B-TAGGING WITHOUT TRACKS: USING A NEURAL NETWORK

TODD HUFFMAN, OXFORD UNIVERSITY

THOMAS RUSSELL, CURRENTLY @ BLOOMBERG L.P.
JEFF TSENG, OXFORD UNIVERSITY

ARXIV:1701.06832

ALSO: 2016 J.PHYS. G: NUCL. PART. PHYS. 43 085001

OUTLINE

Quickly (sorry) go through the "multiplicity jump" technique

Summarize the situation from last year

New things!!

Pile-up and luminous region

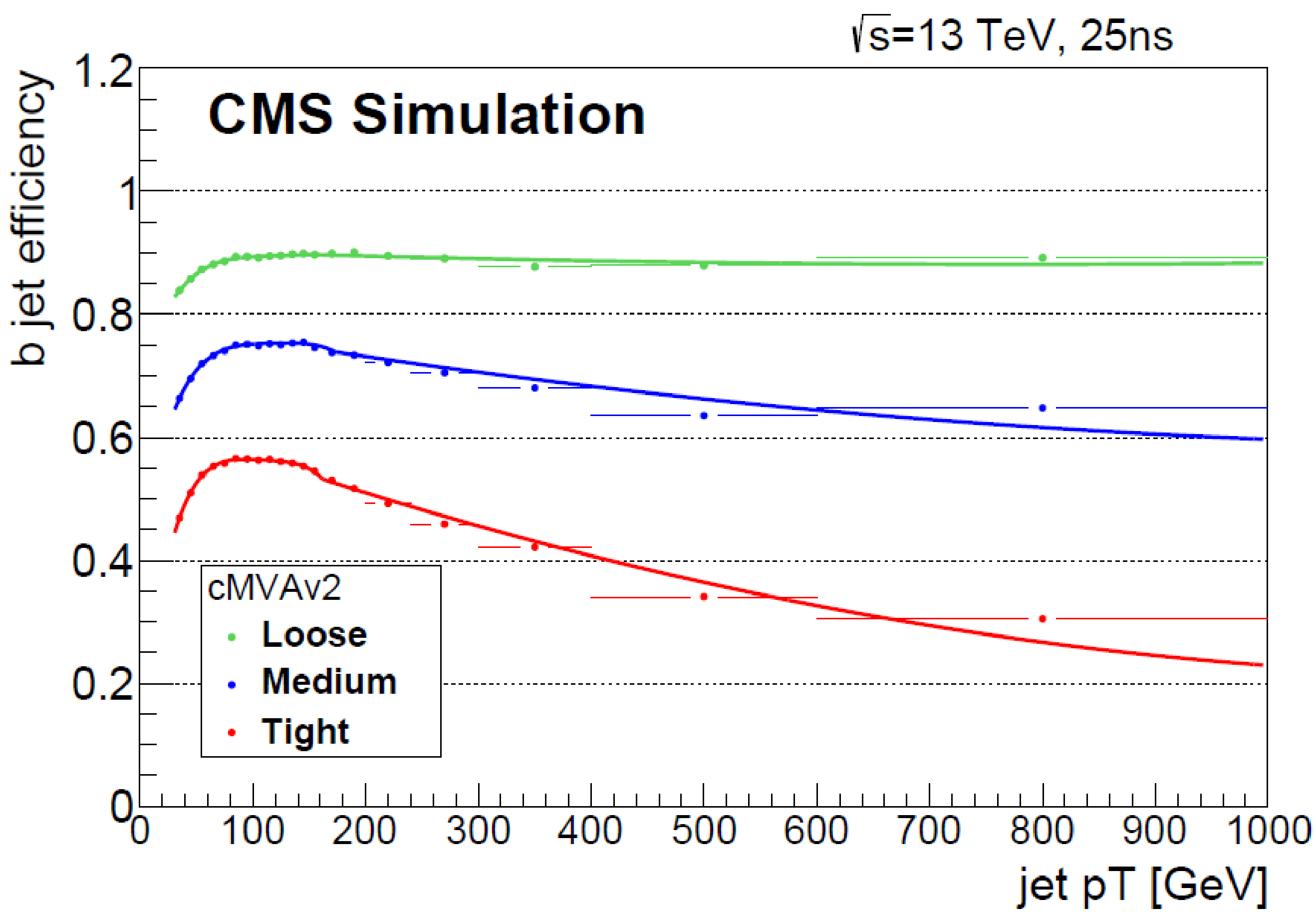
Using an Artificial Neural Net (ANN)

- (It's not "deep" enough to be a DNN)

PROBLEM: TAG HIGH P_T B-JET

Much great work happening

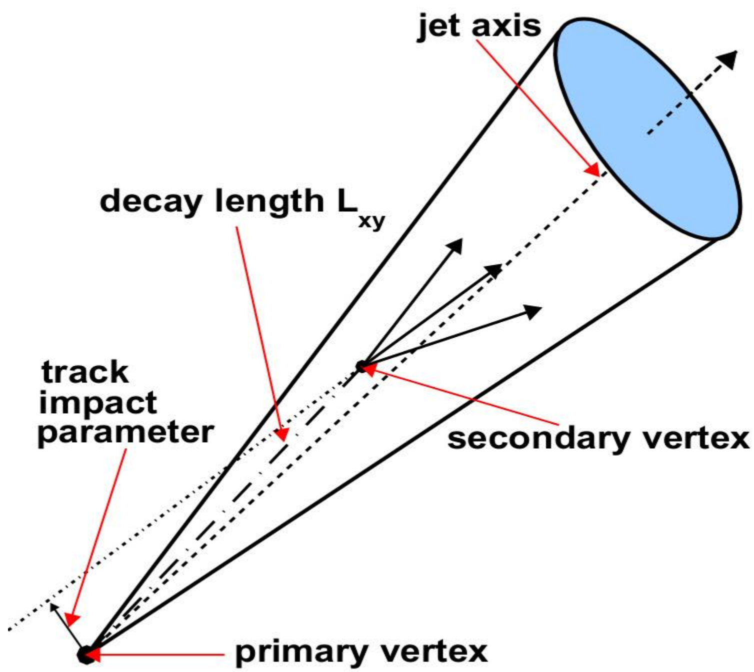
From CMS btag note
CMS PAS BTV-15-001



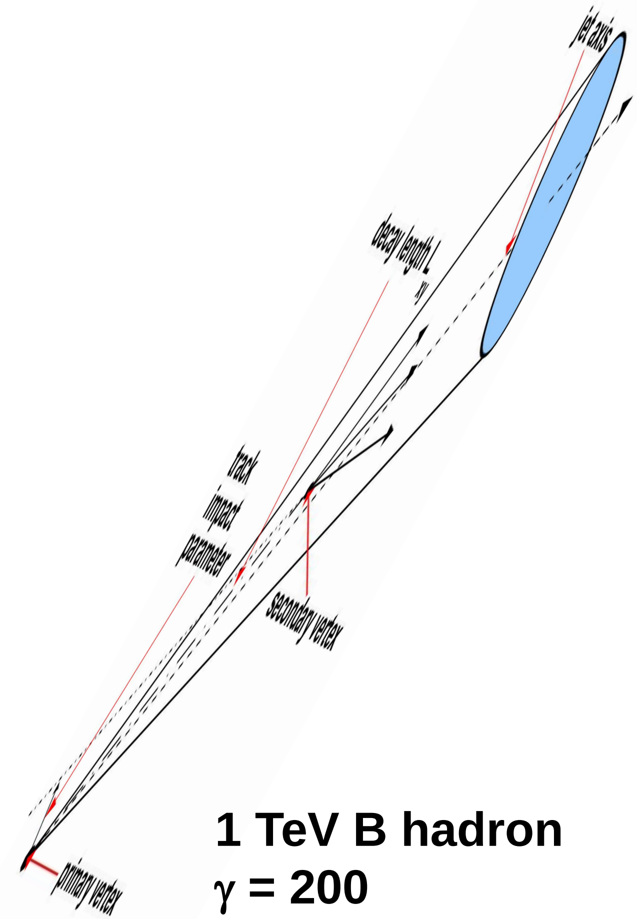
Why it happens

B-TAGGERS FIGHT RELATIVITY

Small cone sizes are prevalent.
ΔR≈0.04 for a B in a 500+ GeV Jet.

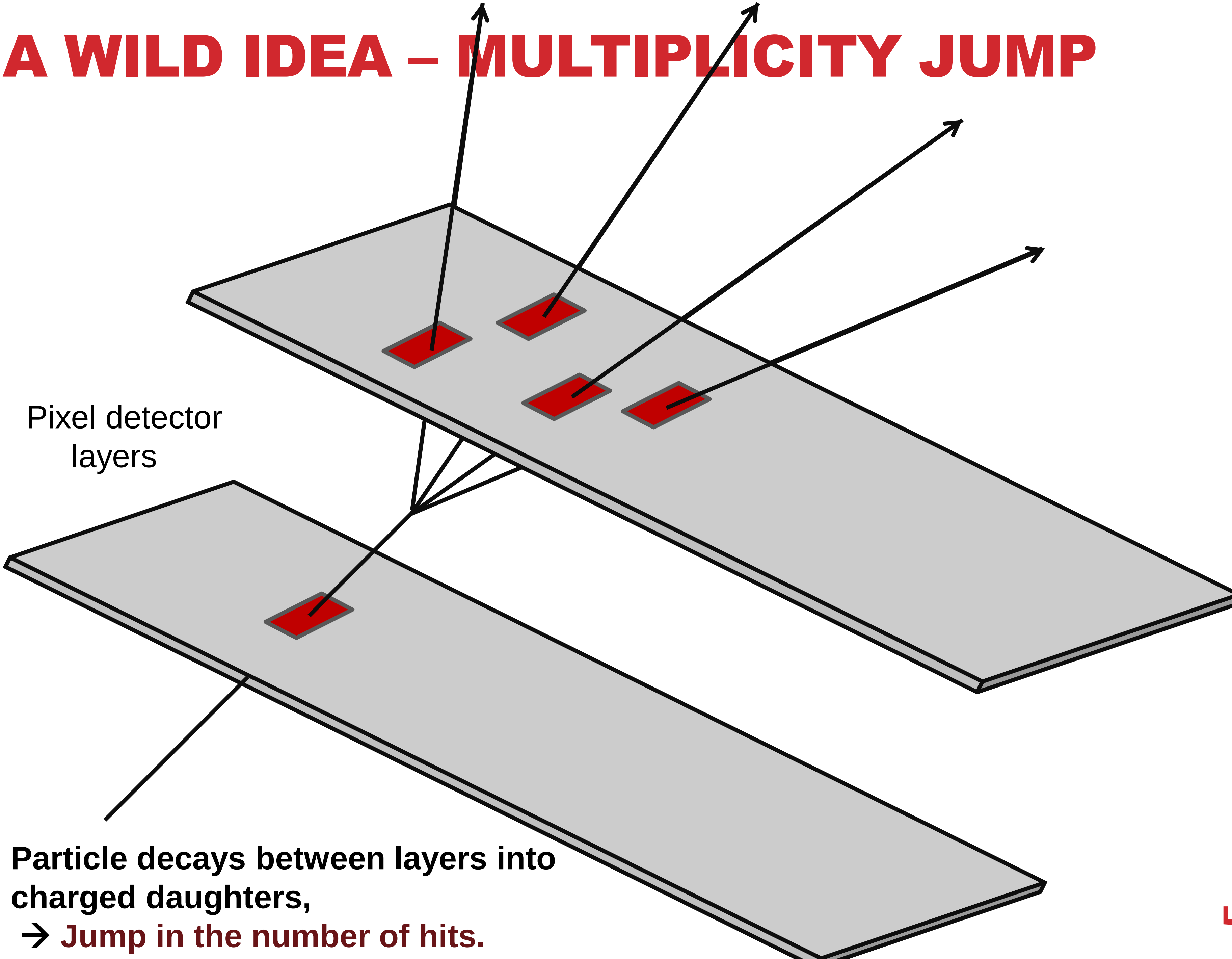


200 GeV B hadron
γ = 40
$\gamma c\tau = 18$ mm



1 TeV B hadron
γ = 200
$\gamma c\tau = 90$ mm
Radius 1st layer = 25 mm

A WILD IDEA – MULTIPLICITY JUMP



Particle decays between layers into charged daughters,
→ Jump in the number of hits.

SIMULATION DETAILS

Generator level simulation → Pythia 8

- pp collider with $\sqrt{s}$ = 13 TeV
- Generate hard QCD; $P_T > 700$ GeV
- Use EvtGen to get B hadron decays correct

Jet simulation → FastJet 3

- Anti-k_T algorithm for forming jets
- Can set jet cone size
  - We've used R = 0.2

Semi-Toy Detector simulation → GEANT4

- Volume → Cylinder 1.4 m radius filled with air, 2T mag. Field
- Silicon layers
  - Active at radii 25, 50, 88, and 122 mm
    - "Pixels" 50 x 400 x 300 μm (ϕ x *z* x *r)*
    - ***inner layer 50 x 250 x 300 μm → IBL-like***
  - *Passive cylinders 2.5 mm thick → X_0=2.5% per layer*

ΔHIT FRACTION → F_I

We *define* a quantity we call "Hit-difference Ratio" or "Hit-ratio" for short → "f_i".

- Use cone $\Delta R < 0.04$ around jet axis from Fastjet.

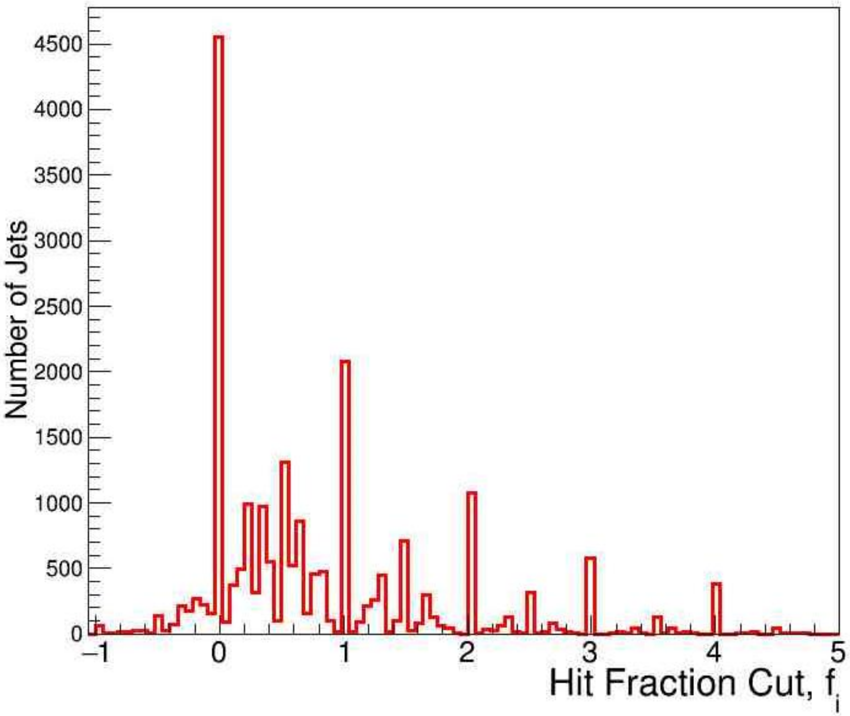
from i^th layer → $f_i = (Nhits_{i+1} - Nhits_i)/Nhits_i$

- **Can only have positive or zero hits, so:**
- **f_i** is bounded from below by -1. & unbounded from above.

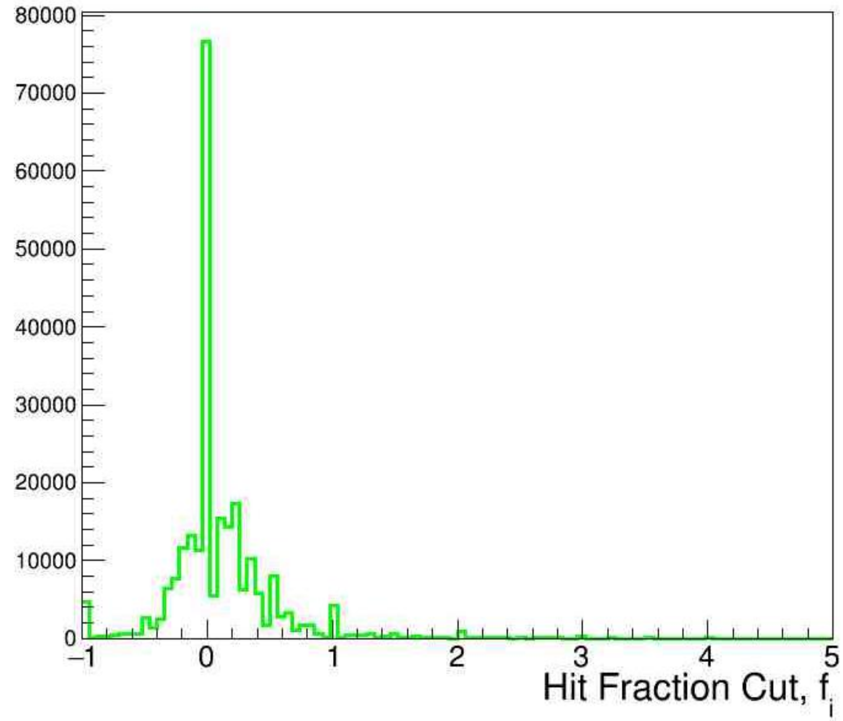
- Have a look at the f_i distribution.
  - Note: This sample → 0.5 to 2.5 TeV jets.

ΔHIT FRACTION = F_I

Jets with a B hadron
All gaps containing a decay



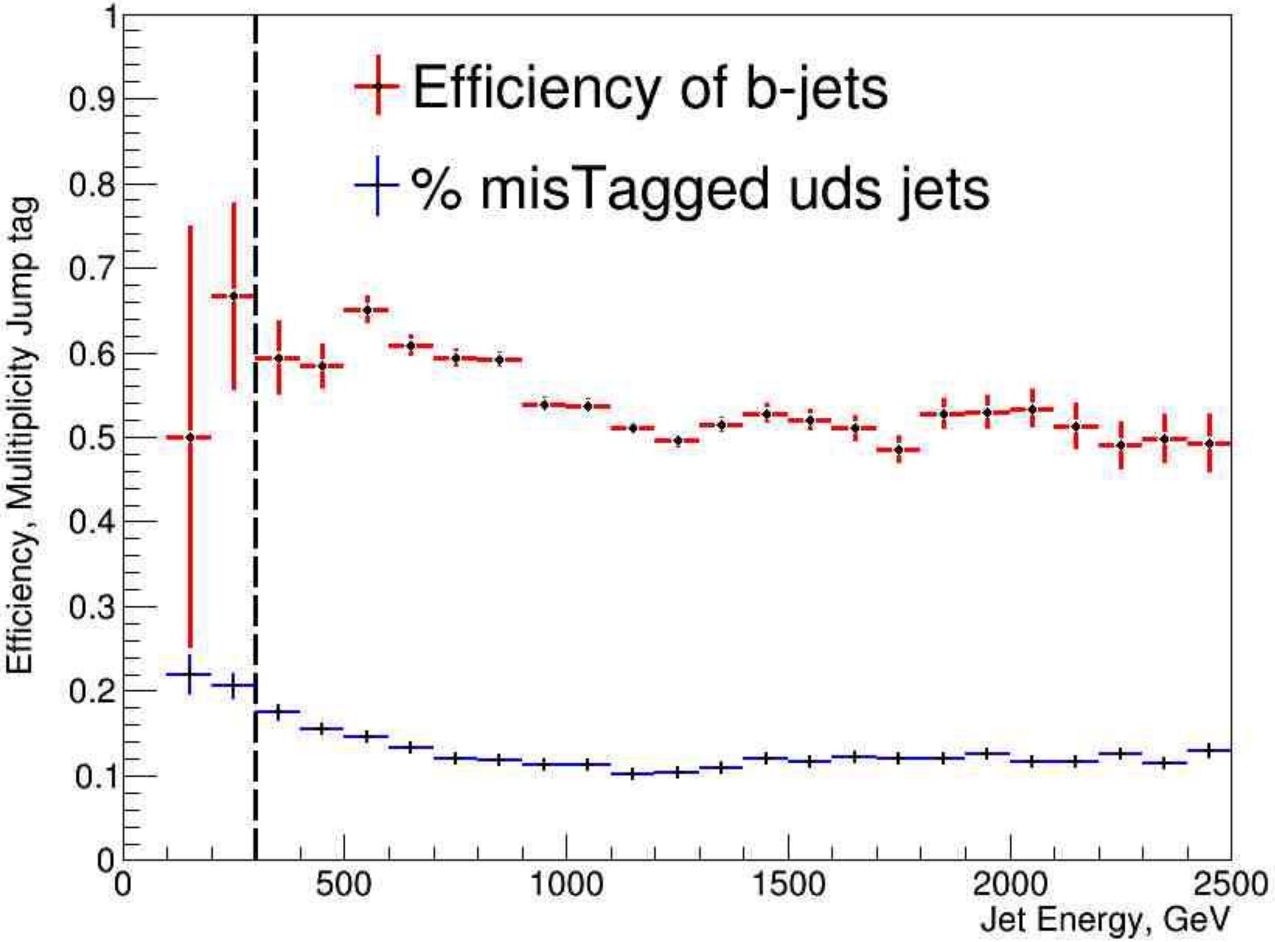
uds Jets



Simulation Z'→u,d,s,c, and b quarks only.
This looked promising → next use it as a cut variable

EFFICIENCY VS. JET ENERGY

Some uds separation – efficient for high Pt jets



NEW THINGS

Clear separation of b and uds jets

- But not what I'd call fantastic
- **Flat efficiency at TeV Jet energy – Party piece**
- No pile-up

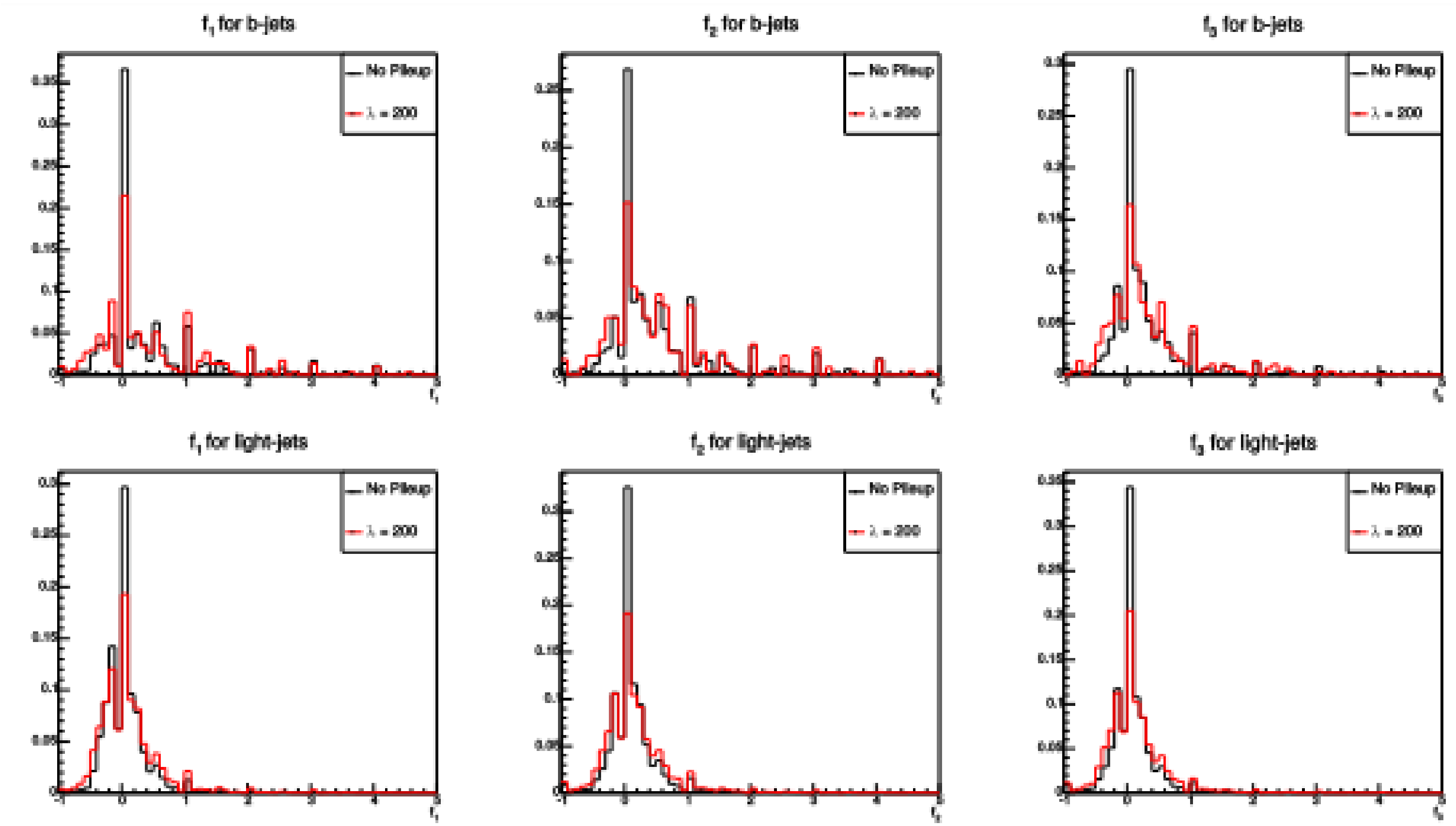
Add pile-up

- Look at up to 200 pile-up events
- Most results use 45 events (Poisson)

Try Neural Net or a Boosted Decision Tree

Pileup - f_i

How does pileup affect our f_i?



Pileup - Conclusion

- Tagger insensitive to increased pileup (tested to pileup with $\lambda = 200$),
  - little change in $f_i > 0$
- Makes sense - most pileup does not result in hits in R<0.04 between layers.

NOTE: All plots now have pileup with $\lambda = 45$.

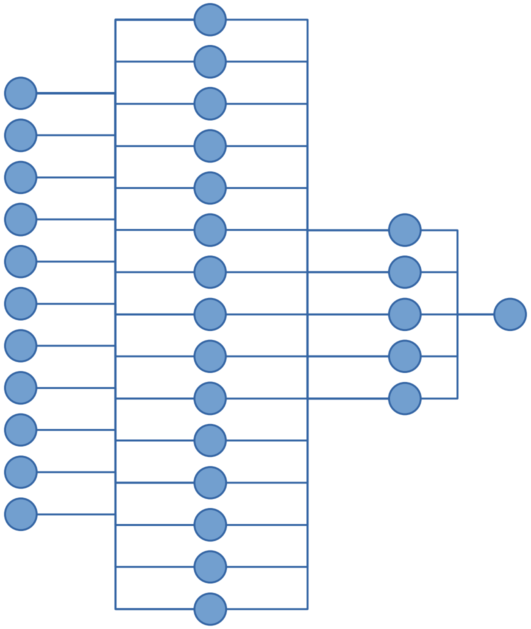
ARTIFICIAL NEURAL NETWORK

Building the ANN

ROC analysis showed that 8-layer ANN was as effective as a 2-layer ANN → not really a "deep learning" problem, Reducible, small feature set.
Fewer node layers means less chance of overfitting.
Used TMVA for a "standard HEP" packaged ANN.
Activation function was tanh(x).

Input Layer:

- Raw hits in each layer
- Multiplicites f_1, f_2, f_3
- max(f_1, f_2, f_3)
- Jet Energy, p_T and mass



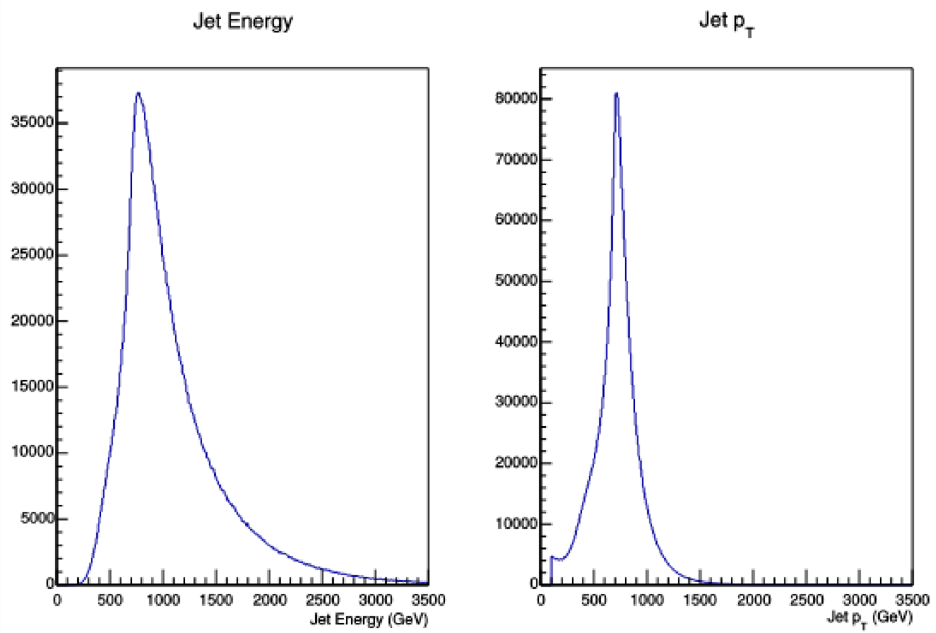
2 Hidden layers:
16 and 5 neurons wide.

(Smaller second hidden layer helps prevent overfitting).

Architecture → educated guessing.
Too many combinations to evaluate

Training Sample

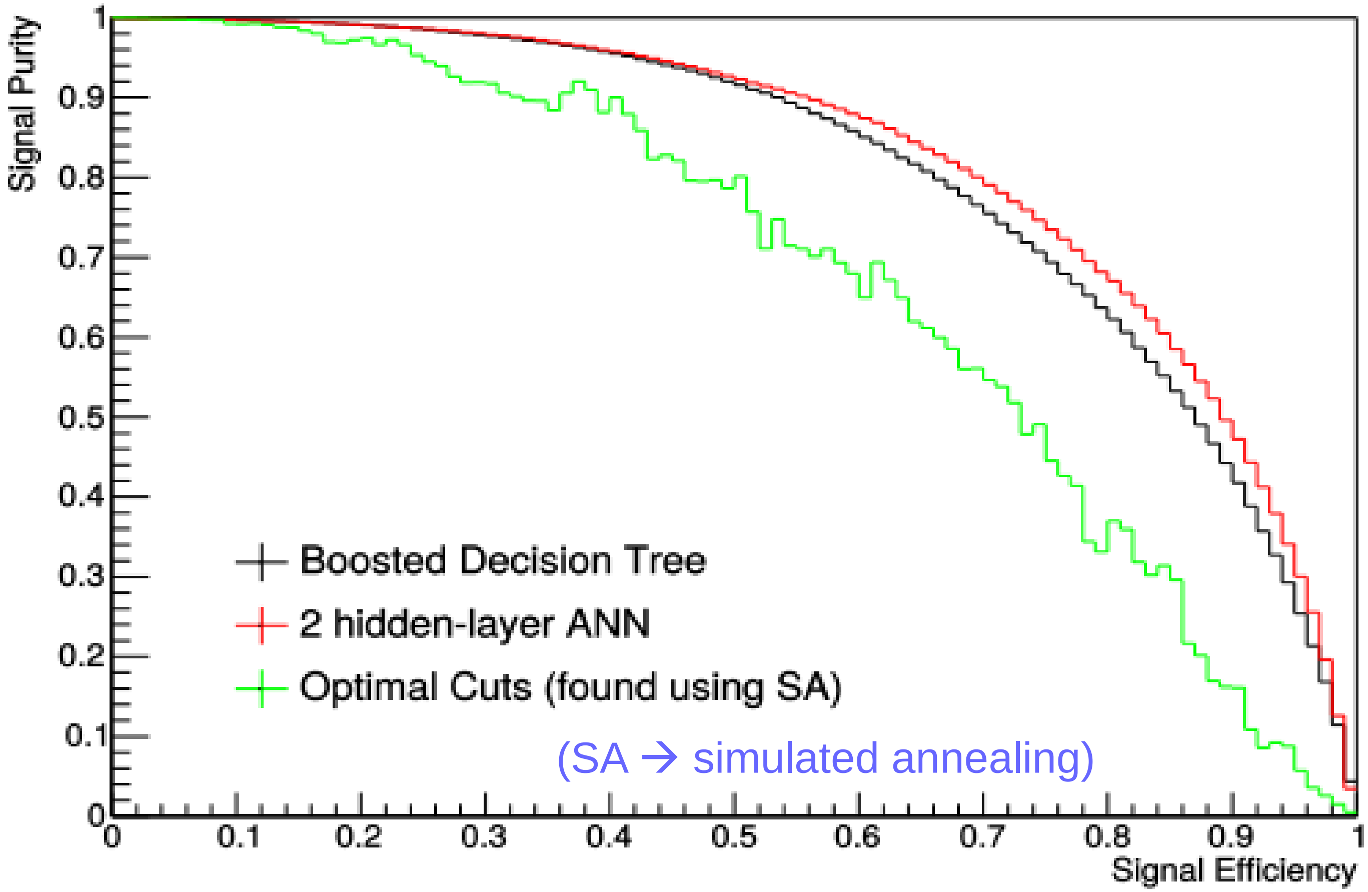
- Training sample had min p_T of interaction at 700GeV.
- 1 million general hard QCD events (anything goes), 300,000 hard QCD events resulting in B-hadron production.



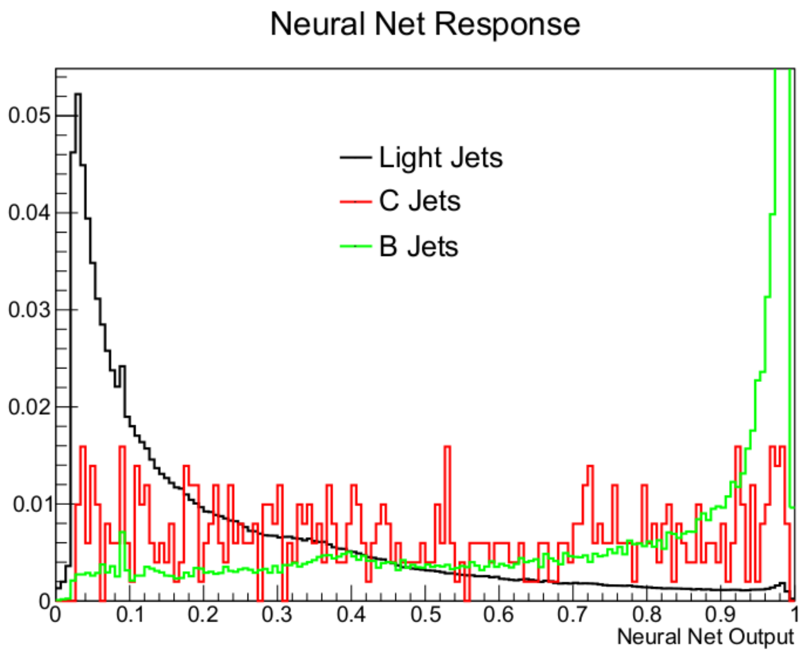
Testing is then done on an independent hard QCD sample.

Choosing a ML technique

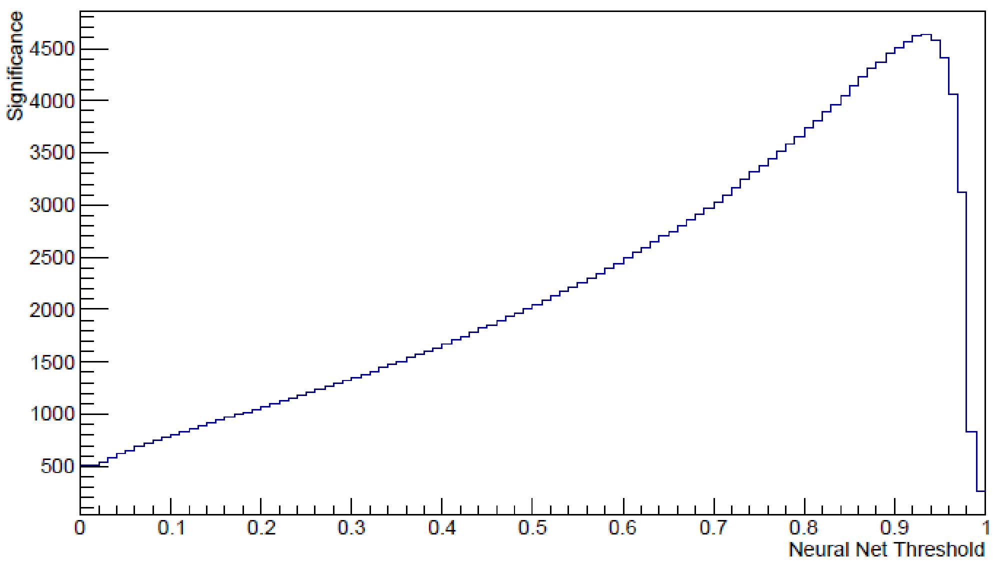
- Each ML technique has a distinct set of advantages/disadvantages.
- ROC curve (Signal Purity vs. Signal Efficiency)



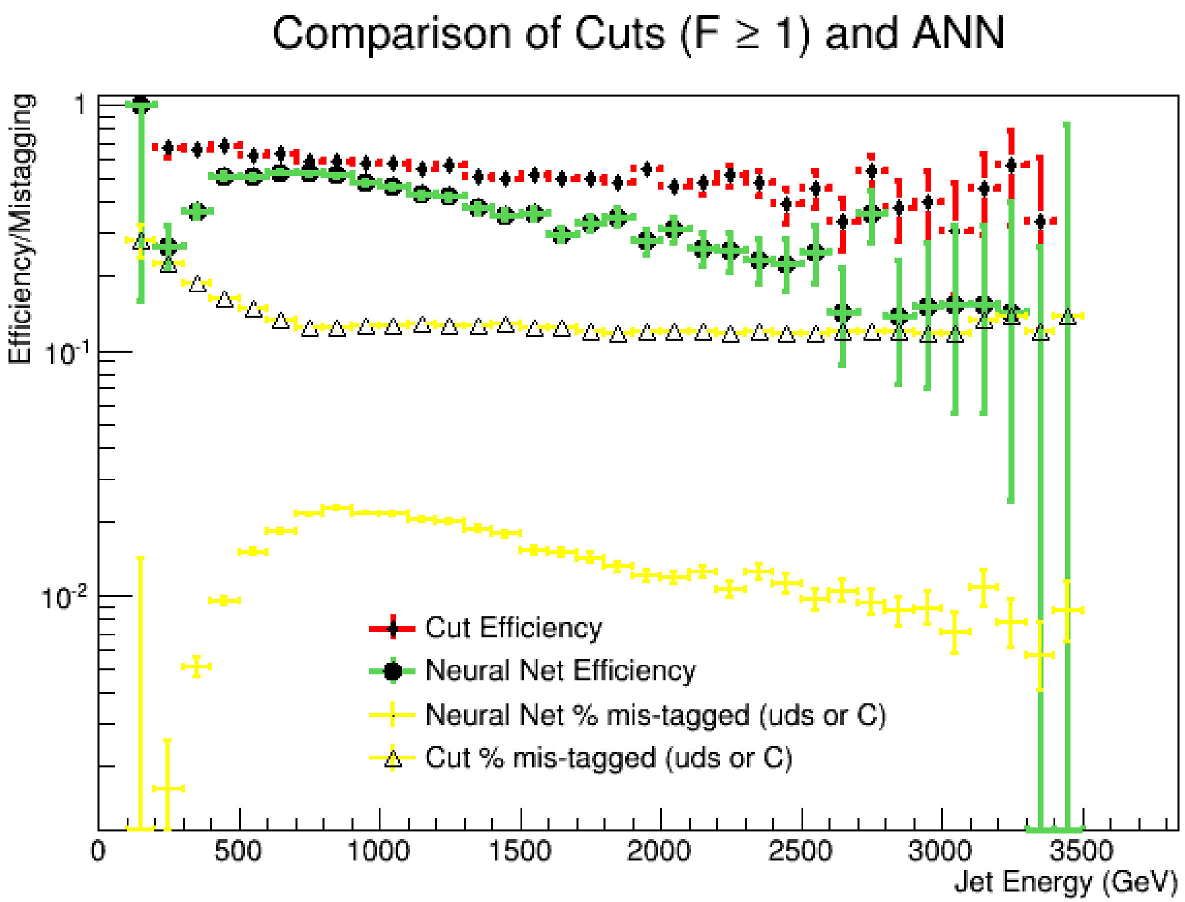
ANN TEST MC RESULTS



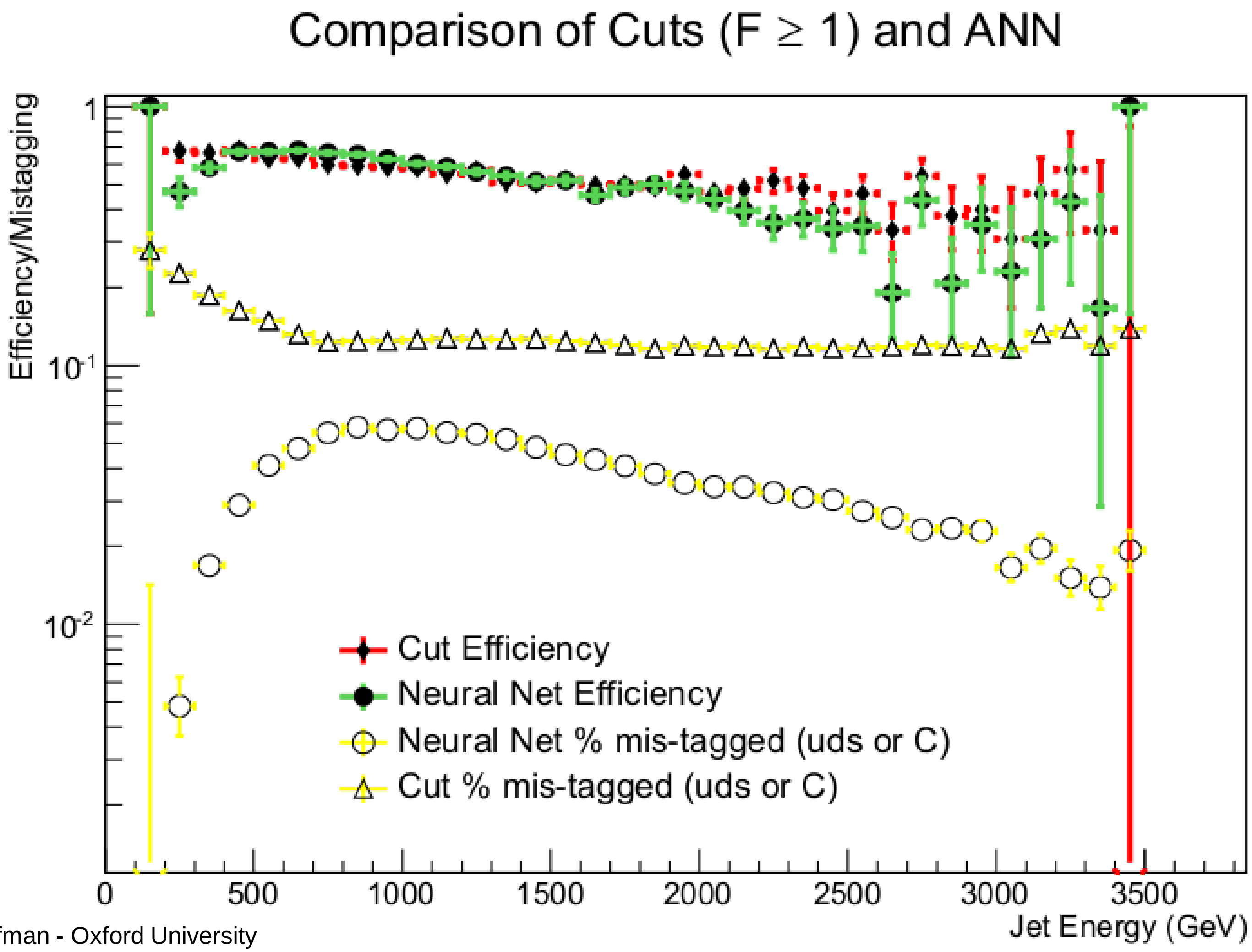
Significance = $\varepsilon_b / \sqrt{\varepsilon_{uds}}$



WITH → ANN OUTPUT>0.9

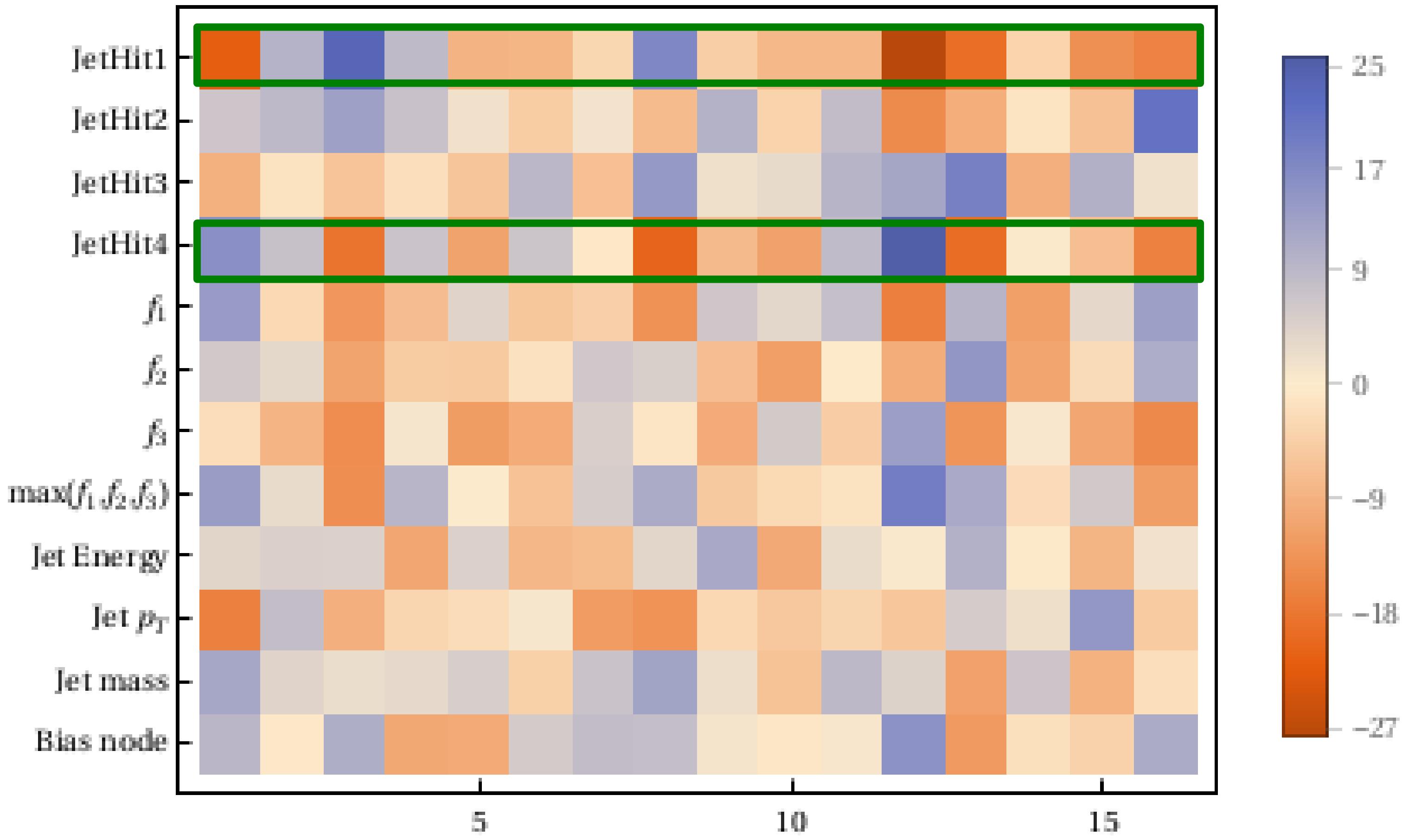


WITH → ANN OUTPUT>0.75



HEAT MAP

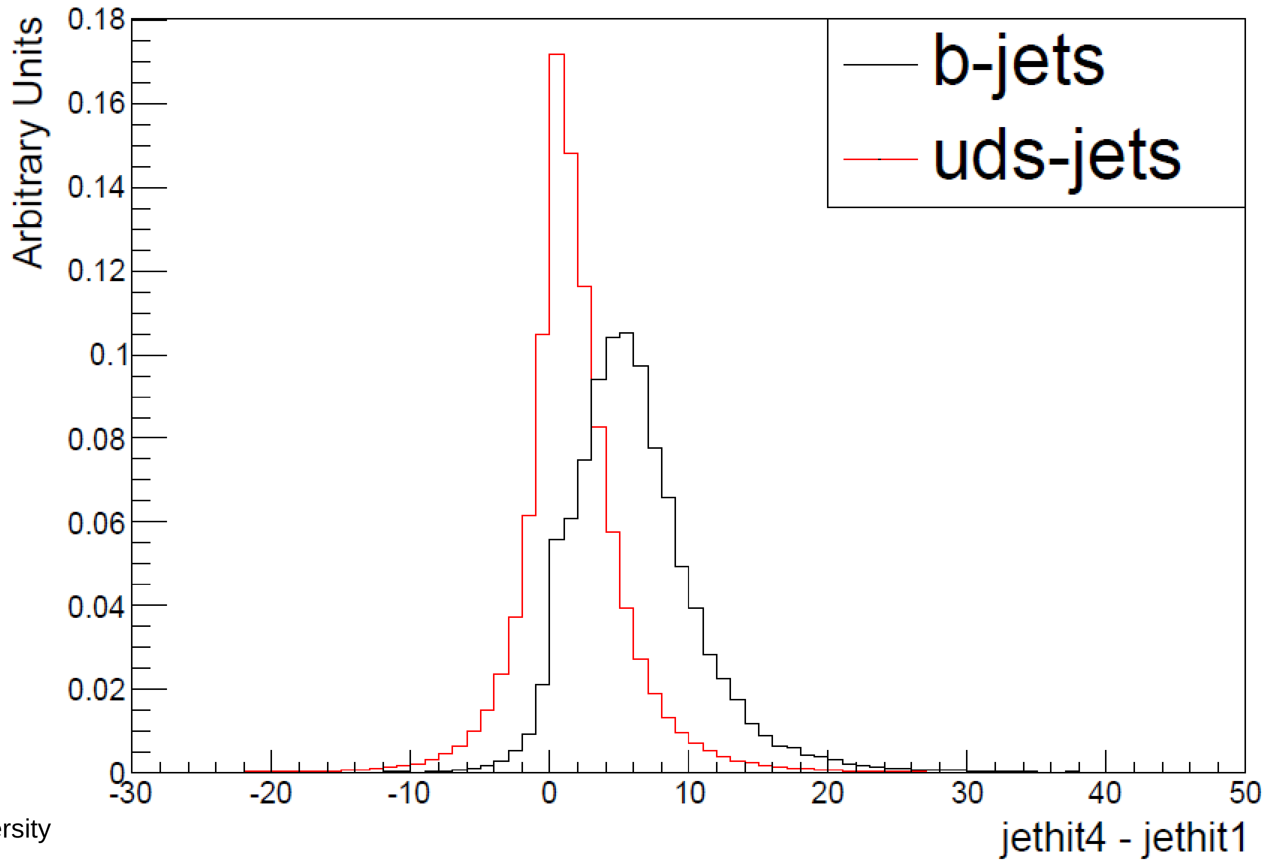
Neural Net seems to find Nhit4 – Nhit1 → good discriminator



ANN INSPIRED

Straight hit difference between layer 4 and layer 1.

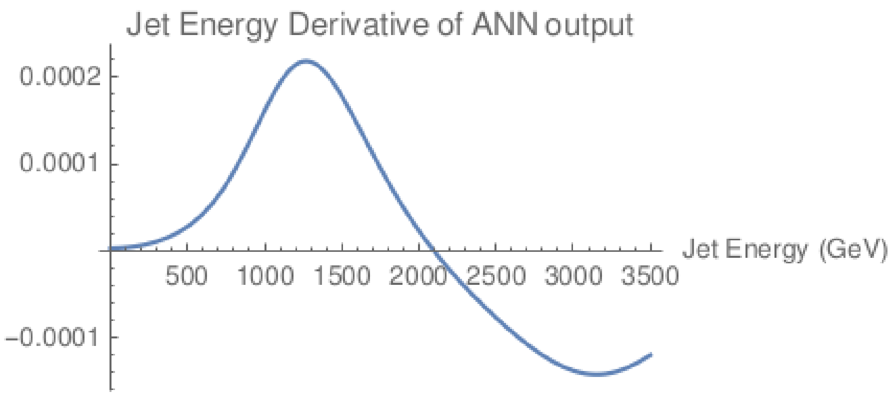
Definite advantage can be seen.



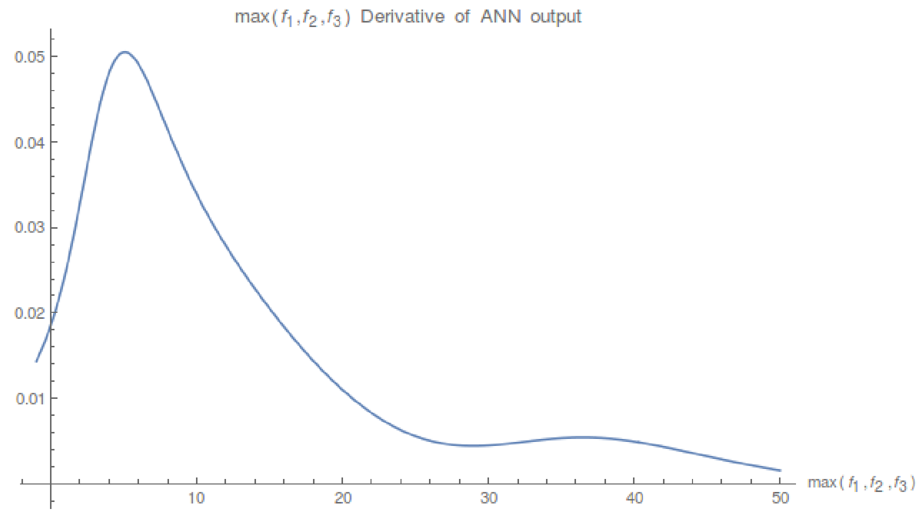
ANN OUTPUT DEPENDENCE
(EXAMPLES)

NN is just a complicated function $h(x_j)$ – Look at $\Delta h/\Delta x_j$

JET ENERGY



MAX(F)



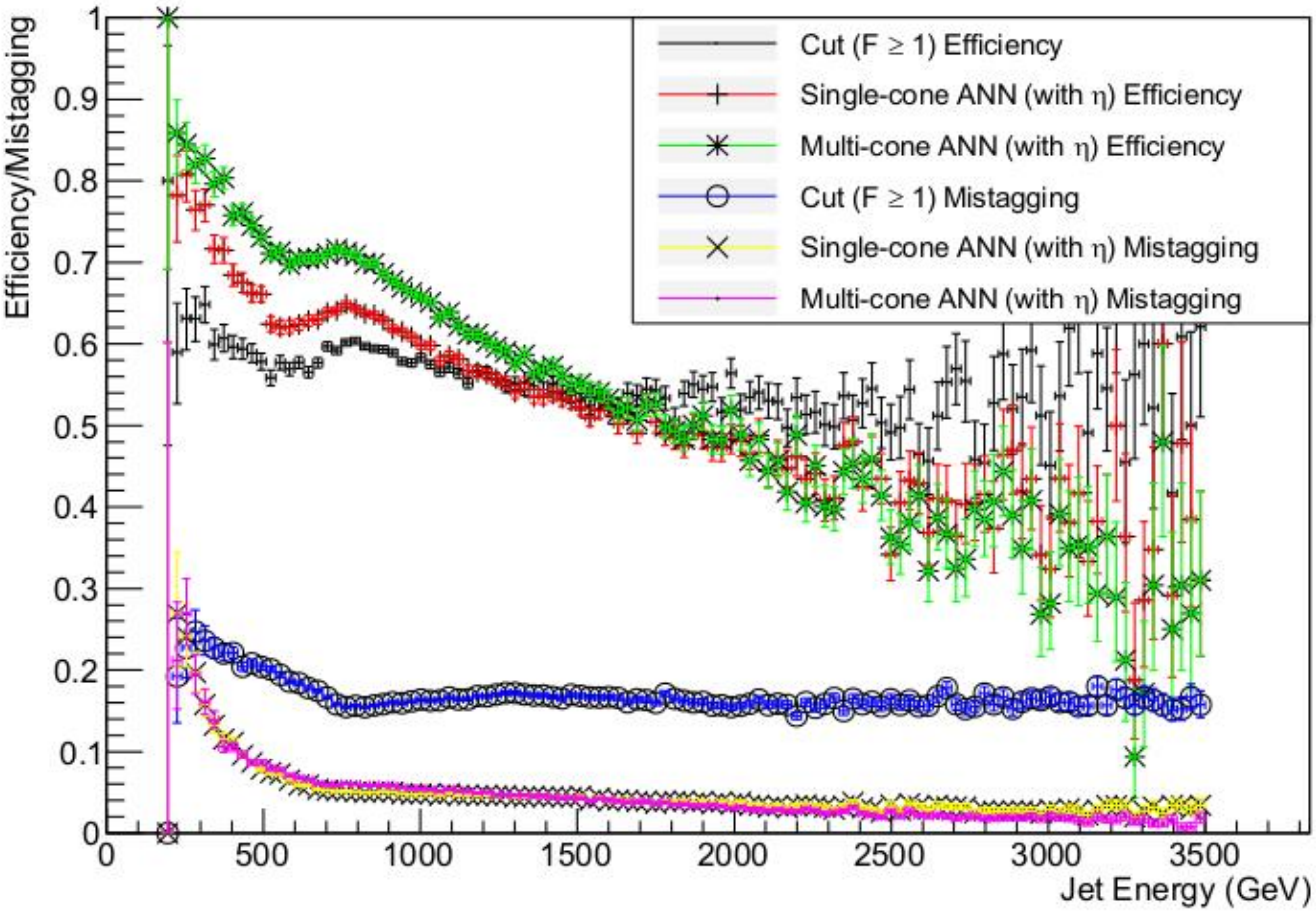
Trained NN output is ~100 times more sensitive to Max(F) than to Jet Energy

Conclude: The Machine is focused on the "right" types of quantities.

LAST THING – MULTIPLE CONE SIZES

Include more B hadron decay products @ lower Energy

$\Delta R<0.04$, $\Delta R<0.1$ and $\Delta R<0.2$ added
Each with own set of inputs



Training sample
Shape dependence?

CONCLUSIONS

Using Mulitvariate techniques significantly improves rejection of uds jets.

Results robust to pile-up

Multiplicity Jump technique more immune to searchlight effect and highly boosted B hadrons.

If continues to show promise, could be added to existing taggers to enhance efficiency at TeV scale jet energies.

Experiments (ATLAS, CMS) are looking into the technique with full simulations → Clear next step.

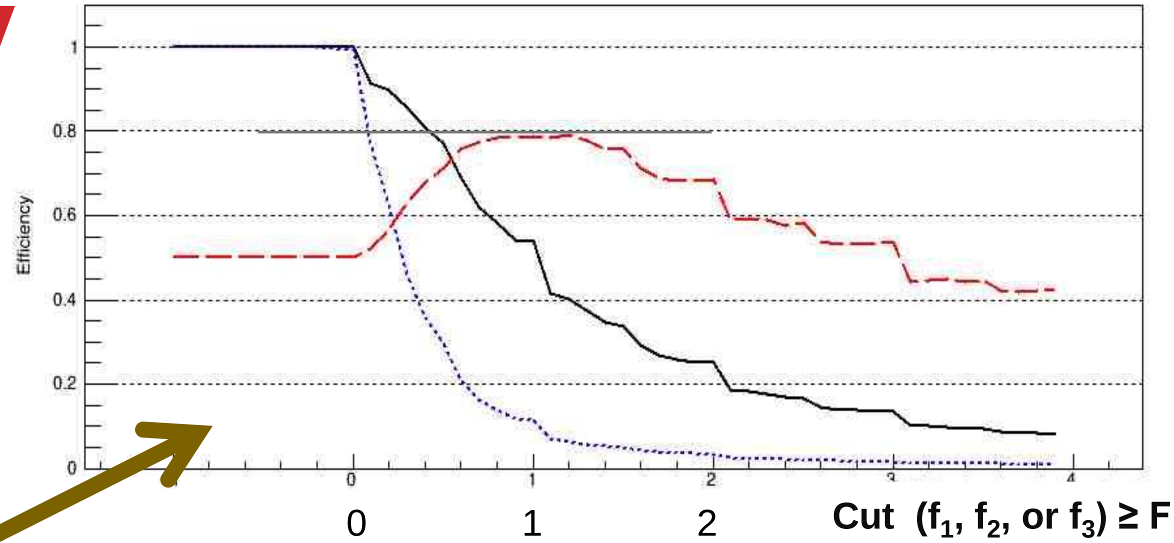
HAND OFF TO ATLAS & CMS!!

THANK YOU! (BACKUP SLIDES)

F_I EFFICIENCY AND PURITY - LAYERS 1234 $M_{Z'}$ = 2.5 TEV

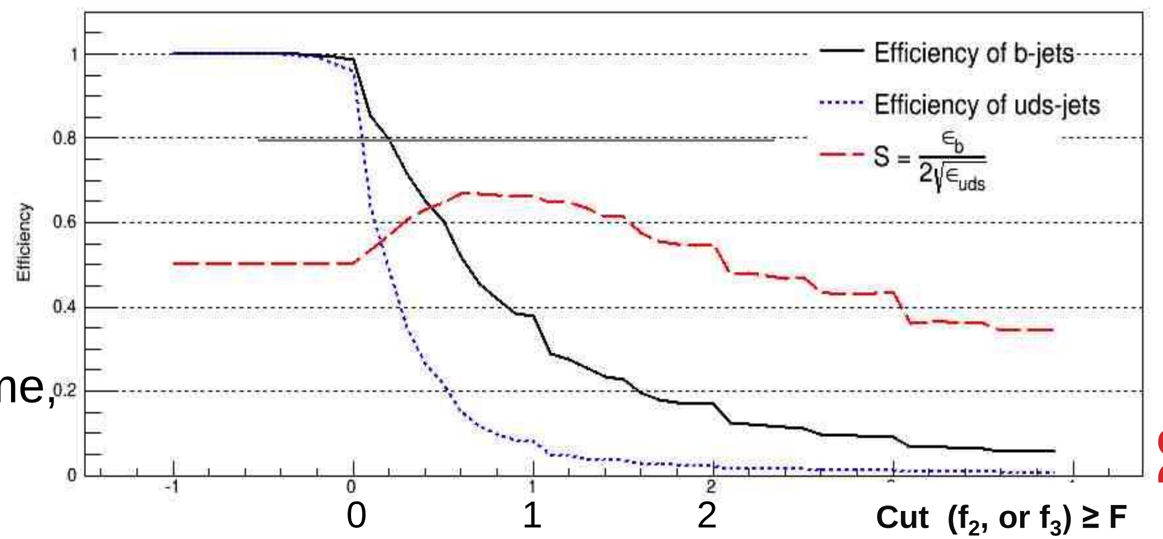
$f_i \geq F=1$

Layers:
1-2 or
2-3 or
3-4



Choose this one!!

Layers:
2-3 or
3-4



Fiducial region → Whole volume,
So can compare.

APPLYING ΔHIT$_F$ CUT

Start at f_i= -1.0 (i.e. no cut at all)

- And Start increasing the cut.

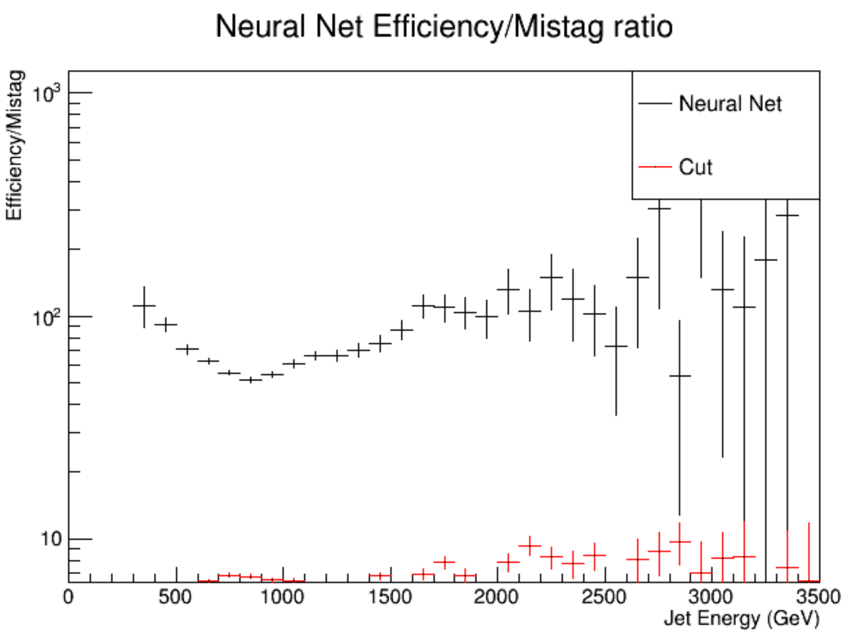
At each cut value, Plot (Number of Events passing cut)/(number of starting events)

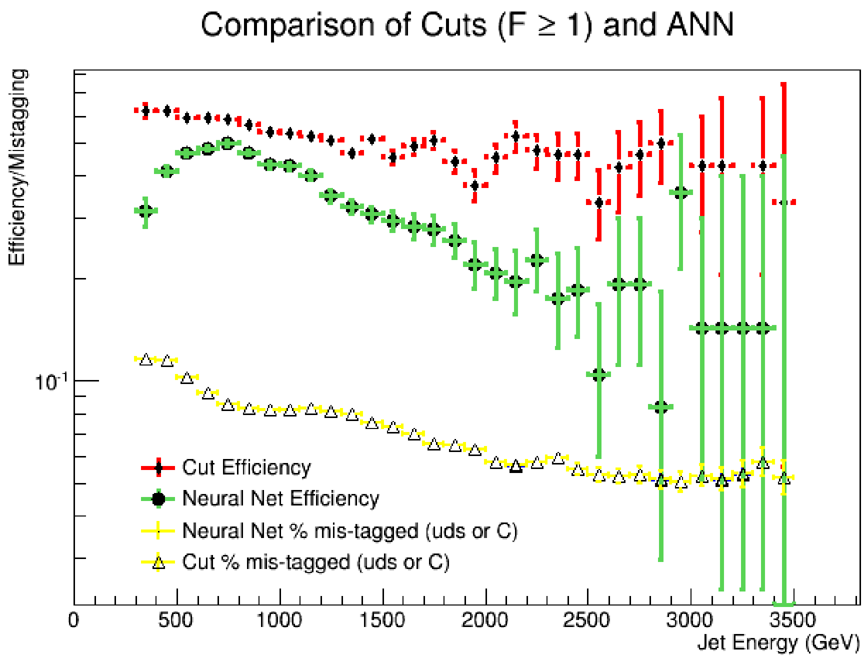
NOTE! Only count B hadron jets where B decayed inside the layers!

"ALL Layers" plot is logical OR of individual layers

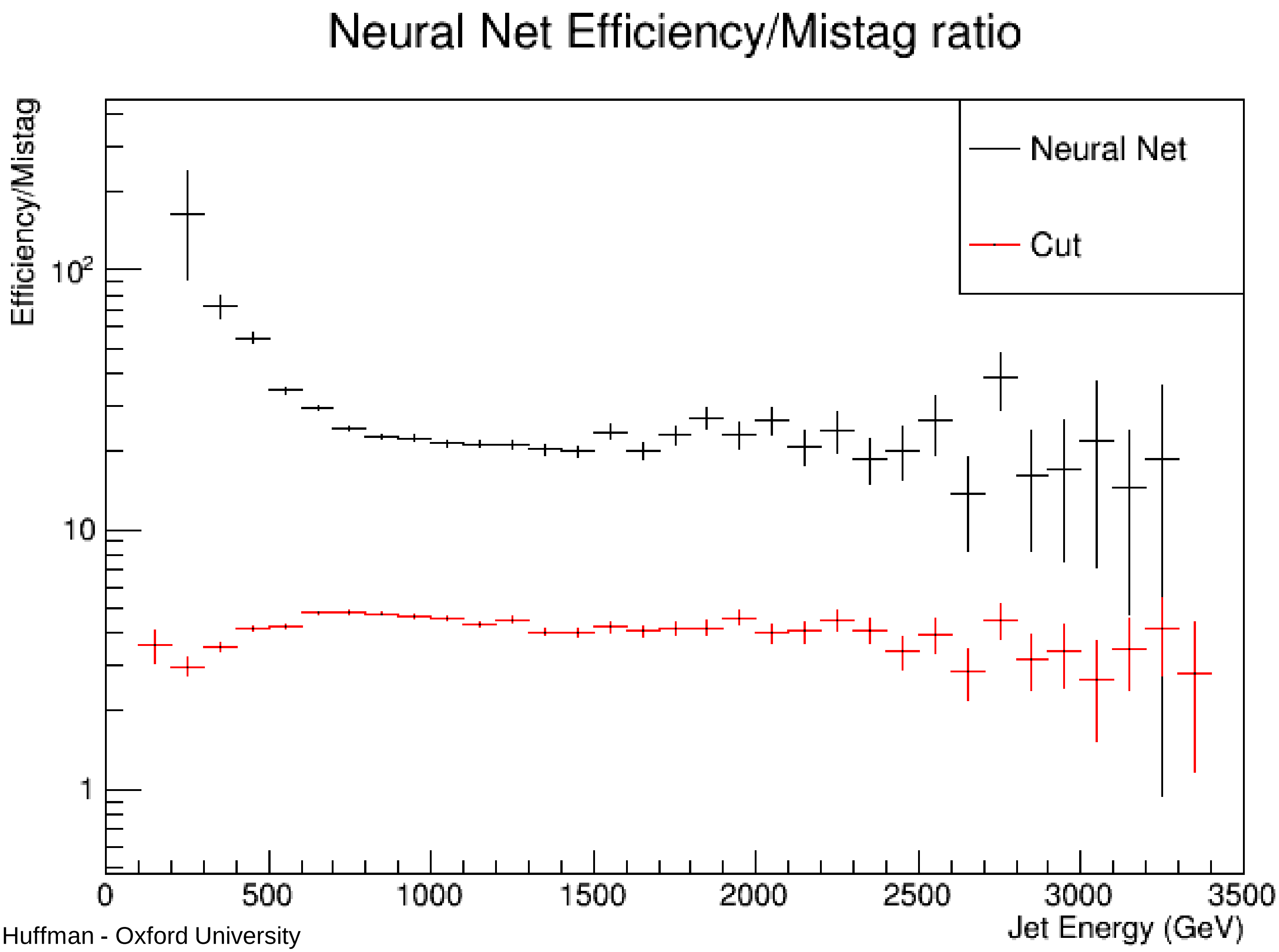
- if any one of the ΔHit$_f$ between any pair of layers passes the cut, the event passes.

DEAD MATERIAL REMOVED – NO EARLY SHOWERS





ANN - STRAIGHT CUT COMPARISON

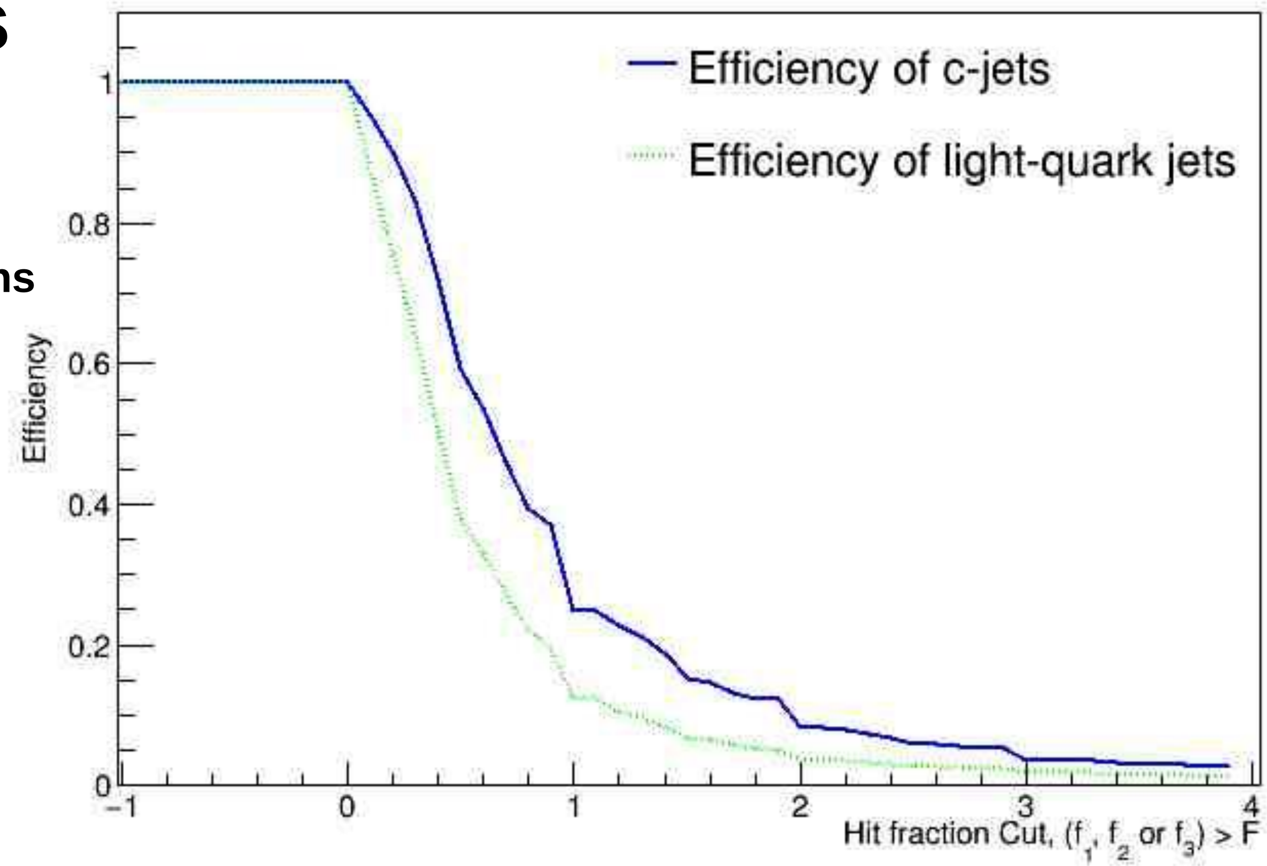


MAX(F_1,F_2,F_3) > ΔHIT

Completeness
Charm jets

2.5 TeV

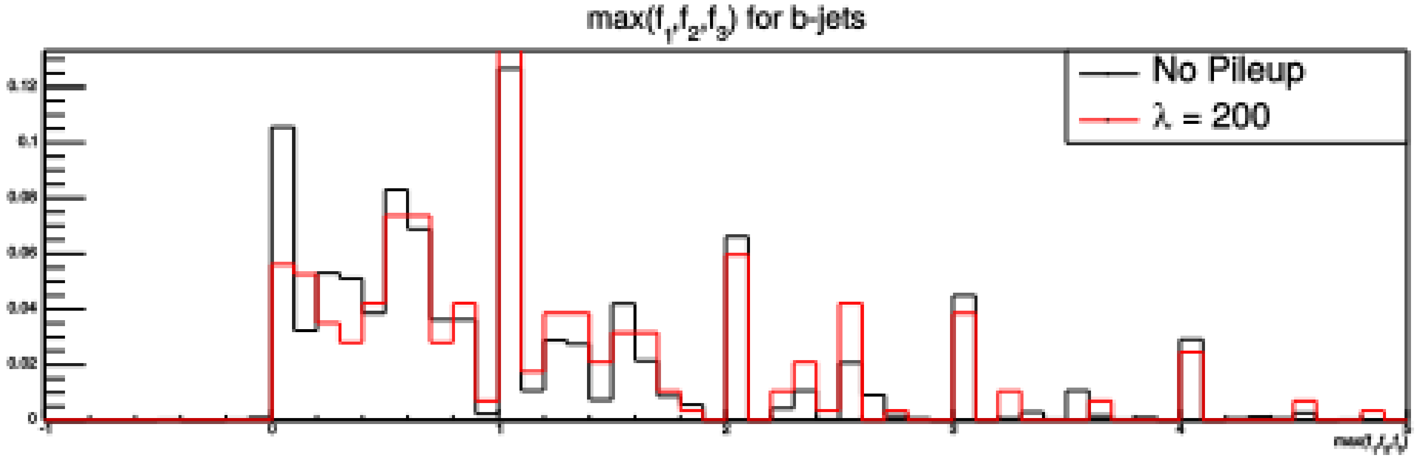
Same set of conditions

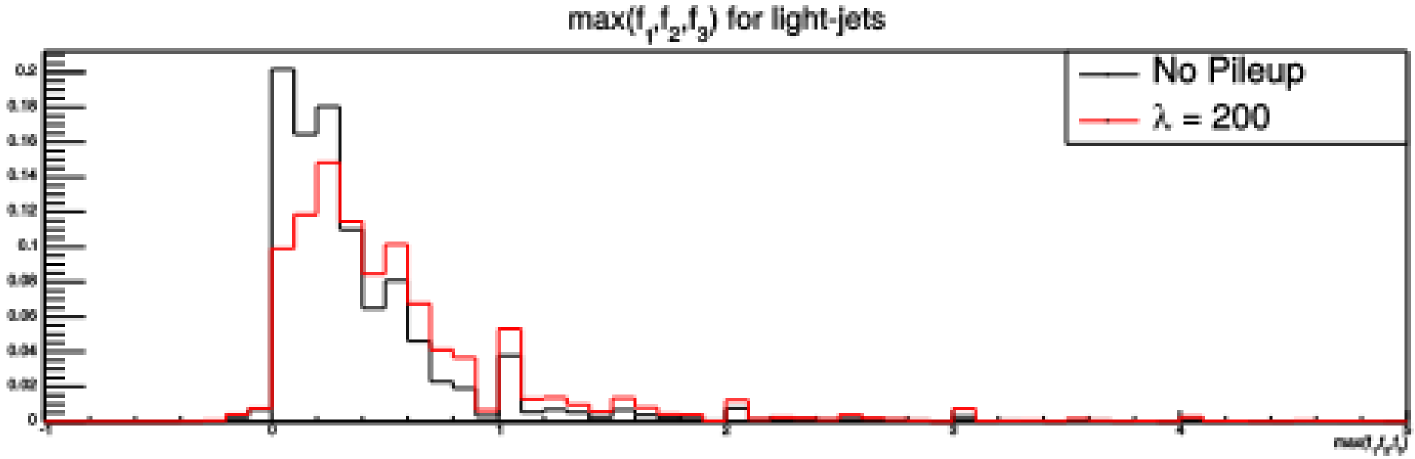


Upgrading the simulation

Making sure it works in a more realistic scenario

Pileup – max(f_1,f_2,f_3)





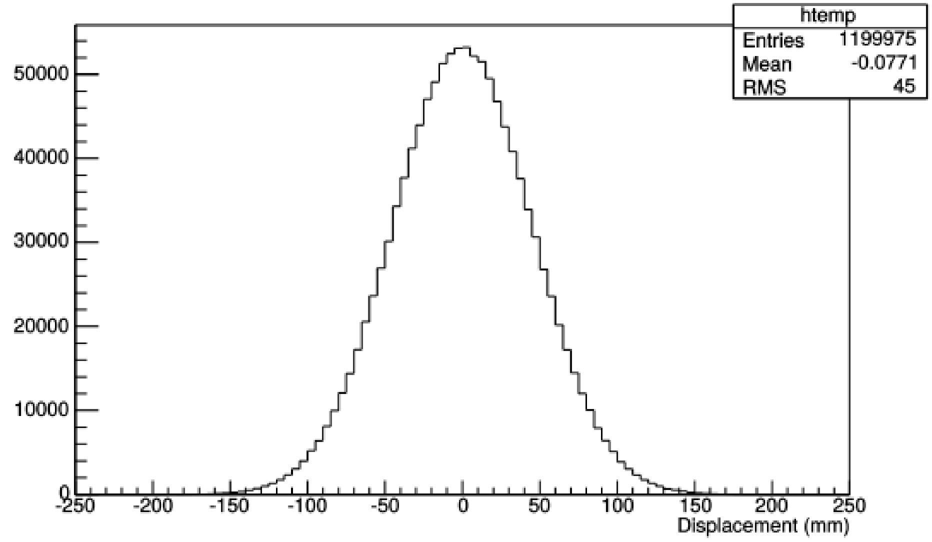
Luminous Region

Events do not always occur at (0,0,0) – modify GEANT4 sim to account for this.
ATLAS has σ_z= 45mm

New Problem – Jet vectors no longer from (0,0,0), have to be careful when selecting pixels.

Easily resolved by finding the event primary vertex. Once this is done, luminous region vs origin makes no difference to the quality of the tagger.

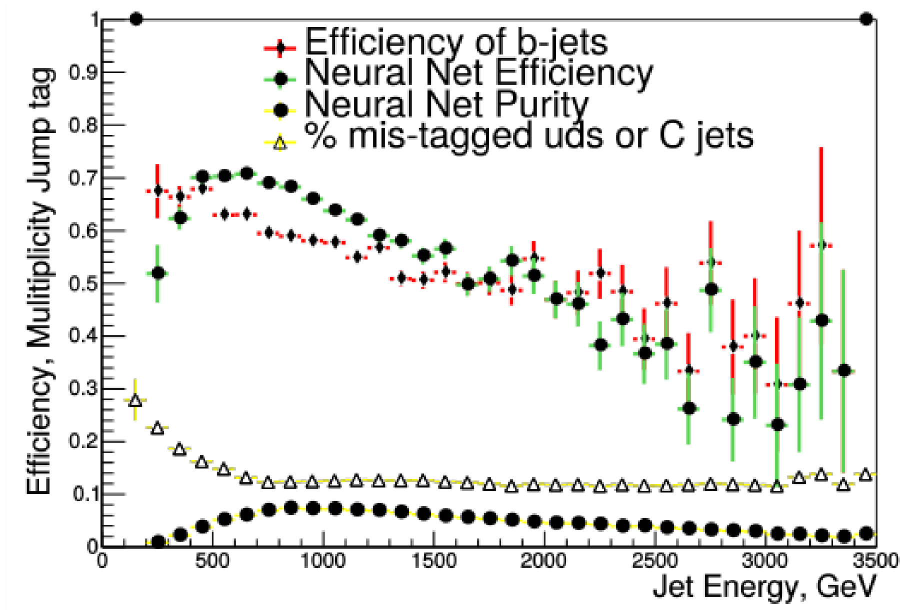
Vertex Z Displacement

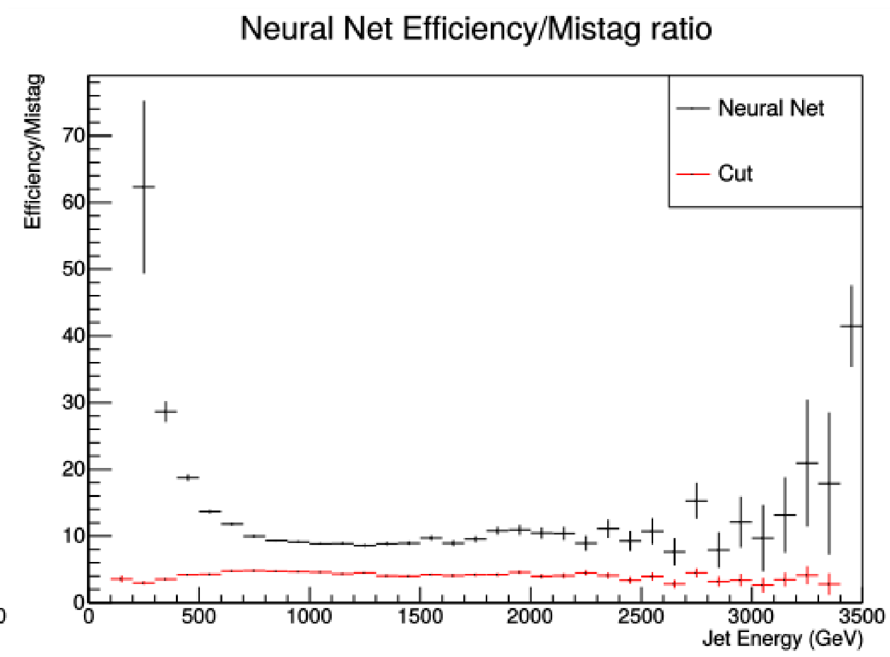


Upgrading the Tagger

We know it works… Let's make it work better!

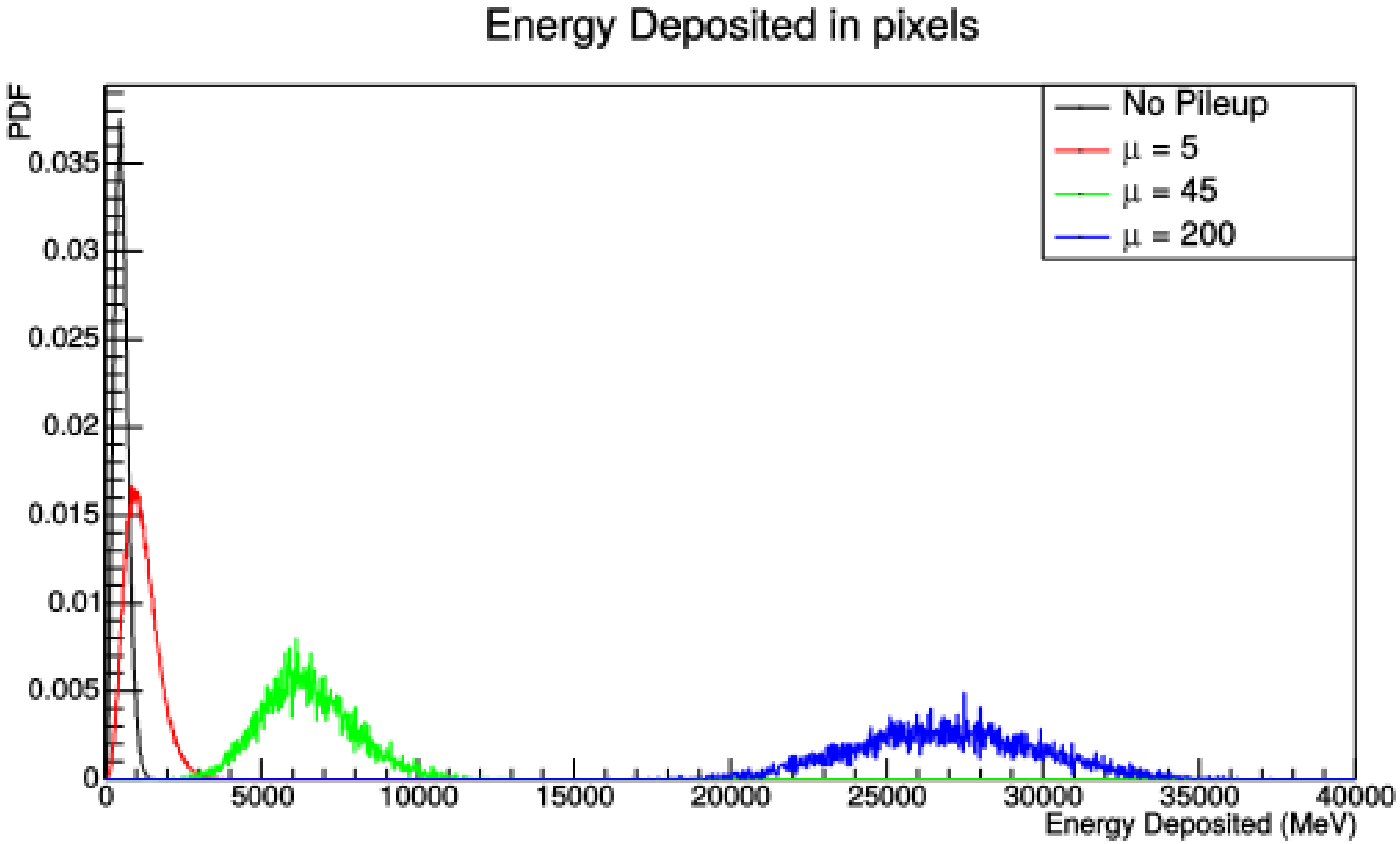
How does it do?





Neural Net Wins!!

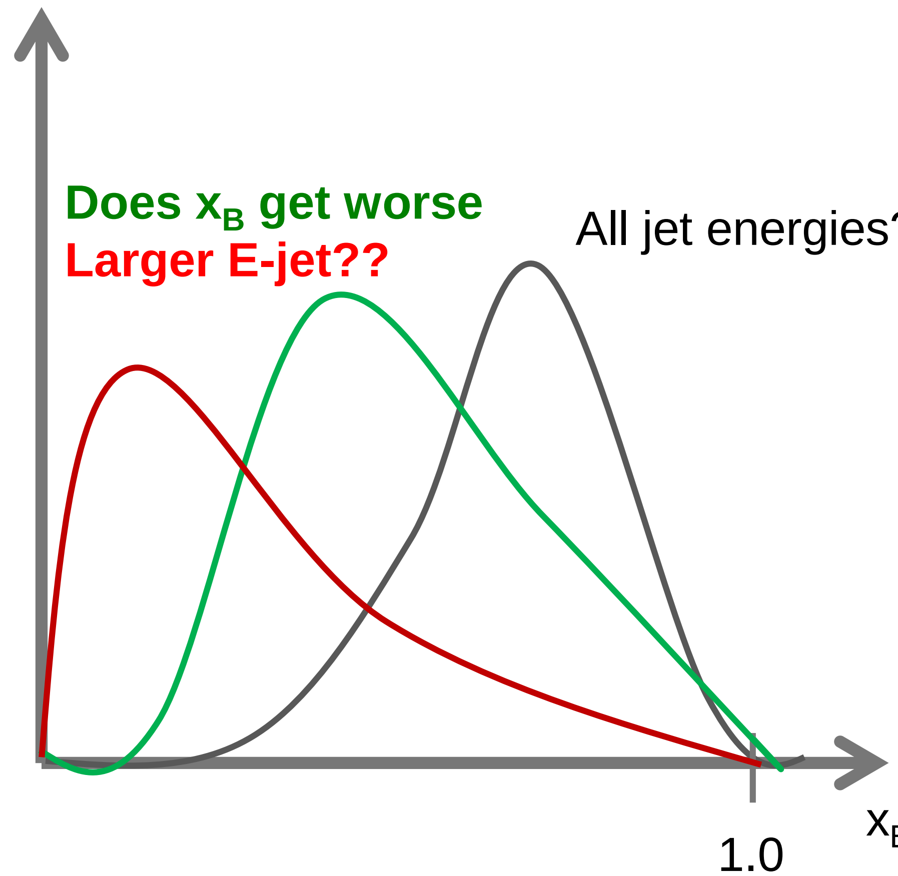
Extra Slides


Energy Deposited in pixels
PDF
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
5000
10000
15000
20000
25000
30000
35000
40000
Energy Deposited (MeV)
No Pileup
μ = 5
μ = 45
μ = 200

ENERGY-FRACTION OF B BARYONS? DOES IT DEPEND ON JET ENERGY?

Energy fraction of B baryons as Jet energy increases

- x_B Logarithmic?
- x_B Non-linear?
- x_B Constant?
- If Jet energy → more tracks
  - Helps taggers



Let's at least find out what simulations say!