Tag Jet Identification

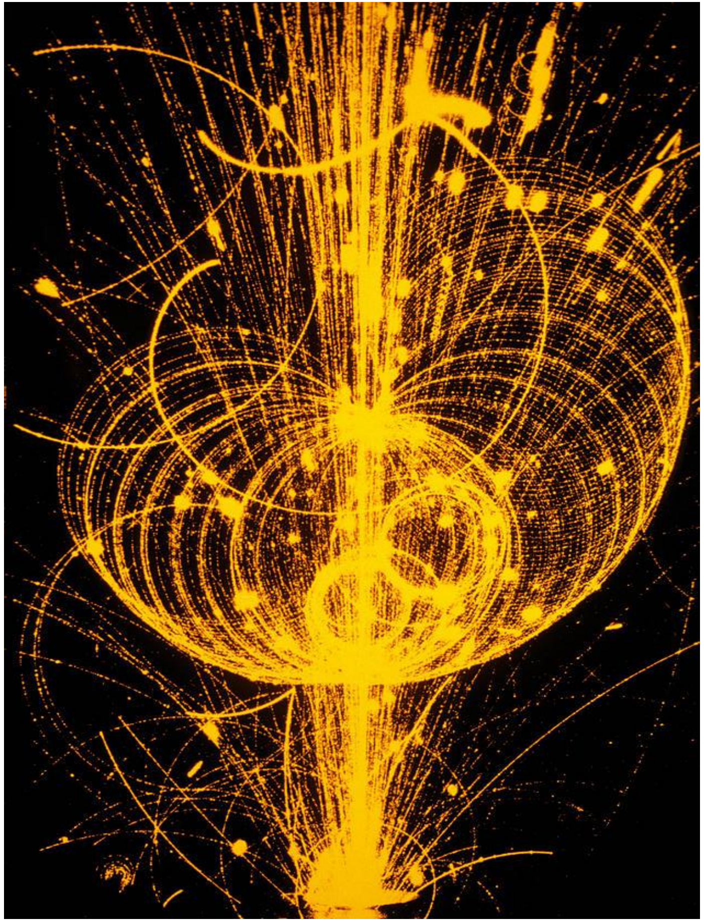
Through the Use of Deep Neural Networks

Anne-Katherine Burns

Machine Learning Seminar, Jefferson Lab

November 6, 2018

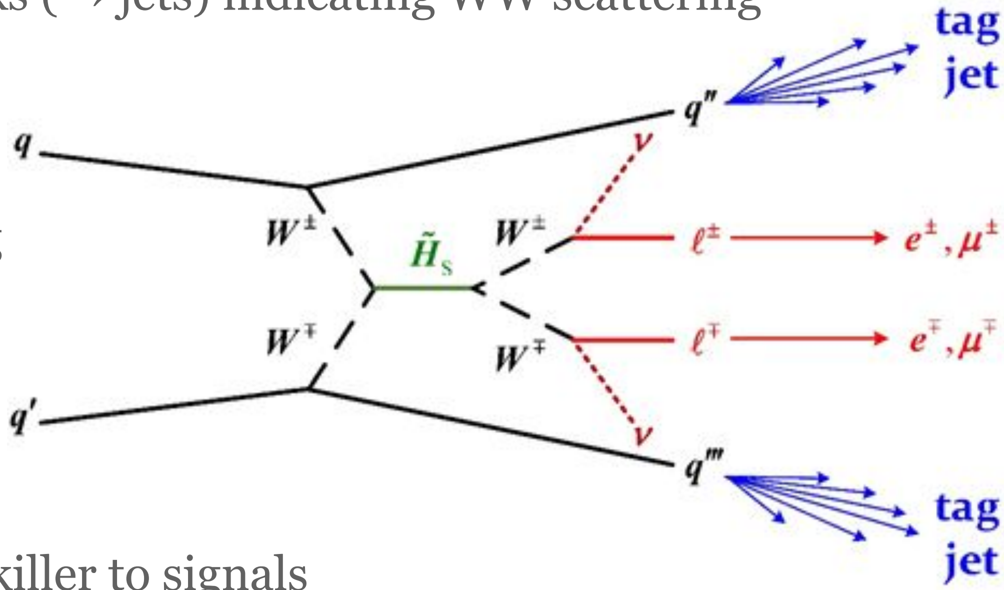
Outline



1. "Cern Art." *Fine Art America*, fineartamerica.com/art/cern.

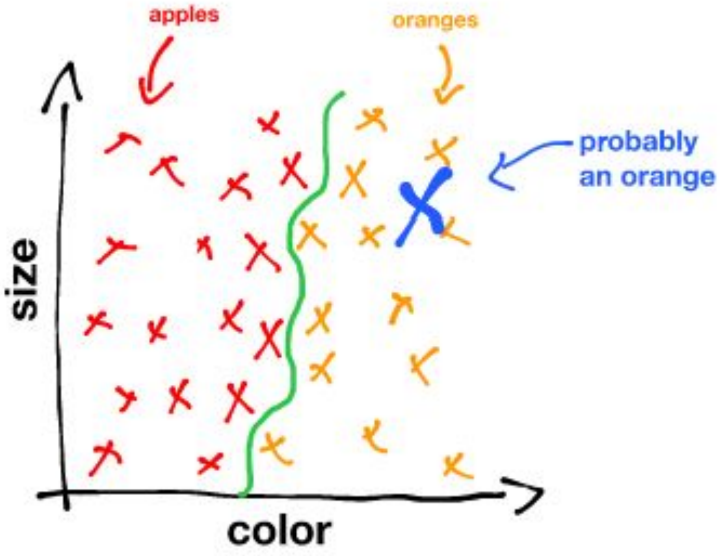
I. Introduction: Physics

- New particle at m = 2.6 TeV produced in hi-lum LHC by Vector Boson Fusion (**VBF**)
- **Goals:**
  - Reconstruct the two “tag” quarks (→ jets) indicating WW scattering
    - Improve signal jet efficiency
    - Employ **clusters and towers** with and without signal timing cuts
    - Apply **pile-up suppression** techniques: jet area, constituent subtraction, and constituent subtraction + soft killer to signals



III. Introduction to Supervised Machine Learning

- Training computer to recognize patterns in data
  - i.e. a certain p_T distribution over rapidity space of jets and pile-up, respectively
- **Neural Networks**, non-linear data modeling tools, are used to identify statistical structure
  - Modeled after biological neural networks, a connected system



- Computer learns to recognize patterns through **training data**

2. "KDnuggets." *KDnuggets Analytics Big Data Data Mining and Data Science*, www.kdnuggets.com/2017/05/machine-learning-crash-course-part-1.html.

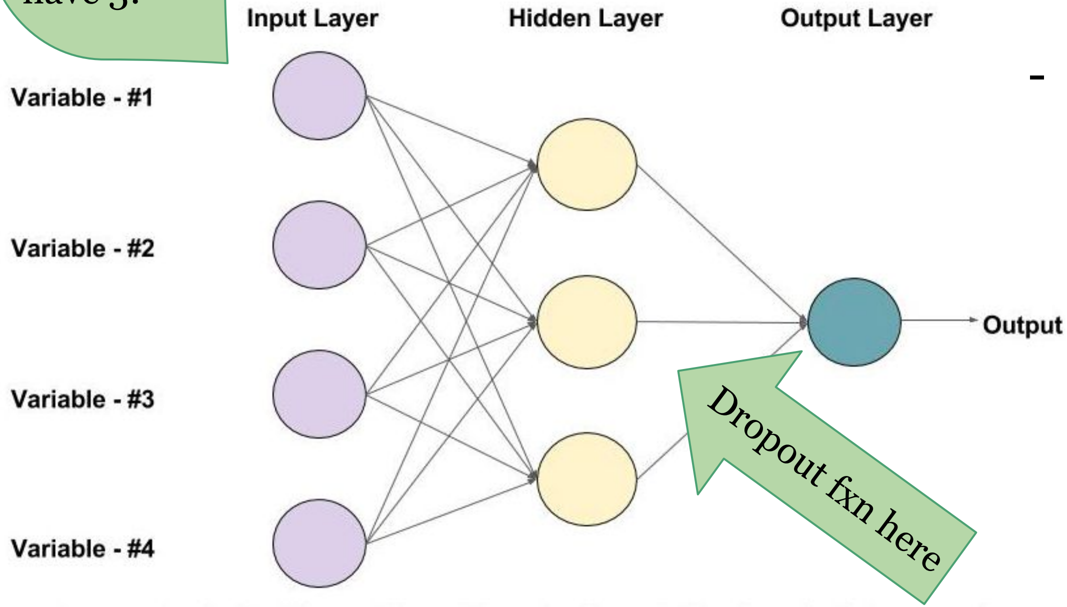
TensorFlow and Tflearn

- Google's open source software library for dataflow programming
  - Especially well suited for designing and implementing **DNN's**
- Most stable for coding in **Python and C**
  - Also provides interfaces for JavaScript, Java, C++, Go, and Swift
- **Tflearn** is a deep learning library built on top of TensorFlow
  - Fully **transparent**
  - Speeds up computation
  - Provides functions for training, evaluation, and prediction

3. TensorFlow. (n.d.). Retrieved from https://www.tensorflow.org/

Algorithms and Configuration

Only 1 hidden layer, we have 3!



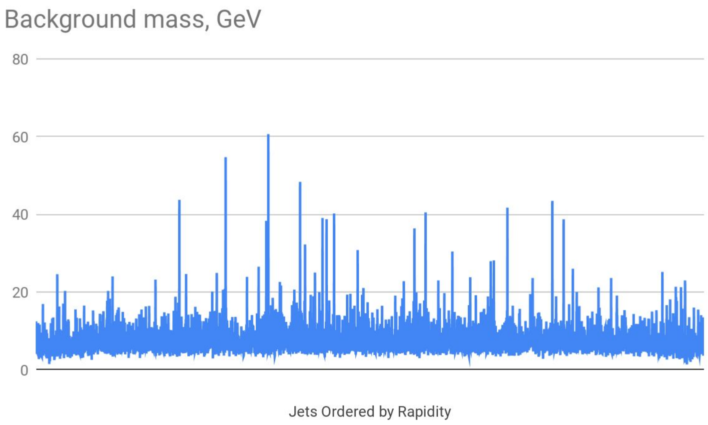
An example of a Feed-forward Neural Network with one hidden layer (with 3 neurons)

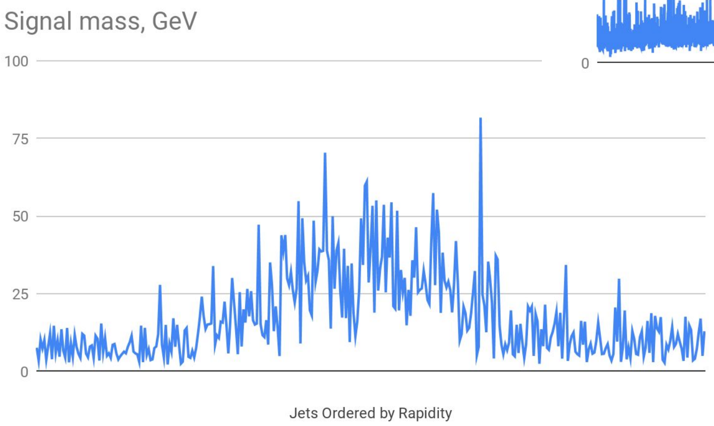
- Looking at a single object
- **Current model:** 4-layer neural network
  - Includes one dropout function (prevents overfitting)
  - TfLearn **DNN** function performs training, prediction, etc.

4. Gupta, Vikas. “Home.” *Learn OpenCV*, 9 Oct. 2017, www.learnopencv.com/understanding-feedforward-neural-networks/.

Relevance of Observables

- **Distinguishability** of observables, based on variance
  1. Mass
  2. p_T
  3. # of constituents
  4. p_T^D
  5. Width



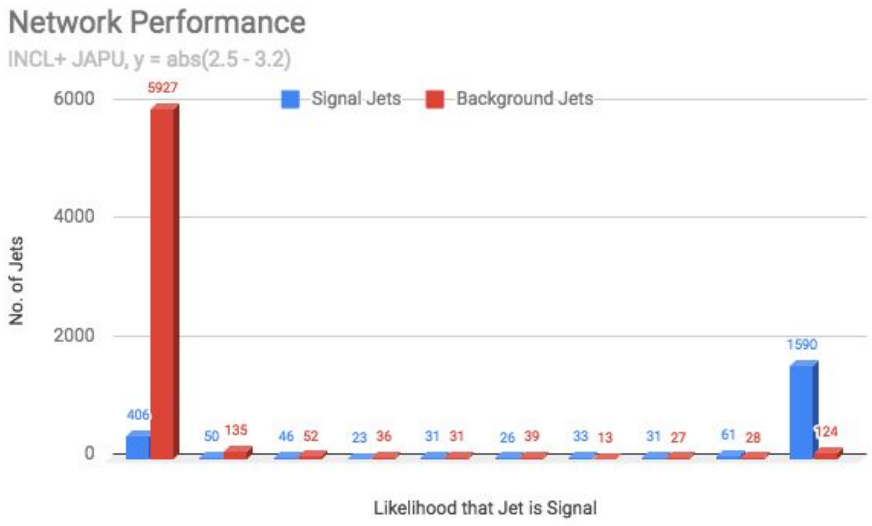


- Network built up **by rapidity region** using one observable at a time by significance, **Accuracy** improved as more observables were added
  - <u>Mass only</u>: **54%** certainty for average signal jet
  - <u>All observables</u>: **100%** certainty for average signal jet

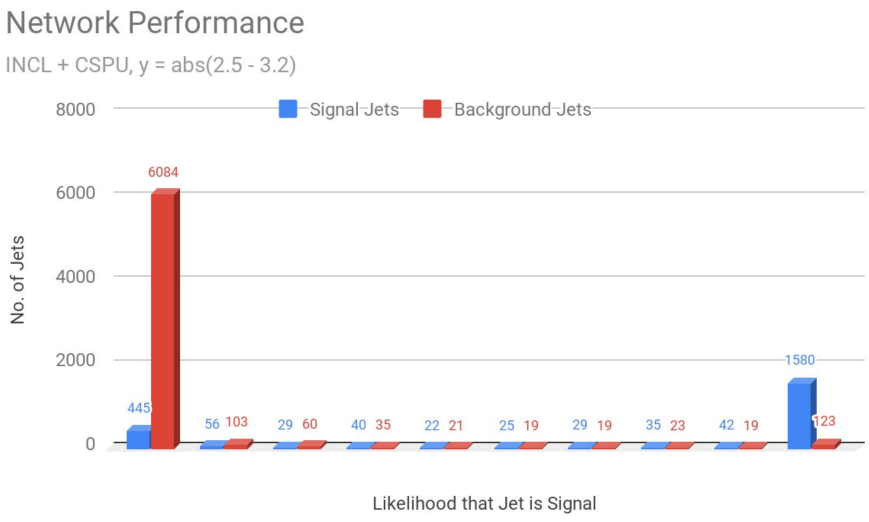
Network Performance

Fine Topo-Towers 0.05 x 0.05

Area-Based Pile-up Suppression



Constituent Subtraction



Fine Topo-Towers 0.05 x 0.05

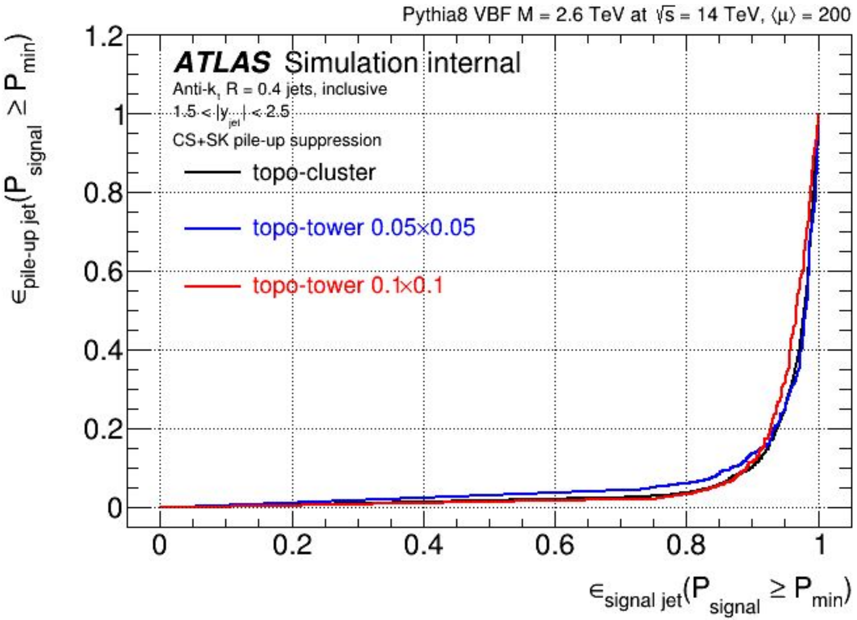
Percentage of Background/Signal Jets Predicted Correctly

Selection	Background	Signal	# training pts.	# test pts.
INCL + JAPU	97.01%	74.66%	4,647	8,709
TIME + JAPU	95.00%	78.10%	4,341	6,168
INCL + CSPU	96.02%	76.99%	4,647	8,809
TIME + CSPU	94.28%	79.31%	4,341	6,168

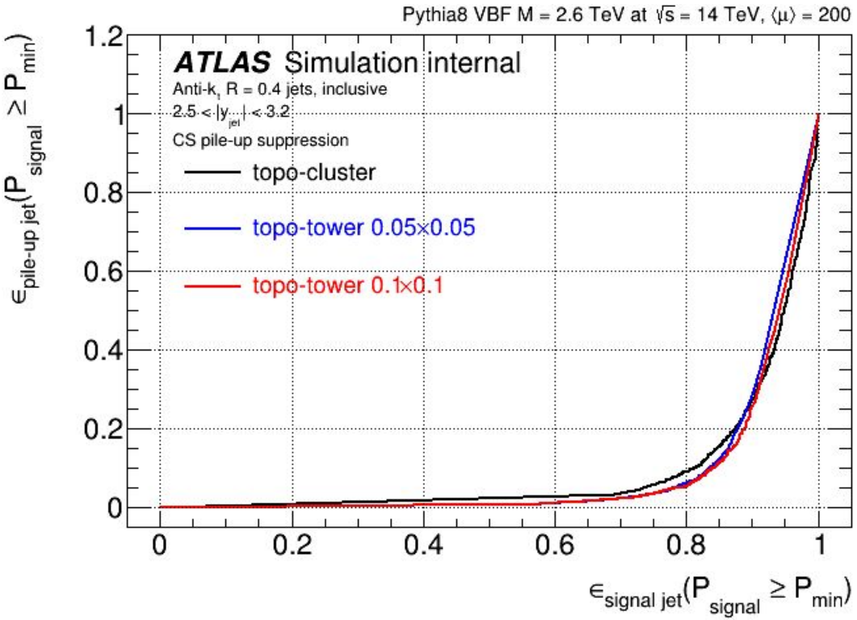
- Cutting at likelihood which maximizes jets identified correctly
- Rapidity Region: y = abs(2.5-3.2)

ROC Curves, ML Algorithm Success

CSSK, 1.5 ≤ y ≤ 2.5



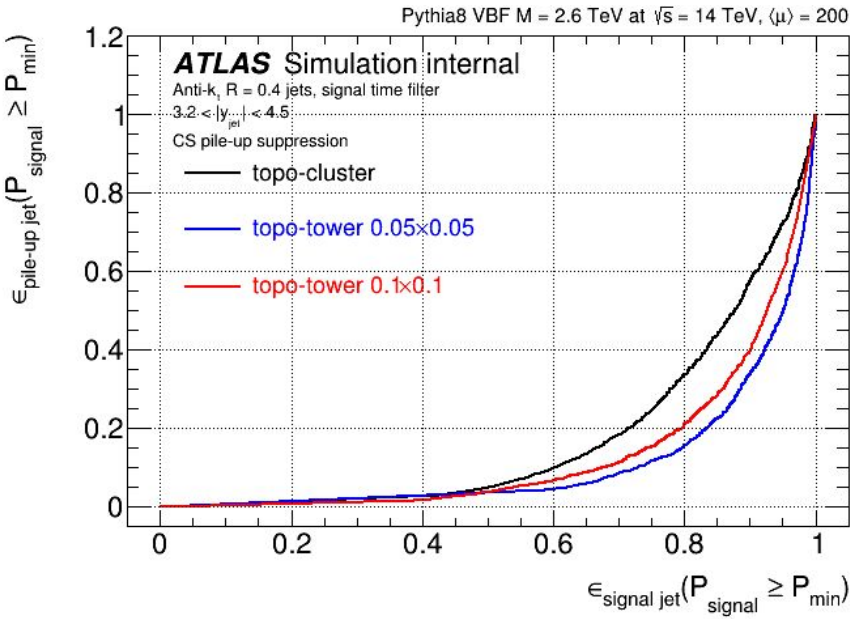
CS, 2.5 ≤ y ≤ 3.2



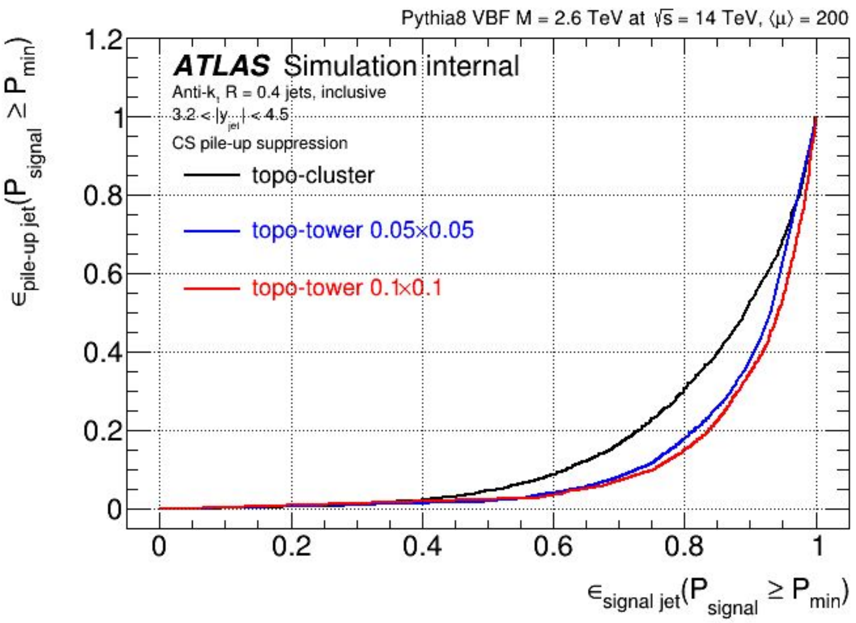
See all ROC curve results here.
More information on ROC curves and other analysis techniques here.

Effectiveness of ML on Different Calorimeter Signals with and without Timing Cuts

CS + TIME, 3.2 ≤ y ≤ 4.5



CS + INCL, 3.2 ≤ y ≤ 4.5



See all ROC curve results here.
More information on ROC curves and other analysis techniques here.

IV. Conclusion and Outlook

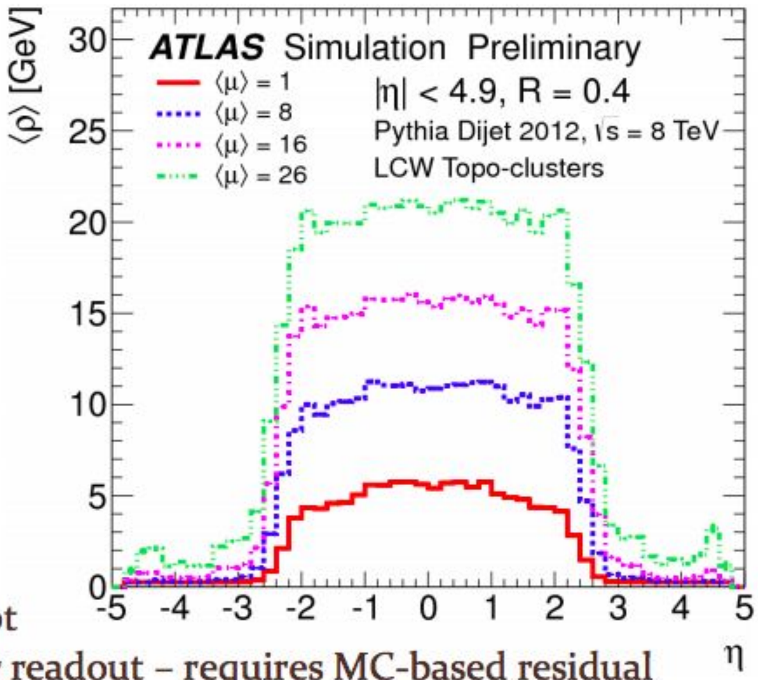
- This project is a first attempt at using machine learning to classify jets for final state with calorimeter clusters, towers, and fine towers
- In the **future** we plan to:
  - Consider only two jets that form the invariant mass and rapidity gap
  - Continue with machine learning implementation and network improvement
  - Use larger training data sets to improve results
  - Consider more selections such as Area Based Pile-up Suppression, Constituent Subtraction, and Soft Killer
- **Why Machine Learning?**

Questions?

Thank you!

Backup: Further Introduction

Calorimeter-based pile-up-jet suppression in Run 3 & beyond



- Extensions of pile-up jet tagging
  - q/g jet tagging in VBF – signal jets are quark-like, pile-up jets are gluon-like
  - Jet shape analysis using e.g. m_{jet}, $p_{\mathrm{T}}^{\mathrm{D}}$ (jet fragmentation function), …
- Jet-area-based pile-up subtraction
  - Well established approach in LHC Run 1 & 2
  - Transverse momentum density ρ not well measured in coarse calorimeter readout – requires MC-based residual corrections in particular in forward region ($|\eta| > 2.5$)
- Constituent-level pile-up suppression prior to jet reconstruction
  - Select calorimeter signal based on features indicating pile-up or generally low signal quality – e.g. timing, significance, …
  - Applying stochastically motivated methods like SoftKiller, Voronoi Suppression, Constituent Subtraction, …

Challenges

Topo-cluster in coarse readout

Deplete (η, φ) space of four-momenta – cell signal collection feature of cluster algorithms

Single topo-cluster catchment area not well defined – e.g. Voronoi in coarse readout can be very large, ρ measurement biased due to few clusters outside of jets/low cluster multiplicity inside of jets ...

Can generate single cluster jets – loss of structural flow information, reduced efficiency of e.g. pile-up jet tagging

Mitigation approaches

Structural flow measures

Transition of jet substructure observables $\longrightarrow$ corresponding topo-cluster moments with similar sensitivity to transverse momentum flow

Topo-cluster $p_{\mathrm{T}}^{\mathrm{D}}$, $m_{\mathrm{jet}} \longrightarrow m_{\mathrm{cluster}}$, jet width $\longrightarrow$ cluster width, ...

Improved cluster area determination

Using e.g. lateral topo-cluster extension moments for area measurement

CaloTowers helpful ?

Non-projective cells in FCal

Different sensitivity of tower signal to transverse momentum flow – complex (geometrical) cell energy sharing between towers

Projective readout

Not much gain expected if tower bin boundaries line up with cell boundaries – simple equal-weight cell energy sharing (TBC)

Topo-clusters

Standard calorimeter signal definition employs noise suppression and local calibration

Full details in Eur.Phys.J. C77 (2017) 490

Calorimeter towers

Two (η, φ) grids

- $\Delta\eta \times \Delta\varphi = 0.1 \times 0.1$ (standard/*coarse*)
- $\Delta\eta \times \Delta\varphi = 0.05 \times 0.05$ (*fine*)
- Cell signals collected using geometrical weights according to cell/tower area overlap in (η, φ) space
- Fixed catchment area in (η, φ)
- Only cells with $E > 0$ considered

Two signal collection strategies

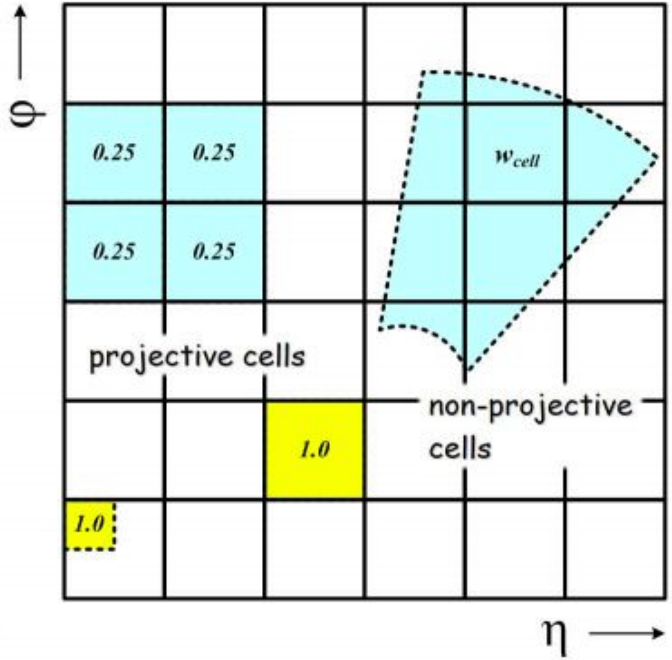
- Inclusive – collect all calorimeter cells
- Topo-towers – collect only cells from topo-clusters (noise suppression!)

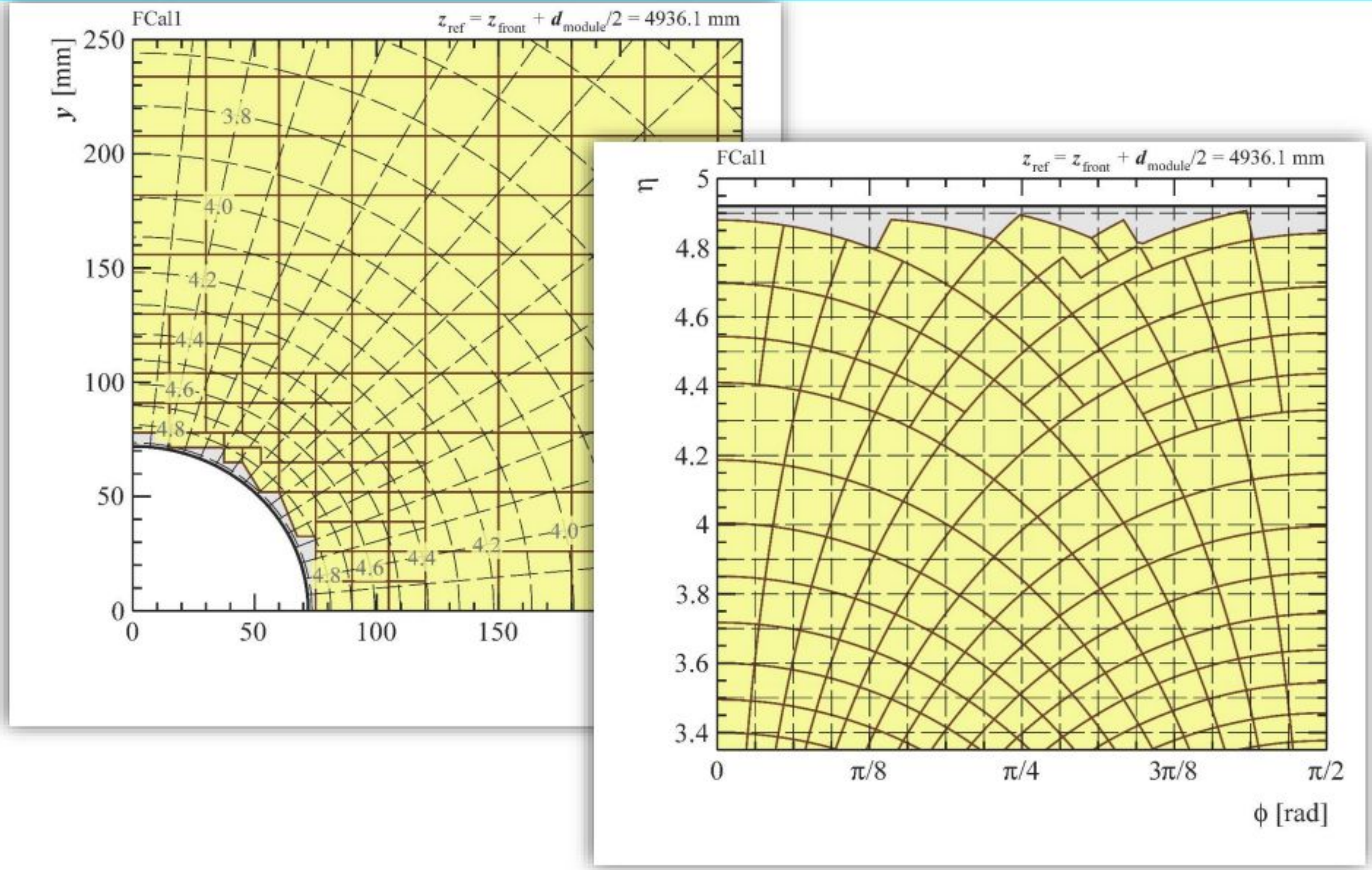
Calibrations

- EM for inclusive and topo-towers
- LCW for topo-towers and topo-clusters (calibrations applied in topo-cluster context, no dedicated topo-tower calibration)

Details

- See Twiki at https://twiki.cern.ch/twiki/bin/viewauth/AtlasSandboxProtected/CaloTowerPerformance




FCal1
$z_{ref} = z_{front} + d_{module}/2 = 4936.1$ mm
y [mm]
250
200
150
100
50
0
0
50
100
150
3.8
4.0
4.2
4.4
4.6
4.8
FCal1
$z_{ref} = z_{front} + d_{module}/2 = 4936.1$ mm
η
5
4.8
4.6
4.4
4.2
4
3.8
3.6
3.4
0
π/8
π/4
3π/8
π/2
ϕ [rad]

$\varrho(\eta)$ Measurement

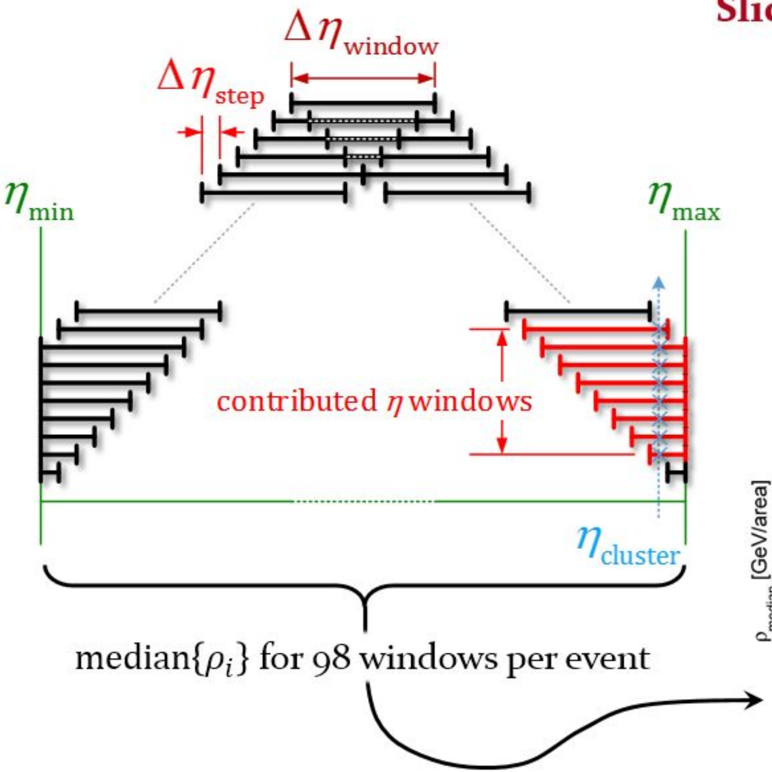
Standard reference

- Jet-based median in central detector region $|\eta| < 2$
  - Anti-k_t $R = 0.4$ jets from topo-clusters on EM ($\rho_{\mathrm{ref}}^{\mathrm{EM}}$) or LCW ($\rho_{\mathrm{ref}}^{\mathrm{LCW}}$) scale
  - FastJet implementation using jets with $p_{\mathrm{T,jet}}^{\mathrm{EM(LCW)}} \geq 0$
  - No η dependence (expected from particle flow in minbias)

Variations for performance evaluation

- Use sliding η windows to collect signals
  - Overlapping windows with nominal width $\Delta\eta_{\mathrm{window}} = 0.8$ centered at a given η_{window} – same detector signal contributes to several windows
  - Slide window in small steps $\Delta\eta_{\mathrm{step}} = 0.1$
  - Adjust left/right window boundary at detector edges
    - $\eta_{\mathrm{window}} - \Delta\eta_{\mathrm{window}}/2 \geq -4.9$ and $\eta_{\mathrm{window}} + \Delta\eta_{\mathrm{window}}/2 \leq 4.9$ – asymmetric windows near detector edges
- Calculate median ρ in each window
  - Use $\rho = p_{\mathrm{T}}/A$ for each topo-cluster/tower in the window
    - Tower areas A well defined by tower grid, Voronoi area used for topo-cluster
  - Median ρ from ***catchment area*** or ***signal area***
    - Catchment area approach includes areas void of signal ($p_{\mathrm{T}} = 0$) window in median (FastJet-like)
    - Signal area approach provides median of signal densities excluding void areas

Introduction: ρ Distribution



Sliding window ρ collector

Window size $\Delta\eta_{\text{window}} = 0.8$

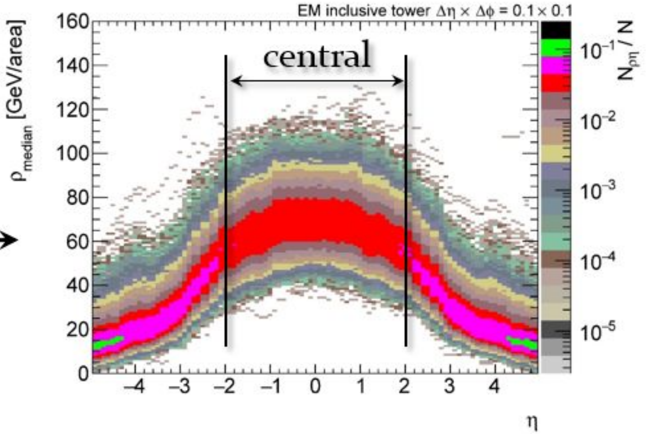
Step size $\Delta\eta_{\text{step}} = 0.1$

$\eta_{\text{min}} < \eta_{\text{window}} < \eta_{\text{max}}$

$\eta_{\text{min}} = -4.9$

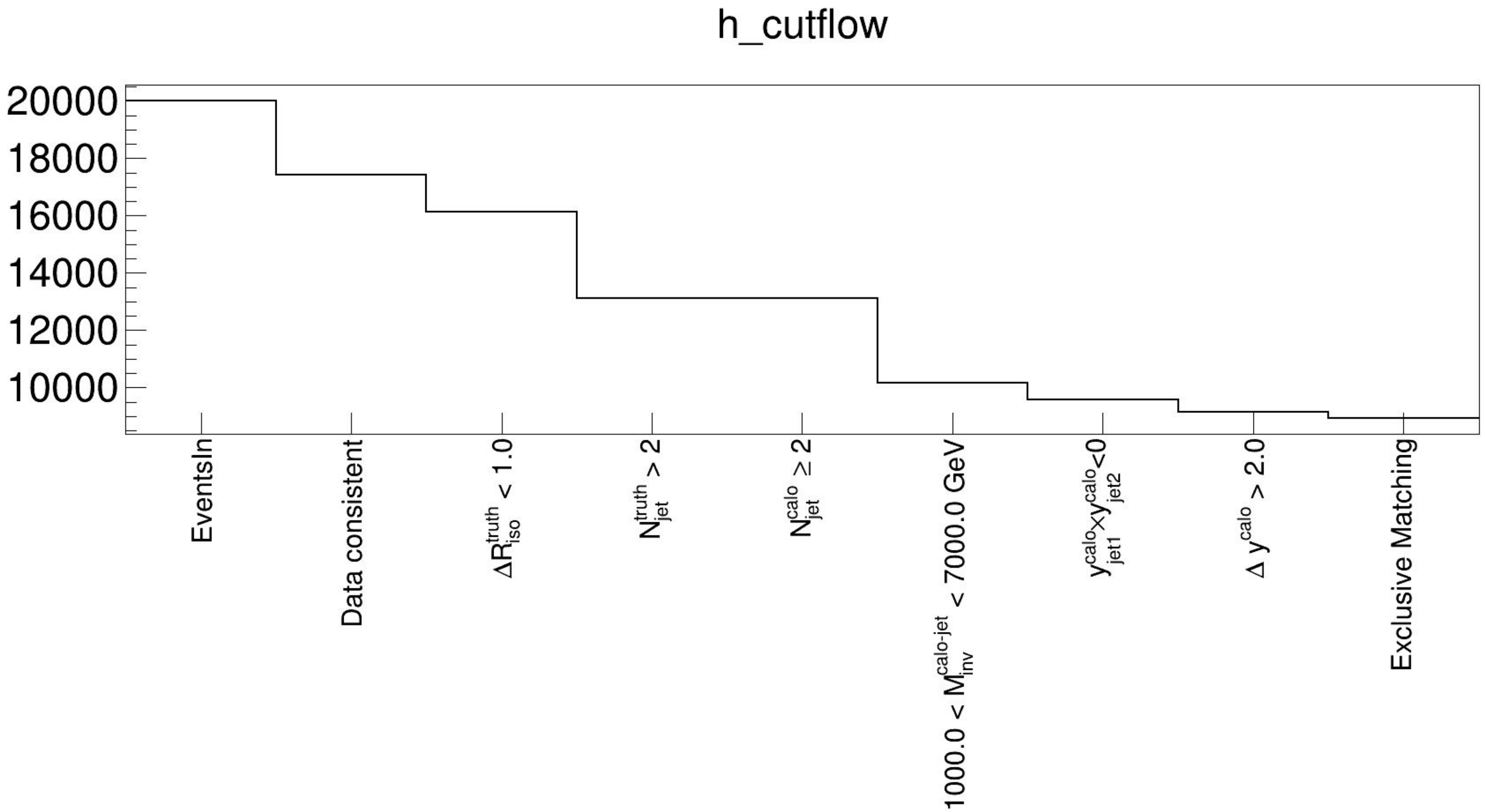
$\eta_{\text{max}} = 4.9$

ρ in η window k is $\rho_k = \text{median}\{\rho_i\}_k$

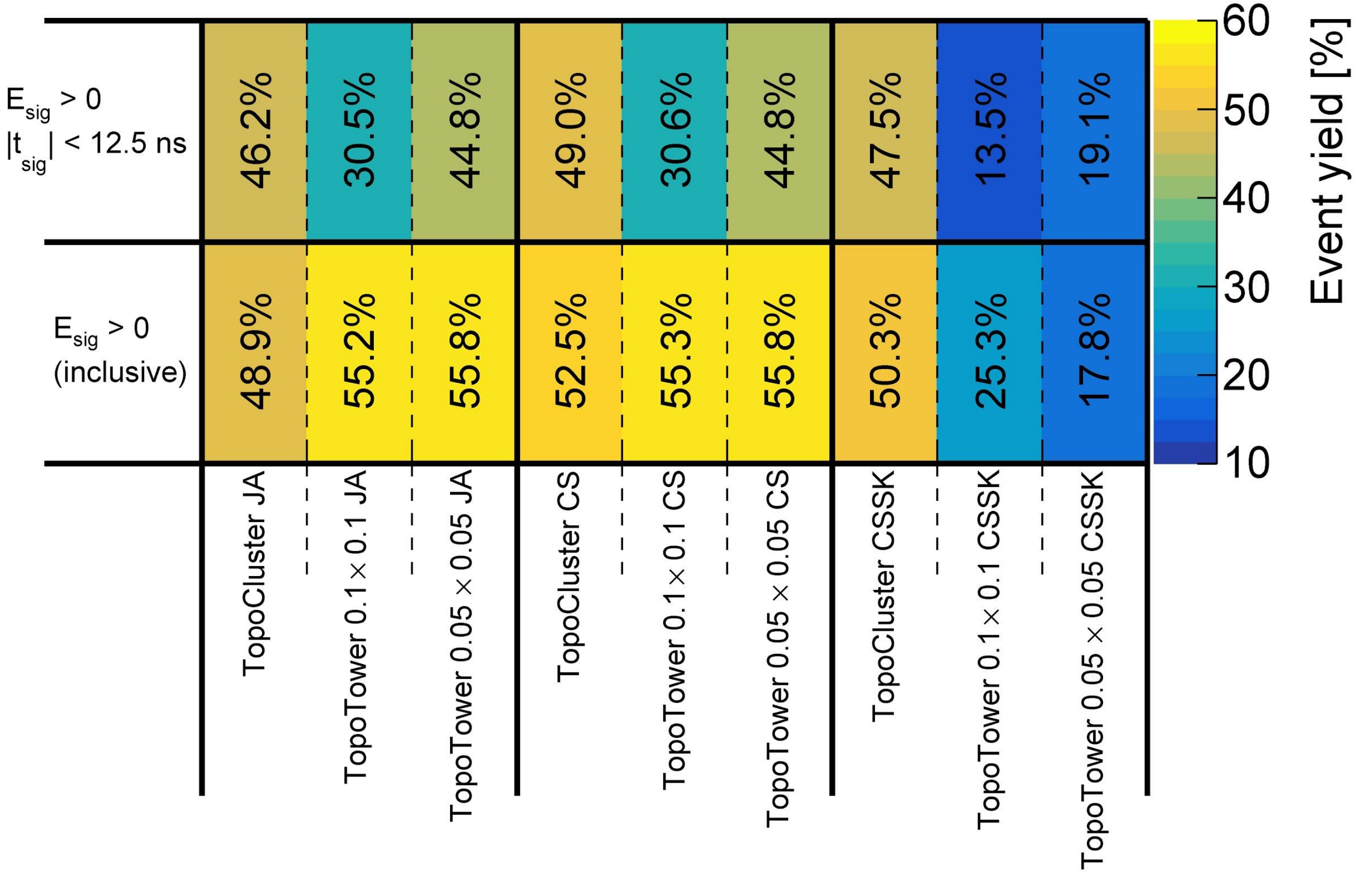


Backup: Event Selection, Jet Reconstruction, and Jet Shapes

Event Selection



Event Efficiencies



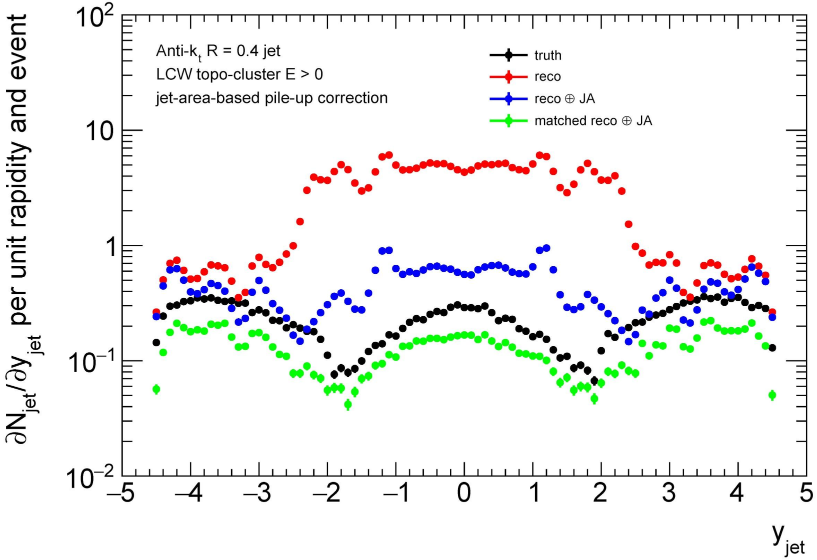
Takeaways:

- Number of events saved after each selection with and without timing cuts

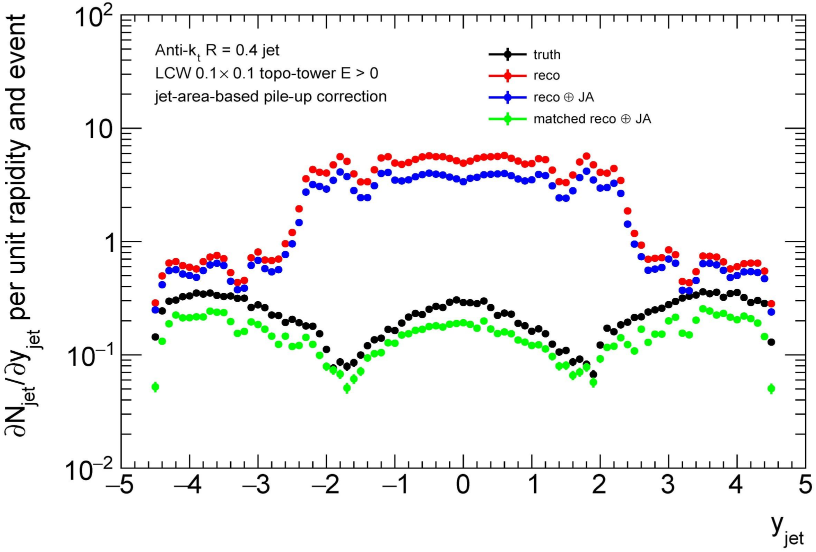
- Constituent subtraction with soft killer may remove too many events to be useful (?)

Jet Rapidity Distribution without Timing Cuts

TOPO-CLUSTERS



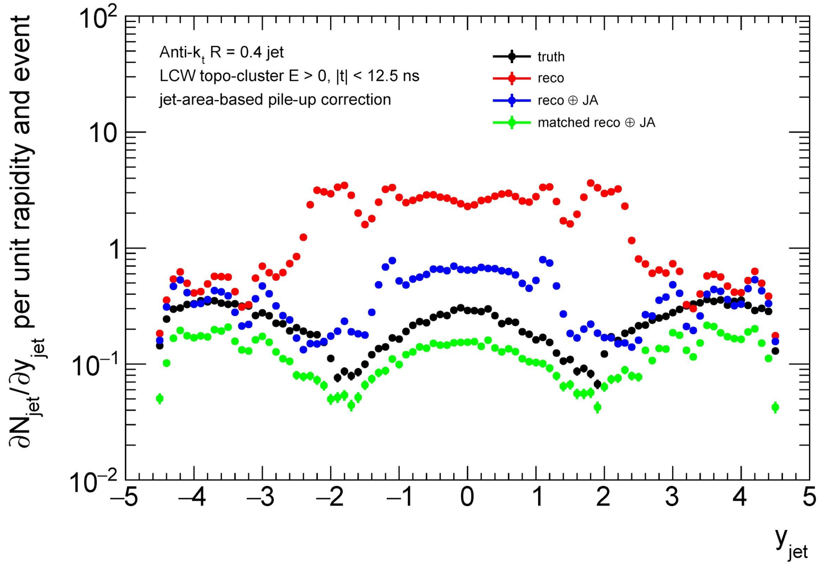
TOPO-TOWERS



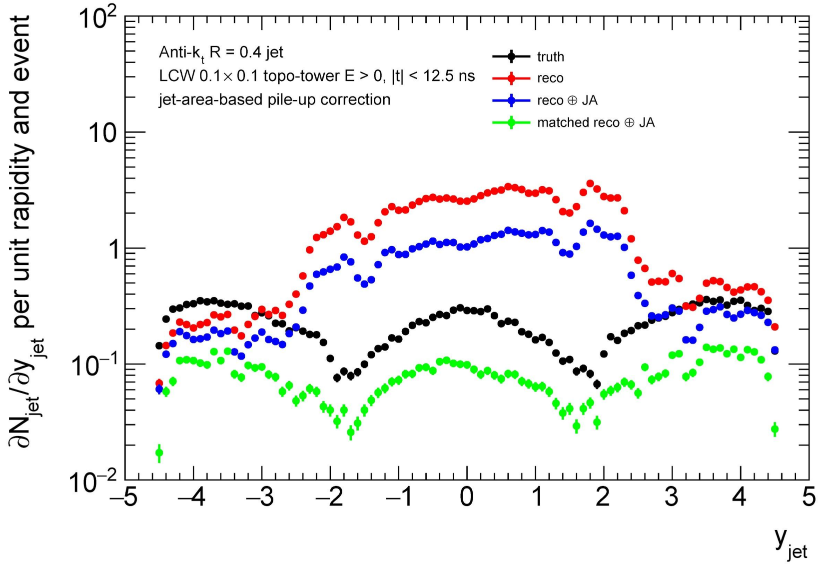
- Note: the effectiveness of the area based suppression is reduced for towers, mostly in the central region

Jet Rapidity Distribution with Timing Cuts

TOPO-CLUSTERS



TOPO-TOWERS

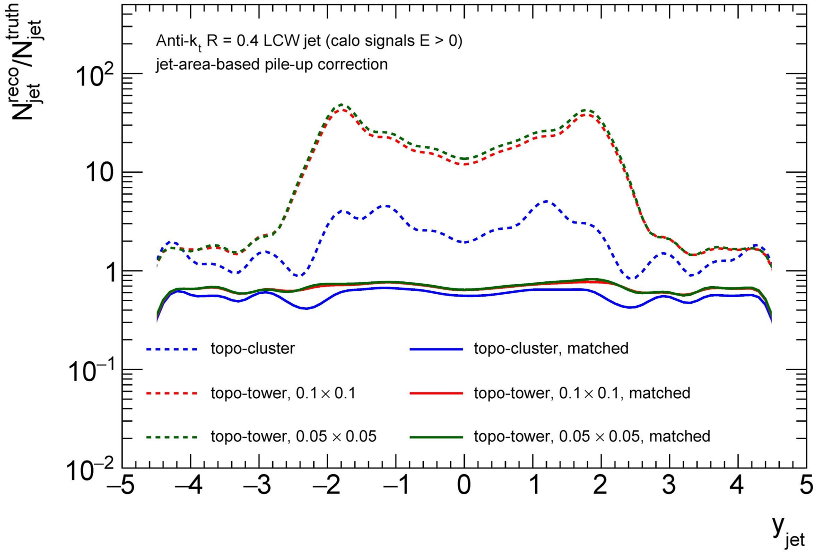


Note: Asymmetry in topotower plot, likely due to timing of signals

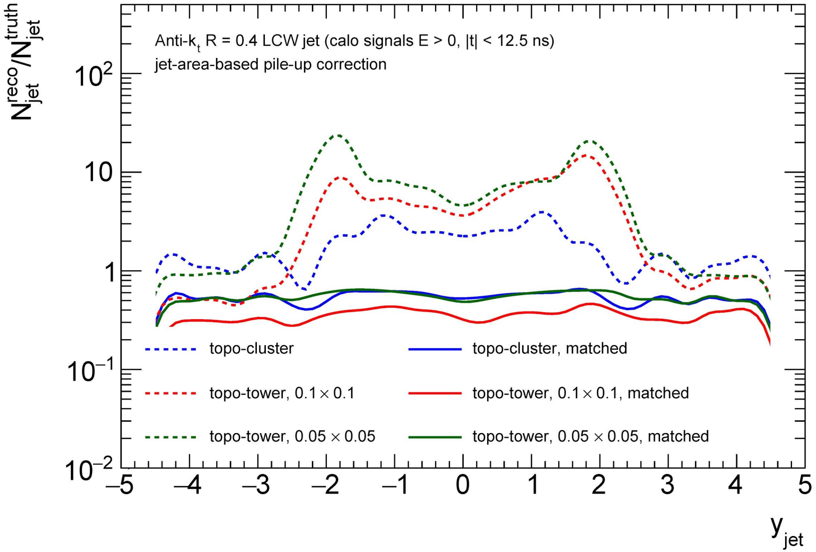
Jet Reconstruction Efficiency

Area Based Suppression

INCLUSIVE



TIMING CUT

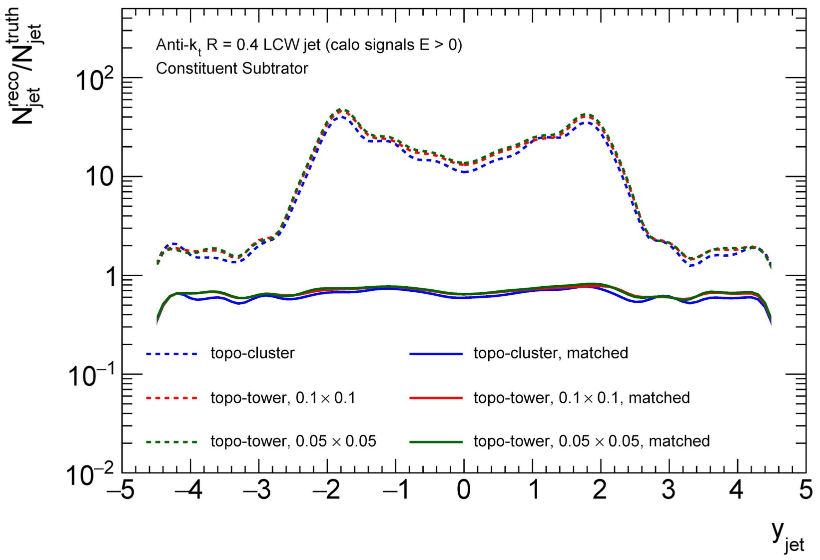


Note: higher efficiency (c.f. matched jets) for towers w/o timing cut; reduced efficiency for 0.1 x 0.1 towers with timing cut

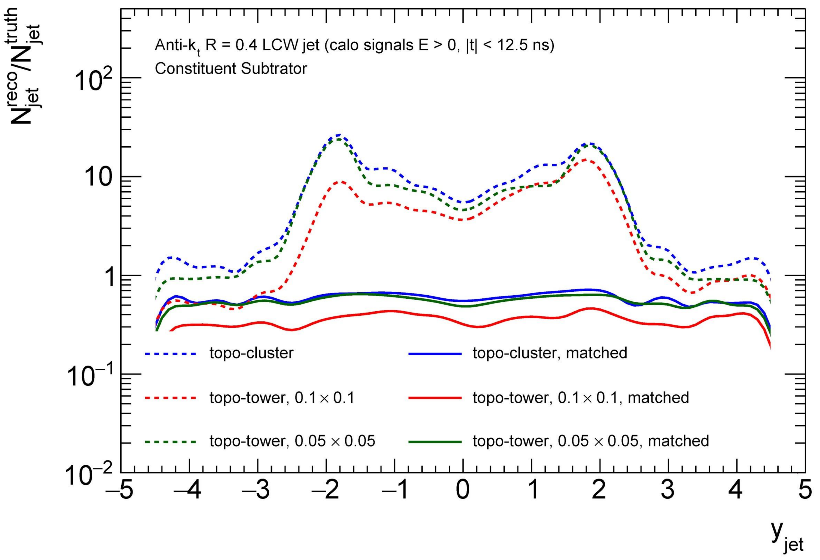
Jet Reconstruction Efficiency

Constituent Subtraction

INCLUSIVE



TIMING CUT

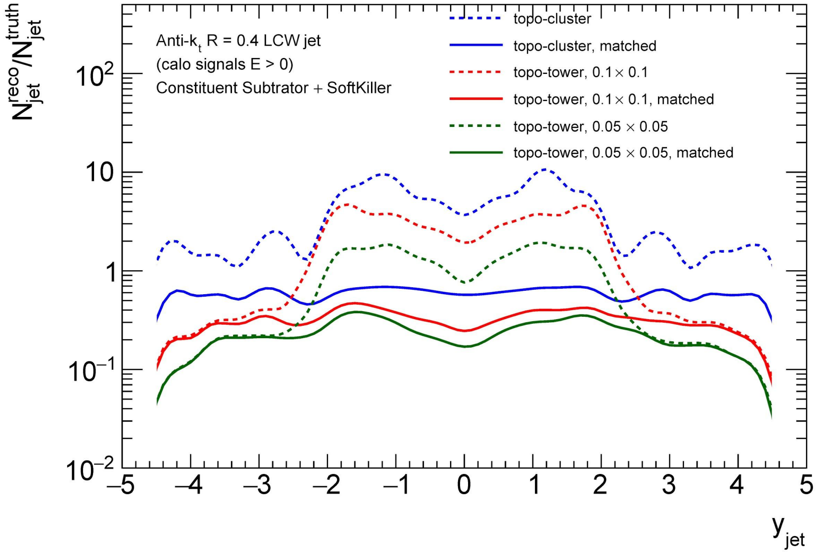


Note: similar efficiency for all calorimeter signals w/o timing cut; reduced efficiency for 0.1 x 0.1 towers with timing cut

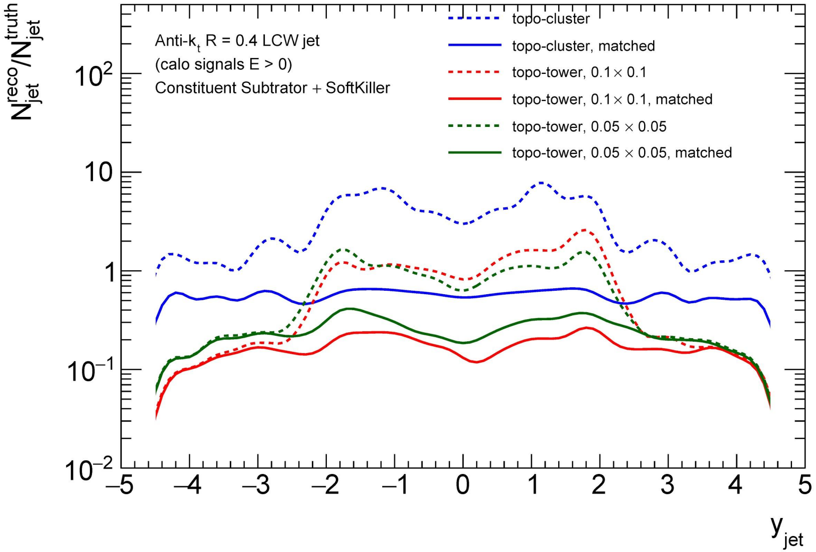
Jet Reconstruction Efficiency

Constituent Subtraction + Soft Killer

INCLUSIVE



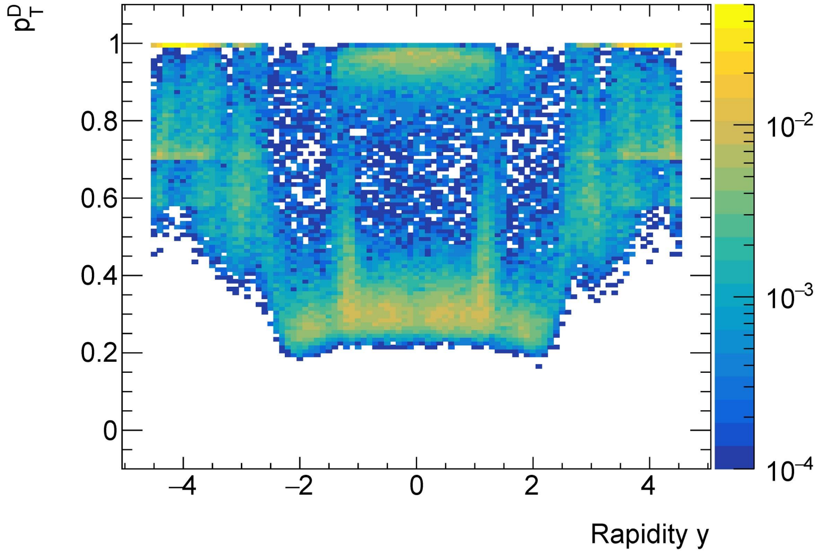
TIMING CUT



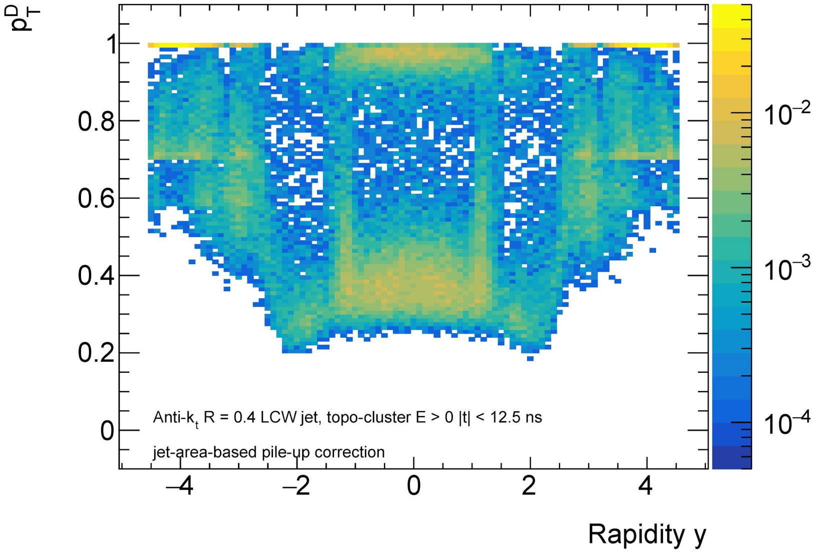
Note: topo-cluster less affected by timing cuts; constituent subtraction + soft killer likely too strong for towers → look at jet shapes & ML-based approaches

Jet Shapes
(p_T^D, topo-cluster)

INCLUSIVE



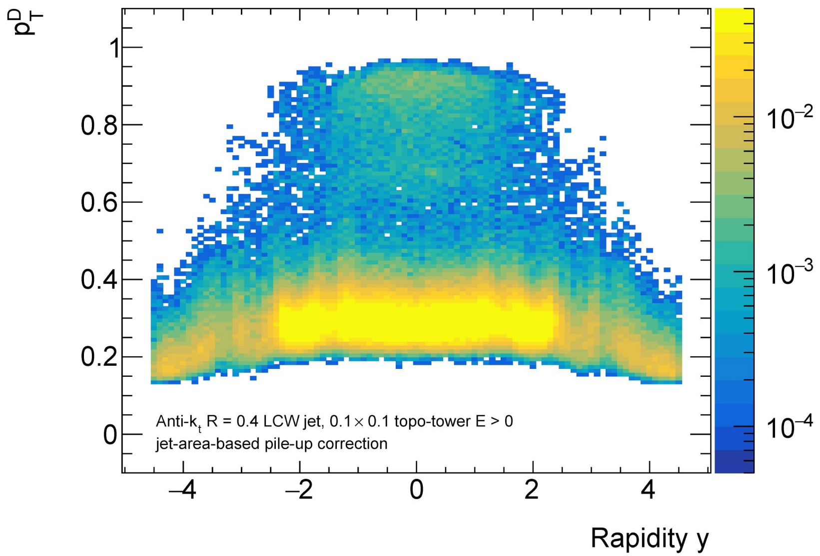
TIMING CUT



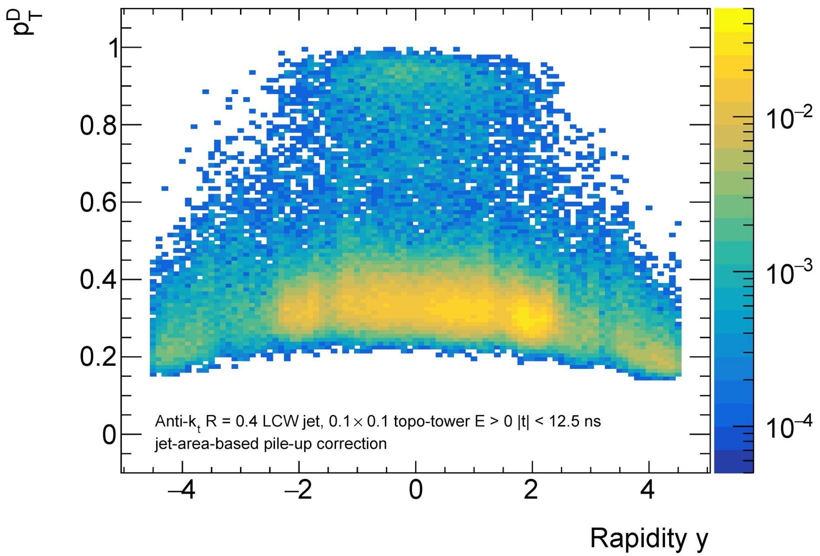
Jet Shapes

(p_T^D, topo-tower 0.1 x 0.1)

INCLUSIVE



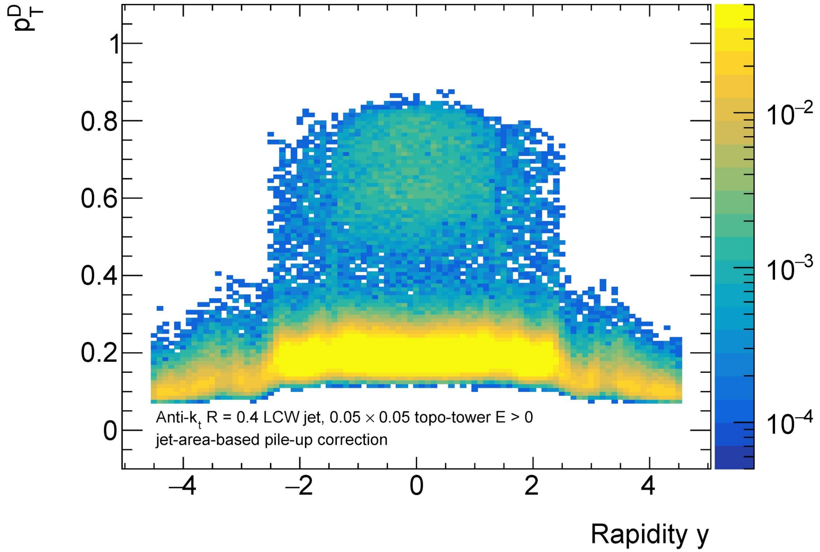
TIMING CUT



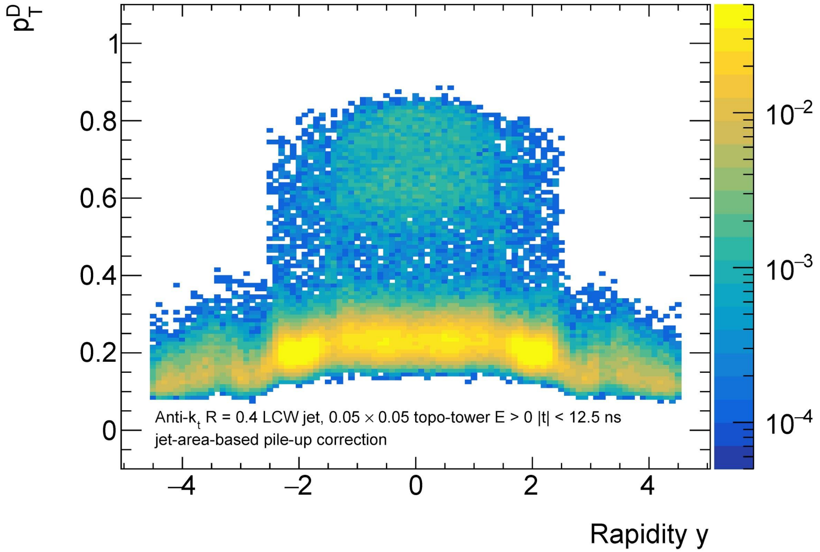
Jet Shapes

(p_T^D, topo-tower 0.05 x 0.05)

INCLUSIVE

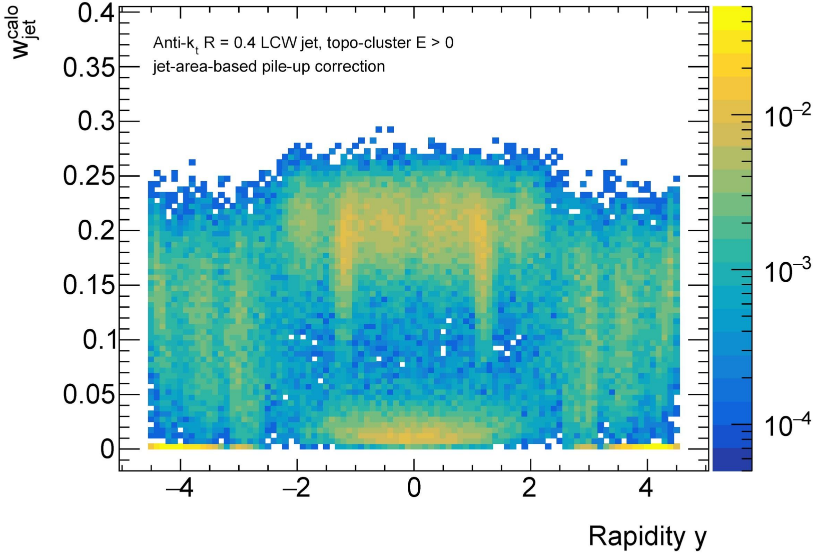


TIMING CUT

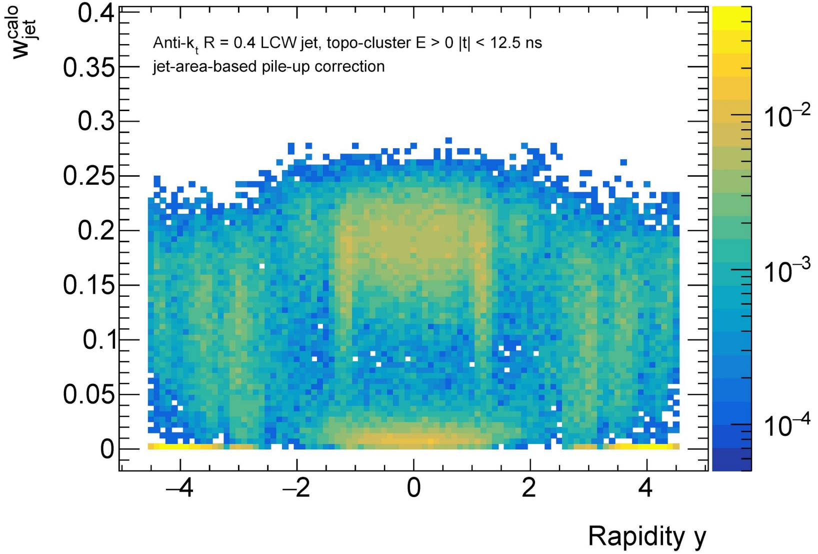


Jet Shapes
(width, topo-cluster)

INCLUSIVE

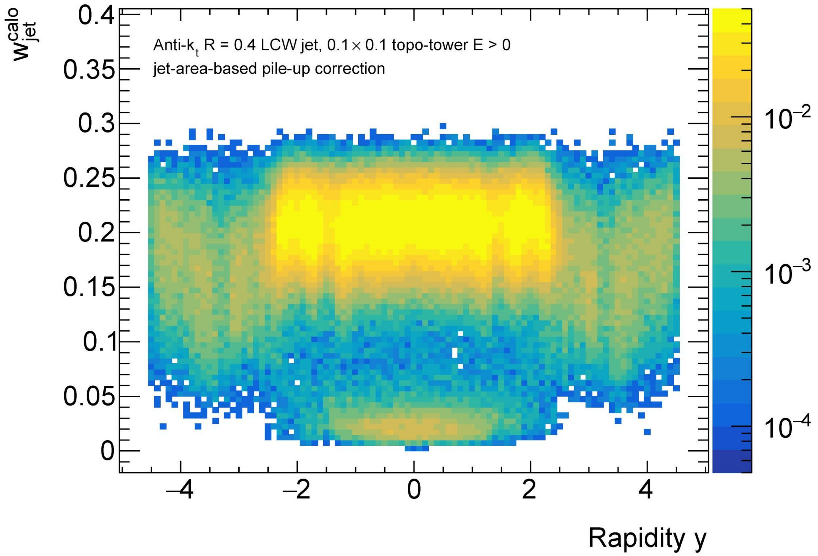


TIMING CUT

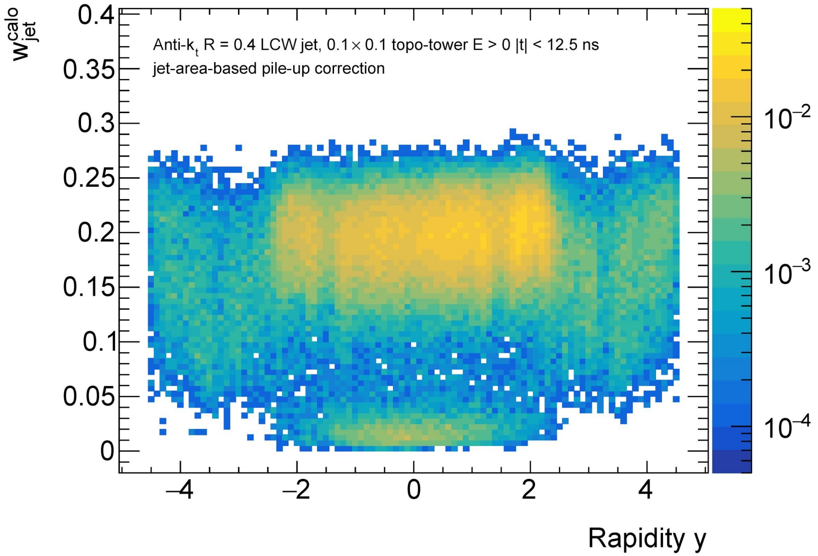


Jet Shapes
(width, topo-tower 0.1 x 0.1)

INCLUSIVE



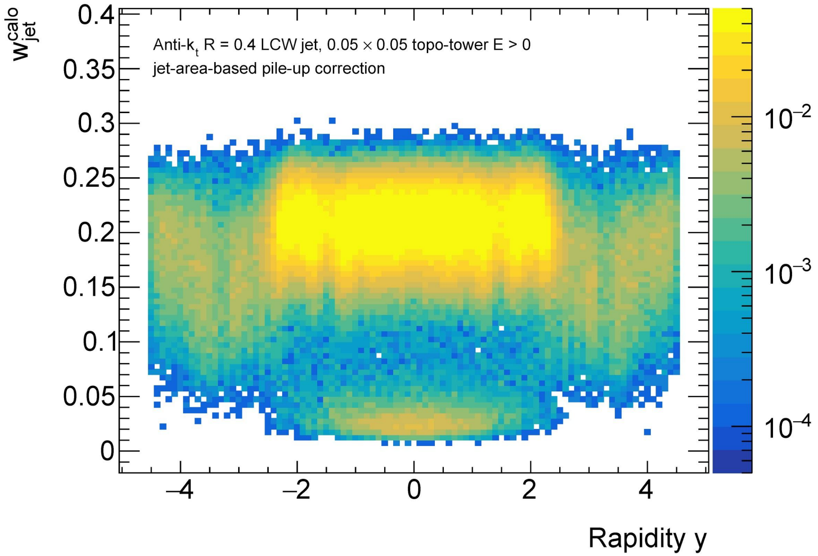
TIMING CUT



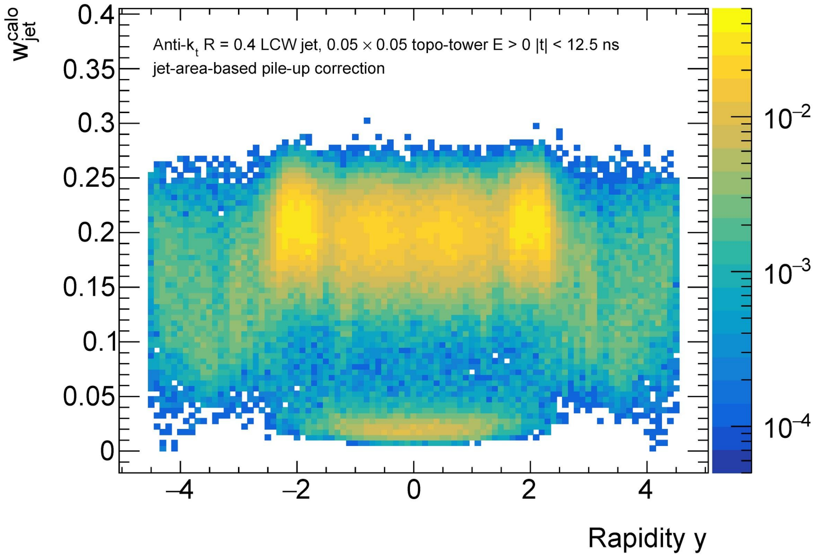
Jet Shapes

(width, topo-tower 0.05 x 0.05)

INCLUSIVE



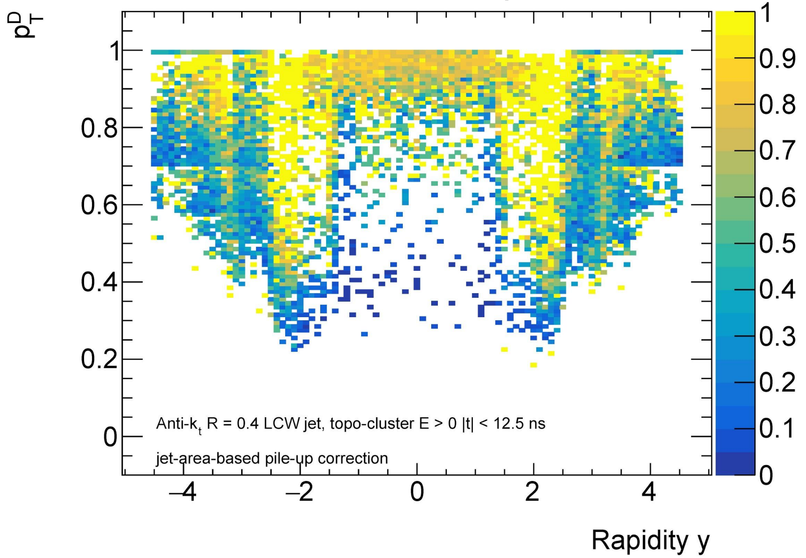
TIMING CUT



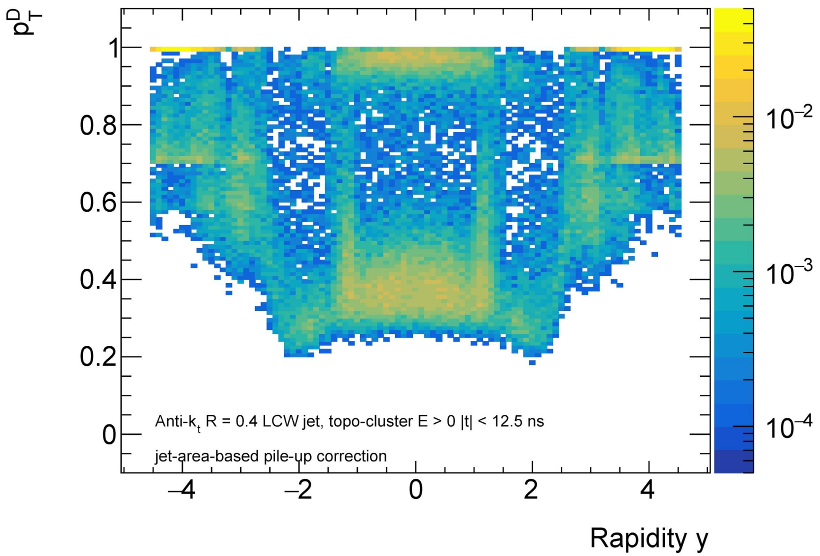
Signal Likelihood in Jet Shapes

(p_T^D, topo-cluster)

RATIO #SIGNAL JETS/#ALL JETS



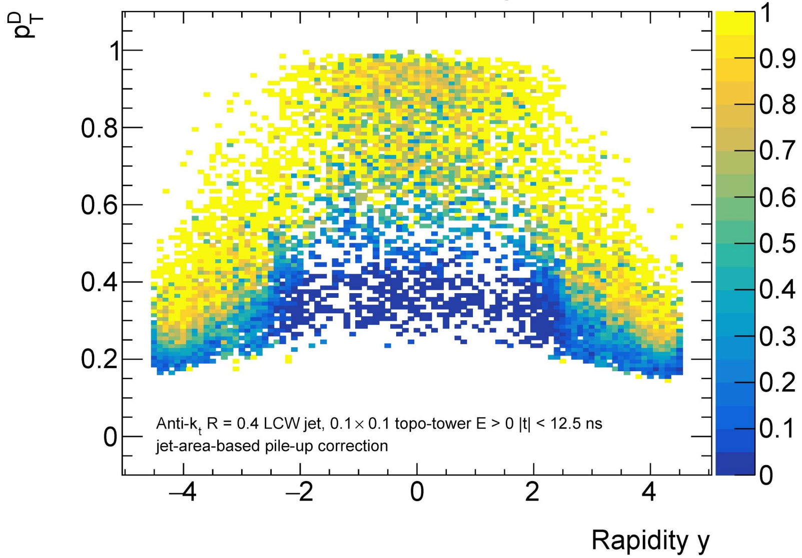
TIMING CUT



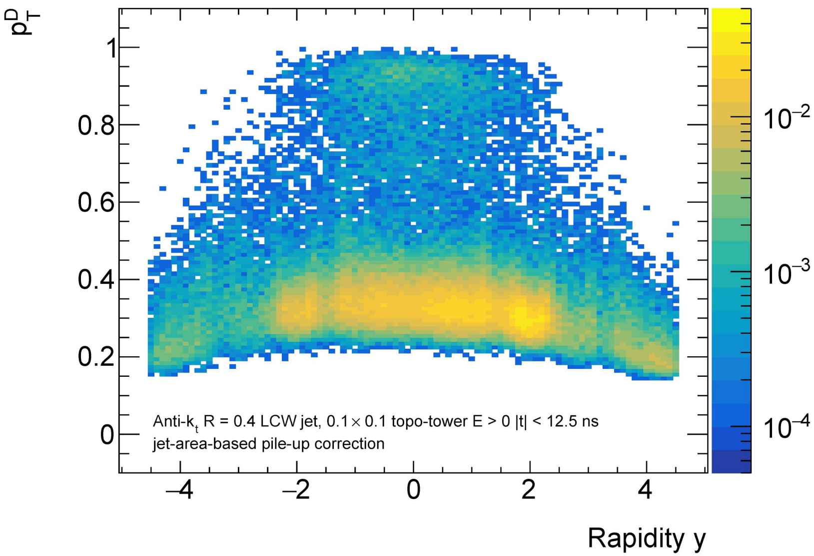
Signal Likelihood in Jet Shapes

(p_T^D, topo-tower 0.1 x 0.1)

RATIO #SIGNAL JETS/#ALL JETS



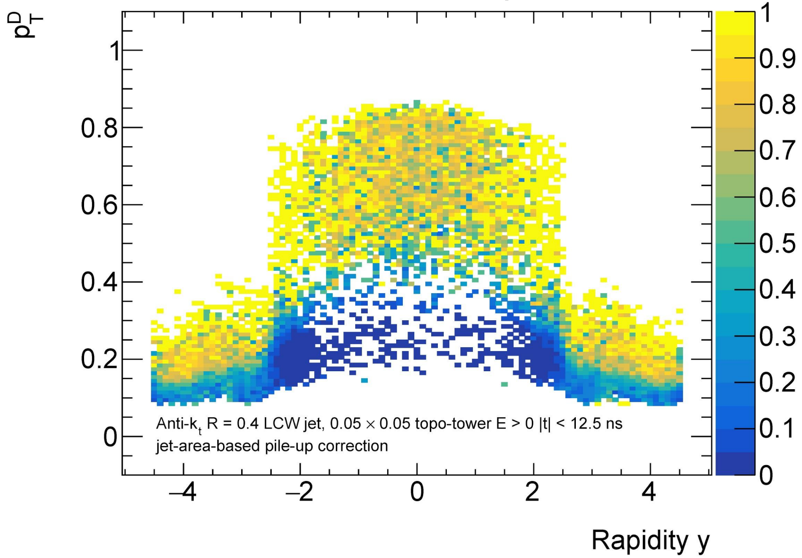
TIMING CUT



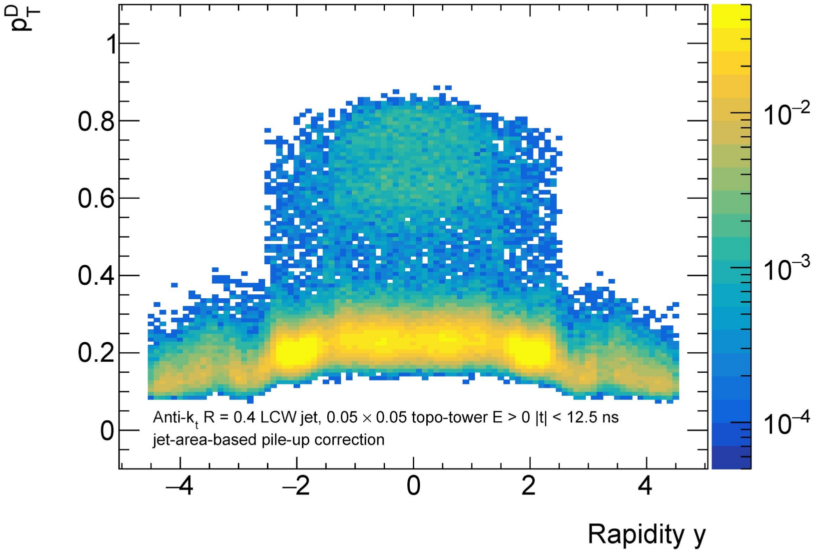
Signal Likelihood in Jet Shapes

(p_T^D, topo-tower 0.05 x 0.05)

RATIO #SIGNAL JETS/#ALL JETS



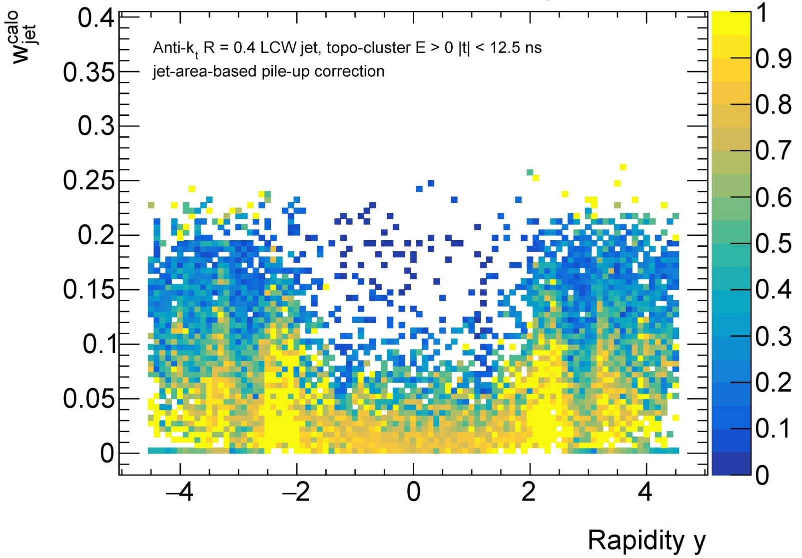
TIMING CUT



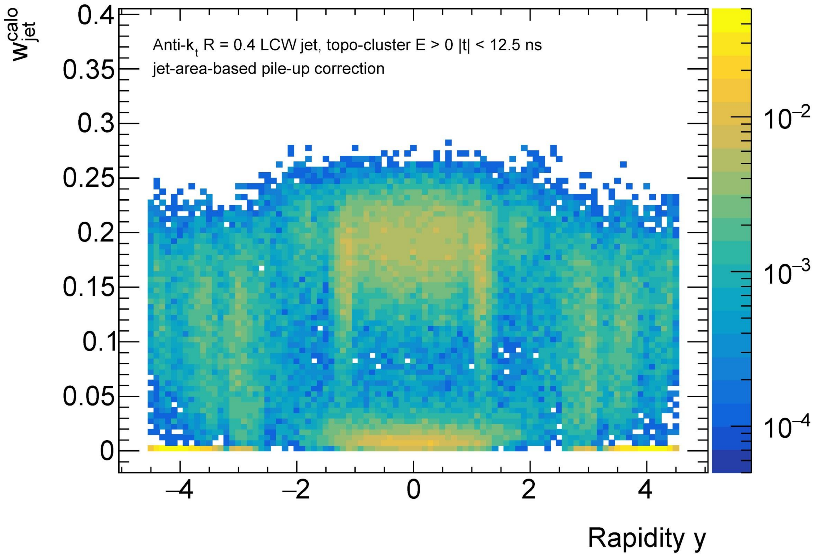
Signal Likelihood in Jet Shapes

(width, topo-cluster)

RATIO #SIGNAL JETS/#ALL JETS



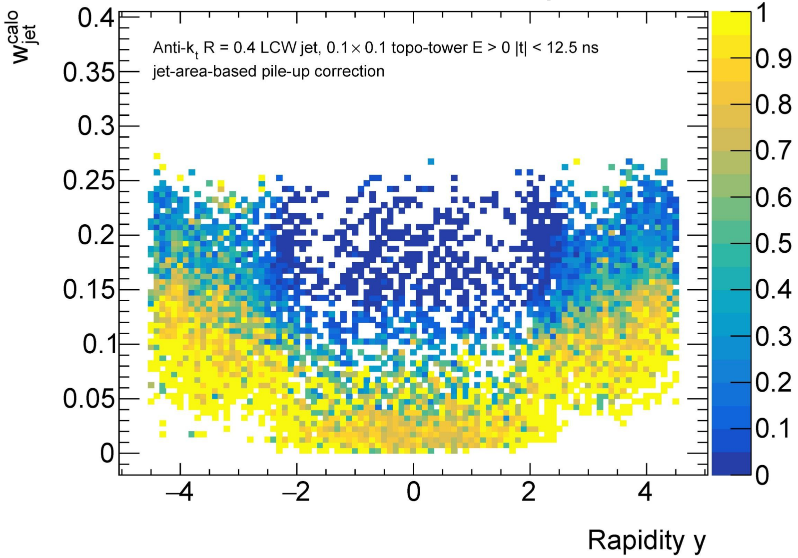
TIMING CUT



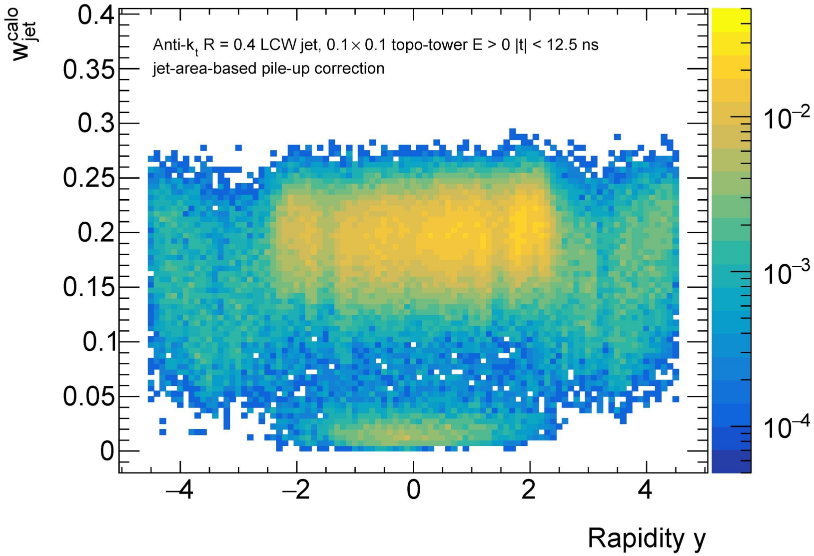
Signal Likelihood in Jet Shapes

(width, topo-tower 0.1 x 0.1)

RATIO #SIGNAL JETS/#ALL JETS



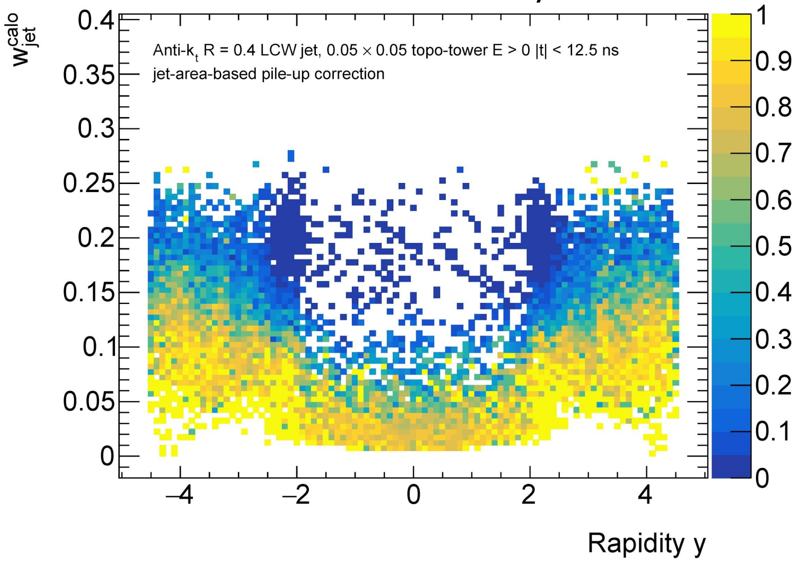
TIMING CUT



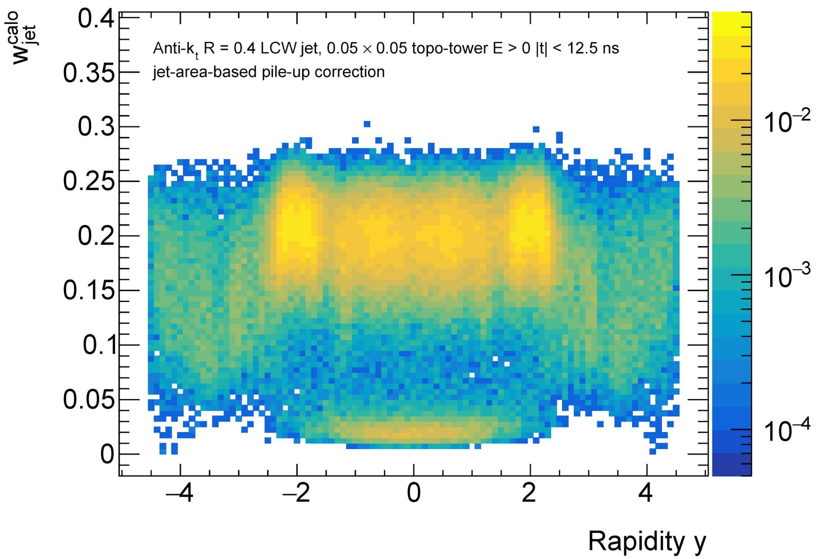
Signal Likelihood in Jet Shapes

(width, topo-tower 0.05 x 0.05)

RATIO #SIGNAL JETS/#ALL JETS



TIMING CUT



Backup: Machine Learning

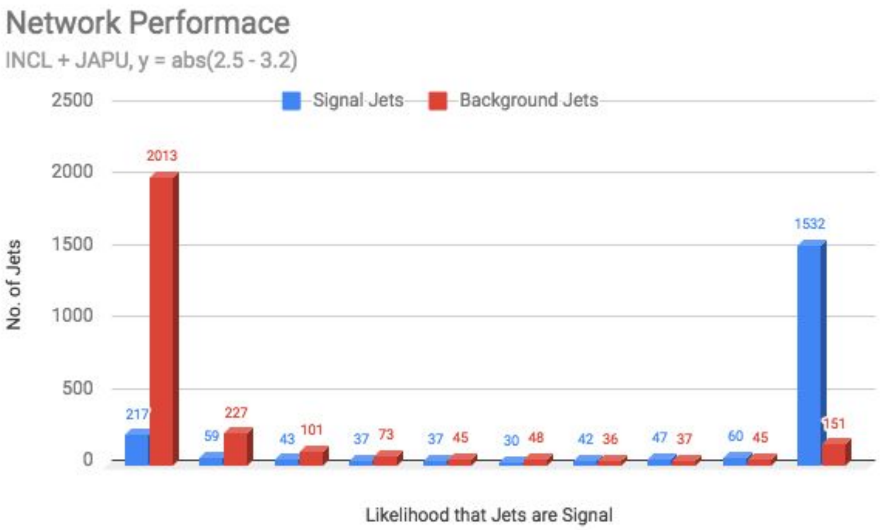
Current Neural Network

```
net = tflearn.input_data(shape=[None, 5])

net = tflearn.fully_connected(net, 32)

dropout1 = tflearn.dropout(net, 0.8)

net = tflearn.fully_connected(dropout1, 2, activation='softmax')

net = tflearn.regression(net)
```

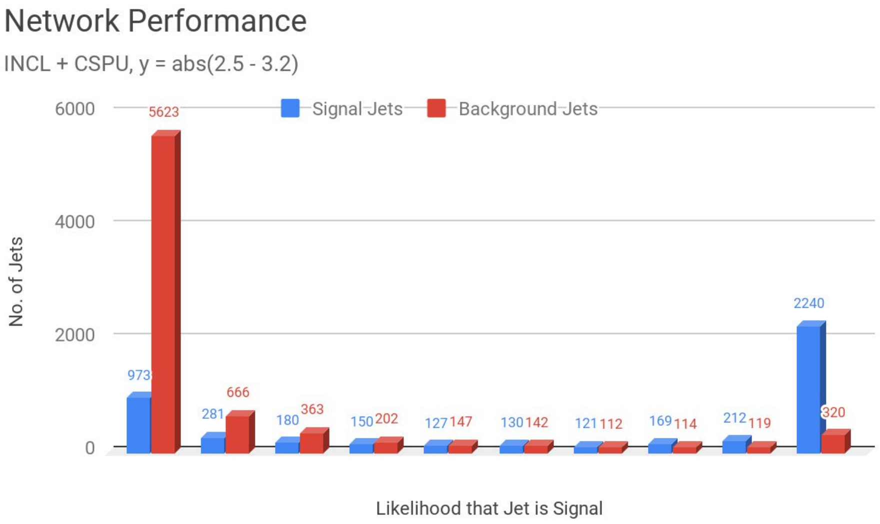
Network Performance

Topo-Clusters

Area-Based Pile-up Suppression



Constituent Subtraction



Topo-Clusters

Percentage of Background/Signal Jets Predicted Correctly

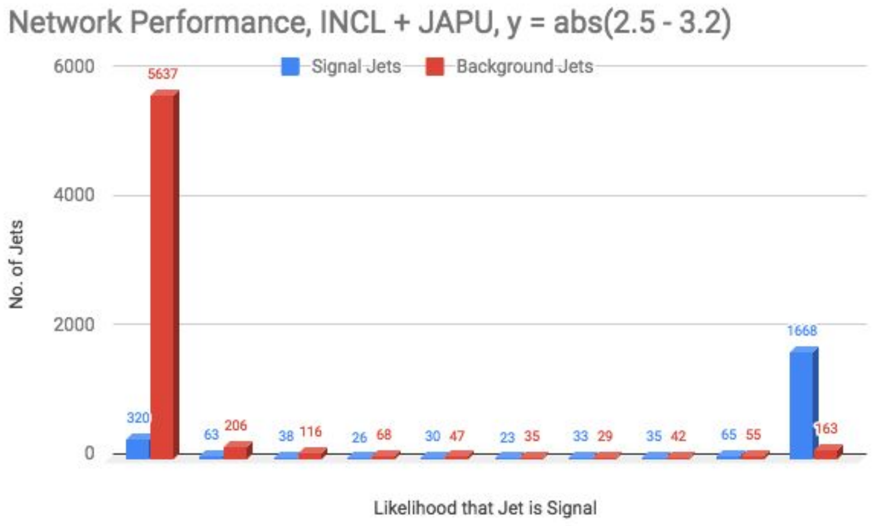
Selection	Background	Signal	# training pts.	# test pts.
INCL + JAPU	90.31%	79.90%	4,183	4,880
TIME + JAPU	92.13%	79.27%	4,093	4,839
INCL + CSPU	89.16%	82.37%	4,527	8,269
TIME + CSPU	71.46%	87.30%	607	7,104

- Cutting at likelihood which maximizes jets identified correctly
- Rapidity Region: y = abs(2.5-3.2)

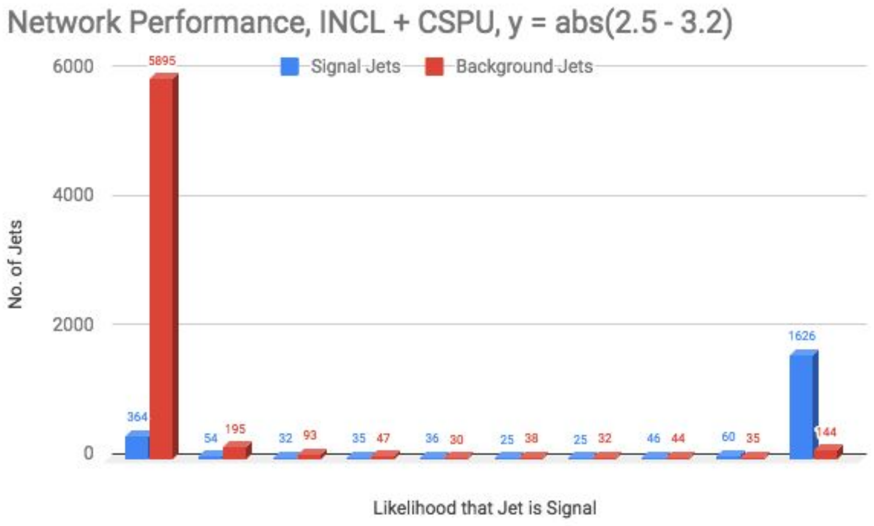
Network Performance

Topo-Towers 0.1 x 0.1

Area-Based Pile-up Suppression



Constituent Subtraction



Topo-Towers 0.1 x 0.1

Percentage of Background/Signal Jets Predicted Correctly

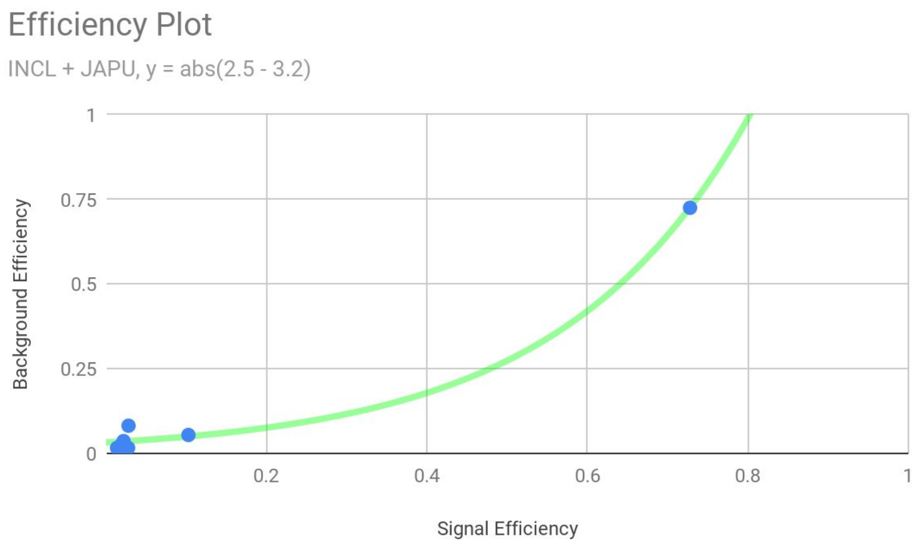
Selection	Background	Signal	# training pts.	# test pts.
INCL + JAPU	92.07%	71.74%	4,663	8,699
TIME + JAPU	81.88%	77.69%	3,661	4,635
INCL + CSPU	88.88%	76.50%	4,665	8,856
TIME + CSPU	84.09%	75.80%	3,659	4,651

- Cutting at likelihood which maximizes jets identified correctly
- Rapidity Region: y = abs(2.5-3.2)

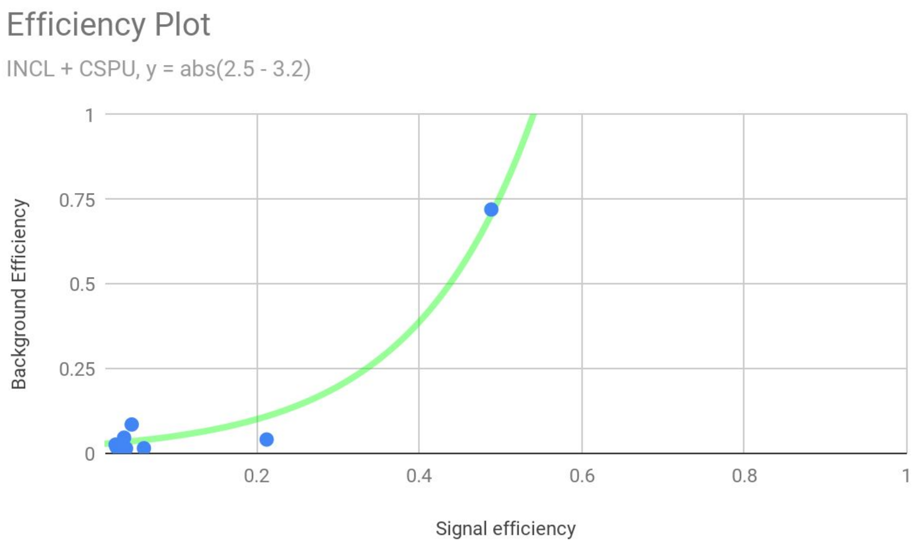
Efficiency Plot

Topo-Clusters

Area-Based Pile-up Suppression



Constituent Subtraction

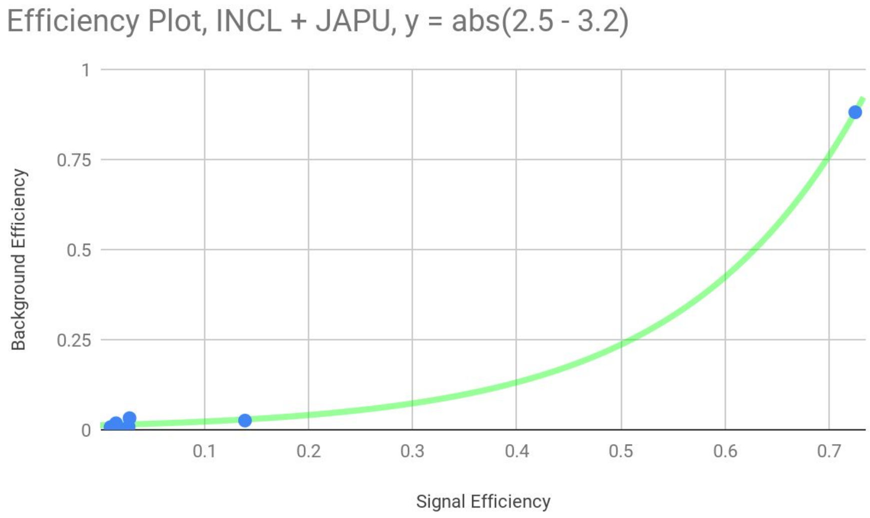


Note: See full Topo-Cluster results here.

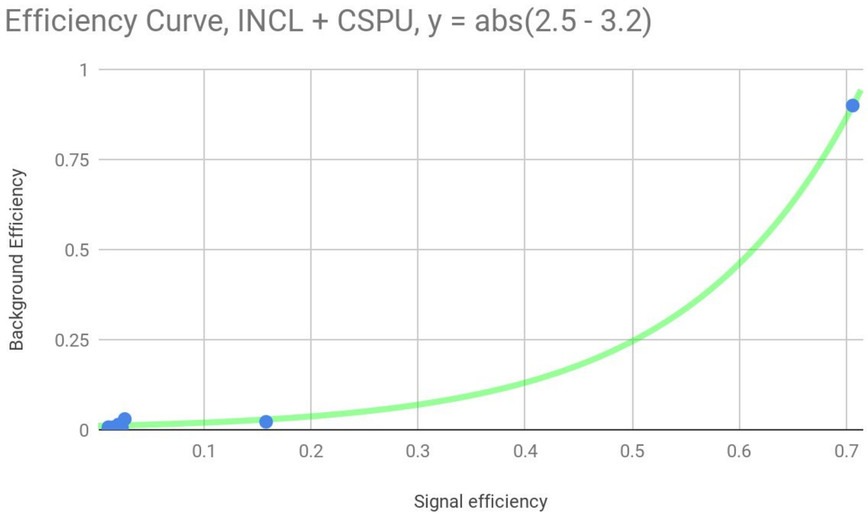
Efficiency Plot

Topo-Towers 0.1 x 0.1

Area-Based Pile-up Suppression



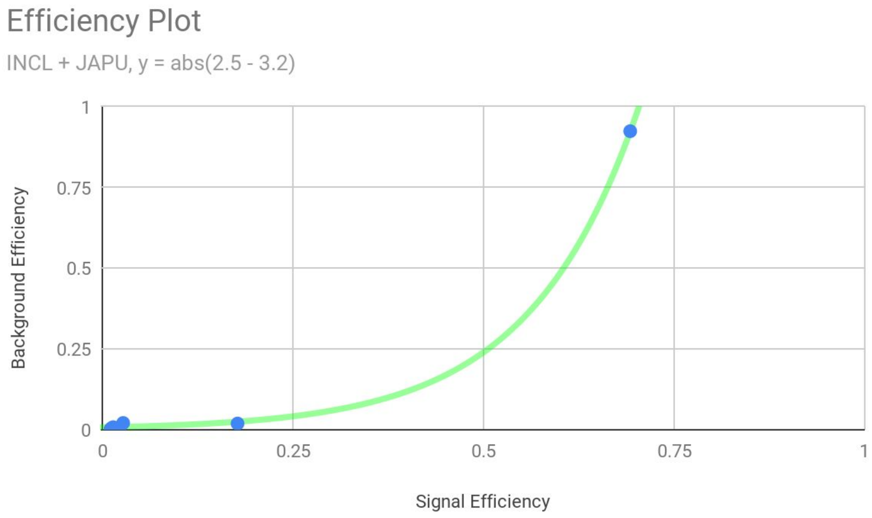
Constituent Subtraction



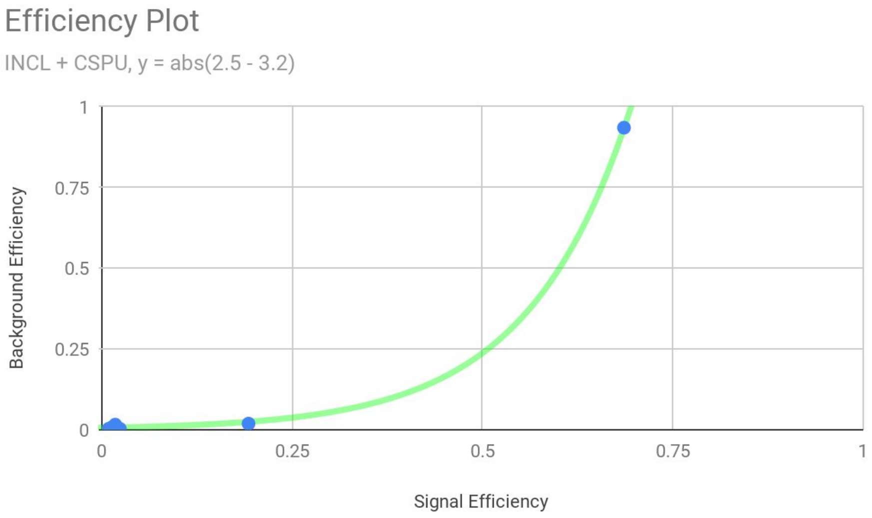
Note: See full Topo-Tower results here.

Efficiency Plot
Fine Topo-Towers 0.05 x 0.05

Area-Based Pile-up Suppression



Constituent Subtraction

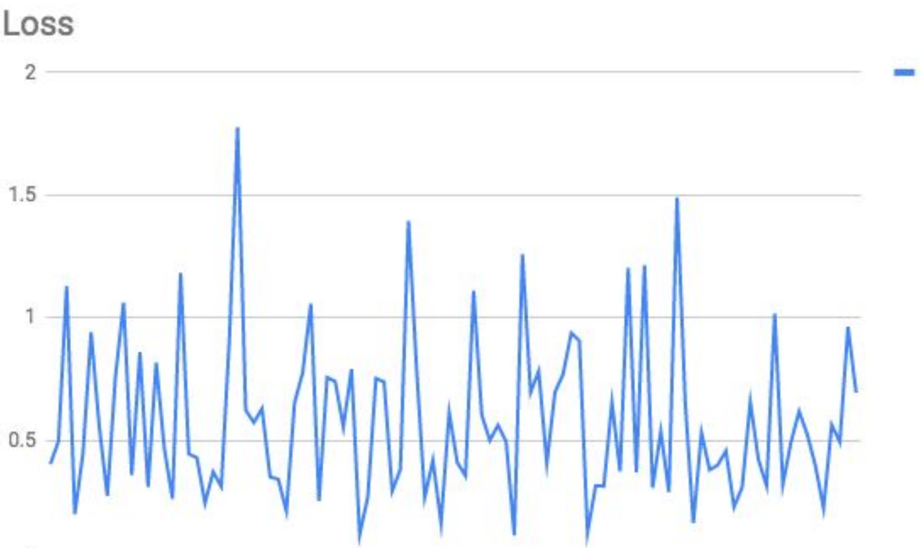


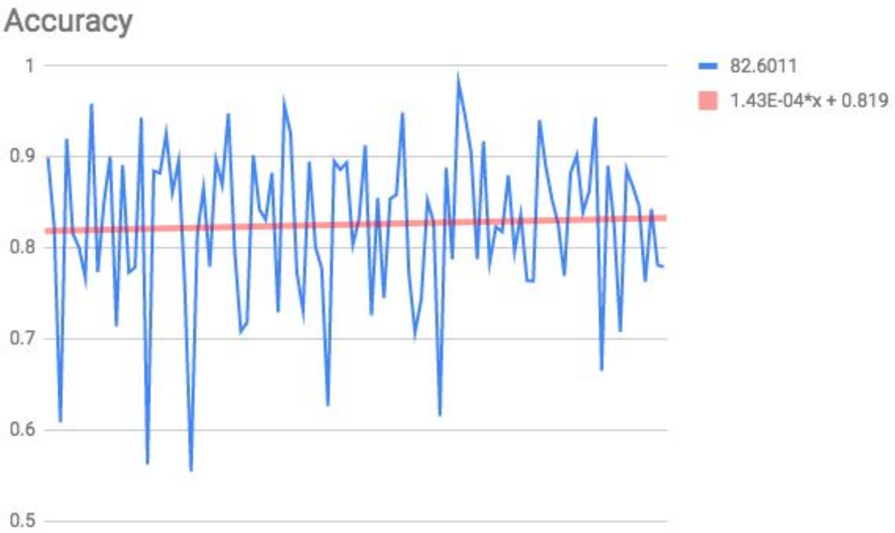
Note: See full Fine Topo-Tower results here.

Results for Different Observable Configurations

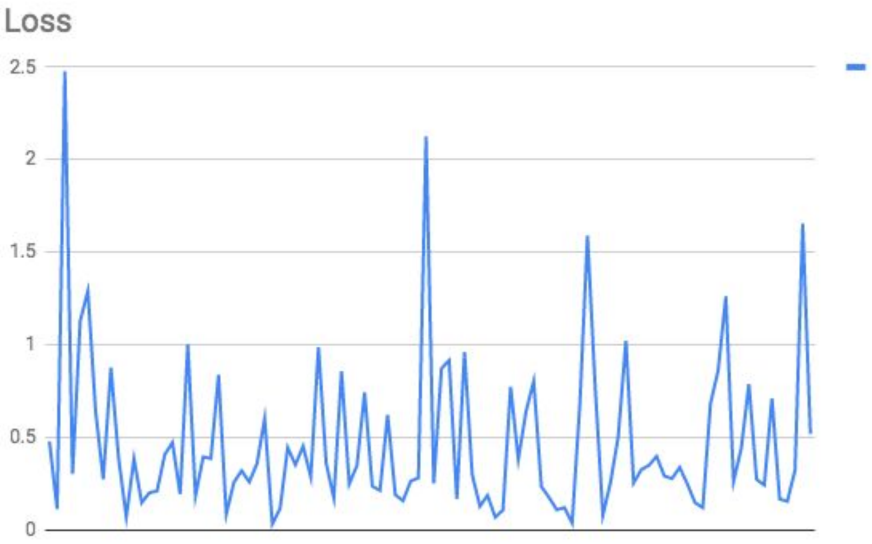
Configuration	Likelihood that Average Signal Jet is a Signal Jet	Likelihood that Average Background is Background	Accuracy
Mass	54.3%	81.7%	79.0%
Mass, p_T	98.6%	82.5%	81.9%
Mass, p_T, #of constituents	100%	100%	81.9%
Mass, p_T, #of constituents, p_T^D	100%	98.5%	84.9%
Mass, p_T, #of constituents, p_T^D, width	100%	99.0%	79.2%

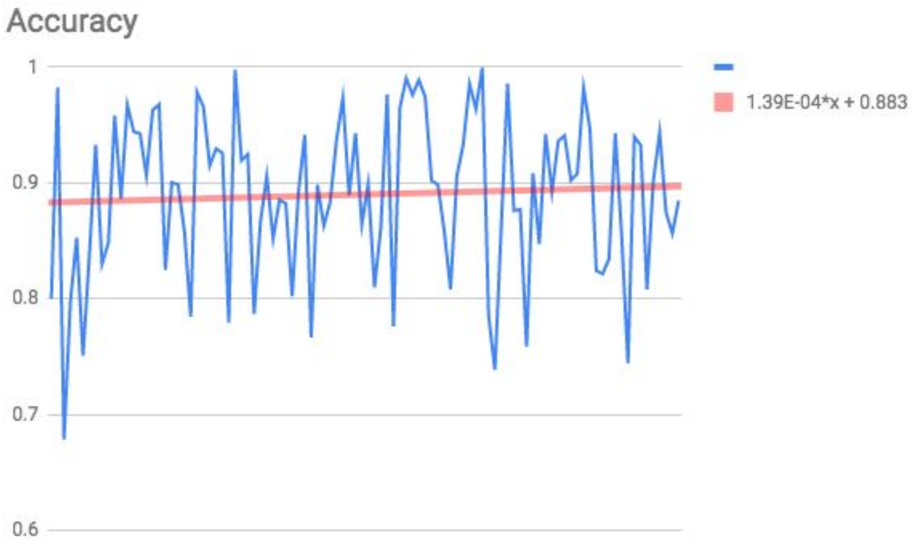
Plots: -4.9 < y < -3.2



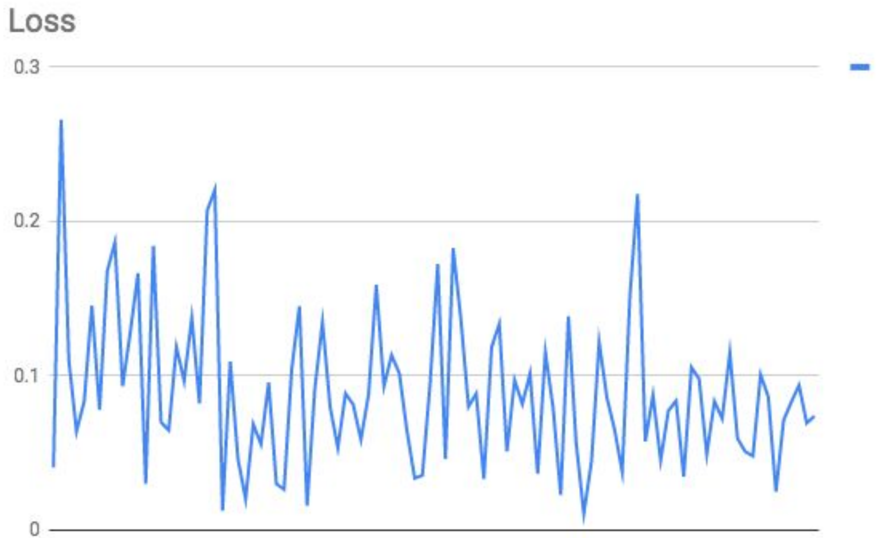


Plots: -3.2 < y < -2.5



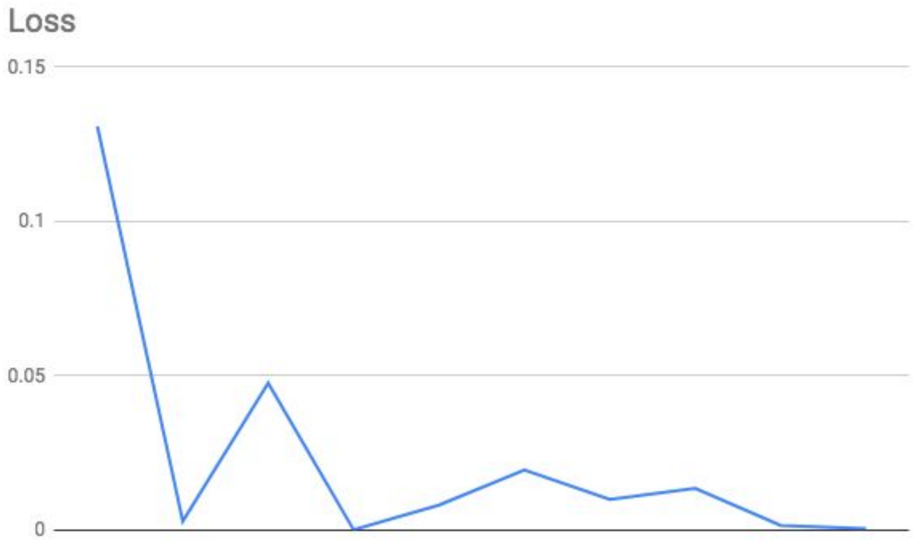


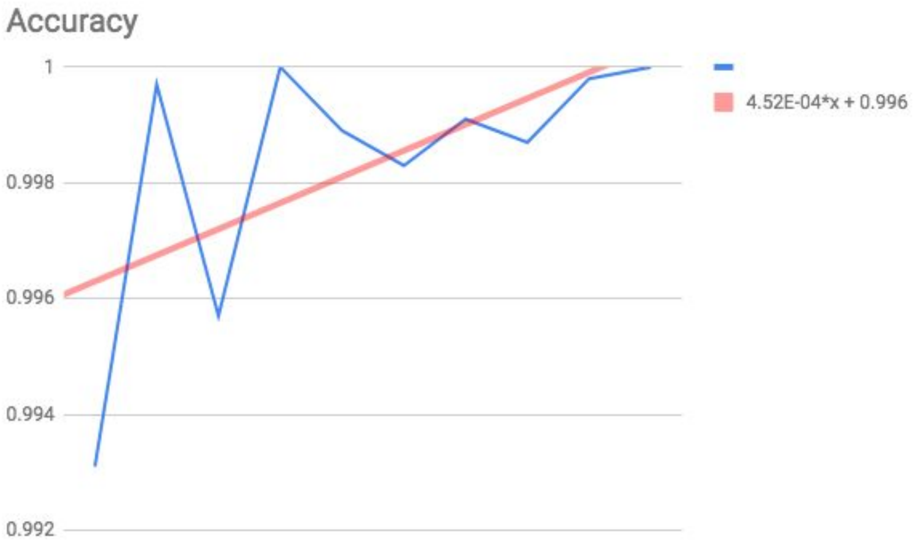
Plots: -2.5 < y < 0


Loss
0.3
0.2
0.1
0

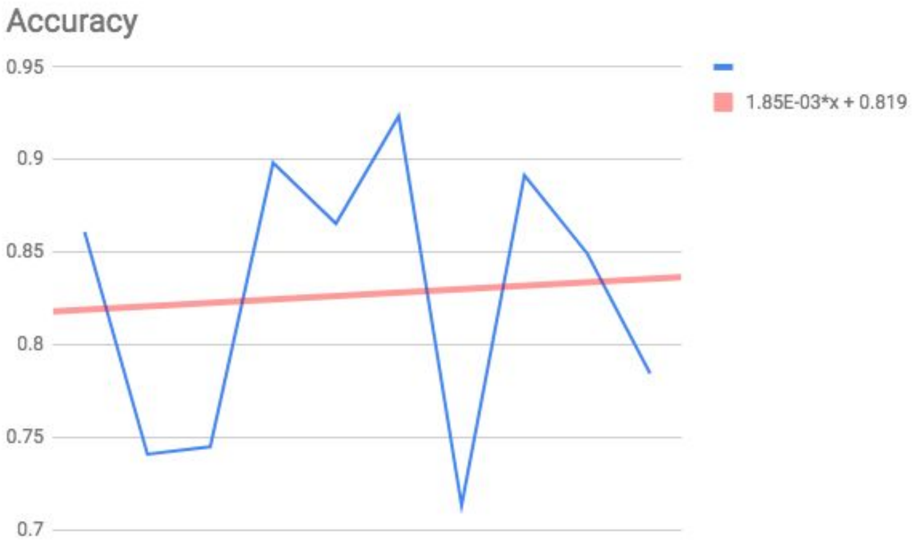

Accuracy Epoch: 10, Batch Size: 1
1
0.98
0.96
0.94
0.92
-4.03E-03*x + 0.994

Plots: 0 < y < 2.5

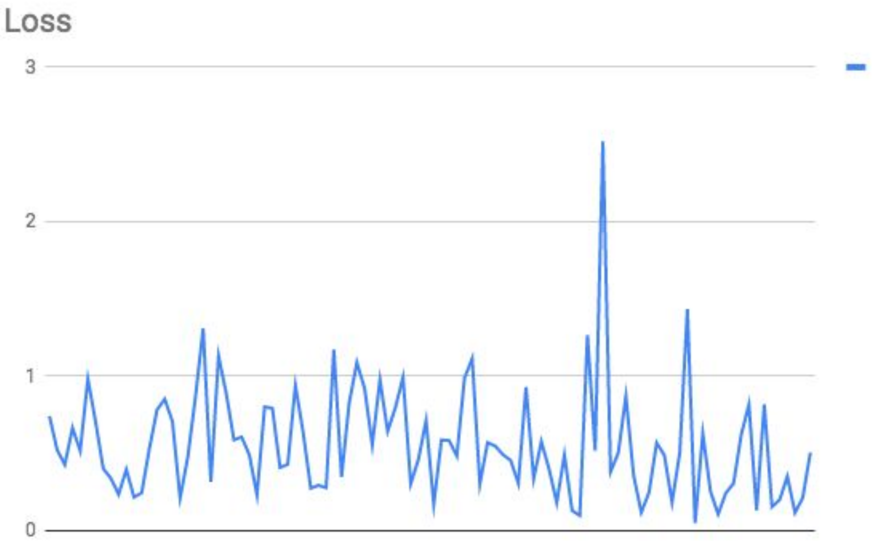


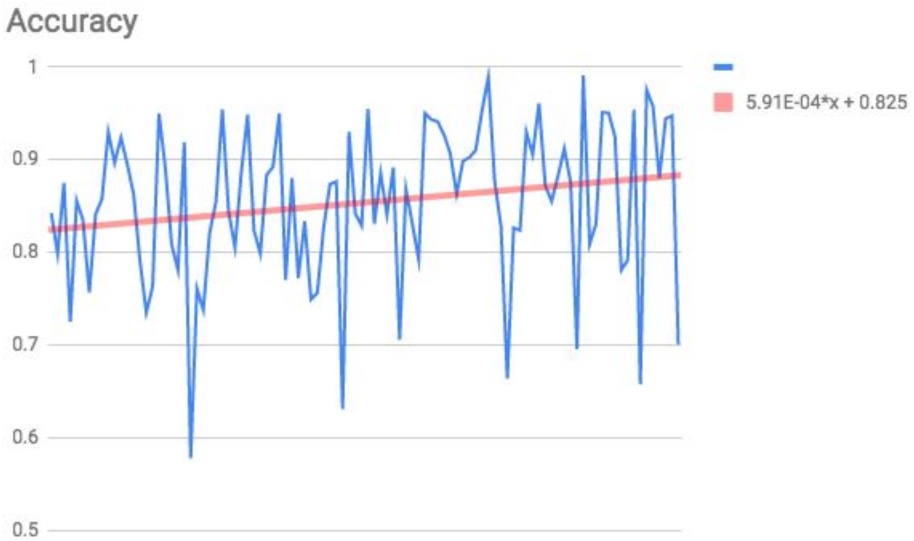


Plots: 2.5 < y < 3.2

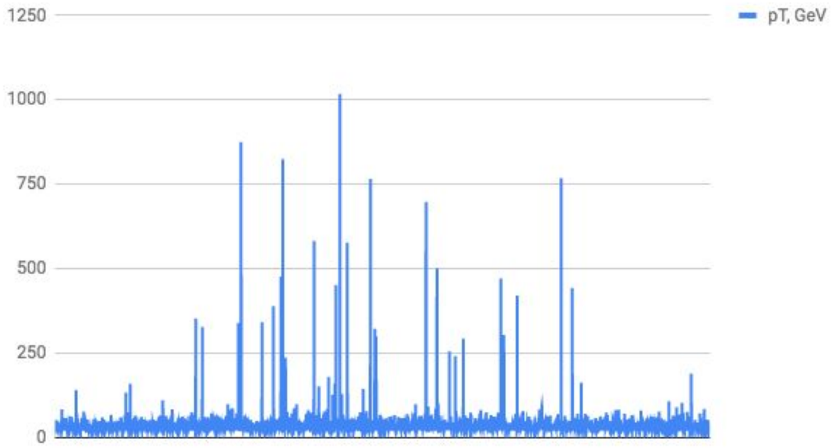


Plots: 3.2 < y < 4.9

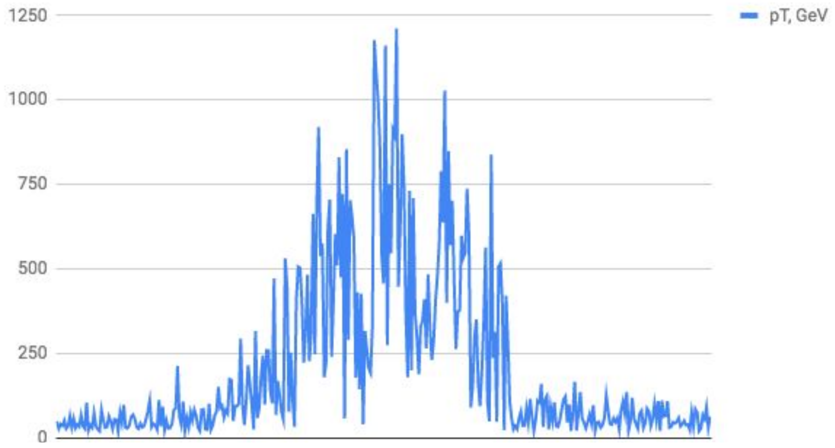




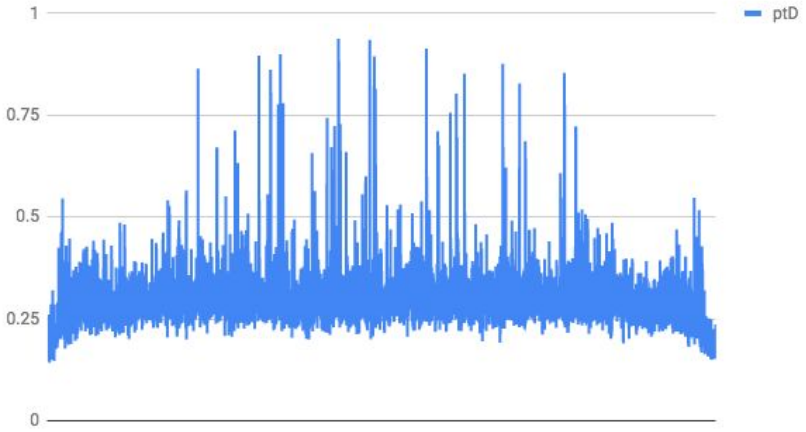
Background pT, GeV



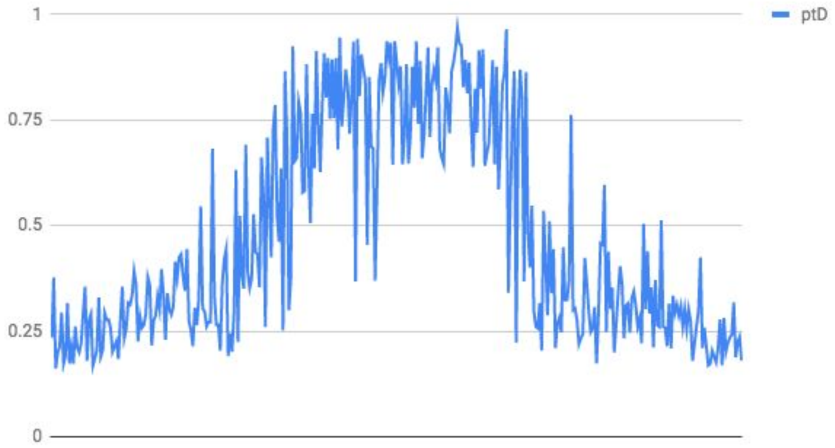
Signal pT, GeV



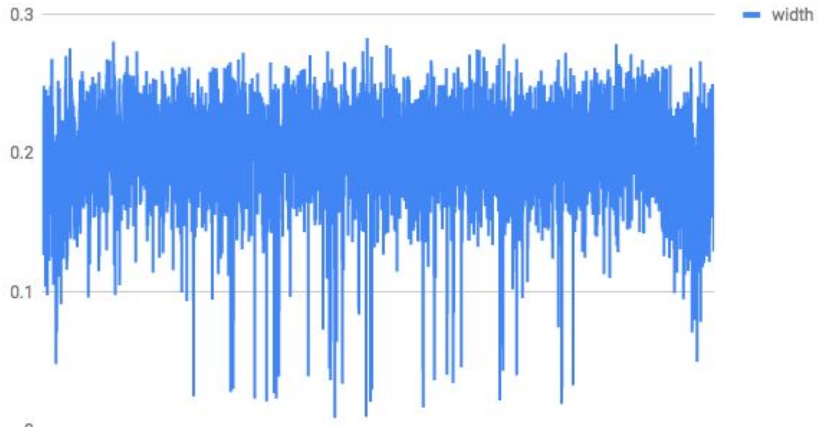
Background ptD



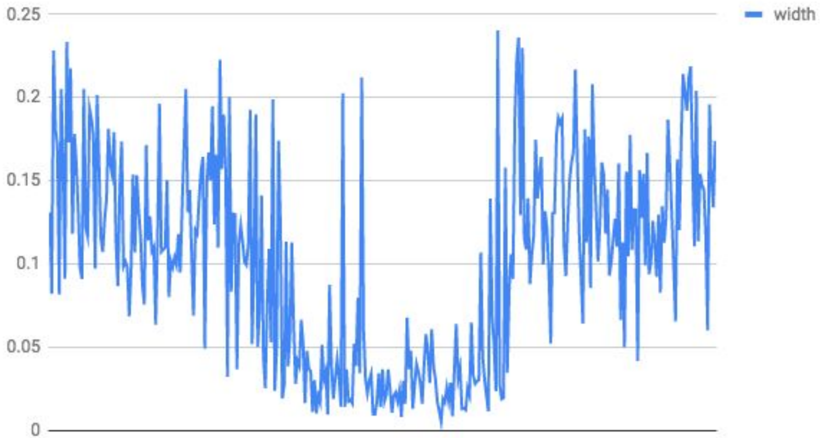
Signal ptD



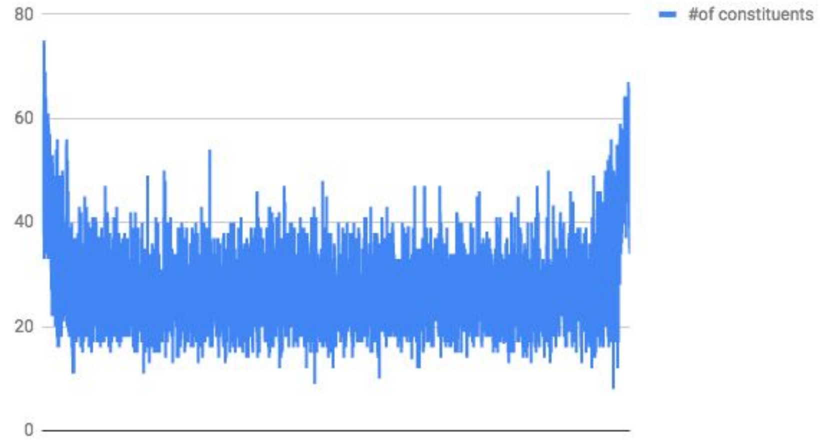
Background width



Signal width



Background #of constituents



Signal #of constituents