University
of Glasgow

Search for new resonances decaying to a Higgs and W/Z boson in the $b\bar{b}$+lν final state for a $\sqrt{s} = 14$ TeV, and testbeam alignment studies using prototypes for the inner tracker upgrade at ATLAS

©

Supervisors:

Prof. Dr. David Britton

Dr. Andy Buckley

School of Physics and Astronomy

College of Science and Engineering

December 2015

Abstract

This report is split into two self contained parts:

Resonance searches: **??** and Alignment studies: **??**.

The review concerned with the resonance searches briefly introduces the detector and the process of Monte Carlo simulations, before delving into the ongoing investigations.

The resolution of the merged $b\bar{b}$ track jets using a new generalised sequential recombination algorithm known as the variable R algorithm is explored. A particular focus is on the alignment of truth Monte Carlo B-hadrons with the reconstructed jet. Initial studies using standard model VH production show that the variable R jets better reconstruct the truth level B-hadrons direction compared to the anti-k_t algorithm.

The expected significance of signal over background for a 1/1.5 TeV resonance predicted by the Heavy Vector Triplet Model (HVT) is determined. This study comprises different discriminating variables being used in a window cut based multivariate analysis. Initial results show that a topological cuts based selection is almost optimum but more discriminating power may be added with some substructure variables.

Alignment studies will be performed on prototypes for the upgrade of the inner tracker of the ATLAS detector. Such prototypes will consist of complex pixel shapes, orientations and other varying environmental factors, which will need innovative methods to create a generic framework for common reconstruction. The EUTelescope framework is just that framework and has been updated to work with a plethora of different geometric and environmental setups. The initial results show the pattern recognition working in tandem with the track fitting and alignment. With results as expected for varying beam energies and magnetic fields.

Contents

Acknowledgements

Chapter 1

Higgs Searches $(H \rightarrow b\bar{b})$

The variable R studies are based on standard model VH production Monte Carlo which is discussed in section **??**. The optimisation studies are for VH resonances and is motivated in section **??**.

1.1 Introduction

An introduction to the process of Monte Carlo simulation and the ATLAS detector is needed before any description of the search can be understood.

1.1.1 Simulation

Monte Carlo simulations are an important aspect to any analysis. Not only do they allow for refinement of analysis techniques but must also be used to compare theoretical expectation with reality.

Simulation can be split into two parts, event generation and detector response. These two parts come sequentially with the final result compared to data as shown in figure **??**.
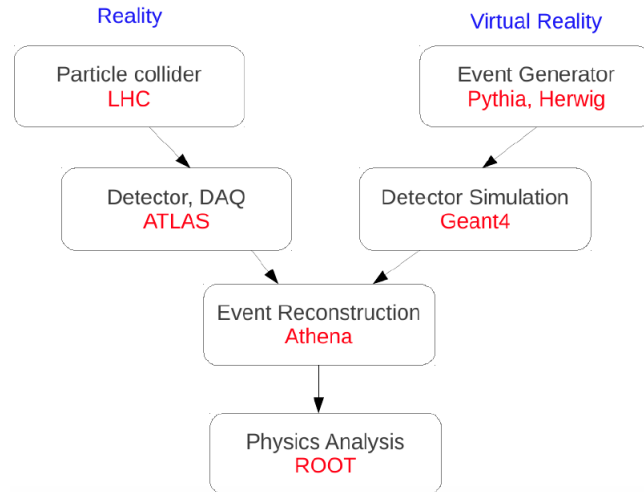
Figure 1.1: Work flow of standard analysis. ©John Morris.

The process of Monte Carlo simulation from matrix elements to our observed decays before simulation of the detector can be described approximately with 5 steps [?]. These steps are shown on figure ??, and use a range of generators and models to arrive at the final state particles this analysis is interested in. The idiosyncrasy of each part are numerous, however the uncertainties associated can be briefly highlighted since they limit the accuracy of our search.

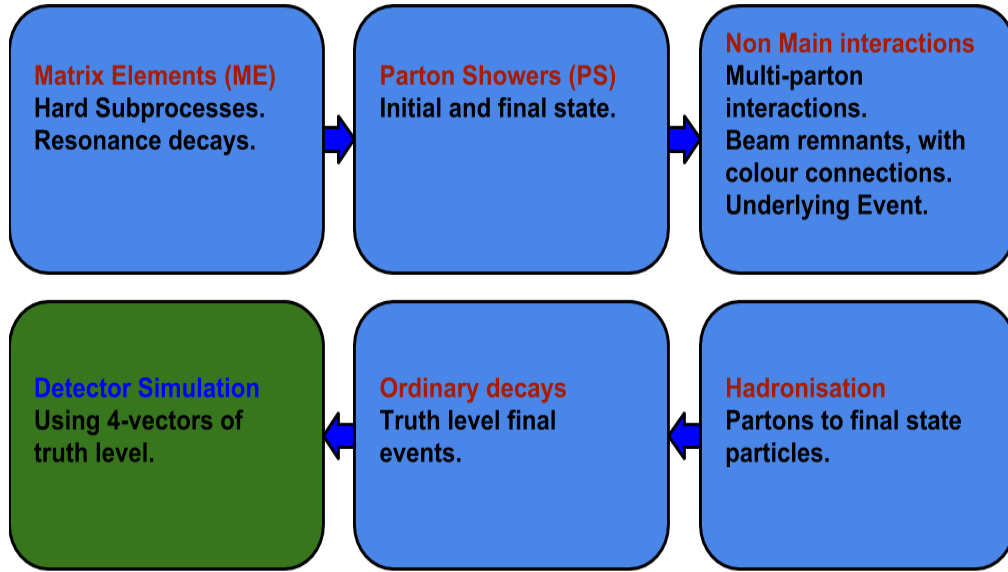Figure 1.2: Rough guide from interaction to truth level particles. The use of the 4-vector truth information in detector simulation will also have many parts.

The very fact that each part can be calculated in a different way leads to theoretical modelling uncertainties. These uncertainties include generator (matrix element calculations), parton density function (PDF), parton shower (PS) and hadronization uncertainties. Generator uncertainties come from the limit on the order in the calculation of the matrix elements. PDF uncertainties come from the use of different parameterisations. At the LHC CTEQ, MSTW and NNPDF are used, each arriving at a different result. PS uncertainties come from the divergencies that must be avoided in any QCD calculation. How this is dealt with depends on the generator. Two common examples would come from the generators PYTHIA and HERWIG, which calculate the parton with the highest p_T and angle respectively. Hadronization occurs at very low energy scales in which perturbation theory is not valid. Therefore a suitable model is needed and two approaches are often used: the Lund string and Cluster models. Each model similarly to PS modelling will produce a different result, therefore this leads to further uncertainties in the overall modelling.

A general practice to determine the total uncertainty is by simply using different models for each calculation and taking the difference, known as the

offset method.

At this stage, known as the truth level, the simulation has produced stable final state particles with 4-vector information and all parent-daughter relations, such as b-tagging from parton level. Now the interaction of the final state particle and the detector must be considered. GEANT4 is used to take the output 4-vectors of the truth level generators and model the detector response. This is a very complex process and involves modelling all subdetectors of the ATLAS experiment. This modelling can be done to different degrees of accuracy to save computation time with the final output matching the output of the ATLAS detector.

1.1.2   The ATLAS Detector

The ATLAS detector is one of the experiments situated at an interaction point round the LHC at CERN.

The standard geometry of the ATLAS detector is a right hand coordinate system with the origin at the intersection point, the z-axis pointing along the beam pipe, the x-axis point towards the LHC centre and the y-axis to the heavens. It's natural to use the cylindrical coordinate system in many situations: this is defined with the axis of rotation round the z-axis (ϕ) and from the z-axis ($\eta = -\ln\tan(\frac{\theta}{2})$). The definition of η is due to it's use to define a Lorentz invariant measure of angular separation: $R_{ij} = \sqrt{\Delta\eta_{ij} + \Delta\phi_{ij}}$.

The experiment can be split into three main subsystems: the inner tracker, the calorimeters and muon spectrometer [?]. The inner tracker is composed of pixel sensors closes to the beam line and strip sensors thereafter. The whole tracker is immersed in a 2T magnetic field used to measure the momentum of charged particle from the curvature. This inner detector makes accurate measurements of position and momentum, up to $\eta < 2.5$, and a transition radiation tracker (TRT) is situated after the silicon detector, up to a $\eta < 2.0$. The pixel detector is essential in b-tagging using extended secondary vertices (SV) from the primary vertex (PV). The TRT is particular useful for electron identification and determination if photons are converted or not.

The calorimeters are located beyond the solenoid and cover the range $|\eta| < 4.9$. With a plethora of detector technologies used, the liquid argon electromagnetic (EM) calorimeters are divided into 3 sections - the barrel, endcap and forward sections - $|\eta| < 1.475$, $1.375 < |\eta| < 3.2$ and $3.1 < |\eta| < 4.9$. The hadronic calorimeters surround the EM calorimeter with a coverage of $|\eta| < 4.9$ and use a combination of scintillating tiles and liquid argon as ac-

tive materials. The muon chamber uses three large air core toroidal magnets, each containing eight superconducting coils to deflect muons, to measure the curvature of the resulting track [**?**].

The mammoth task of reducing the 40 MHz bunch crossing rate to a manageable 300 Hz is dealt with in three stages [**?**]. The first (Level 1) is hardware based and takes input mostly from muon and calorimeter trigger systems. This trigger reduces the rate to less than 75 kHz which is reduced further to ≈300 Hz by the next two trigger systems. These are software based on the ATLAS offline Athena framework.

1.2   The Standard Model $VH \to b\bar{b} + ll/l\nu/\nu\nu$ Decay Channel

The channel $H \to b\bar{b}$ is particularly difficulty to detect due to the large MultiJet (MJ) background at the LHC. Combined with other backgrounds this makes the most prominent decay channel invisible. The association of a vector boson gives a clear signature in the muon and calorimeter trigger systems which allows the identification of the Higgs over many backgrounds.

The focus on the VH production to tag the $H \to b\bar{b}$ decay will remove some but not all backgrounds. Different types of background decays will pollute the signal yield to different extents. The more similar the decay products of the background to the signal, the more difficult to differentiate. The main backgrounds come from (W/Z)+jets and $t\bar{t}$ production with smaller contributions made from single top, diboson production (WW,WZ,ZZ) and MJ [**?**]. Top quarks decay over 99% of the time via $t \to Wb$. With a combination of leptons, neutrinos and actual b-tagged jets for each top decay, this process can be easily confused for signal. The same is true with (W/Z) + jets and diboson production, i.e the final state is similar to the signal final state. The MJ background refers to any event with multiple jets but no prompt isolated electrons or muons [**?**]. This background comes in two forms. One is from large energy fluctuations from jets which acts as E_T^{miss} in the calorimeters and the other as misidentified electrons. The first pollutes the signal in the 0 lepton channel and the other, both lepton channels.

To remove these backgrounds discriminating variables must be used. The invariant mass of the dijet system would be one example. This is the invariant mass of the $b\bar{b}$ pair and must equate the Higgs mass m_H (Note this would be defined in terms of a clustered fat jet) . If this is different from the Higgs

mass then this signifies that the event was background. In the same light other internal and distributive characteristics of the jets can be used. One such substructure variable $\sqrt{d12}$ - which is the kt-splitting scale - is explained in section **??**.

Reliant b-tagging is another important factor for better discriminating power. This is made more difficult at high P_T as the b-jets produced will become merged (unresolved). This is due to the combined momentum of the decay products having to be focused more in one direction. The geometric matching between the truth B-hadron and reconstructed jet is crucial for high b-tagging efficiency. The performance of a new generic sequential recombination algorithm, known as the variable R jet algorithm, is employed to minimize the deltaR (The angular separation) between the B-hadron and the jet.

1.2.1   Variable R Jets

Calorimeter jets are localised clusters in the EM and Hadronic Calorimeters. Track jets are clustered tracks produced in the inner tracker. The most common clustering algorithms used are sequential recombination algorithms, which are best suited to deal with the large QCD background from the pp collisions.

The algorithms use two metrics defined in equations **??**. The p_{Ti}^{2n} terms correspond to the tangential momentum, with each algorithm applying their own (n) parameter which is set to $n = -1$ for the anti-k_t algorithm. The R_0 parameter is a set constant for the full clustering process and plays a role in the jet's size. The ΔR_{ij} parameter is defined in section **??** and is between clusters in the tracker or calorimeter.

$$d_{ij} = min(p_{Ti}^{2n}, p_{Tj}^{2n})\Delta R_{ij} \qquad with \qquad d_{iB} = p_{Ti}^{2n}R_0 \qquad (1.1)$$

A summary of how each algorithm works is as follows: Using their preset value of (R) and (n); compute d_{ij} and d_{iB}; if d_{ij} is the largest then combine jets and repeat; otherwise the jet object (i) is removed as a final jet. This is repeated until the termination point set for each algorithm.

The fixed R parameter throughout the clustering process can be problematic for highly boosted decay products since the jets produced are often merged. This can be mitigated using a clustering parameter R which varies in size with jet momentum [**?**].

$$R_0 = \bar{R}_{eff}(P_T) = \frac{\rho}{P_{Ti}} \qquad (1.2)$$

The variable-R algorithm does this with the parameters defined in equation **??**.

1.2.2 My contribution

The highly boosted decay products of the Higgs will often be unresolved. Therefore, the use of variable-R jets is ideal to resolve the deposited energy from the $b\bar{b}$ decay products. A comparison between the anti-k_t and variable-R algorithm for different input parameters is made. Presented here is the algorithms performance at reconstructing the jet in the same direction as the truth level B-hadron. The signal sample is simulated standard model WH events decaying via $VH \to b\bar{b} + l\nu$. Figures **??** show the geometric displacement between the reconstructed jet and the truth B-hadron.
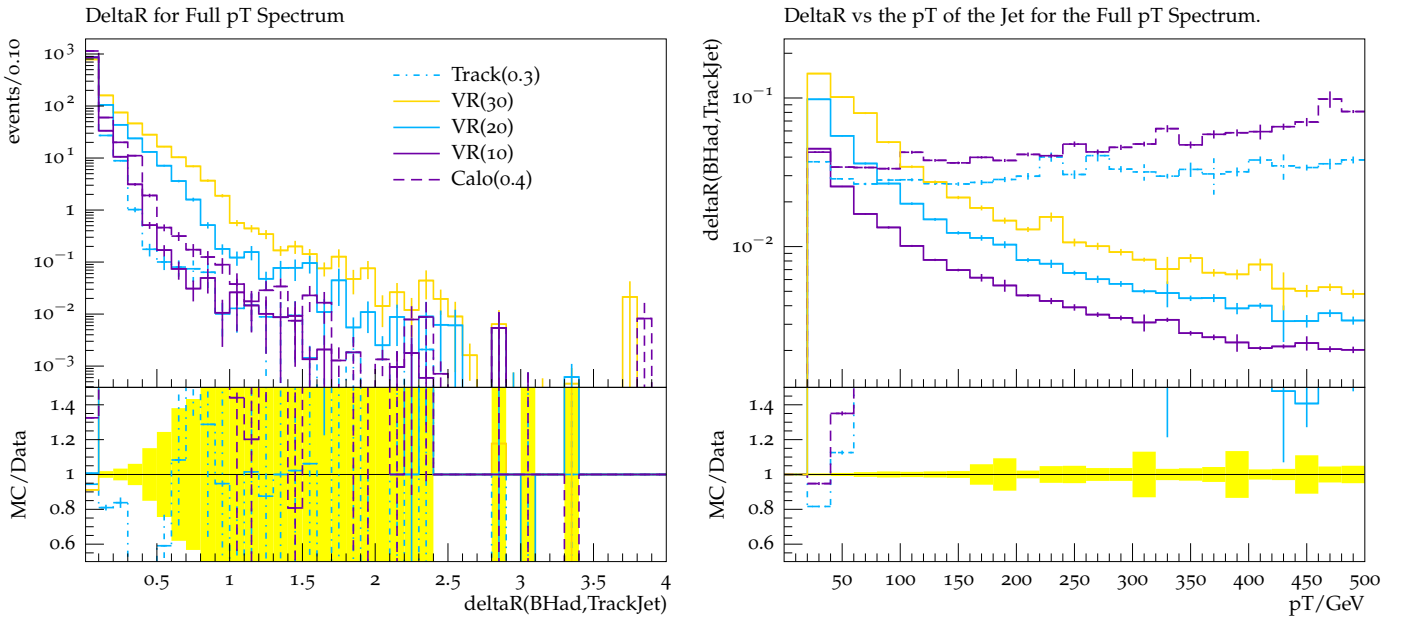

Figure 1.3: The variable deltaR(BHad,Trackjet) is the angular distance. Figure **??** illustrates how well each algorithm with different input variables matches the direction of the B-hadron. Figure **??** shows this performance over the inclusive P_T of the reconstructed jet.

The variable R jet have a smaller angular displacement compared to the B-hadron. Figure **??** shows the variable-R jets with a smaller angular separation at higher P_T compared to the anti-kt algorithm.

1.3 Beyond the Standard Model and $VH \rightarrow b\bar{b}+ll/l\nu/\nu\nu/qq$

With the discovery of a Higgs consistent with the Standard Model the question of naturalness is under reexamination [?]. The naturalness problem arises through the corrections the Higgs mass receives from virtual particles. If the virtual particles can take any momentum to the Planck mass scale then the correction must be huge. This is strange since the Higgs mass is a relatively small, therefore it is natural to think that there must exist a cut off for the maximum momentum taken by the virtual particles. The unnaturalness of the standard model motivates many theories to be put forward. Some of these dynamical electroweak symmetry breaking scenarios predict resonances which decay to a Higgs and vector boson [?]. These will subsequently decay into many different final states, a subsection being $VH \rightarrow b\bar{b}+ll/l\nu/\nu\nu/qq$.

1.3.1 Cut Optimisation and Significance

The setting of exclusion limits and the determination of signal strength needs the removal of the background events and as small as possible uncertainty. The removal of background can be done in many different ways. However a simple yet effective way is to perform a series of window cuts on a set of discriminating variables. This is done using the toolkit for multivariate analysis (TMVA), which is a ROOT integrated environment for the application of multivariate classification [?]. TMVA will maximise the background rejection for a given signal efficiency and produce the cuts corresponding to these values.

Total background rejection is not precisely what is of most interest to this search. Since this does not take into account the relative shapes of signal to background [?]. To shed some light on this the significance must be calculated. The calculation will show larger significance if the signal shape is drastically different from the background. The definition of significance varies from analysis to analysis. This is due to different approximations and assumptions being used.

1.3.2 My contribution

Monte Carlo VH 1.0/1.5 TeV resonance events from the HVT model and backgrounds (W+jets, $t\bar{t}$, Z+jets and single top) with preselection cuts:

- 1 lepton with pt > 25 GeV;

- $E_t^{miss} > 20$ GeV;

- One R=1.0 trimmed anti-kt jet with $p_t > 250$ GeV;

- Two b tagged 0.3 track jets ghost associated (GA) to leading large-R jet;

- Zero b-tagged 0.3 track jets not GA to leading large-R jet;

- B-tagging: Truth labeled b,c and light jets with 75%, 10% and 3% tagging efficiencies respectively;

are used as input to the TMVA. The b-efficiencies are not optimal and could see further improvements.

One topological and two substructure parameters are presented here as the discriminating variables of interest:

$$\frac{VH_{pt}}{H_{pt} + V_{pt}} \qquad \sqrt{d12} \qquad m_H \qquad (1.3)$$

The substructure variables $m_H/\sqrt{d12}$ are from the R=1.0 anti-k_t jet. The splitting scale $(\sqrt{d_{12}})$ is defined as the minimum d_{ij}/d_{iB} before the final clustering step [?]. The topological cut was found to be a powerful discriminating characteristic between signal and background for both resonance masses. This can be seen on figures ?? were each variable was optimised individually. Therefore this variable is used to produce significance plots in combination with the other variables to search for any improvements.
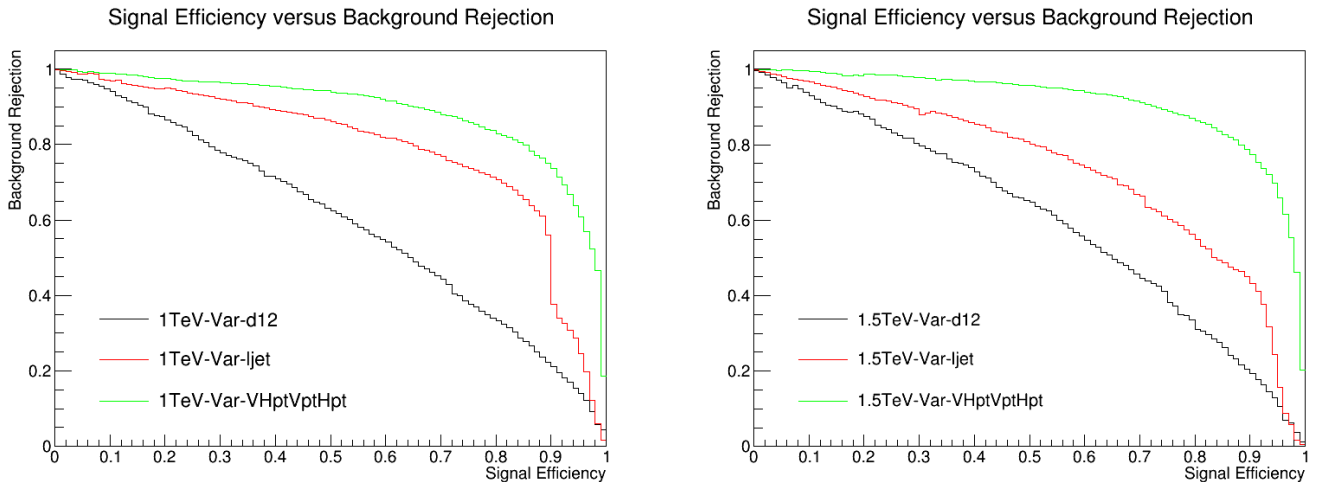

Figure 1.4: Figure ?? shows the background rejection for a given signal efficiency for each cut trained on the 1 TeV resonance sample. Figure ?? shows the same thing trained on a 1.5 TeV resonance. Black: $\sqrt{d12}$ Red: m_H Green: $\frac{VH_{pt}}{H_{pt}+V_{pt}}$

The behaviour of the TMVA must be checked to ensure the validity of the results. One simple way of doing this is checking the cuts the TMVA predicts to ensure they appear realistic. Figures **??/??** show the cuts for each efficiency for the ROC curves **??**. The Higgs mass appears lower than 125 GeV due to the miscalibration of the jets.
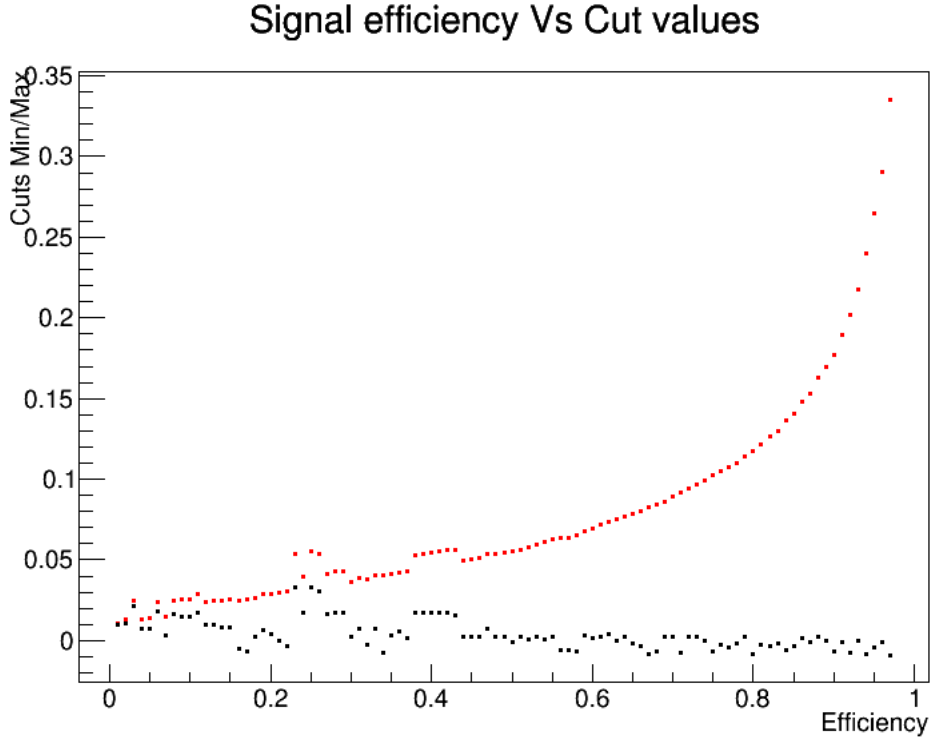

Figure 1.5: The window cut on the topological variable for each efficiency. The cuts window increases from left to right due to a need for greater signal efficiency.
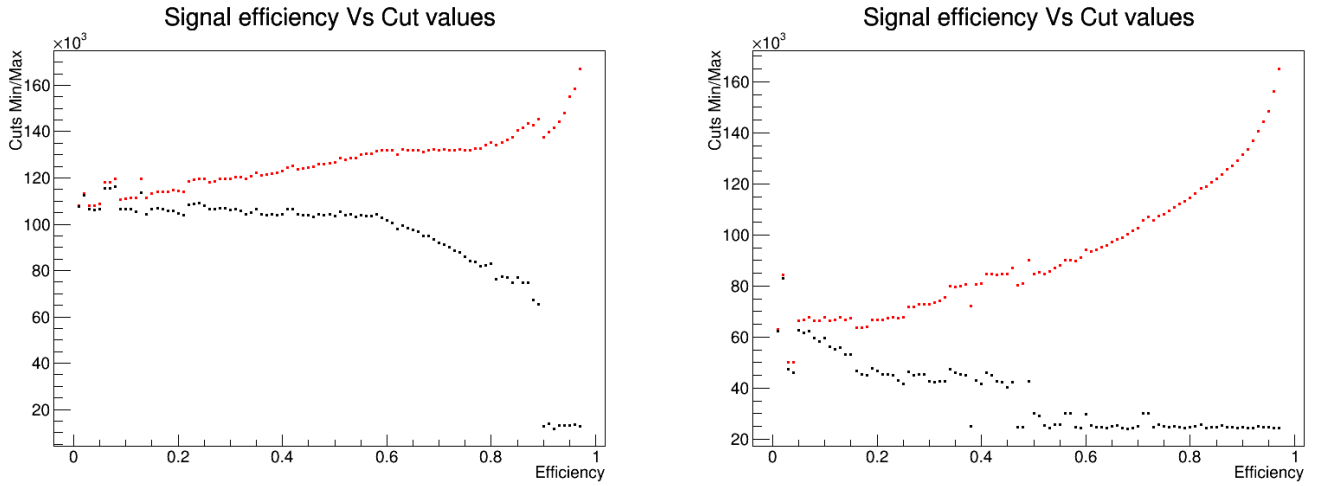
Figure 1.6: Figure **??** shows the cuts for m_H. Figure **??** shows the cuts for $\sqrt{d12}$.

The significance used is defined as as the quadrature sum of $\frac{S}{\sqrt{B}}$ per bin. This definition might not be sufficient due to the small number of signal events which is discussed in section **??**. The significance is calculated over the VH mass spectrum using the cuts for a particular signal efficiency.
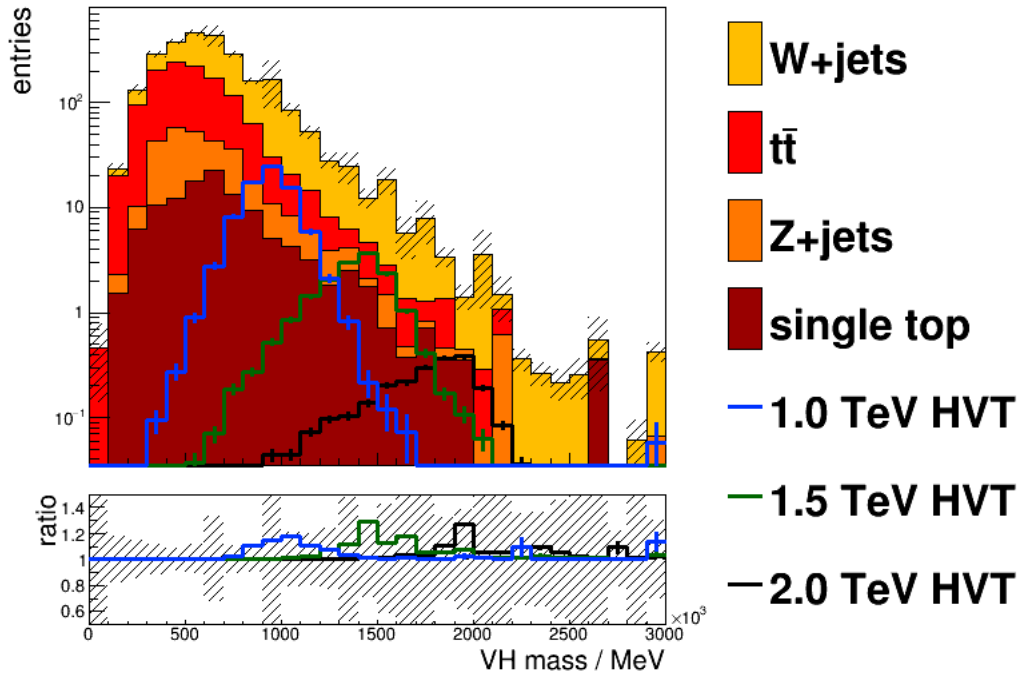
Figure 1.7: The VH mass spectrum with the resonances and backgrounds after preselection.

The search would like to have common selection for the 1 and 1.5 TeV resonances. Therefore only the 1.5 TeV resonance sample is used to train the TMVA and produce the necessary cuts. The cuts for the 1.5 TeV resonance were then used with the 1 TeV sample to separate signal from background and the results are shown on figures **??/??**.
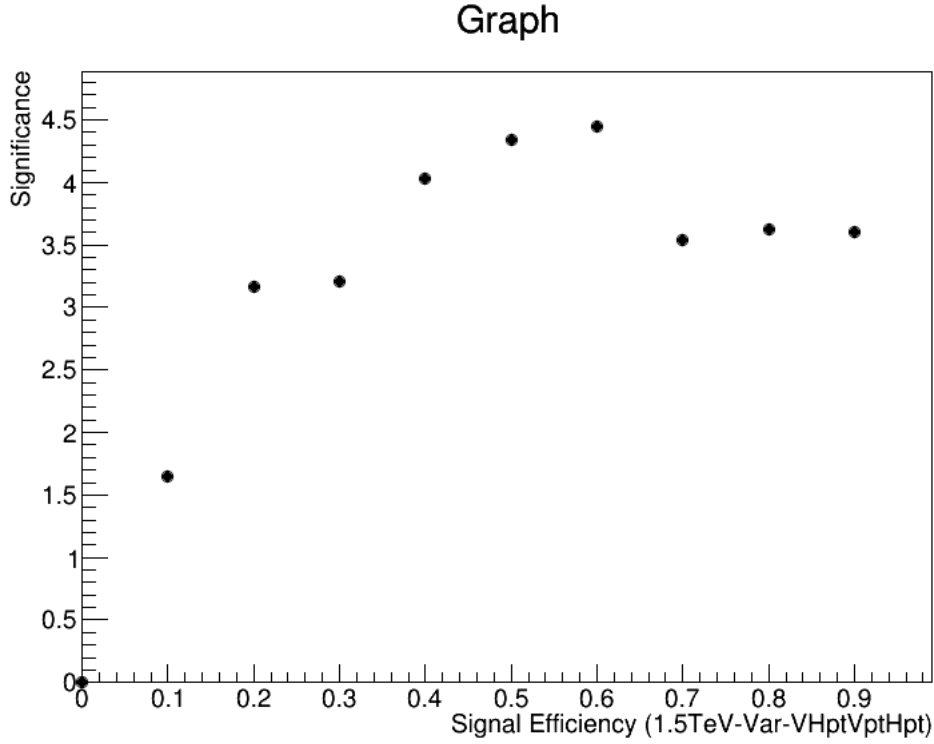
Figure 1.8: The trained TMVA cuts for 1.5 TeV used with the 1 TeV sample for the topological variable alone.

Figure 1.9: The trained TMVA cuts for 1.5 TeV used with the 1 TeV sample for the topological cut combined with $m_{b\bar{b}}$?? and $\sqrt{d12}$??

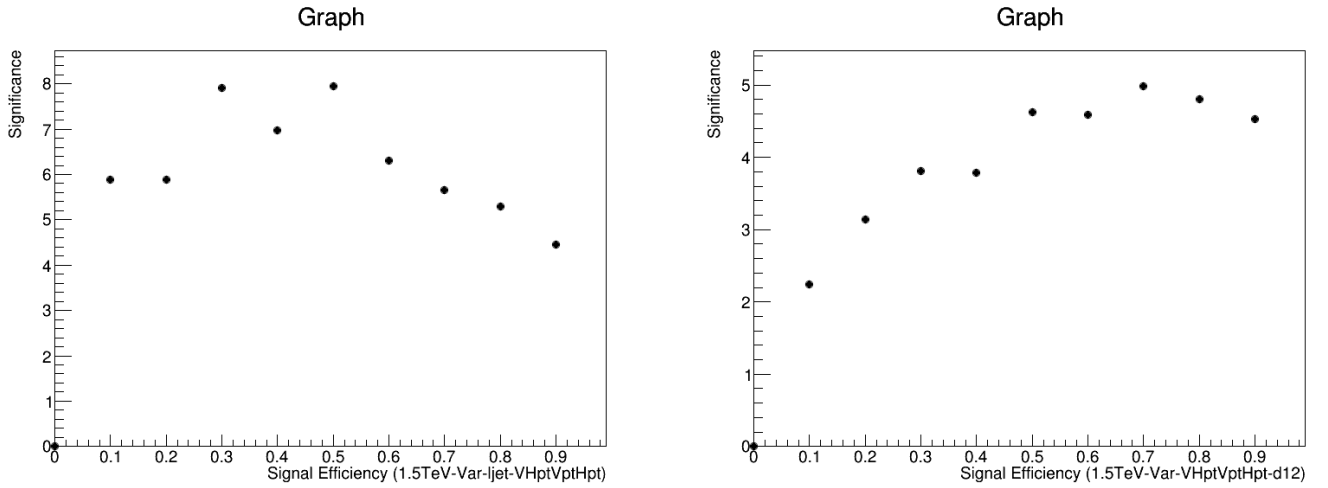The substructure variables both have an improvement on the significance compared to the topological cut alone. The m_H variable has the most drastic improvement. However the $\sqrt{d12}$ could seen further improvement after trimming.

Chapter 2

The integration of the General Broken Lines fitting algorithm into EUTelescope.

2.0.3 The triplet finder

2.0.4 Creating a link between all points

After pattern recognition we assume we have collected all the hits which make up a track. These hits are used to parameterise a track, this is basically a set of parameters which describe a track through the global frame of EUTelescope. This is done in EUTelNav.cc. This track describes the slope and position of the track at a given z global position. A further track parameter is the q/p (Charge over momentum). Now each hit on each plane is linked together using a Jacobian. This links the change in the position, incidence and q/p with all other points. This is the first aspect of the creation of a track.

2.0.5 Propagation from Global to Measurement Frame

The linking of all points in the global frame allows changes in each part of the track to be associated to other parts of the track.

Testbeam studies are carried out to characterise a novel sensor or read-out chip, the so called device under test (DUT). For this purpose, a beam telescope is used. It consists of several reference pixel sensors placed inside a particle beam. The DUT is then placed between the telescope reference sensors. This allows to reconstruct particle tracks on a single particle level and extrapolate them onto the DUT. These tracks can then be used in various subsequent

analyses [**?**].

2.0.6   EUTelelecope

EUTelescope is a highly generic testbeam reconstruction and analysis framework for this purpose [**?**]. Despite being developed hand in hand with the AIDA pixel telescope and its DAQ framework EUDAQ, EUTelescope can be used with any desired DAQ framework. The tracking and alignment in EUTelescope is done in different ways. However there is a common theme in all which can be understood broadly.

2.0.7   Pattern Recognition, Track Fitting and Alignment

To reconstruct a particles trajectory through sensitive and non sensitive material requires consideration of many problems.

The first aspect is the determination of which hits on each sensitive plane make up a certain track. This part of the procedure is an art compared to the others: Many different schemes can be used which are appropriate in different environments. In most cases simple straight or curved trajectory estimates can collect track using a maximum hit distance from the estimate and other topological cuts such as the curvature and incidence. After all hits are collected into tracks combination checks must be performed with most algorithms to ensure of no double counting and merged tracks.

With the set of hits which make up a track determined, the most likely track which produced this collection can be deduced. This track will be in a general case curved (due to magnetic field) and have kinks at points (due to scattering between planes and all other material on the trajectory). The most likely track will be dictated by the spatial resolution of the hit on each detector and the scattering material between each measurement sensor. How this information is combined to form a single track varies, with each method having it's own accuracy. However, a few general principles can be conveyed which are common in all.

> Approximate all the scattering by a series of scattering points. Each point will cause the trajectory of the particle to change due to modelled scattering and can be placed at any location. The kink angle - the angle change of the particles trajectory at that point - will depend on the modelled radiation length at that point. More radiation length will make large kink angles at that point more likely in the final track

model. The modelling of the radiation length will vary from the actual radiation length at that physical point due to the granularity of the scattering points, which is a subset of all points on the trajectory. Each scattering point is used to model the radiation length of all points on the trajectory. This approximation is accurate for most mediums with just a few points, but the more inhomgeneous, the more points required for a more accurate track prediction. Radiation length is defined with relation to the density of the material and energy of the particle. The full radiation length of the particle trajectory must be calculated then split between the scattering points of the trajectory. Each point will get a different amount depending on its physical distance from all near by mass. This way of calculating radiation length is important since it deals with the non linearities within the corrected Highland formula.

With the ability to create a kinked trajectory the hit information is introduced with an estimated spatial resolution which will vary from hit to hit. This can be thought of as a measurement point attached to the measurement plane (the plane defined by the hit pixels or strips) which is part of the trajectory of the track. The measurement points will likely not form a straight line. With the scattering points a track can be formed which best lines up with the measurement points. However, considerations must be made on how large each kink can be made to bend towards each hit. Consider a system with a very large radiation length, then the most likely track is through each hit even if the hit resolution is very poor. As the radiation length decreases to zero, the kinks decrease in angle for all points and the most likely track tends to an unbroken curved trajectory. When the radiation length estimates kinks to bring the track close to the hit within the hits resolution, i.e close but not quite on the hit, the real fun begins. Now the most likely track falls in a no mans land, between the perfect curve and the perfect intersection of all hits. This track can only be determined by use of the precision of the kink angles and hit positions. This precision is determined as the inverse of the covariance matrix and must be propagated from each point to all others. This propagation will link each point with its precision to all other points under some coordinate system. This system is arbitrary, however different systems will have different precision matrices and varying degrees of complexity for propagation.

Lastly, with all the points linked together with the correct propagated

precision and hit information the most likely track can be found. It's advocated that the idea of a track model is kept in mind. So the calculated track is the most likely to have give that series of hits given the prior information of the hit position and radiation length: The track fit is as good as the information provided.

The creation of the most likely track is all well and good when you know where the hits are positioned. However in most cases life is not that simple. At a testbeam it is very difficult to place a detector in a location known to a few micron. Therefore when hits are identified on a plane, only the approximate location of the hit - to ≈ 1 mm - can be identified. The planes must be moved to their correct locations with micron precision to perform track fitting. To gain this level of precision a chicken and egg game appears: The unaligned planes with hits that produce tracks can be used to determine the misalignment, however the planes must be roughly aligned to find the legitimate hits that form a single track. This is solved by splitting the alignment into two parts: pre-alignment and the final micron level alignment. Pre-alignment corrects shifts to ≈ 0.5 mm using the correlations between hits on each plane. This brings the planes close enough to expect the pattern recognition to find the correct tracks (This of course depends on the pattern recognition and environmental factors such as the occupancy per event). The final alignment uses all the track information discussed and then adds the information on how the movement of the measurement planes will change the position of the hits. Now the most likely track can be found by moving the kinks and moving the planes (Contrast this with track fitting which only the scatterer points could move). The correct position of the planes is the one in which all tracks on average maximise their likelihood.

2.0.8 General Broken Lines (GBL) and EUTelescope

When considering the problem of track fitting the creation of measurement and scattering points is common to all techniques. However, how these points are related varies, an example would be the use of a Kalman filter. A Kalman filter is used for many different purposes outside track fitting. However this technique can be combined with a suitable jacobian to link points together under a certain coordinate system to produce an effective track fitter. The salient point to note is that the linking of points with a basic Kalman filter is usually between the position and the incidence of the hits on the measurement

planes (The curvature is also used which is a single parameter for the full track). As each parameter is changed this changes the parameters of all the other points on the trajectory. This method works and is effective, however is not using the underlying track fitting problem's symmetries. Consider how track fitting was described in section **??**, with a series of scattering points moving in space to produce a broken curve which best fitted a series of hits. The scatterers' positions are what must be determined, the incidence angles and curvature are afterthoughts even if they are of the most use for the final analysis.

The General Broken Line (GBL) algorithm does just that: It takes all the information needed expressed in some coordinate system and breaks this down to the relationship between scattering points only. When this is done a simplification of the fit is apparent, with many elements of the final matrix which must be inverted being 0, this is the symmetry present in all track fitting. The most likely position of the scattering points are then determined and this is then related back to parameters such as the incidence and curvature [**?**]. This process is mathematically equivalent to a Kalman filter but is clearly computationally different.

2.1 My contribution

A C++ implementation of GBL exists and has been integrated within EU-Telescope. The integration of GBL within EUTelescope goes beyond just a new track fitting model. It aims to allow the same procedure to fit tracks from a plethora of different DUTs, geometric setups and environmental changes. This includes complex pixel arrangements, magnetic fields and arbitrary DUT orientations. To this end many aspects of EUTelescope had to be updated to deal with these issues in a common way. This is discussed in full in the GBL section of this AIDA note [**?**].

A more practical introduction presented here is to test the fitter in different environments. One important test of a fitter is the effect pattern recognition has on a track fit. The initial pattern recognition estimates of the track should not influence the final fitted track. A way of testing this is to do an energy scan search for different combination of hits that correspond to a different energy. This is done by setting the initial "guessed" energy (E) for the pattern recognition to a particular value and then searching for tracks. Each track will only accept a hit if it falls within a particular radius (R) of it's guess. The

initial pattern recognition tracks will be different for each estimate, however the final GBL track should converge to the correct physical trajectory of the particle.

Data was taken at testbeam at DESY with a set beam energy of 5 and 3 GeV in a magnetic field of 1T. The magnetic field allows the determination of the energy of the particle using the curvature of the track. Figure **??** and **??** show the estimate energy of each track for the 5 and 3 GeV beam.

Figure 2.1: Figure **??** shows the beam energy for a range of pattern recognition inputs, with estimated beam energy 5 GeV and magnetic field 1 T. A cut of $chi2/ndf > 5$ is made to remove poor tracks for figure **??**. Figure **??** shows the chi2 distribution for each fit.

Note the beam is expected to have some natural spread which is shown. Also observe the drop in energy which is present in both from the expected energy. This is due to the passage of the particle through material which houses the magnet. The pattern recognition has no effect on the mean of the distribution and each is only wider due to larger acceptance of tracks in different regions.
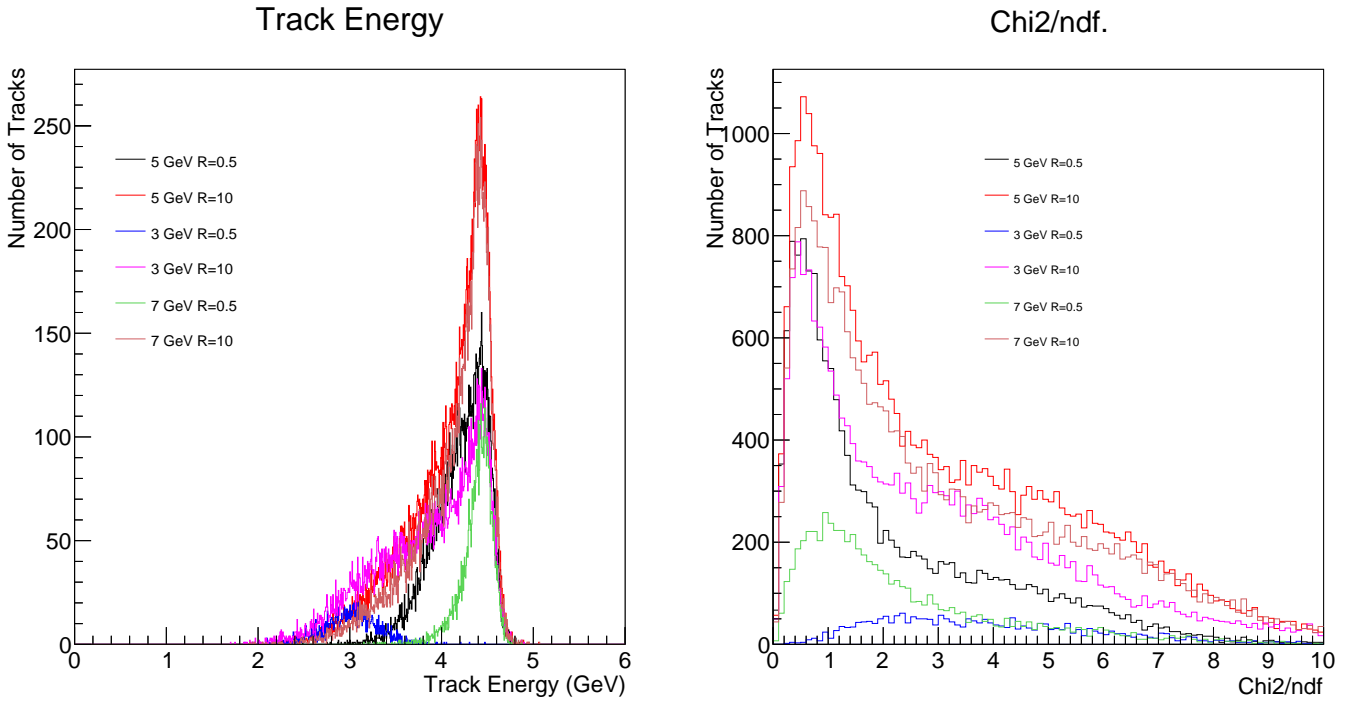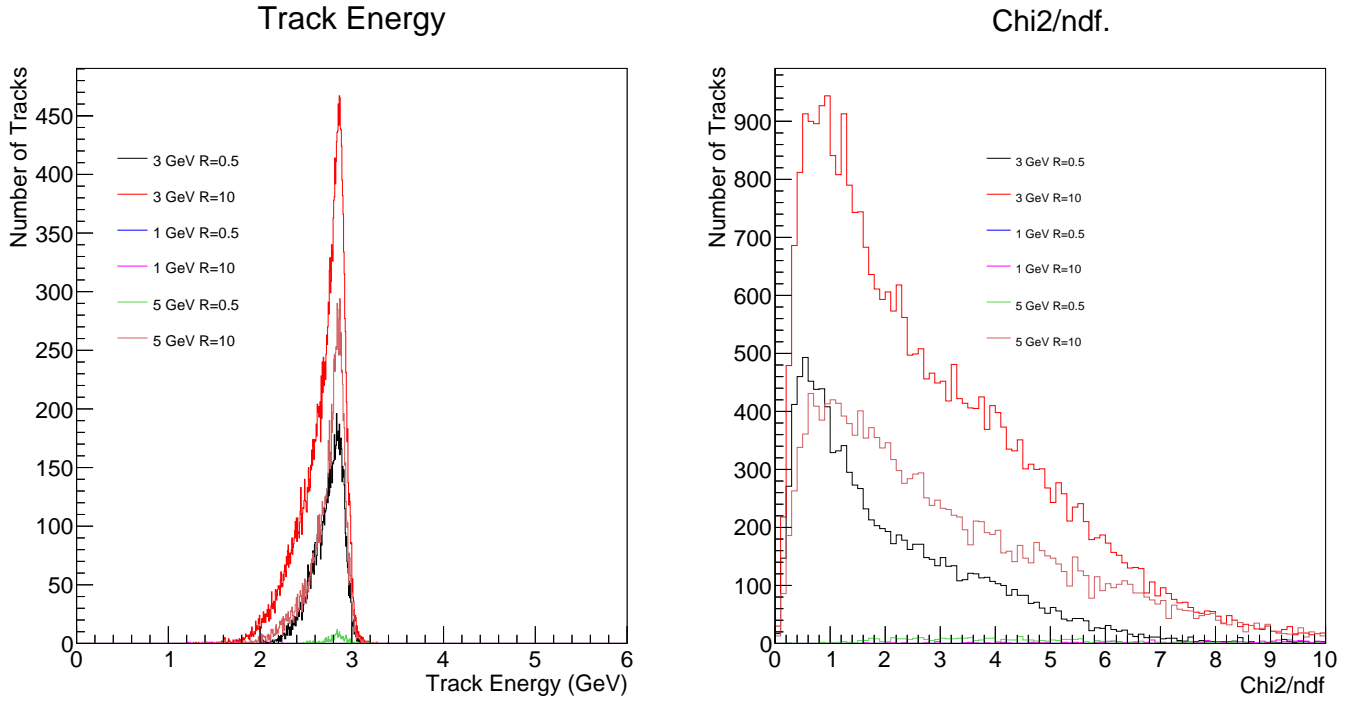
Figure 2.2: Figure **??** shows the beam energy for a range of pattern recognition inputs, with estimated beam energy 3 GeV and magnetic field 1 T. A cut of $chi2/ndf > 5$ is made to remove poor tracks from figure **??**. Figure **??** shows the chi2 distribution for each fit.

Chapter 3

Further Work

The work described below in displayed in an Gantt chart in the appendix **??**.

3.1 Alignment Studies

The alignment project work related to updating the track fitting model is almost complete. The fitter from pattern recognition to alignment has been tested with a range of beam energies and magnetic fields. Introduction of a simple 250x50 μm FEI4 pixel sensor to the setup shows results similar to what is expected. The main work still to be carried out is to extend the fitter to allow strip and more complex pixel geometries to be used with arbitrary orientations. The code base exists, however disparities between the track fitting model and the minimisation code which calculates the final plane positions must be further investigated.

The devices which will be investigated are a series of strip and pixel detectors designed as prototypes for the ATLAS upgrade. The efforts from the point of view of track fitting takes little to do with the complications of data readout to geometric hit creation. It is most concerned with maximising the precision of a track variables of interest for a realistic track model. Initial results should take a few of months. However for a perfected resolution the DUTs radiation length should be estimated. Furthermore a better estimation of the radiation length of a DUT in a setup would improve the track resolution of all future analyses. The ability to do this is already in development and only needs to be completed and tested for each DUT. This should take no more than a couple of months to complete.

Updates to the GBL fitter will allow better track estimates. However, the limited environmental situations this fitter can be applied is almost completely

determined by the pattern recognition. Limits have already been observed with geometries with low angular resolution and high occupancy. These must be tackled to created a truly generic testbeam fitter and already algorithms exist which could be easily integrated to EUTelescope to this end.

1. Complete the full body alignment for complex pixel shapes and generic initial orientations of sensors.

2. Estimate the radiation length of a series of DUTs using the kink angles after scattering.

3.2 Resonance Searches

The VH resonance search expects a result in the fall. The developed analysis code for the TMVA must be improved to include a bin by bin log likelihood estimation. This is needed for a more appropriate estimation of the significance. Further work includes addition discriminating parameters being investigated and how the TMVA performs with different preselection. Already preselection has been seen to produce significant differences in the cuts based classification. How this work progresses depends on how other areas advance. The $\sqrt{d_{12}}$ variable for example was untrimmed, so this must be improved upon before further investigation. B-tagging is another example were improvements are expected with the use of variable R jets.
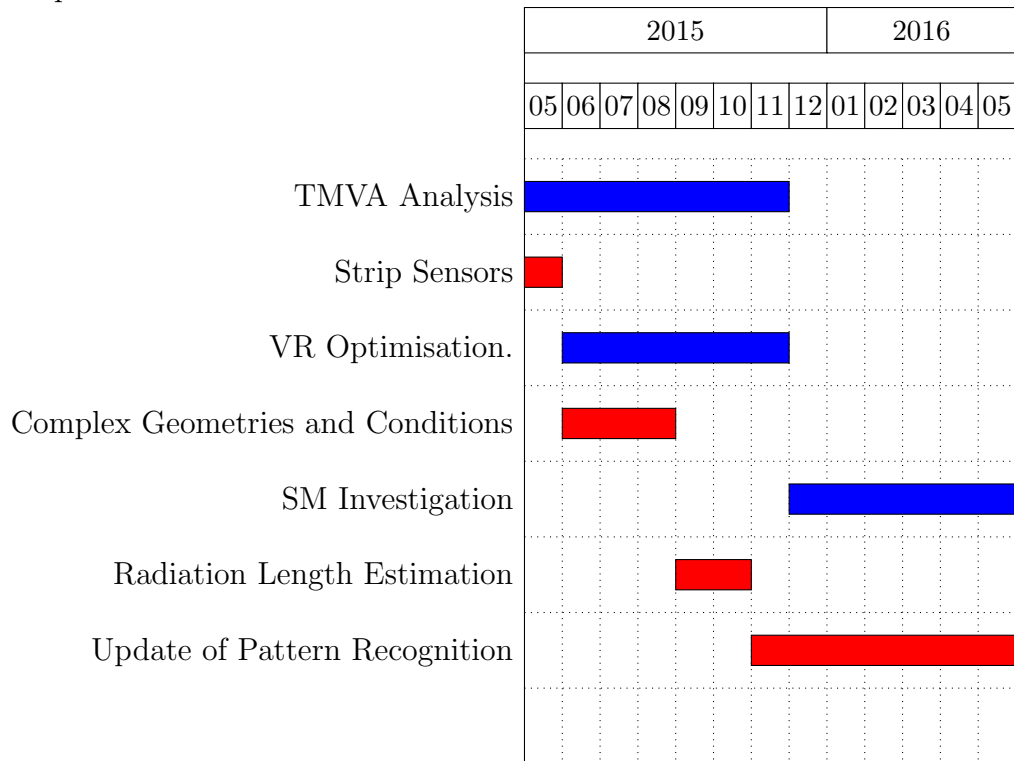
The investigation of variable R jets has begun with the Rivet analysis as shown in section ?? for standard model VH production. These results and other studies indicate that the variable R jets will improve b-tagging efficiencies. To contribute to this study further some time will be needed to learn and eventually contribute ot the DBL/exotics framework.

The aim after November will to be focus on the standard model $H \rightarrow bb$ search. This analysis will be similar in many respects to the VH resonance analysis and much of what is done there will be appied here.

1. Include CP tools to the handlers of the physics objects in the CxAOD-Framework. With a focus on the track jet objects.

2. Include the use of log likelihood functions to determine the significance of the expected signal.

3. Update the TMVA code to work with all new physics objects in the final xAOD derivation.

.0.1 Gantt Chart

Separate areas of work are split by colour. The Gantt chart assumes a split of 20/80 split towards resonace searches in terms of workload.

	2015								2016				
	05	06	07	08	09	10	11	12	01	02	03	04	05

Tasks (Gantt bars):

- **TMVA Analysis** — blue, 05 to 12 (2015)
- **Strip Sensors** — red, 05 (2015)
- **VR Optimisation.** — blue, 06 to 12 (2015)
- **Complex Geometries and Conditions** — red, 06 to 08 (2015)
- **SM Investigation** — blue, 01 to 05 (2016)
- **Radiation Length Estimation** — red, 09 to 11 (2015)
- **Update of Pattern Recognition** — red, 11 (2015) to 05 (2016)

Appendix A

Guide to Resonance Search

A.1   Introduction

This is designed to be an introduction to any analysis and will cover some of the basic ideas needed. However a focus on the the VH analysis give here

A.2   Theory and past searches

Need to be filled in after going through theory course again.

A.3   Data and MC samples

Pileup must be taken into account. This is done via reweighting. The quality of the data is validated using the GRL and DQ flags. The simulation of signal and background can be viewed in **??**.

The k-factor is a factor which relates the leading order to the error compared to the full calculation. The filter term is to relate the number of events expected to the number produced.

To begin to discuss physics objects you need to begin at the xAOD root files created by simulation and data. You can see all the collections and type in an xAOD using:

checkxAOD.py

The ATLAS Muon Combined Performance Group define muons using a series of cut points, this is done for inner detector and muon spectrometer. Both the ID and MS are used in combination to determine the pt and mass of the muon. Note the pt and resolution depend on many factors (magnetic

field, scattering, intrinsic resolution), therefore momentum scale and resolution corrections are also applied. This makes the MC and data follow the same distribution. These were determined using the J/ψ ($c\bar{c}$)/Z resonance. Reconstruction efficiency is allowed to vary with a scale factor which is the ratio of MC/data. This if course depends of pt and η.

Isolation criteria for leptons is needed since semi leptonic/hadronic decays are the main source of non-prompt/misidentified leptons. A tight cut is performed on each lepton by constructing the energy for the track/Calo object inside a R=0.2 cone and then doing a cut which depends on p_t and η. The energy must be over this cut value. A variable R cut is also performed on tracks only for the loosetrackonly requirement. This reduces the size of the cone since the Z $\to$ ll cone will reduce in size for higher pt.

Calorimeter jets are corrected by adding muon inside jet to jet. $H \to b\bar{b}$ reconstructed with 1.0 large R jet, which focuses on pt ¿ 250 GeV since less than this will be in a larger jet. Smaller R jets are used for Etmiss which must use a set eta and pt ¿ 20 GeV for JES calibration to be valid. Note the JES must be compared to MC since the real detector will have some energy loses. This is also an additional uncertainty for the final fit. To reduce jets not from the primary vertices other than hard scattering vertex JVT recommendations for pt ¡ 50 GeV smaller R jets is used. JVT is used to reduce pile up by determining the primary vertex and then determining the amount of energy which comes from this.

Overlapping jets with electrons are removed if within R ¡ 0.2

A.4   Triggers

The trigger efficiency depends on a particular sample. A sample here is a selection of cuts after requiring these triggers. The efficiency is determined by taking a selection of events from the sample and running this through the trigger. The efficiency is just the ratio before and after the trigger. Note the efficiency will depend of some observable, e.g VH mass. A scale factor must also be used to determine the different between MC and data.

A.5   Selection

The number of VHLoose lepton number is varied to keep the 0/1/2 lepton channel orthogonal.

In the two lepton channel the muon momentum resolution deteriorates at high pt so the pt is scaled by $\frac{m_{ZZ}}{m_{\mu\mu}}$.

W mass constraint used in the 1 lepton channel to determine neutrino pt.

Geometric cuts are used in the 0 lepton channel.

After these kinematic cuts then the H $\rightarrow b\bar{b}$ is exploited to increase sensitivity. This is also used to define control regions. This is done looking at the mass and number of b tags.

A.6 Multijet

A.7 Nuisance Check update (NP Pruning)

A.8 CMake Introduction)

A.9 Abbreviations

- EDM $\rightarrow$ Event Data Model

- GRL $\rightarrow$ Good Run List

- JES $\rightarrow$ Jet Energy Scale

- JVT *rightarrow* Jet Vertex Tagger

Bibliography

[1] HEP Analysis. Hep analysis. Available online at `http://www.hep.ucl.ac.uk/~campanel/Post_Grads/2013-2014/JMorris_HEPAnalysis.pdf`, 2012.

[2] atlas collaboration. technical design report. Technical report, atlas collaboration, 1999.

[3] The ATLAS Collaboration. Search for the $b\bar{b}$ decay of the standard model higgs boson in associated (w/z)h production with the atlas detector, 2010.

[4] The ATLAS Collaboration. Measurement of kt splitting scales in $w \to l\nu$ events at $\sqrt{s} = 7$ tev with the atlas detector, 2013.

[5] The ATLAS Collaboration. Search for a new resonance decaying to a higgs boson and a w or z boson in the ll/lν/$\nu\nu$ + b$\bar{b}$ final state with the atlas detector. Available online at `https://cds.cern.ch/record/1669860`, 2013.

[6] Gunter Zech Gerhard Bohm. Introduction to statistics and data analysis for physicists, 2010.

[7] Witold Skiba Gustavo Marques Tavares, Martin Schmaltz. Higgs mass naturalness and scale invariance in the uv. *arXiv:1308.0025*, 0:1–10, 2013.

[8] Claus Kleinwort. General broken lines as advanced track model, 2013.

[9] Simon Owen. Multijet background estimation for supersymmetry searches using the atlas detector at the large hadron collider, 2012.

[10] Sandro Palestini. The muon spectrometer of the atlas experiment. Available online at `http://palesti2.web.cern.ch/palesti2/Documents/SPsienaNote.pdf`, 2003.

[11] I. Rubinskiy. Eutelescope. offline track reconstruction and dut analysis software. *EUDET-Memo-2010-12*, 2010.

[12] Igor Rubinskiy Tobias Bisanz, Alexander Morton. Eutelescope 1.0: Reconstruction software for the aida testbeam telescope, 2015.

[13] A Hoecker P Speckmayer J Stelzer J Therhaag E von Toerne H Voss. Toolkit for multivariate data analysis with root, 2013.

[14] On behalf of the ATLAS Collaboration Will Buttinger. The atlas level-1 trigger system. Journal of Physics: Conference Series 396, 2012.

[15] Katharina Behra Cigdem Issevera Lucy Kogana Jiahang Zhonga. Performance and calibration 1 of jets with variable r, 2014.