Karlsruhe Institute of Technology

A Jet Based Approach to Measuring Soft Contributions to Proton-Proton Collisions with the CMS Experiment

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. Michael Heinrich

Tag der mündlichen Prüfung: 21. Januar 2011

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

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

'The Whole World is our Playground'

A Jet-Based Approach to Measuring Soft Contributions to Proton Collisions with the CMS Experiment

The early running stages of the Large Hadron Collider (LHC) offered perfect conditions to study the Underlying Event (UE) in proton collisions. This term denotes all effects in collider events that are not directly related to the hard partonic interaction. As these effects are present in all collisions, it is necessary to include them in Monte Carlo simulations. It is however not possible to calculate UE contributions with means of perturbation theory, which makes the application of phenomenological models necessary. These models have to be tuned to data, a process that has to be repeated at every newly accessed collision energy. At the start of the LHC program, different tunes have been available that were derived from Tevatron data and extrapolated to LHC energies, yielding large deviations in the predictions of UE contributions, depending on the scaling behaviour of the models.

The influence of different tunes for the underlying event is demonstrated by investigating their influence on the inclusive jet spectrum. This quantity is a basic observable at hadron colliders yet the theoretical prediction in Monte Carlo generators is available to the particle level only in leading order of perturbative QCD. Therefore, correction factors have to be applied to next-to-leading matrix-element calculations to account for effects of the UE and hadronization. It is shown in this thesis that these correction factors strongly depend on the applied UE model and that by using multiple models, a systematic uncertainty on the correction factors can be estimated.

Furthermore, a new and complimentary method to investigate the underlying event is applied for the first time in this work. While traditionally, these effects have been studied by investigating the geometrical area transverse to a leading object in the event, the new approach uses the ratio of the transverse momenta of all jets in the event divided by their areas. This is the first exertion of the concept of jet areas in this context. In order to cover the phase space to particle momenta as low as possible, reconstructed tracks are used as input for the analysis. The silicon tracking detector of the CMS experiment proves to be an outstanding tool for this kind of measurement. Data taken at two different center-of-mass energies are studied: 0.9 TeV and 7 TeV recorded in 2009 and 2010 respectively. It is shown that all Monte Carlo tunes that were produced prior to LHC operation underestimate the event activity, a result that is in line with other analyses. A new model that includes first LHC results is tested as well, showing promising results.

The new method of quantifying soft activity in hadron collisions applied here for the first time offers a new perspective also towards the subtraction of hadronic noise from pile-up events, a task that will benefit from the experience gained in this work.

Contents

Introduction **1**

1 The Standard Model of Particle Physics **5**
1.1 From Rutherford to Deep Inelastic Scattering 5
1.1.1 Cross-Sections 6
1.1.2 Elastic Electron-Proton Scattering 7
1.1.3 Deep Inelastic Scattering and Parton Distribution Functions 8
1.2 Standard Model Principles and Formalism 10
1.2.1 Mathematical Methods 12
1.2.2 Particles and Forces 12
1.2.3 Free Parameters 13
1.3 Quantum Chromodynamics 14
1.3.1 The QCD Lagrangian and the Running Coupling 16
1.3.2 Hard QCD Scattering and Factorization 19
1.4 Monte Carlo Methods 21
1.4.1 Matrix-Element Calculation 22
1.4.2 Parton Shower 22
1.4.3 Hadronization 23
1.4.4 The Underlying Event and Multiple Parton Interactions 23
1.4.5 Monte Carlo Event Generators 25
1.5 Jet Algorithms 27
1.5.1 Cone-Type Algorithms 27
1.5.2 Combinatorial Algorithms 29
1.5.3 Jet Areas 31

2 The CMS Experiment at the Large Hadron Collider **33**
2.1 The Large Hadron Collider 33

2.2 The CMS Experiment . . . 36
2.2.1 The Solenoid Magnet . . . 38
2.2.2 The Tracking System . . . 38
2.2.3 The Electromagnetic Calorimeter . . . 41
2.2.4 The Hadron Calorimeter . . . 41
2.2.5 The Muon System . . . 44
2.2.6 Trigger and Data Acquisition . . . 45
2.3 The Worldwide LHC Computing Grid . . . 46
2.4 Software Concepts . . . 48
2.4.1 The Event Data Model (EDM) . . . 48
2.4.2 CMSSW . . . 49
2.4.3 ROOT . . . 49
2.4.4 Jet Reconstruction with FastJet . . . 50

3 Underlying Event Contributions to Proton Collisions 53
3.1 Underlying Event Contributions to the Inclusive Jet Cross-Section . . . 54
3.1.1 The Inclusive Jet Cross-Section . . . 54
3.1.2 Preparational Study for CMS . . . 55
3.1.3 First Results from CMS . . . 65
3.1.4 Measurements at the Tevatron . . . 67
3.2 The Traditional Approach to Measuring the Underlying Event . . . 68
3.2.1 Measurements . . . 69
3.3 The Jet Area/Median Approach to Measuring the Underlying Event . . . 72
3.3.1 Theoretical Background . . . 72
3.3.2 Proposed Measurement . . . 74

4 Measurement of the Charged Underlying Event Activity with the CMS Detector at 0.9 and 7 TeV with the Jet Area/Median Approach 77
4.1 Data Samples . . . 78
4.1.1 0.9 TeV Data from the 2009 Commissioning Run . . . 78
4.1.2 0.9 TeV Monte Carlo Samples . . . 79
4.1.3 7 TeV Data and Monte Carlo . . . 80
4.2 Track Selection . . . 82
4.3 Jet Reconstruction . . . 87
4.4 Sensitivity . . . 92
4.5 Systematic Uncertainties . . . 94
4.5.1 Tracker Material Budget . . . 96
4.5.2 Tracker Alignment . . . 96
4.5.3 Non-Operational Tracker Channels . . . 97
4.5.4 Trigger Efficiency . . . 98

4.5.5 Vertex Reconstruction . . . 99
4.5.6 Track Reconstruction Efficiency and Fake Rate . . . 99
4.5.7 Variations of the Track Selection . . . 100
4.5.8 Transverse Momentum Resolution . . . 101
4.5.9 Track-Jet Response . . . 102
4.5.10 Overall Systematic Uncertainty . . . 102
4.6 Results . . . 107

5 Summary and Perspectives **111**

A Monte Carlo Tunes **113**
A.1 Pythia 6 . . . 113
A.1.1 Tunable Parameters . . . 114
A.1.2 Ancient Tunes . . . 115
A.1.3 Old Tunes . . . 115
A.1.4 New Tunes . . . 117
A.2 Pythia 8 . . . 118
A.3 Herwig/JIMMY and Herwig++ . . . 118

B Additional Information on Software Setup and Datasets **119**
B.1 The JUEZ data format . . . 119
B.2 Data Access and Software Setup . . . 120
B.2.1 The National Analysis Facility at DESY . . . 120
B.2.2 Grid-Control . . . 121
B.3 Data Samples . . . 121
B.3.1 0.9 TeV Data from the 2009 Commisioning Run and Monte Carlo Samples 122
B.3.2 7 TeV Data and Monte-Carlo . . . 124

C Supplemental Plots **125**
C.1 Track and Charged Particle Distributions . . . 125
C.2 Jet Distributions . . . 128
C.3 Systematic Uncertainties . . . 132

List of Figures **135**

List of Tables **139**

Bibliography **141**

Introduction

Particle physics aims at understanding the fundamental properties of the smallest building blocks of matter and their interactions. Both experimental and theoretical research over the past decades has culminated in the formulation of the Standard Model of Particle Physics. This theory, despite being widely accepted and proving its worth by accurately predicting countless observations in experimental setups, still has a number of unpleasant features. One of the unsolved problems is the mechanism that is responsible for giving particles their masses. In order to confirm or discard the most popular explanation for masses, the Higgs mechanism, enormous energy densities are required in an experiment. Further questions that are unanswered so far and that can be addressed by such a powerful machine are, among others, the nature of the dark matter that is predicted by cosmology and the existence of heavy, yet undetected particles.

The recent start of the physics program of the Large Hadron Collider (LHC) at the European Center for Nuclear Research (CERN) marks the beginning of a new epoch in particle physics. The Standard Model of Particle Physics, which is already one the most thoroughly tested theories in all of science will be probed at a new energy domain. As in any hadron collider, the dominant physical processes follow the laws of quantum chromodynamics (QCD), the theory of the strong interaction. The typical representatives of QCD activity in hadron collider experiments are jets, bundles of collinear particles supposedly originating from the same parton that has taken part in a hard partonic interaction. In order to recombine particles to jets, different clustering algorithms are employed. The implications of jet physics at the LHC are manifold, from basic observables such as spectra of jets in transverse momentum, to the physics of top and bottom quarks that produce jets with a characteristic profile to the search of heavy resonances that are predicted to decay into quarks and manifest themselves as jets.

Furthermore, QCD contributes an irreducible background through pile-up and underlying event to all hadron collissions. Pile-up denotes extra activity in an event either from additional proton interactions in the same bunch crossing or as out-of-time pile-up from neighbouring bunch-crossings. The complicated multi-parton dynamics within the interacting protons that pollute the clear signa-

tures of the hard partonic interactions are referred to as the underlying event.
An important aspect of particle physics is the correct description of the properties of particle collisions with the help of Monte Carlo event generators. Only if one can trust the modelling of known processes, it is possible to exclude or confirm the existence of new particles and phenomena. While the hard partonic processes can be calculated precisely with the techniques of perturbative calculation, soft QCD contributions are excluded from this treatment due to the diverging of the strong coupling constant for small momentum transfers. Thus, phenomenological models have to be applied to predict the soft hadronic activity in all events. Tuning these models is an integral activity in the early stages of every new hadron collider, since the energy scaling behaviour of the models are non-trivial and the interpolations of existing models usually have to be adjusted to the new conditions.

The work presented here comprises studies related to QCD phenomena and the understanding of jet algorithms and underlying event models. The two main contributions are an investigation of the influence of non-perturbative models in Monte-Carlo generators on the inclusive jet cross-section and the first investigation of the new Jet Area/Median approach to measuring soft hadronic activity.
The first chapter of this thesis is dedicated to a brief introduction of the Standard Model of Particle Physics and especially the theory of the strong interaction between colour-charged partons. Additional emphasis is put on the description of the techniques applied in Monte Carlo event generation, one of the most important tools in particle physics. Also, an overview of the most common jet algorithms is given including an explanation of the concept of jet areas in recombinatorial algorithms.
In the second chapter, the LHC and the CMS experiment are introduced as well as the Worldwide LHC Computing Grid (WLCG). This overview includes a discussion of some important reconstruction algorithms and software concepts. The technicalities of jet reconstruction in CMS are presented as well.
The work on the non-perturbative corrections to the inclusive jet spectrum is presented in chapter three. The study on this subject was part of a preperational analysis that paved the way for the actual measurement after the start of the LHC physics program. The techniques investigated in this work have been applied in the results presented at the ICHEP 2010 conference [1] on the inclusive jet cross-section, which is also presented briefly.
In the context of non-perturbative contributions to jet measurements, an introduction is given how these effects are quantified in hadron collisions and how the modelling is steadily improved with measurements at different center-of-mass energies. Also, a recently proposed completely new method of measuring the soft QCD activity is presented, the Jet Area/Median approach [2]. This new technique takes a more holistic path to quantifying the implications of underlying event effects to event topologies.
The fourth chapter is dedicated to the first application of the Jet Area/Median approach on actual detector data at the two center-of-mass energies of 0.9 and 7 TeV. As only little data was available with suitable beam conditions, several compromises had to be made concerning event selection and

trigger setup in comparison with the original proposal for such a measurement. These compromises in turn require an additional adjustment of the proposed observable itself which compensates for the low average activity in the selected events. Yet, even despite the low activity and the necessary adaptation of the observable, the sensitivity to different Monte Carlo models could be proven. The adjacent comparison between data and Monte Carlo after full detector simulation with exhaustive treatment of systematic uncertainties yielded interesting results complementary to the classically derived underlying event observables.

Chapter 1

The Standard Model of Particle Physics

One of the most basic questions in science is the search for the fundamental building blocks of matter. Already in ancient Greece, philosophical discussions were ongoing on whether there are smallest, indivisible units, called *àtomos*, or if matter can be split ad infinitum. With the discovery of the principle of constant proportions, the existence of atoms which are unique for each chemical element, was proved by John Dalton in the early 19th century. The question of electrical charges indicated however, that atoms were composite objects themselves. When J.J. Thompson discovered the electron in his research on cathode rays, an entirely new field of physics was born. His interpretation of these findings led to the *plum pudding* model, which proclaimed atoms to consist of positively charged matter, in which the negatively charged electrons are embedded. Only a few years later, the scientific world was revolutionized by the ideas of quantum mechanics and the foundation was laid for a new era in physics.

1.1 From Rutherford to Deep Inelastic Scattering

In 1903, Ernest Rutherford conducted his famous experiment which refuted the previous atomic models [3]. When he exposed a thin gold foil to alpha rays from a radioactive source, he discovered that many of the alpha particles could traverse the foil without any distraction, while some were scattered even backwards. Analyzing the angular distribution of the outgoing particles, he came to the conclusion, that atoms consist of a very small, positively charged core around which the electrons are orbiting.
Apart from his actual discovery, Rutherford also set the scene for sub-atomic research by probing matter with high energetic particles and interpreting the distribution of the reaction products. According to Louis de Broglie [4], the wavelength of a particle is inverse proportional to its rela-

tivistic energy. Therefore the smaller the structure one wants to investigate with particle rays, the higher the probe energy has to be. This development led to accelerator machines ever increasing in size over the decades. At first, fixed target experiments were used, later collider layouts which have a favourable energy scaling behaviour. At a certain point in increasing collision energy, yet unknown heavy unstable particles started to be produced. So in the 1950s and 60s, a vast number of new particles were observed in the cloud and bubble chambers at the accelerator laboratories. Murray Gell-Mann and George Zweig proposed in 1964 that the discovered hadrons themselves were actually not elementary but composed of smaller entities, called quarks [5]. Originally, the model contained three flavors of quarks: up, down and strange. With the deep inelastic scattering experiments at SLAC[1] which started in 1968, it was revealed that protons indeed have a substructure and contain point-like constituents that were identified with the proposed quarks [6].

1.1.1 Cross-Sections

The cross-section of a physics process describes its probability to occur. For the example of a collider experiment, the cross-section makes it possible to estimate the event rate of a given process when the beam parameters of the accelerator are known. It is commonly used differentially in variables of the phase space such as pseudo-rapidity [2] η or transverse momentum p_{T}. Deviations from existing theories are reflected in changes of counting rates not predicted by the Standard Model which can for example appear as peaks in the cross-section in case of heavy resonances. The cross-section is defined as the reaction rate W per unit of flux Φ of incoming particles and can geometrically be interpreted as the front face of classical scatterers:

$$\sigma = \frac{W}{\Phi} \tag{1.1}$$

While the SI unit for the cross-section is m^2, with the front face of a proton being $R_{\mathrm{p}} \approx 10^{-30}\ \mathrm{m}^2$ it is more convenient to use the unit *barn*, with $1\ \mathrm{b} = 10^{-28}\ \mathrm{m}^2$. In turn, the event rates of particularly well understood processes like Drell-Yan production of muon pairs, where the cross-section is known theoretically to excellent precision, are well suited to measure the luminosity $\mathcal{L}$ of a collider experiment:

$$N = \sigma \int \mathcal{L} dt \tag{1.2}$$

The quantum-mechanical interaction rate for a given process can be calculated with Fermi's Golden Rule [7] from the corresponding transition matrix element $\mathcal{M}_{\mathrm{if}}$, which in turn is calculated from the interaction Hamiltonian, and the available phase space state density for the reaction products

[1] Stanford Linear Accelerator Collider

[2] $\eta = -ln\left(tan\frac{\theta}{2}\right)$, where θ is the angle between the particle momentum and the beam axis

ρ_{f}:

$$W = \frac{2\pi}{\hbar}|\mathcal{M}_{\mathrm{fi}}|^2\rho_{\mathrm{f}} \tag{1.3}$$

$$\mathcal{M}_{\mathrm{fi}} = \langle\Psi_{\mathrm{f}}|\mathcal{H}_{\mathrm{int}}|\Psi_{\mathrm{i}}\rangle \tag{1.4}$$

1.1.2 Elastic Electron-Proton Scattering

In the context of particle physics, collisions that do not destroy the target or the bullet are referred to as elastic scattering processes. The understanding of the structure of the proton can benefit especially from probing it with electrons, which appear point-like in all experiments conducted up to now. Also, at relatively low energies, the electromagnetic interaction between the electron and the protons is well understood and easily calculable. To be able to observe the substructure of the proton, electrons with a de-Broglie wavelength that is smaller than the radius of the proton of about 1 fm, are needed. To describe the elastic interaction between the proton and the electron, it is convenient to use the momentum transfer as a measure for the impact, where k and k' are the four-vectors of the electron before and after the scattering:

$$Q^2 = -q^2 = -(k - k')^2 \tag{1.5}$$

Both the proton and the electron are fermions with intrinsic spin of $\hbar/2$, which has to be accounted

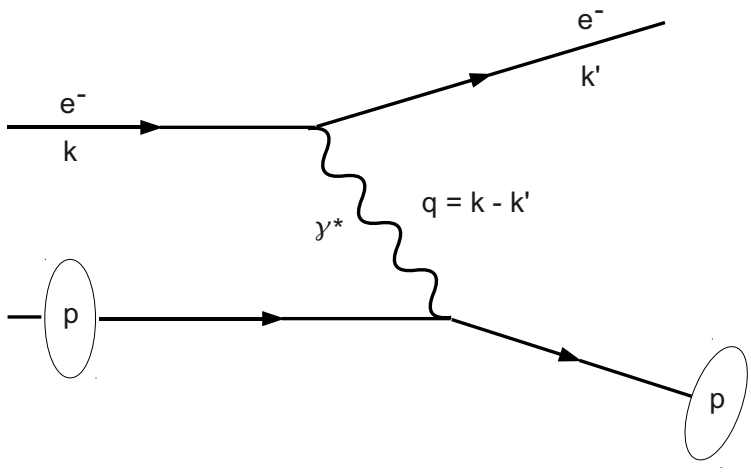


Figure 1.1: *Elastic electron-proton scattering with momentum transfer q. The proton stays intact in the interaction.*

for when calculating the differential scattering cross-section, as the electric charges interact with the magnetic momenta. With the denotion

$$\tau = \frac{Q^2}{4M_{\mathrm{p}}^2c^2}. \tag{1.6}$$

for the kinematic quantities and the total relativistic energy of the fermion

$$W^2 = p^2c^2 + m^2c^4, \tag{1.7}$$

taking into account the the Mott cross-section [8] $\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}}$ of the process

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} = \left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \frac{W^2}{(qc)^4}\left[1 - \frac{v^2}{c^2}sin^2\left(\frac{\theta}{2}\right)\right] \tag{1.8}$$

that describes the elastic scattering of two fermions with one elementary charge each, the Rosenbluth formula can be derived as:

$$\left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right) = \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right)_{\text{Mott}} \cdot \left[\frac{G_E^2(Q^2) + \tau G_M^2(Q^2)}{1+\tau} + 2\tau G_M^2(Q^2)\tan^2\frac{\theta}{2}\right]. \tag{1.9}$$

Here, $G_E^2(Q^2)$ and $G_M^2(Q^2)$ are the electric and magnetic form factors. The latter of which arises from the fact that moving electric charges evoke magnetic fields. Form factors are the Fourier transforms of charge distributions, that can easily be interpreted in the parton picture. Early experiments measuring the form factors of the protons in dependence of Q^2 showed that neither a homogeneous charge distribution nor static point-like charges were observed in the proton. The very complicated dynamics of the partons inside a proton are subject to experimental and theoretical research still today.

1.1.3 Deep Inelastic Scattering and Parton Distribution Functions

Using electron probes with an energy of more than about 5 GeV led to the observation of inelastic scattering in which the proton is destroyed and short-lived excited states are being produced. These can only be explained if the proton indeed consists of smaller building blocks. To calculate the cross-section of processes that invoke excited states, the excitement energy $\nu = E - E'$ which corresponds to the mass of the resonance has to be taken into account:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}\Omega\mathrm{d}E'} = \left(\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}\right) \cdot \left[W_2(Q^2,\nu) + 2W_1(Q^2,\nu)\tan^2\frac{\theta}{2}\right] \tag{1.10}$$

$$W^2c^4 = P'^2 = (P+q)^2 = M^2c^4 + 2Pq + q^2 = M^2c^4 + 2Mc^2\nu - Q^2 \tag{1.11}$$

Here, W_1 and W_2 are the form factors of the electric charge and the magnetic momentum. It was shown already at SLAC that the form factors hardly show a dependence of Q^2. This is a strong indication for the existence of point-like partons inside the proton, taking into account formula 1.8. At this point, *Bjorken scaling* is introduced, which is a measure for the fraction of the overall proton

momentum that is carried by the particular parton in the "infinite momentum frame":

$$x := \frac{Q^2}{2Pq} = \frac{Q^2}{2M\nu} \tag{1.12}$$

With the help of the Bjorken scale, the form factors can be rewritten as:

$$F_1(x, Q^2) = Mc^2 W_1(Q^2, \nu) \tag{1.13}$$
$$F_2(x, Q^2) = \nu W_2(Q^2, \nu). \tag{1.14}$$

The double differential cross section of the inelastic electron-proton scattering can now be written using the inelasticity $\hat{y} = \nu/E$, as:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\mathrm{d}\hat{y}} = \frac{4\pi\alpha^2(s - M^2)}{Q^4} \cdot \left[(1 - \hat{y})F_2(x) + \hat{y}^2 x F_1(x) - \frac{M^2}{s - M^2} x\hat{y} F_2(x)\right] \tag{1.15}$$

The confirmation of this relation [9,10] in particle physics experiments was a triumph for the quark model, which contributed to the formulation of the Standard Model of Particle Physics.
Measuring the fractions of the proton momentum that each quark is carrying, it was found that they add up to only about 0.5, and that they follow particular distributions. This was an early indirect indication for the existence of gluons which carry the undetected fraction of the momentum, even though they do not interact with the electron probes. Also, apart from the expected three quarks, other flavors and also anti-quarks are observed, which arise from gluon splitting processes. More on the strong interaction that binds quarks to hadrons is found in section 1.3. The parton distribution functions (PDFs), that describe the probability to find a parton with a given flavor at a certain Bjorken-x, are still under investigation today by a number of working groups like CTEQ [11] and MSTW [12]. An example from a recent CTEQ publication is shown in Fig. 1.2 for two different values of Q^2. It can be seen that a large portion of the overall proton momentum is carried by low-x gluons and also *sea quarks* can be observed that are not expected in a simple three quark model. These quarks and anti-quarks arise from gluon splitting and thus not only up and down flavors are found in protons but also heavier quarks up to the bottom quark. The determination and use of PDFs allows to write the cross section for each parton separately in case the kinetic energy of the protons is much larger than the proton rest energy of about 1 GeV:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x\mathrm{d}\hat{y}} = \frac{\pi\alpha^2}{sx^2\hat{y}^2}\left[\sum_q (x f_q(x, Q^2) + x f_{\hat{q}}(x, Q^2)\right] \cdot \left[A_q(h, Q^2) + (1 - y)^2 B_q(h, Q^2)\right] \tag{1.16}$$

This partonic cross-section is the basic principle for all theoretical calculations of expected event rates at hadron colliders.

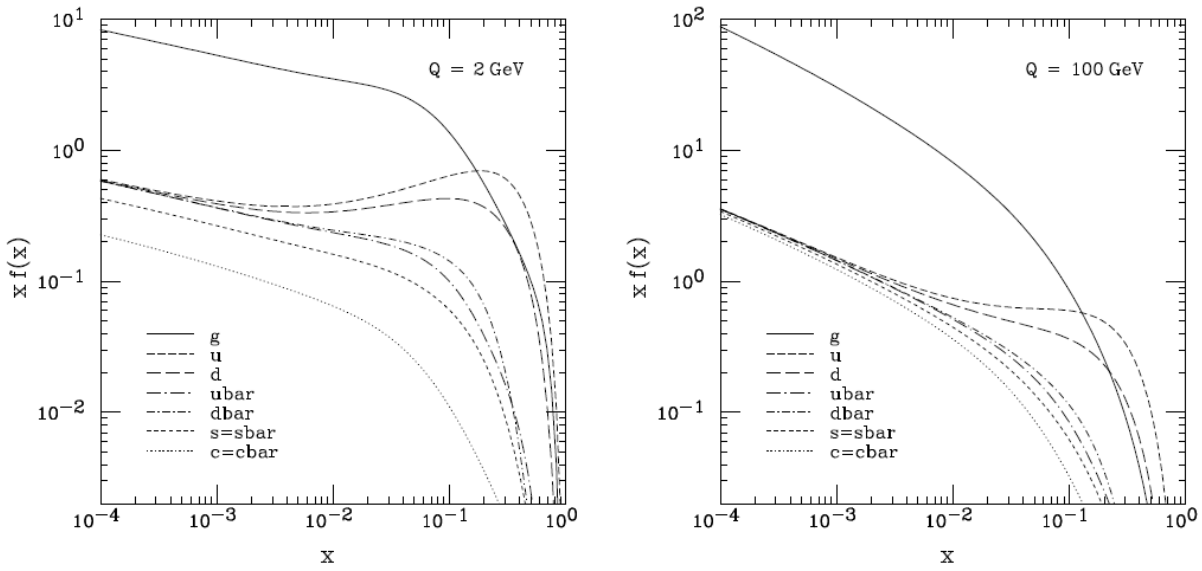


Figure 1.2: *CTEQ6M PDFs for Q=2 GeV and Q=100 GeV. Taken from [13]. It becomes obvious that a large fraction of the proton momentum is carried by low-x gluons.*

1.2 Standard Model Principles and Formalism

The Standard Model of Particle Physics is currently the best description of fundamental interactions and one of the most precisely tested theories in all of science. The following paragraphs are meant to be a brief introduction, with a much more detailed view provided by textbooks such as [14–16].

The Standard Model contains twelve elementary fermions, six quarks and six leptons, each carrying a spin of $\hbar/2$. They can carry different charges, that couple to fields of the three fundamental interactions, the electromagnetic, the weak and the strong interaction. These are the electrical charge, the weak isospin and the color charge. Gravity is not described by the Standard Model. For each of the particles, also a corresponding anti-particle exists, with opposite electric charge, anti-color and opposite weak isospin, yet an identical mass. The elementary fermions are usually arranged in three generations of isospin doublets, of which apart from the neutrinos, only the first generation is stable. All heavier quarks and fermions can only be produced and studied in high energy collisions in either cosmic rays or particle accelerators. Four vector bosons mediate the interactions. An overview of the forces, fermions and bosons is given in Tables 1.1, 1.2 and 1.3. Apart from the aforementioned particles, the Standard Model predicts the existence of an additional spin-0 boson, the Higgs boson, which is supposed to be responsible for giving the other particles their mass. The Higgs boson is the only yet undetected particle within the Standard Model.

Table 1.1: *The three forces of the Standard Model and gravity. The realtive strengths are evaluated at $Q = 1$ GeV. From [17]*

Interaction	Approximate Potential	Parameter Values	Relative Strength
Strong	$\frac{12\pi/23}{Q^2 ln(Q^2/\Lambda^2)}$	$\Lambda \approx 0.2$ GeV	1
Electro-Magnetic	$\frac{\alpha_{\mathrm{em}}}{Q^2}$	$\alpha_{\mathrm{em}} \approx 1/137$	1.4×10^{-2}
Weak	$\frac{\alpha_{\mathrm{em}}}{Q^2 - M_{\mathrm{W}}^2}$	$M_{\mathrm{W}} \approx 80$ GeV/c^2	2.2×10^{-6}
Gravity	$\frac{G_{\mathrm{N}} m_1 m_2}{Q^2}$	$G_{\mathrm{N}} \approx \frac{6.7 \times 10^{-39}}{\mathrm{GeV}^2}$	1.2×10^{-38}

Table 1.2: *The three generations of fermions. In each family the quarks and fermions are forming an isospin-doublet. The quark and fermion types are also referred to as flavor. The mass eigenstates of the quarks do not match with the eigenstates of the weak interaction, so the down-type quarks are notated with primes.*

	Particle			Charge	Color	Spin
Generation	1	2	3			
Quarks	$\begin{pmatrix} u \\ d' \end{pmatrix}$	$\begin{pmatrix} c \\ s' \end{pmatrix}$	$\begin{pmatrix} t \\ b' \end{pmatrix}$	$+2/3e$ $-1/3e$	r, g, b	$\frac{\hbar}{2}$
Leptons	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}$	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}$	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}$	$-e$ 0	-	$\frac{\hbar}{2}$

Table 1.3: *The four gauge bosons of the Standard Model. The world average of the masses as given by the Particle Data Group are found in [18]*

Particle	Charge	Spin	Mass	Interaction	Weak Isospin
Photon γ	0	$\hbar$	0	electro-magnetic	0
Gluon g	0	$\hbar$	0	strong	0
Z^0	0	$\hbar$	91.18 GeV	weak	0
$W^{\pm}$	± 1	$\hbar$	80.40 GeV	weak	± 1

1.2.1 Mathematical Methods

The Standard Model of Particle Physics is a quantum field theory and is thus based on the principles of gauge invariance and local symmetries. The fundamental particles are described as excited states of quantum fields that depend on space-time coordinates. The mathematical foundation is the description of kinematics with Lagrangian field theory in covariant notation.
The Lagrangian formalism starts from the problem of minimizing an action S. A given Lagrangian $\mathcal{L}(\phi_i, \partial_\mu \phi_i)$ is a function of the fields ϕ_i and their first derivatives $\partial_\mu \phi_i$. The action is a function of $\mathcal{L}$:

$$S = \int d^4x \mathcal{L}(\phi_i, \partial_\mu \phi_i) \tag{1.17}$$

The equation of motion can now be written in the Euler-Lagrangian notation:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0 \tag{1.18}$$

Using the Lagrangian formalism as a starting point and including the laws of relativistic quantum mechanics, both the Klein-Gordon equation that describes the dynamics of spin-0 fields (Eq. 1.19) and the Dirac equation for spin-$1/2$ fields (Eq. 1.20) can be derived:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi)(\partial^\mu \Phi) - \frac{1}{2} m^2 \Phi^2 \implies (\partial^\mu \partial_\mu + m^2)\Phi(x) = 0 \tag{1.19}$$

$$\mathcal{L} = \bar{\phi}(i\gamma^\mu \partial_\mu - m)\phi \implies i\partial_\mu \bar{\phi}\gamma^\mu + m\bar{\phi} = 0 \tag{1.20}$$

One of the fundamental principles of theoretical physics, Noether's theorem [19], states that every conservation law is connected to a symmetry. It is therefore convenient to use a model that reflects conserved quantities through symmetries. These symmetry operations do not change the solution of the equation of motion and can be written as:

$$\phi(x) \to \phi'(x) = U\phi(x) \tag{1.21}$$

Though Noether's theorem applies to both global and local symmetries, the Standard Model is based on local ones, that are symmetric under compact Lie groups. Local gauge symmetries in the Standard Model are connected to bosonic gauge fields [15].

1.2.2 Particles and Forces

The first successful formulation of a relativistic gauge theory was quantum electrodynamics (QED) [20]. It required implementing a local U(1) symmetry into the Dirac theory which can naturally explain the electromagnetic interaction of charged particles through photon exchange. In the 1960s, Glashow, Salam and Weinberg combined the electromagnetic force with the weak force, that was

known mainly from radioactive β decay. It is represented by a spontaneously broken $\mathrm{SU(2)}_L \times \mathrm{U(1)}$ symmetry. One particularity of the weak interaction is that it only acts on left-handed fermions and right-handed anti-fermions, which is pointed out by the index L to the SU(2) symmetry. Additionally, the interaction of color charged particles is described by an unbroken SU(3) symmetry, which will be described in more detail in section 1.3.
From group theory it is known, that a SU(2) group has three generators, called W^0, W^1 and W^2, while the U(1) has one, the B^0. The broken symmetry causes the W^0 and B^0 to form two mixing states that can be identified with the Z^0 and the γ and can mathematically be expressed using the weak mixing angle Θ_W:

$$\begin{pmatrix} \gamma \\ Z^0 \end{pmatrix} = \begin{pmatrix} \cos\Theta_\mathrm{W} & \sin\Theta_\mathrm{W} \\ -\sin\Theta_\mathrm{W} & \cos\Theta_\mathrm{W} \end{pmatrix} \begin{pmatrix} B^0 \\ W^0 \end{pmatrix} \tag{1.22}$$

This mixing angle also describes the different masses for the Z and W bosons, which follow the relation

$$M_\mathrm{Z} = \frac{M_\mathrm{W}}{\cos\Theta_\mathrm{W}}. \tag{1.23}$$

The same effect that breaks the electro-weak symmetry also gives the W and Z bosons their mass. This is attributed to the Higgs mechanism, which is a scalar field with a vacuum expectation value that couples to all massive particles with a coupling strength proportional to the particle's mass.
In processes of the electromagnetic and strong interactions, particles do not change their flavor. Only the weak interaction mediates flavor transitions, however lepton family conservation still remains valid. The eigenstates of the weak interaction are not identical to the mass eigenstates of the quarks and the transitions between them are described by the CKM[3] matrix, which strongly favors intra-generation transitions and can also can also describe the observed CP violation in processes of the weak interaction. A number of phenomena in high energy physics are related to the properties of the CKM matrix, for example the relative longevity of hadrons containing b-quarks, that makes them detectable in collider experiments.

1.2.3 Free Parameters

Apart from the non-inclusion of gravity in the Standard Model and other unsolved cosmological questions such as Dark Matter [21] and Dark Energy, little evidence has been gathered that points to physics phenomena that cannot be described with the Standard Model. Another example is the fact, that neutrinos do in fact have a finite mass, which manifests itself in neutrino oscillations. [18]. It is however unpleasant that a lot of parameters that are required in the SM cannot be predicted but have to be determined from measurements. These are in detail:

[3] Cabibbo-Kobayashi-Maskawa

- The masses of the six quarks and three massive leptons. Neutrinos are massless in the SM, an extension that can deal with massive neutrinos requires additional masses and mixing angles.
- The mass of the Higgs boson.
- The three gauge couplings of the electroweak and strong interactions.
- The CKM matrix elements.
- The weak mixing angle.

Furthermore, the Standard Model has a number of not understood features that can not be explained yet. One example is the hierarchy of fermion masses, that ranges from 511 keV for the electron to 173 GeV for the top quark, a range of almost six orders of magnitude. There are a number of proposed extensions to the SM like supersymmetry (SUSY), that provide an explanation for a number of problems such as a natural candidate for Dark Matter by adding a super-partner to each of the SM particles. However, no evidence for SUSY has been found yet. Multiple suggestions exist how supersymmetry is actually realized, with the Minimal Supersymmetric Standard Model (MSSM) being the most popular one. It is predicted, that if the MSSM is true, the lightest supersymmetric particle is stable and neutral and the LHC will be able to produce and detect it.

1.3 Quantum Chromodynamics

Quantum chromodynamics is the theory that describes the strong interaction which is responsible for both forming hadrons from quarks and gluons and for the nuclear force that binds protons and neutrons to nuclei. When quarks and gluons were discovered it became obvious that a new theory was required to describe the very strong cohesion of partons and the absence of free quarks. In order to structurize the many newly discovered hadrons, Murray Gell-Mann suggested in 1964 to order the spin-$1/2$ mesons and baryons into octets and the spin-$3/2$ baryons into a decuplet. However some of the particles of this decuplet had not been observed yet.

The eventual discovery of the Δ^{++} [18] resonance that consists of three up-quarks with a spin of $3/2$ also demanded the introduction of a new quantum number as with the present ones, no anti-symmetric wave function could be formulated for three identical quarks with parallel spins. This was however necessary to fulfill Pauli's exclusion principle and so the new quantum number had to offer at least three possible states. In accordance to the additive color model, these states were denoted as "red", "green" and "blue", a combination of which results in a color-neutral or "white" state. Also to each color charge there is a corresponding anti-color, the three of which can also form color-neutral triplets. In principle, more than three colors would also be possible, measurements of decay widths that depend on the number of colors [22] as a degree of freedom

Table 1.4: *The spin-$1/2$ baryon octuplet and the spin-$3/2$ baryon decuplet. s denotes the strangeness quantum number of the baryon, defined as the number of anti-strange quarks contained. Strange quarks contribute one negative unit of strangeness.*

s	Spin-$1/2$ Baryon Octet			Quark Content		
0	n	p		ddu	duu	
-1	Σ^-	Σ^0 / Λ	Σ^+	dds	dus	uus
-2	Ξ^-	Ξ^0		dss	uss	

s	Spin-$3/2$ Baryon Decuplet				Quark Content			
0	Δ^-	Δ^0	Δ^+	Δ^{++}	ddd	ddu	duu	uuu
-1	Σ^{*-}	Σ^{*0}	Σ^{*+}		dds	dus	uus	
-2	Ξ^{*-}	Ξ^{*0}			dss	uss		
-3	Ω^-				sss			

have shown that in fact, extra colors are excluded.
Two different configurations of color neutral composite objects have been observed: Particles, that form a colorless state from three quarks or three anti-quarks are called baryons and anti-baryons respectively. The proton and the neutron, as long as bound in a nucleus, can be stable. Free protons are also stable, free neutrons decay with a half-life of roughly 15 minutes into a proton, an electron and an anti-electron-neutrino. Mesons on the other hand consist of one quark and one anti-quark, and none of them is stable. Theoretically, bound states of for example four quarks and one anti-quark are possible as well, yet they have not been observed so far.
Even though gluons are massless, the strong interaction has a very short range, basically the size of a proton. This is due to the fact that gluons carry color charge themselves and therefore interact with each other. QCD is thus a non-Abelian gauge theory, described by SU(3) symmetry. This group has eight generators, which form an octet of color-charged gluons and a color-neutral singlet, which is not observed. A possible representation is the following:

Table 1.5: *Color representation of quarks.*

Singlet	$\sqrt{1/3}\,(r\bar{r}+g\bar{g}+b\bar{b})$							
Octet	$r\bar{g}$	$r\bar{b}$	$g\bar{b}$	$g\bar{r}$	$b\bar{r}$	$b\bar{g}$	$\sqrt{1/2}\,(r\bar{r}-g\bar{g}$	$\sqrt{1/6}\,r\bar{r}+g\bar{g}-2b\bar{b}$

1.3.1 The QCD Lagrangian and the Running Coupling

A theory of the properties of quarks and their interaction has to be able to describe and contain the following features that have just been discussed [17]:

- Hadrons are composed of quarks with fractional charges.
- Quarks are spin-$1/2$ fermions.
- They carry one of three possible color charges.
- There is evidence that color charges exhibit an SU(3) symmetry.
- Quarks are subject to a strong interaction.
- Besides quarks, additional partons are contained in hadrons.
- Those partons do not interact via the electro-magnetic or weak force.

The mathematical description of quantum chromodynamics relies on similar principles as the electro-weak theory. The QCD Lagrangian can be written with implied summation over repeated indices as

$$\mathcal{L}_{\text{QCD}} = \sum_q \bar{q}_i (i\gamma^\mu D_\mu - m_q)_{ij} q_j - \frac{1}{4} F^a_{\mu\nu} F^{a\ \mu\nu} \tag{1.24}$$

The field strength tensor and the covariant derivatives read as:

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - g f^{abc} A^b_\mu A^c_\nu \tag{1.25}$$

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu + i g_s T^a_{ij} A^a_\mu \tag{1.26}$$

$$(m_q)_{ij} = m_q \delta_{ij} \tag{1.27}$$

In these equations, A^a_μ are the gluon fields, g_s the gauge coupling, f^{abc} the structure constant and T^a_{ij} the generators of the Lie group. From the expression for the gluon field and the completely anti-symmetric structure constant f^{abc} the non-Abelian gluon-gluon interactions become calculable. The quark-gluon coupling is proportional to the strong gauge coupling g_s which translates to the strong coupling constant as $\alpha_s = \frac{g_s^2}{4\pi}$. The only other free parameters in the Lagrangian are the fermion masses. Taking into account the SU(3) nature of QCD, the eight generators of the symmetry transformation can be expressed using the Gell-Mann matrices that can be found elsewhere [17]. From the couplings it becomes apparent that only three kinds of basic QCD processes are observable along with the propagators of free quarks and gluons: Gluon radiation from a quark and three as well as four gluon vertices. The Feynman graphs of these processes are shown in Fig. 1.3.

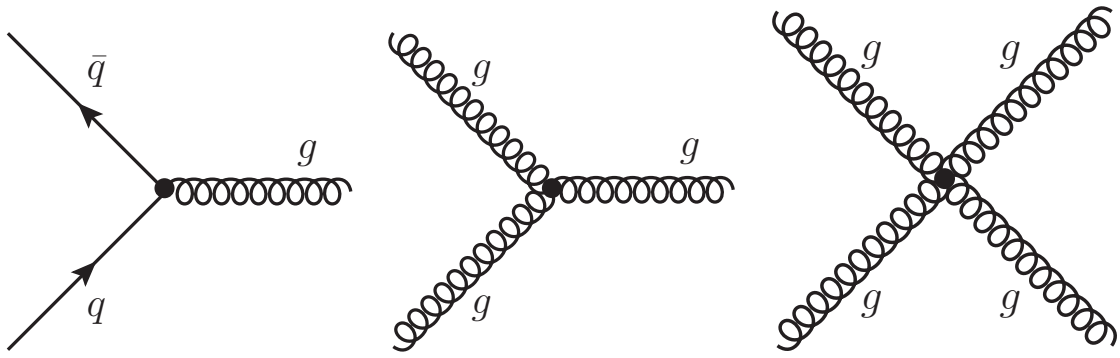


Figure 1.3: *Feynman graphs of the basic QCD interactions*

Gluon self-coupling accounts for potential energy build-up when quarks are separated. Between the quark pair, a gluon string is formed, which breaks apart once enough energy is stored to create a new quark-antiquark pair at the rupture. This explains the absence of free colored particles in nature. The break-up happens at distances of about 1 fm, which is about the size of a proton, explaining the short range of the strong interaction despite its massless force carriers. The observation that rather than extracting partons from a proton, one destroys the proton in scattering experiments

and creates new particles, is called *color confinement*. On the other hand, when the momentum transfer Q^2 is large enough in collider experiments, quarks can be assumed as quasi-free particles, a principle that is called *asymptotic freedom* [23] and is successfully applied in theory calculations, since it allows the application of perturbative techniques.
In contrast to electromagnetic interactions, where screening effects lead to an increasing coupling for small distances and growing energies, asymptotic freedom can be interpreted as *anti-screening*. In the low energy regime, the coupling diverges which makes it impossible to calculate low-Q^2 QCD in the mathematical framework of perturbation theory. In order to deal with divergent terms that arise from gluon self-coupling, renormalization techniques have to be applied, that absorb the problematic terms. It is however necessary to restore gauge invariance by rescaling all involved fields.
Hard QCD interactions, that correspond to large momentum exchange, however happen in an energy region where the strong coupling is small enough to allow perturbative calculations. As mentioned in the previous chapter, α_s is one of the free parameters of the SM and apart from the fermion masses the only one having implications on quantum chromodynamics. Due to its divergence at low Q^2 and the running of the coupling, the definition of a numerical value of the strong coupling constant is always coupled to the corresponding Q. One of the most precise measurements of α_s has been performed at the LEP experiments [24], where the rate of three jet events over two jet events is connected to the probability of hard gluon radiation and thus proportional to the strong coupling at $M(Z^0) = 91.2$ GeV. The current world averages from different measurements can be found in Fig. 1.4 as well as the actual $\alpha_s(Q)$ relation.
The mathematical description of the energy dependence of the strong coupling is a renormalization group equation of the form:

$$\frac{\partial \ln \alpha_s(Q^2)}{\partial \ln Q^2} = \frac{\beta(\alpha_s(Q^2))}{\alpha_s(Q^2)} \tag{1.28}$$

This relation can be re-written as

$$\beta(\alpha_s(Q^2)) = Q^2 \frac{\partial \alpha_s(Q^2)}{\partial Q^2} \tag{1.29}$$

thus defining the β function, which can in turn be expanded in α_s:

$$\beta(\alpha_s(Q^2)) = -\frac{\beta_0}{4\pi}\alpha_s^2(Q^2) - \frac{\beta_1}{8\pi^2}\alpha_s^3(Q^2) + \mathcal{O}(\alpha_s^4(Q^2)) \tag{1.30}$$

The coefficient β_0 depends on the number of quarks flavors n_f that can be produced at the chosen energy scale Q and can be calculated from one-loop QCD theory as

$$\beta_0 = 11 - \frac{2}{3}n_\mathrm{f}. \tag{1.31}$$

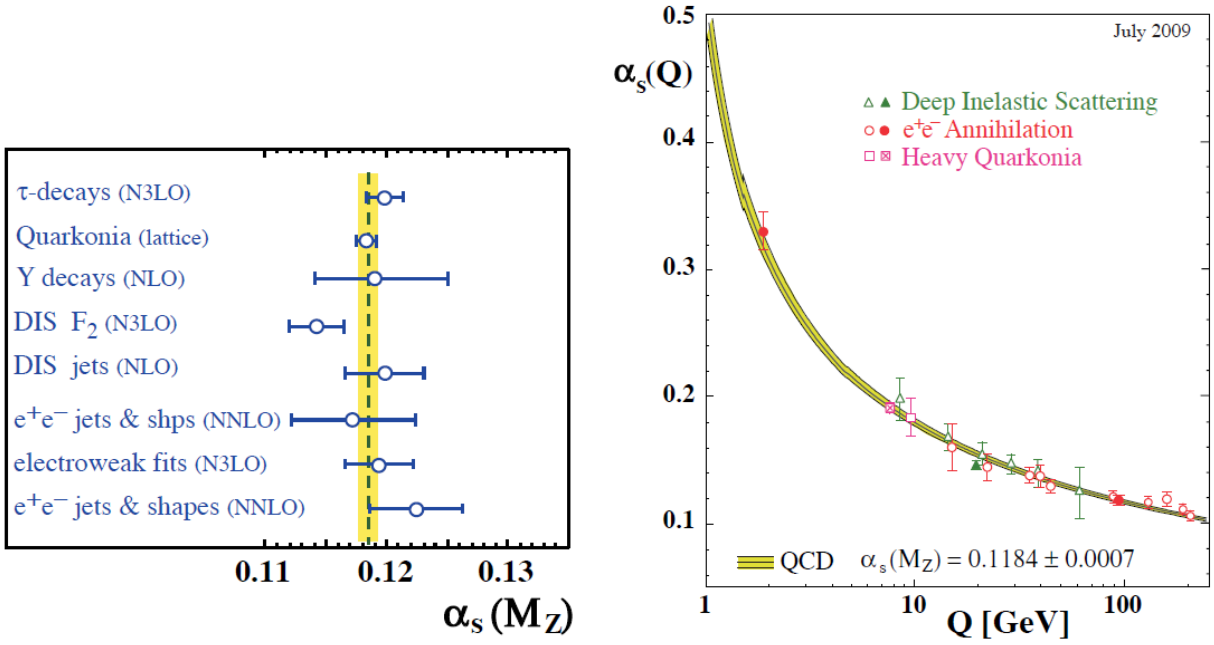


Figure 1.4: *World average of α_s at $M(Z^0)$ and $\alpha_s(Q)$ relation. From [25]*

Now, the leading order solution can be calculated:

$$\alpha_s(Q^2) = \frac{\alpha_s(\mu^2)}{1 + \alpha_s(\mu^2)\frac{\beta_0}{4\pi}\ln\frac{Q^2}{\mu^2}} \tag{1.32}$$

Here, μ is the energy value, at which the renormalization was performed. Often, $M(Z^0)$ is used for this purpose.

1.3.2 Hard QCD Scattering and Factorization

In order to calculate QCD cross-sections, a factorization is necessary so that the hadronic cross-section can be written as a convolution of the partonic cross-section and the parton distribution functions, that have been introduced in section 1.1.3. A relatively easy example of a full calculation including hadrons is the leading-order Drell-Yan production of a muon pair from a quark and an anti-quark via an off-shell γ or Z^0. The Feynman graph for this process is shown in Fig. 1.5.

The cross-section can be calculated in the limit of asymptotic freedom under the assumption that the center-of-mass energy of the hadron collision is growing to infinity. A factorization and integration over the *Bjorken-x* of the annihilating quarks yields:

$$\sigma_{AB} = \iint f_{q,A}(x_q) f_{q,B}(x_q) \hat{\sigma}(q\bar{q} \to \mu^-\mu^+) dx_q dx_q \tag{1.33}$$

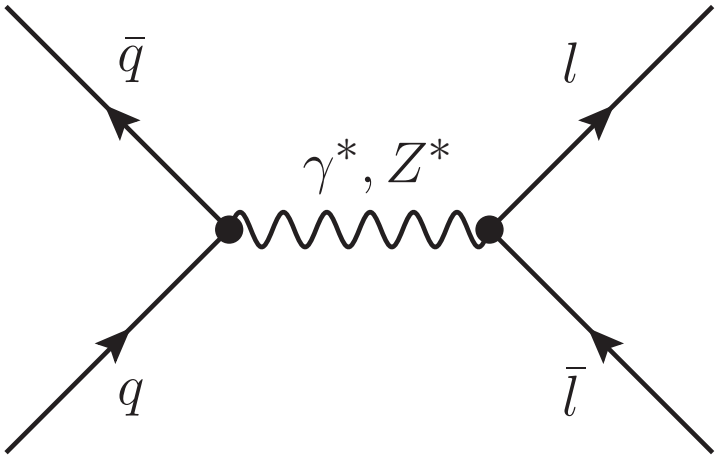


Figure 1.5: *Drell-Yan process* $q\bar{q} \to l\bar{l}$

Large logarithmic perturbative corrections are expected from collinear gluon radiation. These do however also appear in the calculation of the structure functions W_1 and W_2 in deep inelastic scattering (see section 1.1.2). So it is convenient to absorb them already in the definitions of the parton density functions, which leads to logarithmic scaling violation. If one at this point chooses a suitably large momentum scale Q^2, for example the invariant mass of the muon pair, the cross-section including logarithmic correction can be factorized as:

$$\sigma_{AB} = \iint f_{q,A}(x_q, Q^2) f_{q,B}(x_q, Q^2) \hat{\sigma}(q\bar{q} \to \mu^- \mu^+) dx_q dx_q \tag{1.34}$$

The partonic cross-section can be expanded in orders of the coupling constant:

$$\sigma_{AB} = \iint f_{q,A}(x_q, Q^2) f_{q,B}(x_q, Q^2) \left(\hat{\sigma}_0 + \alpha_s(\mu^2)\hat{\sigma}_1 + \dots\right)_{q\bar{q}\to\mu^-\mu^+} dx_q dx_q \tag{1.35}$$

Two different scales appear in this formula: The *factorization scale* that separates the hard and soft component of the interaction and the *renormalization scale* that has already been introduced in the previous chapter. Unless all orders are calculated, the choice of these scales influences the resulting cross-section. Equal values can be chosen for both scales and usually they are picked in the order of the momentum exchange in the process: In the case of Drell-Yan muon production, the invariant mass of the muon pair is the typical choice. In case of $\log \mu \gg \log 1/x$ the DGLAP[4] equations [25] describe the evolution of the PDFs with Q^2, with the help of *splitting functions* that describe the prapability of a parton radiating a gluon. After radiation, the parton carries the momentum fraction z of the original parton. The DGLAP equations read for quarks and gluons

[4]Dokshitzer-Gribov-Lipatov-Altarelli-Parisi

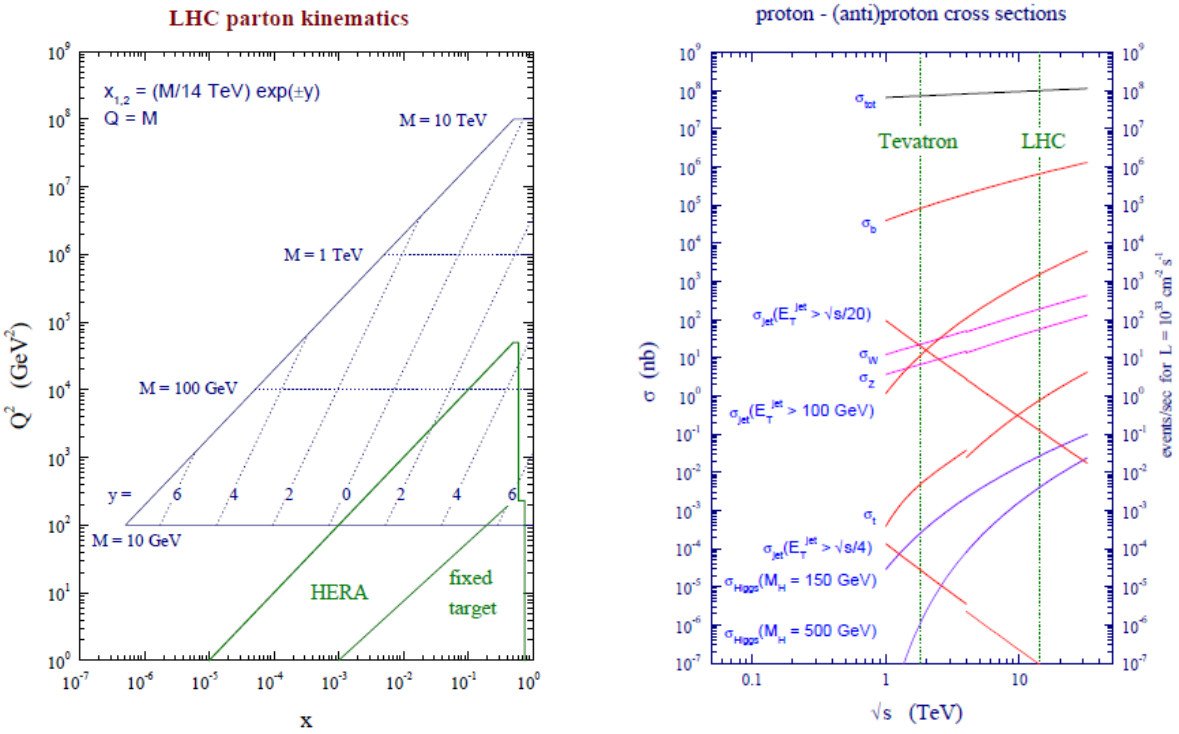


Figure 1.6: *Parton kinematics and cross-sections at Tevatron and the LHC. Taken from [26].*

respectively:

$$\frac{\partial q_i(x,\mu^2)}{\partial \log \mu^2} = \frac{\alpha_s}{2\pi}\int_x^1 \left(P_{q_i q_j}(z,\alpha_s) q_j(\frac{x}{z},\mu^2) + P_{q_i g}(z,\alpha_s) g(\frac{x}{z},\mu^2)\right)\frac{dz}{z} \tag{1.36}$$

$$\frac{\partial q_i(x,\mu^2)}{\partial \log \mu^2} = \frac{\alpha_s}{2\pi}\int_x^1 \left(P_{g q_j}(z,\alpha_s) q_j(\frac{x}{z},\mu^2) + P_{gg}(z,\alpha_s) g(\frac{x}{z},\mu^2)\right)\frac{dz}{z} \tag{1.37}$$

Using these principles, hadron collision cross-sections have been calculated in higher orders of perturbation theory. Results of cross-section calculations at nominal LHC center-of-mass energy of 14 TeV as well as parton dynamics in the $(Q^2 - x)$ plane can be found in Fig. 1.6.

1.4 Monte Carlo Methods

Testing the predictions of the Standard Model and other theories in particle physics against experimental measurements requires predictions of how the theoretical models will manifest themselves in the experimental setup. Therefore, quantum mechanical equations have to be solved that involve probabilistic distributions and the corresponding high dimensional integrals. For these tasks, Monte Carlo methods are especially suited, in which complex integrations are numerically per-

formed with the help of random number generation. In several steps, starting from the evaluation of an interaction matrix, subjecting the outgoing partons to a showering algorithm and finally forming stable hadrons, a complete Monte Carlo event is generated, many of which have to be investigated to study the typical final state distributions of the particular process.

1.4.1 Matrix-Element Calculation

The first step in the generation of a Monte Carlo event is the evaluation of an interaction matrix element. This requires numerically integrating the squared matrix element over the available phase space. This can either be done in leading (LO) or next-to-leading (NLO) and even higher order. As has been pointed out above, leading order, and to a lesser extend also higher order calculations are dependent on factorization and normalization scales and large logarithms have to be re-summed. With their higher precision and reduced scale uncertainties, NLO calculations are a vital contribution for precise predictions in particle physics.
These calculations contain all Feynman diagrams with one additional vertex compared to tree level diagrams. In the case of QCD processes, the coupling is α_s. This leads to additional quarks and/or gluons in the diagram. Depending whether these additional partons appear as external or internal lines, they act as *real* or *virtual* contributions respectively. Virtual contributions require more than one additional vertex and interfere with the tree level diagrams while real contributions allow to access additional phase space regions. For example, the Drell-Yan process in the previous section calculated at tree-level allows no transverse momentum for the virtual boson. Only a balancing parton emission in the opposite direction allows to introduce this.

1.4.2 Parton Shower

Typical collision events in modern hadron colliders show particle multiplicities in the order of a few hundred or even more. Even though these stable particles can in most cases be traced back to only a small number of hard partons, it is still obvious that perturbative matrix element calculations have their limitations as the number of Feynman diagrams contributing to a process grows with the factorial of the number of final state partons.
The common way to get results that are valid in all orders is to apply a parton shower algorithm. The few partons that are produced in the hard interaction are evolved into numerous partons at the QCD factorization scale using the DGLAP formalism. Probabilities for partons to evolve from high scales Q_0 to low scales Q_1 without gluon radiation is expressed by Sudakov Form Factors [25]:

$$\Delta(Q_0, Q_1) \propto \exp(-G_{12}\alpha_s \ln^2(Q_0/Q_1) + \dots) \tag{1.38}$$

With the help of a random number that solves the equation, Q_1 of the first emission is determined. This procedure is repeated until a fixed value in the order of 1 GeV is reached, where the parton

shower stops and the non-perturbative part of the event generation begins. The evolution parameter itself can for example be transverse momentum in case of p_{T}-ordered showers or the radiation angle in case of virtuality ordered showers.
While the parton shower mechanism on the one hand is able to give a very good approximation of the phase space dominated by collinear gluon emission, it has shortcomings in accounting for high energetic radiation and large angles. Matrix element calculations on the other hand can populate these regions of the phase space yet are not able to handle soft radiation. Thus an optimal solution would be a combination of both methods. However in the intermediate region of the phase-space, the problem of double counting arises. Remedies for this problem are given by the CKKW [27] and the MLM [28] methods. Here, the phase space is divided in a region which is covered by matrix element calculations and a region that is covered by the parton shower. The transition between the two is made at a certain virtuality or energy.

1.4.3 Hadronization

The transition from the partonic final state after the parton shower algorithm to stable hadrons cannot be calculated with perturbation theory as in this energy regime, perturbative calculations do not converge. Therefore, a phenomenological model is necessary that builds color-neutral objects from the partonic configuration of the event. Basically, two different approaches are available in Monte Carlo event generators: The Lund-string model [29] uses strings between the quarks that break up and form color-neutral hadrons.
The cluster hadronization model [30] first splits all gluons into quarks and anti-quarks and builds colorless clusters from them which are either split into lighter clusters or interpreted as hadrons. Both models have to be tuned to data in order to correctly reproduce both the particle number and flavor content distributions observed in collider data. Especially the recent re-evaluation of hadronic final states in LEP data [31] has proven to be a rich source of improvement for the hadronization models due to the clean signals.

1.4.4 The Underlying Event and Multiple Parton Interactions

Compared to electron-positron collisions, the initial state in a proton collider is far more complicated due to the compositeness of the proton. One problem that has to be addressed is the fact that the remnants of the protons after parton extraction through hard scattering are instable color-charged objects that radiate gluons. Furthermore, due to the increase in the gluon PDF, that can be observed in Fig. 1.2, the cross-section of additional low-x scatters inside two colliding protons is rapidly rising with the center-of-mass energy of the collision. A sketch of the decomposition of a proton-proton collision is given in Fig. 1.7. Popular approaches to describe the underlying event are always heavily relying on tunable models. Numerous parameters steer the phase space proper-

ties of additional activity in the events. It has to be noted however that the physical definition of

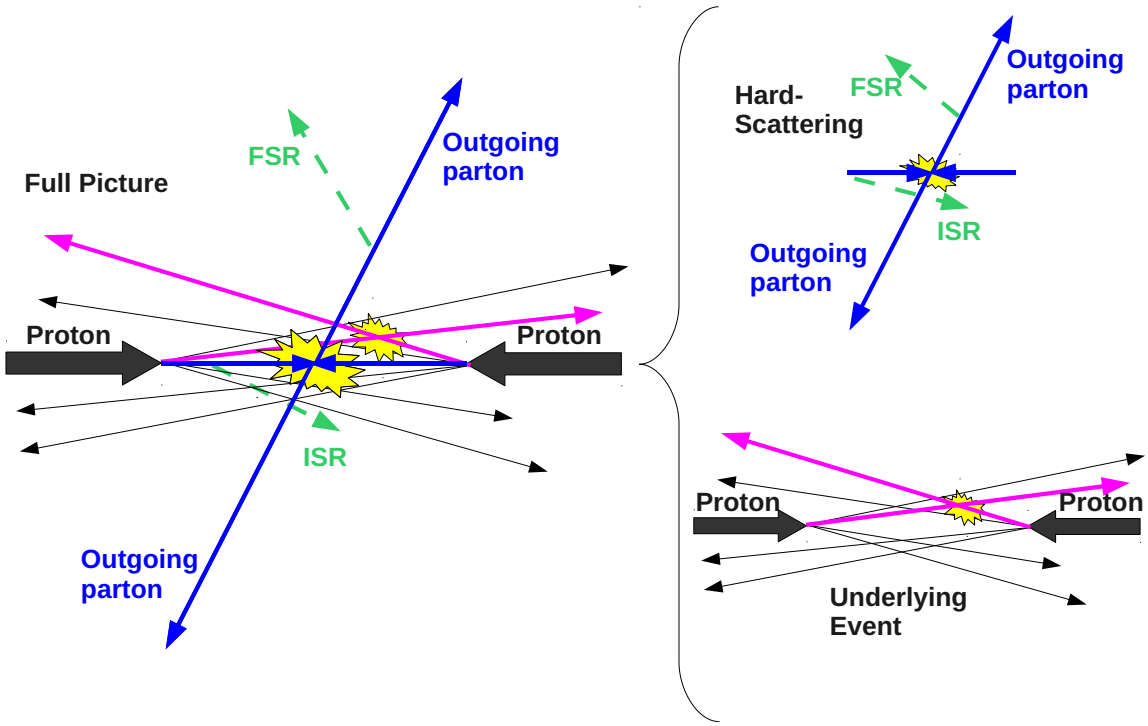


Figure 1.7: *Decomposition of a proton-proton collision. The interaction can be split into a hard scattering of one parton from each proton, that may have emitted initial state radiation and the underlying event. This contains multiple parton interactions and the handling of color-charged beam remnants. From [32].*

the underlying event is more problematic than the treatment in Monte Carlo generation. Chapter 3 is dedicated to this discussion.

The concept of multiple parton interactions (MPI) is based on additional low momentum transfer scatters inside the colliding protons. In a most basic picture, these scatters can be completely independent from the hard parton interaction in the collision and are treated in Monte Carlo event generation as isolated minimum bias QCD parton interactions the products of which undergo parton shower and hadronization the same way as the outgoing partons of the hard interaction. In more sophisticated models, the color-connections of the hard and weakly interacting partons are also considered. One important parameter of MPI models is the lower transverse momentum transfer cutoff, only above which additional scatters are evaluated. This value, together with the gluon PDF at small x determines the average number and distribution of MPI. Also, a scaling factor for the behaviour of this cutoff value depending on the center-of-mass energy is introduced. Further parameters include the mass distribution inside the protons.

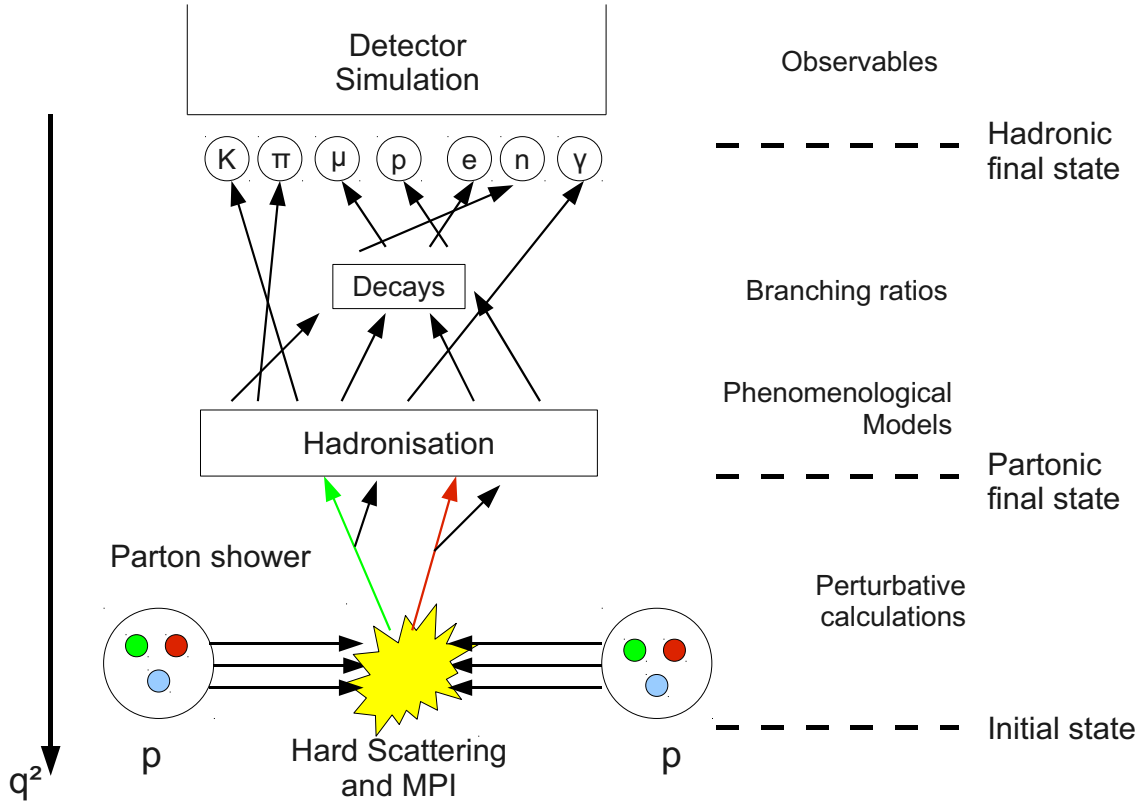


Figure 1.8: *Working principle of a multi-purpose Monte Carlo event generator: Two partons from the incoming protons interact according to the hard matrix element. The outgoing hard partons together with the proton remnants are forwarded to the parton shower to form the partonic final state (PFS). These partons are then hadronized. Particles with a mean lifetime below a particular threshold also are decayed.*

1.4.5 Monte Carlo Event Generators

Multi-purpose Monte Carlo event generators are some of the most important tools in particle physics. They allow to make predictions on distributions expected at experiments employing state of the art theory results, thus being the link between the two fields. Two main packages are available for this task that will be described in the following. Additionally, a large number of specialized event generators exist that concentrate either on particular event topologies and physics channels or on the treatment of properties that are not treated in some general purpose generators like spin correlations of particles [33]. The working principle of a hadron level Monte Carlo generation is shown in Fig. 1.8. The different steps can be executed by multiple generators, rendering it possible to use for example different parton shower algorithms on the same matrix element calculation in order to estimate uncertainties of Monte Carlo predictions.

Pythia

The Pythia multi-purpose Monte Carlo event generator is the most widely used tool in high energy physics. It is currently available in both a Fortran and a C++ version [29, 34]. The Fortran version has been developed since 1977 and the most current version is Pythia 6.4.23. The latest C++ version is Pythia 8.142. Even though Pythia 8 is basically a re-implementation of the existing physics models and processes, it is not yet in a state that is recommended for large scale production and it also still lacks functionalities such as the generation of proton-electron collisions.
Pythia 6 covers an enormous range of physics processes. All Standard Model $2\rightarrow1$ and $2\rightarrow2$ matrix elements are covered including soft and hard QCD processes, heavy flavor production, vector boson and SM Higgs production. Furthermore, many non-Standard Model processes are included, such as supersymmetry and technicolor. In total, Pythia 6 contains more than 300 processes. Special emphasis has also been put on the simulation of soft QCD contributions to proton-proton collisions. As mentioned above, these processes cannot be calculated using perturbation theory but phenomenological models have to be applied to reproduce observed distributions. These models, in the case of Pythia 6 contain a large number of free parameters. Combinations of certain values for all of these parameters are called *tunes*. Detailed information on how these tunes are derived from data can be found in chapter 3 and a couple of examples are presented in appendix A. One of the main features of the Pythia package is its implementation of the Lund-String fragmentation model. Exhaustive discussions of both the program structure and the physics processes of Pythia 6 can be found in [34].
Even though the Fortran programming language still offers good performance and reliability, large scale programs are increasingly hard to maintain due to the non-object-oriented layout. Therefore, the main developing work has shifted to Pythia 8, which is expected to supersede its predecessor in the next years.

Herwig

An alternative to Pythia is the Herwig program which is as well available in both a Fortran [35] and a C++ [36] version. The Fortran implementation, with the most current one being version 6.5, does not natively contain a model for multiple parton interaction, this functionality can however be added through interfacing to the JIMMY [37] program. The C++ version does contain such a model, which in contrast to Pythia relies on only a small number of parameters to describe MPI. Another main difference to Pythia is the employed cluster hadronization model.

1.5 Jet Algorithms

As has been discussed above, color-charged partons that are produced in proton collisions do not appear in a particle detector due to color confinement. Furthermore, through gluon radiation and splitting, the outgoing parton can branch into a large number of stable final state particles. These collimated streams of particles, which can also contain contributions from other processes are referred to as *jets*. Even though these jets are collimated, it is not unambigous from which process a particular final state hadron originates, especially in the presence of the underlying event or even pile-up events. In order to extract information on the hard process, well defined physical objects are required. Jet clustering algorithms rely on geometrical or combinatorial measures that combine particles into jets, ideally originating from the same outgoing parton. They can operate on different levels of reconstructed detector data or Monte Carlo particles, since they only need four-vectors as input. In order to make reliable predictions, that are robust against soft QCD emissions, modern jet algorithms are required to be *collinear* and *infrared safe*. Collinear and infrared unsafe behaviour is shown in Fig. 1.9.

Infrared and collinear unsafe behaviour is typically a problem of cone based algorithms, in which particles in a circle in the $\eta - \phi$ plane around a seed particle are clustered together. It can happen, that an additional soft particle between two separated hard particles can lead to a merging of two jets into one, whereas without the soft radiation they would have been declared two separated jets. In order to circumvent this infrared unsafe behaviour, a threshold on the transverse momentum of the seeds can be introduced. This however can lead to collinear unsafe behaviour, as the splitting of a particle close to the seed threshold can lead to the non-detection of a jet due to collinear splitting. These ambiguities are naturally avoided when no seeds are used in the process of jet clustering.

1.5.1 Cone-Type Algorithms

Cone-type jet algorithms [38] rely on a geometrical interpretation of combining particles to jets. This is an intuitive approach and offers the advantage of fixed geometrical shapes and areas of the jets, that allow for easy area based corrections for residual effects such as pile-up. The opening angle of the cone and therefore the radius of the projection circle is the main parameter of these algorithms. However apart from the problem of infrared and collinear safety, cone type algorithms also suffer from a fundamental problem when handling jets closer than twice the chosen jet radius, an event topology feature that is present in most events at high energy hadron colliders. In that case, an overlap between the jet cones is present, which has to be dealt with in a consistent manner.
Several different implementations of cone-type jet algorithms are available, only one of which, Iterative Cone, is still widely used in the CMS collaboration. The SISCone algorithm was used previously and is the only cone-type algorithm that by design is collinear and infrared safe, how-

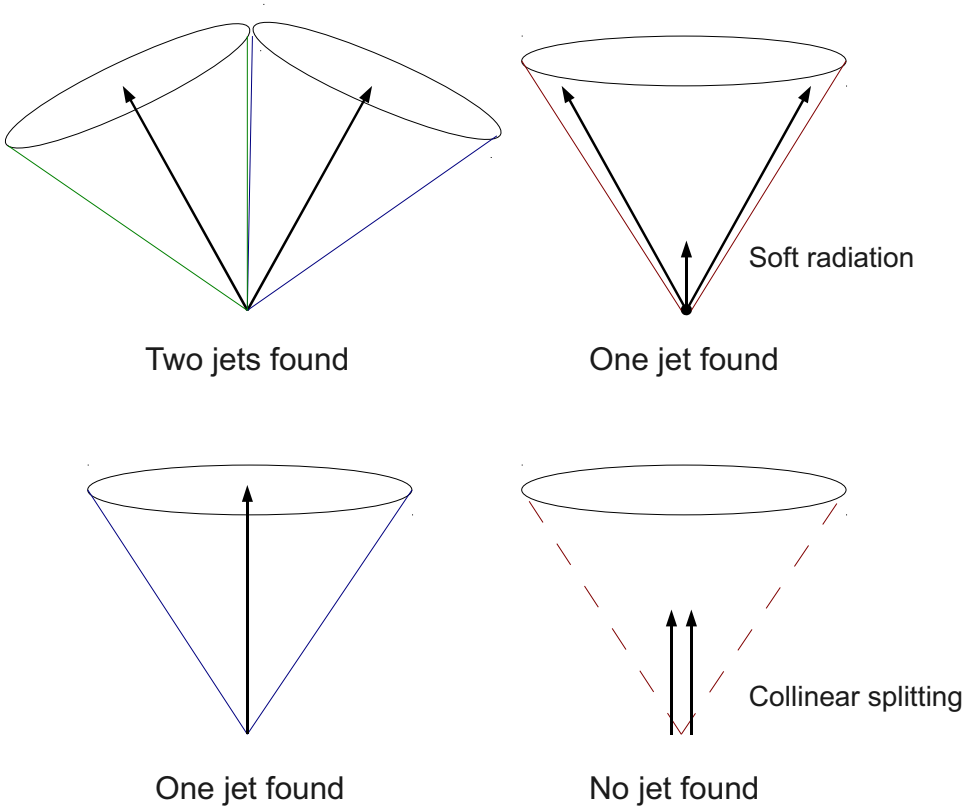


Figure 1.9: *Infrared (upper) and collinear (lower) unsafe behaviour of a jet algorithm. In the infrared case, a soft emission between two stable cones may lead to a change of the algorithm output, in the collinear unsafe case, a collinear splitting can lead to the rejection of the particle as a seed and the non-consideration of the jet altogether.*

ever its computing time does scale very poorly with the number of input particles and the algorithm is therefore not practical in events with a high particle multiplicity.

Iterative Cone

The Iterative Cone algorithm is a fast, yet infrared and collinear unsafe algorithm that is mainly used for triggering purposes. It starts from the highest energetic particle in the event and adds up the four-vectors of all particles inside the chosen jet radius. In the next step, a new circle is drawn around the axis of the four-vector sum of all particles clustered in the previous step. This is repeated until the direction of the proto-jet is no longer changing and the jet is declared stable. The objects belonging to this jet are removed from the list of input particles and the procedure is repeated until the list is empty. A modification of the Iterative Cone is the MidPoint Cone algorithm, that was for example used by the CDF experiment [39]. Here, additional seeds are used that lie in between

stable jets in order to reduce the sensitivity to infrared radiation. As this procedure however does not eliminate the problem of infrared safety in higher orders of perturbative QCD, the MidPoint algorithm has been dismissed by the LHC experiments.

SISCone

The Seedless Infrared-Safe Cone (SISCone) algorithm [40] is the only available cone-type jet algorithm that is infrared and collinear safe in all orders of perturbation theory. This is attained by not using single particles as seeds but pairs of them with a distance smaller than twice the chosen jet radius. The two circles that contain those two particles on their outline are calculated and the transverse momentum content of these circles is determined. Each circle is then pivoted around one of the particles until another particle lies on the outline. This way, only configurations are considered that lie in local maxima of transverse jet momentum. When all stable cones are found, another step is performed in which the overlap of cones is either split between the two or the jets are merged. This step is called *split-merge* procedure and another parameter is introduced that steers the behaviour of the algorithm in such cases.
Even though theoretically well behaved, the SISCone algorithm has shown to be impractical due to its large demands of computing time in events with many input particles and its use has therefore been phased out in the LHC experiments.

1.5.2 Combinatorial Algorithms

The general idea behind combinatorial jet algorithms is not to use a fixed jet geometry but to cluster particles according to their orientation in four-vector space. This approach intrinsically leads to collinear and infrared safe behaviour, however as distances between all particles in an event have to be calculated at every step of the clustering, the original implementation scaled very badly in terms of computing time with an input particle dependency of $t_{\text{clus}}(N) \propto N^3$. It has been shown [41] that with the use of Voronoi diagrams, this can be reduced in events with high particle multiplicities to $t_{\text{clus}}(N) \propto N^2$ or even $t_{\text{clus}}(N) \propto N\ ln(N)$, which leads to an implementation [42], that actually even surpasses the Iterative Cone algorithm in performance.

The Inclusive k_T Algorithm

The inclusive k_T algorithm relies on the distance measures

$$d_{ij} = min(p^2_{\text{T},i}, p^2_{\text{T},j})\frac{\Delta R^2_{ij}}{R^2} \quad \text{with} \quad \Delta R^2_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{1.39}$$

$$d_{iB} = p^2_{\text{T},i} \tag{1.40}$$

where d_{ij} is the distance between two particles i and j in four-vector space and d_{iB} is the distance between i and the beam axis and R is the *resolution parameter* that determines the size of the jets. The jet clustering procedure works in several steps:

- Calculate all d_{ij} and d_{iB}.
- Find the minimum value of all d_{ij} and d_{iB}.
- If the minimum value is a d_{ij}, combine i and j to a single new object and return to the first step. Remove i and j from the list of inputs. Otherwise, if a d_{iB} is the smallest, put the object on the list of final state jets and remove it from the list of inputs altogether.
- Repeat until no particles are left.

Some features of the k_T algorithm have to be noted. Firstly, all particles in the event end up clustered in a jet, which can lead to jets with very small transverse momentum. In experiments, the energy resolution of the detector components becomes the limiting factor for k_T jets in terms of the lower transverse momentum boundary, while in Monte Carlo final state, it is convenient to only consider jets above a certain threshold in transverse momentum. Furthermore, jets containing only one particle contain more information than for cone-type algorithms, as the neighboring jets can influence the shape and size of any k_T jet. From formula 1.39, it is obvious that the k_T algorithm clusters the soft diffuse particles first before these "clouds" of soft contributions are collated to a hard particle. This makes this algorithm especially suited for studies of soft QCD. For many analyses however, a fixed jet shape is desirable which the k_T algorithm does not provide.

The anti-k_T Algorithm

Technically, the anti-k_T algorithm [43] works very similarly to the inclusive k_T algorithm, however it uses a distance measure of

$$d_{ij} = min(p_{T,i}^{-2}, p_{T,j}^{-2}) \frac{\Delta R_{ij}^2}{R^2} \quad \text{with} \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{1.41}$$

$$d_{iB} = p_{T,i}^{-2} \tag{1.42}$$

which leads to a different clustering behaviour. In contrast to the k_T algorithm, the clustering starts with soft particles being attached to close-by hard particles. Through this hierarchy, nearly perfectly circular jets are clustered in case they are well isolated. If two jets are found near each other with substantial difference in their transverse momenta, the harder jet will be a circle, while the softer one will be crescent-shaped. In case they have comparable transverse momenta, they can either be divided along a straight line or merged, depending on the distance of their axes. An example of this behaviour can be seen in Fig. 1.10. This fixed geometry of its jets, together with

the intrinsic infrared end collinear safety and good scaling behaviour in high multiplicity events makes the anti-k_T algorithm a valid choice to replace problematic cone-type jet algorithms.

Generalized k_T Algorithm

As can be seen in equations 1.39 and 1.41, the exponent of the transverse momentum in the distance measure of a combinatorial jet algorithm defines its general behaviour. In a more general form, the equation can be written as

$$d_{ij} = min(p_{T,i}^{2p}, p_{T,j}^{2p})\frac{\Delta R_{ij}^2}{R^2} \quad \text{with} \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{1.43}$$

$$d_{iB} = p_{T,i}^{-2} \tag{1.44}$$

with p being 1 in case of the inclusive k_T and -1 in case of the anti-k_T algorithm. The value of p does not influence the infrared and collinear safety of the algorithms, so the k_T-type jet algorithms can be expressed in a more general form as in equation 1.43. Apart from the two already discussed cases, also a p value of 0 is commonly used, which translates to a purely geometrical distance measure, since the transverse momentum of the particles is not considered. This algorithm is called *Cambridge/Aachen* [44] and is often used for the investigation of jet substructure.

1.5.3 Jet Areas

As previously mentioned, it is considered a strength of cone-type algorithms that they mainly produce jets of a fixed size and shape. With the technique of *active area clustering* [45] it has however become possible to determine the area of any jet clustered with an infrared safe algorithm. For that purpose, a very fine-grained grid of additional artificial input particles, called "ghosts" is added to the physical particles. These ghosts carry a transverse momentum of only 10^{-100} GeV, so they do not influence the output of the jet clustering. Typically, about 10^5 uniformly distributed ghosts are used which requires a sufficiently fast algorithm, and excludes SISCone for practical reasons. The ghosts are treated equally during clustering and the number of ghosts that end up being clustered in a jet is proportional to the area it covers. Figure 1.10 shows typical jet topologies for a Monte Carlo event and four jet algorithms. Jet areas offer the perspective of providing a measure to subtract uniform noise and pile-up effects also from k_T jets. The possibility to examine the soft component of the QCD distribution with the help of jet areas is discussed later in this work.

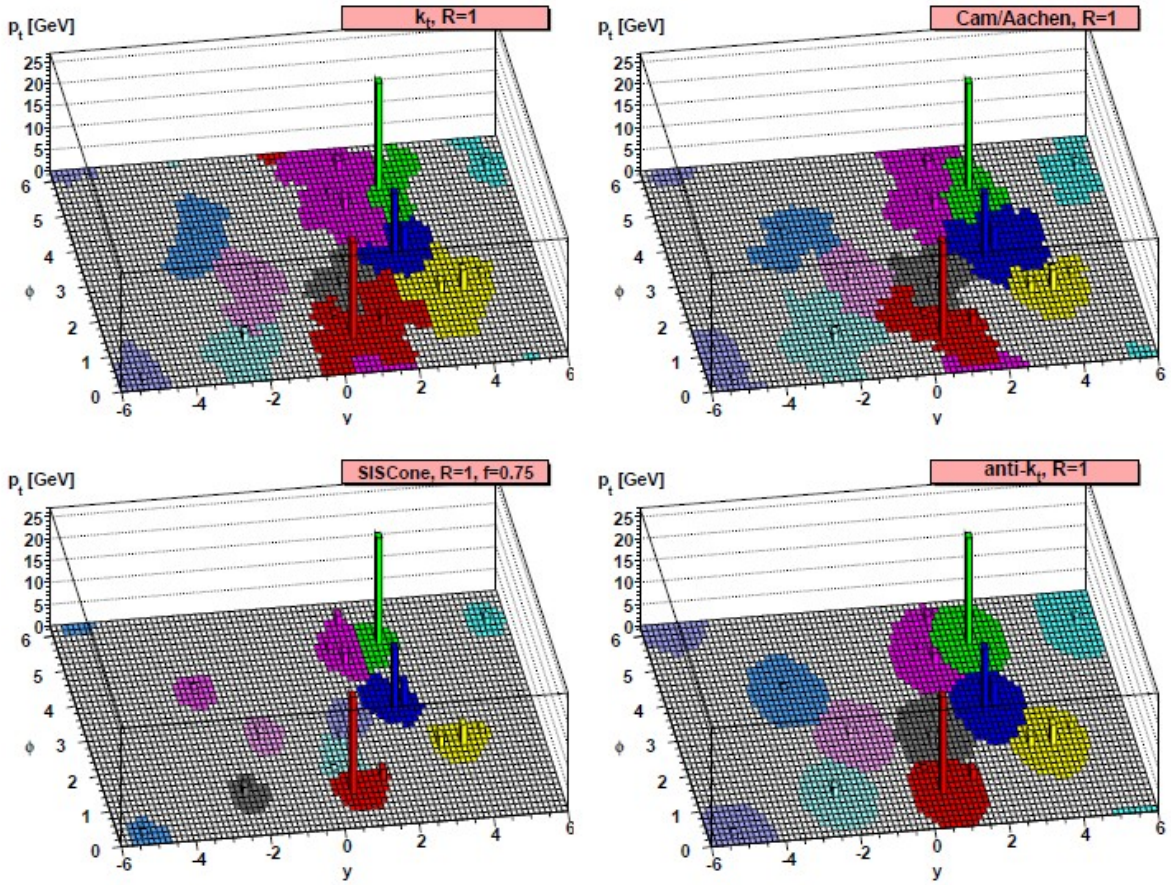


Figure 1.10: *Shapes and areas of jets with different algorithms in $\eta - \phi$. While the k_T and the Cambridge-Aachen algorithms lead to jets with erratic shapes the SISCone and anti-k_T algorithms show a much more circle-like shapes. Note the different shapes of overlapping cones. A uniform grid of* ghost particles *is included in the jet clustering to get a handle on the catchment area of the jets. Taken from [46].*

Chapter 2

The CMS Experiment at the Large Hadron Collider

2.1 The Large Hadron Collider

Housed in the tunnel of the former LEP[1], the Large Hadron Collider (LHC) is the most powerful particle accelerator ever built. To achieve the goal of increasing the production rate of rare particles both the center-of-mass energy and the luminosity of the LHC are unprecedented among hadron colliders.

After the precision measurement of the properties of the weak gauge bosons at LEP and the discovery of the top quark at the Tevatron in 1995, the LHC's main goals are to either find or exclude the existence of the Higgs boson, supersymmetry and other yet undiscovered heavy particles that are not part of the Standard Model. The design of the LHC is centered around superconducting dipole magnets operating at 1.9 K that force two counter-rotating beams of protons on near-circular orbits. To attain that temperature, liquid helium is used in vast quantities. In contrast to proton-antiproton colliders like the Tevatron where both beams can share the same phase-space, two separate rings are required for a proton-proton architecture. The design center-of-mass energy of the colliding beams is 14 TeV with a peak luminosity of 10^{34} cm^{-2}s^{-1}, which requires a nominal dipole field of 8.33 T. After a serious incident in September 2008, where an interconnection between two magnets failed to maintain supraconductivity and caused an electric arc that penetrated the helium system, it was decided to start the LHC physics program with collisions at a center-of-mass energy of 7 TeV. This still is almost three times as powerful as the largest previous hadron collider, the Tevatron. A technical shutdown is scheduled for the end of 2011, after which the design performance of the

[1] Large Electron-Positron Collider

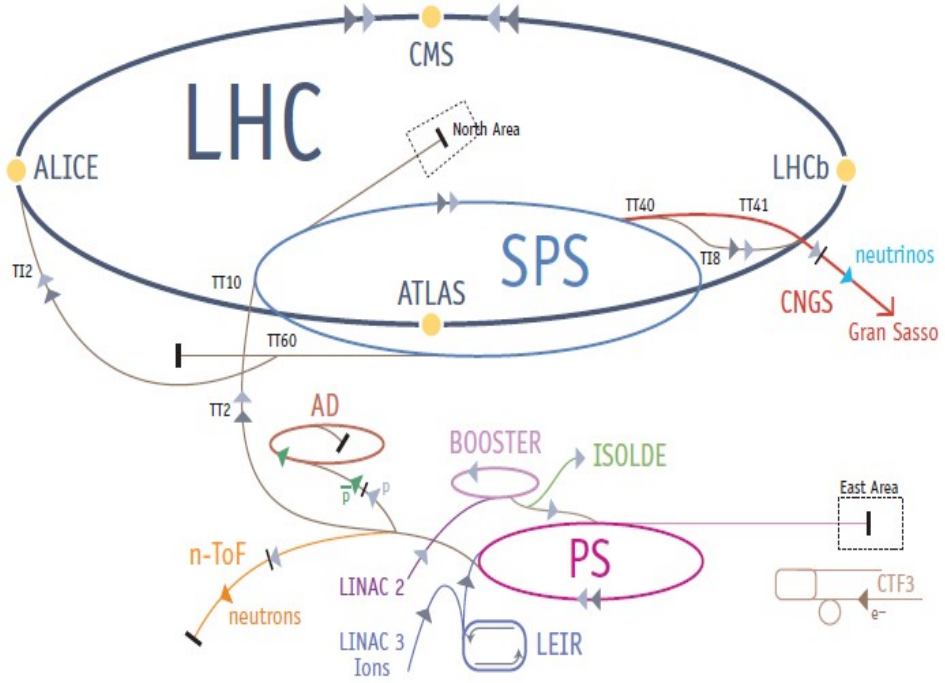


Figure 2.1: *Schematic view of the LHC accelerator complex with its four main experiments. Taken from [47].*

LHC is expected to be reached.

The Large Hadron Collider measures about 27 km in circumference and is installed underground between 45 and 170 meters below the Swiss-French border area near Lake Geneva near the Jura Mountains and the Alps. An overview of the general layout of the accelerator complex can be seen in Fig. 2.1. The pre-acceleration chain involves some of the former CERN colliders such as the PS[2] and the SPS[3] which have been modified to fit this task. The main LHC ring consists of eight straight segments and eight arcs. Each of those sections has an individual cryogenics system so that in case of a section having to be brought to room temperature, not the entire ring is affected. Four main experiments are installed around the crossing points of the two proton beams:

- ALICE [48], a detector designed to investigate collisions of lead nuclei at a center of mass energy of 2.76 TeV per nucleon. Under these extreme conditions, a new state of matter, called quark-gluon plasma is expected to be observed. The high particle density in heavy ion collisions requires extreme radiation hardness of detector components especially close to the interaction point and track reconstruction suitable for thousands of particles in a single event.

- LHCb [49], the only asymmetrical detector at the LHC, specializes in investigating the pro-

[2]Proton-Synchrotron
[3]Super Proton Synchrotron

duction and decay of particles containing b-quarks. The central focus lies on providing the best possible resolution of secondary vertices along the beam-line which are a typical signature of b-quark decays.

- ATLAS [50] and CMS [51] are two multi-purpose detectors with different construction principles and magnetic field designs. Both have a broad physics program including Standard Model and new physics.

ATLAS and CMS are located vis-à-vis at Access Point 1 and 5 respectively, while ALICE is housed in Point 2 and LHCb in Point 8.
In order to achieve the projected luminosity of 10^{34} cm^{-2}s^{-1}, 2808 bunches of 1.15×10^{11} protons will circulate in each direction. At relativistic speed, this corresponds to a collision rate of 40 MHz in the experiments. The relation between number of events of a given process and the luminosity in the experiment is given as

$$\dot{N}_{\text{process}} = \mathcal{L}\, \sigma_{\text{process}} \tag{2.1}$$

with σ_{process} being the cross-section of the specific process. In collider experiments, the luminosity is a function of a number of beam parameters:

$$\mathcal{L} = \frac{N_{\text{b}}^2 n_{\text{b}} f_{\text{rev}} \gamma_{\text{r}}}{4\pi \epsilon_{\text{n}} \beta^*} F \tag{2.2}$$

Here, 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 transverse beam emittance and β^* the beta function at the collision point. Furthermore, in case of a non-zero crossing angle, the geometric luminosity reduction factor F is defined as

$$F = \left(1 + \left(\frac{\theta_{\text{c}} \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{2.3}$$

with the crossing angle θ_{c}, the root mean square of the bunch size in z-direction σ_z and the root mean square of the bunch size in transverse direction σ^*.
There are several factors that limit the performance of the LHC. First of all, the center-of-mass energy is coupled to the radius of the curvature of the proton beams and the maximum attainable magnetic dipole fields. Even though synchrotron radiation is by far less pronounced than in electron positron accelerator rings, it still has to be absorbed by the cryogenics system. Furthermore, the overall stored energy of the proton beams (362 MJ) and magnets (600 MJ) at nominal luminosity and 14 TeV center-of mass energy amasses to 1 GJ. At the end of a running period or in case of an emergency this energy has to be safely absorbed on a very short timescale. Also one has to deal with beam instabilities due to the crossing of the bunches in the interaction points and intra-bunch dynamics of the same-charged protons. During one fill of the collider ring, the luminosity is decreasing, as beam protons constantly collide with remaining gas atoms in the vacuum

Machine Parameter		Nominal Value
Luminosity	$\mathcal{L}$	10^{34} cm^{-2}s^{-1}
Number of bunches	n_{b}	2808
Protons per bunch	N_{b}	10×10^{11}
RMS bunch length	σ_z	7.95 cm
RMS bunch transverse width	σ^*	16.7 µm
Relativistic gamma factor	γ_{r}	7460
Revolution frequency	f_{rev}	11.2 kHz
Bunch crossing frequency	f_{cross}	40 MHz
Crossing angle at interaction point	θ_{c}	285 µrad
Beta function	β^*	0.55 m
Transverse beam emittance	ϵ_{n}	3.75 µm
Luminosity lifetime	τ_{L}	15 h

Table 2.1: *LHC parameters at nominal proton-proton running conditions. Taken from [47]*

pipe and also due the so-called emittance blow-up arising from various kinds of interactions of the beam protons. This puts a limit on the lifetime of a fill and also on the overall time with stable beam conditions.

2.2 The CMS Experiment

The Compact Muon Solenoid experiment (CMS) is one of the two multi-purpose experiments operating at the Large Hadron Collider. Its name stems from the fact that compared to the ATLAS experiment it is significantly smaller in size yet heavier. During the design stages of CMS a special emphasis was placed on the most precise measurement possible of the properties of muons, since the these are important objects in many searches for example for the Higgs boson and supersymmetric particles.
A superconducting solenoid magnet, that creates a homogeneous magnetic field, allows the measurement of the momentum of charged particles via the curvature of their trajectories. The layered design of the CMS detector is shown in Fig. 2.2.

At the moment, the CMS collaboration consists of over 3000 scientists and engineers, from institutes from 38 countries. The detector itself was assembled and tested on site near Cessy, France, from components that were custom made around the world and shipped to CERN. It was subsequently lowered into its cavern slice-wise and re-assembled underground. Due to the delay in LHC operation, extensive runs were performed in which cosmic muons were recorded [54]. The understanding of the geometrical properties of the detector benefitted vastly from this effort even before the start of the LHC physics program.

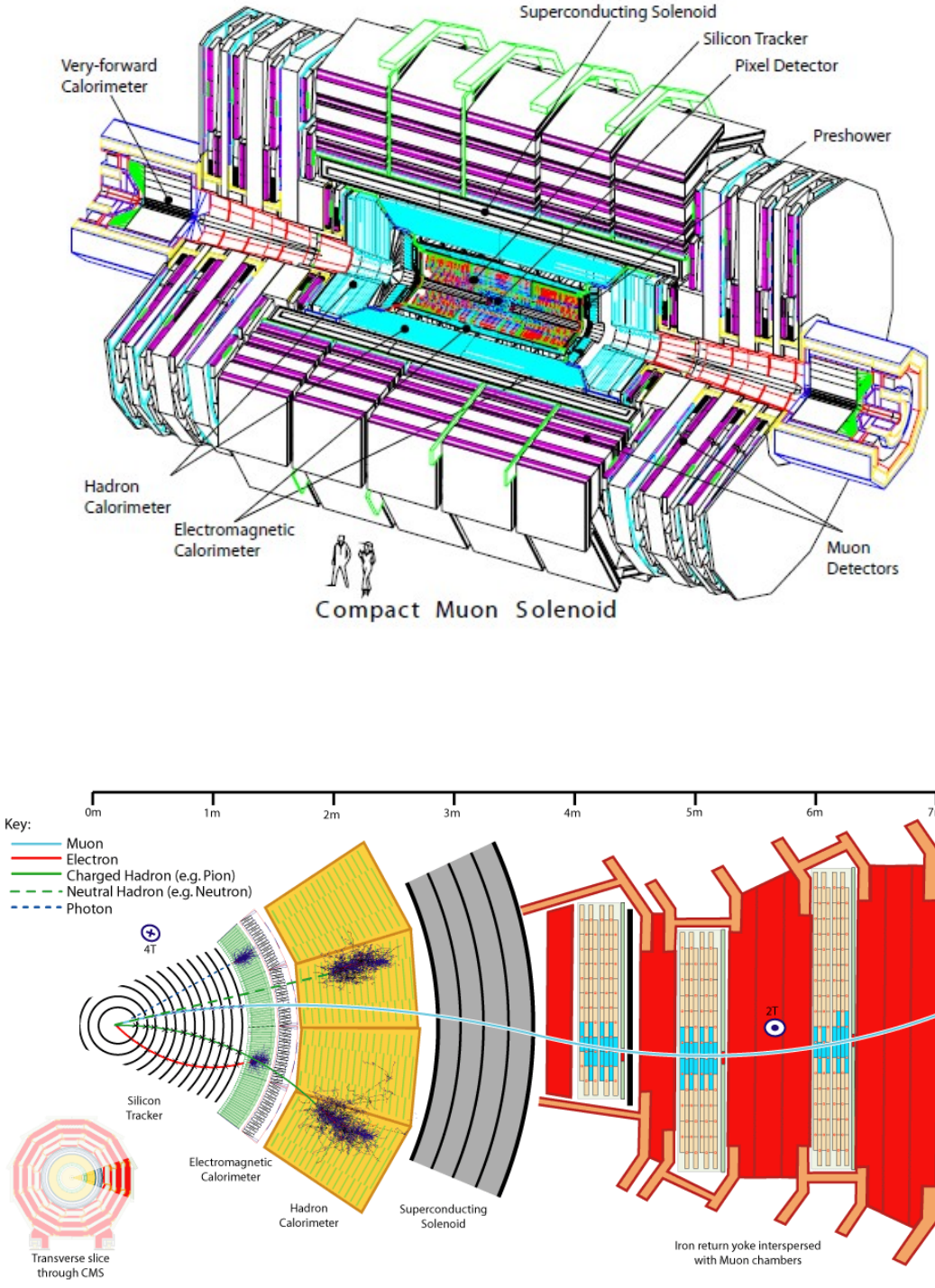


Figure 2.2: *Perspective view of the CMS detector (upper) and slice through the CMS detector and its sub-systems (lower). The trajectories and energy deposits of different kinds of particles are shown as well. Taken from [52] and [53].*

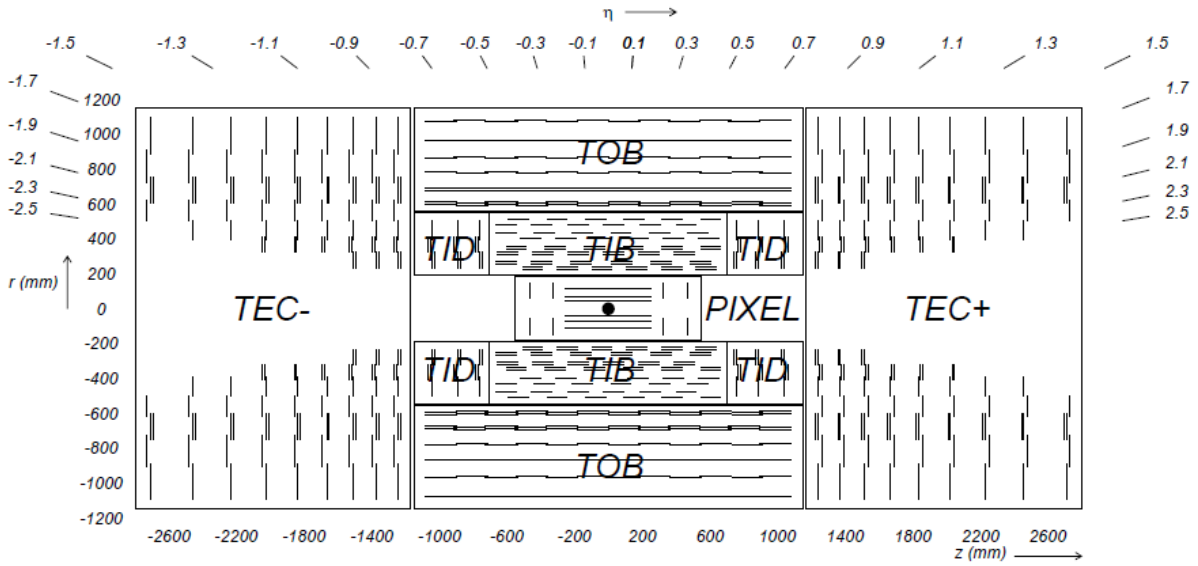
2.2.1 The Solenoid Magnet

CMS uses a superconducting solenoid magnet with a design maximum field of 4T. It is 12.5 m long and has an inner diameter of 6.3 m. At full operational current of 19.14 kA, 2.6 GJ are stored in the magentic field. The magnetic field is returned through an iron yoke that is composed of 5 barrel wheels and 6 endcap disks. The yoke is completely saturated when the magnet is in operation and provides a magnetic field of 1.8 T in the embedded gaseous detectors. For the foreseeable future, the maximum magnetic field is restricted to 3.8 T.

2.2.2 The Tracking System

Design

The innermost detector component of the CMS experiment is the silicon tracking detector, which allows to measure the trajectories of charged particles by combining hit patters from different layers. Two main aims drove the design of the tracker: A fast response time is needed due to the bunch crossing spacing of only 25 ns and the position resolution has to be maximized to allow for a good momentum measurement of the bent tracks even at high energies. Also, given the high flux of ionizing radiation near the interaction point, both the active silicon and the on-board electronics have to be sufficiently radiation hard to withstand years of high luminosity operation. On the other hand, the amount of dead material such as the cooling system has to be kept as low as possible to reduce bremsstrahlung, photon conversion and nuclear interactions. This puts effective limits on the granularity of the silicon structures that can be used in the tracking system.
The 5.8 m long and 2.5 m wide tracking detector relies on two different configurations, silicon pixels and silicon strips. There are three layers of pixels in the barrel region at radii between 4.4 and 10.2 cm and two additional endcap pixel disks that range up to a pseudorapidity of $|\eta| = 2.5$. Overall, 66 million pixels of a size of 100×150 µm cover an area of roughly 1 m^3. An excellent resolution of impact parameters and secondary vertices can be achieved with this setup.
10 layers of silicon strips extend to a radius of 1.1m in the barrel region with 9 additional endcap disk layers. For radii smaller than 55 cm, the strips measure 10 cm $\times 320$ µm $\times 80$ µm. The outermore strips are larger at 25 cm $\times 320$ µm $\times 180$ µm. Some of the modules are mounted double sided with a tilting angle of 110 mrad, which enables the measurement of the third spatial co-ordinate, z in the barrel and r in the endcaps.
The overall active silicon area totals to 198 m^2, the largest such structure ever built. The arrangements of the particular layers as well as the distribution of the material budget can be understood in Fig. 2.3. The obtainable spatial resolution of the tracking system is not equal to the dimensions of the pixels and strips. Due to the Lorentz drift in the magnetic field, mainly two neighbouring channels will share the charge created by the passing through of one ionizing particle. This enhances the resolution to 10×20 µm in the pixel region and 27×230 µm in the strips.



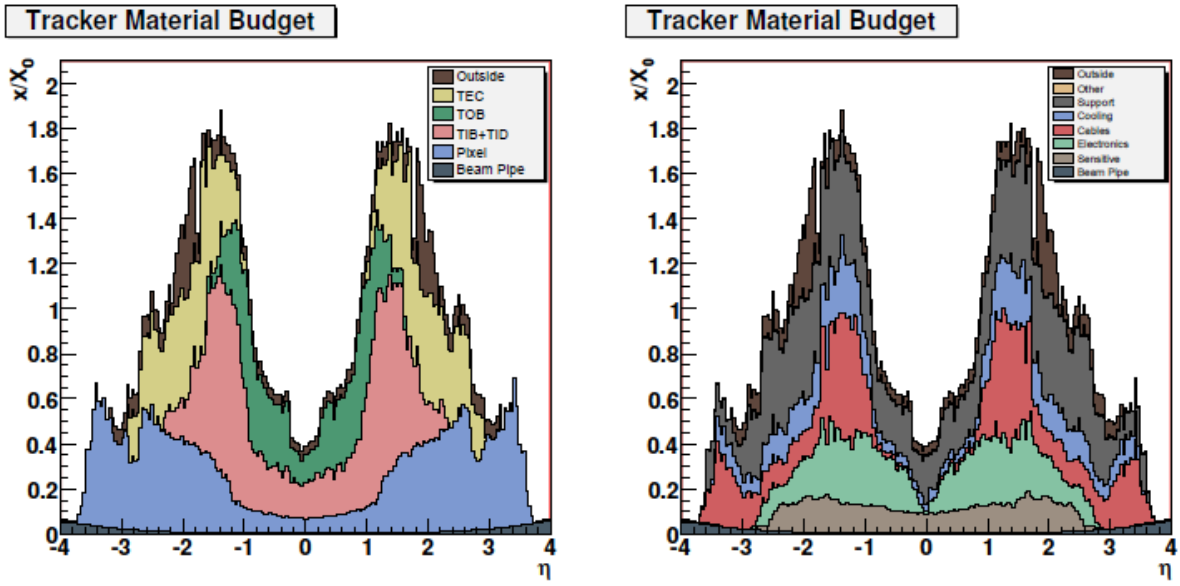


Figure 2.3: *Overview of the CMS silicon detector (upper). The acronyms stand for Tracker Inner Barrel, Tracker Inner Disc, Tracker Outer Barrel and Tracker EndCap.*
Expected tracker material in units of radiation length by sub-detector (lower left) and by functional contribution (lower right). Taken from [52].

Track Reconstruction

The main algorithm to reconstruct tracks in the CMS experiment is a Combinatorial Kalman Filter (CKF), a global least square minimization scheme. It is a local and iterative procedure in which tracks are reconstructed one at a time. The algorithm is robust against energy loss of the traversing particles and multiple scattering. The most common way to reconstruct a track starts with seeds, which are formed by any two pixel hits and constrained by the reconstructed beam spot, which in turn is the average position of all vertices in a given run. Additional restrictions are given by a minimum resolvable transverse momentum of the particle. Pixel hits that act as seed for an eventual track are removed from the collection to avoid duplications. Due to the relatively high interaction rate of hadrons in the tracker and the excellent spatial resolution of the pixels, this *inside-out tracking* is favourable for most configurations.
Starting from the seeds, trajectories are reconstructed layer by layer and each possible combination of hits that is passing a Kalman filter algorithm becomes at track candidate at every step. However, only a certain amount of candidates is kept due to performance restrictions, based on quality measures. The algorithm also accounts for the possibility of a particle not leaving a hit signal in a particular layer of the tracker, a so-called missed hit, but a maximum number of missed hits cannot be exceeded. When the outermost layer is reached, a final fit is made for all possible trajectories. The best combination of hits, taking into account the number of layers that contributed and the χ^2 of the track candidate, is then declared a track and the hits are removed from the collection to avoid redundancy.
With this algorithm, the CMS tracker can reach an efficiency of 98% for muons with a pseudorapidity of $|\eta| < 2.3$. Due to the possibility of nuclear interactions, hadrons have a smaller reconstruction efficiency. Alternative track reconstruction approaches are available in addition to the standard procedure lined out above. They are able to deal for example with non-Gaussian tails of distributions caused by material effects, where the CKF no longer is the optimal choice. A description of these algorithms can be found in [55].
It is obvious that an extremely precise knowledge of the geometrical orientation and positioning of each tracker channel is required for the tracking algorithm to deliver reliable results. Therefore, a database is available that contains alignment information and is queried during the event reconstruction. The effort to align the subsystems already began in the stage of the assembly of the CMS detector by means of photogrammetry. During operation, a laser alignment system is constantly taking data on how the larger structures TIB, TOB, TEC of the tracker are oriented in the CMS co-ordinate system with a precision of about 100 µm. The best knowledge of the tracker alignment however will be retrieved through track based alignment. In an iterative process, the residuals of tracker hits are evaluated and the track quality is improved by alternating repetitions of track reconstruction and alignment algorithms. Prior to the start of LHC operation, millions of cosmic muons were recorded in CMS, both with and without magnetic field. So from the very beginning of taking collision data, a high precision of measuring track quantities was achieved. With growing statistical precision, the goal is to reach position uncertainties of less than 10 µm.

2.2.3 The Electromagnetic Calorimeter

The electromagnetic calorimeter of the CMS detector is designed to identify and measure the momentum of electrons, positrons and photons through their electromagnetic cascades in matter. Due to its well parametrable background, the process $H \rightarrow \gamma\gamma$ is a promising discovery candidate for a light Higgs boson at the LHC despite a low cross-section. To increase sensitivity to this signature, both a good spatial and energy resolution are required from the electromagnetic calorimeter. To keep the ECAL compact, a dense scintillator material with a short radiation length is needed. The material of choice is lead tungstate ($\rho = 8.28$ gcm^{-3}, $X_0 = 0.89$ cm), which is read out with avalanche photo-diodes and vacuum photo-triodes. Additional advantages of the chosen material are its small Molière radius, which is a measure for the transverse extent of the fully contained electromagnetic shower and the fast response time of only 25 ns, in which 80% of the scintillation light is emitted. However, the light yield is small at 4.5 photons per MeV. Through this proportionality factor, the energy of a traversing particle is quantified. In the barrel region, 61200 truncated-pyramid shaped crystals with a front face of 0.0174×0.0174 in $(\eta - \phi)$ are arranged in 36 identical super-modules. The crystal length of 23 cm corresponds to 26 radiation lengths. Together with the 7324 similar crystals in each of the two endcaps the high granularity allows for an excellent position resolution. The ECAL spans out to a radius of 1.55 m and a pseudo-rapidity of $|\eta| = 3.0$ with the barrel-endcap transition at $|\eta| = 1.479$. An overview of the CMS ECAL layout is shown in Fig. 2.4. Before the assembly of the detector, intensive studies of the ECAL components were conducted with test-beams of electrons and neutral pions that decay into a pair of photons. A summary of these efforts can be found in [56] and [57].
In addition to the regular ECAL, a pre-shower detector is installed in the forward direction in front of the endcaps. It both helps identifying neutral pions and increases the resolution in its fiducial volume through its higher granularity. In contrast to the ECAL crystals, the pre-shower detector is a sampling calorimeter consisting of two layers of lead radiators and attached silicons strip sensors. With a thickness that corresponds to $3X_0$, 95% of all incoming photons start their electromagnetic showers in the pre-shower detector.

2.2.4 The Hadron Calorimeter

The CMS hadronic calorimeter follows the purpose of detecting hadrons that have passed through the ECAL without interaction and the tails of the showers of hadrons that started their shower in the ECAL. It completely encloses the ECAL and extends in radius out to the inner limit of the solenoid. An additional calorimeter is placed outside the magnet at central pseudo-rapidity to collect hadronic punch-through of high energetic jets before the muon system. The pseudo-rapidity coverage is extended up to $|\eta| = 5.0$ by means of a forward calorimeter that is installed close to the beam pipe 11 meters from the nominal interaction point.
When interacting with atom nuclei of the detector material, hadrons create showers of mainly pi-

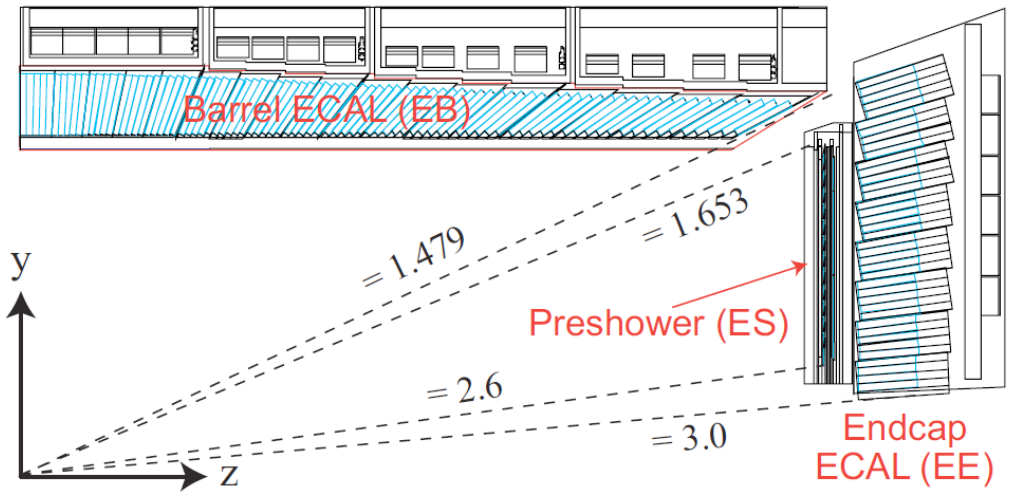


Figure 2.4: *Overview of the layout of the electromagnetic calorimeter of the CMS experiment. Taken from [56].*

ons, kaons and remnants of the destroyed nuclei. These excite the scintillator material and hybrid photo-diodes that are connected via wavelength shifting fibres to read out the signals. The construction architecture of the HCAL is shown in 2.5. It has to be noted that due to the HCAL being exposed to the magnet field, non-magnetic materials have to be used in the forward direction, which is a challenging task for a design of a structure of these dimensions.

In the barrel region (HB), that extends to $|\eta| = 1.3$, 36 identical wedges consisting of four sectors each are mounted in a projective geometry towards the interaction point. Brass is chosen as absorber material with a nuclear interaction length of $\lambda = 16.4$ cm. The individual towers are made of 15 brass plates that are oriented parallel to the proton beams. Each one is 5 cm thick and they are interleaved with 3.7 mm thick scintillator layers. The innermost layer, that is directly attached to the outside of the ECAL to collect hadronic showers created before the HCAL is 9 mm thick. The scintillator fibres are segmented into 16 sectors in pseudo-rapidity, leading to an $\eta - \phi$ front-face of 0.087×0.087. At central pseudo-rapidity, this configuration accounts for 5.4 interaction lengths and over 10 interaction lengths at $|\eta| = 1.3$. Possible shower leakage caused by the relatively thin HB, is recovered by the outer hadron calorimeter (HO). This detector component also reduces hadronic activity in the muon system, which could seriously disturb the measurement.

Overlapping with HB tower 16, the hadronic endcap (HE) covers a pseudo-rapidity region $1.3 < |\eta| < 3.0$. It offers about 10 interactions lengths and a spatial resolution of 0.17×0.17 in $\eta - \phi$. The construction principle closely follows the HB, with brass absorber plates interleaved with scintillator fibres.

Close to the beam pipe to cover high pseudo-rapidities and enhance the hermicity of the phase-space coverage, the hadronic forward detector (HF) is installed. It uses a different approach than the HB and HE. Its design was driven by the very high flux of particles in the forward direction which causes large doses of radiation in this sub-detector. Iron absorber plates and quartz fibres

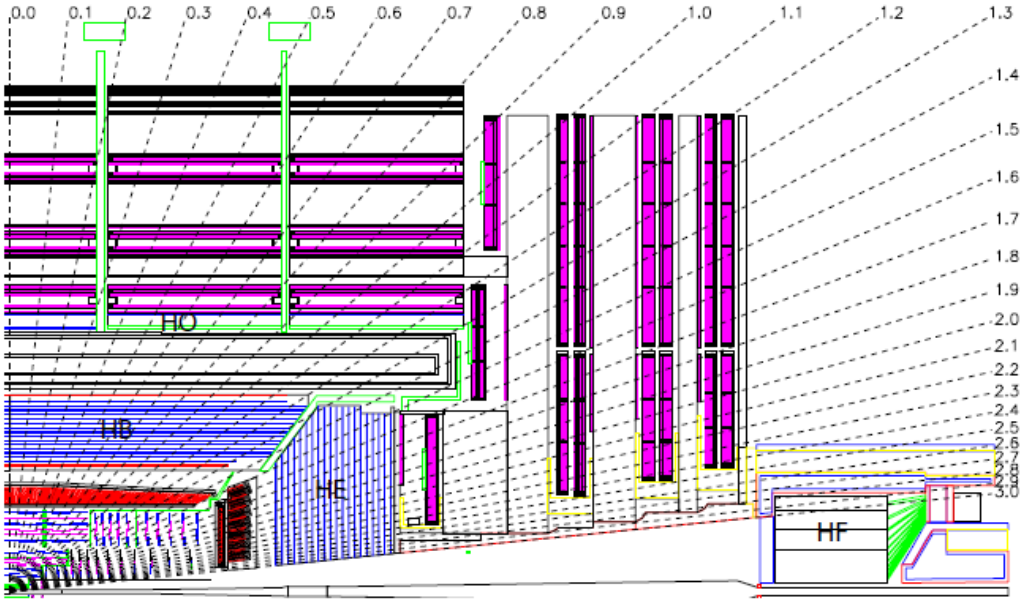


Figure 2.5: *Slice through the CMS hadronic calorimeter. It consists of the Hadronic Barrel (HB), the Hadronic Endcap (HE), the Hadronic Outer (HO) detector and the Hadronic Forward (HF) detector. Taken from [52].*

allow the detection of Cherenkov radiation caused by relativistic secondary particles and are sufficiently radiation hard. However they mainly measure the electromagnetic shower component and therefore their energy resolution suffers.

The calorimetric system in CMS has a number of tasks from identifying isolated electrons and photons up to the reconstruction of jets. It also plays a central role in triggering. The measurement of the overall energy flow in events allows the identification of particles that do not interact with the detector material through an imbalance of the momentum sum of all reconstructed objects, called missing transverse energy (MET). These particles can for example be neutrinos that are produced in weak interactions such as the decay of heavy quarks or bosons. Especially in events that contain supersymmetric particles, large amounts of missing energy are expected.

Two additional detectors are placed outside CMS, the CASTOR[4] experiment and the Zero Degree Calorimeter (ZDC). They cover a pseudorapidity range of $5.2 < |\eta| < 6.6$ and $|\eta| \geq 8.3$ respectively. The design of the two systems relies on quartz fibre scintillators to cope with the high radiation dose. A detailed overview of CASTOR and the ZDC can for example be found in [58].

[4]Centauro And STrange Object Research

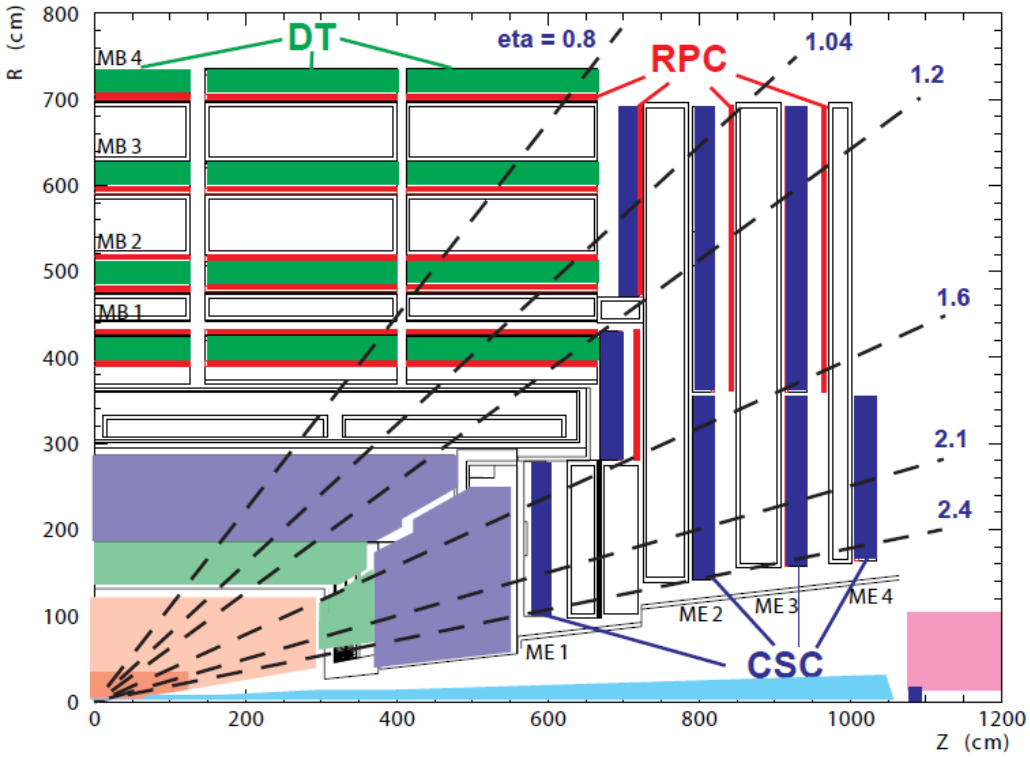


Figure 2.6: *Design overview of the outer CMS muon system [52].*

2.2.5 The Muon System

The precise detection and measurement of muons is a central idea behind the CMS experiment and is reflected in its name. Decay channels with muons do not suffer from the heavy background pollution that many jet based searches have to deal with. Due to their low energy deposit in matter, muons can traverse the calorimeters as well as the magnet coil. Therefore, large arrays of gaseous detectors are interleaved in the magnet return yoke. The layout of the CMS muon system is shown in 2.6. It has to be able to both measure the momenta of muons with the highest possible precision and also be fast enough to contribute to the trigger. Three different types of gaseous detectors are used in the muon system:

- Drift tubes (DT) in the barrel region $|\eta| < 1.2$. They contain a mixture of Argon and Carbon-dioxide and are optimized so that the drift time is never larger than 380 ns. The expected rate of muons is small enough to allow for this relatively large drift time compared to the bunch crossing frequency. The position resolution is in the order of 100 µm.

- Cathode strip chambers (CSC) in the endcaps $0.9 < |\eta| < 2.4$. These are multi-wire proportional chambers that allow a position measurement comparable to the DTs.

- Resistive plate chambers (RPC) in both the barrel and the endcaps $|\eta| < 1.6$ are used to associate a moun to a bunch crossing and operate in avalanche mode. Both the DTs and RPCs contribute independently to the L1 trigger system.

The information from the moun system is combined with tracks from the silicon tracker to allow for an excellent energy resolution even for muons in the TeV range.

2.2.6 Trigger and Data Acquisition

At design luminosity, proton bunches collide in the CMS experiment every 25 ns. At design luminosity, this corresponds to 10^9 proton collisions every second. This event rate is by far not recordable or computable. Therefore, a two step trigger system is in place: The Level-1 (L1) trigger relies on onboard electronics, while the high level trigger reconstructs and investigates whole events that have passed the L1.
To reduce the output frequency from the aforementioned 40 MHz to about 100 kHz, custom-made programmable onboard electronic circuits are directly built into the detector, close to the corresponding read-out channels. Only the calorimeters and the muon chambers contribute to the L1 and thus only energy deposits in small areas of the calorimeters and selected hit patters in the muon chambers can trigger an event to be recorded. The calorimetric information is evaluated locally and finally ranked by the Global Calorimetric Trigger, which reconstructs basic jet quantities or photon and electron candidates. Also, the Global Muon Trigger processes information from all three kinds of muon chambers in which already proto-tracks are reconstructed. All information is then passed to the Global Trigger, which combines the low level information so that also composite objects such as jet multiplicities and missing transverse energy can trigger the L1. As the information of the different sub-detectors has to be stored for a time of 3.2 µs until the L1 has decided whether to read out the entire event, bucket-brigade circuits are used for example to keep tracker information available, while potential muons still traverse the detector before they can trigger the L1 decision. Events passing the L1 are transferred to a filter farm in which standard reconstruction algorithms are run on the detector output and are classified depending on the number of reconstructed objects and their transverse momenta. As the high level trigger is completely software based, it is very flexible to be adjusted to different running conditions of the detector. This includes pre-scales for trigger paths that are expected to fire at a high rate, which means that only a fraction of events passing the particular trigger are accepted to reduce data rates. The HLT reduces the number of events to a few hundred per second, which with a raw event size of about 2MB is the maximum data rate that can reliably be written to storage. The data-flow between trigger and data acquisition is depicted in Fig. 2.7. All raw data are transfered to the main CERN computing centre rom where it is distributed to computer centers aroun the world.

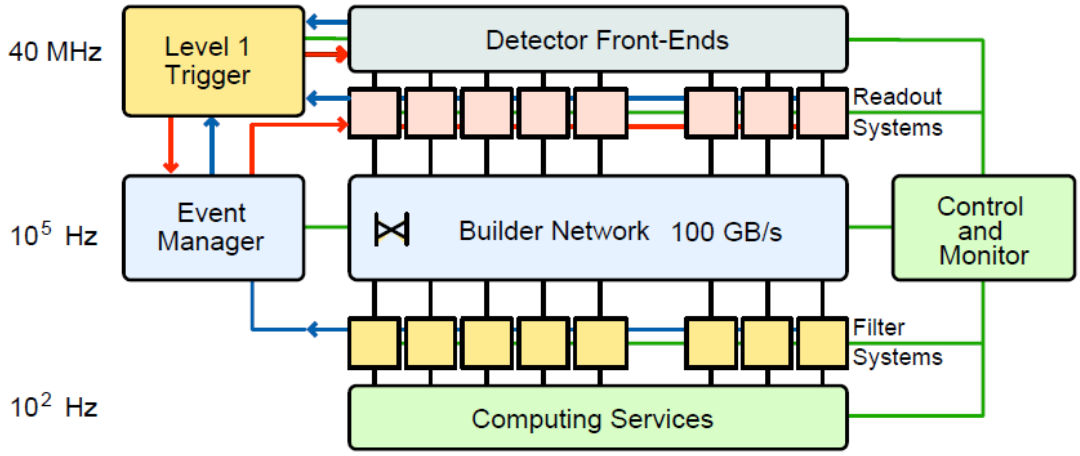


Figure 2.7: *Working principle of the trigger and data acquisition system [52]..*

2.3 The Worldwide LHC Computing Grid

Design Principle

With dozens of institutes and hundreds of scientists located around the world, a decentralized storage and computing concept is required that can deal with the high throughput and workload of modern particle physics. The principle of distributed computing for LHC experiments is closely connected to the idea of grid computing [59] and is governed by the Worldwide LHC Computing Grid (WLCG) organization. A net of computing centres is in place that provides in a tiered structure different services to both central CMS institutions and to individual users such as thousands of CPUs for centrally run large scale Monte-Carlo event simulation and user analysis and in the form of disk space.
The Tier-0 computing centre at CERN receives all recorded data directly from the data acquisition system of the LHC experiments. Large scale storage systems are in place to ensure that a raw copy of each event ever recorded by one of the experiments is available at all times during the running time of the LHC. From here, the datasets are distributed to Tier-1 sites where the bulk of the event reconstruction is performed. Therefore, the Tier-1 centres have a vast number of computing nodes and extensive tape storage available to keep custodial copies of the reconstructed events.
The CMS computing concept excludes end users from running their analyses on Tier-1 resources. For that purpose, each Tier-1 has a number of associated Tier-2s, which provide private storage areas to working groups and individuals that are based geographically close to the particular centre. Copies of datasets of reconstructed events can be requested at Tier-2 centres, which are then transfered on demand from the Tier-1s. Additionally, Tier-3s can be operated for example by university groups that provide flexible, low scale resources to the institute members and are not required to

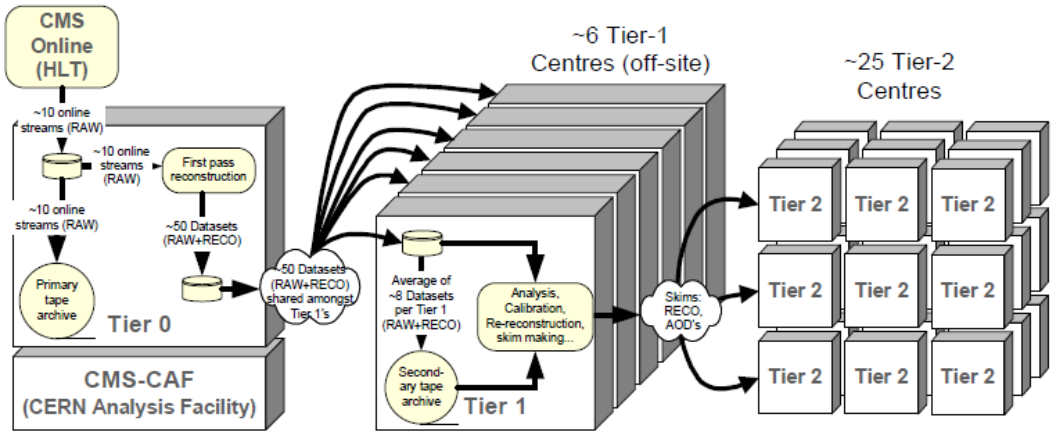


Figure 2.8: *Tiered structure of the Worldwide LHC Computing Grid [52].*

have a 24/7 reliability, as the Tier-0, Tier-1 and Tier-2 have.

Grid Usage

For a large scale computing infrastructure such as the WLCG, security is always an issue. Therefore, all users have to be registered by one or several Virtual Organizations (VO). Before submitting a job to a grid site, the user has to create a proxy certificate, where in the process the virtual organization management servers have to authenticate the creation. This proxy is sent as part of the grid job and identifies the user and in some cases can give him special privileges, for example prioritized access to dedicated national grid resources. Having created the proxy, the user is able to submit his job to the Workload Management System (WMS). He has to provide information on what requirements he has towards software versions, hardware and dataset availability in a job description language (JDL) file. Furthermore, an input sandbox is sent to the WMS, which contains the code the user intends to run. Based on the requirements, the WMS chooses a suitable computing element to which the job is forwarded. Current queue length and proximity to the required datasets are also taken into account for the site selection. The computing elements are portals at the grid sites that act as entry points to local batch queues, which assign the jobs to individual worker nodes (WN). Running on the worker nodes, the jobs have access rights to datasets on mass storage systems as well as write permissions on both local disks and other storage elements (SE) via proxy credentials. Upon successful completion, the grid job sends the output sandbox back to the WMS where it remains until it is fetched by the user. There are several software packages that work as intermediate layers between the grid infrastructure and the end user to provide an easier and more intuitive access to grid resources. One of them is called Grid-Control, which has extensively been used in the scope of this work. A short overview on this is given in App. B.
In addition to the WMS, many other central services are in place that steer the operation of the

WLCG. They range from dataset management infrastructure to sophisticated monitoring tools that help spot possible problems in grid operation. A summary on these applications is given for example in [59].

2.4 Software Concepts

In modern particle physics with hundreds of scientists simultaneously working on one experiment, modular and sustainable software design is vital. In case of the CMS experiment, the bulk of the centrally maintained software is written in C++, which allows for object oriented design where users can plug in their individual analysis code where needed. Databases are in place that provide for example detailed geometric information required for event reconstruction which will change over time.
Together with the recorded data from the detector, also simulated Monte-Carlo events are constantly being produced to check the measured quantities with theoretical predictions and look for possible deviations. As events are completely independent from each other, all tasks concerning event processing are easily parallelizable and benefit from the constant increase in available computing power and infrastructures in the WLCG.

2.4.1 The Event Data Model (EDM)

The basic processing unit in the CMS software model is the event. It corresponds to a recorded bunch crossing. Technically, it is a container that can hold both raw detector output as well as reconstructed physics objects and general information such as the actual event number. Furthermore, additional conditions have to be stored on an Interval-of-Validity (IOV) basis, like information on malfunctioning detector components or alignment constants. The IOV typically corresponds to one run, multiple of which one fill of the LHC is divided into. Due to the decreasing luminosity in one fill, runs are again divided into luminosity sections, typically of 1 to 5 minutes length. The conditions are stored in a database and are queried at different times during reconstruction.
To increase transparency, several well defined configurations, called data-tiers, exist that contain events after a certain processing step:

- RAW contains all detector readout information, including technical trigger bits and metadata. Typically, a RAW file has a size of about 2 MB per recorded event.

- RECO contains reconstructed physics objects that are obtained by running algorithms such as jet finders or pattern recognition on the RAW data. The detector information that was

used to reconstruct these high-level objects is also available. By omitting much of the RAW information, RECO files are only about one third of the size of RAW files per event.

- AOD (Analysis Object Data) contain only minimal content required for particular analyses, obtained from RECO by applying filters. AOD are produced to fit the needs of the end user with special requirements concerning physics content. This results in smaller files of about 0.1MB per event, which are easier to handle.

2.4.2 CMSSW

The common software framework of the CMS collaboration is called CMSSW [60]. It has a highly modular design and a complex dependency structure. The tool SCRAM[5] allows the user to check out a working copy of a particular CMSSW revision and then builds only libraries from modules the user has modified or added. That way, distributed development becomes much easier. The CMSSW framework itself is based on the framework core FWCore and has only a single executable cmsRun for different purposes such as Monte-Carlo event generation, simulation, reconstruction, physics analysis and event visualization. The configuration files that contain steering parameters and a running order for the modules are written in the script language Python. External packages are interfaced to the software and their functionality can be accessed from modules.
The typical workflow of analyzing an event starts from a source. This can either be a file with recorded CMS or Monte-Carlo data, which can also be generated on the fly. Producers can read the event content and add physics objects to the event, a task for which they can access the offline database to retrieve detector and run information. The producer modules can also call external packages like Geant4 [61], which simulates the passage of particles through matter with Monte-Carlo techniques, an essential part of event simulation. Filter modules are capable of reading the event content like trigger information and decide whether to discard the event. User written analyzer modules then write summary information on physics objects in the form of n-tuples or histograms. A defined output which can contain event information in either EDM or other formats is finally written to disk.

2.4.3 ROOT

The Event Data Model of CMSSW is closely connected to the ROOT package [62] and the CMSSW framework makes extensive use of ROOT functions. ROOT offers a large number of tools and libraries for large scale data analysis and is written in C++. It contains methods for statistical analysis, minimization, fitting, histogramming, drawing routines and many more features. Also it has a built-in C++ interpreter and supports script languages like Python. In the present

[5] Software Configuration, Release and Management

analysis, ntuples were used that relied heavily on the ROOT architecture and made use especially of TClonesArrays and TObjects. See App. B for a detailed description of the technicalities.

2.4.4 Jet Reconstruction with FastJet

The fastjet package [42] contains different jet clustering algorithms and is linked to CMSSW as an external package. The different approaches to jet finding are described in detail in Chap. 1.5. The main features of fastjet are the inclusion of the algorithms in a common interface and the first implementation of Voronoi diagrams in jet finding which dramatically speeds up all sequential combination algorithms and raises the possibility of active area clustering.
Different kinds of objects can serve as inputs for the jet clustering algorithms. On Monte Carlo level, usually the four-vectors of all stable particle are subjected to jet clustering, leading to *generator particle jets*. In certain cases however, neutrinos from resonance decay and muons are explicitely excluded, as they leave only a minimal energy deposit in the calorimeters.
The most common choice for the measurement of detector observables are energy deposits in the calorimeters. The position and energy content of the calorimeter towers act as input for the algorithm, the output of which is in this case referred to as *calorimeter jets* [63]. Calorimeter jets are especially robust as they rely only on low-level detector information.
A new approach that is made possible by the excellent tracking detector uses reconstructed particles as input. The following types of particles are distinguished, taking into accoutn information from the tracking detector, the calorimeters and the muon system: Photons, charged hadrons, neutral hadrons, muons and electrons. These *particle flow jets* are expected to have a favourable behaviour concerning energy response and resolution.
Another possible input for jet algorithms are tracks as reconstructed by the tracking detector [64]. It is obvious that this collection only contains information about charged particles. As tracks are however reconstructed in a way that they intrinsically do not suffer from response problems, *track jets* are well suited for precise measurements of low transverse momentum charged activity.

Jet Corrections

As both calorimeter and particle flow objects are not able to account for the complete energy of the jets, a number of corrections to the transverse momenta of the jets are necessary. The CMS collaboration uses a factorized multi-level approach in which only some of the corrections are mandatory, while others can be applied if a particular analysis requires them. The seven levels are in detail:

- Level 1: Offset correction
  Corrects for pile-up and electronic noise [66].

Figure 2.9: *Jet corrections in CMS. A multi-level approach is used, in which the first three levels are required and the others are optional depending on the particular analysis. From [65].*

- Level 2: Relative correction
  Flattens the jet response in pseudo-rapidity [67].
- Level 3: Absolute correction
  Corrects to particle level as a function of p_{T} [68].
- Level 4: EMF correction
  Optional correction for variations of the electromagnetic energy fraction [69].
- Level 5: Flavour correction
  Optional correction for different response of light and heavy quark and gluon initiated jets [70].
- Level 6: UE correction
  Optional correction for underlying event contributions [71].
- Level 7: Parton level corrections
  Optional correction from particle to parton level [72].

A detailed discussion of these corrections and their derivation can be found in [65].

Chapter 3

Underlying Event Contributions to Proton Collisions

Contributions from the underlying event (UE) influence all measurements at hadron colliders. Especially processes that involve jets in the final state suffer from UE contaminations. But also event configurations that seem stable against soft hadronic activity can be influenced, for example through increased fake rates or isolation cuts that are sensitive to extra activity in the event. Thus, for many analyses, the underlying event description by Monte Carlo event generators and the consequent implications on the particular observables contain a potential source of systematic uncertainty. This uncertainty can be estimated by applying multiple Monte Carlo generators or different tunes. It is also obvious that the available models have to be tested and improved regularly during the lifetime of an experiment.

The question of how to separate the UE from the hard interaction in a hadron collision not only arises when it comes to the simulation of UE contributions to signal processes but also when measuring the UE itself. In the following, two approaches to measuring the underlying event activity at hadron colliders are introduced, one of which has been applied recently for the first time with the CMS experiment at the LHC.

In the scope of this work, both this new approach to measuring the UE and the implications of UE activity on the inclusive jet cross-section have been studied. While the latter was part of a preparational analysis before the start of the LHC [73], the techniques developed and established in the preliminary study have also been applied to the first actual measurement of the inclusive jet cross-section in CMS that was published recently [74].

3.1 Underlying Event Contributions to the Inclusive Jet Cross-Section

A benchmark analysis at the LHC is the determination of the inclusive jet cross-section. This quantity expresses the distribution of the hard QCD activity in transverse momentum and pseudo-rapidity and is used for an early verification of the understanding of hard interactions in a new energy regime. Once more data are collected and the systematic and statistical uncertainties are reduced, also constraints on the strong coupling constant and parton density functions can be derived from this quantity. Furthermore, the inclusive jet cross-section is sensitive to contributions from new physics such as contact interactions which lead to a significant deviation of the cross-section from the expected Standard Model behaviour.
Through parton shower techniques as introduced in 1.4.2, multi-purpose Monte Carlo generators such as Pythia are able to generate events with an arbitrary number of final state jets, limited only by the jet clustering resolution. For a precise comparison to theory prediction however, one wants to use the highest order calculations available for the matrix-element. Unfortunately, next-to-leading order calculations of the inclusive jet cross-section are not available in the Monte-Carlo generator MC@NLO [75, 76], thus no particle level event generator is available for this quantity. However, it can be calculated on matrix-element level at NLO precicion with NLOJET++ [82]. Therefore, in order to compare the measured distribution with theory predictions, a multiplicative correction has to be applied to the calculated NLO cross-section [77] that accounts for non-perturbative effects [71].
There is a theoretical ambiguity whether radiation effects are to be included in the scope of the UE, since there is an overlap between higher order matrix element calculations and final and initial state radiation effects. To avoid this double-counting, the non-perturbative corrections to the inclusive jet spectrum only account for multiple parton interaction and hadronization. While it has been shown at the Tevatron [78] that with the choice of jet algorithm, center-of-mass energy and best available Monte Carlo tune applied, the corrections cancel out almost completely, at the LHC with increased particle multiplicities this changes significantly.

3.1.1 The Inclusive Jet Cross-Section

The inclusive jet cross-section can be expressed in bins of transverse momentum Δp_{T} and rapidity Δy as

$$\frac{d^2\sigma}{dp_{\mathrm{T}}dy} = \frac{C_{\mathrm{res}}}{\mathcal{L}\epsilon}\frac{N_{\mathrm{jets}}}{\Delta p_{\mathrm{T}}\Delta y} \tag{3.1}$$

where:

- C_{res} is an *unfolding factor* to correct for detector resolution effects

- $\mathcal{L}$ is the integrated luminosity, ϵ is a factor reflecting event and trigger efficiency

- N_{jets} is the number of jets found in a particular bin.

3.1.2 Preparational Study for CMS

Prior to the start of the LHC physics program, a preparational study has been undertaken, laying out a general plan on measuring and evaluating the inclusive jet cross-section. This study has been documented extensively [71, 73, 77, 79] and the effort on quantifying the impact of the underlying event on this measurement was part of the work presented here.
At the time the analysis was conducted, the start-up energy of the LHC was assumed to be 10 TeV and thus MC samples at this energy were used. Two different jet algorithms were applied: The inclusive k_T with an R-parameter of 0.6 and the SISCone algorithm with $R = 0.7$. The particle level jet spectrum for the k_T algorithm can be found in Fig. 3.1. The expected reach of the analysis was evaluated and found to be around 1.5 TeV for a projected integrated luminosity of 10 fb^{-1}. The lower boundary in transverse momentum was set at 50 GeV, as the jet energy scale is expected to adulterate significantly below this value. The binning in transverse momentum is motivated by the jet energy resolution, with a bin in p_T matching the expected width of a Gaussian fit to the resolution at the center value of the bin. The rapidity binning on the other hand is inspired by the geometry of the CMS detector with two central bins that are contained in the calorimeter barrels, one that is contained in the endcaps, one very forward bin and two transition bins, that are expected to perform a little worse. These distributions were produced with Pythia 6 with the underlying event tune D6T. It can be seen in Fig. 3.1 that that the cross-section falls steeply and spans roughly 10 orders of magnitude. Already at this point it is obvious that uncertainties in the transverse momentum measurement will have a large influence on the resulting spectrum. Another important point is the coverage of the phase space that can be reached with Monte Carlo generators. All Pythia and Herwig implementations generate by default events with respect to their cross-section weight. This would in this case lead to a non-population of the high transverse momentum region due to the extremely small cross-section of these events. Traditionally, this problem has been circumvented by artificially slicing the phase space in bins of $\hat{p_\text{T}}$, which determines the transverse momentum of the hard interaction by implicitly setting cuts on the p_T of the partons originating directly from the hard parton scattering. In each of these bins, the jet spectrum is evaluated individually and the individual spectra are then combined by adding them weighted with their corresponding cross-sections. Another possibility to consistently cover the phase space up to high transverse momenta is given with the so called *flat generation* which re-weights the event probability with a p_T dependent factor, increasing the occurrence rate of rare processes.

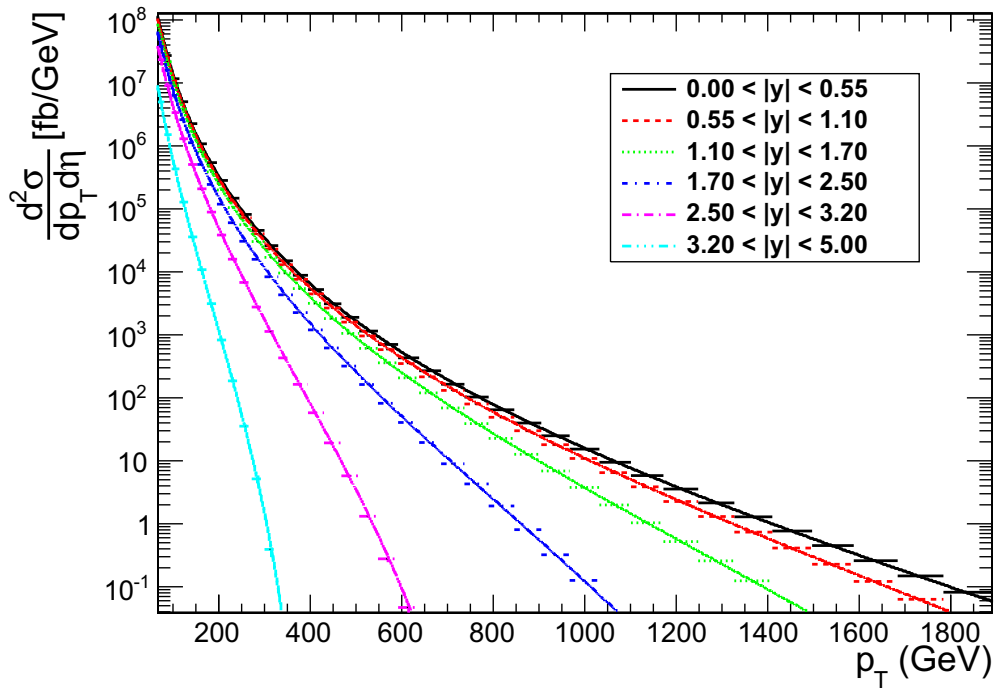


Figure 3.1: *Inclusive jet cross-section at* 10 *TeV from Pythia in the hadronic final state. The jets were clustered with the* k_T *algorithm with an R-parameter of* 0.6 *and different rapidity regions were distinguished. The binning in transverse momentum is motivated by the expected jet resolution of the CMS detector. The fits have been performed using the Ansatz function for inclusive jets.*

Pseudo Data

As at the time of the study, no detector data were available, *pseudo data* were used, that were produced with the Pythia MC generator and the full CMS detector simulation and event reconstruction. For this study, only calorimetric information was used as this is, due to its robustness, an ideal choice for an early analysis on jets. The calorimetric jets were then corrected for calorimetric noise, relative and absolute detector response effects as introduced in chapter 2.4.4. Note that at this point, since all information entering both as pseudo data and as basis for the derivation of the jet corrections, in principle the ratio pseudo data calorimeter jets to Monte Carlo level particle jets should be exactly 1. There are however deviations due to the non-vanishing jet resolution that exhibits tails that cannot be described by Gaussian fits.
The procedure to get rid of these *smearing* effects is rather complicated and known as *unfolding*. It requires to first measure the jet resolution with an asymmetry method, where the measured energy difference in clean di-jet events is evaluated. By definition, this yields the Gaussian core of the resolution. The parameters of the fit depend on the transverse momentum of the jets and the Gaussian distribution is expected to be more narrower for harder jets. The Monte Carlo spectrum is then folded with the p_{T} dependent Gaussian smearing function. By dividing the resulting spectrum with

the original one, a bin-by-bin correction function can be retrieved.

The full CMS detector simulation offers the possibility to simulate malfunctioning detector channels that manifest as noise and influence the measurement as it is expected to happen in the actual data taking. Also cosmic rays and the beam halo can disturb the measured quantities. To get rid of these effects, *cleaning cuts* are used on an event-by-event basis. A robust way to get control of all kinds of unwanted background processes in the case of the inclusive jet spectrum measurement is the "MET/$\sum E_{\mathrm{T}}$" method. As typical hard QCD events contain hardly any neutrinos, it is expected that only a small fraction of the transverse energy of the event remains undetected. Therefore, a large amount of missing transverse energy normalized to the scalar sum of the transverse energy of all jets in the event is an indication for th presence of detector noise, cosmic rays or beam halo. The actual cut can be chosen taking into account measurements of cosmics data and noise simulation and a value of MET/$\sum E_{\mathrm{T}}$ of 0.3 has been shown to clean the data sample very well. Thus, all events with MET/$\sum E_{\mathrm{T}} > 0.3$ are discarded.
An important task when combining the inclusive jet spectrum from detector data is the usage of *trigger streams*. As presented in chapter 2.2.6, triggers are used to enrich "interesting" signatures in the overall recorded data. However, it is necessary to combine the different trigger streams in a consistent manner to avoid double counting of events that contain multiple triggers. For the measurement of the inclusive jet cross-section, single jet triggers are used, that select on the hardest jet found in the event. Different jet triggers are matched to values of transverse momentum at which they are supposed to kick in. To consistently combine the triggered events, it is necessary to identify the triggers' *turn-on points*, which is the value of leading jet transverse momentum at which a certain trigger becomes 99% efficient compared to the next lower one. Each stream is used exclusively in a certain range of transverse momentum.
The measurement of the inclusive jet cross-section suffers from a number of systematic uncertainties. Especially the jet energy scale has a large impact, since a 10% uncertainty on the measurement of jet transverse momentum can lead to a change of up to 100% on the inclusive jet cross-section, due to the steeply falling spectrum. All other uncertainties, for example from luminosity uncertainty and resolution unfolding are a lot less pronounced. The complete picture for the innermost rapidity region can be seen in Fig. 3.2 for both jet algorithms.

Comparison with Theory Predictions

In order to compare measured data with theory predictions, it is desirable to employ the best implementation of the current theory available. Unfortunately, a particle level Monte Carlo generator is not available at next-to-leading order precision for inclusive jets, as this process is not implemented in the MC@NLO [80, 81] generator. There is however the parton-level cross-section integrator NLOJET++, which allows for the calculation of the NLO cross-section. The package fastNLO [83] offers sophisticated techniques to calculate the cross-section numerous times with different PDFs at a vastly improved speed. This eases the determination of systematic uncertainties

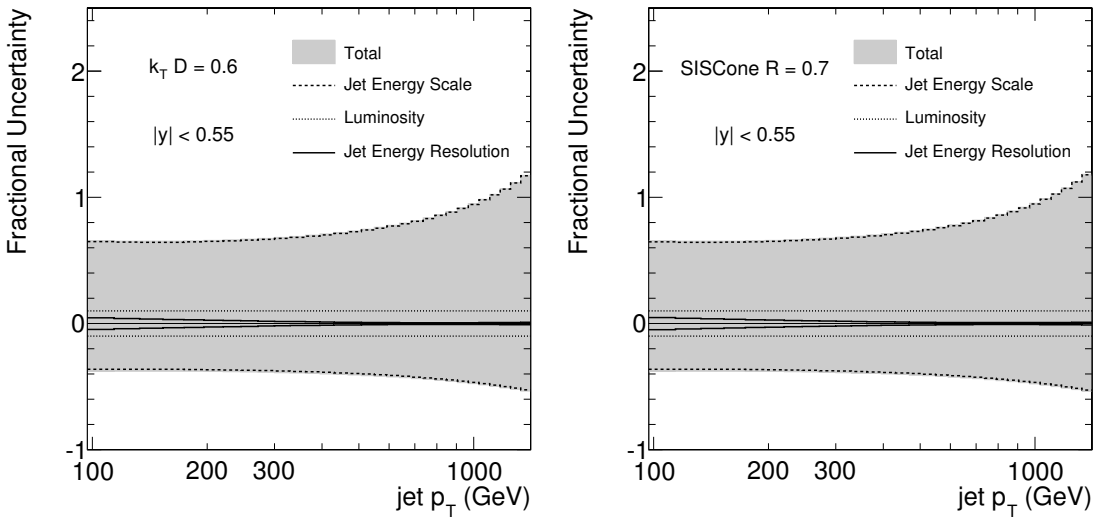


Figure 3.2: *Systematic uncertainties of the inclusive jet cross section for the innermost rapidity region. Inclusive k_T with $R = 0.6$ on the left side and SISCone with $R = 0.7$ on the right side. The jet energy scale is by far the largest contribution with a constant share also from the luminosity uncertainty.*

arising from deviations among PDFs. As the pseudo-data is produced with a leading-order Monte Carlo generator as outlined above, K-factors have to be applied that account for the ratio of NLO and LO calculations. Different parton density functions have to be used as well as different orders of the evolution of the strong coupling constant. The derived K-factors are shown in Fig. 3.3.

Non-perturbative Corrections

As mentioned above, NLO calculations only exist to the parton level. The transition to the particle level has been discussed in chapter 1.4. In contrast to jet corrections such as the absolute and relative correction that are described in 2.4.4, corrections for non-perturbative effects cannot be applied on a jet-by-jet basis. It is therefore necessary to treat the jet spectra as a whole. The procedure works as following: Inclusive jet spectra are generated with different generator settings, either to particle level or to parton level, with or without the treatment of multiple parton interactions. These spectra are then fitted with the Ansatz function for the inclusive jet spectrum

$$f(p_{\mathrm{T}}) = N \cdot p_{\mathrm{T}}^{-a} \cdot \left(1 - \frac{2\cosh(y_{min})p_{\mathrm{T}}}{\sqrt{s}}\right)^{b} \exp(-\gamma p_{\mathrm{T}}), \tag{3.2}$$

which reduces statistical fluctuations. This formula is theoretically motivated and reflects the behaviour of the PDFs and the QCD matrix element. These fits are then divided and the corrections can be applied bin-by-bin in transverse momentum and rapidity. Generally speaking, it would be sufficient to produce only two spectra, one at the particle state and one at the partonic final state without hadronization and multiple parton interaction after the parton shower. It is however bene-

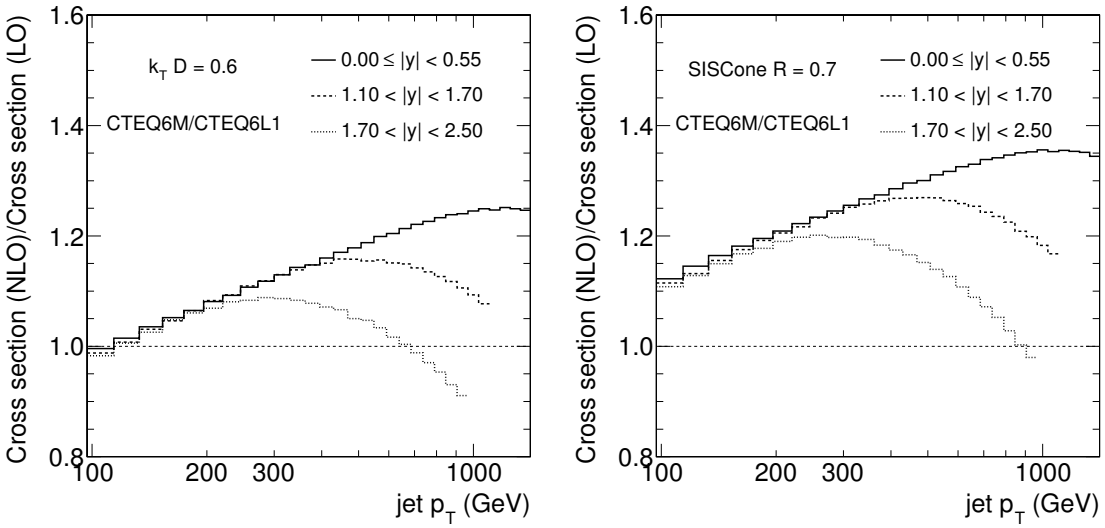


Figure 3.3: *K-factors for the inclusive jet cross-section calculated with NLOJET++ with CTEQ6 PDFs for k_T with $R = 0.6$ (left) and SISCone with $R = 0.7$ for three rapidity bins.*

ficial towards the understanding of the general behaviour of Monte Carlo generators and different tunes and models to study the effects of hadronization and multiple parton interactions independently. Moreover, especially the hadronization corrections are expected to be closely connected to the applied jet algorithm while the MPI corrections, if assumed isotropic, should depend mostly on the jet size parameter.
The influence of hadronization is expected to manifest itself in *out-of-cone effects*, which marks a loss of transverse momentum of a jet through the hadronization step. This effect arises from rather soft particles, that are, unlike their originating partons, not clustered into a given jet. In a more densely populated environment, out-of-cone effects are expected to be less pronounced or even vanishing since the jet algorithm can also catch particles that are not originating from the hard parton one wants to reconstruct. A typical LHC event on Monte Carlo particle level can contain hundreds or thousands of particles, especially in an active event topology such as a hard di-jet event, which is a very densely populated environment in the light of hadronization corrections. To study the interplay between hadronization, jet algorithm and particle densities in detail, an additional, unphysical scenario has been investigated, where multiple parton interactions were disabled in the event generation. This setup is producing a lower particle multiplicity and is expected to exhibit more pronounced out-of-cone effects, resulting in larger hadronization corrections. The distributions are shown in Fig. 3.4 for the Monte Carlo generators Pythia 6 with UE tune D6T and Herwig++ with its default tune for both the inclusive k_T and the SISCone algorithm.

A number of conclusions can be drawn from studying the hadronization corrections. First of all, the string cluster fragmentation model used in Pythia reveals smaller out-of-cone effects than the cluster model used in Herwig++. Concerning the artificial low activity scenario without MPI, the results are as expected: The corrections are about 5 to 10% larger than in the realistic scenario. An-

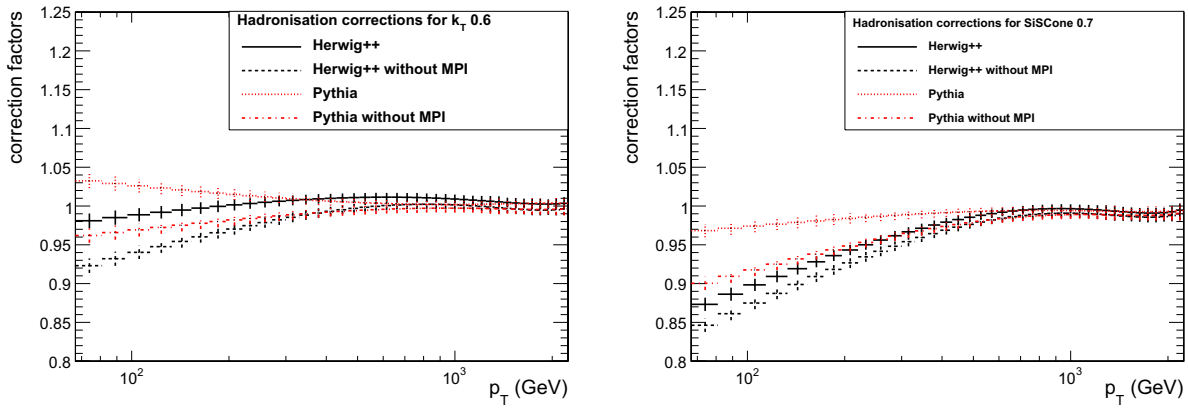


Figure 3.4: *Dependence of the hadronization corrections on the presence of MPI, left for k_T, right for SISCone for the central rapidity region*

other notable feature of the corrections is that the k_T algorithm with its variable jet shape appears more robust against hadronization than the fixed geometrical jet shape of the SISCone algorithm. Finally, for large transverse momenta, all correction functions approach unity, which is expected, as very high energetic jets are generally very focussed and therefore robust against out-of-cone effects.
In contrast to the rather small hadronization corrections, the corrections for multiple parton corrections can be substantial for the inclusive jet spectrum. Especially in the low transverse momentum region, small additional energy contributions can lead to *bin-to-bin migration* effects, that cause a generally harder spectrum. Figure 3.5 shows both the hadronization and MPI corrections for all combinations of the two jet algorithms and the three Monte Carlo generators Pythia 6 with tune D6T, Herwig++ with its default tune and Herwig 6 in combination with the MPI simulation provided by JIMMY. It becomes obvious, that Pythia tune D6T predicts MPI corrections that are almost twice as large for low transverse momenta than both Herwig++ and Herwig+JIMMY, which are almost identical. This comes as no surprise, as the tunes for both generators from the Herwig family employ the same model for MPI, only in slightly different implementations and with minor adjustments.

The combination of both corrections also yields larger values for Pythia, compared to both Herwig implementations. The actual non-perturbative corrections, that were finally applied to the NLO jet spectrum were determined the following way: First, the average of the corrections for both generators of the Herwig family were averaged. As they are employing the same models for hadronization and MPI, their contribution would be exaggerated if treated independently. In the second step, the average of the Herwig-like value and the Pythia value was calculated bin-by-bin in transverse momentum and rapidity. An estimation of the systematic uncertainty is performed by taking half the difference of the combined Herwig and the Pythia value. The final results for both jet algorithms is shown in 3.6. This estimation of a systematic uncertainty is a new method, tested and applied in this context that has since been adopted in several analyses like [84].

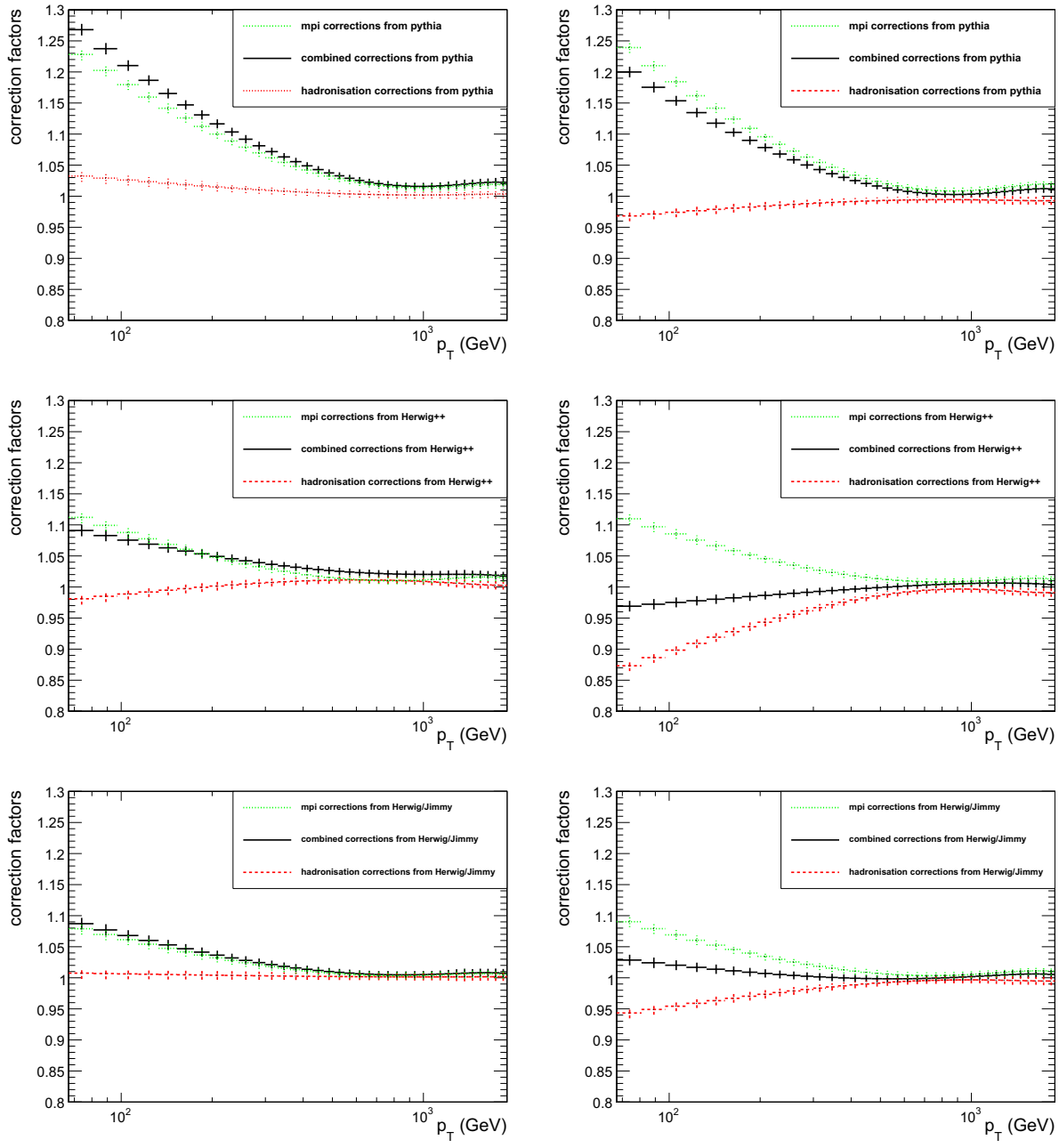


Figure 3.5: *Hadronization and MPI corrections to the inclusive jet cross-section for Pythia (upper row), Herwig++ (middle row) and Herwig6 in combination with JIMMY (lower row). The left plots represent the k_T algorithm with $R = 0.6$ and the right ones the SISCone algorithm with $R = 0.7$.*

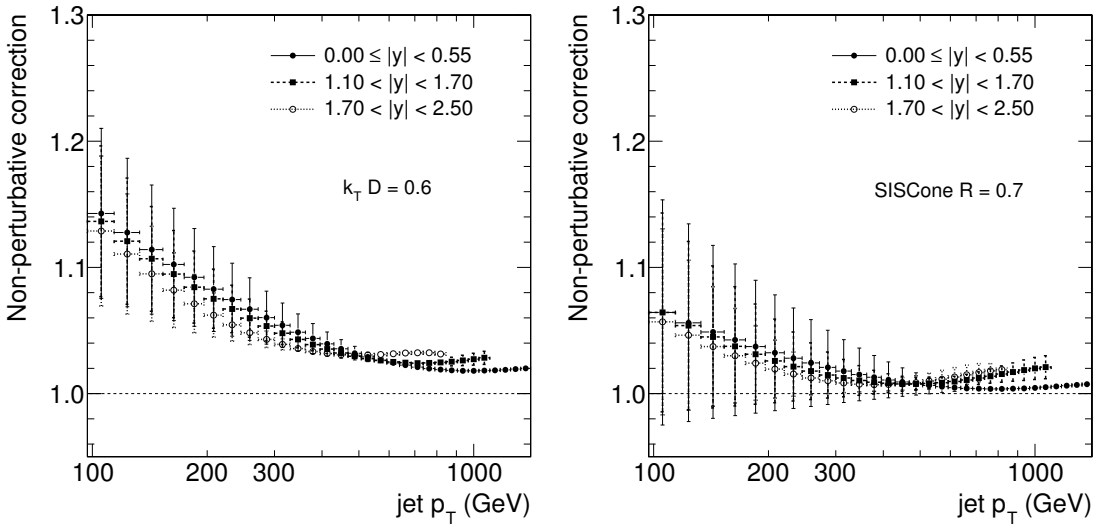


Figure 3.6: *Overall non-perturbative corrections to the inclusive jet cross-section with the k_T algorithm with an R parameter of 0.6 (right) and the SiSCone algorithm with $R = 0.7$ (left) for different rapidity regions.*

An even more detailed look into the interplay of jet algorithms and non-perturbative contributions to the inclusive jet cross-section can be found in [85]. The study presented there contains exhaustive investigations of the influence of the jet size parameter on jet spectra and the implications of different Pythia 6 tunes and has also been part of this work.
From the large difference between the predictions of the two Monte Carlo generators, it is apparent that the extrapolation of the tunes performed with data from the Tevatron is not sufficient. New tunes which contain also results from the LHC are required to reduce the systematic uncertainties and increase the precision of theory predictions for analyses like the inclusive jet cross-section.
In addition to the systematic uncertainty arising from the non-perturbative correction functions, further contributions on the theory side are caused by intrinsic uncertainties of parton density functions and the choices of the NLO renormalization and factorization scales. While the scale uncertainty dominates for higher transverse momenta, the non-perturbative corrections give rise to the dominating systematic uncertainty at small transverse momenta. The complete picture of the fractional uncertainties centered around zero is given in Fig. 3.7. A comparison of the size of systematic uncertainties for the pseudo-data and for the theory prediction is given in Fig. 3.8, together with the expected statistical uncertainty for 10 fb^{-1} of available data. It can be seen, that the jet energy scale uncertainty is still vastly dominant and surpasses also the combined theory uncertainties by far.

As a final result of the preparational study of the inclusive jet cross-section, the sensitivity for contact interactions is demonstrated in 3.10. It is shown, that for contact interactions at a scale of $\Lambda = 3$ TeV, deviations can be seen in the inclusive jet cross-section for values of the transverse momentum larger than 800 GeV.

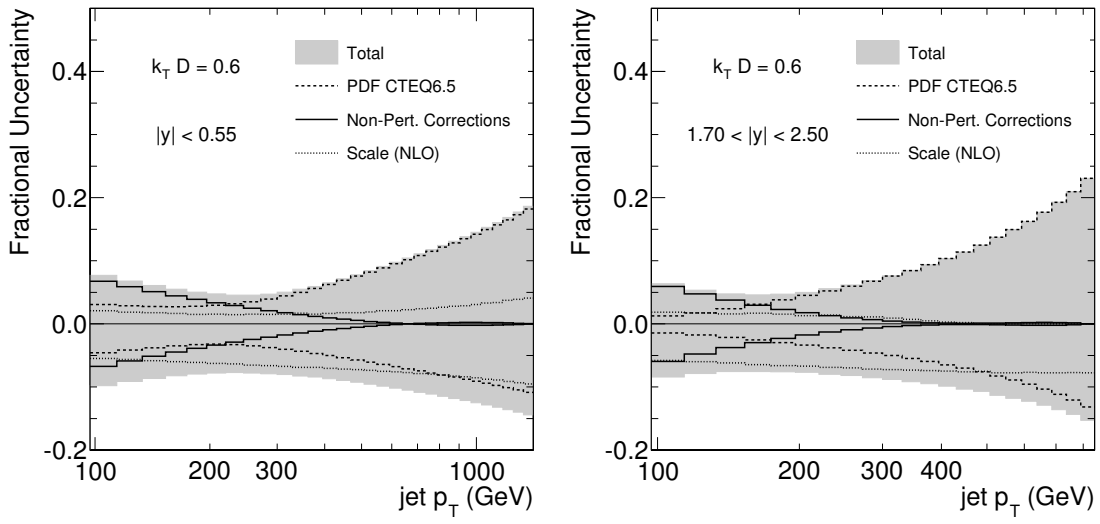


Figure 3.7: *Fractional theory uncertainties on the inclusive jet cross-section. For small transverse momenta, the leading contribution are the non-perturbative corrections, while for large transverse momenta, the uncertainty is dominated by the scale variation uncertainty.*

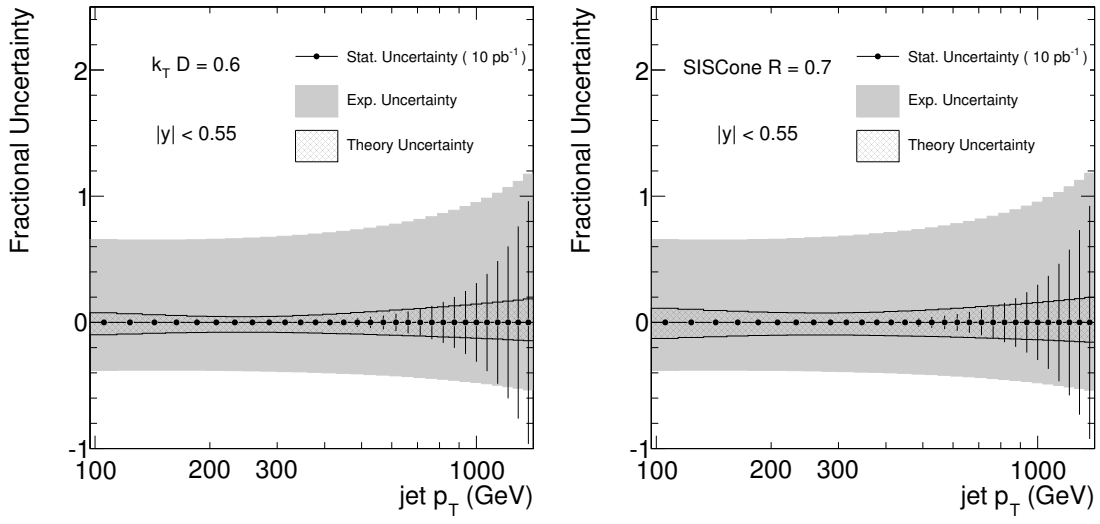


Figure 3.8: *Comparison between systematic uncertainties on the pseudo data and the theory side. Furthermore the expected statistical uncertainty for* $10\ fb^{-1}$ *of available data is given. It is obvious, that the jet energy scale is by far the dominant uncertainty.*

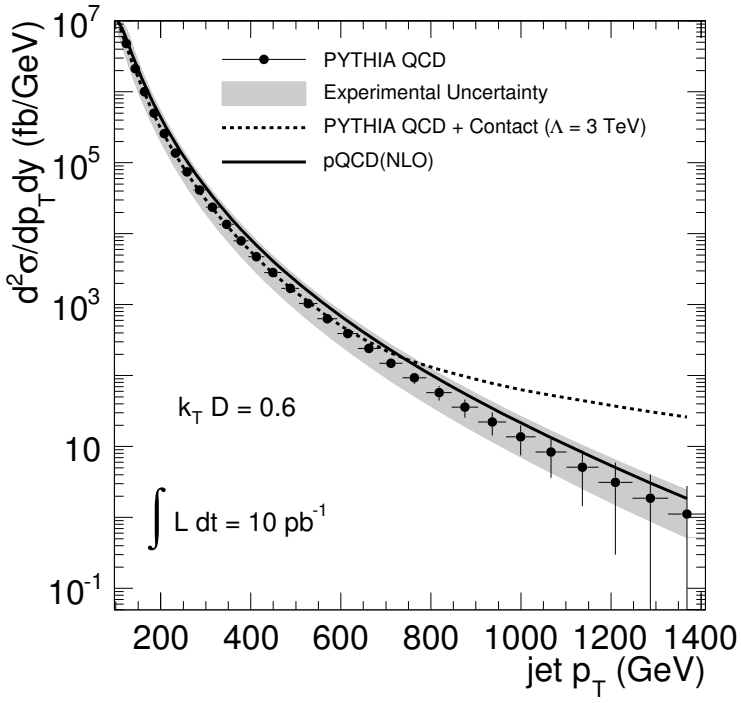


Figure 3.9: *QCD spectrum from pseudo data with combined uncertainties and an additional scenario with contact interactions at $\Lambda = 3$ TeV.*

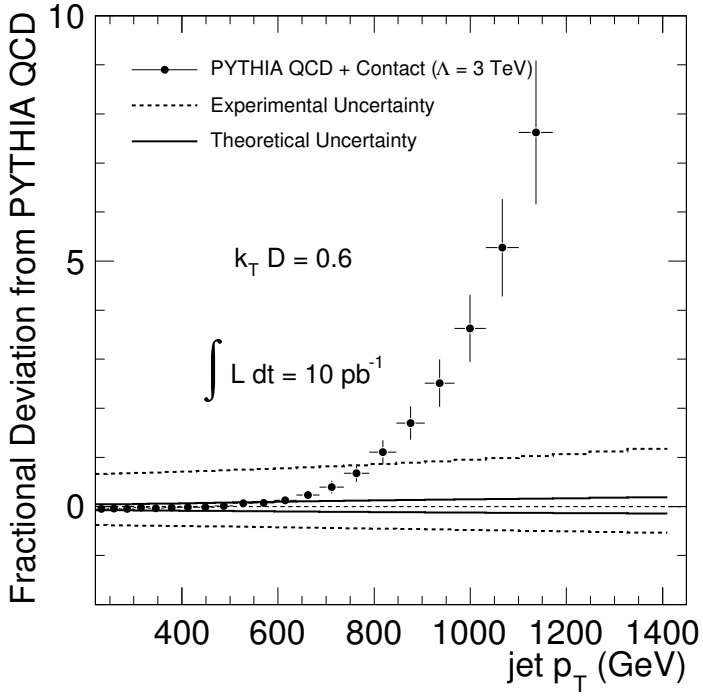


Figure 3.10: *Sensitivity on contact interactions. The ratio between the pure QCD pseudo data and a dataset with contact interactions at $\Lambda = 3$ TeV is displayed.*

3.1.3 First Results from CMS

The first measurement of the inclusive jet cross-section with the CMS experiment has been performed in 2010 with 60 nb^{-1} of data collected at a center of mass energy of 7 TeV. The techniques established and tested in the preparational study have widely been used also when handling actual detector data. The complete study is described in detail in [74]. A new perspective is demonstrated by the usage of *particle flow* jets that take into account tracking information in addition to calorimetric information. This new class of jets allows for an extended reach in low transverse momenta down to 20 GeV. In contrast to earlier studies, the jet algorithm of choice has been the anti-k_T algorithm, which for the measurement of inclusive jets combines the advantages of fast computing times, infra-red and collinear safety and a relatively fixed jet area.
While minor adjustments were necessary to cope with the challenges of a detector operating in its early stages, the strategy to retrieve theory predictions was practically unaltered. It was however decided to omit Herwig 6 in combination with JIMMY from the calculation of the non-perturbative correction functions and use only Pythia 6 tune D6T and the default tune of Herwig++. Due to the extended lower reach in transverse momentum, the region in which the non-perturbative corrections contribute significantly is even more pronounced than in the preperational study. The correction functions and their contribution to the overall theory uncertainty are shown in Fig. 3.11.

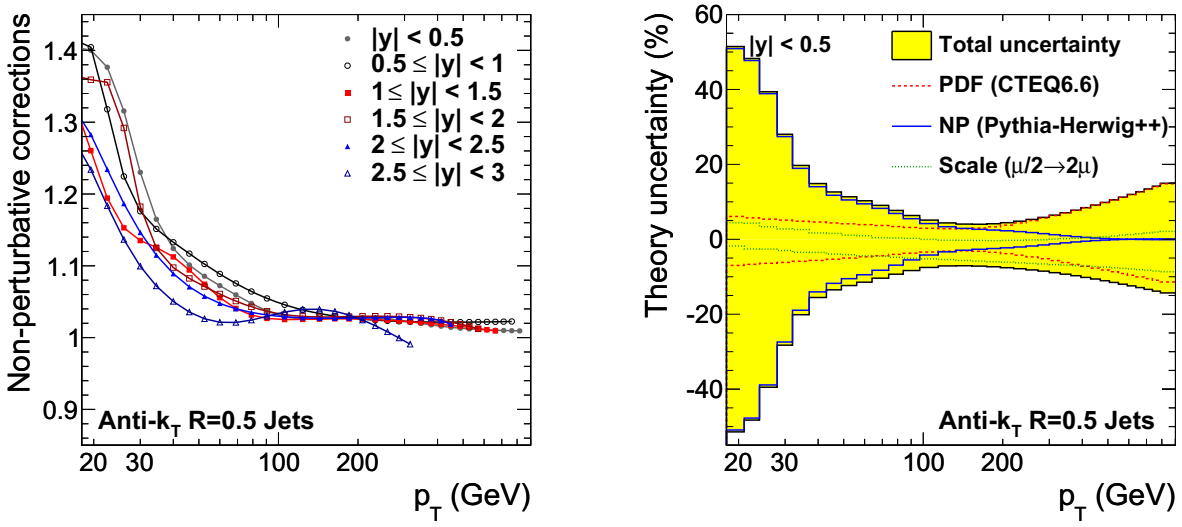


Figure 3.11: *Left: Non-perturbative corrections to the theory predictions of the inclusive jet cross-section at 7 TeV for all rapidity regions. Right: Fractional theoretical uncertainties for the innermost rapidity region.*
Note that the range in transverse momentum is extended further down than in the preparational 10 TeV study. From [74]

Concerning the systematic uncertainties to the theory predictions arising from non-perturbative corrections, which were deduced from taking the average of the correction factors derived from Herwig++ and Pythia as lined out in the previous chapter, the extension to lower transverse mo-

menta magnifies the region in which these uncertainties are dominating. Also the maximum values of the relative uncertainties are increasing up to 50%, which renders precise predictions in this range difficult.
At this point, the need for improved tunes containing measurements from the LHC becomes apparent. New techniques in jet reconstruction allow the probing of a phase space region, that is especially interesting in the light of constraining parton density in the low-x region. In Fig. 3.11 the non-perturbative uncertainty is up to one order of magnitude larger than the PDF uncertainty. It is therefore necessary to produce tunes for both Herwig++ and Pythia 6 that can confidently be used at LHC energies without compromising their consistency with Tevatron measurements.

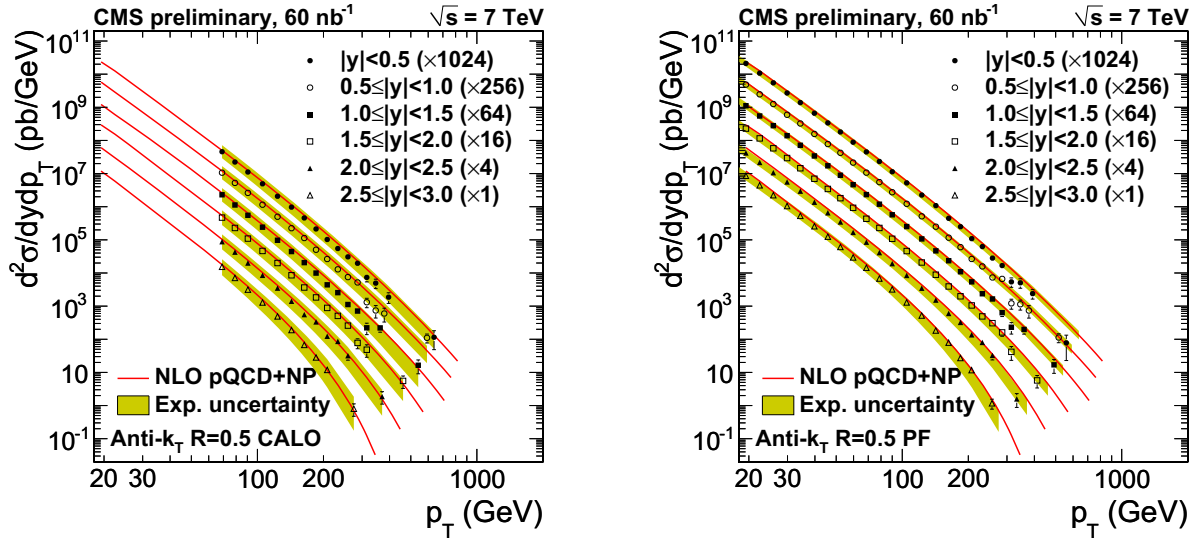


Figure 3.12: *Comparison between measured inclusive jet-cross section and theory predictions at 7 TeV with the CMS detector. Calorimetric jets are shown on the left, with a lower transverse momentum reach of 50 GeV, while particle flow jets on the right reach down to 20 GeV. From [74].*

Figure 3.12 shows the final result of the first measurement of the inclusive jet cross-section with the CMS detector both with calorimetric and particle flow jets. The measurement is consistent with the theoretical prediction within the uncertainties imposed mainly by the jet energy measurement uncertainty. With growing integrated luminosity, the reach in transverse momentum will grow and sensitivity for new physics in the inclusive jet cross-section will be established. On the other hand, in the low transverse momentum region with an improved jet energy scale, probing present parton density functions will become feasible.

3.1.4 Measurements at the Tevatron

A number of articles have been published dealing with the measurement of the inclusive jet cross-section at the Tevatron, both with the CDF [86–88] and the DØ [89–91] experiments. Both run I at 1.8 TeV and run II at 1.96 TeV set the respective limits on heavy resonances and constraints on the low-x gluon PDFs in the inclusive jet cross-section at the time of their publications. An example for the non-perturbative corrections derived with Pythia 6 tune A is given in 3.13 while the latest results on the comparison between measurement and theory are given in 3.14.

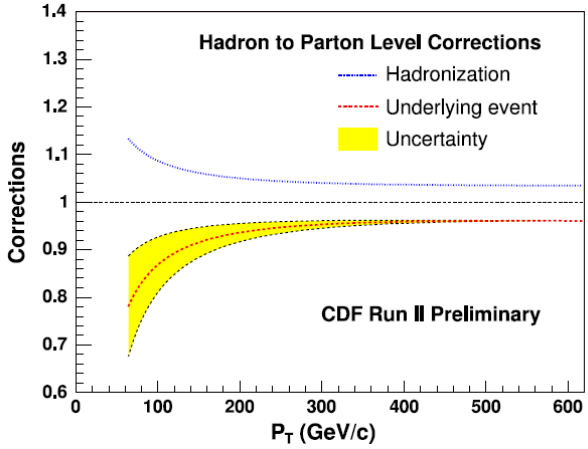


Figure 3.13: *Non-perturbative corrections on the inclusive jet cross-section for CDF II produced with Pythia 6 tune A. The MidPoint cone jet algorithm was used with a size parameter of $R = 0.7$. The label "underlying event" contains what is here referred to as MPI. The two effects basically cancel each other. Taken from [92].*

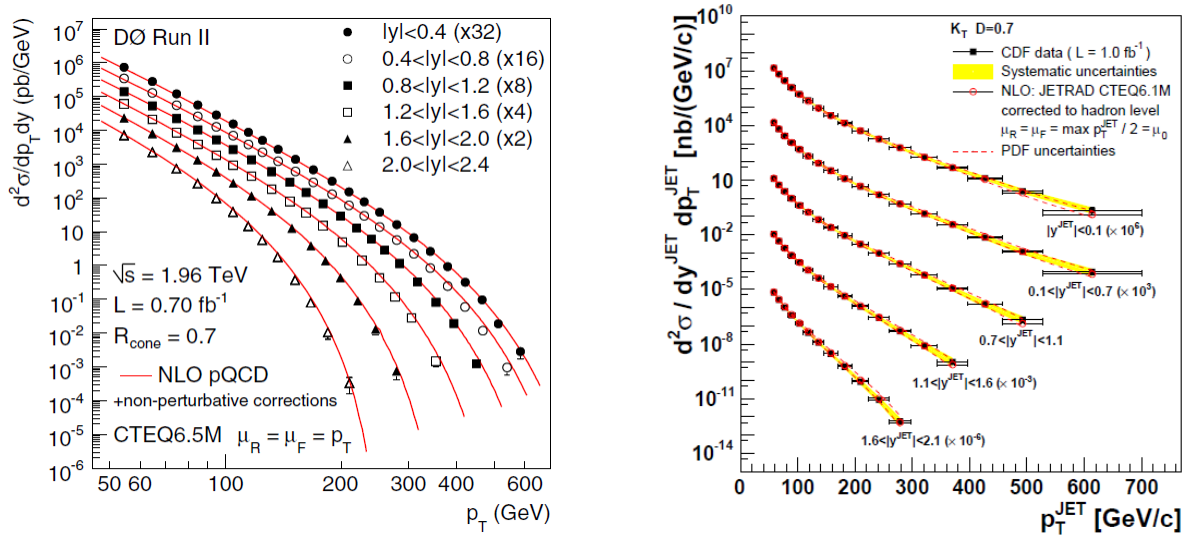


Figure 3.14: *Inclusive jet spectrum from run II of the DØ(left) and CDF experiments (right).Taken from [39] and [93]. Note that the reach of these measurements has already been matched or even surpassed by CMS.*

3.2 The Traditional Approach to Measuring the Underlying Event

Already the term "underlying event" contains part of the problems and ambiguities this phenomenon is regarded with. In principle, in a hadron collision, there is simply one event, whether containing a hard scattering that can be calculated with perturbative QCD or not. As all partons inside a proton are connected via colour fields, every parton interaction with significant momentum transfer will always affect the other partons as well. So, the separation between a hard and a soft component, that is necessary in order to make predictions in the form of generated Monte Carlo events, is strictly speaking already a simplification of the intra-proton dynamic. The parting of the hard and soft component during generation also has implications on the physical understanding of soft QCD: As the hard interaction is treated as the dominant process in a Monte Carlo event generation, in order to study the soft component and be able to model it, one has to find a way to isolate it from influences of the hard scattering in data.

It has been only since run I of the Tevatron that a measurement of the underlying event in hadron collisions has been undertaken. The method applied there relies on a purely geometrical slicing of the event in order to separate the regions in which the contributions from the underlying event can be isolated. A leading object, which can either be a jet, a track or a muon from Drell-Yan production, is identified. Detector acceptance has to be taken into account and appropriate cuts have to be applied. In case the leading object is chosen to be a jet, the leading two jets in transverse momentum are required to be above a certain threshold in p_{T} to ensure the presence of a hard QCD interaction. Then the region perpendicular to this object is investigated, as it is expected to contain only contribution from the underlying event.
This definition however is not safe from hard radiation into the transverse region either from initial or final state partons. Therefore, in the studies performed at Tevatron [94], additional cuts on the direction and transverse momentum of the second and third leading jets are applied to isolate the effects of multiple parton interactions and beam-beam remnants. The resulting event selection is referred to as *back-to-back* topology. A sketch of this event topology and the geometrical slicing is given in Fig. 3.15.
An extension of this measurement also takes into account differences in activity in certain parts of the transverse region. Thus, a *trans-max* and a *trans-min* region are defined which are the halves of the transverse region with the highest and the lowest transverse momentum sum. The trans-max region is especially sensitive to contributions from initial state radiation, an effect that is suppressed by the cuts on the second and third jets in the back-to-back topology. Therefore, these cuts have to be loosened in this extended analysis. In contrast, the trans-min region should contain mostly contributions from beam-beam remnants and multiple parton interactions. This selection is called *leading jet* topology as the transverse momentum cut on the first jet is the only remaining requirement on the event topology.
The measurement aims to cover the transverse momenta of particles as low as possible, which

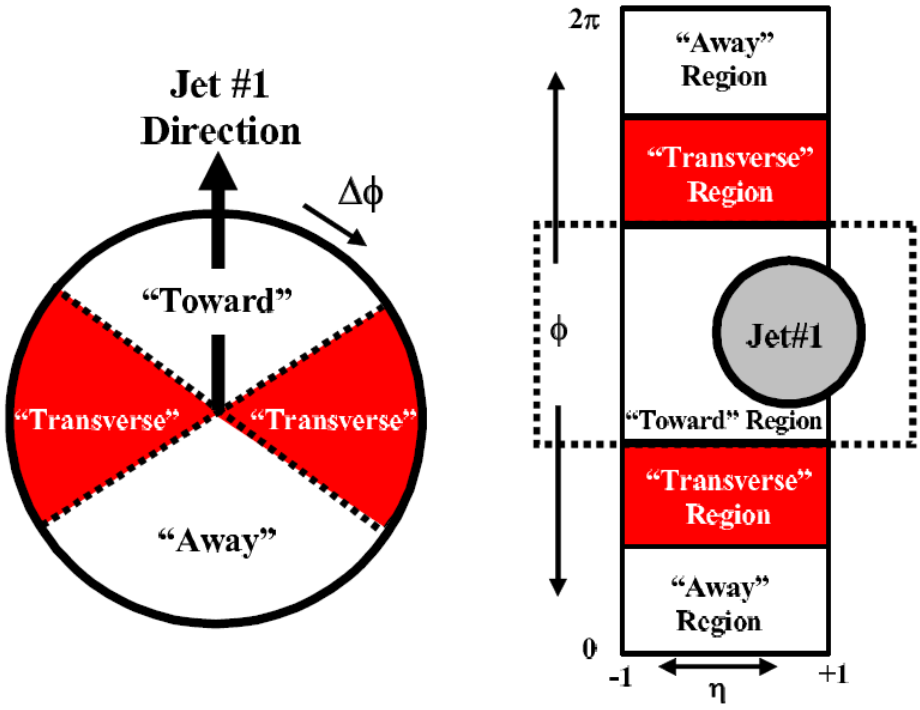


Figure 3.15: *Correlation of directions in an event relative to leading jet. The transverse area is expected to contain information only from the underlying event. Taken from [94].*

can be attained using tracking information. Thus, it is obvious, that only the charged component of the underlying event can be measured as neutral particles cannot be detected with tracking detectors. In the transverse region, the observables of interest are the number of tracks and the sum of their transverse momenta, both usually normalized per unit area in $\eta - \phi$. These quantities are then evaluated as a function of the transverse momentum of the leading object. It is expected that with harder primary interaction and therefore harder leading object, also the underlying event contribution is increasing up to a plateau due to saturation effects.

3.2.1 Measurements

One of the main contributions to the underlying event, the multiple parton interactions, are heavily dependent on the center-of-mass energy of the hadron collision. This behaviour is related to the growing parton density functions for gluons at low Bjorken-x. A sophisticated model for multiple parton interactions must be able to deal with this property. In the Monte-Carlo generator Pythia, see chapter 1.4.5, this challenge is met by applying a center-of-mass dependent cut-off value for MPI simulation. It is therefore necessary to test the existing models at different energies and adjust them where necessary. Prior to the start of the LHC, the underlying event had only been probed at the Tevatron at energies of 1.8 and 1.96 TeV [94, 95] and only recently and at a reduced extend

at 200 GeV by the STAR collaboration at the RHIC [1] accelerator [96]. It is clear that with only one collision energy at disposal, the energy dependence cannot be verified and in fact the different tunes derived from Tevatron measurements predicted a very different behaviour at LHC energy. Also, after the first measurement of the underlying event with the traditional method at the LHC, the main challenge in the tuning process is to find a combination of the parameters that can describe soft QCD contributions at the new energy scale that are compatible with the results from Tevatron.

Tevatron

The concept of the traditional measurement of the underlying event was refined in [97], in which both the structure of the leading jet and the transverse region was studied. Three effects were identified that are contributing to the transverse region: Initial state radiation, beam-beam remnants and multiple parton interactions. Before these studies, Pythia contained basically only one set of parameters accounting for these effects. As it is not clear, which effect contributes to what amount, different tunes were proposed in which the weights of the individual contributions to the phase space differ. This effect also depends on the transverse momentum of the leading jet, as for example, the proton remnants have to absorb more recoil from harder scatterings and therefore fragment differently. The dependency of of the charged multiplicity and momentum sum are shown in 3.16.

The analysis summarized in [97] resulted in the formulation of tune A, the first of the so-called "Pythia 6.2" tunes. It did not yet take into account any measurements including Drell-Yan data. Complementary measurements led to the establishment of a number of additional tunes: AW, DW and DWT.[2] Details and the parameter sets of these tunes can be found in App. A. In addition to these Pythia tunes that all rely on the q^2-ordered parton shower, new tunes were established that include the p_{T}-ordered parton shower [98]. Thus, before the start of the physics program of the LHC, a wide array of tunes was available whose prediction differed significantly. An example is given in Fig. 3.17, which displays the projected transverse charged density for the two most popular tunes, DWT and DW.

CMS

Being a non-subtractable background present in all analyses, the underlying event is one of the measurements performed early in the stage of new collider experiments. Even though the modelling is most important at the actual center-of-mass energy of the physics program, the commissioning run of the LHC in 2009 raised the possibility of getting another important measurement point towards generalizing models valid at all collision energies. The results of this study with the

[1] Relativistic Heavy Ion Collider

[2] Additionally, a tune was established for Herwig 6 in combination with JIMMY.

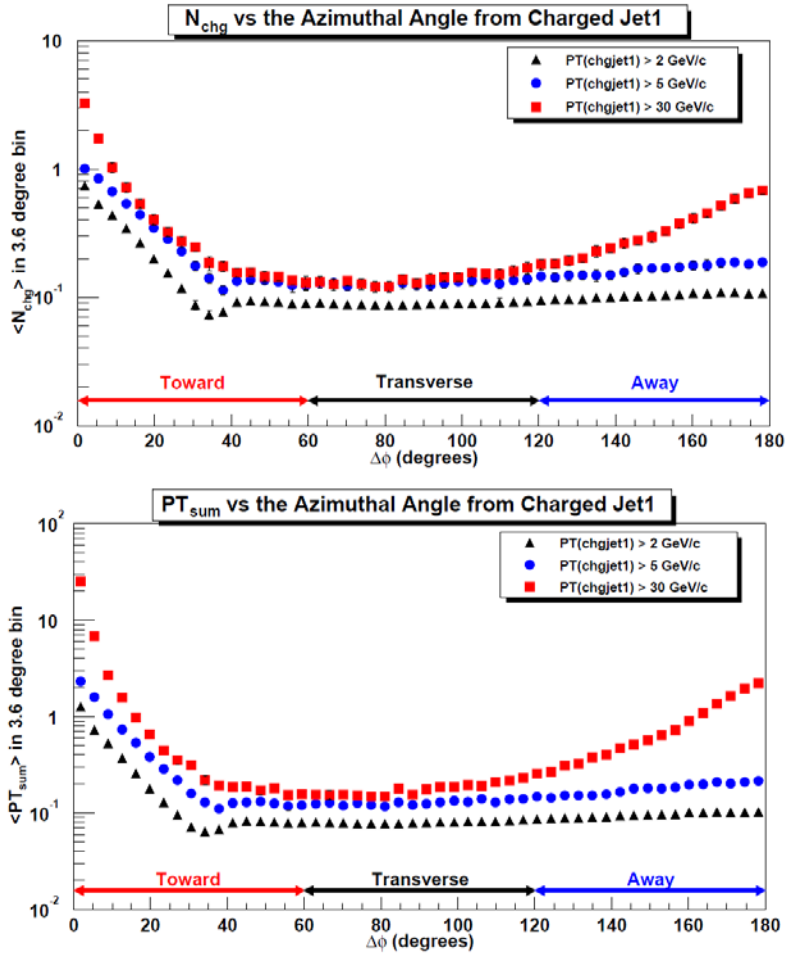


Figure 3.16: *Average charged multiplicity and sum of transverse momenta at* 1.8 *TeV for different transverse momenta of the leading jet. The measured tracks are required to have* $p_T > 0.5$ *GeV and* $|\eta| < 1.0$. *From [97].*

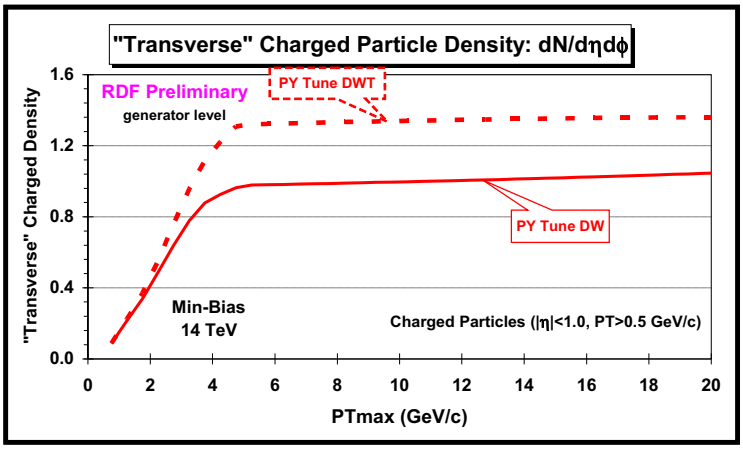


Figure 3.17: *Expected transverse charged density for tunes DW and DWT at the LHC. From [99].*

CMS experiment are presented in [100], with the surprising result, that all available tunes underestimated the charged activity. The new tune CW was introduced thereupon, with a rather extreme MPI scaling, which fails to describe the distributions at 7 TeV [101]. Selected distributions from the 7 TeV study and a comparison with 0.9 TeV are presented in Fig. 3.18. It is visible that neither of the available tunes is able to describe the data points in all details. Especially in the region of low transverse momenta of the leading jet, the Monte Carlo predictions are too low. Pythia 6 tune D6T comes closest to predicting the charged particle multiplicity, but it exaggerates the transverse momentum sum, which is fairly well described by tune DW.

The studies at 7 TeV led to the formulation of the new tunes Z1 and Z2, which in contrast to DW and D6T also feature the p_{T}-ordered parton shower and therefore belong to the "Pythia 6.4" family of tunes. The new tune Z2 is projected to become the new standard Pythia tune in the CMS collaboration for future central Monte Carlo productions, together with tune D6T. Both are built around the CTEQ6L1 parton density function.

3.3 The Jet Area/Median Approach to Measuring the Underlying Event

Recently, a new method to measure the underlying event in hadron collisions has been proposed [2], that follows a completely different path than the traditional approach. In this new method, dubbed the *Jet Area/Median* approach, no explicit exclusion of hard contributions is necessary. It rather relies on the fact, that for example in a hard di-jet event, most of the geometrical space in a particle detector is either completely empty or only populated by soft contributions. Therefore, looking at a measure that takes into account the average activity per unit of area in an event proves to be sensitive to the UE model.

3.3.1 Theoretical Background

As in the traditional approach, an alternative approach to measuring the underlying event must first state what definition of UE it uses. At some point, the concept of multiple parton interactions, despite being successful in practice of Monte Carlo event generation, is questionable as such. The strict separation between a hard interaction, calculated at LO or NLO level and additional soft scatterings ignores the fact, that there is an interference between $2\rightarrow4$ higher order QCD and one hard and one soft interaction. This potential double-counting is a fundamental question in simulating proton collisions, in addition to the obvious colour and momentum conservation for the individual protons that has to hold on parton level. To avoid this ambiguity, the Jet Area/Median approach relies on a more holistic definition of the UE as the traditional approach: It is assumed that at an arbitrarily low p_{T} scale, the geometrical space between the hard partons is completely

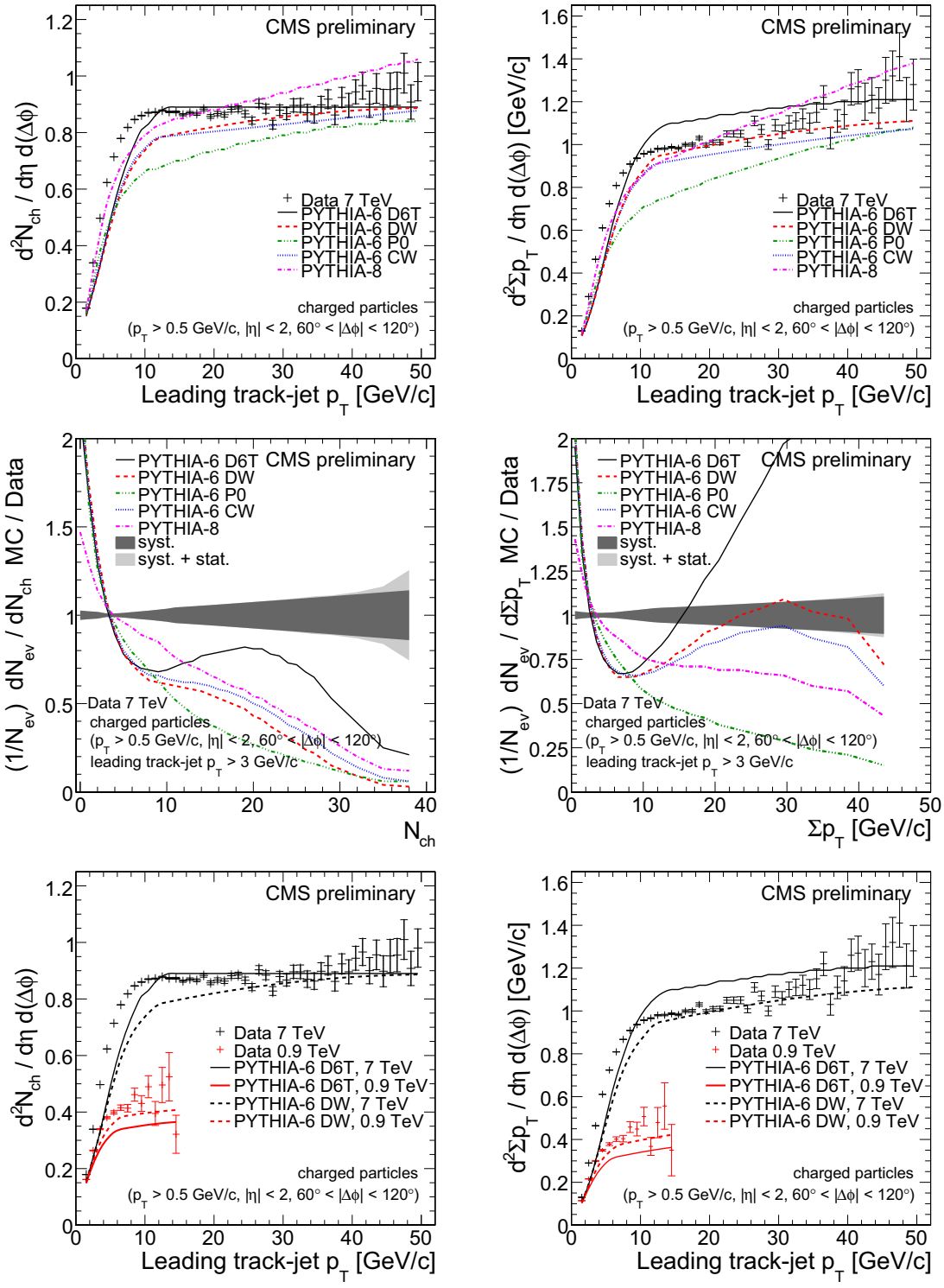


Figure 3.18: *Results from the measurement of the UE with CMS at 7 TeV and comparison to 0.9 TeV from [101]. Only the transverse region of the events is evaluated.*
Upper row: Charged multiplicity and transverse momentum sum in the transverse area for 7 TeV for uncorrected data and different Pythia settings after full detector simulation.
Middle row: Ratios between data points and different Pythia settings.
Lower row: Charged multiplicity and transverse momentum sum comparison between data and MC for 7 and 0.9 TeV.

filled with soft radiation contributions. So the idea is to measure the physical content that fills most of the event and identify this as the underlying event.

3.3.2 Proposed Measurement

Active area jet clustering as introduced in section 1.5.3 is an ideal tool for quantifying an area based measure for physical activity in an event. It allows to evaluate event-by-event the distribution of transverse momentum over jet area of all jets found, including pure ghost jets. A qualitative example for this distribution is given in Fig. 3.19. In [102], the median of this distribution was pro-

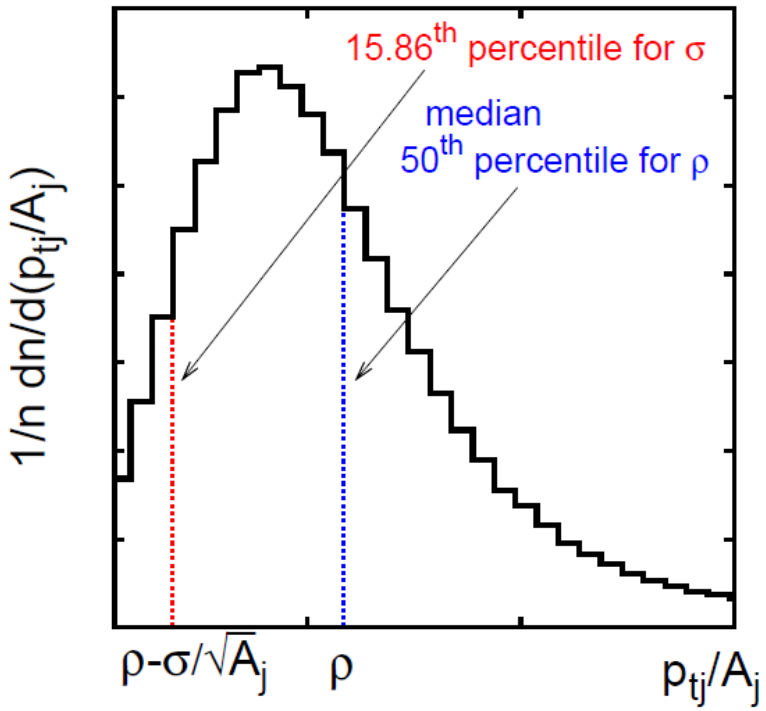


Figure 3.19: *Qualitative picture of jet area over transverse momentum at Monte Carlo level for a single event. Both physical and pure ghost jets are contained in this picture. The median and other quantiles can be derived event-by-event. From [2]*

posed as a measure for uniformly distributed background in an event. This quantity is commonly referred to as ρ:

$$\rho = \underset{j \in \text{jets}}{\text{median}} \left[\left\{ \frac{p_{\text{T}j}}{A_j} \right\} \right] \tag{3.3}$$

The idea behind using the median rather than the average of the distribution is that with this quantity, the analysis is robust against outliers such as hard QCD jets and no additional cuts are necessary to isolate the underlying event. The Jet Area/Median approach therefore is applicable to any event topology without geometrical restrictions. Originally, this method has been proposed

to get a handle on the amount of pile-up activity in high luminosity running [45]. In general, all jet algorithms that allow for active area clustering are suited for this analysis, however the anti-k_T with its fixed cone-like jet shape causes unrepresentative values of p_T/A.
The Monte Carlo study presented in [2] is performed at a center-of-mass energy of 10 TeV and relies on a rather strict event selection to demonstrate the sensitivity of ρ on different Monte Carlo tunes: Two hard jets are required, one with at least 100 GeV and the other with at least 80 GeV of transverse momentum. These leading jets are necessary to identify the presence of a sufficiently hard interaction and are clustered with the anti-k_T algorithm, while the actual calculation of ρ is performed with the Cambridge-Aachen algorithm. This event selection guarantees a fairly crowded event topology. Five different Monte Carlo setups have been used: Pythia 6 tunes S0A, DW and DWT as well as Herwig 6 in combination with JIMMY and, as an unphysical reference, also without JIMMY.

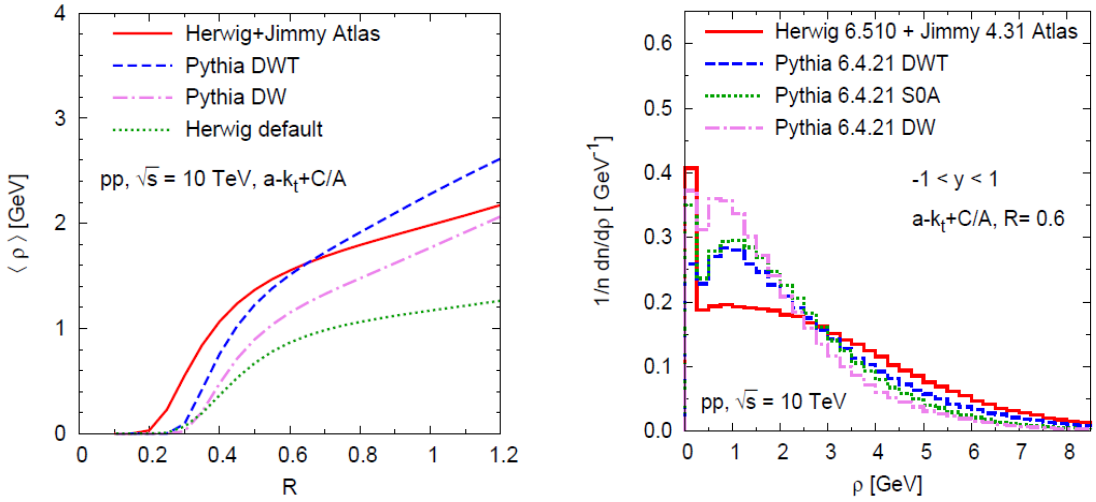


Figure 3.20: *Left: Average value of ρ depending on the size parameter R for the Cambridge-Aachen algorithm. Note the distinct turn-on points for the different tunes.*
Right: ρ distribution for central rapidity event-by-event. Differences between the tunes are clearly visible. From [2]

While the Monte Carlo study presents detailed results on both ρ and related quantities and their behaviour in comparison to the traditional approach to measuring the underlying event, only a number of these conclusions can be summarized here: First, looking at the distribution of the average value of ρ depending on the jet size parameter of the Cambridge-Aachen algorithm shows a distinct turn-on behaviour as can be seen in Fig. 3.20. This value can be interpreted as the jet size parameter above which more than half of the geometrical space of the detector is covered with physical jets. Events in which more than half of the space is covered with ghost jets, that intrinsically have a vanishing transverse momentum, have a median of transverse momentum per jet area that is equal to zero. This is true if the average area of ghost jets is equal to the average area of physical jets, which has been shown to roughly be the case [45]. Obviously, larger jet sizes in an event will lead to a larger value for the average of ρ, as the space filled with ghost jets

will decrease, pushing the value of ρ towards higher values. Another interesting observation is the distribution of ρ at a fixed value of R, also shown in Fig. 3.20, that clearly allows to distinguish between the different Monte Carlo models.
The holistic Jet Area/Median approach to measuring the underlying event proves to have a number of favourable features that makes it an ideal complement for the traditional approach measurement. Any kind of event selection can be applied, with no kinematic cuts necessary whatsoever and the results can be compared with Monte Carlo predictions. With this procedure, the validity of arbitrary tunes can be verified or discarded. In the following, this novel technique will be demonstrated at the first time at a hadron collider with early data recorded with the CMS experiment in 2009 and 2010. Also, the Jet Area/Median technique offers the perspective of subtracting the soft QCD background, both from pile-up and the underlying event, on an event-by-event basis depending on the soft activity in a given event.

Chapter 4

Measurement of the Charged Underlying Event Activity with the CMS Detector at 0.9 and 7 TeV with the Jet Area/Median Approach

Applying the Jet Area/Median approach to measure the charged underlying event activity to actual detector data bears a number of challenges that have not been an issue in the original publication concentrating solely on Monte Carlo information. First and foremost, a suitable data set has to be identified. In this context it is vital that the instantaneous luminosity is rather small so that the measurement is not affected by pile-up. At a collider that obviously aims for a maximization of event rates and thus luminosity, only data from early running stages and commissioning runs are suitable. These early running conditions however have been present for only a short period of time, limiting the available number of events. This has serious implications on the actual event selection, as the rather strict request for the presence of a hard interaction made in the original publication cannot be complied to due to lack of eligible data. Therefore, minimum bias data were used that contain only the restriction that a proton collision has happened at all, characterized by a minimal trigger requirement. This analysis is connected to a number of other early analyses performed with the CMS experiment and for reasons of consistency, such as the traditional UE measurement and the track-jet performance study, a number of analysis related selections have been made in accordance to those measurements. Especially the used datasets have been chosen identically to the traditional approach. This underlines the interpretation of this analysis as a complementary view on the soft contributions to proton-proton interactions.

Furthermore, in the early stages of an experiment, additional efforts have to be made to validate the understanding of tools and detector performance. It is therefore desirable to stay as close as possible to existing or parallel analyses and use as many official tools provided by the experiment software as possible. While track reconstruction is available in centrally hosted CMS data, at least for the 2009 run at 0.9 TeV jet reconstruction from tracks was not available and thus had to be performed in an extra step. In this context, also the active jet area clustering had to be established since this method is applied for the very first time on collider data. Even though track reconstruction is used in many analyses and should therefore be considered a stable and reliable tool, it is nonetheless beneficial to first understand track related observables, especially cross-checking the results with other analysis groups.

4.1 Data Samples

4.1.1 0.9 TeV Data from the 2009 Commissioning Run

While the initial running of the LHC in late 2009 was mainly important as a stepping stone to the new collision energy of 7 TeV in 2010, it also delivered important data at the LHC injection energy of 0.9 TeV. Especially for observables that show a distinct scaling behaviour over energy such as charged multiplicities [103] and of course the traditional underlying event measurement [100] valuable results have been produced.
Due to the low luminosity and reduced collision energy, only minimal trigger requirements had to be met. The technical level-1 trigger bits that were used to identify the presence of an interaction are shown in Table 4.1. Only two systems are contributing: The BPTX (Beam Pickup Timing for the eXperiment), that guarantees the presence of two colliding bunches in the experiment and the BSC (Beam Scintillation Counters) which are sensitive to minimal activity in the detector.

Table 4.1: *The L1 triggers used for the 2009 commissioning data.*

Trigger Bit	Subsystem	Short Description
0	BPTX	both beams passing the IP
40	BSC	1 or more hits on the $+z$ and the $-z$ side
41	BSC	2 or more hits on the $+z$ and the $-z$ side
36	BSC	inner halo detection for beam 2
37	BSC	outer halo detection for beam 2
38	BSC	inner halo detection for beam 1
39	BSC	outer halo detection for beam 1

The logical expression necessary for an event to pass the trigger configuration is
$0 \wedge (40 \vee 41) \wedge \neg(36 \vee 37 \vee 38 \vee 39)$.
Bits 40 and 41 are true if one or two hits are measured in both hemispheres of the detector. Bits 36 to 39 are a veto against ring-like activity patterns that are characteristic for parasitic events arising from collisions of protons with residual gas molecules in the beam pipe or the beam pipe material itself.
With improvements in detector understanding and progress in the reconstruction software, the 2009 collision data were re-reconstructed several times. To be consistent with the traditional UE measurement, the dataset used here is the reprocessing of 19th December with CMSSW version 3_3_6. This reprocessed dataset contains about 20 million events. The trigger requirements already reduce the number of events by about 98%. Furthermore, only runs and luminosity sections identified as valid for analyses by central CMS data quality management have been used. A table containing the actual runs and luminosity sections can be found in App. B. In addition to the trigger selection, supplemental requirements have been imposed on the presence of a reconstructed vertex and the number of tracks used to fit this vertex. The vertex criteria discard an additional 12% of the events passing the trigger selection. The total event number finally entering the analysis after trigger and vertex selection is $225,447$ which is consistent with [100].

Table 4.2: *Numbers of events satisfying the criteria of the different selection steps together with absolute and relative event fractions in the 2009 commissioning data.*

Selection Criterion	Abs. Event Fraction	Rel. Event Fraction	Total No. of Events
recorded	100%	100%	$19,681,382$
BSC coincidence	3.1%	3.1%	$603,730$
beam halo rejection	2.7%	87.0%	$522,465$
BPTX coincidence	2.3%	86.7%	$453,409$
good run selection	1.3%	56.1%	$254,270$
1 primary vertex	1.2%	94.0%	$238,248$
$15\,\mathrm{cm}$ vertex z window	1.2%	99.9%	$238,188$
≥ 3 tracks fitted to vertex	1.1%	94.6%	$225,447$

4.1.2 0.9 TeV Monte Carlo Samples

Several different Monte Carlo samples have been produced to be compared with the 0.9 TeV data. The detector description used for the simulation was based on a realistic estimation of misalignment of the detector components, non-functional readout channels and noise effects. Pythia 6 tunes with both the old and the new model for multiple parton interactions are available as well as the default tune of Pythia 8. Both event generators are run in *MinBias mode* which includes diffractive events and events generated with QCD matrix elements with a minimal momentum transfer. The

enabled Pythia 6 parameters in this production can be found in Table 4.3 while the corresponding settings for Pythia 8 are shown in Table 4.4.

Table 4.3: *Pythia 6 processes in minimum bias Monte Carlo samples with corresponding parameters.*

Parameter	Process
MSUB(11)	$q_i q_j \to q_i q_j$
MSUB(12)	$q_i \bar{q}_i \to q_k \bar{q}_k$
MSUB(13)	$q_i \bar{q}_i \to gg$
MSUB(28)	$q_i g \to q_i g$
MSUB(53)	$gg \to q_k \bar{q}_k$
MSUB(68)	$gg \to gg$
MSUB(92)	single diffraction $AB \to XB$
MSUB(93)	single diffraction $AB \to AX$
MSUB(94)	double diffraction
MSUB(95)	low p_{T} production

Table 4.4: *Pythia 8 processes in minimum bias Monte Carlo samples.*

Parameter
'SoftQCD:minBias = on'
'SoftQCD:singleDiffractive = on'
'SoftQCD:doubleDiffractive = on'

These processes are produced according to their corresponding cross-sections. While the diffractive processes are treated differently, the partonic matrix elements are required to have a minimum momentum transfer of 2.3 GeV. As tune D6T was considered the best available tune at LHC start-up, it acts as a reference tune for many purposes in the analysis, such as the investigation of systematic uncertainties. The event selection on MC was performed identically to detector data, with the according trigger and vertex selection on reconstructed objects and trigger simulation. Per tune, $300,000$ events were processed, in order to reach a statistical precision comparable to data.

4.1.3 7 TeV Data and Monte Carlo

Starting in spring 2010, the first major physics program of the LHC operates at a center-of-mass energy of 7 TeV. However, due to the rapidly increasing luminosity, only data from early runs can be used for analyses that require strictly one collision per bunch crossing. In the first weeks of running, dedicated *Good Collision* data sets were centrally produced, that contain minimum bias triggered collisions from approved running periods with fully operational detector. With the ever

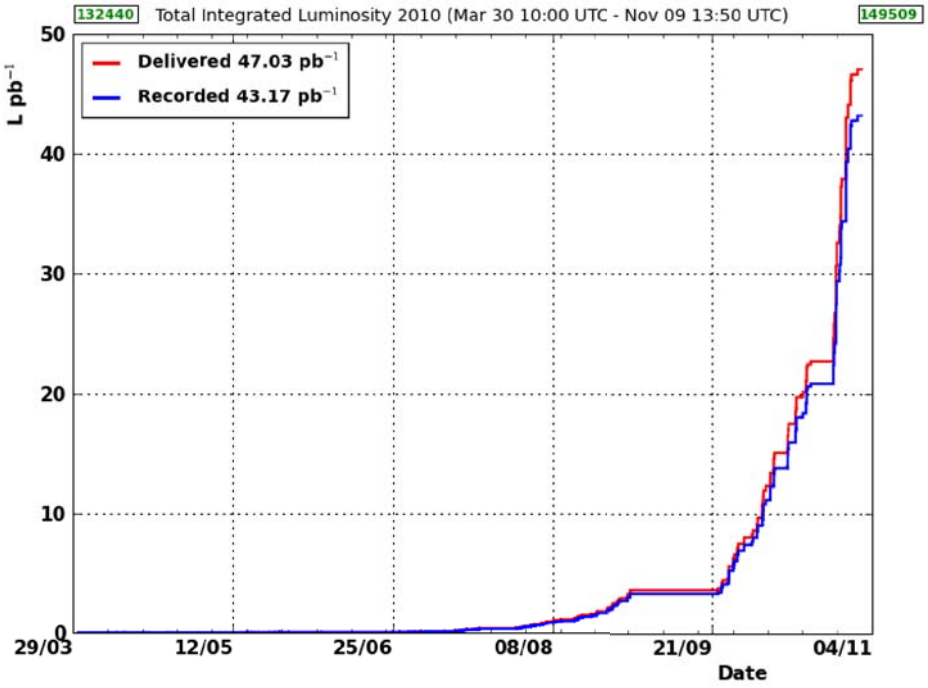


Figure 4.1: *Integrated luminosity over time delivered by the LHC and recorded with the CMS experiment in the 2010 physics program.*

growing number of recorded events, this practice has been discontinued, yet for analyses relying on low luminosity minimum bias data, the Good Collision skims are an ideal starting point. The technical trigger bit criteria are largely identical to the selection at 0.9 TeV, yet additional bits are involved that reject *Beam Splash Events*, which were recorded for testing purposes. Also, a *Beam Scraping* filter is applied that rejects events with a large number of tracks, yet only a small fraction of high purity ones as defined in [104]. These events are caused by a beam halo proton interacting with the tracker material. All further event selection criteria have again been synchronized with the traditional approach analysis [101], containing identical vertex requirements as for 0.9 TeV.
The development of the total integrated luminosity delivered by the LHC and recorded by CMS is shown in Fig. 4.1. Only data recorded until May 6 have used for this analysis. As has been shown for the 2009 0.9 TeV data, this analysis requires only a limited number of events to produce statistically relevant results. Therefore the sufficiently large number of 3 million events was chosen, corresponding to roughly ten times the number of events compared to 0.9 TeV. The full Good Collision dataset contains about 60 million events, details on the data set such as the technicalities of the trigger selection and the DBS database entry can be found in App. B. Corresponding to the recorded data and produced with the same software version, a number of different Monte Carlo datasets were produced with different UE tunes. Details can be found in App. B, while a detailed discussion of tunes is found in App. A. One million events per Monte Carlo dataset were processed, in order to stay competitive with the statistical precision of the increased amount of data

events. Due to limitations of the available computing resources, the private production of a data set with tune Z1 was reduced to $500,000$ events.

4.2 Track Selection

The track selection in this analysis is, like the event selection, closely connected other analyses, yet with a relatively low transverse momentum cut of the tracks of 300 MeV. Additional information on the performance of the CMS tracking system in the 2009 commissioning run as well as the track quality definitions can be found in [104]. The applied selection criteria are in detail:

- High purity track quality, the quantities entering the quality definitions include the number of layers with a hit on the track and the χ^2 of the track among others.
- Transverse momentum $p_{\mathrm{T}} > 0.3$ GeV
- Pseudorapidity $|\eta| < 2.3$
- Transverse impact parameter significance $d_{xy}/\sigma_{d_{xy}} < 5$
- Longitudinal impact parameter significance $d_z/\sigma_{d_z} < 5$
- Relative track p_{T} uncertainty $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} < 5\%$

The track selection for 7 TeV data has been chosen slightly different compared to the 0.9 TeV data, as jets reconstructed from tracks have since become part of the official reconstruction chain, where they are called *TrackJets*. In accordance with [64], the selection criteria were chosen as:

- High purity track quality
- Transverse momentum $p_{\mathrm{T}} > 0.3$ GeV
- Pseudorapidity $|\eta| < 2.3$
- Transverse impact parameter $d_{xy} < 0.2\ cm$
- Longitudinal impact parameter $d_z < 1\ cm$
- Relative track p_{T} uncertainty $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} * max(1, \chi^2/ndof) < 0.2$

Noteworthy changes are the setting of the impact parameters to absolute values. These values have been shown in [64] to be far less restrictive than the relative significance cuts used in the 0.9 TeV analysis.

Obviously, in Monte Carlo events, on fully reconstructed level, the same cuts on tracks have to be applied in order to be able to compare predictions to data. It is however also desirable to have at hand a consistent collection of Monte Carlo level generator particles to investigate the influence of the detector simulation on observables. As the tracking detector can only observe charged particles, neutral Monte Carlo particles are excluded as well as unstable ones. Stability in this context means an average proper lifetime of τ such that $c\tau > 10$ mm. The kinematical cuts are upheld to select particles that are expected to be reconstructed as tracks:

- Charged
- Transverse momentum $p_{\mathrm{T}} > 0.3\,\mathrm{GeV}$
- Pseudorapidity $|\eta| < 2.3$

Figure 4.2 shows the fraction of all and of the charged particles exceeding a certain value of transverse momentum. Only 60% of all particles fulfill $p_{\mathrm{T}} > 300$ MeV and with the charge requirement, this number reduces to 35%. A higher minimum transverse momentum cut of 500 MeV would have further reduced this number to about 20% of all particles.

Comparison of Track and Charged Particle Based Distributions

Figures 4.3 and 4.4 show the basic kinematic distributions of tracks and charged generator particles. The track multiplicity of tracks is rather low, with an average of five at 0.9 TeV. The zero suppression is due to the applied trigger and event selection requirements. As expected, the number of tracks is generally higher at 7 TeV and also tracks with higher transverse momenta are found. A number of additional observations can be made: For both center-of-mass energies, the agreement between charged generator particles and tracks is excellent concerning multiplicity and transverse momentum. In the pseudo-rapidity distributions, geometrical features of the detector are clearly observable. The corresponding plots for ϕ can be found in App. C, also revealing a structure due to detector effects.
At 0.9 TeV, two tunes are able to describe the track multiplicity properly: CW and Z1. There is however an interesting difference in the pseudo-rapidity distributions of the two. CW, as all tunes with the old multiple parton interaction model, has a basically flat behaviour with a slight dip in the central region. Z1 and the other tunes with the new model show a distinct drop-off towards the edges that is also present in data. The transverse momentum spectra are comparable for all tunes.
None of the tunes is able to reproduce the track multiplicity at 7 TeV even though Z1 comes closest. Especially at higher transverse momentum of above 2 GeV, a clear excess is observed in data. The differences in the pseudo-rapidity observed at 0.9 TeV vanish at higher energy and all tunes as well as the track distribution in data are essentially flat.
This first glimpse at the multiplicities and kinematical distributions already underlines what has

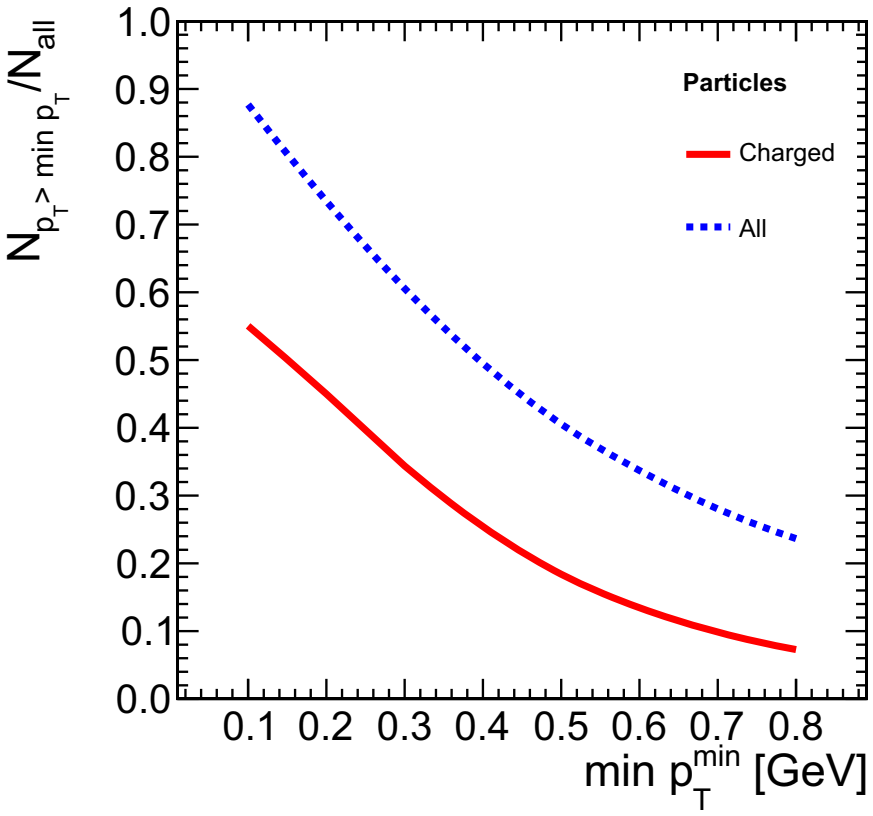


Figure 4.2: *Fractions of all and of charged Monte Carlo particles exceeding a minimal transverse momentum. Produced with Pythia tune D6T at a center-of-mass energy of* 0.9 *TeV.*

been shown in other early CMS analyses [100, 105]: The charged multiplicity is generally underestimated by tunes derived from Tevatron data. The new tune Z1 provides a remedy in the form of an increased activity at 0.9 TeV and performs reasonably well at 7 TeV, yet fails to produce enough tracks at larger transverse momenta. Some interesting remarks can be made on the scaling behaviour of the different tunes: While CW with its steeper scaling factor for multiple parton interactions is performing well at low center-of-mass energy, it is in the range of D6T at 7 TeV concerning multiplicities.

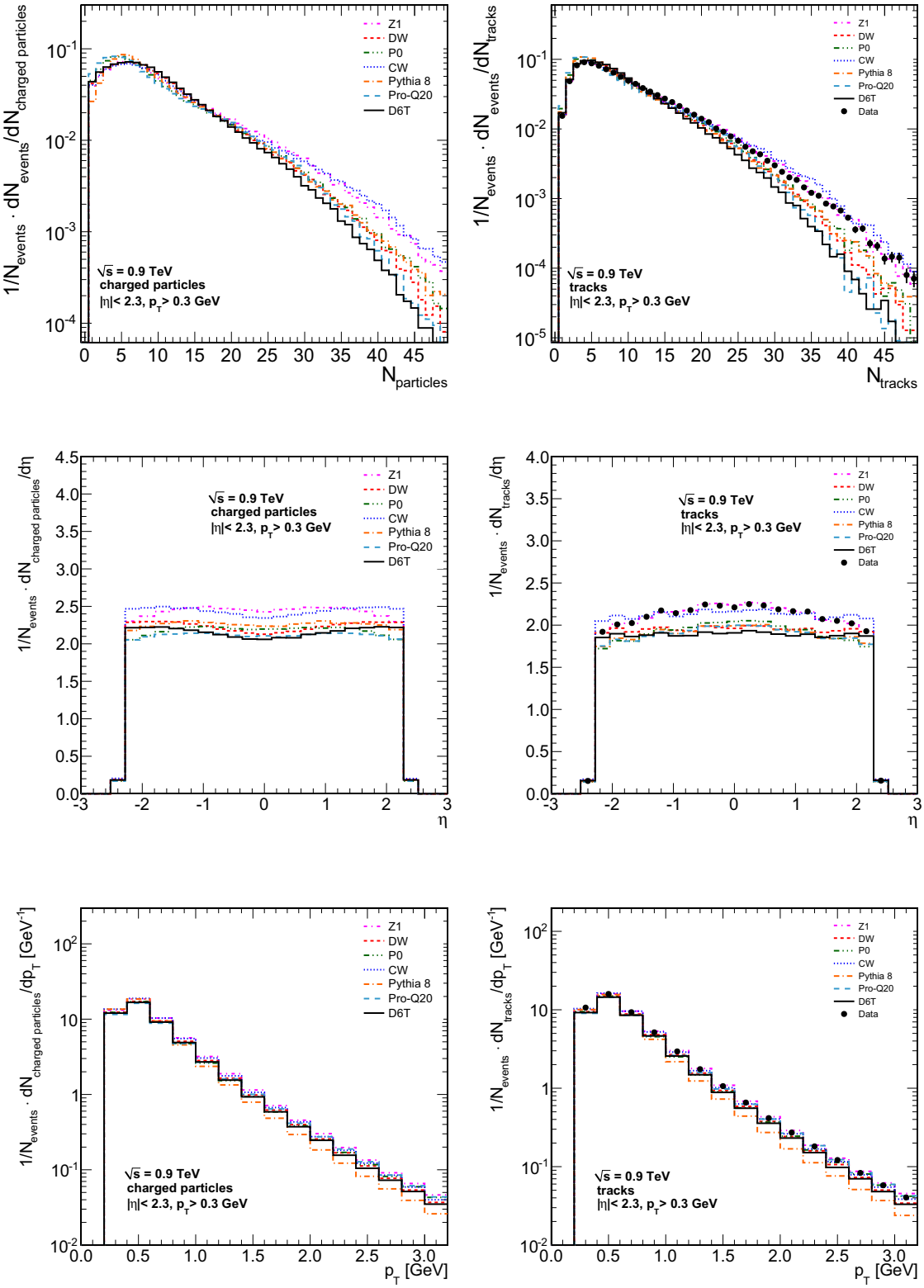


Figure 4.3: *Charged generator particle (left column) and track (right column) distributions for* 0.9 *TeV. Shown are from top to bottom the track multiplicity per event, the pseudo-rapidity distribution and the transverse momentum distribution. The latter two are normalized to the number of events. The same distributions with linear* y*-axes for multiplicity and transverse momentum can be found in*ẽfApp:plots.

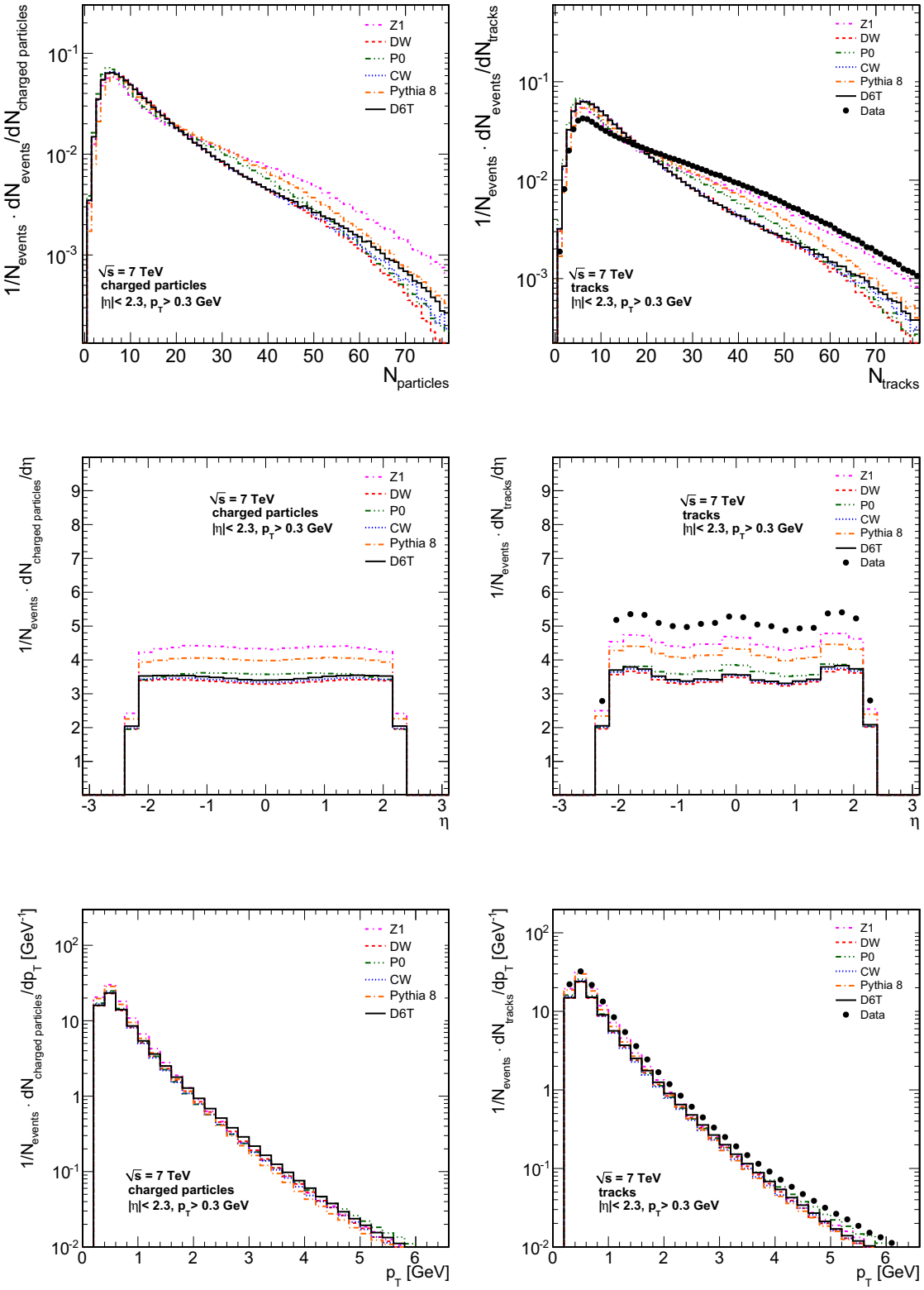


Figure 4.4: *Charged generator particle (left column) and track (right column) distributions for 7 TeV. Shown are from top to bottom the track multiplicity per event, the pseudo-rapidity distribution and the transverse momentum distribution. The latter two are normalized to the number of events. The same distributions with linear y-axes for multiplicity and transverse momentum can be found in*ẽfApp:plots.

4.3 Jet Reconstruction

For an analysis of the diffuse background and the soft QCD activity with the Jet Area/Median approach, a jet algorithm with a variable jet geometry is necessary. As discussed in chapter 1.5, both the inclusive k_T and the Cambridge-Aachen algorithms are suitable. The CMS collaboration has chosen several standard jet algorithms that are present in the standard reconstruction of recorded events. As the Cambridge-Aachen algorithm is not one of them, the choice as reference algorithm is k_T with a resolution parameter of $R = 0.6$. The settings for the active area clustering procedure are following the recommendations from [42], with a *Ghost Area* of 0.01, which corresponds to about 10.000 uniformly distributed ghost particles overlaying the event up to a pseudorapidity of 6.0. Even though this is well beyond the tracker reach, it does not affect the measurement and therefore the standard value was left unaltered. For comparison, other values of R have been studied and results will be presented also for the R-dependence of observables in the following.
Two collections of input particles are used for jet clustering and therefore two types of jets are available for study. The jets clustered from reconstructed tracks are referred to as *track jets* and are available in detector data and Monte Carlo after full simulation. On the other hand *charged particle jets* are clustered from stable charged Monte Carlo particles as defined above. No additional kinematical cuts are imposed on the clustered jets. Of course, the transverse momentum requirement of the tracks and charged Monte Carlo particles is directly reflected in the jet p_T spectrum, with jets only above 300 MeV. To avoid boundary effects in the area determination, a stricter cut in pseudorapidity is imposed on jets compared to tracks. The direction of a jet refers to the jet axis and thus with the choice of a default jet resolution parameter of $R = 0.6$, the jets are required to have a pseudorapidity of $|\eta| < 1.8$. This measure prevents the shapes of the jets to be distorted or unnaturally cut off at the edges of the tracker acceptance.

Comparison of Jet Based Distributions

Due to the rather low multiplicity of tracks in the investigated events, it is expected, that the basic kinematic properties of tracks or charged generator particles and the corresponding jets are similar. In Fig. 4.5 and 4.6 it can be seen, that the vast majority of jets consist of only one or very few tracks or charged generator particles respectively. Even though this picture of a jet may be un-intuitive in the light of considering a jet a spray of collimated particles, it is still algorithmically valid and well defined. Also the jet areas can be computed and they still contain valuable information on the event geometry and distances between clusters of particles or tracks. The multiplicity distributions for jets reveal more of a plateau-like structure with a steep drop-off especially at 7 TeV. This behaviour is caused by a saturation of the available geometrical range with jets of a certain average area.

The general behaviour of the multiplicity distribution reflects the previously discussed track and particle multiplicities of the tunes: At 0.9 TeV both CW and Z1 are able to reproduce the jet multiplicity reasonably well, while at 7 TeV all tunes underestimate this observable, with Z1 coming

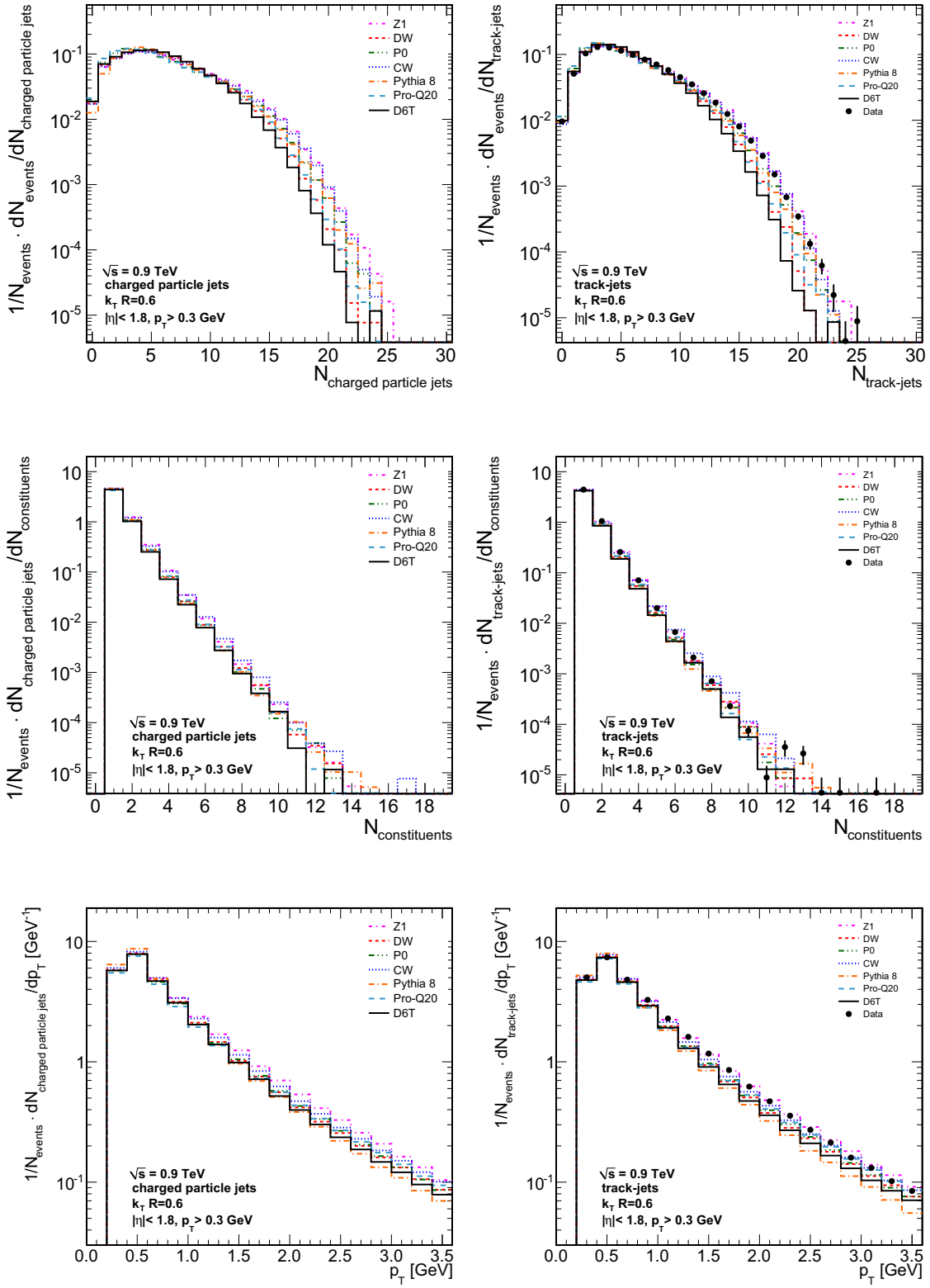


Figure 4.5: *Charged generator particle jets (left column) and track-jets (right column) distributions for* 0.9 *TeV. The jets are clustered with the inclusive* k_T *algorithm with a resolution parameter of* $R = 0.6$*. Shown are from top to bottom the jet multiplicity per event, the constituents multiplicity per jet and the transverse momentum distribution. The latter two are normalized to the number of events.*

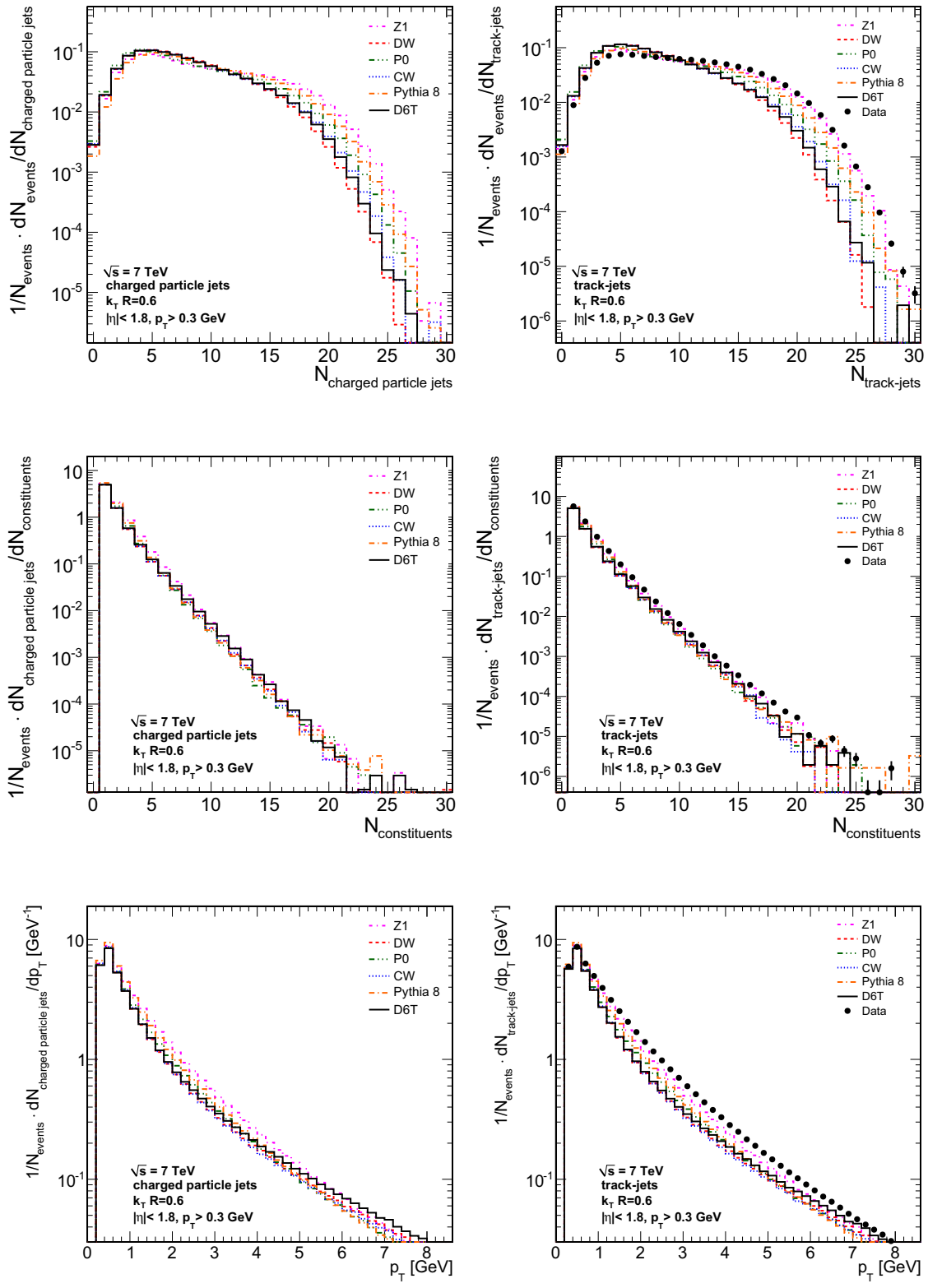


Figure 4.6: *Charged generator particle jets (left column) and track-jets (right column) distributions for 7 TeV. The jets are clustered with the inclusive k_T algorithm with a resolution parameter of $R = 0.6$. Shown are from top to bottom the jet multiplicity per event, the constituent multiplicity per jet and the transverse momentum distribution. The latter two are normalized to the number of events.*

closest. Differences among the tunes are also clearly visible. At the higher energy, especially in the intermediate transverse momentum range, track-jets in data are drastically exceeding all Monte Carlo tunes. For highest transverse momenta of more than 5 GeV, the $\mathrm{p_T}$ distributions of data and tune D6T are approaching each other and also the data points, an indication that in this region the hard production mechanisms are starting to dominate and the soft models are less influential. The corresponding plots for the pseudo-rapidity can be found in App. C.

Jet Areas

The concept of jet areas, has not yet been studied in detail. Jet area distributions for both jet types and both center-of-mass energies are shown in Fig. 4.7. Only jets containing at least one physical particle are shown, which implies that jets consisting purely of ghosts are explicitly excluded. The similarity between all of them is striking. Due to the increased number of events entering at 7 TeV, the reach towards larger areas is increased. The average jet areas are close to $\pi R^2 = 1.13$, which might naively be expected for perfect cone jets. These distributions prove that the jet area measure still behaves reasonably even for low multiplicity events and jets with a very small number of constituents. The fact that all tunes and data exhibit very similar jet areas furthermore indicates that the active area clustering is insensitive to the details of the MPI and hadronization models, which differ among the tunes. Figure 4.8 furthermore demonstrates the behaviour of the jet area distributions for different resolution parameters of the k_T algorithm, namely a relatively wide peak near values of the expected jet size of $A = \pi R^2$.

It has been shown in [45] that the average area of ghost jets is expected to be smaller than the one of jets containing at least one particle. As the final observable of interest of the proposed study is the median of the jet transverse momenta divided by their area for all jets in an event including ghost jets ρ, an important quantity is the event occupancy C defined as:

$$C = \frac{\sum\limits_{j \in \text{physical jets}} A_j}{A_\mathrm{tot}} \tag{4.1}$$

The total available area A_tot acts purely as a normalization factor. In this analysis it is chosen to be 8π, spanning from $-2.0 < \eta < 2.0$ and $0 < \phi < 2\pi$. This value takes into account the pseudo-rapidity cut of $|\eta| < 1.8$ for the jet axis plus some additional space for the jets to spread into. In very rare cases of large activity and many jets pointing towards the edge of the available pseudo-rapidity, the value of C can exceed 1.0 with this definition.
If C is smaller than 0.5 for a given event, ρ is identical 0. Figure 4.9 shows that this is in fact the case for a majority of the events for both jet types and center-of-mass-energies due to the relatively low activity in the selected events. It is unlikely that with the increasing luminosity of the LHC a low pile-up sample with sufficient events for a stricter event selection will be possible for a

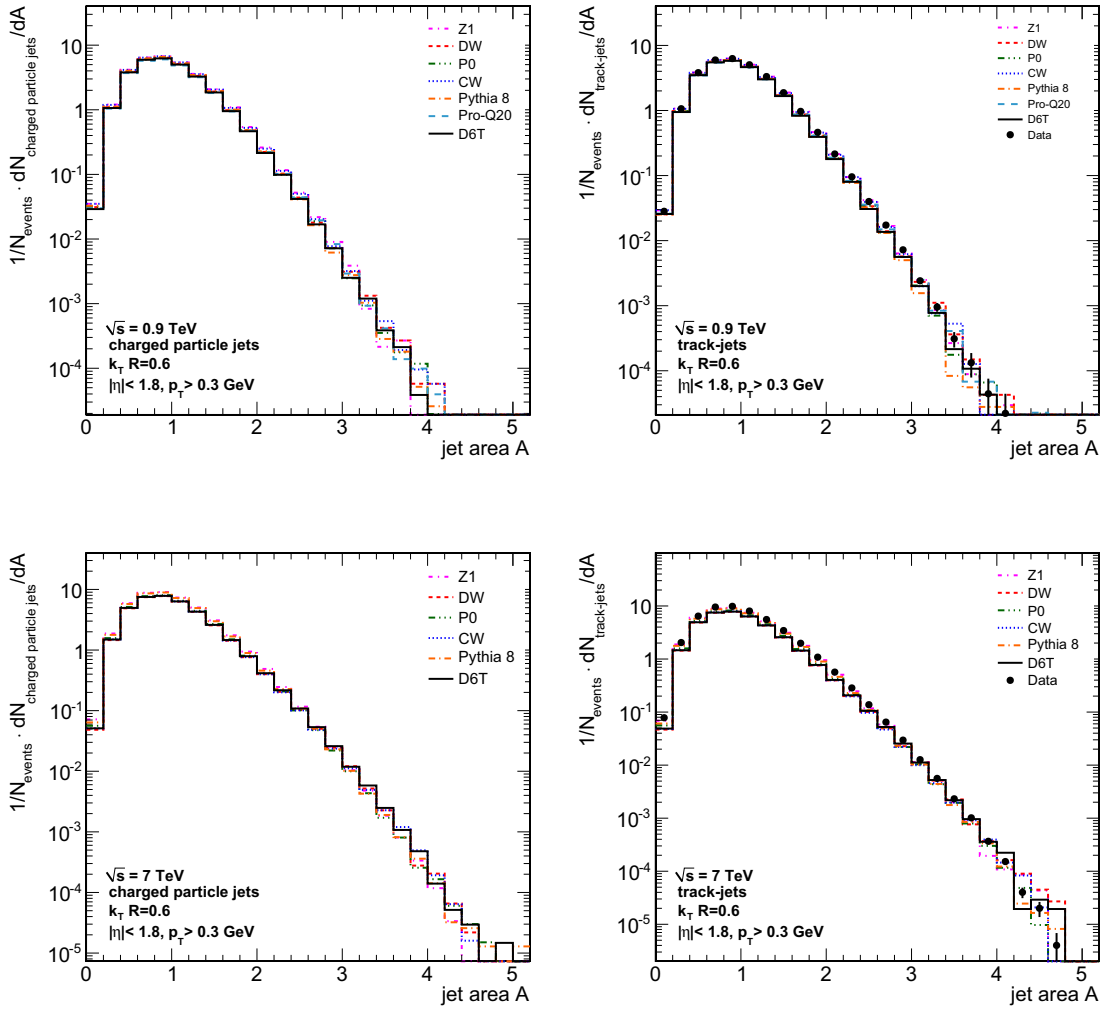


Figure 4.7: *Jet area distributions for* 0.9 *and* 7 *TeV for both jet types. Note the different ranges of the y-axis due to the different number of events entering the distributions.*

study comparable with the hard di-jet topology examined in the original publication. Thus, it is convenient to introduce an adjusted variable ρ' defined as

$$\rho' = \underset{j \in \text{physical jets}}{\text{median}} \left[\left\{ \frac{p_{\text{T}j}}{A_j} \right\} \right] \cdot C \tag{4.2}$$

that explicitly excludes ghost jets but takes into account the event occupancy. The modification of the observable incorporates the "emptiness" of the event by scaling down events with a small number of jets in the ρ' picture. A good example to understand this new variable is the one or two jet case: Assuming a single jet with an area of 1.0, which is close to the average value for k_T jets with $R = 0.6$ and a transverse momentum of 0.5 GeV, the median would be exactly the value of

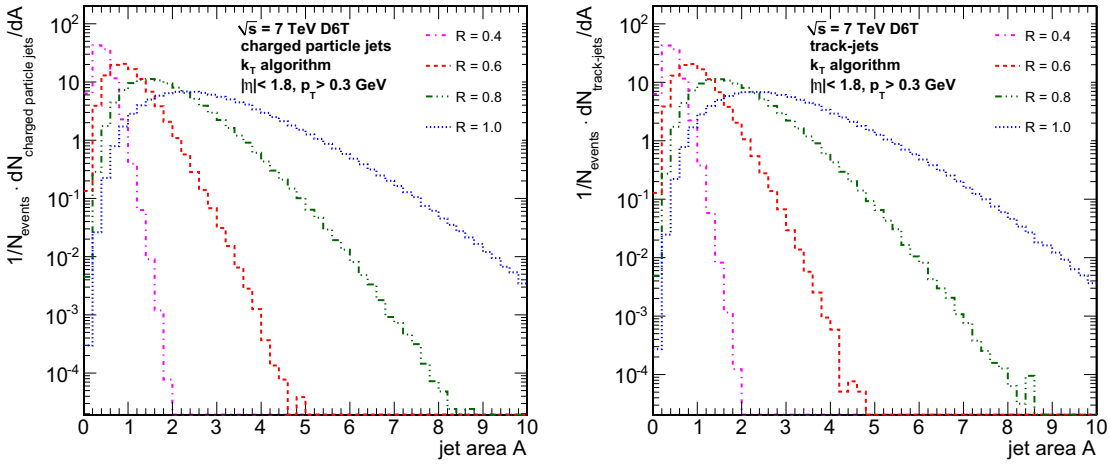


Figure 4.8: *Jet area distributions for different resolution parameters of the k_T algorithm at 7 TeV for charged generator particle jets (left) and track-jets (right). Pythia 6 tune D6T is shown in both plots.*

$\frac{p_T}{A_i}$ of the lone jet i times the event occupancy$\cdot\frac{A_i}{A_{\text{tot}}}$: $\rho' = \frac{0.5\text{ GeV}}{8\pi} \approx 0.02$ GeV. If the event contains a second jet with identical transverse momentum and area, the median remains unchanged, yet the occupancy is doubled, resulting in twice the value of $\rho' \approx 0.04$ GeV. This basic example shows that ρ' is a measure for the overall activity in the event, with an increased event occupancy pushing ρ' towards higher values.

4.4 Sensitivity

Even though the sensitivity of the original ρ variable has been demonstrated with Monte Carlo models, this does not automatically hold for the adjusted ρ'. In Fig. 4.10 and 4.11 the ρ' distributions for charged generator jets are given for 0.9 and 7 TeV respectively as well as the ratios with respect to reference tune D6T of Pythia 6. The ratio plots contain a grey band representing the statistical uncertainty from the number of Monte Carlo events.
Tune D6T is the reference as chosen by CMS. The distributions reveal substantial differences between the tunes: At 0.9 TeV, D6T has an excess of events with $\rho' \approx 0.2$ GeV while all other tunes contain more events than D6T at higher values and thus with increased activity. CW and Z1 even produce twice as many events at the highest $\rho' > 0.8$. This behaviour is expected as these tunes have especially been produced to account for the surprisingly high activity in early CMS measurements.

The picture changes slightly at 7 TeV. First of all, the reach in ρ' is higher, due to the larger number of events and the overall increased activity at higher energies. Also, passing Tevatron energy, the turnaround in multiple parton interaction scaling is visible. While CW with its extremely steep

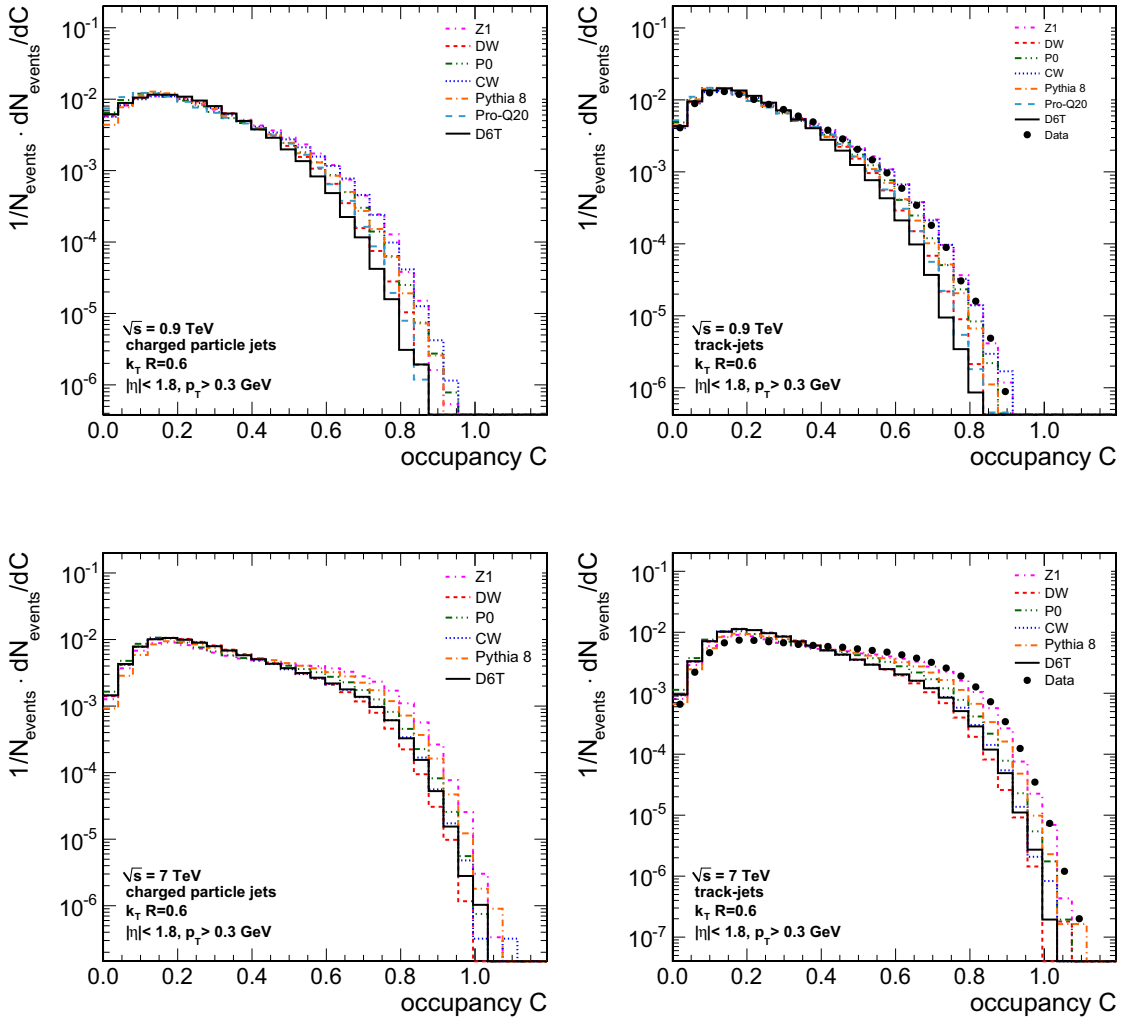


Figure 4.9: *Event occupancies for* 0.9 *and* 7 *TeV.*

scaling predicts much more activity than D6T at 0.9 TeV, it drops off significantly at 7 TeV. P0 and Pythia 8 are very similar for both center-of-mass energies as expected, since the employed Pythia 8 tune is closely connected to the models and values of P0. A very interesting feature is the fact that for very high values of ρ', D6T predicts the highest number of events, even matching Z1, which dominates the intermediate ρ' region.

The behaviour of the different tunes at charged generator level concerning ρ' can also be examined in the light of different jet size parameters. The original proposition did not contain a definite recommendation on this issue, it only stated that the size parameter should not be chosen too small, in order to avoid the short range domain of fragmentation effects. The ρ' variable is expected to be hardly sensitive at all to the jet size parameter since the areas of the jets contribute both in the ratio of transverse momentum and area and also in the event occupancy linearly. These two factors

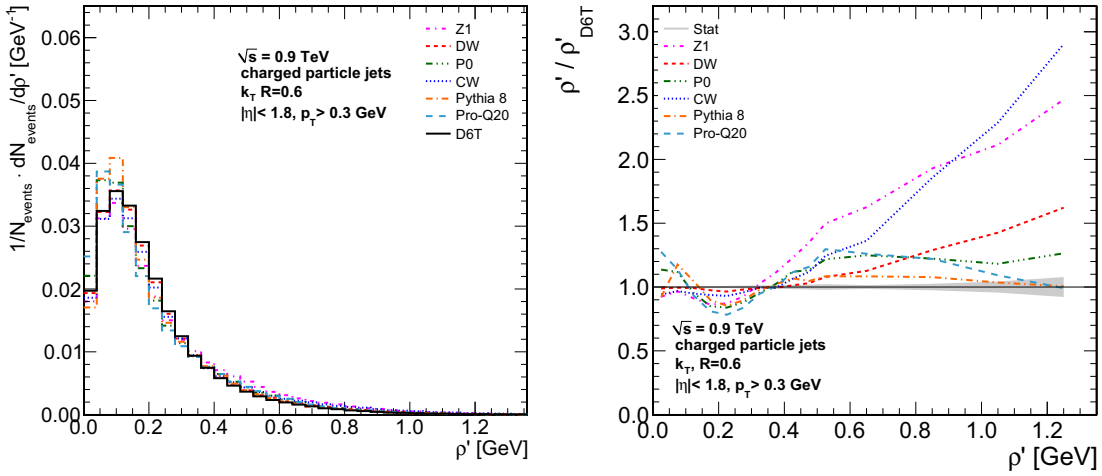


Figure 4.10: *ρ' for charged generator jets at 0.9 TeV and ratios of ρ' distributions with respect to Pythia 6 tune D6T.*

should in principle cancel out as long as, with different size parameters, the same jets are clustered. With many jets containing only one constituent, this should frequently be the case. Figure 4.12 shows the distribution of the average values of the ρ' distribution over the R parameter of the $\mathrm{k_T}$ algorithm at 7 TeV center-of-mass energy. It can be seen, that the prediction actually holds, with the distribution showing a wide plateau for $0.2 < R < 1.2$. Only for very small R a drop-off is observable, proving that $\langle\rho'\rangle$ is robust against the choice of the jet size parameter just as has been proven for the average of $\langle\rho\rangle$.

4.5 Systematic Uncertainties

Even though the sensitivity of ρ' on the underlying event tune has been demonstrated on Monte Carlo particle level, it has to be made sure that systematic uncertainties are not exceeding the differences among the tunes and data. As only track-jets are used, it is not expected that resolution effects are as pronounced as for calorimetric jets, for which explicit unfolding procedures are necessary.
The following sources of systematic uncertainties have been examined at 0.9 TeV:

- Tracker material budget
- Tracker alignment
- Non-operational tracker channels
- Trigger efficiency

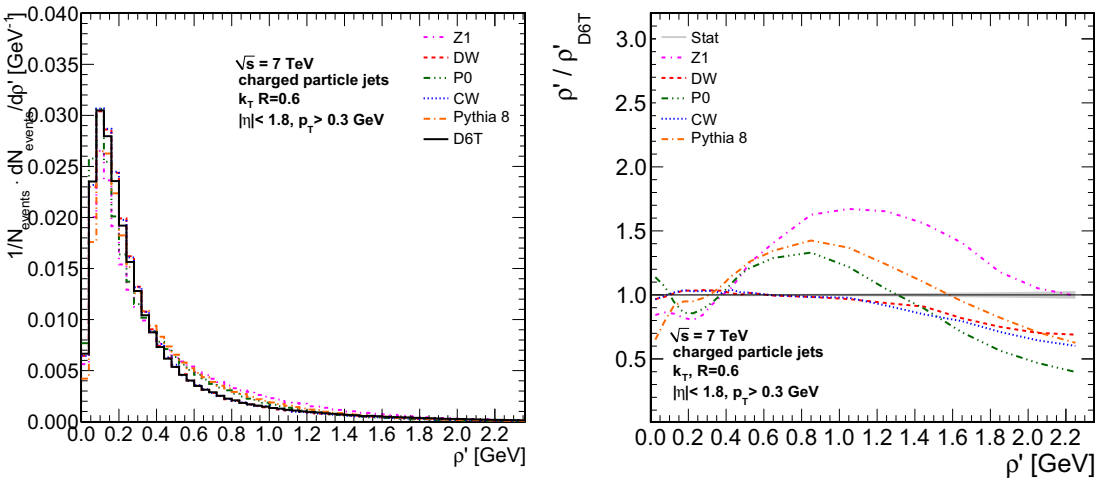


Figure 4.11: *ρ' for charged generator jets at 7 TeV and ratios of ρ' distributions with respect to Pythia 6 tune D6T.*

- Vertex reconstruction
- Track reconstruction efficiency and fake rate
- Variations of the track selection
- Transverse momentum resolution
- Track-jet response

The general procedure to evaluate the impact of these effects has been closely matched to the traditional UE analysis. First of all the ratio of the ρ' distribution on track-jet level between data and the reference Monte Carlo tune D6T is produced. Subsequently, an additional Monte Carlo dataset, also produced with tune D6T yet reflecting one of the aforementioned effects is divided by the distribution from data. Then, the difference between the two ratios is evaluated. In cases, where no ρ' dependency is visible, a straight line fit was performed leading to a constant contribution to the systematic uncertainty. If this constant value is compatible with zero within statistical precision, the statistical uncertainty was added to the absolute value in order to retrieve a conservative estimation of this effect.
For the effects that show a ρ' dependence, a suitable parameterization of the progression was fitted and evaluated bin-by-bin. In case of upwards and downwards variations the average of the absolute values is used as the uncertainty estimation. Finally, all effects were added in quadrature, following the principles of error propagation. For the study at 7 TeV, with the increased Monte Carlo event numbers at hand, the statistical effects were estimated to be reduced compared to 0.9 TeV. The effects found to be compatible with zero at the lower energy,which will be presented in the following, were not considered.

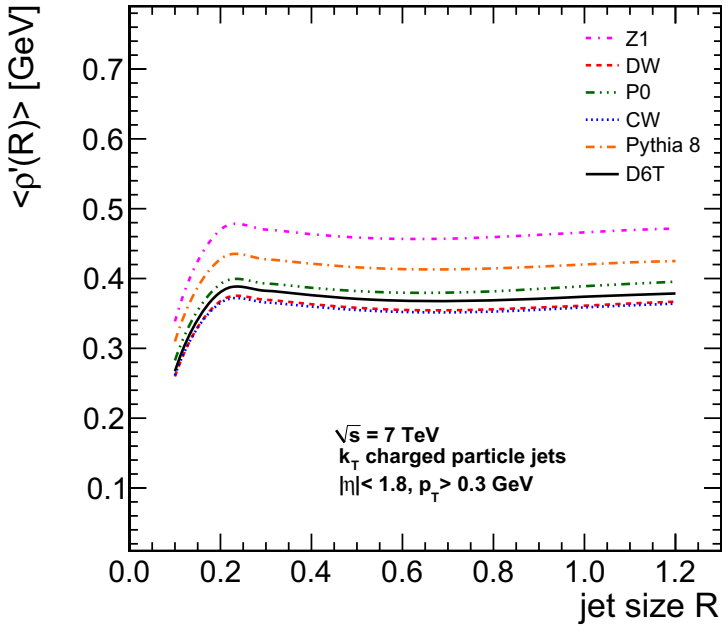


Figure 4.12: *Average of the ρ' distributions over jet size parameter for all employed Pythia tunes at 7 TeV.*

4.5.1 Tracker Material Budget

The exact material budget of the tracker shown in Fig. 2.3 is known only to a certain degree. However it is an important quantity in the simulation of the passage of charged particles through the tracker and influences the measurement of the particle momenta. To study the implications of a wrongly assumed material budget, apart from the unaltered Monte Carlo, two additional datasets were produced, with a $\pm 20\%$ modification of the material density. Additional information on these datasets can be found in App. B. In this setup the expected uncertainty of only $\pm 5\%$ is exaggerated to achieve a detectable variations. Thus, a conversion factor of 0.25 can be applied to the retrieved systematic uncertainty. It turns out however that even with the full variation of 20%, the effect is compatible with zero as can be seen in Fig. 4.13. It can thus safely be assumed that the CMS tracking algorithm is sufficiently robust against imperfect knowledge of the tracker material and the effect can be ignored at 7 TeV.

4.5.2 Tracker Alignment

The geometrical alignment of the tracker sub-components is, even though already very well known from campaigns such as CRAFT [106], a potential source for systematic uncertainty. To examine this possibility, two scenarios were used for Monte Carlo comparison: The *start-up* scenario, which

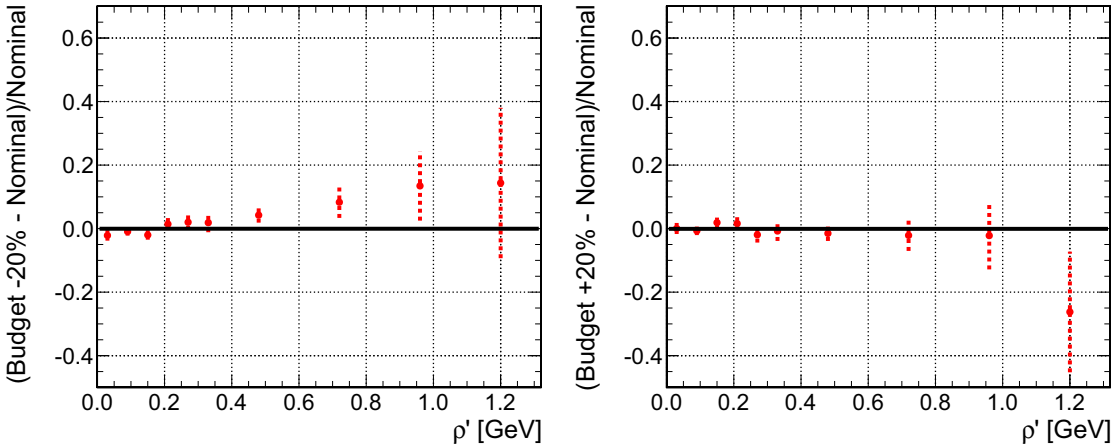


Figure 4.13: *Effect of the tracker material budget uncertainty at* 0.9 *TeV. In the left plot, the material budget was decreased by* 20%*, in the right, it was increased by* 20%*.*

contains the best current knowledge of the actual tracker alignment, was compared to an ideal alignment. As expected, the two turn out to be very similar and no systematic uncertainty is found within statistical uncertainty.

4.5.3 Non-Operational Tracker Channels

Another performance restraint that might influence the track measurement are non-operational tracker channels. The best knowledge of permanently failing channels is integrated in the simulation conditions. However, from run to run, the number and position of channels that are offline for shorter periods of time is very hard to determine. In order to quantify this effect, an additional Monte Carlo sample was produced with all channels operational. The uncertainty on the knowledge of the number of non-operational channels is estimated to be around 5%, thus a conversion factor of 0.05 has to be applied to the effect derived by evaluating the double ratios.

The distribution of the uncertainty estimation at 0.9 TeV in Fig. 4.15 shows a pronounced shift towards higher values of ρ' for the sample with all channels assumed functional. A fit with an error function is performed to the distribution and scaled with the conversion factor bin-by-bin. The final overall effect is 2% for very large and very small ρ' and smaller for intermediate values of $0.1 < \rho' < 0.3$.

In contrast to the 2009 data at 0.9 TeV, at 7 TeV the effect of non-operational tracker channels turns out to be negligible even before applying a conversion factor. This points to an improved performance of the tracking algorithms in the newer version of the reconstruction software.

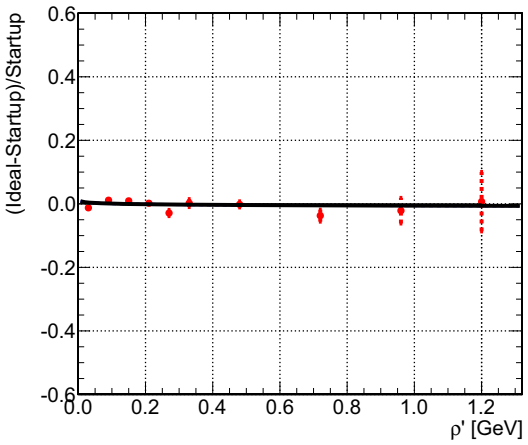


Figure 4.14: *Effect of an ideal tracker alignment at* 0.9 *TeV*

4.5.4 Trigger Efficiency

As lined out in chap. 4.1, level-1 trigger information is used in this analysis to select minimum bias events. It is shown in [103], that the simulation of this trigger is imperfect in the CMS software. The efficiency in Monte Carlo and data of the applied technical triggers 40 and 41 is different compared to another criterion for minimal event activity, an energy deposition in the hadronic forward calorimeters on both sides of the experiment. In data, this efficiency ratio is about 0.82, while in Monte Carlo it is only 0.76. To quantify the uncertainty due to the mismatch in the simulation, a Monte Carlo sample was compared, in which no trigger decision was used at all, effectively investigating all events present in the sample.
As expected, this procedure shifts ρ' towards smaller values as the trigger fails predominantly to accept events with very little activity.

With the difference between the efficiencies in data being about one quarter of the difference between the datasets with and without trigger selection, a scaling factor of 0.25 has to be applied to the retrieved distribution. It turns out, that over a broad range of ρ', the effect is in the order of 10%, resulting in a final systematic uncertainty of roughly 3%.

At 7 TeV, the effect of the imperfect modeling of the trigger shows a similar behaviour, yet to a slightly smaller degree. This is expected due to the increased average activity at a higher center-of-mass energy, resulting in a smaller number of events failing to meet th trigger requirements.

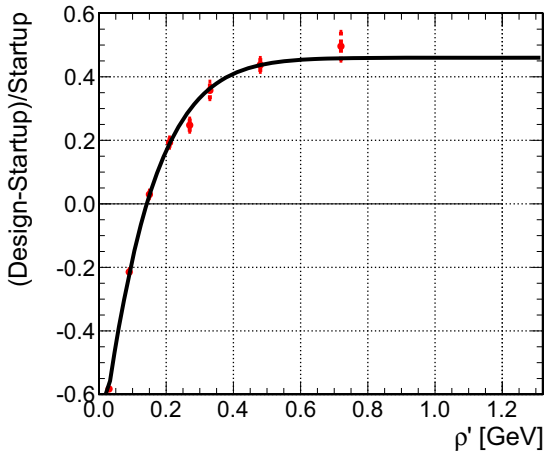


Figure 4.15: *Effect of non-operational tracker channels at* 0.9 *TeV. The* Design *sample assumes ideal conditions with all tracker channels functional.*

4.5.5 Vertex Reconstruction

One of the requirements for the events entering this analysis is the presence of exactly one vertex. The most influential requirement in the vertex reconstruction is the separation to the next vertex candidate in the event in the z-direction. For the 0.9 TeV run, this value was chosen to be at least 10 cm. A variation to values of 5 and 15 cm respectively was done and evaluated in the light of possible systematic uncertainties.

Even though the variations were vastly exaggerated compared to what might be expected for the uncertainty on the vertex separation, the effect is compatible with zero and therefore not investigated in the 7 TeV analysis. The corresponding plots are found in App. C.

4.5.6 Track Reconstruction Efficiency and Fake Rate

As demonstrated in section 4.4, the ρ' variable is very sensitive to the number of tracks in an event. Therefore, the effect of the tracking efficiency and fake rate has to be studied. According to [103], the uncertainty on the simulation of the tracking efficiency was conservatively estimated as 2% while the fake rate uncertainty lied in the order of 0.5%. These estimations are rather pessimistic, as this analysis uses a higher cut on the minimal transverse momentum of tracks, increasing the efficiency and reducing the fake rate. The total uncertainty on the track content of an event was therefore conservatively assumed to be $\pm 2\%$ The following procedure was applied to quantify this effect on ρ': A Monte Carlo sample with tune D6T was split into two independent subsamples. In the first one, each track was dropped with a probability of 2% and the ρ' distribution was compared

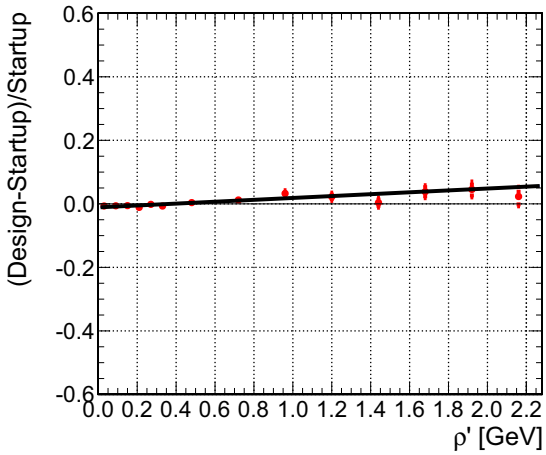


Figure 4.16: *Effect of non-operational tracker channels at 7 TeV. The* Design *sample assumes ideal conditions with all tracker channels functional.*

to the regular distribution. The other subsample was split into two halves, for clarification here named *A* and *B*. For every track already present in the events of *A*, randomly chosen from *B*, was added with a probability of 2%. This way, it is ensured that the additional tracks follow the same kinematic distributions as the ones already present in the sample.

Figure 4.19 shows that the variations in ρ' arising from adding and removing tracks cause an uncertainty that is ρ' dependent. The fit result in the figure, show that 5% are not exceeded at any point of the considered range.

For the 7 TeV study, see Fig. 4.20, the same procedure was applied, again splitting the Monte Carlo sample in two subsamples to add and remove tracks. Due to the higher overall activity and thus increased average number of tracks per event, the effect of the procedure is even smaller than found at 0.9 TeV and compatible with zero.

4.5.7 Variations of the Track Selection

Since the number of tracks in an event potentially has a large impact on ρ', it is sensible to investigate the influence of the track selection cuts on ρ'. The following variations have been studied:

- Pseudo-rapidity cut variation from $|\eta| < 2.3$ to $|\eta| < 2.0$ and $|\eta| < 2.5$
- Variation of the cut on the transverse and longitudinal impact parameters from 5σ to 4σ and 6σ.
- Variation of the transverse momentum significance $\sigma_{p_\mathrm{T}}/p_\mathrm{T}$ cut from 0.05 to 0.03 and 0.07.

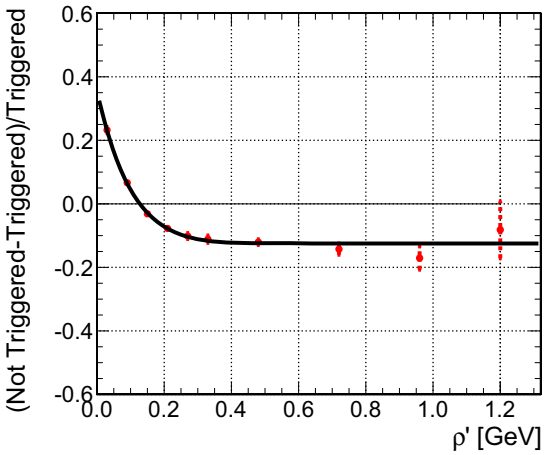


Figure 4.17: *Effect of the imperfect trigger simulation at* 0.9 *TeV. A Monte Carlo sample with no trigger requirement was compared to the regular level-1 trigger requirement of bit 40 or 41.*

- Transverse momentum cut variation from 300 MeV to 270 MeV to 330 MeV.

As can be understood from Fig. C.8, found in App. C, the first three variations do not influence ρ' significantly. The variation of the transverse momentum cut however affects the number of tracks per event significantly and thus the ρ' distribution.

The variation of the transverse momentum cut on tracks systematically shifts the ρ' distribution as can be observed in Fig. 4.21 and 4.22. A lower cut increases the number of tracks in the event and the total activity therefore rises, resulting in a higher ρ' value. A higher cut leads to the opposite observation of a tendency to lower ρ'. The size of the uncertainties are again derived by fitting the ratios and evaluating them bin-by-bin. The effects are below 5% for higher ρ' but substantial for the lowest bins for both, 0.9 and 7 TeV.

4.5.8 Transverse Momentum Resolution

The measurement of the transverse momentum of tracks is subject to potential imprecision. Even though track reconstruction is far less susceptible than calorimetric measurements, this potential effect has to be investigated nonetheless. Figure C.9 in App. C shows that the resolution derived from track-jets in Monte Carlo that are matched to charged particle jets in $\eta - \phi$ can be fitted with a very narrow Gaussian. The width of the Gaussian indicates a resolution uncertainty of below 2%. To quantify the impact of this effect on ρ', a Monte Carlo sample was processed by artificially smearing the transverse momentum of each track-jet with a Gaussian distribution with a width of 5%. This procedure yields a very pessimistic estimation. Nevertheless the effect is negligible, as can be seen in Fig. C.10 in App. C.

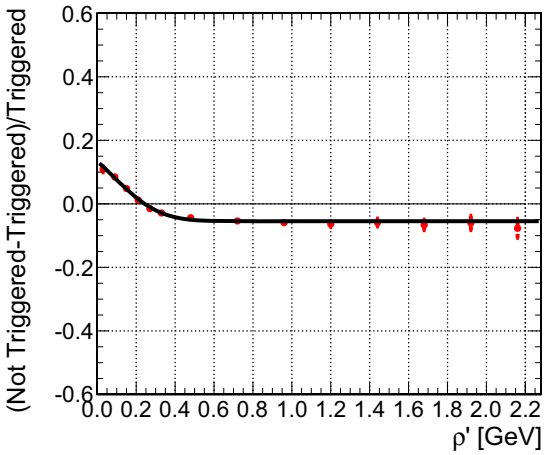


Figure 4.18: *Effect of the imperfect trigger simulation at 7 TeV.*

4.5.9 Track-Jet Response

In addition to a smearing of the jet energy, a possible systematic shift in its measurement is another potential source for uncertainties in the determination of ρ'. To quantify this effect, the distribution of $\sigma_{p_\mathrm{T}}/p_\mathrm{T}$ was investigated. It showed a good agreement with the track resolution that was retrieved from comparing tracks in Monte Carlo with matched charged generator particles. The average value of the distribution turns out to be 1.7%. Two Monte Carlo datasets were prepared to investigate the uncertainty of the track-jet response, in which the transverse momenta of all track-jets were multiplied with 1.017 and 0.983 respectively, accounting for the 1.7% estimated uncertainty. The effect of the track-jet response uncertainty is substantial for both center-of-mass energies and can be parameterized with a logarithmic function. The results are given in Fig. 4.23 and 4.24, amassing to 10% for very low and very high values of ρ' and a smaller effect in between.

4.5.10 Overall Systematic Uncertainty

Following the laws of error propagation, all derived uncertainties were added in quadrature on a bin-by-bin basis to account for ρ'-dependent effects. Only the effects that proved to be significant at 0.9 TeV have been studied at 7 TeV. Table 4.5 gives an overview of the investigated effects, the derivation technique of the numerical values and the size at $\rho = 1.2$ for 0.9 TeV and $\rho' = 2.2$ for 7 TeV center-of-mass energy respectively.

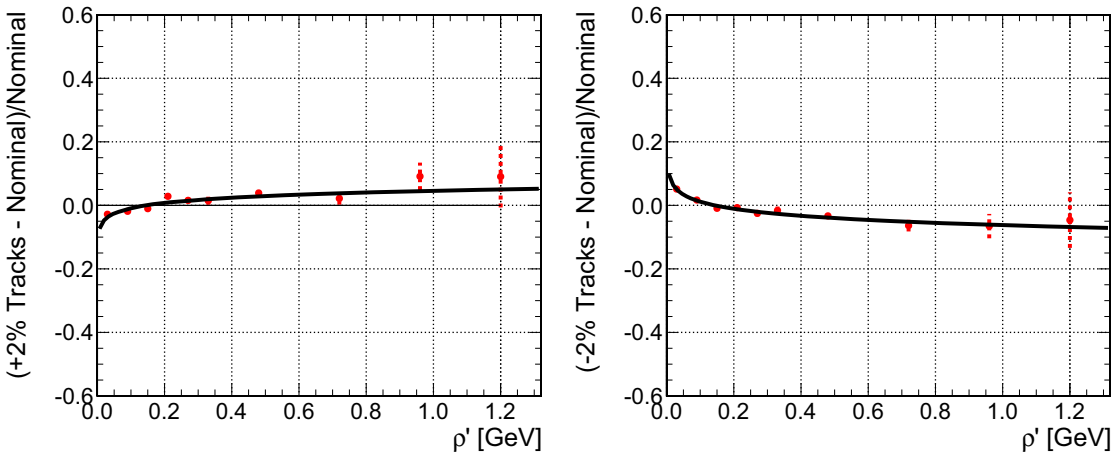


Figure 4.19: *Effects of adding and subtracting tracks at* 0.9 *TeV.*

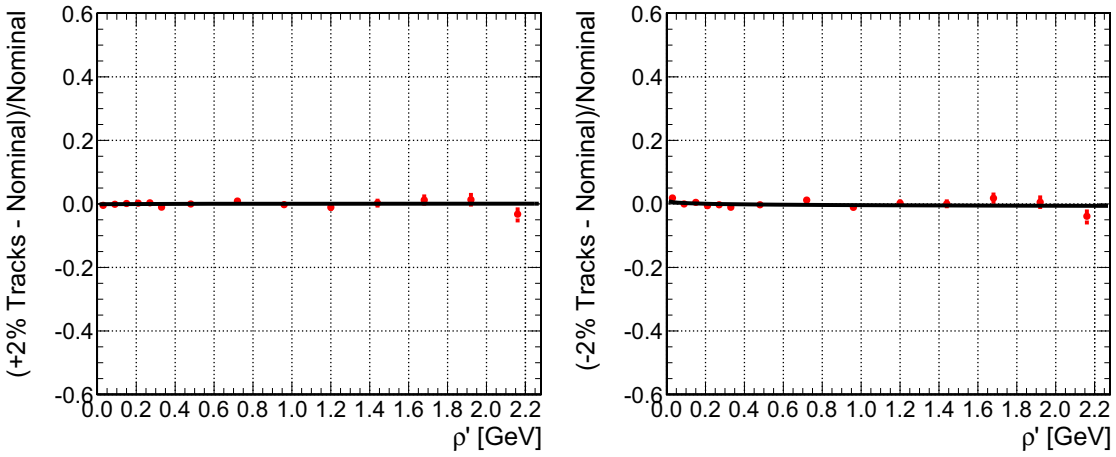


Figure 4.20: *Effects of adding and subtracting tracks at* 7 *TeV.*

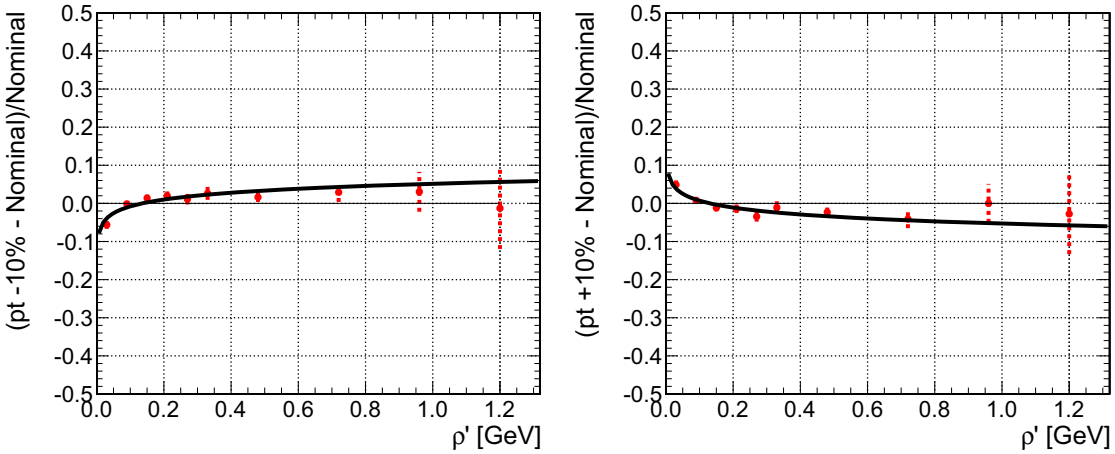


Figure 4.21: *Effect of the variation of the transverse momentum cut on tracks at* 0.9 *TeV*

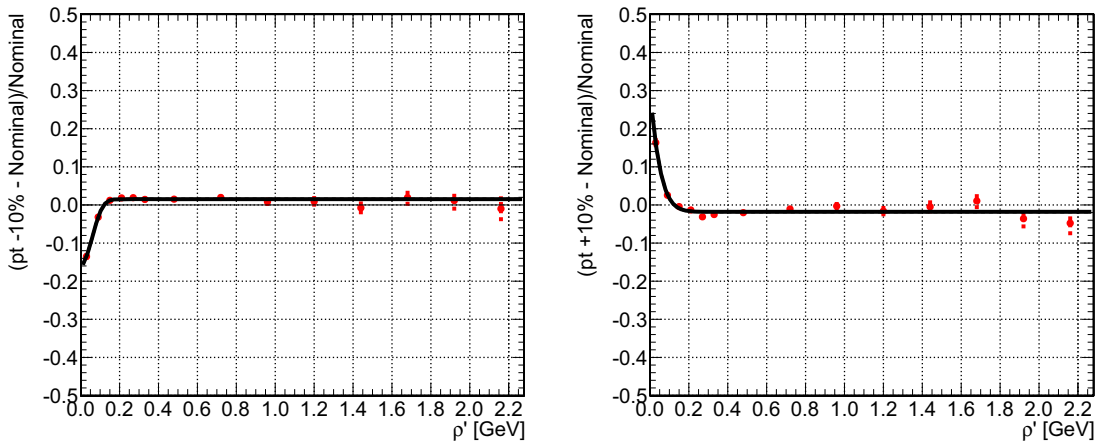


Figure 4.22: *Effect of the variation of the transverse momentum cut on tracks at* 7 *TeV*

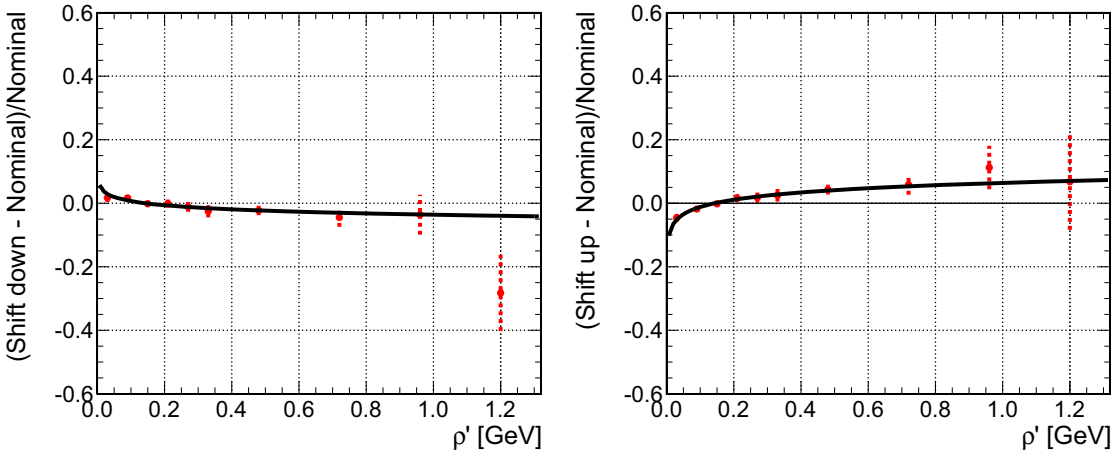


Figure 4.23: *Effect of a systematic shift in jet p_T at 0.9 TeV.*

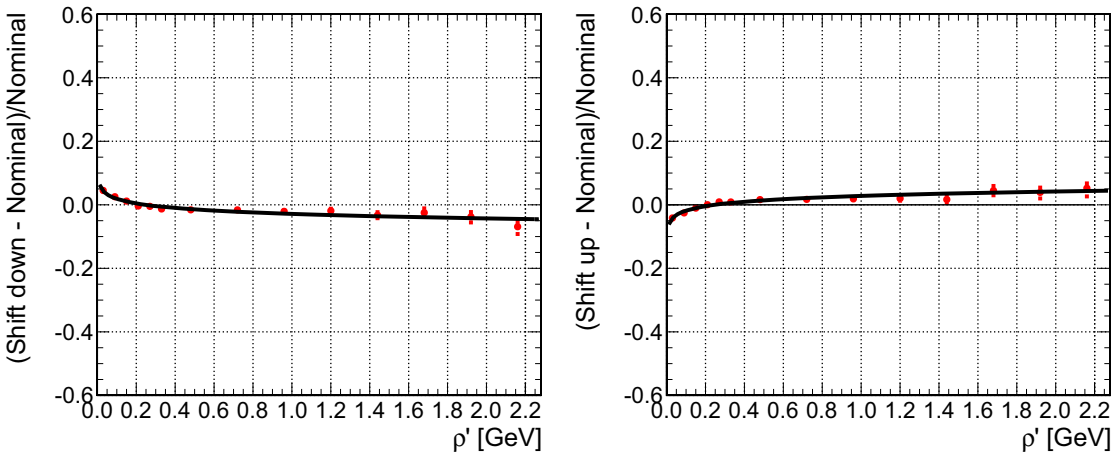


Figure 4.24: *Effect of a systematic shift in jet p_T at 7 TeV.*

Table 4.5: *Overview of the systematic uncertainties at both center-of-mass energies. The size estimation for ρ'-dependent effects corresponds mostly to the maximum considered values of $\rho' = 1.2$ at 0.9 TeV and $\rho' = 2.2$ at 7 TeV. Only effects that have proven to be significant at 0.9 TeV have also been studied at 7 TeV.*

Systematic Effect	Estimation Method	Size at 0.9 TeV	Size at 7 TeV
Tracker material budget	Constant value independent of ρ'	0.2%	
Minimal z separation between vertices		0.5%	
Maximal track $\lvert\eta\rvert$		0.5%	
Significances of track impact parameters		0.5%	
Maximal track p_{T} uncertainty		0.4%	
Track-jet p_{T} resolution		0.5%	
Tracker alignment	Derived bin-by-bin in ρ' from fit	0.6%	
Tracker map of non-operational channels		2.3%	0.5%
Track efficiency & fake rate mismatch		6.0%	1.0%
Minimal track p_{T}		5.8%	2.5%
Track-jet response shift		5.6%	3.8%
Trigger efficiency bias		3.1%	2.5%

4.6 Results

With all systematic effects under control, a comparison between Monte Carlo after full detector simulation including trigger emulation and data is possible.
In Fig. 4.25, the ρ' observable for track-jets is shown for the different Monte Carlo tunes and the 2009 commissioning data as well as the ratio of the Monte Carlo distributions with reference to data. The trends concerning the different event generation models observed with charged particle jets is reflected with track-jets as well: D6T and Pythia 8 predict the lowest number of events with high activity, with Pro-Q20 and P0 slightly above. DW with its favourable MPI energy scaling shows an excess of up to 50% compared to D6T at high values of ρ', while CW and Z1 project more than twice the number of high activity events. Comparing the Monte Carlo predictions to data, it can be observed that none of the tunes is able to reproduce the full ρ' spectrum. While D6T, P0, Pro-Q20 and Pythia 8 have too little events with high activity and an excess around the peak of the ρ' distribution, CW overestimates the number of high activity events. DW and Z1 are closest to reproducing the ρ' spectrum, yet DW is still slightly below and Z1 slightly above the measured distribution from data. For a wide range in ρ' the systematic uncertainty is in the order of 10%, the statistical uncertainty becomes relevant only for highest ρ'.

At 7 TeV, with an expected increase in activity, it is possible to enlarge the observed range in ρ'. In Fig. 4.26, the distribution of ρ' and the ratios of the different tunes to both, the reference tune D6T as well as the data, are shown for track-jets. Compared to the lower center-of-mass energy, the excess of events in data with higher activity is even more striking. Some similarities can be observed: Tune D6T, CW and DW predict the lowest number of events and are practically identical up to $\rho' \approx 1.2$, the maximum value considered at 0.9 TeV. For even higher ρ', D6T exceeds both, P0 and Pythia 8, which are very similar to each other. Z1 dominates the spectrum for a broad range of values, only being matched by D6T for the highest values of $\rho' = 2.2$.
For $\rho' > 0.4$ the data exceed all available Monte Carlo tunes and contain up to twice as many events with high activity. In addition, all tunes, apart from tune D6T, exhibit a slope towards a shrinking number of high activity events with rising ρ'. Z1 is the only tune that is able to at least follow the shape of the data distribution, yet with a constant 15% underestimation. Even though it disagrees with the data curve beyond systematic and statistical uncertainties, it is the tune that best describes the ρ' distribution at 7 TeV.

The results derived with the Jet Area/Median approach are largely in agreement with the traditional UE measurement. It confirms that the soft hadronic activity is underestimated by all tunes that have been available at the start of the LHC program. This effect is present at both center-of-mass energies investigated in this analysis. At 7 TeV, the differences among the tunes are even smaller than their deviations from data. The new tune Z1, which is especially tuned to describe the traditional UE observables at 0.9 and 7 TeV, fails to reproduce the activity seen in this analysis at 7 TeV, yet amongst all other tunes it comes closest.

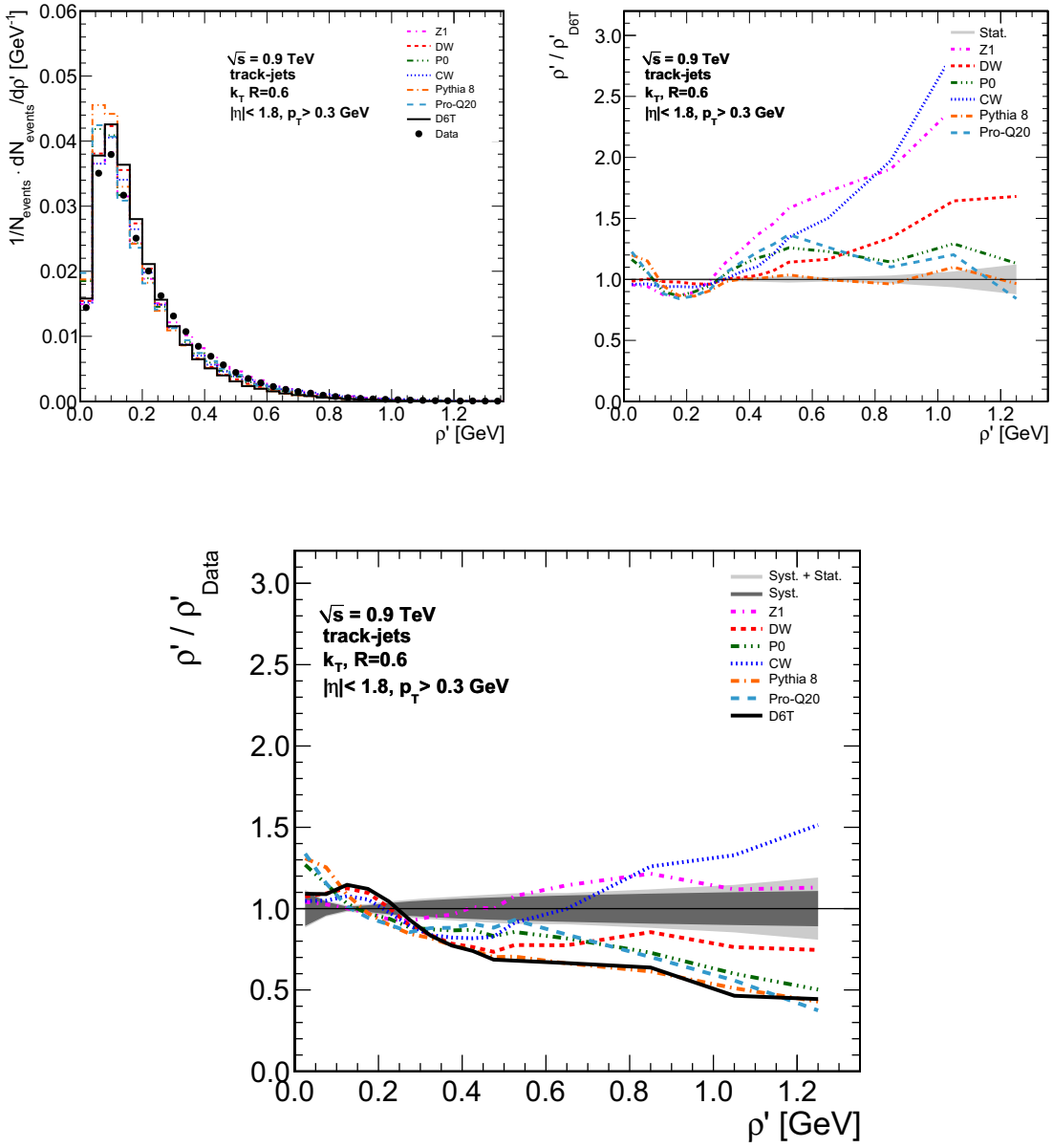


Figure 4.25: *Upper row: ρ' on track-jet level for 0.9 TeV and ratios of Monte Carlo tunes with respect to tune D6T. Lower plot: Ratios of Monte Carlo tunes to data. The light gray bands represent the sum of systematic and statistical uncertainty while the dark grey band represents the systematic uncertainty only.*

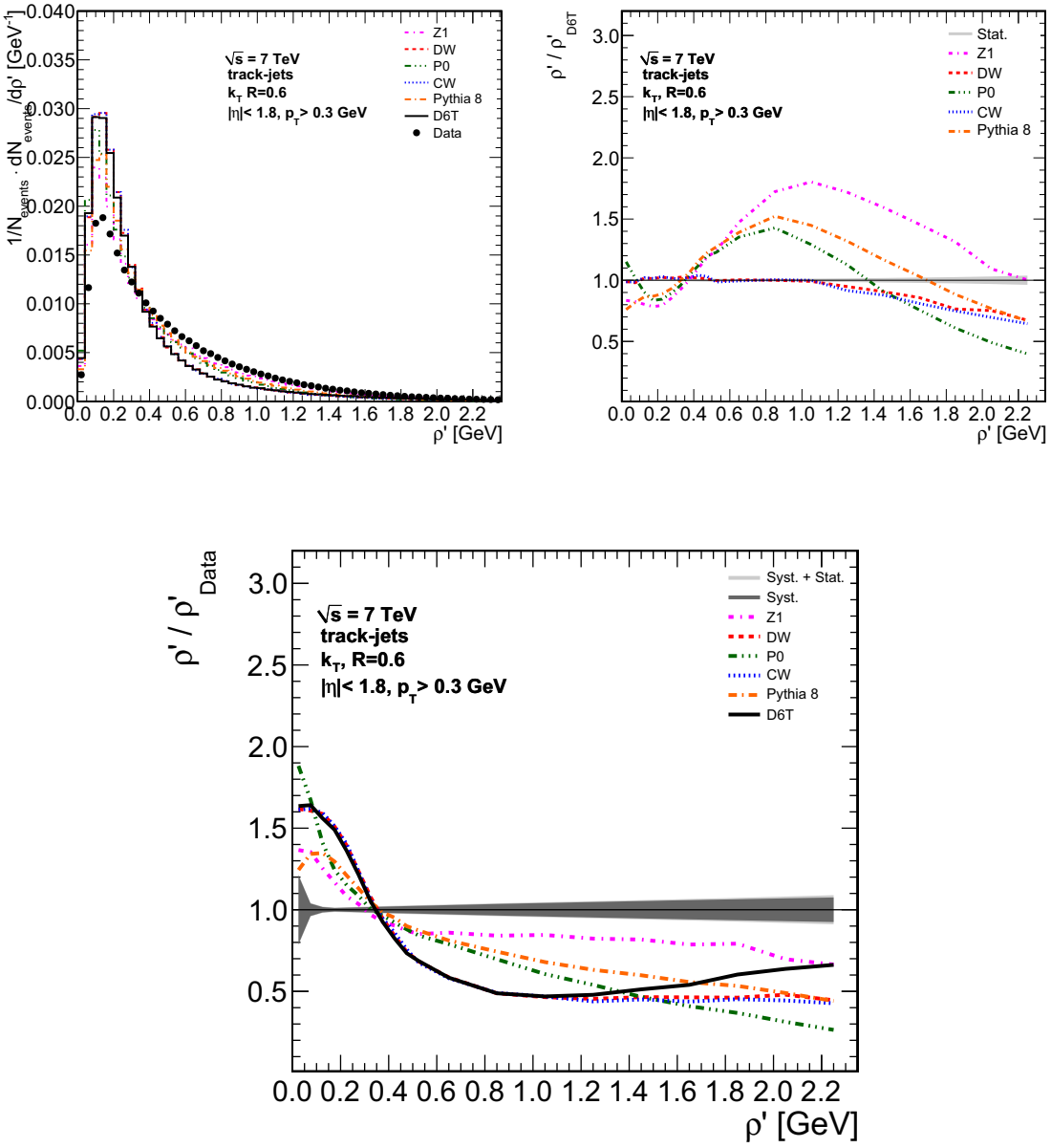


Figure 4.26: *ρ' on track-jet level for 7 TeV and ratios of Monte Carlo tunes with respect to tune D6T. Lower plot: Ratios of Monte Carlo tunes to data. The light gray bands represent the sum of systematic and statistical uncertainty while the dark grey band represents the systematic uncertainty only.*

Chapter 5

Summary and Perspectives

The early data taking stage of the LHC offers ideal conditions to investigate the behaviour of the underlying event (UE) in proton-proton collisions at two different center-of-mass energies. Due to the low instantaneous luminosity, events with one collision per bunch crossing can be studied, without pollution from pile-up. The outstanding performance of the CMS tracking detector is an ideal tool for this task. A new, complementary view on the UE activity has been proposed at the end of 2009. It takes advantage of the concept of jet areas to quantify the activity in events in an infrared and collinear-safe manner by defining an activity density and evaluating the median of the ratio of the transverse momenta over the areas of all jets in each event.

The exertion of the proposed method on data taken with the CMS experiment bears a number of challenges. First and foremost, the active area clustering method has never before been employed in a collider experiment. It is however necessary to adjust the originally proposed observable to account for the low average activity in the events especially at a center-of-mass energy of 0.9 TeV. Comparisons between numerous Monte Carlo tunes prove that the adjusted observable, named ρ', is sensitive to different UE models both on Monte Carlo particle jets and on track jets after full detector simulation with a minimum bias event selection. An extensive study is carried out to quantify the systematic uncertainty of numerous effects. Taking into account these effects and also statistical uncertainties arising from the limited number of events at hand, the Monte Carlo tunes and uncorrected data are still seperatable with clear tendencies observable.

None of the tunes that have been at hand at the start of the LHC is able to describe the ρ' distribution in data properly for both center-of-mass energies. All of them underestimate the charged activity in the events. This result is compatible with the traditional observables of UE contributions in CMS, which also find an underestimation of activity in all recent tunes.

The Jet Area/Median measurement will also play a part in improving Monte Carlo tunes. For now, it can be stated that even with the newest tune, which is a result of the tuning efforts on the traditional UE observables in CMS, the ρ' distribution in minimum bias events cannot be reproduced

in its entirety. A deeper understanding of the deviations between the results presented here and the near-perfect agreement of the new tune with the traditional measurement will require detailed studies of selected features of the UE models in Monte Carlo generators. Thus, the Jet Area/Median approach plays an important role in revealing shortcomings of tunes that are not noted with the traditional UE observables. As all analyses searching for new physics need the most precise modelling of events in Monte Carlo to separate their signals from the Standard Model background, new tunes will be necessary that also predict the Jet Area/Median observables properly.

The Jet Area/Median approach is not restricted to a certain event topology. Generally speaking, this method is so flexible, that every well defined event selection that can be consistently reproduced with Monte Carlo generators can be investigated. This way, ρ' distributions can be produced for different event types, which are ideally all described by the same UE tune. These event topologies include Drell-Yan muon production, with its especially clean signal, and multi-jet final states such as top quark production. Any slicing of the phase space for example in transverse momentum of a leading jet can also be combined with the Jet Area/Median approach.
With the complete data taken by CMS in 2010 available, it will be be interesting to investigate the evolution of the soft hadronic activity with increasing instantaneous luminosity. A run-based evaluation of the ρ' distribution with unaltered trigger configuration is a good starting point. Also, it is interesting to see how ρ' behaves with respect to the number of vertices reconstructed in an event.
The technique tested here is also an ideal candidate for a physically more sophisticated way of correcting measured jets for soft contributions. With the present technique of the multiplicative, multi-level corrections, the sizes of the corrections depend heavily on the running conditions. With the jet area technique at hand, it is possible to correct jets on an event-by-event basis, taking into account the actual activity in the given event and the size of the jet that is to be corrected. Any kind of soft contribution can be dealt with this way. This background subtraction is physically much more sophisticated and has the potential to reduce jet correction uncertainties significantly compared to the present jet correction scheme. Every analysis involving jets will benefit from this reduced uncertainty, enhancing both the precision of Standard Model measurements and the sensitivity of searches for new physics.

Appendix A

Monte Carlo Tunes

The handling of soft quantum chromodynamics effects connected to a hard parton scattering in Monte Carlo event generation is performed with the help of underlying event and fragmentation models. While the perturbative part of a hadron collision should be calculated rather similarly in different event generators, the soft models can vary distinctively. A paradigmic choice has to made between a model that has many free tunable parameters or one that has only a few but a more pronounced physical content. While the first solution can be very flexible and adjusted to many different measurements, it can lack intuitive physical concepts. The other solution however may suffer from its lack of flexibility when trying to combine different measurements in a single tune however it offers a deeper physical concept.

A.1 Pythia 6

The Monte Carlo generator Pythia 6 literally offers dozens of parameters that steer all effects concerning both parton shower, non-perturbative models and the interplay of the two. Over the years, more than 20 different tunes have been produced, that are al still available in the current version of Pythia 6. Prior to the start of the tuning efforts of the CDF experiment, the tunes were connected to a certain version of the generator and referred to for example as "Pythia 6.115 tune" [97]. Since then, the naming of the tunes has become somewhat arbitrary, with selected examples introduced in the following. Three main parameters are the foundations for the tunes and the tunes are usually categorized along them:

- The parton density function: Most current tunes use CTEQ6L, superseding the previous CTEQ5L tunes, alternatively also MRST LO* and others can be used. Especially the low-x

gluon density has a large implication on UE contributions as it determines the MPI cross-section for a given minimum momentum transfer.

- The model for multiple parton interactions: Three models exist, dubbed the “old model”, the “intermediate model” and the “new model”. Their properties will be explained in the following.
- The parton shower model. Although technically not part of the underlying event description, the parton shower is nonetheless closely connected to it and especially in the new model with interleaved showers does influence UE observables.

As Pythia 6 is written in the FORTRAN programing language, the parameters are stored as common block variables, with the typical FORTRAN notation.

A.1.1 Tunable Parameters

Out of the large number of parameters, the following are the most important for the generation of UE contributions [29]:

- MSTP(81): Switch for the multiple parton interaction model with associated treatment of beam-beam remnants. The choices are old, intermediate or new model, all with or without MPI.
- MSTP(82): Switch for the structure of the MPI and matter distribution inside the protons. The default value of 4 represents a model with collisions of variable impact parameters and a double Gaussian matter distribution.
- PARP(82): Regularization scale for MPI. The value is given at the reference center-of-mass energy PARP(89) and is extrapolated following PARP(90).
- PARP(83) and PARP(84): Parameters of the matter overlap chosen in MSTP(82)
- PARP(85) and PARP(86): Probabilities of MPI producing gluons instead of quarks.
- PARP(89): Reference energy scale for PARP(82). Values are typically Tevatron center-of-mass energies.
- PARP(90): Power of energy rescaling for MPI regularization.
- PARP(62): Cut-off scale for space-like parton evolution.
- PARP(64): Scaling value for the squared transverse momentum evolution scale.
- PARP(67): Maximum parton virtuality allowed in virtuality ordered showers.

- MSTP(91): Parton transverse momentum distribution in incoming hadrons. A value of 1 corresponds to a Gaussian with width PARP(91) and upper cutoff PARP(93).
- PARP(91) and PARP(93): Parameters of intrinsic transverse parton momenta MSTP(91).

A.1.2 Ancient Tunes

All Pythia tunes before Tune A are no longer used in practice and only kept for consistency. Note that due to the primitive MPI model, the set of parameters of the tunes differ from the more advanced "old" tunes. The development leading to Tune A can be understood from Tab. A.1.

Parameter	Tune 6.115	Tune 6.125	Tune 6.158	Tune 6.208
MSTP(81)	1	1	1	1
MSTP(82)	1	1	1	1
PARP(81)	1.4	1.9	1.9	1.9
PARP(82)	1.55	2.1	2.1	1.9
PARP(89)		1.0	1.0	1.0
PARP(90)		0.16	0.16	0.16
PARP(67)	4.0	4.0	1.0	1.0

Table A.1: *Ancient Pythia 6 tunes. Empty fields mark a parameter that was not available in the corresponding version.*

A.1.3 Old Tunes

The most commonly used tunes in large scale Monte Carlo production are still using the "old" model of multiple parton interactions. An overview of a number of these tunes is given in Tab. A.2. The main differences among them are the choice of PDF and especially the scaling power of the MPI PARP(90). DWT and D6T employ a steeper scaling which results in a higher MPI cross-section at LHC energies than DW and D6. Even though these tunes were discouraged by RHIC data, they are still the tunes of choice for many analyses. After the evaluation of Minimum Bias 0.9 TeV data at the LHC, tune CW was introduced with a very slow energy scaling that showed good results at that particular analysis but as expected is not suited at all for energies higher than Tevatron.

Parameter	A	B	DW	D6	DWT	D6T	ATLAS	CW
PDF	CTEQ5L	CTEQ5L	CTEQ5L	CTEQ6L	CTEQ5L	CTEQ6L	CTEQ5L	CTEQ5L
MSTP(81)	1	1	1	1	1	1	1	1
MSTP(82)	4	4	4	4	4	4	4	4
PARP(82)	2.0 GeV	1.9 GeV	1.9 GeV	1.8 GeV	1.9409 GeV	1.8387 GeV	1.8 GeV	1.8 GeV
PARP(83)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
PARP(84)	0.4	0.4	0.4	0.4	0.4	0.4	0.5	0.4
PARP(85)	0.9	1.0	1.0	1.0	1.0	1.0	0.33	0.9
PARP(86)	0.95	1.0	1.0	1.0	1.0	1.0	0.66	0.95
PARP(89)	1.8 TeV	1.8 TeV	1.8 TeV	1.8 TeV	1.96 TeV	1.96 TeV	1.0 TeV	1.8 TeV
PARP(90)	0.25	0.25	0.25	0.25	0.16	0.16	0.16	0.3
PARP(62)	1.0	1.0	1.25	1.25	1.25	1.25	1.0	1.25
PARP(64)	1.0	1.0	0.2	0.2	0.2	0.2	1.0	0.2
PARP(67)	1.0	4.0	2.5	2.5	2.5	2.5	1.0	2.5
MSTP(91)	1	1	1	1	1	1	1	1
PARP(91)	1.0	1.0	2.1	2.1	2.1	2.1	1.0	2.1
PARP(93)	5.0	5.0	15.0	15.0	15.0	15.0	5.0	15.0

Table A.2: *Old Pythia 6 tunes, based on the virtuality ordered parton shower and the "old" MPI model.*

The intermediate model of MPI is no longer in use and only kept for consistency.

A.1.4 New Tunes

The Pythia 6 tunes with the "new" MPI model were first introduced in [107] in a work on top quark property estimation at the Tevatron. They usually feature both the new transverse momentum ordered parton shower and colour reconnections, except for tune ProQ20, which still has the q^2-ordered parton shower. Recently, new tunes have been published that make use of the automated tool PROFESSOR [108]. These tunes employ a new fragmentation function that has been derived from re-evaluating hadronic LEP events. Furthermore, a new matter overlap profile has been chosen for MPI for some of the tunes, which can be varied between a Gaussian profile and and an exponential function.

Parameter	S0	P0	ProQ20	Z1	Z2
PDF	CTEQ5L	CTEQ5L	CTEQ5L	CTEQ5L	CTEQ6L
MSTP(81)	21	21	21	21	21
MSTP(82)	5	5	4	4	4
PARP(82)	1.85 GeV	2.0 GeV	1.9 GeV	1.932 GeV	1.832 GeV
PARP(83)	1.6	1.7	0.83	0.356	0.356
PARP(84)			0.6	0.651	0.651
PARP(85)	0.9	0.9	0.86	0.9	0.9
PARP(86)	0.95	0.95	0.93	0.95	0.95
PARP(89)	1.8 TeV	1.8 TeV	1.8 TeV	1.8 TeV	1.8 TeV
PARP(90)	0.16	0.26	0.22	0.275	0.275
PARP(62)	1.0	1.0	2.9	1.025	1.025
PARP(64)	1.0	1.0	0.14	1.0	1.0
PARP(67)	1.0	1.0	2.65	1.0	1.0
MSTP(91)	1	1	1	1	1
PARP(91)	2.0	2.0	2.1	2.0	2.0
PARP(93)	5.0	10.0	5.0	10.0	10.0

Table A.3: *New Pythia 6 tunes, based on the transverse momentum ordered parton shower and the "new" MPI model. Note that tune ProQ20 uses the virtuality ordered shower and that PARP(83) has a different meaning depending on the model chosen in MSTP(82), while PARP(84) is only required for a double Gaussian matter distribution.*

A number of effects have been improved in the new MPI model compared to the old one: Whilst in the old model, only the partons taking part in the hard interaction can radiate ISR, in the new model also partons that interact softly as part on MPI can radiate. These radiation effects are interleaved with the MPI also in a p_{T}-ordered manner, in order to avoid double counting in the PDF space.

A.2 Pythia 8

The tuning efforts for Pythia 8 are currently only in its beginning stages. Mostly, the default tune, also referred to as Tune 1 is used, that is loosely based on tune P0 of Pythia 6 with a similar MPI model as the “new model” of Pythia 6.

A.3 Herwig/JIMMY and Herwig++

The Herwig 6 event generator natively did not include a model for MPI and therefore predicted far too little activity in classical UE measurements [109]. It was however possible to interface the JIMMY package to Herwig 6, providing this functunality. A tune is available that resembles tune A of Pythia 6 and has for example been used in [71].
Herwig++ on the other hand has a similar model as Herwig 6 already built-in. It relies mostly on only two parameters to steer multiple parton interactions [110], one of which is the regularization scale for MPI and the other describes the matter distribution inside the protons. The recommended tune is the default tune of the current version of Herwig++ [111], tuned on Tevatron data with the traditional approach.

Appendix B

Additional Information on Software Setup and Datasets

B.1 The JUEZ data format

Even though the CMSSW framework offers all tools necessary for user analysis, files containing the full reconstructed information tend to be impractical for large scale analysis use due to their size. It is therefore suitable to first run over the centrally hosted RECO and write out the events in a reduced format, containing only information needed for the particluar analysis, thus reducing the storage requirement per event significantly. This process is called *skimming*. The resulting skims can then be run on multiple times in the course of the development of an analysis, yielding for example bare ROOT files, containing only histograms.
Depending on the subjects of their studies, it can be beneficial if multiple working groups join their skimming efforts, producing one common set of skimmed files. A data format called JUEZ [1], has been co-developed and used for this analysis. The concept behing the JUEZ format is to build a date format that is strictly based on ROOT functionality, providing object classes that inherit from the TObject class. These obejcts can be written to a file with the help of TClonesArrays, providing a persistent array of identical TObjects. The following TObjects can be used for event handling that are closely related to the corresponding CMSSW classes, yet allow for a standalone ROOT approach to reading them:

- KAGenParticle: Derived from ROOT TParticle, contains also generator level information such as mother-daughter relations.

[1] Jets, Underlying Event, Z-boson studies

- KAJetID: Contains members and classes for both CaloJet and PFJet identification.
- KAL1: Level-1 trigger information.
- KAParticle: Class for generalized particles: Can contain muons as well as jets and offers special functionality such as jet area information.
- KATrack: Contains reconstructed tracks.
- KAVertex: Vertex information.

Furthermore, arbitrary meta information can be stored in the files that can for example act as internal book-keeping and documentation.
As mentioned above, the first analysis step when using a dedicated skimming format is to run over the centrally hosted datasets and write out skims to a dedicated storage facility. This is still a CMSSW process and all kinds of modules can be run before the actual skimming step. In the case of the present analysis, the jet re-clustering with the active area method has been performed in the same step. Also, all kinds of cuts and event selections can be performed at any time before writing out the files. If a strict preselection is in place and only few events are considered in the analysis, the files can already be small enough to be hosted on a single desktop machine.
The actual analysis is then performed on the JUEZ skim files. From this point on, CMSSW is no longer needed for data access. A standalone version of the JUEZ classes is available via reflex dictionaries that are linked against ROOT executables containing the analysis jobs. These jobs have a relatively short turn-around and can easily be parallelized. Thus they can be repeated numerous times in order to refine the analysis and cope with for example growing datasets. The output of the analysis jobs is stored in the form of histograms containing all the quantities that are investigated, resulting in small files that can easily be merged and require even a lot less storage. At this point, all the information required the analysis is usually gathered in just one file. The histograms in this file is then processed in the final step and formatted appealingly. A typical step at this point can for example be the division of two histograms producing ratio distributions. Storing the final formatted output histograms is perfomed by native ROOT is steered with the help of python configuration files. The turnaround of re-formatting the histograms is in the order of a few seconds, as due to dynamic configuration, no re-compiling is necessary.

B.2 Data Access and Software Setup

B.2.1 The National Analysis Facility at DESY

The standard way of accessing and computing LHC data is closely connected to the Worldwide LHC Computing Grid (WLCG). However, it is possible to provide additional computing ressources

for a specific group of users, for example the members of German institutes contributing to either of the LHC experiments. This approach is taken with the National Analysis Facility (NAF) which is locatet at DESY in Hamburg, which is connected to the Tier-2 centre in place. In addition to the usual grid based approach to computing and storage ressources, also an interactive login on portal machines is possible. A dedicated job queue is available for German users, that uses independent worker nodes but shares the storage facilities with the actual grid site. It is therefore possible to request a certain dataset to the regular Tier-2 storage and access it via the queue. With an additional lustre filesystem in place for local users, very fast processing or skimming becomes feasable. Furthermore full grid functionality can be accessed from the NAF portals, offering ideal working conditions for user analyses.

B.2.2 Grid-Control

For this analysis the job submission tool Grid-Control [112] was used. This is a toolset, completely written in Python that offers a wide range of functionality on top of bare job submission. It can be used for arbitrary computing jobs yet has support for CMSSW as well as dataset based operations and DBS support. Also, it can access storage elements for read-and-write operations via the grid. All configuration files are human-readable, easing their editing and usage.
A very important feature for this analysis was the capability of Grid-Control to submit both to the grid and to a local batch queue for example the PBS queue at the NAF, being able to easily combine the advantages of practically unlimited computing ressources with a short turn-around on less CPU-comsuming jobs.
The Grid-Control package can be downloaded at
`https://ekptrac.physik.uni-karlsruhe.de/trac/grid-control/wiki`
and comes with a number of examples covering a wide range of possible use-cases such as submitting jobs, retrieving their output from a grid-SE, copying files to and from SEs and so on.

B.3 Data Samples

In important aspect of the full detector simulation for Monte Carlo events are the applied detector conditions. The following subdetector conditions have to be provided to the simulation:

- Tracker and Muon alignment
- Hcal Calibration
- Ecal Calibration
- SiStrip calibration

- SiPixel calibration
- DT calibration
- L1 Trigger Conditions

The condition information itself is not contained in and shipped with the CMS software relase but stored in a databased called ORCOF which is accessed during event procession. Defined tags of the conditions of each subsystem are collected into a *Global Tag*, which is also stored in the database. Different global tags can be used with the same software release, increasing flexibility and compatibility. With growing data at hand and in order to keep track of changes in the detector conditions, new global tags are regularly collected. For central Monte Carlo production efforts like the ones used in this analysis, a tag is chosen that corresponds to best knowlegde to the current state of the detector. With the defintion of global tags, it is also possible to easily produce Monte Carlo datasets with different conditions, such as a perfectly aligned tracking detector in order to study systematic uncertainties arising from imperfect detector understanding.

B.3.1 0.9 TeV Data from the 2009 Commisioning Run and Monte Carlo Samples

Table B.1 shows the runs, luminosity sections and bunch crossings that have been identified as valid for this analysis in co-ordinance with [100].

Table B.1: *List of runs, luminosity blocks and bunch crossings that have been used for the 2009 commisiong data.*

Run	Luminosity Block	Bunch Crossings
124020	12-94	51, 151, 2824
124022	60-69	51, 151, 2824
124023	41-96	51, 151, 2824
124024	2-83	51, 151, 2824
124027	24-39	51, 151, 2824
124030	1-31	51, 151, 2824
124230	26-68	51, 151, 232, 1024, 1123, 1933, 2014, 2824, 2905

Table B.2 contains the datasets used in the 0.9 TeV analysis, both the MinimumBias detector data and the Monte Carlo with full detector simulation. The different tunes contain a different number of events, P8 and Z1 were produced privately in addition to the official production. The datasets for the studies of the tracker material budget are not shown here, as they are not officialy published in DBS. The global tag STARTUP_V8K corresponds to the 19th december re-reconstruction. It

contains realistic conditions for 0.9 TeV MinBias production with a beamspot distribution as measured for the collision data in the 2009 commisioning run. Also, the tracker is operated in *peak mode*. The MC_V9B conditions describe a perfectly aligned detector, yet with a realistic map of non-functional tracker channels. The DESIGN_V8A conditions on the other hand has all channels functional yet realisticly smeared alignment and calibration constants. All applied global tags contain the 8E29 HLT menu for low luminosity running. Details on the conditions and the corresponding databases are described in [113].

Table B.2: *DBS entries for 2009 commisiong data and corresponding Monte Carlo data samples. The Z1 and Pythia 8 datasets were produced privately. The V9B and DESIGN datasets were used to study the effects of tracker alignment and dead tracker channels.*

Data Sample	Events
/MinimumBias/BeamCommissioning09-Dec19thReReco_336p3_v2/RECO	$19,681,382$
/MinBias/Summer09-STARTUP3X_V8K_900GeV-v1/GEN-SIM-RECO	$10,951,200$
/MinBias/Summer09-STARTUP3X_V8K_900GeV_P0-v2/GEN-SIM-RECO	$2,195,680$
/MinBias/Summer09-STARTUP3X_V8K_900GeV_DW-v1/GEN-SIM-RECO	$2,048,000$
/MinBias/Summer09-STARTUP3X_V8K_900GeV_ProQ20-v1/GEN-SIM-RECO	$2,278,400$
/MinBias/Summer09-STARTUP3X_V8K_900GeV_P8-priv/GEN-SIM-RECO	$310,000$
/MinBias/Summer09-STARTUP3X_V8K_900GeV_Z1-priv/GEN-SIM-RECO	$310,000$
/MinBiasCW900A/Summer09-STARTUP3X_V8K_900GeV-v1/GEN-SIM-RECO	$2,167,605$
/MinBias/Summer09-MC_3XY_V9B_900GeV-v2/GEN-SIM-RECO	$10,985,000$
/MinBias/Summer09-DESIGN_3X_V8A_900GeV-v1/GEN-SIM-RECO	$550,000$

B.3.2 7 TeV Data and Monte-Carlo

Table B.3 contains the dataset used for the 7 TeV analysis, both the *good collisions* skim and the corresponding Monte Carlo productions. The V26B conditions are matched to the 6th may re-reconstruction of the 2010 commisioning data and contain a realistic detector description. The logical level-1 trigger requirement for the good collission skim reads:
$0 \wedge (40 \vee 41) \wedge \neg(36 \vee 37 \vee 38 \vee 39) \wedge \neg((42 \wedge \neg 43) \vee (43 \wedge \neg 42))$
This includes minimal activity trigger bits as well as a cleaning for beam halo and scraping events.

Table B.3: *DBS entries for Good Collision skims from may 6 rereconstruction and corresponding Monte Carlo data samples. The Z1 datasets was produced privately.*

Data Sample	Events
/MinimumBias/Commissioning10-May6thPDSkim_GOODCOLL-v1/RAW-RECO	61,003,531
/MinBias_TuneD6T_7TeV-pythia6/Spring10-START3X_V26B-v1/GEN-SIM-RECO	10,493,175
/MinBias_TuneP0_7TeV-pythia6/Spring10-START3X_V26B-v1/GEN-SIM-RECO	10,932,196
/MinBias_DWtune_7TeV-pythia6/Spring10-START3X_V26-v1/GEN-SIM-RECO	2,187,040
/MinBias_7TeV-pythia8/Spring10-START3X_V26B-v1/GEN-SIM-RECO	10,764,844
/MinBias_Z1tune-priv_7TeV-pythia6/Spring10-START3X_V26-v1/GEN-SIM-RECO	550,000

Appendix C

Supplemental Plots

In the following, additional plots are presented that back up the conclusions made in chapter 4. These are mostly ϕ-distributions, that are expected to be flat on generator level or show small deviations only from detector geometry artefacts. Also, η-distributions for jets are shown that contain no unexpected features and are thus shown here to increase the readability of the analysis chapter. Also, selected distributions which are shown with a logarithmic y-axis are shown here again with a linear y-axis.
Furthermore, in the second part, all plots from the study of the systematic uncertainties that are consistent with zero can be found in this appendix.

C.1 Track and Charged Particle Distributions

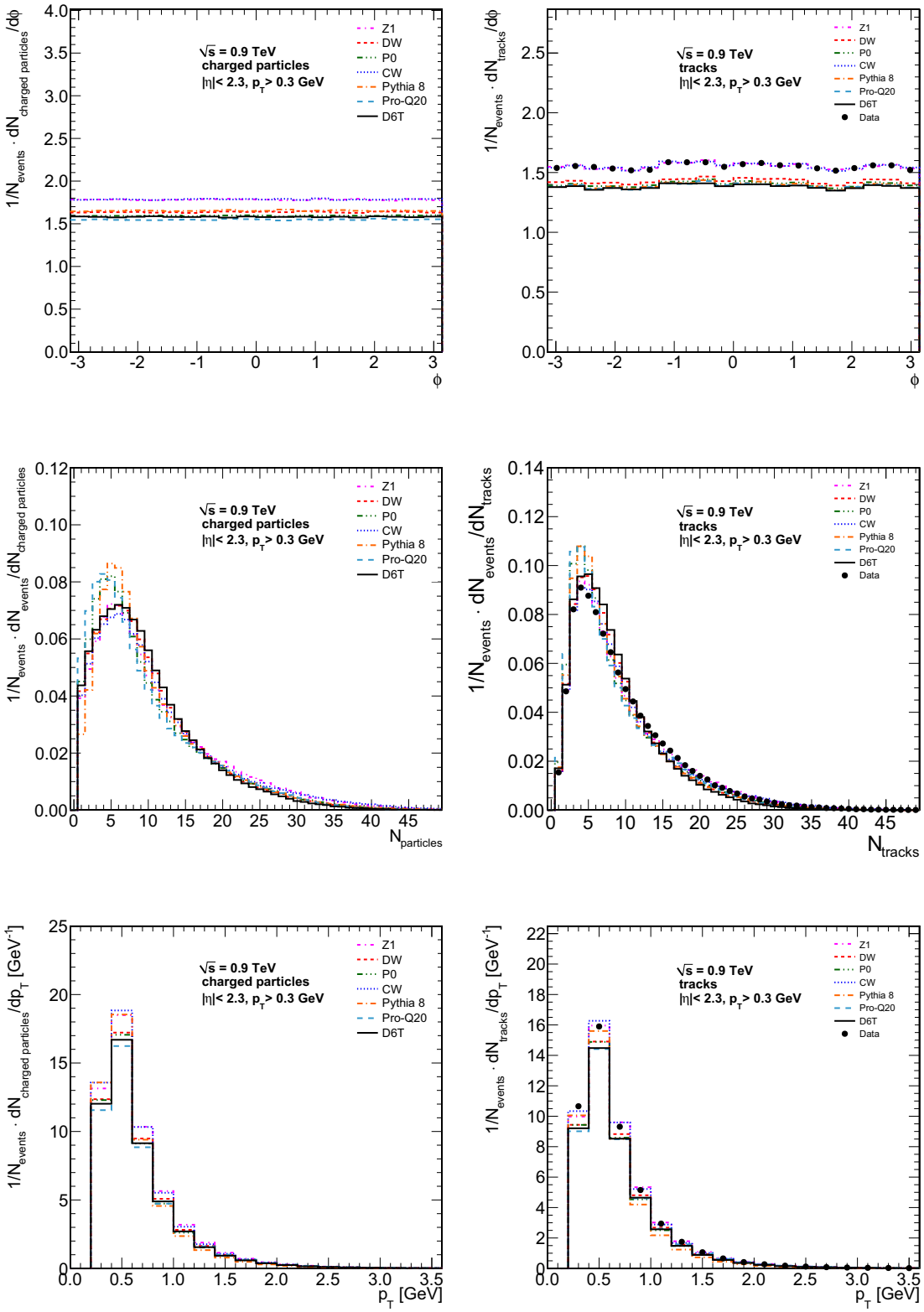


Figure C.1: *Angular distributions of charged particles and tracks at* 0.9 *TeV (upper row). Multiplicity of charged particles and tracks at* 0.9 *TeV (middle row) with linear* y*-Axis. Transverse momentum of charged particles and tracks at* 0.9 *TeV (lower row) with linear* y*-Axis.*

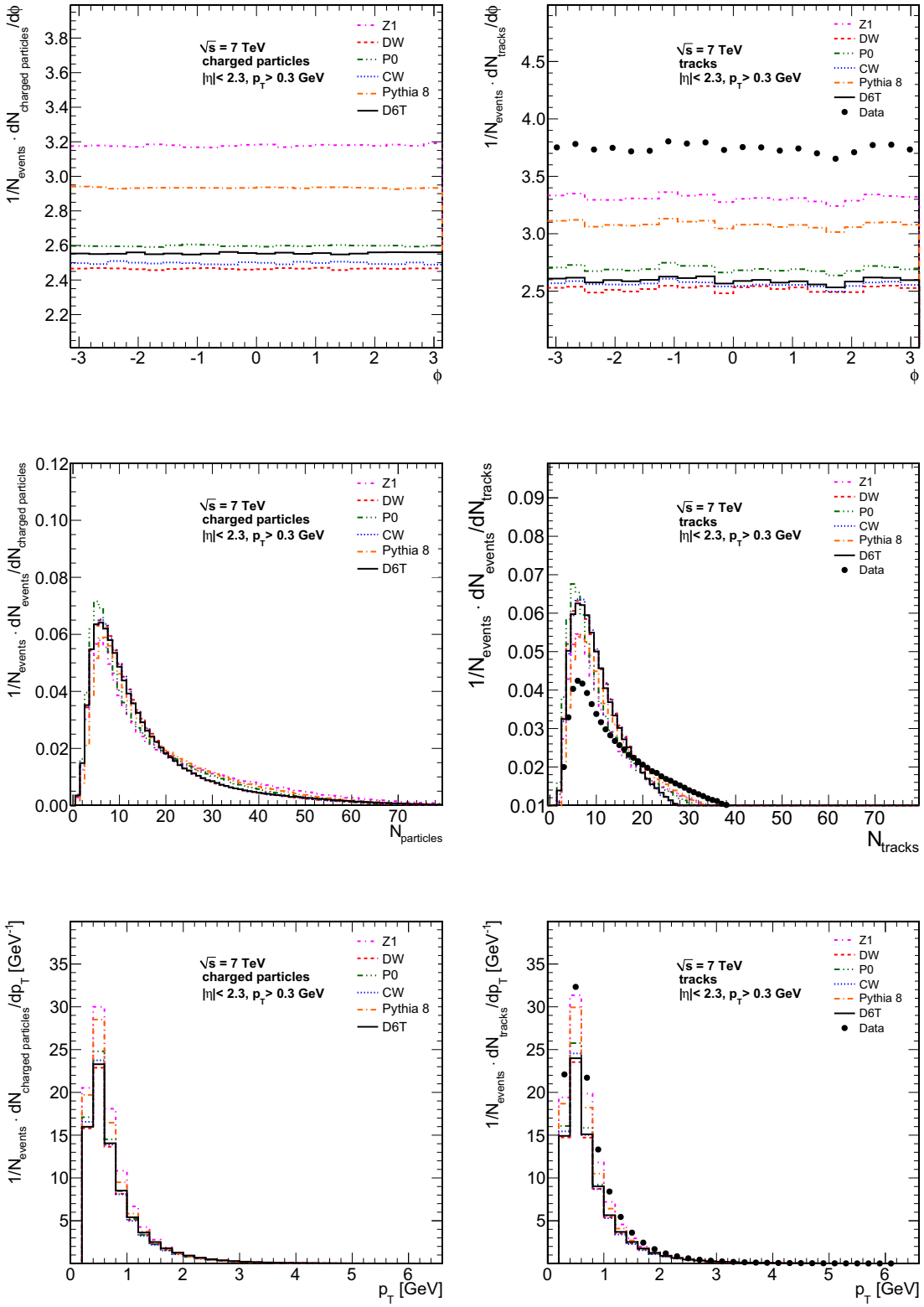


Figure C.2: *Angular distributions of charged particles and tracks at 7 TeV (upper row). Multiplicity of charged particles and tracks at 7 TeV (middle row) with linear y-Axis. Transverse momentum of charged particles and tracks at 0.9 TeV (lower row) with linear y-Axis.*

C.2 Jet Distributions

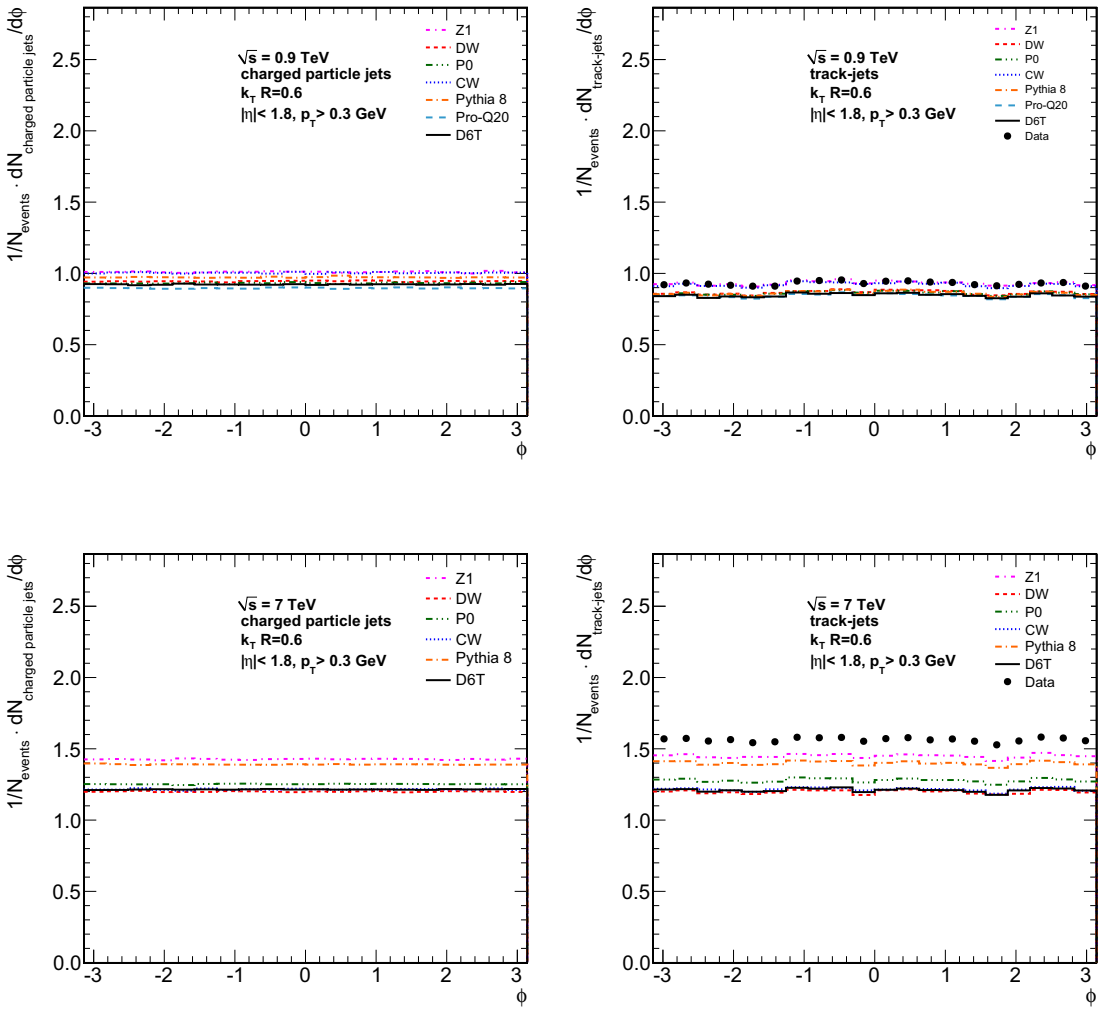


Figure C.3: *Angular distributions of track-jets at both center-of-mass energies.*

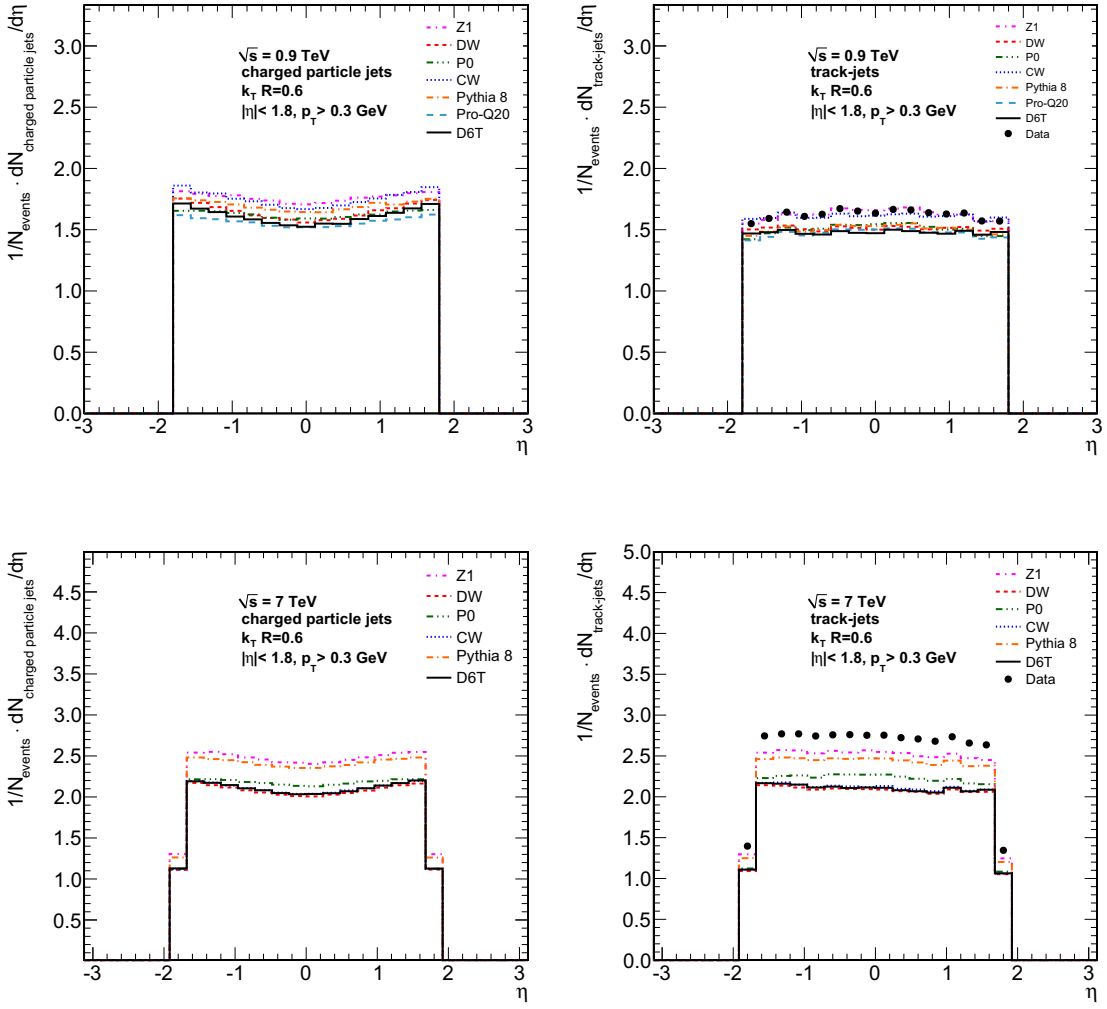


Figure C.4: *Pseudo-rapidity distributions of charged particle jets and track-jets at both center-of-mass energies.*

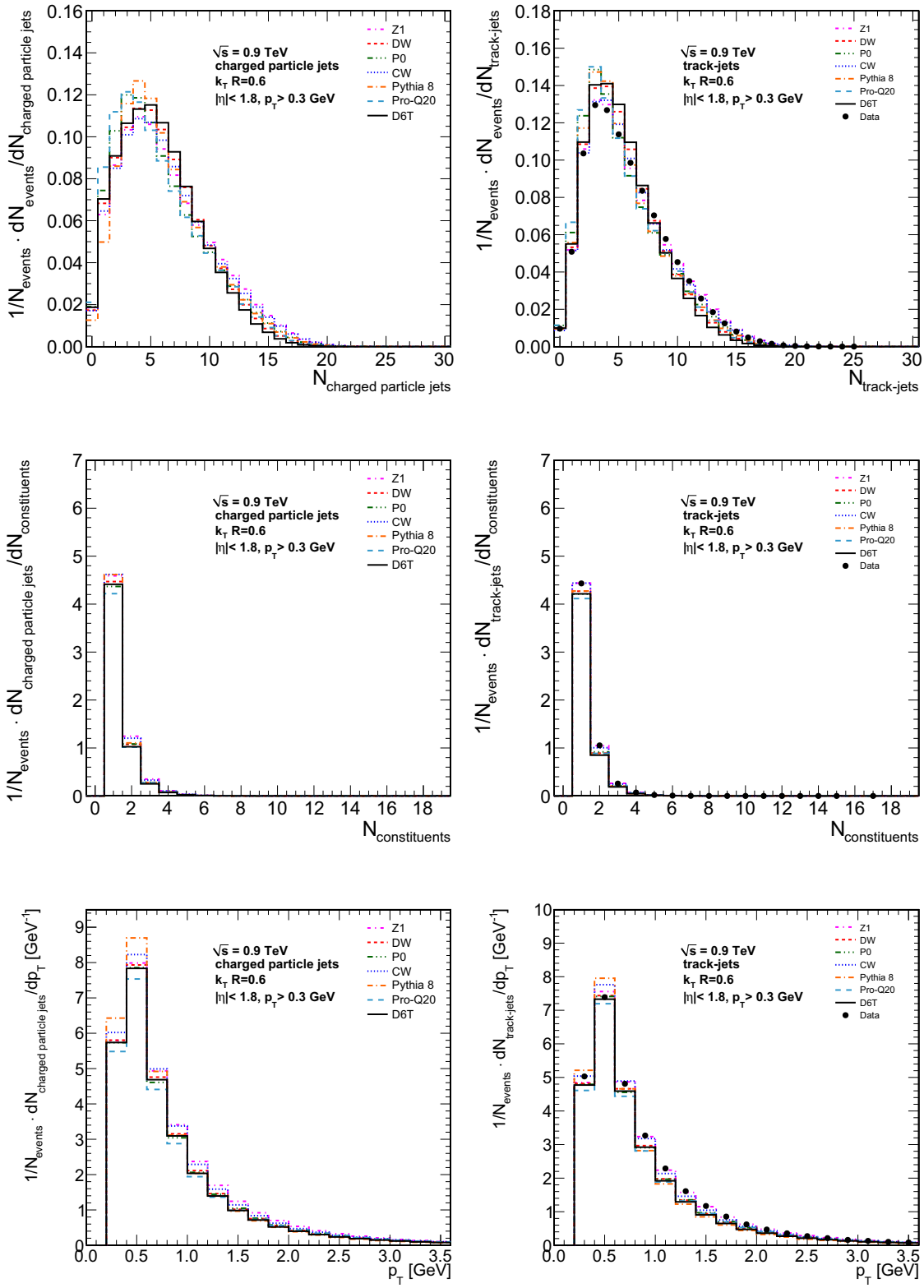


Figure C.5: *Same as Fig. 4.5 with linear y-axes.*

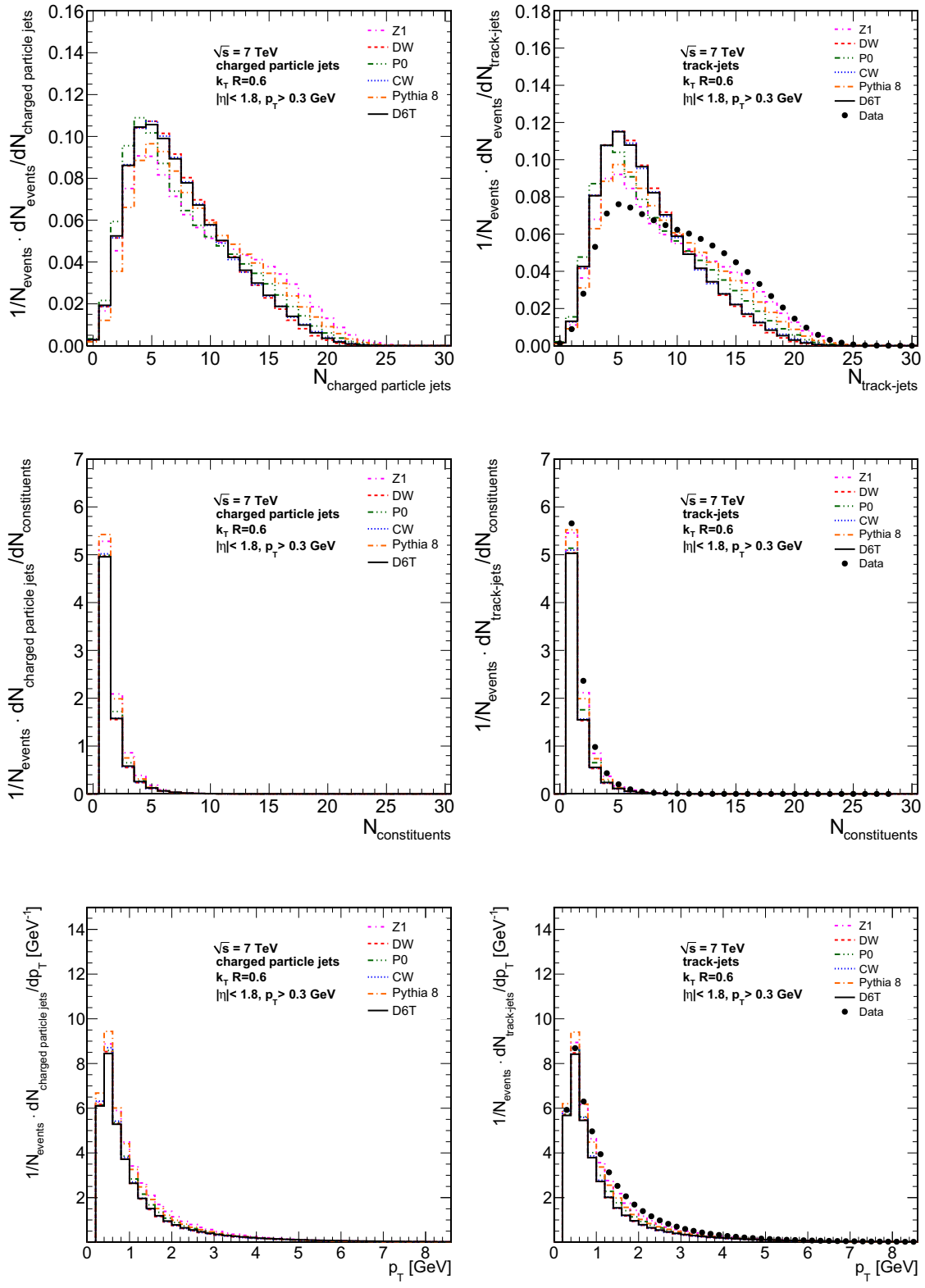


Figure C.6: *Same as Fig. 4.6 with linear y-axes.*

C.3 Systematic Uncertainties

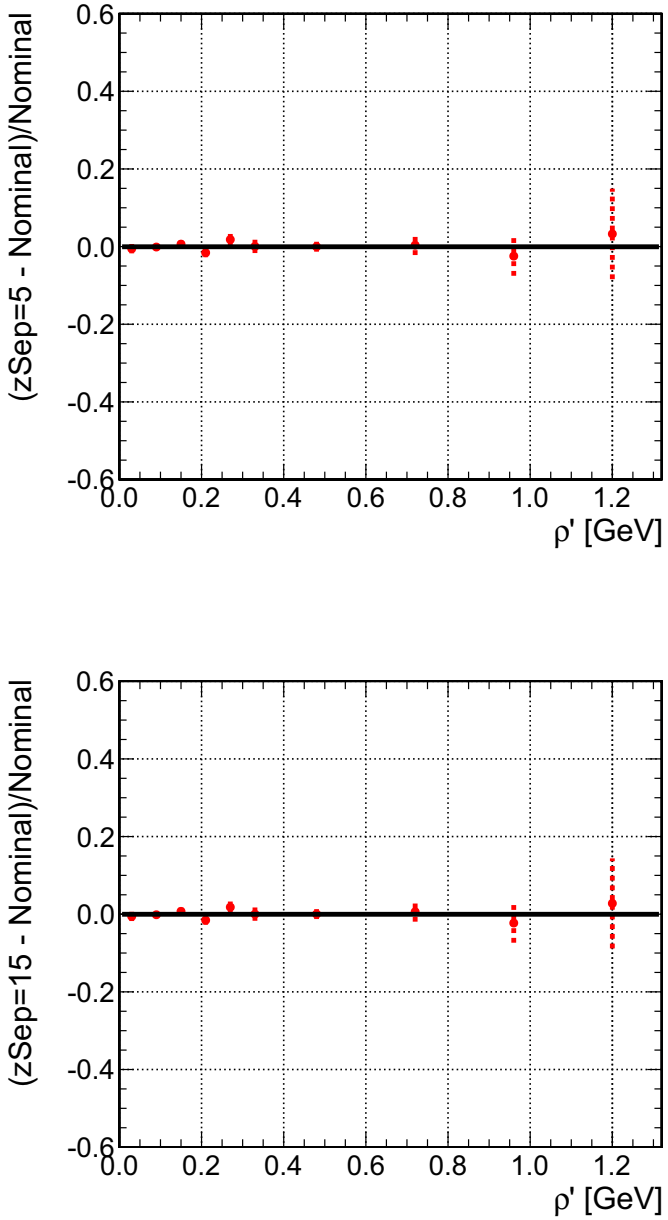


Figure C.7: *Effect of the vertex reconstruction z-seperation requirement at 0.9 TeV. Variations from the default 10 cm to 5 cm (left) and 15 cm (right) were investigated.*

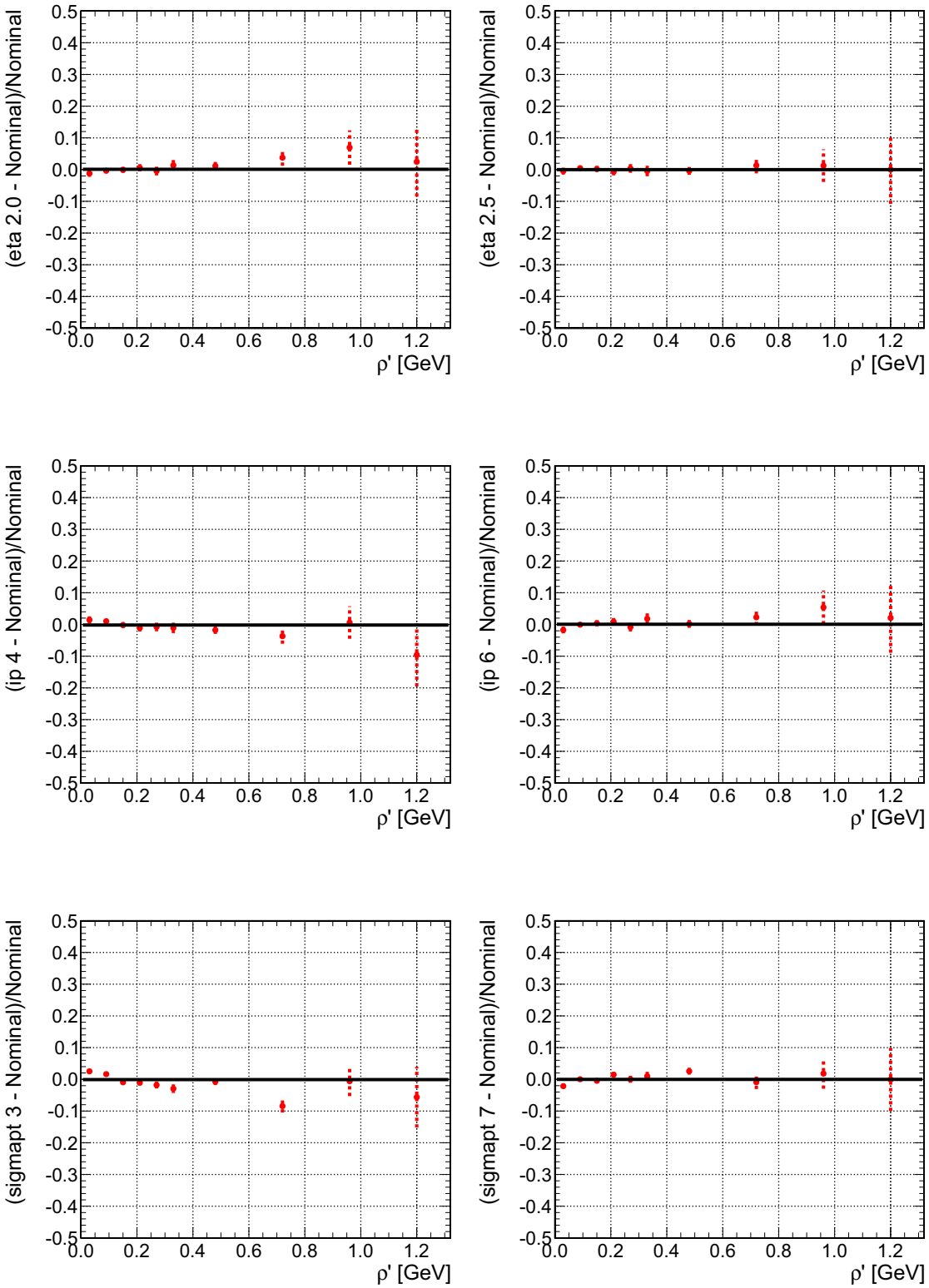


Figure C.8: *Effects of different cut variations on tracks at* 0.9 *TeV.*
Upper row: Variation of pseudo-rapidity cut.
Middle row: Variation of transverse and longitudinal impact parameter.
Lower row: Variation of transverse momentum significance cut.

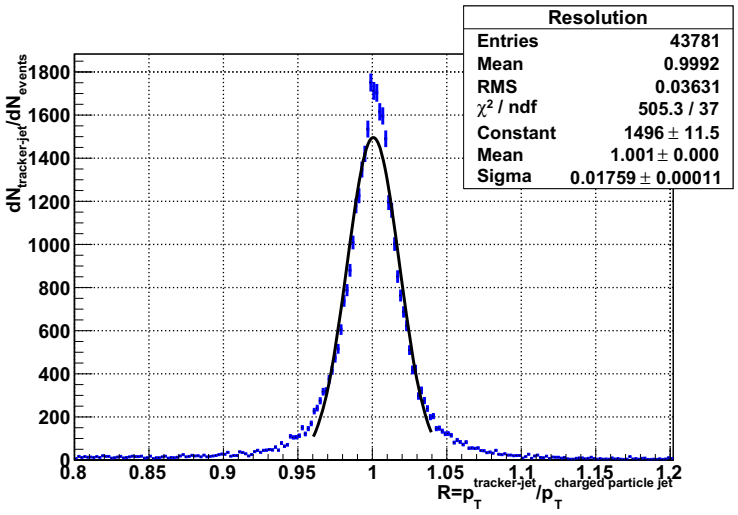


Figure C.9: *Transverse momentum resolution of track-jets.*

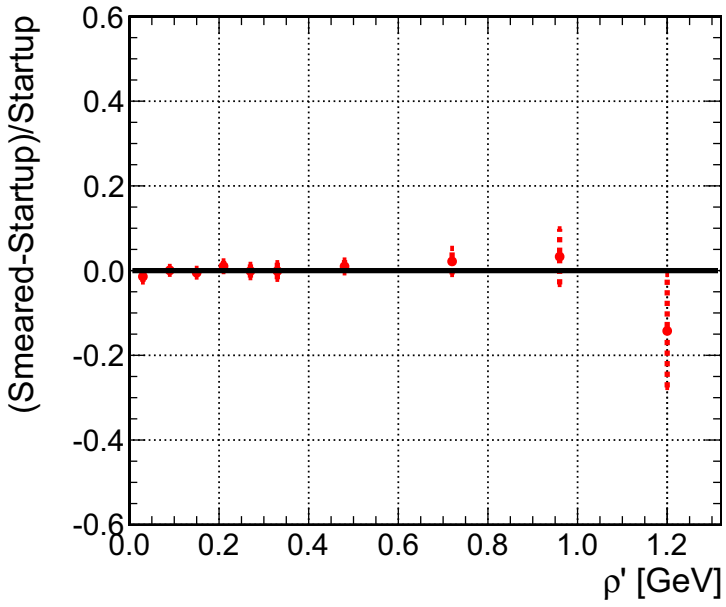


Figure C.10: *Systematic uncertainty derived from artifially smearing the track-jet transverse momentum.*

List of Figures

1.1 Elastic Electron-Proton Scattering . . . 7
1.2 CTEQ6M PDFs for Q=2 GeV and Q=100 GeV . . . 10
1.3 Feynman graphs of the basic QCD interactions . . . 17
1.4 World average of α_s at $M(Z^0)$ and $\alpha_s(Q)$ relation . . . 19
1.5 Drell-Yan process $q\bar{q} \to l\bar{l}$ . . . 20
1.6 Parton kinematics and cross-sections at Tevatron and the LHC . . . 21
1.7 Decomposition of a proton-proton collision . . . 24
1.8 Working principle of a multi-purpose Monte Carlo event generator . . . 25
1.9 Infrared and collinear unsafe behaviour of jet algorithms . . . 28
1.10 Shapes and areas of jets with different algorithms . . . 32

2.1 Schematic view of the LHC accelerator . . . 34
2.2 Perspective view and slice through the CMS detector and its sub-systems . . . 37
2.3 Overview of the CMS silicon tracker and tracker material budget . . . 39
2.4 Overview of the ECAL layout of CMS . . . 42
2.5 Slice through the CMS HCAL . . . 43
2.6 Design overview of the CMS muon system . . . 44
2.7 CMS trigger and DAQ . . . 46
2.8 Tiered structure of the WLCG . . . 47
2.9 Jet corrections in CMS . . . 51

3.1 Inclusive jet cross-section at 10 TeV from Pythia . . . 56
3.2 Systematic uncertainties of the inclusive jet cross section . . . 58
3.3 K-factors for the inclusive jet cross-section . . . 59
3.4 Dependence of the hadronization corrections on MPI . . . 60
3.5 Hadronization and MPI corrections to the inclusive jet cross-section . . . 61
3.6 Overall non-perturbative corrections to the inclusive jet cross-section . . . 62

3.7 Fractional theory uncertainties on the inclusive jet cross-section . . . 63
3.8 Comparison between systematic uncertainties on the pseudo data and the theory side 63
3.9 QCD spectrum from pseudo data . . . 64
3.10 Sensitivity on contact interactions . . . 64
3.11 Non-perturbative corrections and theoretical uncertainties at 7 TeV . . . 65
3.12 Comparison between measured inclusive jet-cross section and theory predictions . 66
3.13 Non-perturbative corrections on the inclusive jet cross-section for CDF II . . . 67
3.14 Inclusive jet spectrum from run II of the DØand CDF experiments . . . 67
3.15 Correlation of directions to leading jet . . . 69
3.16 Average charged multiplicity and sum of transverse momenta at 1.8 TeV . . . 71
3.17 Expected transverse charged density for tunes DW and DWT at the LHC . . . 71
3.18 Results from the measurement of the UE with CMS at 7 TeV and comparison to 0.9 TeV . . . 73
3.19 Jet area over transverse momentum at Monte Carlo level . . . 74
3.20 Average of ρ depending on R and ρ distribution event-by-event . . . 75

4.1 Integrated luminosity over time delivered by the LHC . . . 81
4.2 Fractions of Monte Carlo particles exceeding a minimal transverse momentum . . . 84
4.3 Charged generator particle and track distributions for 0.9 TeV . . . 85
4.4 Charged generator particle and track distributions for 7 TeV . . . 86
4.5 Charged generator jets and track-jets distributions for 0.9 TeV . . . 88
4.6 Charged generator jets and track-jets distributions for 7 TeV . . . 89
4.7 Jet area distributions for 0.9 and 7 TeV . . . 91
4.8 Jet area distributions for different resolution parameters at 7 TeV . . . 92
4.9 Event occupancies for 0.9 and 7 TeV . . . 93
4.10 ρ' for charged generator jets at 0.9 TeV . . . 94
4.11 ρ' for charged generator jets at 7 TeV . . . 95
4.12 Average of the ρ' distributions over jet size parameter . . . 96
4.13 Effect of the tracker material budget uncertainty . . . 97
4.14 Effect of an ideal tracker alignment . . . 98
4.15 Effect of non-operational tracker channels at 0.9 TeV . . . 99
4.16 Effect of non-operational tracker channels at 7 TeV . . . 100
4.17 Effect of the imperfect trigger simulation at 0.9 TeV . . . 101
4.18 Effect of the imperfect trigger simulation at 7 TeV . . . 102
4.19 Effects of adding and subtracting tracks at 0.9 TeV . . . 103
4.20 Effects of adding and subtracting tracks at 7 TeV . . . 103
4.21 Effect of the variation of the transverse momentum cut on tracks at 0.9 TeV . . . 104
4.22 Effect of the variation of the transverse momentum cut on tracks at 7 TeV . . . 104
4.23 Effect of a systematic shift in jet p_{T} at 0.9 TeV . . . 105
4.24 Effect of a systematic shift in jet p_{T} at 7 TeV . . . 105

4.25 ρ' on track-jet level for 0.9 TeV 108
4.26 ρ' on track-jet level for 7 TeV and ratios of Monte Carlo tunes to data 109

C.1 Angular distributions, multiplicity and transverse momentum of charged particles and tracks at 0.9 TeV 126
C.2 Angular distributions, multiplicity and transverse momentum of charged particles and tracks at 7 TeV 127
C.3 Angular distributions of jets 128
C.4 Pseudo-rapidity distributions of jets 129
C.5 Charged generator jets and track-jets distributions for 0.9 TeV with linear y-axes . 130
C.6 Charged generator jets and track-jets distributions for 0.9 TeV with linear y-axes . 131
C.7 Effect of the vertex reconstruction z-seperation requirement 132
C.8 Effects of different cut variations on tracks at 0.9 TeV 133
C.9 Transverse momentum resolution of track-jets 134
C.10 Systematic uncertainty derived from artifially smearing the jet transverse momentum 134

List of Tables

1.1 The three force of the Standard Model and Gravity . . . 11
1.2 The three generations of fermions . . . 11
1.3 The four gauge bosons of the Standard Model . . . 11
1.4 The spin-$1/2$ baryon octuplet and the spin-$3/2$ baryon decuplet . . . 15
1.5 Color representation of quarks . . . 16

2.1 LHC machine parameters at nominal running . . . 36

4.1 The L1 triggers used for the 2009 commissioning data. . . . 78
4.2 Event selection efficiencies for 2009 commissioning data . . . 79
4.3 Pythia 6 processes in minimum bias Monte Carlo samples . . . 80
4.4 Pythia 8 processes in minimum bias Monte Carlo samples . . . 80
4.5 Overview of the systematic uncertainties . . . 106

A.1 Ancient Pythia 6 tunes . . . 115
A.2 Old Pythia 6 tunes . . . 116
A.3 New Pythia 6 tunes . . . 117

B.1 List of runs, luminosity blocks and bunch crossings for 2009 commisiong data . . . 122
B.2 Monte Carlo data sets at 0.9 TeV and 2009 commisioning data set . . . 123
B.3 Good Collision skim and Monte Carlo data sets at 7 TeV . . . 124

Bibliography

[1] ICHEP 2010 Conference
http://www.ichep2010.fr/. 2

[2] M. Cacciari, G. P. Salam, and S. Sapeta, *On the characterisation of the underlying event*, *JHEP* **04** (2010) 065, [0912.4926]. 2, 72, 74, 75

[3] E. Rutherford, *The Scattering of α and β Particles by Matter and the Structure of the Atom*, *Philosophical Magazine* **Series 6, vol. 21** (1911) 669–688. 5

[4] L. de Broglie, *Recherches sur la thorie des quanta*, *Ann. Phys.* **3** (1925) 22. 5

[5] G. Zweig, *An su_3 model for strong interaction symmetry and its breaking; part ii*, . 6

[6] D. H. Coward, H. DeStaebler, R. A. Early, J. Litt, A. Minten, L. W. Mo, W. K. H. Panofsky, R. E. Taylor, M. Breidenbach, J. I. Friedman, H. W. Kendall, P. N. Kirk, B. C. Barish, J. Mar, and J. Pine, *Electron-proton elastic scattering at high momentum transfers*, *Phys. Rev. Lett.* **20** (Feb, 1968) 292–295. 6

[7] P. A. M. Dirac, *The quantum theory of the emission and absorption of radiation*, *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **114** (1927), no. 767 pp. 243–265. 6

[8] N. Mott and H. S. W. Massey, *Theory of Atomic Collision*. Oxford at the Clarendon Press, 1965. 8

[9] M. Breidenbach, J. I. Friedman, H. W. Kendall, E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, L. W. Mo, and R. E. Taylor, *Observed behavior of highly inelastic electron-proton scattering*, *Phys. Rev. Lett.* **23** (Oct, 1969) 935–939. 9

[10] E. D. Bloom, D. H. Coward, H. DeStaebler, J. Drees, G. Miller, L. W. Mo, R. E. Taylor, M. Breidenbach, J. I. Friedman, G. C. Hartmann, and H. W. Kendall, *High-energy inelastic $e-p$ scattering at 6° and 10°*, *Phys. Rev. Lett.* **23** (Oct, 1969) 930–934. 9

[11] J. F. Owens, *Parton distributions in hadrons: The CTEQ program for the global analysis of parton distributions*, . Prepared for Workshop on Deep Inelastic Scattering and QCD (DIS 95), Paris, France, 24-28 Apr 1995. 9

[12] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt, *Parton distributions for the LHC*, `0901.0002`. 9

[13] J. Pumplin *et. al.*, *New generation of parton distributions with uncertainties from global QCD analysis*, *JHEP* **07** (2002) 012, [`hep-ph/0201195`]. 10

[14] C. S. B. Povh, K. Rith and F. Zetsche, *Teilchen und Kerne*. Springer Verlag, 2004. 10

[15] D. J. Griffiths, *Introduction to Elementary Particles*. John Wiley & Sons, Inc., 1987. 10, 12

[16] D. H. Perkins, *Introduction to High Energy Physics*. Cambridge University Press, 2000. 10

[17] G. Dissertori, I. G. Knowles, and M. Schmelling, *High energy experiments and theory*, . Oxford, UK: Clarendon (2003) 538 p. 11, 16, 17

[18] **Particle Data Group** Collaboration, C. Amsler *et. al.*, *Review of Particle Physics*, *Phys. Lett.* **B667** (2008) 1. 11, 13, 14

[19] E. Noether and M. A. Tavel, *Invariant Variation Problems*, *ArXiv Physics e-prints* (Mar., 2005) [`arXiv:physics/0503066`]. 12

[20] R. Feynman, *QED: The Strange Theory of Light and Matter*. Princeton University Press, 1985. 12

[21] E. Aprile and S. Profumo, *Focus on dark matter and particle physics*, *New Journal of Physics* **11** (2009), no. 10 105002. 13

[22] O. Br and U. J. Wiese, *Can one see the number of colors?*, *Nuclear Physics B* **609** (2001), no. 1-2 225 – 246. 14

[23] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, *Phys. Rev. Lett.* **30** (Jun, 1973) 1343–1346. 18

[24] The OPAL Collaboration, *Determination of ing jet rates at lep with the opal detector*, *Eur. Phys. J. C* **45** (2006), no. 3 547–568. 18

[25] K. Nakamura and P. D. Group, *Review of particle physics*, *Journal of Physics G: Nuclear and Particle Physics* **37** (2010), no. 7A 075021. 19, 20, 22

[26] S. Catani, *Aspects of QCD, from the Tevatron to the LHC*, hep-ph/0005233. 21

[27] S. Catani, F. Krauss, B. R. Webber, and R. Kuhn, *Qcd matrix elements + parton showers*, *Journal of High Energy Physics* **2001** (2001), no. 11 063. 23

[28] R. Frederix, *Recent developments in madgraph/madevent v4*, *Nuclear Physics B - Proceedings Supplements* **183** (2008) 285 – 289. Proceedings of the 9th DESY Workshop on Elementary Particle Theory. 23

[29] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [hep-ph/0603175]. 23, 26, 114

[30] G. Marchesini and B. R. Webber, *Simulation of qcd jets including soft gluon interference*, *Nuclear Physics B* **238** (1984), no. 1 1 – 29. 23

[31] P. Z. Skands, *The Perugia Tunes*, 1005.3457. 23

[32] D. Piparo, *Statistical Combination of Higgs Decay Channels and Determination of the Jet-Energy Scale of the CMS Experiment at the LHC*, . 24

[33] Z. Was, *TAUOLA the library for tau lepton decay, and KKMC/KORALB/KORALZ/... status report*, *Nucl. Phys. Proc. Suppl.* **98** (2001) 96–102, [hep-ph/0011305]. 25

[34] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput. Phys. Commun.* **178** (2008) 852, [0710.3820]. 26

[35] G. Marchesini *et. al.*, *HERWIG: A Monte Carlo event generator for simulating hadron emission reactions with interfering gluons. Version 5.1 - April 1991*, *Comput. Phys. Commun.* **67** (1992) 465. 26

[36] M. Bahr *et. al.*, *Herwig++ Physics and Manual*, *Eur. Phys. J.* **C58** (2008) 639, [0803.0883]. 26

[37] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, *Z. Phys.* **C72** (1996) 637, [hep-ph/9601371]. 26

[38] G. Sterman and S. Weinberg, *Jets from Quantum Chromodynamics*, *Phys. Rev. Lett.* **39** (1977) 1436. 27

[39] **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, [0807.2204]. 28, 67

[40] G. P. Salam, *A Practical Seedless Infrared Safe Cone Algorithm*, 0705.2696. 29

[41] M. Cacciari and G. P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, *Phys. Lett.* **B641** (2006) 57, [hep-ph/0512210]. 29

[42] M. Cacciari, G. Salam, and G. Soyez, "FastJet." http://fastjet.fr, 2006. Web Page. 29, 50, 87

[43] G. Soyez, *The SISCone and anti-kt jet algorithms*, 0807.0021. 30

[44] S. Bentvelsen and I. Meyer, *The cambridge jet algorithm: features and applications*, *The European Physical Journal C - Particles and Fields* **4** (1998) 623–629. 10.1007/s100520050232. 31

[45] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, *JHEP* **04** (2008) 005, [0802.1188]. 31, 75, 90

[46] G. P. Salam, *Towards Jetography*, 0906.1833. 32

[47] L. Evans, (ed.) and P. Bryant, (ed.), *LHC Machine*, *JINST* **3** (2008) S08001. 34, 36

[48] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, *JINST* **3** (2008) S08002. 34

[49] LHCb Collaboration, *The LHCb Detector at the LHC*, *JINST* **3** (2008) S08005. 34

[50] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3** (2008) S08003. 35

[51] CMS Collaboration, *CMS Physics Technical Design Report, Volume I: Detector Performance and Software*, vol. LHCC-2006-001 of *Technical Design Report CMS*. CERN, Geneva, 2006. 35, 144

[52] CMS Collaboration, *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004. 37, 39, 43, 44, 46, 47

[53] CMS Outreach
http://cmsinfo.cern.ch/outreach. 37

[54] CMS Collaboration, *Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays*, *JINST* **5** (2010) T03009, [0910.2505]. 36, 148

[55] T. Speer, W. Adam, R. Frhwirth, A. Strandlie, T. Todorov, and M. Winkler, *Track reconstruction in the cms tracker. oai:cds.cern.ch:884424*, Tech. Rep. CMS-CR-2005-014. CERN-CMS-CR-2005-014, CERN, Geneva, Jul, 2005. 40

[56] CMS Collaboration, *CMS Physics Technical Design Report, Volume I: Detector Performance and Software*. Vol. LHCC-2006-001 of Acosta *et. al.* [51], 2006. 41, 42

[57] R. Arcidiacono, *Studies of the cms electromagnetic calorimeter performance in the electron test beam*, Tech. Rep. CMS-CR-2008-099. CERN-CMS-CR-2008-099, CERN, Geneva, Nov, 2008. 41

[58] B. Roland, *Forward Physics Capabilities of CMS with the CASTOR and ZDC detectors*, *ArXiv e-prints* (Aug., 2010) [1008.0592]. 43

[59] C. Eck, J. Knobloch, L. Robertson, I. Bird, K. Bos, N. Brook, D. Dllmann, I. Fisk, D. Foster, B. Gibbard, C. Grandi, F. Grey, J. Harvey, A. Heiss, F. Hemmer, S. Jarp, R. Jones, D. Kelsey, M. Lamanna, H. Marten, P. Mato-Vila, F. Ould-Saada, B. Panzer-Steindel, L. Perini, Y. Schutz, U. Schwickerath, J. Shiers, and T. Wenaus, *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005. 46, 48

[60] G. L. Bayatyan, M. Della Negra, Fo, A. Herv, and A. Petrilli, *CMS computing: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 2005. Submitted on 31 May 2005. 49

[61] GEANT4 Collaboration, *GEANT4: A simulation toolkit*, *Nucl. Instrum. Meth.* **A506** (2003) 250. 49

[62] R. Brun, F. Rademakers, and S. Panacek, *Root, an object oriented data analysis framework*, . 49

[63] CMS Collaboration, *Performance of Jet Algorithms in CMS*, *CMS Physics Analysis Summary* **CMS-PAS-JME-07-003** (2008). 50

[64] CMS Collaboration, *Commissioning of Track-Jets in pp Collisions at* $\sqrt{s} = 7$ *TeV*, *CMS Physics Analysis Summary* **CMS-PAS-JME-10-006** (2010). 50, 82

[65] CMS Collaboration, *Plans for Jet Energy Corrections at CMS*, *CMS Physics Analysis Summary* **CMS-PAS-JME-07-002** (2008). 51

[66] CMS Collaboration, *Offset Energy Correction for Cone Jets*, *CMS Physics Analysis Summary* **CMS-PAS-JME-09-003** (2009). 50

[67] CMS Collaboration, *Determination of the Relative Jet Energy Scale at CMS from Dijet Balance*, *CMS Physics Analysis Summary* **CMS-PAS-JME-08-003** (2009). 51

[68] CMS Collaboration, *Calibration of the absolute jet energy scale with* $Z(\rightarrow \mu^+\mu^-)+$ *jet events at CMS*, *CMS Physics Analysis Summary* **CMS-PAS-JME-09-009** (2009). 51

[69] Esen, S. and Landsberg, G., *MC Truth L4 EMF-Based Factorized Jet Corrections in CMS*, *CMS Analysis Note* (2009). 51

[70] Cammin, J., *Flavor Specific Jet Energy Corrections and Closure Tests for Factorized Jet Energy Corrections*, *CMS Analysis Note* (2009). 51

[71] M. Heinrich *et. al.*, *Non-perturbative Corrections to Inclusive Jet Spectra at 10 TeV*, . 51, 54, 55, 118

[72] CMS Collaboration, *Jet Corrections to Parent Parton Energy*, *CMS Physics Analysis Summary* **CMS-PAS-JME-08-002** (2008). 51

[73] CMS Collaboration, *Initial Measurement of the Inclusive Jet Cross Section at* $10\,\mathrm{TeV}$ *with CMS*, *CMS Physics Analysis Summary* **CMS-PAS-QCD-08-001** (2009). 53, 55

[74] CMS Collaboration, *Measurement of the Inclusive Jet Cross Section in pp Collisions at 7 TeV*, *CMS Physics Analysis Summary* **CMS-PAS-QCD-10-011** (2010). 53, 65, 66

[75] S. Frixione and B. R. Webber, *Matching NLO QCD computations and parton shower simulations*, *JHEP* **06** (2002) 029, [hep-ph/0204244]. 54

[76] Z. Nagy, *Three-jet cross sections in hadron hadron collisions at next-to-leading order*, *Phys. Rev. Lett.* **88** (2002) 122003, [hep-ph/0110315]. 54

[77] M. Heinrich *et. al.*, *NLO Calculations for Inclusive Jets with fastNLO at 10 TeV*, *CMS Analysis Note* (2009). 54, 55

[78] **CDF** Collaboration, T. Aaltonen *et. al.*, *Measurement of* b*-jet Shapes in Inclusive Jet Production in* $p\bar{p}$ *Collisions at* $\sqrt{s}$ *= 1.96-TeV*, *Phys. Rev.* **D78** (2008) 072005, [0806.1699]. 54

[79] M. Heinrich *et. al.*, *Inclusive Jet Cross Section at* 10 *TeV with CMS*, *CMS Analysis Note* (2009). 55

[80] S. Frixione and B. R. Webber, *The MC@NLO event generator*, hep-ph/0207182. 57

[81] S. Frixione and B. R. Webber, *The MC@NLO 3.3 event generator*, hep-ph/0612272. 57

[82] Z. Nagy, *Next-to-leading order calculation of three-jet observables in hadron-hadron collision*, *Phys. Rev.* **D68** (2003) 094002, [hep-ph/0307268]. 54

[83] T. Kluge, K. Rabbertz, and M. Wobisch, "fastNLO — fast pQCD calculations for hadron-induced processes." http://hepforge.cedar.ac.uk/fastnlo, 2005. Web Page. 57

[84] CMS Collaboration, *Dijet Azimuthal Decorrelations and Angular Distributions in pp Collisions at sqrt(s) = 7 TeV*, *CMS Physics Analysis Summary* **CMS-PAS-QCD-10-011** (2010). 60

[85] C. Buttar *et. al.*, *Standard Model Handles and Candles Working Group: Tools and Jets Summary Report*, 0803.0678. 62

[86] **CDF** Collaboration, F. Abe *et. al.*, *Measurement of the inclusive jet cross-section in $\bar{p}p$ collisions at $\sqrt{s} = 1.8$ TeV*, *Phys. Rev. Lett.* **62** (1989) 613. 67

[87] **CDF II** Collaboration, A. Abulencia *et. al.*, *Measurement of the inclusive jet cross section using the k(t) algorithm in p anti-p collisions at $\sqrt{s} = 1.96$- TeV*, *Phys. Rev. Lett.* **96** (2006) 122001, [hep-ex/0512062]. 67

[88] **CDF** Collaboration, A. Abulencia *et. al.*, *Measurement of the inclusive jet cross section in $p\bar{p}$ interactions at $\sqrt{s} = 1.96$-TeV using a cone-based jet algorithm*, *Phys. Rev.* **D74** (2006) 071103, [hep-ex/0512020]. 67

[89] **D0** Collaboration, V. M. Abazov *et. al.*, *Determination of the strong coupling constant from the inclusive jet cross section in ppbar collisions at sqrt(s)=1.96 TeV*, *Phys. Rev.* **D80** (2009) 111107, [0911.2710]. 67

[90] **D0** Collaboration, V. M. Abazov *et. al.*, *Measurement of the inclusive jet cross section in $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$*, *Phys. Rev. Lett.* **101** (2008) 062001, [0802.2400]. 67

[91] **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, [hep-ex/0109041]. 67

[92] 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, [hep-ph/0611148]. 67

[93] **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, [hep-ex/0701051]. 67

[94] **TeV4LHC QCD Working Group** Collaboration, M. G. Albrow *et. al.*, *Tevatron-for-LHC report of the QCD working group*, hep-ph/0610012. 68, 69

[95] **CDF Collaboration** Collaboration, T. Affolder *et. al.*, *Charged jet evolution and the underlying event in proton-antiproton collisions at 1.8 tev*, . 69

[96] G. Webb, *Evaluation of the Underlying Event in Pp Collisions at vs = 200 GEV at Star*, *APS Meeting Abstracts* (Oct., 2009) L9+. 70

[97] CDF Collaboration, *Charged jet evolution and the underlying event in $p\bar{p}$ collisions at 1.8 TeV*, *Phys. Rev.* **D65** (2002) 092002. 70, 71, 113

[98] P. Z. Skands, *The Perugia Tunes*, 0905.3418. 70

[99] R. D. Field, *Predicting "Min-Bias" and the "Underlying Event" at the LHC*, *Minute Particulars & Hidden Symmetries: Chris Quigg Symposium* (Dec., 2009). 71

[100] CMS Collaboration, *First Measurement of the Underlying Event Activity in Proton-Proton Collisions at 900 GeV at the LHC*, *CMS Physics Analysis Summary* **CMS-PAS-QCD-10-001** (2010) [1006.2083]. 72, 78, 79, 84, 122

[101] CMS Collaboration, *Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV*, *CMS Physics Analysis Summary* **CMS-PAS-QCD-10-010** (2010). 72, 73, 81

[102] M. Cacciari and G. P. Salam, *Pileup subtraction using jet areas*, *Phys. Lett.* **B659** (2008) 119, [0707.1378]. 74

[103] CMS Collaboration, *Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 0.9$ and 2.36 TeV*, *JHEP* **02** (2010) 041, [1002.0621]. 78, 98, 99

[104] CMS Collaboration, *Tracking and Vertexing Results from First Collisions*, *CMS Physics Analysis Summary* **CMS-PAS-TRK-10-001** (2010). 81, 82

[105] CMS Collaboration, *Transverse-momentum and pseudorapidity distributions of charged hadrons in pp collisions at $\sqrt{s} = 7$ TeV*, 1005.3299. 84

[106] CMS Collaboration, *Alignment of the CMS Silicon Tracker during Commissioning with Cosmic Rays*, in *JINST* [54], p. T03009, [0910.2505]. 96

[107] C. Buttar *et. al.*, *Les houches physics at TeV colliders 2005, standard model and Higgs working group: Summary report*, hep-ph/0604120. 117

[108] A. Buckley, H. Hoeth, H. Lacker, H. Schulz, and J. E. von Seggern, *Systematic event generator tuning for the LHC*, *Eur. Phys. J.* **C65** (2010) 331, [0907.2973]. 117

[109] R. Field and C. R. Group, *Pythia tune a, herwig, and jimmy in run 2 at cdf*, hep-ph/0510198. 118

[110] M. Bahr, S. Gieseke, and M. H. Seymour, *Simulation of multiple partonic interactions in Herwig++*, *JHEP* **07** (2008) 076, [0803.3633]. 118

[111] M. Bahr *et. al.*, *Herwig++ 2.3 Release Note*, 0812.0529. 118

[112] Grid-Control https://ekptrac.physik.uni-karlsruhe.de/trac/grid-control/. 121

[113] M. De Gruttola, S. Di Guida, D. Futyan, F. Glege, G. Govi, V. Innocente, P. Paolucci, P. Picca, A. Pierro, D. Schlatter, and Z. Xie, *Persistent storage of non-event data in the CMS databases*, *Journal of Instrumentation* **5** (Apr., 2010) 4003–+, [1001.1674]. 123

Danksagung

Mein erster und bedeutendster Dank gebührt Herrn Professor Günter Quast, für die Betreuung dieser Arbeit, seine Geduld, seinen Einsatz für seine Gruppe und seine Bereitschaft, sein unerschöpfliches Wissen zu teilen. Er ermöglichte mir auch meinen Aufenthalt am CERN, der eine wichtige Erfahrung für mich war.
Desweiteren möchte ich Herrn Professor Wim de Boer danken für die Übernahme des Korreferats.

Besonderer Dank gebührt auch Dr. Klaus Rabbertz ohne den diese Arbeit in dieser From nicht möglich gewesen wäre. Seine Kommentare, Anleitungen und seine Motivation haben mich enorm vorangebracht.

Vielen Dank für die hervorragende Zusammenarbeit an Dr. Danilo Piparo und Dr. Andreas Oehler, von denen ich unwahrscheinlich viel lernen konnte.

Danke auch an alle Kollegen mit denen ich das Büro teilte und die auf die eine oder andere Art zu dieser Arbeit beitrugen, sei es durch technische Unterstützung oder durch Korrekturlesen. Im einzelnen sind das Dr. Volker Büge, Dr. Armin Scheurer, Christophe Saout, Oliver Oberst, Kristov Hackstein, Joram Berger, Fred Stober, Thomas Hauth, Manuel Zeise, David Kernert und Stephan Riedel.

Natürlich möchte ich auch allen Mitgliedern des EKP danken, die ich hier nicht namentlich genannt habe, für die Bereitstellung und den Unterhalt der Computerinfrastruktur und die tolle Atmosphäre. Besonderer Dank geht an das Graduiertenkolleg “Hochenergiephysik und Teilchenastrophysik”, für die Bereitstellung der finanziellen Mittel sowie natürlich auch allen Kollegen in der CMS Kollaboration.

Ganz besonders möchte ich meiner Familie und meiner Freundin danken, die mich stets unterstützt und früh verstanden haben, dass “Teilchenphysiker” in der Tat ein Beruf ist. Außerdem natürlich meinen Freunden danke für die gute Zeit.