Measurement of the inclusive jet cross section in proton-proton collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector

Dissertation
zur Erlangung des Grades
Doktor der Naturwissenschaften
am Fachbereich Physik
der Johannes-Gutenberg-Universität

von
Sebastian Eckweiler
geboren in Bad Kreuznach

Mainz, den 17. Juni 2011

Tag der mündlichen Prüfung: 27. Oktober 2011

Dissertation an der Universität Mainz (D77).

Abstract

The dominant process in hard proton-proton collisions is the production of hadronic jets. These sprays of particles are produced by colored partons, which are struck out of their confinement within the proton. Previous measurements of inclusive jet cross sections have provided valuable information for the determination of parton density functions and allow for stringent tests of perturbative QCD at the highest accessible energies.

This thesis will present a measurement of inclusive jet cross sections in proton-proton collisions using the ATLAS detector at the LHC at a center-of-mass energy of 7 TeV. Jets are identified using the anti-k_t algorithm and jet radii of $R = 0.6$ and $R = 0.4$. They are calibrated using a dedicated p_T and η dependent jet calibration scheme. The cross sections are measured for 40 GeV $< p_T \lesssim 1$ TeV and $|y| < 2.8$ in four bins of absolute rapidity, using data recorded in 2010 corresponding to an integrated luminosity of $3\,\mathrm{pb}^{-1}$. The data is fully corrected for detector effects and compared to theoretical predictions calculated at next-to-leading order including non-perturbative effects. The theoretical predictions are found to agree with data within the experimental and theoretic uncertainties.

The ratio of cross sections for $R = 0.4$ and $R = 0.6$ is measured, exploiting the significant correlations of the systematic uncertainties, and is compared to recently developed theoretical predictions. The underlying event can be characterized by the amount of transverse momentum per unit rapidity and azimuth, called ρ_{UE}. Using analytical approaches to the calculation of non-perturbative corrections to jets, ρ_{UE} at the LHC is estimated using the ratio measurement. A feasibility study of a combined measurement of ρ_{UE} and the average strong coupling in the non-perturbative regime α_0 is presented and proposals for future jet measurements at the LHC are made.

Kurzfassung

Der dominierende Prozess in harten Proton-Proton-Kollisionen ist die Produktion von hadronischen Jets. Diese Teilchenschauer werden durch farbgeladene Partonen erzeugt, die aus Ihrem Einschluss im Proton gestreut werden. Bisherige Messungen des inklusiven Jet-Wirkungsquerschnitts haben wertvolle Informationen zur Bestimmung von Partondichteverteilungen geliefert und erlauben präzise Tests pertubativer QCD bei höchsten Energien.

Diese Dissertation dokumentiert eine Messung des inklusiven Jet-Wirkungsquerschnitts in Proton-Proton-Kollisionen mit dem ATLAS Detektor am LHC, bei einer Schwerpunktsenergie von 7 TeV. Jets werden mit dem anti-k_t Algorithmus und den zwei Radien $R = 0.6$ und $R = 0.4$ rekonstruiert. Die Kalibration erfolgt mittels einer dedizierten, p_T- und η-abhängigen Kalibration. Die Wirkungsquerschnittsmessung erfolgt in dem Bereich $40\ \text{GeV} < p_T \lesssim 1\ \text{TeV}$ und $|y| < 2.8$ in vier Bins des Betrags der Rapidität. Die verwendeten Daten wurden im Sommer 2010 aufgenommen und entsprechen einer integrierten Luminosität von $3\,\text{pb}^{-1}$. Die Daten werden vollständig auf Detektoreffekte korrigiert und werden mit theoretischen Vorhersagen nächst zu führender Ordnung einschliesslich nicht-perturbativer Korrekturen verglichen. Die theoretischen Vorhersagen stimmen mit der Messung innerhalb der experimentellen und theoretischen Unsicherheiten überein.

Das Verhältnis der Wirkungsquerschnitte für $R = 0.6$ und $R = 0.4$ wird ebenfalls gemessen und mit kürzlich entwickelten, theoretischen Vorhersagen verglichen. Dieses Verhältnis profitiert von den signifikanten Korrelationen der systematischen Unsicherheiten. Die Wechselwirkung von sekundären Partonen aus den Proton-Resten lässt sich mit Hilfe des Transversalimpulses pro Rapidität und Azimuth, genannt ρ_{UE}, charakterisieren. Mit Hilfe von analytischen, nicht-perturbativen Korrekturen für Jets erlaubt das Wirkungsquerschnitt-Verhältnis eine Bestimmung von ρ_{UE} am LHC. Eine Machbarkeits-Studie einer kombinierten Messung von ρ_{UE} und der gemittelten starken Kopplung im nicht-perturbativen Regime α_0 wird gezeigt und weiterführende Jet-Messungen am LHC werden vorgeschlagen.

Table of contents

1 Introduction

By 400 B.C., the Greek philosopher Demokrit brought up the idea of a basic constituent called *Atom* which makes up all matter and is itself indivisible. But only in the 20^{th} century, the idea of the Atom was revived and gained a modern and more precise interpretation. The field of elementary particle physics continues to pursue Demokrit's thoughts about what matter is composed of and tries to describe nature's laws at their smallest distance scales.

The transition to the modern field of elementary particle physics can be considered to have taken place 1897. At that time the British physicist J. J. Thomson discovered the electron, while studying electrical currents. The electron, still today, is considered an elementary particle. Today's picture of what Atoms are, was founded with Ernest Rutherford's discovery of the proton as the atomic nucleus of hydrogen atoms by 1919. It was completed by the discovery of the neutron by James Chadwick by 1932.

The picture of Atoms being formed of neutrons, protons and electrons lasted approximately until the 1960s. By then, a multitude of experiments in the field of particles physics had lead to the discovery of numerous new, apparently elementary particles. The symmetries and characteristics of these particles were explained by the *quark* model, according to which all these particles are composed of either a quark-antiquark pair or three quarks. These discoveries were enabled by great achievements in the development of particle detectors as well as accelerators. Particles accelerated to increasingly higher energies allowed to probe the structure of the collision targets with increasingly better resolution. At the same time, higher energies made it possible to directly produce new particles that cannot be observed naturally. Consequently, discoveries of new particles or new phenomena often came along with newly built experiments operating at higher energies. Prominent examples of such experiments were UA1 and UA2 placed at the Super-Proton-Antiproton-Synchrotron at CERN, which, in 1983, discovered the $W^{\pm}$ bosons and the Z^0 boson and the DØ and CDF experiment at the Tevatron, which discovered the top quark only 16 years ago. These experiments reflect the latest achievements in the completion of today's so called *Standard Model* of particle physics, which is able to describe a vast amount of experimental data to a very good precision.

Yet, one part of the theoretical picture of the Standard Model still awaits to be discovered. This missing particle is the so called *Higgs* boson, the manifestation of a Higgs field that is supposed to interact with all massive particles. The hunt for this missing piece of the standard model is one of the main motivations for the construction of the **L**arge **H**adron **C**ollider (LHC), which is supposed to provide proton-proton collisions at unprecedented center-of-mass energies of 14 TeV, and currently runs at reduced center-of-mass energies of 7 TeV. A further open question in current particle physics is the composition of matter and energy within the known universe. According to astrophysical observations only 4% of the universe is made of the currently known forms of matter and energy, whereas 23 % consists of currently unknown forms of *dark matter*. Theoretical models of physics beyond the Standard Model predict the existence of new particles, which may form this dark matter. Physics at the LHC here will certainly provide important constraints

for these models.

Complementary to such searches is the exploration of the known Standard Model at the new energy scales accessible at the LHC. Ordinary events at the LHC for instance will be dominated by processes of the strong interaction, the interaction of quarks and gluons. The understanding of these events is a prerequisite in order to be able to decide whether observations at the LHC are compatible with the current Standard Model or not. Events induced via the strong interaction are dominated by the production of so called *jets*. These are sprays of particles which are produced by quarks and gluons which are struck out of the proton. Unambiguous prescriptions how to recombine these particles to a single physical object, are provided by so called *jet algorithms*.

The measurement of the production cross section of such jets provides a direct handle to probe the theory of the strong interaction at the highest accessible energy scales. The reach to very high momentum scales also allows to analyze the proton structure with very high resolution. In analogy to optical microscopy this resolution increases with the momentum of the incident particle. Jet physics at the LHC will hence give access to unprecedented resolutions in probing the structure of protons.

This thesis reports on a measurement of inclusive jet cross sections at the LHC, using the ATLAS detector. In the second chapter the theoretical foundations especially of the strong interaction will be outlined and Chapter 3 will briefly review some of the latest, related measurements by other experiments. In Chapter 4 the theoretical predictions for the measurement will be made. After an introduction to the ATLAS detector in Chapter 5, Chapter 6 will outline how jets in the ATLAS detector are reconstructed and calibrated. Chapter 7 will subsequently present the important characteristics of jets measured in the ATLAS detector and is devoted to the validation of the Monte Carlo simulations which play an important role in the subsequent analysis. Important studies done by the ATLAS collaboration that also impact the presented analysis are reviewed in Chapter 8. The main challenge of the actual analysis is the correction of detector effects that is necessary to provide a measurement that is directly comparable to theoretical predictions. This procedure and the obtained results are described in Chapter 9. The results of the cross section measurement using two variants of jet definitions are examined more closely by measuring the ratio of the two observed cross sections. This analysis benefits from the positive correlations of the experimental uncertainties and is presented in Chapter 10.

2 Theoretical foundations

2.1 The Standard Model of particle physics

The so called *Standard Model of Particle Physics* describes the dynamics and interactions of all currently known elementary particles. It postulates two sorts of particles: The first of which, fermions, are the constituents of all non-elementary matter. The second class of particles, bosons, mediates the forces between the fermions, effectively keeping their bound states together.

Before the establishment of the Standard Model, the known elementary particles included the electron, proton and neutron which form all sorts of atoms matter is made of. After more and more observations of new, apparently elementary particles by the beginning of the 1960's a multitude of such particles was known. They were generally called hadrons and classified into two sub-categories according to their masses: Baryons (*barys*, greek for *heavy*), to which the heavier particles were assigned to and mesons (*mesos*, greek for *medium*), particles with medium mass. In an attempt to explain the observed characteristics and symmetries of these particles in a systematic way, Gell-Mann and Zweig by 1964 postulated the so called quark-model. It claimed that all these particles were bound states of a new class of particles, the so called quarks. They were assumed to come in three different *flavors*, up (u), down (d) and strange (s) quarks and their respective anti-particles. Additionally they were supposed to be carrying an electric charge of either $\pm\frac{1}{3}e$ or $\pm\frac{2}{3}e$, where e is the electron charge. In this quark model, all mesons are composed of a quark-antiquark-pair, while all baryons are bound states of three quarks. Protons (neutrons) for instance, are composed of two up-quarks and one down-quark (two down-quarks and one up-quark). This model was able to almost perfectly describe the spectrum of known hadrons.

However, it also contained several flaws by the time of its postulation. For instance no force was known, that should be holding three quarks together within a baryon, given their partly equal electric charge. The crucial item that brought the need for a significant extension of the quark model, was the existence of, e.g., the Δ^{++} baryon. Being a spin-1/2 particle, quantum dynamics required that its wave function needed to be totally antisymmetric. The fact that the Δ^{++} in the quark model was composed of three up-quarks, whose only degree of freedom was their spin, was an apparent contradiction to Pauli's exclusion principle

The solution to this dilemma was the assignment of a new quantum number to quarks, called *color*. This new kind of charge comes in three different types, which were arbitrarily named after the three base colors, red (r), green (g) and blue (b). Accordingly, anti-quarks carry anti-red ($\bar{\text{r}}$), anti-green ($\bar{\text{g}}$) or anti-blue ($\bar{\text{b}}$) color charge. This new degree of freedom solved the contradiction with the Pauli principle. Along with this charge came the hypothesis of the gluon, carrying a *pair* of color and anti-color, which was introduced as the particle mediating the force between color-charged particles. To accommodate that no free quarks or gluons had been observed, the quark model also postulated that bound states only exist as color singlet states. This feature, together with the observations

that quarks behave almost like free particles inside the proton, came to be known as *confinement*.

To date, three more quarks were discovered: the bottom (b), charm (c) and top (t) quark. Theses quarks, together with the leptons and the bosons form the elementary particles of the Standard Model, seen in Fig. 2.1. Quarks and leptons come in three

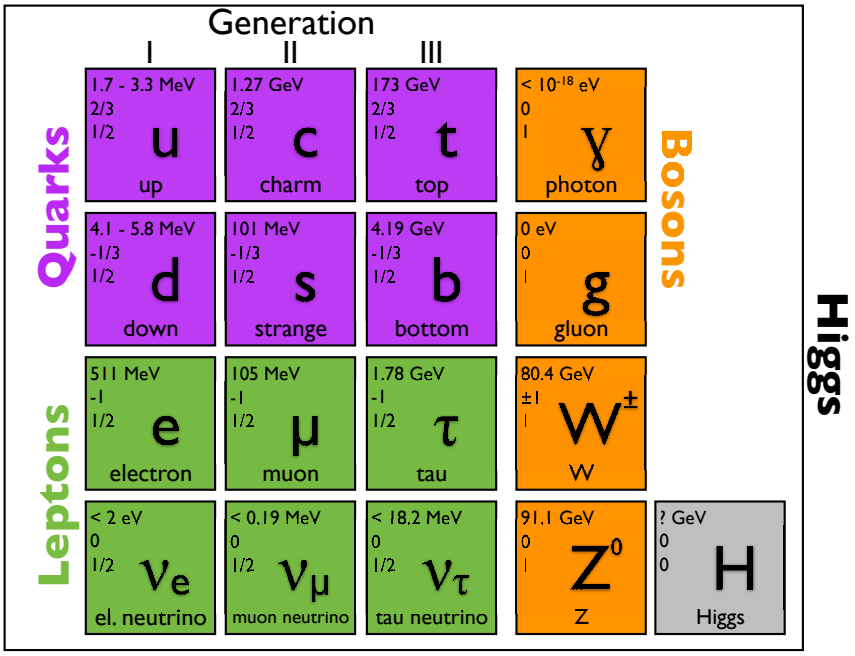

Figure 2.1: *Particles of the Standard Model and their properties. For every fermion shown, there exists the respective anti-fermion with opposite charge. Masses taken from Ref. [1] ([2] for the top quark).*

generations, each consisting of an up-type and a down-type quark along with the charged lepton and its neutrino. The only distinction between the three generations are the masses of the associated particles. Since the particles of the second and third generation are increasingly massive, they all decay into first generation particles. As a consequence, all known stable matter is made of quarks and leptons of the first generation.

2.2 The electroweak interaction

The electroweak theory is the unification of Quantum Electrodynamics (QED) and the weak interaction, established by Glashow, Weinberg and Salam by 1961. The unification postulates, that the weak interaction and electromagnetism are just two manifestations of a single force. In terms of quantum field theory, the electroweak sector is described by the SU(2) × U(1) symmetry group. The ad-hoc requirement of local gauge invariance under these symmetry groups, introduces four new gauge fields to the theory. The physically observable four gauge bosons of the electroweak interaction, the photon, the Z^0 and the $W^{\pm}$ are represented by mixed states of these fields.

At this state in the theory however, all four bosons are massless, which contradicts the experimental observation of only the photon being massless. Hence, the implied symmetries are not realized by nature, but have to be broken. In the current Standard Model, this spontaneous symmetry breaking is caused by the existence of the scalar Higgs

field to which all massive particles couple. Through their coupling to the Higgs field, the Z^0 and the $W^{\pm}$ become massive. Additionally, also the fermions couple to the Higgs field. The strength of this *Yukawa* coupling to the Higgs field determines a fermion's mass: the stronger the interaction with the Higgs field, the larger the mass of the particle.

A consequence of the existence of the Higgs field is the prediction of a massive Higgs boson, which should be observable. This missing part of the Standard Model still awaits its discovery and was one of the main motivations for building the Large Hadron Collider and the associated experiments.

2.3 The strong interaction

According to the name *color* for the newly invented charge of gluons and quarks, the theory of their interactions was named Quantum Chromodynamics (QCD). QCD is based on the color symmetry group SU(3), motivated by the fact that quarks of different color were observed to behave identically. As in the electroweak sector, the requirement of the theory being invariant under local gauge transformations introduces the gauge fields to the theory. The symmetry group being SU(3), implies that there are eight such gauge fields, which represent the gluons. The most distinct feature of QCD is, that, in contrast to QED, the SU(3) symmetry implies that the theory is non-Abelian. For the physics described by the theory, this reflects the fact, that gluons carry color charge, unlike the photon in QED. Consequently, they are also able to interact with themselves, unlike the photon. The full dynamics of the strong interaction is described by the QCD Lagrangian:

$$\mathcal{L} = \sum_f \bar{\psi}_{f,a}(i\gamma^\mu \partial_\mu \delta_{ab} - m_f \delta_{ab} - g_s \gamma^\mu t^C_{ab} \mathcal{A}^C_\mu)\psi_{f,a} - \frac{1}{4} F^A_{\mu\nu} F^{A\mu\nu}, \tag{2.1}$$

where repeated indices are summed over and the γ^μ are the Dirac γ-matrices. Here the $\psi_{f,a}$ denote the quark fields of flavor f with mass m_f and the color index a. The $\mathcal{A}^C_\mu$ with $C = 1, .., 8$ represent the eight gluon fields. The t^C_{ab} are the eight generators of the SU(3) group and $g_s = \sqrt{4\pi\alpha_s}$ is the QCD coupling constant. The field strength tensor $F^A_{\mu\nu}$ is given as:

$$F^A_{\mu\nu} = \partial_\mu \mathcal{A}^A_\nu + \partial_\nu \mathcal{A}^A_\mu - g_s f_{ABC} \mathcal{A}^B_\nu \mathcal{A}^C_\nu \qquad \left[t^A, t^B\right] = i f_{ABC} t^C \tag{2.2}$$

The first two of the three terms enclosed in brackets in Eq. 2.1 correspond to the Dirac-Lagrangian of free fermions with masses m_f, while the third one describes the interaction of quark and gluon fields. The last, non-Abelian term in Eq. 2.2 is responsible for the existence of the gluon self-interaction, which distinguishes QCD from QED. The capability of gluons to interact with themselves leads to the so called asymptotic freedom [3, 4]. Asymptotic freedom manifests itself in the fact, that quarks and gluons behave as free particles at short distances (high momentum scales), while they are confined at large distances (low momentum). The origin of asymptotic freedom is the phenomenon of vacuum polarization. In QED, vacuum polarization is induced by the production of transient e^+e^- pairs in an electric field, originating from a sole, charged particle. These pairs reduce the effective charge of the particle, the larger the distance to the particle gets. This effect is known as *screening* and results in the fact that the electromagnetic coupling constant decreases as the distance increases. Analogue to QED, transient $q\bar{q}$ pairs, produced out of gluons, lead to a screening of color charge. Contrary to QED however, gluons may not only produce $q\bar{q}$ pairs, but also pairs of additional gluons. The

impact of gluons, contrary to the quarks, is to increase the apparent color charge with increasing distance. This effect overwhelms the screening nature of quarks and overall leads to an anti-screening effect. Hence, in QCD the coupling increases with increasing distance or decreasing momentum transfer. In consequence, for large momentum transfers, the strong coupling, denoted by α_s, becomes less strong and quarks inside protons tend to behave as quasi-free particles.

This behaviour separates practical applications of QCD into two regimes: At high momentum transfer, where the coupling constant is sufficiently small ($\mathcal{O}(0.1)$), calculations can be treated perturbatively. At lower momentum scales, where the coupling increases, perturbative calculations are not possible and QCD here is commonly described by suitable phenomenological models. Important aspects of both regimes of QCD will be outlined in the following.

2.3.1 Aspects of perturbative QCD

The QCD Lagrangian describes three fundamental kinds of interactions in QCD: Quark-antiquark annihilation, gluon-gluon fusion and an interaction of four gluons. These interactions can be represented by three fundamental Feynman graphs, as illustrated in Fig. 2.2. Every arbitrarily complex interaction in QCD can be represented by a Feynman

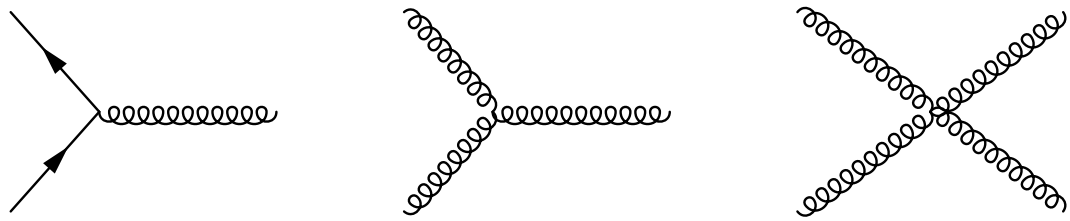

Figure 2.2: *The three fundamental interactions in Quantum Chromodynamics. Solid lines represent quarks and curly lines represent gluons.*

diagram, composed of the three fundamental ones. By using Feynman diagrams a process can not only be visualized, but also its cross section can be calculated using the so called *Feynman rules*. These rules associate exact mathematical prescriptions to each fundamental interaction. These can be combined to calculate the cross section for a process represented by one or multiple diagrams, composed of several fundamental interactions. Each quark-antiquark annihilation or gluon-gluon fusion vertex for instance contributes a factor α_s to the cross section, while the four-gluon vertex comes with a factor α_s^2.

In the framework of perturbative QCD (pQCD), such calculations are expanded in powers of α_s:

$$\sigma^{pQCD} = \sum_i \alpha_s^i \sigma_i \tag{2.3}$$

were the σ_i are contributions to σ^{pQCD} at the i^{th} order of α_s. These calculations provide *partonic* cross sections: cross sections for processes where the incoming and outgoing particles are either quarks or gluons. For the example of inclusive jet production, the leading order (LO) contributions are $2 \to 2$ parton scatterings which are of $\mathcal{O}(\alpha_s^2)$. Next-to-leading order (NLO) contributions at $\mathcal{O}(\alpha_s^3)$ arise from $2 \to 3$ scatterings (real contributions) and $2 \to 2$ scatterings with one loop (virtual contributions). Examples of

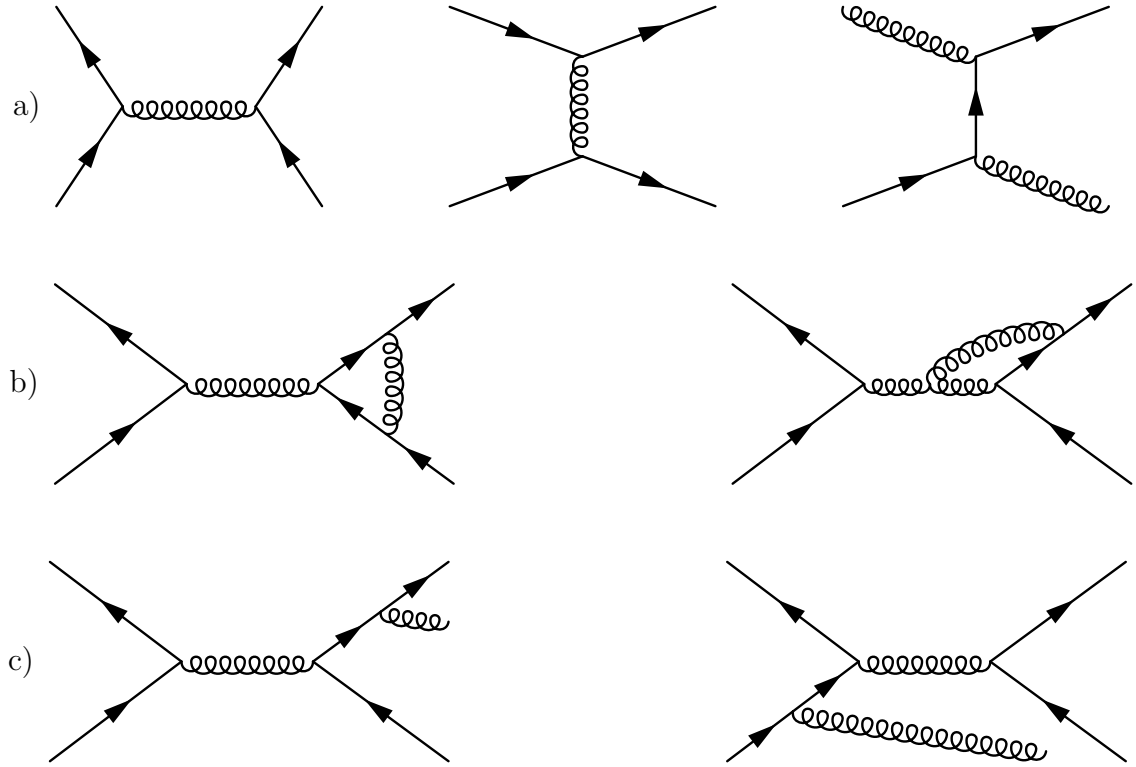

Figure 2.3: *Examples for LO diagrams (a), virtual NLO contribution (b) and real NLO contributions (c) with final state radiation (FSR) on the left and initial state radiation (ISR) on the right.*

LO and NLO diagrams for inclusive jet production are given in Fig. 2.3. The inclusion of the NLO contributions brings considerable complications to the calculation of the matrix elements. These are due to the fact that the probability for a quark to radiate a gluon, diverges for $E_g \to 0$ and $\theta_g \to 0$, with E_g and θ_g being the gluon's energy and its angle with respect to the quark respectively. Hence, both, real and virtual, contributions alone are infrared divergent. The sources of both divergences have the same origins, but come with a different sign and thus cancel each other. This however only holds in case the observable is infrared and collinear safe. That is, the observable $\mathcal{O}$ must not be sensitive to soft or collinear emissions, which can be formulated as in [1]:

$$\begin{aligned} \mathcal{O}_{n+1}(p_1, ..., p_s, ..., p_n) &\to \mathcal{O}_n(p_1, ..., p_n) && \text{for} \quad p_s \to 0 \\ \mathcal{O}_{n+1}(p_1, ..., p_a, p_b, ..., p_n) &\to \mathcal{O}_n(p_1, ..., p_a + p_b, ..., p_n) && \text{for} \quad p_a \parallel p_b, \end{aligned} \tag{2.4}$$

where the first equality implies infrared safety and the latter one collinear safety. Taking into account the NLO contributions significantly reduces the unphysical scale dependency of observables, due to improved convergence of the α_s expansion.

Loop diagrams also introduce the issue of ultraviolet divergences. These are due to the fact that mathematically, the momentum of e.g. a gluon within a loop can be infinite. Integrals over the momentum of particles in loops hence diverge. These integrals are renormalized by the introduction of a cut-off at the so called *renormalization scale* μ_R. This cut-off is absorbed in a renormalized strong coupling constant, which, by this absorption inherits a dependency on μ_R. The strong coupling's world average is quoted

at $\mu_R = M_Z$, the Z^0 mass, to be [5]:

$$\alpha_s(M_Z) \;=\; 0.1184 \;\pm\; 0.0007.$$

Next-to-leading order calculations are the best means to predict additional hard radiation off a $2 \to 2$ process, for instance producing a third jet. At the same time, they do not include possible, numerous soft radiations, which may not produce additional jets, but change the inner structure of jets. Such features are implemented as so called parton showers, which are used along with LO calculations especially in Monte Carlo (MC) event generators. Here the generation of a $2 \to 2$ process is followed by a probability based further splitting of incoming and outgoing partons. The generation of these splittings is based on the so called *Sudakov form factor*, $\Delta(Q_0, Q_1)$, which describes the probability, that a given parton does not split into new partons between the scale Q_0 and Q_1. The minimum scale Q_0 serves as a lower cut-off for emissions, which mitigates the soft and collinear divergences. At the same time Q_0 defines the minimum scale of a physically resolvable emission. The simulation of unresolvable soft or collinear emissions can be considered superfluous. With a proper usage of $\Delta(Q_0, Q_1)$ and the generation of random numbers, splittings are created ordered in Q as $Q_i < Q_{i-1} < ... < Q_0$. The choice of the ordering variable Q herein is an in principle arbitrary one. It is however common to generate the first splitting as the hardest one. Common choices for the ordering quantity Q thus are the p_T of the emitted parton or its angle with respect to the parent parton. The definition of the Sudakov form factor is based on a collinear approximation. Hence, contrary to the NLO calculations, parton showers manage to describe soft emissions rather well, while hard emissions are described less precisely.

A combined usage of NLO calculations with subsequent parton showers is possible, but very complex due to the need of a clear separation of the NLO emissions and the parton shower generation.

2.3.2 Aspects of non-perturbative QCD

Due to the confinement, quarks or gluons cannot exist as free particles. Two colored partons, departing from each other with sufficiently high momentum for this reason have to finally form color-neutral hadrons. This formation of hadrons, commonly referred to as hadronization or fragmentation, happens in the non-perturbative regime of QCD and so far eluded calculations from first principle. Several models describing hadronization in a phenomenological or probabilistic way exist. One prominent and intuitive approach is used in the *Lund string model*, employed by the widely used `Pythia6` MC generator [6], which is illustrated in Fig. 2.4. In this string model, hadronization takes place by cascade-like productions of $q\bar{q}$ pairs out of an imagined color-string, spanning the distance between two departing colored partons. As the two quarks depart from each other, the potential energy stored in the color-string increases until the string breaks and produces a new pair of quarks. Both new quarks are now connected via color-strings to the two initial quarks. Depending on the momenta of the two initial quarks, numerous further breaks may occur. This procedure stops, once only quark-antiquark pairs remain whose relative momentum is low enough to form on-mass-shell hadrons.

In the so called *cluster* model, employed e.g. by the `Herwig++` generator [7], quark-antiquark pairs whose color allows them to form color-singlet states form so called clusters. At the end of the hadronization procedure, for each cluster a new quark-antiquark pair is

created out of the vacuum and these two quarks are used to form hadrons together with the two original quarks of the cluster.

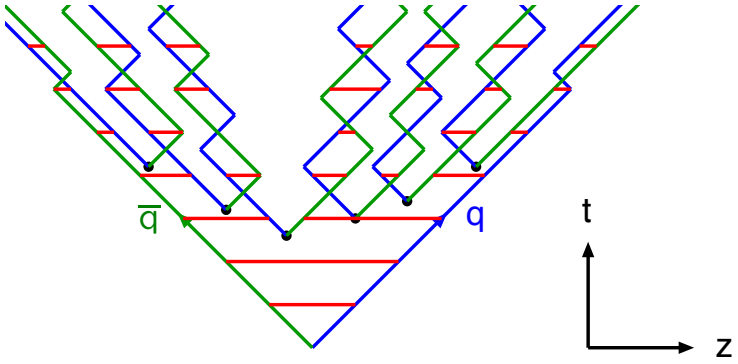

Figure 2.4: *Illustration of the Lund string model for a quark anti-quark pair departing from each other. The red line symbolizes the color string spanning the distance between the quark and anti-quark and eventually breaks up to from a new pair of quarks. Figure taken from Ref. [8].*

For the experiment the phenomenon of hadronization leads to the fact that partons cannot be measured directly in a detector, as it is possible for stable particles. Rather, scattered partons produce numerous stable particles, which become increasingly collimated with increasing momentum of the parent parton. These collimated sprays of particles are called *jets*. Important details of jet identification will be outlined in the last section of this chapter.

2.4 Phenomenology of proton-proton-collisions

The above considerations where restricted to partons. However, since partons are confined, parton collisions are experimentally realized by making e.g. protons collide. This considerably complicates the calculation of physically observable cross sections. The first complication arises from the fact that in the calculation of the partonic cross sections, the momenta of the incoming partons have to be known. Experimentally though, neither the parton's momentum nor its type can be fully controlled. Instead, the colliding partons, constituents of the two colliding protons, only carry a certain fraction of the proton's momentum, denoted by the Bjorken-x variable: $p_{parton} = xp_{proton}$. The probability to find a parton with a certain value of x in a proton is provided by *parton density functions* (PDFs). According to the factorization theorem [9], PDFs are considered to be universal quantities, describing the composition of a proton. Hence, the probabilities to find a parton with a given x inside a proton, are independent of the hard scatter, as indicated in Fig. 2.5. The probability to observe a certain final state thus can be thought of as the probability to first *draw* two partons with a certain momentum from the incoming protons. The convolution of these probabilities with the probability that the scattering of the two partons results in the final state of interest finally gives the probability to observe the process in a proton-proton collision. Mathematically this can be written as:

$$\sigma(pp \to X) = \sum_{a,b} \int_0^1 \mathrm{d}x_a \int_0^1 \mathrm{d}x_b \, \hat{\sigma}_{ab\to X}(x_a, x_b, s, \mu_R^2, \mu_F^2) \, f_{a/p}(x_a, \mu_F^2) \, f_{b/p}(x_b, \mu_F^2) \quad (2.5)$$

$\hat{\sigma}_{ab\to X}$ herein denotes the partonic cross section and it has to be summed over all parton types a and b, with $\sqrt{s}$ being the center-of-mass energy of the two protons. Here the

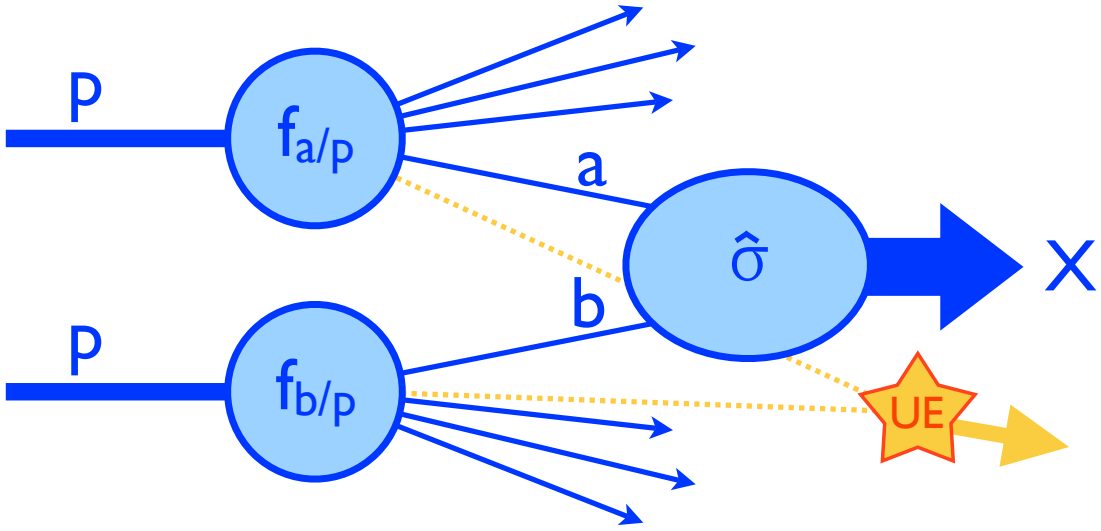

Figure 2.5: *A stylized hadron-hadron collision. The partons a and b enter the hard interactions, whose cross section is denoted by $\hat{\sigma}$. The probabilities to find partons a and b in the proton with a certain momentum fraction is denoted by $f_{a/p}$ and $f_{b/p}$ respectively. Additionally a second pair of softer partons (dashed lines) from the proton remnants may interact to form the underlying event (UE).*

$f_{a/p}(x_a, \mu_F^2)$ represents the PDF: the probability to find the parton of species a with a momentum fraction x_a in the proton. The arbitrary factorization scale determines up to which scale parton emissions are handled by the PDF, rather than considered as part of $\hat{\sigma}_{ab\to X}$. Usually the factorization scale is set equally to the renormalization scale μ_R.

The accommodation of PDFs hence enables the prediction of cross sections measurable in proton-proton-collisions. In a more complete picture of a proton-proton collision, as sketched in Fig. 2.5, however, not only two partons interact. The proton's remnants may also interact, producing the so called *underlying event* (UE) via multiple parton interactions (MPI). These secondary interactions are commonly regarded as $2 \to 2$ QCD processes above a low p_T threshold, denoted by p_T^{min}. For $p_T^{min} \to 0$, the cross section for these processes (σ_{soft}) becomes larger than the total proton-proton cross section σ_{pp}. A common interpretation of this fact is, that the average number of parton-parton interactions, denoted by $\langle N_{int} \rangle$, is simply the ratio of these two cross sections: $\langle N_{int} \rangle = \sigma_{soft}/\sigma_{pp}$. Within this interpretation however, $\langle N_{int} \rangle$ is a function of p_T^{min} and hence diverges for $p_T^{min} \to 0$. To solve also this issue, the wide variety of MC generators employ a range of suppression techniques and further physical arguments to constrain $\langle N_{int} \rangle$ to a finite value. The details of these models and suppression techniques introduces a significant number of parameters, which have to be tuned such, that the MC simulation reproduces real data. Examples for appropriate observables used to perform such tunings are given in Chapter 3.

2.5 The structure of protons

Since the dynamics of PDFs lie within the non-perturbative regime of QCD they cannot be calculated and therefore have to be measured. Today PDFs are determined by a number of collaborations in fits to a wide variety of experimental results. The picture of the proton's composition is such, that its constituents are the three valence quarks (uud),

held together by gluons. The inner dynamics of the proton is governed by splittings of gluons into either quark pairs, forming the so called sea-quark content, or further gluons. The probability for a parton b to undergo either an emission or splitting that yields the parton a with a momentum fraction $f = 1 - z$ is given by the appropriate Altarelli-Parisi splitting function $P_{ab}(z)$ [10]. For instance, the probability that a quark emits a gluon and keeps a fraction z of its original momentum is given by:

$$P_{qq}(z) = \frac{4}{3}\frac{1+z^2}{1-z}. \tag{2.6}$$

The proton can thus be imagined as a compound of ever splitting and recombining partons, leaving the proton as a whole unchanged.

Due to the divergences of the splitting functions in the soft limit ($z \to 1$), increasingly soft splittings are much more abundant. However, at a given momentum scale only splittings down to a certain momentum scale can be resolved. Consequently, the *visible* structure of a proton changes with the momentum scale at which it is probed. This can be seen in analogy to optical microscopy, where the resolution is limited by the wavelength of the light used to inspect a probe. In order to go beyond such a limit, the wavelength has to be decreased, increasing the momentum scale.

The evolution of a certain parton density function $f_{a/p}(x, Q^2)$ as a function of this momentum scale Q is described by the DGLAP evolution equations (see e.g. [1]). Exemplary parton density results are shown in Fig. 2.6 at two very different scales, demonstrating the evolution of the proton structure. Though in the original quark model one would

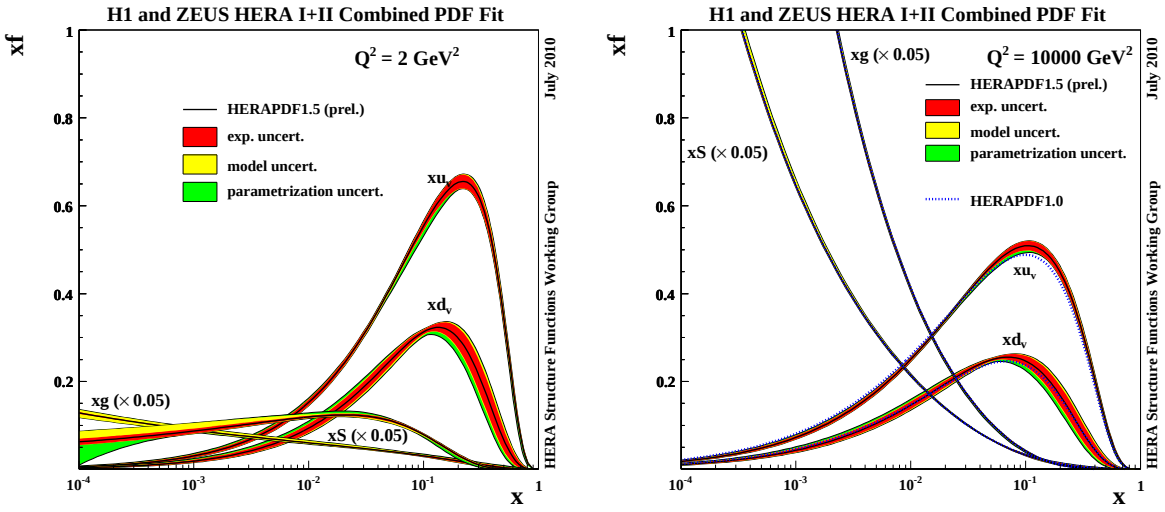

Figure 2.6: *Proton PDFs at $Q^2 = 2$ GeV^2 (left) and $Q^2 = 10000$ GeV^2 (right), taken from Ref. [11, 12]. Parton densities multiplied by x are shown:* xg $= x\, f_{g/p}(x, Q^2)$. S *refers to the summed PDFs of all sea-quarks, whereas* g, $\mathrm{u_v}$ *and* $\mathrm{d_v}$ *show the gluon density and the two valence-quark densities respectively.*

expect that the three valence quarks in total carry the full proton momentum, in fact only $\approx 50\,\%$ is carried by the valence- and sea-quarks. The remaining momentum is carried by the gluons. The most dramatic evolution in the PDF between the two different scales shown in Fig. 2.6 can be observed for the low-x gluon and sea-quark PDFs, which increase by more than one order of magnitude for $x \lesssim 10^{-2}$. This is a result of the fact

that with increasingly high momentum transfer more and more soft $g \to gg$, $g \to q\bar{q}$ and $q \to qg$ splittings can be resolved. The inner structure of protons hence depends on how closely it is examined. Though the DGLAP equations are able to describe the evolution of a certain PDF between two scales, they cannot be used to actually determine a PDF without a starting scale.

This instead is done in PDF fits. Here DGLAP evolution is the essential component to handle input data measured at widely different momentum scales. Mostly, parton densities are described by a parametric Ansatz with a certain set of free parameters, which is valid at a certain starting scale $Q_0 = \mathcal{O}(1\text{ GeV})$. In a single iteration of a fit procedure, for each input data the PDF Ansatz is evolved to the appropriate momentum scale and theoretical predictions are compared to the data. Considering the large spread of momentum scales in data used in PDF fits, DGLAP evolution can be considered to be extremely successful.

As illustrated by the colored bands in Fig. 2.6, the determination of PDFs is connected with considerable uncertainties. These uncertainties arise from a number of different sources, which are also taken into account in the uncertainty bands shown. Firstly, there are of course experimental uncertainties on the PDFs, that are a direct consequence of the limited precision of the input data, shown as red bands in Fig. 2.6. Secondly, most PDFs rely on a parametrized form for each individual PDF, which constrains the results to a certain degree. The resulting uncertainty from an altered parametrization on the PDFs in Fig. 2.6 is shown as a green band. The model uncertainties shown refer to uncertainties induced by the choice of e.g. the heavy quark masses. These uncertainties to some degree of course impact all PDF determinations. However, the chosen approaches for the estimation of PDF uncertainties differ among the various collaborations performing PDF fits.

The choice of the $\alpha_s(M_Z)$ value is an additional parameter, whose treatment differs among the PDF collaborations. It is either used as a fixed, external parameter or as an additional free parameter, which is fitted along with the PDF itself. For both cases, the uncertainty on the PDF is estimated by comparing PDFs extracted with varying values of $\alpha_s(M_z)$. In addition to the sources of uncertainty mentioned above, another issue is the choice of the input data, that in general is different for various PDFs.

A consequence of these items is, that in general it cannot be expected that predictions using different PDFs agree within their associated uncertainties. Hence, the more realistic estimate of PDF induced uncertainties on any observable should include a variety of PDF sets, to find the actual envelope of the predicted values for the observable.

2.6 Algorithms for jet identification

The identification of jets is only trivial in the very simple scenario of an event with only two partons, as implied by the LO diagrams in Fig. 2.3a). Already at the parton level, NLO scenarios as in Fig. 2.3c) pose the question, whether a gluon from FSR or ISR *belongs* to one of the two outgoing quarks or not. Avoiding these ambiguities is the major requirement for jet algorithms, which should provide a well defined and unambiguous set of rules how to combine single particles into jets.

An additional requirement for jet algorithms may be called resilience. This requirement arises from several issues. The most prominent is that jet algorithms have to be applied to a variety of different input data. Three types of input data are going to be used in the

upcoming analysis. According to the different stages of event simulation, these are:

- **Parton level jets**: Jets at the parton level use simulated parton momenta, that are either generated in (N)LO calculations or by MC event generators, without executing the hadronization step.
- **Particle level jets**: Here, momenta of simulated, color-neutral particles are used. These are simulated using MC event generators including the hadronization step.
- **Reconstructed jets**: For reconstructed jets experimental (either simulated or real) inputs are used, which may be tracks of charged particles or calorimeter signals.

A good jet algorithm has to be made such that it provides stable results at all levels.

Two particularly subtle issues are infrared and collinear safety, as defined in Eq. 2.4. In terms of calculations, unsafe algorithms may spoil the cancellation of real and virtual divergences. Experimentally, infrared emissions can be mimicked for instance by electronic noise as well as by soft particles, which may also originate from the underlying event. Also collinear topologies may occur, e.g. due to energy depositions which are split among neighboring calorimeter cells. Infrared and collinear safety hence are not only abstract flaws, but may induce real and most importantly avoidable pathologies in jet events.

Since first measurements of jets in electron-positron collisions a variety of jet algorithms has been established. It is possible to subdivide all the established algorithms into two distinct classes: Firstly, there is the class of *cone*-algorithms. These algorithms identify jets on a geometrical basis, by grouping together particles which lie within a cone (a circle in the η-ϕ-plane) of certain radius R. Generally, cone algorithms can be summarized as follows:

Step 1) Chose the highest p_T particle in the event as a starting point, called *seed*.

Step 2) Place a cone at the seed's position.

Step 3) Recombine* all particles within the cone and place a new cone along the direction of the recombined momentum. Repeat this procedure until the position of the new cone is stable with respect to the firstly placed cone. Call the recombined object a *proto-jet* and remove it from the list.

Step 4) Repeat steps 1 to 3 until the list is empty.

Step 5) The split & merge procedure: Find all proto-jets sharing energy and merge them if the shared energy exceeds a certain threshold. Else, associate all shared particles to the closest proto-jet.

Due to the usage of seeds, which are only considered above a certain threshold, none of the seed-based cone algorithms is infrared nor collinear safe. Exemplary topologies of infrared and collinear unsafe behaviour of cone algorithms are illustrated in Fig. 2.7. In latest analyses at the Tevatron, the so called Midpoint algorithm was used [13]. This algorithm additionally uses all midpoints between two stable cones as seeds for new cones, before entering the split & merge procedure. This feature aimed to provide a better infrared behaviour, but only reduced the existing problems by one order (see e.g. [14]). Recently,

*Different methods for recombination exist. Here, to recombine a set of particles, their four-momenta are summed. Other recombination schemes exist (cf. [13]), but will not be used in this analysis.

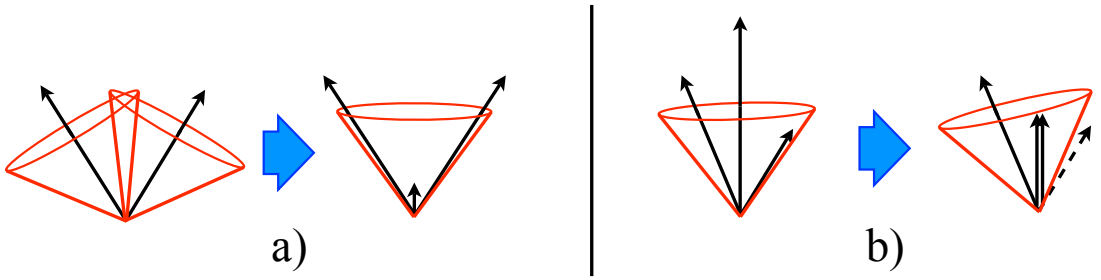

Figure 2.7: *Examples of pathologies caused by infrared (a) and collinear (b) emissions. In a), the additional soft (infrared) particle in between the two harder ones is an additional seed for a cone that includes all three particles. In b) the collinear splitting of the hardest particle leads to a different ordering of seeds. The first stable cone is found without including the dashed particle.*

the *SISCone* algorithm [14] was developed, which, as the only cone based algorithm, features perfect infrared and collinear safety.

The second class of algorithms employs methods to sequentially recombine pairs of particles. Most algorithms of this kind can be described by a common algorithm, which has its origin in the *inclusive k_t algorithm* [15]. Algorithms of this class are generally defined by an abstract distance measure between two particles. Particles which are closest in terms of the defined distance measure are recombined first. Similar to the cone algorithms, also these algorithms make use of a basic radius parameter R, which controls the size of the jets. Introducing a second parameter t, all algorithms inheriting the general aspects of the k_t algorithm can be described as follows:

Step 1) For each pair of particles, define: $d_{\mathrm{ij}} = min(p_{T,i}^t, p_{T,j}^t)\dfrac{\Delta R_{ij}^2}{R^2}$, with ΔR_{ij} being the distance between the particle i and j in η-ϕ-space: $\Delta R_{ij} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

Step 2) For each particle, define: $d_{\mathrm{iB}} = p_{T,i}^t$.

Step 3) Find the smallest of all d_{ij} and d_{iB}.

Step 4) If it is a d_{ij}, recombine particle i and j.

Step 5) If it is a d_{iB}, call it a jet and remove it from the list.

Step 6) Repeat from step no. 1, until no particle is left.

All algorithms of the k_t family are infrared and collinear safe, which is due to the fact that they do not need seeds. Inconveniently large computing times of k_t algorithms, that were still an issue for Tevatron analyses [13], have been solved by more recent implementations [16, 17]. For $t = 2$, this algorithm reproduces the original k_t algorithm. The motivation behind this choice is, that the most soft and collinear radiations are subsequently undone. The k_t algorithm hence effectively tries to reverse QCD radiations. A purely geometrical approach is chosen by the *Cambridge/Aachen* algorithm [18, 19], for which $k = 0$. More recently the anti-k_t algorithm [20], using $t = -2$ was developed, which clusters the hardest particles in an event first. This choice produces very circular jets, as it is the intention of cone algorithms, and at the same time it inherits all benefits of sequential recombination, which will become apparent in the following. The anti-k_t algorithm hence unifies the advantages of cone-algorithms, namely their regular shape, and the infrared and collinear

safety of the clustering algorithms. Due to these reasons, this analysis will in the following use the anti-k_t algorithm.

3 Overview of recent QCD measurements

In order to be able to put the upcoming analysis into perspective of current physics, it is necessary to provide a brief overview of the most recent results in this field. For this purpose, a few selected results that are closely connected to this analysis are presented in the following.

Historically, measurements of jets had a very significant impact on the understanding of QCD. First observations of two-jet events have been made in e^+e^- collisions at the Stanford Linear Collider in 1975 [21]. These very first observations of the hadronization phenomenon confirmed the theoretical picture of confinement. Not much later, the first three-jet events observed in 1979 at PETRA [22] proved the existence of the hypothesized gluon.

In the more recent past, with QCD being a well-established theory, measurements of jet observables help to understand a variety of its aspects. Jet physics directly probes the framework of pQCD up highest scales of transverse momentum, due to the large cross sections. Besides, jet observables, such as inclusive cross sections, help to constrain the proton structure in PDF fits and allow for a measurement of the strong coupling constant at very high energy scales. In the following a small overview of results from the HERA and the Tevatron experiments shall be given.

The underlying event, an inevitable component of jets in hadron-hadron collisions, has significant impact on jet analyses. For this reason, also a recent measurement of the underlying event at the Large Hadron Collider is also presented.

3.1 Results from HERA

At HERA electrons with an energy of 27.5 GeV were brought to collision with 920 GeV protons. Here, in scatterings via charged and neutral currents, the quark content can be probed very directly in the measurement of the proton's structure functions. Being the only electron/positron-proton collider, the results of HERA so far are unique. The reconstruction of the electron's momentum allows for a very precise reconstruction of the kinematics of the underlying scattering. For instance, the measurable momentum-loss of the electron allows to determine the x of the scattered parton in the proton very precisely.

The two HERA collider experiments, H1 and ZEUS, extensively measured the proton structure functions. The measurements spanned a very wide range of x and Q^2 [23]. The kinematic range for the neutral current processes was $10^{-7} < x < 0.65$ with $0.045 < Q^2 < 30000$ GeV^2. For the charged current processes the ranges were slightly smaller. These measurements were able to stringently test the evolution of PDFs described by the DGLAP evolution equations. The success of these equations is impressively demonstrated by the excellent description of the data by next-to-leading order QCD calculations (see e.g. [23]). Today, the large amount of data points from the HERA experiments builds

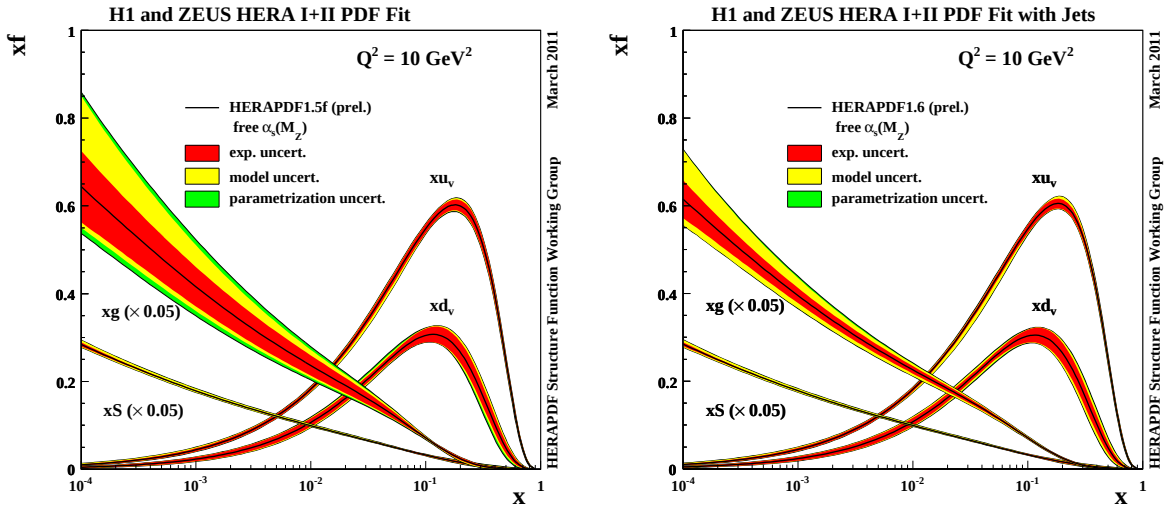

Figure 3.1: *Results of PDF fits without (left) and with (right) jet data. Figure taken from Ref. [11].*

the basis for the understanding of the proton structure. The inclusion of jet data here results in a significantly more precise determination especially of the gluon PDF. This is illustrated in Fig. 3.1, which shows two sets of PDFs determined with and without the incorporation of HERA jet data. The uncertainty on the gluon PDF for instance at low x after the incorporation of the jet data significantly decreases from approximately 20 % to 10 %. This highlights the value of the jet cross sections for the determination of PDFs.

3.2 Results from the Tevatron

At the Tevatron protons and anti-protons are brought to collision at a center-of-mass energy of 1.96 TeV. Compared to the analyses performed at HERA, measurements at Tevatron are naturally much more similar to what can be expected at the Large Hadron Collider (LHC). Inclusive jet cross sections at the Tevatron have been measured by the DØ and CDF collaborations using the Midpoint cone algorithm with $R = 0.7$ as a function of transverse momentum up to approximately 600 GeV. CDF uses a slightly larger dataset of $1.1\,\mathrm{fb}^{-1}$ [24], compared to DØ, which used only $0.7\,\mathrm{fb}^{-1}$ [25]. Both measurements, shown in Fig. 3.2, are in very good agreement with NLO pQCD predictions. The experimental uncertainties at the highest transverse momenta for both measurements are dominated by the jet energy scale uncertainty. For the CDF measurement, the maximum experimental uncertainties are approximately 35 %, whereas the DØ measurement is somewhat more precise and quotes an experimental uncertainty of approximately 20 % in the central rapidity bin. Both measurements meanwhile are incorporated in recent global PDF fits and contribute to an improved uncertainty especially on the gluon PDF at high x.

Besides testing pQCD, inclusive cross section measurements may also hint at new physics. Particularly models predicting quark substructure that may become visible at a certain scale, would manifest themselves in significantly amplified jet cross sections. First inclusive jet cross sections measured at the Tevatron in Run I [26] for instance, showed such higher cross sections at largest transverse energies. However, the excess could also be well explained by an enhanced gluon PDF in the high x region and led to a significant

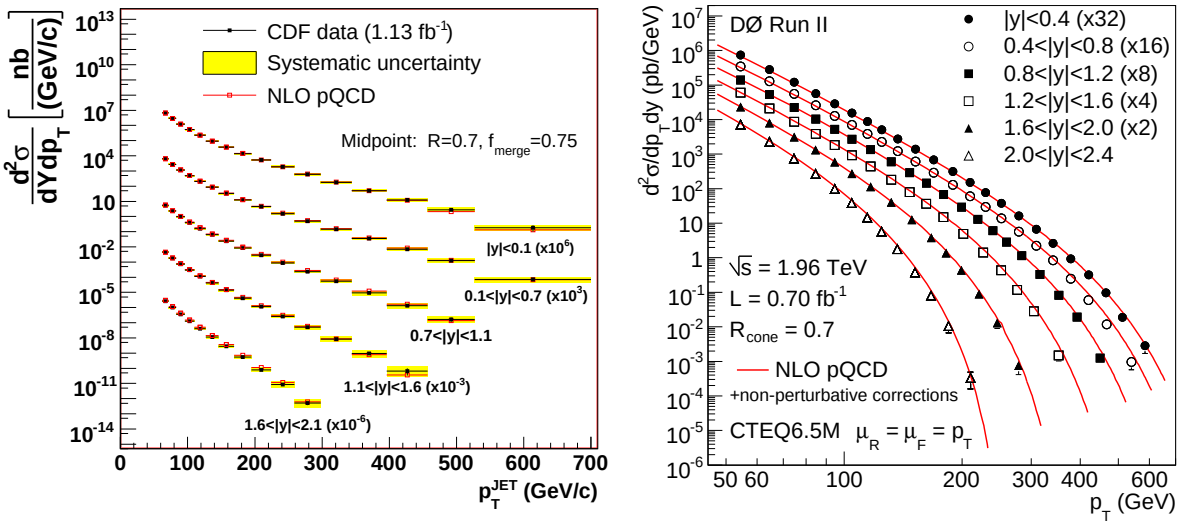

Figure 3.2: *Inclusive jet cross section measurements by CDF (left, from Ref. [24]) and DØ (right, from Ref. [25]).*

change of subsequent PDFs. This illustrates the necessity of precisely known PDFs in the view of searches for deviations from the known Standard Model.

Besides cross section measurements, a variety of other jet observables were measured at the Tevatron (see e.g. [28] for a recent review). Only two examples are the measurements of jet shapes and angular correlations in two-jet events.

The measurement of jet shapes [29] probes the inner structure of jets, whose simulation is an important test of hadronization models and MC generator tunings. The angular correlation of jets allows to test theoretical predictions at NLO. The dominant contribution to jet production are $2 \to 2$ scatterings, which result in two-jet topologies in which both jets are back-to-back, namely separated in azimuth by $\Delta\phi = \pi$. The additional production of third and fourth jets leads to modifications of these topologies, bringing the two leading jets closer together. This was measured by the DØ collaboration [27] and the results are shown in Fig. 3.3. They are compared to LO ($2 \to 3$) and NLO calculations of three jet observables. As expected, the LO computation, which includes only three jets in the final state, is only able to describe the data at intermediate values of $\Delta\phi$. The NLO calculation is able to describe the observations very well across the full range of $\Delta\phi$.

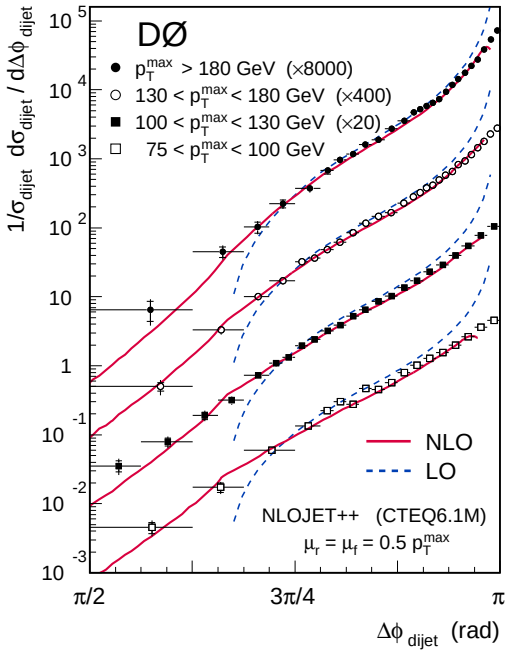

Figure 3.3: *Measurement of the di-jet angular decorrelation. The separation in azimuth between the two leading jets in an event is denoted by $\Delta\phi_{dijet}$. Figure taken from Ref. [27]*

3.3 Measurements of underlying event characteristics

Measurements of the underlying event have become standard analyses at hadron colliders. They provide unique information for the proper tuning of MC generators, which is a crucial ingredient for all other kinds of analyses.

In a naive approach, the underlying event can be considered to be independent of the hard scattering, which in general exhibits a back-to-back, two-jet structure. It can hence be expected that the dynamics in the region perpendicular to the two-jet system, is dominated by the underlying event. This assumption is the basis of the majority of underlying event analyses, which measure the characteristics of particle production in certain azimuthal regions with respect to the hard scattering. A common choice is, to

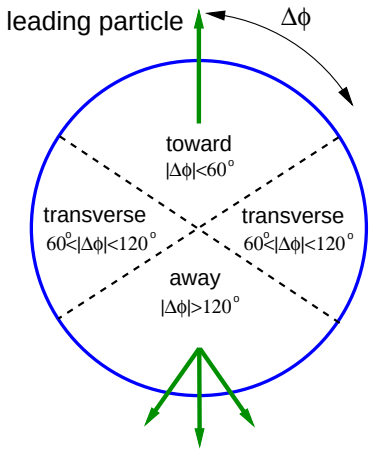

Figure 3.4: *Illustration of the separation of the full azimuthal range into four distinct regions. Figure taken from Ref. [30].*

define the direction of the hard process by selecting the highest p_T object in an event, as illustrated in Fig. 3.4, and to define four distinct azimuthal regions as follows: The *toward*-region, which is identified by the direction of the leading particle and the *away*-region, which is oriented to the opposite direction. This defines the remaining two *transverse* regions, which are oriented perpendicular to the hard scattering.

With this definition, it is expected that the observed particles in the transverse region mainly originate from the underlying event. Commonly used observables include the number of particles observed and features of their transverse momentum distribution. Figure 3.5 shows the measurement of such observables done by the ATLAS collaboration at the LHC using data at a center-of-mass energy of 7 TeV [30]. The left panel shows the scalar sum of the transverse momenta of all particles in the transverse region, normalized to the area in η-ϕ-space as a function of the leading particle's p_T. This observable reveals that the assumption of the underlying event being independent of the hard process is not completely true. The activity in the transverse region evidently scales with the momentum of the leading particle, especially at very low momenta. Only at higher p_T^{lead} the distribution flattens and hence appears to be less correlated with the hard process. The comparisons with various MC generator predictions reveal that they tend to simulate a too soft behaviour of the underlying event.

The right-hand side of Fig. 3.5 shows the average number of particles observed versus their angle with respect to the leading particle. Here, a W-like shape can be observed. Its peaks at $\Delta\phi = 0$ and $\Delta\phi = \pi$ to arise from the particles produced in the hard scattering.

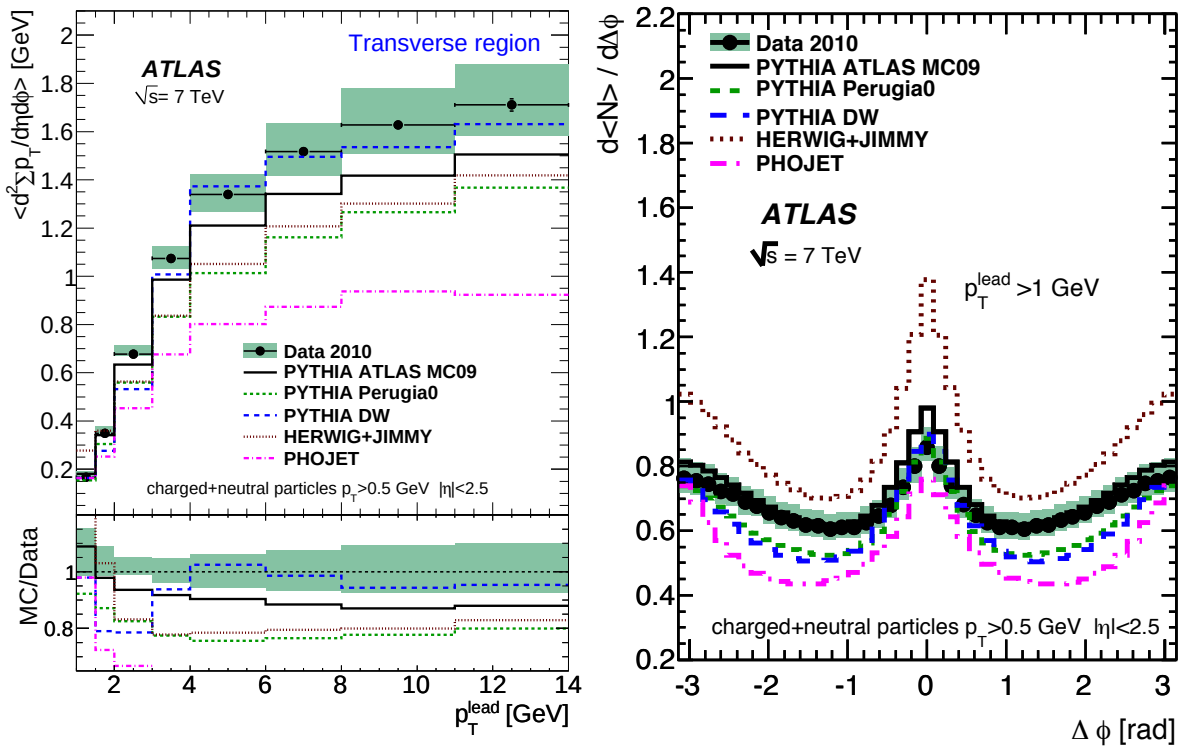

Figure 3.5: *Left: The amount of transverse momentum per unit pseudorapidity and azimuth in the transverse region, as function of the leading track's momentum. Right: The average number of particles observed versus their azimuthal angle with respect to the leading particle. Both figures taken from Ref. [30].*

On the contrary, a significant fraction of the particles produced around $\Delta\phi \approx \pi/2$ is assumed to originate from the underlying event. Similar measurements [31], restricted to charged particles only, have already been used to establish new MC generator tunes, specific for the environment at the LHC.

4 Theoretical predictions

4.1 Theoretical predictions at parton level

The theoretical predictions for the inclusive jet cross section are calculated at next-to-leading order (NLO) using the `NLOJet++` [32, 33] program in combination with the CTEQ6.6 PDF set [34]. `NLOJet++` effectively performs a MC integration of the integral in Eq. 2.5, including final states with up to four partons and one loop. For the MC integration `NLOJet++` firstly generates a randomized event of initial and final state partons along with the cross section calculated at tree level as an event weight. In subsequent iterations, NLO corrections to the tree level event are generated along with their (positive or negative) weights. This results in changed kinematics for instance in the case of the generation of initial state radiation (ISR) or final state radiation (FSR). Once all contributions to a given tree level event have been calculated, a new event is generated. To obtain the cross section of an arbitrary observable, this observable is calculated from the existing parton momenta in each iteration. The total cross section is then given as the sum of all per-event weights divided by the number of tree level events created. Depending on the details of the observable, in particular the binning, very high numbers of iterations ($\mathcal{O}(10^9) - \mathcal{O}(10^{10})$) have to be processed in order to achieve satisfactory cancelation between real and virtual NLO contributions. For the calculation of the inclusive jet cross section, the tree level events contain two final state partons and a third one for the real NLO contributions.

Jets are clustered using the anti-k_t algorithm within the FastJet library [16, 17, 20] with the two radii $R = 0.6$ and $R = 0.4$. For the calculation of the cross section in each iteration the renormalization (μ_r) and factorization (μ_f) scales were chosen as the p_T of the jet with the highest transverse momentum in the event, denoted by p_T^{lead}. The double differential, inclusive jet cross section is defined as:

$$\frac{d^2\sigma(p_T, y)}{dp_T dy} = \frac{\sum_i N_{jets,i}(p_T, y)\, w_i}{N_{tree} \Delta p_T \Delta y}, \tag{4.1}$$

where N_{tree} is the number of tree-level events, the sum runs over all events i and $N_{jets,i}$ and w_i are the number of jets in a given bin found for the i^{th} event with a weight w_i. Results of these calculations are shown in Fig. 4.1 for various bins of rapidity.

4.2 Theoretical uncertainties

In a general picture, the predictions of course are firstly limited by the fact that they are only computable at NLO. Approximations of the NNLO contribution [35] reveal, that the correction arising from contributions beyond NLO can be expected to be rather small compared to the impact of the NLO contribution on the LO prediction. Uncertainties on the NLO prediction arise from the choice of μ_r and μ_f, the value of the strong coupling constant α_s and the PDF uncertainty.

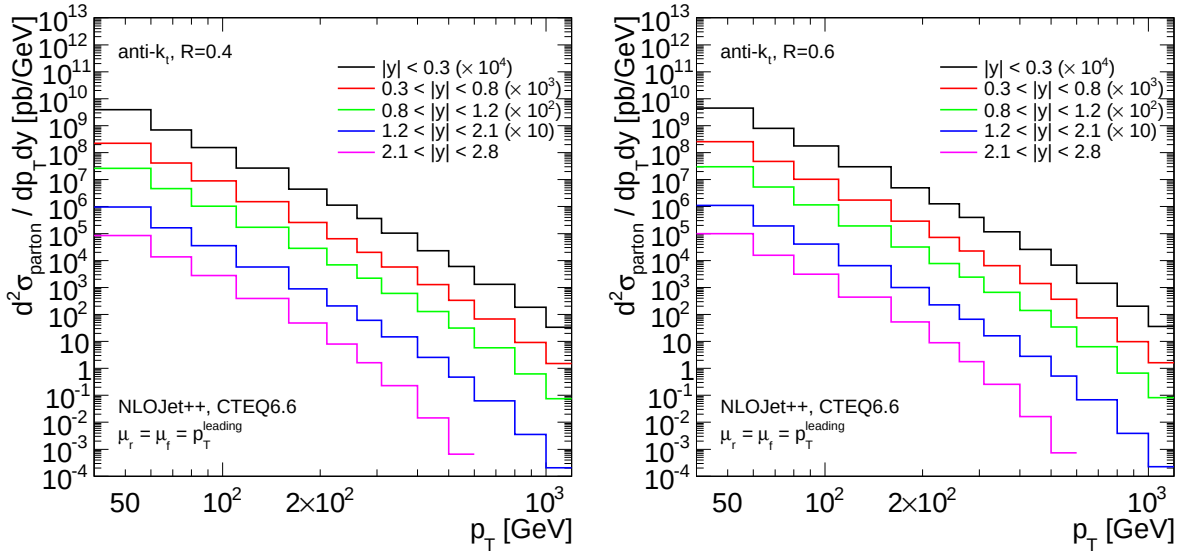

Figure 4.1: *Theoretical predictions for the inclusive jet cross sections at parton level. As indicated, the histograms were multiplied by increasing powers of ten for illustrative reasons.*

All three uncertainties are studied using the `APPLGRID` program [36]. The use of `APPLGRID` allows to store the perturbative coefficients in three-dimensional histograms, so called grids, in bins of x_1, x_2 and Q^2. Here x_1 and x_2 are the two Bjorken-x variables of the incoming partons and $Q^2 = (p_T^{\mathrm{lead}})^2$ is the squared momentum scale of the event. For the concrete application, these perturbative coefficients are the weights, which, multiplied by the parton densities, give the cross section for a given event. The convolution with the PDF, as well as the choice of α_s and scaling of μ_r and μ_f by constant factors can be done subsequently and independently of the program used to fill the grids.

In the following the uncertainties are derived using `APPLGRID` to calculate the cross sections with a certain parameter altered. The deviations observed from the nominal cross section will then be considered as the systematic uncertainty associated to the parameter in question.

The scale uncertainties are estimated by increasing and decreasing both scales independently by a factor of two. The ratios of the resulting cross sections to the nominal one are displayed in Fig. 4.2. In each bin of p_T the maximum deviation from the nominal cross section is considered the relative uncertainty. This is done separately for deviations to higher and lower cross sections, resulting in an asymmetric uncertainty. It amounts to approximately 5 % at low p_T, increasing to $\approx$ 20 % at 2 TeV. It is interesting to note that the scale uncertainties turn out to differ quite significantly between the two jet algorithms. This is in particular the case for the lowered renormalization scale, while the impact of constant μ_r and varying μ_f is very similar. The radius dependency of the μ_r uncertainty can be explained by the increased sensitivity of $R = 0.4$ jets to the NLO contributions. This is due to the fact that for smaller jets more three-parton topologies will be clustered to three jets, while they are clustered into two jets for $R = 0.6$.

Similarly the uncertainties due to the choice of α_s were estimated. The nominal CTEQ6.6 PDF uses $\alpha_s(M_Z) = 0.118$. Alternative PDF sets are provided, which were fitted using $\alpha_s(M_Z) = 0.116$ and $\alpha_s(M_Z) = 0.120$. These are recommended by the CTEQ collaboration for an uncertainty estimate at 95 % confidence level [37]. One should note that a change between the given PDF sets not only changes α_s, but also the PDF to a

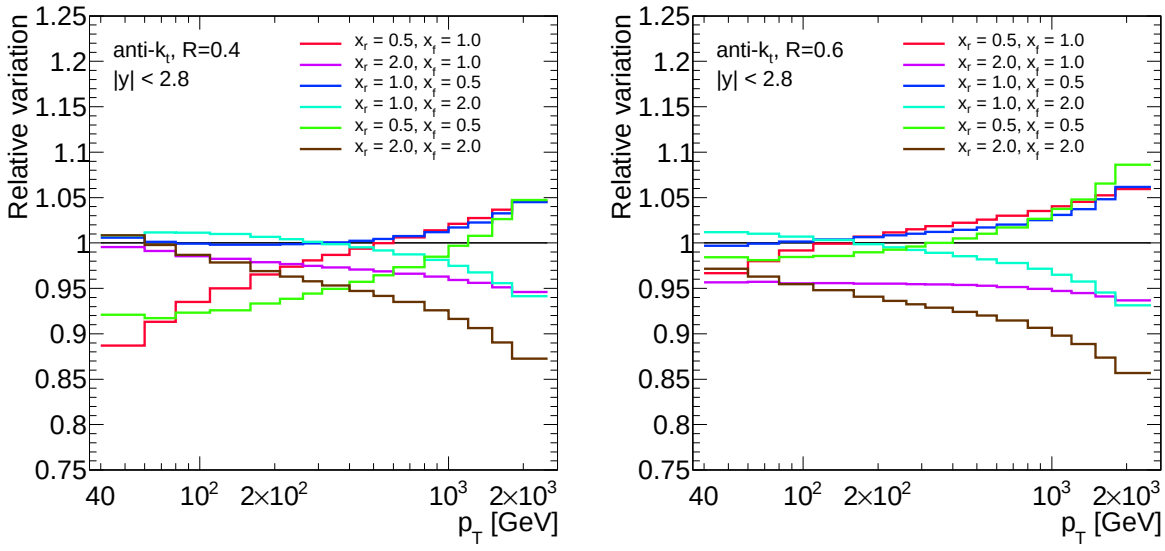

Figure 4.2: *Relative deviations of the cross sections induced by scaling μ_r (μ_f) by a factor x_r (x_f) as indicated. Separately for positive and negative deviations from unity, the maximum deviation in each p_T bin is taken as the systematic uncertainty.*

certain degree. The variations found are shown in Fig. 4.3 a). The uncertainties, again taken asymmetrically as the deviation from unity, are on the order of 3 % to 5 %. As expected, these uncertainties are on the same order as the actual uncertainty on α_s^2, being approximately 3.4 %, which is the order with which the dominant contribution to the jet cross section comes.

The PDF uncertainties are evaluated according to the recommendations of the CTEQ group [38] for the evaluation of asymmetric uncertainties. The provided PDF set includes 22 PDF eigenvectors with their positive and negative variation. The cross section predictions are calculated for each of these 44 PDF variations. Positive and negative deviations arising from each eigenvector are summed quadratically. The result of this procedure and the relative cross sections using the single eigenvectors are shown in Fig. 4.3 b). The uncertainty obtained ranges from $\approx 3\,\%$ to $^{+32\,\%}_{-22\,\%}$ at very high tranverse momenta. Due to the complex transformation of the PDF fit parameters to the eigenvectors, a physical interpretation of a single eigenvector is unfortunately not possible. For a fixed factorization scale, the gluon PDF uncertainty starts to increase significantly at lower x values than the uncertainties of u and d quarks. One can thus expect the gluon PDF to contribute significantly to the uncertainty at high transverse momenta.

Finally all three uncertainty components are combined in Fig. 4.4, by adding all contributions quadratically. The resulting total uncertainty is dominated by the PDF uncertainty for $p_T \gtrsim 300$ GeV and lies between 4 % and 35 %. Except for the observed differences in the impact of the renormalization scale, the uncertainties for both jet sizes are very similar.

4.3 Corrections to particle level

The predictions in Fig. 4.1 and the above described associated uncertainties apply to the parton level. In particular they neither include hadronization effects nor the underlying

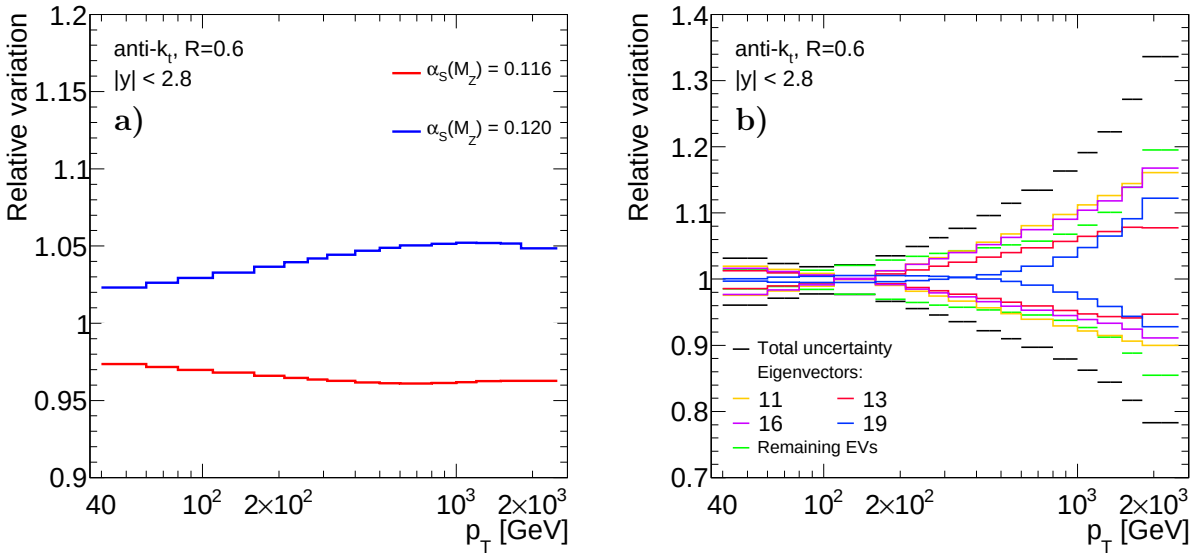

Figure 4.3: **a)** *Relative variations induced by the usage of PDFs extracted with the indicated α_s values. The deviations to higher and lower cross section are taken as the asymmetric α_s uncertainty.* **b)** *Cross section variation induced by PDF uncertainties. Only the four eigenvectors (EVs) with the highest contributions are shown explicitly. Uncertainties from remaining eigenvectors are summarized.*

event (UE) inevitably present in hadron-hadron interactions. Both have to be included in the theoretical prediction in order to allow for a meaningful comparison to data, which will not be corrected for either of the two effects.

As both effects originate in the non-perturbative regime of QCD there is no equally well founded basis for their computation as for the partonic cross sections*. The method commonly used to incorporate both effects is to apply a bin-by-bin correction factor to the parton level prediction. With this non-perturbative correction factor, denoted by K, the final theoretical prediction of the cross sections can be written as:

$$\frac{d^2\sigma(p_T,y)}{dp_T dy} = \frac{d^2\sigma_{parton}(p_T,y)}{dp_T dy} \times K(p_T,y), \tag{4.2}$$

where the partonic cross section σ_{parton} is taken from Fig. 4.1. These correction factors are extracted from MC simulations, which allow for event generation with and without the simulation of hadronization and underlying event. K is then derived as the ratio of the cross section including UE or hadronization (or both) and the cross section at parton level. Using *HAD* as a short hand for hadronization, this can symbolically be written as:

$$\begin{aligned} K(p_T,y) &= \frac{\sigma''(p_T,y;\text{with UE, with HAD})}{\sigma''(p_T,y;\text{no UE, no HAD})} \\ K_{UE}(p_T,y) &= \frac{\sigma''(p_T,y;\text{with UE, no HAD})}{\sigma''(p_T,y;\text{no UE, no HAD})} \\ K_{HAD}(p_T,y) &= \frac{\sigma''(p_T,y;\text{no UE, with HAD})}{\sigma''(p_T,y;\text{no UE, no HAD})}. \end{aligned} \tag{4.3}$$

*An exception from this statement could be the recently developed techniques in [39], which allow for an analytical computation of hadronization effects using only few experimental input.

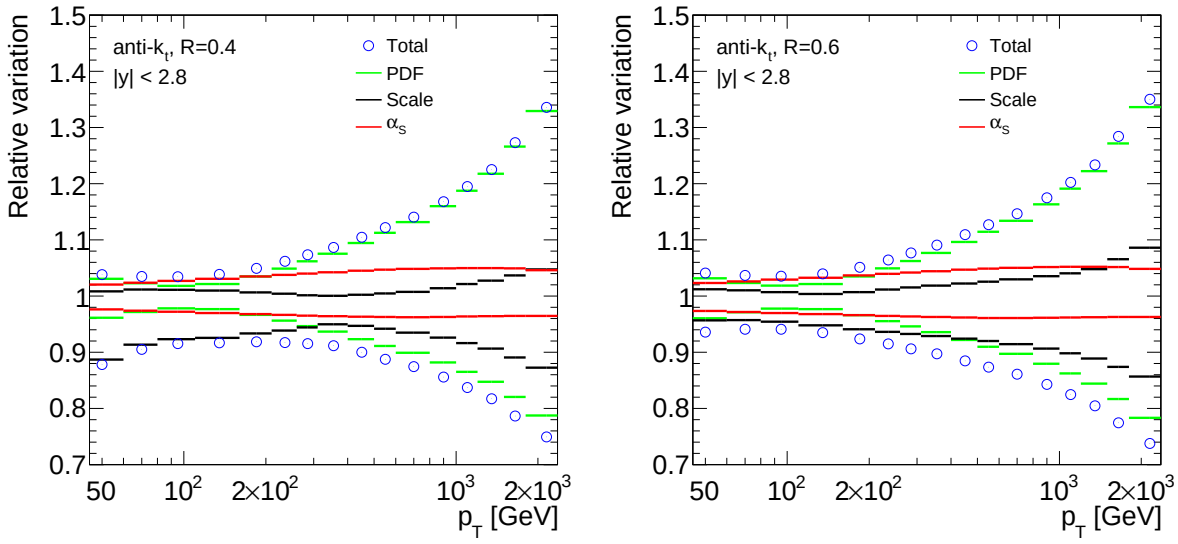

Figure 4.4: *Single contributions to the theoretical uncertainty on the partonic cross section and the resulting total uncertainty for both jet sizes.*

Here $\sigma''(p_T, y) = \dfrac{d^2\sigma(p_T, y)}{dp_T dy}$ is used for brevity. It should be noted that this procedure is theoretically not perfectly well founded. A major caveat is that UE and hadronization in a MC generator are introduced on top of an intermediate state, which would have to match the parton final state within `NLOJet++`. This is however not the case, due to the fact that the NLO contributions are not fully contained in common LO event generators. Instead they are only modeled to a certain degree by the use of parton showers. These parton showers also produce a rather large amount of partons in the final state, compared to up to only three partons in `NLOJet++`. The ideal situation would be to interface the full NLO calculation properly to an event generator handling the remaining steps. This possibility is for example provided by the `MC@NLO` project [40] for various processes, however not including inclusive jet production. Only very recently inclusive jet production has been implemented in `POWHEG` [41].

Due to this situation the bin-by-bin correction is kept. To account for the caveats mentioned above the correction is derived in a variety of MC generator tunes and two different generators, in order to gain a realistic estimate of the associated uncertainty. The MC generators used for this purpose are `Herwig++` [7] using a recent tune to ATLAS UE measurements available at [42] and `Pythia6` using a selection of the so called Perugia tunes [43] and the ATLAS tune MC09 [44].

In principle the simulation of K would be sufficient in order to correct the parton level predictions to the particle level. It is however instructive to first study the impact of UE and hadronization seperately. Subsequently K and its associated systematic uncertainty will be derived.

The results for K_{UE} and K_{HAD} are collected in Fig. 4.5 for both jet sizes. In general the hadronization's effect is to move energy from inside the jet outside of its scope, thus lowering the transverse momentum and the observed cross section at fixed p_T. The size of this effect thus decreases as the jet gets wider, which is confirmed by the results in Fig. 4.5. The impact of the UE can basically be described by an average energy density added to the event on top of the hard parton scattering. Depending on the jet size, portions of

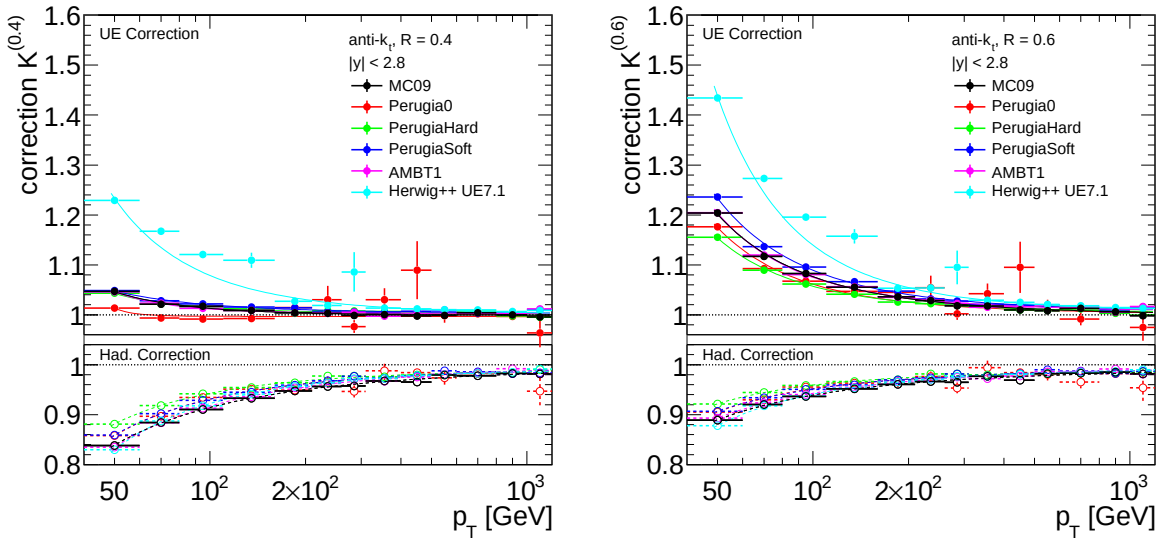

Figure 4.5: *UE and hadronization corrections for various* `Pythia6` *and* `Herwig++` *tunes for anti-k_t6 (left) and anti-k_t4 jets (right). The lower sections shows the hadronization corrections, the upper ones contain the UE corrections. The MC predictions are fitted using Eq.* 4.4.

this background are included in the jets of the hard scattering. One can thus conclude that the impact of the UE on a single jet's p_T increases with its area. This corresponds to an increase of the UE impact by a factor of 2.25 going from $R = 0.4$ to $R = 0.6$ jets, taking into account that anti-k$_t$ jet areas are to a very good approximation circular [20]. The impact on the cross section and thus on the non-perturbative correction is however amplified by the very steeply falling cross section. Consequently, in particular at low p_T K_{UE} is multiple times larger for $R = 0.6$ jets than for the smaller jet radius.

The differences in the UE correction between the `Pythia6` tunes and `Herwig++` are strikingly large. Yet at the same time both generators manage to describe ATLAS data to a certain degree: For `Herwig++` this is demonstrated at [42], while comparisons with various `Pythia6` tunes are included in [30]. The latter measurement however implies that the `Pythia6` tunes used can be considered to include too few UE activity in general, which slightly favours the `Herwig++` prediction. The spread in the hadronization corrections is in turn much smaller. Nevertheless the `Herwig++` result shows the largest hadronization effect. In the physically more sensible scenario with UE and hadronization, the two rather large `Herwig++` corrections can be expected to cancel each other to a certain degree.

In order to take eventual correlations between the UE and hadronization properly into account, K is evaluated by switching on UE and hadronization in parallel, as indicated in Eq. 4.2. The corrections K obtained this way are displayed in Fig. 4.6. It can be noted that K decreases the partonic cross section for small jets while it significantly enhances the one for large jets. The major reason for this is the large impact of the UE. The large spread between `Pythia6` and `Herwig++` in this final correction is only slightly smaller. The inclusion of `Herwig++` increases the uncertainty significantly with respect to the set of corrections obtained with `Pythia6`. This is considered to provide a much more realistic estimate of the uncertainty. In particular `Herwig++` can be expected to cover uncertainties which are not covered by the various `Pythia6` tunes, such as eventual features due to the different models used in both generators. An exemplary, common feature of all `Pythia6` tunes is the asymptotical behaviour for very high transverse momenta, which differs only

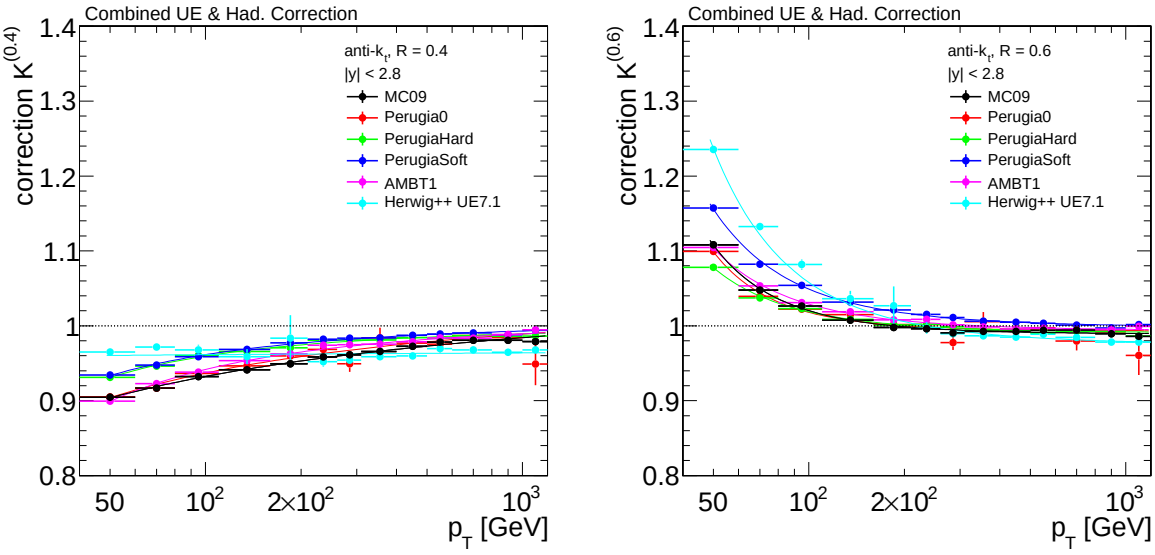

Figure 4.6: *Non perturbative correction for various* `Pythia6` *and* `Herwig++` *tunes for anti-k_t6 (left) and anti-k_t4 jets (right). The MC predictions are fitted using Eq.* 4.4.

very little among the various tunes. Here `Herwig++` predicts a different behaviour, which obviously is not within the scope of the `Pythia6` tunes.

To finally obtain smooth non-perturbative corrections, each correction is fitted using a logarithmic power law,

$$f(p_T) = A + B \log(p_T/\text{ GeV})^{-C}, \tag{4.4}$$

with A, B and C being free parameters. This parametrization fits all observed corrections well. The predictions obtained using the `Pythia6` MC09 tune is used as the nominal correction. The distance towards the maximum (minimum) fitted correction in each p_T bin is taken as the upper (lower) systematic uncertainty.

4.4 Theoretical predictions at particle level

The results of the two previous chapters complete the theoretical predictions and provide the baseline comparison to the results of the analyses presented later on. Figure 4.7 displays the calculated cross sections at particle level along with their systematic uncertainties. The total uncertainty obtained is dominated by the uncertainty on the parton level prediction over a large range of transverse momentum. Only at low p_T the non-perturbative correction poses a significant contribution, which is mainly due to the `Herwig++` results.

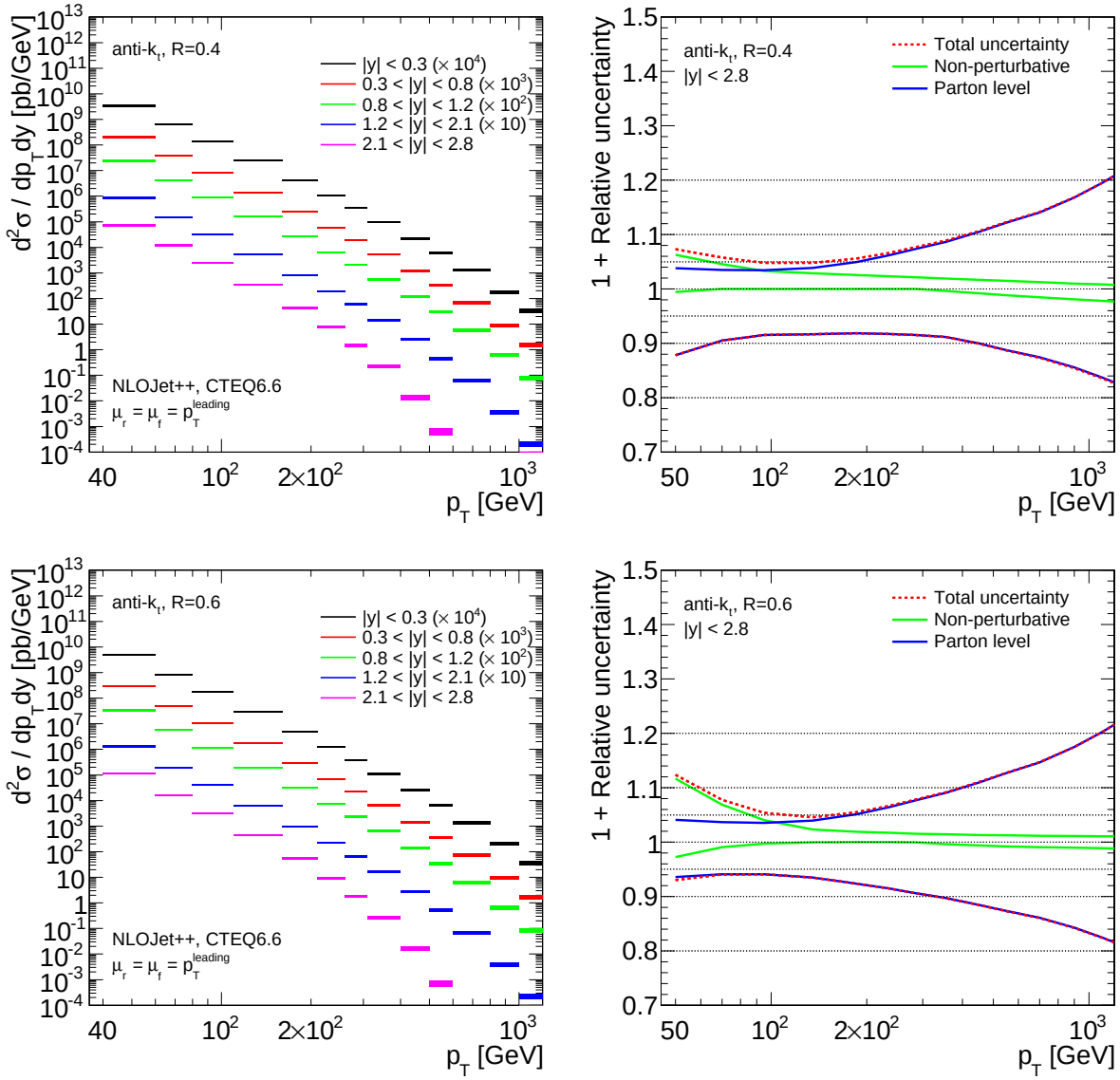

Figure 4.7: *Theoretical predictions for the inclusive jet cross section (left) and the associated systematic uncertainties (right). The top row shows the results for $R = 0.4$ jets, the bottom row the ones for $R = 0.6$ jets.*

5 The ATLAS experiment at the LHC

The Large Hadron Collider (LHC) was built to gain access to unprecedented energies and luminosities in hadron-hadron collisions. It was built in the tunnel that previously housed the Large Electron Positron collider near Geneva/Switzerland on the grounds of the European Organization for Nuclear Research (CERN). It was designed to primarily provide proton-proton (pp) collisions at a maximum center of mass energy of 14 TeV and instantaneous luminosities of up to $10^{34}\,\mathrm{s}^{-1}\mathrm{cm}^{-2}$. The instantaneous luminosity quantifies how many proton-proton pairs per unit area and time pass through the interaction region. It allows to determine the rate R_p at which of processes with a given cross section σ_p occur:

$$R = L \times \sigma \tag{5.1}$$

The LHC was built as a synchrotron collider, using superconducting magnets. Due to the synchrotron nature, the proton beams are not continuous but consist of bunches of protons that are brought to collision in the interaction points of the experiments. Single proton bunches contain $\mathcal{O}(10^{11})$ protons and have an approximate transverse spread of $\mathcal{O}(10\mu\mathrm{m})$ and a longitudinal one of $\mathcal{O}(1\,\mathrm{cm})$. In each bunch crossing one or more pairs of protons from both bunches may collide, where the probability to observe a certain number of pp-collisions follows a Poissonian distribution. These multiple pp-collisions are referred to as pile-up and the mean of the Poissonian is generally denoted by μ. The value of μ is mainly determined by the number of protons in the bunches and their collimation in the interaction regions. For further technical details on the accelerator the reader is referred to [45].

Experiments taking data at the LHC are ALICE [46], LHCb [47], LHCf [48] and TOTEM [49], whose detectors are built for very specific physics purposes. Additionally, there are two general purpose experiments: The CMS [50] and the ATLAS experiment [51]. The ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) detector is fully operational since Summer 2008 and taking collision data since November 2009. The declared aim of the ATLAS experiment is the exploration of physics in the new energy regime accessible at the LHC. The general design principle of the detector, shown in Fig. 5.1 thus was universality and a as large coverage as possible. Due to the very universal physics goal, the detector comprises a large variety of different subdetectors. Similarly to previously built particle detectors at colliders, the detector consists of several layers. From the inside out, these can be grouped into the inner detector, which is contained in a solenoidal magnetic field and contains the tracking system. Surrounding the tracking system are the electromagnetic and hadronic calorimeters, measuring the particle's energies. The outer envelope of the detector is formed by the muon system, which is immersed in a toroidal magnetic field.

The design parameters of the specific detector sub-components followed the physics expectations as well as the general environment at the LHC. The performance of the calorimetry system was designed to ensure the best possible resolution in view of important signatures, such as Higgs bosons decaying into two photons. The tracking system has to be highly granular to be able to cope with the very high occupancy caused by numerous

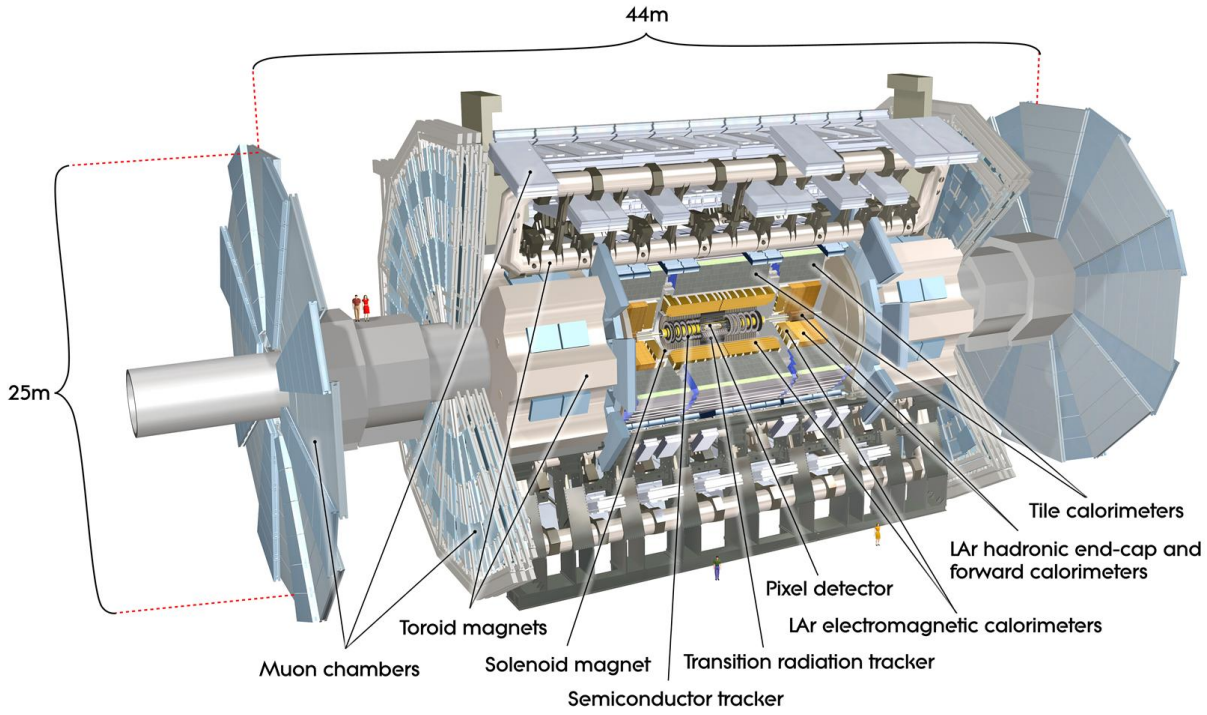

Figure 5.1: *View of the ATLAS detector.*

proton-proton-collisions in one bunch crossing. The detector read-out and the trigger system have to cope with the interaction rates of 40 MHz. The very high interaction rates also bring new challenges for the data acquisition, distribution and analysis.

In the following all the systems that participate in the data taking are outlined. The focus will be on subdetectors and features that will be especially important for the upcoming measurement, such as the calorimetry and calorimeter trigger.

5.1 The ATLAS coordinate system

The coordinate system used within ATLAS is based upon a right-handed, Cartesian coordinate system, whose z-direction is oriented counter-clockwise along the beam axis. The positive x-axis points towards the center of the LHC ring and the y-axis points upwards. The origin of the coordinate system is defined as the nominal interaction point. Further, frequently used and occasionally more convenient coordinates are listed below:

- Radial distance to the beam axis: $r = \sqrt{x^2 + y^2}$.
- Azimuthal angle: $\phi = \arctan(x/y)$.
- Polar angle: $\theta = \operatorname{arccot}(z/r)$.
- Pseudorapidity: $\eta = -\ln(\tan\theta/2)$.
- Rapidity: $y = \frac{1}{2} \ln\left[\frac{E + p_z}{E - p_z}\right]$.

- Distance in the η-ϕ-plane: $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.
- Quantities in the transverse plane indicated by the T subscript, such as p_T, E_T etc., are the quantity's projection onto the x-y-plane.

5.2 The inner detector

The inner part of the detector tracks particles bent within the solenoidal magnetic field of 2 T and thus can measure their momentum. The momentum resolution of the inner detector was designed to be $\sigma_{p_T}/p_T = 0.05\,\% \, p_T/1\ \text{GeV} \oplus 1\,\%$, within a coverage of up to $|\eta| < 2.5$. This is done using a combination of pixel and silicon microstrip (SCT) detectors, surrounded by the transition radiation tracker (TRT) as laid out in Fig. 5.2. Closest to the beam-pipe, at a radial distance from the beam axis of approximately 45 mm,

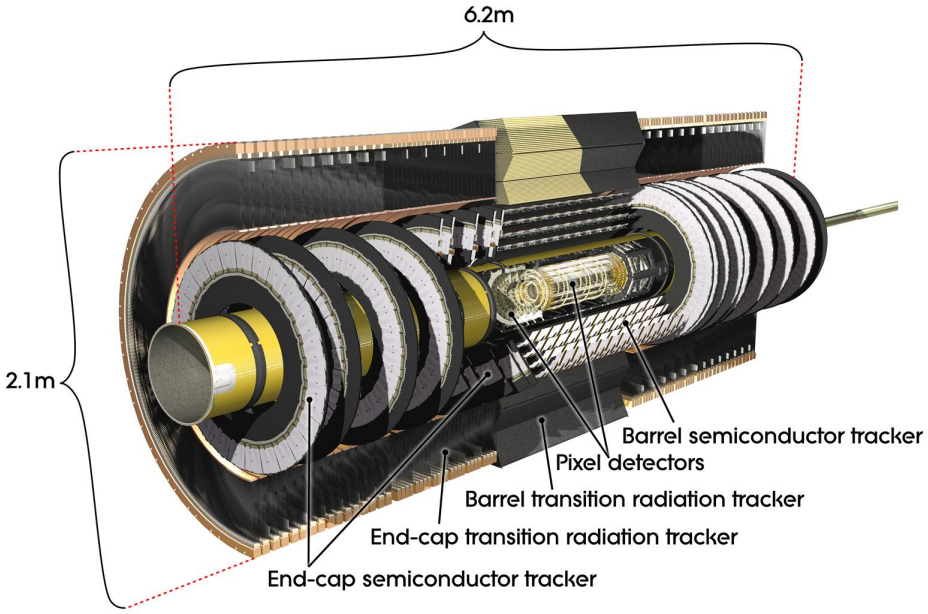

Figure 5.2: *Sectional view of the ATLAS inner detector and its sub-components.*

where the occupancy is highest, the tracking is done using silicon pixel detectors with a minimum pixel size in $r\phi \times z$ of $50{\times}400\ \mu\text{m}^2$. The measurement of hits in these pixels is possible with a resolution of 10 μm in the r-ϕ plane and 115 μm in the z direction. The pixel detectors are placed in three layers on concentric cylinders around the beam-pipe and three disks perpendicular to the beam-pipe up- and downstream of the nominal interaction point. Adjacent to the pixel detector are four SCT layers which span the radial distance of up to 50 cm away from the beam axis. Each SCT layer consists of two sub-layers in which the strips are arranged with a stereo angle of 40 mrad to be able to also measure the z-position. The last inner detector system for a particle to pass is the TRT, which consists of straws arranged in 73 planes in the barrel (cf. Fig. 5.2), parallel to the beam axis (160 in the end-cap in radial direction). The TRT only measures a hit's r-ϕ coordinate, with a precision of 130 μm.

The read-out of the inner detector has to cope with information out of approximately 87 million electronic channels, of which 80 million are contributed by the pixels, 6 million by the SCT and approximately 350000 by the TRT. The combination of precision

measurements by the semiconductor detectors and the TRT coordinates at rather large radii results in a very robust pattern recognition and high precision. Besides the provision of additional spacepoints to the track reconstruction the transition-radiation induced by electrons in the TRT enhances the electron identification capabilities.

5.2.1 Reconstruction of charged particles and vertices

The details of the ATLAS track reconstruction are outlined in detail in [52]. Here only, very briefly, the general scheme shall be described.

The very first step of the track reconstruction consists of converting the hit information of the pixel and SCT detectors into three dimensional representations, the so called *space-points*. Pairs of space-points of the pixel detectors are used to find track seeds. These seeds already provide a crude estimate of the direction of the track. Hits on subsequent detector layers in the proximity of this direction are successively added to the track fit. The reconstruction of tracks within the silicon detectors is followed by the extension to the TRT.

For the reconstruction of vertices, first all vertex candidates and their associated tracks are found. In a second step the vertex is fitted using the uncertainty information of all associated tracks. In this step, also all tracks are refitted, implementing the constraint that they originate from the same interaction point.

The primary vertex is an event is defined as the one associated with the hardest scattering, while additional vertices in the event are considered as pile-up vertices. The identification of the primary vertex in an event is based on the sum of squares of the transverse momenta of all associated tracks. The vertex that maximizes this sum is considered the primary vertex in the event.

5.3 The calorimeter system

The ATLAS calorimeters were designed to be able to cope with the large variety possible new physics might bring. The ATLAS calorimetry, shown in Fig. 5.3, consisting of electromagnetic (em.) and hadronic calorimeters, comprises two basic detector concepts: The em. calorimetry, using LAr-lead sampling calorimeters and the hadronic calorimeter, which is realized as iron-scintillator detectors and LAr-copper sampling calorimeters. The criteria that the calorimeters had to match, are as follows:

- Coverage up to very large pseudorapidities is needed to be able get an as complete picture as possible of a pp-collision. This feature is especially important for the resolution of the missing transverse energy, which is an important event feature in many models of physics beyond the standard model. The ATLAS calorimeters reach pseudorapidities of up to $|\eta| \approx 4.9$, which corresponds to an angle with respect to the beam axis of 14 mrad.

- Energy resolution is a major benchmark of all calorimeters. The energy resolution of the em. calorimeter is expected to allow for a measurement of the Higgs boson's mass with a resolution of 1 %, e.g. in the decay into two photons. The required, fractional energy resolution of the calorimeters for this purpose has to be better than $10\,\%/\sqrt{E/1\text{ GeV}} \oplus 1\,\%$. For the hadronic calorimeters similar requirements apply based on the prospects for measurements of the top quark mass or Higgs

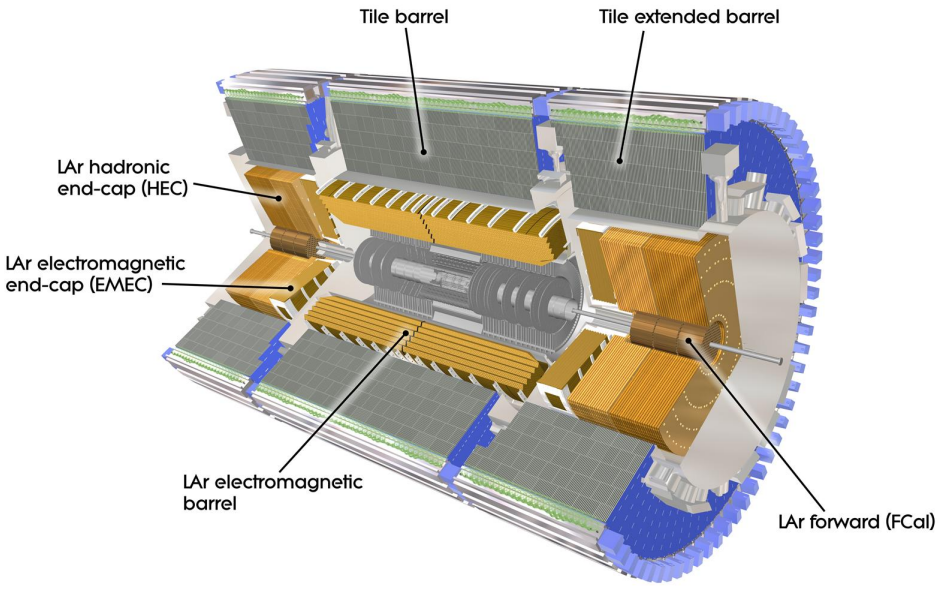

Figure 5.3: *Left: Sectional view of the ATLAS calorimeters. Figure taken from Ref. [51].*

decays which may include hadronically decaying bosons. The targeted resolutions for jet energy reconstruction with the combination of em. and hadronic calorimeters is $50\,\%/\sqrt{E/1\text{ GeV}} \oplus 3\,\%$.

- Complementary to a very good energy resolution, the linearity of the absolute energy scale for the em. calorimeters is expected to be below 0.5 %, while for the jet reconstruction a scale uncertainty of 2 % is the minimum aim.

- Angular resolution poses a contribution to the resolution of invariant masses of two (or more) particles. To keep this contribution small, a high resolution in the measurement of angular positions requires highly granular segmentations along η and ϕ.

The electromagnetic calorimeters are built as sampling calorimeters with liquid Argon (LAr) as active medium and lead as the absorber. The hadronic calorimeters in the barrel ($|\eta| < 1.0$) and extended barrel ($|\eta| < 1.7$), the *tile* calorimeters, consist of iron absorber plates interleaved with plastic scintillator tiles. In the end-caps the hadronic calorimeters are realized as LAr-copper sampling calorimeters.

5.3.1 Geometry of the electromagnetic calorimeters

The barrel LAr calorimeters cover pseudorapidities up to $|\eta| < 1.475$ and consist of two half-barrels joined at $\eta = 0$. They are built in an accordion geometry as sketched in Fig. 5.4, which has the benefit of enabling full azimuthal symmetry and a fast read-out of the signal at the rear or at the front of the electrodes. Longitudinally the calorimeter in the barrel is segmented threefold, with the middle layer (*layer 2*) generally absorbing the majority of a particle's energy. The first layer is very finely segmented in η, allowing for very good spatial resolution.

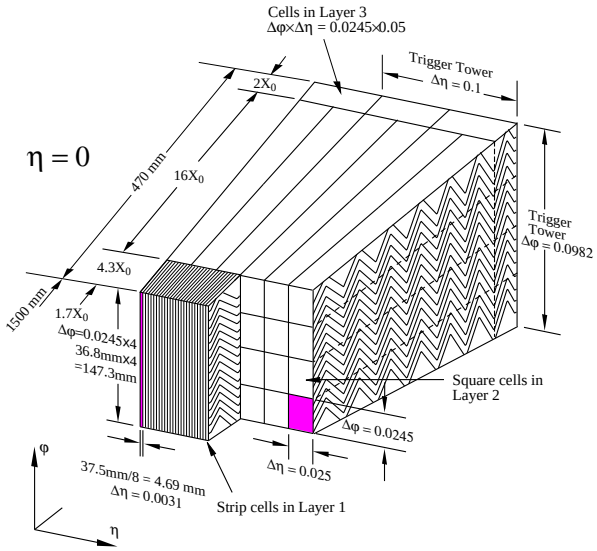

Figure 5.4: *Schematic of the LAr accordion architecture in the barrel. Incident particles in this view enter the calorimeter from the lower left corner. Figure taken from Ref. [51].*

The end-cap calorimeters consist of two wheels, the outer of which covering $1.375 < |\eta| < 2.5$ and the inner extending the coverage up to 3.2. The accordion structure is continued in the end-caps, were the waves are parallel to the radial direction and run axially. Also the longitudinal segmentation from the barrel region is propagated, and the granularity in the middle layer is 0.025×0.025 as in the barrel. As indicated in Fig. 5.4 four times four of these highly granular calorimeter cells can be grouped into so called *trigger towers*, which are projective sums of all cells in a certain η-ϕ direction. These towers are used in the trigger system to allow for a very fast, but rather coarse read-out of the calorimeters. The thickness of the em. calorimetry is approximately 23 radiation lengths (X_0) in the barrel and well above $25\,X_0$ for $|\eta| > 1.5$, ensuring a proper containment of em. showers up to highest energies.

Non-uniformities in ϕ direction thanks to the accordion geometry do not exist. As a function of pseudorapidity the transition from the barrel to the end-cap around $|\eta| \approx 1.45$ exhibits an unavoidable interrupt in the calorimeter's uniformity. Within the barrel region, much smaller non-uniformities exists around $\eta = 0$, where the two half-barrels are joined and at $|\eta| \approx 0.8$, where the absorber thickness within the calorimeter changes. The total em. calorimetry comprises approximately 160000 calorimeter cells which are read out.

5.3.2 Geometry of the hadronic calorimeters

The tile calorimeters

The tile calorimeters cover pseudorapidities of up to $|\eta| < 1.7$, with a division at $|\eta| \approx 0.8$ into the tile barrel calorimeter and the extended tile ($|\eta| > 0.8$). As the em. calorimetry, it is longitudinally segmented into three parts, with the middle one being the thickest. The geometry of one detector slice in ϕ is shown in Fig. 5.5. The scintillating tiles are not projective in η, but lie within the r-ϕ plane. Particles at central rapidity thus travel parallel to the tiles and with an increasing angle for increasing pseudorapidities. The read-out happens through wavelength shifting fibers which run along the thin edges of the tiles, guiding the scintillation light to photomultipliers and the read-out electronics at the outer

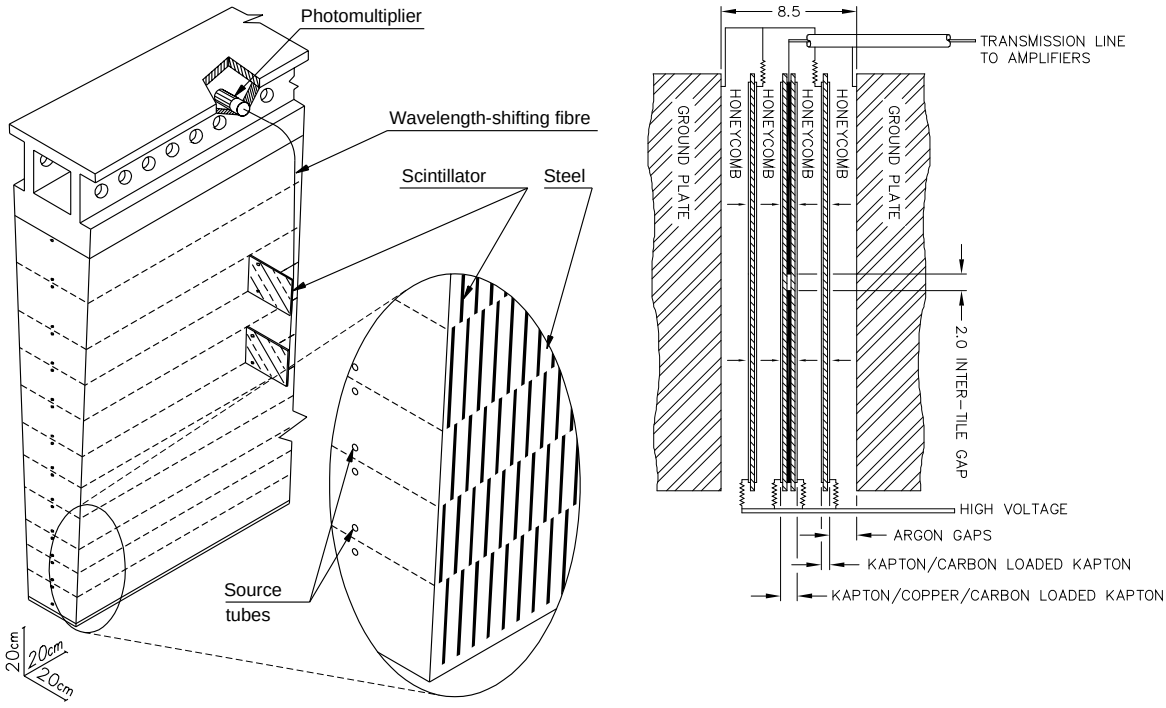

Figure 5.5: *Left: Schematic of the geometry of the tile calorimeter modules. Right: Schematic of the calorimeter architecture in the hadronic end-cap. Figure taken from Ref. [51].*

end. By properly grouping signals from individual tiles as input to one photomultiplier, an approximately projective read-out is achieved.

Holes were drilled into each azimuthal segment, parallel to the beam axis. Along these channels a ^{137}Cs source may be injected into the system, whose signal can be used for calibration purposes. The number of interaction lengths particles traverse until the end of the tile calorimeters is approximately 10, except for the transition region around $|\eta| \approx 0.8$. In total, the tile calorimeters contribute approximately 10000 read-out channels to the total calorimeter read-out.

The hadronic end-cap

In the hadronic end-cap (HEC) the calorimetry uses a completely different architecture than in the barrel. Here the calorimeter uses copper as absorber and LAr as active medium and thus is also contained within the LAr cryostat in which the em. end-cap is placed. The full end-cap is constructed as two wheels, one behind the other, both covering $1.5 < |\eta| < 3.2$.

The absorbers are arranged as flat plates in the r-ϕ plane, separated in z by four honeycomb structures which are immersed in the liquid argon (cf. Fig. 5.5). The total gap between two absorber plates is subdivided by three electrodes. Distinct calorimeter cells are defined by pads, which are etched on the central electrode foil between two absorbers. The geometry of these pads also makes it possible to define the read-out cells to be projective. The number of interaction lengths up to the end of the hadronic end-cap is approximately 12. Approximately 6000 electronic channels are used to read-out the hadronic end-cap calorimeters.

Instrumentation in the barrel to end-cap transition

The transition region between the barrel and end-cap calorimetry is occupied with service structures and power supplies for the inner detector and the LAr calorimeters. To nevertheless be able to detect energy depositions of particles traversing this region, special detectors were inserted as depicted in Fig. 5.6. In the radially outermost of the detector

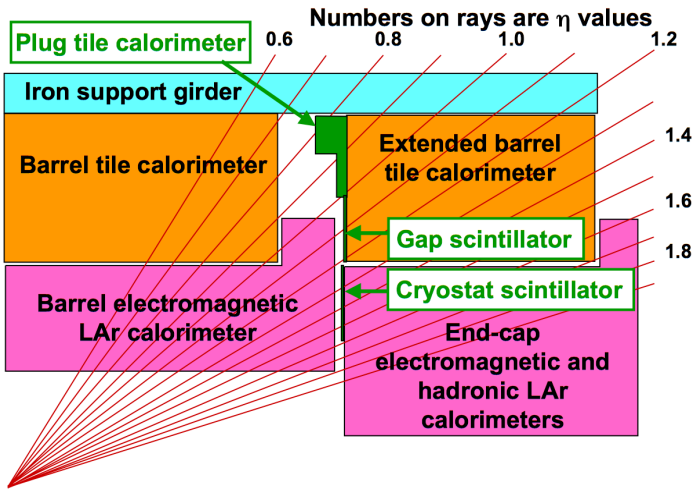

Figure 5.6: *Schematic view in the r-z-plane of the calorimeters in the barrel to end-cap transition region. Figure taken from Ref. [51].*

the tile *plug*-calorimeter, a tile calorimeter section reduced in size, provides additional coverage. The so called *gap scintillators* were attached to the tile calorimeter surface, covering the region of $1 < |\eta| < 1.2$. The more forward region of $1.2 < |\eta| < 1.6$ is covered by the cryostat scintillators, which are attached to the tile calorimeters and read out through their electronics.

The energy deposited in these scintillators allows to estimate the amount of energy lost in the not instrumented regions in front of the scintillators.

5.3.3 Forward calorimetry

Calorimetry beyond $|\eta| > 3.2$ is provided by the forward calorimeter (FCAL), which is also located in the cryostats of the end-cap calorimeters and covers the pseudorapidity of $3.1 < |\eta| < 4.9$. The close vicinity to the beam-pipe exposes the FCAL to exceptionally high radiation, which lead to a design with very small liquid-argon gaps. The FCAL is longitudinally segmented into three parts, the first of which being optimized for em. calorimetry. This first segment is made of stacked copper plates, in which holes were drilled parallel to the beam axis. Copper rods with a slightly smaller diameter are centered in these tubes, leaving a thin gap which is filled with LAr.

The two hadronic segments use the same architecture as the em. segment. But, in order to achieve as many interactions lengths as possible, as much copper as possible was replaced by tungsten: The copper rods where replaced by rods made of tungsten and the hadronic segments are only held together by two copper end-plates which embrace a matrix of tungsten slugs forming the necessary tubes.

5.3.4 Calorimeter signal reconstruction and calibration

The determination of a particle's energy based on the calorimeter response is a non-trivial task. For all LAr based calorimeters, the raw detector response is a current pulse, whose integral is proportional to the energy the particle deposited in the calorimeter. This analogue signal is shaped and digitized, resulting in a signal of ADC (analogue-to-digital converter) counts. The full pulse time in the LAr is approximately 400 ns. To a avoid a long integration time, this pulse is filtered to a bipolar shape. This allows for a faster readout by sampling only five values of the pulse in intervals of one bunch crossing (25 ns). The bipolar shaping is optimized to minimize the sum of electronic and pile-up noise arising from signals from the following bunch crossings.

In the very first calibration step, the conversion from ADC counts to μA needs to be obtained. This conversion is determined using charge injection systems mounted on the calorimeter front-end electronics. These are capable of injecting a precisely known charge into the system and thus allow to extract the conversion constant. The calibration from μA to eV is obtained from test-beam measurements, where the detector is exposed to particles of very well known energy.

The calibration of the tile calorimeter involves one more step, since the light yield within a scintillator is converted by the photomultipliers to an analogue pulse. For the calibration and monitoring of the photomultipliers a laser system, whose light is guided via clear plastic fibers to the photo-cathodes is used. Additionally it is possible to inject a ^{137}Cs source through channels in the tile calorimeters, probing the full chain from light collection via photomultipliers to the digitized current pulse.

To obtain the conversion from the digitized photomultiplier output in ADC counts to pC a charge injection system is used, similarly to the procedure in the LAr systems. The absolute calibration from pC to eV is obtained from test-beam measurements.

Results from electron and pion test-beams are shown in Fig. 5.7. In these test-beam

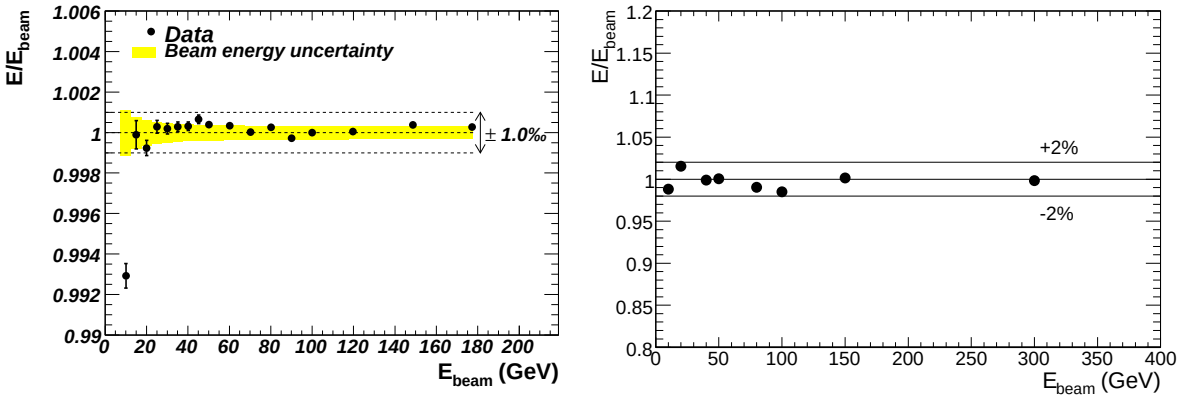

Figure 5.7: *Left: Linearity of the electron response, all points are normalized to the value at $E_{beam} = 100$ GeV. Right: Linearity of the measured pion beam energy for a combined LAr and tile calorimeter segment. Figure taken from Ref. [51].*

results the aims outlined in the beginning of this section are achieved to a satisfactory degree. The linearity obtained for electrons is on the level of 1 ‰, which is well below the aim of 0.5%. The response for pions in the combined LAr and tile calorimeter is constant within 2 %.

The fractional energy resolutions measured are $10\,\%/\sqrt{E/1\text{ GeV}} \oplus 0.4\,\%$ for electrons.

For charged pions the resolution obtained amounts to $52\,\%/\sqrt{E/1\ \mathrm{GeV}} \oplus 3\,\%$. These results meet the physical requirements well.

5.4 The muon system

Due to their mass, muons traverse the inner detector and calorimeter loosing only very few energy. The outer envelope of the ATLAS detector thus is formed by the muon tracking system. Once outside the calorimeters, muon's are bent in the toroidal magnetic field of 0.5 Tesla (1 Tesla) in the barrel (endcap) region. While tracks in the inner detector are bent in the r-ϕ-plane, the toroidal field leads to a bending in the r-z-plane.

The design of the muon system follows the performance goal of a momentum resolution of approximately 10 % for 1 TeV tracks.

The full muon system consists of two general detector types. Precision measurements are provided by monitored drift tubes (MDTs) in the central region, while cathode strip chambers (CSCs) are used in the end-cap region. The MDTs are approximately 3 cm thick tubes filled with a Ar/CO_2 mixture and work as classical drift tubes, in which ionized electrons drift to the wire which goes along the center of the tube. The CSCs are multiwire proportional chambers, which are arranged in wheels at each side of the interaction point. Each wheel is composed of 16, slightly overlapping wedge shaped segments.

For triggering purposes different detector types are used. These allow for a faster read-out at the cost of reduced precision.

The reader is referred to [51] for more details on the muon system.

5.5 The trigger system

Though the LHC may be able to provide bunch crossings at a rate of up to 40 MHz, events can only be recorded with a maximum rate of approximately 300 Hz. This reduction in rate has to be achieved with a very fast, but at the same time efficient pre-selection of events. This pre-selection is done by the trigger system.

At ATLAS the trigger is realized in three stages: the Level-1 (L1) trigger, the Level-2 (L2) trigger and the event filter (EF). The two latter are grouped into the High-Level trigger (HLT). Each of the three stages is able to analyze events increasingly more detailed, allowing for an increasingly precise selection using more sophisticated algorithms. The largest reduction in the trigger rate has to be achieved by L1. For this purpose the L1 system is built from custom hardware, enabling trigger decisions within 2.5 μs. The maximum L1 accept rate is 75 kHz. The output of L1 consists of information about interesting regions in the detector, such as regions containing high-p_T objects, which are forwarded to L2. The L2 trigger is implemented as software algorithms running on computer farms and can spend up to 40 ms for making the trigger decision. At L2 the event rate is reduced to approximately 3 kHz. Events passing L2 are forwarded to the EF, at which the full event information can be used. Here, the trigger decision may take up to four seconds per event.

If the rate reduction achieved with a certain trigger is insufficient but at the same time the trigger requirement cannot be changed, so called *prescale* factors are used. The introduction of a prescale factor p for a trigger means that only every p^{th} event fulfilling the trigger requirement, is recorded. This concept is especially important for triggers selecting low p_T events, whose cross section is too large to record all events fulfilling the

trigger requirement. In these cases the prescales allow for a selection of only a fraction of these events.

During the processing of an event within the various trigger stages, the detector data with full granularity is stored in pipelines within the Data Acquisition (DAQ) path (cf. Fig. 5.8). If the event is accepted by the trigger at all levels, this data is read out and finally written to disk.

5.5.1 The Level-1 trigger

The time constraint of making trigger decisions within 2.5 μs is the major design criterion for the L1 trigger system. This requirement can only be met using coarser detector information than available. As a consequence the L1 trigger only incorporates calorimeter signals and information from the muon trigger chambers, whereas no tracking information can be used. The general logic flow of the system is illustrated in Fig. 5.8. The system

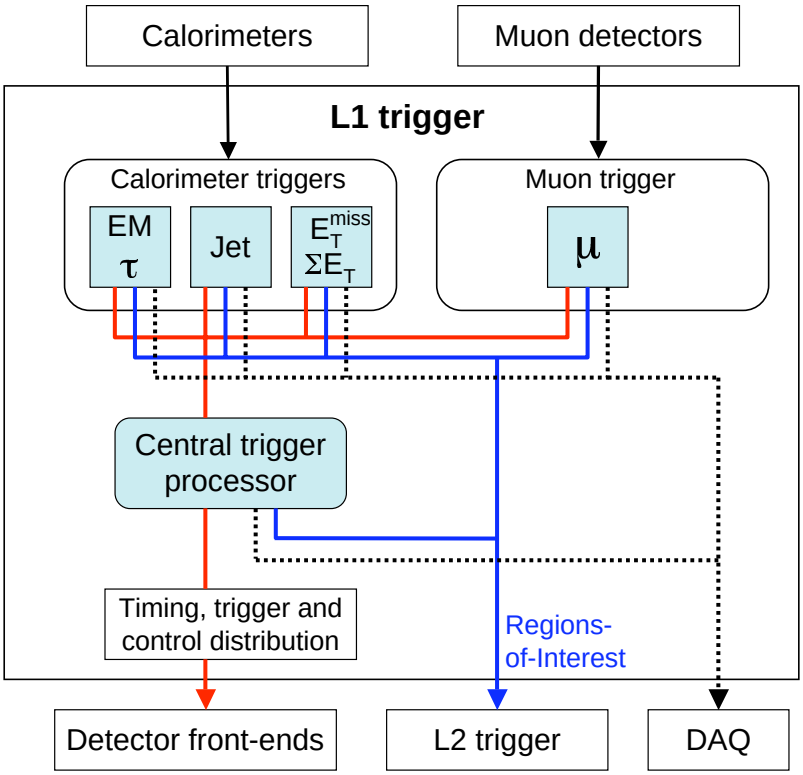

Figure 5.8: *Diagram of the data flow in the Level-1 trigger system. Data from the calorimeters and the muon trigger chambers is processed within the calorimeter and muon trigger. The results from both sub-systems are sent to the central trigger processor, which makes the final trigger decision for current event. In case of a positive trigger decision, the data of regions of interest are sent to the L2 trigger. Figure taken from Ref. [51].*

consists of two sub-systems: the calorimeter trigger and the muon trigger. Both systems count the number of physical objects, such as jets and electrons, passing certain, programmable transverse energy thresholds. The calorimeter trigger additionally calculates the global event features of summed transverse energy ($E_{T,Sum}$) and missing transverse energy ($E_{T,Miss}$), which can be used in the CTP. As the muon trigger will not be used in the analysis, the focus here will be on the calorimeter trigger.

The calorimeter trigger

The calorimeter trigger is of exceptional importance for the upcoming analysis, since it selects all the events used. To be able to make a trigger decision within 2.5 μs, the trigger cannot use the calorimeter information at full granularity. Instead, approximately 7000 trigger towers (cf. Fig. 5.4) are used, which come with a granularity of 0.1×0.1 in η-ϕ in most parts of the detector ($|\eta| < 2.5$). After the analogue read-out the trigger towers are provided to the pre-processor. Here the analogue calorimeter signals are digitized, timed in and calibrated. The pre-processor also builds the so called *jet elements* for later usage by the jet/energy-sum processor (JEP), by summing up 2×2 trigger towers in η-ϕ. In the JEP, previously separated em. and hadronic trigger towers are summed.

Jets at the L1 trigger are found by the JEP, using a sliding window algorithm. This algorithm uses so called regions of interest (ROIs), which are formed by 2×2 jet elements. To find jets at L1, those ROIs are identified that are a local maximum. The definition of a local maximum is best described by the illustration in Fig. 5.9. The jet's energy is determined by summing up all jet elements within a certain window surrounding the ROI, as depicted in Fig. 5.9. This window may be configured to encompass 2×2, 3×3 or 4×4 jet elements which are centered on the ROI, or chosen as to maximize the jet E_T in the case of the 3×3 window. Consequently, jets at L1 are square objects of size 0.4×0.4, 0.6×0.6 or 0.8×0.8 in η-ϕ. The latter and largest window size is the most commonly used. The area of such a L1 jet thus approximately corresponds to that of anti-k_t jets with $R = 0.45$.

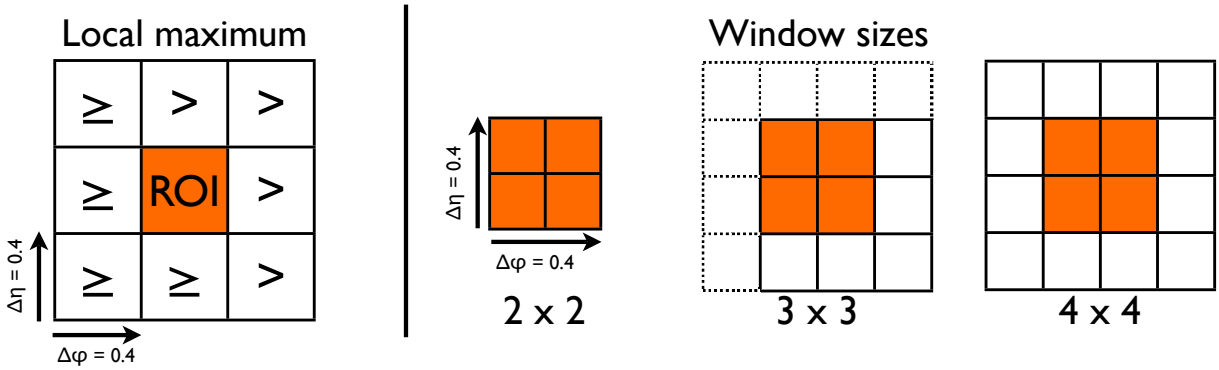

Figure 5.9: *Left: Requirements for a ROI to be a local E_T maximum. Its transverse energy has to compare to all neighboring ROIs as indicated. Right: Possible window configurations for the determination of the jet E_T at L1. The colored region is the ROI, which is in the center of the smallest and largest window. For the middle window size the centering is not possible, but four positions within the indicated, dashed grid exist. The position is chosen as the one that maximizes E_T.*

Electron, photon and τ candidates are found by the cluster processor (CP). Here also appropriate sliding window algorithms are used with finer granularity than for jets and additional isolation criteria are applied. They are described in detail in [51].

The configuration of the calorimeter trigger consists of defining a set of E_T thresholds, to which the ROIs found are compared. Due to the time and hardware constraints, only a limited amount of such thresholds can be configured. Eight thresholds can be defined for central jets ($|\eta| < 3.2$) and four thresholds for $\eta < -3.2$ and for $\eta > 3.2$ forward jets can be configured. Additionally twelve thresholds can be programmed in total for $E_{T,Miss}$ and $E_{T,Sum}$. For the cluster processor, in total 16 thresholds may be defined.

The number of ROIs exceeding each threshold are provided to the central trigger processor (CTP). Based on these multiplicities, the CTP makes the global L1 trigger decision, according to the defined trigger signatures. The naming and definition of these signatures happens according to multiplicity and threshold value. For instance a signature named "L1_nJx" would correspond to a jet ROI threshold, requiring n jet ROIs with transverse energies above x GeV in the event in order to pass the trigger. Up to 256 such signatures, including combinations of jet, electron and muon multiplicities, can be defined, making the system very flexible.

The minimum bias triggers

The L1 calorimeter and muon triggers are based on the selection of high p_T objects. For the selection of low p_T objects, so called minimum bias triggers are used to also select almost arbitrarily soft pp-collisions. One such trigger system is composed of the *minimum bias trigger scintillators* (MBTS), which are mounted as disks in front of the end-cap calorimeters, covering $2.09 < |\eta| < 3.84$. In order to select hard scattering events, the loosest requirement applicable using the MBTS is to require in total one hit in any of the scintillators.

A different approach to the selection of pp-collisions are the Zero-Degree calorimeters (ZDC) [53], which are mounted very close to the beam-pipe at $z = \pm 140\,\mathrm{m}$, thus covering pseudorapidities $|\eta| > 8.3$. Due to their coverage up to very high pseudorapidities, the ZDC can be used to efficiently select collision events based on coincident hits in both ZDC systems.

Level-1 trigger signatures

For completeness here the L1 trigger signatures are listed, which will be used in the upcoming analysis. The calorimeter trigger signatures are the fundamental ones, selecting the majority of QCD jet events. The MBTS triggers will be used as reference triggers in the determination of trigger efficiencies and for the selection of the lowest p_T events. The ZDC trigger will serve as a reference trigger to cross-check the trigger efficiency of the MBTS trigger. All items are summarized in Table 5.1.

Signature	Requirement	Window-Size
L1_ZDC	Coincident hits	-
L1_MBTS_1	One hit in total	-
L1_J5	$E_{T,\mathrm{ROI}} > 5$ GeV	4×4
L1_J15	$E_{T,\mathrm{ROI}} > 15$ GeV	8×8
L1_J30	$E_{T,\mathrm{ROI}} > 30$ GeV	8×8
L1_J55	$E_{T,\mathrm{ROI}} > 55$ GeV	8×8

Table 5.1: *Level-1 trigger signatures used in the analysis.*

5.5.2 The High-Level trigger

Upon acceptance by the L1 trigger, only data within identified ROIs is transmitted to the L2 trigger system. This data amounts to only 1 % to 2 % of the full event data. In contrast to L1, here the data within the ROIs can be used at full granularity. Thus for instance

jet finding can be done using the full calorimeter information and more sophisticated jet algorithms and calibrations can be used.

At the event filter finally the full event data is available, thus algorithms very similar to the ones used in the final reconstruction can be employed.

For the data selection of the upcoming analysis, the HLT was configured to not reject events but only run in a passive mode. In this mode, all trigger algorithms are executed and their decision is stored. But the event is accepted only based on the L1 decision, irrespective of the L2 and EF algorithm's results. For this reason the usage of HLT information is not necessary and for further details of the HLT it is referred to [51].

5.6 Data acquisition and computing

Data taken is organized in luminosity blocks and runs. One run corresponds to the data taken between the start and end of a certain recording period, which commonly coincides pretty well with the injection of a new fill into the LHC and the dump of beam. Depending on the beam lifetime, individual runs last between one and several hours. Each run is subdivided into so called luminosity blocks, the smallest chunk of data for which an integrated luminosity value is defined. Luminosity blocks last approximately 2 minutes.

Events accepted by the trigger are written to disk at the local computing infrastructure at CERN. In order to guarantee the accessibility of the data for all ATLAS members, the computing infrastructure is based upon pyramid-like data distribution system, arranged in several "Tiers" [54]. Tier-0 is the data storage and computing system at CERN, where all data taken is kept in its raw format. Here also the major reconstruction and processing of the data is done.

From the Tier-0 more and more derived forms of the raw data are distributed to Tier-1's, which also keep a significant amount of raw data. Via the Tier-2's and the Tier-3's, the data is increasingly more refined and the data volume is reduced. The increasing amount of higher tiers allows to provide the recorded data to all physicists being a member of ATLAS collaboration.

5.7 Physics and detector simulation

The simulation of physical processes and the response of the detector to a given event is a crucial ingredient to most analyses of high energy particles physics. Specifically many corrections that have to be applied to data are derived relying to a certain degree on the simulation.

5.7.1 Physics simulation

These simulations can generally be viewed to consist of two parts. Firstly events have to be generated, where the nomenclature *event* commonly refers to a collection of stable* particles and their four-momenta. These events can be simulated using one of numerous existing MC event generators. For the simulations that will be used in this analysis, these generators are `Pythia6` and `Herwig++`. Both generators implement LO QCD matrix elements, but are complementary in the modeling of parton showers, hadronization and the

*In the ATLAS simulation framework all particles with a lifetime $\tau > 10\,\text{ps}$ are considered stable.

underlying event. For the parton showering for instance `Pythia6` uses a p_T ordered shower, while `Herwig++` uses angular ordering. As briefly outlined Section 2.3, the hadronization in `Pythia6` uses the Lund string model, while `Herwig++` models the hadronization using a cluster model.

Since many aspects of the event simulation cannot be calculated from first principle, they are handled by phenomenological models. These generally need a number of parameters to be adjusted properly, in order to provide a sensible description of real data. This in particular holds for the parameters controlling parton showering, hadronization and the underlying event. A certain, well defined set of such parameters is commonly called a *tune*. For `Pythia6`, the default tune used, if not stated otherwise, is the MC09 tune, done by the ATLAS collaboration [44] which uses the MRST LO* PDF [55]. `Herwig++` is used with a underlying tune to ATLAS measurements, done by the `Herwig++` authors [42].

The most commonly used MC samples in the following will be the so called *dijet* MC samples, generated using `Pythia6`. In these samples, all processes relevant for QCD jet production, namely all $2 \rightarrow 2$ parton scatterings are simulated. In order to obtain reasonable statistics also for jets at highest p_T, various sub-samples are generated in bins of $\hat{p}_T$, which denotes the transverse momentum transfer in the hard scattering[†]. For analysis purposes these sub-samples are mixed, weighted by their corresponding cross sections. The limits used for $\hat{p}_T$ and the cross sections of the sub-samples are listed in Table 5.2. For each Jx sample approximately 1.4 million events are simulated and processed through the full detector simulation.

Sub-sample	$\hat{p}_{T,min}$ [GeV]	σ [nb]
J0	8	9.85×10^{6}
J1	17	6.78×10^{5}
J2	35	4.10×10^{4}
J3	70	2.20×10^{3}
J4	140	8.77×10^{1}
J5	280	2.35
J6	560	3.36×10^{-2}
J7	1120	1.37×10^{-4}
J8	2240	6.21×10^{-9}

Table 5.2: *Listing of the dijet sub-samples used and their corresponding cross sections. For a given sample, the upper limit for $\hat{p}_T$ is the lower limit of the next higher sub-sample. For J8, no upper limit on $\hat{p}_T$ is imposed.*

5.7.2 Detector simulation

In the second step of the simulation generated events are *placed* in the simulated experimental environment, namely the ATLAS detector. The GEANT framework [56, 57] is used to simulate the interactions of the generated particles traversing the single detector components. This includes the particle's interactions with material and the response of detector electronics to the simulated signal. Upon completion of the full simulation chain, the available events and their data structure is identical to that of real data, allowing for a reconstruction using the very same algorithms.

[†]More precisely, in terms of the mandelstam variables $\hat{s}$, $\hat{t}$ and $\hat{u}$ of the partonic $2 \rightarrow 2$ process, it can be written as: $\hat{p}_T = \hat{t}\,\hat{u}/\hat{s}$

As the complexity of the full simulation is beyond the scope of this thesis, only very few details will be addressed here. More details can for instance be found in [58] for the general simulation framework and in [59] for details on the simulation of calorimeter showers.

The simulation of hadronic showers, due to the multitude of processes involved, is less well understood than for instance the modeling of electromagnetic showers. In the GEANT framework a variety of phenomenological and parametrized models exist. A certain combination of models used is denoted a *physics list* and commonly consists of a few different models, which handle the various types of interactions. The default physics list was established using data from testbeam runs [60], where fully instrumented slices of the final ATLAS calorimeters were exposed to protons and charged pions from the Super Proton Synchrotron at CERN. In this context the calorimeter performance can be studied in great detail since the energy and species of the incoming particle is known.

Most importantly it was possible to study the average response and the resolution, as illustrated in Fig. 5.10. These results concluded that the best calorimeter simulation

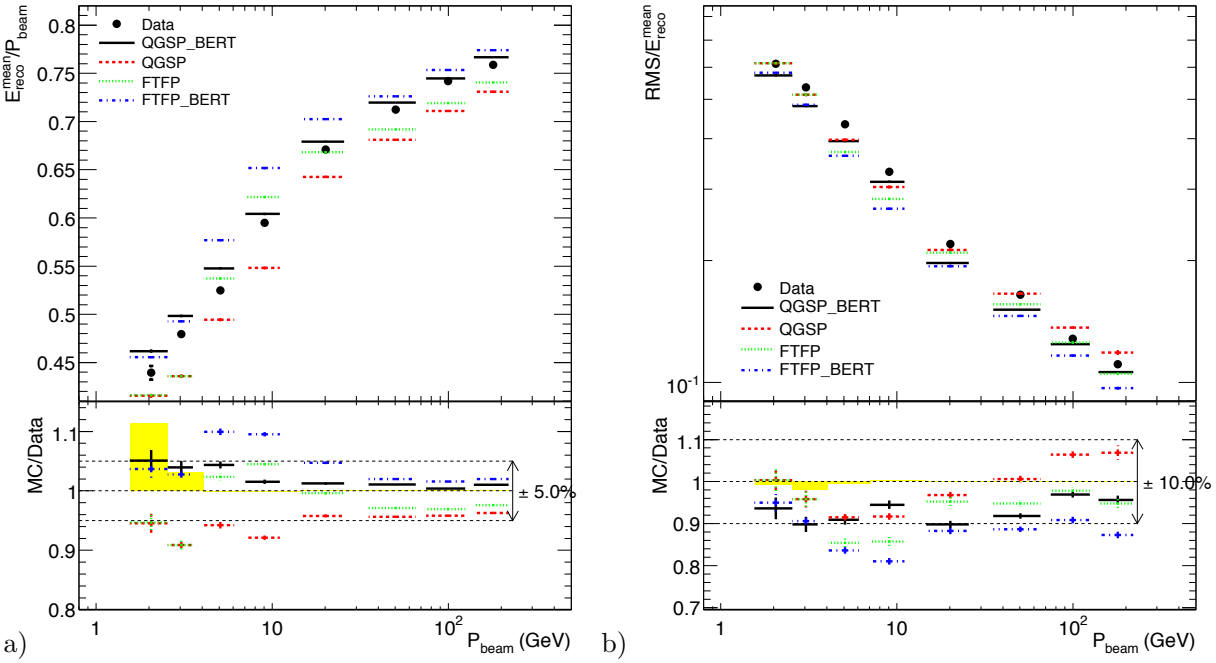

Figure 5.10: *Mean measured energy (a) and resolution (b) in the reconstruction of charged pions as a function of the incident pion momentum. Figure taken from Ref. [59].*

is achieved with the QGSP_BERT physics list, which uses a Quark-Gluon-String (QGS) model for the description of high energy hadron-nucleon interactions and the Bertini (BERT) cascade for hadron interactions below 10 GeV (for details see [59]).

6 Jet reconstruction and calibration

This chapter covers the technical aspects of jet reconstruction and calibration. It will cover all details of how the basic quantities as provided by the calorimetric reconstruction described in Section 5.3.4 are refined to finally end up being jet constituents.

The general calorimeter reconstruction provides the input to the jet reconstruction in form of cell energies calibrated at the electromagnetic scale. An average event may contain numerous cells with negative energy due to noise fluctuations. Care has to be taken of these negative energy cells before they can be fed into a jet finding algorithm. This is due to the fact that commonly used jet algorithms require physical four-vectors as input, namely four-vectors with $E \geq 0$. This is only one very technical reason why the ATLAS jet reconstruction prefers so called *topological clusters* as the input to jet finding. Other, physical advantages will come up in the following.

6.1 The formation of topological clusters

The topological clusters used in the jet reconstruction are intentionally designed to collect cell energies in the calorimeter that belong to one incident particle that might have spread its energy among several calorimeter cells. For this purpose the algorithm described in the following clusters single calorimeter cells into three-dimensional blobs in the detector. The algorithm follows the idea to grow clusters around so called *seed* cell by an iterative attachment of the neighboring cells. The algorithm can be summarized as follows:

1. **Finding seeds:** The seed cells are required to have an energy significance above a reasonably high seed threshold t_{seed}. In this step all cells fulfilling this requirement,

$$\frac{|E_{cell}|}{\sigma_{cell}} > t_{seed}, \tag{6.1}$$

   are found. Here $|E_{cell}|$ is the absolute value of the cell's energy and σ_{cell} denotes its expected noise. The noise includes electronic noise and the contribution expected from pile-up. All seed cells identified in this way form preliminary clusters, so called *proto-clusters*.

2. **Adding neighbors:** Sorted by descending energy significance, for each proto-cluster the neighboring cells are identified. As long as neighboring cells whose energy significance exceeds $t_{neighbor}$ are found, the proto-cluster is extended by these cells and the procedure is repeated. Proto-clusters sharing at least one neighboring cell are merged. This way clusters grow until all neighboring cells' significance is below $t_{neighbor}$.

   As neighboring cells all eight surrounding cells in the same calorimeter layer are considered. In addition also cells in adjacent calorimeter layers or systems partly overlapping in η and ϕ are regarded as neighbors.

3. **Adding surrounding cells:** Sorted by descending energy significance, for each proto-cluster all surrounding cells exceeding the cell threshold t_{cell} are attached. Surrounding cells adjacent to two proto-clusters are attached to the first proto-cluster processed, thus the one providing the higher energy significance. Proto-clusters in this step are thus not merged and are converted to final clusters.

4. **Cluster splitting:** The above algorithm may obviously merge showers of more than one particle into only one cluster. Given a resolvable separation between the two the cluster splitting can undo this using a search for local maxima in all clusters found with the procedure above. Cells being local maxima have to fulfill three criteria:

   - The cell's energy has to exceed the one of all surrounding cells.
   - The cell's energy has to exceed 500 MeV.
   - The cell must have at least four neighboring cells in the parent cluster.

   For clusters containing more than one local maximum the clustering procedure is repeated using the local maxima as seeds and considering only the cells within the parent cluster. In contrast to the default clustering procedure this second pass does not allow for merging. Shared neighboring cells are attached to the local maximum providing the higher energy significance. This procedure results in exactly one cluster per local maximum.

The full algorithm is explained in more technical detail in Ref. [61]. The parameters for the three different thresholds regulating the clustering are summarized in Table 6.1. The

Threshold	Value
t_{seed}	4
$t_{neighbor}$	2
t_{cell}	0

Table 6.1: *Threshold values for the topological clustering algorithm.*

seed threshold of $t_{seed} = 4$ is the main parameter for the suppression of clusters purely formed by noise, as it determines the probability to find pure noise clusters. One can estimate the expected number per event N_{noise} of such clusters as

$$N_{noise} = N_{cells}\sqrt{\frac{2}{\pi}}\int_{t_{seed}}^{\infty} e^{-t^2/2}dt, \tag{6.2}$$

where N_{cells} is the total number of calorimeter cells taken into account and the complementary error function describes the probability for each cell that its noise fluctuates beyond four standard deviations. Considering the large amount of calorimeter cells in the ATLAS detector of approximately 180000 this equation predicts approximately 12 noise clusters per event.

Thanks to the low noise in the ATLAS calorimeters the contribution from pure noise clusters to jets built later on can be expected to be small. The noise can be measured using randomly triggered and consequently empty events. The results of such measurements can be seen in Fig. 6.1. The expected and actually measured distribution of clusters in collision events as a function of pseudorapidity is shown in Fig. 6.2 for events selected with a minimum bias trigger. The observed differences can be attributed to an imperfect

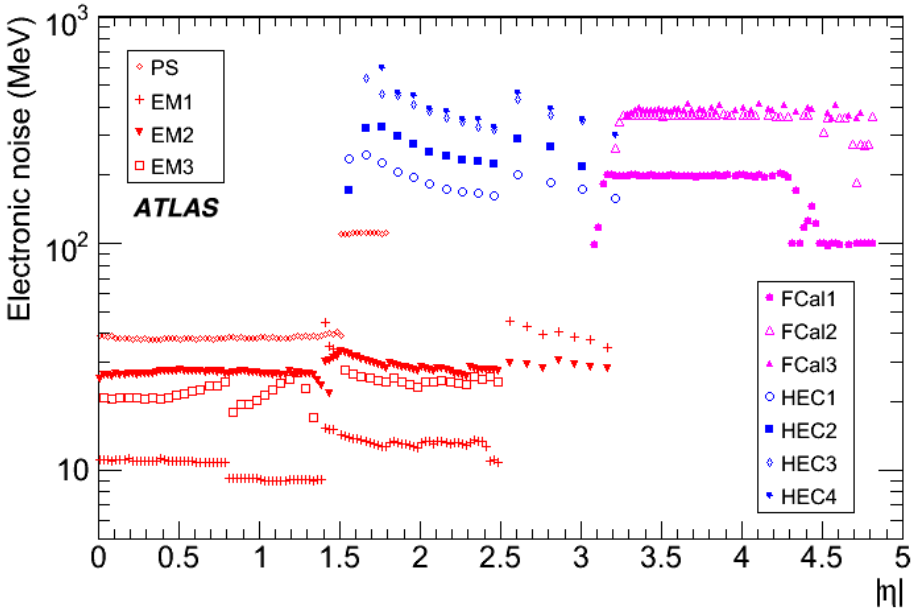

Figure 6.1: *Electronic noise in randomly triggered events at the electromagnetic scale in individual cells for each layer of the calorimeter as a function of $|\eta|$. Results are averaged over ϕ. Figure taken from Ref. [62]).*

modeling of the noise and the amount of dead material in front of the calorimeters. Here the MC description shows an overall quite good agreement, with maximum deviations at the level of 10%.

6.2 Jet identification

As input to a jet finding algorithm it is necessary to be able to provide the cluster collection in a four-vector representation. For this purpose each cluster is assigned a massless four-vector with energy equal to the clusters energy and its position being the energy centroid of the cluster constituents:

$$\begin{aligned} E_{cluster} &= \sum_i E_i \\ \eta_{cluster} &= \frac{1}{E_{cluster}} \sum_i E_i \eta_i \\ \phi_{cluster} &= \frac{1}{E_{cluster}} \sum_i E_i \phi_i \\ M_{cluster} &= 0. \end{aligned} \tag{6.3}$$

In this representation the collection of clusters can easily be interfaced to a jet algorithm. Jets are found using the anti-k_t algorithm with two different jet radii $R = 0.4$ and $R = 0.6$. For this purpose the ATLAS reconstruction software uses the publicly available FastJet [16] library that performs the jet finding and recombination. The result is a collection of four-vectors representing the jet's momentum.

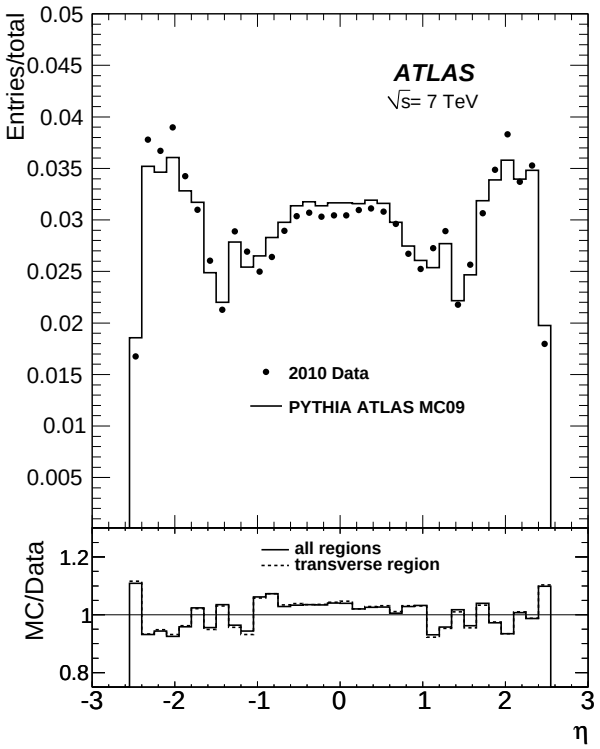

Figure 6.2: *Distribution of cluster pseudorapidity in data and MC. Both histograms are normalized to unity. Figure taken from Ref. [30].*

6.3 Jet energy calibration

Jets as the direct output of the jet finding algorithm are at the calibration level that the calorimeter reconstruction provides, namely the electromagnetic (*em.*) scale. The fact that jets not only consist of em. interacting particles makes a dedicated calibration for jets necessary.

6.3.1 Jet energy reconstruction

The global aim of the jet calibration is to calibrate jets such that the average reconstructed jet p_T is the *true* incident jet transverse momentum. The reason why this achievement is non-trivial lies within the fact that jets are a mixture of hadronically and electromagnetically interacting particles. Electromagnetic showers can well be modeled by cascades of π^0's going to photons which themselves deposit energy by electron-positron production. The complete shower energy thus is visible to an ionization medium in a calorimeter. Due to these fortunate features of em. showers the need for a dedicated jet calibration is mainly due to the properties of hadronic showers, which are outlined in the following.

The key feature of hadronic cascades is that significant energy fractions are lost in a non-ionizing manner, such as binding energy losses from nuclear spallations and recoil. Depending on the realization of a given calorimeter, these energy losses are not detectable and such calorimeters are called *non-compensating*. Due to π^0 production in secondary collisions with sufficient energy, a fraction f_{π^0} of the energy is transferred to an electromagnetic sector within the shower [63]. This transfer is irreversible because of the short π^0 lifetime. f_{π^0} thus gains a subtle energy dependence because the higher the energy, the

more secondary collisions occur that transfer energy to the em. sector. As shown in [63] this can be described by a power law of the form $f_{\pi^0}(E) = 1 - (E/E_0)^{m-1}$. Simulations for f_{π^0} are shown in Fig. 6.3. Using the energy dependence of f_{π^0} the response e.g. of a

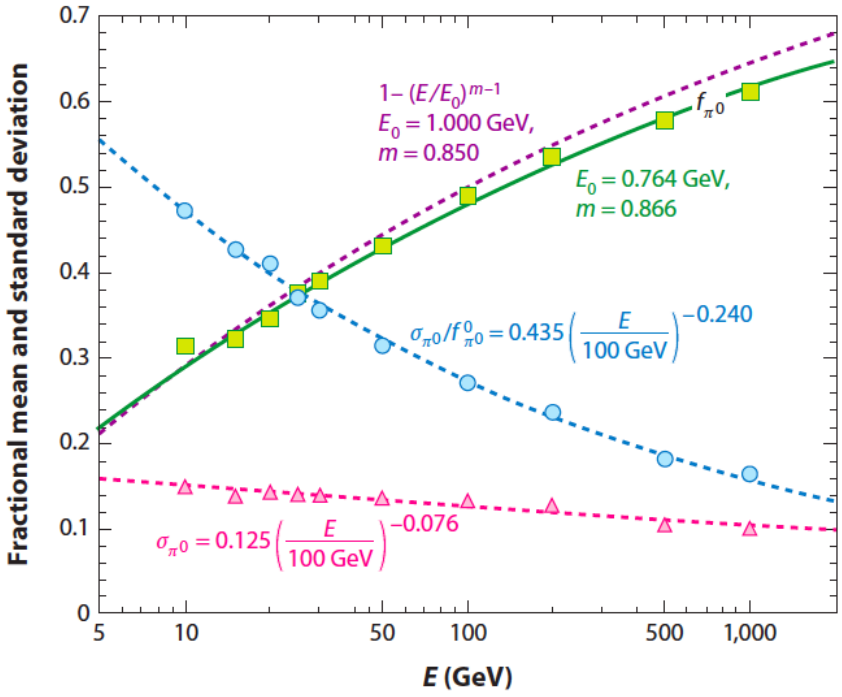

Figure 6.3: *Energy dependence of f_{π^0} and its standard deviation. Figure taken from Refs. [64, 65].*

charged pion relative to that of an electron can be formulated as [64]:

$$\pi/e = 1 - a\ E^{m-1}. \tag{6.4}$$

The generalization from a single charged pion to a jet is done in [64] and is found to approximately obey the same functional form. This idealization deliberately ignores eventual effects from magnetic fields or jet finding issues such as out of cone losses. Nevertheless Eq. 6.4 predicts the general features of the jet calibration: The correction can be expected to be smooth, following a power law. The response will asymptotically approach unity for infinite jet energies and the largest correction will be necessary at low p_T.

6.3.2 p_T- and η-dependent jet calibration

Of course the calibration used in the experiment can only access the measured signals in the calorimeter. For this reason the most straight forward jet calibration uses the information that the jet reconstruction at this stage can provide, namely the four-vector of the jet calibrated at the em. scale. The calibration scheme that will be introduced, consists of a set of parametrizations that determine a scalar factor for each jet, by which the four-momentum of the jet is multiplied [66]. The parametrizations make use of the jets' positions in the detector, since it can be expected that different subdetectors will need different calibrations. And of the jets' transverse momenta, which will be necessary since - as outlined above - the response can be expected to show a significant energy dependence.

Mathematically the calibration procedure can thus be formulated as

$$p_{jet}^{calib} = p_{jet} \times f_{calib}(p_T, \eta), \tag{6.5}$$

where p_{jet} refers to the four-vector representation of the jet. The calibration factors are derived from MC. To extract the response from MC a mapping between the generated particle jets and the reconstructed jets has to be established. This is achieved by a matching procedure, in which each particle jet is assigned the reconstructed jet that is closest in η-ϕ-space. The distance in this space is defined as

$$\Delta r = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}, \tag{6.6}$$

where $\Delta\phi$ and $\Delta\eta$ denote the separation of the two jets in ϕ and η respectively. Then the response at the per-jet level is defined as the ratio between the particle jet's energy E^{true} and the reconstructed jet's energy E^{reco}. We define its most probable value which will be extracted using Gaussian fits as the *response*:

$$\langle \mathtt{R} \rangle = \left\langle \frac{E^{\text{reco}}}{E^{\text{true}}} \right\rangle, \tag{6.7}$$

where the typographical $\mathtt{R}$ is used to distinguish the response quantity from the jet radius parameter. Being able to extract the response in this way, it is straight forward to define the calibration factor used in Eq. 6.5, which is the inverse of the response:

$$f_{calib}(p_T, \eta) = \langle \mathtt{R} \rangle^{-1} (p_T, \eta). \tag{6.8}$$

So far there was no distinction between particle and reconstructed jets e.g. in the dependency of f_{calib}. However, in order to be able to apply the calibration to data where particle jets are not available, there is no choice but to use p_T^{reco} and η^{reco}. This leaves one subtlety to be taken care of: the specification of the procedure to find the most probable response value. The easiest way to go would be to derive the response as $\langle \mathtt{R} \rangle_{reco} (p_T^{\text{reco}}, \eta^{\text{reco}})$, where $\langle ... \rangle_{reco}$ specifies that the most probable value is found in bins of measured transverse momentum and pseudorapidity. This however has one disadvantage: While a Gaussian distribution of $\mathtt{R}$ for a fixed range of E^{true} can be expected, the distribution is distorted once this quantity is binned in E^{reco}.

This can easily be illustrated in a toy MC, where E^{true} is randomized according to a steeply falling probability ($\propto E^{-4}$ and $\propto E^{-2}$) and E^{reco} according to an arbitrary resolution and a mean of one. The resulting distributions of $\mathtt{R}$, either in a fixed range of E^{true} or E^{reco}, are displayed in Fig. 6.4. The distortion observed is due to jets whose response is either very high or very low, such that they migrate out of or into the E^{reco} range looked at. The result is a shift in the mean of the distribution as well a distortion of the shape, which becomes slightly asymmetric. This leads to the fact that though the input energy scale used is one, the Gaussian mean of the $\mathtt{R}_{reco}$ distribution is significantly shifted upwards by approximately 3%. In particular this effect gets more pronounced the steeper the input spectrum is, because low p_T jets with high response will simply be much more abundant than high p_T jets with low response. In the above example the difference in the observed mean doubles when going from the E^{-2} spectrum to E^{-4}. In the context of jet calibration this is of exceptional importance, since it has to cope with the very steeply falling jet p_T spectrum.

To bypass this effect the calibration is done using bins of p_T^{true} in a two-step procedure:

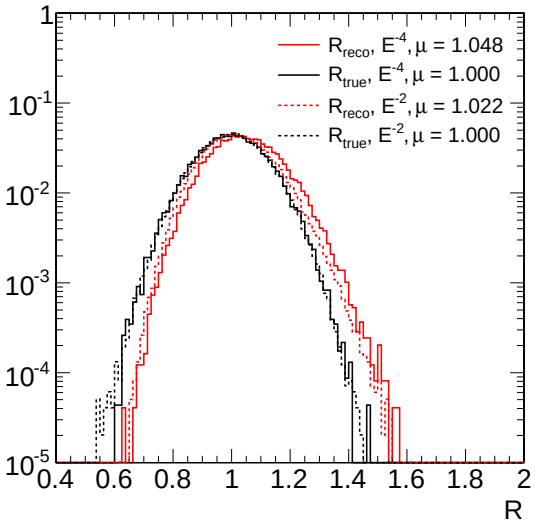

Figure 6.4: *Response distribution for two choices of binning: Binning in E^{reco} (red) and E^{true} (black). The solid histograms refer the to toy study done with an input spectrum proportional to E^{-4}, while for the dashed versions E^{-2} was used. The legend also displays the mean for all four distributions.*

1. In a first round the response is extracted in bins of p_T^{true}, which is denoted by $\langle \mathtt{R} \rangle_{true} (p_T^{\text{true}}, \eta^{\text{reco}})$.

2. The second step essentially is a transformation of the x-axis, in order to be able to use the extracted response as a function of p_T^{reco}: Each true jet's p_T^{true} is replaced by p_T^{est} being the best estimate of the reconstructed transverse momentum for this true jet. Using the response extracted in step no. 1 this is:

$$p_T^{est} = p_T^{\text{true}} \times \langle \mathtt{R} \rangle_{true} (p_T^{\text{true}}, \eta^{\text{reco}}). \tag{6.9}$$

   So now the response, which remains $\mathtt{R} = E^{\text{reco}}/E^{\text{true}}$, is extracted in bins of p_T^{est}. The distortion effect observed thus does not exist and the Gaussian response and proper mean value are preserved. The result now is $\langle \mathtt{R} \rangle_{est} (p_T^{\text{est}}, \eta^{\text{reco}})$. The application of the calibration can now simply follow Eq. 6.5 with the calibration function

$$f_{calib}(p_T, \eta) = \frac{1}{\langle \mathtt{R} \rangle_{est} (p_T^{\text{reco}}, \eta^{\text{reco}})},$$

   since the p_T^{est} was constructed to be the most probable value of p_T^{reco}.

6.3.3 Technical realization

Following the above described general scheme the derivation of this calibration in the experiment is described in the following.

The calibration that will be used throughout the remaining analysis was extracted using the `Pythia6` dijet MC samples. The detector description and simulation herein include all knowledge about the detector status at that time. As the calibration was designed to be applied using an already existing framework within the ATLAS software, the technical details closely followed previous implementations:

- The dependency of f_{calib} on η is accounted for by introducing 45 bins in $|\eta|$: 44 of these bins are 0.1 units of pseudorapidity wide, covering $|\eta| < 4.4$, while the 45$^{\text{th}}$ bin includes all remaining jets.

- For each bin of absolute pseudorapidity, numbered by *ieta*, the dependency of f_{calib} on p_T is accounted for by a polynomial of the form:

$$f_{calib}^{ieta}(p_T) = 1 + \sum_{i=1}^{4} \frac{p_i^{ieta}}{\log\left(p_T/\text{GeV}\right)^i}. \tag{6.10}$$

  The choice of inverse logarithms is based on the fact that the function in this way can safely be extrapolated to high p_T. It also reflects the expectation of an response of 1 for p_T going to infinity. The set of p_i^{ieta}'s are free parameters in these functions which are going to be fitted to the response extracted from MC.

The response in the MC is extracted as the most probable value of R in a given $(p_T^{\text{true}}, \eta^{\text{reco}})$ bin. For this purpose the R distribution in each of these bins is fitted twice with a Gaussian, yielding $\langle \mathtt{R} \rangle_{true}\,(p_T^{\text{true}}, \eta^{\text{reco}})$ as illustrated in Fig. 6.5. The first fit encom-

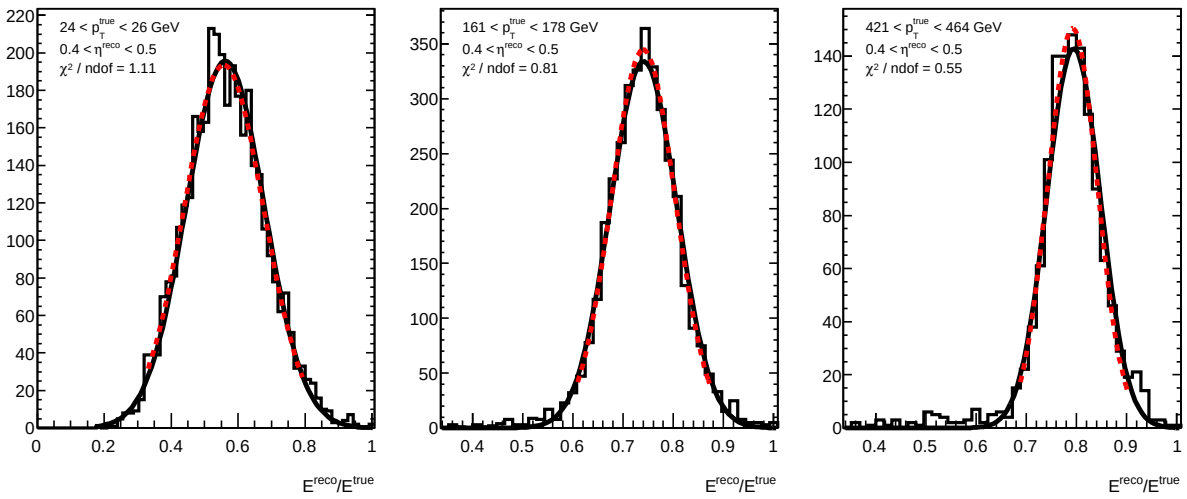

Figure 6.5: *Exemplary Gaussian fits for anti-k_t, $R = 0.6$ jets in different p_T^{true} bins done to determine the most probable response value. The solid black curve shows the first fit using the full range on the x-axis, while the second one (red, dashed curve) is used to find the Gaussian mean.*

passes the full range of R. Using the Gaussian μ and σ from this first fit a second one is performed within $\mu \pm 2\sigma$ in order to find the peak position. The fit result for μ in this second fit is considered the response in this certain bin and is plotted as a function of p_T^{true} and p_T^{est} in the first and second calibration step respectively.

Several technical items have to be taken care of during this procedure:

- **Jet matching:** In order to only map sensible pairs of particle and reconstructed jets to each other an upper limit on their distance is set: $\Delta r < 0.3$.

- **Jet reconstruction threshold:** Due to file size constraints the jet reconstruction was configured to only keep jets with $p_T > 4$ GeV. This fact limits the possibility to

gain information on the response of very low p_T^{true} jets as some of the reconstructed jets might fall below this threshold. For the example of a p_T^{true} bin starting at 10 GeV this implies that response values below 0.4 will be missed, because jets having such a low response will be filtered out. For this reason the Gaussian fits were only performed in a region where this cut does not influence the response.

- **Calibration of low p_T jets:** In consequence of the reconstruction threshold there is a lower limit for p_T^{true} (and thus p_T^{est}) at which a solid value for the response can be extracted. This directly raises the question of how to treat jets below this limit during the calibration process. Due to the nature of the parametrization in Eq. 6.10 the fits to the response tend to become unstable very quickly below the lowest p_T used in the fit. As a result, a naive evaluation of the parametrization below this threshold might return non-physical values for the response. For this reason the lowest p_T used to evaluate the function is set to 10 GeV. Consequently all jets below 10 GeV are calibrated with the same calibration factor:

$$f_{calib}(p_T < 10 \text{ GeV}, \eta) = f_{calib}(10 \text{ GeV}, \eta).$$

- **Jet isolation:** As the aim of the calibration is the correction of the calorimeter non-compensation and out-of-cone losses, only pairs of jets are used which are isolated with respect to other jets in the event. This requirement is aimed to reject cases where one particle jet is split into two reconstructed jets or – the other way round – two particle jets are merged into one reconstructed jet. These topologies will have either very high or very low response, since the energy in the second jet (on either particle or reconstructed level) will be lost in the comparison of the jets' energies. For this purpose only jets are considered whose next-to-nearest jet with $p_T > 7$ GeV is more than 2.5 times the jet radius away: $\Delta r_2 > 2.5\, R$.

Results of this procedure can be seen in Fig. 6.6, where the left plot shows the response extracted in the first calibration step and the right one is the result of the second step. The effect of the x-axis transformation is most clearly visible in the low p_T region, where the response is lowest. Apart from that, the general shape of course does not change significantly due to the very smooth response. As indicated by the $\chi^2/ndof$ on these figures, the parametrizations do not fit perfectly but certainly to sufficient degree. Besides the fits using Eq. 6.10 a fit using the Groom function introduced in Eq. 6.4 is shown as well. It is interesting to see that the general shape of the MC response is modeled pretty well by this function, while it cannot follow the generally stronger curvature. This does not surprise since the assumptions made to derive this formula were, as stated earlier, very idealized. Thus the fit using the less well motivated but more flexible polynomial is preferred.

In order to get a better overview of all the fits involved for the full set of 45 parametrizations, Fig. 6.7 shows the deviations between MC response and the fits, averaged over the according η bins. These histograms give confidence in the good impression gained from Fig. 6.6 and thus promise a good result. The scatter around zero is on the level of 2 ‰ and the oscillating structure implies that this is not an issue of the parametrization. These oscillations can be blamed on the fact that the Pythia dijet sub-samples here are used unweighted, due to complications in the Gaussian fits that would come along with a weighted usage. The associated drawback of the unweighted usage however is, that the mixture of jets originating from different dijet sub-samples is unphysical in the transition

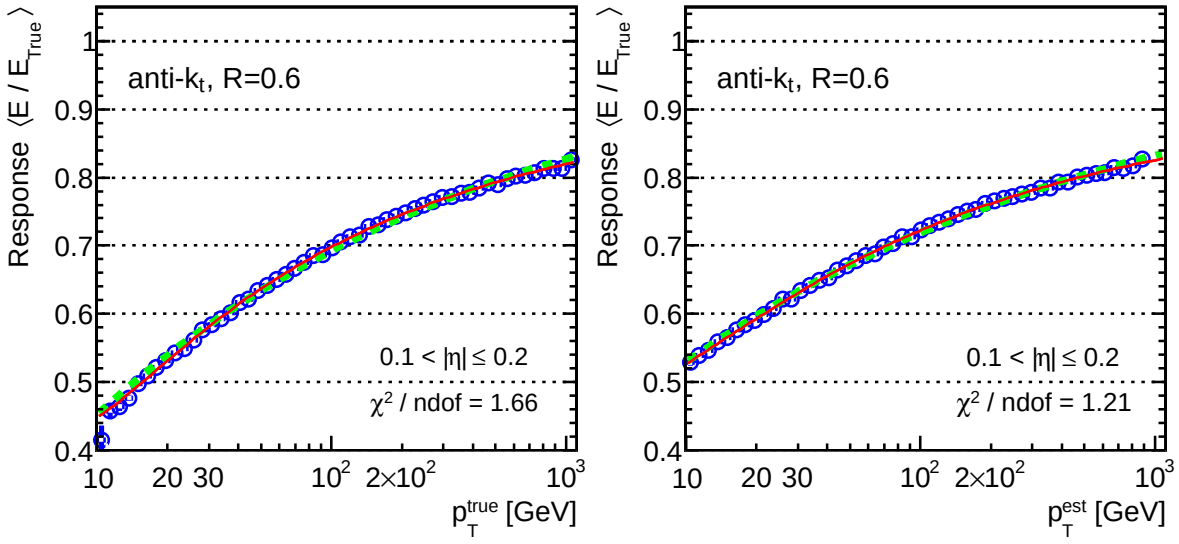

Figure 6.6: *Blue open circles shows the response extracted in bins of p_T^{true} in two pseudorapidity regions. The red line represents the fit result of f_{calib} whereas the green, dashed line shows a fit using Eq.* 6.4.

regions from one to another sub-sample. This leads to the periodic drops in the response. As this effect is only on a level of less than 2 ‰ and since the polynomial does a good job in averaging out these oscillations the procedure is left unchanged. The resulting two-dimensional calibration factors are displayed in Fig. 6.8. As expected the calibration factors for both jet sizes show the same structures in general, like a decreased response in transition regions between certain detectors, such as around $|\eta| \approx 1.4$ and $|\eta| \approx 3.2$. Besides, certain features are more pronounced for smaller jets, like the increase of the calibration factors at low p_T around $|\eta| \approx 3.2$.

As a closure test the calibration derived from the fits in Fig. 6.6 is applied on the identical MC sample than it was derived on. The resulting, calibrated jets are then fed through the very same calibration algorithm once again. The result is shown in Fig. 6.9. Linearity above 30 GeV is achieved within approximately ±1% for both jet sizes and for most regions of pseudorapidity. While the results are in fact fully satisfactory, they are not perfect. In particular the response for $R = 0.4$ jets shows a very clear trend to too high response at low p_T^{true}. This effect can be blamed on a subtlety of the isolation requirement: During the derivation of the calibration jets whose next-to-nearest jet is too close and has $p_T^{uncalib} > 7$ GeV are rejected. In the closure test $p_T^{calib} > 7$ GeV is applied in the isolation check. Considering the fact that $p_T^{calib} > p_T^{uncalib}$, this means that the isolation criteria itself becomes stricter. The stricter isolation of course leads to a slightly higher response, since in general less energy is lost outside of the jets. As this can be considered a constant effect, the impact on the observed jet response quickly falls with $1/p_T$. It is also remarkable that the same effect is already very much reduced going from $R = 0.4$ to $R = 0.6$ jets, because split or merged jet topologies occur less often.

6.3.4 Validation of the jet calibration

As a final validation of the calibration derived above, the jet response is finally cross checked after the calibration has been included in the default reconstruction software.

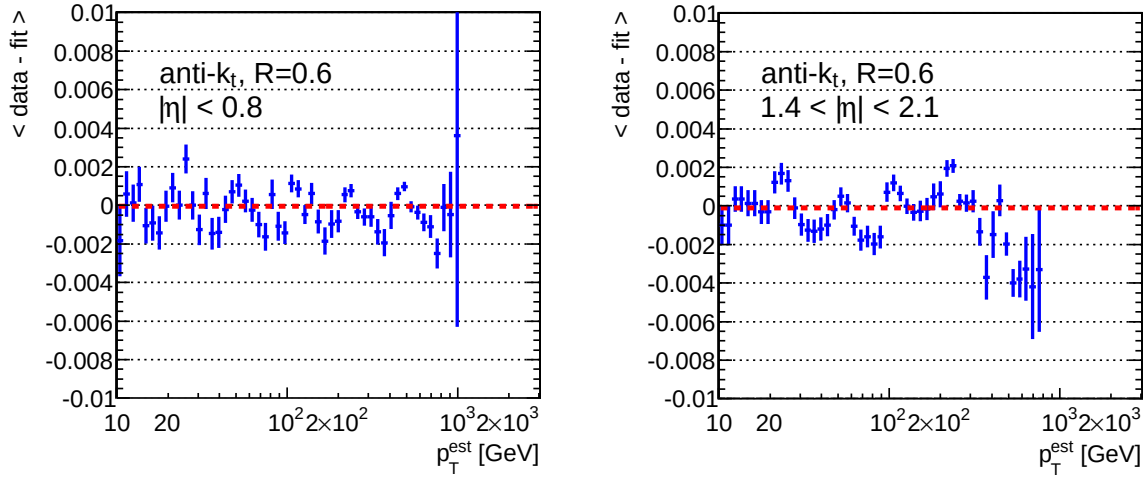

Figure 6.7: *Averaged deviations of the fits from the response as predicted by the MC. The solid red line represents a fit of the deviations to a constant function.*

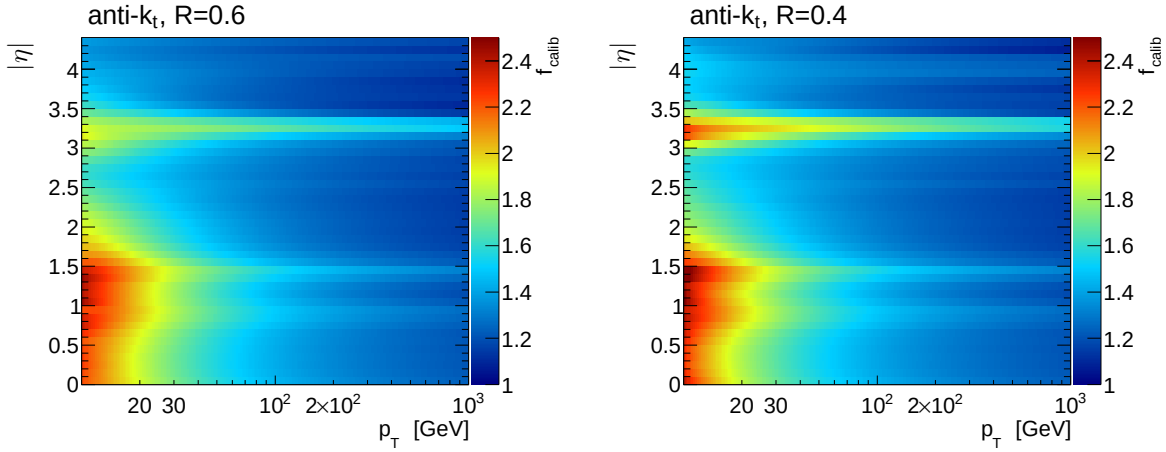

Figure 6.8: *Jet calibration factors as functions of p_T and η for both jet sizes.*

This was done in Fig. 6.10, that shows the linearity in an significantly larger MC sample than was used for the derivation. Also in this final crosscheck the calibration gives very good results which are close to the performance seen in the closure test. One should however be aware of the fact that though the technical performance of the jet calibration can get close to the ideal response to within 1-2%, this is by far no guarantee that the jet energy scale in data is known with the same precision. The calibration fully relies on a realistic MC simulation. As will be outlined in more detail in Chapter 8 the precision to which it can be claimed that the MC fulfills this prerequisite will be significantly larger than the remaining imperfections seen in the closure test.

6.4 Jet energy resolution

Besides the response considerations made above, the second main benchmark of energy reconstruction in general is the resolution, which will be a key ingredient of the upcoming analysis. In the energy measurement of jets the resolution is the sum of many effects.

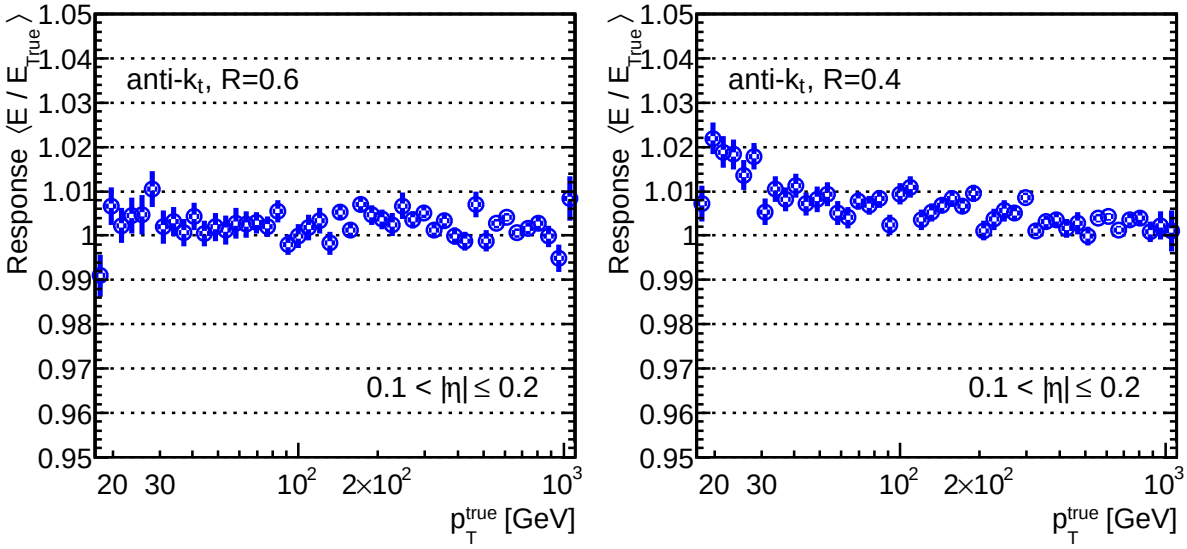

Figure 6.9: *Response after the calibration derived in Fig. 6.6.*

The dominating one can be expected to be the large fluctuations of the em. component in the jet and the resulting shower in the calorimeter. Two distinct effects lead to the fluctuations of this component: Firstly, there is the fluctuation of f_{π^0}, the size of which is illustrated in Fig. 6.3. It shows the standard deviation of f_{π^0}, which is found to be very large, up to a fractional width of almost 50% at low energies. Secondly the particle spectrum produced in the fragmentation of the jet varies significantly on a jet-by-jet basis. Minor, but not negligible contributions to the resolution may arise from out of cone losses, energy lost in not instrumented detector regions including leakage and the intrinsic calorimeter resolution.

Here the jet p_T resolution is studied in the MC. During the extraction of the linearity in bins of p_T as illustrated in Fig. 6.10, Gaussian fits were used to find the most probable value of the response. At the same time these fits yield the width of the Gaussian. This width σ divided by the mean μ is quoted as the relative p_T resolution*. In Fig. 6.11 the p_T resolutions are plotted as a function of p_T. For later usage the fractional p_T resolutions are fitted using:

$$\frac{\sigma}{p_T}(p_T) = \frac{a}{\sqrt{p_T}} \oplus \frac{b}{p_T} \oplus c. \tag{6.11}$$

In this parametrization a accounts for the statistical fluctuations within a calorimeter shower, which decreases proportional to $1/\sqrt{E}$. Energy independent contributions like the electronic detector noise are included in b. Fractionally constant contributions due to detector imperfections, such as leakage and albedo losses, are included in c. Dealing with jets, all parameters also absorb the contributions due to out-of-cone losses, which can however be expected to roughly scale as $1/E$, since jets with increasing energy become more narrow. Though originally this form of parametrization is used as a function of energy, it is also able to describe the jet p_T resolution as function of p_T.

The results of these fits are summarized in Table 6.2. The resolution improves going

*The division by μ partly corrects for an eventually wrong response, which could fake changes in the resolution: In a scenario where each jet's energy is scaled by $\frac{1}{2}$, the width of the response distribution would shrink to 50% as well, though in fact there is no real improvement in resolution.

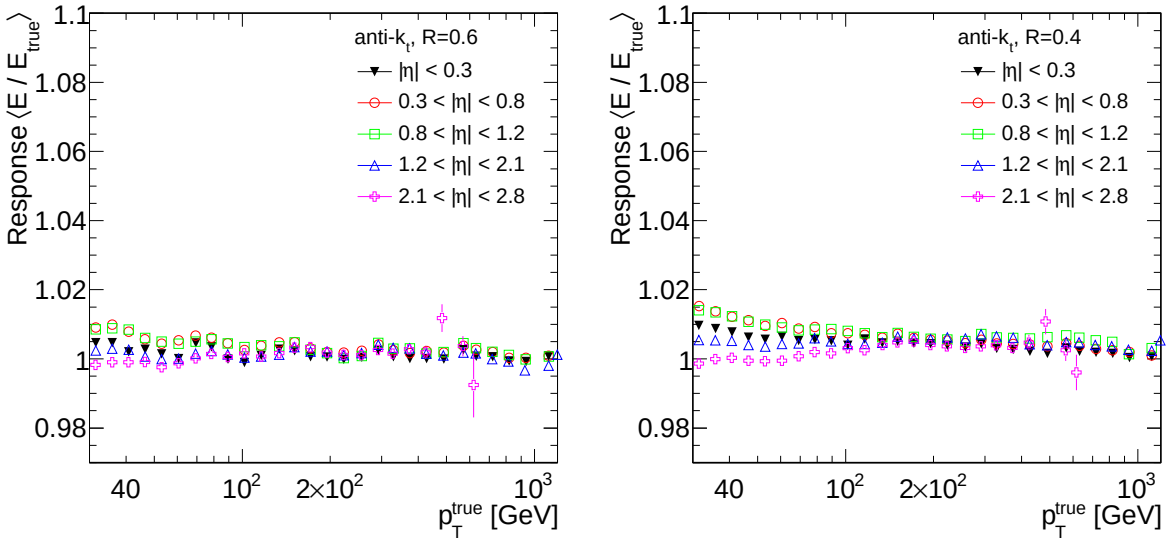

Figure 6.10: *Calibrated jet energy response as derived from the MC simulation.*

to the more forward rapidity regions, which can be attributed to the fact that the resolution intrinsically is a function of energy, rather than transverse momentum. As the jet energy increases while going to the more forward bins, the resolution improves. It is also worthwhile mentioning that the resolution for larger jets is slightly superior to the one using a smaller jet radius. This can be attributed to the fact that especially at low p_T out of cone losses are more important for $R = 0.4$ jets. The fit results confirm this, as they differ most in the p_T dependent terms, while the constant term c is very similar for both jet radii and ranges between 5.6% in the central η region, decreasing to 4.5% in the most forward regions looked at.It should be noted that the calibration scheme introduced

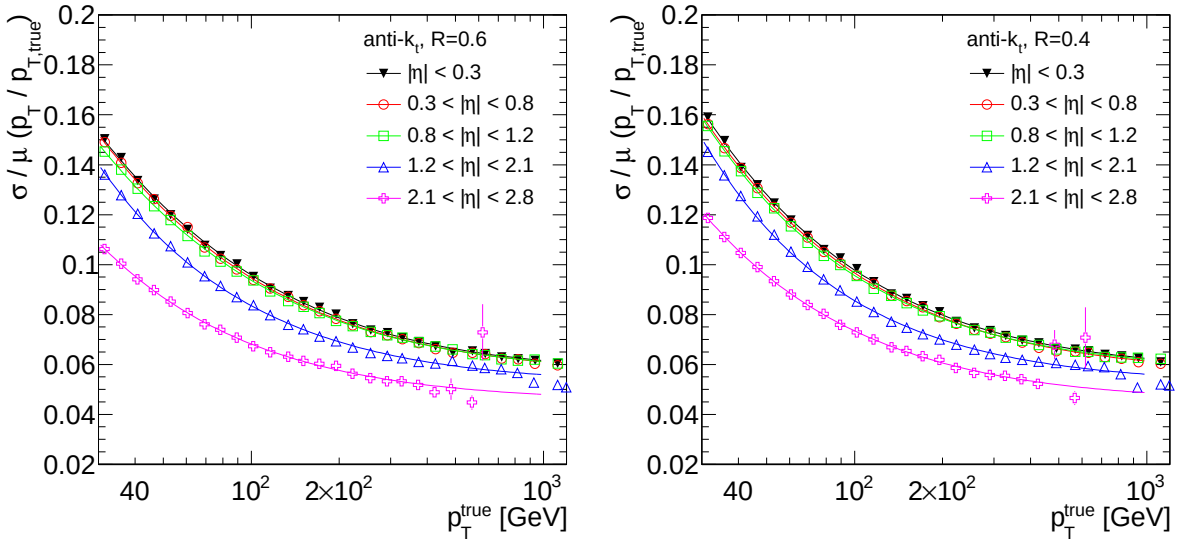

Figure 6.11: *Relative jet p_T resolutions as extracted from the MC in various bins of pseudorapidity. Lines indicate the fits done using Eq.* 6.11.

cannot improve the resolution despite a small benefit that arises from the equalization of the response across different detector regions. Besides, using only p_T as input to the

anti-k_t, $R = 0.6$

η-region	a [GeV$^{-1/2}$]	b [GeV]	c [%]	χ^2/ndof
$\lvert\eta^{jet}\rvert < 0.3$	0.774 ± 0.005	0.435 ± 0.416	5.64 ± 0.02	38.3 / 24
$0.3 < \lvert\eta^{jet}\rvert < 0.8$	0.758 ± 0.004	0.885 ± 0.155	5.61 ± 0.02	43.7 / 24
$0.8 < \lvert\eta^{jet}\rvert < 1.2$	0.748 ± 0.005	0.427 ± 0.359	5.64 ± 0.02	28.4 / 24
$1.2 < \lvert\eta^{jet}\rvert < 2.1$	0.624 ± 0.004	1.884 ± 0.051	5.22 ± 0.02	79.5 / 24
$2.1 < \lvert\eta^{jet}\rvert < 2.8$	0.495 ± 0.007	1.207 ± 0.089	4.54 ± 0.05	38.7 / 21

anti-k_t, $R = 0.4$

η-region	a [GeV$^{-1/2}$]	b [GeV]	c [%]	χ^2/ndof
$\lvert\eta^{jet}\rvert < 0.3$	0.775 ± 0.005	1.648 ± 0.097	5.74 ± 0.02	35.7 / 24
$0.3 < \lvert\eta^{jet}\rvert < 0.8$	0.771 ± 0.004	1.534 ± 0.079	5.63 ± 0.02	47.9 / 24
$0.8 < \lvert\eta^{jet}\rvert < 1.2$	0.745 ± 0.005	1.769 ± 0.080	5.73 ± 0.02	27.6 / 24
$1.2 < \lvert\eta^{jet}\rvert < 2.1$	0.632 ± 0.004	2.447 ± 0.038	5.23 ± 0.02	113.9 / 24
$2.1 < \lvert\eta^{jet}\rvert < 2.8$	0.561 ± 0.007	1.406 ± 0.084	4.53 ± 0.05	27.0 / 21

Table 6.2: *Resolution fit parameters extracted from Fig. 6.11*

calibration does not provide sufficient information for a real improvement in jet energy resolution.

In order to achieve better resolutions in general quantities that make it possible to estimate whether a jet will have either a low or high response on a jet-by-jet level have to be found. Various calibration schemes using such information exist (see e.g. [67]) and have been shown to be able to improve the constant term c by approximately 30% down to 4% [68].

7 Jets in the ATLAS detector

Before proceeding to the actual analysis, this chapter will provide an overview of the performance of jet reconstruction in data and the MC simulation. After a brief outline of the necessary event selection steps, the criteria to reject jets that are for instance due to noise are discussed. With a proper selection, the subsequent studies will turn to basic features of jet reconstruction such as kinematic distributions and the inner structure of jets.

7.1 Data and event selection

The data used for this analysis was collected in summer 2010 and includes a total of 75 runs. Runs taken under similar accelerator conditions are grouped into so called data periods A to F, where each refers to a well defined list of runs.

In order to analyze only events taken under stable and good conditions, the ATLAS collaboration maintains a data quality monitoring system. This ensures control of the detector conditions at every time of data taking.

7.1.1 Data quality requirements

Data quality within ATLAS can be understood as a set of flags for various subdetectors that indicate to which degree it was working properly at the time of data taking. To reflect the eventual development of such states within a given run, this information can be changed on the basis of a luminosity block. For each category or subdetector the data quality flags are defined as green, yellow or red:

- **Green:** This flag is set if there is nothing to object about the state of the subdetector. Data taken under green conditions can be used right away for physics analyses.
- **Yellow:** Yellow conditions are set if the existing problems can still be recovered, such as non-nominal high voltage at the calorimeters, which can be accounted for in the calibration. Data taken under these circumstances generally is still valuable for physics analyses.
- **Red:** Red detector conditions usually indicate that the affected detector was not running at all or under unrecoverable circumstances. These defects occur mainly due to tripped power supplies which leave the affected system switched off completely. This data is usually lost for any analyis making use of affected detector systems.

In order to provide this information for all necessary parts of an analysis, the data quality is monitored online and offline. The online monitoring concentrates on hardware related detector issues, while the offline monitoring includes regular sanity checks of e.g. kinematic distributions of basic objects like leptons and jets.

This information is provided in form of so *c*alled good runs lists (GRL), built such that all detector parts participating in the reconstruction of event features used in this analysis are required to be functioning well at the time of data taking. For this analysis, starting from general items to more specific ones, the following list of data quality categories were required to be green:

- Stable beams: The proton beams are required to be under stable conditions. Though collisions may also occur under non-stable conditions, several detector parts are as a safety measure only running during stable conditions.
- Level-1 trigger system: Since all triggers depend on a functioning Level-1 trigger system, it is crucial that this part of the data taking machinery is functioning properly.
- Inner detector: As a prerequisite for a proper vertex reconstruction all inner detector parts are required to be under green conditions.
- Combined performance jet flags: This item is closely related to the final analysis. The combined performance flags determine whether the calorimeter system was operating well. In addition to the proper functioning of the calorimeter hardware, this would also include eventual problems in the jet reconstruction software, such as bad calibration constants. For this purpose a multitude of histograms are monitored during data taking, such as kinematic distributions e.g. of leptons and jets and associated quantities.

7.1.2 Trigger selection

All events have to satisfy at least one of the inclusive jet triggers L1_J5, L1_J15, L1_J30, L1_J55 or the L1_MBTS_1 trigger as described in Section 5.5. In order to select an event sample that is not biased by the trigger used, for each trigger the p_T region in which its efficiency is sufficiently close to 100% is determined. These minimum p_T values are increasingly far beyond the actual threshold value applied at Level-1 and can be read off Table 7.1. In the following, for jets with a given p_T, the event is always required to have passed the highest inclusive jet trigger for which $p_{T,min} < p_T$. This requirement and the actual values for $p_{T,min}$ also drive the choice of the bin-edges for observables binned in p_T.

The determination and validation of these trigger plateau regions will be done in Section 7.5. Following the rather technical data selection made so far, the event selection

Trigger	$p_{T,min}$ [GeV]	L [nb^{-1}]	Avg. prescale
L1_MBTS_1	< 20	0.62	5000
L1_J5	60	32.0	97.5
L1_J15	110	553	5.64
L1_J30	160	2000	1.56
L1_J55	260	3120	1

Table 7.1: *List of triggers used and the extracted minimum p_T for 100% efficiency. Also shown is the integrated luminosity per trigger and its luminosity-weighted average precsale.*

described in the following starts looking at physical objects in the reconstructed events in order to reject non-collision backgrounds.

7.1.3 Primary vertex selection

The primary vertex per event is chosen as the vertex with the maximum $p^2_{T,Sum}$, defined as the quadratic sum of the transverse momenta of tracks associated to the vertex:

$$p^2_{T,Sum} = \sum_{i=0}^{nTracks} p^2_{T,i} \tag{7.1}$$

A very loose, but sufficient requirement is a minimum number of associated tracks of $N_{PV}^{Tracks} \geq 5$. This excludes possible vertices built from one track left by a cosmic particle traversing the detector close to the nominal interaction point. Since the vertexing algorithm expects tracks to originate from the nominal interaction point, such a track could accidentally be split in two tracks, which might form a fake vertex. A second source for non-collision backgrounds are interactions of a proton with residual particles in the beam pipe. Studies with looser selection criteria than the one presented, showed that the fraction of such background contributions is below 10^{-4} and thus negligible (see [69]).

7.1.4 Missing E_T significance

The primary vertex selection may only reject events that solely consist of non-collision background. Non-collision backgrounds however may also contaminate an otherwise good event. Here the fact that QCD events do not contain *real** missing transverse energy ($E_{T,Miss}$) can provide a handle to identify events significantly contaminated by non-collision backgrounds. The $E_{T,Miss}$ quantity measures the transverse imbalance of an event as defined by

$$E_{T,Miss} = \sqrt{E^2_{Miss,X} + E^2_{Miss,Y}}, \tag{7.2}$$

where $E_{Miss,X}$ and $E_{Miss,Y}$ are the vectorial sums of the x- and y-components of all energy depositions in the calorimeter. Thus in an idealized detector both quantities vanish due to transverse momentum conservation, whereas in reality neutrinos and muons are invisible to the calorimetry. Energy depositions from non-collision backgrounds in general spoil this momentum conservation and so those events will show high $E_{T,Miss}$. For this reason it is suitable to apply an upper limit on $E_{T,Miss}$ in a QCD analysis. It is however not sufficient to specify a global limit on $E_{T,Miss}$, since the $E_{T,Miss}$ resolution increases with the total amount of energy deposited in the calorimeter ($E_{T,Sum}$), which also causes a higher average $E_{T,Miss}$. The $E_{T,Miss}$ resolution generally behaves approximately proportional to $1/\sqrt{E_{T,Sum}}$. Thus, instead of $E_{T,Miss}$, it is convenient to use $E^{Sig}_{T,Miss}$, defined as:

$$E^{Sig}_{T,Miss} = \frac{E_{T,Miss}}{\sqrt{E_{T,Sum}}}. \tag{7.3}$$

The remaining energy dependence is accounted for by parameterizing the $E^{Sig}_{T,Miss}$ threshold as a function of the leading jet's p_T. The maximum allowed $E^{Sig}_{T,Miss}$ is extracted from MC as the value where the selection's inefficiency becomes negligible, namely below 1 ‰. Two exemplary histograms used to derive these values are displayed in Fig. 7.1. The agreement between data and QCD MC is excellent in the bulk of the distributions, which

* *Real* $E_{T,Miss}$ is referred to as the energy invisible to the calorimetry. Namely this is energy carried by neutrinos or muons.

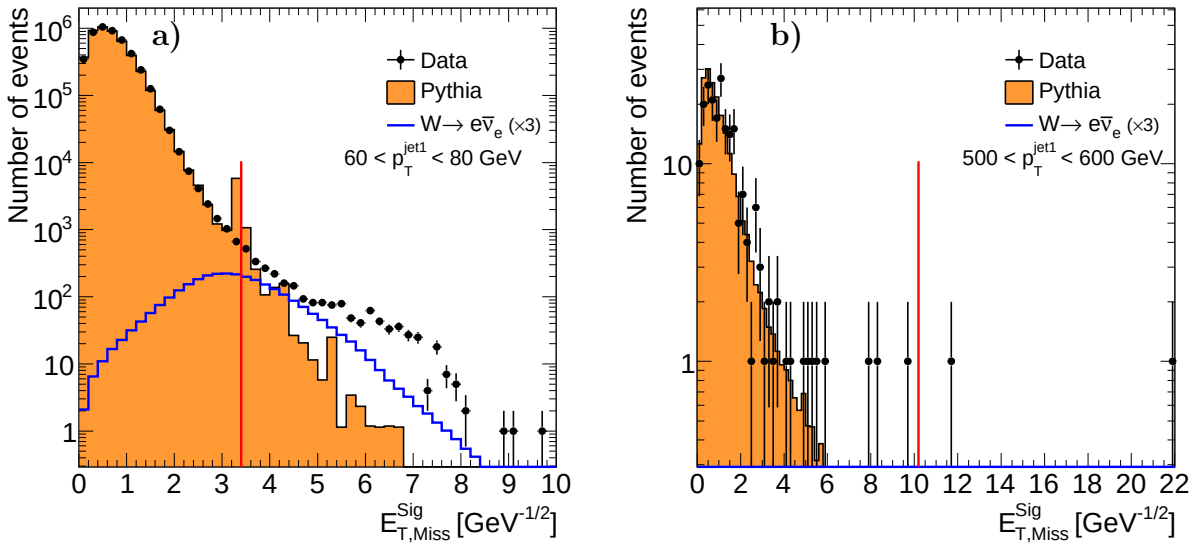

Figure 7.1: *$E^{Sig}_{T,Miss}$ distributions for two exemplary leading jet p_T slices. To account for contributions from events with real $E_{T,Miss}$ $W \to e\bar{\nu}_e$ MC is overlayed with an increased cross section in order to mimic the expected contribution by $W \to \mu\bar{\nu}_\mu$ and $W \to \tau\bar{\nu}_\tau$. The red line indicates the maximum allowed $E^{Sig}_{T,Miss}$ value.*

makes the use of MC legitimate. Only at the very high $E^{Sig}_{T,Miss}$ values data starts to differ significantly from MC. The peak structure at $E^{Sig}_{T,Miss} \approx 3.4$ is due to single MC events with a high weight. Not the full difference between data and MC in these histograms can be accounted to contaminations by non-collision backgrounds: At this point the QCD MC (filled histograms in Fig. 7.1) does not contain the contribution from events with real $E_{T,Miss}$ like events containing $W \to l\bar{\nu}_l$ decays, whose $E^{Sig}_{T,Miss}$ distribution is overlayed as well. The result of the complete procedure is shown in Fig. 7.2, which overlays the cut-value as a function of leading jet p_T on top of the $E^{Sig}_{T,Miss}$ distribution seen in data. The cut applied in the analysis ranges from ≈ 2.5 to ≈ 15, for the events containing the highest p_T jets. The upwards rising tail in this figure can be accounted to events from, or contaminated by cosmic or beam-halo particles in the calorimeter. Two exemplary events being rejected due to a too high $E^{Sig}_{T,Miss}$ are shown in Appendix A.3.

It should be made clear that the major purpose of this selection is to reject cosmic events producing very energetic jets. In a very unfortunate scenario these could significantly contribute to the highest p_T bins in the analysis, where only very few jets will be counted. Nevertheless it does not harm to construct the selection in a consistent manner also for the low p_T events, where the fraction of rejected events can be expected to be negligible.

7.2 Jet selection

The event selection is followed by a selection of jets, which have to satisfy a set of quality criteria as well. They all aim at a rejection of jets that are heavily affected or purely originating from calorimeter defects, mainly such like sporadic noise bursts in certain detector regions. The quantities used to identify such jets are the following:

- **Jet time $\mathbf{t}_{\mathrm{jet}}$**: The jet time quantity is computed as the absolute value of the energy weighted mean of the constituent cell's time information, which is measured relative

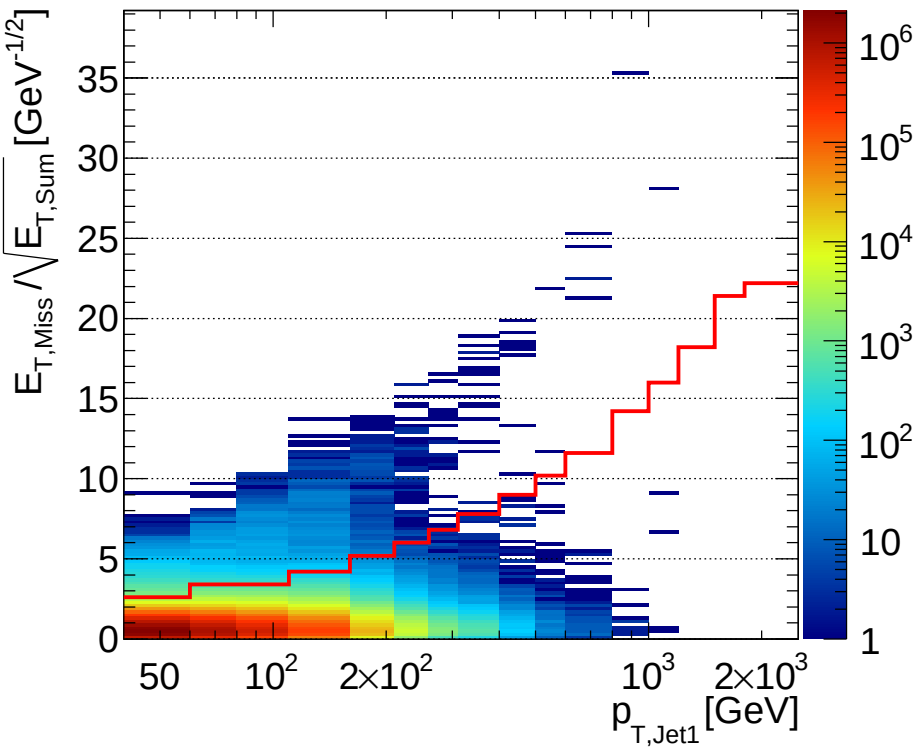

Figure 7.2: *Distribution of the observed $E_{T,Miss}^{Sig}$ values as a function of the leading jet p_T and the maximum allowed value (red line).*

to the nominal event time. This quantity allows an identification of cosmic jets, since their impact time is uncorrelated to the bunch crossing time.

- **N_{90}**: This is the minimum number of cells containing 90% of the jet's energy. Very small values of N_{90} hint to jets made up of few, noisy or hot calorimeter cells since the shower of a proper jet can be expected to spread its energy over a large amount of jets.

- **f_{HEC}, f_{EM}, $f_{TileGap3}$**: These are the fractions of the jet's energy deposited in the hadronic end cap, the electromagnetic calorimeters and the tile gap scintillators (the *cryostat scintillators* shown in Fig. 5.6) respectively. All three quantities are based on the mapping of the constituent cells to the various subdetectors. For f_{EM}, values too close to zero hint to jets made up of noise in the nearby hadronic calorimeters. Values close to 1 for f_{HEC} or $f_{TileGap3}$ point to problematic jets which deposited only very few energy outside these two regions.

- **Jet quality Q_{jet}**: The jet quality is the fraction of energy contributed to the jet by cells whose pulse shape was found to differ significantly from the predicted one. Pulse shapes seen in hot or noisy calorimeter cells are unphysical, thus this fraction becomes significant for jets heavily affected by such cells. Besides these effects, the pulse shapes also depend on the energy deposited in a cell, thus also high energy jets tend to have $Q_{jet} \neq 0$. This energy dependence is not yet modeled in the MC, as can be seen in Fig. 7.3. For this reason the cuts on this quantity will be chosen in a purely data-driven manner to avoid a rejection of high p_T jets.

These quantities are used to form cuts to reject mainly jets that are heavily affected or purely made from noise. These cuts were derived by the ATLAS jet performance working group [70] and are outlined here for completeness.

The jet time offers the first possibility to further reject jets from non-collision backgrounds since their time will not correlate with the bunch crossing time. Figure 7.3 a) shows the absolute value of the jet time in data compared to MC. The timing cut applied was chosen very loosely to $t < 50$ ns which is the time between two bunch crossings. The MC simulates the timing resolution in the calorimeter reconstruction while signals that are in fact delayed with respect to the nominal collision time are not simulated.

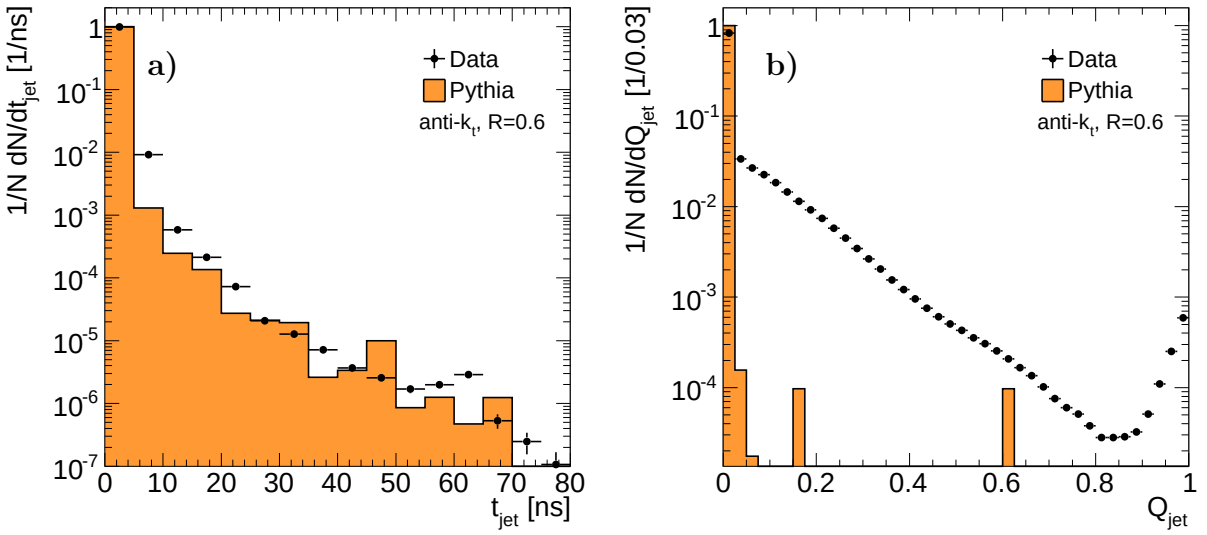

Figure 7.3: *The two distributions compare the jet timing and jet quality variable compared in data and MC. The modeling of the jet quality in the MC does not take increasing deviations from the nominal pulse shapes into account.*

Numerous noise jets were found to occur due to coherent noise in the electromagnetic calorimeter. Looking at Fig. 7.4 these jets show $f_{EM} \approx 1$. Since they are faked by noise in the em. calorimeter there is no associated energy deposition in the hadronic calorimeter. Since especially low p_T jets also tend to be absorbed completely in the em. calorimeter the cut designed to reject such jets also uses Q_{jet} as illustrated in the figure. Though the MC does not describe the Q_{jet} variable at all, as demonstrated in Fig. 7.3, the inefficiency due to is considered negligible due to the placement of the cut close to the boundary of the distribution.

Besides the electromagnetic calorimeter mainly the hadronic endcap is the source of noise jets which occur due to noise bursts. A first possibility to indentify these jets in the hadronic endcap is N_{90}. Since the cells in the hadronic endcap are comparably large and the noise frequently appears only in isolated cells, jets faked by this problem frequently are built of only a single cell, as can be seen in Fig. 7.5. The distributions show a wide distribution along f_{HEC} and a peak at $f_{HEC} \approx 0$ which is due to jets that are outside the fiducial region of the hadronic endcap. The simulation in this case predicts only a very tiny number of real jets removed by this cut.

Jets built from noise that spreads over several cells and real, soft jets overlayed on noisy cells will however make it past the N_{90} cut. These remaining, often high p_T jets can be identified using the $Q_{jet} - f_{HEC}$ plane where they appear in the region of $Q_{jet} >$

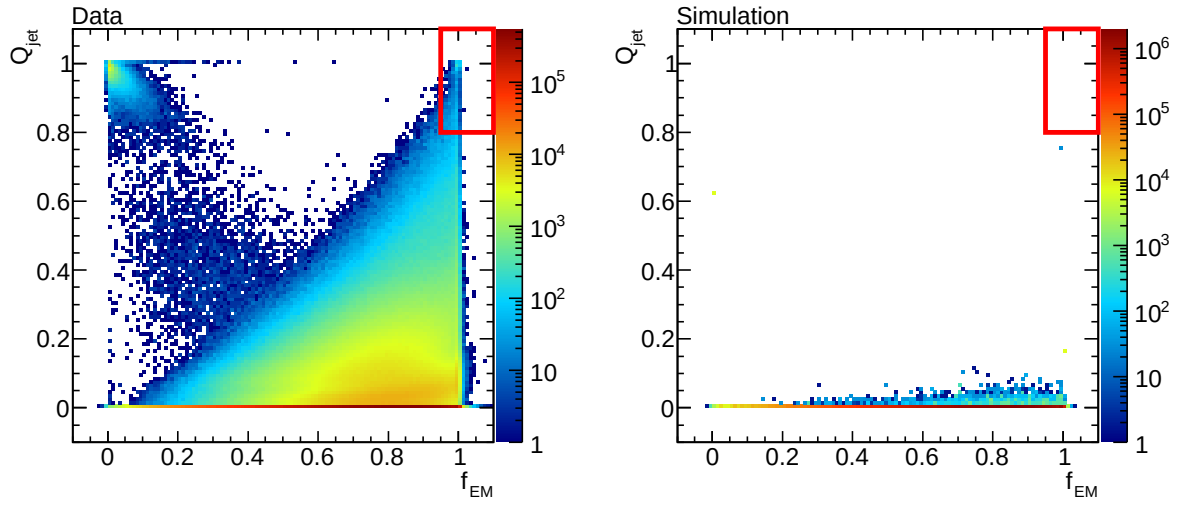

Figure 7.4: *The cuts indicated by the red boxes reject jets that are faked by coherent noise in the electromagnetic calorimeter. The simulated distribution does not model the Q_{jet} variable, so no jets are rejected by this cut in the MC.*

0.2 & $f_{HEC} > 0.5$ as shown in Fig. 7.6. The cut designed to remove these jets follows the boundary of the distribution along $f_{HEC} = 1 - Q_{jet}$.

Using all the cuts described as above it is believed that jets from noise are sufficiently well rejected from the analysis since the bulk of the remaining jets follows the simulation well.

Two last jet selection cuts are not designed to remove fake jets but rather jets that are known to be badly measured: For dead cells in the calorimeter the reconstruction software employs an algorithm that estimates the dead cell's energy using the information in neighboring cells. While this algorithm cannot harm in general, there are pathological cases where this correction amounts to a very significant fraction of the jet's energy due to unphysical extrapolations from small neighboring cells. Jets where this correction, denoted by f_{Corr}, contributes more than half of the jet's energy are rejected.

Finally jets that deposited more than half of their transverse momentum in the cryostat scintillators are excluded from the analysis. The calibration of these very small scintillators is not completely understood and thus these jets are rejected from the analysis as a safety measure.

The fraction of jets removed by these last two cuts is negligible as can be estimated from Fig. 7.7. To summarize, there are three cuts removing the major source of fake jets, being noisy jets originating in the electromagnetic calorimeter and the hadronic endcap. The timing cut is used to reject remaining jets from non-collision backgrounds. The two lastly mentioned cuts remove not well understood jets. All five cuts and their purpose are documented in Table 7.2.

In conclusion Table 7.3 summarizes the number of events and the number of jets in all subperiods analyzed after the various event cuts mentioned in this section. An important detail on the observed number of $R = 0.4$ and $R = 0.6$ jets in Table 7.3 is, that the ratio of the two numbers is not constant. This is due to the fact that this ratio is p_T dependent and in later periods the low p_T component has a less stronger impact, since the low threshold triggers were becoming increasingly heavy prescaled. This changes the raw p_T spectrum and hence also the ratio of the two numbers.

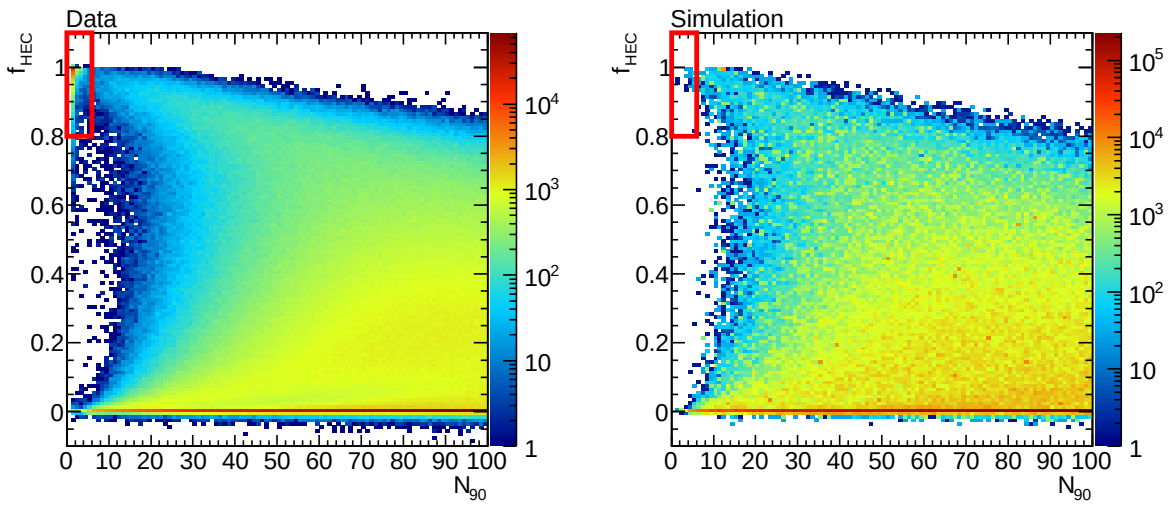

Figure 7.5: *Distribution of the jet's hadronc endcap energy fraction and N_{90}. Jets faked by noise bursts in the hadronic endcap can be removed by rejecting jets in the indicated regions of f_{HEC} and N_{90}.*

Figure	Cut	Designed to remove:
7.3	$t_{jet} > 50$ ns	Non-collision backgrounds.
7.4	$Q_{jet} > 0.8$ & $f_{EM} > 0.95$	Jets built from coherent noise.
7.5	$f_{HEC} > 0.8$ & $N_{90} \leq 5$	Jets faked by noise in the hadronic endcap.
7.6	$f_{HEC} > 1 - Q_{jet}$	
7.7	$f_{TileGap3} > 0.5$ $f_{Corr} > 0.5$	Badly measured jets.

Table 7.2: *Summary of the cleaning cuts applied to reject jets faked by either calorimeter noise or non-collision backgrounds.*

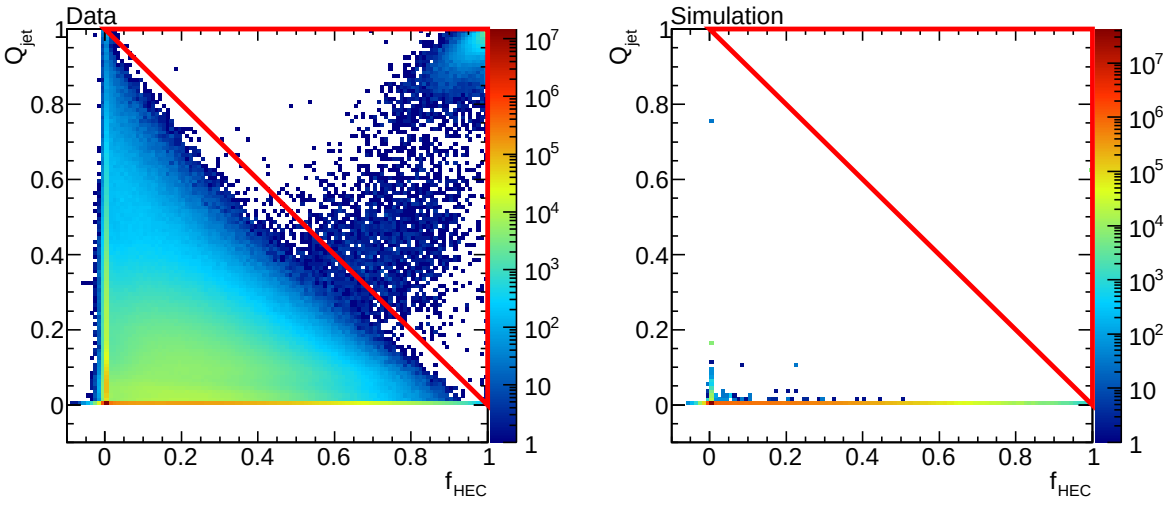

Figure 7.6: *The jet quality and f_{HEC} distributions provide the second cut designed to remove noisy jets within the hadronic endcap. These jets can be assumed to consist solely of noisy cells in the HEC.*

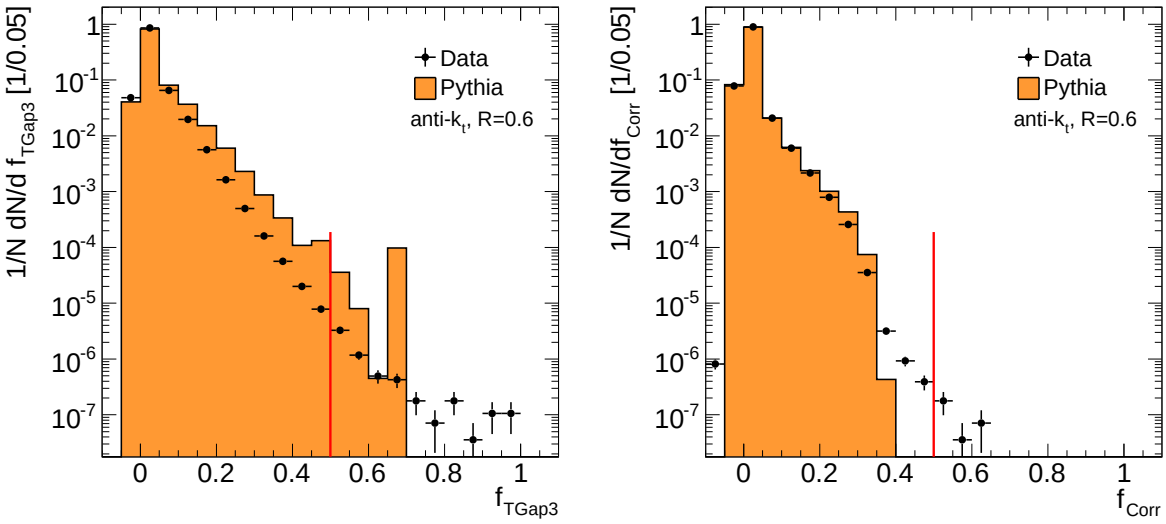

Figure 7.7: *In contrast to the other jet cleaning cuts, these cuts are more targeted to remove jets that are inherently not well measured. As implied by the distributions shown, these cuts remove only a negligible amount of jets.*

Period	Total	Good Runs List	Vertex Cuts	Trigger	Good $E_{T,Miss}^{Sig}$	One jet 0.6	One jet 0.4	#Jets 0.6	#Jets 0.4
B	74991267	57287618	19833997	23823006	19827612	240372	160724	64241	45304
C1	21521862	19146569	5190009	5936974	5186895	202693	136312	48336	34133
C2	2780083	2259464	449582	474098	449084	34177	23827	7199	5167
D1	31153947	22343989	3000708	3048861	2995224	614533	473412	51387	36764
D2	24567606	20365482	1678042	1718209	1674307	678674	549244	26452	19358
D3	17102149	14610059	1727611	1803464	1724934	665863	542192	24470	18225
D4	22768613	18865759	2837333	2926424	2832592	1583470	1303147	42170	31953
D5	7755715	7497542	1070955	1111756	1069598	563638	462335	15548	11782
D6	12477402	10980771	2946681	2998836	2942755	1960138	1607456	45699	34976
E1	8946127	8294211	2111562	2287076	2107767	1016944	900341	26849	20519
E2	6522403	6246728	2922037	3041064	2919338	1757108	1456395	43295	33084
E3	15191406	13305658	4172216	4438376	4167843	2266878	1959816	57167	43904
E4	6086324	3996077	706096	770691	705019	339824	328966	8982	6967
E5	5843108	5330050	969506	1052860	968182	475983	462409	12689	9757
E6	6291442	6018100	1023523	1092007	1022013	571823	557436	14595	11430
E7	8099817	7075882	1484400	1597531	1482834	766897	717535	19898	15365
F1	14013603	12845314	2297599	2470831	2293822	1483530	1455698	35973	28279
F2	26297734	24061912	2969102	3189207	2964476	1628349	1587815	37947	29510
Total	312410608	260531185	57390959	63781271	57334295	16850894	14685060	582897	436477

Table 7.3: *The number of events or jets after the several selection steps. The columns two to five display the numbers of events retained after the according event selection step. The following two columns show the numbers of events containing at least one good jet (passing the jet selection criteria) above with* $p_T > 40$ GeV *and with* $|y| < 2.8$*, with requiring the trigger responsible for this jet. The last two columns count the number of good jets, which fulfill the kinematic requirements and where the event fulfills the appropriate trigger requirements. Only jets counted in the last two columns enter the analysis.*

7.3 Comparisons of data to Monte Carlo simulations

The aim to measure the inclusive jet cross section will involve making corrections to the data that are derived purely from MC. This is only a valid approach if the simulation models the data sufficiently well. Whether this is the case shall be examined in this section, where uncorrected jet distributions in data are compared to MC. For this purpose only events and jets are used which pass the selections described in the previous section.

7.3.1 Topological clusters

As outlined earlier the input needed for the jet reconstruction are the topological calorimeter clusters formed prior to the jet finding. Properties of these inputs are thus the first characteristics that can be compared in the simulation and in data. Figure 7.8 a) shows the pseudorapidity distribution of clusters with $p_T^{cl} > 5$ GeV, contained in jets. This distribution shows the characteristic shape for quantities with a requirement of a minimum transverse momentum: These usually show a rather flat shape in the center of the detector ($|\eta| \lesssim 1$) where the $|p_T|$ to $|p|$ conversion is still moderate. For higher $|\eta|$ the production of objects above a p_T threshold is increasingly strong suppressed. High transverse momenta of 40 GeV e.g. correspond to total momenta of more than 1 TeV at $|\eta| \approx 4$. The distribution is reasonably well modeled by the simulation, with small differences around $|\eta| \approx 2$. The p_T distribution of clusters in jets is displayed in Fig. 7.8 b), summed over all jets within $|\eta| < 2.8$ and shows an excellent agreement with the simulation over six orders of magnitude, up to the highest cluster energies of 200 GeV seen in data. Sensitivity to the internal structure of jets is already provided by examining the number of cluster constituents N^{cl} in jets. This is done in Fig. 7.9 which counts the number of clusters with $p_T^{cl} > 1$ GeV in jets with $p_T > 40$ GeV. On average these jets in data have approximately 6.2 clusters, while the MC shows a slightly lower average of approximately 6.0 clusters per jet. But nevertheless the simulation describes data reasonably well, including the long tail to higher N^{cl}. This points to a good simulation of the fragmentation as well as a realistic description of calorimeter noise which could contribute noise clusters to jets.

7.3.2 Jet kinematics and internal structure

Having a solid understanding of the input quantities to the jet finding process, the MC description of the observables needed for the targeted measurement can be validated.

Figure 7.10 compares the angular distributions of observed jets to the MC prediction. For ϕ clearly a flat distribution due to the ϕ symmetry of the experimental setup is expected. Here Fig. 7.10 a) reveals that this is not perfectly true in data and also not in the MC simulation. Randomly distributed hardware defects such as defective read-out modules of the calorimeter may introduce inefficiencies in certain ϕ regions. Such defects are however also included in the simulation, so nevertheless a good agreement in these quantities can be expected. The ϕ distributions in data and MC show that these effects (e.g. around $\phi \approx -2$) are simulated quite well.

Less good agreement can be observed in the pseudorapidity distribution shown in Fig. 7.10 b). Most jets above a given p_T threshold are produced centrally, due to the same reasons mentioned for Fig. 7.8. Besides the general shape one remarkable feature in Fig. 7.10 b) is the slight excess of jets in data at $|\eta| \approx 2.0$. This excess, which would also be noticeable without a direct comparison to MC, can be attributed to an underesti-

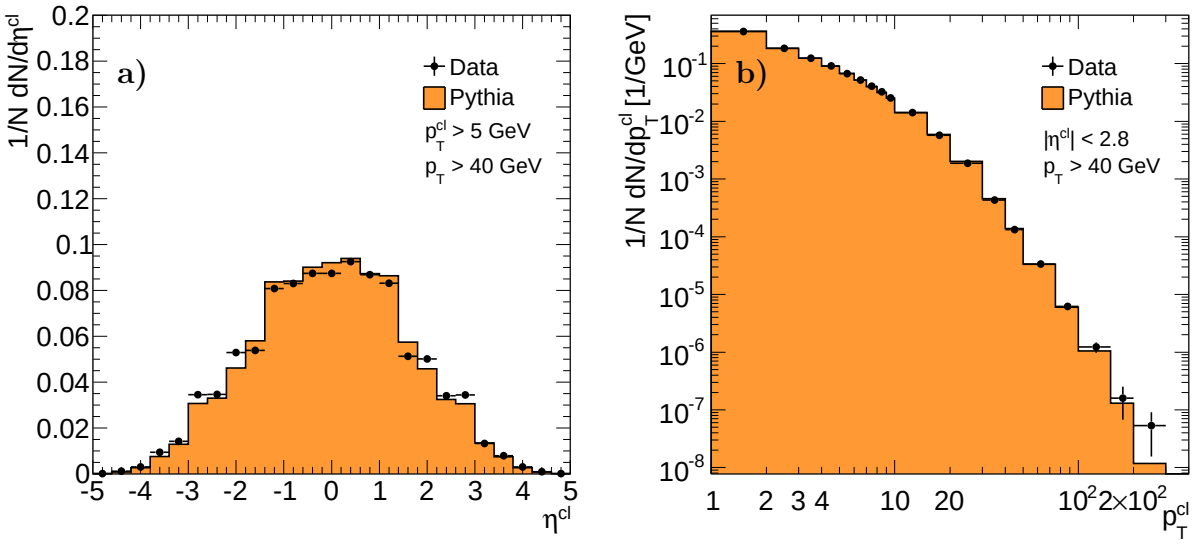

Figure 7.8: *Distributions of transverse momentum (b)) and pseudorapidity (a)) of topological clusters in jets.*

mated energy scale in this region in the MC. This assumption is backed up by so called intercalibration studies exercised by the ATLAS collaboration. These studies exploit the balancing in p_T in two-jet events, by using jets in a fixed detector region as a reference compared to the balancing jet in the detector region of interest. This way the relative energy scale between two calorimeter systems can be determined. Figure 7.11, taken from Ref. [71], shows a result of this study which reveals an underestimate of around 4 to 5% in energy scale. Taking into account the very steeply falling p_T spectrum of QCD jet production, this lifts a large amount of jets above the p_T limit of 40 GeV used in Fig. 7.10 b). An approximation for the increase in the number of jets above a given threshold can be made using the knowledge that the jet p_T distribution roughly falls according to a power law. The change due to a shift in p_T can than be estimated using the derivative:

$$\frac{dN_{jets}}{dp_T} \propto p_T^{-A} \tag{7.4}$$

$$\rightarrow \frac{dN_{jets}}{d^2p_T^2} \propto -A \times p_T^{-A-1} \tag{7.5}$$

With A ranging from 5 to 6 in the observed jet p_T spectrum the difference in energy scale of approximately 5% can be expected to have an impact of roughly 25% to 30%, which agrees with the actual excess seen in Fig. 7.10 b).

To be able to identify eventually remaining hot or dead regions in the calorimeter, it is very helpful to look at the two-dimensional distribution of jets in η-ϕ-space. This is done in Fig. 7.12. Here the detector simulation shows a very good agreement in the modeling of the few dead calorimeter regions. Only three small regions were not yet included in the simulation since these regions were functioning well at the start of data taking. Besides, the map for data also reveals the features that could be observed in Fig. 7.11 like the dip around $\phi \approx -2$.

Finally, Fig. 7.13 compares jet p_T distributions seen in data and MC. For this purpose jets not selected by the highest inclusive jet trigger have to be scaled according to the corresponding trigger, which can be read off Table 7.1. The number of jets quoted this

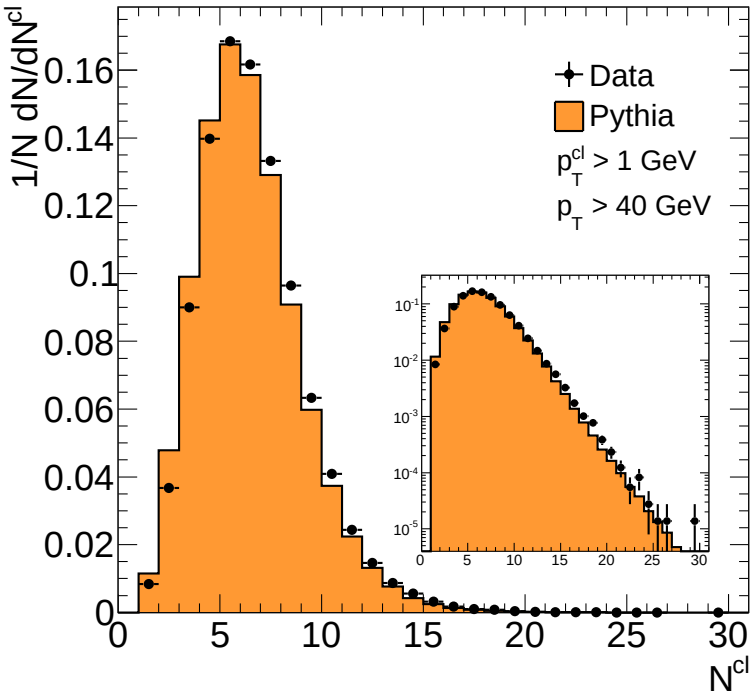

Figure 7.9: *Distributions of the number of constituents per jet. The inlay shows the same distribution on a logarithmic scale.*

way represents the expected number of jets in a given p_T bin, assuming the trigger would not have been prescaled. In order to quantify the agreement in shape, the MC distribution is normalized to the integrated number of jets seen in data. The ratio between data and MC reveals that data produces up to 50 % more jets at a p_T close to 1 TeV. Rather than to a problem of detector simulation this points to a difference in physics simulated in the MC. These are due to the leading order simulation done by Pythia and is also strongly affected by the choice of the PDF, which is CTEQ6L1 for the `Pythia6` simulation with the MC09 tune.

Another interesting quantity to check is the width of jets, defined as:

$$\langle \mathrm{R_c} \rangle = \frac{\sum E_c \, \Delta R(\vec{j}, \vec{c})}{\sum E_c}, \tag{7.6}$$

where the sums run over all jet constituents with energy E_c and distance ΔR in η-ϕ-space to the jet axis. For a jet built with a radius of $R = 0.6$ the naively expected maximum width would thus amount to $\langle \mathrm{R_c} \rangle = R/2$ for an imaginary jet made of only two constituents separated by $\Delta R = 0.6$. This quantity provides a hint to the inner structure of jets, which probes the agreement of jet fragmentation in data and MC. A reasonable description of the jet substructure to a certain degreee also gives confidence in a similarly good simulation of jet energy response and resolution. This is due to the fact that resolution and energy scale can be expected to differ between narrow and wider jets, since the energy of wider jets is distributed among more particles. Figure 7.14 shows this quantity for jets with $R = 0.4$ and $R = 0.6$. The width for $R = 0.6$ is larger than for the more narrow jets, which is of course the expected behaviour. It is also interesting to note that while the jet resolution parameter used is 0.6 and 0.4 respectively the width is in

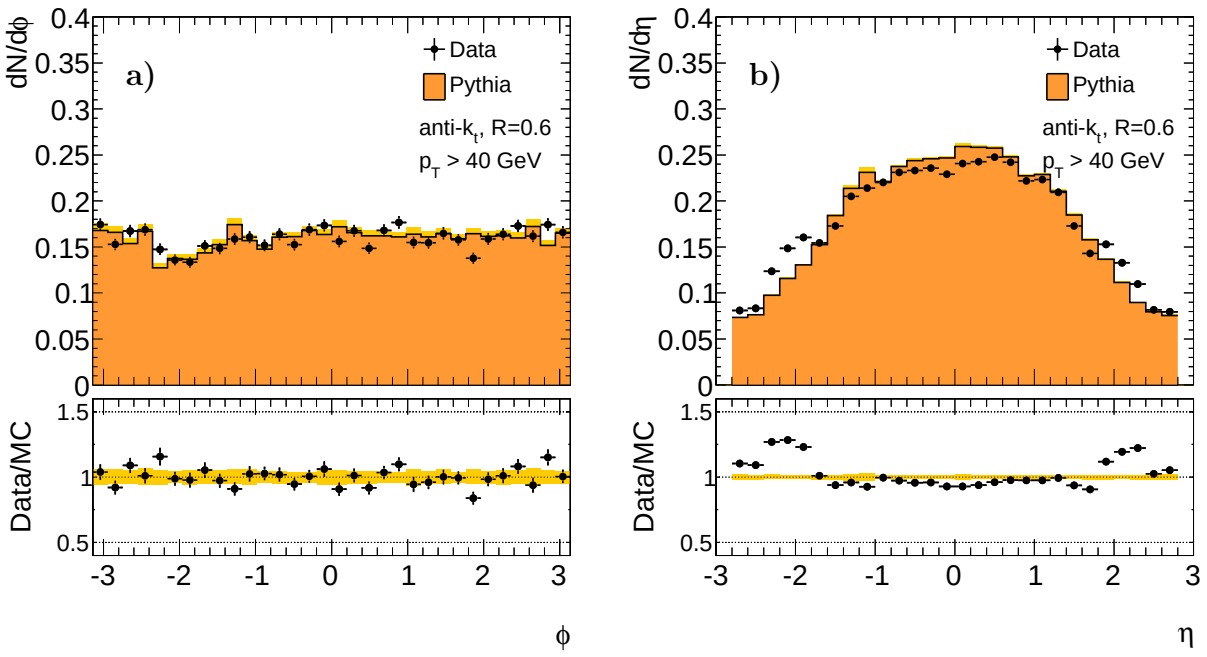

Figure 7.10: *Distributions of jet azimuth and pseudorapidities. The orange, filled area reflects the statistical uncertainty on the MC simulation.*

general significantly below 50% of the resolution parameter, which reflects that the radial distribution of energy within a jet is centered at its axis and drops off towards larger radii. Both jet figures reveal that jets in data are broader than the simulation predicts. This can be taken as feature of the specific MC tune, whose fragmentation can probably be improved in the future. While $\langle R_c \rangle$ itself does only provide information about the average, energy weighted distance of jet constituents, a slightly more sophisticated observable is the differential radial shape, which provides a more direct probe of the jet substructure. Differential jet shapes have already been measured by the ATLAS collaboration [72]. Its results, consistently with the ones presented here, are, that jets in data are slightly wider than predicted by the default MC tune.

7.4 Estimation of the impact of pile-up

As mentioned earlier the full dataset analyzed consists of runs taken over a period of roughly four months. During these four months the detector and much more the accelerator conditions changed over time. The most extreme development took place in the instantaneous luminosity provided by the LHC. The integrated luminosity as a function of the average number of pp-collision per crossing μ is displayed in Fig. 7.15. This shows that the dominant fraction of data was taken with μ relatively flatly ranging from 0.7 to 1.6 interactions per bunch crossing. A large fraction of luminosity increase is due to a higher number of protons in a single bunch and more importantly better collimation the bunches. Both effects result in an increase of the probability for more than one pp-collision taking place in one bunch crossing.

Along with the increase in μ, the probability that jets from any secondary collision coincide with a jet from the primary interaction is increased. Such coincidences add energy into a jet and thus on average increase the cross section at a given p_T. For each

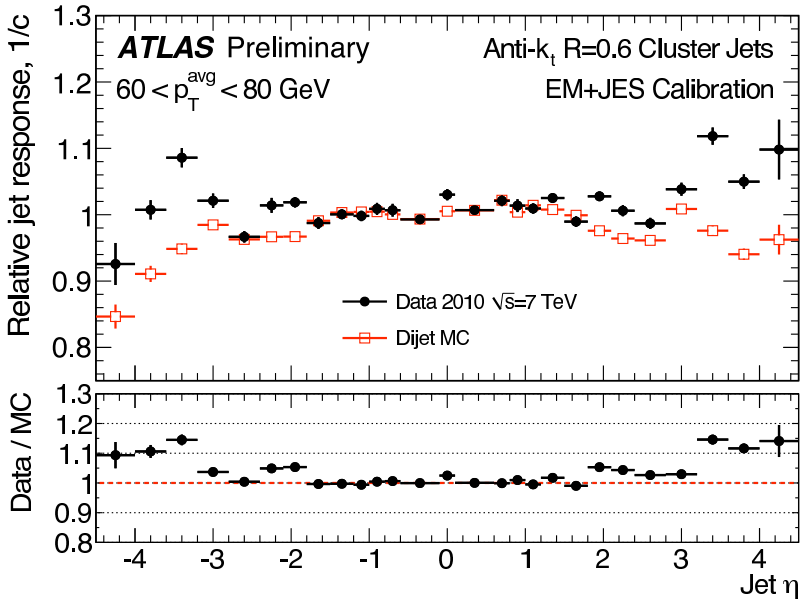

Figure 7.11: *Results of the η intercalibration studies [71]. Measured is the relative response with respect to two reference regions in the central detector.*

additional pp-collision, the energy brought into a jet E^{PU} from pile-up events is estimated as the average energy density in the calorimeter times the jet's area:

$$\left\langle E^{PU} \right\rangle = \rho(\eta, N_{pp}) \times \langle A_{jet} \rangle \,, \tag{7.7}$$

where $\langle A_{jet} \rangle$ is a measure of the jets area. This takes into account that the energy density ρ can be expected to change as function of pseudorapitidy. The relative impact of this effect can thus be expected to decrease for high p_T, since it will scale as $1/p_T$. It should be noted that this effect can in principle be corrected for, making use of a slightly modified version of Eq. 7.7, as described in [73]. During the early stage of data taking this extra calibration was not fully validated yet and for that reason is not used. The impact can nevertheless already be estimated to be very small, since the data used in this analysis was taken under moderate pile-up conditions with $\mu \lesssim 2$, as can be seen in Fig. 7.15. As reported in [73] this corresponds to E_T^{PU} of less than 1 GeV, which - for jets with $p_T > 40$ GeV- corresponds to less than 2.5% of their total transverse energy. This small but nevertheless existing contribution instead will be taken care of by an appropriate treatment within the systematic uncertainties.

7.4.1 Impact on jet rates

To assess the impact of this effect in a purely data driven way,

$$N_{jets}(\mu) = \frac{\sum N_{lb}^{jets}(\mu)}{\sum L_{lb}(\mu)/1 \text{ nb}} \tag{7.8}$$

is computed as the number of jets normalized by luminosity. Here $N_{lb}^{jets}(\mu)$ is the number of jets accumulated in a given luminosity block lb in a certain μ bin and $L_{lb}(\mu)$ is its integrated luminosity. The sums range over all luminosity blocks analyzed. One would expect that with increasing pileup the number of jets above a given p_T threshold increases

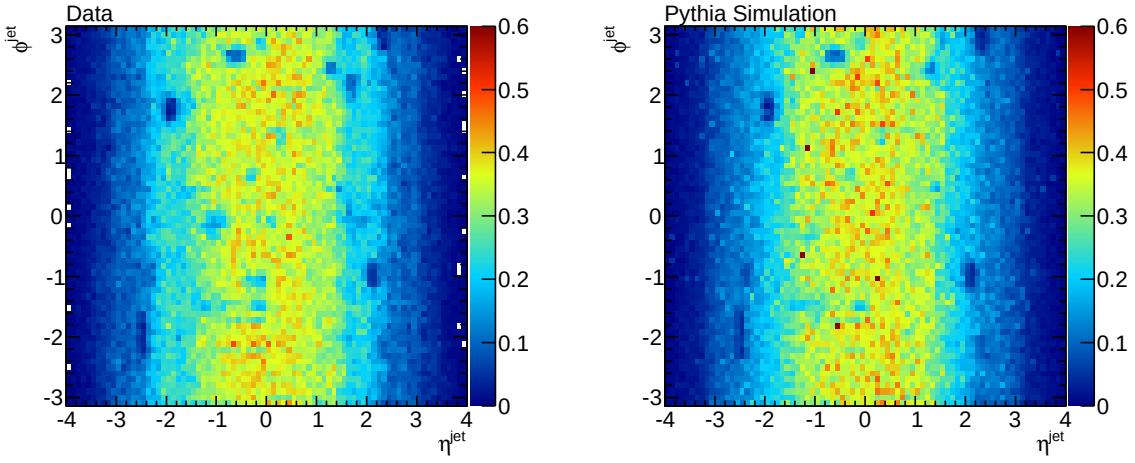

Figure 7.12: *η, ϕ distributions of jets in data (left) and simulated jets (right). Arbitrary units are used for the z-axes.*

with μ, since higher E_T^{PU} will carry more jets above the p_T threshold. To which degree this is the case is presented in Fig. 7.16 for a low and high p_T bin. The statistical uncertainties displayed are the uncorrelated components arising from $\sum N_{lb}^{jets}(\mu)$ only. The uncertainty assigned to $L_{lb}(\mu)$ is zero, accounting for the fact that the luminosity uncertainties are fully correlated. Consequently all data points share a common normalization uncertainty which is not of interest for stability considerations. The error bars shown thus reflect only the uncorrelated, purely statistical uncertainty on the number of jets counted. The result reveals that an actual increase of the jet rate in fact cannot be observed.

7.4.2 Jet width dependency on pile-up

Similar to the jet rate studies performed above, an impact of pile-up on the jet width can be expected too. Contributions from secondary pp-interactions can be expected to increase the average width of jets. Since the distribution of energy from minimum bias is uncorrelated with the primary scattering, the energy will be randomly distributed with respect to the jets in the event. Thus these contributions will increase the jet width, putting the approximately Gaussian energy profile onto an increasing background. Figure 7.17 displays this dependencies for small and large jets. Also shown is a straight line fit to the average width in every μ bin, which reveals that there is indeed a small slope towards increasing widths. The straight line fit also reveals, that the effect is significantly smaller for small jets, which can be expected due to the fact that a smaller jet area will of course collect less energy from secondary pp-interactions.

7.5 Determination of trigger efficiencies

In this section the efficiencies of the several triggers used in the analysis are derived. These efficiencies are the basis for the determination of the $p_{T,min}$ thresholds (cf. Table 7.1), which were already used in the previous sections.

To extract trigger efficiencies from data, the so called *bootstrapping* method is employed. Here the efficiency of a trigger B is determined using a second trigger A as a baseline, assuming that trigger A provides an unbiased sample of events. The trigger efficiency can

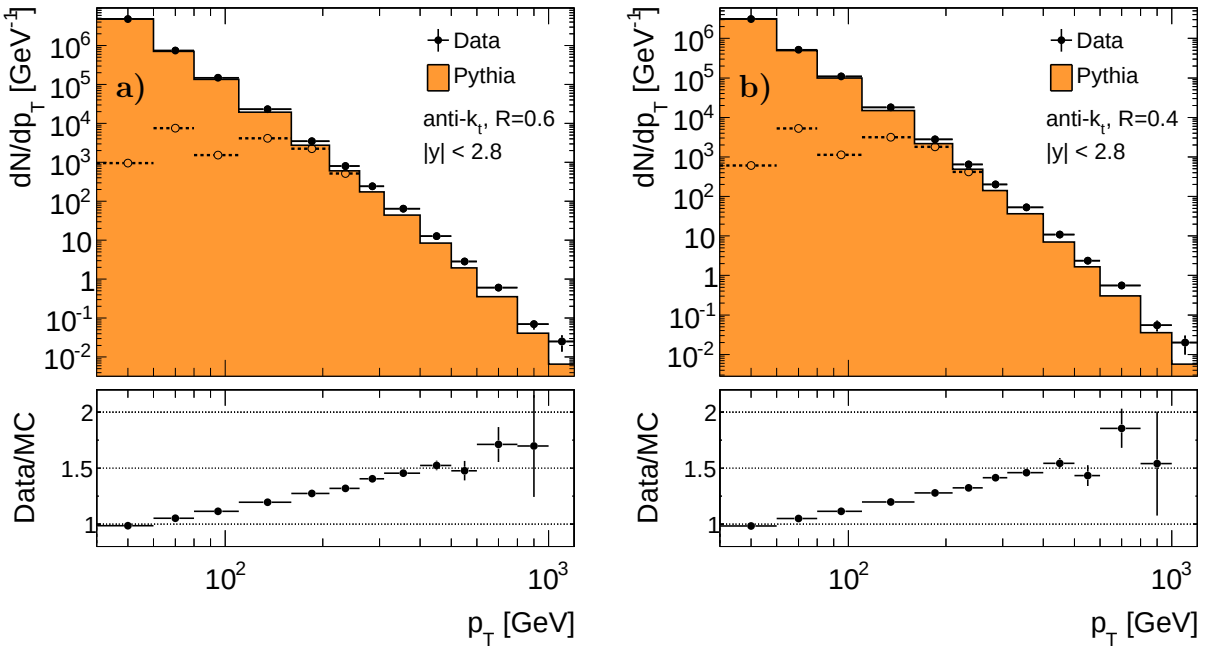

Figure 7.13: *Distributions of transverse jet momenta. The dashed histogram represents dN_{raw}/dp_T, with N_{raw} being the number of jets before correction of the average prescale of the corresponding trigger. The solid histogram represents the expectation without prescale.*

then be defined as

$$\varepsilon = \frac{N_{A\,\&B}}{N_A}, \tag{7.9}$$

where N_A is the total number of events delivered from trigger A and $N_{A\,\&B}$ is the number of events (among N_A) requiring the trigger B. Once the efficiency for trigger B is known it can be used as the baseline trigger for a next trigger of interest. In this section the efficiencies for the triggers listed in Table 7.1 are determined. The efficiencies are extracted as a per-event efficiency in bins of the leading jets' p_T and η.

For the lowest threshold trigger used in the analysis, L1_MBTS_1, the L1_ZDC trigger is used to select the sample of events on which the efficiency is calculated. The results in Fig. 7.18 show that this trigger is fully efficiency already at a jet p_T of 20 GeV across all rapidity regions. This qualifies this trigger as the safest possibilty to trigger low p_T jets and also makes it the best choice for a baseline trigger in the subsequent determination of the jet trigger efficiencies. A little benefit can be gained by using L1_MBTS_2 for the determination of the trigger efficiencies, as it provides a slightly larger event sample.

Due to the heavy prescales applied to the minimum bias triggers (≈ 5000) they can however only be used up to L1_J30 since the number of jets in the interesting p_T region for higher threshold triggers becomes very little. As a consequence L1_J15 is used as the baseline trigger for L1_J55. The efficiencies for all jet triggers are shown in Fig. 7.19 and Fig. 7.20. The results show relatively broad turn-on curves. Out of Fig. 7.19 the $p_{T,min}$ thresholds can be extracted. They determine the p_T value at which the trigger gets sufficiently close to 100% efficiency: L1_J5 reliably selects jets only above 60 GeV, which leaves the minimum bias triggers as the only possibility to select lower p_T jets in a fully efficient way. For the subsequent thresholds 110 GeV, 160 GeV and 260 GeV are chosen as the $p_{T,min}$.

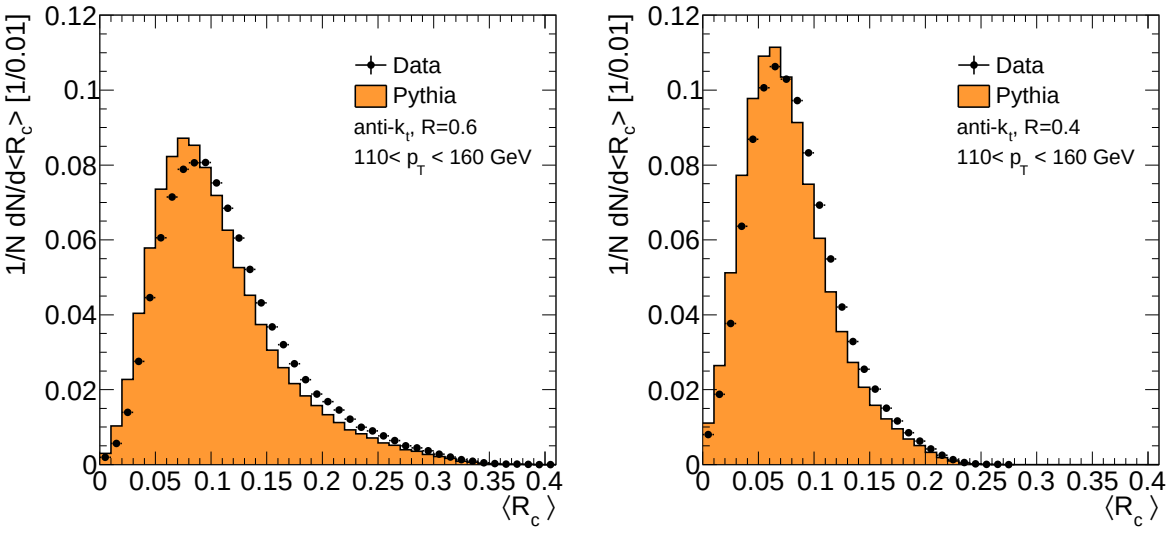

Figure 7.14: *Jet width as seen in data compared to the simulation.*

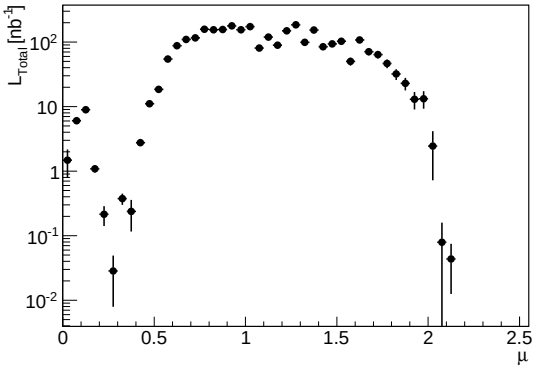

Figure 7.15: *The integrated luminosity in the dataset analyzed, as a function of average number of interactions per bunch crossing μ.*

The in general relatively large gap between the threshold value applied at the L1 trigger and $p_{T,min}$ can be attributed to the many limitations of the sliding window algorithm employed at L1 compared to the offline jet reconstruction. For instance the calibration of signals is only done at the calorimeter tower level. Thus it is not possible to use a dedicated jet calibration. In addition the area of $R = 0.6$ jets is approximately two (or seven, for L1_J5) times larger than the according window size at Level-1, which means that very significant amounts of energy are lost outside the Level-1 ROI window.

The comparison with MC in the above studies is done for completeness, though agreement is not a prerequisite for the analysis. One can observe that the plateau region in MC in general is reached a little earlier than in data. Table 7.4 summarizes the results of this section in form of the extracted $p_{T,min}$ thresholds.

The above results are also fully valid for the smaller jet sizes. The efficiences for $R = 0.4$ jets, which for brevity were not shown, are consistently higher than or equal to the efficiencies of the $R = 0.6$ jets. This is due to the different area of the jets, while the window size at the trigger level remains fixed. For a $R = 0.6$ jet the amount of energy deposited outside the ROI window is significantly higher than for the smaller jets. This

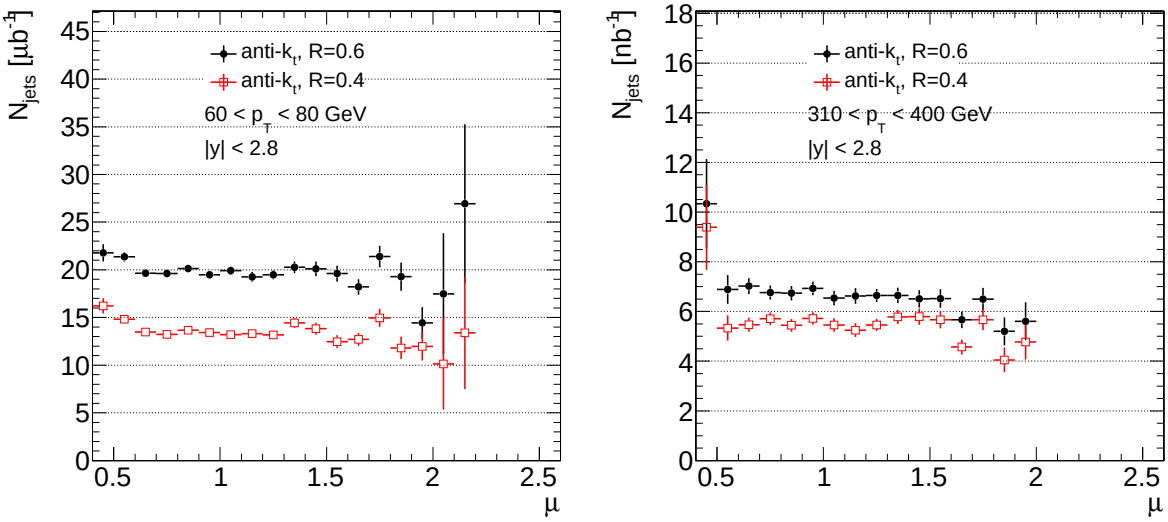

Figure 7.16: *Stability of the number of jets counted per* μb^{-1} *in two bins of* p_T *as a function of mean number of interactions per collision.*

Trigger	$p_{T,min}$ [GeV]	Baseline trigger
L1_MBTS_1	< 20	L1_ZDC
L1_J5	60	L1_MBTS_2
L1_J15	110	L1_MBTS_2
L1_J30	160	L1_MBTS_2
L1_J55	260	L1_J15

Table 7.4: *List of triggers used and the extracted minimum* p_T *for 100% efficiency. Also shown is the baseline trigger used to extract the efficiency.*

in turn results in a lower efficiency. As a consequence, the $p_{T,min}$ tresholds from Table 7.4 applied to small jets can be considered even more conservative.

7.5.1 Systematic uncertainties

As observed above, the determination and modeling of trigger efficiencies is not straight forward. The only question important for the upcoming analysis is, to which degree the assumption of negligible inefficiency above the $p_{T,min}$ thresholds is valid. This issue is addressed by a calculation of the trigger efficiencies in an alternative manner. Instead of a per-event efficiency the *tag & probe* method estimates the efficiency to trigger on a single jet within an event. For this purpose in each event the Level-1 ROI with the highest transverse energy is considered the ROI responsible for triggering the event and is called $\mathrm{ROI_{max}}$. Subsequently, for all jets in the event the closest Level-1 ROI is found. If the ROI is sufficiently close to the jet ($\Delta R(jet, ROI) < 0.4$) and if the ROI is **not** the $\mathrm{ROI_{max}}$, it is checked whether this ROI passed the trigger threshold of interest. If the closest ROI is the leading ROI, the jet is skipped and not counted in the number of total jets N^{jets}_{total}. The per-jet efficiency is then calculated as:

$$\varepsilon_{jet}(p_T) = \frac{N^{jets}_{ROIpassed}(p_T)}{N^{jets}_{total}(p_T)} \tag{7.10}$$

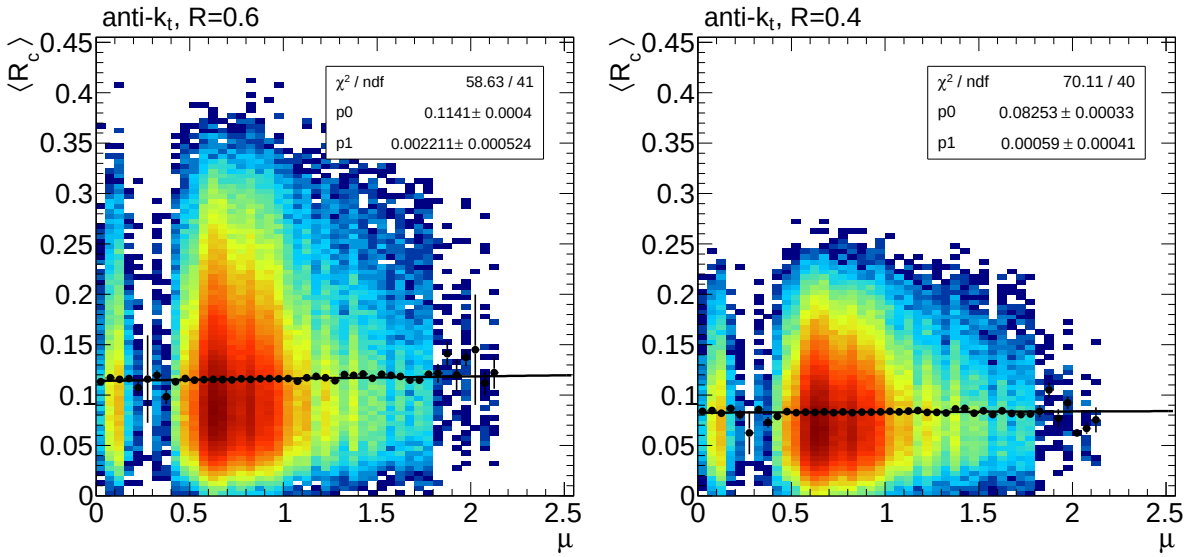

Figure 7.17: *Jet width as a function μ for large (left) and small (right) jets with* $160 < p_T < 210$ GeV. *Colors represent the abundance of jets, black dots show the mean along the y-axis for each μ bin. The means are fitted with a straight line of the form $f(x) = p_0 + xp_1$.*

In Fig. 7.21 this efficiency is compared in the region above $p_{T,min}$ for the various calorimeter triggers used. The observed differences are well below 0.2% in all but one rapidity region and for all calorimeter triggers used. The larger discrepancies appearing in the region of $1.2 < y < 2.1$ arise from a temporary trigger configuration subtlety: In this region the trigger system was configured differently during the subperiods E1 to E4. During these periods in a narrow pseudorapidity window only a small part of the calorimeter was connected to the trigger system, thus lowering the trigger efficiency. This effect naturally is much more visible in the per-jet efficiencies. This is because the average QCD event can be expected to contain at least two jets of similar p_T, which could be triggered on. Thus inefficiencies in certain detector regions are partly compensated for by additional jets in the event, when per-event efficiencies are studied. The event efficiency can in principle be written as a function of the per jet efficiencies of all jets in an event:

$$\varepsilon_{event} = 1 - \prod_i (1 - \varepsilon_{jet,i}). \tag{7.11}$$

As a consequence, the tag & probe per-jet efficiency by construction is smaller than (or equal to) the per-event efficiency calculated with the bootstrapping method. For an examplary event containing only two jets with $\varepsilon_{jet,i} = 98\%$, the efficiency to trigger the whole event would thus be 99.96%.

It should be emphasized that the trigger requirements in the analyses presented are done on a per-event basis. Per-jet efficiencies thus are an pessimistic estimate of the actual efficiency entering the analyses.

To summarize, these results back up the assumption of negligible inefficiency above the $p_{T,min}$ thresholds introduced. The inefficiency in the barrel to endcap transition region is accounted for by an 1% systematic uncertainty. This is considered a very conservative estimate given the excellent agreement of both methods in the remaining rapidity regions.

The minimum bias trigger used for the lowest p_T bin is considered to be fully efficient.

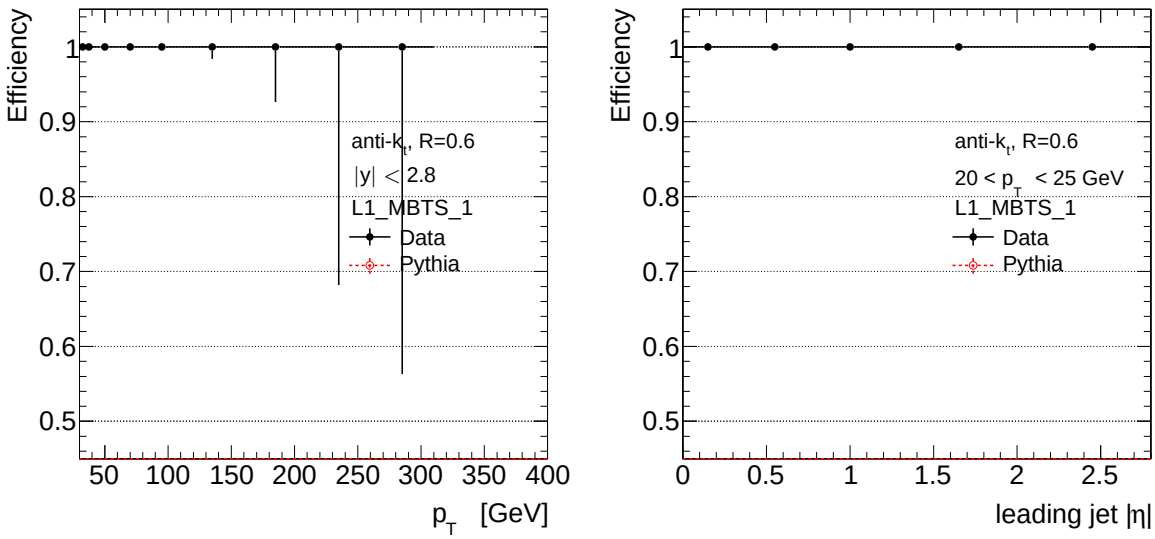

Figure 7.18: *Trigger efficiencies for the* L1_MBTS_1 *trigger, versus the transverse momentum of the leading jet in an event. As the ZDC trigger was not included in the simulation of the MC, this method cannot be applied to the MC.*

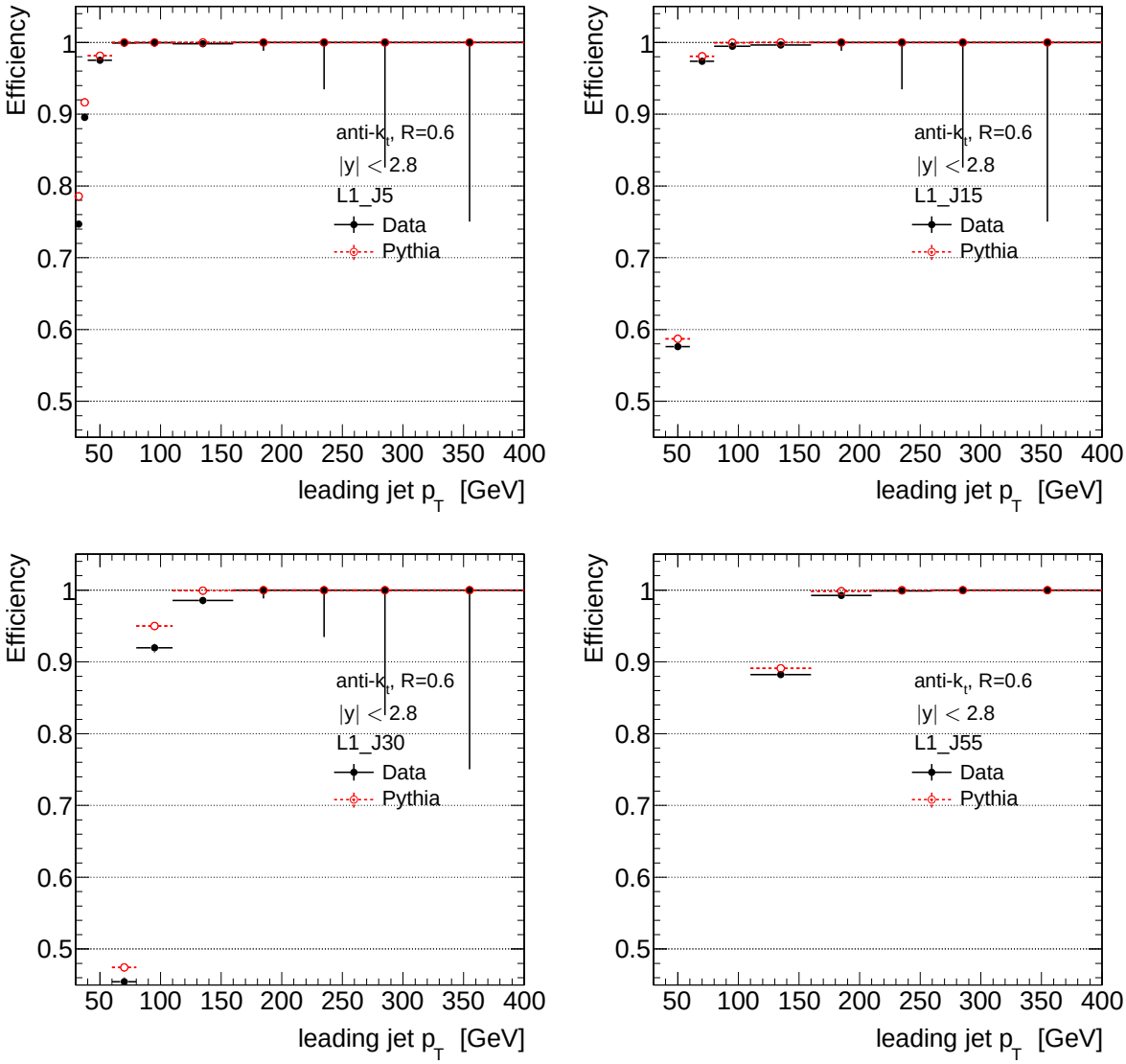

Figure 7.19: *Trigger efficiencies of the inclusive jet triggers versus the transverse momentum of the leading jet in an event.*

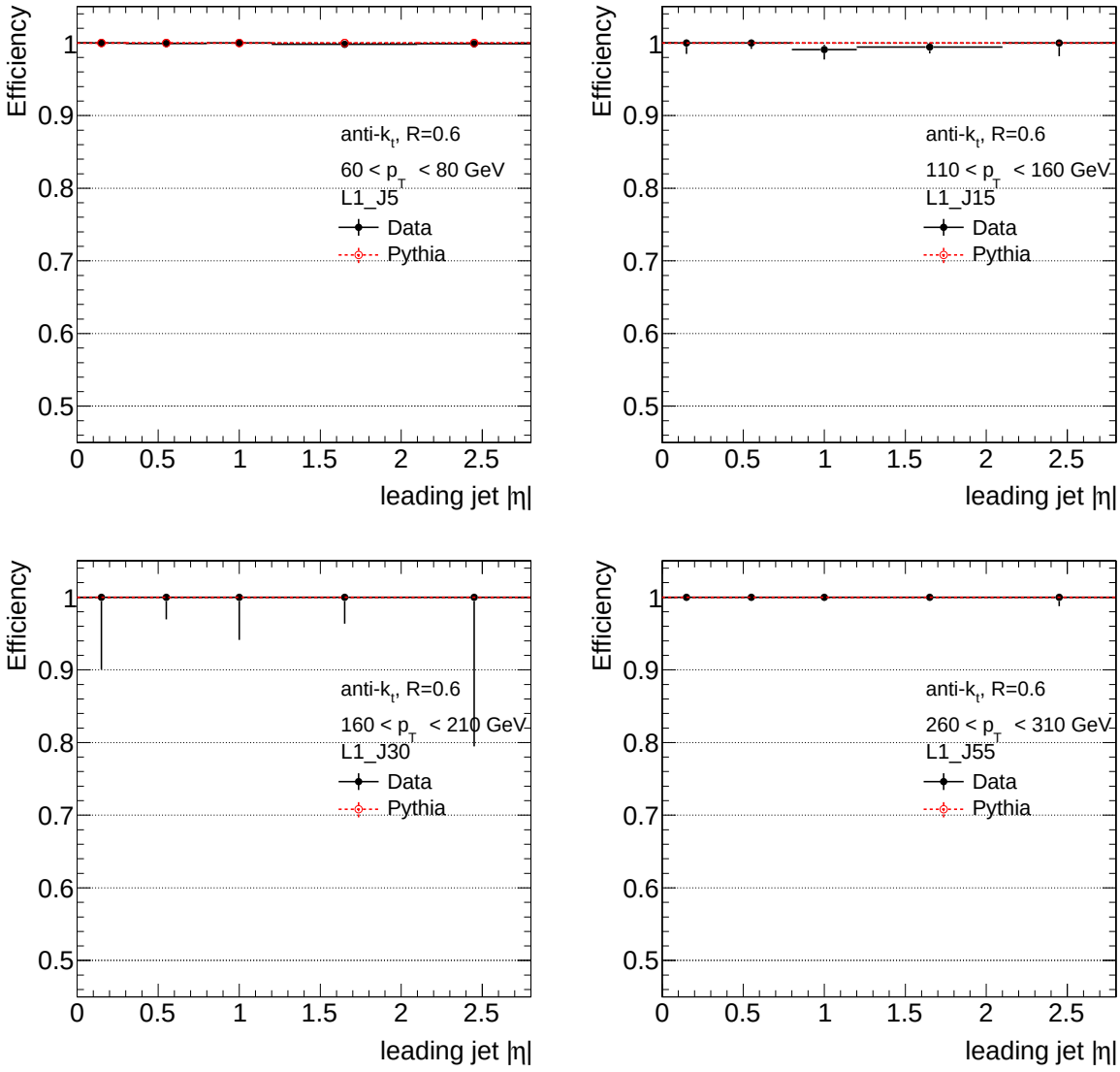

Figure 7.20: *Trigger efficiency of the inclusive jet triggers versus the leading jet's pseudorapidity.*

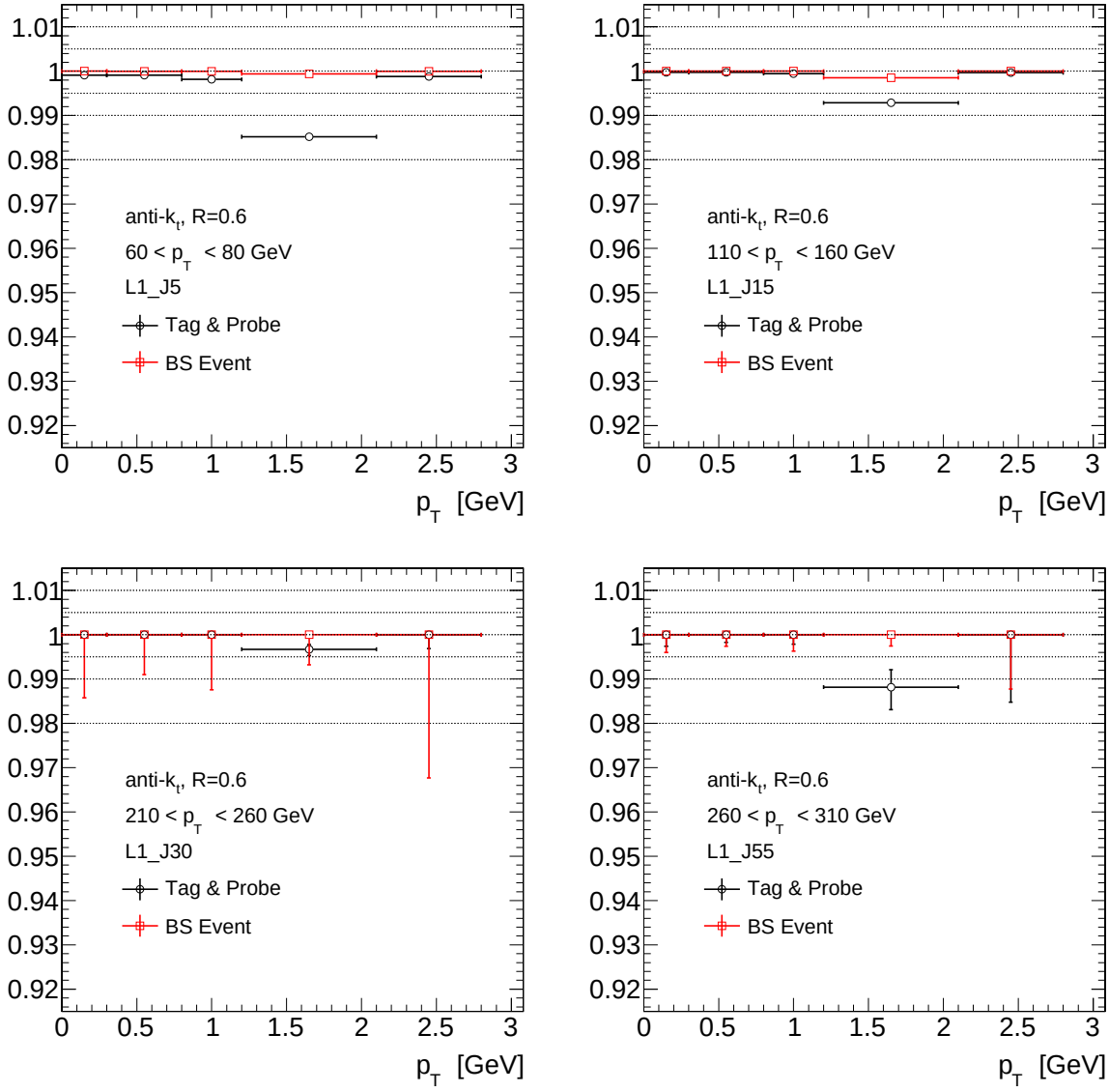

Figure 7.21: *Comparison of jet trigger efficiencies using the tag & probe method and the per-event efficiencies using the bootstrapping method, shown in Fig. 7.20.*

8 Jet energy scale and resolution uncertainty

This chapter will briefly outline the results of extensive studies carried out by the ATLAS collaboration to assess the systematic uncertainties of jet energy scale and jet energy resolution.

8.1 Jet energy scale uncertainty

The jet energy scale (JES) uncertainty is estimated using a mixture of MC studies and data driven methods where possible. To gain an estimate of the JES uncertainty, generally the jet response (e.g. as shown in Fig. 6.10) is studied for instance in MC simulations in which parts of the simulation affecting the jet energy scale are changed. The deviations Δ_{JES} from the nominal response $R_{nom}(p_T, \eta)$, defined as

$$\Delta_{JES}(p_T, \eta) = |1 - \frac{R_{var}(p_T, \eta)}{R_{nom}(p_T, \eta)}|, \tag{8.1}$$

[74], with R_{var} being the response in these varied scenarios, are taken as the systematic uncertainty.

The studies that will be referred to, are presented in [75]. One can divide the sources of uncertainty into two categories: Firstly, uncertainties related to the simulation of jets and their environment at generator level. And secondly, uncertainties related to the detector description and, most importantly, the models describing the interaction of particles with the calorimeter. The major contributions to both categories are examined in the following.

8.1.1 Uncertainties at generator level

The impact of various generator models and tunings is estimated studying the JES in accordingly varied MC samples. These include `Pythia6` MC with tunings that specifically vary the jet fragmentation and the underlying event contributions. Both aspects change the jet composition and shape, which can have an impact on the JES. The observed difference in Fig. 8.1 turns out to be small and does not exceed 1%. The impact of an entirely different MC generator was also studied using MC samples generated using the combination of `Alpgen`, `Herwig` and `Jimmy` [76–78]. Compared to `Pythia6`, these samples employ different techniques and models in every step of the event generation: Starting from `Alpgen`'s multi-parton matrix element calculation to the parton shower done by `Herwig` and the underlying event simulated by `Jimmy`. The impact on the JES is slightly more significant than the `Pythia6` tunes implied, namely $\approx 2\%$ below 60 GeV but below 1% for jets beyond 100 GeV.

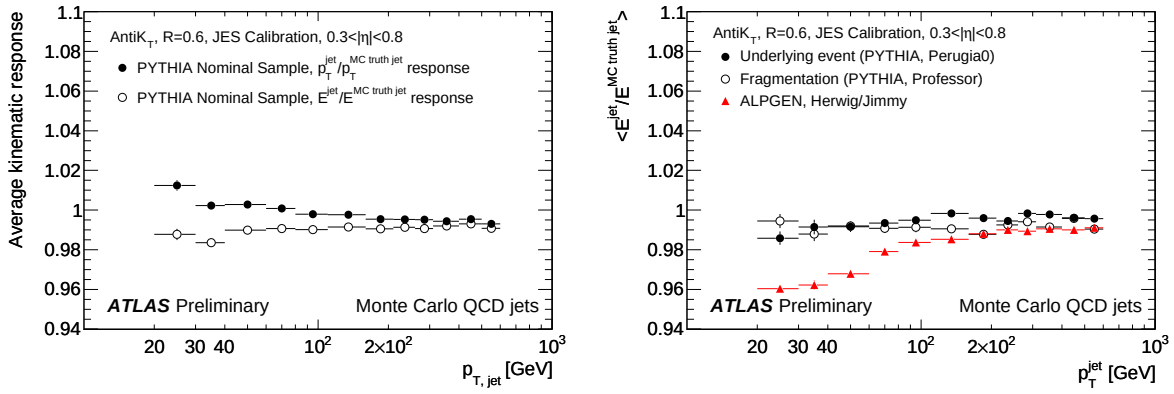

Figure 8.1: *Jet response in the nominal (left) and in various different MC samples (right) (figure taken from Ref. [74]). The differences observed between nominal and alternative MC are taken as the systematic uncertainty.*

8.1.2 Uncertainties in the detector simulation

The dominant source of uncertainty lies within the modeling of hadronic showers in the calorimeter (cf. Section 5.7). In [74] the differences in response using alternative physics lists were studied in the MC and with $\approx 6\%$ were found to be the dominant contribution to the total JES uncertainty.

The accumulation of more data in 2010 allowed for a significant decrease in the JES uncertainty with respect to first studies [74] by using data driven methods to assess these uncertainties. Instead of purely relying on MC, the JES uncertainty was updated in [75] using measurements of the response of charged hadrons [79]. This analysis attempts to mimic the testbeam situation by finding single, isolated charged hadrons using the tracking system. The tracks of these particles are extrapolated to the calorimeter and the energy E measured there can be probed against the momentum measurement p from the inner detector.

Measuring the average ratio of these two quantities, $\langle E/p \rangle$, in data and MC, one can quantify how well the calorimeter response is simulated. Exemplary results of these studies are shown in Fig. 8.2, revealing a very good agreement of data and MC on the level of approximately 3%.

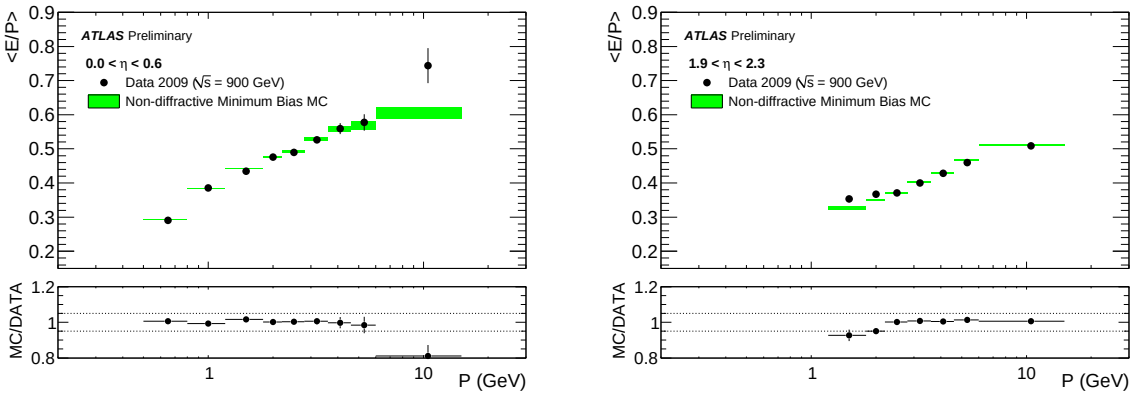

Figure 8.2: *Singe particle response as measured in data and MC. Figures taken from Ref. [79].*

The propagation from single particle response uncertainty to the JES uncertainty uses

pseudo-experiments [79]: The ATLAS simulation allows to map every energy deposit in the calorimeter to the truth particle that induced it. Making use of this functionality the total energy of a jet can be decomposed into the contribution from each particle. Each particle's energy deposit is then varied within the uncertainties appropriate to the particle species. The deviation of the response due to these variations is taken as the systematic uncertainty.

8.1.3 Final jet energy scale uncertainty

The above mentioned, major contributions to the JES uncertainty are summed appropriately including few additional, not mentioned contributions which are described in [75]. This results in the JES uncertainty displayed in Fig. 8.3 which shows that in the central detector region the JES uncertainty is around 4%. The uncertainty in the more forward region uses additional information from intercalibration studies which increases the uncertainty by approximately 1% for jets beyond 60 GeV.

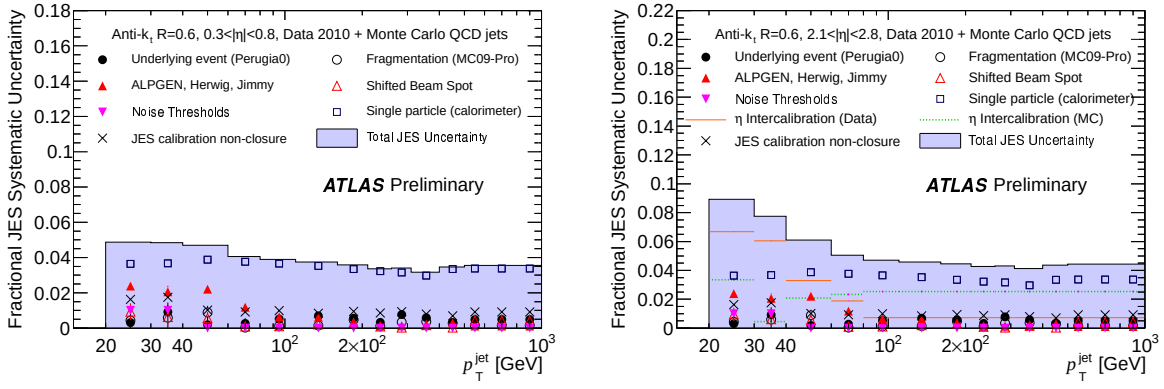

Figure 8.3: *Contributions of various sources (markers) to the total (filled histogram) JES uncertainty. Figure taken from Ref. [75].*

8.2 Jet energy resolution uncertainty

To which degree the jet energy resolution is well described in the simulation is evaluated in a direct comparison to the resolution measured in data. The studies summarized below are outlined in much more detail in [68].

The measurement of the jet energy resolution in data is exercised using two methods, each exploiting the p_T balance in clean two-jet events.

8.2.1 Resolution measurement using the two-jet asymmetry

The first method [80] makes use of the transverse momentum asymmetry in two-jet events, defined as

$$A = \frac{p_{T,1} - p_{T,2}}{\frac{1}{2}\left(p_{T,1} + p_{T,2}\right)}, \tag{8.2}$$

where $p_{T,1}$ and $p_{T,2}$ are the *randomly* ordered transverse momenta of the two leading jets. Under the assumption that both jets were produced with equal transverse momenta and

that they thus are subject to equal resolutions, one can write down the variance in A as:

$$\sigma_A^2 = \frac{\sigma_{jet}^2}{2p_T^2} \qquad \text{or :} \qquad \frac{\sigma_{jet}}{\bar{p}_T} = \sqrt{2}\sigma_A. \tag{8.3}$$

Hence, the relative jet energy resolution can easily be derived measuring the asymmetry distribution. Exemplary distributions are shown in Fig. 8.4, illustrating a good agreement between data and MC.

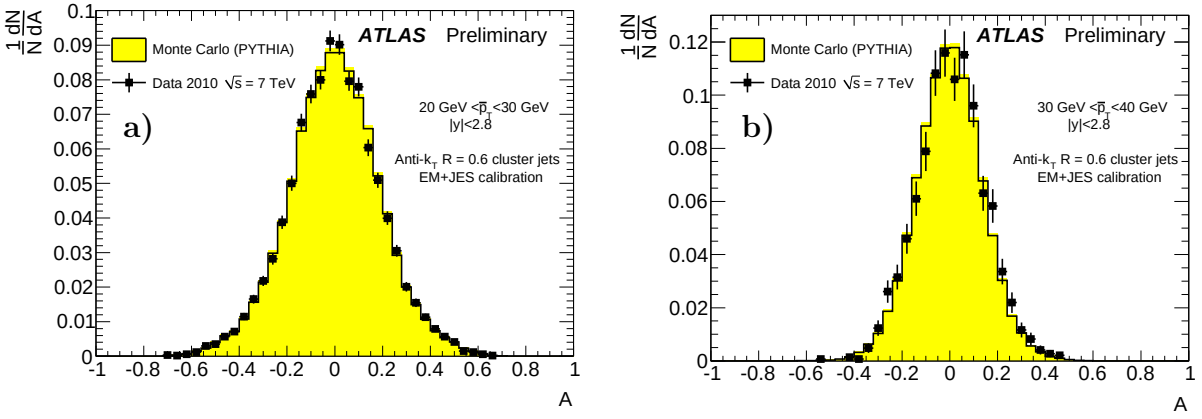

Figure 8.4: *Asymmetry distribution for two bins of the average p_T bins. Figure taken from Ref. [68].*

The imbalance introduced by additional jets in the event is corrected for by measuring σ_A using a varying veto against such jets: Events where there is an additional jet with $p_T > p_{T,3}^{cut}$ are rejected and $p_{T,3}^{cut}$ is varied between 5 GeV and 20 GeV. Finally each σ_A as a function of $p_{T,3}^{cut}$ is extrapolated to $p_{T,3}^{cut} \to 0$ using a straight line fit as displayed in Fig. 8.5.

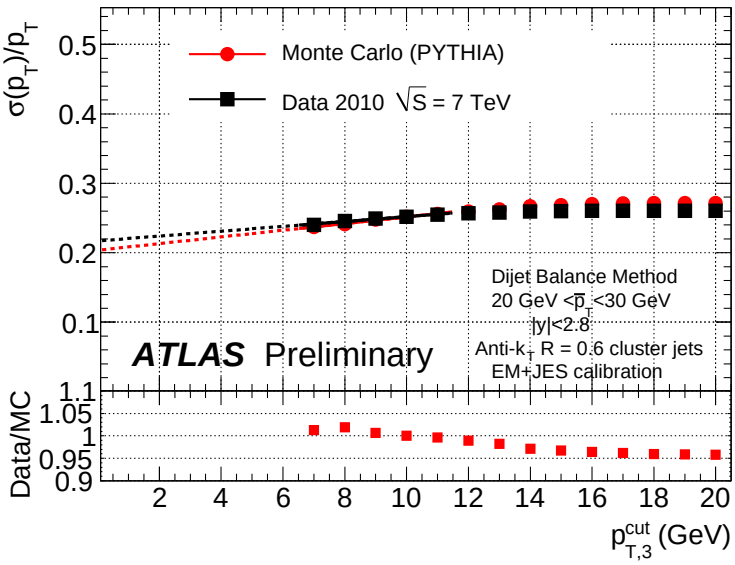

Figure 8.5: *Relative p_T resolution versus $p_{T,3}^{Cut}$. Figure taken from Ref. [68].*

8.2.2 Resolution Measurement using the bisector method

The second method [81] employs a slightly different strategy, which nevertheless also starts from the basic assumption of p_T balance between the two jets in a clean two-jet event. The sum of the transverse momenta of the two leading jets is split up in two distinct components: $p_{T,\Psi}$, along the bisector of the two jets (see illustration in Fig. 8.6) and $p_{T,\eta}$ orthogonal to it and thus approximately parallel to the jet axis. The idea behind this

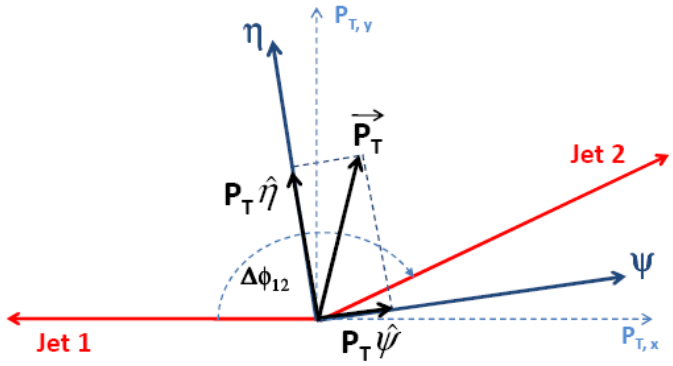

Figure 8.6: *Sketch of the construction of $p_{T,\eta}$ and $p_{T,\Psi}$. The η axis defines the azimuthal bisector between the two jets, which leaves the orthogonal Ψ axis approximately parallel to the two-jet-system. Figure taken from Ref. [68].*

splitting in two contributions is that fluctuations of the p_T balance at particle level will occur to equal parts in $p_{T,\eta}$ and $p_{T,\Psi}$. Imbalances due to mismeasurement of either of the two leading jets however will mainly increase the variance of $p_{T,\eta}$. It can be shown, that the difference between the widths of both quantities σ_Ψ and σ_η provides a handle to the actual jet p_T resolution:

$$\frac{\sigma_{jet}}{\bar{p}_T} = \frac{\sqrt{\sigma_\Psi^2 - \sigma_\eta^2}}{\sqrt{2}\bar{p}_T\sqrt{|\cos\Delta\phi_{12}|}} \tag{8.4}$$

The benefit of this procedure being that no explicit correction of soft radiation effects is necessary.

8.2.3 Combined results

The results of both methods are displayed in Fig. 8.7 and compared to MC. The interpretation of the results is limited by significant statistical uncertainties, but it can be concluded that differences between data and MC are on the order of 10 %. Meanwhile, the ATLAS collaboration has updated these studies using the full dataset of 2010, which amounts an integrated luminosity of approximately $33\,\mathrm{pb}^{-1}$. The reach in p_T in this study is much larger and the statistical precision has improve accordingly.

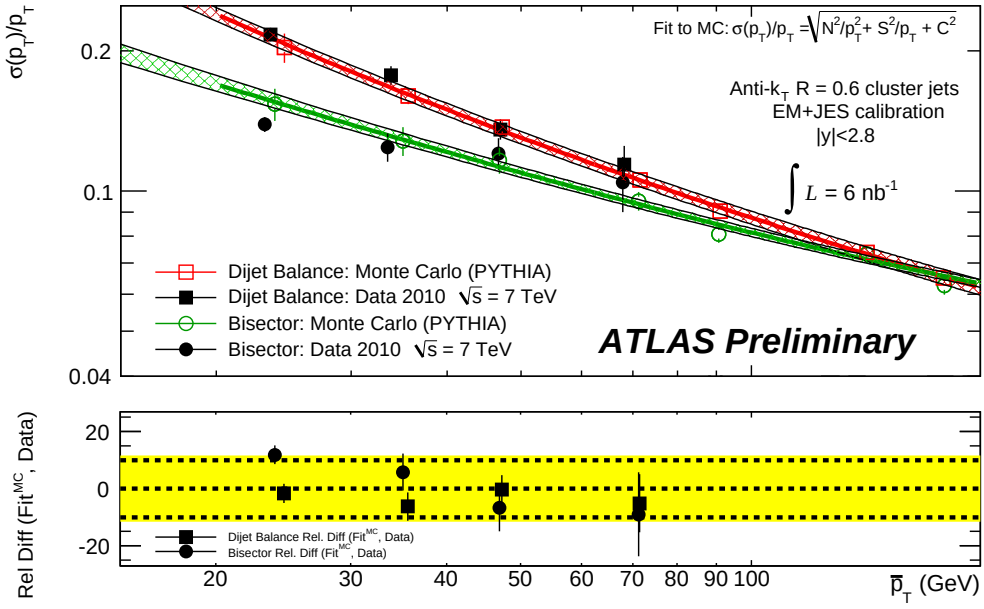

Figure 8.7: *Comparison of resolution as extracted from MC and data. Figure taken from Ref. [68].*

9 The cross section measurement

This chapter covers the details of the measurement of the jet cross section. After briefly reviewing the general concept and aim of the analysis, the largest fraction of this chapter is devoted to the derivation and validation of the necessary corrections to data. Subsequently the results are compared to the theoretical predictions established in Chapter 4.

9.1 Measurement principle

The aim of this analysis is the measurement of the inclusive, double differential jet cross section, as defined by:

$$\frac{d\sigma}{dp_T dy} = \frac{1}{L}\frac{N_{jets}(p_T, y)}{\Delta p_T \Delta y} \tag{9.1}$$

Here N_{jets} is the number of jets counted in a given bin of rapidity and transverse momentum, with bin widths Δy and Δp_T respectively and L is the integrated luminosity of the analyzed dataset.

As outlined earlier in Section 5.5 especially low threshold triggers are running in a prescaled mode. This running mode is especially necessary for the inclusive jet triggers, since the large cross section would otherwise produce too high rates for many triggers.

As a result, the totally accessible p_T range has to be split up into distinct regions for this measurement, where each is assigned to a certain trigger. Each jet with a given p_T is only taken into account if the event fullfilled the trigger assigned to the specific range of p_T. These intervals per trigger have already been introduced in the previous chapter and are summarized in Table 7.1. For the cross section measurement this results in the use of various values for the integrated luminosities, as these differ for the differently prescaled triggers. As a result of this, Eq. 9.1 has to be slightly modified to account for this fact and may be written as

$$\frac{d\sigma}{dp_T dy} = \frac{1}{L(p_T)}\frac{N_{jets}(p_T, y)}{\Delta p_T \Delta y}, \tag{9.2}$$

where $L(p_T)$ is the luminosity for the trigger assigned to the specific p_T range and as a consequence contains an implicit p_T dependence. In practice this results in the measurement being split up into five separate ones, each using data selected by only one trigger and being restricted to a certain range of p_T. Figure 9.1 shows the uncorrected number of jets in bins of p_T for each of these five triggers. The large differences of jet yields at fixed p_T but for different triggers observed in this distribution illustrate the necessity of splitting into different triggers: relying on fewer triggers would either require substantial corrections of trigger inefficiencies or reduce the available statistics substantially. Especially in the high p_T region were the statistical uncertainty starts to become significant there is no alternative to the usage of the highest non-prescaled trigger. The minimum bias trigger is the only trigger whose efficiency was measured to be 100% in the lowest p_T bin just above 40 GeV. This is why this trigger is used for the lowest p_T bin despite its comparably low statistics. Nevertheless the statistical uncertainty can be expected to

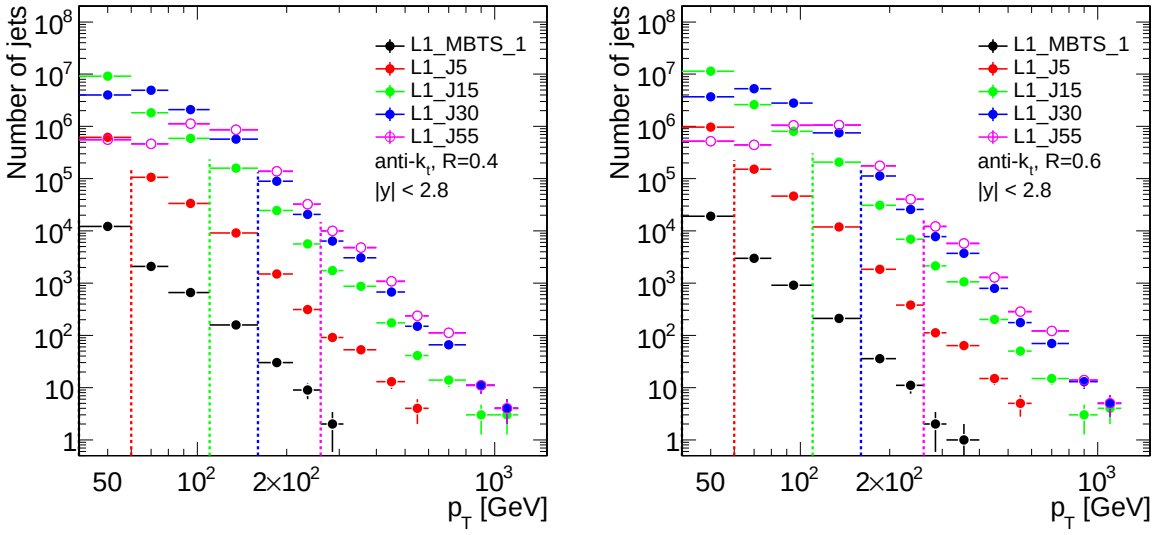

Figure 9.1: *Jet yield as a function of p_T per trigger for small and large anti-k_t jets. Shown are the raw, integer numbers of jets collected in each p_T bin, not normalized to the bin width. The dashed lines indicate the p_T tresholds from which on a certain trigger is used up to the next higher one.*

be negligible in view of the systematic uncertainties. The same argument holds against a lowering of the $p_{T,min}$ thresholds extracted in Section 7.5. Though it would in fact be possible to increase the statistics in certain bins significantly by lowering the $p_{T,min}$ thresholds, the anticipated advantage on the final result is negligible.

9.2 Analysis cuts and efficiencies

As described in detail in the previous chapter several steps of event and jet selection take place until the final event and jet sample is obtained:

1. **Trigger selection:** Each event is required to fullfill at least one of the triggers introduced in Table 7.1.

2. **Vertex selection:** The primary vertex in the event is required to have at least five tracks associated: $N_{PV}^{Tracks} \geq 5$.

3. **Missing $\mathbf{E_T}$ significance:** The $E_{T,Miss}^{Sig}$ quantity has to be below a threshold which is parameterized as a function of the leading jet's transverse momentum, as displayed in Fig. 7.2.

4. **Jet selection:** All jets have to fullfill the quality criteria listed in Table 7.2.

Each of these steps may also reject proper events or jets and thus introduce an inefficiency that needs to be corrected for. These effects are studied in the following:

- **Trigger efficiencies:** As outlined in Chapter 7 the momentum regions for each trigger were specifically selected as the region where the trigger can be assumed to be 100% efficient. For this reason no correction for trigger inefficiency is applied.

- **Vertex selection efficiency:** The vertex selection, requiring a minimum number of five tracks being associated to the primary vertex, neither does introduce significant inefficiencies that would need to be corrected. This can be concluded from Fig. 9.2, that shows the efficiencies of the various selection cuts as a function of p_T as they are extracted from the MC. For this reason no nominal correction for the jet vertex

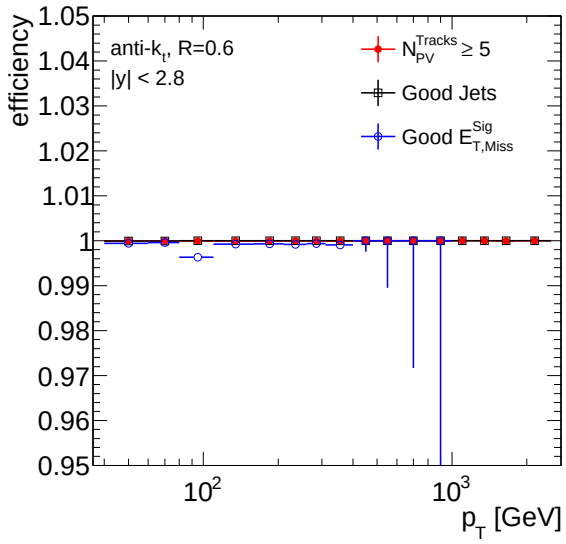

Figure 9.2: *Effiencies of event and jet selection cuts. The efficiency at* $p_T \approx 100$ GeV *can be attributed to a single MC event with a high weight, that is rejected.*

selection is performed.

- **Missing E_T significance selection:** This selection was designed to have no significant inefficiency. The inefficiency in the derivation of the cut values was constructed to be 1 ‰ and is also confirmed in Fig. 9.2. The 1 ‰ inefficiency is considered negligible.

- **Jet selection efficiencies:** The inefficiency of the jet selection cuts is estimated using the MC simulation. This is justified by the good modeling of all associated jet cleaning variables in the MC. The only exception from this statement is the jet quality variable Q_{jet} as demonstrated in Section 7.2. The impact of this flaw in the MC is believed to have a negligible impact on the jet selection efficiency for the following reasons:

  1. Both jet cleaning cuts using the Q_{jet} were exactly for this reason placed at the boundaries of the distributions in data as illustrated in Fig. 7.4 and Fig. 7.6. For both cleaning cuts Q_{jet} is also combined with another jet quantity that is well modeled in the MC.

  2. Several events containing jets classified as bad were scanned by eye and the jets were found to be indeed bad.

As a result of the very loose selection cuts there is no need for a correction of event or jet selection inefficiencies. However, although the nominal inefficiencies are small, there are sources of systematic uncertainties on each cut which are investigated in the following.

9.2.1 Systematic uncertainties

Trigger efficiency

The systematic uncertainty on the trigger efficiencies was derived in Section 7.5 and amounts to 1%.

Vertex selection

The ineffciency on the vertex selection is extracted from the N_{PV}^{Tracks} distribution of the events analyzed, as it is shown in Fig. 9.3. To estimate the number of events lost due to the $N_{PV}^{Tracks} \geq 5$ cut, the low N_{PV}^{Tracks} region is extrapolated towards zero, using a fit with a Landau function. This is illustrated in Fig. 9.3 where part of the N_{PV}^{Tracks} distribution for events passing the L1_J5 trigger is shown. Under the pessimistic assumption that all

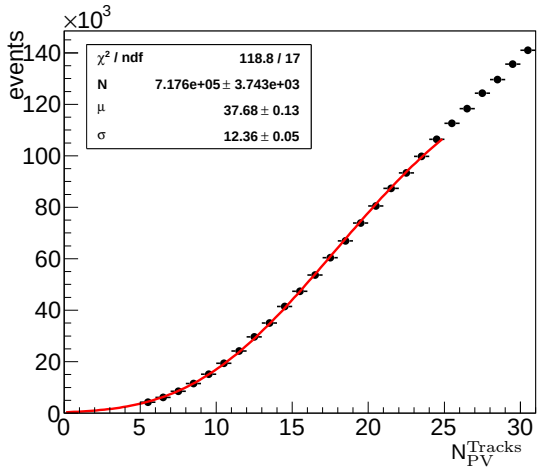

Figure 9.3: *The distribution of the number of tracks per vertex in events triggered by* L1_J5 *and a fit extrapolating to* $N_{PV}^{Tracks} < 5$*. The fit used a landau function with an overall normalization given by parameter N, the most probable value given by* μ *and a width* σ*.*

those events are signal events, an upper limit on the inefficiency $\bar{\varepsilon}_{vtx}$ of this cut can be extracted as:

$$\bar{\varepsilon}_{vtx} = \frac{N_{rej}}{N_{rej} + N_{sel}} \tag{9.3}$$

with :

$$N_{rej} = \int_0^5 f(N_{PV}^{Tracks})\, dN_{PV}^{Tracks}$$

$$N_{sel} = \int_5^\infty \frac{dN_{Events}}{dN_{PV}^{Tracks}} dN_{PV}^{Tracks},$$

where N_{rej} is the integral of the fit function and N_{sel} the number of events in the histogram. This value turns out to be below 1 ‰. This inefficiency is taken as a conservative estimate of the uncertainty.

$E_{T,Miss}^{Sig}$ selection

The $E_{T,Miss}^{Sig}$ selection was designed to reject only very few good events, so an obvious check in data is to verify this expectation. A first impression on how many events are rejected

can be gained from Fig. 7.2, which shows that only few events are above the cut at high p_T. To quantify this impression, the fraction of jets in events which are accepted by the $E_{T,Miss}^{Sig}$ cut is measured. In analogy to the estimation of the vertex selection efficiency, this quantity will serve as a lower limit of the actual efficiency ε_{met}*:

$$\varepsilon_{met} \geq \varepsilon_{met}^{all} = \frac{N_{jets}(\text{Good } E_{T,Miss}^{Sig})}{N_{jets}(\text{All})} \tag{9.4}$$

Figure 9.4 shows $1-\varepsilon_{met}^{all}$ in bins of p_T and y and confirms the expectation of a low number of events being rejected. In the high p_T regions the cut does indeed reject a significant

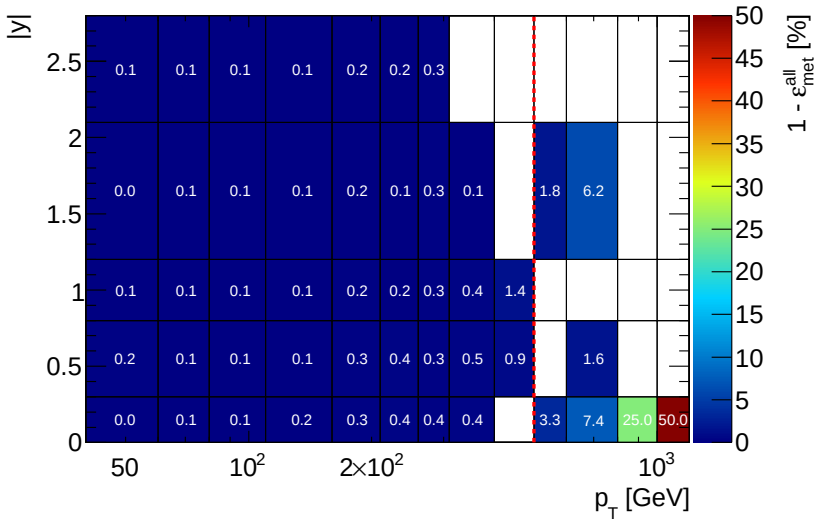

Figure 9.4: *Fraction of jets in events rejected by the $E_{T,Miss}^{Sig}$ requirement. The dashed, red line indicates $p_T = 500$ GeV.*

fraction of jets. It would however be overly conservative to assign for instance $\varepsilon_{met}^{all} = 0.5$ as an systematic uncertainty, without taking a look at the rejected events. Fortunately only eight events are rejected for $p_T > 500$ GeV, which allows an inspection of these events in event displays. Seven of this eight events can unambiguously be identified as cosmic events overlayed on a very soft pp-collision and only a single jet in an event that is only just above the $E_{T,Miss}^{Sig}$ cut could be assumed be a proper jet lost for the analysis. The eighth event contributes a single jet to the $500 < p_T < 600$ GeV bin, at central rapidity. In this bin 58 jets are counted, thus one additional jet would correspond to an increase of the cross section by 1.7 %. In view of the statistical uncertainty of 13 % in this bin, the very small increase in the cross section, that would result from incrementing the number of jets by one is considered negligible. Thus for the high p_T region it can be concluded that the inefficiency of the $E_{T,Miss}^{Sig}$ cut is negligible. Identical checks have been made for the jets rejected in the $400 < p_T < 500$ GeV bin, finding only a single proper jet being rejected, in view of approximately 400 jets counted in total.

Concentrating on the lower p_T region ($p_T < 400$ GeV), where the number of jets is of course significantly higher, an inefficiency below 0.5% can be observed. This number also represents an upper limit on the inefficiency for this selection and is used as a conservative estimate of the uncertainty associated with the $E_{T,Miss}^{Sig}$ selection.

*The limit would be the true efficiency under the assumption that all rejected events were signal events.

Jet selection

The jet selection inefficiency taken from the MC is trusted to a large degree. Uncertainties however can be assumed to be associated with the cuts involving the Q_{jet} variable whose significant energy dependence is not at all modeled in the MC. For this purpose the Q_{jet} distributions are examined in bins of p_T, of which two are shown in Fig. 9.5. The majority

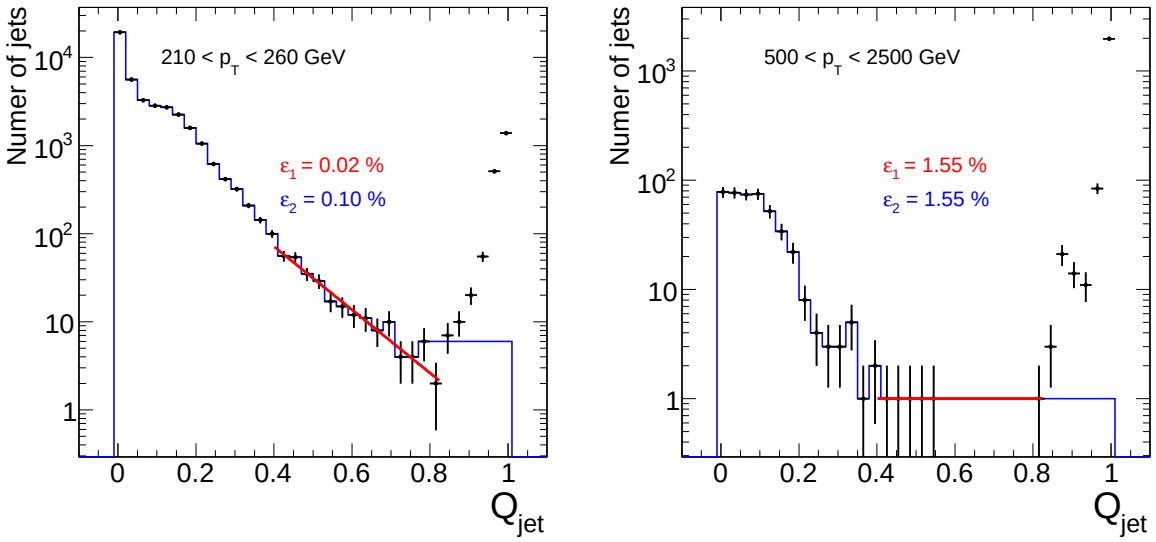

Figure 9.5: *Jet quality distributions for two bins of p_T. The red line shows a fit of the form $f(x) = ae^{-(x-0.2)/A} + be^{-(x-0.2)/B}$, which results in the inefficiency ε_1. The blue histogram is identical to the black data points, but continues as a constant beyond 0.8. It yields the inefficiency ε_2.*

of jets has a jet quality $Q_{jet} \equiv 0$, due to the fact that none of the constituent cells has a quality value above the critical threshold. Besides, there is an approximately exponentially decreasing population for $Q_{jet} > 0$ down to a minimum around the cut value of $Q_{jet} \approx 0.8$. In order to estimate the inefficiency of the cut, the distribution below 0.8 is extrapolated up to 1 using an exponential parameterization. The integral of the function from 0.8 to 1 is considered the number of proper jets that are lost. A more pessimistic variation of this principle is illustrated by the blue histograms drawn in Fig. 9.5, which extrapolates into the cut region with a constant, set to the bin content just below 0.8. The inefficiencies estimated in both ways are also shown and are well below 0.5% for $p_T < 500$ GeV.

For higher p_T values both methods become unstable due to the lack of statistics of intermediate Q_{jet} values. But at the same time this results in two well separated populations of good and bad jets. As there is no reason to suspect a significant fraction of good jets in the high Q_{jet} population, a fully efficient selection of jets in these p_T regions can be assumed.

As a result of these considerations a flat uncertainty of 0.5% is assumed for the jet cleaning inefficiency.

9.3 Resolution unsmearing

In contrast to the negligible selection inefficiencies, the impact of the finite detector resolution needs to be corrected in this measurement. The calorimeter's energy resolution

that translates into a resolution in our observable p_T blurs the quantity that is in fact the aim of the measurement. Let N_{true} and N_{reco} denote the number of jets produced and reconstructed within a certain bin of p_T and y. In general these two quantities are not necessarily equal, since the reconstructed quantities may end up in different bins of p_T and y.

For the case of the jet cross section this has a very subtle effect that is due to the very steeply falling spectrum: For a fixed p_T bin the falling cross section always leads to more jets migrating into the bin from the left than jets migrating out of this bin to the left. For any analysis, that applies a lower p_T cut, the net effect of this bin-to-bin migration for a steeply falling spectrum is an increase in the measured number of jets. Transferred to the cross section observable defined in Eq. 9.1 this can mathematically be described by a convolution as:

$$\frac{d\sigma(p_T^{\text{reco}}, y^{\text{reco}})}{dp_T^{\text{reco}} dy^{\text{reco}}} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{d\sigma(p_T^{\text{true}}, y^{\text{true}})}{dp_T^{\text{true}} dy^{\text{true}}} \times g(p_T^{\text{true}}, y^{\text{true}}, p_T^{\text{reco}}, y^{\text{reco}})\, \mathrm{d}p_T^{\text{true}}\, \mathrm{d}y^{\text{true}}. \quad (9.5)$$

In this formulation $g(p_T^{\text{true}}, y^{\text{true}}, p_T^{\text{reco}}, y^{\text{reco}})$ is the probability that a jet with true transverse momentum of p_T^{true} and rapidity y is reconstructed with p_T^{reco} and y^{reco}. The transfer function alone for p_T under the assumption of a Gaussian resolution can be formulated as follows:

$$g(p_T^{\text{true}}, p_T^{\text{reco}}) = \frac{1}{\sigma(p_T^{\text{true}})\sqrt{2\pi}} \exp\left[-\left(\frac{p_T^{\text{true}} - p_T^{\text{reco}}/R(p_T^{\text{true}})}{2\,\sigma(p_T^{\text{true}})}\right)^2\right], \quad (9.6)$$

where $\sigma(p_T^{\text{true}})$ is the absolute resolution in p_T depending on p_T^{true} and $R(p_T^{\text{true}})$ is the response as defined in Chapter 6. Similarly the rapidity transfer function can be composed of a rapidity response, which can be assumed to be one, and the rapidity resolution. It will be shown later, that the rapidity mismeasurement has an almost negligible impact on the cross section measurement compared to the p_T resolution effects. For this reason the discussion will focus on the latter.

In view of a binned measurement, $g(p_T^{\text{true}}, p_T^{\text{reco}})$ can also be written as a matrix g_{ij}, containing the probability that a jet from a given p_T^{true} bin i migrates to the j^{th} p_T^{reco} bin. These probabilities can be extracted from the MC and are displayed in Fig. 9.6. According to the probability definition every row of fixed p_T^{true} sums up to unity. As expected the matrix shows that the most probable, reconstructed bin is the original p_T^{true} bin, which is indicated by the fact that the diagonal matrix elements are the maximum ones for each p_T^{true} bin. One can also notice that due to the steeply falling spectrum the migrations to the neighboring low p_T bin are stronger than to the high p_T neighbor. This can be attributed to the fact that within the relatively wide bin the p_T distribution will be strongly peaked at the lower bin edge.

Extending the view beyond the neighboring bins, it is also remarkable that the fluctuation to too high reconstructed p_T^{reco} values is constrained to the next-to-neighboring bin, whereas the low response tails regularly span more than four bins. This is at least partly an artifact of the increasing bin-size. A jet with fixed p_T^{true} will be reconstructed with $p_T^{\text{reco}} = p_T^{\text{true}} + p_{T,Random}$, where $p_{T,Random}$ denotes the random mismeasurement by the calorimeter and fluctuates equally to positive and negative values. Due to the increasing bin-sizes however, the fluctation in *units* of bins will be asymmetric and larger towards lower p_T bins in this analysis. The residual tails towards extremely low response can be attributed to properties of the energy measurement in a calorimeter where several effects

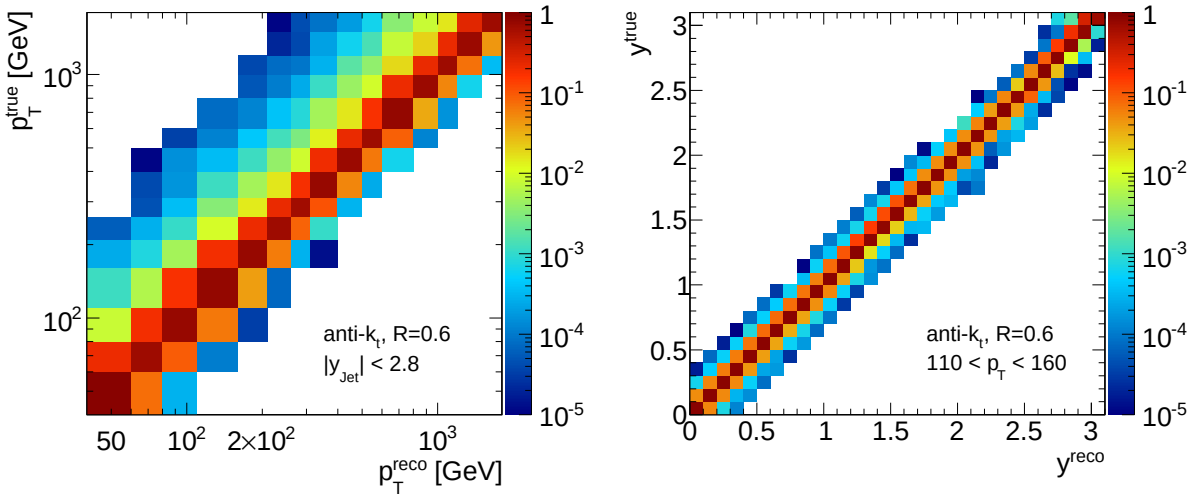

Figure 9.6: *Response matrices for rapidity and transverse momentum as extracted from the MC simulation.*

can lead to decrease in measured energy while the probability for too high reconstructed energies in general is lower. It is worth mentioning that this is in fact very fortunate: Migrations from high p_T to much too low p_T can hardly have a significant impact, since the cross section at low p_T is much higher. In that sense systematic migrations across several bins to too high p_T would have a very strong impact.

The rapidity response matrix reveals that, compared to the bin-widths of the analysis, the resolution for the rapidity measurement, is significantly better than for the p_T measurement. At $y^{\text{true}} \approx 1.4$ the increased migrations to $y^{\text{reco}} < y^{\text{true}}$ arise from the fact this regions corresponds to the transition between the barrel and the endcap calorimeters. The distribution of inactive material in this regions leads to slight bias in the reconstruction of a jet's rapidity. Similar features can also be observed at the boundary of the analysis region at $y^{\text{true}} \approx 2.8$.

9.3.1 Bin-by-bin unfolding

The aim of the *unfolding* procedure will be to extract the true quantities from the measured one. For this purpose a bin-by-bin correction factor is extracted from the MC, that will be applied to the measured cross section $\sigma_{reco}(p_T, y)$. These unfolding factors are defined as follows:

$$C\left(p_T^{\text{reco}}, y^{\text{reco}}\right) = \frac{\sigma_{true}(p_T, y)}{\sigma_{reco}(p_T, y)} \tag{9.7}$$

$$= \frac{N_{true}(p_T, y)}{N_{reco}(p_T, y)}, \tag{9.8}$$

where the rewriting from cross sections to the raw number of jets makes use of the fact that the luminosity cancels out in the ratio. Taking this correction into account Eq. 9.2 becomes

$$\frac{d\sigma}{dp_T dy} = \frac{1}{L(p_T)} \frac{N_{jets}(p_T, y) \times C\left(p_T, y\right)}{\Delta p_T \Delta y}, \tag{9.9}$$

and will be the final formula for the calculation of the cross section. Before extracting these correction factors from the MC it is crucial to be confident in the MC description of the data. There are two major items that affect the resulting unfolding factors:

1. **Jet response:** The jet energy scale and resolution are the crucial items that define the transfer matrix shown in Fig. 9.6. The nominal unfolding corrections will be derived under the assumption that these items are properly modeled in the MC. The precision to which this statement can be made was investigated in Chapter 8 and will be accounted for in the assessment of the systematic uncertainties.

2. **Simulated cross section:** Besides the transfer function, Eq. 9.5 contains the simulated, true cross section that is smeared by the detector effects which are defined by the transfer function. In order to get correct unfolding factors it is necessary that the MC does describe the cross section shape in data well, since more or less steep shapes will change the net effect of the migrations induced by a fixed transfer function.

In Chapter 7 it became obvious that the nominal MC used does not satisfy the latter requirement (see Fig. 7.13). A derivation of the unfolding factors using the unchanged MC would thus result in unfolding corrections which would not be the best estimate.

Reweighting of the MC simulation

One possible solution for this problem is to reweight the MC in order to improve the agreement in shape. The starting point for this procedure is the data to MC ratio in Fig. 7.13 which is fitted using a smooth function given by:

$$w(x) = \begin{cases} 1 + a\log(x) + b\log(x)^2 & \text{for } \log(x) > \frac{-a}{2b} \\ 1 - \dfrac{a^2}{4b} & \text{else.} \end{cases} \tag{9.10}$$

This piecewise definition takes care of a smooth extrapolation to p_T below the fit range: From the point where the derivative of the function vanishes it is continued as a constant. The fit is performed using the data to MC ratio for $R = 0.6$ jets in the inclusive rapidity bin as illustrated in Fig. 9.7. The reweighting function found is able to reproduce the observed ratio between data and MC to a very good degree.

In contrast to the derivation of the weight, the weight is not applied as a function of p_T to the MC. Instead the weight is used as a per-event-weight by which the existing pythia Jx weight (as introduced in Chapter 4) per event is multiplied. For this purpose the $w(x)$ function is evaluated as a function of $\hat{p}_T$, which is the transverse momentum of the simulated $2 \to 2$ hard parton scattering. This method is considered to be safer with respect to a reweighting based on jet quantities, which might not only change the general shape in p_T but could also be sensitive to event topologies. Doing the reweighting as described is more similar to commonly used *PDF reweighting* procedures.

The result of this procedure using the parameters extracted in Fig. 9.7 is displayed in Fig. 9.8. While the agreement is not perfect, it is significantly improved especially for the inclusive rapidity bin which was used for the extraction of the reweighting function. The difference between data and the reweighted MC is below 10% for all rapidity bins except for the most forward one and below 5% for the inclusive rapidity bin. These remaining differences and their impact on the unfolding factors will be included in the estimation of the systematic uncertainties. The MC used in the following will always refer to the reweighted MC if not stated otherwise.

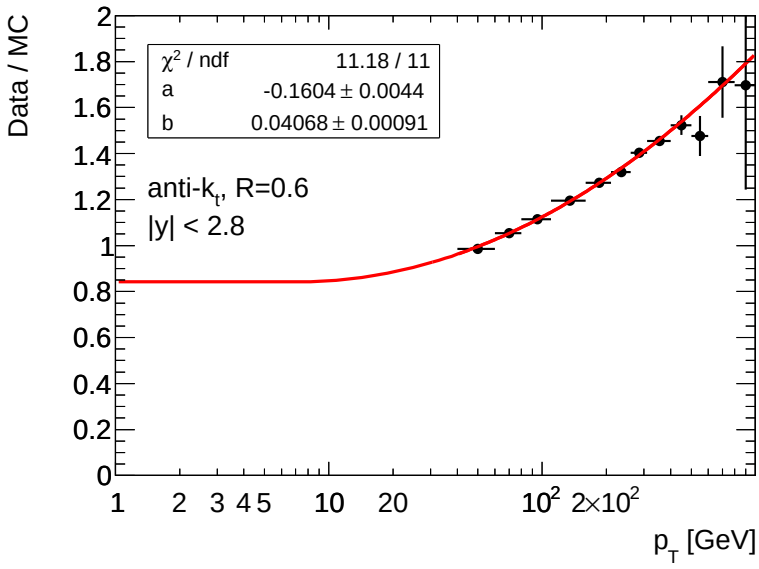

Figure 9.7: *Fit of the reweighting function defined in Eq.* 9.10 *to the data to MC ratio.*

Extraction of unfolding factors

The extraction of the correction factors $C\left(p_T^{\text{reco}}, y\right)$ from the (reweighted) MC is now straight forward. The resulting unfolding factors are displayed in Fig. 9.9. There are some interesting observations to make on the shape of the correction factors versus p_T: Firstly, the expectation of $N_{reco} > N_{true}$ is confirmed. Secondly there is a characteristic p_T dependence than can be attributed to the p_T evolution of the relative jet energy resolution described in Eq. 6.11 and the slope of the jet cross section. At low p_T, where the fractional p_T resolution is worse, also the absolute value of the slope is maximized. Both effects themselves lead to an amplification of the migration effect and thus the difference between N_{reco} and N_{true} is maximized in this region. At high values of p_T close to 1 TeV the cross section slope flattens and at the same time the fractional resolution approaches the minimum defined by the constant term. Accordingly the net effect of the migrations are minimized.

One subtle detail in the calculation of the unfolding factors is the associated statistical uncertainty. For a proper calculation, in principle the correlation between N_{true} and N_{reco} would have to be known. Of course this correlation cannot assumed to be zero, but unfortunately it is also not straight forward to get a solid numerical value for it.

A workaround for this can be constructed using information provided by the response matrix. For this purpose, the efficiency ε_{true} and the purity p_{reco} are defined. Both quantities are based upon the number of jets generated and reconstructed in the same p_T bin (N_{ep}) and are defined in Fig. 9.9. The efficiency denotes the probability that a jet with p_T^{true} is reconstructed with p_T^{reco} being in the same bin. Similarly the purity is the fraction of reconstructed jets in a given bin, whose corresponding true jet p_T lies in the same bin. One can then define the numbers of jets that do not have their corresponding true or reconstructed jet in the same p_T bin:

$$N_{ep} = \varepsilon_{true}\, N_{true} = p_{reco}\, N_{reco} \tag{9.11}$$

$$N_{\bar{e}} = N_{true}\,(1 - \varepsilon_{true}) \tag{9.12}$$

$$N_{\bar{p}} = N_{reco}\,(1 - p_{reco}), \tag{9.13}$$

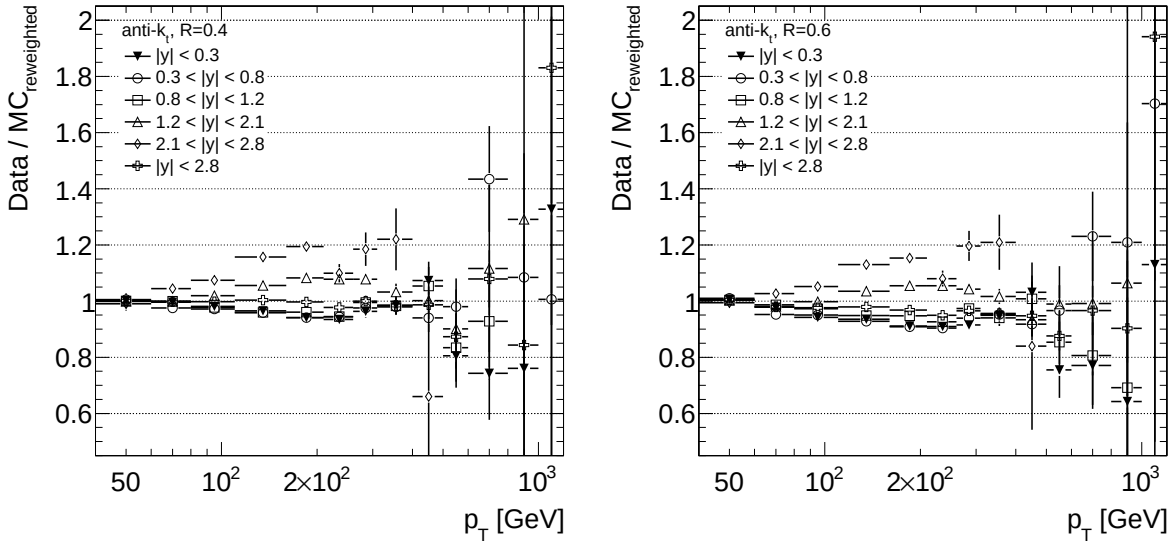

Figure 9.8: *Ratios of the jet p_T distributions in data and MC after the above described MC reweighting procedure for large and small jets.*

where the existing p_T dependence of all included quantities is omitted for brevity. With the help of these two quantities Section 9.3.1 can be rewritten as follows:

$$C = \frac{N_{true}}{N_{reco}} = \frac{N_{\bar{e}} + N_{ep}}{N_{\bar{p}} + N_{ep}} \tag{9.14}$$

In the latter formulation the three N's are now considered uncorrelated and the error propagation is straight forward. The values for purity and efficiency can be read off the response matrix in Fig. 9.6 and range between 60% and 70%. This method is already employed in the calculation of the uncertainties in Fig. 9.9.

Systematic uncertainties

As outlined earlier, the bin-by-bin unfolding is sensitive to eventually not well modeled features in the simulation, namely the jet response in energy and angle and the simulated true cross section. The uncertainties of these items result in an uncertainty on the derived unfolding factors. To assess these uncertainties, the general procedure is as follows:

Depending on the source of uncertainty the MC used to extract the unfolding factors will be changed at the analysis level, meaning that the actual MC generation is not changed. This will result in sets of correction factors that can in general be expected to differ from the ones obtained from the nominal MC. To assess the impact of these changes on the resulting cross section, in principle the reconstructed cross section in the nominal MC would have to be unfolded using unfolding factors from an altered MC. The difference between the cross sections obtained this way and the nominal ones can than be taken as the associated uncertainty. Fortunately this procedure can be cut short by comparing the unfolding factors directly and propagating their relative difference onto the cross section measurement. This procedure is numerically equivalent since the unfolding solely consists of a scalar multiplication.

Dependency on the simulated cross section The unfolding method to a certain degree depends on the cross section in the simulation. By construction, in the limit where

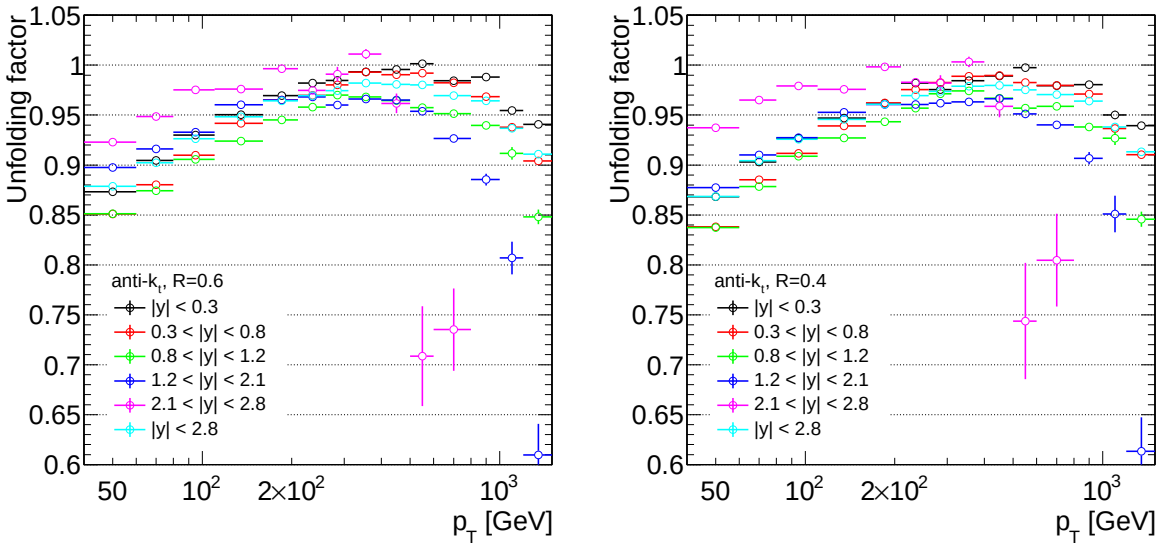

Figure 9.9: *Unfolding corrections versus the reconstructed jet p_T for large and small jet resolution paramter for the inclusive rapidity bin.*

the reconstructed p_T distribution coincides with the one predicted by the MC the result of the unfolding procedure will be the simulated cross section shape. This effect however can be controlled and the eventual bias it might introduce will be estimated using a variety of reweighting functions to alter the simulated cross section. The major item of discussion in this procedure is how large the variations have to be chosen in order to cover the possible variations.

The best hint at the real shape of the true cross section is given by the shape of the measured p_T spectrum in data. This can be understood by taking into account that the actual change in shape going from true to reconstructed jets is given by the unfolding factors shown in Fig. 9.9. The deviations from unity in the unfolding factors are in fact much smaller than the ones observed between unweighted simulation and data (see Fig. 9.7). For this reason it is sensible to do the variations in shape on the same order of magnitude observed in the comparisons of data to simulation.

The reweighting scenarios used are displayed in Fig. 9.10. The shapes are varied such that the relative cross section between 40 GeV and 1 TeV is approximately increased or decreased by a factor of 2, which conservatively covers the data to MC comparisons. The functional form used is

$$w(x) = 1 + \left(\frac{\log\left(x/20\ \text{GeV}\right)}{\log\left(50\right)} \right)^{P}, \tag{9.15}$$

whose only physical motivation is that changes in shape are expected to be smooth and logarithmic as a function of p_T. The parameter P is used to vary the slope of the weighting function. The parameters besides P were chosen such that the function value is 1 at 20 GeV and reaches 2 at 1 TeV, thus spanning the kinematic region of the measurement. The resulting weight parameterizations for a set of P's are drawn in Fig. 9.10. As in the reweighting introduced for the nominal MC this weight is multiplied onto the existing event weight as a function of $\hat{p}_T$.

The differences with respect to the nominal unfolding factors are shown in Fig. 9.11, which reveals that the variation of the true cross section shape has only very limited impact. The deviations are below 2% throughout the full p_T range. The p_T dependence

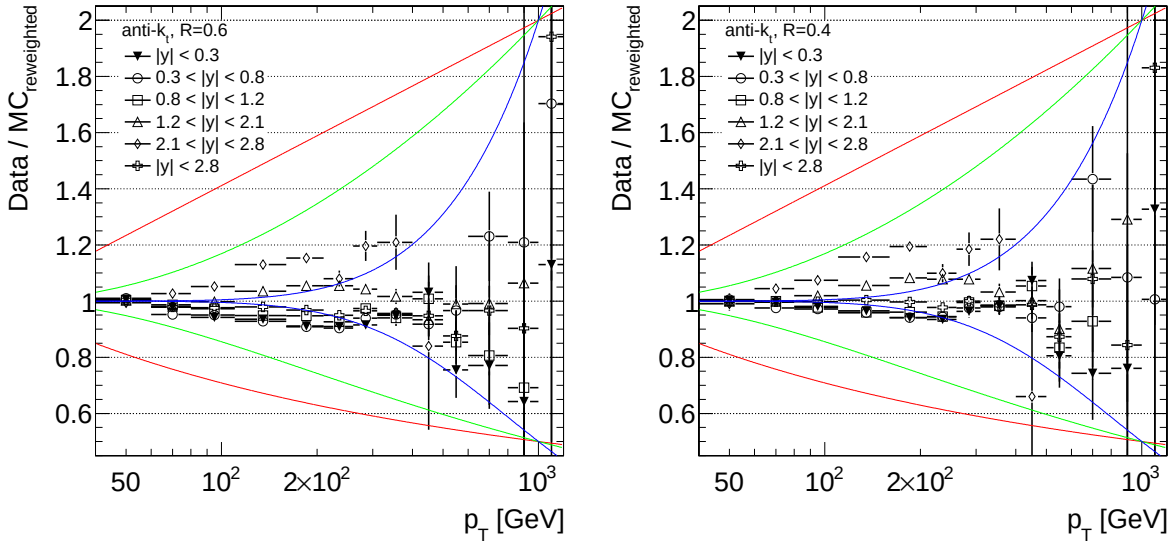

Figure 9.10: *Scenarios used to reweight the MC truth cross section overlayed on the data to MC comparison from Fig. 9.8. The colors refer to different values of P in Eq. 9.15, namely 1 (red), 2 (green) and 6 (blue). For each parameter the existing event weight was either multiplied or divided by the $w(x)$*

of the uncertainty is obviously an artificial one, because it is determined by the choice of P: the uncertainty will turn out to be maximal where the derivative of $w(x)$ is largest, which is either at the low or high p_T end of the momentum range studied. For this reason the systematic uncertainty assigned will be constant and taken as the largest deviation found in Fig. 9.11. Qualitatively Fig. 9.11 does not change for the single rapidity bins.

Jet energy resolution As can be read off from Eq. 9.5 the second ingredient to the convolution is the jet energy transfer function defined in Eq. 9.6. For the systematic uncertainties the crucial item within the latter definition is $\sigma(p_T^{\text{true}})$, the resolution as a function of p_T^{true}. The results presented in Chapter 8 showed that the resolution in MC agrees with data within 10%. In order to estimate the influence of a wrongly simulated jet energy resolution in the MC, the resolution in MC is artificially worsened.

The procedure to increase the resolution by a factor f_{res} is as follows:

1. Take a jet with transverse momentum p_T and rapidity y.
2. Look up the nominal, fractional resolution $\sigma_{rel}(p_T, y)$ for this jet using Table 6.2.
3. The smeared resolution has to fullfill: $\sigma_{smear} = f_{res}\,\sigma_{rel}$. This can be achieved by multiplication of p_T with a factor r, that is randomized according to a normal distribution centered around 1 and with the width σ_{rndm} being:

$$\sigma_r = \sigma_{rel}\sqrt{f_{res}^2 - 1}, \tag{9.16}$$

such that:

$$\sigma_{smear} = \sqrt{\sigma_{rel}^2 + \sigma_r^2} = \sqrt{\sigma_{rel}^2 + (f_{res}^2 - 1)\sigma_{rel}^2} = f_{res}\,\sigma_{rel}. \tag{9.17}$$

So, in the final step, randomize r with a with σ_r and change p_T according to:

$$p_T \rightarrow r \times p_T \tag{9.18}$$

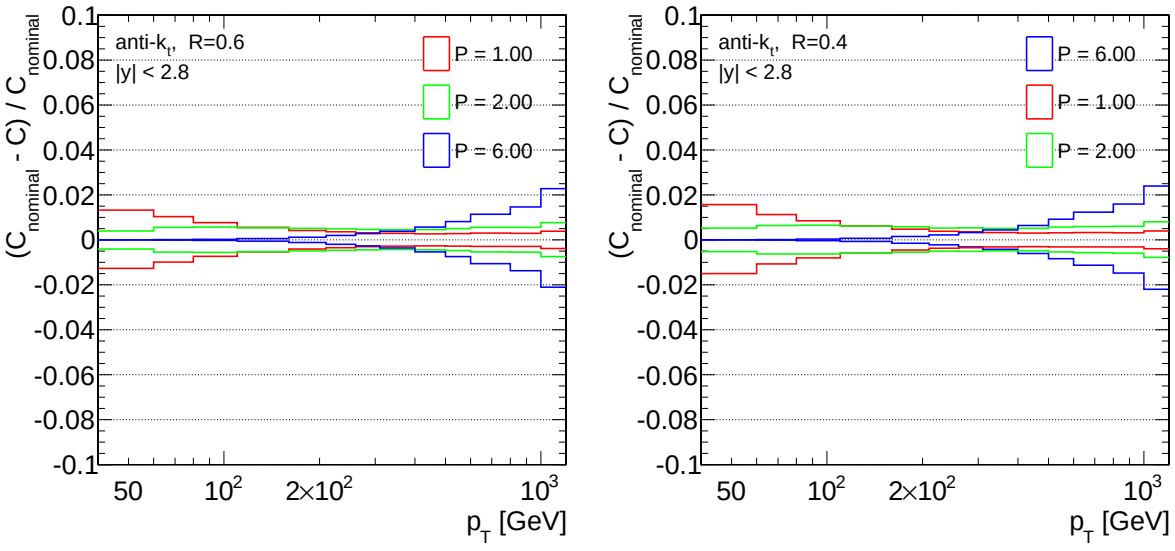

Figure 9.11: *Difference in unfolding corrections induced by a shape reweighting. The reweighting uses Eq.* 9.15 *with the indicated value (positive and negative) for P.*

This was done for three values of f_{res} in order to worsen the nominal resolution by 5%, 10%, 15% and 20%. In order to minimize statistical fluctuations of the varied correction factors with respect to the nominal ones, the above procedure can be repeated multiple times per jet. This way the varied sample inherits the statistical fluctuations of the nominal one, but the deviations due to the change in resolution can be estimated with improved precision. The extracted variation in the unfolding factors are shown in Fig. 9.12. It is

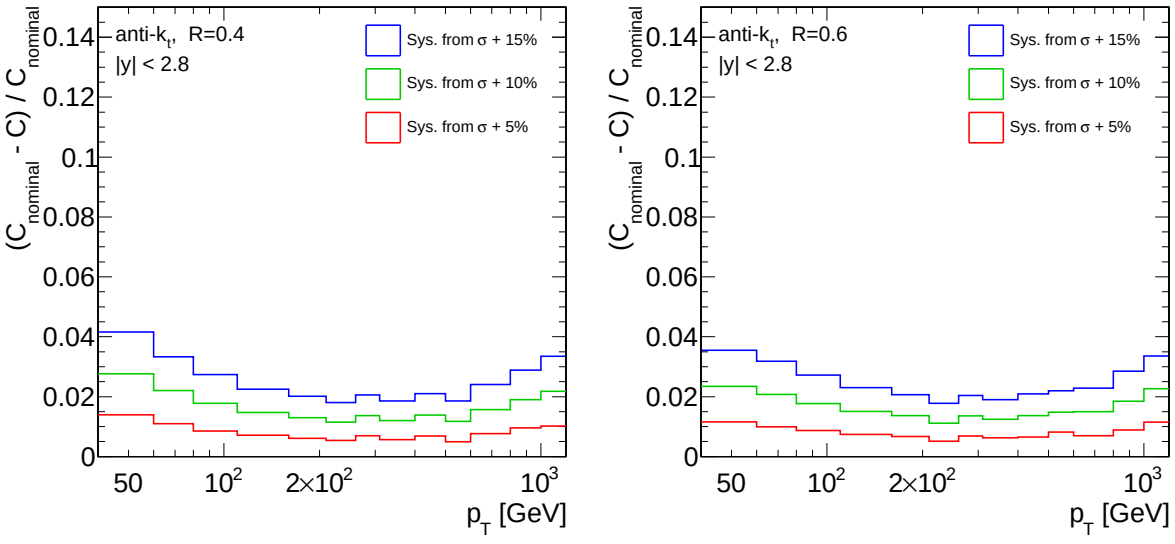

Figure 9.12: *Difference in unfolding corrections from resolution smearing.*

worthwhile mentioning that in contrast to Fig. 9.11 the general behaviour as a function of p_T in this case can be considered a real effect. This is due to the fact that, while the size of the resolution smearing is constant, this has a different impact at low and high p_T. The reason for this is the same, that is driving the shape of the nominal unfolding factors, namely the running of the relative energy resolution and the variation in the cross section

shape.

It is also striking that for fixed p_T the variation in the unfolding factors is to a very good approximation linear. While with the method described above it is not possible to improve the resolution in MC, this linearity allows for a symmetric assignment of the systematic uncertainty extracted from Fig. 9.9. This feature also allows to estimate the systematic uncertainty associated with an arbitrary resolution uncertainty.

Compared to the uncertainties found in Fig. 9.11 the uncertainties are slightly larger, ranging between 1% and 3% for an uncertainty on the resolution of 10%.

Rapidity resolution The mismeasurement of the jet rapidity can, in a way analogous to the energy mismeasurement, also affect the measured cross section $\sigma(p_T, y)$: Jets with mismeasured rapidity may end up in a different rapidity bin and thus change measured the cross section in both bins. However there are two aspects that make this effect almost negligible with respect to the p_T migrations investigated so far:

1. **Spatial resolution:** The precision of the measurement of rapidity is dominated by the measurement of θ. The resolution compared to the bin width here is much better than for the energy measurement and ranges between approximately 40 mrad in the central detector and 10 mrad in the most forward bin.

2. **Cross section dependency:** The cross section dependency for fixed p_T and varying y compared to the p_T dependence is much less steep. While the cross section varies over more than seven orders of magnitude as a function of p_T it does much less as a function of y.

The impact of a mismodeled θ resolution is estimated by increasing it by 5%. The procedure is identical to the one employed for the increase in p_T resolution. The results in Fig. 9.13 reveal that the additional smearing of θ obviously does not change the resulting correction factors in a significant manner. It should be noted that also for this excercise each jet was used with a smeared θ 100 times. The only tendency that can be observed is that the smearing of θ seems to decrease the visible cross section to a very slight degree. Due to this situation the systematic uncertainty associated is obtained by fitting the deviations with a constant value.

Results

Putting all three sources of systematic uncertainties together, the final unfolding factors are displayed in Fig. 9.14. As shown before, the corresponding uncertainties for $R = 0.4$ jets are almost identical and thus are not shown for brevity. The single contributions to the systematic uncertainties are illustrated too, amplified by a factor of ten for optical reasons. It should be noted that the relative uncertainty on the unfolding factors turn out to be small. For $p_T > 100$ GeV the sum of the uncertainties is around 2% in all rapidity bins except for the most forward bin, where the θ resolution studies increase this value to approximately 3%. One can also observe that the systematic uncertainty due to the p_T resolution uncertainty is the dominant contribution in all rapidity bins. As found out in the studies above this uncertainty will benefit directly from any improvement in this direction due to the linear dependence of resolution and unfolding factor uncertainty.

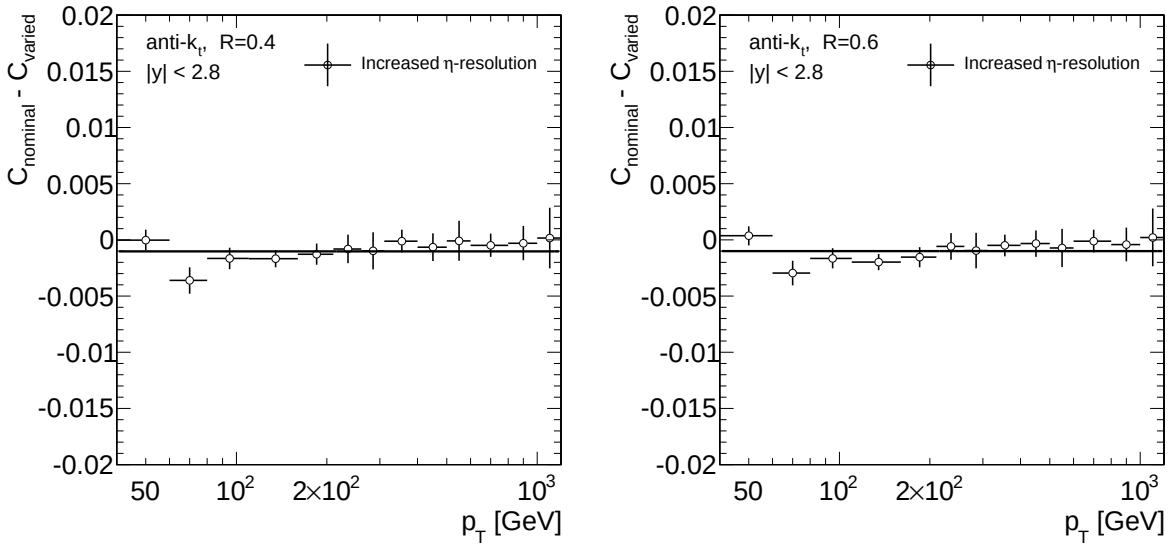

Figure 9.13: *Difference in unfolding corrections from η-smearing.*

9.4 Systematic uncertainties

For completeness here all ingredients to Eq. 9.9 are listed once more:

- $C\left(p_T, y\right)$: The uncertainties on the unfolding factors have been evaluated in the previous section.
- $L(p_T)$: the uncertainty on the luminosity is taken from the corresponding luminosity measurement done by the ATLAS collaboration, which is documented in [82]. The measurement is performed using so called *van der Meer scans*, in which the transverse profile of the proton beams is measured. In combination with a measurement of the beam currents, this allows to extract a visible cross section that can be used to convert event rates measured in other runs to instantaneous luminosities. The uncertainty amounts to 3.4%, which is dominated by the uncertainty on the measurement of the beam currents and is hence fully correlated among all triggers and runs.
- $N_{jets}(p_T, y)$: The number of jets in a given p_T and y bin is the quantity to which several sources of systematic uncertainty contribute. These will be evaluated in the following.

The uncertainties on $N_{jets}(p_T, y)$ in principle inherit all the uncertainties associated to the single event and jet selection steps introduced in Chapter 7. As outlined in Section 9.2 the nominal inefficiency on all selection cuts is assumed to be negligible. To which precision this statement can be made for all involved cuts will be examined in the following.

9.4.1 Jet energy scale uncertainty

At first sight the above mentioned uncertainties could be considered complete. Variations in the efficiencies mentioned above would directly lead to an increase or decrease in the number of jets counted. The major source of uncertainty however arises from the fact

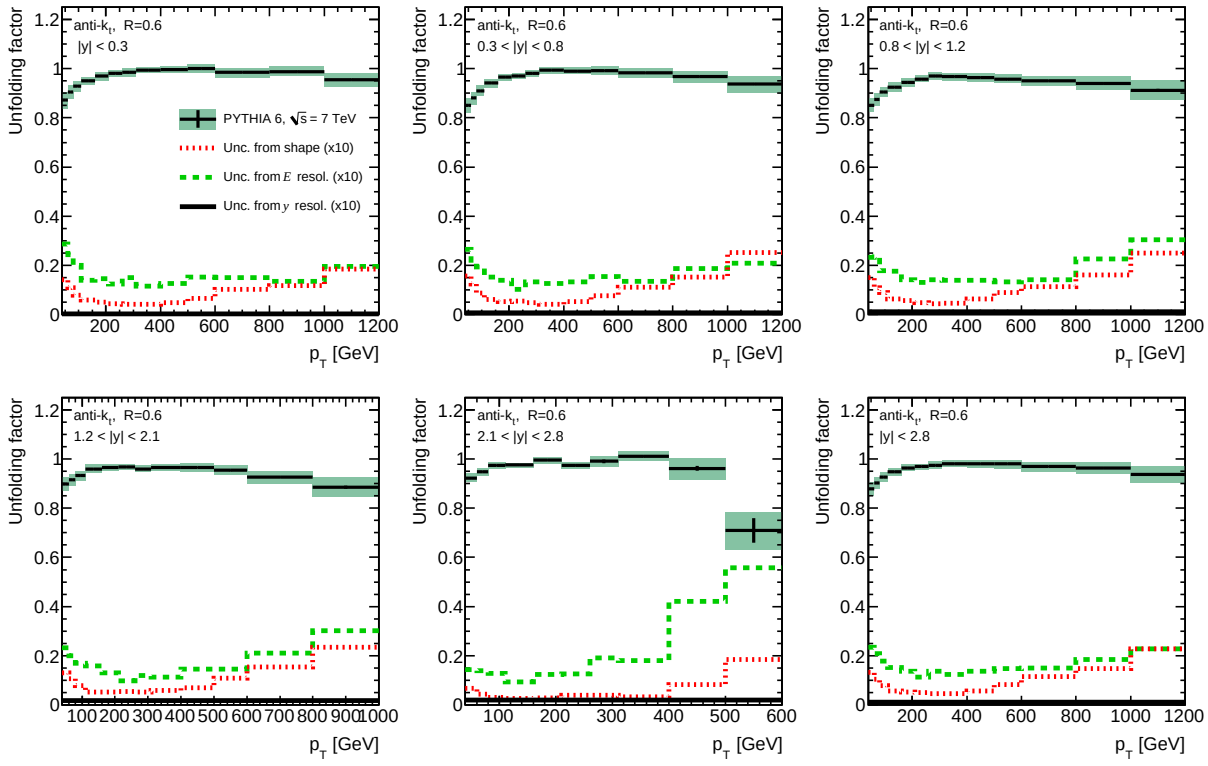

Figure 9.14: *Unfolding corrections versus jet p_T including systematic uncertainties.*

that the average, measured p_T could be off. The degree to which the average, measured p_T is trusted is quantified by the jet energy scale uncertainty introduced in Chapter 8. To assess the impact of this uncertainty, the cross section is extracted from MC varying each jet's p_T by the jet energy scale uncertainty taken from Fig. 8.3. The resulting change in the cross section is taken as the systematic uncertainty, asymmetrically for the increase and decrease of the jet energy scale. The relative JES uncertainties are approximately 4%, and lead to relative uncertainties on the cross section that are a factor of four to five larger, depending on the slope of the cross section. This results in uncertainties on the measured cross section which range between 20 % and 30 %. Figure 9.15 shows these uncertainties together with the remaining ones.

9.4.2 Total systematic uncertainties

Table 9.1 summarizes all systematic uncertainties. Among the constant uncertainties, the luminosity uncertainty of 3.4 % is the dominating one, while the remaining ones are on the order of 1 %. In the end however, all constant contributions are significantly exceeded by that of the jet energy scale. This is also illustrated in Fig. 9.15, which shows all uncertainty components and the resulting total systematic uncertainty as a function of p_T.

9.5 Results

Having established the unfolding procedure and its associated uncertainty, all ingredients to the determination of the actual cross sections in data are ready. The measured cross sec-

Source	Size
Luminosity	3.4 %
Trigger efficiency	1 %
Vertex selection	1 ‰
$E_{T,Miss}^{Sig}$	0.5 %
Jet cleaning	0.5 %
Total (constant)	3.6 %
Unfolding	2 % - 3 %
Jet energy scale	20 % - 30 %

Table 9.1: *Table of systematic uncertainties.*

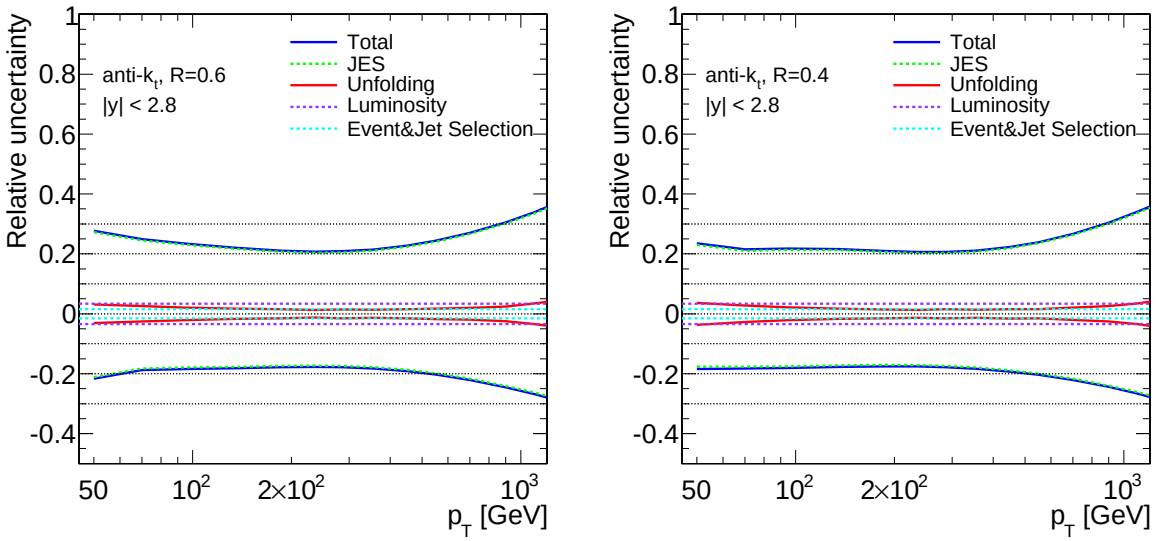

Figure 9.15: *Summary of systematic uncertainties.*

tions can hence finally be compared with the theoretical predictions outlined in Chapter 4. For the inclusive rapidity bin, $|y| < 2.8$, this is done in Fig. 9.16 for both jet radii. The measured cross sections at low p_T are of the order of 20 nb, steeply falling down to below 1 fb at $p_T \approx 1$ TeV. For both jet sizes there is good agreement with the NLO prediction, over eight orders of magnitude, within the systematic uncertainties. The measurement is systematically limited up to approximately 800 GeV where the statistical uncertainties become increasingly important. While of course in general the cross sections are very similar for both jet radii, it can be observed that especially at low p_T the $R = 0.6$ cross section is significantly enhanced. Taking a closer look at the direct comparison in the data to theory ratio reveals, that, for $R = 0.6$, the measured cross section is approximately 5 % above the predicted cross section in most regions of transverse momentum. For $R = 0.4$ the behaviour is the other way round. Here the measured cross sections are consistently below the prediction, by approximately 3 %- 5 %. Considering that all systematic uncertainties can be assumed to be largely correlated, this is quite surprising. Possible reasons and more detailed studies of the correlations will follow in the next chapter.

The measurements in different rapidity regions are displayed and compared to the theoretical prediction in Fig. 9.17 ($R = 0.6$) and Fig. 9.18 ($R = 0.4$). The slopes of the cross sections generally decrease going to the more forward rapidity regions. This

effect becomes most pronounced going from the second-last rapidity bin to the last one, leading to a smaller reach in transverse momentum of up to only approximately 450 GeV. Taking a closer look at the ratios, the results obtained in the inclusive rapidity region are reproduced, with the exception being the amplified statistical fluctuations at highest p_T. These show a slightly alternating behaviour, going from one rapidity bin to another, which is however considered to be a purely statistical effect. Also here, the theoretical prediction for $R = 0.6$ slightly undershoots the measurement, while for $R = 0.4$ it lies above the measured cross section.

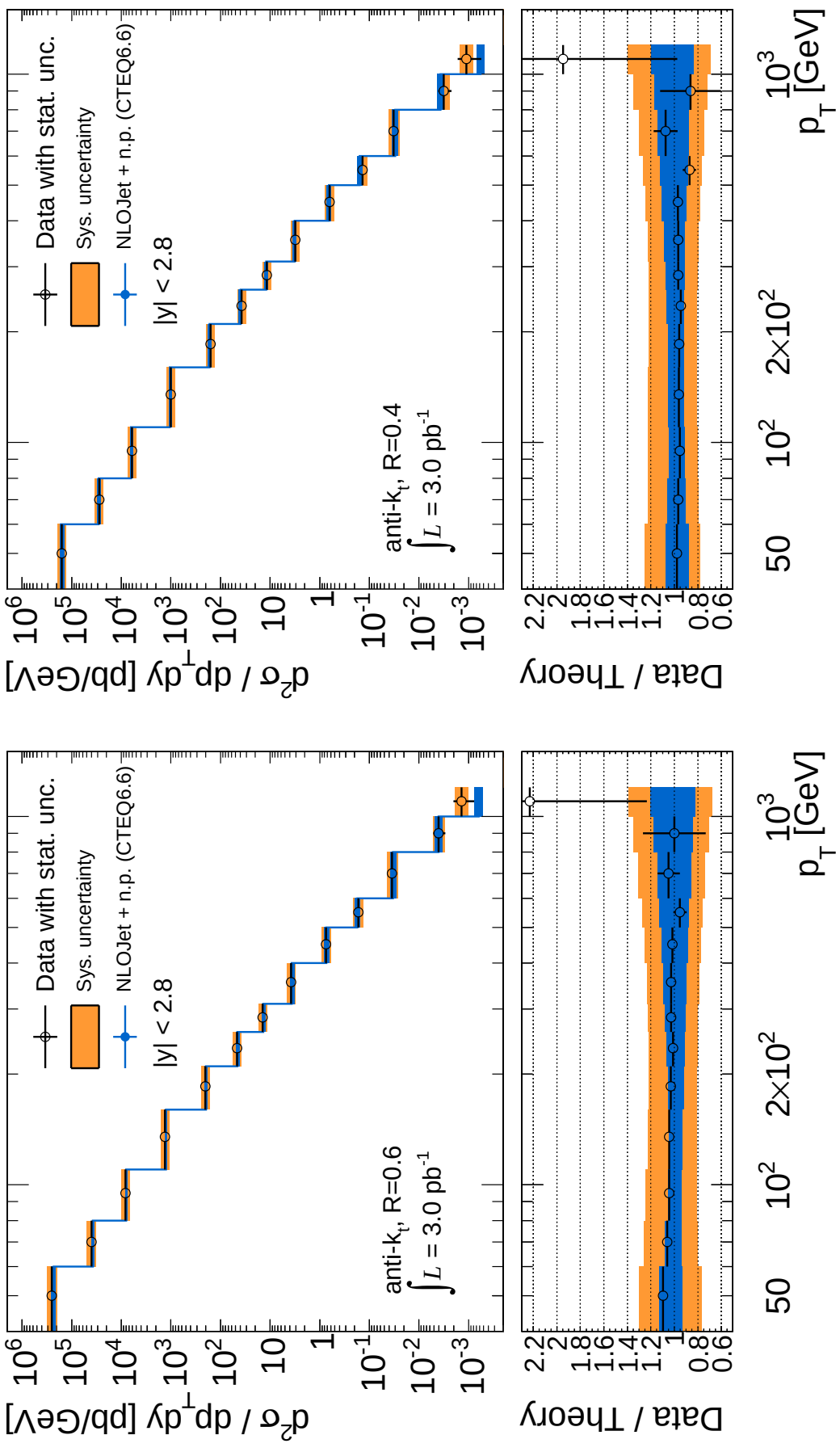

Figure 9.16: *Measured double differential cross section compared to NLO calculations with non. pert. corrections. In the ratios, the orange band shows the quadratic sum of the systematic uncertainties on data and theory (blue).*

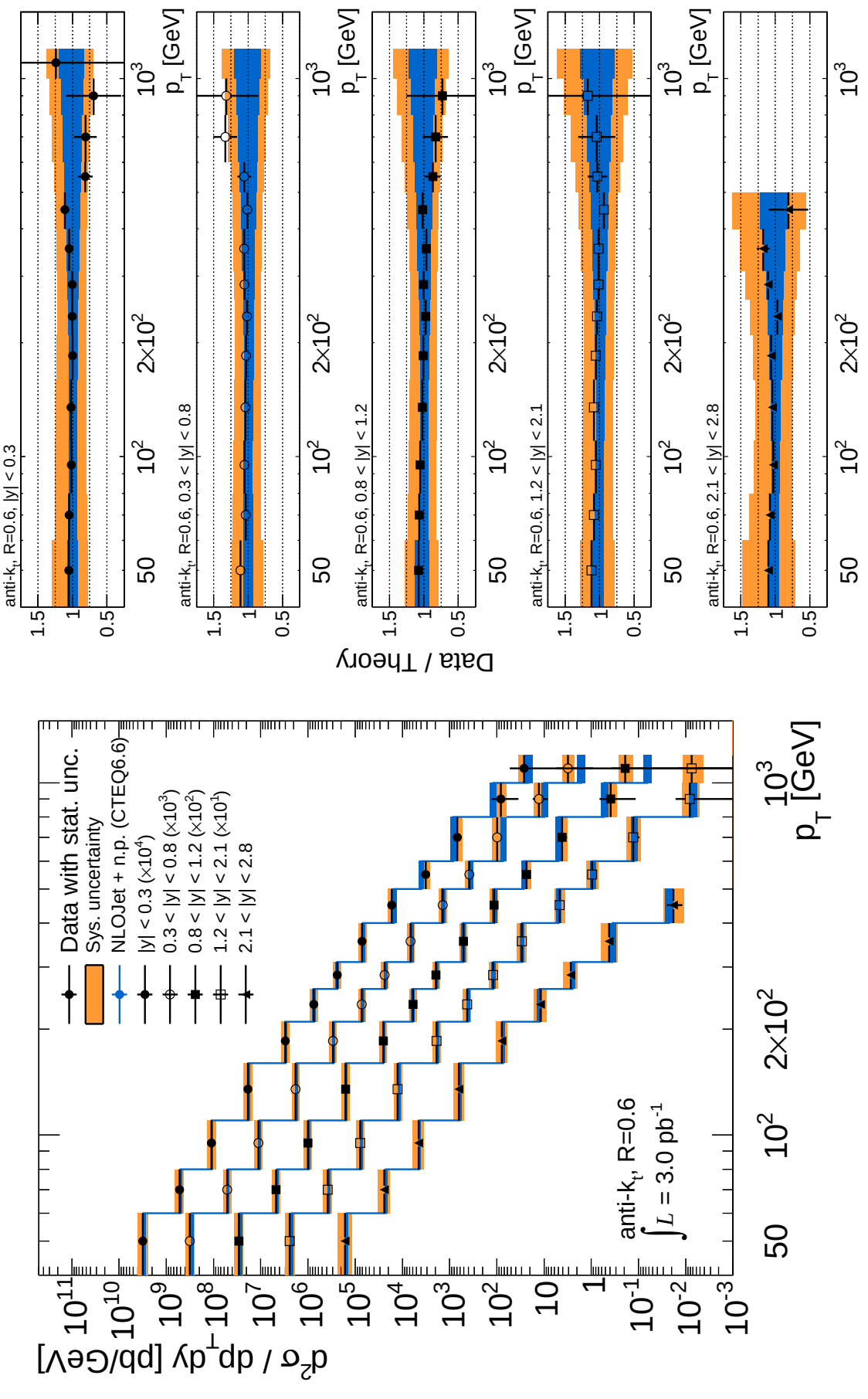

Figure 9.17: *Measured double differential cross section compared to NLO calculations with non. pert. corrections. In the ratios, the orange band shows the quadratic sum of the systematic uncertainties on data and theory (blue).*

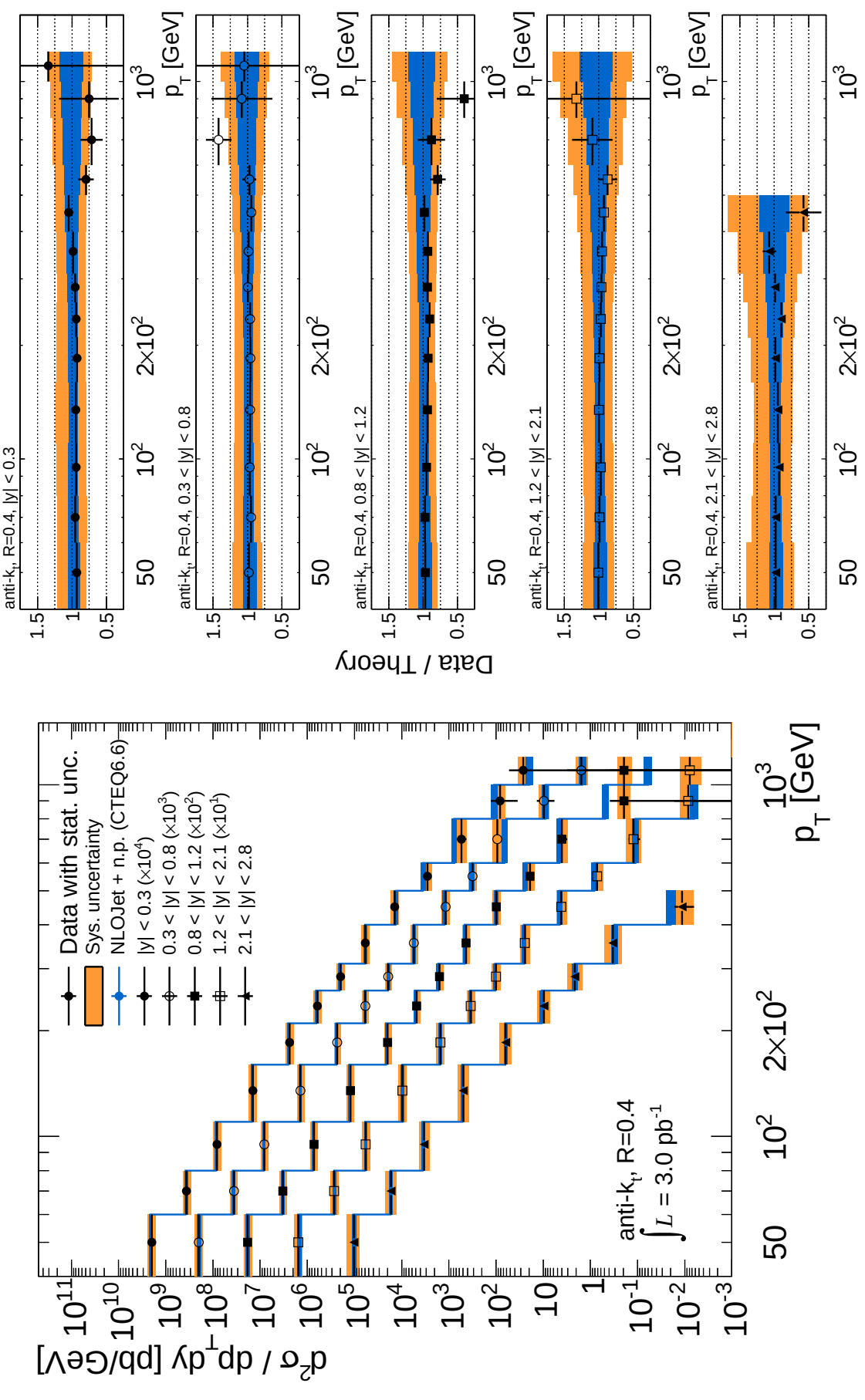

Figure 9.18: *Measured double differential cross section compared to NLO calculations with non. pert. corrections. In the ratios, the orange band shows the quadratic sum of the systematic uncertainties on data and theory (blue).*

The choice of the PDF is the major ingredient to the above comparisons to NLO theory. Hence, a comparison of the measured cross sections to the predictions using various different PDFs is interesting. For this purpose theoretical predictions were calculated with a small sample of alternative PDFs, using `APPLGRID`. The CT10 [83] PDF provided by the CTEQ collaboration is an updated version of the CTEQ6.6 PDF, which for instance includes the latest Tevatron inclusive jet cross section measurements that were not used in the derivation of the CTEQ6.6 PDF. The MSTW2008 PDF [84] are constructed similarly to the ones by the CTEQ collaboration, but for instance include the strong coupling constant as a free parameter, rather than as a fixed, external parameter. The neural network PDF NNPDF2.1 [85, 86] is quite different from the other PDFs, since it uses a very large number of free parameters ($\mathcal{O}(200)$) and thus is much more flexible in adapting to the input data. The authors also claim that this flexibility results in a more realistic estimate of the PDF uncertainties. The theoretical uncertainties (at 90 % CL) on each PDF set have been evaluated using the recommendend prescription of each collaboration, which are summarized in [87]. In Fig. 9.19 the predictions of the different PDFs, normalized to the CTEQ6.6 prediction are shown. Leaving the data out of the considerations at

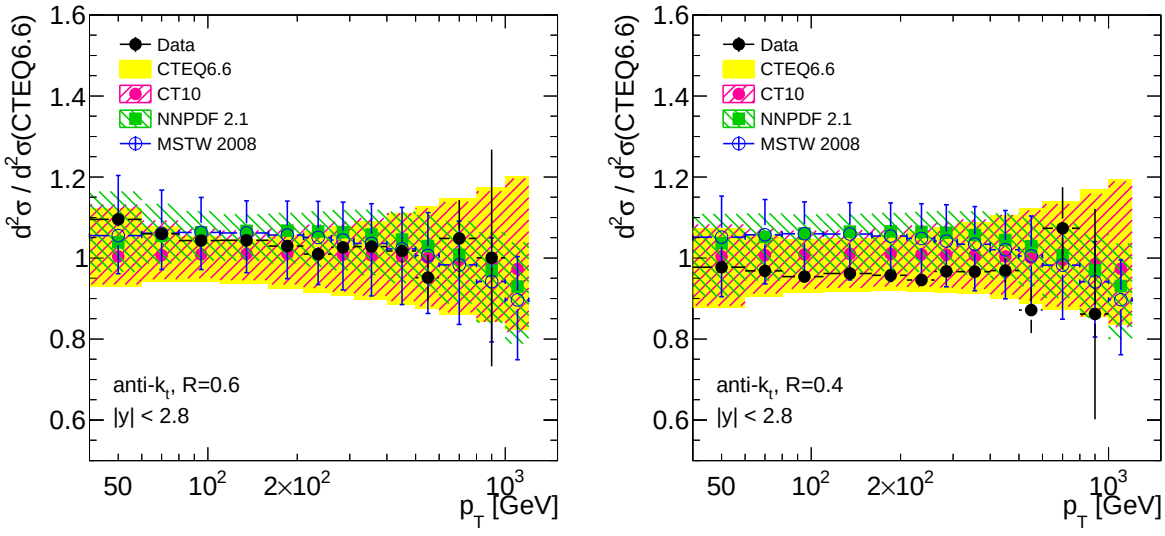

Figure 9.19: *Ratios of theoretical predictions using various PDFs for anti-k_t, $R = 0.6$ jets (left) and $R = 0.4$ (right). The measured cross sections and the theoretical predictions are divided by the theoretical prediction using the CTEQ6.6 PDF. The systematic uncertainties on data are not shown. Systematic uncertainties on the theoretical predictions were calculated at a confidence level of 90 %.*

first, the predictions of the various PDFs all agree within their systematic uncertainties. Only around transverse momenta of 1 TeV they start to deviate from each other, with NNPDF and the MSTW prediction showing a stronger decrease at highest p_T. The MSTW predictions agree best with data for $R = 0.6$, while for $R = 0.4$ all theoretical predictions overshoot the data. The data, being statistically powerful only up to approximately 600 GeV is not able yet, to clearly favour one PDF set over another. Comparisons to the various PDFs in the single rapidity regions are omitted here for brevity and instead moved into Appendix A.2.

10 Measurement of the cross section ratio

The inclusive cross sections are probing perturbative QCD and will, with increased precision, be useful for global PDF fits. In Section 9.5 it was already observed, that, while each measured cross section is within the uncertainties of the theoretical predictions, the comparison of data to theory is different for small and large jets. These observations are already motivation enough for a measurement of the ratio of jet cross sections with different radii, including a proper treatment of the correlations of experimental and theoretical uncertainties. Such an analysis can be expected to be able to tell, whether the observed differences are indeed compatible with the theoretical predictions or not.

The difference can be expected to arise from the underlying event and hadronization effects, which were shown to have a significant jet radius dependency. The ratio will provide a better handle to both effects than the single cross section measurements. This is due to the fact that uncertainties strongly affecting the overall normalization, e.g. the JES uncertainty and luminosity, are correlated to a large degree and thus will be decreased in the ratio. As will be seen in the following, this statement almost equally holds for the theoretical uncertainties. A proper treatment of the correlations is the main subject of the analysis described in this chapter.

Starting from the results obtained in Chapter 4, the theoretical predictions for the cross section ratio will be investigated first. Subsequently a thorough study of the correlations of the experimental uncertainties precedes the measurement of the ratio in data.

10.1 Theoretical predictions

The transition from single cross sections to the ratio, using the results from Chapter 4 is straightforward. At parton level the ratio can simply be calculated using the partonic cross sections already calculated and shown in Fig. 4.1:

$$\mathcal{R}_{parton}(p_T, y; 0.6, 0.4) = \frac{d^2\sigma_{parton}(p_T, y; 0.6)}{dp_T dy} \Big/ \frac{d^2\sigma_{parton}(p_T, y; 0.4)}{dp_T dy} \,. \tag{10.1}$$

This ratio is shown in Fig. 10.1 and ranges between 1.15 at low p_T and 1.1 at higher p_T. Since the ratio does only show a very little rapidity dependency, the analysis will in the following concentrate on the inclusive rapidity bin $|y| < 2.8$, in order to exploit the full statistics available. The ratio in this inclusive bin in the following is denoted as

$$\mathcal{R}_{parton}(p_T; 0.6, 0.4) = \frac{d\sigma_{parton}(p_T; 0.6)}{dp_T} \Big/ \frac{d\sigma_{parton}(p_T; 0.4)}{dp_T} \equiv \frac{\sigma_{parton}(p_T; 0.6)}{\sigma_{parton}(p_T; 0.4)}. \tag{10.2}$$

Figure 10.1 b shows the theoretical uncertainties on the ratio. As expected, the PDF and α_s uncertainties almost vanish, due to the correlation between the two jet algorithms. The scale uncertainties are not significantly reduced. This is due to the fact that in particular

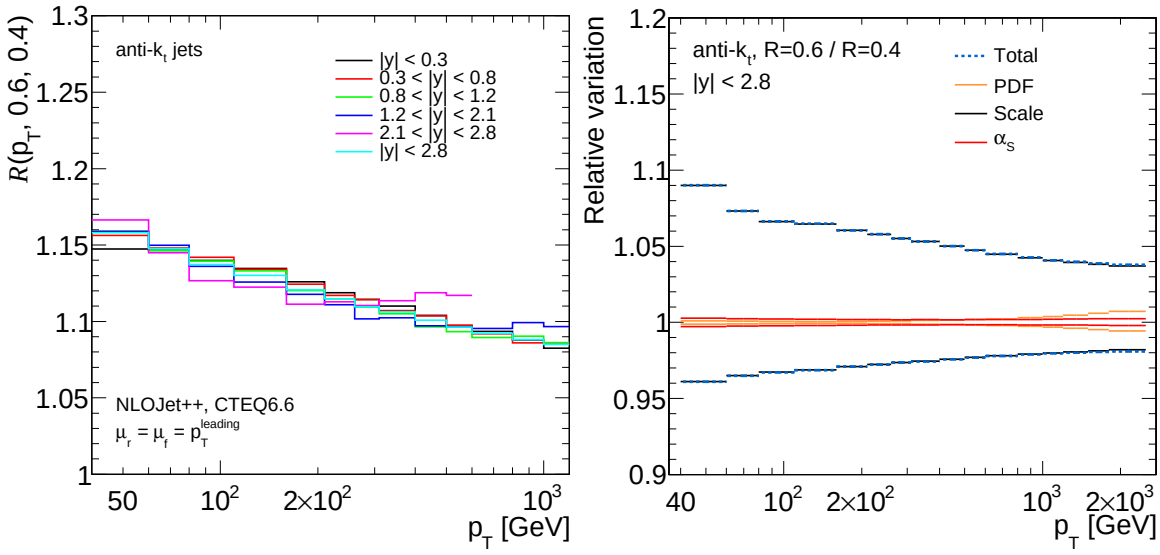

Figure 10.1: *Results for the cross section ratio at parton level (left), obtained from the results shown in Fig. 4.1 and the associated systematic uncertainties (right). The predictions for the ratio at high p_T in the most forward rapidity bin slightly suffers from non-optimal convergence of the MC integration performed by* `NLOJet++`.

the uncertainties arising from the change of the renormalization scale are significantly larger for $R = 0.4$ jets (cf. Fig. 4.2).

It is worthwhile noting that the computation of the cross section ratio in this way is not anymore at NLO, as it is the case for each single cross section. This is due to the fact that the LO contributions to the cross sections (the $2 \to 2$ parton processes) are independent of the jet radius. Thus the ratio is trivially one at this order. The first, non-trivial ratio arises from $2 \to 3$ parton configurations, which thus is the leading order. For the next higher order, the $\mathcal{O}(\alpha_s^4)$ contributions, particularly those with three or four partons in the final state, would have to be included. A computation at this order using Eq. 10.2 would thus require each single cross section to be known at NNLO.

The two-loop $\mathcal{O}(\alpha_s)^4$ contributions to the single cross section, which prevent their calculation at NNLO, do not contribute to the cross section ratio, as there are only two partons in the final state. Thus it is possible to calculate the ratio at NLO, including the zero- and one-loop $\mathcal{O}(\alpha_s^4)$ contributions. For this purpose, the perturbative expansion has to be done explicitly for the cross section ratio. This was done in [88] and is only briefly summarized below. The first step consists of going back to the general, perturbative expansion of the cross section with a jet radius R, that can be written as:

$$\sigma_{pQCD}(p_T; R) = \sum_{n=2}^{4} \alpha_s^n \sum_{p=0}^{n-2} \sigma^{(n,p)}(p_T; R) + \mathcal{O}\left(\alpha_s^5\right), \tag{10.3}$$

where $\sigma(p_T; R)$ is used as a shorthand for $\mathrm{d}\sigma(p_T; R)/\mathrm{d}p_T$. Here $\sigma^{(n,p)}$ denotes the $\mathcal{O}(\alpha_s^n)$ contribution with p loops. In the second step, the ratio of two cross sections,

$$\mathcal{R}_{pQCD}(p_T; R_1, R_2) = \sigma_{pQCD}(p_T; R_1) \,/\, \sigma_{pQCD}(p_T; R_2),$$

is itself expanded explicitly in powers of α_s. Exploiting the important feature, that contributions with only two partons in the final state are independent of the jet radius, this

becomes:

$$\begin{aligned}\mathcal{R}_{pQCD}(p_T; R_1, R_2) = 1 &+ \alpha_s \frac{\Delta\sigma^{(3,0)}(p_T; R_1, R_2)}{\sigma^{(2)}(p_T)} \\ &- \alpha_s^2 \frac{\sigma^{(3)}(p_T; R_2)\Delta\sigma^{(3,0)}(p_T; R_1, R_2)}{[\sigma^{(2)}(p_T)]^2}, \\ &+ \alpha_s^2 \frac{\Delta\sigma^{(4,0)}(p_T; R_1, R_2) + \Delta\sigma^{(4,1)}(p_T; R_1, R_2)}{\sigma^{(2)}(p_T)},\end{aligned} \tag{10.4}$$

with $\Delta\sigma^{(n,p)}(p_T; R_1, R_2) = \sigma^{(n,p)}(p_T; R_1) - \sigma^{(n,p)}(p_T; R_2)$ and $\sigma^{(n)}(p_T; R) = \sum_{p=0}^{n-2} \sigma^{(n,p)}(p_T; R)$. Most importantly, using this expansion, there is no two-loop contribution left, since $\Delta\sigma^{(4,2)}(p_T; R_1, R_2) \equiv 0$. This makes the calculation of $\mathcal{R}_{pQCD}(p_T; R_1, R_2)$ possible, for instance with `NLOJet++`, allowing to make predictions at next-to-leading order. Numerical results of such calculations presented in the following, are taken from Ref. [89]. The benefit of this procedure is, that Eq. 10.4 includes all $\mathcal{O}(\alpha_s^2)$ contributions, while the ones in the last line of Eq. 10.4 are missing in the calculation of the *direct ratio* (using Eq. 10.2).

10.1.1 Non-perturbative corrections

The necessary non-perturbative corrections to $\mathcal{R}_{parton}(p_T; R_1, R_2)$ also can be derived using the results from Chapter 4. The final prediction for the ratio at particle level can be written as:

$$\begin{aligned}\mathcal{R}(p_T; 0.6, 0.4) &= K^{(0.6/0.4)}(p_T) \frac{\sigma_{parton}(p_T; 0.6)}{\sigma_{parton}(p_T; 0.4)}, \\ \text{with} \quad K^{(0.6/0.4)}(p_T) &= \frac{K^{(0.6)}(p_T)}{K^{(0.4)}(p_T)}.\end{aligned} \tag{10.5}$$

With $K^{(0.6)}$ and $K^{(0.4)}$ being the non-perturbative corrections for the larger and smaller jet size respectively, taken from Fig. 4.5. The importance of the non-perturbative corrections for the ratio arises from the distinct jet radius dependency of the UE and hadronization effects: In a simplified picture the UE adds an isotropic amount of energy to the event. The amount of UE contribution to a jet thus increases with the jet's area.

In contrast, the hadronization correction arises from out-of-cone losses and consequently will decrease with increasing jet radius. In [39] it was established, that this correction is proportional to the inverse of the jet radius.

For the alternative Ansatz followed in [88], the UE and hadronization corrections are calculated in an analytical manner. Starting from a jet at parton level, UE and hadronization lead to a radius dependent, average shift of a jet's transverse momentum denoted by $\langle \delta p_T(R) \rangle$. The correction to the cross section, arising from such a p_T shift, then can be derived as:

$$K^{(R)}(p_T) = \frac{\sigma_{LO}(p_T + \langle \delta p_T(R) \rangle)}{\sigma_{LO}(p_T)}, \tag{10.6}$$

and the resulting correction for the cross section ratio thus is

$$K^{(0.6/0.4)}(p_T) = \frac{\sigma_{LO}(p_T + \langle \delta p_T(0.6) \rangle)}{\sigma_{LO}(p_T + \langle \delta p_T(0.4) \rangle)}, \tag{10.7}$$

where $\sigma_{LO}(p_T)$ is again calculated using `NLOJet++`. The usage of the leading order cross section here is motivated (as in [88]), by the usage of the LO cross section for the derivation of the hadronization correction in [39].

The hadronization corrections were derived as described in [88]. The shift in p_T is derived in [39] and found to be:

$$\langle \delta p_T(R) \rangle_{\mathrm{HAD}} = \frac{-2C_R}{R} \frac{2M}{\pi} \mathcal{A}(\alpha_0, p_T). \tag{10.8}$$

Here C_R is the Casimir factor, which is $C_F = 4/3$ for quarks and $C_A = 3$ for gluon jets, resulting in larger hadronization corrections for gluon jets, than for light jets. M is the so called Milan factor [90], which, for anti-$\mathrm{k_t}$ jets, is 1.49. $\mathcal{A}(\alpha_0, p_T)$ includes all the non-perturbative behaviour:

$$\mathcal{A}(\alpha_0, p_T) = \frac{\mu_I}{\pi} \left[\alpha_0(\mu_I) - \alpha_{\mathrm{s}}(p_T) - \frac{\beta_0}{2\pi} \left(\log\left(\frac{p_T}{\mu_I}\right) + \frac{K}{\beta_0} + 1 \right) \alpha_s^2(p_T) \right], \tag{10.9}$$

with $\beta_0 = (11C_A - 2n_f)/3$ and $K = C_A \left(\frac{67}{16} - \frac{\pi^2}{6}\right) - \frac{5}{9} n_f$. It is a function of the average strong coupling $\alpha_0(\mu_I)$ below a certain scale μ_I. This average, non-perturbative coupling is the only, external parameter which has to be taken from experiment in order to calculate the hadronization corrections. For the upcoming results, taken from Ref. [89], $\alpha_0(\mu_I)$ was taken from event shape studies [91].

For the UE, $\langle \delta p_T(R) \rangle_{\mathrm{UE}}$ can be approximated by the average energy density ρ_{UE} of the UE multiplied by the jet area. For anti-$\mathrm{k_t}$ jets, this area is simply that of a circle of radius R [39], leading to:

$$\langle \delta p_T(R) \rangle_{\mathrm{UE}} = \pi R^2 \rho_{\mathrm{UE}}, \tag{10.10}$$

A value for ρ_{UE} can be extracted from recent analyses of the UE by the ATLAS collaboration [30], already shown in Fig. 3.5. These studies only range up to transverse momenta of 14 GeV for the leading particle in the event. Based on these results, the provided results [89] estimated $\rho_{\mathrm{UE}} = (1.8 \pm 0.5)$ GeV. However, due to the limited momentum range one can expect a further, a priori unknown increase up to the momentum range analyzed in this study. This is reflected by a rather conservative uncertainty associated to ρ_{UE}.

Figure 10.2 shows the UE and hadronization corrections, using the analytical Ansatz described above and the various MC simulations already used in Chapter 4. The analytical results show a very good agreement with the results obtained from the much more complex MC simulations. This demonstrates that such corrections can in fact be computed in an analytical manner, requiring only two experimentally measured parameters. Additionally, only ρ_{UE} depends on the experimental environment, while α_0 is a universal parameter.

Looking at the MC results, it can be observed that there is no significant cancellation in the uncertainty of the non-perturbative corrections. The spread at lowest p_T, for the total correction, is on the order of 10 %, with `Herwig++` predicting the largest ratio between $K^{(0.6)}$ and $K^{(0.4)}$.

To summarize the results of this Section, $\mathcal{R}_{pQCD}(p_T; R_1, R_2)$ at LO (up to $\mathcal{O}(\alpha_{\mathrm{s}})$) and NLO (up to $\mathcal{O}(\alpha_s^2)$) is shown in Fig. 10.3 compared to the direct ratio. As expected the direct ratio lies between $\mathcal{R}_{pQCD}(p_T; R_1, R_2)$ at LO and NLO. The reason being that the $\mathcal{O}(\alpha_s^2)$ contributions with three partons and two partons plus one loop are included in the direct ratio. Nevertheless it can be observed that the full NLO computation significantly increases the ratio by approximately 8 % at low p_T and by approximately 4 % at the high

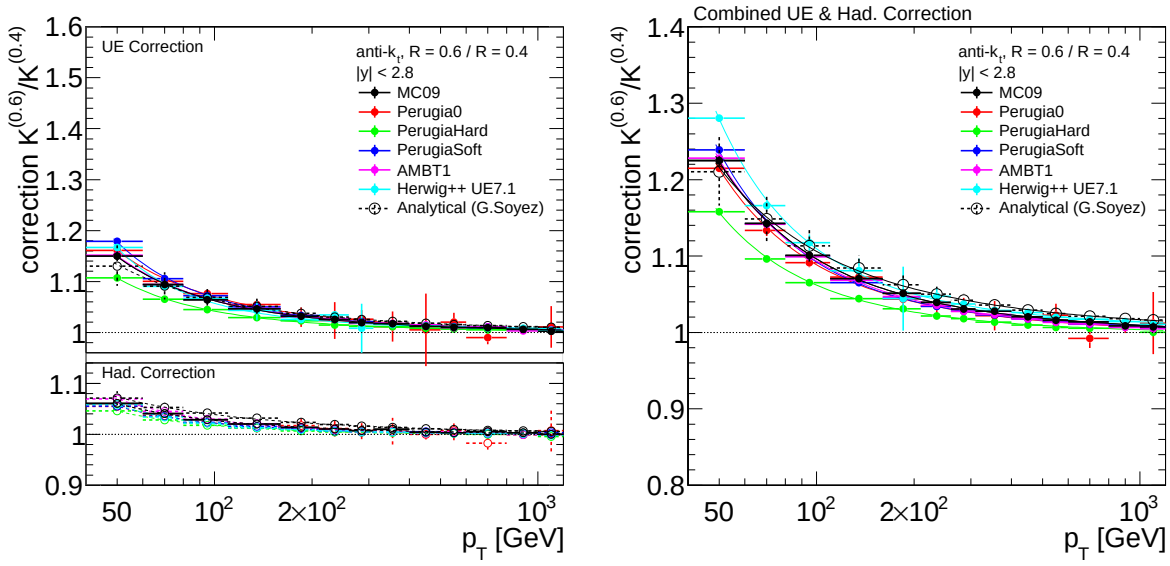

Figure 10.2: *Left: UE and hadronization corrections to the cross section ratio. Right: Combined corrections from various MC generators and tunes. The envelope of all predictions is taken as the systematic uncertainty. Analytical results from [89].*

p_T end. This can hence be concluded to arise from the $\Delta\sigma^{4,0}$ and $\Delta\sigma^{4,1}$ contributions in Eq. 10.4.

It is also striking that the theoretical uncertainty on the direct ratio, predominantly arising from the scale uncertainty, is multiple times larger than quoted for the LO and NLO calculations. This is expected to be due to the additional $\mathcal{O}(\alpha_s^2)$ contributions in Eq. 10.4, which improve the convergence of the perturbative series.

Figure 10.3 also confirms that the non-perturbative corrections have a significant impact on the jet cross section ratio. Their inclusion leads to a ratio of approximately 1.5 at lowest p_T and of 1.1 at highest p_T.

10.2 Experimental uncertainties and correlations

As mentioned in the beginning of this chapter the main challenge in the measurement of the ratio is a proper treatment of the systematic uncertainties and their correlations. The starting point for these studies is the experimental definition of the ratio, which is, in analogy to Eq. 10.2:

$$\mathcal{R}(p_T; 0.6, 0.4) = \frac{d\sigma(p_T; 0.6)}{dp_T} \Big/ \frac{d\sigma(p_T; 0.4)}{dp_T} \equiv \frac{\sigma(p_T; 0.6)}{\sigma(p_T; 0.4)}. \tag{10.11}$$

Looking back at the existing systematic uncertainties in Fig. 9.15, there are three non-trivial components whose correlations have to be treated: the JES, the unfolding correction and the event and jet selection efficiencies. The luminosity uncertainty can be excluded from the very start, since it is fully correlated and identical for the two cross sections.

The remaining uncertainties are studied in more detail in the following.

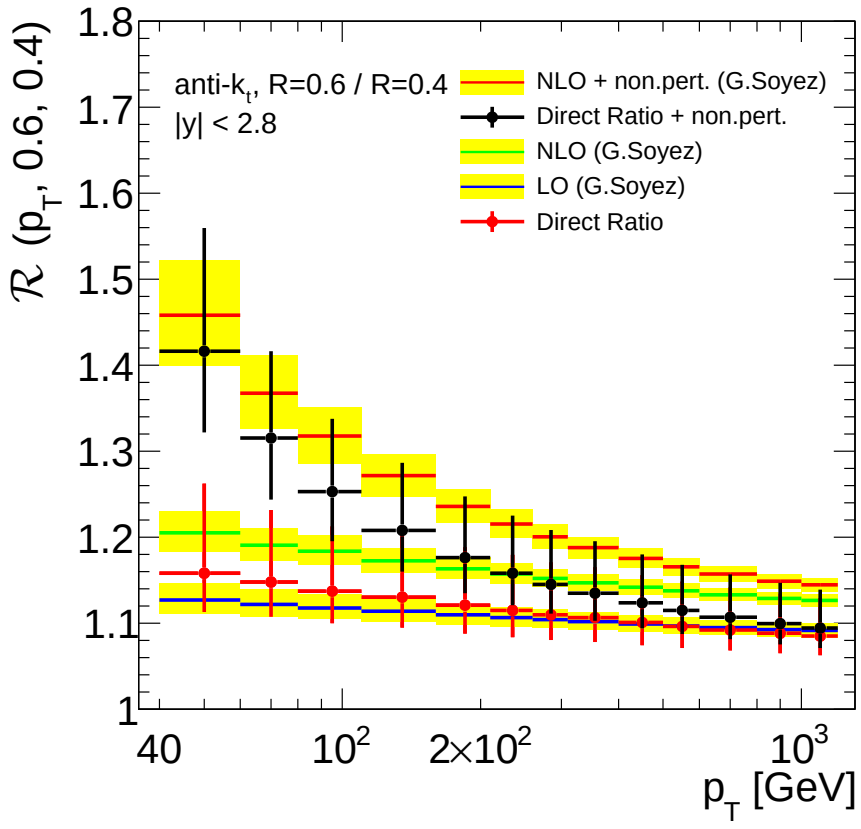

Figure 10.3: *Comparisons of the jet cross section ratio predictions taken from Ref. [89] (red, green, blue with yellow uncertainty bands) and the direct ratio (markers). The two upper histograms include the corrections from UE and hadronization (non. pert.).*

10.2.1 Event and jet selection efficiency uncertainties

The uncertainty on the event selection efficiency, which was shown to be very small, is taken as fully correlated between both cross sections and thus vanishes for the cross section ratio. This is a justified assumption since the single cross sections are both measured in the very same event sample. Additionally, the event selection cuts do not make use of jet quantities which could bias the event selection for either of the two cross section measurements. For the only exception, the $E_{T,Miss}^{Sig}$ selection, it was checked that the event sample selected using the leading $R = 0.4$ jet differs in a negligible amount from the default selection using the leading jet with $R = 0.6$.

Similarly the uncertainty of the jet selection efficiency is also taken to be fully correlated between the two cross section measurements. The only source for this uncertainty, the absent modeling of the jet quality in the MC, can be expected to not differ significantly between the two jet sizes.

10.2.2 Unfolding uncertainties

As the already unfolded cross sections are used to construct the ratio, this implicitly includes the application of unfolding factors to the ratio. These are simply the ratio of

the two single cross section unfolding corrections:

$$C^{0.6/0.4}(p_T, y) = \frac{C^{0.6}(p_T, y)}{C^{0.4}(p_T, y)}. \tag{10.12}$$

Having already exercised the variations of MC truth shape, jet energy resolution and jet rapidity resolution in Section 9.3.1, the same procedure is employed here to assess the impact of these variations on the cross section ratio. First of all it is instructive to see what the unfolding correction for the ratio, $C^{0.6/0.4}(p_T, y)$, actually look like (Fig. 10.4). The

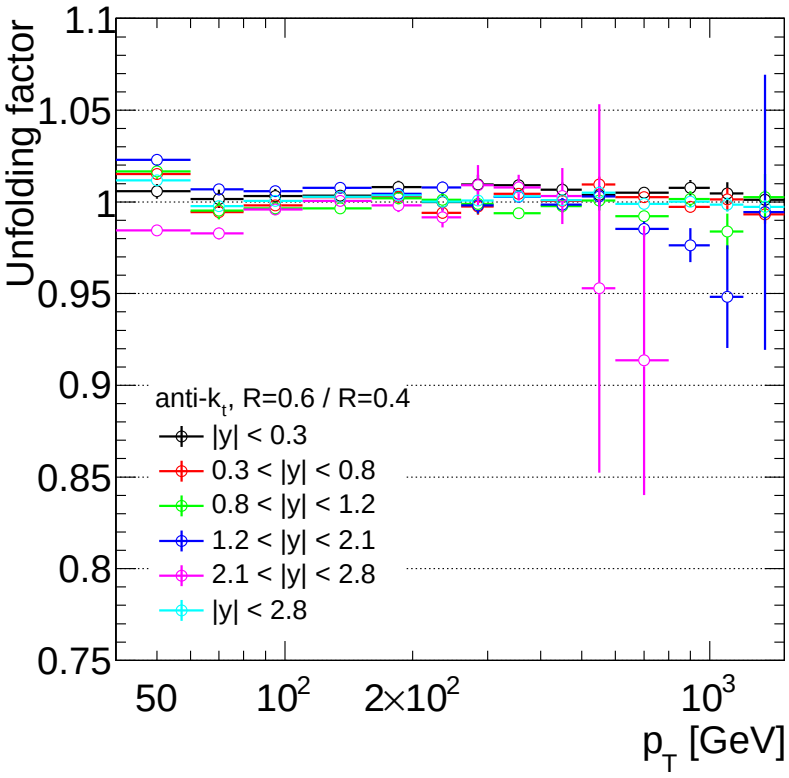

Figure 10.4: *Ratios of the unfolding factors presented in Fig. 9.9.*

impression of pretty similar unfolding corrections that can be gained from Fig. 9.9 here is confirmed quantitatively. The ratio of the two corrections is within ±1 % over most of the p_T range. This is not surprising since, as stated in Section 9.3.1, the unfolding correction is determined by the interplay between jet energy response (including resolution) and the simulated, true cross section shape. Both ingredients are very similar for large and small jets. For the truth cross section this can easily be understood, considering the fact that its shape is determined by pQCD and the proton PDFs. The jet algorithm only plays a very minor role. Concerning the resolution, it was already observed in Table 6.2 that large jets only have a very slightly better resolution. This is a result of the energy profile inside jets, which is strongly peaked at the jet's axis. The only effect by which the resolution of smaller jets may be degraded with respect to larger jets, is the increase of out-of-cone losses, which are of course more important, the smaller the jet's radius is.

The variations in $C^{0.6/0.4}(p_T, y)$ induced by changes of the truth cross section shape and jet energy resolution are shown in Fig. 10.5. The variations in shape have an impact well below 0.5 %. It should be mentioned that this only holds for a correlated change in the

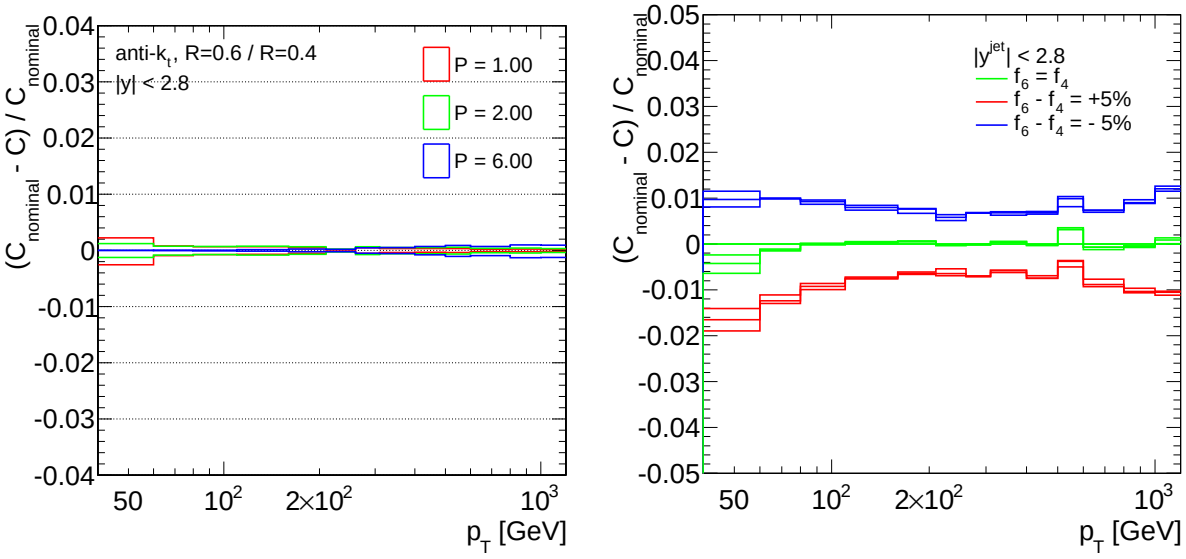

Figure 10.5: *Difference in unfolding corrections from shape reweighting (left) and from worsening the jet energy resolution (right). Details of both procedures are outlined in Section 9.3.1.*

truth cross section shapes for both jet sizes. However, following the considerations above, there is no sensible effect that could be expected to change the cross section for one jet definition in a significantly different way than for another.

Concerning the resolution uncertainty, the fact that the predominant energy fraction is deposited within the core of the jet, leads to the conclusion that the correlation can be expected to be very close to 100 %. However, the more conservative approach chosen here, is to vary the resolutions of both jet sizes independently to a certain degree. As the overall jet energy resolution uncertainty was estimated to be 10 %, using a difference in the variations of ±5 % can be considered a conservative choice, as it would imply a correlation of only 50 %. The increase of the jet energy resolution for the jet algorithm with size R is denoted as f_R. As in Section 9.3.1,

$$\{f_4\} = \{f_6\} = \{0\,\%, 5\,\%, 10\,\%, 15\,\%\}$$

is used and all combinations satisfying $|f_4 - f_6| \leq 5\,\%$ are examined. The resulting variations of the unfolding correction are shown in Fig. 10.5, which splits the total of nine combinations into three categories, where f_4 is either equal, larger or smaller than f_6. The differences between all three categories are on the 1 % level. The variations within each category are found to be negligible, except for the very first bin which shows variations on the order of 0.5 %.

Considering these results, a constant 1 % uncertainty is assigned on the unfolding for the ratio measurement.

10.2.3 Uncertainties due to pile-up

Pile-up may have an impact on the observed cross section ratio, since it can be expected to affect large jets stronger than smaller ones. Consequently, for a given pile-up scenario, the number of large jets can be expected to increase more than the number of small jets. Whether this effect is significant, is cross-checked in analogy to the already exercised stability studies in Fig. 7.16. Figure 10.6 shows the observed ratio of the number of large

and small jets, as a function of the average number of interactions per bunch crossing. No significant slope of the ratio as a function of μ can be observed. Hence it can be concluded, that there is no significant impact of pile-up on the measured cross section ratio.

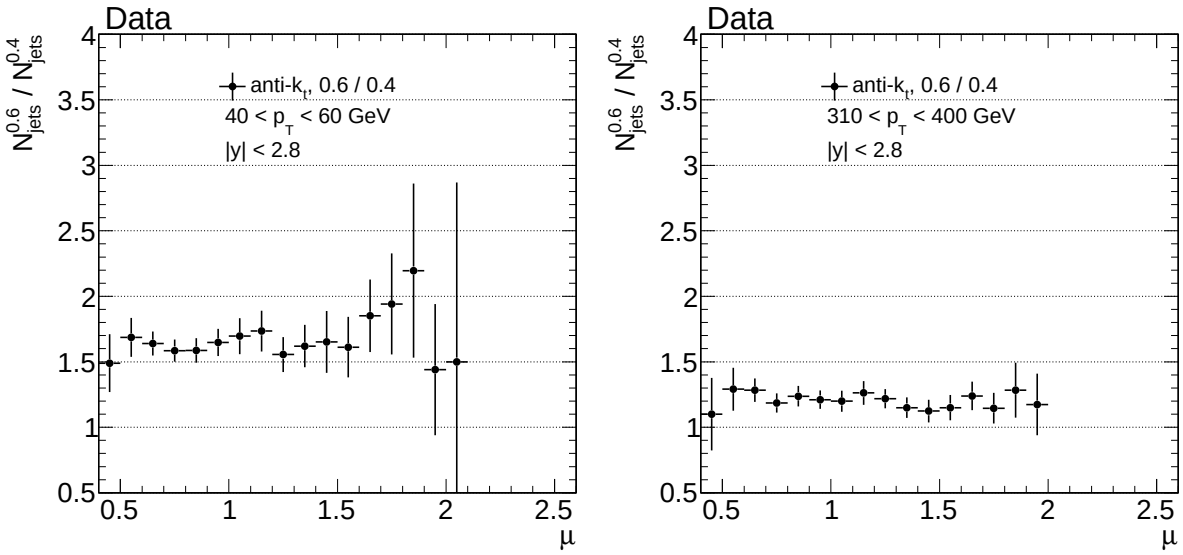

Figure 10.6: *Stability of the observed ratio of the number of large and small jets, for two different bins of transverse momentum. The indicated χ^2/NDF values are the result of fitting a constant to the observations.*

10.2.4 Jet energy scale uncertainties

The for sure existing, but a priori unclear correlation between the JES uncertainties for large and small jets is the most important ingredient to the measurement of the cross section ratio. As observed in the previous chapter, the JES uncertainty is the predominant contribution to the overall uncertainty for each jet cross section.

Dealing with a systematic uncertainty, there is unfortunately no direct statistical method to get a handle on the correlation. Instead the correlation will be estimated using the two approaches described in the following. The observed agreement of both results will allow for an estimate of the inherent uncertainties and advantages of one method over the other.

Decomposition of JES uncertainty components

The first approach uses a decomposition of the total JES uncertainty on the cross section measurement into its various components. Each contribution, as provided by [75], will either be taken as fully correlated or fully uncorrelated, based upon physical reasoning. The overall correlation, ρ_{JES} can then be determined as:

$$\rho_{JES} = \frac{\sum_i \rho_i \sigma^i_{0.4} \sigma^i_{0.6}}{\sigma_{0.4}\sigma_{0.6}} \qquad \text{with} \qquad \sigma_R = \sqrt{\sum_i (\sigma^i_R)^2}, \tag{10.13}$$

where σ^i_R denotes the uncertainty arising from the i^{th} contribution on the cross section measurement with radius R, σ_R is the total JES uncertainty on the cross section and $\rho_i \pm 1$

or $\rho_i = 0$. Below and in Table 10.1 the single uncertainties and their assumed correlations are listed.

- **Detector Modeling:** JES uncertainties due to effects of detector simulation are taken as fully correlated, since jets with both radii are measured within the identical detector. This in particular applies to the dominant calorimeter uncertainty, but also to the uncertainties due to a varying beamspot or alternative noise thresholds.
- **Underlying Event & Pile-up:** The inclusion of the Perugia tune is meant to estimate the impact of a different UE modeling. This modeling is independent of the jet algorithm used for this analysis. The same considerations apply to the pile-up uncertainty*. Thus these uncertainties are taken as fully correlated.
- **Fragmentation:** The Professor tune is expected to reflect the change in response due to varying fragmentation. The fragmentation effectively controls the width of jets and thus e.g. the out of cone losses. A given fragmentation model might describe these effects well for large jets, but not for smaller ones. For this reason these uncertainties are assumed to be uncorrelated.
- **Alpgen-MC:** Since the Alpgen MC employs different models at all levels of the event generation, the correlation of the induced JES uncertainty cannot be quantified. For this reason the conservative choice of $\rho = 0$ was made.
- **Intercalibration Results:** The intercalibration results are different from the remaining uncertainties. They cannot be assumed to probe only a distinct aspect of jet energy response, since they are using real data. For this reason they do not fit into the decomposition scheme exercised with MC and are excluded from these studies.

Though the η intercalibration results are not considered in the estimate of the correlation, they are included in the determination of the actual uncertainty.

Source	Correlation ρ_i [%]	Uncertainty [%]
Calorimeter	100	20 - 30
Noise thresholds	100	0
Beamspot	100	2
Perugia tune (UE)	100	3
Pile-up	100	<4
Professor tune (Fragmentation)	0	2
Alpgen MC	0	2 - 10
η-intercalibration	Not included	-
MC-intercalibration	Not included	-

Table 10.1: *Contributions to the total JES uncertainty on a single cross section measurement.*

The overall correlation calculated this way, ranges between 85 % at the lowest p_T and approximately 100 % towards the high p_T limit, as displayed in Fig. 10.7. The correlation for the most central bin is strikingly low compared to the remaining rapidity regions. This is due to the fact that the JES uncertainty arising from the Alpgen-MC in this

*In addition, pile-up has a negligible effect on the measured ratio, as displayed in Fig. 10.6.

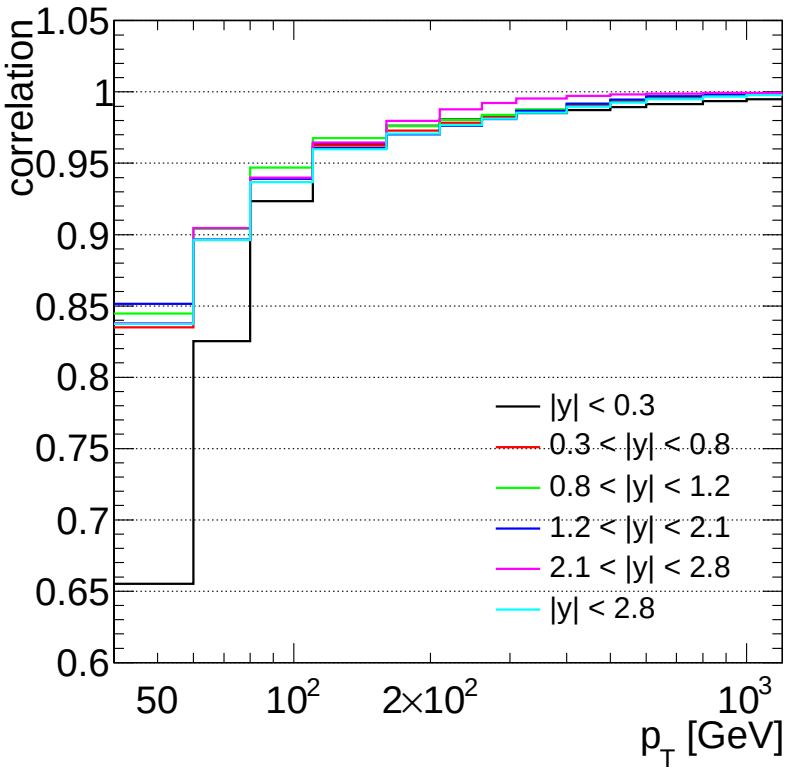

Figure 10.7: *Correlation of the jet energy scale uncertainties between large and small jets, with the assumed correlations in Table 10.1.*

region happens to be larger. As this effect does not exist in any other rapidity bin this is considered a fluctuation in the estimate of the JES uncertainty.

Except for this effect, the picture is as expected from general considerations: In an idealized picture a large jet ($R = R_1$) can be imagined as the sum of a small jet ($R = R_2$) plus an annulus of width $R_1 - R_2$. The source of any uncorrelated energy response thus can only arise from the energy deposited in this annulus. As the energy profile of jets becomes more narrow with increasing momentum, this fraction of energy decreases, thus increasing the correlation.

Jet response correlation

An alternative, statistically driven method starts from the definition of the jet energy scale uncertainty itself. Based upon Eq. 8.1, Δ^4_{JES} and Δ^6_{JES} denote the JES uncertainties of $R = 0.4$ and $R = 0.6$ jets respectively. The correlation between the two can be written as:

$$\mathrm{corr}\left(\Delta^{(0.4)}_{JES}, \Delta^{(0.6)}_{JES}\right) = \mathrm{corr}\left(1 - \frac{\mathtt{R}^{(0.4)}_{var}}{\mathtt{R}^{(0.4)}_{nom}}, 1 - \frac{\mathtt{R}^{(0.6)}_{var}}{\mathtt{R}^{(0.6)}_{nom}}\right). \tag{10.14}$$

Here the nominal response, e.g. $\mathtt{R}^{(0.4)}_{nom}$, is considered to be fixed and the varied responses $\mathtt{R}^{(0.4)}_{var}$ and $\mathtt{R}^{(0.6)}_{var}$ are treated as the *random variables*. For correlations of random variables X and Y however the following equation holds:

$$\mathrm{corr}\left(a_1 + b_1 X, a_2 + b_2 Y\right) = \mathrm{corr}\left(X, Y\right). \tag{10.15}$$

Thus Eq. 10.14 can be simplified to

$$\mathrm{corr}\left(\Delta_{JES}^{(0.4)}, \Delta_{JES}^{(0.6)}\right) = \mathrm{corr}\left(\mathtt{R}_{var}^{(0.4)}, \mathtt{R}_{var}^{(0.6)}\right). \tag{10.16}$$

Standard estimators of correlation coefficients cannot deal with single random variables, therefore the mean values used here are substituted by the distribution of single jet measurements from which they are derived: $\mathtt{R}_{var}^{(0.4)} = 1/N_{jets}\sum_{jets}\mathtt{R}_{jet}^{(0.4)} = \left\langle \mathtt{R}_{jet}^{(0.4)} \right\rangle$ and for $R = 0.6$ jets respectively. Taking this into account the correlation becomes:

$$\mathrm{corr}\left(\Delta_{JES}^{(0.4)}, \Delta_{JES}^{(0.6)}\right) = \mathrm{corr}\left(\left\langle \mathtt{R}_{jet}^{(0.4)} \right\rangle, \left\langle \mathtt{R}_{jet}^{(0.6)} \right\rangle\right) = \mathrm{corr}\left(\mathtt{R}_{jet}^{(0.4)}, \mathtt{R}_{jet}^{(0.6)}\right) \equiv \rho_{JES}. \tag{10.17}$$

Here, due to the linearity of covariances, the correlation of averages can be replaced by the correlation of the elements averaged over. Thus the correlation of the average response is identical to the correlation between $\mathtt{R}_{jet}^{(0.4)}$ and $\mathtt{R}_{jet}^{(0.6)}$ on a jet-by-jet basis, denoted by ρ_{JES}.

Since these correlations might depend on the MC used to extract them, they will be studied in a variety of simulations.

This is done in analogy to the jet calibration described in Chapter 6, starting with the nominal `Pythia6` MC09 MC sample. In each event large and small jets at particle and reconstructed level are matched to each other, thus yielding four versions of a jet and the two response values:

$$\mathtt{R}_{jet}^{(0.4)} = E_{reco}^{(0.4)}/E_{true}^{(0.4)} \qquad \text{and} \qquad \mathtt{R}_{jet}^{(0.6)} = E_{reco}^{(0.6)}/E_{true}^{(0.6)}. \tag{10.18}$$

Only isolated topologies are considered (as introduced in Section 6.3.3), since the different split and merge behaviour for large and small jets would compromise the comparison between the two. Distributions of $\mathtt{R}_{jet}^{(0.4)}$ versus $\mathtt{R}_{jet}^{(0.6)}$ are shown in Fig. 10.8, revealing

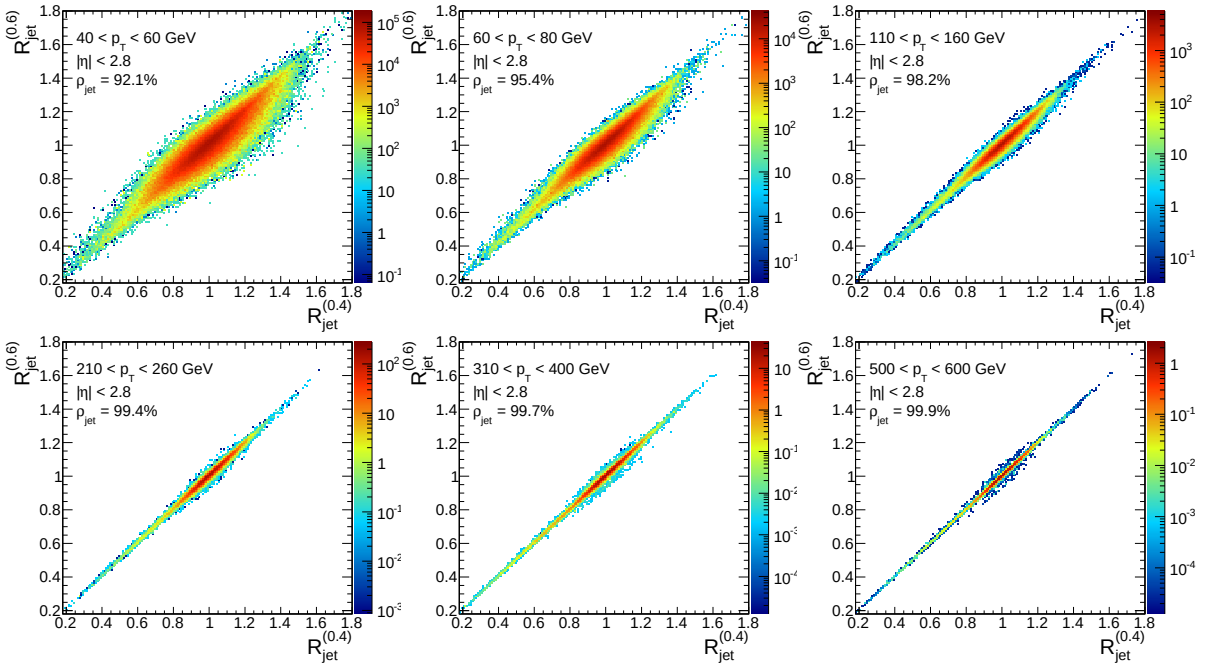

Figure 10.8: *Distributions of $R_{jet}^{(0.6)}$ versus $R_{jet}^{(0.4)}$ in various bins of p_T. Also quoted is the correlation within each histogram. Units on the z-axis are arbitrary.*

an increasingly high correlation for increasing transverse momentum regions. While at

low p_T there is a very significant spread around $\mathtt{R}_{jet}^{(0.6)} = \mathtt{R}_{jet}^{(0.4)}$, this vanishes very rapidly, leaving an almost perfect straight line with negligible spread above transverse moment of approximately 500 GeV.

As already mentioned it is necessary to ascertain that these correlations are equal in various MC samples. Thus the above studies are also exercised using different MC samples, including the ones used to assess the JES uncertainty. Figure 10.9 summarizes the results from all MC samples and confirms that the correlations within varying MC samples are very similar. The spread in the MC predictions at low p_T is on the order of 1 %. The

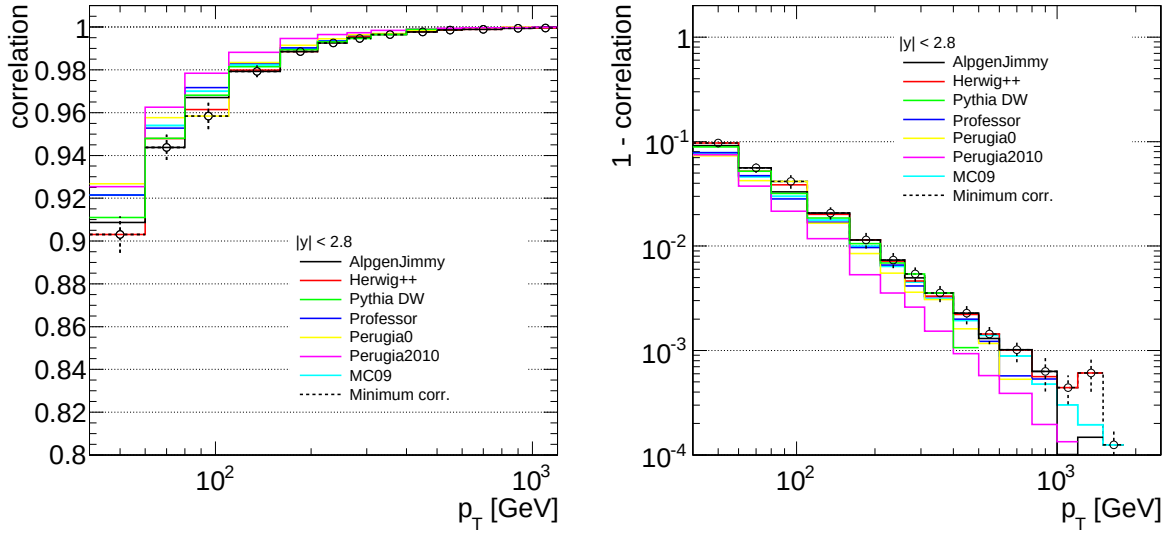

Figure 10.9: *Correlation ρ_{JES} as a function of transverse momentum as predicted by various MC simulations. The error on the minimum displayed in each bin was chosen as the standard deviation of the MC predictions.*

`Pythia6` Perugia2010 tune systematically predicts higher correlations than all other MC predictions. The correlation at lowest p_T values is found to be around 90 %, quickly rising towards almost 100 % for higher momenta. The minimum correlation found in each p_T bin is used as the result from this procedure.

Comparison of correlations

Finally, the correlations extracted with both methods are compared in Fig. 10.10. It is striking that the correlation found using the jet response is consistently and significantly higher than the estimate using the JES uncertainty decomposition. Especially at high p_T this can be explained by the limited precision to which the single JES uncertainty components reflect the true uncertainties arising from certain sources. This precision is firstly limited by the statistical uncertainties that enter the comparison between nominal and varied JES in the determination of the JES uncertainty [74]. Additionally, the JES uncertainties are simply taken to be the absolute difference in the JES, without an attempt to smooth out statistical fluctuations. Thus a determination of such uncertainties on the sub-percent level cannot be expected. It would however be necessary in order to resolve correlations to a similar precision. This becomes even more important as the JES uncertainties from Fig. 8.3 are amplified by a factor of approximately five when they are propagated to the jet cross section. An example for this is the underlying event uncertainty

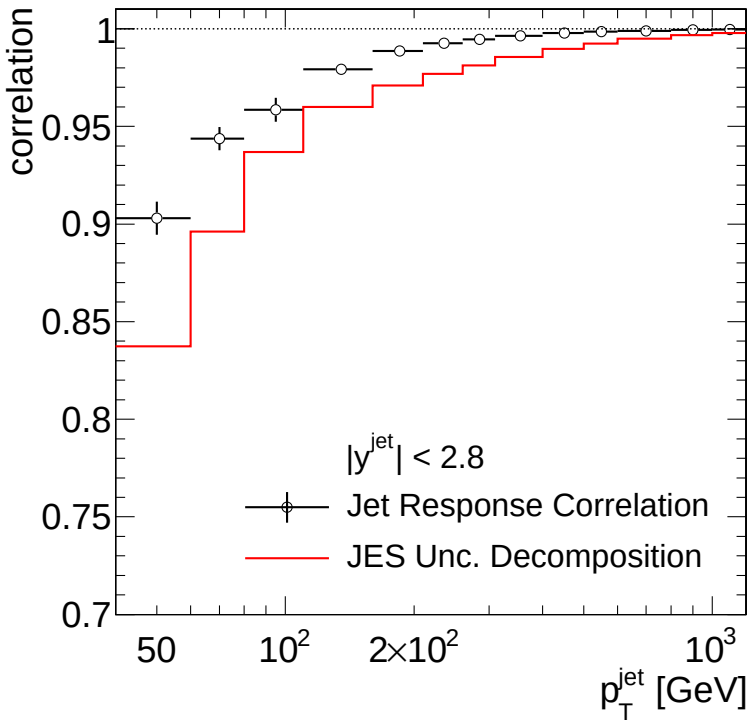

Figure 10.10: *Comparison of correlations extracted with the two indicated methods. The error bars on the jet response correlation points indicate the spread of the various MC samples studied.*

which is estimated using the Perugia0 tune. One can expect that the underlying event in fact plays an negligible role for jets in the TeV range. Nevertheless the uncertainty propagated from Fig. 8.3 is rather constant at approximately 2 %. To conclude, these uncertainties should rather be interpreted as upper limits.

In the low p_T region the significant JES difference in the Alpgen MC sample was taken as fully uncorrelated. This was an intentionally conservative choice, giving rise to a significant contribution with no correlation.

These considerations render the correlation extracted with the JES uncertainty decomposition an approximate lower limit. For this reason the response correlations obtained with the second method are used in the following.

10.2.5 Total systematic uncertainties

Using the correlations as estimated above, the systematic uncertainty on the ratio is propagated from the single cross section uncertainties via standard error propagation, starting from Eq. 10.11.

$$\left[\frac{\delta\mathcal{R}_{sys}(p_T,0.6,0.4)}{\mathcal{R}(p_T,0.6,0.4)}\right]^2 = \left(\delta^{(0.6)}_{JES}(p_T)\right)^2 + \left(\delta^{(0.4)}_{JES}(p_T)\right)^2 - 2\rho_{JES}(p_T)\,\delta^{(0.4)}_{JES}(p_T)\,\delta^{(0.6)}_{JES}(p_T) + \delta^2_{Unfolding}. \tag{10.19}$$

Here $\delta^{(0.6)}_{JES}$ and $\delta^{(0.4)}_{JES}$ denote the relative JES uncertainty on the jet cross sections (see Fig. 9.15) and the relative unfolding uncertainty, $\delta^2_{Unfolding}$, derived in Section 10.2.2 is added in quadrature. The resulting, total systematic uncertainties for the jet cross

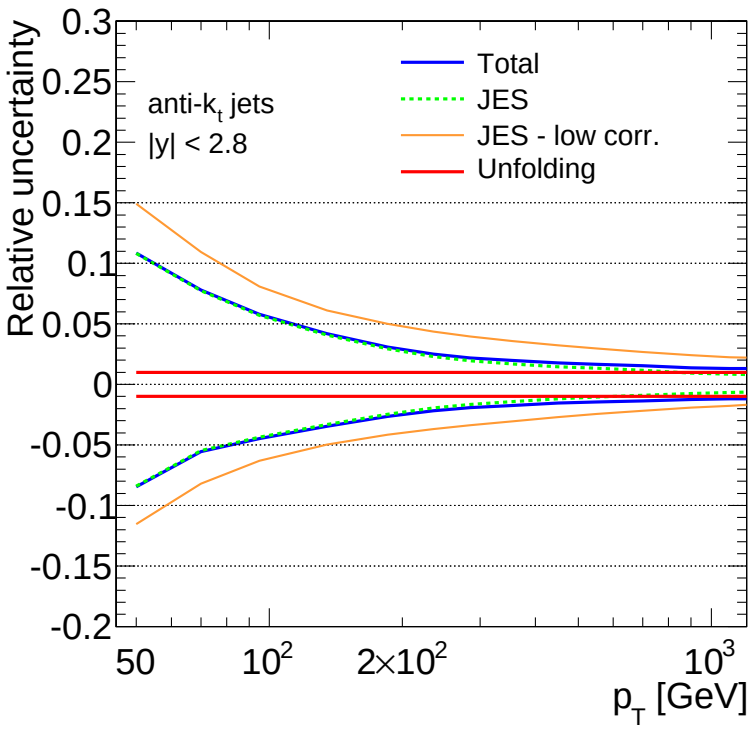

Figure 10.11: *Systematic uncertainties on the jet cross section ratio. For completeness the JES uncertainties using the correlations from Fig. 10.9 are shown as well (solid, orange line).*

section ratio is shown in Fig. 10.11. Compared to Fig. 9.15, there is a significant decrease in the uncertainties compared to the measurement of a single cross section. The total uncertainty is approximately 10 % at lowest p_T values, decreasing down to 2 % at 1 TeV. The uncertainty resulting from the usage of the more conservative choice for the correlation is significantly larger, namely between 15 % at lowest p_T and approximately 4 % at high p_T.

10.2.6 Statistical correlation

From Fig. 10.11 it can be concluded that the systematic uncertainties will dominate in the low p_T ranges. Looking back at the results in Fig. 9.19, significant statistic uncertainties start to occur at transverse momenta above 400 GeV. In analogy to Section 10.2.5, the statistical uncertainty on the ratio is given by:

$$\left[\frac{\delta\mathcal{R}_{stat}(p_T,0.6,0.4)}{\mathcal{R}(p_T,0.6,0.4)}\right]^2 = \left(\delta_{stat}^{(0.6)}(p_T)\right)^2 + \left(\delta_{stat}^{(0.4)}(p_T)\right)^2 - 2\rho_{stat}(p_T)\,\delta_{stat}^{(0.4)}(p_T)\,\delta_{stat}^{(0.6)}(p_T). \tag{10.20}$$

It is thus necessary to also study the statistical correlation ρ_{stat} between the two jet algorithms. This correlation boils down to the correlation of the number of $R = 0.6$ and $R = 0.4$ jets in a given (p_T, y) bin: $\mathrm{corr}\left(N^{(0.6)}, N^{(0.4)}\right)$. It can be estimated with a series of pseudo-experiments, exploiting the quite good statistics of the nominal MC sample. For this purpose the full MC sample is split into 100 sub-samples. Denoting the number of large (small) jets in the i^{th} sub-sample by $N_i^{(0.6)}$ ($N_i^{(0.4)}$), this yields a series of

pairs $\{N_i^{(0.6)}, N_i^{(0.4)}\}$. Using these 100 pairs of numbers the correlation between the two quantities can easily be calculated.

Unfortunately the uncertainty on the correlation using only 100 pseudo-experiments is relatively large. To gain an impression of the fluctuations in the correlation the above procedure is repeated 20 times. The randomized composition of the 100 sub-samples is altered each time. Each repetition j then yields an estimate ρ_j for the correlation:

$$\rho_j = \mathrm{corr}\left(\{N_i^{(0.6)}\}_j, \{N_i^{(0.4)}\}_j\right). \tag{10.21}$$

The result of this procedure is displayed in Fig. 10.12, which shows the 20 correlation values extracted from each set of pseudo-experiments. In each bin the mean of the cor-

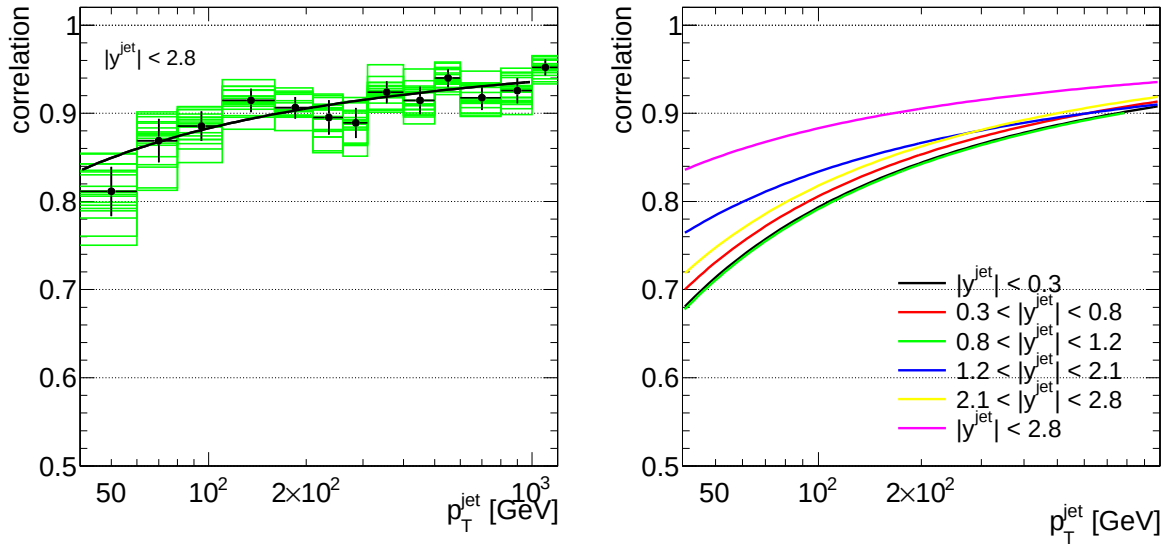

Figure 10.12: *Left: Statistical correlation between the number of* 0.6*-jets and* 0.4*-jets. Black markers indicate the mean of the correlations and their standard deviation is quoted as the uncertainty on the mean. The solid line shows the fit to the means. Right: Fit results in the various bins of rapidity.*

relations is shown as well, its uncertainty assigned being the standard deviation of the observed correlations. The expected, relatively large spread in the correlations ρ_j is confirmed: the standard deviations are of the order of 4 % at low p_T, decreasing towards a few percent at high p_T. A fit using a polynomial of inverse logarithms is used to smooth out remaining statistical fluctuations. This procedure yields minimum and maximum correlations of approximately 70 % and 90 % respectively in all pseudorapidity bins and across the observed p_T region.

One can also observe that the correlation in general is higher in the inclusive rapidity bin than in each single rapidity bin. This can be considered a real effect: Every energy deposition in the calorimeter above a certain threshold will for sure end up being a jet constituent, irrespective of the jet algorithm. As a consequence, starting from a $R = 0.4$ jet, there will always be a corresponding a $R = 0.6$ jet sharing at least one constituent with the smaller jet. It is then a matter of how much constituents the two jets share, whether the corresponding larger jet ends up in the same (p_T, y) bin and thus contributes to a higher correlation. As the probability for ending up in the same bin increases with the bin-width, the correlation increases as well. On a smaller scale this feature can also be

observed for the bin-width in p_T. Due to the larger statistical uncertainties, it is however considered reasonable to smooth out these fluctuations using the fit as described above.

10.3 Results

After the discussion of the theoretical predictions and the uncertainties on data, the results can be examined. In Fig. 10.13 the results obtained in data are compared with the theoretical predictions. The direct ratio predicts a consistently smaller ratio than observed in data across all p_T bins. The significant uncertainties on this calculation cover these differences, especially in combination with the experimental uncertainties. Here, it should also be taken into account that the experimental and theoretical uncertainties are to a large degree correlated among the p_T bins. $\mathcal{R}_{pQCD}(p_T; 0.6, 0.4)$ is in slightly better agreement with the data and at the same time its theoretical uncertainty is much smaller as well. This favours the calculations suggested in [88] over the naive approach of taking the direct ratio. In turn, this confirms the necessity of taking into account the NLO contributions to the ratio.

These issues also highlight that the inclusion of these contributions would in very principle be necessary in order to achieve a coherent description of jet cross sections of variable jet radii.

The measurement is limited by the experimental uncertainties up to approximately 500 GeV, where statistic uncertainties become significant.

Besides the comparison to theoretical predictions it is also interesting to compare the measurement to the predictions of various MC generators as shown in Fig. 10.14. These include the generators and tunes that have already been used to derive the non-perturbative corrections. The comparison to data here might help to disqualify certain tunes and thus decrease the uncertainty on the non-perturbative corrections. A first observation to make is, that the `Pythia6` PerugiaSoft tune does not describe the data across a wide range in p_T. The deviations become especially pronounced at high transverse momenta where the experimental uncertainties shrink. Better descriptions are achieved using the Perugia0 and PerugiaHard tunes. The best agreement within the `Pythia6` tunes can be observed for the AMBT1 tune. This is expected since this tune was obtained from early ATLAS minimum bias data. Nevertheless all these tunes are consistently below data.

The overall best description is provided by the `Herwig++` UE tune, which also included ATLAS UE data. It almost perfectly fits the data across the whole range of p_T.

10.4 Prospects for a measurement of ρ_{UE} and α_0

The theoretical predictions at NLO described in Section 10.1 make use of two parameters, which are provided by experiments. The first one is ρ_{UE}, the transverse momentum density of the UE which enters the according correction (Eq. 10.10). Secondly, there is α_0, the average coupling in the non-perturbative regime, which is part of $\mathcal{A}(\alpha_0, p_T)$ in Eq. 10.8. The qualitatively very good agreement of the theoretical predictions with data suggest, that the measurement can be used to measure both parameters entering the predictions. Since the NLO parton-level prediction is fixed, such an analysis would essentially measure the non-perturbative corrections on top of the parton level model. Subsequently, α_0 and ρ_{UE} can be determined by finding the non-perturbative corrections resulting in the best agreement with data.

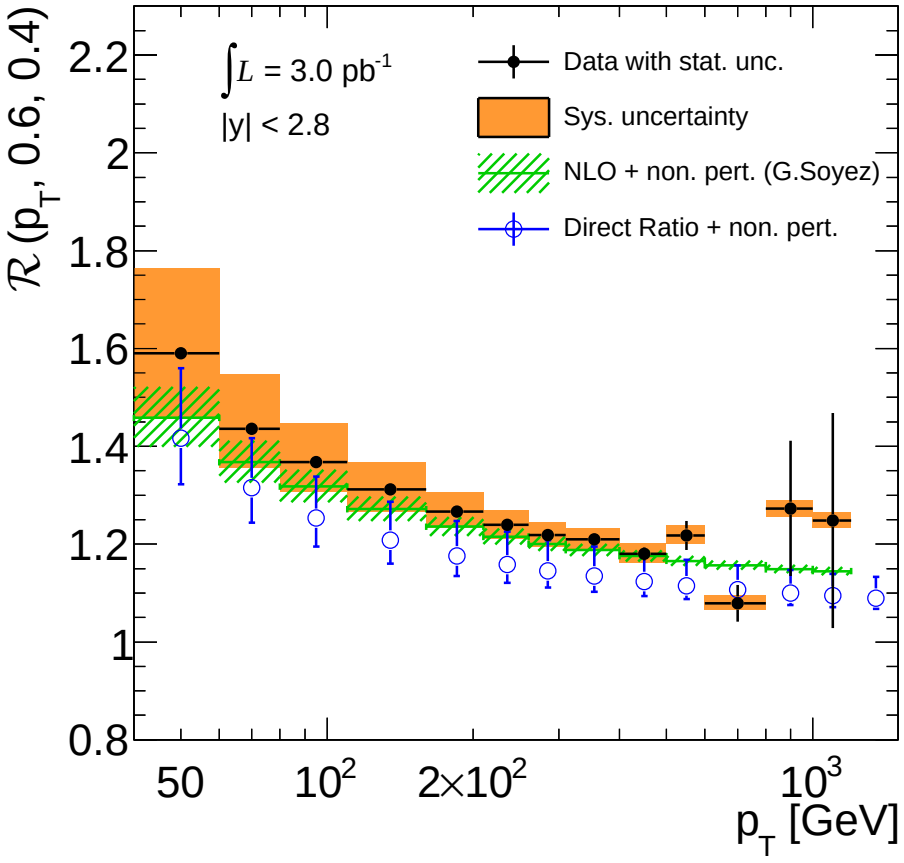

Figure 10.13: *The measured cross section ratio compared to theoretical predictions including non-perturbative corrections. Statistical uncertainties on data are indicated by error bars, the orange filled area represents the systematic, experimental uncertainties. NLO prediction taken from Ref. [89].*

Before exercising such a fit using the available data, the dependencies and correlations of both parameters are outlined in the following. A good starting point is Eq. 10.7. As a first step, the cross section at a shifted p_T value is approximated as:

$$\sigma_{LO}(p_T+\langle\delta p_T(R_1)\rangle) = \sigma_{LO}(p_T)+\sigma'_{LO}(p_T)\times\langle\delta p_T(R_1)\rangle \quad \text{with:} \quad \sigma'_{LO} = \frac{d\sigma_{LO}(p_T)}{dp_T}. \tag{10.22}$$

Hence, Eq. 10.7 becomes:

$$K^{(R_1/R_2)}(p_T) = \frac{\sigma_{LO}(p_T) + \sigma'_{LO}(p_T) \times \langle\delta p_T(R_1)\rangle}{\sigma_{LO}(p_T) + \sigma'_{LO}(p_T) \times \langle\delta p_T(R_2)\rangle}, \tag{10.23}$$

where only $\langle\delta p_T(R)\rangle$ is a function of the jet radius. To first order this can be expanded to:

$$\begin{aligned} K^{(R_1/R_2)}(p_T) &\approx 1 + \frac{\sigma'_{LO}(p_T) \times \langle\delta p_T(R_1)\rangle}{\sigma_{LO}(p_T)} - \frac{\sigma'_{LO}(p_T) \times \langle\delta p_T(R_2)\rangle}{\sigma_{LO}(p_T)} \\ &= 1 + \frac{\sigma'_{LO}(p_T)}{\sigma_{LO}(p_T)} \left[\langle\delta p_T(R_1)\rangle - \langle\delta p_T(R_2)\rangle\right], \end{aligned} \tag{10.24}$$

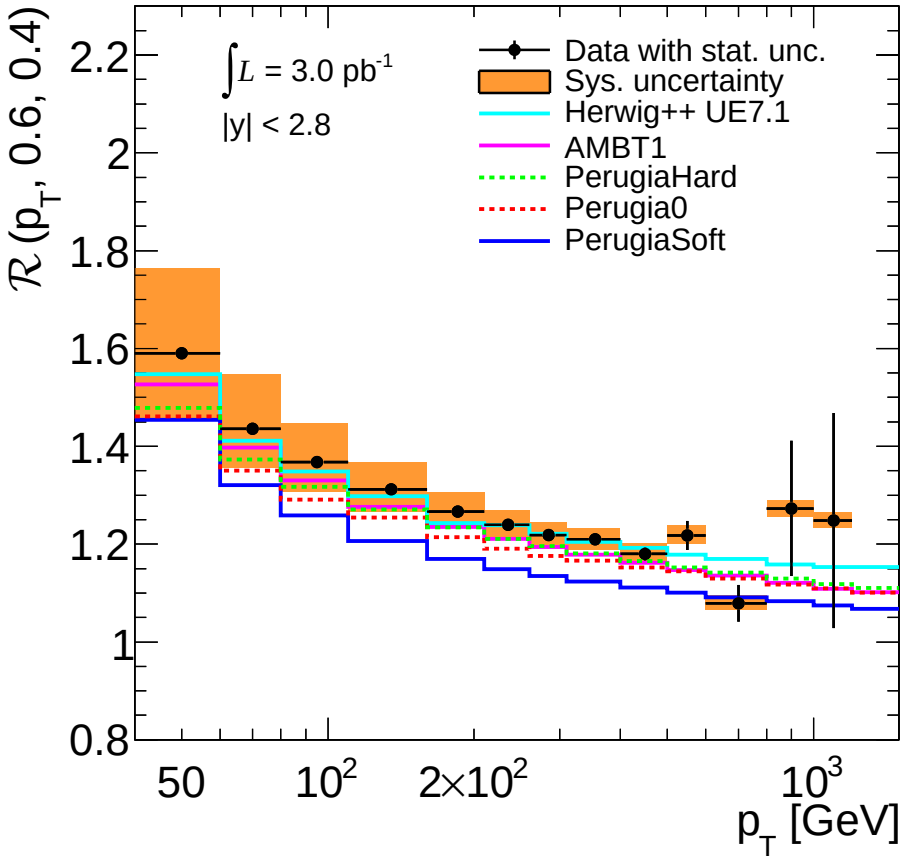

Figure 10.14: *The measured cross section ratio compared to various MC generators. Statistical uncertainties on data are indicated by error bars, the orange filled area represents the systematic, experimental uncertainties.*

In order to stress the interesting dependencies on α_0 and ρ_{UE}, $\langle \delta p_T(R) \rangle$ can be written as:

$$\begin{aligned} \langle \delta p_T(R) \rangle &= \langle \delta p_T(R) \rangle_{HAD} + \langle \delta p_T(R) \rangle_{UE} \\ &= \frac{F_{HAD}(\alpha_0)}{R} + F_{UE}(\rho_{\mathrm{UE}}) R^2, \end{aligned} \tag{10.25}$$

where F_{HAD} and F_{UE} include all constant factors from Eq. 10.8 and Eq. 10.10 respectively. With these notations Eq. 10.23 becomes:

$$K^{(R_1/R_2)}(p_T, \alpha_0, \rho_{\mathrm{UE}}) = 1 + \frac{\sigma'_{LO}(p_T)}{\sigma_{LO}(p_T)} \left[\frac{F_{HAD}(\alpha_0)(R_2 - R_1)}{R_2 R_1} + F_{UE}(\rho_{\mathrm{UE}})(R_1^2 - R_2^2) \right]. \tag{10.26}$$

For a first study using the measured cross section ratio, this equation holds valuable information. Namely, that a simultaneous measurement of ρ_{UE} and α_0 is not possible. This is due to fact that for a fixed combination of R_1 and R_2, data can only constrain the linear combination of $F_{HAD}(\alpha_0)$ and $F_{UE}(\rho_{\mathrm{UE}})$. For an actual measurement of both parameters it will thus be necessary to include at least one additional cross section measurement with a different jet radius. In this case the different R-dependence of $\langle \delta p_T \rangle_{HAD}$ and $\langle \delta p_T \rangle_{UE}$ will allow to constrain ρ_{UE} and α_0 separately.

To see the radius dependencies more clearly, for a fixed non-perturbative correction $K^{(R_1/R_2)}(p_T, \alpha_0, \rho_{\mathrm{UE}}) = K^{(R_1/R_2)}(p_T)$, Eq. 10.26 can be reordered. It then becomes obvious that the dependency between $F_{HAD}(\alpha_0)$ and $F_{UE}(\rho_{\mathrm{UE}})$ in this case follows a straight line:

$$F_{UE}(\rho_{\mathrm{UE}}) = \frac{\sigma'_{LO}(p_T)}{\sigma_{LO}(p_T)} \frac{K^{(R_1/R_2)}(p_T) - 1}{(R_1^2 - R_2^2)} + \frac{F_{HAD}(\alpha_0)}{R_1 R_2 (R_1 + R_2)}. \tag{10.27}$$

Inserting the constants included in $F_{HAD}(\alpha_0)$ and $F_{UE}(\rho_{\mathrm{UE}})$, ρ_{UE} can be written as a function of α_0:

$$\begin{aligned} \rho_{\mathrm{UE}}(\alpha_0) &= \rho^0_{UE} - m(R_1, R_2)\, \alpha_0 \\ \text{with :} \quad m(R_1, R_2) &= -\frac{4 C_R M \mu_I}{\pi^3\, R_1 R_2 (R_1 + R_2)}. \end{aligned} \tag{10.28}$$

where all contributions independent of α_0, R_1 and R_2 are absorbed in ρ^0_{UE}. For an arbitrary choice of α_0, Eq. 10.27 determines the ρ_{UE} resulting in any chosen value for the total, non-perturbative correction. This result states that, for a scenario of fixed R_1 the slope $m(R_1, R_2)$ decreases approximately quadratically for increasing R_2.

This is cross-checked by calculating $K^{(0.6/0.4)}(p_T, \alpha_0, \rho_{\mathrm{UE}})$ on a grid of ρ_{UE} and α_0 using `NLOJet++`, following Eq. 10.7. For this purpose, the non-perturbative corrections are factorized into the UE and the hadronization correction:

$$\begin{aligned} K^{(0.6/0.4)}(p_T, \alpha_0, \rho_{\mathrm{UE}}) &= K^{(0.6/0.4)}_{HAD}(p_T, \alpha_0) \times K^{(0.6/0.4)}_{UE}(p_T, \rho_{\mathrm{UE}}) \\ &= \frac{\sigma_{LO}(p_T + \langle \delta p_T(0.6) \rangle_{HAD})}{\sigma_{LO}(p_T + \langle \delta p_T(0.4) \rangle_{HAD})} \times \frac{\sigma_{LO}(p_T + \langle \delta p_T(0.6) \rangle_{UE})}{\sigma_{LO}(p_T + \langle \delta p_T(0.4) \rangle_{UE})}. \end{aligned} \tag{10.29}$$

Subsequently, for a fixed p_T, the lines of constant non-perturbative corrections,

$$K^{(0.6/0.4)}(p_T, \alpha_0, \rho_{\mathrm{UE}}) = K^{(0.6/0.4)}(p_T, 0.5, 2.3\ \mathrm{GeV}), \tag{10.30}$$

are found, where $\rho_{\mathrm{UE}} = 2.3$ GeV and $\alpha_0 = 0.5$ were chosen arbitrarily. For this purpose all jets were assumed to be gluon initiated jets. This substantially simplifies the analytical calculations. These contours are displayed in Fig. 10.15 together with the expectation from Eq. 10.27. The agreement found is good, except for small deviations in the absolute slope, which can be expected to arise from the various approximations made in the derivation of Eq. 10.27. The slope's evolution as a function of R_2 is almost perfectly described. Highest sensitivities to ρ_{UE} and α_0 are to be expected in configurations where the crossing angle between the straight lines in Fig. 10.15 is maximized. This encourages the usage of a wide variety of jet radii.

10.4.1 Determination of ρ_{UE} with fixed α_0

The data analyzed does not allow to measure α_0 and ρ_{UE}, due their linear connection. However, it allows to measure for instance ρ_{UE}, using α_0 as external input. For the following studies $\alpha_0(\mu_I = 2\ \mathrm{GeV}) = 0.50^{+0.07}_{-0.04}$ as measured at JADE [91] is used. To quantify the agreement between data and the theoretical prediction, for every $K^{(R_1/R_2)}(p_T, \alpha_0, \rho_{\mathrm{UE}})$ the χ^2 quantity with respect to data is calculated as follows:

$$\begin{aligned} \chi^2(\rho_{\mathrm{UE}}, \alpha_0) &= \sum_i \left(\frac{y_i - f_i(\rho_{\mathrm{UE}}, \alpha_0)}{\sigma_i} \right)^2 \\ \text{with :} \quad f_i(\rho_{\mathrm{UE}}, \alpha_0) &= \mathcal{R}_{pQCD}(p_T^{(i)}, 0.6, 0.4)\ K^{(0.6/0.4)}(p_T^{(i)}, \alpha_0, \rho_{\mathrm{UE}}), \end{aligned} \tag{10.31}$$

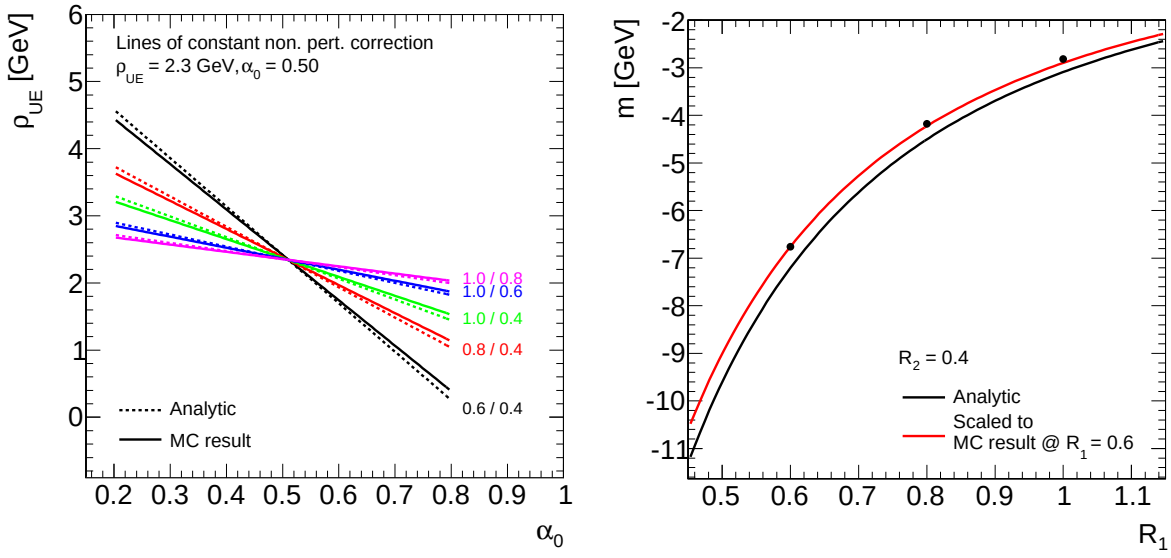

Figure 10.15: *Left: Lines of constant non-perturbative corrections to the cross section ratio, for various combinations of jet radii. Dashed lines represent the results from Eq.* 10.27*, which was used to calculate the slope. The intercept on the ordinate was set such that the line passes through* $(\alpha_0 = 0.5, \rho_{\mathrm{UE}} = 2.3\ \mathrm{GeV})$*. Right: The slopes for three choices of* R_1*, with fixed* $R_2 = 0.4$ *(dots) and the expectation according to Eq.* 10.27 *(black line). The red line was fixed to reproduce the MC result for* $m(0.6, 0.4)$*, keeping the* R_2 *dependence from Eq.* 10.27*.*

where i enumerates the bins of transverse momentum. This χ^2 definition does not take into account experimental systematic uncertainties nor uncertainties on the theoretical predictions. These will be treated in subsequent steps. The statistical uncertainties σ_i on the measurement are considered uncorrelated. For the hadronization correction, a physical mixture of quark and gluon jets is achieved by extracting the fraction of gluon jets from MC simulations using `Pythia6`, without the simulation of ISR, FSR, UE and hadronization. For the calculation of the correction using `NLOJet++`, the decision whether a jet is assumed to be gluon or quark-initiated, is then randomized according to this fraction.[†]

The result of this procedure is the two-dimensional $\chi^2(\alpha_0, \rho_{\mathrm{UE}})$ distribution shown in Fig. 10.16. As expected from the previous considerations, this distribution does not exhibit a real minimum. Rather it shows a valley of minimum χ^2 following a straight line. In order to determine ρ_{UE} for fixed α_0, the value of ρ_{UE} is found that minimizes $\chi^2(\rho_{\mathrm{UE}}, 0.5)$. The value found amounts to $(2.97\ \mathrm{GeV} \pm 0.02)$ GeV, where the uncertainty quoted corresponds to the shift in ρ_{UE}, resulting in an increase of χ^2 by one. It thus reflects the purely statistical uncertainty on ρ_{UE}. The minimum value of χ^2 found is approximately 31, for 13 fitted data points, corresponding to a probability of 3 ‰. Reasons for the χ^2 value being rather large can be expected to arise from the neglected theoretical uncertainties, which are to a certain degree statistically limited. Furthermore, a perfect description of the data within the very small statistical uncertainties cannot necessarily be expected, given the various assumptions that go into the analytical, non-perturbative corrections. It is also not guaranteed that variations of the data within its systematic uncertainties still yield a result, that perfectly fits to a theoretical prediction.

[†]Due to the implementation of `NLOJet++`, this workaround is necessary, since the species of partons in a generated event are not defined.

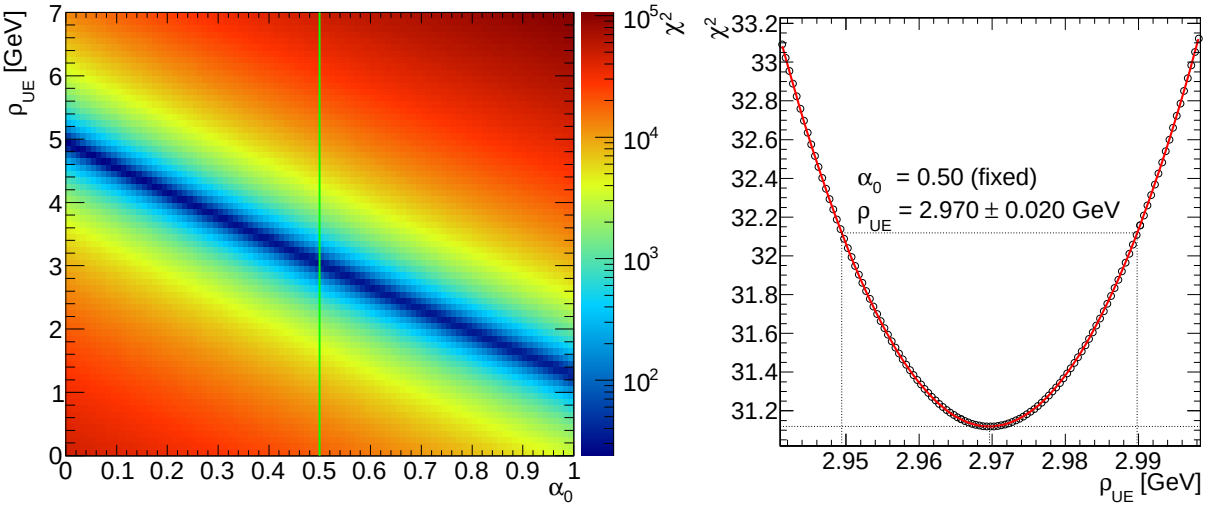

Figure 10.16: *Left: Distribution of χ^2 as a function of ρ_{UE} and α_0. The green line indicates the fixed α_0 value used. Right: Projection of the two-dimensional distribution along the green line. The dashed lines indicate the locations of the best estimate for ρ_{UE}.*

The systematic uncertainties are taken into account using pseudo-experiments: In each pseudo-experiment, randomized shifts to data are applied according to the systematic uncertainties. For this purpose they are assumed to be Gaussian and are symmetrized, by averaging the upper and lower uncertainty. These randomized shifts also account for the correlation between the single p_T bins. In principle, these correlations can be extracted by constructing the full covariance matrix for the cross section ratio (see Appendix A.1 for details). Unfortunately not all the necessary information to do so is available. The missing information is the correlation of the JES uncertainty between large and small jets, across different p_T bins. The estimation of the JES uncertainty correlation in this chapter only allowed to estimate this correlation for identical p_T bins and cannot be generalized further. Instead, the JES uncertainty correlations between single p_T bins of the calorimeter JES uncertainty are used, which were derived in single particle studies when estimating the JES uncertainty (cf. Chapter 8 and [92]). These correlations are very similar for both jet sizes, differing within less than 1 % in the kinematic region accessible in this analysis. For this reason, they are considered as the currently best estimate of the JES uncertainty correlation between p_T bins in the ratio measurement. These correlations are almost 100 % for adjacent p_T bins, decreasing to approximately 30 % between the lowest p_T bin and 1 TeV.

Each pseudo-experiment yields a result for ρ_{UE} and the distribution of these results is displayed in Fig. 10.17. The results are approximately Gaussian distributed, which is a direct consequence of the assumption that the systematic uncertainties are Gaussian as well. The non-Gaussian behaviour for low and negative values of ρ_{UE} arises from pseudo-experiments in which the data fluctuates downward. Since the fixed value of α_0 already results in a certain non-perturbative correction larger than one, the fluctuations of data often lie beneath the theoretical prediction, even without the underlying event correction. This also explains the large amount of unphysical results of $\rho_{\mathrm{UE}} < 0$. The core of the distribution is used to extract the systematic uncertainty for ρ_{UE}, as the width of a Gaussian fit. This yields an uncertainty of 1.3 GeV, which amounts to a relative uncertainty of approximately 43 %.

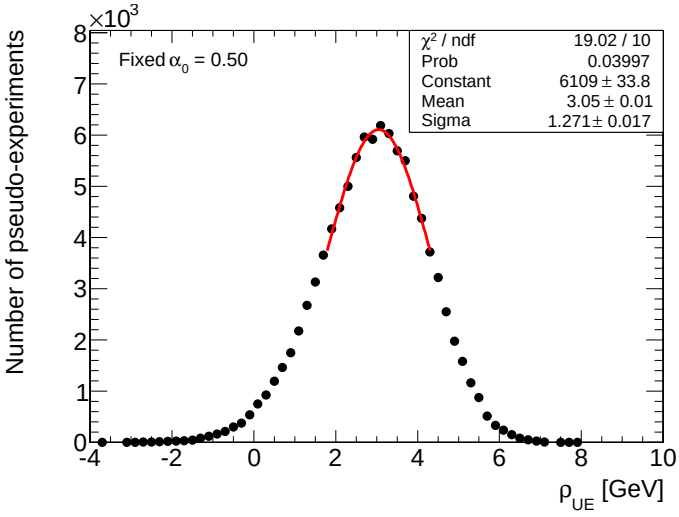

Figure 10.17: *Distribution of the ρ_{UE} values in each pseudo-experiment. The width of the distribution is taken as the systematic uncertainty.*

In order to assess the uncertainty induced by the uncertainty on the external value of α_0, it is varied by the uncertainties quoted in [91]. The resulting variation in ρ_{UE} of $^{+0.2}_{-0.3}$ GeV is taken as the corresponding systematic uncertainty. The uncertainty in the parton level prediction of $\mathcal{R}_{pQCD}(p_T; R_1, R_2)$ is propagated to ρ_{UE} by varying it within the scale uncertainties and repeating the fit. The variation induced in the result for ρ_{UE} amounts to ± 0.4 GeV.

One simplification that was made in the calculation of the analytic non-perturbative corrections is, that there is no dispersion in $\langle \delta p_T \rangle_{HAD}$ and $\langle \delta p_T \rangle_{UE}$. The impact of this assumption is estimated by randomizing the p_T shifts according to a Gaussian with a fractional width of 80 %. Hence, the average non-perturbative correction remains unchanged. The width of the Gaussian was chosen according to UE analyses of the standard deviation of the transverse momentum density in the transverse plane [93]. As there are no theoretical predictions for the dispersion of the hadronization correction, the same procedure is applied there as well. The variation found for ρ_{UE} amounts to 0.2 GeV and is taken as an additional uncertainty.

The systematic uncertainties on the ρ_{UE} determination are summarized in Table 10.2. The dominant uncertainty is the one induced by the experimental uncertainty.

Source	Uncertainty [GeV]
Experimental	1.3
$\alpha_0 = 0.50^{+0.07}_{-0.04}$	$^{+0.2}_{-0.3}$
μ_r, μ_f scales	0.4
Dispersion	0.2

Table 10.2: *Systematic uncertainties in the determination of ρ_{UE}.*

For $\alpha_0(\mu_I = 2\ \mathrm{GeV}) = 0.50^{+0.07}_{-0.04}$, the underlying event density measured this way amounts to:

$$\rho_{\mathrm{UE}} = \left[3.05 \pm 0.02\,(\mathrm{stat.}) \pm 1.27\,(\mathrm{exp.})\ ^{+0.49}_{-0.54}\,(\mathrm{model})\right] \mathrm{GeV}, \tag{10.32}$$

where all but the experimental, systematic uncertainties are grouped into the *model* uncertainty. This result hence prefers a larger underlying event density, than one can conclude

from studies done at ATLAS so far (see Fig. 3.5), from which $\rho_{\mathrm{UE}} = (1.8 \pm 0.5)$ GeV was extracted for the analytic underlying event corrections. Still, this value is compatible within the significant systematic uncertainties with the presented measurement. Additionally, it should be noted that UE studies so far have only been carried out at lower transverse momentum scales, which makes extrapolations necessary to estimate the UE at larger transverse momenta. Furthermore, analyses of the underlying event, in practice, have so far only been carried out in the region transverse to the hard two-jet system in an event. The completely different Ansatz followed here instead attempts a model-dependent measurement of the UE right within the products of the hard scattering. Taking into account, that the UE is in fact not completely independent of the hard scattering, it can be concluded that previous measurements and the presented one are not measuring the very same quantity. Hence, a perfect agreement between the two results should not be expected. Rather, this measurement should be taken as a complementary approach, probing slightly different features of the underlying event.

For future measurements of ρ_{UE} with fixed α_0 there is still room for improvement. The presented methods in this chapter are considered to provide a reasonable and still conservative estimate of the JES uncertainty correlation. Nevertheless, it seems desirable to establish a JES uncertainty, whose ingredients more directly aim at physical properties of jets, thus making a more precise estimate of the correlations possible.

Improved results may also be gained by not treating the systematic uncertainties as Gaussian distributed. The Gaussian assumption for instance results in pseudo-experiments exhibiting a increasing, rather than decreasing, cross section ratio at low transverse momenta. An improved fitting procedure should aim to restrict the measured values of ρ_{UE} and α_0 to physically sensible values above zero.

For the dispersion of the underlying event correction the Gaussian assumption made to assess the systematic uncertainty is a very idealized one. Measurements of e.g. the differential distributions of the transverse momentum density in the transverse plane, might provide a more realistic Ansatz.

Since there are no theoretical predictions for the size of the dispersion of the hadronization correction, it was treated using an dispersion identical to that used for the underlying event correction. Tough the impact of the assumed dispersion is small compared to the experimental uncertainties, future measurements would benefit from theoretical predictions for this issue.

10.4.2 Feasibility study for a combined measurement of ρ_{UE} and α_0

As outlined in the beginning of this chapter, a measurement of ρ_{UE} and α_0 needs at least one more cross section measurement with a different jet radius as input. In this section the feasibility of a such a combined fit is examined using the results above, together with theoretical considerations. This is done by simulating additional ratio measurements as follows: It is assumed that Eq. 10.27 is able to properly describe the slopes of the minimum in the χ^2 planes. This was demonstrated in Fig. 10.15 to a satisfactory degree. Hence, χ^2 planes for different combinations of R_1 and R_2 than the ones used, could be simulated by rotating the *measured* χ^2 plane in Fig. 10.16 to another slope, according to R_1 and R_2. Denoting the slope of the χ^2 valley in Fig. 10.16 as m_{data}, the slope $m(R_1, R_2)$ of an alternative ratio can generally be written as:

$$m(R_1, R_2) = m_{data} \frac{0.6 \times 0.4 \times (0.6 + 0.4)}{R_1 R_2 (R_1 + R_2)}. \tag{10.33}$$

Accordingly, the modified intercept on the ordinate is constructed as

$$b(R_1, R_2) = \rho_{\mathrm{UE}}^{true} - m(R_1, R_2)\alpha_0^{true}. \tag{10.34}$$

For this purpose $\rho_{\mathrm{UE}}^{true} = 2.97$ GeV (according to Fig. 10.16) and $\alpha_0^{true} = 0.5$ are taken as the *true* values, which are used as input to the following pseudo-experiment. They define the center of the rotation applied to the χ^2 planes. These values hence should also be the result of the pseudo-experiment, in order to validate the assumption, that a combined measurement of both parameters is indeed feasible. In the following example $R_2 = 0.4$ was kept as the fixed, smaller radius, while $R_1 = 0.9$ and $R_1 = 1.2$ were chosen as additional radii. Instead of a rotation of the χ^2 plane, a slightly simpler technique is employed: The χ^2 distribution in each α_0 bin is shifted such that the resulting slope and intercept on the ordinate obey Eq. 10.33 and Eq. 10.34 respectively. Taking into account the very large number of bins necessary for this study, this was found to be a computationally easier, yet satisfactory procedure.

The two distributions accordingly transformed are shown in Fig. 10.18 along with the original distribution. Making the further assumption, that all three fits are uncorrelated,

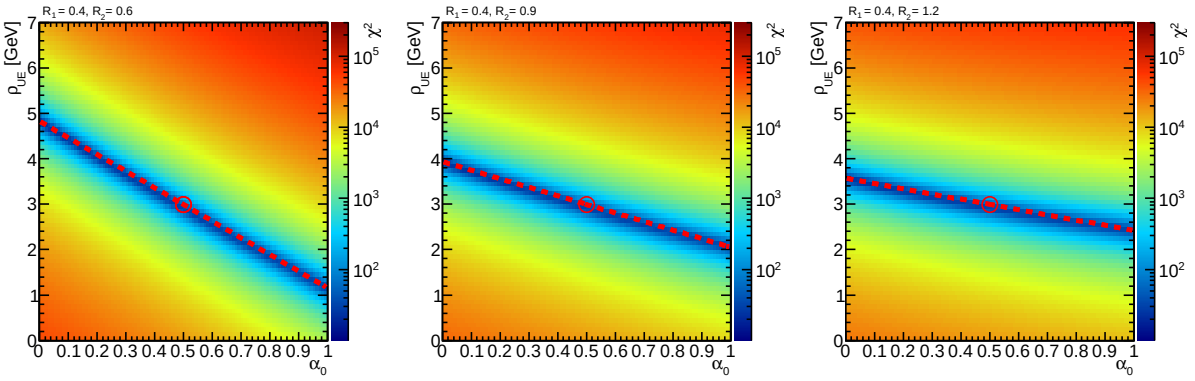

Figure 10.18: *From left to right: The observed χ^2 distribution for $(R_1, R_2) = (0.4, 0.6)$ and simulated distributions for $(R_1, R_2) = (0.4, 0.9)$ and $(R_1, R_2) = (0.4, 1.2)$. The open circle marks the input values used for the pseudo-experiments of $\rho_{\mathrm{UE}}^{true} = 2.97$ GeV and $\alpha_0^{true} = 0.5$.*

the χ^2 values calculated for each ratio can simply be added:

$$\chi^2_{combined}(\rho_{\mathrm{UE}}, \alpha_0) = \sum_{R_1 \in \{1.2, 0.9, 0.6\}} \chi^2(\rho_{\mathrm{UE}}, \alpha_0; R_1, 0.4). \tag{10.35}$$

Though this is an unrealistic assumption, it is legitimate for the purpose of this pseudo-experiment. While correlations can be expected to change the actual minimum χ^2 values, they cannot be expected to have an impact on the bare existence of such a minimum. The distribution of the χ^2 sum is shown in Fig. 10.19. As expected, in this combination the different slopes add up to a well localized χ^2 minimum. The fit results for both parameters are extracted as the values where $\chi^2_{combined}$ has its global minimum. The uncertainties are obtained from the one-dimensional $\chi^2_{combined}$ distributions, which scan $\chi^2_{combined}$ as function of one parameter, keeping the other one fixed, as illustrated in Fig. 10.20. This yields the result of this pseudo-experiment:

$$\rho_{\mathrm{UE}} = 2.975 \pm 0.012 \quad \text{and} \quad \alpha_0 = 0.498 \pm 0.005.$$

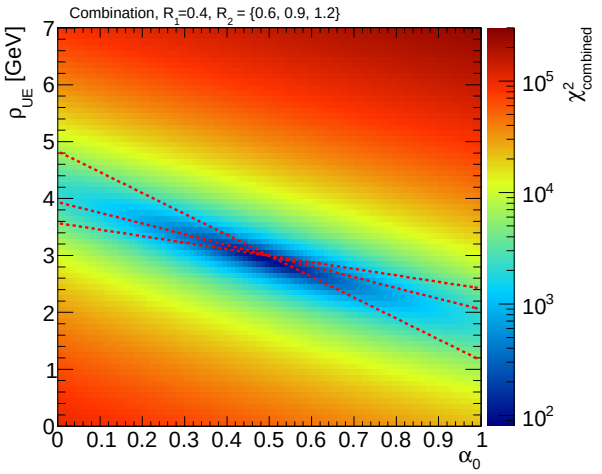

Figure 10.19: *Distribution of the combined χ^2, using the single distributions from Fig. 10.18. Dashed lines correspond to the slopes found for the single distributions.*

These results are in very good agreement with the input parameters used. Due to the neglected correlations and the simplicity of the toy study, the extracted uncertainties cannot be taken as a serious estimate of the statistical uncertainties in an actual measurement. For a future combined measurement, additionally to the possible improvements mentioned at the end of the previous subsection, two important issues have to be addressed:

- **Fits to alternative ratios:** Before exercising the fit to various ratios at the same time, the theoretical predictions firstly will have to be able to describe each single ratio properly. This is a major assumption for the combination of the χ^2-planes in Fig. 10.18, but cannot be taken for granted.

- **Correlations of systematic uncertainties:** As the results in this subsection were meant as a first feasibility study, the systematic uncertainties were ignored.

  To incorporate the systematic uncertainties properly, more detailed analyses of the correlations between the JES uncertainty of various jet radii will be needed. The methods established in this chapter for instance cannot be generalized to assess the correlations between different jet radii and different bins of transverse momentum. For the JES uncertainty correlation determination using pseudo-experiments based on single-particle studies, the extension to a correlation matrix between various jet sizes and across different p_T bins can be considered to be straight forward. For the correlations of other effects, such as the fragmentation and the underlying event new methods will have to be developed.

  Finally, the introduction of additional cross sections will add yet another dimension of correlation, which will have to be taken care of.

With a proper treatment of these items, it seems reasonable to consider a measurement of α_0 possible. The measurement of ρ_{UE} in this new fashion may provide a new access to the UE, complementary to common measurements of the UE at low transverse momenta.

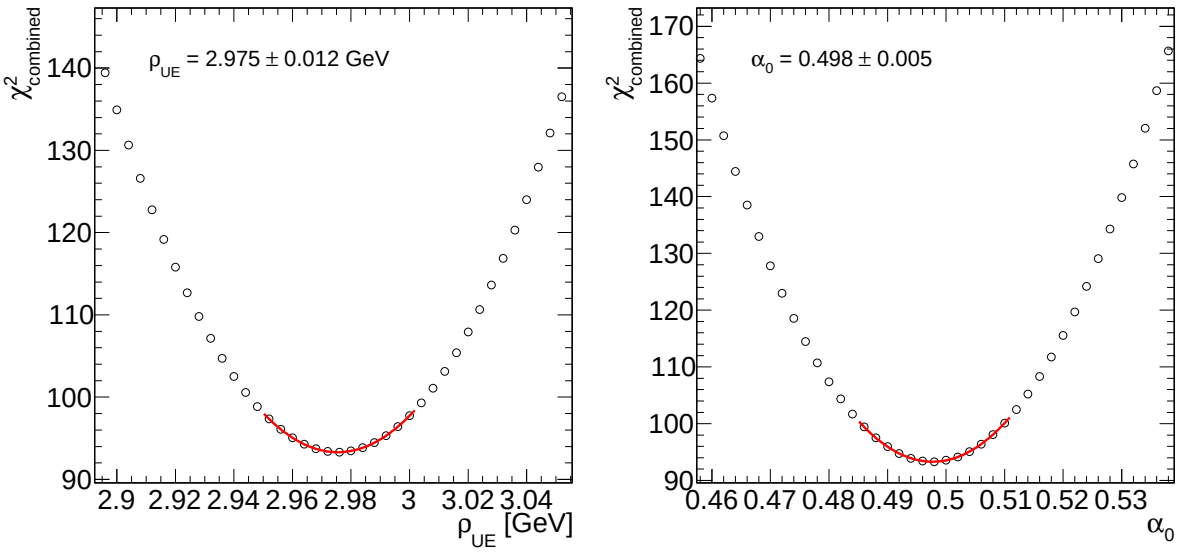

Figure 10.20: $\chi^2_{combined}$ *distributions around the minimum, as a function of* ρ_{UE} *(left) and* α_0 *(right).*

11 Conclusions and outlook

Inclusive jet cross sections are one of the most direct probes of pQCD at the highest accessible momentum scales. Due to the large production cross section, they are a natural observable at new experiments which extend the kinematic reach with respect to previous ones. They additionally allow to probe the proton's structure at highest Q^2 and hence at highest resolution.

This thesis presented an inclusive jet cross section measurement in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. Jets were identified using the anti-k_t jet algorithm with two different jet sizes. The measurement was performed double differentially in bins of the jet transverse momentum and bins of absolute rapidity. Kinematically, the measurement was constrained to jets with $p_T > 40$ GeV and $|y| < 2.8$. The amount of data corresponds to $3\,\mathrm{pb}^{-1}$, recorded in Summer 2010.

Jets with transverse momenta of up to 1 TeV were observed, going beyond the kinematic reach of earlier analyses at the Tevatron accelerator. The measured data has been fully corrected for detector effects, allowing for meaningful comparisons to theoretical predictions. The experimental systematic uncertainty ranges between 20 % and 30 % and is the dominant uncertainty up to transverse momenta of approximately 700 GeV. The systematic uncertainty itself is dominated by the limited knowledge of the absolute jet energy scale calibration. Data was compared to theoretical predictions, calculated at next-to-leading order using the CTEQ6.6 PDF set. The predictions were found to agree within the experimental and theoretic uncertainties. Comparisons to different PDF sets showed that due to the experimental precision, data cannot favour one PDF set over another yet.

Significant differences in the predictions by various PDF sets start to appear beyond transverse momenta of 1 TeV. Continuous running of the LHC at increasingly higher instantaneous luminosities will quickly increase the kinematic reach and improve the statistical precision in this region. By the beginning of June 2011, ATLAS already recorded data corresponding to more than $1\,\mathrm{fb}^{-1}$. Decreasing the systematic uncertainties will be the most important task for future jet cross section measurements. The estimate of the absolute jet energy scale calibration in the full 2010 dataset was already improved with respect to the one used in the analysis presented. To increase the value of the results for the purpose of global PDF fits, the evaluation and inclusion of the correlations of systematic uncertainties is a further aim. Several steps into these directions are already undertaken by the ATLAS collaboration. This already lead to updated preliminary results of the jet cross section measurements using the full 2010 dataset, which uses a factor of ten more data than the analysis presented. Complementary to the accumulation of more data, recent studies also significantly extend the coverage in pseudorapidity up to $|y| < 4.4$. This significantly enhances the sensitivity in probing the proton's structure at low values of the Bjorken-x.

Current LHC plans foresee a further increase of the center-of-mass energy up to the originally designed value of 14 TeV by 2014. This drastic increase will come along with a large increase of the kinematic range for jet cross section measurements. Cross sections

in the currently accessible region at the same time will be increased by up to two orders of magnitude, as is illustrated in Fig. 11.1.

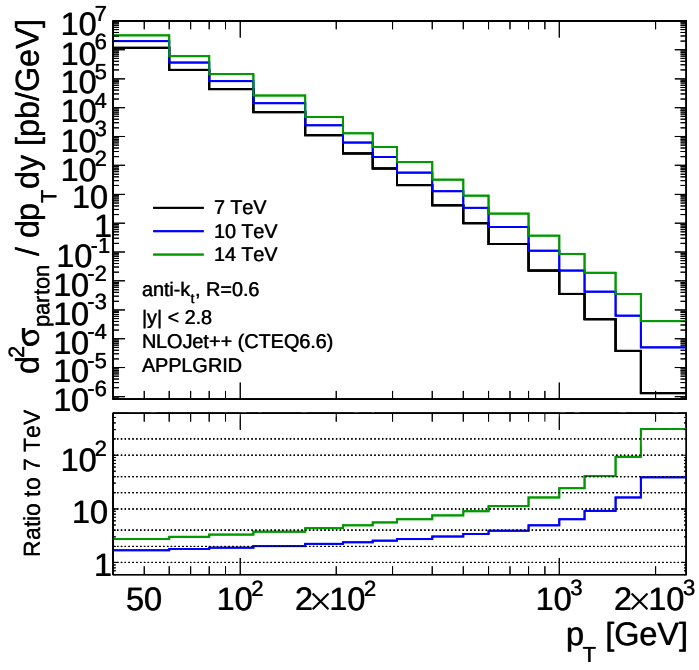

Figure 11.1: *Expected inclusive jet cross sections at parton level as in this analysis and for center-of-mass energies of 10* TeV *and 14* TeV*. The extrapolation to higher center-of-mass energies was performed using* `APPLGRID`*.*

The measurement of inclusive jet cross sections has been carried out using two different jet radii, in an otherwise identical analysis. For the first time, a measurement of the ratio of two cross sections has also been performed, properly taking into account the correlations of the experimental uncertainties. The correlations found are above 85 % and lead to a significant decrease in the relative, systematic uncertainty for the ratio measurement, down to a maximum of 10 %. It was observed, that inclusive cross section predictions at next-to-leading order do not describe this ratio well and suffer from significant scale uncertainties. Newer approaches were found to provide a more realistic description of the data including significantly smaller uncertainties. The cross section ratio also confirmed recent theoretical work on the analytical treatment of non-perturbative corrections to be working well.

Inspired by theoretical work, analytical non-perturbative corrections have been used to measure ρ_{UE}. The result is higher than current estimates from ATLAS underlying event studies, but is compatible within the significant systematic uncertainties. A feasibility study has been performed to assess the possibility of a concurrent measurement of ρ_{UE} and the non-perturbative strong coupling α_0. Such a measurement has been found to be possible, but it requires the inclusion of a least third cross section measurement using a different jet radius.

An extended analysis using multiple jet cross section ratios will require further work. Calibrations of jets with different radii have to be derived. A framework for the calibration of jets of various radii was developed within the ATLAS collaboration, as a direct

extension of the calibration method presented in this analysis. Most importantly however, the correlations between jet energy scale uncertainties for various jet sizes have to be understood. Methods to assess these correlations have been developed in this thesis and are able to be significantly extended to the usage with updated jet energy scale uncertainties. The increasingly better performance of data driven methods for JES uncertainty estimates will certainly allow a new, more precise assessment of the correlations. The increased amount of data available for future analyses will result in improved statistical uncertainties and will allow for a precise measurement of the ratio up to 1 TeV. In view of the small systematic uncertainties, it may hence provide an interesting access to the underlying event and non-perturbative effects at new energy scales.

A possible extension of the analysis presented might be to measure cross section ratios, in certain bins of transverse momentum, versus the jet's radius, using an arbitrary jet radius as reference. Contrary to the thesis presented, this analysis does not require a complete study of the inclusive cross sections. As observed in Chapter 10, particularly the unfolding procedure, an essential ingredient to the single cross section measurement, could be omitted, allowing for an easier construction of the cross section ratio at particle level. The measurement of the ratio versus p_T would allow to directly probe the theoretical predictions for the size of non-perturbative corrections as a function of the jet radius. The already significant spread observed among the various predictions from MC generators can be expected to grow with the spread of the jet radii used. Future ratio measurements may hence also be provided as input for generator tuning efforts. In this context, ratios of inclusive jet cross sections can be considered a new, complementary approach to features of the underlying event and a direct measurement of non-perturbative effects.

List of acronyms

ADC	Analogue to Digital Converter
ALICE	A Large Ion Collider Experiment
ATLAS	A Toroidal LHC Apparatus
CDF	Collider Detector at Fermilab
CMS	Compact Muon Spectrometer
CTP	Central Trigger Processor
EF	Event Filter
FCAL	Forward CALorimeter
FSR	Final State Radiation
HLT	High-Level Trigger
ISR	Initial State Radiation
JES	Jet Energy Scale
L1	Level-1 Trigger
LHC	Large Hadron Collider
LO	Leading Order
MBTS	Minimum Bias Trigger Scintillators
MC	Monte Carlo
NDF	Number of Degrees of Freedom
NLO	Next-to-Leading Order
NNLO	Next-to-Next-to-Leading Order
NNPDF	Neural Net Parton Density Function
PDF	Parton Density Function
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
ROI	Region of Interest

SCT	Semiconductor Tracker
TRT	Transition Radiation Tracker
UE	Underlying Event
ZDC	Zero-Degree Calorimeter

Bibliography

[1] Particle Data Group Collaboration, K. Nakamura et al., *Review of particle physics*, J. Phys. **G37** (2010) 075021.

[2] CDF and DØ Collaboration, *Combination of CDF and* DØ *Results on the Mass of the Top Quark using up to 5.6 fb^{-1} of data*, arXiv:1007.3178 [hep-ex].

[3] D. J. Gross and F. Wilczek, *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (1973) no. 26, 1343–1346.

[4] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) no. 26, 1346–1349.

[5] S. Bethke, *The 2009 Wolrd Average of $\alpha_s(M_Z)$*, Eur. Phys. J. **C64** (2009) 689–703, arXiv:0908.1135 [hep-ph].

[6] T. Sjöstrand, S. Mrenna, and P. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, 0603175 [hep-ph].

[7] M. Bahr, S. Gieseke, M. Gigg, D. Grellscheid, K. Hamilton, et al., *Herwig++ Physics and Manual*, Eur.Phys.J. **C58** (2008) 639–707, arXiv:0803.0883 [hep-ph].

[8] A. Buckley et al., *General-purpose event generators for LHC physics*, arXiv:1101.2599 [hep-ph].

[9] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization for short distance hadron-hadron scattering*, Nuclear Physics B **261** (1985) 104 – 142.

[10] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl.Phys. **B126** (1977) 298.

[11] *HERA Combined Results Website*, https://www.desy.de/h1zeus/combined_results.

[12] V. Radescu, *Combination and QCD Analysis of the HERA Inclusive Cross Sections*, in *Proceedings of the 35th International Conference of High Energy Physics*. PoS(ICHEP 2010)168, 2010.

[13] G. C. Blazey et al., *Run II jet physics*, arXiv:0005012 [hep-ex].

[14] G. P. Salam and G. Soyez, *A practical Seedless Infrared-Safe Cone jet algorithm*, JHEP **05** (2007) 086, arXiv:0704.0292 [hep-ph].

[15] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. D **48** (1993) no. 7, 3160–3166.

[16] G. P. Salam, M. Cacciari, and G. Soyez.
http://www.lpthe.jussieu.fr/~salam/fastjet/.

[17] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B641** (2006) 57–61, arXiv:0512210 [hep-ph].

[18] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *Better Jet Clustering Algorithms*, JHEP **08** (1997) 001, arXiv:9707323 [hep-ph].

[19] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross sections in deep-inelastic scattering*, arXiv:9907280 [hep-ph].

[20] M. Cacciari, G. Salam, and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **0804** (2008) 063, arXiv:0802.1189.

[21] G. Hanson et al., *Evidence for Jet Structure in Hadron Production by e^+e^- Annihilation*, Phys. Rev. Lett. **35** (1975) no. 24, 1609–1612.

[22] P. Soding, B. Wiik, G. Wolf, and S. Wu, *The First evidence for three jet events in e+ e- collisions at PETRA: First direct observation of the gluon*, . Talk given at Award Ceremony to the 1995 EPS High Energy and Particle Physics Prize.

[23] H1 and ZEUS Collaboration, F. D. Aaron et al., *Combined Measurement and QCD Analysis of the Inclusive ep Scattering Cross Sections at HERA*, JHEP **01** (2010) 109, arXiv:0911.0884 [hep-ex].

[24] CDF Collaboration, T. Aaltonen et al., *Measurement of the Inclusive Jet Cross Section at the Fermilab Tevatron p-pbar Collider Using a Cone-Based Jet Algorithm*, Phys. Rev. **D78** (2008) 052006, arXiv:0807.2204 [hep-ex].

[25] DØ Collaboration, V. M. Abazov et al., *Measurement of the inclusive jet cross-section in p anti-p collisions at $\sqrt{s}$ =1.96* TeV, Phys. Rev. Lett. **101** (2008) 062001, arXiv:0802.2400 [hep-ex].

[26] G. C. Blazey and B. L. Flaugher, *Inclusive jet and dijet production at the Tevatron*, Annual Review of Nuclear and Particle Science **49** (1999) no. 1, 633–685. http://www.annualreviews.org/doi/abs/10.1146/annurev.nucl.49.1.633.

[27] DØ Collaboration, V. M. Abazov et al., *Measurement of dijet azimuthal decorrelations at central rapidities in p anti-p collisions at $\sqrt{s}$ = 1.96* TeV, Phys. Rev. Lett. **94** (2005) 221801, arXiv:0409040 [hep-ex].

[28] A. A. Bhatti and D. Lincoln, *Jet Physics at the Tevatron*, Annual Review of Nuclear and Particle Science **60** (2010) no. 1, 267–297.

[29] CDF Collaboration, D. Acosta et al., *Study of jet shapes in inclusive jet production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D **71** (2005) no. 11, 112002.

[30] ATLAS Collaboration, *Measurements of underlying-event properties using neutral and charged particles in pp collisions at 900 GeV and 7 TeV with the ATLAS detector at the LHC*, arXiv:1103.1816 [hep-ex].

[31] ATLAS Collaboration, *Charged-particle multiplicities in pp interactions measured with the ATLAS detector at the LHC*, New J.Phys. (2010) , arXiv:1012.5104 [hep-ex].

[32] S. Catani and M. H. Seymour, *A general algorithm for calculating jet cross sections in NLO QCD*, Nucl. Phys. **B485** (1997) 291–419, arXiv:9605323 [hep-ph].

[33] Z. Nagy, *Next-to-leading order calculation of three jet observables in hadron hadron collision*, Phys. Rev. **D68** (2003) 094002, arXiv:0307268 [hep-ph].

[34] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, et al., *Implications of CTEQ global analysis for collider observables*, Phys.Rev. **D78** (2008) 013004, arXiv:0802.0007 [hep-ph].

[35] M. Rubin, G. P. Salam, and S. Sapeta, *Giant QCD K-factors beyond NLO*, JHEP **09** (2010) 084, arXiv:1006.2144 [hep-ph].

[36] T. Carli et al., *A posteriori inclusion of parton density functions in NLO QCD final-state calculations at hadron colliders: The APPLGRID Project*, Eur. Phys. J. **C66** (2010) 503–524, arXiv:0911.2985 [hep-ph].

[37] H.-L. Lai et al., *Uncertainty induced by QCD coupling in the CTEQ global analysis of parton distributions*, Phys. Rev. **D82** (2010) 054021, arXiv:1004.4624 [hep-ph].

[38] D. Stump, J. Huston, J. Pumplin, W.-K. Tung, H.-L. Lai, S. Kuhlmann, and J. F. Owens, *Inclusive jet production, parton distributions, and the search for new physics*, Journal of High Energy Physics **2003** (2003) no. 10, 046. http://stacks.iop.org/1126-6708/2003/i=10/a=046.

[39] M. Dasgupta, L. Magnea, and G. P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, JHEP **0802** (2008) 055, arXiv:0712.3014 [hep-ph].

[40] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP **06** (2002) 029, arXiv:0204244 [hep-ph].

[41] S. Alioli, K. Hamilton, P. Nason, C. Oleari, and E. Re, *Jet pair production in POWHEG*, arXiv:1012.3380 [hep-ph].

[42] *The Herwig++ LHC-UE7-1 Tune*, http://projects.hepforge.org/herwig/trac/wiki/MB_UE_tunes.

[43] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[44] ATLAS Collaboration, *ATLAS Monte Carlo tunes for MC09*, ATL-PHYS-PUB-2010-002, 2010.

[45] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001.

[46] ALICE Collaboration, K. Aamodt et al., *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[47] LHCb Collaboration, A. Alves et al., *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[48] LHCf Collaboration, O. Adriani et al., *The LHCf detector at the CERN LHC*, JINST **3** (2008) S08006.

[49] TOTEM Collaboration, G. Anelli et al., *The TOTEM experiment at the CERN LHC*, JINST **3** (2008) S08007.

[50] CMS Collaboration, R. Adolphi et al., *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.

[51] ATLAS Collaboration, *The ATLAS Experiment at the CERN LHC*, JINST **3** (2008) S08003.

[52] ATLAS Collaboration, T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007, CERN, Geneva, Mar, 2007.

[53] P. Jenni, M. Nessi, and M. Nordberg, *Zero Degree Calorimeters for ATLAS*, Tech. Rep. LHCC-I-016. CERN-LHCC-2007-001, CERN, Geneva, Jan, 2007.

[54] R. W. L. Jones and D. Barberis, *The evolution of the ATLAS computing model*, Journal of Physics: Conference Series **219** (2010) no. 7, 072037. `http://stacks.iop.org/1742-6596/219/i=7/a=072037`.

[55] A. Sherstnev and R. S. Thorne, *Parton Distributions for LO Generators*, Eur. Phys. J. **C55** (2008) 553–575, `arXiv:0711.2473 [hep-ph]`.

[56] S. Agostinelli et al., *G4–a simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) no. 3, 250 – 303.

[57] J. Allison et al., *Geant4 Developments and Applications*, IEEE Transactions on Nuclear Science **53** (2006) 270–278.

[58] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur.Phys.J. **C70** (2010) 823–874, `arXiv:1005.4568 [physics.ins-det]`.

[59] E. Abat et al., *Response and Shower Topology of 2 to 180 GeV Pions Measured with the ATLAS Barrel Calorimeter at the CERN Test-beam and Comparison to Monte Carlo Simulations*, Tech. Rep. ATL-CAL-PUB-2010-001, CERN, Geneva, May, 2010.

[60] E. Abat et al., *Study of energy response and resolution of the ATLAS barrel calorimeter to hadrons of energies from 20 to 350* GeV, Nucl. Instrum. Meth. **A621** (2010) no. 1-3, 134 – 150.

[61] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, Tech. Rep. ATL-LARG-PUB-2008-002, CERN, Geneva, Apr, 2008.

[62] ATLAS Collaboration, *Readiness of the ATLAS Liquid Argon Calorimeter for LHC Collisions*, Eur. Phys. J. **C70** (2010) 723–753, `arXiv:0912.2642`.

[63] T. Gabriel, D. Groom, P. Job, N. Mokhov, and G. Stevenson, *Energy dependence of hadronic activity*, Nucl. Instrum. Meth. **A338** (1994) no. 2-3, 336 – 347.

[64] D. E. Groom, *Energy flow in a hadronic cascade: Application to hadron calorimetry*, Nucl. Instrum. Meth. **A572** (2007) no. 2, 633 – 653.

[65] J. E. Brau, J. A. Jaros, and H. Ma, *Advances in Calorimetry*, Annual Review of Nuclear and Particle Science **60** (2010) no. 1, 615–644.

[66] D. Lopez Mateos, E. W. Hughes, and A. Schwartzman, *A Simple p_T- and η-Dependent Monte Carlo-Based Jet Calibration*, Tech. Rep. ATL-PHYS-INT-2009-077, CERN, Geneva, Aug, 2009.

[67] S. Eckweiler, *Performance of jet reconstruction and calibration in first ATLAS data at a centre-of-mass energy of 7 TeV*, in *Proceedings of the 35th International Conference of High Energy Physics*. PoS(ICHEP 2010)034, 2010.

[68] ATLAS Collaboration, *Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy sqrts = 7 TeV*, Tech. Rep. ATLAS-CONF-2010-054, CERN, Geneva, Jul, 2010.

[69] ATLAS Collaboration, *Charged-particle multiplicities in pp interactions at $\sqrt{s}$ = 900 GeV measured with the ATLAS detector at the LHC*, Phys. Lett. **B688** (2010) 21–42, `arXiv:1003.3124 [hep-ex]`.

[70] Atlas Collaboration, *Measurement of inclusive jet and dijet cross sections in proton-proton collisions at 7 TeV centre-of-mass energy with the ATLAS detector*, Eur.Phys.J. **C71** (2011) 1512, `arXiv:1009.5908 [hep-ex]`.

[71] ATLAS Collaboration, *In-situ pseudo-rapidity inter-calibration to evaluate jet energy scale uncertainty and calorimeter performance in the forward region*, Tech. Rep. ATLAS-CONF-2010-055, CERN, Geneva, Jul, 2010.

[72] ATLAS Collaboration, *Study of Jet Shapes in Inclusive Jet Production in pp Collisions at $\sqrt{s}$ = 7 TeV using the ATLAS Detector*, `arXiv:1101.0070 [hep-ex]`.

[73] ATLAS Collaboration, *In-situ jet energy scale and jet shape corrections for multiple interactions in the first ATLAS data at the LHC*, Tech. Rep. ATLAS-CONF-2011-030, CERN, Geneva, Mar, 2011.

[74] ATLAS Collaboration, *Jet energy scale and its systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s}$ = 7 TeV and measured with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2010-056, CERN, Geneva, Jul, 2010.

[75] ATLAS Collaboration, *Update on the jet energy scale systematic uncertainty for jets produced in proton-proton collisions at $\sqrt{s} = 7$ TeV measured with the ATLAS detector*, Tech. Rep. ATLAS-CONF-2011-007, CERN, Geneva, Feb, 2011.

[76] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **07** (2003) 001, `arXiv:0206293 [hep-ph]`.

[77] G. Corcella, I. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al., *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*, JHEP **0101** (2001) 010, arXiv:0011363 [hep-ph].

[78] J. M. Butterworth and M. H. Seymour, *JIMMY4: Multiparton Interactions in Herwig for the LHC*, October, 2004. http://projects.hepforge.org/jimmy/.

[79] ATLAS Collaboration, *ATLAS Calorimeter Response to Single Isolated Hadrons and Estimation of the Calorimeter Jet Scale Uncertainty*, Tech. Rep. ATLAS-CONF-2010-052, CERN, Geneva, Jul, 2010. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2011-028.

[80] DØ Collaboration, B. Abbott et al., *High-p_T jets in $\bar{p}p$ collisions at $\sqrt{s} = 630$ GeV and 1800 GeV*, Phys. Rev. **D64** (2001) 032003, arXiv:0012046 [hep-ex].

[81] UA2 Collaboration, *Measurement of jet production properties at the CERN Collider*, Physics Letters B **144** (1984) no. 3-4, 283 – 290.

[82] ATLAS Collaboration, *Updated Luminosity Determination in pp Collisions at root(s)=7 TeV using the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2011-011, CERN, Geneva, Mar, 2011.

[83] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, et al., *New parton distributions for collider physics*, Phys.Rev. **D82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[84] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur.Phys.J. **C63** (2009) 189–285, arXiv:0901.0002 [hep-ph].

[85] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, et al., *Impact of Heavy Quark Masses on Parton Distributions and LHC Phenomenology*, arXiv:1101.1300 [hep-ph].

[86] R. D. Ball, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, et al., *A first unbiased global NLO determination of parton distributions and their uncertainties*, Nucl.Phys. **B838** (2010) 136–206, arXiv:1002.4407 [hep-ph].

[87] S. Alekhin, S. Alioli, R. D. Ball, V. Bertone, J. Blumlein, et al., *The PDF4LHC Working Group Interim Report*, arXiv:1101.0536 [hep-ph].

[88] G. Soyez, *A simple description of jet cross-section ratios*, Phys. Lett. **B698** (2011) 59–62, arXiv:1101.2665 [hep-ph].

[89] G. Soyez. Private communication.

[90] Y. L. Dokshitzer, A. Lucenti, G. Marchesini, and G. Salam, *On the universality of the Milan factor for 1 / Q power corrections to jet shapes*, JHEP **9805** (1998) 003, arXiv:9802381 [hep-ph].

[91] P. Movilla Fernandez, *Alpha(S) and power corrections from JADE*, arXiv:0209022 [hep-ex].

[92] 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, Mar, 2011.

[93] Atlas Collaboration, *Measurement of underlying event characteristics using charged particles in pp collisions at* $\sqrt{s} = 900\ GeV$ *and 7 TeV with the ATLAS detector*, `arXiv:1012.0791 [hep-ex]`.

[94] R. J. Barlow, *Statistics: a guide to the use of statistical methods in the Physical Sciences*. WILEY, 1991.

A Appendix

A.1 Construction of the full covariance matrix for the cross section ratio

The construction of the full covariance matrix is possible by strictly following standard error propagation (see e.g. [94]). The covariance of a series of functions $\{f_i\}$ of variables $\{x_i\}$ is given by:

$$cov(f_k, f_l) = \sum_i \sum_j \left(\frac{\partial f_k}{\partial x_i}\right) \left(\frac{\partial f_l}{\partial x_j}\right) \mathrm{cov}(x_i, x_j) \tag{A.1}$$

For the application to the jet cross section ratio, the function f_i is the ratio of cross sections in the i^{th} momentum-bin and the variables $\{x_i\}$ are the two measured cross sections in each bin.

$$\begin{aligned} \{f_j\} &= \{R_i\} \\ \{x_j\} &= \{y_i^6, y_i^4\} \\ R_i &= \frac{y_i^4}{y_i^6} \end{aligned} \tag{A.2}$$

Here, y_i^4 and y_i^6 are the measured cross sections in the i^{th} momentum-bin for $R = 0.4$ and $R = 0.6$ respectively. Thus, for n bins of transverse momentum there are n functions f_j and $2n$ variables x_j. Due to the definition of the ratio, the derivatives in Eq. A.1 are mostly zero:

$$\text{for all } \; k \neq i: \quad \frac{\partial f_k}{\partial y_i^4} = \frac{\partial f_k}{\partial y_i^6} = 0 \tag{A.3}$$

Thus only three different contributions arise from the sums in Eq. A.1. Namely these are two quadratic ones for $(x_i, x_j) = (y_k^6, y_l^6)$ or $(x_i, x_j) = (y_k^4, y_l^4)$. And in addition, there is the mixed contribution for $(x_i, x_j) = (y_k^6, y_l^4)$ or $(x_i, x_j) = (y_k^4, y_l^6)$, which are equal. Accordingly for the concrete application Eq. A.1 can be written as:

$$\begin{aligned} \mathrm{cov}(R_k, R_l) = \left(\frac{\partial R_k}{\partial y_k^6}\right) \left(\frac{\partial R_l}{\partial y_l^6}\right) \mathrm{cov}(y_k^6, y_l^6) \;&+\; \left(\frac{\partial R_k}{\partial y_k^4}\right) \left(\frac{\partial R_l}{\partial y_l^4}\right) \mathrm{cov}(y_k^4, y_l^4) \\ &+\; 2\left(\frac{\partial R_k}{\partial y_k^6}\right) \left(\frac{\partial R_l}{\partial y_l^4}\right) \mathrm{cov}(y_k^6, y_l^4) \end{aligned} \tag{A.4}$$

This can be further simplified, by switching to relative uncertainties, which is easily achievable by division by $1/(R_k R_l)$ and the insertion of the derivatives:

$$\frac{\mathrm{cov}(R_k, R_l)}{R_k R_l} = \frac{\mathrm{cov}(y_k^6, y_l^6)}{y_k^6 y_l^6} + \frac{\mathrm{cov}(y_k^4, y_l^4)}{y_k^4 y_l^4} - 2\,\frac{\mathrm{cov}(y_k^6, y_l^4)}{y_k^6 y_l^4} \tag{A.5}$$

Denoting the measurement uncertainties on y_i^R as σ_i^R, this can explicitly be written as:

$$\frac{\mathrm{cov}(R_k, R_l)}{R_k R_l} = \rho_{kl}^6 \frac{\sigma_k^6 \sigma_l^6}{y_k^6 y_l^6} + \rho_{kl}^4 \frac{\sigma_k^4 \sigma_l^4}{y_k^4 y_l^4} - 2\rho_{kl}^{64} \frac{\sigma_k^6 \sigma_l^4}{y_k^6 y_l^4} \tag{A.6}$$

This prescription formally treats all possible correlations in the correct way. For $k = l$ Eq. A.6 results in the equations 10.19 and 10.20 using the appropriate uncertainties and correlations. For $k \neq l$ in the case of the systematic uncertainties, the lack of knowledge about ρ_{kl}^{64} leads to problems described in Section 10.4.1.

A.2 Inclusive cross section PDF comparisons in different rapidity regions

Here, in addition to Fig. 9.19, comparisons of the measured cross sections to various PDF sets are shown.

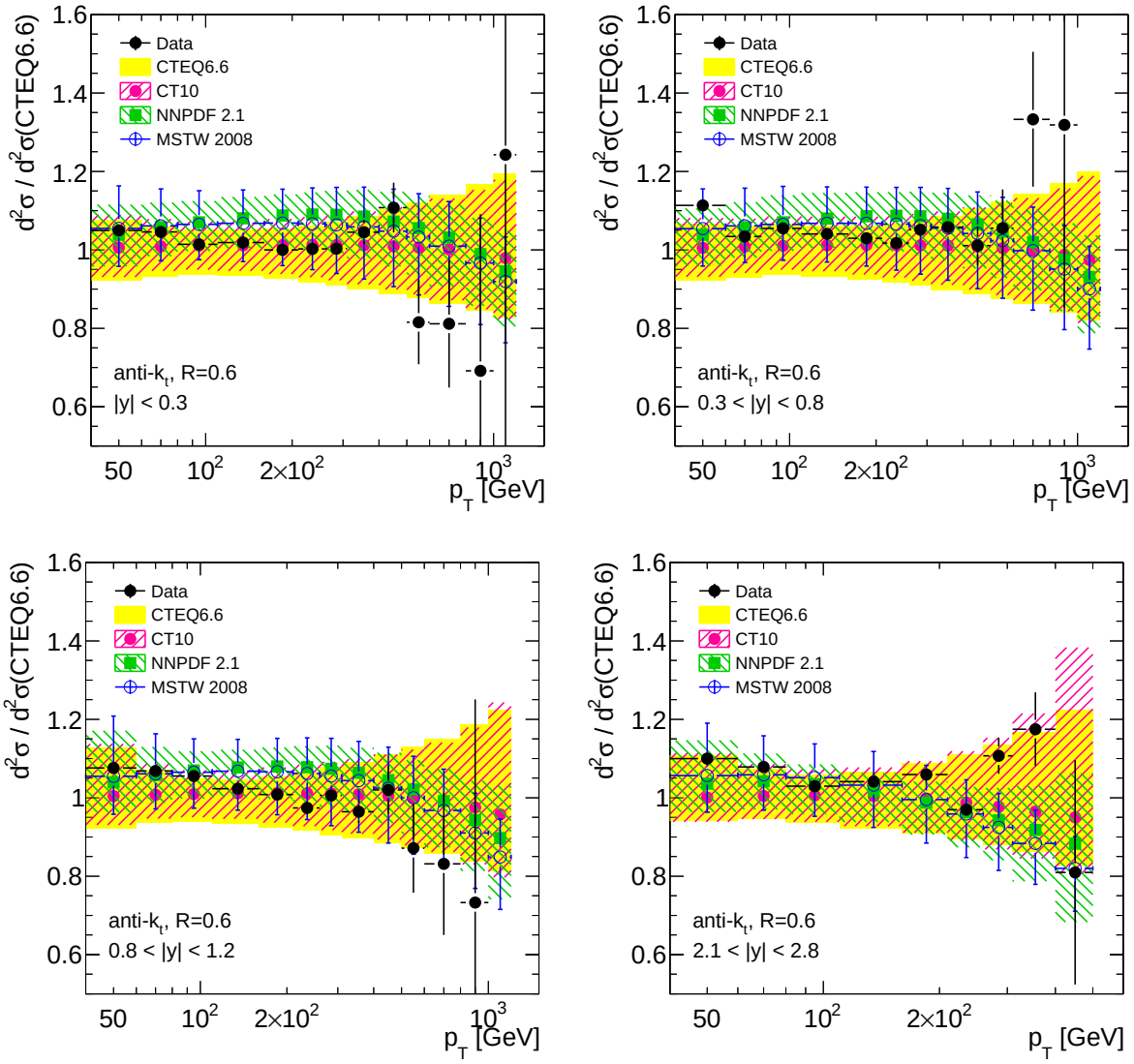

Figure A.1: *Ratios of theoretical predictions using various PDFs for anti-k_t, $R = 0.6$ jets. The measured cross sections and the theoretical predictions are divided by the theoretical prediction using the CTEQ6.6 PDF. The systematic uncertainties on data are not shown. Systematic uncertainties on the theoretical predictions were calculated at a confidence level of 90 %.*

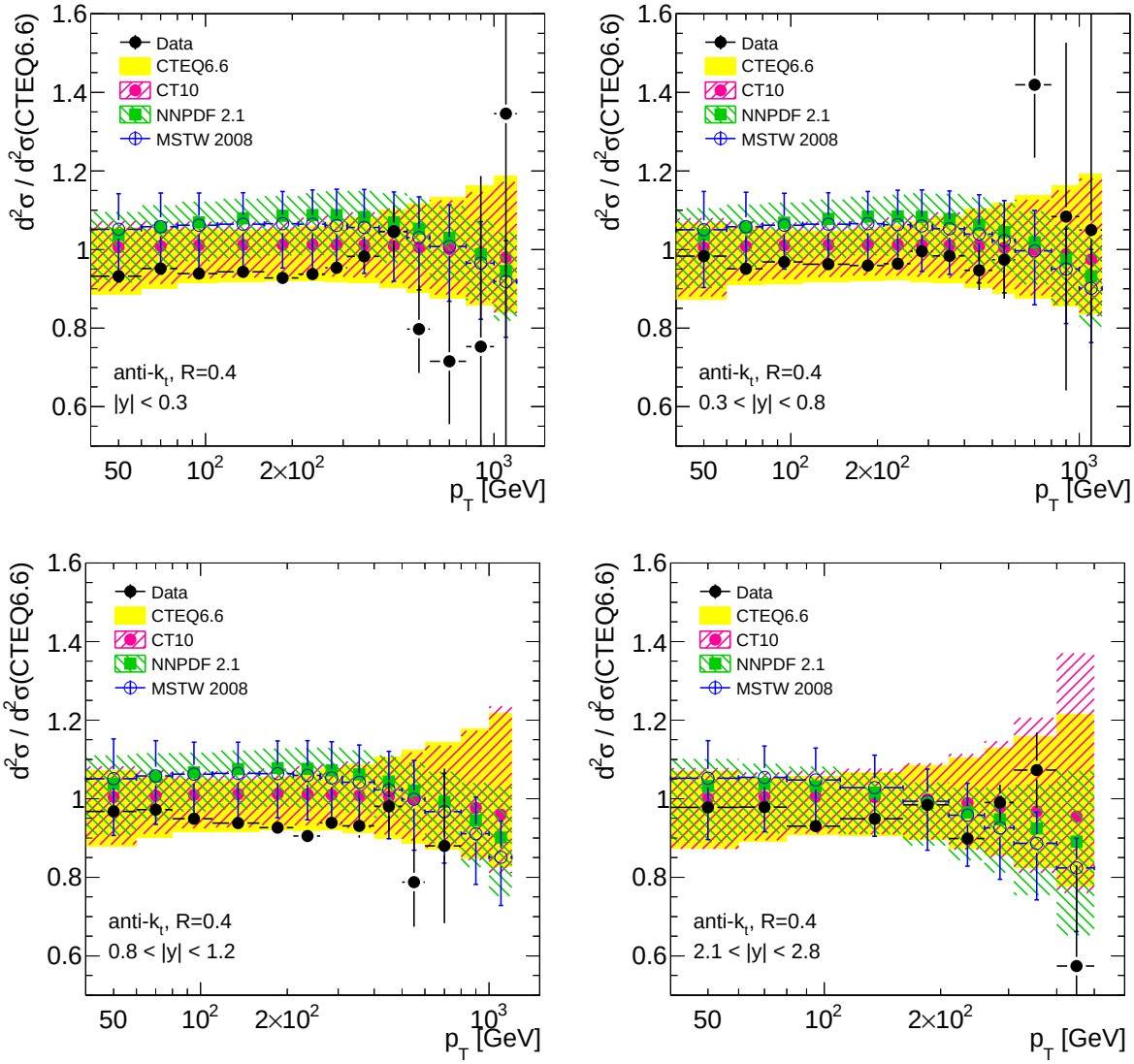

Figure A.2: *Ratios of theoretical predictions using various PDFs for anti-k_t, $R = 0.4$ jets. The measured cross sections and the theoretical predictions are divided by the theoretical prediction using the CTEQ6.6 PDF. The systematic uncertainties on data are not shown. Systematic uncertainties on the theoretical predictions were calculated at a confidence level of 90 %.*

A.3 Displays of events with high $E_{T,Miss}^{Sig}$

Here, two exemplary events that are rejected by the $E_{T,Miss}^{Sig}$ selection are shown. The first of which, shown in Fig. A.3, is a very unambiguous example of a very soft collision event that is overlayed with energy depositions by cosmic particles.

In these event displays, the red area symbolizes the Tile calorimeter, while the inner, green area represent the em. calorimeters. The calorimeter's segmentation is shown is well. Yellow areas within the calorimeters indicate deposited energy, where the area is proportional to the energy. Histogram structures surrounding the two calorimeter systems symbolize these energy depositions as well. These are shown in green for the em. calorimeter and red for the hadronic calorimeters.

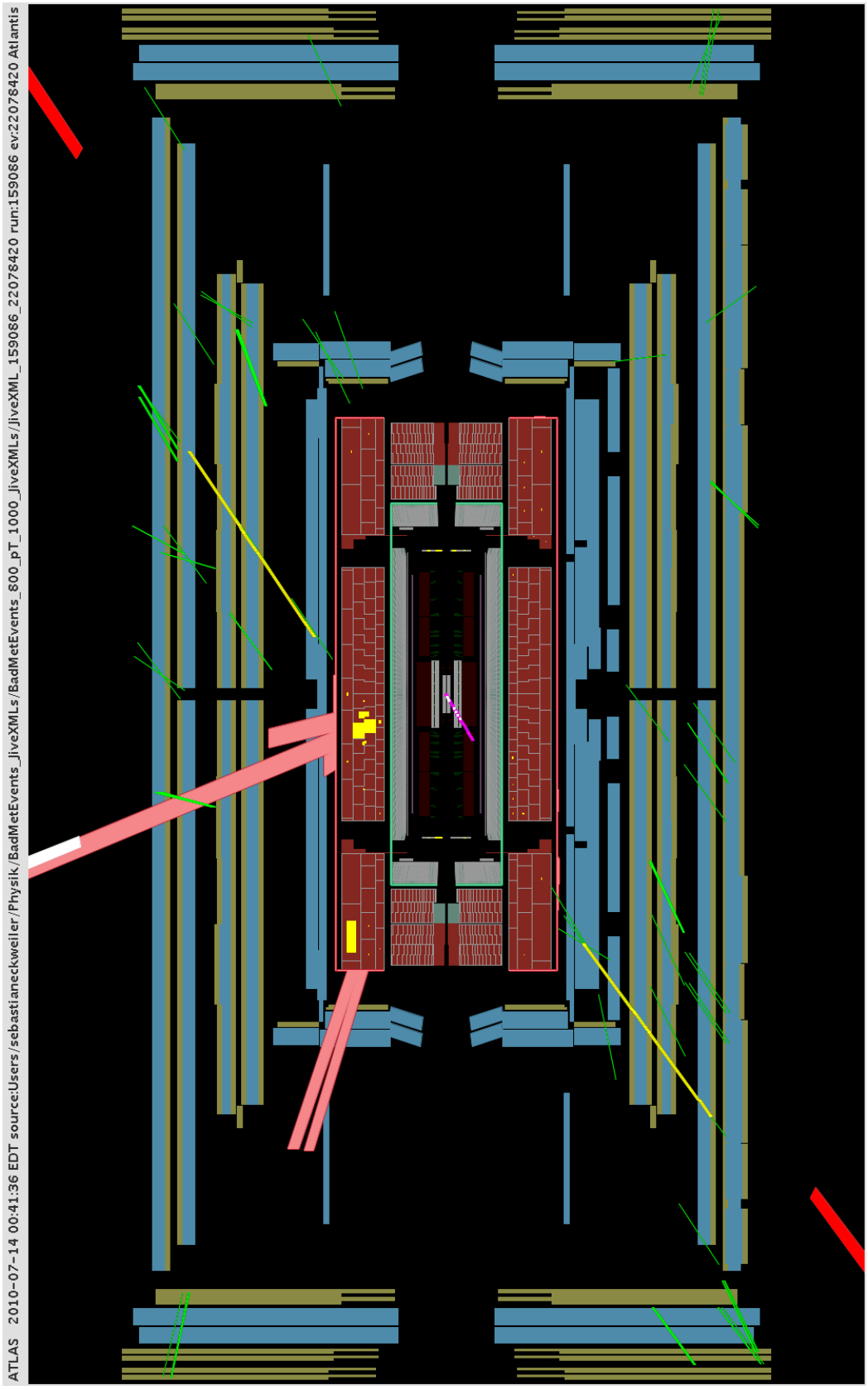

Figure A.3: *Display of an event with $E_{T,Miss}^{Sig} = 35.3$ GeV$^{-1/2}$ in the r-z-projection. The green line segments represent hits in the muon system, which are combined to muon track segments (yellow lines). In this particular event, the energy depositions in the upper half of the calorimeter are most probably induced by cosmic particles. The energy depositions in the calorimeter where reconstructed as jets with transverse momenta of approximately 1 TeV for the more central jet and 500 GeV.*

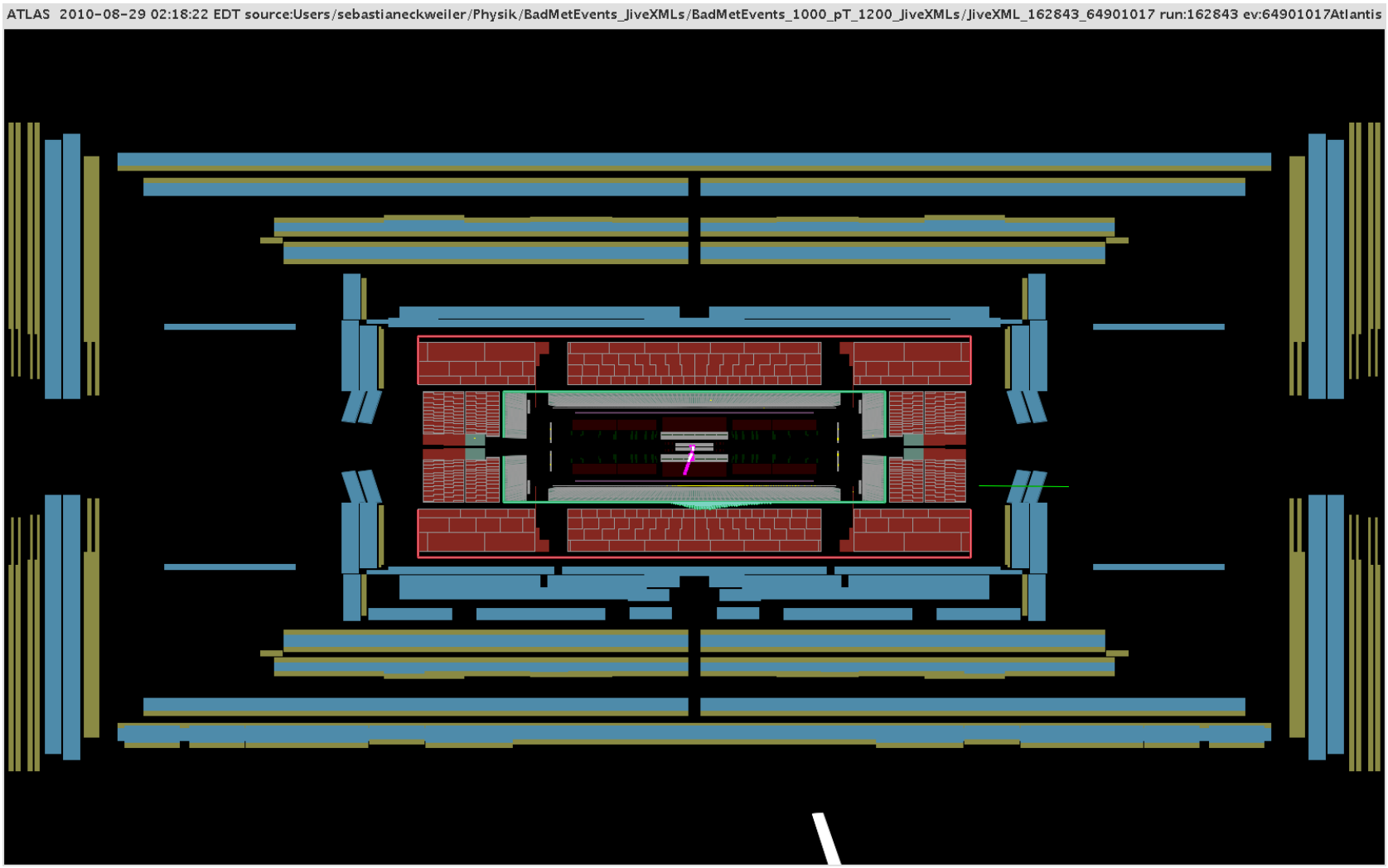

Figure A.4: *Display of an event with* $E_{T,Miss}^{Sig} = 28.0\ \mathrm{GeV}^{-1/2}$. *This event can very clearly be classified as one dominated by a particle out of the beam halo. The beam halo particle travels very precisely in parallel to the beam axis, within the em. calorimeter, leaving an energy deposit of approximately* 1.2 TeV, *that is reconstructed as a jet. The histogram representation of the em. energy deposits clearly shows the longitudinal structure of the electromagnetic shower. A very soft collision event at the same time leads to the event's acceptance.*

Danksagungen

Aus Datenschutzgründen entfernt.