Clustering algorithms for b-jets from BSM Higgs Bosons

Billy Ford

University of Southampton

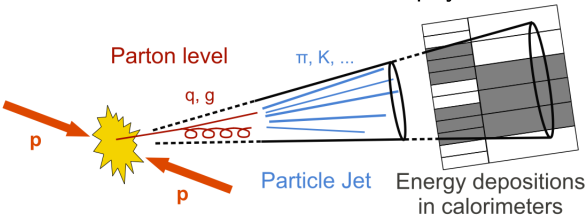
IPPP YTF 11, 2018

Introduction

- Why are jets useful to study?
- How do we classify jets?
- What's so special about b-jets?
- How can this help finding new physics?
- What can we do better?

Jets at the LHC

- Why are jets important? We don't have the luxury of seeing individual hard scatterings at the LHC!
- Improving methods of jet classification allows us to map the underlying particle theory to experimentally detectable phenomena
- Even better, we can look for new physics!



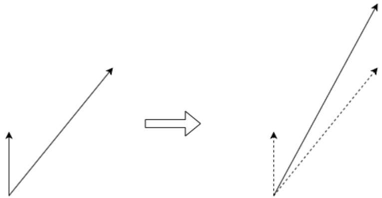
Jet Clustering Algorithms

- We form jets from particle tracks using so called **Jet Clustering Algorithms**
- Sequentially combine final state hadrons together into jets, based on seperation/energy
- There are many different algorithms suck as anti-k$_T$ ($n = -2$) and the Cambridge-Aachen ($n = 0$)
- Begs the question - which is best for a particular search? (We'll come back to this!)

$$d_{ij} = min(k_{Ti}^n, k_{Tj}^n)\frac{\Delta_{ij}^2}{R^2} \ ; \ d_{iB} = k_{Ti}^n$$

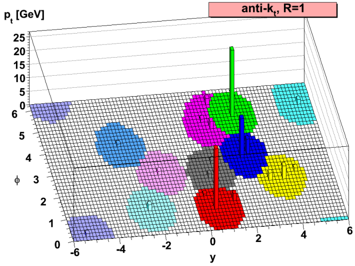
Jet Clustering Algorithms (cont.)

- So we have $d_{ij} = min(k_{Ti}^n, k_{Tj}^n)\frac{\Delta_{ij}^2}{R^2}$ and $d_{iB} = k_{Ti}^n$
  - $\rightarrow$ d_{ij} is the seperation between two particles i and j
  - $\rightarrow$ d_{iB} is the 'beam distance', R the (user input) cone size
- Where $\Delta_{ij}^2 = (y_i - y_j)^2 - (\phi_i - \phi_j)^2$ is the angular separation
  - $\rightarrow$ y_i is the rapidity
  - $\rightarrow$ ϕ_i is the azimuth
- Algorithm will compute all d_{ij} and d_{iB}'s
  - $\rightarrow$ If smallest value is one of the d_{ij}, combine i and j and repeat
  - $\rightarrow$ If smallest is d_{iB}, declare i a jet and remove it



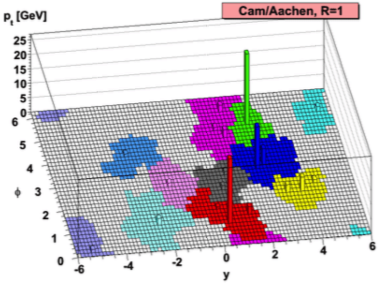
The anti-k_T

- How do the different algorithms work?
- The anti-k_T algorithm has the form $d_{ij} = min(\frac{1}{k_{Ti}^2}, \frac{1}{k_{Tj}^2})\frac{\Delta_{ij}^2}{R^2}$
- Will favour clustering hard particles due to $\frac{1}{k_T^2}$ term
- Tends to form circular shaped jets



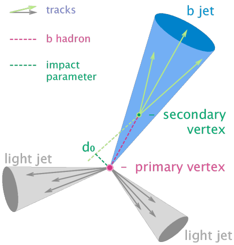
The Cambridge-Aachen

- Has the form $d_{ij} = \frac{\Delta_{ij}^2}{R^2}$
- No momentum dependence, clusters purely on geometric position of particles
- Good at resolving jet substructure



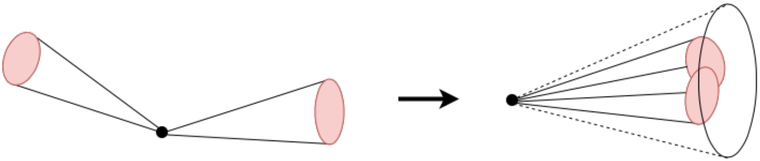
b-jets at the LHC

- Mass gap between b and light quarks gives b decay products high p_T so b-jets are wider in size
- b-mesons travel some distance before decaying, can detect lighters mesons with **displaced vertices** as coming from b's
- A Higgs Boson is expected to decay to a $b\bar{b}$ pair most of the time (we finally saw this in the summer)!



Fat Jets

- Basic conservation of momentum - rest particle would produce back to back jets
- In boosted configurations, jet pairs become more and more collinear
- Is possible for these jets to overlap - **fat jets**!
- This behaviour is prevalent in certain signals!

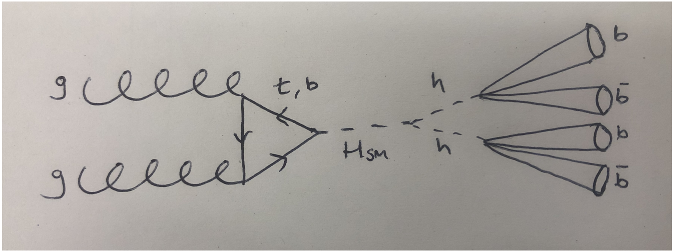


Using Jets

- We've seen how jets arise and their importance at the LHC
- How exactly do we use them to study new physics?
- What physics can we look for?
- What problems arise in searching for BSM signals?

The Hunt for New Physics!

- We consider the **2 Higgs Doublet Model**, which contain additional Higgs particles
- In particular, an SM-like Higgs can split into two lighter Higgs', which each in turn produce two b-jets, so we have a $bb\bar{b}\bar{b}$ 4-jet final state
- With the right masses, this configuration can be susceptible to boosted merging



Enter Monte Carlo!

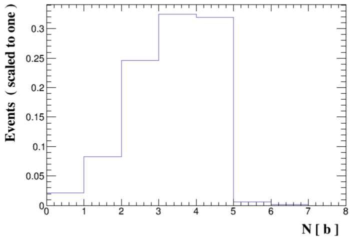
- Use Madgraph5 (interfaced with Pythia8 and Fastjet) to generate samples of Monte Carlo data for this BSM process
- Give us insight into the jet multiplicity distribution, as well as kinematics
  - $\rightarrow$ Do the different algorithms yield different multiplicity distributions?

Enter Monte Carlo! (cont.)

- Can also investigate the effect of cone size to find a suitable value
  - $\rightarrow$ Too small and we can over cluster larger jets into separate objects
  - $\rightarrow$ Too large and we risk sweeping up multiple jets into single, merged objects
- Optimizing the cone size and best algorithm gives better resolution of the of jet distribution and kinematics

Why is this difficult?

- There are however difficulties with generating these signal events
- To see the full signal isn't always possible - frequently lose events to detector cuts on p_T and η
- We also get pick up additional b-jets - 4b final state somehow yielding > 4 jets!
  - $\rightarrow$ Are these large jets incorrectly clustered into 2 (or more)?
  - $\rightarrow$ Could be coming from ISR $g \rightarrow b\bar{b}$ jets?

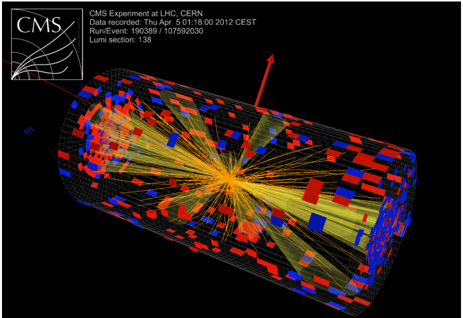


Okay, so why is this Useful?

- In solving this we have assembled toolkit to generate interesting data samples of **BSM** physics
- We can even go further and fix jet multiplicity to investigate kinematics of boosted events
- Generality - Many models have extended Higgs sectors, simple to extend results to different models

The Need for Improvement

- This type of channel is difficult to explore, overwhelming b-jet background!
- How well can we distinguish merged jets from individual, wider jets?
- Is a particular clustering algorithm better suited for this signal?
- We will compare performance of anti-k_T and Cambridge-Aachen

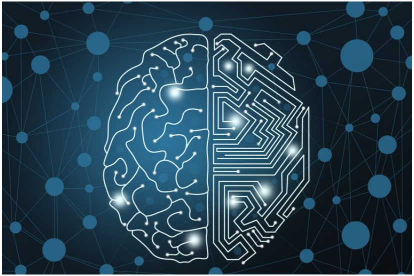


Going Even Further

- Can we do even better than this?
- Clustering algorithms still aren't perfect, merged jets problematic
- Underlying physics tells us there are still always 4 jets, so we should be able to see them!
- We have a set of event data at the parton level, hadron level and detector level...

Going Even Even Further

- ...What if we feed this into some **Artificial Intelligence** agent!
- Use our stream of Monte Carlo data to train some AI algorithm, see what it can learn
- Let loose on experimental data now LHC run II has come to an end
- Ambitious project, but very exciting applications for Particle Phenomenology!



Thanks for listening!