Measurement of the pair production of b-jets in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector

von
Andrea Neusiedl
geboren in Kaiserslautern

Mainz, den 21.Mai 2012

Tag der Prüfung: 19.11.2012

Kurzfassung

In hadronischen Kollisionen entstehen bei einem Großteil der Ereignisse mit einem hohen Impulsübertrag Paare aus hochenergetischen Jets. Deren Produktion und Eigenschaften können mit hoher Genauigkeit durch die Störungstheorie in der Quantenchromodynamik (QCD) vorhergesagt werden. Die Produktion von *bottom*-Quarks in solchen Kollisionen kann als Maßstab genutzt werden, um die Vorhersagen der QCD zu testen, da diese Quarks die Dynamik des Produktionsprozesses bei Skalen wieder spiegelt, in der eine Störungsrechnung ohne Einschränkungen möglich ist. Auf Grund der hohen Masse von Teilchen, die ein *bottom*-Quark enthalten, erhält der gemessene, hadronische Zustand den größten Teil der Information von dem Produktionsprozess der Quarks. Weil sie eine große Produktionsrate besitzen, spielen sie und ihre Zerfallsprodukte eine wichtige Rolle als Untergrund in vielen Analysen, insbesondere in Suchen nach neuer Physik. In ihrer herausragenden Stellung in der dritten Quark-Generation könnten sich vermehrt Zeichen im Vergleich zu den leichteren Quarks für neue Phänomene zeigen. Daher ist die Untersuchung des Verhältnisses zwischen der Produktion von Jets, die solche *bottom*-Quarks enthalten, auch bekannt als *b*-Jets, und aller nachgewiesener Jets ein wichtiger Indikator für neue massive Objekte. In dieser Arbeit werden die Produktionsrate und die Korrelationen von Paaren aus *b*-Jets bestimmt und nach ersten Hinweisen eines neuen massiven Teilchens, das bisher nicht im Standard-Modell enthalten ist, in dem invarianten Massenspektrum der *b*-Jets gesucht. Am Large Hadron Collider (LHC) kollidieren zwei Protonenstrahlen bei einer Schwerpunktsenergie von $\sqrt{s} = 7$ TeV, und es werden viele solcher Paare aus *b*-Jets produziert. Diese Analyse benutzt die aufgezeichneten Kollisionen des ATLAS-Detektors. Die integrierte Luminosität der verwendbaren Daten beläuft sich auf 34 pb^{-1}. *b*-Jets werden mit Hilfe ihrer langen Lebensdauer und den rekonstruierten, geladenen Zerfallsprodukten identifiziert. Für diese Analyse müssen insbesondere die Unterschiede im Verhalten von Jets, die aus leichten Objekten wie Gluonen und leichten Quarks hervorgehen, zu diesen *b*-Jets beachtet werden. Die Energieskala dieser *b*-Jets wird untersucht und die zusätzlichen Unsicherheit in der Energiemessung der Jets bestimmt. Effekte bei der Jet-Rekonstruktion im Detektor, die einzigartig für *b*-Jets sind, werden studiert, um letztlich diese Messung unabhängig vom Detektor und auf Niveau der Hadronen auswerten zu können. Hiernach wird die Messung zu Vorhersagen auf nächst-zu-führender Ordnung verglichen. Dabei stellt sich heraus, dass die Vorhersagen in Übereinstimmung zu den aufgenommenen Daten sind. Daraus lässt sich schließen, dass der zugrunde liegende Produktionsmechanismus auch in diesem neu erschlossenen Energiebereich am LHC gültig ist. Jedoch werden auch erste Hinweise auf Mängel in der Beschreibung der Eigenschaften dieser Ereignisse gefunden. Weiterhin können keine Anhaltspunkte für eine neue Resonanz, die in Paare aus *b*-Jets zerfällt, in dem invarianten Massenspektrum bis etwa 1.7 TeV gefunden werden. Für das Auftreten einer solchen Resonanz mit einer Gauß-förmigen Massenverteilung werden modell-unabhängige Grenzen berechnet.

Abstract

In hadronic collisions, a large amount of processes with large momentum transfer produce a pair of high-p_{T} jets. Their production rate and event properties can be predicted with good precision using perturbative Quantum Chromodynamics (QCD). The production of bottom-quarks in such collisions is a benchmark process in perturbative QCD because they probe the underlying strong dynamics at a well-defined scale. Because of their large mass, bottom-flavoured particles hold the most direct correspondence between the parton-level production and the observed hadron level. The large pair production rate of bottom-quarks and their corresponding decay products makes them important as background source for many analyses including searches for new physics. Besides this, quarks of the third generation could take an exceptional position among the quarks concerning the sensitivity to new massive objects. Studies on the fraction of jets containing bottom-flavoured particles, known as b-jets, relative to all-flavour jets could reveal such new phenomena. In this thesis the production rate of and the correlation between pairs of b-jets is measured. The invariant dijet mass spectrum is searched for indications for a new resonance in context of physics beyond the Standard Model. At the Large Hadron Collider (LHC) two proton beams at a centre-of-mass energy of $\sqrt{s} = 7$ TeV collide, producing a large number of such pairs of b-jets. This measurement makes use of the data recorded with the ATLAS detector. The total integrated luminosity available for the analysis is about 34 pb^{-1}. b-jets are identified via their long lifetime and the reconstruction of their charged decay products. For this analysis differences between jets originating from light objects, like gluons and light quarks, compared to jets containing bottom-flavoured objects have to be taken into account. The jet energy scale of b-jets is established and the additional uncertainty on the jet energy measurement is determined. Detector effects in the jet reconstruction special to b-jets are studied in order to make the measurement independent of the detector and to correct them to hadron level. Then the cross section measured is compared to next-to-leading order Monte-Carlo predictions. These next-to-leading order predictions are in agreement with ATLAS data. Consequently, the underlying production mechanism is confirmed to also be valid in this new energy regime of the LHC. However, first indications for deviations in the description of next-to-leading processes are discovered. Up to an invariant dijet mass of about 1.7 TeV no evidence for a new resonance decaying into a pair of high-p_{T} b-jets is found. Therefore, model-independent upper limits are calculated for signals following a Gaussian distribution with different signal widths.

Für Mama.

Contents

1 Introduction

The study of matter and its motion has founded the discipline of physics way back in the 6th century BC. The concept of elementary particles forming all matter is very old. It took until the end of the 19th century to discover and understand sub-atomic particles like the electron. Thereafter the knowledge of the structure of matter increased rapidly. So did the scattering of electrons on nucleons confirm the existence of sub-nuclear particles, the quarks. Today physicists describe the universe with the help of these elementary particles and the forces which govern them.

The experiments to test and reveal these smallest particles from which matter is built of have become huge and complex in terms of design. These experiments are in general comprised of a particle accelerator and a particle detector. Particles are brought to almost the speed of light and two particle beams travelling in opposite directions are brought into collision. Around the collision point, a particle detector surveys the particles emerging from the collision.

The Large Hadron Collider (LHC), which was built at the European Organisation for Nuclear Research (CERN) in Geneva, is the world's largest and most powerful particle accelerator. It has been designed to collide two proton beams at a centre-of-mass energy of 14 TeV with very high luminosity. LHC is intended to test the current understanding of high-energy physics and the properties of interactions at extreme energy densities. It started regular operation in 2009. First collisions at a centre-of-mass energy at 7 TeV have been recorded in early 2010. The operation of this accelerator has successfully continued in 2011 and 2012.

One of the experiments, which detects the collision products, is the ATLAS experiment (A Toroidal LHC ApparatuS). This particle detector is comprised of three main components to measure the trajectories, the energy and specific particle types of collision products.

Quarks are charged particles with half-integral spin and are thought of as point-like particles. There are six different types of quarks currently known and they differ in their masses. The heavy types of these quarks are the most precisely measured objects because they take an exceptional position among the quarks due to their large masses. Their behaviour and consequently the behaviour of composite particles containing such heavy quarks is significantly different from lighter types. Thus, experiments are able to distinguish between light- and heavy-quark objects.

The second heaviest quark is called bottom-quark (b-quark). Evidence of its existence was found in 1977 at the Fermilab for the first time. The long lifetime of composite particles containing this quark type lets them travel a macroscopic distance before decaying (on average 3 mm in the transverse plane at $p_\mathrm{T} \approx 50$ GeV). The development and introduction of silicon technology in particle detectors improved the identification of such decays and in turn of particles containing the bottom-quark. In consequence of the refinements and achievements of the experimental results, there has been also progress on the theoretical side in order to explain the experimental outcome.

In nature any type of quark cannot exist as a free particle. In particle collisions, where constituents of the protons scatter on each other with a large momentum transfer, only

bound quark-systems, called hadrons, and their decay products are registered by the detector. These particles emerging from the collision are not uniformly distributed but collimated in narrow cones. These bundles of particles are known as jets. They are reconstructed with the help of the particle detector and used to probe the properties of the original proton constituent, which took part in the scatter during the collision, and the underlying dynamic of the interaction. One impact of the heavy mass of the bottom-quark is that there is a more direct connection between the observed quantities and the underlying dynamics.

In this analysis, the production of bottom-quark pairs in proton-proton collisions at the LHC is studied by identifying jets with bottom-quark content. The data used for this study has been recorded by the ATLAS experiment in 2010. In Chapter 2 the elementary particles and their interactions as described by the Standard Model are presented as well as the Monte-Carlo techniques used to generate a realistic event sample of such interactions. Chapter 3 focuses on the bottom-flavour production in hadronic collisions and closes with an overview on the history of the bottom-quark. Chapters 4 and 5 describe the LHC machine and running configuration, the ATLAS experiment and the reconstruction methods used. In Chapter 6 the techniques for identifying b-jets are discussed, Chapter 7 presents the selection criteria for the measurement of the production rate of bottom-quark pairs. Chapter 8 evaluates the main components of this measurement: the jet energy scale and the b-tagging calibration. Chapter 9, finally, presents the cross section measurement in data, followed by the description of the next-to-leading order predictions (Chapter 10) which are subsequently compared to data in Chapter 11. After quantifying the agreement between data and the current QCD theory, it is searched for hints on new phenomena in the dijet mass spectrum. Furthermore, model independent limits are calculated for resonance signals with a Gaussian shape. With the conclusions from this analysis and the outlook on the LHC era Chapter 13 closes the thesis.

2 Theoretical introduction

2.1 The Standard Model

The Standard Model of particle physics is our current understanding of elementary particles and all forces except gravity [1]. This theory contains all known elementary particles which are believed to be the fundamental particles without an internal structure and is able to describe the electroweak and strong interactions between them. Its understanding is mainly derived from three sources of experimental information: scattering experiments, measurements of decays and of bound states. Since the 1960s this theory has been evolving and withstanding every experimental test. The Standard Model has become an extraordinarily well understood foundation to describe the universe.

Despite all its success there are short-comings. The Standard Model does not incorporate the gravitational force nor an explanation for the physics of dark energy, which is the most accepted theory to describe our expanding universe. The remarkable difference measured in the rotation curves of galaxies provides evidence for matter which is not seen by regular observation methods. There are no particles known to the Standard Model which could make up for this effect. The theory also includes some unnatural hierarchies between couplings and masses which cannot be accommodated for in the Standard Model. Without a doubt, our profound knowledge has to be extended by new physics with the help of the upcoming experiments.

In the following sections the elementary particles and fundamental forces which are described by the Standard Model are introduced. The calculation and modelling of scattering processes of hadrons is explained and finally new physics models are discussed briefly.

2.1.1 The elementary particles

All matter is built from spin $\frac{1}{2}$ particles: six leptons (l), six quarks (q) and their antiparticles ($\bar{l}$, $\bar{q}$). Generally, particles with spin $\frac{2n+1}{2}(n = 0, 1, 2, ...)$ are known as fermions. The stable material surrounding us consists of the lightest quarks and leptons: up-, down-quarks, electrons and their partner neutrino. These particles and their antiparticles are the so-called first generation. The second and third generations are heavier copies of the first one. The force carriers of the strong (gluons g) and electroweak (Z,$W^{\pm}$ and γ) interaction have spin 1. Generally, particles with integral spin are known as bosons. All mentioned particles are shown in Fig. 2.1. Within this thesis natural units are used, i.e. $c = \hbar = 1$. The quarks (coloured red in Fig. 2.1) carry fractional electric charges. All up-type quarks (up (u), charm (c) and top (t)) carry charge $+\frac{2}{3}|e|$ and all down-type quarks (down (d), strange (s) and bottom(b)) $-\frac{1}{3}|e|$. The masses of the quarks increase from left to right. The quark types, or flavours, are coarsely separated into light-flavour (light) quarks consisting of u-, d- and s-quark with masses in the region of MeV and into heavy-flavour quarks with masses in the GeV regime covering the remaining flavours. The heaviest quark is the top-quark with a mass of 173.2 ± 0.9 GeV [2]. Besides the electric charge all quarks carry

a colour charge: blue, red or green.

There are two types of leptons, leptons with electric charge $\pm 1\ e$ and neutral leptons known as neutrinos (ν). The neutral neutrino and charged lepton of the same flavour are paired. The masses of the charged leptons increase from left to right, as well. The tau lepton (τ) is the heaviest lepton with a mass of 1.78 GeV [3]. In the Standard Model the neutrinos are massless but the observation of neutrino-oscillation indicate a non-zero mass. The measurement of the β-spectrum near its endpoint has set an upper limit on the neutrino masses of the order of $\mathcal{O}(1\ \text{eV})$ [4]. The photon γ, the $W^\pm$- and Z-boson are the force carriers of the electroweak interaction. The photon is massless, the other electroweak bosons carry masses of 80.399$\pm$0.023 GeV for the W-bosons and 91.1876$\pm$0.0021 GeV for the Z-boson [3]. The mediators of the strong interaction are the gluons which are massless and carry a colour charge. The yet unmentioned particle is the Higgs boson H which is until now only predicted to exist. The Higgs boson and the Higgs field are postulated to explain the observed mass hierarchies.
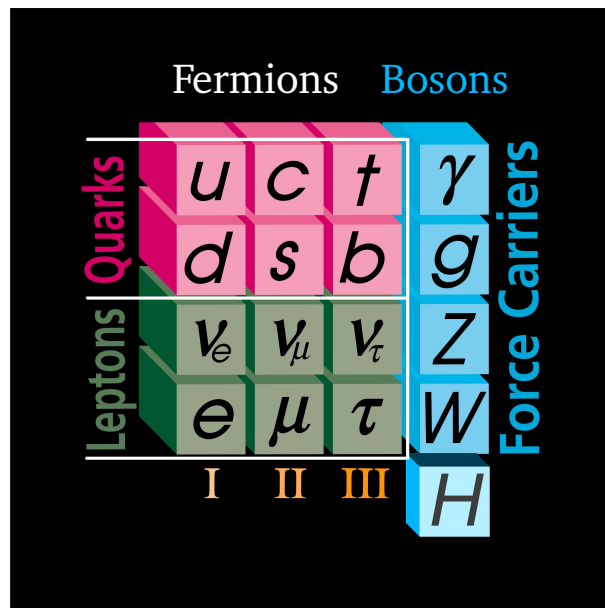
Figure 2.1: *The Standard Model of elementary particles including the predicted Higgs boson which is to be found. Original picture taken from Ref. [5].*

2.1.2 Electroweak interactions

The formalism of the electroweak interactions unites the electromagnetic and the weak force. Compared to the strong force which is discussed in the following chapter the electromagnetic and especially the weak force exhibit a smaller strength. Table 2.1 shows a rough estimate of the relative force strengths evaluated for $q = 1$ GeV. The strong interaction as well as the electromagnetic interaction form bound states of fermions, the weak interaction does not. It is nevertheless relevant for many decay processes in nature. The electromagnetic interactions can be reduced to the elementary process shown in Fig. 2.2: a charged particle, e.g. an electron, emits or absorbs a photon.[1]

The electromagnetic force has unlimited range due to the massless mediator (photon).

---

[1]For all Feynman-graphs the time will flow from left to right.

Force	Strength	Theory	Mediator
Strong	1	Chromodynamics	Gluon
Electromagnetic	10^{-2}	Electrodynamics	Photon
Weak	10^{-6}	Flavordynamics	$W^{\pm}$, Z

Table 2.1: *Overview on the fundamental forces in the Standard Model. Taken from Ref. [6].*
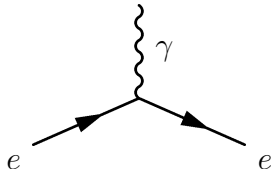
Figure 2.2: *Photon emission (or absorption).*

The photon couples to the electric charge e and the coupling constant of each vertex is $\sqrt{\alpha}$. In the limit of non-relativistic electron-proton scattering the Rutherford formula provides a good description of the kinematics (neglecting spin):

$$\frac{d\sigma}{d\Omega} = \frac{e^4}{4m_e^2 v^4} \frac{1}{\sin^4(\theta/2)}, \tag{2.1}$$

where m_e and v are the mass and the velocity of the electron respectively. The angle θ is called the scattering angle and represents the deflection of the electron trajectory.

In the relativistic description of electron-electron scattering (Feynman-graphs Fig. 2.3) by Quantum Electrodynamics (QED), the fermions are described by spinors u representing solutions of the Dirac equation, and the photon is described by a field strength tensor and the Klein-Gordon equation for massless particles. The momentum transfer q between the incoming electrons is given by the momentum difference of $p_1 - p_3$ for the un-crossed Feynman-graph (Fig. 2.3 left) and $p_1 - p_4$ for the crossed Feynman-graph (Fig. 2.3 right) respectively. The amplitude $\mathcal{M}_1$ of the un-crossed Feynman-graph can be evaluated as follows with the coupling constant $g_e = \sqrt{4\pi\alpha}$:

$$\begin{aligned}
\mathcal{M}_1 = (2\pi)^4 \int d^4 q \delta^{(4)}(p_1 - p_3 - q)\delta^{(4)}(p_2 - p_4 + q) \\
[\bar{u}(p_3)(ig_e\gamma^{\mu})u(p_1)] \frac{-ig_{\mu\nu}}{q^2} [\bar{u}(p_4)(ig_e\gamma^{\nu})u(p_2)] .
\end{aligned} \tag{2.2}$$

The amplitude $\mathcal{M}_2$ for the crossed Feynman-graph can be derived from $\mathcal{M}_1$ by exchanging the momenta $p_3 \leftrightarrow p_4$. There are four important parts in the amplitude formula: the phase space integration, the delta distributions establishing energy and momentum conservation, the vertex factors between the spinors of the two leptons and the photon propagator connecting both vertices. The overall squared amplitude $|\mathcal{M}|^2 = |\mathcal{M}_1 + \mathcal{M}_2|^2$ includes contributions from both Feynman-graphs and an interference term between them.

The common aspect of the weak and electromagnetic interaction is that each electromagnetic process contains a neutral weak contribution from a Z-boson exchange. As long as the
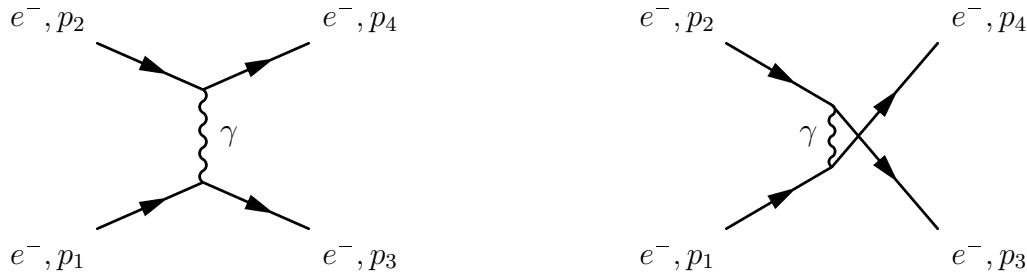
Figure 2.3: *Electron-electron scattering.*

momentum transfer of a scattering process is far below the large mass of the Z-boson, the electromagnetic part is overwhelming. For completeness, in the Feynman-graphs in Fig. 2.3 the Z-boson has to be added to the force mediating photon γ. The amplitude for the Z-boson exchange is calculated similar to Eq. 2.2 but with a vertex factor $\frac{ig_Z}{2}\gamma^\mu(c_V^f - c_A^f\gamma^5)$ and the propagator $\frac{i(g_{\mu\nu} - q_\mu q_\nu/(m)^2}{q^2 - (m)^2}$. The vertex factor also exhibits an important aspect of the weak interaction, it adds an axial vector $(\gamma^\nu\gamma^5)$ to a vector coupling (γ^ν) and therefore does not conserve parity. The consequence of parity violation is that the weak interactions distinguish between left-handed and right-handed particles. The strength of the Z-boson's axial and vector couplings described by c_V^f and c_A^f is particle dependent. All fermions are affected by the weak force. Lepton-neutrino scattering confirmed (Fig. 2.4) the existence of the Z boson. The large masses of both the charged mediators and the neutral mediator reduce the range of the weak forces to very small distances. These weak currents are lepton-flavour conserving (lepton number) but can change the quark flavour and also connect quarks of different generations. Examples of such processes are shown by the Feynman-graphs in Fig. 2.5: the neutron decay into a proton changes the d-flavour into a u-flavour (left) and the generation transition of s- to a u-flavour (right). These graphs are based on a simplified picture where the other quarks act as spectators. The pure underlying quark-process-coupling to a W-boson is given by $G_W = \frac{-ig_W}{2\sqrt{2}}\gamma^\nu(1 - \gamma^5)$ and analogously $g_w = \sqrt{4\pi\alpha_w}$. The W boson only interacts with left-handed fermions and right-handed antifermions, described by $c_V^f = c_A^f = 1$. The basic idea to describe the observed generation transitions is that the weak eigenstates (d', s' and b') are not identical to the quark mass states (d, s and b) but rather linear combinations of them. The mechanism explaining these generation transitions is written down in the so-called CKM-matrix (Cabibbo-Kobayashi-Maskawa), which connects the weak and the mass eigenstates via this unitary matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} \mathbf{V_{ud}} & V_{us} & V_{ub} \\ V_{cd} & \mathbf{V_{cs}} & V_{cb} \\ V_{td} & V_{ts} & \mathbf{V_{tb}} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.3}$$

The diagonal elements describe transitions within one generation and are all nearly one. The non-diagonal conversions, especially between first and third generation, are highly suppressed and the suppression factor V_{ij} alters the effective coupling as indicated in Fig. 2.5.

The electroweak unification explains the electromagnetic and weak interactions by one
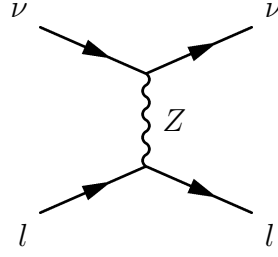
Figure 2.4: *Lepton-neutrino scattering.*

Coupling: $G_W V_{ud}$ '<'   Coupling: $G_W V_{us}$

Figure 2.5: *Flavor changing charged currents: within the first generation (left) and between first and second generation (right).*

single theoretical system and one single coupling given by the electric charge e. At high energies (or small distances), both interactions are symmetric (possess the same strength) and at low energies the symmetry between these interactions is broken due to the massive mediator bosons of the weak processes. This is called electroweak symmetry breaking. The success of this theory lies in the connection the electromagnetic and weak coupling constants and in fact correctly predicts the particle-dependent Z-boson couplings. This formalism introduces a new quantum number: the weak isospin T and its third component T_3. The electroweak unification introduces four mediating bosons: an isospin-triplet W_μ^1, W_μ^2, W_μ^3 with the coupling constant g and an isospin-singlet B_μ with a different coupling strength g'. Right-handed fermions do not interact with the W-bosons and are described as isospin-singlets. Table 2.2 lists the new quantum number T for the multiplets of fermions.

	Multiplets of fermions			T	T_3	$z_f = \frac{Q}{e}$
Leptons	$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L$	$\begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L$	$\begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L$	$\frac{1}{2}$	$+\frac{1}{2}$ $-\frac{1}{2}$	0 -1
	e_R	μ_R	τ_R	0	0	-1
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}_L$	$\begin{pmatrix} c \\ s' \end{pmatrix}_L$	$\begin{pmatrix} t \\ b' \end{pmatrix}_L$	$\frac{1}{2}$	$+\frac{1}{2}$ $-\frac{1}{2}$	$+\frac{2}{3}$ $-\frac{1}{3}$
	u_R d_R	c_R s_R	t_R b_R	0	0	$+\frac{2}{3}$ $-\frac{1}{3}$

Table 2.2: *Overview on weak isospin for all fermions.*

The two known neutral currents (Z and γ) are described by orthogonal linear combi-

nations of W_μ^3 and B_μ. Their mixing is described by the electroweak mixing angle θ_W (known as the Weinberg-angle) and is not predicted by theory, but is an experimentally determined input value to the theory:

$$|\gamma\rangle = \cos\theta_W |B_\mu\rangle + \sin\theta_W |W_\mu^3\rangle, \tag{2.4}$$

$$|Z\rangle = -\sin\theta_W |W_\mu^3\rangle + \cos\theta_W |B_\mu\rangle. \tag{2.5}$$

The charged W-bosons are represented by a combination of W_μ^1 and W_μ^2. The mixing angle [3] in the $\overline{\text{MS}}$-scheme [7] is:[2]

$$\sin^2\theta_W = 0.23116 \pm 0.00013. \tag{2.6}$$

The electric charge and the weak coupling constants are connected as follows:

$$\frac{g'}{g_w} = \tan\theta_W, \ g_e = g_w \cdot \sin\theta_W. \tag{2.7}$$

The coupling strength of the Z-boson to a fermion is given by the following equation:

$$g_Z(f) = \frac{g_w}{\cos\theta_W} \cdot (T_3 - z_f \sin^2\theta_W). \tag{2.8}$$

A mixing of different states should only happen if the involved states have similar energies and hence masses. This is a shortfall of the theory of the electroweak unification discussed so far and is theoretically explained by a spontaneous symmetry breaking. Therefore, a complex doublet of scalar fields is predicted, the so-called Higgs field. The choice of the vacuum expectation value of the Higgs field breaks the $SU(2) \otimes SU(1)$ symmetry. The isospin triplet W_μ absorbs three out of four Higgs field components and they become massive. The W-B mixing (cf. Eq. 2.5) results in a massless photon and a massive Z. This leaves one single neutral Higgs boson remaining which is searched for. The mass of the Higgs is a free parameter of the model but different sources of electroweak precision data and direct searches in the context of the Standard Model have excluded the Higgs Boson in various mass ranges. Recent results from the ATLAS experiment exclude the Higgs Boson in the context of the Standard Model in the mass ranges from 110.0 GeV to 117.5 GeV, 118.5 GeV to 122.5 GeV, and 129 GeV to 539 GeV [8]. The coupling strength of the Higgs field to all particles is proportional to their masses, so it preferably decays into the heaviest available quark or lepton pair or pair of bosons.

2.1.3 Strong interactions

The theory of the strong interactions is Quantum Chromodynamics (QCD) [9]. It describes the interactions between quarks and gluons. The gluons are the force mediating, massless bosons for the strong interaction acting on quarks, antiquarks and other gluons. The quarks are the physical representations of the flavour $SU(3)$ gauge theory. The gluons couple to the so-called colour charge: quarks have three different possible colours (green (g), blue(b) and red (r)) and gluons have 8 different possible combinations (colour and anticolour) of the three fundamental colours. The gluons are represented by the Gell-Mann

---

[2]$\overline{\text{MS}}$-scheme is the modified minimal subtraction renormalisation scheme. Renormalisation schemes regulate UV divergences like closed loop momentum integrals to render such integrals finite. The chosen scheme defines the counterterms and the set of renormalisation conditions which are introduced to treat these divergences.

$\bar{Q}, \bar{b}$ Q', r

$r\bar{b}$

Q, r $\bar{Q}', \bar{b}$
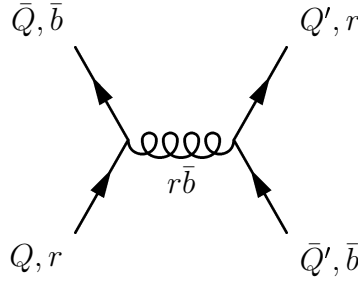
Figure 2.6: *Strong interaction between quarks Q with colour charge red (r) and $\bar{Q}$ with colour charge antiblue (b). The colour charge is conserved.*

matrices (3×3 matrices) which are an analogy to the Pauli-matrices of $SU(2)$. Unlike the photons in QED gluons also carry a colour charge and therefore the field strength tensor $F^A_{\alpha\beta}$ contains a self-coupling term $gf^{ABC}A^B_a A^C_\beta$ leading to triplet and quartic gluon vertices:

$$F^A_{\alpha\beta} = \left[\delta_\alpha A^A_\beta - \delta_\beta A^A_\alpha - g_s f^{ABC} A^B_\alpha A^C_\beta \right], \tag{2.9}$$

where A^A_β represents the gluon fields, g_s the strong coupling constant and f^{ABC} the structure constants. The indices A, B, C run over the eight colour degrees of freedom of the gluon field. A complete quark state specification requires a Dirac spinor and a three-element vector giving its colour. A strong interaction as shown in Fig. 2.6 changes the colour state of a quark, not its flavour. The matrix element is calculable in analogy to Eq. 2.2 in QED. For this purpose the coupling constant has to be exchanged ($g_e \leftrightarrow g_s(\alpha_s)$) and an extra factor carrying the colour charge of the gluon at each vertex needs to be accounted for. Important experimental results were that the constituents inside a hadron behave like point-like free particles if they are probed at high momentum transfers and that all observed particles are colourless bound states. These two concepts in QCD are known as asymptotic freedom and confinement. The behaviour of quarks and gluons as point-like free constituents inside hadrons founded the parton model. Partons are identified with quarks and gluons. This parton model is based on the diminishing coupling constant at high energy (asymptotic freedom). The confinement, on the other hand, implies that only hadrons, not individual quarks, can be observed as free particles. At low energy and large separation of quarks the coupling gets strong enough to create pairs of quarks and antiquarks. The phenomenon of asymptotic freedom is described by the running of the coupling constant of the strong interaction α_s. Here the dependence of a physical observable R on the momentum transfer Q^2 and a mass scale μ (renormalisation scale) to remove ultraviolet divergences is absorbed into the running of the coupling constant $\alpha_s(Q^2)$. The running of the coupling constant itself is determined by the renormalisation group equation:

$$Q^2 \frac{\partial \alpha_s}{\partial Q^2} = \beta(\alpha_s), \tag{2.10}$$

where $\beta(\alpha_s)$ is the β-function with the expansion in QCD [6]: $\beta(\alpha_s) = -b\alpha_s^2(1 + \mathcal{O}(\alpha_s))$. If both $\alpha_s(Q^2)$ and $\alpha_s(\mu^2)$ are in the perturbative region the renormalisation group equation can be solved by neglecting higher order terms. The solution is then given by:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2) b \ln \frac{Q^2}{\mu^2}}. \tag{2.11}$$

For increasing Q^2 the running coupling $\alpha_s(Q^2)$ decreases logarithmically. The perturbative region is reached at scales $Q \approx 1 - 2$ GeV. In contrast to QED where the coupling
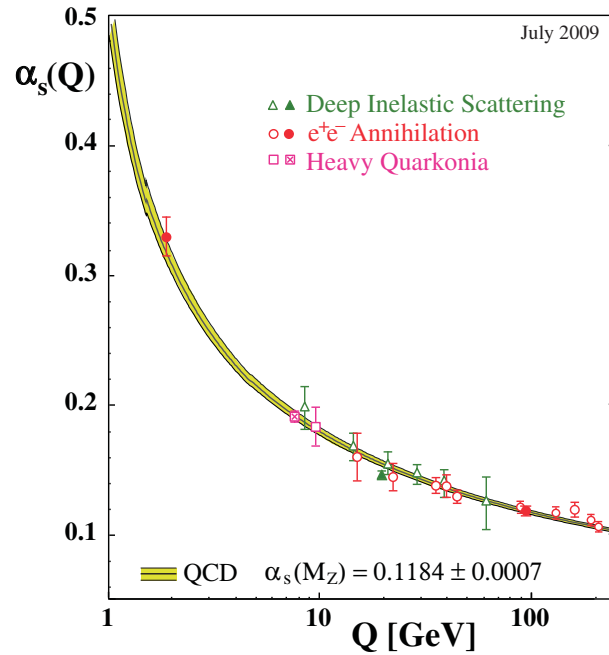
Figure 2.7: *Summary of measurements of α_s as a function of the respective energy scale Q. Plot is taken from Ref. [10].*

decreases with decreasing Q^2, a positive sign of b as found for Eq. 2.11 lets the strong coupling increase with decreasing Q^2 as seen in Fig. 2.7. Only in regions, where α_s is sufficiently small, processes can be calculated with means of perturbative QCD (pQCD). In the regime of large distances the theory is non-perturbative and pQCD is inapplicable. However static properties of QCD, like hadron masses and lifetimes [3], are calculable with methods of lattice QCD.

The process of building colourless bound states which are observed in nature from single produced partons is called hadronisation. As α_s increases the partons reorganise themselves and multiply to colour-singlet states illustrated in Fig. 2.8. Hadronisation is a dynamic
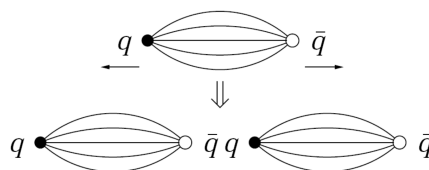
Figure 2.8: *Visualisation of the principle of confinement by creating extra partons.*

process on a large timescale involving larger values of α_s which can only be described by phenomenological models and is discussed in Section 2.3.3.

2.1.4 Hadron-hadron interactions

Hadron-hadron collisions may intuitively be separated into two classes:

- Soft interactions with small momentum transfers: These have typically large cross

sections of the order $\mathcal{O}(\text{mb})$ at the LHC and they are sensitive to long-distance effects. Due to their non-perturbative nature phenomenological models are used for their description. They are divided into several sub-categories: elastic, diffractive and soft inelastic interactions.

- Hard interactions with large momentum transfers probing the internal structure of the hadrons: These processes have, compared to the first class, small cross sections. They are the main focus in this analysis and described in the following section. The large momentum transfer allows them to be described perturbatively.

The current predictions for physical observables in hadron-hadron interactions are based on the fundamental attribute of QCD, namely factorisation. Therefore a hadron-hadron interaction can be decomposed into different parts illustrated in Fig. 2.9. Factorisation states that these different components can be treated independently from each other. The component which can be calculated in perturbative QCD is called the hard scatter. The other components are the parton showers (PS), the underlying event (UE), the beam remnants and the hadronisation (HAD). In the parton shower approach possible histories of the initial and final states are handled. The prediction from parton showers represents an approximate perturbative treatment of QCD dynamics of soft and collinear configurations to all orders. A parton shower converts a highly virtual parton into a low virtual parton with either positive or negative virtuality. With positive virtuality the state describes a parton after final-state radiation at energies of about 1 GeV ready to hadronise, a state with negative virtuality describes a parton before initial-state radiation emerging from the incoming hadron. Hadronisation happens subsequently after the final-state parton shower at energies of the order 1 GeV and contains the transition from a partonic final state to a hadronic observable. Therefore detectable event features are described by the means of collimated sprays of hadrons known as jets, which ideally are associated with the primary partons. The principles used to identify jets are presented in Section 5.1. The beam remnants are part of the non-perturbative components and describe the remaining constituents of the hadrons not taking part in the hard scatter. The underlying event, which is not illustrated in Fig. 2.9 for readability, treats the additional activity in an event which is not directly associated with the hard scatter. This component describes multiple parton interactions (MPI) which would produce additional hard scatters. The underlying event is an important element of the environment at high-energy experiments polluting the hard scatter.

An inelastic hadron collision results in a short-distance scattering process described by the cross section σ_{ab} between the quarks and gluons producing new partons with momenta p_3 and p_4 as shown on the left side of Fig. 2.9. The parton density functions $f_i(x_i)$ (PDF) describe the probability of finding a constituent i with momentum fraction x_i defined as $p_i = x_i \cdot P_i$ in the hadron. These parton density functions and the formalism of the hard scatter of a hadron-hadron interaction including important observables in high-energy physics are discussed in detail in the next Sections 2.1.4.1 and 2.1.4.2. The parton shower, the underlying event, remaining beam remnants and the hadronisation are needed to fully describe an event. The process of parton branching within a parton shower is calculated with pQCD. The subsequent generation of such branchings creates a parton shower and is taken care of by a Monte-Carlo generator. The description of the underlying event, the beam remnants and hadronisation are based on different models. All these components of a hadron-hadron interaction in context of Monte-Carlo techniques are discussed in detail in Section 2.3.
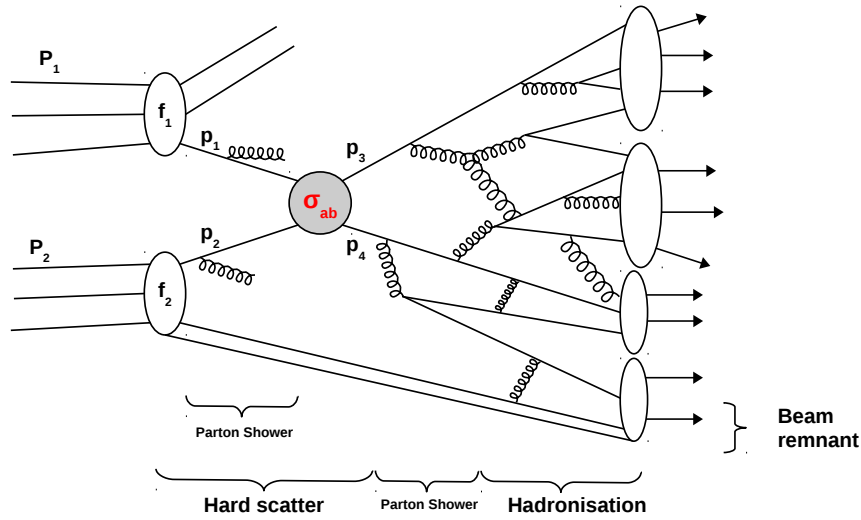
Figure 2.9: *Illustration of a hadron-hadron interaction with incoming hadrons with momenta P_1 and P_2 and outgoing final-state hadrons.*

2.1.4.1 Parton density functions

The parton density functions are part of the non-perturbative region of QCD and their functional form must be extracted from fits on a variety of experimental data. Two important groups doing such global fits are the CTEQ [11] and MSTW [12] collaborations. At leading order PDFs describe the longitudinal momentum fractions x at momentum transfer Q^2 carried by the different partons in a given hadron. In a qualitative picture the proton consists of three valence quarks (uud) which are responsible for the electric charge. In addition, measurements proof the existence of a sea of quark-antiquark pairs and gluon constituents. The probability of a specific process in a hadron-hadron collision as seen in Fig. 2.9 is the combination of the partonic cross section and the PDFs. PDF fits from MSTW for the proton at two different scales Q^2 are shown in Fig. 2.10. Both of these PDF fits show the distributions of the momentum fraction x of the valence quarks (u, d), the sea quarks ($\bar{u}, \bar{d}, c, \bar{c}, s, \bar{s}, b, \bar{b}$) and of the gluon constituents (g).[3] The PDFs are a function of the momentum fraction x as well as of the momentum transfer Q^2 at which the hadron is probed. The gluon density, probed at different momentum transfer scales Q^2, is dominating at low values of x. In a hadron-hadron interaction processes involving incoming low x gluons are favoured. As the probing scale increases the number of gluons which are resolved into quark-antiquark pairs with small momentum fraction increases. This enhances the perturbatively produced quark-antiquark pairs contributing to the sea distribution. On the contrary, the fraction of partons carrying a large momentum fraction (valence quarks) decreases because $q \rightarrow q + g$ processes are resolved. Both effects cause the densities to fall more steeply and to populate lower regions in x. The evolution of the PDFs is determined by the set of DGLAP equations [13] which are solved in pQCD.

---

[3]Strictly speaking also u and d quarks are of course part of the seq-quarks.

The PDFs should be extracted at the same fixed-order as the partonic cross section is calculated for factorisation.
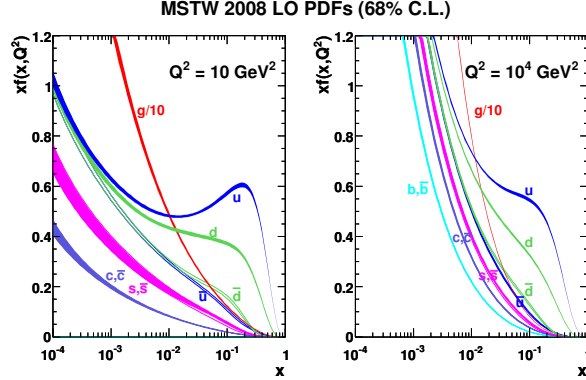
Figure 2.10: *Extracted PDFs at leading order for the proton from the MSTW collaboration taken from Ref. [12].*

2.1.4.2 The hard scatter

The factorisation theorem of QCD allows the treatment of the hard scatter as a pure perturbative strong interaction by cleanly separating long- and short-distance processes by a factorisation scale μ_f. Processes with smaller scales than the factorisation scale are factored into the parton distribution functions $f_i(x_i, \mu_f^2)$ as being part of the hadron structure. Therefore the cross section for a hard scatter initiated by two hadrons with four-momenta P_1 and P_2 can be written as:

$$\sigma(P_1, P_2) = \sum_{ij} \int dx_1 dx_2 f_1(x_1, \mu_f^2) f_2(x_2, \mu_f^2) \sigma_{ij}\left(p_1, p_2, \alpha_s(\mu_r^2), \frac{Q^2}{\mu_r^2}, \frac{Q^2}{\mu_f^2}\right). \qquad (2.12)$$

The hard scale is given by Q. The renormalisation scale μ_r was introduced in Section 2.1.3. σ_{ij} is the partonic cross section of the scattering of partons of types i and j. For a specific final state with partons k and l the cross section includes the matrix element $M_{ij \to kl}$ and the phase space terms. The incoming hadrons have a spectrum of longitudinal momenta determined by the PDFs. As a consequence, the centre-of-mass frame of the hard scatter is boosted with respect to the incoming hadrons. Therefore, it is convenient to use variables with additive boost transformation along the z direction like the rapidity y or pseudorapidity η:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}, \ \eta = -\ln \tan \frac{\theta}{2}, \qquad (2.13)$$

where θ is the polar angle. The polar angle θ and the azimuthal angle ϕ are defined with respect to the beam direction along the z-axis and x-axis, respectively. The outgoing partons can be fully described by the rapidity y, the transverse momentum p_T and the azimuthal angle ϕ:

$$p^\mu = (E, p_x, p_y, p_z) = (m_T \cosh y, p_T \sin \phi, p_T \cos \phi, m_T \sinh y), \ m_T = \sqrt{p_T^2 + m^2}. \quad (2.14)$$
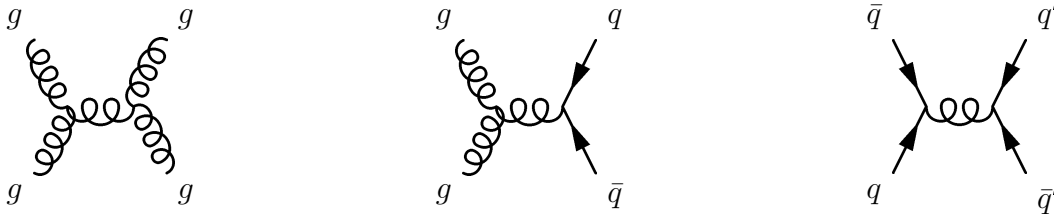
Figure 2.11: *A subset of possible $2 \rightarrow 2$ scattering processes: (left) gluon fusion, (middle) gluon annihilation, (right) quark-antiquark annihilation.*

Because of momentum conservation, the two final-state particles k and l for a $2 \rightarrow 2$ scattering are produced with equal momenta of opposite direction in the centre-of-mass frame. If the transverse momenta of the incoming partons is negligible, those two partons are back-to-back in azimuth ($\Delta\phi = \pi$) and balanced in transverse momentum in the laboratory frame. The Feynman-graphs in Fig. 2.11 show a subset of possible $2 \rightarrow 2$ processes.

The partonic cross section for all possible $2 \rightarrow 2$ scatterings with quarks and gluons in the initial and final state is:

$$\frac{d^3\sigma}{dy_3 dy_4 dp_T^2} = \frac{1}{16\pi\hat{s}^2} \sum_{i,j,k,l=q,\bar{q},g} \frac{f_i(x_1,\mu_f^2)}{x_1} \frac{f_j(x_2,\mu_f^2)}{x_2} \times \overline{\sum} |M(ij \rightarrow kl)|^2 \frac{1}{1+\delta_{kl}} \qquad (2.15)$$

with $\overline{\sum}$ including the average and sum over initial- and final-state spins and colours respectively and $\hat{s} = (p_1 + p_2)^2$ being one of the Mandelstam variables. The energy-momentum conservation fixes the values of x_1 and x_2 when the transverse momenta and the rapidities of the outgoing partons are measured. The following relation holds:

$$x_1 = \frac{m_T}{\sqrt{s}}(e^{y_3} + e^{y_4}), \; x_2 = \frac{m_T}{\sqrt{s}}(e^{-y_3} + e^{-y_4}). \qquad (2.16)$$

The scattering angle θ^* in the centre-of-mass frame is directly correlated to the rapidity difference $y_3 - y_4$ of the two outgoing partons in the laboratory frame. Thus, the angular distribution is sensitive to the underlying $2 \rightarrow 2$ matrix element. For convenience, the angular distribution is expressed in the variable $\chi = \exp|y_3 - y_4|$ to study deviations from the underlying QCD kinematics [14]. The relation to the centre-of-mass scattering angle is:

$$\chi = \frac{1 + \cos\theta^*}{1 - \cos\theta^*}. \qquad (2.17)$$

In addition to the large gluon densities in the proton, the dominant parton-parton interaction, comparing the matrix elements, is the scattering process $gg \rightarrow gg$ as shown in Fig. 2.12. Therefore, proton-proton interactions at high energies are dominated for a large energy range by gluon-gluon scatterings. The full matrix element for this process is the squared sum of all four Feynman-graphs shown in Fig. 2.12. The singularities in the cross section arising from t- and u-channel lead to the dominance of those scattering channels. The χ-dependence of the subprocess $gg \rightarrow gg$ which was calculated and is presented in Fig. 2.13, is approximately constant over a large range. Therefore, $d\sigma_{gg \rightarrow gg}/d\chi \approx$ constant is a flat distribution. This linearity is only disturbed at $\chi \lesssim 3$ where s-channel contributions become more important. This behaviour is in agreement with the underlying kinematics of pure Rutherford scattering (peaking at small θ^*) which is a t-channel exchange of a spin-1 particle just like the t-channel shown in Fig. 2.12.

Figure 2.12: *Diagrams for $2 \to 2$ processes involving two gluons in the initial and final state. From left to right: s-channel, t-channel, u-channel and quartic vertex.*
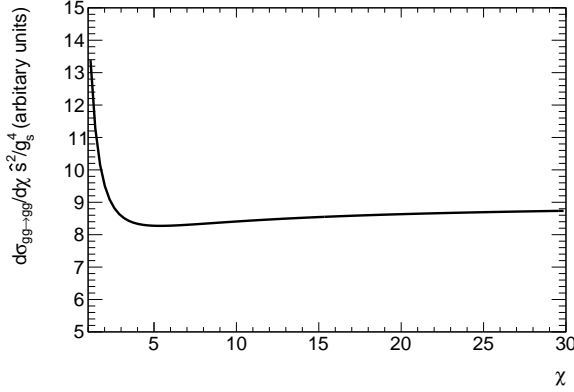
Figure 2.13: *The χ-dependence of the invariant matrix element squared $\frac{d\sigma}{d\chi} \frac{\hat{s}^2}{g_s^4}$ for the subprocess $gg \to gg$.*

Furthermore, the contributions from the other important $2 \to 2$ subprocesses can be estimated qualitatively with respect to the dominant $gg \to gg$ process. This is due to the similar behaviour of $\frac{d\sigma}{d\chi} \frac{\hat{s}^2}{g_s^4}$. Before the inclusion of the PDFs, with the help of colour factors, the contributions from the other $2 \to 2$ subprocesses behave effectively like:

$$ qq \to qq : gq \to gq : gg \to gg \approx 1 : \frac{9}{4} : \left(\frac{9}{4}\right)^2 . \tag{2.18} $$

Therefore, the angular distribution of χ is insensitive to the relative weighting of each process and thus uncertainties arising from the PDFs are minimised.

The subprocess $q\bar{q} \to q'\bar{q}'$ (quark-antiquark annihilation) is only realised by the s-channel exchange of a gluon. Therefore this subprocess exhibits the specific angular distribution shape of the s-channel whereas the subprocesses in Eq. 2.18 are a mixture of different channels. This becomes important when searching for new particles or resonances. Fig. 2.14 shows the χ dependence of the matrix elements squared $= \frac{d\sigma}{d\chi} \frac{\hat{s}^2}{g_s^4}$ for the process $q\bar{q} \to q'\bar{q}'$. Shown is also the shape comparison to the subprocess $gg \to gg$ (dashed line) which is scaled to converge with the matrix element from $q\bar{q} \to q'\bar{q}'$ for large χ. Unlike the dominant t-channel subprocesses shown before, in pure s-channel processes the outgoing quarks are correlated in rapidity and $d\sigma/d\chi$ peaks at low χ.

The currently available precision of calculations in pQCD of hadron-hadron interactions is of the order $\mathcal{O}(\alpha_s^3)$, called next-to-leading order (NLO). Fig. 2.11 shows the so-called Born-level Feynman-graphs, these are leading order (LO) contributions. To obtain the NLO predictions, corrections need to be added: $\sigma_{\text{Born}} + \alpha_s(\mu_r^2) \cdot \sigma_1$. Fig. 2.15 shows the necessary processes and matrix elements at NLO. The first correction contributing to σ_1 is coming from one additional gluon radiation off any of the outgoing partons, the so-called
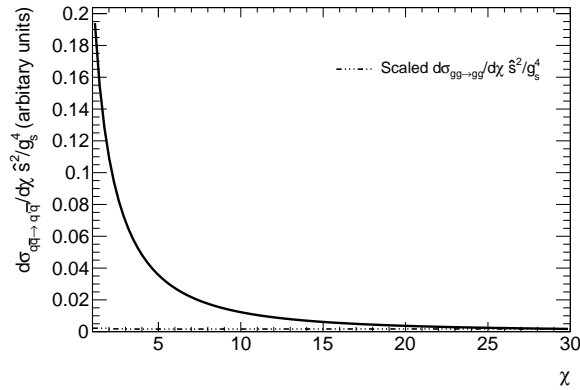
Figure 2.14: *The χ-dependence of the invariant matrix element squared $\frac{d\sigma}{d\chi}\frac{\hat{s}^2}{g_s^4}$ for the subprocess $q\bar{q} \rightarrow q'\bar{q}'$. The invariant matrix element squared for the subprocess $gg \rightarrow gg$ is shown as dashed line. It is scaled to converge at large χ.*

real emission. These real emission contributions correspond to the lowest order matrix element of $2 \rightarrow 2 + g$ or generally $2 \rightarrow 3$. The last contribution is referred to as virtual emission and involves the emission of a gluon which is re-absorbed by the other parton (as shown in Fig. 2.15) or by its emitter itself. Both additional terms include infrared divergences (soft and collinear singularities) which cancel each other. The importance of considering higher orders in pQCD calculations is established by the fact that NLO corrections can lead to contributions as big as the leading order prediction.
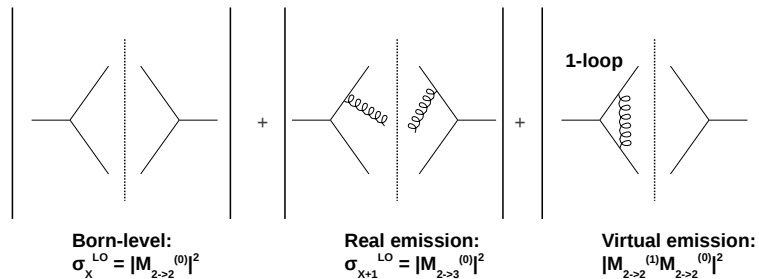
Figure 2.15: *Cross section up to NLO (for simplicity real emission from other leg and loops on the same leg are not drawn).*

2.2 Extensions of and departures from the Standard Model

The observed deficiencies of the Standard Model lead to new theoretical developments. There are mainly two categories of developments: pure extensions of the Standard Model and new models in which the Standard Model is only an effective theory. All of these theories have one thing in common: they urge to be in agreement with all known phenomena. In the energy regime of past experiments, the Standard Model predictions were in consent with the experimental data. Currently, new kinematic regions are explored by the Large

Hadron Collider. There is a large variety of new physics models which can explain known experimental phenomena by new underlying theories. Grand Unified Theories (GUT) predict that at high-energy, the three gauge interactions of the Standard Model, namely the electromagnetic, weak, and strong interactions, are merged into one single interaction invoking higher dimensional gauge groups or left-right symmetric models [15, 16, 17]. These models could explain the quantisation of charges and the mass hierarchies. Another popular model predicts supersymmetric partners for each particle. For every type of boson (fermion) a corresponding fermion (boson) exists carrying the same internal quantum numbers. Besides solving the hierarchy problem, this model could help solving cosmological questions [18].

Until now quarks are believed to be point-like [19]. The best limits on the internal structures from the Tevatron end at around 900 GeV probing scales of approximately 10^{-17} cm. The Large Hadron Collider enables probing for a substructure in the TeV regime. The observed mass hierarchy and generational structure of the quarks motivate models of compositeness. If a quark is a composite particle, excited states [20] are expected. The Randall-Sundrum model containing extra dimensions can connect the gravitational force with the Standard Model and explain the observed mass hierarchies [21].

The common aspect of these models is the prediction of new heavy particles or resonances produced via s-channel processes. They would decay into stable particles which would be observed by deviations in measured cross sections with respect to Standard Model predictions. A promising channel to search for such effects is observing the production of two energetic partons which are reconstructed as jets. The measurement of the invariant mass of the decay products is sensitive to new phenomena. In particular new resonances and new unstable particles are expected to produce configurations which are more isotropic in rapidity than QCD events which result mostly from spin-1-gluon exchanges in the t-channel. Therefore, the angular distribution as a function of χ is highly sensitive to such signatures and has the advantages of being independent from the underlying decay width and calorimeter energy resolution. Both discussed variables are sensitive to such new phenomena and would show an excess of events (cf. Fig. 2.14) at low χ and high invariant mass.

2.3 Monte-Carlo models and generators

Generally, Monte-Carlo (MC) techniques are based on a random number generator to sample random variables following a probability function. These techniques have become indispensable in high-energy physics firstly because the calculation of a cross section includes a vast number of integrations on phase space, helicities, spins and colours. Secondly realistic QCD predictions of hadronic final states are required to be comparable to experimental data.

There are two sorts of MC event generators: fixed-order and parton shower generators. Their structures are directly linked to the factorisation in pQCD. The fixed-order generators which are used for this analysis are next-to-leading order and they produce parton-level and not hadron-level predictions as observed in experiments. They are discussed separately in Section 10.

Parton shower event generators [22] simulate events to the point of having a fully specified set of final-state hadrons. These are the main tools for experimentalists because their output can be linked to any detector simulation. With their help detector designs and experimental

strategies are developed and optimised. The generation of the hard subprocess is according to the leading order matrix elements providing a good description of wide-angle and high energy particles. The technical implementation is basically the same as for the fixed-order generators. However, to provide a complete picture of a subprocess any fixed order is not sufficient. Consequently, the leading order matrix element has to be complemented by small-angle and/or low energy emissions described by the parton shower to evolve to a final parton state. The hadronisation then forms colourless hadronic states. The following Sections 2.3.1 and 2.3.4 give an overview of the event generators commonly used and their implementation of the parton shower, hadronisation and underlying event. The problems arising from connecting the parton shower and the hard scatter at next-to-leading order is be briefly presented in Section 2.3.5.

2.3.1 Event generators

Two of the main parton shower event generators are PYTHIA (version 6.423) [23] and HERWIG++ (version 2.5.0) [24] and its FORTRAN-sibling HERWIG (version 6.510) [25].[4] Both show great success in reproducing experimental data. But not only the generator and its version is important, also the tune. A tune is an accumulation of settings for event generation, hadronisation and underlying event which are adjusted to a certain sample of experimental data. The default PYTHIA tune is the so-called *ATLAS Minimum Bias Tune 1* (AMBT1) [26], which is the first tune of PYTHIA to LHC data recorded with the ATLAS detector. Similar for HERWIG, the first tune is called *ATLAS Underlying Event Tune 1* (AUET1) [27]. The default HERWIG++ tune is the so-called *LHC Underlying Event Tune 1* (LHC-UE7-1). All mentioned tunes are made with the MRSTLO* [28] PDF set.[5] As long as not stated differently these tunes are called MC10 tunes using a centre-of-mass energy of $\sqrt{s} = 7$ TeV and used throughout the analysis for PYTHIA, HERWIG and HERWIG++.

2.3.2 Parton showers

The parton shower evolution converts a parton with a high virtuality t coming from the hard scatter to a final-state parton of low virtuality t_0 ready for the hadronisation phase. This evolution can either be ordered in terms of the opening angle of the parton branching as it is done in HERWIG and HERWIG++ or the transverse momentum of the process as in PYTHIA. Each parton coming from the hard scatter at scale t undergoes successive steps of branchings by a random generator until it reaches a cut-off scale t_0 below which pQCD no longer is valid and the evolution is terminated. The probability that a parton evolves from scale t_1 to t_2 without a branching is the Sudakov factor. The branching processes are $q \rightarrow gq$, $g \rightarrow gg$ and $g \rightarrow q\bar{q}$. They are described by a set of so-called splitting functions P calculated in pQCD obeying the DGLAP evolution equations. They determine the probability that a parton b which evolves from a scale t to $t + \delta t$ converts into partons a

---

[4]Those version numbers are the default choice. If a different version is used it will be stated explicitly.

[5]These are modified LO PDFs. Cross sections and acceptances calculated with LO PDFs differ in both shape and normalisation from calculations using NLO PDFs. LO* PDFs try to incorporate properties to reduce these deficiencies. In general LO* PDFs should behave as LO as $x \rightarrow 0$ and as close to NLO as possible as $x \rightarrow 1$

and a' carrying momentum fractions z and $(1 - z)$ respectively [6]:

$$P(b \rightarrow a(z) + a'(1 - z) \text{ when } t \rightarrow t + \delta t) = \frac{\delta t}{t} \frac{\alpha_s(t, z)}{2\pi} P_{ab}(z). \qquad (2.19)$$

The concept of parton showers has the advantage that leading order is sufficient for the matrix element describing the hard scatter. Its success is to describe jet sub-structure and event shapes. But there are shortcomings due to its approximate nature. Interference between initial-state and final-state radiation is not handled. Additionally, the emission of hard partons at large angles and high energy is not well accounted for and this leads to deviations in the rate of observed hard partons.

2.3.3 Hadronisation

The hadronisation step of a hadron-hadron interaction is not in the perturbative regime but is needed for full simulation of final states. The hadronisation of a parton is presumed to be universal e.g. not depending on the environment of the production. The formation of hadrons is described by means of phenomenological models which are tuned to experimental data. The function describing hadronisation is called the fragmentation function $f(z)$ parametrised as a function of the momentum fraction carried by the hadron with respect to the parton. Parametrisations for these functions differ depending on the kinematic arguments used to derive them. Also here the DGLAP equations can be used for scale evolution. Similar to PDF fitting these functions are extracted from and tuned by experimental data. Hadronisation models are based on the observation that global features of hadronic systems (multiplicities, angular distributions and inclusive energy spectra etc.) calculated at the parton level, agree very well with the measured ones. This concept is called parton-hadron duality, in which hadron-level flow of energy, momentum and flavour should follow parton-level. In the following paragraphs the two main models used by PYTHIA and both HERWIG implementations are discussed [22].

2.3.3.1 The cluster model

The pre-confinement property of pQCD [29] is the motivation for using cluster objects formed at the end of the parton shower. The key idea is to enforce that remaining gluons at the shower cut-off scale are split non-perturbatively into quark-antiquark pairs beforehand. The colour flow in the parton shower causes colour-singlet pairs of a quark and antiquark to end up close in phase space. The mass spectrum of these colour-singlet clusters is asymptotically independent of the hard subprocess type, scale and the beam energy. The hadronisation step combines adjacent colour lines which form physical clusters. This approach is illustrated in Fig. 2.16(a) and is the core of the hadronisation in both HERWIG implementations. If a cluster is too heavy it splits into two clusters, otherwise these excited mesonic states decay directly into two hadrons according to the two-hadron density of states. Clusters too light for such a decay form into one hadron directly. This principle presents a well-founded and very simple attempt to describe hadronisation without using a fragmentation function. Nevertheless, there are special cases, like suppression of the production of baryons and heavy flavors, that need an intervention within this model.

(a) Via the cluster model   (b) Via the string model
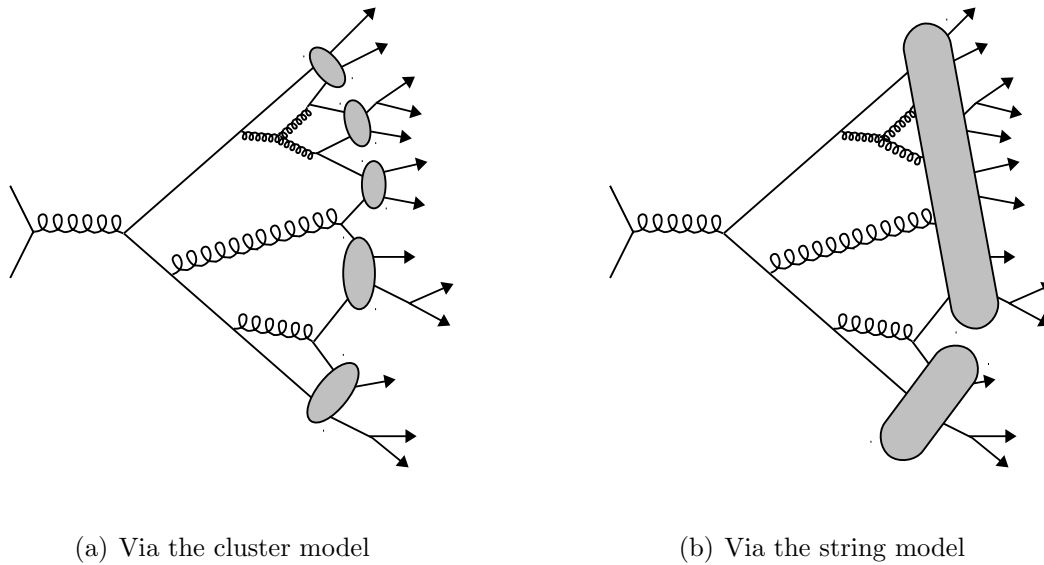
Figure 2.16: *A hard process followed by parton shower and hadronisation.*

2.3.3.2 The string model

The linear confinement at large distances is the starting point for the string model. Calculations [30] have shown that the energy stored in the colour dipole field between a charge and an anticharge increases linearly with the distance of the charges. So two partons moving apart from each other are the extremes of colour flux tubes. With increasing distance a tube is stretched and the rising potential may break the string and create new $q\bar{q}$ pairs. The probability of breaking the string is given by the quantum mechanical probability of tunnelling through a potential barrier. Hence, the fragmentation to heavy quark pairs is suppressed due to their higher masses. Charm and heavier quarks are not expected to be produced in those soft fragmentations. The fragmentation continues until all strings end with quarks and antiquarks and all systems reach a cut-off scale. Then the string transforms, according to its mass and flavour content, into a hadron or excited hadronic state. A schematic picture of a multi-hadronic state formed by the string model is shown in Fig. 2.16(b). One important property of this model is collinear and infrared safety. The PYTHIA generator uses this hadronisation model. Although this complex model has proven its ability to reproduce data very well, whether it is its underlying physical concept or its flexibility based on the large number of unrestricted parameters is unknown.

2.3.4 The underlying event and beam remnants

In hadron collisions a description of the beam remnants and multiple parton interactions (MPI) is needed. The beam remnant (non-perturbative) is colour-connected to the hard interaction, and forms part of the same fragmenting system. The activity in a hadron-hadron collision is heavily dependent on the centre-of-mass energy and MC tunes are adjusted consequently. The underlying event physics are built from perturbative models based on the $2 \to kl$ with $k,l = q,g$ scattering processes. In a hadron-hadron collision a

beam particle contains a multitude of partons. Consequently more than one separate scatter can be observed. An effective $p_{\perp\min}$ scale is introduced, below which the perturbative cross section is assumed to completely vanish or to be strongly damped. The average number of interactions per collision is then given by the ratio between the total hadronic cross section σ_{kl} and the total partonic cross section σ_{tot}:

$$\langle n\rangle(p_{\perp\min}) = \frac{\sigma_{kl}(p_{\perp\min})}{\sigma_{tot}(p_{\perp\min})} \tag{2.20}$$

with $\langle n\rangle(p_{\perp\min})$ giving the average of the distribution of the number of parton-parton interactions above $p_{\perp\min}$ per hadron-hadron collision. In PYTHIA, multiple interactions are ordered in $p_{\perp}$ and the sum of x fractions of each successive interaction can never be greater than unity. In the HERWIG++ and HERWIG model, the generation step is stopped once the energy-momentum conservation is broken.

2.3.5 NLO matrix element generators with shower programs

The approximative nature of parton showers can be improved by NLO calculations of the hard kinematics and cross sections. But these NLO calculations have problems when trying to describe collinear and soft partons whereas the parton showers are doing well in describing those. As a consequence the results from parton shower event generators are only reliable for the shape of distributions, not the absolute normalisation. Clearly a combination of both techniques would improve the description of any partonic state. However, adding a NLO matrix element generator to a parton shower is not trivial. But it is the only way to overcome the deficiencies of NLO calculations. More sophisticated prescriptions are needed due to the problems of double counting jets which are either generated by real emission or along with the parton shower and correctly covering phase space. There are two groups of approaches. The first approach introduces a merging scale where any parton produced above that scale is generated by the NLO matrix element and, conversely, any parton produced below is generated by the shower. In the second approach which is adapted for some of the NLO generators used in this analysis (see Section 10) higher-order corrections to an inclusive process are integrated into the parton shower.

3 Heavy flavour production

Heavy flavour production is a benchmark process in perturbative QCD because it probes the underlying strong dynamics at a well-defined scale. The mass scale of the heavy quarks $m_Q \gtrsim 1$ GeV is above the application limit for pQCD. Because of this heavy-flavour jets are of intrinsic interest because they hold the most direct correspondence between the parton-level production and the observed hadron level. At the LHC QCD jet and dijet production is the dominant process. The $b\bar{b}$ jet pairs are produced abundantly with a cross section of about 300 μb as shown in Fig. 3.1. Studying the agreement of $b\bar{b}$ dynamics and production is not only interesting in the context of tests of our knowledge of pQCD, but also in the context of searches for new heavy objects and deviations from the Standard Model. Heavy-flavour quarks could arise from new heavy objects and studies of the fraction of heavy-quark jets relative to all-flavour (inclusive) jets could reveal this. Besides they and their decay products, especially leptons, are background to many searches for new physics. Besides their abundant production b-quarks are also very important for top-quarks studies because the top-quark directly decays into a b-quark without interacting with other quarks. This chapter will discuss the production and fragmentation of b-quarks and emphasise the differences to light flavour quarks. Finally, past measurements on $b\bar{b}$ dijet production will be presented.

3.1 Bottom flavour in hadronic collisions

With an approximate bottom quark mass of $m_b \approx 4.75$ GeV the corresponding value of $\alpha_s(m_b)$ is around 0.24 (see Fig. 2.7) and therefore perturbation theory is expected to hold.[1] Restricting the final state to $b\bar{b}$ leaves two sets of $2 \rightarrow 2$ processes of Fig. 2.11 to be considered at leading order: gluon fusion and light-quark annihilation. Within pQCD calculations, the light-flavour quarks are considered massless which introduces a singularity in the propagators and the need for an appropriate cut-off scale. In contrast, the propagators for bottom production cannot develop such poles. Furthermore, the fragmentation functions can therefore be analysed to a greater extent with perturbative means. The main difference to light quark fragmentation is that a heavy quark is not slowed down as much when picking a light quark to form a hadronic state. Therefore the b-hadron retains about 70% of the original b-quark momentum.

In the following the dijet cross section of b-jets with contributing subprocesses is presented at NLO. Then aspects of the measurement of heavy-flavours in jets are discussed followed by the presentation of parametrisations for the heavy quark fragmentation functions. Finally, techniques for identifying bottom-flavour in jets are explained. For these purposes jets are formed by the clustering algorithm anti-k_t [32, 33] with radius parameter $R = 0.4$ in simulated events with the MC10 tune at the level of stable particles. Details on the jet reconstruction can be found in Section 5.1. The definition of a b-jet in MC simulation is as

---

[1]This mass definition uses the pole mass as reference: 4.5-5 GeV. The pole mass of the heavy quark is defined as the position of the pole in the quark propagator in perturbation theory.
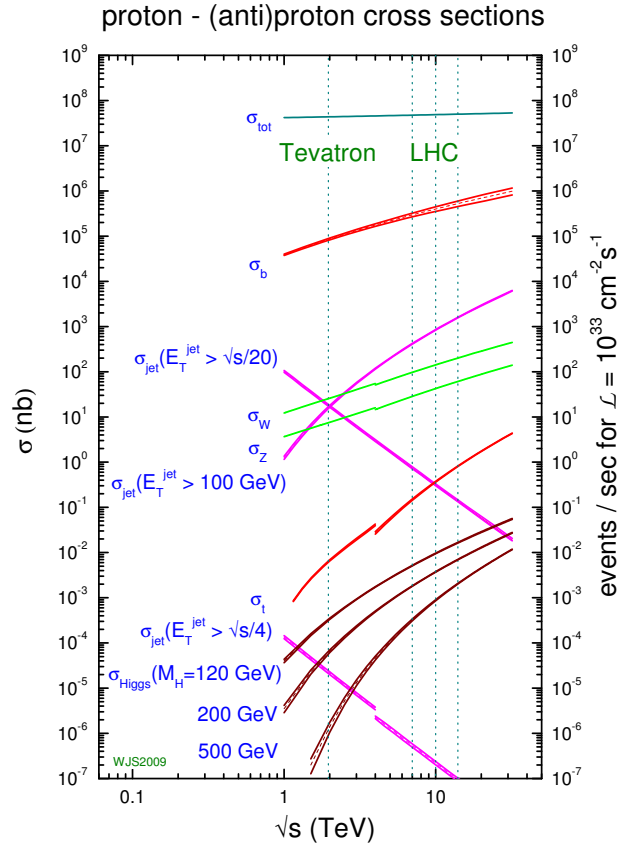
Figure 3.1: *Standard Model cross sections as a function of centre-of-mass energy $\sqrt{s}$. Indicated are the Tevatron centre-of-mass energy of 1.96 TeV and the possible centre-of-mass energies of 7, 10 and 14 TeV of the LHC. The Tevatron is a synchrotron that accelerates protons and antiprotons. The LHC accelerates two beams of protons. Taken from Ref. [31].*

follows: a jet containing any bottom-flavoured hadron with at least $p_{\mathrm{T}}^{b-\mathrm{hadron}} > 5$ GeV and at a maximum distance $\Delta R < 0.3$ to the jet axis will be labelled a b-jet. Subsequently, if a jet does not contain a b-hadron it is searched for a c-hadron and under the same conditions the jet is labelled a c-jet. If both matches fail the jet is automatically labelled a light-jet representing the cases of light-quark and gluon-initiated jets.[2] From now on quantities not specific further represent implicitly jet quantities.

3.1.1 Theory of bottom-quark production

In the partonic cross section two mass effects need to be considered affecting the matrix element and the phase-space integration. The latter is restricted by energy conservation and the cross section vanishes for $\sqrt{\hat{s}} \leqslant 2m_b$. The subprocesses contributing to the matrix element at LO, i.e. $O(\alpha_s^2)$ are:

$$gg \to b\bar{b},$$
$$q\bar{q} \to b\bar{b}$$

as shown in Fig. 3.2.

---

[2]The cases in which electrons or taus fake jets are neglected.

So the cross section is sensitive to the gluon content down to low x-values of the incoming hadrons. The matrix elements [9] calculated with these Feynman-graphs are

$$\frac{1}{g^4}\overline{\sum}|M(q\bar{q}\to b\bar{b})|^2 = \frac{4}{9}\left(\tau_1^2 + \tau_2^2 + \frac{4m_b^2}{2\hat{s}}\right), \tag{3.1}$$

$$\frac{1}{g^4}\overline{\sum}|M(gg\to b\bar{b})|^2 = \left(\frac{1}{6\tau_1\tau_2} - \frac{3}{8}\right)\left(\tau_1^2 + \tau_2^2 + \frac{4m_b^2}{2\hat{s}} - \frac{4m_b^2}{\hat{s}\tau_1\tau_2}\right) \tag{3.2}$$

with $\tau_{1,2} = \frac{1\mp\beta\cos\theta^*}{2}$, the heavy-quark velocity $\beta = \sqrt{1 - \frac{4m^2}{\hat{s}}}$ and the scattering angle θ^* (cf. Eq. 2.1). Using the variables in Eq. 2.16 and Eq. 2.17 these matrix elements and the cross section can be expressed in terms of Δy or χ in analogy to Paragraph 2.1.4.2. Additional mass-terms need to be added to the cross section formula but the overall conclusions from Fig. 2.13 and Fig. 2.14 are still valid. The cross section reveals the following rapidity dependence [9, 34]:

$$d\hat{\sigma} \propto \frac{1}{(1 + \cosh(\Delta y))^2}\overline{\sum}|M(ij\to b\bar{b})|^2, \tag{3.3}$$

with

$$\frac{1}{g^4}\overline{\sum}|M(q\bar{q}\to b\bar{b})|^2 = \frac{4}{9}\left(\frac{1}{1 + \cosh\Delta y}\right)\left(\cosh\Delta y + \frac{m_b^2}{m_T^2}\right), \tag{3.4}$$

$$\frac{1}{g^4}\overline{\sum}|M(gg\to b\bar{b})|^2 = \frac{1}{24}\left(\frac{8\cosh\Delta y - 1}{1 + \cosh\Delta y}\right)\left(\cosh\Delta y + 2\frac{m_b^2}{m_T^2} - 2\frac{m_b^4}{m_T^4}\right). \tag{3.5}$$

Both cross sections only depend on Δy and not on the rapidities themselves. Eq. 3.3 shows that the cross section is damped at large Δy. Furthermore, the light-quark annihilation via the s-channel is peaked at small rapidity differences and decreases very rapidly with

(a) s-channel

(b) left to right: t-channel, u-channel and s-channel
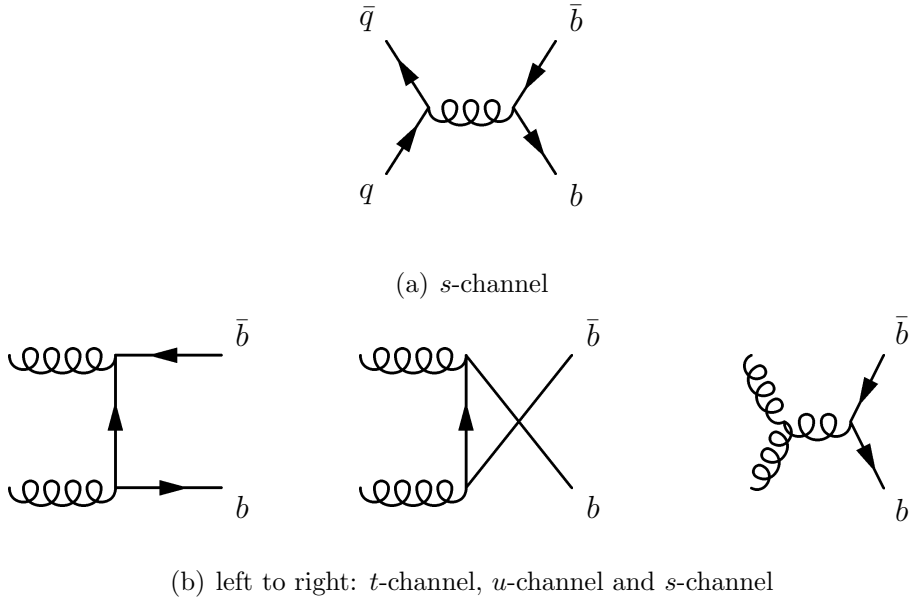
Figure 3.2: *Lowest-order Feynman-graphs for $b\bar{b}$ production: (top) light-quark annihilation, (bottom) gluon fusion*

increasing Δy. In the gluon-fusion process the t- and u-channel poles in the matrix elements at vanishing scattering angle θ^* favour a large Δy but this behaviour is damped by the $1/\hat{s}^2 \propto \frac{1}{(1+\cosh \Delta y)^2}$ term. Therefore the dependence on the rapidity difference for the gluon-fusion process leads to a much broader distribution peaked at around $\Delta y \approx 1$. The resulting rapidity distributions showing this behaviour are presented in Fig. 3.3(a) and Fig. 3.4(a) with the corresponding projections along the x-axis. The projection is made from the range $-2 \leqslant y_2 < -1.6$. The projection is therefore peaked at around $y_1 \approx -1$ to -0.6.

At LO and small dijet invariant mass gluon fusion is the dominating process making

(a)

(b)
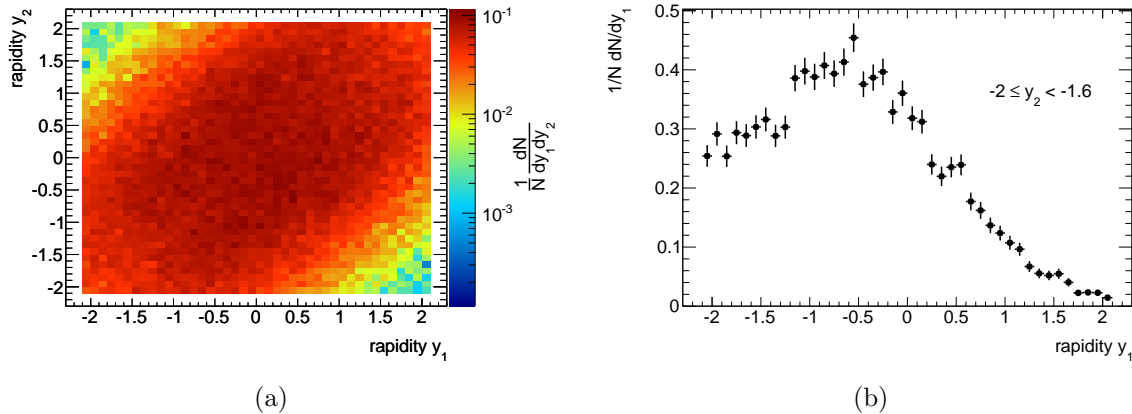
Figure 3.3: *(a) Rapidity distribution of gluon fusion $b\bar{b}$ events generated with* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV with b-hadrons selected with $p_T^{b-hadron} > 5$ GeV. (b) Projection along x-axis illustrating the cross section dependence on rapidity.*

(a)

(b)
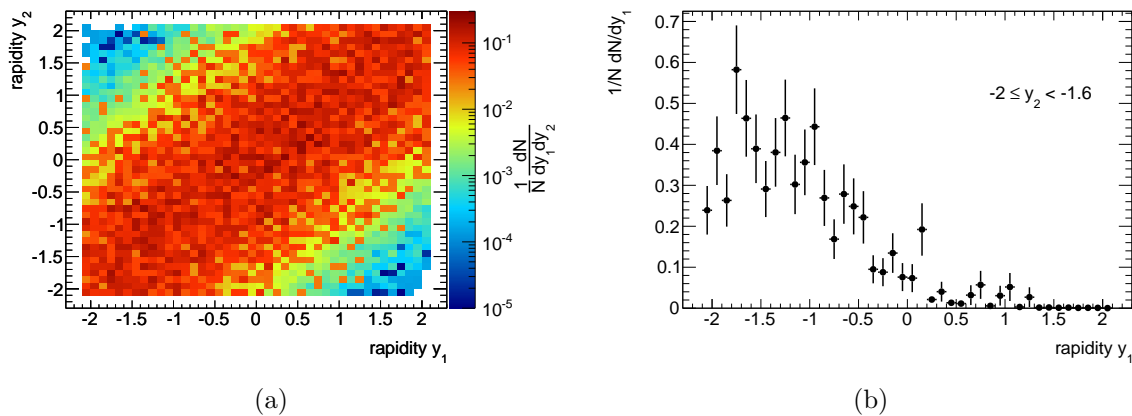
Figure 3.4: *(a) Rapidity distribution of light-quark annihilation $b\bar{b}$ events generated with* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV with b-hadrons selected with $p_T^{b-hadron} > 5$ GeV. (b) Projection along x-axis illustrating the cross section dependence on rapidity.*

up about 90% of all events and producing mostly central configurations. This large contribution from gluons in the initial state leads to a significant uncertainty in the production cross section arising from the gluon PDF uncertainty at low x. With increasing

invariant mass of the outgoing two-parton system the light-quark annihilation becomes more important.

The two-parton system is reconstructed with the help of the jets (cf. Section 5.1). The two highest-p_T jets, referred to as leading and sub-leading jets, are selected as the dijet system and identified as the two outgoing partons. At an invariant dijet mass of about 1.8 TeV more than 50% of the events are produced by light-quark annihilation as can be seen in Fig. 3.5.[3] The dijet mass range up to 5 TeV in Fig. 3.5 spans over 12 orders of magnitude in the cross section. The point, at which the major fraction of events is produced by light-quark annihilation, is determined by the slope of the gluon PDF. Dedicated studies on the underlying production mechanism using such characteristics could improve our knowledge on the PDFs.

The NLO contributions to the cross section contain the real and virtual emissions to
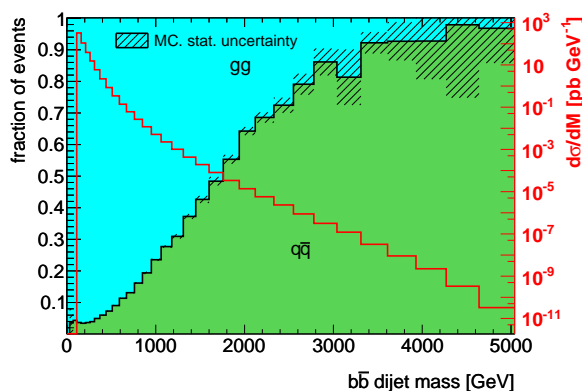
Figure 3.5: *The $b\bar{b}$ dijet mass cross section event fractions as a function of dijet mass calculated at LO with* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV ($p_\mathrm{T} > 40$ GeV, $|y| < 2.1$) show the contributions from gluon fusion (gg) and light-quark-annihilation ($q\bar{q}$). Superimposed is the dijet mass cross section as a function of dijet mass.*

the LO graphs, examples are shown in Fig. 3.6. Additional processes calculated with leading-order precision appear at NLO. In summary the processes contributing at NLO are:

$$gg \rightarrow b\bar{b}g,$$
$$q\bar{q} \rightarrow b\bar{b}g,$$
$$qg \rightarrow b\bar{b}q,$$
$$\bar{q}g \rightarrow b\bar{b}\bar{q}.$$

The subprocess with the initial partons g and q (or $\bar{q}$) appears first at NLO level. In addition there are new NLO process which lead to configurations identical to those from LO with real emission. These are illustrated with the help of the Feynman-graphs in Fig. 3.7. These diagrams can also be realised with the initial state gg. Traditionally, the production processes at NLO are grouped into three categories: flavour creation (FCR), flavour excitation (FEX) and gluon splitting (GSP). This grouping is intuitive and a useful concept which in terms of pQCD however disregards interference terms. FCR contains the

---

[3]The selected phase space of $p_\mathrm{T} > 40$ GeV, $|y| < 2.1$ is motivated in the following measurement of ATLAS data and will be explained later.

Born-level processes and in addition all diagrams with a gluon radiation to those $2 \rightarrow 2$ terms. FEX illustrated on the left side in Fig. 3.7 describes diagrams in which an initial state gluon splits into a $b\bar{b}$ pair before the interaction with the parton from the other colliding hadron. GSP on the other hand (right side of Fig. 3.7) consists of the diagrams where a gluon splits into a $b\bar{b}$ pair after the interaction. FEX and GSP both first appear at NLO level. All-order event generators include GSP processes not from first principle in the matrix element but they are realised by initial- and final-state radiation in the parton shower step. Most all-order event generators include FEX processes already in the matrix element. The kinematic properties of these processes are mirrored in the simulation but the normalisation is inaccurate. The importance of these NLO processes in bottom-flavour studies can be numerically estimated by the fragmentation process $g \rightarrow b\bar{b}$ with occurs in a fraction $\alpha_s(m_b^2)/2\pi$ of events. Its frequency increases with the centre-of-mass energy and the large number of gluon-gluon processes at the LHC makes FEX and GSP the dominating process.

The Born-level processes produce back-to-back configurations and the $\Delta\phi$ distribution is only a sharp peak at π. Gluon radiation populates the region below π and FEX and GSP processes contribute to all regions. The angular distributions are very sensitive to the different production mechanisms. This is presented with the help of simulated events from PYTHIA with the MC10 tune in Fig. 3.8, where the b-hadrons are selected after final-state radiation with at least $p_T^{b\text{-hadron}} > 5$ GeV. To be able to distinguish between different production processes to study their kinematic properties the following restriction is imposed: in each event only the hardest $b\bar{b}$ production process is selected. This requirement is implied in the naming scheme of the histogram axis labels: σ_{hardest} (cf. Fig. 3.8). This restriction is illustrated with the following example: if a GSP process generated in the parton shower phase coincides with a FCR hard scatter process, the GSP process will be discarded.

The next-to-leading order perturbative series of the partonic cross section at an energy

Figure 3.6: *A subset of Feynman-graphs representing NLO corrections to $b\bar{b}$ production: (left) one-loop correction to gluon fusion and (right) real emission correction to light-quark annihilation.*

Figure 3.7: *New contributions at NLO level from gluon splitting (right) and flavour excitation (left). Both graphs also can be realised with the initial state gg.*
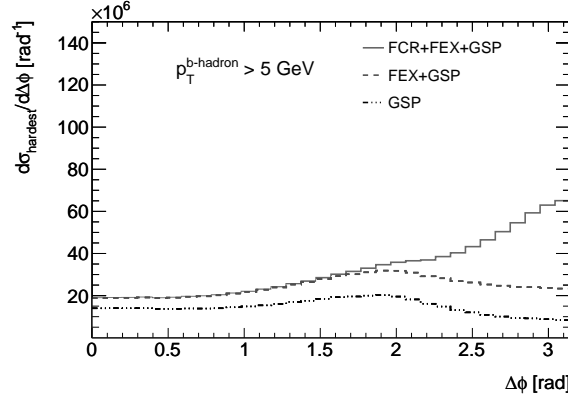
Figure 3.8: *Azimuthal distance between simulated $b\bar{b}$ hadrons with $p_{\mathrm{T}}^{b-hadron} > 5$ GeV after the parton shower predicted by* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV . Only the hardest $b\bar{b}$ production process is selected per event ($\sigma_{hardest}$).*

scale Q (= factorisation scale μ_f) is given by:

$$\hat{\sigma}_{ij} = \alpha_2(\mu^2)A + \alpha_s^2(\mu^2)\left(B + A\beta_0 \ln \frac{\mu^2}{Q^2}\right) + \mathcal{O}(\alpha_s^4). \tag{3.6}$$

The renormalisation is performed at arbitrary scale μ to compensate mass singularities. The dependence on μ of the partonic cross section is shown in Eq. 3.6 and enters through logarithms of the ratio between μ and the energy scale Q in the NLO term. Deviations due to the variation of the scales are of the order $\mathcal{O}(\alpha_s^4)$ but can nevertheless cause a sizable effect on the cross section. Variations of the scales are one way to probe the sensitivity of predictions to these uncalculated terms. In principle the choice of μ and Q is arbitrary but has to be sensible or large logarithms can develop. Therefore choosing μ of the order of Q, the logarithms are kept small. For differential cross sections it is reasonable to chose μ^2 proportional to $m^2 + Q^2$, where Q is an invariant of the two-parton system (e.g. p_{T}, $\sum p_{\mathrm{T}}(\mathrm{parton}_i)$ etc. which are invariant under longitudinal boosts). In case of the bottom-quark the inclusion of higher orders like GSP almost doubles the cross section and therefore does not stabilise the calculation against scale changes. GSP and FEX are only calculated with leading order accuracy and therefore it is expected that the predictions for $b\bar{b}$ pairs will be subject to large theoretical uncertainties arising from scale choices.

3.1.2 Bottom flavour in jets

The measurement of a jet containing one or more bottom hadrons focuses on the properties of these jets, regardless of the momentum fraction of the jet carried by the bottom hadron(b-hadron). From first principles a bottom-flavoured jet should be described by a finite-order QCD calculation more precisely than a light-flavoured jet. The mass of the bottom quark acts as a cut-off against final state collinear divergences. At leading order the b-jet coincides with the bottom-hadron itself. At the next-to-leading order the b-jet can be the heavy hadron, or it can contain a heavy hadron and a light companion, or even both the b- and the $\bar{b}$-hadron.

The detecting methods require certain acceptance regions due to the detector geometry and resolution and reconstruction limits on objects like tracks, vertices and jets. A typical

lower limit on the transverse momentum in the reconstruction of jets is $p_T < 20$ GeV due to jet identification and calorimeter properties. Consequently, also the b-hadrons are required to have a certain amount of transverse momentum. In a leading-order picture, the transverse momentum has to be at least 20 GeV. This causes two effects: the number of measured b-hadrons is obviously reduced and the angular distribution is changed compared to Fig. 3.8. This is due to the more heavily suppressed contributions from GSP and FEX than from FCR and the enhancement of the collinear regions $\Delta\phi \to 0$ and $\Delta R \to 0$ as seen in figures Fig. 3.9(a) and Fig. 3.9(b). In the jet reconstruction with distance parameter R, this leads to the inability to resolve two b-hadrons which are closer than $2R$. In general every cross section measurement is sensitive to the distance parameter R of the jet reconstruction due to such resolution effects. In case of the b-hadrons, because of the large fraction of $b\bar{b}$ pairs produced via FEX and GSP, this effect is even more important. The inclusive b-jet cross section as a function of p_T is dominated by b-jets arising from GSP. This can be concluded from Fig. 3.10. Below 200 GeV the main source of b-jets are jets which contain one b-hadron produced by GSP, above 200 GeV both b-hadrons are merged into one jet. These configurations are consequently lost in a dijet measurement where the separate reconstruction of both bottom-flavours is required. Therefore the dijet measurement is enhancing the FCR contributions and the resulting effect on the measurement of the invariant dijet mass is shown in Fig. 3.11(a) by the differential cross section and in Fig. 3.11(b) by the fraction of events from different processes. In total, the subprocesses appearing at NLO contribute at most 10% to the dijet cross section due to the requirement of a dijet topology.

The dominating fraction of FCR in the dijet cross section can also be seen in the cross section as a function of angular variable χ (cf. Eq. 2.17). The shape of this cross section is given by the FCR processes. With increasing angular separation in χ the cross section decreases. The contribution from GSP falls more rapidly over almost 2 orders of magnitude than the other two. In GSP processes the b-hadrons are strongly correlated in rapidity.

(a)   (b)
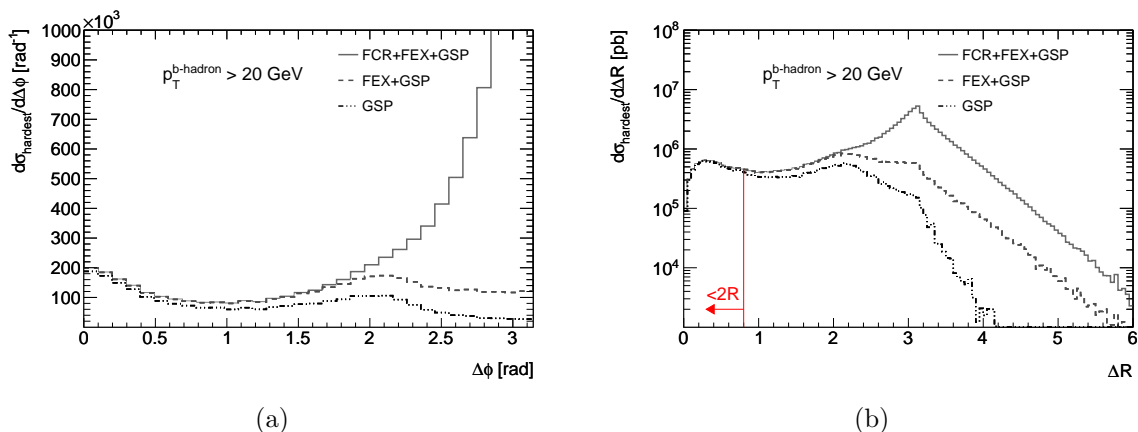
Figure 3.9: *(a) Azimuthal distance between simulated $b\bar{b}$ hadrons with $p_T^{b-hadron} > 20$ GeV and (b) $\eta - \phi$ correlation expressed as $\Delta R = \sqrt{\eta^2 + \phi^2}$ between simulated $b\bar{b}$ hadrons with $p_T^{b-hadron} > 20$ GeV after the parton shower predicted by* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV. Only the hardest $b\bar{b}$ production process is selected per event ($\sigma_{hardest}$).*
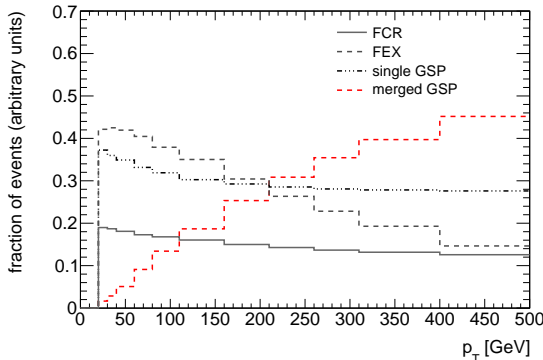
Figure 3.10: *Fractional inclusive b-jet p_T distribution of different production processes with $p_\mathrm{T} > 20$ GeV predicted by* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV. Distinguished are GSP, FEX and FCR processes. Only the hardest $b\bar{b}$ production process is selected per event ($\sigma_{hardest}$).*

3.1.3 Bottom-quark fragmentation function

Fragmentation is the process describing the transition between partons (cf. Section 2.3.3) to sprays of hadrons reconstructed in an experiment. In an MC generator this is handled in the hadronisation step where the production of partons is calculated down to a cut-off scale and finally the hadronic final states are modelled by different schemes. Measurements [35, 36] have confirmed early that hadrons formed by heavy quarks carry a larger part of the energy of the initial heavy quark than in case of a light quark. The momentum spectrum of hadrons, often referred to as the fragmentation function, is parametrised by a scaling function $f(z)$ which describes the development of the longitudinal fragmentation process where z is the longitudinal momentum fraction carried by the hadron. There are a variety of parametrisations available derived by different kinematic arguments describing fragmentation function measurements. In PYTHIA the choice of the fragmentation function can be steered by configuration parameters. The most commonly used functions are implemented in PYTHIA. Light quarks are treated with the symmetric Lund function [37]. In HERWIG no separate fragmentation parametrisation is needed because the fragmentation is handled by the phase-space scheme of the two-body decays.

Phenomenological arguments for this non-perturbative part are used to derive the so-called Peterson et al. [38] fragmentation function for heavy quarks:

$$f(z) \propto \frac{1}{z\left(1 - \frac{1}{z} - \frac{\epsilon}{1-z}\right)^2},\tag{3.7}$$

where the free parameter ϵ is related to the effective quark masses m_q of the meson $\epsilon = (m_q/m_b)^2$ and is derived from data. This function estimates the transition matrix element with the help of the energy difference between the incoming partons and outgoing hadrons. But high-energy physics data of past decades disfavours this fragmentation function. This was observed for example in the measurement obtained by the SLD collaboration [36]. This measurement is performed on hadronic Z decays where the b-hadrons are weakly decaying. Reconstructed is the variable $x_B = E_B/E_\mathrm{beam}$ where E_B is the energy of the weakly decaying b-hadron and E_beam the energy of the electron-positron beam. This variable is sensitive to the fragmentation and hadronisation model. The results
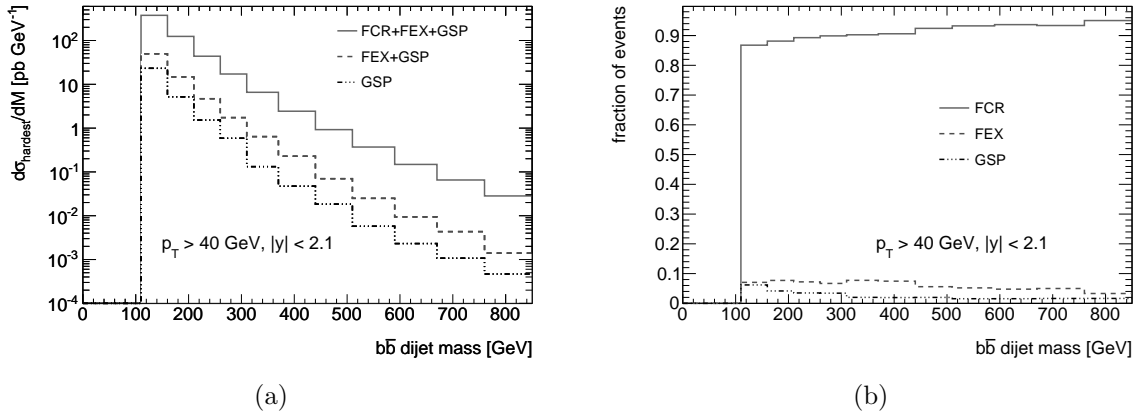
Figure 3.11: *(a) Simulated $b\bar{b}$ dijet cross section as a function of dijet mass and (b) corresponding fractions of production processes in the $b\bar{b}$ dijet cross section as a function of dijet mass in* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV and $p_T > 40$ GeV and $|y| < 2.1$. Only the hardest $b\bar{b}$ production process is selected per event ($\sigma_{hardest}$).*
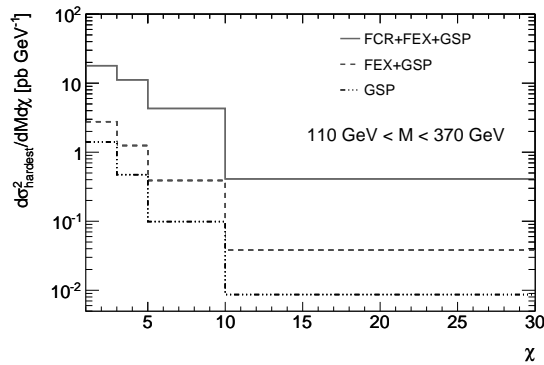
Figure 3.12: *Dijet cross section as a function angular variable χ (cf. Eq. 2.17) predicted by* PYTHIA *using a centre-of-mass energy $\sqrt{s} = 7$ TeV and $p_T > 40$ GeV and $|y| < 2.1$. Only the hardest $b\bar{b}$ production process is selected per event ($\sigma_{hardest}$).*

presented in Fig. 3.13 show that the Petersen et al. function (bottom) is not favoured by data. A much better description is given by the symmetric Bowler function [39] (cf. Fig. 3.13 (top)). Recent simulated data from PYTHIA with the MC10 tune is using the symmetric Bowler function to describe the fragmentation of heavy quarks:

$$f(z) \propto \frac{1}{z^{1+r_Q \cdot b \cdot m_Q^2}} (1-z)^a e^{-\frac{bm_T^2}{z}}, \qquad (3.8)$$

where a, b and r_Q are free parameters of the model. The symmetric Bowler model contains the transverse mass m_T of the hadron created in the hadronisation process and the transverse mass m_Q of the heavy quark at the end of the string in the string fragmentation process. The default ATLASMC10 fragmentation parameters are $a = 0.3$, $b = 0.58$ GeV^{-2} and $r_Q = 0.75$.
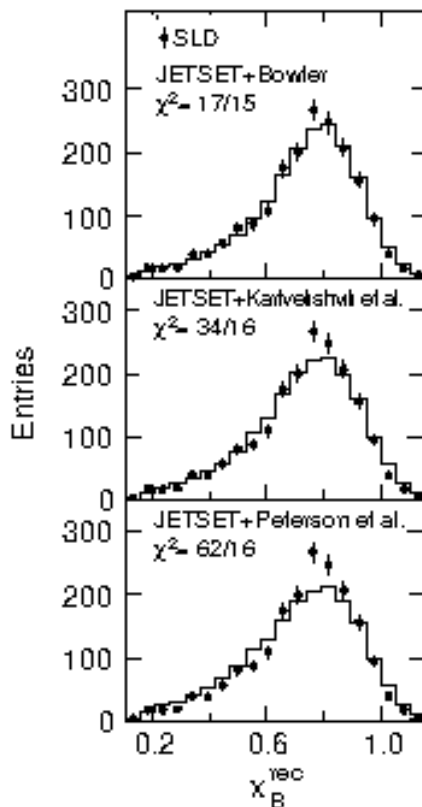
Figure 3.13: *Distribution of the reconstructed ratio x_B between the energy of the weakly decaying b-hadron and the beam energy (points) in hadronic Z-decays measured by the SLD collaboration. The data was collected between 1996 and 1997 at a centre-of-mass energy of $\sqrt{s} \approx 92$ GeV. The predictions of three models are shown: (top) Bowler, (middle) Kartvelishvili et al. [40] and (bottom) Peterson et al.. Subset of histograms taken from Ref. [36].*

3.1.4 Bottom-hadron decays

The ability to identify jets with bottom-flavours is based on the following properties of the b-hadron and its decay. Firstly, because of the long lifetime of b-hadrons, of the order of 1.5 ps, they will travel a significant distance in the transverse plane (≈ 3 mm in a jet with $p_T = 50$ GeV) before decaying. Therefore they will produce secondary vertices displaced from the primary interaction point. Furthermore, b-hadrons are massive objects producing decay products with large transverse momentum with respect to the jet axis and large opening angle allowing a separate reconstruction. Therefore the kinematics of the decay products originating from b-hadrons allow a distinction between decay products from lighter c-, light-quark and gluon jets. In addition, the decay mode into leptons (so-called semi-leptonic decay) has a branching ratio of about 10% [3] and has the advantage of very efficient and accurate lepton (= muon and electron) reconstruction at high-energy experiments. In summary, b-jets can be identified with the help of the measurement of displaced tracks and reconstruction of secondary vertices (SV) and by identifying leptons from hadron decays in a jet.

3.2 History of bottom measurements and motivating measurements of $b\bar{b}$ pair production

The discovery of the bottom quark at Fermilab near Chicago in 1977, where it manifested itself in a colourless $b\bar{b}$ bound state with a mass of roughly 10 GeV, initiated more than 30 successful years of b-physics measurements. This bound state was detected in the dimuon spectra at the E288 experiment [41]. In the following years the first measurement of the bottom production cross section was performed by the UA1 collaboration [42] at the proton-antiproton collider Sp$\bar{\text{p}}$S. This measurement (shown in Fig. 3.14) of the semileptonic bottom decays into muons calculated the integrated bottom-quark cross section at $\sqrt{s} = 630$ GeV. Although this result shows a slight excess to the NLO QCD prediction, it
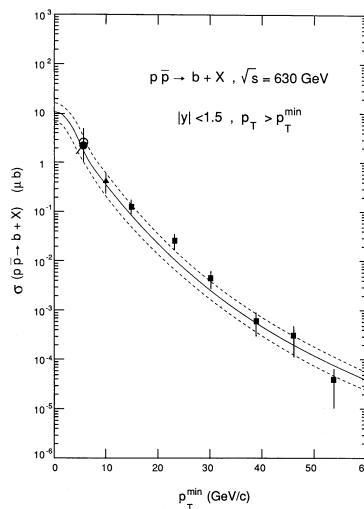
Figure 3.14: *UA1 measurement ($\sqrt{s} = 630$ GeV) of the inclusive b-quark integrated cross section for $|y| < 1.5$ as a function of the transverse momentum threshold. The data correspond to a integrated total luminosity of 4.7 pb^{-1}. The curve is the NLO QCD prediction and the dashed lines represent the theory uncertainty. Taken from Ref. [42].*

is in agreement with theory within systematic uncertainties. The measurements performed by the next generation of high-energy experiments at the Tevatron namely the Collider Detector at Fermilab (CDF) [43] and the D0 experiment [44] are not in agreement with QCD predictions. Their first measurements [45, 46] showed a factor 2-4 excess in the measured data with respect to theory [47]. This has led to many developments both in theoretical calculations and in experimental research. Changes in inputs to theoretical and experimental analyses have caused the largest impact. The more precise measurement of PDFs especially of the gluon have improved the pQCD predictions as well as new techniques for matching NLO calculations with parton showers. The other part with enormous progress is the understanding and proper transfer of the fragmentation function of heavy quarks from e^+e^- data to hadronic collisions. As a consequence of these improvements not only does the differential cross section of b-hadron production show a remarkable agreement with pQCD prediction in Fig. 3.15 but also the inclusive b-jet cross section as a function of p_T shown in Fig. 3.16 is in agreement with the NLO prediction. But this latter agreement

is mostly due to the large theoretical uncertainties and systematic uncertainties in the measurement. Especially in the high p_T region, the shape between data and pQCD NLO prediction does not agree well indicating enhancements of higher order contributions or non-perturbative fragmentation effects of gluons. As demonstrated in Fig. 3.16, the usage of jets enables probing a larger transverse momentum range due to the independence from an exclusive decay measurement. In addition a jet measurement is less subject to fragmentation and hadronisation effects due to the clustering of associated particles. In the b-jet measurement shown in Fig. 3.16 the theoretical uncertainties already play a major role at a centre-of-mass energy less than one third of the LHC centre-of-mass energy. The measurement of bottom-flavour pair production and their angular correlations as presented in Section 3.1.1 may help to understand the production mechanisms, test QCD and provide means to search for new physics. In the following section a brief summary of past measurements is presented.
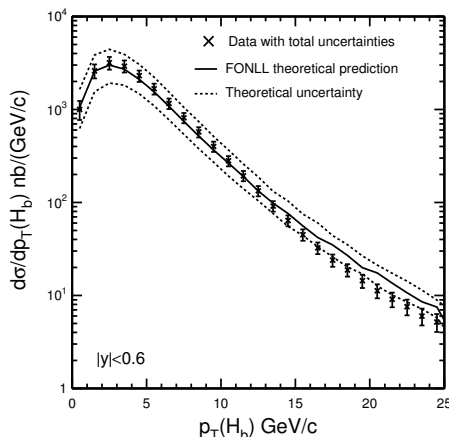
Figure 3.15: *Differential cross-section distribution of b-hadron production as a function of transverse momentum $p_\mathrm{T}(H_b)$ measured with the CDF detector at $\sqrt{s} = 1.96$ TeV with a total luminosity of 39.7 pb^{-1}. The solid line is the central theoretical values using the fixed order plus next-to-leading logarithms (FONLL) calculations [48], the dashed line is the theoretical uncertainty. FONLL calculations contain the NLO calculation and next-to-leading logarithm resummation. Taken from Ref. [49].*

3.2.1 Past measurements of $b\bar{b}$ pair production properties

The advantage of measuring the $b\bar{b}$ dijet systems lies in the inclusion of correlation information on the underlying QCD production. Similarly to the measurements presented on the inclusive bottom-flavour production the dijet measurements beginning with the UA1 experiment revealed deviations from QCD predictions. The first measurement on $b\bar{b}$ correlations performed by the UA1 collaboration [51] with the help of a data sample containing two muons shown in Fig. 3.17 exhibits an excess of 30-40% compared to NLO QCD predictions. In Fig. 3.17 the QCD predictions are scaled by 30% and 40% for the left and right plot, respectively. However, this disagreement is still covered by the theoretical uncertainty. Also here more recent experiments at the Tevatron [52, 53, 54] have exposed the inability of QCD predictions to describe data. As an example in Fig. 3.18
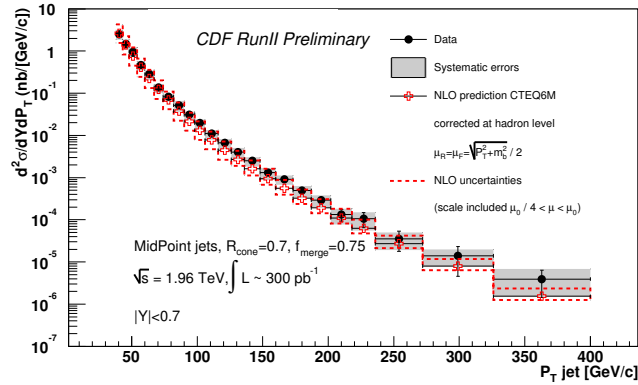
Figure 3.16: *Inclusive b-jet differential cross section as a function of jet p_T at $\sqrt{s}$ = 1.96 TeV measured with the CDF detector and data corresponding to a total luminosity of 300 pb^{-1}. Taken from Ref. [50].*

the differential cross section as a function of $\Delta\phi$ is measured with the CDF detector in a data sample where one b-quark was identified by its decay into a muon and the other one via displaced tracks. Neither is the cross section shape nor the normalisation correctly described by the QCD prediction. The data is showing again an excess of the factor of 2. With the new developments the QCD predictions were considerably improved and therefore also the consistency between data and theory. The CDF measurement [55] of $b\bar{b}$ dijets, presented in Fig. 3.19 shows remarkable agreement between data and theory for the dijet cross sections as a function of leading jet E_T and invariant dijet mass. Furthermore, the agreement in the angular correlation which is sensitive to the production processes, is improved almost all the way down to the region $\Delta\phi < 1$ when comparing data to the NLO generator MC@NLO [56]. This is a fixed-order generator with the capability to be connected to a parton shower generator (HERWIG). This measurement has again proven the importance of including the NLO processes GSP and FEX. Besides these measurements the angular correlation as a function of rapidity difference and the ratio between the production of all jets and bottom-flavoured jets is a powerful tool for searching for new physics coupling to the third generation and probing key properties of the underlying QCD formalism.
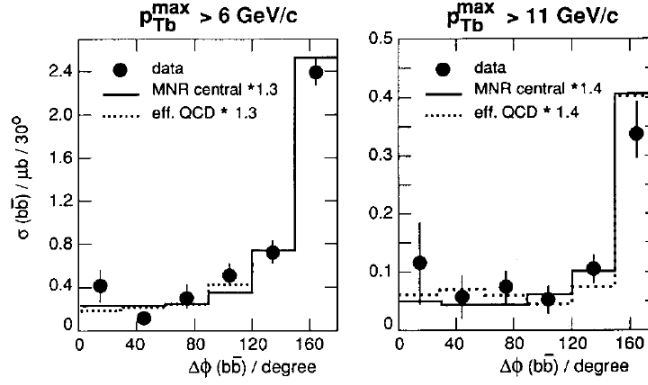
Figure 3.17: *Measurement of the $b\bar{b}$ angular correlation using a data sample containing two muons with the UA1 detector with a p_T of the higher-p_T quark > 6 GeV (left) and > 11 GeV (right). Data is compared to $\mathcal{O}(\alpha_s^3)$ QCD predictions (MNR) and to the effective QCD Monte-Carlo (eff. QCD). The predictions are scaled by 30% and 40% respectively for the left and right plot. The theoretical uncertainty on the NLO prediction is about 40%. Taken from Ref. [51].*
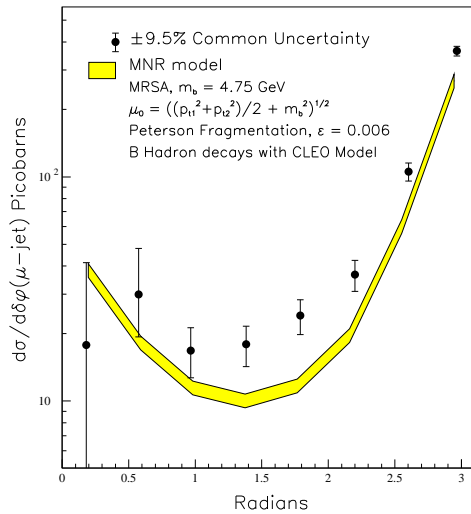
Figure 3.18: *The differential cross section measured with the CDF detector as a function of $\delta\phi$ between an identified muon from a b-hadron decay and a b-jet selected by displaced tracks for $p_T^\mu > 9$ GeV, $|\eta^\mu| < 0.6$, $E_T^b > 10$ GeV and $|\eta^b| < 1.5$. The Data corresponds to a integrated luminosity of 15.08 pb^{-1} and was taken at a centre-of-mass energy of $\sqrt{s} = 1.8$ TeV. Data is compared to theoretical $\mathcal{O}(\alpha_s^3)$ QCD predictions. Taken from Ref. [52].*
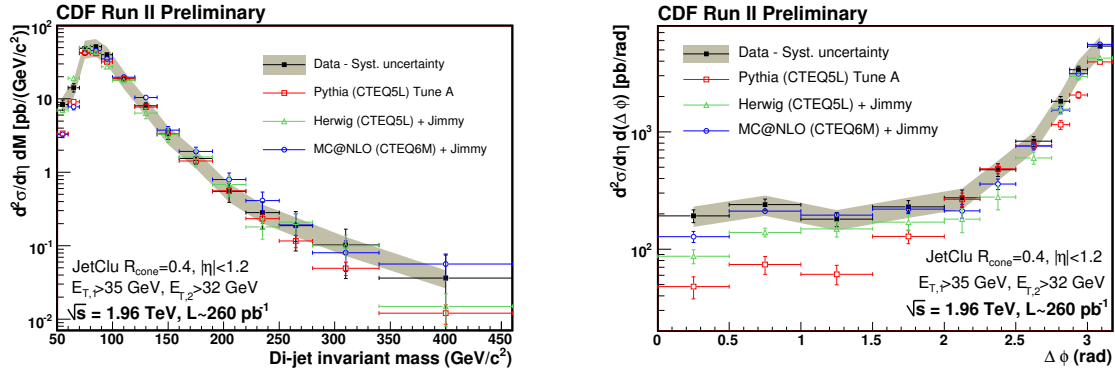
(a) $b\bar{b}$ differential cross section as a function of dijet $\Delta\phi$ correlation

(b) $b\bar{b}$ differential cross section as a function of dijet invariant mass

Figure 3.19: *The double differential dijet cross section of b-jets performed with the CDF detector ($|\eta| < 1.2$). The b-jets are identified via a reconstructed displaced, secondary vertex with $E_\mathrm{T} > 20$ GeV. The data corresponds to a total luminosity of 260 pb^{-1} at a centre-of-mass energy of $\sqrt{s} = 1.96$ TeV. The shaded area represents the total systematic uncertainty. Taken from Ref. [55].*

4 The Large Hadron Collider and the ATLAS detector

The Large Hadron Collider (LHC) is a hadron-hadron collider located underground at the European Organisation for Nuclear Research (CERN) facility in Geneva, Switzerland [57]. The LHC accelerates two hadron beams -either protons or lead ions- in opposite directions in the accelerator ring about 100 m underground. There are in total six experiments making use of the hadron beam accelerated by LHC. Each experiment is unique and specialised to a specific topic in particles or high-energy physics. There are two large experiments, ATLAS (A Toroidal LHC ApparatuS) and CMS (Compact Muon Solenoid), which are general-purpose detectors and used to analyse the properties of proton-proton collisions at high luminosity. Building two experiments with similar abilities ensures independent cross-confirmation of new discoveries. Besides these experiments, there are two medium-size experiments, ALICE (A Large Ion Collider Experiment) and LHCb (LHC-beauty), which are specialised detectors exploring specific phenomena in particle physics. These four experiments are complemented by the experiments TOTEM (TOTal Elastic and diffractive cross section Measurement) and LHCf (LHC-forward). They focus on particles very close to the beam pipe of LHC. All experiments are located underground around the ring of the LHC.

In this chapter the LHC is briefly discussed, a focus is set on the machine configuration used in the data taking in 2010 with proton beams. This is followed by a description of the ATLAS detector collecting data from collision, its trigger and data management system. Finally, the simulation of the detector response is explained which is used to generate large-scale MC simulation samples. With these a wide range of physics processes and scenarios can be studied before events are recorded. In these samples the connection between the particle's properties as measured by the detector and the particle which entered the detector is recoverable. They store the so-called MC truth information of the particles entering the detector.

4.1 The Large Hadron Collider

The accelerator is hosted by the underground tunnels of the former Large Electron-Positron collider (LEP). The accelerator ring has a circumference of 26.7 km. The LHC has been designed to run at a beam energy of 7 TeV corresponding to a centre-of-mass energy of 14 TeV and an instantaneous luminosity of 10^{34} $s^{-1}cm^{-2}$. In vacuum beam pipes bunches of protons are accelerated. Before a proton bunch is injected into the LHC, it has been accelerated to 450 GeV by two successive acceleration steps. The number of protons in one single bunch in the LHC is about $1 \cdot 10^{11}$. Bunches in the LHC travel in radio frequency buckets which have a minimum distance of 25 ns. The beam energies reached by LHC are based on superconductive magnets to guide and focus the proton beams in the ring.

The proton bunches can collide at four experiments. A collision is evoked by dipole

magnets bending the beams to cross each other. Quadrupole magnets are used to squeeze the proton bunches to collide with a transverse size of about 16 μm and a bunch length of a few centimetres at the centres of the ATLAS and CMS detector. Once stable beam with colliding bunches is established the experiments start collecting data. At best such conditions are preserved over several hours.

4.2 The machine parameters in 2010 ATLAS data

Proton-proton collisions at a centre-of-mass energy of 7 TeV which have been recorded between March and October 2010 are analysed. In Fig. 4.1(a) the time line of the total integrated luminosity delivered by the LHC is presented. The luminosity is registered after successful acceleration to the desired beam energy (*delivered all*) and after the beam conditions are stabilised (*delivered stable*). After this last step the experiments start to record data. In the end a total integrated luminosity of about 45 pb^{-1} has been recorded by ATLAS. The minimum distance of succeeding proton bunches was at minimum 150 ns in 2010 data.
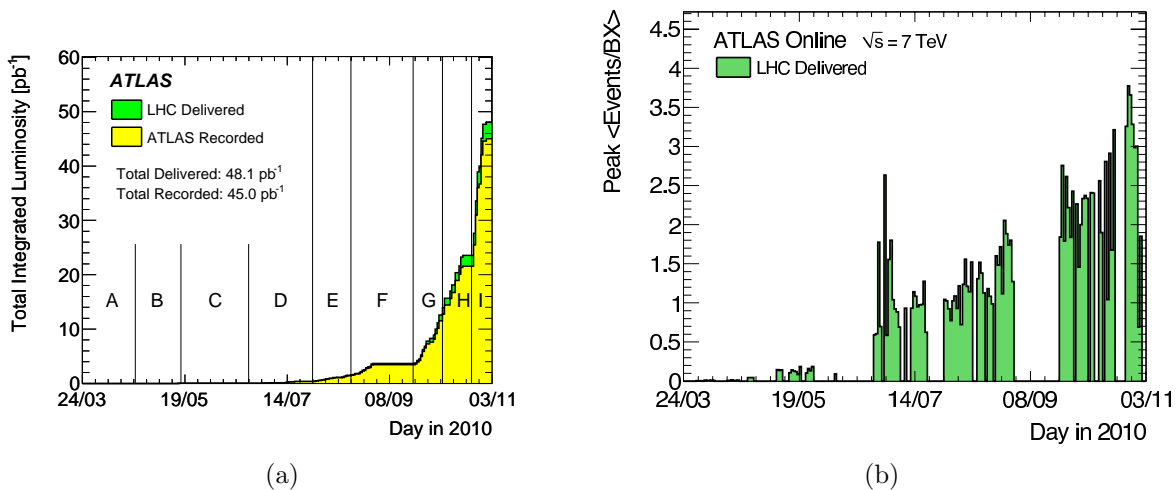
Figure 4.1: *(a) Measurement of the total integrated luminosity of the LHC run period in 2010 taken from Ref. [58]. (b) The peak number of interactions per bunch crossing as measured online by the ATLAS luminosity detectors [59] taken from Ref. [60].*

Due to changes in the machine configurations and detector operation the data taking period of ATLAS is grouped into different sub-periods labelled with capital letters. In total data from 2010 is sub-divided into data periods A to I. One effect of the progression in the commission of machine parameters is the evolution of the maximum number of simultaneous, inelastic interactions (peak) derived from the online luminosity measurement. This is presented in Fig. 4.1(b). Until the end of the run period of LHC in November, the number of simultaneous interactions increased to about 3. The major part of the luminosity therefore has been recorded with more than one interaction per bunch-crossing. This increase is caused by the reduction of the transverse size of the beam and the growth in the number of protons per bunch. Consequently the instantaneous luminosity increased from $9.3 \cdot 10^{26}$ to $2.1 \cdot 10^{32}$ s^{-1}cm^{-2}.

The determination of the luminosity is a combined effort of the accelerator and the experiment [59, 61]. Machine parameters need to be determined and the observed interaction rate per crossing needs to be measured. The essential machine parameters are the beam currents and the transverse profiles of the colliding beams. ATLAS measures the observed interaction rate per crossing with a variety of detectors and several counting techniques. The beam currents are measured with the help of electric instrumentation. The transverse profiles are determined using so-called van der Meer scans [62]. For the 2010 data set a systematic uncertainty of 3.4% has been determined.

4.3 The ATLAS experiment

ATLAS [63] is a general-purpose detector designed to cope with the conditions at the high-luminosity accelerator LHC and to measure the properties of particles produced in the proton-proton and heavy-ion collisions at unprecedented energies and luminosities. The ATLAS experiment follows two goals. Firstly, to perform precision tests of the Standard Model and secondly, to study physics beyond the Standard Model.

The design of the detector has been driven by the aim at the discovery of new phenomena. As their appearance may occur in any process and at any energy, this imposes a set of general requirements on the performance of the detector. Besides radiation hardness and fast processing requirements for the LHC detectors in general, these physics goals demand for a large acceptance in pseudorapidity as well as for a almost full azimuthal angle coverage. A good charged-particle momentum resolution and a robust track reconstruction are essential as well as an accurate energy measurement for objects such electrons, photons, jets and for the global measurement of missing transverse energy. A good muon identification and momentum resolution are in addition fundamental abilities for such goals. Further details on the physics program can be found in Ref. [64].

The detector layout of ATLAS is presented in Fig. 4.2. The detector is symmetric in forward-backward direction with respect to the nominal interaction point. The detector consists of three sub-systems which are arranged concentric around the interaction point. The main components are: the tracking system which reconstructs the trajectories of charged particles, the sampling calorimeters which measure the energy of particles and the muon system consisting of the muon detectors and a magnet system detecting muons which originate from the interaction point. Beside these main components, ATLAS has a number of detectors to perform luminosity measurements. These are placed in the forward region: LUCID, a LUminosity Cherenkov Integrating Detector, and ALFA (Absolute Luminosity for ATLAS).

The coordinate system used in this analysis has the origin in the interaction point. The z-axis is parallel to the beam axis and the x-y plane transverse to it. The positive y-axis is defined as pointing upwards, the positive x-axis is pointing towards the centre of the acceleration ring of the LHC. The polar angle θ is measured with respect to the beam pipe and the azimuthal angle ϕ is measured with respect to the x-axis. Frequently used are the quantities rapidity y and pseudorapidity η which are identical in the high-energy limit. The pseudorapidity is related to the polar angle as follows: $\eta = -\ln\tan\frac{\theta}{2}$. The rapidity is defined as $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$, where E denotes the energy and p_z is the component of the momentum along the z-axis. Transverse quantities are defined in the x-y plane which is perpendicular to the beam axis. The detector is sub-divided into three regions: barrel region, end-cap region and forward region.

The subsystems of the ATLAS detector are discussed briefly in the following, subsequently the trigger system is presented which selects event for recording upon predefined requirements. At last the data management and monitoring system and the tools used to simulate the detector response are discussed.
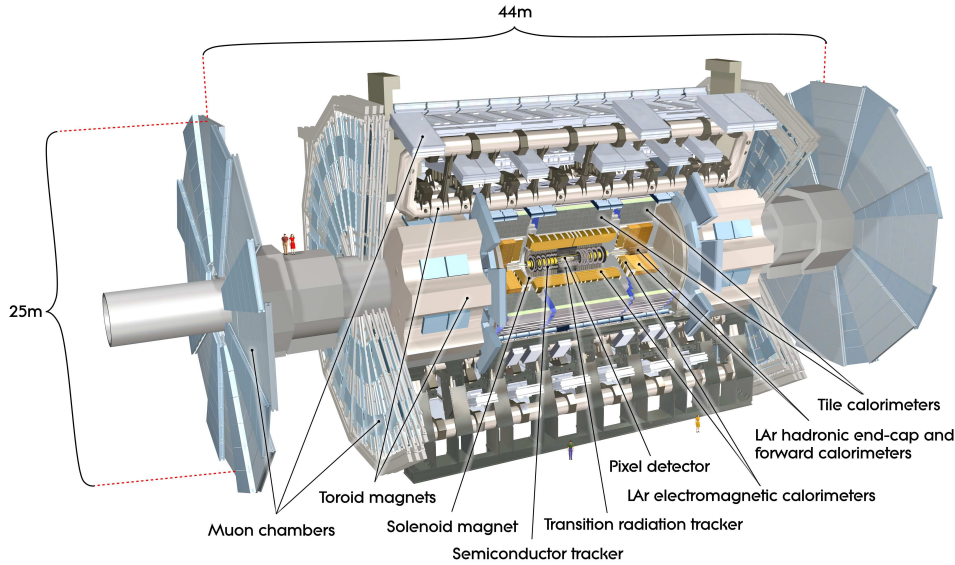
Figure 4.2: *Cut-away view of the ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tonnes. Taken from Ref. [63].*

4.3.1 The inner detector

The inner detector provides the identification of charged particles and measurements of the vertex at which the hard interaction took place, so-called primary vertex (PV), and vertices originating from the decay of long-lived particles, so-called secondary vertices (SV). The inner detector extends up to a pseudorapidity of $|\eta| < 2.5$ and is immersed into a 2 T solenoidal magnetic field which extends over a length of 5.3 m. The overall envelop of the detector is $R < 115$ cm and $|z| < 351$ cm as can be seen in Fig. 4.3.

The ID consists of three independent sub-detectors. At inner radii, the silicon pixel detector (Pixel) and the silicon microstrip tracker (SCT) deliver high-resolution tracking capabilities. At larger radii, the transition tracker comprises many layers of gaseous straw tubes interleaved with transition radiation material (TRT).

The Pixel detector and the SCT are arranged on concentric cylinders around the beam axis in the barrel region and in the end-cap region on disks perpendicular to the beam axis. The innermost layer of the Pixel detector at a radius of about 5 cm carries the name b-layer. It holds the highest precision of the reconstruction of secondary vertices which can occur in decays of bottom-flavoured hadrons. The highest precision in the position measurement are achieved in the barrel are 10 μm in the R-ϕ plane and 115 μm in z-direction. The Pixel detector can measure discrete three-dimensional space-points, whereas the SCT detector measures stereo pairs of microstrip layers. The inherent accuracies from the SCT in the barrel are 17 μm in the R-ϕ plane and 158 μm in z-direction.

The TRT which is constructed from straw tubes of 4 mm diameter, only provides R-ϕ
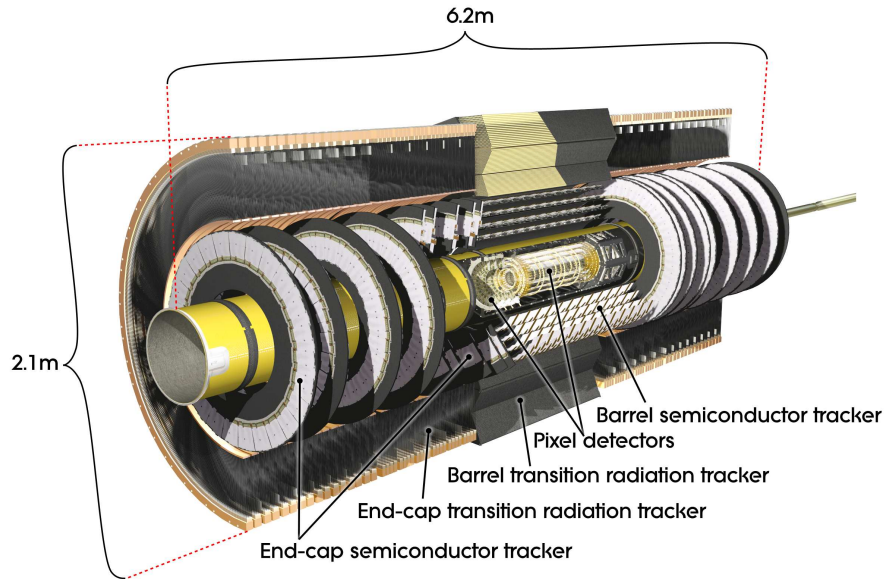
Figure 4.3: *Cut-away view of the ATLAS Inner Detector. Taken from Ref. [63].*

information, for which it has an intrinsic accuracy of 130 μm per straw. The TRT being at large radii provides continuous tracking to enhance the patter recognition ability as well as the momentum resolution. The inclusion of transition radiation material allows an electron identification complementary to the calorimeters. The general goal on the performance of the track momentum measurement is $\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\% p_\mathrm{T}/$ GeV $\oplus 1\%$. The approximate operational fraction of readout channels in the inner detector is larger than 95.9% for all three sub-systems.

4.3.2 The calorimeters

A schematic view of the ATLAS electromagnetic and hadronic calorimeters is presented in Fig. 4.4. The calorimetry consists of the electromagnetic (EM) calorimeter which is closest to the interaction point covering the region $|\eta| < 3.2$, a hadronic calorimeter covering $|\eta| < 3.2$ and forward calorimeters extending to $3.1 < |\eta| < 4.9$.

The EM calorimeter is divided into a barrel (EMB, $|\eta| < 1.475$) part with two half-barrels and two end-cap components (EMEC, $|\eta| < 3.2$) realised as wheels. This calorimeter is a lead-liquid argon (LAr) detector with accordion-shaped electrodes and lead absorber plates over its full coverage. The architecture provides full azimuthal symmetry. Each part has at most three layers in depth. The barrel part of the EM calorimeter has typically a 0.025×0.025 granularity in $\eta \times \phi$ space whereas the end-cap wheels have coarser lateral granularities.

In the region of $|\eta| < 1.8$, a presampler is placed in front of the EM calorimeter to estimate the energy lost of particles in the material between the interaction point and the calorimeters.

The hadronic calorimeter surrounds the EM calorimeters. The hadronic calorimeters use two different technologies. The region $|\eta| < 1.7$ of the hadronic calorimetry, which consists of the Tile barrel and Tile extended barrel, is a sampling calorimeter using steel as absorber and scintillating tiles as the active material. Both divisions are divided azimuthally into

64 modules. The scintillating tiles are oriented perpendicularly to the beam direction in three layers and staggered in radial depth. The $\eta \times \phi$ segmentation is typically 0.1×0.1. Between the barrel and the extended barrel there is a gap of about 60 cm for supporting structures. Scintillators cover this gap in the region $1.0 < |\eta| < 1.2$. The crack scintillators are located in the front of the LAr end-caps to cover the transition between barrel and end-cap calorimeter. In the end-caps the hadronic calorimeters (HEC, $1.5 < |\eta| < 3.2$) use the LAr technology with absorbers made of copper housed in two independent wheels. It overlaps partially with the extended barrel tile calorimeters as well as with the calorimetry in the forward region ($|\eta| > 3.1$). Each wheel is divided into two segments in depth and build of four layers per end-cap. The forward calorimeters (FCal) also use the LAr technology. They consist of three modules each. The first module uses copper as absorber and is optimised for electromagnetic measurements, while the other two modules are made of tungsten absorbers which measure predominantly hadronic interactions. The FCal calorimeters are cylindrical in shape with a hole in the middle for the beam pipe. The chosen LAr technology also provides the radiation hardness needed in the proximity of the beam. The architecture of the calorimeters achieves a seamless coverage and provides the shielding for the muon system. On average the total amount of active material at the end of the active calorimetry over the full acceptance region is at least 10 interactions lengths.[1] The fraction of operational calorimetry for 2010 is at least 99.5%. The general performance
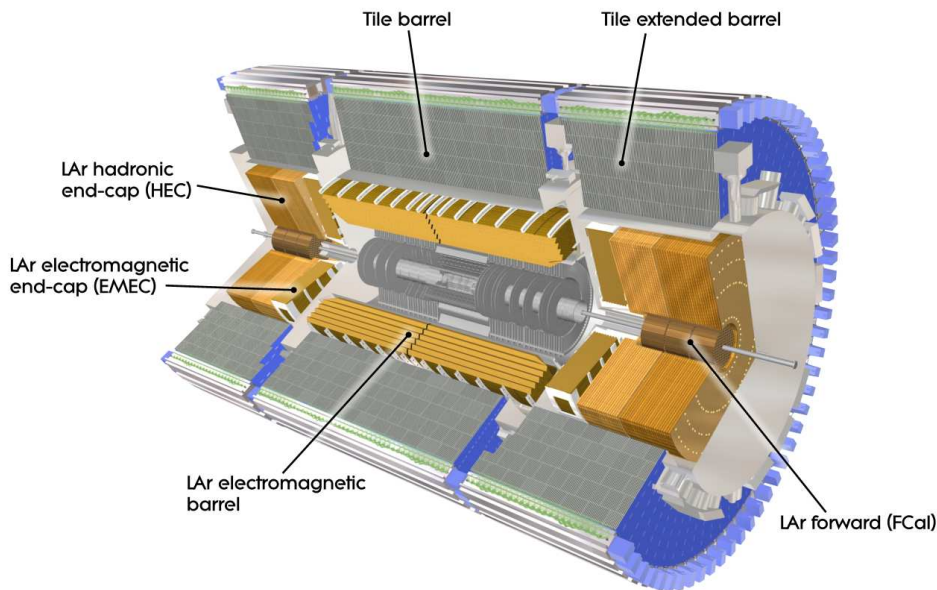
Figure 4.4: *Cut-away view of the ATLAS calorimeter system. Taken from Ref. [63].*

goals of the ATLAS calorimeters using the presented architecture is to achieve a resolution of $\sigma_E/E = 10\%/\sqrt{E/\text{ GeV}} \oplus 0.7\%$ in the electromagnetic calorimeter for electrons and photons and of $\sigma_E/E = 50\%/\sqrt{E/\text{ GeV}} \oplus 3\%$ in the hadronic barrel and end-cap calorimeters for the jet energy measurement.

---

[1]The interaction length is the mean path length of a particle before undergoing an interaction which is not elastic nor quasi-elastic.

4.3.3 The muon system

The calorimeter system is surrounded by the muon spectrometer presented in Fig. 4.5. The muon measurement is based on the magnetic deflection of muon tracks in the large superconducting air-core toroid magnets. The spectrometer is instrumented with separate trigger and high-precision tracking chambers. Over the range of $|\eta| < 1.4$, the magnetic field is provided by a large barrel toroid and for $1.4 < |\eta| < 2.7$ by two smaller end-cap magnets. The bending power is characterised by the field integral $\int B dl$, where B is the magnetic field component transverse to the muon direction and along the trajectory of the muon. The barrel toroid provides a bending power of 1.5 to 5.5 Tm and the end-cap toroids approximately 1 to 7.5 Tm. The performance goal of the muon system for the momentum resolution is $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 10\%$ at $p_{\mathrm{T}} = 1$ TeV.

The track measurement of muons is performed over the full coverage in three layers. In
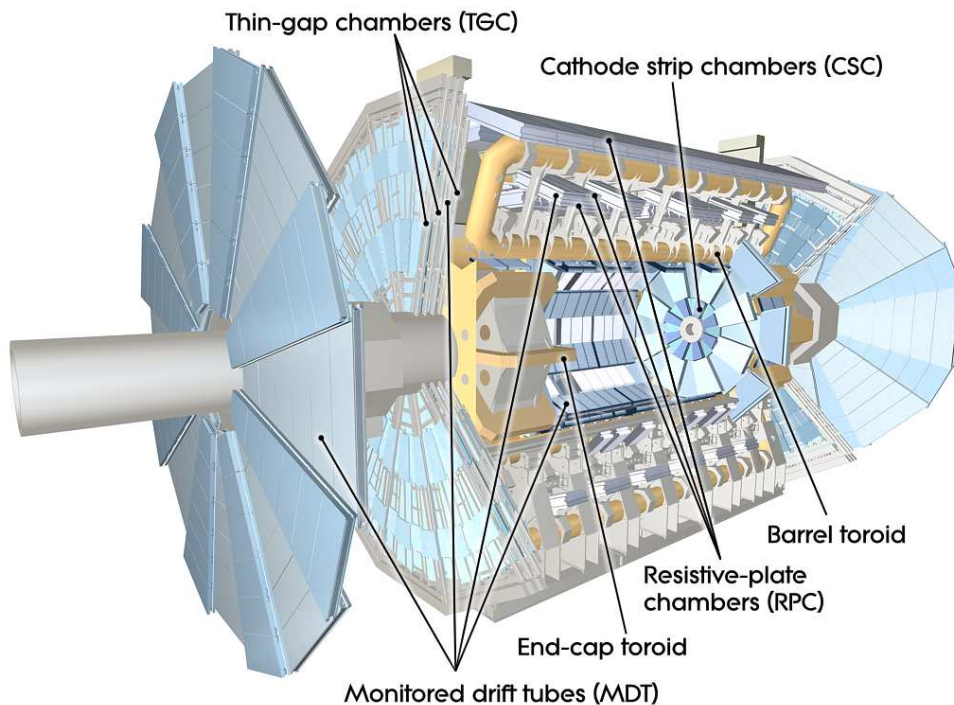
Figure 4.5: *Cut-away view of the ATLAS muon system. Taken from Ref. [63].*

the barrel these layers are arranged cylindrically, in the end-caps vertically. The track coordinates in the principal bending direction of muons are over most of the pseudorapidity coverage measured by Monitored Drift Tubes (MDT). At large pseudorapidity the measurement is realised by multiwire proportional chambers, called Cathode Strip Chambers (CSC). Beside the instrumentation for the precision measurements, the muon trigger system uses Resistive Plate Chambers (RPC) in the barrel and Thin Gap Chambers (TGC) in the end-cap regions to provide fast track information. The trigger chambers measure both coordinates of the track, one in the bending plane and one in the non-bending plane.

4.3.4 The trigger system

At design luminosity and a bunch spacing of 25 ns, the mean number of interactions per crossing is 23. Consequently, the interaction rate would be approximately 1 GHz and the data rate 60 TB/s. However, the data recording rate is limited to 200 Hz or equivalent 300 MB/s. The trigger system is therefore designed to reduce the event rate down but ensuring a maximum event selection efficiency. However, for short time periods in 2010 it was possible to maintain output rates up to 600 Hz.

The ATLAS trigger system consists of three levels of event selection: Level-1 (L1), Level-2 (L2) and Event Filter (EF). The L2 and event filter together form the High-Level Trigger (HLT). The L1 trigger is implemented using custom-made hardware components, whereas HLT is mainly based on software running on commercial computers. The most important tasks and data rates of the trigger system is illustrated in the schematic overview in Fig. 4.6. The first trigger stage is L1. The event information processed by L1 comes from the muon trigger system and from the calorimeters and has a reduced granularity. The L1 decision must reach the buffers in which events are temporarily stored within 2.5 μs. The maximum L1 rate which can be handled by the readout system is 75 kHz. Events not selected by the L1 trigger are automatically removed from the system. The L2 trigger is seeded from the information from L1, denoted as RoI (Region-of-Interest) and reconstructs the RoIs with the help of the full granularity of the detector. L2 reduces the event rate below 3.5 kHz and the L2 decision has to be made in 40 ms. The EF fully reconstructs and analyses the events accepted by L2 in about 2 s. On these events, it selects events down to a rate of 200 Hz.
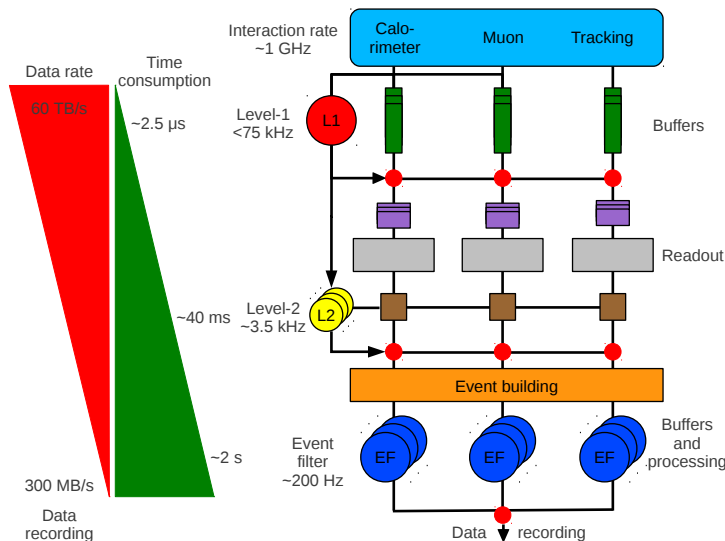
Figure 4.6: *Block diagram of the ATLAS trigger system.*

4.3.4.1 Level-1 trigger

The L1 trigger performs an initial event selection based on the information from the calorimeters and muon system. The calorimeter trigger at L1 (L1Calo) aims at the identification of high-p_T objects such as electrons, photons, decay products of τ's and jets, as well

as events with large total transverse energy ($\sum E_\mathrm{T}$) and missing transverse energy ($E_\mathrm{T}^\mathrm{miss}$). The muon trigger at L1 uses signals from the muon trigger chambers (TGC and RPC). The trigger searches for high-p_T muons and applies to identified candidates six different p_T thresholds. The information from both sub-systems are collected in the Central Trigger Processor (CTP) which stores programmable trigger menus with up to 256 distinct trigger items. The trigger items can be any combination of requirements on the input data. The L1 trigger mainly counts multiplicities of trigger objects. In addition the CTP also has inputs from specialised detectors which mostly cover the forward regions for calibration and monitoring issues.

The L1 jet trigger: The L1Calo triggers are the main subject for measuring inclusive jets as done in this analysis. The L1Calo system digitises the calorimeter signals with a reduced granularity of $\eta \times \phi = 0.1{\times}0.1$. The smallest objects used to find jets are $\eta \times \phi = 0.2{\times}0.2$. The electromagnetic and hadronic calorimeter signals are added and based on this information a feature search is performed in overlapping, sliding windows. The jet algorithm identifies E_T sums within overlapping windows of sizes of $\eta \times \phi = 0.4{\times}0.4$, $0.6{\times}0.6$, $0.8{\times}0.8$. E_T sums with $|\eta| < 3.2$ are compared to 8 pre-defined jet energy thresholds, this comparison avoids by construction double-counting. Such local maxima fulfilling any defined threshold give rise to a RoI object. Every defined jet energy threshold in the rapidity region $|\eta| < 3.2$ is identified with a single jet trigger item, named L1_JX, where X stands for the trigger threshold in GeV at L1. The L1_JX item is flagged as passed, if the L1Calo trigger identified at least one jet object with the corresponding jet transverse energy. In 2010, except for the single jet trigger with the lowest threshold the sliding window algorithm used a dimension of $\eta \times \phi{=}0.8{\times}0.8$. The jet trigger with lowest jet energy threshold was running with a window size of $\eta{\times}\phi{=}0.4{\times}0.4$.

4.3.4.2 High-level trigger

L2 and EF are comprised of computer farms. As soon as L1 accepted an event, pipeline memories which hold the data are activated to write out the corresponding event into the data acquisition system (DAQ). The DAQ system receives and stores this data. L2 then requests the detailed detector data on identified RoI's. L2 processes each single RoI. Those events not fulfilling the L2 criteria are expunged from the system. Each event accepted is transferred to the event building stage. The reconstructed events are analysed by the event filter and either are written to mass storage or dropped.

The HLT jet trigger: The detailed calorimeter information available at the HLT allows for higher precision in the identification of jet objects. The EF jet trigger was not activated during the data taking period in 2010. The L2 jet trigger was activated starting from September 2010. The L2 jet reconstruction uses a cone algorithm with $R = 0.4$ iterating over cells from the RoI region ($\eta \times \phi = 1.0{\times}1.0$). In 2010 data, the L2 jet trigger did not apply any corrections to the energy measured in the non-compensating calorimeters. In correspondence to L1 eight L2 thresholds are defined, in general placed 15 GeV above the L1 trigger. The L2 jet trigger items are denoted as L2_jX in the following, where X stands for the jet energy threshold in GeV at L2.

4.3.5 Triggers for physics analyses

All triggers used or mentioned in this analysis are required to be in coincidence with a beam crossing in which both proton bunches are filled. This is ensured by the signals from electrostatic beam pick-up devices [58]. To control the trigger rates on each trigger item prescale factors (PS) can be applied, such that only 1 in N events passing the trigger requirements is accepted at the corresponding trigger level. A defined trigger menu is accompanied with a set of prescales to deal with a range of luminosities. Prescales are auto-generated based on a set of rules taking into account the priorities for each trigger. The most important triggers are the triggers to select a physics signal. If possible, they should not be prescaled. However, especially the high rate of QCD events make it impossible to retain unprescaled single jet triggers with increasing luminosity. The signal triggers used in this analysis are the single jet triggers at L1 with the thresholds at 5, 10, 15, 30, 55, 75 and 95 GeV and in later data period the single jet triggers at L2 with the following thresholds at 15, 25, 30, 45, 70, 90 GeV.

Furthermore, two other triggers sensitive to a minimum of detector activity are used. The first one is based on two scintillator wheels of 32 counters mounted between the ID and LAr end-cap calorimeter on each side of the interaction point. This sub-system is called Minimum Bias Trigger Scintillators (MBTS) and their information is analysed at L1. The trigger item uses the acronym L1_MBTS and requires one single scintillator counter over threshold on either side. The second trigger uses the measurement of the Zero Degree Calorimeter (ZDC) covering $|\eta| \geqslant 8.2$ which is located ± 140 m away from the interaction point in the ATLAS detector. It is a tungsten-quartz fiber calorimeter to detect neutral particles. The trigger item is referred to as L1_ZDC and requires one hit over threshold on either side.

Besides these physics triggers, there are also supporting triggers. One of them is a random trigger which is used to collect data for calibration, background and pile-up studies. It is a hardware based trigger from the random clock in the CTP which generates an accept signal independent from the event topology.

4.3.6 Data management

The main components of the data acquisition system are: readout, L2 trigger, event builder, event filter as well as configuration and control. The readout system receives the data at L1 trigger rate from the on-detector frond-end systems. It stores the data until a final trigger decision off-detector in readout buffers and sends data fragments on request to the high-level trigger levels. Events passing all three stages of the trigger system are moved to permanent storage at the CERN computing centre. In addition to the storing and movement of data, the DAQ system provides the configuration, control and monitoring of the hardware and software components which together provide the data-taking functionality.

The Detector Control System (DCS) is the basis for coherent and safe operation of the ATLAS detector. This system is interfaced to all sub-detectors and technical infrastructure of the experiment. It controls, monitors and archives the operational parameters from systems such as cooling, voltage supply and cryogenics. It also handles the communication with vital systems outside the ATLAS detector for example to the LHC accelerator.

Besides the control over the hardware, networking and other facilities, it is vital to be able to diagnose the quality of the data recorded rapidly. Therefore the ATLAS data quality

monitoring framework is setup, which incorporates online (during data taking) and offline (after data taking) data quality operations. For the online monitoring a small fraction of recorded events are reconstructed promptly and fed into data quality operations in an express stream. These operations generate data distributions which represent the quality of the data. On these distributions automated tests are performed and their results are continuously stored and transferred to the ATLAS operation team.

4.3.7 Detector simulation

The simulation of ATLAS data is performed within the ATLAS simulation framework [65] using the GEANT4 software toolkit [66]. The simulation chain consists of different steps: generation of the event and immediate decays using the event generators presented in Section 2.3, simulation of the detector and physics interactions, digitisation of the energy deposited in the active material of the detector into voltages and currents. The last two steps are performed by GEANT4. The output of the simulation chain into detector signals is identical to the ATLAS data acquisition system and in turn reconstructed with the same reconstruction software as data. Simulation samples, which have gone through this full chain, are referred to as full-simulation samples. The ATLAS detector geometry used for the simulation, digitisation and reconstruction is built from databases including the information on the geometry of the system, materials involved, external electromagnetic fields and the response of sensitive detector components. GEANT4 models hadronic interactions with the detector material using the Quark Gluon String model (QGSP) [67, 68, 69, 70, 71] for the fragmentation of the nucleus and the Bertini cascade model [72, 73, 74, 75] for the description of the interactions of hadrons in the nuclear medium [76]. This hadronic interaction model has been validated with test-beam measurements using a slice of ATLAS calorimeters [77]. An agreement within a few percent has been found.

The implementation of the ATLAS geometry and the amount of material is to the best knowledge but causes, due to the imperfect description, systematic uncertainties. The description of the geometry and amount of inactive materials will be improved with data. Besides the default description, full-simulation samples are generated with geometries and material descriptions which knowingly include distortions and changes in the material budget to investigate the impact of these systematics.

4.4 Monte-Carlo simulation samples

The baseline Monte-Carlo full-simulation sample is generated with PYTHIA using the MC10 tune and contains high-p_T jets produced via the strong interaction. The default ATLAS geometry used reflects the current best knowledge. Studies of the material of the inner detector (in front of the calorimeters) are performed and if necessary systematic studies are done using full-simulation samples with extra material added. The baseline MC samples do not contain pile-up because all physical objects measured in data are corrected for any effect from pile-up.

In order to optimise the production of events according to the steeply falling inclusive cross section, the phase space is sliced in orthogonal divisions according to the transverse momentum $\hat{p}_T$ allowed in the hard scatter between two partons. The divisions are named J with a serial number starting at zero. Each division can efficiently be generated with a small amount of statistics in contrast to the full phase space. Physical distributions are

finally produced by adding all divisions according to weights. These weights are given by a pre-defined luminosity L_J, the cross section σ_J and the available statistics N_J: $w_J = \frac{L_J \sigma_J}{N_J}$. The pre-defined luminosity is default taken as 100 pb^{-1}. The statistical uncertainties on the MC events are given by the root of the sum of weights squared.

The available statistics and cross sections of the nominal MC samples generated with PYTHIA and HERWIG++ are summarised in Table 4.1. The luminosity of the MC samples with PYTHIA is about 10000 fb^{-1}. In Appendix B detailed information on the simulation samples can be found.

MC dataset	$\hat{p}_T$ [GeV]	Cross section [nb]	Number of events
PYTHIA			
J0	8-17	9860100	1398438
J1	17-35	678140	1397889
J2	35-70	40981	1397991
J3	70-140	2193.1	1396590
J4	140-280	87.707	1395487
J5	280-560	2.3502	1391492
J6	560-1120	$3.3618 \cdot 10^{-2}$	1394670
J7	1120-2240	$1.3744 \cdot 10^{-4}$	1386025
HERWIG++			
J0	8-17	9613900	397799
J1	17-35	743660	396897
J2	35-70	44307.0	398498
J3	70-140	2357.60	399598
J4	140-280	94.236	396443
J5	280-560	2.58130	397094
J6	560-1120	$3.9439 \cdot 10^{-2}$	397597

Table 4.1: *Summary of baseline* PYTHIA *samples and* HERWIG++ *samples used.*

Because the transverse size and position of the beam crossings change per run and period, the MC samples are reweighted to match the z-distribution of primary vertices in data. In data, the z-distribution of primary vertices which describes a Gaussian distribution, has a width of $\sigma_z = 60$ mm. The corresponding raw z-distributions measured in data and MC are plotted as well as the reweighting factor applied to MC can be found in Appendix B Fig. B.1.

For further systematic checks, pile-up samples which reflect the LHC bunch configuration and pile-up situation in periods G to I are studied. Details on these samples are included in Appendix B. The nominal PYTHIA samples are adapted to data from periods A to F and is referred to as non-pile-up sample.

5 Data reconstruction in ATLAS

The previous chapter has given an overview of the ATLAS detector with an emphasis on the most relevant subsystems for this measurement. The reconstruction of the physical objects from data used in this analysis is presented here. The digitised, raw data from the subsystems are processed in different steps by the reconstruction software. The reconstruction of jets is discussed including the presentation of the jet calibration scheme used in ATLAS. The reconstruction of vertices and tracks is explained as well as first results from measurements on properties critical to the performance of the tracking system are shown which are of use to this analysis. In the final section details on the reconstruction of muons are given as they will be relevant for the calibration of b-tagging-algorithms.

5.1 Jet reconstruction

Jet clustering algorithms are the main tools to analyse high-energy partons which fragment and emerge as collimated sprays of hadrons. There are two types of algorithms: cone and recombination. They are used to provide from different levels a common view on event properties. These different levels can be the parton level from fixed-order generators, the particle level from simulations including parton shower and hadronisation processes or the calorimeter (detector) level from reconstructed calorimeter information from the ATLAS detector. Between the parton level and the particle level the reconstructed jets are changed by hadronisation and underlying-event effects. The hadronisation may cause particles originating from the parton from the hard scatter not to be clustered into the same jet. The underlying event adds particles to the event not emerging from the hard scatter and possibly increases the energy of each jet. In high-luminosity colliders like the LHC, one single bunch crossing may produce several separate events (several protons interacting with each other), called pile-up events. The reconstructed jets on calorimeter level are also effected by pile-up which is dependent on accelerator parameters. Pile-up produces independent overlapping energy depositions in the calorimeter.

The running of jet algorithms on particle level with stable particles in the MC simulation determines the set of truth jets, e.g. variables are referred to as $p_{\mathrm{T}}^{\mathrm{truth}}$. The default definition of stable particles in this analysis includes all particles with a lifetime longer than 10 ps excluding muons and neutrinos. Besides this definition particle jets can also be constructed from stable particles and non-interacting (NI) muons and neutrinos. The additional usage of the non-interacting particles will be stated explicitly and this collection of jets denoted as NI jets.

In an experiment the measurement of jets starts with the energy depositions in the cells of the electromagnetic and hadronic calorimeters. From these energy depositions calorimeter clusters which are considered as input to jet finding are reconstructed. These jets formed by the calorimeter energy depositions are reconstructed at the electromagnetic scale (EM) which correctly reconstructs energy in an electromagnetic shower. The energy measurement of jets is affected by the following detector effects:

- The non-compensating calorimeters of ATLAS only partially measure the energy deposited by hadrons by nuclear interactions.

- Supporting structures of the sub-detectors and material in front of the calorimeters cause energy losses.

- The finite calorimeter coverage causes incomplete enclosure of the hadron shower.

- Energy depositions are lost because charged particles are captured by the magnetic field not reaching the calorimeters and are bent out of the reconstructed jet.

- Signal losses occur due to noise thresholds and inefficiencies in the calorimeter clustering.

A jet energy calibration corrects for those effects and relates the simulated true energy E_{truth} of the stable particles and the reconstructed energy E_{reco} in the calorimeters. A truth jet and a reconstructed calorimeter jet are associated with each other if their jet axes are separated in angular space ΔR by less than 0.3. The jet reconstruction performance is typically expressed in terms of jet energy resolution and jet energy response. The energy resolution is given by the width of the distribution of the relative difference between reconstructed jet energy E_{reco} and simulated true energy E_{truth} of the matched truth jet:

$$\frac{\sigma}{E} = \sqrt{\left\langle \left(\frac{E_{\text{reco}} - E_{\text{truth}}}{E_{\text{truth}}} \right)^2 \right\rangle - \left\langle \frac{E_{\text{reco}} - E_{\text{truth}}}{E_{\text{truth}}} \right\rangle^2}. \tag{5.1}$$

Analogously the resolutions for dijet mass, rapidity and azimuthal angle can be determined. The jet energy response $\mathcal{R}$ is determined by the ratio of the reconstructed jet energy E and the matched truth jet energy E_{truth}:

$$\mathcal{R} = \langle \frac{E_{\text{reco}}}{E_{\text{truth}}} \rangle. \tag{5.2}$$

The jet p_{T} response is defined analogous to the jet energy response, related are not the reconstructed energies but the transverse momenta. In the following section jet algorithms and the reconstruction of the calorimeter clusters which are input to the jet finding at detector level are discussed and the jet energy calibration scheme is presented.

5.1.1 Jet algorithms

A jet algorithm takes hundreds of constituents measured in an event and maps them into a few jets. Jet constituents can be the partons, the particles or the reconstructed calorimeter clusters. A jet definition contains the algorithm itself, its parameters and a choice of a recombination scheme. An important property for properly relating experimental results with theoretical calculations is the infrared and collinear safety of a jet algorithm. The final set of reconstructed jets should not be sensitive to the addition of an infinitely soft parton or to the branching of a parton into a collinear pair of partons. This ensures the cancellation of infrared divergences between real and virtual emission of soft gluons.

5.1.1.1 The cone algorithms

Modern jet cone algorithms [78] are fully defined by the cone radius R, the overlap parameter f, the number of iterations N_{pass} and a minimum transverse momentum $p_{\mathrm{T},min}$ in the recombination step. A jet is identified by a cone,defined in angular space (η, ϕ), around the direction of the dominant energy flow. To find the direction of the dominated energy flow some cone algorithms start to search for a seed (different sets of seeds can be used) defining the starting point, and some ones start without a seed (= seedless). In the starting point the centre of a cone with radius R is placed. All constituents in an event are written into a list. The algorithm continues by adding all constituents that lie within the cone with radius R around the starting point (or the former iterated centre of the cone) and re-calculates the centre of the E_{T}-weighted cone. This procedure is repeated until the centres of the cone of iteration k and iteration $k-1$ are identical and consequently the cone considered stable. Each stable cone is called a protojet. The constituents in the stable cone are removed from the event list and the next stable cone is iterated until no constituents are left or the number of repetitions is bigger than N_{pass}. Then the recombination step is performed with overlap parameter f and a minimum transverse momentum. This step is called split-merge procedure and describes the method handling overlapping jets. Iteratively all protojets ordered by the scalar sum of the constituents transverse momenta $\tilde{p}_{\mathrm{T}} = \sum_{i \in \mathrm{protojet}} |p_{\mathrm{T},i}|$ above a minimum transverse momentum $p_{\mathrm{T},min}$ which overlap with another protojet are either split or merged. Two protojets are merged if the fraction of shared transverse momentum of the jet with smaller $\tilde{p}_{\mathrm{T}}$ is bigger than f. Otherwise the jets are split and the constituents are assigned according to their distance to the protojet axis. After this step the protojets defined by stable cones become the final set of reconstructed jets.
With few exceptions, jet cone algorithms are not infrared and collinear safe.

5.1.1.2 The clustering algorithms

The clustering algorithms use sequential recombination of the constituents parametrised by a given power of the energy scale in the distance measure. Realisations of such jet algorithms are the k_t [79, 80] and anti-k_t [32, 33] algorithms. The anti-k_t algorithm used in this analysis is a k_t algorithm with negative power in the distance measurement. Both jet algorithms are defined by:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R^2}, \tag{5.3}$$

$$d_{iB} = k_{ti}^{2p}, \tag{5.4}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti} is the transverse momentum of constituent i and d_{ij} is the distance between constituents i and j and d_{iB} the distance between constituent i and the beam. The algorithm identifies and starts from the smallest of all possible distances d_{ij} between all constituents in the event. As long d_{ij} is smaller than d_{iB} the constituents i and j are combined and their four-momenta are added. The clustering stops if $d_{iB} < d_{ij}$ and a final jet is created from these constituents. The constituents which formed a jet are removed from the constituent list and with the remaining constituents and their distances d_{ij} the algorithm is repeated until the constituent list is empty. For $p = 1$ Eq. 5.4 describes the k_t and for $p = -1$ the anti-k_t algorithm.
The negative sign in the anti-k_t algorithm causes the hardest constituents (= constituents

with the highest transverse momenta) to be merged in the first iteration. Subsequently softer constituents are clustered with the hard ones. This sequential arrangement and the transverse momentum influencing the distance measure lead to circular hard jets. In case of close-by jets the harder jet of both stays circular and the nearby soft jet misses a lens-shaped region as illustrated in Fig. 5.1. In contrast, k_t algorithms with a positive sign in Eq. 5.4 cluster in the first iteration constituents with small transverse momenta and merge in the end constituents with highest transverse momenta.

So from first principles these two algorithms do not create overlapping jets. In addition
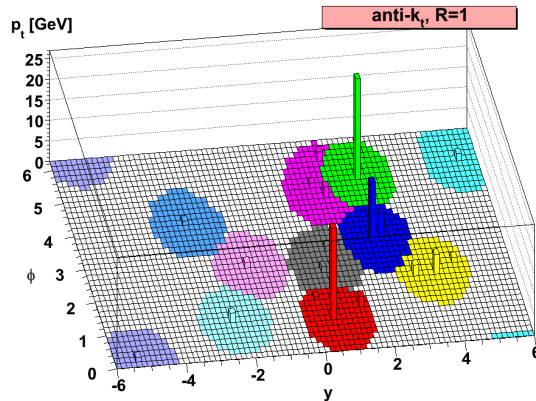
Figure 5.1: *This plot is made from a sample of parton-level events generated with* HERWIG *and clustered with the anti-k_t jet algorithm. It illustrates the active catchment area of the resulting hard jets and nearby soft jets. Taken from Ref. [33].*

they have the advantage of infrared and collinear safety. The optimal radius parameter to reconstruct jets which have a good momentum reconstruction is a compromise. To gather all objects emerging (created by hadronisation) from the parton initiating a jet demands a large jet radius. In consequence of a large jet radius the contamination from the underlying event increases which counteracts a good momentum reconstruction. Taking both effects into account the optimal jet radius ranges between 0.4 and 0.6 [81]. In this analysis, anti-k_t jets with distance parameter $R = 0.4$ are used.

5.1.2 The reconstruction of jet constituents from calorimeter cells

The jet constituents chosen here are topological calorimeter clusters [82] with an energy $E > 0$ GeV consisting of a group of calorimeter cells. These dynamically-grown topological clusters aim to reconstruct the energy of a single particle which deposits its energy in many calorimeter cells. The advantage of this algorithm is the efficient suppression of the calorimeter noise. The noise in each calorimeter cell is estimated beforehand as the absolute value of the energy deposited in the calorimeter cell divided by the root mean square (RMS) of the energy distribution measured in events triggered at random bunch crossings.

This cluster forming algorithm is designed to imitate a typical shower development and leads to clusters with a variable number of cells. The algorithm starts from a seed cell whose signal-to-noise ratio is above a threshold of 4. Iteratively neighbouring cells which fulfil a signal-to-noise ratio of 2 are added. Finally, all adjacent calorimeter cells are added to the existing cluster. The building of topological clusters is finished after a splitting

step in which overlapping showers from close-by particles are separated. The individual particles that initiated the shower may have produce well-separated local maxima in each topological cluster. Such local maxima are defined as those clustered cells which satisfy $E > 500$ MeV and whose neighbouring cells have all less energy. These local maxima are used as seeds in a new iteration of cluster building and the original cluster is split into more topological clusters, one for each local maximum found. At the end of this cluster building process, each shared cell is added to the adjoining clusters with weights, which are functions of their energies and distances to the centre of the cluster centroids. The energy of a topological cluster is equal to the energy sum of all clustered cells belonging to it. The clusters are assigned a mass of zero and the direction is reconstructed from the absolute energy-weighted averages of pseudorapidities and azimuthal angles of the constituent cells.

5.1.3 Jet calibration in ATLAS

The jet energy calibration scheme used in this analysis is called the EM+JES scheme [60]. The reconstructed jets in this analysis have this energy calibration applied. This simple scheme corrects the jet energy E^{EM} derived at the electromagnetic scale with the help of correction factors from MC simulation, which are a function of the jet energy and pseudorapidity. This calibration scheme as simple as it is allows a direct evaluation of the systematic uncertainties and is referred to as the jet energy scale uncertainty. The calibration consists of three steps:

Correction due to pile-up: The additional energy reconstructed in jets arising from pile-up is corrected by subtracting an average amount of energy from each jet. This energy correction is calculated with the help of two quantities: the average number of jet constituents $\langle N_{\mathrm{constituents}}\rangle$ and the average additional energy C_{pileup} in each jet constituent. The latter quantity is a function of the number of reconstructed primary vertices, N_{PV}, the detector region encoded in η and the bunch spacing. The number of reconstructed primary vertices is a measure of the number of additional proton-proton interactions in the same bunch crossing. The spacing between consecutive filled bunches is considered because events from the previous bunch crossing may have impact on the jet measurement in the subsequent bunch crossing. This additional energy is measured directly in events selected by the L1_MBTS trigger from the average energy deposited in a non-noise-suppressed jet constituent with respect to events with one primary vertex and no additional energy. The average number of jet constituents which indicates the jet area is measured as a function of rapidity in data triggered by the single jet triggers. With the help of both quantities the additional energy in a jet is given by:

$$E_{\mathrm{T}}^{\mathrm{corrected}} = E_{\mathrm{T}}^{\mathrm{measured}} - \langle N_{\mathrm{constituents}}(\eta)\rangle \mathcal{C}_{\mathrm{pileup}}(N_{\mathrm{PV}}, \eta, \text{bunch spacing}). \quad (5.5)$$

For jets with $|y| < 2.1$ the pile-up correction increases with the number of reconstructed primary vertices ($N_{\mathrm{PV}} = 2, ..., 5$) from approximately 0.2 GeV to a maximum of 2 GeV.

Vertex correction: Calorimeter jets and their constituents are reconstructed using the geometrical centre of the ATLAS detector as the default. After the event reconstruction in each event the jet four-momenta are corrected by adjusting the coordinates of each topological cluster to point to the vertex of the primary hard scatter. The kinematic observables of the clusters are re-calculated and the jet direction is re-defined. This

correction improves the angular resolution.

Jet energy and direction correction: The final step of the calibration scheme is to restore the reconstructed jet energy to the mean energy of the truth jet and to correct direction biases in the jet reconstruction. These may arise from poorly instrumented detector regions measuring a lower energy than better instrumented ones. This non-uniformity in the energy measurement biases the direction reconstruction of topological clusters. This η-correction is derived with the help of simulated events calculating the shift in the η reconstruction $\Delta \eta = \eta_{\text{truth}} - \eta_{\text{reco}}$. The size of this correction on the jets is very small ($\Delta \eta < 0.01$) for most of the detector regions except the transition regions ($\Delta \eta < 0.03$) at around $|\eta| \approx 0.8$ and $|\eta| \approx 1.3$.

The energy calibration is derived from MC simulations on all isolated calorimeter jets that have an isolated matching truth jet. The isolation of a jet is fulfilled if there is no other jet with $p_{\text{T}} > 7$ GeV within $\Delta R = 1$. The EM-scale energy response $\mathcal{R} = \frac{E^{\text{EM}}}{E_{\text{truth}}}$ for each matched jet pair is measured in bins of the truth jet energy E_{truth} and uncorrected calorimeter jet detector pseudorapidity. The response in each (E_{truth},η)-bin is fitted by a Gaussian distribution to determine the peak position $\langle R \rangle$ and the average calorimeter jet energy $\langle E^{\text{EM}} \rangle$. For every detector region a function $\mathcal{F}_{\text{calib}}(E^{\text{EM}})$ of the jet response is obtained by fitting the points $(\langle E^{\text{EM}} \rangle, \langle \mathcal{R} \rangle)$. This fitted function then determines the correction factor to relate the measured jet energy in the calorimeter to the hadronic scale:

$$E^{\text{EM+JES}} \equiv E = \frac{E^{\text{EM}}}{\mathcal{F}_{\text{calib}}(E^{\text{EM}})|_\eta}. \tag{5.6}$$

The values of the correction factor $\frac{1}{\mathcal{F}_{\text{calib}}(E^{\text{EM}})|_\eta}$ decrease with higher jet p_{T} and range from about 1.7 for $p_{\text{T}} = 40$ GeV in the central region to 1.2 for $p_{\text{T}} = 1000$ GeV in the end-cap region. They are illustrated in Fig. D.1 in the appendix. Final jets calibrated with the EM+JESscheme are saved and input to physics analysis if they fulfil $p_{\text{T}} > 7$ GeV which approximately corresponds to $p_{\text{T}}^{\text{EM}} > 4$ GeV. Information of soft jets below this kinematic cut are not available. From now on, reconstructed jet quantities, e.g. p_{T}, use implicitly the EM+JES calibration.

5.2 Tracking

The reconstruction of trajectories from charged particles is the main ingredient to identify the hard scatter and long-lived hadrons decaying in a significant distance from the hard scatter. Charged particles produced in a hard scatter emerge from one interaction point. Their trajectories point back to this point of origin which is called the primary vertex. For example a decay vertex from a heavy flavour hadron is called a secondary vertex (SV). The reconstruction of such secondary vertices is described in Section 6.2. A charged particle traversing the inner detector generates hits along its flight path in each sub-detector. From these hits tracks are reconstructed and from the final set of tracks primary vertices are formed.

There are several effects degrading the tracking performance. Due to the high activity in high-energy proton-proton collisions there may be ambiguities in associating a hit to a track. The alignment of the inner detector components is crucial for track finding as well as a realistic description of the material along the flight paths of the particles. A systematic uncertainty on the modelling of the track reconstruction in simulation is caused

by the presence of unaccounted material leading to additional hadronic interactions.

In the following sections, the reconstruction of tracks and of primary vertices is outlined. In the last section studies on the performance of the ATLAS tracking system with 2010 ATLAS data are presented.

5.2.1 The tracking algorithm

The default tracking algorithm [83] starts from the inner silicon sub-detectors and is called inside-out track reconstruction. In a pre-processing step the hit information of the Pixel and SCT detectors are transformed into three-dimensional space-points. These space-points are processed for seeding the track candidate search. Track seeds are formed from a combination of space-points. Based on these track seeds, track candidate creation is started. The track candidate creation involves track fitting and is performed on the hit information itself because the formation of the space-points in the SCT has an inherent uncertainty. The located seeds provide a directional information from which the detector elements for the further search of associated hits to this track candidate are determined. Successive hits along the track candidate are simultaneously added and fitted by a fitter formalism [84] to help to predict the next expected hit position.

The next step in the track reconstruction involves the solving of ambiguities between different tracks. Many track candidates share hits, are incomplete and may even describe fake tracks. Therefore quality criteria need to be established to rank each track candidate. Track candidates are refit using a refined reconstruction geometry with a more detailed material description. Then these are classified by a scoring strategy making use of different characteristics of a track. In general, missing and shared hits between tracks result in a downgrading of these tracks and hits are weighted according to the precision of the sub-detectors. Shared hits are assigned to the track with the higher score, while the remaining track is refit disregarding this hit. Below a certain quality criterion reconstructed tracks are excluded from further processing.

Finally, the track reconstructed from the silicon sub-detector information is extended into the TRT. Two steps are necessary for this: finding compatible sets of TRT hits and a track refit with the full set of hit information from all three sub-detectors. Given a track candidate from the Pixel and SCT detectors, the track is extrapolated into the TRT detector volume. This extrapolation is done over a large distance and within the magnetic field and is therefore subject to large uncertainties. Therefore the search window for TRT points which are collected for further processing is large. For an associated extension the track is refit allowing to change the initial silicon track. The original and the new track are compared using the quality score strategy. The silicon track is kept in case the quality score is lower for the extended version. Finally, all tracks above a certain minimum p_T^{track} are saved. In case of this analysis, all tracks with at least $p_T^{track} > 100$ MeV are kept.

5.2.2 Primary vertex finding

The reconstruction of the interaction vertex is based on the reconstruction of charged-particle tracks as presented in the previous section. The reconstruction of primary vertices (PV) is done in two steps [85]: a primary vertex finding algorithm identifying vertex candidates and associated reconstructed tracks and a vertex fitting algorithm for reconstructing the vertex position and the corresponding error matrix. With the help of an identified primary vertex the associated tracks are refit constraining their point of origin. Tracks

entering the primary vertex reconstruction are subject to the following quality requirements:

- $p_\mathrm{T}^\mathrm{track} > 150$ MeV,

- $|d_0^\mathrm{BS}| < 4$ mm,

- $\sigma(d_0^\mathrm{BS}) < 5$ mm,

- $\sigma(z_0^\mathrm{BS}) < 10$ mm,

- at least 4 hits in the SCT detector,

- at least 6 hits in the silicon detectors.

Here d_0^BS and z_0^BS denote the transverse and longitudinal impact parameters respectively of tracks with respect to the centre of the luminous region, so-called beam-spot (BS), with their corresponding uncertainties $\sigma(d_0^\mathrm{BS})$, $\sigma(z_0^\mathrm{BS})$ from the track fit. A detailed description of the beam-spot determination can be found in[86]. Impact parameters are defined by the distance between the point of closest approach to a point of interest in the transverse plane and along the z-axis for the transverse and longitudinal impact parameter respectively. The impact parameters are shown in Fig. 5.2 with respect to the primary vertex. The sign of z_0^PV is given by the distance between z_PV and z_SV in the ATLAS coordinate system. The sign of d_0^PV is given by the geometrical relation between these unit vectors $\mathrm{sign}(d_0^\mathrm{PV}) = (\hat{p} \times \hat{d}) \cdot \hat{L}$, where $\hat{p}$ is the unit vector pointing in direction of the transverse momentum of the track, $\hat{d}$ is the unit vector pointing from the primary vertex to the point of closest approach and $\hat{L}$ the unit vector in direction of the angular momentum of the particle's trajectory in the magnetic field. The selection cuts based on the impact parameters are meant to remove a large fraction of tracks originating from secondary interactions. The luminous region is on a regular basis determined during data taken by the distribution of the primary vertices reconstructed and is archived in a database. In the final reconstruction this recorded information is utilised to calculate the impact parameters.

The vertex finding starts by selecting all tracks compatible with the interaction region and the quality criteria listed above. The iterative vertex finding approach then continues by searching for a vertex seed. A vertex seed is a global maximum in the distribution of z coordinates of the tracks with respect to the point of closest approach to the beam-spot centre. The vertex position is determined by the adaptive vertex fitting algorithm [87] taking into account the seed and the associated tracks around it. This fitter is a robust χ^2-based fitting algorithm which handles outlying tracks by down-weighting their contribution to the vertex progressively. Tracks incompatible with the vertex by more than approximately 7σ in transverse and longitudinal directions are able to be used as new vertex seeds. This procedure is repeated until no new vertices are found or the list of reconstructed tracks is empty. The measurement of the beam-spot constrains the available seeds and reduces the number of outliers in the reconstruction of the primary vertex. Consequently the constraint on the primary vertex to be in agreement with the beam-spot reduces beam-related background and background from cosmic particles. The reconstructed vertices are ordered in descending order by $\sum(p_\mathrm{T}^\mathrm{track})^2$, which is strongly correlated to the total number of tracks at the vertex. The hard-scatter vertex is distinguished from soft interactions by the higher average $p_\mathrm{T}^\mathrm{track}$ of its tracks and correspondingly the primary vertex at top of the list is selected.
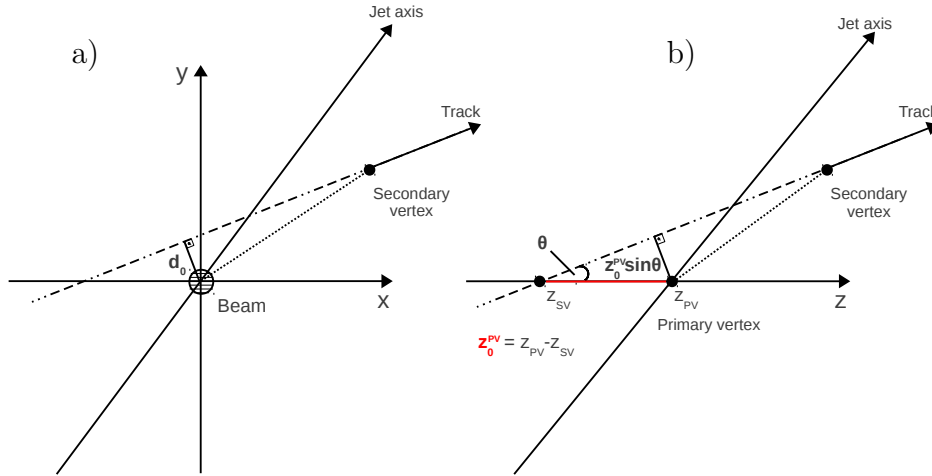
Figure 5.2: *Definition of the impact parameters with respect to the primary vertex: a) transverse and b) longitudinal impact parameter.*

5.2.3 Measurements of tracking performance

This section presents a selection of measurements on the performance of the tracking system relevant to this analysis. The track and vertex reconstruction depend on correct alignment information, detector geometry and material description which makes a thorough check on their description necessary. These performance studies which began with the measurement of cosmic particles improved the understanding of the detector and consequently the tracking performance. The first measurement of tracking properties in collisions was done with data selected by the minimum bias triggers [88]. The tracks from charged particles in the kinematic range $|\eta| < 2.1$ and $p_T^{\text{track}} > 500$ MeV are selected in a data sample triggered by the single-arm minimum-bias trigger and ensured data quality. The events are required to have one primary vertex consistent with the beam-spot and at least one good quality track. A good track needs to satisfy the following criteria:

- $p_T^{\text{track}} > 500$ MeV,

- a minimum number of silicon detector hits,

- transverse and longitudinal impact parameters calculated with respect to the event primary vertex $|d_0^{\text{PV}}| < 1.5$ mm and $|z_0^{\text{PV}}| \cdot \sin\theta < 1.5$ mm.

The impact parameter cuts are used to remove secondary interactions mostly due to hadronic interactions, photon conversions and decays of long-lived particles. Fig. 5.3 shows that the average number of silicon hits on tracks as a function of pseudorapidity is well described by the simulation.

The following performance study [90] has grown more mature and uses data-driven methods to derive systematic uncertainties. It extended the considered tracks to the full acceptance range of the inner detector of $|\eta^{\text{track}}| < 2.5$ and lowered the p_T^{track} threshold cut for the track reconstruction. This updated study has refined their track selection cuts, for details refer to [90]. The track reconstruction efficiency was estimated from simulation and the

(a)   (b)
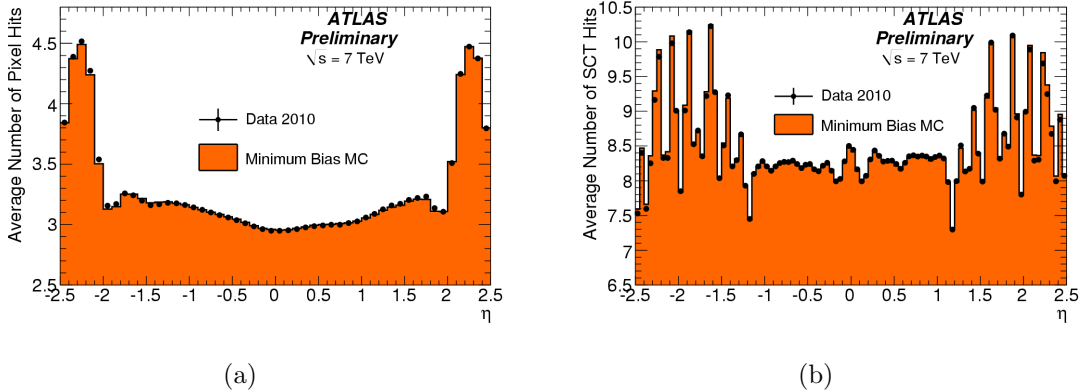
Figure 5.3: *The number of hits measured in the silicon detectors with the ATLAS detector at $\sqrt{s} = 7$ TeV: a) Pixel, b) SCT is compared to* PYTHIA *prediction with the older ATLAS tune MC09 [89]. The data corresponds to an integrated luminosity of 6.8 μb^{-1}. Taken from Ref. [88].*

level of agreement between data and simulation was determined to derive a systematic uncertainty. This efficiency is defined as

$$\epsilon_{\text{track}}(p_{\text{T}}^{\text{track}}, \eta) = \frac{N_{\text{reco}}^{\text{matched}}(p_{\text{T}}^{\text{track}}, \eta)}{N_{\text{gen}}(p_{\text{T}}^{\text{track}}, \eta)}, \tag{5.7}$$

where $p_{\text{T}}^{\text{track}}$ and η are generated particle properties. $N_{\text{reco}}^{\text{matched}}$ is the number of reconstructed tracks matched to a generated charged particle and N_{gen} is the number of generated charged particles as a function of $p_{\text{T}}^{\text{track}}$ and η. The association between a reconstructed track and a generated charged particle uses a matching algorithm in angular space with matching parameter $\Delta R < 0.15$. The resulting track reconstruction efficiency is shown in Fig. 5.4. The rise of the efficiency with increasing $p_{\text{T}}^{\text{track}}$ is due to an indirect minimum $p_{\text{T}}^{\text{track}}$ requirement on the tracks and a minimum number of detector layers, which are passed through, coming from the minimum numbers of silicon hits allowed.

The imperfect description of the detector, in particular the material budget, leads to the main systematic uncertainty which needs to be considered.

Two data-driven methods are used to determine the agreement between simulation and data concerning the material distribution [90]. The first method reconstructs the invariant mass of K_s^0 mesons decaying into oppositely charged pions, the second compares track lengths in data and simulation. For the most accurate description both methods are combined. The first method is most sensitive at small radii, while the other method probes the detector volume beyond the Pixel detector. The comparison of the data to MC simulations with different material budgets determines the level of unaccounted material in the inner detector.

The track length method validates the material description in MC by studying the rate at which tracks which were found in the Pixel detector can be extended into the SCT detector. The comparisons between MC simulations with different material descriptions and data indicate the level of imperfection of the material description beyond the Pixel detector.

The considered decay products of K_s^0 mesons undergo ionisation and multiple scattering when traversing the inner detector. The reconstruction of the invariant mass (width and
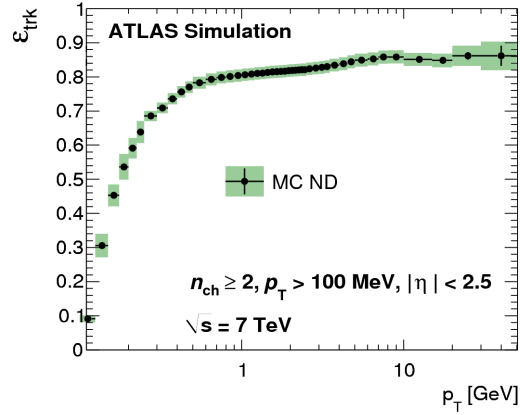
Figure 5.4: *The track reconstruction efficiency as a function of p_T^{track} for $\sqrt{s} = 7$ TeV is derived from non-diffractive (ND) minimum bias MC[1]. The efficiency is shown for the following phase space cuts: the number of charged particles in the event $n_{ch} \geqslant 2$, $p_T^{track} > 100$ MeV and $|\eta^{track}| < 2.5$. The statistical errors are shown as black lines, the total errors as green shaded areas. Taken from Ref. [90].*

mean) as a function of radius in data and MC should have been equally biased by ionisation and hadronic interactions. The peak and the width of the invariant mass spectra is fitted and the fitted K_s^0 mass relative to the nominal sample as a function of radius is determined. Fig. 5.5 shows the resulting behaviour of the fitted mass ratio with respect to the nominal sample. In data the mass is in very good agreement with the nominal simulation sample for small radii, but with increasing radii data exhibits a negative slope.

The agreement between data and nominal MC simulation varies in each detector region.
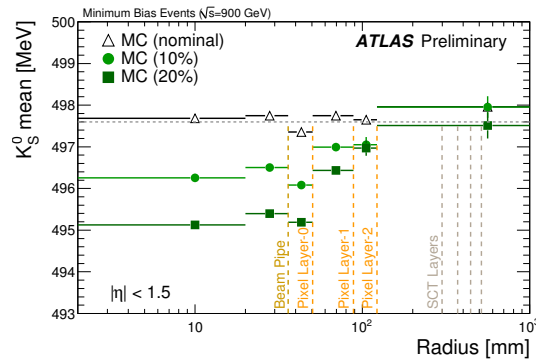
Figure 5.5: *Fitted K_s^0 mass ratios as a function of the decay radius for data and various MC simulations with PYTHIA relative to the nominal MC sample. Data was taken at $\sqrt{s} = 900$ TeV and corresponds to the full data set from 2009 LHC running (400000 events). The K_s^0 candidates are required to have a decay radius within the bin boundaries. The two pions are not allowed to have silicon hits before the decay radius and at least one silicon hit in every following layer. The marker's errors are statistical only. The vertical dashed lines indicate radial positions of material in the Inner Detector barrel. Taken from Ref. [91].*

Whereas the barrel is very well described, the transition region shows small deviations

---

[1]Non-diffractive (soft inelastic) processes are modelled from low-p_T lowest-order perturbative QCD $2 \rightarrow 2$ scatters.

in both methods which are covered by 10% uncertainty ($\Delta_{10\%}$) on the Pixel and SCT detector material. In the K_s^0 invariant mass measurement the unaccounted material between the Pixel and SCT detector is described by the data over MC ratio of the track length measurement ($\Delta_{\text{track length ratio}}$). The end-caps show fairly high deviations in the track length method which is accounted for by a 20% increase in material in the gap between Pixel and SCT detector ($\Delta_{20\%\text{gap}}$) in combination with the 10% uncertainty on the Pixel and SCT detector material. The deviations in the track reconstruction efficiency Δ_{material} between the nominal and MC samples with extra material, which show the best agreement between data and MC, are taken as systematic uncertainty:

$$\Delta_{\text{material}}(K_s^0) = \Delta_{10\%} \oplus \Delta_{\text{track length ratio}}, \tag{5.8}$$
$$\Delta_{\text{material}}(\text{track length}) = \Delta_{10\%} + \Delta_{20\%\text{gap}}. \tag{5.9}$$

The final systematic uncertainty selected is given by the more conservative choice between these data-driven methods, and is summarised in Table 5.1. With these methods the systematic uncertainties arising on tracking related observables are obtained. The resulting systematic uncertainties are given in Table 5.1.

Systematic uncertainty on tracking efficiency			
$\lvert \eta^{\text{track}} \rvert < 1.3$	$1.3 < \lvert \eta^{\text{track}} \rvert < 1.9$	$1.9 < \lvert \eta^{\text{track}} \rvert < 2.3$	$2.3 < \lvert \eta^{\text{track}} \rvert < 2.5$
2%	3%	4%	7%

Table 5.1: *Summary of the tracking efficiency systematics due to material interactions for tracks above 500 MeV [90].*

5.3 Muon reconstruction

This section describes briefly the reconstruction of muons originating from semi-leptonic decays of *b*- and *c*-hadrons, so-called soft muons. The term soft refers to the rather low momentum of these muons compared to those coming from the decay of heavy particles (such as *Z* or *W*). More details on the general muon reconstruction can be found in [92]. Soft muons are reconstructed with the help of both the muon system and the inner detector. The combination of these two systems improves the momentum resolution and reduces the mis-identifications of muons. Muons in this collection are reconstructed using two complementary methods:

- Combined muon: a standalone muon spectrometer track is associated to an inner detector one on the basis of a χ^2 match. The combined track parameters are determined from either a statistical combination of both tracks or a complete refit of the full track.

- Tagged muon: a reconstructed inner detector track identifies a tagged muon if the track extrapolated to the muon spectrometer, using detailed information about the material and the magnetic fields, can be associated with a track segment in one or more muon chambers.

The combination of these methods increases the performance of the muon identification at low p_T because the latter algorithm identifies muons which do not have enough energy to cause hits in the full spectrometer. The reconstruction efficiency for combined muons in 2010 ATLAS data is about 98%.

For the reconstruction of the muons two chains are available:

- Chain 1: the reconstruction begins locally in each muon chamber by the search for straight line track segments in the bending plane pointing to the centre of ATLAS. Track segments are extended to the next station to build search roads. A minimum of two track segments in different muon stations are required to be able to form a muon track. These tracks can be statistically combined to an inner detector track using matching cuts to build a combined track. Unused segments found by this process are used to tag extrapolated inner detector tracks and so to reconstruct tagged muons.

- Chain 2: on the full muon spectrometer a track pattern recognition algorithm performs independent transforms on both the bending and non-bending projections. The segment finding is seeded from patterns in the two projections. A track candidate is formed from segments compatible with a curved track and belonging to the same pattern. Segments from the outermost chambers seed a track fit with all segments in the next chambers closer to the interaction point. Combined muons are obtained by fitting the full track consisting of the standalone muon track and the associated inner detector track. Tagged muons can be found with the help of all track segments build by the pattern recognition process.

6 Bottom-tagging in ATLAS

The identification of jets with bottom-flavours (b-tagging) takes advantages of several properties of bottom-hadrons as discussed in Section 3.1.4. The identification relies on the precise measurement of charged tracks in the inner detector of ATLAS and on the reconstruction of muons in the muon spectrometer. The relevant jet information for b-tagging purposes is the jet direction. Only jets fulfilling $p_T > 15$ GeV and which are within the inner detector acceptance ($|\eta| < 2.5$) are subject to the b-tagging algorithm. These jets are called *taggable* jets. The strategy of labelling jets in MC simulation to be a b-jet, c-jet or light jet was presented in Section 3.1.

6.1 Bottom-tagging algorithms

Each b-tagging algorithm produces a weight distribution where the heavy flavours are visually separated from other flavours. In the simplest case the b-jet identification is an one-dimensional selection cut along this distribution. After this selection cut the remaining jets are called b-tagged. To estimate the performance of a b-tagging algorithm the following quantities are defined:

The b-tagging efficiency ϵ_{btag} is defined as the fraction of taggable jets labelled as b-jets which are tagged by the b-tagging algorithm under study:

$$\epsilon_{\text{btag}} = \frac{\#(b\text{-jet}\&b\text{-tagged})}{\#(b\text{-jet})}. \tag{6.1}$$

The b-tagging rejection R_{btag} is defined as the inverse of the mistag rate ϵ_{mistag} which is given by the fraction of taggable jets not labelled as b-jets which are tagged by the b-tagging algorithm under study:

$$\frac{1}{R_{\text{btag}}} = \epsilon_{\text{mistag}} = \frac{\#(\text{no } b\text{-jet}\&b\text{-tagged})}{\#(\text{no } b\text{-jet})}. \tag{6.2}$$

The mistag rate of light jets ϵ_{light} (of c-jets ϵ_c) is given by the fraction of taggable light-jets (c-jets) which are tagged by the b-tagging algorithm under study:

$$\epsilon_{\text{light}(c)} = \frac{\#(\text{light-jet}(c\text{-jet})\&b\text{-tagged})}{\#(\text{light-jet}(c\text{-jet}))}. \tag{6.3}$$

The b-tagging purity f_{pur} is defined as the fraction of b-tagged-jets which are also labelled a b-jet:

$$f_{\text{pur}} = \frac{\#(b\text{-jet}\&b\text{-tagged})}{\#(b\text{-tagged})}. \tag{6.4}$$

These three performance estimators are dependent on the data sample used and on the kinematic cuts applied. The performance estimated from simulation of b-tagging algorithms is strongly depended on the MC generator, its incorporated heavy flavour

description and on the detector description. Therefore the *b*-tagging algorithms require a dedicated data-driven calibration. Based on this calibration each analysis has to obtain their *b*-tagging efficiency and purity.

The spatial algorithms, the secondary vertex tagger and the impact parameter taggers, which use tracks and subsequently reconstructed vertices, are the most powerful ones and the first choice for an analysis. The soft lepton taggers are very important because their overlap to the spatial algorithms is minimal. The most important realisations for this analysis of a secondary vertex tagger, impact parameter tagger as well of the soft lepton tagger are discussed in the following. Their usage is preferred due to their robustness, simplicity and on an uncomplicated validation with data. This is because the algorithms are based on well-measurable distributions and do not involve complicated mathematical methods. Consequently their performance is limited. However, the usage and validation of more advanced techniques is on-going. Among these techniques are the combination of individual taggers and the extension to two-dimensional discriminant distributions and the validation of a multivariate approach and a neural network.

6.2 Secondary vertex tagger

The *b*-tagging algorithm discussed in this section is the main tool in this analysis. It reconstructs the displaced vertex from a *b*-hadron decay with the help of tracks which are significantly displaced from the primary vertex [93]. A heavy-flavour hadron decay within a jet and the resulting secondary vertex with its tracks is shown in Fig. 6.1. This secondary vertex is reconstructed from tracks associated to a jet with an angular matching requiring $\Delta R < 0.4$. Tracks, which can be associated by $\Delta R < 0.4$ to two jets, are assigned to the closest one. Tracks which fulfil the criteria listed in Table 6.1 are used in the vertex finding. These selection cuts are different to other impact parameter based taggers due to a SV0-specific optimisation.

The reconstruction of the secondary vertex starts by building all two-track vertices

	Selection cuts
p_T^{track}	> 0.5 GeV
$\lvert d_0^{\text{PV}} \rvert$	< 2 mm
$\lvert z_0^{\text{PV}} \sin\theta \rvert$	< 2 mm
$\sigma(d_0^{\text{PV}})$	< 1 mm
$\sigma(z_0^{\text{PV}})$	< 5 mm
χ^2/DoF	< 3
Number of Pixel hits	$\geqslant 2$
Number of SCT hits	$\geqslant 4$
Number of silicon hits	$\geqslant 7$

Table 6.1: *Track selection requirements used by the SV0 tagging algorithm. d_0^{PV} and z_0^{PV} are the impact parameters of the track in the transverse plane and the longitudinal direction respectively. They are derived with respect to the reconstructed primary vertex. The χ^2/DoF denote the property of the track fit. DoF stands for the number of degrees of freedom.*
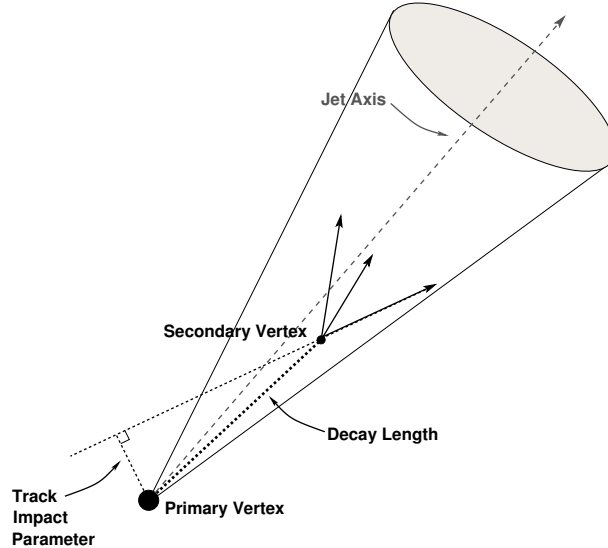
Figure 6.1: *Illustration of a heavy flavour hadron decay within a jet. The distance between a secondary vertex (SV) and the primary vertex (=decay length) indicates the presence of a long-lived particle. The secondary vertex is reconstructed from tracks which are significantly displaced from the primary vertex. Taken from Ref.[93]*

considering only tracks not compatible with the primary vertex by requiring the three-dimensional impact parameter significance $L/\sigma(L)$ with respect to the primary vertex to be bigger than 2.3. In addition the sum of the impact parameter significance of the two tracks has to be greater than 6.6. Furthermore the quality of the vertex fits must be good ($\chi^2 < 4.5$) and the vertices must all be incompatible with the primary vertex.The remaining two-track vertices are probed whether their mass is consistent with a K_s^0 meson, a Λ^0 baryon or a photon conversion or whether their radius is consistent with the radius of one of the three Pixel detector layers (material interactions). If any of the former criteria holds the vertex is removed. Finally for each jet, the tracks from the two-track vertices, which survived the selection cuts, are combined and fitted to a single secondary vertex. Iteratively the vertex fit removes the track with the largest χ^2 contribution until the fit probability of the vertex is greater than 0.001, the vertex mass is less than 6 GeV and the largest χ^2 contribution from any track is 7 or less. Tracks which failed the selections during the building of the vertices are re-considered in the vertex fit and if suitable re-incorporated into the selection. The resulting discriminatory variable, called SV0 weight, is the three-dimensional signed decay length significance $L_{\mathrm{SV}}/\sigma(L_{\mathrm{SV}})$ of the secondary vertex. The sign of $L_{\mathrm{SV}}/\sigma(L_{\mathrm{SV}})$ is given by the sign of the projection of the decay length vector on to the jet axis. The uncertainty $\sigma(L_{\mathrm{SV}})$ is the quadratic sum of the uncertainty from the secondary vertex fit and the uncertainty from the primary vertex.

A jet is considered b-tagged if it contains a secondary vertex, reconstructed with the SV0 tagging algorithm, with a SV0 weight greater than 5.85: this is the operating point that yields a 50% b-tagging efficiency in simulated $t\bar{t}$ events. The SV0 weight distribution of the leading jets with $p_{\mathrm{T}} > 40$ GeV and $|y| < 2.1$ in simulation using a QCD dijet sample is shown in Fig. 6.2. The single contributions from different flavoured jets are drawn stacked on top of each other. This tagging-algorithm is well-understood and robust. Its performance is decreased due to tracking related effects like material interactions and pattern-recognition problems near the jet core where the track density is the highest.
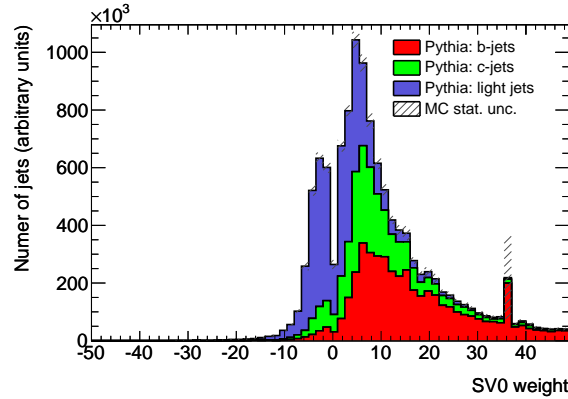
Figure 6.2: *The SV0 weight which is identical to the three-dimensional signed decay length significance $L_{\text{SV}}/\sigma(L_{\text{SV}})$ is shown obtained from PYTHIA QCD simulation. Events with at least two jets fulfilling $p_{\text{T}} > 40$ GeV and $|y| < 2.1$ are selected. Jets originating from b-hadrons compared to other other flavours tend to have larger significances. The single contributions are drawn stacked on top of each other.*

6.3 Impact parameter tagger

The jet probability b-tagging algorithm (JetProb) [94] uses the transverse impact parameters of all tracks fulfilling certain quality cuts. This algorithm will be used to investigate a possible systematic due to the choice of the b-tagging algorithm. For b-tagging purposes the sign of the impact parameters of tracks associated to a jet are re-calculated to reflect the lifetime. The transverse and longitudinal impact parameters are shown in Fig. 5.2. In context of flavour tagging the sign of these are determined by the sign of the scalar product of the jet direction vector and the vector pointing from the primary vertex to the point of closest approach. All reconstructed tracks are subject to resolution effects. The experimental resolution produces positive and negative signs randomly distributed for tracks originating from the primary vertex. But long-lived particles produced in the primary vertex tend to generate decay products with positive signs.

The tracks are associated to a jet with a p_{T}-dependent angular matching in $\Delta R < \text{A}(p_{\text{T}})$. The association cut is given by $\text{A}(p_{\text{T}}) = 0.239 + e^{-1.22 - 1.64 \cdot 10^{-5} \cdot p_{\text{T}}/\text{ MeV}}$. The functional form is optimised to associate the majority of decay products even at low jet p_{T} and to reproduce the collimation of jets with increasing jet p_{T}. Consequently the cut is loosened for low p_{T} due to the more isotropic behaviour of the decay products and tightened for increasing p_{T}. For this b-tagging algorithm an accurate measurement of the impact parameters is essential. Therefore, the track selection includes the requirement that each selected track contains a hit in the innermost tracking layer, the so-called b-layer. Due to its small radius, it provides the most valuable tracking information to derive the impact parameters. The tracks associated to a jet are selected if they fulfil the criteria listed in Table 6.2. In addition two-track vertices formed from these selected tracks (cf. Section 6.2) compatible with long-lived particles, photon conversions and material interactions are identified and the corresponding tracks rejected. The signed transverse impact parameter significance is shown in Fig. 6.3 for all tracks considered by the JetProb tagging algorithm obtained from PYTHIA simulation for light, c- and b-jets. The tagging algorithm compares all signed impact parameter significances $d_0/\sigma(d_0)$ of the selected tracks in a jet to a resolution function $\mathscr{R}$ for prompt tracks. In order to calculate the probability $\mathscr{P}_{\text{track}}$ for one track i

	Selection cuts
$p_\mathrm{T}^\mathrm{track}$	> 1 GeV
$|d_0^\mathrm{PV}|$	< 1 mm
$|z_0^\mathrm{PV} \sin\theta|$	< 1.5 mm
Number of Pixel hits	$\geqslant 2$
Number of b-layer hits	$\geqslant 1$
Number of silicon hits	$\geqslant 7$

Table 6.2: *Track selection requirements used by the impact parameter tagging algorithm. d_0^PV and z_0^PV are the impact parameters of the track in the transverse plane and the longitudinal direction respectively. They are derived with respect to the reconstructed primary vertex.*
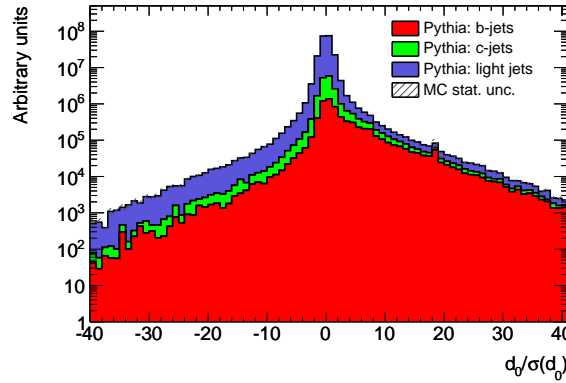
Figure 6.3: *The signed transverse impact parameter significance $d_0/\sigma(d_0)$ of all tracks considered by the JetProb tagging algorithm in simulated QCD events obtained from* PYTHIA. *The jets are selected with the requirements $p_\mathrm{T} > 40$ GeV and $|y| < 2.1$. The single contributions are drawn stacked on top of each other.*

that its origin is the primary vertex, the following quantity is calculated:

$$\mathscr{P}_{\mathrm{track}\,i} = \int_{-\infty}^{-\left| d_0^i / \sigma^i(d_0) \right|} \mathscr{R}(x)dx. \tag{6.5}$$

The resolution function is determined from experimental data with the help of tracks with negative impact parameters, which are dominated from light-flavour jets and reflect the experimental resolution smearing of the impact parameter. The negative side of the signed transverse impact parameter distribution is symmetrised around zero and fitted to build the resolution function $\mathscr{R}$. The jet probability $\mathscr{P}_{\mathrm{jet}}$, which discriminates between b-jets and light jets, is the combination of individual probabilities $\mathscr{P}_{\mathrm{track}}$ of the N tracks with positive transverse impact parameters and is given as follows:

$$\mathscr{P}_{\mathrm{jet}} = \mathscr{P}_0 \cdot \sum_{k=0}^{N-1} \frac{(-\ln \mathscr{P}_0)^k}{k!}, \tag{6.6}$$

where

$$\mathscr{P}_0 = \prod_{k=0}^{N-1} \mathscr{P}_{\mathrm{track}\,i} \,. \tag{6.7}$$

The individual track probabilities $\mathscr{P}_{\text{track}}$ are multiplied to form the product $\mathscr{P}_0$. The jet probability $\mathscr{P}_{\text{jet}}$, called JetProb weight, is calculated from the product $\mathscr{P}_0$ weighted by a factor dependent on the track multiplicity. The JetProb weight reflects the probability that a jet contains no tracks originating from a decay of a long-lived particle. A light jet would correspondingly have a probability near 1. The distribution $-\log_{10}(\text{JetProb weight})$ of the JetProb tagging weight is shown in Fig. 6.4. In this parametrisation the light jets are accordingly localised around zero. With the JetProb algorithm jets are b-tagged if they fulfil $-\log_{10}(\text{JetProb weight}) > 3.25$. This is the operating point for the JetProb tagging algorithm that yields a 50% b-tagging efficiency in simulated $t\bar{t}$ events. This operating point is compatible with the SV0 weight cut at 5.85.
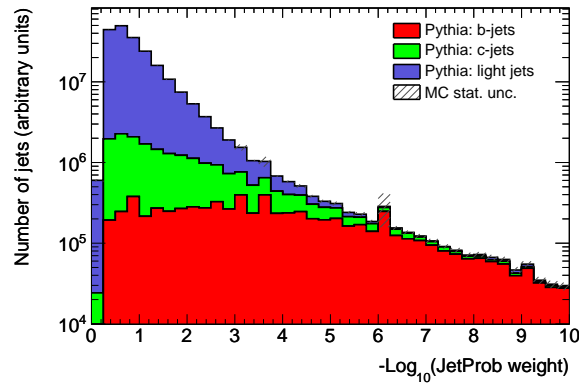
Figure 6.4: *The JetProb weight as a function of $-\log_{10}(\text{JetProb weight})$ shown of the leading jets with $p_{\text{T}} > 40$ GeV and $|y| < 2.1$ in dijet events are shown obtained from QCD* PYTHIA *simulation. The single contributions are drawn stacked on top of each other.*

6.4 Soft muon tagger

The soft lepton tagging algorithm [64] takes advantages of the presence of a lepton (= muon or electron) within a jet originating from a semi-leptonic decay of a b-hadron. The usage of muons is preferred over electrons due to the higher background in the electron sample from QCD, the larger acceptance of the muon reconstruction and deficiencies in the electron reconstruction caused by dead material and the poor performance of the transition region between barrel and end-cap cryostats. The requirement of the presence of a muon limits the bottom-flavour identification by the semi-leptonic branching ratio, however, this method provides a complement method which is mostly statistically independent from the more effective lifetime taggers discussed above. Because of this the soft muons are extensively used for cross-calibration of the lifetime taggers which is presented later (cf. Section 8.2.1).

The key objects for this algorithm are the reconstructed muons [92]. Muons are associated to the closest jet in angular space which fulfils $\Delta R < 0.5$ in the event. Additionally, the following requirements are made to the muons:

- $|d_0^{\text{PV}}| < 4$ mm,

- $p_{\mathrm{T}}^{\mathrm{muon}} > 4$ GeV,

- Matching between the muon spectrometer and inner detector tracks $\chi^2/\mathrm{ndof} < 10$. ndof stands for the number of degrees of freedom.

This tagging algorithm uses as discriminant the p_{T}^{rel} variable which is the momentum of the muon transverse to the combined muon-jet axis, where the jet axis is corrected for the presence of the muon, shown in Fig. 6.5. If more than one muon is associated to the jet, the muon with the highest transverse momentum is chosen. The muons from
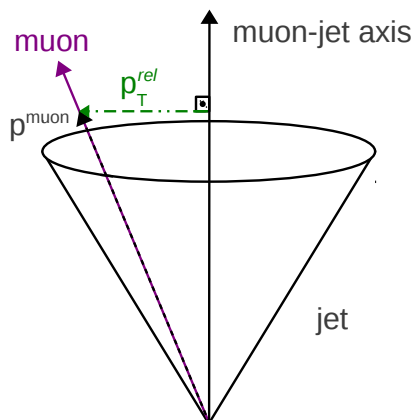
Figure 6.5: *Illustration of the momentum of the muon transverse to the combined muon-jet axis: p_{T}^{rel}.*

bottom decays have a significantly larger transverse momentum relative to the jet axis than muons from light hadron decays. Fig. 6.6 presents this behaviour obtained from QCD PYTHIA simulation, where b-jets tend to have a broad p_{T}^{rel} distribution with the centre shifted to higher p_{T}^{rel} values. These different shapes of the p_{T}^{rel} distributions are used in an one-dimensional likelihood ratio to discriminate b-jets from c- and light jets. The likelihood ratio uses the normalised variable $p_{\mathrm{T}}^{rel}/(p_{\mathrm{T}}^{rel}+0.5$ GeV$)$. For each jet with the help of the muon with the highest transverse momentum the likelihood ratio is given by:

$$Q = \frac{\epsilon_b^0 \times L(p_{\mathrm{T}}^{rel}|b)}{\epsilon_{\mathrm{light}}^0 \times L(p_{\mathrm{T}}^{rel}|\mathrm{light})}, \tag{6.8}$$

where ϵ_b^0 ($\epsilon_{\mathrm{light}}^0$) is the fraction of b-jets (light jets) that contain a muon satisfying the selection cuts and $L(p_{\mathrm{T}}^{rel}|b)$ ($L(p_{\mathrm{T}}^{rel}|\mathrm{light})$) is the probability density function for a b-jet (light jet). The discriminating power of the p_{T}^{rel} variable is diminishing with increasing jet p_{T} due to the stronger collimation of the jets. This causes the muon to be almost collinear with the jet axis and the p_{T}^{rel} distributions start to match very strongly.

6.5 Summary

The taggers presented in detail in the former sections are chosen due to their robustness. The default b-tagging algorithm in the analysis is the SV0 tagging algorithm. A possible
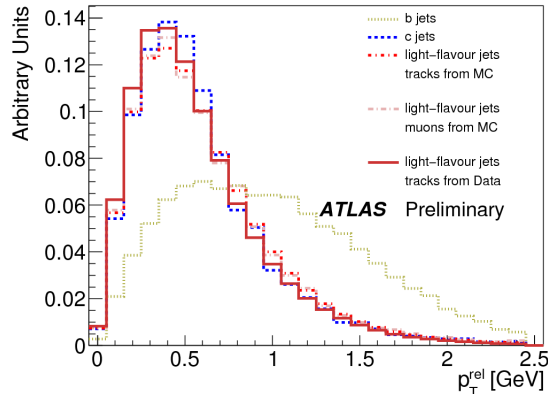
Figure 6.6: *The p_T^{rel} distribution for b-, c- and light-flavour jets obtained by QCD* PYTHIA *simulation. The distributions are all normalised to unit area. Jets are considered with $p_T > 25$ GeV and $|\eta| < 2.5$ and the combined muon track has to satisfy the selection cuts in Table 6.1. Taken from Ref. [95].*

dependence of the measured cross section on the tagging algorithm is checked with the help of the JetProb algorithm. The working point of 50% b-tagging efficiency in simulated $t\bar{t}$ events is used for both lifetime taggers. Correspondingly a jet is b-tagged if the SV0 weight is bigger than 5.85 or if $-\log_{10}(\text{JetProb weight})$ is bigger than 3.25. Their estimated performance from $t\bar{t}$ events is shown in Fig. 6.7. The light-jet rejection is presented as a function of the tagging efficiency for jets with $p_T > 20$ GeV and $|\eta| < 2.5$. The light-jet rejection is decreasing with increasing b-tagging efficiency. An efficiency increase is achieved by lowering the selection cut and correspondingly selecting more jets which do not contain a bottom-flavoured hadron. The advantage of the robustness of the SV0 and JetProb tagging algorithm is achieved at the expenses of a low light-jet rejection with respect to the advanced tagging algorithms. The high-performance taggers SV1, IP3D, JetFitter and their combinations IP3D+SV1 and IP3D+JetFitter can achieve a higher light jet rejection of the order of nearly 2000 at 50% b-tagging efficiency. The SV1 tagger takes advantages of further secondary vertex properties with discriminating power like the invariant mass of the charges tracks in the secondary vertex. The SV1 weight is a the outcome of a likelihood ratio which takes into properties of the secondary vertex and their correlation. On the other is the IP3D tagger the advanced version of the JetProb tagger and utilises the two-dimensional distribution of the signed transverse and longitudinal impact parameter significances. The IP3D tagger is also based on hypothesis tests with a likelihood ratio. The JetFitter algorithm combines the information from the SV1 tagger with the decay length significance evaluated with the help a more advanced fitting algorithm. For more details on the advanced methods please refer to [64, 96]. At the chosen working point of 50% in the simulated $t\bar{t}$ sample the SV0 algorithm corresponds to a light-jet rejection of about 300 which means that for every 300 light-jets one is mis-identified as a b-jet. The light-jet rejection of the JetProb algorithm is approximately 150, one light-jet out of 150 is mis-identified as a b-jet. The rejection power of a b-tagging algorithm rapidly decreases with rising jet p_T. This degradation is caused by the track reconstruction. The track density increases with the jet p_T due to the higher activity and stronger collimation of the jet which enhances the overlap between tracks and the

probability of mis-reconstruction. Therefore a light jet can more easily fake tracks with a significant lifetime causing the light-rejection power to be deteriorated. On the other hand also the *b*-tagging efficiency is a function of jet p_T. Firstly the efficiency increases at low p_T (approximately 20 to 100 GeV), stays approximately constant and is degraded at high jet p_T (approximately 300 GeV). Slower *b*-hadrons generate more isotropic decay products, these may fail to be associated to the corresponding jet or to fulfil the track selection cuts (e.g. $p_T^{track} > 1$ GeV) of the *b*-tagging algorithm in question. With increasing jet p_T and hence increasing *b*-hadron p_T the decay products are boosted in direction of the *b*-hadron. This increases the fraction of tracks from decay products to be associate with the jet and this simply increases the efficiency. The degradation of the efficiency at high jet p_T is caused by the track reconstruction deficiencies which also decrease the light-jet rejection.
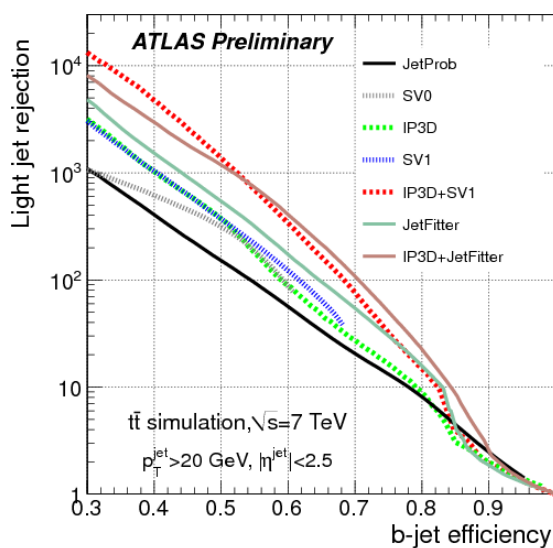
Figure 6.7: *Light-jet rejection as a function of the b-jet tagging efficiency for the early tagging algorithms (JetProb and SV0) and for the advanced algorithms, based on simulated $t\bar{t}$ events from the MC@NLO generator [56]. Details on the advanced algorithms SV1, IP3D, JetFitter and their combinations can be found in [64, 96]. Taken from Ref. [96].*

7 Data and event selection

After giving a description of the detector, the definition of jets and the methods to identify bottom-flavoured jets, the following chapter presents details on the data samples, data cleaning and the event and trigger selection of the analysis. Starting from the trigger system which selects data events online, basic quality requirements on detector level determine the integrated luminosity available for the analysis. Subsequently on event level the reconstructed objects which are subject to this analysis undergo cleaning cuts. Finally, from the sample of reconstructed jets (containing all flavours), in the following called inclusive jets, the jets with bottom-flavoured content are extracted and serve as basis for the cross section measurement.

7.1 Data samples and quality

The data sample which is available for this analysis contains proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 7$ TeV recorded from March to October 2010 with the ATLAS detector. It corresponds to a total integrated luminosity of L=45.0±1.5 pb^{-1}. This data sample is divided into distinct periods A to I corresponding to time periods with constant accelerator configuration (cf. Section 4.2) and detector operation. The analysed sample is selected with the help of the hardware-based calorimeter jet triggers. The early data periods A to F only make use of the lowest trigger level (L1). With increasing luminosity L1 triggers with lower jet energy thresholds were prescaled starting from period C. Finally, the more advanced jet algorithms on the Level-2 (L2) trigger level have become the default selection criteria starting from period G (except for the unprescaled jet trigger with lowest jet energy threshold). In order to maximise the available integrated luminosity different jet triggers are used to cover the entire jet p_T spectrum. The chosen trigger strategy ensures that the event selection by the single jet trigger is nearly 100% efficient. To ensure the quality of the events hardware-based and analysis-based quality criteria are used. The hardware-based criteria requires that the LHC delivers stable beams with a beam energy of 3.5 TeV. The second criterion is the operation of the ATLAS detector or of chosen detectors components with nominal setup, e.g. magnets, inner detector, calorimeter and muon system. If these basic selection criteria are met, an offline analysis-based inspection of a standard set of distributions leads to a data quality assessment for each individual sub-detector, trigger and each type of reconstructed physics object. In the end the final result stated in five different colours (green, yellow, red, grey and black) of each data quality inspection is recorded in a condition database. For details see Appendix C and [97]. The green colour indicates that a sub-system performed with no major problems. For this analysis good data quality is required on detector and analysis level for the tracking, calorimeter and muon system and for the jet triggers. A complete list of requirements is presented in Appendix C. The total integrated luminosity with good data quality is 33.9±1.2 (≈ 34) pb^{-1} which is a fraction of about 72% of the total recorded luminosity. Table 7.1 and Table 7.2 show an overview of the available integrated luminosities with good

data quality per data period and single jet trigger. The luminosities have a systematic uncertainty of 3.4% [61] and include possible prescales of the trigger items.

Period	L1_J5	L1_J10	L1_J15	L1_J30	L1_J55	L1_J75	L1_J95
A	0.3	0.3	0.3	0.3	0.3	0.3	0.3
B	7.7	7.7	7.7	7.7	7.7	7.7	7.7
C	7.9	8.2	8.2	8.2	8.2	8.2	8.2
D	10.1	67.5	278.7	278.7	278.7	278.7	278.7
E	2.9	19.6	223.4	965.2	965.2	965.2	965.2
F	1.4	9.2	8.3	666.1	1741.2	1741.2	1741.2
Total	30.4	112.6	526.6	1926.2	3001.3	3001.3	3001.3

Table 7.1: *Overview of total integrated luminosities of the L1 single jet triggers with good data quality in nb^{-1} for data periods A to F.*

Period	L2_j15	L2_j25	L2_j30	L2_j45	L2_j70	L2_j90	L1_J95
G	1.0	10.9	35.8	182.3	5211.7	4476.3	5531.9
H	0.4	4.5	4.5	20.6	435.1	486.0	6939.3
I	0.7	7.5	7.5	33.0	245.1	798.6	18411.4
Total	2.1	22.9	47.9	235.9	5892.0	5761.0	30882.5

Table 7.2: *Overview of total integrated luminosities of the L2 single jet triggers and the lowest unprescaled L1 trigger with good data quality in nb^{-1} for data periods G to I.*

7.2 Quality criteria for jets

In a first step the data quality cuts at the detector level ensure a general event quality, subsequently a dedicated object cleaning at event level is performed to solve remaining issues. Firstly in the case of reconstructed jets, events with a hard interaction may contain jets not originating from the collision itself but from background sources. Secondly jets may be seeded from real energy depositions but located in problematic calorimeter regions [98]. The following sources can create reconstructed jets not associated to the hard interaction, so-called bad jets:

- Non-collision and cosmic background:

    - Beam-gas events, where an interaction happened between a proton in the bunch and the residual gas within the beam pipe over the length of the ATLAS detector.
    - Muons or pions in the halo of the beam, originating from interactions which occurred not in the proximity of the ATLAS detector with the material from technical equipment in and around the beam-line.

- Cosmic ray muons traversing the ATLAS detector in coincidence with a proton-proton collision and experiencing catastrophic energy loss.[1]

- Calorimeter noise:

  - Sporadically fake calorimeter signals in the HEC calorimeter can occur. In this case the neighbouring calorimeter cells around the fake signal have a negative energy due to the capacitive coupling. The sum of negative energy cells (neg_{E}) is a measure for such fake signals. The reconstructed jets of these signals have a large energy fraction (f_{HEC}) in the HEC calorimeter and can in addition be identified with the help of poor signal shapes. The calorimeter quality ($f_{HECquality}$) is a measure for the amount of energy contributing to a jet from cells with a poor signal shape. A poor signal shape is identified by differences in the sampling with respect to a reference pulse shape.

  - Coherent noise induced by noise bursts in the electromagnetic calorimeter lead to a mis-reconstruction of jets which have a large energy fraction in the electromagnetic calorimeter (f_{EM}). In addition these cells exhibit a poor signal shape. The calorimeter quality ($f_{EMquality}$) is a measure for the amount of energy contributing to a jet from cells with a poor signal shape. A poor signal shape is identified by differences in the sampling with respect to a reference pulse shape.

To reject non-collision and cosmic background jets several properties are checked. The reconstructed jet time (t_{jet}) deduced from the reconstructed calorimeter cell times weighted with the square of the cell energies is required to be in agreement with the event time reported from the trigger. For jets within the tracking acceptance a minimal charged fraction of the jet's transverse momentum (f_{ch}) has to be measured by tracks associated to each jet. A minimal energy deposition of each jet has to be reported by the closest calorimeter (f_{EM}) sub-system to the interaction region. In addition a maximum energy fraction (f_{max}) in any single calorimeter layer is imposed on each jet.

To reject non-collision and cosmic background jets in data two quality selections are used:

- **Loose** selection is designed to conserve more than 99% of the jets originating from a hard interaction and real energy deposition.

- **Medium** selection requires stricter quality criteria than the loose selection but induces a larger inefficiency.

A measurement using inclusive jets requires the medium selection because this selection is capable to reject and control jets in the high transverse momentum region more efficiently. In case of a measurement of b-jets the loose selection criteria can be used because a b-tagged jet and its associated tracks fulfil high quality criteria which exclude any jet originating from a background source or calorimeter noise.

In addition, jets originating from problematic calorimeter regions are rejected. The two main issues leading to such jets are:

---

[1]With increasing muon energy radiative effects in the energy loss mechanism start to become more significant. The probability for such a large energy loss in the calorimeter increases, often referred to as catastrophic.

	Loose	Medium
Had. noise	$(f_{\mathrm{HEC}} > 0.5$ and $\mid f_{\mathrm{HECquality}} \mid > 0.5)$ or $\mid neg_{\mathrm{E}} \mid > 60$ GeV	Loose or $f_{\mathrm{HEC}} > 1 - \mid f_{\mathrm{HECquality}} \mid$
Coherent EM noise	$f_{\mathrm{EM}} > 0.95$ and $f_{\mathrm{EMquality}} > 0.8$ and $\mid \eta \mid < 2.8$	Loose or $f_{\mathrm{EM}} > 0.9$ and $f_{\mathrm{EMquality}} > 0.8$ and $\mid \eta \mid < 2.8$
Non-collision and cosmic background	$\mid t_{\mathrm{jet}} \mid > 25$ ns or $(f_{\mathrm{EM}} < 0.05$ and $f_{\mathrm{ch}} < 0.05$ and $\mid \eta \mid < 2)$ or $(f_{\mathrm{EM}} < 0.05$ and $\mid \eta \mid \geqslant 2)$ or $(f_{\max} > 0.99$ and $\mid \eta \mid < 2)$	Loose or $\mid t_{\mathrm{jet}} \mid > 10$ ns or $(f_{\mathrm{EM}} < 0.05$ and $f_{\mathrm{ch}} < 0.1$ and $\mid \eta \mid < 2)$ or $(f_{\mathrm{EM}} > 0.95$ and $f_{\mathrm{ch}} < 0.05$ and $\mid \eta \mid < 2)$

Table 7.3: *Selection criteria used to reject jets from background sources. Recommendations are taken from [98].*

- The energy measurement of jets which contain dead or deactivated calorimeter cells is corrected by extrapolating the energy deposit from the energy in the neighbouring cells. The energy measurement of jets with a significant fraction of extrapolated energy f_{corr} is not trusted and rejected if $f_{\mathrm{corr}} > 50\%$.

- Jets with a large fraction of energy deposition in the scintillators in the gap between the hadronic calorimeters in the barrel and end-cap f_{gap} are discarded due to the not-validated calibration of this sub-system if $f_{\mathrm{gap}} > 50\%$.

These cuts are established by the analysis of events selected by the L1_MBTS trigger [98]. Jet properties are compared to minimum bias PYTHIA simulation after successive cleaning cuts. The power of the cleaning cuts are presented in Fig. 7.1. Events with good data quality and colliding proton bunches (DQ+Collisions) are selected. The jet p_{T} distribution is showed for different jet cleaning selections. The underlying jet p_{T} distribution does not require any cleaning cut and shows a large tail towards high transverse momenta which is not described by simulation. The two cleaning selections are improving the agreement between data and simulation until data and simulation sit on top of each other.

7.3 Trigger efficiency and usage concept

The hardware-based calorimeter jet triggers are used to select the events. These are discussed in detail in Section 4.3.5. The use of prescales, which are larger than 1 to downscale event rates, deforms the recorded jet p_{T} spectrum. Because of the limited bandwidth, prescales of triggers with lower jet energy thresholds are larger. Due to inefficiencies of the trigger with respect to the offline analysis events are lost. In an analysis these effects of limited efficiencies and rate downscaling must be corrected to restore the original statistics and properties. To optimise the available event statistics different triggers are combined to cover different kinematic regions. Here the determination of the trigger efficiency is presented and subsequently the strategy of the trigger usage is discussed. This strategy combines triggers according to the trigger efficiencies such that it provides a fully efficient selection. Consequently, only the downscaling by the prescales must be corrected for in the analysis which reduces possible systematic uncertainties.
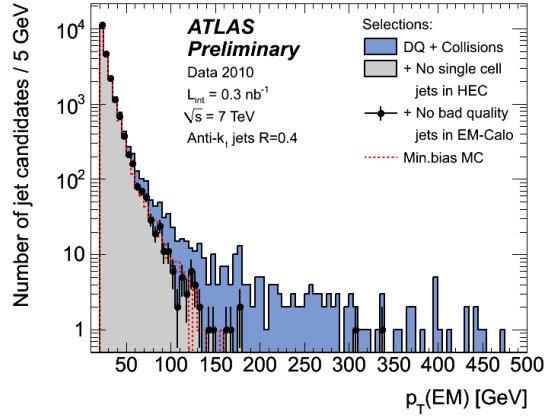
Figure 7.1: *Inclusive anti-k_t jet p_T distribution with distance parameter R = 0.4 after successive cleaning cuts: requiring no jets reconstructed from one noisy hadronic calorimeter cell, rejecting bad quality jets. The jets are reconstructed at the electromagnetic scale. The prediction from minimum bias PYTHIA simulation is superimposed and normalised to the number of jets in data. The outliers above the transverse momentum of 300 GeV are real jets from one single dijet event. Taken from Ref. [98].*

7.3.1 Trigger efficiency measurement

The trigger efficiency is measured on an event basis with the help of an unbiased event sample fulfilling the given analysis selections. The efficiency per event is defined as the probability that the jets in a given event would satisfy the trigger. The trigger efficiency is determined from data because the trigger is not emulated perfectly in simulation. The unbiased event sample is selected with the help of a reference trigger. The efficiency of a trigger as a function of an event property q is given by:

$$\epsilon_{\text{trig}}(q) = \frac{\text{number of events selected by both triggers}(q)}{\text{number of events selected by reference trigger}(q)}. \tag{7.1}$$

For this analysis the trigger efficiencies for the L1 jet triggers in periods A to F are determined as well as the trigger efficiencies of the L2 jet triggers in periods G to I. The efficiencies are determined for jets within the rapidity region $|y| < 2.1$ in events with good data quality. Furthermore events are selected if they contain a primary vertex with at least 10 associated tracks (cf. Section 7.4). The trigger efficiencies are always calculated with respect to jets calibrated with the EM+JES scheme.

Periods A to F: The L1 jet trigger efficiencies are measured with respect to the independent L1_MBTS trigger, which provides an unbiased reference sample. The inefficiency of the L1_MBTS trigger for the events considered is negligible [99]. This method is called orthogonal trigger method. The trigger efficiency for one L1_JX trigger as a function of reconstructed jet p_T, calibrated with the EM+JES scheme, is given by:

$$\epsilon_{\text{trig,L1\_JX}}(p_T) = \frac{\text{jet } p_T \text{ distribution that passes L1\_JX and L1\_MBTS criteria}}{\text{unbiased jet } p_T \text{ distribution that passes L1\_MBTS}}. \tag{7.2}$$

This efficiency is determined for all prescaled triggers and the lowest unprescaled trigger. A jet trigger efficiency distribution is illustrated in Fig. 7.2 as a function of reconstructed

jet p_T of a jet trigger imposing a minimum energy requirement in the calorimeter. The distribution can be described by the shape of a step-function with smoothened turn-on behaviour (Fermi-function). At a certain point the efficiency reaches a plateau from which on the trigger is fully efficient. The plateau is defined as the point where the efficiency reaches over 99%.

The jet trigger efficiencies obtained with the reference trigger L1_MBTS for the data periods A to F are presented in Fig. 7.3. The efficiencies are plotted and the plateaus are indicated by the vertical lines. The dashed lines are extracted from a fit to the efficiency curves. The statistical errors on the trigger efficiencies are obtained by Bayes' theorem calculated at 68.3% confidence-level. [100] This method determines the most probable value of the efficiency given the observed events N_{total} and N_{pass} events out of N_{total} passing the threshold of the trigger. This probability is given by Bayes' theorem: the probability that ϵ_{trig} is the true efficiency given N_{pass} and N_{total} events is the product of the probability of observing N_{pass} events given the efficiency ϵ and the probability of the true efficiency within the allowed integral with obeying the correct normalisation. The process is assumed to be binomial and the efficiency values between $[0, 1]$ are equally qualified so that the most probable value is set to be $N_{\text{pass}}/N_{\text{total}}$. From the probability distribution the 68.3% confidence-level is extracted and used as upper and lower statistical uncertainty. This method is strictly speaking valid for events with weights equal to one and not applicable to QCD simulation. The plateau regions extracted from this plot are written down in Table 7.4.
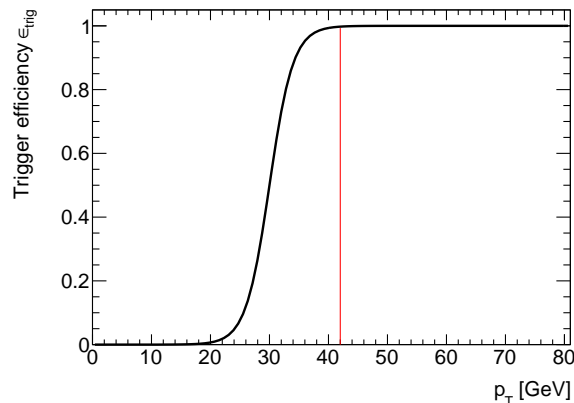
Figure 7.2: *Illustration of a trigger efficiency distribution as a function of reconstructed jet p_T.*

Periods G to I: For data periods G to I the L2 jet algorithms which are seeded by L1 were activated. The L2 trigger efficiencies include the L1 trigger efficiencies. The activation of the L2 jet algorithms and consequently the possible rejection of jets changes the trigger efficiencies.

Due to the large downscaling of the MBTS triggers, they lack the necessary statistics to serve as reference triggers. To obtain the L2 trigger efficiencies two different methods were used:

- Bootstrap method: The unbiased reference sample for each trigger L2_jX (for example L2_j30) is provided by a lower threshold trigger L2_jX$_{\text{low}}$ (for example L2_j20) which

Figure 7.3: *Trigger efficiencies for the L1 single jet triggers used in periods A to F for jets with $|y| < 2.1$. The dashed lines are drawn to guide the eyes.*

Trigger	Trigger plateau [GeV]	Average prescale (A to F)
L1_J5	40	98.79
L1_J10	80	26.66
L1_J15	110	5.70
L1_J30	160	1.56
L1_J55	210	1

Table 7.4: *Beginning of trigger plateaus, at which the trigger efficiency reaches over 99%, expressed in jet p_T and the corresponding trigger for data periods A to F. Due to a misconfiguration in period G which persisted for a short time period L2_j70 recorded a larger fraction of events compared to the L2_j90.*

is on plateau in the region of interest where the higher trigger reaches the plateau:

$$\epsilon_{\mathrm{trig,L2\_jX}}(p_T) = \frac{\text{jet } p_T \text{ distribution that passes L2\_jX and L2\_jX}_{\mathrm{low}}}{\text{unbiased jet } p_T \text{ distribution that passes L2\_jX}_{\mathrm{low}}}. \tag{7.3}$$

- Tag-and-probe method: The unbiased reference sample for each trigger L2_jX is obtained from events which pass any L2_jX. From these events one of the triggered RoIs and its corresponding offline jet which served as tag jet are removed. The remaining probe jets are used to measure the trigger efficiency. Each remaining jet is matched in angular space ($\Delta R < 0.4$) with an L1 RoI and subsequently with an L2 RoI:

$$\epsilon_{\mathrm{trig,L2\_jX}}(p_T) = \frac{\text{probe jet } p_T \text{ distribution of jets with a passing L2\_jX RoI}}{\text{probe jet } p_T \text{ spectrum from any L2 jet trigger}}. \tag{7.4}$$

The bootstrap method is the default choice to determine the trigger efficiencies but it has the disadvantage that it is inapplicable for the lowest single jet trigger. To measure the trigger efficiency of L2_j15 the tag-and-probe method is used. This method determines the efficiency per jet, while the bootstrap method returns the efficiency per event. The efficiency per jet is defined as the probability that a specific jet would satisfy the trigger

condition. The difference between per event and per jet efficiencies is that while in the first case all jets in an event that passed the trigger are positively contributing to the efficiency, in the second case only reconstructed jets matched in angular space to a trigger RoI.

The resulting trigger efficiencies for the data periods G to I are presented in Fig. 7.4 as function of reconstructed jet p_T, calibrated with the EM+JES scheme. The trigger efficiencies are plotted in the kinematic range in which the reference trigger is on plateau. Below the plateau region of the reference trigger the efficiency is biased and unphysical. The vertical lines indicate the beginning of each plateau. The trigger efficiency obtained from the tag-and-probe method for L2_j15 does not reach the plateau at 40 GeV. The subsequent plateaus have not changed with respect to periods A to F. Nevertheless in context of this dijet measurement the inefficiency of L2_j15 for two jets within $|y| < 2.1$ and $p_T > 40$ GeV is smaller than 0.002. This assumption is checked in the full event sample recorded. It is evaluated that in every event with two jets above 40 GeV within $|y| < 2.1$, one of these jets had a L2 RoI which at least 15 GeV. Therefore the trigger plateaus stay unchanged with respect to periods A to F and are summarised in Table 7.5.
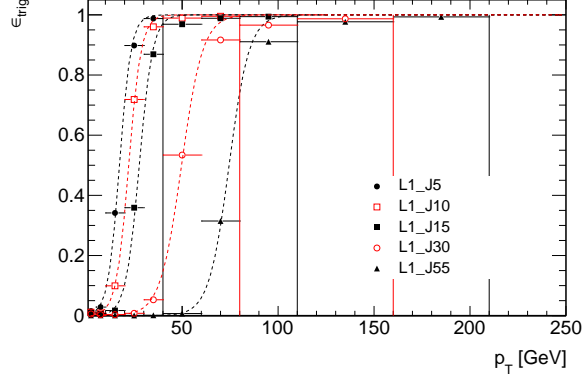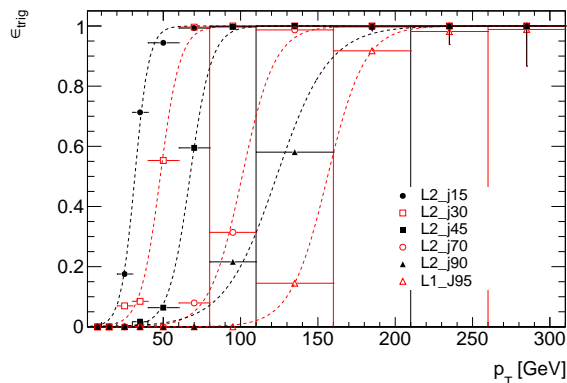

Figure 7.4: *Trigger efficiencies for the L2 single jet triggers used in periods G to I for jets with $|y| < 2.1$. The dashed lines are drawn to guide the eyes.*

Trigger	Trigger plateau [GeV]	Average prescale (A to F)
L2_j15	40	14581.10
L2_j30	80	645.03
L2_j45	110	130.89
L2_j70	160	5.24
L2_j90	210	5.36
L1_J95	> 260	1

Table 7.5: *Beginning of trigger plateaus, at which the trigger efficiency reaches over 99%, expressed in jet p_T and the corresponding trigger for data periods G to I.*

7.3.2 Concept of trigger usage

The concept of trigger usage divides the phase space into distinct regions in terms of leading jet p_T and selects events only by a single trigger in each region, while all other events are dropped. For the smallest possible statistical uncertainty and a nearly fully efficient event selection the trigger in each phase space region is used which provides the largest number of events and is on plateau. These selection criteria are easily deduced from Table 7.4 and Table 7.5. The starting points of the plateaus of subsequent triggers determine the phase space divisions. The original statistics of the selected data sample is restored by applying to each event the weight given by

$$w_{\text{trig}} = \frac{\text{prescale}}{\epsilon_{\text{trig}}}. \tag{7.5}$$

This weighting corrects for the unphysical properties in data caused by the trigger configuration but the data statistic is artificially increased. The statistical uncertainty on the data points does not follow Poisson statistics. The statistical error is given as square root of the sum of w_j weight squares.

The measurement of the dijet cross sections starts by determining the number of events $N(M, p_T)$ as a function of dijet mass and leading jet p_T. The number of events as a function of dijet mass and leading jet p_T is illustrated in Fig. 7.5. Each division in leading jet p_T is full-efficiently measured by one single jet trigger. Every p_T-bin in Fig. 7.5 has its own trigger luminosity L_i which intrinsically includes the rate downscaling by the set of prescales and trigger efficiency $\epsilon_{\text{trig},i}$. The trigger efficiencies for the chosen phase space regions are set equal to one.
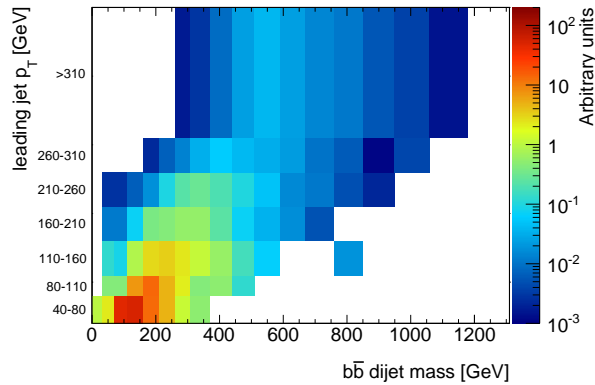
Figure 7.5: *Shown is the number of events $N(M, p_T)$ as a function of dijet mass and leading jet p_T according to the distinct phase space divisions from the trigger efficiencies.*

7.4 Event selection for cross section measurement

Events passing the good data quality criteria and fulfilling the trigger strategy are selected for further processing. The subsequent offline selection steps which are presented here include firstly basic collision-event characteristics and the jet object cleaning requirements. Ultimately analysis-dependent selections on the event topology are made. The complete event selection is given in the following:

1. **Event selection:** Each event is required to have a primary vertex which is reconstructed in agreement with the beam spot measurement (cf. Section 5.2.2) with at least ten associated tracks to ensure a good primary vertex resolution. The distribution of the z-coordinate of the primary vertex in simulated data is reweighted to be in agreement with data (cf. Section 4.4). The width of the z_{PV} distribution is approximately 6 cm.

2. **Jet selection:** Within these events reconstructed jets calibrated with the EM+JES scheme are selected if their transverse momentum is above $p_{\mathrm{T}} > 40$ GeV and they are within the rapidity region $|y| < 2.1$. The single jet trigger with lowest threshold reaches its plateau at 40 GeV. The limited acceptance in rapidity ensures that the jets are completely contained in the tracking volume. Hence tracks with $dR < 0.4$ to the jet axis are fully reconstructed. As the colliding bunches have a spatial width, an interaction may happen displaced to the centre of the detector, tracks may be lost from detection by this shift. However, the effect of the width of the z_{PV} distribution is estimated and found to be negligible ($|\Delta y| < 0.02$).

3. **Cleaning selection:** Jets are excluded if they do not fulfil the jet quality criteria (loose or medium). An event containing a rejected jet is still assumed to be a valid QCD event.

4. **Inclusive dijet selection:** Events are discarded if the number of remaining jets for the analysis is smaller than 2. At this point the event sample is an inclusive dijet sample.

5. $b\bar{b}$ **dijet selection:** The leading jets are selected as bottom-flavour jets by the chosen b-tagging algorithm if they fulfil the b-tagging weight cut. The b-tagging selection cuts are summarised in Table 7.6. For convenience in the rest of the analysis, b-tagged jets refer to SV0 b-tagged jets.

b-tagging algorithm	b-tagging weight cut
SV0	5.85
$-\log_{10}(\mathrm{JetProb})$	3.25

Table 7.6: *Summary of b-tagging weight cuts in use.*

The selection of events in simulation is equivalent to the presented selection steps except for the data quality requirement, the trigger selection and the jet cleaning cuts. Because the evaluation of data and jet quality is not applicable to simulated data. Furthermore, the data is corrected for the trigger downscaling.

Table 7.7 presents the number of events and jets after each selection cut. The trigger selection cut rejects more than 90% of the events. About 3% of all events are selected as a dijet event, containing at least two clean jets with $p_{\mathrm{T}} > 40$ GeV and $|y| < 2.1$. Finally, the prescale-corrected number of events with leading jets which are b-tagged is about 250000.

The jet p_{T} distributions after the selection cuts numbered two to four is presented in Fig. 7.6. Without any jet cleaning in data (cf. Fig. 7.6 (2.)), there is a significant

	Number of events	Number of jets
Data Quality	$2.13 \cdot 10^8$	$7.42 \cdot 10^8$
Trigger selection	$1.35 \cdot 10^6$	$6.70 \cdot 10^6$
Event selection	$1.34 \cdot 10^6$	$6.67 \cdot 10^6$
Inclusive dijet events	$7.74 \cdot 10^5$	$1.88 \cdot 10^6$
SV0 b-tagged dijets (5)	256713	

Table 7.7: *Cut flow summary for events and jets.*

excess of jets in data at high jet p_T. After the jet cleaning cuts presented in Fig. 7.6 (3.) the comparison between data and PYTHIA simulation improves as expected from former studies presented in Fig. 7.1 (taken from Ref. [98]). Finally, the jet p_T distribution of events with at least two clean jets with $p_T > 40$ GeV and $|y| < 2.1$ is presented in Fig. 7.6 (4.). The agreement between data and MC is good. The conventions for the binning of physical quantities in this analysis with non-uniform bin widths are written down in Appendix A. The leading and sub-leading jet are referred to as jet 1 and jet 2, the remaining jets are numerically numbered.
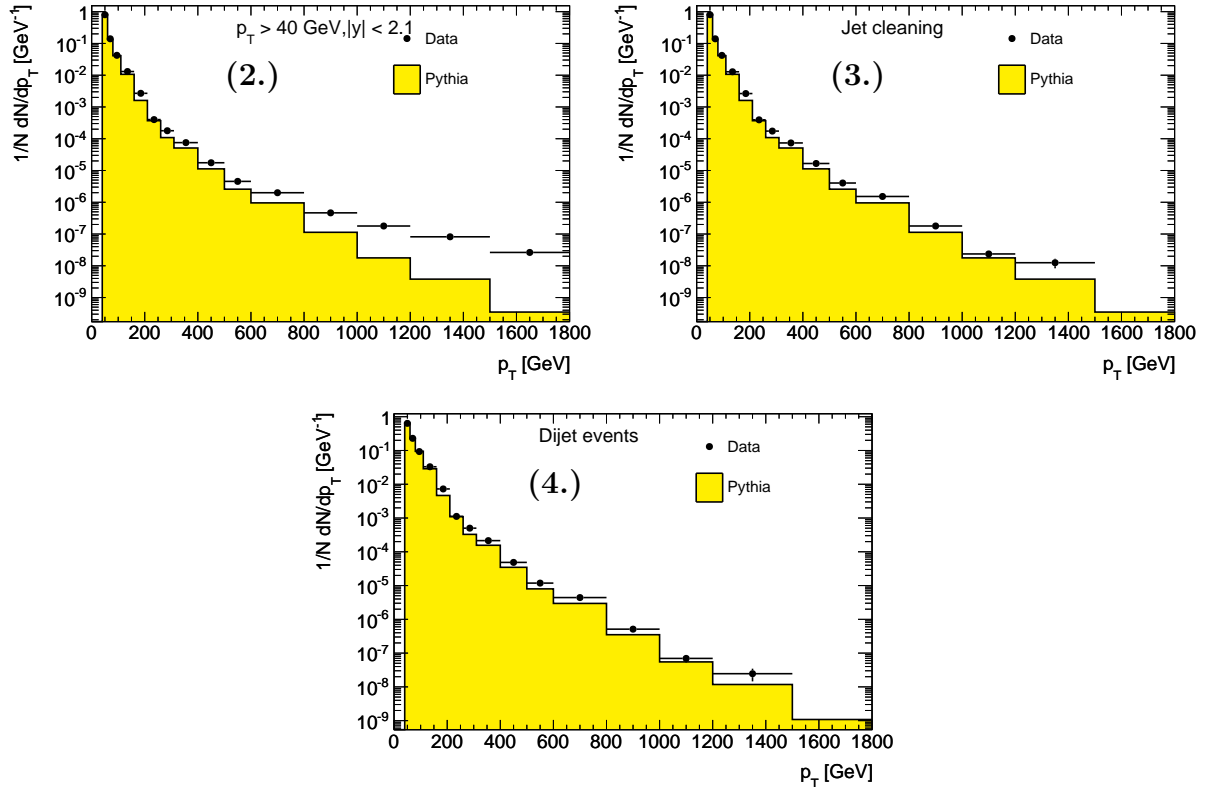
Figure 7.6: *Jet p_T distributions changing with cut flow from cut number 2 (top left) to cut number 4 (bottom left) as defined on page 83.*

8 Object understanding and calibration

In this chapter the event selection presented in the preceding chapter is used to determine the data sample. In Section 8.1 the description of ATLAS data by the PYTHIA simulation is discussed. The focus for this study is on the reconstructed jets and jet related tracking properties. The simulation helps to determine several important quantities like efficiencies, calibrations and correction factors. These can only be trusted if a thorough check on the simulation is performed. Furthermore, possible systematics arising from the uncertainty on the modelling have to be addressed. In Section 8.2 the b-tagging calibration for the SV0 and JetProb algorithms are presented. The extracted calibration is extended and validated to higher jet-p_T regions for this measurement. In Section 8.3 the jet energy scale uncertainty of EM+JES calibrated jets is discussed and the determination of the uncertainty for b-tagged jets is presented.

8.1 Jet and track properties in data and simulation

This dijet cross section measurement relies on an adequate description of data by the simulation. The default PYTHIA samples (cf. 4.4) are used to derive any necessary correction for the cross section measurement which is not deducible from data. This section studies the agreement of analysis related properties between data corresponding to a total integrated luminosity of about 34 pb^{-1} and the PYTHIA predictions. The event selection summarised in Section 7.4 is applied to data and to simulated events. This agreement is quantified by the ratio between data and MC and has to be evaluated taking statistical uncertainties into account from both samples. With the help of inclusive jets, the compatibility of the leading-order pQCD calculations for the hard scatter between protons implemented in the parton shower generator PYTHIA is investigated. The properties of tracks from charged particles associated to these jets are discussed because they are key ingredients for the b-tagging algorithms. Along the selection of b-tagged jets, the discriminant weight distribution of both lifetime taggers is compared between data and simulation. In the next step, all b-tagged jets found in the inclusive data sample are studied and the phase space coverage of the available dataset is illustrated.

8.1.1 Properties of inclusive jets

The production of jets in proton-proton collisions reflect the distinctive signature of short-distance interactions between the partons in a leading-order picture of the proton. The leading jets are identified with the partons emerging from the hard scatter. Additional jets in the event are produced by final- and initial-state radiation, multi-parton interactions and fragmentation. The following comparisons between data and MC are done separately for the leading jets and additional jets. The statistical uncertainty on the MC distributions

are represented by grey hatched areas. Data and MC distributions are either normalised to unit area or the MC distribution is normalised to data.

The properties of the selected dijet sample is presented in the following. The event and object selection for an inclusive jet sample is detailed in Section 7.4. In Fig. 8.1 the leading jets p_T and dijet mass M distributions are presented for reconstructed jets with $|y| < 2.1$ and $p_T > 40$ GeV. The statistics of the simulated data events for inclusive jets is high (cf. Section 4.4) and the statistical uncertainty is hardly visible. The jet p_T spectrum (Fig. 8.1(a)) extends to 1.5 TeV and contains 1.5 million jets. The corresponding measurement of the invariant dijet mass M of these jets in Fig. 8.1(b) covers the phase space to about 3 TeV. The phase space restrictions cause the unphysical decrement of the number of events below $M < 110$ GeV. The observed agreement between data and MC for the inclusive p_T and dijet mass distributions is good but it gets worse with increasing jet p_T. The prediction from PYTHIA tends to underestimate the distribution at high p_T and dijet mass M and consequently overestimates the jets and events in the low p_T and dijet mass range, respectively. This deficiency caused by the limitations to describe the hard scatter will give rise to a systematic uncertainty in the measurement. With this first data at a centre-of-mass energy of 7 TeV from the stable running periods of the LHC in 2010, the measurements from the Tevatron on inclusive jets are surpassed [101] already. The azimuthal and rapidity correlations are determined between the leading jets. At

(a) Jet p_T distribution of leading jets   (b) Dijet mass M distribution of leading jets
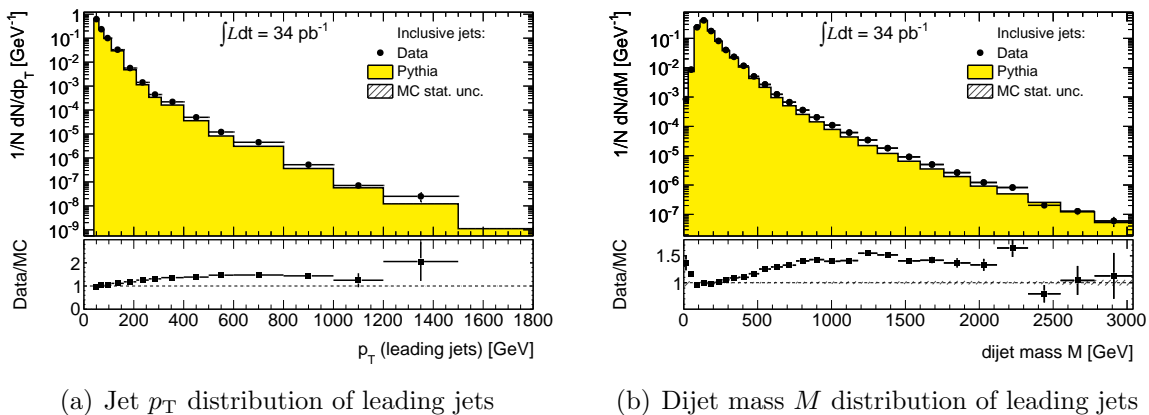
Figure 8.1: *Properties of the leading jets in an inclusive dijet sample in data and QCD* PYTHIA *samples.*

leading-order the partons emerging out of the hard scatter are produced back-to-back. The angular correlation is peaked at $\Delta\phi = \pi$. The angular correlation of the leading jets is presented in Fig. 8.2(a). Next-to-leading order corrections modify the angular correlation and populate the lower regions in $\Delta\phi$. PYTHIA models this higher order effects partly through the parton showers. This approach leads to an adequate description of the shape of the angular correlations between the leading jets above $\pi/2$. Nevertheless, PYTHIA underestimates the contributions at regions with low angular separation. In this region the leading jets, which are selected, originate from a parton pair, which was created in the parton shower. The unsteady shape of the prediction below $\Delta\phi < 1$ is caused by the creation mechanism, namely parton splittings, which are an inefficient procedure from the statistical point of view. The comparison of the PYTHIA prediction and data is as expected (cf. Section 2.3.2). The rapidity correlation (Fig. 8.2(b)) has a maximum at

$\Delta\eta \approx 0$ and follows a falling distribution with increasing $|\Delta\eta|$. This correlation shows a very good agreement between data and simulation within the statistical uncertainty.

The jet multiplicity and the inclusive p_T distribution of the additional jets are shown

(a) Angular correlation of leading jets

(b) Rapidity correlation of leading jets
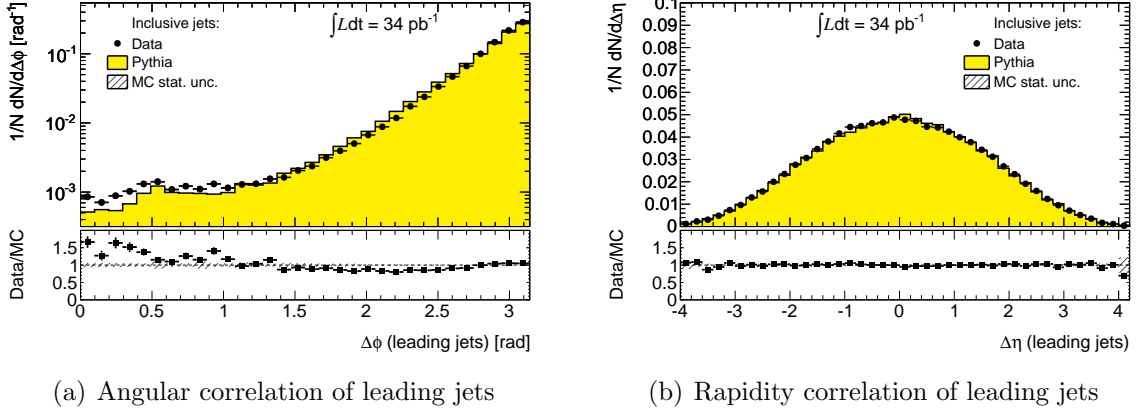
Figure 8.2: *Angular and rapidity correlation of the leading jets in an inclusive dijet sample.*

in Fig. 8.3. Events with exactly 2 jets, fulfilling the cleaning cuts and the phase space restrictions, make up more than 90% of the selected events. PYTHIA describes the jet multiplicity well. Also the p_T distribution of additional jets is well reproduced with the same remarks on the hardness of the distribution as for the leading jets above. Furthermore, the ΔR distributions between the leading jets and the additional jets show good agreement between data and MC (see Appendix E).

Concluding the observations made so far, the description of inclusive jets in data with

(a) Jet multiplicity in inclusive dijet sample

(b) p_T spectrum of additional jets
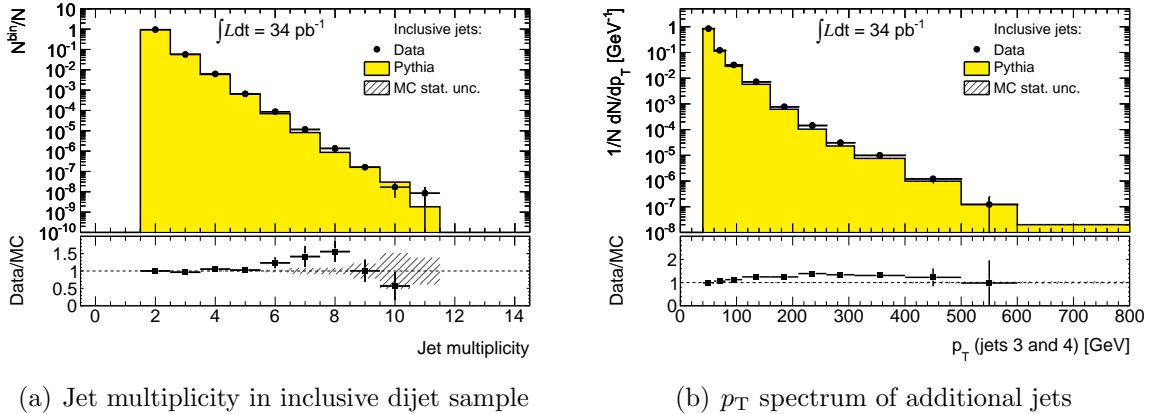
Figure 8.3: *Description of the jet multiplicity and p_T distribution of the additional jets in an inclusive dijet sample.*

the leading-order parton shower generator PYTHIA is good. The deficiencies confirm the expectations from Section 2.3. The underestimation of high p_T jets has to be taken into account in the systematics.

8.1.2 Properties of tracks associated to jets

As was seen, the scattering process of the proton constituents and its kinematics is well described by pQCD and the default generator in this analysis. The structure of the jets generated in the fragmentation process is more complex. With the help of the *b*-tagging track selection cuts presented in Table 6.2 used in the JetProb tagger, the modelling of charged tracks associated to jets is studied. The *b*-tagging track selection requires well-measured tracks and is designed to reject fake tracks as well as tracks from long-lived particles or material interactions.

Hadronisation produces particles with limited transverse momentum with respect to the

(a) All reconstructed tracks   (b) *b*-tagging quality tracks
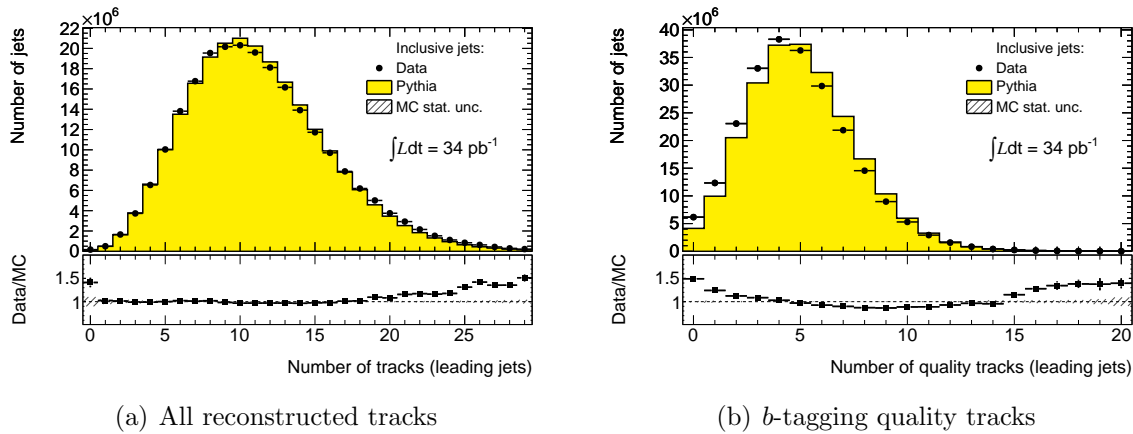
Figure 8.4: *Track multiplicities for differently selected tracks associated to a jet: (a) all reconstructed tracks considered, (b) only b-tagging quality tracks considered.*

parton direction. With increasing transverse momentum of the jet a large fraction of transverse momentum of the jet is concentrated in the jet core and the jet gets more collimated. The contribution of perturbative QCD radiation becomes more significant and counteracts the collimation. Accordingly, the average track multiplicity increases with jet p_T. As studies have shown, jets are in general broader in data than in MC [102]. The track multiplicity for the leading jets predicted by PYTHIA and compared to data is shown in Fig. 8.4. This figure presents the multiplicities for an unrestricted and for the *b*-tagging track quality selections. The multiplicities in Fig. 8.4(a) are almost identical, whereas for the *b*-tagging track selection they are almost 5% lower for data than for MC. The stricter *b*-tagging selection presented in Fig. 8.4(b) rejects tracks and the average track multiplicity decreases. The modelling of the track multiplicity is dependent on the description of the tracking system in simulation and on the hadronisation which is strongly dependent on the chosen phenomenological model.

The transverse impact parameter and its significance distributions for tracks associated to a jet and fulfilling the *b*-tagging selection cut is presented in Fig. 8.5. This distribution is the basis for the lifetime tagger JetProb and the measurement of the impact parameters is essential for finding secondary vertices. It has to be emphasised that the PYTHIA prediction is in excellent agreement with data. This implies several facts: It can be deduced that the disagreement in the track multiplicities are rather caused by the imperfection of the physics tune (cf. Section 2.3.1) than from the modelling of the track reconstruction. Furthermore, the pattern recognition in a high-density environment like the jet core is

(a) Transverse impact parameter

(b) Transverse impact parameter significance
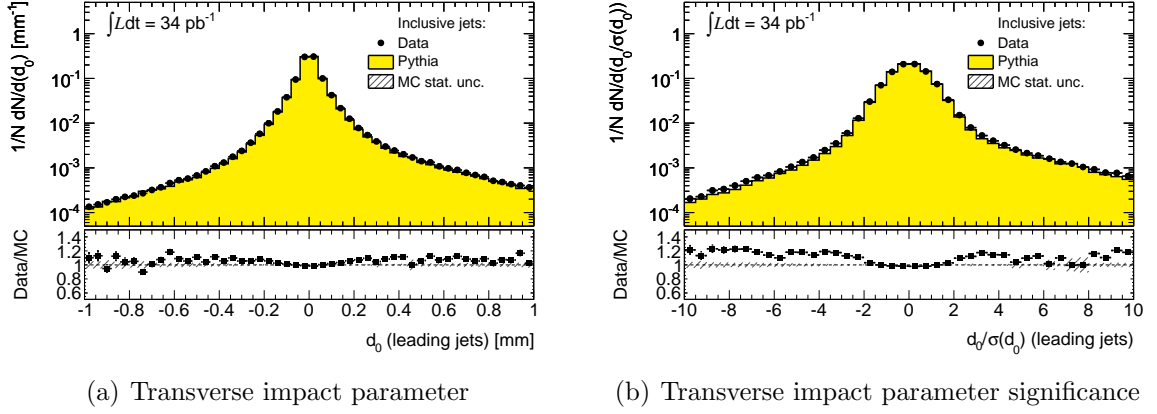
Figure 8.5: *The reconstructed transverse impact parameter and its significance are presented for tracks associated to a jet and fulfilling b-tagging quality cuts.*

well reproduced by the simulation as well as the hit sharing among tracks. In Fig. 8.5(b) the ratio data/MC shows that there is a slight disagreement between data and MC of about 20% at high $|d_0/\sigma(d_0)|$ indicating that the simulated distribution is slightly narrower. From the perfect agreement of the transverse impact parameter distribution, it can be concluded that this slight disagreement comes from the resolution modelling of the primary and secondary vertex. This can be caused by small residual misalignments in the detector (not perfectly described by the simulation) or differences in the primary vertex resolution. In Appendix E Fig. E.3 the longitudinal impact parameter distribution can be found. Finally, the reconstructed transverse momentum of tracks associated with a jet and fulfilling
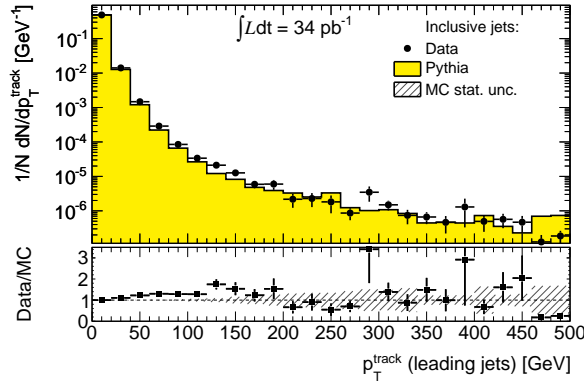
Figure 8.6: *Reconstructed transverse momentum distribution of tracks associated with a jet and fulfilling b-tagging quality cuts.*

b-tagging quality cuts is studied. These cuts require a minimum transverse momentum of the tracks, this ensures that mainly tracks are selected from a hard process rather than from the underlying event. In Fig. 8.6 it can be seen that the transverse momentum of tracks reconstructed in data and MC agrees well and describes an exponentially falling spectrum.

The deviations in the track multiplicity are due to the imperfection of the modelling of the complex processes during fragmentation. The input variables for the b-tagging

algorithms are very well described by the simulation. The performance of the tagging algorithms will therefore be well reproduced by simulation. The discrepancy between the *b*-tagging efficiency and mistagging rate in data and MC will be small. Nevertheless, as the performance of the *b*-tagging algorithm depends on the fraction of transverse momentum of the jet carried by charged tracks and neutral particles, as well as the track multiplicity and track and vertex resolutions, a *b*-tagging calibration is necessary.

8.1.3 Properties of *b*-tagged jets

The discriminant weight distributions produced by the SV0 and JetProb taggers are presented and compared to simulated events separated into the different jet categories: *b*-jets, *c*-jets and light-jets. The $p_{\mathrm{T}}^{\mathrm{muon}}$ spectrum of the identified soft muon with highest transverse momentum is studied as a function of jet p_{T}. Finally, properties of identified *b*-tagged jets with the SV0 tagging algorithm are examined and the covered phase space is illustrated.

To identify *b*-jets, the weight distributions, which are presented in Fig. 8.7, are used. Jets on which these *b*-tagging algorithms failed to find a secondary vertex or to calculate a jet probability are excluded. This failure can have a diversity of reasons. There are no tracks available for the algorithms, the secondary vertex fit is not successful or the likelihood function is undefined for the impact parameter measured. Consequently, the number of jets with no failure in MC is reweighted to the number of jets with no failure in data. The sample of jets shown by the SV0 weight distribution and by the JetProb are not identical, but do have a sizable overlap (¿50%). The description of the weight distribution by Pythia is in remarkable agreement within 10% with data especially in the region above the *b*-tagging selection cut of 5.85 which marks the 50% efficiency in the simulated $t\bar{t}$ sample.

In the reconstruction of the jets based on the calorimeter, identified muons are not taken

(a) SV0 *b*-tagging algorithm   (b) JetProb *b*-tagging algorithm
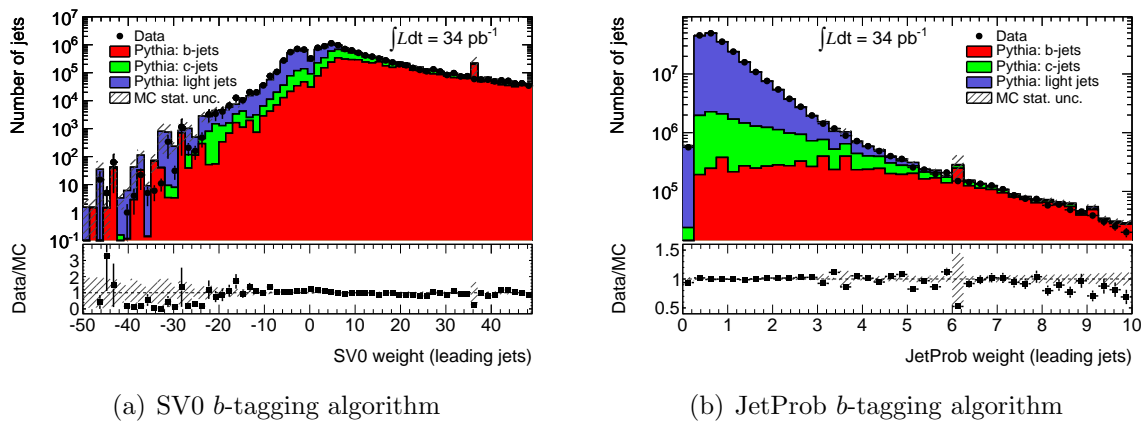
Figure 8.7: *Discriminant weight distribution of the b-tagging algorithms SV0 and JetProb for the taggable, leading jets. The single contributions are drawn stacked on top of each other.*

into account and the escaping neutrinos carry away part of the *b*-hadron's energy. In case of a semi-leptonic *b*-hadron decay with a branching ration of about 10% the transverse momentum loss due to the escaping muon is on average 10% according to Pythia. These detector effects need to be corrected for and therefore the modelling of the hadron decay

is of interest. The $p_\mathrm{T}^\mathrm{muon}$ spectrum of identified soft muons with respect to the beam axis in data (cf. Section 6.4) compared to the simulation indicates how well the hadron decay is described in terms of the fraction of momentum carried by the muon and respectively by the neutrino. The distributions of the transverse momenta in simulation of the soft muons in Fig. 8.8 as a function of jet p_T show reasonable agreement with data.

The sample of b-tagged jets in the following has applied the calibration derived in

(a) Jet p_T: 80-110 GeV

(b) Jet p_T: 210-260 GeV
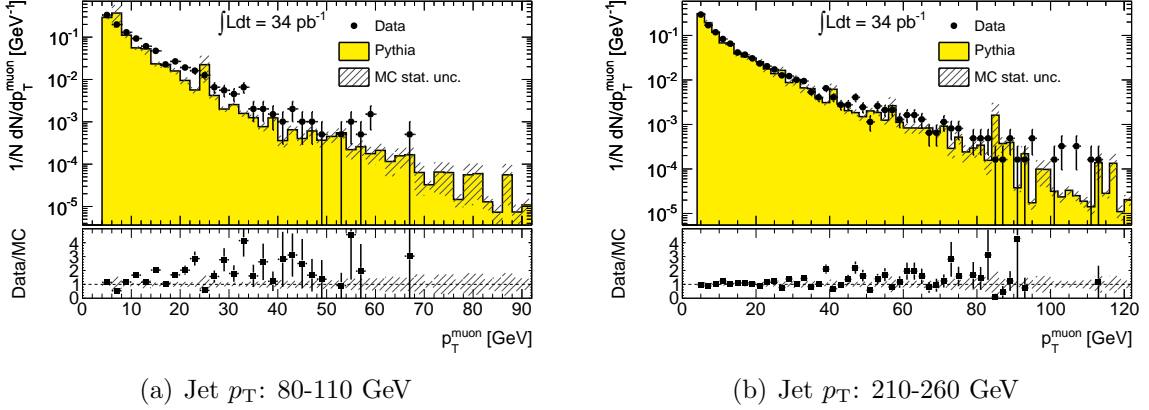
Figure 8.8: *Subset of muon p_T distributions for different p_T ranges according to trigger plateaus: (a) jets with p_T between 80 and 110 GeV, (b) jets with p_T between 210 and 260 GeV.*
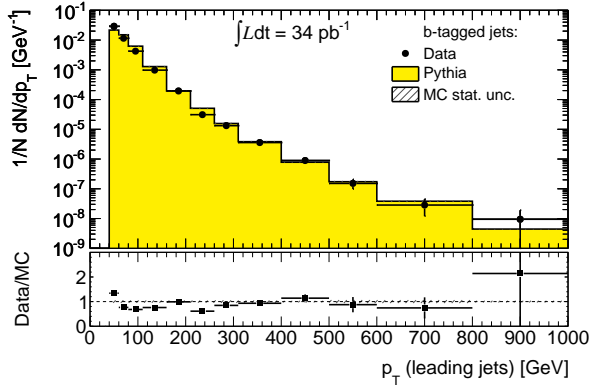
Figure 8.9: *The transverse momentum distribution of the leading jets from events with double b-tagged leading jets.*

Section 8.2. The properties of the b-tagged jets used to measure the dijet cross section are presented in Figs. 8.9 and 8.10. The b-tagged dijets are a sub-sample of the inclusive jets presented in Fig. 8.1(b). The transverse momentum spectrum extends to 900 GeV and the MC sample each bin is separated into three contributions: mistagged light- and c-jets and correctly tagged b-jets. Among all b-jets only a fraction dependent on the b-tagging efficiency is identified. The spectrum is reasonably well described by the simulation. The same is true for the ϕ and η distributions of b-tagged jets in Fig. 8.10. In case of the η distribution it has to be taken into account that rapidity and pseudorapidity are not the same. Therefore, the phase space restriction to the rapidity region below 2.1 is smeared

(a) ϕ distribution
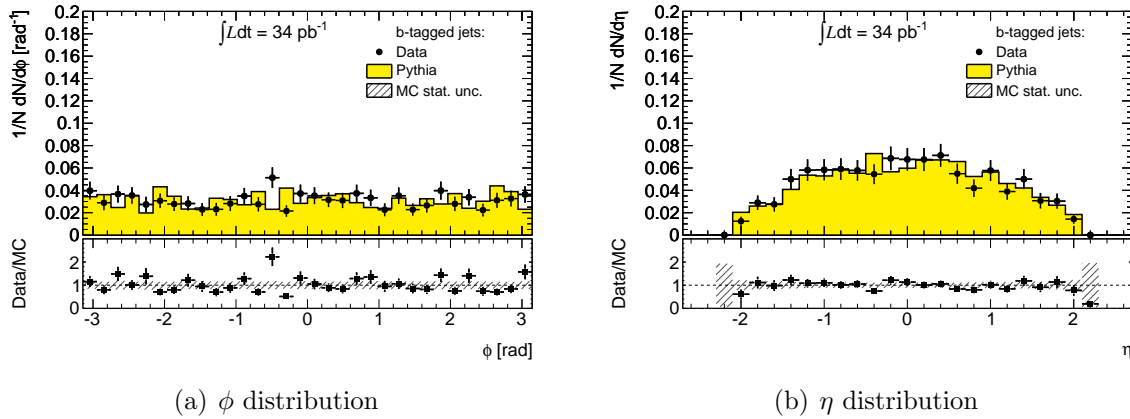
(b) η distribution

Figure 8.10: *Distribution of the number of b-tagged jets as a function of (a) ϕ and (b) η from events with double b-tagged leading jets.*

at the acceptance limits in terms of pseudorapidity. These distributions also reveal the deficiency of the available simulation events, the available statistic is not sufficient. This raises important issues for this analysis, as can be seen later.

In summary, the studies presented justify that the simulated events can be used as a baseline to derive the necessary corrections and systematics for this cross section measurement. The few remaining deficiencies in the simulation modelling have to be taken into account in systematic uncertainties. The final remark is that the presented studies focus on certain important issues in the context of this analysis. This good agreement cannot be extended blindly onto other aspects. In the next sections the primary properties and uncertainties of the b-tagging algorithms and the jet energy measurement are discussed, which are the main focus for this measurement.

8.2 Calibration of bottom-tagging algorithms

The performances of the tagging algorithms SV0 and JetProb in data and simulation are not identical for the reasons we have seen in the former section. Several effects cause differences between data and MC. Simulated resolutions differ from data due to poorly understood geometrical effects which are not modelled in the simulation. A small fraction of the inner detector modules is disabled from time to time. The estimated performance in the simulation depends on the modelling of fragmentation and the description of charged tracks, so consequently on the tune used in the simulation. The measurement of the b-tagging efficiency and mistag rate in data is translated into scale factors that correct the b-tagging performance in simulation to that observed in data:

$$\kappa_{\epsilon_b} = \frac{\epsilon_b^{\mathrm{data}}}{\epsilon_b^{\mathrm{sim}}}, \ \ \kappa_{\epsilon_{\mathrm{light}}} = \frac{\epsilon_{\mathrm{light}}^{\mathrm{data}}}{\epsilon_{\mathrm{light}}^{\mathrm{sim}}}. \tag{8.1}$$

Because both the b- and c-hadrons are long-lived hadrons, it is assumed for the moment that b- and c-tagging efficiencies show a similar behaviour. These scale factors have the advantage of universality in the sense that are applicable to differently selected analysis samples with different event topologies. With the help of the scale factors the efficiency and

mistag rate can be deduced with the help of simulation. The scale factors are determined as a function of jet p_T and η for the reasons presented in Section 6.5. Several methods are available to derive the scale factors. Their applicability is mostly dependent on the available statistics.

The calibration of the efficiency used in this analysis (see Section 8.2.1), due to the limited statistics, is making use of the semi-leptonic decays of b-hadrons into muons. With the help of the soft muons ($p_T^{\mathrm{muon}} > 4$ GeV) and the measured p_T^{rel} distribution the scale factors can be determined. This method is referred to as the p_T^{rel} method. As discussed in Section 6.4 this method breaks down at transverse momenta of the jets of about 150 GeV. The extension of the calibration and its validity for this cross section measurement are presented in Section 8.2.2. Methods favoured in case of high statistics, which are also applicable for high p_T jets, are using $t\bar{t}$ events where the event selections make use of semi-leptonic and dilepton signatures in the $t\bar{t}$ decay channels.[1]

The mistag rate is determined with the help of two methods which are performed on an inclusive jet sample. The use of the inclusive jet sample makes both methods less dependent on statistical effects. The so-called SV0 mass method measures with the help of the reconstructed mass distribution of secondary vertices the fraction of light-jets before and after applying the tagging cut and the efficiency calibration. The negative tag method relies on the fact that the impact parameter or the decay length from light-jets due to the finite resolution of the tracking system is measured as originating in front or behind the primary vertex with respect to the jet axis. This mis-measurement occurs equally distributed in each direction. With the help of the negative side of the impact parameter or the decay length distribution the amount of mistakenly tagged light-jets can be estimated. The difference in the measured and simulated negative tail of the impact parameter and the decay length distribution are corrected by the mistagging scale factor. The results of both independent methods are combined in order to give the best possible estimation and to cover the largest possible kinematic region.

8.2.1 Calibrating the lifetime tagging algorithms

For all methods, distributions in simulation, which are known to disagree with data, are reweighted to reflect the measured distribution to reduce the impact of them on the systematics. The primary vertex z_{PV} distribution is reweighted in the same manner as shown in Section 4.4. In addition the inclusive jet p_T spectrum as presented in Fig. 8.1(a) is harder in data than in simulation and this discrepancy is corrected for by a bin-by-bin reweighting in the simulation samples used for the scale factor determination. The calibration is done for anti-k_t jets with a radius $R = 0.4$ and a minimum p_T of 20 GeV and within the rapidity $|y| < 2.5$. Events are selected if their primary vertex consists of more than 9 tracks. The following sections go into detail on the efficiency and mistag calibration [103].

---

[1]The top(antitop)-quark decays almost 100% into $W+b(\bar{b})$. In semi-leptonic signatures one of the W-bosons decayed into a lepton-pair and the other into a quark-antiquark pair, in dilepton signatures both decayed into a lepton-pair.

8.2.1.1 Efficiency calibration with muons

In the p_T^{rel} method events are collected using the single jet triggers (cf. Sections 4.3.4.1 and 4.3.5) with an additional requirement that a reconstructed muon is spatially matched to the jet object. p_T^{rel} distributions are constructed for b-,c, and light-jets with the help of data and MC. From the p_T^{rel} distributions for each jet flavour templates are derived which describe the shape of the distributions. These templates are used in a likelihood fit which adjusts the relative contributions such that their sum best describes the p_T^{rel} shape in data. The fraction of b-jets is determined from the p_T^{rel} distributions in data before the SV0 b-tagging requirement (f_b) shown in Fig. 8.11(a) and afterwards (f_b^{tag}) shown in Fig. 8.11. The corresponding p_T^{rel} distribution requiring the JetProb tagging selection cut can be found in Appendix F Fig. F.1. The b-tagging efficiency is then derived from

$$\epsilon_b^{\text{data}} = \frac{f_b^{\text{tag}} \cdot N^{\text{tag}}}{f_b \cdot N}, \tag{8.2}$$

where N and N^{tag} are the total number of jets in those samples. The data sample, which

(a) Before b-tagging (b) After SV0 b-tagging cut
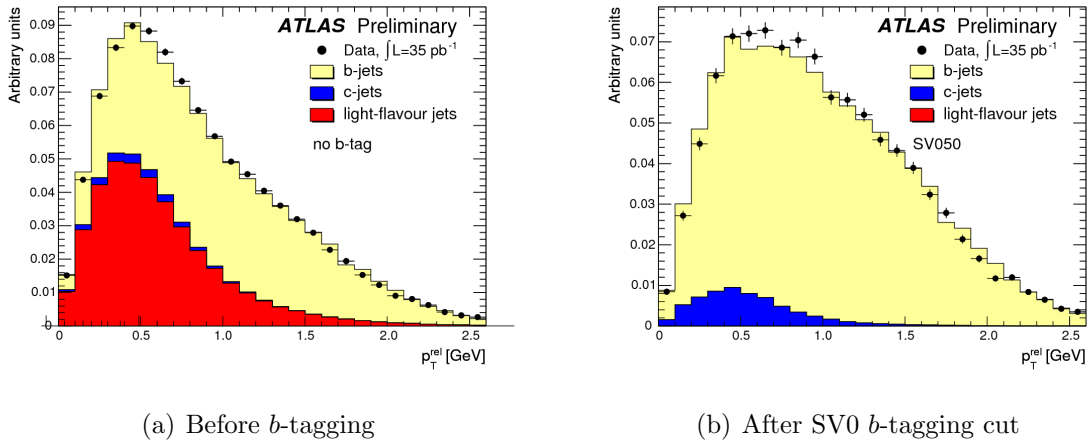
Figure 8.11: *Example of a template fit in data to the p_T^{rel} distribution before b-tagging (a) and after the SV0 b-tagging cut (b) at 50% b-tagging efficiency. The template fit is performed on all jets. Uncertainties shown are for data statistics only. The discrepancies between the data and the sum of the templates are fully covered by the systematic uncertainties on the template shapes. Taken from Ref. [103].*

is used to perform the fit, has the additional requirement to contain at least one jet with a decay length significance (SV0 weight) bigger than 1 (tag) in order to increase the heavy-flavour content in the sample. The corresponding tag jet is discarded for the efficiency measurement and the fit is making use of the probe jet. This requirement diminishes the light-flavour content and is used to minimise the dependency on the modelling and the knowledge on the angular resolution of the p_T^{rel} variable in light-jets.

The b- and c-flavour templates are derived from PYTHIA simulation. The light-jet template is derived from a data sample dominated by light-jets. This data sample contains only events in which no jet is b-tagged by the IP3D+SV1 tagger using a working point with 80% b-tagging efficiency derived in simulated $t\bar{t}$ events (see Fig. 6.7).

Systematic uncertainties affecting this measurement have either a direct impact on the p_T^{rel}

distributions or indirectly by changing the sample compositions or their kinematics. The uncertainties are evaluated by repeating the p_T^{rel} template fits on data with the modified template distributions. Three sources of systematic uncertainties can be identified: the jet measurement, systematics from the method itself and simulation related uncertainties. The uncertainty of the jet energy measurement is affecting the p_T^{rel} sample composition and kinematics. Systematics arising from the method are the limited statistics available to build the templates, the template modelling and the usage of MC derived templates for the b- and c-flavours. Furthermore, it is assumed that the result of this calibration method is also approximately valid on hadronically decaying b-hadrons. The validity of this assumption is dependent on the description of track multiplicity in hadronically decaying b-hadrons in simulated data. The difference between data and MC when comparing the muon and the inclusive jet sample is taken as systematic uncertainty and this causes a sizable uncertainty. Uncertainties caused by the imperfection of modelling the data by the PYTHIA simulation are evaluated by reweighting the corresponding variables to agree with data. In this case the b-hadron direction and the low region of the muon p_T spectrum are not perfectly described. Finally, uncertainties in the simulation concerning branching ratios of semi-leptonic decays, b-fragmentation, modelling of the kinematics of b-hadron decays and the fraction and modelling of $b\bar{b}$ pairs from gluon splitting are assessed with the help of the simulation.

Table F.1 and Table F.2 in Appendix F summarise the different contributions in detail. More details on the mentioned evaluation of the uncertainties are given in Ref. [103].

The resulting scale factors for the SV0 tagger using the 50% efficiency working point and their uncertainties are given in the following:

$$20 < p_T < 30 \text{ GeV} : 0.85 \pm 0.06 \text{ (stat)} \pm 0.09 \text{ (syst)}$$
$$30 < p_T < 60 \text{ GeV} : 0.91 \pm 0.03 \text{ (stat)} \pm 0.07 \text{ (syst)}$$
$$60 < p_T < 90 \text{ GeV} : 0.88 \pm 0.04 \text{ (stat)} \pm 0.03 \text{ (syst)}$$
$$90 < p_T < 140 \text{ GeV} : 0.88 \pm 0.09 \text{ (stat)} \pm 0.08 \text{ (syst)}$$

The detailed numbers for the corresponding working point of the JetProb tagger can be looked up in Appendix F. The scale factor for c-jets is assumed to be the same as that for b-jets but with the uncertainty doubled. Dedicated studies are needed to constrain the c-tagging properties further but are not available at the moment. Both scale factors are illustrated in Fig. 8.12 as a function of jet p_T up to 140 GeV. The scale factors for both taggers are all close to one in all bins of jet p_T. The relative total uncertainty ranges from 4% to 14%. The largest contributions to the systematic uncertainties come from the jet energy measurement, the template statistics and the uncertainty from the assumption of the validity of the calibration on hadronically decaying b-hadrons. The systematic uncertainty on the b-tagging efficiency is large ($\approx 6-20\%$) and affects the dijet cross section measurement significantly. In prospects of much larger available statistics from LHC operation in 2011 and 2012, this uncertainty can be reduced greatly with the help of the calibration method using $t\bar{t}$ events. Furthermore, the calibration can be improved by reducing the jet energy scale uncertainty and by increasing the knowledge on properties of b-fragmentation and decay.

(a) SV0 tagger for SV0 weight > 5.85

(b) JetProb tagger for JetProb weight > 3.25
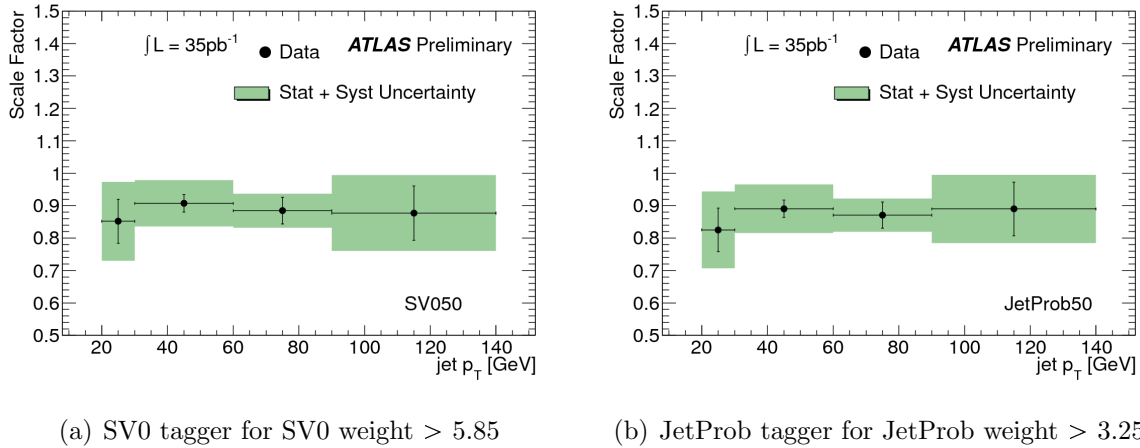
Figure 8.12: *The b-tagging efficiency scale factors from the p_{T}^{rel} method as a function of jet p_{T} for the SV0 tagger (a) and the JetProb tagger (b) using the 50% efficiency working point from simulated $t\bar{t}$ events. Taken from Ref. [103].*

8.2.1.2 Calibration of the mistag rate

The calibration measurements for the mistag rate use an inclusive jet sample selected by the single jet triggers. The mistag rate is the fraction of jets originating from light-jets which are tagged by the b-tagging algorithm.[2] The mistag calibration is derived as a function of the jet p_{T} and rapidity. Two independent methods are used which are combined to give the final result.

SV0 mass method: Similar to the p_{T}^{rel} method, from a discriminant variable, which separates the different flavours, templates are constructed and with their help the fraction of light-jets retained after the b-tagging selection is determined. In this case the templates are constructed from the invariant mass of the charged particles associated to the secondary vertex, denoted as SV0 mass. An SV0 mass distribution is shown in Fig. 8.17. The templates are derived from simulation and fitted to the experimental data after having applied the b-tagging selection cut. The fits determine the fraction of flavours in the b-tagged sample. With the help of the efficiency calibration the number of heavy flavours (b and c) before the tagging selection can be calculated from: $N_{(b,c)} = \frac{N_{(b,c)}^{\mathrm{tag}}}{\epsilon_{(b,c)}}$. The number of light-jets in the data sample before tagging is obtained by subtracting the other flavours from the total number of jets: $N_{\mathrm{light}} = N_{\mathrm{data}} - N_b - N_c$. The mistag rate for light-jets measured in data is given by:

$$\epsilon_{\mathrm{light}} = \frac{N_{\mathrm{light}}^{\mathrm{tag}}}{N_{\mathrm{light}}}. \tag{8.3}$$

Negative tag method: A jet is considered as negatively b-tagged if it fulfils the inverse relation, for example SV0 weight $\leqslant -5.85$. The mistag rate $\epsilon_{\mathrm{light}}$ (as defined in Section 6.1) is estimated by the negative tag rate $\epsilon_{\mathrm{inc}}^{\mathrm{neg}}$ of the inclusive jet sample. Both rates would be identical, if the negative part of the distribution used for flavour discrimination is identical for all jet flavours and the tagging distribution for light-jets was perfectly symmetric

---

[2]Jets originating from u, d or s quarks or from gluons are jointly referred to as light-jets.

around zero (see Fig. 8.7). Therefore, the following corrections have to be applied in order to correct for these discrepancies. The negative tag rate consists of contributions from b-, c- and light-jets. Firstly, the negative tag rate for b- and c-jets differs from the negative tag rate for light-jets. This is because resolution effects on heavy-flavour jets which have measurable lifetimes cause a sign flip of the discriminating variable and this increases the negative tag rate with respect to the light mistag rate. The negative tag rate has to be corrected with a factor $k_{\mathrm{hf}} = \frac{\epsilon_{\mathrm{light}}^{\mathrm{neg}}}{\epsilon_{\mathrm{inc}}^{\mathrm{neg}}} < 1$. Secondly, the negative tag rate misses the effects of material interactions and long-lived particles included in the light mistag rate. The negative tag rate has to be corrected with a factor $k_{\mathrm{ll}} = \frac{\epsilon_{\mathrm{light}}}{\epsilon_{\mathrm{light}}^{\mathrm{neg}}} > 1$. In summary, to extract the light mistag rate from the negative tag rate two corrections factors, which are derived from simulation, need to be considered:

$$\epsilon_{\mathrm{mistag}}(\equiv \epsilon_{\mathrm{light}}) = \epsilon_{\mathrm{inc}}^{\mathrm{neg}} \cdot k_{\mathrm{hf}} k_{\mathrm{ll}}. \tag{8.4}$$

Systematics: Sources of systematic uncertainties on both methods are instrumental effects and uncertainties on the modelling of physics by MC. Both sources cause either modified shapes of the templates or modifications in the correction factors k_{hf} or k_{ll}. The uncertainty by the simulation statistics, the uncertainty on the b-tagging efficiency calibration and the uncertainty on the fraction of jets containing long-lived particles or material interactions are considered as systematic uncertainties in both methods. For the SV0 mass method also the uncertainties on the shape of the constructed templates are taken into account. Discrepancies in the modelling of the track multiplicity and impact parameter resolutions are corrected by reweighting MC to fit data and to evaluate the impact. The resulting relative systematic uncertainties are included in Appendix F. In addition the dependency of the scale factors on the data periods are derived to account for changes in the trigger configuration (for example trigger prescale) and changes in the pile-up conditions. The observed shifts in the resulting scale factors are taken into account in the final result and a systematic uncertainty contributing between 10% and 40% is assigned.

Combination of both methods: For the jet p_{T} bins above 200 GeV only the negative tag method is used due to the available statistics in data and simulation. For each data period the scale factors of both analyses, which show good overall agreement, are combined under the assumption of uncorrelated uncertainties. The final scale factor is derived taking the shifts between the data periods into account in the following way:

$$\kappa_{\epsilon_{\mathrm{light}}}^{\mathrm{data/sim}}(p_{\mathrm{T}}) = \frac{\max(\kappa_{\epsilon_{\mathrm{light}}}^{\mathrm{data/sim}}(p_{\mathrm{T}})) + \min(\kappa_{\epsilon_{\mathrm{light}}}^{\mathrm{data/sim}}(p_{\mathrm{T}}))}{2}, \tag{8.5}$$

where $\max(\kappa_{\epsilon_{\mathrm{light}}}^{\mathrm{data/sim}}(p_{\mathrm{T}}))$ and $\min(\kappa_{\epsilon_{\mathrm{light}}}^{\mathrm{data/sim}}(p_{\mathrm{T}}))$ denote the maximum and minimum scale factor obtained from the average scale factor of both methods over the data periods. The uncertainty, which is the observed spread of scale factors over the data periods, covering the data period dependence is added in quadrature to the systematic uncertainties described above. The combined mistag scale factors and their uncertainties are given in Table 8.1 and illustrated in Fig. 8.13(a) for the SV0 tagging algorithm at the 50% b-tagging working point. The corresponding results are summarised in detail for the JetProb algorithm in Table F.7 and illustrated in Fig. 8.13(b). The mistag calibration scale factors are in agreement with unity within their systematic uncertainty. The major systematic uncertainties arise from the data period dependency of the scale factors, the simulation statistics, the uncertainties

Jet p_{T} [GeV]					
40-60	60-90	90-140	140-200	200-300	300-500
$|\eta| < 1.2$					
1.0±0.2	1.1±0.1	0.9±0.1	1.1±0.1	1.3±0.3	0.9±0.5
$1.2 < |\eta| < 2.1$					
1.2±0.2	1.0±0.2	1.0±0.2	1.0±0.3	0.9±0.6	2±1

Table 8.1: *The mistag rate scale factors for the SV0 tagger for the selection cuts yielding 50% efficiency in $t\bar{t}$ events as function of jet p_{T} and $|\eta|$. The quoted uncertainties include the statistical and systematic uncertainties.*

(a)   (b)
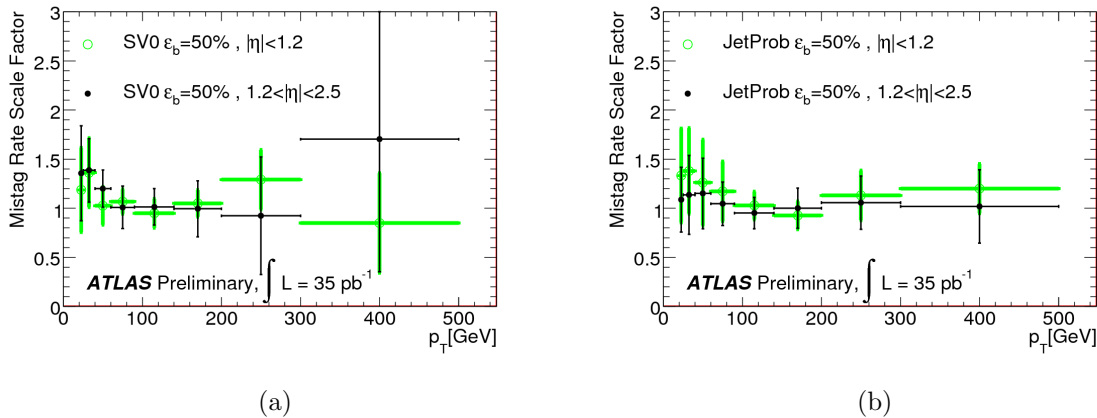
Figure 8.13: *The mistag scale factors after combining the results from the SV0 mass and negative tag method for the SV0 tagging algorithm (a) and the JetProb tagging algorithm (b) for the selection cuts yielding 50% efficiency in $t\bar{t}$ events for $|\eta| < 1.2$ (green) and $1.2 < |\eta| < 2.5$ (black). Taken from Ref. [103].*

on the impact parameter resolution and from the variations in the fractions of long-lived particles and material interactions in jets. The total uncertainties on the scale factor range from approximately 30% at 40 GeV to 10% for intermediate jet p_{T} and are dominated by systematic uncertainties. The total uncertainty in the high-p_{T} region approximately amounts to 50%.

8.2.2 Extension of the calibration and validation

The *b*-tagging calibration with 2010 ATLAS data yielded scale factors close to one with total uncertainties in the range between 5% and 50%. The efficiency scale factors do not show any dependence within the uncertainty on jet p_{T} or rapidity. This is the justification to extend the *b*-tagging efficiency calibration for this cross section measurement to higher kinematic regions. If necessary, additional systematic uncertainties have to be addressed. The calibration corrects for effects concerning the track and vertex resolutions and includes systematic uncertainties which are determined by reweighting simulation to fit data and

(a) 40 GeV $< p_\mathrm{T} <$ 140 GeV

(b) 40 GeV $< p_\mathrm{T} <$ 140 GeV

(c) 140 GeV $< p_\mathrm{T} <$ 300 GeV

(d) 140 GeV $< p_\mathrm{T} <$ 300 GeV
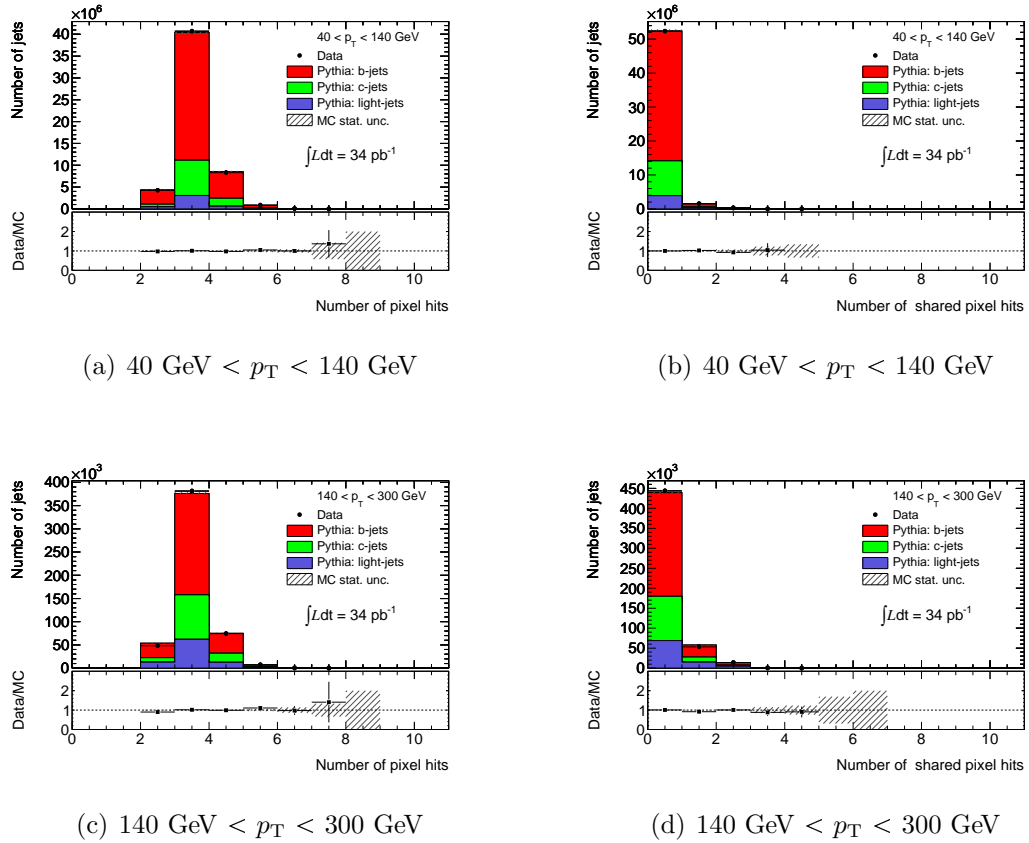
Figure 8.14: *Distribution of the number of pixel hits on tracks in different kinematic regions, for both hit categories: (a,c) all hits, (b,d) shared hits. The tracks are considered by the SV0 tagging algorithm of b-tagged jets. Simulated data is normalised to the number of b-tagged jets in data. The single contributions are drawn stacked on top of each other.*

by examining the uncertainties on the simulation modelling. The ansatz to extend the calibration is to apply the last measured scale factor to jets with $p_\mathrm{T} > 140$ GeV and to double the systematic uncertainty. The agreement between data and simulation is checked between the kinematic regions which are already validated with respect to the calibration discussed in the former section and the kinematic regions beyond. In the following the calibration is extended as explained above and the properties and description of *b*-tagged jets are studied in detail.

The properties of the tracks, associated to *b*-tagged jets fulfilling the event selection and subject to the SV0 tagging algorithm, are investigated depending on the jet p_T. Compared are the following regions: 40 GeV $< p_\mathrm{T} <$ 140 GeV, 140 GeV $< p_\mathrm{T} <$ 300 GeV and $p_\mathrm{T} > 300$ GeV. Important to the validity of the *b*-tagging calibration is the probability and properties of shared hits because they play an important role in degrading the *b*-tagging performance. They mostly influence the impact parameter distribution in the tails (cf. Fig. G.1). Hit sharing in the pixel detector is especially critical to *b*-tagging performance. A track with shared hits, denoted as *shared track*, is defined by having at least one shared hit in the pixel or at least two shared hits in the SCT detector. A track without shared hits is referred to as *good track*. The hit multiplicities measured

(a)

(b)
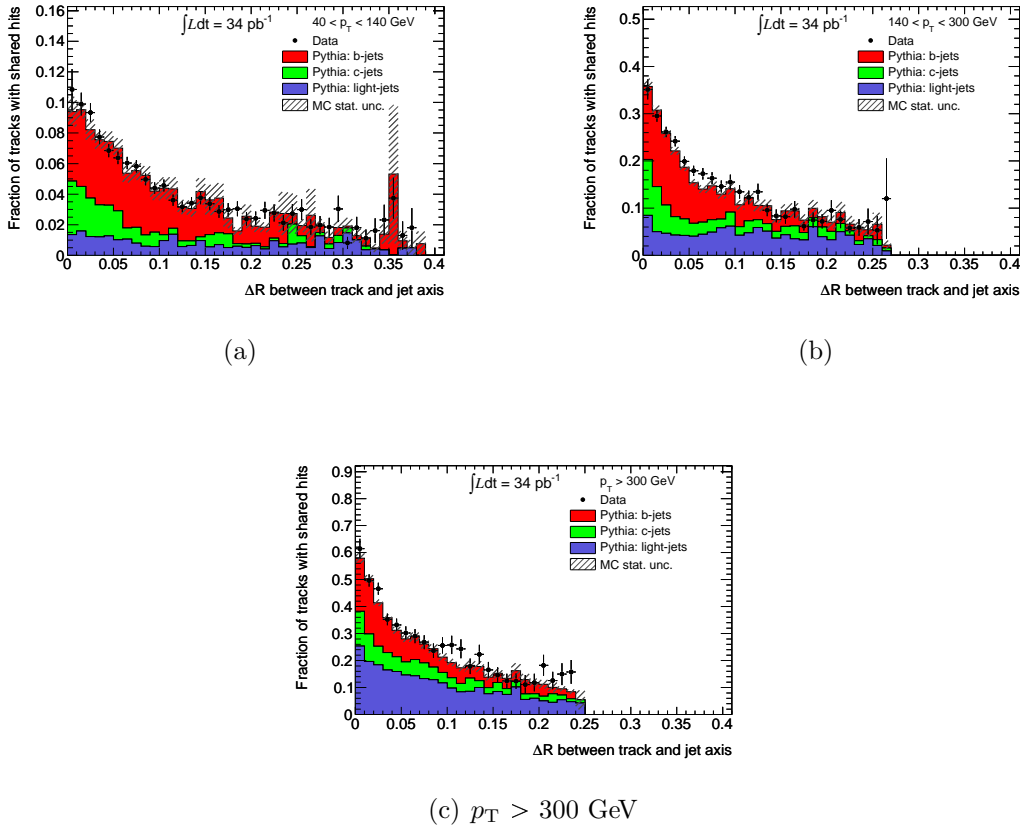
(c) $p_{\mathrm{T}} > 300$ GeV

Figure 8.15: *Distribution of the fraction of tracks with shared hits in b-tagged jets as a function of their distance ΔR to the axis of the jet. Presented are three kinematic regions: (a) 40 GeV < p_{T} < 140 GeV, (b) 140 GeV < p_{T} < 300 GeV and (c) p_{T} > 300 GeV. The single contributions are drawn stacked on top of each other.*

in the pixel detectors for both track categories as a function of the jet p_{T} are shown in Fig. 8.14. The hit multiplicities for the upper most kinematic region is included in Appendix G. On average a track consists of 3 pixel hits of which most are not shared between different tracks. These three space points measured at very small radii from the interaction point are essential for b-tagging. The hit multiplicities for the other inner detector subsystems are included in Appendix G. In general a track also consists of 8 SCT hits and 34 hits in the TRT. The distribution of the number of shared hits for the b-layer, the pixel and SCT detector is peaked at zero for all jet p_{T} ranges. The agreement between simulation and data is very good in general, the number of shared hits is well modelled. There is no significant discrepancy between the calibrated region (Fig. 8.14 top) and the region with the extended calibration (Fig. 8.14 bottom). To validate that the simulation correctly reflects the degradation of the b-tagging performance in the higher kinematic regions, the increase of the fraction of shared hits as a function of their distance from the axis (transverse profile) of the jet in angular space is studied. Besides the frequency and distribution of tracks with shared hits among the sub-systems in the inner detector, this transverse profile is important. The prediction of PYTHIA for the fraction of shared hits in tracks associated to b-tagged jets is compared to data in Fig. 8.15. It is clearly

Figure 8.16: *SV0 weight (decay length significance) distribution of b-tagged jets in events where the leading jets are both b-tagged in bins of reconstructed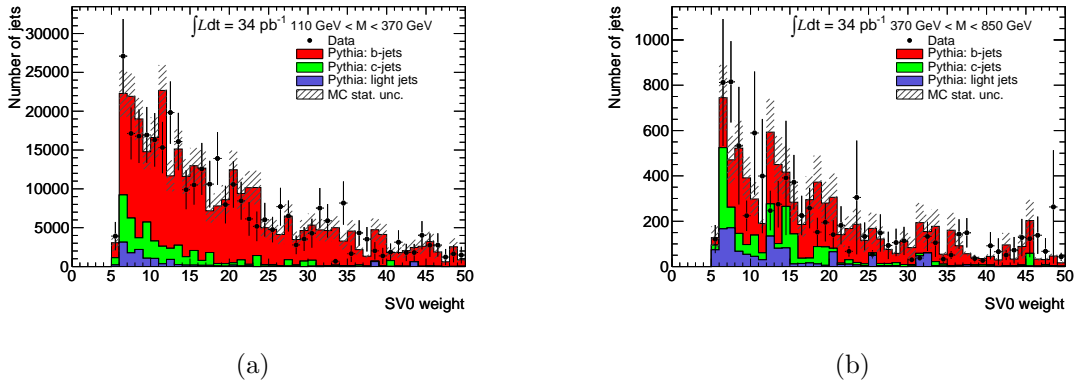 dijet mass: (a) 110 GeV < M < 370 GeV and (b) 370 GeV < M < 850 GeV. The single contributions are drawn stacked on top of each other.*

visible that the fraction of tracks with shared hits is larger for tracks closer to the axis of the jet. These distributions also reflect the fact that the light-jet rejection is strongly dependent on the jet p_{T}. The fraction of mistagged light-jets is heavily increasing in each subsequent kinematic region. The fraction of shared hits among the tracks associated to a jet is more than doubled between subsequent kinematic regions. The fraction increases from about 10% to 60% at $\Delta R \to 0$ with increasing jet p_{T} bin. This effect is well modelled by the simulation in the region covered by the b-tagging calibration as well as in the region probed above 140 GeV.

The former paragraph discussed the entire sample of b-tagged jets, the next distributions focus on the properties of b-tagged jets in events where the leading jets are both b-tagged and which are used in the cross section measurement. The SV0 weight distribution of these b-tagged jets is presented in Fig. 8.16 in bins of reconstructed dijet mass M. The dijet mass bins approximately represent the kinematic regions with validated calibration (110 GeV < M < 370 GeV) and the kinematic region subject to this validation (370 GeV < M < 850 GeV). In contrast to the fraction of light-jets and c-jets remaining in Fig. 8.15 the requirement of a double b-tag in events improves the purity and the fraction of light-jets is reduced. In the higher mass region the number of events is strongly decreased, but the overall agreement between data and simulation is good. The mass of the reconstructed secondary vertex among those b-tagged jets is presented in Fig. 8.17. These templates nicely demonstrate the discriminating power of the secondary vertex mass.

Each b-tagged jet is a single entry in the histogram, namely two entries per event. But these figures cannot show the correlation between the b-tagged jets nor is it possible to extract the number of signal events for the $b\bar{b}$ cross section measurement. The SV0 mass of light-jets peaks in general below 1 GeV, whereas c- and b-jets contain a secondary vertex with an average mass of 2 GeV. Both discussed quantities, the SV0 weight and the SV0 mass distribution separated into different mass regions, are well described by the simulation. Both distributions are important for the cross section measurement. The SV0 mass is well modelled because the tracking in simulation can be trusted. In Appendix G

several additional cross checks are shown including the description of jets with a SV0 weight below 5.85 which also consolidates the validity of the calibration for jets with $p_T > 140$ GeV.

In summary, the cross section measurement makes use of the *b*-tagging efficiency calibration

(a) 110 GeV $< M <$ 370 GeV

(b) 370 GeV $< M <$ 850 GeV
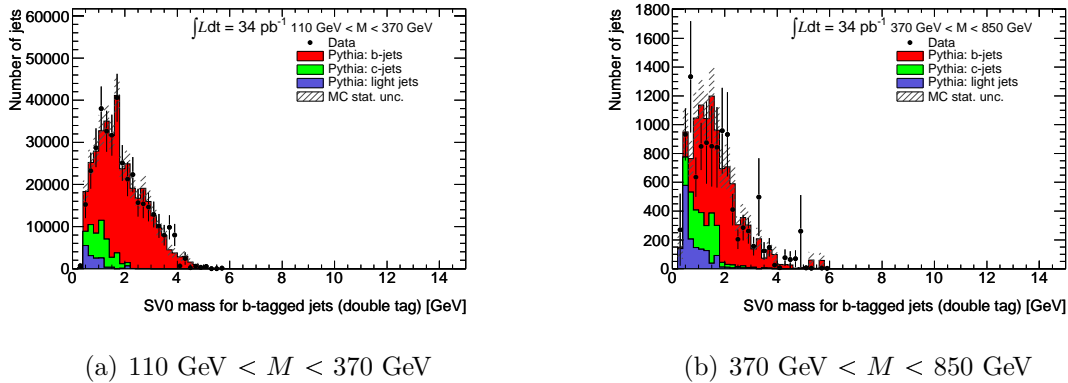
Figure 8.17: *The distribution of the invariant mass of the charged tracks in the secondary vertex in events where both leading jets are b-tagged. The single contributions are drawn stacked on top of each other.*

extended for jets with a transverse momentum higher than 140 GeV. The validity of this assumption is proven by the cross checks presented. It can be estimated that doubling the systematic uncertainty is sufficient and conservative enough. In Section 9 it will shown that an extension of the mistag scale factors is not necessary and that they are not of concern to the cross section measurement. But the extension of the mistag scale factors is included in the histograms presented. Furthermore, in the following description of systematics for the analysis the extension is addressed once more and, if necessary, additional systematic uncertainties have to be assigned.

The resulting calibration scale factors are the following:

$$20 < p_T < 30 \text{ GeV} : 0.85 \pm 0.06 \text{ (stat)} \pm 0.09 \text{ (syst)}$$
$$30 < p_T < 60 \text{ GeV} : 0.91 \pm 0.03 \text{ (stat)} \pm 0.07 \text{ (syst)}$$
$$60 < p_T < 90 \text{ GeV} : 0.88 \pm 0.04 \text{ (stat)} \pm 0.03 \text{ (syst)}$$
$$90 < p_T < 140 \text{ GeV} : 0.88 \pm 0.09 \text{ (stat)} \pm 0.08 \text{ (syst)}$$
$$p_T > 140 \text{ GeV} : 0.88 \pm 0.09 \text{ (stat)} \pm 0.16 \text{ (syst)}$$

8.3 Jet energy scales and resolution

The algorithms used to obtain the four-momenta of jets and the calibration to correct jet energies at the hadronic scale have already been presented. The precision of this established energy scale determines the accuracy of many measurements. A jet energy scale uncertainty of 1% which is the goal of ATLAS in the long term results in an uncertainty of about 10% on the jet cross section in the intermediate p_T range. For this analysis, the inclusive jet

energy scale uncertainty ranges from approximately 2% to 4% for jets within $|\eta| < 2.1$ and between $p_\mathrm{T} > 40$ GeV and $p_\mathrm{T} < 800$ GeV. The following section discusses this uncertainty and its origin as well as the determination of the jet energy resolution with a data-driven method. In principle, the jet energy uncertainty depends on the initial parton type by which a jet was initiated. This arises from the different fragmentation behaviour of different parton types. However, it is not derived separately for each parton type since in the data this is a priori unknown. In Section 3.1.3 it was already presented that the fragmentation of b-jets is harder compared to light-quark and gluon jets. Due to the presence of the b-hadron the particle composition in such jets differs with respect to the majority of jets on which the energy scale uncertainty has been derived from. To obtain an estimate of an additional uncertainty on b-jets a data-driven method is tested making use of tracking information.

8.3.1 Inclusive jet energy scale uncertainties

The jet energy scale (JES) uncertainty [60] for inclusive jets which are calibrated with the EM+JES calibration scheme (cf. Section 5.1.3) is determined as function of jet p_T and η. It is estimated from combinations of data-driven techniques [104], a single pion test-beam measurement [105] and analyses based on MC simulations. Details on the nominal PYTHIA MC samples, on which the calibration is derived from, are given in Section 4.4. The sources of the systematic uncertainties and their effect on the response (cf. Eq. 5.2) of EM+JES calibrated jets are discussed. The systematic uncertainty has the following contributions:

- The uncertainty due to the calibration method.

- The uncertainty due to the calorimeter response.

- The uncertainty due to the detector simulation.

- The uncertainty due to the physics model and parameters embedded in the MC event generators.

- The uncertainty due to the relative calibration of jets with $|\eta| > 0.8$.

Calibration method: After applying the EM+JES calibration scheme on the sample on which the calibration is derived from, small deviations from unity in the response $\mathcal{R}$ (cf. Eq. 5.2) at low p_T remain (non-closure). The reconstructed jets are therefore not completely restored to particle level. Firstly, this is caused by the assumption in the EM+JES calibration that each constituent in a jet needs the same average compensation. Secondly, the calibration does not distinguish between jet energy and momentum, any jet with a non-zero jet mass is slightly distorted in transverse momentum by the EM+JES energy calibration. The systematic uncertainty due to the non-closure is taken as the largest deviation of the response from unity between energy and p_T.

Calorimeter response: The jet energy response and the scale uncertainty is derived from the response and the uncertainty of the single particles interacting in the calorimeters which contribute to a jet. With the help of the simulation the energy deposited in each calorimeter cell can be linked to the initiating particles. The uncertainties for charged hadrons are measured by:

- The measurement of the single hadron energy in a cone around an isolated track with respect to the track momentum (E over p measurement) in the momentum range from $0.5 < p^\mathrm{track} < 20$ GeV.

- The measurement of the response of single charged pions of a momentum between 20 and 350 GeV in the combined test-beam performed in 2004, where a full slice of the ATLAS calorimeter has been assembled.

The absolute electromagnetic energy scale is established using $Z \rightarrow ee$ decays for the electromagnetic calorimeters and using the energy loss of minimally ionising muons in the Tile calorimeter. The uncertainty found on this electromagnetic scale is considered for all other particles in the jets not covered by the E over p measurement. Additional uncertainties which are taken into account are related to the calorimeter acceptance, the uncertainties for particles with $p_\mathrm{T}^\mathrm{track} > 400$ GeV and uncertainties connected to neutral hadrons.

Detector simulation: Uncertainties due to the detector simulation are caused by the modelling of the electronic noise and the material description in the simulation.

As described in Section 5.1.2, the topological clusters are reconstructed using the signal-to-noise ratio of calorimeter cells. Hence, the mis-modelling of the noise in the simulation can lead to differences in shape and in the appearance of fake clusters. The noise in the calorimeters is time-dependent due to for example small instabilities in the voltage supply, whereas the noise injected in the simulation is not. Electronic noise in the ATLAS calorimeters is dependent on the architecture of the calorimeter. In the electromagnetic calorimeters the noise is of the order of 10 to 100 MeV, in the hadronic calorimeters the noise is of the order of 100 to 600 MeV.

Therefore, the modelling of noise cannot completely reflect the noise in data. This effect is estimated by using the RMS of the noise measured in data to reconstruct jets from topological clusters in simulation. The maximum observed change in the jet response between the nominal sample and the modified sample is used to estimate the uncertainty on the jet energy measurement.

The jet calibration was derived using the geometry simulated in the nominal MC sample (cf. Section 4.4). Generated MC samples with variations in the material budget are studied. In general, the data-driven analyses are not directly affected by this because they use reconstructed data. Nevertheless the track selection in the E over p measurement and all particles not included in the data-driven E over p measurement with $p > 20$ GeV are affected by discrepancies in the material description.

Physics modelling in the event generator: The contribution to the jet energy scale uncertainty from the modelling of the fragmentation, the underlying event and different MC models are obtained with the help of the following samples:

- ALPGEN+HERWIG+JIMMY: The ALPGEN generator [106] interfaced with HERWIG and JIMMY is used to determine the contributions to the uncertainty arising from differences in the modelling of the hard process generation and the soft physics modelling. For these samples an MC10 tune is used with the CTEQ6L1 parton distribution [11]. Compared to the nominal PYTHIA simulation, this sample is different in every single generation step: modelling of the multi-parton interactions, the parton shower, the fragmentation model and the underlying event.

- Perugia2010 tune: Besides the MC10 tune, the Perugia tunes [107, 108] are independent tunes which are derived from recent hadron collider data. Adjustments to final-state radiation, to the production of particles with strangeness and to measurements of jet fragmentations are included. This tune is used to mainly investigate effects of soft physics modelling.

The jet response for each sample is determined and the ratio of the nominal response to the response for each of the two samples is used to estimate a systematic uncertainty. **Intercalibration:** The JES uncertainty for forward jets in the end-cap and forward regions ($|\eta| > 0.8$) is determined using the uncertainty from central jets ($|\eta| < 0.8$) from the single particle response and from the systematic variations of the MC simulations. The uncertainty is transferred into the forward region with the help of transverse momentum balance in dijet events. The events used require at least one central jet. Events are selected for this dijet balance method, if they posses a perfect back-to-back dijet topology $|\phi_1 - \phi_2| > 2.6$. In case of an additional jet, the transverse momentum carried by this third jets is restricted to a small fraction of the average transverse momentum of the leading jets $p_{\mathrm{T,3}}^{\mathrm{EM}} < 15\% \frac{p_{\mathrm{T,1}} + p_{\mathrm{T,2}}}{2}$. The central jet is exploited to measure the forward jet's relative response. The following effects contribute to the total jet energy scale uncertainty for forward jets:

- Deviations between the relative jet response in data and simulation are considered in the systematic uncertainty by adding those in quadrature.

- The total uncertainty from the reference jets are kept as baseline.

As part of the jet energy calibration, a correction to each jet is applied to correct for the additional energy contributed by pile-up interactions. To estimate the remaining uncertainty in the jet energy scale after this correction the following studies are done: the effect of variations in the trigger selection on the average additional energy C_{pileup} in each jet constituent is studied, the construction of the jet correction from the average number of jet constituents $\langle N_{\mathrm{constituents}}(\eta) \rangle$ is validated, jets formed from charged particle tracks ($p_{\mathrm{T}}^{\mathrm{track}} > 0.5$ GeV and $|\eta| < 2.5$) which are input to the anti-k_t algorithm are matched to calorimeter jets to measure the variations in the corrected calorimeter jet energies as a function of the number of primary vertices N_{PV}, the non-closure of the correction at the constituents level is evaluated as a function of the number of primary vertices N_{PV}.[3] Beyond the limited coverage of the track-jets, the dijet balance method is also here used to transfer the uncertainty in the more forward regions. The total systematic uncertainty from the pile-up correction is a function of N_{PV} and is the quadratic sum of all uncertainties.

Summary of the jet energy scale uncertainty: The total uncertainty on the jet energy scale is the quadratic sum of all individual contributions. In the central region ($|\eta| < 0.8$), the uncertainty as function of jet p_{T} and $|\eta|$ is composed from:

- The final non-closure uncertainty which is accounting for the uncertainty on the calibration procedure is the maximum deviation from unity between the energy and p_{T} response.

- The calorimeter response uncertainty is estimated from the propagation of single particle uncertainties to the jets.

- Contributions to the uncertainties using Monte-Carlo samples with a systematic variation are accounted for as described in the following:
  The ratio between the response in the sample $\mathcal{R}_{\mathrm{var}}$ with the corresponding variation

---

[3] These are so-called track-jets. They are insensitive to pile-up effects and can therefore be used as reference objects to study the pile-up correction.

and the response in the nominal sample $\mathcal{R}_{\mathrm{nom}}$ is used as follows to estimate the deviations caused by systematic variation:

$$\Delta_{\mathrm{JES}}(p_{\mathrm{T}}, \eta) = \left| 1 - \frac{\mathcal{R}_{\mathrm{var}}(p_{\mathrm{T}}, \eta)}{\mathcal{R}_{\mathrm{nom}}(p_{\mathrm{T}}, \eta)} \right|. \tag{8.6}$$

This deviation is determined for the energy as well as for the p_{T} response. The final contribution to the JES uncertainty is given by the larger deviation in either response to the final JES systematic uncertainty due to the specific systematic effect:

$$\Delta_{\mathrm{JES}}(p_{\mathrm{T}}, |\eta|) = \max(\Delta_{\mathrm{JES}}^{\mathrm{E}}(p_{\mathrm{T}}, \eta), \Delta_{\mathrm{JES}}^{p_{\mathrm{T}}}(p_{\mathrm{T}}, \eta)). \tag{8.7}$$

In addition the uncertainties from the material description are adjusted. The uncertainty estimated from additional material in the inner detector for the E over p measurement is scaled by the average fraction of particles forming the jet that are within the acceptance $p < 20$ GeV. The uncertainty from the additional material in front of the calorimeter for particles not covered in the E over p measurement is scaled by the average fraction of such particles within a jet.

For each (p_{T}, η)-bin, the uncertainty contributions from the calorimeter, the jet calibration non-closure, and systematic Monte-Carlo simulation variations are added in quadrature. Furthermore, for jets in the regions beyond $|\eta| > 0.8$, the intercalibration contribution is added in quadrature to the total JES uncertainty with the exception of the non-closure term that is taken from the specific η-region. The contribution to the uncertainty due to additional proton-proton interactions is added event-by-event, because it depends on the number of primary vertices in the event. In the following the JES uncertainties assume only one single interaction in the event. For jets with pseudorapidity $|\eta| < 2.1$ and one additional interaction the pile-up uncertainty is less than 2% and monotonically decreasing with jet p_{T}. The uncertainty increases with each additional pile-up vertex, with four vertices the uncertainty is overall less than 6%.
This uncertainty derived on a QCD inclusive jet sample reflects the specific composition of gluon- and quark-initiated jets in QCD. Differences in the event topology from this mixture can lead to a dependence on the jet energy scale. Such analyses like the one presented have to account for these differences due to a different jet flavour mixing.
Finally, the fractional jet energy scale systematic uncertainty and the individual contributions are shown for two selected pseudorapidity regions in Fig. 8.18. The jet energy scale uncertainty which is the baseline for jets considered in this analysis is between 2% and 4%. This does not include the uncertainty from the pile-up correction. The uncertainty from the non-closure for all pseudorapidity regions and jets with $p_{\mathrm{T}} > 30$ GeV is smaller than 1%. The uncertainty is driven in the high-p_{T} region by the uncertainty on the calorimeter response from the single particle response and adds 1.5% up to 4% to the uncertainty. The uncertainty from the noise modelling in MC is negligible for jets with $p_{\mathrm{T}} > 45$ GeV and from the detector modelling well below 1% for all kinematic regions. At low p_{T} the uncertainty from the MC modelling accounts for approximately a constant contribution of 1.5% and decreases with increasing jet p_{T}. In summary, the maximum uncertainties in different η regions is presented using representative jet transverse momenta of 20, 200 and 1.5 TeV.

(a) in the barrel region
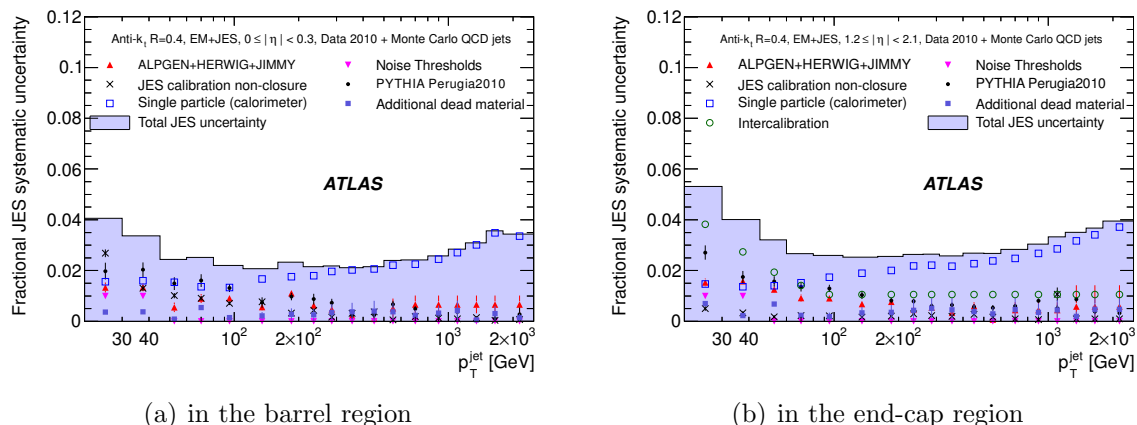
(b) in the end-cap region

Figure 8.18: *Fractional jet energy scale systematic uncertainty as a function of p_T for jets in the pseudorapidity region $0.3 < |\eta| < 0.8$ in the calorimeter barrel (a) and in the region $1.2 < |\eta| < 2.1$ in the calorimeter end-cap. The total uncertainty is shown as the solid light shaded area. The individual sources are also shown together with uncertainties from the fitting procedure if applicable. Taken from Ref. [60].*

8.3.2 Jet energy resolution

The energy smearing of particles and correspondingly for jets of fixed energy is due to the limited energy resolution $\sigma(E)/E$ of the calorimeters from fluctuations inherent in the development of showers, and by instrumental and calibration limits. For jets an additional aspect influences the resolution. The particle composition within the jet has an impact on the total jet energy resolution. Fluctuations in the shower development are due to the probabilistic nature of interactions of the particles in the jet with the calorimeter material. This statistical process improves with increasing energy ($\propto 1/\sqrt{E}$). An energy-independent term due to instrumental effects contributes to the energy resolution ($\propto 1/E$) and a constant term is limiting the resolution at very high energies due to non-linearities and non-uniformities. The total energy resolution is consequently given by:

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{8.8}$$

where a, b and c are determined by the calorimeter parameters. In analogy the transverse momentum resolution is defined and is for simplicity also referred to as jet energy resolution.

8.3.2.1 Resolution of inclusive jets

For the jet energy measurement the knowledge on the jet energy resolution is important as it defines the dijet mass resolution. The currently applied EM+JES calibration is not designed to improve the energy resolution of the jets. The calibration only shifts the average response of jets closer to unity. The modelling of the jet energy resolution in simulation has to be validated using data-driven methods in order to be able to correct these smearing effects for the cross section measurement.

The jet energy resolution in data is measured using the transverse momentum balance of the leading jets in QCD dijet events [109]. At leading order the transverse momentum

η region	Maximum fractional JES Uncertainty		
	$p_T = 20$ GeV	200 GeV	1.5 TeV
$0 < \|\eta\| < 0.3$	4.1%	2.3%	3.1%
$0.3 < \|\eta\| < 0.8$	4.3%	2.4%	3.3%
$0.8 < \|\eta\| < 1.2$	4.4%	2.5%	3.4%
$1.2 < \|\eta\| < 2.1$	5.3%	2.6%	3.5%
$2.1 < \|\eta\| < 2.8$	7.4%	2.7%	

Table 8.2: *Summary of the maximum* EM+JES *jet energy scale systematic uncertainties for different p_T and η regions from the Monte-Carlo simulation based study for anti-k_t jets with $R = 0.4$. Taken from Ref. [60].*

of the two jets in a dijet topology is conserved. An asymmetry between the transverse momenta of the two leading jets $A(p_{T,1}, p_{T,2})$ is defined as

$$A(p_{T,1}, p_{T,2}) = \frac{p_{T,1} - p_{T,2}}{p_{T,1} + p_{T,2}}. \tag{8.9}$$

The asymmetry is described by a Gaussian distribution and is sensitive to the jet energy resolution. The relation between the Gaussian σ of the asymmetry distribution and the relative jet resolution for jets in the same rapidity region is given by:

$$\sigma_A = \frac{\sqrt{\sigma(p_{T,1})^2 + \sigma(p_{T,2})^2}}{\langle p_{T,1} + p_{T,2} \rangle} \simeq \frac{\sigma_{p_T}}{\sqrt{2} p_T} \Rightarrow \sqrt{2} \sigma_A = \frac{\sigma_{p_T}}{p_T}. \tag{8.10}$$

The average p_T of the dijet system is given by $p_T^{\mathrm{avg}} = \frac{p_{T,1} + p_{T,2}}{2}$. To be less sensitive to soft radiation effects and to enrich the events with back-to-back configurations, the following selection cuts are required: the number of jets in the event is required to be smaller than 4, $\Delta\phi$(leading jets) > 2.6 and the transverse momentum of the possible third jet has to be smaller than 10 GeV at the electromagnetic scale. A subset of measured asymmetries for inclusive jets with $|y| < 1.2$ is given in Fig. 8.19. The data and simulation prediction are overlaid and found to be in reasonable agreement. The asymmetry has the shape of a Gaussian distribution and each distribution is fitted to extract σ_A. In order to correct for effects due to the presence of additional soft jets not detected in the calorimeter or not reconstructed by the jet reconstruction, the asymmetry is recomputed allowing for a third jet with a series of allowed p_T values $p_{T,3}^{\mathrm{EM}} < \{5,7,8,10,15,30\}$ GeV. The fitted σ of the asymmetry as a function of the threshold value on the third jet is fitted with a straight line to extrapolate down to $p_{T,3} \to 0$. This extrapolation gives a data-driven estimate on the expected resolution on an ideal dijet topology. The dependence of the asymmetry on the presence of a third jets is presented by two examples in Fig. 8.20. These figures show the relative resolution extracted from the asymmetry distributions as a function of threshold value $p_{T,3}$. The asymmetry in each jet p_T region is recalculated for these six threshold values: 5, 7, 8, 10, 15 and 20 GeV. A soft radiation correction factor is obtained from the ratio between the extrapolated value at $p_{T,3} \to 0$ and the value at 10 GeV. This
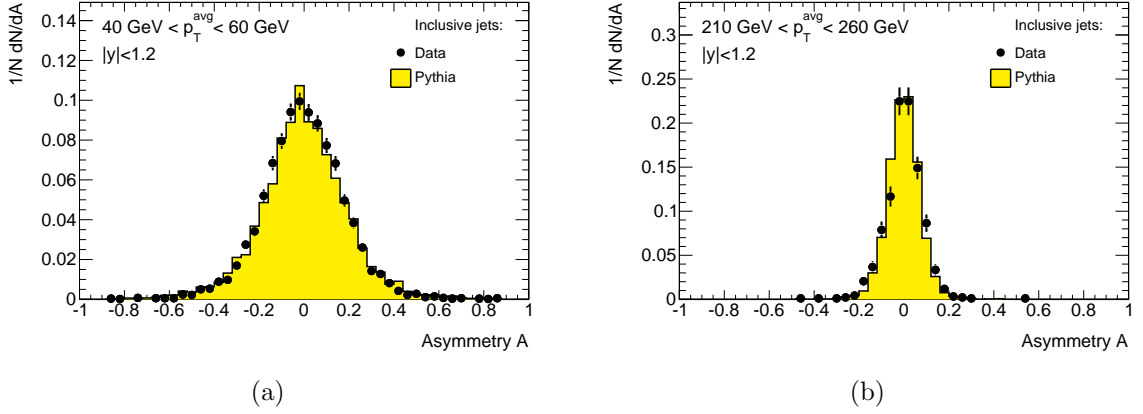
(a)   (b)

Figure 8.19: *Asymmetry distributions for inclusive jets with $|y| < 1.2$ as defined in Eq. 8.9 for two different p_T^{avg} bins: (a) $40 < p_T^{avg} < 60$ GeV, (b) $210 < p_T^{avg} < 260$ GeV.*

correction is then applied to the measurement satisfying $p_{T,3} < 10$ GeV:

$$K_{soft} = \frac{(\sigma(p_T)/p_T)_{p_{T,3} \to 0 \text{ GeV}}}{(\sigma(p_T)/p_T)_{p_{T,3} \to 10 \text{ GeV}}}. \tag{8.11}$$

The correction factor is a rising function of p_T^{avg} and converges for high transverse momenta towards 1. Over the full range, it takes values between 0.8 and 1. If, due to statistics, the correction factor should marginally become bigger than one, it is forced to one. The resulting fitted transverse momentum resolutions are presented in Fig. 8.21. The shape of the resolutions obtained in data and simulation are in agreement. The lower parts of Fig. 8.21 show the differences between the resolutions obtained from data and simulation. The dotted lines indicate a $\pm10\%$ difference. The resolution in data is underestimated by about 10%. With the help of the fitted resolution curves it is estimated that this tendency continues for jets with $p_T > 300$ GeV. More detailed studies of the resolution on ATLAS

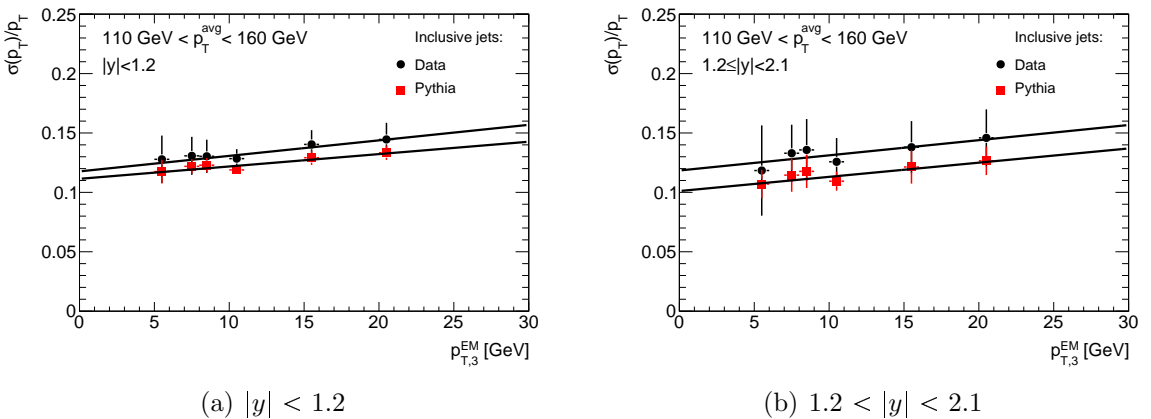(a) $|y| < 1.2$ (b) $1.2 < |y| < 2.1$

Figure 8.20: *Fitted resolution extracted from the asymmetry distributions as a function of the threshold value applied on $p_{T,3}$ for the p_T region : (a) $|y| < 1.2$, (b) $1.2 < |y| < 2.1$. Only statistical uncertainties are shown.*
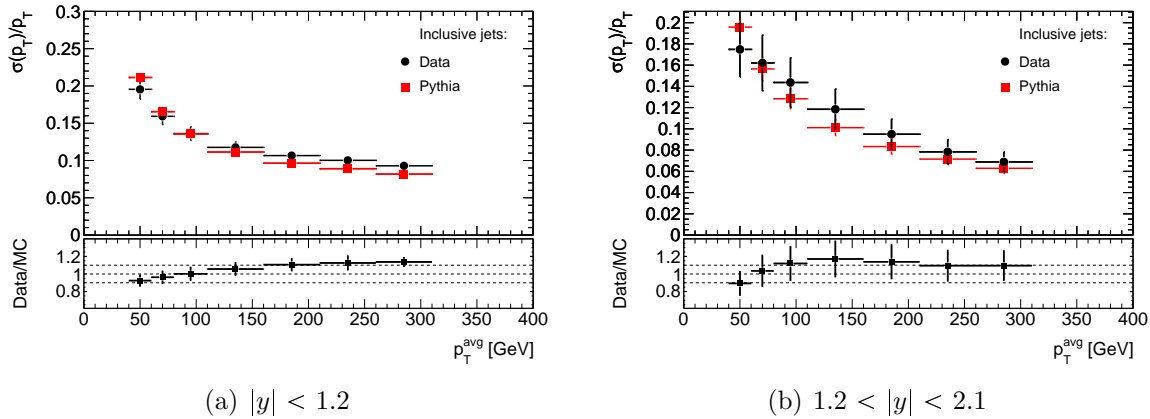
(a) $|y| < 1.2$ (b) $1.2 < |y| < 2.1$

Figure 8.21: *Jet momentum resolution measured using the dijet balance in QCD events as a function of p_T^{avg} between 40 and 300 GeV. The soft radiation effects are corrected. The dashed lines in the lower plots indicated a deviation of $\pm 10\%$.*

data can be found here [109]. The p_T difference measured with the dijet balance is not only due to the resolution affects in the detector but also includes energy losses in the reconstruction of each jets due to the showering and fragmentation (so-called out-of-cone effects). With the help of the simulation this effect can be corrected for but this does not change the relative deviation of about 10% between data and simulation. Such a correction effects the lower p_T region more strongly and yields a correction factor of about 10% for jets around 40 to 60 GeV.

The resolution embedded in the ATLAS simulation measured from the jet response of reconstructed jets matched to particle jets for an inclusive jet sample is presented in Fig. 8.22. The resolution is identified with the standard deviation of the Gaussian distributions of the jet response. This resolution derived from the simulation is not expected to be in perfect agreement with the outcome of the dijet balance method which is corrected for soft radiation and energy losses due to the showering. This is because the jets entering in the jet response calculation are taken from all event topologies. With the help of the resolution extracted from simulation in Fig. 8.22 pseudo-experiments will later be used to estimate the uncertainty arising from the mis-modelling of the jet energy resolution in simulation. The jet resolution in simulation is degraded conservatively by 10% which is the outcome of the dijet balance measurement for all rapidity and kinematic regions.

8.3.2.2 Resolution of jets with heavy-flavour content

Due to the limited statistics detailed studies on the basis of the dijet balance method for b-tagged jets are not possible. The uncertainty covering the deviation in the resolution between data and simulation from inclusive jets is already conservative. There is no evidence that this deviation is much larger or different for b-jets therefore the knowledge from the dijet balance method is transferred to b-tagged jets. The resolution of b-jets is extracted from simulation the same way as discussed for the inclusive jet sample and presented in Fig. 8.23. The dashed lines represent the outcome of a fit describing the resolution curve with Eq. 8.8. The bin sizes used in the dijet mass cross section measurement are chosen accordingly to the dijet mass resolution to contain approximately
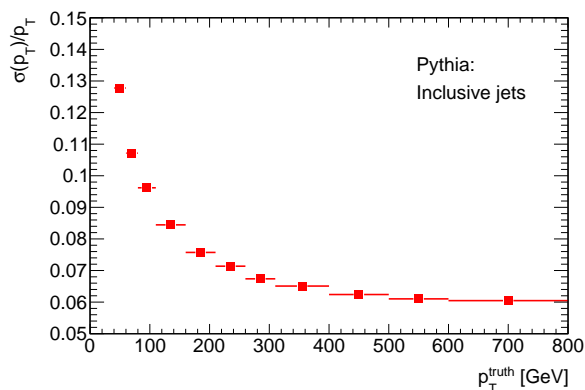
Figure 8.22: *Jet momentum resolution as a function of p_T^{truth} derived from simulation. The jets entering in this calculation fulfil the same requirements as for the jet response calculation. The error bars reflect the uncertainties on the fit parameter.*

95% ($=2\sigma$) of the reconstructed events. Details on this binning are given in Appendix A. With the help of the resolution of b-jets shown in Fig. 8.23(a) the uncertainty on the dijet cross section will be evaluated using pseudo-experiments. The resolution of b-jets is similar to the resolution of inclusive jets in the low p_T region. At around 300 GeV the resolution at calorimeter level of b-jets is better than for inclusive jets. At about 800 GeV the resolution for b-jets is about 15% better. Due to the dominating presence of gluon-initiated jets, the resolution of inclusive jets is worse than that of jets from heavy-quarks. Gluon-initiated jets tend to produce more particles (The colour charges predict that the probability of gluon emission of a gluon relative to that of a gluon emission off a quark is $= \frac{9}{4}$.) which are also softer compared to particles in quark-initiated jets. This already leads to an effect in the resolution due to the statistical nature of a shower development. Moreover, slower particles tend to bend more in a magnetic field. Therefore gluon-initiated jets are wider and may be subject to higher losses in the jet reconstruction because particles are bent outside the area of the jet algorithm.

The inclusion of non-interacting particles into the resolution determination has the effect of degrading the jet p_T resolution as shown in Fig. 8.23(b). The distribution of the relative difference between reconstructed jet energy E_{reco} and simulated true energy E_{truth} has approximately a Gaussian shape. The inclusion of non-interacting particles enhances the negative tail and makes the distributions more asymmetric. A Gaussian core is still retained. In the worst case the resolution is degraded by about 25% in the high p_T region.

8.3.3 Energy scale uncertainties of jets with heavy-flavour content

For this cross section measurement of b-jets with the help of the SV0 tagging algorithm, an additional jet energy scale uncertainty is derived with the help of tracking information in data. The tracks are calibrated independently from the calorimeter and serve as reference objects. This uncertainty is strictly speaking only valid for SV0 b-tagged jets. The total jet energy scale uncertainty is given by:

$$\text{Total JES uncertainty} = \text{inclusive JES uncertainty} \oplus b\text{-jet JES uncertainty}. \qquad (8.12)$$
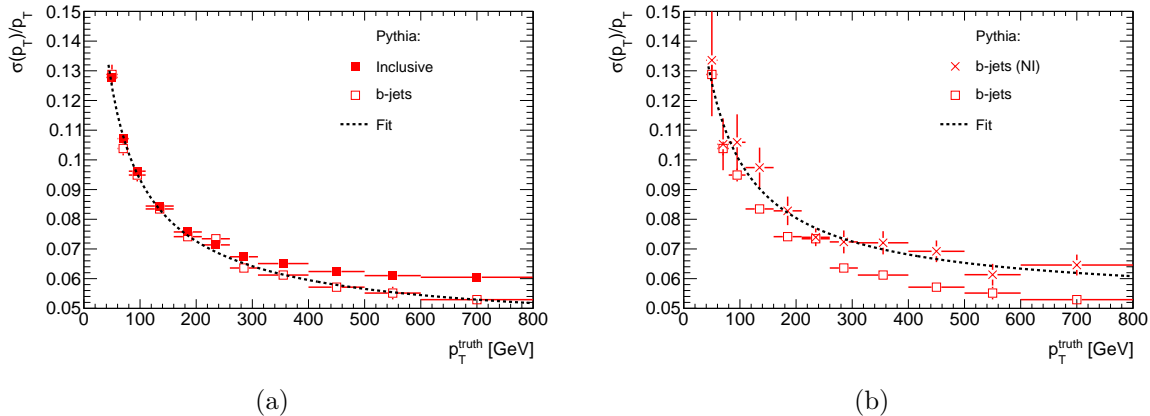
(a) (b)

Figure 8.23: *Jet p_T resolution as a function of p_T^{truth} derived from simulation. The jets entering in this calculation fulfil the same requirements as for the jet response calculation. (a) Inclusive and b-jets are compared at calorimeter level, (b) b-jets are compared at calorimeter level and at particle level including non-interacting particles (NI). The error bars reflect the uncertainties on the fit parameter.*

All b-jets are treated equally, irrespective of their decay mode and calibrated with the EM+JES scheme. The missing energy carried away by neutrinos and muons is not corrected for directly in the energy measurement. But this correction is performed in a single step where the energy of the calorimeter jet is corrected with the help of the energy of the matched particle jet in simulation including the energy of the neutrino and the muon. In the course of these studies, the jet response of b-jets and b-tagged jets is compared to understand the impact of the EM+JES calibration. This is followed by the studies aiming to obtain an additional uncertainty on the jet flavour for this analysis.

8.3.3.1 Jet energy response of jets with heavy-flavour content

The jet p_T response (cf. Eq. 5.2) is calculated as a function of the transverse momentum of the particle jet for each subset of jet flavours. The impact of the EM+JES calibration is studied because the calibration constants are driven by the flavour mixture of the QCD sample. To obtain the jet p_T response, reconstructed jets are matched to truth jets requiring $\Delta R < 0.4$. If a reconstructed or truth jet fails an isolation criteria, namely that another jet with $p_T^{\text{EM}} > 7$ GeV is found within $\Delta R < 1$, this jet pair is discarded. The response in a given p_T^{truth} bin is obtained by fitting a Gaussian distribution and identifying the response as the Gaussian mean from the fit. The behaviour of the jet p_T response of b-jets compared to the inclusive response is presented in Fig. 8.24. The errors on the obtained response values are given by the uncertainty on the Gaussian mean from the fit result and reflect the available statistics in each jet flavour sample. The response presented in Fig. 8.24(a) is calculated from particle jets which represent the calorimeter level (cf. Section 5.1). The response is calculated from the default PYTHIA samples for truth jets with $p_T^{\text{truth}} > 20$ GeV and $|y| < 2.1$.

The response of the inclusive QCD jet sample shows deviations from unity of less than 1% and gives rise to the non-closure uncertainty contribution presented in Section 8.3.1. The response of the light-jet sample in which the light-quark and gluon-initiated jets are joined is in almost perfect agreement with the inclusive sample as expected. These two

samples have the largest overlap and the contribution of the other heavier quark initiated jets in the inclusive sample is at the per cent level.

The response of the *c*- and *b*-jets can be understood when separating the inclusive sample into gluon- and quark-initiated jets. The fragmentation properties of gluon jets cause that the correction to restore gluon jets to hadronic scale has to be much larger than for quark jets. Gluon jets show a larger track multiplicity and a broader jet shape compared to quark jets. These properties can also be used to distinguish between quark- and gluon-initiated jets. Therefore the energy of the light-quark jets in the inclusive sample is as overcompensated as the one from the *c*- and *b*-jets (cf. Fig. D.2 and [110]). The *b*-jet response rises to 2% above the inclusive response and decreases with increasing transverse momentum of the particle jet. This behaviour is driven by the EM+JES correction factor shown in Fig. D.1. At about 400 GeV the response of quark initiated jets dominate the correction factors in the EM+JES calibration and the *b*-jet response reaches the inclusive response and ranges between 0.99 and 1. The same holds for the response of *c*-jets. In case of particle jets which also include muons and neutrinos (cf. Section 5.1) as in Fig. 8.24(b) the response of jets initiated from quarks with an appreciable branching ratio into muons and neutrinos (*b*- and *c*-quarks) is reduced to approximately 3% for *b*-jets. This is because the particle jet's momentum in the denominator of the response is enlarged with respect to the detectable energy in the calorimeters. The response of the inclusive and light-jet sample change within a few per mill with respect to the particle jets excluding muons and neutrinos.

In the following, the impact of the *b*-tagging requirement on the *b*-jets is studied.

(a) Calorimeter level   (b) Particle level (NI)
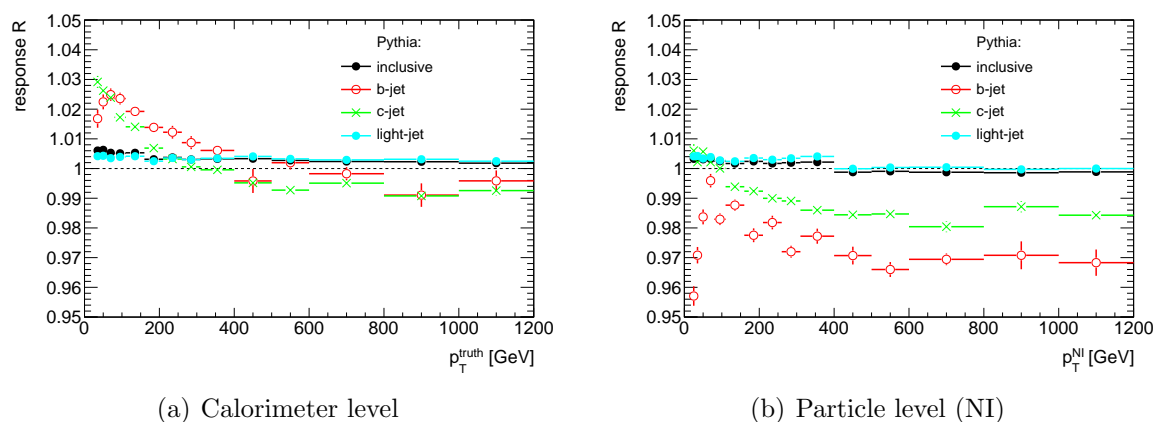
Figure 8.24: *Average simulated jet response (of jets calibrated with* EM+JES *scheme) for different jet flavours as a function of* $p_T^{\text{truth}} > 20$ *GeV ((a)) and* $p_T^{\text{NI}} > 20$ *GeV ((b)) for all jets within* $|y| < 2.1$. *The transverse momentum* p_T^{NI} *in simulation is calculated using in addition the momenta from the neutrinos and muons. The error bars represent the error obtained by the Gaussian fit and approximately reflects the statistics of each jet flavour.*

The selection of *b*-jets using charged tracks might bias the jet energy scale due to the selection of a well measurable subset of *b*-jets with a specific fragmentation. The first separation into different sub-samples is done using the SV0 tagging algorithm in Fig. 8.25(a). These sub-samples contain different mixtures of jet flavours. The response of the not-tagged and inclusive sample is as expected identical. The sample of *b*-tagged jets is approximately composed of about 50% *b*-jets. The composition varies between different p_T

(a)   (b)

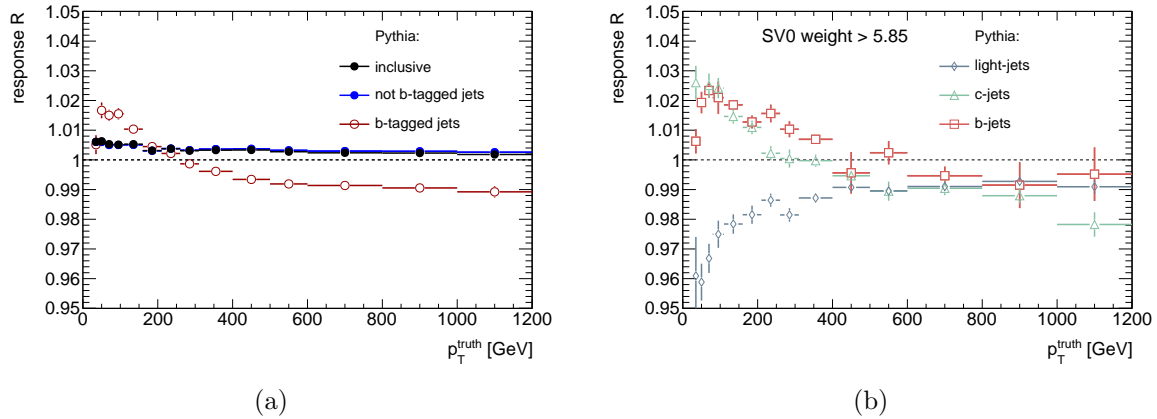**Figure 8.25:** *Average simulated jet response (of jets calibrated with* EM+JES *scheme) for different jet samples selected with the b-tagging selection cut as a function of $p_{\mathrm{T}}^{\mathrm{truth}} > 20$ GeV for all jets within $|y| < 2.1$. (a) Jets are separated by the b-tagging selection cut. (b) b-tagged jets are in addition separated into their jet flavour. The error bars represent the error obtained by the Gaussian fit and approximately reflects the statistics of each jet flavour.*

bins. Nevertheless, the *b*-tagged sample exhibits roughly the same behaviour as the sample of *b*-jets in Fig. 8.24(a). The rise at the beginning of the p_{T} spectrum is damped and the response reaches more rapidly unity. This is due to the mistagged light jets whose fraction is increasing with p_{T}. In the data-driven methods the jet energy scale of a *b*-tagged sample is studied as well as that of *b*-tagged samples with enriched and depleted *b*-jet fractions and their agreement is discussed. The response of these *b*-tagged jets is again split into their jet flavours in Fig. 8.25(b). The response of mistagged light jets starts at 96% and rises up to 99% in the high p_{T} region. This behaviour indicates that the tagging algorithm selects among the light-jets a significant subset of jets which show a reduced response in the EM+JES calibration. This can be explained by the fact that mistagged light-jets with a significant signed decay length significance are more likely to be faked by light-jets with an increased track multiplicity. This implies that a larger part of the jet's energy is contained in charged particles and therefore hadrons, whose reconstructed energy is intrinsically lower and rises slowly with increasing $p_{\mathrm{T}}^{\mathrm{truth}}$. But it is very important to note that tagged *b*-jets in Fig. 8.25(b) and *b*-jets in Fig. 8.24(a) show an identical response within the uncertainty of the Gaussian fit as well as the *c*-jets. In conclusion the ability to tag a *b*-jet and reconstruct a secondary vertex from charged tracks does not alter the energy scale. This is the foundation to study the properties of *b*-jets with the aid of a sample of *b*-tagged jets.

8.3.3.2 Jet energy scale uncertainty of jets with heavy-flavour content

The aim is to extract the additional jet energy scale uncertainty which accounts for the differences in fragmentation of jets with bottom-flavour content. The data-driven method chosen uses the total momentum of charged particles associated with the jet measured in the inner detector. A track is associated to a jet if $\Delta R(\mathrm{track}, \mathrm{jet}) < 0.4$, where track parameters are evaluated at the distance of closest approach to the primary vertex. It is examined how well Monte-Carlo describes *b*-tagged jets and their associated tracks in the data. The tracking system is independently calibrated with respect to the calorimeters and can be

used as an independent estimate of the jet energy scale. The transverse momentum of a jet is compared to the total transverse momentum measured in tracks associated to the jet:

$$r_{\text{trk}} \equiv \frac{|\sum \vec{p}_{\text{T}}^{\text{track}}|}{p_{\text{T}}}. \tag{8.13}$$

This ratio determines the calorimeter jet energy scale for jets contained in the tracking coverage. Depending on the jet selection, it is be separated between r_{trk} with inclusive jets and b-tagged jets. This measurement uses jets with 20 GeV $< p_{\text{T}} <$ 600 GeV. The L1_MBTS trigger provides the low p_{T} jet measurement between 20 and 40 GeV, referred to as minimum bias data. The remaining jets are selected using the single jet triggers as explained in Section 7.3, referred to as jet trigger data. Tracks are selected with the following criteria, where d_0 and z_0 are calculated with respect to the primary vertex:

$$p_{\text{T}}^{\text{track}} > 1 \text{ GeV}, \ N_{\text{pixel}} \geqslant 1, \ N_{\text{SCT}} \geqslant 6,$$
$$|d_0^{\text{PV}}| < 1.5 \text{ mm}, \ |z_0^{\text{PV}} \sin\theta| < 1.5 \text{ mm}. \tag{8.14}$$

Each considered jet contains at least one single quality track. These selection cuts are in good agreement with the b-tagging quality cuts. To ensure that the track matching to the jet is not ambiguous, jets are required to be separated by a distance $\Delta R > 0.8$, if they are not, only the harder jet is retained. The harder jet is well-defined as it was shown in Section 5.1.1.2 and circular whereas the softer one is deformed. This analysis is split up in two y-bins: $|y| < 1.2$ and $1.2 < |y| < 2.1$. To quantify the agreement between data and simulation further, the track multiplicity N_{trk} and the distribution of the transverse momentum measured in charged particles $d_{\text{trk,Jet}}$ as a function of the distance r from the centre of the jet are studied in Fig. 8.26 and Fig. 8.27. The $d_{\text{trk,Jet}}$ distribution is calculated from the sum of the transverse momenta of all tracks which lie within the annulus between r and $r + \Delta r$:

$$d_{\text{trk,Jet}}(r) = \frac{1}{N_{\text{trk}} \sum_{0 \leqslant r' < 0.4} p_{\text{T}}^{\text{track}}} \frac{d \sum_{r \leqslant r' < r+\Delta r} p_{\text{T}}^{\text{track}}}{\Delta r}. \tag{8.15}$$

The track multiplicity and the $d_{\text{trk,Jet}}$ distributions are presented for b-tagged jets between 40 GeV $< p_{\text{T}} <$ 60 GeV. The conclusions made from these plots also hold for the inclusive jet sample and the other rapidity region. The reference event selection in this jet p_{T} region uses the jet trigger but if statistics allow the corresponding measurement is also presented using data selected by the L1_MBTS trigger. The agreement of both trigger selections indicate that there is no bias associated with the trigger used. The distributions are normalised to the same area. The track multiplicity presented in Fig. 8.26(a) is reasonably well described. On average a b-tagged jet between 40 GeV $< p_{\text{T}} <$ 60 GeV contains 6 quality tracks. The distribution of the number of tracks gets broader with increasing jet p_{T} and the mean track multiplicity increases. At around 400 to 600 GeV on average a jet contains about 16 tracks but there are also jets with 40 tracks measured. The mean track multiplicity in data and simulation as a function of jet p_{T} agrees within about 5% as shown in Fig. 8.26(b). The level of agreement is independent from the transverse momenta of the jet. The distribution of the transverse momenta as a function of the distance from the centre of the jet is presented for b-tagged jets with 40 GeV $< p_{\text{T}} <$ 60 GeV in Fig. 8.27(a). Within $r < 0.1$ from the jet axis the largest transverse momentum fraction is found. The fraction decreases with increasing distance from the jet axis. The simulation

reproduces the transverse momentum flow. The average distance of the highest transverse momentum fraction as a function of jet p_T is presented in Fig. 8.27(b). With higher transverse momentum of the jet this distance decreases as expected due to the collimation of the jets. Conservatively, the level of agreement is estimated to be at most 20%. The simulation predicts for $p_T > 100$ GeV a smaller mean distance than observed in data. However, this again indicates that the density of the charged particles and transverse jet shape is well modelled. Furthermore, it can be concluded that the transverse spread of the jets and the loss of tracks at the edge of the tracking coverage is under control.

The estimation of the total jet energy scale uncertainty for b-jets utilises the r_{trk} distribu-

(a)   (b)
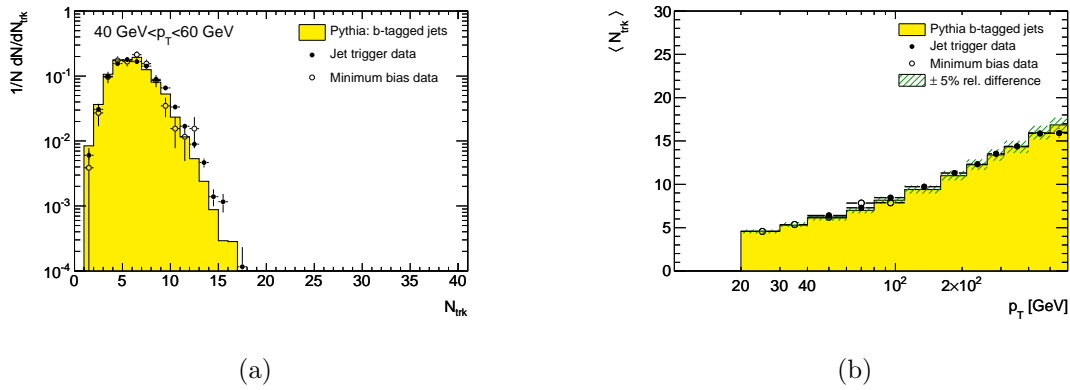
Figure 8.26: *The charged track multiplicity distribution for 40 GeV < p_T < 60 GeV and rapidity range |y| < 1.2 is presented in (a) for b-tagged jets. The mean track multiplicity as a function of jet p_T for b-tagged jets is shown in (b).*

(a)   (b)
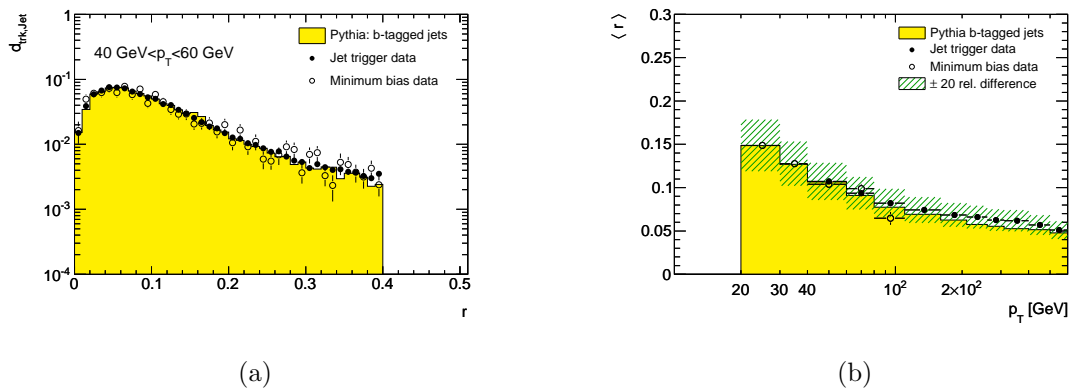
Figure 8.27: *The distribution of the fraction of transverse momentum observed in charged particles as a function of the distance from the centre of the jet for b-tagged jets and (a) 40 GeV < p_T < 60 GeV and rapidity range |y| < 1.2 and in (b) the corresponding mean value of this distribution as a function of jet p_T.*

tions in simulation and data. The r_{trk} distribution for b-tagged jets with 60 GeV < p_T < 80 GeV
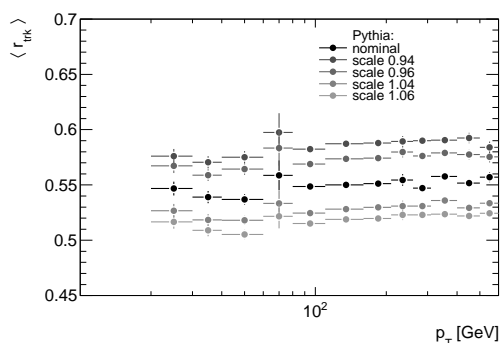
Figure 8.28: *Sensitivity of the mean r_{trk} from simulation to variations of the jet energy scale by ± 4 and 6%.*

for both rapidity regions are presented in Fig. 8.29. The missing r_{trk} distributions for *b*-tagged jets as well as the complete set of distributions of the inclusive jet sample are included in Appendix H. The agreement between data and simulation is excellent. The r_{trk} variable can become larger than 1 in the tails due to problems in the patter recognition in the track finding. From these r_{trk} distributions the mean and the statistical error of the mean value are obtained as a function of jet p_{T}. Typical r_{trk} mean values are around 0.5 to 0.7. The r_{trk} distributions get broader with increasing jet p_{T} and extend to r_{trk} values of almost two for the highest measured kinematic range. To demonstrate the

(a) $|y| < 1.2$

(b) $1.2 < |y| < 2.1$
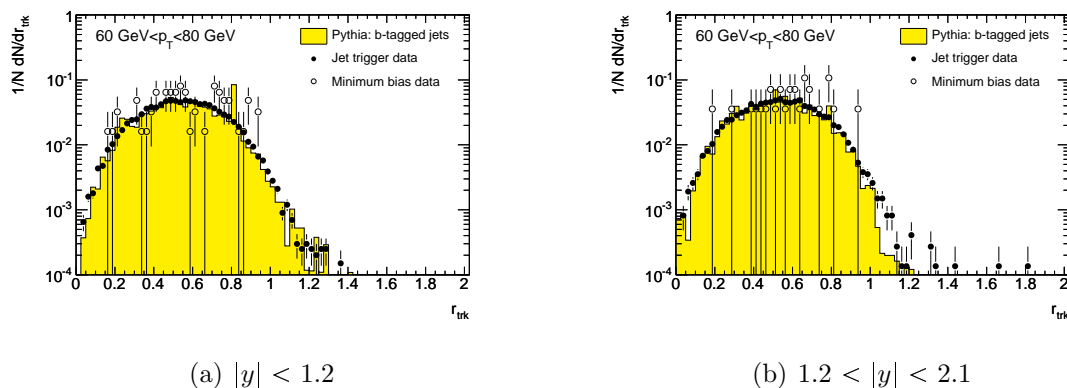
Figure 8.29: *r_{trk} distributions for b-tagged jets with 60 GeV $< p_{\mathrm{T}} <$ 80 GeV for both rapidity regions.*

sensitivity of the total transverse momentum measured in the tracking system to the jet energy scale, variations of $\pm 4\%$ and$\pm 6\%$ to the jet energy scale are applied in simulation and the mean r_{trk} distribution is recalculated. The r_{trk} distributions as a function of jet p_{T} with shifted jet energy scales are presented in Fig. 8.28. A 6% shift in the jet energy scale results in an approximately 9% variation in the mean of the r_{trk} distribution. It can be concluded that the level of agreement between the jet energy scale in data and simulation can be determined with the help of the tracking system. Its sensitivity is reflected in the

mean r_trk distributions in data and simulation. Therefore the correspondence of the jet energy scale is quantified using the double ratios for the corresponding jet selections applied:

$$R_{r_\text{trk},\text{inclusive}} \equiv \frac{\left[\langle r_\text{trk}(\text{incl})\rangle\right]_\text{Data}}{\left[\langle r_\text{trk}(\text{incl})\rangle\right]_\text{MC}}, \tag{8.16}$$

$$R_{r_\text{trk},b\text{-jet}} \equiv \frac{\left[\langle r_\text{trk}(b\text{-jet})\rangle\right]_\text{Data}}{\left[\langle r_\text{trk}(b\text{-jet})\rangle\right]_\text{MC}}. \tag{8.17}$$

An identical jet energy scale in data and simulation results in an $R_{r_\text{trk},\text{X}}$ distribution close to unity. Therefore it is possible to constrain the energy scale and to determine the uncertainty. In addition it is possible to determine the uncertainty on the jet energy scale. The systematic uncertainties on the $R_{r_\text{trk},\text{inclusive}}$ distribution are a direct validation of the jet energy scale uncertainties presented in Section 8.3.1. A detailed study on inclusive jets can be found in [111]. The goal is to derive an additional uncertainty for b-tagged jets taking recourse to the uncertainty on inclusive jets. This can be achieved with the aid of the following distribution:

$$R' \equiv \frac{R_{r_\text{trk},b\text{-jet}}}{R_{r_\text{trk},\text{ inclusive}}}. \tag{8.18}$$

Remaining systematics on R' estimate the uncertainty contribution arising from the heavy-flavour fragmentation. Together with the well founded inclusive jet energy scale uncertainty in Section 8.3.1, the total jet energy scale uncertainty needed for the cross section measurement can be estimated then with the help of Eq. 8.12 by identifying the b-jet JES uncertainty with the systematic uncertainty derived on R'.

The systematic uncertainties on $R_{r_\text{trk},\text{inclusive}}$ and $R_{r_\text{trk},b\text{-jet}}$ arise either from tracking related uncertainties, jet related uncertainties and from the simulation modelling. In addition the b-tagging calibration and uncertainty needs to be evaluated for $R_{r_\text{trk},b\text{-jet}}$. All contributions considered are assumed to be uncorrelated. The systematic uncertainties are described in the following:

1. Simulation modelling: The dependency of the r_trk distribution on the simulation is examined with the help of the HERWIG++ generator which incorporates a different shower and especially a different fragmentation model (cf. Section 2.3). The variation observed in the r_trk distribution between PYTHIA and HERWIG++ samples are symmetrised and taken as a systematic uncertainty. Due to the insufficient statistics of the HERWIG++ samples concerning the b-tagging selection, the variations on $\left[r_\text{trk}(b\text{-jet})\right]_\text{MC}$ are fitted with a constant function.

2. b-tagging calibration: For the $R_{r_\text{trk},b\text{-jet}}$ measurement the b-tagging scale factors for the efficiency and mistag rate are varied correlated within their systematic uncertainty in simulation. The r_trk distribution is re-evaluated and the resulting shifts are added in quadrature to the systematic uncertainty for $R_{r_\text{trk},b\text{-jet}}$.

3. Material description: The knowledge of the tracking efficiency modelling in simulation was evaluated in detail [91], shortly presented in Section 5.2.3 and summarised in Table 5.1. The systematic uncertainty on the tracking efficiency for isolated tracks with $p_\text{T}^\text{track} > 500$ MeV increases from 2% for central tracks to 4% for tracks in the end-cap region. The resulting effect on r_trk is rapidity-dependent and leads to a 2% uncertainty for $|y| < 1.2$ and a 3.1% uncertainty for $1.2 < |y| < 2.1$.

4. Tracking in jet core: High track densities in the jet core influence the tracking efficiency due to shared hits, fake tracks and lost tracks. From previous studies in Section 5.2.3 it was concluded that the effects of shared hits and the number of fake tracks is well described in simulation whereas the loss of tracks is not well understood. As already described in Eq. 5.7 reconstructed tracks are associated to a truth track and vice versa. A relative systematic uncertainty of 50% conservatively estimates the effect of the loss of efficiency using pseudo-experiments. The average inefficiency in each jet p_{T} bin is estimated from simulation by counting the reconstructed and matched tracks associated to a jet with $p_{\mathrm{T}} > 1$ GeV and the corresponding generated tracks with $p_{\mathrm{T}} > 1$ GeV. For each jet the variation in terms of percentages in the ratio $\frac{\sum p_{\mathrm{T,reco}}^{\mathrm{track}}}{\sum p_{\mathrm{T,gen}}^{\mathrm{matched}}}$ is determined by a random generator deciding according to the relative uncertainty of 50% if a truth track should be added or discarded from the denominator. The relative shift in the ratio r_{trk} is added in quadrature to the systematic uncertainty.

5. Jet resolution: The jet energy resolution in simulation is about 10% better than in data (cf. Fig. 8.21). The resolution in simulation is degraded according to this maximum deviation for all p_{T} and rapidity bins. The effect of a resolution improvement is estimated by the variation in case of the degradation. In pseudo-experiments a random energy that corresponds to a resolution smearing of 10% is added (or subtracted) to each jet. The resulting average shift on the r_{trk} distribution is evaluated and added in quadrature to the systematic uncertainty.

The total systematic uncertainties for inclusive and b-tagged jets are presented for both rapidity regions in Figs. 8.30 and 8.31. The systematic uncertainty for inclusive jets is between 2% and 6%. With increasing p_{T} the uncertainty decreases and is maximal in the lowest p_{T} bin. The systematic uncertainty arising from the tracking in the jet core is negligible for the lower rapidity bin and rises to about 1% for the rapidity region between $1.2 < |y| < 2.1$. The uncertainty based on the material description is constant as a function of jet p_{T} and is the largest contribution at $p_{\mathrm{T}} > 50$ GeV. The behaviour of the uncertainty based on the jet resolution smearing and the generator modelling is similar. Both contributions decrease to less than 1% with increasing p_{T} and make up the largest part of the uncertainty in the lower kinematic regions. In the course of the studies detailed in [111] the uncertainty on r_{trk} from pile-up is negligible and to a large extent this uncertainty will cancel out for the measurement of R'. Most of the description of the systematic uncertainties on the $R_{r_{\mathrm{trk}},inclusive}$ can be transferred to the fractional systematic uncertainties of $R_{r_{\mathrm{trk}},b\text{-}jet}$ in Fig. 8.31. The systematic uncertainty is between 3% and 5%. In addition the uncertainty from the b-tagging calibration adds about 2% at most to the uncertainty. Finally, Figs. 8.32 and 8.33 present the measurement of $R_{r_{\mathrm{trk}},inclusive}$ and $R_{r_{\mathrm{trk}},b\text{-}jet}$. Both measurements show that the jet energy scale in data and simulation agree within systematic uncertainties. The agreement is of the level of 2% for inclusive jets and b-tagged jets with $|y| < 1.2$ and 4% for b-tagged jets in the outermost rapidity region. The dotted lines in both figures indicate this agreement. With the help of the tracking information the jet energy scale predicted by the simulation is validated and an independent scale uncertainty is obtained.

Two further cross checks provide confidence in the estimation of the jet energy scale uncertainty of b-tagged jets. The estimation of this additional uncertainty is based on the fact that b-jets which are capable of being b-tagged do not show a systematic difference
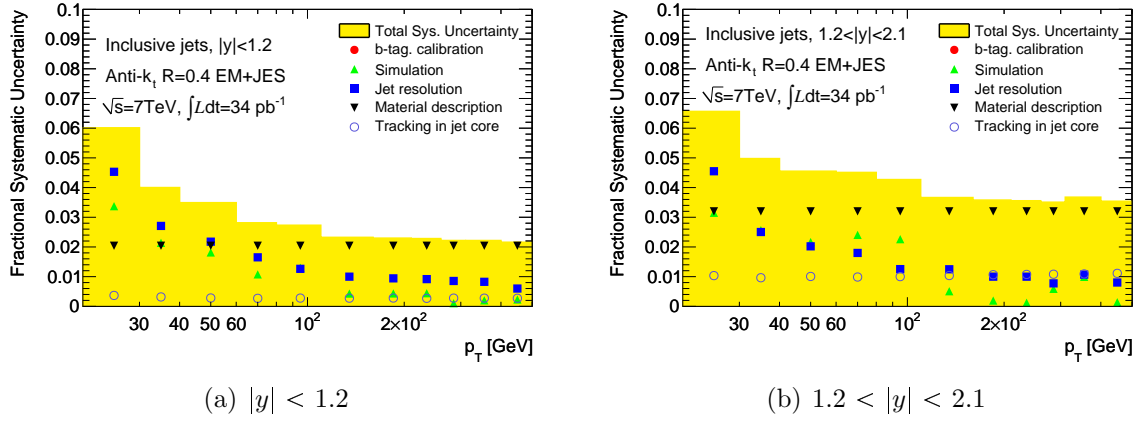
(a) $|y| < 1.2$

(b) $1.2 < |y| < 2.1$

Figure 8.30: *The fractional systematic uncertainties on $R_{r_{\mathrm{trk}},inclusive}$ as a function of jet p_{T} for (a) $|y| < 1.2$ and (b) $1.2 < |y| < 2.1$.*
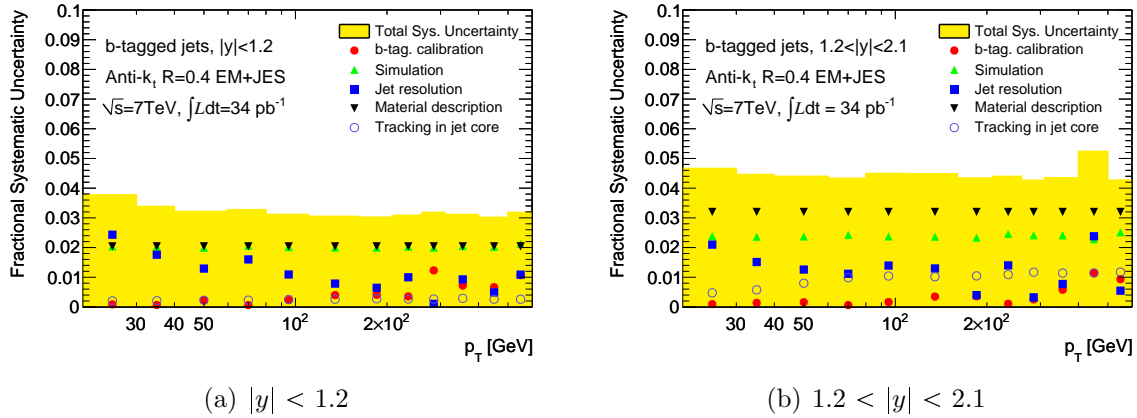
(a) $|y| < 1.2$

(b) $1.2 < |y| < 2.1$

Figure 8.31: *The fractional systematic uncertainties on $R_{r_{\mathrm{trk}},b\text{-}jet}$ as a function of jet p_{T} for (a) $|y| < 1.2$ and (b) $1.2 < |y| < 2.1$.*

in their energy response. To study this, the fraction of *b*-jets among the *b*-tagged jets is changed and the agreement between data and simulation is checked. The following sub-samples are used for reasons presented in Section 8.2.2: SV0 mass $\geqslant 1.5$ GeV and SV0 mass < 1.5 GeV. The former sample has an enhanced fraction of *b*-jets and the latter is depleted. The ratio of the mean value of r_{trk} in data and simulation is presented for both sub-samples in Fig. 8.34. It can be seen that both selections either enhancing or depleting the fraction of *b*-jets in the sample are in agreement with unity within the systematic uncertainties. The level of agreement is identical to the one found in the full sample. This affirms that the relation found in simulation between Fig. 8.24(a) and Fig. 8.25(b) is valid. Furthermore, as muons from a semi-leptonic decay of a bottom-flavoured hadron are reconstructed as charged tracks it is possible to use an additional selection on soft muons in the sample of *b*-tagged jets to ascertain the relation between the jet energy response of *b*-jets on calorimeter level and particle level. The transverse momentum of the soft muon was already studied in Section 8.1.3 and was found to be in agreement between data and simulation. Relating the mean r_{trk} values between data and simulation for *b*-tagged

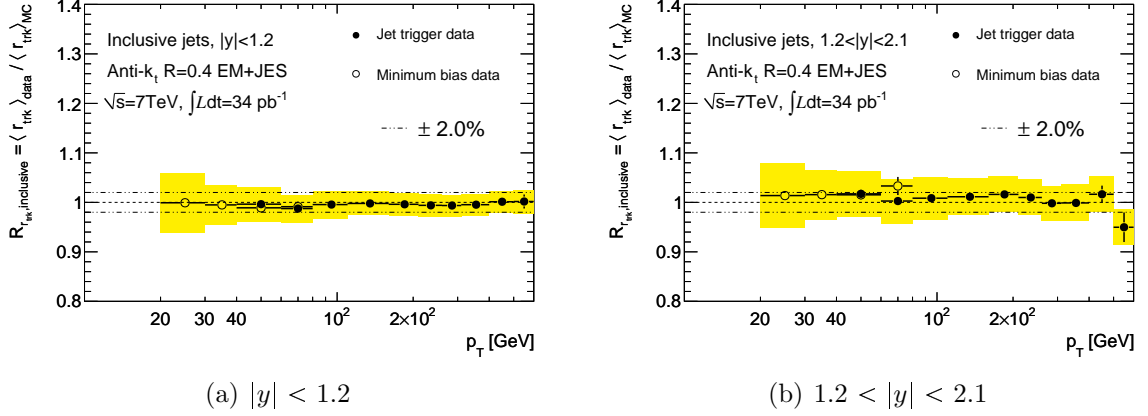**Figure 8.32:** *The ratio of the mean value of r_{trk} in data and simulation for inclusive jets as a function p_{T} for (a) $|y| < 1.2$ and (b) $1.2 < |y| < 2.1$.*
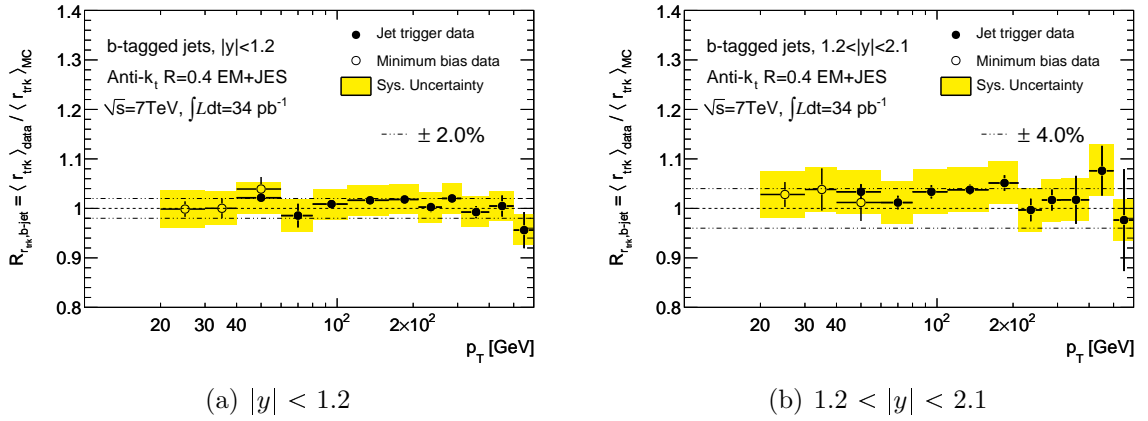
Figure 8.33: *The ratio of the mean value of r_{trk} in data and simulation for b-tagged jets as a function p_{T} for (a) $|y| < 1.2$ and (b) $1.2 < |y| < 2.1$.*

with a soft muon the energy scale can be tested further for the impact of the neutrino. In Fig. 8.35(a) the impact of the muon can be seen in the change of the slope of the mean value distribution. The fraction of energy carried by charged tracks is increased due to the presence of the muon. The mean r_{trk} value at small p_{T} is increased by a factor of approximately 50%. At 500 GeV the increase shrinks to about 9%. Within the statistical errors and a 2% relative difference data and simulation mean values agree with each other. In addition to rule out any opposing effect on the full r_{trk} measurement, the mean values of r_{trk} as a function of p_{T} for the selection with a veto on a soft muon is shown in Fig. 8.35(b). The mean values are similar to the ones from the full measurement as the events with a soft muon are rather rare as the statistical errors have indicated in Fig. 8.35(a). The level of agreement is also of the order of 2% or better.

For the calculation of the systematic uncertainties on R' the following arguments are considered:

1. Simulation modelling: The deviation between R' calculated from data and PYTHIA and $R'_{\mathrm{HERWIG++}}$ calculated from data and HERWIG++ simulation are the basis of
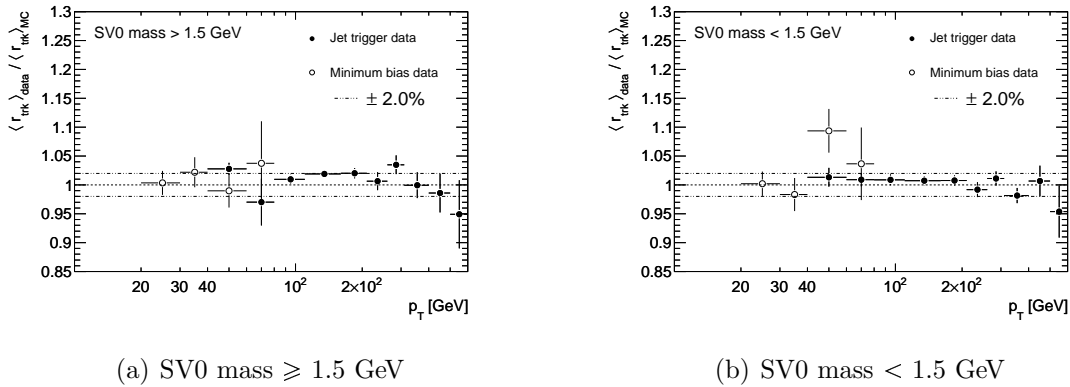
(a) SV0 mass $\geqslant$ 1.5 GeV

(b) SV0 mass < 1.5 GeV

Figure 8.34: *The ratio of the mean value of r_{trk} in data and simulation for b-tagged jets with $|y| < 1.2$ separated into two sub-samples: (a) SV0 mass $\geqslant$ 1.5 GeV and (b) SV0 mass < 1.5 GeV.*

the systematic uncertainty. The remaining deviations reflect the uncertainty due to the different fragmentation as the contributions from the underlying event and the parton shower cancel out. Due to the limited statistics in case of the b-tagging selection the deviations are fitted with step functions with disjunctive intervals. The intervals are chosen to describe the behaviour of the deviations. The first interval is between 20 GeV and 40 GeV where the difference to the HERWIG++ prediction is significant. The second interval is between 40 GeV and 600 GeV and is the important kinematic range for the cross section measurement.

2. *b*-tagging calibration: The b-tagging calibration uncertainty is propagated to R'.

3. Material description: This uncertainty on isolated tracks is coming from an external measurements and the uncertainties on $R_{r_{\mathrm{trk}},inclusive}$ and $R_{r_{\mathrm{trk}},b\text{-}jet}$ are fully correlated and cancel.

4. Tracking in jet core: It is assumed that both uncertainties are fully correlated and to first order cancel. This assumption is exactly valid for high p_{T} jets; for low p_{T} jets the second order deviations are estimated to be about 0.1%.

5. Jet resolution: It is assumed that both uncertainties are fully correlated and to first order cancel. This assumption is exactly valid for high p_{T} jets; for low p_{T} jets the second order deviations are estimated to be about 0.2%.

The significant systematic uncertainties on R' are due to the modelling in simulation and the b-tagging calibration. Both are added in quadrature. The results on R' and the fractional systematic uncertainties are presented in Fig. 8.36 and Fig. 8.37. The jet energy scale of b-tagged jets relative to the scale of inclusive jets is in agreement to the level of 2% and 2.5% for jets with $|y| < 1.2$ and $1.2 < |y| < 2.1$ respectively. Within the statistical and systematic uncertainties the energy scale of b-tagged jets with respect to inclusive jets is in agreement with unity. The uncertainty from the simulation modelling is almost dominating in each p_{T} and rapidity region. The systematic uncertainty is between 2% and 4%. The focus for this analysis is on the interval between 40 GeV and 600 GeV. An additional uncertainty for b-tagged jets is estimated with the help of charged tracks relating the total

(a) Jets with an soft muon

(b) Jets without a soft muon
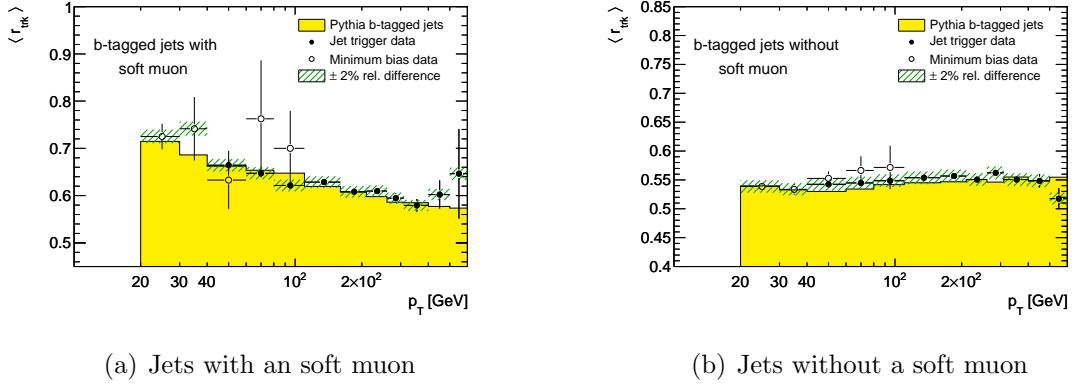
Figure 8.35: *The mean value of r_{trk} in data and simulation for b-tagged jets with $|y| < 1.2$ separated into two sub-samples: (a) b-tagged jets with an soft muon and (b) b-tagged jets without a soft muon.*

transverse momentum in charged particles to the transverse momentum of a jet measured in the calorimeter. The uncertainty on the fitting parameters for estimating the simulation modelling uncertainty show a sensitive to the fit result of $\pm\ 0.2\%$. To be conservative and cover fitting systematics, an overall constant jet energy scale uncertainty contribution for b-tagged jets is set to 2.5%. Furthermore, the jet energy scale is proven to be in agreement between data and simulation. It is expected that differences from fragmentation become less important at high jet p_{T}, therefore it is justified to safely extrapolate the additional 2.5% jet energy scale uncertainty to higher p_{T} regions for this analysis. In Table 8.3 a summary of representative values of the systematic uncertainty on the jet energy scale uncertainty for b-jets is given deduced from the systematic uncertainty on the jet energy scale uncertainty for inclusive jets and the discussed R' measurement.
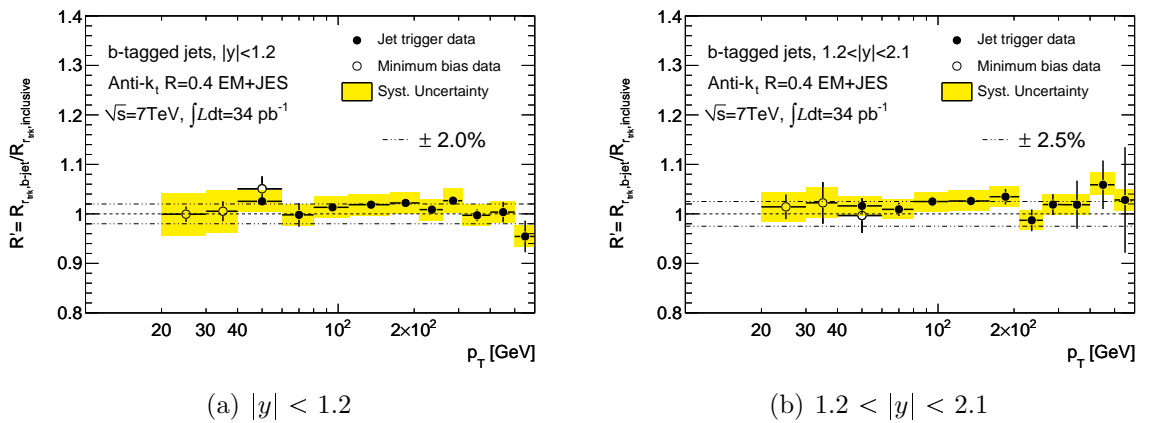
(a) $|y| < 1.2$

(b) $1.2 < |y| < 2.1$

Figure 8.36: *The ratio R' (Eq. 8.18) calculated from inclusive and b-tagged for (a) $|y| < 1.2$ and (b)$1.2 < |y| < 2.1$. The dashed lines indicate the estimated agreement between data and simulation.*

p_T [GeV]	JES Uncertainty [%]			
	$0 < |\eta| < 0.3$	$0.3 < |\eta| < 0.8$	$0.8 < |\eta| < 1.2$	$1.2 < |\eta| < 2.1$
40	4.2 %	4.0 %	4.2 %	4.9 %
60	3.5 %	3.4 %	3.7 %	3.7%
80	3.3 %	3.5 %	3.6 %	3.6 %
110	3.2 %	3.5 %	3.5 %	3.6 %
160	3.4 %	3.4 %	3.6 %	3.6 %
210	3.3 %	3.5 %	3.6 %	3.7 %
260	3.3 %	3.5 %	3.6 %	3.7%
310	3.3 %	3.4 %	3.5 %	3.6 %
400	3.3 %	3.5 %	3.6 %	3.7 %
500	3.5 %	3.5 %	3.6 %	3.7 %
600	3.5 %	3.6 %	3.7 %	3.8 %
800	3.6 %	3.8 %	3.9 %	4.0 %
1000	3.8 %	3.9 %	4.0 %	4.2 %

Table 8.3: *Summary of the* EM+JES *jet energy scale systematic uncertainties for b-jets for different p_T and η regions for anti-k_t jets with $R = 0.4$.*

(a) $|y| < 1.2$ (b) $1.2 < |y| < 2.1$

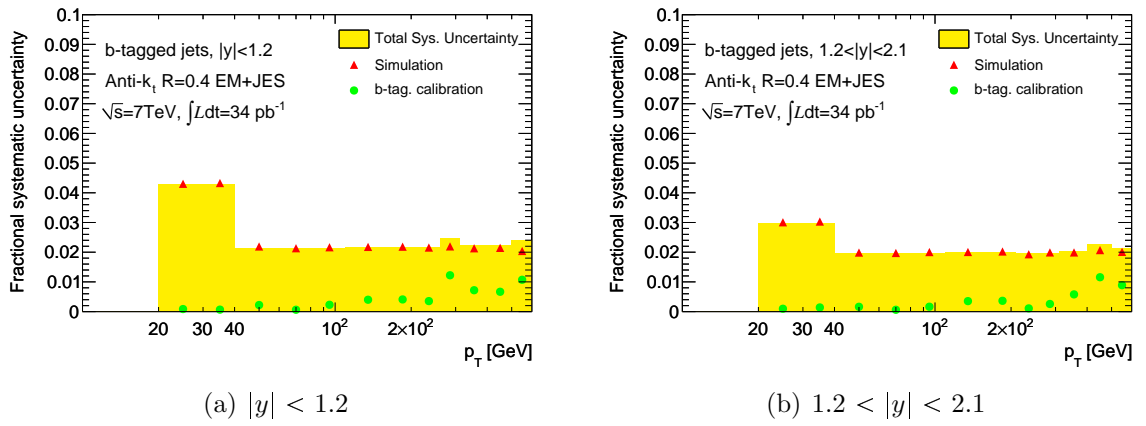**Figure 8.37:** *The fractional systematic uncertainties for R′ calculated from inclusive and b-tagged jets for (a) $|y| < 1.2$ and (b)$1.2 < |y| < 2.1$.*

9 The cross section measurement

This chapter explains the dijet cross section measurement and describes the methods used to obtain it. The previous chapters have introduced the physical objects and the approaches to properly define and reconstruct them (cf. Section 5). The event and object selection suitable for the cross section measurement has been discussed in Section 7. The adequacy of the simulation has been thoroughly studied and the most important systematic uncertainties arising from it have been identified and quantified in the previous chapter. In the next section the definitions of the cross sections and the terminology used in the context of the measurement are given. In Section 9.2 the corrections and corresponding systematic uncertainties for all experimental effects are presented. Finally, the dijet cross sections which are at this point corrected to particle level are presented and the systematic uncertainties summarised. The final cross sections are directly comparable to NLO predictions.

9.1 Cross section definition

The event and jet selection as well as the trigger selection was presented in Section 7 and is the basis for the cross section measurement. Jets are reconstructed using the anti-k_t algorithm with resolution parameter $R = 0.4$ and calibrated with the EM+JES scheme. The relatively long lifetime of hadrons containing a b-quark is exploited to obtain a jet sample enriched in b-jets. A jet is identified as a b-jet if the decay length significance (SV0 weight) is above 5.85.

The $b\bar{b}$ dijet cross section is measured for the leading and sub-leading jets in the event as a function of the invariant mass M, the azimuthal angle difference $\Delta\phi$ and the angular variable $\chi = \exp|y_1 - y_2|$ of the dijet system for jets with $p_T > 40$ GeV and $|y| < 2.1$. The cross section as a function of dijet mass is plotted only above the minimum mass where the measurement is no longer biased by the p_T phase space cut on the leading jets. Therefore the measurement begins above a dijet mass of 110 GeV. The spectrum below this mass is very sensitive to the jet energy scale uncertainty and is also omitted for all other cross section measurements. The angular variable χ was motivated in Section 2.1.4. The dijet cross section as a function of χ is measured up to a maximum value of $\chi < 30$ and an additional acceptance cut is set on $\frac{1}{2}|y_1 + y_2| < 1.1$ to reject events where both jets are boosted into the forward or backward direction. This acceptance cut is illustrated in Fig. I.1. The dijet cross section as function of the angular variable χ is measured in two different mass regions: 110 GeV $< M <$ 370 GeV and 370 GeV $< M <$ 850 GeV. The definitions of the cross sections are given by:

$$\frac{d\sigma}{dM} = \frac{1}{\Delta M} \cdot \left. \frac{P_{b\bar{b}} \cdot C_{\text{unf}}}{\epsilon_{\text{sel}} \cdot \epsilon_{\text{clean}} \cdot \epsilon_{b\bar{b}} \cdot \epsilon_{\text{trig}}} \right|_{M} \cdot \frac{N(M)}{L}, \tag{9.1}$$

$$\frac{d\sigma}{d\Delta\phi} = \frac{1}{\Delta\phi} \cdot \left. \frac{P_{b\bar{b}} \cdot C_{\text{unf}}}{\epsilon_{\text{sel}} \cdot \epsilon_{\text{clean}} \cdot \epsilon_{b\bar{b}} \cdot \epsilon_{\text{trig}}} \right|_{\Delta\phi} \cdot \frac{N(\Delta\phi)}{L}, \tag{9.2}$$

$$\frac{d^2\sigma}{d\chi dM} = \frac{1}{\Delta\chi\Delta M} \cdot \left. \frac{P_{b\bar{b}} \cdot C_{\text{unf}}}{\epsilon_{\text{sel}} \cdot \epsilon_{\text{clean}} \cdot \epsilon_{b\bar{b}} \cdot \epsilon_{\text{trig}}} \right|_{\chi, M} \cdot \frac{N(\chi, M)}{L}. \tag{9.3}$$

The trigger efficiency ϵ_{trig} has been studied in Section 7.3 and was proven to be identical to 1.0 for the trigger selection chosen. The luminosity L used in the measurement has been presented in Tables 7.1 and 7.2, and the corresponding systematic uncertainty was discussed in Section 4.2 and amounts to $\pm 3.4\%$. N denotes the number of events measured as a function of dijet mass, azimuthal difference and angular variable χ. The binning of the cross section in these variables is given in Appendix A. The binning in terms of dijet mass is motivated by the dijet mass resolution as discussed in Section 8.3.2. The binning of the angular and azimuthal variables is explained later in this chapter. The event selection efficiency ϵ_{sel} and the jet cleaning efficiency ϵ_{clean} are studied in the following and possible systematic uncertainties arising from them are discussed. The identification of b-jets using the b-tagging algorithm SV0 requires two corrections to the measurement. Firstly, the b-tagging efficiency $\epsilon_{b\bar{b}}$ has to be evaluated using the b-tagging calibration and applied to correct the number of b-tagged jets to the number of b-jets in data. Secondly, the number of b-tagged jets needs to be corrected for the additional mistagged jets in the jet sample with the help of the b-tagging purity $P_{b\bar{b}}$. Both corrections bear systematic uncertainties which need to be evaluated. The correction factors discussed so far take values between zero and one. The final correction factor C_{unf} not bound between zero and one corrects for detector efficiencies and resolutions in a single step. This correction is obtained from a bin-by-bin unfolding method using simulated events from PYTHIA. All these factors are derived separately as a function of dijet mass, $\Delta\phi$ and $\chi(M)$ from simulated events for the different measurements. For readability this is omitted in its usual form in the cross section definitions above, only the vertical lines with the corresponding subscripts are given as indicators.

On the basis of the dijet cross section measurement of b-jets, the ratio to inclusive dijets as a function of dijet mass is calculated. The following cross section definition for inclusive dijet cross section is used:

$$\frac{d\sigma_{\text{incl}}}{dM} = \frac{1}{\Delta M} \cdot \left. \frac{C_{\text{unf}}^{\text{incl}}}{\epsilon_{\text{sel}}^{\text{incl}} \cdot \epsilon_{\text{clean}}^{\text{incl}} \cdot \epsilon_{\text{trig}}^{\text{incl}}} \right|_{M} \cdot \frac{N_{\text{incl}}(M)}{L}. \tag{9.4}$$

The cross section ratio is then defined as:

$$\frac{d\sigma_{b\bar{b}}/dM}{d\sigma_{\text{incl}}/dM} = \left. \frac{P_{b\bar{b}} \cdot C_{\text{unf}}^{b\bar{b}} \cdot \epsilon_{\text{sel}}^{\text{incl}} \cdot \epsilon_{\text{clean}}^{\text{incl}} \cdot \epsilon_{\text{trig}}^{\text{incl}}}{C_{\text{unf}}^{\text{incl}} \cdot \epsilon_{\text{sel}}^{b\bar{b}} \cdot \epsilon_{\text{clean}}^{b\bar{b}} \cdot \epsilon_{b\bar{b}} \cdot \epsilon_{\text{trig}}^{b\bar{b}}} \right|_{M} \frac{N_{b\bar{b}}(M)}{N_{\text{incl}}(M)}, \tag{9.5}$$

where $d\sigma_{b\bar{b}}/dM$ is identical to $d\sigma/dM$ in Eq. 9.1 and the corresponding correction factors are given an additional superscript to track their origin.

9.2 Correction factors and systematic uncertainties

This section presents the corrections applied to the cross section measurement to correct to particle level for comparison to NLO predictions and discusses the systematic uncertainties on the measurement.

9.2.1 Event and jet cleaning selection

The event selection efficiency is determined with the help of the simulated sample of inclusive jets. The inclusive jet sample is obtained using the same selection cuts without requiring the b-tagging cuts. It is a nearly perfect representation of the kinematic properties of a b-tagged jet sample with the important advantage of more statistics. Also later on if applicable the inclusive jet sample is used as representative sample to evaluate the effect of systematic uncertainties. The efficiency is calculated from the number of events as a function of dijet mass without and with requiring at least ten tracks associated to the primary vertex but keeping the remaining selection cuts in place (cf. Section 7.4). The event selection efficiency is the bin-wise ratio between these two numbers of events. The error on the event selection efficiency from simulation is calculated from binomial error calculation. The result of this calculation is presented as (1- ϵ_{sel}). The inefficiency is shown in Fig. 9.1(a) as a function of inclusive dijet mass. The event selection efficiency is above 99.9% for 110 GeV $< M <$ 1000 GeV. The fluctuations in the inefficiency are caused by the statistics in the simulated samples. The efficiencies as functions of $\Delta\phi$ and χ exhibit as expected the same behaviour. A primary vertex is unambiguously defined by a minimum of three tracks. With increasing jet p_{T} the number of tracks associated to a jet is also increasing. Therefore a measurement using jets with p_{T} lower than 40 GeV would have a higher inefficiency due to the minimum number of tracks required. On average one jet between 40 and 60 GeV already contributes 6 tracks. Hence, it does not surprise that the inefficiency is small. The impact on the cross section measurement is negligible and the efficiency inserted into Eq. 9.1 to Eq. 9.3 is $\epsilon_{\mathrm{sel}} = 1$. As the track multiplicity in data is 5% higher than in simulation, the event selection efficiency is determined with a looser (6) and a tighter (12) cut on the number of tracks in the primary vertex. The deviation between both efficiencies is calculated and was found to be smaller than $\Delta\epsilon_{\mathrm{sel}} < 0.0005$. This systematic uncertainty is neglected due to its small impact.

For completeness the jet reconstruction efficiency has been determined in other studies [101] and reveals that jets with at least 30 GeV transverse momentum are reconstructed with approximately 100% efficiency. The jet cleaning efficiency is determined on data using $b\bar{b}$ dijet events. The b-tagging selection cuts remove all jets arising from background sources and calorimeter noise. The $b\bar{b}$ dijet mass distribution is reconstructed with and without the jet cleaning cuts. The resulting efficiency is presented in Fig. 9.1(b). The selected event samples are completely identical to each other and in the $b\bar{b}$ dijet cross section measurement no b-jet is rejected due to the jet cleaning cuts. The resulting efficiency is $\epsilon_{\mathrm{clean}} = 1$. Consequently, the measurements using the other variables $\Delta\phi$ and χ are also not affected. Because the b-jet selection is not affected by the cleaning cuts no systematic uncertainty has to be evaluated.
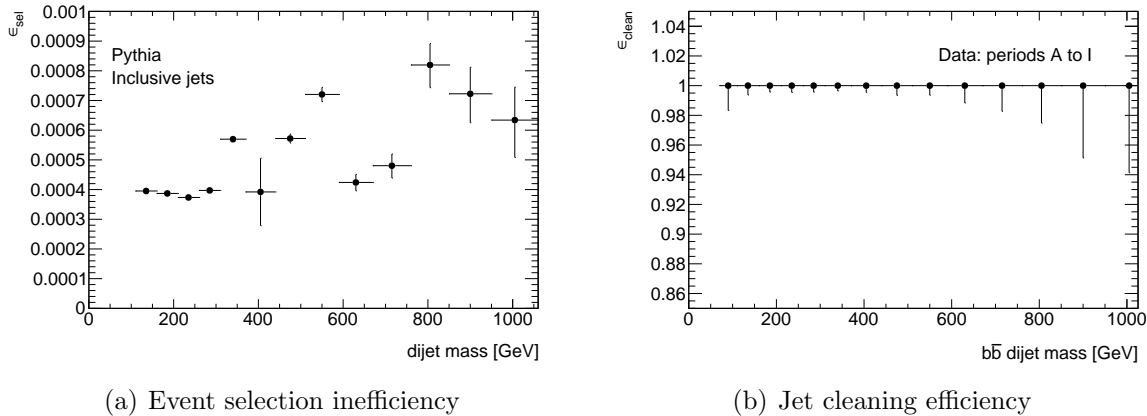
(a) Event selection inefficiency

(b) Jet cleaning efficiency

Figure 9.1: *(a) Event selection inefficiency* $(1 - \epsilon_{\mathrm{sel}})$ *as a function of inclusive dijet mass, (b) Jet cleaning efficiency* $\epsilon_{\mathrm{clean}}$ *as a function of $b\bar{b}$ dijet mass for data periods A to I.*

9.2.2 Trigger selection

Two sources of systematic uncertainties on the trigger efficiency are evaluated. First systematics arising from the choice of the reference sample and of the method are studied. Then effects on the trigger efficiency caused by differences between inclusive jets and b-tagged jets are discussed.

Uncertainties on the measurement: The orthogonal trigger method using L1_MBTS is the baseline for the determination of the trigger efficiency in data periods A to F. Events selected by the minimum bias trigger L1_MBTS are assumed to be unbiased. The trigger requires a single hit in one of the Minimum Bias Trigger Scintillators covering the end-cap calorimeter region at both sides of the detector. This trigger efficiency is compared to the trigger efficiency, which is obtained with respect to the L1_ZDC trigger, as shown in Fig. 9.2(a). This trigger uses the Zero-Degree Calorimeter placed in the very forward region just after the LHC beam pipe splits in two. The comparisons between both trigger efficiencies are shown for the trigger with the lowest threshold and the unprescaled trigger with the lowest threshold. The assumption that the minimum bias trigger is unbiased and orthogonal with respect to jets is proven to be correct for the jets in the plateau regions. The plateaus are reached in the same p_{T} bin and therefore the efficiencies are in agreement to less than the one per cent level because a plateau is defined above the 99% trigger efficiency. In case of the L1_J55 trigger the comparison to the trigger efficiency measured with the bootstrap method with the help of L1_J30 is presented in Fig. 9.2(b). The bootstrap method yields the more precise measurement in terms of statistical error. The trigger efficiencies in the data periods G to I were obtained using the bootstrap method which can determine the efficiency at event level. The L2_j15 trigger is an exception and has to be measured with the tag-and-probe method. The unbiased triggers are highly prescaled in this data period and the available statistics are very low. A systematic check on L2_j15 with the help of the L1_MBTS trigger shows that they are in agreement within their statistical errors. The trigger efficiencies of the triggers with higher thresholds obtained from the bootstrap method are tested with the help of the tag-and-probe method. Both trigger efficiencies are presented in Fig. 9.3 for the triggers L2_j30 and L1_J95. The beginning of the plateaus regions are in agreement of the order of less than 1%. The deviations between both methods arise from the fact that the tag-and-probe method
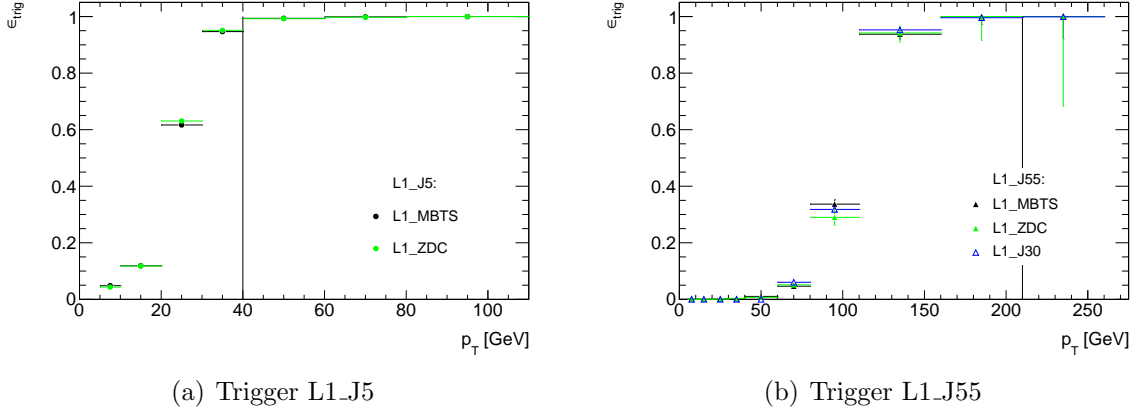
(a) Trigger L1_J5

(b) Trigger L1_J55

Figure 9.2: *(a) Systematic check of the trigger efficiency of L1_J5 calculated from L1_MBTS with the help of the orthogonal trigger L1_ZDC in data periods A to F, (b) Systematic check of the trigger efficiency of L1_J55 calculated from L1_MBTS and L1_ZDC with the help of the bootstrap method using L1_J30 in data periods A to F. The vertical lines indicate the trigger plateaus.*

determines per jet trigger efficiencies. In summary, the trigger efficiencies are tested to be in agreement at the one per cent level or better.

Systematic effects due to the *b*-tagging selection: The subset of *b*-tagged jets may

(a) Trigger L2_j30
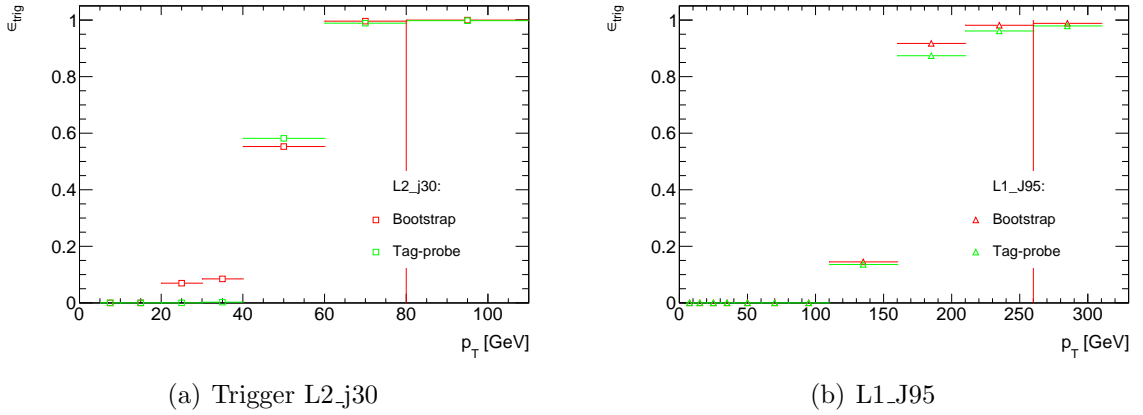
(b) L1_J95

Figure 9.3: *(a) Systematic check of the trigger efficiency of L2_j30, which has been calculated with the bootstrap method, with the help of the tag-and-probe method in data periods G to I, (b) Systematic check of the trigger efficiency of L1_J95, which has been calculated with the bootstrap method, with the help of the tag-and-probe method in data periods G to I. The vertical lines indicate the trigger plateaus.*

have a slightly different trigger efficiency than inclusive jets due to their production mechanism and different fragmentation. Due to the different production mechanism for *b*-jets the rapidity correlation between the leading jets is different with respect to inclusive jets. As the trigger efficiency depends on the jet rapidity as presented in Ref. [112] (Fig. 2), the integrated trigger efficiency for jets within $|y| < 2.1$ may be affected by this. To conservatively estimate the uncertainty, the extreme case is considered in which all

b-tagged jets are assumed to be reconstructed in the region with highest inefficiencies. The maximum difference between the trigger efficiencies for jets $|y| < 1.2$ and $1.2 < |y| < 2.1$ is assigned as a systematic uncertainty and added in quadrature to the previously discussed uncertainty. In periods A to I the difference $\epsilon_{|y|<1.2} - \epsilon_{1.2<|y|<2.1}$ is identical to zero for all triggers except L1_J5 or respectively L2_j15. An asymmetric uncertainty of 1.2% is assigned in the p_T region of 40 to 60 GeV for periods A to F and an uncertainty of 1% in periods G to I. Furthermore, it is checked if the identification of RoIs is biased by the *b*-tagging selection. The trigger efficiency was separately determined from the sample of *b*-tagged jets and compared to the full data set of periods A to F. The necessary information on *b*-tagging for data periods G to I is not accessible in the available trigger sample. The resulting trigger efficiencies for the available data set and the *b*-tagging selection are presented in Fig. 9.4 for the triggers L1_J5 and L1_J55. The effect of a different fragmentation for L1_J5 is visible in the segment of the curve with very high slope. The efficiency measured with the *b*-tagging selection is significantly higher than the efficiency in the inclusive sample. This may come from the harder fragmentation and the requirement of high p_T tracks in the *b*-tagging track selection, so that jets with this signature in the calorimeter may more easily be identified as a local energy maximum. But this effect has almost vanished in case of the L1_J55 trigger. More importantly when both trigger curves approach the plateau region, the difference in the efficiency measurement vanishes and both efficiencies are in agreement within the statistical errors. Because the cross section measurement does not rely on the trigger efficiency below the plateau region, the relevant trigger efficiencies of the full data set and the *b*-tagging selection are found to be identical. The trigger selection is not biased with respect to the data sample used in the analysis. There are no restrictions to translate this directly onto the data periods G to I. This will be proven explicitly in Section 9.2.7.

In summary, an additional uncertainty of 1.2% and 1% for $|y| < 1.2$ and $1.2 < |y| < 2.1$ respectively is assigned to the jet p_T region of 40 to 60 GeV to cover possible biases in the rapidity distribution between inclusive and *b*-tagged jets.

(a) Trigger L1_J5

(b) Trigger L1_J55
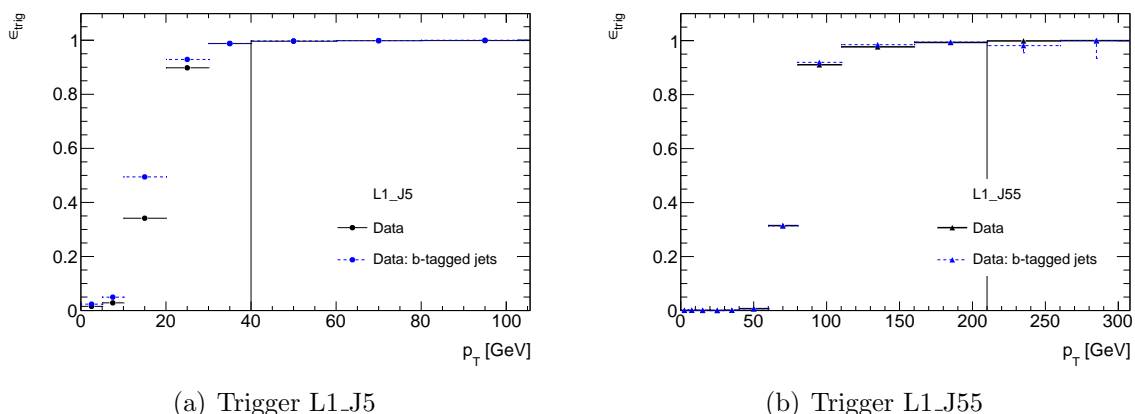
Figure 9.4: *Systematic check of the trigger efficiencies between the inclusive jet sample and a b-tagged jet sample using the bootstrap method for L1_J55 and the L1_MBTS trigger for L1_J5 in periods A to F: (a) for trigger L1_J5, (b) for trigger L1_J55. For technical reasons, periods G to I are not accessible for this systematic check. The vertical lines indicate the trigger plateaus.*

9.2.3 *b*-tagging efficiency

To obtain the *b*-tagging efficiency $\epsilon_{b\bar{b}}$ the *b*-tagging calibration which was presented in Section 8.2 is applied to simulation. For each *b*-jet which is *b*-tagged in simulation and selected by the analysis cuts, the scale factor $\kappa_b(p_T)$ depending on jet p_T is extracted and applied as a weight. Consequently, a dijet event has an *b*-tagging efficiency weight which is the product of each individual scale factor . The determination of the jet flavour is described in Section 3.1. For each desired variable the *b*-tagging efficiency is evaluated by using Eq. 6.1. The efficiency is determined from the fraction of *b*-jets which are also *b*-tagged as a function of dijet mass, $\Delta\phi$ and χ. To evaluate the systematic uncertainty arising from the calibration itself the weights which are applied to each jet are varied according to their total uncertainty. The variations are conservatively taken as fully correlated including the statistical uncertainty. For each variation the shifted efficiency is evaluated and the maximum deviation from the nominal efficiency is calculated. The maximum deviation is symmetrised in order to be less sensitive to statistical fluctuations and assigned as systematic uncertainty.

The resulting efficiencies and systematic uncertainties are presented in Figs. 9.5 and 9.6.
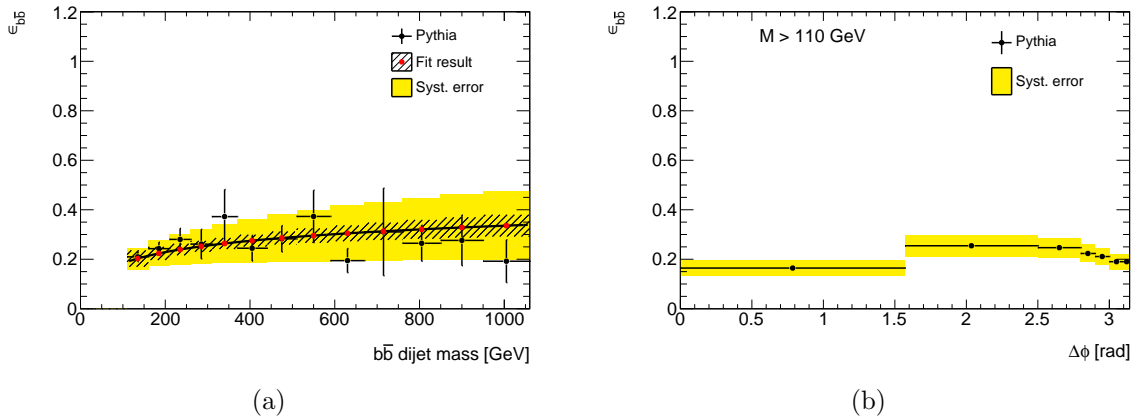

(a) (b)

Figure 9.5: *The b-tagging efficiency $\epsilon_b\bar{b}$ as a function of (a) dijet mass and (b) $\Delta\phi$. The efficiency as function of the dijet mass is fitted with the function: $a \cdot x^b + c$ with $0 < b < 1$. The error bands (hatched area) on the fit result describe the 95% confidence-level on each bin centre.*

The efficiency as a function of dijet mass in Fig. 9.5(a) exhibits a positive slope over the full dijet mass range. As discussed in Section 6.5 the efficiency increases with p_T at low p_T until a plateau is reached. The working point which is used yields a 50% *b*-tagging efficiency integrated over jet p_T. Naively, the $b\bar{b}$ efficiency should be 50%×50%=25%. This assumption is proven to describe the efficiency well and shows that the efficiencies for both jets are largely uncorrelated. The efficiency as a function of dijet mass ranges from 20% to 30%. In order to be less sensitive to statistical fluctuations the efficiency as a function of dijet mass is fitted by the function $a \cdot x^b + c$ with $0 < b < 1$.[1] In case of the other variables $\Delta\phi$ and χ the expected behaviour of the efficiency is not well known and these are not fitted. The statistical uncertainty which drives the fit result as a function of dijet mass is calculated under the assumption of uncorrelated variables. The remaining *b*-tagging efficiencies are calculated according to the following calculation assuming that the true

---

[1]The resulting parameters are $a = 5.0 \pm 2.2$, $b = 1.2 \cdot 10^{-2} \pm 3.8 \cdot 10^{-2}$ and $c = -5.1 \pm 2.2$.
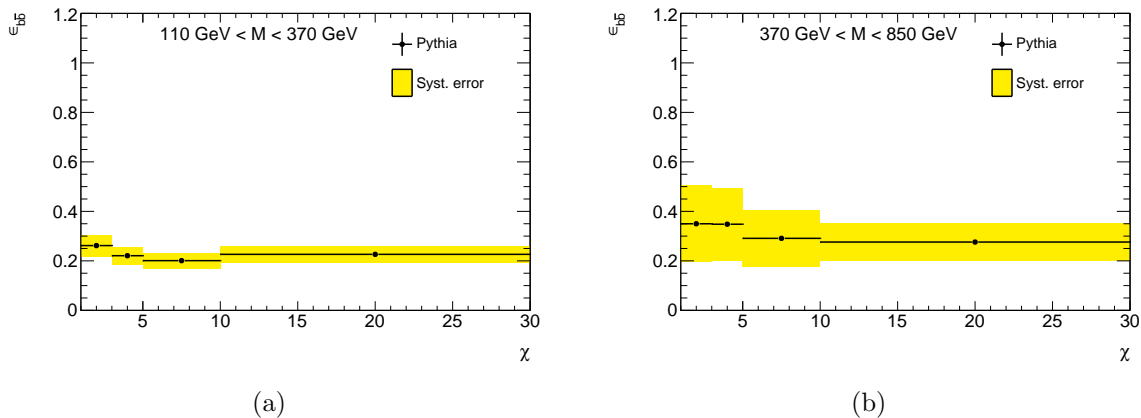
Figure 9.6: *The b-tagging efficiency $\epsilon_{b\bar{b}}$ as a function of angular variable χ in mass regions (a) 110 GeV < M < 370 GeV and (b) 370 GeV < M < 850 GeV.*

efficiency is estimated from the measured efficiency:

$$\epsilon_{b\bar{b}} = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{not-pass}}} \rightarrow \delta\epsilon_{b\bar{b}} = \sqrt{\frac{N_{\text{pass}}N_{\text{not-pass}}}{N_{\text{total}}^3}}, \tag{9.6}$$

where $N_{\text{total}} = N_{\text{pass}} + N_{\text{not-pass}}$, and where $N_{\text{not-pass}}$ and N_{pass} are assumed to be independently Poisson distributed.

In addition to the systematic uncertainty arising from the b-tagging calibration and in case of the efficiency as a function of dijet mass, the 95% confidence-level on the fit result in each bin is assigned as the systematic uncertainty on the fit. At low dijet mass the uncertainty on the fit result dominates the total systematic uncertainty. The maximum of the uncertainty from the fitting results is approximately ± 0.05. The systematic uncertainty due to the b-tagging calibration dominates for dijet masses above 150 GeV. This systematic uncertainty increases with dijet mass because the uncertainty is increasing with p_{T}. The highest total uncertainty on the scale factor is about 20% for jets with $p_{\text{T}} > 140$ GeV (cf. Section 8.2.2). In case of dijet events in which both transverse momenta of the jets are above 140 GeV the maximum expected uncertainty is 40%. This maximum is reached at $M = 1$ TeV as can be read off Fig. 9.5(a). Table 9.1 gives an overview of the systematic uncertainties on the efficiency measurement, where the efficiency as a function of dijet mass is chosen as representative:

The b-tagging efficiencies as function of the angular variable χ and of the azimuthal angle between the leading jets are nearly constant because the full spectrum of jet p_{T} values contributes to each bin. These efficiencies are summarised in Table I.1 and Table I.2. They range approximately between 20% and 30%. The efficiency as a function of $\Delta\phi$ decreases slightly when approaching the back-to-back configuration. The systematic uncertainty on the efficiency as a function of $\Delta\phi$ is also nearly constant. The fractional systematic uncertainty is about 20% for the efficiency measurement in $\Delta\phi$. The efficiencies as a function of the angular variable χ in Fig. 9.6 tend to slightly decrease with higher χ values. The mass region 110 GeV $< M <$ 370 GeV is dominated by jets with lower p_{T}. Therefore the systematic uncertainty is smaller than for the mass region 370 GeV $< M <$ 850 GeV. For the lower mass region the uncertainty varies between 10% and 20%, for the higher

Dijet mass bin [GeV]	b-tagging efficiency	Systematic uncertainty	Syst. fractional uncertainty [%]
110-160	0.20	± 0.04	20
160-210	0.22	± 0.05	22
210-260	0.24	± 0.06	25
260-310	0.25	± 0.07	28
310-370	0.26	± 0.08	31
370-440	0.28	± 0.09	32
440-510	0.29	± 0.10	34
510-590	0.30	± 0.10	33
590-670	0.30	± 0.11	36
670-760	0.31	± 0.12	39
760-850	0.32	± 0.12	38
850-950	0.33	± 0.13	39
950-1060	0.34	± 0.14	41

Table 9.1: *b-tagging efficiency as a function of dijet mass.*

mass region between 25% and 40%.

9.2.4 b-tagging purity

The b-tagging purity which is explained in Eq. 6.4 is determined from the ratio between the number of events with two b-jets which pass the b-tagging requirement and the number of events with two b-tagged jets. The sample of b-tagged jets also includes mistagged c-jets, light-quark and gluon-initiated jets and their contribution needs to be estimated. This purity is extracted from template fits with data, fitting the distribution of the sum of the secondary vertex masses of the two b-tagged jets. The discriminating power of this sum and the different contributions to it are illustrated in Fig. 9.7 integrated over all dijet masses. The events are categorised in three samples with the help of the identification of the jet flavour in simulation: true $b\bar{b}$ events, events with only one true b or $\bar{b}$ and anything else ($b(\bar{b})X$) and events with no true b-jet (XY). The purities of both jets are correlated. If a b-tagged jet in an event has been found already, the probability that a second b-tagged jet corresponds to a real jet with bottom-flavoured content is 90%. The purity in $b\bar{b}$ events is therefore only deducible from an event variable describing this correlation directly (two dimensional) or indirectly by a combination of variables. The sum of the SV0 masses of both jets is a good discriminator describing the correlation indirectly, although due to the presence of neutral particles and energy losses from detector effects the b-hadron invariant mass cannot be fully reconstructed. The core area of the distribution of $b(\bar{b})X$ events lies between the other two, since one component of each event contains a secondary vertex with an invariant mass spectrum of a bottom-flavoured jet. Because of the $b\bar{b}$ fraction is of major interest, denoted as b-contribution, and because of the statistical fluctuations in the

other two samples which contain mistagged events, the $b(\bar{b})X$ and the XY contributions are merged into a single non-b contribution. The binning of the SV0 mass distribution is optimised to reduce statistical fluctuations in simulation and to ensure that every SV0 mass sum distribution in data contains at least few bins with 15 events minimum. With both contributions a linear combination $c_1 \cdot b\bar{b} + c_2 \cdot (XY + b(\bar{b})X)$ where $c_1 = 1 - c_2$ can be found to describe the full SV0 mass sum distribution.

The template fit finds the best suited linear combination to describe data with the help of a binned maximum likelihood method, the so-called *TFractionFitter* [113]. The b-contribution describes the signal template were both jets are matched to a b-hadron at particle level and the other component describing the background is formed by the non-b contribution. To be able to perform these template fits enough data events in each analysis bin are required to ensure the stability of the template fit. It was found that mass bins with a half width of twice the detector mass resolution are sufficient. But the template fit stability limits the cross section measurement as a function of dijet mass to 760 GeV. The templates for higher mass regions have lost their discriminating power due to the increasing mistag rate. Due to the mistag rate the contribution $b(\bar{b})X$ with one b-jet shifts the core area of the template for the non-b contributions towards the b-templates. This causes a large overlap and the instability of the template fit due to the similarities in the template shapes. In the following the measurement of the cross section as a function of dijet mass is limited to 760 GeV. Furthermore, this requirement sets the bin widths of the cross section measurements as a function of $\Delta\phi$ and χ and limits the $\Delta\phi$ measurement to the region above $\pi/2$ due to a lack of statistics. The chosen bin widths in the $\Delta\phi$ measurement are also required to be twice as large as the $\Delta\phi$ resolution [114]. Because the SV0 tagging algorithm possesses a large mistag rate, it is not possible to measure the $b\bar{b}$ dijet cross section without requiring a second b-tag. Without this second b-tag, the statistics could be increased to a large amount but the discriminating power of the SV0 mass vanishes because the non-b contribution becomes overwhelming. The advantage of this data-based approach could not be preserved.

In the following the construction of both templates is described and the final results on the
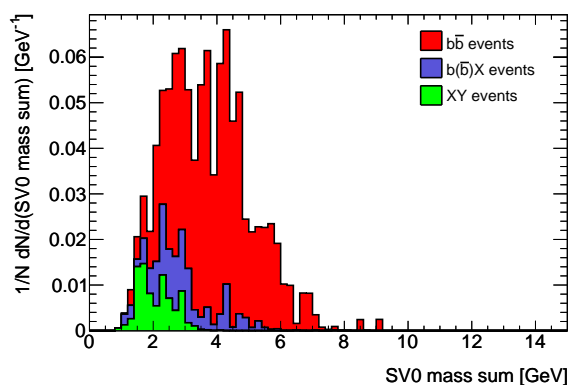
Figure 9.7: *The distribution of the sum of the SV0 masses for all selected b-tagged dijet events with jets $p_T > 40$ GeV. Events are separated into three categories: true $b\bar{b}$ events, events with only one true b or $\bar{b}$ and anything else ($b(\bar{b})X$ events) and events with no true b-jet denoted as XY events. The single contributions are drawn stacked on top of each other.*

b-tagging purity are presented. This is followed by the discussion and presentation of the

systematic uncertainties. Finally, additional cross checks on the systematic uncertainties, the fitting method and implementation are presented.

To reduce the deficiencies of the simulated samples and their technical combination, high statistics b- and non-b templates are constructed from the simulated distributions by fitting them. Signal and background templates are needed for every bin in the cross section measurements. The fit used to extract the shape of a template uses either a Gaussian or a Landau distribution with mean μ and sigma σ. Both shapes are well motivated. Templates, which have no impurities and are only subject to resolution effects, should be well described by a Gaussian shape. Templates like the non-b templates have a core region, which is defined by the dominant fraction of light-jets, and a tail from the events with a c- and/or b-contribution.

As the shapes may vary with phase space range the optimal fitting function is determined automatically for each template by fitting the sum of both possible functions with a fraction factor f: $f \cdot \text{Gaussian} + (1-f) \cdot \text{Landau}$. The fraction factor then determines the optimal fitting function with which the template is then re-fitted. For a Gaussian shape is decided if $f > 0.5$ and otherwise for a Landau distribution. This selection is justified because the shapes of the templates clearly favour one fit function and the values of f are always either close to zero or close to one. The distributions are fitted with a Gaussian distribution in 85% of all cases. Two examples of fits on the b- and non-b contributions are given in Fig. 9.8 for the mass range 160 GeV $< M <$ 210 GeV. Appendix I contains all fits on b- and non-b contributions. The lack of statistics in the templates is clearly visible. With the help of the fitting procedure the contributions below 1 GeV and above 5 GeV are determined because this phase space is insufficiently filled. An additional advantage of this method is that it can constrain a bin content with the help of the neighbouring bins more than the statistical error would allow. The outcome of the fits is used to generate high statistics templates so that the statistical error on the likelihood fit is completely dominated by the available data statistics.

The systematic uncertainty on the fitting procedure is evaluated with the help of the

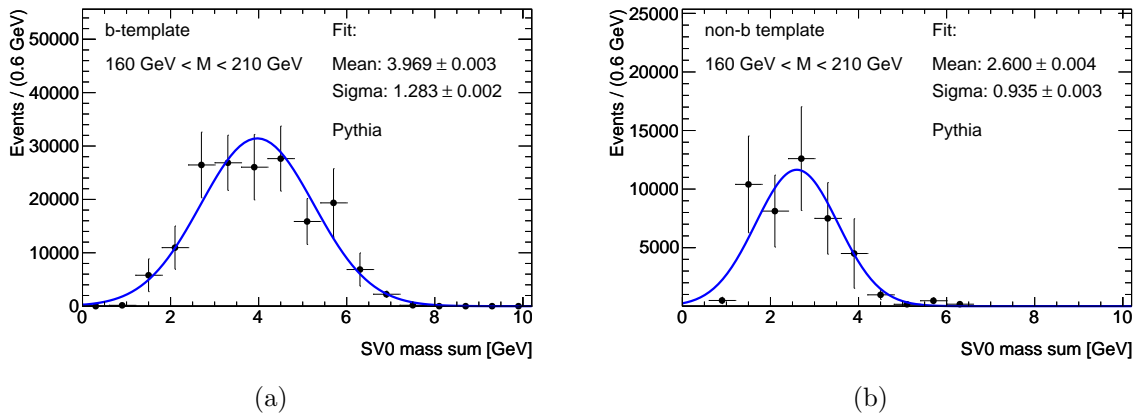

Figure 9.8: *Examples of construction of templates for (a) b and (b) non-b contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of dijet mass. Shown are the raw templates and the fit result for the mass range 160 GeV < M < 210 GeV.*

error parameters from the fit. Because both fitting functions are well motivated and in

each case the optimal shape is used, no additional uncertainty on the choice of fitting function is obtained. The covariance matrix is used to generate the error ellipses of the fit parameters mean μ and sigma σ. Two examples of the error ellipses are illustrated in Fig. 9.9. The results of the fitting procedures represent the centre of the ellipses. In case of Fig. 9.9(a) the correlation between the fit parameters is small and the ellipse is almost circular, the correlation in Fig. 9.9(b) is negative and large. Pseudo-experiments are performed according to the error ellipse returning the pair (μ,σ) to evaluate the 68% confidence-level of the fit result in each bin. Consequently, variations of the fit prediction are obtained and serve as basis for variations in the template construction. Two examples of variations are presented in Fig. 9.10 for a *b*- and a non-*b* template in the $\Delta\phi$ range 2.8 rad to 2.9 rad and in the χ range 3 to 5, respectively. The dashed lines indicate the variations in the template construction and the continuous black line is the nominal fit prediction.

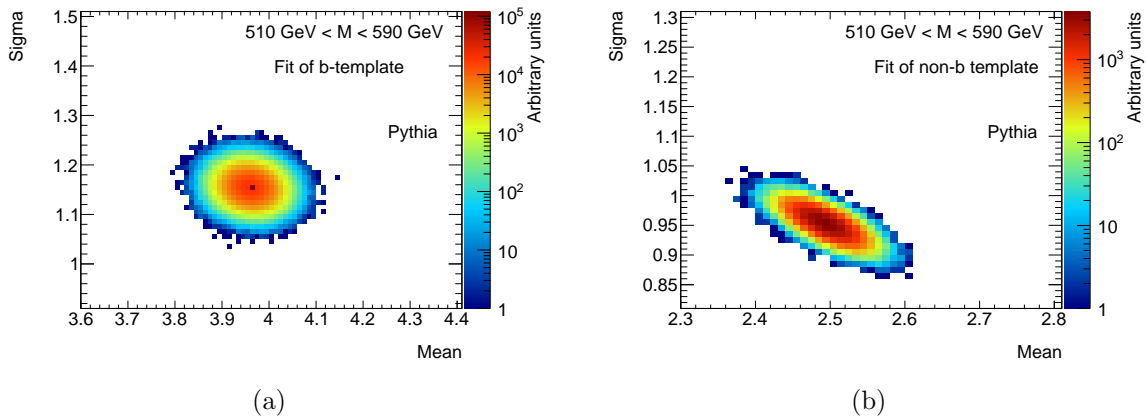Using the constructed templates the template fits are performed on data in bins of dijet

(a) (b)

Figure 9.9: *Generated error ellipsis for an example of templates (a) b-template and (b) non-b template for the dijet mass range 510 GeV < M < 590 GeV.*

mass, $\Delta\phi$ and χ. Examples of these fits are collected in Fig. 9.11: (top) in bins of dijet mass, (middle) $\Delta\phi$ and (bottom) χ. The non-*b* and *b*-template are drawn stacked on top of each other. The template fits not shown here are included in Appendix I. The statistical uncertainty on the purities reflect the available statistics in data because the constructed templates contribute a negligible statistical uncertainty. The statistical uncertainties are re-calculated from the covariance matrix [115]. With decreasing data statistics the statistical uncertainty on the likelihood results grows for example with increasing dijet mass. The lower parts of the plots show the ratio between the data distribution and the summed simulation templates with applied fractions from the likelihood fit. The overall agreement between data and simulation and the fit quality gives high confidence in the fitting procedure. The systematic uncertainty arising from the fitting procedure is determined by re-doing the template fit on data with the varied templates introduced in Fig. 9.10. The effects of the variations are added in quadrature to the statistical uncertainty of the likelihood fit.

The bin widths for the SV0 mass sum distribution may cover shifts in the SV0 mass distributions between data and simulation. The variations on the templates from the 68% confidence-level on the fitting parameters do not cover possible shifts in the scale of
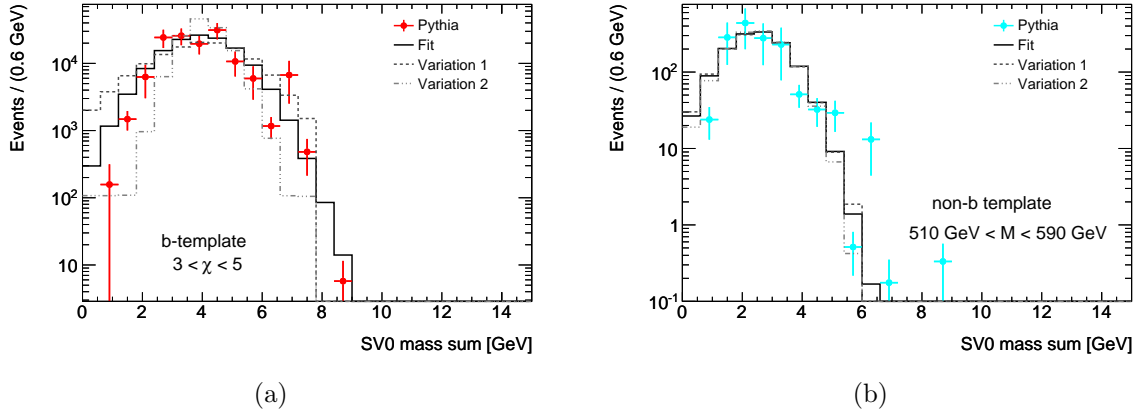
(a)   (b)

Figure 9.10: *Shown are the raw templates from* Pythia *simulation and the constructed templates from the fitting procedure with the overlaid variations (dashed lines). The variations reflect the 68% confidence-level for the constructed templates.*

the invariant mass of the secondary vertex. Therefore the template fits are re-done with templates shifted by half a bin width. The resulting deviations are added in quadrature to the systematic uncertainty obtained.

The total uncertainty on the b-tagging purity consists of the variations on the shape and the scale of the templates and the statistical uncertainty on the template fits themselves. The total uncertainty is symmetrised. In cases where the systematic uncertainties would shift the b-tagging purity outside the definition interval $[0, 1]$, the uncertainty is not symmetrised. The resulting purities and systematic uncertainties for the dijet mass and $\Delta\phi$ bins are presented for the measured phase space in Fig. 9.12 and detailed information on the purities and the systematic uncertainties are summarised in Table 9.2 for the measurement as a function of dijet mass. Further details on the remaining variables are given in Table I.3 and Table I.4. The purities measured in dijet mass bins range from 50% in the high mass region to 90% in the low mass region. The maximum fractional systematic uncertainty is 30%. The purity decreases with increasing dijet mass, reaches a plateau at around 60% and increases in the two last bins. The systematic uncertainties are driven by the lack of statistics in the simulation. Only in the last four dijet mass bins, the statistical uncertainty competes with the systematic uncertainty and at last dominating at 670 GeV to 750 GeV. At low dijet mass the mistag rate is very small and the construction of the non-b templates suffers from the small fraction of inclusive jets which enter into these distributions. Consequently, the non-b templates suffer from large systematic uncertainties due to the fit variations. Since the variations in the non-b template cause the largest effect on the b-tagging purity, the uncertainty at the low mass region becomes very large. Variations affecting the tails of the non-b templates, which pollute the core region of the b-templates, have a very large effect on the b-tagging purity. The systematic uncertainty then decrease with increasing dijet mass. On the other side of the phase space the b-tagging efficiency decreases and the b-templates suffer from small statistics. Therefore the systematic uncertainties tend to grow slightly again with dijet mass due to the fit uncertainties on the b-templates. The purities in the $\Delta\phi$ bins range from 55% to 70%, and when approaching the back-to-back configuration the purities are decreasing. The systematic uncertainty is about 20% to 30%. The statistical uncertainties

are negligible compared to the systematic uncertainty. In Fig. 9.13 the purities in bins of the angular variable χ are presented for both mass regions. The purities vary between 50% and 70% for 110 GeV $< M <$ 370 GeV and 30% and 40% for 370 GeV $< M <$ 850 GeV. The second lowest bin in both mass regions is suffering from large systematic uncertainties due to the badly distributed simulation events and the systematic uncertainty reaches up to 50% in the lower mass region and up to 90% in the higher mass region. Except for those bins the fractional systematic uncertainty is about 30% for the lower mass region and 50% in the higher mass region.

Systematic checks of the method:

Dijet mass bin [GeV]	b-tagging purity	Total uncertainty	Fractional uncertainty [%]
110-160	0.90	$^{+0.07}_{-0.27}$	$^{+8}_{-30}$
160-210	0.72	± 0.17	24
210-260	0.59	± 0.20	33
260-310	0.59	± 0.13	21
310-370	0.65	± 0.15	24
370-440	0.69	± 0.17	25
440-510	0.53	± 0.13	25
510-590	0.49	± 0.12	24
590-670	0.64	± 0.19	30
670-760	0.81	± 0.16	19

Table 9.2: *b-tagging purity as a function of dijet mass.*

Systematic uncertainties based on physical arguments like the dependency of the non-b template on the correct modelling of the c- and light-jet contribution can hardly be estimated because their effects are hidden below large statistical uncertainties in the simulated contributions. Nevertheless, the systematic shifts in the templates are evaluated due to the uncertainty on the amount of light- and c-jet contributions. Both contributions are multiplied separately by a factor of 3 and 0.33. From these distributions new templates are constructed and used in the template fit on data. The resulting deviations between the nominal and the shifted purities is plotted in Fig. 9.14 on top of the systematic uncertainty assigned by the uncertainty on the fitting procedure. This nominal systematic uncertainty covers the deviations from the templates with weighted contributions. Therefore the physics motivated systematic uncertainties are well estimated by the assigned systematic variations. The uncertainty from the fitting procedure partly covers the systematic from the modelling of the non-b contributions but has the advantage of being well defined. The variations in the templates from the fitting procedure mainly influence the tails. Most importantly the rising edge at low SV0 mass sum dominated by the non-b template and the falling edge of the non-b template reaching into the core area of the b-template as well as the falling edge of the b-template describing the very high SV0 mass sum distribution are influenced. These are the critical parts in the template fit. These parts are also mainly
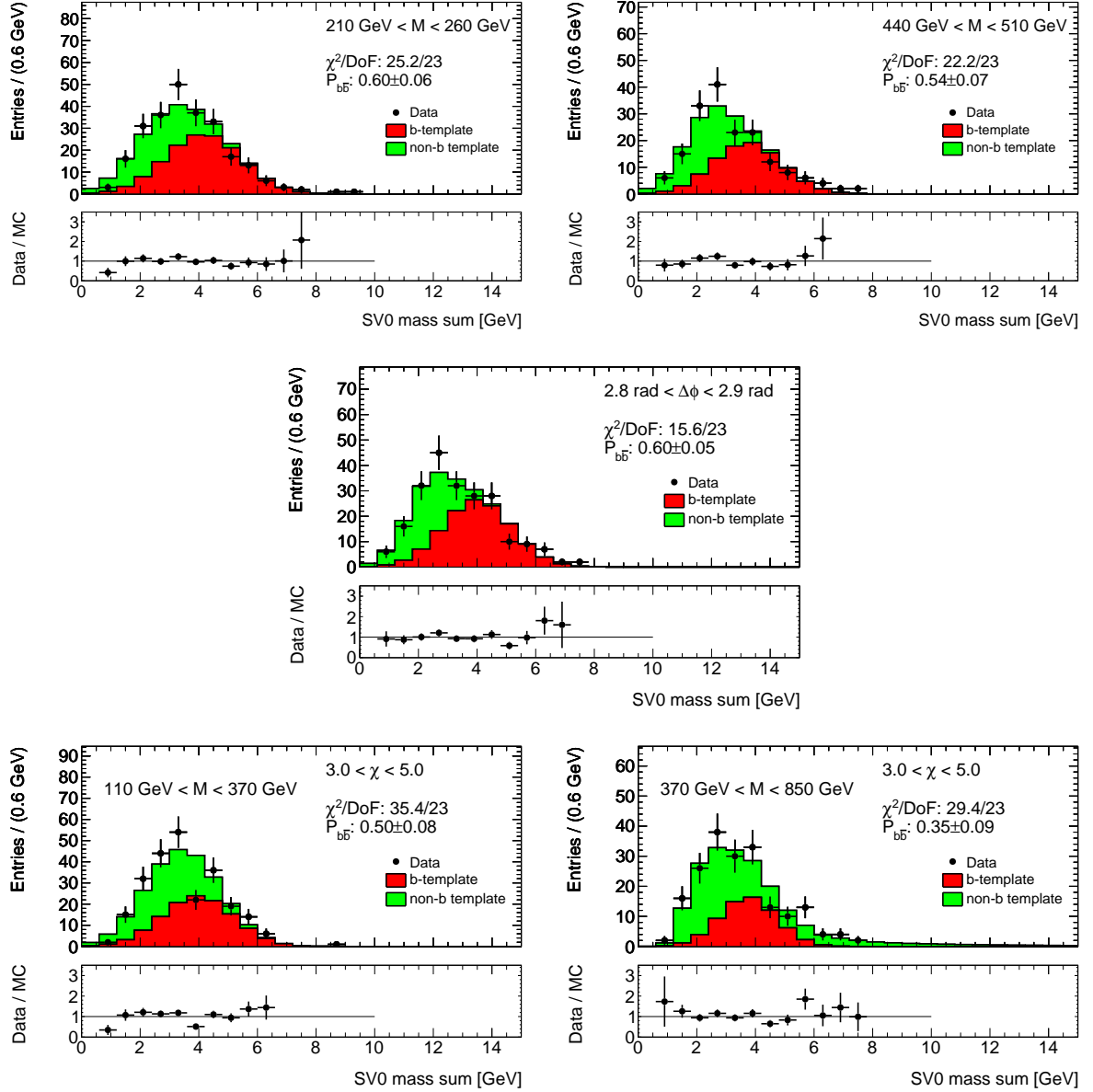
Figure 9.11: *Examples of the final template fits on data using a binned maximum likelihood fit (TFractionFitter): template fits in a bin of dijet mass (top) for a low mass bin (top left) and a high mass bin (top right), a template fit in a bin of $\Delta\phi$ (middle) and a template fit in a bin of χ (bottom) for the lower mass range (bottom left) and the upper mass range (bottom right). The uncertainty on $P_{b\bar{b}}$ denotes the statistical uncertainty from the likelihood fit. DoF stands for the number of degrees of freedom. The single contributions are drawn stacked on top of each other.*
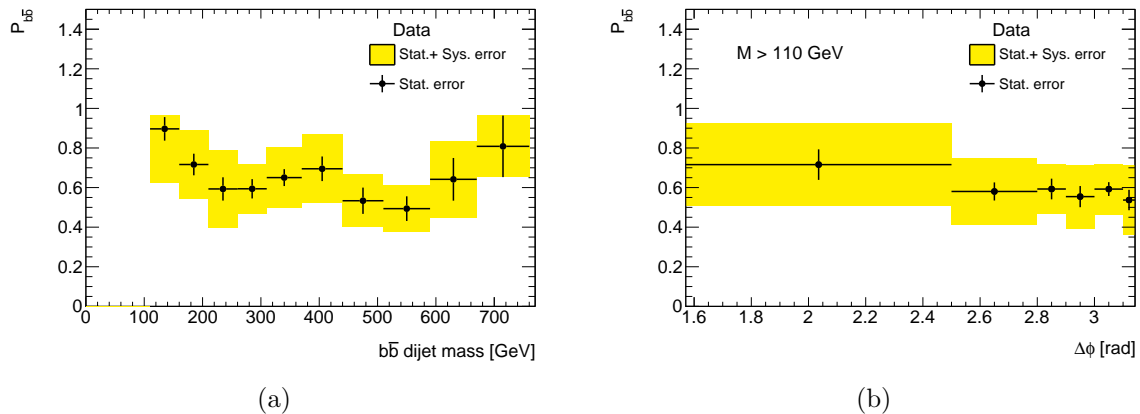
Figure 9.12: *The b-tagging purity $P_{b\bar{b}}$ as a function of (a) dijet mass and (b) the azimuthal difference $\Delta\phi$ extracted from the template fits. The statistical error are extracted from the likelihood fit. The systematic uncertainty contribution to the error band are extracted with the help of variations of the templates within the 68% confidence-level.*

the focus in the physics motivated systematics but are not sufficiently constrained by the simulated distributions. The templates with weighted contributions can only represent the limited phase space filled by the simulation. In this sense the uncertainty from the fitting procedure is well-defined because it is independent from such uncontrollable effects.

The quark multiplicity inside a jet modifies the *b*-tagging performance. A jet including two *b*-quarks is more likely to be *b*-tagged because the jet contains on average more tracks and has a higher probability that a secondary vertex is reconstructed. The uncertainty on the modelling and the frequency of the multiplicity of *b*-quarks is included in the *b*-tagging calibration. The effect on a measurement which requires two *b*-tagged jets is therefore overestimated because the sensitivity to such jets are largely reduced. This may also affect the modelling of the SV0 mass. A higher number of tracks attached to the secondary vertex can change the shape of the SV0 mass distribution. This effect is studied with the help of the SV0 mass distributions in bins of track multiplicity in the secondary vertex. Also here it is expected that the effect on a dijet measurement is small compared to an inclusive *b*-jet measurement. These distributions studied are included in Appendix I as Fig. I.11 and show no deviation between the track multiplicity and the shape of the SV0 mass distribution in data and simulation.

To validate the chosen templates the purities obtained by the nominal templates are compared to the purities in simulation. As the purities in simulation suffer from the lack of Monte-Carlo statistics, it is expected that the purities agree within their uncertainties. In Fig. 9.15(top left) it is shown how both purities agree with each other. In the lower part of the plot it is visible that the purities agree within their uncertainties and that the overall characteristics are reproduced by simulation. Furthermore, the results obtained with the nominal templates are compared to the results of template fits with a simplified version of the templates. The templates integrated over the dijet mass region used in the analysis are the basis for two generic templates of each type: *b*-template and non-*b* template. The SV0 mass distribution for *b*-jets is a Gaussian distribution with mean at around 2 GeV and for background jets the shape follows mainly a Landau distribution with a mean around 1 GeV. From this a simplified, generic template for each sort of events can be generated by adding
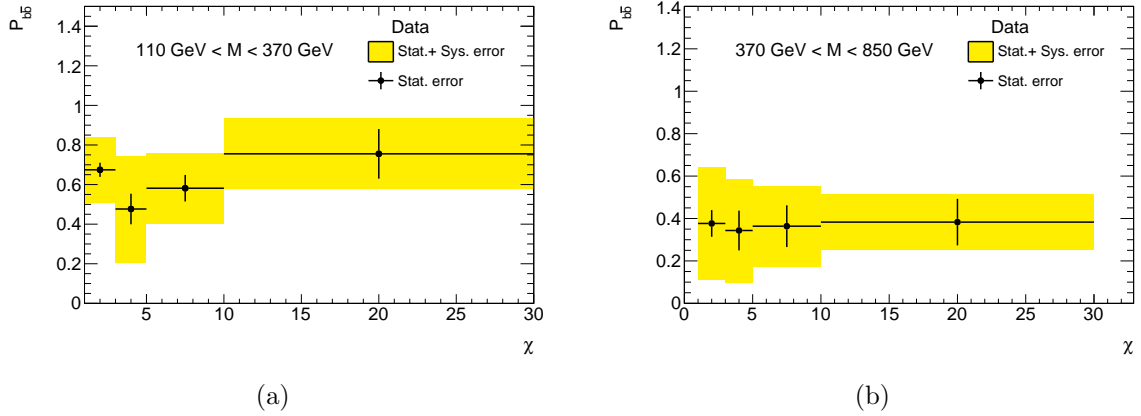
Figure 9.13: *The b-tagging purity $P_{b\bar{b}}$ as a function of the angular variable χ for (a) 110 GeV < M < 370 GeV and (b) 370 GeV < M < 850 GeV extracted from the template fits. The statistical error are extracted from the likelihood fit. The systematic uncertainty contribution to the error band are extracted with the help of variations of the templates within the 68% confidence-level.*

two Gaussian distributions for the *b*-template and two Landau distributions for the non-*b* template. They are simplified in the sense that they do not account for any change in the non-*b* template concerning the relative contributions from *c*- and light-jets nor between the XY and $b(\bar{b})X$ contributions. The mistag rates accordingly develop independent from each other with dijet mass. The template fits on data are evaluated with these generic template and the outcome is presented in Fig. 9.15(top right). Also here the relations between purities and dijet mass range is reproduced and the purities agree well within their uncertainties.

At last the implementation of the likelihood fit in the *TFractionFitter* is compared to the *RooFit* package [116]. This comparison is shown in Fig. 9.15(bottom). The results deviate from each other within only a small fraction of the statistical uncertainties. Even more in both methods the statistical uncertainties agree in size and the available statistics set in each bin is sufficient to ensure the stability of the template fits. In all discussed cases the conclusions hold also in terms of variables $\Delta\phi$ and χ.

9.2.5 Jet energy scale

Besides the *b*-tagging related systematic uncertainties the jet energy scale is the dominant experimental uncertainty in jet measurements. In Section 8.3.3 the jet energy scale uncertainty for *b*-jets has been presented and Table 8.3 summarised the uncertainty as a function of jet p_T and rapidity. The jet energy scale uncertainty includes contributions from inclusive jets and *b*-jet specific ones estimated with the help of tracking information. The minimum jet energy scale uncertainty on *b*-jets is 3.2% and the maximum amounts to 4.9%. To evaluate the effect on the cross section measurement with the help of simulation, the energies of all jets at the beginning of the event selection are re-scaled. The variation is fully correlated in each jet p_T bin. The shifts between the measured cross section with nominal scale and the cross sections with varied scales are determined. It is found that the variations in the sample of *b*-tagged jets are largely influenced by statistical
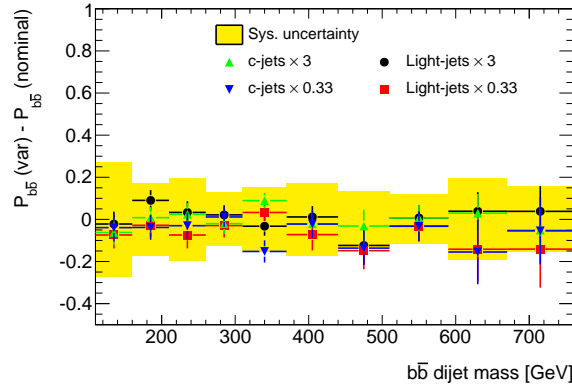
Figure 9.14: *Difference between the purities obtained by templates with an varied c- and light-jet contribution and the nominal purities presented in Table 9.2. The differences are compared to the nominal systematic uncertainty obtained from the uncertainty on the fitting procedure.*

fluctuations. Therefore the effect of the jet energy scale on the cross section measurement of b-jets is calculated with the help of the inclusive jet sample. It has been verified that the uncertainty on $b\bar{b}$ events is of the same order and can be estimated by inclusive jets. The variation which increases (decreases) the energy scale of jets causes that events with two jets near the selection cut of 40 GeV are selected (discarded) because the transverse momenta of both jets are increased (decreased). On the other side events with two jets with a dijet mass near 760 GeV migrate out of (into) the valid phase space. In case of the dijet cross section as a function of dijet mass the steeply falling cross section amplifies the effect of such migrations. In case of the variation which increases (decreases) the energy scale the cross section in each mass bin is increased (decreased). The fractional variation in the cross section as a function of dijet mass is presented in Fig. 9.16 and summarised in Table 9.3. The resulting systematic uncertainties due to the jet energy scale are between 10% to 20%. The systematic uncertainty decreases slowly with increasing dijet mass because the jet energy scale uncertainty decreases with increasing transverse momentum. The region with rising uncertainty above 500 GeV is hardly touched in this measurement. As a reminder Fig. 7.5 illustrates the composition of each dijet mass bin in terms of the transverse momentum of the leading jet. The effect of the jet energy scale uncertainty is higher when the energy scale is increased because of the steeply falling spectrum in which the migration is larger from the right edge of each bin. The detailed numbers for the jet energy scale uncertainty on the cross section measurements as a function of $\Delta\phi$ and χ are given in Tables 9.4 and 9.5. The systematic uncertainties of the cross section measurement as a function of $\Delta\phi$ are between 12% and 19%. The uncertainties decrease slightly for back-to-back configuations. The systematic uncertainties of the cross section measurement as a function of χ are between 10% and 20%. In the lower mass region the uncertainty is larger due to the larger fraction of jets with p_T between 40 GeV and 100 GeV which contribute the largest jet energy scale uncertainty. The higher mass region exhibits almost a constant systematic uncertainty due to the jet energy scale between 10% to 13%.

9.2.6 Unfolding correction

In order to compare to higher-order QCD calculations, the measured cross section measurement has to be corrected for detector effects after which the measurement is independent
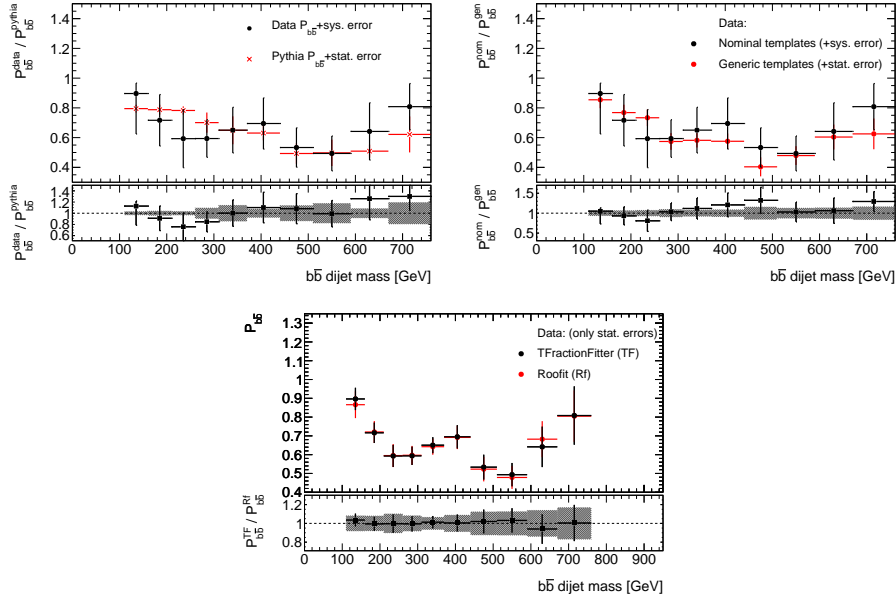
Figure 9.15: *The outcome of the nominal template fits are compared to: (top left) the purities in simulation, (top right) the purities in data with different templates and (bottom) the purities in data obtained with a different fitting programme (RooFit [116]).*

from the detector used. The jet energy calibration restores the jet to the hadronic scale on calorimeter level but does not account for smearing effects or jet migration between bins in a measurement as a consequence of the finite energy resolution of the calorimeter and the jet energy calibration. To restore the measured cross section to particle level a correction is necessary which is referred to as unfolding correction. A bin-by-bin unfolding based on the PYTHIA simulation is used to correct for all detector effects in a single step. The correction factors are calculated from *b*-jets reconstructed from calorimeter clusters and *b*-jets at particle level reconstructed from final-state particles including non-interacting particles, namely neutrinos and muons. The jet flavour is determined as discussed in Section 3.1 from an angular matching to *b*-hadrons. The unfolding correction

Figure 9.16: *The fractional jet energy scale uncertainty on the $b\bar{b}$ dijet cross section as a function of dijet mass determined on simulated events.*

Dijet mass bin [GeV]	Fractional JES uncertainty [%]
110-160	$^{+17}_{-15}$
160-210	$^{+17}_{-16}$
210-260	$^{+18}_{-15}$
260-310	$^{+18}_{-16}$
310-370	$^{+20}_{-16}$
370-440	$^{+18}_{-16}$
440-510	$^{+17}_{-16}$
510-590	$^{+19}_{-16}$
590-670	$^{+21}_{-17}$
670-760	$^{+18}_{-17}$

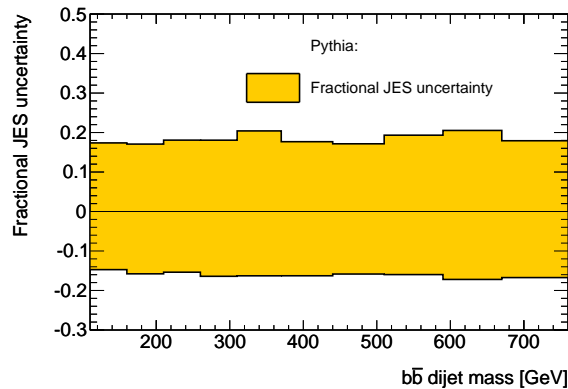**Table 9.3:** *Jet energy scale uncertainty as a function of dijet mass.*

$\Delta\phi$ bin [rad]	Fractional JES uncertainty [%]
1.6-2.5	$^{+21}_{-18}$
2.5-2.8	$^{+20}_{-16}$
2.8-2.9	$^{+19}_{-16}$
2.9-3.0	$^{+17}_{-15}$
3.0-3.1	$^{+15}_{-14}$

Table 9.4: *Jet energy scale uncertainty as a function of the difference in azimuthal angle $\Delta\phi$.*

is determined from the bin-by-bin ratio of the calorimeter jet and particle jet cross sections:

$$C_{\text{unf}}(X) = \frac{d^n\sigma/(dX_1...dX_n)(\text{particle } b\text{-jets})}{d^n\sigma/(dX_1...dX_n) \text{ (calorimeter } b\text{-jets})} = \frac{N_{b\bar{b}}(X)(\text{particle } b\text{-jets})}{N_{b\bar{b}}(X) \text{ (calorimeter } b\text{-jets})}, \quad (9.7)$$

where X stands for dijet mass M, $\Delta\phi$ or $\chi(M)$ and N for the number of events in each bin at calorimeter or particle level. The correction factors are applied to each measured cross section to obtain the particle-level measurement. The correction includes effects from acceptance cuts and from the jet energy resolution, as well as the underestimation of the b-hadron energy due to lost neutrinos and muons. In Section 9.2.8 these contributions are disentangled with the help of the inclusive jet cross section.

The most important systematic uncertainty arises from the discrepancies between the cross section shape in data and simulation. PYTHIA simulation as discussed in Section 8.1.1 is not able to describe the inclusive jet p_T and dijet mass distribution. So the underlying particle-level prediction of PYTHIA underestimates the distribution at high p_T and consequently overestimates the number of jets in the low p_T range. On the basis of Fig. 8.1(a) an event weight is computed to cover the deviations between data and PYTHIA simulation. The event weight is calculated as a function of leading jet transverse momentum at particle level p_T^truth with the help of the following function:

χ bin	110 GeV $< M <$ 370 GeV	370 GeV $< M <$ 850 GeV
	Fractional JES uncertainty [%]	
1-3	$^{+15}_{-14}$	$^{+17}_{-15}$
3-5	$^{+17}_{-15}$	$^{+17}_{-15}$
5-10	$^{+18}_{-16}$	$^{+17}_{-15}$
10-30	$^{+20}_{-18}$	$^{+18}_{-16}$

Table 9.5: *Jet energy scale uncertainty as a function of angular variable χ for both mass ranges.*

$$\text{event weight} = 1 + p \cdot \left(\frac{p_{\mathrm{T}}^{\mathrm{truth}} / \text{ GeV} - 17}{1103} \right)^{q}, \tag{9.8}$$

with parameters $p = 1, -0.5$ and $q = 0.33, 1, 3$. The event weights are illustrated in Fig. 9.17(a). These event weights change the number of jets in each p_{T} bin as a function of leading jet p_{T} at particle level. In case of the dijet cross section as a function of dijet mass the resulting effect of the event weights is presented in Fig. 9.17(b) and the differences in the cross section cover the deviations in Fig. 8.1(b). The tendency of the discrepancy in Fig. 8.1(a) is also visible in comparison to other parton shower generators. In the appendix the comparison is shown for the HERWIG++ generator in Fig. I.12(a). Therefore the discrepancy is dominated by missing next-to-leading order contributions. Consequently, this systematic error cannot be estimated with the help of other parton shower generators but with the help of data. Furthermore, involving different generators into the estimation of the systematic uncertainty introduces a duplication in the systematic uncertainty due to the jet energy scale. Deviations in the unfolding correction between different generators are partly due to different, simulated jet energy scales.

Besides this systematic error on the modelling of the cross section shape, the uncertainty on the modelling of the detector resolution needs to be accounted for. The size of the unfolding correction is directly linked to the finite detector resolution which is simulated. In addition to the energy resolution and uncertainty presented in Section 8.3.2 also the angular resolutions are important for the cross section measurements. According to Eq. 5.1 the resolutions of the azimuthal angle and the rapidity are determined with the help of the simulation in which jets at generator level are matched in angular space to reconstructed jets. The resolutions are presented in Fig. 9.18 as a function of jet p_{T}. The resolutions are determined through fits in bins of p_{T} with a Gaussian distribution on the difference between the generated and reconstructed value normalised to the generated value.

In the following the contributions to the systematic uncertainty are discussed in detail and presented separately for the different cross sections.

Systematic contributions to the unfolding correction factors as a function of dijet mass:

- The unfolding factors for the dijet mass are fitted by an exponential function $a \cdot e^{b \cdot M / \text{ GeV}}$ to limit the effect of missing statistics.[2] The statistical error is calculated assuming uncorrelated quantities. An exponential function is chosen as it is flexible

---

[2]The resulting parameters are $a = 1.5 \cdot 10^{-1} \pm 3.1 \cdot 10^{-2}$ and $b = -6.2 \cdot 10^{-5} \pm 4.5 \cdot 10^{-5}$.
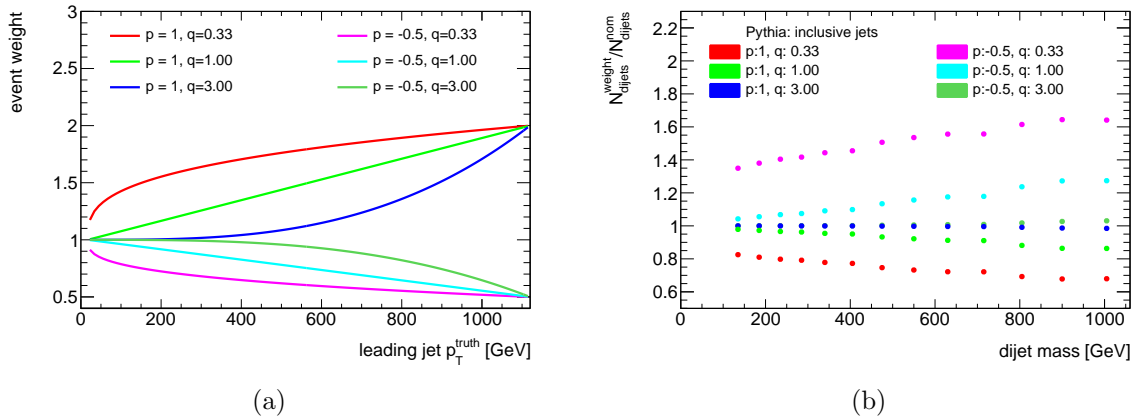
Figure 9.17: *The discrepancies between data and simulation are the basis for the event weights with are parametrised with the function Eq. 9.8. (a) shows the chosen parametrisations as a function of leading jet p_T^{truth} at particle level. The resulting effect on the dijet cross section as a function of dijet mass is presented in (b).*

enough to fit the unfolding correction of the dijet cross section as well as the unfolding corrections from the b-jet cross section as a function of p_T [117, 118] and the unfolding corrections from the inclusive analysis [119]. There is no theoretical prediction for the functional form of the unfolding correction factors. The uncertainty on the fit prediction is estimated from the 95% confidence-level on the fit result and contributes about 4% to the systematic uncertainty. The resulting unfolding correction is presented in Fig. 9.20(top left).

- The resulting effect of the cross section shape variations by the function in Eq. 9.8 on the unfolding correction factors for the dijet mass are around 2% and included in Appendix I in Fig. I.13. Due to the limited statistics the varied unfolding corrections are fitted by the exponential function $a \cdot e^{b \cdot M/\text{ GeV}}$ and the resulting deviation from the nominal unfolding correction is calculated. The deviations are symmetrised and the maximum deviation observed is added in quadrature to the systematic uncertainty.

- The energy resolution in data is about 10% worse than simulated in PYTHIA (cf. Section 8.3.2). In pseudo-experiments a randomised energy amount that corresponds to a resolution smearing of 10% is added to each reconstructed jet at calorimeter level . Due to the limited statistics this systematic uncertainty is evaluated using inclusive jets. The deviation in the unfolding correction for the energy resolution smearing is presented in Fig. 9.19 for three different smearing factors: 5, 10 and 15%. The assigned systematic uncertainty which is added in quadrature is based on the 10% resolution smearing and contributes at most 1% to the systematic uncertainty.

- Uncertainties on the unfolding correction for the cross section measurement as a function of dijet mass from the angular resolutions on azimuthal angle and rapidity are negligible.
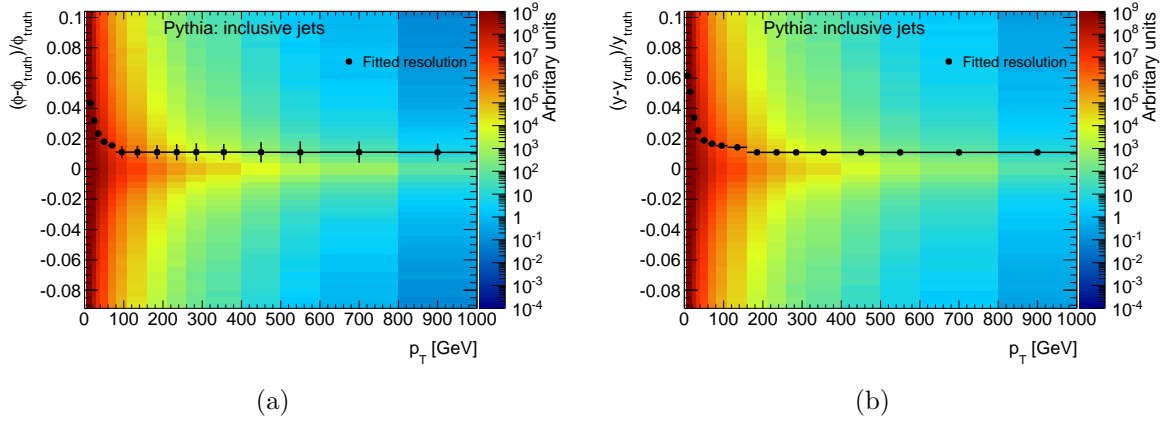
Figure 9.18: *The resolutions of the azimuthal angle and the rapidity determined with the help of the simulation. The resolutions are identical to the Gaussian width of the distribution of the difference between generated and reconstructed value normalised to the generated one. The coloured density distribution in each figure shows the fractional resolution as a function of leading jet $p_{\mathrm{T}}^{\mathrm{truth}}$.*

The methods and description of the evaluation of the systematic uncertainties in case of the cross section as a function of dijet mass are equivalent for the variables $\Delta\phi$ and χ. For those two cases the unfolding correction factors are not fitted and no systematic uncertainty due to the fit needs to be assigned. The statistical errors on the unfolding correction is calculated by separating the number of events in each bin into three statistically uncorrelated numbers:

$$C_{\mathrm{unf}}\,(\text{bin i}) = \frac{N(\text{truth\&reco}) + N(\text{truth\&!reco})}{N(\text{truth\&reco}) + N(\text{!truth\&reco})}, \qquad (9.9)$$

where $N(\text{truth\&reco})$ is the number of events in which both reconstructed dijet masses lie in the same bin, and $N(\text{truth\&!reco})$ and $N(\text{!truth\&reco})$ are the number of events in which the reconstructed dijet mass at calorimeter level and the dijet mass at particle level do not correspond to either other. Either the reconstructed event has migrated out of this bin ($N(\text{truth\&!reco})$) or a reconstructed event from a different bin has migrated into this bin ($N(\text{!truth\&reco})$). With the help of this separation error propagation for uncorrelated variables is applicable.

The remaining contributions are the shape variation and the resolution uncertainty which are both added in quadrature. The resolution uncertainty on the angular variables is estimated with the help of different sources. In preliminary studies with early data the deviation between the resolution in data and simulation was estimated to be 10% by matching topological clusters (cf. Section 5.1.2) to tracks [114]. In between improvements in simulation have increased the agreement [120]. For this analysis and phase space, the uncertainty is estimated to be 5%. The contributions are described in more detail in the following.

Systematic contributions to the unfolding correction factors as a function of $\Delta\phi$:

- The effects of the cross section shape variations on the unfolding correction factors as a function $\Delta\phi$ are within 1.5%.
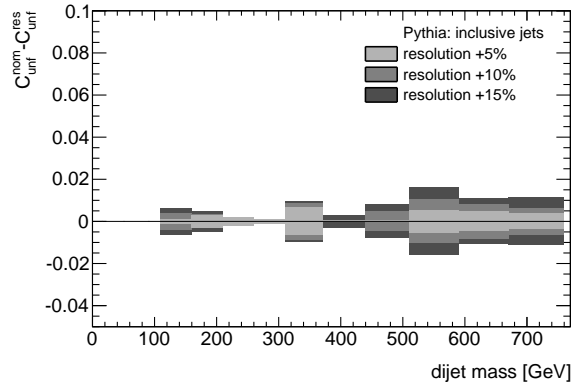
Figure 9.19: *Effect of the resolution smearing on reconstructed jets at calorimeter level on the unfolding correction as a function of dijet mass. The systematic uncertainty on the $b\bar{b}$ unfolding correction is evaluated using inclusive jets.*

- The resolution of the azimuthal angle ϕ is determined from simulation. The resolution is smeared by 5%. In pseudo-experiments a randomised smearing term is added to each reconstructed jet at calorimeter level and the cross section as a function of $\Delta\phi$ is re-evaluated. The effect is of the order of 1% to 5%.

Systematic contributions to the unfolding correction factors as a function of χ:

- The effects of the cross section shape variations on the unfolding correction factors as a function χ are within 1.5%.

- The resolution of the angular variable χ is determined from simulation. The resolution is smeared by 5%. In pseudo-experiments a randomised smearing term is added to each reconstructed jet at calorimeter level and the cross section as a function of χ is re-evaluated. The effect is of the order of 1% to 2%.

The individual contributions to the systematic uncertainties are symmetrised and finally summed in quadrature. The unfolding correction factors and the resulting systematic error bands are collected in Fig. 9.20. The corrections range from 20% to about 55% at most. The corrections predominately account for the energy lost due the semi-leptonic b-hadron decays. The size of these corrections are in agreement with the unfolding corrections determined from HERWIG++ simulation which can be found in Fig. I.12(b) with uses a different model for fragmentation and decays. Previous studies discussing the transverse momenta of muons and the jet energy scale with respect to the total transverse momentum in tracks have shown that the missing energy is described appropriately (cf. Sections 8.1.3 and 8.3.3) in PYTHIA simulation. The final systematic uncertainties on the unfolding correction as a function of dijet mass is of the order of 5% independent of the dijet mass and dominated by the uncertainty assigned to the fit prediction. This is reasonable because the unfolding correction is not well constrained due to the lack of statistics. The systematic uncertainties on the unfolding correction as a function of $\Delta\phi$ range between 1% and 7%. The maximum uncertainty is reached in the last four bins of the measurement. The systematic uncertainty on the unfolding correction as a function of χ is nearly 1% in the lower mass region 110 GeV $< M <$ 370 GeV and between 1% and 2% in the higher mass region 370 GeV $< M <$ 850 GeV.
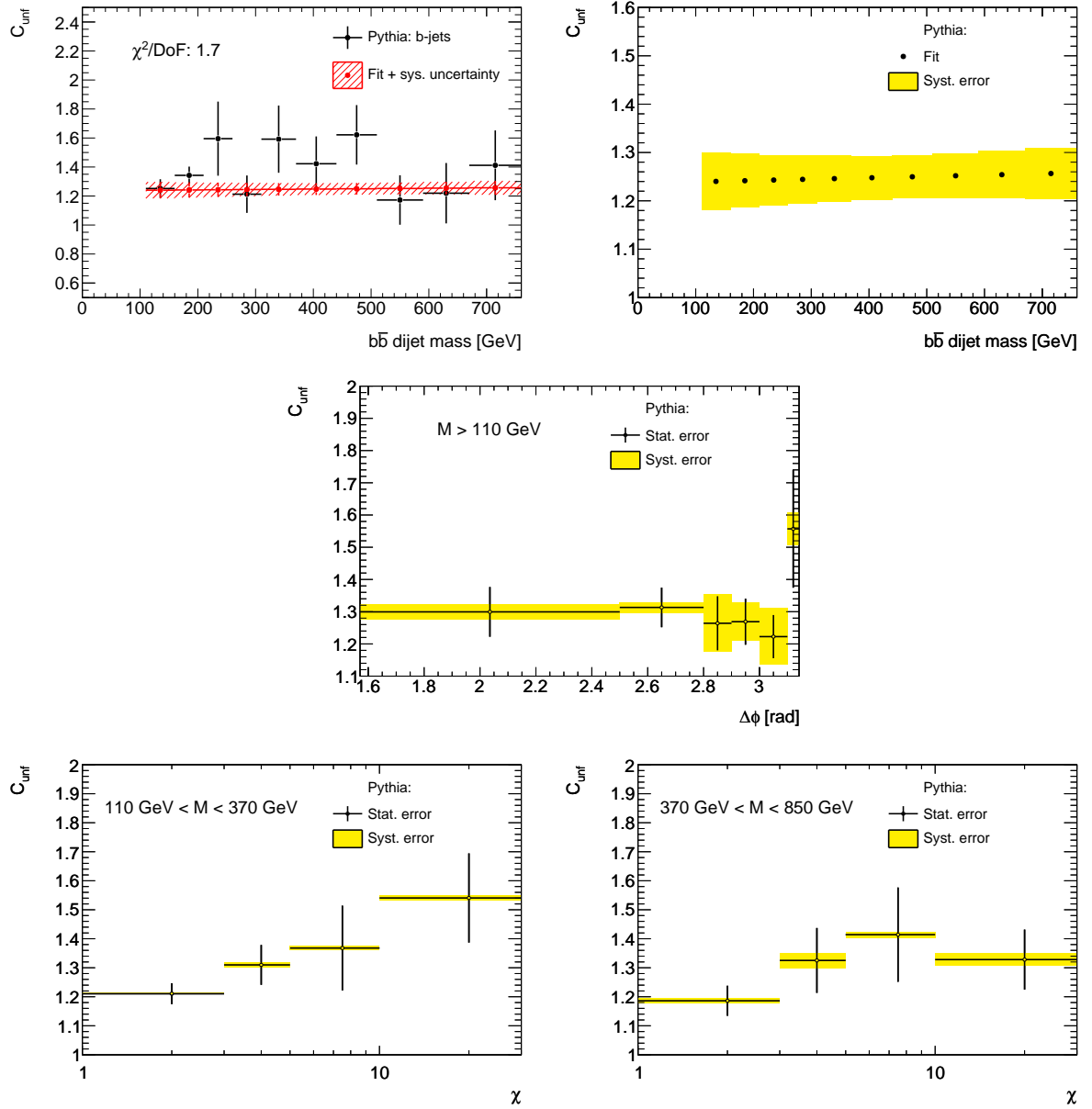
Figure 9.20: *The unfolding correction factors C_{unf} necessary to correct the measured $b\bar{b}$ cross sections to particle level with the corresponding systematic uncertainties: (top left) the fit of the unfolding correction as a function of dijet mass with a systematic uncertainty from the 95% confidence-level, (top right) the resulting unfolding correction as a function of dijet mass with the total systematic uncertainty, (middle) the unfolding correction as a function of $\Delta\phi$, (bottom) the unfolding correction as a function of χ for the lower mass region 110 GeV < M < 370 GeV (left) and the higher mass region 370 GeV < M < 850 GeV (right).*

9.2.7 Systematics from pile-up and b-tagging algorithm

The data set in use for this cross section measurement consists of mainly three different run blocks. With start of the data taking, each beam crossing contained on average 0.15 interactions until the number of protons per bunches increased and the size of the bunches decreased. With periods E and F, the average number of interactions increased to 1.5. At the end of the data taking (periods G to I) the instantaneous luminosity increased to $2.07 \cdot 10^{32}$ cm^{-2} s^{-1} and the number of interactions per beam crossing was on average 3. In Section 5.1.3 it was discussed how the bias in the jet reconstruction due to pile-up is treated and corrected. The jet energy scale uncertainty accounts for differences and uncertainties in the pile-up correction. So the jet energy measurement was adapted to cope with the pile-up conditions at the LHC and the nominal simulation samples are without pile-up. Further studies on the inclusive jet cross sections in [119] have shown that the cross section does not depend on the number of reconstructed vertices in the events which is a direct measure of the additional interactions. This section discusses possible biases from pile-up on the b-tagging identification and secondary vertex properties which are used to deduce the b-tagging purity. Finally, the independence of the measured cross section from the choice of the b-tagging algorithm is shown.

The identification of b-jets in the spatial algorithms is based directly on the reconstruction of tracks and consequently depend on primary vertices. With additional interactions a larger number of particles are emerging simultaneously from the beam crossing and a higher density of tracks and jets is created in the detector. In consequence, the distances between objects are reduced and in the extreme case the objects overlap and affect the reconstruction of each other. Therefore tracks of nearby objects may influence and even facilitate the reconstruction of secondary vertices which are important for the b-tagging algorithm used. This could have an effect directly on the b-tagging efficiency and mistag rate. Indirectly the track multiplicities in the secondary vertex could additionally change the invariant mass distribution of the secondary vertices. Both issues are studied by investigating their behaviour as a function of instantaneous luminosity in data and by comparing simulation samples with two different pile-up scenarios: the default PYTHIA sample and a pile-up sample with the configuration of data periods G to I (cf. Section 4.4). It can be seen in Fig. 9.21(a) that within the statistical uncertainties the b-tagging efficiency determined in the pile-up and in the nominal PYTHIA sample are identical. The presence of additional jets reconstructed in the event does not alter the identification power of the b-tagging algorithm in the pile-up scenario in ATLAS 2010 data. This assumption is verified by Fig. 9.21(b) measured from data. The average number of b-tagged jets with respect to the total number of reconstructed jets is determined as a function of instantaneous luminosity. It is distinguished between data periods A to D with instantaneous luminosities below 10^{31} cm^{-2}s^{-1} and data periods E to I with instantaneous luminosities between 10^{31} cm^{-2}s^{-1} and 10^{33} cm^{-2}s^{-1}. The latter distribution is shifted by 0.2 on the vertical scale. The average number $\langle n \rangle$ as a function of instantaneous luminosity is fitted by a constant function. The matching fit results confirm that the average number of b-tagged jets with respect to the total number of reconstructed jets does not depend on the amount of pile-up in the data period.

With the help of Fig. 9.22 two properties of the reconstructed secondary vertex in data as a function of instantaneous luminosity are studied. The shapes of the invariant mass of the charged particles associated to the secondary vertex in Fig. 9.22(a) for any dijet mass are compared in the data periods A to D ($L < 10^{30}$ cm^{-2}s^{-1}), E to F
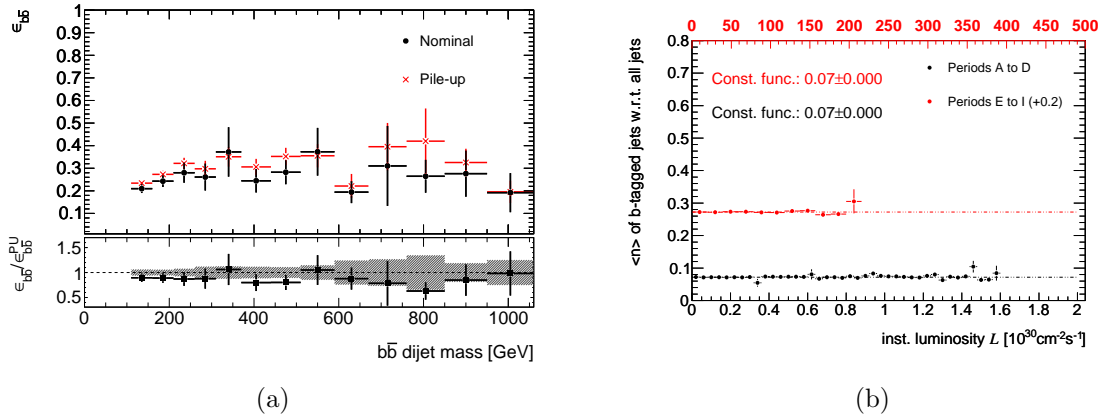
Figure 9.21: *(a) The b-tagging efficiency $\epsilon_b\bar{b}$ deduced from different simulation samples using different pile-up scenarios. (b) The fraction of b-tagged jets with respect to all jets in data as a function of instantaneous luminosity from periods A to I. The data points for periods E to I have an offset of 0.2 for visibility.*

$(10^{30}$ cm^{-2}s$^{-1} < L < 10^{31}$ cm^{-2}s$^{-1})$ and G to I $(10^{31}$ cm^{-2}s$^{-1} < L < 10^{33}$ cm^{-2}s$^{-1})$ with different instantaneous luminosities. The overlaid ratio with respect to the data period without pile-up shows that the shape is unchanged with respect to the amount of simultaneous interactions. Within the reconstruction step of the secondary vertex two-track vertices are built. Under the influence of nearby and overlapping tracks the number of two-track vertices could change depending on the amount of pile-up interactions. In Fig. 9.22(b) the average number of two-track vertices is determined as a function of instantaneous luminosities and fitted by a constant function. The average number of two-track vertices increases slightly from periods A to D to periods E to I which are shifted by 0.2 on the vertical scale. The increase is not significant because the deviation of 0.01 between the fit results is covered by fit systematics due to the choice of the binning. In summary, within the amount of pile-up in data 2010 the b-tagging identification is not significantly biased.

The independence of the cross section measurement from the pile-up can also be seen in the comparison between the cross section measurement [117] using 3 pb^{-1} and the measurement presented in this analysis. The ratio between these cross section measurements presented in Fig. 9.23(a) shows the agreement between them within the systematic uncertainties. The simulation samples with MC09 contain several deficiencies in the description of data [95]. The deficiencies were correctly treated and the observed discrepancies between the cross section measurements are partly due to the underlying tune and detector description used. Finally, the cross section as a function of dijet mass deduced with the SV0 b-tagging algorithm is compared to the measurement with the JetProb algorithm. The main event selection steps are identical except the b-tagging selection cut. The b-tagging efficiency and purity are deduced solely on simulated events. A b-tagged jet from the JetProb algorithm does not necessarily contain a reconstructed secondary vertex. The assigned SV0 mass to such jets is zero. Therefore the SV0 mass distribution of b-tagged jets from the JetProb algorithm is strongly influenced by tracking efficiencies, vertex reconstruction efficiencies and systematic uncertainties from the tracking selection cuts. It is not possible to use a data-driven method based on the invariant mass of the secondary vertex. Nevertheless the

(a) (b)
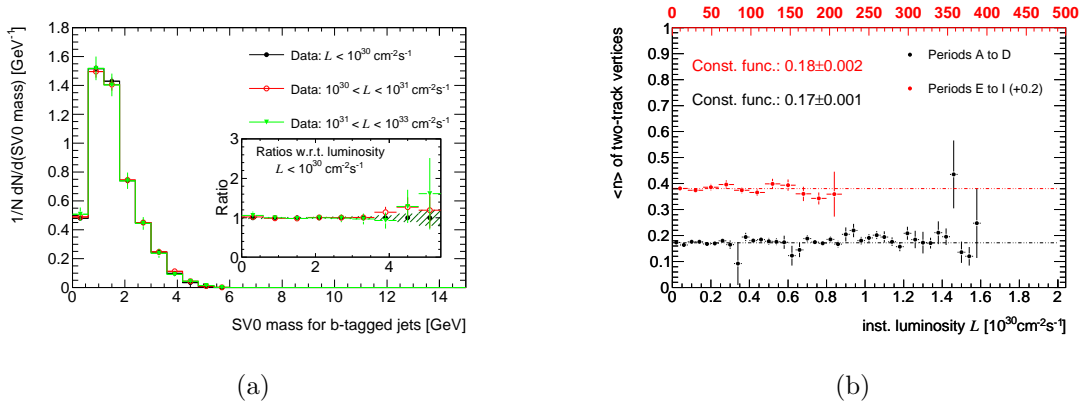
Figure 9.22: *(a) The SV0 mass distribution for any dijet mass separately measured in 3 different data periods with small (periods A to D), medium (periods E and F) and high (periods G to I) instantaneous luminosity. The ratio of the SV0 mass distribution, which is presented in the inlay, is calculated with respect to (w.r.t) the instantaneous luminosities below 10^{31} $cm^{-2}s^{-1}$. (b) The average number of reconstructed two-track vertices as a function of instantaneous luminosity from periods A to I. The data points for periods E to I have an offset of 0.2 for visibility.*

sample of *b*-tagged jets between both *b*-tagging selections are strongly correlated (more than 50%). To calculate a representative statistical error the dijet events selected each are taken to be 50% correlated. The unfolding correction is identical in both measurements. The resulting comparison between both measurements is presented in Fig. 9.23(b). Within the systematic uncertainty of the SV0 based measurement both cross sections agree. According to this the choice of the *b*-tagging algorithm does not affect the measured cross section.

(a) (b)
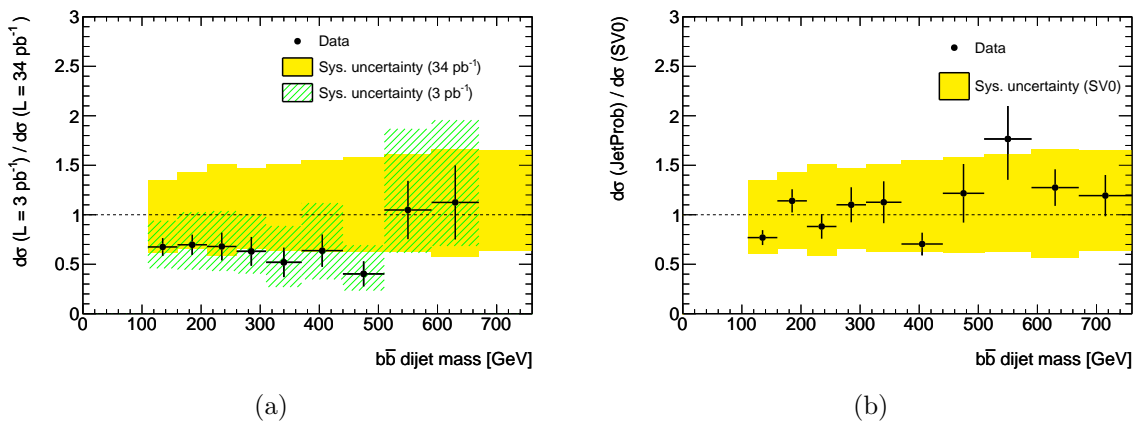
Figure 9.23: *(a) The comparison between the cross section measurement as a function of dijet mass with 3 pb^{-1} (data periods A to F) [117] and the full 34 pb^{-1} ATLAS data from 2010.*

9.2.8 The cross section ratio measurement

This section describes the measurement of the inclusive cross section briefly as the most important aspects and approaches have been discussed in case of the $b\bar{b}$ cross section and are completely analogous. The systematic uncertainties on the inclusive cross section and the determination of the systematic uncertainties on the ratio measurement are presented. The trigger strategy, event selection and luminosity used are identical. Hence their systematic uncertainties cancel out in the cross section ratio. The jet cleaning for the inclusive cross section is based on the medium selection (cf. Section 7.2) in contrast to the loose selection used in case of b-jets. The medium selection makes it necessary to correct for the inefficiency of the selection because the inefficiency is larger than 1%. The cleaning efficiency as a function of jet p_{T} and η has to be measured using data because some cleaning variables are not reproduced in simulation. The efficiency for identifying real jets is measured with a tag-and-probe method [121]. With the help of well-balanced, back-to-back dijet events similar to the ones used in Section 8.3.2 in which the tag jet is within $|\eta| < 2.0$ and fulfils the jet quality criteria, the fraction of probe jets being rejected by the jet quality cuts is measured. A dijet event is well-balanced if $(p_{\mathrm{T},1} - p_{\mathrm{T},2})/p_{\mathrm{T}}^{\mathrm{avg}} < 0.4$. The event selection ensures to select signal events without jets from background sources or noise bursts. The systematic uncertainty on this efficiency is obtained by varying the selection criteria on the tag jet. The jet selection inefficiency is below 4% for jets considered in the measurement and decreases with p_{T} at smaller than 1% for $p_{\mathrm{T}} > 100$ GeV in all rapidity regions. This inefficiency as a function of jet p_{T} is applied to simulated events. For each jet entering the dijet mass measurement a random numbers between 0 and 1 is generated. If the random number in an event exceeds the corresponding efficiency, the event is discarded and the fraction of events removed is measured. In Fig. 9.24 the cleaning efficiency and its systematic uncertainty as a function of dijet mass for inclusive jets is presented. In Table I.8 details of this efficiency are given. The inefficiency in the dijet cross section measurement is at most 4% at low dijet mass and decreases to less than 1% at 300 GeV. The systematic uncertainty is no more than 1% for all dijet mass bins. The dijet cross sections are corrected for these inefficiencies according to Eq. 9.4 and the systematic uncertainty on the efficiency is taken as a systematic uncertainty on the cross section.

The unfolding factors are calculated in analogy bin-by-bin from PYTHIA simulation and the same systematic uncertainties are evaluated as mentioned in Section 9.2.6 except for the fit systematics. The unfolding correction factors for inclusive jets are not fitted. The resulting unfolding correction with its systematic uncertainty can be seen in Fig. 9.25. The unfolding correction for inclusive dijets differs from the one on $b\bar{b}$ dijets by more than 20%. Apart from the correction due to the finite detector resolution the energy losses from neutrinos and muons seem to account for the additional and dominant 20% of the correction. The unfolding factors for inclusive jets are in agreement with unity for the dijet mass range measured. The corrections and the systematic uncertainties are summarised in detail in Table I.9. The systematic uncertainty ranges from 1.8% to 2.3%. The dominant contribution to the uncertainty is the uncertainty on the cross section shape and adds about 2% whereas the resolution uncertainty accounts for approximately 1%.

The uncertainty arising from the jet energy scale is determined in analogy to the approach discussed in Section 9.2.5. The applied uncertainties are presented coarsely in Table 8.2. It is expected that the uncertainty on the inclusive measurement compared to the $b\bar{b}$ one is slightly smaller because the contribution to account for the b-jet modelling is unnecessary.
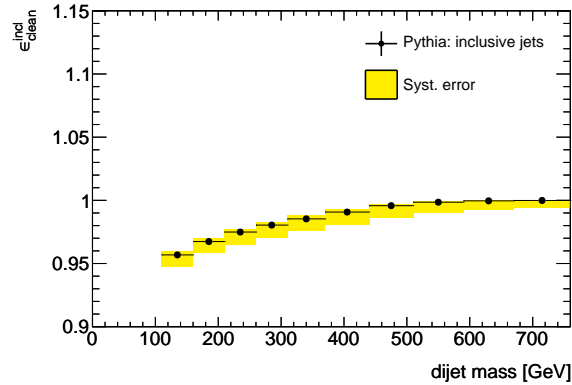
Figure 9.24: *The efficiency for jet identification as a function of dijet mass for inclusive jets and the medium cleaning selection. The black circles indicate the efficiency measured using the tag-and-probe method. The systematic uncertainty on the efficiency is obtained by varying the tag jet selection.*
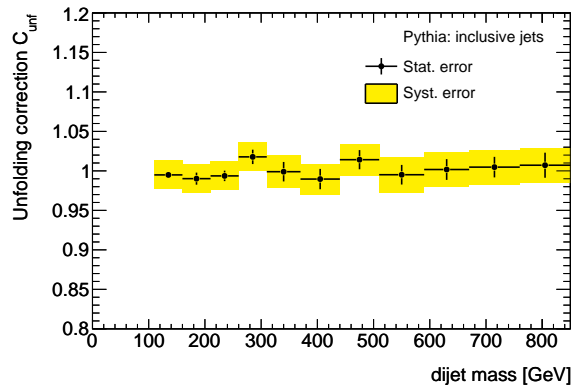
Figure 9.25: *The unfolding correction factors necessary to correct the measured dijet cross section as a function of dijet mass to particle level with the corresponding systematic uncertainty arising from the uncertainty on the jet energy resolution and cross section shape.*

The resulting systematic uncertainties on the inclusive dijet cross section measurement as a function of dijet mass are given in Table 9.6. The systematic uncertainties range from about 11.5% to 14% in the dijet mass range measured.

At this point all the necessary correction factors to measure the inclusive dijet cross section as a function of dijet mass have been presented. The cross section measurement is done according to Eq. 9.4. The systematic uncertainties discussed are added in quadrature and arise from the luminosity, the jet energy scale, the unfolding and the jet cleaning. The cross section ratio in Eq. 9.5 can be simplified by neglecting corrections which were found to be identical between both measurements or identical to unity. The cross section ratio is then given by:

$$\frac{d\sigma_{b\bar{b}}/dM}{d\sigma_{\mathrm{incl}}/dM} = \left.\frac{P_{b\bar{b}} \cdot C_{\mathrm{unf}}^{b\bar{b}} \cdot \epsilon_{\mathrm{clean}}^{\mathrm{incl}}}{C_{\mathrm{unf}}^{\mathrm{incl}} \cdot \epsilon_{b\bar{b}}}\right|_{M} \frac{N_{b\bar{b}}(M)}{N_{\mathrm{incl}}(M)}, \tag{9.10}$$

Dijet mass bin [GeV]	Fractional uncertainty [%]
110-160	$^{+14}_{-12}$
160-210	$^{+14}_{-13}$
210-260	$^{+14}_{-13}$
260-310	$^{+14}_{-13}$
310-370	$^{+16}_{-13}$
370-440	$^{+13}_{-13}$
440-510	$^{+13}_{-12}$
510-590	$^{+14}_{-12}$
590-670	$^{+14}_{-13}$
670-760	$^{+13}_{-12}$

Table 9.6: *Jet energy scale systematic uncertainty as a function of dijet mass for inclusive jets.*

The calculation of the systematic uncertainty on the cross section ratio follows these assumptions:

- The systematic uncertainty arising from the jet cleaning on inclusive jets and from the b-tagging efficiency and purity on the $b\bar{b}$ measurement are uncorrelated. These are handled by the error propagation of uncorrelated uncertainties and added in quadrature.

- To first order the systematic uncertainties on the unfolding correction from the cross section shape and resolution smearing are fully correlated with the exception of the fitting procedure for the $b\bar{b}$ dijets. This part of the systematic error is uncorrelated and considered by the error propagation of uncorrelated uncertainties and added in quadrature to the other systematic uncertainties. The uncertainty from the fitting procedure contributes about 4%.

- The systematic uncertainties from the jet energy scale (JES) are correlated. The uncertainty on the cross section ratio is evaluated by pseudo-experiments. The jet energy scale uncertainty for b-jets is the quadratic sum of the JES uncertainty on inclusive jets and the heavy flavor component. So the contribution based on inclusive jets is fully correlated and the other is uncorrelated. Two Gaussian distributions are generated with the width of the jet energy scale uncertainties: the relative inclusive jet energy scale uncertainty and the relative b-jet energy scale uncertainty of 2.5%. Each pseudo-experiment generates a Gaussian distributed random number w_1 which describes the inclusive cross section jet energy scale variation and another Gaussian distributed random number w_2 which describes the variation due to the 2.5% additional contribution for b-jets. The ratio measurement symbolised by $\frac{A}{B}$ is then varied by $\frac{A \cdot (1+(w_1+w_2))}{B \cdot (1+w_1)}$. Several million pseudo-experiments are generated. An example of the variations of the central value by the pseudo-experiments is presented in Fig. 9.26 for the asymmetric jet energy scale uncertainties. The outcome of

the pseudo-experiments are finally fitted by a Gaussian distribution and the 68% confidence-level is assigned as the jet energy scale uncertainty on the cross section ratio. The jet energy scale uncertainty on the ratio is of the order of 10% to 15% and recorded in detail in Table 9.7.
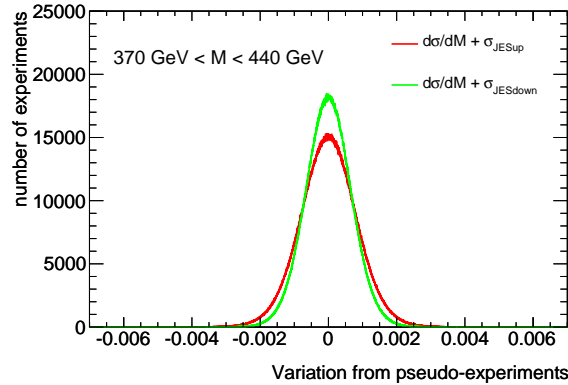
Figure 9.26: *Pseudo-Experiment variations of the jet energy scale according to two Gaussian distributions in the dijet mass bin 370 GeV < M < 440 GeV.*

Dijet mass bin [GeV]	Fractional uncertainty [%]
110-160	$^{+10}_{-9}$
160-210	$^{+10}_{-10}$
210-260	$^{+10}_{-10}$
260-310	$^{+11}_{-11}$
310-370	$^{+13}_{-11}$
370-440	$^{+11}_{-11}$
440-510	$^{+12}_{-11}$
510-590	$^{+13}_{-11}$
590-670	$^{+14}_{-13}$
670-760	$^{+13}_{-12}$

Table 9.7: *Jet energy scale systematic uncertainty as a function of dijet mass for the ration measurement between inclusive jets and b-jets.*

9.3 Results of the cross section measurements

The cross section measurements using the ATLAS detector and recorded data from 2010 which correspond to a total integrated luminosity of 34 pb^{-1} are presented. The previous sections supply the necessary correction factors and have discussed the systematic uncertainties arising in the measurements:
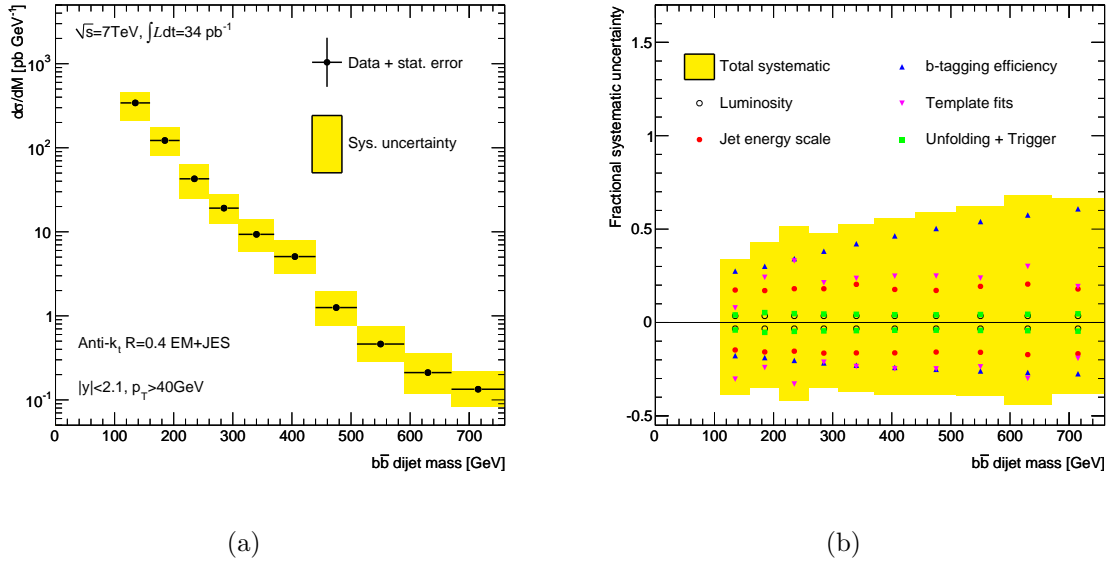
Figure 9.27: *Differential cross section measurement of $b\bar{b}$ dijets as a function of dijet mass (a) with the ATLAS detector and data corresponding to an integrated luminosity of 34 pb^{-1}. The fractional systematic uncertainty is presented as a function of dijet mass (b). The shaded area shows the combined uncertainty and the magnitude of the individual components are presented by the markers.*

- The uncertainty on the luminosity.

- The uncertainty on the trigger selection.

- The uncertainty on the jet energy scale.

- The uncertainty on the b-tagging efficiency.

- The uncertainty on the unfolding correction.

Among those corrections the unfolding procedure has been presented which corrects all cross sections to hadron level. The cross sections are measured with jets with transverse momenta above 40 GeV and within the rapidity $|y| < 2.1$. The cross section is measured only above a minimum dijet mass not biased by residual kinematic effects. The systematic uncertainties presented in Section 9.2 are all added in quadrature.

For the statistical error calculation on the unfolded cross sections the statistical error from the trigger efficiency and from the event selection efficiency are neglected due to their small contribution. For all cross section measurements on b-jets the statistical uncertainty on the b-tagging purity and calibration are incorporated into the systematic uncertainty. A jet cleaning in the $b\bar{b}$ measurement is unnecessary therefore no statistical error needs to be accounted for. Furthermore for the $b\bar{b}$ cross section measurement as a function of dijet mass the statistical uncertainties are treated differently compared to the other measurements as discussed in the previous section. The statistical uncertainty on the b-tagging efficiency as well as on the unfolding correction as functions of dijet mass are transferred into the systematic uncertainty by the fitting procedure. For the measurement as a function of $\Delta\phi$ and χ the statistical uncertainties on the b-tagging efficiency and

unfolding correction are calculated using error propagation and the correlations between the number of events. The contributions from the statistical uncertainties are propagated to the cross section measurement:

$$\delta\left(\frac{d\sigma_{b\bar{b}}}{d\Delta\phi \text{ or } d\chi}\right) = \frac{d\sigma_{b\bar{b}}}{d\Delta\phi \text{ or } d\chi} \cdot \sqrt{\left(\frac{\delta N}{N}\right)^2 + \left(\frac{\delta C_{\text{unf}}^{b\bar{b}}}{C_{\text{unf}}^{b\bar{b}}}\right)^2 + \left(\frac{\delta\epsilon_{b\bar{b}}}{\epsilon_{b\bar{b}}}\right)^2}, \qquad (9.11)$$

where $\delta N/N$ is the fractional error on the number of events, $\delta C_{\text{unf}}^{b\bar{b}}/C_{\text{unf}}^{b\bar{b}}$ is the fractional error on the unfolding corrections and $\delta\epsilon_{b\bar{b}}/\epsilon_{b\bar{b}}$ the one on the b-tagging efficiency.

For the inclusive cross section measurement as a function of dijet mass the statistical contributions are as follows:

$$\delta\left(\frac{d\sigma_{\text{incl}}}{dM}\right) = \frac{d\sigma_{\text{incl}}}{dM} \cdot \sqrt{\left(\frac{\delta N}{N}\right)^2 + \left(\frac{\delta C_{\text{unf}}^{\text{incl}}}{C_{\text{unf}}^{\text{incl}}}\right)^2 + \left(\frac{\delta\epsilon_{\text{clean}}^{\text{incl}}}{\epsilon_{\text{clean}}^{\text{incl}}}\right)^2}, \qquad (9.12)$$

where $\delta N/N$ is as defined above, $\delta C_{\text{unf}}^{\text{incl}}/C_{\text{unf}}^{\text{incl}}$ is the fractional error on the unfolding corrections and $\delta\epsilon_{\text{clean}}^{\text{incl}}/\epsilon_{\text{clean}}^{\text{incl}}$ the one on the jet identification efficiency.

For the ratio measurement the statistical uncertainty consists on the one hand of the terms from the error propagation from the statistical error on the unfolding corrections and on the jet identification efficiency for inclusive jets and on the other hand from the statistical error on the number of events from the $b\bar{b}$ measurement. As the $b\bar{b}$ dijets are a real subset of the inclusive jets entering in the cross section measurement the event samples are partly correlated. But from the raw event numbers the correlation is determined to be of the order of per mill and has a negligible effect on the statistical error. Therefore the statistical error from the counting statistics is calculated for uncorrelated events overestimating the uncertainty but beyond the precision of the measurement.

The measurements as a function of $\Delta\phi$ and χ are normalised by the cross section, integrated over the same phase space. This minimises experimental uncertainties. Therefore the uncertainty arising from the luminosity cancels out in these measurements and the normalisation reduces the uncertainties due to the jet energy scale, unfolding correction and jet cleaning uncertainty, but the measurement stays sensitive to differences in the shapes of the cross sections.

Cross section as a function of dijet mass: The cross section measurement as a function dijet mass is presented in Fig. 9.27 together with the fractional systematic uncertainty on the cross section. The cross sections spans over almost four orders of magnitude and is measured up to a dijet mass of 760 GeV. The cross section as a function of dijet mass is a steeply falling function as expected from QCD. The cross section measured at around 400 GeV seems to stand out. This exceptional behaviour is already present in the raw data spectra for each trigger. In Chapter 12 this disagreement is studied and quantified. With only 34 pb^{-1} of integrated luminosity this measurement extends the dijet mass cross section from Tevatron with an integrated luminosity of 260 pb^{-1} by more than 300 GeV (cf. Section 3.2). The inclusive cross section of b-jets as a function of transverse momentum has been measured and is presented along with the dijet cross section in [118]. The total systematic uncertainty ranges from 30% to 60%. On average the largest contributions are the b-tagging related systematics. The b-tagging calibration driving the b-tagging efficiency uncertainty causes the dominant contribution above 250 GeV. Because the b-tagging calibration has to be extended to measure the cross section of b-jets in this kinematic
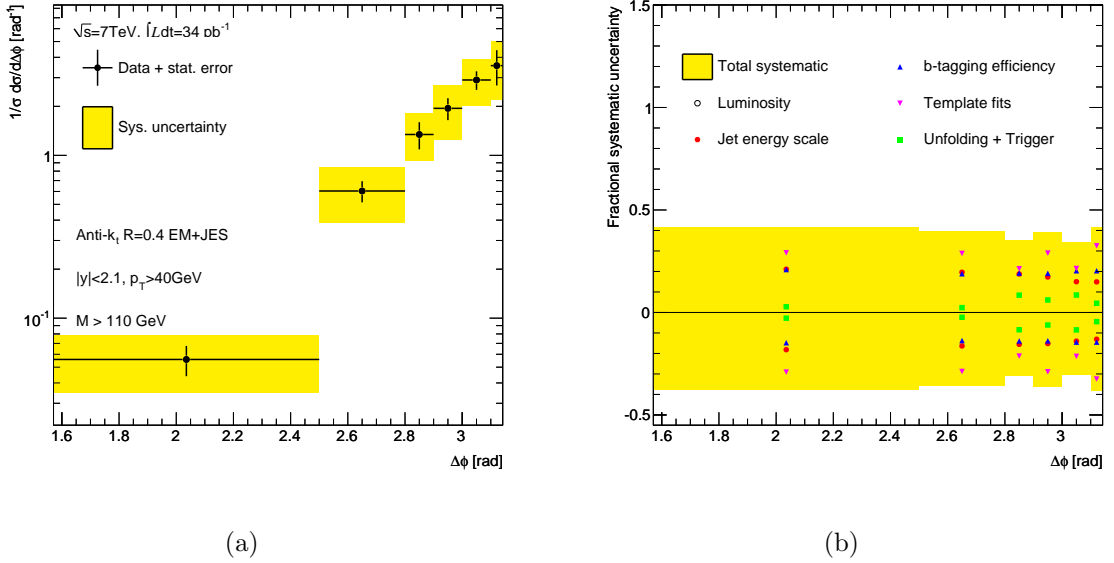
Figure 9.28: *Differential cross section measurement of $b\bar{b}$ dijets as a function of dijet azimuthal angle (a) with the ATLAS detector and data corresponding to an integrated luminosity of 34 pb^{-1}. The fractional systematic uncertainty is presented as a function of $\Delta\phi$(b). The shaded area shows the combined uncertainty and the magnitude of the individual components are presented by the markers.*

region the uncertainty was conservatively estimated. The statistical uncertainties on the cross section measurement as a function of dijet mass range from about 2% to 8%. The details on the cross section measurement are written down in Table J.1.

Cross section as a function of $\Delta\phi$: The cross section as a function of $\Delta\phi = |\phi_1 - \phi_2|$ normalised to the total cross section measured in this phase space is presented in Fig. 9.28 along with the fractional systematic uncertainty. By normalising the cross section uncertainties are reduced and only the shape of the cross section is determined. The cross section falls rapidly over two orders of magnitude with decreasing angular separation $\Delta\phi$.The fractional systematic uncertainty as a function of $\Delta\phi$ shows a homogeneous picture as almost all uncertainties are approximately constant over the measured phase space. Due to the mixing of different jet p_T values in each $\Delta\phi$ bin the systematic uncertainties represent an average. Especially, the b-tagging efficiency uncertainty is constant and adds only 20% uncertainty. The systematic uncertainty on the measurement amounts to 40% over the measured phase space. The dominant uncertainty in this measurement is the uncertainty on the b-tagging purity from the template fits.

The details on the cross section measurement are written down in Table J.2. The statistical uncertainties on the cross section measurement range from about 16% to 25%. The increase of the statistical error in contrast to the measurement as a function of dijet mass is due to the fact that neither the b-tagging efficiency nor the unfolding correction are fitted. In case of the dijet cross section as a function of dijet mass, the statistical uncertainties are taken into account by the fitting uncertainties and inserted into the systematics.

Cross section as a function of χ: The cross sections as a function of the angular variable $\chi = \exp|y_1 - y_2|$ normalised to the corresponding total cross section are presented in Fig. 9.29 and Fig. 9.30. The cross sections are plotted until $\chi < 30$ and in this region
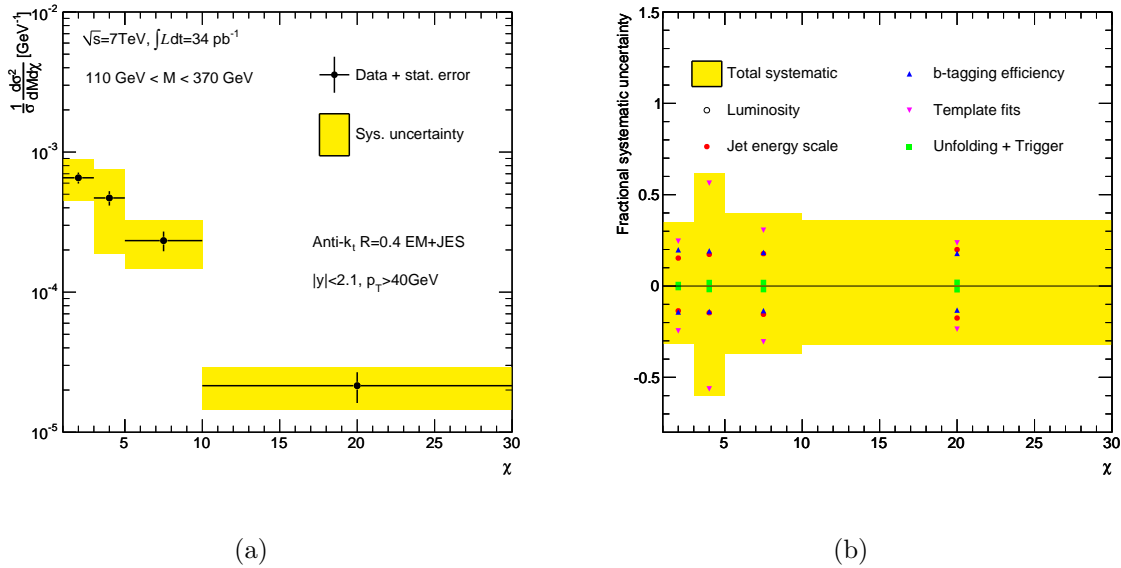
Figure 9.29: *Double differential cross section measurement of $b\bar{b}$ dijets as a function of angular variable χ in the mass region 110 GeV < M < 370 GeV (a) with the ATLAS detector and data corresponding to an integrated luminosity of 34 pb^{-1}. The fractional systematic uncertainty is presented as a function of chi (b). The shaded area shows the combined uncertainty and the magnitude of the individual components are presented by the markers.*

the cross sections decrease with increasing χ. A side effect of the chosen binning in χ which increases exponentially with increasing χ from the point of statistics in the purity templates is that migration effects are a minor issue. The magnitude of the systematic uncertainty depends on the mass region in which the cross section as a function of χ is measured. In the mass region 110 GeV < M < 370 GeV the uncertainties tend to be smaller. They range from 30% to 60%. In the second χ bin the non-b template suffers from statistical problems. In the higher mass region the b-tagging calibration uncertainty and the template fit uncertainties dominate and add up to effects of more than 100% in the asymmetric systematic uncertainty. The systematic uncertainty in the higher mass region goes down to a minimum of 40%. The details on the cross section measurement are written down in Table J.3 and Table J.4. The statistical uncertainties on the cross section measurement as a function of angular variable χ range from about 10% to 25%. Also here the statistical uncertainties from the simulation on the b-tagging efficiency and unfolding correction are accounted for in the statistical error on the measurement.

Inclusive cross section and ratio measurement as a function of dijet mass: Finally, the inclusive cross section measurement followed by the ratio measurement are presented. The measurement of the inclusive cross section as a function of dijet mass is presented in Fig. 9.31. The measurement is preformed in the same phase space region 110 GeV < M < 760 GeV which is used in the $b\bar{b}$ measurement. The inclusive cross section falls also almost four orders of magnitude. In this measurement the jet energy scale uncertainty is the dominating uncertainty and contributes about 12% to 15%. The next significant uncertainty is the uncertainty on the luminosity determination. The remaining uncertainties add about 2% to the total uncertainty. The statistical error on the measurement is maximum about 2% at around 700 GeV and below per mill level at low
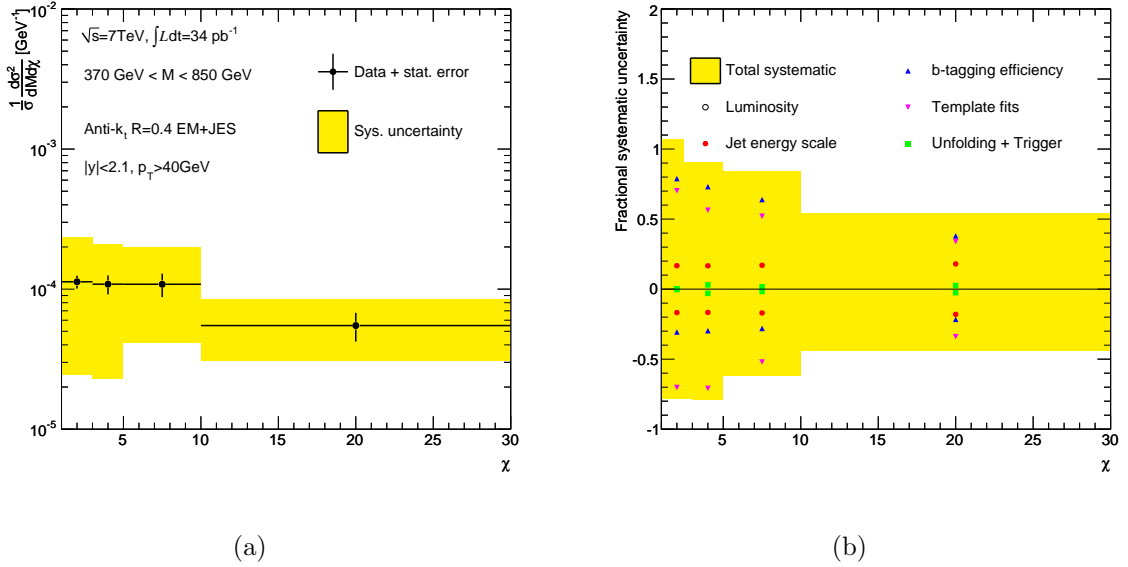
Figure 9.30: *Double differential cross section measurement of $b\bar{b}$ dijets as a function of angular variable χ in the mass region 370 GeV < M < 850 GeV (a) with the ATLAS detector and data corresponding to an integrated luminosity of 34 pb^{-1}. The fractional systematic uncertainty is presented as a function of χ (b). The shaded area shows the combined uncertainty and the magnitude of the individual components are presented by the markers.*

dijet masses. The details on the cross section measurement are written down in Table J.5. The inclusive cross section measurement is used to determine the ratio to the $b\bar{b}$ dijet cross section. The ratio measurement is presented in Fig. 9.32 as a function of dijet mass. The fraction of $b\bar{b}$ dijets decreases slightly from about 1% to 0.3% with increasing dijet mass. Also here, the ratio measured at around 400 GeV stands out in the ratio caused by the $b\bar{b}$ dijet measurement. The statistical uncertainty ranges from about 2% at lower dijet mass to about 9%. The total systematic uncertainty on the ratio measurement is about 30% to 60%. The systematic uncertainties not related to b-tagging are smaller than 13%. The jet energy scale uncertainty is reduced by the correlation between both jet measurements. The remaining part of the uncertainty is contributed by the b-tagging efficiency and purity. At high dijet masses the b-tagging efficiency is the dominant systematic uncertainty.

The magnitudes of the systematic uncertainties in the measurements are summarised in Table 9.8
The dominating systematic uncertainties are related to the b-tagging performance. With the magnitude of data recorded in 2011 by the ATLAS detector, the b-tagging calibration can be improved with the help of the sample of $t\bar{t}$ events. There are three methods to the measure the b-tagging efficiency in a sample of $t\bar{t}$ events. Compared to the calibration in events with a semi-leptonic decay into a muon, these methods do not break down at transverse momenta of 200 GeV and inherently include the hadronic decays of the b-hadrons. The extension of the calibration from the semi-leptonic decay to all possible b-hadron decay modes contributes a large fraction to the systematic uncertainty. Additionally, the methods based on $t\bar{t}$ events have the advantages of a high-purity sample solely based on kinematic selections and the possibility to cross-check the calibration with the total
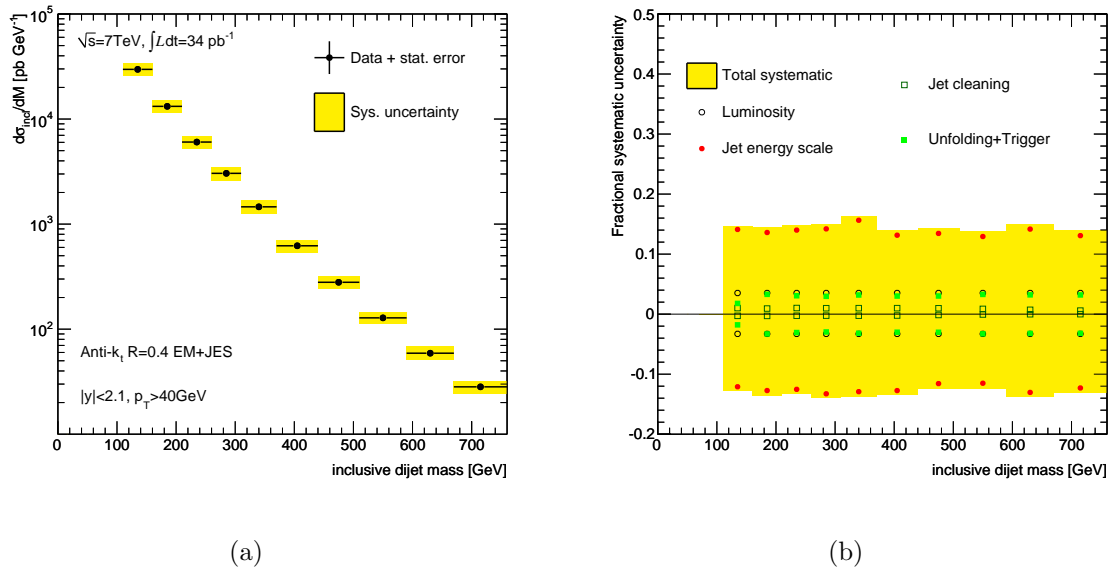
165

(a) (b)

Figure 9.31: *Differential cross section measurement of inclusive dijets as a function of dijet mass (a) with the ATLAS detector and data corresponding to an integrated luminosity of 34 pb^{-1}. The fractional systematic uncertainty is presented as a function of dijet mass (b). The shaded area shows the combined uncertainty and the magnitude of the individual components are presented by the markers.*

measured cross section of top pairs in ATLAS.

The usage of more advanced algorithms to identify b-jets is another step towards a precision measurement of b-jets. These algorithms in general make use of two-dimensional distributions and heavy-flavour properties. With equally good b-tagging efficiency an increase in light jet rejection of a factor of 3 can be achieved. These high-performance taggers will allow for searches of new phenomena with low production cross sections.

Uncertainties in relation to the purity determination can be reduced by improving the theoretical uncertainties on c-quark production and modelling as well as on bottom fragmentation. With the help of dedicated studies the b-tagging properties of charm-flavoured hadrons can be understood and better constrained in measurements.

The uncertainty determination on the luminosity in this early data taking period has been experimentally very challenging and the luminosity uncertainty has been measured as precisely as by accelerators which are operating for years.

The jet energy scale uncertainty on the cross section measurement contributes about 15% to 20%. This precision has been established using various techniques in the first year of proton-proton collisions. It is a long time project to reduce the jet energy scale significantly by the use of more sophisticated methods. With increasing integrated luminosity the calibrations of jets using data-driven methods like the transverse momentum balance in γ+jet and Z+jet events become the focus of the jet energy scale uncertainty determination. For the cross section ratio measurement it is also of use if the b-tagging related contributions in the systematic uncertainty decrease to take advantage of the correlation of experimental and theoretical systematic uncertainties in the measurement.

(a)   (b)
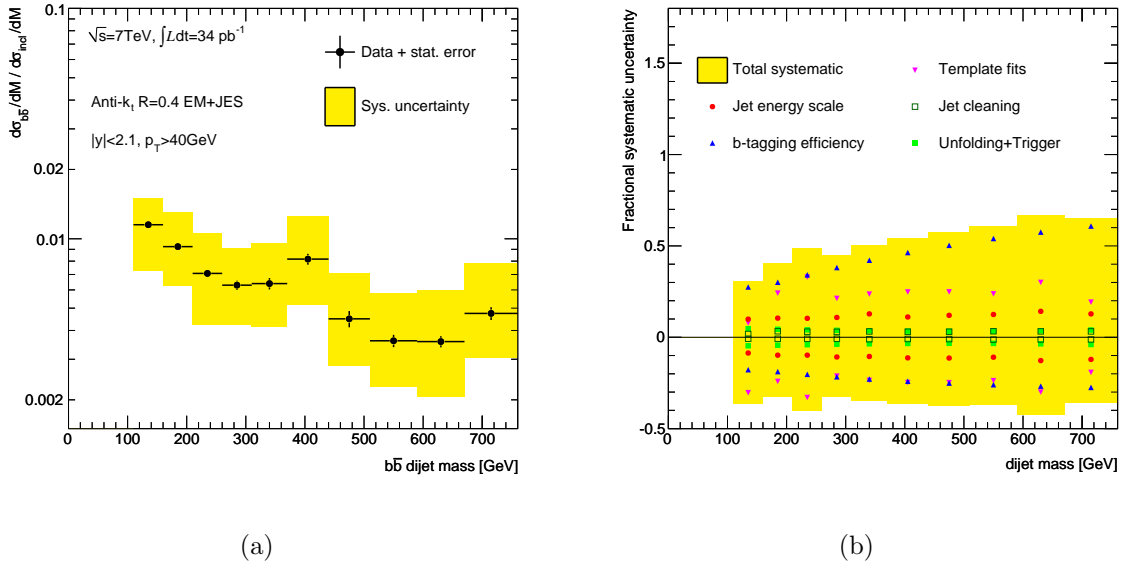
Figure 9.32: *Differential cross section ratio measurement between $b\bar{b}$ and inclusive dijets as a function of dijet mass (a) with the ATLAS detector and data corresponding to an integrated luminosity of 34 pb^{-1}. The fractional systematic uncertainty is presented as a function of dijet mass (b). The shaded area shows the combined uncertainty and the magnitude of the individual components are presented by the markers.*

Systematic uncertainty	Contribution to cross section			
	Dijet mass	Ratio	$\Delta\phi$	χ
Jet energy scale	15-20%	9-14%	10-20%	$\approx 20\%$
b-tagging efficiency	30-60%	30-60%	$\approx 20\%$	20-80%
b-tagging purity	20-30%	20-30%	20-30%	30-70%
Luminosity	3.4%	3.4%	3.4%	3.4%
Unfolding and Trigger	< 5%	< 5%	< 10%	< 5%
Jet cleaning	-	$\approx 1\%$	-	-

Table 9.8: *Summary of the important systematic uncertainties on the dijet cross section of b-jets using the SV0 b-tagging algorithm and of inclusive dijets.*

10 Next-to-leading order calculations

This chapter discusses briefly the methods used to perform NLO calculations and presents the NLO predictions which are compared to the measured cross sections presented in the previous chapter. Three different NLO generators are used for this purpose and explained in the following. Finally, the generators and their predictive power is compared and summarised.

10.1 Introduction

The leading-order predictions of parton shower generators like PYTHIA and HERWIG can in general describe the main features of partonic processes and give a first estimate of their cross sections. Divergences in QCD calculations are absorbed by introducing unphysical renormalisation and factorisation scales as discussed in Section 2.1.3. The inherent uncertainty in a lowest-order calculations due to the scale dependence is large. In contrast, calculations at higher orders show reduced uncertainties and in addition give control over the uncertainties due to the scale choice (cf. Eq. 3.6). A calculation to all orders, if possible, would be independent from the choice of these scales. The extension to next-to-leading order (if available) is the logical step. In addition, as shown in case of $b\bar{b}$ production (cf. Section 3.1.1), new partonic processes may contribute at NLO emphasising the need for higher order calculations.

The next-to-leading order dijet cross section can be written as:

$$\sigma = \sigma_{\mathrm{LO}} + \sigma_{\mathrm{NLO}}, \tag{10.1}$$

where the LO term contains the Born approximation for 2 partons in the final state. In Section 2.1.4 the NLO contributions have been discussed and in correspondence the NLO term in Eq. 10.1 is given by:

$$\sigma_{\mathrm{NLO}} \equiv \int d\sigma_{\mathrm{NLO}} = \int_{2+1 \text{ partons}} d\sigma_{\mathrm{R}} + \int_{2 \text{ partons}} d\sigma_{\mathrm{V}}, \tag{10.2}$$

where $d\sigma_{\mathrm{R}}$ denotes the real corrections from the 3-parton final states and $d\sigma_{\mathrm{V}}$ denotes the virtual corrections to the 2-parton final states. The renormalisation procedure has been carried out on the virtual correction to remove ultraviolet divergences. On their own both integrals are divergent (in four dimensions) due to low-momentum and small-angle regions, although their sum is finite. Among different schemes, which deal with these divergences to calculate next-to-leading order jet cross section, the subtraction scheme [122] is widely used and is implemented into the core of all three NLO generators used. The general idea behind the subtraction method is to add and subtract counterterms $d\sigma_{\mathrm{A}}$ which are a proper approximation of the real emission amplitude $d\sigma_{\mathrm{R}}$. The NLO contributions are rewritten as follows:

$$d\sigma_{\mathrm{NLO}} = [d\sigma_{\mathrm{R}} - d\sigma_{\mathrm{A}}] + d\sigma_{\mathrm{A}} + d\sigma_{\mathrm{V}}. \tag{10.3}$$

The counterterms need to have the same behaviour in each phase-space point as $d\sigma_\text{R}$ to act as a local counterterm. The first term in Eq. 10.3 is integrable in four dimensions and the remaining singularities are associated with the last two.[1] If the analytical integration of $d\sigma_\text{A}$ over the one-parton subspace leads to the same poles as in the integration of $d\sigma_\text{V}$, the counterterm can cancel the divergences of the virtual corrections. The final structure of the calculation is then given by:[2]

$$\sigma_\text{NLO} = \int_{2+1 \text{ partons}} [d\sigma_\text{R} - d\sigma_\text{A}] + \int_{2 \text{ partons}} \left[d\sigma_\text{V} + \int_{1 \text{ parton}} d\sigma_\text{A} \right]. \qquad (10.4)$$

The final challenging task is to obtain the actual form of the counterterm. The so-called dipole factors deliver a general construction prescription outlined in detail in Ref. [123]. In case of initial-state hadrons additional soft and collinear singularities in the final state are absorbed in the PDFs and give rise to extra counterterms.

This procedure is guaranteed to be valid for cross sections of hadronic observables as long as they are collinear and infrared safe which is obeyed by the anti-k_t jet algorithm.

10.2 The NLO generators and predictions

This section presents the configurations and generators which are used to obtain the NLO predictions. The generators themselves are described briefly. The resulting NLO cross sections are prepared in context of the measurements using the same event selection criteria from the previous chapters.

The NLO predictions which are compared to the measurement are obtained by these three generators: MCFM [124, 125], MC@NLO [56] and POWHEG [126, 127]. The MCFM generator is a NLO fixed-order generator which calculates the cross section at parton level. In addition to the MCFM prediction corrections are necessary, so-called non-perturbative corrections (non-perp.), to include the effects of multiple interactions, parton shower and hadronisation. The other two NLO generators MC@NLO and POWHEG combine the precision of NLO matrix element calculations with the treatment of higher orders, soft and collinear configurations via the parton shower approximation. The predictions of both generators are interfaced to a parton shower program which then finally produces the prediction at hadron level. On either the parton level from MCFM or the hadron-level from MC@NLO and POWHEG the anti-k_t jet algorithm reconstructs jets with the radius parameter $R = 0.4$. On these events the same phase space selection cuts are applied as on data: $p_\text{T} > 40$ GeV and $|y| < 2.1$. On both levels the b-jets are identified by the angular matching $\Delta R < 0.3$ to either b-quarks or b-hadrons ($p_\text{T}^b > 5$ GeV). The basis for the NLO calculation is the MSTW 2008 NLO PDF [12].[3] This PDF set contains the uncertainties on the PDFs evaluated at the 68% confidence-level. To evaluate the impact of the choice of a different PDF the CTEQ66 [129] PDF is used. The parton shower generator used to obtain the corrections for the parton-level prediction from MCFM and to generate the parton shower for the predictions from POWHEG is PYTHIA with the AMBT1 tune (cf. Section 2.3.1). MC@NLO connects to HERWIG to obtain the hadron-level prediction. The b-quark mass is set to $m_b = 4.95$ GeV for all computations. In these calculations the requirement of $p_\text{T} > 40$ GeV diminishes the dependence of the theoretical result on

---

[1]With dimensional regularisation in $d = 4 - 2\epsilon$ dimensions the singularities turn into poles in ϵ.

[2]In the limit of $\epsilon \to 0$ in context of dimensional regularisation.

[3]The PDF sets are provided via the LHAPDF interface [128].

the inserted b-quark mass. Furthermore, given the large mass of the b-quark and the phase space probed, effects from intrinsic transverse momenta of the partons in the hadron should be negligible.

The theoretical uncertainties on the NLO calculations arise from the experimental uncertainties on the measurements on which the PDFs are extracted from, from the uncertainties on α_s in the NLO calculation and in the PDF fits, from the factorisation and renormalisation scales and from the choice of the parton shower generator and tune to generate hadron-level predictions. These four uncertainties are in general added in quadrature, if not it is stated otherwise. The uncertainties on the total cross sections from each NLO generator are mostly below 0.5% and neglected.

The error propagation of experimental uncertainties to the PDF uncertainties is done using the Hessian method [12], which is based on linear error propagation from a set of eigenvectors. The MSTW 2008 PDF set includes 20 eigenvectors. These are evaluated in positive and negative direction around the central value. The uncertainty on the cross section is the quadratic sum of each variation in the positive and negative direction.

The renormalisation and factorisation scales are set to be equal, and are denoted by μ_0. The scale uncertainties are evaluated by varying the renormalisation or the factorisation scale μ_0 by a factor of 2 ($= 2\mu_0$ and $\frac{1}{2}\mu_0$). The systematic uncertainty arising from the scale variation are estimated by the maximum deviation of each cross section from the renormalisation or factorisation scale variations.

The uncertainty on the cross section from the uncertainty on α_s can be evaluated either directly with the help of PDF fits obtained with a varied α_s [130] value or in case of the MSTW PDFs in combination with the PDF uncertainties [131]. In the direct case the PDF uncertainty on the cross section is added in quadrature to the uncertainty from α_s. In case of MSTW the $PDF+\alpha_s$ uncertainty is extracted taking the correlations into account from the envelop of five PDF predictions of PDF sets and their eigenvector sets in which α_s is displaced by $\{-1\sigma, -\sigma/2, 0, +\sigma/2, +1\sigma\}$.

The uncertainty on the parton shower model and on the tune is evaluated by comparing the cross section predictions from different tunes and parton shower generators. The envelop of these predictions is taken as systematic uncertainty on the cross section.

The calculation of NLO predictions is CPU-intensive. Therefore the number of calculated NLO predictions are a compromise between CPU power and necessity. Theoretical uncertainties are rated proportional to the importance of their contribution to the prediction. The significant scale dependence of the NLO calculations for bottom production is due to large contributions from higher orders. In the high energy limit large corrections can occur in diagrams where the gluon exchanged in the t-channel becomes soft. At high transverse momentum of the produced b-quark the quark mass is negligible and the quark behaves like a massless particle. This introduces significant contributions of radiation processes of gluons. The processes FEX and GSP which appear at NLO are calculated without considering higher orders. The inclusion of such higher orders can reduce the scale dependence. In comparison to the scale uncertainties the PDF uncertainty on the $b\bar{b}$ cross section is marginal although the gluon parton density carries large uncertainties at large momentum fractions x. In light of the importance of the α_s uncertainties and the amount of CPU time needed it is considered to be impracticable to evaluate this uncertainty. The uncertainty induced by α_s is therefore neglected. The evaluation of the uncertainty from the non-perturbative corrections is performed in detail for the MCFM prediction. With the help of observations from the non-perturbative corrections for MCFM, this uncertainty on the other generators is estimated due to the limitation in computing power.

The following subsections go into details on the NLO generators used.

10.2.1 The MCFM Generator

MCFM (Monte Carlo for FeMtobarn processes) is a parton-level program providing NLO predictions for a range of processes. The final state for an NLO calculation at the matrix element level contains at most three partons. The anti-k_t algorithm (R = 0.4) forms jets from these partons. For the calculation of the $b\bar{b}$ cross sections the scale μ_0 is set to the scalar sum of the p_T's of all the partons in the event. The number of integrations necessary for sufficient statistics is about 10 million. To illustrate the effect of the inclusion of higher orders the MCFM program calculates the cross section at LO with a LO PDF and at NLO with a NLO PDF. The comparison between LO and NLO is presented as functions of dijet mass and χ in Fig. 10.1. Higher orders, generally speaking, reduces the cross section at small dijet masses and increases it at the high dijet mass. The overall effect of the higher orders on the cross section as a function of dijet mass is comparable to the ratio between the leading-order Pythia prediction and data in Fig. 8.1(b). Because of the use of LO* PDFs which incorporates some of the properties of NLO PDFs in Pythia the differences seen in Fig. 10.1 are larger. The higher orders seem to reshuffle the events as a function of χ to lower values of rapidity difference.

The parton-level predictions have to be complemented by the non-perturbative corrections
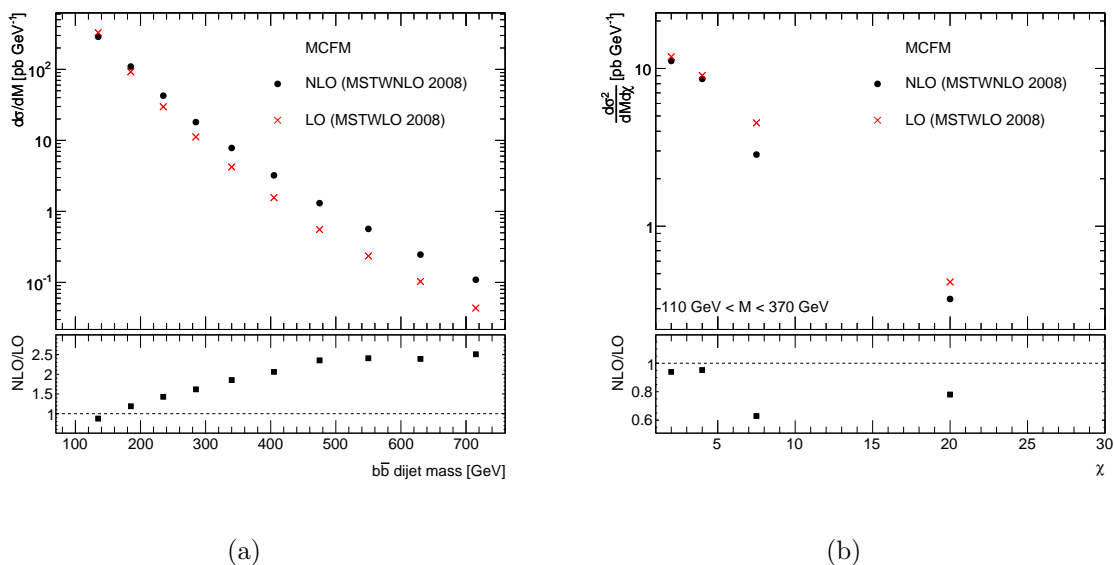
Figure 10.1: *The dijet cross section of b-jets at parton level calculated with* MCFM *at leading and next-to-leading order with the corresponding PDF set from MSTW 2008 (a) as a function of dijet mass and (b) as a function of the angular variable χ for 110 GeV < M < 370 GeV. The ratio of both predictions is illustrated in the lower part of the plot.*

which include multiple interactions, parton shower and hadronisation.

10.2.1.1 Non-perturbative corrections

The non-perturbative corrections are obtained by a parton shower generator using the predictions at different levels of the simulation. The corrections are obtained by the ratio P_{NP}:

$$P_{NP} = \frac{\frac{d^n \sigma(X_1,...,X_n)}{dX_1...dX_n}\Big|_{\text{shower+had.+UE}}}{\frac{d^n \sigma(X_1,...,X_n)}{dX_1...dX_n}\Big|_{\text{shower}}}. \tag{10.5}$$

The b-jets before hadronisation (had.) are identified by the angular matching ($\Delta R < 0.3$) to b-quarks, after hadronisation the b-jets are identified with respect to b-hadrons. This procedure includes effects in which the b-quarks in the process of hadronisation may be disconnected from the jet. The predictions after the parton shower is set equivalent to the NLO prediction at parton level. The effects of multiple interactions (UE) add activity to the event and can add energy to the jet and increase the measured cross section. On the other hand the effects from hadronisation in which hadrons are formed from the unphysical partons may decrease the cross section. To simulate the $b\bar{b}$ cross section in PYTHIA correctly, it is required to generate inclusive jets and to filter out the desired events. This makes the calculations computing time intensive, in order to achieve a smooth phase space coverage of $b\bar{b}$ events. For each necessary cross section prediction more than 600 million events are generated. In this sense it is reasonable to restrict the number of tunes and other parton shower generators studied to obtain the systematic uncertainties on the corrections. As mentioned before the default choice is PYTHIA and the AMBT1 tune. With the help of PYTHIA the effects on the non-perturbative corrections from different tunes and b-fragmentation functions are examined. The following variations are studied:

- The Perugia tunes (cf. Section 8.3.1 and [107, 108]) are used to obtain LHC independent corrections which are tuned to LEP and Tevatron data. Distinguished are two tunes: the Perugia *hard* tune has a higher amount of activity from perturbative physics and counter-balances this with reduced production of particles from non-perturbative sources, the Perugia *soft* tune behaves the opposite way.

- The b-fragmentation function used in PYTHIA is the Lund-Bowler function presented in Section 3.1.3 Eq. 3.8. Two variations of this function's parameters are studied. The chosen parameters are: $a = 0.3$, $b = 0.58$ GeV^{-2} and $r_Q = 1$ denoted as Bowler (high) and $a = 0.3$, $b = 0.58$ GeV^{-2} and $r_Q = 0.5$ denoted as Bowler (low) in the following. In addition the non-perturbative corrections are calculated using the Peterson fragmentation function Eq. 3.7 with the parameter $\epsilon = 0.005$. Only the fragmentation parameters are changed, the remaining parameters for PYTHIA are taken from the AMBT1 tune.

- The corrections are also obtained using the HERWIG++ generator with the LHC-UE7-2 tune.[4]

The envelop of these six non-perturbative corrections around the default PYTHIA correction is used as systematic uncertainty. The different corrections are shown in Figs. 10.2 and 10.3. The non-perturbative correction not shown here are included in Appendix Fig. K.1. In case of the dijet mass the corrections despite the large statistics generated are suffering from large fluctuations (see Fig. 10.2(a)). In case of the angular cross sections this effect is smaller due to the smaller number of bins and the composition of different kinematic

---

[4]Due to problems in HERWIG++ version 2.5.0 and LHC-UE7-1 which are used in the MC10 simulation samples, for this standalone calculation the version 2.5.1 and LHC-UE7-2 are used.

regions. To smoothen the correction as a function of dijet mass, the correction is fitted and the resulting correction is presented in Fig. 10.2(b). The non-perturbative corrections are close to unity with at most 10% deviation. Therefore both effects influencing the corrections with opposite behaviour seem to balance each other mostly.

The lower uncertainty on the correction is based on the prediction of HERWIG++. This prediction consistently yields the largest effect onto the parton-level prediction from MCFM. Because both tunes AMBT1 and LHC-UE7-2 have been adjusted to LHC data, it is probable that the large deviation is mostly due to the different hadronisation concepts. This assumption is supported by the differences seen in jet fragmentation studies in [132]. The variations based on PYTHIA seem to stay very close to the default correction. The five variations in PYTHIA have bigger correction factors in the lower mass region and vice versa in the higher mass region above about 400 GeV at which point the correction factors intersect.

Using the non-perturbative corrections, the resulting MCFM predictions are presented in
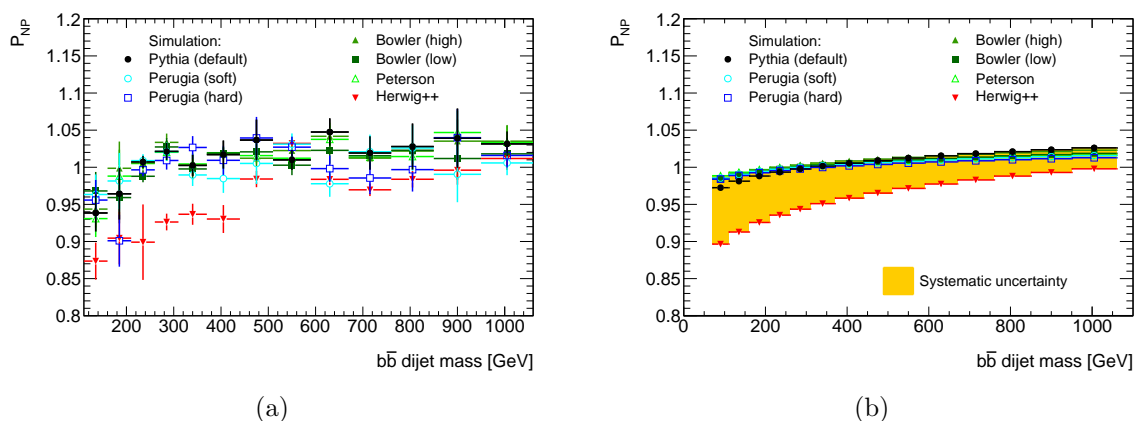
Figure 10.2: *The non-perturbative corrections as a function of dijet mass for the measurement of the dijet cross section of b-jets: (a) the original corrections and (b) the fitted corrections with resulting uncertainty envelop around the default PYTHIA correction.*

Figs. 10.4 and 10.5. The PDF uncertainty on the MCFM prediction is calculated directly by MCFM with the central prediction. The statistical uncertainties are negligible. In case of the cross section as a function of the azimuthal angle between the leading jets the results from pQCD from MCFM at large $\Delta\phi \approx 3$ are excluded because fixed-order perturbation theory fails to describe the data in this region due to the dominating contribution of soft processes.

The theoretical uncertainties range between 5% and 25%. Clearly visible is the dominating effect of the scale uncertainties in all cross sections. The scale uncertainty is at most 20%. On average the scale uncertainty contributes between 5% and 20% to the total uncertainty. The contribution of the PDF uncertainty increases with the transverse momenta and ranges between 2% and 6% in all cross sections. The contribution from the non-perp. corrections ranges between 2% and 20% in rare cases.

Figure 10.3: *The non-perturbative corrections as a function of (a) azimuthal difference and (b) the angular variable χ in the mass range $370\ GeV < M < 850\ GeV$ with resulting uncertainty envelop around the default* PYTHIA *correction.*

10.2.2 The MC@NLO generator

MC@NLO can generate exclusive events. The calculation is done with the NLO matrix element and the soft and collinear emissions are handled by a parton shower. The implementat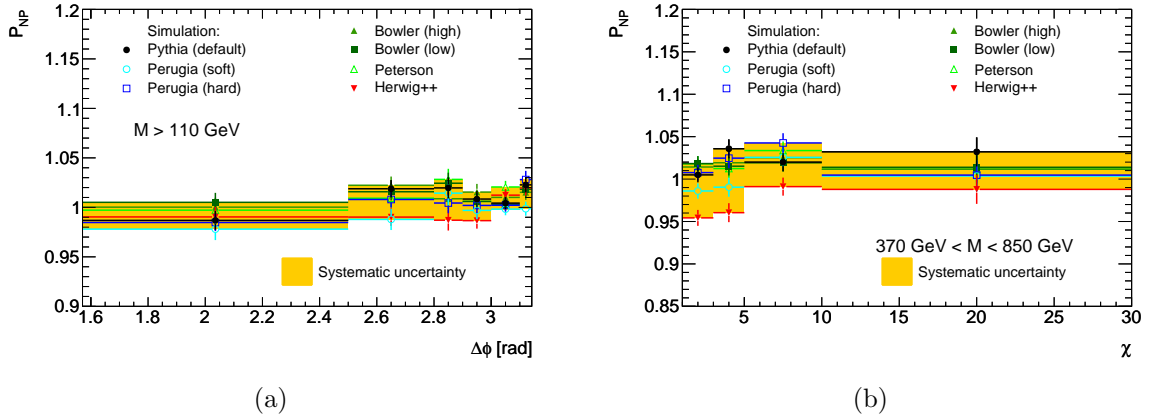ion of MC@NLO allows to insert the NLO calculations to the parton shower generators HERWIG and HERWIG++ in general. In case of $b\bar{b}$ production only the HERWIG simulation is usable. To overcome the problem of double counting configurations which can occur in the matrix element calculation or in the parton shower process, the NLO calculation is modified. MC@NLO can deliver events with negative weights. It should be noted that these negative-weight events have a different origin than the negative contributions that appear in NLO computations. In case of the $b\bar{b}$ production the number of negative events is sizable about 20%. To balance this effect more statistics is necessary. All calculations of MC@NLO contain at least 900 million events which needs a large computing effort. The scale used to compute the predictions is set to the transverse mass of the heavy quark in the rest frame of the two-parton system: $m_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2}$. The core of the NLO generator uses a modified version of the subtraction method to take into account the terms that are generated by the parton shower. For more details see [56]. The pure NLO calculation with the matrix element is modified to avoid the double counting using the Sudakov form factor from HERWIG. The system can undergo an arbitrary number of emissions with probability controlled by the Sudakov form factor and the amount of emissions which occur twice is estimated and subtracted from the NLO calculation before inserting them into the parton shower program.

The implementation of bottom flavour production has to take into account that it is included in the evolution of initial-state and final-state showers in the parton shower programs, consistent with the FEX and GSP processes. Switching off these parts in the parton shower responsible for bottom flavour would lead to inconsistencies. Therefore the implementation of MC@NLO is based on FCR processes only at orders α_s^2 and α_2^3. The prescription of the production formulae is given in [133]. However, all contributions from GSP and FEX cannot be included, like GSP diagrams in which the gluon emits another gluon before splitting. Nevertheless, besides the advantages of predictions of exclusive final states as measured in a detector and the inclusion of soft and collinear contributions,
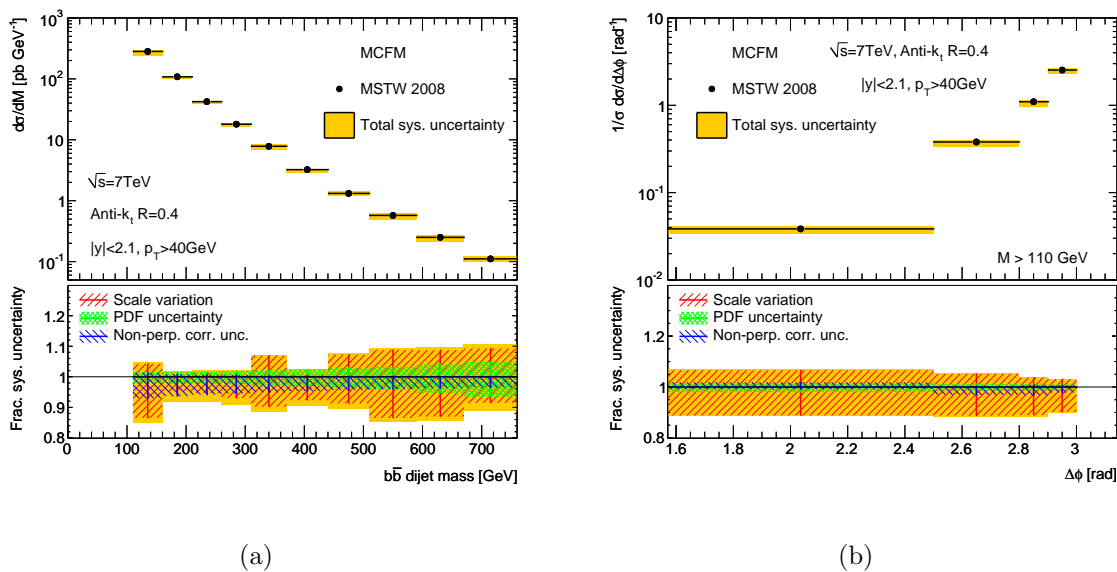
Figure 10.4: *Dijet cross sections calculated at NLO with the help of* MCFM *as a function of (a) dijet mass and (b) azimuthal difference $\Delta\phi$ with the total theoretical uncertainty. The contributions of individual uncertainties are presented by the shaded ares in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

GSP contributions are taken into account fully efficient in contrast to the implementation in standard parton shower predictions.

The PDF uncertainty on the MC@NLO prediction is obtained with the procedure explained above. The predictions according to the different eigenvectors i are calculated using a re-weighting technique. This calculation is different to the MCFM calculation. With the information of the parton's identity (id), their momentum fraction x and the scale Q^2 of the hard process a reweighting factor W for each eigenvector i and event upon the available central prediction with PDF S_0 can be calculated by [134]:

$$\text{reweighting factor } W^i = \frac{f_{id}(x_1, Q, S_i) f_{id}(x_2, Q, S_i)}{f_{id}(x_1, Q, S_0) f_{id}(x_2, Q, S_0)}, \tag{10.6}$$

where W^0 is identical to unity, S_i denotes the i-th eigenvector and f represents the probability to find the partons 1 and 2 with momentum fraction x in the proton. This method is theoretically correct in the limit of a full phase space coverage with high statistics. Because each event is simulated only once, the kinematics are unchanged and there is no residual statistical variation in uncertainty. However, it should be taken care when applying this method because the re-weighted events do not correctly modify the Sudakov form factors used to avoid double counting. The re-weighting can only affect the hard process. However, the impact of this was shown to be negligible (two orders of magnitude below the PDF uncertainty) [135].

 The resulting MC@NLO prediction with the total systematic uncertainty obtained only via the PDF uncertainty are presented in Figs. 10.6 and 10.7. The PDF uncertainties obtained for the MSTW 2008 PDFs are comparable to the PDF uncertainty from MCFM. This supports the validity of the PDF re-weighting approach. For the cross section as a function of dijet mass the PDF uncertainty increases with dijet mass and is at most
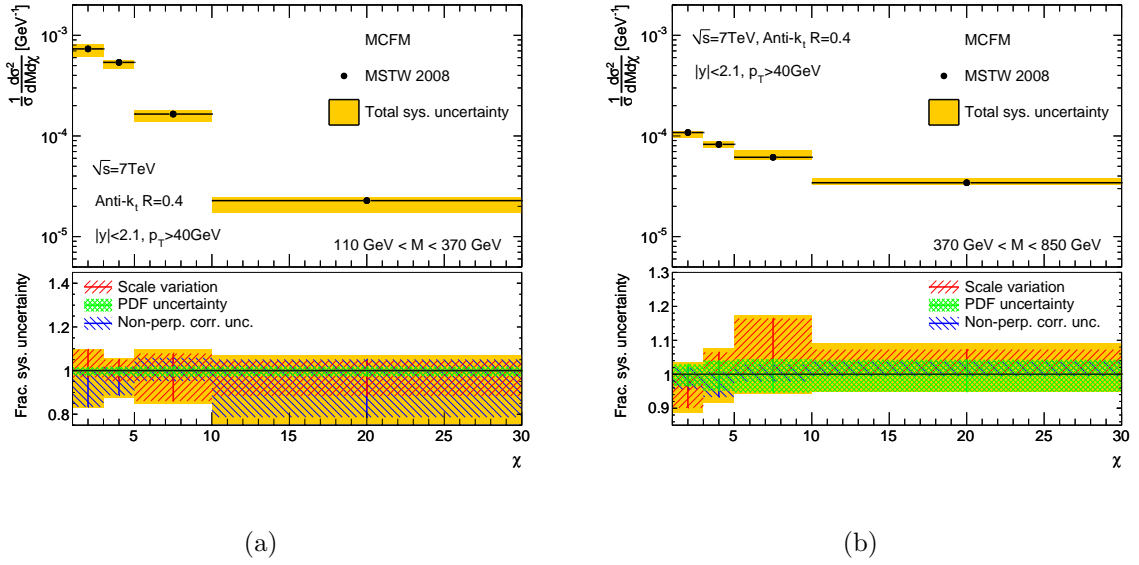
Figure 10.5: *Dijet cross sections calculated at NLO with the help of* MCFM *as a function of χ in the mass region (a) 110 GeV < M < 370 GeV and (b) 370 GeV < M < 850 GeV with the total theoretical uncertainty. The contributions of individual uncertainties are presented by the shaded ares in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

about 6%. For the cross sections as functions of angular variables the PDF uncertainty is in general reduced and at minimum about 1.5%. In contrast to MCFM, MC@NLO is capable of predicting the region around π due to the parton shower approach from HERWIG. As expected the cross section as a function of $\Delta\phi$ peaks at π. Due to negative events the statistical precision of this prediction is a bit worse than compared to the others.

10.2.3 The POWHEG method

The method used in POWHEG [136, 137, 138] (POsitive Weight Hard Event Generator) to calculate the NLO prediction is the most recent development. This generator is used to calculated not only the dijet cross sections of *b*-jets but also the prediction for the inclusive dijet cross section and furthermore the ratio between *b*-jets and inclusive jets. The approach taken by the POWHEG generator is similar to the method in MC@NLO but is meant to overcome its deficiencies.

POWHEG takes advantage of the leading-log approximation in parton shower generators but improves the NLO accuracy of the hard emission to combine the best features of both methods. The drawbacks from MC@NLO are that the NLO generator approach is specific to a particular parton shower generator and may generate negative weighted events. POWHEG overcomes this by the implementation of its own Sudakov form factor. The strategy to achieve this is correcting NLO by the hardest emission and this has to be generated first. Details on the treatment of p_T-ordered and angular order parton showers are given in [127]. It can be interfaced to HERWIG and PYTHIA. But it has to be mentioned that the approach in POWHEG requires minor modification to the parton shower generators which are currently not implemented. However, the effects of these
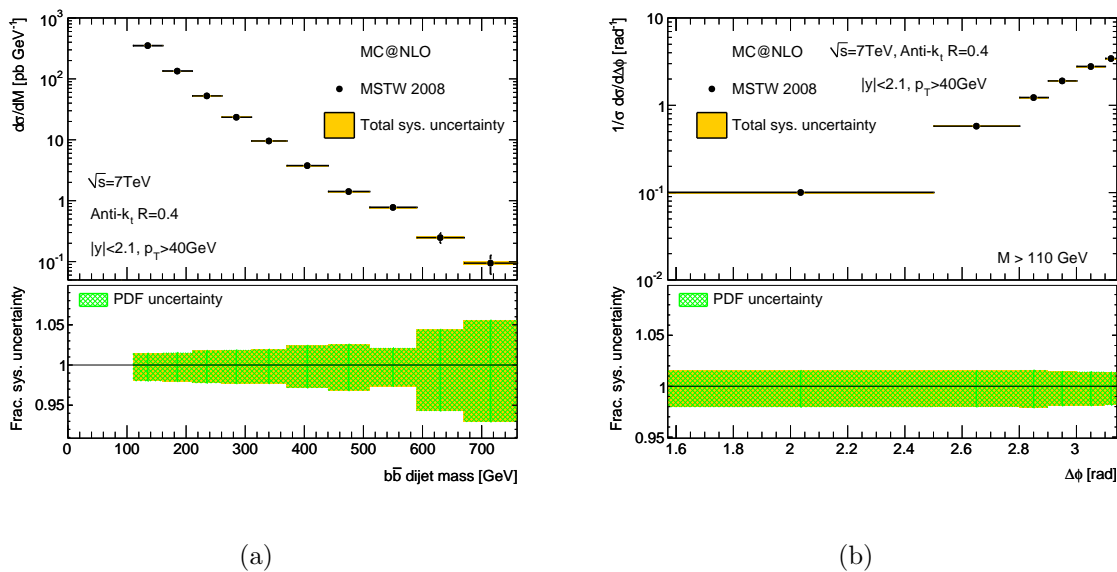
Figure 10.6: *Dijet cross sections calculated at NLO with the help of* MC@NLO *as a function of (a) dijet mass and (b) azimuthal difference* $\Delta\phi$ *with the total theoretical uncertainty. The total theoretical uncertainty consists only of the contribution from the PDF uncertainty and is presented by the area in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

adjustments are expected to be small [139].

The scale used in POWHEG is identical to choice made in MC@NLO ($m_{\mathrm{T}} = \sqrt{m^2 + p_{\mathrm{T}}^2}$). For this NLO prediction the scale variations are calculated and give rise to a contribution to the theoretical uncertainty as described above. The uncertainty on non-perturbative corrections are dominated by the difference of PYTHIA to HERWIG++ and this difference may serve as a reasonable estimate for this uncertainty. Due to the necessary statistics of more than 700 million events, the uncertainty from the non-perturbative corrections are therefore estimated only from HERWIG.

The resulting NLO predictions for the cross sections of b-jets are given in Figs. 10.8 and 10.9. In these predictions the PDF uncertainties are missing. It was found that the PDF re-weighting procedure applied to POWHEG does not give consistent results. This may be due to the fact that the real contributions are evaluated with different structure functions and different x's than in the hard process. It was concluded that it was better to not follow the procedure. Consequently, only the comparison to a different PDF (CTEQ66) was studied, another solution did not seem manageable in terms of computing time. In Fig. 10.11(a) lower part the deviations to CTEQ66 are at most 10%. The total uncertainty on the cross section as a function of dijet mass Fig. 10.8(a) are at most 25% of which the scale uncertainty is dominating. The total uncertainty on the measurement with $\Delta\phi$ are overall smaller than 6% and both uncertainties are equivalent. The cross sections as function of χ have total uncertainties in the range between 10% and 25%. Also here the scale uncertainties are dominant.

In the last two years the dijet production has been implemented into POWHEG, about four years after the implementation of $b\bar{b}$ pair production. In the implementation each event is built by first producing the underlying Born configuration, $2 \rightarrow 2$ scattering. From
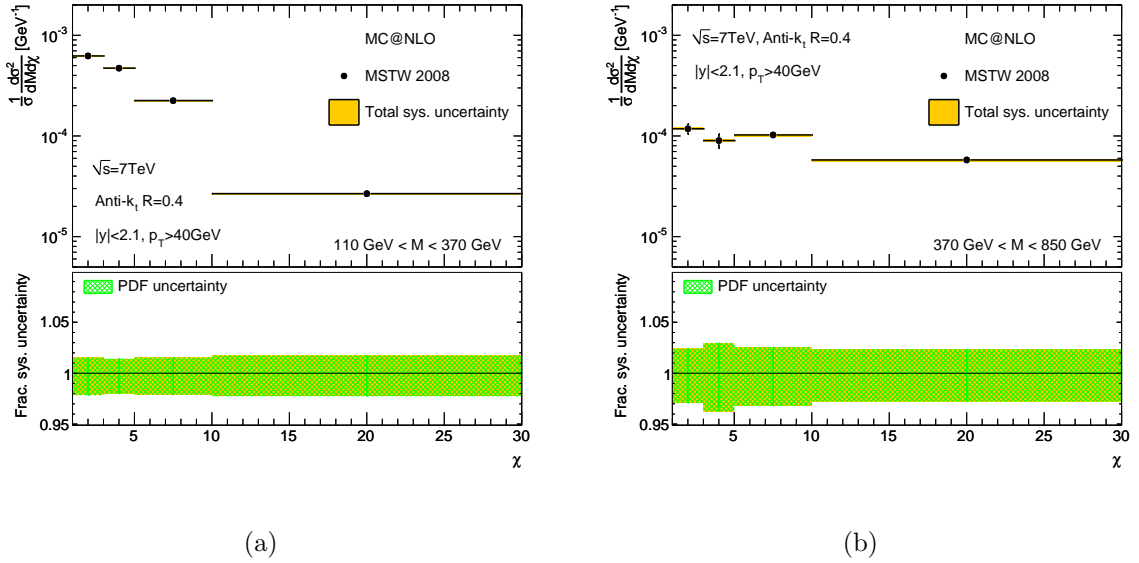
Figure 10.7: *Dijet cross sections calculated at NLO with the help of* MC@NLO *as a function of* χ *in the mass region (a) 110 GeV < M < 370 GeV and (b) 370 GeV < M < 850 GeV with the total theoretical uncertainty. The total theoretical uncertainty consists only of the contribution from the PDF uncertainty and is presented by the area in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

this Born configuration the $p_{\rm T}$ of the system is used as factorisation and renormalisation scale. To control the divergences in the inclusive cross section a generation cut of 1 GeV is applied. In addition, the POWHEG event generation is capable of weighting events to populate the high-$p_{\rm T}$ regions of phase space with less statistics. When this weighting procedure is activated, events are distributed according to the differential cross section multiplied by

$$S(k_{\rm T}) = \left(\frac{k_{\rm T}^2}{k_{\rm T}^2 + k_{\rm T,supp}^2} \right)^3 , \qquad (10.7)$$

where $k_{\rm T}$ is the transverse momentum of the final-state partons. In case of the inclusive dijet production the suppression factor was set to $k_{\rm T,supp} = 250$ GeV. This weighting procedure is unfortunately not tested and bears problems in case of $b\bar{b}$ production and was abandoned in this case. The necessary statistics is reduced to about 50 million events. Evaluated are the scale uncertainty and with the help HERWIG the uncertainty from the non-perturbative corrections.

The resulting inclusive dijet cross section as a function of dijet mass is presented in Fig. 10.10. The total uncertainty ranges between 4% and 20%. The scale uncertainty contributes at most 10% at medium and high dijet mass. The non-perturbative corrections contribute at most 20% in negative direction in general.

For the ratio measurement POWHEG is used exclusively. This ensures complete correlation between the scale uncertainties in both predictions and allows to propagate the scale uncertainty directly onto the ratio measurement. To illustrate the effects of the change in the parton shower model and the effects of a different PDF onto the ratio measurement Fig. 10.11 is given. The POWHEG calculations are repeated in case of the $b\bar{b}$ measurement
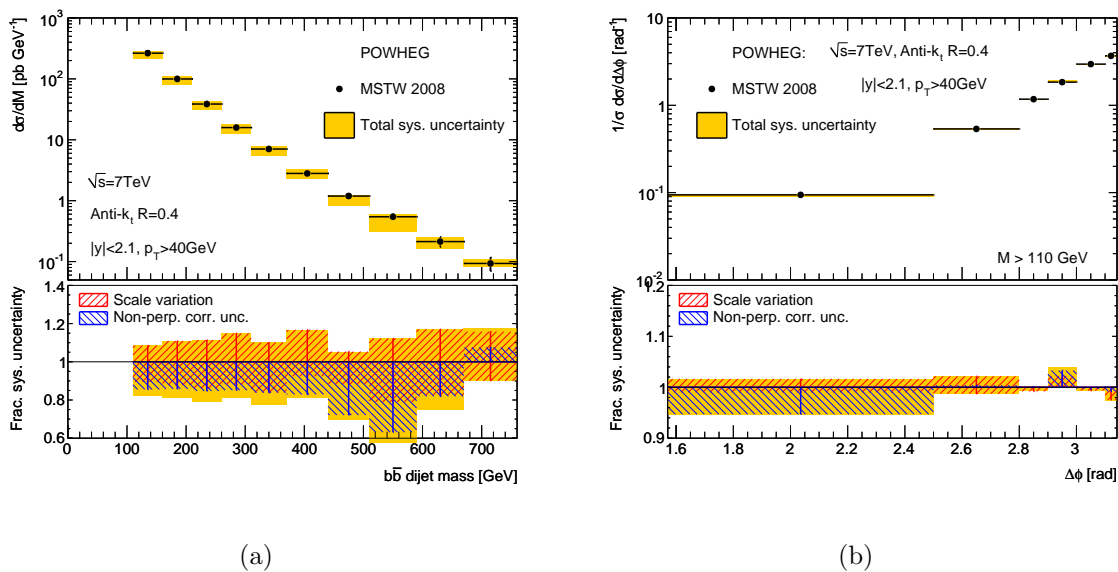
Figure 10.8: *Dijet cross sections calculated at NLO with the help of* POWHEG *as a function of (a) dijet mass and (b) azimuthal difference $\Delta\phi$ with the total theoretical uncertainty. The total theoretical uncertainty consists only of the contributions from the scale uncertainty and the uncertainty on the non-perturbative corrections. The shaded areas present the individual contributions in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

as well as the inclusive cross section with the CTEQ66 PDF because to some extend the PDF uncertainties from the eigenvectors from MSTW 2008 cancel in the ratio measurement. Due to the limitation in computing time, the PDF uncertainty on the ratio measurement is estimated using the PDF CTEQ66. Both plots in Fig. 10.11 indicate the same HERWIG behaviour with respect to the non-perturbative corrections in Fig. 10.2(b). Both HERWIG predictions are about at most 20% below the corresponding PYTHIA prediction. In case of the POWHEG calculation with CTEQ66 the prediction is also up to 10% below the MWST 2008 prediction. This effect seems to be larger in the cross sections of *b*-jets.

 The resulting ratio measurement of dijets between *b*-jets and inclusive jets is presented in Fig. 10.12. The total uncertainty on the measurement is at most 20% but in general reduced by the correlations between the uncertainties on the individual cross sections. The similar tendencies in the predictions with CTEQ66 reduce the PDF uncertainty to about 5%. The remaining scale uncertainties are caused by opposite effects of the scale uncertainties in the individual cross sections. The scale uncertainty ranges from a minimum of 8% to 20% in the last bin of the measurement. Also the strong correlations in the ratios of the prediction with PYTHIA parton shower to the prediction with HERWIG parton shower reduce the uncertainty on the non-perturbative correction. Except for five bins, this uncertainty is negligible, these bins show a different behaviour with respect to the default prediction. Besides the advantage of reducing experimental uncertainties also the theoretical uncertainties are positively influenced.
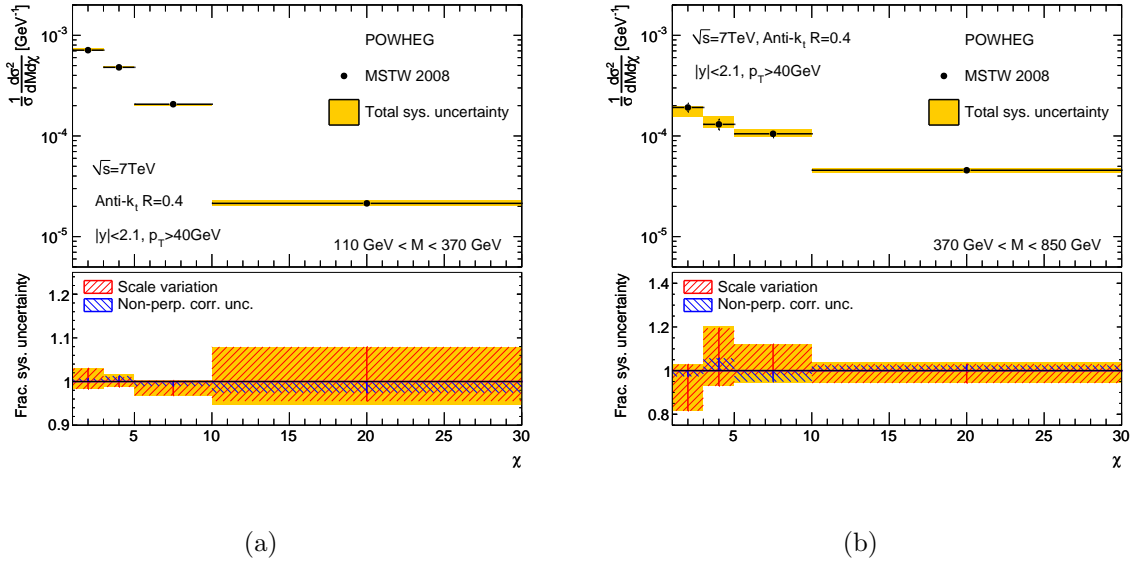
(a)   (b)

Figure 10.9: *Dijet cross sections calculated at NLO with the help of* POWHEG *as a function of* χ *in the mass region (a) 110 GeV < M < 370 GeV and (b) 370 GeV < M < 850 GeV with the total theoretical uncertainty. The total theoretical uncertainty consists only of the contributions from the scale uncertainty and the uncertainty on the non-perturbative corrections. The shaded areas present the individual contributions in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

10.3 Summary

The performed NLO calculations and uncertainty assessments represent a compromise between their importance and the estimated computing power needed. The PDF uncertainties approximately amount to 6% fractional uncertainty at most. The uncertainty arising from the choice of a parton shower model is about 10% for MCFM or 20% for POWHEG at most depending on the models which are compared. The dominating uncertainty is the scale variation which contributes in general 15% to 20% fractional uncertainty.

In terms of computing power the fixed order NLO calculation from MCFM have some advantages. The PDF uncertainties are calculated directly without consuming additional computing power and the scale variations are also not time consuming. But in addition the non-perturbative corrections are necessary. The limitations of the fixed-order prediction are the inability to produce predictions in regions in which soft processes are important and the necessity of non-perturbative corrections in which the NLO accuracy is exchanged with the prediction from the leading-order parton shower generator.

The procedure of PDF re-weighting is proven to be problematic on POWHEG. But the similarities between MC@NLO and POWHEG allow at least an estimation of them for the POWHEG prediction. So it can be roughly assumed that the PDF uncertainties calculated for MC@NLO also hold for the POWHEG calculations. They are not explicitly added to the calculation but it is kept in mind that they ought to contribute as well.

With the ability to produce inclusive dijet as well as $b\bar{b}$ predictions, the focus for the NLO calculations has been relying on POWHEG. Therefore the scale uncertainties have been calculated for POWHEG (about 4.5 billion events).

Figure 10.10: *Inclusive dijet cross sections calculated at NLO with the help of* POWHEG *as a function of dijet mass with the total theoretical uncertainty. The total theoretical uncertainty consists only of the contributions from the scale uncertainty and the uncertainty on the non-perturbative corrections. The shaded areas present the individual contributions in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

The dominating effect of HERWIG++ in the non-perturbative corrections has given the possibility to estimate these uncertainties with reasonable computing effort on POWHEG. The effects on the MC@NLO calculation from a change in the HERWIG tune have been neglected.

Differences between MC@NLO and POWHEG are due to the treatment of higher orders [140]. They are tracked down to the role of how the radiation of the matrix element is generated, sub-leading terms in the parton shower may differ in the two approaches, due to reshuffling of the splitting process in the parton shower and to different scaling choices in both generators. But both still are proven to be at NLO accuracy. New implementations and improvements are on their way.

From the fact, that the difference seen in the non-perturbative corrections between PYTHIA and HERWIG++ for MCFM are consistently smaller than the ones evaluated with the help of POWHEG predictions using HERWIG's and PYTHIA's parton showers, may be concluded that the minor modifications needed in the parton shower generators lead to opposite effects in PYTHIA and HERWIG parton showers. This was reported to the authors of the generator and is under investigation.
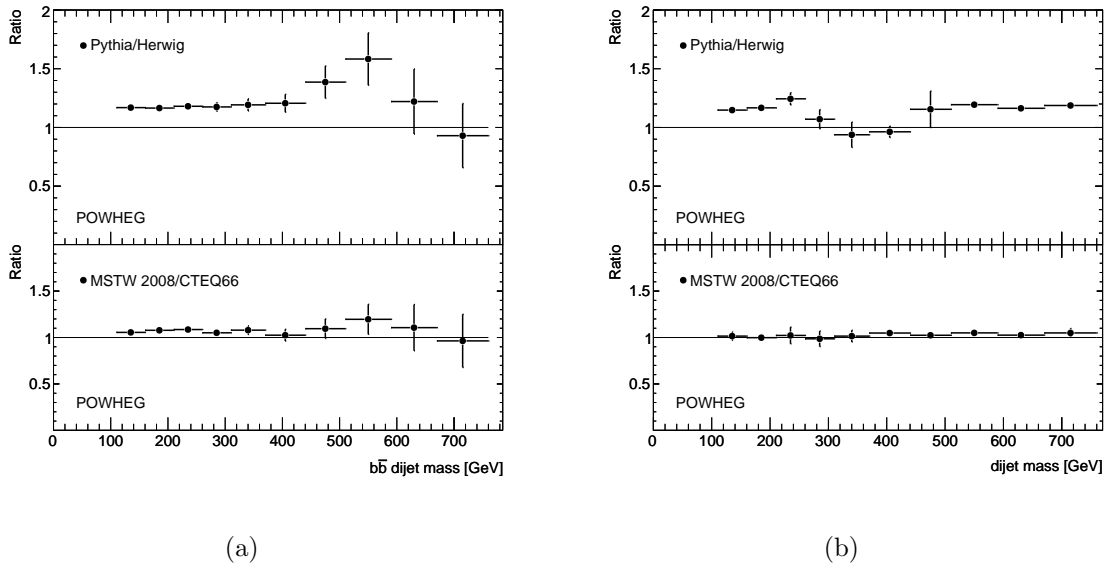
Figure 10.11: *Ratio of the default* POWHEG *prediction using* PYTHIA *with MSTW 2008 PDF to (upper plot) the* HERWIG *prediction using the same PDF and (lower plot) to the prediction using* PYTHIA *but the CTEQ66 PDF: (a) dijet cross section of b-jets as a function of dijet mass and (b) dijet cross section of inclusive jets as a function of dijet mass. The statistical errors are calculated uncorrelated.*

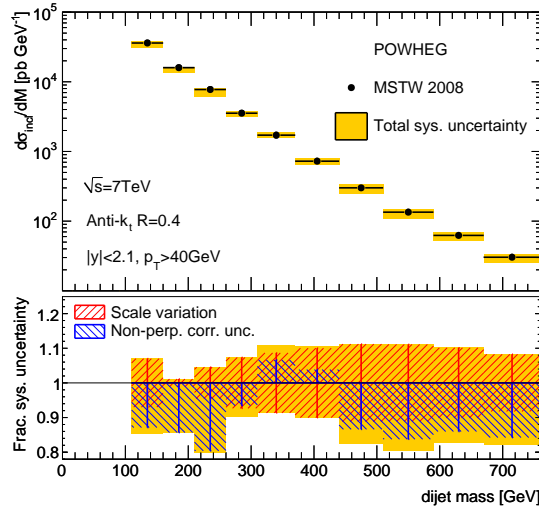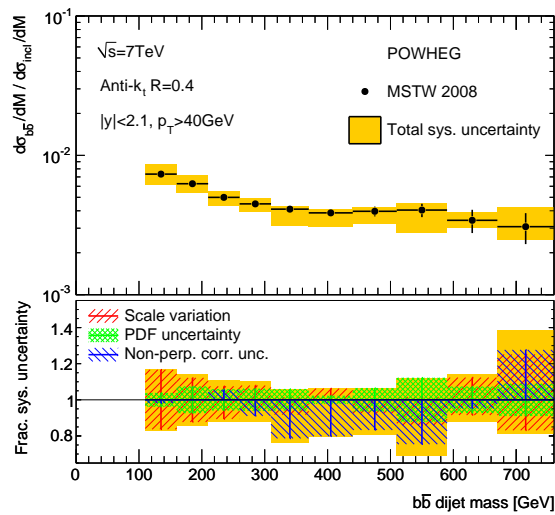**Figure 10.12:** *Ratio of the dijet cross sections of b-jets and inclusive jets calculated at NLO with the help of* POWHEG *as a function of dijet mass with the total theoretical uncertainty. The total theoretical uncertainty consists of the contributions from the scale uncertainty, the PDF uncertainty evaluated with CTEQ66 and the uncertainty on the non-perturbative corrections. The shaded areas present the individual contributions in the lower part of the plots. The markers on the central prediction included the statistical uncertainty.*

11 Results of the cross section measurements

The previous chapters have gone into detail on the object reconstruction and understanding of jets and b-jets. The trigger menu used to record the data, the event and object selection cuts have been presented. These cuts are amongst other things determined by the triggers and the trigger efficiencies to make optimal use of the trigger plateau regions and to minimise systematic uncertainties on the trigger selection. The event and jet selection cuts are influenced by b-tagging requirements to stay as conform as possible to the b-tagging calibration. At last the detector layout dictates immutable conditions for the reconstruction. Furthermore, migration effects are minimised if possible by the choice of the binning.

The most important systematic uncertainties on these cross section measurements arise from the jet reconstruction and b-jet identification. Major parts of the jet and b-jet energy scale uncertainty determination have been discussed. The extension of the b-tagging calibration which is the basis for this measurement has been verified with the aid of data. Additional small uncertainties are coming from the luminosity and event and trigger selection. The measurement has been corrected for detector effects to make the measurement experiment independent and the correction has been proven to be mostly unbiased from the simulation model used. It has been validated that the small amount of pile-up interactions in the data have a negligible effect on the measurement.

The measurement is compared to the calculations at next-to-leading order presented in the previous chapter and the PYTHIA prediction. For all NLO predictions the theoretical uncertainties have been estimated. For the measurements as a function of azimuthal difference and angular variable χ the cross sections are normalised to the measured cross section which reduces the experimental as well as the theoretical uncertainties. The theoretical uncertainty is dominated mostly by the choice of the scales and the effect is found to be approximately 20% for all kinematic regions whereas the PDF uncertainty is at most 10%.

The measurement of the one-particle inclusive b-jet cross section has been performed which serve as a strong test of pQCD [118]. However, several important features of the production mechanisms of b-jet pairs can only be probed with the help of the correlation between the quark and antiquark. Measurements at hadron colliders are in addition sensitive to fragmentation effects and can deliver valuable insights in non-perturbation processes.

This measurement includes the $b\bar{b}$ pairs produced via $t\bar{t}$ production. The inclusive $t\bar{t}$ cross section measured in ATLAS is about 170 pb^{-1} [141]. Assuming that the major part of the b-jets are produced with a transverse momentum between 60 and 80 GeV the contribution to the inclusive b-jet cross section (≈ 120 nb) is at most per mill level. Furthermore, as the measurement select jets with $p_T > 40$ GeV and measures the invariant mass $M > 110$ GeV, the major fraction of events from $Z \to b\bar{b}$ (≈ 5 nb) decays are cut away. Besides these known contributions, the large QCD background and the prescaling of QCD events in the trigger selection diminishes the sensitivity of this measurement to the Higgs decay mode into $b\bar{b}$ although the branching ratio into $b\bar{b}$ for a Higgs boson at low masses around

50 GeV to 150 GeV is large. The discovery potential in the other possible decay modes into vector boson pairs or photon pairs is enhanced due to the cleanness of the signals.

11.1 Dijet cross sections of b-jets

The measurement of $b\bar{b}$ dijets studies to a large extend the kinematic properties of the quarks arising in flavour creation (FCR) processes. Since the measurement refers to leading and sub-leading jets, the event selection enhances this contribution. In addition the major part of $b\bar{b}$-quark pairs from gluon splitting are merged into one jet and thus are not resolved into a dijet event topology.

The b-tagging templates and the purity determination dictates the maximal phase space studied in this measurement. Furthermore, they also define for the most part the available bin sizes. But it has been shown that as much as the b-jet identification influences choices made in the measurement the choice of the b-tagging algorithm does not alter the final cross section.

The following sections present in detail the resulting comparisons between data and MC predictions. In case of the NLO predictions the MSTW 2008 PDF is used and in order to obtain the particle-level prediction the hadronisation and shower model from PYTHIA is used from the MC10 tune. The PYTHIA prediction itself uses the MRSTLO* PDFs. In all cases b-jets are defined at particle level with an angular matching to a b-hadron within $\Delta R < 0.3$. The cross section in ATLAS data is compared to four QCD predictions, one from PYTHIA at leading order and the other three from MCFM, POWHEG and MC@NLO at next-to-leading order. From these predictions the central values are overlaid on top of the measurement without any statistical uncertainties.

11.1.1 Cross section as a function of dijet mass

The dijet cross section of b-jets as a function of dijet mass is presented in Fig. 11.1 for jets with $p_\mathrm{T} > 40$ GeV and $|y| < 2.1$. The cross section is steeply falling over almost four orders of magnitude which is well reflected by all MC predictions. Independent from the order at which the calculation was done all four MC predictions seem to describe the general features of the cross section. However, the PYTHIA prediction is not expected to reproduce the correct normalisation. Therefore, leading-order prediction of PYTHIA is scaled to the measured integrated cross section in data. The scaling factor needed is 0.83. The right hand side in Fig. 11.1 presents the ratios of data to each MC prediction. The statistical uncertainty from the measurement and from the MC predictions are propagated into the ratio assuming uncorrelated errors. The systematic uncertainty on the measurement ranges from 40% to 60%. From Fig. 3.11(b) it was deduced that the NLO contributions from gluon splitting (GSP) and flavour excitation (FEX) contribute at most 10% to the measured dijet cross section due to the emphasis on $2 \rightarrow 2$ configurations. In consideration of the systematic uncertainty and the portion of GSP/FEX processes, it can be concluded that it is probable that even without those contributions from GSP and FEX an agreement between data and MC is achievable. But it should be another time stressed that categorising the events is strictly speaking incorrect.

The dijet cross section from MCFM at leading order with a leading-order PDF which has been presented in Fig. 10.1 would not be in agreement with data. Therefore, NLO corrections are essential to obtain consistency between the measured cross section and

Figure 11.1: *The $b\bar{b}$ dijet cross section as a function of dijet mass for b-jets with $p_T > 40$ GeV and $|y| < 2.1$ is presented on the left hand side. The data are compared to MC predictions from POWHEG, MC@NLO, MCFM and PYTHIA. Only the central values of the theoretical predictions are drawn. The leading-order PYTHIA prediction is scaled to the measured to the measured integrated cross section. On the right hand side the corresponding ratios between data and the MC simulations are presented. The shaded areas around the MC prediction indicate the systematic theoretical uncertainty (except for PYTHIA). The statistical error is calculated from the uncorrelated statistical errors from MC and data.*

QCD. However, PYTHIA would be in agreement even without the scaling factor, this can be concluded from Fig. 11.1 (right bottom). This agreement is achieved by the parton shower approach and the usage of a PDF which mimics NLO properties. All four MC predictions are in agreement with data within the systematic uncertainties from data and the systematic theoretical uncertainties from the scale choice. The different size of the theoretical uncertainties has been explained in the previous chapter.

11.1.2 Cross section as a function of $\Delta\phi$

The dijet cross section of b-jets normalised to the total measured cross section as a function of azimuthal difference $\Delta\phi$ is presented in Fig. 11.2 for jets with $p_T > 40$ GeV and $|y| < 2.1$. The cross section shows that the majority of events exhibit a pronounced back-to-back configuration in the transverse plane which is generally well reproduced by the QCD calculations. The cross section decreases rapidly with diminishing azimuthal separation. The predictions from the fixed-order calculation of MCFM are excluded for the region near π. Due to the dominating soft processes in this region and contributions from logarithmic terms, the fixed-order prediction is divergent. The systematic uncertainty on data is about 40% all kinematic regions.

With the help of the studies on heavy-flavour production, the relative contributions to this cross section are understood. In Fig. 3.9(a) the distribution of the azimuthal distance from different production mechanisms between the b-hadrons is illustrated. The leading-order production via the flavour creation (FCR) requires back-to-back configurations and

dominates the region around $\Delta\phi = \pi$. The NLO contributions to FCR (real and virtual emission) shift the angular separation to smaller values. Until a angular separation of $\Delta\phi \approx 2$ the FCR contributions dictate the shape of the cross section. Below this region the NLO production mechanisms FEX and GSP which are approximately flat for the entire phase space $0 < \Delta\phi < \pi$ are responsible for the cross section shape and value. The measurement is only understandable with the contributions from FEX and GSP.

The cross section for small azimuthal angle differences is affected by the separation in η-ϕ space of the b-hadrons. This determines the possibility of the reconstruction algorithms to resolve them into different physical objects, namely here jets. The cross section therefore depends heavily on the resolution parameter of the jet reconstruction.

The normalisation of the cross sections reduces experimental and theoretical uncertainties and allows only a shape comparison between data and MC. Consequently, there is no need to extract a PYTHIA scaling factor. In the region above $\Delta\phi \geqslant 2.5$ all four MC predictions show an agreement with data as can be seen in the ratios on the right side of Fig. 11.2. The statistical uncertainty from the measurement and from the MC predictions are propagated into the ratio assuming uncorrelated errors. Among those four POWHEG, MC@NLO and PYTHIA show excellent agreement, even if only the statistical uncertainty is considered. In the region below $\Delta\phi < 2.5$ MCFM is still in agreement but underestimates the relative cross section by a factor of almost 2. Whereas the other predictions show a deviation to data of maximal twice the uncertainty. The smallest deviation is between data and the POWHEG prediction. The cause of this deviation may be enlightened with the help of two measurements.

The angular correlations between the $B\bar{B}$-hadrons with $p_T^B > 15$ GeV and $|\eta^B| < 2.0$ have been measured of the CMS collaboration [142]. This measurement uses displaced secondary vertices to identify the $B\bar{B}$-hadrons. The measurement is performed with a total integrated luminosity of 3.1 pb^{-1}. For three leading jet p_T regions, the differential $B\bar{B}$ production cross section as a function of ΔR and $\Delta\phi$ is measured and compared to predictions of PYTHIA and MC@NLO. The CMS measurement exhibits a substantial enhancement of the cross section with respect to MC predictions at small angular separation ($\Delta\phi < 1$) and a small deficiency in the intermediate region around $\Delta\phi \approx 2$. In these regions the PYTHIA prediction is even closer to data than the prediction of MC@NLO. The overshoot seen in the lowest bin of the cross section as a function of $\Delta\phi$ in this analysis may be qualitatively explained by the CMS measurement. When approaching the region below $\Delta\phi < 2$, the processes FEX and GSP become important. The CMS measurement predicts a stronger negative slope in the cross section as a function of $\Delta\phi$ which decreases the number of events in the intermediate region around $\Delta\phi \approx 2$ and reduces the measurable dijet cross section. This would lead to exactly the same effect as seen in this analysis.

Besides this, also the inclusive b-jet measurement [118] in ATLAS have seen a remarkable difference in the track multiplicities and secondary vertex mass. With an increasing number of tracks in the secondary vertex, the secondary vertex mass measured in data and reconstructed in the simulation begin to disagree. The secondary vertex mass measured in data is shifted to larger values. It seems reasonable to assume that jets containing two b-hadrons likely have a large numbers of tracks measured in the secondary vertex and this fraction of jets seems to be underestimated in the simulation as it can be concluded from the CMS measurement.

However, it cannot be confirmed that the deviation seen in this analysis can be completely attributed to these two observations nor quantitatively explained by them. These observations can only serve as a guidance.
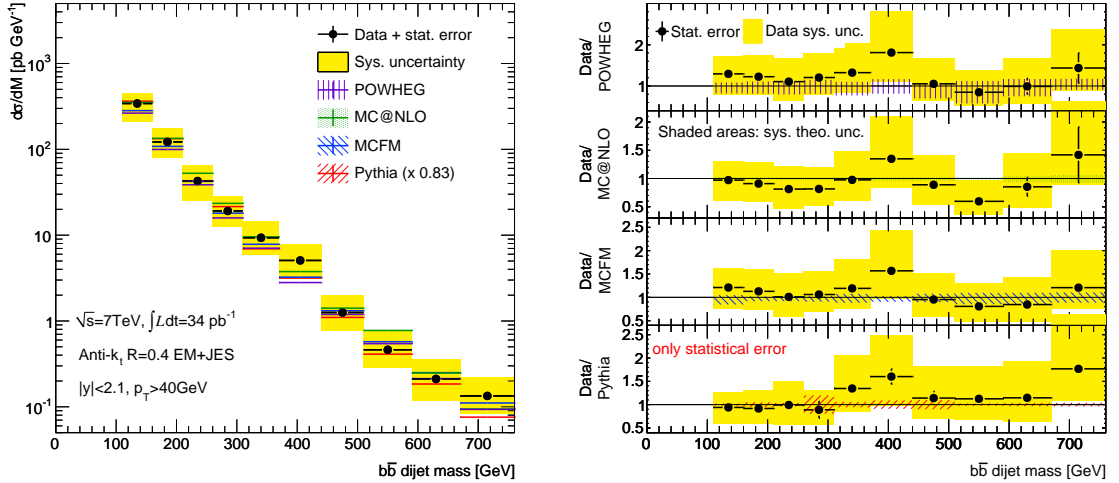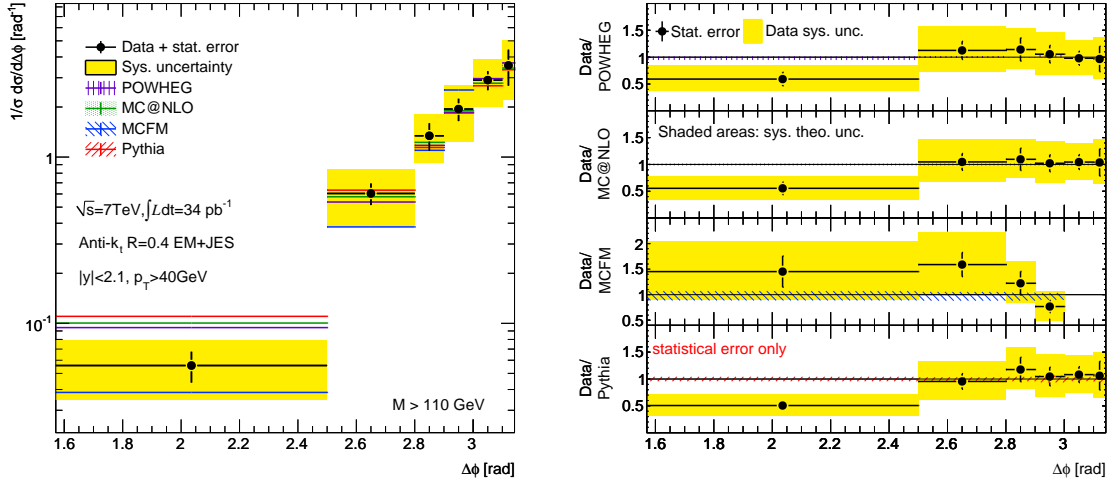
Figure 11.2: *The $b\bar{b}$ dijet cross section as a function of $\Delta\phi$ for b-jets with $p_T > 40$ GeV and $|y| < 2.1$ is presented on the left hand side. The data are compared to MC predictions from* POWHEG, MC@NLO, MCFM *and* PYTHIA. *Only the central values of the theoretical predictions are drawn. On the right hand side the corresponding ratios between data and the MC simulations are presented. The shaded areas around the MC prediction indicate the systematic theoretical uncertainty (except for* PYTHIA). *The statistical error is calculated from the uncorrelated statistical errors from MC and data.*

11.1.3 Cross section as a function of angular variable χ

The dijet cross section of b-jets normalised to the total measured cross section as a function of angular variable χ is presented in Figs. 11.3 and 11.4 in two different mass regions for jets with $p_T > 40$ GeV and $|y| < 2.1$. The maximum allowed range due to the rapidity selection cut would be $e^{4.2} \approx 67$. The measurement restricts the allowed rapidity separation to $|y_1 - y_2| < 3.4$. The separation in different mass regions increases the sensitivity to the behaviour of the partonic cross section. For fixed $\hat{s}$ the partonic cross section $d\hat{\sigma}/d\chi$ is approximately constant. Deviations from QCD in context of new resonances are more likely to give rise to events peaking at low χ. The correspondence between $\hat{s}$ and dijet mass M makes it preferable to separate a measurement into different dijet mass bins. Due to statistical restrictions, this measurement makes use of only two dijet mass regions: 110 GeV $< M <$ 370 GeV and 370 GeV $< M <$ 850 GeV.

In hadron-hadron interactions the PDFs need to be considered and are convoluted with the partonic cross section. The momentum fractions x of the cross section are equivalent to parameters dependent from the rapidities. The cross section $d\sigma/d\chi$ measured in a specific mass region belongs to a nearly constant partonic cross section. But each χ bin samples events in which only the rapidities vary according to the PDF convolution. To control this contribution and to ensure the conformity of all χ bins in terms of PDF weighting factors, the acceptance cut $\frac{1}{2}|y_1 + y_2| < 1.1$ enforces the same $(y_1 + y_2)$ range in all bins. This cut rejects part of the events in the χ range between 1 and 10. This in addition also controls the PDF uncertainties due to the rejection of events in which both jets are

Figure 11.3: *The $b\bar{b}$ dijet cross section as a function of χ in the mass region 110 GeV < M < 370 GeV for b-jets with p_T > 40 GeV and $|y|$ < 2.1 is presented on the left hand side. The data are compared to MC predictions from* POWHEG, MC@NLO, MCFM *and* PYTHIA. *Only the central values of the theoretical predictions are drawn. On the right hand side the corresponding ratios between data and the MC simulations are presented. The shaded areas around the MC prediction indicate the systematic theoretical uncertainty (except for* PYTHIA*). The statistical error is calculated from the uncorrelated statistical errors from MC and data.*

boosted strongly in the same direction.

On the left side of Figs. 11.3 and 11.4 the measurements are overlaid with the corresponding MC predictions. The cross sections as a function of χ are decreasing with increasing rapidity separation. In the lower mass region the cross section falls more than one order of magnitude. In the higher mass region this slope is relaxed and the cross section descents only a factor of about 2. On first sight the MC predictions are all in agreement with data. The systematic uncertainty on data is about 40% to 55% in the lower mass region and 50% to 100% in the higher mass region.

At leading order the χ distributions consist of two contributions: from light-quark annihilation (cf. Eq. 3.4) and gluon fusion (cf. Eq. 3.5). The characteristic properties in terms of rapidity correlation of these two contributions are discussed in Section 3.1.1. The cross section for light-quark annihilation is a rapidly decreasing function of Δy (cf. Fig. 3.4(b)) due to the fact that this subprocess proceeds entirely through an s-channel propagator exchange. On the other hand the partonic cross section of the gluon-fusion subprocess dominated by t- and u-channel exchange diagrams is a broad distribution as a function of rapidity difference with a maximum at $|\Delta y| \approx 1$ (cf. Fig. 3.3(b)). The first contribution via the s-channel exchange peaks strongly at $\chi = 1$ and decreases rapidly with increasing χ. The second contribution describes a much slower decrease with increasing χ and is much more flat than the first contribution. The next factor which needs to be considered is the product of the parton momentum densities which is in general not a function of Δy. The χ cross section in the lower mass region falls almost two orders of magnitude until $\chi = 30$, in the higher mass region the decrease of the cross section is damped. Here
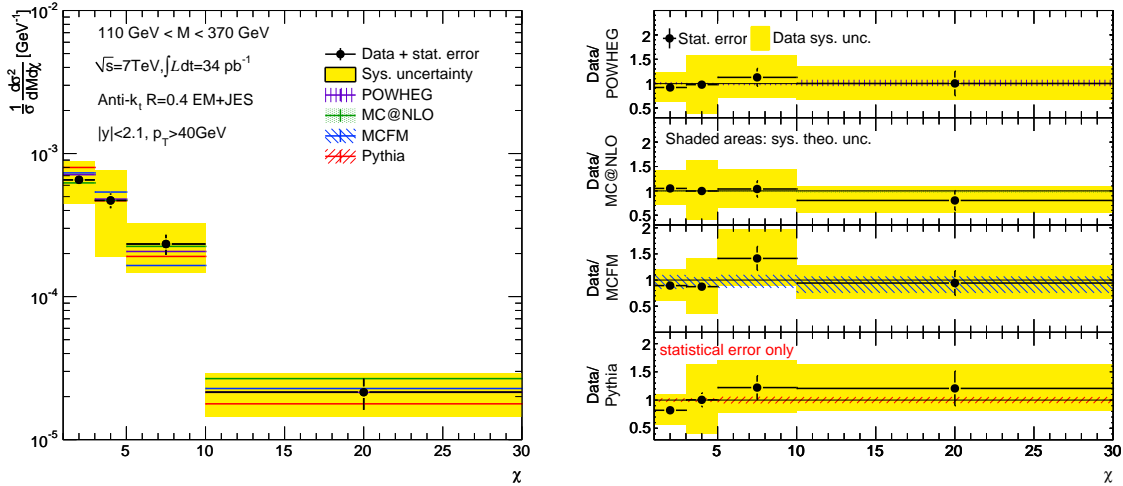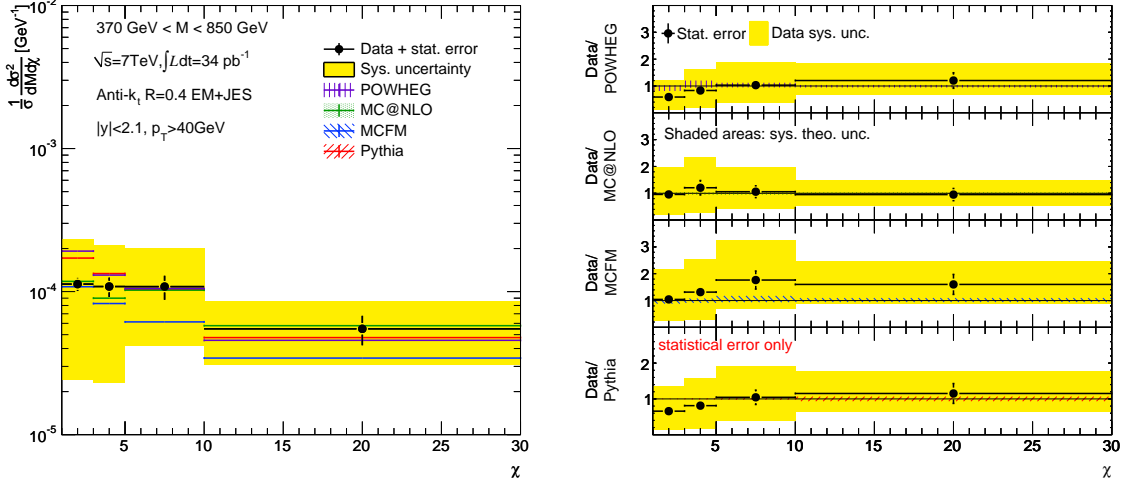
Figure 11.4: *The $b\bar{b}$ dijet cross section as a function of χ in the mass region 370 GeV < M < 850 GeV for b-jets with p_T > 40 GeV and $|y|$ < 2.1 is presented on the left hand side. The data are compared to MC predictions from POWHEG, MC@NLO, MCFM and PYTHIA. Only the central values of the theoretical predictions are drawn. On the right hand side the corresponding ratios between data and the MC simulations are presented. The shaded areas around the MC prediction indicate the systematic theoretical uncertainty (except for PYTHIA). The statistical error is calculated from the uncorrelated statistical errors from MC and data.*

two effects from the parton density functions play a role. It is useful to parametrise the PDFs with the following function: $xf(x, Q^2) = cx^{-\alpha}(1-x)^{\beta}$ with $0 < \alpha < 1$ [12, 34]. The parameters c, α and β are unique to each parton at a given scale Q^2, at high x gluons can be described by $\beta = 5$ and quarks with $\beta = 3$ assuming $Q^2 = 5$ GeV^2.

At small x_1 and x_2 the behaviour of the product of the parton momentum densities $x_1 x_2 f_1(x_1, Q^2) f_2(x_2, Q^2)$ supports the positive correlation peaked at $\Delta y = 0$ of the outgoing partons $((1-x)^{\beta} \approx 1)$ using the Eq. 2.16:

$$(x_1 x_2)^{-\alpha} \propto (1 + \cosh(\Delta y))^{-\alpha} \tag{11.1}$$

The product $(1-x_1)^{\beta_1}(1-x_2)^{\beta_2}$ itself cannot be expressed in terms of Δy. The damping effect of the product $(1-x_1)^{\beta_1}(1-x_2)^{\beta_2}$ cannot be neglected if x grows. With increasing invariant dijet masses the momentum fractions of the partons rise, the factors $(1-x_1)^{\beta_1}$ and $(1-x_2)^{\beta_2}$ cannot be neglected and produce an anti-correlation between the outgoing partons forcing Δy away from zero. The structure function and the cross section at Born level favour opposite configurations. The rapidity correlation in the gluon-fusion process counteracts the influence of the gluon structure functions and the favoured harder momenta spectra of valence quarks counteract the strong and narrow correlation of the light-quark annihilation process. In the lower mass region the structure functions do not counteract strongly against the rapidity correlations from the cross sections. In the higher mass region the parton momenta are larger and both effects influence the shape of the χ cross sections shifting the χ values away from one and flattening the cross section shape. The higher orders leading to configurations with three emerging partons, in which the $b\bar{b}$ pair is not back-to-back or a heavy quark at large angle to the beam direction and the

other close to the incident proton, are likely to broaden the rapidity correlation whereas the gluon splitting contribution tends to accentuate a peak at small rapidity differences. In Fig. 3.12 the expected subprocess contributions are illustrated and it is shown that the pure NLO subprocesses are a small fraction.

On the right side of Figs. 11.3 and 11.4 the ratios of the corresponding cross sections to the MC predictions are shown. The statistical uncertainty from the measurement and from the MC predictions are propagated assuming uncorrelated errors into the ratio. The agreement of the MC predictions with data is very good. In most of the cases the agreement is achieved even at the level of the statistical uncertainties. It can be concluded as QCD powerfully describes the data in the kinematic region measured, that there seems to be not much space left for new phenomena. The agreement in the angular and rapidity related cross sections indicate that the slight bump registered in the dijet mass cross section at around 400 GeV is completely caused by experimental effects.

11.1.4 Summary

In all measured cross sections an agreement between data and MC predictions is observed. In some cases the agreement is at a few per cent level and covered by the statistical uncertainty on data.

The NLO corrections are essential to understand the measured cross sections. This is especially true in case of the cross sections as a function of azimuthal angle between the leading jets and of the correlation in polar angle. However, it has been observed that the predictions from PYTHIA give reliable distributions. The PYTHIA prediction cannot reproduce the normalisation but delivers a good description for shape comparisons. Therefore the widely spread use of PYTHIA to obtain all relevant corrections seems to be justified.

MC@NLO which uses HERWIG for the modelling of the parton shower shows good agreement with data. But the predictions of POWHEG using HERWIG reveal large difference compared to predictions in which the parton shower was modelled with PYTHIA. So far no explanation for this has been found.

At a unprecedented center-of-mass energy of 7 TeV higher order processes become more important and measurements may indicate deficiencies in the theoretical calculations and simulation. The measurement on $B\bar{B}$-hadrons from CMS and the deviations seen in the lowest bin of $\Delta\phi$ indicate such deficiencies.

The measured dijet mass spectrum of $b\bar{b}$ dijets extends to around 1.5 TeV (cf. Fig. 3.5). The presented cross sections which make use of the SV0 b-tagging algorithm only cover partly this available statistics. Solely the $\Delta\phi$ cross section contains all these events. However from the agreement found between NLO calculations and data, it can be concluded that no evidence for new phenomena has been found. This will be quantified more precisely in Chapter 12.

11.2 Ratio measurement between $b\bar{b}$ dijets and inclusive dijets

This section presents the inclusive dijet cross section and the ratio measurement between $b\bar{b}$ and inclusive dijets. Both measurements as a function of dijet mass are presented

in Fig. 11.5. The top plots in this figure on both sides show the measurement overlaid with predictions from POWHEG and PYTHIA. The plots in the lower part on both sides present the ratio between data and the corresponding MC prediction. The statistical uncertainty from the measurement in data and from the MC predictions are propagated assuming uncorrelated errors. The PYTHIA prediction for the inclusive cross section has small statistical errors, too small to visualise them on the ratio plot. The prediction from PYTHIA is scaled by 0.65 to the integrated cross section in data for the inclusive cross section on the left side.

The inclusive dijet cross section is a steeply falling function of the dijet mass. The cross section spans over almost four orders of magnitude. The measurement is restricted to the phase space region explored in the $b\bar{b}$ measurement. The dijet mass spectrum in data extends to more than 2 TeV. The jet energy scale uncertainty dominates the systematic uncertainty on the measurement. This uncertainty accounts to 15% on average. The prediction from POWHEG overestimates the cross section in the lower and medium dijet mass region by 20%. Within the systematic theoretical uncertainty the POWHEG prediction and data are in agreement. The scaling factor of PYTHIA is even further away from unity than in the case of the $b\bar{b}$ cross section. It is eye-catching from the comparison of Fig. 11.1 (right bottom) and Fig. 11.5 (right bottom), that the PYTHIA prediction of the $b\bar{b}$ cross section is better than the prediction of the cross section of inclusive jets. Contribution to this fact is, that for $b\bar{b}$ production in most of the parton shower generators the flavour excitation matrix element is included in the hard process. So part of the additional processes contributing at NLO can be accounted for in PYTHIA. The deviation in PYTHIA for inclusive jets ranges from a deficit of 5% to an overshoot of 30% at high dijet masses.

The ratio measurement on the right side of Fig. 11.5 shows a fraction of $b\bar{b}$ dijets of about 0.3% to 1% decreasing with increasing dijet mass. The systematic uncertainty on the measurement is dominated by b-tagging related uncertainties and amounts to 40% to 60%. The prediction from PYTHIA is not scaled. In the measured phase space the major contribution to the inclusive dijet cross section comes from gg, gq initial states which dominate until far above 1 TeV whereas the $b\bar{b}$ production is dominated by the gluon-fusion process. With increasing dijet mass the fraction of light-quark annihilation ($q\bar{q}$) processes leading to $b\bar{b}$ final states increases by three (cf. Fig. 3.11(b)). Simultaneously the fraction of general quark initial states ($q\bar{q}, qq, \bar{q}\bar{q}$) in which all possible channels contribute becomes more important in the inclusive case. Therefore the drop of the fraction of $b\bar{b}$ pairs is explained by the influence of the parton densities.

The ratios between data and MC predictions for the fraction of $b\bar{b}$ dijets show that both MC simulations underestimate this fraction. The POWHEG prediction is on average 50% below data in the lower and medium dijet mass region. Above 450 GeV the prediction and data seem to converge. The deviation at around 400 GeV seen before in the $b\bar{b}$ cross section has a strong influence in the ratio measurement. It causes a deviation between once and twice the uncertainty in both MC predictions. This exceptional behaviour is already present in the raw data spectra for each trigger. The fraction of $b\bar{b}$ dijets predicted by PYTHIA seems to be on average 20% to 30% underestimated.

Such a ratio measurement is another approach sensitive to deviations between data and QCD. Any model of new phenomena bears the possibility to be sensitive to the different quark generations and to favour one over the other in its decay modes. The relative fraction between inclusive jets and a chosen type of jets is a excellent contact point. However, in the measured phase space the QCD predictions are able to describe data.

From an experimental point of view the advantage of a ratio measurement is that it
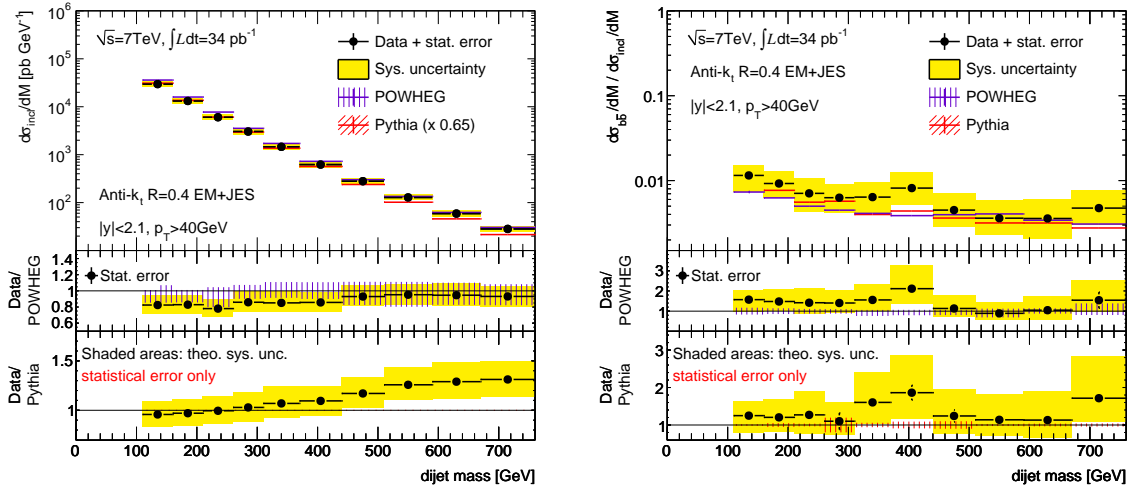
Figure 11.5: *The inclusive dijet cross section as a function of dijet mass with $p_T > 40$ GeV and $|y| < 2.1$ is presented on the left hand side. On the right hand side the ratio between the dijet cross sections of b-jets and inclusive jets with $p_T > 40$ GeV and $|y| < 2.1$ is presented. The data are compared to MC predictions from POWHEG and PYTHIA. Only the central values of the theoretical predictions are drawn. The leading-order PYTHIA prediction is scaled to the measured integrated cross section in case of the inclusive dijet cross section. The lower parts of the plots present the corresponding ratios between data and the MC simulations. The shaded areas around the MC prediction indicate the systematic theoretical uncertainty (except for PYTHIA). The statistical error is calculated from the uncorrelated statistical errors from MC and data.*

reduces theoretical and experimental systematic uncertainties. This is exploited in case of the theoretical uncertainties. But the reduction in case of data is marginal. This comes from the overwhelming contribution of b-tagging related uncertainties which are present exclusively on the $b\bar{b}$ measurement. As long as these dominate, the effect of the reduction and cancellation of uncertainties like luminosity and jet energy scale uncertainty on the total uncertainty is only at a few per cent.

12 Search for resonances in the cross section

The measurement of the dijet cross section of b-jets as a function of dijet mass has shown that the production of pairs at high-energies from two incoming hadrons is well described in terms of QCD in the Standard Model. The dijet mass spectrum predicted from QCD falls smoothly with increasing dijet mass. Many extensions beyond the Standard Model predict new massive particle states, which are localised at a given mass and could be observed as resonances in the dijet mass spectrum. In addition a new force manifesting itself at very high centre-of-mass energies could enhance the event rate.

Models in which quarks are composite particles [20] predict excited quarks states which decay into dijets. Models which embed the QCD into a large gauge group and higher symmetries often contain yet unobserved heavier gauge bosons [143, 16, 17]. Also models of extra dimensions [21] can manifest itself in an excess of dijet events at an invariant mass.

To test the measured mass spectrum, a statistical method sensitive on the presence or absence of a resonance in the mass spectrum is used. In this method the region above 760 GeV which had to be ignored in the cross section measurement is also exploited. Recent ATLAS results on the search for resonances in the inclusive dijet spectrum have not found an evidence for a resonance signal [144]. Without the observation of a signal from new physics, an exclusion limit on their production can be obtained. Recently ATLAS excluded most of the important models with new physics below about 2 TeV [145].

In the following this statistical method and its application on the $b\bar{b}$ cross section measurement is explained. Changes to the event selection and the trigger efficiency as a function of the dijet mass for the search are briefly presented. Finally, a model independent limit on the event yield of a new resonance is determined from data and put into context of a hypothetical signal of a new resonance predicted from a model beyond the Standard Model.

12.1 Statistical method

The goal of the statistical method, called BUMPHUNTER [146], is to find the most significant excess of events above background (bkg) using a window of varying width. Such localised effects are colloquially called *bumps*. The advantage of this method, is that it respects the trial factors (also called *look elsewhere effect*). In case of any hypothesis test where the data is tested for its consistency to a hypothesis, the trial factors account for the possibility that any bin (= trial) can contain a discovery by mere fluctuations, reflecting the fact that the resonance has an unknown mass. A brief description of BUMPHUNTER is given and its application to the $b\bar{b}$ dijet mass spectrum is explained.

12.1.1 The BUMPHUNTER method

The BUMPHUNTER algorithm is in detail explained in Ref. [146] in the context of other hypothesis tests. The brief description here focuses on the aspects important for this investigation. BUMPHUNTER is a hypothesis test, which quantifies the consistency of a hypothesis with an observation. A discovery, which is generally a disqualification of a null hypothesis, is usually claimed in case of the inconsistency to a background hypothesis H_0 (=null hypothesis) like the Standard Model. But the statistical method only establishes the inconsistency of data to the hypothesis H_0, it does not address the amount of signal residing in the bump. So in general any hypothesis test follows three steps:

1. The data D are compared to a hypothesis H_0, resulting in a quantity t_0, called *test statistic*, which describes the differences between both. By convention a bad agreement is given by a large t_0.

2. Pseudo-data are generated allowing for bin-wise Poisson fluctuations, following the H_0 hypothesis. For each pseudo-data spectrum the *test statistic t* is calculated which quantifies the difference between the pseudo-data and the hypothesis H_0.

3. Then the probability is calculated that the *test statistic t* is equal or greater than the *test statistic t_0* from data assuming the hypothesis H_0. This probability is referred to as *p*-value:

$$p\text{-value} \equiv P(t \geqslant t_0 | H_0). \tag{12.1}$$

This *p*-value is not a statement about how likely H_0 is, but it is identical to the probability to mistakenly rule out the background hypothesis (Type-I error).

Besides the observed data distribution D and the background hypothesis H_0, the BUMPHUNTER method requires a minimal set of information to calculate the *test statistic t*. The minimum and maximum number of consecutive bins is required in which the consistency of data is checked against the background hypothesis. These numbers define the range of widths allowed for the search window. The mass region measured in data is determined on which BUMPHUNTER evaluates the *test statistic*. Quality criteria on an excess found in data may be imposed, for example that an excess has to be surrounded by non-discrepant regions. With these information the BUMPHUNTER algorithm calculates the *test statistic* and the *p*-value with the following steps:

- Iterate over the mass range by shifting the configured search window and varying the search window between the minimum and maximum width. For each window width and for each mass, the data or pseudo-data events d are compared with the background events b. For each window i the *p*-value$_i$ following this prescription is calculated:

$$p\text{-value}_i = \begin{cases} 1 & \text{if } d_i \leqslant b_i \\ \Gamma(d_i, b_i) = \sum_{n=d_i}^{\infty} \frac{b_i^n}{d_i!} e^{-b_i} & \text{otherwise} \end{cases} \tag{12.2}$$

$$\tag{12.3}$$

- Among all calculated p-values find the smallest value and set the *test statistic* of the BUMPHUNTER method according to:

$$t = -\log p\text{-value}_{\min}, \tag{12.4}$$

where the $p\text{-value}_{\min}$ is the smallest of all p-values found in the previous step. The mass region with the smallest p-value is the window with the smallest probability, that the deviation from the background is caused by a statistical fluctuation.

The output of BUMPHUNTER consists of the interval with the most interesting properties (smallest p-value), the p-value according to Eq. 12.1 and the significance $\sigma_p = \sqrt{2} \cdot \text{Erf}^{-1}(1 - 2 \cdot p\text{-value})$ of this p-value which are corrected for the trial factor.[1] A negative significance corresponds to a $p\text{-value} > 0.5$ and significant deviations are characterised by small p-values (< 0.001), corresponding to significances above 3.

Systematic uncertainties on the background may shift the background and decrease the significance of any bump. To account for these uncertainties BUMPHUNTER performs pseudo-experiments in which the background is allowed to shift according to a Gaussian distribution with width of the uncertainty. In case of several uncertainties, they need to be merged into one uncertainty by adding them in quadrature. The BUMPHUNTER algorithm searches for the maximum of the product of two probabilities, namely from the most interesting bump and from the Gaussian probability for observing the shifted background.

12.1.2 Application to $b\bar{b}$ dijet mass spectrum

The investigation on significant deviations from the Standard Model (=background hypothesis) is performed on the dijet mass spectrum of b-jets with a minimum window size of one mass bin. In general the maximum window size corresponds to half the mass range spanned by the data. No quality requirement on the surrounding region of an observed excess is set.

The background normalisation and shape is extracted from the observed spectrum in data by fitting a smooth function [19]

$$f(x) = p_0(1 - x)^{p_1} x^{p_2 + \ln(x) \cdot p_3}, \tag{12.5}$$

where $x = M/\sqrt{s}$ and $p_{\{0,1,2,3\}}$ are free parameters. The parameters are constrained to describe the behaviour of the dijet cross section. The function falls monotonically and describes the behaviours in the limit $f(0) = +\infty$. Although this function is not inspired by theory, it has proven to describe the dijet mass spectrum well [144].

To properly use the BUMPHUNTER method, data and background distributions are required to follow the Poisson statistics. This means that the spectrum with prescale-corrected events is not applicable to this hypothesis test. Therefore, the prescale-corrected spectrum is only used to extract the background hypothesis H_0 with the above mentioned fit. The result from the fit is subsequently corrected back to raw background events by scaling bin-wise with an effective prescale $ps_{\text{eff}} = \frac{N^{\text{data}}_{\text{uncorrected}}}{N^{\text{data}}_{\text{corrected}}}$ obtained from data. The asymmetric

---

[1]The function Erf^{-1} is the inverse of the error function. The error function maps the p-value to the significance σ_p and is given by: $p\text{-value} = \int_{\sigma_p}^{\infty} \frac{1}{\sqrt{2 \cdot \pi}} e^{-x^2/2} dx$. The Erf^{-1} function gets the p-value as input and returns the corresponding σ_p.

statistical uncertainty $N^{+\Delta N_{\text{high}}}_{-\Delta N_{\text{low}}}$ on the raw data events is taken from the 68% confidence-level on Poisson-distributed events $N^{\text{data}}_{\text{uncorrected}}$ [3, 147]:

$$
\Delta N_{\text{low}} = N^{\text{data}}_{\text{uncorrected}} - \frac{1}{2} F^{-1}_{\chi^2} \left(\frac{68.27\%}{2}; 2 \cdot N^{\text{data}}_{\text{uncorrected}} \right),
$$

$$
\Delta N_{\text{high}} = \frac{1}{2} F^{-1}_{\chi^2} \left(1 - \frac{68.27\%}{2}; 2 \cdot \left(N^{\text{data}}_{\text{uncorrected}} + 1 \right) \right) - N^{\text{data}}_{\text{uncorrected}}, \qquad (12.6)
$$

where $F^{-1}_{\chi^2}$ is the quantile of the χ^2 distribution or the inverse of the cumulative distribution of Poisson distributed events.[2] The statistical uncertainty on the prescale-corrected events is calculated from error propagation with trigger prescale ps:

$$
\text{Variance} \left(\sum_{\text{trigger } i} N_i \cdot ps_i \right) = \sum_{\text{trigger } i} \text{Variance}(N_i \cdot ps_i) = \sum_{\text{trigger } i} ps_i^2 \cdot \text{Variance}(N_i), \quad (12.7)
$$

where the $\text{Variance}(N_i)$ is given by the error of Poisson distributed events N_i evaluated at 68% confidence-level (cf. Eq. 12.6). The statistical error on the effective prescales is evaluated from the error propagation of the statistical uncertainties on $\frac{N^{\text{data}}_{\text{uncorrected}}}{N^{\text{data}}_{\text{corrected}}}$.

To demonstrate the BUMPHUNTER method, it is shown that an artificially injected signal, whose cross section is smaller than the QCD cross section, is immediately tracked by the BUMPHUNTER method. An arbitrary background hypothesis is chosen with parameters $p_1 = 1 \cdot 10^{-4}$, $p_2 = 10$, $p_3 = 6.3$ and $p_4 = 0.1$. This background hypothesis is fluctuated bin-by-bin according to a Poisson distribution to simulate a data observation. Two signals following a Gaussian distribution are injected into this simulated data with a cross section of 0.2 per mill of the total cross section:

- With a mean at 480 GeV and a width of 20 GeV.

- With a mean of 1000 GeV and a width of 100 GeV.

Examples of the BUMPHUNTER results for the solely Poisson fluctuated data and for the two data distributions with injected signals are presented in Fig. 12.1(left). In the spectrum without an injected signal the determined most interesting interval should be equally distributed over the dijet mass range and the p-values fluctuating around the value 0.5, which is equivalent to a significance of zero. On the top of the right hand side of Fig. 12.1, the mean value of the p-value distribution as a function of dijet mass after 1000 pseudo-experiments is presented. In the dijet mass spectra with injected signal Fig. 12.1(middle and bottom), the p-values in the pseudo-experiments are not uniformly distributed. They are all below 0.1 ($\equiv \sigma_p = 1.3$) and the most interesting interval is Gaussian distributed around the injected signal mass.

12.2 Search for an excess in the dijet mass spectrum

New resonances in the dijet mass spectrum of b-jets are searched for with the help of BUMPHUNTER. This section presents the adjustments and extensions to the former cross

---

[2]The quantile x_a is the value of the random variable x at which the cumulative distribution is equal to a. The cumulative distribution function $F(a)$ is the probability that $x \leqslant a$: $F(a) = \int_{-\infty}^{a} f(x)dx$, where $f(x)$ is the probability density function. Taken from Ref. [3].
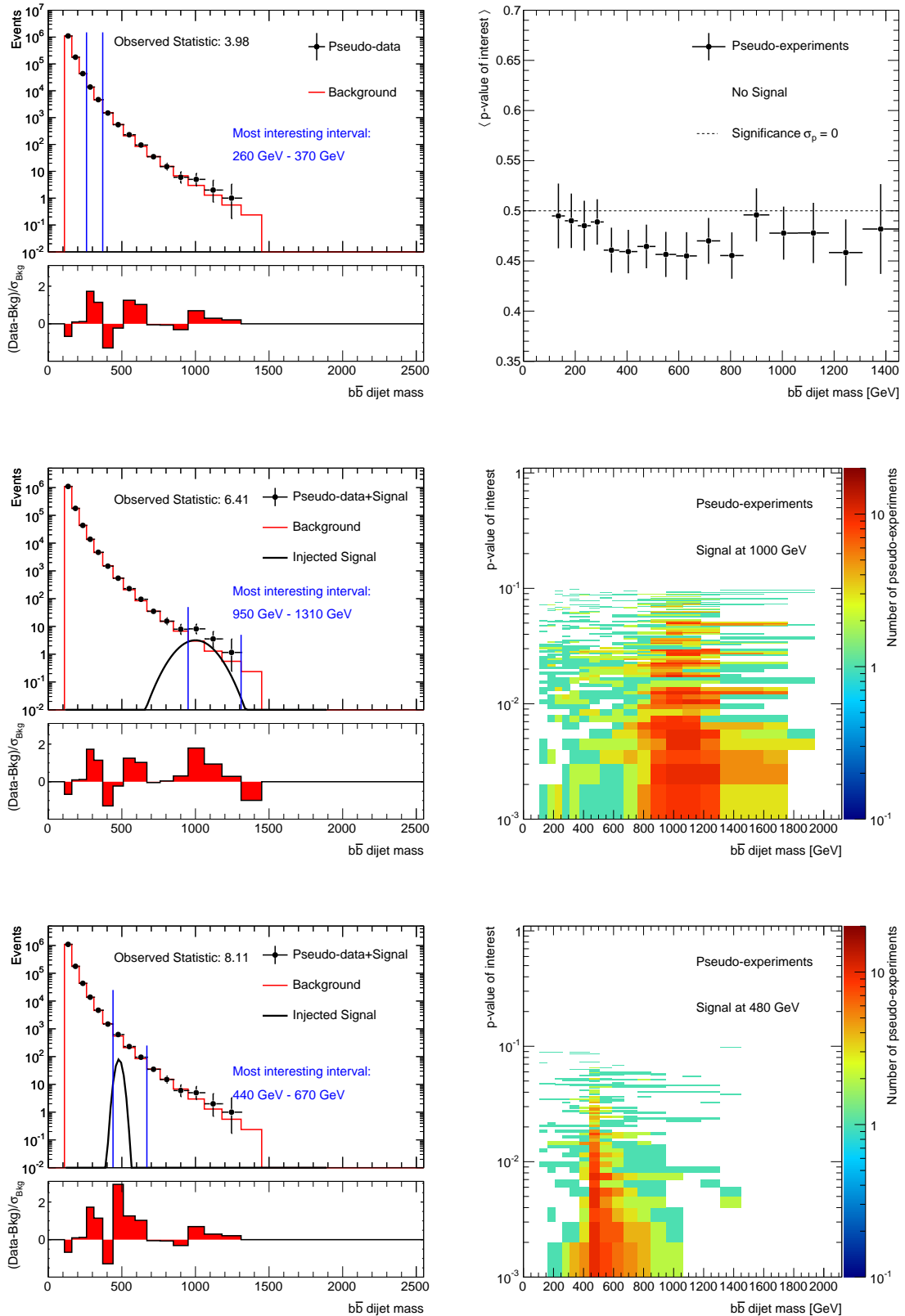
Figure 12.1: *The* BUMPHUNTER *results for pseudo-data generated by Poisson fluctuations with no injected signal (top), with injected signal with a mean at 1000 GeV (middle) and with injected signal with a mean at 480 GeV (bottom). On the left side a typical examples of a* BUMPHUNTER *result is presented. In the top right corner the mean p-value as a function of dijet mass is presented and in the middle and the bottom on the right side the distributions of the p-values as a function the most interesting interval are presented.*

section measurement to increase the sensitivity to new phenomena.

12.2.1 Event selection and trigger

For the searches the BumpHunter method is given the uncorrected mass spectrum without the application of b-tagging related correction factors as well as the bin-by-bin unfolding. This is important as these corrections are capable of concealing new phenomena because they are derived under assumption of QCD only. In addition to the mass spectrum, which has been used to derive the dijet cross section, two other selections have been developed to increase the sensitivity to s-channel processes. A full mass spectrum is derived using all jet triggers but restricting the rapidity correlation to $|\Delta y| < 1$ between the two leading jets. This diminishes the dominating contribution of $b\bar{b}$ pairs from gluon fusion in the t-channel and the selection focuses on the region in which s-channel processes such as decays of heavy resonances are important. This selection criteria is motivated by the rapidity correlation of the outgoing partons in the gluon-fusion process shown in Fig. 3.3(b). Furthermore, to study the higher mass regions with a homogeneous selection, an additional event selection based on the unprescaled trigger with lowest threshold is validated. The trigger efficiency for the L1_J95 trigger is determined using the bootstrap method (cf. Section 7.3) as a function of inclusive dijet mass for the event selection presented in Section 7.4 and using the requirement of $|\Delta y| < 1$. The trigger plateau for L1_J95 is reached at a dijet mass of 440 GeV as can be seen in Fig. 12.2. To account for possible biases between inclusive and b-jets due to different rapidity correlations a safety margin is chosen. The measurement of b-tagged jets is performed for dijet masses above 510 GeV. In summary three searches are
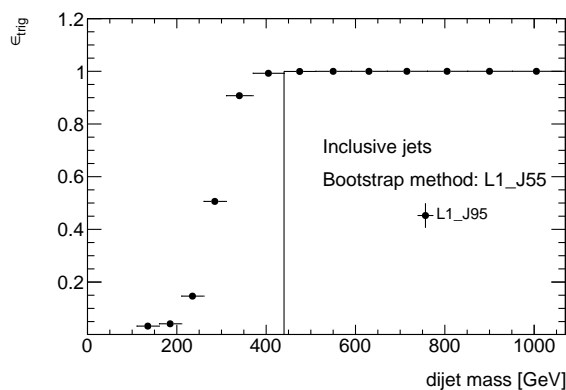
Figure 12.2: *Trigger efficiency of the L1_J95 as a function of inclusive dijet mass determined with the bootstrap method and L1_J55. The events are required to obey $|\Delta y| < 1$.*

performed:

- On the full mass spectrum used in the cross section measurement (≈ 250000 events).

- On the full mass spectrum with an acceptance cut on the rapidity correlation $|\Delta y| < 1$ (≈ 75000 events).

- On the mass spectrum obtained with the unprescaled trigger L1_J95 and an acceptance cut on the rapidity correlation $|\Delta y| < 1$ (≈ 210 events).

12.2.2 Results on the search for resonances

The first step is to extract the background hypothesis with the dijet mass function given in Eq. 12.5. All three spectra with the corresponding fit are presented in Fig. 12.3. The fitted parameter values are presented in Table L.1. The data are well described by the corresponding fit with $\chi^2/\text{DoF} \approx 1$ and with p-values greater than 0.35.[3]

To evaluate the uncertainty on the fit result in each bin, pseudo-experiments are performed in which the data points are fluctuated according to their statistical uncertainty and this new spectrum is re-fitted. The root mean square of all fit results in each bin is determined and is identified with the fit uncertainty. Compared to the statistical uncertainty from the prescale correction, this uncertainty is small and neglected in the following.

The distributions of the BumpHunter statistic can be seen on the left hand side in Fig. 12.4. Besides the broad distribution of the pseudo-experiments, the BumpHunter statistic for data is indicated by the blue arrow. From these distributions the p-value for each event selection can be obtained by using Eq. 12.1. In the full spectrum (top) without any restrictions the significance is sizable in contrast to the full spectrum with the rapidity difference cut (middle) and the mass spectrum measured with L1_J95 (bottom) in which the BumpHunter statistic of data is well below the median of the pseudo-data. The resulting p-values and the corresponding significances are summarised in Table 12.1. The p-value of the full spectrum is 0.02, whereas the other selections have p-values above 0.4 and the distributions are in good agreement with the background hypothesis. On the right hand side in Fig. 12.4 the data distributions and the most interesting intervals in data are indicated by the vertical lines. None of the interesting intervals between those three selections are identical, supporting the conclusion that no significant deviation from QCD is observed. The plots on the bottom of Fig. 12.4 show the difference between observed and expected events normalised to the uncertainty on the background expectation.

Selection	Interval [GeV]	p-value	Significance
Full spectrum	1310-2330	0.02	2.02
Full spectrum with $|\Delta y| < 1$	310-370	0.44	0.14
Spectrum with L1_J95 and $|\Delta y| < 1$	510-590	0.53	-0.06

Table 12.1: *Summary of the* BumpHunter *results for different fitting procedures.*

From this method, it is concluded that no evidence is found for new phenomena in the dijet mass spectrums with different sensitivity to s-channel processes. The deviations seen in the $b\bar{b}$ dijet mass spectrum and in the ratio measurement around 400 GeV are not significant. Due to time constrains the effects of the systematic uncertainties are not evaluated. However, qualitatively speaking their inclusion will decrease the significances even more. The deviations without the uncertainties are already not significant and they

---

[3]DoF stands for the number of degrees of freedom.
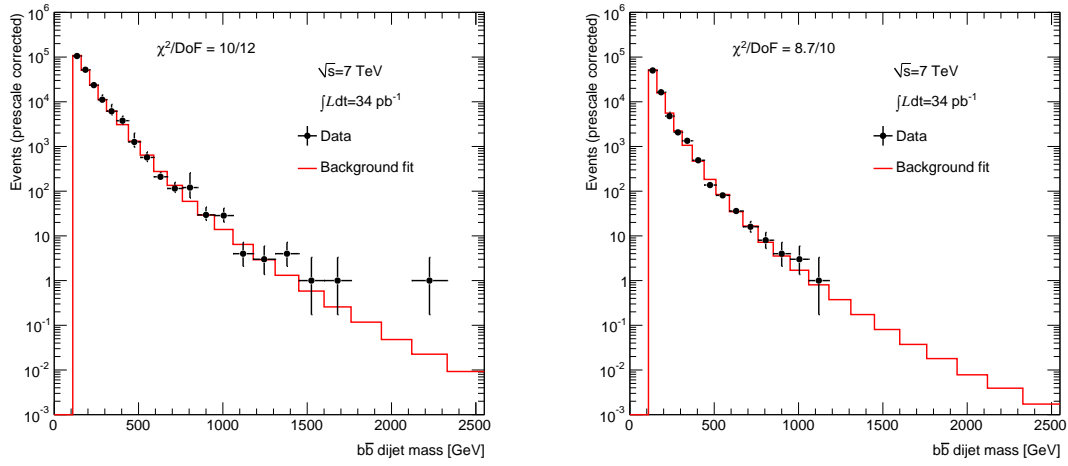
(a) Full spectrum

(b) Full spectrum with $|\Delta y| < 1$
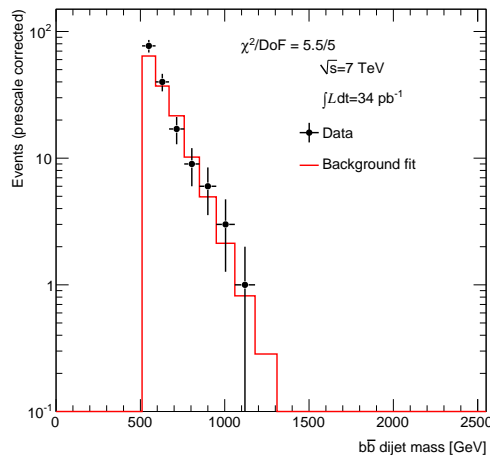
(c) Mass spectrum from L1_J95

Figure 12.3: *Fits on observed mass spectrum in data to obtain the background hypotheses on (a) the full spectrum, (b) the full spectrum with $|\Delta y| < 1$ and (c) the mass spectrum measured with L1_J95.*

will not be afterwards.
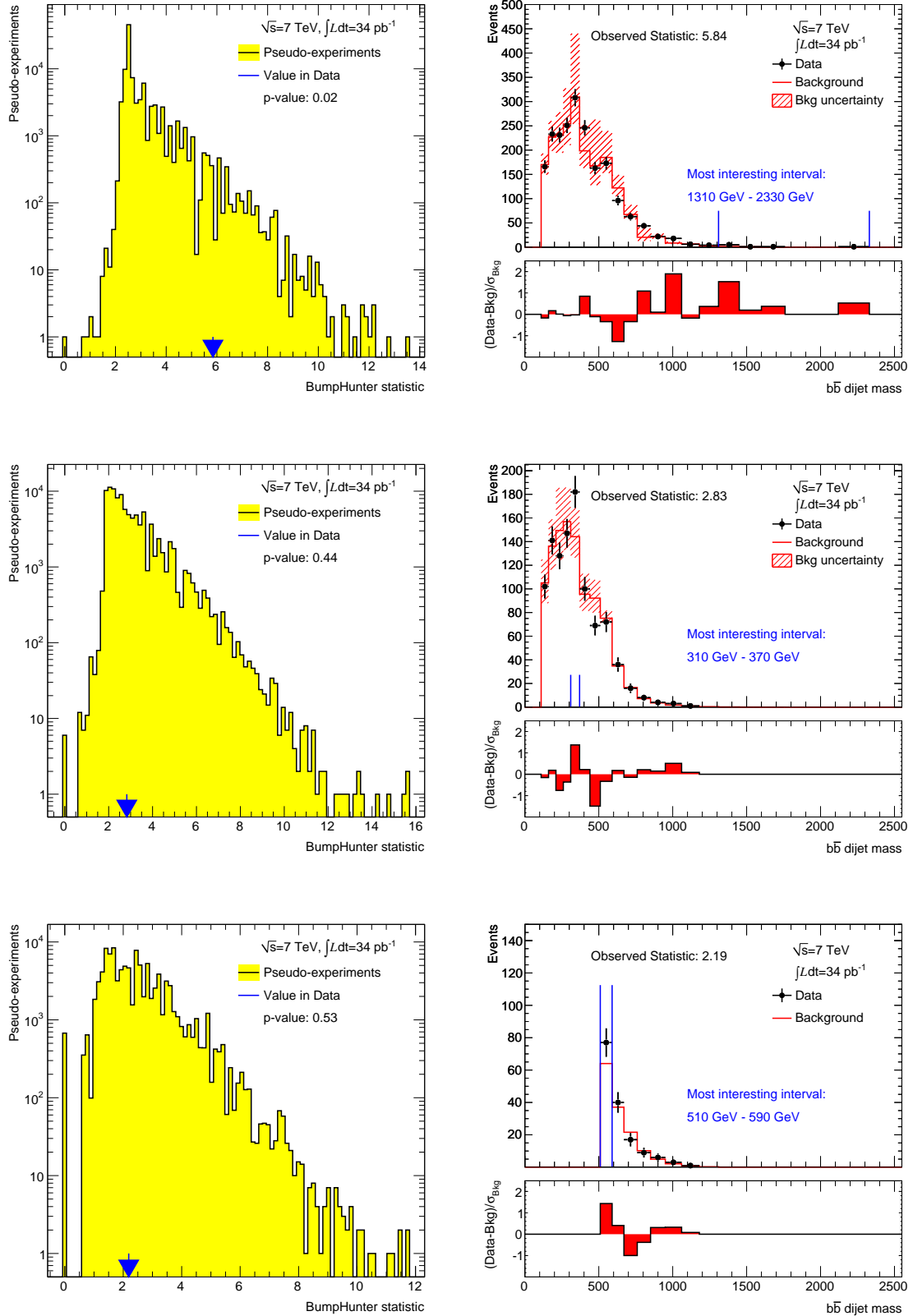
Figure 12.4: *From top to bottom: Full spectrum, full spectrum with $|\Delta y| < 1$ and mass spectrum from L1_J95 and with $|\Delta y| < 1$. On the left side the resulting* BumpHunter *statistics from the pseudo-experiments and the* BumpHunter *statistic observed in data (arrow) are presented. On the right side the mass spectra with most interesting interval are presented. The plots on the bottom show the difference between data and background hypothesis normalised to the total uncertainty on the background.*

12.3 Model independent limit setting

There is no evidence for a resonance signal in the dijet mass spectrum of b-jets. However, from the observed data spectrum in context of the Standard Model (= background hypothesis) it is possible to deduce a limit on the maximum number (= maximum cross section) of events reconstructed from a new resonance which are excluded by the data. Using the assumption that any new resonance is likely to be detected as a Gaussian signal from the calorimeters, a limit can be calculated which is independent from any model. Such a limit could be used to estimate the current sensitivity of the measurement to any model. A recipe for this application is briefly described and finally the model for a heavy gauge boson, called Z', assuming the same couplings to leptons and quarks as that of the Standard Model Z is tested with the help of the model-independent limit [148].

12.3.1 Bayesian limit on generic Gaussian signals

To calculate a limit on a resonance signal with Gaussian shape, Bayesian methods are applied. Further details on Bayesian limit setting can be found in Ref. [149]. The likelihood function $\mathscr{L}_m$ dependent on the signal mass m is given by

$$\mathscr{L}_m(d|b, s_m) = \prod_i \frac{(b_i + s_{m,i})^{d_i}}{d_i!} e^{-(b_i + s_{m,i})}, \tag{12.8}$$

where the product runs over all bins i. The likelihood function describes the probability to observe d data events with the condition of b background events and s_m signal events from a massive resonance with mass m. This likelihood function is maximised with respect to the data events. Assuming a flat prior for all possible number of signal events, the 95% confidence-level (CL) to exclude a signal with strength s_m or bigger is calculated by integrating the likelihood over the prior. The 95% CL upper limit is parametrised by the mass of the resonance and this limit is usually dependent on the signal width. This limit calculates the maximum number of events of a given resonance which are at 95% confidence-level excluded. This is then usually expressed in $s_m(\text{excluded})/L = \sigma \times BR \times Acc$, where BR stands for the branching ratio of the resonance into $b\bar{b}$. In case of a measurement which is subject to a selection efficiency as the b-tagging algorithm, the result of the Bayesian limit is identical to $s_m(\text{excluded})/L = \sigma \times BR \times Acc \times \text{efficiency}$.

A hypothetical resonance which decays into all flavours of quark-antiquark pairs can have a non-zero branching ratio into top-antitop pairs if $m_{Z'} > 2m_t$. The $b\bar{b}$ dijets from these decays are also likely to be identified by the b-tagging algorithm. Depending on the transverse momenta of the top- and antitop-quark, which increase with increasing Z' mass, the decay products become more collimated. As long as they are well separated ($m_{Z'} \gtrsim 2m_t$), the invariant mass distribution of the $b\bar{b}$ dijets does not reflect the invariant mass of the resonance due to the decay of the top. This solely leads to a broad distribution which would contribute to the Gaussian signal mainly in the tails. If the resonance mass is in the region of 1 TeV, merged objects become more likely and the invariant mass shifts towards the resonance mass.

Besides these irreducible top-antitop events, signal events from light-flavour jets (u, d, s) and c-jets can be mis-tagged. A simple calculation shows that the fraction of mis-tagged events in the signal is negligible: Assuming SV0 b-tagging performance and a signal with a cross section of 10 pb which decays democratically into all types of

quarks ($m_{\text{sig}} > 350$ GeV), the fraction of events reconstructed from this signal consist of: $\sigma_{\text{sig}} \cdot L_{ATLAS} \cdot \frac{BR_{\text{light}} \cdot \epsilon_{\text{mistag}} + BR_c \cdot \epsilon_c}{BR_{b\bar{b}} \cdot \epsilon_{b\bar{b}}} = \frac{1}{17} \ll 1$. It should be noted that their effect would only be beneficial for the limit calculation in the first place, so disregarding them makes the calculation conservative.

In summary, in the core region of the Gaussian signal events from top-antitop decays are negligible and the fraction of signal events which are mis-tagged is small. So the limit can be simply expressed in terms of $\sigma \times BR \times Acc \times b$-tagging, where the *b-tagging* term stands solely for the *b*-tagging efficiency.

The limit calculation uses the measured prescale-corrected dijet mass spectrum with the rapidity correlation selection cut $|\Delta y| < 1$. The background hypothesis is again extracted from a fit with Eq. 12.5 to data. The systematic uncertainties affecting the Bayesian limit are the statistical uncertainty on the data points arising from the prescale correction, the uncertainty on the luminosity, the uncertainty on the signal templates from the jet energy scale and the jet energy resolution.

Each data point is shifted according to the statistical uncertainty and the background fit is redone. From the sample of fit results the root mean square in each bin is taken as systematic uncertainty. The systematic uncertainty on the luminosity is 3.4%. The systematic uncertainties related to the jet reconstruction affect the signal templates only. The signal shape for each signal mass is varied according to the jet energy scale uncertainty and the uncertainty on the jet energy resolution. These uncertainties are evaluated in the limit setting by varying all the sources according to a Gaussian probability distribution and convoluting these with the likelihood function.

The signal templates for masses M between the 200 GeV and 1300 GeV are generated with six different signal widths $\sigma(M)$. The signal widths are expressed in multiples of the detector resolution. The signal widths chosen are: σ_{detector}, $2\sigma_{\text{detector}}$, $2.5\sigma_{\text{detector}}$, $3\sigma_{\text{detector}}$, $3.5\sigma_{\text{detector}}$ and $5\sigma_{\text{detector}}$. The detector resolution has been calculated with the help of PYTHIA QCD simulation samples with a full-simulation of the ATLAS detector. The resolution as a function of dijet mass is fitted and this parametrisation is used to determine the signal width as function of multiple detector resolutions. The resulting parametrisation can be found in Appendix Fig. L.1. For signal widths of the detector resolution the ratio $\sigma(M)/M$ ranges from 5% to 15% for high and low resonances mass respectively.

To estimate the shift of the signal template due to the jet energy scale uncertainty, it is assumed that the emerging jets from the decay are balanced back-to-back in transverse momentum ($p_{T,1} \approx p_{T,2} \approx M/2$) and the uncertainty on the transverse momentum is the maximum jet energy scale uncertainty determined in the rapidity region $|y| < 2.1$. The resulting shift of the core region of the Gaussian signal is between 3.5% and 4.8%. The variation from the uncertainty on the jet energy resolution is estimated to be 15% for all signal masses. The uncertainty from the jet energy resolution causes the signal to become narrower or broader. An example of a resulting signal template with a resonance mass of 350 GeV and a signal width of twice the detector resolution and its variations is presented in Fig. 12.5(a). The resulting observed limit in data from these signal templates as a function of the invariant dijet mass for six different signal widths is shown in Fig. 12.5(b). The expected limit and the 68% uncertainty band for a signal template with a signal width of the detector resolution is calculated from the background distribution using statistical fluctuations. The area above each curve is excluded from data. The curves are falling with increasing dijet mass. In general, it is expected that a signal template with a broad signal width has a worse exclusion limit than a signal template with a more narrow signal width. Consequently, the curves should be ordered from top to bottom with decreasing

signal width. However, fluctuations in the same direction in data, which extend over more than one bin, can change this. If two or more bins show a deficit, then the limit on a signal with broad signal width is improved. As can be seen in Fig. E.3(b) between 200 GeV and 500 GeV there a neighbouring bins which fluctuate in the same direction and their appearance can be tracked in the limit curves as well. The expected limit has been calculated with a small number of pseudo-experiments due to time constraints. This makes the uncertainty band less smooth and lets the error band increase in the last bin due to this small statistics.

These limits can be used for any new physics model with a new massive particle decaying

(a)   (b)
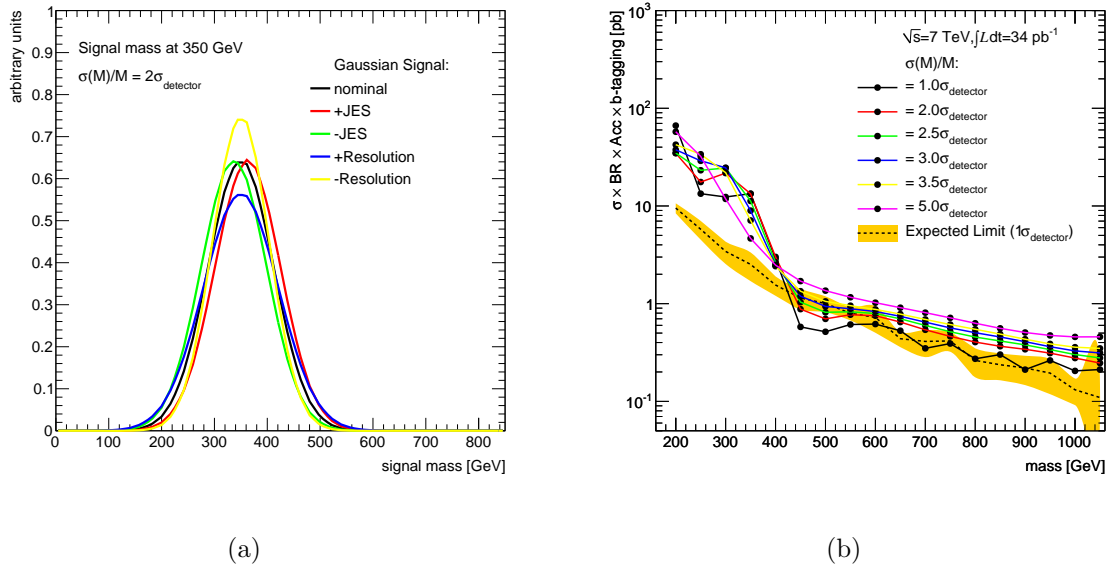
Figure 12.5: *(a) Example of assumed signal shape at an resonance mass of 350 GeV and a width of twice the detector resolution with the variations in the shape from the systematic uncertainties jet energy scale (JES) and jet energy resolution. (b) The 95% CL upper limits on $\sigma \times BR \times Acc \times b$-tagging observed in data for simple Gaussian resonances decaying into $b\bar{b}$ dijets as a function of the invariant dijet mass. The expected limit (dashed line) and the 68% uncertainty band are computed using statistical uncertainties alone.*

into $b\bar{b}$ dijets. The signal events and corresponding reconstructed jets from a new resonance have to be selected according to the event selections applied to obtain the limit. The dijet mass spectrum for these events need to obtained. The dijet mass distribution has to be convoluted with the detector resolution and resulting tails have to be removed because a Gaussian shape has been assumed. From the resulting Gaussian core of the resonance the acceptance of the signal has to be calculated and for the core the b-tagging performance has to be estimated. Then the theoretical value for $\sigma_{\text{Theory}} \times BR_{\text{Theory}} \times Acc_{\text{Theory}} \times b$-tagging$_{\text{Theory}}$ can be determined and conservatively compared to the observed limit in Fig. 12.5(b) by choosing the limit curve with corresponding signal width by rounding up and the nearest mass point from rounding down. This procedure is used in the following section to obtain the theoretical curve for a heavy gauge boson decaying into $b\bar{b}$.

12.3.2 Limit on new heavy gauge boson Z'

The model-independent limit which was obtained in the previous section is compared to a signal from a Z' decay. This hypothetical gauge boson refers to a new massive copy of the well-known Z boson in the context of the Standard Model. MC events containing Z''s are generated with PYTHIA without the full-detector simulation under the assumption of identical couplings to leptons and quarks as the Z boson. The resonance width of the Z' therefore increases linearly with its mass: $\Gamma_{Z'} = \left(\frac{M_{Z'}}{M_Z}\right)\Gamma_Z$.

Firstly, a range of Z' masses is generated with $\sigma_{\text{Theory}} \times BR_{\text{Theory}}$ and the events are selected in accordance to the data in the observed limit. For all masses the generated mass, the dijet mass of anti-k_t $R = 0.4$ jets on particle level and on calorimeter level are obtained. The dijet mass at calorimeter level reflects the energy of particles visible to the calorimeters. In Fig. 12.6(a) all three masses for a Z' at 420 GeV are presented. The generated mass with its small width describes a Breit-Wigner function. The reconstructed dijet masses of jets with resolution parameter $R = 0.4$ at particle level already measure a smaller mass due to particle losses from the hadronisation as well as a more distinct tail for smaller masses. The exclusion of neutrinos and muons which are not part of the visible energy at calorimeter level make these effects even worse. The next step is to convolute the reconstructed dijet mass with the detector resolution using pseudo-experiments in which each jet is smeared according to the resolution presented in Fig. 8.23(a). As can be seen in Fig. 12.6(b) for a Z' signal at 420 GeV the detector resolution shifts the asymmetric distribution to smaller masses and makes the signal symmetric. In the core region the resonance looks like a Gaussian distribution. The mass window in which the signal is Gaussian is extracted and is used as an additional selection cut on the signal. The events which pass this additional selection cut are used to calculate the acceptance Acc_{Theory}. Because the jet calibration applied to data is determined from truth jets at calorimeter level, it is a valid approximation to treat these smeared jets as pseudo-data.

To estimate the b-tagging performance, the full-simulated QCD PYTHIA events are used. The b-tagging efficiency for the SV0 algorithm is obtained as a function of jet p_{T} only for b-jets originating from the hard scatter to suppress the b-jets from GSP which are not contained in the signal sample. For the rapidity region $0.8 < |y| < 1.2$ the b-tagging efficiency is shown in Fig. 12.7(a) as a function of dijet mass. The efficiencies are fitted by a Landau function because all efficiency curves favour this shape and the obtained χ^2's are close to one. The b-tagging efficiencies are slowly decreasing with increasing rapidity. All curves are generally rising in the lower p_{T} region to a maximum between 0.5 to 0.7 at around 200 GeV and decreasing with increasing p_{T} to about 0.2 to 0.3 at around 800 GeV. Each event in the mass window is weighted by the product of both efficiencies for each jet. From the unweighted dijet mass spectrum and the efficiency-weighted spectrum, which are shown for a Z' mass at 420 GeV in Fig. 12.7(b), the b-tagging efficiency ϵ_{Theory} is obtained. The effective b-tagging efficiency for the case of the Z' signal at 420 GeV as shown in Fig. 12.7(b) is about 0.33.

The effective b-tagging$_{\text{Theory}}$ values for all generated dijet masses are presented in Fig. 12.8 and summaries in Table L.2. The b-tagging efficiency as a function of signal mass shows the same behaviour as the b-tagging efficiency as a function of p_{T}, it rises to around 0.35 at a mass of 350 GeV and then decreases to values smaller than 0.18 at 1300 GeV. This is due to the rising transverse momentum of the decay products.

The resulting theoretical values of $\sigma_{\text{Theory}} \times BR_{\text{Theory}}$ and Acc_{Theory} for all generated mass points are presented in Appendix as Fig. L.2. The values of $\sigma_{\text{Theory}} \times BR_{\text{Theory}}$ starting
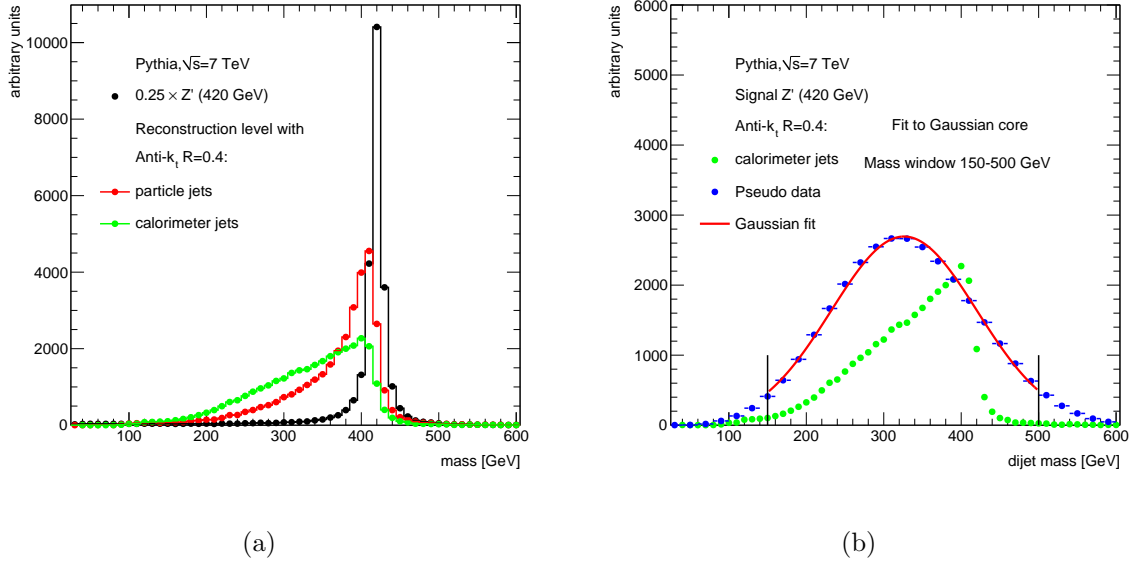
(a)   (b)

Figure 12.6: *(a)* Pythia *Z′ signal at various levels: generated Z′ mass at 420 GeV, after applying the event and jet selection of anti-k_t jets at particle level and at calorimeter level. (b) Generating from the reconstructed Z′ signal at calorimeter level pseudo-data which simulate the detector resolution. The core region with Gaussian shape of this pseudo-data is extracted via a fit and determines the mass window.*

at around 100 pb decrease rapidly with increasing signal mass over more than 4 orders magnitude until the highest signal mass of 1200 GeV. The acceptance falls with increasing signal mass from 20% to 10%. This has several reasons. The production of Z' off the resonance peak increases with increasing mass due to the sea quark PDF and the fraction of events in the mass window, which defines the core region with a Gaussian shape, decreases with increasing mass. A sample of Toy-MC predictions for three different Z' masses are plotted with the observed data dijet mass spectrum in Fig. 12.9(a). With increasing signal mass the signal widths and the shift between the generated and reconstructed mass with detector resolution increase rapidly. For the smallest signal mass in Fig. 12.9(a) the measurement observed more than 10000 events in the mass bin of interest and the predicted number of events are 450.

With these numbers the theoretical curve of $\sigma_{\text{Theory}} \times BR_{\text{Theory}} \times Acc_{\text{Theory}} \times b\text{-tagging}_{\text{Theory}}$ is obtained and compared to the observed limit in data. This is presented in Fig. 12.9(b). The points of the Z' signals represent the reconstructed dijet mass, not the generated resonance mass. In Table L.2 the generated and the corresponding reconstructed masses are presented. Any exclusion region obtained has to be translated into generated resonance mass. The Z' signal points are coloured corresponding to the signal width obtained in the Toy-MC. For the Z' signal with smallest resonance mass, the signal width already is more than twice the detector resolution. It can be seen that with the available data set the Z' with identical quark couplings as the Z cannot be ruled out in any mass region. In special cases in which the Z' coupling to leptons is suppressed or even zero, the sensitivity of this observed limit may be increased. In the hypothetical case in which the branching ratio into $b\bar{b}$ of a new heavy gauge boson is 100%, the sensitivity is increased by a factor of $\approx 100\%/15.12\% \approx 6.6$ [3]. The theoretical values of the Z' signal are shifted up by this
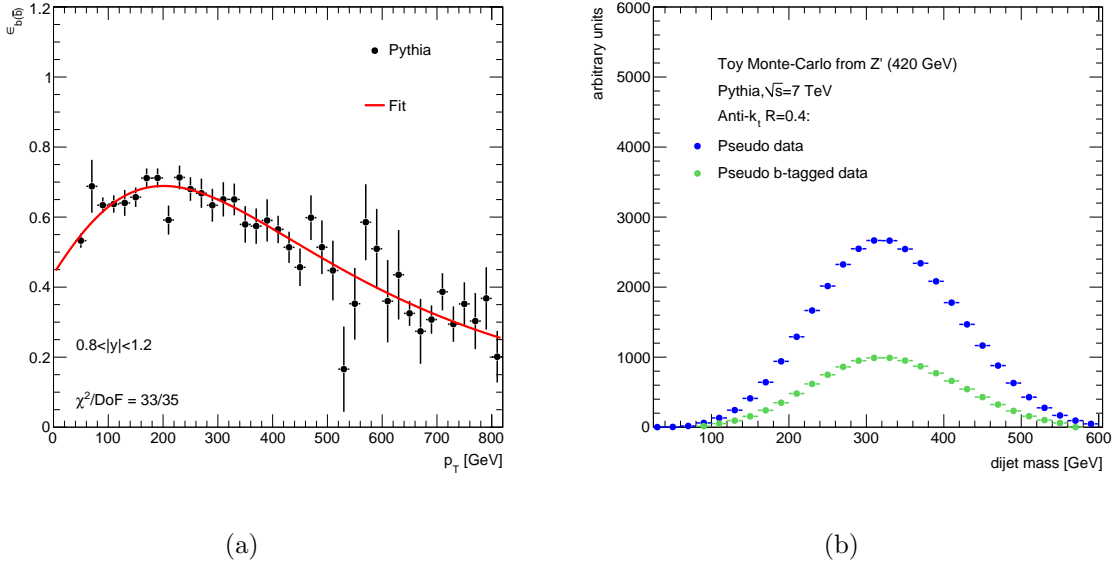
(a)   (b)

Figure 12.7: *(a) From QCD* Pythia *simulation the b-tagging efficiency as a function of jet p_T in four different rapidity ranges is calculated using b-hadrons produced in the hard scatter. (b) The b-tagging performance is simulated on pseudo-data to determine the b-tagging efficiency on signal events and the resulting pseudo-b-tagged data is plotted against the pseudo-data.*

factor in Fig. 12.9(b). In this case an exclusion limit for a heavy gauge boson with a mass around 260 GeV can be found. This exclusion region obtained from Fig. 12.9(b) is already translated into terms of Z' mass using Table L.2.

Compared to exclusion limits with charged leptons, the experimental sensitivities in the hadronic decay channels are much weaker. The sensitivity in the lepton sector arises from trigger selections and smaller systematic uncertainties on the measurement. With charged leptons ATLAS has excluded a Z' with the same lepton and quark couplings as the Z boson up to 2.21 TeV [150]. Measurements on the hadronic decay channels using inclusive dijets have excluded the following regions: $M_{Z'} > 237$ GeV [151] and 320 GeV $< M_{Z'} < 740$ GeV [19]. A measurement from CDF on a final state with $b\bar{b}$ using data corresponding to an integrated luminosity of 87 pb^{-1} is also not able to exclude the Z' [152]. Their sensitivity is in the same regime as this measurement.

An improvement in the sensitivity in searches with $b\bar{b}$ final states can be achieved with a more advanced b-tagging algorithm. A realistic achievement is an improvement of 50% to 70% in the b-tagging efficiency with constant light-jet rejection (cf. Fig. 6.7). This means roughly an improvement of 40%. However, pseudo-experiments, which estimate the 95% CL upper limit assuming this b-tagging performance, show that with 2010 ATLAS data ($L = 34$ pb^{-1}) the Z' cannot be excluded. Of much more importance, is therefore the amount of integrated luminosity recorded in 2011 data. The expected limit with an integrated luminosity of 5 fb^{-1} is calculated to demonstrate the sensitivity of 2011 ATLAS data. The expected limit for reconstructed resonances with a signal width of 2.5 and 3.5 detector resolutions is presented in Fig. 12.10. These signal widths are chosen due to Fig. 12.9(b) as they are important for masses below 500 GeV. In this limit calculation, the performance of the b-tagging algorithm is assumed to be identical to the SV0 algorithm and the jet energy scale uncertainty is improved by 20%. These are rather conservative
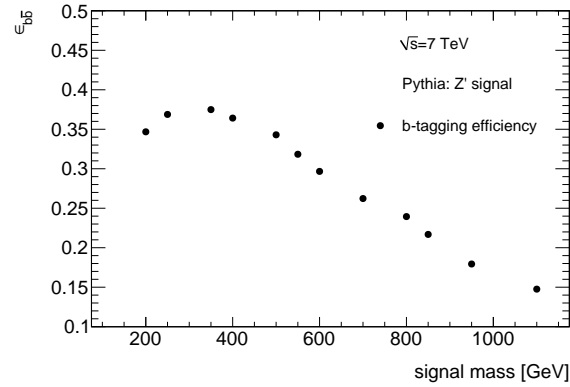
Figure 12.8: *Effective b-tagging efficiency for each generated resonance as a function of signal mass.*

assumptions. The expected limit shows that with the recorded data from 2011, the Z' can be excluded in final states with pairs of b-jets. The exclusion region can be estimated to be around 260 GeV $< M_{Z'} < 700$ GeV for 2011 ATLAS data. This prognosis for the exclusion region obtained from Fig. 12.10 is already translated into terms of Z' mass using Table L.2.

(a)   (b)
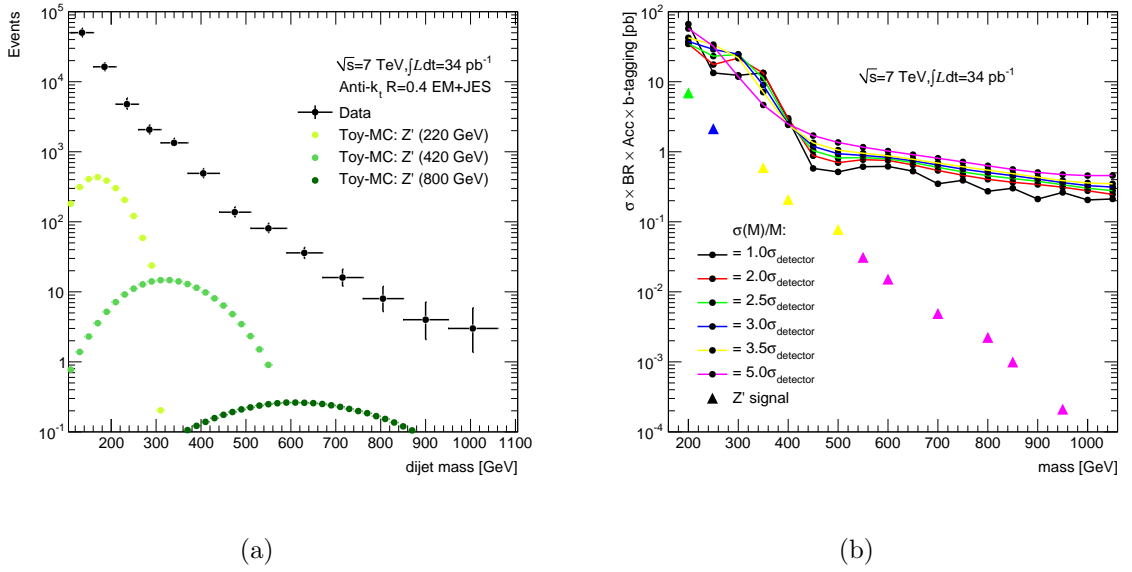
Figure 12.9: *(a) The measured dijet mass spectrum of b-jets is shown with predictions from Toy-MC of* PYTHIA *Z' for masses 220, 420 and 800 GeV using the SV0 b-tagging algorithm. (b) The 95% CL upper limits on $\sigma \times BR \times Acc \times b\text{-}tagging$ for simple Gaussian resonances decaying into $b\bar{b}$ dijets as a function of the invariant dijet mass and the determined $\sigma \times BR \times Acc \times b\text{-}tagging$ values for the heavy gauge boson Z'. The colours of the Z' signals indicate the corresponding signal width obtained in the Toy-MC.*
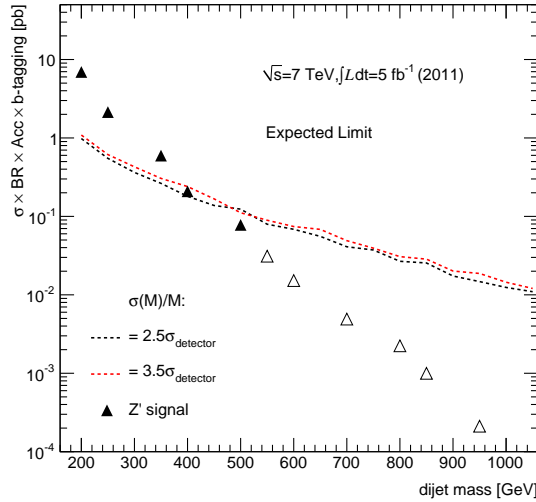
Figure 12.10: *Expected limit for data corresponding to a total integrated luminosity of 5 fb^{-1} which has been recorded by ATLAS in the LHC running period in 2011. The evaluated expected limits are calculated for two signal widths: $2.5\sigma_{\text{detector}}$ and $3.5\sigma_{\text{detector}}$. The empty signal markers indicate that the signal width is larger than the ones for which the expected limit is calculated.*

13 Conclusion and outlook

The Large Hadron Collider (LHC) began regular operation at a centre-of-mass energy of 7 TeV in 2010. The run period in 2010 set up a new world record in the beam energy at a particle accelerator. The commissioning of the LHC offers the opportunity to explore unprecedented kinematic regimes. With this machine precision measurements in the TeV regime of the Standard Model (SM) and the search for new phenomena in the multi-TeV region are achievable. ATLAS is a multi-purpose detector built for high-energy particle physics at the LHC. The full data set recorded by ATLAS during 2010 LHC operation at $\sqrt{s} = 7$ TeV corresponds to an integrated luminosity of 45 pb^{-1}.

This thesis presented the measurement of the production rate of bottom-quark pairs with the ATLAS experiment. It made intensive use of the capabilities of the calorimeters and the precision of the tracking system of the ATLAS detector. The $b\bar{b}$ cross section is determined as a function of the invariant dijet mass, the azimuthal angle between the $b\bar{b}$ pairs and their polar angle correlation in the variable χ. Bottom-flavoured jets are measured above a transverse momentum of 40 GeV and within the rapidity region $|y| < 2.1$. For the cross section measurements the data is corrected to particle level, making it independent of detector effects. It is then compared to next-to-leading order QCD predictions. Physics phenomena beyond the Standard Model of particle physics have been searched for in the $b\bar{b}$ invariant mass spectrum, and the 95% confidence-level upper limits on a resonance signal with a Gaussian shape have been calculated.

Jet production is by far the dominant process at the LHC. In proton-proton collisions, jet studies can therefore usually benefit from small statistical uncertainties. The total jet production rate is of interest to test the QCD description as well as to understand them as a background source in various numbers of analyses. But jets can also be used to search for new phenomena which would manifest themselves in an excess or deficit of events at a given invariant dijet mass or in an excess at high masses. Bottom-flavoured jets are an excellent tool to test QCD. They are the most direct probe of the hard scatter and its kinematic properties. Their production at large transverse momentum is only subject to small contamination from non-perturbative effects. The fragmentation of bottom-flavour hadrons in high-p_{T} jets can be studied by means of perturbation theory since the momenta involved are typically large enough. The ratio of heavy-quark jets relative to generic jets could be an indicator for new physics if the coupling to quarks is sensitive to the quark generation.

For the measurement of jets with bottom-flavour content three ingredients are essential: the absolute scales of the calorimeter measurement of the energies of inclusive jets and of b-jets, and the calibration of the heavy-flavour identification (b-tagging). In ATLAS, the inclusive jet energy scale is established using data-driven techniques and test-beam measurements and used as a baseline in this analysis. For the cross section measurement performed in this thesis, an additional contribution to the jet energy scale uncertainty taking into account the difference in fragmentation between bottom-flavour jets and light-quark and gluon-initiated jets has been derived with the help of the tracking system. The b-tagging calibration is determined with the help of semi-leptonic b-hadron decays into muons. The

resulting scale factors, which adjust the *b*-tagging performance in Monte-Carlo simulation (MC) to comply with data, are shown to be valid assuming a conservative uncertainty in the higher kinematic regions which were aimed for in this measurement. The uncertainties on this measurement are dominated by systematic uncertainties related to *b*-tagging. The second largest uncertainties arise from the jet energy scale uncertainty.

The MC studies have shown that the dijet cross section of *b*-jets is dominated by the production via flavour creation. The next-to-leading-order calculations (NLO) are capable of describing the measured differential cross sections very well. They are also essential to understand the differential cross sections. It has been shown that the QCD production mechanism is also valid in the energy regime of the LHC. However, first evidence for deficiencies in the description of data have been found in the azimuthal angle correlation between *b*-jets. These have been discussed in the context of other discrepancies found at the LHC in the description of *b*-hadron production and modelling. As good as the next-to-leading QCD calculations are, there is still room for improvements. One goal would be to describe higher jet multiplicities at leading order (LO) while maintaining the NLO accuracy for the hard scatter. New approaches are investigated to achieve next-to-leading-order accuracy for parton shower algorithms. However, the measurement and the determination of corrections could have not been so straightforward if not for the good modelling of the underlying event, the multiple parton interactions and hadronisation in the parton shower generators. The studies on tracks within jets have impressively shown the power of the phenomenological models on the non-perturbative processes during hadronisation.

Additionally, the same data was used to search for physics beyond the SM. Three different event selections with different sensitivities to processes in the *s*-channel, which join into an intermediate state, have been investigated for the presence of an excess or a deficit in the invariant mass spectrum. It has been shown that there is no evidence for such new physics. A model independent limit on the event yield of a process forming a resonance with a Gaussian shape with different widths has been calculated at 95% confidence-level. A neutral heavy gauge boson with the same couplings as the Z boson cannot be excluded in any mass region. In scenarios in which the heavy gauge boson exceptionally decays only into $b\bar{b}$ pairs, the sensitivity is increased and may even be excluded using this data set of 34 pb^{-1}.

In 2011 an integrated luminosity of 5 fb^{-1} has been recorded by the ATLAS experiment at $\sqrt{s} = 7$ TeV. After surveying the perspectives for 2012, it has been decided to operate the LHC at $\sqrt{s} = 8$ TeV in 2012. This run period has begun in April 2012. In 2013 the LHC will shutdown to prepare running at design center-of-mass energy of 13-14 TeV. This includes months of repair work and an intensive training of the magnets to reach their maximum magnetic field.

Improvements from the *b*-tagging performance and the jet energy scale uncertainty are expected in measurements using 2011 and 2012 data. High-performance *b*-tagging algorithms can achieve an increase in efficiency from 50% to 70% with the same light jet rejection as the SV0 *b*-tagging algorithm used in this analysis. These new algorithms will allow for refined searches of new phenomena with low production cross sections. Equivalently, the light jet rejection at the point of 50% *b*-tagging efficiency can be increased by a factor of 3. The inclusive jet energy scale uncertainty has been improved during the last year and already reached the 1% level in the central rapidity region for intermediate jet p_{T}. As a consequence a reduction of the jet energy scale uncertainty on the cross section of almost 50% in the low dijet mass region can be achieved in an inclusive jet measurement.

To take advantage of the correlation of experimental and theoretical uncertainties in a cross section ratio measurement the b-tagging related uncertainties have to be reduced to the level of the jet energy scale uncertainties. When this is accomplished, the production ratio of heavy-quark jets relative to generic jets can get very sensitive to new physics. The additional contribution to account for the differences in heavy-flavour fragmentation has already been reduced on the basis of MC studies. This additional uncertainty is about 2% for jets with $p_{\mathrm{T}} \approx 100$ GeV and decreases to below 1% for higher-p_{T} jets. The jet energy scale for b-jets could also be improved and validated using (anti-)top quark decays for the intermediate transverse momentum region. The prospects of the search with the integrated luminosity available in 2011 data have been shown. It can be expected that with 2011 data first limits on the heavy gauge boson with the same couplings as the Z boson in the $b\bar{b}$ decay channel can be extracted.

Because the number of interactions per beam crossing has increased to more than 20 in data in 2011 and to almost 30 in 2012, the analyses face new challenges. But still the performance of ATLAS is expected to increase in terms of calibration and simulation efforts leading to smaller experimental uncertainties. The coming years particle physics will be strongly affected by the results of the LHC and its experiments.

List of acronyms

ALFA	Absolute Luminosity for ATLAS
ALICE	A Large Ion Collider Experiment
AMBT1	ATLAS Minimum Bias Tune 1
ATLAS	A Toroidal LHC ApparatuS
AUET1	ATLAS Underlying Event Tune 1
BC	Before Christ
BKG	Background
BR	Branching Ratio
BS	Beam Spot
CDF	Collider Detector at Fermilab
CERN	Centre Européen pour la Recherche Nucléaire
CKM	Cabibbo-Kobayashi-Maskawa
CL	Confidence level
CMS	Compact Muon Solenoid
CPU	Computer Processing Unit
CSC	Cathode Strip Chambers
CTEQ	Coordinated Theoretical-Experimental Project on QCD
CTP	Central Trigger Processor
DAQ	Data Acquisition
DCS	Detector Control System
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi
DoF	Number of Degrees of Freedom
DQ	Data Quality
EF	Event Filter
EM	Electromagnetic
EMB	Electromagnetic Barrel Calorimeter
EMEC	Electromagnetic End-cap Calorimeter
FCal	Forward Calorimeter
FCR	Flavour Creation
FEX	Flavour Excitation
FONLL	Fixed Order plus Next-to-Leading Logarithms

GSP	Gluon Splitting
GUT	Grand Unified Theory
HAD	Hadronic
HEC	Hadronic End-cap Calorimeter
HLT	High-level Trigger
ID	Inner Detector
JES	Jet Energy Scale
LAr	Liquid Argon
LEP	Large Electron-positron Collider
LHAPDF	The Les Houches Accord PDF Interface
LHC	Large Hadron Collider
LHCb	LHC-beauty
LHCf	LHC-forward
LO	Leading Order
LUCID	LUminosity Cherenkov Integrating Detector
MBTS	Minimum Bias Trigger Scintillators
MC	Monte-Carlo
MCFM	Monte Carlo for Femtobarn Processes
MDT	Monitored Drift Chambers
MNR	Mangano-Nason-Ridolfi
MPI	Multiple Parton Interactions
MRST	Martin-Roberts-Stirling-Thorne
MSTW	Martin-Stirling-Thorne-Watt Parton Distribution Functions
NI	Non-interacting
NLO	Next-to-leading Order
NNLO	Next-next-to-leading Order
NP	Non-perturbative
PDF	Parton Density Functions
POWHEG	POsitive Weight Hard Event Generator
pQCD	Perturbative QCD
PS	Prescale
PV	Primary Vertex
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics

QGSP	Quark Gluon String model
RMS	Root Mean Square
RoI	Region-of-Interest
RPC	Resistive Plate Chambers
SCT	Silicon Microstrip Tracker
SLD	Stanford Linear Collider Large Detector
SM	Standard Model
SV	Secondary Vertex
TGC	Thin Gap Chambers
TOTEM	Total Elastic and Diffractive Cross Section Measurement
TRT	Transition Radiation Tracker
UE	Underlying Event
ZDC	Zero Degree Calorimeter

A Appendix: Analysis binning

Following conventions for the variable binning are used:

- Binning of jet p_T in GeV: {0,5, 10, 20, 30, 40, 60, 80, 110, 160, 210, 260, 310, 400, 500, 600, 800, 1000, 1200, 1500, 1800, 2500}

- Binning of dijet mass in GeV: { 0, 30, 70, 110, 160, 210, 260, 310, 370, 440, 510, 590, 670, 760, 850, 950, 1060, 1180, 1310,1450, 1600, 1760, 1940, 2120, 2330, 2550, 2780, 3040, 3310, 3610, 3930, 4270, 4640, 5040, 5470, 5940, 6440, 7000}

- Binning of $\Delta\phi$ in rad: {0,$\pi/2$,2.5,2.8,2.9,3.0,3.1,π}

- Binning of $\chi = e^{|y_1 - y_2|}$: {1,3,5,10,30}

B Appendix: Monte-Carlo samples

The beam spot used in the MC samples with the MC10 tune is the measured beam spot from data period D. The ATLAS geometry used in the samples is called GEO-16-00-00 with particle ID (s934).
Baseline PYTHIA samples:

- No pile-up setup (150 ns bunch spacing) with reconstruction tags r1653.

- Pile-up: double bunch trains with 225ns separation, within trains there are 8 filled bunches with 150ns bunch separation and < number of pile-up interactions >= 2.2 with reconstruction tag r1833.

MC dataset	$\hat{p}_T$ [GeV]	Cross section [nb]	Number of events
J0	8-17	9860100	16343508
J1	17-35	678140	7384565
J2	35-70	40981	2795084
J3	70-140	2193.1	2792379
J4	140-280	87.707	2786179
J5	280-560	2.3502	2773476
J6	560-1120	$3.3618 \cdot 10^{-2}$	2774601
J7	1120-2240	$1.3744 \cdot 10^{-4}$	1381026

Table B.1: *Summary of pile-up* PYTHIA *samples studied.*
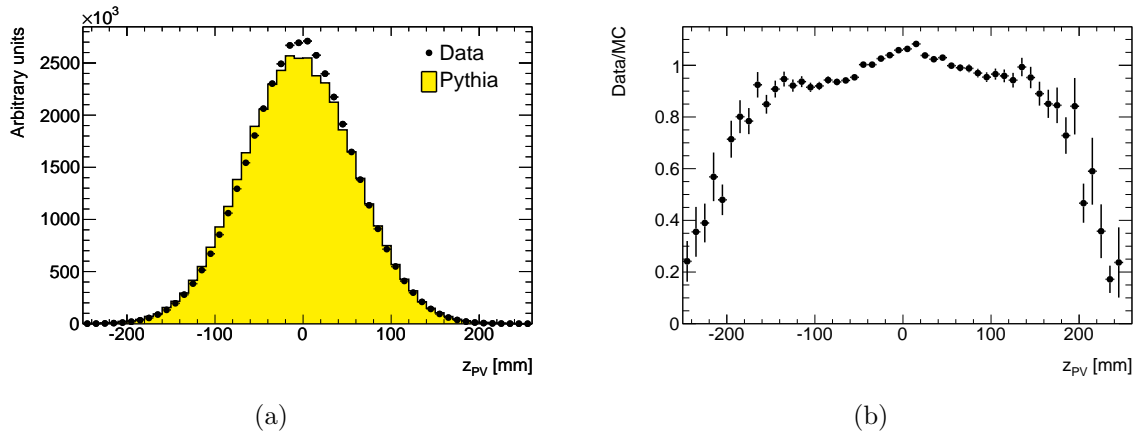
Figure B.1: *(a) The z-distribution of reconstructed primary vertices z_{PV} in data and* PYTHIA *simulation. (b) The reweighing factors as a function of z_{PV} which are applied to* PYTHIA *MC.*

C Appendix: Data quality requirements

Colour scheme used in the data quality assessment [97]:

- Green colour: Data can be non-restrictively used for physic analysis.

- Yellow colour: Data is not used for public analysis unless explicitly stated. This status indicates that the corresponding data are flawed relative to green data

quality but possibly recoverable. A recovery may be possible due to adjusted offline reconstruction.

- Red colour: Unrecoverable data which is lost due to major problems for data analysis.

- Grew colour: For very short runs with insufficient statistics or due to problems in the monitoring or assessment software, it may occur that a clear statement cannot be made.

- Black colour: System disabled.

A summary of the data quality requirements on accelerator parameters, detector operation and the analysis-based quality criteria in use in this analysis:

- Accelerator parameters: LHC delivered stable beams with an energy of 3.5 TeV.

- Detector-level quality criteria:
  - The L1 trigger processor has green data quality.
  - The L1 calorimeter trigger has green data quality.
  - The toroid magnet is at nominal voltage.
  - The solenoid magnet is at nominal voltage.
  - The Pixel, SCT and TRT work at nominal parameters.
  - The luminosity monitors and counters work at nominal parameters.
  - Good beam spot data is available with a full beam spot determination including the (resolution-corrected) beam spot size.

- Analysis-based quality criteria:
  - The monitored tracking properties behave as expected.
  - The muon algorithms and monitored muon reconstruction properties behave as expected.
  - For every calorimeter region, the inspected jet and missing transverse energy variables behave as expected.

D Appendix: Jet energy calibration

The values of the correction factor $\frac{1}{\mathcal{F}_{\mathrm{calib}}(E^{\mathrm{EM}})|_\eta}$ are illustrated in Fig. D.1 for three representative η-intervals.

The jet energy response of jets in the inclusive sample is derived separately for gluon- and quark-initiated jets and presented in Fig. D.2.
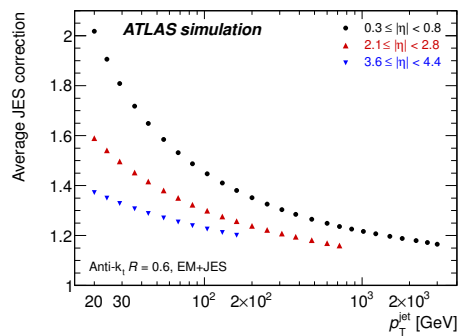
Figure D.1: *Average jet energy scale correction as a function of the calibrated jet transverse momentum for three representative η-intervals obtained from the nominal Monte-Carlo simulation sample. The correction is only shown over the accessible kinematic range. Taken from Ref. [60].*
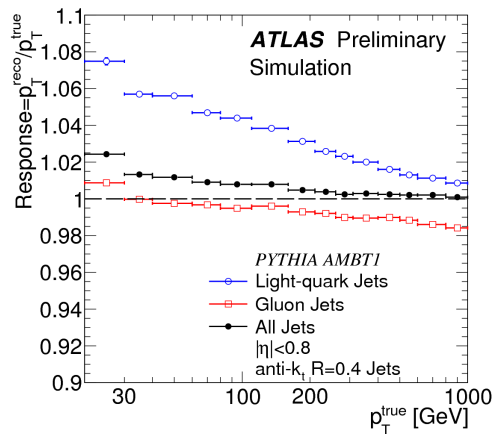
Figure D.2: *Response as a function of particle jet transverse momenta ($p_{\mathrm{T}}^{\mathrm{truth}}$) for all jets in the QCD sample (black solid circles), gluon-jets (red open squares), and light-quark jets (blue open circles) falling in the barrel in MC simulation. Jets are reconstructed with the anti-k_t algorithm with distance parameter R=0.4 and calibrated with the EM+JES calibration scheme. Taken from Ref. [110].*

E Appendix: Monte-Carlo comparisons

(a)

(b)
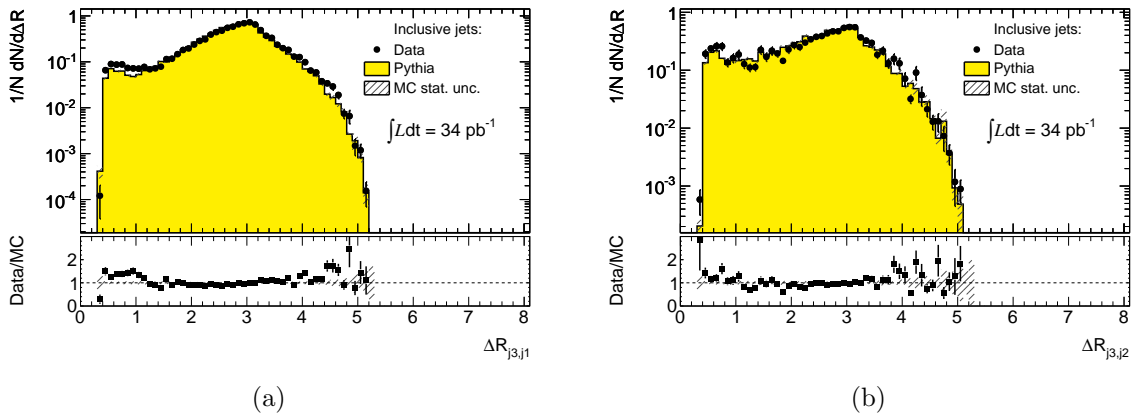
Figure E.1: *The ΔR distribution between the third leading jet in the event and the leading jets.*

(a)

(b)
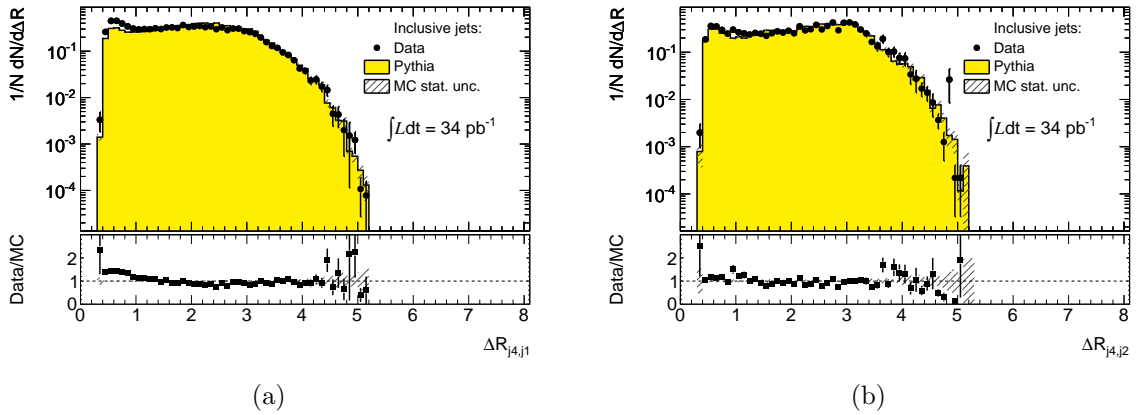
Figure E.2: *The ΔR distribution between the fourth leading jet in the event and the leading jets.*
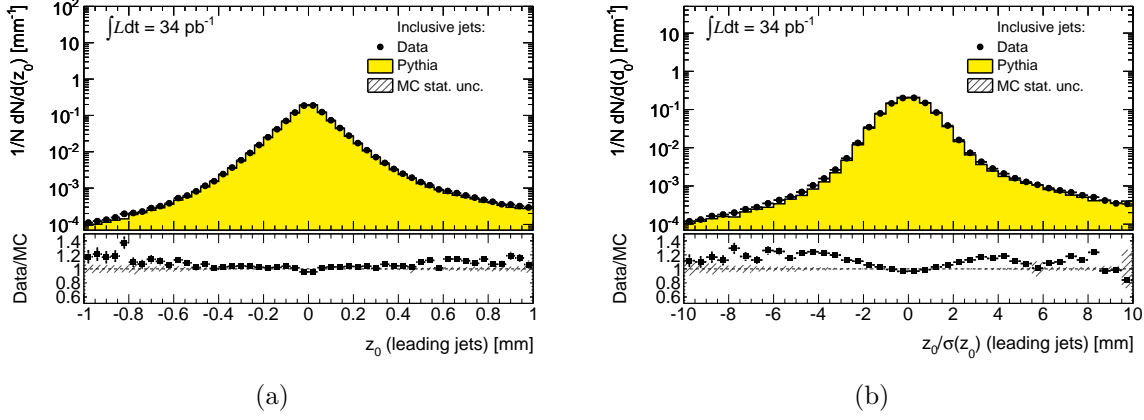
Figure E.3: *The reconstructed longitudinal impact parameter ((a)) and its significance ((b)) are presented for tracks associated to a jet and fulfilling b-tagging quality cuts.*

F Appendix: Calibration of bottom-tagging algorithms

Summary of the systematic uncertainties on the b-tagging efficiency measurement with the p_{T}^{rel} method given in Table F.1 and Table F.2.

The resulting scale factors for the JetProb tagger using the 50% efficiency working point and their uncertainties are given in the following:

$$20 < p_{\mathrm{T}} < 30 \text{ GeV} : 0.83 \pm 0.06(\text{stat}) \pm 0.09(\text{syst})$$
$$30 < p_{\mathrm{T}} < 60 \text{ GeV} : 0.89 \pm 0.03(\text{stat}) \pm 0.07(\text{syst})$$
$$60 < p_{\mathrm{T}} < 90 \text{ GeV} : 0.87 \pm 0.04(\text{stat}) \pm 0.03(\text{syst})$$
$$90 < p_{\mathrm{T}} < 140 \text{ GeV} : 0.89 \pm 0.09(\text{stat}) \pm 0.07(\text{syst})$$

Systematic uncertainties from the SV0 mass method for the mistag rate calibration of SV0 and JetProb tagging algorithm at 50% efficiency are given in Table F.3 and Table F.4. Systematic uncertainties from the negative tag method for the mistag rate calibration of SV0 and JetProb tagging algorithm at 50% efficiency are given in Table F.5 and Table F.6. The resulting mistag rate for the JetProb tagger is given in Table F.7.
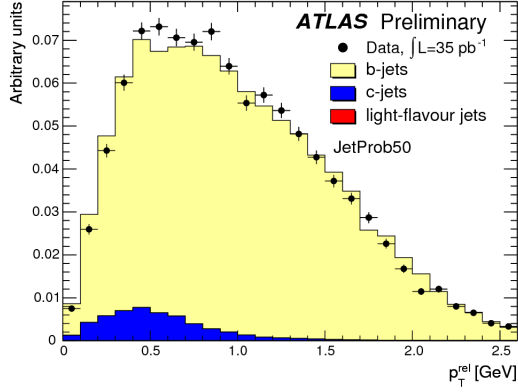
Figure F.1: *Example of a template fit in data to the p_T^{rel} distribution after the JetProb b-tagging cut. The template fit is performed on all jets. Uncertainties shown are for data statistics only. The discrepancies between the data and the sum of the templates are fully covered by the systematic uncertainties on the template shapes. Taken from Ref. [103].*

	Jet p_T [GeV]			
Source	20-30	30-60	60-90	90-140
Statistical	7.5	2.9	4.6	9.8
Measurement:				
Jet Energy Scale	4.3	0.5	1.3	1.2
Method:				
Template statistics	5.1	1.4	1.8	2.2
Non-b template modelling	0.8	0.8	0.7	0.3
b-template modelling	1.4	0.2	0.3	0.1
Validity for all b-hadrons	3.5	5.6	0.7	2.9
MC model dependence:				
b-hadron direction	2.5	0.1	1.4	3.6
Muon $p_\mathrm{T}^{\mathrm{muon}}$-spectrum	4.7	3.6	1.8	3.3
b-fragmentation	0.4	0.1	< 0.1	0.5
b-decay modelling (spectrum)	5.3	0.3	0.6	0.1
b-decay modelling (branching fractions)	1.0	0.4	0.1	< 0.1
Modelling of b/c-production	2.4/1.1	0.1/1.4	0.6/1.0	0.8/7.2
Total	11.1	7.1	3.6	9.5

Table F.1: *Relative statistical and systematic uncertainties for the SV0 tagger at working point SV0 weight > 5.85.*

	Jet p_{T}[GeV]			
Source	20-30	30-60	60-90	90-140
Statistical	7.7	2.9	4.6	9.5
Measurement:				
Jet Energy Scale	2.3	0.5	1.3	0.5
Method:				
Template statistics	5.4	1.4	1.8	2.5
Non-b template modelling	0.8	0.8	0.7	0.5
b-template modelling	1.4	0.2	0.4	0.2
Validity for all b-hadrons	4.9	5.7	0.5	3.0
MC model dependence:				
b-hadron direction	1.7	0.4	1.4	3.9
Muon $p_{\mathrm{T}}^{\mathrm{muon}}$-spectrum	4.7	4.6	1.8	2.4
b-fragmentation	0.4	0.1	< 0.1	0.4
b-decay modelling (spectrum)	5.3	0.3	0.6	0.4
b-decay modelling (branching fractions)	1.0	0.4	0.1	< 0.1
Modelling of b/c-production	2.9/1.3	0.1/1.6	0.6/1.1	0.5/4.1
Total	11.2	7.7	3.6	7.4

Table F.2: *Relative statistical and systematic uncertainties for the JetProb tagger at working point JetProb weight > 3.25.*

	Jet p_{T}[GeV]					
Source	40-60	60-90	90-140	140-200	200-300	300-500
	$\|\eta\| < 1.2$					
Simulation statistics	14.7	16.1	8.4	7.7	7.7	7.1
b-tagging efficiencies	3.0	3.9	3.2	2.8	2.4	1.4
b-, c-template shapes	1.6	1.7	1.0	0.7	0.3	0.6
Light-template shapes	6.8	7.1	9.2	11.0	12.4	16.7
Long-lived particles and material interactions	3.2	3.2	3.2	3.2	3.2	3.2
Total	16.9	18.4	13.2	14.1	15.1	18.5
	$1.2 < \|\eta\| < 2.5$					
Simulation statistics	9.6	10.1	8.5	8.3	11.5	24.5
b-tagging efficiencies	1.8	2.2	2.1	2.0	2.2	2.5
b-, c-template shapes	0.5	0.3	0.4	0.1	0.4	1.2
Light-template shapes	5.9	6.4	11.0	9.9	11.3	18.7
Long-lived particles and material interactions	3.2	3.2	3.2	3.2	3.2	3.2
Total	11.9	12.5	14.4	13.4	16.6	31.1

Table F.3: *Relative systematic uncertainties (in %) of the SV0 mass measurement for the SV0 tagger at working point SV0 weight > 5.85.*

Source	Jet p_T [GeV]					
	40-60	60-90	90-140	140-200	200-300	300-500
$\|\eta\| < 1.2$						
Simulation statistics	6	7.6	3.5	3.3	4.0	4.7
b-tagging efficiencies	2.2	3.4	2.8	2.6	3.0	2.8
b-, c-template shapes	0.7	1.6	0.9	1.0	1.0	1.0
Light-template shapes	2.4	2.8	3.8	4.7	6.2	6.5
Long-lived particles and material interactions	3.2	3.2	3.2	3.2	3.2	3.2
Total	7.5	9.5	6.7	7.1	8.6	9.1
$1.2 < \|\eta\| < 2.5$						
Simulation statistics	6.6	4.2	3.9	3.7	4.5	13.4
b-tagging efficiencies	1.6	1.6	2.0	1.6	2.0	0.8
b-, c-template shapes	0.6	0.7	0.9	0.4	0.2	0.2
Light-template shapes	2.1	2.6	3.1	3.6	3.9	2.9
Long-lived particles and material interactions	3.2	3.2	3.2	3.2	3.2	3.2
Total	7.8	6.1	6.3	6.3	7.0	14.1

Table F.4: *Relative systematic uncertainties (in %) of the SV0 mass measurement for the JetProb tagger at working point JetProb weight > 3.25.*

Source	Jet p_T [GeV]					
	40-60	60-90	90-140	140-200	200-300	300-500
$\|\eta\| < 1.2$						
Simulation statistics	7.2	5.3	4.8	4.3	4.0	3.8
b-tagging efficiencies	4.1	5.9	6.6	7.1	8.8	8.4
Heavy flavour fraction	1.0	2.0	2.1	2.3	2.9	3.0
Track multiplicity	1.0	0.4	0.2	0.3	1.1	2.2
Long-lived particles and material interactions	1.5	2.0	1.8	0.8	0.5	0.5
Impact parameter resolution	3.6	3.0	12.7	5.4	9.8	7.1
Total	9.3	9.0	15.4	10.2	14.1	12.3
$1.2 < \|\eta\| < 2.5$						
Simulation statistics	16.6	8.8	8.8	7.8	5.1	10.4
b-tagging efficiencies	3.9	3.7	6.6	4.0	4.3	3.9
Heavy flavour fraction	1.1	0.9	2.5	1.2	1.2	1.2
Track multiplicity	1.7	4.8	2.1	0.7	1.2	2.2
Long-lived particles and material interactions	2.3	0.7	1.0	0.5	0.7	0.2
Impact parameter resolution	7.3	13.4	6.5	16.7	19.2	12.1
Total	18.8	17.2	13.2	18.9	20.4	16.6

Table F.5: *Relative systematic uncertainties (in %) of the negative tag measurement for the SV0 tagger at working point SV0 weight > 5.85.*

Source	Jet p_T[GeV]					
	40-60	60-90	90-140	140-200	200-300	300-500
	$|\eta| < 1.2$					
Simulation statistics	3.8	2.8	2.3	2.1	1.9	1.6
b-tagging efficiencies	10.0	11.8	13.0	15.2	15.4	12.8
Heavy flavour fraction	3.3	4.0	4.2	4.9	4.7	4.1
Track multiplicity	6.2	0.3	0.1	0.3	0.4	1.3
Long-lived particles and material interactions	2.0	1.1	0.6	0.8	1.0	0.5
Impact parameter resolution	6.0	0.6	3.1	2.5	2.8	2.4
Total	14.3	12.8	14.2	16.3	16.5	13.8
	$1.2 < |\eta| < 2.5$					
Simulation statistics	4.4	3.2	3.2	2.9	2.2	3.4
b-tagging efficiencies	4.5	5.4	7.4	7.5	7.2	5.6
Heavy flavour fraction	1.2	1.7	2.5	2.9	2.5	2.1
Track multiplicity	0.7	1.0	0.4	0.4	0.5	1.6
Long-lived particles and material interactions	1.0	0.9	0.9	0.3	0.6	0.4
Impact parameter resolution	13.8	3.5	6.5	0.5	7.8	9.4
Total	15.3	7.5	10.8	8.5	11.1	11.7

Table F.6: *Relative systematic uncertainties (in %) of the negative tag measurement for the JetProb tagger at working point JetProb weight > 3.25.*

Jet p_T[GeV]					
40-60	60-90	90-140	140-200	200-300	300-500
$|\eta| < 1.2$					
1.3$\pm$0.4	1.2$\pm$0.3	1.0$\pm$0.1	0.9$\pm$0.1	1.1$\pm$0.3	1.2$\pm$0.3
$|\eta| < 1.2$					
1.2$\pm$0.4	1.1$\pm$0.2	1.0$\pm$0.2	1.0$\pm$0.2	1.1$\pm$0.3	1.0$\pm$0.4

Table F.7: *The resulting mistag rate scale factors for the JetProb tagger for the selection cuts yielding 50% efficiency in $t\bar{t}$ events as function of jet p_T and $|\eta|$.*
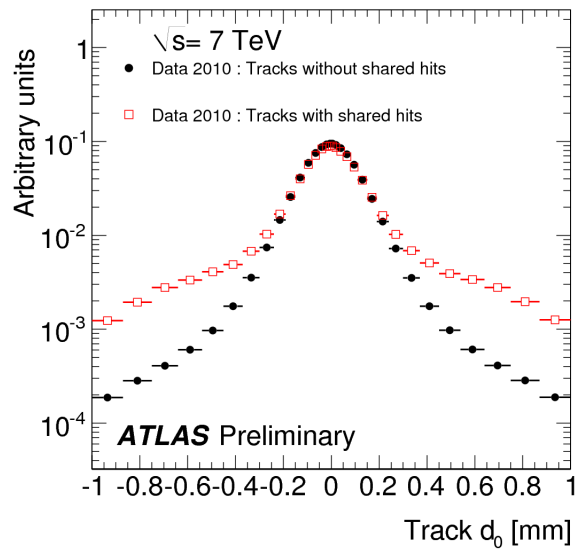
G Appendix: Extension of calibration

Figure G.1: *Distribution of the transverse impact parameter in experimental data for tracks without shared hits and for tracks with shared hits. Both plots are normalised to unity. Taken from Ref. [153].*
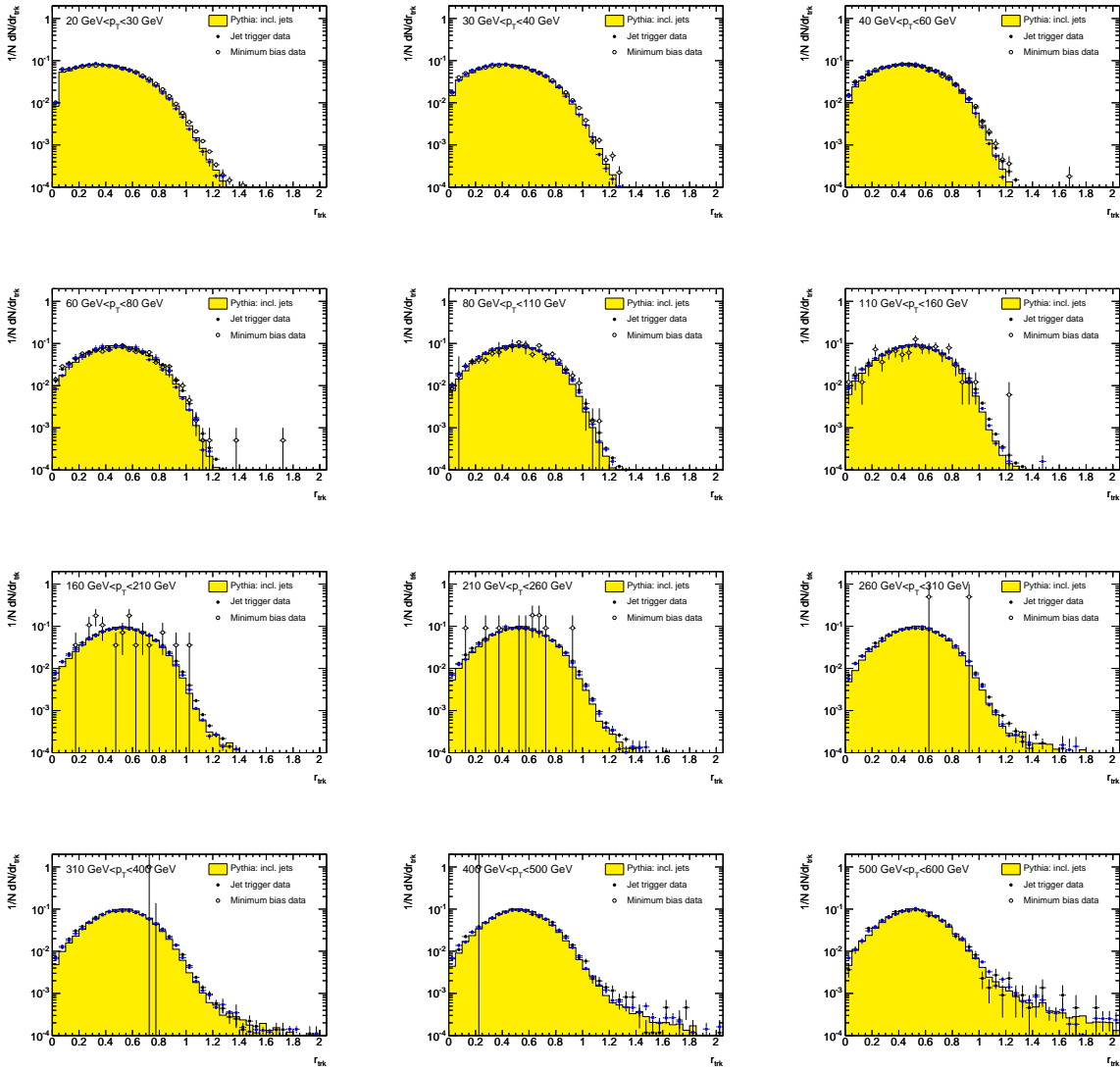
H Appendix: Jet energy scale

Figure H.1: r_{trk} *distributions for all* p_{T} *bins for the inclusive jet sample for* $|y| < 1.2$.
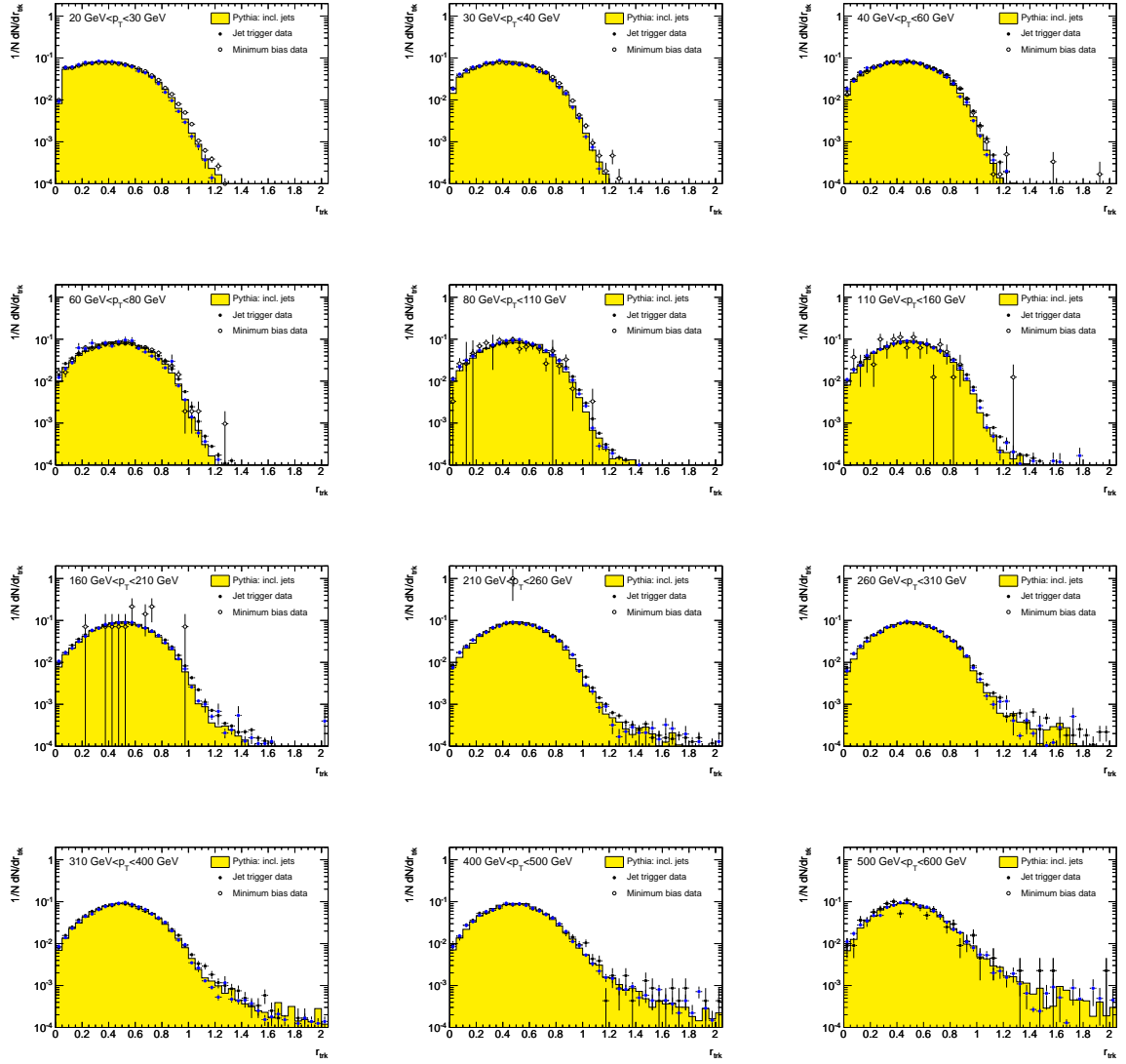
Figure H.2: r_{trk} *distributions for all* p_{T} *bins for the inclusive jet sample for* $1.2 < |y| < 2.1$.
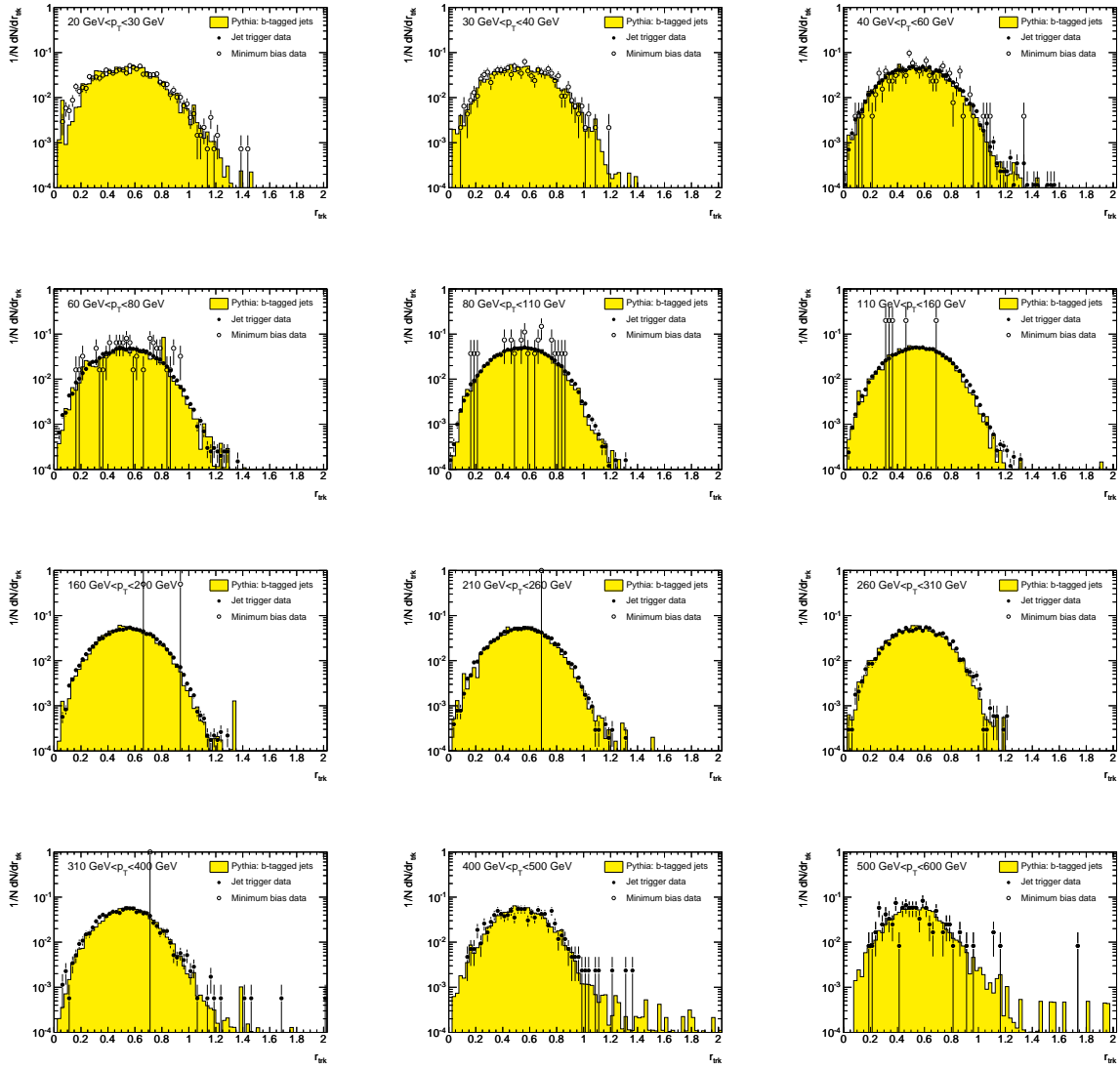
Figure H.3: r_{trk} *distributions for all* p_{T} *bins for the b-tagged jet sample for* $|y| < 1.2$.
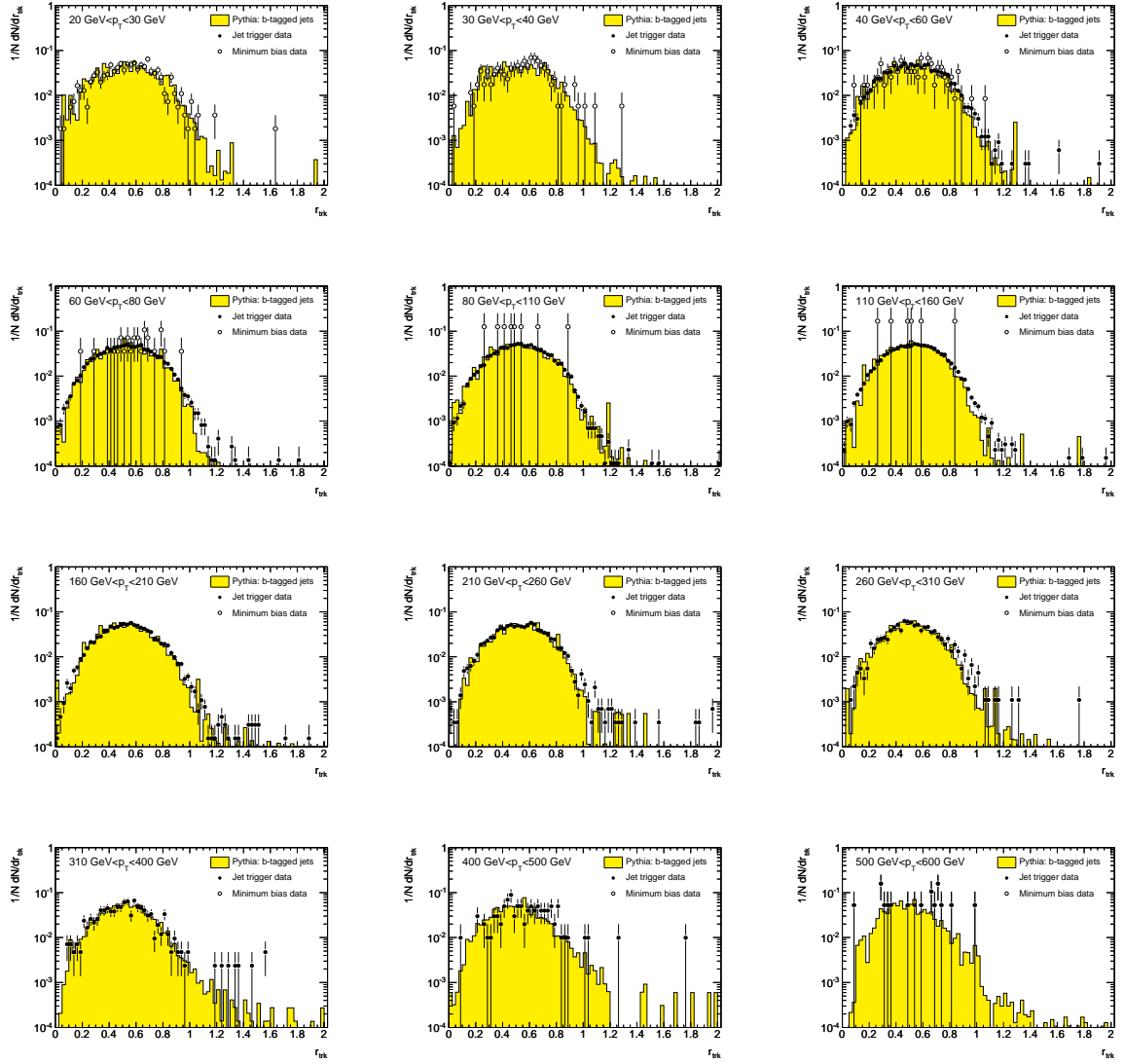
Figure H.4: r_{trk} *distributions for all* p_T *bins for the b-tagged jet sample for* $1.2 < |y| < 2.1$.

I Appendix: Correction factors and systematics
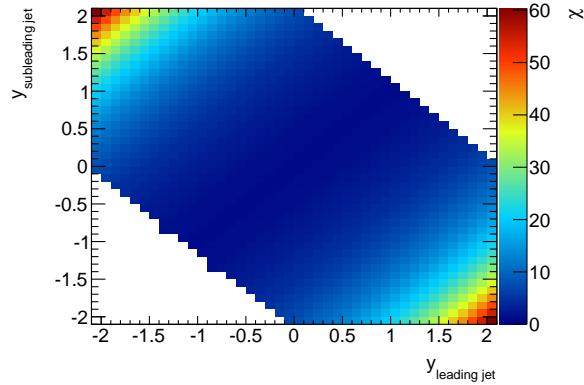
Figure I.1: *Illustration of the acceptance cut $\frac{1}{2}|y_1 + y_2| < 1.1$ for the measurement of the dijet cross section as a function of the angular variable χ.*

$\Delta\phi$ bin [rad]	b-tagging efficiency	Statistical error	Fractional sys. uncertainty [%]
$\pi/2$-2	0.25	± 0.00	± 17
2-3	0.25	± 0.00	± 16
3-3	0.22	± 0.00	± 16
3-3	0.21	± 0.00	± 16
3-π	0.19	± 0.00	± 17

Table I.1: *The b-tagging efficiency as a function of the difference in azimuthal angle $\Delta\phi$. The statistical error is of the order of $1 \cdot 10^{-4}$.*

χ bin	110 GeV $< M <$ 370 GeV			370 GeV $< M <$ 850 GeV		
	b-tagging efficiency	Stat. error	Sys. un-certainty [%]	b-tagging efficiency	Stat. error	Sys. un-certainty [%]
1-3	0.26	± 0.00	17	0.35	± 0.01	44
3-5	0.22	± 0.00	16	0.35	± 0.01	42
5-10	0.20	± 0.00	16	0.29	± 0.01	39
10-30	0.23	± 0.00	15	0.28	± 0.00	28

Table I.2: *The b-tagging efficiency as a function of angular variable χ for both mass ranges.*

$\Delta\phi$ bin [rad]	b-tagging purity	Total uncertainty	Fractional tot. uncertainty [%]
$\pi/2$-2.5	0.72	± 0.21	29
2.5-2.8	0.58	± 0.17	29
2.8-2.9	0.59	± 0.13	21
2.9-3.0	0.55	± 0.16	29
3.0-π	0.59	± 0.13	21

Table I.3: *The b-tagging purity and systematic uncertainty as a function of $\Delta\phi$.*

Figure I.2: *Construction of templates for b-contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of dijet mass.*

Figure I.3: *Construction of templates for non-b contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of dijet mass.*

Figure I.4: *Construction of templates for b-contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of $\Delta\phi$.*
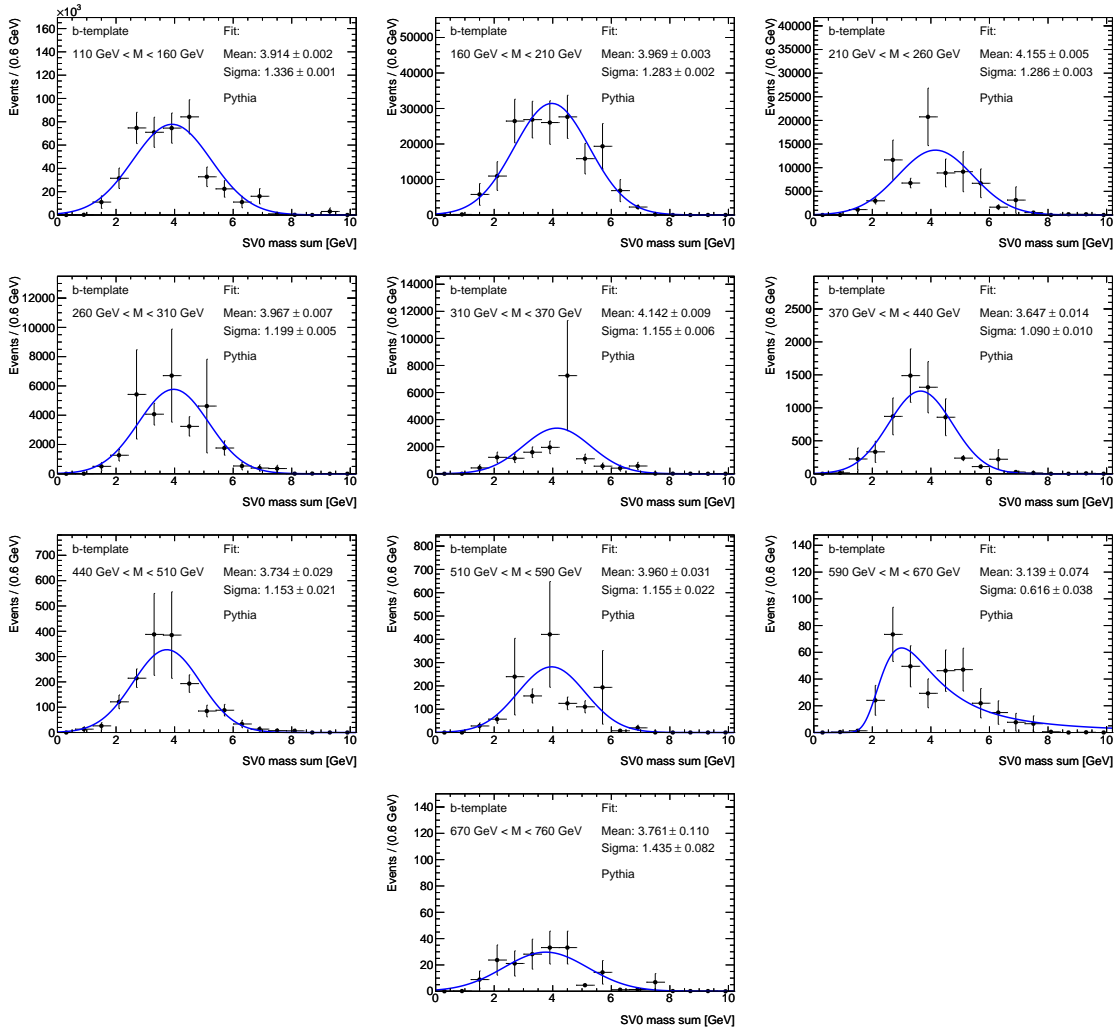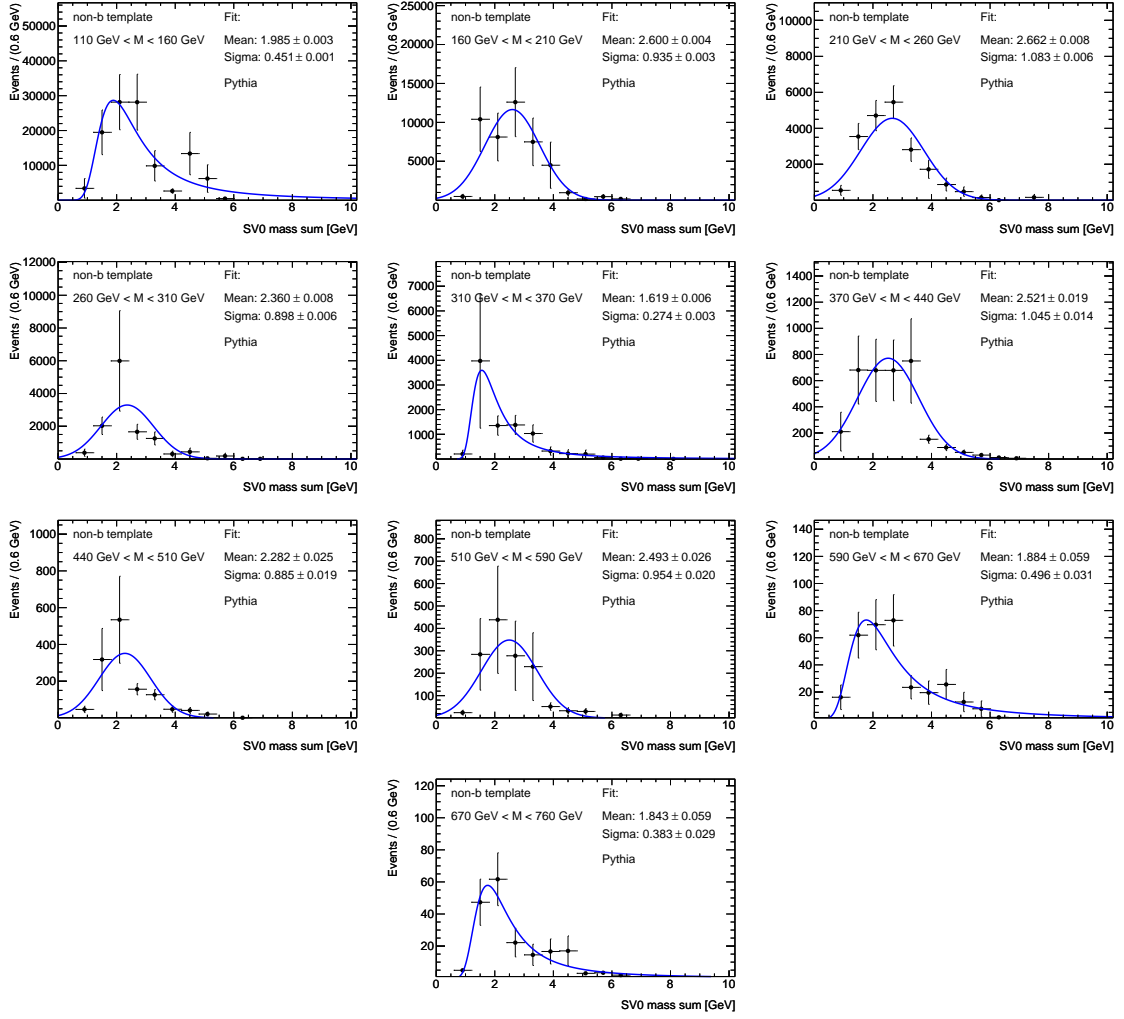
Figure I.5: *Construction of templates for non-b contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of $\Delta\phi$.*

χ bin	b-tagging purity	Total uncertainty	Fractional tot. uncertainty [%]
110 GeV $< M <$ 370 GeV			
1-3	0.67	± 0.17	25
3-5	0.48	± 0.27	56
5-10	0.58	± 0.18	31
10-30	0.76	± 0.18	24
370 GeV $< M <$ 850 GeV			
1-3	0.38	± 0.26	17
3-5	0.34	± 0.24	55
5-10	0.36	± 0.19	27
10-30	0.38	± 0.13	29

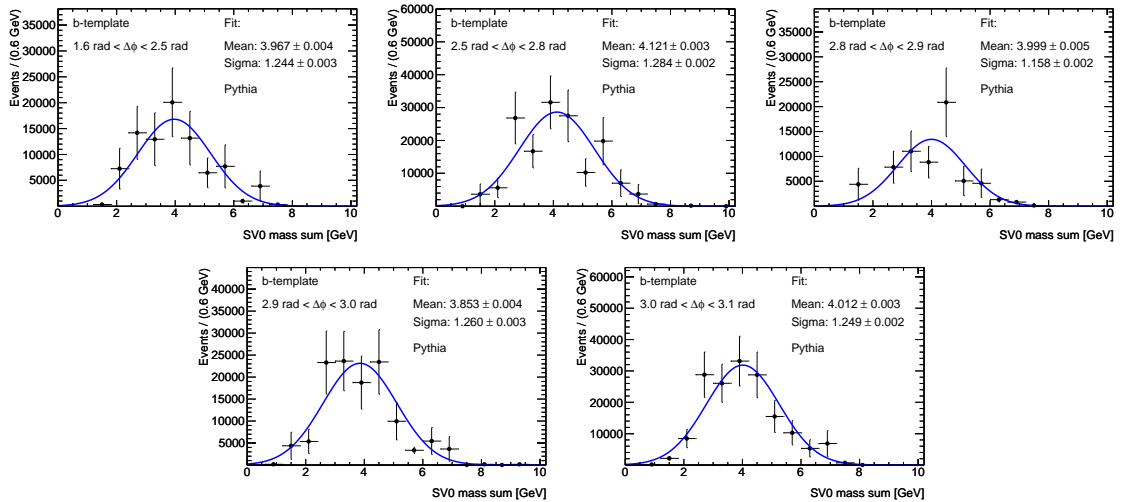**Table I.4:** *The b-tagging purity and systematic uncertainty as a function of χ.*
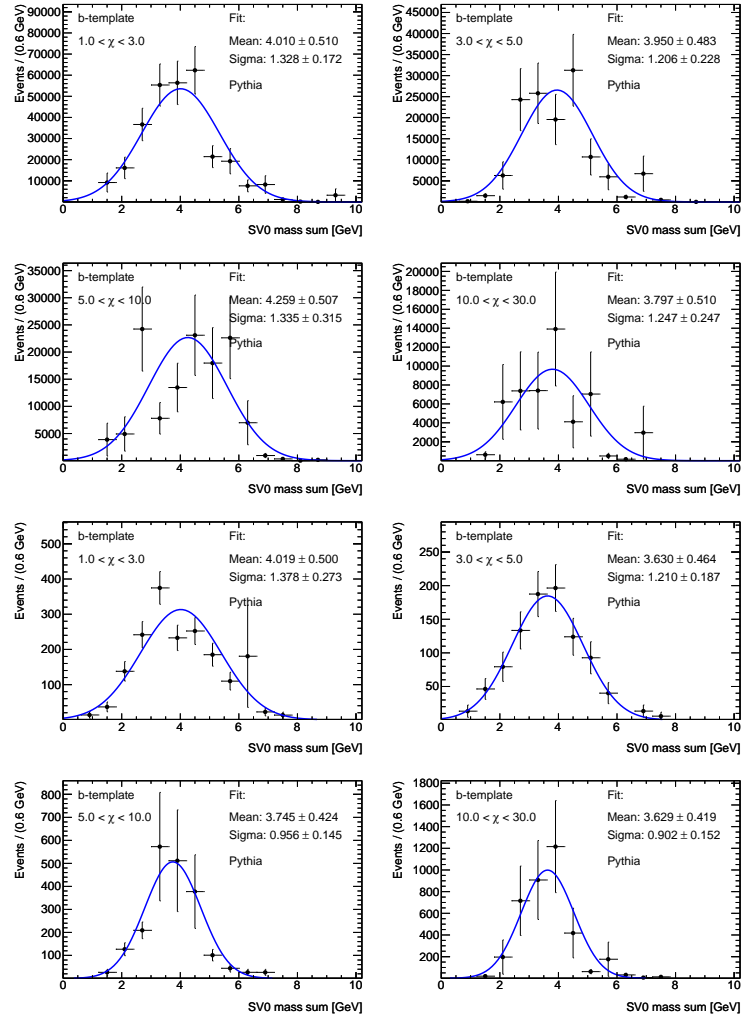
Figure I.6: *Construction of templates for b-contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of χ for the lower and higher mass region.*
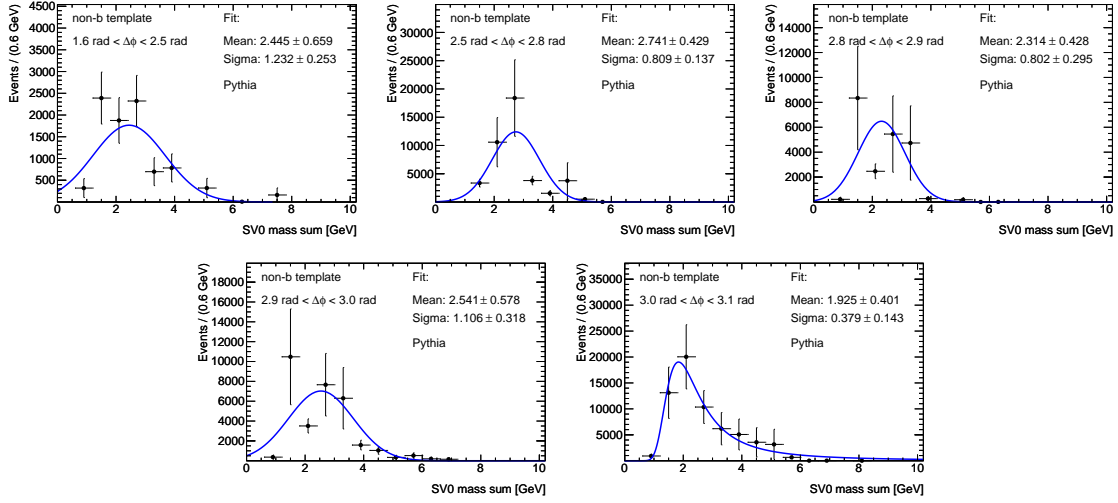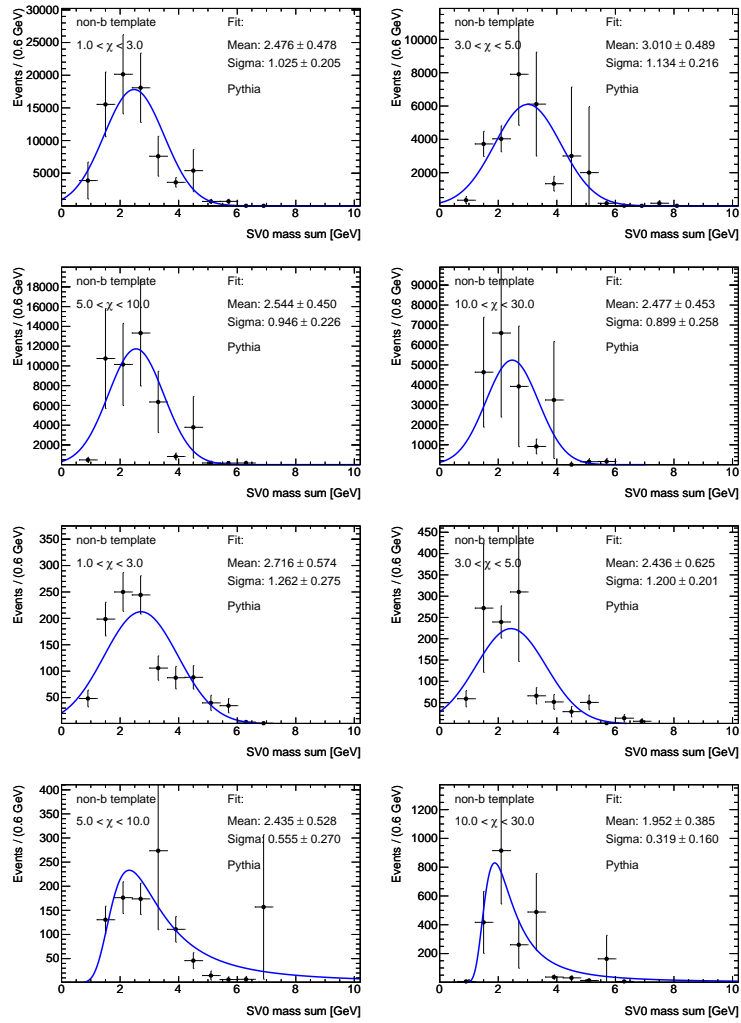
Figure I.7: *Construction of templates for non-b contribution in the SV0 mass sum distribution from simulation for the measurement of the dijet cross section as a function of χ for the lower and higher mass region.*
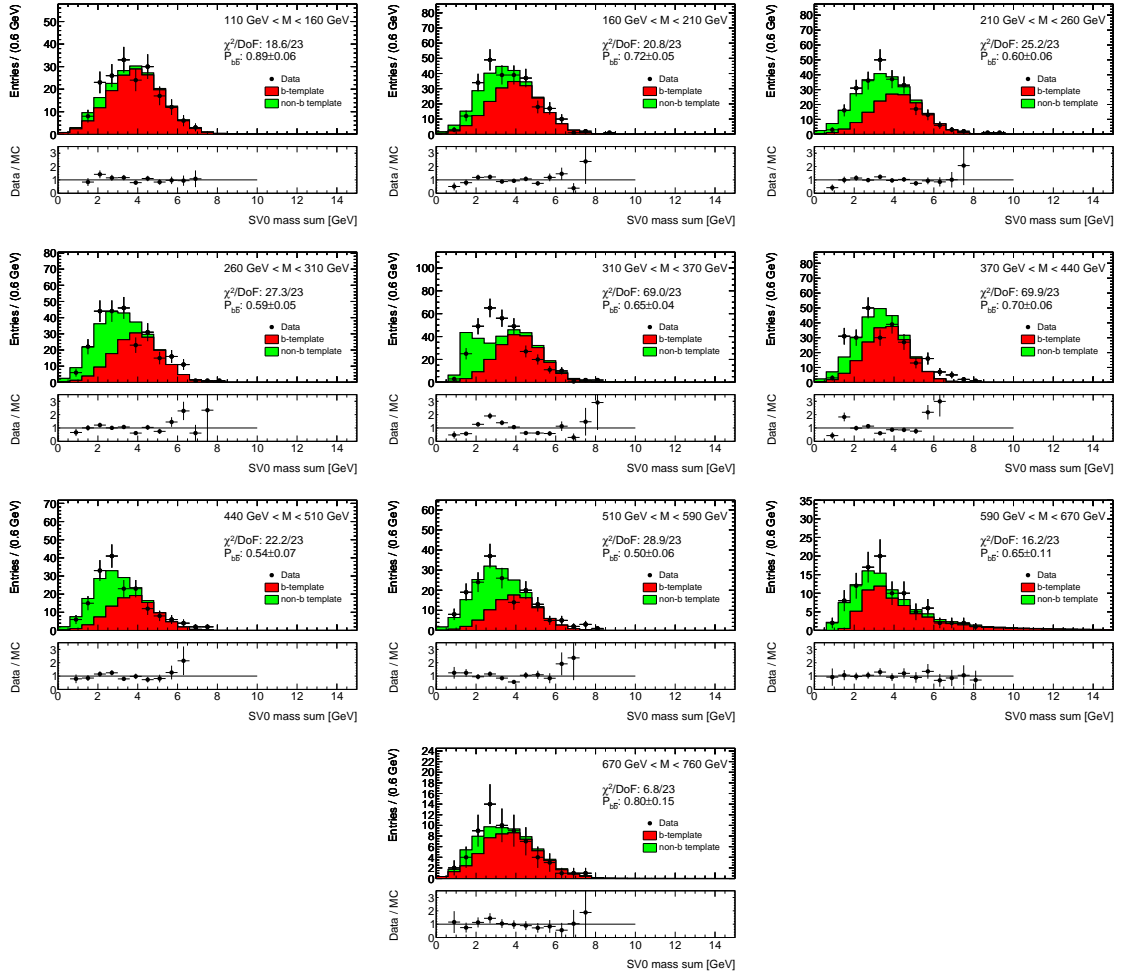
Figure I.8: *Final template fits on data using a binned maximum likelihood fit (TFractionFitter) in bins of dijet mass. The uncertainty on $P_{b\bar{b}}$ denotes the statistical uncertainty from the likelihood fit. DoF stands for the number of degrees of freedom. The single contributions are drawn stacked on top of each other.*
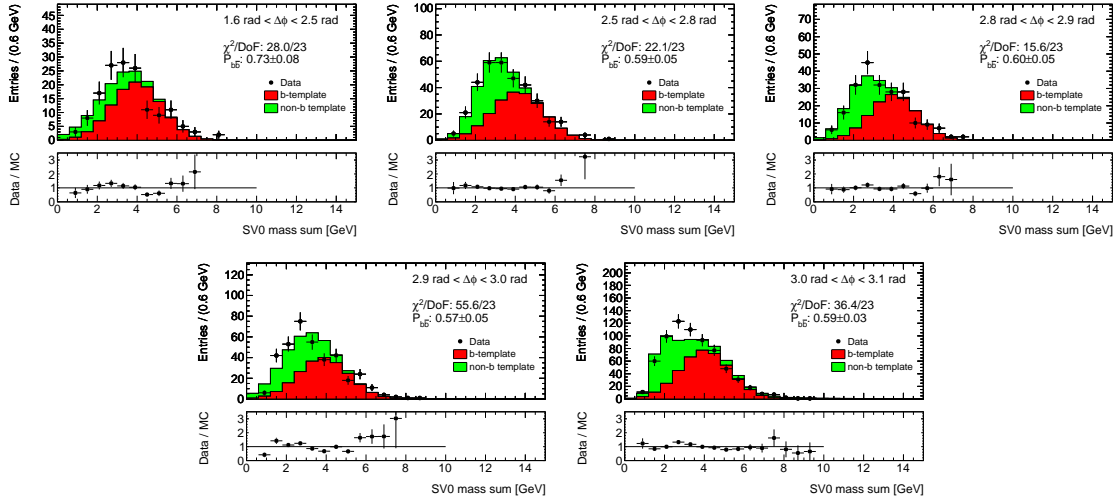
Figure I.9: *Final template fits on data using a binned maximum likelihood fit (TFractionFitter) in bins of $\Delta\phi$. The uncertainty on $P_{b\bar{b}}$ denotes the statistical uncertainty from the likelihood fit. DoF stands for the number of degrees of freedom. The single contributions are drawn stacked on top of each other.*
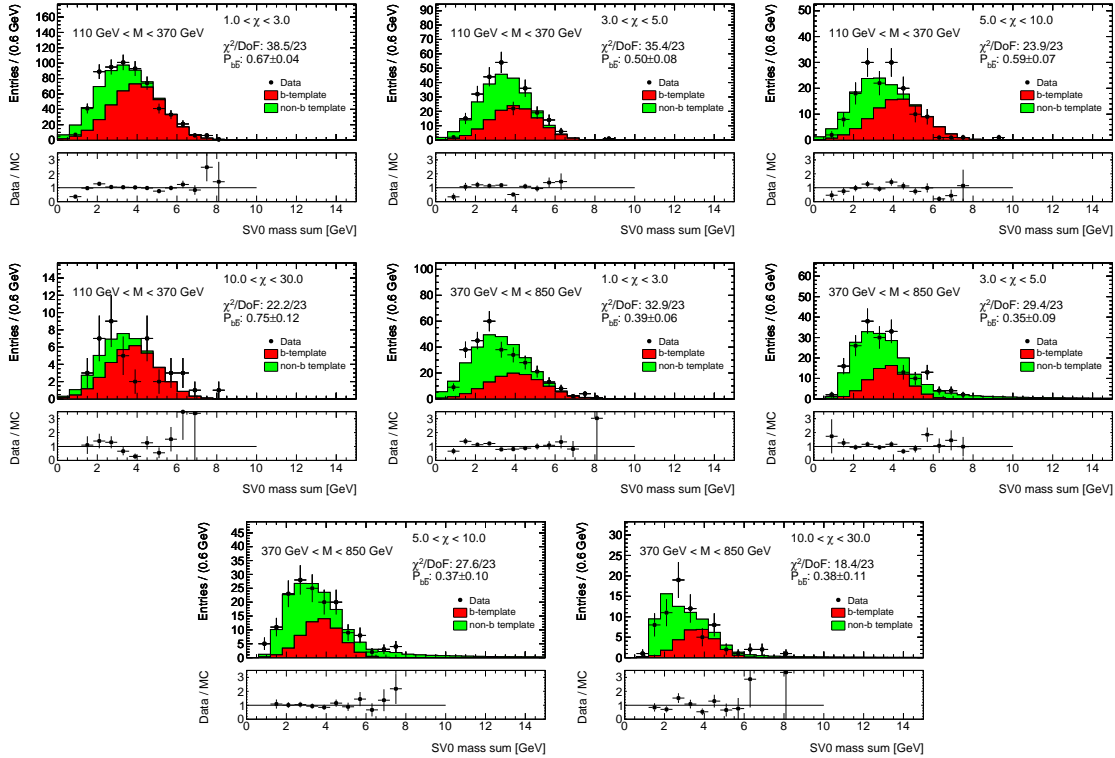
Figure I.10: *Final template fits on data using a binned maximum likelihood fit (TFractionFitter) in bins of χ. The uncertainty on $P_{b\bar{b}}$ denotes the statistical uncertainty from the likelihood fit. DoF stands for the number of degrees of freedom. The single contributions are drawn stacked on top of each other.*
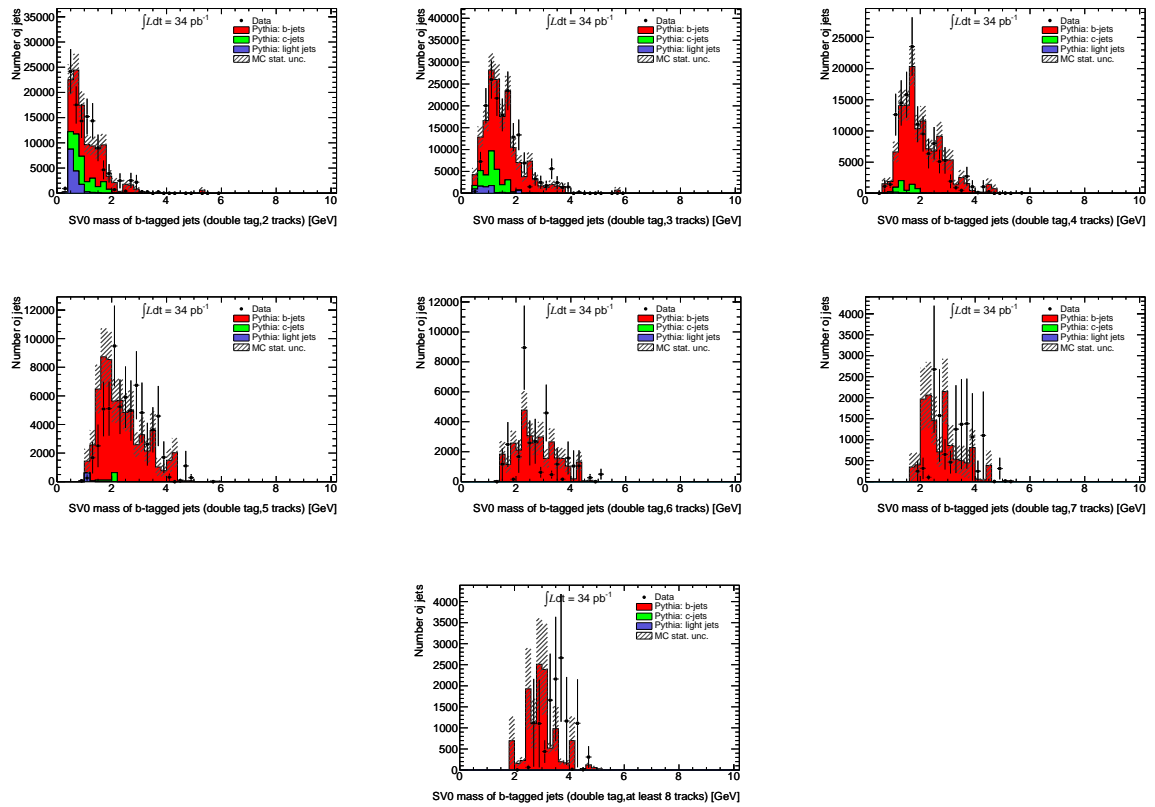
Figure I.11: *Distribution of the invariant mass of the charged tracks in bins of track multiplicity in the secondary vertex (SV0 mass) found in the b-tagged jets in events in which both leading jets are b-tagged. Simulated data is normalised to the number of b-tagged jets in data. The single contributions are drawn stacked on top of each other.*

(a)   (b)
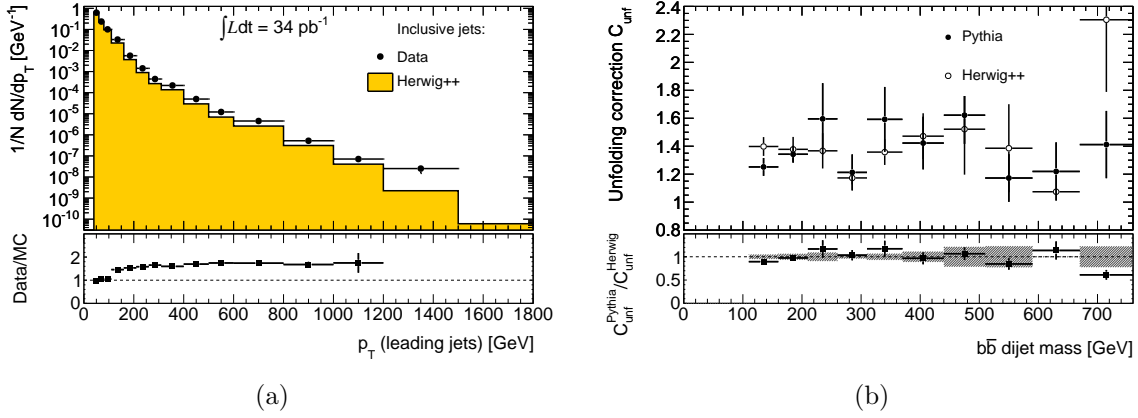
Figure I.12: *(a) The leading jet p_T distribution compared to* HERWIG++ *simulation on the basis of Fig. 8.1(a) and (b) the corresponding unfolding correction $C_{unf}^{Herwig++}$ compared to* PYTHIA *on the basis of Fig. 9.20.*
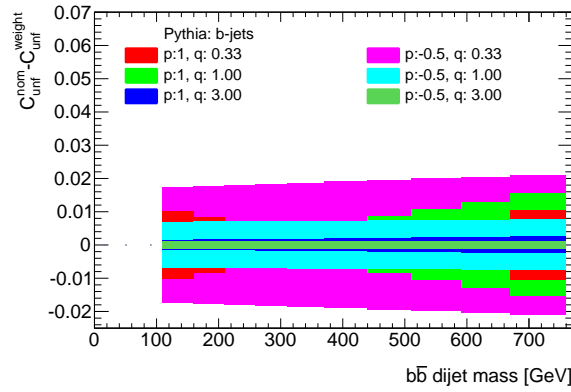
Figure I.13: *The resulting effect on the unfolding correction factors from the shape variations from the function Eq. 9.8 for the $b\bar{b}$ cross section as a function of dijet mass. The variations are fitted and the deviation to the nominal unfolding correction symmetrised.*

Dijet mass bin [GeV]	C_{unf}	Sys. uncertainty	Fractional sys. uncertainty [%]
110-160	1.24	± 0.06	5
160-210	1.24	± 0.06	4
210-260	1.24	± 0.05	4
260-310	1.24	± 0.05	4
310-370	1.25	± 0.05	4
370-440	1.25	± 0.05	4
440-510	1.25	± 0.04	4
510-590	1.25	± 0.05	4
590-670	1.25	± 0.05	4
670-760	1.26	± 0.05	4

Table I.5: *The unfolding correction and systematic uncertainty as a function of dijet mass.*

$\Delta\phi$ bin [rad]	C_{unf}	Stat. error	Fractional sys. uncertainty [%]
$\pi/2$-2	1.30	± 0.08	1.8
2.5-3	1.31	± 0.06	1.3
2.8-3	1.26	± 0.08	7.1
2.9-3	1.27	± 0.07	4.7
3.0-π	1.22	± 0.07	7.1

Table I.6: *The unfolding correction and systematic uncertainty as a function of $\Delta\phi$.*

χ bin	C_{unf}	Stat. error	Fractional sys. uncertainty [%]
110 GeV $< M <$ 370 GeV			
1-3	1.21	± 0.04	0.2
3-5	1.31	± 0.07	0.7
5-10	1.37	± 0.15	0.6
10-30	1.54	± 0.15	0.6
370 GeV $< M <$ 850 GeV			
1-3	1.19	± 0.05	0.7
3-5	1.33	± 0.11	2.0
5-10	1.41	± 0.16	0.7
10-30	1.33	± 0.10	1.6

Table I.7: *The unfolding correction and systematic uncertainty as a function of χ.*

Dijet mass bin [GeV]	Cleaning efficiency	Statistical error	Fractional sys. uncertainty[%]
110-160	0.96	± 0.00	$^{+0.3}_{-1.0}$
160-210	0.97	± 0.00	$^{+0.3}_{-0.9}$
210-260	0.97	± 0.00	$^{+0.3}_{-1.0}$
260-310	0.98	± 0.00	$^{+0.3}_{-1.0}$
310-370	0.99	± 0.00	$^{+0.3}_{-1.0}$
370-440	0.99	± 0.00	$^{+0.2}_{-1.0}$
440-510	1.00	± 0.00	$^{+0.1}_{-1.0}$
510-590	1.00	± 0.00	$^{+0.1}_{-0.9}$
590-670	1.00	± 0.00	$^{+0.0}_{-0.7}$
670-760	1.00	± 0.00	$^{+0.0}_{-0.6}$

Table I.8: *The jet cleaning efficiency and systematic uncertainty as a function of dijet mass for inclusive jets.*

Dijet mass bin [GeV]	Unfolding correction	Statistical error	Fractional sys. uncertainty [%]
110-160	0.99	±0.00	±1.8
160-210	0.99	±0.01	±1.8
210-260	0.99	±0.01	±1.8
260-310	1.02	±0.01	±1.8
310-370	1.00	±0.01	±2.1
370-440	0.99	±0.01	±1.9
440-510	1.01	±0.01	±2.0
510-590	1.00	±0.01	±2.3
590-670	1.00	±0.01	±2.2
670-760	1.00	±0.01	±2.2

Table I.9: *The unfolding correction and systematic uncertainty as a function of dijet mass for inclusive jets.*

J Appendix: Cross section measurement

Dijet mass bin [GeV]	Central value [pb GeV^{-1}]	Statistical error [pb GeV^{-1}]	Systematic error [pb GeV^{-1}]
110-160	341.69	±6.55	+115.82 - 131.44
160-210	122.36	±2.94	+52.23 - 42.82
210-260	42.90	±1.48	+21.94 - 18.04
260-310	19.29	±0.82	+9.18 - 6.74
310-370	9.28	±0.49	+4.90 - 3.44
370-440	5.02	±0.25	+2.80 - 1.94
440-510	1.24	±0.10	+0.73 - 0.48
510-590	0.44	±0.03	+0.27 - 0.17
590-670	0.20	±0.01	+0.14 - 0.09
670-760	0.12	±0.01	+0.08 - 0.05

Table J.1: *Results from the dijet cross section measurement of b-jets as a function of invariant dijet mass.*

$\Delta\phi$ bin [rad]	Central value [rad^{-1}]	Statistical error [rad^{-1}]	Systematic error [rad^{-1}]
$\pi/2$-2.5	0.06	±0.01	+0.02 - 0.02
2.5-2.8	0.60	±0.09	+0.24 - 0.22
2.8-2.9	1.34	±0.25	+0.47 - 0.42
2.9-3.0	1.94	±0.30	+0.76 - 0.70
3.0-3.1	2.90	±0.38	+0.99 - 0.89
3.1-π	3.56	±0.87	+1.47 - 1.36

Table J.2: *Results from the dijet cross section measurement of b-jets as a function of $\Delta\phi$.*

χ bin	Central value [GeV^{-1}]	Statistical error [GeV^{-1}]	Systematic error [GeV^{-1}]
1-3	$65.5 \cdot 10^{-5}$	$\pm 5.9 \cdot 10^{-5}$	$+23.0 \cdot 10^{-5}$ $- 20.6 \cdot 10^{-5}$
3-5	$47.0 \cdot 10^{-5}$	$\pm 5.5 \cdot 10^{-5}$	$+29.2 \cdot 10^{-5}$ $- 28.1 \cdot 10^{-5}$
5-10	$23.3 \cdot 10^{-5}$	$\pm 3.7 \cdot 10^{-5}$	$+9.3 \cdot 10^{-5}$ $- 8.6 \cdot 10^{-5}$
10-30	$2.1 \cdot 10^{-5}$	$\pm 0.5 \cdot 10^{-5}$	$+0.8 \cdot 10^{-5}$ $- 0.7 \cdot 10^{-5}$

Table J.3: *Results from the cross section measurement as a function of χ for 110 GeV < M < 370 GeV.*

χ bin	Central value [GeV^{-1}]	Statistical error [GeV^{-1}]	Systematic error [GeV^{-1}]
1-3	$11.3 \cdot 10^{-5}$	$\pm 1.2 \cdot 10^{-5}$	$+12.1 \cdot 10^{-5}$ $- 8.8 \cdot 10^{-5}$
3-5	$10.9 \cdot 10^{-5}$	$\pm 1.7 \cdot 10^{-5}$	$+10.2 \cdot 10^{-5}$ $- 8.5 \cdot 10^{-5}$
5-10	$10.8 \cdot 10^{-5}$	$\pm 2.1 \cdot 10^{-5}$	$+9.1 \cdot 10^{-5}$ $- 6.7 \cdot 10^{-5}$
10-30	$5.5 \cdot 10^{-5}$	$\pm 1.3 \cdot 10^{-5}$	$+3.0 \cdot 10^{-5}$ $- 2.4 \cdot 10^{-5}$

Table J.4: *Results from the cross section measurement as a function of χ for 370 GeV < M < 850 GeV.*

Dijet mass bin [GeV]	Central value [pb GeV^{-1}]	Statistical error [pb GeV^{-1}]	Systematic error [pb GeV^{-1}]
110-160	29706.92	± 216.75	$+4362.04$ $- 3762.42$
160-210	13207.69	± 165.69	$+1874.75$ $- 1753.06$
210-260	6039.29	± 73.98	$+880.88$ $- 790.08$
260-310	3041.64	± 49.82	$+449.45$ $- 419.44$
310-370	1460.01	± 30.89	$+236.41$ $- 196.89$
370-440	621.58	± 14.82	$+85.67$ $- 82.63$
440-510	279.33	± 7.00	$+39.31$ $- 34.05$
510-590	128.07	± 3.24	$+17.44$ $- 15.60$
590-670	59.03	± 1.61	$+8.72$ $- 8.04$
670-760	28.23	± 0.76	$+3.88$ $- 3.65$

Table J.5: *Results from the dijet cross section measurement of inclusive jets as a function of invariant dijet mass.*

Dijet mass bin [GeV]	Central value [pb GeV^{-1}]	Statistical error [pb GeV^{-1}]	Systematic error [pb GeV^{-1}]
110-160	$11.5 \cdot 10^{-3}$	$\pm 0.2 \cdot 10^{-3}$	$+3.5 \cdot 10^{-3}$ $- 4.2 \cdot 10^{-3}$
160-210	$9.3 \cdot 10^{-3}$	$\pm 0.3 \cdot 10^{-3}$	$+3.7 \cdot 10^{-3}$ $- 3.0 \cdot 10^{-3}$
210-260	$7.1 \cdot 10^{-3}$	$\pm 0.3 \cdot 10^{-3}$	$+3.5 \cdot 10^{-3}$ $- 2.9 \cdot 10^{-3}$
260-310	$6.3 \cdot 10^{-3}$	$\pm 0.3 \cdot 10^{-3}$	$+2.9 \cdot 10^{-3}$ $- 2.1 \cdot 10^{-3}$
310-370	$6.4 \cdot 10^{-3}$	$\pm 0.4 \cdot 10^{-3}$	$+3.2 \cdot 10^{-3}$ $- 2.2 \cdot 10^{-3}$
370-440	$8.1 \cdot 10^{-3}$	$\pm 0.4 \cdot 10^{-3}$	$+4.4 \cdot 10^{-3}$ $- 3.0 \cdot 10^{-3}$
440-510	$4.4 \cdot 10^{-3}$	$\pm 0.4 \cdot 10^{-3}$	$+2.6 \cdot 10^{-3}$ $- 1.7 \cdot 10^{-3}$
510-590	$3.4 \cdot 10^{-3}$	$\pm 0.2 \cdot 10^{-3}$	$+2.1 \cdot 10^{-3}$ $- 1.3 \cdot 10^{-3}$
590-670	$3.5 \cdot 10^{-3}$	$\pm 0.2 \cdot 10^{-3}$	$+2.3 \cdot 10^{-3}$ $- 1.5 \cdot 10^{-3}$
670-760	$4.3 \cdot 10^{-3}$	$\pm 0.3 \cdot 10^{-3}$	$+2.8 \cdot 10^{-3}$ $- 1.6 \cdot 10^{-3}$

Table J.6: *Results from the dijet cross section ratio measurement of b-jets and inclusive jets as a function of invariant dijet mass.*
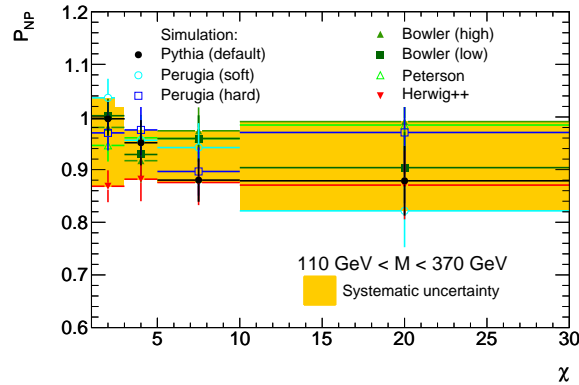
K Appendix: NLO calculation

Figure K.1: *The non-perturbative corrections as a function of the angular variable χ in the mass range 110 GeV < M < 370 GeV with resulting uncertainty envelop around the default* PYTHIA *correction.*

L Appendix: Resonance search

Selection	p_0	p_1	p_2	p_3
Full spectrum	$1.6{\cdot}10^{-10} \pm 2.6 \cdot 10^{-10}$	$5.5{\cdot}10^{-11} \pm 8.1$	13.3±1.1	1.4±0.2
Full spectrum with $\lvert \Delta y \rvert < 1$	$1.1{\cdot}10^{-10} \pm 2.6 \cdot 10^{-10}$	$6.2{\cdot}10^{-11} \pm 31.8$	11.7±1.5	1.0±0.2
Spectrum with L1_J95	$1.6{\cdot}10^{-10} \pm 2.3 \cdot 10^{-10}$	$13.1{\cdot}10^{-11} \pm 11.2$	21.5±1.6	3.8±0.5

Table L.1: *Summary of the background fits for the* BUMPHUNTER *on different event selections.*
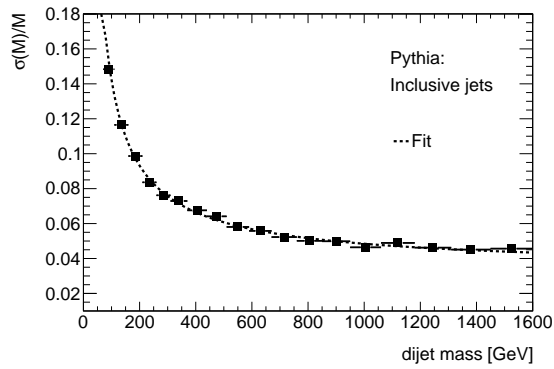
Figure L.1: *Relative resolution as a function of dijet mass determined with the help of* Pythia *QCD samples.*
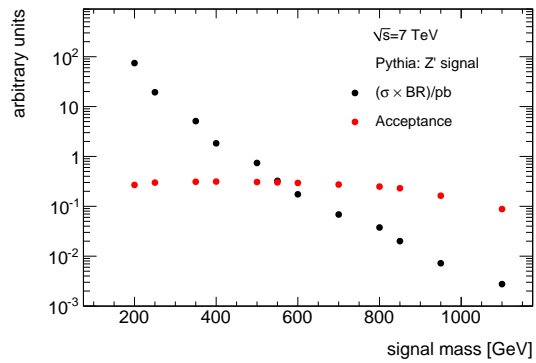

Figure L.2: *The acceptance and the product of cross section and branching ratio for each resonance mass of a Z′ boson are presented.*

Generated Z' mass	Reconstructed Z' mass	SV0 b-tagging efficiency
220	175.11	0.34
310	242.134	0.37
420	327.659	0.37
520	403.999	0.36
620	480.782	0.34
720	555.921	0.32
800	616.385	0.30
930	713.374	0.26
1020	780.272	0.24
1120	847.952	0.22
1300	969.715	0.18
1500	1098.49	0.15

Table L.2: *Summary of the generated and corresponding reconstructed mass of a Z' boson and of the SV0 b-tagging efficiency obtained by Toy-MC.*

Bibliography

[1] Griffiths D 2008 *Introduction to elementary particles*

[2] Lancaster M (Tevatron Electroweak Working Group) 2011 (*Preprint* `1107.5255`)

[3] Nakamura K *et al.* 2010 *J. Phys. G 37, 075021*

[4] Kraus C, Bornschein B, Bornschein L, Bonn J, Flatt B *et al.* 2005 *Eur.Phys.J.* **C40** 447–468 and references therein (*Preprint* `hep-ex/0412056`)

[5] Elementary particles Fermilab 95-759 http://www.fnal.gov

[6] Dissertori G, Knowles I G and Schmelling M 2003 *Quantum Chromodynamics High Energy Experiments and Theory*

[7] Bardeen W A, Buras A J, Duke D W and Muta T 1978 *Phys. Rev. D* **18**(11) 3998–4017 URL `http://link.aps.org/doi/10.1103/PhysRevD.18.3998`

[8] The ATLAS Collaboration An update to the combined search for the standard model higgs boson with the atlas detector at the lhc using up to 4.9 fb^{-1} of *pp* collision data at $\sqrt{s} = 7$ TeVTech. Rep. ATLAS-CONF-2012-019 CERN Geneva

[9] Ellis R K, Stirling W J and Webber B R 1996 *QCD and Collider Physics*

[10] Bethke S 2009 *Eur. Phys. J.* **C64** 689–703 (*Preprint* `0908.1135`)

[11] Pumplin J *et al.* 2002 *JHEP* **07** 012 (*Preprint* `hep-ph/0201195`)

[12] Martin A D, Stirling W J, Thorne R S and Watt G 2009 *Eur. Phys. J.* **C63** 189–285 (*Preprint* `0901.0002`)

[13] Altarelli G and Parisi G 1977 *Nucl. Phys.* **B126** 298

[14] Boelaert N and Akesson T 2010 *Dijet angular distributions in proton-proton collisions at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 14$ TeV* Ph.D. thesis Lund

[15] Hewett J L and Rizzo T G 1989 *Phys. Rept.* **183** 193

[16] Mohapatra R N and Pati J C 1975 *Phys. Rev.* **D11** 566–571

[17] Senjanovic G and Mohapatra R N 1975 *Phys. Rev.* **D12** 1502

[18] Martin S P 1997 (*Preprint* `hep-ph/9709356`)

[19] Aaltonen T *et al.* (CDF Collaboration) 2009 *Phys.Rev.* **D79** 112002 (*Preprint* `0812.4036`)

[20] Baur U, Spira M and Zerwas P M 1990 *Phys. Rev.* **D42** 815–824

[21] Randall L and Sundrum R 1999 *Phys. Rev. Lett.* **83** 3370–3373 (*Preprint* `hep-ph/9905221`)

[22] Buckley A *et al.* 2011 *Physics Reports* **504** ISSN 0370-1573

[23] T Sjostrand and S Mrenna and PZ Skands 2006 *JHEP* **05** 026 (*Preprint* `hep-ph/0603175`)

[24] Bahr M *et al.* 2008 *Eur. Phys. J.* **C58** 639–707 (*Preprint* `0803.0883`)

[25] Corcella G *et al.* 2001 *JHEP* **01** 010 (*Preprint* `hep-ph/0011363`)

[26] The ATLAS Collaboration Charged particle multiplicities in p p interactions at $\sqrt{s} = 0.9$ and 7 TeV in a diffractive limited phase-space measured with the atlas detector at the lhc and new pythia6 tune Tech. Rep. ATLAS-CONF-2010-031 CERN Geneva

[27] Buckley A G and Schulz H First tuning of herwig/jimmy to atlas data Tech. Rep. ATL-COM-PHYS-2010-620 CERN Geneva

[28] Sherstnev A and Thorne R S 2008 *Eur.Phys.J.* **C55** 553–575 (*Preprint* `0711.2473`)

[29] Amati D and Veneziano G 1979 *Physics Letters B* **83** 87 – 92 ISSN 0370-2693 URL `http://www.sciencedirect.com/science/article/pii/0370269379908967`

[30] 't Hooft G 1974 *Nucl. Phys.* **B75** 461

[31] Stirling W J private communication

[32] M Cacciari and G P Salam and G Soyez The fastjet package http://fastjet.fr/

[33] M Cacciari and G P Salam and G Soyez 2008 *JHEP* **04** 63 (*Preprint* `arxiv:0802.1189`)

[34] Berger E L 1988 *Phys. Rev. D* **37**

[35] Abbiendi G *et al.* (OPAL Collaboration) 2003 *Eur.Phys.J.* **C29** 463–478 (*Preprint* `hep-ex/0210031`)

[36] Abe K *et al.* (SLD Collaboration) 2000 *Phys.Rev.Lett.* **84** 4300–4304 (*Preprint* `hep-ex/9912058`)

[37] Andersson B, Gustafson G, Ingelman G and Sjöstrand T 1983 *Physics Reports* **97** 31 – 145

[38] Peterson C, Schlatter D, Schmitt I and Zerwas P M 1983 *Phys. Rev.* **D27** 105

[39] Bowler M 1981 *Z.Phys.* **C11** 169

[40] Kartvelishvili V G, Likhoded A K and Petrov V A 1978 *Phys.Lett.* **B78** 615

[41] Herb S W, Hom D C, Lederman L M, Sens J C, Snyder H D, Yoh J K, Appel J A, Brown B C, Brown C N, Innes W R, Ueno K, Yamanouchi T, Ito A S, Jöstlein H, Kaplan D M and Kephart R D 1977 *Phys. Rev. Lett.* **39**(5) 252–255 URL `http://link.aps.org/doi/10.1103/PhysRevLett.39.252`

[42] Albajar C *et al.* 1991 *Physics Letters B* **256** 121 – 128 ISSN 0370-2693 URL http://www.sciencedirect.com/science/article/pii/037026939190228I

[43] Abe F *et al.* (CDF) 1988 *Nucl. Instr. Meth.* **A271** 387–403

[44] Abachi S *et al.* (D0) 1994 *Nucl. Instrum. Meth.* **A338** 185–253

[45] Abbott B *et al.* (D0 Collaboration) 2000 *Phys.Rev.Lett.* **84** 5478–5483 (*Preprint* hep-ex/9907029)

[46] Abe F *et al.* 1993 *Phys. Rev. Lett.* **71**(15) 2396–2400 URL http://link.aps.org/doi/10.1103/PhysRevLett.71.2396

[47] Mangano M L 2005 *AIP Conf.Proc.* **753** 247–260 14 pages, Presented at the 2004 Hadron Collider Physics Workshop, East Lansing, MI, June 2004 Report-no: CERN-PH-TH/04-210 (*Preprint* hep-ph/0411020)

[48] Cacciari M, Frixione S, Mangano M L, Nason P and Ridolfi G 2004 *JHEP* **0407** (*Preprint* hep-ph/0312132)

[49] Acosta D e a (CDF Collaboration) 2005 *Phys. Rev. D* **71** 032001

[50] The CDF Collaboration Measurement of the inclusive *b*-jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeVTech. rep. Fermilab cDF note 8418

[51] Albajar C *et al.* (UA1) 1994 *Z. Phys.* **C61** 41–52

[52] Abe F *et al.* (CDF Collaboration) 1996 *Phys.Rev.* **D53** 1051–1065 (*Preprint* hep-ex/9508017)

[53] Abbott B *et al.* (D0 Collaboration) 2000 *Phys.Lett.* **B487** 264–272 (*Preprint* hep-ex/9905024)

[54] Abe F *et al.* (CDF) 1997 *Phys. Rev.* **D55** 2546–2558

[55] The CDF Collaboration Measurement of the $b\bar{b}$ cross section using a dedicated trigger in $p\bar{p}$ collisions at 1.96 TeVTech. rep. Fermilab cDF note 8939

[56] Frixione S and Webber B R 2002 *JHEP* **0206** 029 (*Preprint* hep-ph/0204244)

[57] Brüning O S, Collier P, Lebrun P, Myers S, Ostojic R, Poole J and Proudlock P 2004 *LHC Design Report* (Geneva: CERN)

[58] Aad G *et al.* (Atlas Collaboration) 2012 *Eur.Phys.J.* **C72** 1849 (*Preprint* 1110.1530)

[59] Aad G *et al.* (ATLAS) 2011 *Eur. Phys. J.* **C 71** 1630 (*Preprint* 1101.2185)

[60] The ATLAS Collaboration 2011 (*Preprint* 1112.6426)

[61] The ATLAS Collaboration Updated luminosity determination in *pp* collisions at $\sqrt{s} = 7$ TeV using the atlas detector Tech. Rep. ATLAS-CONF-2011-011 CERN Geneva

[62] van der Meer S Calibration of the effective beam height in the isr. oai:cds.cern.ch:296752 Tech. Rep. CERN-ISR-PO-68-31. ISR-PO-68-31 CERN Geneva

[63] Aad G *et al.* 2008 *J. Instrum.* **3** S08003. 437 p also published by CERN Geneva in 2010

[64] Aad G *et al.* (The ATLAS Collaboration) 2009 (*Preprint* `0901.0512`)

[65] Aad G *et al.* (ATLAS Collaboration) 2010 *Eur.Phys.J.* **C70** 823–874 submitted to Eur. Phys. J. C (*Preprint* `1005.4568`)

[66] Asai M 2006 *Trans.Amer.Nucl.Soc.* **95** 757

[67] Folger G and Wellisch J P 2003 MOMT007 (*Preprint* `0306007`)

[68] Amelin N S *et al.* 1991 *Phys. Rev. Lett.* **67** 1523–1526

[69] Amelin N S, Csernai L P, Staubo E F and Strottman D 1992 *Nucl. Phys.* **A 544** 463–466

[70] Bravina L V, Csernai L P, Levai P, Amelin N S and Strottman D 1994 *Nucl. Phys.* **A 566** 461–464

[71] Bravina L V 1995 *Phys. Lett.* **B 344** 49–54

[72] Bertini H W 1969 *Phys. Rev.* **A 188** 1711–1730

[73] Guthrie M P, Alsmiller R G and Bertini H W 1968 *Nucl. Instrum. Meth.* **66** 29–36

[74] Guthrie M P and Bertini H W 1971 *Nucl. Phys.* **A 169** 670–672

[75] Stepanov N V Statistical modeling of fission of excited atomic nuclei. 2. calculation and comparison with experiment Tech. Rep. In Russian ITEP Moscow (*Preprint* `ITEP-88-55`)

[76] Ribon A *et al.* Status of GEANT4 hadronic physics for the simulation of lhc experiments at the start of the lhc physics program Tech. Rep. CERN-LCGAPP-2010-02

[77] Abat E, Abdallah J M, Addy T N, Adragna P, Aharrouche M *et al.* 2010

[78] Salam G P 2007 (*Preprint* `0705.2696`)

[79] Catani S, Dokshitzer Y L, Seymour M H and Webber B R 1993 *Nucl. Phys.* **B406** 187–224

[80] Ellis S D and Soper D E 1993 *Phys.Rev.* **D48** 3160–3166 (*Preprint* `hep-ph/9305266`)

[81] Soyez G 2010 *JHEP* **1007** 075 (*Preprint* `1006.3634`)

[82] Lampl W, Laplace S, Lelas D, Loch P, Ma H, Menke S, Rajagopalan S, Rousseau D, Snyder S and Unal G 2008 Calorimeter clustering algorithms: Description and performance Tech. Rep. ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003 CERN Geneva

[83] Cornelissen T, Elsing M, Fleischmann S, Liebig W, Moyse E and Salzburger A Concepts, design and implementation of the atlas new tracking (newt) Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002 CERN Geneva

[84] Frühwirth R 1987 *Nucl. Instrum. Methods Phys. Res., A* **262** 444. 19 p

[85] The ATLAS Collaboration Performance of primary vertex reconstruction in proton-proton collisions at $\sqrt{s}$ =7 TeV in the atlas experiment Tech. Rep. ATLAS-CONF-2010-069 CERN Geneva

[86] The ATLAS Collaboration Characterization of interaction-point beam parameters using the *pp* event-vertex distribution reconstructed in the atlas detector at the lhc Tech. Rep. ATLAS-CONF-2010-027 CERN Geneva

[87] Fruhwirth R, Waltenberger W and Vanlaer P 2007 *J. Phys.* **G34** N343

[88] The ATLAS Collaboration Charged particle multiplicities in *pp* interactions at $\sqrt{(s)} = 7$ TeV measured with the atlas detector at the lhc Tech. Rep. ATLAS-CONF-2010-024 CERN Geneva

[89] The ATLAS Collaboration Atlas monte carlo tunes for mc09 Tech. Rep. ATL-PHYS-PUB-2010-002 CERN Geneva

[90] The ATLAS Collaboration 2011 *New J.Phys.* **13** (*Preprint* `1012.5104`)

[91] The ATLAS Collaboration Study of the material budget in the atlas inner detector with k_s^0 decays in collision data at $\sqrt{s} = 900$ GeV Tech. Rep. ATLAS-CONF-2010-019 CERN Geneva

[92] The ATLAS Collaboration Muon performance in minimum bias *pp* collision data at $\sqrt{s} = 7$ TeV with atlas Tech. Rep. ATLAS-CONF-2010-036 CERN Geneva

[93] The ATLAS Collaboration Performance of the atlas secondary vertex b-tagging algorithm in 7 TeV collision data Tech. Rep. ATLAS-CONF-2010-042 CERN Geneva

[94] The ATLAS Collaboration Performance of impact parameter-based b-tagging algorithms with the atlas detector using proton-proton collisions at $\sqrt{s} = 7$ tev Tech. Rep. ATLAS-CONF-2010-091 CERN Geneva

[95] The ATLAS Collaboration Calibrating the *b*-tag and mistag efficiencies of the sv0 *b*-tagging algorithm in 3 pb^{-1} of data with the atlas detector Tech. Rep. ATLAS-CONF-2010-099 CERN Geneva

[96] The ATLAS Collaboration Commissioning of the atlas high-performance b-tagging algorithms in the 7 TeV collision data Tech. Rep. ATLAS-CONF-2011-102 CERN Geneva

[97] Baak M *et al.* Data quality status flags and good run lists for physics analysis in atlas Tech. Rep. ATL-COM-GEN-2009-015 CERN Geneva

[98] The ATLAS Collaboration Data-quality requirements and event cleaning for jets and missing transverse energy reconstruction with the atlas detector in proton-proton collisions at a center-of-mass energy of *sqrts* = 7 TeVTech. Rep. ATLAS-CONF-2010-038 CERN Geneva

[99] Aad G *et al.* (ATLAS Collaboration) 2010 *Phys.Lett.* **B688** 21–42 40 pages (pages 24-40 appendix with ATLAS collaboration author list), 4 figures. The data points are in the HEP-data database at: http://hepdata.cedar.ac.uk/View/8591806 (*Preprint* `1003.3124`)

[100] Paterno M Calculating efficiencies and their uncertainties Tech. rep. FNAL/CD/CEPA/SLD

[101] collaboration T A 2010 *Eur. Phys. J. C* **71** 1512. 67 p

[102] The ATLAS Collaboration Properties of jets and inputs to jet reconstruction and calibration with the atlas detector using proton-proton collisions at $\sqrt{s} = 7$ TeVTech. Rep. ATLAS-CONF-2010-053 CERN Geneva

[103] The ATLAS Collaboration Calibrating the b-tag efficiency and mistag rate in 35 pb^{-1} of data with the atlas detector Tech. Rep. ATLAS-CONF-2011-089 CERN Geneva

[104] The ATLAS Collaboration In-situ pseudorapidity intercalibration for evaluation of jet energy scale uncertainty using dijet events in proton-proton collisions at $\sqrt{s} = 7$ TeVTech. Rep. ATLAS-CONF-2011-014 CERN Geneva

[105] The ATLAS Collaboration Atlas calorimeter response to single isolated hadrons and estimation of the calorimeter jet scale uncertainty Tech. Rep. ATLAS-CONF-2011-028 CERN Geneva

[106] Mangano M L, Moretti M, Piccinini F, Pittau R and Polosa A D 2003 *JHEP* **0307** 001 (*Preprint* `hep-ph/0206293`)

[107] Skands P Z 2009 (*Preprint* `0905.3418`)

[108] Skands P Z 2010 *Phys. Rev.* **D82** 074018 (*Preprint* `1005.3457`)

[109] Romeo G, Schwartzman A, Piegaia R, Carli T and Teuscher R Jet energy resolution from in-situ techniques with the atlas detector using proton-proton collisions at a center of mass energy $\sqrt{s} = 7$ TeVTech. Rep. ATL-COM-PHYS-2011-240 CERN Geneva

[110] The ATLAS Collaboration Light-quark and gluon jets in atlas Tech. Rep. ATLAS-CONF-2011-053 CERN Geneva

[111] The ATLAS Collaboration Atlas jet energy scale uncertainties using tracks in proton proton collisions at $\sqrt{s} = 7$ tev Tech. Rep. ATLAS-CONF-2011-067 CERN Geneva

[112] The ATLAS Collaboration Performance of the atlas jet trigger in the early $\sqrt{s} = 7$ TeV data Tech. Rep. ATLAS-CONF-2010-094 CERN Geneva

[113] Barlow R and Beeston C 1993 *Computer Physics Communications* **77** 219–228

[114] Aad G *et al.* (ATLAS Collaboration) 2011 *Phys.Rev.Lett.* **106** 172002 (*Preprint* `1102.2696`)

[115] Nappi A *Computer Physics Communications* (*Preprint* `0803.2711`)

[116] Verkerke W and Kirkby D *RooFit Users Manual v2.07*

[117] The ATLAS Collaboration Measurement of the inclusive and dijet cross section of *b*-jets in *pp* collisions at $\sqrt{s}$ = 7 TeV with the atlas detector Tech. Rep. ATLAS-CONF-2011-056 CERN Geneva

[118] Aad G *et al.* (ATLAS Collaboration) 2011 *Eur.Phys.J.* **C71** 1846 long author list - awaiting processing (*Preprint* `1109.6833`)

[119] Aad G *et al.* (Atlas Collaboration) 2011 *Eur.Phys.J.* **C71** 1512 (*Preprint* `1009.5908`)

[120] Aoun S, Boudreau J, Fleckner J, Hirsch F, Jung C, Mueller J, Strandberg S and Talby M 2011 Calibration of *b*-jet tagging efficiency using the p_t^{rel} method Tech. Rep. ATL-COM-PHYS-2011-043 CERN Geneva

[121] The ATLAS Collaboration Measurement of inclusive jet and dijet cross sections in proton-proton collision data at 7 TeV centre-of-mass energy using the atlas detector Tech. Rep. ATLAS-CONF-2011-047 CERN Geneva

[122] Ellis R K, Ross D A and Terrano A E 1981 *Nuclear Physics B* **178** 421 – 456 ISSN 0550-3213 URL `http://www.sciencedirect.com/science/article/pii/0550321381901656`

[123] Catani S and Seymour M H 1996 *Phys.Lett.* **B378** 287–301 (*Preprint* `hep-ph/9602277`)

[124] Campbell J M and Ellis R K 2000 *Phys.Rev.* **D62** 114012 (*Preprint* `hep-ph/0006304`)

[125] Campbell J M, Ellis R K and Williams C *MCFM 6.0 - A Monte Carlo for FeMtobarn processes at Hadron Colliders* http://mcfm.fnal.gov/

[126] Frixione S, Nason P and Oleari C 2007 *JHEP* **0711** 070 (*Preprint* `0709.2092`)

[127] Nason P 2004 *JHEP* **0411** 040 (*Preprint* `hep-ph/0409146`)

[128] Whalley M R, Bourilkov D and Group R C 2005 (*Preprint* `hep-ph/0508110`)

[129] Nadolsky P M *et al.* 2008 *Phys. Rev.* **D78** 013004 (*Preprint* `0802.0007`)

[130] Lai H L *et al.* 2010 *Phys. Rev.* **D82** 054021 (*Preprint* `1004.4624`)

[131] Martin A, Stirling W, Thorne R and Watt G 2009 *The European Physical Journal C - Particles and Fields* **64**(4) 10.1140/epjc/s10052-009-1164-2 URL `http://dx.doi.org/10.1140/epjc/s10052-009-1164-2`

[132] Aad G *et al.* (ATLAS Collaboration) 2011 *Eur.Phys.J.* **C71** 1795 (*Preprint* `1109.5816`)

[133] Frixione S, Nason P and Webber B R 2003 *JHEP* **0308** 007 (*Preprint* `hep-ph/0305252`)

[134] Bourilkov D, Group R C and Whalley M R 2006 (*Preprint* `hep-ph/0605240`)

[135] Gieseke S 2005 *JHEP* **0501** 058 (*Preprint* `hep-ph/0412342`)

[136] Alioli S, Nason P, Oleari C and Re E 2010 *JHEP* **1006** 043 (*Preprint* `1002.2581`)

[137] Frixione S, Nason P and Ridolfi G 2007 *JHEP* **0709** 126 (*Preprint* `0707.3088`)

[138] Alioli S, Hamilton K, Nason P, Oleari C and Re E 2011 *JHEP* **04** 081 (*Preprint* `1012.3380`)

[139] Latunde-Dada O, Gieseke S and Webber B 2007 *JHEP* **0702** 051 (*Preprint* `hep-ph/ 0612281`)

[140] Nason P and Webber B 2012 27 pages, 20 figures (*Preprint* `1202.1251`)

[141] Aad G *et al.* (ATLAS Collaboration) 2012 (*Preprint* `1202.4892`)

[142] Khachatryan V *et al.* (CMS Collaboration) 2011 *JHEP* **1103** 136 (*Preprint* `1102. 3194`)

[143] Altarelli G, Mele B and Ruiz-Altaba M 1989 *Z.Phys.* **C45** 109

[144] Aad G *et al.* (ATLAS Collaboration) 2012 *Phys.Lett.* **B708** 37–54 (*Preprint* `1108. 6311`)

[145] The ATLAS Collaboration Search for new physics in dijet mass and angular distributions using 4.8/fb of *pp* collisions at $\sqrt{s} = 7$ TeV collected by the atlas detector Tech. Rep. ATLAS-CONF-2012-038 CERN Geneva

[146] Choudalakis G 2011 34 pages, 91 figures, prepared for PHYSTAT2011 (*Preprint* `1101.0390`)

[147] Aggarwal R and Caldwell A 2012 *Eur.Phys.J.Plus* **127** 24 (*Preprint* `1112.2593`)

[148] Altarelli G, Mele B and Ruiz-Altaba M *Z. Phys. C* **47**(4) URL `http://dx.doi. org/10.1007/BF01552335`

[149] Aad G *et al.* (ATLAS Collaboration) 2011 *New J.Phys.* **13** 053044 and references therein (*Preprint* `1103.3864`)

[150] The ATLAS Collaboration Search for high-mass dilepton resonances with 5 fb^{-1} of *pp* collisions at $\sqrt{(s)} = 7$ TeV with the atlas experiment Tech. Rep. ATLAS-CONF-2012-007 CERN Geneva

[151] Alitti J *et al.* (UA2 Collaboration) 1993 *Nucl.Phys.* **B400** 3–24

[152] Abe F *et al.* ((CDF Collaboration)) 1999 *Phys. Rev. Lett.* **82**(10) 2038–2043 URL `http://link.aps.org/doi/10.1103/PhysRevLett.82.2038`

[153] The ATLAS Collaboration Tracking studies for *b*-tagging with 7 TeV collision data with the atlas detector Tech. Rep. ATLAS-CONF-2010-040 CERN Geneva

Danksagung

Aus Datenschutzgründen entfernt.