CERN School of Computing

*From Quark to Jet: A Beautiful Journey*
*Lecture* ***2***

# Jet Clustering, Classification, and Personal Computing

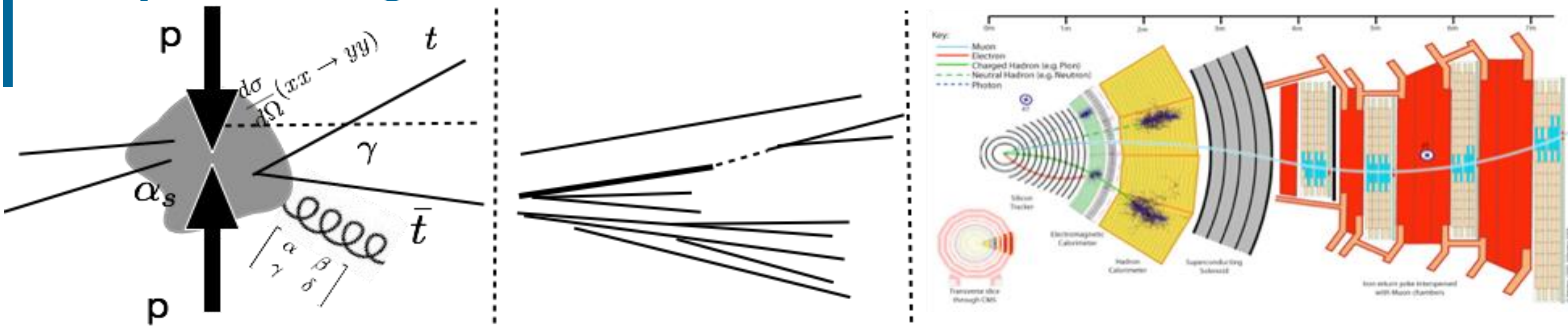
**Tyler Dorland**

**Deutsches Elektronen-Synchrotron (DESY)**

**Inverted CERN School of Computing, 24-25 February 2014**

CERN School of Computing

# Explaining the Title: An outline



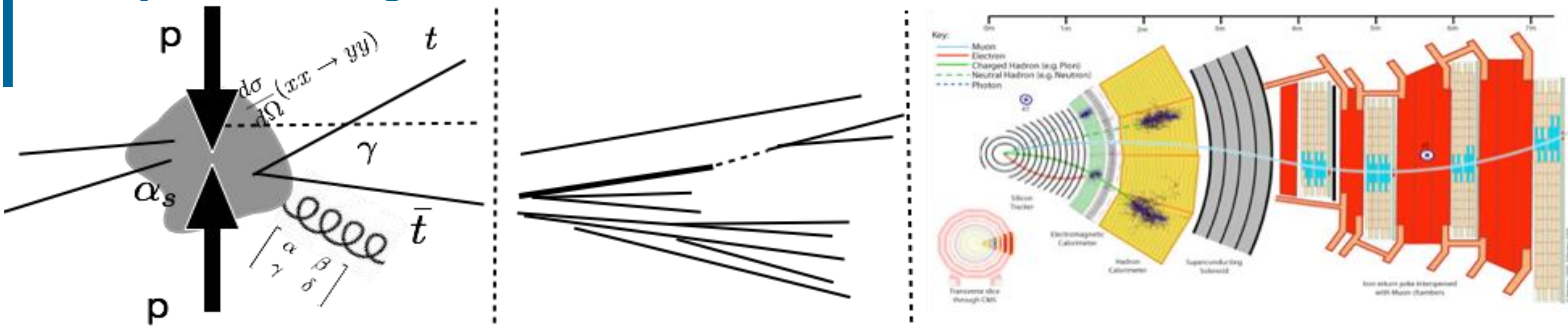

model uncertainties    hadronization schemes    experimental uncertainties

parton    particle    Reconstructed

- **Theoretical**

Quarks
Mathematical Objects:
Matrices, operators, etc.

# Explaining the Title: An outline




model uncertainties

hadronization schemes

experimental uncertainties

parton

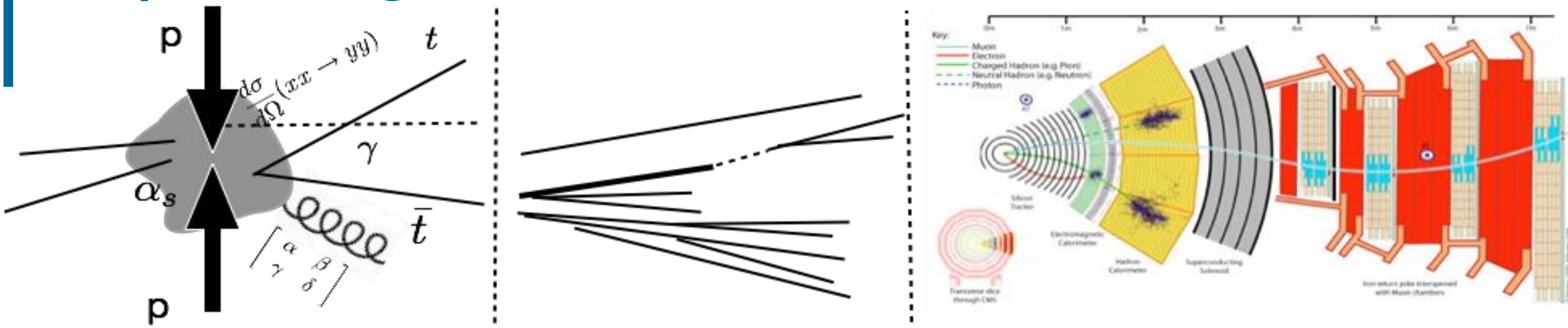
particle

Reconstructed

- **Theoretical**
- **Hadronization**

| Quarks | Particles |
| --- | --- |
| Mathematical Objects: Matrices, operators, etc. | intermediate and quasi-final state objects |

# Explaining the Title: An outline




model uncertainties
hadronization schemes
experimental uncertainties

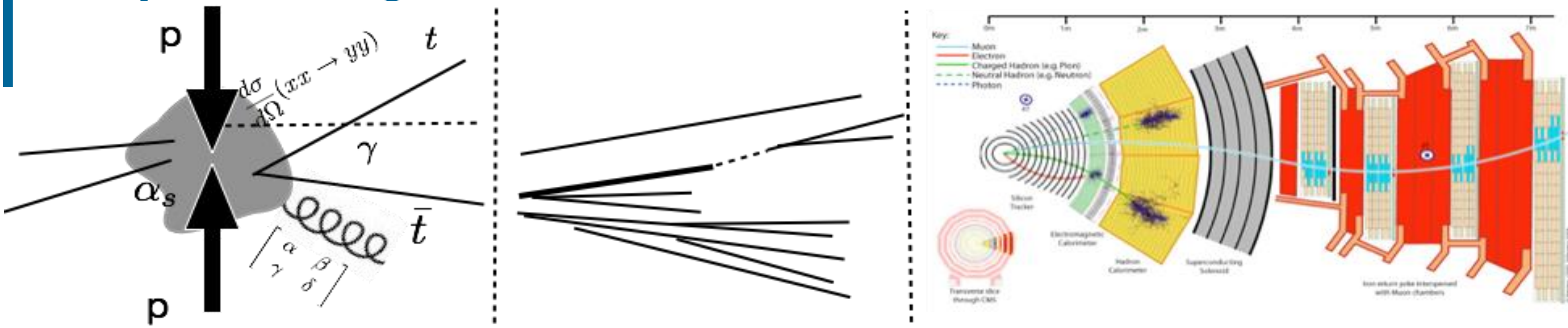
parton
particle
Reconstructed

- **Theoretical**
- **Hadronization**
- **Reconstruction**

| Quarks | Particles | Jets |
| --- | --- | --- |
| Mathematical Objects: Matrices, operators, etc. | intermediate and quasi-final state objects | Energy deposits in detector used to recreate particles |

CERN School of Computing

# Explaining the Title: An outline




model uncertainties
hadronization schemes
experimental uncertainties

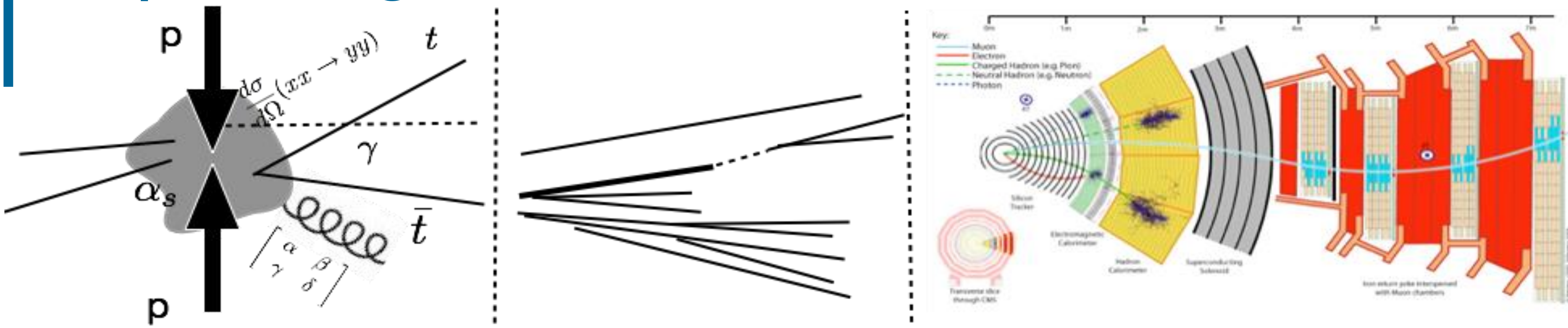
parton
particle
Reconstructed

- **Theoretical**
- **Hadronization**
- **Reconstruction**

| Quarks | Particles | Jets |
|---|---|---|
| Mathematical Objects: Matrices, operators, etc. | intermediate and quasi-final state objects | Energy deposits in detector used to recreate particles |

CERN School of Computing

# Explaining the Title: An outline




model uncertainties

parton

hadronization schemes

particle

experimental uncertainties

Reconstructed

- **Theoretical**
- **Hadronization**
- **Reconstruction**

| Quarks | Particles | Jets |
|---|---|---|
| Mathematical Objects: Matrices, operators, etc. | intermediate and quasi-final state objects | Energy deposits in detector used to recreate particles |

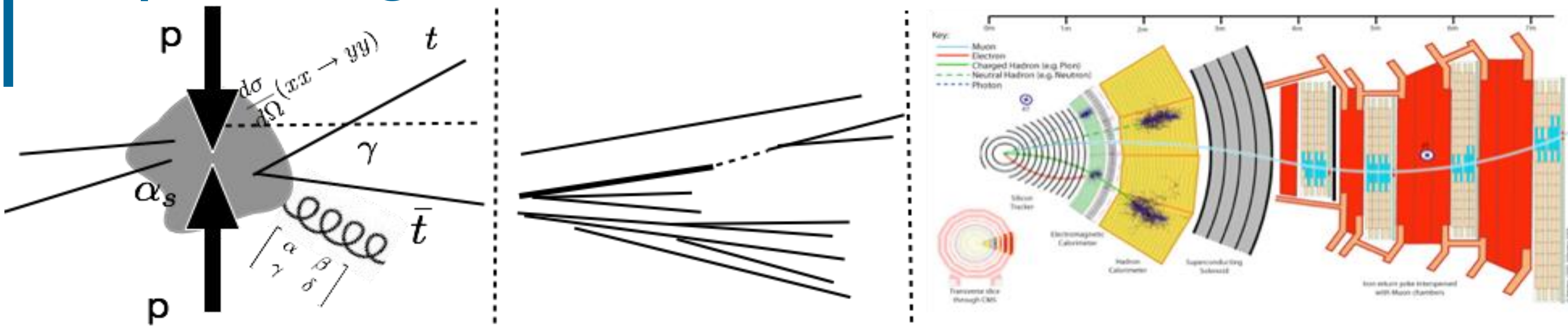
Huge numbers of complex equations

Entirely Simulated, particles are subjected to decay conditions

Detector simulation, Algorithmic reconstruction

CERN School of Computing

# Explaining the Title: An outline




model uncertainties

parton

hadronization schemes

particle

experimental uncertainties

Reconstructed

- **Theoretical**
- **Hadronization**
- **Reconstruction**

Quarks

Mathematical Objects:

complex equations

Particles

intermediate and quasi-

particles are subjected to decay conditions

Jets

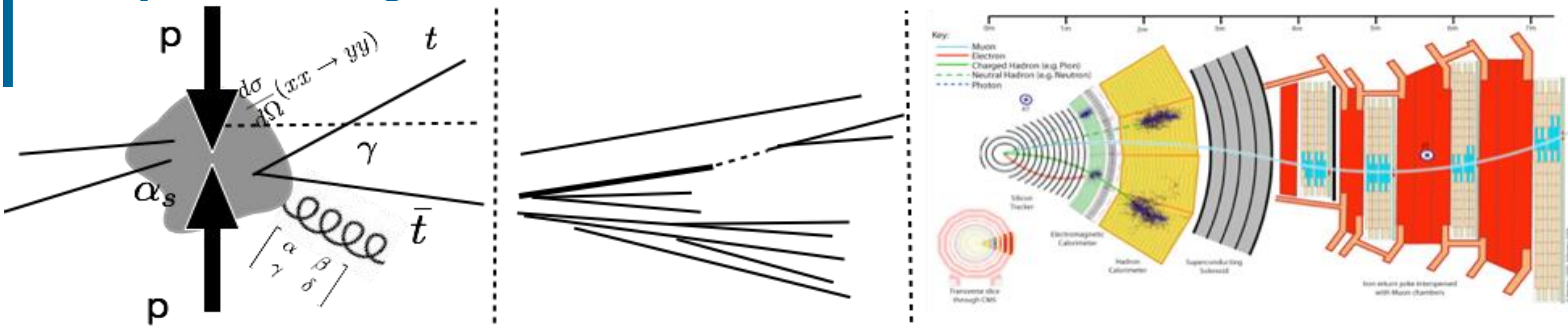
Energy deposits in detector

Algorithmic reconstruction

Different computing solutions used to tackle the unique challenges at each step

CERN School of Computing

# Explaining the Title: An outline




model uncertainties

parton

hadronization schemes

particle

experimental uncertainties

Reconstructed

- **Theoretical**
- **Hadronization**
- **Reconstruction**

| Quarks | Particles | Jets |
| --- | --- | --- |
| Mathematical Objects: | intermediate and quasi- | Energy deposits in detector |

Different computing solutions used to tackle the unique challenges at each step

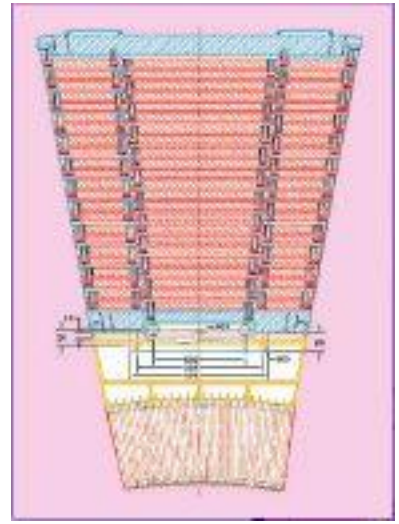
The Second lecture will explore jet clustering, classification and how they are adapted for use on personal scale computers

CERN School of Computing

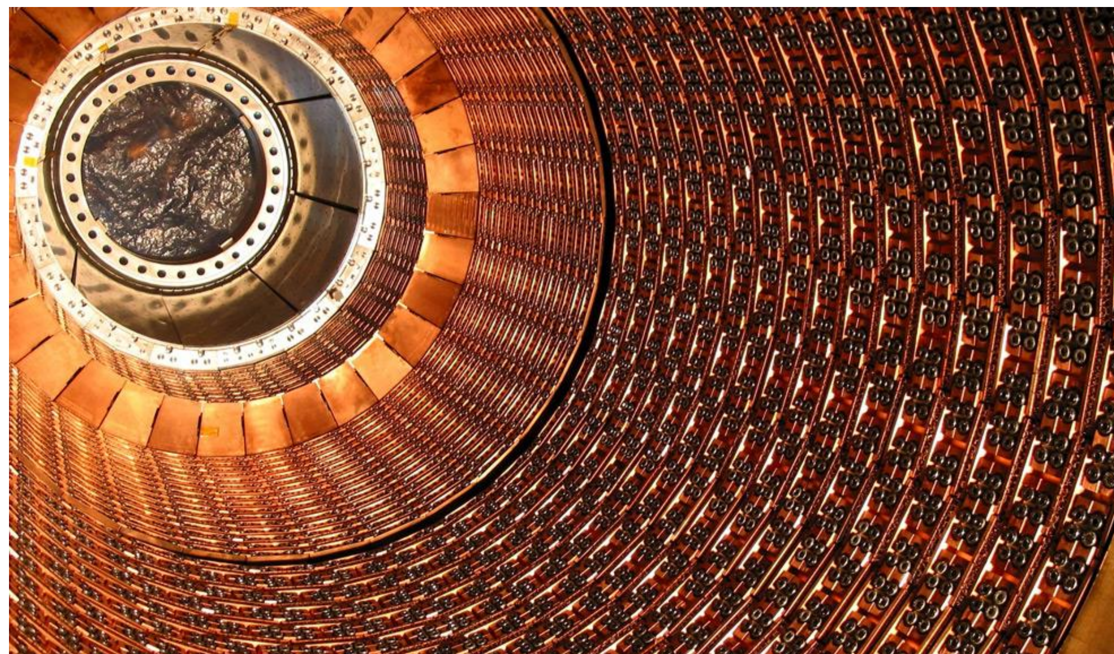
# Ask Questions here

- **Use theory to make predictions for observables of particles**
- **Design detectors to detect these observables**
- **Reconstruction algorithms to remake the objects**

# Calorimeters



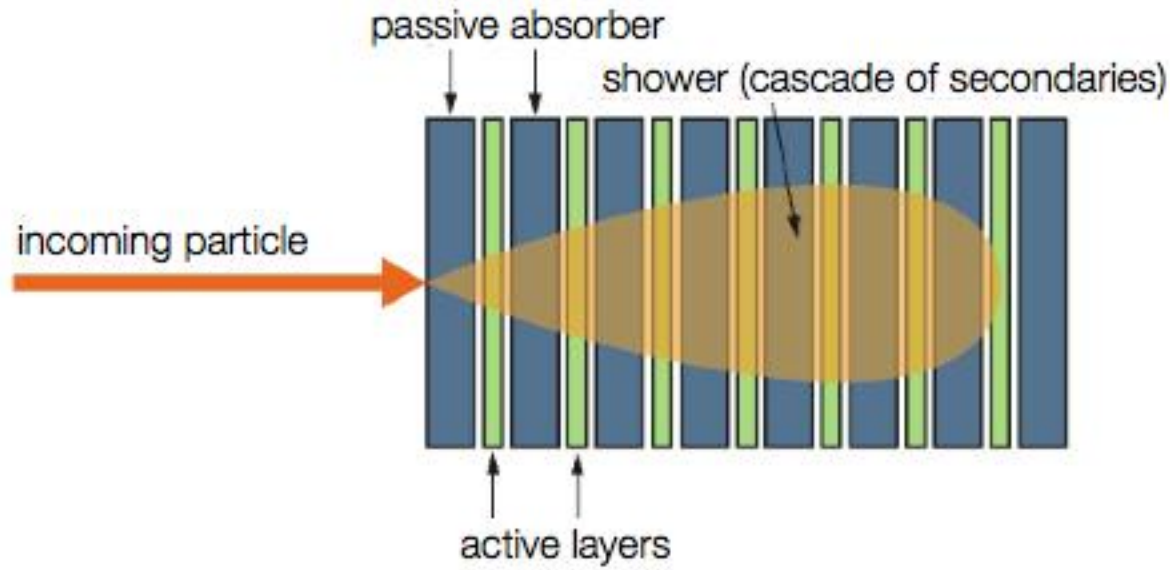
- **Calorimeters are designed to capture the energy of particles**
- **Two different types are in use at LHC**
  - Homogeneous – Capture all of the energy of the incident particle
  - Sampling – capture a portion the incident energy and make a correction

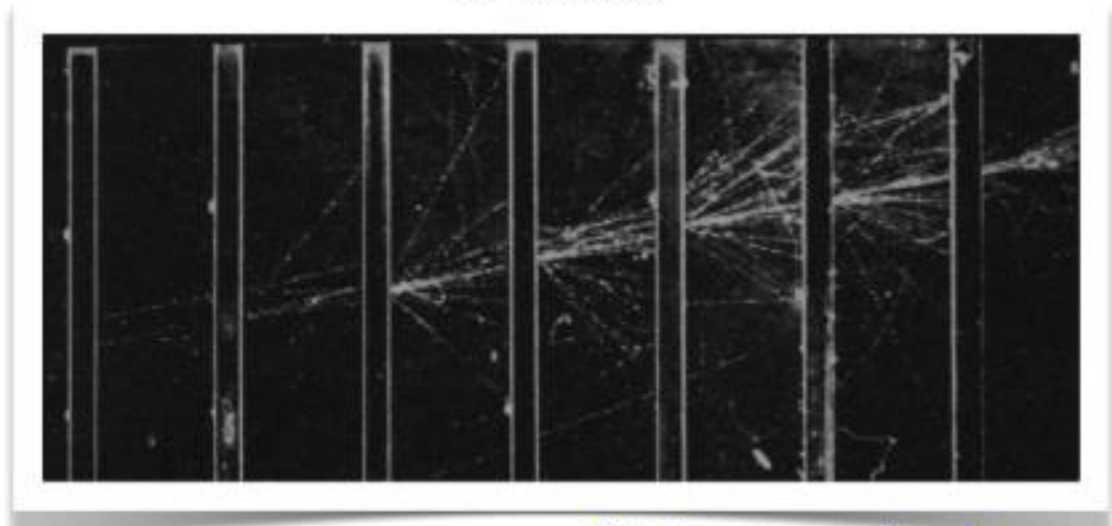


# Sampling Calorimeter

- **Sampling calorimeters have a sensitive layer sandwiched between to heavy absorber layers**
  - Absorber layers useful to create showers of secondary particles

- **Useful for hadrons because a homogeneous detector would be too large**
  - Worse resolution, though



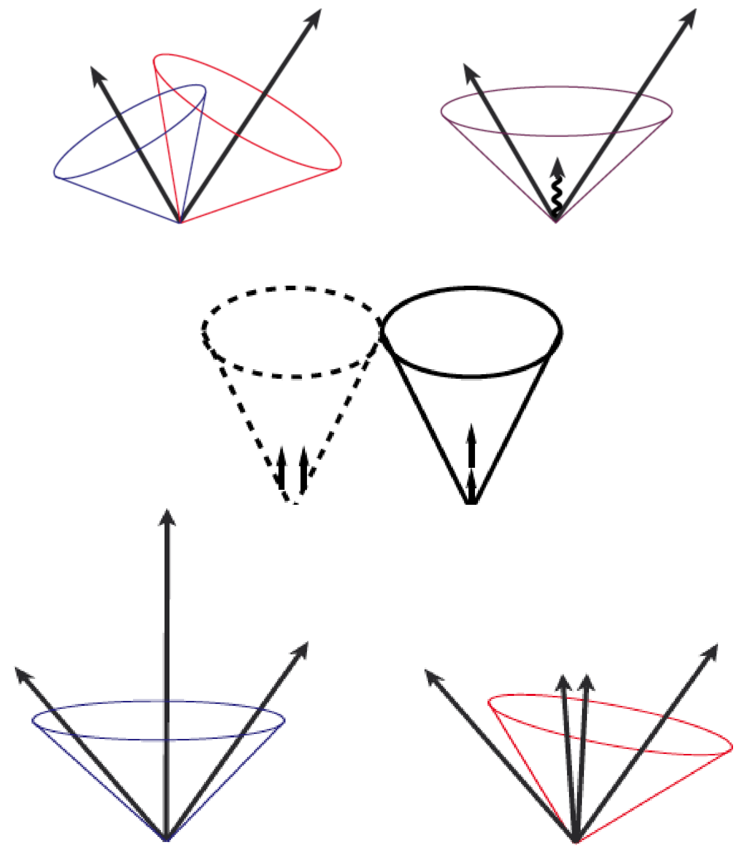
Electromagnetic shower

CERN School of Computing

# Jet clustering

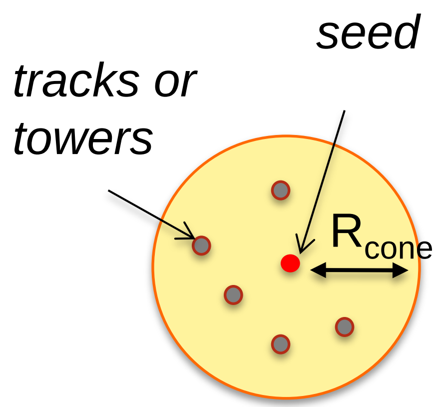
Because we are measuring decay products we must find a way to cluster them together to accurately represent the original particle

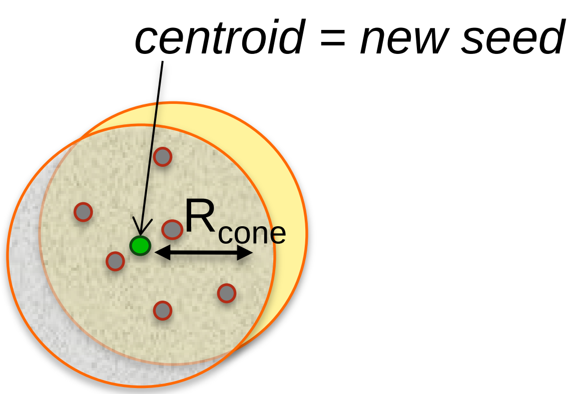
- **A few theoretical considerations**
- ***Infrared Safe***
  - Should not be sensitive to soft radiation
- ***Collinear Safe***
  - Should not be sensitive to collinear radiation



# Cone Algorithms

- **“Seed” defines approximate jet direction**
- **All energy deposits within a given radius are put into the jet**
- **The centroid is determined summing all particles within the cone**
- **The centroid becomes the new seed**
    - Iterated until stable

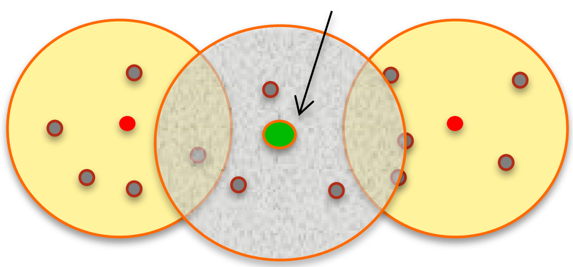


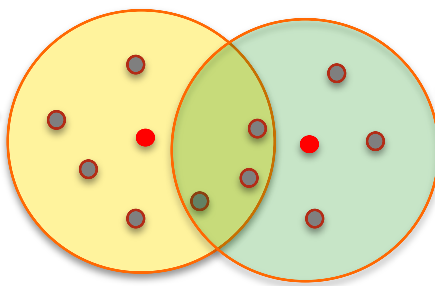
CERN School of Computing

# Mid-point cone algorithm

- **Search for missing jets using the midpoint of all jet pairs as a seed**




- **If there is a stable cone consider the energy deposits shared between the two jets ($E_{Shared}$)**




- **Take f = $E_{Shared}/E_{jet2}$**
  - If f> 50% merge the jets; else split the jets

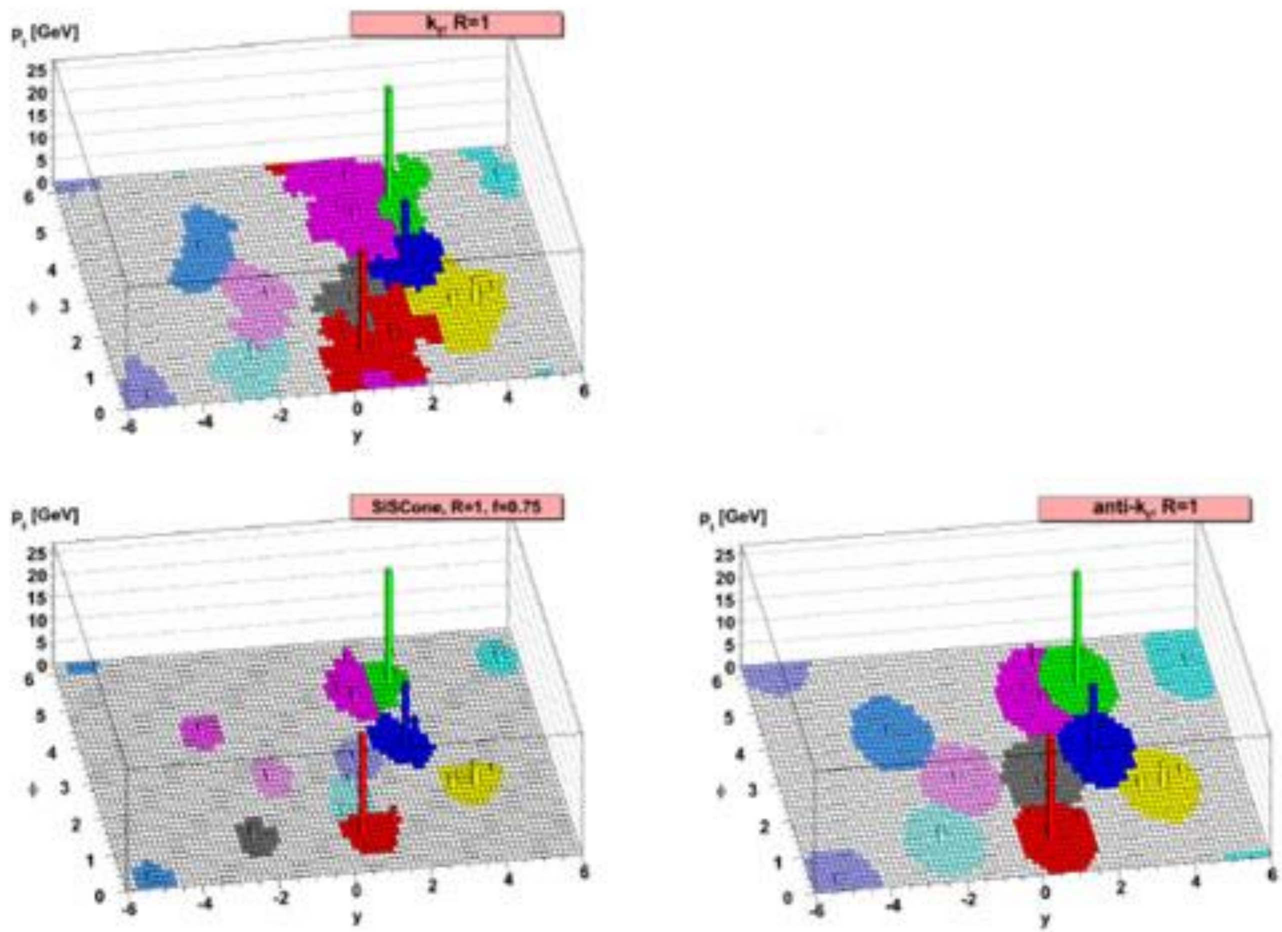
CERN School of Computing

# $K_T$ algorithm

- **Begin with a list of hits and calorimeter towers**
- **Calculate**
  - For each precluster i: $d_i = p_{T,i}^2$
  - For each pair (i,j): $d_{ij} = min(p_{T,i}^2, p_{T,j}^2)\frac{\Delta y^2 + \Delta\phi^2}{D^2}$
- **Find the minimum, $d_{min}$, of all $d_i$ and $d_{i,j}$**
- **If $d_{min}$ is a $d_{i,j}$, remove preclusters i and j from the list and replace with a new merged precluster**
- **If $d_{min}$ is a $d_i$, precluster I is not "mergeable" and can be added to the list of jets**
- **Repeat until list is exhausted**

iCSC CERN School of Computing

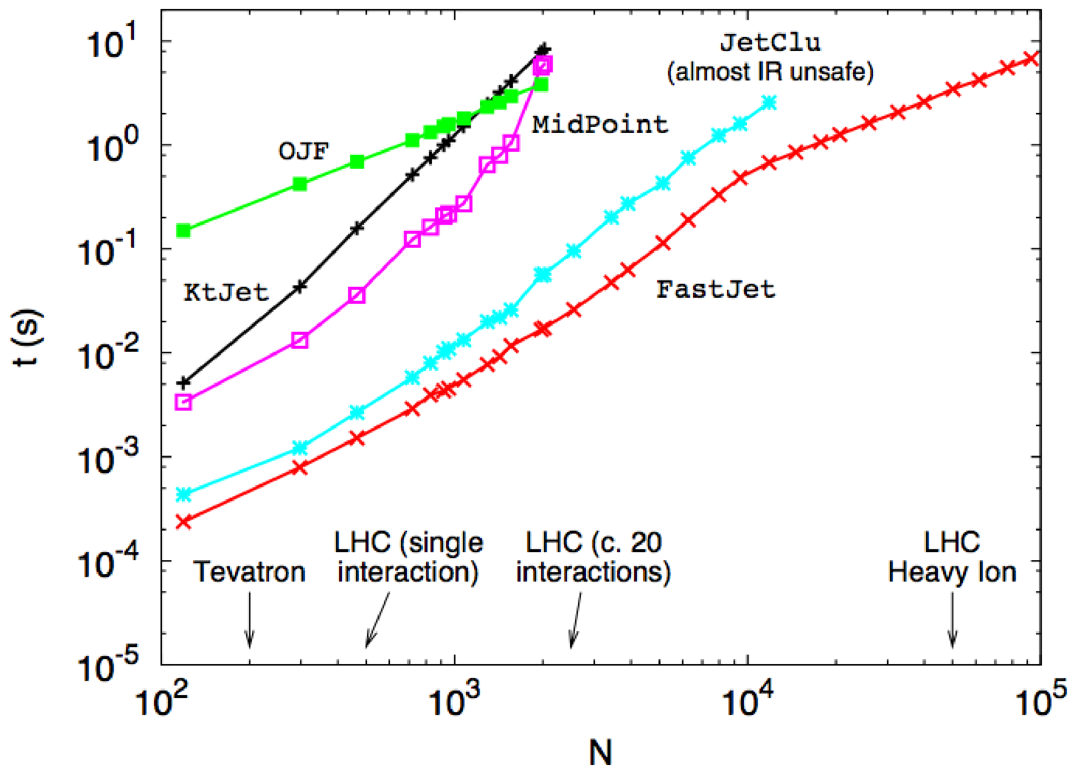
# Some different Examples



iCSC CERN School of Computing

# FastJet Software

- **FastJet (fastjet.fr) is a commonly used tool in HEP that calculates cones for many different algorithms**
- **It also contains many algorithmic improvements**
  - $K_T$ algorithm is $O(N^3)$
  - FastJet KT is O(NlnN)
  - *Identical* results
- **Used in many experiments, so it's input structure is independent**



CERN School of Computing

# Fast Jet Input

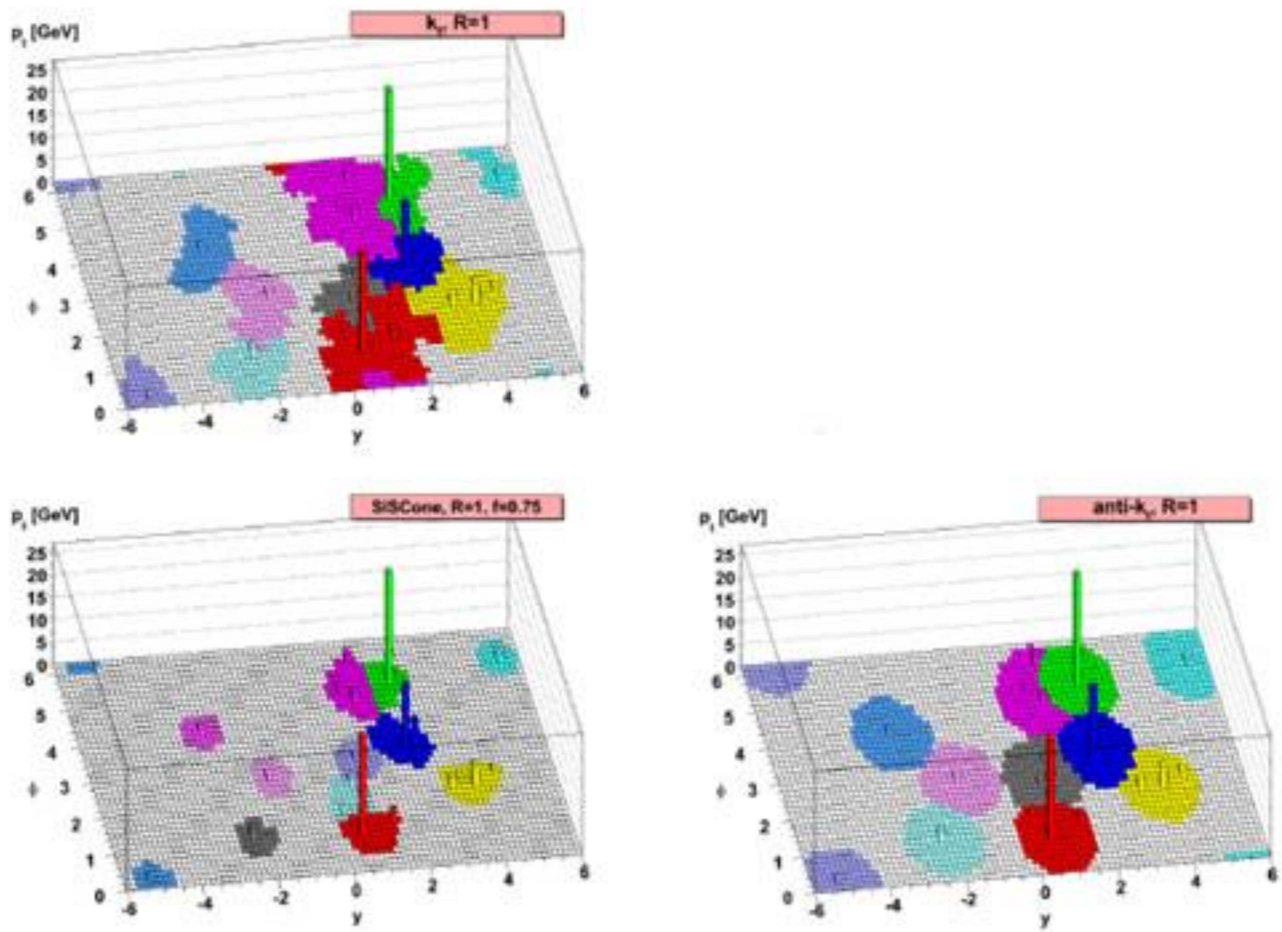
- **FastJet is only a clustering algorithm**
  - It doesn't need information about the detector, beam, etc.
- **The only thing it needs as input are momentum 4-vectors**

```
// an event with three particles:   px     py   pz       E
particles.push_back( PseudoJet(   99.0,  0.1,  0, 100.0) );
particles.push_back( PseudoJet(    4.0, -0.1,  0,   5.0) );
particles.push_back( PseudoJet(  -99.0,    0,  0,  99.0) );
```

- **It only outputs 4-vectors** (improtant later)

```
Clustering with Longitudinally invariant anti-kt algorithm with R = 0.7
and E scheme recombination
        pt y phi
jet 0: 103 0 0
    constituent 0's pt: 99.0001
    constituent 1's pt: 4.00125
jet 1: 99 0 3.14159
    constituent 0's pt: 99
```

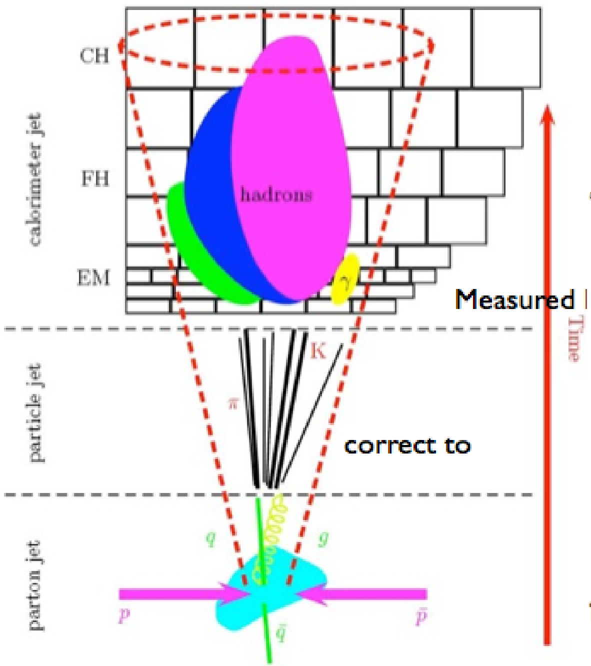
# Some different Examples



CERN School of Computing

# Corrections

However all of these objects need corrections for various reasons. Jets need to be corrected for detector and algorithm inefficiencies

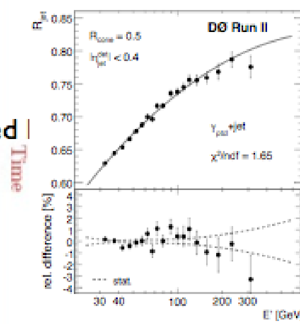



Applied to all jets

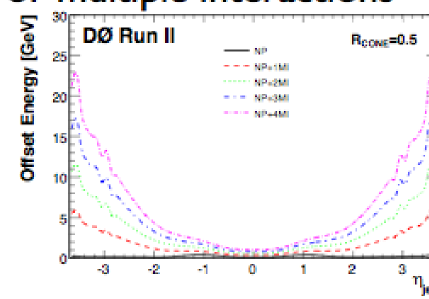
We want to correct to the particle level (before any detector interaction) using this algorithm:

$$E_{jet}^{ptcl} = \frac{E_{jet}^{raw} - O}{F_\eta \times R \times S}$$

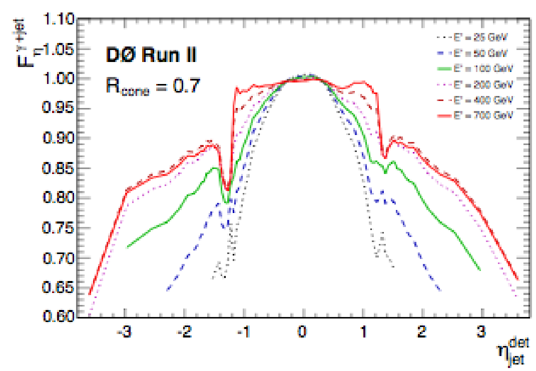
Response: This term estimates the energy deposited in nonsensitive regions or inefficiencies in the CC

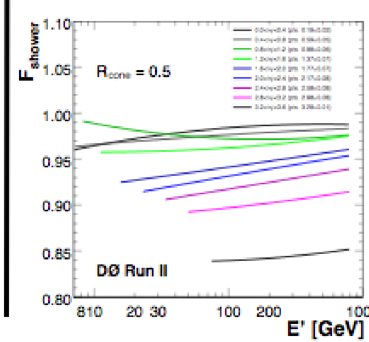



Offset: Energy deposits from noise, electronics or multiple interactions




Then corrects the response of forward regions to that of the CC

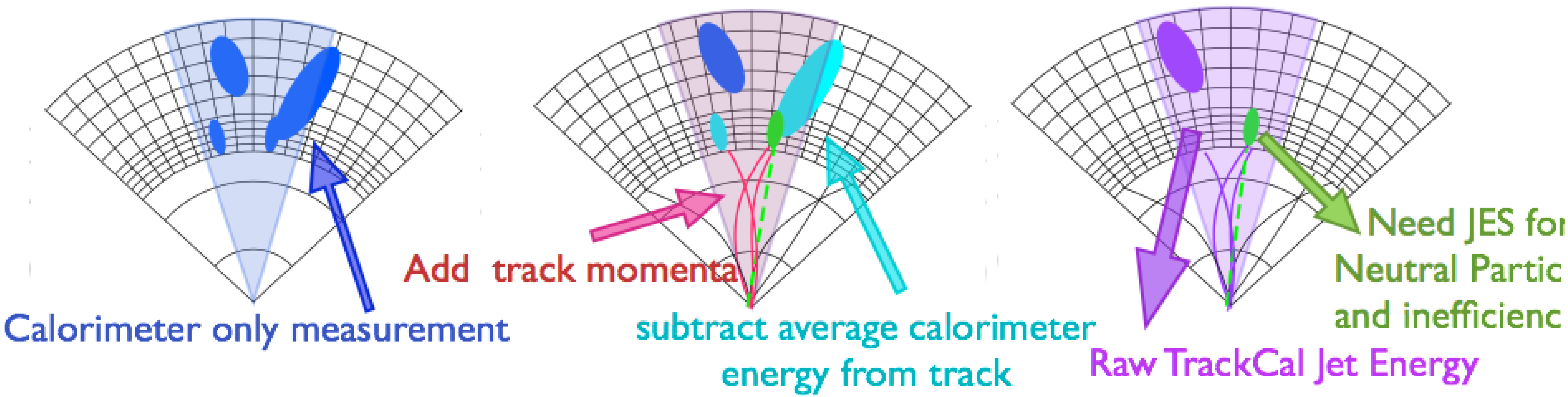



Showering: Energy not accounted for by the jet cone algorithm

CERN School of Computing

# Combining Measurements




- **It can be advantageous to use different parts of the detectors for different measurements for the constituents of the jets**
- **For example, we can sometime replace a calorimeter measurement with a tracker measurement associated to it**
  - For low momenta, the tracker measurements are more precise

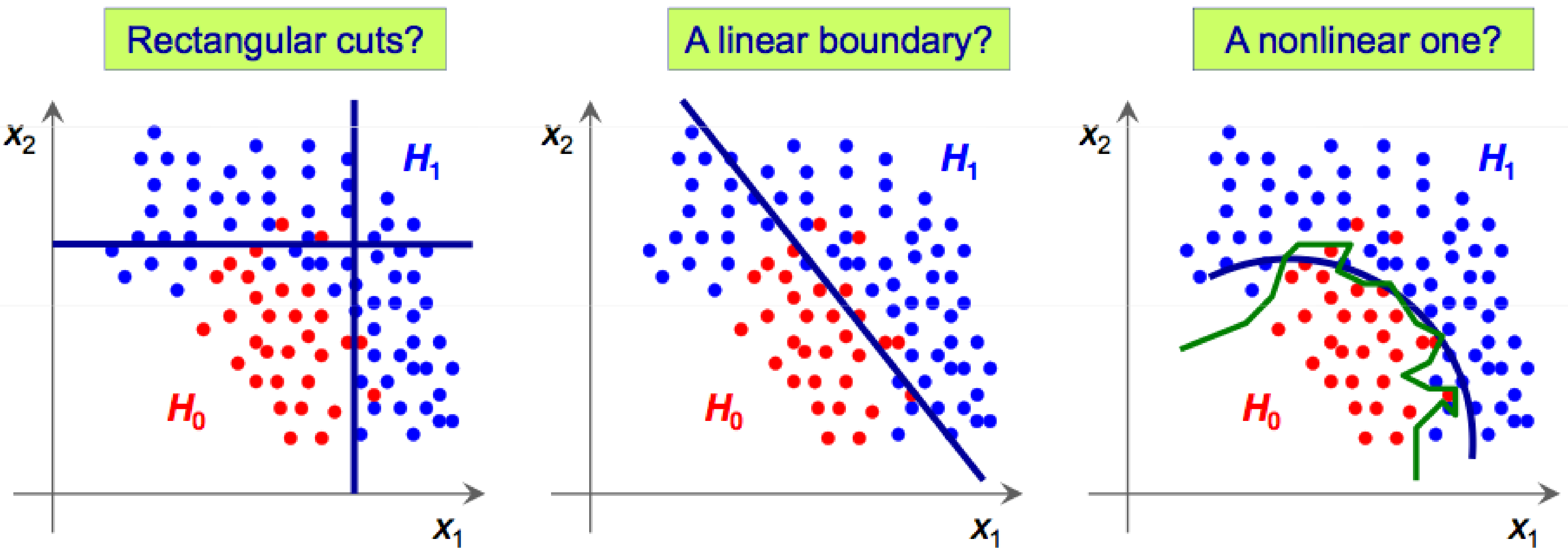
# Jet Questions

- **Calorimeters are designed to collect all particle energy**
- **Particles are scattered and not individually measured**
- **Reconstruction attempt to safely reconstruct the original energy**

CERN School of Computing

# Classifcation Techniques

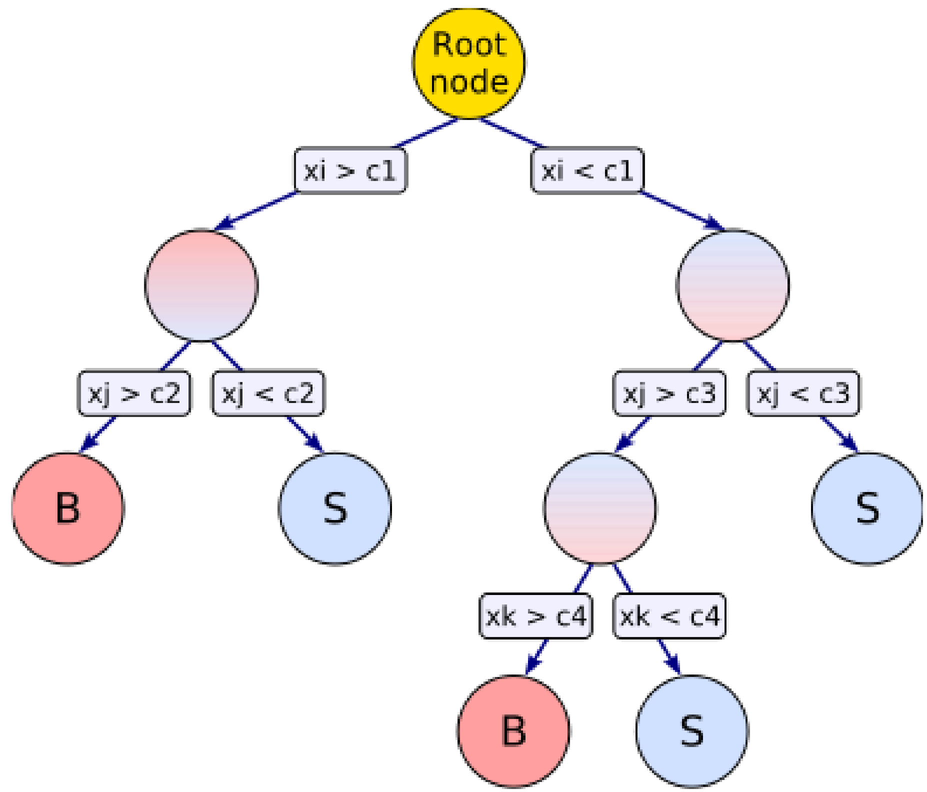
- **Found two variables**
- **What is the best cut**



CERN School of Computing

# Multivariate Techniques

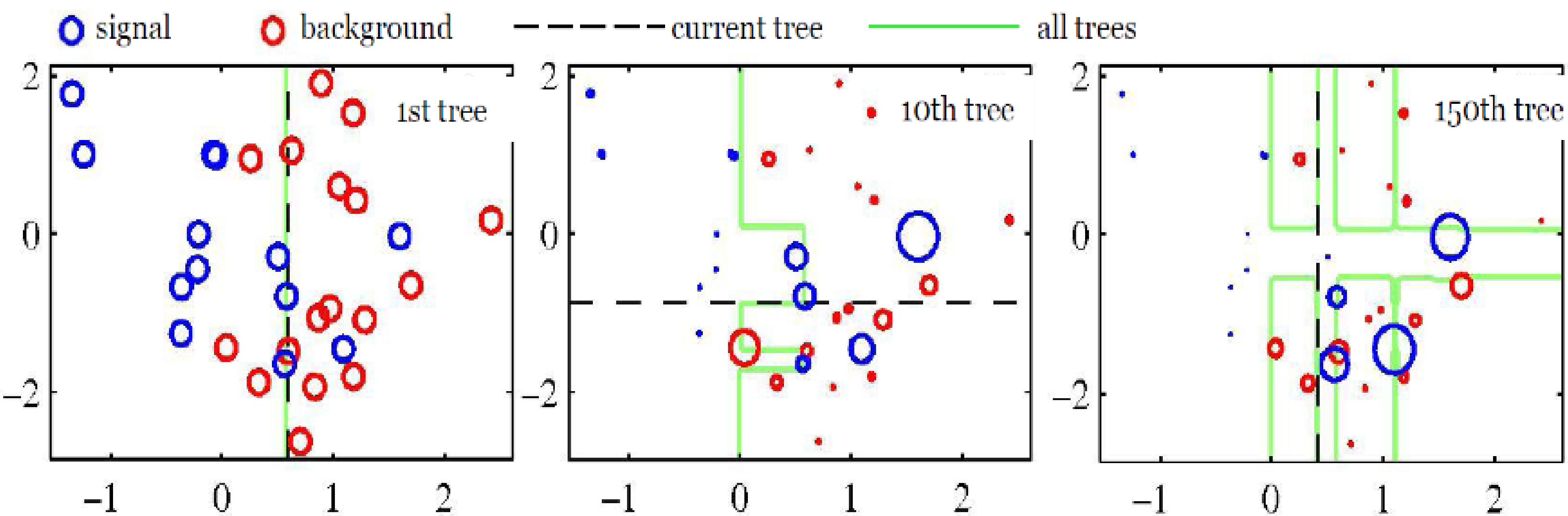
- **Finally we can combine all these measurements in to very power multivariate analysis (MVA) techniques**
  - These can give a measure of how likely a jet is to be a b-jet
- **One technique is a decision tree that makes a series of cuts on different input variables**
  - Then reclassified by the Gini index p(1-p)



CERN School of Computing

# Decision trees - weighting

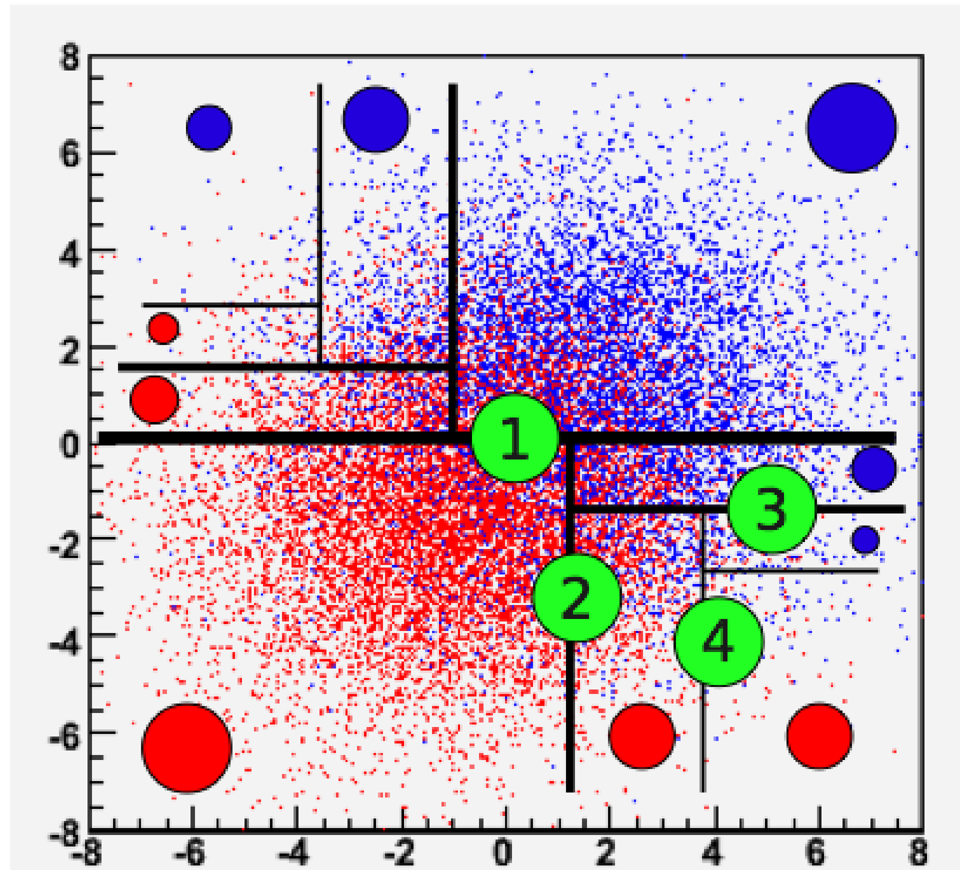
- **The tree is trained against a known truth (from MC)**
- **Misclassified events are given a larger weight then retrained**



CERN School of Computing

# Sample Result

- **In the end we get a better discriminator than any simple cut**



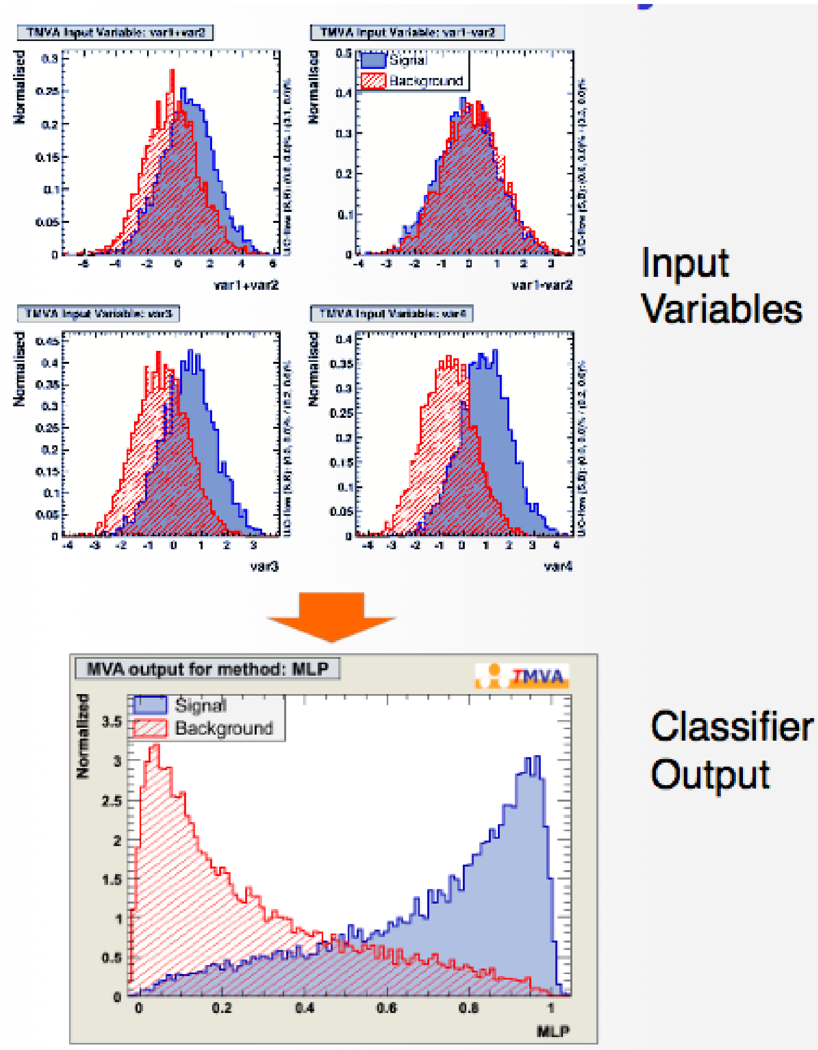
CERN School of Computing

# TMVA Software

- The Toolkit for Multivariate Analysis (TMVA [http://tmva.sourceforge.net/]) is the standard software for implementation of classification/regression techniques in HEP
- It contains algorithms for many Classification/Regression techniques
- Standard Tool for BDTs etc in HEP

# TMVA Inputs

- **TMVA again is experiment independent**
  - It doesn't want or need details about your detector, beam, etc.
- **All it takes for input are histograms** (important for latter)
  - Or properly formatted vectors
- **It returns a function such that all it will need are the input variables of the event**



CERN School of Computing

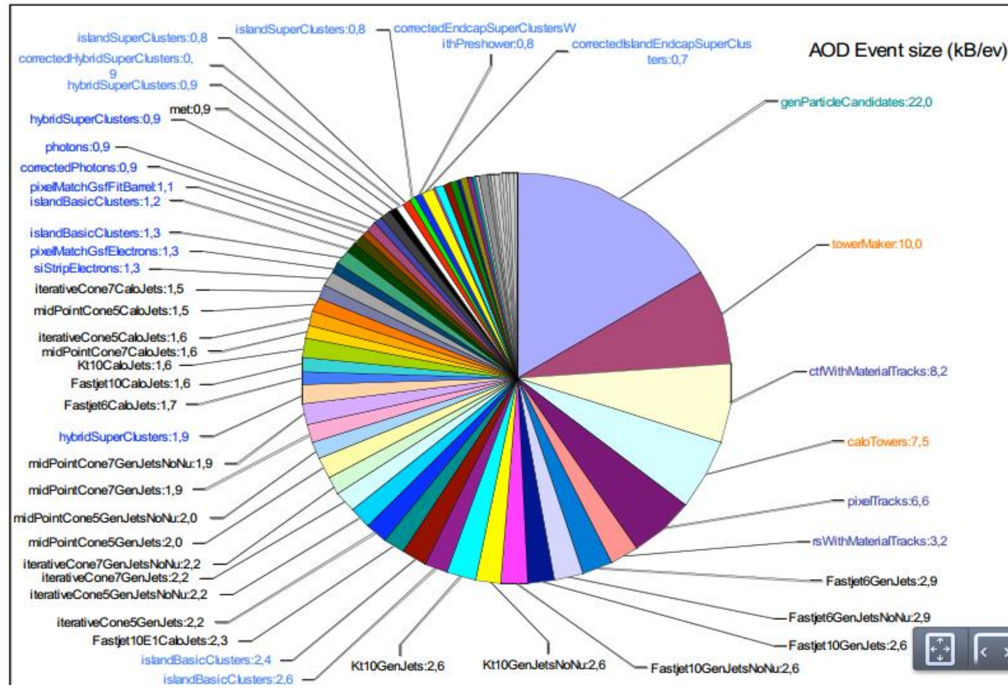
# Classification Questions Questions

- **Calorimeters are designed to collect all particle energy**
- **Particles are scattered and not individually measured**
- **Reconstruction attempt to safely reconstruct the original energy**

CERN School of Computing

# Laptops – A Powerful Tool in HEP

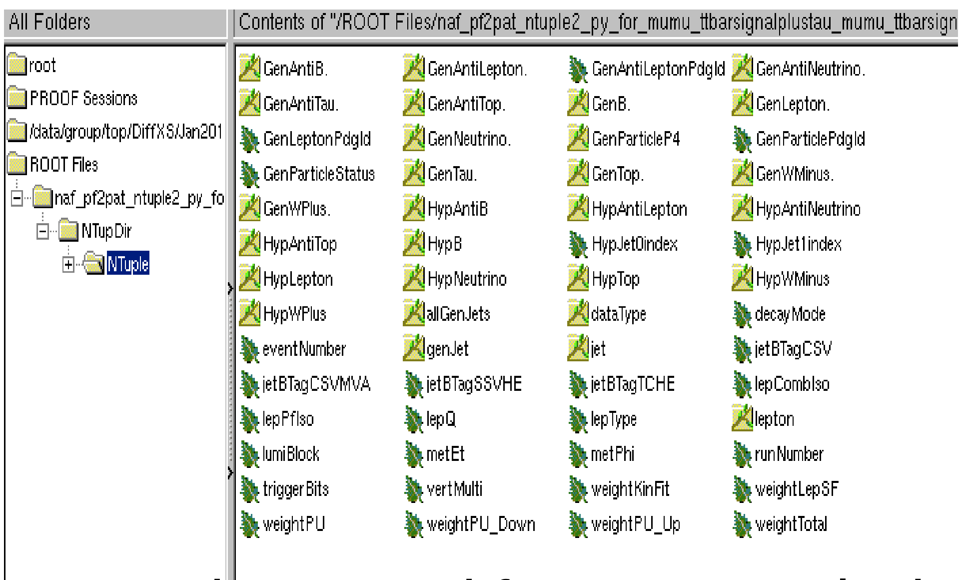
- **LHC experiments store 10s of petabytes of information**
  - But not all of it is useful for every part of every analysis

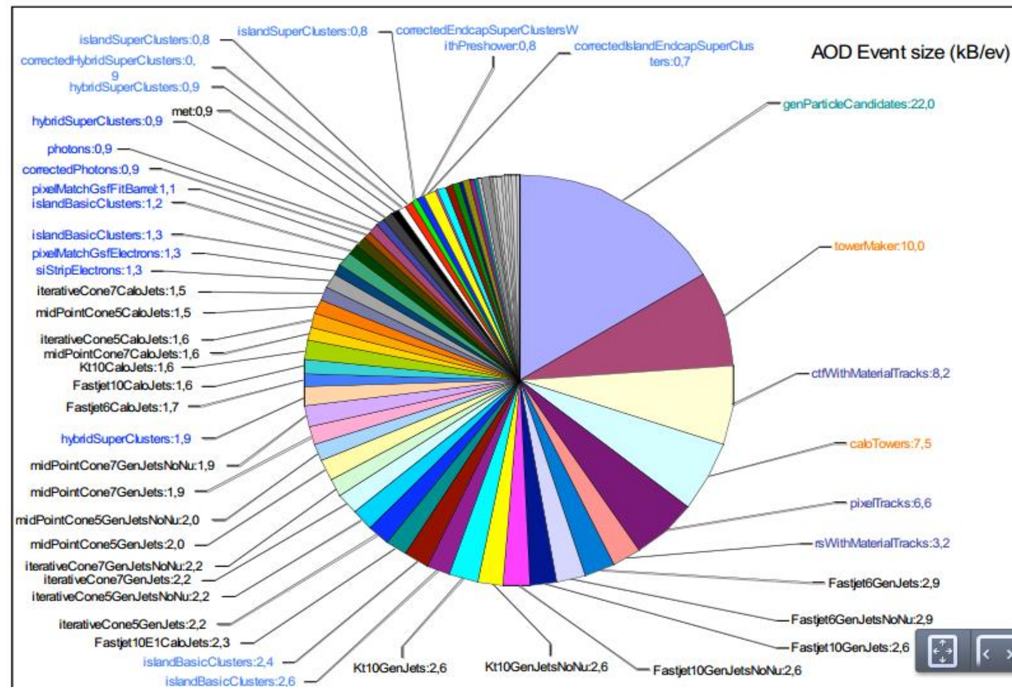


CERN School of Computing

# Laptops – A Powerful Tool in HEP

- **LHC experiments store 10s of petabytes of information**
  - But not all of it is useful for every part of every analysis




Subset used for a top analysis

# Event Flow

- **What you could do**

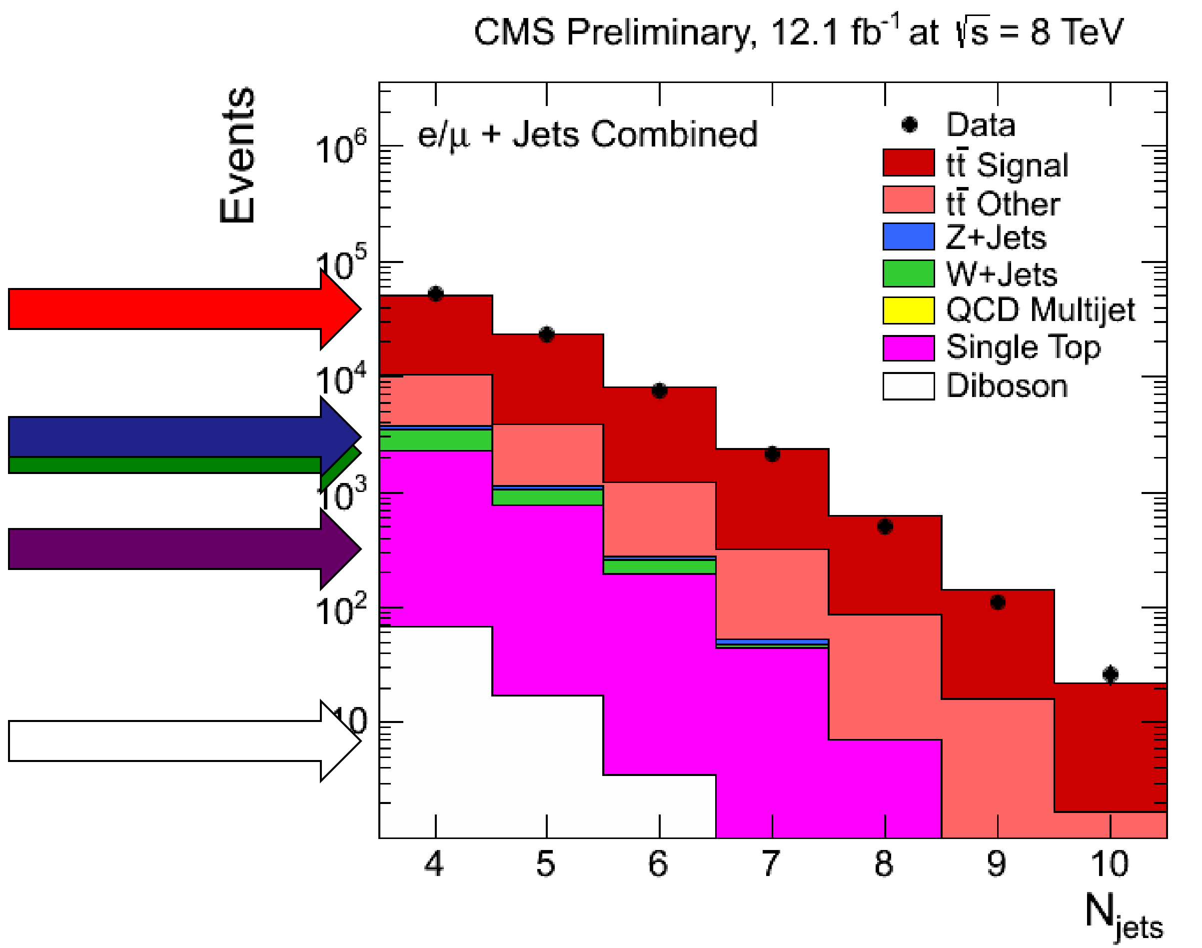
| | |
|---|---|
| Full event reconstruction, TMVA, Jet clustering, Mathematical operations, Corrections, studies | Plotting and Final Results |

- **Or we could introduce intermediate steps**

| | | | |
|---|---|---|---|
| Event Reconstruction | Object selection<br>Reduce size of trees | TMVA, or reclustering, or studies | Final Results |

# MC Samples

- **Many different samples for many different backgrounds**
- **Sometimes over 20 samples per analysis**
- **Each can be treated separately**



CERN School of Computing

# Mass Parallelization

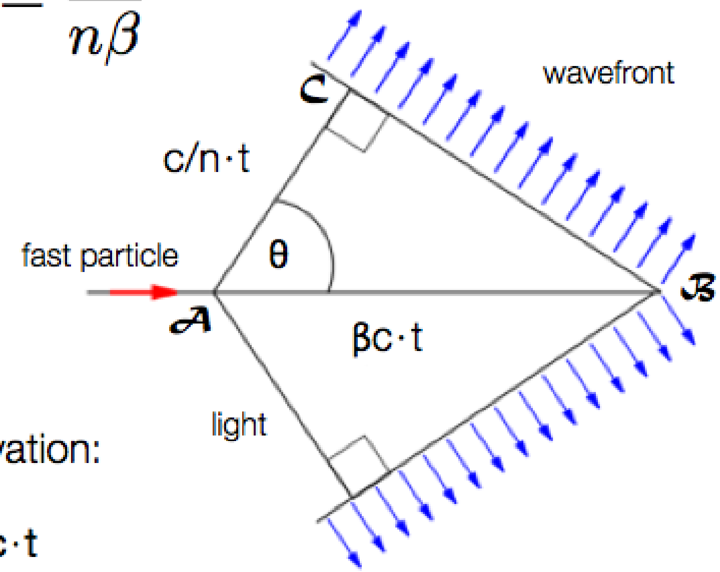
| | | | |
|---|---|---|---|
| Event Reconstruction | Object selection<br>Reduce size of trees | TMVA, or reclustering, or studies | Final Results |
| Event Reconstruction | Object selection<br>Reduce size of trees | TMVA, or reclustering, or studies | Final Results |
| Event Reconstruction | Object selection<br>Reduce size of trees | TMVA, or reclustering, or studies | Final Results |
| Event Reconstruction | Object selection<br>Reduce size of trees | TMVA, or reclustering, or studies | Final Results |
| Event Reconstruction | Object selection<br>Reduce size of trees | TMVA, or reclustering, or studies | Final Results |

# Particle ID - RICH

- **Cherenkov raditation is emitted when a particle passes through a medium and is initially going faster than the speed of light in that medium**
- **A ring of light is emitted that is proportional to the momentum**

Cherenkov angle:

$$\cos\theta_c = \frac{1}{n\beta}$$




Simple
Geometric derivation:

$AB = \beta c \cdot t$

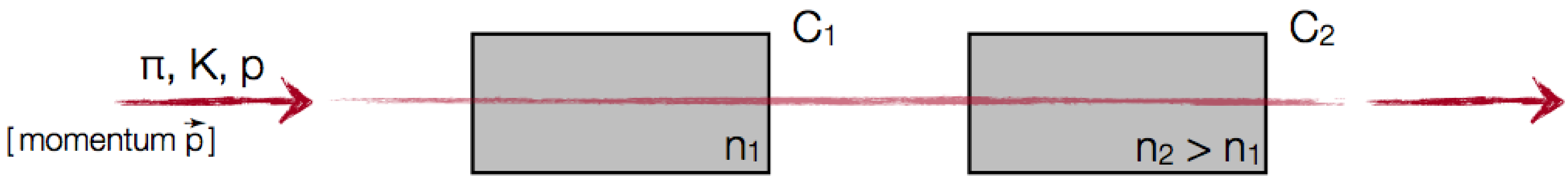
$AC = c/n \cdot t$

$\cos\theta = AC / AB = c/n \cdot t / (\beta c \cdot t)$

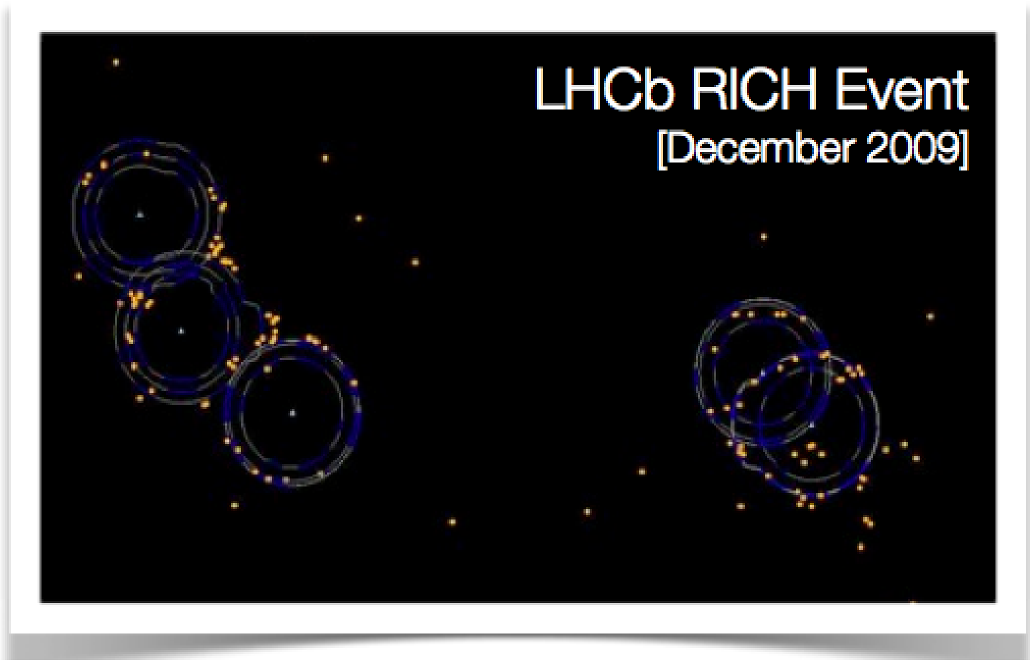
$= 1/n\beta$

iCSC CERN School of Computing

# Particle ID - RICH

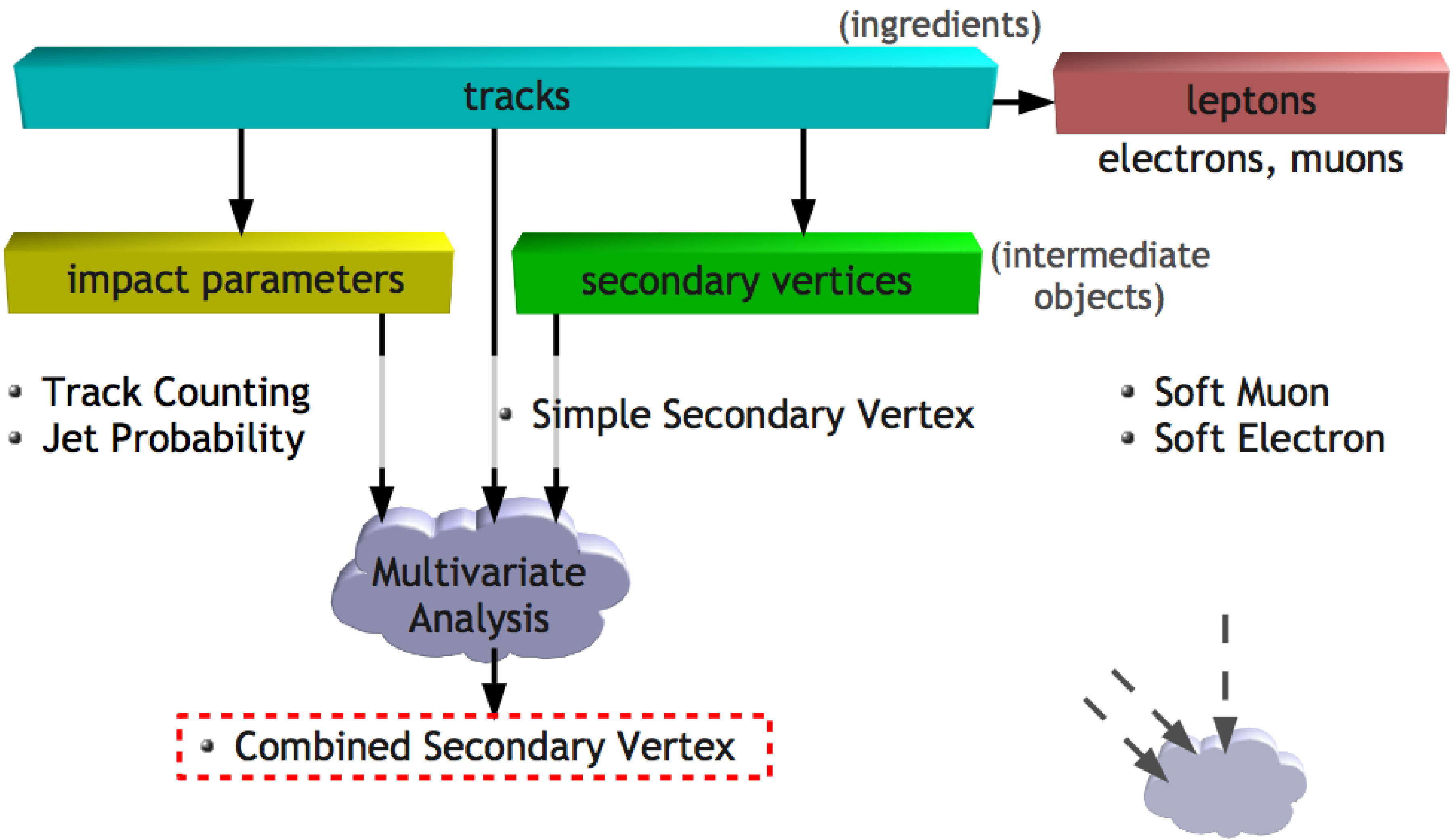



- **By choosing the correct media, we can use this as a for of particle identification**
  - Light in $C_1$ and $C_2$ = Pion
  - Light in just C1 = Kaon
  - No light = proton



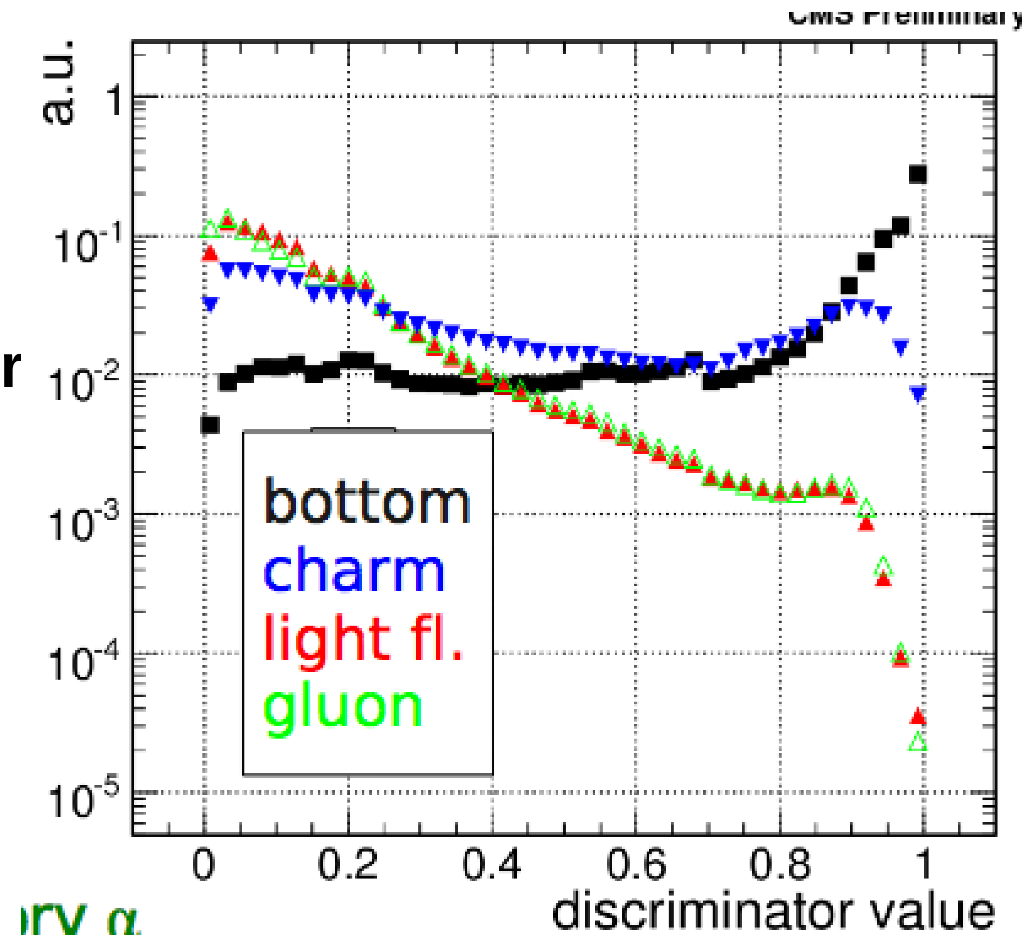
CERN School of Computing

# B-tagging Inputs

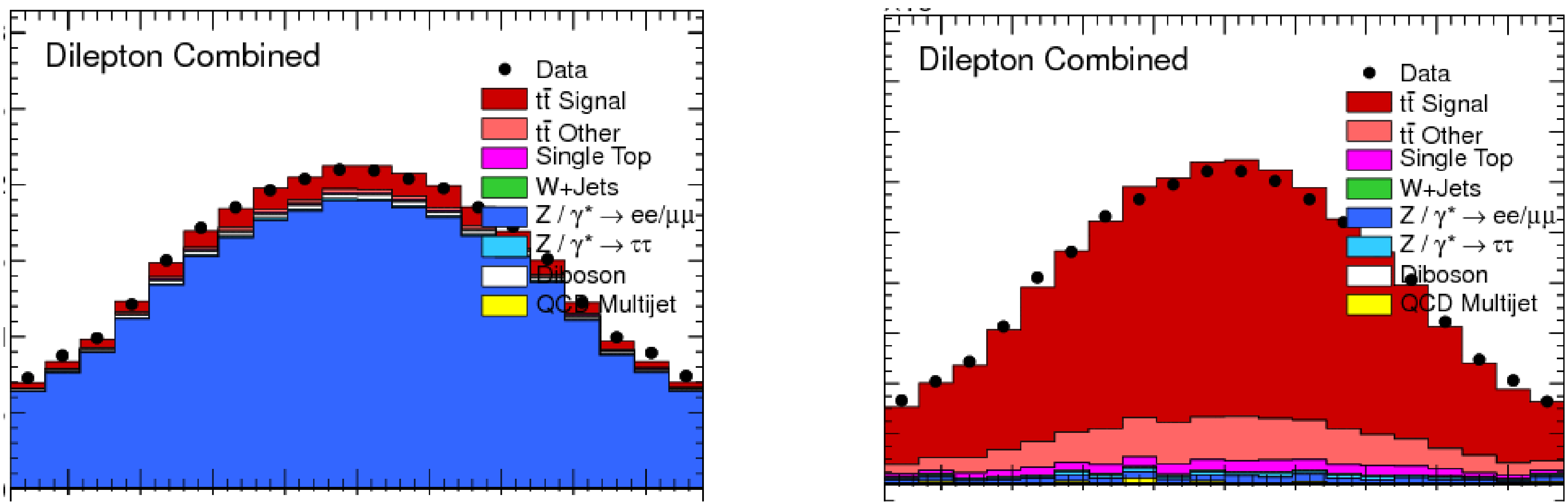


# Output

- **Using the inputs from multiple parts of the detector we can make a better judgement on if the jet we are measuring came from a b-quark or another source**



# Physics Analysis




- **B-tagging can reduce some background by over 90%**
- **As well as increase the signal-to-noise ratio!**

CERN School of Computing

# Conclusions

- **Jet algorithms are designed to make up for the inefficiencies created by finite resolution of our calorimeters**
- **These algorithms are well suited for personal computing**
- **Aspects from many different portions of the detector can be combined using statistical tools such as decision trees to determine how likely it is that a particular jet came from a b-quark**
- **Thanks for listening!**

# Backup

CERN School of Computing

# Data Storage - NTuples