Machine Learning in Jet Physics

Sreedevi Varma

King's College London

September 26, 2019

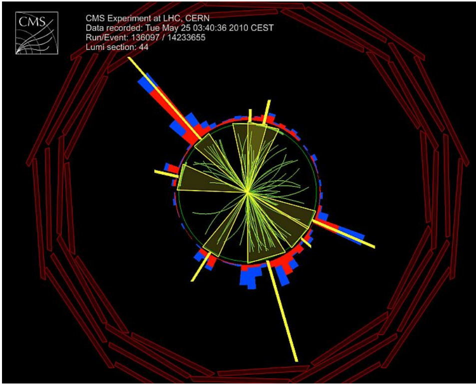
Outline

1. Jets
   - N-Subjettiness
   - Jet-Images
2. Machine Learning
   - Neural Networks
   - Convolutional Neural Networks
3. Top/QCD classification
   - Deep Neural Network on N-subjettiness variables
   - Convolutional Neural Network on Jet Images
4. Conclusions

What is a Jet?

Jets - Definition

- Jets are collimated stream of particles produced by particle collisions.
- Jets are used to interpret complex hadronic activities.



Why Jets?

- We get information of regions of the detector by clustering the particles into jets.
- Jets show properties of the initial hard-process and are used to classify quark-initiated jets from gluon initiated jets.
- Jets are an efficient tool in the classification of the hadronic decay of heavy particles and hadronic activity of a QCD processes in the final state.

N- Subjettiness

- N-subjettiness is a jet shape used for tagging boosted objects.
- Variables quantifying the amount of radiation contained within a jet (event) is aligned along different (sub)jet axes.

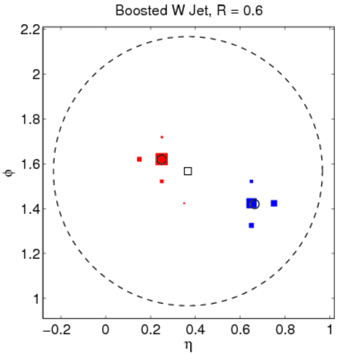
$$\tau_N^{(\beta)} = \frac{1}{p_{T,J}} \sum_{i \in J} p_{T,i} \min \left\{ R_{1i}^{\beta}, R_{2i}^{\beta}, \ldots, R_{Ni}^{\beta} \right\}$$

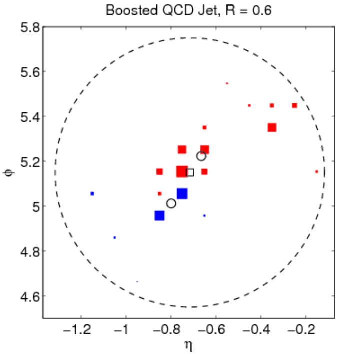
where,

- R_{Ni} is the distance in the $\eta - \phi$ plane of the jet constituent i to the axis N.
- p_T is the transverse momentum.
- β is an angular exponent.

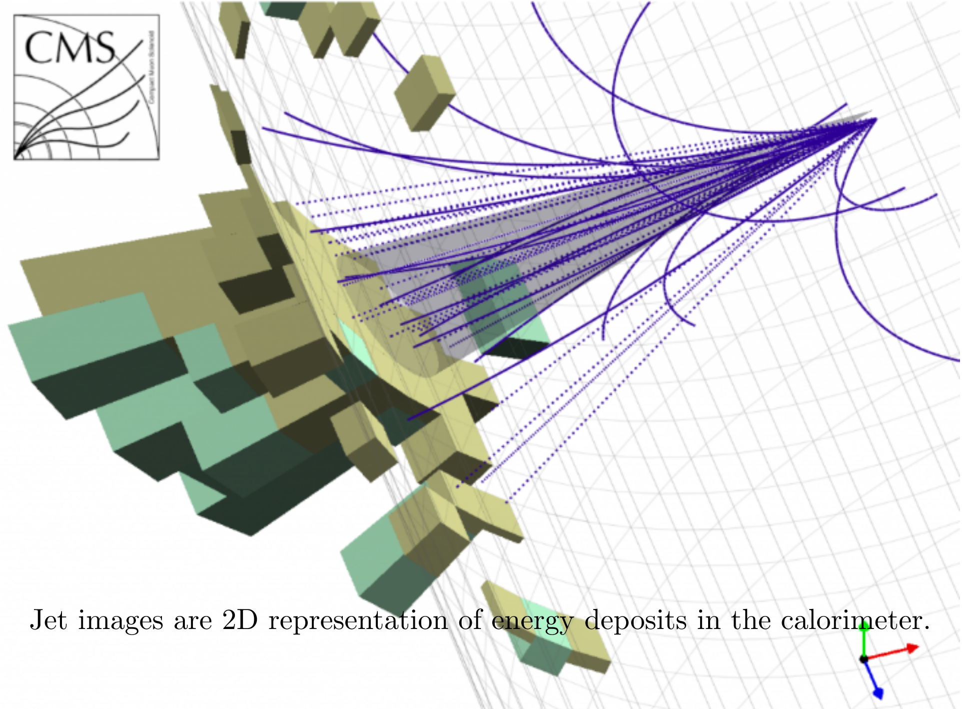
K. Datta, A.Larkoski, How Much Information is in a Jet?, JHEP06, 073 (2017), doi:10.1007/JHEP06(2017)073, 1704.08249

- τ_{21}=τ_2/τ_1 is a n-subjettiness variable for 2-pronged jets.





J. Thaler and K. Van Tilburg, Identifying Boosted Objects with N-subjettiness, JHEP03, 015 (2011), doi:10.1007/JHEP03(2011)015, 1011.2268.



Jet images are 2D representation of energy deposits in the calorimeter.

Preproccessing - I

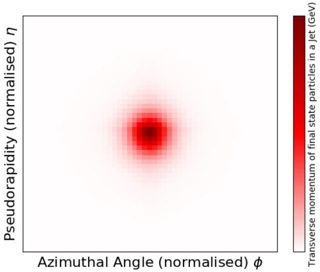
1. Centering: p_T weighted centroid of the jet is brought to $(0, 0)$.
2. Crop: Crop the image with $(\eta, \phi)\epsilon(-R, R)$.
3. Normalize: Total pixel intensity of the image is $\sum I_{ij} = 1$.
4. Zero-center: $I_{ij} \to I_{ij} - \mu_{ij}$, where μ_{ij} is the average of the training set.
5. Standardize: $I_{ij} \to I_{ij}/(\sigma_{ij} + r)$ where σ_{ij} is the standard deviation of the training set and $r = 10^{-5}$.

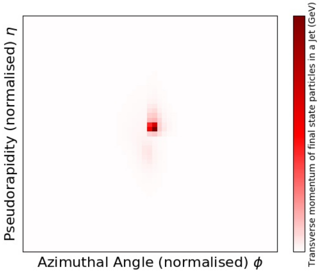
P. T. Komiske, E. M. Metodiev, M. D. Schwartz, Deep learning in color: towards automated quark/gluon jet discrimination, JHEP 01, 110 (2017), doi: 10.1007/JHEP01(2017)110, 1612.01551

Preproccessing- II

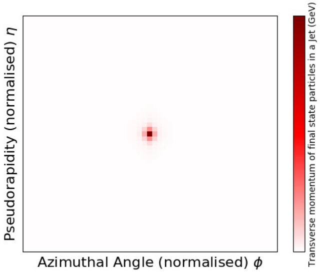
1. Centering: p_T weighted centroid of the jet is brought to $(0,0)$.
2. Rotation: Rotate the jet such that p_T weighted principal axis is vertical.
3. Flip: Images are flipped along L-R and U-D axes.
4. Normalize: Total pixel intensity of the image is $\sum I_{ij} = 1$.

S. Macaluso, D.Shih, Pulling Out All the Tops with Computer Vision and Deep Learning, JHEP 10, 121 (2018), doi:10.1007/JHEP10(2018)121, 1803.00107





Average top quark signal sample image after two different sets preprocessing steps.



Average qcd background sample image after two different sets preprocessing steps.

Machine Learning

- The most important part of machine learning is to build a model that can predict the results correctly.
- A hypothesis function is defined over the input variables to predict the output variable ($h : x \to y$).
- Cost function measures the accuracy of this hypothesis.
- With large number of variables, a linear hypothesis function fails to predict the results.

That's why we use Neural Networks!

Neural Networks

The hypothesis function of a neural network can be written as,

$$h_\theta(x) = a^{(j+1)} = g(\Theta^{(j)} a^{(j)})$$

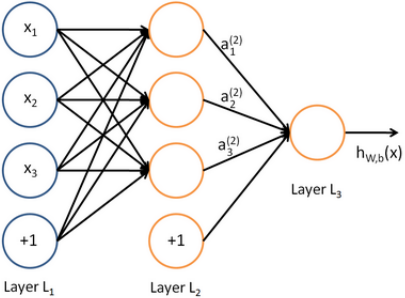


Figure 1: Neural Network architecture

Convolutional Neural Network

- Convolutional Neural Network (CNN) are neural networks for image recognition and image classification.
- CNN scans over the two dimensional pixel intensities of an RGB image.

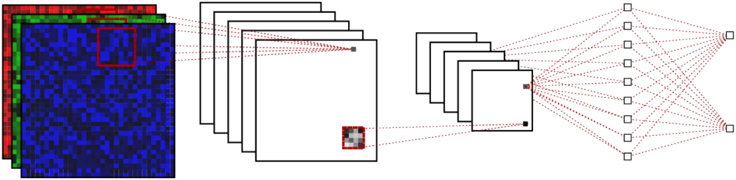


Figure 2: Convolutional Neural Network

P. T. Komiske, E. M. Metodiev, M. D. Schwartz, *Deep learning in color: towards automated quark/gluon jet discrimination*, JHEP 01, 110 (2017), doi: 10.1007/JHEP01(2017)110, 1612.01551

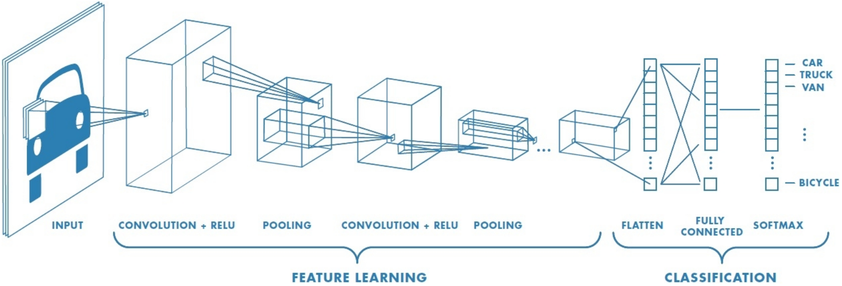
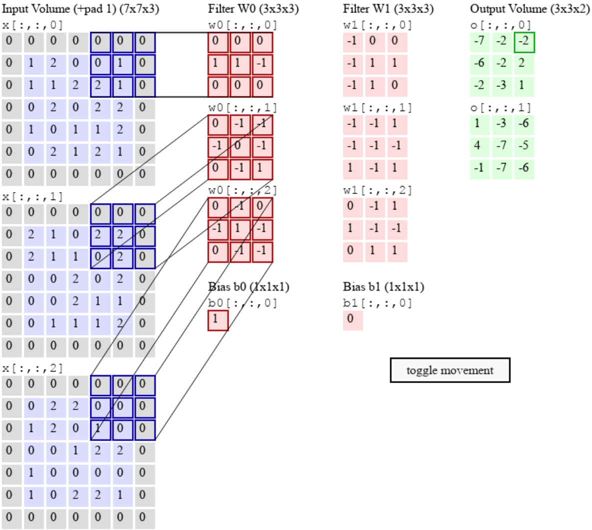


Figure 3: Components of a CNN



Convolutional Neural Networks for Visual Recognition, Available: http://cs231n.github.io/neural-networks-3/anneal

Our Work: 1807.04769

We compare the results of two machine learning techniques used to classify top jets with QCD jets:

Convolutional Neural Network on Jet Images

- Image recognition techniques to classify signal and background.

Deep Neural Network on N-subjettiness variables

- Physics motivated variable learned using DNN.

L. Moore, K. Nordström, S. V, M. Fairbairn, Reports of my demise are greatly exaggerated: N-subjettiness taggers take on jet images, SciPost Phys. 7, 036 (2019), doi: 10.21468/SciPostPhys.7.3.036, 1807.04769

Processes

Top quark tagging

- Signal : $pp \to W_1^- t$, $t \to W_2^+ b$, $W_2^+ \to jj$, $W_1^- \to e^- \bar{\nu}_e$
- Background : $pp \to W^- j, W \to e^- \bar{\nu}_e$
- p_T ranges: $[350, 400]$ GeV, $[500, 550]$ GeV and $[1300, 1400]$ GeV
- Radius $= 0.8$ ($[1300, 1400]$ GeV), 1.5
- Pseudorapidity $|\eta| < 2.5$ ($[1300, 1400]$ GeV) and $|\eta| < 1.0$

Machine Learning on N-subjettiness variables

The input variables for the Deep Neural Network are,

$$\left\{\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \ldots, \tau_{M-2}^{(0.5)}, \tau_{M-2}^{(1)}, \tau_{M-2}^{(2)}, \tau_{M-1}^{(1)}, \tau_{M-1}^{(2)}\right\}$$

- This covers M-body phase space with $3M-4$ observables.
- Jet mass is also included as an input to the neural network.

K. Datta, A.Larkoski, How Much Information is in a Jet?, JHEP06, 073 (2017), doi:10.1007/JHEP06(2017)073, 1704.08249

The network consists of,

- Four fully connected hidden layers.
- First two with 300 nodes and a dropout regularisation of 0.2.
- Last two with 100 nodes and a dropout regularisation of 0.1.
- Activation: ReLU.

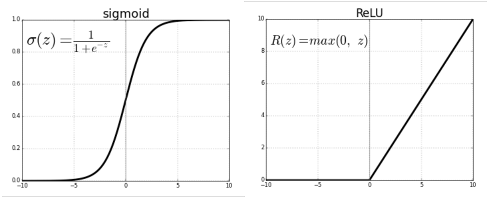


Figure 4: ReLU and Sigmoid activations

- Optimizer: Adam.
- Learning rate $\alpha = 0.001$.

Machine Learning on Jet-Images - CNN

- Jet images of size 51×51 (37×37 for pT in $[1300, 1400]$ GeV)).
- 3 convolutional layer and 2 fully connected layer.
- ReLU activation.

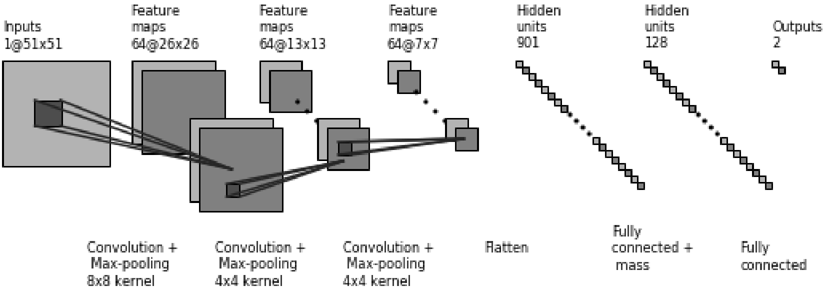


Figure 5: Network architecture of the CNN. This figure was generated by adapting the code from `https://github.com/gwding/draw_convnet`.

Machine Learning on Jet-Images - CNN1

- CNN1 has two blocks each with two convolutional layers.
- A maxpooling layer is applied after each block.

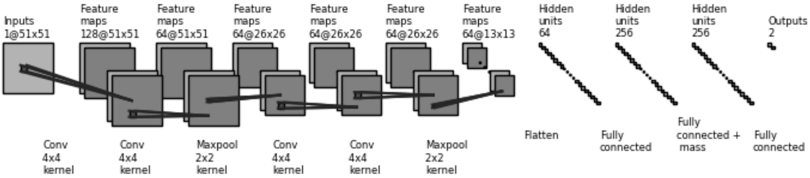


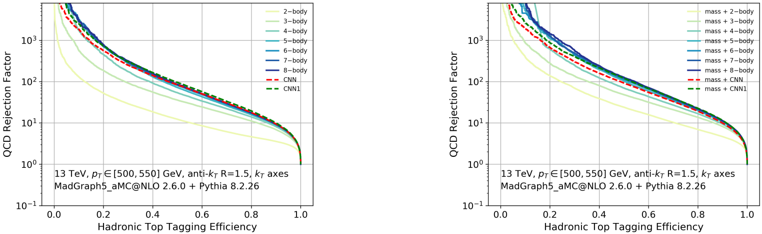
Figure 6: Network architecture of the CNN1. This figure was generated by adapting the code from `https://github.com/gwding/draw_convnet`

.

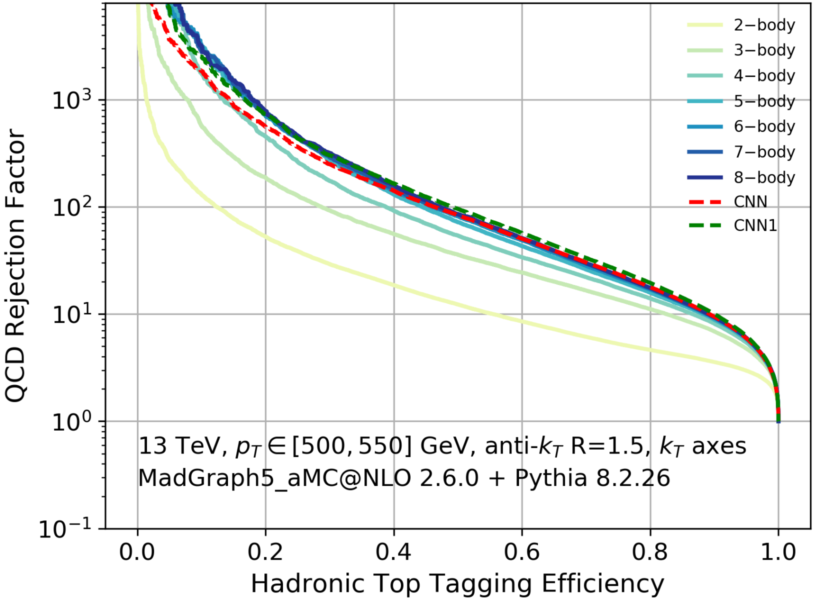
- Convolutional Neural Network is trained on Tensorflow using NVIDIA GeForce 1080Ti GPU on Cuda 9.0 platform.
- The network is trained over 50 epochs with a learning rate α of 0.001.
- 2M jet images are used for training, 100k images for validation and 200k images are used for testing.

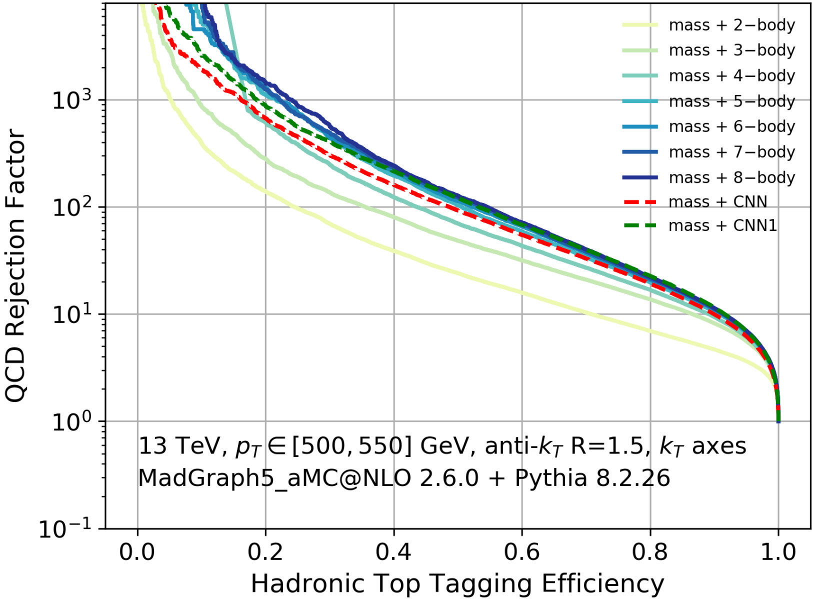
Conclusions

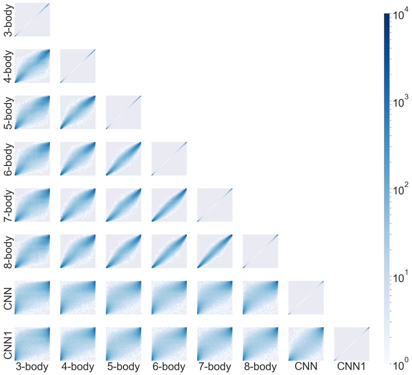
- Mass as an additional information improves the performance of the network.
- As the ROC curves are matching, we can conclude that the information learned from both methods are the same.



ROC curves for top quark tagging without mass on the left and with mass on the right, for $p_T \in [500, 550]$ GeV.


QCD Rejection Factor
10^3
10^2
10^1
10^0
10^{-1}
0.0
0.2
0.4
0.6
0.8
1.0
Hadronic Top Tagging Efficiency
2−body
3−body
4−body
5−body
6−body
7−body
8−body
CNN
CNN1
13 TeV, $p_T \in [500, 550]$ GeV, anti-k_T R=1.5, k_T axes
MadGraph5_aMC@NLO 2.6.0 + Pythia 8.2.26


QCD Rejection Factor
10^3
10^2
10^1
10^0
10^{-1}
0.0
0.2
0.4
0.6
0.8
1.0
Hadronic Top Tagging Efficiency
mass + 2−body
mass + 3−body
mass + 4−body
mass + 5−body
mass + 6−body
mass + 7−body
mass + 8−body
mass + CNN
mass + CNN1
13 TeV, $p_T \in [500, 550]$ GeV, anti-k_T R=1.5, k_T axes
MadGraph5_aMC@NLO 2.6.0 + Pythia 8.2.26



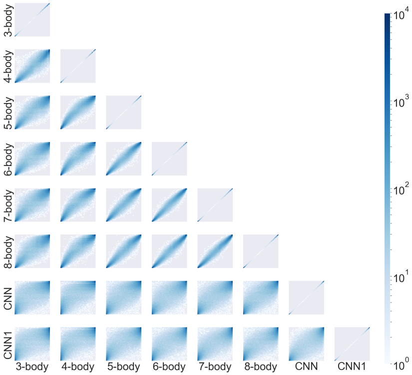


Figure 7: Correlation plots for top quark tagging without mass on the left and with mass on the right, for $p_T \in [500, 550]$ GeV.

Thank you!

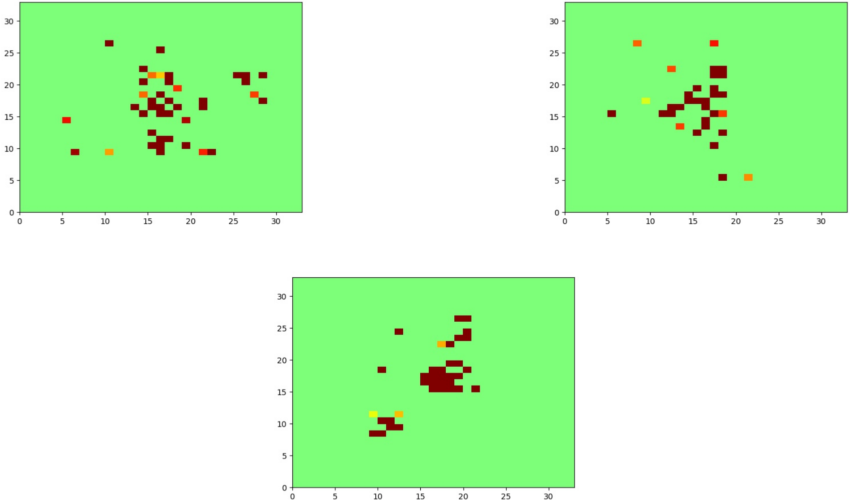


Figure 8: Sample top images

Measuring Jets

- Jet has a four-momentum. $E = \sum_i E_i \quad \overrightarrow{p} = \sum_i \overrightarrow{p}_i$
- Transverse momentum of a jet:

$$p_T^{JET} = \sqrt{p_x^2 + p_y^2}$$

- The radius of the jet (R) is given by,

$$R^2 = (\eta_i - \eta^{JET})^2 + (\phi_i - \phi^{JET})^2$$

- Position in the collider in two coordinates:
  - Pseudorapidity of the jet (η):

$$\eta^{JET} = -ln(tan\frac{\theta}{2})$$

    where θ is the polar angle and $cos\theta = \frac{\sqrt{p_x^2 + p_y^2}}{p_z}$

  - Azimuthal angle of the jet (ϕ):

$$\phi^{JET} = tan^{-1}(\frac{p_y}{p_x})$$

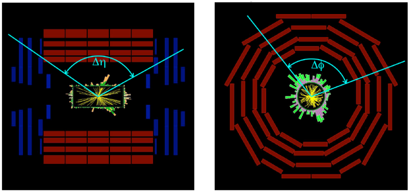


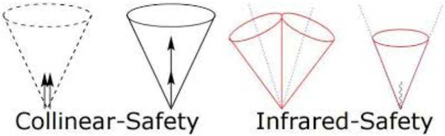
Figure 9: Coordinates

Jet Clustering

- The final state particles from the collisions are clustered using jet algorithms.
- Jet algorithms have different types:
  - Cone Algorithms
  - Clustering Algorithms

Sequential clustering algorithms

- Sequential clustering algorithms are the most commonly used algorithms today.
    - Combines particles according to the distance between them.
    - Infrared and collinear (IRC) safe.
        - Infrared safety: The outcome is not affected by the emission of a low energy (soft) gluon.
        - Collinear safety: The outcome is not affected when the gluons are emitted in a very close angle to the parton in the event.



R. Ellis, W. Stirling, and B. Webber,QCD and Collider Physics, ser. Cambridge Monographson Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2003.

Sequential clustering algorithms

- Cluster particles which have smallest distance between them in the momentum space.

$$d_{ij} = min(p_{ti}^a, p_{tj}^a) \times \frac{R_{ij}^2}{R}$$

where R is the radius of the cone and R_{ij} is the distance between particles in (η, ϕ) space.

$$R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$

$$d_{iB} = p_{ti}^a$$

R. Atkin, Review of jet reconstruction algorithms, Journal of Physics: Conference Series, vol.645, no. 1, p. 012008, 2015. [Online]. Available:http://stacks.iop.org/1742-6596/645/i=1/a=012008

K_t

K_t:

- a =2.

$$d_{ij} = min(p_{ti}^2, p_{tj}^2) \times \frac{R_{ij}^2}{R}$$

  and $d_{iB} = p_{ti}^2$

- Clusters soft particles first.
- Good at resolving subjets.

R. Atkin, Review of jet reconstruction algorithms, Journal of Physics: Conference Series, vol.645, no. 1, p. 012008, 2015. [Online]. Available:http://stacks.iop.org/1742-6596/645/i=1/a=012008

Anti-K_t

Anti-K_t:

- a =-2.

$$d_{ij} = min(1/p_{ti}^2, 1/p_{tj}^2) \times \frac{R_{ij}^2}{R}$$

  and $d_{iB} = 1/p_{ti}^2$
- Clusters hard particles first.
- Good resolving power.

R. Atkin, Review of jet reconstruction algorithms, Journal of Physics: Conference Series, vol.645, no. 1, p. 012008, 2015. [Online]. Available:http://stacks.iop.org/1742-6596/645/i=1/a=012008

Cambridge/Aachen

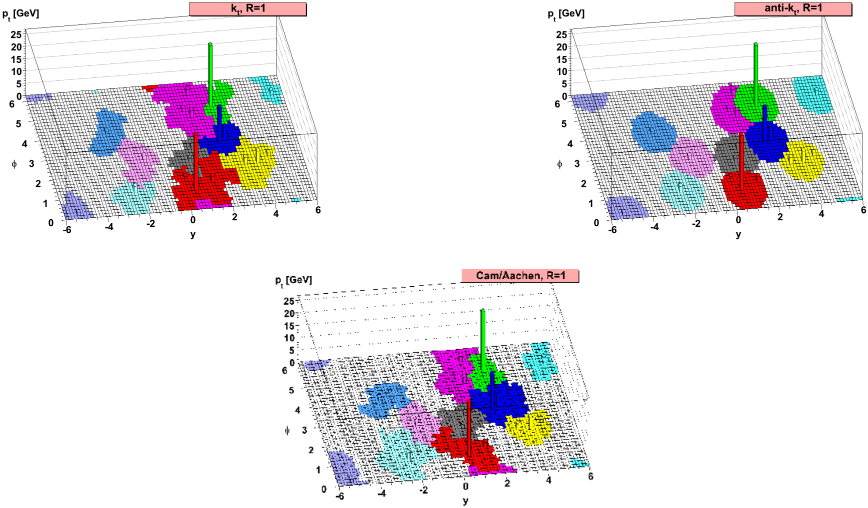
Cambridge/Aachen:

- a =0.

$$d_{ij} = \frac{R_{ij}^2}{R}$$

  and $d_{iB} = 1$
- Both variables are independent of momentum.
- Best suited for studying the substructure.

R. Atkin, Review of jet reconstruction algorithms, Journal of Physics: Conference Series, vol.645, no. 1, p. 012008, 2015. [Online]. Available:http://stacks.iop.org/1742-6596/645/i=1/a=012008



Matteo Cacciari, Gavin P. Salam, Gregory Soyez, The anti-kt jet clustering algorithm, arXiv: arXiv:0802.1189 [hep-ph]

Jet-Substructure

- Jet substructure techniques exploit the internal structure of a jet.
- Two classes of jet- substructure techniques are:
  - Jet grooming : Eliminate extra energy deposits in the jet coming from pile-up, ISR and the uderlying event.
  - Jet tagging : Defining observables and distributions to classify signal and background jets.

Deepak Kar, Jet substructure: a discovery tool, Available: http://events.saip.org.za/getFile.py/access?resId=30materialId=6confId=53

ROC-Curves

- Receiver operating characteristic (ROC) is used to visulaize the performance of a binary classifier.

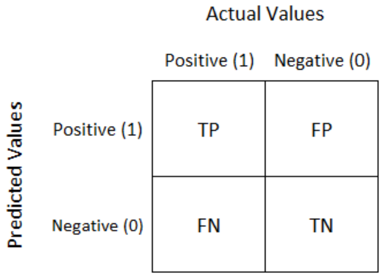


Figure 10: Confusion Matix

ROC-Curves

- Plot the true positive rates (TPR) and false positive rates (FPR) for every possible classification threshold to obtain a ROC curve.

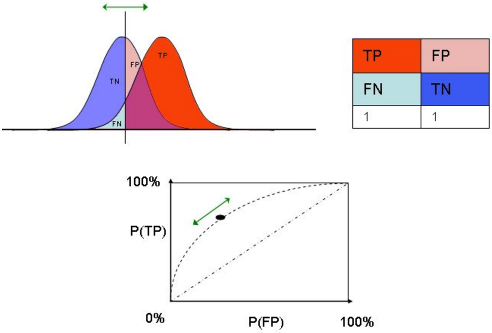


Figure 11: ROC-curve