Higher Order QCD Radiation in Top Pair Production with the CMS Detector

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von
Diplom-Physiker Alexander Floßdorf
aus Essen

Hamburg
2009

Gutachter der Dissertation:	Prof. Dr. Joachim Mnich Priv.-Doz. Dr. Hannes Jung
Gutachter der Disputation:	Prof. Dr. Joachim Mnich Jun.-Prof. Dr. Johannes Haller
Datum der Disputation:	31. August 2009
Vorsitzender des Prüfungsausschusses:	Dr. Georg Steinbrück
Vorsitzender des Promotionsausschusses:	Prof. Dr. Robert Klanner
Leiter des Departments Physik:	Prof. Dr. Joachim Bartels
Dekan der MIN-Fakultät:	Prof. Dr. Heinrich Graener

Abstract

The Large Hadron Collider at CERN will collide protons with a centre-of-mass energy of up to $\sqrt{s} = 14\,\text{TeV}$, thereby offering the opportunity to explore a wide range of physics topics.

In this thesis the effects of QCD radiation in top pair events are examined. Due to the large top mass, top pairs are well suited for an investigation of gluon emissions. An extensive study comparing different radiation models implemented in Monte Carlo event generators is presented. The transverse momentum distribution of the $t\bar{t}$ system is rather sensitive to radiation influences and therefore analysed in detail. As hard emissions can be associated with jets, a thorough investigation of these jets is performed. The transverse momentum of hard jets and the rapidity distribution of the hardest jet in the $t\bar{t}$ rest frame are examined.

Moreover an analysis of samples incorporating different radiation models after the full CMS detector simulation is presented, studying the same observables as on generator level. The potential of the CMS experiment to distinguish between different models is estimated and a method to obtain the underlying transverse momentum distribution of the $t\bar{t}$ system is described.

Zusammenfassung

Am Large Hadron Collider am CERN werden Protonen bei einer Schwerpunktsenergie von bis zu $\sqrt{s} = 14\,\text{TeV}$ zur Kollision gebracht. Dadurch wird die Erforschung vielfältiger neuer Physikprozesse ermöglicht.

In dieser Arbeit werden die Eigenschaften von QCD-Strahlung in Top-Paarereignissen untersucht. Aufgrund der hohen Masse des Top-Quarks eignen sich Top-Paare besonders gut für die Analyse von Gluon-Abstrahlungen. Es wird eine detaillierte Studie präsentiert, die verschiedene in Monte-Carlo-Generatoren implementierte Modelle miteinander vergleicht. Aufgrund der Sensitivität auf Strahlungseinflüsse wird insbesondere die Transversalimpulsverteilung des Top-Paares untersucht. Es ist möglich harte Abstrahlungen mit Jets zu assoziieren. Sowohl der Transversalimpuls solcher harten Jets als auch die Rapidität des führenden Jets im Ruhesystem des $t\bar{t}$-Systems werden ausführlich diskutiert.

Des Weiteren wird eine Analyse verschiedener Modelle nach der vollständigen CMS-Detektorsimulation durchgeführt, wobei die gleichen Observablen wie in der Generatorstudie betrachtet werden. Daraus wird das Potential des CMS-Experimentes abgeschätzt, zwischen den verschiedenen Modellen unterscheiden zu können. Außerdem wird eine Methode vorgestellt, die es ermöglicht, die zugrunde liegende Transversalimpulsverteilung des $t\bar{t}$-Systems zu bestimmen.

Contents

Chapter 1

Introduction

1.1 Overview

In this thesis higher order QCD radiation in top pair events at the LHC is examined. Different physics models implemented in Monte Carlo event generators are compared. Therefore the most important observable is the transverse momentum of the $t\bar{t}$ system. Furthermore jets associated with radiated gluons are investigated.

In this chapter the context associated with the performed analysis is presented. First the experimental environment is introduced, followed by a brief description of the actual state of knowledge. Then the latest achievements concerning the top quark are summarised.

Following, in chapter 2 some basic properties of the theory of QCD which builds the foundation of Monte Carlo event generators are explained. Then some aspects of proton collisions needed for the further analysis are introduced in chapter 3. In chapter 4 Monte Carlo event generators are described with a focus on different models for gluon radiation. In chapter 5 a detailed comparison of different physics models that are implemented in the various Monte Carlo generators is presented. Top pair events are examined with emphasis on the transverse momentum of the $t\bar{t}$ system and the properties of additional jets associated with gluon emissions.

In chapter 6 the attention is drawn to the CMS detector. Reconstructed physics objects are introduced in chapter 7 and the selection and reconstruction of semi-muonic $t\bar{t}$ events produced with different Monte Carlo models is described in chapter 8. In chapter 9 the reconstructed top pairs are then analysed and a comparison of the observables examined in chapter 5 on generator level is performed. Furthermore the potential of CMS to distinguish between different models and to obtain the underlying distribution of the transverse momentum of the $t\bar{t}$ system is discussed.

1.2 Large Hadron Collider

The *Large Hadron Collider* (LHC) [1, 2, 3] is a proton–proton ring collider built by the *European Laboratory for Particle Physics CERN* (Conseil Européen pour la Recherche Nucléaire). The 27 km long accelerator is located about 100 m underground at the French-Swiss border near Geneva in the old *Large Electron Positron Collider* (LEP) tunnel [4, 5].

The LHC is designed to collide protons with an energy of 7 TeV each, thus a centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ will be achieved. At this energy the total inelastic proton–proton cross section reaches $\sigma \approx 125\,\text{mb}$ [6]. Collisions will take place with a frequency of 40 MHz so proton bunches will meet every 25 ns. The design luminosity is $\mathcal{L} = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, but in the first years the LHC will be operated at a lower luminosity of about $\mathcal{L} = 2 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$. During this *low luminosity phase* it is planned to collect several $10\,\text{fb}^{-1}$ of data.

Figure 1.1 shows the cross section for several processes over the centre-of-mass energy for p–p and p–$\bar{p}$ collisions. The regions of the proton–anti-proton collider *Tevatron* [7] and the LHC are marked by dashed lines. As the Tevatron is operated at $\sqrt{s} = 1.96\,\text{TeV}$, the centre-of-mass energy available at the LHC is increased relatively by a factor of seven. Even though the total cross section does not change very much over this energy range, the cross sections for several physics processes increase significantly, e. g. for top production. In combination with the high luminosity this results in an amount of collected data much higher than for any former experiments. High statistics and the high centre-of-mass energy offer the opportunity for many interesting studies. Besides proton interactions lead–lead collisions with a centre-of-mass energy of $\sqrt{s} = 5.5\,\text{TeV}$ and a luminosity of $L = 10^{27}\,\text{cm}^{-2}\text{s}^{-1}$ will be investigated.

The proton beams collide at four interaction points, that are shown in figure 1.2. *ATLAS* (A Toroidal LHC Apparatus) [9] and *CMS* (Compact Muon Solenoid) [10] are two multi purpose detectors for exploring a wide range of physics topics. A detailed description of the CMS detector is given in chapter 6. Besides, *LHCb* (Large Hadron Collider beauty) [11] is an experiment specialised on physics connected with B-hadrons while *ALICE* (A Large Ion Collider Experiment) [12] is a detector specially designed for heavy ion physics. Furthermore two experiments for forward physics and the measurement of the total cross section are located next to ATLAS and CMS. *LHCf* (Large Hadron Collider forward) [13] is built close to ATLAS while *TOTEM* (TOTal Elastic and diffractive cross section Measurement) [14] is built close to CMS.

1.3 Standard Model of Particle Physics

The goal of particle physics is a consistent description of all fundamental particles and their interactions. The present state of knowledge is summarised in the *standard model of particle physics*. It describes the properties of the known fundamental particles and their interactions by *quantum field theories* [16].

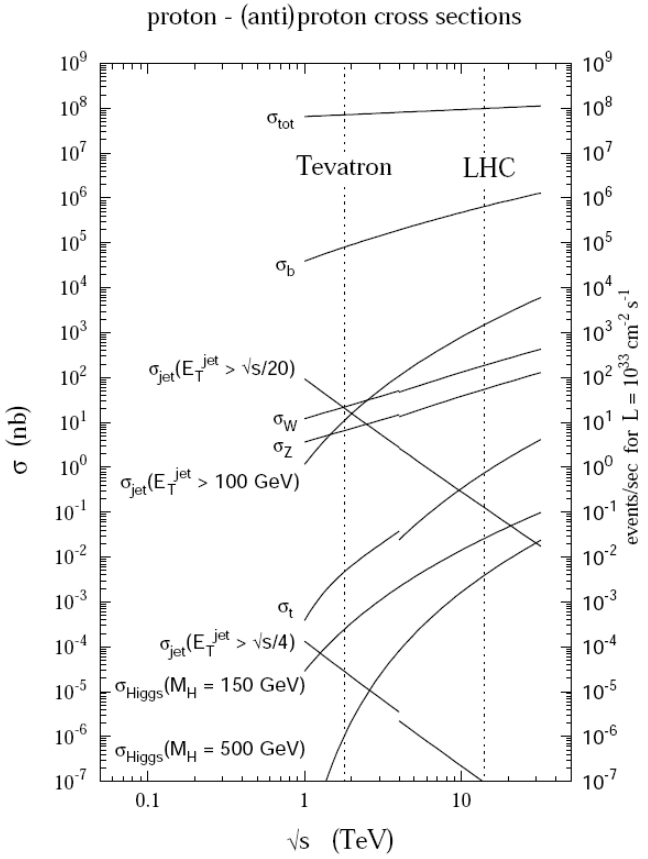


Figure 1.1: Cross sections for different processes over the centre-of-mass energy $\sqrt{s}$ in the hadron–hadron system. [8]

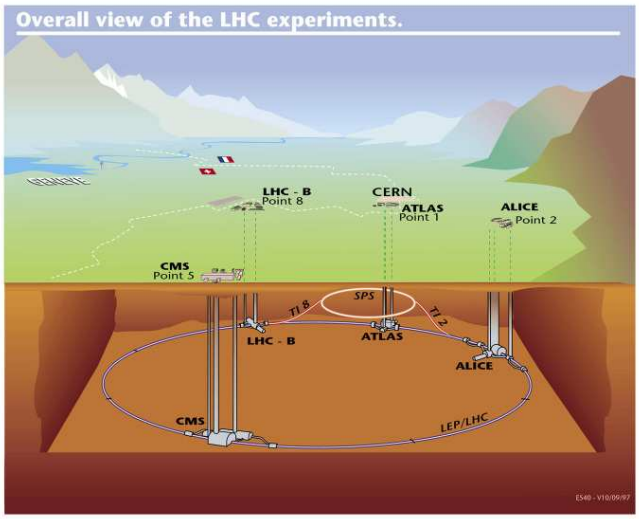


Figure 1.2: Sketch of the LHC ring indicating the location of the four main experiments: ATLAS, ALICE, CMS and LHCb. [15]

The elementary particles known today are six *leptons*, six *quarks* and their *anti-particles*. The leptons are *electron* (e), *muon* (μ), *tau* (τ) and three *neutrinos* (ν). The quarks are called *up*, *down*, *charm*, *strange*, *top* and *bottom* and are symbolised by the first letter of their name. All of these particles are fermions, i. e. they have *spin* $\frac{1}{2}$. Leptons and quarks can be arranged in three so-called *families*.
Leptons:

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix} \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix} \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix} .$$

Quarks:

$$\begin{pmatrix} u \\ d \end{pmatrix} \begin{pmatrix} c \\ s \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix} .$$

The particles of the second and third doublets differ from the ones in the first doublet only in their masses. These masses vary by several orders of magnitude. For the neutrinos only upper limits can be determined but from neutrino oscillation experiments it is known that neutrinos have a non-vanishing mass [17].

The particles can interact with each other via four different forces. The standard model includes the description of the *electromagnetic*, *weak* and *strong force*. The former two interactions can be unified in the *electroweak* theory [18, 19, 20]. The interactions are mediated by *gauge bosons*: the *photon* for the electromagnetic force, W- and Z-*bosons* for the weak force and *gluons* for the strong force. The particles are affected by *gravity* as well but up to now it is not possible to to combine all four interactions consistently into one model. At energies available at colliders gravity is very weak compared to the other forces. Therefore it can be safely neglected for a proper treatment of elementary particles.

For the electromagnetic force the coupling strength is proportional to the particles' electric charge. In units of the elementary charge $e = 1.6 \cdot 10^{-19}\,\mathrm{C}$, e, μ and τ have a charge of -1 whereas the neutrinos are neutral. Hence they do not participate in electromagnetic interactions. Up-type quarks have a charge of $+\frac{2}{3}$ and down-type quarks $-\frac{1}{3}$. The respective anti-particles have the opposite electric charges.

All leptons and quarks participate in weak interactions. It turns out that the quarks are not eigenstates of the weak interaction, thereby transitions between the three generations become possible. These transitions are governed by the *CKM matrix* [21, 22] that allows to compute the transition probability for each up-type and down-type quark combination. The electroweak theory induces a relation between two quarks and two leptons in a way that they build *weak isospin* doublets according to the notation above. The weak force is mediated by massive Z-, W^+ and W^--bosons. The two types of W-bosons are identical except for their electric charge which is ± 1, respectively. In contrast to photons and Z-bosons, W-bosons carry electric and weak charge, hence photons and Z-bosons couple to them.

The strong force affects only quarks. The corresponding charge is called *colour* and appears in three different states. The strong force is mediated by eight types

of gluons. Gluons carry colour charge themselves and hence self-interactions are possible. A brief introduction into the theory describing the strong interactions is given in chapter 2.

The formulation of the standard model discussed so far does not include any masses. It is possible to introduce particle masses via the *Higgs mechanism* [23], which requires a new particle: the *Higgs boson*. This particle has not been observed in former experiments and the search for the Higgs boson is one of the most important tasks for the LHC.

Several experimental measurements indicate the limitations of the standard model. For example the matter described by the standard model accounts for only a few percent of the existing energy in the universe [24]. Furthermore a matter–anti-matter asymmetry is observed in the universe that is not consistent with the actual models. Finally, it is not possible to unify the different forces and to include gravity into one consistent theoretical description. These observations lead to the development of several extended theories covering physics beyond the standard model. The energy range available at the LHC will offer the opportunity to check many of these models.

1.4 Top Quark Physics

The top quark was discovered at the Tevatron in 1995 [25, 26]. With a mass of $m_t = 171.2 \pm 2.1\,\text{GeV}$ [27] it is almost as heavy as a gold atom, thereby being the heaviest known elementary particle. Its mass is well above Λ_{QCD} and hence it is possible to describe top quark production and decay by perturbative means. Due to its large mass the top can decay electroweakly in an on-shell W-boson accompanied by a down-type quark. The CKM matrix element V_{tb} is close to one and predicts the decay

$$t \to Wb$$

to happen in over 99% of all cases. This decay process induces a proper lifetime of the top of about $5 \cdot 10^{-25}\,\text{s}$. As hadronization takes place at time scales of order $3 \cdot 10^{-24}\,\text{s}$, the top quark decays before the fragmentation process starts. Consequently no bound top states are built and the decay products of a bare quark can be investigated. [28]

In figure 1.3 the leading order Feynman diagrams for top pair production at hadron colliders are shown. In all cases the propagators are proportional to $p_T^2 + m_t^2$ [29], therefore the top mass is an important parameter setting the scale for the production process. At the LHC about 90% of the top pairs are expected to be produced via gluon–gluon fusion while only 10% are caused by quark–anti-quark annihilation [28]. The next-to-leading order cross section for $t\bar{t}$ production is predicted to be $908^{+82}_{-85}(\text{scales})^{+30}_{-29}(\text{PDFs})\,\text{pb}$ [30]. At the LHC about one $t\bar{t}$ pair per second will be produced at the low luminosity phase.

Top pairs decay according to

$$t\bar{t} \to W^+W^-b\bar{b}$$

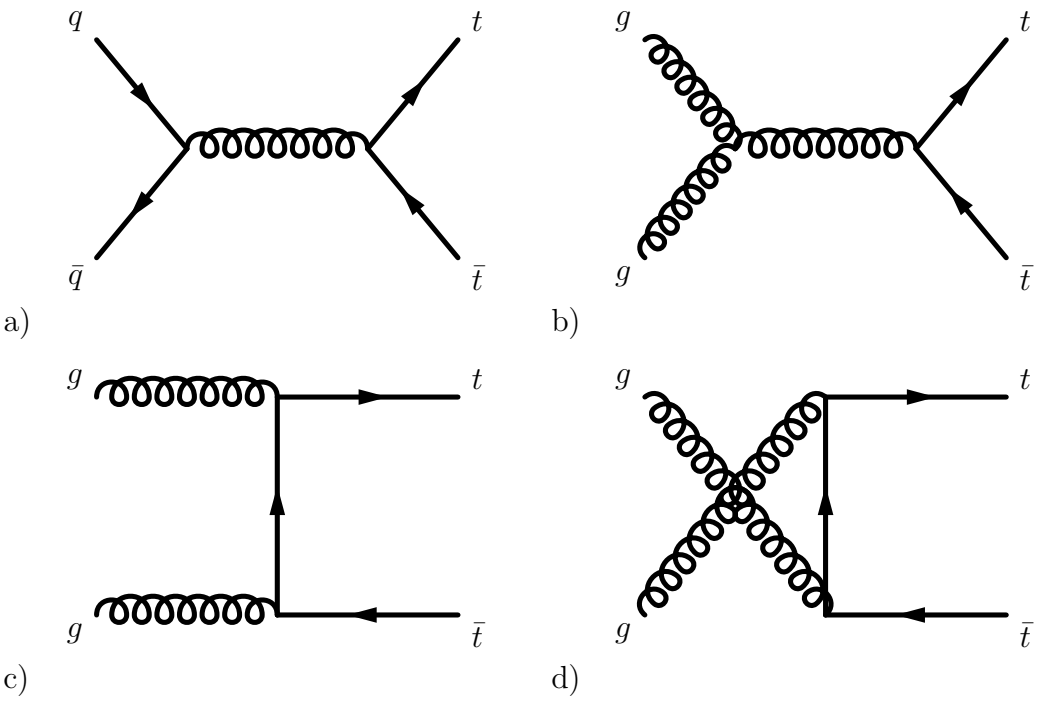


Figure 1.3: Leading order Feynman diagrams for $t\bar{t}$ production. a) shows the diagram for $q\bar{q}$ annihilation while b) – d) depict the s-, t- and u-channel for gluon fusion, respectively.

and are classified by the decays of the W-bosons. In case both W-bosons decay into $q\bar{q}'$ the decay channel is called *fully hadronic*, for both W-bosons decaying leptonically ($W \rightarrow l\nu$) the channel is called *di-leptonic*. If one W-boson decays leptonically and one hadronically the decay is called *semi-leptonic*. The branching ratios are influenced by the fact that each quark can appear in three different colour states, leading to branching fractions of 10.5%, 43.5% and 45.5% for di-leptonic, semi-leptonic and fully hadronic decays, respectively [28]. In this analysis the *semi-muonic* channel with a branching ratio of 14.5% is investigated. Here one of the W-bosons decays according to $W \rightarrow \mu\nu$ while the other one decays hadronically. This decay provides a clear signal and a relatively high reconstruction efficiency and background rejection rate are achievable (compare chapter 8).

In 2009 the observation of single-top quarks at the Tevatron was reported [31, 32]. The production process of single-tops is directly sensitive to the CKM matrix element V_{tb} and probes a different kinematic region, thereby complementing the knowledge about top quarks.

At the LHC a lot of studies will be performed determining the quantum numbers of the top directly. Furthermore discrepancies between theoretical predictions and experimental results open the possibility to discover new physics beyond the standard model. Due to its large mass the top quark plays an important role in the exploration of the high energy scales the LHC will reach.

Among many other observables the differential cross sections of top quark pairs will be of high interest. The transverse momentum of a top pair is caused by effects beyond leading order and thus higher order contributions can be examined. Besides, the transverse momentum spectrum offers the possibility to look for physics beyond the standard model as additional couplings may influence the transverse momentum distribution. In this thesis the transverse momentum of $t\bar{t}$ pairs is studied in detail.

Chapter 2

Quantum Chromodynamics

Quantum Chromodynamics (*QCD*) describes the interaction of quarks and gluons mediated by the strong force. In this chapter the fundamental properties of QCD are summarised. A detailed discussion of all aspects described in this chapter can be found in [29] and [33].

2.1 QCD Lagrangian and Feynman Rules

QCD is a non-abelian Yang-Mills theory describing the interaction of quarks and gluons. Each quark appears in three different charged states, called colour. The properties of coloured particles are described by the non-abelian group SU(3), which is characterised by eight generators. Demanding local gauge invariance of the quark fields introduces eight gauge fields, the gluons. The non-abelian algebra of SU(3) leads to non-vanishing structure constants, reflecting that the gauge fields are colour-charged themselves. As a consequence gluon self-interactions take place giving rise to three and four gluon vertices in the theory.

The Lagrangian density for the quarks is:

$$\mathcal{L}_{quark} = \sum_{flavours} \bar{q}_j(x) \left[i\partial\!\!\!/ - m \right] q_j(x) + g_s T^a_{jk} \bar{q}_j(x) A\!\!\!/^a q_k(x) \tag{2.1}$$

with the following denotations:

- q: quark of a certain flavour,
- j, k: indices running over all colours,
- m: quark mass,
- g_s: strong coupling constant,
- T: generator of the group SU(3),
- a: index running over all generators,
- A: gluon field.

The terms describe a free propagating coloured quark with mass m and the interaction of a quark with a gluon field. For the gauge field the Lagrangian is:

$$\mathcal{L}_{gauge} = -\frac{1}{2}\left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right)\left(\partial^\mu A^{\nu a} - \partial^\nu A^{\mu a}\right) + g_s f_{abc}(\partial_\mu A^a_\nu) A^{b\mu} A^{c\nu} - \frac{g_s^2}{4} f_{abc} f_{ade} A^{b\mu} A^{c\nu} A^d_\mu A^e_\nu \quad (2.2)$$

with

- f_{abc}: structure constants of SU(3).

The first term describes a free propagating gluon while the other two terms represent three and four gluon self-interaction vertices, respectively.

The requirement of renormalizability forbids terms of mass dimension higher than four while local gauge invariance forbids all other terms except one that does not appear in the Lagrangians above. Today it is not clear whether this so called θ term contributes, but even if it does, it will not give rise to any perturbative physics, so it can be safely neglected. [33]

With the classical QCD Lagrangian $\mathcal{L}_{classical} = \mathcal{L}_{quark} + \mathcal{L}_{gauge}$ problems are encountered when formulating the Feynman rules. Additional terms have to be included to fix the gauge and for the choice of a covariant gauge *ghosts* have to be introduced to theory. Ghosts cancel unphysical polarization states of gluons in loop diagrams and ensure correct gauge-invariant results of all calculations. The gluon propagator now depends on an arbitrary gauge parameter ξ which changes the calculations but leaves all physical observables unchanged. Thus the complete QCD Lagrangian is:

$$\mathcal{L}_{QCD} = \mathcal{L}_{classical} + \mathcal{L}_{gaugefix+ghost}\,. \quad (2.3)$$

From this Lagrangian Feynman rules can be deduced which describe propagating quarks and gluons, a quark–gluon vertex and three and four gluon vertices. It is worth noting that the coupling g_s is the same for quark–gluon and gluon–gluon couplings (as well as for ghost–gluon interactions).

Quarks are associated with three different colour states, labled *red* (r), *green* (g) and *blue* (b), while anti-quarks are associated with anti- colour states, *anti-red* ($\bar{r}$), *anti-green* ($\bar{g}$) and *anti-blue* ($\bar{b}$). An examination of the SU(3) generators shows that gluons can be regarded as particles carrying a combination of colour and anti-colour.

A more detailed description of the fundamentals of QCD can be found in [29] and [33]. A complete treatment of the Lagrangian and deduction of the Feynman rules is performed in [16].

2.2 Renormalization

The theory describing the interaction of fundamental fermions by Feynman rules is referred to as *perturbative QCD*. Quarks can build composed state which are

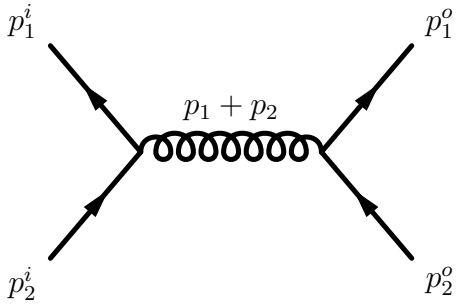


Figure 2.1: Leading order Feynman diagram.

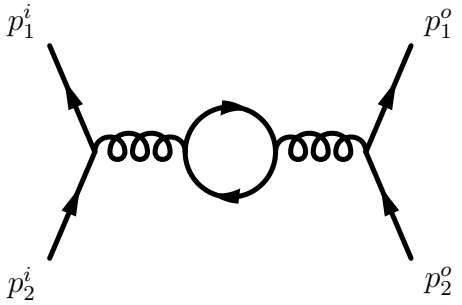


Figure 2.2: Feynman diagram containing a loop.

called hadrons. Introducing these hadrons in the picture raises the need for a *non-perturbative QCD* description as those complex particles are not governed by the QCD Lagrangian. The typical radius of a hadron, e. g. the proton, is of order $1\,\mathrm{fm} \sim 1/200\,\mathrm{MeV}$. It is not possible to distinguish different partons in a hadron so the perturbative description is not valid any more at this scale. It is common to introduce a scale $\Lambda_{QCD} \sim 200\,\mathrm{MeV}$ at which perturbative physics break down and non-perturbative effects dominate. In this and the following sections the perturbative part will be analysed while non-perturbative effects will be introduced later.

For fundamental particles the *cross section* of a specified process is given by

$$\sigma = \int \frac{1}{flux} \times |\mathcal{M}|^2 \times \mathrm{d}\Phi \tag{2.4}$$

using the denotations:

- σ: cross section,
- $flux$: incoming particle flux,
- $\mathcal{M}$: matrix element calculated from Feynman rules,
- $\mathrm{d}\Phi$: phase space volume over which is integrated.

The *matrix element* $\mathcal{M}$ describes the physics involved. It can be visualised by a *Feynman diagram* that graphically represents the Feynman rules for the calculation of a process. Figure 2.1 shows a tree level Feynman diagram for a fundamental process. Moreover, higher order corrections have to be taken into account, like loops (figure 2.2), real (figure 2.3) and virtual corrections (figure 2.4). These additional diagrams cause ultraviolet and infrared divergences.

To handle the ultraviolet divergences that arise from virtual and loop diagrams, these diagrams are calculated using the method of *dimensional regularization*. Each loop or virtual correction contributes a term proportional to the strong coupling $\alpha_s = g_s^2/4\pi$. As each propagator of a loop or correction can contain a loop or additional virtual corrections inside, diagrams to all orders in α_s

contribute. The resulting expressions consist of a finite and a divergent part. For each order in α_s a counter term has to be added to the Lagrangian to cancel the divergent parts from the loop diagrams. This procedure is called *renormalization*. The counter terms can have arbitrary finite parts in addition. This is the reason for the existence of different renormalization schemes, as different choices of the finite parts are convenient for different applications.

The result from a calculation using dimensional regularization depends additionally on an arbitrary unit mass μ that is needed to give the expression the correct mass dimension. As a consequence parameters like quark masses and the strong coupling α_s become dependent on μ. By construction any observable calculated to all orders in α_s is independent of the choice of μ. Real calculations include only a finite number of Feynman diagrams, so all results depend on the choice of the renormalization scheme and the choice of μ.

The scale μ for a calculation is usually referred to as *renormalization scale*. The most important dependence is on the strong coupling. α_s can be calculated to n^{th} order, thus absorbing the effects from loops up to order n. Therefore renormalization accounts for effects at very small distances, i. e. on very small time scales. Adjusting $\alpha_s(\mu)$ it is possible to remove physics happening at time scales $t \ll 1/\mu$ from the perturbative calculation. Effectively this procedure sums physical effects at time scales smaller than $1/\mu$.

2.3 Running Coupling

For the parameters influenced by renormalization, the so called *renormalization group equations* can be derived. These equations describe the dependence of the parameters on μ and allow to compute the evolution of any parameter to a particular order. The calculation of the dependence of α_s leads to the following equation, where for μ the virtuality Q is used:

$$Q^2 \frac{\mathrm{d}\alpha_s(Q^2)}{\mathrm{d}Q^2} = \beta(\alpha_s(Q^2)) = -\alpha_s^2(\beta_0 + \beta_1\alpha_s + \beta_2\alpha_s^2 + \ldots)\,. \tag{2.5}$$

The coefficients β_i are needed to derive $\alpha_s(Q^2)$ to a specified order in perturbation theory. Solving the renormalization group equation for the coupling α_s gives its value including corrections up to the specified order. As reference scale Λ_{QCD} is chosen, i. e. the scale at which perturbative QCD breaks down. Using the boundary condition $\alpha_s(\Lambda_{QCD}) = \infty$ the *next-to-leading order* (NLO) expression for α_s is:

$$\alpha_s(Q^2) = \frac{1}{1+\beta_0 \ln(Q^2/\Lambda_{QCD}^2)}\left(1 - \frac{\beta_1 \ln\left[\ln(Q^2/\Lambda_{QCD}^2)\right]}{\beta_0^2 \ln(Q^2/\Lambda_{QCD}^2)} + \frac{\beta_1^2}{\beta_0^4 \ln(Q^2/\Lambda_{QCD}^2)}\left[\left(\ln\left[\ln(Q^2/\Lambda_{QCD}^2)\right] - \frac{1}{2}\right)^2 - \frac{5}{4}\right]\right). \tag{2.6}$$

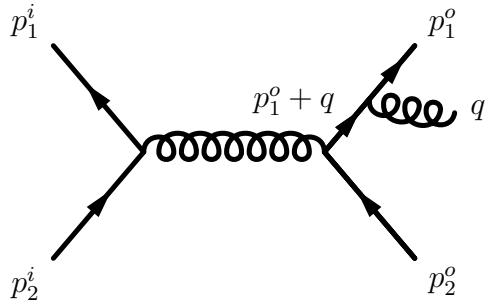


Figure 2.3: Feynman diagram with an additional real emission.

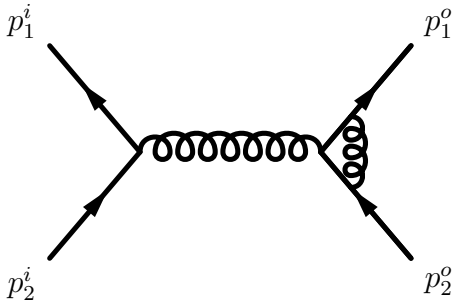


Figure 2.4: Feynman diagram with a virtual vertex correction.

In practice α_s has to be determined from data. It is common to take $\alpha_s(M_Z^2)$ as reference value for comparisons and to derive α_s values at other scales. Working at *leading order* (LO) one can solve the renormalization group equation to obtain:

$$\alpha_s(Q^2) = \frac{\alpha_s(Q_0^2)}{1 + \beta_0 \alpha_s(Q_0^2) \ln(Q^2/Q_0^2)} \,. \tag{2.7}$$

To get a reasonable result with a finite number of terms, the scale in α_s is chosen to be the virtuality Q^2 of the process of interest. As α_s is a function of Q^2, it is referred to as *running coupling*.

According to equation 2.6 $\alpha_s(Q^2)$ will increase if Q^2 decreases. This is true for $\beta_0 > 0$, which is the case for a theory with six quark flavours as it is assumed here. A smaller Q^2 corresponds to a larger distance, thus α_s increases with the distance of two coloured particles. As a consequence free colour charges cannot be observed above a distance $1/\Lambda_{QCD}$ due to the increased force acting between them. This effect is known as *confinement* leading to the existence of colour neutral hadrons containing colour charged quarks and gluons. The singularity of equation 2.6 at $Q = \Lambda_{QCD}$ is a consequence of the fact that too large values of α_s invalidate equation 2.5.

2.4 Higher Order Corrections

Figures 2.3 and 2.4 show diagrams with real and virtual corrections, respectively. Due to the fact that gluons are massless, there is in both cases the possibility of soft contributions, i. e. the additional gluon may have vanishing energy: $E_g \to 0$. In addition for real emissions the possibility of collinear contributions exists: $\sphericalangle(emitter, gluon) \to 0$. Calculating the diagrams of a certain order shows that the real and the virtual parts both give rise to infrared divergences. These divergences cancel for every order giving an allover finite result. So both real and virtual diagrams have to be taken into account for any calculation. The specification of the order for a calculated observable, e. g. the total cross section, refers to the inclusion of all diagrams with real and virtual corrections up to this order in α_s.

Figure 2.5: Collinear real emission diagram.

2.5 Factorization

Investigating higher order corrections in hadron–hadron collisions, one finds that there are infrared divergences due to real emissions that are not canceled by the virtual corrections. Collinear gluon emissions as shown in figure 2.5 are associated with an almost vanishing quark propagator, meaning that this quark is almost on its mass shell. These low-virtuality quarks have a very long life time and thus travel a long distance before any emission takes place ($x \propto 1/p$). Hence such an emission is considered not to be part of the hard parton interaction but it is associated with the incoming or outgoing hadron.

After the complete calculation at a given order in α_s using the procedure of dimensional regularization indeed only poles due to collinear emissions remain. The basic idea is to distinguish between long distance physics, associated with collinear emissions from almost real partons, and short distance physics, i. e. the hard process. One defines a so called *parton density function* (pdf) to separate long distance effects. The divergent parts of collinear emissions are absorbed in the parton density function. But there is some arbitrariness in the choice of constant terms put in the pdf, leading to different factorization schemes, like in the case of renormalization (compare section 2.2). Due to the usage of dimensional regularization another unit mass μ_F is introduced, which is referred to as *factorization scale*. This scale separates long and short distance physics. For convenience renormalization and factorization scale are often chosen the same. Hence the pdf depends on the chosen scheme and the choice of the factorization scale. In addition there is a dependence on the momentum fraction x of the momentum of the initial hadron as possible splittings depend on the momentum of the initial parton. The incoming parton is assumed to carry a fraction x of the hadron's momentum p.

A good variable for looking at two partons coming from a collinear splitting is the relative transverse momentum p_T. Then an emission causes a factor of the form

$$\frac{\alpha_s(p_T^2)}{2\pi} \int_{\mu_F^2}^{s} \frac{\mathrm{d}p_T^2}{p_T^2} = \frac{\alpha_s(p_T^2)}{2\pi} \ln\left(\frac{s}{\mu_F^2}\right) \tag{2.8}$$

in the cross section. $\sqrt{s}$ denotes the centre-of-mass energy of the process. In this integral μ_F shields the singularity which has been moved to the pdf, but it still contributes a large logarithm. For a complete calculation this integral has to be folded with a pdf.

2.6 DGLAP Equations

A higher value of Q^2 resolves structures at small distances, thus each parton does not contain only one single parton of a dedicated flavour but a Q^2 and x dependent number of all flavours. Summing up all contributions from one single parton will give one net parton of the specific flavour, e. g. probing an up quark there is the possibility of finding gluons or other quark flavours but there remains one net up quark. Furthermore a hadron does not consist only of valence quarks but additionally of sea quarks and gluons due to vacuum fluctuations. The pdfs can be considered giving the probability of finding a particular parton with momentum fraction x when probing a hadron with Q^2. Replacing the factorization scale μ_F in a pdf by an arbitrary scale μ, e. g. the virtuality Q^2, reveals a dependence of the pdfs on x and μ.

The interpretation above suggests to examine a parton radiating many gluons, thereby changing its momentum fraction. Calculating the influence of one additional gluon emission from a gluon for a process already including n emissions gives:

$$\mathrm{d}\sigma^{n+1} = \frac{\mathrm{d}p_T^2}{p_T^2}\mathrm{d}x\frac{\alpha_s}{2\pi}P_{qq}(x) \times \mathrm{d}^n\sigma\,. \tag{2.9}$$

In equation 2.9 a coefficient function depending on the momentum fraction x appears, the so called *splitting kernel* $P_{qq}(x)$. This function describes the splitting $q \to qg$. It will be discussed in more detail below. The computation of all possible radiations implies the need to order the emissions. Assuming that each emission has a higher p_T than the preceding when approaching the hard process means a strong ordering in p_T:

$$Q^2 \gg p_{Tn}^2 \gg \ldots \gg p_{T2}^2 \gg p_{T1}^2 \gg Q_0^2\,, \tag{2.10}$$

where Q^2 is the virtuality of the hard process and Q_0^2 a lower cutoff. Applying the ordering of equation 2.10 to equation 2.9 gives:

$$\sigma \propto \int_{Q_0}^{Q^2} \frac{\mathrm{d}p_{Tn}^2}{p_{Tn}^2}\frac{\alpha_s(p_{Tn}^2)}{2\pi} \cdots \int_{Q_0}^{p_{T3}^2} \frac{\mathrm{d}p_{T2}^2}{p_{T2}^2}\frac{\alpha_s(p_{T2}^2)}{2\pi} \int_{Q_0}^{p_{T2}^2} \frac{\mathrm{d}p_{T1}^2}{p_{T1}^2}\frac{\alpha_s(p_{T1}^2)}{2\pi}\,. \tag{2.11}$$

As each integral over $\mathrm{d}p_T^2/p_T^2$ causes a numerically large logarithm, the total cross section for n emissions is enhanced proportional to

$$\sigma \propto (\alpha_s)^n \ln^n\left(\frac{Q^2}{Q_0^2}\right)\,. \tag{2.12}$$

These large logarithms are referred to as *leading logarithms* (LL). In case $p_{Tn}^2 \ll p_{Tn-1}^2$ there is no denominator of order p_{Tn}^2, thus there is one power less in the logarithms for the same power in α_s, known as *next-to-leading logarithms* (NLL).

Not only gluon emission by quarks and gluons but also gluon splittings in $q\bar{q}$ pairs affect the pdfs. Analyzing the properties of pdfs leads to the powerful

DGLAP (Dokshitzer, Gribov, Lipatov, Altarelli, Parisi) equations [34, 35, 36]:

$$\mu^2\frac{\partial q_i}{\partial\mu^2}\left(x,\mu^2\right)=\int_x^1\frac{\mathrm{d}z}{z}\frac{\alpha_s}{2\pi}\Bigg[P_{q_iq_j}(z,\alpha_s)q_j\left(\frac{x}{z},\mu^2\right)+P_{q_i\bar{q}_j}(z,\alpha_s)\bar{q}_j\left(\frac{x}{z},\mu^2\right)+P_{q_ig}(z,\alpha_s)g\left(\frac{x}{z},\mu^2\right)\Bigg]$$

$$\mu^2\frac{\partial \bar{q}_i}{\partial\mu^2}\left(x,\mu^2\right)=\int_x^1\frac{\mathrm{d}z}{z}\frac{\alpha_s}{2\pi}\Bigg[P_{\bar{q}_iq_j}(z,\alpha_s)q_j\left(\frac{x}{z},\mu^2\right)+P_{\bar{q}_i\bar{q}_j}(z,\alpha_s)\bar{q}_j\left(\frac{x}{z},\mu^2\right)+P_{\bar{q}_ig}(z,\alpha_s)g\left(\frac{x}{z},\mu^2\right)\Bigg]$$

$$\mu^2\frac{\partial g}{\partial\mu^2}\left(x,\mu^2\right)=\int_x^1\frac{\mathrm{d}z}{z}\frac{\alpha_s}{2\pi}\Bigg[P_{gg}(z,\alpha_s)g\left(\frac{x}{z},\mu^2\right)+\sum_{f=q,\bar{q}}P_{gf}(z,\alpha_s)f\left(\frac{x}{z},\mu^2\right)\Bigg]. \tag{2.13}$$

The DGLAP equations describe the evolution of a parton over the scale μ, summing up the leading logarithmic contributions to all orders in α_s. x again denotes the fraction of the hadron's momentum.

The Altarelli-Parisi splitting kernels $P_{ij}(z,\alpha_s)$ are associated with branchings $j\to i+X$ where parton i gets the momentum fraction z of j's momentum. These splitting kernels can be expanded in a power series in α_s:

$$P(z,\alpha_s)=P^{(0)}(z,\alpha_s)+\frac{\alpha_s}{2\pi}P^{(1)}(z,\alpha_s)+\left(\frac{\alpha_s}{2\pi}\right)^2P^{(2)}(z,\alpha_s)+\ldots. \tag{2.14}$$

By including higher orders of the splitting kernel, one accounts for higher order effects in α_s, i. e. fluctuations within fluctuations. In [33] the splitting kernels are given explicitly to NLO. The flavour and spin dependences are discussed in detail in [29].

As discussed above, collinear emissions give rise to logarithmic enhancements of the form $\alpha_s\ln(Q^2/Q_0^2)$ while soft emissions lead to enhancements in $\alpha_s\ln(1/x)$. Furthermore the contributing regions can overlap causing double logarithmic contributions of the form $\alpha_s\ln(Q^2/Q_0^2)\ln(1/x)$. The leading contributions of soft radiation are associated with a strong ordering similar to equation 2.10: $x\ll x_n\ll\ldots\ll x_1\ll x_0$. It can be shown that the DGLAP equations sum correctly the leading contributions from the $\alpha_s\ln(Q^2/Q_0^2)$ and $\alpha_s\ln(Q^2/Q_0^2)\ln(1/x)$ terms. The $\alpha_s\ln(1/x)$ terms are governed by the *BFKL* equation [37, 38]. Both equations can be modified to include NLL contributions that are suppressed by a factor α_s.

Thus it is possible to factorise long and short distance physics when calculating cross sections for hadron–hadron collisions. The pdf $f_{a/h}(x,\mu_F^2)$ provides the probability for finding a parton a with momentum fraction x inside hadron h when using the factorization scale μ_F. Therefore it is possible to calculate the

cross section for partons c and d in the final state in hadron collisions:

$$\mathrm{d}\sigma\left(h_1 h_2 \to cd\right) = \int_0^1 \mathrm{d}x_1 \mathrm{d}x_2 \sum_{a,b} f_{a/h_1}(x_1, \mu_F^2) f_{b/h_2}(x_2, \mu_F^2) \times \mathrm{d}\hat{\sigma}^{(ab\to cd)}(Q^2, \mu_F^2)\,. \quad (2.15)$$

Here the hard process is separated from effects at time scales $t \gg 1/\mu_F$, which are moved into the pdf. It should be noted that factorization is not proven but only an assumption.

2.7 Fragmentation

Equation 2.15 serves for calculating the partonic final state in hadron collisions. While it is sufficient for computing lepton production, it has to be extended for quark and gluon production. Confinement, introduced in section 2.3, is the reason that free colour charges cannot be observed at distances greater $1/\Lambda_{QCD}$. Hence colour charged particles produced in a collision described by equation 2.15 have to be converted into hadrons. This procedure is called *hadronization* or *fragmentation*. The final partons entering this non-perturbative hadronization process are described by *fragmentation functions* D^h. In analogy to the DGLAP equations, treating the evolution of partons going into the hard process, similar equations can be formulated for describing the evolution of a parton from the hard interaction to a final state directly before hadronization takes place. These equations account for radiative corrections in the final state. Again factorization holds for this process, giving the following μ_F dependent equations:

$$\mu^2 \frac{\partial D_{q_i}^h}{\partial \mu^2}\left(x, \mu^2\right) = \int_x^1 \frac{\mathrm{d}z}{z} \frac{\alpha_s}{2\pi} \left[P_{q_j q_i}(z, \alpha_s) D_{q_j}^h\left(\frac{x}{z}, \mu^2\right) + P_{\bar{q}_j q_i}(z, \alpha_s) D_{\bar{q}_j}^h\left(\frac{x}{z}, \mu^2\right) + P_{g q_i}(z, \alpha_s) D_g^h\left(\frac{x}{z}, \mu^2\right) \right]$$

$$\mu^2 \frac{\partial D_{\bar{q}_i}^h}{\partial \mu^2}\left(x, \mu^2\right) = \int_x^1 \frac{\mathrm{d}z}{z} \frac{\alpha_s}{2\pi} \left[P_{q_j \bar{q}_i}(z, \alpha_s) D_{q_j}^h\left(\frac{x}{z}, \mu^2\right) + P_{\bar{q}_j \bar{q}_i}(z, \alpha_s) D_{\bar{q}_j}^h\left(\frac{x}{z}, \mu^2\right) + P_{g \bar{q}_i}(z, \alpha_s) D_g^h\left(\frac{x}{z}, \mu^2\right) \right]$$

$$\mu^2 \frac{\partial D_g^h}{\partial \mu^2}\left(x, \mu^2\right) = \int_x^1 \frac{\mathrm{d}z}{z} \frac{\alpha_s}{2\pi} \left[P_{gg}(z, \alpha_s) D_g^h\left(\frac{x}{z}, \mu^2\right) + \sum_{f=q,\bar{q}} P_{fg}(z, \alpha_s) D_f^h\left(\frac{x}{z}, \mu^2\right) \right]. \quad (2.16)$$

The fragmentation functions represent the probability of finding a hadron h with momentum fraction x in the final state. It is worth noting that the indices for the splitting kernels are inverted compared to the DGLAP equations, equation 2.13. While the original DGLAP equations treat a spacelike evolution of partons,

equations 2.16 are timelike. The splitting kernels are identical for both kind of equations at leading order but differ beyond.

The preceding sections present the means for treating a hadron–hadron collision correctly from initial to final hadrons. The theorem of factorization allows the separation of a process in perturbative and non-perturbative steps. First incoming hadrons described non-perturbatively by pdfs can be evolved perturbatively using the DGLAP equations. Then the hard interaction of two partons can be calculated by the means of perturbative QCD, i. e. by the Feynman rules. Afterwards the usage of fragmentation functions provides a way to evolve the final partons perturbatively before the non-perturbative hadronization process starts. Thereby it is possible to draw a complete picture of hadron–hadron processes. This will be done in chapter 4.

Finally it is necessary to handle the remnants of the incoming hadrons, i. e. the partons inside the hadron that do not participate in the hard interaction. These *beam remnants* cannot be treated properly by analytical means of QCD, though they play an important role in hadron–hadron collisions. Sometimes more than two partons interact and single partons can interact more than once. Moreover the excited remnants carry colour charge forcing them to interact with the partons arising from the hard process to produce colour neutral hadrons in the final state. In section 4.4 a possible treatment of the beam remnants will be discussed.

Chapter 3

Proton–Proton Interactions

3.1 Proton Collisions

A proton is a colour neutral hadron, consisting of three *valence quarks*, namely two up quarks and one down quark. Thus it has an electric charge of +1 in units of the elementary charge. Due to the uncertainty principle additional quark–anti-quark pairs can be observed in the proton, called *sea quarks*. The proton can take part in elastic interactions in which it acts as one single particle, or in inelastic ones, wherein it is "broken up" and single partons are scattered. For the inelastic collisions at a scale Q^2 the interacting parton has to be evolved according to the DGLAP equations 2.13. Therefore not only quarks of different flavours but also gluons can enter the hard process. In a figurative sense each parton is surrounded by a cloud of other partons that can be resolved at a specific scale. About 50% of the proton momentum is carried by quarks and 50% by gluons. [33]

For the proton parton density functions are determined giving the probability of finding a special kind of parton at a certain scale Q^2 with a momentum fraction x inside. The measurements of different experiments serve as input for global fits determining the pdfs for the proton. The fits are performed utilizing matrix elements of a particular order in perturbation theory. Respecting the rules of factorization, pdfs obtained at a specific order in perturbation theory can be used only for predictions at this order, thus NLO pdfs have to be combined with NLO matrix elements for meaningful predictions. Several groups provide independent fits. In figure 3.1 pdfs obtained from a NLO fit are depicted, showing a significant decrease of the valence quark densities and an increase of gluon and sea quark densities at small x values. In chapter 5 pdfs from *Alekhin* [39], the *CTEQ* [40, 41] and the *MRST* collaboration [42] are compared.

At colliders not single hadrons but whole bunches collide. If two hadron bunches hit each other, more than one hadron–hadron interaction can take place. For LHC in the majority of all collisions only one hard scattering is expected. Nevertheless, additional soft proton–proton scatterings can cause additional hadrons and thus jets in the detector contaminating the measurement. These interactions are referred to as *pileup*. During the low luminosity phase at LHC approximately five pileup events will overlay each hard collision [44].

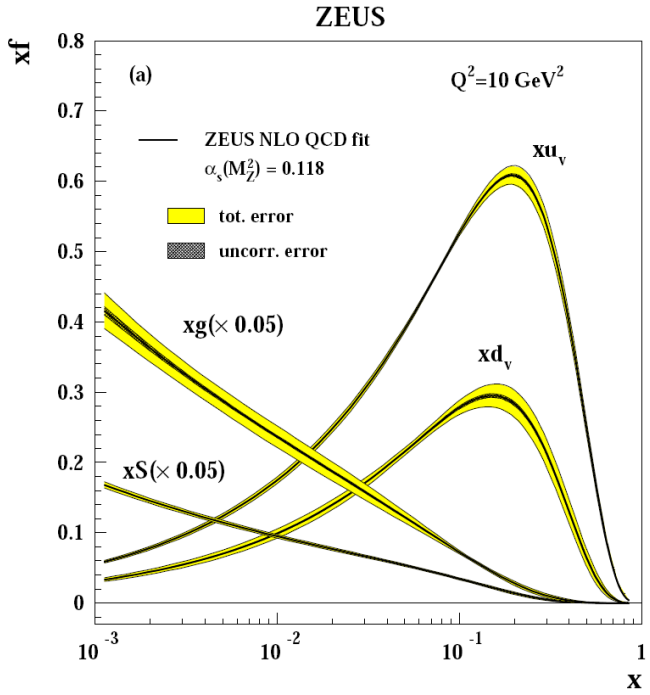


Figure 3.1: Gluon, sea and valence quark distributions obtained from a NLO fit to ZEUS data. [43]

Finally the Fermi motion of the partons has to be respected, providing the partons inside the hadron with a transverse momentum. This motion can equip the interacting partons with a small transverse momentum that is passed to the outgoing particles.

3.2 Infrared Safe Observables

An observable is infrared safe if a quark or gluon is treated the same by the theory as a quark or gluon accompanied by an arbitrary number of soft gluons and collinear gluons and $q\bar{q}$ pairs, i. e. $|q\rangle = |q + n_1 g_s + n_2 g_{\parallel} + n_3 q\bar{q}_{\parallel}\rangle$ must give the same contribution. Experimentally the result of a measurement must be independent of any changes in the detector granularity or resolution. This implies that any infrared safe observable must be a linear combination of parton four momenta. An observable from an n-parton final state can be written in the form:

$$O = \sum_n \frac{1}{n} \int \mathrm{d}\Phi^n \overline{\sum} \left|\mathcal{M}^{(n)}(p_i)\right|^2 S_n(p_i) \tag{3.1}$$

$$\text{with} \quad \begin{cases} S_n = 1 & \text{for } O = \sigma \\ S_n = \delta(X - \chi_n(p_i)) & \text{for } O = \frac{\mathrm{d}\sigma}{\mathrm{d}X}\,. \end{cases}$$

In the case of infrared safety an observable has the property

$$\chi_{n+1}(p_1, \ldots, (1-\lambda)p_n, \lambda p_n) = \chi_n(p_1, \ldots, p_n)\,. \tag{3.2}$$

In sections 2.5, 2.6 and 2.7 the picture of soft and collinear partons accompanying a quark or gluon probed at Q^2 was developed. According to equation 3.2 infrared safety now implies that the contribution to a physical observable of a single parton and a parton surrounded by an arbitrary number of soft and collinear partons should be the same. [33]

3.3 Kinematic Variables

In this section some standard variables used at hadron colliders are briefly presented.

The incoming beams are defined to be along the z-axis, leading to the definition of the transverse momentum

$$p_T = \sqrt{p_x^2 + p_y^2}\,. \tag{3.3}$$

The polar angle is denoted by θ while the azimuthal angle is denoted by ϕ. The momentum of the colliding partons is not balanced in z-direction and θ is not invariant under boosts. For detectors at hadron colliders the so-called *pseudo-rapidity* η is introduced, defined by

$$\eta = -\ln\tan\frac{\theta}{2}\,, \tag{3.4}$$

which is a directly measurable quantity. A comparison with the *rapidity* y, defined by

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right)\,, \tag{3.5}$$

shows that pseudo-rapidity and rapidity are identical for massless particles. It is possible to introduce a distance measurement between two particles i and j according to

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}\,. \tag{3.6}$$

ΔR is independent of boosts along the z-axis and can be directly extracted from measured quantities. This comes from the fact that rapidities can be added linearly. A comprehensive overview of these variables and their properties can be found in [45].

3.4 Jet Algorithms

One single parton from a hard process can lead to a huge number of particles in the final state due to radiation, fragmentation and interactions with the detector. A good entity for the analysis of hadronic final states is a *jet*, which is a

combination of final particles or detector output to one single parton-like object by combining the input's four momenta. The concept of jet algorithms is driven by the assumption that particles originating from one initial parton are close in space. While a jet algorithm tries to project many particles onto one single parton, it is important to mention that a jet is not the same as a parton though it is sometimes treated that way.

A jet is an object that can be built of partons, hadrons or calorimeter output from detectors. At each step a jet is a different kind of object and an one-to-one mapping is not fully valid. Nevertheless a jet is constructed in a way that it shows as little sensitivity to QCD effects as possible. A jet is a physical observable and thus should fulfil the requirement of infrared safety according to the definition given in section 3.2, i. e. it should be insensitive to collinear splittings and soft radiation.

3.4.1 Recombination Schemes

For jet algorithms the merging of two four vectors into a single one induces an arbitrariness. The different methods have different advantages and deficiencies, but basically in each procedure a piece of the information is lost, leaving the choice of which information to preserve in full. Throughout this thesis the *E-scheme* is used in which the four momenta are added linearly. This results in an invariant mass of the final object that may have no physical interpretation, e. g. if the jet is associated with a light quark. However, this procedure does not affect the direction of the jet. Alternative recombination schemes are discussed in [46].

3.4.2 Sequential Clustering

The idea of *sequential clustering* is the inversion of the branching process by a successive combination of particles, starting by those with the smallest relative scale. The particular algorithm used here is the *inclusive k_T algorithm* [47], defined by the following procedure:

1. For each pair of particles i and j calculate

$$d_{ij} = \min(k_{Ti}^2, k_{Tj}^2)\frac{\Delta R_{ij}^2}{R^2}\,, \tag{3.7}$$

   where k_{Ti} denotes the transverse momentum of i with respect to the beam axis and R is a parameter of the algorithm. In addition for each particle its beam distance $d_{ib} = k_{Ti}^2$ is calculated.

2. If d_{ij} is the minimum of all values combine particles i and j to a single particle. If d_{ib} is the minimum call i a jet and remove it from the list.

3. Iterate until no particles are left.

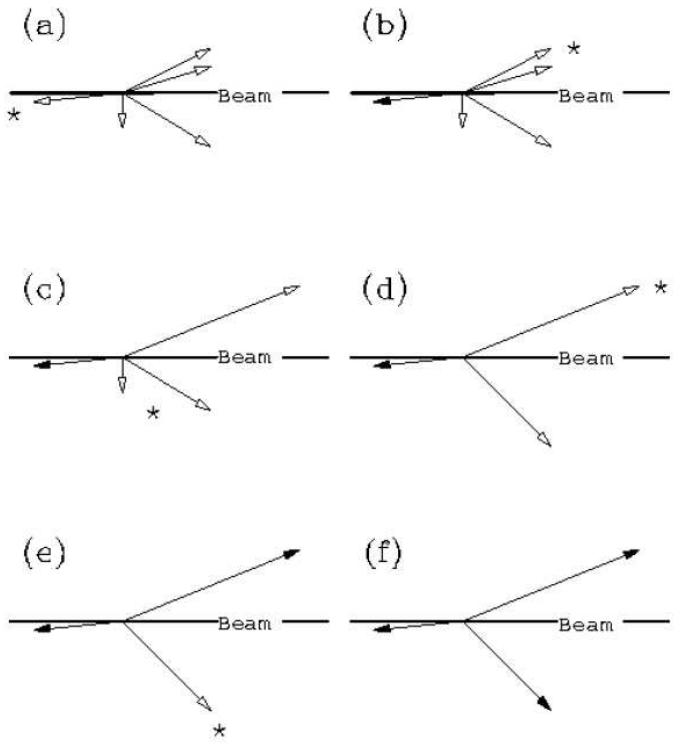


Figure 3.2: Schematic representation of the sequential clustering algorithm. The asterisk marks the particles (arrows) processed in the particular step. Either two particles are merged or a stable jet is defined. The latter ones are represented by filled arrows. [48]

R is a distance-like parameter giving the minimum separation in y and ϕ of any jets. This procedure is depicted in figure 3.2. A k_T algorithm with parameter R is denoted by KT and ten times the value of R, e. g. KT4 for $R = 0.4$. By construction sequential clustering algorithms are infrared safe.

3.4.3 SISCone

The second big category of jet algorithms are *cone jets*, which are driven by the idea of a directed, centered energy flow. All particles or detector cells inside a cone of a given value ΔR are combined to a jet candidate, and the combined four momentum is calculated. The direction of this four vector is used as axis and a new cone is drawn around it. This procedure is iterated until the axis is considered stable.

Most cone jet definitions suffer from infrared unsafety [49]. Here the *Seedless Infrared Safe Cone* jet algorithm (SISCone) is used which cures these problems. Each distinct particle contents within a stable cone is considered a jet candidate. As particles may be contained in more than one jet candidate, a *split-merge* procedure like proposed in [48] is used. Analogously to the k_T algorithm, a

SISCone jet definition with $\Delta R = 0.4$ is denoted with SC4. A full definition of the SISCone algorithm and a detailed construction description are given in [49].

Chapter 4

Monte Carlo Event Generators

Monte Carlo event generators are computer programs calculating physics processes numerically. According to the procedure presented in section 2.7, they combine perturbative and non-perturbative physics. Hard matrix elements (ME) are computed to a particular order in perturbation theory while the parton evolution governed by the DGLAP equations is done via the *parton shower* (PS) approach. Experimentally determined pdfs are taken as input for the parton evolution. Non-perturbative models are applied for the description of *multiple interactions* (MI) and the treatment of the beam remnants, whereas the evolution of the partons from additional interactions is treated again by the perturbative parton shower. A *hadronization model* converts the produced partons in colour neutral hadrons. In a figurative sense the whole process can be written as

$$(\mathrm{ME} \oplus \mathrm{MI}) \otimes \mathrm{PS} \otimes \mathrm{Hadronization}\,.$$

Monte Carlo generators compute single collisions (events), by this providing the opportunity not only to calculate inclusive quantities like cross sections, but also to check different observables on an event-by-event basis.

After a very brief introduction in Monte Carlo integration (4.1) a detailed description of the parton shower approach is given (4.2). Afterwards different possibilities of the combination of hard matrix elements and the parton shower are presented (4.3) and the non-perturbative components are discussed (4.4 and 4.5). Finally the Monte Carlo generators used for this work are introduced (4.6 to 4.9).

4.1 Monte Carlo Integration

Monte Carlo integration is a numerical method of integration. The basic idea of integrating a function like

$$I = \int_{x_1}^{x_2} f(x)\mathrm{d}x = (x_2 - x_1)\langle f(x)\rangle \tag{4.1}$$

can be generalised to

$$I \approx I_N = (x_2 - x_1)\frac{1}{N}\sum_{i=1}^{N} f(x_i)\,. \tag{4.2}$$

The accuracy of equation 4.2 is governed by the *central limit theorem*:

$$I \approx I_N \pm \sqrt{V_N/N}\,, \tag{4.3}$$

where V_N denotes the variance of I_N. Thus the error on an integral solved via the Monte Carlo method is proportional to $1/\sqrt{N}$. There are plenty of other numerical integration methods available that converge much faster, but extending the formalism to more dimensions shows that the error for the Monte Carlo method remains to be $1/\sqrt{N}$ while the convergence for all other methods gets worse. Monte Carlo integration offers a good possibility to calculate observables, e. g. cross sections like in equation 2.15, numerically. A detailed discussion of Monte Carlo integration and its general implementation in Monte Carlo event generators can be found in [50, 51]. The event generators described in this chapter are based on this integration method.

4.2 Parton Showers

For the following sections the splitting kernels introduced in equation 2.14 have to be modified to enable a numerical treatment. A more careful investigation in section 2.6 would have shown that for writing down the splitting kernels explicitly the so-called *plus-prescription* is needed. It serves for handling the singularities correctly and ensures flavour and energy conservation in analytical calculations. In Monte Carlo generators these quantities can be tracked explicitly. The basic structure of the equations is not affected by this change, but for any calculations it is necessary to ensure the use of the correct splitting kernels depending on the assumptions made for the derivation of the equations. More information on the plus-prescription and its usage can be found in [29].

The following discussion of parton showers is based on [29, 33, 50]. A full description of all aspects summarised in this section can be found therein.

4.2.1 Parton Evolution

The DGLAP equations 2.13 account for branchings of partons in high energetic processes. In a Monte Carlo program these branchings are simulated explicitly by the parton shower approach. As a high energetic particle resolves smaller distances, thus seeing the substructure of hadrons as well as of partons, it is necessary to evaluate the parton entering the hard process at the correct scale. As equations 2.13 and 2.16 are not suited for a numerical approach, they have to be rewritten using the modified splitting kernels mentioned above. An abstract evolution scale t is introduced for the description of the evolution. The reason for this will become apparent in the next sections. The DGLAP equations modified

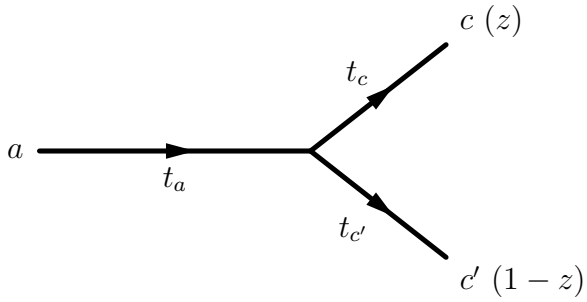


Figure 4.1: Timelike splitting of a parton a in two partons c and c'.

for a numerical treatment are:

$$t\frac{\partial}{\partial t}f_{a/h}(x,t) = \int_x^{1-\epsilon'_a(t)} \frac{\mathrm{d}z}{z}\frac{\alpha_s(t,z)}{2\pi}P_{ab}(z)f_{b/h}\left(\frac{x}{z},t\right) - f_{a/h}(x,t)\sum_{a\to cc'}\int_{\epsilon_c(t)}^{1-\epsilon_{c'}(t)}\mathrm{d}z\frac{\alpha_s(t,z)}{2\pi}P_{ca}(z) \tag{4.4}$$

$$t\frac{\partial}{\partial t}D_a^h(x,t) = \int_x^{1-\epsilon'_b(t)} \frac{\mathrm{d}z}{z}\frac{\alpha_s(t,z)}{2\pi}P_{ba}(z)D_b^h\left(\frac{x}{z},t\right) - D_a^h(x,t)\sum_{a\to cc'}\int_{\epsilon_c(t)}^{1-\epsilon_{c'}(t)}\mathrm{d}z\frac{\alpha_s(t,z)}{2\pi}P_{ca}(z) \tag{4.5}$$

Equation 4.4 is spacelike and gives the x and scale t dependent number of partons of flavour a found for an incoming parton of flavour h. Every parton b that branches in a process $b \to aa'$ increases the number of available partons of flavour a while each branching $a \to cc'$ decreases it. The two contributing terms can thus be viewed as source and sink for parton flavour a. For the timelike equation 4.5 a similar interpretation exists, leading to the final hadrons of type h resulting from partons of flavour a. Figure 4.1 illustrates a timelike branching of a parton a into two partons c and c', that gain a momentum fraction z and $1-z$ of the initial momentum of a, respectively.

The cut-off parameters ϵ_i depend on the choice of the evolution variable t and the kinematics of the evolution. z denotes an energy fraction for which different choices of its exact definition and the particular evolution frame are available. For the evolution scale t a lower limit exists at which perturbative QCD is invalidated. So it is necessary to introduce a cut-off t^0 for every parton flavour to stop the evolution well above this limit. ϵ serves as a resolution criterion, shielding too soft radiation. Thus in equations 4.4 and 4.5 the unresolvable real emissions are isolated below (above) the z cut-off. By adding these contributions to the virtual diagrams both the real and the virtual part become finite (compare section 2.4).

4.2.2 Sudakov Form Factors

The differential branching probability for $a \to bc$ is given by

$$\mathrm{d}P_a = \sum_{b,c} \frac{\mathrm{d}t}{t} \mathrm{d}z \frac{\alpha_s}{2\pi} P_{a\to bc}(z) \tag{4.6}$$

in terms of t and momentum fraction z. Using the definition

$$I_{a\to bc}(t) = \int_{\epsilon_b(t)}^{1-\epsilon_c(t)} \mathrm{d}z \frac{\alpha_s}{2\pi} P_{a\to bc}(z) \tag{4.7}$$

one can write the probability for an emission in the infinitesimal interval $\mathrm{d}t$ as $\sum_{b,c} I_{a\to bc}(t)\mathrm{d}t/t$. Hence the probability for no emission is given by $1 - \sum_{b,c} I_{a\to bc}(t)\mathrm{d}t/t$. A parton evolving from a scale t_0 can branch at a scale $t > t_0$ only if it has not branched before. So the non-emission probability for all infinitesimal time intervals between t_0 and t has to be taken into account, leading to a no-branching probability between t_0 and t:

$$\Delta_a(t, t_0) = \exp\left(- \int_{t_0}^{t} \frac{\mathrm{d}t'}{t'} \sum_{b,c} I_{a\to bc}(t') \right). \tag{4.8}$$

The probability for branching at the scale t is consequently given by the product of the probability that no branching occurred before t and that a branching takes place at t:

$$\frac{\mathrm{d}P_a}{\mathrm{d}t} = \Delta_a(t, t_0) \times \left(\frac{1}{t} \sum_{b,c} I_{a\to bc}(t) \right). \tag{4.9}$$

The factor $\Delta_a(t, t_0)$ is called *Sudakov form factor*. It ensures that no emission happens before t and thus serves for conservation of the total probability. The second term is the probability for a branching at the scale t.

4.2.3 Timelike Parton Shower

Using the Sudakov form factor defined in equation 4.8, it is possible to rewrite the timelike DGLAP equation (equation 4.5):

$$t\frac{\partial}{\partial t}\left(\frac{D_a^h(x,t)}{\Delta_a(t, t_0^a)} \right) = \int_x^{1-\epsilon_c(t)} \frac{\mathrm{d}z}{z} \frac{\alpha_s(z,t)}{2\pi} P_{ba}(z) \frac{1}{\Delta_a(t, t_0^a)} D_b^h\left(\frac{x}{z}, t\right). \tag{4.10}$$

Solving equation 4.10 yields:

$$\begin{aligned} D_a^h(x,t) &= \Delta_a(t, t_a^0) D_a^h(x, t_a^0) \\ &+ \int_{t_a^0}^{t} \frac{\mathrm{d}t_1}{t_1} \int_x^{1-\epsilon_c(t_1)} \frac{\mathrm{d}z_1}{z_1} \Delta_a(t, t_1) \frac{\alpha_s(t_1, z_1)}{2\pi} P_{ba}(z_1) \Delta_b(t_1, t_b^0) D_b^h\left(\frac{x}{z_1}, t_b^0\right) \\ &+ \dots . \end{aligned} \tag{4.11}$$

Equation 4.10 is the master equation for timelike parton showers used in Monte Carlo generators. This shower type is used for *final state radiation* describing the emissions of timelike partons, i. e. with $t > 0$, coming from the hard matrix element. The first term represents a parton of flavour a evolving from the starting scale t down to the cut-off t_a^0 without emitting any resolvable radiation. This behaviour is governed by the Sudakov form factor $\Delta_a(t, t_a^0)$. Then parton a enters the non-perturbative hadronization process. The second term describes a parton of flavour a evolving down to scale t_1 without any resolvable emission. At t_1 a process $a \to bc$ takes place where a parton of type b receiving momentum fraction z_1 is produced. This parton then evolves down to its cut-off t_b^0 without any resolvable emission. The following terms all include one splitting more than their predecessor. The partons arising from a splitting are all evolved according to equation 4.11 leading to a tree-like structure of successive branchings. After all possible branchings have been simulated, the partons are put on their mass-shell and enter the hadronization process.

This evolution is implemented in Monte Carlo programs as follows: First the Sudakov form factor is evaluated to select a branching scale for a parton evolving from t_1 to t_2 without resolvable emission by solving

$$\Delta_a(t_2, t_1) = R\,, \tag{4.12}$$

where R denotes a random number in $[0, 1]$. If a scale below t_a^0 is selected, the parton will be put on its mass-shell. The selection of the flavour in which the initial parton branches is done by a comparison of the different branching probabilities given in equation 4.7. If a branching $a \to bc$ occurs, the momentum fraction $z = z_2/z_1$ of b has to be determined in a second step, solving

$$\int_x^{z_2/z_1} \mathrm{d}z \frac{\alpha_s}{2\pi} P_{ba}(z) = R' \int_x^{1-\epsilon_b(t_2)} \mathrm{d}z \frac{\alpha_s}{2\pi} P_{ba}(z)\,, \tag{4.13}$$

where R' denotes another random number in $[0, 1]$. Applying this procedure iteratively determines all branchings in the final state radiation process.

An important issue is the choice of the argument of the strong coupling. Using the transverse momentum $p_T^2 \approx z(1-z)Q^2$ as argument of α_s at leading order, equation 2.7, supplies an additional term of order $\ln(1-z)$. This term is present in the second order splitting kernel, thus it is possible to resum not only the leading order but even some of the next-to-leading order contributions via the leading order parton shower. As a consequence Monte Carlo generators give a much more sophisticated description than an analytical calculation on the same level.

With the choice of p_T as argument of α_s the Sudakov form factor becomes:

$$\Delta_a(t, t_a^0) \propto \left(\frac{\alpha_s(t)}{\alpha_s(t_a^0)} \right)^{c \ln(t/\Lambda_{QCD}^2)}\,, \tag{4.14}$$

where c is a constant. This term goes to zero faster than any inverse power in t. From this can be seen that it becomes more and more unlikely to evolve down

in t without any resolvable emission as the Sudakov Form factor favours the soft phase space region. The parton shower is indeed built to account especially for these soft and collinear emissions, thus it may give unsatisfying results for hard radiation. This will be discussed further in section 4.3.

4.2.4 Spacelike Parton Shower

For a parton originating from a hadron the evolution process is different. Starting with an almost on-shell parton the scale has to be evolved to the one of the hard process. As four-momentum conservation is respected at every branching $a \to bc$, at least one of the partons b and c must have a spacelike virtuality $t < 0$. For this *initial state radiation* the evolution of the parton entering the hard process is described spacelike, meaning that the evolution stops at $t = -t_{hardprocess}$. The other parton is either put on its mass shell or further evolved by the timelike algorithm, i. e. it has $t > 0$.

From equation 4.4 an equation similar to 4.11 can be deduced for spacelike partons. Unfortunately the method described in section 4.2.3 is highly inefficient as it is not possible to steer the evolution by simple means to get the correct momentum fraction needed for the hard process. This problem is solved by introducing a *backward evolution* that starts at the hard process scale and than evolves $-t$ back to lower values. By modifying the Sudakov form factor from equation 4.8 to

$$\Pi_a(t, t_a^s, x) = \Delta_a(t, t_a^s) \frac{f_{a/h}(x, t_a^s)}{f_{a/h}(x, t)}\,, \tag{4.15}$$

it is possible to write a modified equivalent of equation 4.11 for the spacelike case. Manipulating the solution from 4.4 it is possible to rewrite the obtained series with the modified Sudakov form factor [33], yielding:

$$\begin{aligned} 1 = {} & \Pi_a(t, t_a^s, x) \\ & + \int_{t_a^s}^{t} \frac{\mathrm{d}t_1}{t_1} \int_x^{1-\epsilon_c^s(t_1)} \frac{\mathrm{d}z_1}{z_1} \Pi_a(t, t_1, x) \frac{\alpha_s(t_1, z_1)}{2\pi} \\ & \qquad \times P_{ab}(z_1) \frac{f_{b/h}(x/z_1, t_1)}{f_{a/h}(x, t_1)} \Pi_b\left(t_1, t_b^s, \frac{x}{z_1}\right) \\ & + \dots\,. \end{aligned} \tag{4.16}$$

The spacelike Sudakov form factor from equation 4.15 can be interpreted as the probability that a parton a present in hadron h evolves back from scale t to scale t_a^s without any resolvable emission. The DGLAP evolution suppresses partons with high momentum fractions x, which is expressed by $f_{a/h}(x, t) < f_{a/h}(x, t_a^s)$. So the pdf ratio reflects that a parton with high x is more unlikely to have undergone a preceding branching.

Equation 4.16 is the normalised sum of all possibilities to find a parton of flavour a after the evolution. In addition to the modified Sudakov factors the splitting kernels are accompanied by pdfs guiding the evolution back to the correct

parton content in the hadron. The probability for parton a to originate from a parton of flavour b depends on the incoming hadron, e. g. for a gluon coming from a proton it is much more likely that its parent is a u- than a d-quark. With this knowledge an interpretation of equation 4.15 analogous to the timelike case, equation 4.11, is possible. A parton of flavour a entering the hard process at a scale t is evolved backwards to the spacelike cut-off scale t^s. The first term describes an evolution of a back to h without any resolvable branchings. The second term represents the evolution of a back to a scale t_1 where it was produced via the branching $b \to aa'$. Parton b carrying the momentum fraction $x/z_1 > x$ originates from h at a scale t_b^s.

In principle equation 4.4 can be solved to gain the pdf $f_{a/h}$. In real life the pdfs are taken from fitted experimental data and evaluated at the desired x and t values using the DGLAP equations. Then these pdfs are used to steer the backward evolution taking into account that the hadron composition depends on the scale t, reflecting that at higher scales smaller distances can be resolved.

The implementation of initial state radiation in Monte Carlo generators is done analogously to the final state case. The timelike Sudakov form factor is replaced by the spacelike one and the selection of the momentum fraction is done proportional to $\alpha_s(t,z)P_{ba}(z)f_{b/h}(x/z,t)$, respecting the influence of the pdfs of the initial hadron.

4.2.5 Coherent Parton Branching

An external line of a Feynman diagram with momentum p and mass m emitting a gluon q causes a propagator of the form

$$\frac{1}{(p \pm q)^2 - m^2} = \frac{\pm 1}{2\omega E(1 - \cos\theta_{pq})}\,, \tag{4.17}$$

where E denotes the energy of the emitting particle, ω the energy of the gluon and θ_{pq} the angle between both particles. Such a propagator can give rise to soft and collinear enhancements. This section focuses on the treatment of soft gluons. For highly virtual partons $(p^2 - m^2)$ does not go to zero as ω approaches zero. So off-shell particles do not cause soft enhancements in the cross section.

For a multi-parton state it is possible to emit a soft gluon but its parent cannot be identified. If the gluon's wavelength is too large to resolve the internal structure of the multi-parton state, the contributions of all possible emitters have to be added. Considering an n-particle state emitting a soft gluon with four-momentum q the cross section for this process has the property

$$\mathrm{d}\sigma^{n+1} \propto \sum_{i \neq j} \left(\omega^2 \frac{p_i \cdot p_j}{(p_i \cdot q)(p_j \cdot q)} \right) \times \left(\frac{\mathrm{d}\omega}{\omega} \right) \times \mathrm{d}\sigma^n\,. \tag{4.18}$$

This formula describes the radiation of a dipole, so each pair of partons can be regarded as an antenna emitting radiation. This is the *colour dipole model* where the emissions give the typical bremsstrahlung spectrum proportional to $\mathrm{d}\omega/\omega$.

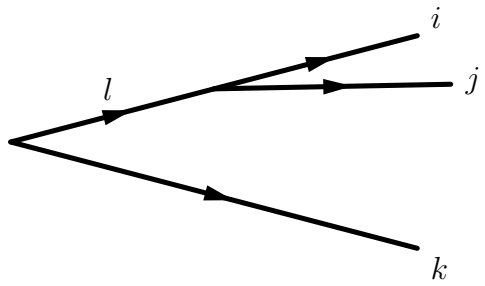


Figure 4.2: Three partons are present in the final state. For emissions associated with a dipole ik or jk a gluon cannot resolve the internal structure of i and j if the emission angle is greater than θ_{ij}. i and j can be combined to a quasi-parton l.

It is possible to decompose the dipole description in a way that the emission of a gluon q can be associated with either parton i or j. The emissions from parton i can be written in the form:

$$\left(\omega^2 \frac{p_i \cdot p_j}{(p_i \cdot q)(p_j \cdot q)}\right)^{(i)} \propto \frac{1}{1-\cos\theta_{iq}} \,. \tag{4.19}$$

The angle allowed for emissions is confined to:

$$\frac{m_i}{E_i} < \theta_{iq} < \theta_{ij} \,. \tag{4.20}$$

Equations 4.19 and 4.20 are the basis of *angular ordered* parton emissions. Parton i can emit a gluon q only at an angle θ_{iq} lower than the angle between i and its colour partner j. At θ_{ij} there is not a rigorous step function applied, but angles smaller than this value are highly preferred.

The particle mass prevents emissions in the particle's direction leading to a *dead cone* around heavy particles. Again this is not a step function but the emission rate decreases rapidly at $\theta_{iq} \sim m/E$.

Next the radiation pattern of three partons i, j and k forming a colour singlet will be briefly discussed. Making the assumption that i and j are close in angle their momenta can be replaced by their sum $p_l = p_i + p_j$ for interaction with k, so the emission of a parton q can be decomposed in a convenient way. There are two terms representing the contributions of a dipole consisting of i and j proportional to their colour charge like before. For angles above θ_{ij} the contributions of i and j sum coherently, for a soft gluon it is not possible to resolve their internal structure. As these two partons are close in angle, they can be described by a single quasi-parton l building a dipole with k, compare figure 4.2. This procedure can easily be extended to an arbitrary number of gluons, building the fundament for angular ordered parton cascades, which handle wide angle emissions first. A full treatment of the three-parton case can be found in [29], the n-parton case is discussed in detail in [33].

With these ingredients it is now possible to give more precise meaning to the evolution variable t. The emission of an additional gluon in an n-parton state from parton a can be described by the formula:

$$\mathrm{d}\sigma^{n+1} = \mathrm{d}\sigma^{n} \frac{\mathrm{d}t}{t} \mathrm{d}z \frac{\alpha_s}{2\pi} P_{aa} \,. \tag{4.21}$$

For the correct description of coherent branchings imposing angular ordering the variable

$$\zeta_{ij} = \frac{p_i \cdot p_j}{E_i E_j} = 1 - \cos\theta_{ij} \tag{4.22}$$

is introduced. The last relation is exact only for massless particles i and j. It can be shown that the soft emission is described correctly by the Altarelli-Parisi splitting kernels if ζ is used as evolution scale [33]. Thus a successive parton shower accounts correctly for the soft emissions if it is ordered in angle, i. e. in ζ. This analysis was carried out for timelike emissions, but it can be repeated for the spacelike case in an analogous way.

Up to now the argument for the cut-offs t^0 and t^s was based on the identification of t with the virtual mass of the partons Q^2. Equation 2.10 describes the strict ordering in p_T^2 necessary for the correct summation of the leading logarithmic enhancements via the DGLAP equations. It turns out that Q^2, p_T^2 and ζ are well suited as evolution variables, leading to a different ordering of the parton cascade. In [52] results from multi-jet events were presented claiming evidence for the observation of colour coherence effects as only an angular ordered shower was capable to describe the measured data correctly.

Figure 4.3 shows a simplified representation of the phase space for parton emissions. The variables chosen are the transverse momentum p_T of an emission and the rapidity relative to the emitting parton while the azimuth is not considered. The three ordering variables used explore the phase space from different starting points thereby preferring different emission types. The evolution variables used are p_T^2, $Q^2 \sim p_T^2/z(1-z)$ and $E^2\theta^2 \sim p_T/z^2(1-z)^2$. An evolution in a particular variable will prefer a high value for itself in the first emission. For example, a p_T^2 ordered shower (shown in figure 4.3 on the left) starts from the highest p_T, favouring a high p_T in the first emission. All successive emissions will have a lower p_T while for the other shower types the first one is not obliged to have the highest p_T. The same arguments hold for the other evolution variables as well. As the soft poles are not isolated by any limit in angle, angular ordered showers need an additional cut-off.

Usually ordering in virtuality is less restrictive than ordering in angle or transverse momentum. Which of the latter choices is more restrictive depends on the phase space region. The results and goodness of the parton shower depend on the phase space region of interest. It should be mentioned that angular ordering reduces the available phase space in comparison to the other two variables. This so-called *dead zone* will be explained in section 4.6, but it should be emphasised that it is not connected to the dead cone caused by heavy quark masses.

All three choices, Q^2, p_T^2 and angle, are exploited by the Monte Carlo programs used in this thesis, compare sections 4.6 to 4.9. For completeness it should be

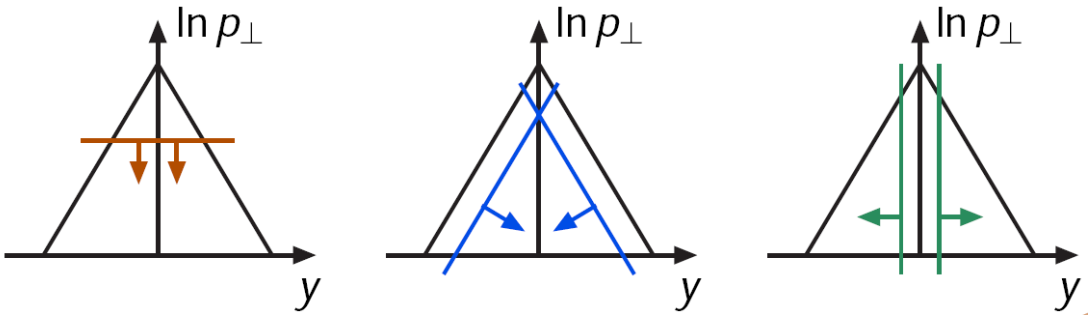


Figure 4.3: A simplified phase space picture for transverse momentum p_T and rapidity y is shown. The picture illustrates the direction in which the phase space is filled for the different evolution variables p_T^2 (on the left in red), Q^2 (in the middle in blue) and angle $\sim E^2\theta^2$ (on the right in green). [53]

mentioned that implementations of the full colour dipole model in Monte Carlo event generators exist [54]. For these generators a separation in initial and final state radiation is not necessary. Currently no generator using the colour dipole approach is fully ready for hadron–hadron events.

4.3 Matrix Element Matching

In section 2.4 the issue of real higher order corrections to a given process was raised, leading to multi-parton final states. Different approaches exist for the description of these parton emissions. The first one is the usage of all diagrams up to a certain order. Up to now this is implemented to NLO, so processes with one additional parton in the final state can be described completely. The second approach is the calculation of Feynman diagrams containing additional real partons, neglecting any diagrams containing virtual corrections. As virtual diagrams are not considered, the divergences from the real diagrams are not canceled. To fix this problem a minimum relative transverse momentum for the emitted partons is employed. A difficulty is the problem of combining different multiplicities to ensure the correct treatment of any observables. The last method used is the parton shower approach described in section 4.2. By this an arbitrary multiplicity of final state partons can be generated though the total cross sections is not affected by the shower, i. e. the cross section remains to be leading order. The parton shower does not guarantee to agree with higher order matrix elements, thus observables might differ in the various approaches.

NLO calculations and the calculation of matrix elements with additional emissions are both correct to a given order in α_s while they suffer from a limited number of final state partons. Moreover the latter method reveals a cut-off dependence. On the other hand, the parton shower is not a fully correct description in α_s but cut-off independent and allows for arbitrary parton multiplicities. As the shower model is built to treat the soft and collinear phase space regions, it

suffers from shortcomings in the description of hard radiation. Thus it is necessary to correct the parton shower as discussed in section 4.3.1 or to combine different methods as described in sections 4.3.2 and 4.3.3.

4.3.1 Matching to the Hard Interaction

The parton shower described in section 4.2 is built respecting factorization. Besides the lower cut-off values t^0 and t^s, there is an upper limit which is related to the hard matrix element, $t_{max} \sim t_{hard}$. The value of the variable t depends on the ordering of the shower and its implementation. This shower that respects factorization will be called *standard shower* in this thesis. It is constructed to describe the collinear and soft limit correctly, but it may fail to describe hard radiation properly due to the phase space limit. There is the possibility to enlarge the phase space by multiplying the upper shower limit with a number bigger than one. A more rigorous approach is to choose the whole centre-of-mass energy available in the hadron–hadron collision as maximum value for the initial state radiation, i. e. $t_{max} = s$. This choice is referred to as *power shower*. The final state shower is not affected by this as the energy available after the hard interaction is limited by the system's invariant mass. It should be noted that the power shower option is not motivated by physical concepts but it is an ad-hoc alternative.

Using the NLO real emission matrix element, it is possible to reweight the first emission in the shower. The probability of a hard emission on NLO tree level at a hard scale t with momentum fraction z is $\mathrm{d}P^{(1)}_{hard}(t,z)$. The emission probability for the parton shower is given by equation 4.6. For the first emission this probability is multiplied by a constant c such that $c\cdot\mathrm{d}P^{(1)}_{PS}(t,z) > \mathrm{d}P^{(1)}_{hard}(t,z)$ for the whole phase space, i. e. for all possible hard interaction scales t and momentum fractions z. The first emission from the parton shower is kept with a probability

$$\frac{\mathrm{d}P^{(1)}_{hard}(t,z)}{c\cdot\mathrm{d}P^{(1)}_{PS}(t,z)}\,, \tag{4.23}$$

therefore correcting the distribution of the first emission to tree level. Sometimes this method is used even for subsequent branchings.

4.3.2 NLO Subtraction Method and Parton Shower Matching

Feynman diagrams have been calculated to NLO for a large number of processes, so it is possible to use NLO results in Monte Carlo event generators. One problem arising is the treatment of the divergences of the real and virtual contributions. For any numerical implementation both parts have to be separately finite. The *subtraction method* is one possibility to achieve the cancellation of the poles in the two terms. To the NLO cross section

$$\sigma_{NLO} = \sigma_{LO} + \sigma_{virtual} + \sigma_{real} \tag{4.24}$$

a term containing the singularities from the real emission part is added and subtracted in a way that both $\sigma_{real} \rightarrow \sigma_{real} - \sigma_{sub}$ and $\sigma_{virtual} \rightarrow \sigma_{virtual} + \sigma_{sub}$ become finite. This corresponds to a cut-off that isolates soft and collinear emissions by effectively shifting them to the non-emission part of the cross section. Using the modified expressions for the real and virtual contributions, the non-emission cross section $\sigma^{(0)}_{NLO} = \sigma_{LO} + \sigma_{virtual}$ and the cross section for a real emission $\sigma^{(1)}_{NLO} = \sigma_{real}$ are finite separately. These equations offer the possibility to use numerical methods for NLO event generation with weights $\sigma^{(0)}_{NLO}/\sigma^{total}_{NLO}$ and $\sigma^{(1)}_{NLO}/\sigma^{total}_{NLO}$, respectively.

Interfacing this result with a Monte Carlo parton shower to include higher orders in α_s leads to difficulties with diverging event weights. Furthermore the final result suffers from *double counting* as the parton shower fills phase space regions already covered by the matrix element emission. To cure these problems the subtraction described above is modified in a way that the first emission from the parton shower is added and subtracted as well, thus leading to

$$\mathrm{d}\sigma^{(0)}_{NLO} = \left(\mathrm{d}\sigma^{(0)}_{LO} + \mathrm{d}\sigma^{(0)}_{virtual} + \mathrm{d}\sigma^{(1)}_{PS}(t_1)\right) \times \Delta(t_1, t^0)\,, \tag{4.25}$$

$$\begin{aligned}\sigma^{(1)}_{NLO} = &\left(\mathrm{d}\sigma^{(1)}_{real}(t_1) - \mathrm{d}\sigma^{(1)}_{PS}(t_1)\right) \times \Delta(t_1, t^0) \\ &+ \left(\mathrm{d}\sigma^{(0)}_{LO} + \mathrm{d}\sigma^{(0)}_{virtual} + \mathrm{d}\sigma^{(1)}_{PS}(t_1)\right) \times \frac{\alpha_s}{2\pi} P\Delta(t, t_1)\Delta(t_1, t^0)\,. \end{aligned} \tag{4.26}$$

The real and virtual matrix element contributions in equations 4.25 and 4.26 are the finite parts constructed above, i. e. the unresolvable emissions on matrix element level are absorbed in $\sigma^{(0)}_{virtual}$. The Sudakov form factor in equation 4.25 ensures that no radiation happens, hence leading to the cross section for no additional parton. Equation 4.26 contains contributions from the corrected real emission part as well as from the parton shower. The emission is assumed to happen at a scale t_1. The first emission is of order α_s, so $\mathrm{d}\sigma^{(1)}_{PS}$ contributes only in the corrected real emission term while its contribution is of order α_s^2 in the parton shower term. The phase space regions available for the parton shower are different for the two contributions to $\mathrm{d}\sigma^{(1)}_{NLO}$ as made obvious by the Sudakov form factors. So the parton shower term added and subtracted is not identical beyond NLO. This *modified subtraction scheme* does not suffer from double counting and reproduces analytical NLO results to a reasonable extent. [55, 56]

In some cases $\mathrm{d}\sigma^{(1)}_{PS}$ is bigger than $\mathrm{d}\sigma^{(1)}_{real}$, leading to negative event weights. With sensible choices for the actual implementation it is possible to keep the number of these events with negative weights at a reasonable small level. It should be mentioned that the subtraction term $\mathrm{d}\sigma^{(1)}_{PS}$ depends on the implementation of the shower. Consequently the NLO calculation has to be redone for a change in the parton shower applied. [55]

4.3.3 Combining Different Multiplicities

The cross section for a process containing n additional partons q_i in the final state on tree level is given by:

$$\mathrm{d}\sigma^{(n)} = \mathrm{d}\Phi^n \overline{\sum} \left|\mathcal{M}^{(n)}(p_i, q_1, \ldots, q_n)\right|^2 . \tag{4.27}$$

To each matrix element a parton shower can be applied to generate additional radiation. The goal is to separate the phase space in a way that each emission happening above a scale μ is calculated on matrix element level while each emission below μ originates from the parton shower. Then several exclusive cross sections $\mathrm{d}\tilde{\sigma}^{(n)}$ can be merged to a single one. Thereby it must be taken care that no phase space region is used twice. To avoid double counting each multiplicity has to be reweighted with the running α_s and Sudakov form factors. By adding Sudakov form factors to equation 4.27 one obtains:

$$\mathrm{d}\tilde{\sigma}^{(n)}(\mu) = \mathrm{d}\Phi^n \overline{\sum} \left|\mathcal{M}^{(n)}(p_i, q_1, \ldots, q_n, \mu)\right|^2 \times \Delta(t, t_1, \mu) \times \ldots \times \Delta(t_n, t^0, \mu) . \tag{4.28}$$

The Sudakov form factor for $\mathrm{d}\tilde{\sigma}^{(0)}$ is simply $\Delta(t, t^0)$, thus the probability for no emission. In equation 4.28 the μ dependence of the Sudakov form factors and the matrix elements has been made explicit. Thus the complete procedure for merging matched cross sections of different multiplicities can be formulated:

1. Generate events of a chosen multiplicity n.

2. Construct the emission scales t_i of all partons.

3. Reweight the whole event with the running coupling $\prod_i^n \frac{\alpha_s(t_i)}{\alpha_s(t)}$.

4. Construct the Sudakov form factors for all emissions and reweight according to equation 4.28.

5. Add the parton shower to the event and veto every emission with a scale above μ. For the highest multiplicity N used the veto scale has to be set to t_N to avoid under estimating the number of emissions.

6. Add up all samples $\mathrm{d}\tilde{\sigma}^{(n)}$.

The different matching scale for the shower in the highest multiplicity bin is necessary to account correctly for higher emission multiplicities not generated by any matrix element. So the shower must be able to populate the phase space region not covered on tree level.

There are several matching schemes available differing in their actual implementation. A detailed description and a comparison of their performance can be found in [57]. For the generators used within the scope of this thesis only the *MLM matching* scheme is needed.

4.3.4 MLM Matching

The MLM matching [57, 58] routine differs from the procedure given in section 4.3.3 as the parton shower is applied to each event without any reweighting with analytical Sudakov form factors. Instead a veto algorithm is applied to reject all events with double counting.

For a given multiplicity n partons are generated on matrix element level above a defined threshold. Then a shower using the whole phase space is added. The final parton state (before hadronization) is clustered back and compared to the initial partons from the matrix element. Each high-energetic cluster has to match exactly one initial parton on tree level. Any event which has remaining clusters above the chosen threshold is vetoed. It suffers from double counting as the phase space region covered by the matrix element is again populated by the shower. Vetoing these events means weighting with a non-emission probability above the scale μ, which corresponds to a Sudakov form factor.

The MLM matching routine as implemented in the Monte Carlo generator ALPGEN, described in section 4.8, uses a p_T^{min} and η^{max} cut-off for the partons generated on tree level. Two partons are constrained to be separated by at least R^{min}. N denotes the highest multiplicity used. The algorithm is as follows:

1. Generate events of a chosen multiplicity n. Each additional parton on matrix element level fulfills $p_T^{parton} > p_T^{min}$ and $|\eta^{parton}| < \eta^{max}$, while $\Delta R < R^{min}$ for each parton pair.

2. Construct the emission scales t_i of all partons.

3. Reweight the whole event with the running coupling $\prod_i^n \frac{\alpha_s(t_i)}{\alpha_s(t)}$.

4. Add the parton shower to each event.

5. The final parton state is clustered by a cone jet algorithm with cone size $R^{cluster}$. Starting with the highest p_T parton each tree level parton is matched to the cluster closest in ΔR, where $\Delta R < R^{match}$ has to be fulfilled. The event is vetoed if any parton is not matched to a cluster or a cluster with $E_T > E_T^{min}$ remains. For $n = N$ additional clusters are allowed to remain as long as they have less energy than all clusters matched to a parton.

6. Add up all samples $\mathrm{d}\tilde{\sigma}^{(n)}$.

The default values used in ALPGEN are $E_T^{min} = 1.2 \cdot p_T^{min}$, and $R^{min} = R^{cluster} = R^{match} = 0.7$. This procedure can be used with a k_T *jet* algorithm as well. The samples with $n < N$ are called exclusive samples while the sample with $n = N$ is called inclusive sample as it contains orders higher than n due to the relaxed matching criterion.

4.4 Underlying Event and Multiple Interactions

In proton–proton collisions one parton from each proton participates in the hard interaction. A beam remnant with a net colour charge remains. Depending on whether a valence quark, sea quark or gluon is taken from the proton, the remnant has different properties. In all cases it is colour connected with the hard scattering, which has to be considered in the hadronization step.

The partons remaining in the beam remnant might interact with partons from the other beam remnant, leading to more than one parton–parton interaction in one collision. These *multiple interactions* lead to additional activity besides the process of interest. The rate of parton–parton interactions is a function of p_T. For the cross section σ of parton–parton interactions one finds:

$$\sigma_{hard}(p_{T\,min}) = \int_{p_{T\,min}^2}^{s/4} \frac{\mathrm{d}\sigma}{\mathrm{d}p_T^2} \mathrm{d}p_T^2 \propto \frac{\mathrm{d}p_T^2}{p_T^4}\,. \tag{4.29}$$

So $\sigma_{hard}(p_{T\,min}) \to \infty$ for $p_{T\,min} \to 0$. Equation 4.29 is true for free partons, but an infinitely soft gluon cannot resolve the colour structure of a proton. Comparison with data shows that a cut-off $p_{T\,min} \approx 2 - 3\,\mathrm{GeV}$ gives a reasonable description, by this eliminating the problem of the soft divergences.

In this thesis the contributions of beam remnants and multiple interactions are summarised as *underlying event*, i. e. all contributions not associated with the hard process or the shower evolution. Often another definition of this term is used also including parton shower contributions but within this work the above definition will be used exclusively. A detailed discussion of the underlying event can be found in [50].

4.5 Hadronization

To get a full description of a physics process, the partons originating from (semi-) hard interactions and the parton shower have to be converted into stable hadrons. The phase transition of low-virtual quarks to colour neutral hadrons is described by non-perturbative models. They serve for a proper treatment of the fragmentation functions.

The hadronization process happens at scales of order Λ_{QCD} whereas the hard scattering and the parton shower are associated with scales $Q \gg \Lambda_{QCD}$. This offers the possibility to factorise the perturbative part and the non-perturbative hadronization step. All models are based on the hypothesis of parton-hadron duality, i. e. the fact that quantum numbers on the hadron level follow those on the parton level. For example the flavour of a hadron in the final jet is associated with the flavour of the original parton, e. g. a b-quark on parton level causes a B-hadron on hadron level.

The emission of particles with $p_T < \Lambda_{QCD}$ is not resolvable in the hadronization process as the separation of emitter and radiated particle is still below $R \sim 1/\Lambda_{QCD}$ when hadronization takes place. Nevertheless, there are a lot of

emitted partons with $p_T \sim \Lambda_{QCD}$ which are called *gluers*. After the parton shower all partons are dressed with a cloud of gluers that can blanch the partons' colour fields, thus allowing high-energetic, colour neutral systems to separate. The colour flow is essential to understand the existence of separated jets consisting of colour neutral hadrons.

A lot of the primary hadrons formed during hadronization are not stable and decay further. Typical resonances have short life times and decay within distances of order $2 - 200\,\text{fm}$, so a clear distinction between hadronization and resonance decays is not possible. This is different for hadrons containing heavy quarks, i. e. c- and b-quarks that travel distances in the order of $10^{12}\,\text{fm}$ before their decay. These decays can indeed be treated independently of the hadronization process. More details about hadronization can be found in [29, 33].

Three different kinds of phenomenological hadronization models exist: *independent hadronization*, *cluster hadronization* and *string hadronization*. The first of these models has several deficiencies and reveals problems when confronted with data. The latter two groups will be presented in the next subsections with a strong emphasis on the models actually implemented in the Monte Carlo event generators presented in sections 4.6 and 4.7.

4.5.1 Cluster Model

Cluster models are based on the observation that the invariant mass spectrum of two colour-connected quarks is steeply falling and independent of both type and scale of the hard process. The distribution is peaked at low values of order a few times the parton shower cut-off. This suggest clusters built from pairs of quarks that decay independently.

After the parton shower is terminated, all remaining gluons are forcibly split in $q\bar{q}$ pairs and colour singlets of two neighboured quarks are built, which can be seen from figure 4.4. This procedure follows the structure of the parton shower as colour-connected partons are close in phase space. The clusters decay independently of each other isotropically in their rest frame into a pair of mesons or baryons conserving flavour and momentum. For a cluster consisting of q_1 and $\bar{q}_2$, a $q\bar{q}$ pair (or di-quark pair) from the vacuum is taken to allow a decay in two mesons with contents $q_1\bar{q}$ and $q\bar{q}_2$. Clusters containing a quark from the hard process remember the quark's direction and the decay product retains the original direction, by this respecting parton-hadron duality. Clusters that are too massive to justify a decay in two hadrons are split in sub-clusters, while clusters too light for a two-body decay undergo a one-body decay by exchanging momentum with neighbouring clusters. Clusters with b- or c-quarks are treated slightly different to improve the description of heavy flavour fragmentation. [29, 33, 59, 60]

The cluster model is a minimalistic model that does not need any adjustable fragmentation function.

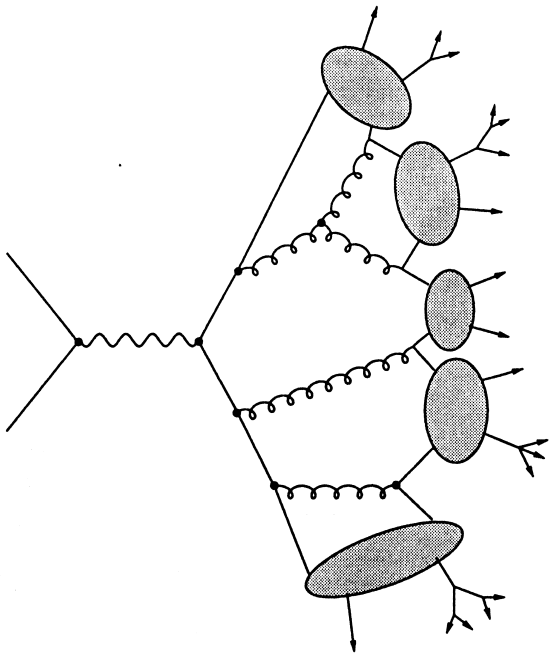


Figure 4.4: Schematic representation of the cluster hadronization model. [29]

4.5.2 String Model

When two colour-connected quarks move away from each other, the colour field between them grows linearly with their distance. It can be regarded as a colour flux tube stretched between the two quarks with a transverse dimension of order Λ_{QCD}, which is very small compared to the string's length. Thus no transverse excitations occur which leads to the picture of a relativistic string with no transverse degrees of freedom and a constant energy per unit length of about 1 GeV/fm. As the potential between the two quarks grows with their distance, at some point a new $q\bar{q}$ pair is produced from the vacuum leading to two separate strings. The string breaks further until each piece corresponds to an on-shell meson. The possibility to create either di-quark pairs from the vacuum or two quark pairs with different colours one after another enables baryon production via the string break-up.

Introducing gluons complicates the structure of the string as they enter carrying energy and momentum, thus leading to kinks in the string, compare figure 4.5. It is worth noting that this treatment of gluons is infrared safe as collinear and soft gluons do not change the string configuration. After all partons are assigned to a colour neutral string, each string is broken up iteratively starting at one of its ends. Postulating that breaking up the string from both sides should be equivalent constrains the shape of the fragmentation function that serves for energy and longitudinal momentum assignment. For the choices of flavour, transverse momentum, etc., a lot of free parameters are needed in this model. [29, 33, 50]

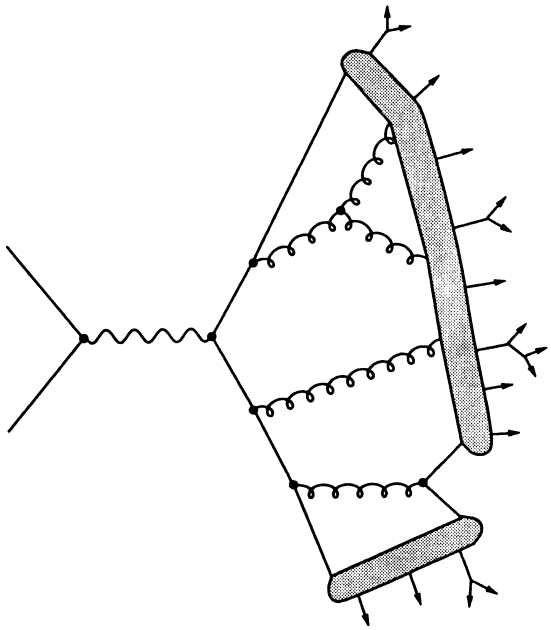


Figure 4.5: Schematic representation of the string hadronization model. [29]

The *string effect* predicts that for a $q\bar{q}g$ system building a string the region between the two quarks is less populated with hadrons after the break-up than the regions between (anti-) quark and gluon. The gluon changes the configuration of the string away from the $q\bar{q}$ axis, see figure 4.5. This fact can be motivated from the dipole picture as well, as the amount of radiation emitted by a $q\bar{q}$ dipole is suppressed by a factor of $1/N_C$ compared to a qg dipole. [33]

Currently different string models are available, from which the best known one is the *Lund model*. Their advantage is the big number of free parameters that allows a good tuning of these models to data. On the other hand they suffer from the number of free parameters as a lot of assumptions are needed for a complete description of the hadronization process.

4.6 HERWIG

HERWIG is a multi purpose event generator that incorporates the full sequence of hard interaction, parton showering, underlying event and hadronization. It can be used for many Standard Model as well as Beyond Standard Model processes at e^+e^-, ep, pp and $p\bar{p}$ colliders.

4.6.1 HERWIG6

The HERWIG6 event generator is a leading order generator containing an angular ordered parton shower, by this including colour coherence effects, as discussed in section 4.2.5.

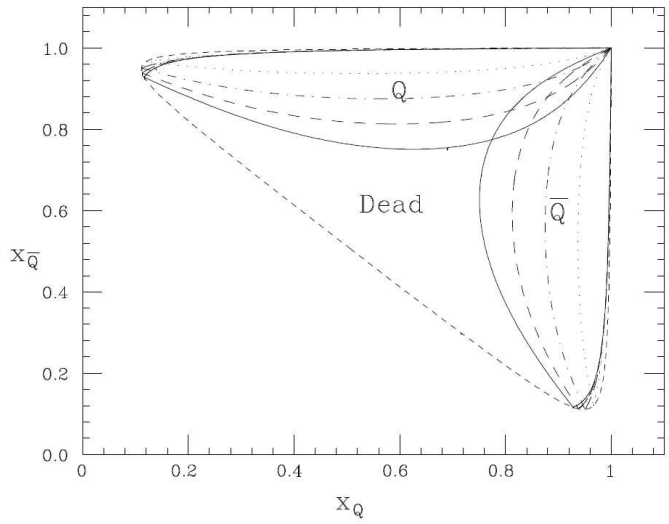


Figure 4.6: Dalitz plot for $b\bar{b}$ production at $s = M_Z^2$. The allowed phase space for the first emission including the dead zone is indicated. The different line shapes belong to different angles between the two quarks. [56]

The parton shower applied in HERWIG6 is ordered in ζ as introduced in equation 4.22. Thus the phase space populated by the shower is restricted to an angular ordered region, resulting in a dead zone, shown in figure 4.6. For timelike final state radiation the phase space boundary is given by $\zeta < 1$, for spacelike initial state radiation by $\zeta < z^2$. The evolution is decreasing in ζ and therefore decreasing in angle. For massless partons at small angles the relation $\zeta \approx 1/2\theta^2$ holds. In a region where the momentum of the emitter fulfils $p \ll p_{hard}$ and $p \gg \Lambda_{QCD}$ the coherent formulation of the parton shower includes even next-to-leading contributions. The radiation from the parton is approximately confined in a cone around it due to angular ordering.

The initial state shower is implemented as backward evolution of a spacelike parton, though all emitted partons have timelike virtualities. In HERWIG6 all timelike partons are evolved until they reach their phase space boundary.

The dead zone not populated by the shower corresponds to hard large angle emissions, that often do not contribute to leading order and are outside angular ordered cones. Nevertheless, one emission in this phase space region can be simulated by first order matrix elements. As two or more emissions in this region happen very rarely, these are neglected. Emissions in the dead zone do not contribute any soft or collinear logarithmic enhancements that have to be handled by the parton shower. This matrix element correction is available only for few processes, including top quark decays but not top pair production.

HERWIG6 provides only a standard shower that respects factorization. For the spacelike case, emissions with $p_T > p_{T\,hard}$ are vetoed to avoid double counting.

HERWIG6 includes an underlying event model based on a parameterization of UA5 data. Depending on $\sqrt{s}$ the multiplicity distribution of charged clusters is

modeled. In addition to this model, an external package called JIMMY is available which is briefly described in the next section.

The whole event is hadronised using the cluster hadronization model described in section 4.5.1. [59, 60]

4.6.2 JIMMY

JIMMY offers an alternative approach for underlying event simulation in HERWIG6, using the matter distribution inside the hadron in the impact parameter space. The number of multiple interactions is calculated according to the overlap of two hadrons obeying Poisson statistics. All additional scatterings are assumed to happen independently and are calculated with hard matrix elements, but neither any ordering of the interactions nor any pdf rescaling is applied. For all partons originating from multiple interactions the parton shower is applied. [61, 62]

The model depends on the assumed hadron radius and a minimum p_T for the additional interactions to tame the divergence for $p_T \to 0$. For JIMMY different tunes to Tevatron data exist. [63, 64]

4.6.3 HERWIG++

HERWIG++ is a C++ version of the HERWIG event generator, whose earlier versions were programmed in Fortran. It includes some improvements compared to HERWIG6, e. g. a completely covariant formulation of the parton shower and mass-dependent splitting functions. The latter ones lead to a correct description of the dead cone region described in section 4.2.5. For HERWIG++ JIMMY is the standard description of the underlying event while the UA5 model is still implemented for comparative studies. [65]

4.7 PYTHIA

PYTHIA is a multi purpose event generator including a broad variety of physics processes. It provides the full generation chain including matrix elements, parton shower, underlying event and hadronization for e^+e^-, ep, pp and $p\bar{p}$ collisions.

4.7.1 PYTHIA6

PYTHIA6 is a leading order generator, i. e. it uses leading order matrix elements. PYTHIA6 provides two different parton shower approaches: one ordered in invariant mass and another one ordered in transverse momentum.

In PYTHIA the hard process chosen is associated with the highest p_T. This is done to avoid double counting of cross sections as processes like jet production might also be caused by the parton shower or the underlying event. Considering the relation $Q^2 = p_T^2/z(1-z)$ for Q^2 ordered timelike showers gives a multiplicative factor of about four for the matching to the matrix element. For spacelike showers this issue is more complicated and any factor between one and four seems

to be reasonable, but again four is chosen as default. So the maximum virtuality of the first emission in both initial and final state radiation is $Q^2_{max} = 4 \cdot Q^2_{hard}$. This is the standard shower option in PYTHIA. Additionally it is possible to use a power shower for the initial state radiation with $Q^2_{max} = s$. For the final state radiation a matrix element matching is performed to correct the first emission to matrix element accuracy. For initial state radiation this is only implemented for few dedicated processes, which do not include $t\bar{t}$ production. Comparison with data has shown that coherence effects have to be taken into account for a correct process description using Q^2 evolution [52]. While the first emission is not affected, for each subsequent branching angular ordering is imposed by the application of the veto algorithm, leading to decreasing angles in the shower evolution. As explained in section 4.2.5 a soft gluon might not resolve the net colour charge of two quarks that are close in angle.

The approach of ordering the shower in p_T is inspired by the colour dipole model. While the partons are still evolved separately, the recoil of a branching is absorbed by the partner from the preceding branching, even if both are not colour connected. Again a standard and a power shower are implemented.

For the underlying event two different models are used depending on the shower model. For the Q^2 ordered shower the so-called "old" underlying event model is used, in which only for the hardest interaction the partons are showered and colour-connected to the beam remnants via strings. All interactions are ordered in $x_T = 2p_T/E_{CMS}$ whereas all additional scatterings have to be softer than the hard process. The number of interactions is given by a Poisson distribution. For the matter distribution in the proton a double Gaussian is assumed, so the number of scatterings can be computed depending on the impact parameter, i. e. the overlap of the two protons. After each interaction the pdf is rescaled to take the changed parton contents in the proton remnant and its energy and momentum loss into account.

For the p^2_T shower a "new" model exists in which initial state radiation and multiple interactions are interleaved. Both are evolved in decreasing p_T with separate Sudakov form factors. Every time an emission or additional interaction happens at $\hat{p}_T$, the evolution is restarted with $\hat{p}_T$ being the new maximum for both. Even for the initial state radiation the energy is taken from the beam remnant, thus the pdf has to be rescaled in both cases as x changes. In this model initial and final state radiation are available for all scatterings and all particles have a colour junction to the beam remnant. The backward evolution scheme raises the possibility of two particles having the same ancestor, so a joined interactions takes place. Up to now this issue is not considered in this underlying event description. Finally it should be noticed that ordering emissions and scatterings in p^2_T does not imply any ordering in time.

After all interactions are calculated and the shower is performed, the Lund string fragmentation model that is described in section 4.5.2 is used to hadronise the event.

A full description of PYTHIA6 is given in [50].

4.7.2 PYTHIA8

While PYTHIA6 was programmed in Fortran, PYTHIA8 is a rewrite in C++. PYTHIA8 provides only the p_T^2 ordered parton shower and the new underlying event model already described in the former section. As an improvement initial and final state radiation are both interleaved with the multiple interactions, and colour reconnections are used by default. The only deficiency is that due to the relatively short period PYTHIA8 exists up to now no tuning to data has been performed. Though its defaults are chosen in a way that its behaviour is quite similar to tune S0 for PYTHIA6. [66]

4.8 ALPGEN

To give an accurate description of multi-jet final states, ALPGEN uses leading order matrix elements for higher order processes like e. g. $t\bar{t}$ plus additional jets.

Events with different parton multiplicities are generated for a certain process and a matching and merging procedure is applied, as discussed in section 4.3.3. Thereby radiation of high energetic partons can be described on matrix element level. Depending on the process different multiplicities are available, for $t\bar{t}$ production up to six additional partons can be generated. These partons have to fulfil requirements like having a minimum p_T and maximum η. ALPGEN generates events on parton level that can be further processed by either HERWIG6 or PYTHIA6 using the MLM matching routine described in section 4.3.4. ALPGEN still uses leading order matrix elements and thus is supposed to reproduce leading order results for fully inclusive observables like jet multiplicity or cross section. [67]

4.9 MC@NLO

MC@NLO is a generator incorporating NLO matrix elements. It generates NLO events on parton level which can be passed to HERWIG6 for further processing, i. e. showering and hadronization.

Thereby MC@NLO combines NLO matrix elements with a parton shower using the subtraction method presented in section 4.3.2. As mentioned before the subtraction method based on equations 4.25 and 4.26 depends on the applied parton shower, therefore the implementation can be used only with this particular shower option.

According to section 4.3.2, this procedure results in events with negative weights but their number can be kept reasonably small. The actual implementation leads to weights corresponding to ± 1, so that the generated events are unweighted up to the sign. [55, 56, 68]

Chapter 5

Generator Analysis

5.1 Transverse Momentum of the $t\bar{t}$ System

Leading order Feynman diagrams describe the production of a $t\bar{t}$ system balanced in the transverse directions, therefore the $t\bar{t}$ system does not gain any transverse momentum. The transverse momentum of the $t\bar{t}$ system is caused by additional radiation, or more technically by real higher order contributions. As α_s is large compared to the electroweak coupling, the transverse momentum spectrum is dominated by gluon radiation. Another weak effect providing the $t\bar{t}$ system with a transverse momentum is the transverse motion of the partons inside the protons, but this effect is small compared to the influences of radiation. Studying the p_T spectrum of the $t\bar{t}$ system thus offers the possibility to examine the influence of real higher order QCD contributions.

The high mass of the top quarks causes the $t\bar{t}$ production to happen at a relatively high energy of at least $2 \cdot m_{top}$. The evolution of the partons to this high scale allows a large number of emissions to take place. Due to the large proton–proton centre-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$, top pairs can be produced from high-energetic partons, providing a large phase space for gluon radiation. So for the incoming partons both is given, a large available phase space region and an evolution to a high scale. The produced top quarks radiate as well, but the available phase space is much smaller in this case. While for the initial state the whole centre-of-mass energy $\sqrt{s}$ is available, the final state is restricted by the energy involved in the hard interaction, reduced by the energy needed for the production of the massive top quarks. Consequently the transverse momentum distribution is governed by initial state radiation, while the effects of final state radiation are much smaller. All investigations shown here include both initial and final state radiation, but parameter changes are applied for the initial state only, as variations in the final state case turned out to be comparably small.

In this chapter the predictions for the p_T distribution of the different generators introduced in sections 4.6 to 4.9 are presented. The goal is a comparison of different parton shower implementations with models that incorporate higher order matrix elements matched to a parton shower. The following figures show the transverse momentum of the $t\bar{t}$ system, where figure a) always shows the region

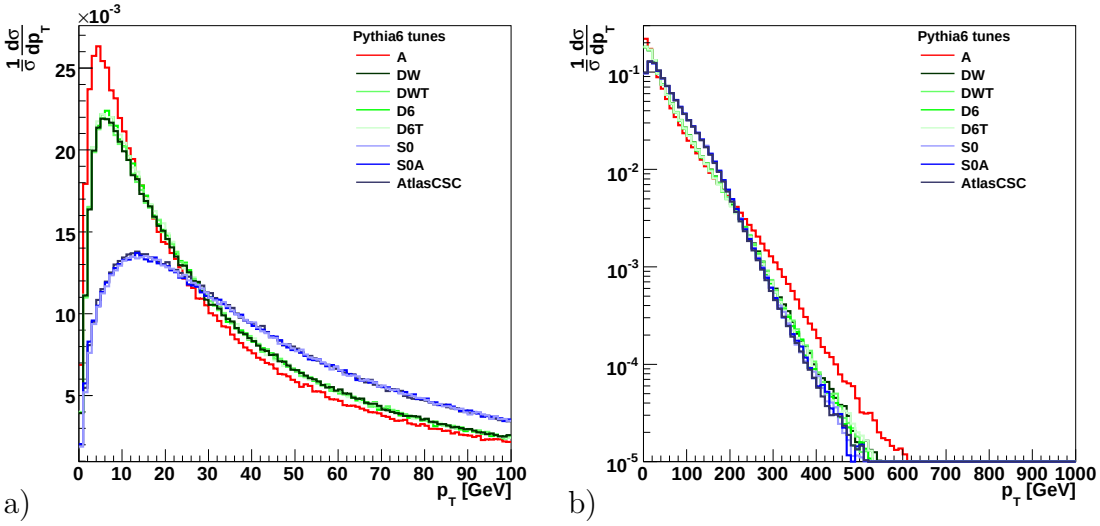


Figure 5.1: $t\bar{t}$ p_T spectrum for different PYTHIA6 tunes. a) shows the soft region, b) the tail on a logarithmic scale. For tunes showing a similar behaviour similar colours are used.

where the peak of the spectrum is located while figure b) shows the tail of the distribution on a logarithmic scale. The soft region with the peak is influenced by the strong coupling α_s which depends on the scales involved in the emission. Technically this dependence is incorporated in the Sudakov form factor used during the evolution. The tail of the transverse momentum spectrum is governed by the upper phase space boundary chosen for the parton shower, i. e. the phase space region where the hardest emission can take place. During the investigation both regions of interest will be considered. In the beginning the emphasis is put on the peak region, while the tail of the distribution will be investigated in depth later.

In the first step changes carried out in the scales in the Sudakov form factor will be analyzed. PYTHIA6 offers a big variety of tunes in which scales in the Sudakov form factor are modified. For HERWIG6 tunes to Tevatron data exist for the JIMMY underlying event model. Here the shower parameters are not tuned, hence the p_T distribution is not affected by the application of different JIMMY tunes. For PYTHIA8 no special tune exists so far and HERWIG++ employs an implementation of JIMMY like HERWIG6. Consequently PYTHIA6 will be used for this analysis step.

Figure 5.1 shows the p_T spectrum for different PYTHIA6 tunes [50, 69, 70]. These tunes are based on comparisons to Tevatron data, especially via minimum bias studies, and tune both the parton shower and the underlying event model simultaneously. In the models applied multiple interactions have no connection to the hard scattering, thus the $t\bar{t}$ p_T distribution is not affected by the actual underlying event description. Therefore the p_T spectrum of the $t\bar{t}$ pair provides a good check for the shower tuning.

Table 5.1: Pythia6 tunes and their most important properties. The keywords "old" and "new" refer to the Q^2 ordered parton shower with the old underlying event (UE) model and the p_T^2 ordered parton shower with the new underlying event model, respectively. A complete description of them is given in section 4.7.1. If no other information is given the tunes were made with the CTEQ5l pdf.

Tune	model	Properties
A	old	Large starting scale
DW	old	Moderate scalings
DWT	old	Similar to DW, more UE activity
D6	old	Like DW for CTEQ6l
D6T	old	Like DWT for CTEQ6l
S0	new	Many multiple interactions
S0A	new	Like S0 with with energy scaling from tune A
Atlas-CSC	new	CTEQ6l

There are tunes using the Q^2 ordered shower and the old underlying event model and tunes using the p_T^2 ordered shower and the new underlying event model. For the former one tunes A, DW, DWT, D6 and D6T are investigated while for the latter one tunes S0, S0A and Atlas-CSC are used. The tunes for the p_T^2 ordered shower and the new underlying event model apply additionally colour reconnections for the final partons. It should be emphasised that every tune depends on a specific pdf. Table 5.1 summarises the tunes of interest and their main features.

In figure 5.1a) three different blocks of distributions can be identified. Tune A shows a significant soft behaviour due to a relatively high scale used in α_s for the shower. All three tunes using the p_T^2 ordered shower and the new underlying event model show a very similar behaviour and have a much harder spectrum than the other tunes applying the old model. Adjusting the tunes to other pdfs, namely from CTEQ5l to CTEQ6l, does not imply any differences in the distributions as expected. Figure 5.1b) shows that the tail of the p_T distribution looking almost identical for all tunes except for tune A. Tune A uses $Q^2_{max} = 4 \cdot Q^2_{hard}$ as starting scale for the shower making a larger phase space region available for radiation. Hence the tail of the p_T distribution is slightly harder compared to the other tunes.

Next, the influence of matrix element matching will be studied, examining the effects of several Alpgen settings. Alpgen applies the MLM matching procedure to combine higher order real matrix elements with the parton shower of Pythia6 or Herwig6. The additional partons produced by Alpgen have a minimum p_T (section 4.8) while emissions with lower p_T are left to the parton shower. Hence the choice of the *matching scale* determines the number of emissions generated by Alpgen and the parton shower, respectively. Figure 5.2

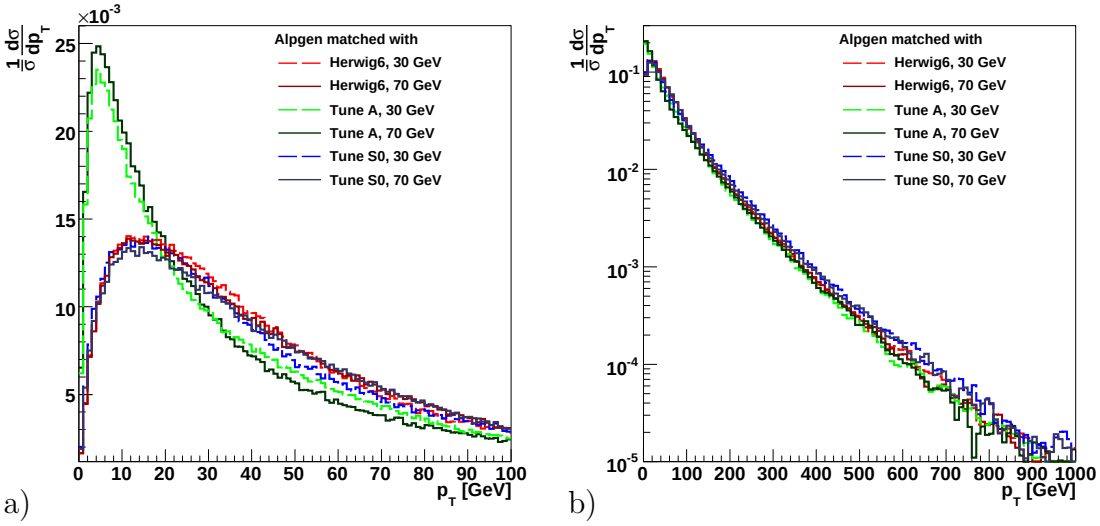


Figure 5.2: $t\bar{t}$ p_T spectrum for ALPGEN matched with different parton showers and matching scales. a) shows the soft region, b) the tail on a logarithmic scale. The soft region is dominated by the parton shower configuration while the tail is completely governed by ALPGEN.

shows the $t\bar{t}$ p_T spectra for ALPGEN samples in which ALPGEN accounts only for the hardest emission above the matching scale while all other emissions are left to the parton shower. The matching scales chosen for comparison are 30 GeV and 70 qGeV. These samples are showered with the HERWIG6 event generator and two different tunes of PYTHIA6, where both a Q^2 and a p_T^2 ordered shower are used. The peak region is driven by the parton shower and therefore depends on the generator tuning. A slight dependence on the matching scale can be observed due to the different treatment of emissions in the region $30\,\text{GeV} < p_T < 70\,\text{GeV}$. While the p_T spectrum gets slightly harder for PYTHIA6 tune A, it gets a little softer for tune S0 when going to a lower matching scale. This shift accounts for the difference in the number of generated emissions, so compared to ALPGEN tune A produces less hard radiation whereas tune S0 produces more. Figure 5.2b) shows that the tail of the distribution is mainly controlled by ALPGEN. Only a very slight dependence on the parton shower can be observed, leading to a little harder spectrum for the p_T^2 ordered choice, but the effect turns out to be small.

Figure 5.3 shows ALPGEN samples matched with PYTHIA6 tune S0 for different parton multiplicities. The number of partons refers to the number of hard extra emissions above the matching scale generated by ALPGEN matrix elements. The number of emissions does not influence the p_T distribution of the $t\bar{t}$ pair very much. An effect can be observed only for a matching scale of 30 GeV, where a second emission causes a small increase in the soft region. As mentioned before PYTHIA6 tune S0 estimates a higher number of hard emissions than ALPGEN, resulting in more soft contributions from the latter one. As in most cases high-p_T emissions happen only rarely, most events in the combined ALPGEN samples

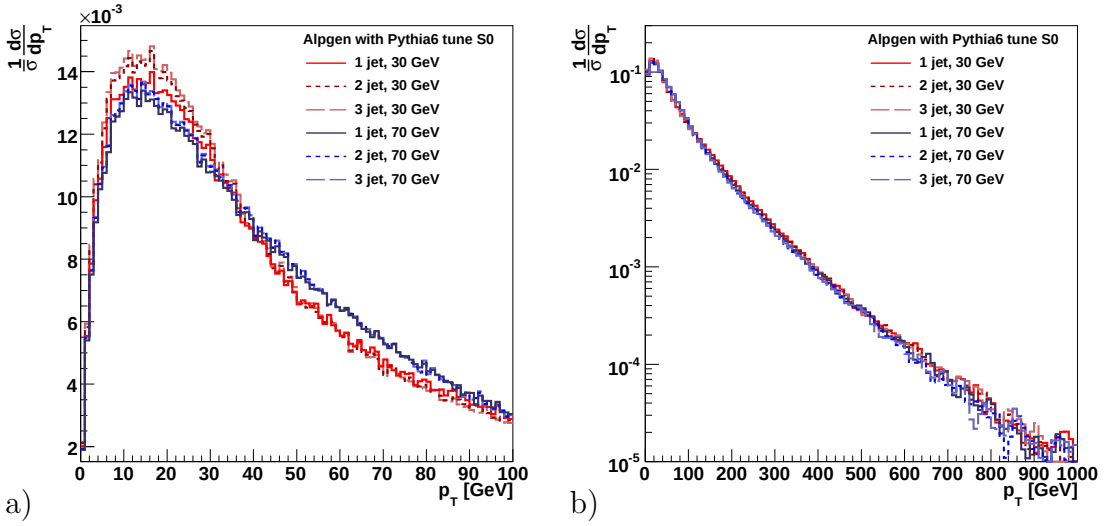


Figure 5.3: $t\bar{t}$ p_T spectrum for Alpgen with Pythia6 tune S0 with different parton multiplicities generated by Alpgen. a) shows the soft region, b) the tail on a logarithmic scale. In the soft region a dependence on the matching scale can be observed, but the effect of the actual parton multiplicity turns out to be small.

contain zero or only one hard emission. This is why the structure of the samples with one, two and three partons is similar. The majority of the events in the three samples is identical.

As already mentioned, the available phase space region plays an important role for the modeling of the tail of the $t\bar{t}$ system's p_T distribution. Hence different choices for the upper phase space limit are compared, first for the angular ordered shower type including a comparison of different matching schemes, then for the Q^2 and p_T^2 ordered type.

Next, a comparison of the analytical NLO result and the MC@NLO generator will be discussed. MC@NLO applies NLO matrix elements and the subtraction scheme discussed in section 4.3.2. So the first hard emission is generated to NLO accuracy. Figure 5.4 compares MC@NLO and Herwig6 with the analytical NLO prediction. Due to multiple emissions the Monte Carlo generators show less entries with low p_T values. The tail of MC@NLO distribution coincides with the analytical NLO result.

The angular ordered shower implementations are compared in figure 5.5, showing MC@NLO, Alpgen with Herwig6, bare Herwig6 and Herwig++. The peak region depicted in 5.5a) illustrates the dependence of the MC@NLO subtraction method on the actually implemented parton shower. The MC@NLO curve is very similar to the pure Herwig6 result. For Alpgen it becomes clear that the soft region is completely modeled by the shower with no visible influence from emissions above the matching scale. The tail of the p_T distribution shows no

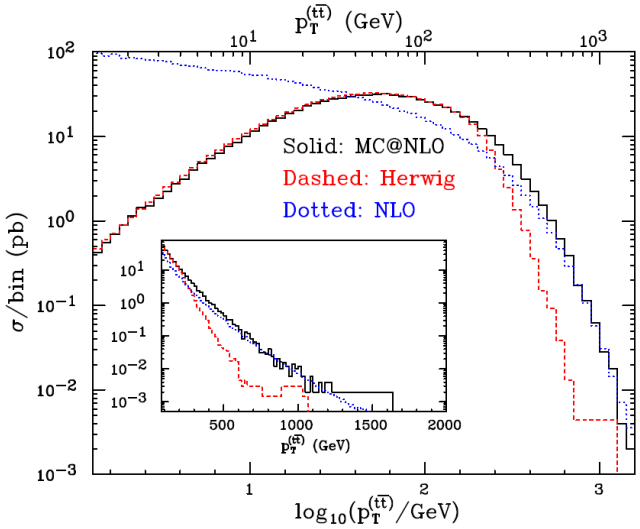


Figure 5.4: The transverse momentum of the $t\bar{t}$ system is shown in a double logarithmic representation, comparing HERWIG6, MC@NLO and the analytical NLO result. [56]

significant difference between ALPGEN and MC@NLO, in good agreement with the results presented in [58]. Furthermore the advanced angular ordered shower implemented in HERWIG++ shows only small differences compared to the old HERWIG6 implementation.

In figure 5.6 Q^2 and p_T^2 ordered shower implementations are shown. The power and standard shower option from PYTHIA8 are compared to PYTHIA6 tune S0, whereas both p_T^2 ordered shower implementations are configured to yield comparable results. In the soft region the shape of the $t\bar{t}$ p_T distribution remains unaffected by the upper phase space boundary. The small deviations are caused by the normalization since with the power shower some events are produced with a high p_T. The difference between PYTHIA8 and tune S0 in the peak region is small. The tail of both p_T^2 ordered showers agrees for both the standard and the power shower version.

In addition the power shower option for tune A incorporating a Q^2 ordered shower and a matched version for tune S0 are presented. From 5.6b) it can be seen that both power showers are harder than the matched version. For the power shower the upper phase space boundary in the spacelike evolution is increased to $\sqrt{s} = 14\,\text{TeV}$, leading to a huge phase space available for additional radiation. In contrast the phase space cut of the standard shower is set by the factorization scale which is usually the scale involved in the hard process. Thus the standard shower is much softer.

The fact that the p_T^2 ordered power shower results in a harder spectrum than the Q^2 one can be motivated with figure 4.3. The p_T^2 ordered evolution favours a large p_T for the first emission while the Q^2 one favours a large Q^2. Thereby the first emission of the p_T^2 ordered shower does more often possess a high p_T. The

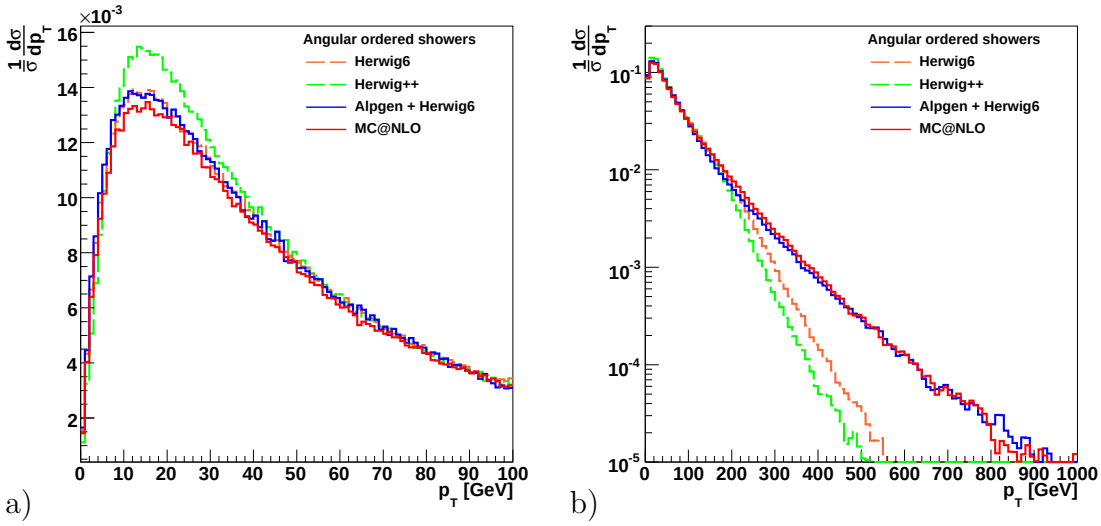


Figure 5.5: $t\bar{t}$ p_T spectrum for different Herwig associated generators. a) shows the soft region, b) the tail on a logarithmic scale. Slight differences between Herwig6 and Herwig++ can be spotted. MC@NLO and Alpgen show a very good agreement.

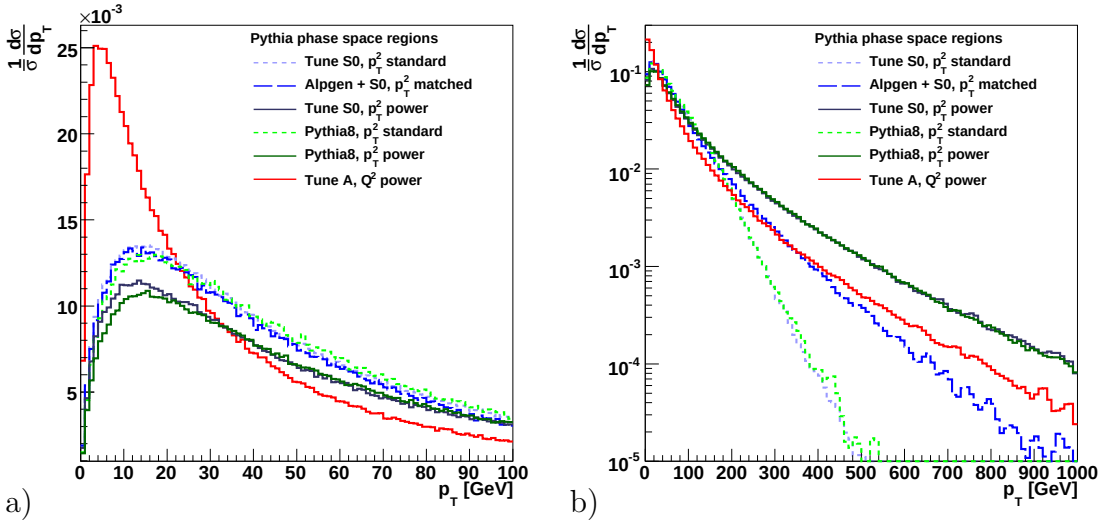


Figure 5.6: $t\bar{t}$ p_T spectrum for different Pythia versions. a) shows the soft region, b) the tail on a logarithmic scale. Pythia8 and Pythia6 tune S0 show very good agreement. It can be observed that the tail of the distribution is much harder for the p_T^2 ordered power shower compared to the Q^2 one.

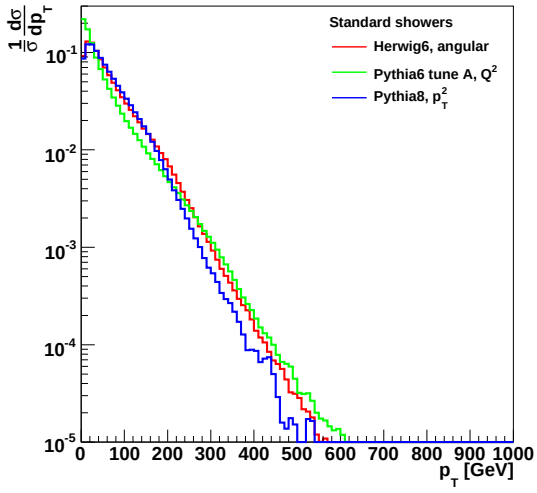


Figure 5.7: $t\bar{t}$ p_T spectrum for the standard showers in different ordered parton showers. The kink in the distribution shows the dependence on the upper phase space limit for the different evolution variables. As the p_T of the $t\bar{t}$ system is directly sensitive to the p_T of the radiation, the p_T^2 ordered shower reveals the strongest dependence.

transverse momentum of the $t\bar{t}$ system is directly sensitive to the p_T of the hard emissions and therefore to the way the phase space is filled by the parton shower.

Figure 5.7 demonstrates another dependence of the $t\bar{t}$ p_T on the phase space. Around a p_T value of the top mass the distribution of the p_T^2 ordered shower shows a kink, reflecting the dependence on the phase space boundary of the shower. The p_T of the $t\bar{t}$ pair is sensitive to the p_T of the emissions, so the upper phase space limit of the shower influences this variable as well. A weaker kink can be observed for the angular ordered shower while it is least pronounced for the Q^2 case. The p_T of the emissions depends on the phase space boundary for those evolution variables as well but not as strong as if the p_T itself is used for the evolution.

Each tune for a generator is done for a specific pdf. Nevertheless, each configuration can be run with another pdf, exploring the influence of the pdf choice. Figure 5.8 depicts the p_T spectrum for different pdfs for PYTHIA6 tune A utilizing a power shower. Even for this extreme configuration no dependence on the pdf can be observed. This was checked for all generators used in this analysis.

Besides the pdf, the strong coupling α_s can be changed in the evolution. PYTHIA8 power shower samples with variations of the value of α_s at M_Z^2 for the spacelike shower are presented in figure 5.9. The change of the reference value of the strong coupling directly influences the Sudakov form factor responsible for the soft region, leading to a significant change in the distribution. The tail of the distributions remains unaffected. The same effect could be observed for the other generators as well.

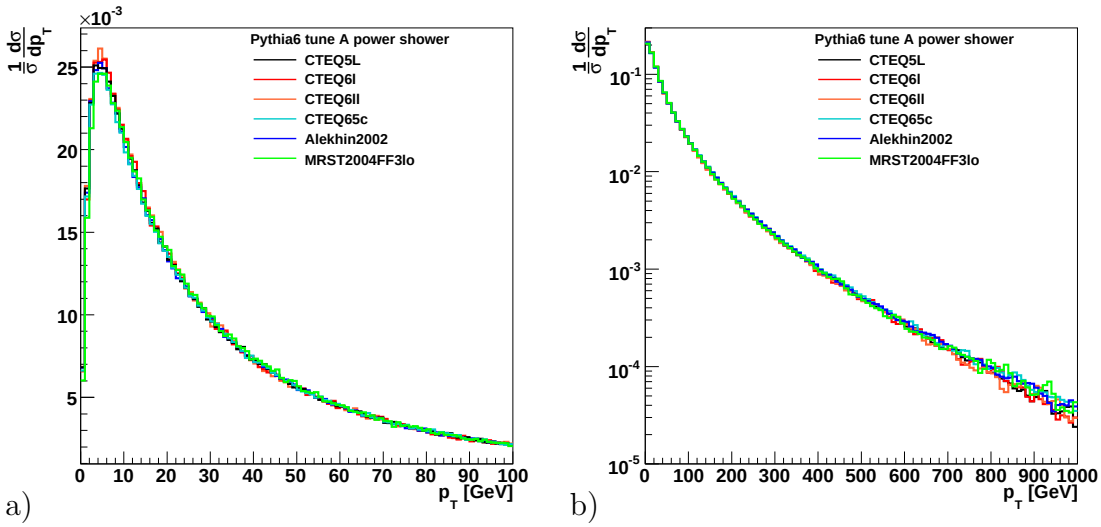


Figure 5.8: $t\bar{t}$ p_T spectrum for different pdfs for Pythia6 tune A. a) shows the soft region, b) the tail on a logarithmic scale. The influence of the pdf choice turns out to be negligible.

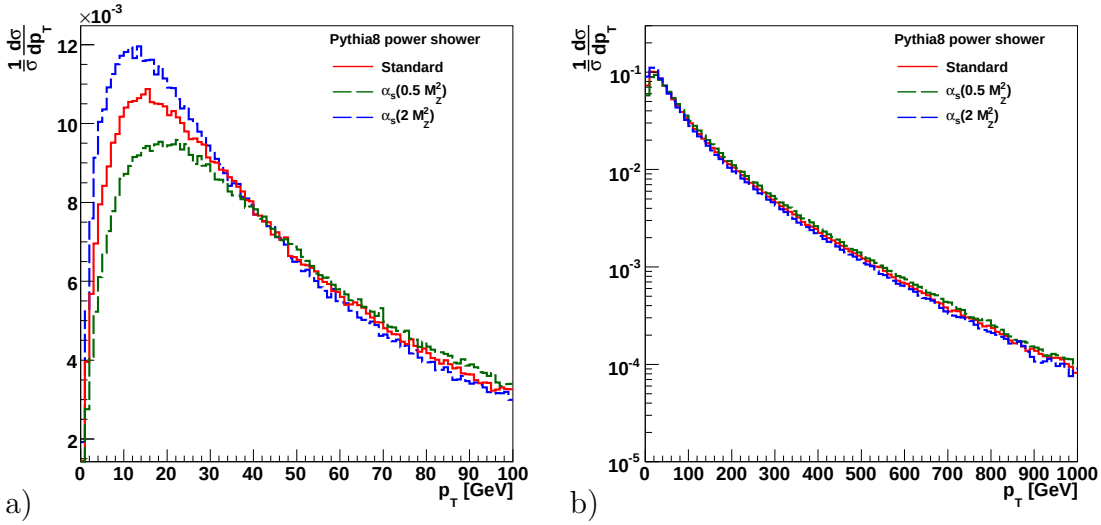


Figure 5.9: $t\bar{t}$ p_T spectrum for different scale choices in Pythia8. a) shows the soft region, b) the tail on a logarithmic scale. A different value of α_s at M_Z^2 has a clear influence in the peak region.

5.2 Parton Jets

For a deeper investigation of the hard emissions jet algorithms are run on the partonic final state. Therefore hadronization is switched off and the top quarks are set stable immediately before their decay. Multiple interactions can take place and the particles are evolved via the parton shower until they reach their lower cut-off value. The top quarks are excluded from the list of particles and a jet algorithm is run on the partons. The jets obtained from this procedure will be referred to as *parton jets*. If not stated otherwise the KT4 algorithm is used. For the results presented in sections 5.2.1 and 5.2.2 no dependence on the algorithm was observed when comparing different settings for the k_T and SISCone jet algorithm.

5.2.1 Transverse Momenta of Parton Jets

QCD radiation is one of the main sources for additional jets, hence it is expected that the hard emissions will manifest themselves in hard jets. In this case hardness refers to the transverse momentum. Figure 5.10 shows the p_T distribution of the two hardest jets on a logarithmic scale for several PYTHIA configurations. The hardest jet shows a strong correspondence to the transverse momentum of the $t\bar{t}$ system. Again the power shower allows for much more high-p_T radiation while the standard shower is much softer, showing again a kink due to the upper phase space boundary. The fact that the p_T^2 ordering favours high-p_T radiation is again clearly visible when comparing the two evolution scales for the power shower option. The Q^2 shower still produces more hard radiation than is done on matrix element level.

For the second hardest jet the same structure is present but the tail of the Q^2 ordered power shower is even slightly softer than the matched p_T^2 version. Furthermore the spectrum of the second hardest jet is much steeper. Both aspects show that at least the hard p_T tail of the $t\bar{t}$ system is in most cases governed by the hardest emission. It was checked that the underlying event model does not influence the high-p_T jets as expected, but leads to differences in the spectra only in the low-p_T region. So for most of the cases where the $t\bar{t}$ system possesses a high transverse momentum a jet in the opposite direction can be found, leading to a balanced p_T of the whole system.

In figure 5.11 the p_T of the leading jets for the generators presented already in figure 5.5 is shown. While the standard showers of HERWIG6 and HERWIG++ behave as expected, an interesting difference can be found when comparing ALPGEN with MC@NLO. While the $t\bar{t}$ p_T distribution is almost identical for both generators the hardest emission from MC@NLO is significantly softer. This effect was already reported in [58]. For the second jet the difference between MC@NLO and HERWIG6 is almost negligible. The subtraction scheme used for the first emission does not influence the hardness for the second emission. On the contrary, ALPGEN sets a new starting scale for the shower depending on the scale of the emission already calculated on matrix element level, thereby leaving more phase

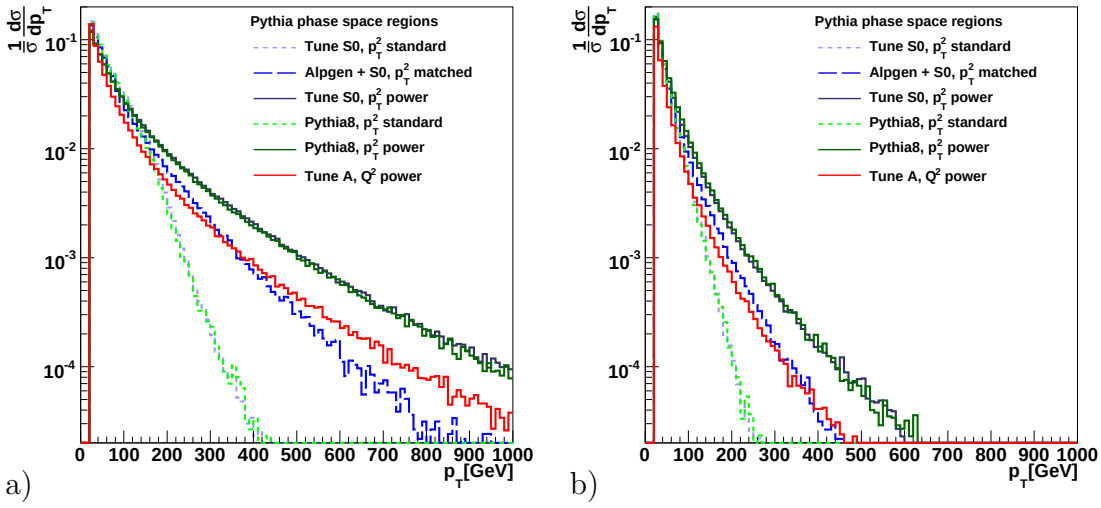


Figure 5.10: The p_T distribution of the first (a) and second (b) hardest jet on a logarithmic scale for different PYTHIA versions is shown. The distribution of the hardest jet follows the one obtained from the $t\bar{t}$ system. The second hardest jet shows a significant dependence on the ordering, as the second hardest jet of the Q^2 power shower has a p_T comparable to the one from the matched shower.

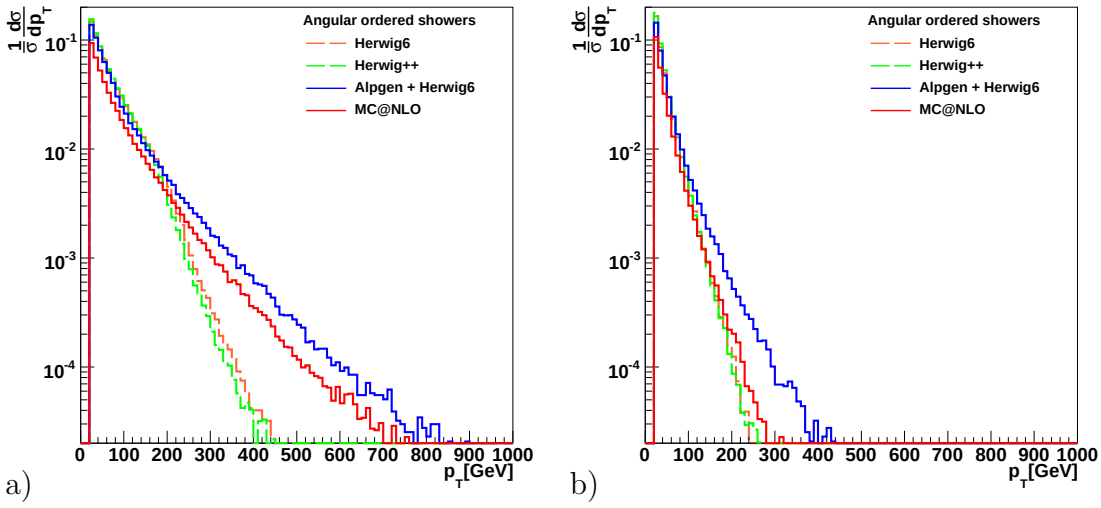


Figure 5.11: The p_T distribution of the first (a) and second (b) hardest jet on a logarithmic scale for different HERWIG versions is shown. In spite of the good agreement in the p_T spectrum of the $t\bar{t}$ system, MC@NLO and ALPGEN show a significant difference in the p_T distribution of the hardest jet. The subtraction method does not influence the second hardest emission whereas ALPGEN sets a new starting scale for the shower, leaving more phase space available.

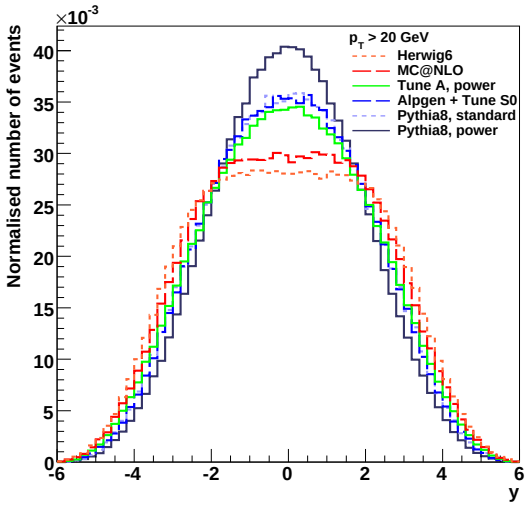


Figure 5.12: Rapidity distribution of the leading jet for different generators, normalised to the actual number of events. In contrast to the other generators the HERWIG shower causes a plateau in the central region.

space available for the second one. It is interesting that despite the difference in the jet spectra the $t\bar{t}$ spectrum agrees very well for both generators.

5.2.2 Rapidity of Parton Jets

Constructing jets from the partonic final state with excluding the top-quarks offers the possibility to determine the direction of the first emission. While there is no preference concerning the azimuth, the pseudo-rapidity of the jets shows a clear structure. As the jets are built using the E-scheme they obtain a mass and rapidity and pseudo-rapidity are not identical. Nevertheless, the jet masses are not very high and both variables do not differ much. This fact will be exploited later when jets on detector level are examined and compared to the generator level.

Figure 5.12 shows the rapidity of the hardest jet for different generator configurations with a jet p_T above 20 GeV. In all cases a similar structure is obtained, but the shape differs in the region $-2 < y < 2$. The p_T^2 ordered power shower shows the most significant bump while HERWIG6 and MC@NLO form a plateau in this region. This effect gets even more pronounced when looking at the rapidity of the leading jet in the frame of the $t\bar{t}$ system, i. e. the quantity $\Delta y = y_{jet} - y_{t\bar{t}}$, presented in figure 5.13.

Sub-figure 5.13a) shows the case for all jets with $p_T > 20$ GeV, exposing a strong dip structure for the HERWIG6 and MC@NLO case. The other generators show a more or less distinct bump, preferring high-p_T emissions at low rapidities, i. e. at wide angles. The structure apparent for HERWIG is caused by the existence of the dead zone explained in section 4.6. Representing the available phase space in variables $x = M^2_{Q\bar{Q}}/s$ and $y = \cos\theta_g$ as done in figure 5.14, the origin of the

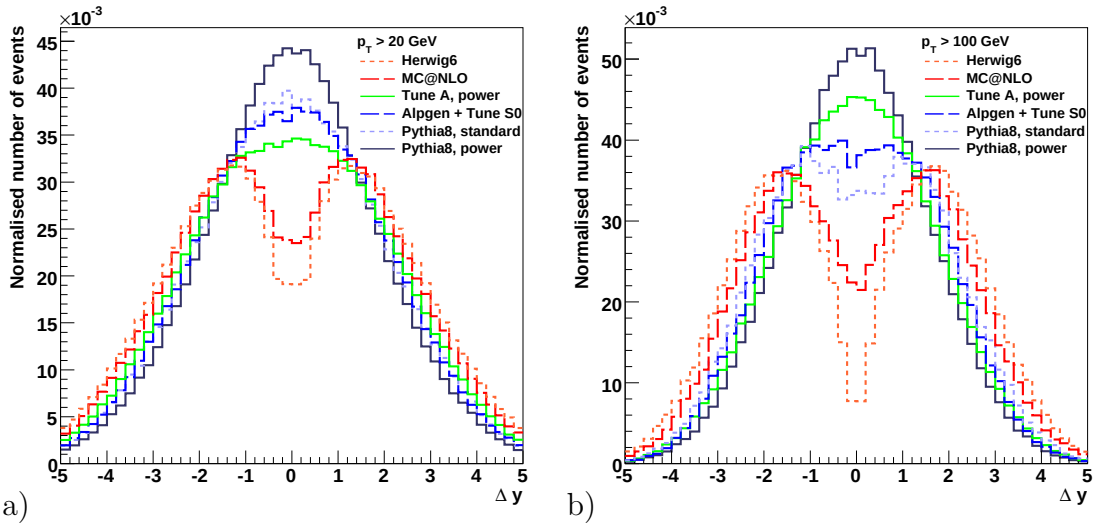


Figure 5.13: Rapidity of the leading jet with respect to the $t\bar{t}$ system: $\Delta y = y_{jet} - y_{t\bar{t}}$. The dead zone from the Herwig shower causes a dip in the central region. The distributions are normalised to the actual number of events.

dip becomes clear. In figure 5.15 a sketch of a $t\bar{t}$ pair emitting a gluon is shown. For a high energetic gluon the invariant mass of the $t\bar{t}$ system is small compared to the available energy, i. e. x is small. The emission angle of the gluon with respect to the $t\bar{t}$ system θ_g is limited due to the dead zone as shown in figure 5.14. The phase space is indeed restricted such that hard wide angle emissions are forbidden due to the phase space boundaries implied by angular ordering. For the Q^2 ordered shower angular ordering is imposed which seems to cause the difference between the Q^2 and p_T^2 ordered distributions.

Figure 5.13b) shows the relative rapidity of the leading jet for jets fulfilling $p_T > 100\,\text{GeV}$. It is obvious that the dip structure gets even more pronounced for harder jets. MC@NLO follows the Herwig6 distribution though the emissions on matrix element level account for some emissions into the dead region. Nevertheless, the subtraction method used makes the first emission sensitive to the parton shower applied, leading to a dip structure for MC@NLO as well. The fact that the Herwig6 distribution does not go exactly to zero is owed to multiple emissions. The consideration of only very hard jets in 5.13b) exhibits a slight dip structure for the standard shower options of Pythia as well, indicating that here some part of the phase space is missed as well.

While the effect of the dead zone caused by Herwig6 is present for MC@NLO as well, the case of Alpgen combined with Herwig6 is more complicated. Figure 5.16 shows the rapidity of the leading jet for Alpgen showered with Herwig6. A slight dependence on the matching scale can be observed, leading to a rise in the central region for a lower matching value. In this case more emissions are calculated by the Alpgen matrix elements, filling the phase space left out

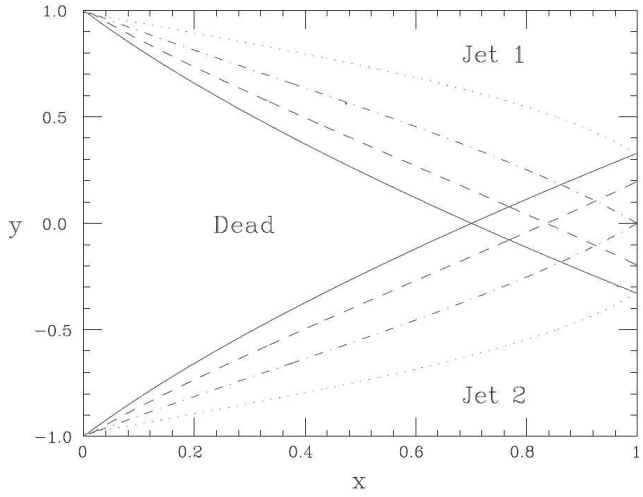


Figure 5.14: Phase space indicating the dead zone for $b\bar{b}$ at $s = M_Z^2$ in $x = M_{Q\bar{Q}}^2/s$ and $y = cos\theta_g$. The different line shapes belong to different angles between the two quarks. For lower values of x more energy is available for the emission. Coincidently lower values of x restrict the emission angle θ_g, leading to a dead zone in the phase space and suppressing hard emissions at wide angles. [56]

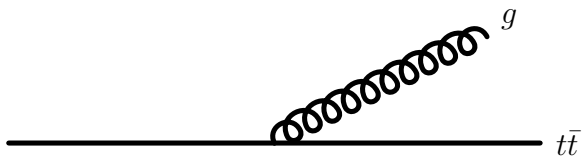


Figure 5.15: Schematic representation of $t\bar{t}$ system emitting a gluon.

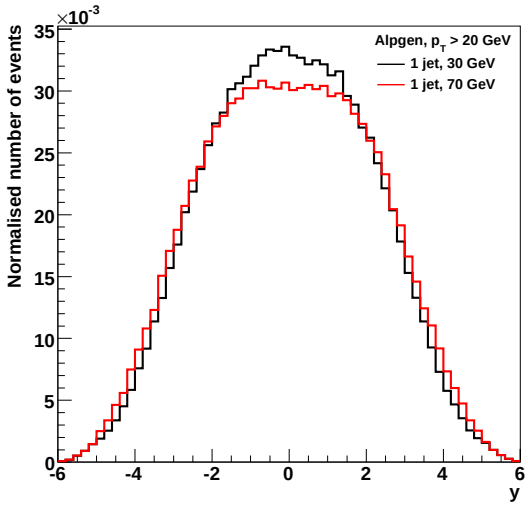


Figure 5.16: Rapidity distribution of the leading jet for ALPGEN with HERWIG6 with different matching scales. The lower matching scale causes more radiation generated by ALPGEN.

by the HERWIG6 shower. This gets even more obvious when looking at figure 5.17 where again the Δy distribution for jets above 20 GeV and 100 GeV is presented. In 5.17a) a clear difference can be observed due to the fact that for a lower matching scale ALPGEN controls a larger fraction of the emissions. Figure 5.17b) shows the case where the jet cut is well above the matching scale, so the jets considered here are expected to be caused by ALPGEN emissions. The two curves agree very well and the dip structure vanishes. A comparison between the ALPGEN and MC@NLO curves emphasises the strong dependence of the MC@NLO subtraction method on the parton shower. It should be mentioned that HERWIG++ shows a dip structure similar to HERWIG6, as both HERWIG versions contain a dead zone, though they are not fully identical. A detailed comparison and explanation of the dead zones and Δy spectrum for Higgs events is given in [71].

5.2.3 Parton Jet Multiplicities

The observed jet multiplicity does not depend on one single aspect. In this section the influences of phase space boundaries, soft emissions and the underlying event are investigated, as well as the impact of the chosen jet algorithm.

Figure 5.18 shows the jet multiplicities for different underlying event models for HERWIG6 for a p_T cut of 20 GeV and 50 GeV. In sub-figure b) the same distribution as in a) is depicted on a logarithmic scale. The different tunings [63, 64] are performed to describe minimum bias data from Tevatron. The extrapolation to LHC energies reveals a huge difference when checking the charged particle multiplicities. In the saturated region tuning A predicts about twice the number of particles compared to the Atlas tune. In this study tuning A leads indeed to

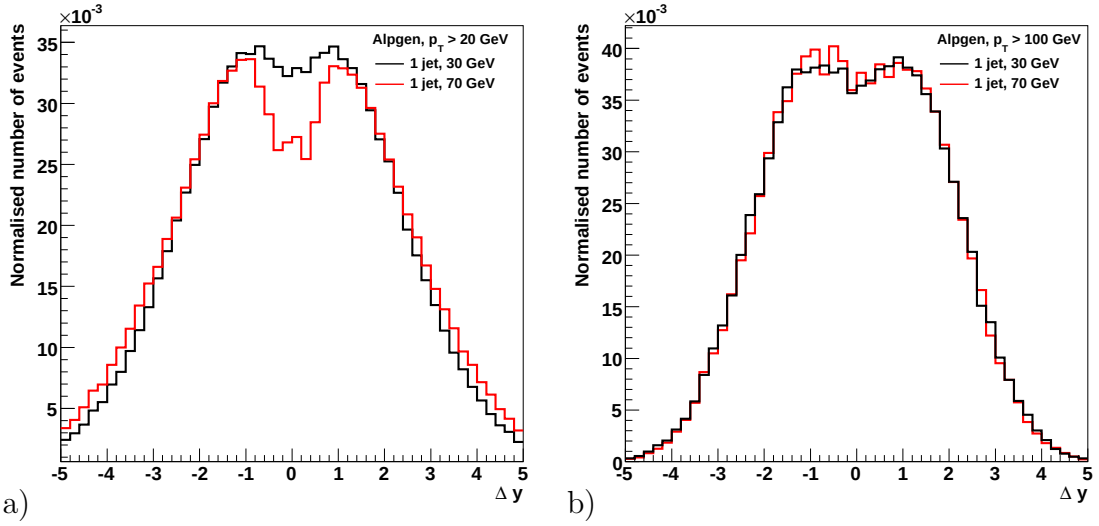


Figure 5.17: Rapidity of the leading jet with respect to the $t\bar{t}$ system $\Delta y = y_{jet} - y_{t\bar{t}}$ for ALPGEN with HERWIG6. a) For a jet p_T cut of 20 GeV the matching scale exhibits a significant difference, as for the lower matching scale ALPGEN generates more emissions by this filling the gap caused by the dead zone. b) For a jet p_T cut of 100 GeV the matching scales, both well below 100 GeV, do not have any influence.

higher jet multiplicities than the other models. The UA5 model gives the lowest contribution to additional jets as expected. For jets with $p_T > 50$ GeV shown in figures 5.18c) and 5.18d) the difference between the different underlying event models turns out to be negligible. The contributions from the underlying event are mainly in soft regions and have only limited influence on high-p_T objects.

Figure 5.19 presents the jet multiplicities for different PYTHIA6 tunes and phase space boundaries again for a transverse momentum cut of 20 GeV and 50 GeV. Figures 5.19a) and 5.19c) both show significant differences in the jet multiplicities depending on the tunes. The simultaneous tuning of underlying event and soft emissions causes a dependence of the multiple interactions and the Sudakov form factor on the specific tune. Tune A leads to little soft radiation compared to the other tunes which causes a large number of events with zero additional jets. Here again the scaling of α_s in the Sudakov form factor is important (compare figure 5.1).

In addition the choice of the upper phase space boundary has impact on the number of jets, preferring higher multiplicities. Especially 5.19d) underlines the influence of the upper phase space limit on the number of events with many jets. This is not only the case for the power shower, but also for the standard shower option. The tune A standard shower option has a higher starting scale and produces more events with high jet multiplicities than tune DW standard. As already seen in the analysis of the $t\bar{t}$ p_T distribution the evolution variable affects the p_T of the emitted particles. The p_T^2 ordered shower favours high p_T

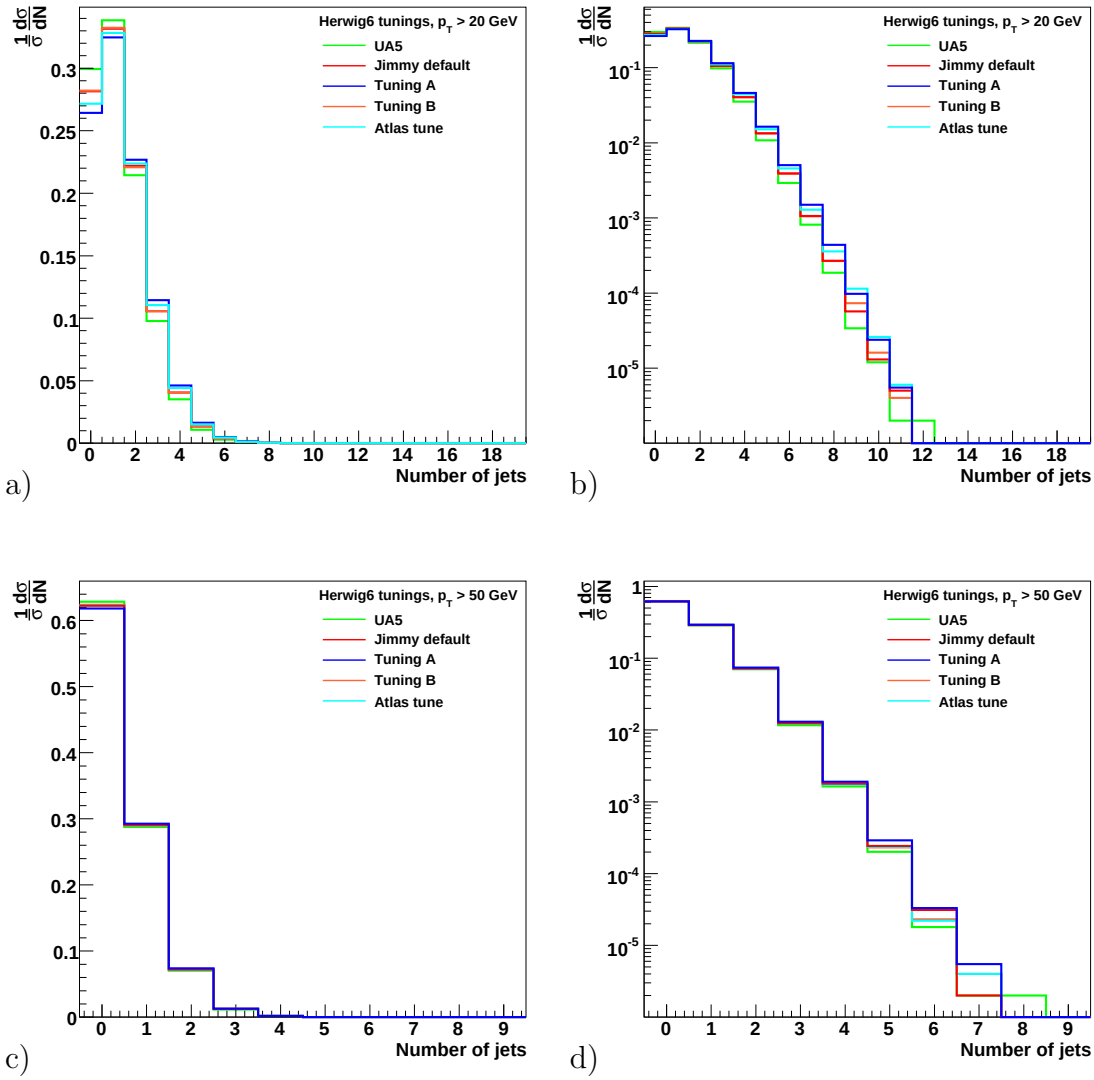


Figure 5.18: Jet multiplicities for the KT4 algorithm for different Herwig6 tunings. Jets above 20 GeV are shown in a) on a linear and in b) on a logarithmic scale. Jets above 50 GeV are shown in c) on a linear and in d) on a logarithmic scale.

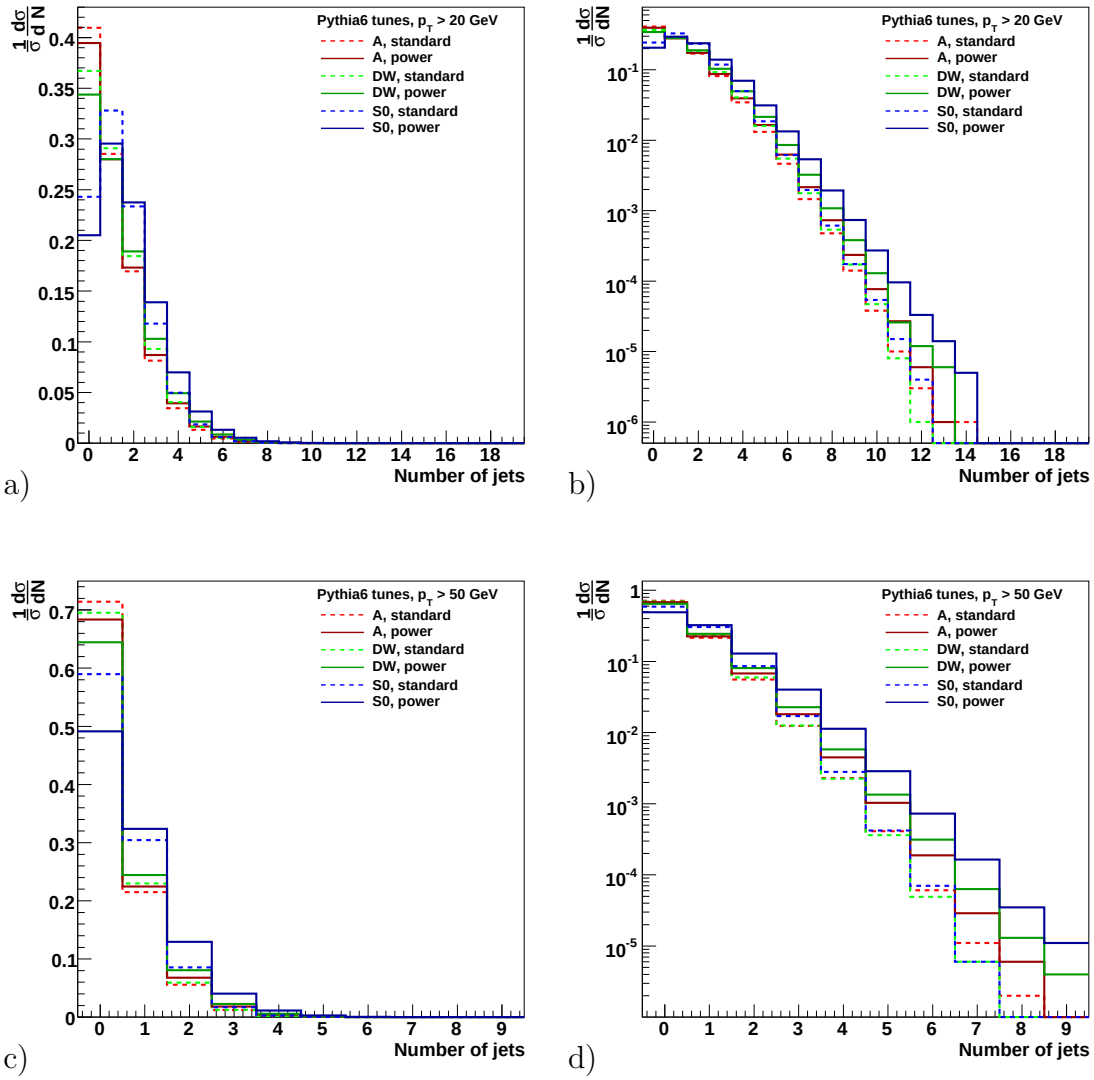


Figure 5.19: Jet multiplicities for the KT4 algorithm for different PYTHIA6 tunings. Jets above 20 GeV are shown in a) on a linear and in b) on a logarithmic scale. Jets above 50 GeV are shown in c) on a linear and in d) on a logarithmic scale.

values in the evolution which seems to cause higher jet multiplicities compared to the other evolution variables.

The jet multiplicities presented in figures 5.18 and 5.19 were obtained using the KT4 jet algorithm. Increasing the R parameter to 1.0 (KT10) changes the jet multiplicities significantly. This is shown in figures 5.20a) and 5.20b) where again different Jimmy tunings for Herwig6 are compared. Due to the higher value of the R parameter the jet algorithm combines more particles into a single jet. Thus more particles from the underlying event are collected, leading to a stronger dependence on the actual underlying event description. Figures 5.20c) and 5.20d) show that jets with $p_T > 50\,\text{GeV}$ are clearly less affected by a change of the R parameter, because most of the particles caused by the underlying event are low-energetic and hence the relative impact on high-p_T jets is much weaker. Checking the SISCone algorithm with SC4 and SC7 reveals no big difference and the jet multiplicities are reasonably stable, showing a behaviour comparable to the one of the KT4 algorithm.

Concerning physics effects, the jet multiplicity turns out to depend on a combination of the available phase space, the scale in the Sudakov form factor and the underlying event model applied. So different choices of these parameters can give comparable results. It follows that it is difficult to study one specific effect investigating only the jet multiplicity. For a particular study additional information is needed.

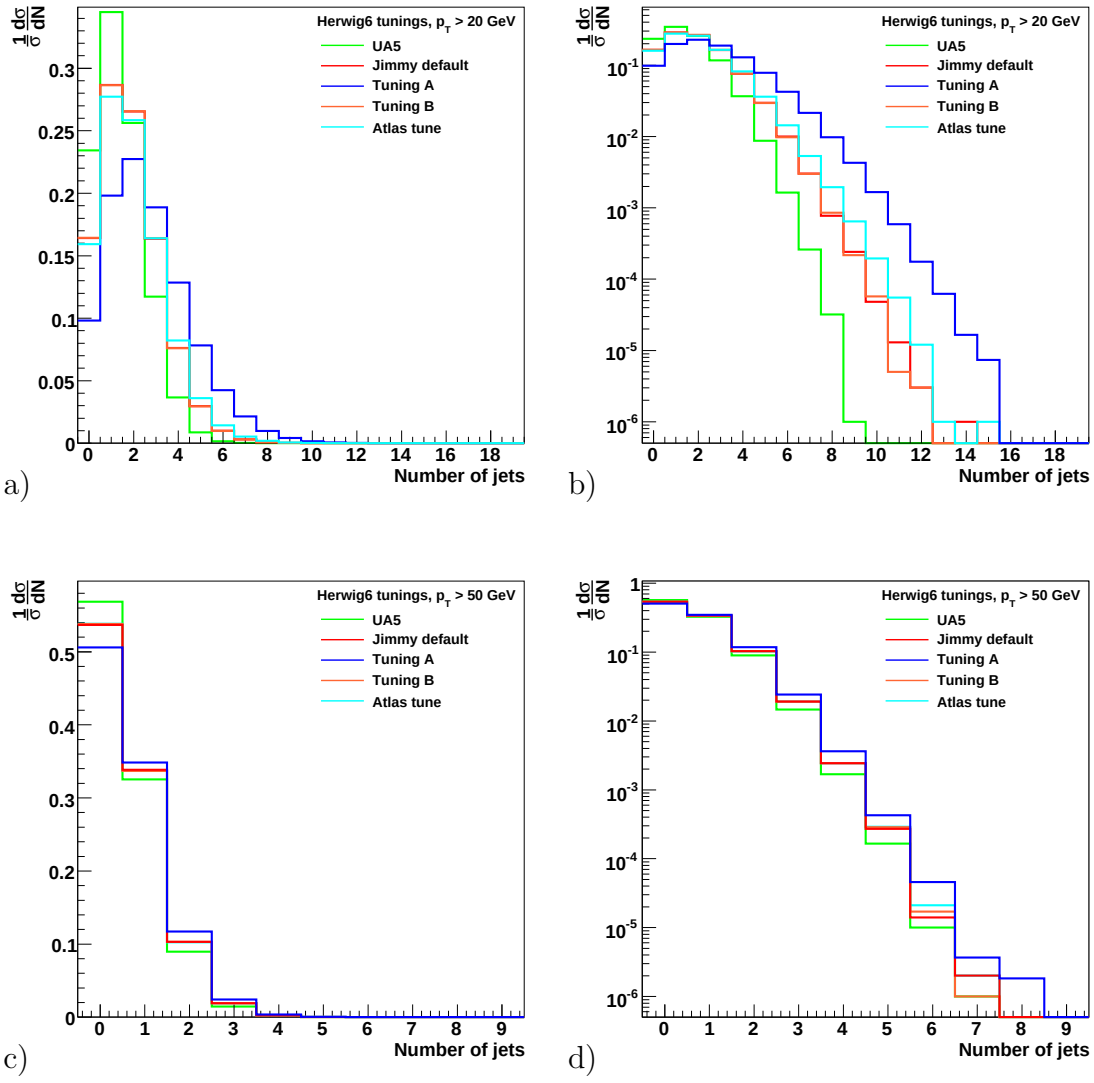


Figure 5.20: Jet multiplicities for the KT10 algorithm for different HERWIG6 tunings. Jets above 20 GeV are shown in a) on a linear and in b) on a logarithmic scale. Jets above 50 GeV are shown in c) on a linear and in d) on a logarithmic scale.

Chapter 6

CMS Detector

The CMS detector is designed for a precise measurement of particles produced in proton–proton collisions at the LHC. With a rate of 40 MHz, i. e. one collision every 25 ns, 10^9 inelastic events per second are expected for the LHC design luminosity. This results in approximately 1,000 particles emerging from the interaction region every 25 ns. The detector features required for these conditions are good time resolution and high granularity to gain a low occupancy, as well as radiation hardness to minimise the damage due to the intense high-energetic particle flux.

The CMS detector shown in figure 6.1 has a traditional onion-like structure: The interaction point is surrounded from inside to outside by a *tracking system*, an *electromagnetic calorimeter*, a *hadronic calorimeter* and finally a *muon system*.

6.1 Magnet

The detector design is driven by the choice of the magnetic field. To gain a large bending power and thus the feasibility to measure momenta to very high precision a superconducting solenoid with a magnetic field strength of 4 T has been chosen. Outside the magnet an iron yoke of 10 ktons weight returns the magnetic field, yielding a field strength of about 2 T. The magnet leads to a compact design of the detector, that has a length of 21.6 m, a diameter of 14.6 m and a complete weight of 12,500 tons. Tracker and both calorimeters are accommodated inside the magnet while the muon detectors are integrated in the iron return yoke, compare figure 6.1. Details about the magnet are given in [73].

6.2 Tracker

The tracking detector is used for the measurement of charged particles' trajectories in the magnetic field and thus enables the determination of their charge sign and momentum. Hence it provides very important information for the reconstruction of charged leptons and hadrons. As mentioned above a short response time and high granularity are needed for precision measurements. Furthermore the

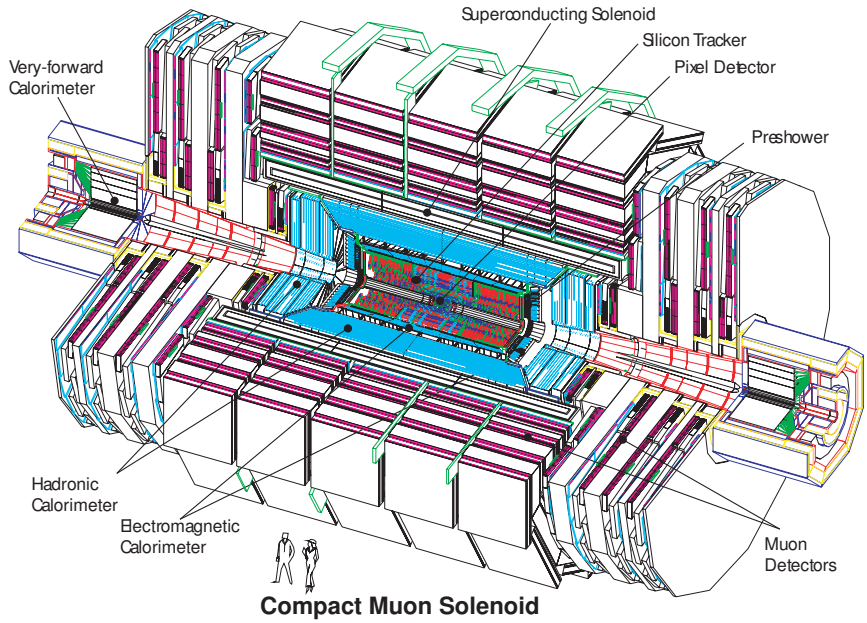


Figure 6.1: The CMS detector. [72]

number of interactions with the tracker material has to be kept at a minimum to avoid any distortions of the measurement, so the amount of matter in the tracker should be as small as possible. In CMS the whole tracking system is built from silicon, resulting in an active area of 200 m^2. The principle of semi-conductor detectors is described in [74].

A schematic sketch of the tracker is shown in figure 6.2. The tracking detector covers a pseudorapidity range up to $|\eta| < 2.5$. Around the beampipe the barrel of the *pixel detector* is placed, consisting of three layers of silicon pixels. It is complemented by two layers in the pixel endcaps on each side. There are altogether 66 million pixels of a typical size of $100 \times 150\,\mu\mathrm{m}^2$ at a distance between 4.4 cm and 10.2 cm to the beampipe. The pixel detector serves for precision measurements of trajectories near the primary interaction point, by this providing the means to secondary vertex reconstruction. The occupancy of each pixel is kept at 10^{-4} per bunch crossing.

Outside the pixel a *silicon strip detector* built of 9.3 million strips is mounted. It is divided in two barrel parts, *tracker inner barrel* (TIB) and *tracker outer barrel* (TOB), which are complemented by endcaps called *tracker inner disk* (TID) and *tracker end cap* (TEC) (compare figure 6.2). The occupancy of the strips is between 1% and 3% per bunch crossing. The dimensions of the whole tracker are a length of 5.8 m and a diameter of 2.5 m.

Each sufficiently energetic particle emerging from the interaction point, that does not curl due to the magnetic field, traverses at least three layers in the pixel detector and nine layers in the strip tracker, of which four provide two-dimensional information. The material crossed by a particle corresponds to $0.4 - 1.8$ radiation lengths, depending on the direction of flight.

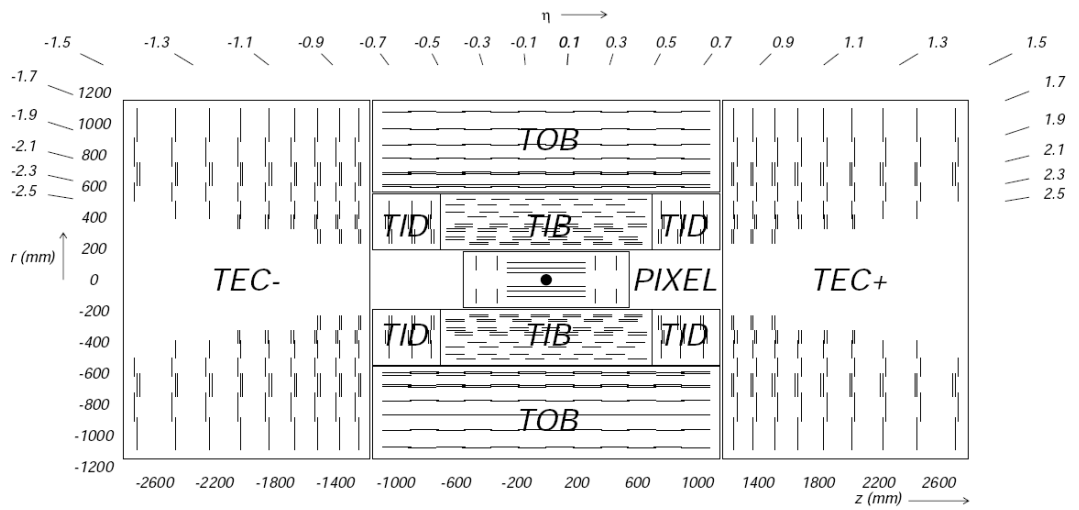


Figure 6.2: Schematic cross section of the CMS silicon tracker. [10]

In figure 6.3 the expected resolution performance for muons is shown. For high momentum tracks the resolution of the transverse momentum is around $1 - 2\%$ up to $|\eta| \approx 1.6$ while the transverse impact parameter can be measured to about $10\,\mu$m accuracy, together allowing a very precise track reconstruction. The track reconstruction efficiency yields 99% for muons up to $|\eta| \approx 2$. Due to a higher interaction rate with the tracker material, this number drops to $80 - 90\%$ for hadrons as shown in figure 6.4.

The silicon strip detector is operated at a temperature of $-10°$C to reduce radiation damage and to allow its operation over the full lifetime of the experiment which is expected to be ten years. A full description of the CMS tracker can be found in [10, 75, 76].

6.3 Electromagnetic Calorimeter

The electromagnetic calorimeter is built for precision measurements of electrons and photons, covering a pseudo-rapidity range up to $|\eta| = 3$. It is composed of a barrel and an endcap as shown in figure 6.5. The material used is lead tungstate, which has a density of $8.28\,\mathrm{g/cm^3}$, a Molière radius of 2.2 cm and a *radiation length* of $X_0 = 0.89\,\mathrm{cm}$. With over 75,000 crystals the electromagnetic calorimeter provides a high granularity of $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$ in the barrel region, decreasing to 0.05×0.05 in the endcap. The number of radiation lengths is about $25 \cdot X_0$. The electromagnetic shower causes light emission which is detected by photodiodes located at the back of the crystals. 80 % of the light emission happens within 25 ns, which means that the scintillation time is of the order of the LHC bunch crossing time.

The calorimeter is operated at a temperature of 18°C to establish an equilibrium between radiation damage and recovery, for which a stability of 0.05°C is

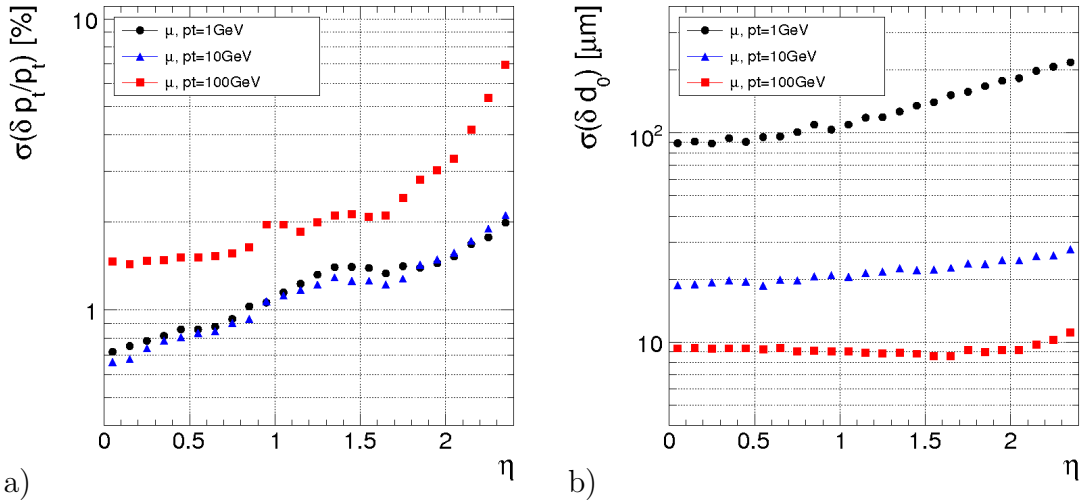


Figure 6.3: a) shows the resolution of transverse momentum, b) the resolution of the transverse impact parameter for muons with a transverse momentum of 1, 10 and 100 GeV, respectively. The tracks have been reconstructed without the muon system. [72]

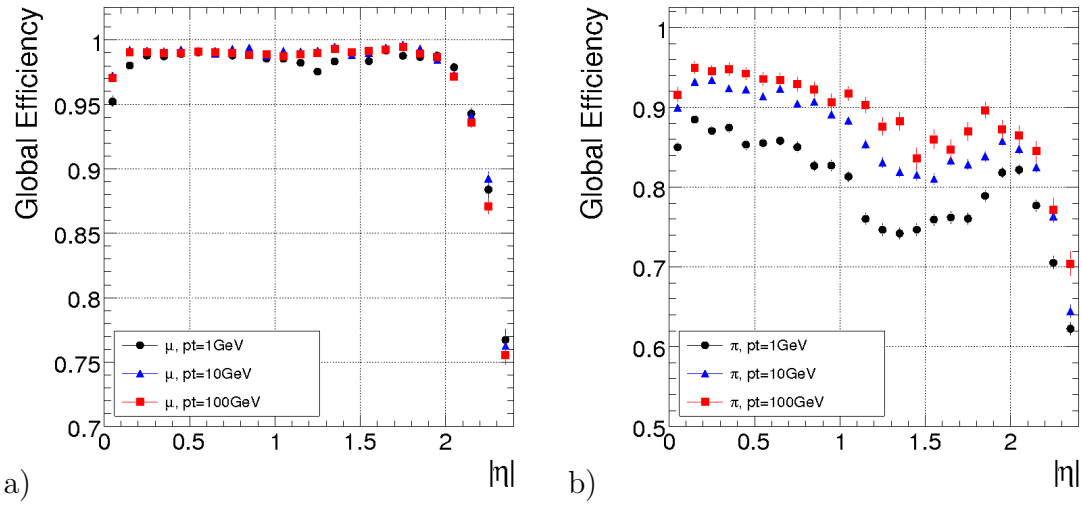


Figure 6.4: Global track reconstruction efficiency for a) muons and b) pions with a transverse momentum of 1, 10 and 100 GeV, respectively. [72]

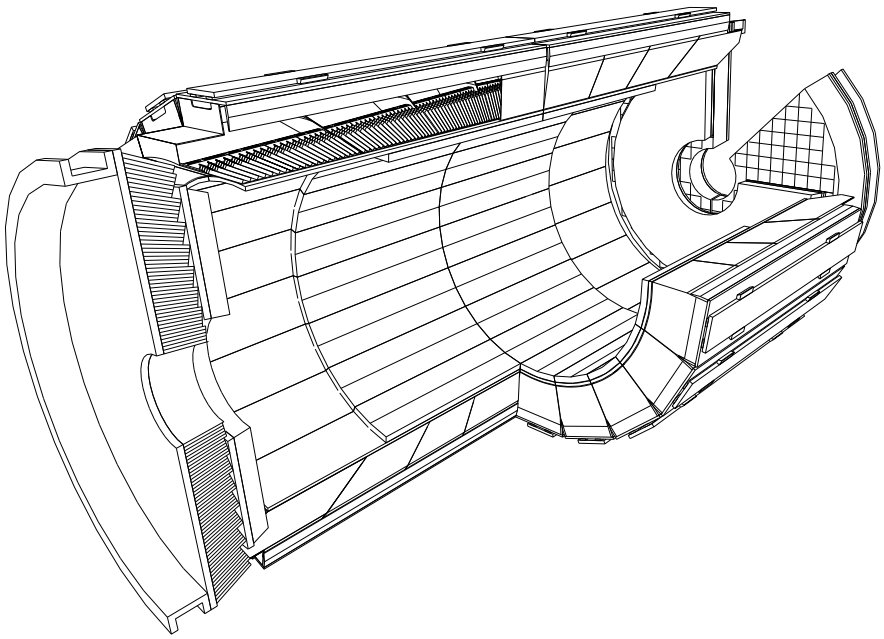


Figure 6.5: Schematic view of the electromagnetic calorimeter. [77]

aimed. Since the tracker is operated at $-10°$C, the electromagnetic calorimeter and the tracker are thermally decoupled.

In the region $1.65 < |\eta| < 2.6$ an additional preshower device is mounted, built as a two-layer sampling calorimeter providing three radiation lengths. Its tasks are a better pion identification and an improved distinction between electrons and *minimum ionizing particles*.

The overall energy resolution of the electromagnetic calorimeter for electromagnetic showers is given by

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \ , \tag{6.1}$$

providing a very accurate measurement of electromagnetic objects. A detailed description of this detector component can be found in [77].

6.4 Hadronic Calorimeter

The hadronic calorimeter measures the produced hadrons, by this providing the major input for the reconstruction of jets and missing transverse energy (section 7.3). The latter quantity defines a handle on particles invisible to the detector such as neutrinos or weakly interacting exotic particles. As the LHC is a hadron–hadron accelerator most of the particles in the final state will be of hadronic nature.

The arrangement of the hadronic calorimeter can be seen in figure 6.6. It is divided in *barrel* (HB), *endcap* (HE), *outer part* (HO) and *forward part* (HF).

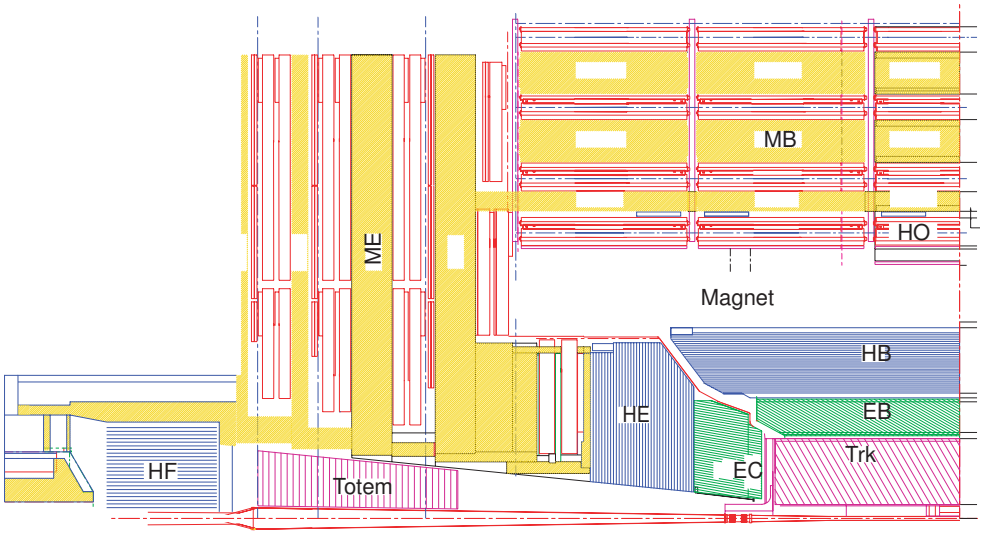


Figure 6.6: A quadrant of the CMS detector with labeled detector components. The arrangement of the subsystems of the hadronic calorimeter is shown in blue. [72]

The segments HB, HE and HO are constructed as sampling calorimeters of brass absorber plates and plastic scintillators. The scintillation light is collected by wavelength shifting fibres and finally read out via photodiodes. HB and HE both consist of 16 layers of active material. For the HO the magnet coil acts as additional absorber increasing the number of hadronic *interaction lengths* (λ_I). Particles traversing barrel and outer part pass at least $11.8 \cdot \lambda_I$ while the endcaps account for approximately $10 \cdot \lambda_I$. It is worth noting that about $1.1 \cdot \lambda_I$ are provided by the electromagnetic calorimeter. The segmentation for HB and HO is $\Delta\eta \times \Delta\Phi = 0.087 \times 0.087$ whereas this number increases for the endcap up to 0.17×0.17 at $|\eta| = 3$ successively.

The forward calorimeter, covering the range from $|\eta| = 3$ to $|\eta| = 5.2$, operates in a region where a large amount of energy is deposited. Thus its design is guided by the necessity to survive in this harsh environment. It is built of a steel absorber with quartz fibres as active medium. Charged shower particles emit Cherenkov light that is detected by photo-multipliers. The granularity of the HF is $\Delta\eta \times \Delta\Phi = 0.175 \times 0.175$.

Combining the power of electromagnetic and hadronic calorimeter leads to a resolution for pions [72] of

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{120\%}{\sqrt{E}}\right)^2 + (6.9\%)^2 \ . \tag{6.2}$$

Compared to the electromagnetic calorimeter both granularity and energy resolution are worse for the hadronic one. Therefore the precision of combined calorimeter output is limited by the hadronic calorimeter. More information about the hadronic calorimeter is given in [78].

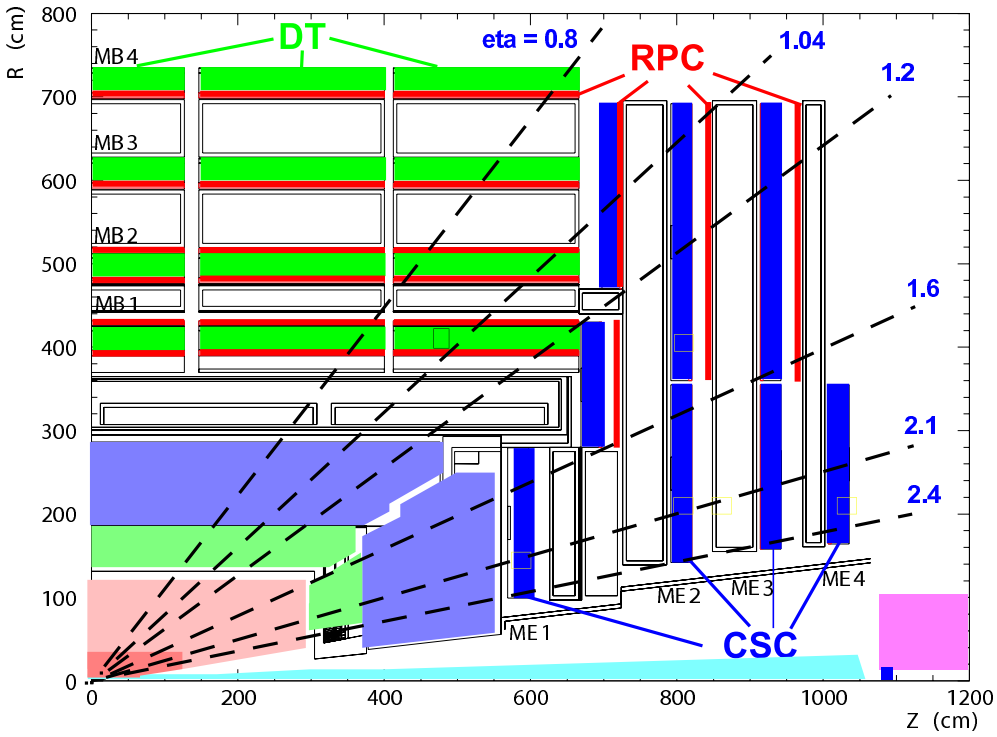


Figure 6.7: A quadrant of the CMS detector showing the placement of the different muon detectors. [72]

6.5 Muon System

Muons are less affected by radiation loss than electrons and deposit only a small fraction of their energy inside the calorimeters. They are the only particles visible to the detector that are not fully showered in the calorimeters. Hence special detectors for the measurement of muons are placed outside the calorimeters. Thereby muons can be identified with high efficiency and are very well suited for triggering and high-precision measurements.

The muon chambers are installed inside the iron yoke that returns the magnetic field. Inside the muon chambers the magnetic field has the opposite direction with respect to the tracker. The iron return yoke additionally serves for the rejection of hadrons.

Three different types of gaseous detectors are used for the muon system. In the barrel region ($|\eta| < 1.2$) *Drift Tube* (DT) chambers are mounted, arranged concentrically in four stations around the beam pipe. Each station consists of two to three so-called super layers which are built each of four layers of rectangular drift cells, staggered by half a cell to avoid insensitive regions. There are cells in which the wires are arranged along the beam line and and those with wires orthogonal to it, leading to precise measurements of all coordinates. In the worst case a muon might pass only two of the four stations, due to dead zones in between the chambers.

In the endcap a strong, non-uniform magnetic field and a high background contamination lead to the choice of *Cathode Strip Chambers* (CSC). These are multi-wire proportional chambers consisting of six anode wire planes among seven cathode panels. The wires' orientation is azimuthally so they define the tracks' radial component. Three respectively four stations cover the region from $0.9 < |\eta| < 2.4$. The arrangement of DTs and CSCs is shown in figure 6.7.

Due to uncertainties in background rates and to assure the measurement of the correct beam crossing *Resistive Plate Chambers* (RPC) are used as additional detectors in barrel and endcap region. These are gaseous parallel-plate detectors which have a very good time resolution, by this being extremely important for the trigger system. In addition they are well suited for the association of muon tracks to the corresponding bunch crossing and serve for resolving ambiguities in the muon track finding. Six layers of RPCs are mounted in the barrel region, whereas only three stations are placed in the endcap, covering a region up to $|\eta| = 1.6$. An extension up to $|\eta| = 2.1$ and an increase to four stations is foreseen in the near future. A complete description of the muon system is given in [79].

6.6 Trigger

The LHC will be operated with a bunch crossing rate of 40 MHz while the number of events per second that can be stored is limited to the order of 100. Thus a selection of the events to be recorded is necessary. The *trigger* system at CMS is arranged in two separate steps: the *Level-1* trigger and the *High Level Trigger* (HLT). The Level-1 trigger is implemented via programmable electronics and works with coarsely segmented information from calorimeters and muon chambers. All information from a collision is stored in pipelines in the front-end electronics of the sub-detectors. Due to the pipeline size a trigger decision has to be taken within 3.2 μs, so information of interest can be flagged for read-out. Events that pass the Level-1 trigger are completely read out and processed further by the HLT.

The calorimeter trigger uses combined input from both electromagnetic and hadronic calorimeter to construct electron and photon candidates and to set τ-flags. A special cluster algorithm is run to build jets and additional information concerning minimum ionizing particle signatures (compare section 7.1) and isolation flags are sent to the muon trigger. The muon trigger system combines information of all three muon detector types to construct muon candidates from hit patterns, already making use of the calorimeter information passed to it. The global Level-1 trigger combines the information from the two subsystems and the decision whether the event is discarded or a read out is taken. The output rate of the Level-1 trigger is of the order of about 100 kHz.

The HLT has to reduce the rate by another factor of 1,000. A computer farm with about 2,500 CPUs is used to run a special version of the CMS software to calculate more complicate entities to decide which events are of interest. With

growing knowledge and understanding of the detector the algorithms employed here will develop over time. The chosen events are stored to disk for further analysis. Details of the trigger and data acquisition chain are given in [80] and [81].

Chapter 7

Reconstruction of Physics Objects

The signal samples of interest are semi-muonic $t\bar{t}$ events. After the decay one muon, one neutrino and four quarks are present, in most cases two b- and two light quarks. The detector objects are therefore reconstructed muons, jets including b-tags and missing transverse energy. The reconstruction of these objects with the CMS detector is described in this chapter.

7.1 Muons

In a first step only the muon system is used for the reconstruction of muons, utilizing a Kalman filter [82] and applying a vertex-constrained fit to the trajectories. The muon candidates identified in the muon system serve as input for the muon reconstruction including tracker information. The reconstruction in the tracker is done again via a Kalman filter. After all found trajectories in the tracker have been cleaned from ambiguities, a global refit of them is done including the information from the muon system. The efficiency for high-p_T muons is very high as can be seen from figure 7.1. A detailed description of muon reconstruction is given in [83].

The HLT makes use of reconstructed muons as a large fraction of the physics of interest are associated with high-p_T leptons. In the low luminosity phase an inclusive single muon trigger with a muon p_T cut of 19 GeV will result in an expected muon rate of approximately 25 Hz. [72]

To suppress muons originating from hadron decays, isolation criteria are defined. By this muons accompanied by too many additional tracks or too much calorimeter activity can be rejected. As a consequence not only muons originating from hadrons are affected but muons that are incidentally close to a jet are rejected as well. Nevertheless, isolation procedures are extremely powerful to reduce the background for semi-muonic $t\bar{t}$ decays. The method used in this analysis is based on [84]. Using muons with $p_T > 30$ GeV an isolation requirement is defined for both tracker and calorimeters. Figure 7.2a) shows the relative p_T

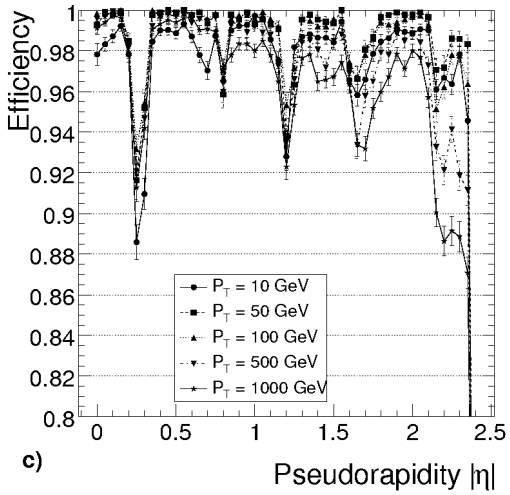


Figure 7.1: Reconstruction efficiency of muons built from tracker and muon system information for different transverse muon momenta. [72]

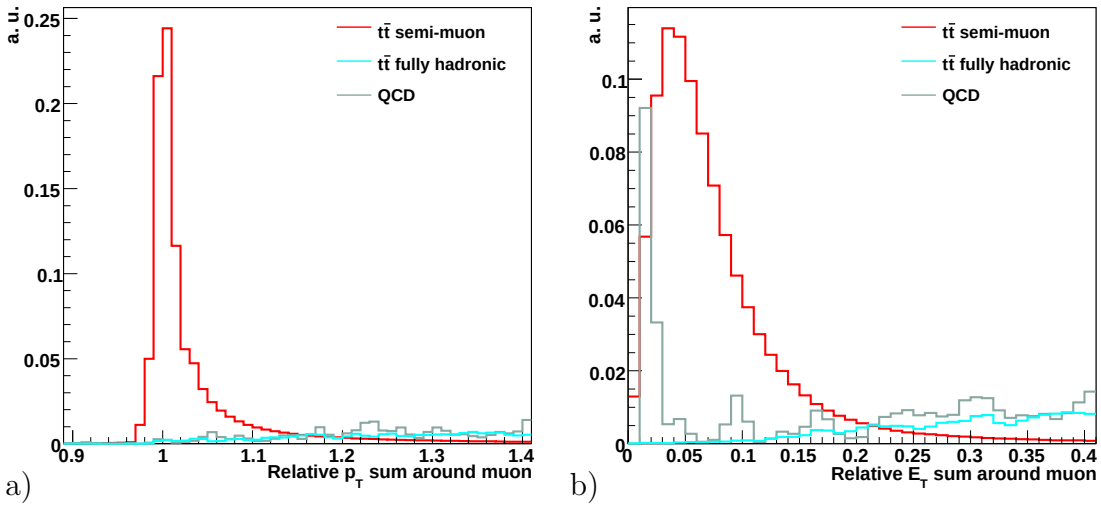
sum of tracks around the muon direction inside a cone with $\Delta R = 0.3$ for signal events, $t\bar{t}$ fully hadronic decays and QCD events. In figure 7.2b) the relative E_T sum of calorimeter towers inside a cone of $\Delta R = 0.3$ is shown for the same samples. Muons from W- and Z-decays cause distributions very similar to the semi-muonic $t\bar{t}$ case, so for those samples an isolation criterion is not well suited for rejection. According to the two distributions in figure 7.2 an isolated muon is required to fulfil:

$$0.95 \cdot p_T(\mu) \; < \sum_{\Delta R=0.3} p_T(tracks) \; < \; 1.15 \cdot p_T(\mu)\,, \tag{7.1}$$

$$0.02 \cdot p_T(\mu) \; < \sum_{\Delta R=0.3} E_T(towers) \; < \; 0.2 \cdot p_T(\mu)\,. \tag{7.2}$$

7.2 Jets

The jets are reconstructed using the information from both electromagnetic and hadronic calorimeter. Therefore the energy depositions are combined into over $4,000$ objects called *calorimeter towers*. Due to the different granularity of the two calorimeters the tower dimensions are limited by the segmentation of the hadronic calorimeter. The jets are built from these calorimeter towers. The towers are treated as massless particles whose direction is defined by the interaction point



a) b)

Figure 7.2: a) shows the p_T sum of all tracks inside a cone of $\Delta R = 0.3$ around the muon direction, divided by the p_T of the reconstructed muon. b) shows the E_T sum from the calorimeter towers inside a cone of $\Delta R = 0.3$ around the reconstructed muon direction divided by the muon p_T.

and the tower centre. The algorithms, described in section 3.4, are the same as for the generator case.

7.2.1 Jet Calibration

A perfect jet would be a projection onto a parton originating from the hard process. However, final state radiation may lead to a loss of particles that are not collected by the jet algorithm whereas initial state radiation, underlying event and pileup provide additional particles that may enter the jet. The hadronization process causes an additional smearing of the particles, worsening the jet description further. Besides the physics issues, detector effects play an important role as resolution, calibration and noise of calorimeter cells lead to non-linear effects in the jets' energy scale. To reconstruct the jet direction as precisely as possible, the E-scheme is chosen as recombination scheme (compare section 3.4.1). While a jet's direction can be reconstructed reasonably well, a correction to its energy is needed.

Several proposed procedures for jet calibration are discussed in [72]. Once the CMS experiment is running, data driven jet calibrations will be applied. In this analysis a Monte Carlo driven method is used. Reconstructed jets are calibrated by a comparison to generator jets constructed from the hadronic final state. Each jet that has been corrected for detector effects can be corrected for influences from physics in a second step, calibrating the jet to the original parton.

The calibration method used here follows the one developed in [84]. For this analysis jets are corrected to generator level using signal events. The considered

raw jets have to fulfil $p_T > 10\,\text{GeV}$ and $|\eta| < 2.4$, i. e. they have to be inside the region covered by the tracker. Jets are built from calorimeter towers on detector level and from the hadronic final state on generator level. Each b-quark from a top decay and each quark originating from a subsequent W-decay are matched to generator jets applying a criterion of $\Delta R < 0.1$. Afterwards the successfully matched jets are matched to the jets from calorimeter towers. As an example the transverse momentum ratio for matched KT4-b-jets is shown in figure 7.3. A dependence on the transverse momentum as well as on the detector region is found, illustrating the transition from calorimeter barrel to endcap. Separate correction factors are determined for jets associated with b-quarks and quarks from W-decays, respectively. An *unbinned likelihood fit* using the fitting program *Minuit* [85] inside the *ROOT* framework [86] is used for the determination of a p_T calibration factor. The fit is performed in p_T and η simultaneously. The fitted functionals for the b-jets are:

$$\frac{22.5}{p_T} + \begin{cases} 6.9 \cdot 10^{-2} \cdot |\eta|^2 + 1.1 & \text{, if } |\eta| < 1.4 \\ -0.4 \cdot \eta + 1.7 & \text{, if } 1.4 \leq |\eta| < 2.4\,. \end{cases} \tag{7.3}$$

For the jets associated with the W-decay the fit result is:

$$\frac{15.4}{p_T} + \begin{cases} 6.4 \cdot 10^{-2} \cdot |\eta|^2 + 1.1 & \text{, if } |\eta| < 1.4 \\ -0.3 \cdot \eta + 1.7 & \text{, if } 1.4 \leq |\eta| < 2.4\,. \end{cases} \tag{7.4}$$

To gain a generator-independent fit, all four signal samples, which are discussed in detail in section 8.1, are considered in this fit. Figure 7.4 shows the closure test for KT4-b-jets, indicating a reasonable correction of the transverse momentum. This procedure has been iterated with all used samples, showing no generator dependence. However, a comparison with the original parton shows non-negligible differences between the different samples. As not even the direction of the additional correction is equal, the application of a factor reflecting the physics effects is not undertaken. Hence the jets are corrected only for detector effects. In [44] the accuracy of the data driven calibration is estimated to be of order 3 – 5%. The procedure applied here yields an accuracy that is within ten percent for the important p_T regions.

7.2.2 B-tagging

Almost 100% of all top quarks decay into a W-boson and a b-quark. By tagging jets originating from b-quarks, background rejection and reconstruction efficiency can be enhanced. *B-tagging* is possible as B-hadrons have a relatively long lifetime of $\tau \approx 1.6\,\text{ps}$ corresponding to a decay length of $c \cdot \tau \approx 500\,\mu\text{m}$. Thus jets initiated by b-quarks may originate from a *secondary vertex* that can be separated from the primary vertex, i. e. the vertex associated with the hard parton–parton scattering. An additional handle can be constructed from the fact that about 20% of all B-hadrons decay leptonically, leading to leptons inside b-jets. The third possibility

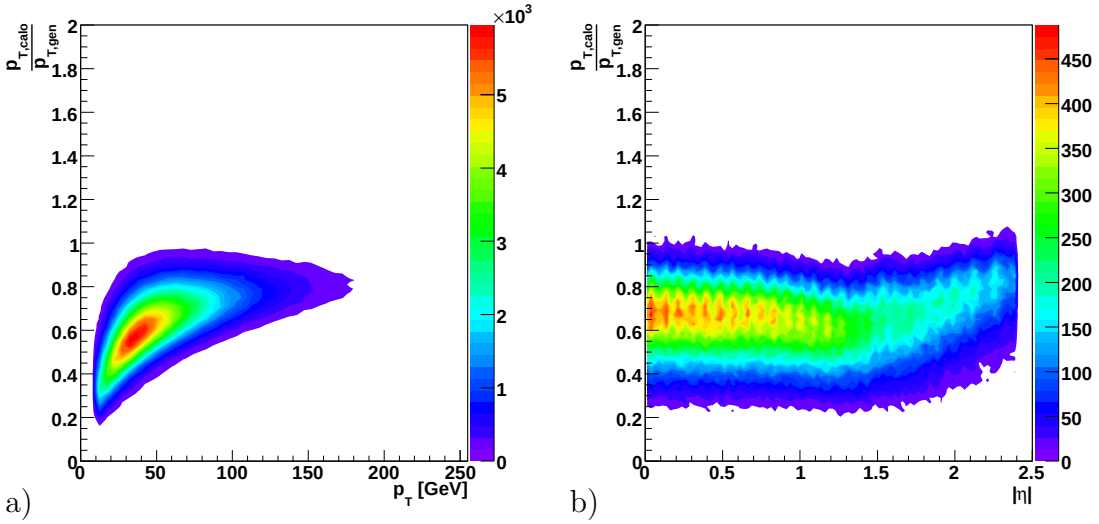


Figure 7.3: Transverse momentum ratio of calorimeter and generator jets over p_T and η of the calorimeter jet, respectively.

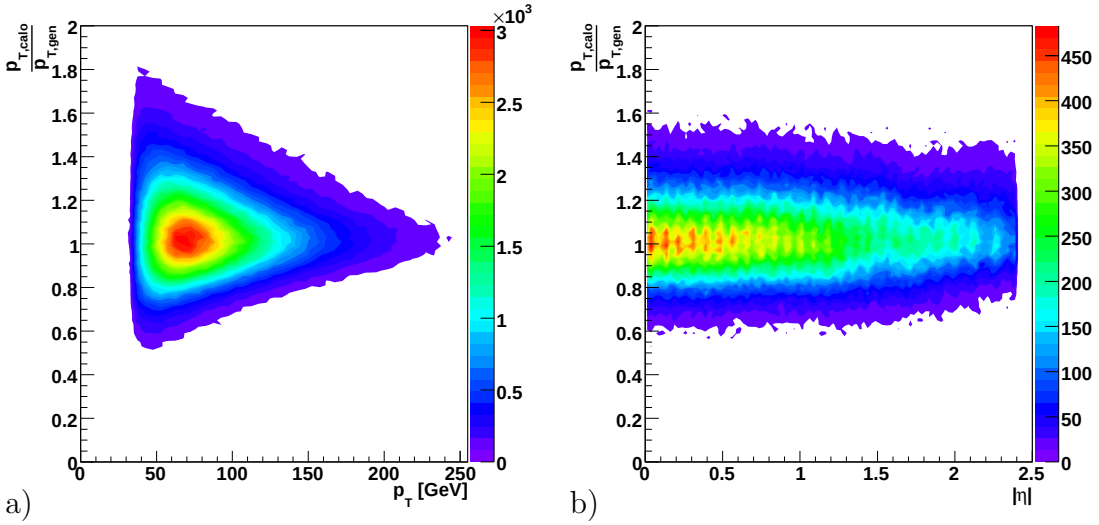


Figure 7.4: Transverse momentum ratio of calorimeter and generator jets over p_T and η of the calorimeter jet, respectively.

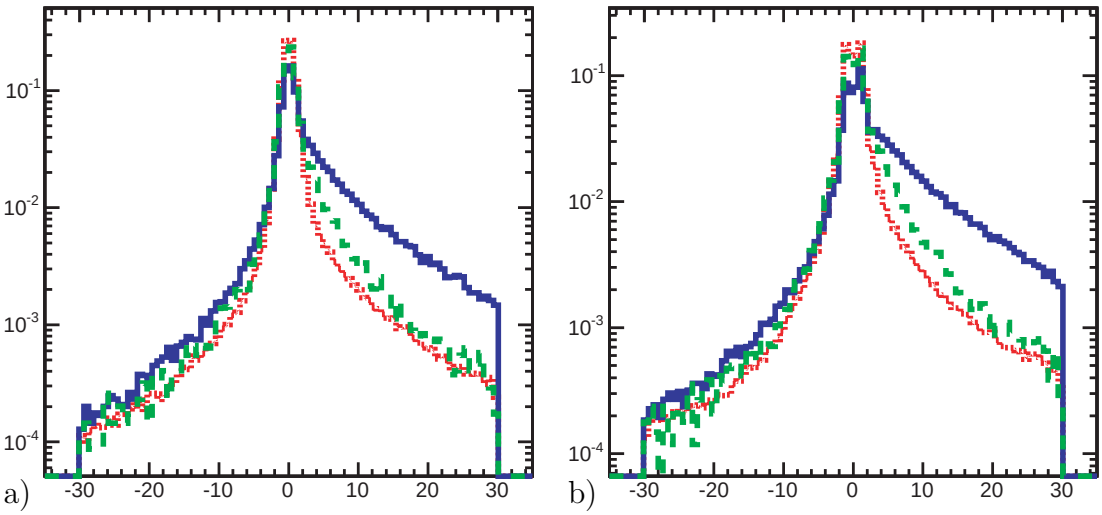


Figure 7.5: a) Transverse and b) three dimensional impact parameter distribution for tracks belonging to b- (blue), c- (green) or light jets (red). [87]

to distinguish b-jets from others is based on the high charged track multiplicity associated with B-decays. On average five charged tracks originate from the decay of a B-hadron. [72]

The b-tagging algorithm applied here makes use of the charged tracks coming from B-hadron decays. For each jet all tracks inside a cone of $\Delta R = 0.3$ around the jet axis are determined and their transverse and three dimensional impact parameters are computed. The impact parameter distributions of tracks originating from the primary vertex follow a Gaussian centred at zero. For tracks originating from secondary vertices this distribution is shifted. Figure 7.5 shows the transverse impact parameter values for tracks belonging to jets of different flavours. As b-jet tracks do often not originate from the primary vertex higher impact parameters are encountered more often leading to an asymmetric shape. From this distribution it is possible to define a probability for a track to come from the primary vertex. Combining these impact parameter driven probabilities for all high-quality tracks results in one final discriminator value for the jet, reflecting its tracks' compatibility with the primary vertex. Figure 7.6 shows the resulting discriminator distribution for different flavoured jets. A low discriminator value corresponds to a high compatibility with the primary vertex. An in-depth discussion of this algorithm and its performance can be found in [87].

7.3 Missing Transverse Energy

Neutrinos (and maybe some exotic particles) are invisible to the detector. As the actual parton–parton collision's energy is unknown, it is not possible to reconstruct the missing energy component along the z-axis. In the transverse region each event is supposed to be balanced, thus a *missing transverse energy* $\not{E}_T$ due

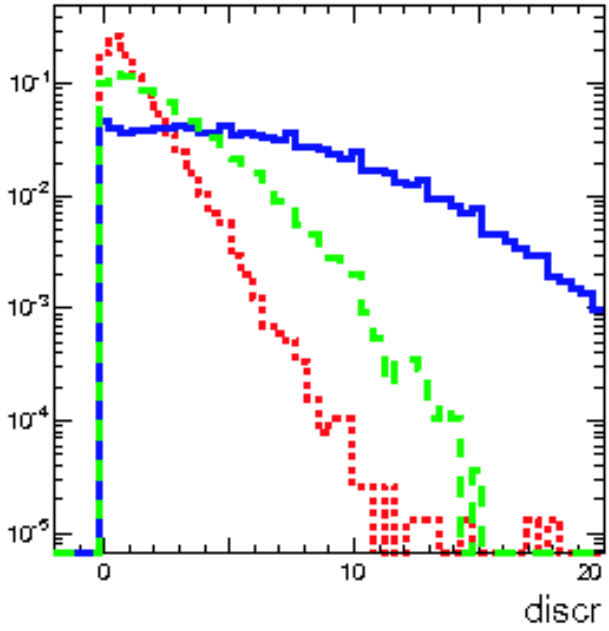


Figure 7.6: Discriminator calculated of the probabilities of the jets' tracks to origin from the primary vertex for b- (blue), c- (green) or light jets (red). [87]

to undetected particles can be determined. The measurement of this quantity is quite difficult as the detector is not hermetic and particles with $|\eta| > 5$ are not taken into account for its computation. In addition the underlying event of a collision and pile-up events spoil the measurement.

The missing energy is computed from the calorimeter cells for x- and y-direction separately and finally $\not\!\!E_T$ is calculated. For events with high $\not\!\!E_T$, often due to one single high-energetic, undetected particle, its direction, i. e. the ϕ component, can be determined quite accurately. The quality of the direction measurement is reduced for low values of $\not\!\!E_T$. The calorimeter cell noise and their non-linear response introduce a huge uncertainty to this observable, especially in case the "true" value is low. In spite of all the difficulties it provides a reasonable starting point to estimate the unmeasured neutrino p_T in $t\bar{t}$ events, so it is used within this analysis. The different corrections presented in [88] did not improve the $\not\!\!E_T$ performance in this analysis, so none of them is applied.

Chapter 8

Selection and Event Reconstruction

The semi-muonic decay channel will be analyzed, thus all processes leading to one muon and at least four jets have to be considered as possible background as they might result in a signal-like signature. The backgrounds taken into account are leptonically decaying W- and Z-bosons accompanied by jets and QCD events containing leptons. It is expected that other $t\bar{t}$ decay modes than the semi-muonic one contribute the largest amount of background.

The semi-muonic $t\bar{t}$ decay is chosen for this analysis because of its clear signature. Compared to the other $t\bar{t}$ decay channels this one is expected to allow the highest reconstruction efficiency and background suppression.

In this chapter the applied event selection procedure is described. To avoid any generator-dependent bias a robust cut based selection is performed using only fundamental variables. The selection and reconstruction procedure applied is based on the studies presented in [84].

8.1 Data Samples

As no measured data is available at the time of writing, a feasibility study is presented. For this the full CMS detector simulation and the CMS analysis framework *CMSSW* [72, 89] are used. The focus of this work is the investigation of additional radiation in $t\bar{t}$ events. Therefore four different generators were used for signal simulation: Herwig6, MC@NLO, Pythia6 and Pythia8.

Herwig6 and MC@NLO were run with default settings applying tuning B for the Jimmy underlying event generation. Pythia6 was run in default mode as well, which implies the application of the Q^2 ordered power shower option and tune A with a slightly different energy scaling for the soft underlying event. For Pythia8 no tunes exist so far, but it behaves very similar to tune S0 utilizing a p_T^2 ordered power shower. So the four samples include two different power showers, a matched shower and a standard shower, as well as angular, Q^2 and p_T^2 ordering. In addition the two Pythia versions cover the two most extreme

behaviours of the Sudakov form factor obvious in the shape of the $t\bar{t}$ transverse momentum distribution at small values.

According to the NLO cross section from [30], the amount of signal data produced for each generator corresponds to an integrated luminosity of $7\,\mathrm{fb}^{-1}$. Background samples of interest include other $t\bar{t}$ decay modes, leptonically decaying W- and Z-bosons accompanied by additional jets and QCD multi-jet events. The background samples used were produced during the CSA07 challenge [90, 91]. The production was done using ALPGEN for $t\bar{t}$ and W/Z+jets samples and PYTHIA6 for QCD. As the production was done only for $1\,\mathrm{fb}^{-1}$ with a LO cross section, all background events gain an additional weight for scaling the total number of events to an NLO $t\bar{t}$ cross section for $7\,\mathrm{fb}^{-1}$.

8.2 Preselection

The goal of the preselection step is a reasonable reduction of the background before the computing time consuming reconstruction step is run. Furthermore the existence of all ingredients needed for the reconstruction is required. The semi-leptonic muon channel investigated here consists of one muon, one neutrino and four quarks in the final state. Due to the almost exclusive decay $t \to Wb$ two b-quarks are expected in the final state. Analyzing the decay products on detector level, one expects to observe one muon, two b-jets, two non-b-jets and $\not\!\!E_T$ due to the neutrino. As mentioned in section 7.3, $\not\!\!E_T$ suffers from different uncertainties so it is not used for the preselection. For the presented analysis a KT4 jet algorithm is applied. The choice of this particular jet algorithm is motivated in sections 9.1.1 and 9.1.3.

The applied cut values are motivated by N-1 distributions, i. e. the distributions after all other cuts. Figure 8.1 shows the N-1 distributions for the four different signal samples.

For the used background samples soft prefiltering cuts were applied to reduce the amount of data on the storage system. This changes the distributions of the cut variables, but it does not influence the final cut efficiencies. All selection cuts are chosen much tighter than the prefiltering criteria. Figure 8.2 shows the N-1 distributions for signal and background samples. MC@NLO is used as signal for the comparison, where the same prefiltering cuts have been applied as to the background, leading to a slightly different distribution compared to 8.1. The cuts are chosen based on the presented distributions.

8.2.1 Muon Selection

Each event is required to contain exactly one muon with $p_T > 30\,\mathrm{GeV}$ that fulfils the isolation requirement defined in equations 7.1 and 7.2. The distribution 8.1a) shows an almost identical behaviour for all signal samples. Especially the fact that the biggest difference occurs among HERWIG and MC@NLO shows the independence of this cut to the underlying event model as the same tuning is

used for both samples. The two PYTHIA samples with different underlying event models are located between HERWIG and MC@NLO. Di-leptonic $t\bar{t}$, W and Z events with one muon in the detector acceptance region are not affected by this cut whereas the number of all other event types is reduced significantly, which can be seen from figure 8.2a).

8.2.2 B-tag Selection

Due to the large branching ratio $t \to Wb$ two b-jets are expected in most $t\bar{t}$ events. Figure 8.1b) shows the distribution of the highest b-tag for the different signal samples. A slight difference may be noted between the two PYTHIA samples and HERWIG and MC@NLO. This might be caused by the different fragmentation functions used. In 8.2b) the $t\bar{t}$ backgrounds are omitted as no distinction between the different $t\bar{t}$ decay channels can be made from b-tags. Asking for the highest b-tag in an event to exceed 0.5 reduces the QCD, W and Z background significantly.

As two b-jets are needed for the reconstruction step, the existence of a second b-tag above zero is required. This requirement does not induce any additional distinction power and rejects only very few events from both signal and background.

8.2.3 Jet Multiplicity Selection

Figures 8.1c) and 8.2c) show the number of jets above 25 GeV, figures 8.1d) and 8.2d) the number of jets above 35 GeV, both for $|\eta| < 2.4$. In this case the respective other jet cut has not been applied as the jet selection is treated in one step. The signal plots show the expected difference for the jet multiplicity, leading to more high-p_T jets per event for PYTHIA8 and to more events with only few high-p_T jets for PYTHIA6. From figures 8.2c) and d) it can be seen that the jet cuts reduce the amount of di-leptonic $t\bar{t}$, QCD and W and Z events. As the jet multiplicity cut should be independent of the other cuts applied, one might expect the same jet multiplicity distribution for the signal and the $t\bar{t} \to e/\tau +$ jets case. If both samples were generated with the same generator settings, this would be true. Here the difference of MC@NLO and ALPGEN + PYTHIA6 can be seen.

Three jets with $p_T > 35$ GeV and a fourth jet with $p_T > 25$ GeV are requested for an event to be selected.

8.2.4 Preselection Summary

The applied preselection cuts discussed in sections 8.2.1 to 8.2.3 are:

- exactly one isolated muon with $p_T > 30$ GeV,
- highest b-tag above 0.5, second b-tag is present,
- at least three jets with $p_T > 35$ GeV in $|\eta| < 2.4$,
- at least four jets with $p_T > 25$ GeV in $|\eta| < 2.4$.

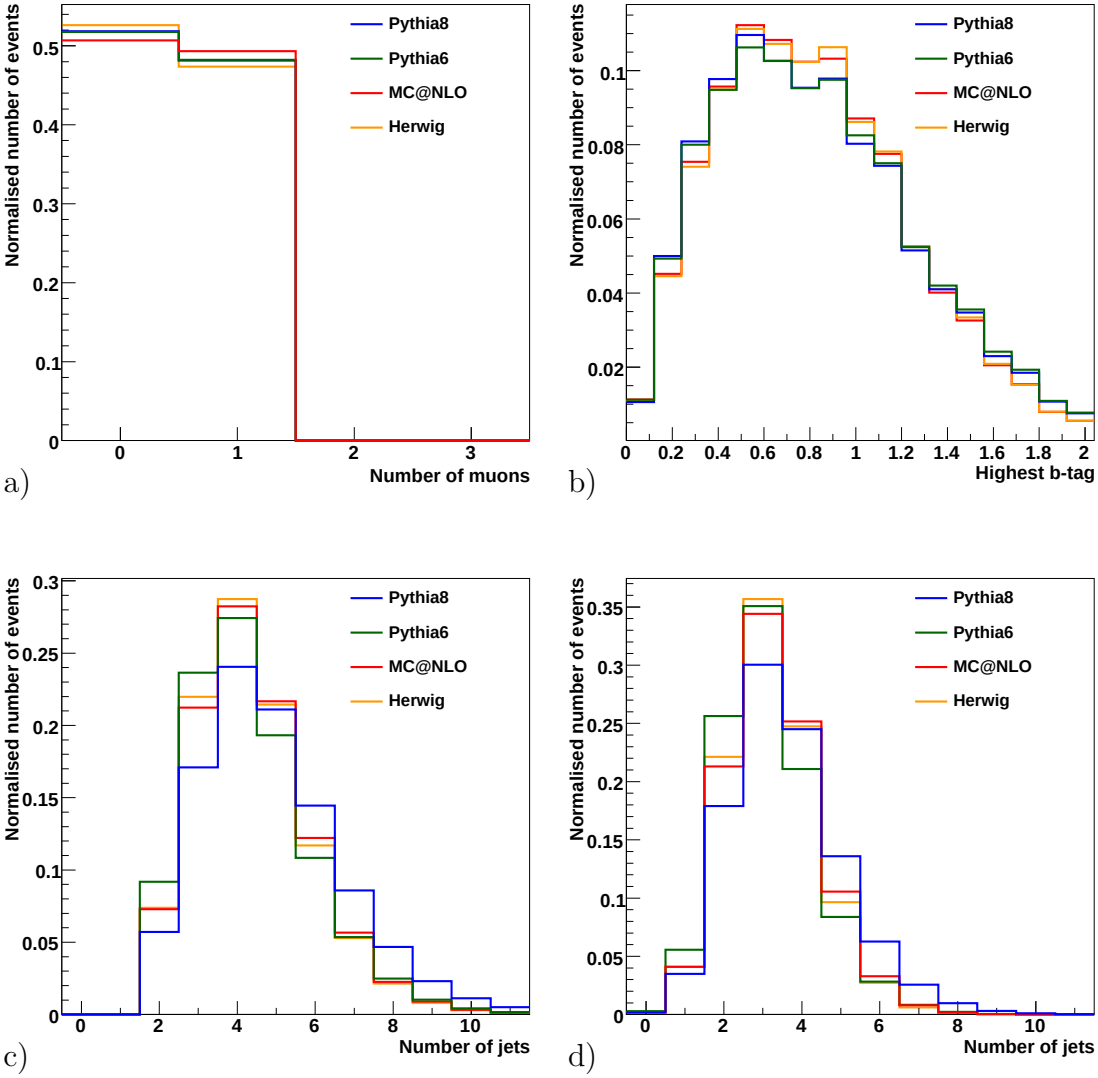


Figure 8.1: N-1 plots for the applied preselection cuts comparing the four different signal samples. a) shows the number of isolated muons, b) the highest b-tag, c) the number of jets with $p_T > 25\,\text{GeV}$ and d) the number of jets with $p_T > 35\,\text{GeV}$.

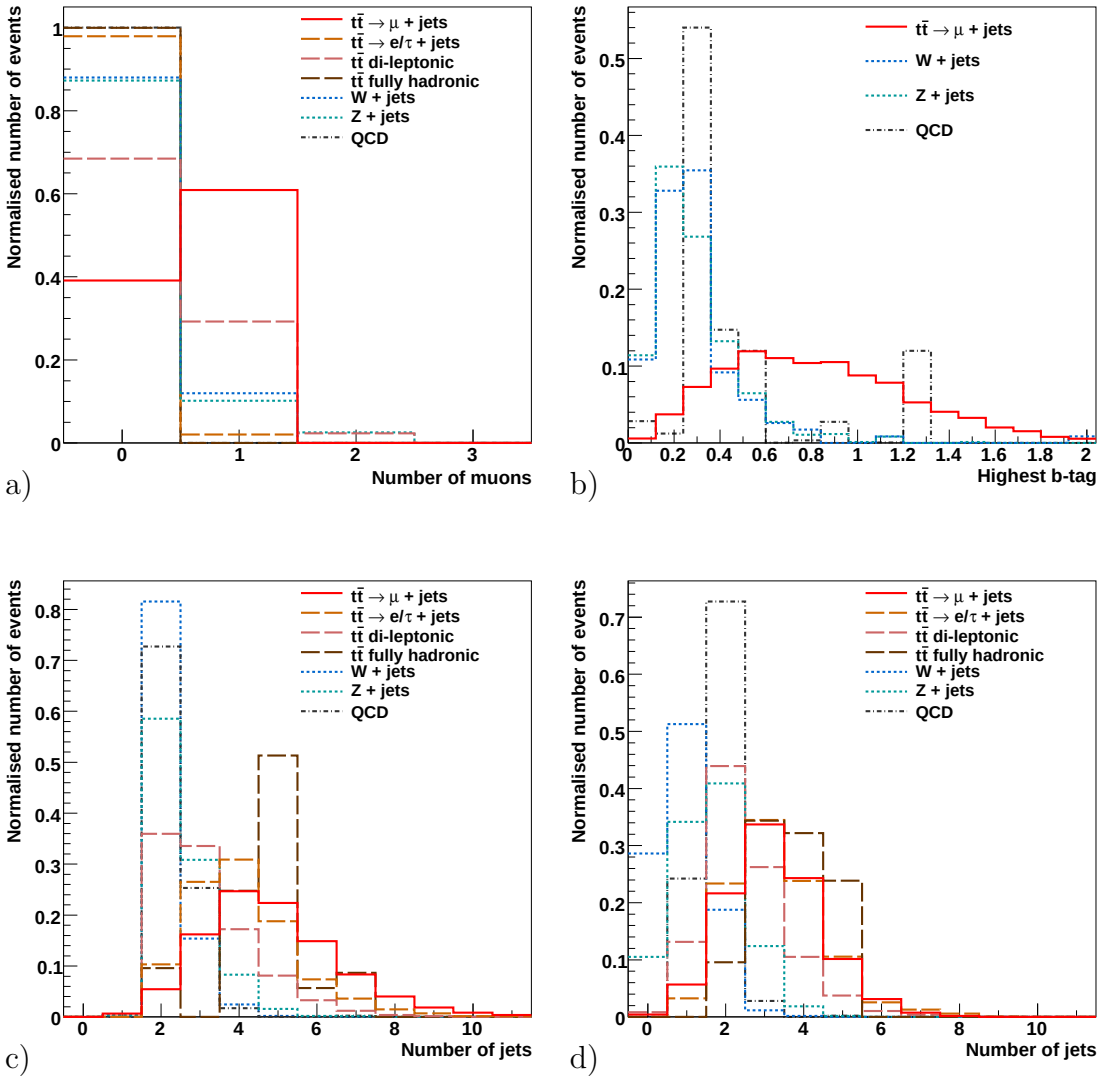


Figure 8.2: N-1 plots for the applied preselection cuts comparing signal with various backgrounds. a) shows the number of isolated muons, b) the highest b-tag, c) the number of jets with $p_T > 25\,\text{GeV}$ and d) the number of jets with $p_T > 35\,\text{GeV}$.

Table 8.1: The number of initial and preselected events and the preselection efficiency are given for the different backgrounds.

Sample	Initial Events	Preselected Events	Efficiency
$t\bar{t}$ di-leptonic	$7.0 \cdot 10^5$	23177	$3.3 \cdot 10^{-2}$
$t\bar{t}$ fully hadronic	$2.8 \cdot 10^6$	152	$5.4 \cdot 10^{-5}$
$t\bar{t} \rightarrow e/\tau +$ jets	$1.88 \cdot 10^6$	8499	$4.5 \cdot 10^{-3}$
$W +$ jets	$8.1 \cdot 10^8$	18416	$2.3 \cdot 10^{-5}$
$Z +$ jets	$8.1 \cdot 10^7$	1152	$1.4 \cdot 10^{-5}$
QCD	$7.7 \cdot 10^{11}$	2812	$3.7 \cdot 10^{-9}$

Table 8.2: The number of initial and preselected events are given for the different signal samples. Furthermore preselection efficiency and yielded signal-to-background ratio for the preselection step are presented.

Generator	Initial Events	Preselected Events	Efficiency	S/B Ratio
HERWIG	941630	186044	0.197	3.4
MC@NLO	941630	192115	0.200	3.5
PYTHIA6	941630	169061	0.180	3.1
PYTHIA8	941630	215031	0.228	4.0

Table 8.1 shows the reduction of the different backgrounds achieved with these cuts. The total number of background events is reduced from 10^{11} to about $5 \cdot 10^4$. The efficiency of the preselection for the signal samples and the yielded signal-to-background ratio are summarised in table 8.2. The different jet multiplicities in the signal samples have a huge impact on the preselection efficiency, leading to a signal-to-background ratio in a range from 3.1 to 4. Especially the low additional radiation amount of the PYTHIA6 tune reduces the number of jets and hence the number of events passing the jet cuts. The comparatively soft underlying event scaling for this generator enhances this effect.

8.3 Reconstruction

For the reconstruction of the top quarks a *least square fit* with external constraints is applied using the *TKinFitter* [92]. All measured quantities are varied according to their precision and candidates for the top and the anti-top are constructed. The accuracy of the measurements are given by the experimental resolutions. First these resolutions needed for the kinematic fit are determined and then the general procedure of the fit is explained.

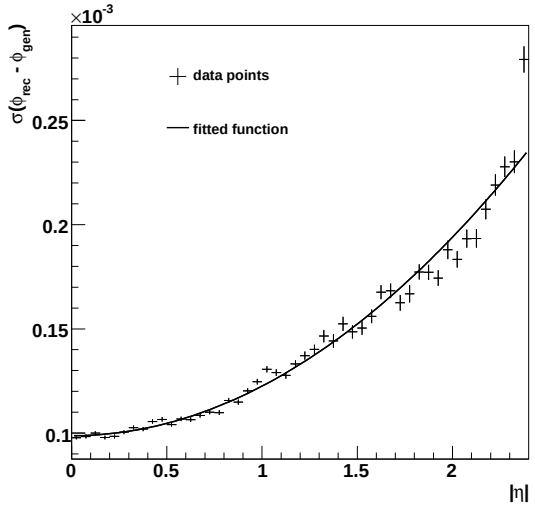


Figure 8.3: Resolution in the azimuthal angle ϕ of muons as a function of $|\eta|$. The fit has been performed with ROOT.

8.3.1 Object Resolutions

The resolutions of all objects utilised in the construction are needed as input for the kinematic fit. The resolutions are determined by a comparison of reconstructed objects and generator objects. Later it will be possible to quantify the resolutions from experimental data as discussed in [72]. The objects' parameterization is done in p_T, η and ϕ.

The reconstructed objects are matched to generator objects via a ΔR criterion. In most cases only one dominant dependence on the objects' variables is found and a binned χ^2 fit with ROOT [86] can be performed. Only for the muon p_T dependences on both p_T and η are observed and an unbinned likelihood fit with Minuit [85] is done. Like for the jet calibration the generator dependence of the resolutions was studied and it turned out to be negligible. As an example of a fitted resolution function the ϕ resolution of muons is shown in figure 8.3.

The muon resolutions are:

$$\sigma_{p_T}(p_T,\eta)[\mathrm{GeV}] = 1.7\cdot 10^{-4}\cdot p_T^2[\mathrm{GeV}^2] + 5.1\cdot 10^{-1}\cdot|\eta| + 1.1\cdot 10^{-4}\,, \tag{8.1}$$

$$\sigma_{\eta}(\eta) = 6.8\cdot 10^{-5}\cdot|\eta|^2 - 2.1\cdot 10^{-4}\cdot|\eta| + 4.2\cdot 10^{-4}\,, \tag{8.2}$$

$$\sigma_{\phi}(\eta) = 2.4\cdot 10^{-5}\cdot|\eta|^2 + 9.9\cdot 10^{-5}\,. \tag{8.3}$$

The $\not\!\!E_T$ resolutions are:

$$\sigma_{E_T}(E_T)[\mathrm{GeV}] = 0.24\cdot E_T[\mathrm{GeV}] + 9.4\,, \tag{8.4}$$

$$\sigma_{\phi}(E_T) = \frac{43.2}{E_T[\mathrm{GeV}]} + 5.0\cdot 10^{-2}\,. \tag{8.5}$$

The resolutions of b-jets with $p_T > 20\,\text{GeV}$ are:

$$\sigma_{p_T}(p_T)[\text{GeV}] = 1.7 \cdot \sqrt{p_T[\text{GeV}]} - 2.9\,, \tag{8.6}$$

$$\sigma_\eta(p_T) = \frac{0.9}{p_T[\text{GeV}]} + 2.4 \cdot 10^{-2}\,, \tag{8.7}$$

$$\sigma_\phi(p_T) = \frac{-9.7}{p_T^2[\text{GeV}^2]} + \frac{2.5}{p_T[\text{GeV}]} + 6.2 \cdot 10^{-3}\,. \tag{8.8}$$

The resolutions of jets from a W-decay with $p_T > 20\,\text{GeV}$ are:

$$\sigma_{p_T}(p_T)[\text{GeV}] = 1.8 \cdot \sqrt{p_T[\text{GeV}]} - 3.8\,, \tag{8.9}$$

$$\sigma_\eta(p_T) = \frac{0.6}{p_T[\text{GeV}]} + 2.6 \cdot 10^{-2}\,, \tag{8.10}$$

$$\sigma_\phi(p_T) = \frac{-15.2}{p_T^2[\text{GeV}^2]} + \frac{2.0}{p_T[\text{GeV}]} + 8.8 \cdot 10^{-3}\,. \tag{8.11}$$

8.3.2 Kinematic Fit

The accuracy of the measured quantities is expressed by a covariance matrix using the resolutions from the last section. During the fit corrections are calculated for each component of the particles' momenta using the covariance matrix for weighting. This means that the corrections correspond to the accuracy of the measurements. For example the momentum of the well measured muon should be shifted less than the momentum of the neutrino. The momenta are changed until the requirements given below are fulfilled. This procedure is motivated by the idea that no quantity is measured correctly, so its value can be shifted in a reasonable range. The covariance matrix is assumed to be diagonal where the resolutions presented in 8.3.1 are used for its calculation. A detailed description of the applied procedure is given in [51].

The input to the fit are the five highest p_T jets as candidates for jets originating from the W-decay, the three highest b-tagged jets as candidates for the b-jets, the isolated muon and $\not\!\!E_T$ for the neutrino. For the jets and the muon the three momenta are used while for $\not\!\!E_T$ only $\not\!\!E_x$ and $\not\!\!E_y$ are available. So for the neutrino input values can be computed for E_T and ϕ. The neutrino η is treated as unmeasured quantity.

Here the semi-muonic decay is analyzed, so the external constraints implemented are:

- invariant mass of muon and neutrino candidate has to equal the W mass,
- invariant mass of muon and neutrino candidate and one b-tagged jet has to equal the top mass,
- invariant mass of two jets has to equal the W mass,
- invariant mass of these two jets and one additional b-tagged jet has to equal the top mass.

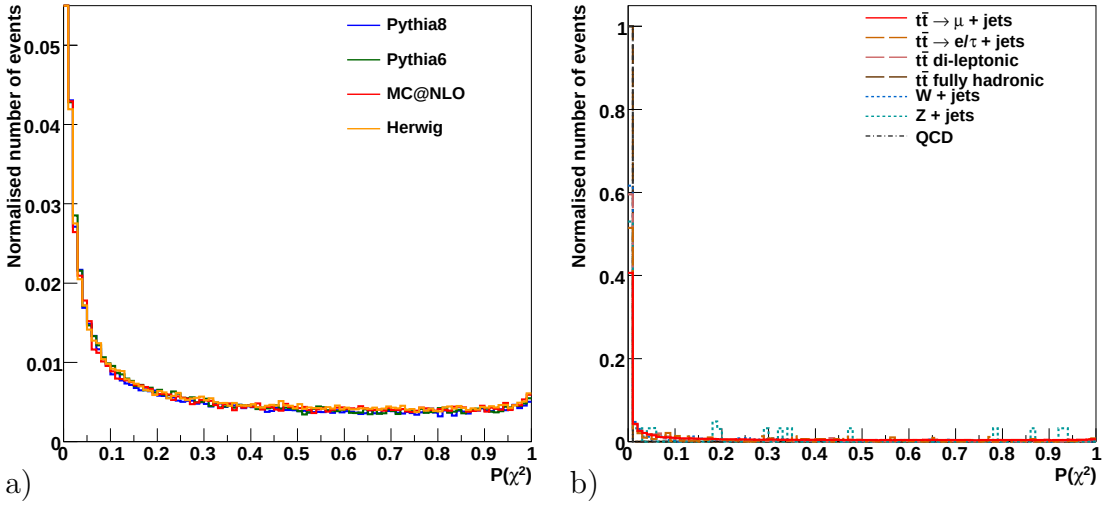


Figure 8.4: In a) the $P(\chi^2)$ distribution is shown for all signal samples. The peak in the first bin is cut to emphasise the flatness of the distribution for higher values. In b) the $P(\chi^2)$ distribution of the backgrounds is compared to the signal, showing a significant difference in the number of entries in the first bin. The first bin contains the events that are discarded by the final selection cut.

The kinematic fit computes shifted candidates for the muon, neutrino and jets. Furthermore W-boson and top quark candidates are calculated. The top and the anti-top candidate are used to construct a possible $t\bar{t}$ system.

The fit is required to converge within 100 iterations otherwise the tried combination is rejected. If more than one converged fit is found the fit with the lowest χ^2 value is chosen. The χ^2 of the fit follows a χ^2 distribution so it can be used to calculate the probability for obtaining this value or even a higher one:

$$P(\chi^2) = 1 - F(\chi^2)\,, \tag{8.12}$$

where $F(\chi^2)$ denotes the integral of the χ^2 distribution. This probability is expected to be a flat distribution between zero and one. Figure 8.4a) shows the probability distribution for the four signal samples. The ratio of entries in the first bin varies from about 39% for HERWIG to 42.5% for PYTHIA8. The fact that the distribution is flat between 0.2 and 1 justifies the selection of the event with the lowest χ^2 value. Figure 8.4b) compares the signal (MC@NLO is chosen as an example) with the different backgrounds, revealing a significant difference in the first bin. The background distributions are almost flat in the range $P(\chi^2) > 0.2$ like the signal, but most of them suffer from very low statistics. A small value for $P(\chi^2)$ is associated with a high value for χ^2 and a low quality of the fit. The huge number of entries extremely close to zero is thus considered to be not signal- but background-like.

Table 8.3: The number of initial and selected events and the selection efficiency are given for the different backgrounds.

Sample	Initial Events	Selected Events	Efficiency
$t\bar{t}$ di-leptonic	$7.0 \cdot 10^5$	4532	$6.4 \cdot 10^{-3}$
$t\bar{t}$ fully hadronic	$2.8 \cdot 10^6$	0	$< 5.0 \cdot 10^{-6}$
$t\bar{t} \to e/\tau + \text{jets}$	$1.88 \cdot 10^6$	1905	$1.0 \cdot 10^{-3}$
$W + \text{jets}$	$8.1 \cdot 10^8$	2915	$3.6 \cdot 10^{-6}$
$Z + \text{jets}$	$8.1 \cdot 10^7$	188	$2.3 \cdot 10^{-6}$
QCD	$7.7 \cdot 10^{11}$	0	$< 1.8 \cdot 10^{-11}$

8.4 Selection

The probability value from the χ^2 distribution serves as final selection cut. Since the relative number of events with values very close to zero is much higher for background samples. All events are requested to fulfil

$$P(\chi^2) > 0.01\,. \tag{8.13}$$

Table 8.3 shows the combined efficiency of preselection and this cut for the background. It should be noted that all events with no stable fit result are discarded as well. The total number of background events is reduced to 9,540 with this step. While the observation of no fully hadronic decaying $t\bar{t}$ event passing the selection is trustworthy, this is quite optimistic for QCD events. Due to computational limitations only a small fraction of the expected QCD events could be simulated, leading to high weights for the individual events. When the LHC is running, a QCD background description from data will be used.

Table 8.4 summarises the efficiency and signal-to-background ratio for the signal samples. The ratio of preselected events passing the reconstruction and final selection is lowest for PYTHIA8. The large number of additional jets allows a large fraction of events to pass the preselection cuts, but the jets caused by radiation and underlying event do in most cases not fulfil the constraints of the kinematic fit, leading to a drop in the fitting efficiency and to a higher ratio of fit probability values near zero (compare section 8.3.2).

The selection efficiency and the yielded signal-to-background ratio vary with the different signal generators and their settings. It is obvious that another choice of the generator for the background would lead to changed results, too. Thus for a reasonable determination of both signal and background efficiencies a good understanding of the physics involved and the generator properties are mandatory.

Table 8.4: The number of initial and selected events is given for the different signal samples. Furthermore selection efficiency and signal-to-background ratio for the full selection are presented.

Generator	Initial Events	Selected Events	Efficiency	S/B Ratio
Herwig	941630	57927	0.062	6.1
MC@NLO	941630	57765	0.061	6.1
Pythia6	941630	49256	0.052	5.2
Pythia8	941630	61018	0.065	6.4

Table 8.5: The number of selected events with one additional jet is given for the different signal samples and the combined background. Furthermore selection efficiency and signal-to-background ratio for the signal samples are presented.

Generator	Selected Events	Efficiency	S/B Ratio
Herwig	51929	0.055	6.9
MC@NLO	51939	0.055	7.0
Pythia6	43342	0.046	5.8
Pythia8	55655	0.058	7.4
Background	7472	–	–

8.5 Specific Selections

In the next chapter not only the reconstructed $t\bar{t}$ system but also the additional hardest jet will be investigated. Asking for an additional jet slightly reduces the number of events fulfilling all requirements. Jets from the whole detector region are taken into account, thus the pseudo-rapidity region is increased to $|\eta| \leq 5$. The additional jet is required to have an uncalibrated transverse momentum of at least 10 GeV. So the additional selection criterion is:

- at least one additional jet in $|\eta| \leq 5$ with an uncalibrated p_T above 10 GeV.

The actual numbers are given in table 8.5.

Selecting the additional hardest jet from all jets not used for the $t\bar{t}$ reconstruction introduces a huge combinatorial background. Therefore additional selection requirements are applied to improve the number of correctly chosen jets for the rapidity study presented in section 9.2.2. As shown in section 5.2.2 the dip structure in the relative rapidity distribution present in Herwig becomes more pronounced for high-p_T jets. Consequently for this analysis jets with high transverse momentum are used. The jet is required to have an uncalibrated minimum transverse momentum of 50 GeV. As shown in chapter 5 the transverse momentum of high-p_T $t\bar{t}$ pairs is governed by the hardest emission. Thus the jet and the $t\bar{t}$ system are often arranged back-to-back. Here the $t\bar{t}$ system and the additional

Table 8.6: The number of selected events fulfilling the requirements specified in the text are given for the different signal samples and the combined background. Furthermore selection efficiency and yielded signal-to-background ratio for the signal samples are presented.

Generator	Selected Events	Efficiency	S/B Ratio
HERWIG	7178	$8 \cdot 10^{-3}$	5.7
MC@NLO	7963	$8 \cdot 10^{-3}$	6.3
PYTHIA6	5762	$6 \cdot 10^{-3}$	4.6
PYTHIA8	9328	$1.0 \cdot 10^{-2}$	7.4
Background	1258	–	–

jet are required to fulfil $\pi - 0.5 < \Delta\phi_{t\bar{t}-jet} < \pi + 0.5$. Furthermore top and anti-top quark are produced back-to-back in the $t\bar{t}$ rest frame. Applying a boost due to a hard emission narrows the angle between them significantly, leading to the requirement $|\Delta\phi_{t-\bar{t}} - \pi| > 0.3$. These three additional cuts reduce the number of events significantly as summarised in table 8.6. Nonetheless, the combinatorial background decreases extremely. Summarizing, the additional selection criteria are:

- at least one additional jet in $|\eta| \leq 5$ with an uncalibrated p_T above 50 GeV,
- $t\bar{t}$ system and hard additional jet are arranged back-to-back: $\pi - 0.5 < \Delta\phi_{t\bar{t}-jet} < \pi + 0.5$,
- top and anti-top are not back-to-back, fulfilling: $|\Delta\phi_{t-\bar{t}} - \pi| > 0.3$.

These criteria serve very well for the necessities of the rapidity study but it should be noted that the applied cuts reduce the used phase space significantly.

Chapter 9

Results

The events selected and reconstructed according to chapter 8 are analysed in this chapter and the distributions obtained from the fitted $t\bar{t}$ system are presented. First the p_T spectrum of the $t\bar{t}$ pair is discussed. Afterwards the properties of the hardest additional jet are analysed, investigating its transverse momentum and its rapidity with respect to the $t\bar{t}$ system.

9.1 Transverse Momentum of the $t\bar{t}$ System

Figure 9.1 shows the p_T distributions of the $t\bar{t}$ system on generator level for the four investigated data samples. As mentioned in chapter 8 the parameters were chosen to give the most different predictions for this distribution. The PYTHIA6 distribution differs significantly from the other ones in the soft region, especially the difference to PYTHIA8 is very large. In this region the distribution is driven by the Sudakov form factor. The tail of the p_T spectrum is different for all of the used generators that cover a standard, a matched and two power shower options. This region is mainly influenced by the upper phase space boundary for the hardest emission. More detailed explanations for the different behaviour of these models are given in chapter 5.

Figure 9.2 shows the normalised p_T distributions obtained from the reconstructed $t\bar{t}$ pair including statistical errors. In sub-figure 9.2a) the soft region of the $t\bar{t}$ p_T is presented, showing a big distortion compared to the input distributions. The result is affected by two large effects. First, the selection of low-p_T $t\bar{t}$ pairs is relatively suppressed, which will be described in detail below. Furthermore not all $t\bar{t}$ pairs are reconstructed correctly, resulting in migration effects. A migration to higher p_T values is preferred, so the very soft p_T bins are underpopulated and the shape of the different samples becomes very similar. Nevertheless, the relative number of the entries in each bin is still correlated to the input distribution though the difference has decreased significantly. Due to the different selection and reconstruction efficiencies all results are normalised to emphasise the shape of the distribution. Compared to the other samples the total number of entries is much lower for PYTHIA6.

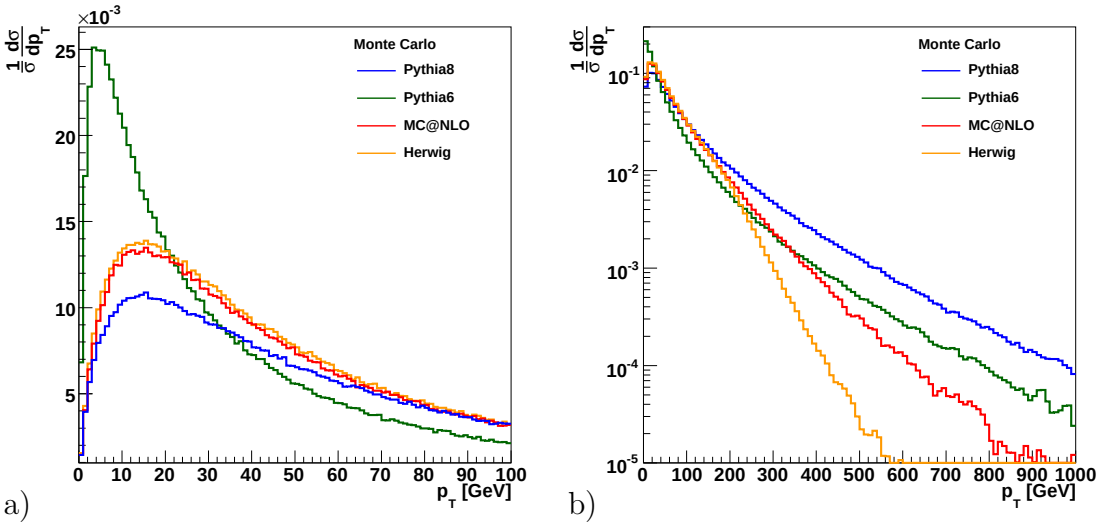


Figure 9.1: $t\bar{t}$ p_T spectrum on generator level for the different generators. a) shows the soft region, b) the tail on a logarithmic scale. The distributions differ significantly in both regions. More details about these spectra can be found in chapter 5.

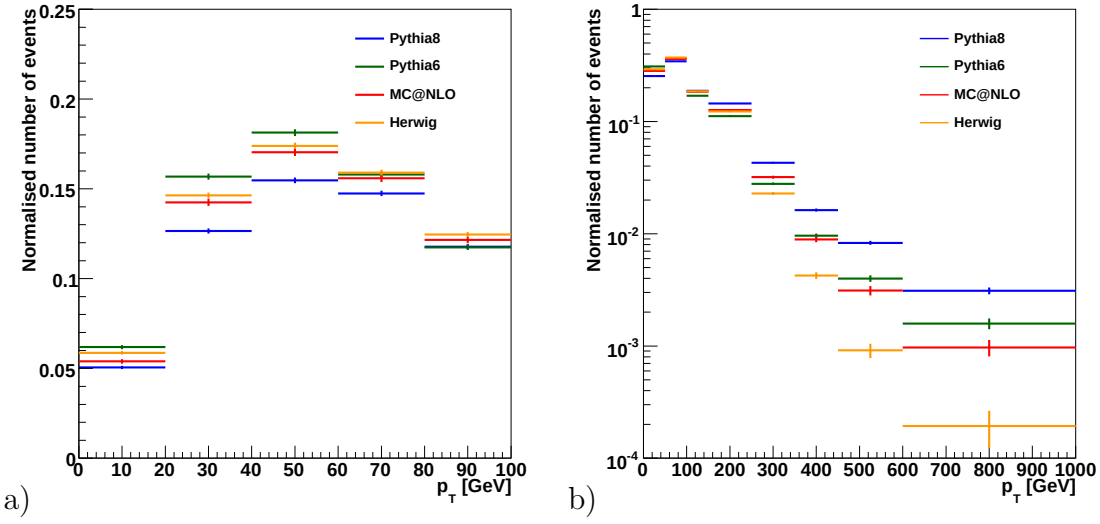


Figure 9.2: a) shows the soft region of the normalised p_T distribution of the reconstructed $t\bar{t}$ pair, b) the tail on a logarithmic scale, both including statistical uncertainties.

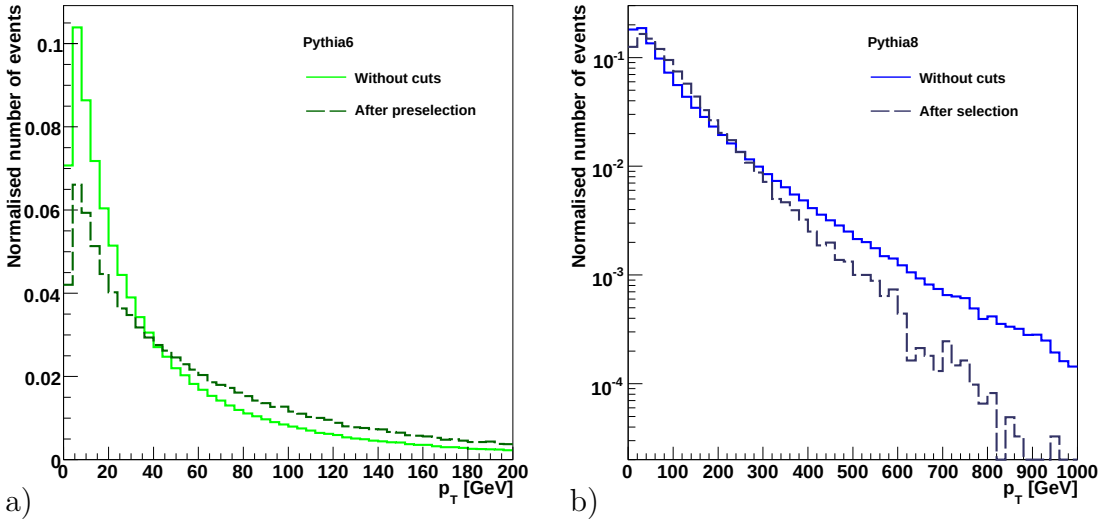


Figure 9.3: a) shows the soft region of the normalised $t\bar{t}$ p_T distribution for Pythia6 on generator level before and after the preselection. The preselection affects the shape of the distribution in the soft region significantly. b) depicts the tail of the $t\bar{t}$ p_T distribution for Pythia8 on generator level before any cuts and after the final selection. In the high-p_T region less events are fitted than for low and intermediate p_T values.

In figure 9.2b) the tail of the $t\bar{t}$ p_T distribution is presented. Here the difference of the signals is more pronounced. The different input distributions can be reconstructed qualitatively.

In the following sections the bias induced by the selection procedure and the systematic uncertainties are discussed. Finally an unfolding procedure is applied to obtain the "true" underlying spectrum.

9.1.1 Selection Bias

Figure 9.3a) shows the soft region of the Pythia6 $t\bar{t}$ p_T spectrum on generator level before and after the application of the preselection. The preselection does not only reduce the total number of events but also changes the shape of the distribution as events in the low-p_T region are more often rejected by the cuts. This is caused by the jet multiplicity cuts that demand at least three (four) jets above 35 GeV (25 GeV). As the $t\bar{t}$ system gains its transverse momentum via radiation, a high-p_T $t\bar{t}$ pair is more often accompanied by high-p_T jets. These additional jets can help the event to pass the preselection cuts, so the higher acceptance is caused by jets not associated with the hard process. The jet multiplicity of an event is correlated with the p_T of the $t\bar{t}$ system. Thus more events with high-p_T $t\bar{t}$ pairs pass the preselection due to "wrong" jets. The impact of this fact is most prominent for Pythia6 because of the high number of events with only very soft radiation.

In figure 9.3b) the tail of the $t\bar{t}$ p_T distribution is shown for PYTHIA8, comparing the sample before any cuts and after the complete selection. While the preselection does not affect the tail of the distribution the kinematic fit leads to a significant drop of reconstructed events in this region. The top quarks are highly boosted in the transverse plane. As this boost is transmitted to their decay products, hadrons associated with different quarks can be assigned to one single jet. Consequently the $t\bar{t}$ system cannot be reconstructed properly in this case and the efficiency of the kinematic fit is significantly reduced.

The choice of the particular jet algorithm has a significant influence on the number of times hadrons originating from more than one quark are combined into one single jet. It turns out that a small R parameter or respectively a small cone radius are best suited to minimise the number of such wrong combinations and to yield a high reconstruction efficiency in the high-p_T region. The KT4 algorithm used in this analysis turns out to show the best performance in this region.

The selection and reconstruction of the $t\bar{t}$ pair thus distorts the shape of the distribution. For a specific analysis of the soft region a more advanced selection procedure is needed, that does select especially low-p_T events. For a correct reconstruction of the tail the inclusion sub-jets may improve the number of reconstructible events.

9.1.2 Systematic Uncertainties

The most important sources for systematic uncertainties are briefly discussed. In chapter 5 the dependence on generator settings was presented. The goal of this study is a comparison of different signal samples, so systematic uncertainties from generator settings are not relevant. The focus is set on the most significant influences of the selection and reconstruction. For the systematic treatment an accuracy is assumed that is expected after the analysis of at least $10\,\mathrm{fb}^{-1}$ of data. First the considered sources for uncertainties are presented and then the actual size of the particular errors is discussed.

Jet Energy Scale

The jet energy scale is varied according to [44]:

$$\frac{\sigma_E}{E} = \begin{cases} 10\% & \text{for } p_T < 20\,\mathrm{GeV} \\ 10\% - 7\% \cdot (p_T - 20\,\mathrm{GeV})/30\,\mathrm{GeV} & \text{for } 20\,\mathrm{GeV} < p_T < 50\,\mathrm{GeV} \\ 3\% & \text{for } p_T > 50\,\mathrm{GeV}\,. \end{cases} \quad (9.1)$$

It affects the number of events passing the preselection jet cuts and the fit result.

Missing Transverse Energy

The $\not{E}_T$ is varied independently from the jet energy scale within 10 % according to [44]. The number of selected events remains almost unchanged.

B-Tagging

The accuracy of the b-tagging is expected to be of order 5 % as stated in [44]. In the reconstruction the three highest b-tagged jets are used, so the most interesting variation is expected to come from a change in the order of the b-tags. All b-tags were smeared with a Gaussian with a width of 5 % of the respective b-tag and processed again.

Background

It is not fully known how good Monte Carlo generators are able to model the investigated physics processes at the LHC. Especially the simulation of QCD events implies a big uncertainty. Following the conservative approaches, the selected background is varied within 50 %.

Pileup

The number of pileup events is directly related to the total cross section. Following [6] the uncertainty of the total cross section at the LHC is of order 20 %. As the number of pileup events is given by a Poisson distribution this systematic is studied by varying the mean of this distribution by 20 %. An increasing number of pileup events is expected to result in a larger number of jets, therefore increasing the combinatorial background.

A dedicated study was performed with particular data samples showing that the number of fitted events grows with the number of pileup events as expected. Variations of the shapes of the resulting distributions are within the statistical uncertainties. Due to the limited statistics of theses samples, their accuracy cannot be compared to the other event samples used. Hence uncertainties caused by pileup effects are not included in the errors shown in the following figures. Nonetheless, pileup uncertainties are expected to influence this analysis and a good understanding and proper treatment of pileup are important.

Systematic Influences on the Transverse Momentum of the $t\bar{t}$ System

The systematic effects are studied for both binnings presented in figure 9.2. First the soft region will be examined. In figure 9.4 the relative shifts for the different systematic influences are shown. The key words "positive" and "negative" refer to the direction of the variation in case of the jet energy scale and $\not{E}_T$. The dependence on the jet energy scale is due to the different number of events passing the selection. The shape of the p_T spectrum remains almost unaffected. The effect is slightly higher for Pythia6 due to the lower number of events passing the selection. The variation of $\not{E}_T$ does not change the number of reconstructed events very much but the shape is distorted compared to the original distribution. A good description of $\not{E}_T$ is hence very important for a good reconstruction. The variation of the b-tag shows almost no influences on the final distribution. For

the background variation only the shift upwards is shown. Compared to the other generators the impact on PYTHIA6 is again higher due to the lower number of events after the selection. The relative contribution to the final distribution is shown in figure 9.5a). In all cases the largest uncertainties are caused by the background and jet energy scale variations.

In sub-figure 9.4g) the total combined error is shown which is of order 15%. A slight dependence on the generators can be observed reflecting the number of events passing the selection. Consequently the relative error is highest for PYTHIA6 and lowest for PYTHIA8.

The relative systematic uncertainties for the binning chosen in figure 9.2b) are presented in figure 9.6 showing a similar behaviour as for the soft region. For the high-p_T region the relative errors grow due to low number of entries in the last bins. Only for PYTHIA8 the relatively high number of entries even in the last bin leads to a throughout stable result. In some plots the entry for HERWIG is missing in the last bin as the relative error becomes very large. For HERWIG the standard shower option with a soft tail was used, leading to very few events in the high-p_T region. Therefore the number of reconstructed events with a transverse momentum above 600 GeV is very small. Like for the binning discussed before it can be observed that the largest uncertainties are caused by the background and the jet energy scale. The error due to $\not\!\!E_T$ grows with increasing p_T and becomes of comparable size to the former two uncertainties for $p_T > 350\,\text{GeV}$.

Sub-figure 9.6g) depicts the total relative uncertainty, showing a significant rise with growing p_T for all generators except PYTHIA8. In the last bin the total error for HERWIG exceeds one. Especially the background causes a big uncertainty for HERWIG as the tail of the normalised background distribution is placed between the tails of MC@NLO and PYTHIA6. In figure 9.5b) the relative contribution of the background to the HERWIG spectrum is shown. The background is much harder than the HERWIG distribution and consequently the last bin in figures 9.2b) and 9.5b) is completely dominated by background events. It should be emphasised that the background is scaled by a large factor due to lower statistics. Therefore it is not possible to describe the shape of the distribution in the low populated bins in the tail correctly. Background samples with higher statistics are needed.

The uncertainties for the separate bins presented in figures 9.4 and 9.6 are treated as uncorrelated. In section 9.1.3 correlations between different bins will be included.

The selection bias and especially the systematic migration to higher reconstructed p_T values distort the reconstructed spectra compared to the generator input. The reconstructed distributions show all a very similar shape. Figure 9.7 shows the same distributions as figure 9.2 but in this case the combined uncertainties including statistical and systematic errors are shown for MC@NLO. As the errors are of comparable size for all samples, the actual distinction power can be deduced. In the soft region depicted in 9.7a) a distinction between the two PYTHIA versions is possible, but HERWIG and MC@NLO are compatible with

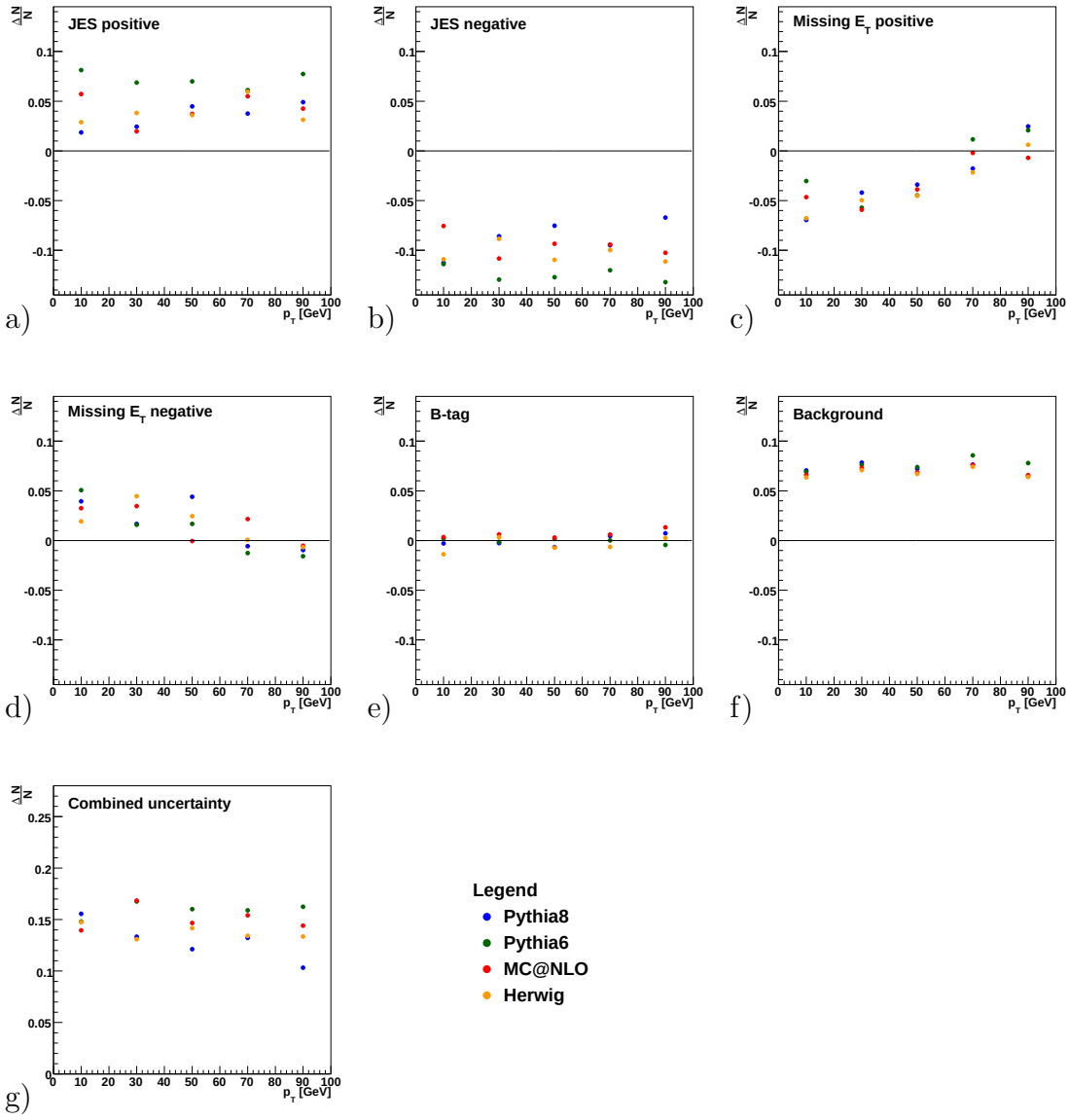


Figure 9.4: Relative deviations for different systematic changes for the soft region of the p_T spectrum of the $t\bar{t}$ system. For the jet energy scale and $\not{E}_T$ the key words "positive" and "negative" refer to the direction of the applied variation. In sub-figure g) the combined relative systematic uncertainty is depicted. The binning is chosen according to figure 9.2a).

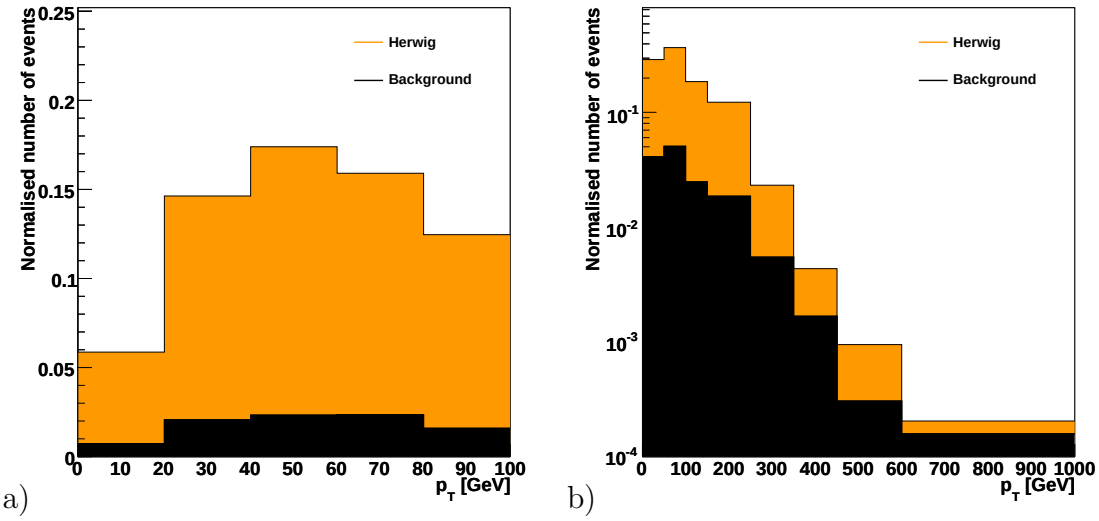


Figure 9.5: The reconstructed p_T distribution obtained from HERWIG is depicted, showing the impact of the background on the final distribution. a) shows the soft region, b) the tail on a logarithmic scale. In the high-p_T region the impact of the background is very large for HERWIG while it is much smaller for the other generators with harder p_T tails.

all other generators. In the high-p_T bins depicted in figure 9.7b) the samples are distinguishable in many cases. It follows that a compatibility check between data and different Monte Carlo models is feasible.

9.1.3 Unfolding

Unfolding of a measured distribution offers the opportunity to extract the "true" underlying distribution of this variable by taking statistical errors and acceptance corrections into account. The possibility to obtain an unfolded distribution is estimated in this section based on the method described in [51].

Monte Carlo studies allow the determination of a transfer matrix A that reflects the transition of a true value x to a measured value y. Unfolding a measured y distribution can be done by solving the least square problem

$$\chi^2 = (Ax - y) \cdot W[y] \cdot (Ax - y)^T . \tag{9.2}$$

Here $W[y]$ denotes the inverse of the covariance matrix of the measured quantity y. The solution of equation 9.2 may show large oscillations. It is possible to apply an additional regularization criterion, leading to a smooth result. The actual equation that needs to be solved is:

$$\chi^2 = (Ax - y) \cdot W[y] \cdot (Ax - y)^T + \tau \cdot L(x - x_0) \cdot L(x - x_0)^T . \tag{9.3}$$

x_0 denotes the bias distribution gained from a Monte Carlo sample. Several regularization conditions $L(x - x_0)$ are possible. Here the difference $x - x_0$ is

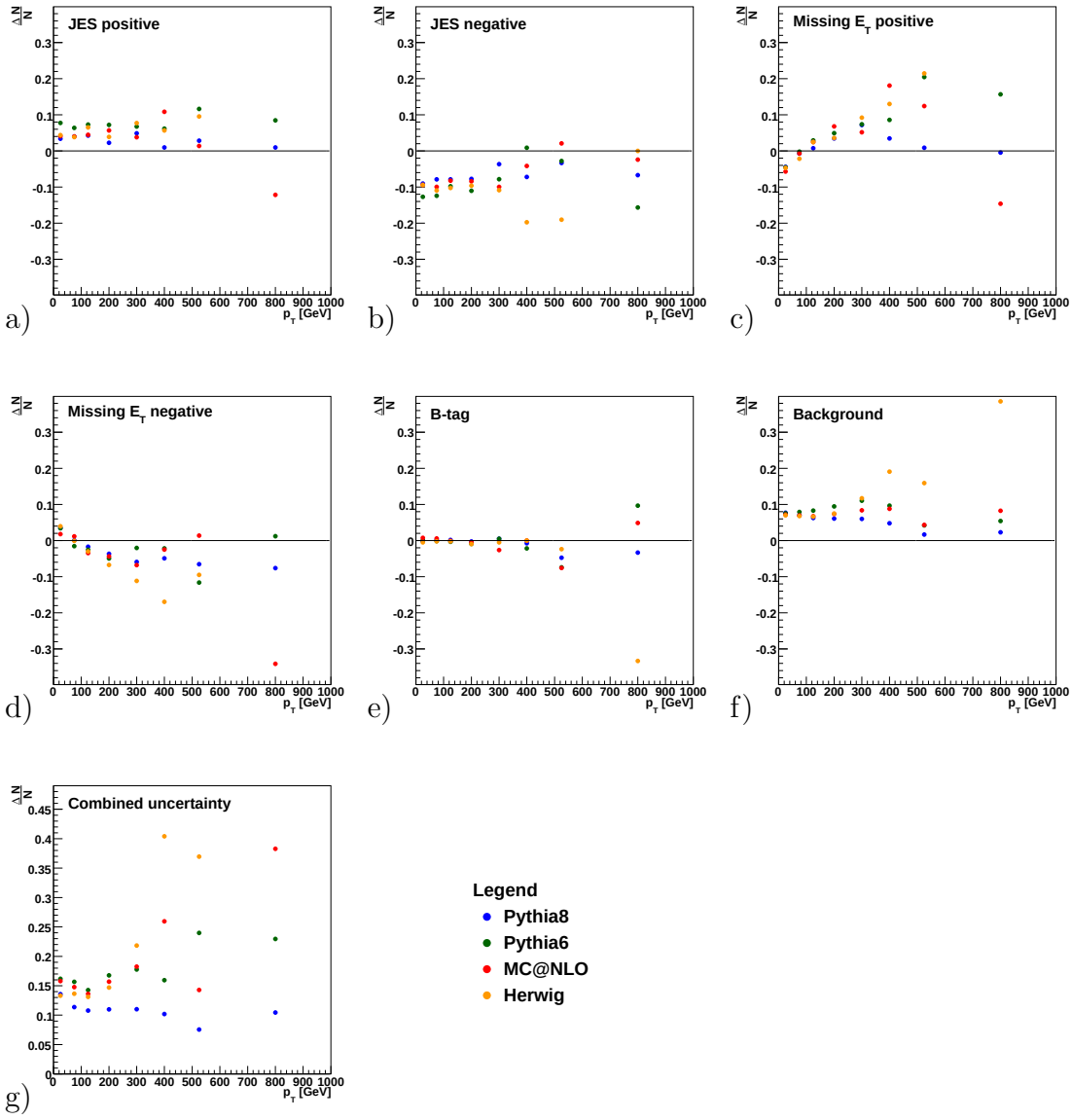


Figure 9.6: Relative deviations for different systematic changes for the the p_T spectrum of the $t\bar{t}$ system focusing on the tail of the distribution. For the jet energy scale and $\not{E}_T$ the key words "positive" and "negative" refer to the direction of the applied variation. In sub-figure g) the combined relative systematic uncertainty is depicted. The binning is chosen according to figure 9.2b).

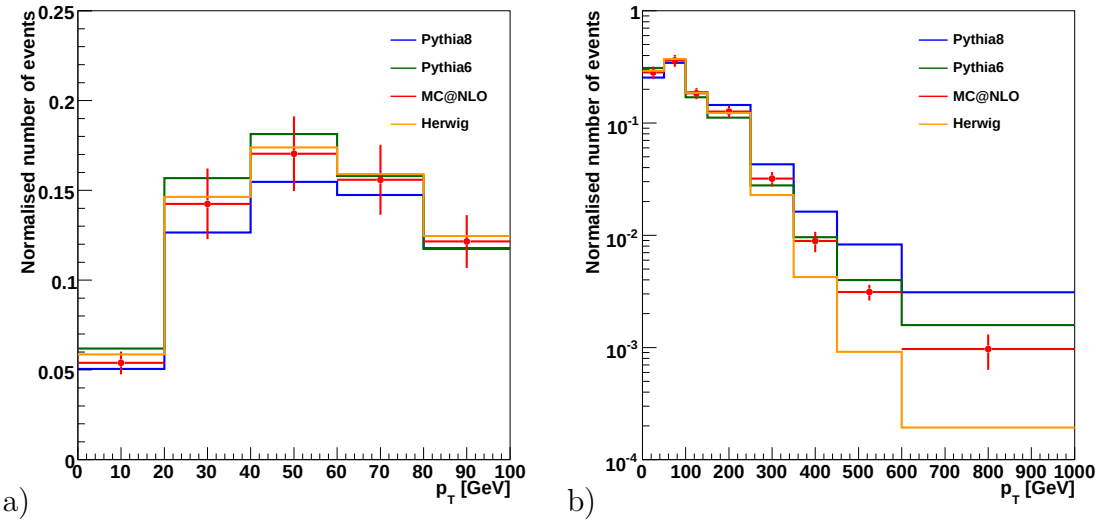


Figure 9.7: a) shows the soft region of the normalised p_T distribution of the reconstructed $t\bar{t}$ pair, b) the tail on a logarithmic scale. For MC@NLO the combined uncertainties are drawn to estimate the distinction power of the final result. The combined uncertainties are of comparable size for all signal samples.

minimised. The regularization parameter τ is determined applying the *L-Curve criterion* [93]. A good choice of τ leads to a solution which is neither dominated by the errors given via the covariance matrix nor by the regularization criterion. Details about unfolding procedures can be found in [51]. The program used here is *TUnfold* [94]. The transfer matrix is chosen to have sixteen bins for the parton level and ten bins for the reconstructed distribution. To reduce the correlations between adjacent bins, finally two neighboured bins are combined into one single bin, so five unfolded bins are obtained.

This unfolding method is applied to the measured p_T spectra. One of the samples is used to compute the transfer matrix for the unfolding procedure. For each systematic an alternative transfer matrix is computed and the full covariance matrix is calculated. Thereby not only the absolute number of events in a bin but also the migration effects are taken into account. For the $\not{E}_T$ and jet energy scale systematic two alternative migration matrices are used. Background subtraction and an assumed uncertainty of 50% are included as well. Figures 9.8 and 9.9 show the unfolded spectra in which MC@NLO and respectively PYTHIA8 have been used for the computation of the transfer matrix. The covariance matrices are given in tables 9.1 to 9.6.

In figures 9.8a) and 9.8b) the unfolded spectra including error bars are shown. The total number of entries is scaled to the expected number for $7\,\mathrm{fb}^{-1}$. As MC@NLO is used for the computation of the migration matrix, the scaling is done according to MC@NLO, leading to a difference in the total number for the various generators due to the different selection efficiencies. Figures 9.9a) and 9.9b) show the the unfolded distributions with PYTHIA8 as reference generator.

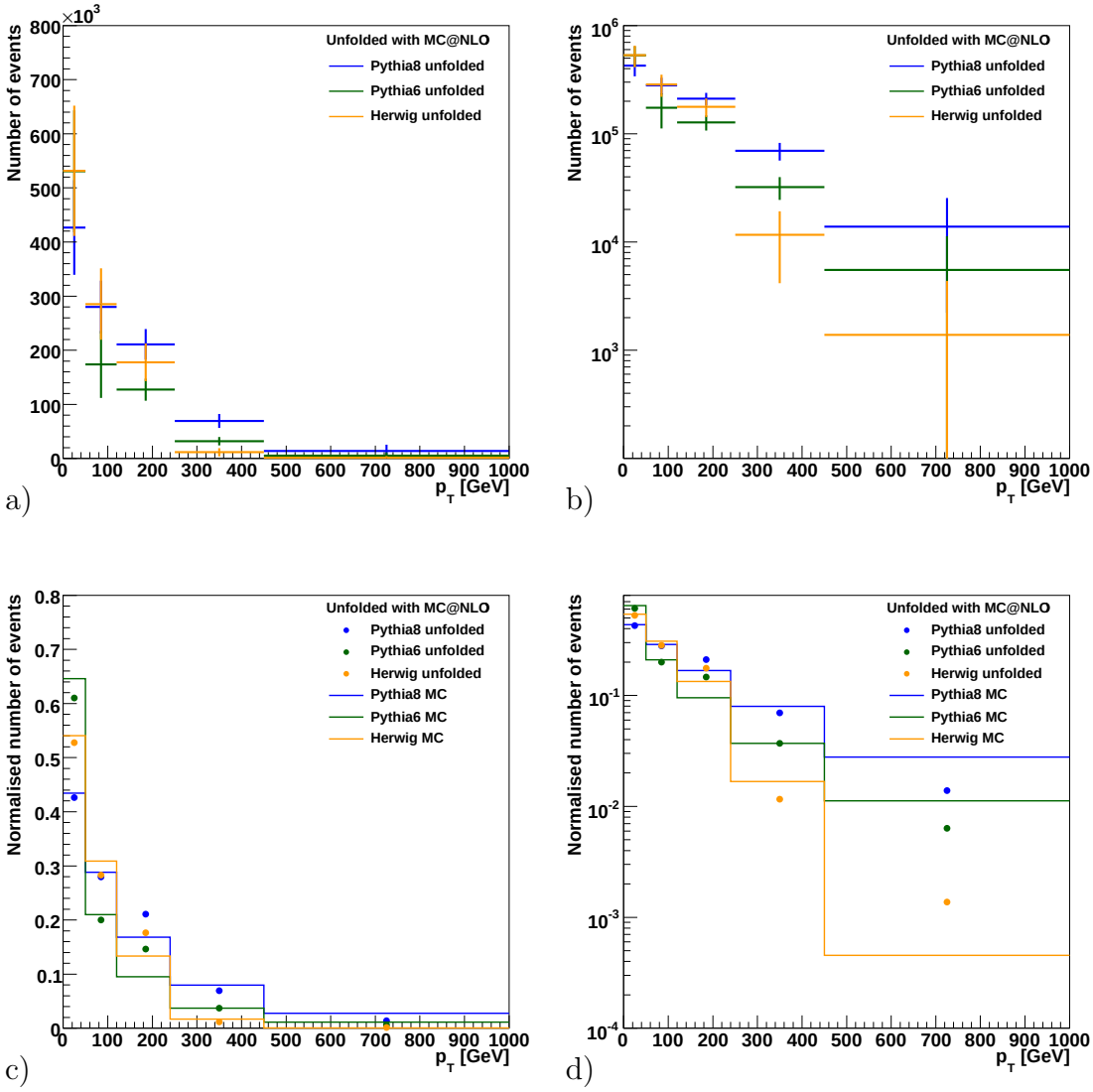


Figure 9.8: p_T spectrum of the $t\bar{t}$ system unfolded with MC@NLO. a) and c) show the soft region, b) and d) the tail on a logarithmic scale. a) and b) present the unfolded distribution including a scaling to the expected number of entries. c) and d) show the normalised result in direct comparison to the distribution on generator level.

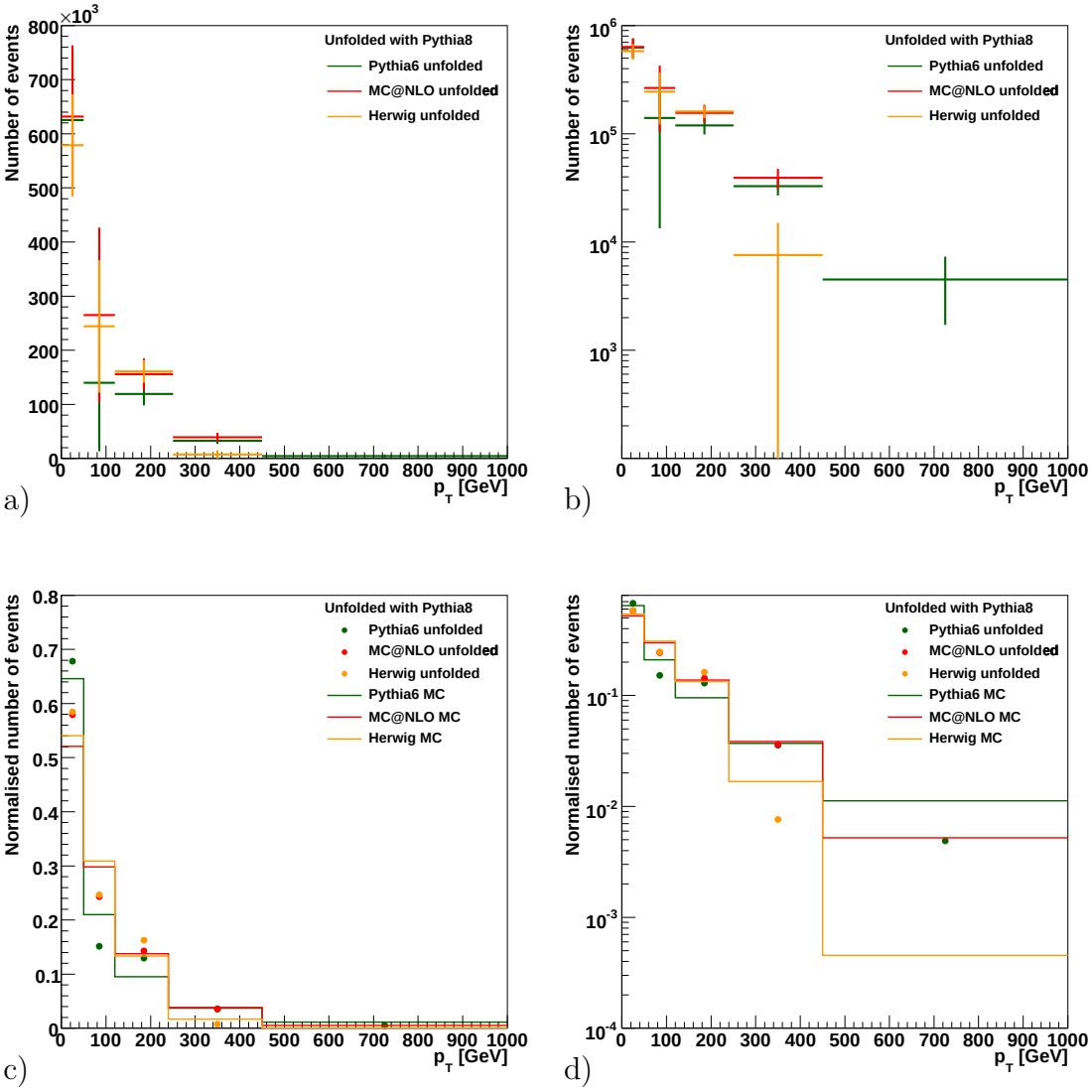


Figure 9.9: p_T spectrum of the $t\bar{t}$ system unfolded with PYTHIA8. a) and c) show the soft region, b) and d) the tail on a logarithmic scale. a) and b) present the unfolded distribution including a scaling to the expected number of entries. c) and d) show the normalised result in direct comparison to the distribution on generator level.

Table 9.1: Covariance matrix for Herwig unfolded with MC@NLO.

Interval [GeV]	$0-50$	$50-120$	$120-250$	$250-450$	$450-1000$
$0-50$	$1.4\cdot10^{10}$	$-1.7\cdot10^{9}$	$1.0\cdot10^{9}$	$-1.8\cdot10^{8}$	$-1.5\cdot10^{8}$
$50-120$	$-1.7\cdot10^{9}$	$4.4\cdot10^{9}$	$8.1\cdot10^{8}$	$3.2\cdot10^{7}$	$-1.1\cdot10^{8}$
$120-250$	$1.0\cdot10^{9}$	$8.1\cdot10^{8}$	$1.2\cdot10^{9}$	$1.4\cdot10^{8}$	$-4.2\cdot10^{7}$
$250-450$	$-1.8\cdot10^{8}$	$3.2\cdot10^{7}$	$1.4\cdot10^{8}$	$5.6\cdot10^{7}$	$1.5\cdot10^{6}$
$450-1000$	$-1.5\cdot10^{8}$	$-1.1\cdot10^{8}$	$-4.2\cdot10^{7}$	$1.5\cdot10^{6}$	$9.0\cdot10^{6}$

Table 9.2: Covariance matrix for Pythia6 unfolded with MC@NLO.

Interval [GeV]	$0-50$	$50-120$	$120-250$	$250-450$	$450-1000$
$0-50$	$1.3\cdot10^{10}$	$-1.6\cdot10^{9}$	$3.7\cdot10^{8}$	$-1.3\cdot10^{8}$	$-1.0\cdot10^{8}$
$50-120$	$-1.6\cdot10^{9}$	$3.8\cdot10^{9}$	$3.7\cdot10^{8}$	$1.4\cdot10^{8}$	$-1.5\cdot10^{8}$
$120-250$	$3.7\cdot10^{8}$	$3.7\cdot10^{8}$	$4.2\cdot10^{8}$	$4.8\cdot10^{7}$	$-2.0\cdot10^{7}$
$250-450$	$-1.3\cdot10^{8}$	$1.4\cdot10^{8}$	$4.8\cdot10^{7}$	$5.9\cdot10^{7}$	$1.6\cdot10^{7}$
$450-1000$	$-1.0\cdot10^{8}$	$-1.5\cdot10^{8}$	$-2.0\cdot10^{7}$	$1.6\cdot10^{7}$	$3.4\cdot10^{7}$

Table 9.3: Covariance matrix for Pythia8 unfolded with MC@NLO.

Interval [GeV]	$0-50$	$50-120$	$120-250$	$250-450$	$450-1000$
$0-50$	$7.6\cdot10^{9}$	$7.6\cdot10^{6}$	$-3.6\cdot10^{7}$	$-6.1\cdot10^{7}$	$9.4\cdot10^{7}$
$50-120$	$7.6\cdot10^{6}$	$2.6\cdot10^{9}$	$7.9\cdot10^{8}$	$5.7\cdot10^{7}$	$-2.8\cdot10^{8}$
$120-250$	$-3.6\cdot10^{7}$	$7.9\cdot10^{8}$	$7.9\cdot10^{8}$	$6.0\cdot10^{7}$	$-1.4\cdot10^{8}$
$250-450$	$-6.1\cdot10^{7}$	$5.7\cdot10^{7}$	$6.0\cdot10^{7}$	$1.7\cdot10^{8}$	$3.4\cdot10^{7}$
$450-1000$	$9.4\cdot10^{7}$	$-2.8\cdot10^{8}$	$-1.4\cdot10^{8}$	$3.4\cdot10^{7}$	$1.4\cdot10^{8}$

Table 9.4: Covariance matrix for HERWIG unfolded with PYTHIA8.

Interval [GeV]	$0-50$	$50-120$	$120-250$	$250-450$	$450-1000$
$0-50$	$8.9\cdot10^{9}$	$-4.4\cdot10^{9}$	$-2.0\cdot10^{8}$	$-4.7\cdot10^{8}$	$2.5\cdot10^{7}$
$50-120$	$-4.4\cdot10^{9}$	$1.5\cdot10^{10}$	$1.9\cdot10^{8}$	$6.0\cdot10^{7}$	$-1.9\cdot10^{8}$
$120-250$	$-2.0\cdot10^{8}$	$1.9\cdot10^{8}$	$4.4\cdot10^{8}$	$2.8\cdot10^{6}$	$-1.5\cdot10^{7}$
$250-450$	$-4.7\cdot10^{8}$	$6.0\cdot10^{7}$	$2.8\cdot10^{6}$	$5.6\cdot10^{7}$	$-2.5\cdot10^{6}$
$450-1000$	$2.5\cdot10^{7}$	$-1.9\cdot10^{8}$	$-1.5\cdot10^{7}$	$-2.5\cdot10^{6}$	$4.7\cdot10^{6}$

Table 9.5: Covariance matrix for MC@NLO unfolded with PYTHIA8.

Interval [GeV]	$0-50$	$50-120$	$120-250$	$250-450$	$450-1000$
$0-50$	$1.0\cdot10^{10}$	$-6.0\cdot10^{9}$	$-5.5\cdot10^{8}$	$-3.9\cdot10^{8}$	$-1.1\cdot10^{8}$
$50-120$	$-6.0\cdot10^{9}$	$1.5\cdot10^{10}$	$8.3\cdot10^{7}$	$2.9\cdot10^{8}$	$-1.5\cdot10^{8}$
$120-250$	$-5.5\cdot10^{8}$	$8.3\cdot10^{7}$	$5.1\cdot10^{8}$	$4.0\cdot10^{7}$	$-2.2\cdot10^{7}$
$250-450$	$-3.9\cdot10^{8}$	$2.9\cdot10^{8}$	$4.0\cdot10^{7}$	$4.0\cdot10^{7}$	$5.5\cdot10^{5}$
$450-1000$	$-1.1\cdot10^{8}$	$-1.5\cdot10^{8}$	$-2.2\cdot10^{7}$	$5.5\cdot10^{5}$	$8.0\cdot10^{6}$

Table 9.6: Covariance matrix for PYTHIA6 unfolded with PYTHIA8.

Interval [GeV]	$0-50$	$50-120$	$120-250$	$250-450$	$450-1000$
$0-50$	$1.6\cdot10^{10}$	$-6.8\cdot10^{9}$	$-5.1\cdot10^{8}$	$-2.0\cdot10^{8}$	$-1.7\cdot10^{8}$
$50-120$	$-6.8\cdot10^{9}$	$1.6\cdot10^{10}$	$2.5\cdot10^{8}$	$2.1\cdot10^{8}$	$-2.3\cdot10^{7}$
$120-250$	$-5.1\cdot10^{8}$	$2.5\cdot10^{8}$	$4.5\cdot10^{8}$	$5.0\cdot10^{7}$	$-1.3\cdot10^{7}$
$250-450$	$-2.0\cdot10^{8}$	$2.1\cdot10^{8}$	$5.0\cdot10^{7}$	$3.5\cdot10^{7}$	$-1.0\cdot10^{5}$
$450-1000$	$-1.7\cdot10^{8}$	$-2.3\cdot10^{7}$	$-1.3\cdot10^{7}$	$-1.0\cdot10^{5}$	$7.8\cdot10^{6}$

The big distortion of the measured spectrum shown in figure 9.2a) and the corresponding errors lead to large uncertainties in the first bins. Using MC@NLO as reference it is possible to unfold all bins whereas Pythia8 massively underestimates the number of entries from Herwig and MC@NLO in the last bin. This is caused by the comparably large number of entries from Pythia8. In addition the background sample has to be split into two subsamples for a correct treatment, thereby increasing the weights of the background events to even higher values. For a better unfolding result a much more precise background description is needed.

In sub-figures c) and d) the unfolded result is compared to the generator prediction. In all cases the principal shape is described correctly and in almost all cases the theoretical prediction lies within the error bars of the unfolded result. Thus it is possible to unfold a measured distribution even if the "wrong" Monte Carlo model is applied. Reducing the number of events for the computation of the transfer matrix worsens the agreement of unfolded and predicted distributions considerably.

The covariance matrices (tables 9.1 to 9.6) are fully dominated by systematic errors, the statistical errors are about one order of magnitude smaller. For such a measurement systematic effects must be known to a higher precision than assumed here. Furthermore this procedure suffers extremely from the underpopulation of the very soft region. $50-70\%$ of all events are assumed to be produced with a transverse momentum below 50 GeV, so any error on this region propagates to the other regions changing the whole distribution. Especially a distribution similar to Pythia6 is affected by this.

It is important to note the large (anti-)correlations between the different bins shown in tables 9.1 to 9.6. Especially between the low and high populated bins a large interrelation is visible. This is caused by the migration already visible in figure 9.2a) where the low-p_T region is extremely underpopulated. As the events generated in the low-p_T region are reconstructed with a higher p_T, the different bins cannot be regarded independently. Over the full range a migration to higher p_T values is preferred. For the last bin there is always at least one off-diagonal element with an absolute value higher than its actual variance.

All results presented here are obtained using the KT4 jet algorithm due to reasons discussed in section 9.1.1. Performing the unfolding procedure with events reconstructed with other jet algorithms shows that the final unfolding result is not influenced very much. As stated above a small R parameter or respectively a small cone radius increase the number of reconstructed high-p_T $t\bar{t}$ pairs. Consequently the error on the last bin and the covariances associated with this bin are much smaller. In a comparison of KT4, KT6, SC4, SC5 and SC7 it turns out that in most cases the KT4 algorithm leads to the lowest covariances. The worst result is achieved with the SC7 algorithm, whereas the performances of KT6, SC4 and SC5 are only slightly worse. A direct comparison of the reconstructed result to the generator prediction shows only small differences.

Basically an unfolding of a measured data sample can be accomplished, though it will be quite challenging. An adequate understanding of the systematic influ-

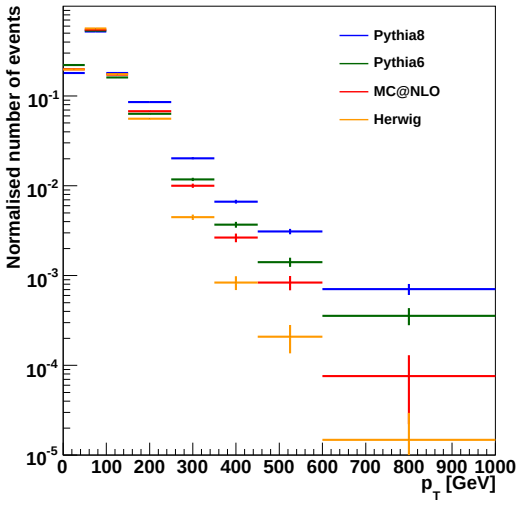


Figure 9.10: Normalised transverse momentum of the hardest reconstructed jet on detector level in each event with statistical errors. The selection procedure is merely used for event tagging.

ences and a good background description are needed. In addition the migration matrices have to be computed from Monte Carlo samples with high statistics.

9.2 Reconstructed Jets

9.2.1 Transverse Momenta of Reconstructed Jets

Jets in $t\bar{t}$ events originate from the decay products of the top quarks, radiation, multi parton interactions and from pileup events. According to the studies presented in chapter 5 gluon emissions can cause high-p_T jets. Choosing the hardest jet from each selected event leads to the p_T spectrum depicted in figure 9.10. Except for small differences, mostly due to normalization, the distributions look identical up to transverse momenta of order 150 GeV. For higher p_T values the expected distribution due to hard emissions is observed, offering the possibility to gain a hint on the available phase space only by checking the hardest jet in events identified as $t\bar{t}$ events.

In section 8.5 an extended selection asking for an additional jet is discussed. Using this selection an investigation of additional hard jets becomes possible.

The hardest jet that is not used for the reconstruction of the $t\bar{t}$ system is assumed to be caused by a hard emission. The transverse momentum distributions of these jets are shown in figure 9.11. The peak region is significantly different to the one shown in figure 9.10 but the tails of the distributions reveal a very similar behaviour. The transverse momentum of the additional hardest jet is sensitive to the phase space for hard emissions. In both figures the jets are not calibrated, which leads to only a qualitative statement. After a calibration of gluon jets is

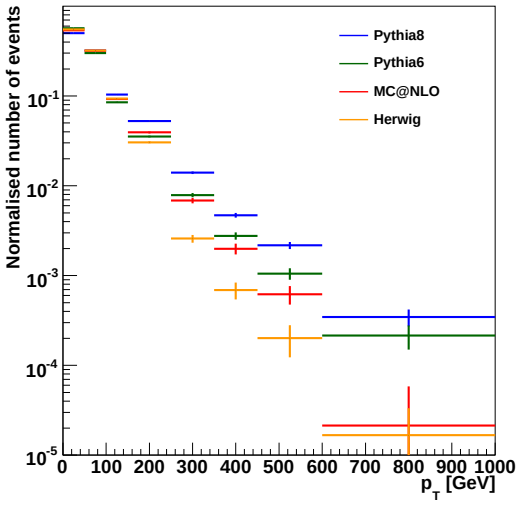


Figure 9.11: Normalised transverse momentum of the hardest jet which is not used for the $t\bar{t}$ reconstruction on detector level with statistical errors.

applied, these spectra may help in identifying the upper phase space boundary for hard emissions.

The relative shifts due to systematic effects are presented in figure 9.12. As in the case of the $t\bar{t}$ system huge deviations can be observed for the high-p_T bins, again showing sensitivity to the total number of event reconstructed in the bin. For Herwig the total error becomes larger than one already at a p_T of order 300 GeV as shown in figure 9.12g). The last bins of the Herwig and MC@NLO distributions are again governed by background events with high weights and even for Pythia6 the error becomes very large. The background distribution of the jet p_T is very hard, located somewhere between Pythia6 and Pythia8. The actual influence on the Herwig spectrum is presented in figure 9.13a). The systematic influences from $\not{E}_T$ and b-tagging are negligible except for the bins in which statistical effects lead to large deviations. In figure 9.13b) the same distributions as in figure 9.11 are presented but the combined uncertainties for MC@NLO are shown to study the distinction power between the different samples. In many cases the different samples can be distinguished from each other.

9.2.2 Rapidity of Reconstructed Jets

For the analysis of the rapidity distribution the extended selection with very hard cuts discussed in section 8.5 is applied.

Figure 9.14 shows the absolute value of the hardest additional jet's rapidity with respect to the $t\bar{t}$ system on generator level. All selection cuts discussed in section 8.5 are applied. The cut on the jet p_T results in a smaller difference between the p_T^2 and Q^2 ordered showers. The imposed angular ordering for the Q^2 ordered shower is designed especially for the soft and collinear region, thus its influence is smaller for high-p_T radiation.

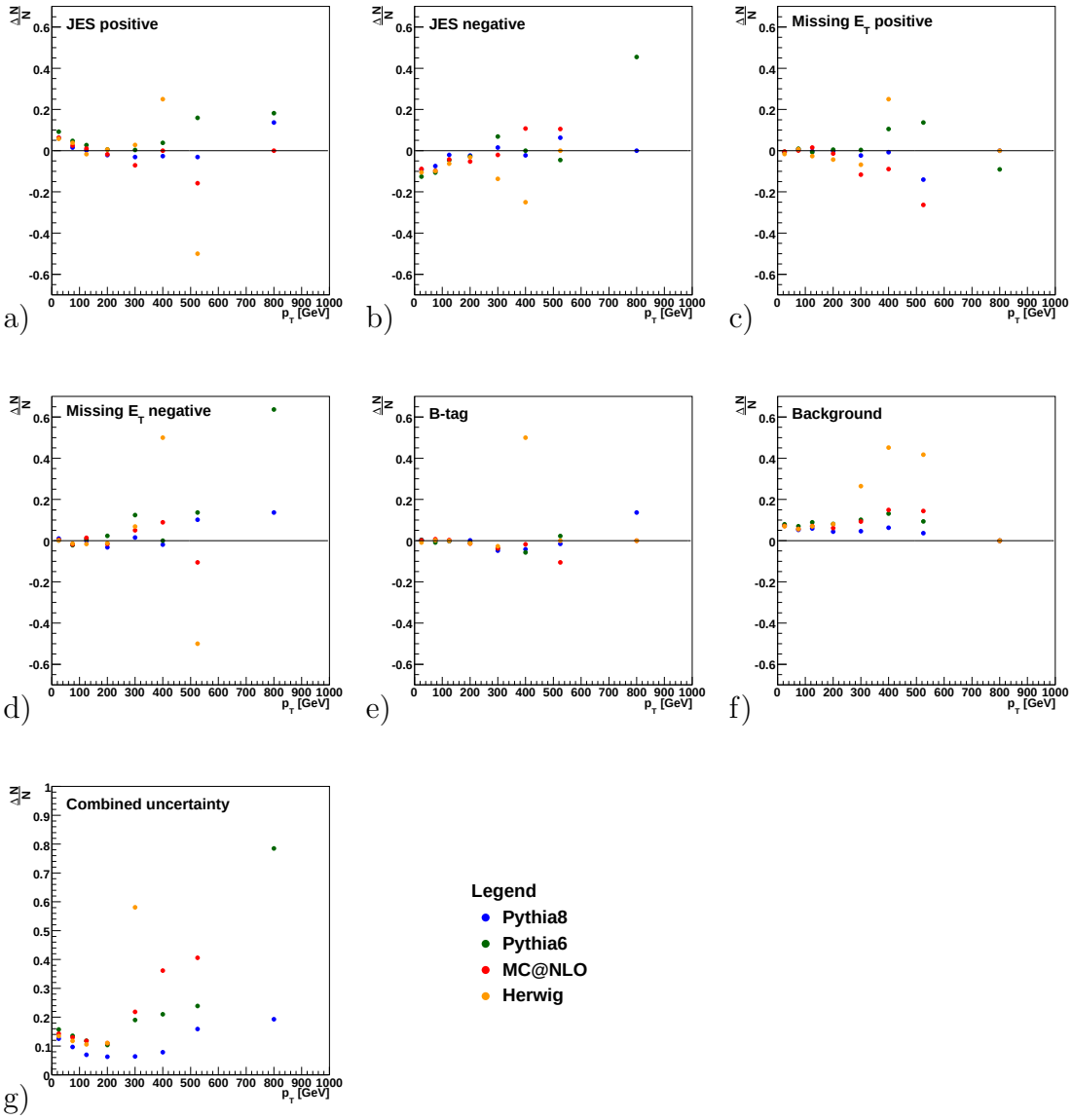


Figure 9.12: Relative deviations for different systematic changes for the p_T spectrum of the additional hardest jet focusing on the tail of the distribution. For the jet energy scale and $\not{E}_T$ the key words "positive" and "negative" refer to the direction of the applied variation. In sub-figure g) the combined relative systematic uncertainty is depicted. The binning is chosen according to figure 9.11. In some cases the uncertainties exceed one.

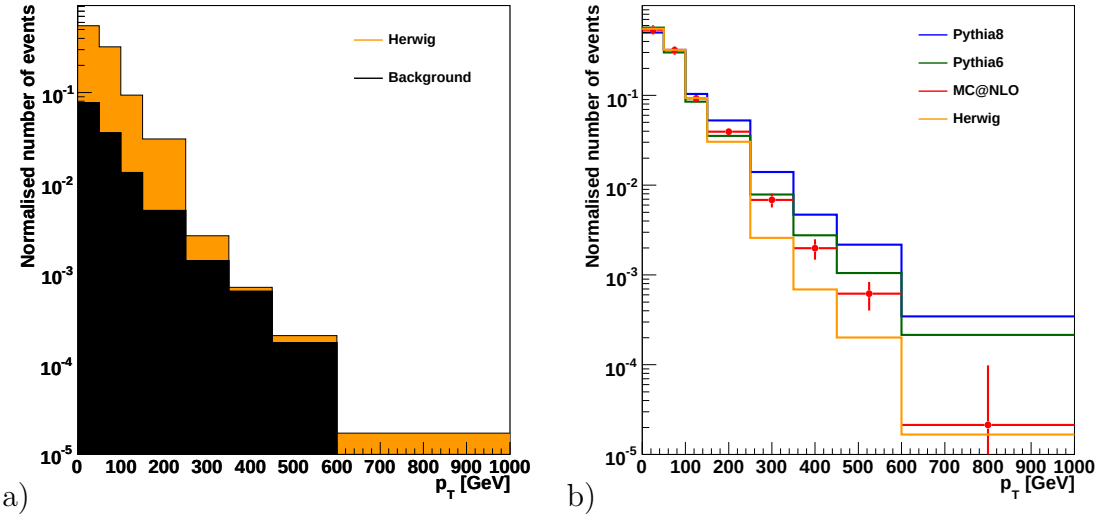


Figure 9.13: Normalised transverse momentum of the hardest jet which is not used for the $t\bar{t}$ reconstruction. In a) the influence of the background on the Herwig result is shown. In b) the combined uncertainties for MC@NLO are drawn to estimate the distinction power of the final result. In the low-p_T region the combined uncertainties are of comparable size for all signal samples.

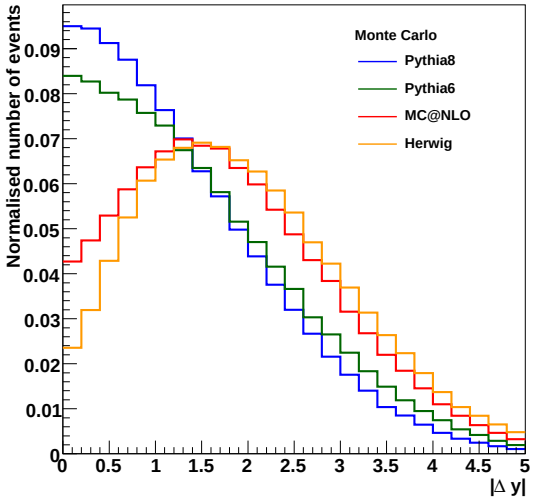


Figure 9.14: Absolute rapidity of the leading jet with respect to the $t\bar{t}$ system $\Delta y = y_{jet} - y_{t\bar{t}}$ on generator level for all four generators including the cuts described in section 8.5.

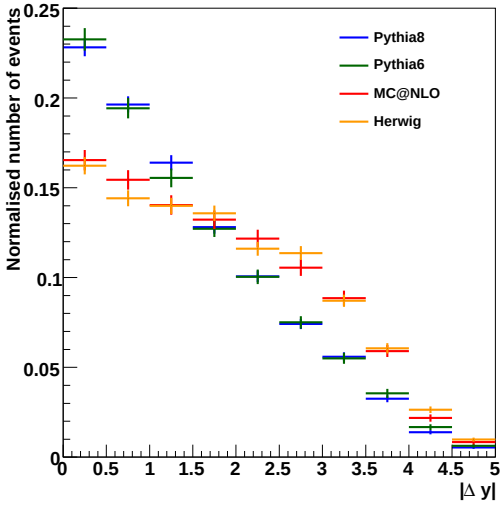


Figure 9.15: Absolute rapidity of the additional leading jet with respect to the reconstructed $t\bar{t}$ system $\Delta y = y_{jet} - y_{t\bar{t}}$ on detector level.

In figure 9.15 the relative rapidity of the first jet is shown for the reconstructed events. The distributions from PYTHIA6 and PYTHIA8 are indistinguishable but clearly different from HERWIG and MC@NLO. The latter two do not show the dip structure observed on generator level but there are significantly less events with a low relative rapidity compared to the PYTHIA samples. It turns out that the background distribution on its own has a similar shape as the two PYTHIA versions. The impact of the background is illustrated in figure 9.16a) using the HERWIG signal as example. Furthermore combinatorics prefer low relative rapidity values, thereby physics and combinatorial background both fill the low relative rapidity region. Figures 5.13 and 9.14 show that the distributions of PYTHIA6 and PYTHIA8 become more similar for higher values of the jet p_T. As the cut of 50 GeV is applied to uncalibrated jets, the emissions associated with these jets are expected to have an even higher transverse momentum. Thus the agreement of the reconstructed distributions of the two PYTHIA versions is not surprising.

The systematic shifts are presented in figure 9.17. The uncertainties due to $\not{E}_T$ and b-tagging are governed by statistical fluctuations. For the jet energy scale a trend can be identified but again the fluctuations are very large. The background and the total uncertainties show a clear structure depending on the total number of entries in a bin. For low relative rapidities the relative errors are larger for HERWIG and MC@NLO and for high values the uncertainties are larger for the PYTHIA samples, depending on the total number of events in the particular region.

Within the errors the curves are distinguishable and the influence of the phase space boundaries due to angular ordering can be examined. Figure 9.16b) shows the same distributions as figure 9.15, including the combined statistical and systematic uncertainties for MC@NLO. The uncertainties are of comparable size

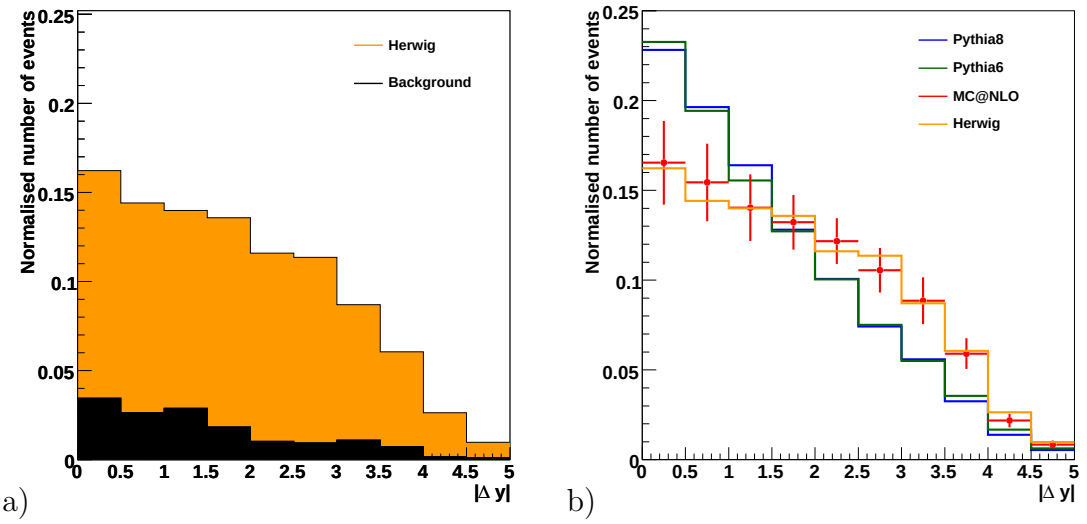


Figure 9.16: Absolute rapidity of the additional leading jet with respect to the reconstructed $t\bar{t}$ system $\Delta y = y_{jet} - y_{t\bar{t}}$ on detector level. In a) the impact of the background on the Herwig result is shown. In b) the combined uncertainties for MC@NLO are drawn to estimate the distinction power of the final result. The combined uncertainties are of comparable size for all signal samples.

for all samples, thus a clear distinction between the angular ordered and the non-angular ordered results is feasible.

9.3 Conclusions

As the results of this chapter show, a distinction between the different investigated radiation models is possible. Once LHC is running, data can be compared to the various models.

The reconstructed transverse momentum spectrum of the $t\bar{t}$ system is distorted due to a selection bias and huge migration effects. For the soft region, that is mainly affected by the migration of events, a better handling of the systematic uncertainties is needed. The hard tail can be reconstructed well and the different models are distinguishable. Here the shape of the background distribution is important and must be known to high precision. In spite of the high systematic uncertainties and the large migration effects, the different models can be distinguished from each other within the systematic uncertainties in most bins. An unfolding of a measured distribution seems to be feasible but a good understanding of the background and a careful treatment of the systematic uncertainties are needed.

The transverse momentum distribution of the hardest additional jet shows a good sensitivity to the upper phase space limit in the the high-p_T range of the distribution. In addition the rapidity with respect to the $t\bar{t}$ system can be determined well after the application of hard cuts. Even though the fraction of

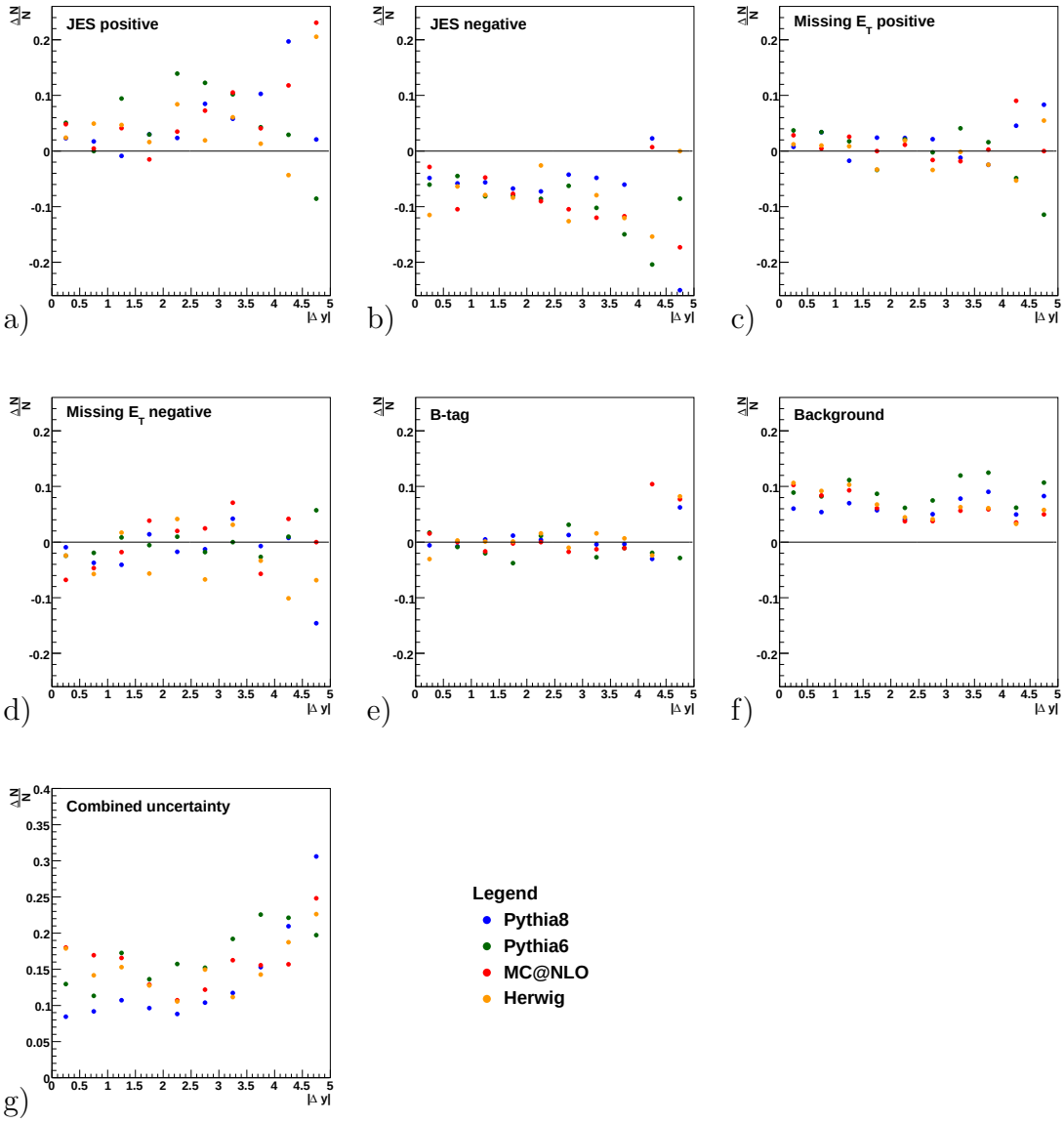


Figure 9.17: Relative deviations for different systematic changes for the the absolute rapidity spectrum of the additional hardest with respect to the $t\bar{t}$ system. For the jet energy scale and $\not{E}_T$ the key words "positive" and "negative" refer to the direction of the applied variation. In sub-figure g) the combined relative systematic uncertainty is depicted. The binning is chosen according to figure 9.15.

events used for the rapidity study is quite low, a clear distinction between the models with and without a dead zone due to colour coherence effects can be accomplished.

The presented analysis offers the opportunity to compare different radiation models to data. Since a good understanding of systematic uncertainties and background description are needed, a lot of experience has to be gained by CMS before a realization of this study is possible.

Chapter 10

Summary and Outlook

10.1 Summary

Due to the large top mass, top pairs are well suited for the investigation of QCD radiation at the LHC, especially of high-p_T emissions. The transverse momentum of the $t\bar{t}$ system is directly sensitive to QCD radiation and offers the opportunity to investigate completely different phase space regions. The soft peak region is governed by the Sudakov form factor leading to a sensitivity on the evolution scale and α_s. The structure of the high-p_T tail depends on the available phase space region. Real emissions lead to additional jets, so a correlation between the kinematics of the $t\bar{t}$ system and high-p_T jets is expected. An examination of these observables allows to compare different radiation models implemented in Monte Carlo event generators.

The results from the generator study show that the soft region of the p_T of the $t\bar{t}$ system reveals big differences depending on the actual settings for the parton shower. Different tunes with different parameter choices exist. In the high-p_T region MC@NLO and ALPGEN both agree with each other. The standard parton showers show a much softer behaviour due to the phase space limits set by the factorization scale. In contrast the power showers exploiting the whole available phase space show a much harder tail. Here a sensitivity to the actual ordering in the shower is observed: The p_T^2 ordered shower simulates more hard emissions. Investigating additional jets shows a clear correlation between the hardest jet and the transverse momentum of the $t\bar{t}$ system. In all scenarios the p_T of the $t\bar{t}$ pair is mainly driven by the hardest emission. The rapidity distribution of the hardest jet in the frame of the $t\bar{t}$ system shows a dip structure for the angular ordered shower. The reason is a dead zone in the phase space that is caused by the inclusion of colour coherence effects. The jet multiplicity turns out to depend on various aspects, so for the analysis of specific effects additional information is needed.

A selection and reconstruction of top pairs after the full CMS detector simulation introduces several difficulties. While the background can be suppressed to a reasonable level, the selection is sensitive to the kinematic configuration of the simulated $t\bar{t}$ system. Nonetheless, the p_T spectra of the reconstructed $t\bar{t}$ pairs

differ for the various inputs and a compatibility check between data and different Monte Carlo models is possible. Applying an unfolding procedure allows to obtain the underlying spectrum with a reasonable precision. The errors are dominated by systematic uncertainties. Especially a good understanding of the jet energy scale and a trustworthy computation of $\not{E}_T$ are needed. The high-p_T tail of the hardest jet in each $t\bar{t}$ event is governed by the hardest additional emission. Moreover, the hardest additional jet not used for the $t\bar{t}$ reconstruction can be examined. Both options show a clear sensitivity on the upper phase space boundary for hard emissions. Finally the rapidity of the additional jet in the frame of the reconstructed $t\bar{t}$ system reflects the available phase space regions. The dip structure of the relative rapidity distribution from the angular ordered shower can be identified.

The CMS experiment offers the opportunity to compare the different models to data. The deviations of the models studied in this analysis could be identified on the reconstruction level. Once LHC is running, it will be possible to measure these distributions with an accuracy high enough to differentiate between the various models. Both the transverse momentum spectrum of the $t\bar{t}$ system and the properties of the additional jets are very well suited for a determination of the radiation parameters. Therefore higher order calculations can be checked and possible modifications of the spectra due to contributions of non-standard model physics processes may be identified.

10.2 Outlook

On generator level additional models can be investigated. Different schemes for matrix element matching and alternative parton shower evolutions based on approaches different from the DGLAP equations exist. An extension of the presented analysis with these generators would lead to an increased understanding of the underlying physics.

On reconstruction level a diminishment of systematic uncertainties is very important. Additional Monte Carlo samples with sufficient statistics could facilitate the inclusion of pileup effects in the treatment of systematic uncertainties. Furthermore, background samples with higher statistics would help to gain a better understanding of the high-p_T region. Performing two separate analyses focusing on to the low- and high-p_T regions, respectively, it may be possible to obtain better reconstructed results in the particular region. Finally, the performance of the unfolding could be improved with a larger Monte Carlo signal sample for the computation of the transfer matrix.

Appendix A

Technical Details

In the following some technical specifications of the used software and some generator settings are summarised.
Used generator versions for the studies presented in chapter 5:

- ALPGEN: version 2.13,
- HERWIG6: version 6.510 with JIMMY version 4.31,
- HERWIG++: version 2.2.0,
- MC@NLO: version 3.3 with HERWIG6 version 6.510,
- PYTHIA6: version 6.416 with a modification in *pyveto* [95] to enable matching of ALPGEN events with a p_T^2 ordered shower,
- PYTHIA8: version 8.100.

The four signal samples analysed in chapters 8 and 9 were produced with the CMS software version CMSSW_1_6_12. For MC@NLO the CTEQ6m pdf was used while for the other generators the corresponding LO pdf CTEQ6l was utilised. The top mass was set to $m_t = 175\,\text{GeV}$ for all generators.
Used generator versions for the production of these signal samples:

- HERWIG6: version 6.510 with JIMMY version 4.2,
- MC@NLO: version 3.3 with HERWIG6 version 6.510 and JIMMY version 4.2,
- PYTHIA6: version 6.409,
- PYTHIA8: version 8.100.

The used background samples were produced during the CSA07 challenge [90] with the CMS software version CMSSW_1_5_2. The $t\bar{t}$, W and Z plus jets backgrounds were generated with ALPGEN [91] while the QCD background was generated with PYTHIA6. The changes made during the development from CMSSW version 1_5_2 to 1_6_12 do not affect the reconstructed objects of interest [96].

Bibliography

[1] (Ed.) Bruning, Oliver S. et al. LHC design report. Vol. I: The LHC main ring. CERN-2004-003-V-1.

[2] (Ed.) Buning, O. et al. LHC Design Report. 2. The LHC infrastructure and general services. CERN-2004-003-V-2.

[3] (Ed.) Benedikt, M., (Ed.) Collier, P., (Ed.) Mertens, V., (Ed.) Poole, J., and (Ed.) Schindl, K. LHC Design Report. 3. The LHC injector chain. CERN-2004-003-V-3.

[4] LEP Design Report. Vol. 1. The LEP Injector Chain. CERN-LEP/TH/83-29.

[5] LEP Design Report: Vol. 2. The LEP Main Ring. CERN-LEP-84-01.

[6] P. V. Landshoff. The total cross section at the LHC. *Acta Phys. Polon.*, B39:2063–2094, 2008.

[7] TeVI Group. Design Report Tevatron 1 project. FERMILAB-DESIGN-1984-01.

[8] (Ed.) Altarelli, G. and (Ed.) Mangano, Michelangelo L. Standard model physics (and more) at the LHC. Proceedings, Workshop, Geneva, Switzerland, May 25-26, October 14-15, 1999. CERN-2000-004.

[9] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[10] R. Adolphi et al. The CMS experiment at the CERN LHC. *JINST*, 0803:S08004, 2008.

[11] A. Augusto Alves et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[12] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 0803:S08002, 2008.

[13] O. Adriani et al. The LHCf detector at the CERN Large Hadron Collider. *JINST*, 3:S08006, 2008.

[14] G. Anelli et al. The TOTEM experiment at the CERN Large Hadron Collider. *JINST*, 3:S08007, 2008.

[15] `http://lhcathome.cern.ch/images/LHCdesign.jpg`. Accessed: July 2009.

[16] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[17] Q. R. Ahmad et al. Direct evidence for neutrino flavor transformation from neutral-current interactions in the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 89:011301, 2002.

[18] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[19] Abdus Salam. Weak and Electromagnetic Interactions. Originally printed in Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held 1968 At Lerum, Sweden*, Stockholm 1968, 367-377.

[20] Steven Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[21] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531–533, 1963.

[22] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[23] Peter W. Higgs. Broken Symmetries And The Masses Of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, 1964.

[24] D. N. Spergel et al. First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. *Astrophys. J. Suppl.*, 148:175–194, 2003.

[25] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. *Phys. Rev. Lett.*, 74:2626–2631, 1995.

[26] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[27] C. Amsler et al. Review of particle physics. *Phys. Lett.*, B667:1, 2008.

[28] Werner Bernreuther. Top quark physics at the LHC. *J. Phys.*, G35:083001, 2008.

[29] R. K. Ellis, W. J. Stirling, and B. R. Webber. *QCD and Collider Physics*. Cambridge University Press, 1996.

[30] Matteo Cacciari, Stefano Frixione, Michelangelo L. Mangano, Paolo Nason, and Giovanni Ridolfi. Updated predictions for the total production cross sections of top and of heavier quark pairs at the Tevatron and at the LHC. *JHEP*, 09:127, 2008.

[31] T. Aaltonen et al. First Observation of Electroweak Single Top Quark Production. 2009.

[32] V. M. Abazov et al. Observation of Single Top Quark Production. 2009.

[33] Günther Dissertori, Ian Knowles, and Michael Schmelling. *Quantum Cchromodynamics, High Energy Experiments and Theory.* Oxford University Press, 2003.

[34] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl. Phys.*, B126:298, 1977.

[35] V. N. Gribov and L. N. Lipatov. Deep inelastic e p scattering in perturbation theory. *Sov. J. Nucl. Phys.*, 15:438–450, 1972.

[36] Yuri L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov. Phys. JETP*, 46:641–653, 1977.

[37] E. A. Kuraev, L. N. Lipatov, and Victor S. Fadin. The Pomeranchuk Singularity in Nonabelian Gauge Theories. *Sov. Phys. JETP*, 45:199–204, 1977.

[38] I. I. Balitsky and L. N. Lipatov. The Pomeranchuk Singularity in Quantum Chromodynamics. *Sov. J. Nucl. Phys.*, 28:822–829, 1978.

[39] S. I. Alekhin. Parton distributions from deep-inelastic-scattering data. *Phys. Rev. D*, 68(1):014002, Jul 2003.

[40] H. L. Lai et al. Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions. *Eur. Phys. J.*, C12:375–392, 2000.

[41] J. Pumplin et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.

[42] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne. Parton distributions incorporating QED contributions. *Eur. Phys. J.*, C39:155–161, 2005.

[43] S. Chekanov et al. A ZEUS next-to-leading-order QCD analysis of data on deep inelastic scattering. *Phys. Rev.*, D67:012007, 2003.

[44] G. L. Bayatian et al. CMS Technical Design Report, Volume II: Physics Performance. *J. Phys.*, G34:995–1579, 2007.

[45] John C. Collins. Light cone variables, rapidity and all that. 1997.

[46] J. M. Butterworth, J. P. Couchman, B. E. Cox, and B. M. Waugh. KtJet: A C++ implementation of the K(T) clustering algorithm. *Comput. Phys. Commun.*, 153:85–96, 2003.

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

[48] Gerald C. Blazey et al. Run II jet physics. 2000.

[49] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007.

[50] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 05:026, 2006.

[51] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. B. G. Teubner Stuttgart Leipzig, 1998.

[52] F. Abe et al. Evidence for color coherence in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$TeV. *Phys. Rev. D*, 50(9):5562–5579, Nov 1994.

[53] Leif Lönnblad. Matching and merging Matrix Elements with Parton Showers, 2008. Lecture series given at the CTEQ-MCnet summer school in Debrecen, Hungary.

[54] Leif Lonnblad. ARIADNE version 4: A Program for simulation of QCD cascades implementing the color dipole model. *Comput. Phys. Commun.*, 71:15–31, 1992.

[55] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. *JHEP*, 06:029, 2002.

[56] Stefano Frixione, Paolo Nason, and Bryan R. Webber. Matching NLO QCD and parton showers in heavy flavour production. *JHEP*, 08:007, 2003.

[57] Johan Alwall et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473–500, 2008.

[58] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, and Michele Treccani. Matching matrix elements and shower evolution for top- quark production in hadronic collisions. *JHEP*, 01:013, 2007.

[59] 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.*, 67:465–508, 1992.

[60] G. Corcella et al. HERWIG 6.5: an event generator for Hadron Emission Reactions With Interfering Gluons (including supersymmetric processes). *JHEP*, 01:010, 2001.

[61] J. M. Butterworth, Jeffrey R. Forshaw, and M. H. Seymour. Multiparton interactions in photoproduction at HERA. *Z. Phys.*, C72:637–646, 1996.

[62] J. M. Butterworth and M. H. Seymour. *Jimmy4: Multiparton Interactions in Herwig for the LHC*, 2005. Draft. `http://projects.hepforge.org/jimmy`. Last updated: November 2005. Accessed: July 2009.

[63] S. Alekhin et al. HERA and the LHC - A workshop on the implications of HERA for LHC physics: Proceedings Part B. pages 192–217, 2005.

[64] Borut Kersevan. MC Generators for LHC at ATLAS, 2007. Presentation at Generator Services monthly meeting. `http://indico.cern.ch/materialDisplay.py?contribId=1&materialId=slides&%confId=15961`.

[65] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.

[66] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.

[67] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.

[68] Stefano Frixione and Bryan R. Webber. The MC@NLO 3.2 event generator. 2006.

[69] R. D. Field. Studying the underlying event at CDF. Presented at 33rd International Conference on High Energy Physics (ICHEP 06), Moscow, Russia, 26 Jul - 2 Aug 2006.

[70] Peter Skands and Daniel Wicke. Non-perturbative QCD effects and the top mass at the Tevatron. *Eur. Phys. J.*, C52:133–140, 2007.

[71] Keith Hamilton, Peter Richardson, and Jon Tully. A Positive-Weight Next-to-Leading Order Monte Carlo Simulation for Higgs Boson Production. *JHEP*, 04:116, 2009.

[72] G. L. Bayatian et al. CMS physics: Technical Design Report. CERN-LHCC-2006-001.

[73] G. Acquistapace et al. CMS, The Magnet Project: Technical Design Report. CERN-LHCC-97-10.

[74] W. R. Leo. Techniques For Nuclear And Particle Physics Experiments: A How To Approach. Berlin, Germany: Springer (1987) 368 p.

[75] CMS, Tracker Technical Design Report. CERN-LHCC-98-06.

[76] Addendum to the CMS tracker TDR. CERN-LHCC-2000-016.

[77] CMS: The Electromagnetic Calorimeter. Technical Design Report. CERN-LHCC-97-33.

[78] CMS: The Hadron Calorimeter Technical Design Report. CERN-LHCC-97-31.

[79] CMS, The Compact Muon Solenoid. Muon Technical Design Report. CERN-LHCC-97-32.

[80] S. Dasu et al. CMS. The TriDAS Project. Technical Design Report, Vol. 1: The Trigger Systems. CERN-LHCC-2000-038.

[81] (ed.) Sphicas, P. CMS: The TriDAS Project. Technical Design Report, Vol. 2: Data Acquisition And High-level Trigger. CERN-LHCC-2002-026.

[82] Greg Welch and Gary Bishop. An Introduction to the Kalman Filter. Technical report, Chapel Hill, NC, USA, 1995.

[83] G. Abbiendi et al. Muon Reconstruction in the CMS Detector. 2008. CMS AN-2008/097.

[84] Christoph Rosemann. Measurement of Top Quark Properties from Pair Production and Decay with the CMS Detector. 2009. DESY-THESIS-2009-006.

[85] F. James and M. Roos. Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations. *Comput. Phys. Commun.*, 10:343–367, 1975.

[86] Rene Brun et al. ROOT, An Object-Oriented Data Analysis Framework. 2008. Version 5.18. `http://root.cern.ch`.

[87] A. Rizzi, F. Pall, and G. Segneri. Track impact parameter based b-tagging with CMS. 2006. CMS Note 2006/019.

[88] S. Esen et al. $\not\!\!E_T$ Performance in CMS. 2007. CMS An-2007/041.

[89] C. D. Jones, L. Lista, and B. Hegner. Analysis environments for CMS. *J. Phys. Conf. Ser.*, 119:032027, 2008.

[90] CSA07 Production, 2007. `https://twiki.cern.ch/twiki/bin/view/CMS/GeneratorProduction2007CSA07`. Last updated: December 2007. Accessed: July 2009.

[91] ALPGEN Production in Summer07, 2007. `https://twiki.cern.ch/twiki/bin/view/Main/AlpgenSummer07`. Last updated: January 2009. Accessed: July 2009.

[92] Verena Klose and Jan Erik Sundermann. KinFitter - A Kinematic Fit With Constraints. 2004. BABAR Analysis Document #1061.

[93] P. C. Hansen. The L-curve and its use in the numerical treatment of inverse problems. *Computational Inverse Problems in Electrocardiology*, 2000. http://www.imm.dtu.dk/ pch/TR/Lcurve.ps.

[94] Stefan Schmitt. TUnfold, 2009. `http://www.desy.de/~sschmitt`.

[95] Stephen Mrenna. Private communication, 2008.

[96] Benedikt Hegner (CMS Physics Tools Convener). Private communication, 2009.

Acknowledgements

Finally I would like to thank all the people who supported me while I worked on this thesis:

- Joachim Mnich for offering me the opportunity to work on this interesting topic and for his support during the last years.
- Hannes Jung for many helpful discussions and explanations. He taught me a lot about Monte Carlo generators and physics in general.
- Christoph Rosemann for his support in many different fields and a lot of very fruitful discussions. I learned a lot from him especially about selection and event reconstruction strategies. His help was essential and without it the detector study would have never reached the current status.
- Benedikt Hegner for his mental support, many interesting and productive discussions and his help in questions concerning the CMS software.
- Peter Richardson, Mike Seymour and Peter Skands who supervised me during my MCnet short term studentship at CERN. In many interesting discussions I learned a lot concerning Monte Carlo generators and the underlying physics. Big parts of the work presented in chapter 5 were achieved with their support.
- Sven Meyer for carefully reading this manuscript.
- Last but not least my parents who supported me over the whole time. There is no way to express how much I owe to them.