STRATEGY FOR AN INITIAL MEASUREMENT OF THE INCLUSIVE JET CROSS SECTION WITH THE CMS DETECTOR

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl.-Phys. Andreas Oehler
aus Öhringen

Tag der mündlichen Prüfung: 4.12.2009

Referent: Prof. Dr. G. Quast
Institut für Experimentelle Kernphysik

Korreferent: Prof. Dr. W. de Boer
Institut für Experimentelle Kernphysik

Contents

Titlepage **I**

Contents **i**

1 Introduction **1**

2 The Theory of the Strong Interaction **5**
2.1 The Standard Model 6
2.2 Cross Sections 8
2.3 Quantum Chromodynamics 9
2.3.1 The QCD Lagrangian 11
2.3.2 The running coupling $\alpha_S(Q^2)$ of the Strong Interaction 14
2.3.3 Hadrons 18
2.3.4 The QCD Improved Parton Model 19
2.4 Hadron-Hadron scattering at the LHC 20
2.4.1 Jet Cross Sections 22
2.5 From Quarks to Hadrons 23
2.5.1 Parton Shower 24
2.5.2 Hadronisation 26
2.6 Jet Algorithms 30
2.6.1 Collinear Safety 32
2.6.2 Infrared Safety 33
2.6.3 Classification of Jet Algorithms 34
2.6.4 The $\mathbf{k_T}$ Algorithm 35

2.6.5 The Seedless Infrared-Safe Cone Algorithm . . . 36
2.7 Latest Results from the Tevatron . . . 38
2.7.1 Jet Reconstruction . . . 38
2.7.2 Theory Comparison . . . 39

3 The CMS Experiment 43
3.1 Large Hadron Collider . . . 43
3.2 CMS Collaboration . . . 46
3.3 Compact Muon Solenoid . . . 48
3.3.1 The Detector Coordinate System . . . 48
3.3.2 Tracking System . . . 51
3.3.3 Electromagnetic Calorimeter (ECAL) . . . 52
3.3.4 Hadron Calorimeter (HCAL) . . . 52
3.3.5 Superconducting Solenoid . . . 54
3.3.6 Muon Chambers . . . 54
3.3.7 Data Acquisition and Trigger . . . 55
3.4 Worldwide LHC Computing Grid (WLCG) . . . 55
3.4.1 Grid Computing . . . 56
3.4.2 WLCG Structure . . . 57
3.4.3 Grid Services . . . 57
3.4.4 Grid Usage . . . 59
3.4.5 CMS Data Management . . . 60
3.5 Datastructure and Software . . . 62
3.5.1 Event Data Model . . . 62
3.5.2 ROOT . . . 62
3.5.3 matplotlib . . . 63
3.5.4 SCRAM . . . 63
3.5.5 The CMSSW Framework . . . 63
3.5.6 Datatiers . . . 64
3.6 Monte Carlo Methods . . . 66
3.6.1 Multi-Purpose Event Generators: PYTHIA and HERWIG(++) . . . 66
3.6.2 Signal Generators and Alternative Approaches . . . 67
3.6.3 Detector Simulation . . . 68
3.6.4 NLOJET++ . . . 68
3.6.5 fastNLO . . . 68

4 Prediction of the Inclusive Jet Cross Section 71
4.1 Calculations in NLO perturbative QCD . . . 72
4.1.1 Intrinsic Uncertainties . . . 72
4.1.2 Binning in Rapidity and Jet Transverse Momentum . . . 75
4.1.3 Scale Uncertainty . . . 76

4.1.4 The Strong Coupling α_S . . . 80
4.1.5 PDF Uncertainties . . . 81
4.1.6 K-Factors and Comparison to PYTHIA . . . 85
4.2 Non Perturbative Corrections . . . 94
4.2.1 Monte Carlo Generators and Tunes . . . 94
4.2.2 Spectrum Construction for Inclusive Jets . . . 96
4.2.3 Derivation of the Correction Factors and their Uncertainties . . . 97

5 Measurement of the Inclusive Jet Cross Section with CMS 103
5.1 Observable . . . 104
5.2 Jet Reconstruction . . . 106
5.2.1 Jet Energy Determination . . . 106
5.2.2 Jet-by-Jet MC Truth Comparison . . . 108
5.2.3 Jet Energy Resolution . . . 109
5.3 Event Clean-up . . . 113
5.4 Trigger Requirements and Spectrum Construction . . . 113
5.5 Expected Statistical Precision . . . 116
5.6 Experimental Corrections . . . 116
5.6.1 Jet Energy Scale . . . 117
5.6.2 Resolution Unsmearing . . . 117
5.7 Experimental Systematic Uncertainties . . . 119
5.7.1 Jet Energy Scale Uncertainty . . . 120
5.7.2 Luminosity . . . 122
5.7.3 Unsmearing Uncertainty . . . 123
5.7.4 Comparison to Particle Jets . . . 123
5.8 Theory Prediction Summary . . . 128
5.9 Non-perturbative Corrections . . . 128
5.10 Final Comparison . . . 130

6 Conclusion and Outlook 137

A Appendix - Theoretical Details 141
A.1 The Dirac γ Matrices . . . 141
A.2 The Gell-Mann Matrices . . . 142

B Appendix - Technical Details 143
B.1 Jet Algorithms . . . 143
B.1.1 The $\mathbf{k_T}$ Algorithm . . . 143
B.1.2 SISCone . . . 144
B.2 Grid Services . . . 146
B.3 Usage of fastNLO tables . . . 148

B.3.1 The FORTRAN Version 149
B.3.2 The C++ Version 150
B.4 The Data Samples 151
B.4.1 The Cosmic Data Samples 151
B.4.2 The Monte Carlo Simulation Samples 152

List of Figures **155**

List of Tables **159**

Bibliography **161**

Acknowledgements **177**

Chapter 1

Introduction

Particle physics experiments study the currently smallest known objects and their interactions. Conducted at energy densities comparable to the situation shortly after the big bang high energy physics experiments reveal the basic properties of the fundamental building blocks of matter within the universe. Understanding these phenomena from first principles would finally help to explain how the universe developed into the currently observed state.

Our most up-to-date knowledge about the fundamental forces and particles is, however, far from being complete. Almost all current information is condensed into the Standard Model of particle physics. It is incredibly successful in describing all known interactions tested at particle physics experiments during the last 50 years. Yet, it is a model which requires several parameters that have to be determined by experiments, and it is not able to describe gravity. Additionally, a central component, which manifests itself in the existence of the Higgs boson, could not be experimentally confirmed up to now.

More open questions are posed by astronomical observations. There are strong indications for the existence of invisible matter distributions in the universe ("dark matter"), which manifest themselves in the observed motion of solar systems within galaxies. New theoretical models have been developed to explain these observations by introducing particles with masses above the reach of current collider experiments. Furthermore, the universe is expanding with increasing acceleration. This raises the question about the so called "dark energy" which seems to drive the expansion.

Advancing on the way to find a more fundamental explanation for the observed properties of matter, interactions, and the structure of the universe, which might be possible in form of a grand unified theory (GUT), continuously new experiments are required to test

the theories in subsequent energy density regions. The next step on this path is the Large Hadron Collider (LHC) and the associated experiments at the **C**onseil **E**uropéen pour la **R**echerche **N**ucléaire (CERN), Switzerland, which are introduced in Chapter 3 with a special focus on the CMS experiment.

Most searches at the LHC will focus on the analysis of collisions between protons, which are not elementary particles. Therefore, a good knowledge of the proton substructure described by parton density functions (PDFs) is vital for interpreting experimental observations. For reactions with large momentum transfers, these can be understood as the interaction among the individual proton constituents. Each struck quark or gluon manifests itself in the detector as a collimated stream of particles, usually referred to as a jet.

Quantum chromodynamics (QCD), the theory of the strong interaction, is the theoretical basis for the mathematical description of proton-proton collisions. The emergence of jets as well as the possibility to make use of universal parton distribution functions to describe proton-proton scattering follows as a fundamental feature of this theory. This is outlined in Chapter 2 together with an introduction to supplementary phenomenological models and to jet algorithms, which relate the measurements to the quarks and gluons of the theoretical calculations.

The inclusive jet cross section differential in transverse momentum and rapidity predicts the production rate of jets for a given luminosity. A precise measurement of this observable constrains the proton PDFs and allows simultaneously an extraction of the strong coupling constant α_S, one of the free parameters of the Standard Model. Deriving these fundamental quantities is a vital part of testing the validity of an extrapolation to the new energy ranges and is absolutely necessary before claims of new physics observations from jets could be published. Especially an untested extrapolation of PDFs is a prominent example, which lead to all kinds of speculations in the past [1]. However, current PDFs are much better constrained for the LHC scenario due to the HERA measurements than they were at the start-up of the Tevatron.

A prediction of the inclusive jet cross section in form of calculations at NLO and the respective uncertainties are presented in Chapter 4. The results are derived in phase space ranges adapted to the geometry and expected performance of the CMS detector. Such calculations do not reflect the fully hadronised final state, therefore the required additional corrections to the NLO predictions are addressed as well.

Measuring the inclusive jet cross section is an important test to demonstrate the capabilities of the detector and analysis framework to correct for experimental effects. Due to the large amount of high p_{T} jets that will be produced at the LHC over a good part of the studied phase space, the observable can be determined with rather low statistical uncertainty early on. However, the experimental challenge is to handle the influence of the energy determination within the calorimeter system of the detector, which has a strongly non-linear response with respect to the incoming particle type and thus directly affects the measured jet energies. The analysis presented in Chapter 5, evaluates an initial measurement of this observable within the experimental uncertainty given by early available, data

driven correction methods. The results are compared to the previously presented theoretical predictions. A comparison of the dominant uncertainties allows a prospect for the impact of such a measurement on our current knowledge about physics at the TeV scale.

Chapter 2

The Theory of the Strong Interaction

High energetic interactions of two protons are usually studied in experiments which collide parallel proton beams at an interaction point. The process is understood as a 2-2 scattering process due to the strong interaction of single partons. During the interaction, the involved partons gain momentum in the plane transverse to the beam direction. In the following they radiate off additional partons which finally hadronise together with the initial parton. Due to the characteristic of the strong interaction, described by quantum chromodynamics (QCD), these splitting products are not homogeneously distributed in phase space, but strongly collimated into one direction. Such a collimated stream of particles is usually referred to as a jet.

During the development of a theory of the strong interaction, it was not initially obvious, that such a jet-like structure would be an underlying property of the interaction. This became obvious when first evidence for a jet structure in hadron production of e^+e^- annihilation at 6.2 and 7.4 GeV was found at SLAC in 1975 [2]. Inspired by this experimental finding, Sterman and Weinberg showed 1977 [3] that the jet structure is a fundamental feature of quantum chromodynamics. With rising momentum transfer, QCD processes can increasingly well be described using perturbative methods while in parallel the jet structure of the events becomes more and more obvious. Therefore, given large enough momentum transfer of the interacting partons, jet observables are ideal candidates to test the predictions of perturbative QCD (pQCD) and our knowledge of parton distribution functions (PDFs).

Such an observable is the inclusive jet cross section, which describes the production rate of all jets produced in specific regions of the phase space. Therefore, it provides an

Table 2.1: *Fundamental fermions of the Standard Model. In case of the down-type quarks, the third component of the weak isospin is given for the Cabibbo-rotated mass-eigenstates.*

fermions	family			electromagnetic	colour	spin	weak isospin	
	1	2	3	charge [units of e^+]	charge	J	$T_{3,L}$	$T_{3,R}$
leptons	ν_e	ν_μ	ν_τ	0	-	$1/2$	$+1/2$	-
	e^-	μ^-	τ^-	-1	-	$1/2$	$-1/2$	0
quarks	u	c	t	$+2/3$	r,g,b	$1/2$	$+1/2$	0
	d	s	b	$-1/3$	r,g,b	$1/2$	$-1/2$	0

ideal test of detector understanding, as it can be used to differentially describe a vast part of the accessible phase space in transverse momentum. It is very sensitive to the absolute energy measurement and its resolution. In the following, a short introduction to the Standard Model of Particle Physics will be given, followed by a more detailed description of quantum chromodynamics with a focus on LHC physics. Finally, the current experimental findings in the field of the inclusive jet cross sections from the Tevatron will be summarised.

2.1 The Standard Model

Our current knowledge on the fundamental particles and the forces in-between them is accumulated in the Standard Model of particle physics. There are twelve fundamental particles which build up the known matter in the universe (tab. 2.1). For each of these particles, an antiparticle exists which has the same mass, but opposite electric charge, the corresponding anti-colour, and opposite component of the weak isospin. The particles are arranged in three generations. The known stable matter of the universe solely consists of up-quarks, down-quarks, neutrinos and electrons. Matter consisting out of other particles is usually unstable and can only be produced temporarily within particle accelerators, supernovae, or other objects which can reach comparable energy densities.

Currently there are four known interactions, namely the strong interaction, the weak interaction, the electromagnetic interaction and gravity, from which all but gravity are described by the Standard Model. Each interaction is described by the exchange of their dedicated gauge bosons (tab. 2.2). Especially the actual production of $W^\pm$ and Z^0 vector bosons with the expected properties lead to increasing confidence in the Standard Model. The relative strength of each interaction varies with the distance and strongly depends on the respective peculiarities of the interactions. The reach of the electromagnetic force is in principle unlimited due to the massless mediating photons. However, over macroscopic distances the original charge is shielded by vacuum polarisation effects. The impact of the weak and the strong interaction is limited to distances well below $1\ \mathrm{fm}$, which is the radius

Table 2.2: *Elementary interactions and intermediate gauge vector bosons of the Standard Model, J^P is spin parity quantum number*

interaction	coupling with	intermediate boson	boson mass (GeV/c^2)	J^P
strong	colour	8 gluons (g)	0	1^-
electomagn.	el. charge	photon (γ)	0	1^-
weak	weak charge	$W^{\pm}$, Z^0	≈ 100	1

of a proton. That is why they could not be observed on macroscopic scales. While the reach of the weak interaction is limited by the masses of the gauge bosons, the reach of the strong interaction is limited by the self-coupling of the gluons. The distance dependence can directly be mapped into an energy scale dependence of the interaction by the Heisenberg uncertainty principle. Thus, the interactions are dependent on the energy scale of the processes. The relative strength of the interactions lead to the naming of the strong interaction which at momentum transfers of $Q = 1$ GeV is about 100 times stronger than the electromagnetic interaction and about 10^6 times stronger than the weak interaction.

Using quantum field theory, the fundamental particles of the Standard Model are described as exited states of space-time coordinate dependent quantum fields. The dynamics of these fields are described using Lagrangian field theory and requiring local gauge invariance. Following Noether's theorem, symmetries and conservation laws are closely linked, thus it was only natural to use a model based on symmetry groups, where discrete symmetries are the conserved quantum numbers of the theory. Evolving from the Dirac theory, the first relativistic gauge theory developed was quantum electrodynamics (QED), which successfully implemented a U(1) symmetry naturally by requiring local gauge invariance to describe the electromagnetic interactions of fermions via a mediating photon field. Motivated by the success of this description, a relativistic gauge theory for spin $\frac{1}{2}$ particles with local gauge invariance was already formulated by Yang and Mills for SU(2) in 1954, trying to describe the proton and the neutron. They extended the scalar symmetry of U(1) by using a matrix instead. However this was not successful, and the theory just came to life within quantum chromodynamics (QCD) and electroweak theory much later. Using this formalism, the Standard Model actually uses a spontaneously broken $SU(2)_L \otimes U(1)_Y$ symmetry for the electroweak theory and an unbroken $SU(3)_C$ colour gauge theory for the strong interaction. More details of QCD will be outlined in 2.3, however the interested reader is pointed to the excellent literature on the topic, e.g. [4–7].

Although the Standard Model has been extraordinary successful up to now in describing the observations of experiments, there are several fundamental parameters which have to be determined by experiments. The total number of parameters and their specific nature depends on the actual formulation of the model and how new developments, like the neutrino mixing matrix, are handled. Additionally, the Higgs mechanism, introduced in order

to describe the electroweak symmetry breaking in form of a local gauge invariant theory, is not yet confirmed. Thus a potential discovery, or a possible exclusion of the Higgs Boson is one of the major goals of the experiments at the LHC (Sec. 3.1). Several observations, like CP violating processes or more basically the actual values of particle masses, which are not explained, point out that the Standard Model is not the end of the story. New models beyond the Standard Model could not yet be confirmed, however, the non existence of the Higgs Boson would definitely require new physical concepts.

2.2 Cross Sections

For the comparison of collider experiments with theory, a quantity is needed which is calculable from the theory and also measurable by experiments. Such a quantity is the cross section, which relates the particle flux Φ and the interaction rate per target particle W:

$$\sigma = \frac{W}{\Phi} \, . \tag{2.1}$$

The unit of cross section is defined as the barn: $1\mathrm{b} = 10^{-28}\,\mathrm{m}^2$. This is roughly the geometrical cross section of a nucleus of mass number $A = 100$.

The interaction rate of a specific process is connected to the transition matrix element $|M_{if}|$ of the interaction and the energy density ρ_f of the final states available in phase space using Fermi's golden rule:

$$W = \frac{2\pi}{\hbar} |M_{if}|^2 \rho_f \, . \tag{2.2}$$

To gain a deeper understanding of the processes, cross sections are calculated and measured differentially. This can be either differential or double-differential in angle, energy, or other suitable quantities. Usually one is restricted to regions of the phase space within the limits of the experimental apparatus or the theoretical description. To relate machine properties of colliders to the number of produced events, the luminosity $\mathcal{L}$ is used which in case of the LHC is given in section 3.1. This quantity depends on parameters of the design like the number of protons per bunch, the number of bunches, properties of the beam optics, and the energy of the beam. The actual instantaneous luminosity will have to be determined experimentally and therefore the cross sections are estimated using

$$\sigma = \frac{W}{\mathcal{L}} \, . \tag{2.3}$$

Different approaches are used to determine the luminosity experimentally. One method exploits the fact, that the total pp or $p\bar{p}$ cross section is related to the nuclear elastic forward scattering amplitude $f_{\mathrm{el}}(0)$ via the Optical Theorem [8]. With specialised adaptations of

the experimental settings and using special forward detectors, which can operate at a minimal distance to the beam, the rates of elastic (N_{el}) and inelastic (N_{inel}) collisions can be measured. Together with differential elastic rate dN_{el}/dt at a squared four momentum transfer $t = 0$ the luminosity and the total cross section can be extracted:

$$\sigma_{\mathrm{tot}} = \frac{16\pi}{1+\rho^2} \cdot \frac{dN_{\mathrm{el}}/dt|_{t=0}}{N_{\mathrm{el}} + N_{\mathrm{inel}}}, \quad \mathcal{L} = \frac{1+\rho^2}{16\pi} \cdot \frac{(N_{\mathrm{el}} + N_{\mathrm{inel}})}{dN_{\mathrm{el}}/dt|_{t=0}} \,. \tag{2.4}$$

In this case $\rho = \frac{\mathrm{Re}[f_{\mathrm{el}}(0)]}{\mathrm{Im}[f_{\mathrm{el}}(0)]}$ needs at first to be taken from theory, but later it might be be even measured via the interference between Coulomb and hadronic contributions to the elastic scattering cross section. A systematic contribution to the uncertainty of the measurement is introduced by the required extrapolations to $t = 0$. This measurement is part of the scientific programme of the TOTEM experiment [9].

Another possibility to measure the luminosity, or cross check the values retrieved by the above methods is making use of the clear signature of W- or Z-boson decays. Given the precise knowledge on these physical quantities from previous experiments and relying on the validity of the Standard Model, one can extract the luminosity. Other methods exist to monitor the instantaneous luminosity in parallel to the measurements with less accuracy.

2.3 Quantum Chromodynamics

Quantum chromodynamics is the theory of the strong interaction, which is the binding force of hadrons. In a diminished form the force is also responsible for holding together the nuclei. After having revealed the substructure of the proton, which lead to the quark parton model, it became necessary to describe the constituent particles and their interactions with a new theory. First observations indicated that the existence of quarks alone was not the whole story. Bound quark states, like the Δ^{++} resonance consisting of 3 up quarks with parallel spin, required the introduction of an additional quantum number, subsequently named colour, in order to save the anti-symmetry of the wave function, as required by the Pauli-Principle for bound fermion states. Additionally this "spin-statistics problem" occurs with the other barions made of three equal quarks with parallel spin, like the Δ^- (3 d quarks) and the Ω^- (3 s quarks). So QCD introduces three colours (red, green, blue) and three anti-colours (anti-red, anti-green, anti-blue), which allow to describe the observations. A possible representation of these colour states as a singlet and an octet is given in Table 2.3. The singlet is colourless, so it does not contribute to the interaction, while the other 8 gluons make up a representation of SU(3).

Using only three colours was not the only possibility to solve the above puzzle. The decay rate of the $\pi^0 \rightarrow \gamma\gamma$ theoretically depends on the number of quark colours N_c. However, also pointed out by Abbas [10], this only gives knowledge about the number of colours, if the charges of the quarks can be determined independently. Assuming the

Table 2.3: *A possible representation of the colour singlet and triplet of SU(3) symmetry is listed below. For actual calculations these representations require further symmetrisations.*

symmetry	representation							
octet	$r\bar{g}$	$r\bar{b}$	$g\bar{b}$	$g\bar{r}$	$b\bar{r}$	$b\bar{g}$	$\sqrt{1/2}(r\bar{r}-g\bar{g})$	$\sqrt{1/6}(r\bar{r}+g\bar{g}-2b\bar{b})$
singlet	$\sqrt{1/3}(r\bar{r}+g\bar{g}+b\bar{b})$							

charges given by the quark model, which is supported by measurements performed in deep inelastic scattering, one finds [6]:

$$\Gamma(\pi^0 \to \gamma\gamma) = 7.63 \text{ eV} \left(\frac{N_c}{3}\right)^2 . \tag{2.5}$$

The experimental result of 7.84 ± 0.56 eV gave reason to believe, that N_c should actually be three. The strongest evidence, that there should be three colours is given by the ratio

$$R_\gamma = \frac{\sigma(\mathrm{e^+e^-} \to \text{hadrons})}{\sigma(\mathrm{e^+e^-} \to \mu^+\mu^-)} = N_c \sum_q e_q^2 = N_c \frac{11}{9} . \tag{2.6}$$

Measurements of R_γ yield $N_c \approx 3.2$, which indicates that three colours are preferred. The reason for the value being slightly above 3.0 can be explained by higher order QCD corrections.

From deep inelastic electron-nucleon scattering experiments, it was found, that only half of the proton momentum was actually carried by the quarks. The missing momentum could be attributed to the field bosons of the strong interaction. So QCD is a Yang-Mills gauge theory [11] based on a non abelian SU(3) symmetry. Thus the mediating bosons, the gluons, can, in contrast to photons, couple to other gluons because they also carry colour charge.

No coloured objects exist on the macroscopic scale, because gluon self coupling enforces, that enough energy is gathered in the colour field between coloured objects to create new separated colourless particles from the vacuum in case the colour connection reaches distances of the order of 1 fm. This behaviour is usually referred to as "Confinement". However, the exact theoretical description of this effect is not complete yet. The strong coupling constant α_S is becoming too strong at small momentum transfers Q^2, which prevents the use of the usual perturbative approach.

"Asymptotic Freedom" is another property of QCD which allows partons to be treated as quasi free particles at sufficiently large Q^2. From this originates the success of the quark parton model describing the findings of collider experiments. For the discovery of

the Asymptotic Freedom as a fundamental property of the strong interaction, the Nobel Prize in Physics was awarded to Gross, Politzer and Wilczek in 2004. Thus, starting off as an idea to describe the experimentally motivated quark parton model, QCD became the standard theory for the strong interaction. Up to now it has been very successful in describing experimental findings.

2.3.1 The QCD Lagrangian

The full theory of fermionic quarks and bosonic gluons was formulated in 1973 by Fritzsch, Gross, Wilczek, and Weinberg [12–14]. The action is defined in terms of a Lagrangian density which for a single flavour of non-interacting quarks is given by

$$S = i \int d^4x \mathcal{L}(x) \qquad \text{with} \qquad \mathcal{L} = \bar{q}_j(x)(i\partial\!\!\!/ - m)q_j(x). \tag{2.7}$$

The index j on the Dirac four-spinors q_j runs over the $N_c = 3$ quark colours. The slash notation is used to describe $\partial\!\!\!/ = \gamma^\mu \partial_\mu = \gamma_\mu \partial^\mu$, implying the Einstein sum rule to sum over equal indices of the Dirac γ- matrices (see Appendix A.1). Additionally, $c = \hbar = 1$ is used and the charge is given in the Heaviside-Lorentz system. The equation of motion in Lagrangian field theory is given by the Euler-Lagrange equations for the arbitrary fields ψ_i:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi_i)} \right) = \frac{\partial \mathcal{L}}{\partial \psi_i} \, . \tag{2.8}$$

The basis of gauge theories is, that a given quantum state ψ is invariant under the symmetry transformations of the given group:

$$\psi \rightarrow U\psi \, . \tag{2.9}$$

Within the given SU($N_c = 3$) symmetry, U can be written as a function of $N_c^2 - 1 = 8$ real parameters θ_a

$$U(\theta) = e^{i\theta_a T^a} \tag{2.10}$$

where the index a runs over all generators $T^a = \frac{1}{2}\lambda^a$ of the Lie group, with the Gell-Mann matrices λ^a (see Appendix A.2). The transformation $U(\theta(x)) = U(x)$ describes a local symmetry transformation when it depends on the position x, and a global one if it does not. Equation (2.7) is not invariant under a local gauge transformation due to the derivative. In order to make the Lagrangian density invariant, $N_c^2 - 1$ real valued gauge fields A_a^μ are introduced and ∂^μ is replaced by the covariant derivative D^μ,

$$D^\mu = \partial^\mu + i g_s A^\mu \qquad \text{with} \qquad A^\mu = A_a^\mu T^a, \tag{2.11}$$

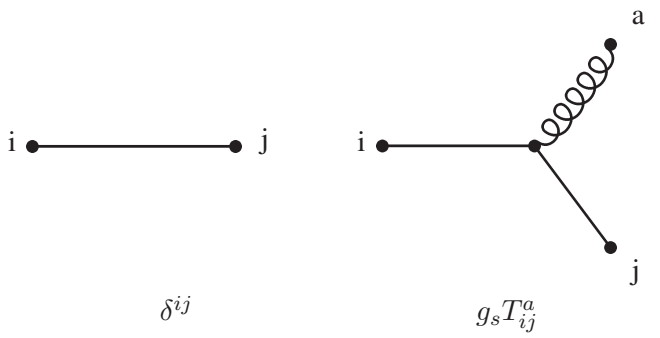

Figure 2.1: *The terms of $\mathcal{L}_{\text{quark}}$ (2.14) are shown with their coupling strength for the gluon colour a and quark colours i and j: a free quark field and the quark gluon coupling. The represented terms are summed over all quark flavours.*

introducing the gauge coupling g_s and the gluon fields A^μ_a. Local gauge invariance requires the transformation property

$$\begin{aligned} & & D^\mu(A') &= U(x)D^\mu(A)U(x)^{-1} \\ \text{equivalent to} & & D^\mu(A')q'(x) &= U(x)D^\mu(A)q(x) \end{aligned} \tag{2.12}$$

which is realised if the A^μ field transforms as:

$$A^\mu \to U(x)A^\mu U(x)^{-1} + \frac{i}{g_s}\left[\partial^\mu U(x)\right]U(x)^{-1} . \tag{2.13}$$

Due to the second, inhomogeneous term, non vanishing gauge field configurations can be produced from the vacuum ($A^\mu = 0$). The above considerations enable us to formulate the locally gauge invariant Lagrangian density for the quark fields:

$$\begin{aligned} \mathcal{L}_{\text{quark}} &= \bar{q}_j(x)[i\not{D} - m]_{jk}q_k(x) \\ &= \bar{q}_j(x)[(i\not{\partial} - m)\delta_{jk} - g_s\not{A}_a T^a_{jk}]q_k(x) . \end{aligned} \tag{2.14}$$

The Lagrangian (2.14) only describes the interaction of the quarks with the field as an external source. This includes a quark propagator and the coupling of a quark to a gluon, whose strength is proportional to $g_s T^a_{ij}$, depicted in Figure 2.1.

In order to describe the dynamics of the theory, derivative terms ($\partial^\nu A^\mu_a$) have to be introduced which sustain gauge and Lorentz invariance of the expression. The commutators of the covariant derivatives contain some combinations of derivative terms:

$$[D_\mu, D_\nu] \equiv ig_s F_{\mu\nu} \quad \Longrightarrow \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + ig_s[A_\mu, A_\nu] \tag{2.15}$$

$$\text{or, taking components} \quad F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g_s f^{abc} A^b_\mu A^c_\nu . \tag{2.16}$$

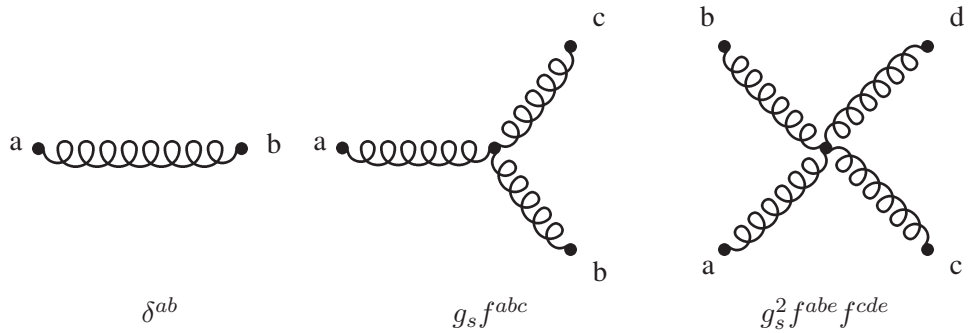

Figure 2.2: *The terms corresponding to $\mathcal{L}_{\text{gauge}}$ (2.17) are shown with their coupling strength for gluon colours a,b,c,d,e = 1,...,8: gluon propagator, three gluon vertex, and the four gluon vertex.*

This defines the gauge field strength (Lorentz) tensor $F_{\mu\nu}$. Which has non-trivial gauge transformation properties as a tensor under SU(N_c), as given in 2.12. However, a suitable Lorentz and gauge invariant term can now be constructed to add to the Lagrangian density,

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2} Tr\{F_{\mu\nu} F^{\mu\nu}\} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu}. \tag{2.17}$$

This term describes the free gluon field and the gluon self couplings including a three gluon vertex, which couples proportionally to $g_s f^{abc}$ and a four gluon vertex coupling proportionally to $g_s^2 f^{abc} f^{cde}$, as depicted in Figure 2.2. Additional terms are possible which still sustain gauge and Lorentz invariance. Following the same procedure like in QED, there is a certain additional freedom in choosing the quantisation of the gluon field due to the gluon self coupling, which leads to the so called "ghost-terms". The additional freedom manifests itself in additional ghost fields η^a and an arbitrary parameter ξ. The full Lagrangian in covariant gauge ($\partial_\mu A^{a\mu} = 0$) reads:

$$\begin{aligned}
\mathcal{L}_{\text{QCD_cov}} \;=\; & \bar{q}_j(x)[i\partial\!\!\!/ - m]q_j(x) \\
& -\frac{1}{2}\left[(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu}) + \frac{1}{\xi}(\partial^\mu A^a_\mu)(\partial^\nu A^a_\nu)\right] \\
& +(\partial^\mu \eta^{a\dagger})(\partial_\mu \eta^a) - g_s T^a_{jk} \bar{q}_j(x) A\!\!\!/^a q_k(x) + g_s f_{abc}(\partial_\mu \eta^{a\dagger})\eta^b A^{c\mu} \\
& +g_s f_{abc}(\partial_\mu A^a_\nu) A^{b\mu} A^{c\nu} - \frac{g_s^2}{4} f_{abc} f_{ade} A^{b\mu} A^{c\nu} A^d_\mu A^e_\nu \,.
\end{aligned} \tag{2.18}$$

A special choice of a gauge $n \cdot A^a = 0$ ("axial" or "physical" gauge) with a fixed four-vector n, removes all couplings of ghosts with all other fields, thus they can be ignored in the calculation. One finally retains a simpler form of the Lagrangian density of QCD as a

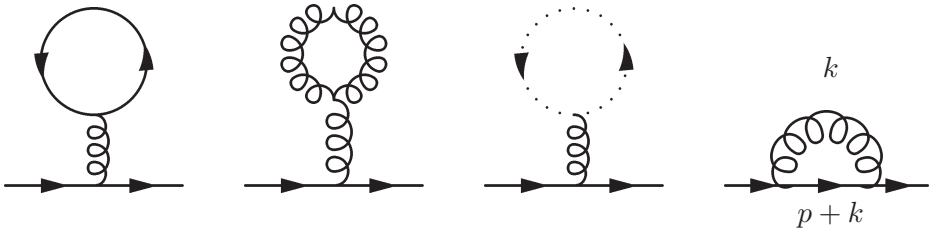

Figure 2.3: *Contributions to the quark self-energy at leading non-trivial order, $\mathcal{O}(\alpha_S)$, due to the three tadpole-type diagrams and the quark-gluon loop diagram are presented. However, the first three do not contribute due to the vanishing colour factors.*

sum of the terms (2.14) and (2.17):

$$\mathcal{L}_{\text{QCD_phys}} = -\frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu} + \bar{q}_j(x)[(i\partial\!\!\!/ - m)\delta_{jk} - g_s A\!\!\!/_a T^a_{jk}]q_k(x)\ . \tag{2.19}$$

2.3.2 The running coupling $\alpha_S(Q^2)$ of the Strong Interaction

Interactions are described by coupling constants, which can be chosen to be dimensionless to enable a comparison of the different couplings. In analogy to the electromagnetic coupling represented by the fine-structure constant α, which describes the coupling of photons to the electric charge e, the strong coupling constant α_S was introduced for the strong interaction. It describes couplings between gluons and the colour "charge" $g_s = \sqrt{4\pi\alpha_S}$ introduced in Section 2.3.1. A brief overview of the peculiarities of the coloured coupling will be given in the following.

Applying calculations of the standard perturbative QCD means operating with "naked" charges g (or masses m) of point particles, where one has to deal with ultraviolet divergences occurring in the integrations of loop momenta $d^4k/(2\pi)^4$. These occur for example as gluon or quark "self-energies" or vertex corrections. An example for one loop self-energy contributions to the quark propagator is given in Figure 2.3, additional diagrams contribute to the gluon self-energies.

To finally derive finite results from such calculations a renormalisation procedure needs to be introduced. For each ultraviolet divergence, an additional counter term is added to the Lagrangian which fixes the divergences at the cost of additional, couplings and interactions. One can absorb the new terms by renormalising the fields, masses and couplings. Unfortunately this spoils the gauge symmetry of the theory, which is an important precondition for the formal proof of the renormalisability of the theory. In an unrenormalisable theory, additional counter terms would be needed in every order of perturbation. However, gauge invariance can be restored by requiring every gauge coupling to be equal. This symmetry leads to the so called Slavnov-Taylor identities which finally allow the introduction of a single renormalisation factor and a unique gauge coupling g_{s0}. This way one

finally finds a finite Lagrangian, which in fact has exactly the same form as 2.19 written in terms of rescaled fields ψ_0, A_0 and parameters g_{s0}, m_0. In case of covariant gauge, the additional ghost fields and couplings also need to be rescaled. The renormalised strong coupling α_S as well as the quark masses become dependent of the momentum transfer Q^2, which means a "running coupling constant" is retained. The rescaled parameters are the ones measurable in experiments, as the "naked" ones cannot be resolved by definition. Additionally, the schemes introduce a dependence on an arbitrary unphysical unit mass μ_R which sets the scale for the problem. There is a certain freedom of choice concerning the counter terms; a specific choice of them is referred to as renormalisation scheme. The $\overline{MS}$ scheme [15] is mass independent and the most popular one, as it has some features which simplify the calculations. However, physical quantities may not depend on the arbitrary choice of the renormalisation scheme or the scale choice for μ_R. So a group structure exists, which encapsulates the transformations connecting quantities g_s, m, ψ, etc. in different schemes and scales. Assuming an amplitude Γ for an operator describing the 'scattering' of n_ψ (anti)quarks and n_A gluons, where the counter term is proportional to itself, so that the renormalisation is multiplicative (otherwise the representation would be more complicated):

$$\Gamma_0(\alpha_{s0}, m_0, \xi_0, Q) = Z_\psi^{-\frac{n_\psi}{2}} Z_A^{-\frac{n_A}{2}} \Gamma(\mu_R, \alpha_S, m, \xi, Q) \; . \tag{2.20}$$

With the renormalisation coefficients Z_ψ and Z_A, arising due to the renormalisation scheme. So like the physical quantity on the left side, also the renormalised quantity may not depend on μ_R:

$$\begin{aligned} 0 &\stackrel{!}{=} \mu_R \frac{d}{d\mu_R} \left[Z\Gamma(\mu_R, \alpha_S, m, Q) \right] \\ 0 &= \left\{ \mu_R \frac{\partial}{\partial \mu_R} + \beta \frac{\partial}{\partial \alpha_S} + m\gamma_m \frac{\partial}{\partial m} + \xi\delta_\xi \frac{\partial}{\partial \xi} - n_\psi \gamma_\psi - n_A \gamma_A \right\} \Gamma \; . \end{aligned} \tag{2.21}$$

The first term reflects the explicit μ_R dependency, while the remainder takes care of any implicit dependencies via $g_s(\mu_R)$ or $m(\mu_R)$. Equation 2.21 defines the functions which give the μ_R dependence in the respective renormalisation scheme, like

$$\beta\left(\alpha_S, \frac{m}{\mu_R}\right) = \mu_R \frac{\partial \alpha_S}{\partial \mu_R} \tag{2.22}$$

$$\gamma_m\left(\alpha_S, \frac{m}{\mu_R}\right) = \frac{\mu_R}{m} \frac{\partial m}{\partial \mu_R}, \tag{2.23}$$

and the other implicit dependencies of μ_R. So the β-function gives the scale dependence of α_S and γ_m the dependence of the quark mass. The evolution of these functions in terms

of Q^2 in a mass independent scheme reads:

$$Q^2 \frac{d\alpha_S(Q^2)}{dQ^2} = \beta\left(\alpha_S(Q^2)\right) = -\alpha_S^2(\beta_0 + \beta_1\alpha_S + \beta_2\alpha_S^2 + ...) \quad \text{and} \tag{2.24}$$

$$\frac{Q^2}{\overline{m}(Q^2)} \frac{d\overline{m}(Q^2)}{dQ^2} = \gamma_m\left(\alpha_S(Q^2)\right) = -\alpha_S^2(\gamma_0 + \gamma_1\alpha_S + \gamma_2\alpha_S^2 + ...)\,. \tag{2.25}$$

The coefficients can be calculated, e.g. in the $\overline{MS}$ scheme. However the coefficients β_0, β_1, and γ_0 are the same for any mass independent scheme:

$$\beta_0 = \frac{33 - 2n_f}{12\pi}, \tag{2.26}$$

$$\beta_1 = \frac{153 - 19n_f}{24\pi^2}, \tag{2.27}$$

$$\gamma_0 = \frac{1}{\pi}. \tag{2.28}$$

In this notation n_f is the number of quark flavours that can be produced with the available Q^2. So exploiting the above relations, if α_S is known at a given scale Q_0 it can be evaluated at any other scale. The one-loop solution reads

$$\alpha_S(Q^2) = \frac{\alpha_S(Q_0^2)}{1 + \beta_0\alpha_S(Q_0^2)\ln(\frac{Q^2}{Q_0^2})}\,. \tag{2.29}$$

This is a falling function for increasing Q^2 which is due to the non-abelian SU(3)-β function being negative (for $n_f < 17$) and enables perturbative calculations at large enough $Q \gg \lambda \approx 200\ \mathrm{MeV}$. In QED with abelian photons, the respective β-function is positive, which leads to growing electric charge with increasing Q^2. Historically, a dimensional parameter Λ has been defined to be the point, where the strong coupling would diverge and used this to parametrise α_S. However, there are some disadvantages of this choice: Λ is not dimensionless, it depends on n_f, and several different definitions for Λ exist in the literature. Moreover it depends on the renormalisation scheme, which has therefore always to be given for a certain Λ value. Therefore it is currently most common to evaluate α_S at $M_Z = 91.2\ \mathrm{GeV}$, the mass of the neutral Z-boson. The current (2009) value given by the particle data group [16, 17] is:

$$\alpha_S(M_Z^2) = 0.1176 \pm 0.002. \tag{2.30}$$

In Figure 2.4 several current measurements extrapolated to $Q = M_Z$ are shown on the left, the points measured at their respective scales are given on the right. These support the theory of the running coupling, with the expected feature of decreasing strength with increasing scale or decreasing distance ($\ll 1\ \mathrm{fm}$). The explanation is similar to electro-

magnetic shielding effects, although the shielding occurs in the case of the strong coupling at small distances, mainly due to gluon self-coupling. This leads to an asymptotic freedom, which allows to treat quarks as quasi-free particles at high energies. It enables the treatment of hadron collisions by a decoupling of the small distance hard interaction from the low energetic large distance one (proton radius $\approx 1fm$). The phenomenological quark parton model is theoretically backed up this way and the use of parton distribution functions in the context of a QCD improved parton model can thus be introduced. At large distances ($\sim$ 1 fm) or small Q^2 respectively, the strong coupling increases tremendously and therefore quarks are confined within the structure of the proton. In fact the values of α_S derived with perturbative QCD calculations diverge at Q^2 values of about 200 MeV. However, this Landau pole lies outside the region where perturbative QCD can be applied, because at scales below 1 GeV α_S can no longer be assumed to be small. Particles measured after hard interactions in a detector are always colourless hadrons, because as the hard partons reach distances $>$ 1 fm enough energy has been gathered up by the colour field to produce additional $q\bar{q}$-pairs or gluons, finally leading to a fragmentation into the observed colourless hadrons. Such observations are currently described by phenomenological non-perturbative models.

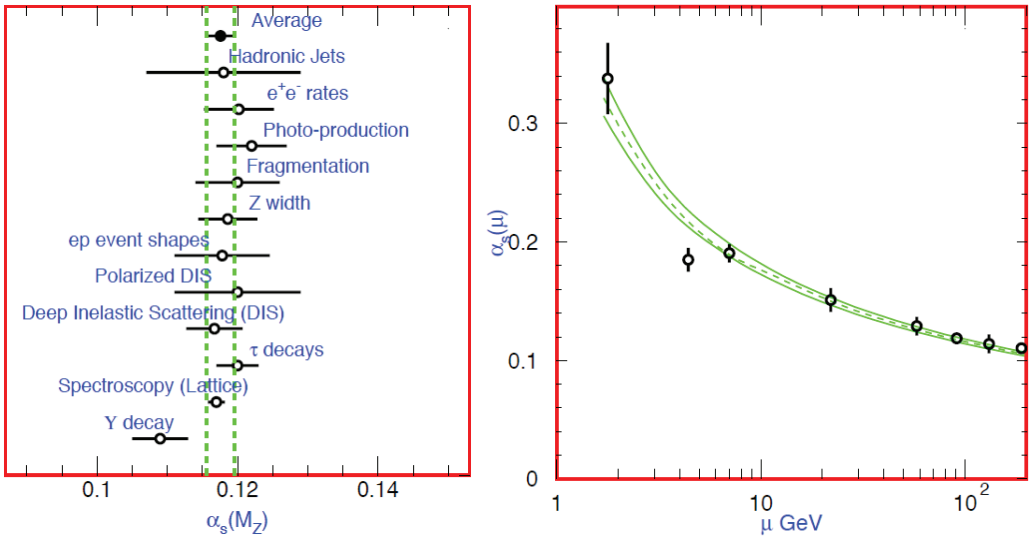

Figure 2.4: *A summary of the values of $\alpha_s(M_Z)$ from various processes extrapolated from the measurement scaled to $\mu = M_Z$ are shown on the left. This plot also contains the PDG average. The measured running of the strong coupling α_S is shown on the right. The lines show the central values and the $\pm 1\sigma$ limits of the PDG average. The data are in increasing order of μ, τ width, Υ decays, deep inelastic scattering, e^+e^- event shapes at 22 GeV from the JADE data, shapes at TRISTAN at 58 GeV, Z width, and e^+e^- event shapes at 135 and 189 GeV. Both plots are taken from PDG [17].*

2.3.3 Hadrons

All coloured objects are confined together with other coloured objects to finally build colourless objects. These objects are called hadrons. Up to now two types of hadrons are known: Baryons which are fermions made out of either three quarks or three anti-quarks, and mesons which consist of a quark and an anti-quark. The known properties of the hadrons can be constructed from their constituent quarks following conservation laws and are studied by observations of their decay products. Heavy baryons or mesons, especially the ones consisting of heavy quarks, are unstable and decay quickly. Lifetimes are about 10^{-17}s for electromagnetic decays and $< 10^{-20}$s for strong decays. Despite the mentioned constraints for the production of hadrons, up to now the properties of over 100 hadrons are registered within the particle listing of the particle data group. This is a result of having six different flavours for quarks as well as for anti-quarks, and regarding the fact that in particle physics also excited states of hadrons are treated as new particles, because their excitation energy reaches the order of magnitude of the self energy of the ground state.

The proton is one of the basic building blocks of all stable matter in the universe. Due to its charged nature, and easy availability, it is the ideal candidate for studying the properties of baryonic objects. Details about the proton structure are revealed mainly in deep inelastic scattering (DIS) of electrons and protons, described by the exchange of a virtual photon. Due to the Heisenberg uncertainty principle, the energy of the probing photon needs to be larger than $1\ \mathrm{GeV}$ in order to resolve the inner structure of the proton. In such experiments, the deviation from elastic scattering usually is described by the "Bjorken" x

$$x = \frac{Q^2}{2M_h\nu} = \frac{Q^2}{2Pq}, \tag{2.31}$$

with the momentum transfer $Q^2 = (P - q)^2$ between proton and photon, the mass of the proton M_h, and the Lorentz invariant energy transfer $\nu = \frac{Pq}{M_h}$. In fact from such experiments it was found, that the proton has a substructure of point-like partons. The insight could be derived from the scaling behaviour of the measured structure functions, which for DIS became independent of Q^2, some early results were published in [18]. Together with notions gathered from classifying observations made at hadron spectroscopy, this gave rise to the first quark parton model [19, 20] which was invented already before the formulation of QCD. As a quasi-classical model it describes the hadron as a collection of independent partons, off which a lepton can scatter via the exchange of a vector boson. The "Callan-Gross-relation" [21]

$$2xF_1(x) = F_2(x) \tag{2.32}$$

which only holds for fermionic particles, was found to be fulfilled between the parton and momentum density functions F_1 and F_2. This finally identified the quarks as fermions, and

lead to the constituent quark model, which added quantum numbers to the constituents and suggested a relationship between the measured structure functions [22]. At this point it is instructive to know, that at large Q^2, where parton masses can be neglected, the Bjorken x can be interpreted as the momentum fraction of the proton carried by the quark, which took part in the interaction. So in the constituent quark model, which neglects constituent masses, each of the constituents carries a part of the proton momentum, following a probability distribution $f_i(x_i)$, which gives the probability that a constituent i carries a momentum fraction x_i. This means, that the probability for x_i to fall into the infinitesimal range $[x, x + dx]$ is given by $f_i(x)dx$. These distributions are usually referred to as parton distribution functions (PDFs). It was found that the scaling behaviour of structure functions and of the PDFs is best described as a function of the Bjorken x. One of the initial problems of the constituent quark model was the fact, that only half of the protons momentum had been found to be carried by the quarks. It was only later, that the QCD improved parton model solved this problem in a consistent fashion, introducing gluons as additional constituents of the proton.

2.3.4 The QCD Improved Parton Model

Together with the improved constituent quark model, QCD provides a framework for the independent constituents observed in DIS by the asymptotic freedom, and with the gluon a candidate for the observed missing transverse momentum of the hadrons with respect to the measured quark momenta. It also gives a basis to describe the hard interaction of the constituents with other particles at sufficiently high Q^2. However, the quantum theoretical part of the hadron structure is missing. This was overcome with Altarelli introducing pQCD corrections [23], which introduced an essential feature of the current improved parton model: It could be shown, that the large distance interactions, which describe the hadron as a whole, are indeed universal for the type of hadron, and can be separated from the short range interaction of the specific parton. This separation actually makes the parton distribution functions (PDFs) a scale dependent function of the momentum transfer Q^2, and the Bjorken x. In this picture, the partons themselves consist of further "daughter" partons, appearing when increasing the probing Q^2. The scale dependence is governed by the DGLAP equations which were introduced by Dokshitzer, Gribov, Lipatov, Altarelli and Parisi [24–26]. Certainly it remains true, that the PDFs cannot be calculated from first principles until sufficient non-perturbative techniques for QCD will be available. However, measured parton distribution functions at one scale can be evaluated at another scale using the DGLAP equations [27]:

$$\begin{aligned}
\frac{\partial q_i(x,\mu^2)}{\partial \log \mu^2} &= \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\left[P_{q_iq_j}(z,\alpha_s)q_j\left(\frac{x}{z},\mu^2\right) + P_{q_ig}(z,\alpha_s)g\left(\frac{x}{z},\mu^2\right)\right], \\
\frac{\partial g(x,\mu^2)}{\partial \log \mu^2} &= \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\left[P_{gq_j}(z,\alpha_s)q_j\left(\frac{x}{z},\mu^2\right) + P_{gg}(z,\alpha_s)g\left(\frac{x}{z},\mu^2\right)\right], \qquad (2.33)
\end{aligned}$$

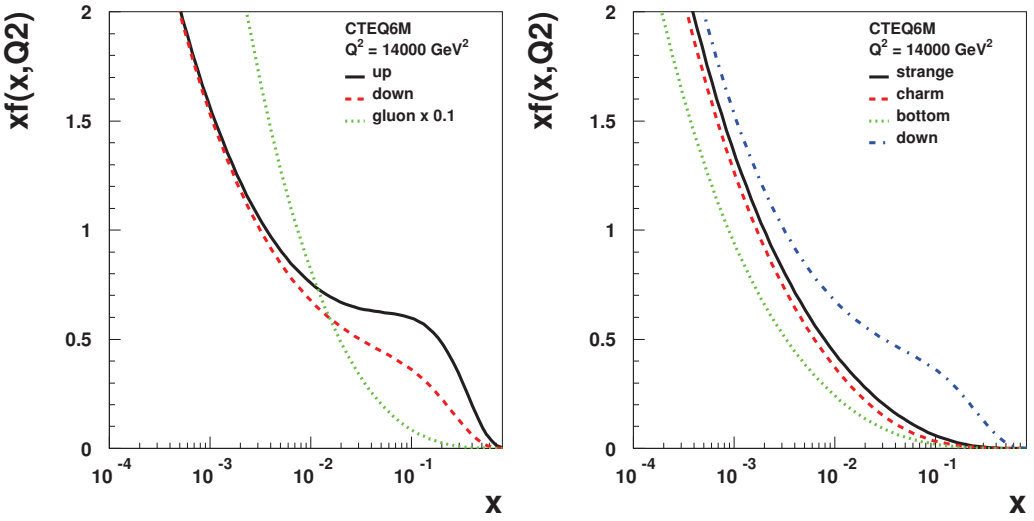

Figure 2.5: *The parton momentum distribution is shown for u-, d-quarks, and gluons (reduced by a factor of 10) on the left and d-,s-,c-,b-quarks on the right. This has been derived using the CTEQ6M PDFs and the online tool of the Durham university for a typical LHC high p_{T} jet scale of $Q \approx 118$ GeV.*

where the Altarelli-Parisi splitting functions $P_{ab}(z, \alpha_S(\mu^2))$, are associated with the branchings $b \to aX$ and can be expanded as power series in α_S:

$$P_{ab}(x, \alpha_S) = P_{ab}^{(0)}(x) + \frac{\alpha_S}{2\pi} P_{ab}^{(1)}(x) + \cdots . \tag{2.34}$$

Various methods exist to measure parton density functions with all kinds of experiments. Finally, the results are being combined to officially released functions, which is currently done by several independent groups, e.g. MSTW [28] or CTEQ [29]. These functions contain theoretical and experimental uncertainties, which are provided within error-functions. In order to estimate uncertainties of observables derived from the PDFs, these error-functions have to be evaluated. The distributions in Figure 2.5 are extracted from CTEQ6M and plotted by the online tool of the Durham university [30].

2.4 Hadron-Hadron scattering at the LHC

The design properties for the LHC foresee two bunches of protons colliding at a centre of mass energy of $\sqrt{s} = 14$ TeV. Most interactions will involve low Q^2 elastic scattering of protons. However these interactions do not have a large transverse momentum transfer, and are not the main focus for searches of new physics. The processes we are interested in

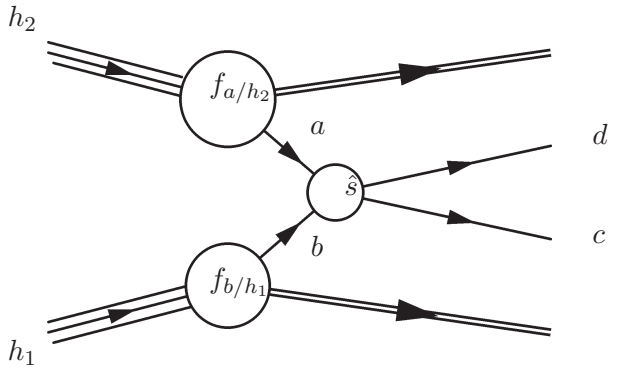

Figure 2.6: *Hadron-hadron collision is shown with the hard interaction between partons a and b.*

involve processes with large Q^2. In this region the strong interaction can be described by perturbative QCD. For $Q > 1\ \mathrm{GeV}/c$ the De Broglie wavelength

$$\lambda = \frac{h}{p}, \tag{2.35}$$

drops below the proton radius of one fm. With sufficiently large momentum transfer, the interactions can be described as incoherent elastic scattering interactions between the constituents of the protons (Fig. 2.6). So in fact the interacting parton a (b) carries a sufficiently large momentum fraction to probe the inner structures of the other proton h_b (h_a). In deep inelastic scattering experiments with electrons probing the proton structure, it has been found, that the total hadron momentum is split up into fractions x_i between the constituents i following parton density functions $f_i(x_i)$. These functions give the probability $f_i(x)dx$, that the momentum of parton i falls into the infinitesimal range $[x, x + dx]$. Following the QCD improved parton model (Sec. 2.3.4), PDFs allow a decoupled handling of the long ranged hadron and short ranged parton interactions, which in fact makes the PDFs universal, thus allows PDFs determined from $e^- p$ interactions to be used in pp interactions. So the cross section of pp into two particles c and d is given by:

$$\begin{aligned} d\sigma^{(h_1 h_2 \to cd)} &= \int_0^1 dx_1 dx_2 \sum_{a,b} f_{a/h_1}(x_1, \mu_F^2) f_{b/h_2}(x_2, \mu_F^2) \\ &\times \ d\hat{\sigma}^{(ab \to cd)}(Q^2, \mu_F, \mu_R), \end{aligned} \tag{2.36}$$

where the indices a,b refer to the partons (q,$\overline{q}$,g) of the hadrons h_1 and h_2. In the hard process the parton momenta are given by $p_a^\mu = x_1 p_{h_1}^\mu$ and $p_b^\mu = x_2 p_{h_2}^\mu$. In most cases $x_1 \neq x_2$, so the hard interaction is boosted with $\beta = (x_1 - x_2)/(x_1 + x_2)$ with respect to the laboratory frame of $h_1 h_2$. However, measuring in transverse momenta and Lorentz in-

variant rapidity[1] y keeps arising effects small. The sum is over all partonic processes which produce particles c and d. So finally, the hadronic cross section is made out of parameters and parametrised functions, which need to be determined by experiments and renormalisation and factorisation scales which introduce unphysical scales μ_R and μ_F, from which the final result should be independent. These scales come together with schemes based on perturbative QCD which introduce running couplings and quark masses and which need to be used consistently for the PDFs as well as for the calculation of the hard matrix elements $ab \to cd$, because otherwise the cancellation of ultraviolet and collinear divergences would not work. It is common to use $Q^2 = \mu_R^2 = \mu_F^2$, in the case of jet cross sections, it is often set to the squared transverse momentum p_{T}^2 of the jet.

2.4.1 Jet Cross Sections

The collimated stream of particles observed after a hard interaction in the direction of a struck quark is commonly referred to as a jet. In case of calculations in pQCD, in leading order one would not need to apply any specific jet algorithm, as the final state simply consists of two partons, which are produced back-to-back, thus each of them represents a jet. This assumes that the resolution parameter of the jet algorithm is chosen reasonably. Already at NLO, the additional possible radiation introduces a more complex event structure, which could be two jets, in case the radiation is soft or collinear, or three jets, in which the radiation forms a separated jet. Thus it is not straight forward to unambiguously classify an event as a two or as a three jet event. As low energetic or collinear emissions happen randomly, one also does not want the final event description to be dependent on such features. Therefore it is required to define an observable which is robust against such radiations. This is fulfilled by collinear and infrared safe jet algorithms (see Sec. 2.6), which cluster emitted particles together which are close in phase space according to the respective algorithmical definition. These algorithms have to be included into the phase-space integration in equation 2.36. The advantage is, that this procedure is also applicable to measurements, where detectors measure the energy deposits of the hadronised particle streams. Usual QCD dijet events at the LHC contain ≈ 1000 particles or even more energy deposits of the detectors calorimeter systems, which need to be clustered into jets. These clusterings should ideally produce the same results as clustering partons from direct pQCD calculations. In real measurements some additional corrections need to be applied in order to correct for non-perturbative effects, which cannot be described in pQCD (see Section 4.2). More details on this, the theoretical explanations of the building of such final states, and the technicalities of jet algorithms are given in the following sections.

To illustrate the influence of the PDFs at the perturbative level, Figure 2.7 shows the decomposition of the total jet cross section into the seven possible partonic sub processes for $p\bar{p}$ collisions at the Tevatron on the left hand side and for pp collisions at the LHC on

[1] For a fourvector with energy E and longitudinal momentum component p_z parallel to the hadron direction, rapidity is usually defined as $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$.

the right hand side at central rapidities. The latter have been derived with next-to-leading order precision using the fastNLO [31] package, more details on the calculation and the results for $\sqrt{s} = 10$ TeV are given in sections 4.1 and 5.10. The fractional contributions are drawn versus the scaling variable $x_T = 2p_{\mathrm{T}}/\sqrt{s}$. As expected, large percentages from gluon induced processes can be seen in the low $\mathrm{x_T}$ region. As expected, the higher the transverse momentum the more valence quark contributions dominate, that is $q\bar{q}$ in case of the Tevatron and qq partonic reactions in case of the LHC. One has to keep in mind though that due to sum rules the PDFs are coupled and cannot be considered separately.

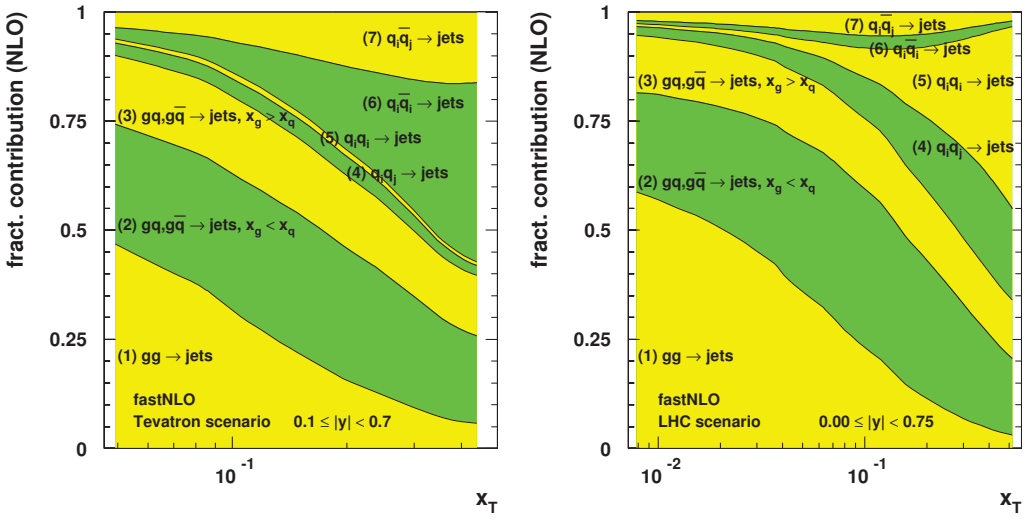

Figure 2.7: *Decomposition of the total jet cross section into the partonic processes for $p\bar{p}$ collisions at the Tevatron ($\sqrt{s} = 1.96$ TeV, left) and pp collisions at the LHC ($\sqrt{s} = 14$ TeV, right), respectively. The fractional contributions are shown comulatively versus the scaling variable $x_T = 2p_{\mathrm{T}}/\sqrt{s}$ for the k_{T} jet algorithm.*

2.5 From Quarks to Hadrons

Desirably, all processes from the hard interaction to the final hadrons, electrons and photons, which can be observed within the detector, should be calculable from first principles. Unfortunately, there are several complications due to properties of the strong interaction which prevent this. We have already seen that only in energy ranges > 4 GeV perturbative methods can be used. However, those methods are very computing intensive when trying to go to higher orders in perturbation. Most calculations are therefore executed with leading order (LO) or next-to-leading order (NLO) precision. Few processes could be studied using next-to-next-to-leading order (NNLO) calculations. These calculations only represent the processes of the actual hard interaction. Experimentalists need to have a theoretical predic-

tion which can be applied to the observed particles including hadrons, leptons and photons. Thus, additional theoretical models are required in order to emulate the missing orders and to describe the low Q^2 behaviour. These models introduce additional parameters, which need to be derived from measurements by tuning the modelling of multiple observables to the respective measurements. Such additional predictions are essential to establish a detailed model of the detector as well as allowing to derive additional corrections in order to compare pQCD calculations to the measurements.

The products of the hard process radiate off new particles in a collinear or soft way producing jets, which directly follows from the internal structure of QCD, as was first shown by Sterman and Weinberg [3] in order to describe the observations at experiments at sufficiently high energies [2].

As long as the energy of the particles is large enough, radiations can be modelled using the collinear parton shower approximation. With particle energies evolving below the limit where perturbative methods can be used, an additional model has to be applied to describe the final production of hadronic particles. At this point phenomenological hadronisation models are used which have been tuned to best describe experimental data. This is shown in Figure 2.8, where a schematic view of a common event generation process with attached detector simulation is given, as it is used for example within the CMS experiment to model all aspects of the collisions and the resulting processes. The proton remnants are treated by additional dedicated models which describe e.g. multi-parton interactions (MPI [32]). The understanding of the physics described by all these models is of major interest to all studies at the LHC, because the initial process for all produced particles is a strong interaction. Therefore, the determination of the actual centre of mass energy $\sqrt{\hat{s}}$ of the processes strongly depends on how well these interactions are understood. This is the major difference to e^+e^--colliders like LEP, where the initial state of the interactions was well defined due to the point-like nature of the electrons and anti-electrons. Furthermore, such high energetic multi-jet-events are the main background for most physics studies comprising Higgs-searches as well as the search for new physics beyond the Standard Model. A more detailed view on parton shower as well as hadronisation models will be given in the following.

2.5.1 Parton Shower

In the 1970s and 1980s the parton-shower approach for the description of jet radiation has been developed. Motivated by the fact that the phase space for collinear and infrared radiation is enhanced, this method uses a collinear approximation. Starting with the initial 2-2 process every outgoing parton is assigned a mass of the order of Q^2, which can then be perturbatively related to states at the lowest scale Λ_{QCD} via the DGLAP equations (eqn. 2.33). Thus, the decay of a hard parton is then calculated using the Altarelli-Parisi splitting kernel. This kernel can be transformed into a Sudakov form factor (eqn. 2.37), which gives a probability of evolving from a higher scale to a lower scale without the emission

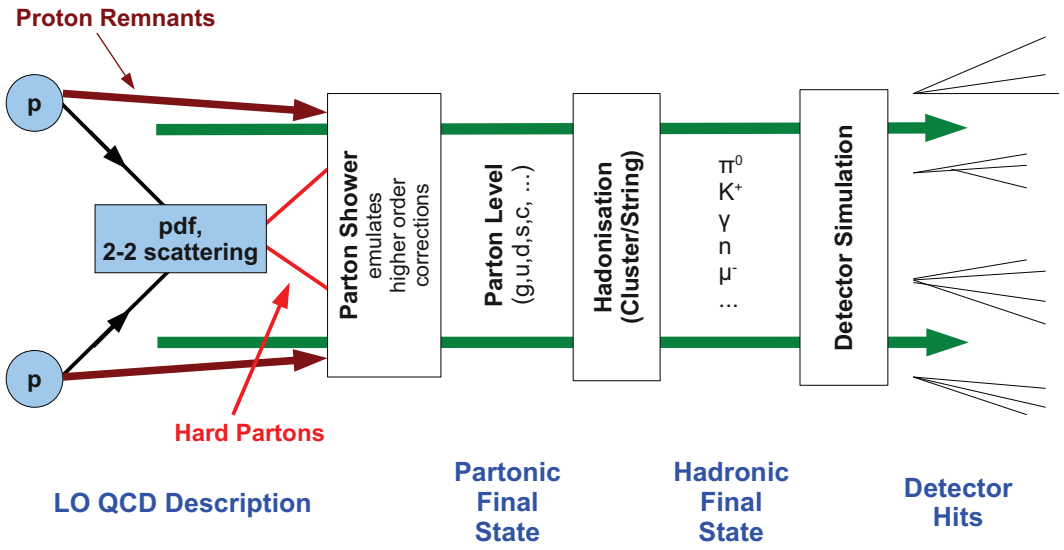

Figure 2.8: *This Figure shows the typical event generation process for a QCD event in pp collisions using a multi purpose LO event generator, which is used as input for the experiment specific detector simulation.*

of a gluon greater than a given value. Parton showering introduces two new scales. One for initial-state parton showering, where the evolution proceeds backwards from the hard scale of the process to the cutoff scale, with the Sudakov form factors being weighted by the parton distribution functions at the relevant scales [33]:

$$\Delta(t) \equiv \exp\left[-\int_{t_0}^{t} \frac{dt'}{t'} \int \frac{dz}{z} \frac{\alpha_S}{2\pi} P(z) \frac{f(\frac{x}{z}, t)}{f(x, t)}\right] . \tag{2.37}$$

Here t is the hard scale, t_0 is the cutoff scale, and $P(z)$ is the splitting function for the branching under consideration. The other scale is given for the shower in the final state, where the Sudakov form factor has a similar form, but without the PDF weighting. The scales themselves are set depending on the specific implementation.

Based on these probability functions Monte-Carlo methods can be used to consecutively simulate the chain of sequential radiations. This chain develops in the direction of the momentum of the initial hard parton, energy-momentum conservation is adjusted at the end. In the process successive values of an evolution variable t, a momentum fraction z, and an azimuthal angle ϕ are generated, along with the flavours of the partons emitted during the showering. The choice of the evolution variable t depends in the specific implementation of the model, as it is used within PYTHIA [34], HERWIG [35], or HERWIG++ [36]. Addition-

ally, different methods are implemented to describe the angular ordering of the emission. Angular ordering represents an attempt to simulate more precisely those higher order contributions that are enhanced due to soft gluon emission (colour coherence). Fixed-order calculations explicitly account for colour coherence, while parton shower Monte Carlos including colour-flow information model it only approximately.

The method is also called "leading-log-approximation" because only the leading terms of the perturbative expansion remain due to the cut-off parameter t_0. With increasing angle the emitted energy decreases in a logarithmic way. Narrow jets have most gluons emitted in the forward direction, therefore this cut solely results in a small energy loss. However wide angle emissions are not correctly described by this method, thus additional models are implemented in the generators to fill the respective phase space. However, it is not possible to directly couple higher order matrix elements to these parton shower models.

Several new approaches have been developed lately [37] to allow the use of parton shower models for an appropriate description of higher jet multiplicities. One problem which occurs when attaching parton shower models to multijet matrix-elements (tree level) or NLO loop calculations is the problem of double counting. Certain contributions are calculated twice, once in the higher orders of perturbation or included in the extended matrix element, and additionally they are reproduced within the emulation of the higher orders of the parton shower model. Two different methods are used to tackle this problem. The CKKW (Catani, Krauss, Kuhn, Webber [38, 39]) method, currently implemented in SHERPA [40], uses a k_T algorithm (see Section 2.6) on the higher multiplicity matrix level calculations in order to eliminate infrared divergences, and apply an additional reweighting of the decay histories. A similar approach is the MLM (M.L. Mangano) prescription implemented in ALPGEN [41, 42] which uses a cone based algorithm to identify the double-counted contributions. MADGRAPH [43, 44] uses a hybrid approach between the implementations in ALPGEN and SHERPA and implements the phase space matching via k_{T} algorithm.

MC@NLO [45] uses NLO loop calculations and attaches a parton shower from which finally double counted contributions are being subtracted. However, not all processes needed for an inclusive jet cross section calculation are implemented yet.

2.5.2 Hadronisation

After interaction energies during the decay chain fall below the limit $t_0 \approx 1 - 4\,\mathrm{GeV}$, parton shower calculations are no longer applicable, because α_S becomes too large. At this stage hadronisation or fragmentation models have to be attached to finally produce "final state" hadrons, thus cancelling the dependence on t_0, which was chosen arbitrarily. The overall assumption is, that the hadronisation process, operating at low momentum transfers, should not change the large scale topology of the event, which is determined by effects which can be described perturbatively. However, the non-perturbative effects cannot be neglected, especially when precision measurements should be performed. To get

a raw estimate on experimental uncertainties one has to consider several of them. Because the current implementations are tuned to describe the data of current experiments, LHC data can be expected to be somewhere between the extrapolation of those models to the new energy scale.

- **Independent Fragmentation**
  In the 1970s the independent fragmentation model was developed, mainly by Field and Feynman [46]. In this model gluons are first replaced by a light $q\bar{q}$-pair, thus it finally works on $q\bar{q}$-pairs only. Then the idea is to iterate a sequence of universal branchings, $|q_1> \rightarrow |q_1> + h(|q_1, \bar{q}_2>)$, by creating a $q_2\bar{q}_2$ from the vacuum. The produced hadron carries away a momentum fraction z, and is then considered either a final state particle or treated as unstable resonance and further decayed into final state mesons. This procedure is run with every quark until the leftover quark is below some cut-off energy. So with very few parameters, using a Gaussian distribution for the momentum left over to the remaining quark in the cascade, the model gives good agreement with data from e^+e^- experiments at moderate energies. Unfortunately, the model has no strong theoretical underpinning, so that the details are rather arbitrary. Quite some desired physical properties need to be adjusted afterwards, like momentum conservation, or the broken quantum number conservation due to the leftover quark. The flavour decomposition created in the branchings is also empirically adjusted to reflect the observed ones. Especially due to its large cutoff parameter $Q_0 \sim 3\ \mathrm{GeV}$ and the lack of infrared and collinear safeness, it is no longer widely used.

- **Cluster Fragmentation**
  The currently best known cluster fragmentation model is implemented in the simulation program HERWIG and HERWIG++ [47–49]. It is based on the preconfinement property of the angular-ordered parton shower. Thus it takes the remnants of the preceding parton shower, builds colour singlet clusters of $q\bar{q}$-pairs and decays them into hadrons following flavour conservation. Gluons which remain at the end of the parton shower process are split isotropically into $q\bar{q}$-pairs of light quarks. Together with the remaining quarks, they are bound together to colour singlet states with neighbouring $q\bar{q}'$ pairs. The arising singlets are thus formed into clusters, with the momentum given by the sum of the constituent partons. Preconfinement implies that pairs of colour-connected neighbouring particles have the tendency to be arranged within limited extension in both coordinate and momentum space, thus the formed clusters have a mass distribution that falls rapidly with growing masses and is asymptotically independent of the overall energy scale Q^2 and the hard interaction. A typical mass of a few GeV suggests associating clusters with 'super-resonances', that decay independent of one another and according to the available phase space into hadron resonances. This is modelled via simplified dynamics ignoring spin correlations which leads mainly to an isotropic two body decay of a cluster into two

hadrons. During the decay, the constituent flavours of the cluster are conserved and an additional $q\bar{q}$ pair is popped from the vacuum to fill the empty spaces in the new hadrons

$$\mathrm{Cl}(q_1\bar{q}_2) \rightarrow |h_1> + |h_2> = |q_1\bar{x}> + |x\bar{q}_2> \ . \tag{2.38}$$

The new quarks are created to grant a correct colour flow; spin, momenta, and masses are adjusted in order to fulfil the respective conservation laws. If the clusters are too light for a two body decay, they are decayed into a single hadron with the excess momenta being redistributed among neighbouring clusters. If the mass M of the cluster is too large, an isotropic decay is no longer reasonable, therefore an anisotropic fission mode is used to split the clusters into two lighter ones popping a $q\bar{q}$ pair from the vacuum, thus it is reminiscent of the string fragmentation model, to be described later. The directions of motion of the new clusters are aligned along the original $q_1 - \bar{q}_2$ axis. This is repeated recursively until the clusters are allowed to decay into final hadrons via the isotropic method. The cluster fission is applied in HERWIG++, if the cluster mass M fulfils

$$M^{\mathrm{Cl_{pow}}} \geq \mathrm{Cl}_{\mathrm{max}}^{\mathrm{Cl_{pow}}} + (m_1 + m_2)^{\mathrm{Cl_{pow}}}, \tag{2.39}$$

with the arbitrary cutoff mass $\mathrm{Cl_{max}}$, which is set according to the quark flavours of the cluster to values of $\sim 3 - 4\ \mathrm{GeV}$. Also the exponent $\mathrm{Cl_{pow}}$ is chosen according to the quark flavours with values $\sim 1 - 1.5$. Only up, down and strange quarks are inserted into the split quarks following additional parameters, steering the probabilities. The cluster content of heavy quark flavours like c or b quarks is fixed by an additional model, to reflect multiplicities observed in experiments. The flavours of the final particles are assigned from probability functions derived from experimentally determined branching ratios as they are provided by the PDG.

- **Lund/String Fragmentation**
  The Lund/String fragmentation model initially developed at the University of Lund [50] is one of the most popular string hadronisation models especially because it is implemented in the multi purpose generator PYTHIA. It is motivated by the properties of the colour field between two coloured particles, which loose energy to the colour field between them, as they move apart. This field is supposed to collapse due to self-interactions into a string-like configuration with a uniform energy per unit length, or string tension $\kappa \approx 1\ \mathrm{GeV/\ fm} \approx 0.2\ \mathrm{GeV}^2$, corresponding to a linear confining potential like the QCD confinement for quarks. The transverse size of a string, $\langle r_\perp^2 \rangle$, is therefore negligible, allowing the dynamics to be described by a massless, one-dimensional, relativistic string possessing no transverse excitations. The equations of motion for the string then lead to the so called yo-yo mode, where a massless quark-antiquark pair forming the endpoints of the string, seen from the

strings centre of mass frame, oscillate repeatedly outwards and inwards at the speed of light, passing through each other and transferring energy to and from the string [51]. Colourless string segments are formed between neighbouring partons of the parton shower output, each segment terminating on a quark and an antiquark. As additional solution to the string model, so called "kinks" show up (fig. 2.9), which represent gluons. This includes soft and hard radiation of gluons – the stronger the kink, the harder is the radiated gluon. The hadronisation is now described by an iterative spontaneous $q\bar{q}$ production in the intense colour field, which breaks the string more and more apart, producing hadrons following certain flavour and spin selection rules.

The pair production invokes quantum mechanical tunnelling. If the q and $\bar{q}$ have no common mass or transverse momentum, they can classically be created at one point and then pulled apart by the field. If the quarks have mass and/or transverse momentum, the q and $\bar{q}$ must classically be produced at a certain distance, so that the field energy between them can be transformed into the sum of the two transverse masses $\mathrm{m}_\perp$. Quantum mechanically, the quarks may be created in one point (in order to sustain local flavour conservation) and then tunnel out to the classically allowed region. In terms of a common transverse mass $m_\perp$ of the q and the $\bar{q}$, the tunnelling probability is given by

$$\exp\left(-\frac{\pi m_\perp^2}{\kappa}\right) = \exp\left(-\frac{\pi m^2}{\kappa}\right)\exp\left(-\frac{\pi p_\perp^2}{\kappa}\right). \tag{2.40}$$

Since the string is assumed to have no transverse excitations, this $p_\perp$ is locally compensated between the quark and the antiquark of the pair. The formula also implies a suppression of heavy quark production $u : d : s : c \approx 1 : 1 : 0.3 : 10^{-11}$, charm and heavier quarks are hence not expected to be produced in the soft fragmentation. This Gaussian model is somehow oversimplifying the process, thus the "predicted" value of $\langle p_\perp \rangle = \sqrt{\kappa/\pi} = 0.25$ GeV has to be adjusted to ≈ 0.4 GeV in order to describe the LEP observations. So due to the tunnelling model, the transverse mass of the produced hadron $m_\perp$ is already fixed, thus only the longitudinal momentum can be chosen freely. However, further constraints come into this by requiring longitudinal boost invariance and that the process should be independent of the start and the endpoint of the splitting, the Lund symmetric fragmentation function (LSFF) needs to be chosen to determine probability that the fraction z of the momentum taken by the split off hadron:

$$\begin{aligned} f(z) &= \frac{N_{\alpha\beta}}{z} z^{a_\alpha}\left(\frac{1-z}{z}\right)^{a_\beta}\exp\left(-b\frac{m_\perp^2}{z}\right) \\ &\overset{\alpha=\beta}{=} \frac{N}{z}(1-z)^a \exp\left(-b\frac{m_\perp^2}{z}\right). \end{aligned} \tag{2.41}$$

Here α labels the parent quark flavour and β the daughter quark flavour, which are commonly chosen to be the same. Thus one parameter a remains for each quark flavour. In principle only mesons could be produced by the model, as described above, but also baryon production is described, as the q or $\bar{q}$ can also be a diquark system. After having chosen the quark flavour via the tunnelling model, and the amount of the longitudinal momentum is assured by the LSFF, a choice for the properties of the formed hadrons has to be made. Hadrons are formed based on an attempt to model, via flavour and spin selection rules, the supposed dynamics of the process. Therefore quite a lot of parameters need adjustment, so that the produced state is compatible with observations. This for example comprises the imprecisely determined masses m of the (di)quarks to be used in 2.40 for different flavours ($m_u, m_s, m_{uu}, m_{us}, m_{ss}$). Additional features are build into the baryon production. As mentioned above, the machinery is capable of handling diquarks, producing $qq'\bar{q}\bar{q}'$ pairs instead of $q\bar{q}$, tunnelling from the vacuum. In the diquark model, the baryon and anti-baryon are always produced as nearest neighbours along the string. Furthermore, a popcorn model exists which in the simple case creates one additional meson between the baryons produced by diquarks, and multiple mesons in the advanced case (Fig. 2.10). If the energy of the string is not large enough to utilise the above mentioned methods, they are treated as clusters, comparable to the cluster model described previously.

2.6 Jet Algorithms

The confinement does not allow coloured objects like quarks or gluons to be separated from each other by distances larger than about one fermi. The energy invested in the colour field between these partons is used up to create quark anti-quark pairs from the vacuum until all these partons have been integrated into colourless hadrons. As a consequence, collimated streams of hadrons, which are customarily named jets, are observed in a detector while at the origin a small number of partons had been created in the hard collision. In order to reestablish the link between the observed particles and the hard process, algorithms are defined to group particles that are supposed to come from the same hard parton into a jet. As the infrared and collinear emissions happen on a random basis, their occurence should not change the final observable. Thus the clustering procedure of the jet algorithm must be robust against such random influences and unambigously project out the initial parton direction. Collinear and infrared sensitive situations are depicted in sections 2.6.1 and 2.6.2. To assure that the algorithms can be run on different levels of calculation, generation or reconstruction, the objects used as input to the algorithms must have comparable properties. A four-vector-like character, comprising a three dimensional direction and an energy component, is most suitable for this purpose. Such quantities can be derived either from calculated particle properties or from calorimeter measurements, using location and

amount of the energy deposit. Four-momenta of calorimeter objects are usually constructed to be mass-less, however, they are usually combined from different calorimeter cells into so called “towers” before passed on to the jet algorithm. Depending on the recombination schemes used during tower building or during the jet clustering, jet input or the final state jets will be massive four-vectors. The required ingredients of a jet algorithm to combine input objects are:

- a distance measure to define the separation between objects,
- a procedure to decide whether objects are to be combined,
- a recombination scheme explaining how to combine objects.

Although several recombination schemes exist, straight forward four-vector addition is used uniformly within the CMS collaboration and thus also throughout this work. The history of the two main categories of jet algorithms, namely the subsequent clustering algorithms and the cone type algorithms is presented in 2.6.3.

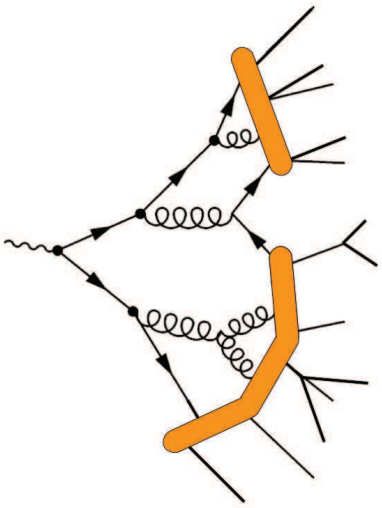

Figure 2.9: *Schematic representation of the string model. Quarks form endpoints of strings, whereas gluons produce kinks in them.*

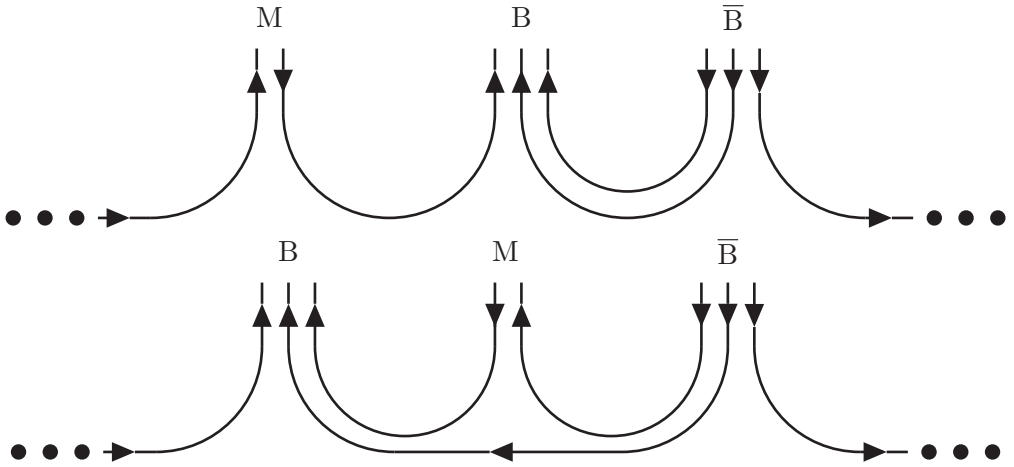

Figure 2.10: *Baryon production in the Lund model is shown via the diquark model (top) and simple popcorn model (bottom) leading to* $\mathrm{MB\overline{B}}$ *and* $\mathrm{BM\overline{B}}$ *(*B*= baryon,* M*=meson).*

2.6.1 Collinear Safety

Jet evolution is usually happening in a small cone, which was the reason why the collinear approximation could be used to calculate high energetic parton showers (Sec. 2.5.1). This evolution contains for instance gluon radiation collinear to a high energetic parton. For the jet definition, which has to be able to connect the Hadronic Final State to the initial partonic process, such additional collinear radiation may not lead to a different interpretation. Collinear safety means that all effects due to collinear splitting during calculation, hadronisation, or during measurements within the detector may not change the output of the algorithm. So the two situations pictured in Figure 2.11 should always produce a single jet. Algorithms which produce zero or two jets, in case of the left picture, are not collinear safe. This depicts the problem of thresholds in seeded algorithms: Although both seeds together might be large enough to be considered a seed, thus also producing a jet, no jet might be produced, as both are below the threshold. Another unwanted feature is, that collinear radiation leads to ambiguities, whether one or two jets will be created. The presented situation can happen due to the splitting of the jet energy into two neighbouring parts of the calorimeter system. In case of calculations in pQCD starting from NLO, collinear unsafety leads to singularities which spoil the convergence of the MC integration.

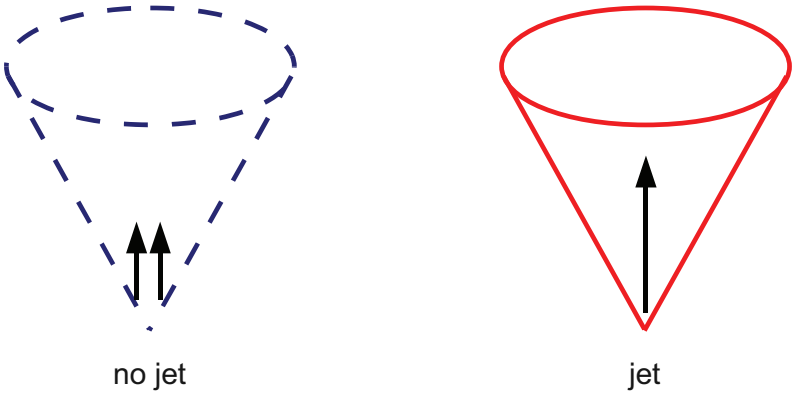

Figure 2.11: *An example for collinear unsafe behaviour of jet algorithms is shown: In both cases it is assumed, that the hard interaction led to the production of 1 jet. The right picture shows the situation if the algorithm constructed one jet. In the left picture the contributions of one jet are splitted up into two equal parts. This can happen due to collinear radiation effects or simply be a result of the granularity of the calorimeter, when a jet exactly hits the middle of two calorimeter towers. In this worst case scenario no jet at all would be produced by a seeded algorithm, as both small jets would fall below the threshold.*

2.6.2 Infrared Safety

Soft gluon radiation can also happen during the part of the processes described by parton showers or hadronisation. These gluons only have a small amount of energy, thus they should not change the output of the jet algorithm. This feature is intrinsically fulfilled within the k_T algorithms because the resolution parameter already manages that lower energetic parts are easily merged with higher energetic contributions, while the cone type algorithms are not intrinsically safe of such a behaviour. With a soft gluon being emitted in the middle of two cone jets, this could lead to sufficient energy overlap of the two initial cones, that the algorithms would merge the two cones into one, as pictured in Figure 2.12. The construction of jets around the midpoint seeds should avoid this unwanted behaviour within the midpoint cone algorithm [52]. However, this only fixed the infrared safety temporarily, considering an additional second radiation would already spoil the midpoint fix. Thus, only a seedless cone algorithm, like SISCone [53] is safe to all orders in perturbation theory.

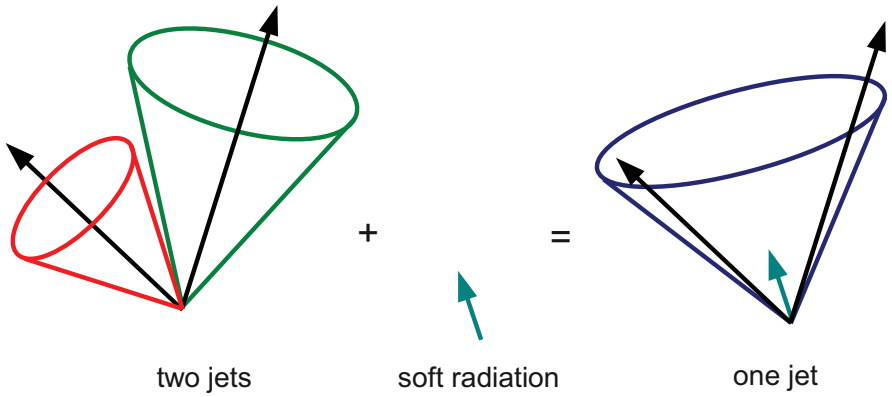

Figure 2.12: *Infrared unsafe behaviour of jet algorithms is illustrated: If an additional soft gluon changes the output of the algorithm an infrared unsafe algorithm has been used.*

2.6.3 Classification of Jet Algorithms

Two principal types of algorithms are in common use:

- **Subsequent Clustering Algorithms:**
  Subsequent clustering algorithms iteratively combine objects that have the smallest distance of all pairwise combinations possible. The latter have predominantly been used in e^+e^- and $e^{\pm}p$ collisions, first in form of the Jade algorithm [54, 55] and nowadays as k_T algorithm [56–59]. In this work an implementation [60] is used, which applies special techniques to speed up the clustering of N particles significantly. It is working at a complexity $\mathcal{O}(N^2)$ or even $\mathcal{O}(N \log N)$ at high multiplicities, and is algorithmically exactly equivalent to the traditional implementations which had usually complexity $\mathcal{O}(N^3)$. The algorithm is known to be collinear and infrared safe. The algorithmic prescription is given in more detail in subsection 2.6.4. More subsequent clustering algorithm types have emerged lately, like the Anti-k_{T} algorithm [61], however they will not be focused within this work.

- **Cone Type Algorithms:**
  Cone type algorithms [3] traditionally have been employed in hadron-hadron collisions where objects are clustered together that are close in angle around a high-energetic seed. The main reason behind the popularity of cone jet algorithms is due to their relatively simple geometric shape. This eases the treatment of the underlying event and pile-up. To first order, both effects contribute a uniformly distributed amount of energy to the event, therefore also to each jet. The prior knowledge of the jet area is therefore a convenient property for the determination and application of

the respective corretions. Especially this is the reason for their popularity at hadron colliders. However the actual simple implementations have certain draw backs. Usually at first, a seed is chosen from which the clustering starts. Unfortunately, due to ambiguity within the definition of the seeds, or a respective (noise) threshold in addition to those, the final output becomes dependent on the choice for the seeds or at least can change significantly if an infrared energy deposit is added to the event. The second issue is the treatment of overlapping cones. Due to the initial cone creation around seeds, specific energy deposits can become part of multiple cones. Thus a dedicated split-merge procedure is implemented for disambiguation. Unfortunately, implementations which do not handle these issues properly, ensuring collinear and infrared safeness, cannot be used in pQCD calculations. In fact extra singularities show up due to that, or the calculations simply do not converge anymore. Several approaches have been made, for example in the midpoint-cone algorithm, in order to fix this. However in most cases the algorithmic prescription then becomes much more complicated, the simple jet shape gets lost and even strange effects, like the so called "dark towers" could arise. In this case the additional fixes result in energy deposits, which do not get clustered into any final jet. Additionally, the infrared safeness could only be secured to one additional order in the coupling. The SISCone algorithm is applied in this work. It is an implementation of the cone-type algorithms, which is infrared and collinear safe and has a comparable behaviour as the midpoint-cone algorithm, as far as sufficiently inclusive observables are concerned. However, the infrared and collinear safeness comes at a certain cost in complexity and run-time requirements. More details are given in subsection 2.6.5.

2.6.4 The $\mathrm{k_T}$ Algorithm

The k_T algorithm uses a resolution parameter to decide about the way input objects are combined into jets. This parameter also depends on the distance ΔR and additionally on the energy of the input object. This additional dependence gives an invariant-mass like interpretation of these parameters, which has several advantages with respect to the cone-approach, especially when used with perturbative calculations including higher orders. In fact the algorithm mimics a walk backwards through the QCD evolution of a parton showering process and hadronisation clustering particles radiated from an original parton. The algorithm included within the CMS Software (Sec. 3.5.5) is described in [60], and technically the k_T jet finder of the fastJet package [62] is directly interfaced to the CMS software. The first implementation of this interface was developed in the scope of the presented work. The retrieved jet final state of this jet finder is identical with the original prescription in [63]. A method to find the nearest neighbour of a jet using methods given in modern computer science literature, without calculating the distance between all possible combinations of two inputs, speeds up the new implementation ($\mathcal{O}(N \ln N)$) with respect to the old one ($\mathcal{O}(N^3)$). For increased clarity several names of variables within this work are deviating

from the original description. There are several possible choices for the resolution parameters. The most common settings for hadron-hadron collisions have been used throughout this work, because they are best comparable to the most frequently used cone algorithms in this field. The choice of resolution parameters is referred to as the "ΔR-scheme". For each input object h_k with transverse momentum p_T, two distances d_{kB} and d_{kl} are defined in the following way:

$$d_{kB} = p_{\mathrm{T},k}^2 \tag{2.42}$$

which relates the beam distance of each object directly to the transverse momentum p_T and

$$d_{kl} = \min(p_{\mathrm{T},k}^2, p_{\mathrm{T},l}^2) R_{kl}^2 \tag{2.43}$$

with

$$R_{kl}^2 = (y_k - y_l)^2 - (\Phi_k - \Phi_l)^2, \tag{2.44}$$

where d_{kl} defines the distance between the input objects h_k and h_l. In contrast to the cone type algorithms, these definitions weight angular distances with the transverse momenta of the contributing objects. This relates the final description of the event in a better way to the underlying physical dynamics, than a simple geometrical interpretation does. The algorithmical prescription itself is given in the Appendix B.1.1 together with some additional modes to run the algorithm for special purposes.

2.6.5 The Seedless Infrared-Safe Cone Algorithm

The Seedless infrared-safe cone algorithm (SISCone) [53] was developed, with the idea to provide a jet algorithm that has the main features of the midpoint-cone algorithm with respect to jet observables, but without the shortcomings in the areas of infrared- and collinear safety or dark towers. As one problem has been identified to stem from the usage of seeds, seedless cone algorithms have been developed. Initial implementations usually required $N2^N$ iterations to find jets among N particles, which made them unusable in hadron colliders.

The version actually used at CMS speeds this up to $N^2 \ln N$, which leads to a negligible running time with respect to other reconstruction tasks, like track finding at the expected LHC pp hadron multiplicities of $\mathcal{O}(1000)$. However, in the current implementation it cannot be used for heavy ion collisions, due to its extensive memory requirements and soaring computing time requirement (multiplicities $\mathcal{O}(40000)$, $\approx 1\mathrm{h}/\mathrm{event}$).

The algorithm is split into two main steps. At first all possible combinations of the input into cones have to be found. This is also the basis for the infrared safety makes

sure, that no possible jet could be overseen during the clustering. A hash is build by an exclusive-or ("checkxor") of all integers identifying the constituents of the protojets and saved to memory. This procedure runs with $Nn\ln n$ complexity for n typical particles clustered to one jet. Using the hashing procedure equal protojets can be identified using an xor procedure. Due to this procedure the algorithm also has a quite large memory consumption of the order $\mathcal{O}(Nn)$. A less memory consuming version exists, but the trade-off is a significantly larger computation time ($\mathcal{O}(Nn^{3/2})$).

Within protojets, the input objects are not unambiguously attributed to one protojet, so during the second, so called "split-merge" step, input objects need to be assigned unambiguously to final jets. This is realised by splitting or merging of overlapping cones dependent on the fraction f of overlapping $\tilde{p}_{\mathcal{T}} = \sum_{i\in\text{jet}} |p_{\mathrm{T,i}}|$ content. Here, $\tilde{p}_{\mathcal{T}}$ is the scalar sum of the constituents p_{T}, which is chosen to ensure boost invariance and IR safety. Additionally to the prescription in pseudo-code in the Appendix B.1.2, some additional technical tweeks are needed to reach the final execution time.

2.7 Latest Results from the Tevatron

The latest tests of perturbative QCD at the highest available energies using the inclusive jet cross section have been conducted by the Tevatron experiments DØ and CDF. Both experiments lately published results of data taken at a centre of mass energy of $\sqrt{s} = 1.96\ \mathrm{TeV}$ in $p\bar{p}$ collisions. DØ [64] presented jets for an integrated luminosity of $0.7\ \mathrm{fb}^{-1}$ subdivided into six bins in rapidity up to $|y| < 2.4$. CDF [65] accumulated $1.13\ \mathrm{fb}^{-1}$ of integrated luminosity. Five bins have been used to subdivide the rapidity range of $|y| < 2.1$.

These analyses set the starting point of the inclusive jet measurement at the LHC. Additionally they are also included in the global PDF fits (e.g. [66]), which will be used within the first LHC analyses.

2.7.1 Jet Reconstruction

Both experiments use the Tevatron Run II midpoint cone algorithm within their jet reconstruction. CDF also published results using the k_T clustering algorithm [67], which as well had been used in former publications of both collaborations. However, only midpoint results are presented within this section. The two experiments use dedicated measurements to improve the calorimeter response, which directly affects the experimental jet input. However, even with a fully corrected response of single cells, the dead material in between those cells cannot be corrected for. Therefore, reconstructed jets still need additional corrections to fully remove systematic biases of the detectors.

At first, jet-by-jet corrections account for the energy loss in the detector material and due to the shower evolution which leads to particles escaping the jet cone. Also contributions from pile-up and multiple parton interactions are removed, which are measured to be $0.97\ \mathrm{GeV}$ per jet in the CDF measurement.

In case of DØ the corrections are mainly based on γ-jet and dijet balancing measurements, which exploit the fact, that due to transverse momentum conservation the balanced jet should have the same transverse momentum. γ-jet balancing even allows an absolute energy scale determination, as the photon can be measured much better in the electromagnetic calorimeter than the balanced hadronic jet. Additionally, the different response of quark and gluon jets as well as the energy loss due to undetectable particles, like muons or neutrinos, are determined from Monte Carlo studies. CDF derives all jet corrections using PYTHIA followed by a full detector simulation and evaluates uncertainties from a comparison to HERWIG results. However, they show in additional studies, that their simulation reflects the true detector response.

The next class of corrections is finally applied to the whole spectrum in order to account for event selection inefficiencies and the smearing due to the finite detector resolution. Both experiments select their events from prescaled trigger streams and apply noise suppression cuts dependent on the transverse component of the fourvector sum of all measured energies ($\not{E}_\mathrm{T}$). As the beamspot varies quite sizeably at the Tevatron, CDF requires the event vertex

to lie within ± 60 cm around the centre of the detector along the beampipe. DØ requires ± 50 cm to retain the functionality of the projective calorimeter geometry. DØ is using a four-parameter Ansatz function to parametrise the p_{T} and y dependence of the spectrum folded with a parametrisation of the respective resolution, which is measured from data via dijet balance. The unsmearing corrections are then extracted from a fit of the smeared Ansatz to the measured jet spectrum. In contrast to that, CDF again fully relies on its detector simulation and derives the corrections using a PYTHIA spectrum which is passed through the full detector simulation and weighted to reflect the measured spectrum. The correction factors are extracted bin-wise from a comparison to the pure PYTHIA spectrum without detector simulation.

2.7.2 Theory Comparison

For a final comparison to the best available pQCD prescription, both experiments rely on NLO calculations. These spectra do not reflect the fully hadronised final state of the measurement, so additional corrections need to be applied to enable a reasonable comparison. Both studies use fastNLO (see Sec. 3.6.5, 4.1) within their theory evaluation. While CDF uses the CTEQ61M parton distribution to derive the central NLO values and uncertainties and selects the renormalisation and factorisation scale as $\mu_R = \mu_F = \frac{1}{2} p_{\mathrm{T}}^{\mathrm{Jet}}$, DØ relies on CTEQ65 and chooses the scales as $\mu_R = \mu_F = p_{\mathrm{T}}^{\mathrm{Jet}}$. Each study also compares to a calculation using the MRST2004 PDF, which does not significantly change the results. The required non-perturbative corrections are derived from studying the tuned PYTHIA results with and without non-perturbative effects. CDF uses PYTHIA tune A for this while DØ relies on PYTHIA tune QW together with CTEQ6.5M and a two loop α_S calculation. Uncertainties are derived from the deviation to the HERWIG results.

A comparison of the measured cross section with the respective NLO calculations is presented in Figure 2.13. Detailed ratios of the theoretical predictions and the measurements is given in Figure 2.14 together with the respective systematic uncertainties from theory and experiment. The uncertainty due to the luminosity, which directly translates into an uncertainty of the total scale of the spectra, is not included. The experiments give about six percent uncertainty on the luminosity. Overall it has been observed, that the theory describes the measurements quite well within the given uncertainty bands. However, the NLO description in most cases is slightly larger than the measured points and the ratio is not flat over the presented range in transverse momentum. Not surprisingly, larger deviations are observed at the edges of the studied phase space, where the experimental and statistical uncertainties are large.

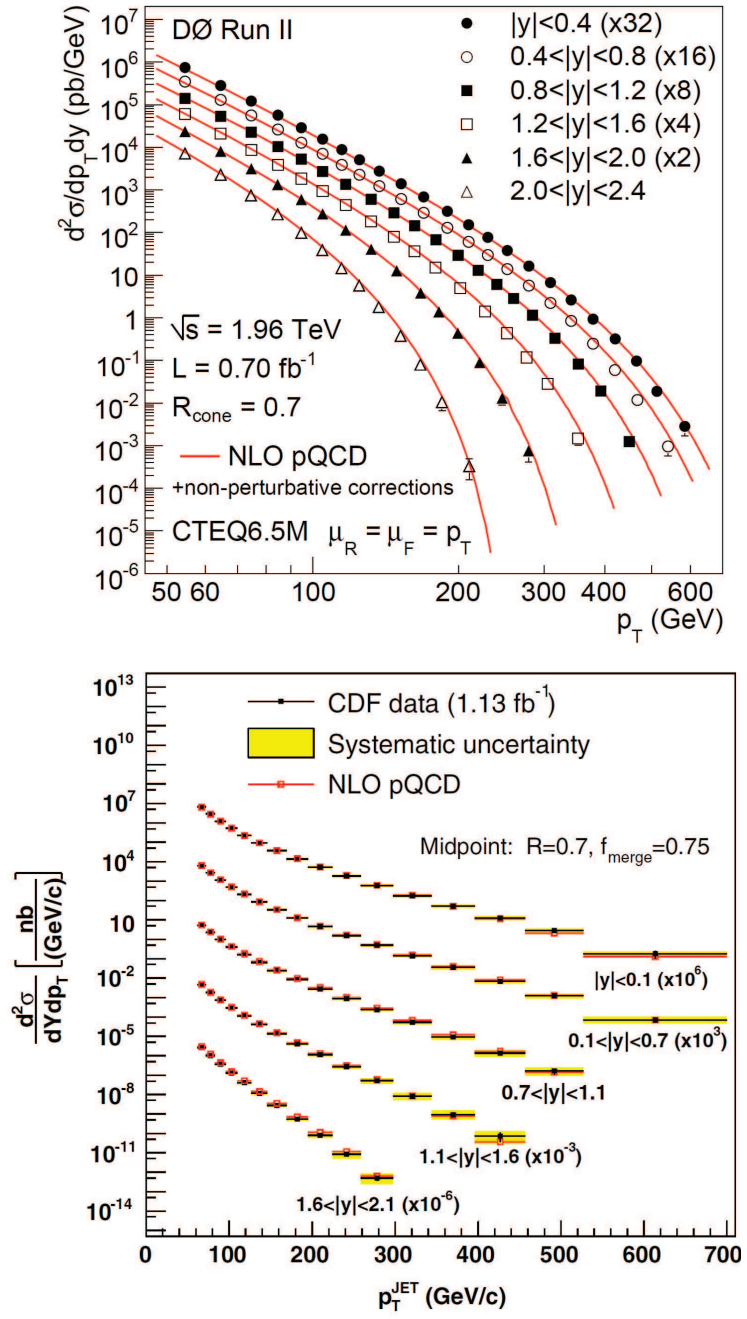

Figure 2.13: *The inclusive jet cross section measured from data and compared to calculations in NLO is presented as published by DØ (top [64]) and CDF (bottom [65]).*

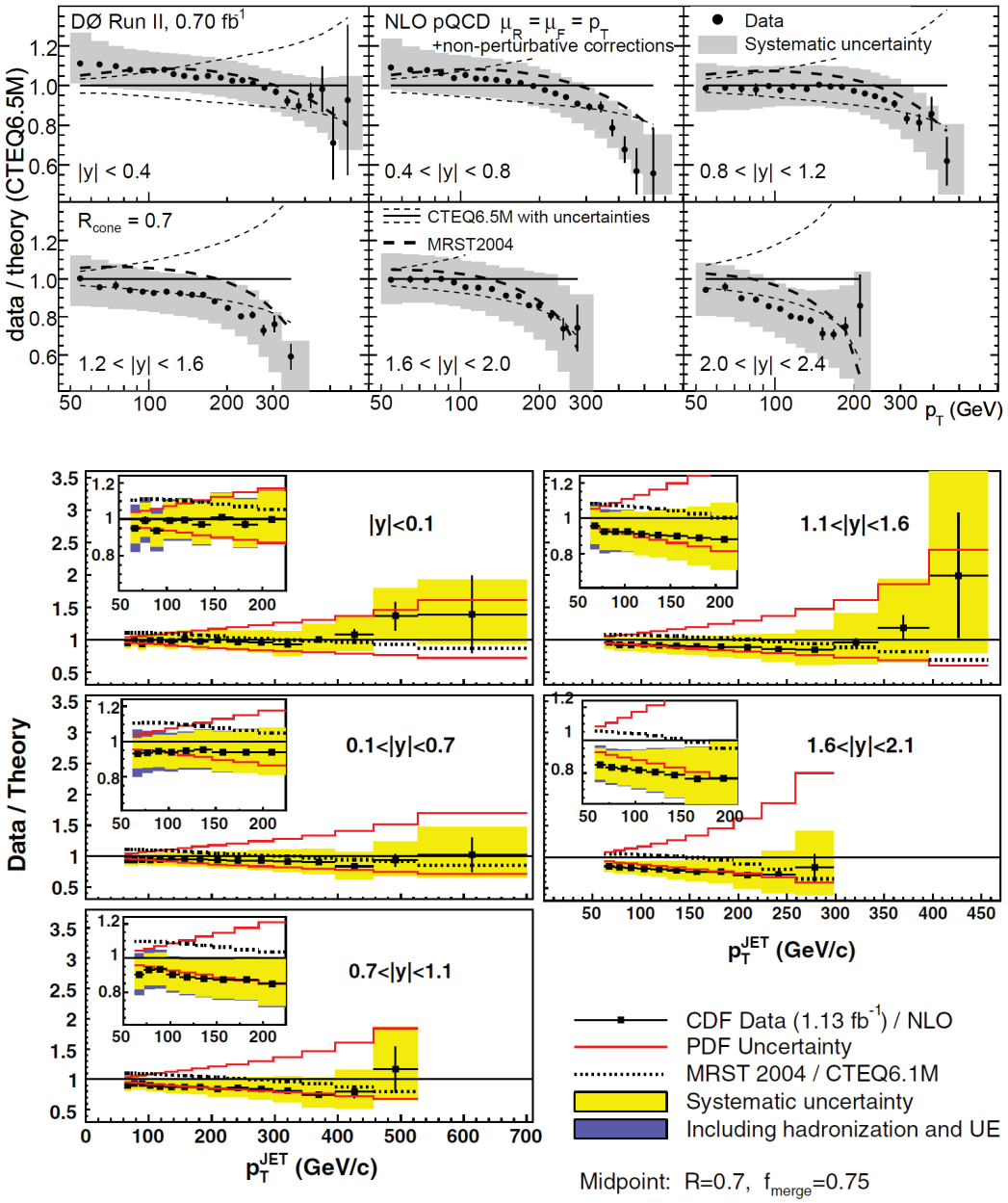

Figure 2.14: *The inclusive jet cross section measured from data and compared to calculations in NLO is presented as published by DØ (top [64]) and CDF (bottom [65]). The ratios of data and theory are given together with the respective dominant uncertainties.*

Chapter 3

The CMS Experiment

3.1 Large Hadron Collider

After exploiting the energy range up to a centre of mass energy of $200\ \mathrm{GeV}$ at LEP[1] for several years it has become more and more evident that for discoveries like the Higgs boson and new physics beyond the Standard Model larger energy scales would have to be probed. Therefore, the Large Hadron Collider (LHC) [68] was planned and built. Although reaching a centre of mass energy up to $\sqrt{s} = 1.96\ \mathrm{TeV}$ the experiments at the Tevatron, situated at the Fermi National Accelerator Laboratory near Batavia/Illinois, USA, were not able to detect the Higgs boson up to now. The experiments currently even exclude a Standard Model Higgs boson in the mass range between 160 and $170\ \mathrm{GeV}$ [69].

The LHC underground structure together with the cooling state of the magnets in October 2009 is given in Figure 3.1. After the switch-on in 2008, the LHC crew was able to circulate beams with an energy of $450\ \mathrm{GeV}$ around the ring. However, while trying to run the magnets at currents necessary to control a $5\ \mathrm{TeV}$ beam, a severe accident occurred. Due to this, the LHC needed to be switched off and is currently under repair. The restarting schedule foresees first collisions at injection beam energies of $450\ \mathrm{GeV}$. After that, collisions at a centre of mass energy of $\sqrt{s} = 7\ \mathrm{TeV}$ are planned. If everything goes well, this might be extended to $10\ \mathrm{TeV}$. During the following shutdown the machine will be prepared for design energies and it will ultimately collide beams of 2808 bunches of about 1.15×10^{11} protons at a centre of mass energy of $\sqrt{s} = 14\ \mathrm{TeV}$ leading to a design lu-

[1] **L**arge **E**lectron **P**ositron Collider

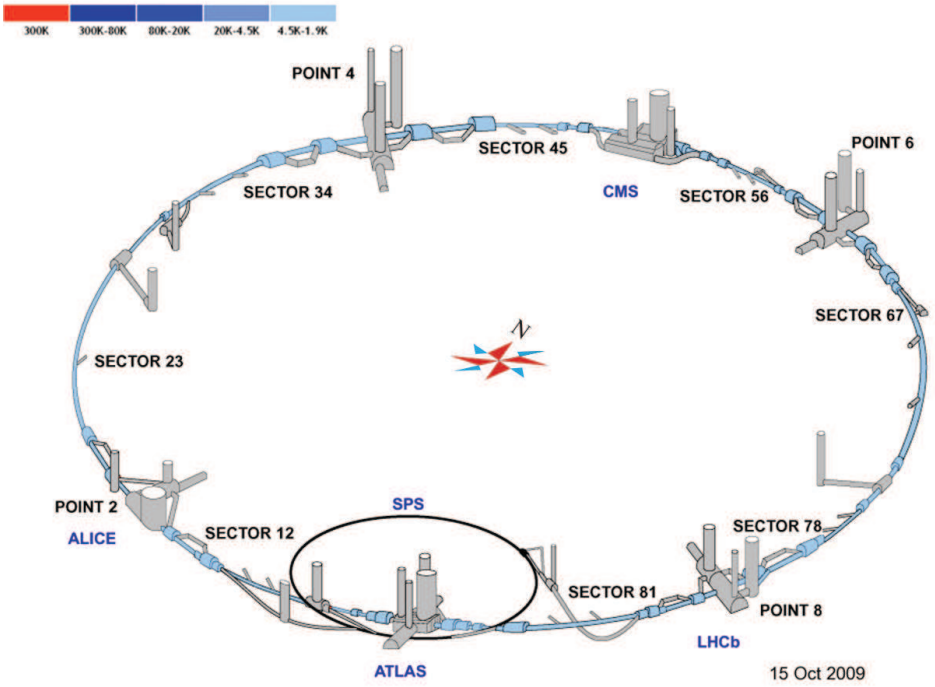

Figure 3.1: *The LHC underground structure [70] with the cooling state of the magnets in October 2009 [71] is shown, points without experiment labels are mainly for LHC service purposes, beam dumps are near Point 6, Beam injections at Points 2 and 8.*

minosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. In order to get new insights into the state of the early universe, beams with 592 bunches of 7×10^7 lead nuclei will be accelerated and brought to collision with an energy of $5.52\ \text{TeV}$ thus probably producing a Quark Gluon Plasma with energy densities comparable to the the situation shortly after the Big Bang. For these lead nuclei collisions a design luminosity of $10^{27}\ \text{cm}^{-2}s^{-1}$ is forseen.
The LHC is situated in the former tunnel of the LEP collider at CERN (**C**onseil **E**uropéen pour la **R**echerche **N**ucléaire) situated near Geneva, Switzerland, astride the Franco-Swiss border. About 5000 superconducting coils consisting of niobium-titan, cooled down to liquid helium temperatures and producing magnetic fields of about 8 Tesla are used to hold the beams on track within the $27\ \text{km}$ circumference of the tunnel in a depth of 50 to $175\ \text{m}$ below the surface.
At four distinct points of the tunnel the beams are brought to collision such that they can be investigated by the experiments ALICE [72], ATLAS [73], CMS[74] and LHCb [75]. In addition to these primary experiments situated at the interaction points, TOTEM [9] equips the tunnel before and after the CMS cavern with so called roman pots, extreme forward detectors, at distances of about $150\ \text{m}$ to $400\ \text{m}$ to the interaction point. This is done to achieve an extreme η coverage for studies of inelastic proton-proton collisions, where only one proton stays intact after the collision, as well as elastic collisions to improve the knowledge on the proton structure. Both will contribute to the luminosity measurement required by the CMS experiment.

The aim of the LHC is to provide the experiments with collision events at sufficiently high centre of mass energies $\sqrt{s}$. Therefore the beam injected into the LHC from the SPS (**S**uper **P**roton **S**ynchrotron) storage ring with an energy of $450\ \text{GeV}$ is accelerated using a radio frequency of $400\,\text{MHz}$ within niobium sputtered cavities to the desired value. The event rate $\dot{N}_{\text{event}}$ measured at the experiments for a dedicated process with the cross section σ_{process} is given as

$$\dot{N}_{\text{event}} = \mathcal{L}\sigma_{\text{process}}. \tag{3.1}$$

The beam at the LHC is kept on track around the ring mainly by the dipole bending magnets. A specialized beam optic consisting of further quadrupole magnets, and specialised higher order magnets is used to control the beam density and focus the beams onto the collision points. The ideal arrangement of the beam optical devices leads to a maximum number of particles circling around on a stable orbit, which emerges due to the interaction of the beam with the machine. Particles with too strong deviations from this stable conditions get lost within the very first circulations. The beam shape can be described as a three dimensional ellipsoid, which follows from solving the differential equation describing the beam within the machines beam optics. The solutions also contain so called betatron oscillations, which characterise the particles trajectories around the equilibrium path given by the machines parameters. The betatron function β^* is part of these solutions, and is strongly dependent on the beam optics arrangements. It is related to the transverse

envelope of the beam by the emittance ϵ:

$$E_{\text{max}}(z) = \sqrt{\epsilon}\sqrt{\beta^*(z)}\,, \tag{3.2}$$

where z describes the distance along the beam pipe. The emittance resembles the transverse "temperature" of the beam. The interested reader is pointed to the excellent literature on this topic, e.g. [76]. In case of a symmetric collider like the LHC, the instantaneous luminosity $\mathcal{L}$ is given as:

$$\mathcal{L} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi\epsilon_n \beta^*} F\,, \tag{3.3}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalised transverse beam emittance, β^* the value of the betatron function at the collision point, and F the geometric luminosity reduction due to the crossing angle at the interaction point:

$$F = \left[1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right]^{-\frac{1}{2}}\,, \tag{3.4}$$

with θ_c being the full crossing angle at the interaction point, σ_z the root mean square of the bunch lengths, and σ^* the root mean square of the transverse beam size at the interaction point. The above expression assumes round beams, with $\sigma_z \ll \beta$, and with equal beam parameters for both beams. The revolution frequency f_{rev} is the frequency with which a single proton circulates around the 27 km circumference of the ring. The design parameters for the LHC and beam crossing at CMS for $\sqrt{s} = 14$ TeV are given in Table 3.1. These parameters are design estimates, however using equations 3.3 and 3.4 with the parameters given in Table 3.1 leads to $\mathcal{L} = 9.91 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$, which is quite close to the desired value.

3.2 CMS Collaboration

As of June 2008 there are 3600 people working for CMS, 3000 of them are scientists and engineers. They come from 183 institutes in 38 countries, spanning Europe, Asia, North- and South America, and Australasia [78]. The collaboration was founded to build and maintain the CMS (**C**ompact **M**uon **S**olenoid) detector as well as to exploit the data derived from it. One of the main focuses is on the search for the Higgs-Boson as well as revelations of new physics beyond the Standard Model. With contributing people coming from all over the globe, storing and analysing the huge amounts of data produced by CMS is one main challenge for the experiment.

Table 3.1: *The LHC machine parameters are given for the design beam energy and the planned peak luminosity. Values from [68, 74, 77],* $\gamma_r = \frac{E_{\text{beam}}}{m_p}$.

Quantity	Unit	Collision at 7 TeV
Luminosity $\mathcal{L}$	$\frac{1}{\text{cm}^2\text{s}}$	10^{34}
Protons per bunch N_b	1	1.15×10^{11}
Number of bunches n_b	1	2808
revolution frequency f_{rev}	kHz	11.2
rel. γ_r	1	7460
normalised transverse emittance ϵ_n	μm $\cdot$ rad	3.75
crossing angle at IP θ_c	μrad	285
bunch length σ_z	ns	0.265
bunch length σ_z	m	0.0795
bunch transverse width σ^*	μm	16.7
value of β^* at the IP	m	0.55

3.3 Compact Muon Solenoid

In contrast to ALICE, or LHCb, the multi-purpose detectors CMS and ATLAS are designed to cover a broad variety of physics analyses including the search for new physics as well as more precise measurements of already known quantities of the Standard Model.
With a length of $21.5\ \mathrm{m}$, a diameter of $15\ \mathrm{m}$ and a total weight of 12500 tons, the CMS detector (fig. 3.2) consists of several parts dedicated to measure different properties of the particles produced within the collisions. The detector parts were assembled at the surface above the interaction point and were lowered down into the cavern, where everything was ready for data taking in 2008. The tracker as well as the calorimeters are placed within a $13\ \mathrm{m}$ long superconducting solenoid providing a magnetic field of 3.8 Tesla. The muon chambers are situated in the region outside the superconducting solenoid, within the return yoke of the magnet. A magnetic field strength of about 1.8 Tesla bends tracks of charged particles escaping the inner detector parts. Huge efforts were necessary to construct the detector in a way that it is capable to achieve the planned tasks. A brief overview of the detector components is given in the following sections. More detailed information can be found in the CMS Physics Technical Design Reports [77, 79] or other technical documentations, where [74] gives the latest official update.

3.3.1 The Detector Coordinate System

The CMS coordinate system is used to identify points within the detector:

- Cartesian coordinates with the centre at the nominal interaction point:
  - The x-axis is horizontal, pointing south towards the LHC centre.
  - The y-axis is vertical pointing upwards.
  - The z-axis is horizontal pointing west, tangentially to the beamline.
- Polar coordinates:
  - r: the radial distance to the beamline
  - ϕ: the azimuthal angle measured in the x/y-plane. Whereas $\phi = 0$ is pointing to the $+x$-axis and $\phi = \pi/2$ to the $+y$-axis.
  - Θ: the polar angle with respect to the z-axis. Whereas $\Theta = 0$ corresponds to the $+z$-direction and $\Theta = \pi$ to the $-z$-axis.
  - η: the pseudo-rapidity. It is defined as $\eta = -\ln\left[\tan(\frac{\Theta}{2})\right]$. This leads to an η of 0 for particles moving perpendicular to the beamline and the beamline itself has a pseudo-rapidity of $+\inf$ in $+z$ and $-\inf$ in $-z$ direction.
  - y: the rapidity. It is defined for a fourvector with longitudinal momentum component p_z parallel to the beampipe and energy E as $y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z}$.

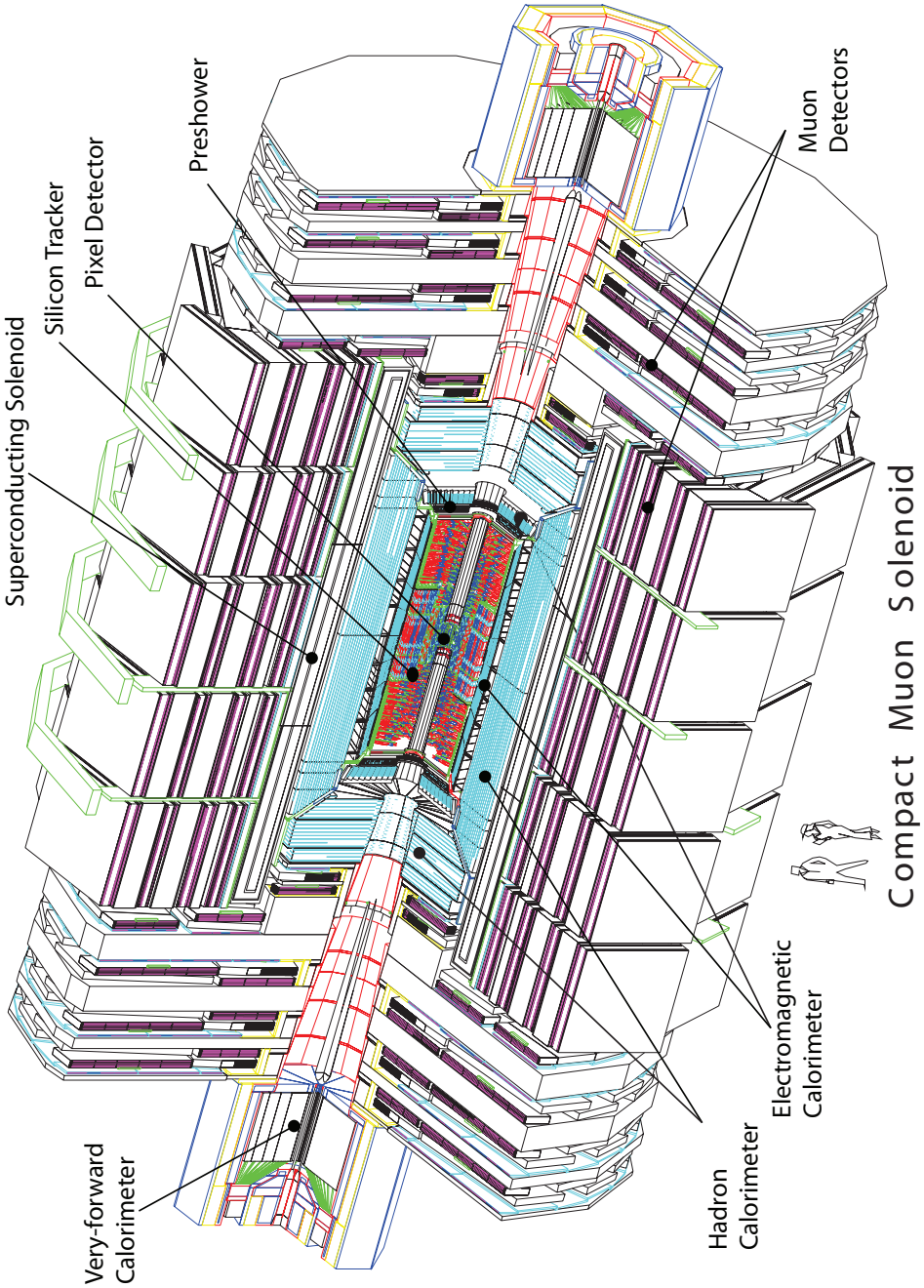

Figure 3.2: *The presented Figure gives an insight into the CMS detector with all important parts labelled.*

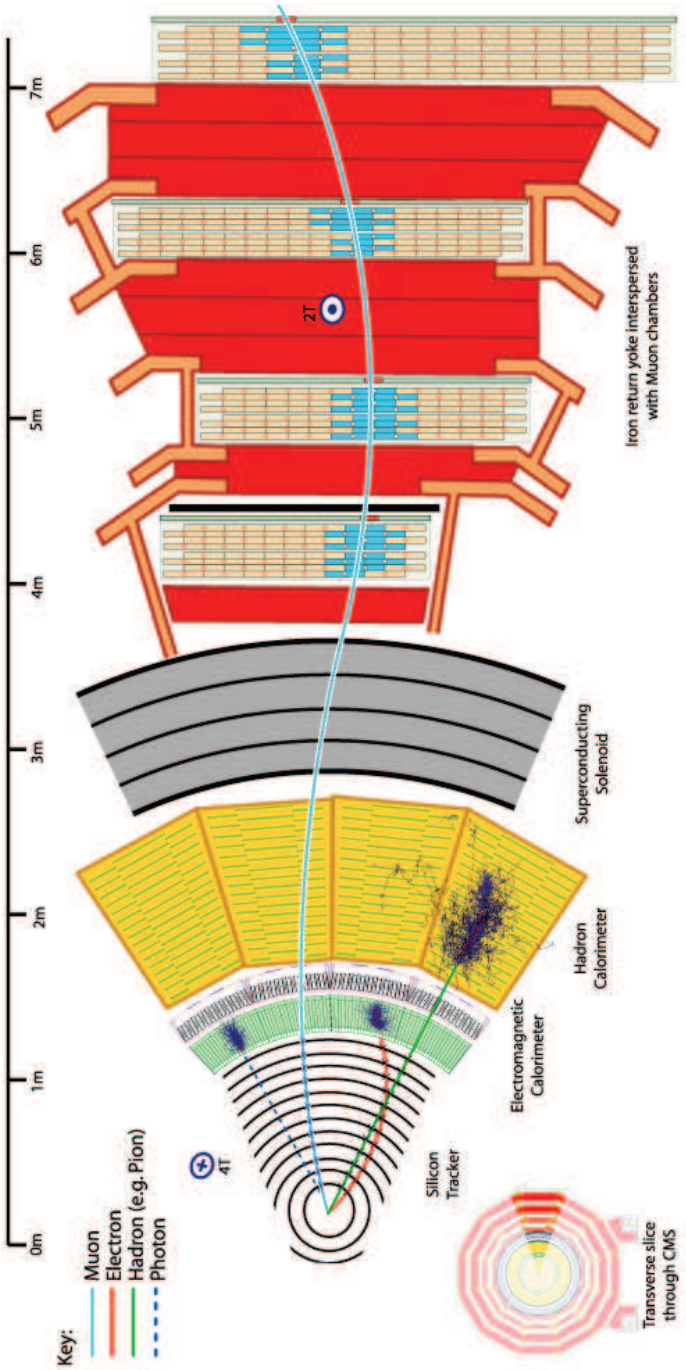

Figure 3.3: *Profile of the CMS detector with tracks of an electron, a photon, a hadron and a muon traversing the detector producing showers or hits in the active material.*

3.3.2 Tracking System

Measuring momentum and determining decay vertices of particles is the task of this detector part, therefore it is placed directly next to the beam pipe. The tracking system comprises a volume given by a cylinder of length $5.8\ \mathrm{m}$, diameter $2.8\ \mathrm{m}$ and providing coverage up to $|\eta| < 2.4$. Having numerous silicon strip and pixel detectors, it is optimised for spatial resolution leading to a good pattern recognition, an essential feature for good track reconstruction. Particles traversing the detector material deposit energy within consecutive detector layers. A pattern recognition algorithm is applied to find deposits originating from the same particle, and which are compatible with the equation of motion for the particle using the properties of the magnetic field. Often there are several possible trajectories, so the algorithm has to find the most probable one and declare it as a track. The default multipurpose track-finding in CMS uses a Kalman filter technique [80]. With this in hand it is possible to determine charge and momentum of particles with finite charge.

Time and spatial resolution have to be good enough to avoid measurements of overlaying tracks which is a tremendous task since there are about 1000 particles per event. In addition to that, pile-up events have to be taken into account: With LHC running at design luminosity, there will be multiple proton-proton interactions happening within a single bunch crossing leading to even more particles entering the tracking system simultaneously. Furthermore, the material of the system has to withstand a high level of radiation around the beam pipe for several years. Thus an all-silicon tracking system was chosen, since it best fulfils these requirements. This device is split up into two subsystems, namely the Silicon Pixel Detector and the Silicon Strip Detector, which are explained below.

Silicon Pixel Tracker

Estimating the vertex of the hard interaction and additional decay vertices of particles is the main task for this innermost part of CMS. Being set next to the beam pipe, it consists of 3 concentric barrel layers of hybrid pixel detectors arranged at radii of $4.4\ \mathrm{cm}$, $7.3\ \mathrm{cm}$ and $10.2\ \mathrm{cm}$. The size of the pixels is $(100 \times 150)\ \mu\mathrm{m}^2$. At each side of the barrel layers an endcap disk extending from 6 to $15\ \mathrm{cm}$ in radius, is placed at $|z| = 34.5\ \mathrm{cm}$ and $46.5\ \mathrm{cm}$. The barrel comprises 768 pixel modules arranged into half-ladders of 4 identical modules each, adding up to a total of 66 million pixels. Due to the magnetic field, the electrons within the depleted semiconductor are distracted during their way to the readout electronics by the Lorentz force. Therefore the charge measured will be finally shared between mostly two readout channels, thus allowing to improve the r-ϕ-resolution by using the known magnetic field strength. Also the endcap disks benefit from the Lorentz effect. This leads to a measured spatial resolution of about $10\ \mu\mathrm{m}$ for the r-ϕ-measurement and about $20\ \mu\mathrm{m}$ for the z measurement.

Silicon Strip Tracker

The barrel region of this tracker is divided into two parts. An inner part with four layers out of which the first two are made with 'stereo' modules in order to provide both r-ϕ and r-z coordinates. This leads to a single point resolution of about $27\ \mu\mathrm{m}$ in r-ϕ and $230\ \mu\mathrm{m}$ in z. The outer part comprises 6 layers, of which the first two have a 'stereo' configuration leading to a single point resolution of about $40\ \mu\mathrm{m}$ in r-ϕ and $530\ \mu\mathrm{m}$ in z. The endcap region is covered by two tracker parts. There are 9 disks placed in the region $120\ \mathrm{cm} < |z| < 280\ \mathrm{cm}$ and 3 small disks in the gap between endcap and barrel region. Also within these parts the 'stereo' arrangement of modules is used to provide an optimal resolution. The 15400 modules cover an area of $210\ \mathrm{m}^2$ in total adding up to 9.6 million silicon strips.

3.3.3 Electromagnetic Calorimeter (ECAL)

Electrons, positrons, and photons create an electromagnetic shower, a cascade of photons and electron-positron pairs, in the calorimeter material. Compton scattering and the photoelectric effect due to the shower particles, as well as the incident particle, pass energy to the calorimeter. The total energy of the particle, which initially entered the detector, is proportional to the total sum of all the small energy deposits within the detector. Single energy deposits within the scintillator material lead to the emission of photons. Their energy can finally be measured using silicon avalanche photodiodes in the barrel and vacuum photodiodes in the endcaps. The electromagnetic calorimeter is a hermetic and homogeneous calorimeter comprising 61200 lead tungstate ($\mathrm{PbWO_4}$) crystals mounted in the central barrel part, and 7324 crystals in each of the two endcaps. Lead tungstate scintillating crystals were chosen because of their high density leading to short radiation ($X_0 = 0.89\ \mathrm{cm}$) and Molière ($2.2\ \mathrm{cm}$) lengths, in order to stop all electrons and photons within a minimal depth of the material. This enabled the CMS collaboration to build a compact calorimeter within the solenoid. Fast light emittance (80% of the light is emitted within $25\ \mathrm{ns}$), fine granularity and radiation resistance were other factors which contributed to the decision to use this material. Finally the ECAL has to provide an adequate energy resolution to reconstruct Higgs bosons, which decayed into two photons. Since the energy resolution of the ECAL is better than the energy resolution of the Hadron Calorimeter it can be used to calibrate hadronic jets using the energy of a photon being produced back to back to a high energetic jet.

3.3.4 Hadron Calorimeter (HCAL)

Hadrons traverse the ECAL largely without hadronic interactions. These particles are stopped not until the hadron calorimeter where they create hadronic showers as a result of the interaction of the incident particles with the nuclei of the absorber material. These showers consist mostly of pions, kaons, nucleons, and fragments of the nuclei. The main

purpose of the HCAL is energy measurement with an appropriate spatial resolution to contribute to the overall energy measurement of the calorimeter system. Like in the ECAL, the energy is determined using scintillator material connected to wavelength shifting fibres to carry the light to the readout system made of multi-channel hybrid photodiodes. In the hadron forward calorimeter quartz fibres are used as scintillating material and to transport the light, which is finally detected by photomultipliers.

As the HCAL is also placed within the magnetic field, the design is mainly influenced by magnetic parameters and the requirement to provide a good overall coverage for the measurement of missing transverse energy (E_T^{miss}). From energy and momentum conservation the sum of all transverse energy components within the collision must vanish. The non-vanishing part during a measurement of this summation is called missing transverse energy and is used to determine parameters of particles like neutrinos or unmeasured muons, which have either few or no interactions with the detector material at all, and thus can not be detected directly. The HCAL system consists of 4 parts: hadron barrel (HB), hadron outer (HO), hadron endcap (HE) and hadron forward (HF). Brass has been chosen as absorber material since it has a reasonably short nuclear interaction length, is easy to machine and is non-magnetic.

Test beam results of the complete calorimeter layout (including the ECAL) indicate, that an energy resolution $\sigma_E/E = 100\%/\sqrt{E} \oplus 4.5\%$ is achievable between 30 GeV and 1 TeV [81]. More details on the components of the HCAL is given in the following:

Hadron Barrel Calorimeter (HB)

The barrel part of the HCAL consists of 64 segments covering the pseudo-rapidity region of $|\eta| < 1.4$. Each plane in the η segmentation is split up into a ring of 36 segments covering the complete ϕ-region. This results in 2304 towers with a segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. The towers themselves are comprised of 15 brass plates of 5 cm thickness with 3.7 mm scintillator plates between them. The first scintillator plate has a thickness of 9 mm to produce better light output.

Hadron Outer Calorimeter (HO)

The hadron outer detector contains scintillators with a thickness of 10 mm, which line the outside of the outer vacuum tank of the coil and cover the region with $|\eta| < 1.26$. This increases the effective thickness of the HCAL to over 10 nuclear interaction lengths leading to a better coverage for high energetic tails providing improved E_T^{miss} measurements. The HO is physically located inside the barrel muon system and is hence constrained by the geometry and construction of that system.

Hadron Endcap Calorimeter (HE)

Each HE covers a pseudo-rapidity range of $1.3 < |\eta| < 3.0$ with 14 towers, which have a ϕ-segmentation varying from 5° in the lower η-regions to 10° in the higher regions. Within the HE the η segmentation of towers varies from 0.087 to 0.35 with increasing $|\eta|$. The total number of HE towers is 2304.

Hadron Forward Calorimeter (HF)

The coverage between pseudo-rapidities of 3.0 and 5.0 is provided by the hadron forward (HF) calorimeter located $11.2\ \mathrm{m}$ from the interaction point. It consists of $1.65\ \mathrm{m}$ deep steel absorbers and embedded radiation-hard quartz fibres, which provide a fast collection of the emitted Cherenkov light within the fibres. Two different fibre lengths ($1.43\ \mathrm{m}$ and $1.65\ \mathrm{m}$) are used for adjacent, parallel fibres, in order to disentangle signals originating from the central interaction and ones originating e.g. from cosmic interactions. There are 13 towers in η, all with a size given by $\Delta\eta \approx 0.175$, except for the lowest-η towers with $\Delta\eta \approx 0.1$ and the highest-η towers with $\Delta\eta \approx 0.3$. The ϕ segmentation of all towers is 10°, except for the highest-η one which has $\Delta\phi = 20^\circ$. This leads to 900 towers and 1800 channels in the 2 HF modules.

3.3.5 Superconducting Solenoid

The solenoid is about $13\ \mathrm{m}$ long and has an inner diameter of $5.9\ \mathrm{m}$ producing a magnetic field of up to 4 Tesla in the inner region. The return field is strong enough to saturate the magnetisation of the $1.5\ \mathrm{m}$ iron of the return yoke, which is distributed radially. The width of the yoke ranges from $\approx 3.5\ \mathrm{m}$ to $7\ \mathrm{m}$ and is interspersed with muon chambers. In total there will be 2.7 GJ of energy stored in the magnetic field. This high magnetic field strength is needed to gain sufficient momentum resolution for high energetic particles within the compact volume of CMS. The main feature of the solenoid is the use of a high-purity aluminium-stabilised NbTi conductor maintained at liquid helium temperatures ($\sim 4\ \mathrm{K}$). The yoke has to keep the return flux of the magnet and additionally ensure the stability of the system holding the magnetic pressure of $P = \frac{B_0^2}{2\mu_0} = 6.4\,\mathrm{MPa}$. It is also part of the quench protection system, as the induced eddy currents allow to lead away the energy stored in the magnetic field in case of a quench of the magnet.

3.3.6 Muon Chambers

The muon chambers are situated at the very outside region of the detector, because of all particles observed by the CMS detector, muons have the weakest interaction with matter. Four muon stations of the barrel region are integrated into the return yoke of the magnet, benefiting from the returned magnetic field for momentum measurements.

Three types of gaseous detectors are used to identify and measure muons with the precision needed for instance to identify the Higgs decaying into 4 muons, a golden channel for this search. The choice of the detector technologies has been driven by the very large surface to be covered, and by the qualities of the radiation environment, in which the specific compounds of the muon chambers will have to be run. In the barrel region ($|\eta| < 1.2$), where the neutron induced background is low, drift tube (DT) chambers are used. Cathode strip chambers (CSC) are deployed in the two endcaps, since the muon rate as well as the neutron induced background are high, and a strong magnetic field is present. These strip chambers cover the region up to $|\eta| < 2.4$. Additional resistive plate chambers (RPC) are used both in the barrel and the endcap regions. These RPCs are operated in avalanche mode to ensure good operation at high rates. RPCs provide fast response with good time resolution but with coarser position resolution than the DTs and CSCs. RPCs can therefore be used to identify unambiguously the correct bunch crossing. Making use of these subsystems the muon system provides the Level-1 trigger system (Sec. 3.3.7) with two independent and complementary sources of information in order to reduce event rates in a most sensible manner.

3.3.7 Data Acquisition and Trigger

At design luminosity, the proton bunches will cross 40 million times each second leading to an interaction rate of $\sim 10^9$ Hz. Since only a data rate of 300 Hz can be written to archival material for pp runs [78] a sophisticated trigger system has to be implemented. The CMS trigger and data acquisition system consists of 4 parts: the detector electronics, the Level-1 trigger processors (calorimeter, muon and global), the readout network, and the online event filter system, which is actually a computer farm that executes the software for the High-Level Triggers (HLT). The Level-1 trigger electronics mainly consist of custom Application Specific Integrated Circuits (ASICs), semi-custom and gate-array ASICs, Field Programmable Gate Arrays (FPGAs), and Programmable Logic Devices (PLDs). During the $\approx 3.2\mu$s needed by the Level-1 trigger electronics to decide whether an event is kept, the full event information is buffered by the electronics of several devices. After a decision is made, the event information is compressed and finally passed through to the HLT at a rate of $\sim$100 kHz. More sophisticated and flexible algorithms can be used to reduce the event rate to a final value of 300 Hz, which is then written to the storage system. With an expected raw, zero-suppressed, event size of about 1.5 MB on average, the system output will sum up to large amounts of data to be stored and processed.

3.4 Worldwide LHC Computing Grid (WLCG)

Proton-proton collisions at the TeV scale result in about 1000-2000 particles per event. In addition, the need for maximal sensitivity for new physics increases the aimed at resolution

and event rate of the detectors at modern high energy physics experiments. Taking these points into account finally leads to large data rates to be handled, stored and processed in order to exploit their physical content.

At the LHC there will be 40 million pp bunch crossings per second resulting in about 5 hard interactions happening simultaneously and leading to an event size of about 1.5 MB in case of the CMS detector. This would mean to write 60 TB of data to the storage system each second, which is neither a possible nor a sensible thing to do within the given tasks and abilities of the collaboration. Even a sophisticated trigger system (see Section 3.3.7), which reduces the written event rates to the required 300 Hz, leads to a final data stream of 450 MB$/$s. Such data streams are about 4 times of what a usual 1 TB SATA disk ($\approx 120\,$MB$/$s) is able to handle, and assuming it had the required capabilities it would be filled within 39 minutes, which would sum up to 1100 new disks needed each month. Further storage requirements arise as hardware failures cannot be safely excluded, the experimental results have to be stored in a redundant manner to reduce the risk of losing data, and multiple copies need to be held for effective processing.

Finally, all data have to be accessible in order to be analysed repeatedly by about 3600 CMS members from all around the world. With the requirement of these physicists to run CPU and input demanding analysis code, it becomes obvious, that, for the analysis of real data, as well as for production, storage, and analysis of simulated events, a dedicated computing infrastructure has to be implemented.

The LHC experiments decided to cope with these challenges via Grid technologies, which is a special way of using decentralised worldwide computing and mass storage resources and making them accessible from all over the globe.

3.4.1 Grid Computing

The main idea of Grid computing was spread around the world with Ian Foster and Carl Kesselman publishing their book “The Grid - Blueprint for a New Computing Infrastructure” in 1998 [82]. Their vision was to provide distributed resources for transparent public use based on a standardised interface. With respect to the everyday analysis of a particle physicist this would mean using large amounts of computing and mass storage resources as well as data spread over multiple computing centres just as if it would be hosted on one single computer.

As it became obvious for LHC experiments, that only such a wide spread of computing and storage resources would be affordable, the idea of Grid computing was taken over to enable physicists to cope with the described computing needs in a most convenient and efficient way.

3.4.2 WLCG Structure

Having these ideas in mind, the WLCG (**W**orldwide **L**HC **C**omputing **G**rid) [83] was planned in a multi-tiered architecture (Figure 3.4) mainly driven by the need of data distribution, which leads to a natural assignment of computing tasks for the Tiers. The Tier 0 centre at CERN receives the data output directly from the Data Acquisition (DAQ) System (see chap. 3.3.7) of the experiments and distributes them to Tier 1 centres worldwide. These Tier 1 centres usually have large computing power and huge mass storage systems including tape systems to be able to do standard event reconstruction, store the data output of the experiments, and keep custodial copies of dedicated samples for backup purposes. Several Tier 2 centres are associated to each Tier 1 centre. The main focus of Tier 2 centres are user analyses and the production of simulated data. They need a fast broadband network connection to their Tier 1 centre to make the simulated events available at the Tier 1 site and to receive the specialised datasets for ongoing analysis but without the need to store them over a longer period. Tier 0, 1 and 2 centres are planned to have a 24/7 reliability in providing their service. In addition small computing clusters or Tier 3 centres maintained by university groups can provide services to the Grid, like it is implemented at the IEKP [84]. They are mainly foreseen for the needs of local users and for educational purposes and do not require a service reliability around the clock. Although one can establish dedicated links to special Tier 2 centres, Tier 3 centres are mainly used in a more flexible way by users submitting jobs from them and retrieving their final output to their local storage elements.

All Grid centres are steadily improved and new centres are constantly being attached to the Grid to reach the desired amount of computing and storage resources. The structure of the dedicated links in Figure 3.4 only resemble a minimum configuration of the CMS part of the WLCG. Of course specific links can be established between all displayed centres, following computing needs and facility capabilities. Especially Tier 2 and Tier 3 resources are interesting for MC production jobs, which do not need datasets to be present at the centres. The WLCG is currently used for processing of the data constantly taken from the measurements of cosmic particles, as well as Monte Carlo production and analysis on a regular basis. Particular tests and monitoring efforts are in place to keep an overview of the Grid status and availability.

3.4.3 Grid Services

Having all those computing centres with large computer farms at disposal, software services are needed to enable users to access the data and to use the computing resources. For this purpose, several publicly funded projects were founded in the past to develop software for the WLCG. The main idea of these software projects is to build a so called “middleware” system, providing the user with Grid access without the need to know and actively control all underlying services. This middleware makes Grid services work just

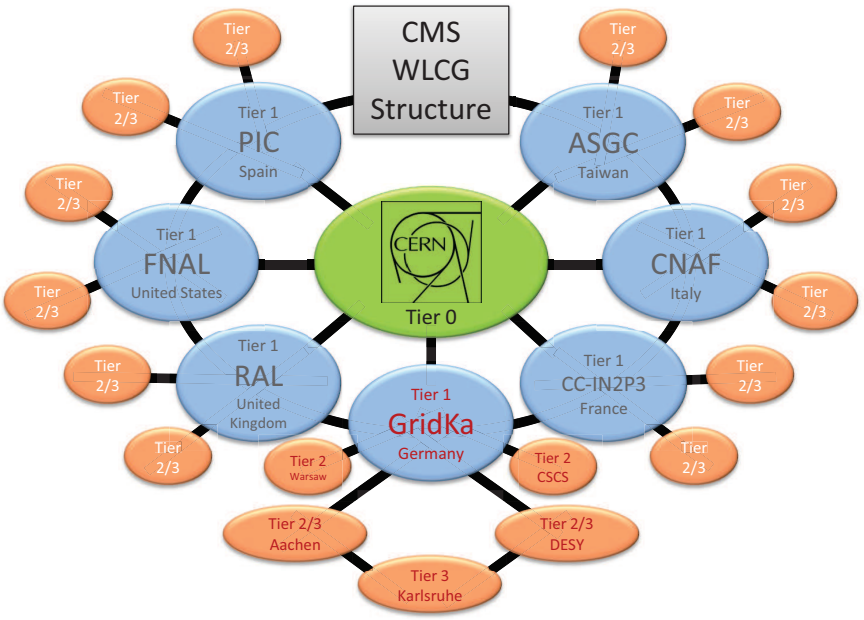

Figure 3.4: *Tier structure of the CMS part of the Worldwide LHC Computing Grid[85]. Tier 1 centres are directly linked to the Tier 0 at CERN, while dedicated connections are established from Tier 1 centres to their local Tier 2 centres with attached Tier 3 centres. The structure has a dedicated focus on the situation of the German centres.*

like the TCP/IP protocols[2], which do not need to be aware whether they are currently being delivered via telephone line, wireless network, satellite or ethernet connections. Special requirements for such a Grid environment are security concerns, heterogeneous hardware, and assorted software being used on different Tier centres. Special hardware and software requirements bind Tier centres to run Scientific Linux [86] on x86 hardware, which reduces respective complications to a minimum. In contrast to that security is a big task as one has to accomplish two things which contradict each other: Grid users should have access granted to all Grid sites and resources needed in a most uncomplicated way, but the large number of computing nodes and their broadband Internet access must not be opened to abuse.

Currently several middleware softwares are used, which in principle are compatible. While the OSG (**O**pen **S**cience **G**RID [87]) is mainly used in the USA, most European sites make use of gLite, developed by EGEE (**E**nabling **G**rids for **E**-scienc**E**) [88].

The middleware has to provide services which have to be run either on every Grid site or just at specific locations. Because WLCG is used by all the LHC experiments, there is a special need to personalise Grid access and services in order to ensure a reasonable splitting of the funding of the computing resources. For this purpose each user is individually registered and attributed to a certain virtual organisation (VO) which controls computing resources. These VOs are for example ATLAS or CMS. The personalised certificates follow the X.509 standard and enable administrators to unambiguously identify each individual user. The specific Grid services provided by the middleware are described in more detail in the Appendix B.2.

3.4.4 Grid Usage

Analysis software is usually developed at local computing clusters or user desktops. Within CMS this software commonly consists of scripts written in bash, Python, tcsh, or other script languages combining several user analysis programs written in FORTRAN or C++ using software frameworks like ROOT or the experiment software (e.g. CMSSW). After the software is developed and tested locally, it can finally be sent to the Grid. For this purpose a Grid job has to be built. In case of gLite, this mainly consists of a `.jdl`-file describing the job using Job Description Language (JDL). In this file the following information has to be provided:

- The location of the local files, which have to be sent to the remote site via “Input Sandbox”. This mostly contains analysis programs, scripts and configuration files.
- Information on required hardware and software, which have to be provided at the target Grid site, or additional requirements, for example the specific target site. This is evaluated by the so called “Workload Management Service (WMS)”.

[2]The TCP/IP protocol suite is of major importance to the Internet. It is named after the Transmission Control Protocol (TCP) and the Internet Protocol (IP), which were the first protocols included.

- Location of files produced by the job on the Grid site, which have to be transferred back via "Output Sandbox".

When the job is submitted, the software of the User Interface builds the Input Sandbox out of the files listed in the `.jdl`-file and sends them to the WMS, which then passes it on to a specific site fulfilling all requirements. Finally arrived at a worker node of the target site, the Input Sandbox is extracted and the script or binary specified by the job description is executed in a traditional Linux bash or tcsh environment. When the job is finished, the output files specified in the `.jdl`-file are sent back via Output Sandbox and can be retrieved by the user at his User Interface. In addition to that, files stored at Storage Elements are used to either transfer programs to the Grid sites or retrieve output. This also reduces the load of the WMS, because the sandboxes need to be transferred via the WMS and are also stored there until the job is finished. This is why sandbox sizes are usually limited to $10\,\mathrm{MB}$ by the WMS configuration.

In order to achieve suitable statistical precision, physics analysis usually have to be run on millions of events with a time consumption from seconds up to minutes per event. On the one hand this is a result of the specific analysis, but on the other hand peculiarities of the particular Grid site, on which the job is executed, also have strong influence on the actual duration of the job processing. This is why usually about $10-1000$ jobs are used to process one dataset. With these numbers of jobs it is not a trivial task to keep track of sending and retrieving all jobs, especially when it comes to resubmitting failed ones. For this task self made scripts can be developed. A more advanced example of such a script is grid-control [89], which condenses the gathered Grid experience of our working group into ≈ 5000 lines of Python code. It is strongly modularised and capable of handling user defined as well as CMSSW framework (Section 3.5.5) grid jobs and has interfaces to multiple batch systems. The official CMS tool to submit framework jobs to the grid is CRAB [90]. For large scale data processing additional CMS specific tools have been developed.

3.4.5 CMS Data Management

The available Grid models provide services for data management in form of storage elements and the Replica Location Service. They allow a setup of registering files in global VO-wide databases and are identified with the global identification GUID. However these services never really scaled to the extend of the CMS requirements. Therefore, a number of CMS-specific distributed computing services operate above the generic Grid layer. The datasets are accessible using Grid file protocols like srm or Gridftp and stored on standard Grid storage elements. However, the RLS service is replaced by the Dataset Bookkeeping System (DBS [91]), where all datasets are registered in a central database together with metadata, like sizes, content, or the sites where they can be accessed. The data management itself does not handle single files, but datasets, which are subdivided into blocks of files, with an aimed at size of $\mathcal{O}(1\mathrm{TB})$. Only large block sizes enable a proper handling

in case of data transfer and storage on tape systems. Datasets are grouped in physical boundaries, like detector runs or a usual Monte Carlo production sizes ($\mathcal{O}(10\text{k}) - \mathcal{O}(10\text{M})$ events).

DBS queries can be conducted via web-page or via Python-API call to the Data Location Service for automated processing. The files contained in a dataset are returned starting from a relative path which has to be resolved to an absolute physical path within the storage system of the side. For this purpose each job performs a look-up in the Local File Catalogue of the respective grid site. This task is implemented into the CMS software package CMSSW, to be described in the following (Section 3.5.5). In order to transfer and register datasets the specialised service PhEDEx [92] consisting of several agents running at the participating sites is set up. These agents are mainly software daemons and steer the transfer of the data. They are usually installed and maintained by local CMS teams at each site. This system also allows specialised routing over several sites, in case certain Internet or Grid problems make this necessary. The system manages the tasks of scheduling, monitoring and verifying the movement of data in conjunction with the storage interfaces at the CMS sites, ensuring an optimal use of the available bandwidth. The baseline mode of operation for the data management system is that the collaboration will explicitly place datasets at defined sites, where they will remain for access by CMS applications until removed.

3.5 Datastructure and Software

With current experiments in high energy physics having become more and more complex, a broad variety of tools has been developed to exploit the huge amount of valuable data acquired with modern particle detectors like CMS. Tools are required to efficiently store and process the data, as well as a framework which allows hundreds of collaborators to work together and develop their own analysis. Therefore, most tools are developed using the object oriented programming language C++ and open source tools to build and handle code (gcc, CVS, subversion, ...) and provide additional algorithms for data analysis and statistical evaluation, like ROOT. The usual analysis handles these tools and the corresponding workflows using script languages like bash, tcsh, Perl or Python[93].

Including simulated events, CMS is expected to produce about 20 petabytes of data per year, which will have to be analysed repeatedly within the iterative process of improving the understanding of the detector and the measured physics. New simulations, with updated detector conditions, improved algorithms, and LHC machine expectations are steadily conducted to be best prepared for physics data after first collisions at the LHC and to further improve reconstruction algorithms and the understanding of the detector response during the running of the experiment.

A dedicated data structure to store the accumulated information is required as well as customised software, which at CMS steers the high level trigger farm, enables event reconstruction and data analysis. This is included in the software package CMSSW to be described later.

3.5.1 Event Data Model

The central concept of the CMS data model is the event. An event comprises the recorded raw data and conditions from a single triggered bunch crossing, and the new data derived from it by applying reconstruction algorithms. The same information is derived in the simulation case using MC technologies. In this case additional MC truth information is saved. The model also requires additional provenance information, storing the processing history of the objects contained in the event file, including physically relevant parameter settings during execution.

3.5.2 ROOT

The analysis strategy of CMS makes excessive use of ROOT tools for data storage and processing. ROOT succeeded PAW [94], a FORTRAN based program, as an interactive data analysis software for high energy physics. The ROOT system provides a set of object oriented frameworks written in C++ with all the functionalities needed to handle and analyse large amounts of data in a very efficient way.

Having the data defined as a set of objects, specialised storage methods are used to get direct access to the separate attributes of the selected objects, without having to touch the bulk of the data. Included are various histogramming methods, curve fitting, function evaluation, minimisation, graphics and visualisation classes to allow to set up an analysis system that can query and process the data interactively or in batch mode. In addition, the ROOT framework can be included into every user analysis code based on the C++ programming language.

3.5.3 matplotlib

The Python package matplotlib [95] is a 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. It can be used in Python scripts, the Python and ipython shell (ala matlab or mathematica), web application servers, and six graphical user interface toolkits. This convenient tool has been used to visualise the majority of the figures in Section 4.2 and Chapter 5.

3.5.4 SCRAM

All CMS related software projects are strongly modularised and have complicated relations and dependencies between each other. This software structure is necessary to enable a flexible use and development within a large group of geographically dispersed collaborators both in offline and online context. The **S**oftware **C**onfiguration, **R**elease **A**nd **M**anagement software SCRAM [96] was developed in order to enable people to work together in such a framework. It is a build and runtime environment, which eases to maintain the hundreds of libraries and plugins belonging to the CMS software framework. The advantage of using shared object libraries and plugins is on one hand side the simple modular exchange of parts of the software, without the need to rebuild everything, and the reduced memory consumption by only loading required modules during runtime. The usual collaborator is required to have multiple revisions of the software in parallel use, in these cases SCRAM eases to switch between multiple versions.

3.5.5 The CMSSW Framework

Making use of the previously introduced tools SCRAM and ROOT, the basis of the CMSSW framework is developed implementing the CMS Event Data Model. It is designed for event reconstruction and data analysis. The running of the modules is steered by configuration files written in the object oriented script language Python[93]. The framework has an interface to multi-purpose Monte Carlo event generators and includes a simulation part (Sec. 3.6.3), which uses GEANT4 [97, 98] to simulate particles passing through matter. For this purpose, the full simulation of CMS detector and its response is included. Also

a simplified fast simulation is available, which produces final events in seconds, while the full simulation takes several minutes, depending on the simulated detector occupancy.

The running of the framework starts from a given "input source". This source contains event information from data streams directly obtained from the DAQ, or read from stored framework files. Additional information is added about the event setup, which can be retrieved from database(s). Such databases store the detector status, e.g. status of detector modules (broken, switched on/off) or information about Monte Carlo events. This information can then be used by subsequently executed framework modules, which can also produce additional data to be utilised by successive modules. The order and the configuration parameters for each module is set in a configuration file. All modules can read information available in the event. Dedicated producer modules can add information in form of framework objects, mostly containing reconstructed physical quantities, to the event. Filter modules are available and are mainly used for the high level trigger. Analysing the available event content, they return a boolean result, which can be used to stop a sequence of modules configured to run consecutively at a certain point of unfulfilled conditions. Analyser modules are dedicated to user analysis and thus produce summary information from an event collection, like histograms or user defined data streams. Finally the output can be written to disk, possibly split into different streams of ROOT files containing different parts of the event content, as configured by the respective output module. External packages, like reconstruction algorithms as e.g. fastJet [62], or the simulation package GEANT4, are linked into the framework and used by dedicated modules. This way event generation and Monte Carlo simulation can be naturally included into the workflow.

3.5.6 Datatiers

In order to reduce the data contained in the files, different data tiers are defined for dedicated steps of the data workflow.

RAW format

RAW events contain the full recorded information from the detector, plus a record of the trigger decision and other metadata. An extension of the RAW data format is used to store the output of CMS Monte Carlo simulation tools. The RAW data are expected to be of a size of $1.5\,\mathrm{MB/event}$ for detector data, and $2.0\,\mathrm{MB/event}$ for simulated data.

RECO format

Reconstructed (RECO) data are produced applying several levels of pattern recognition and analysis algorithms to the RAW data, producing high level physics objects like tracks of charged particles or jets of hadrons. These objects are later used for the physics analysis. The RECO format will only contain as much RAW information as required to reprocess

the high level objects with new detector conditions or calibration constants. With the bulk of RAW data removed and RECO objects added, this data tier is expected to occupy around $0.5\,\mathrm{MB/event}$.

AOD format

AOD (**A**nalysis **O**bject **D**ata) is a compact analysis format, solely containing the highest level objects needed for analysis filtered out of dedicated RECO streams. This data format is planned to be the main analysis target of the physicist, accessed on dedicated Tier 2 centres. The format is planned to occupy about $0.1\,\mathrm{MB/event}$.

Non-event data

In order to best analyse the event data, additional non-event information is required. This contains construction data, generated during the construction of the detector, equipment management data, configuration data, comprising programmable parameters related to detector operation, and conditions data, including calibration, alignment and detector status information. The latter is most important when producing higher level objects for particle physical interpretation. Such conditions are required and produced by both offline and online applications and have a well-defined **I**nterval **O**f **V**alidity (IOV), as calibration and alignment constants might vary or detector parts might even stop working during data taking.

The events and conditions are recorded in subsections determined by machine and detector parameters. After the protons have been injected into the LHC and accelerated to the desired energies, collisions are produced until the luminosity falls below useable values, which is mainly due to beam loss from collisions.

The period during which measurements are possible is usually referred to as one “fill”, which is expected to last between 5 and 12 hours, depending on how fast the next beam can be filled (estimates vary between one and seven hours). The actual data taking period of the experiment is called a “run”, which ideally is as long as one fill, but one fill could be split into multiple runs in case of problems with the detector. As the luminosity constantly falls during one run, runs are usually subdivided into so called “luminosity sections” (LS), where the instantaneous luminosity can be regarded as constant. The duration is counted in multiples of beam orbits, and expected to be between 1 and 5 minutes of data taking. This is a reasonable “human scale” period, in which also operators can react on changes and document their findings. All condition data stored are synchronised with the LS boundaries. If the boundaries of a LS cannot be clearly defined, the LS is rejected.

The non-event data are held in a number of central Oracle databases for access by offline and online applications. In order not to overload the database, the conditions are replicated to the offline sites by the Frontier [99] system, a distributed network of caching http proxy

servers. Thus analysis jobs on the Grid sites access the conditions by reading from the local proxy reducing required network bandwidth and load of the central database.

3.6 Monte Carlo Methods

Monte Carlo (MC) methods rely on repeated random sampling during computation and are especially useful for problems with a large number of coupled degrees of freedom, like modelling and calculation of most realistic particle interactions or simulating the response of a complicated apparatus like the CMS detector on an event by event basis. Additionally these methods profit from the fast increase of affordable computing resources, especially due to the fact, that single events are independent from each other. So the huge number of events can easily be split into multiple jobs working on few events each. So on one hand, MC methods play an important role in the planning phase of the experiment or a specific analysis in order to estimate the expected performance. On the other hand it is an indispensable tool to determine the detector resolution necessary for an unfolding procedure of detector effects. In case of searches for new physics, it is absolutely necessary to know how the aimed at signature looks like in the detector, in order to configure triggers and implement analysis cuts to reduce the huge background and extract signal-like events for analysis.

The simulation chain is usually started with a description of the high energetic collision and the subsequent particle production using a Monte Carlo event generator. The theoretical basis of this has been outline previously (Sec. 2.1). As far as the simulation of the detector response is concerned, solely event generation methods are suitable which produce final state particles like hadrons, leptons, or photons. Only such final states reflect the observations within the real detector. For the purpose of event generation various software packages can be used, which usually solve the task by adding parton shower and hadronisation models to initial fixed order calculations in perturbative QCD (Sec. 2.5). The output of such generators is used as input for the detector simulation within CMSSW (Sec. 3.5.5), where it is later passed to a dedicated module based on the simulation package GEANT4 (Sec. 3.6.3).

3.6.1 Multi-Purpose Event Generators: PYTHIA and HERWIG(++)

Two multi-purpose generators were considered within this work. In both cases leading order (LO) calculations are used to derive coloured partons from the hard interaction. Starting with these partons, the event generation packages produce colourless objects like hadrons, leptons and photons. A perfect detector would exactly measure the properties of these particles at a real experiment.

The by far predominant part of currently available simulated events, which have been subjected to the full CMS detector simulation, were initially produced with PYTHIA [34].

This program for high energy physics event generation has been written in FORTRAN and is maintained by the Lund University theory group since 1978. It is designed to describe multi-particle production in collisions between elementary particles as detailed as experimentally observable. Due to the used MC-methods this also includes statistical fluctuations. A specially tuned parton shower calculation is used to generate Bremsstrahlung corrections and higher order corrections. The result is thereafter left to hadronise using the Lund/String hadronisation model (Sec. 2.5.2). This simulation is a highly non-trivial process and is strongly tuned to describe measurements of past detectors. To allow for a realistic description of the processes, several models are included for distinct physical aspects. Short distance interactions of quarks, leptons and gauge bosons can be described as well as initial and final state radiation, multiple interactions and beam remnants.

An alternative to PYTHIA is HERWIG (**H**adron **E**mission **R**eactions **W**ith **I**nterfering **G**luons) [35], which also uses LO calculations with added parton shower and hadronisation models. It has own tunes and an implementation of a parton shower which uses a cluster hadronisation model (Sec. 2.5.2). As the theoretical basis describing hadronisation is rather limited, a possibility to exctract uncertainties is comparing PYTHIA with HERWIG outputs. Recently a C++ version of HERWIG has been released under the name HERWIG++ [36]. It has a new implementation and tune of the underlying event model JIMMY [100], which can be combined with the FORTRAN HERWIG. Since 2008 HERWIG++ is used by the CMS MC production team, and significant numbers of HERWIG++ events are available with full detector simulation.

3.6.2 Signal Generators and Alternative Approaches

Other available generators mainly focus on special processes only and need parton shower and hadronisation models of the multi purpose generators to be attached. One example for additional processes added to PYTHIA is TopReX [101]. Another interesting task is attaching parton shower calculations to QCD-calculations which lead to higher jet multiplicities. Examples for this are SHERPA, MC@NLO, MADGRAPH, or ALPGEN (see chap. 2.5.1). The latter has multi-jet matrix elements implemented and is currently used to produce samples of simulated events for the CMS detector, to achieve better understanding of physics with multiple hard jets, which also contribute to signals for the Higgs search. Special methods exist to match parton showers to these multi matrix element calculations, which allows these generators to be used together with multi purpose MC generators leading to hadronised final states. These matching procedures remove double-counted radiations, which could be produced during parton shower as well as already being included into the matrix elements. Unfortunately, such methods currently only exist for LO perturbative matrix elements, NLO approaches (MC@NLO or POWHEG) are only available for some processes.

3.6.3 Detector Simulation

The detector simulation of CMSSW is based on the simulation package GEANT4 [97, 98]. GEANT4 is a C++ toolkit for simulating the passage of particles through matter. It follows the FORTRAN based GEANT3 and includes a complete range of functionality including tracking, geometry, physics models and hits. The physics processes offered cover a comprehensive range, including electromagnetic, hadronic and optical processes, a large set of long-lived particles, materials and elements, over a wide energy range starting, in some cases, from 250 eV and extending in others to the TeV energy range. It has been designed and constructed to expose the physics models utilised, to handle complex geometries, and to enable its easy adaptation for optimal use in different sets of applications. The toolkit is the result of a worldwide collaboration of physicists and software engineers. It has been created exploiting software engineering and object-oriented technology and implemented in the C++ programming language. It is widely used in applications in particle physics, nuclear physics, accelerator design, space engineering and medical physics.

3.6.4 NLOJET++

NLOJET++ [102–104] is a QCD event generator for hadron-hadron collisions developed by Zoltán Nagy, which can calculate one-, two-, and three-jet observables at next-to-leading order. In case of the three-jet or inclusive jet cross section this extremely reduces the renormalisation and factorisation scale dependence with respect to a leading order calculation. This additional precision is exploited for the inclusive jet cross section within this work (Sec. 4.1). A slightly modified Catani-Seymor [105] dipole formalism is used in the calculation to cancel infrared divergences, allowing maximal precision and flexibility during phase space generation. Although Monte Carlo integration is also a vital part of this program, individual events suited for detector simulation can not be produced by this. When using the results for comparison to measured data, also additional corrections need to be applied, which is focused in section 4.2.

3.6.5 fastNLO

Since precise computations in NLO, e.g. using the previously described NLOJET++, are very time consuming, a more efficient set-up in form of the fastNLO project [106, 107] has been setup. It allows the fast re-derivation of the considered cross section for arbitrary input parton distribution functions and α_S values. This is done by separating the PDF dependency from the hard matrix element calculation by interpolating the PDFs between fixed support points in fractional proton momentum x so that the PDF dependency can be evaluated a posteriori from one complete calculation. The initial calculation with a first PDF has still to be done in the common way. The fastNLO package is attached to NLOJET++, which performs the initial perturbative calculation in next-to-leading order. More

details on the parametrisation uncertainty, physics output and computing performance is given in section 4.1 and 5.8.

Chapter 4

Prediction of the Inclusive Jet Cross Section

A fundamental aim of experimental research is the comparison of new data with the best theory predictions at disposal at a given moment in time. In case of the inclusive jet cross section the best current prescription is available to us in the form of next-to-leading order calculations in QCD. These calculations require as input the strong coupling constant α_S and in case of pp-collisions also parton distribution functions. In this chapter the dominant uncertainties of the theoretical description are discussed. Additionally a method to derive non-perturbative corrections to the NLO result is presented. Such corrections are required to make a comparison to measurements, which can only be done at the hadronised final state. The binning and setup for the calculations is chosen dedicated to the presented measurement plans in Chapter 5, and is therefore driven by the properties of the CMS detector.

4.1 Calculations in NLO perturbative QCD

For one computation of jet cross sections in perturbative QCD to next-to-leading order the program NLOJET++ [102] is employed which would be completely sufficient for a simple comparison to data. Unfortunately, this type of calculations is rather time consuming even with recent processor technologies. For example, a NLO integration with 10^8 events takes about five hours of CPU time for the k_T algorithm, if the fast implementation of [60] is used in the inclusive mode with $D = 0.6$. The SISCone algorithm ($R = 0.7$) takes twice as long. For the inclusive jet scenario described in this section the computation has been set up according to Table 4.1 with a total of 30 billions of LO as well as NLO events for the k_T and SISCone algorithm requiring about 1650 resp. 3840 hours of CPU time. By submitting multiple jobs to the Grid the task has been parallelised which in addition provides a handle to estimate the statistical uncertainty of the Monte Carlo integrations (see sec. 4.1.1).

If the calculation is repeated with different conditions like another set of parton distribution functions as available in the LHAPDF package [108, 109], the full computing time has to be invested again. Fortunately, a more efficient set-up in the form of the fastNLO project [106, 107] is available. There, the very time consuming step of a precise jet cross section computation in NLO is done only twice. At the expense of interpolating the PDFs between fixed support points in fractional proton momentum x the PDF dependency can be separated from the hard matrix element calculation and a re-derivation of the considered cross section for arbitrary input PDFs and α_S values can be done quickly (microseconds). The second calculation labelled as *reference* in Table 4.1 is required only in order to check the quality of the approximation which can be improved upon if necessary.

To derive for example the PDF uncertainty from fits providing error PDFs like in CTEQ-6.5 [110] this method is extremely helpful. In order to include jet data into global PDF fits like in MSTW2008 [66] it is indispensable.

In the next section, a detailed account of the uncertainties intrinsic to the Monte Carlo integration of NLOJET++ and the approximation in fastNLO is presented. Following that, a short survey is given of the final output of a computation and how it can be used to derive the uncertainties due to the finite order in perturbative QCD, the measured precision of the strong coupling α_S and the global PDF fits which are all described in the subsequent sections. More details on the technical usage of the derived tables is given in the Appendix B.3. The last section deals with the so-called k factors and comparisons to the LO Monte Carlo generator PYTHIA 6.4 [111].

4.1.1 Intrinsic Uncertainties

In order to accelerate by several orders of magnitude the **repeated** derivation of NLO jet cross sections with for example different PDFs, it is avoided to redo the time consuming matrix element calculation contained in NLOJET++ by storing all relevant information in a table with bins in the fractional hadron momenta x_i and x_j of the interacting partons

order	LO		NLO	
jet algorithm	k_T 0.6	SISCone 0.7	k_T 0.6	SISCone 0.7
# jobs	30		300	
# events/job	10^9		10^8	
# reference jobs	20		100	
# events/ref. job	10^9		10^8	
primary raw table size	0.44 MB (gzipped)		2.5 MB (gzipped)	
ref. raw table size	0.94 MB (gzipped)		3.1 MB (gzipped)	
final table size	—		6.8 MB (gzipped)	
final table size	—		9.8 MB (gzipped)	
total CPU time/h ($\approx$)	$150 + 200$	$240 + 130$	$1.5\,\mathrm{k} + 1.0\,\mathrm{k}$	$3.6\,\mathrm{k} + 0.9\,\mathrm{k}$

Table 4.1: *Setup of the primary and reference calculations required for the inclusive jet cross section to NLO for both, the k_T and SISCone algorithm.*

i and j. In the setup for the inclusive jet cross sections described here, twelve bins for each fractional momentum $x_{i,j}$, equidistant in $\sqrt{\lg(1/x)}$, have been used. To replace the PDF weights a bi-cubic interpolation (linear at the edges) has been implemented, where in addition a global re-weighting function of the type

$$w(x) = \frac{(1 - 0.99x)^3}{x^{3/2}} \tag{4.1}$$

was employed to dampen strong variations at small and large x. For more general details see [106].

Since fastNLO in comparison to the original NLO code contains an additional approximation, the most important check to be performed is to estimate the quality of this approximation with respect to the original results. This can be done with special reference jobs where for identical events the approximation table as well as the full result including PDF weights and factors of α_S are stored at the same time. The difference between the re-derivation of the cross sections according to the fastNLO scheme and the full calculation in each bin is then taken as the algorithmic uncertainty. Of course, this uncertainty should be kept so small that it is negligible for all practical purposes. Here it is ensured that it is smaller than the statistical uncertainty that is always associated with programs using Monte Carlo integration techniques such as NLOJET++. In Figures 4.1 and 4.2 both uncertainties are presented for the NLO calculations as given in Table 4.1. In both cases the algorithmic uncertainty is below ≈ 2.5 per mille whereas the statistical one is mostly below one per cent. If a higher precision is needed both uncertainties can easily be reduced by increasing the number of bins in x and/or increasing statistics at the expense of more computing time and larger table sizes.

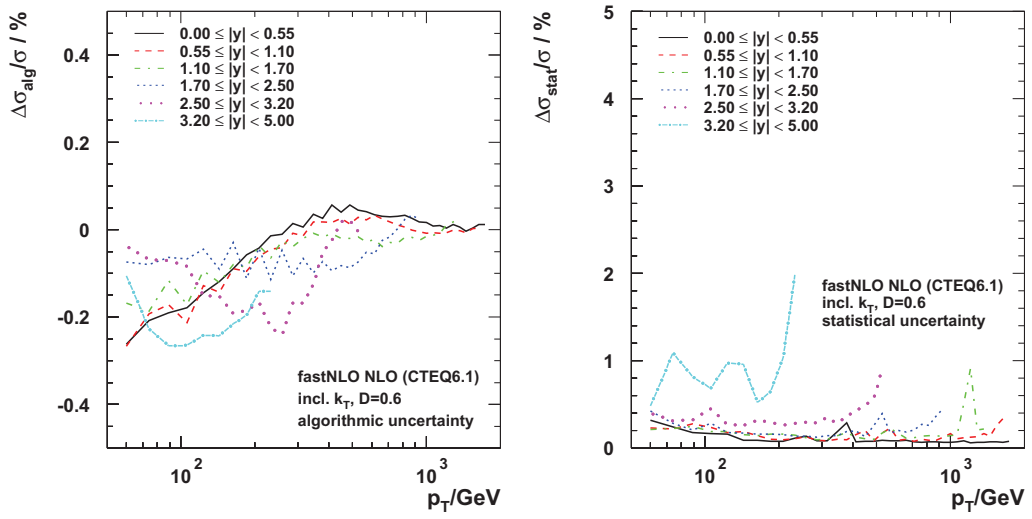

Figure 4.1: *Relative algorithmic and statistical uncertainties of the NLO inclusive jet cross section in per cent for all bins in rapidity from* fastNLO *for the k_T algorithm.*

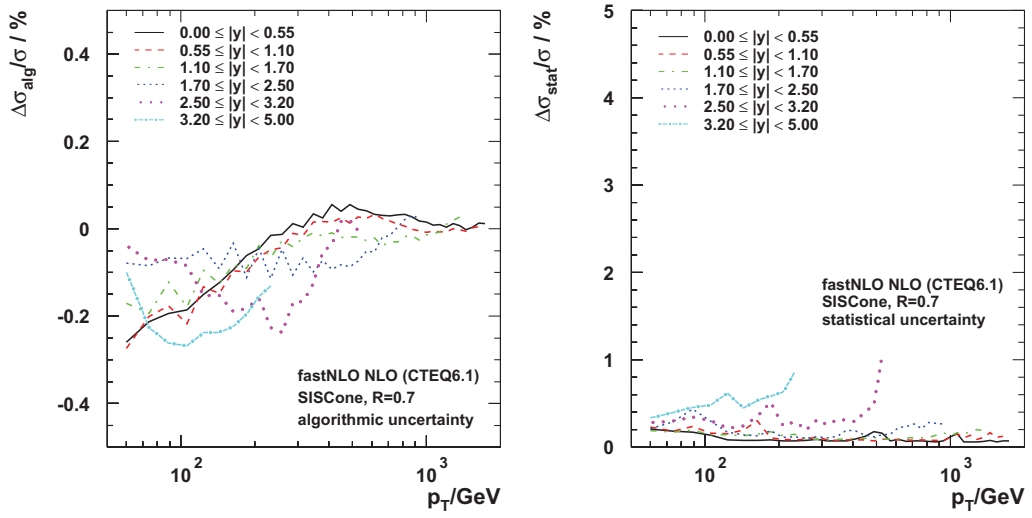

Figure 4.2: *Relative algorithmic and statistical uncertainties of the NLO inclusive jet cross section in per cent for all bins in rapidity from* fastNLO *for the SISCone algorithm.*

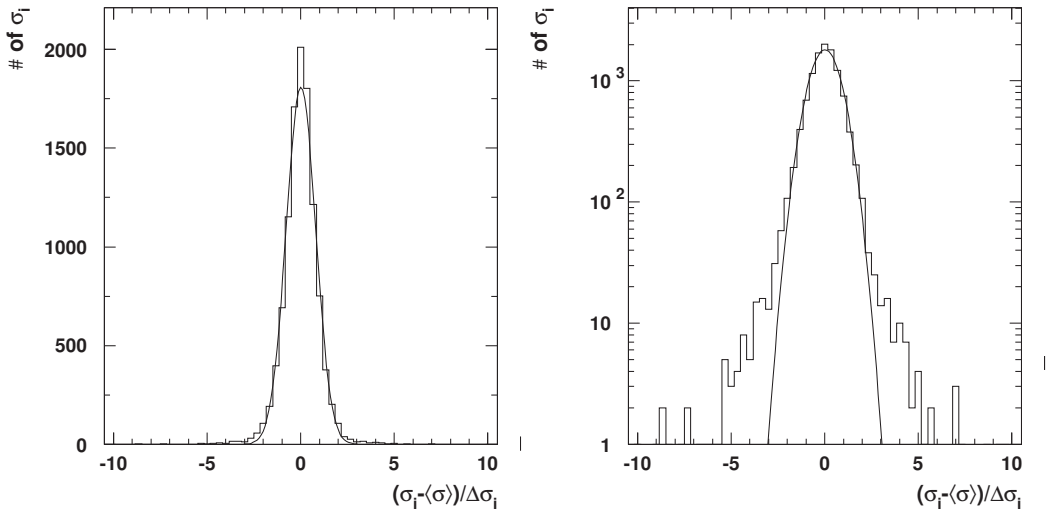

Figure 4.3: *Distribution of individual NLO inclusive jet cross section calculations σ_i around their means $\langle\sigma\rangle$ for all p_{T} bins at central rapidity in linear (left) and logarithmic scale (right). All differences have been scaled by their corresponding standard deviation s_i. Additionally, a Gaussian fit is shown.*

In Figures 4.3 and 4.4 examples are shown for the distribution of the individual cross section evaluations σ_i around their means $\langle\sigma\rangle$. Since all differences have been scaled by the corresponding standard deviations s_i, the histograms are centred with means of 0 and root-mean-square (RMS) deviations of one. Especially from the logarithmic plots it is obvious that Gaussian fits do not describe the tails. Hence, the assumption of Gaussian uncertainties in using their widths of 0.78 resp. 0.62 in Figures 4.3 and 4.4 would lead to an underestimation of the statistical uncertainty. Therefore statistical uncertainties of the NLO cross sections with NLOJET++ and fastNLO are derived from the spread of a large number of statistically independent calculations as

$$\sigma = \frac{1}{n}\sum_{i=1}^{n}\sigma_i \pm \frac{\mathrm{RMS}(\sigma_i)}{\sqrt{n-1}}. \tag{4.2}$$

4.1.2 Binning in Rapidity and Jet Transverse Momentum

For every NLO calculation it has to be defined beforehand how the available phase space is divided up. The binning in absolute rapidity $|y|$ used in these scenarios has been defined according to the CMS detector geometry and is given in Table 4.2, which is identical to the binning used in Chapter 5. The binning in jet p_{T} follows approximately its assumed resolution in the CMS calorimeters and is listed in Table 4.3. The upper limits considered

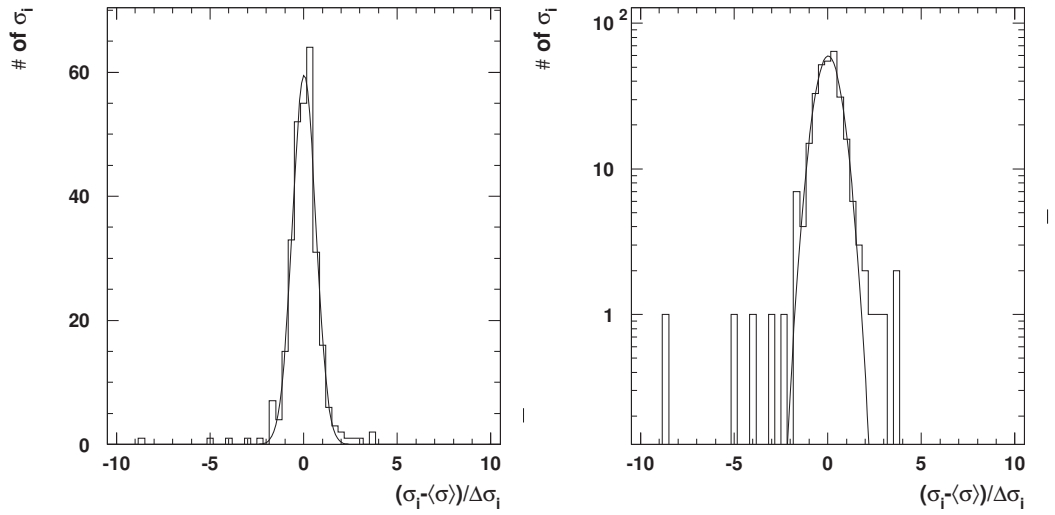

Figure 4.4: *Distribution of individual NLO inclusive jet cross section calculations σ_i around their mean $\langle\sigma\rangle$ for the p_{T} bin of $53\ \mathrm{GeV}$ to $67\ \mathrm{GeV}$ at central rapidity in linear (left) and logarithmic scale (right). All differences have been scaled by their corresponding standard deviation s_i. Additionally, a Gaussian fit is shown.*

in $p_{\mathrm{T,jet}}$ are set as appropriate for an integrated luminosity $\mathcal{L}_{\mathrm{int}}$ of $100\ \mathrm{pb}^{-1}$ at $10\ \mathrm{TeV}$. The expected reach is defined as the upper boarder of the last bin in which one expects at least one jet for an integrated luminosity of $100\ \mathrm{pb}^{-1}$. This has been determined using PYTHIA particle jets applying the full parton shower and hadronisation procedure, which is also used in section 4.2 and Chapter 5.

4.1.3 Scale Uncertainty

Every calculation in perturbative QCD *per se* is merely an approximation to the full theory. Generically, a cross section formula to order n in the strong coupling for hadron-hadron collisions can be written as

$$\sigma_{\mathrm{h_1 h_2}} = \sum_{k=0}^{n} \alpha_S^k(\mu_R) \sum_{\text{flavour } i} \sum_{\text{flavour } j} c_{i,j,k}(\mu_R, \mu_F) \times f_i^{\mathrm{h_1}}(x_1, \mu_F) \times f_j^{\mathrm{h_2}}(x_2, \mu_F), \quad (4.3)$$

which depends on

- the strong coupling constant α_S
- perturbative coefficients $c_{i,j,k}$

$\lvert y_{min}\rvert$	$\lvert y_{max}\rvert$	expected reach in p_{T} for $100\ \mathrm{pb}^{-1}$	calorimeter part
0.00	0.55	1684 - 1784 GeV	central barrel, y region 1
0.55	1.10	1588 - 1684 GeV	central barrel, y region 1
1.10	1.70	1327 - 1410 GeV	barrel/endcap transition
1.70	2.50	905 - 967 GeV	endcap
2.50	3.20	507 - 548 GeV	endcap/forward transition
3.20	5.00	272 - 300 GeV	forward

Table 4.2: *Detector geometry based binning in absolute rapidity. The reach in p_{T} is defined as the jet p_{T} bin in which at least one jet is expected to be observed with $100\ \mathrm{pb}^{-1}$ of integrated luminosity.*

maximum $\lvert y\rvert$	lower limits in $p_{\mathrm{T,jet}}$/ GeV									
5.00	53	67	81	97	114	133	153	174	196	220
5.00	245	272	300							
3.20	330	362	395	430	468	507	548			
2.50	592	638	686	737	790	846	905	967		
1.70	1032	1101	1172	1248	1327	1410				
1.10	1497	1588	1684							
0.55	1784									

Table 4.3: *Resolution based binning in jet transverse momentum $p_{\mathrm{T,jet}}$ together with the respective reach in $\lvert y_{\mathrm{max}}\rvert$.*

- the parton density functions $f_i^{\mathrm{h}_1}(x_1, \mu_F)$, $f_j^{\mathrm{h}_2}(x_2, \mu_F)$ of the hadrons h_1 and h_2
- the renormalisation and factorisation scales μ_R and μ_F
- the momentum fractions x_1, x_2.[1]

Now it is justified to ask how good an approximation to order n in α_S is. Unfortunately, an exact answer to that can only be given if either the full result is known or at least the next term of the perturbative series implying a next-to-next-to-leading order (NNLO) computation of the inclusive jet cross section. For a more detailed discussion of this topic see [17]. Both are not at disposal so a fallback solution has to be applied using scale variations where *scale* means the renormalisation scale μ_R and the factorisation scale μ_F above, both with dimensions of a momentum. They enter into Equation 4.3 via the evaluation of the strong coupling $\alpha_S(\mu_R)$ due to the renormalisation of ultraviolet divergences, and via the evaluation of the parton density functions $f_i(x, \mu_F)$ due to the factorisation into short-distance hard scatters and long-distance hadronic physics. In the full theory the dependence of the cross section on these scales would be exactly cancelled via the μ_R and μ_F dependence of the perturbative coefficients $c_{i,j,k}$. In a truncated series, however, a residual dependency remains which is assumed to reflect to some extent the sensitivity to missing higher order contributions and which should therefore decrease with more terms added to the perturbative series.

Since one lacks sufficient information one cannot associate a definite confidence interval with the cross sections for different choices of these scales. So it can only be considered as a conventional recipe to choose a relevant physical momentum scale of the concerned process and to vary it by a factor of one half and two in order to derive a scale uncertainty. Here, the p_{T} of each individual jet is taken as central value. For the case of a simultaneous variation of μ_R and μ_F the described uncertainty can then be evaluated by comparing the central result of a cross section with the ones for $\mu_R = \mu_F = p_{\mathrm{T,jet}}/2$ and $\mu_R = \mu_F = 2 \cdot p_{\mathrm{T,jet}}$.

Because it is easily possible that the two variations both lead to only an increase (or decrease) of a cross section, the uncertainty can be one-sided. In this case, always the larger deviation is taken. In the Figures 4.5 and 4.6 the relative scale uncertainties of calculations in LO and NLO are presented for the k_{T} (left plots) and the SISCone algorithm (right plots). To leading order the cross section variation is caused almost completely by the change of μ_R in α_S such that both algorithms give, as expected, the same scale uncertainties of 20 to 40% increasing with p_{T} and rapidity. At NLO the uncertainties are reduced to 5 to 10% with apparently larger deviations for the SISCone algorithm.

[1] An integration over the momentum fractions x_1 and x_2 is implicitly assumed and not shown in the formula.

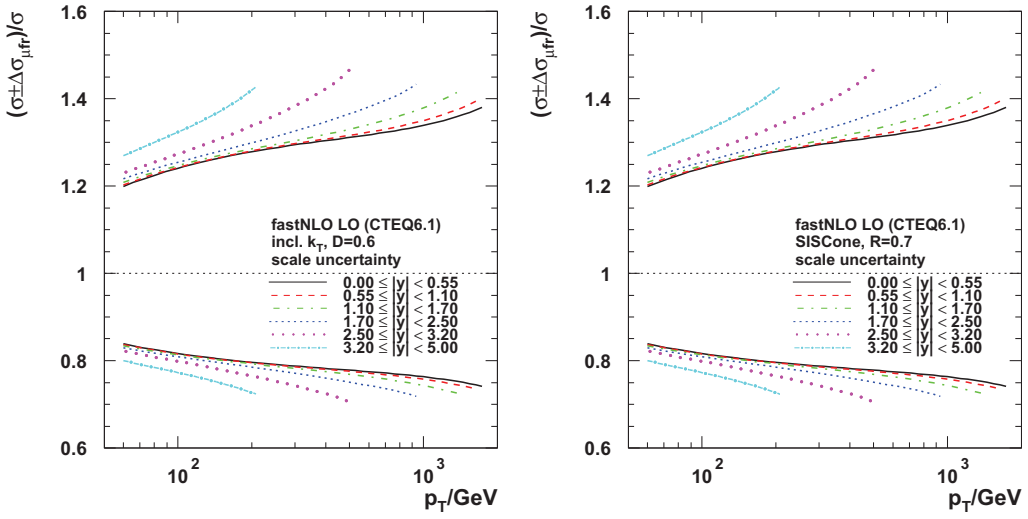

Figure 4.5: *Relative scale uncertainties of the inclusive jet cross section in LO for all bins in rapidity for the k_{T} (left) and SISCone (right) algorithm.*

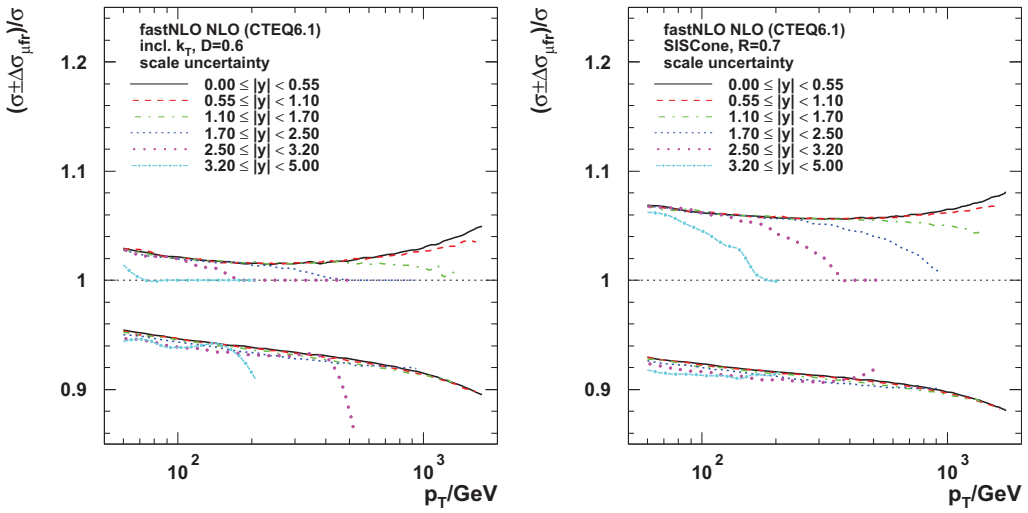

Figure 4.6: *Relative scale uncertainties of the inclusive jet cross section in NLO for all bins in rapidity for the k_{T} (left) and SISCone (right) algorithm.*

4.1.4 The Strong Coupling α_S

In order to deduce an explicit formula for the renormalised strong coupling α_S one expands the β function of the Renormalisation Group Equation (RGE), see [17], in powers of α_S:

$$\frac{\beta(\alpha_S)}{4\pi} = -\sum_{n=0}^{\infty} \beta_n \left(\frac{\alpha_S}{4\pi}\right)^{n+2} . \tag{4.4}$$

The first two coefficients are found to be

$$\beta_0 = \frac{33 - 2N_f}{3} , \tag{4.5}$$

$$\beta_1 = \frac{306 - 38N_f}{3} , \tag{4.6}$$

where N_f is the number of active flavours. Starting with β_2 further coefficients depend on the applied renormalisation scheme.
Retaining only the first term, the 1-loop solution for $\alpha_S(Q)$ can be written as

$$\alpha_S(Q) = \frac{\alpha_S(\mu_R)}{1 + \frac{\beta_0}{2\pi}\alpha_S(\mu_R) \ln\left(\frac{Q}{\mu_R}\right)} , \tag{4.7}$$

where Q denotes the relevant scale of the hard process. Using this equation, α_S can be evaluated at any (sufficiently high) scale provided it is known at one point, say M_Z.
Following ref. [112] and applying the 2-loop equation according to

$$\alpha_S(Q) = \frac{\alpha_S(M_Z)}{1 + \alpha_S(M_Z) \cdot L^{(n)}\left(\frac{Q}{M_Z}\right)} \tag{4.8}$$

where

$$L^{(1)}\left(\frac{Q}{M_Z}\right) = \frac{\beta_0}{2\pi} \ln \frac{Q}{M_Z} \tag{4.9}$$

one reproduces the 1-loop formula above and

$$L^{(2)}\left(\frac{Q}{M_Z}\right) = \left(\frac{\beta_0}{2\pi} + \frac{\beta_1}{8\pi^2}\alpha_S(M_Z)\right) \ln \frac{Q}{M_Z} \tag{4.10}$$

gives the 2-loop result. For the application with PDFs for NLO cross sections the 2-loop evolution is required. In Figure 4.7 left the sensitivity of the inclusive jet cross section at NLO to the evolution order is shown. With respect to the default 2-loop result significant differences are exhibited for the simple 1-loop formula, which should not be used for this purpose. Only a very small influence can be observed when going to a 3-loop evolution.

Figure 4.7 right presents the imposed changes of the inclusive jet cross section when the uncertainty on $\alpha_S(M_Z) = 0.1176 \pm 0.0020$ from the Particle Data Group value [17] is taken into account. Within the p_{T} range of 50 GeV to 1800 GeV an increase (decrease for smaller $\alpha_S(M_Z)$) of 4.0 to 2.5% can be seen.

However, this is not the whole story, since the strong coupling α_S enters already in the PDF fitting procedure. With their latest release of new PDFs the CTEQ collaboration has also provided a series of fits for different assumed values of $\alpha_S(M_Z) = 0.112, 0.114, 0.122$ and 0.125 which are to be compared with the central result for $\alpha_S(M_Z) = 0.118$ [113]. In order to demonstrate the difference between just changing $\alpha_S(M_Z)$ *a posteriori* and completely different PDF fits, Figure 4.8 shows on the left the variation of the inclusive jet cross section at central rapidity for the series of $\alpha_S(M_Z)$ values above with respect to always the central CTEQ6.6 PDF, and on the right for the PDF series CTEQ6.6a. Obviously, the full re-fits lead to significantly higher uncertainties at high p_{T} and smaller ones at low p_{T} as naively expected. Since SISCone gives very similar results only the figures for the k_{T} algorithm are included in this section.

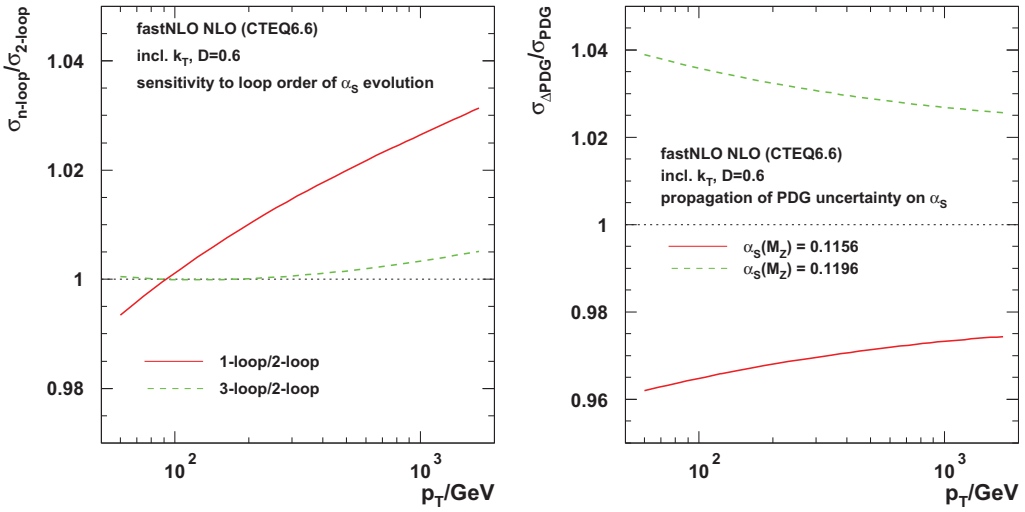

Figure 4.7: *Sensitivity of the inclusive jet cross section in NLO to the loop order of the evolution of the strong coupling α_S (left) and propagation of the uncertainty on $\alpha_S(M_Z) = 0.1176 \pm 0.0020$ from the Particle Data Group value [17] (right). Both are for the k_{T} algorithm at central rapidity, results for SISCone are very similar.*

4.1.5 PDF Uncertainties

In this section we will have a look at the next ingredient of the cross section Formula 4.3, the PDFs. Apart from special purposes global fits are performed to a multitude of experi-

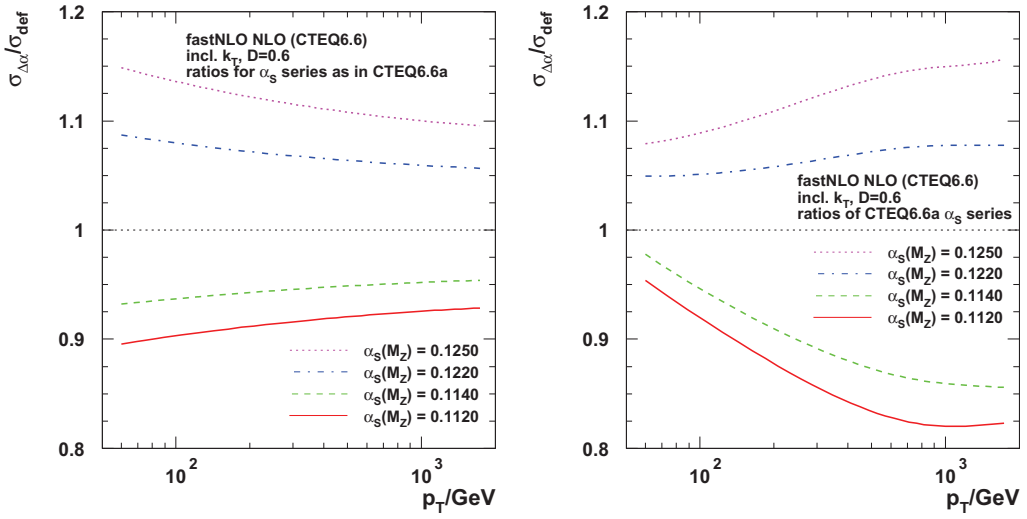

Figure 4.8: *Ratios of the inclusive jet cross sections in NLO for a series of $\alpha_S(M_Z)$ values for the k_T algorithm at central rapidity. On the left always the central CTEQ6.6 PDFs were taken and only $\alpha_S(M_Z)$ was changed, on the right the fit series CTEQ6.6a was used where $\alpha_S(M_Z)$ was different already in the fit procedure.*

mental datasets, each with their own particular uncertainties. Originally, the fits were done without propagating uncertainties or providing any information about the precision which lead to the effect that some deviations in experimental high p_T jet data, reported in [1], could be interpreted as hints for new physics. Finally, the discrepancies could, however, be accommodated for by adaptations in the gluon density of the proton.

In the meantime much more sophisticated ways to assess the uncertainties inherent to the PDF fits have been developed and the required PDF variations are available in the LHAPDF package [108, 109] whose most recent version is 5.8.0. In particular the following PDF sets have been looked at:

1. CTEQ6.1 [114] (used only to check algorithmic precision, see Section 4.1.1)
2. CTEQ6.5 [110] (our default)
3. CTEQ6.6 [113]
4. MSTW2008 [66]
5. HERAPDF0.1 [115, 116]
6. GJR08FF [117]
7. NNPDF1.0 [118]

To represent and propagate the uncertainties of these fits two primary schemes are employed. The most popular one orthogonalises the uncertainty source matrix and offers in addition to a central fit result (set member 0, δ_0) further $2n$ set members δ_k with the $+$ and $-$ variations for each of the n orthogonalised uncertainty sources. This is described in more detail for example in [119] or [66]. The (asymmetric) PDF uncertainties of the quantity of interest can then be extracted according to Equations 4.11 and 4.12:

$$\Delta\sigma^{+}_{\mathrm{PDF}} = \sqrt{\sum_{k=1}^{n} \left(\max\left[\sigma(\delta_k^{+}) - \sigma(\delta_0), \sigma(\delta_k^{-}) - \sigma(\delta_0), 0\right]\right)^2} \tag{4.11}$$

$$\Delta\sigma^{-}_{\mathrm{PDF}} = \sqrt{\sum_{k=1}^{n} \left(\max\left[\sigma(\delta_0) - \sigma(\delta_k^{+}), \sigma(\delta_0) - \sigma(\delta_k^{-}), 0\right]\right)^2} \tag{4.12}$$

All the PDF sets listed above fall into this category except NNPDF1.0. The latter applies a Monte Carlo method to represent the allowed parameter space such that the central result and its uncertainty can be determined as mean and RMS of a large number of member functions. This procedure is similar to the approach used in Section 4.1.1 of this chapter to estimate the statistical uncertainty of the jet cross sections. Also here a zeroth member is provided corresponding to a pre-calculated average. From the sample to average it has therefore to be excluded, see especially [118].

In order to have comparable results from the different groups providing global fits, in all cases the default value and evolution was taken for α_S as defined by the corresponding PDF set. Since the tolerances for deviations from the central results, however, are not all defined consistently — in some cases (CTEQ, MSTW) 90% confidence intervals are chosen to set the limits on fit variations, in others (HERAPDF1.0, NNPDF1.0) 1σ-contours are used — a simple scaling factor has been applied to achieve a more equal footing for the comparisons. To adapt 1σ-contours to the 90% confidence levels the resulting uncertainties have been multiplied by a factor of $\sqrt{2}/\mathrm{erf}^{-1}(0.90) \approx 1.64485$. In case of the GJR08FF PDF set the factor of $1/0.47$ as given in [117] was applied. In the figures these are both labelled as *CL90*.

Figure 4.9 presents the inclusive jet cross sections versus p_{T} for the k_{T} (left) and the SISCone algorithm (right) for all considered bins in rapidity. The coloured bands correspond to the PDF uncertainty as derived from the CTEQ6.5 PDF set. Obviously, both algorithms look very similar on a logarithmic scale. For a better visibility, the ratios of the inclusive jet cross section in NLO for the SISCone with respect to the k_{T} algorithm are provided in Figure 4.10 in all rapidity regions showing an about 15 to 8% higher cross section for SISCone.

Concerning PDF uncertainties in general both exhibit a similar behaviour. For a closer look, in Figure 4.11 the relative PDF uncertainties for the inclusive jet cross section at

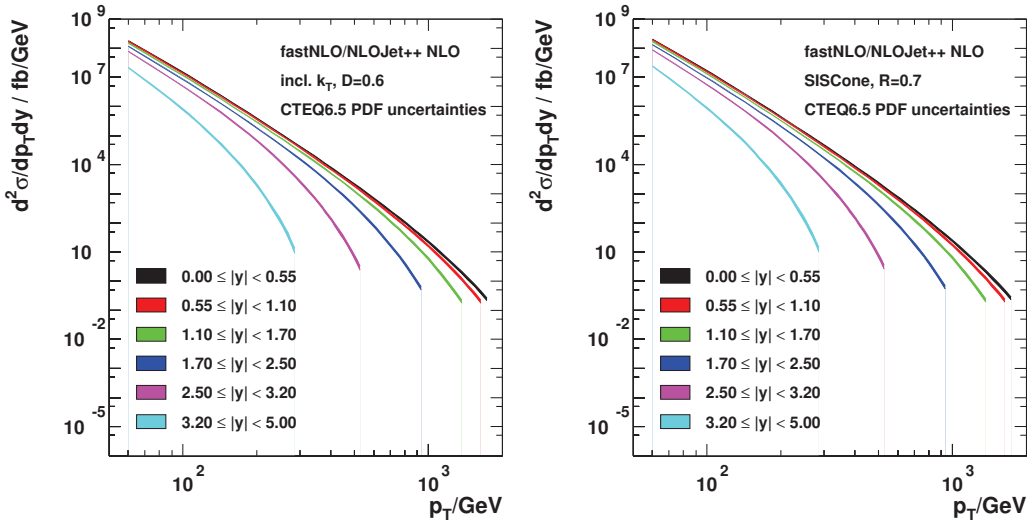

Figure 4.9: *Inclusive jet cross sections in NLO for the k_{T} (left) and SISCone algorithm (right) in all rapidity regions. The coloured bands correspond to the PDF uncertainty as derived from the CTEQ6.5 PDF set.*

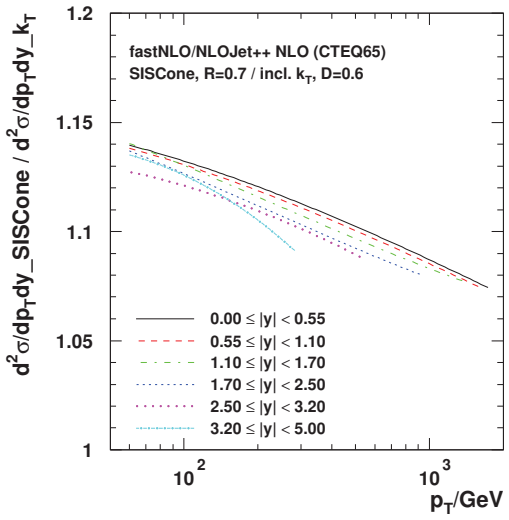

Figure 4.10: *Ratio of the inclusive jet cross section in NLO for the SISCone with respect to the k_{T} algorithm in all rapidity regions.*

NLO are drawn in all rapidity regions. From the top row it can be concluded that k_{T} on the left and SISCone on the right, both for CTEQ6.5, do not differ significantly. Hence, the other four plots show only the k_{T} algorithm, now for the MSTW2008, HERAPDF0.1[2], GJR08FF and NNPDF1.0 PDF sets. Without going into details it is somewhat astonishing that especially between CTEQ6.5 and MSTW2008 at CL 90 there are factors of two to four between the two predictions.

In addition to the uncertainties propagated in the global fit PDF sets, also details like the selected measurements, parametrisation, theoretical assumptions etc. entering the various global fits lead to discrepancies in the extracted parton densities and hence in the predicted cross sections. In order to illustrate this, the central results of all PDF sets for the innermost rapidity are shown with respect to CTEQ6.5. Deviations of up to 10%, 20% in case of GJR08FF, are exhibited which partially lie outside the error band of just the CTEQ6.5 set.

4.1.6 K-Factors and Comparison to PYTHIA

Since NLO programs like NLOJET++ mostly are cross section integrators and not full Monte Carlo event generators the question remains, how non-perturbative corrections available in these generators can be dealt with. In the concrete case of the inclusive jet cross section this topic is discussed in Chapter 4.2. Another question frequently raised in this context is, how precise the cross section of a simulated Monte Carlo sample (mostly in LO) is and what *K-factor* should be used. This sample might be foreseen for an estimation of the background to another physics process and the point to address is by which factor this background could possibly be underestimated due to the fact that the Monte Carlo program only performed a LO calculation. Then a so-called K-factor is defined which is represented by the ratio

$$K := \frac{\sigma_{\mathrm{NLO}}}{\sigma_{\mathrm{LO}}} \tag{4.13}$$

and serves as a scaling factor to increase (or decrease) the weight of the simulated events.

However, the above equation does not completely define this factor! Most importantly, the PDF and its order have not been specified, neither in the numerator nor in the denominator. Three principal possibilities exist:

1. The LO and NLO PDFs of one set are used consistently with the corresponding LO and NLO matrix elements.

2. The same NLO PDF is used in both cases so that observed differences are due to the change in matrix elements alone.

3. Two completely different PDF sets are used. This could be useful if the ratio of a recent calculation with respect to an older LO PDF is required.

[2]For HERAPDF0.1 only the experimental uncertainty sources were considered.

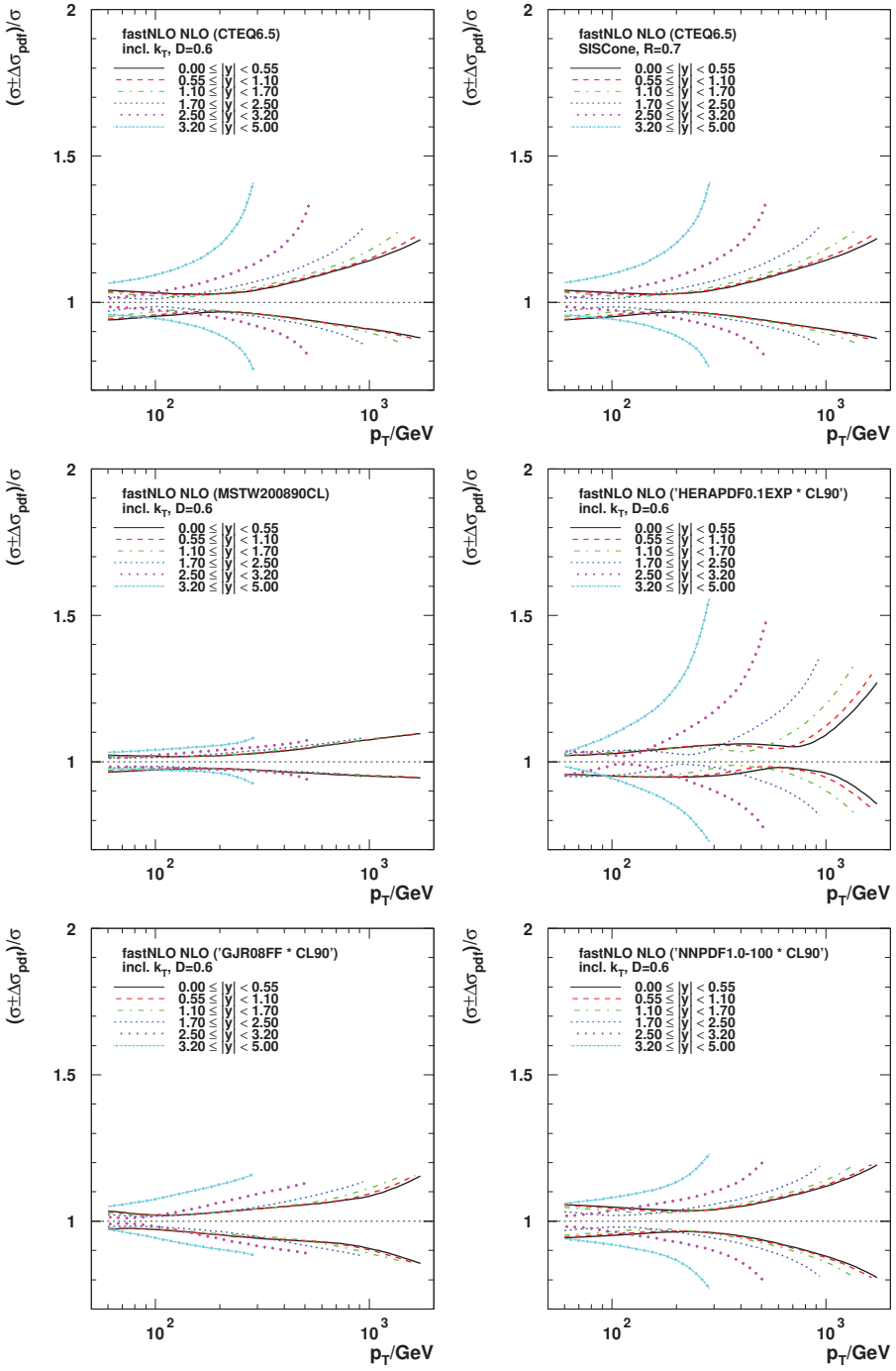

Figure 4.11: *Relative PDF uncertainties for the inclusive jet cross section at NLO in all rapidity regions. On the top for CTEQ6.5 PDFs for the k_{T} (left) and SISCone algorithm (right). The other four show only the k_{T} algorithm for the MSTW2008, HERAPDF0.1, GJR08FF and NNPDF1.0 PDF sets.*

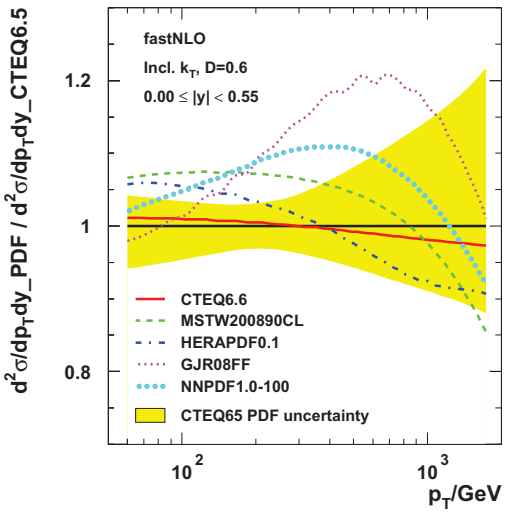

Figure 4.12: *Ratios of the inclusive jet cross section for k_{T} for different PDF sets over CTEQ6.5 at central rapidity. The yellow band corresponds to the PDF uncertainty as derived from the CTEQ6.5 PDF set.*

Following a discussion whether one can achieve more precise Monte Carlo generator results with LO or NLO PDFs, also different approaches have been considered, see [120] for many more details on this topic.

To demonstrate the difference between the possibilities one and two, Figure 4.13 shows the K-factors for both cases with the CTEQ6 PDF set for the k_{T} (top) and SISCone algorithm (bottom). As the LO and NLO PDFs have been fitted to data with severe differences on the theory side significant discrepancies are exhibited as expected. For the sake of checking the consistency of mock-up data from PYTHIA with CTEQ6L1 versus NLOJET++ the right plots have to be taken as K-factors.

Exploiting the fastNLO table a bit further one can also look into the K-factors separately for the underlying subprocesses (accessing different parton densities ...) of the reaction $pp \rightarrow$ jets. For a better explanation the seven possibilities are listed in Table 4.4 from which the symmetries $H_n(x_1, x_2) = H_n(x_2, x_1)$ for $n = 1, 4, 5, 6, 7$ and $H_2(x_1, x_2) = H_3(x_2, x_1)$ become obvious.[3] For practical reasons the subprocesses two and three are stored in a modified way though, such that H_2 comprises all combinations of qg, gq, $\bar{q}g$, $g\bar{q} \rightarrow$ jets with $x_g < x_{q,\bar{q}}$. For H_3 the same holds for $x_g > x_{q,\bar{q}}$. Taking this subdivision into account the respective K-factors with CTEQ6.5 NLO are shown in Figures 4.14 and 4.15 for the k_{T} algorithm at all investigated rapidities.

In addition, Figure 4.16 collects all K-factors at central rapidity into one plot for the k_{T} (left) and the SISCone algorithm (right). So not only are they significantly different for the

[3] In case of $p\bar{p}$ the two exchanges $H_4 \leftrightarrow H_7$ and $H_5 \leftrightarrow H_6$ are required.

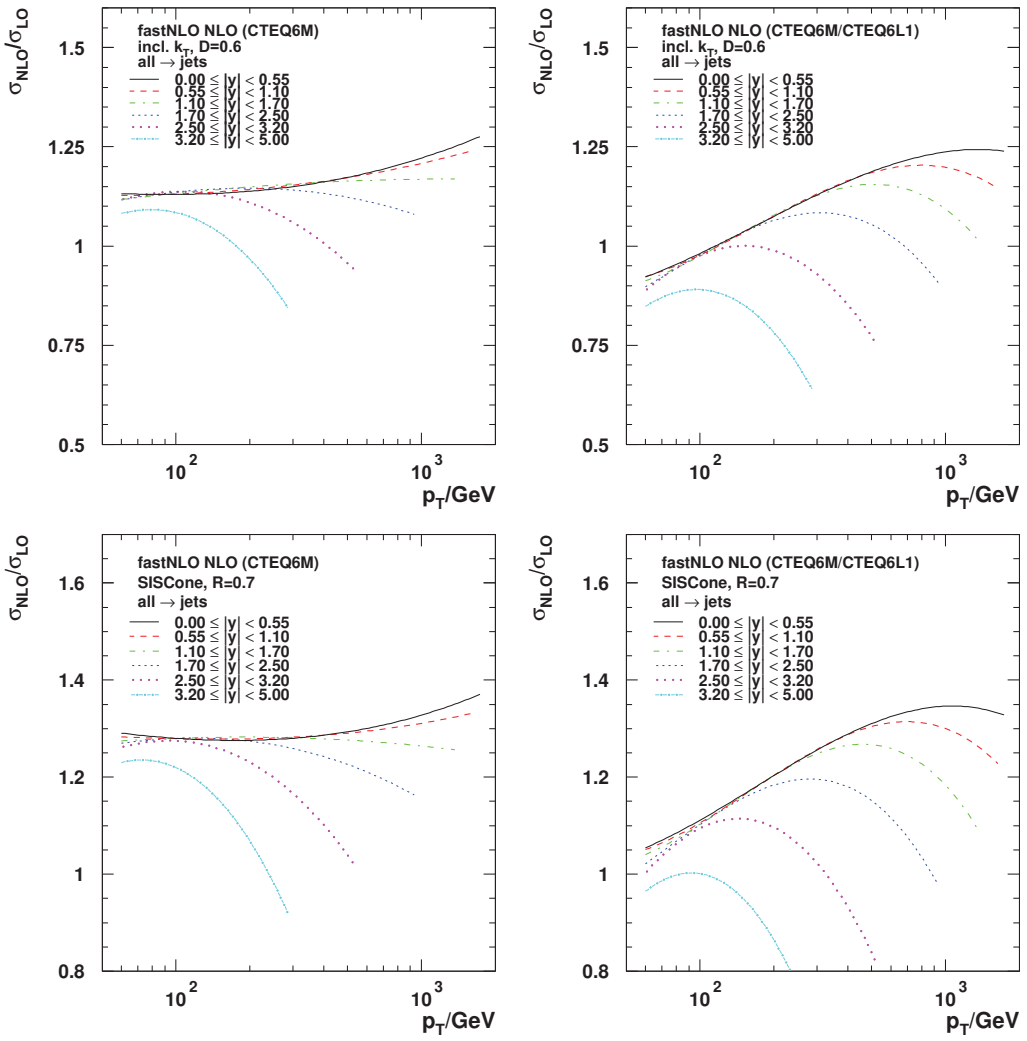

Figure 4.13: *NLO to LO cross section ratios (K-factor) for the k_{T} (top) and SISCone algorithm (bottom) when both times the NLO PDF CTEQ6M is used (left) or the LO PDF CTEQ6L1 in the denominator.*

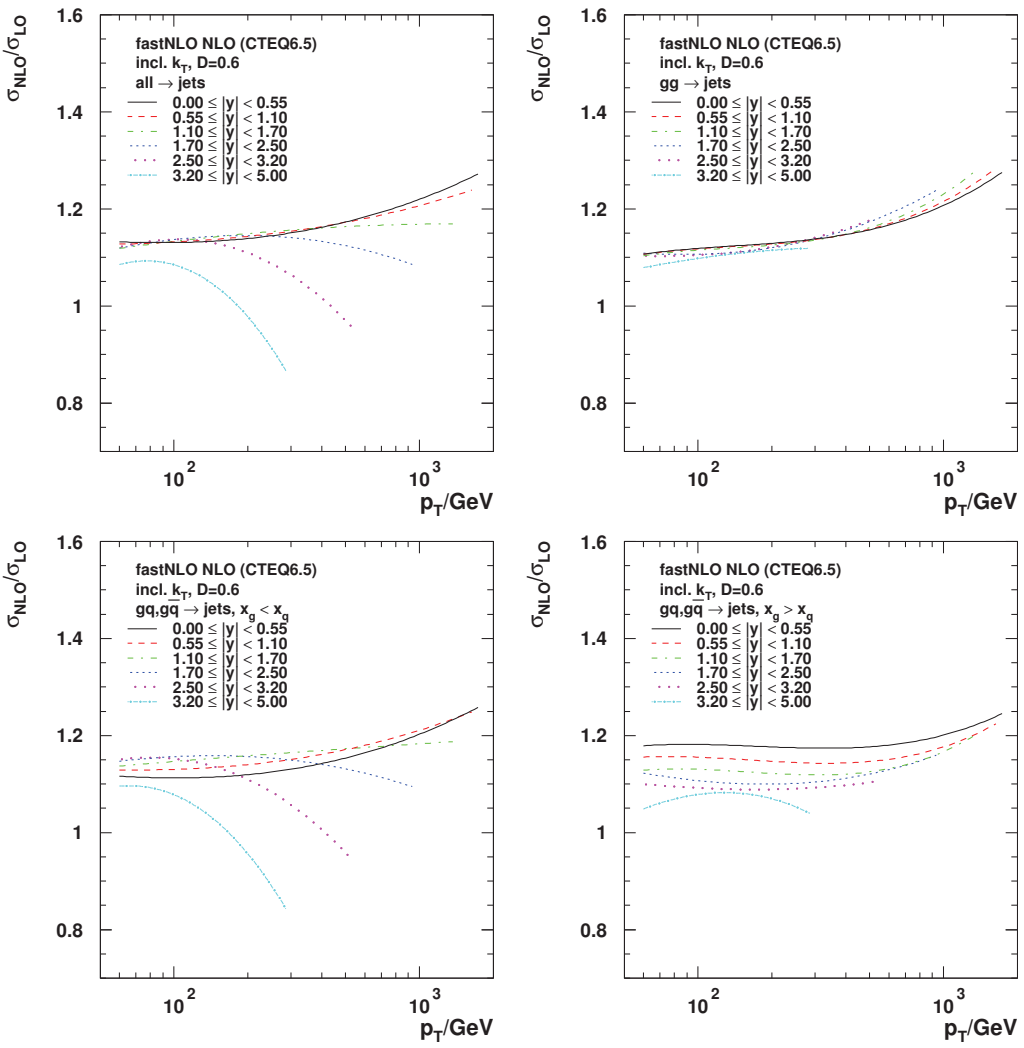

Figure 4.14: *NLO to LO cross section ratios (K-factor) for the k_{T} algorithm with CTEQ6.5 for all (top left) and the three gluon initiated subprocesses according to Table 4.4.*

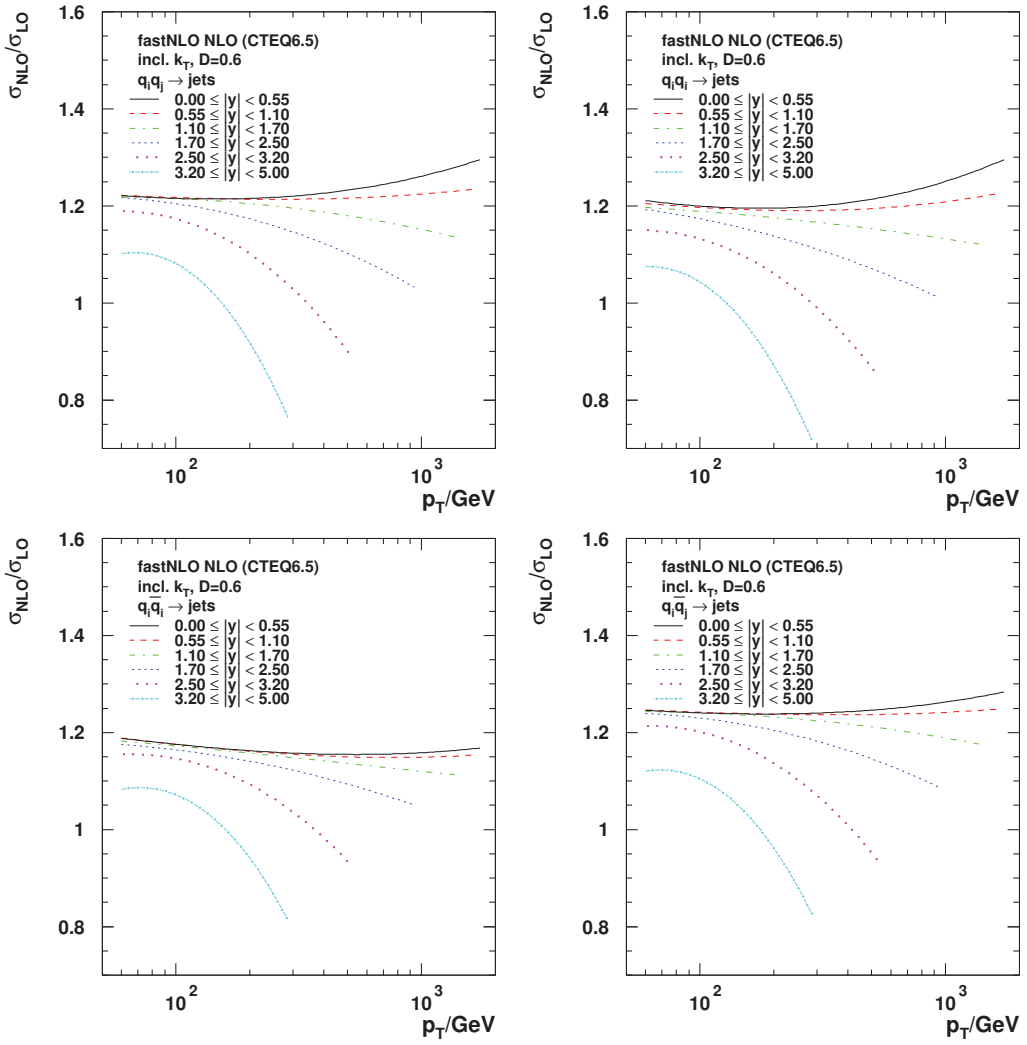

Figure 4.15: *NLO to LO cross section ratios (K-factor) for the k_{T} algorithm with CTEQ6.5 for the three quark and/or anti-quark initiated subprocesses according to Table 4.4.*

gg	$\rightarrow$	jets					$\propto$	$H_1(x_1, x_2)$
qg	$\rightarrow$	jets	plus	$\bar{q}g$	$\rightarrow$	jets	$\propto$	$H_2(x_1, x_2)$
gq	$\rightarrow$	jets	plus	$\bar{q}g$	$\rightarrow$	jets	$\propto$	$H_3(x_1, x_2)$
$q_i q_j$	$\rightarrow$	jets	plus	$\bar{q}_i \bar{q}_j$	$\rightarrow$	jets	$\propto$	$H_4(x_1, x_2)$
$q_i q_i$	$\rightarrow$	jets	plus	$\bar{q}_i \bar{q}_i$	$\rightarrow$	jets	$\propto$	$H_5(x_1, x_2)$
$q_i \bar{q}_i$	$\rightarrow$	jets	plus	$\bar{q}_i q_i$	$\rightarrow$	jets	$\propto$	$H_6(x_1, x_2)$
$q_i \bar{q}_j$	$\rightarrow$	jets	plus	$\bar{q}_i q_j$	$\rightarrow$	jets	$\propto$	$H_7(x_1, x_2)$

Table 4.4: *Subprocesses*

seven subprocesses, but also between the two considered jet algorithms. To combine the presented curves into the black overall correction, one needs to know the relative decomposition of the total cross section which is presented in Figure 4.17 versus the scaled jet momentum $x_T := 2p_{\mathrm{T}}/\sqrt{s}$.

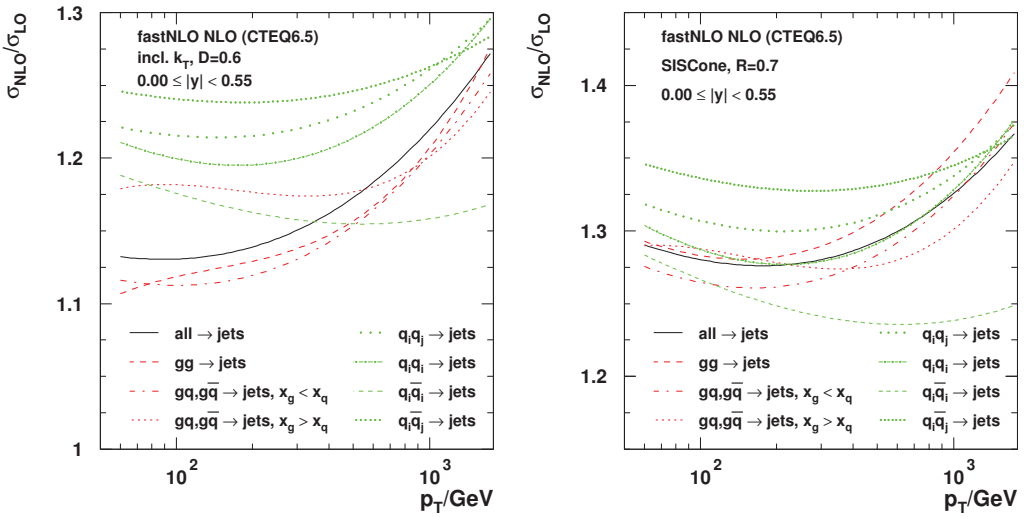

Figure 4.16: *NLO to LO cross section ratios (K-factor) for the k_{T} (left) and SISCone algorithm (right) with CTEQ6.5 for all subprocesses at central rapidity.*

Finally, the simulated Monte Carlo event samples representing the data in the inclusive jets analysis in Chapter 5 have been generated with the D6T tune for PYTHIA, an update to the tune DWT [121], and employed the CTEQ6L1 leading-order PDF set [119]. This means that the correction factors for a consistency check with NLOJET++ can be read directly from the right plot of Figure 4.13 above. Nevertheless a small residual discrepancy remains. Especially at LO these can be observed as shown in Figure 4.18, where 20 to 40% variations correspond to the LO scale uncertainties. The final comparison of PYTHIA to NLOJET++ (triangle marker) exhibits a nice consistency flat over the whole p_{T} range

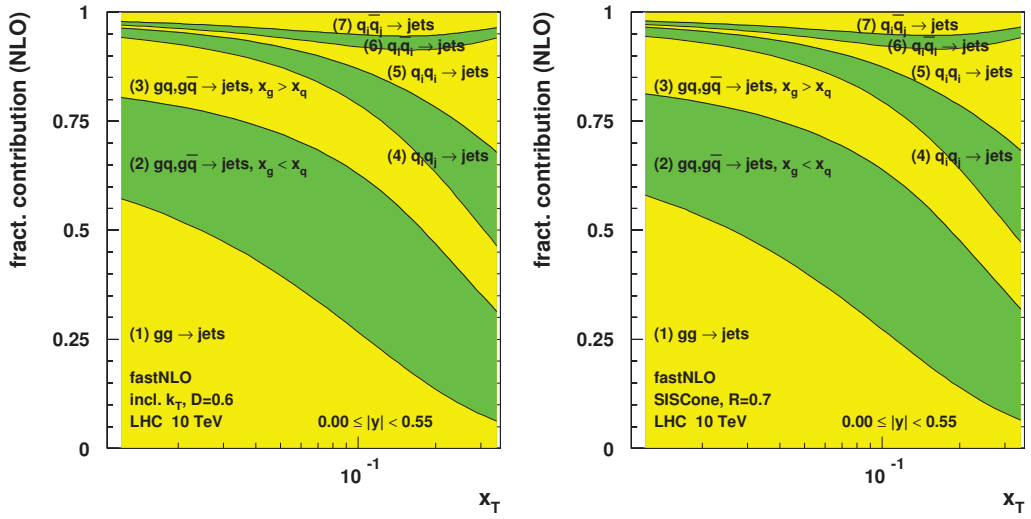

Figure 4.17: *Relative decomposition of the NLO cross section into the seven subprocesses of Table 4.4 for the k_T (left) and the SISCone algorithm (right) at central rapidity is depicted cumulatively.*

and only 3 to 4% below exact closure at unity. The difference in the ratio is found to be attributed to the different handling of the $\Lambda^4_{\overline{\text{MS}}}$ paramter for four active quark flavours, which in the fastNLO case is the value which was also used during the fit of the evaluated CTEQ6L1 PDF ($\Lambda^4 = 215\,\text{MeV}$). However, in the PYTHIA case within the CMSSW framework, the Λ^4 paramter was fixed to 192 MeV without considering PDF informations. This is confirmed by the two curves, where the NLOJET++ result with the PYTHIA like choice lies directly above the PYTHIA ratio, and the NLOJET++ result with the default choice and the PYTHIA α_S definition closes exactly at one. This also rules out further differences due to the α_S formula or scale choices.

The presented study is also available as an internal document of the CMS experiment [122].

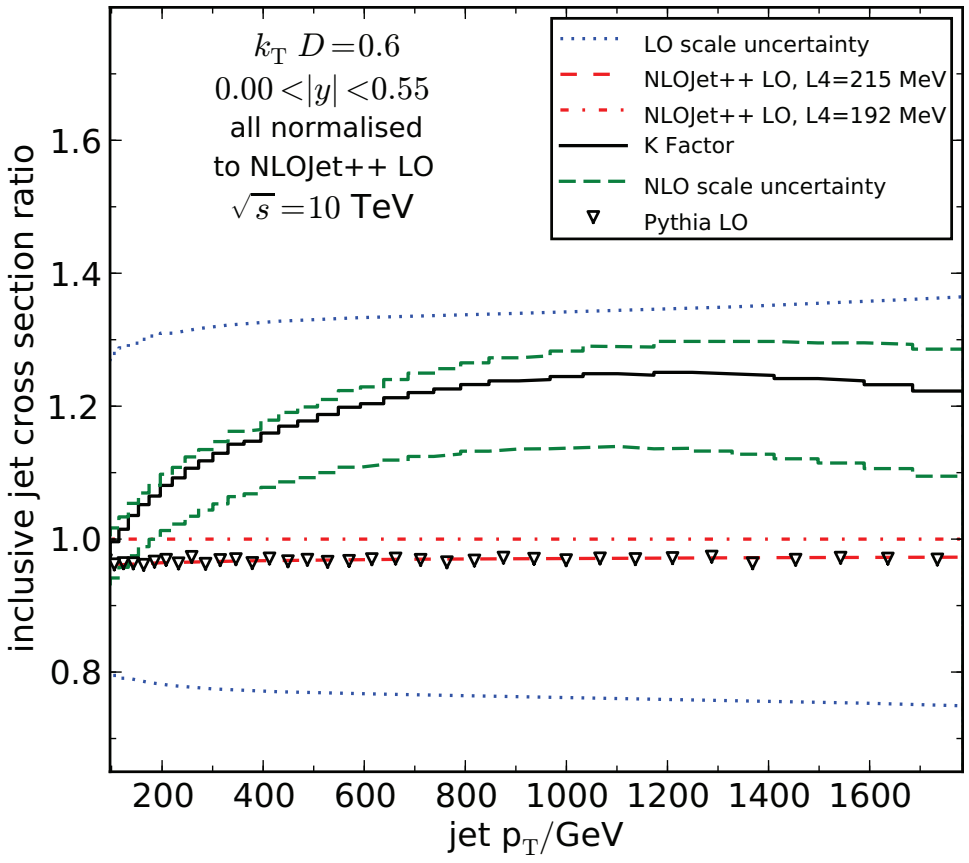

Figure 4.18: *Cross section ratios are shown of the LO* PYTHIA *Monte Carlo generator and* NLOJET++ *in LO. For the LO derivations the CTEQ6L1 PDF was used. The different λ_4 choices during the α_S evaluation explain the observed offset between the* PYTHIA *cross section and the default choice of* NLOJET++ *used at LO. Scale uncertainties are derived from the deviations between $\mu_R = \mu_F = \frac{1}{2} p_{\mathrm{T,jet}}$ and $2 \cdot p_{\mathrm{T,jet}}$ and can be interpreted as uncertainty band of the K-Factor or the central LO closure line respectively.*

4.2 Non Perturbative Corrections

Experimental data are usually corrected for all kind of detector effects to the so-called *particle level* where they finally can be compared directly to theory expectations in the form of generated Monte Carlo events. In the case of inclusive jet measurements at hadron colliders, theory calculations in next-to-leading order of perturbative QCD can be derived as described in section 4.1. Unfortunately, no full NLO event generator exists as it is available in the form of MC@NLO [123] for other processes. Therefore one has to account for non-perturbative effects by applying additional correction factors to the NLO theory curves as it was done for example in the relevant analyses from the CDF and DØ experiments at the Tevatron [64, 65, 67, 124]. A general overview with a more elaborate description of this point can also be found in [33]. In the following a method is presented which derives such non-perturbative corrections and their uncertainty for the inclusive jet measurements at LHC energies by comparing different LO Monte Carlo event generators.

4.2.1 Monte Carlo Generators and Tunes

In the following, full Monte Carlo event generators were employed to derive the non-perturbative corrections:

- PYTHIA 6.4 [111]
- HERWIG++ 2.2 [49, 125]
- HERWIG 6 [126] together with JIMMY [100]

All of these start their event generation with leading-order matrix elements, which are then followed by a parton shower algorithm, treatment of multiple parton interactions (MPI), and finally a hadronisation step. Whilst the matrix element calculation and the parton shower can be performed with perturbative techniques, multiple parton interactions and hadronisation require phenomenological models which have to be tuned to data from previous experiments to make sensible predictions for LHC energies. These predictions, however, can vary significantly depending on the used model, even if the models describe the same data used for the tuning equally well. Hence, leading-order Monte Carlo generators make up for their lack of predictive power compared to next-to-leading order calculations by emulating higher-order effects with parton showers and incorporating tunable non-perturbative approaches.

Multiple Parton Interactions

In this context, one frequently encounters the term *Underlying Event* (UE), under which it is customary to subsume all effects that are not directly connected to the hard parton scattering. Hence, this term comprises both, initial and final state radiation as well as multiple parton interactions, i.e. perturbative *and* non-perturbative parts of the event modelling.

Since the corrections to be derived, however, consider the leading-order plus parton shower theory to be on an equal footing as the next-to-leading order calculation and therefore only deal with the non-perturbative part, it would be misleading to say that the Underlying Event is corrected for. So this term will be avoided and it will only be mentioned in the context of the MC tunes where it is justified to do so, since also the parton shower technique has at least one tunable parameter which is the scale cut-off where the shower development is stopped.

Multiple parton interactions are additional scatters that take place in the hadron collision and are only weakly related to the hard interaction.[4] Both PYTHIA and HERWIG++ handle multiple parton interactions by performing a number of minimum bias parton scatters above a certain threshold in transverse momentum exchange. This threshold, together with the chosen parton density function, defines an average number of additional interactions that take place in an event and is one of the most important parameters in a tuning procedure. Other parameters include the matter distribution in the proton and variables that steer the amount of initial and final state radiation that is spread into the event. HERWIG 6 natively does not include MPI but only a simple soft interaction model that is based on an approach pioneered by the UA5 experiment [127]. The lacking functionality can, however, be complemented by linking with the JIMMY package that also inspired the model in HERWIG++. More details can be found e.g. in [128].

Hadronisation

The second primary difference of the Monte Carlo generators of the PYTHIA and the HERWIG family is their treatment of the hadronisation step in the event generation. PYTHIA uses the Lund string model which assumes colour-charged strings between the partons coming from the parton shower step. As those partons move apart, the strings break and new quark-anti-quark pairs appear at the breaking points. In contrast, both, HERWIG 6 and HERWIG++, employ the cluster hadronisation model, in which colour-neutral clusters are formed from neighbouring quarks, which then decay isotropically into hadrons. More details about the specific models are given in section 2.5.2.

Tunes and Event Generation Setup

In order to compare multiple parton interaction and hadronisation effects one has to be aware of the parameter sets used for the event generation. Several tunes are available for PYTHIA. In the recent official MC productions of the CMS collaboration tune D6T [129] is used. The CTEQ6L1 [119] parton density function was used during the estimation of the parameters and is therefore linked to the parameter set.

[4] Some Monte Carlo generators take colour-connections between the partons of the hard interaction and the MPI into account.

For HERWIG++, the default tune [130] was used, which is built around the MRST2001 PDFs [131–133]. In the case of HERWIG + JIMMY, a recent tune is applied, which has been developed in the ATLAS collaboration [134]. Note that here the minimal transverse momentum for secondary scatterings becomes dependent on the centre of mass energy similar as parameter `PARP(90)` in the PYTHIA implementation. The tune is different from the ones reported in [128].

For the derivation of the hadronisation and multiple parton interaction corrections, it is necessary to produce two independent datasets: One dataset which contains the hadronic final state as given by the used Monte Carlo generator and one dataset which contains the partonic final state, i.e. the partons from the generator before the hadronisation step is performed and in which multiple parton interactions are disabled.[5] For PYTHIA this is achieved by setting the parameters `MSTP(81)=0` and `MSTJ(1)=0` to switch off MPI and hadronisation respectively. In HERWIG++, the lines

```
set /Herwig/Shower/ShowerHandler:MPI No
set /Herwig/EventHandlers/LHCHandler:HadronizationHandler NULL
```

have to be added to the configuration file. The final state particles in both cases have been clustered with the k_T algorithm [57–59] with a D-parameter of 0.6 and the SISCone algorithm [53] with a jet size parameter of 0.7 and a split-merge parameter of 0.75. For both algorithms the implementation as in ref. [60] has been used.

4.2.2 Spectrum Construction for Inclusive Jets

With the steeply falling inclusive jet cross section, it is mandatory to find a way to populate all of the desired phase space. Monte Carlo generators would by default generate events with respect to their cross section weight and thus practically no events containing jets with high transverse momenta would show up. There is, however, the possibility to consistently slice the phase space into pieces already during the evaluation of the hard matrix elements, thus allowing to generate events in a given range of transverse momentum of the outgoing partons of the hard scattering. This technique is commonly referred to as $\hat{\mathrm{p}}_\mathrm{T}$ slicing. For every one of the three generators, 18 $\hat{\mathrm{p}}_\mathrm{T}$ bins with one million events each were used to achieve sufficient statistical precision also at the upper edges of the bins. The jet spectra were then evaluated for each bin separately and subsequently added with weights according to the cross section of the given bin. With this method a comparable coverage of the whole inclusive jet spectrum for transverse momenta between $50\ \mathrm{GeV}$ and $4000\ \mathrm{GeV}$ is achieved. The choice in transverse momentum binning for the corrections is related to the expected jet resolution of the CMS detector and is given in Table 4.3. To take into account the differences in jet response in the detector, a binning in rapidity was chosen according to the geometry of the calorimeters see Table 4.2.

[5] To gain a deeper insight, additional datasets were produced which contain either multiple parton interactions, but no hadronisation and vice versa.

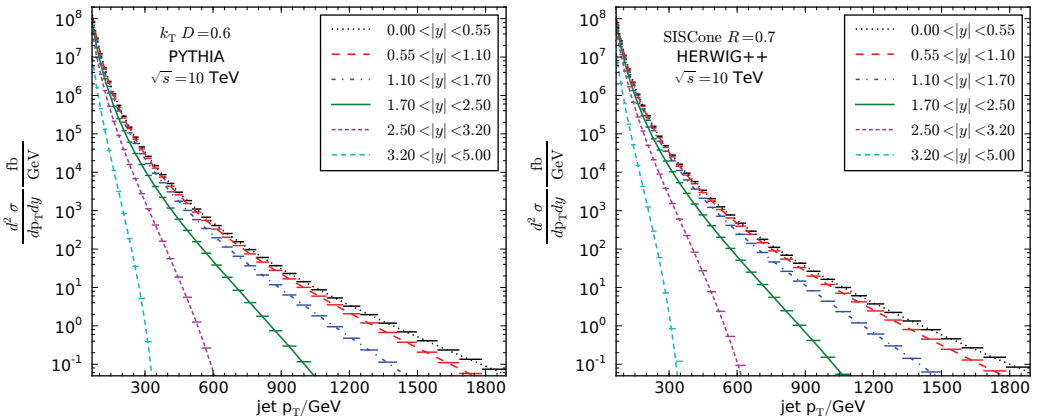

Figure 4.19: *Inclusive jet spectra from* PYTHIA *with* k_T 0.6 *for all rapidity bins with fits (left) and from* HERWIG++ *with SISCone* 0.7 *(right)*

This analysis is focused on the jet transverse momentum range, that will be attainable with the first $100\ \mathrm{pb}^{-1}$ of LHC data. In this range, a fitting procedure using an Ansatz function was performed on the inclusive jet spectrum to reduce statistical fluctuations in the spectrum ratios that will finally serve as correction factors. The specific Ansatz function 5.6 is also used within the measurements described in the next chapter. The parametrisation has been successfully used in inclusive jet measurements in the past [64, 135, 136]. The results are shown in Figure 4.19 for PYTHIA and HERWIG++.

As an estimation of the statistical uncertainty, the Monte Carlo uncertainty from the individual p_T bins was simply transferred to the data points of the fit curve. This method is overestimating the error, since fitting the distribution is reducing the number of degrees of freedom for the error function. This is not a problem, however, since the systematical uncertainty from the differences in the predictions of the compared Monte Carlo models is much larger than the statistical one, if sufficient numbers of Monte Carlo events are at disposal.

4.2.3 Derivation of the Correction Factors and their Uncertainties

Unlike lower level jet corrections like η-balance, see also [137], the corrections for multiple parton interactions and hadronisation can not be applied on a jet-by-jet basis. Therefore it is necessary to examine the jet spectra as a whole. The effects of multiple parton interactions and hadronisation on the inclusive jet spectrum, however, are not independent. In a generally more densely populated environment, jet algorithms tend to collect additional particles into jets that are not originating from the hard parton that the jet algorithm is supposed to reconstruct. In a sparsely populated environment, which is achieved in this

analysis by disabling multiple parton interactions, jets rather tend to lose particles due to out-of-cone effects. Especially cone-type jet algorithms with their fixed jet geometry are known to expose this feature. Figure 4.20 shows this delicate interplay of jet algorithm, multiple parton interactions and hadronisation. The curves which only state the generator name are the ratio between the full model (LO, PS, MPI, and hadronisation) and the model without hadronisation (LO, PS, and MPI only). The curves marked "without MPI" are the ratios of the model without MPI but with hadronisation (LO, PS, and hadronisation) and the same without hadronisation. Due to the fact, that the tunes are derived for the full model, the latter scenario is not strictly physically motivated but used to analyse the performance of jet algorithms in a sparsely populated environment. The cluster hadronisation model of HERWIG++ and HERWIG + JIMMY seems to be generally more prone to remove transverse momentum from jets. For higher transverse jet momenta all curves approach one, which is expected from the fact that harder jets are generally more focused.

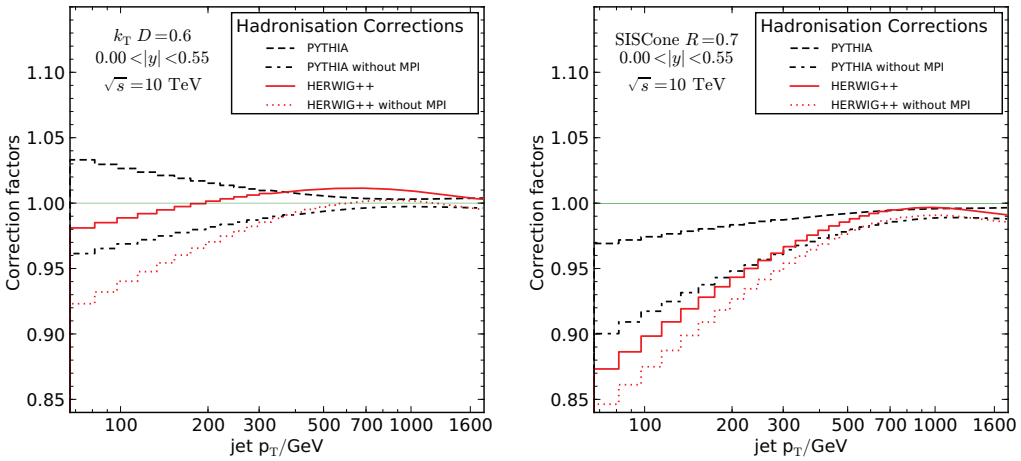

Figure 4.20: *Dependence of the hadronisation corrections on the presence of MPI, left for k_T, right for SISCone for the central rapidity region.*

The multiple parton interactions themselves, however, can shift the steeply falling inclusive jet spectrum significantly. A small shift in transverse momentum of jets at the upper edges of p_T bins leads to bin migration, which has a large impact on spectrum ratios, especially in the low p_T-range. The corrections for multiple parton interactions are very similar for both jet algorithms, as with the chosen jet size parameters they are also expected to cluster jets of roughly the same size. This becomes obvious from Figure 4.21. The MPI contributes an almost constant amount of energy per solid angle. Therefore the amount of MPI induced transverse momentum fraction of a jet depends on the jet size [138, 139]. For higher jet transverse momenta, the influence of multiple parton interactions vanishes

and the curves approach one. So for the combination of both hadronisation and multiple parton interaction corrections it is safe to assume that for jets above several hundred GeV of transverse momentum, the corrections become very small.

The shift in the transverse momentum spectrum is a bit smaller for HERWIG++ and HERWIG + JIMMY, which for the overall combined corrections together with the hadronisation results in significantly larger correction values for PYTHIA. Since HERWIG++ and HERWIG + JIMMY employ the same models for both multiple parton interactions and hadronisation, only with slightly different tunes and implementations, the average of the two is taken and treated as a combined "HERWIG-like" correction. To combine this correction with the PYTHIA prediction, again simply the average of the two is used. So overall, PYTHIA contributes one half and HERWIG++ and HERWIG + JIMMY one quarter each to the overall combined corrections. The estimation of the systematical error then is the spread between the PYTHIA value on one hand and the "HERWIG-like" value on the other hand, which can amass up to 10% for the k_{T} and even 20% for the SISCone algorithm (Fig. 4.22). The statistical uncertainties for the generator specific corrections are below one per cent in the whole accessible phase space apart from the kinematic edges. The uncertainties on the PYTHIA corrections are given in Figure 4.23. As the same phase space coverage is used for the other generators, the values for HERWIG++ and HERWIG + JIMMY are almost identical.

For the inner rapidity bins up to $|y| < 2.5$, on which the first LHC measurements will concentrate, the distribution ratios behave as expected (Fig. 4.24). In the outermost region, however, the statistical uncertainty increases very fast, so the absolute reach for a determination of the non perturbative corrections in this region is limited.

The presented study is also available as an internal document of the CMS experiment [140].

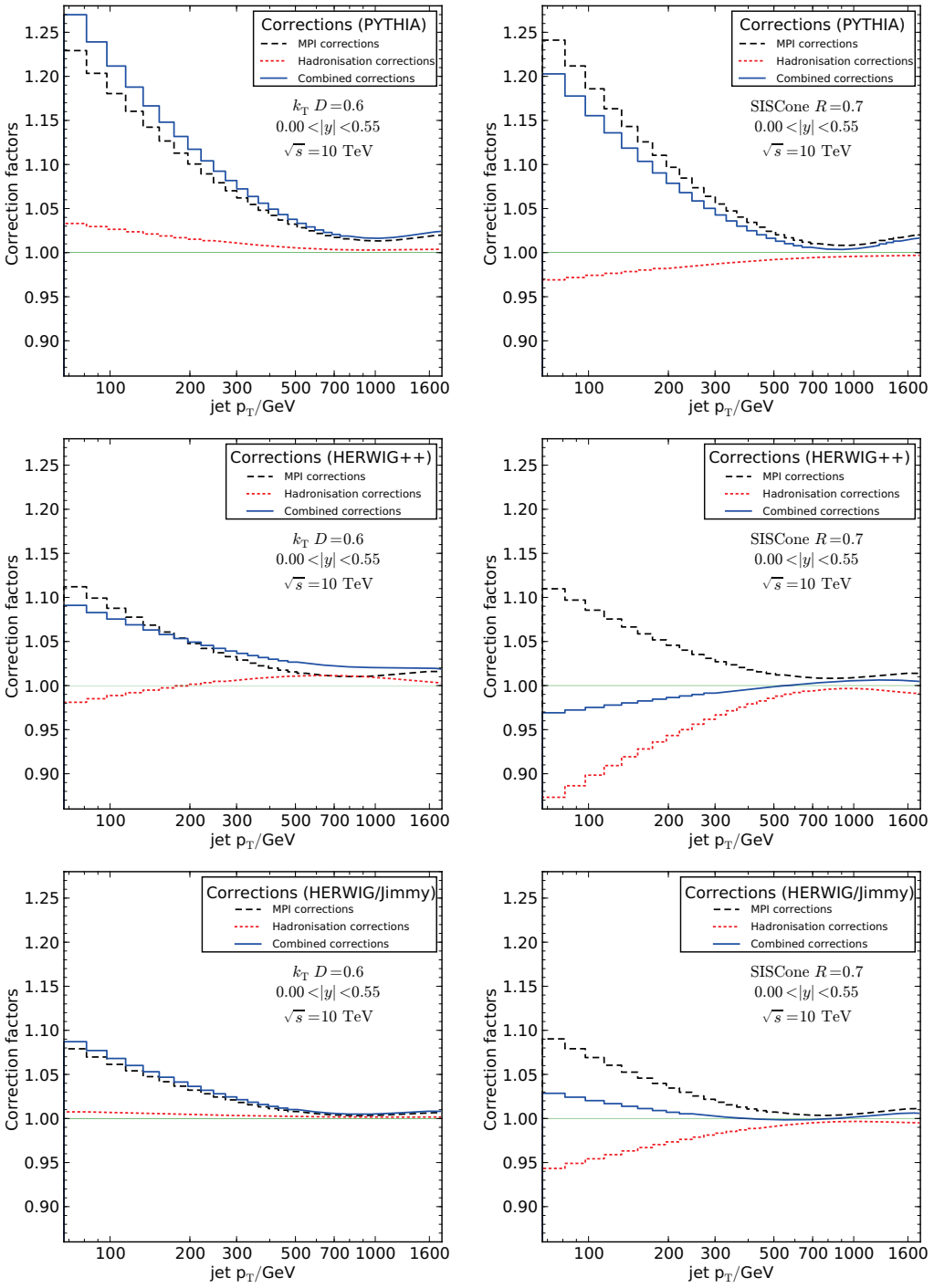

Figure 4.21: *MPI, hadronisation and overall correction for* PYTHIA *(upper row),* HERWIG++ *(middle row) and* HERWIG + JIMMY *(lower row) for the innermost rapidity bin. The left column corresponds to the k_T and the right one to the SISCone algorithm.*

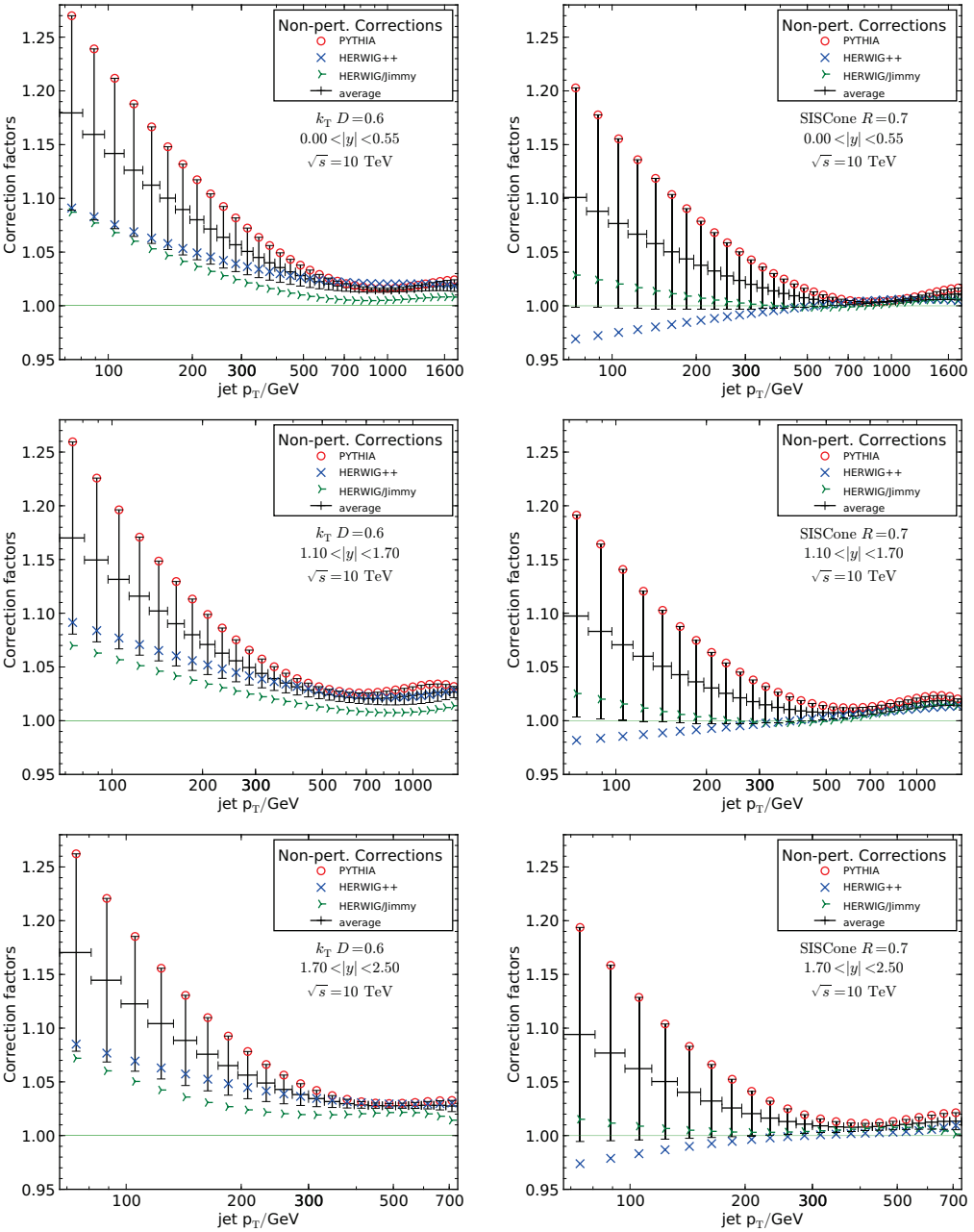

Figure 4.22: *Combined overall corrections for k_T (left) and SISCone (right) are shown for selected rapidity bins. The black error bars correspond to the difference between* PYTHIA *and the average of* HERWIG++ *and* HERWIG + JIMMY*, which is assumed to be the systematical uncertainty of the corrections. Statistical uncertainties are not shown for clarity, however they become significant when approaching the kinematic limits at large p_T and outer rapidities. Note the different scale on the x-axis for the different rapidity regions.*

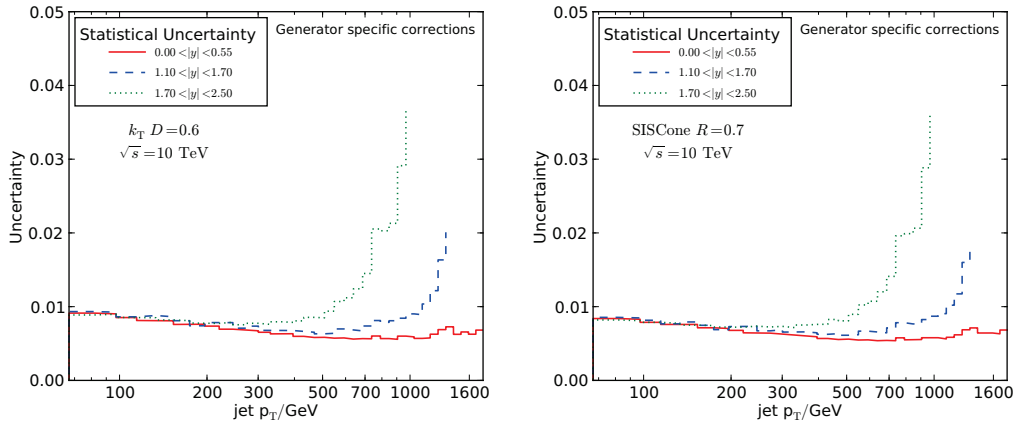

Figure 4.23: *The statistical uncertainty on the generator specific non-perturbative corrections is depicted for k_T (left) and SISCone (right) for selected rapidity bins in case of the* PYTHIA *generator. The uncertainties arise due to the chosen phase space coverage during the generation of the samples, and therefore the uncertainties for* HERWIG++ *and* HERWIG + JIMMY *are almost identical to those of the* PYTHIA *samples.*

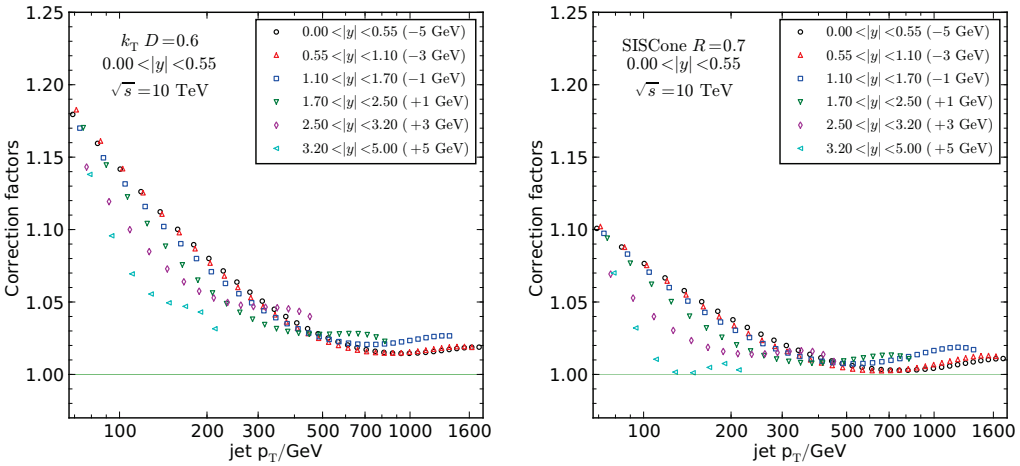

Figure 4.24: *Combined overall correction for k_T (left) and SISCone (right) for all rapidity bins for the p_T range given in Table 4.2. The points are slightly shifted on the p_T axis by the given values for better conspicuity.*

Chapter 5

Measurement of the Inclusive Jet Cross Section with CMS

With the start-up of LHC, a new frontier of energy will be surpassed and it is not at all clear that an extrapolation of our current knowledge in form of the Standard Model will suffice to describe the new measurements. Therefore, first analyses will have to check whether the LHC-data are compatible with the measurements of other experiments, like those at LEP or the Tevatron. Additionally, studies of well established observables are vital to gain a deeper understanding of the detector and enable the optimisation of the detector simulation. The next step will be to check measurements at the new energy regions against predictions. Of special interest will be those parts of the phase space where theoretical uncertainties, probably enhanced by the required extrapolation to untested energy regions, are dominating. Within the scope of this work, the plan of an initial measurement of the inclusive jet cross section with the CMS detector from first data is presented. The study uses fully simulated events in order to test experimental correction procedures and estimate their uncertainties. Additionally, a comparison of these corrected pseudo-measurements with dedicated calculations in next-to-leading order is shown. The comparison setup follows the procedures documented in [64, 65, 67] to set the initial LHC findings on an equal footing with current results from the Tevatron experiments.

Measuring the inclusive jet cross section and comparing it to perturbative QCD in the highest available order is the first step towards a measurement of parton distribution functions and the strength of the strong coupling α_S using jet data at the LHC.

Parton distribution functions are fundamental ingredients for understanding experimen-

tal results of all collision experiments involving hadrons in the initial state and are therefore of major interest to all experiments at the LHC.

Besides comparing experimental results to current PDFs derived from past experiments, new regions of phase-space will be opened. At the LHC, smaller fractional parton momenta x will be accessible as well as interactions with larger absolute momentum exchange than ever before. Especially the gluon density is a yet rather imprecisely known quantity.

Measured and unfolded jet cross sections can be fitted to theoretical calculations with varied underlying parton distribution functions. This procedure is sensitive to uncertainties of the theory, so it is advised to use the highest available accuracy, which means using NLO or higher order precision in the calculations of perturbative QCD for the fitting procedure. Conducting these calculations within a realistic amount of computing time, dedicated methods, like implemented in the fastNLO package, are required. More details on this are described in section 3.6.5 and section 4.1.

The cross section analysis is performed on fully simulated events which are adopted as pseudo data representing proton-proton collisions in the CMS detector at $\sqrt{s} = 10$ TeV. The events were generated with PYTHIA 6.4 which uses the recent tune D6T [129] employing the CTEQ6L1 [119] parton density function. The theoretical underpinning of the parton distributions and the tunes was already discussed previously (Chap. 4). More technical details about the samples analysed are given in the Appendix, Section B.4.

5.1 Observable

The differential inclusive jet cross section is measured in bins of the jet transverse momentum p_{T} and rapidity y and is defined by the basic formula 5.1:

$$\frac{d^2\sigma}{dp_{\mathrm{T}}dy} = \frac{C_{\mathrm{res}}}{\mathcal{L}\cdot\varepsilon}\cdot\frac{N_{\mathrm{jets}}}{\Delta p_{\mathrm{T}}\cdot\Delta y}\,, \tag{5.1}$$

where:

- N_{jets} is the number of jets counted in a bin,
- $\mathcal{L}$ is the integrated luminosity,
- ε is the efficiency of the event clean-up and any ID cuts,
- C_{res} is the resolution unsmearing correction factor,
- Δp_{T} and Δy are the p_{T} and rapidity bin sizes respectively.

The binning in absolute rapidity $|y|$ used in this analysis has been defined according to the CMS detector geometry [74] and is given in Table 5.1. The jet p_{T} intervals are based on the jet energy resolution expected for the CMS detector, such that the width of each bin

is roughly proportional to the absolute resolution at its centre. The exact border values are given in Table 5.2 together with the expected reach for an integrated luminosity of $10\ \mathrm{pb}^{-1}$. These upper limits are defined as the upper boarder of the last bin, in which at least one jet is expected. The estimation is based on particle jets derived from fully hadronised PYTHIA events.

Within this chapter only jets up to $y < |2.5|$ are analysed, as this detector region will also be the main candidate for initial measurements. It contains the highest p_{T} part of the phase space, and due to the overlap with the tracker, a steep improvement of understanding systematic effects in this region can be expected. For clarity, the bin $0.55 < |y| < 1.10$ is omitted in the following, as it behaves very similarly to the inner most bin.

$\lvert y_{min}\rvert$	$\lvert y_{max}\rvert$	expected reach in p_{T} for $10\ \mathrm{pb}^{-1}$	calorimeter region
0.00	0.55	1327 - 1410 GeV	central barrel, y region 1
0.55	1.10	1248 - 1327 GeV	central barrel, y region 1
1.10	1.70	1032 - 1101 GeV	barrel/endcap transition
1.70	2.50	790 - 846 GeV	endcap
2.50	3.20	468 - 507 GeV	endcap/forward transition
3.20	5.00	272 - 300 GeV	forward

Table 5.1: *Binning in absolute rapidity based on the detector geometry. The reach in p_{T} is defined as the jet p_{T} bin in which at least one jet is expected to be observed with $10\ \mathrm{pb}^{-1}$ of integrated luminosity.*

maximum $\lvert y\rvert$	lower limits in $p_{\mathrm{T,jet}}$/ GeV									
5.00	53	67	81	97	114	133	153	174	196	220
5.00	245	272	300							
3.20	330	362	395	430	468	507				
2.50	548	592	638	686	737	790	846			
1.70	905	967	1032	1101						
1.10	1172	1248	1327							
0.55	1410									

Table 5.2: *Resolution based binning in jet transverse momentum $p_{\mathrm{T,jet}}$ together with the respective reach in $|y_{\mathrm{max}}|$.*

5.2 Jet Reconstruction

In this analysis, jets are reconstructed either from energy depositions in the CMS projective calorimeter towers [74] (excluding the outer hadron calorimeter which is not part of the default jet reconstruction) or from all stable Monte Carlo particles. For the calorimetric jets the reconstruction thresholds of *Scheme B* [141] are applied. In addition, the raw energy of all jets observed in the forward region ($|\eta| > 3$) is artificially lowered by a constant factor of 0.7. This is done during the reconstruction in order to roughly equilibrate the trigger rates between the central and forward calorimeters.

For the purpose of the inclusive jet cross-section measurement two jet finding algorithms are employed: The inclusive k_{T} with a resolution parameter $D = 0.6$ and the SISCone with a radius of $R = 0.7$ and overlap threshold of 0.75. Both algorithms are infrared and collinear safe. The properties of infrared and collinear safety are necessary in order to compare with theory calculations. The specific choice of the parameters ($D = 0.6$ and $R = 0.7$) is a compromise between the need to maximise the energy collected into a jet and a still sufficient resolution for multi-jet events.

5.2.1 Jet Energy Determination

The default jet energy corrections (JEC) at CMS correct on average the observed jet energy to the energy of the final state particle jet [137]. They consist of three steps applied in sequence: The *Offset* correction [142] removes the energy added to the jet due to electronic noise and pile-up. The *Relative* correction removes the pseudorapidity dependence of the jet energy response and the *Absolute* correction restores the response to unity as a function of p_{T}. Currently, the jet energy corrections are derived from Monte Carlo truth by matching reconstructed jets with generated particle jets. The response is recorded in bins of particle jet p_{T} and η and the extracted correction factors are expressed as a function of the reconstructed jet quantities. Once collision data are available, the jet energy corrections will be derived from direct measurements. The relative correction, for example, will be extracted from di-jet balancing [143] and the absolute correction from γ+jet [144] or Z+jet [145, 146] balancing.

The pre-calibration of the hadron calorimeter comes from test beam measurements with charged pions of $50\ \mathrm{GeV}$ [147], while the pre-calibration of the electromagnetic calorimeter is established with test beam electrons. In CMS, the calorimeters will be calibrated using *in situ* measurements from collision data. The absolute scale of the ECAL will be determined from $Z \to e^+e^-$ events while isolated $\pi^0 \to \gamma\gamma$ events will be used to achieve a uniform azimuthal response. Random triggers, which do not depend on actual measurements, are used in order to collect a sample of events which can be regarded as unbiased as far as instrumental effects on the event selection are concerned. Such *Zero-Bias* and *Minimum-Bias* events, will be employed for the HCAL calibration to equalise the response in ϕ for each η ring. The absolute scale of the hadron barrel (HB) and the hadron endcap

(HE) calorimeters will be re-derived *in situ* using single isolated tracks exploiting the fact that their momenta are measured accurately by the CMS tracker. The hadron forward calorimeter (HF) will be calibrated using di-jet events, exploiting the transverse momentum balance. It should be noted however, that the underlying calorimeter calibration will not lead to (almost) calibrated jets because of the non-linear HCAL response. For this reason, the independent jet energy calibration described in the previous paragraph is mandatory. For the CMS simulated data used in this analysis, the calorimeter calibration is taken from test beam measurements.

The uncertainty of the jet energy scale is critical for the measurement of the inclusive jet cross section. With $10\ \mathrm{pb}^{-1}$ at $10\ \mathrm{TeV}$ pp collisions it is expected that enough data will be available in order to measure the jet energy scale with $\gamma + jet$ events up to $p_{\mathrm{T}} \sim 600\ \mathrm{GeV}$ [144]. At the same time, jets will be observed up to $p_{\mathrm{T}} > 1\ \mathrm{TeV}$. In the p_{T} range where *in situ* measurements will be available, the systematic uncertainty can be constrained to better than 10%. In the high transverse momentum region, the jet energy scale will have to be extrapolated [145] relying on MC simulations that need to be tuned to the direct measurements. This may lead to an increase of the systematic uncertainty.

In the present analysis an estimate of 10% uncertainty in the jet energy scale is used which includes the preliminary uncertainties on data-driven techniques. A more detailed discussion of this dominant systematic uncertainty will only be possible after jet energy corrections have been derived from actual collision data. For illustration, the magnitude of the JEC as a function of the observed jet p_{T} is shown in Fig. 5.1 and its pseudorapidity dependence is shown in Fig. 5.2.

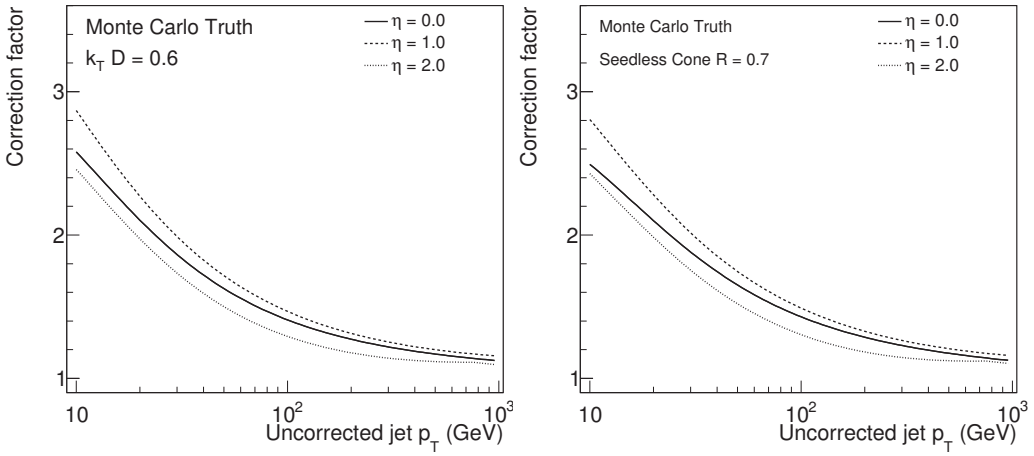

Figure 5.1: *Jet energy correction (JEC) as a function of the uncorrected jet p_{T} for the k_{T} (left) and the SISCone algorithm (right)[148].*

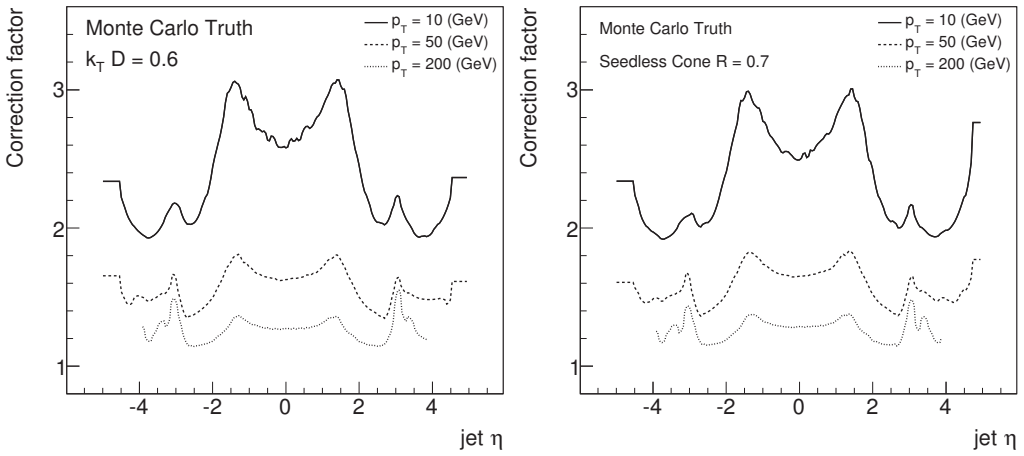

Figure 5.2: *Jet energy correction (JEC) as a function of the reconstructed jet η for the k_{T} (left) and the SISCone algorithm (right)[148].*

5.2.2 Jet-by-Jet MC Truth Comparison

In order to understand the remaining detector effects after jet energy scale corrections are applied, a matching procedure is used. For this purpose particle jets are matched to the full detector simulated and JES corrected ones. Jet pairs are declared a match, if they are the closest in

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta y)^2}. \tag{5.2}$$

Only jets with $p_{\mathrm{T}} > 10$ GeV are used within the matching. For further studies only pairs with $\Delta R < 0.5$ are used. The relative difference in p_{T} of the matched jets is recorded in histograms binned in p_{T} of the particle jets. In this case 25 bins were chosen, as a trade off between maximising the number of matches within each bin in order to achieve a reasonable description of the shape, and a reasonably small bin width in order to describe trends in mean and peak width with least smearing. The calorimeter response to single particles has a finite resolution which manifests itself in the finite width of the jet energy response distributions, which is not corrected for by the JES corrections and therefore can be observed within this matching procedure. An example of three such distributions from dedicated bins is given in Figure 5.3. As expected, the distributions have a rather Gaussian core and the width is reduced with increasing jet p_{T}. However, especially the peaks at large p_{T} show an additional shoulder to lower values, which emerges at about three orders of magnitude below the peak value. This is attributed to energy escaping detection from the calorimeter (*punch-through* effect), which adds an asymmetric effect to the resolution. For our purposes it is sufficient to parametrise each peak by a Gaussian given as:

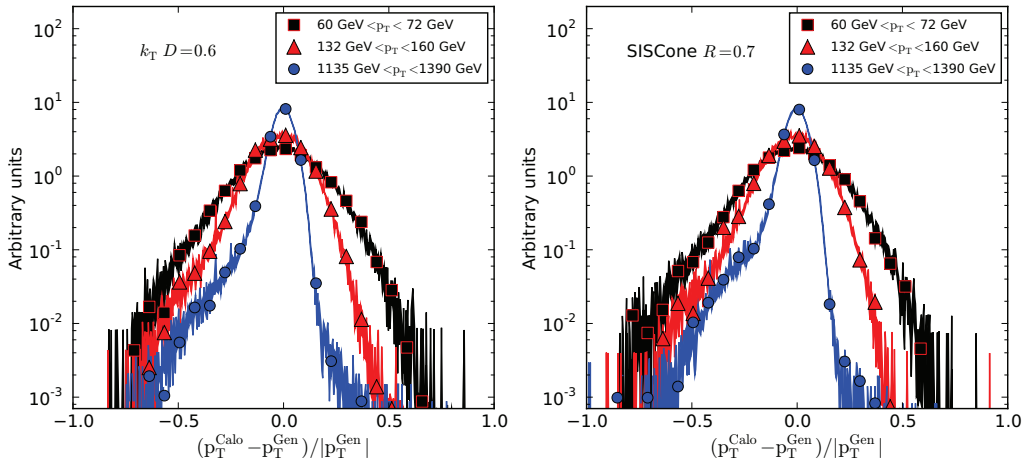

Figure 5.3: *Distribution of the relative deviation of matched particle ("Gen") and JES corrected full simulated jets ("Calo"). For improved clarity, markers are not given for every point. The bandwidth reflects the statistical precision. points.*

$$N_i \frac{1}{\sigma_i \sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right). \tag{5.3}$$

The additional parameter N_i is used, as due to the width of the distribution, the area normalisation of the histograms at low p_T not always reaches the full range within the observed range. At larger p_T the fitting procedure returns N_i as unity. The Gaussians are fitted within ± 1.5 of the root mean squared around the mean of the histograms, for bins with p_T larger than 112 GeV, ± 2.0 of the root mean squared is used. The result of the fitting is demonstrated for dedicated bins in Figure 5.4. The means μ_i of the fitted Gaussians in Figure 5.5 shows to which extent the applied jet calibration is able to correct the jet response on average. For the central rapidity bin a small residual undercalibration of $\approx -1.5\%$ at ≈ 100 GeV is observed, which is reduced to -0.5% at ≈ 1 TeV. The endcap bin shows a little stronger undercalibration of $\approx -2\%$ at ≈ 100 GeV which even transfers into a slide overcalibration at the edge of the reachable phase space with $10\ \mathrm{pb}^{-1}$. The observed effects are well contained within the expected JES uncertainty of 10%.

5.2.3 Jet Energy Resolution

The knowledge of the jet energy resolution is important for the unfolding of the measured differential inclusive jet cross section to the particle level because it leads to a significant distortion of the steeply falling QCD spectrum.

Currently, the jet energy resolution is determined from MC truth by matching the reconstructed jets with the corresponding particle jets, as it is described in the previous section.

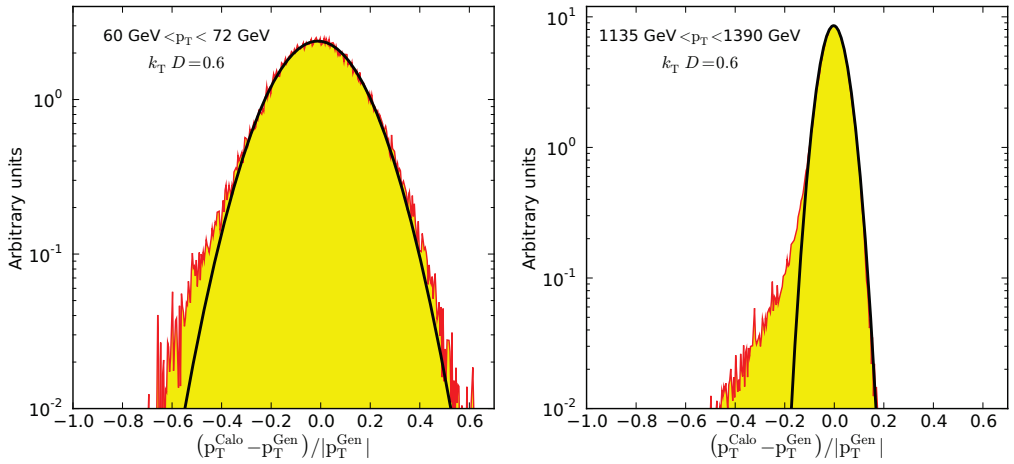

Figure 5.4: *Distribution of relative deviation of matched particle level and JES corrected full simulated jets for two p_{T} bins, together with the fit results for the k_{T} algorithm in logarithmic representation.*

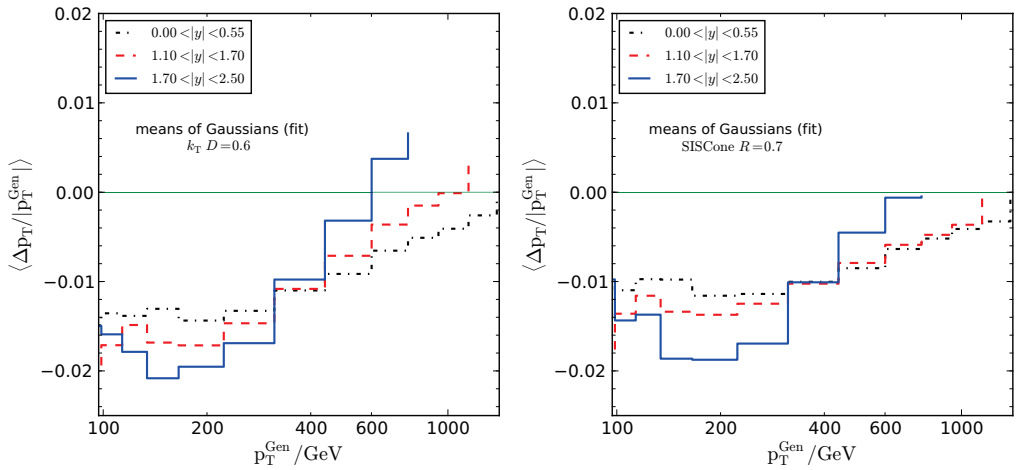

Figure 5.5: *The means of the Gaussian fitted to the p_{T} difference of matched calorimeter and particle level jets for the k_{T} (left) and the SISCone algorithm (right).*

Additional studies of jet properties are given in [149]. The relative resolution $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$, which is given by the σ-parameters of the Gaussians fitted to the detector response, is parametrised as a function of the particle jet p_{T} with the typical calorimetric expression:

$$\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}} = \sqrt{\frac{N^2}{p_{\mathrm{T}}^2} + \frac{S^2}{p_{\mathrm{T}}} + C^2}\,, \tag{5.4}$$

where the parameters C, S, N are given in table 5.3 for the k_{T} and in table 5.4 for the SISCone algorithm. The fit results are rather similar between the two algorithms and shown in Figure 5.6. It should be noted that the procedure outlined above describes only the Gaussian core of the jet energy resolution. As it is shown in Fig. 5.4 there is an observed non-Gaussian behaviour at the tails of the response distributions which becomes visible at more than two orders of magnitude below the peak value. However, a parametrisation of the full response distribution for CMS calorimeter jets is not yet available and thus only the Gaussian core will be used within this work. In the context of the large energy scale uncertainties this shows to be more than sufficient to describe the observed effects. The parametrised jet energy resolution is shown in Figure 5.6 for selected rapidity regions for the case of the k_{T} and the SISCone algorithm. The rapidity resolution is derived with the same method, the result is given in Figure 5.7. In this case the single resolution peaks have a very Gaussian shape and no significant tails could be observed.

As soon as collision data become available, CMS plans to measure the jet energy resolution from di-jet events, utilising the *Di-jet Asymmetry Method* [150]. This method by construction only helps to get a handle on the Gaussian core of the distribution, thus other methods will have to be developed to describe more features of the jet resolution.

rapidity region	N	S	C
$0.00 \leq \|y\| < 0.55$	4.21	1.22	0.0320
$1.10 \leq \|y\| < 1.70$	5.98	1.174	0.0412
$1.70 \leq \|y\| < 2.50$	4.27	0.836	0.0250

Table 5.3: *Jet resolution parameters, from MC truth, in different rapidity regions for the k_{T} algorithm (D=0.6).*

rapidity region	N	S	C
$0.00 \leq \|y\| < 0.55$	3.25	1.28	0.0316
$1.10 \leq \|y\| < 1.70$	4.73	1.26	0.0398
$1.70 \leq \|y\| < 2.50$	4.79	0.823	0.0314

Table 5.4: *Jet resolution parameters, from MC truth, in different rapidity regions for the SISCone algorithm (R=0.7).*

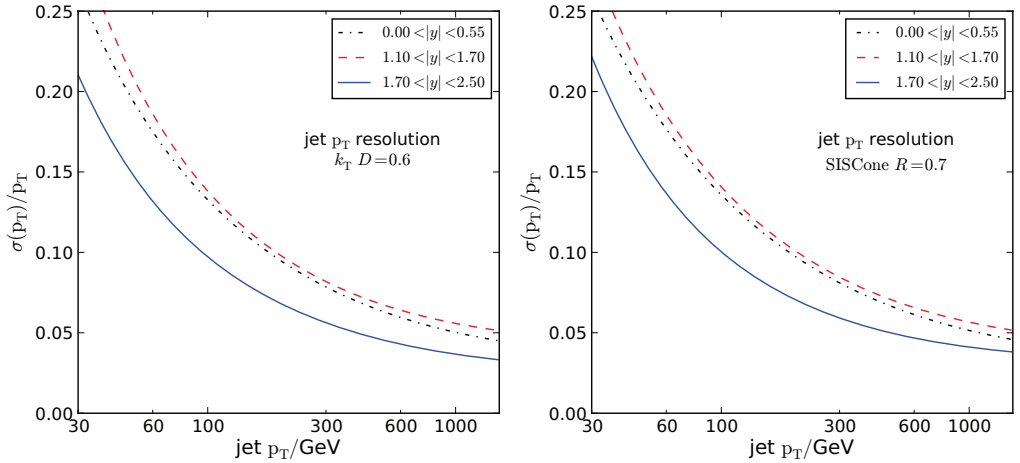

Figure 5.6: *Parametrised relative jet energy resolution as a function of the particle jet p_T for jets reconstructed with the k_T algorithm (left) and SISCone algorithm (right) for three selected rapidity bins.*

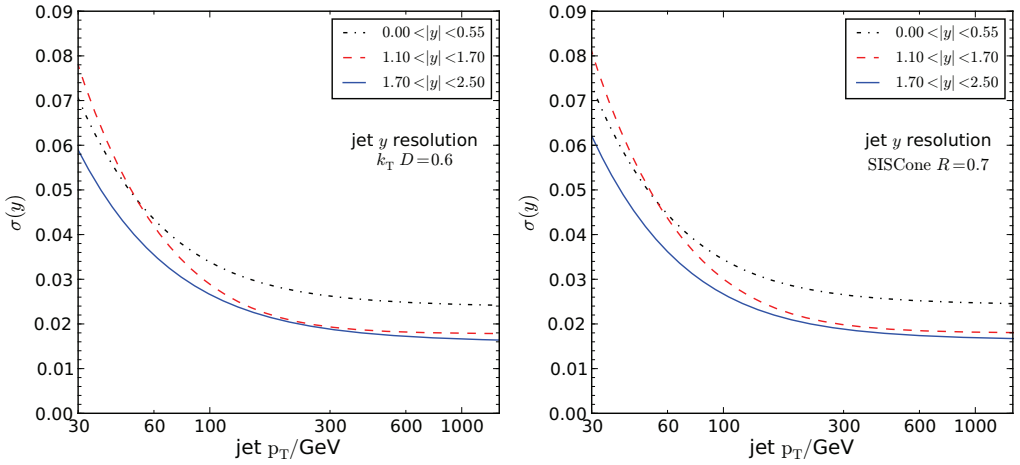

Figure 5.7: *Parametrised jet rapidity resolution as a function of the particle jet p_T for jets reconstructed with the k_T algorithm (left) and SISCone algorithm (right) for three selected rapidity bins.*

5.3 Event Clean-up

In data, large calorimetric signals originating from noise, beam halo energy deposits, or cosmic ray showers will be observed in addition to jets from the hard scattering of the beam protons. All these sources of noise and non-collision data can produce large amounts of transverse energy E_{T} that is not balanced by any partner in a physical scattering process and that appears as so-called missing E_{T} (MET) corresponding to the absolute value of the vector sum of all transverse tower energies. In order to remove the unbalanced events an upper limit on the relative missing E_{T}, MET/$\sum E_{\mathrm{T}}$, where $\sum E_{\mathrm{T}}$ is the scalar sum of transverse tower energies, is planned to be imposed. The distribution of this quantity in QCD simulated events and in cosmic data is shown in Figure 5.8, as it is pointed out in [151]. Due to the finite jet energy resolution, real collision events typically give rise to small imbalances, while noise events are maximally unbalanced in the transverse plane and lead to high values of MET/$\sum E_{\mathrm{T}}$.

Noise rejection studies with the cosmic data indicate that an event clean-up selection of only using events with MET/$\sum E_{\mathrm{T}} < 0.3$ has very high noise rejection power while at the same time preserves almost all events with sufficiently hard jets. In Fig. 5.8 (right) the cosmic data rejection and the efficiency for selecting simulated QCD data are shown as a function of the MET/$\sum E_{\mathrm{T}}$ cut, for a sample of events where the leading jet has raw $p_{\mathrm{T}} > 30$ GeV. The selection efficiency for this sample is $> 95\%$ and dedicated studies indicate that it becomes greater than 99% for events with corrected leading jet $p_{\mathrm{T}} > 100$ GeV.

The preliminary cut value will be re-evaluated when collision data become available. In addition, jet identification criteria are under study in an attempt to further reject fake jets.

5.4 Trigger Requirements and Spectrum Construction

After reconstructing jets from triggered collision events, the next step towards the measurement of the inclusive jet cross section is the combination of the triggered data to form a continuous spectrum. Data will be accumulated with the single jet triggers which fire when the corrected leading jet p_{T} in an event is above a certain threshold. Each single jet trigger path consists of a Level 1 trigger requirement and a High Level Trigger (HLT) condition [147]. Table 5.5 lists the trigger streams that are foreseen to be used for the inclusive jet cross-section measurement. The associated pre-scales correspond to early data taking conditions with an instantaneous luminosity of about $\mathcal{L} = 10^{31}\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In the actual data taking the trigger pre-scales will be measured independently in order to verify that their values are the advertised ones. The turn-on point for each trigger is defined as the value of p_{T} where it becomes at least 99% efficient with respect to the preceding single jet trigger. For this purpose the turn-on curve is described using the parameters a and b within

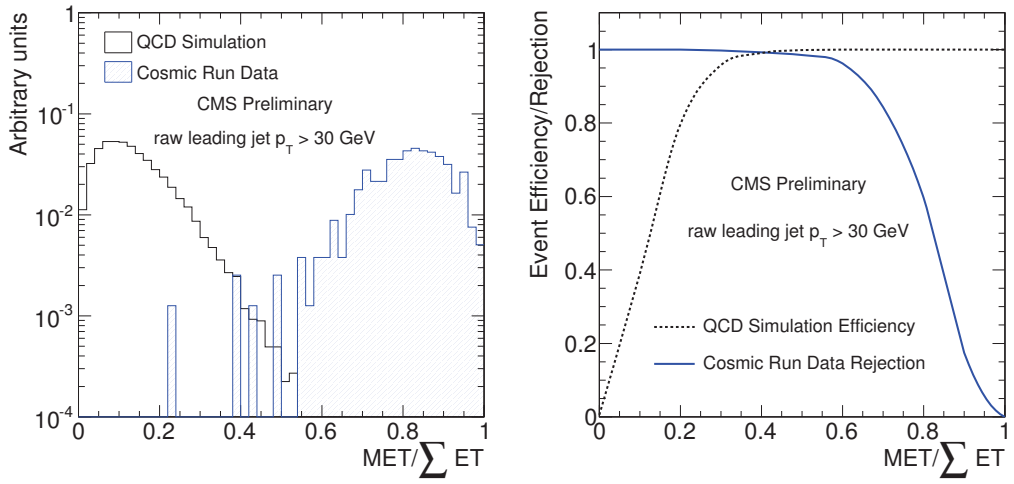

Figure 5.8: *Missing E_{T} over $\sum E_{\mathrm{T}}$ distribution for QCD and cosmic data (left) and QCD event efficiency and cosmic event rejection as a function of the MET/ $\sum E_{\mathrm{T}}$ cut (right) [151].*

trigger	p_{T} threshold (GeV)	pre-scale
$HLT_L1Jet15$	15	10000
HLT_Jet30	30	2500
HLT_Jet50	50	50
HLT_Jet80	80	10
HLT_Jet110	110	1

Table 5.5: *Single jet trigger streams and pre-scales for early data taking conditions. The p_{T} thresholds refer to corrected values.*

the argument $u = ax + b$ of the error function

$$\mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-t^2} dt \,. \tag{5.5}$$

The turn-on point is finally the x value, where the function $\mathrm{erf}(u) = 0.99$. The efficiency of the lowest threshold trigger will be determined by comparison to the Minimum Bias trigger. In Table 5.6 the turn-on point for each trigger is shown for different selected rapidity regions up to $|y| = 2.5$ in the case of the k_{T} algorithm. The left side of Figure 5.10 shows as an example of the turn-on curves in the central rapidity region with $|y| < 0.55$. The corresponding turn-on points for the SISCone algorithm are similar within the range given by the jet energy resolution.

Once the trigger turn-on points are identified, the spectrum in each rapidity bin is constructed from a combination of the trigger streams in such a way that each p_{T} bin receives

contributions from exactly one fully efficient trigger (the one with the highest threshold and therefore smallest pre-scale). The algorithm for the spectrum construction ensures that there is no double counting of events that happened to fire more than one trigger. The contributions to the spectrum out of the single trigger streams is depicted in Fig. 5.10 on the right side. The trigger yield for those streams is given for central rapidity in Figure 5.9 on the left. Additionally the yield for the SISCone algorithm is given on the right. The expected absolute number of jets which go into the spectrum reconstruction, and the directly related relative statistical uncertainty of the inclusive jet cross section including the respective trigger pre-scales are shown for the k_{T} algorithm in Fig. 5.11.

Trigger	$0.00 < \|y\| < 0.55$	$1.10 < \|y\| < 1.70$	$1.70 < \|y\| < 2.50$
HLT_Jet30	44	44	40
HLT_Jet50	57	61	59
HLT_Jet80	87	90	91
HLT_Jet110	116	121	122

Table 5.6: *Turn-on points in corrected jet p_{T} (in GeV) above which each trigger stream is at least 99% efficient with respect to the stream with next lower threshold. The values given here refer to the k_{T} algorithm.*

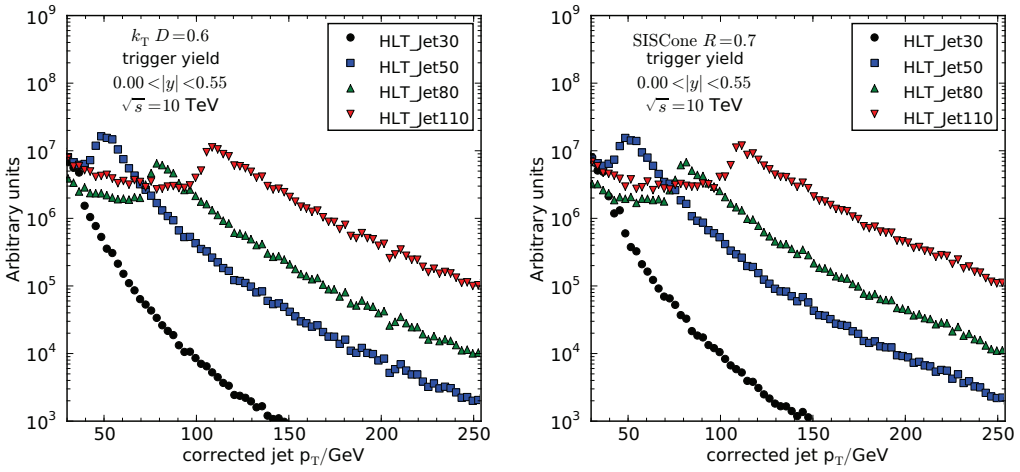

Figure 5.9: *The prescaled trigger yield of all trigger streams in central rapidity is shown for the k_{T} algorithm (left) and SISCone (right). For clearity only every third bin is marked.*

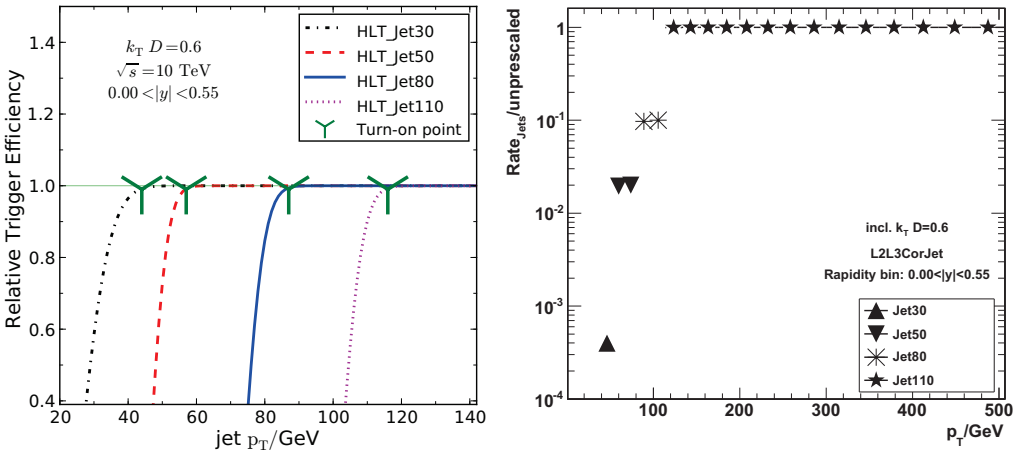

Figure 5.10: *Example: The turn-on for the k_T algorithm in central rapidity for the used trigger streams is given on the left. Each stream is normalised to the preceding one. The markers give the estimated turn-on point from which on the trigger is more than 99% efficient and used in the spectrum. The decomposition of the full spectrum into p_T ranges served by a single trigger stream of the k_T algorithm using the example of central rapidity including the respective pre-scales is depicted on the right.*

5.5 Expected Statistical Precision

This analysis is expected to run the first time with an integrated luminosity of $10\ \mathrm{pb}^{-1}$, thus the statistical uncertainties are evaluated for this scenario. Scaling the jet cross section as given by the PYTHIA particle jet cross section to the expected number of jets, and taking into account the trigger pre-scales from Tab. 5.5, the number of expected jets and the relative statistical precision is evaluated. The result for the k_T algorithm is depicted in Fig. 5.11. It can be seen, that the expected statistical uncertainty ranges from some per mille at low jet p_T and reaches the 10% level for central rapidity at $\approx$ 800 GeV, which is already above the Tevatron reach of $\approx$ 600 GeV. The expectations for the SISCone algorithm do not significantly differ.

5.6 Experimental Corrections

The jet yield measured from calorimeter towers requires corrections to account for inefficiencies and, non-linearities of energy measurements. Additionally finite resolutions in energy and direction of the detectors, as well as the event and jet selections have to be corrected for. This section is dedicated to the most important corrections for the jet energy scale and the jet energy resolution, the associated uncertainties will be discussed in section 5.7.

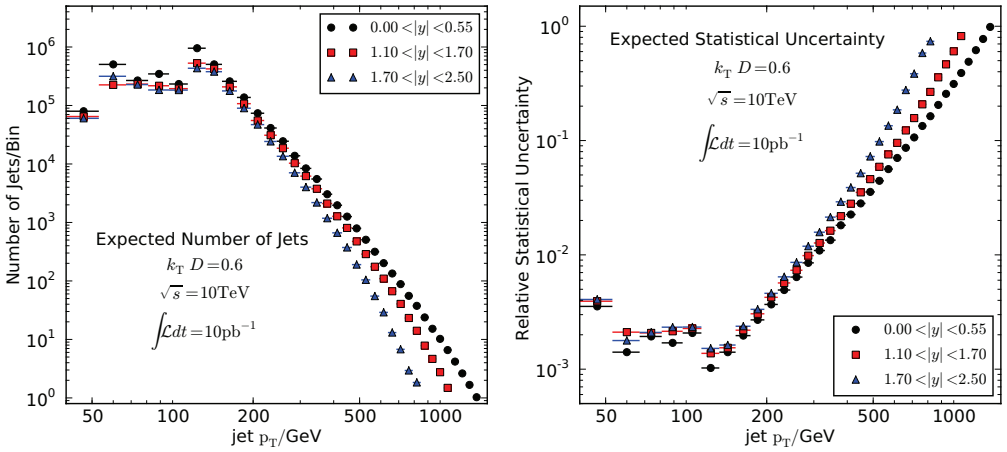

Figure 5.11: *Expected number of jets/* GeV *(left) and relative statistical uncertainty (right) for the* k_T *algorithm with* $10\ \text{pb}^{-1}$ *of integrated luminosity estimated with fully hadronised* PYTHIA *events. Expected trigger pre-scales are taken into account in both cases.*

5.6.1 Jet Energy Scale

The most important treatment for the jets is the energy calibration which adjust the p_T of the measured jet so that on average inefficiencies and losses are accounted for. Despite the large size of the correction the application of the JEC recovers most of the particle level differential cross section and the resulting spectrum is denoted as partially corrected. It is already reasonably close ($O(10\%)$) to the final result. The left side of Figure 5.12 shows the observed yield for uncorrected jets relative to the particle level from MC truth, and on the right side the same ratio after the JEC is displayed. All ratio plots are cleaned of statistical fluctuations by fitting the Ansatz function (Equation 5.6) before dividing.

5.6.2 Resolution Unsmearing

The measured inclusive jet cross section vs p_T is the convolution of the actual particle jet spectrum, folded with the finite p_T and y resolutions of the detector. While the effect of the y resolution can be neglected to first order, the p_T smearing effect must be corrected for. The unsmearing corrections for the jet p_T spectra in these studies are derived using the *Ansatz Method*, which has been successfully employed at the Tevatron [64, 135] is described below.

The starting point for this method is a functional description of the unknown particle jet cross section:

$$f(p_T) = N \cdot p_T^{-a} \cdot \left(1 - \frac{2\cosh(y_{min})p_T}{\sqrt{s}}\right)^b \exp(-\gamma p_T)\ . \tag{5.6}$$

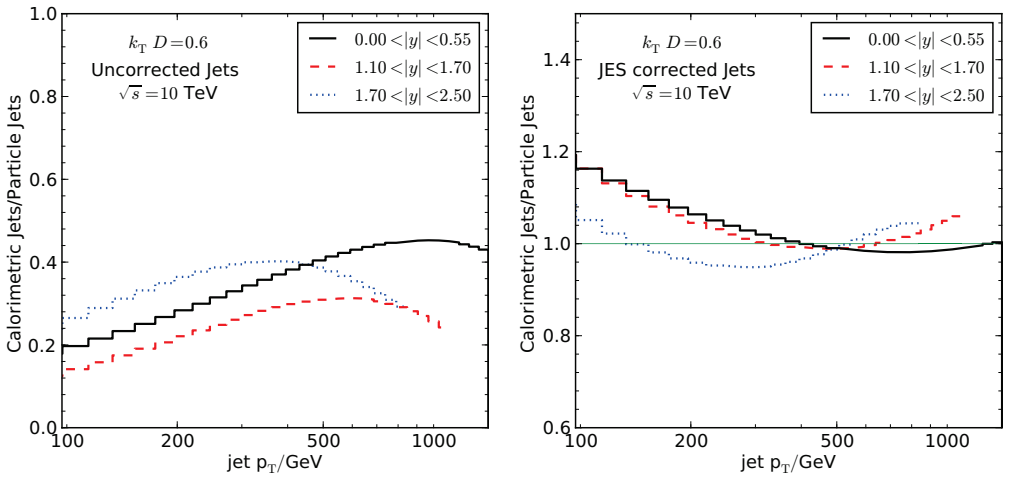

Figure 5.12: *Fractional differential jet yield for uncalibrated (left) and JES corrected calorimeter jets (right) normalised to the MC truth.*

This particular function is theoretically motivated. The $\left(1 - \frac{2\cosh(y_{min})p_T}{\sqrt{s}}\right)^b$ term reflects the behaviour of the parton densities at high proton momentum fractions x as well as the cross section dependence on rapidity. The p_T^{-a} imitates the QCD matrix element. The measured cross section is defined as the convolution of the particle jet spectrum with the detector resolution:

$$F(p_T) = \int_0^\infty f(p'_T) R(p'_T, p_T) dp'_T \, . \tag{5.7}$$

In order to apply this method it is necessary to model the p_T resolution. The simplest assumption is a Gaussian approximation:

$$R(p'_T, p_T) = \frac{1}{\sqrt{2\pi}\sigma(p'_T)} \exp\left[-\frac{(p_T - p'_T)^2}{2\sigma^2(p'_T)}\right] , \tag{5.8}$$

where $\sigma(p'_T)$ is a parametrisation of the relative σ dependence of p'_T (Equation 5.4).

Although the Gaussian model for the resolution is reasonable enough, it fails to describe the full shape, i.e. the tails as shown in Fig. 5.3. Due to the steeply falling nature of the inclusive jet p_T spectrum, the mismodelling of the resolution tails introduces a systematic bias on the unsmearing correction. In case of high p_T tails, this would strongly enhance the effect, however, the observed low p_T tails for large jet p_T will not have such a large impact. More details on this will be revealed in a final comparison to particle jets in section 5.7.4.

Once the measured spectrum is fitted with the smeared Ansatz function, the unsmearing correction is calculated as a multiplicative factor for every bin as:

$$C_{bin} = \frac{\int_{bin} f(p_{\mathrm{T}})dp_{\mathrm{T}}}{\int_{bin} F(p_{\mathrm{T}})dp_{\mathrm{T}}} \,. \tag{5.9}$$

In Figure 5.13 the measured cross sections in different rapidity bins are fitted successfully with the smeared ansatz function as indicated by the fit quality plots (Fig. 5.14). The final unsmearing correction factors are shown in Figure 5.15. It can be seen, that the corrections are about 20% at low p_{T} and central rapidity, and decrease to 5% at higher p_{T}. As the relative resolution in the end cap is better, also the unsmearing corrections show this behaviour. At higher p_{T} in the more forward rapidity bins, the spectrum becomes more steeply falling which leads to larger corrections in this region.

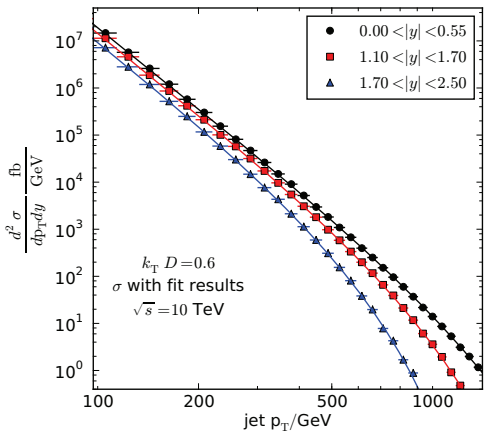

Figure 5.13: *JES corrected inclusive jet cross section vs p_{T} for selected rapidity ranges, fitted with the smeared ansatz function.*

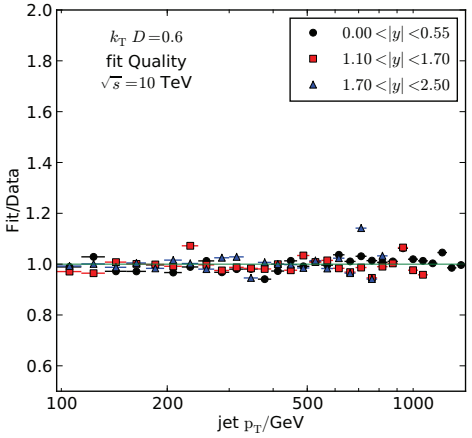

Figure 5.14: *Quality of the fit: fractional difference between the fitting curve and the "data" points.*

Finally these corrections are applied to the partially corrected spectrum (after JEC), which results in the spectra given in Fig. 5.16.

5.7 Experimental Systematic Uncertainties

The major sources of systematic uncertainty for the cross-section measurement are the jet energy scale (JES), the luminosity and the jet energy resolution (JER), while an initially negligible contribution comes from the jet angular resolution.

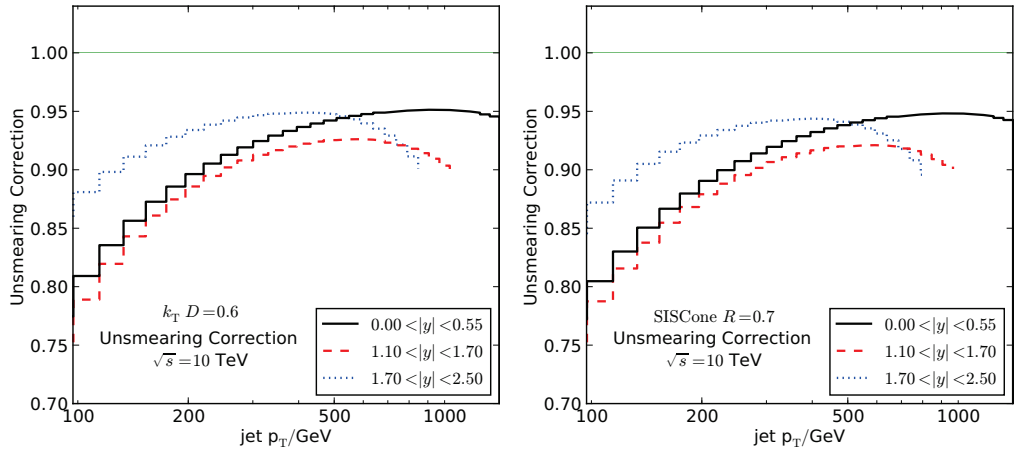

Figure 5.15: *Unsmearing correction factors for different rapidity ranges.*

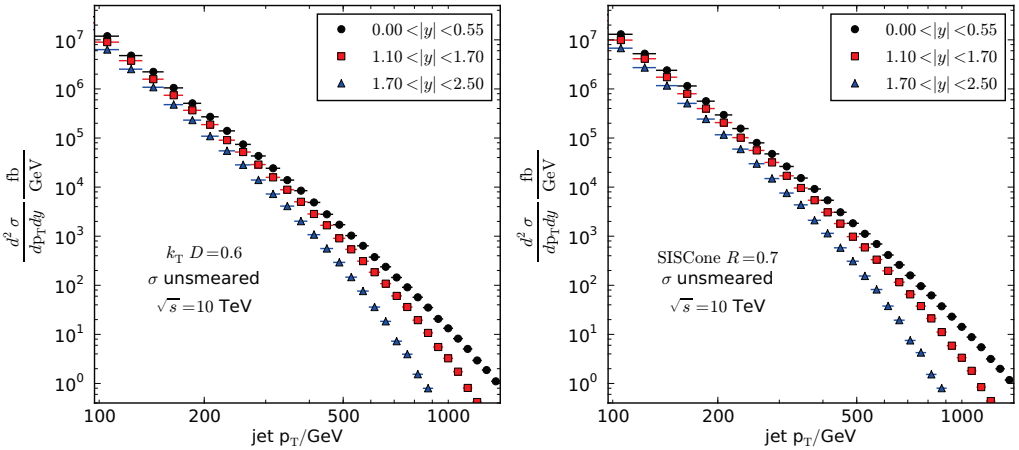

Figure 5.16: *Inclusive jet cross section vs p_T corrected for energy scale and unsmeared for different rapidity ranges for the k_T algorithm (left) and SISCone algorithm (right) .*

5.7.1 Jet Energy Scale Uncertainty

Due to the fact that the analysis presented here is targeting the first $10\ \mathrm{pb}^{-1}$ of data taking, it becomes clear that the JES uncertainty is by far the dominant one.

The inclusive jet cross section measurement is sensitive to the JES due to the steeply falling nature of the spectrum with increasing $p_{T,\mathrm{jet}}$. A rough estimate of the dependence of the cross section uncertainty on JES can be obtained from the approximate expression

$d\sigma/dp_{\mathrm{T}} \sim p_{\mathrm{T}}^{-a}$ which yields the uncertainty as:

$$\frac{\delta(d\sigma/dp_{\mathrm{T}})}{d\sigma/dp_{\mathrm{T}}} \sim -a\frac{\delta p_{\mathrm{T}}}{p_{\mathrm{T}}} \; . \tag{5.10}$$

Given that $a \sim 10$, it can be derived that a 10% JES uncertainty is translated into an uncertainty in the jet cross section of $\sim 100\%$. For the observed spectrum the gradient changes with p_{T}, therefore even the constant JES uncertainty does not translate as a constant fraction. In the context of the presented analysis the scenario proposed by the CMS JetMET group is evaluated. According to the studies performed it is suggested that a flat 10% JES uncertainty is the most educated guess for the start-up of the experiment [137].

One way to treat the JES uncertainty is to vary explicitly the jet calibration constant, according to each scenario, on a jet by jet basis and repeat the cross section measurement. However, although this method is straightforward, it suffers from the statistical fluctuations, especially at higher p_{T}. Alternatively, one can use the smeared ansatz function which has been fitted to the measured spectrum (Equation 5.7). Each jet p_{T} bin is determined by its boundary values p_{T}^{l}, p_{T}^{h} and the measured cross section is:

$$\left(\frac{d^2\sigma}{dp_{\mathrm{T}}dy}\right)_{bin} = \frac{1}{p_{\mathrm{T}}^{h} - p_{\mathrm{T}}^{l}} \int_{p_{\mathrm{T}}^{l}}^{p_{\mathrm{T}}^{h}} F(p_{\mathrm{T}})dp_{\mathrm{T}} \; . \tag{5.11}$$

For shifted jet p_{T} values due to the JES uncertainty, the measured cross section is:

$$\left(\frac{d^2\sigma}{dp_{\mathrm{T}}dy}\right)^{\pm}_{bin} = \frac{1}{p_{\mathrm{T}}^{h} - p_{\mathrm{T}}^{l}} \int_{\left[1\pm\delta(p_{\mathrm{T}}^{l})\right]p_{\mathrm{T}}^{l}}^{\left[1\pm\delta(p_{\mathrm{T}}^{h})\right]p_{\mathrm{T}}^{h}} F(p_{\mathrm{T}})dp_{\mathrm{T}} \; , \tag{5.12}$$

where $\delta(p_{\mathrm{T}})$ is the JES uncertainty according to the particular scenario. Using the above results the fractional systematic uncertainty of the jet cross section is

$$\alpha(\pm) = \frac{\left(\frac{d^2\sigma}{dp_{\mathrm{T}}dy}\right)^{\pm}_{bin}}{\left(\frac{d^2\sigma}{dp_{\mathrm{T}}dy}\right)_{bin}} - 1 \; . \tag{5.13}$$

The fractional systematic uncertainties due to the JES are shown at the end of the section in Figures 5.17 and 5.18, in two different rapidity bins, and in comparison to the other leading experimental uncertainties for the k_{T} and SISCone algorithm. The uncertainty on the cross section grows for higher rapidities, because the steepness of the spectrum increases even more.

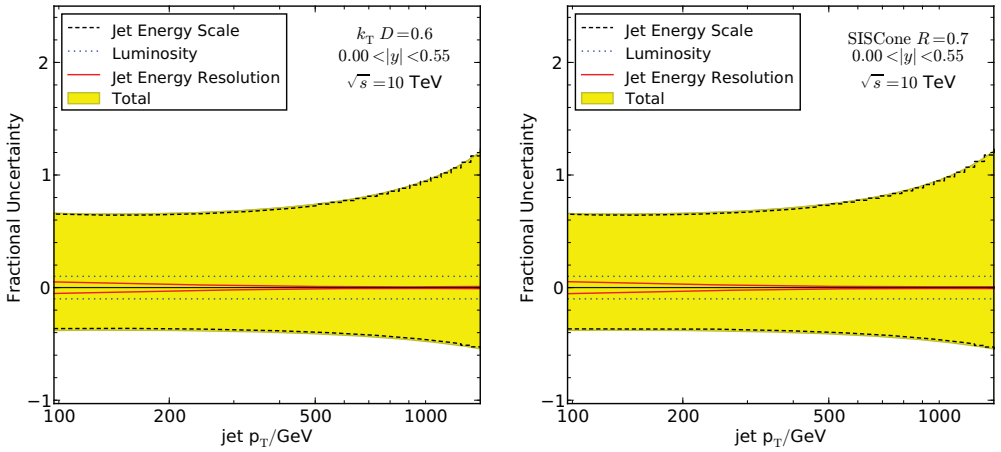

Figure 5.17: *The dominant experimental uncertainties due to the jet energy scale, the jet energy resolution as well as the luminosity is given for the central rapidity bin for both jet algorithms.*

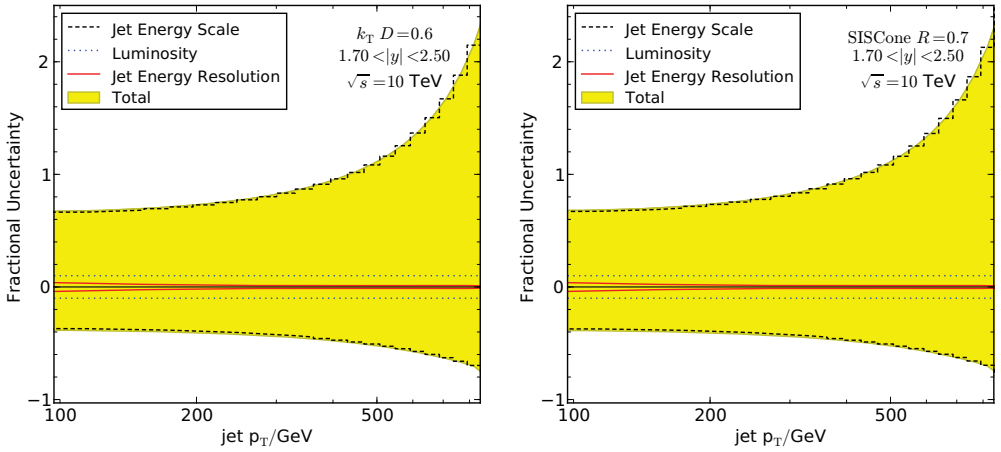

Figure 5.18: *The dominant experimental uncertainties due to the jet energy scale, the jet energy resolution as well as the luminosity is given for the forward region and both jet algorithms.*

5.7.2 Luminosity

For the early data taking period a 10% uncertainty on the luminosity and hence on the normalisation of the inclusive jet cross section is assumed. The uncertainty is directly propagated into the measurement as a flat contribution of 10%.

5.7.3 Unsmearing Uncertainty

The unsmearing procedure relies on a proper knowledge of the jet p_{T} resolution. Therefore the uncertainties on the jet p_{T} resolution propagate through to the unsmearing corrections. A determination of the unsmearing uncertainty thus requires varying the resolution parameters used in the unsmearing procedure within a given uncertainty range. Motivated by the study of the jet p_{T} resolution measurement with the Asymmetry Method [150] an uncertainty of 10% on the knowledge of the relative jet resolution is assumed.

The unsmearing is thus performed using a $\sigma_{Gaussian}$ which is varied by $\pm 10\%$ around the parametrisation given in section 5.2.3. The resulting effect on the cross section ratios is included in Figure 5.17 for the k_{T} algorithm and in Figure 5.18 for the SISCone algorithm together with the other uncertainties. The variation of the Gaussian width leads to a symmetric misestimation of the cross section by about four percent at $100\ \mathrm{GeV}$ down to one percent at one TeV.

After having corrected for a smearing due to the jet p_{T} resolution, there is the question about the dependence on the finite resolution in y. In Figure 5.19 the effect is studied by smearing y of the jets with a Gaussian. The width is taken to be 0.035 for $|y| < 1.70$, 0.025 for outer rapidities. These values are are slightly above the residual value, that the rapidity resolution shows above $100\ \mathrm{GeV}$ (Fig. 5.7). From the Figure it can be seen, that such a smearing almost leaves the cross section ratios in the central region untouched. In the transition and endcap regions the effects are about one percent, which is below the statistical uncertainties due to the Monte Carlo modelling. Therefore this effect can be neglected. Not surprisingly, the deviations rise at the kinematic and statistical limits at large p_{T} for the transition and endcap region.

5.7.4 Comparison to Particle Jets

Finally the effect of all corrections for detector effects on the inclusive jet cross section described in the previous sections are analysed by comparing the corrected calorimetric cross sections to the particle jet cross section. Ideally all ratios shown in the following should give a flat line closing at 1.0. Figure 5.20 shows the ratio of the corrected inclusive jet cross section normalised to the particle jet cross section. After applying all corrections finally a flat spectrum is retained, which for the central rapidity bin is at $\approx 90\%$ for the whole p_{T} range. Achieving closure here is a delicate interplay of having the correct JES correction and estimating the correct JER. As shown while discussing the experimental systematic effects, already an undercalibration of 1% can lead to such deviations. In Fig. 5.5 a deviation in the mean jet p_{T} response of one to two percent for the central rapidity bin is observed. A measurement of this curve to the same accuracy from data over the whole p_{T} spectrum will most likely never be available. However, a linear interpolation of the observed deviations is now used for demonstration purposes to correct the jet transverse momenta on a jet-by-jet basis. The result shows that the closure improves significantly. It can be observed, that the

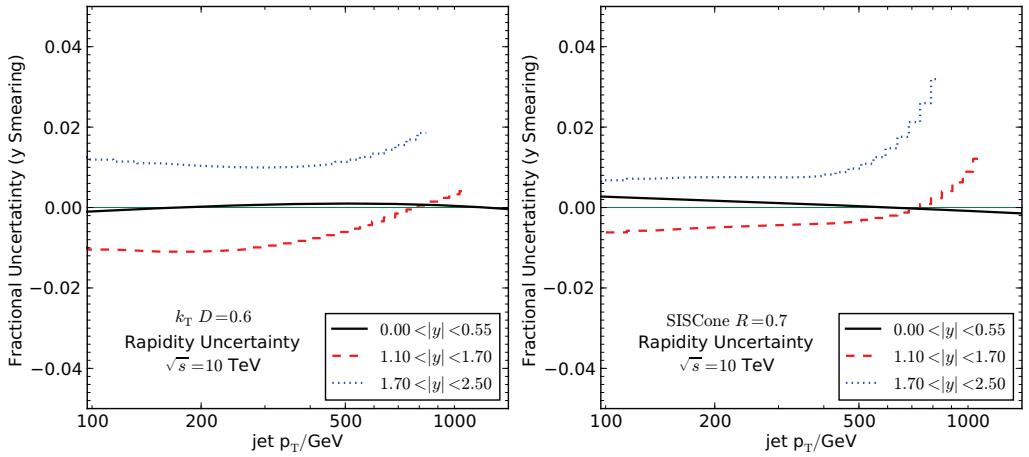

Figure 5.19: *The fractional uncertainty due to a smearing of y is evaluated by artificially smearing y of the jets by a Gaussian. A width of* 0.035 *has been used for* $|y| < 1.7$, 0.25 *for outer rapidities. The effect is shown to be at the order of one percent, larger differences are observed at the kinematic and statistical limits.*

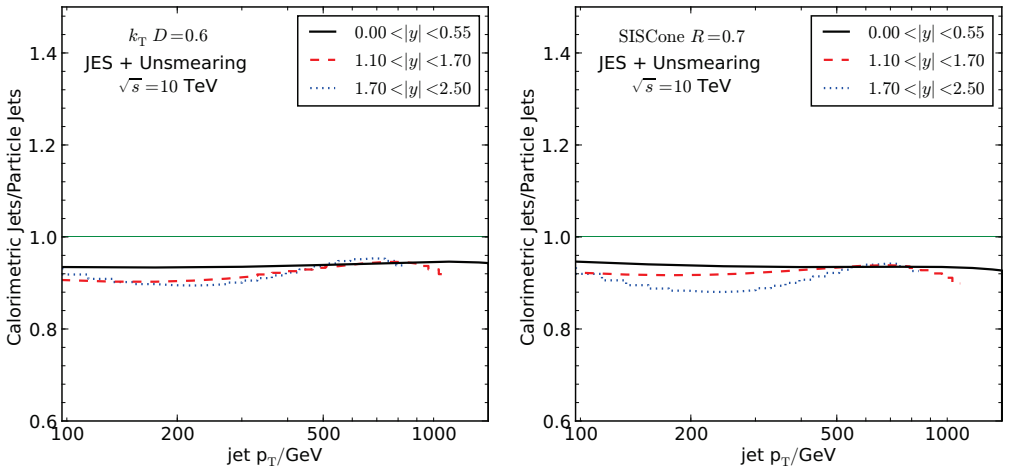

Figure 5.20: *Relative differential jet rate for JES corrected and unsmeared calorimeter* k_T *jets (left) and SISCone jets (right) normalised to the particle jet rate.*

additional corrections move the comparison to the 1.0 line almost over the whole p_T range (Fig. 5.21). Only at large p_T still some deviations remain. These are enlarged for outer rapidities.

A mismodelling of the resolution within the unsmearing procedure could also cause deviations in this comparison. In section 5.2.2 a slight asymmetric deviation of the Gaussian

shape for higher p_{T} was observed. As the Gaussian does not account for this, the derived corrections might be slightly smaller than the required ones. In order to include this observation, in Figure 5.22 the Gaussian means derived from the fitting procedure are exchanged by the means of the resolution histograms for $p_{\mathrm{T}} > 614\ \mathrm{GeV}$. Compared to the previously used corrections, the means of the histograms show more undercalibration than the means of the fitted Gaussians. The Figure also contains lines, which interpolate between the bins. These curves have been used to additionally correct the spectrum. The result in Figure 5.23 now also closes at larger p_{T} for central rapidity. The transition and endcap regions still show some deviations at high p_{T}, where the steeper spectra enhance the differences. The remaining differences at intermediate and low p_{T} are well within the fluctuations of the sample, which is also reflected by the deviations in the fit quality given in Figure 5.14.

So the estimated jet energy resolution together with a well motivated additional jet energy correction finally lets us correct for the dominating systematic effects up to several percent with the presented method. Therefore this demonstrates, that the applied unsmearing procedure itself works, given a proper input for the JES corrections and $\sigma_{\mathrm{Gaussian}}$ parametrisation. From the successful closure it was found, that the region between 50 and $600\ \mathrm{GeV}$ can be well described using a Gaussian description. The observed deviation due to the non-Gaussian tails were much smaller than the experimental uncertainties, and only emerged for large jet p_{T}, therefore the resulting additional uncertainties are regarded to be already included in the 10% JES uncertainty. So the presented unsmearing procedure can well be used within a start-up scenario. At larger p_{T}, at least for central rapidity, differences between using the Gaussian means and using the histogram means indicate, that probably in this region an improved JER description might improve the procedure. The fact, that the effect could also be corrected by an average shift in the JES without modifying the JER shape, shows the importance of controlling the JES first.

The additional corrections which could be derived given MC truth information and using the exact same event generator and detector simulation, cannot be obtained from real data. Thus it is important to use all possible data driven methods to improve the jet energy scale corrections and the Monte Carlo description. As soon as the JES uncertainty could be pushed to a level of a few percent, also an improved description of the resolution within an unsmearing procedure will be required, which will then also include an unfolding for the y resolution. This especially becomes more important when extending the analysis to larger rapidity bins. For all further presentations of the cross section only the standard JES corrections will be applied, as currently there are no methods available to derive the additional corrections from data. Dedicated γ-jet studies might shed some light into this, however the interesting regions at high p_{T} cannot be probed by this. Additional improvements could be gained exploiting tracks of charged particles, this however, is restricted to the fiducial volume of the tracker ($|y| < 2.5$) and also to jets with low and intermediate p_{T}. For advanced future studies such corrections might therefore only be derived from a very well tuned Monte Carlo simulation.

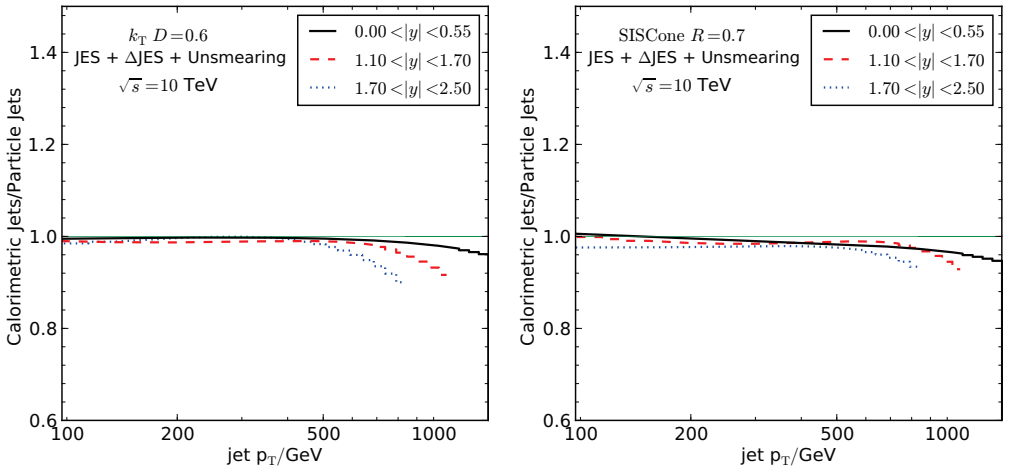

Figure 5.21: *Relative differential jet rate for calorimeter k_T jets normalised to the particle jet rate. The calorimeter jets have standard JES corrections and an additional correction derived from Figure 5.5 applied. Finally the unsmearing procedure was applied.*

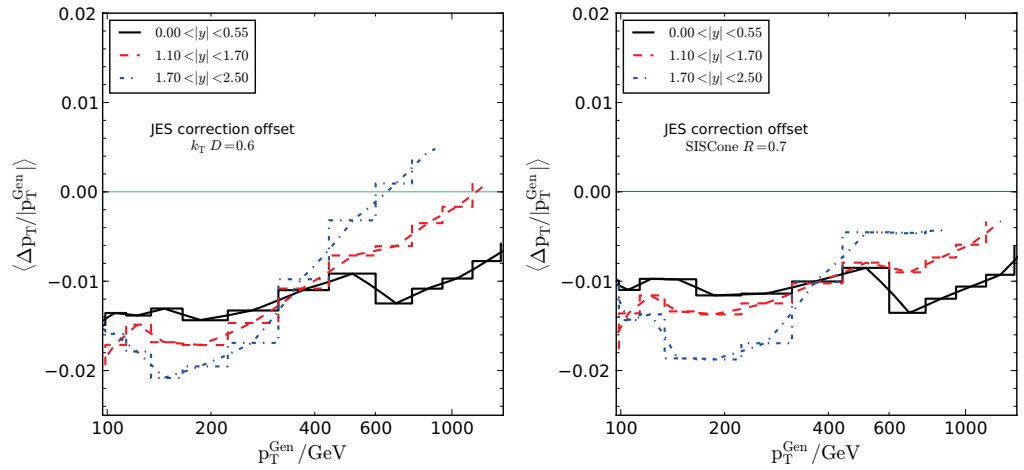

Figure 5.22: *The means of the Gaussian fitted to the p_T difference of matched calorimeter and particle jets for the k_T (left) and the SISCone algorithm (right) for $p_T < 614$ GeV. The mean of the total response histograms is used above. The additional lines indicate the interpolation used within the additional jet correction.*

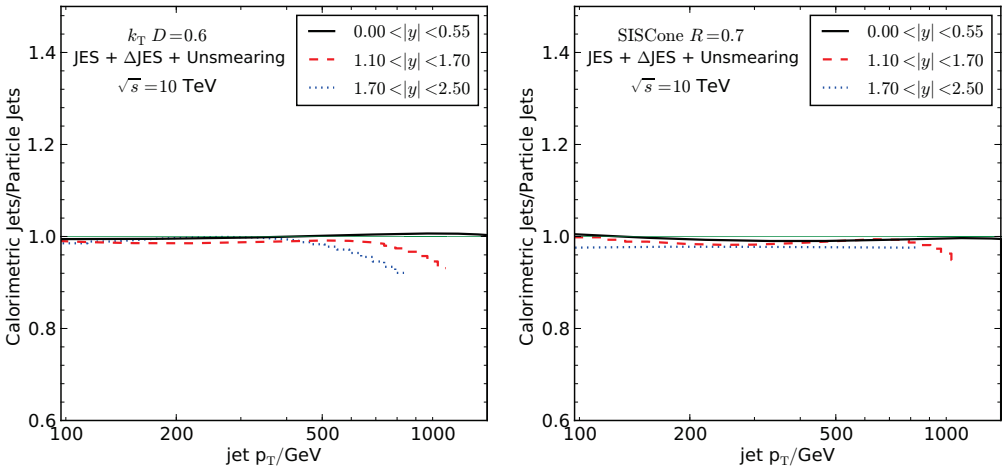

Figure 5.23: *Relative differential jet rate for calorimeter k_T jets normalised to the particle jet rate. The calorimeter jets have standard JES corrections and an additional correction derived from Figure 5.22 applied. Finally the unsmearing procedure was applied.*

5.8 Theory Prediction Summary

The currently best theoretical predictions for inclusive jet measurements are next-to-leading order calculations in perturbative QCD. For this analysis they have been performed in a dedicated setup with the same binning as used in the described measurement. Uncertainties due to factorisation and renormalisation scales as well as due to PDFs have been evaluated. Electroweak corrections, which might also contribute significantly for high p_{T} have not been regarded. The procedure is in detail described in Chapter 4.

Lacking collision data, simulations employing the leading-order Monte Carlo generator PYTHIA are taken as a substitute. In contrast to what would be done with real data, the pseudo data require an additional treatment in order to be compared with the NLO theory. This is demonstrated in Fig. 5.24 where the ratios of the inclusive jet cross section of NLOJET++ in NLO and of the different stages of the QCD di-jet event generation of PYTHIA are shown with respect to NLOJET++ in LO for the k_{T} algorithm. To account for the difference in pQCD precision (LO vs. NLO), employed parton density (CTEQ6L1 [114] vs. CTEQ6M) and the order of the evolution of the strong coupling α_S (1-loop vs. 2-loop), K factors analogous to the NLOJET++ NLO over NLOJET++ LO curve of Fig. 5.24 have been applied to the PYTHIA cross sections. The K factors are presented in Fig. 5.25.

Following the procedure adopted for the Tevatron measurements of inclusive jet cross sections [33, 64] the NLO predictions are modified by additional non-perturbative corrections which correspond to the difference between the "PYTHIA (LO and pert. corr.)" and the "PYTHIA particle jets" curves in Fig. 5.24. The effect of the PYTHIA hadronisation model alone can be judged by comparing "PYTHIA unhadronised" with "PYTHIA particle jets". More details including HERWIG++ as alternative MC generator are given in the next section.

In addition, Fig. 5.24 illustrates the size of perturbatively motivated corrections (pert. corr.) in PYTHIA due to initial and final state radiation as well as parton showers. The deviation of the PYTHIA + pert. corr. curve from one, i.e. the LO reference, leads to a residual underestimation of the inclusive jet cross section in the comparison of the PYTHIA pseudo data which is visible for example in Fig. 5.27 and Fig. 5.30 left.

5.9 Non-perturbative Corrections

In order to go from the partons of a NLO calculation to final state hadrons the additional steps of parton showering, hadronisation, decays, and multiple parton interactions have to be performed. In particular for hadronisation and multiple parton interactions only phenomenological models exist that currently can solely be used together with LO matrix elements for the inclusive jet cross section.[1] Therefore correction factors for the non-perturbative steps have to be applied to the NLO result as explained in Chapter 4. For this

[1] In MC@NLO [123, 152] the QCD process required for inclusive jets unfortunately is not yet implemented.

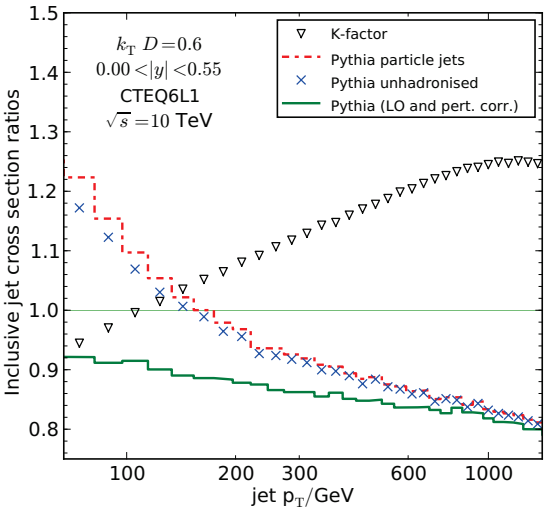

Figure 5.24: *Inclusive jet cross section ratios of* NLOJET++ *in NLO and of the different stages of the QCD di-jet event generation of* PYTHIA *with respect to* NLOJET++ *in LO for the* k_T *algorithm.*

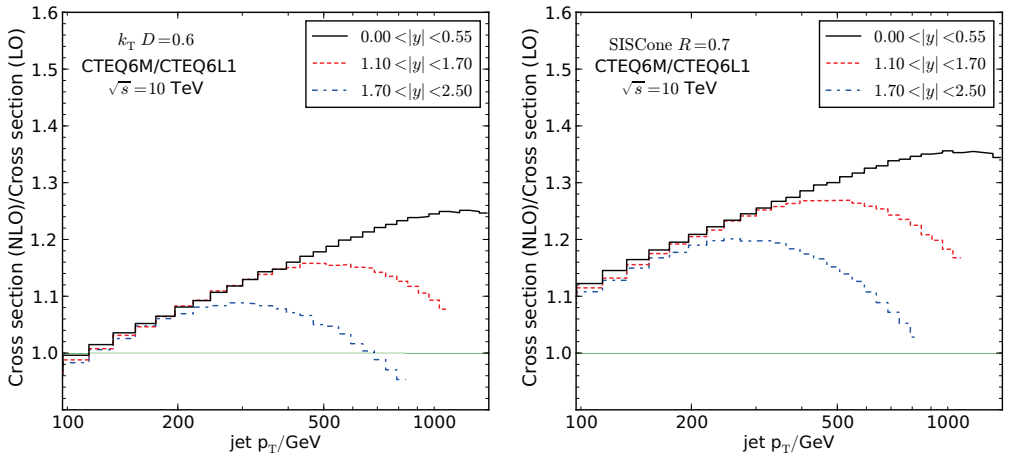

Figure 5.25: K *factors applied to the* PYTHIA *pseudo data in order to account for the difference in pQCD precision, PDF and loop order of the evolution of* α_S *for the* k_T *(left) and the SISCone algorithm (right).*

analysis the correction factors shown in Fig. 5.26 are determined as the average between the predictions derived from the PYTHIA and HERWIG++ [49, 125] event generators where half the spread between the two is adopted as the associated systematic uncertainty. Overall we observe smaller correction factors but larger uncertainties for the SISCone algorithm.

Once first tunes of the MC generators at LHC energies are available, these corrections and their uncertainty have to be revisited. More details on the procedure of deriving the corrections are presented in Chapter 4.

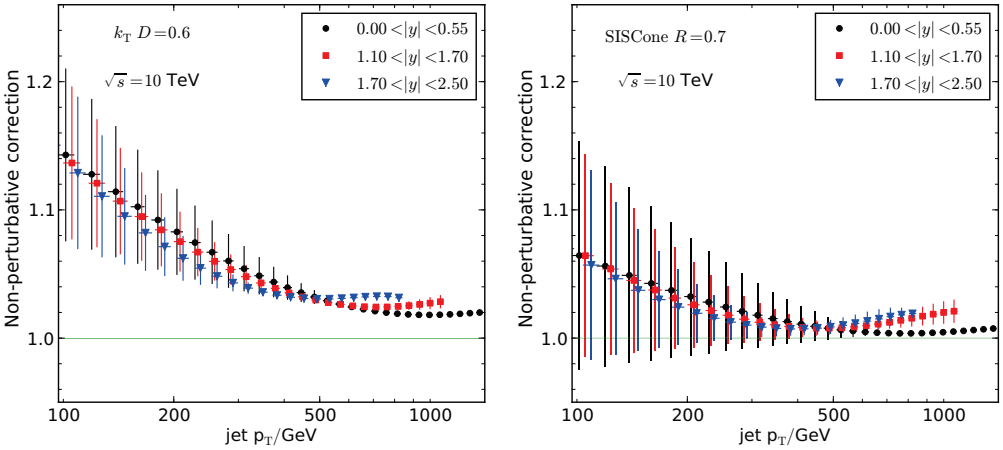

Figure 5.26: *Non-perturbative corrections to NLO QCD calculations for the k_T (left) and the SISCone algorithm (right). The error bars correspond to half the spread between the predictions from* PYTHIA *and* HERWIG++*. For improved readability, the points are drawn shifted with respect to the plotted x-value: the red points with the square marker are centred to the true value, the others are shifted by ± 4 GeV.*

5.10 Final Comparison

Finally a comparison of the inclusive jet cross section is given for both the k_T as well as the SISCone algorithm between NLO and pseudo data derived from full detector simulation and detector specific corrections applied. In order to achieve a fair comparison, the simulated measurement as well as the NLO results have the respective theoretical corrections applied, which have been discussed in the previous section.

Figure 5.27 shows the comparison between the proposed measurement and the theory predictions. As expected from Figure 5.24 and 5.20 it is observed, that the NLO calculation exhibits a slightly higher cross section than the one retrieved from our pseudo data. In addition, the hadronisation corrections are the ones one would use for a comparison with real data, thus they do not correspond to a pure PYTHIA modelling. A summary of all considered theoretical systematic uncertainties is given in Fig. 5.28 for two different rapidity bins and for both jet algorithms. The scale uncertainty in NLO has been evaluated by varying the renormalisation and factorisation scale simultaneously from the default setting of $p_{T,\mathrm{jet}}$ to $p_{T,\mathrm{jet}}/2$ and $2 \cdot p_{T,\mathrm{jet}}$. As these uncertainties are mainly uncorrelated, their total effect is the quadratic sum of each subcontribution.

A comparison to the quadratic sum of the experimental uncertainties is depicted in the Figure 5.29. The attached vertical lines represent the statistical uncertainty for an estimated luminosity of $10\ \mathrm{pb}^{-1}$, estimated from the PYTHIA particle level jet cross section. The additional lines show the total uncertainty assuming a JES uncertainty of six and three percent, to give an idea how this picture might evolve in the future. After years of gaining experience with their detectors, the Tevatron collaborations currently publish their measurements of the inclusive jet cross section with a given JES uncertainty of about three percent. In the presented study, leaving all other contributions untouched, this means, that the ten percent luminosity uncertainty already claims 50% of the given total experimental uncertainty at low p_{T}. However, one might expect to also reduce this number by some percent, the TOTEM experiment even claims final uncertainty of 1%. Therefore it can be expected, that the given theoretical uncertainties could be touched. It should be remarked, that such an optimistic JES uncertainty would be really hard to reach above $p_{\mathrm{T}} \approx 600\ \mathrm{GeV}$, where γ-jet or other methods cannot provide a direct measurement of the jet energy scale and one would have to rely on extrapolation methods, like using two low p_{T} jets balanced against one high p_{T} jet [153].

For the early data assumptions, the estimated experimental systematic effects are dominating over the theoretical uncertainties demonstrating that the measurement is systematically limited. Nevertheless the consistency with the underlying theory can be tested and with increasing integrated luminosity and a more precise jet energy scale determination, the theoretical uncertainties will be probed. The statistical uncertainty becomes significant only at higher jet transverse momenta that are beyond the reach of any previous experiment. In this region, signs of new physics that produce a large deviation from the QCD predictions such as contact interactions due to quark compositeness (Fig. 5.30, [151, 154]) can be sought for, even with relatively large experimental systematic uncertainties.

The central findings of this study, which has been presented here in an updated version, are also publicly available within a Physical Analysis Summary of the CMS Collaboration [151].

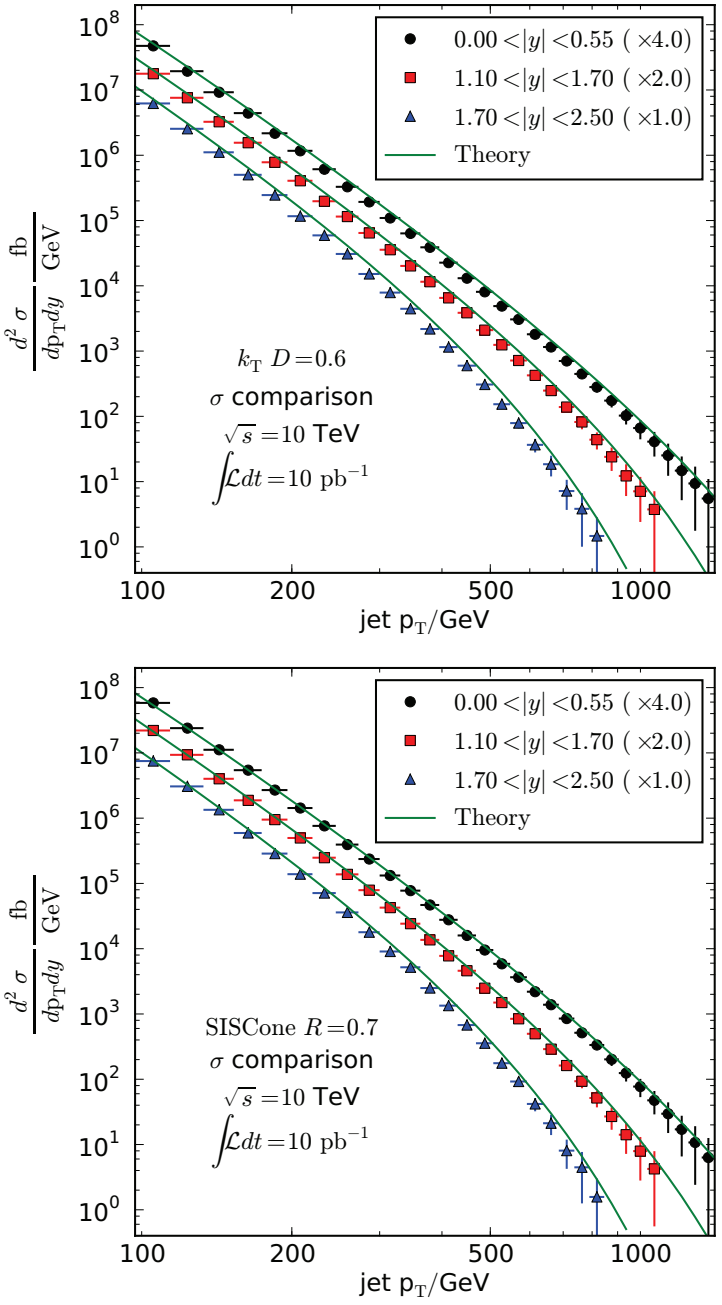

Figure 5.27: *Comparison between the corrected measured spectra and the theory predictions for the k_T (top) and the SISCone (bottom) algorithm. For better visibility the spectra have been multiplied by factors of 1, 2, and 4. Uncertainties of the "data"-points refer to the expected statistical uncertainty for* $10\ \mathrm{pb}^{-1}$ *only.*

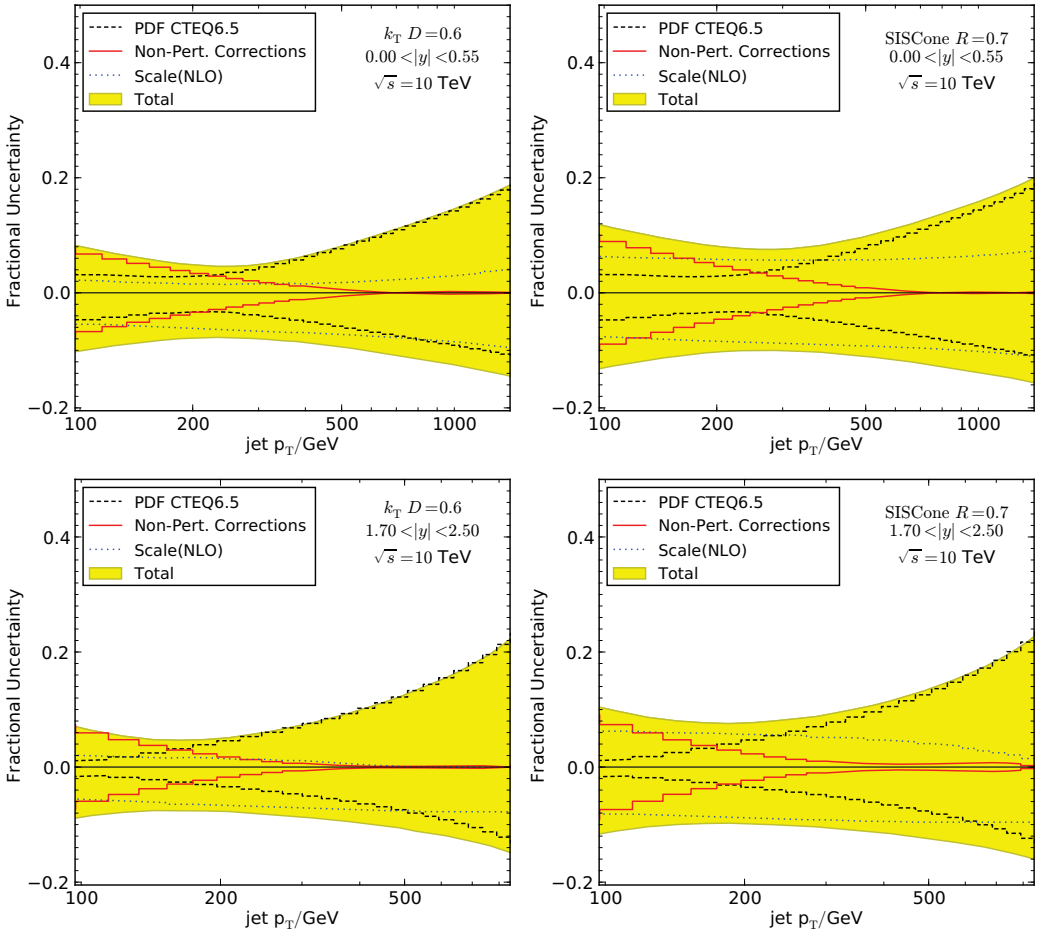

Figure 5.28: *Fractional theory uncertainties are shown centred around 0 in a central and an endcap rapidity bin for the k_T jet algorithm (left) and SISCone (right). Included are uncertainties due to the non-perturbative corrections, PDF uncertainties and scale uncertainty.*

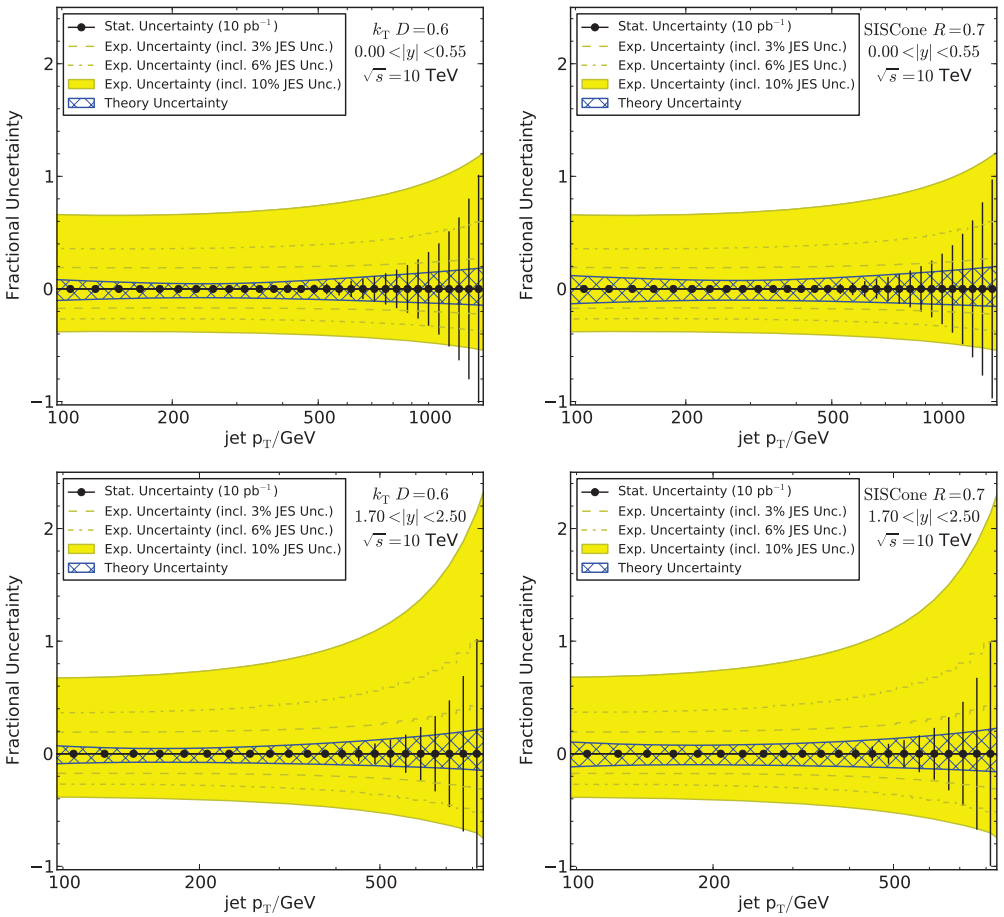

Figure 5.29: *Comparison of the total fractional uncertainties from theory and experiment and the expected statistical uncertainty with 10* pb^{-1} *for the* k_T *(left) and the SISCone algorithm (right). Additionally the total experimental uncertainty is given including 3 and 6 percent JES uncertainty to give a future perspective with improved detector understanding.*

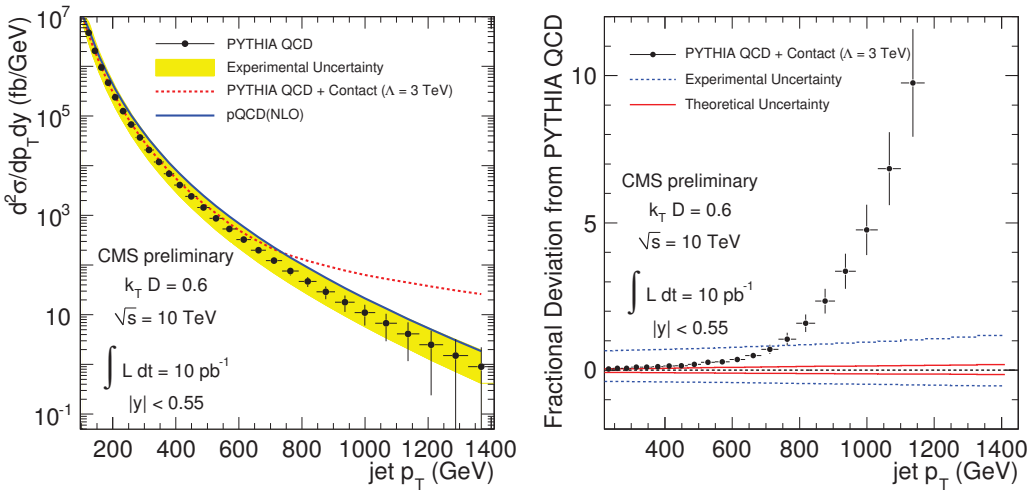

Figure 5.30: *Left: measured QCD spectrum (PYTHIA+CMS simulation, no K-factors) with experimental systematic uncertainty compared with theory (NLO incl. non-perturbative effects) and to QCD+*3 TeV *contact interaction term. Right: fractional difference of the QCD+contact interaction term from theory, compared to the experimental and theoretical uncertainties [151].*

Chapter 6

Conclusion and Outlook

A strategy to measure the inclusive jet cross section already with first LHC data and to compare it to calculations in quantum chromodynamics to the highest available precision is presented in this work. A central aspect is to evaluate the dominant uncertainties for this study, given that only data driven methods will be reliable for first analyses. On the other hand, the dominant theoretical uncertainties due to parton density functions and the unphysical scale dependence of the calculations are derived in order to allow a prediction of the possible impact of such early measurements on the current theoretical knowledge.

Already with an integrated luminosity of $10\ \mathrm{pb}^{-1}$ the statistical limitations due to the population of the phase space have been found to be small enough such that jets with large enough transverse momenta will be available to double the reach of the Tevatron, which currently is at $700\ \mathrm{GeV}$. The suggested CMS start-up measurement is found to be dominated by the experimental uncertainties over the theoretical ones by about a factor of six.

The experimental uncertainties are by far driven by the jet energy scale. The strong non-linear response of the calorimeter system and the uncertainties of the data driven correction methods lead to the current assumption that the JES uncertainty will be $\approx \pm 10\%$ for first data. In the inclusive jet cross section case, this translates into an uncertainty of $+60\%$ and -40% at $100\ \mathrm{GeV}$ and increases to $+100\%$ and -60% at the edge of the available phase space for central rapidity. For the outer regions in rapidity, the increase is even larger due to the steeper falling cross section at the kinematic limits. The uncertainty due to the luminosity is expected to contribute 10% over the whole jet p_{T} range and the uncertainty due to the unfolding procedure has been found to be about four percent at

$100\ \mathrm{GeV}$ and can be neglected for larger transverse momenta. However, in this scenario a possible uncertainty due to non-Gaussian tails in the jet resolution are attributed to the 10% JES uncertainty and are therefore not included in the evaluation of the unsmearing uncertainty.

In the theory case, non-perturbative corrections need to be applied to the bare NLO result to allow a comparison with the measurement. Such corrections contribute the largest uncertainties for the low p_T region and almost vanish at large p_T. They amount to about six percent at central rapidity and in the jet p_T range of 100 to $200\ \mathrm{GeV}$. These corrections will be available with significantly improved precision once the Monte Carlo generators can be tuned to LHC collision data and especially when NLO programs become available, which perform a matching to parton showers and allow a consistent interfacing to hadronisation models. At large jet p_T the uncertainty due to the parton distribution takes over. In case of the CTEQ65 PDFs, an uncertainty of $+20\%$ and -10% for central rapidity at $1.6\ \mathrm{TeV}$ is observed. However, the asymmetric scale uncertainty reaches comparable values at the TeV scale for negative deviations. So in total the studied theoretical uncertainties sum up to slightly less than $\pm 10\%$ at low p_T and rise to $+20$ and -15% at the edges of the statistically reachable phase space for the $10\ \mathrm{pb}^{-1}$ scenario.

Given those uncertainties it can be concluded, that only very large deviations from the Standard Model prediction can be observed within the presented first data scenario. Such deviations are expected by models which assume a compositeness of quarks, which would exceed the predicted inclusive jet cross section by a factor of ten for a contact interaction scale of $\Lambda = 3\ \mathrm{GeV}$.

Besides from searching for such exceptional signals it is important to sufficiently study the inclusive jet cross section right from the start. It covers a wide range of the reachable energy range of the experiment with relatively small statistical uncertainty and it is strongly affected by detector effects. Therefore it is an ideal observable to improve upon the required correction methods. While it should be sufficient to collect collision data for some continuous weeks at reasonably high luminosity, it will probably take years and the hard work of the whole collaboration to reduce the jet energy scale uncertainty to a few percent.

Projecting the given scenario into the future and assuming JES and luminosity uncertainties of the order of current Tevatron publications, it can be well expected to reach the order of the theoretical uncertainties already within the $\sqrt{s} = 10\ \mathrm{TeV}$ runs. Then it will be possible to run fits of α_S and constrain the parton density functions. Such measurements will be a fundamental test of the Standard Model. Especially in case of the gluon density at high fractional parton momenta x, improvements of current values can be expected. Eventually this can be further improved by looking at other jet related quantities e.g. jet rates, or by including other processes into the fitting procedure like W charge asymmetry or Drell-Yan reactions to fix the low x gluon density. Measurements at the design centre of mass energy of $14\ \mathrm{TeV}$ will finally allow to shed light into even larger jet p_T regions.

Chapter A

Appendix - Theoretical Details

A.1 The Dirac γ Matrices

The gamma matrices are required to satisfy the Clifford algebra,

$$\gamma^{\mu} \qquad \text{with} \qquad \{\gamma^{\mu}, \gamma^{\nu}\} = \eta^{\mu\nu} \tag{A.1}$$

with $\eta^{\mu\nu} = \eta_{\nu\mu} = \text{diag}(1, -1, -1, -1)$ being the metric tensor of special relativity. As the Dirac theory describes spin $\frac{1}{2}$ particles, it does not come as a surprise that the γ-matrices are made up from the Pauli spin matrices. One example for the notation as 4x4 matices is the "Bjorken and Drell" convention [155]:

$$\gamma^0 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix}, \tag{A.2}$$

where 1 and 0 denote the unitary or 0 matrix in 2x2, $\sigma^i (i = 1, 2, 3$ rsp. $x, y, z)$ are the pauli matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{A.3}$$

A.2 The Gell-Mann Matrices

To formulate the fundamental generators of SU(3) the generators can be written as $T^a = \lambda_a/2$, with the Gell-Mann matrices λ_a:

$$\begin{aligned}
\lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \\
\lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \\
\lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad \lambda_8 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \frac{1}{\sqrt{3}}.
\end{aligned} \tag{A.4}$$

The structure constants of the group ($f^{abc} = f_{abc}$) are defined through the commutator relations

$$[T^a, T^b] = i f^{abc} T^c. \tag{A.5}$$

They are totally antisymmetric ($f_{bac} = f_{acb} = -f_{abc}$) in their indices and the non-zero values out of the total $8 \cdot 8 \cdot 8 = 512$ are

$$\begin{aligned}
f_{123} &= 1 \\
f_{458} &= f_{678} = \sqrt{3}/2 \\
f_{147} &= f_{165} = f_{246} = f_{345} = f_{376} = f_{257} = 1/2
\end{aligned} \tag{A.6}$$

including the respective permutations of the indices.

Chapter B

Appendix - Technical Details

B.1 Jet Algorithms

An introduction to the jet algorithms used within this work is given in section 2.6. In this section the specific algorithmical prescriptions are discussed in more detail.

B.1.1 The $\mathrm{k_T}$ Algorithm

The algorithm introduced in section 2.6.4 clusters the input fourvectors according to a resolution parameter, which in this case follows a scheme known as the "ΔR-scheme". For each input object h_k with transverse momentum p_T, two distances d_{kB} and d_{kl} are defined in the following way:

$$d_{kB} = p^2_{\mathrm{T},k} \tag{B.1}$$

which relates the beam distance of each object directly to the transverse momentum p_T and

$$d_{kl} = \min(p^2_{\mathrm{T},k}, p^2_{\mathrm{T},l}) R^2_{kl} \tag{B.2}$$

with

$$R^2_{kl} = (y_k - y_l)^2 - (\Phi_k - \Phi_l)^2, \tag{B.3}$$

where d_{kl} defines the distance between the input objects h_k and h_l.

The algorithm works iteratively, using the following steps:

1. A list $L_{objects}$ is created, which contains information about the value of the beam distance d_{kB} and all distances between the objects d_{kl} for each input object. An additional jet resolution parameter D can be used to scale this distances at this stage, to enable some kind of radius like adjustment as it the case for cone algorithms.

2. In this step, the object with the smallest value among d_{kB} and d_{kl} is selected. If d_{kB} is the smallest, h_k is declared as jet and moved to the list L_{jets} of jets. If d_{kl} is the smallest, the objects h_k and h_l are merged with respect to the selected recombination scheme, producing a new object within the list $L_{objects}$. h_k and h_l are then removed from $L_{objects}$.

3. As long as there are objects within $L_{objects}$, steps 1 and 2 are repeated. Finally the list L_{jets} contains all produced jets and is passed on as output of the algorithm. In these jets information of all initial input objects is contained.

The procedure above describes the "inclusive" mode to run this algorithm, as it was used within this work, and is the mode best comparable to the cone algorithms. In addition to the "inclusive" mode, there is the possibility to run this algorithm in the "exclusive" mode, where all objects closest to the beam are added to a beam jet. This method introduces an additional parameter d_{cut}, which stops step 2 as soon as $\min(d_{kB}, d_{kl}) < d_{cut}$ is fulfilled. All objects are defined as jets which have not been added to the beam jet up to this point. A third possibility is to define the number N_{jet} of jets to be produced. In this case, the algorithm is run until N_{jet} jets have been constructed. The resolution parameter of the last produced jet then gives a value which finally characterises the event.

B.1.2 SISCone

The SISCone algorithm was introduced in section 2.6.5. It can be seperated into two main seqences, which will be given in a pseudo-code prescription below.

Protojet Identification

Unique protojets are defined as circles with radius R, which enclose a specific set of input objects. The procedure to find all stable jets runs as follows:

1. For any group of collinear particles, merge them into a single particle.
2. **for** particle i = 1 . . . N **do**
3. Find all particles j within a distance $2R$ of i. If there are no such particles, i forms a stable cone of its own.

4. Otherwise for each j identify the two circles for which i and j lie on the circumference. For each circle, compute the angle of its centre C relative to i, $\zeta = \arctan \frac{\Delta\Phi_{i,C}}{\Delta y_{i,C}}$.
5. Sort the circles found into increasing angle ζ.
6. Take the first circle in this order, and call it the current circle.
   Calculate the total momentum and find the cones ("checkxor") that it defines in the initially defined hash table.
   Consider all 4 permutations of edge points being included or excluded.
   Call these the "current cones".
7. **repeat**
8. **for** each of the 4 current cones **do**
9. If this cone has not yet been found, add it to the list of distinct cones.
10. If this cone has not yet been labelled as unstable, establish if the in/out status of the edge particles (with respect to the cone momentum axis) is the same as when defining the cone; if it is not, label the cone as unstable.
11. **end for**
12. Move to the next circle in order. It differs from the previous one either by a particle entering the circle, or one leaving the circle. Calculate the momentum for the new circle and corresponding new current cones by adding (or removing) the momentum of the particle that has entered (left); the checkxor can be updated by XORing with the label of that particle.
13. **until** all circles considered.
14. **end for**
15. **for** each of the cones not labelled as unstable do
16. Explicitly check its stability, and if it is stable, add it to the list of stable cones (protojets).
17. **end for**

In case of cocircular points, which means at least three points lie on the same circle of radius R, a certain ambiguity is introduced into the procedure at step 4, which is then handled by considering all possible combinations of inclusion/exclusion.

The Split-Merge Part

The split-merge part is basically that adopted for Run-II of the Tevatron [52] with some slight modifications. It is run with input of the stable cones found via B.1.2. The disambiguated, scalar $\tilde{p}_{\mathrm{T}}$ based formulation of a Tevatron Run-II type split-merge procedure with overlap threshold parameter f and transverse momentum threshold $p_{\mathrm{T,min}}$. To ensure boost invariance and IR safety, for the ordering variable and the overlap measure, it uses of $\tilde{p}_{\mathrm{T,jet}} = \sum_{i \in \mathrm{jet}} |p_{\mathrm{T,i}}|$, i.e. a scalar sum of the particle transverse momenta (as in a "p_{T}" recombination scheme). Additionally a threshold $p_{\mathrm{T,min}}$ is introduced, below which pro-

tojets are discarded. This procedure allows an infrared and collinear safe way of removing a large number of low p_{T} stable cones in pile-up polluted environments. In pseudo-code the prescription is:

1. **repeat**
2. Remove all protojets with $p_{\mathrm{T}} < p_{\mathrm{T,min}}$.
3. Identify the protojet (i) with the highest $\tilde{p}_{\mathrm{T}}$.
4. Among the remaining protojets identify the one (j) with highest $\tilde{p}_{\mathrm{T}}$ that shares particles (overlaps) with i.
5. **if** there is such an overlapping jet **then**
6. Determine the total $\tilde{p}_{\mathrm{T,shared}} = \sum_{k \in i\&j} |p_{\mathrm{T,k}}|$ of the particles shared between i and j.
7. **if** $\tilde{p}_{\mathrm{T,shared}} < f\tilde{p}_{\mathrm{T,j}}$ **then**
8. Each particle that is shared between the two protojets is assigned to the one to whose axis it is closest. The protojet momenta are then recalculated.
9. **else**
10. Merge the two protojets into a single new protojet (added to the list of protojets, while the two original ones are removed).
11. **end if**
12. If a newly produced protojet coincides with an existing one, maintain the new protojet as distinct from the existing copy(ies).
13. **else**
14. Add i to the list of final jets, and remove it from the list of protojets.
15. **end if**
16. **until** no protojets are left.

B.2 Grid Services

The Worldwide LHC Computing Grid is a central part of the CMS computing model. The main introduction is given in Section 3.4, here the specific services required for the grid operation are focused in some more detail.

Grid Wide Services

The following services are provided for the whole WLCG, and therefore do not need to be installed at every Grid site. Some of them, like the VO Server or the Replica Location Service, may only be run once or need special ways of synchronising between each other to avoid confusions. This is a common problem to distributed databases, where the same request can lead to different statements if no proper care is taken of synchronisation issues.

- **VO Server**

All users belonging to a certain VO are registered at this server. Each user has a personalised Grid certificate which has to be approved by the VO and is used for authentication purposes. Special roles for Grid users and the associated rights for Grid usage are controlled by the attached Virtual Organisation Membership Service (VOMS).

- **Replica Location Service**
  Files can be stored on various Grid storage elements. To be able to access these datasets from every Grid site, they have to be registered to the database of the Replica Location Service (RLS). Datasets are given special identification strings (GUID) which can be used later to access the data from an arbitrary Grid site. In addition to that, user specified strings, so called Logical Filenames (LFN), can be used. They have to be globally unique within one VO. This service is provided by the RLS, which only exists once per VO.

- **Information Systems**
  Information about existing Grid resources and their status as well as information about single Grid jobs is being gathered by the Information Systems. Several Systems exist to collect this information and to provide it to the end-user or other Grid services.

- **Workload Management Service (WMS)**
  Before jobs can be actually processed, they have to be sent to the WMS first. By using data from the Information Systems, this service is aware of free computing resources on the Grid. Users can attach requirements to their jobs, regarding for example memory demands, CPU-time or requisite software installations, which have to be provided by the target site. Considering these requirements the WMS is able to find the compatible Grid resource and will pass the job to a suitable site.

Site Wide Services

The following services have to be provided individually by each Grid site. Some are only for local usage, others will also offer their service to the Grid.

- **User Interface**
  User access to the LHC Grid is provided by the User Interface usually installed on dedicated machines or on a user's laptop. User interfaces must be provided with a set of tools to authenticate within the Grid, submit and monitor jobs as well as retrieve their output. Users authenticate themselves using their Grid certificate and retrieve a so called "Grid proxy" with a limited lifetime. As Grid services can only be used with a valid proxy, jobs will be killed when the respective proxy is expired.

- **Storage Element**
  The actual storing of files within the Grid is closely related to the Storage Element. It works as a portal between other Grid services and the local storage systems at each site. To access files on the Grid only the information provided by the RLS is needed. The physical accessing or writing of files is steered by the Storage Element. In case of storing files, Grid services are supposed to register the file to the RLS and write them to a physical storage using the Storage Element. All types of storage systems from hard disk to tape storage and other mixed storage area network installations can be attached to the Storage Element. Due to the physical storage in the backend and the throughput considerations, several implementations of storage elements exist, with optimisations for available backends and requirements of the site. While small Tier 3 sites can work with easy maintainable implementations, large Tier 1 sites require better scalability, which also increases the complexity of the storage elements.

- **Computing Element**
  Like Storage Elements can work as portals to local storage systems, Computing Elements (CE) are Grid-widely visible portals to the local batch systems. The CE retrieves jobs from the Resource Broker and submits them to the local worker nodes, typically using the local batch system. Worker nodes are usually single computers in a computing cluster. Batch systems like TORQUE [156] handle distribution and monitoring of the jobs. These systems have been in use for years in computing centres worldwide, so there is huge expertise in this field, of which the Grid infrastructure can profit. It is also the easiest way to implement Grid computing in traditional computing centres, with respect to their local policies. The Computing Element is attached to the local batch system in such a way, that there is no need for Grid users to be aware of certain peculiarities of the batch systems on which their job actually runs. The CE also takes care of returning the default output back to the WMS, from which the user finally retrieves it. Larger output is usually written to storage elements by the jobs.

B.3 Usage of fastNLO tables

The final result of a scenario prepared with NLOJET++ and fastNLO is a set of tables each of which contains all required information to rederive an inclusive cross section for different PDFs, α_S, etc.. One set comprises one primary table, one reference table and as many tables for the statistical uncertainty as there were NLO jobs, see Table 4.1. For almost all purposes it is sufficient to dispose of the primary table, the code to reevaluate it and one histogram file including additional histograms with the algorithmic and statistical uncertainties. For the two scenarios generically named `fnl1310` and `fnl1308`, i.e. inclusive jets for proton-proton collisions at 10 TeV with a minimum of $p_{\mathrm{T,jet}}$ of 50 GeV and the

individual $p_{\mathrm{T,jet}}$ as the hard scale of the process for the k_{T} algorithm with $D = 0.6$ resp. the SISCone algorithm with $R = 0.7$ they are made available from the following TWiki page:

`https://twiki.cern.ch/twiki/bin/view/CMS/CMSfastNLO`

In order to use it one has to download the desired table file, `fnl1310.tab.gz` or `fnl1308.tab.gz`, the `pdfunc.tar.gz` package for FORTRAN code creating HBOOK histograms or the `pdfuncpp.tar.gz` C++ package for writing ROOT histograms.
The sample histograms files are named
`fnl1310_cteq65_aspdf_full.hbk` and
`fnl1308_cteq65_aspdf_full.hbk`.
In both cases you need to have LHAPDF installed, for the Fortran version in addition you need CERNLIB.

B.3.1 The FORTRAN Version

In order to use the FORTRAN version do the following:

- 'gunzip' the table file: `gunzip table.tab.gz`
- Unpack the pdfunc.tar.gz package: `tar xzf pdfunc.tar.gz`
- Set the environment variables `CERNLIB` and `LHAPDF` to point to the directories where the libraries can be found.
- Define properly the `FC` variable in the provided Makefile to point to your compiler, e.g. `g77`, `f77` or `gfortran`.
- Type `make` and then `./pdfunc -h` to get a summary of the command syntax:

```
##############################################
# PDFUNC
##############################################
# Example program to compute PDF uncertainties
# using a fastNLO table and PDFs from LHAPDF
##############################################
#

Usage: ./pdfunc [arguments]
  NLO input table, def. = table.txt
  HBOOK output file, def. = fastnlo.hbk
  PDF set, def. = cteq65.LHgrid
  PDF path, def. = \$(LHAPDF)/../share/lhapdf/PDFsets
  alpha_s calc., def. from PDF set
```

When successful you will get a printout of all NLO cross sections with lower and upper PDF uncertainties[1] and you will find histograms in the HBOOK file corresponding to the cross sections vs. $p_{\mathrm{T,jet}}$ in $\mathrm{fb}/\ \mathrm{GeV}$, i.e. divided by the bin size in p_{T} and absolute rapidity $|y|$. Note that all calculations are done in double precision, however, HBOOK histograms store numbers only in single precision. Also, the printed numbers of the FORTRAN and the C++ version are only identical up to single precision due to some inconsistencies in LHAPDF. In the histogram numbers the exact content is encoded as specified in Table B.1. As an example, the NLO cross section at central rapidity can be found in histogram 0300100.

histogram no.	allowed values	meaning
Oxxxxxx	$\mathrm{O} = 0, 1, 2$	Full cross section, LO, NLO correction
xSxxxxx	$\mathrm{S} = 1, \ldots, 4$	1: $\mu_r = \mu_f = p_{\mathrm{T,jet}}/4$
		2: $\mu_r = \mu_f = p_{\mathrm{T,jet}}/2$
		3: $\mu_r = \mu_f = p_{\mathrm{T,jet}}$ (default)
		4: $\mu_r = \mu_f = 2 \cdot p_{\mathrm{T,jet}}$
xxPxxxx	$\mathrm{P} = 0, \ldots, 7$	0: all subprocesses
		1–7: the seven possible q, $\bar{q}$, g subprocesses
xxxxRxx	$\mathrm{R} = 1, \ldots, 6$	bin in absolute rapidity $\|y\|$
xxxxxxU	$\mathrm{U} = 0, \ldots, 5$	0: central result
		1: lower PDF uncertainty
		2: upper PDF uncertainty
		3: statistical uncertainty derived from RMS
		4: statistical uncertainty derived from maximal spread
		5: algorithmic uncertainty

Table B.1: *Setup of the primary and reference calculations required for the inclusive jet cross section to NLO for both, the k_{T} and SISCone algorithm.*

B.3.2 The C++ Version

For the C++ version one has to complete these steps:

- 'gunzip' the table file: `gunzip table.tab.gz`
- Unpack the pdfuncpp.tar.gz package: `tar xzf pdfunc.tar.gz`

[1] Do not use this program with e.g. the NNPDF1.0 PDF which provides uncertainties in a different way than the CTEQ or MSTW groups!

- Set the environment variables `LHAPDF` to point to the directories where the libraries can be found.
- Run the auto configuration:
  `./autogen.sh -with-lhapdf=$LHAPDF -enable-maintainer-mode CXXFLAGS="-O3 -Wall"`
- Type `make`, `make install` and then `bin/pdfunc -h` to get a summary of the command syntax:

```
Syntax: bin/pdfunc [OPTIONS]

Without options: bin/pdfunc <input table> <outputfile> <pdfsetname>

With options: -h, --help Show this help message
              -i, --input = <filename> Filename of input table
              -o, --output = <filename> Write histogramms to
                                        <filename>
              -p, --pdfset = <pdfset> Name of pdfset
                                      (e.g. cteq65.LHgrid)
              -a, --alphas = <calctype> Type of Calculation to use.
                             (0=default 2 loop, 1 = 1 loop,
                              2 = alphaS from PDF)
              -s, --sourcepdf= <path> Path to pdfsets
```

The file `PDFSETDIR` is read for the path to the PDF sets. If it does not exist $LHAPDF is evaluated. When successful you will find similar histograms in the ROOT file as described above for the FORTRAN/HBOOK case.

B.4 The Data Samples

B.4.1 The Cosmic Data Samples

The noise suppression procedure to reject events based on a cut on MET/$\sum E_{\mathrm{T}}$ in section 5.3 is motivated by studies of signals due to cosmic particles [151]. The data was collected by the CMS detector from mid October till mid November 2008 with the magnetic field of $\approx 4\,\mathrm{T}$ turned on but without beam in the LHC. Under these conditions events can be triggered by real particles, usually muons created in cosmic ray air showers, or, for example, by electronic noise, which provides an excellent environment to study event clean-up. The data run number under consideration is 68021 and is contained in the dataset `/Cosmics/Commissioning08-PromptReco-v2/RECO`. The pixel detector, tracker, electromagnetic calorimeter, hadron calorimeter (barrel and endcap) and the

muon system were turned on. Only one of the jet triggers, HLT_Jet30, with a threshold of 30 GeV on the calorimetric energy was active for this run.

B.4.2 The Monte Carlo Simulation Samples

The samples used to perform the cross section study were produced between Summer 2008 and Spring 2009. The events were generated with the following Monte Carlo generator:

A - PYTHIA 6.4 [111], DBS identification:
/QCDDiJetPt*to*/Summer08_IDEAL_V9_AODSIM_v1/AODSIM

B - PYTHIA 6.4, DBS identification:
/QCDpt*/Summer08_IDEAL_V11_redigi_v1/AODSIM

In the DBS identifications '*' has to be replaced by the various phase-space limits. For the PYTHIA events the recent underlying event tune D6T [129] is used. It differs from its predecessor tune DWT [121] only in the choice of the CTEQ6L1 [119] over the CTEQ5L [157] parton density function (PDF) and an adjusted minimum transverse momentum exchange of the multiple parton interactions to keep their cross section constant [129].

After the generation the events are passed through the full detector simulation using CMSSW 2.1.7 and 2.1.8 for the reconstruction of samples A. Although being simulated with version 2.1.7, version 2.2.1 was used for the reconstruction of the samples B. The phase space of the samples A has been divided up into 21 bins of the transverse momentum of the hard interaction $\hat{p}_T$ as given in Tab. B.2. Due to limitations in the event generation with zero minimal required p_T of PYTHIA (Minimum bias mode), the first sample does not have an upper $\hat{p}_T$ cut so that overlapping events have to be removed by imposing manually such an upper cut. As each sample corresponds to a different integrated luminosity, the statistical representation of the phase space by a dataset exhibits some characteristic discontinuities as shown in Fig. B.1. The single samples in B do not have an upper p_T head cut imposed, thus this needs to be treated when reconstructing the jet cross section from all samples. The samples B contain about 31 million events in total and are only used for the dedicated jet resolution studies in Chapter 5.2.

As the MC data partially contained physically identical events, it is remarked that these had to be removed from the samples in an additional selection step. The numbers given in the table reflect the numbers of unique events used for the study. The differential jet cross sections are finally derived from the cleaned MC samples using the integrated cross section of each generated $\hat{p}_T$ range.

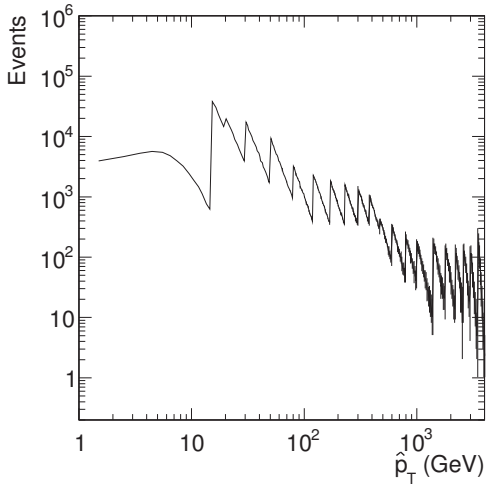

Figure B.1: *Monte Carlo generated phase space coverage of sample A with* PYTHIA *in* 21 $\hat{p}_{\mathrm{T}}$ *bins.*

min. $\hat{p}_T$ in GeV	Events	Cross Section in pb
0 to ∞	103838	$5.156 * 10^{10}$
15 to 20	137880	$9.494 * 10^{8}$
20 to 30	101880	$4.010 * 10^{8}$
30 to 50	133200	$9.470 * 10^{7}$
50 to 80	102600	$1.220 * 10^{7}$
80 to 120	51165	$1.617 * 10^{6}$
120 to 170	50085	$2.560 * 10^{5}$
170 to 230	51840	$4.833 * 10^{4}$
230 to 300	54000	$1.062 * 10^{4}$
300 to 380	60048	$2.635 * 10^{3}$
380 to 470	51840	$7.221 * 10^{2}$
470 to 600	27648	$2.410 * 10^{2}$
600 to 800	28620	$6.250 * 10^{1}$
800 to 1000	20880	$9.421 * 10^{0}$
1000 to 1400	23100	$2.344 * 10^{0}$
1400 to 1800	27676	$1.569 * 10^{-1}$
1800 to 2200	22848	$1.381 * 10^{-2}$
2200 to 2600	22560	$1.296 * 10^{-3}$
2600 to 3000	22800	$1.140 * 10^{-4}$
3000 to 3500	20880	$8.432 * 10^{-6}$
3500 to ∞	34200	$1.815 * 10^{-7}$

Table B.2: *Coverage of the phase space in 21 exclusive bins in $\hat{p}_T$. The integrated event cross section for each sample is listed as given by* PYTHIA. *The lower border of the following bin corresponds to the upper border of the preceding bin. The bins from 0 and 3500 reach to infinity $\hat{p}_T$.*

List of Figures

2.1 Schematic view of $\mathcal{L}_{\text{quark}}$ . . . 12
2.2 Schematic view of $\mathcal{L}_{\text{gauge}}$ . . . 13
2.3 Quark self-energy . . . 14
2.4 The running coupling α_S . . . 17
2.5 Distributions of parton momenta . . . 20
2.6 Hadron-hadron collision . . . 21
2.7 Decomposition of the total jet cross section into the partonic Processes . . . 23
2.8 Event simulation of multi purpose generators . . . 25
2.9 Lund/String Model . . . 31
2.10 Lund diquark and popcorn model . . . 32
2.11 Collinear unsafe behaviour of jet algorithms . . . 33
2.12 Infrared unsafe behaviour of jet algorithms . . . 34
2.13 Incl. jet cross section from Tevatron, log-log comparison . . . 40
2.14 Incl. jet cross section from Tevatron, ratios . . . 41

3.1 LHC underground structure . . . 44
3.2 Overview of the CMS detector . . . 49
3.3 Profile of the CMS detector . . . 50
3.4 Tier structure of the WLCG . . . 58

4.1 Uncertainties of the fastNLO calculations for the k_{T} algorithm . . . 74
4.2 Uncertainties of the fastNLO calculations for the SISCone algorithm . . . 74
4.3 Stat. unc. nlo calc. - all bins. . . . 75
4.4 Stat. unc. nlo calc. - individual bins. . . . 76

4.5 Scale uncertainties, LO . . . 79
4.6 Scale uncertainties, NLO . . . 79
4.7 Sensitivity on the order of α_S . . . 81
4.8 Sensivity on the central value of α_S . . . 82
4.9 Inclusive jet cross section - NLO . . . 84
4.10 SISCone-k_{T} comparison (NLO) . . . 84
4.11 PDF uncertainties - NLO . . . 86
4.12 PDF variation - central values, NLO . . . 87
4.13 K-factors, CTEQ6M/CTEQ6L1 . . . 88
4.14 K-factors, CTEQ65, all processes . . . 89
4.15 K-factors, CTEQ65, specific processes . . . 90
4.16 K-factors, CTEQ65, process comparison . . . 91
4.17 Decomposition of subprocesses . . . 92
4.18 PYTHIA - NLOJET++ comparison . . . 93
4.19 Inclusive jet spectra from PYTHIA and HERWIG++ with fits . . . 97
4.20 Dependence of the hadronisation corrections on MPI . . . 98
4.21 MPI, hadronisation and overall correction for HERWIG + JIMMY . . . 100
4.22 Combined overall corrections for inner rapidity . . . 101
4.23 Statistical uncertainties for detector specific corrections . . . 102
4.24 Combined overall corrections for all rapidities . . . 102

5.1 JEC(p_{T}) . . . 107
5.2 JEC(η) . . . 108
5.3 Jet p_{T} Resolution - individual peaks . . . 109
5.4 Jet p_{T} resolution, individual bins with Gaussian fit . . . 110
5.5 Jet p_{T} resolution - Gaussian means . . . 110
5.6 Jet p_{T} resolution - parametrised curve . . . 112
5.7 Jet rapidity resolution - parametrised curve . . . 112
5.8 Missing E_{T} over $\sum E_{\mathrm{T}}$ distributions . . . 114
5.9 Trigger yield . . . 115
5.10 Trigger turn-on, individual streams, spectrum composition . . . 116
5.11 Expected number of jets . . . 117
5.12 JES calibration effect on the inclusive jet cross section . . . 118
5.13 Incl. jet cross section with smeared ansatz fit . . . 119
5.14 Smeared Ansatz - fit quality . . . 119
5.15 Unsmearing correction factors . . . 120
5.16 Unsmeared incl. jet cross section . . . 120
5.17 Experimental uncertainties - central y bin . . . 122
5.18 Experimental uncertainties - endcap bin . . . 122
5.19 Fractional y-uncertainty . . . 124
5.20 JES+unsmearing cor. calojets v.s. particle jets . . . 124

5.21 JES+unsmearing+Gaussian mean cor. calojets v.s. particle jets 126
5.22 Hybrid JES deviation 126
5.23 JES+unsmearing+hybrid cor. calojets v.s. particle jets 127
5.24 NLOJET++ compared to PYTHIA generation stages 129
5.25 K-factors, CTEQ6M/CTEQ6L1 129
5.26 Non-perturbative corrections 130
5.27 Full NLO comparison 132
5.28 Theory uncertainties 133
5.29 Theory, experimental, and statistical uncertainties 134
5.30 Comparison to contact interactions 135

B.1 PYTHIA Monte Carlo phase space coverage 153

List of Tables

2.1 Fundamental fermions of the Standard Model 6
2.2 Elementary interactions 7
2.3 The colour representation of QCD 10

3.1 LHC Machine Parameters 47

4.1 NLO calculation computing resources 73
4.2 Rapidity binning, 100/pb 77
4.3 p_{T} binning, 100/pb 77
4.4 List of subprocesses 91

5.1 Rapidity binning, 10/pb 105
5.2 p_{T} binning - 10/pb 105
5.3 Jet p_{T} resolution - fit parameters k_{T} 111
5.4 Jet p_{T} resolution - fit parameters SISCone 111
5.5 Single jet trigger streams 114
5.6 Trigger turn-on points 115

B.1 NLO calculation configuration 150
B.2 PYTHIA Monte Carlo phase space coverage 154

Bibliography

[1] **CDF** Collaboration, F. Abe et al., "Inclusive jet cross section in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV," *Phys. Rev. Lett.* **77** (1996) 438–443, arXiv:hep-ex/9601008. doi:10.1103/PhysRevLett.77.438. (Citations on pages 2, 82).

[2] G. Hanson et al., "Evidence for Jet Structure in Hadron Production by e^+e^- Annihilation," *Phys. Rev. Lett.* **35** (1975) 1609. (Citations on pages 5, 24).

[3] G. Sterman and S. Weinberg, "Jets from Quantum Chromodynamics," *Phys. Rev. Lett.* **39** (1977) 1436. (Citations on pages 5, 24, 34).

[4] Povh, Rith, Scholz, and Zetsche, "Teilchen und Kerne". Springer Verlag, 2004. (Citation on page 7).

[5] D. J. Griffiths, "Introduction to elementary particles". John Wiley & Sons, Inc., 1987. (Citation on page 7).

[6] G. Dissertori, I. Knowles, and M. Schmelling, "Quantum Chromodynamics - High Energy Experiments and Theory". Oxford Science Publications, 2003. (Citations on pages 7, 10).

[7] R. K. Ellis, W. J. Stirling, and B. Webber, "QCD and Collider Physics". CAMBRIDGE, 1996. (Citation on page 7).

[8] G. Newton, R., "Optical theorem and beyond," *Am.J.Phys* **44** (1976) 639–642. doi:10.1119/1.10324. (Citation on page 8).

[9] **TOTEM** Collaboration, G. Anelli et al., "The TOTEM experiment at the CERN Large Hadron Collider," *JINST* **3** (2008) S08007. doi:10.1088/1748-0221/3/08/S08007. (Citations on pages 9, 45).

[10] A. Abbas, "On the number of colours in quantum chromodynamics," arXiv:hep-ph/0009242. (Citation on page 9).

[11] R. Mills and C.-N. Yang, "Conservation of Isotopic Spin and Isotopic Gauge Invariance," *Phys. Rev.* **96** (1954) 191. doi:10.1103/PhysRev.96.191. (Citation on page 10).

[12] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys. Lett.* **B47** (1973) 365–368. doi:10.1016/0370-2693(73)90625-4. (Citation on page 11).

[13] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. 1," *Phys. Rev.* **D8** (1973) 3633–3652. doi:10.1103/PhysRevD.8.3633. (Citation on page 11).

[14] S. Weinberg, "Nonabelian Gauge Theories of the Strong Interactions," *Phys. Rev. Lett.* **31** (1973) 494–497. doi:10.1103/PhysRevLett.31.494. (Citation on page 11).

[15] W. A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta, "Deep Inelastic Scattering Beyond the Leading Order in Asymptotically Free Gauge Theories," *Phys. Rev.* **D18** (1978) 3998. doi:10.1103/PhysRevD.18.3998. (Citation on page 15).

[16] Particle Data Group, "Physical Constants." http://pdg.lbl.gov/2009/reviews/rpp2009-rev-phys-constants.pdf, 2009. (Citation on page 16).

[17] **Particle Data Group** Collaboration, C. Amsler et al., "Review of particle physics," *Phys. Lett.* **B667** (2008) 1. doi:10.1016/j.physletb.2008.07.018. (Citations on pages 16, 16, 17, 78, 80, 81, 81).

[18] E. D. Bloom et al., "RECENT RESULTS IN INELASTIC ELECTRON SCATTERING,". Presented at the XV International Conference on High Energy Physics, Kiev, USSR, Aug 26-Sep4, 1970. (Citation on page 18).

[19] J. D. Bjorken and E. A. Paschos, "Inelastic Electron Proton and gamma Proton Scattering, and the Structure of the Nucleon," *Phys. Rev.* **185** (1969) 1975–1982. doi:10.1103/PhysRev.185.1975. (Citation on page 18).

[20] R. P. Feynman, "Photon-hadron interactions,". Reading 1972, 282p, published: Addison-Wesley, Menlo Park, California. (Citation on page 18).

[21] J. Callan, Curtis G. and D. J. Gross, "High-energy electroproduction and the constitution of the electric current," *Phys. Rev. Lett.* **22** (1969) 156–159. doi:10.1103/PhysRevLett.22.156. (Citation on page 18).

[22] F. E. Close, "AN INTRODUCTION TO QUARKS AND PARTONS,". Academic Press/london 1979, 481p. (Citation on page 19).

[23] G. Altarelli, "Partons in Quantum Chromodynamics," *Phys. Rept.* **81** (1982) 1. doi:10.1016/0370-1573(82)90127-2. (Citation on page 19).

[24] Y. L. Dokshitzer, "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics," *Sov. Phys. JETP* **46** (1977) 641–653. (Citation on page 19).

[25] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.* **B126** (1977) 298. doi:10.1016/0550-3213(77)90384-4. (Citation on page 19).

[26] V. N. Gribov and L. N. Lipatov, "Deep inelastic e p scattering in perturbation theory," *Sov. J. Nucl. Phys.* **15** (1972) 438–450. (Citation on page 19).

[27] J. C. Collins and D. E. Soper, "The Theorems of Perturbative QCD," *Ann. Rev. Nucl. Part. Sci.* **37** (1987) 383–409. (Citation on page 19).

[28] MSTW parton distribution functions http://durpdg.dur.ac.uk/hepdata/mrs.html/. (Citation on page 20).

[29] CTEQ parton distribution functions http://www.phys.psu.edu/~cteq/. (Citation on page 20).

[30] Online tool for pdf-plotting, Durham University http://durpdg.dur.ac.uk/hepdata/pdf3.html/. (Citation on page 20).

[31] T. Kluge, K. Rabbertz, and M. Wobisch, "fastNLO - fast pQCD calculations for hadron-induced processes." http://hepforge.cedar.ac.uk/fastnlo. to be published. (Citation on page 23).

[32] T. Sjostrand and P. Z. Skands, "Transverse-momentum-ordered showers and interleaved multiple interactions," *Eur. Phys. J.* **C39** (2005) 129–154, arXiv:hep-ph/0408302. doi:10.1140/epjc/s2004-02084-y. (Citation on page 24).

[33] J. M. Campbell, J. W. Huston, and W. J. Stirling, "Hard interactions of quarks and gluons: A primer for LHC physics," *Rept. Prog. Phys.* **70** (2007) 89, arXiv:hep-ph/0611148. doi:10.1088/0034-4885/70/1/R02. (Citations on pages 25, 94, 128).

[34] PYTHIA (and JETSET) webpage, Lund University http://www.thep.lu.se/~torbjorn/Pythia.html. (Citations on pages 25, 66).

[35] HERWIG http://hepwww.rl.ac.uk/theory/seymour/herwig/. (Citations on pages 25, 67).

[36] S. Gieseke, "The new Monte Carlo event generator Herwig++," arXiv:hep-ph/0408034. (Citations on pages 25, 67).

[37] D. A. Kosower, "QCD at the Dawn of the LHC Era". arXiv:hep-ph/0604015, 2006. (Citation on page 26).

[38] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, "QCD Matrix Elements + Parton Showers," *JHEP* **11** (2001) 063, arXiv:hep-ph/0109231. (Citation on page 26).

[39] F. Krauss, "Matrix elements and parton showers in hadronic interactions," *JHEP* **08** (2002) 015, arXiv:hep-ph/0205283. (Citation on page 26).

[40] SHERPA webpage, Institute of Theoretical Physics, Dresden University for Technology http://www.sherpa-mc.de. (Citation on page 26).

[41] ALPGEN webpage, M. L. Mangano, F. Piccini, A. D. Polosa, M. Moretti, R. Pittau http://mlm.home.cern.ch/mlm/alpgen. (Citation on page 26).

[42] 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:hep-ph/0206293. (Citation on page 26).

[43] J. Alwall et al., "MadGraph/MadEvent v4: The New Web Generation," *JHEP* **09** (2007) 028, arXiv:0706.2334. (Citation on page 26).

[44] J. Alwall, S. de Visscher, and F. Maltoni, "QCD radiation in the production of heavy colored particles at the LHC," *JHEP* **02** (2009) 017, arXiv:0810.5350. doi:10.1088/1126-6708/2009/02/017. (Citation on page 26).

[45] S. Frixione, B. Webber, MC@NLO webpage `http://www.hep.phy.cam.ac.uk/theory/webber/MCatNLO`. (Citation on page 26).

[46] R. D. Field and R. P. Feynman, "A parametrization of the properties of quark jets," *Nucl. Phys.* **B136** (1978) 1. `doi:10.1016/0550-3213(78)90015-9`. (Citation on page 27).

[47] B. R. Webber, "A QCD Model for Jet Fragmentation Including Soft Gluon Interference," *Nucl. Phys.* **B238** (1984) 492. `doi:10.1016/0550-3213(84)90333-X`. (Citation on page 27).

[48] G. Corcella et al., "HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes)," *JHEP* **01** (2001) 010, `arXiv:hep-ph/0011363`. (Citation on page 27).

[49] M. Bahr et al., "Herwig++ Physics and Manual," *Eur. Phys. J.* **C58** (2008) 639–707, `arXiv:0803.0883`. `doi:10.1140/epjc/s10052-008-0798-9`. (Citations on pages 27, 27, 94, 129).

[50] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton Fragmentation and String Dynamics," *Phys. Rept.* **97** (1983) 31–145. `doi:10.1016/0370-1573(83)90080-7`. (Citation on page 28).

[51] X. Artru, "Classical String Phenomenology. 1. How Strings Work," *Phys. Rept.* **97** (1983) 147. `doi:10.1016/0370-1573(83)90081-9`. (Citation on page 29).

[52] G. C. Blazey et al., "Run II jet physics," `arXiv:hep-ex/0005012`. (Citations on pages 33, 145).

[53] G. P. Salam and G. Soyez, "A practical Seedless Infrared-Safe Cone jet algorithm," *JHEP* **05** (2007) 086, `arXiv:0704.0292`. (Citations on pages 33, 36, 96).

[54] **JADE** Collaboration, W. Bartel et al., "Experimental Studies on Multi-Jet Production in e+ e- Annihilation at PETRA Energies," *Z. Phys.* **C33** (1986) 23. `doi:10.1007/BF01410449`. (Citation on page 34).

[55] **JADE** Collaboration, S. Bethke et al., "Experimental Investigation of the Energy Dependence of the Strong Coupling Strength," *Phys. Lett.* **B213** (1988) 235. `doi:10.1016/0370-2693(88)91032-5`. (Citation on page 34).

[56] S. Catani, Y. L. Dokshitzer, M. Olsson, G. Turnock, and B. R. Webber, "New clustering algorithm for multi - jet cross-sections in e+ e- annihilation," *Phys. Lett.*

B269 (1991) 432–438. doi:10.1016/0370-2693(91)90196-W. (Citation on page 34).

[57] S. D. Ellis and D. E. Soper, "Successive combination jet algorithm for hadron collisions," *Phys. Rev.* **D48** (1993) 3160–3166, arXiv:hep-ph/9305266. (Citations on pages 34, 96).

[58] S. Catani, Y. L. Dokshitzer, and B. R. Webber, "The K-perpendicular clustering algorithm for jets in deep inelastic scattering and hadron collisions," *Phys. Lett.* **B285** (1992) 291–299. doi:10.1016/0370-2693(92)91467-N. (Citations on pages 34, 96).

[59] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, "Longitudinally invariant K(t) clustering algorithms for hadron hadron collisions," *Nucl. Phys.* **B406** (1993) 187–224. doi:10.1016/0550-3213(93)90166-M. (Citations on pages 34, 34, 96, 96).

[60] M. Cacciari and G. P. Salam, "Dispelling the N**3 myth for the k(t) jet-finder," *Phys. Lett.* **B641** (2006) 57–61, arXiv:hep-ph/0512210. (Citations on pages 34, 35, 72, 96).

[61] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_T jet clustering algorithm," *JHEP* **04** (2008) 063, arXiv:0802.1189. doi:10.1088/1126-6708/2008/04/063. (Citation on page 34).

[62] M. Cacciari, G. Salam, and G. Soyez. fastJet homepage http://www.lpthe.jussieu.fr/~salam/fastjet/. (Citations on pages 35, 64).

[63] J. M. Butterworth, J. P. Couchman, B. E. Cox, and B. M. Waugh, "KtJet: A C++ implementation of the K(T) clustering algorithm," *Comput. Phys. Commun.* **153** (2003) 85–96, arXiv:hep-ph/0210022. doi:10.1016/S0010-4655(03)00156-5. (Citation on page 35).

[64] **D0** Collaboration, V. M. Abazov et al., "Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$TeV," *Phys. Rev. Lett.* **101** (2008) 062001, arXiv:0802.2400. doi:10.1103/PhysRevLett.101.062001. (Citations on pages 38, 40, 41, 94, 97, 103, 117, 128).

[65] **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. doi:10.1103/PhysRevD.78.052006. (Citations on pages 38, 40, 41, 94, 94, 103, 103).

[66] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, "Parton distributions for the LHC," arXiv:0901.0002. (Citations on pages 38, 72, 82, 83).

[67] **CDF - Run II** Collaboration, A. Abulencia et al., "Measurement of the Inclusive Jet Cross Section using the k_{T} algorithm in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV with the CDF II Detector," *Phys. Rev.* **D75** (2007) 092006, arXiv:hep-ex/0701051. doi:10.1103/PhysRevD.75.092006. (Citations on pages 38, 94, 103).

[68] Evans, Lyndon, (ed.) and Bryant, Philip, (ed.), "LHC Machine," *JINST* **3** (2008) S08001. doi:10.1088/1748-0221/3/08/S08001. (Citations on pages 43, 47).

[69] **CDF** Collaboration, "Combined CDF and DZero Upper Limits on Standard Model Higgs-Boson Production with up to 4.2 fb-1 of Data," arXiv:0903.4001. (Citation on page 43).

[70] "LHC Design Report Volume II, The LHC Infrastructure and General Services (CERN-2004-003-v2)". 2004. (Citation on page 44).

[71] LHC structure and cooling state, October 2009, from CERN archive server http://doc.cern.ch//archive/electronic/cern/others/PHO/photo-bul//bul-pho-2009-098.jpg. (Citation on page 44).

[72] **ALICE** Collaboration, K. Aamodt et al., "The ALICE experiment at the CERN LHC," *JINST* **0803** (2008) S08002. doi:10.1088/1748-0221/3/08/S08002. (Citation on page 45).

[73] **ATLAS** Collaboration, G. Aad et al., "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST* **3** (2008) S08003. doi:10.1088/1748-0221/3/08/S08003. (Citation on page 45).

[74] **CMS** Collaboration, R. Adolphi et al., "The CMS experiment at the CERN LHC," *JINST* **0803** (2008) S08004. doi:10.1088/1748-0221/3/08/S08004. (Citations on pages 45, 47, 48, 104, 106).

[75] **LHCb** Collaboration, A. A. Alves et al., "The LHCb Detector at the LHC," *JINST* **3** (2008) S08005. doi:10.1088/1748-0221/3/08/S08005. (Citation on page 45).

[76] F. Hinterberger, "The physics of particle accelerators and ion optics. (In German),". Berlin, Germany: Springer (1997) 365 p. (Citation on page 46).

[77] **CMS** Collaboration, G. L. Bayatian et al., "CMS physics: Technical design report,". CERN-LHCC-2006-001. (Citations on pages 47, 48).

[78] CMS Collaboration, "Public Web Page." `http://cms.web.cern.ch/cms/Collaboration/index.html`, 2009. (Citations on pages 46, 55).

[79] **CMS** Collaboration, G. L. Bayatian et al., "CMS technical design report, volume II: Physics performance," *J. Phys.* **G34** (2007) 995–1579. `doi:10.1088/0954-3899/34/6/S01`. (Citation on page 48).

[80] R. Kalman, "A New Approach to Linear Filtering and Prediction Problems," *Transaction of the ASME, Journal of Basic Engineering* (1960) 35–45. (Citation on page 51).

[81] "CMS The Hadron Calorimeter Project, Technical Design Report (CERN/LHCC 97-31)". 1997. (Citation on page 53).

[82] C. K. Ian Foster, "The Grid - Blueprint for a New Computing Infrastructure". Morgan Kaufmann Publishers, 1998/1999. (Citation on page 56).

[83] LHC Computing Grid Technical Design Report LCG-TDR-001, CERN-LHCC-2005-024, 20 June 2005. (Citation on page 57).

[84] V. Buege, U. Felzmann, C. Jung, H. Stadie, and A. Vest, "The Tier 2/3 prototype center at IEKP". Institut für Experimentelle Kernphysik, Universität Karlsruhe IEKP-KA/2006-03, 2006. (Citation on page 57).

[85] CMS WLCG Structure: Thanks to A. Scheurer for providing the picture. (Citation on page 58).

[86] Scientific Linux `http://www.scientificlinux.org/`. (Citation on page 59).

[87] Open Science Grid. homepage: `http://www.opensciencegrid.org/`. (Citation on page 59).

[88] Enabling Grids for E-sciencE `http://public.eu-egee.org/`. (Citation on page 59).

[89] F. Stober et al., "Presentation: grid-control - The Swiss Army knife of job submission tools." `https://indico.desy.de/getFile.py/access?contribId=5&resId=1&materialId=slides&confId=2295` , 2009. GridKA School. (Citation on page 60).

[90] CRAB (Cms Remote Analysis Builder) `http://cmsdoc.cern.ch/cms/ccs/wm/www/Crab/`. (Citation on page 60).

[91] **CMS** Collaboration, Fanfani et al., "Distributed Data Management in CMS." "http://cdsweb.cern.ch/record/933704, Proceedings of the conference on computing in high energy physics 2006, Mumbai, India. (Citation on page 60).

[92] **CMS** Collaboration, J. Rehn et al., "PhEDEx high-throughput data transfer management system,". Proceedings of the conference on computing in high energy physics 2006, Mumbai, India. (Citation on page 61).

[93] Python `http://www.python.org/`. (Citations on pages 62, 63).

[94] Physics Analysis Workstation PAW `http://wwwasd.web.cern.ch/wwwasd/paw/`. (Citation on page 62).

[95] matplotlib `http://matplotlib.sourceforge.net/`. (Citation on page 63).

[96] SCRAM (Software Configuration, Release And Management) `http://cmsdoc.cern.ch/Releases/SCRAM/doc/scramhomepage.html`. (Citation on page 63).

[97] Geant Simulation Toolkit `http://wwwasd.web.cern.ch/wwwasd/geant4/geant4.html`. (Citations on pages 63, 68).

[98] **GEANT4** Collaboration, S. Agostinelli et al., "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303. `doi:10.1016/S0168-9002(03)01368-8`. (Citations on pages 63, 63, 68, 68).

[99] S. Kosyakov et al., "Frontier : High Performance Database Access Using Standard Web Components in a Scalable Multi-tier Architecture," *Computing in High Energy Physics and Nuclear Physics 2004, Interlaken, Switzerland, 27 Sep - 1 Oct 2004* (2004) 669. (Citation on page 65).

[100] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, "Multiparton interactions in photoproduction at HERA," *Z. Phys.* **C72** (1996) 637–646, `arXiv:hep-ph/9601371`. `doi:10.1007/s002880050286`. (Citations on pages 67, 94).

[101] S. R. Slabospitsky and L. Sonnenschein, "TopReX generator (version 3.25): Short manual," *Comput. Phys. Commun.* **148** (2002) 87–102, `arXiv:hep-ph/0201292`. `doi:10.1016/S0010-4655(02)00471-X`. (Citation on page 67).

[102] Z. Nagy, "Three-jet cross sections in hadron hadron collisions at next-to-leading order," *Phys. Rev. Lett.* **88** (2002) 122003, arXiv:hep-ph/0110315. doi:10.1103/PhysRevLett.88.122003. (Citations on pages 68, 72).

[103] Z. Nagy, "Next-to-leading order calculation of three-jet observables in hadron-hadron collision," *Phys. Rev.* **D68** (2003) 094002, arXiv:hep-ph/0307268. doi:10.1103/PhysRevD.68.094002. (Citation on page 68).

[104] Z. Nagy. NLOJET++ Homepage http://www.cpt.dur.ac.uk/~nagyz/nlo++/. (Citation on page 68).

[105] S. Catani and M. H. Seymour, "A general algorithm for calculating jet cross sections in NLO QCD," *Nucl. Phys.* **B485** (1997) 291–419, arXiv:hep-ph/9605323. doi:10.1016/S0550-3213(96)00589-5. (Citation on page 68).

[106] T. Kluge, K. Rabbertz, and M. Wobisch, "Fast pQCD calculations for PDF fits," *arXiv* (2006) arXiv:hep-ph/0609285. (Citations on pages 68, 72, 73).

[107] T. Kluge, K. Rabbertz, and M. Wobisch, "fastNLO - fast pQCD calculations for hadron-induced processes." http://hepforge.cedar.ac.uk/fastnlo, 2005. Web Page. (Citations on pages 68, 68, 72, 72).

[108] M. R. Whalley, D. Bourilkov, and R. C. Group, "The Les Houches Accord PDFs (LHAPDF) and Lhaglue," arXiv:hep-ph/0508110. (Citations on pages 72, 82).

[109] M. R. Whalley, D. Bourilkov, and R. C. Group, "LHAPDF - the Les Houches Accord PDF Interface." http://hepforge.cedar.ac.uk/lhapdf, 2005. Web Page. (Citations on pages 72, 72, 82, 82).

[110] W. K. Tung et al., "Heavy quark mass effects in deep inelastic scattering and global QCD analysis," *JHEP* **02** (2007) 053, arXiv:hep-ph/0611254. (Citations on pages 72, 82).

[111] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175. (Citations on pages 72, 94, 152).

[112] W. T. Giele, E. W. N. Glover, and J. Yu, "The Determination of α_S at Hadron Colliders," *Phys. Rev.* **D53** (1996) 120–130, arXiv:hep-ph/9506442. doi:10.1103/PhysRevD.53.120. (Citation on page 80).

[113] P. M. Nadolsky et al., "Implications of CTEQ global analysis for collider observables," *Phys. Rev.* **D78** (2008) 013004, arXiv:0802.0007. doi:10.1103/PhysRevD.78.013004. (Citations on pages 81, 82).

[114] D. Stump et al., "Inclusive jet production, parton distributions, and the search for new physics," *JHEP* **10** (2003) 046, arXiv:hep-ph/0303013. (Citations on pages 82, 128).

[115] H. Collaboration and Z. Collaboration, "H1 and ZEUS Combination preliminary results." http://www-h1.desy.de/publications/H1ZEUSpreliminary.html, 2008. Web Page. (Citation on page 82).

[116] H. Jung et al., "Proceedings of the workshop: HERA and the LHC workshop series on the implications of HERA for LHC physics," arXiv:0903.3861. (Citation on page 82).

[117] M. Gluck, P. Jimenez-Delgado, and E. Reya, "Dynamical parton distributions of the nucleon and very small-x physics," *Eur. Phys. J.* **C53** (2008) 355–366, arXiv:0709.0614. doi:10.1140/epjc/s10052-007-0462-9. (Citations on pages 82, 83).

[118] **NNPDF** Collaboration, R. D. Ball et al., "A determination of parton distributions with faithful uncertainty estimation," *Nucl. Phys.* **B809** (2009) 1–63, arXiv:0808.1231. doi:10.1016/j.nuclphysb.2008.09.037. (Citations on pages 82, 83).

[119] J. Pumplin et al., "New generation of parton distributions with uncertainties from global QCD analysis," *JHEP* **07** (2002) 012, arXiv:hep-ph/0201195. (Citations on pages 83, 91, 95, 104, 152).

[120] A. Sherstnev and R. S. Thorne, "Parton Distributions for LO Generators," *Eur. Phys. J.* **C55** (2008) 553–575, arXiv:0711.2473. doi:10.1140/epjc/s10052-008-0610-x. (Citation on page 87).

[121] D. E. Acosta, F. Ambroglini, P. Bartalini, A. de Roeck, L. Fano, R. Field, and K. Kotov, "The Underlying Event at the LHC," Technical Report CMS-NOTE-2006-067. CERN-CMS-NOTE-2006-067, CERN, Geneva, Jun, 2006. (Citations on pages 91, 152).

[122] **CMS** Collaboration, M. Heinrich, A. Oehler, and K. Rabbertz, "NLO calculations for Inclusive Jets with fastNLO at 10 TeV," *CMS Internal Analysis Note* **CMS-AN-2009/058** (2009). (Citation on page 92).

[123] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations," *JHEP* **06** (2002) 029, arXiv:hep-ph/0204244. (Citations on pages 94, 128).

[124] **D0** Collaboration, V. M. Abazov et al., "The inclusive jet cross-section in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV using the k_T algorithm," *Phys. Lett.* **B525** (2002) 211–218, arXiv:hep-ex/0109041. doi:10.1016/S0370-2693(01)01441-1. (Citation on page 94).

[125] S. Gieseke et al., "Herwig++ 2.0 release note," arXiv:hep-ph/0609306. (Citations on pages 94, 129).

[126] G. Corcella et al., "HERWIG 6.5 release note," arXiv:hep-ph/0210213. (Citation on page 94).

[127] **UA5** Collaboration, G. J. Alner et al., "The UA5 High-Energy anti-p p Simulation Program," *Nucl. Phys.* **B291** (1987) 445. doi:10.1016/0550-3213(87)90481-0. (Citation on page 95).

[128] S. Alekhin et al., "HERA and the LHC - A workshop on the implications of HERA for LHC physics: Proceedings Part A," arXiv:hep-ph/0601012. (Citations on pages 95, 96).

[129] R. Field, "Monte-Carlo Generators for CMS." CMS MC Generators Meeting, 2007. (Citations on pages 95, 104, 152, 152).

[130] M. Bahr, S. Gieseke, and M. H. Seymour, "Simulation of multiple partonic interactions in Herwig++," *JHEP* **07** (2008) 076, arXiv:0803.3633. doi:10.1088/1126-6708/2008/07/076. (Citation on page 96).

[131] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "MRST2001: Partons and alpha(s) from precise deep inelastic scattering and Tevatron jet data," *Eur. Phys. J.* **C23** (2002) 73–87, arXiv:hep-ph/0110215. doi:10.1007/s100520100842. (Citation on page 96).

[132] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "Uncertainties of predictions from parton distributions. I: Experimental errors. ((T))," *Eur. Phys. J.* **C28** (2003) 455–473, arXiv:hep-ph/0211080. doi:10.1140/epjc/s2003-01196-2. (Citation on page 96).

[133] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "NNLO global parton analysis," *Phys. Lett.* **B531** (2002) 216–224, arXiv:hep-ph/0201127. doi:10.1016/S0370-2693(02)01483-1. (Citation on page 96).

[134] B. Kersevan, "Monte Carlo Generators for LHC at ATLAS." "http://epiphany.ifj.edu.pl/epiphany.2007/pres/bkersevan.pdf", 2007. Cracow Epiphany Conference on Precision Physics and Monte Carlos for LHC. (Citation on page 96).

[135] **D0** 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:hep-ex/0012046. doi:10.1103/PhysRevD.64.032003. (Citations on pages 97, 117).

[136] **D0** Collaboration, M. Voutilainen, "Inclusive jet cross section measurement at D0," arXiv:hep-ex/0609026. (Citation on page 97).

[137] **CMS** Collaboration, "Plans for Jet Energy Corrections at CMS," *CMS Physics Analysis Summary* **CMS-PAS-JME-07-002 - http://cms-physics.web.cern.ch/cms-physics/public/JME-07-002-pas.pdf** (2008). (Citations on pages 97, 106, 121).

[138] M. Dasgupta, L. Magnea, and G. P. Salam, "Non-perturbative QCD effects in jets at hadron colliders," *JHEP* **02** (2008) 055, arXiv:0712.3014. doi:10.1088/1126-6708/2008/02/055. (Citation on page 98).

[139] C. Buttar et al., "Standard Model Handles and Candles Working Group: Tools and Jets Summary Report," arXiv:0803.0678. (Citation on page 98).

[140] **CMS** Collaboration, M. Heinrich, M. Kaur, A. Oehler, K. Rabbertz, and M. Zubin Mehta, "Non-perturbative Corrections to Inclusive Jet Spectra at 10 TeV," *CMS Internal Analysis Note* **CMS-AN-2009/033** (2009). (Citation on page 99).

[141] R. Demina, J. Dolen, C. Justus, P. Tipton, M. Zielinski, A. Bhatti, and R. Harris, "Calorimeter Cell Energy Thresholds for Jet Reconstruction in CMS," Technical Report CMS-NOTE-2006-020. CERN-CMS-NOTE-2006-020, CERN, Geneva, Dec, 2005. (Citation on page 106).

[142] **CMS** Collaboration, "Offset Energy Correction for Cone Jets," *CMS Physics Analysis Summary* **CMS-PAS-JME-09-003 - http://cms-physics.web.cern.ch/cms-physics/public/JME-09-003-pas.pdf** (2009). (Citation on page 106).

[143] **CMS** Collaboration, "Determination of the Relative Jet Energy Scale at CMS from Dijet Balance," *CMS Physics Analysis Summary* **CMS-PAS-JME-08-003 - http://cms-physics.web.cern.ch/cms-physics/public/JME-08-003-pas.pdf** (2009). (Citation on page 106).

[144] **CMS** Collaboration, "Jet energy calibration with photon+jet events," *CMS Physics Analysis Summary* **CMS-PAS-JME-09-004 - http://cms-physics.web.cern.ch/cms-physics/public/JME-09-004-pas.pdf** (2009). (Citations on pages 106, 107).

[145] **CMS** Collaboration, "Jet energy correction using $Z \to e^+e^-+$ jet p_T balance and procedure for combining data driven corrections," *CMS Physics Analysis Summary* **CMS-PAS-JME-09-005 - http://cms-physics.web.cern.ch/cms-physics/public/JME-09-005-pas.pdf** (2009). (Citations on pages 106, 107).

[146] **CMS** Collaboration, "Calibration of the absolute jet energy scale with $Z \to \mu^+mu^-+$ jet events at CMS," *CMS Physics Analysis Summary* **CMS-PAS-JME-09-009 - http://cms-physics.web.cern.ch/cms-physics/public/JME-09-009-pas.pdf** (2009). (Citation on page 106).

[147] **CMS** Collaboration, D. Acosta, M. Della Negra, L. Foà, A. Hervé, and A. Petrilli, eds., "CMS Physics: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 2006. (Citations on pages 106, 113).

[148] **CMS** Collaboration, M. Heinrich, P. Jindal, K. Kousouris, D. Mason, A. Oehler, and K. Rabbertz, "Inclusive Jet Cross Section at 10 TeV with CMS," *CMS Internal Analysis Note* **CMS-AN-2009/034** (2009). (Citations on pages 107, 108).

[149] **CMS** Collaboration, "Performance of Jet Algorithms in CMS," *CMS Physics Analysis Summary* **CMS-PAS-JME-07-003 - http://cms-physics.web.cern.ch/cms-physics/public/JME-07-003-pas.pdf** (2008). (Citation on page 110).

[150] **CMS** Collaboration, "Jet Reconstruction Performance at CMS," *CMS Physics Analysis Summary* **CMS-PAS-JME-09-007 - http://cms-physics.web.cern.ch/cms-physics/public/JME-09-007-pas.pdf** (2009). (Citations on pages 111, 123).

[151] **CMS** Collaboration, "Initial Measurement of the Inclusive Jet Cross Section at 10 TeV," *CMS Physics Analysis Summary* **CMS-PAS-QCD-08-001 - http://cms-physics.web.cern.ch/cms-physics/public/QCD-08-001-pas.pdf** (2009). (Citations on pages 113, 114, 131, 131, 135, 151).

[152] S. Frixione and B. R. Webber, "The MC@NLO 3.3 event generator," `arXiv:hep-ph/0612272`. (Citation on page 128).

[153] F.-M. Stober, "Study of Three-Jet Topologies and first Event Simulation with the Event Generator Herwig++ for the CMS experiment at the LHC," *IEKP Note, Diploma Thesis* **IEKP-KA/2008-28** (2008). (Citation on page 131).

[154] E. Eichten, K. D. Lane, and M. E. Peskin, "New Tests for Quark and Lepton Substructure," *Phys. Rev. Lett.* **50** (1983) 811–814. `doi:10.1103/PhysRevLett.50.811`. (Citation on page 131).

[155] J. Bjorken and S. Drell, "Relativistic Quantum Mechanics and Relativistic Quantum Fields". McGraw-Hill, New York, 1964. (Citation on page 141).

[156] Tera-scale Open-source Resource and QUEue manager TORQUE `http://www.clusterresources.com/products/torque/`. (Citation on page 148).

[157] **CTEQ** Collaboration, H. L. Lai et al., "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions," *Eur. Phys. J.* **C12** (2000) 375–392, `arXiv:hep-ph/9903282`. `doi:10.1007/s100529900196`. (Citation on page 152).

Acknowledgements

I wish to sincerely thank my referee Prof. Dr. Günter Quast. His guidance and support allowed me to present this work after a very interesting and instructive period. This included twelve exciting months at CERN, which gave me insight into some of the complex internal structures of the CMS Collaboration. At this point I want to thank Prof. Dr. Thomas Müller who also helped to make this possible. Additionally, I wish to thank Prof. Dr. Wim de Boer for accepting the task of being my co-referee.

Most of all I have to thank my adviser Klaus Rabbertz. Due to his outstanding commitment especially in the field of QCD related physics it was a pleasure to work with him. He always kindly answered all sorts of questions and with his close proofreading very much helped in improving the current state of the presented thesis.

Additionally, I thank all the people of the Karlsruhe CMS group at the EKP. It was a real pleasure to be part of this group, which always granted a pleasant working atmosphere. There was always someone interested in discussing all kinds of physical and technical questions. Out of this several software projects have been started, which I hope also in the future will help students to increase their physics productivity and ease some technical frustrations.

For their commitment to this and for the close proofreading of this thesis I have to thank Volker Büge, Joram Berger, Kristov Hackstein, Michael Heinrich, Peter Krauss, Oliver Oberst, Danilo Piparo, Christophe Saout, Armin Scheurer, Fred Stober, Anja Vest, Matthias Wolf, and Manuel Zeise.

I also have to thank all the people who contributed to the CMS Physics Analysis Sum-

mary “Initial Measurement of the Inclusive Jet Cross Section at $10\ \mathrm{TeV}$”. Therefore I have to thank the CMS QCD group and the members of the review committee. Especially I thank Michael Heinrich, Pratima Jindal, Kostas Kousouris, Klaus Rabbertz, and Manuk Zubin Mehta for their valuable contributions.

At this point I want to say thank you to all people that I had the pleasure to learn to know and work with during my stay at CERN for the great time.

I would like to thank all the members of the EKP-admin team for successfully maintaining the computing cluster, the grid services, and the desktop systems.

Many thanks to all other members of the Institut für Experimentelle Kernphysik and all those I might have forgotten for the great time.

Last, but not least, I thank my parents very much for their support and confidence in me.