Double B-hadron Jet Tagging and Identification of Gluon to $b\bar{b}$ jets with the ATLAS Detector

Lic. María Laura González Silva

Tesis Doctoral en Ciencias Físicas

Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Noviembre 2012

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

Double B-hadron Jet Tagging and Identification of Gluon to $b\bar{b}$ jets with the ATLAS Detector

Trabajo de Tesis para optar por el título de

Doctor de la Universidad de Buenos Aires en el área Ciencias Físicas

por **María Laura González Silva**

Director de Tesis: Dr. Ricardo Piegaia

Consejero de estudios: Dr. Daniel Deflorian

Lugar de Trabajo: Departamento de Física (CONICET-UBA)

Buenos Aires, 2012

AGRADECIMIENTOS

Agradezco a...

Abstract

This thesis describes a method that allows the identification of double B-hadron jets originating from gluon-splitting. The technique exploits the kinematic differences between the so called "merged" jets and single B-hadron jets using track-based jet shape and jet substructure variables combined in a multivariate likelihood analysis. The ability to reject b-jets from gluon splitting is important to reduce and to improve the estimation of the b-tag background in Standard Model analyses and in new physics searches involving b-jets in the final state. In the simulation, the algorithm rejects 95% (50%) of merged B-hadron jets while retaining 50% (90%) of the tagged b-jets, although the exact values depend on the jet p_T.

Contents

Chapter 1

Introduction

The first years of proton-proton collisions at a centre of mass energy of 7 TeV delivered by the Large Hadron Collider and recorded by the ATLAS experiment have provided data to explore quantum chromodynamics (QCD) at scales never reached before. Precision measurements of strong interactions are interesting in their own right, but, in addition, QCD provides one of the main backgrounds to many New Physics measurements; furthermore, it is also through tests of QCD that New Physics may be discovered. Hadronic jets are a fundamental ingredient for precision tests of QCD: understanding and measuring their performance is crucial in the LHC environment.

A wide range of physics signatures, within the Standard Model predictions (SM) and Beyond the Standard Model (BSM), contain jets originating from bottom (b) quarks. The ability to identify jets containing b-hadrons[1] is therefore important for the high-p_T physics program of the ATLAS experiment. b-tagging algorithms rely on the relatively long decay length of

[1]Due to QCD confinement the experimental signature of quarks and gluons are not the quarks and gluons themselves but a spray of "colorless" hadrons. In the case of the b-quarks, the so called b-hadrons are observed.

b-hadrons that gives rise to large impact parameter tracks and displaced decay secondary vertices; or on the presence of a soft lepton within the jet, the product of the semileptonic b-decay. These algorithms, however, do not provide information on the number of b-hadrons within the jet. In particular, they tag jets containing a $b\bar{b}$ pair, with no net heavy flavour, which do not correspond to the intuitive picture of a b-jet as a jet containing a single b-quark or antiquark.

b-jets containing two b-hadrons, henceforth called "merged", and single b-jet, the jets containing only one b-hadron, are depicted in Fig. 1.1.

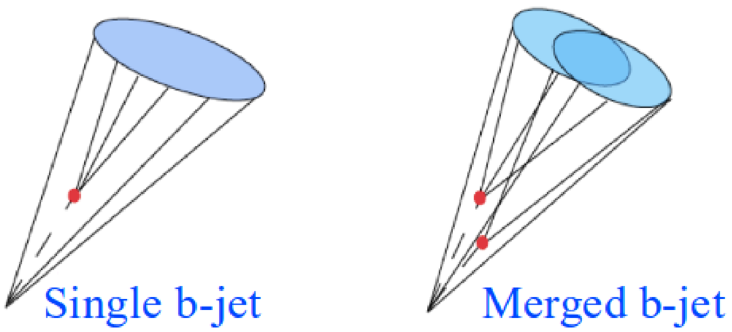

Figure 1.1: b-tagging algorithms select jets originating both from the fragmentation of single b-quark ("single" b-jet, left image) or from the splitting of a gluon into a pair of close-by $b\bar{b}$ quarks ("merged" b-jets, right image).

The ability to single out merged b-jets expected to be produced in QCD mainly by gluon splitting, has several applications. Here we briefly discuss two cases, the measurement of QCD b-quark production and the reduction of background in SM and Beyond the Standard Model (BSM) analyses.

The measurement of the inclusive b-jet spectrum

Studies of QCD bottom production are important because of the correspondence between parton level production and the observed hadron level, and their potential to provide information on the b-quark parton distribution function, a component of the proton structure thought to be generated entirely perturbatively from the QCD evolution equations of the other flavours. The theoretical calculation of the inclusive b-jet spectrum presents rather important uncertainties ($\sim 50\%$), considerably larger than those for the light jet inclusive spectrum ($\sim 10-20\%$) [1]. These arise from the poor convergence of the perturbative series, as evidenced by the large value of the K-factor (NLO/LO), $K = 6 - 10$, in the p_T range covered by the LHC. While at LO only the so-called "flavour creation" channel is present, at NLO two new channels open up, often referred to as "flavour excitation" (FEX) and "gluon splitting" (GSP), see Fig. 2.1. NLO effects are included approximately in LO parton showering models, such as Herwig or Pythia. The various channels can be approximately separated in a parton shower Monte Carlo generator such as these, where one can determine the underlying hard process from the event record. It is found that the LO channel has a much smaller contribution than the FEX and the GSP channels, which receive strong enhancement from collinear logarithms [?].

Ref. [2] proposes a new observable to free the heavy-flavour spectrum calculation from collinear logarithms, and improve the accuracy of the theoretical prediction, by not including in the production cross-section the contribution from double b-hadron jets. Final-state logarithms are removed by employing a recently developed jet reconstruction scheme, the flavour-k_t algorithm [3], which maintains the correspondence between partonic flavour and jet flavour. Specifically, jets containing a b-quark and a b-antiquark, which in a parton shower MC generator are produced $\sim 95\%$ of the time by the GSP channel, are labeled in an IR-safe way as light jets and removed from the b-jet spectrum. The initial-state (FEX) collinear logarithms can be resummed by using a b-quark parton distribution functions. With this algorithm the K-factor for the differential heavy-jet spectrum cross-section is shown not to exceed a value of $K = 1.4$, with a factor of four reduction in the theoretical (scale variation) uncertainties.

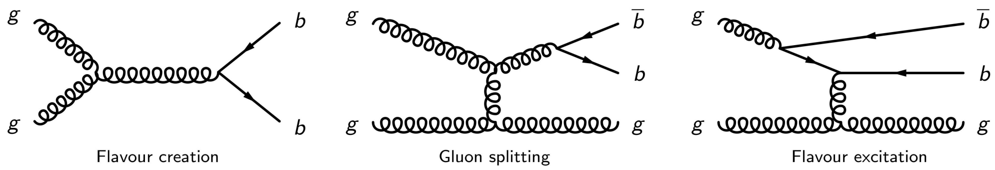

Figure 1.2: Representative diagrams of the three channels contributing to QCD b-quark production up to NLO. The flavour creation channel (left) is the only one present at LO. At NLO, two new channels open up, referred to as gluon splitting (center) and flavour excitation (right).

Rejection of background in Standard Model analyses and beyond-SM searches

Efficient tagging of merged b-jets can provide an important handle to understand, estimate and/or reject b-tagged backgrounds to SM and new physics searches at the LHC.

SM physics analyses that rely on the presence of single b-jets in the final state, such as top quark physics (either in the $t\bar{t}$ or the single top channels) or associated Higgs production ($WH \to \ell\nu b\bar{b}$ and $ZH \to \nu\nu b\bar{b}$), suffer from backgrounds that can be in part removed by a merged b-jet tagger. These are the reducible background from QCD, which can produce double b-hadron jets as discussed in the previous subsection, and the irreducible background due to W bosons plus b-jets. While at LO only single b-jets are present in $W + b$ production, at NLO jets containing two b-hadrons are expected due to the contribution of diagrams containing a $gb\bar{b}$ vertex. The relevance of merged b-jets in this channel is supported by NLO calculations of the production of W bosons and two jets with at least one b-quark at the LHC [4]. For jets with $p_T > 25$ GeV, and $|\eta| < 2.5$, they indicate that the cross section for $W(b\bar{b})j$ is almost a factor of two higher than $Wb\bar{b}$, where $(b\bar{b})$ denotes the case in which the two b-quarks are merged into the same jet.

Jets containing a single b-quark or antiquark also enter in many BSM collider searches, notably because b-quarks are produced in the decays both of heavy SM particles, top quarks, the Z boson and the Higgs boson, and of particles appearing in proposed extensions of the SM. The ability to distinguish single b-jets from jets containing two b-hadrons is thus here of wide application to reduce SM backgrounds giving rise to close-by $b\bar{b}$ pairs.

The average multiplicity of $b\bar{b}$ pairs from gluons in hadronic Z^0 decays was first measured by DELPHI experiment at LEP. The analysis was performed by looking for secondary vertex production in 4-jet events. The average rate was found to be (0.21±0.11(stat.)±0.09(syst.))% [5].

The rate of gluon splitting into bottom quarks, $g \to b\bar{b}$, was also measured by the Stanford Linear Collider (SLC) Large Detector (SLD), in hadronic Z^0 decays collected from 1996 to 1998 [6]. Following a similar procedure, the rate per hadronic event was found to be $(3.7\pm0.71(\text{stat.})\pm0.66(\text{syst.}))\times10^{-3}$.

The probability for secondary production of a bottom quark pair from a gluon per hadronic Z^0 decay in e^+e^- annihilation,

$$g_{b\bar{b}} = \frac{N(Z^0 \to q\bar{q}g, g\bar{b})}{N(Z^0 \to hadrons)} \tag{1.1}$$

is expected to be very small, since the gluon must have sufficient energy to produce the bottom quark pair.

At the Tevatron, on the other hand, 50% of the B hadrons are due to the gluon splitting process, and a larger fraction is expected to contribute at the LHC.

There are two possible strategies to attempt to identify b-jets containing two b-hadrons in hadronic collisions. One of them relies on the direct reconstruction of the two b-decay secondary vertices [7]. This allows the measurement of the angular separation between the b-hadrons, but suffers from the low efficiency of a double b-tag requirement plus additional reconstruction inefficiencies at small angular separation between the two b-hadrons. In this thesis we develop an alternative method that does not rely on explicit vertex finding, but exploits the substructure differences between single and merged b-jets, combining them in a multivariate analysis.

Chapter 2 describes the theoretical framework, with emphasis in the theory of the strong interactions and the aspects that are important for the understanding of the hadronic final state in hadronic collisions. The LHC and the ATLAS detector components are described in Chapter **??**, together with a summarization of the detector conditions during 2011 data taking. Chapter **??** details how jet reconstruction and calibration are performed at ATLAS and describes the procedure for the identification of b-quark jets. Chapter **??** presents the analysis of jet shape and substructure variables for the discrimination between single and double b-hadron jets. The validation of the variables in 2011 data is also included. The construction of the multivariate discriminator and the discussion of the systematic uncertainties are presented in Chapter **??**.

TO DO Preliminary results for the measurement of the fraction of double b-hadron jets in data are discussed in Chapter **??**.

Finally, chapter **??** summarizes the results.

Chapter 2

Theoretical framework

In this chapter a short overview of the theory of elementary particles and fundamental interactions is presented, with emphasis on the strong interactions and the description of the hadronic final state in hadron collisions.

2.1 The Standard Model

The Standard Model (SM) is a quantum field theory that describes the behavior of all experimentally-observed particles under the influence of the electromagnetic, weak and strong forces[1]. In this model, all forces of nature are the result of particle exchange. The force mediators interact on the particles of matter, and, in some cases, due to the non-Abelian character of the theory[2], with each other.

The fundamental building blocks of matter predicted by the theory are

[1]In principle gravitational forces should also be included in the list of fundamental interactions but their impact is fortunately negligible at the distance and energy scales usually considered in particle physics experiments.

[2]The transformations of the symmetry group do not commute in the case of the QCD and weak groups.

fermions with spin 1/2:

- six leptons (and their antiparticles), organized in three families,

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix} \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix} \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}$$

- and six quarks (and their antiparticles), organized in three families,

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}$$

.

The six types of quark are also known as the six quark flavors. Collectively, the u (up), d (down), and s (strange) quarks are frequently referred to as the light quarks. The heaviest quark of the Standard Model, the quark t (top), was the last to be found [8] [9]. These particles are considered point-like, as there is no evidence of any internal structure of leptons or quarks to date.

In addition, the model contains the vector bosons which are the carriers of the the fundamental forces:

- a gauge boson for the electromagnetic interactions, the photon γ;
- three gauges bosons for the weak interactions, $W^\pm$ and Z^0;
- eight gauge bosons for the strong interactions, called gluons.

The exact symmetry (see below) of the SM predicts massless particles. One possible mechanism for breaking this symmetry is the existence of a massive scalar Higgs field that has non-zero vacuum expectation value [10]. Very recently, a Higgs-like particle was discovered by ATLAS and CMS experiments at the LHC [11, 12]. This scalar boson completes the table of Standard Model particles.

Gauge invariance, defined as the invariance of the theory under local transformations, is a fundamental property of the SM. In the theory, the electromagnetism (Quantum electrodynamics), the weak interaction, and the strong nuclear force (Quantum Chromodynamics) are all derived from imposing Lorentz invariant symmetries onto the interacting fields.

The theory of Quantum electrodynamics (QED) describes the interaction of charged particles via the exchange of one (or more) photon. It is formulated by imposing a $U(1)$ or rotational symmetry onto the simplest field lagrangian that obeys the correct equation of motion. The full theory of QED was developed by Richard Feynman and others throughout the 1940s [13]. The structure of the SM is in a sense a generalisation of this theory, extending the gauge invariance of electrodynamics to a larger set of conserved currents and charges.

The symmetry associated to the weak interaction is the $SU(2)$ symmetry, which corresponds to rotations of 2-dimensional vectors. The latter combines with the $U(1)$ symmetry from QED to produce additional gauge fields. The gauge fields merge with the gauge field from QED to form W^+, W^- and Z^0 bosons that are the carriers of the weak force. Unlike the photon, which is massless, the $W^\pm$ and Z^0 bosons have masses close to 80 and 90 GeV, respectively [14] [15]. Due to these large masses, the weak force has a short range and is feeble at low energies. At masses higher than the Z mass, the electromagnetic and weak forces unify into a single force, known as the electroweak force [16] [17] [18].

The current theoretical theory of the strong interactions began with the identification of the elementary fermions that make up the proton and other hadrons. In 1963, Gell-Mann and Zweig propose the quark model [19] [**?**] [20], which asserts that these particles are in fact composites of smaller constituents. Mesons were expected to be quark-antiquark bound states and Baryons were interpreted as bound states of three quarks, all with fractional charges. The quark model was formalized into the theory of Quantum Chromodynamics (QCD) by Harold Fritzsch and Murray Gell-Mann [?] [21] in 1973, who proposed that quarks carried an additional quantum number called color. Whitout color charge, it would seem that the quarks inside some hadrons exist in identical quantum states, in violation of the Pauli exclusion principle (this was indeed the problem of the quark model as proposed by Gell-Mann and Zweig). The color theory extends the electroweak Lagrangian to be symmetric under $SU(3)$ transformations, which introduces eight new physical gauge fields, the gluons. Due to Richard Feynman's parton model nomenclature [22], both quarks and gluons are commonly referred to as partons.

Another problem of the quark model was that free particles with fractional charges were never found. The answer to why we never see free quarks or gluons outside of a hadron, together with the tools for performing theoretical calculations in QCD are given the next sections.

2.2 Perturbative QCD

As described above, the fundamental actors of the theory of the strong interactions are quarks and gluons or, collectively, partons. Partons are confined in hadron like particles, like the proton, but, act free at sufficiently small scales. This behaviour is called asymptotic freedom. The essence of asymptotic freedom is that the strong force couples particles together more strongly as the distance between them increases. The experimental consequence of asymptotic freedom is that quarks and gluons require interactions with high energy probes to be ejected from nucleons, and they cannot be observed directly.

First indications of the presence of quarks resulted from the measurement of deep inelastic lepton-hadron scattering. The momentum transfer, Q^2, between the probe particles (leptons) and the target hadron is analogous to the distance scale within the hadron being measured. The variation of the strength of the coupling with the energy is referred to as the "running" of coupling constant.

The low value of the strong coupling constant at high-energies permits the use of perturbative techniques to calculate physical processes. As one goes higher in the perturbative expansion, each term contains an additional factor of the coupling constant, α_s. Since α_s depends on the energy, it must be evaluated at some energy scale, close to the energy scale involved in the process. For instance, at an energy of 15 GeV, the strong coupling constant takes on an approximate value of 0.1, thus, from an expansion of an infinite number of terms, only a few need to be computed. The complexity of the process determines the precision of the calculation that can be permormed. For example, predictions for the cross section for events with three partons in the final state are only available up to leading-order (LO). For inclusive parton production, calculations are typically performed at next-to-leading order (NLO). Feynman diagrams are used in the computation of the multiple terms in the expansion, they are graphical representations of each term. Example of Feynman diagrams for QCD b-quark production up to NLO are shown intance Fig. 2.1.

With this formalism, the cross-section for the interaction of partons can be computed up to fixed-order in perturbation theory. In the case of hadron colliders, such as the LHC, is the factorization theorem [23] that allows the

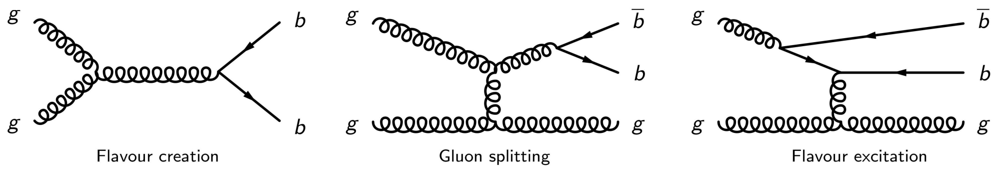

Figure 2.1: Representative diagrams of the three channels contributing to QCD b-quark production up to NLO.

perturbative calculations for parton interactions to be extended to proton-proton collisions.

In the simplest picture, a proton is a combination of three quarks: two up quarks and one down quark. The reality, however, is much more complex. In the proton there are gluons constantly being emitted and absorbed, causing quark/antiquark pairs of many flavors to be briefly produced and destroyed. The up and down quarks of the standard hadron model are called valence quarks, while the virtual quark/antiquark pairs are known as sea quarks. Both valence quarks and sea quarks, along with the gluons, share the total momentum of the hadron. The distribution of the momentum fraction, x, carried by each parton is expressed as a probability to find a particular parton with a given x. The latter is known as the Parton Distribution Function (PDF).

For cross-section calculations, the PDFs are evaluated at a factorization scale, μ_f, which can be thought of as the scale that separates short-distance, perturbative physics from long-distance, non perturbative physics. Any variation in the computed cross sections due to different choises of the energy scales can be interpreted as an uncertainty due to the unknown higher-order corrections in the cross-section calculation.

The evolution of the PDF on Q is given by the DGLAP equations, published separately in the 1970s by Yuri Dokshitzer, Vladimir Gribov and Lev

Lipatov, and Guido Altarelli and Giorgo Parisi [24]. The DGLAP equations are derived by noting that the PDFs should be independent of the factorization scale f . This gives a precise mathematical form to the dependence. The dependence on x, however, must be obtained by fitting possible cross section predictions to data from hard scattering experiments.

2.3 Jet physics

Due to confinement the experimental signature of quarks and gluons are the final state "coloreless" hadrons[3]. The packet of particles produced tends to travel collinearly with the direction of the initiator quark or gluon. The result is a collimated "spray" of hadrons (also photons and leptons) entering the detector in place of the original parton; these clusters of objects are what we define as jets. The first evidence for jet production was observed in e^+e^- collisions at the SPEAR storage ring at SLAC in 1975 [25].

The evolution from a single parton to an ensemble of hadrons accurs through the processes of parton showering and hadronization. Since the strong coupling constant grows with increasing distance between color charges, a strong color potential forms as the parton from the "hard" (high Q^2) scattering process separates from the original hadron. This large potential causes quark/antiquark pairs ($q\bar{q}$) to be created, each carrying some of the energy and momentum of the original partons. As these new partons move away from one another, yet more color potentials are formed, and the process repeats. Thus from one parton a shower of partons appears, traveling along the same direction as the original. This process continues until there is no longer enough energy to create additional $q\bar{q}$ pairs, and instead the remaining

[3]We use "colorless" to mean a singlet representation of the color group.

partons combine to form stable hadrons. Since this progression involves successively lower energies and lower momentum transfers, perturbative QCD cannot describe the full process. The hadronization process then cannot be calculated from first principles, but has to be modelled.

2.3.1 Monte Carlo tools

Knowing QCD predictinos is crucial in the design of methods to search for new physics, as well as for extracting meaning from data. Different techniques can be used to make QCD predictions at hadron colliders, and in particular at the LHC. The so called Matrix Element Monte Carlos use direct perturbative calculations of the cross-section matrix elements for each relevant partonic subprocesses. LO and NLO calculations are available for many processes. These "fixed-order predictions" include the first terms in the QCD perturbative expansion for a given cross-section; as more terms are involved in the expanstion, an improvement in the accuracy of the prediction is expected. The complexity of the calculations increased significantly with the number of outgoing legs, limiting available results to those with at most three outgoing partons. Matrix element MC programs include ALPGEN [26], MADGRAPH [27] and others.

An alternative approach is applied by the so called Monte Carlo parton shower programs. These simulation programs use LO perturbative calculations of matrix elements for $2 \to 2$ processes, relying on the parton shower to produce the equivalent of multi-parton final state. PYTHIA [28] and HERWIG++ [29] are the most commonly used parton shower Monte Carlos together with SHERPA.

The Monte Carlo generators must account for and correctly model the showering of partons. To approximate the energy-evolution of the shower, the DGLP equations that describe the evolution of the PDFs with changing energy scale can be used. The separation of radiation into initial- (before the hard scattering process takes place) and final-state showers is arbitrary, but sometimes convenient. In both initial- and final-state showers, the structure is given in terms of branchings $a \to bc$: $q \to qg$, $q \to q\gamma$, $g \to gg$ and $g \to q\bar{q}$. Parton b carries a fraction z of the energy of the mother energy and parton c carries the remaining $1 - z$ (the term "partons" is including the radiated photons). In turn, daughters b and c may also branch, and so on. Each parton is characterized by some evolution scale, which gives an approximate sense of time ordering to the cascade. In the initial-state shower, the evolution scale values are gradually increasing as the hard scattering is approached, while these values decrease in the final-state showers. The evolution variable of the cascade in the case of PYTHIA generator, Q^2, has traditionally been associated with the m^2 of the branching partons[4]. In the recent version of PYTHIA a $p_\perp$-ordered shower algorithm, with $Q^2 = p_\perp^2$ is available, and the shower evolution is cut off at some lower scale Q_0 typically arond 1 GeV for QCD branchings. HERWIG++ provides a shower model which is angular-ordered.

There are two leading models for the description of the non-perturbative process of hadronization, after parton showering. PYTHIA uses the Lund string model of hadronization to form particles [30]. This model involves stretching a colour "string" across quarks and gluons and breaking it up into hadrons.HERWIG++ utilizes the cluster model of hadronization. In this model each gluon is split into a $q\bar{q}$ pair and then quarks and anti-quarks are grouped into colourless "clusters", which then give the hadrons.

[4]The final-state partons have $m^2 > 0$. For initial-state showers the evolution variable is $Q^2 = -m^2$, which is required to be strictly increasing along the shower.

Hadronization models involve a number of non-perturbative parameters. The parton-shower itself involves the non-perturbative cut-off Q_0^2 . These different parameters are usually tuned to data from the LEP experiments.

In addition to the hard interaction that is generated by the Monte Carlo simulation, it is also necessary to account for the interactions between the incoming proton remnants. This is usually modelled through multiple extra $2 \to 2$ scattering, occurring at a scale of a few GeV. This effect is known as multiple parton interactions (MPIs). In addition, these partons may radiate some of their energy, either before or after the hard interaction. All the parton interactions, which are note calculated form the hard scattering process, are grouped together in the term underlying event. The modelling of the underlying event is crucial in order to give an accurate reproduction of the (quite noisy) energy flow that accompanies hard scatterings in hadron-collider events.

It should be stressed that these multiple parton interactions are a separate effect from the multiple proton interactions that may occur in each collision event in the LHC. These multiple proton collisions are referred to as pileup, and are not included in the definition of the underlying event.

No precise model exists to reproduce the unerlying event activity. This activity is instead also adjusted to reproduce available experimental data. A specific set of chosen parameters for a generator is referred to as a "tune".

The two Monte Carlo generators used in this analysis are summarized below, indicating the particular versions and tunes that were implemented.

Pythia

PYTHIA event generator has been used extensively for e^+e^-, ep, $pp/p\bar{p}$ at LEP, HERA, and Tevatron, and during the last 20 years has probably been the most used generators for LHC physics studies. PYTHIA contains an extensive list of hardcoded subprocesses, over 200, that can be switched on individually. These are mainly 2→1 and 2→2, some 2→3, but no multiplicities higher than that. Consecutive resonance decays may of course lead to more final-state particles, as will parton showers.

As mentioned above, in this MC generator, showers are ordered in transverse momentum [31] both for ISR and for FSR. Also MPIs are ordered in p_T [32]. Hadronization is based solely on the Lund string fragmentation framework.

For the results presented in this thesis simulated samples of dijet events from proton-proton collision processes were generated with P YTHIA 6.423 [28]. The ATLAS AMBT2 tune of the soft model parameters was used [33]. This tune attemps to reproduce the ATLAS minimum bias charged particle multiplicity and angular distribution measurements and the ATLAS measurements of charge particle and p_T density obseved collinear and transverse to the high-energy activity.

For systematic comparisons, a set of additional tunes, called the Perugia tunes [34] were also used. These tunes utilize the minimum bias and p_T density measurements of CDF to model the underlying event, hadronic Z^0 decays from LEP to model the hadronization and final state radiation, and Drell Yann measurements from CDF and $D0$ to model the initial state radiation. In particular, the Perugia 2011, which is a retune of Perugia 2010 [35] including 7 TeV data (Mar 2011).

Herwig++

HERWIG++ [29] is based on the event generator HERWIG (Hadron Emission Reactions With Interfering Gluons), which was first published in 1986 and was developed throughout the era of LEP. Herwig was written in Fortran, and the new generator, Herwig++ developped in C++. Some distinctive features of Herwig++ are: Angular ordered parton showers and cluster hadronization. Hard and soft multiple partonic interactions to model the underlying event and soft inclusive interactions [?].

This MC generator was used for systematic uncertainties studies. The version utilized was versin 2.4.2 released in 2009.

In order to use events produced by Monte Carlo generators to model events that one might observe with the detector, the output of these generators is passed through a detector simulation model. ATLAS uses the GEANT4 [36] toolkit to simulate the passage of particles through the detector material. This includes models for the production of additional particles caused by inelastic scattering off of electrons and nuclei, as well as ionization and absorption by active detector elements.

2.3.2 Jet algorithms

As described above, quarks and gluons cannot be directly observed. Almost inmediatly after being produced, a quark or gluon hadronises, leading to a collimated spray of energetic hadrons, a jet. By measuring the jet energy and direction one can close to the idea of the original parton. But one parton may form multiple experimentally observed jets, for example due to a hard gluon emission plus soft and collinear showering. Then, in comparing data to theory and MC programs predictions a set of rules for how to group particles into jets is needed. A jet algorithm, together with a set of parameters and a recombination scheme (how to assign a momentum to the combination of two particles) form a jet definition.

By using a jet definition a computer can take a list of particle momenta for an event (be they quarks and gluons, or hadrons, or calorimeter depositions), and return a list of jets. One important point to remark is that the result of applying a jet definition should be insensitive to the most common effects of showering and hadronization, namely soft and collinear emissions. This is illustrated in Fig. 2.2.

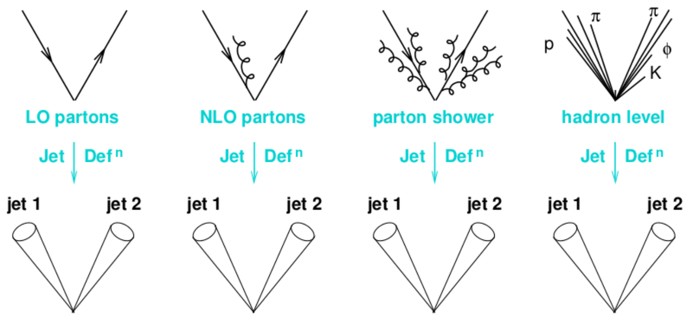

Figure 2.2: The application of a jet definition to a variety of events that differ just through soft/collinear branching and hadronization should give identical jets in all cases [37].

Tradiontally, jet algorithms have been classified into two categories: cone algorithms and sequential recombination algorithms.

Cone-like algorithms are based on the collinear nature of gluon radiation and the parton shower described above. The decay products of and emission from a hard quark or gluon will tend to form a cone of particles in the $\eta-\phi$ plane as they propagate. An cone algorithm will work as follows [5]: first, it sorts all particles in the event according to their momentum, and identifies the one with largest p_T. This is referred to as seed particle. Then a cone of radius R in $\eta-\phi$ is drawn around the seed. The direction of the sum of the momenta of those particles is identified and if it doesn't coincide with

[5] This is how CMS cone algorithm, used for the preparation for the LHC running, works.

the seed direction then the sum is used as a new seed direction, and iterates until the sum of the cone contents coincides withthe previus seed (this type of algorithm is cone "iterative" cone since it iterates the cone direction). This is how a stable cone is reached. A difficulty and major drawback in this procedure is the use of the transverse momentum of the particle to select the first seed. This definition is collinear unsafe, i.e. a splitting of the hardest particle into a nearly collinear pair can have the consequence that another, less hard particle, pointing in a different direction suddenly becomes the hardest in the event, leading to a different final set of jets. There are many other variants of cone algorithms, and nearly all suffer from problems of either collinear safety, or infrared safety (an extra soft particle creates a new seed, which can lead to an extra stable cone being found). A fix for these problems came in a algorithm called Seedless Infrared Safe Cone (SISCOne) [38].

Recombination algorithms, on the other hand, are both collinear and infrared safe. And for this reason, they can be used in calculations to any order in perturbation theory. The term recombination is used given that these algorithms work as if they were inverting the sequence of splittings of the parton shower. In general, recombination algorithms operate by successively combining pairs of particles using a distance metric, d_{ij}. At hadron colliders, due to the fact that one of the incoming partons may continue along the beam, for every pair of particles this metric is compared to a so-called "beam distance", d_{iB}, and only when $d_{ij} < d_{iB}$ the particle pair is combined and considered for subsequent clustering steps.

ATLAS (and also CMS) has chosen anti-k_t [39] algorithm as the default jet algorithm for use in physics analysis. This recombination algorithm as well as the Cambridge-Achen algorithm [**?**], or C/A are extensions of the original k_t algorithm developed for the analysis of multi-jet events at e^+e^-

colliders [40] and subsequently extended for use at hadron colliders [41] [42]. In this thesis, the k_t algorithm was used for jet substructure studies, see section 2.3.3.

The orginal k_t algorithm implements the following (2.1) distance metric between particles i and j,

$$d_{ij} = \frac{2E_iE_j(1 - cos\theta_{ij})}{Q^2} \tag{2.1}$$

where Q is the total energy in the event, E_i is the energy of particle i and θ_{ij} the angle between particles i and j. In the collinear limit, d_{ij} is related to the relative transverse momentum between particles i and j (hence the name k_t algorithm), normalized to the total visible energy. The particles are combined if the minimum d_{ij}, d_min, is below a certain threshold, y_{cut}. The jet multiplicity depends on the value of y_{cut}, as a lower value will result in more soft or colliinear emissions surviving as jets. As mentioned above, for hadron colliders, the notion of a beam distance is added. A distance scale, $\Delta R = \sqrt{\Delta y^2 + \Delta\phi^2}$, is introduced to define the typical radius for a jet, effectively replacing y_{cut}. In this case for every pair of particles a new distance is define, (2.2),

$$d_{ij} = min(p_{ti}^2, p_{tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{2.2}$$

and the beam distance, $d_{iB} = p_{ti}^2$. The algorithm proceeds by searaching for the smallest of the d_{ij} and the d_{iB}. If it is a d_{ij} then particles i and j are recombined into a single new particles. If it is a d_{iB} then i is removed from the list of particles, and called a jet. This is repeated until no particles remain.

As opposed to cone algorithms, for the k_t algorithm, the jets have quite irregular shapes, and particles with $\Delta R_{ij} > R$ can still be clustered within the jet. This is a problem when, for example, an irregularly shaped jet happens to extend into poorly instrumented detector regions. Another drawback of this definition is that soft particles are clustered first. This has the potential to introduce complications when the detector noise of energy density fluctuations are large.

A feature of the k_t algorithm that is attractive is that it not produces jets but also assigns a clustering sequence to the particles within the jet. It is possible then to undo the clustering and look inside the structure of the jet. This has been exploited in a range of QCD studies, and also in searches of hadronic decays of boosted massive particles and will be used here for the search of two-pronged jets in gluon splitting.

The prescription above may be generalized beyond the k_t algorithm. By inverting the power law in the particle distance metric, d_{ij}, the anti-k_t algorithm is obtained. The particle distance metric used by this algorithm is,

$$d_{ij} = min(p_{ti}^{-2}, p_{tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \tag{2.3}$$

and the beam distance, $d_{iB} = p_{ti}^{-2}$. This definition results in the clustering of the hardest emissions first. This has several benefits in the context of high-luminosity hadron collisions.

Note that the anti-k_t algorithm does not provide useful information on jet substructure if a jet contains two hard cores, then the k_t (or C/A) algorithms first reconstruct those hard cores and merge the resulting two subjets. The anti-k_t will often first cluster the harder of the two cores and then gradually aglomerate the contents of the second hard core.

These algorithms, and more, are implemented in FASTJET [43] software package for jet-finding.

2.3.3 Jet substructure

The study of a quantity related to the distributions and multiplicity of particles in the event phase space lead to the first evidence of jet structure, as pointed out in ref. [25]. In general, all final hadronic states in $pp/p\bar{p}/e^+e^-$ collisions can be explored in terms of the structure and shape of the event energy flow by means of so called "event shape" variables. This family of variables attempt to extract information about the global geometry of an event, usually distinguishing between di-jet events and multijet final states. Such variables have been successfully utilized in many SM measurements and BSM searches, see for example [44][45].

Although very useful, event shape variables are not sensitive to the detailed structure and distribution of energy inside a particular jet in the event. In new physics searches, tools for the identification of individual objects that might be signature of new particles are desired. For example, when an unstable massive particle with large p_T decays hadronically, the final state may be composed of a number of nearly collinear jets. These jets may be merged by a jet finder; a method for selecting these jets would allow for the study of their properties. This interest lead to the development of a wide range of jet substructure techniques in the recent years.

Jet substructure methods probe the internal structure of jets from a detailed study of its constituents (see chapter**??**). These techniques have been first thought for distinguishing boosted hadronic objects from the background of jets initiated by light quarks and gluon, see for example [46], but they have been also used succesfully in other applications, including separating quark jets from gluon jets [47] and identifying boosted decay producs in new physics [48].

Jet shapes, which are event shape-like observables applied to single jets, are an effective tool to measure the structure of individual jets [49].The shape of a jet no only depends on the type of parton (quark or gluon) but is also sensitive to non-perturbative fragmentation effects and underlying event contributions [50].

In the particular case of the present analysis, several distinguishing characteristics between jets originating from b-quarks and jets originating from the the splitting of a gluon into a $b\bar{b}$ pair can be determined using the techniques of jet substructure.

Jet width

The jet width is part of a set of continuous variables that try to distinguish individual particles/subjets within the jet as a smooth funcion of $(\delta\eta, \delta\phi)$ away from the jet axis, in order to form combinations like geometric moments. This particular combination sums the distances between the jet constituents and its axes, weighted by the constituent p_T, and then normalized to the total p_T of the jet. The compact definition is

$$Jet\ width = \frac{\sum_{i=1}^{N} {p_T}^{const_i}\, \Delta R(const_i, jet)}{\sum_{i=1}^{N} {p_T}^{const_i}} \tag{2.4}$$

where N is the total number of calorimeter or track constituents. This observable is also highly correlated to the mass of the jet.

This linear radial moment is a measure of the width or "girth" [51] of the jet. Under the assumption of central jets with massless constituents at small angles, this linear moment is identical to jet broadening [52], defined as the sum of momenta transverse to the jet axis normalized by the sum of momenta. While jet broadening is natural at an e^+e^- collider, the linear radial moment is more natural at the LHC.

An alternative approach to measuring the width is to use the angular separation of the two hardest constituents inside jets. This has the advantage of effectively removing any dependence on the shower development within the calorimeter and focuses on the hard component of the jet.

Eccentricity

In defining a jet moment there are several ways to weight the momentum and define the center of the jet. We have defined the jet width as the first moment of the transverse energy with respect to the jet axis; another example of useful combination is the jet pull [51]. But it is also natural to look at higher moments, such as those contained in the covariance tensor,

$$C = \sum_{i \in jet} \frac{p_T^i |r_i|}{p_T^{jet}} \begin{pmatrix} \Delta y_i^2 & \Delta y_i \Delta \phi_i \\ \Delta \phi_{ii} & \Delta \phi_i^2 \end{pmatrix}.$$

Here, $r_i = (\Delta y_i, \Delta \phi_i) = c_i - J$, where $J = (y_J, \phi_J)$ is the location of the jet and c_i is the position of a cell or particle with transverse momentum p_T^i. The eigenvalues $a \geq b$ of this tensor are similar to the semimajor and semiminor axes of an elliptical jet. The jet eccentricity, defined below, is a combination of these eigenvalues, and it is a measure of how elongated is the area of a jet.

$$e = \sqrt{\frac{(a^2 - b^2)}{a}} \tag{2.5}$$

Jet Mass

The jet mass, like the linear radial moment, also depends on the radiation pattern of the event. It is the most basic observable for disinguishing massive boosted objects from jets originating from quarks or gluons. The latter are expected to be dominated by wide-angle emissions, with increase probability to see high mass jets initiated from gluons as opposed to quarks [53].

NEED TO COMPLETE THIS.

Subjet multiplicity

With the development of the k_t algorithm, subjets were first used in the description of the hadronic final state in e^+e^- annihilation, such as the study of the jet multiplicity at different energy scales [54]. By using the sequential recombination algorithms introduced in the previus section, it is straightforward to define a "subjet algorithm" in which the structure of the jet's constituents is resolved using either the same jet finder algorithm or a new one with a fixed (smaller) distance parameter.

The subjet multiplicity − the number of subjets within a jet − provides information on the distribution of energy and multiplicity of particles within a jet. For instance, in [55] the result of meassuring this "radiation variable" on quark- and gluon-initiated jets indicates that gluon-initiated jets tend to have on average higher subjet multiplicity. This result is consistent with the QCD prediction that gluons radiate more than quarks. In the case of this and different other analyses the k_t algorithm is rerun for subjet finding.

As an alternative to fixed distance parameter subjets, it is also possible to undo the last step in the recombination sequence [41] in order to identify the decay products of an object. This approach is used in seveal jet grooming procedures[6], see for instance [57].

It is also possible to extend the use of individual subjets in conjunction with more traditional jet shape variables. Using these tools, an inclusive jet shape based on the substructure topology of a single jet, "N-subjettiness" [58] is defined.

[6]Jet grooming comprises dedicated techniques to remove uncorrelated radiation within a jet. A review of these procedures can be found in [56].

N-subjettiness

As mentioned above, the N-subjettiness [58] is a jet shape that describes the energy flow within a jet. It quantifies the degree to which radiation is aligned along specified subjet axes. This jet shape was adapted from the event shape N-jettiness [59].

Given candidate subjets directions determined by an external algorithm such as the exclusive k_t procedure [60], the variables is defined as,

$$\tau_N^{(\beta)} = \frac{1}{\sum_k p_{Tk}\,(R_0)^\beta} \sum_k p_{Tk} (\min\{\Delta R_{j1,k},\, \Delta R_{j2,k}, ..., \, \Delta R_{jN,k}\})^\beta \tag{2.6}$$

The sum runs over the k constituent particles in a given jet where $p_{T,k}$ are their transverse momenta, and $\Delta R_{j1,k}$ is the distance between the candidate subjet $j1$ and a constituent particle k. R_0 is the characteristic jet radius used in the original jet clustering algorithm. The exponential weight, β, can optionally be applied to the angular distance computed between the subjets and the jet constituents.

This jet shape was designed to identify boosted N-prong hadronic decays. With $\beta = 1$, the definition above indicates that jets with $\tau_N \approx 0$ have all their radiation aligned with the candidate subjet directions and therefore have N (or fewer) subjets. Jets with $\tau_N \gg 0$ have a large fraction of their energy distributed away from the candidate subjet direction and therefore have at least $N + 1$ subjets.

To separate boosted hadronic objects from the QCD jet background, one could use the complete set of τ_N (with different values of β) in a multivariate analysis. However, [58] showed that a simple cut on the ratio τ_N/τ_{N-1} provides excellent discrimination power for N-prong hadronic objects. In particular, τ_2/τ_1 can identify boosted W/Z and Higgs bosons, with the angular weighting exponent $\beta = 1$ providing the best discrimination.

Since eq. 2.6 is linear in each of the constituent particle momenta, this variable is an infrared- and colliner-safe observable. In subsequent work [61], Thaler and van Tilburg showed that the initial step of choosing candidate subjet axes is in fact unnecessary. In particular, the quantity in equation 2.6 can be minimised over the candidate subjet directions, further improving boosted object discrimination.

The definition of N-subjettiness is not unique, and different choises can be sued to give different weights to the emissions within a jet. There generalizations of N-subjettiness are similar to different "angularities" [62] used in $e^+e^- \rightarrow$hadrons measurements.

2.4 Heavy flavor jet production

Heavy flavour quark production in hadronic collisions may be subdivided into three classes depending on the number of heavy quarks participating in the hard scattering. The hard scatter is defined as the 22 subprocess with the largest virtuality (or shortest distance) in the hadron-hadron interaction [63].

- Quark pair creation: two heavy quarks are produced in the hard subprocess. Being Q the heavy flavor quark, at leading order this process is described by $gg \rightarrow Q\bar{Q}$ and $q\bar{q} \rightarrow Q\bar{Q}$.
- Heavy flavor quark excitation: the heavy flavour quark excitation can be depicted as an initial state gluon splitting into a heavy quark pair, where one of the heavy quarks subsequently enters the hard subprocess.
- Gluon splitting: no heavy quarks participate in the hard subprocess in this case, but they are produced in $gQ\bar{Q}$ branchings in the parton shower.

Example of Feynman diagrams for QCD b-quark production up to NLO were shown intance Fig. 2.1.

The definition above is not strict, but can be used as a basis for the understanding of the characteristics of heavy flavour quark production.

Bibliography

[1] S. Frixione and M.L. Mangano. Heavy quark jets in hadronic collisions. *Nucl.Phys.*, B483:321–338, 1997.

[2] Andrea Banfi, Gavin Salam, and Giulia Zanderighi. Accurate qcd predictions for heavy-quark jets at the tevatron and lhc. *JHEP*, 0707:026, 2007.

[3] Andrea Banfi, Gavin Salam, and Giulia Zanderighi. Infrared safe definition of jet flavour. *Eur.Phys.J.C*, 47:022, 2006.

[4] John M. Campbell, R.Keith Ellis, F. Maltoni, and S. Willenbrock. Production of a W boson and two jets with one b^- quark tag. *Phys.Rev.*, D75:054015, 2007.

[5] P. Abreu et al. Measurement of the multiplicity of gluons splitting to bottom quark pairs in hadronic Z0 decays. *Phys.Lett.*, B405:202–214, 1997.

[6] Toshinori Abe et al. A Preliminary measurement of the gluon splitting rate into b anti-b pairs in hadronic Z0 decays. 1999.

[7] CDF Collaboration. Measurements of Bottom Anti-Bottom Azimuthal Production Correlations in Proton-Antiproton Collisions at s**(1/2)= 1.8 TeV. *Phys.Rev.D*, 71:38, 2005.

[8] F. Abe et al. Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab.

[9] S. Abachi et al. Search for high mass top quark production in $p\overline{p}$ collisions at $\sqrt{s} =$ 1.8 tev.

[10] P.W. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12(2):132 – 133, 1964.

[11] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. 2012.

[12] A. Sherstnev and R. S. Thorne. Parton distributions for LO generators. *European Physical Journal C*, 55:553–575, June 2008.

[13] F. J. Dyson. The radiation theories of tomonaga, schwinger, and feynman. *Phys. Rev.*, 75:486–502, Feb 1949.

[14] M Banner et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider. *Physics Letters B*, 122(56):476 – 485, 1983.

[15] G. Arnison et al. Experimental observation of lepton pairs of invariant mass around 95 gev/c2 at the cern sps collider. *Physics Letters B*, 126(5):398 – 410, 1983.

[16] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.

[17] A. Salam and J.C. Ward. Electromagnetic and weak interactions. *Physics Letters*, 13(2):168 – 171, 1964.

[18] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967.

[19] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8(3):214 – 215, 1964.

[20] G. Zweig. AN SU(3) MODEL FOR STRONG INTERACTION SYMMETRY AND ITS BREAKING. 2. 1964.

[21] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365 – 368, 1973.

[22] Richard P. Feynman. Very high-energy collisions of hadrons. *Phys. Rev. Lett.*, 23:1415–1417, Dec 1969.

[23] John C. Collins, Davison E. Soper, and George F. Sterman. Factorization of Hard Processes in QCD. *Adv.Ser.Direct.High Energy Phys.*, 5:1–91, 1988.

[24] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear Physics B*, 126(2):298 – 318, 1977.

[25] G. Hanson, G. S. Abrams, A. M. Boyarski, M. Breidenbach, F. Bulos, W. Chinowsky, G. J. Feldman, C. E. Friedberg, D. Fryberger, G. Goldhaber, D. L. Hartill, B. Jean-Marie, J. A. Kadyk, R. R. Larsen, A. M. Litke, D. Lüke, B. A. Lulu, V. Lüth, H. L. Lynch, C. C. Morehouse, J. M. Paterson, M. L. Perl, F. M. Pierre, T. P. Pun, P. A. Rapidis, B. Richter, B. Sadoulet, R. F. Schwitters, W. Tanenbaum, G. H. Trilling, F. Vannucci, J. S. Whitaker, F. C. Winkelmann, and J. E. Wiss. Evidence for jet structure in hadron production by e^+e^- annihilation. *Phys. Rev. Lett.*, 35:1609–1612, Dec 1975.

[26] Michelangelo L. Mangano, Fulvio Piccinini, Antonio D. Polosa, Mauro Moretti, and Roberto Pittau. Alpgen, a generator for hard multiparton processes in hadronic collisions. *Journal of High Energy Physics*, 2003(07):001, 2003.

[27] Johan Alwall, Pavel Demin, Simon de Visscher, Rikkert Frederix, Michel Herquet, Fabio Maltoni, Tilman Plehn, David L. Rainwater, and Tim Stelzer. Madgraph/madevent v4: the new web generation. *Journal of High Energy Physics*, 2007(09):028, 2007.

[28] Torbjorn Sjostrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[29] M Bhr, S. Gieseke, M.A. Gigg, A. Grellscheid, K. Hamilton, O. Latunde-Dada, S Pltzer, P Richardson, M.H Seymour, M Sherstnev, et al. Herwig++ physics and manual. *Eur.Phys.J.C*, 58:68, 2008.

[30] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(23):31 – 145, 1983.

[31] R. Corke and T. Sjöstrand. Improved parton showers at large transverse momenta. *European Physical Journal C*, 69:1, 2010.

[32] T. Sjöstrand and P. Z. Skands. Transverse-momentum-ordered showers and interleaved multiple interactions. *European Physical Journal C*, 39:129, 2005.

[33] Atlas tunes of pythia 6 and pythia 8 for mc11. Technical Report ATL-PHYS-PUB-2011-009, CERN, Geneva, Jul 2011.

[34] Peter Z. Skands. The Perugia Tunes. 2009.

[35] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. *Phys. Rev. D*, 82:074018, 2010.

[36] S. Agostinelli et al. Geant4 a simulation toolkit. *Nucl. Inst. Meth. Section A*, 506(3):250 – 303, 2003.

[37] Salam, G.P. Ellements of QCD for hadron colliders. *CERN-2010-002*, Jan 2011.

[38] Gavin P. Salam and Grgory Soyez. A practical seedless infrared-safe cone jet algorithm. *Journal of High Energy Physics*, 2007(05):086, 2007.

[39] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[40] W. Bartel, L. Becker, R. Felst, D. Haidt, G. Knies, H. Krehbiel, P. Laurikainen, N. Magnussen, R. Meinke, B. Naroska, et al. Experimental studies on multijet production in e^+e annihilation at PETRA energies. *EPJ C Particles and Fields*, 33:8, 1986.

[41] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[42] S. Catani, Y.L. Dokshitzer, H. Seymour, and B.R. Webber. Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions. *Nucl. Phys.*, B406:187, 1993.

[43] M Cacciari and G.P. Salam. Dispelling the N^3 myth for the k_t jet-finder. *Phys. Lett. B}, 661:057, 2006.*

[44] G. Abbiendi et al. Measurement of alpha(s) with Radiative Hadronic Events. 2007.

[45] *Georges Aad et al. Measurement of event shapes at large momentum transfer with the ATLAS detector in pp collisions at sqrt(s) = 7 TeV. 2012.*

[46] *ATLAS Collaboration. Atlas sensitivity to the standard model higgs in the hw and hz channels at high transverse momenta.* ATL-PHYS-PUB-2009-088*, Aug 2009.*

[47] *Jason Gallicchio and Matthew D. Schwartz. Quark and gluon tagging at the lhc.* Phys. Rev. Lett.*, 107:172001, Oct 2011.*

[48] *Graham D. Kribs, Adam Martin, Tuhin S. Roy, and Michael Spannowsky. Discovering higgs bosons of the mssm using jet substructure.* Phys. Rev. D*, 82:095012, Nov 2010.*

[49] *Stephen Ellis, Christopher Vermilion, Jonathan Walsh, Andrew Hornig, and Christopher Lee. Jet shapes and jet algorithms in scet.* Journal of High Energy Physics*, 2010:1–83, 2010. 10.1007/JHEP11(2010)101.*

[50] *G. Aad et al. Study of jet shapes in inclusive jet production in pp collisions at* $\sqrt{s} = 7$ TeV *using the atlas detector.* Phys. Rev. D*, 83:052003, Mar 2011.*

[51] *Jason Gallicchio and Matthew D. Schwartz. Seeing in color: Jet superstructure.* Phys. Rev. Lett.*, 105:022001, Jul 2010.*

[52] *S. Catani, G. Turnock, and B.R. Webber. Jet broadening measures in e+e annihilation.* Physics Letters B*, 295(34):269 – 276, 1992.*

[53] *Leandro G. Almeida, Seung J. Lee, Gilad Perez, Ilmo Sung, and Joseph Virzi. Top quark jets at the lhc.* Phys. Rev. D*, 79:074012, Apr 2009.*

[54] S. Catani, Yu.L. Dokshitzer, F. Fiorani, and B.R. Webber. Average number of jets in e+e annihilation. Nuclear Physics B, 377(3):445 – 460, 1992.

[55] R. Snihur. Subjet multiplicity in quark and gluon jets at d0. Nuclear Physics B - Proceedings Supplements, 79(13):494 – 496, 1999. ¡ce:title¿Proceedings of the 7th International Workshop on Deep Inelastic Scattering and QCD¡/ce:title¿.

[56] A. Abdesselam, E. Bergeaas Kuutmann, U. Bitenc, G. Brooijmans, J. Butterworth, et al. Boosted objects: A Probe of beyond the Standard Model physics. Eur.Phys.J., C71:1661, 2011.

[57] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Techniques for improved heavy particle searches with jet substructure. Phys. Rev. D, 80:051501, Sep 2009.

[58] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. JHEP, 1103:015:026, 2011.

[59] Iain W. Stewart, Frank J. Tackmann, and Wouter J. Waalewijn. n jettiness: An inclusive event shape to veto jets. Phys. Rev. Lett., 105:092002, Aug 2010.

[60] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. Phys. Rev. D, 48:3160–3166, Oct 1993.

[61] Jesse Thaler and Ken Van Tilburg. Maximizing boosted top identification by minimizing n-subjettiness. Journal of High Energy Physics, 2012:1–33, 2012. 10.1007/JHEP02(2012)093.

[62] *Carola F. Berger, Tibor Kúcs, and George Sterman. Event shape˘energy flow correlations.* Phys. Rev. D*, 68:014012, Jul 2003.*

[63] *E. Norrbin and T. Sjostrand. Production and hadronization of heavy quarks.* Eur.Phys.J.*, C17:137–161, 2000.*