Chapter 1

A measurement of jet substructure observables on $Z + b\bar{b}$ final states

(Lund) Jet fuel can't melt steel beams.

– George W. Bush

The main focus of this thesis is a measurement of the Lund Jet Plane and other substructure observables on $Z + b\bar{b}$ final states. The analysis focuses on boosted and resolved topologies. For comparison, final states containing $Z + jj$, where j indicates a light-flavour jet, are also considered. This chapter will introduce the measurement in detail. First, a description of the observables measured will be given. Subsequently, a details review of selections, final states, data and simulation samples utilised will be given.

1.1 Motivation

This analysis aims to study the properties of b-jets. Currently, little is known about the radiation pattern within these jets, Talk about importance of jet substructure, why we want to understand b-jets, uses in tagging, background in searches or for higgs, precision physics, better mc modelling, etc.

1.2 Observables of interest

1.2.1 The primary Lund Jet Plane

The primary Lund Jet Plane is the main observable which this analysis focuses on. The observable is particularly interesting due to a number of neat properties it possesses.

As the name would suggest, the Lund Jet Plane is tightly linked to parton shower development. The precursors to the Lund Plane were Lund diagrams, used to describe the kinematic distribution of radiation from a QCD dipole

within a y-k_t plane [**?**]. The kinematically allowed region of phase space forms a triangle bounded by $|y| < \ln(\sqrt{s}/k_t)$, leading to the plane's characteristic shape.

As the field of jet substructure began to take hold, the Lund diagram provided a way to visualise the effect of various techniques used to simplify analytic calculations, such as *trimming*, *pruning* and *grooming* [**?**, **?**]. The description was later adapted for jets in [**?**]. In this work, the authors used the plane to create a two-dimensional representation of the clustering history, and therefore structure, of a jet.

The primary Lund Jet Plane is formed starting from an anti-k_t clustered jet, which is then reclustered using the Cambridge-Aachen (C/A) algorithm [**?**]. After reclustering, the jet is de-clustered, working backwards through the C/A algorithm following the hardest branch at each splitting, to obtain a list of the branchings from the base of the jet. The variables $\ln(k_t/\text{GeV})$ and $\ln(1/\Delta)$ where

$$\begin{cases} \Delta^2 = (y_a - y_b)^2 + (\phi_a - \phi_b)^2 \\ k_t = p_{Tb}\Delta \end{cases} \tag{1.1}$$

are plotted for each splitting. Here, a and b refer to the two branches after the splitting, with $p_{Ta} > p_{Tb}$. In the limit $p_{Ta} \gg p_{Tb}$ and $\Delta \ll 1$, k_t refers to the transverse momentum of subjet b with respect to the emitter $a + b$.

The lack of momentum weighting which characterises the C/A algorithm allows for the algorithm to mimic the angular ordering of $1 \to 2$ splittings in QCD. Anti-k_t jets are not suitable as they are lacking this key angular-ordering which mimics the perturbative evolution of the final state and which allows us, in some sense, to work "backwards" through the shower. As each emission is plotted in the Lund Plane, the evolution through the parton shower down to the hadronisation scale emerges, as each QCD emission is found at larger angles and lower k_t than the last.

Figure 1.1 shows the various kinematic regions identifiable within the primary LJP. Towards the left edge of the plane, an area characterised by large Δ initial state radiation is found. A hard-collinear region where the two splittings have roughly equal p_T follows the diagonal edge of the triangle. This edge, in fact, is due to the kinematic limit $k_t < \frac{1}{2}p_{T,jet}\Delta$. In the bottom left corner, there is a region which originates from multiparton interactions and the underlying event, whose emissions within the jet are characterised by a large angle and small k_t. A non-perturbative region is found at small k_t, and lastly soft-collinear radiation falls towards the centre of the plane.

For the sake of completeness, we would like to mention that a splitting is characterised not just by the (Δ, k_t) coordinates, but also by

$$\begin{cases} z = \frac{p_{Tb}}{p_{Ta}+p_{Tb}} \\ m^2 = (p_a + p_b)^2 \\ \psi = \tan^{-1}\frac{y_b - y_a}{\phi_b - \phi_a}. \end{cases} \tag{1.2}$$

There exist also alternative representations of the Lund Plane, where instead of $\ln(k_t/\text{GeV})$, $\ln(1/z)$ is shown. These other Lund coordinates have been used to train graph neural network based taggers, such as LundNet [**?**].

It is also possible to define Lund Planes beyond the primary, by considering further splittings off of the emitted subjets when present. This leads to the

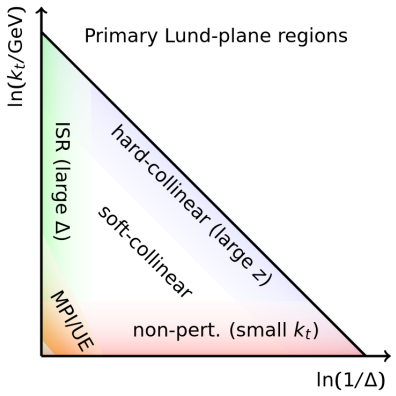

Figure 1.1: The kinematic regions within the primary Lund Jet Plane [**?**].

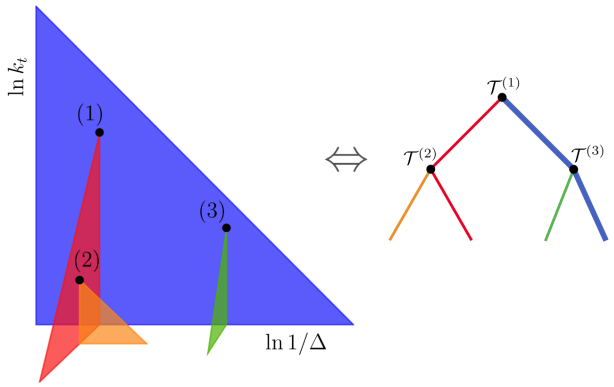

Figure 1.2: The primary Lund Plane for a hypothetical jet with emissions (1) and (3), and the secondary Lund Plane (2) from branch (1) and corresponding tree structure of the jet showing this evolution. Each splitting can be charachterised by the set of Lund variables $\mathcal{T}$, given by Eq. 1.1 and Eq. 1.2 [**?**].

definition of secondary planes, tertiary planes, etc. This process is shown in Figure 1.2.

The Lund Plane has been measured on inclusive jets by the ATLAS, CMS and ALICE collaborations [**?**, **?**, **?**]. ATLAS has additionally released measurements for the Lund Plane for top and W jets [**?**] and a measurement on the Lund subjet multiplicity [**?**].

1.2.2 Modified jet angularities

The generalised angularities [**?**, **?**, **?**], henceforth referred to as just *angularities*, are a class of observables designed for quark/gluon jet discrimination. They are defined as

$$\lambda_\alpha^\kappa = \sum_i \left(\frac{p_{Ti}}{p_{Tjet}}\right)^\kappa \left(\frac{\Delta R_i}{R_0}\right)^\alpha \tag{1.3}$$

where the sum runs over all constituents of an anti-k_t jet of radius R_0, and ΔR_i indicates the distance between the i-th constituent and the jet axis. λ_α^κ is IR-safe for $\kappa \geq 1$, and we shall assume $\kappa = 1$ going forward, unless otherwise specified.

A few angularities are distinguished enough to have their own name. These include λ_1, known as the *jet width* and λ_2, known as the *jet thrust*, amongst others. The idea behind these and other angularities is to emphasise various aspects of a jet's substructure in order to highlight features typical of quark or gluon jets. By tuning the β parameter, one can accentuate either wide-angle or collinear radiation, while tuning the κ parameter has the same effect on harder or softer hadrons. Ultimately, the difference between quark and gluon jets originates from the fact that $C_A > C_F$, implying that gluons are more likely to emit soft and collinear gluons than quarks.

The phenomenology of such observables has been extensively studied at colliders for processes such as Z+jets [**?**] at the LHC, and they have been measured in proton-proton collisions by the ATLAS, CMS and ALICE collaborations on a variety of energies and final states [**?**, **?**, **?**, **?**].

Recently, a proposal for modified angularities has been put forth [**?**].

The variants are defined by reclustering a jet with the C/A algorithm using the winner-takes-all (WTA) scheme. Two vectors are defined:

$$n_0 = (\cosh y, \cos\phi, \sin\phi, \sinh y) \tag{1.4}$$

$$n = \left(\frac{m_{T,n}}{p_{T,n}}\cosh y, \cos\phi, \sin\phi, \frac{m_{T,n}}{p_{T,n}}\sinh y\right) \tag{1.5}$$

where ϕ and y are the azimuthal angle and rapidity of the WTA axis, and $m_{T,n} = p_{T,n} + m_n$ is the transverse mass of the particle aligned with the WTA axis. The (E, p_x, p_y, p_z) convention is used to define the vectors.

The variants are then defined as

$$\dot{\lambda}_0^\alpha = \sum_i \frac{p_{T,i}}{p_T}\left(\frac{2p_i\cdot n_0}{p_{T,i}R^2}\right)^{\frac{\alpha}{2}} \qquad \dot{\lambda}^\alpha = \sum_i \frac{p_{T,i}}{p_T}\left(\frac{2p_i\cdot n}{p_{T,i}R^2}\right)^{\frac{\alpha}{2}} \tag{1.6}$$

$$\hat{\lambda}_0^\alpha = \sum_{i\neq n} \frac{p_{T,i}}{p_T}\left(\frac{2p_i\cdot n_0}{p_{T,i}R^2}\right)^{\frac{\alpha}{2}} \qquad \hat{\lambda}^\alpha = \sum_{i\neq n} \frac{p_{T,i}}{p_T}\left(\frac{2p_i\cdot n}{p_{T,i}R^2}\right)^{\frac{\alpha}{2}}. \tag{1.7}$$

where the sums in Eqs. 1.7 excludes the particle aligned with the WTA axis. These variants are specifically designed to bring to light the mass of the particle aligned with the WTA-axis. Considering jets seeded by heavy flavour partons and suppress the decay of the corresponding hadrons, the particle aligned with the WTA axis will be a hadron containing the same flavour of quark which seeded the jet. As such, the observables defined in Eq. 1.6 will contain an *explicit mass dependence*. In the quasi-collinear limit, for example,

$$\dot{\lambda}_0^\alpha \approx \sum_{i\neq n} \frac{p_{T,i}}{p_T}\left(\frac{m_i^2}{p_{T,i}^2R^2} + \frac{\Delta R_i^2}{R^2}\right)^{\frac{\alpha}{2}} + \left(\frac{m_n^2}{p_{T,n}^2R^2}\right)^{\frac{\alpha}{2}}. \tag{1.8}$$

The end result is a set of observables (Eq. 1.7) sensitive to quark mass.

Show plots from paper??

1.2.3 Other variables of interest

Colour ring, D2,boosted kinematic variables, etc.

1.3 Object selections and event definitions

1.3.1 Data Pre-selection

This analysis makes use of data recorded with the ATLAS experiment during Run 2, corresponding to the years 2015-2018. The events consist of pp collisions at $\sqrt{s}$ =13 TeV during stable beam conditions with all subsystems of the detector operational and passing all data quality requirements. Overall, the data amounts to 139 fb^{-1}.

1.3.2 Definition of signal regions

As mentioned, the analysis aims to understand the substructure of b-jets and does so by making comparisons to light jets. To this aim, the signal corresponds to Z-boson production in association with two jets, either both b-tagged or both anti-b-tagged. Leptonic decays of the Z-boson are considered, either to an e^+e^- or $\mu^+\mu^-$ pair.

Two topologies are considered: a resolved topology where the jets are sufficiently separated to be identified as two separate anti-k_t jets of radius $R = 0.4$. In this case, PFlow jets are considered and the substructure observables are measured on these jets. A boosted topology is also considered, where, in this case, the jets are produced at high p_T and as a result are too close together to be fully resolved. In this case, two VR-trackjets, both b- or anti-b- tagged, are required to lie within a large-R LCTopoJets of radius $R = 1.0$, which is taken to be the signal jet for the measurement. Overall, we are left with four signal regions: resolved 2B, resolved 0B, boosted 2B, and boosted 0B.

1.3.3 Lepton selections

Electron candidates are identified from energy clusters in the electromagnetic calorimeter that are linked to reconstructed tracks in the Inner Detector. Candidates must be found within the calorimeter, i.e. within $|\eta| < 2.47$, excluding the transition region between 1.37 ¡ $|\eta|$ 1.52. In order to qualify as electrons, they must meet the TightLH identification criterion and satisfy additional requirements on their transverse and longitudinal impact parameters: $|z_0 \sin(\theta)| < 0.5$ mm and $|d_0/\sigma(d_0)| < 5.0$. To further minimize background from non-prompt electrons, conversions, and hadrons, candidates are required to be meet isolation criteria and must pass the Tight_VarRad isolation working point. This equates to having both the calorimeter energy sum and p_T sum of the tracks within a variable radius cone (up to a maximum radius of $\Delta R = 0.2$) centred around the electron less than $0.06E_T$, where E_T indicates the transverse energy. Lastly, electrons must meet specific criteria regarding the shape of the electromagnetic shower, track quality, and track alignment with respect to the calorimeter cluster. A likelihood based method with a "Tight" working point is utilised to satisfy these selections.

Muon candidates are identified by correlating tracks found within the inner detector to tracks or track segments in the muon spectrometer. They must meet impact parameter criteria of $|z_0 \sin(\theta)| < 0.5$ mm and $|d_0/\sigma(d_0)| < 3.0$ and pass the "Medium" identification working point, based on quality requirements on the tracks in the inner detector and the muon spectrometer. In addition, muons must pass the "FCTight FixedRad" working point to pass isolation requirements. The p_T sum of tracks in a variable radius cone (maximum radius $\Delta R = 0.3$) centred around the muon must be no more than 0.04 times the muon p_T. Lastly, muons must be found at pseudorapidities $|\eta| < 2.5$.

Both electrons and muons must have a $p_T < 27$ GeV.

1.3.4 Jet selections

Resolved SRs

In the resolved signal regions, jets are clustered using the anti-k_t algorithm as implemented in the FastJet package with PFlow objects are constituents. Jets in these signal regions must satisfy $p_T > 20$ GeV and $|y| < 2.5$. The jet vertex tagger (JVT) discriminant is used to reject jets originating from pileup vertices, with a "Tight" working point, corresponding to a JVT value above 0.5 for jets in the p_T range $20\,\text{GeV} < p_T < 60\,\text{GeV}$ and with pseudorapidity $|\eta| < 2.4$.

Jets in the resolved signal regions are (anti) b-tagged using the DL1r b-tagging discriminant. Jets are considered (anti) b-tagged if they (fail) pass the 70% efficiency working point, corresponding to 70% efficiency in identifying jets containing b-hadrons. This working point is also characterised by a c-jet rejection rate of 10, and a light jet rejection rate of 417, indicating that 1/10 or 1/417 jets will be mistagged, respectively.

Boosted SRs

Two jet collections are utilised in the boosted regimes. Large-R jets of radius $R = 1.0$, clustered with the anti-k_t algorithm from the FastJet package. These jets are clustered with locally-calibrated topological (LCTopo) clusters of energy from the hadronic calorimeter and are trimmed with parameters $R_{\text{sub}} = 0.2$ and $f_{\text{cut}} = 0.05$. Large-R jets are tagged through the use of VR-subjets matched to the jet of interest.

1.3.5 Track selections

1.3.6 Event selection

1.4 Lund plane (and angularities) reconstruction

1.5 Flavor fit

1.6 bkg characterisation/ttbar rejection

1.7 B hadron reconstruction