Identification and tagging of double b-hadron jets with the ATLAS Detector

María Laura González Silva

Doctoral Thesis in Physics

Physics Department

University of Buenos Aires

November 2012

UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Física

Identificación y etiquetado de jets con dos hadrones b con el detector ATLAS.

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 de Florian

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

Buenos Aires, 2012

AGRADECIMIENTOS

Quiero agradecer a mi director, Ricardo Piegaia, por su enseñanza constante, dedicación y amistad y a todos aquellos que trabajaron junto conmigo en el experimento ATLAS, Gastoncito Romeo (mi compañero de aventuras desde el comienzo), Gustavo Otero y Garzón, Hernán Reisin y Sabrina Sacerdotti. Un especial agradecimiento a Ariel Schwartzman y su equipo por las ideas y el soporte técnico constante. Quiero agradecer al Laboratorio CERN, al Experimento ATLAS, al programa HELEN y al programa e-Planet por permitirme ser parte de un proyecto increible.

Quiero agradecer también a mis compañeros de grupo y oficina, Javier Tiffenberg, Yann Guardincerri, Pablo Pieroni y Orel Gueta por estar siempre dispuestos a darme una mano. Quiero agradecer el apoyo de mis compañeros de la carrera, especialmente a mis amigos Cecilia Bejarano y Tomas Teitelbaum. Quiero agradecer a los amigos que hice a lo largo de estos años en mis visitas al Laboratorio CERN, y a mis colegas y amigos de la Universidad de la Plata. Un especial agradecimiento a Fernando Monticelli.

Quiero agradecer a mis amigos de la vida por continuar a mi lado a pesar de las ausencias. Agradezco profundamente a toda mi familia por su apoyo y comprensión, especialmente a Cristina Silva, Lorena González y Juan Martín Alba.

Finalmente, quiero agradecer al CONICET y a la Fundación Exactas por hacer posible la realización de esta tesis.

Abstract

Esta tesis describe un método que permite la identificación de jets que contienen dos hadrones b, que se originan en la división de un gluon en un par $b\bar{b}$. La técnica desarrollada explota las diferencias cinemáticas entre los llamados jets "merged" y los genuinos jets b, usando variables que describen la estructura interna y la forma de los jets, construídas a partir de las trazas asociadas a los mismos. Las variables con mayor poder discriminador son combinadas en un análisis de multivariable. Poder identificar y remover jets b que provienen de la división de un gluon es importante para la estimación y la redución del fondo a señales de física dentro del Modelo Estándar y en nueva física. El algoritmo diseñado rechaza, en eventos simulados, el 95% (50%) de los jets "merged", mientras que retiene el 50% (90%) de los jets b genuinos.

Palabras clave: Experimento ATLAS, Jets, Subestructura de Jets, QCD, Producción de jets b, Etiquetado de Jets b.

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.

Keywords: ATLAS Experiment, Jets, Jet Substructure, b-jet Production, QCD, Gluon Splitting, b-tagging.

Contents

Chapter 1

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.

1.1 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 [1, 2]. 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 [3].

Very recently, a Higgs-like particle was discovered by ATLAS and CMS experiments at the LHC [4]. 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, 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 [5]. 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 [6, 7]. 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 [8, 9, 10].

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 proposed the quark model [11, 12, 13], 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 [14] 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 [15], 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.

1.2 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 hadrons, 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 performed. 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.

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, the factorization theorem [16] allows the 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 choices 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 [17]. The DGLAP equations are derived by noting that the cross-section of a process should be independent of the factorization scale μ_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.

1.3 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 [18].

The evolution from a single parton to an ensemble of hadrons occurs 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 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 full parton shower and hadronization process then cannot be calculated from first principles, but has to be modelled.

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

1.3.1 Monte Carlo tools

Knowing QCD predictions 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 increases significantly with the number of outgoing legs. Matrix element MC programs include ALPGEN [19], MADGRAPH [20] 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 [21] and HERWIG++ [22] are the most commonly used parton shower Monte Carlos.

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" includes 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, 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 [23]. 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.

Hadronization models involve a number of âĂŸnon-perturbativeâĂŹ 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

[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.

$2 \to 2$ 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 additional parton interactions, which are not involved in 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 underlying 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 $2\to1$ and $2\to2$, some $2\to3$, 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 trans-

verse momentum [24] both for ISR and for FSR. Also MPIs are ordered in p_T [25]. 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 PYTHIA 6.423 [21]. The ATLAS AMBT2 tune of the soft model parameters was used [26]. 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 observed collinear and transverse to the high-energy activity.

For systematic comparisons, a set of additional tunes, called the Perugia tunes [27] 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 [28] includes 7 TeV data from 2011 data taking.

Herwig++

HERWIG++ [22] is based on the event generator HERWIG (Hadron Emission Reactions With Interfering Gluons), which was first published in 1986 and was developed throughout the LEP era. 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, and 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 version 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 [29] 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.

1.3.2 Jet algorithms

As described above, quarks and gluons cannot be directly observed. Quarks and gluons hadronise, leading to a collimated spray of energetic hadrons, a jet. By measuring the jet energy and direction one can get 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. 1.1.

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

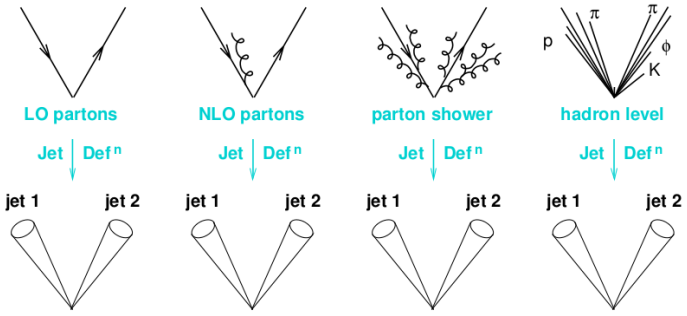

Figure 1.1: The application of a jet definition to a variety of events that differ just through soft/collinear branching and hadronization should give identical jets in all cases [30].

Cone-like algorithms are based on the collinear nature of gluon radiation and the parton shower described above. The decay products of quarks and gluons and their emissions will tend to form a cone of particles in the $\eta - \phi$ plane[5] as they propagate. An cone algorithm will work as follows[6]: 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 a 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 the seed direction then the sum is used as a new seed direction, and it iterates until the sum of the cone contents coincides with the previous seed (this type of algorithm is "iterative" cone since it iterates the cone direction). This is how

[5]The azimuthal angle ϕ is measured around the beam axis, and the polar angle θ is the angle from the beam axis. The pseudorapidity is defined as $\eta = lntan(\frac{\theta}{2})$. The transverse momentum p_T is defined in the plane transverse to the beam motion. The distance ΔR in the pseudorapidity-azimuthal angle space is defined as $DeltaR = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. See section **??** for more details on the ATLAS Coordinate system.

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

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) [31].

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 appropiate 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 the anti-k_t [32] algorithm as the default jet algorithm for use in physics analysis. This recombination algorithm as well as the Cambridge-Achen algorithm [**?**] (C/A) are extensions of the original k_t algorithm developed for the analysis of multi-jet events at e^+e^- colliders [33] and subsequently extended for use at hadron colliders [34, 35]. In this thesis, the k_t algorithm was used for jet substructure studies, see section 1.3.3.

The original k_t algorithm implements the following (1.1) distance metric between particles i and j

$$d_{ij} = \frac{2E_iE_j(1-\cos\theta_{ij})}{Q^2}, \tag{1.1}$$

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 collinear 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 defined, (1.2),

$$d_{ij} = \min(p_{ti}^2, p_{tj}^2)\frac{\Delta R_{ij}^2}{R^2} \tag{1.2}$$

and the beam distance, $d_{iB} = p_{ti}^2$. The algorithm proceeds by searching 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 po-

tential 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 does not produce jets but it also assigns a clustering sequence to the particles within the jet. It is possible then to undo the clustering and to look back at the shower development history. This has been exploited in a range of QCD studies, and also in searches of hadronic decays of boosted massive particles and it 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,

$$d_{ij} = \min(p_{ti}^{-2}, p_{tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \tag{1.3}$$

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 [36] software package for jet-finding.

1.3.3 Jet substructure

The first evidence of jet structure resulted from the study of the spacial distribution and multiplicity of particles in the event phase space in hadron

production in e^+e^- collisions [18]. Generally, all final hadronic states in both 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 the so called "event shape" variables. This family of variables attempts 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 [37][38].

Although very useful, event shape variables are not sensitive to the detailed structure and distribution of energy inside a particular jet. In SM and new physics searches, tools for the identification of individual objects that might be signature of new particles are desired. At the LHC, many of the particles considered to be heavy at previous accelerators will be frequently produced with a transverse momentum greatly exceeding their rest mass, like the electro-weak gauge bosons $W^{\pm}$ and Z, the top quark, the Higgs boson (or bosons) and possibly other new particles in the same mass range. These boosted objects, produced either by recoil against other energetic objects or from decays of even heavier BSM particles, upon decay can give rise to a highly collimated topology too close to be resolved by standard jet algorithms. A method for selecting these jets would allow for the study of their properties. This interest led to the development of a wide range of sophisticated tools in the last years [39, 40] that allow the analysis of the substructure of the ensuing jet and reveal its heavy-particle origin.

Jet substructure methods probe the internal structure of jets from a detailed study of its constituents. These techniques have been first implemented for distinguishing boosted SM hadronic objects from the background of jets initiated by light quarks and gluons, see for example [41], but they have been also succesfully used in other applications, including separating quark

jets from gluon jets [42] and identifying boosted decay producs in new physics searches [43].

Jet shapes, which are event shape-like observables applied to single jets, are an effective tool to measure the structure of individual jets [44].The shape of a jet not only depends on the type of parton (quark or gluon) but is also sensitive to non-perturbative fragmentation effects and underlying event contributions [45].

In chapter **??**, several distinguishing characteristics between jets originating from single b-quarks and jets containing two close-by b-hadrons are determined using the techniques of jet substructure.

1.4 Heavy flavor jet production

Heavy flavor quarks enter in many collider searches, notably because they are produced in the decays of various SM particles (top quarks, the Z boson and the Higgs boson, if light), and of numerous particles appearing in proposed extensions of the SM. 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 $2 \to 2$ subprocess with the largest virtuality (or shortest distance) in the hadron-hadron interaction [46].

- Heavy flavor creation (FCR): two heavy quarks are produced in the hard subprocess. Being Q the heavy flavor quark, at leading order this process is described by $gg \to Q\bar{Q}$ and $q\bar{q} \to Q\bar{Q}$.
- Heavy flavor excitation (FEX): 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 (GSP): no heavy quarks participate in the hard subprocess in this case, but they are produced in $g \to Q\bar{Q}$ branchings in the parton shower.

Example of Feynman diagrams for QCD b-quark production up to NLO are shown in Fig. 1.2.

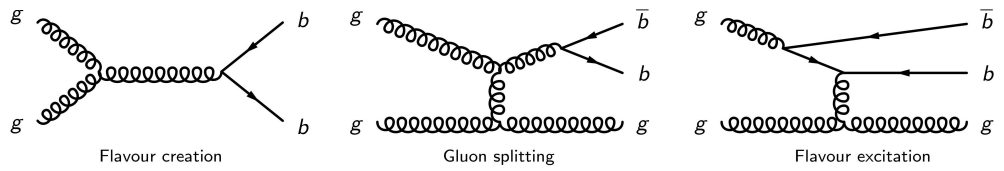

Figure 1.2: Representative diagrams of the three channels contributing to QCD b-quark production up to NLO. The flavour creation channel (left) is the only one present at LO. At NLO, two new channels open up, referred to as gluon splitting (center) and flavour excitation (right).

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

The simplest and most fundamental measurement of heavy-quark jet production is the inclusive heavy-quark jet spectrum, which is dominated by pure QCD contributions. Studies of QCD bottom production are important in their own right 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 ($\sim 10 - 20\%$) [47].

A review of the origin of these uncertainties is presented by Banfi, Salam and Zanderighi in reference [48]. They arise from the poor convergence of the perturbative series, as evidenced by a large value of the K-factor, the ratio of the next-to-leading order (NLO) to the leading order (LO) cross section, in the p_T range covered by the LHC. This is illustrated in Fig. 1.3. The observed K values (6 to 10) indicate that the NLO result cannot be an accurate approximation to the full result. It is for this reason that the scale dependence (middle panel in Fig. 1.3) is large.

The fact that the perturbative series is very poorly convergent is related to the different channels for heavy quark production. While at LO only the FCR channel is present, at NLO the FEX and GSP channels open up[7]. In the gluon splitting process, one of the final-state light partons (at NLO always a gluon) splits collinearly into a $b\bar{b}$ pair that a clustering algorithm can classify within the same jet. A jet containing both b and $\bar{b}$ is considered to be just a b-jet in standard definitions.

The various channels can be approximately separated in a parton shower Monte Carlo generator such as HERWIG or PYTHIA. These MC generators include NLO effects, and one can determine the underlying hard process from the event record. Their relative contributions to the total b-jet spectrum are shown in the bottom panel of Fig. 1.3. 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, going as $\alpha_s^2(\alpha_s \ln(p_T/m_b))^n$ for flavour excitation [17] and $\alpha_s^2 \cdot \alpha_s^n \ln^{2n-1}(p_T/m_b)$ for gluon splitting ($n \geq 1$) [50].

[7]It is sometimes stated that it makes no sense, beyond LO, to separately discuss the different channels, for example because diagrams for separate channels interfere. However, each channel is associated with a different structure of logarithmic enhancements, $\ln^n(p_T/m_b)$, and so there is distinct physical meaning associated with each channel.

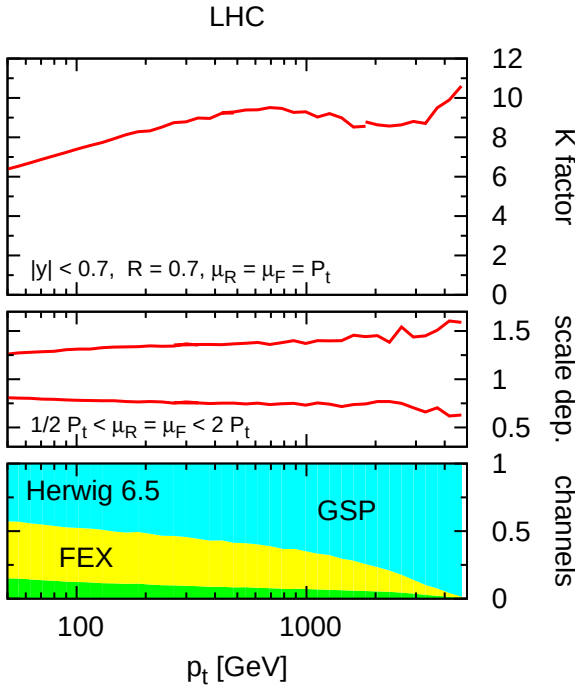

Figure 1.3: Top: K-factor for inclusive b-jet spectrum taken from [48], clustering particles into jets using the k_t jet-algorithm [35] with R=0.7, and selecting jets in the central rapidity region ($|y| < 0.7$). Middle: scale dependence obtained by simultaneously varying the renormalisation and factorisation scales by a factor two around p_T, the transverse momentum of the hardest jet in the event. Bottom: breakdown of the Herwig [49] inclusive b-jet spectrum into the three major underlying channels, flavor creation (FCR) flavor excitation (FEX) and gluon splitting (GSP).

Ref. [48] 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 jets. Final-state logarithms are removed by employing a recently developed jet reconstruction scheme, the flavour-k_t algorithm [51], 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.

Succesfully identifying jets with two b-hadrons, the products of the b-quark or b-antiquark hadronization, can also 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, and associated Higgs production: $WH \to \ell\nu b\bar{b}$ and $ZH \to \nu\nu b\bar{b}$), suffer from the reducible background from QCD, which can produce double b-hadron jets as discussed above, and the irreducible background due to W bosons produced in association with b-quarks. Figure 1.4 shows the two diagramas for $W + b$ production. While at LO only single b-jets are present, at NLO jets containing two b-hadrons are expected due to the contribution of a diagram containing a $gb\bar{b}$ vertex. The b-quark pair is produced at small angles and

can be often reconstructed as one merged jet.

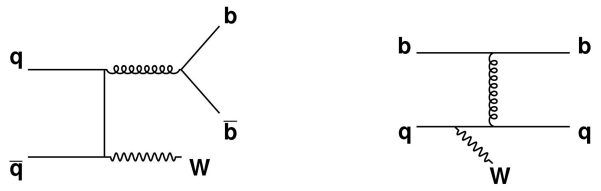

Figure 1.4: Feynman diagrams for W production in association with b quarks.

The relevance of double b-hadron jets is supported by NLO calculations of the production of W bosons and two jets with at least one b quark at the LHC for jet $p_T > 25$ GeV, and $|\eta| < 2.5$ [52] indicate that the cross section for $W(b\bar{b})j$ is almost a factor of two higher than $Wb\bar{b}$, and about a third of Wbj, where $W(b\bar{b})j$ 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. An example is the search for supersymmetry in the framework of generic R-parity conserving models [53]. The superpartners of quarks and gluons could be copiously produced via the strong interaction at the LHC. The partners of the right- and left-handed quarks, $\tilde{\mathrm{q}}_L$ and $\tilde{\mathrm{q}}_R$, can mix to form two mass eigenstates and, since mixing is proportional to the corresponding fermion masses, it becomes more important for the third generation producing sbottom and stop significantly lighter than the other squarks. In this model, thus, sbottom and stop production is expected to dominate. As they chain decay to b-quarks and the lightest supersymmetric particle, the signature for this channel is missing transverse energy plus (single) b-jets. 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 to close-by $b\bar{b}$ pairs.

The study of $b\bar{b}$ jets from gluon splitting is an ideal testbed for exploring jet substructure in data, as it provides a large supply of boosted, merged jets. Furthermore, understanding $g \to b\bar{b}$ jets is important as they are themselves the background to boosted object searches, like $Z \to b\bar{b}$ or $H \to b\bar{b}$. Understanding the much more common QCD events with double b-hadron jets will be essential before attempting to measure more rare final states.

Gluon splitting Measurements

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 $(0.21\pm0.11(\text{stat.})\pm0.09(\text{syst.}))\%$ [54].

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 [55]. Following a similar procedure, the rate per hadronic event was found to be $(3.7\pm0.71(\text{stat.})\pm0.66(\text{syst.}))\times10^{-3}$. In order to improve singal/background ratio, topological information was used.

Bibliography

[1] F. Abe et al. Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab. *Phys. Rev. Lett.*, 74:2626–2631, Apr 1995.

[2] S. Abachi et al. Search for high mass top quark production in $p\overline{p}$ collisions at $\sqrt{s} =$ 1.8 tev. *Phys. Rev. Lett.*, 74:2422–2426, Mar 1995.

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

[4] 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.

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

[6] 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(5âĂŞ6):476 – 485, 1983.

[7] 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.

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

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

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

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

[12] Zweig, G. An SU(3) model for strong interaction symmetry and its breaking. 1964.

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

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

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

[16] 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.

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

[18] G. Hanson, G. S. Abrams, A. M. Boyarski, M. Breidenbach, F. Bulos, W. Chinowsky, G. J. Feldman, C. E. Friedberg, D. Fryberger, G. Gold-

haber, 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.

[19] 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.

[20] 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.

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

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

[23] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2âĂŞ3):31 – 145, 1983.

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

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

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

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

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

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

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

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

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

[33] 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.

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

[35] 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.

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

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

[38] 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.

[39] A. Abdesselam et al. Boosted objects: a probe of beyond the standard model physics. *The European Physical Journal C - Particles and Fields*, 71:1–19, 2011.

[40] A. Altheimer et al. Jet Substructure at the Tevatron and LHC: New results, new tools, new benchmarks. 2012.

[41] 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.

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

[43] 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.

[44] 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.

[45] 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.

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

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

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

[49] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *JHEP*, 0101:010, 2001.

[50] M.H. Seymour. Heavy quark pair multiplicity in e+eâĹŠ events. *Nuclear Physics B*, 436(1âĂŞ2):163 – 183, 1995.

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

[52] 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.

[53] ATLAS Collaboration. Search for supersymmetry in pp collisions at $\sqrt{s} = 7$TeV in final states with missing transverse momentum, b-jets and no leptons with the ATLAS detector. *ATLAS-CONF-2011-098*, 2011.

[54] 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.

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