Jet clustering algorithms for b-jets from BSM Higgs Bosons

Billy Ford

University of Southampton

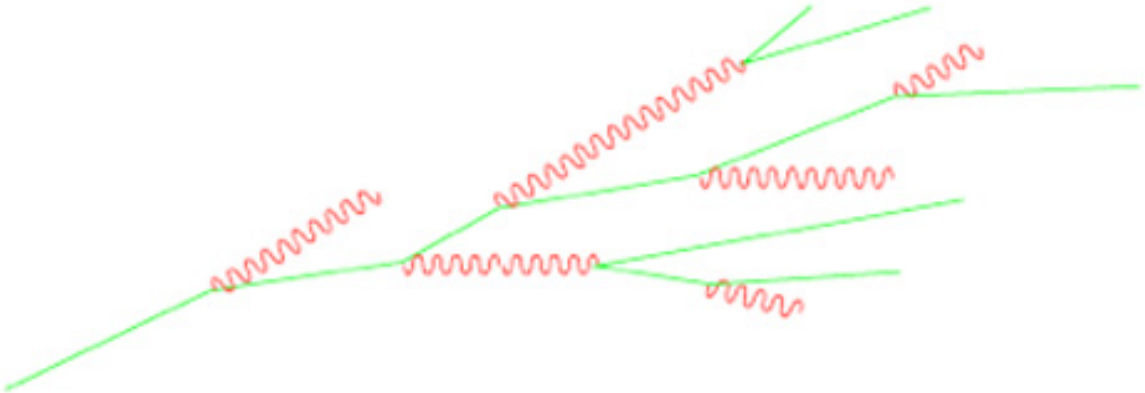
NExT Seminar Meeting, November 2020

Introduction

- Ultimate motivation is in search of new physics **B**eyond the **S**tandard **M**odel
- Shortcomings of SM are well established and understood
- We question whether experiments can increase sensitivity to certain processes of BSM physics
- Begin with review of some key concepts surrounding study

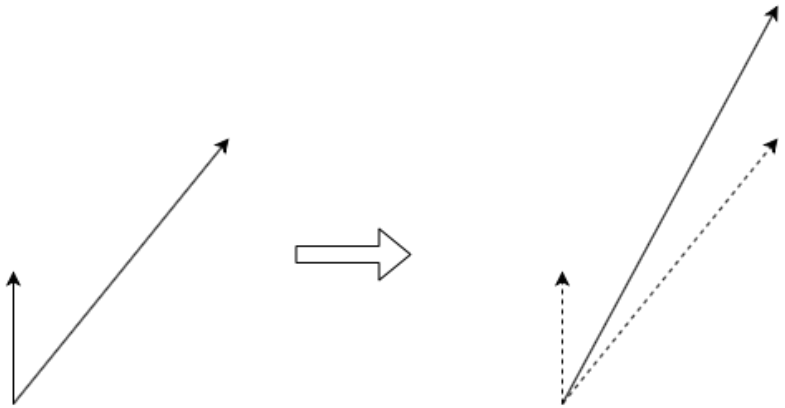
Introduction - Jet Clustering Algorithms

- How exactly is a jet defined? - We require a **jet definition**, i.e. a specific map between the hadronic sprays in a detector back into a distinguishable object
- Rich history associated with the development of jet clustering algorithms
- Exist two main types - **cone algorithms** and **sequential recombination algorithms**
- We focus solely on sequential recombination algorithms in our study...



Introduction - Jet Clustering Algorithms (cont.)

- Sequential recombination work by essentially rewinding back the hadronisation/showering process
- Combine particles four-momenta together based on separation/p_T into **pseudojets**, and then continue combining particles/pseudojets together
- When some cut off is reached, a pseudojet can be declared a jet and removed from the sample
- The process is repeated until no pseudojets remain

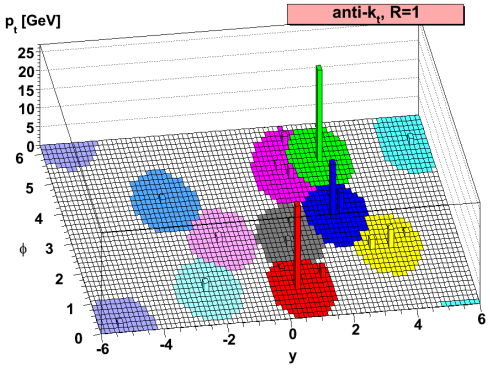


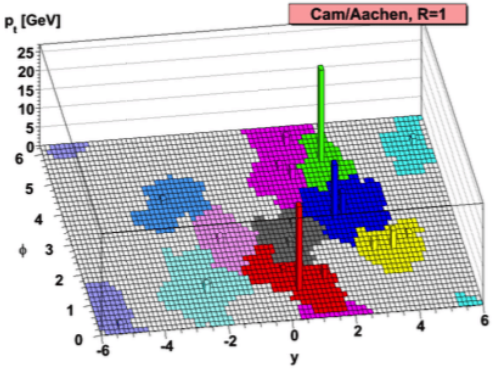
Introduction - The Generalised-k_T algorithm

- Most modern implementations of jet clustering algorithms today stem from the generalised k_T algorithm
- The distance measure between two particles (i and j) is $d_{ij} = min(p^n_{Ti}, p^n_{Tj})\frac{\Delta^2_{ij}}{R^2}$
  - $\rightarrow$ R is an input parameter which scales how large the resulting jets will be (acts as a sort of cut off)
  - $\rightarrow$ Δ_{ij} is the distance between i and j in (η, ϕ) space
- Also use $d_{iB} = p^n_{Ti}$, the beam distance, which acts as a cut off
- Each iteration of algorithm will compute all possible 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, remove it and continue until everything has been discarded as a jet

Introduction - The anti-k_T and Cambridge-Aachen Algorithms

- The two most common descendants of the generalised-k_T are the anti-k_T and the Cambridge-Aachen (CA) algorithms
- Both take the exact form as in the last slide, with the exponent of the transverse momenta set to $n = -2$ (anti-k_T) and $n = 0$ (CA)
- CA uses a purely geometric d_{ij}
- Anti-k_T prefers to cluster around a hard center of a jet





Introduction - Two Higgs Doublet Models (2HDMs)

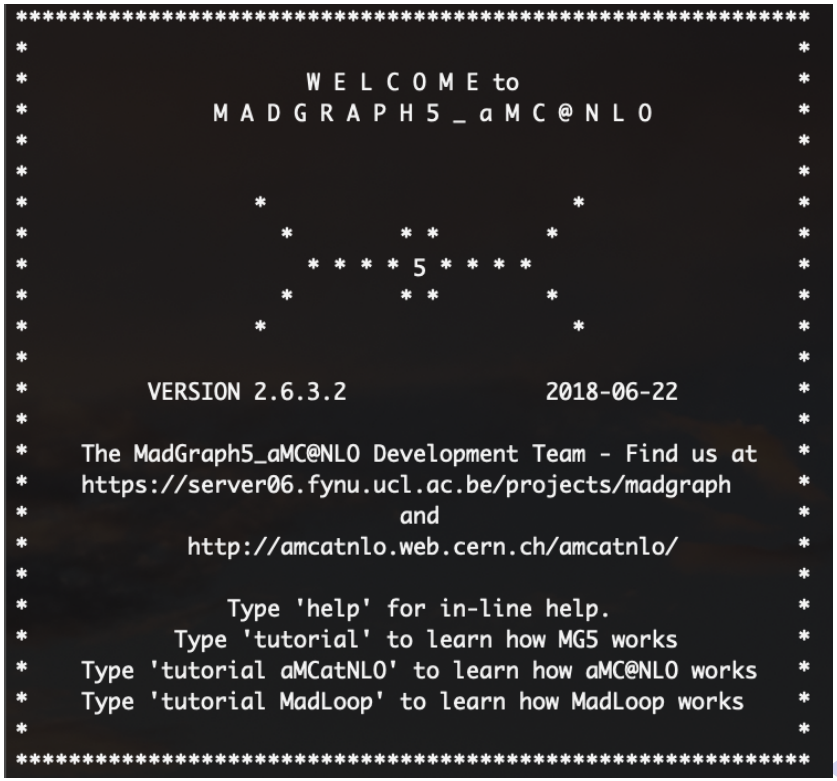
- Simple extension to the SM - addition of a second Higgs doublet such that Φ_a $(a = 1, 2)$
- As a result, electroweak symmetry breaking (EWSB) yields five physical Higgs states
  - $\rightarrow$ CP-even scalars; h and H (where conventionally $m_H > m_h$, the discovered 125GeV Higgs can be identified with either of these
  - $\rightarrow$ CP-odd pseudoscalar A
  - $\rightarrow$ A pair of charged states $H^{\pm}$ with mixed CP properties
- Where kinematically allowed, we can therefore have decays of the form $H \rightarrow hh$, and further on both of the lighter Higgses can decay into a pair of b-jets
- Are modern jet reconstruction techniques optmised to look for such topologies?

Methodology - Selecting a Benchmark

- Before analysing events we must select a suitable set of parameters in the 2HDM framework for our model
- We work in scenario where $m_H = 125\text{GeV}$, so for $H \to hh$ decays we require $m_h < \frac{m_H}{2}$
- To find experimentally interesting points (i.e. not excluded), we scan over parameters using 2HDMCalculator, interfaced with HiggsBounds and HiggsSignals, to check against experimental constraints
- Passing points are then checked against flavour constraints using SuperISO
- Two benchmark points are selected, with $m_h = 40, 60\text{GeV}$ respectively

Methodology - Simulation Details

- We generate samples of $\mathcal{O}(10^5)$ events for the signal process $gg \to H \to hh \to bb\bar{b}\bar{b}$ using MadGraph5 and showered in Pythia8
- Jet reconstruction, cutflow and analysis are them performed with MadAnalysis5 interfaced with FastJet



Methodology - Cutflow

- The following cutflow is applied, firstly at the particle level, and then at jet level, to provide a simple simulation of a detector:

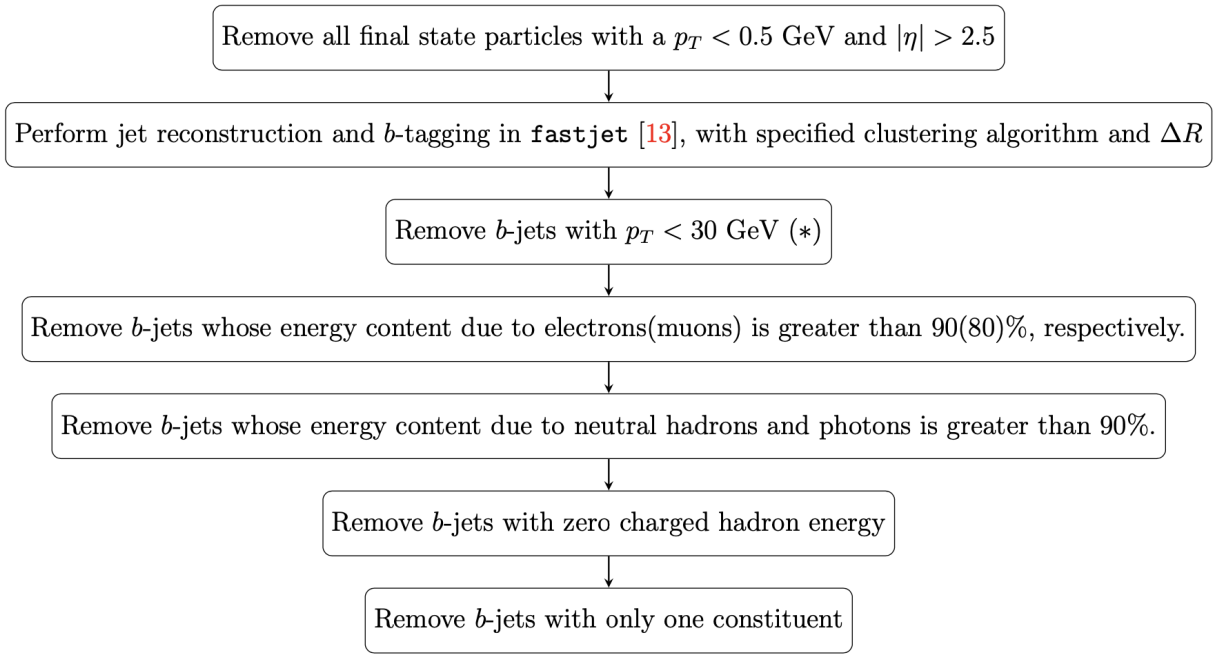
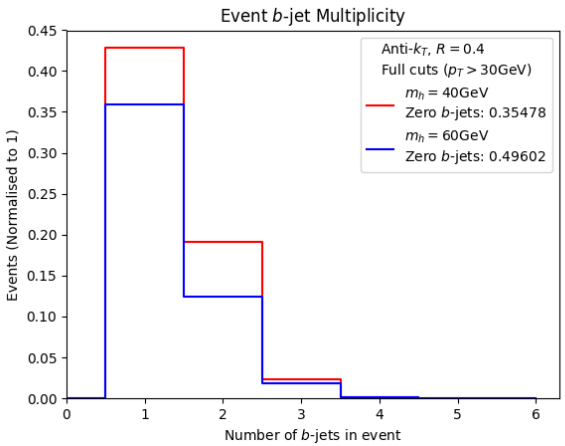
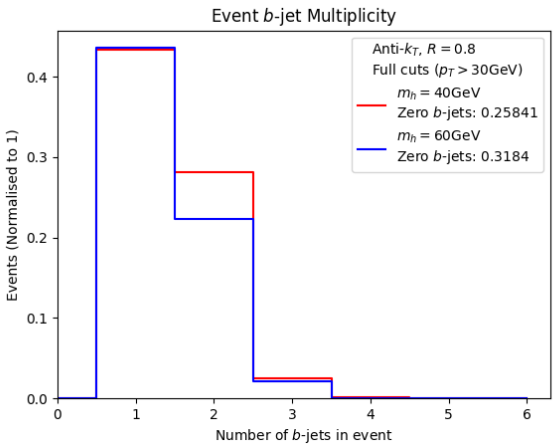


Figure 5: Description of our initial procedure for jet clustering, b-tagging and selection of jets. Notice that the starred cut $(*)$ will eventually be modified in our optimised b-jet selection. Also note that our analysis is performed at particle rather than detector detector, so MC truth information is used for cuts on jet constituents.

Results - Standard Clustering

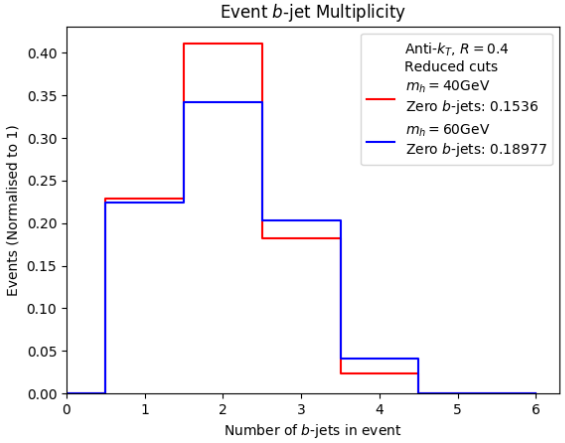
- We firstly investigate the visibility of the signal when using a standard set of jet reconstruction parameters - anti-k_T with $R = 0.4, 0.8$
- Check b-jet multiplicity in events
- We know there are four b-quarks at parton level, how many jets however do we see after jet reconstruction, tagging and cutflow?
- Lots of jets are lost to cutflow! - Impossible to reconstruct signal!

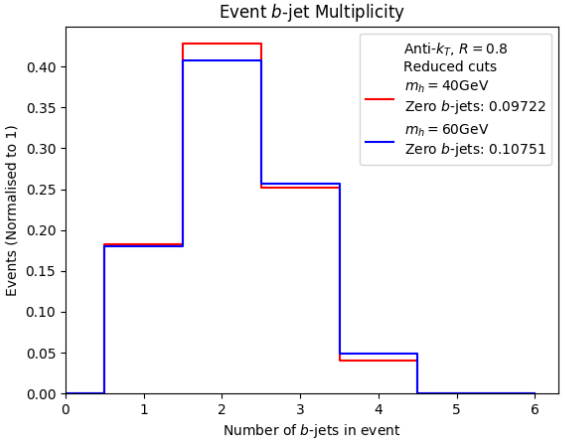




Results - Standard Clustering (cont.)

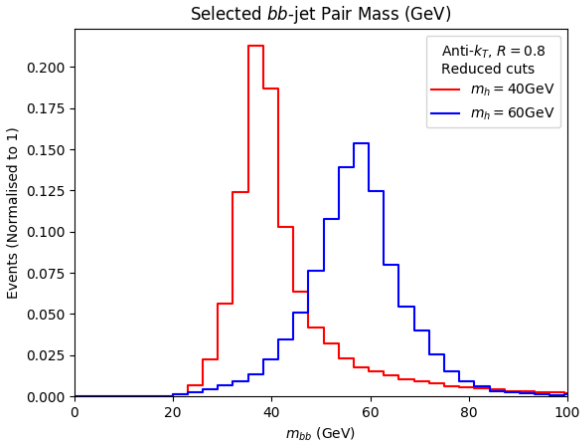
- Replace p_T cut with one slightly lower
- p_T of first, second, third, fourth (p_T ranked) b-jets must be $> 20, 15, 15, 15$GeV respectively
- Much better statistics - can actually reconstruct signal!

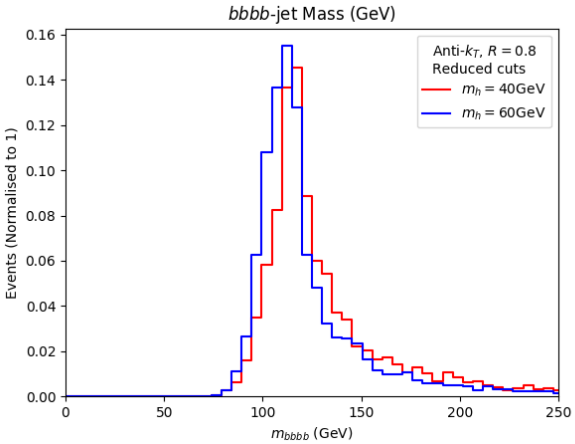




Results - Standard Clustering (cont.)

- We can reconstruct Higgs masses m_h and m_H in the invariant dijet and fourjet masses as seen in plots where we have used $R = 0.8$
- Statistics still rather small, can we do ever better?





Intermission - Variable-R Jet Clustering

- With further improvements in mind we consider an alternative, more recent variation of the above clustering - the so-called **variable-R** clustering
- As name suggests, we no longer use fixed cone, but instead allow the cone size to change depending on the particles we are reconstructing..

Intermission - Variable-R Jet Clustering (cont.)

- Convenient to rewrite $d_{ij} = \min(p_{Ti}^n, p_{Tj}^n)\Delta_{ij}^2$ and $d_{Bi} = p_{Ti}^n R^2$
- Fixed R parameter is replaced with the function $R_{\text{eff}i} = \frac{\rho}{p_{Ti}}$ where ρ is now the (dimensionful) input parameter
- d_{Bi} is therefore suppressed for larger p_T objects - more likely to be declared jets over combining with more particles

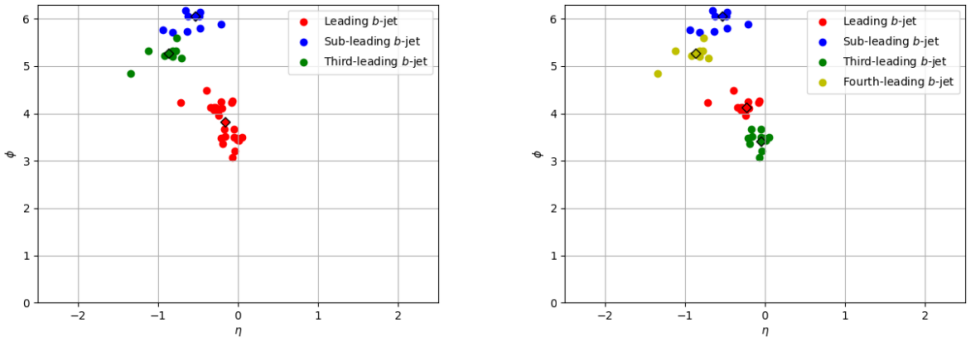
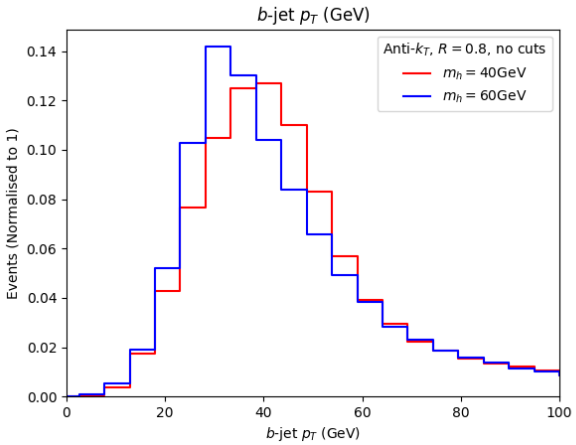
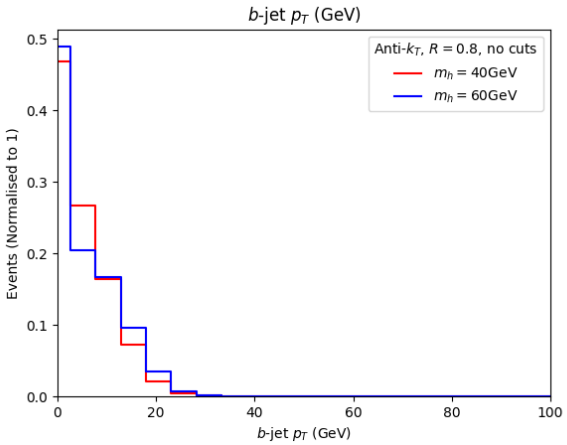


Figure 3: Same plot as in Fig. 2, however, here, the given event is clustered into three b-jets when a fixed $R = 0.8$ is used (left) and four b-jets when we use a variable-R approach (right).

Intermission - Variable-R Jet Clustering (cont.)

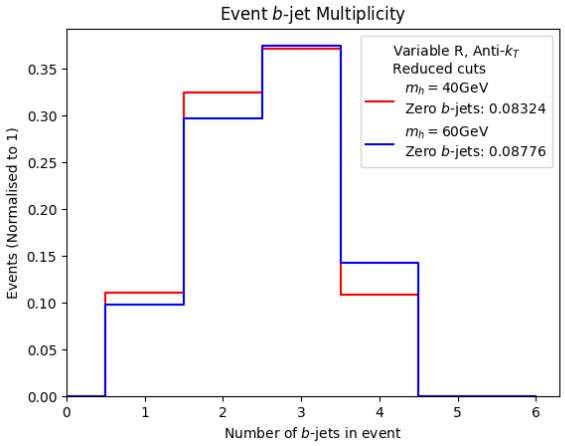
- To demonstrate why this could be useful, consider the p_T of the leading and sub^3-leading b-jet in a signal event
- Vast difference - so need a cone which will balance between narrow high-p_T jets and more spread low-p_T jets in same event

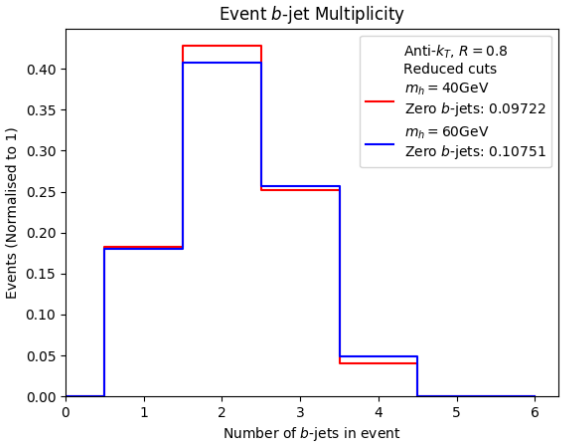




Results - Variable-R

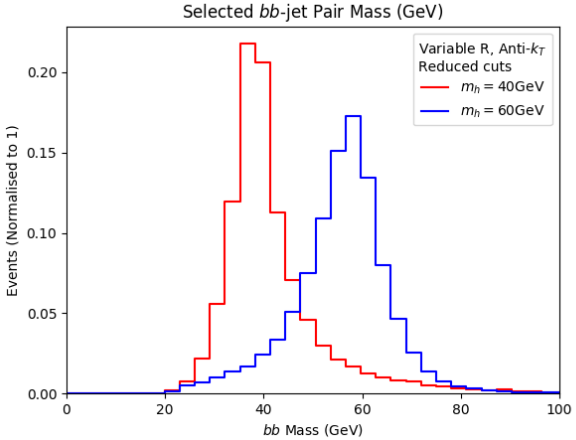
- With variable-R implemented in our analysis, we can compare the performance with a fixed-R
- Firstly compare b-jet multiplicity against fixed-R
- Can see variable-R reconstruct b-jets at a significantly higher rate - more than double the number of events that contain all four expected b-jets

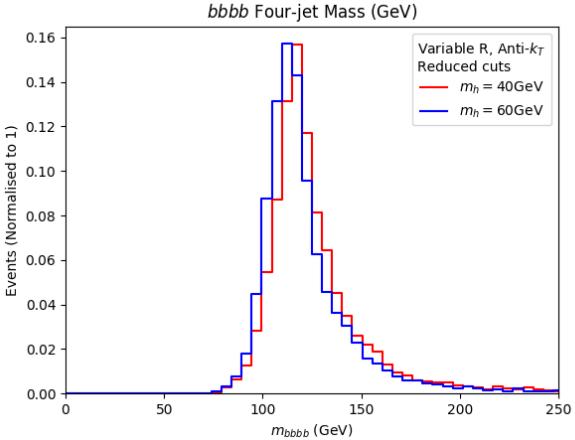




Results - Variable-R (cont.)

- Can also look at the Higgs mass reconstruction of jets clustered with variable-R
- Sharp visible peaks of signal are present
- To gain meaningful insight we must consider backgrounds..





Results - Signal vs Background Significance Rates

- We build a selection analysis to pick out events that look 'signal-like'
- Firstly we require a way of pairing together to corresponding b-jets from the same Higgs (that therefore should construct m_h)
  - $\rightarrow$ To do so all possible bb pairs are constructed, and the one which minimises $|m_{bb} - m_h|$ is chosen (as is the other therefore by default)
- Events are then selected as signal if they pass the following criteria:

Select events that contain exactly four b-jets

↓

Remove event if $|m_{bbbb} - m_H| > 20$ GeV

↓

Using di-jet pairings chosen in above analysis

↓

Remove event if $|m_{bb} - m_h| > 15$ GeV

Results - Signal vs Background Significance Rates (cont.)

- We run the analysis for the signal and the leading backgrounds, which are found to be: $pp \to t\bar{t}$, $pp \to bb\bar{b}\bar{b}$ and $pp \to Zb\bar{b}$
- The number of passing events are counted, and the significance for a given algorithm is computed as:

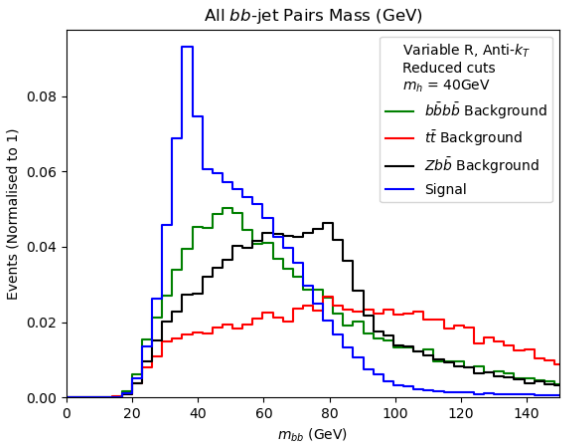
$$\Sigma = \frac{N(S)}{(N(B_{t\bar{t}})+N(B_{bb\bar{b}\bar{b}})+N(B_{Zb\bar{b}}))^{1/2}}$$

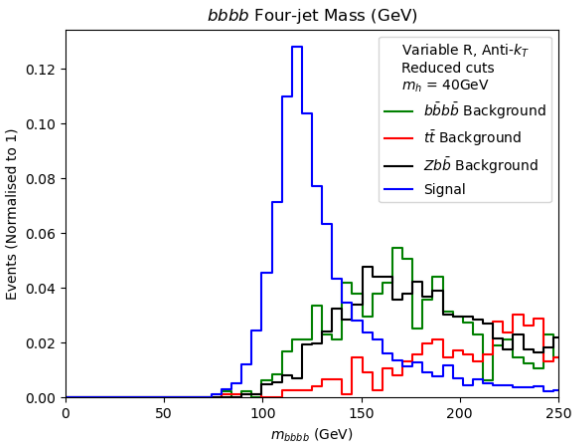
- When running over signal and all backgrounds our results are as follows, using a luminosity of $\mathcal{L} = 140\text{fb}^{-1}$

	variable-R, $\rho = 20$ GeV	$R = 0.4$	$R = 0.8$
40 GeV	0.358	0.061	0.160
60 GeV	6.676	2.074	3.138

Conclusions and Review

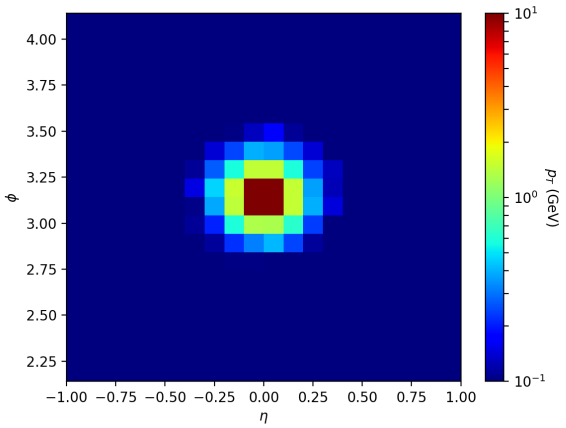
- We have demonstrated that certain 2HDM-II topologies of experimental interest are suppressed with current jet reconstruction parameters
- We establish that a variable-R jet clustering can outperform fixed-R implementations currently in use
- Can read more details of our study in our paper; arXiv:2008.02499

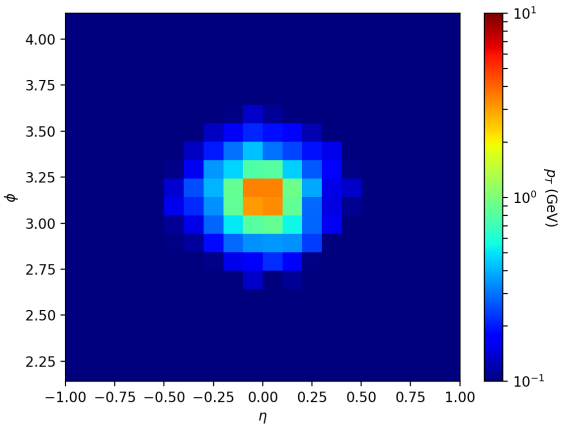




Future Work

- While results gathered are for specific model/process - results can be applied to any BSM models with four-b final states
  - $\rightarrow$ For example the rich parameter space of 2HDMs, scenarios with $m_h = 125$GeV etc
  - $\rightarrow$ Alternatively can consider $H \rightarrow AA \rightarrow bb\bar{b}\bar{b}$ where allowed
- Can variable-R be useful in machine learning for jets?
  - $\rightarrow$ We are currently investigating whether variable-R jets can be better distinguished from backgrounds in image recognition problems





Thanks for listening!