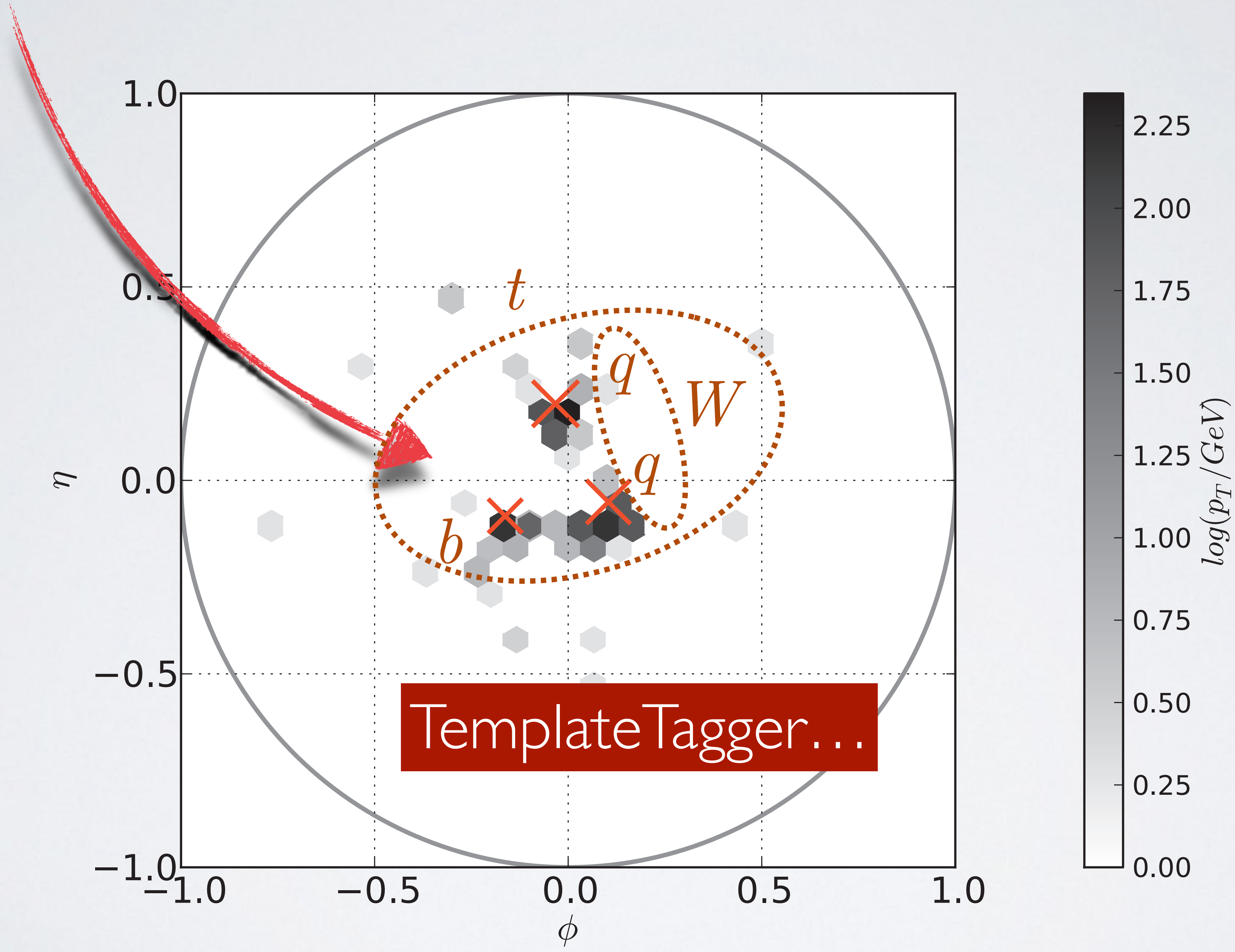
Subjets - recluster the event with a smaller cone and exploit correlations between subjets.



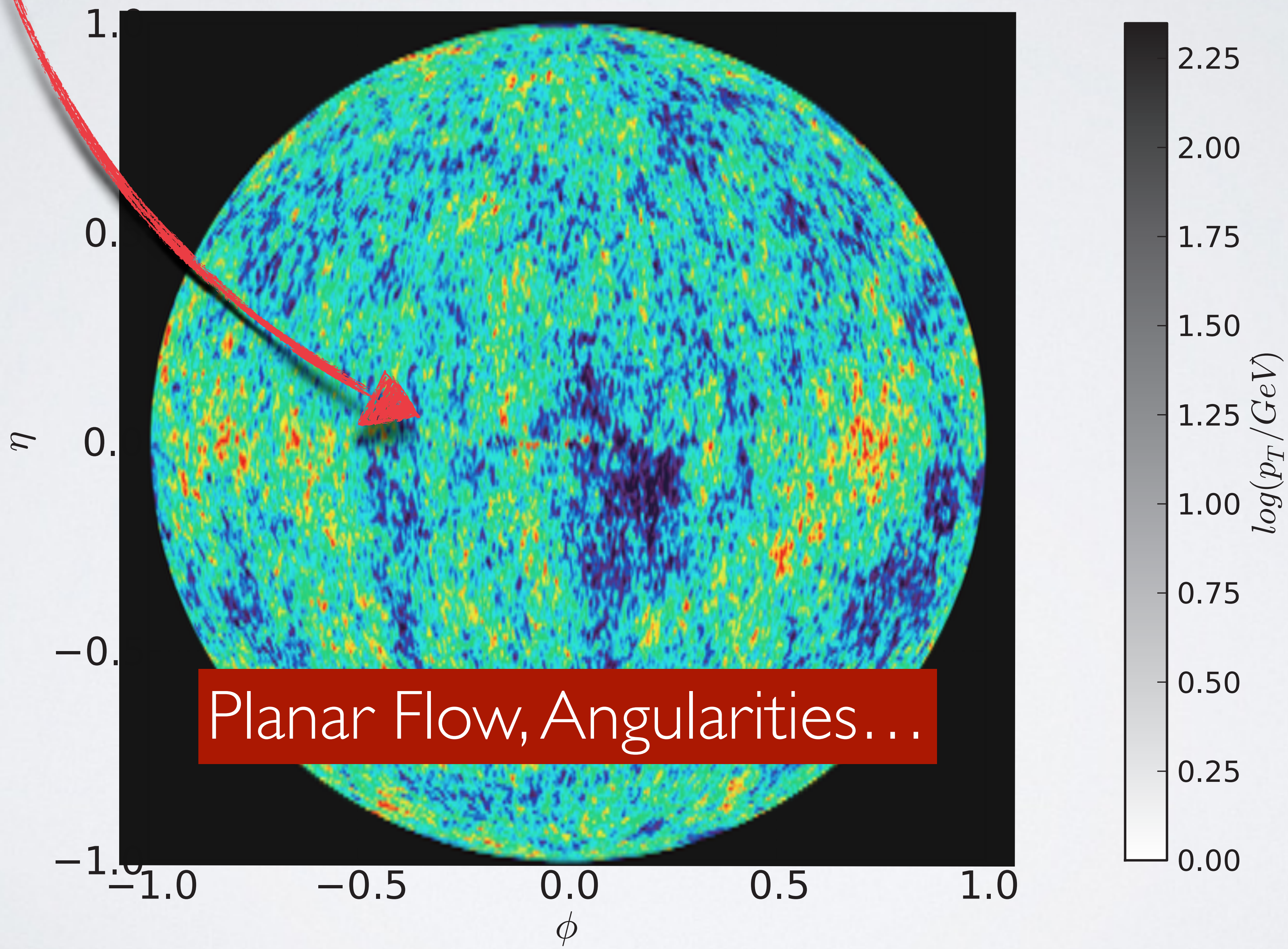
Clustering history- exploit the **differences in steps** which led to the jet.



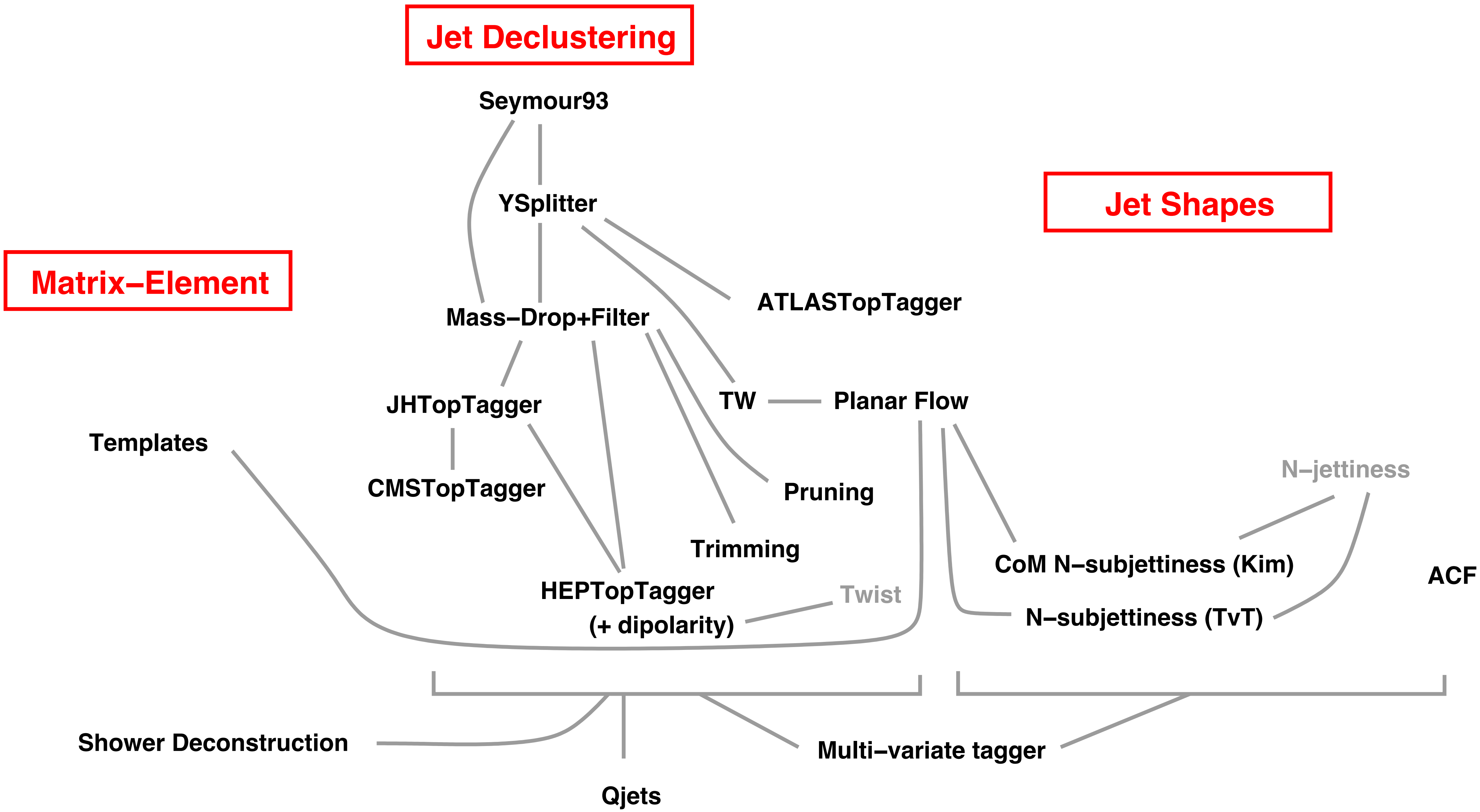
Partons - Interpret the jet as a **partonic structure** with kinematic properties of some heavy boosted object.



Energy distribution - the picture is essentially some distribution $f(\eta, \phi)$. Look at the moments of the distribution

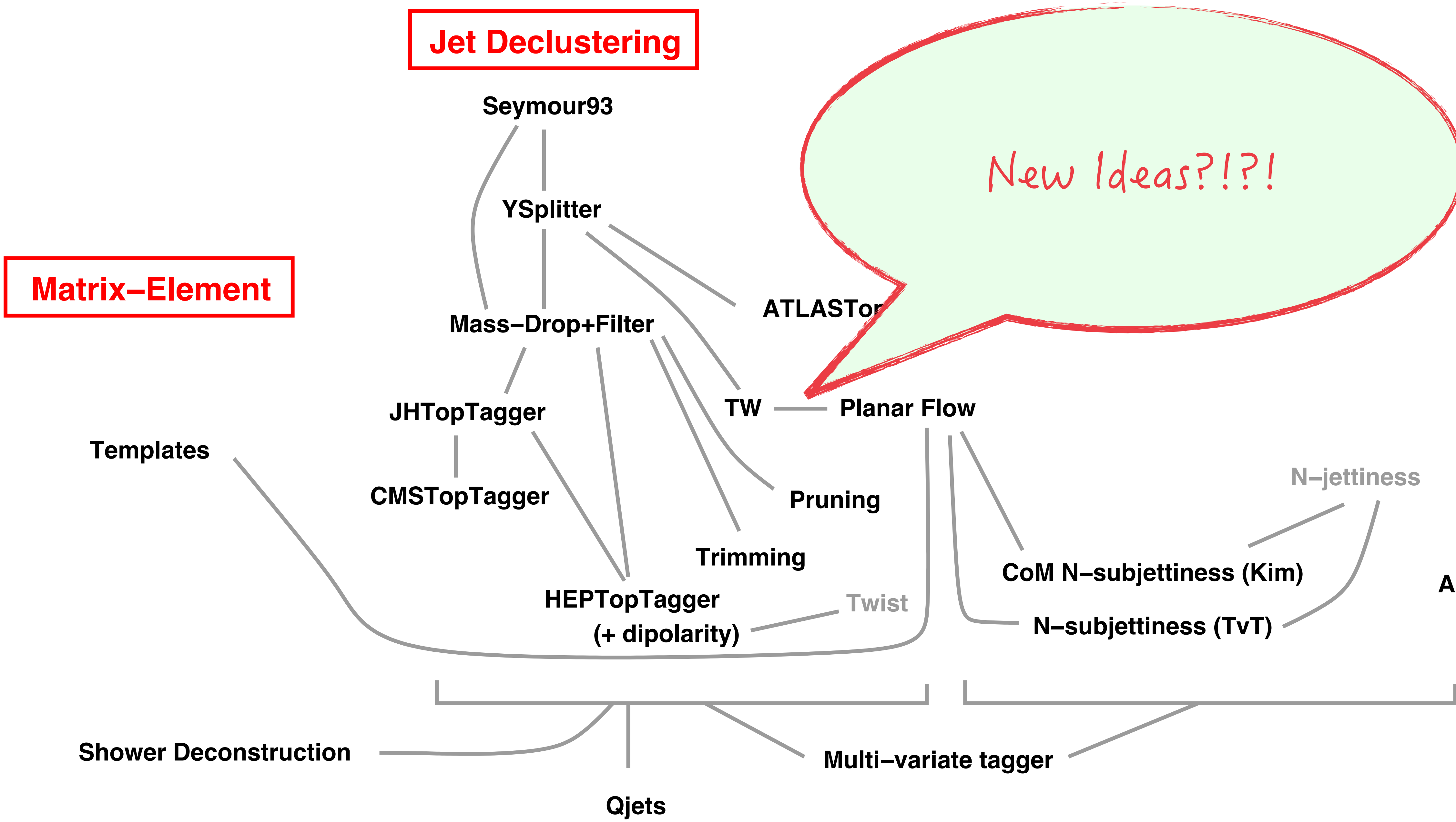


Some taggers and jet-substructure observables



apologies for omitted taggers, arguable links, etc.

Some taggers and jet-substructure observables

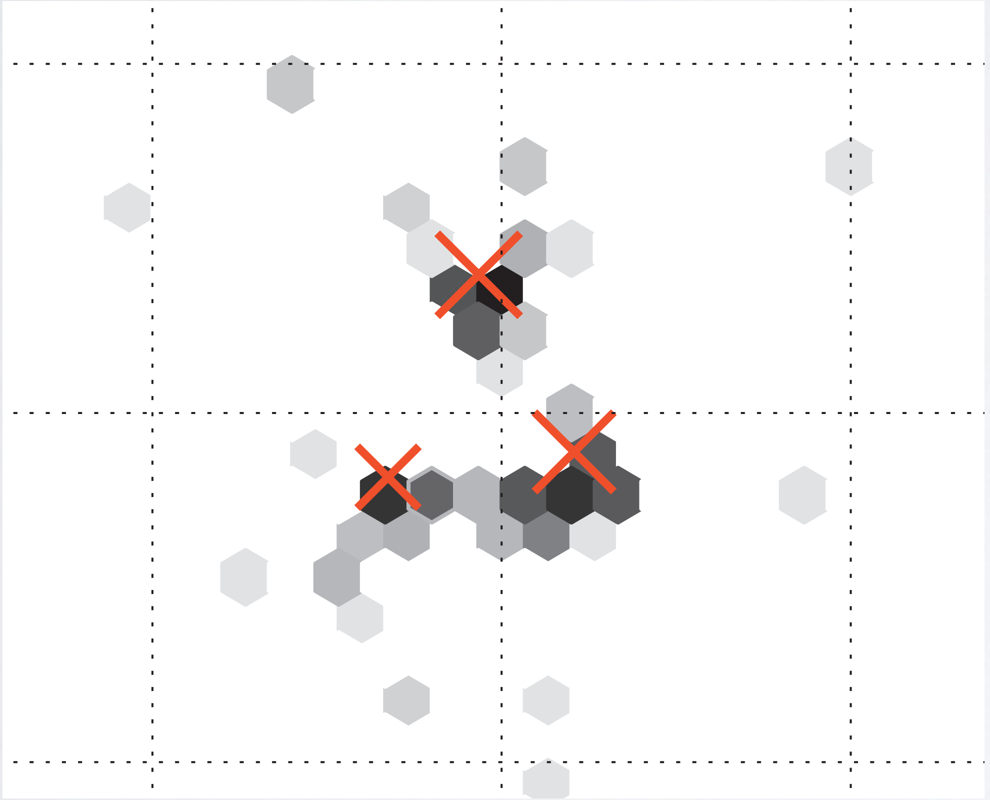


apologies for omitted taggers, arguable links, etc.

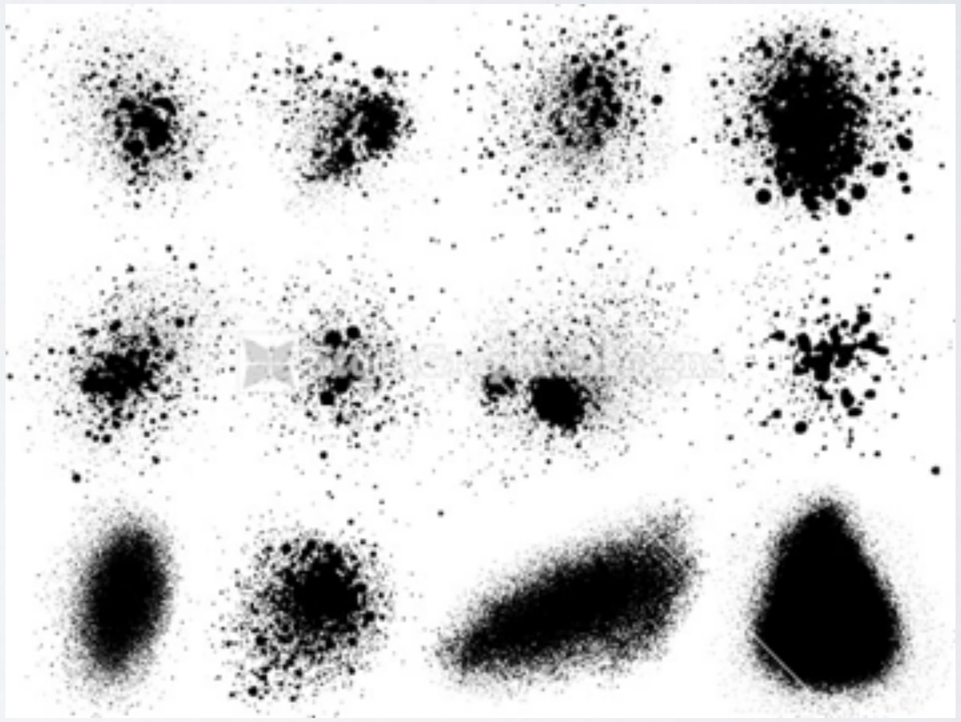
(Crazy) New idea…

Treat a jet as a **“splash pattern”** or image.

Use **image/pattern recognition** technology
to classify “splash patterns”.



=



**jet “splash patterns” contain all
of calo. information.**

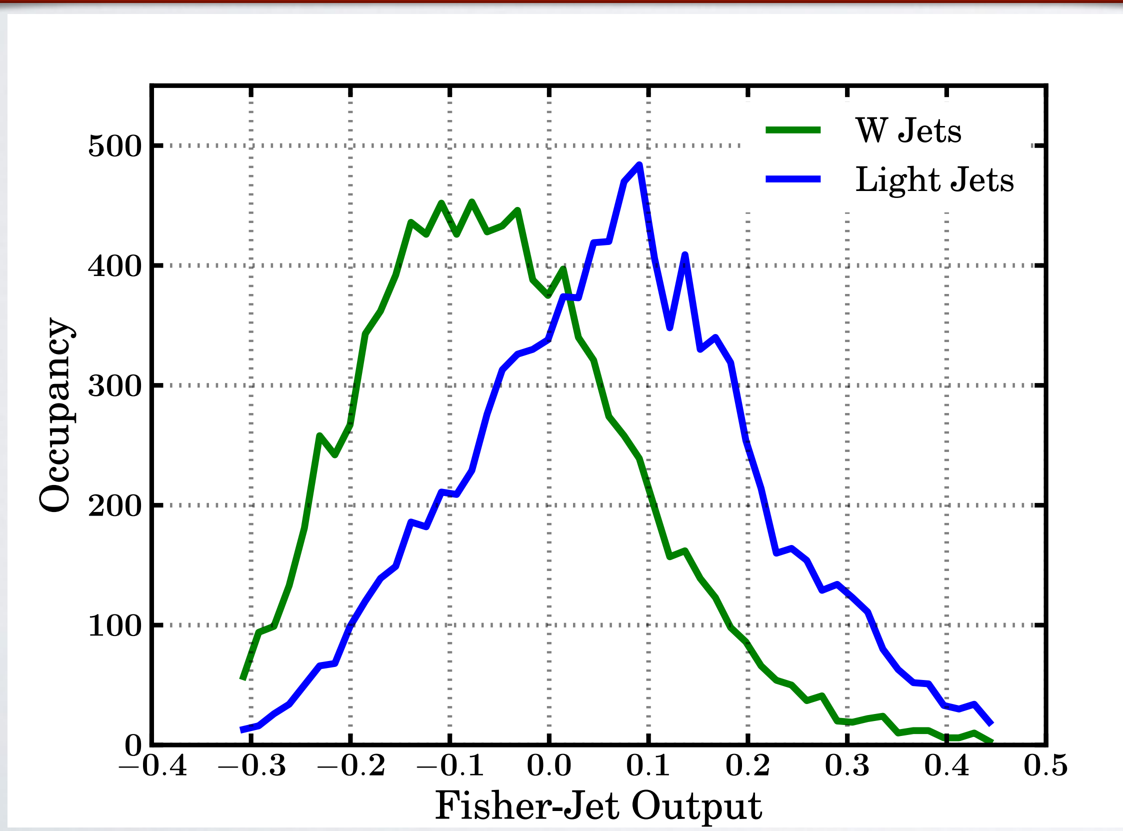
A couple of years ago a similar idea appeared

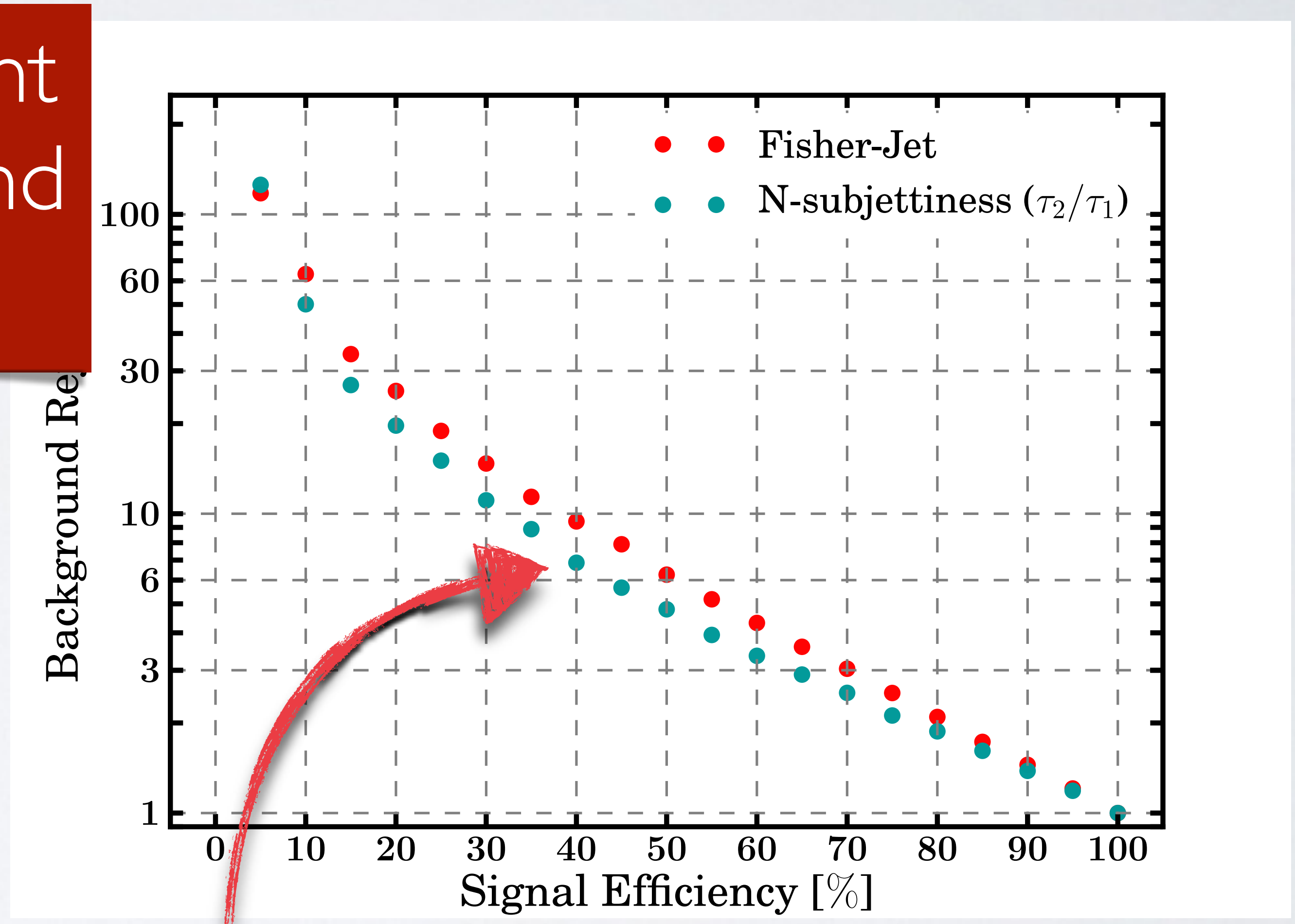
Jet-Images:
Computer Vision Inspired Techniques for Jet Tagging

Josh Cogan[a] Michael Kagan[a] Emanuel Strauss[a] Ariel Schwartzman[a]

JHEP 1502 (2015) 118
(See also BOOST2013 talk by J. Cogan)

Use of the linear Fisher discriminant for the purpose of W/h-tagging and q/g discrimination.



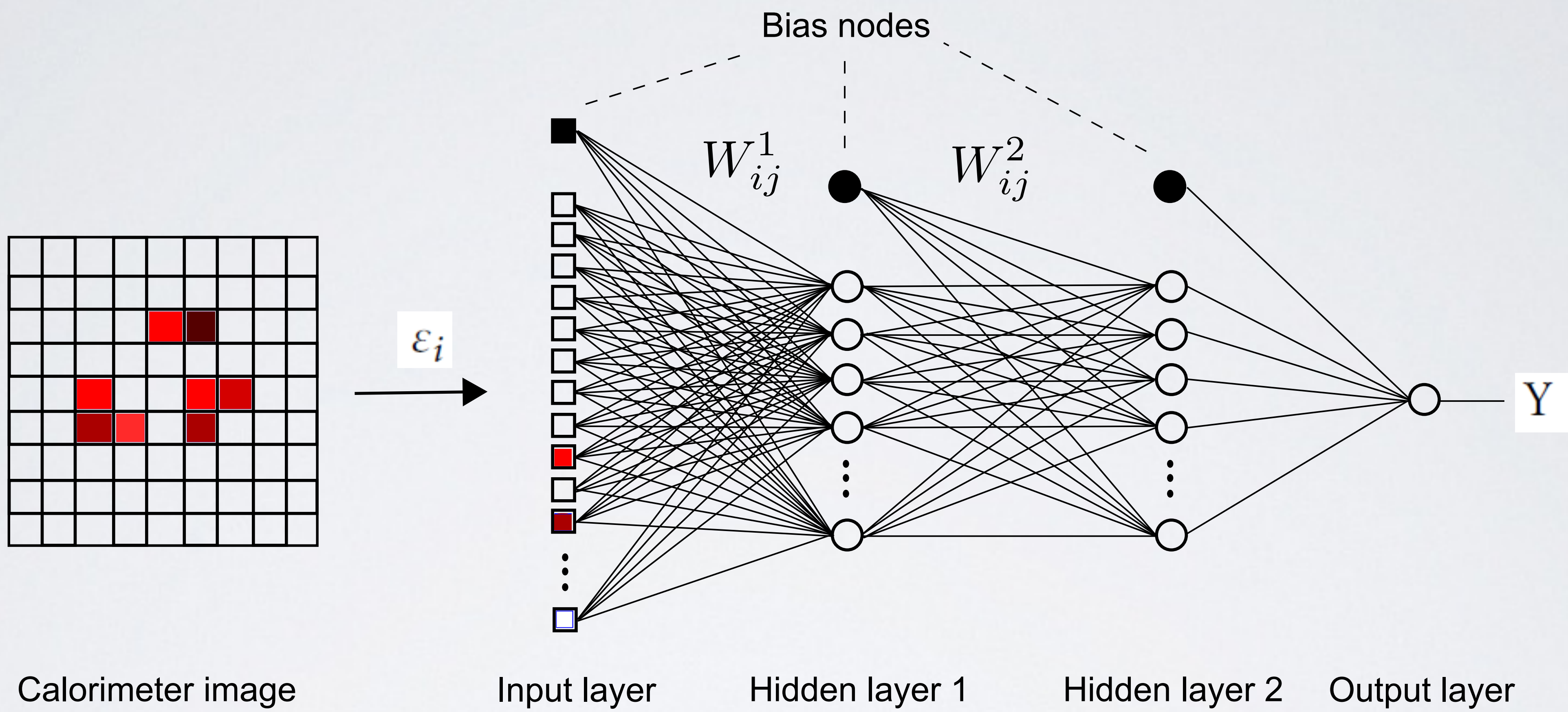


(a)

In some cases out-performed N-subjetiness!

We tried using Neural Networks (NN) to classify "splash patterns"

Feed the entire jet **(splash pattern)** as an **array of pixels**.



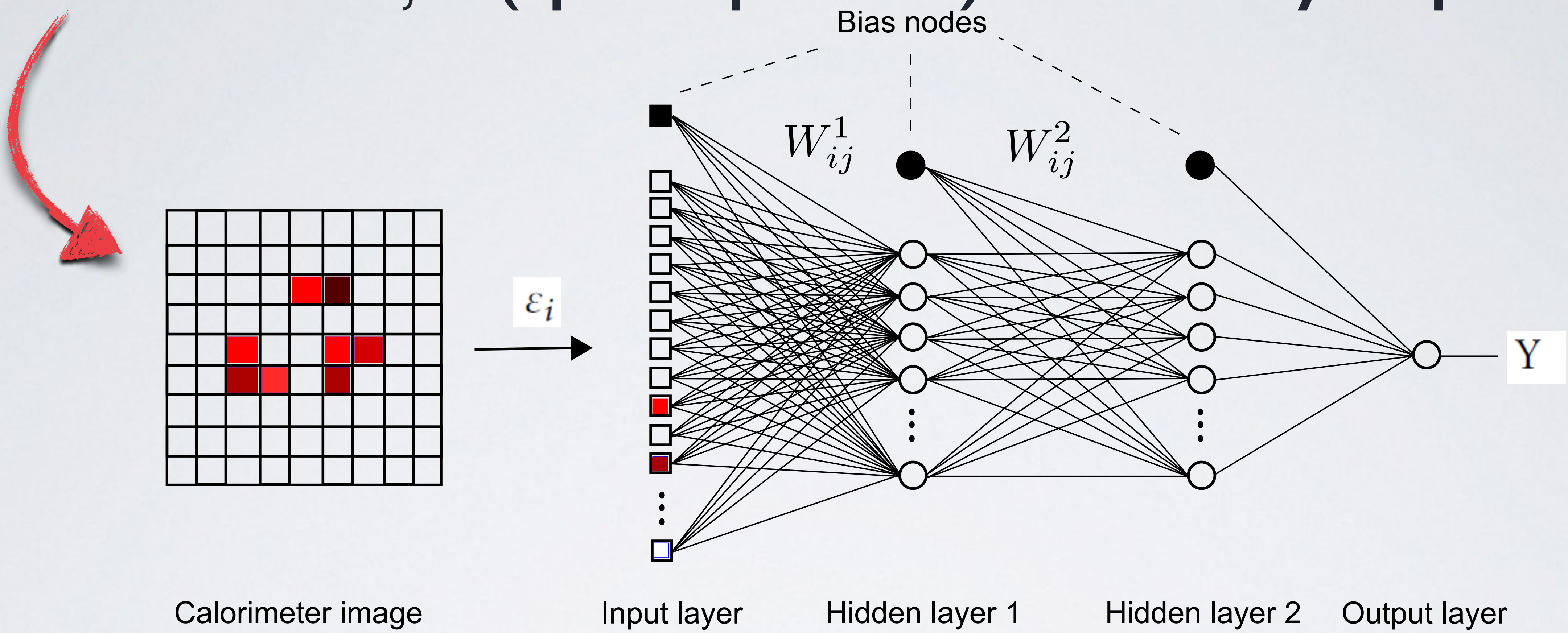
At each layer compute a weighted sum:

$$\epsilon_i \to h_i^{(1)} = f(W_{ij}^{(1)}\epsilon_j + b_i^{(1)}) \to \cdots \to h_i^{(l)} = f(W_{ij}^{(l)} h_j^{(l-1)} + b_i^{(l)}) \to Y = f(W_j^{(O)} h_j^{(l)} + b^{(O)})$$

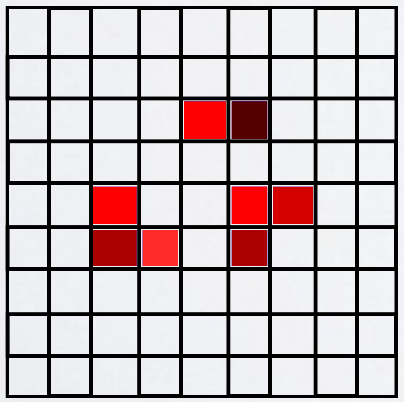
$$f(z) = \frac{1}{1+e^{-z}}$$

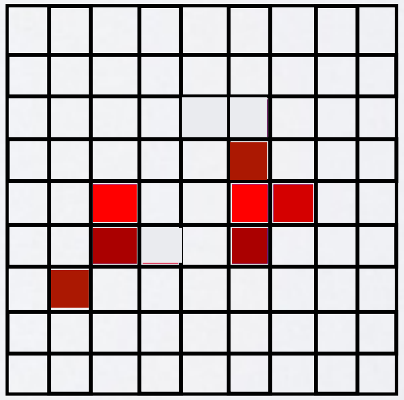
We tried using Neural Networks (NN) to classify "splash patterns"

Feed the entire jet **(splash pattern)** as an **array of pixels**.



Adjust (train) the weights W_{ij} to give:

Top  $Y = 1$

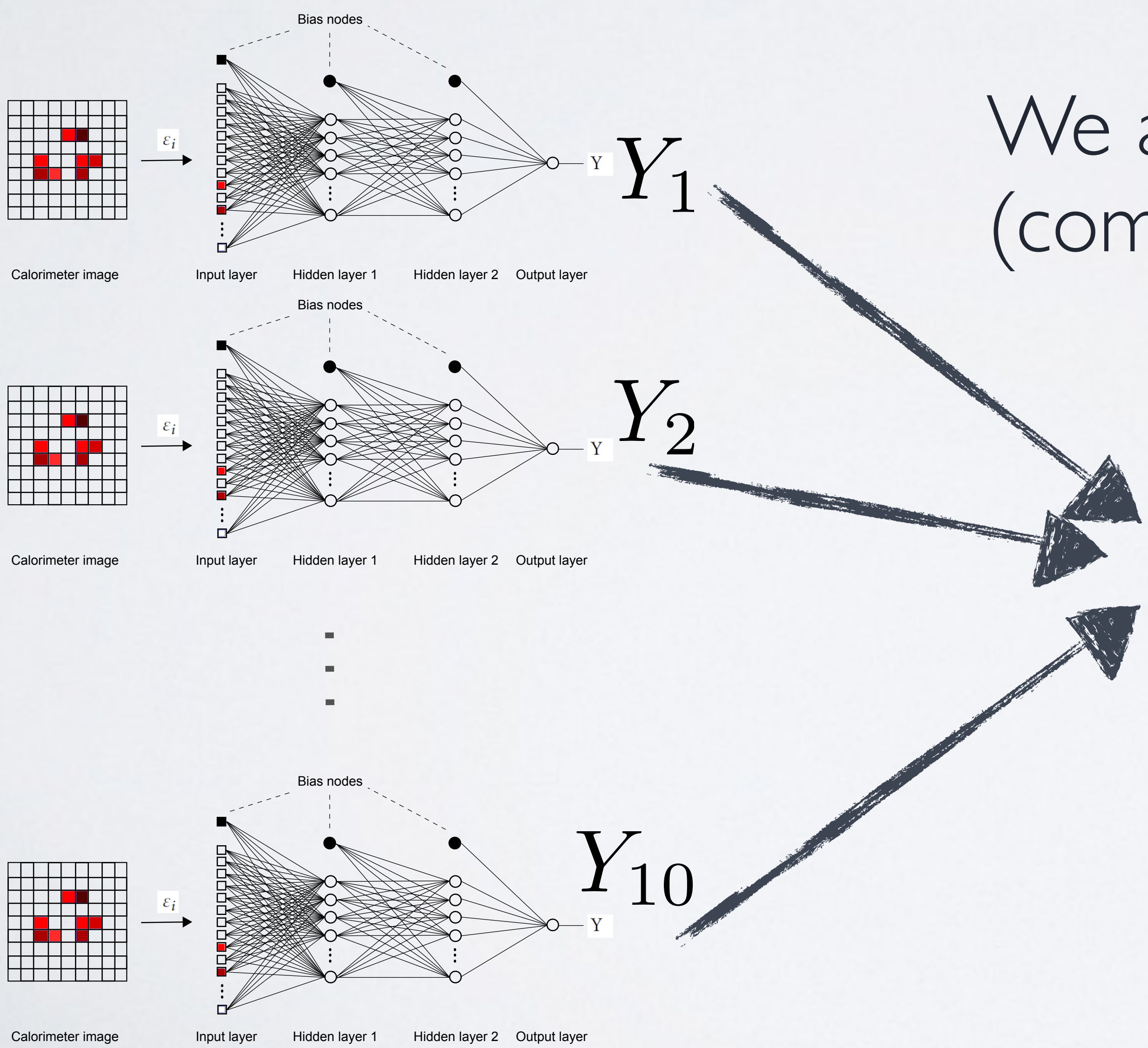
QCD  $Y = 0$

*** We use the standard back-propagation algorithm with gradient descent

In training the NN, we found that a single NN often **mis-indentifies some fraction of tops**

Solution: Train another NN with the same training sample, but force the weights of the mis-identified jets to be larger.

Allows the NN to "focus" more on the jets which failed the previous classification.



We actually did this B = 10 times (computing power is cheap!)

$$\boxed{O = \frac{1}{B}\sum_{i=1}^{B} Y_i.}$$

top "classifier"

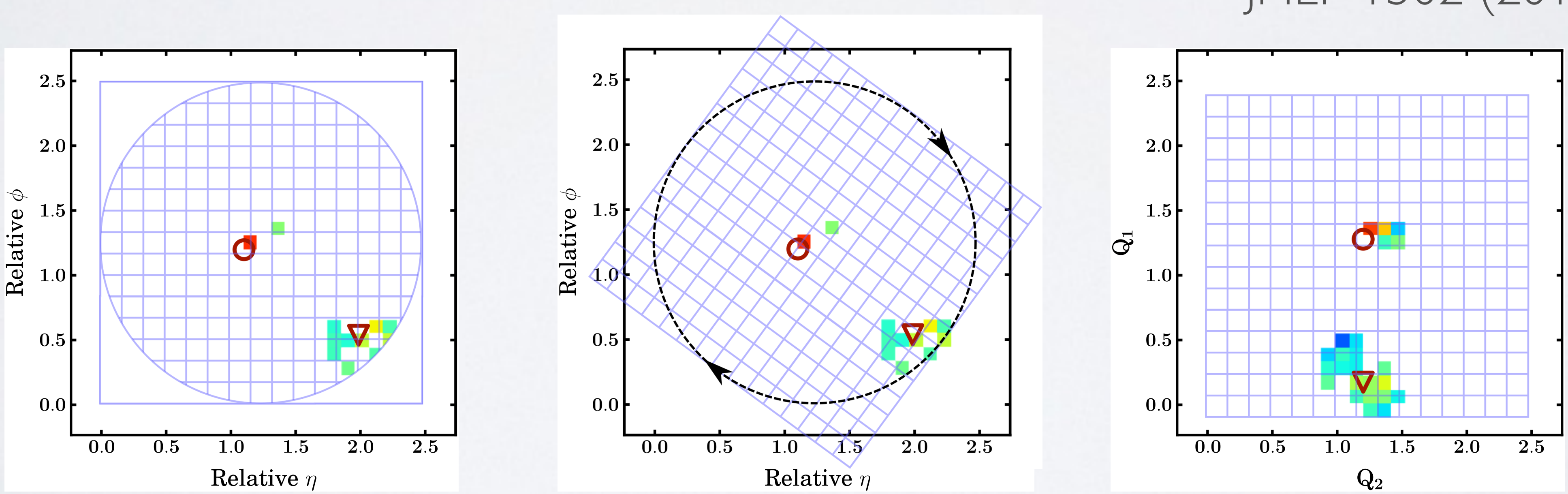
Pre-processing of jets

Often, pattern recognition algorithms **require some pre-processing of input data**.

Jet splash patterns are uniformly distributed with respect to the angle around the jet axis.

E.g: Linear Fisher discriminant requires to rotate each image into the same frame (Not necessarily trivial)

JHEP 1502 (2015) 118



(a) Jet-image prior to rotation

(b) Rotated pixel grid

(c) Jet-image after projection onto rotated grid, before translation

Pre-processing of jets

Often, pattern recognition algorithms **require some pre-processing o**

Je ... the

ar

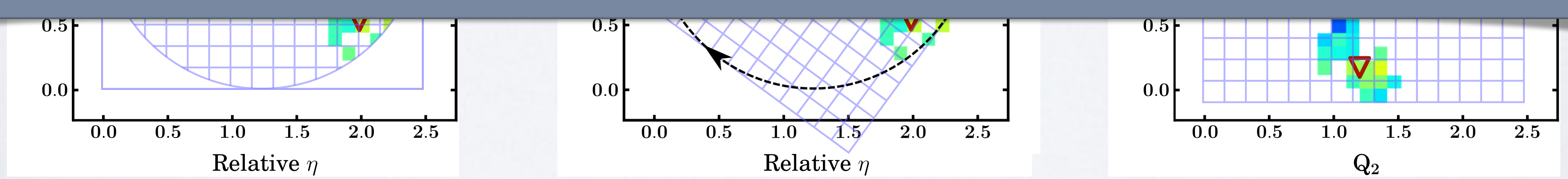
E. ... into the

sa

) 118

We have checked that we do not need to pre-process the jets for NN to work

However, the training sample size needs to be much larger if the jets are not pre-processed (**NN needs to learn about the angle around the jet axis**)



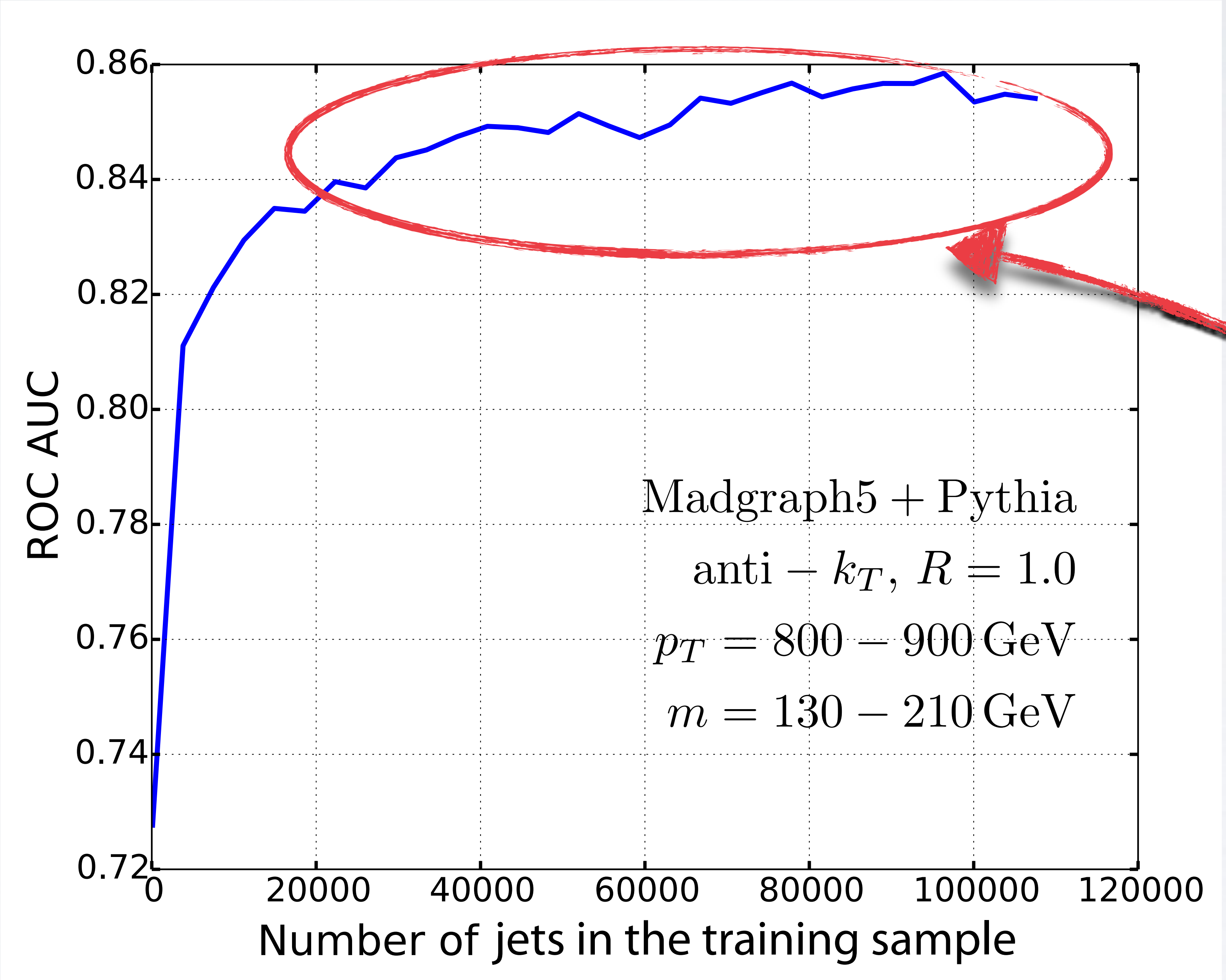
(a) Jet-image prior to rotation

(b) Rotated pixel grid

(c) Jet-image after projection onto rotated grid, before translation

How large should a training sample be?

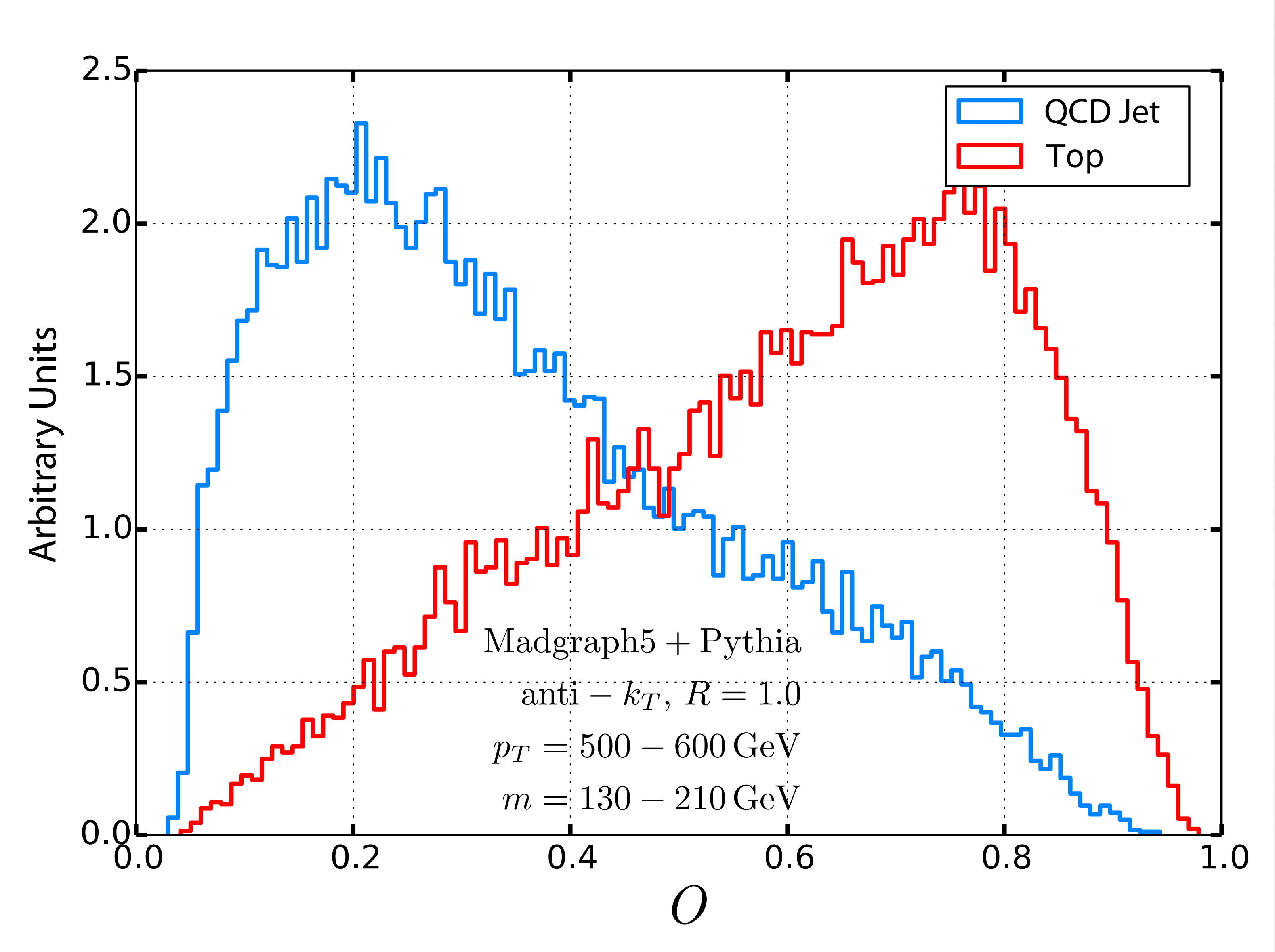
test sample of 50000 pre-processed top jets



Only about 20000 events needed to train the NN

Let's look at a few examples…

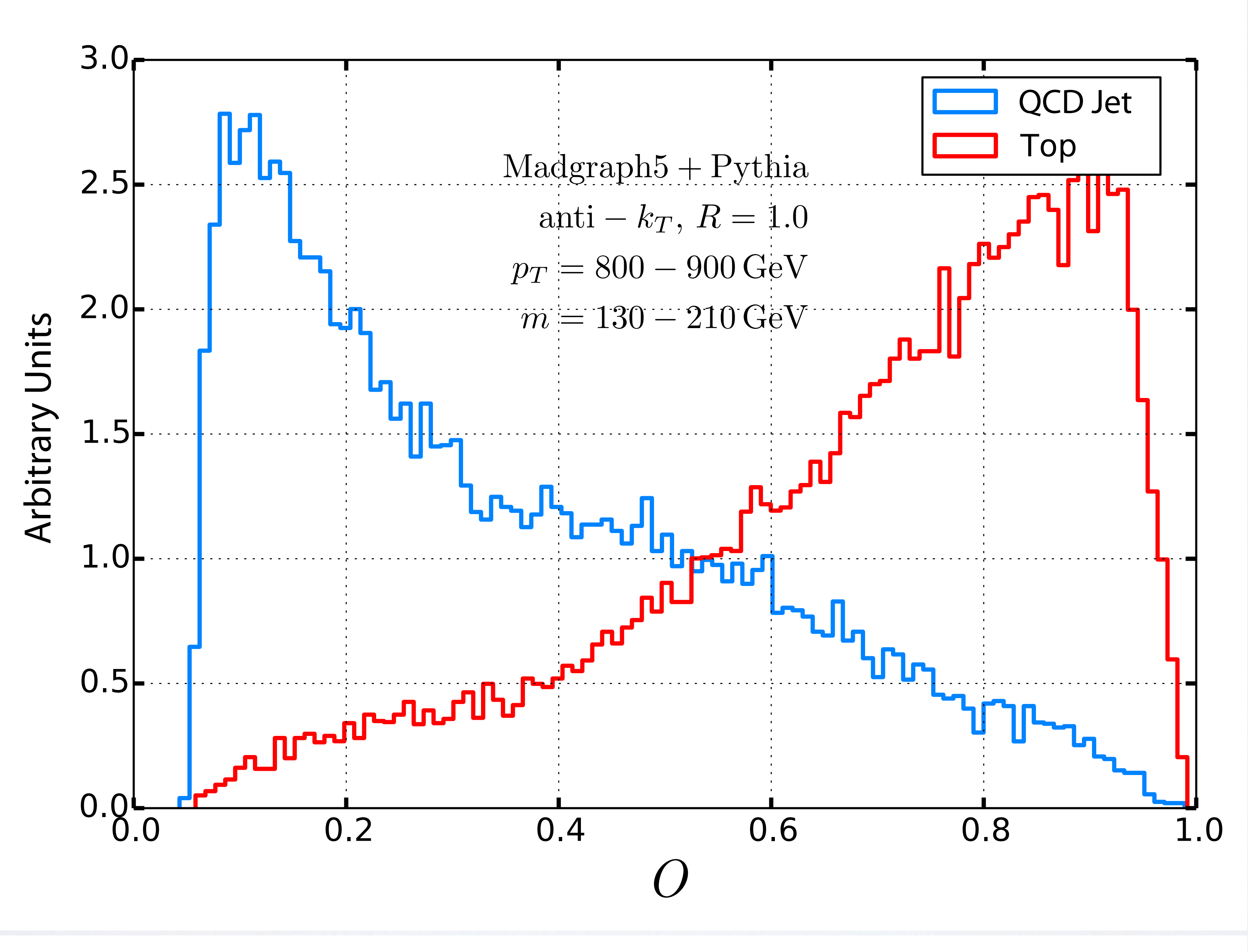
$$p_T = 500 - 600\,\mathrm{GeV}$$



Good signal/background separation.
So far so good…

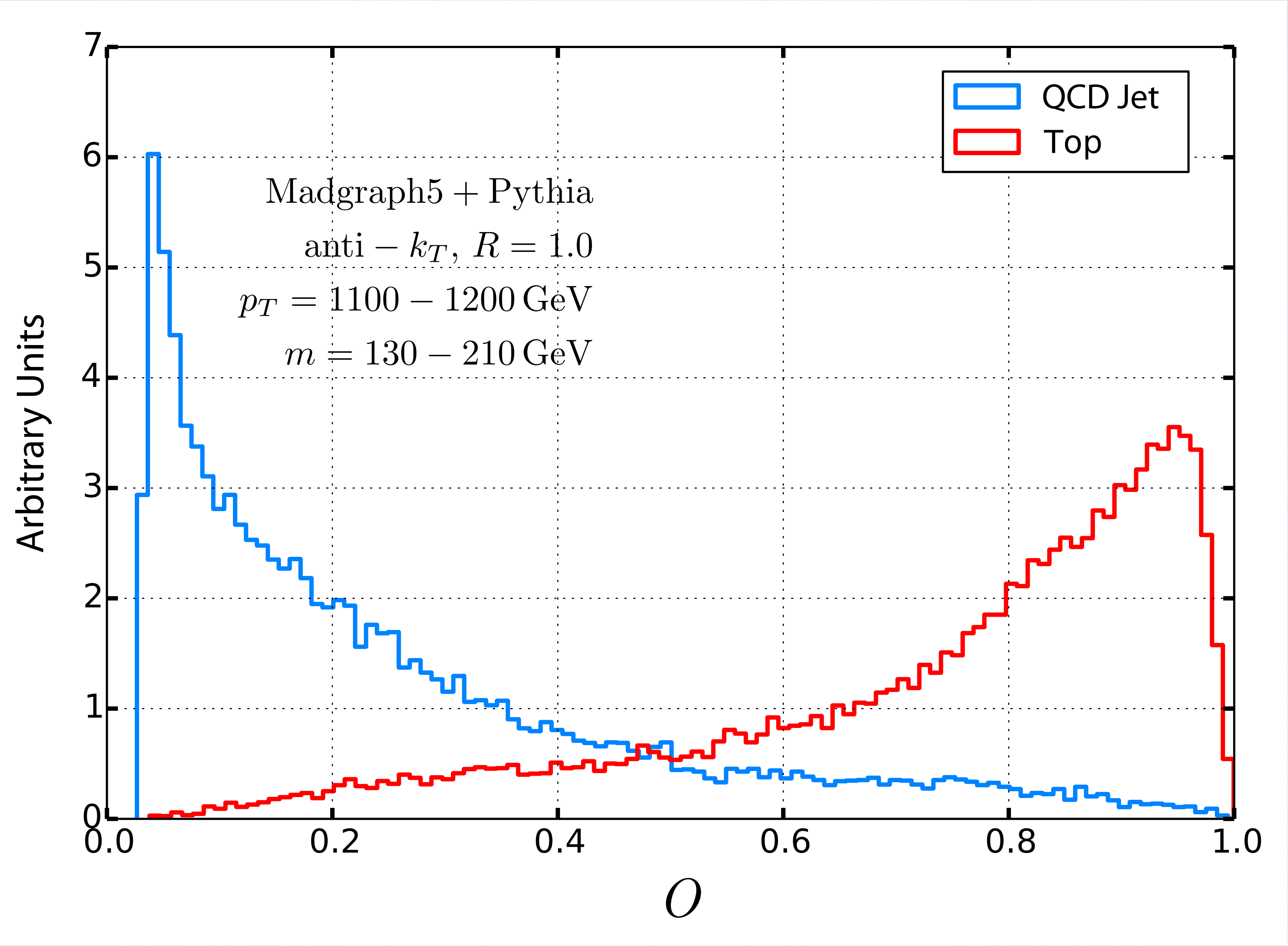
Let's look at a few examples…

$$p_T = 800 - 900\,\mathrm{GeV}$$

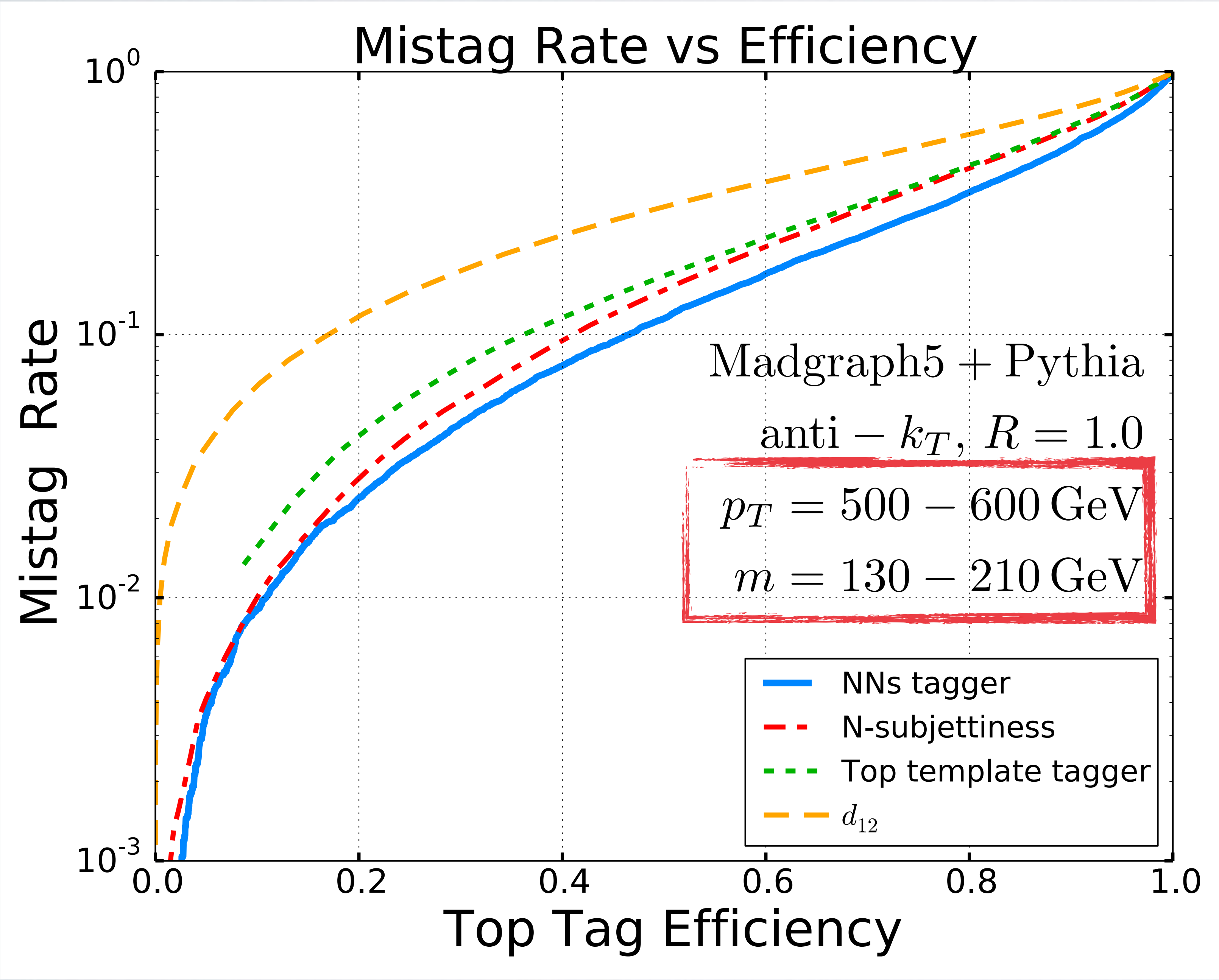


Let's look at a few examples…

$$p_T = 1100 - 1200\,\mathrm{GeV}$$



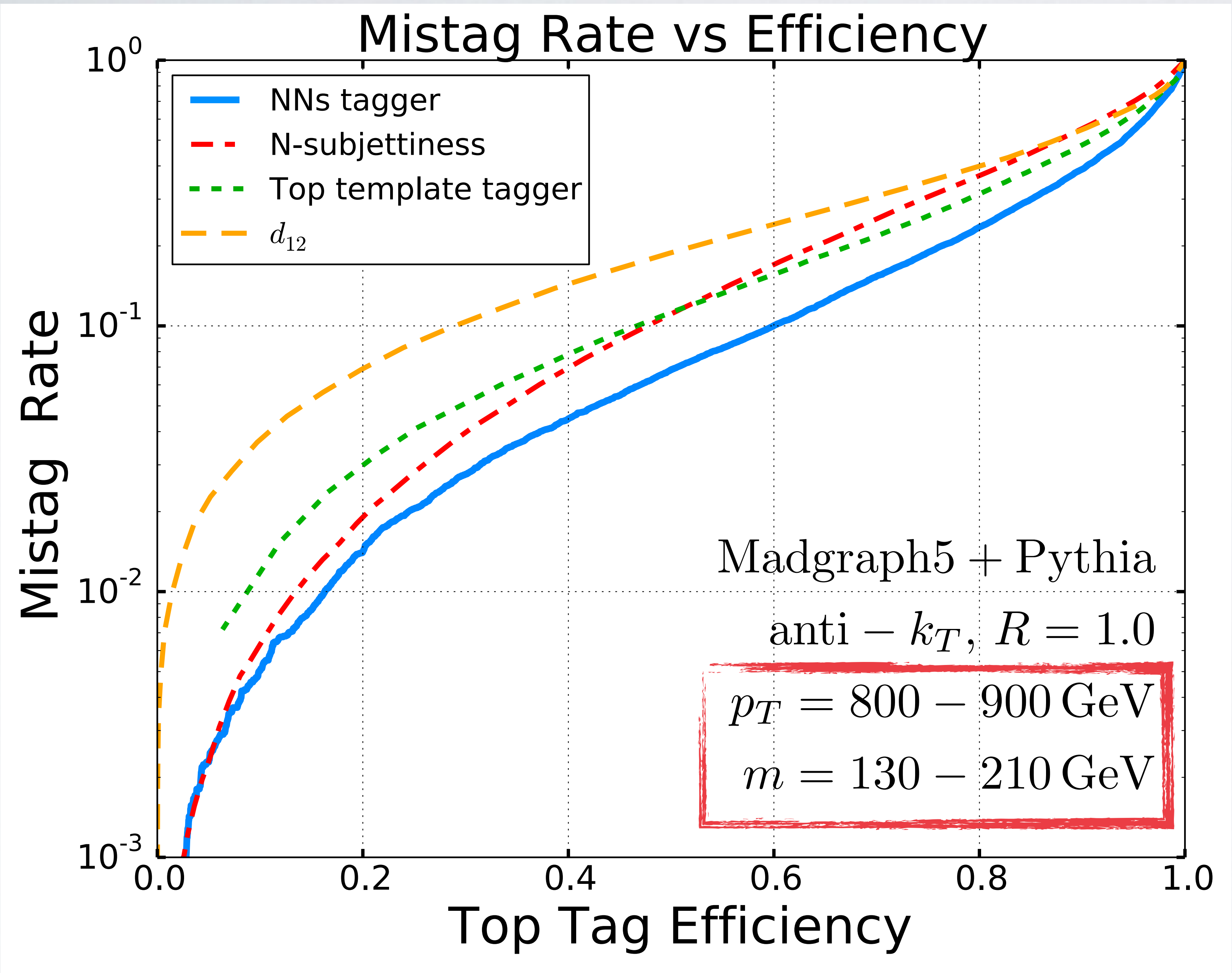
Comparison with other top taggers



$$\mathrm{Eff} = \frac{N^{\mathrm{top}}_{\mathrm{top}}}{\mathcal{N}_{\mathrm{top}}}, \quad \mathrm{Mistag} = \frac{N^{\mathrm{top}}_{\mathrm{QCD}}}{\mathcal{N}_{\mathrm{QCD}}}$$

NN top tagger performance better or comparable to some existing techniques!

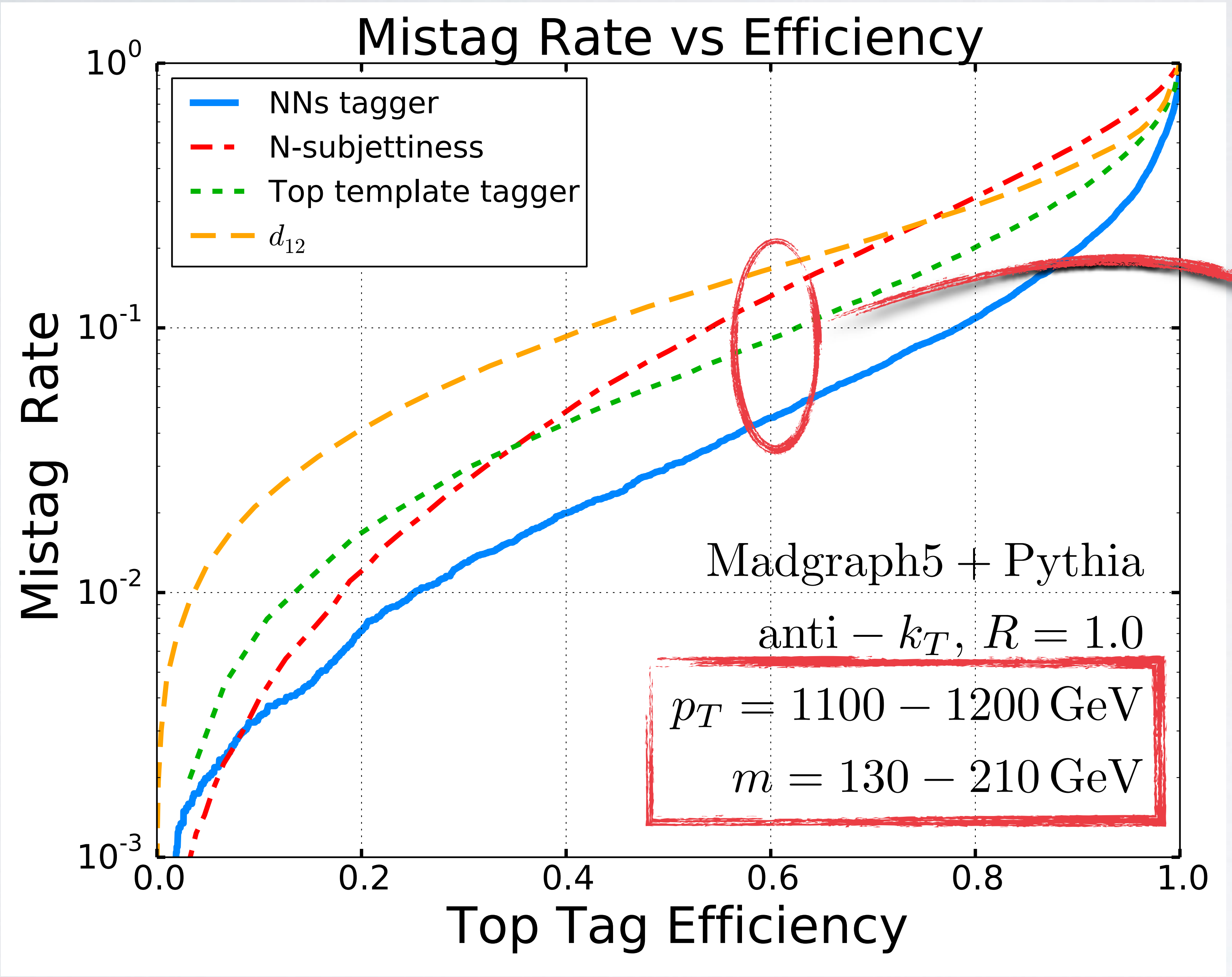
Comparison with other top taggers



$$\mathrm{Eff} = \frac{N^{\mathrm{top}}_{\mathrm{top}}}{N_{\mathrm{top}}}, \quad \mathrm{Mistag} = \frac{N^{\mathrm{top}}_{\mathrm{QCD}}}{N_{\mathrm{QCD}}}$$

NN top tagger performance better or comparable to some existing techniques!

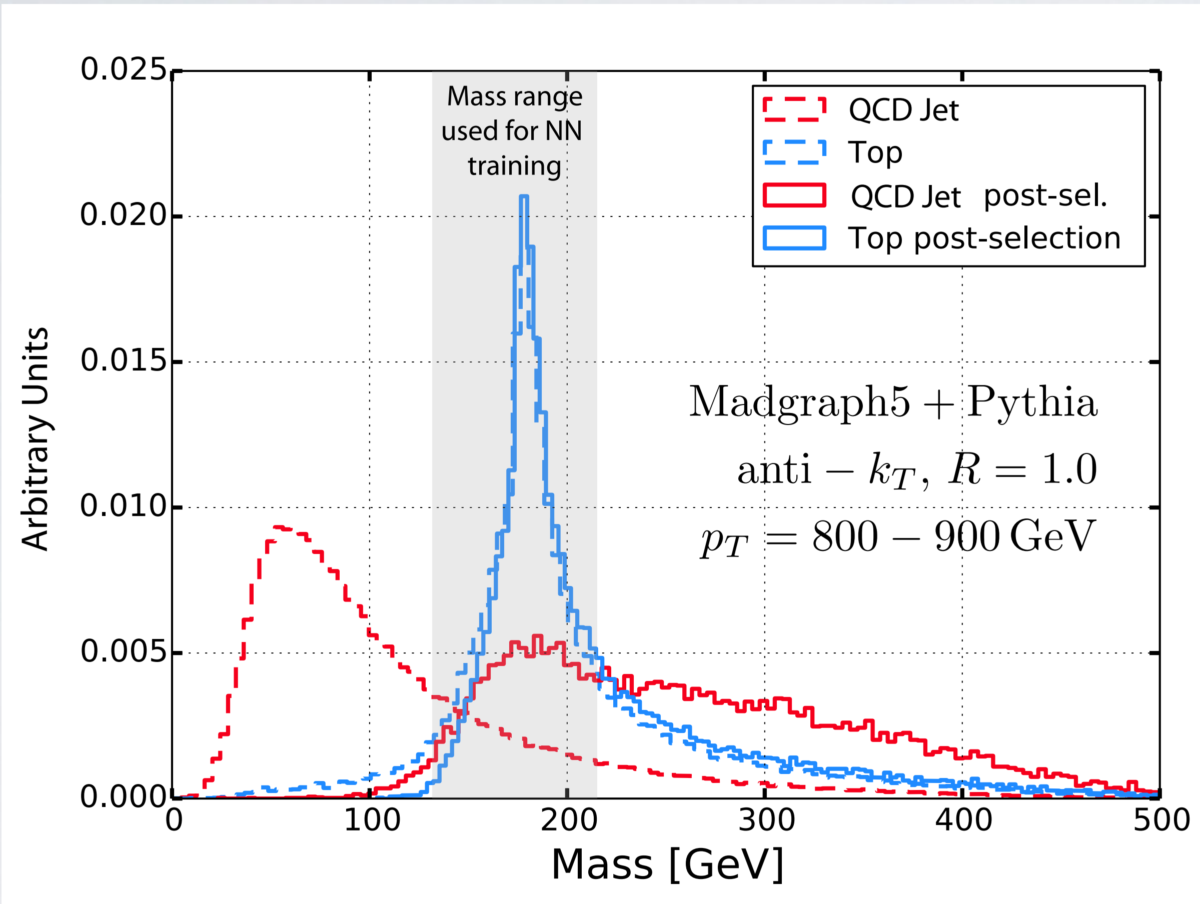
Comparison with other top taggers

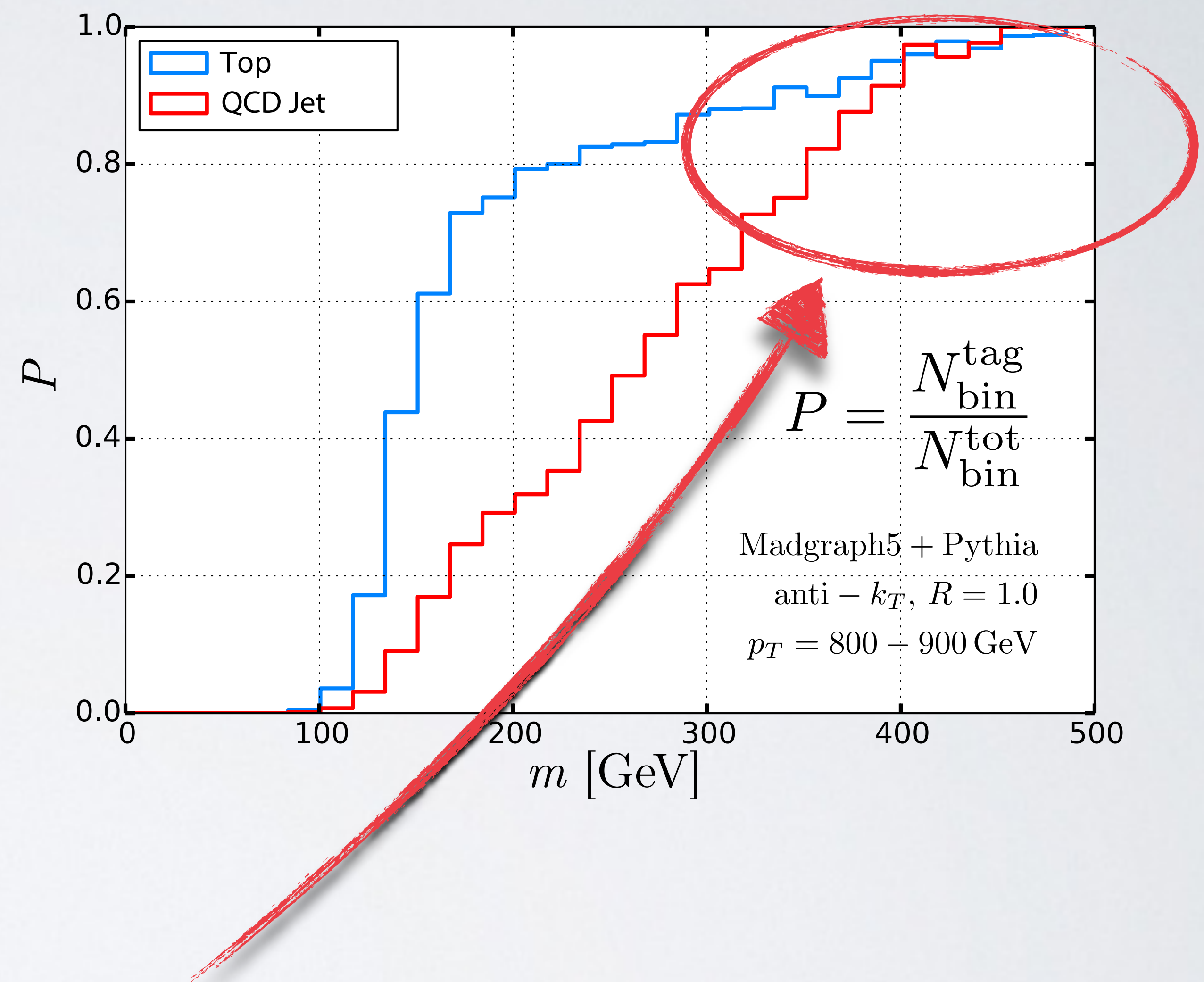


$$\mathrm{Eff} = \frac{N^{\mathrm{top}}_{\mathrm{top}}}{N_{\mathrm{top}}}, \quad \mathrm{Mistag} = \frac{N^{\mathrm{top}}_{\mathrm{QCD}}}{N_{\mathrm{QCD}}}$$

~ factor of 2 -3 improvement over existing methods!

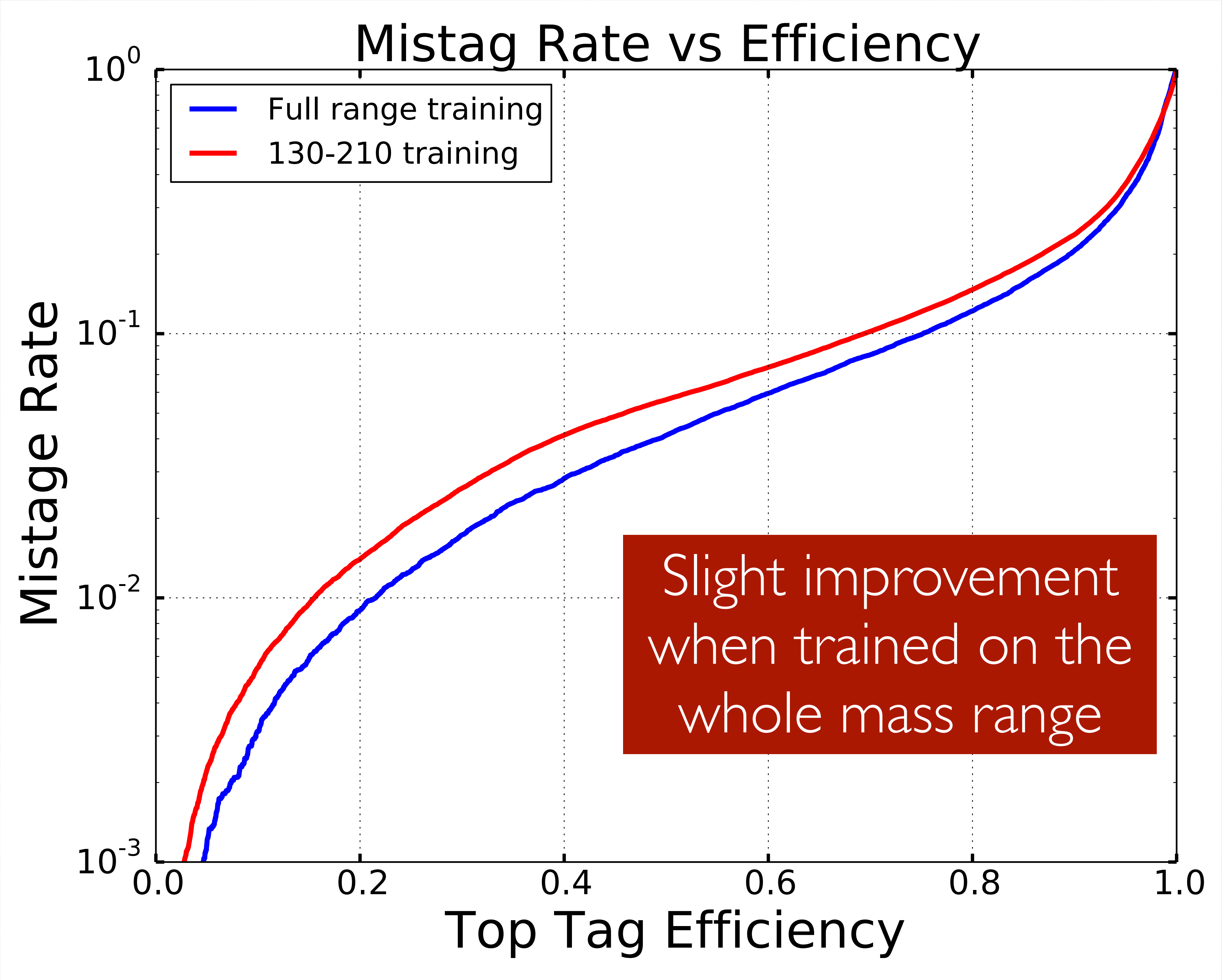
NN tagger filters jet mass.



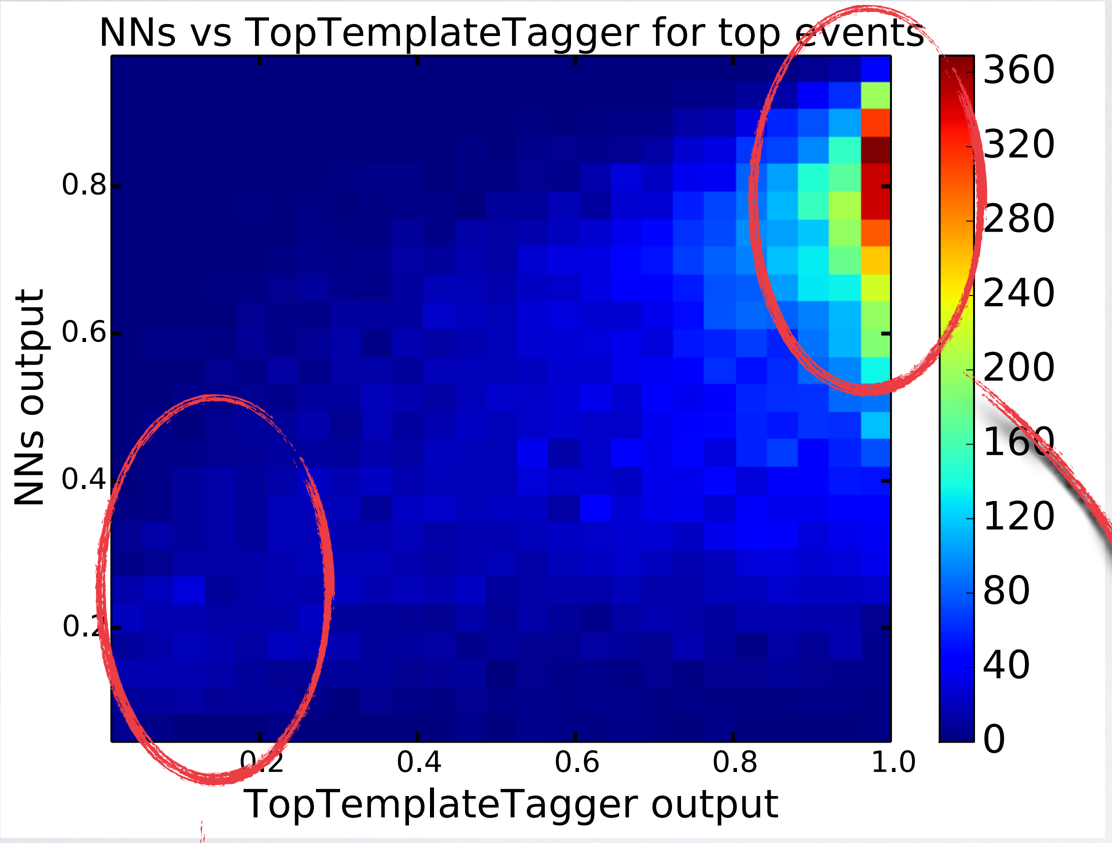


Very interesting that high mass events always pass selection

We also tried training on the full mass range

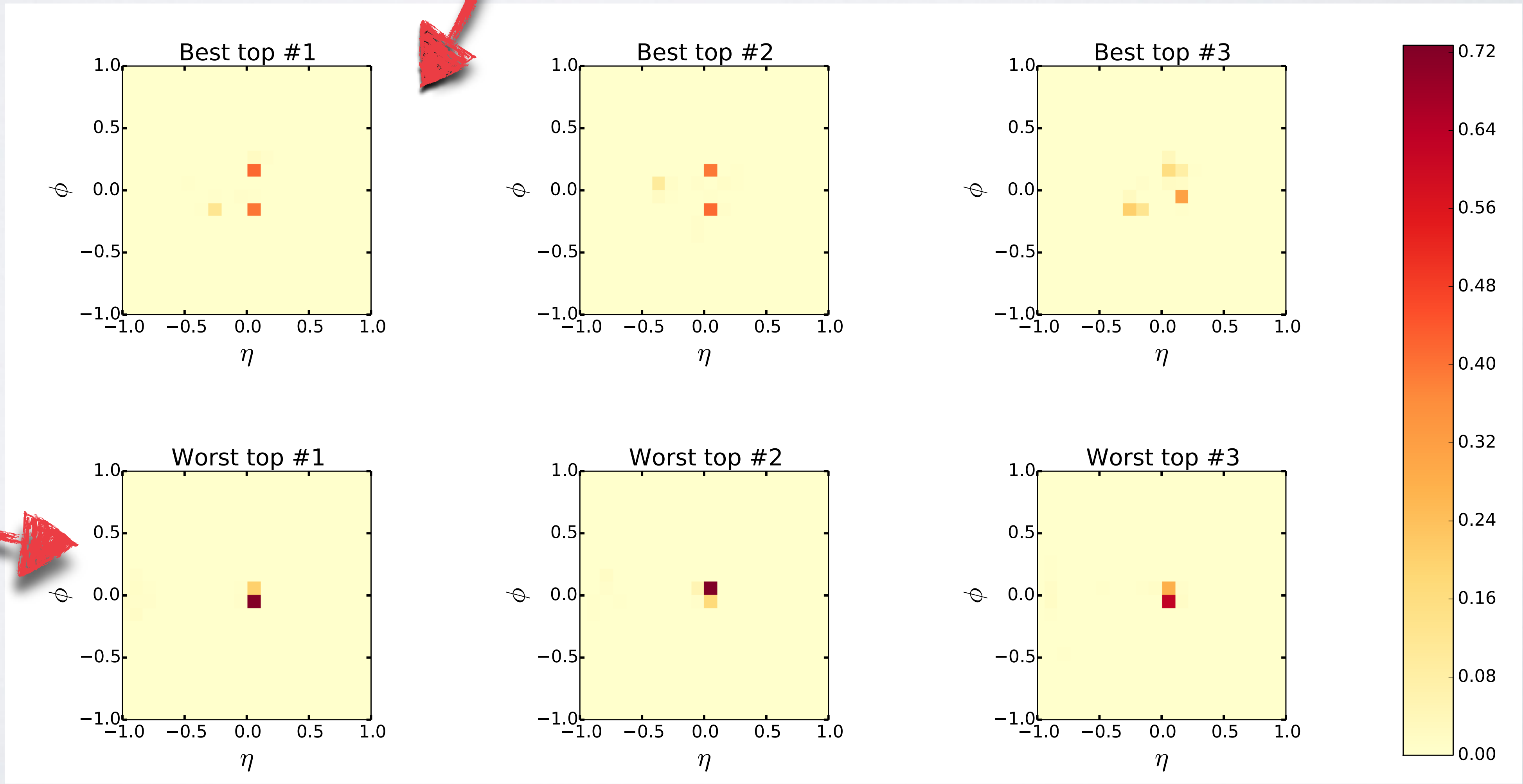


Correlation with other taggers

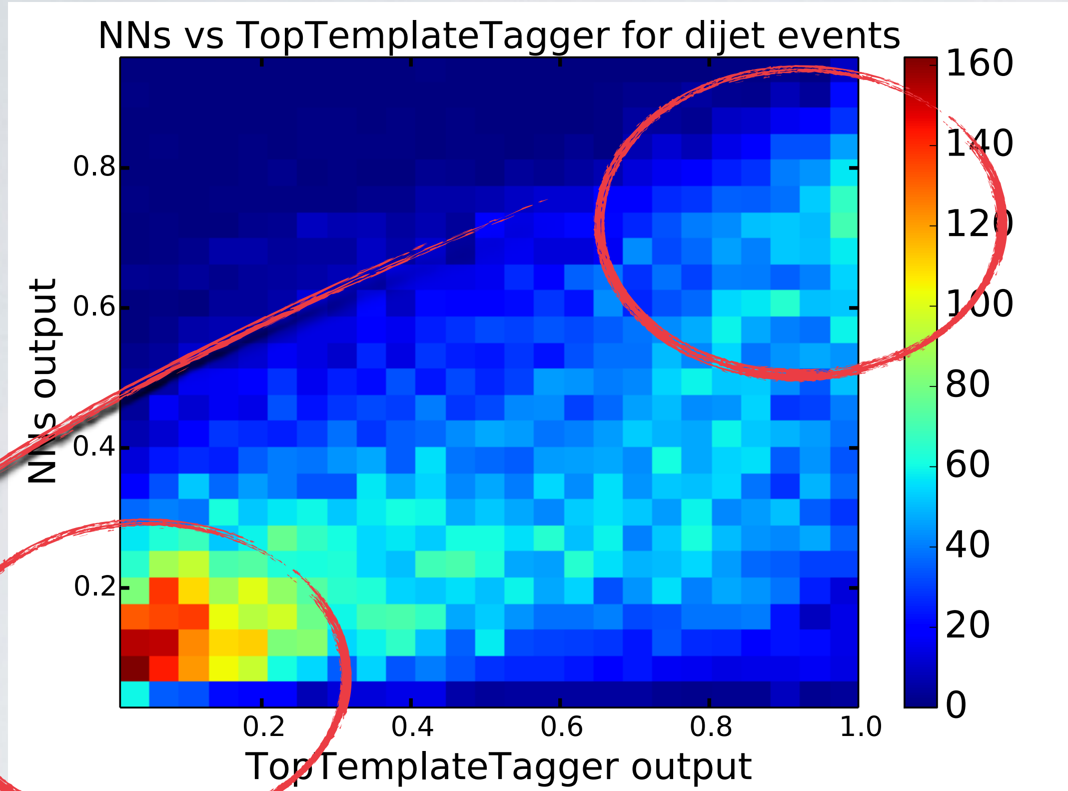


Both the NN tagger and TemplateTagger see tops for what they are

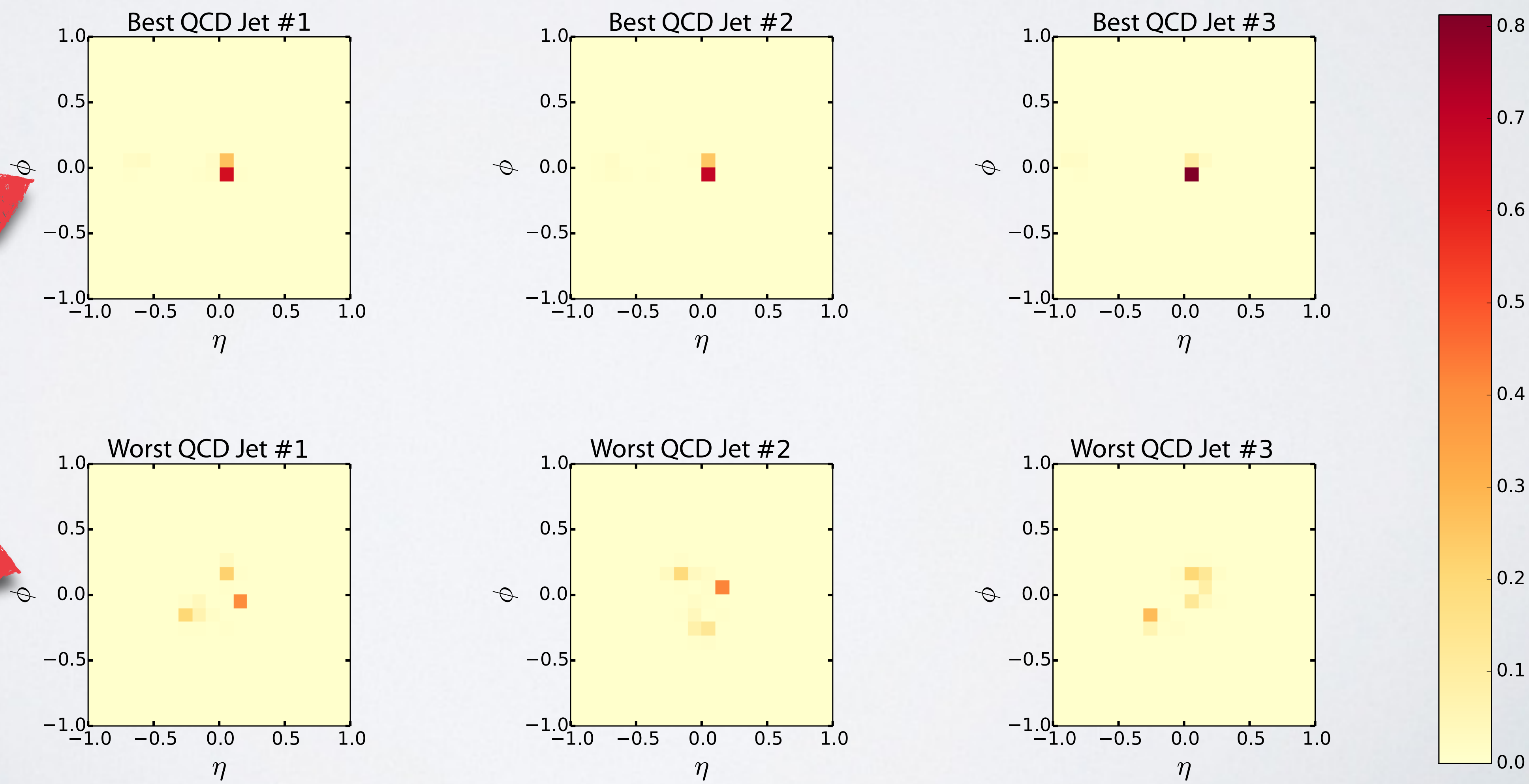
NN sensitive to “prongs" of the jet



Background outputs are also correlated.



NN sensitive to “prongs" of the jet



There are still **many things to understand** about the NN tagger…

… but, the first results are already **extremely interesting,** suggesting that image processing technology is useful in boosted jet tagging!

Many interesting things also remain to be done!

- add more information to the NN (b-tag, tracking information …).
- **What is the best input for NN**?!??
- study in more detail correlations between NN and other taggers.
- currently studying **boosted boson tagging.**
- **quark/gluon** discrimination.
- experimenting with **NN architecture**, optimization of parameters etc.
- Try other pattern recognition technology (we are about 20 years behind).
- **….**

THANK YOU!