The Underlying Event in Proton-Proton Collisions

Dissertation
zur Erlangung des Doktorgrades
des Department Physik
der Universität Hamburg

vorgelegt von
Florian Bechtel
aus Speyer

Hamburg 2009

Gutachter der Dissertation: Prof. Dr. Peter Schleper
Prof. Dr. Joachim Mnich

Gutachter der Disputation: Prof. Dr. Peter Schleper
JProf. Dr. Johannes Haller

Datum der Disputation: 15. Mai 2009

Vorsitzender des Prüfungsausschusses: Dr. Georg Steinbrück

Vorsitzender des Promotionsausschusses: Prof. Dr. Robert Klanner

Dekan der MIN-Fakultät: Prof. Dr. Heinrich H. Graener

Abstract

In this thesis, studies of the underlying event in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 10$ TeV are presented. Crucial ingredient to underlying event models are multiple parton-parton scatters in single proton-proton collisions. The feasibility of measuring the underlying event was investigated with the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC) using charged particles and charged-particle jets.

Systematic uncertainties of the underlying event measurement due to detector misalignment and imperfect track reconstruction are found to be negligible after $\int \mathcal{L}\mathrm{d}t = 1\ \mathrm{pb}^{-1}$ of data are available. Different model predictions are compared with each other using fully simulated Monte Carlo samples. It is found, that distinct models differ strongly enough to tell them apart with early data.

Zusammenfassung

In dieser Arbeit werden Studien des *Underlying Event* in Proton-Proton-Stößen bei einer Schwerpunktsenergie von $\sqrt{s} = 10$ TeV vorgestellt. Entscheidender Bestandteil von Underlying-Event-Modellen sind mehrfache Parton-Parton-Stöße in einzelnen Proton-Proton-Kollisionen. Die Machbarkeit einer Messung des Underlying Event wurde untersucht mit dem *Compact Muon Solenoid* Detektor am *Large Hadron Collider* unter Benutzung von geladenen Teilchen und Jets geladener Teilchen.

Systematische Unsicherheiten der Underlying-Event-Messung aufgrund fehlerhafter Detektor-Alignierung und fehlerhafter Spurrekonstruktion stellen sich als vernachlässigbar heraus, sobald $\int \mathcal{L} \mathrm{d}t = 1\ \mathrm{pb}^{-1}$ an Daten verfügbar sind. Verschiedene Modellvorhersagen werden verglichen anhand vollständig simulierter Monte-Carlo-Datensätze. Es zeigt sich, daß der Unterschied zwischen verschiedenen Modellvorhersagen groß genug ist, um diese mit den ersten Daten voneinander abzugrenzen.

Contents

Introduction

The *Standard Model of particle physics* is a theory of electroweak [1,2,3] and strong interactions [4] and of all known elementary particles taking part in these interactions. Electroweak and strong interactions are described by a gauge theory with the gauge group $SU(3) \times SU(2) \times U(1)$. Up to energies of the order of 100 GeV, experimental tests of the Standard Model are consistent with its predictions.

According to the Standard Model, electroweak symmetry breaking is mediated by a $SU(2)$-doublet, the Higgs-field, which, after symmetry breaking, produces a new scalar particle - the *Higgs boson* [5,6,7,8]. To this date, no experimental evidence for the Higgs boson has been found [9]: A detailed study of the TeV-scale is expected to unravel the nature of electroweak symmetry breaking and prove or refute the existence of the Standard-Model-Higgs-boson.

At any rate, the Standard Model cannot be a complete theory of fundamental interactions: Neutrinos have been observed to oscillate [9], the nature of dark matter is unknown [10] and there is no Standard-Model explanation for the observed asymmetry in the universe's matter and antimatter content. Extensions to the Standard Model, which explain these observations, typically involve new particles and new interactions. Moreover, the Standard Model falls short of accounting for fundamental questions: How and why do fermion masses and their mixings come about? Why are three gauge coupling constants needed to describe electroweak and strong interactions? And why does no Standard-Model-mechanism protect the Higgs boson mass from enormous radiative corrections (*gauge hierarchy problem*) [11,12,13]? Among the suggestions to cure the last point are models like dynamical electroweak symmetry breaking à la Technicolor [14,15,16,17,18], low-energy supersymmetry [19,20,21], extra dimensions [22,23,24,25,26] and "little Higgs solutions" [27]. While these models avoid huge radiative corrections to the Higgs mass, every one of them predicts new phenomena to manifest themselves at the TeV-scale.

Thus, a wealth of reasons exists why the TeV-scale should be investigated. A proton-proton collider as the LHC with a beam energy of 7 TeV at a design luminosity of $\mathcal{L} = 10^{34}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ meets all requirements to study new energy regions. This energy will be seven times the center-of-mass energy of the Tevatron. What's more, its luminosity is one-hundred times higher than what has been achieved with other hadron colliders: Accordingly, accelerator and detectors had to be designed with great care to withstand the extreme conditions under nominal operation.

As explained in the following, every physics study at the LHC is in dire need to thoroughly understand all features of "ordinary" QCD processes, i. e. minimum-bias events and jet production, in order to efficiently select final states of interest. This understanding involves to not only investigate the hard scatter, but to also measure and interpret the *underlying event* - the soft component of a collision which adds further particles to the overall event activity.

The detection of almost any physics object is influenced by the underlying event: Jets are clustered differently with additional particles in the event. Jet energies are modified if particles from the underlying event are clustered to the jet. Isolation criteria may be spoiled by the underlying event activity. Eventually, the underlying event also influences the measurement of charged particles. Additional charged particles raise the radiation-level imposed on the CMS detector, in particular on the inner tracking system, leading to a decrease in lifetime of the silicon sensors. Furthermore, charged particles produce additional hits in the CMS tracking detectors, thereby aggravating the identification of trajectories and impact parameters. In short, every physics analysis at the LHC is to some extent affected by the underlying event. It is thus crucial to determine the level of underlying event activity and tune event generators as early as possible after LHC startup.

This thesis describes strategies to characterize the underlying event in proton-proton collisions measured with the Compact Muon Solenoid (CMS) detector at the LHC. In particular, the feasibility of constraining existing models with early collision data is studied. First, the relevant aspects of QCD that underly all cross section predictions for the LHC are summarized (Chap. 1). Chapter 2 introduces two popular models describing the underlying event at the Tevatron. Next, the relevant design and performance parameters of the LHC and the CMS detector are summarized along with reconstruction methods for the key objects used in the underlying event analysis (Chap. 3). Computing is a major challenge due to the huge data sets expected from the LHC. A considerable part of this work was thus devoted to computing issues. Detector-level studies are carried out using large sets of fully simulated Monte Carlo samples. Generation and simulation settings are described in Chap. 4. Observables suitable for the characterization of key parameters of the underlying event and its components are compared in Chap. 5. The sensitivity to model parameters is studied with hadron-level quantities. In Chapter 6, we investigate the reconstruction performance of charged particles and charged-particle jets with the CMS inner tracking system. Jets clustered from tracks are particularly interesting when analyzing early LHC runs, where the CMS hadron calorimeter is not yet calibrated to sufficient precision. In addition, at low transverse momenta, the CMS tracking system performs better than the CMS calorimeters and is thus better suited to study the soft physics aspects that dominate the underlying event.

Chapter 7 picks up the discussion of observables suitable to characterize the underlying event: Different models are studied using the full detector simulation under two alignment conditions. Emphasis is put on anticipating experimental uncertainties. We conclude by recommending a list of observables to be measured at LHC startup and by proposing additional observables to comprehend the nature of proton-proton collisions at the LHC (Chap. 8).

Chapter 1

Hard Interactions of Quarks and Gluons

Quantum Chromodynamics (QCD) is the underlying theory of hard and soft high-energy hadron collisions. Rates of *hard* processes and event properties are predicted by perturbation theory to good precision. Yet, perturbation theory fails to predict *soft* process rates and event properties such that phenomenological models with parameters adjusted to existing data are employed. As soft interactions accompany many hard processes, precision measurements require a sound understanding of both scales. The perturbative framework for the calculation of hard-scattering processes has been reviewed in [28] and is summarized here.

1.1 Hard-scattering formalism and the QCD factorization theorem

The QCD factorization theorem rests on extending parton-model concepts (originally developed in the context of deep-inelastic scattering) to hadron-hadron collisions [29]. It is illustrated here for the *Drell-Yan-process* (depicted in Fig. 1.1), i. e. muon pair production by quark-antiquark annihilation. The QCD factorization theorem states that the hadronic cross section σ_{AB} is a convolution of the partonic cross section $\hat{\sigma}(q\bar{q} \to \mu^+\mu^-)$ with the parton distribution functions $f_{q/A}(x)$ (f is the probability density to find a quark q with momentum fraction x with respect to the quark's mother hadron A)

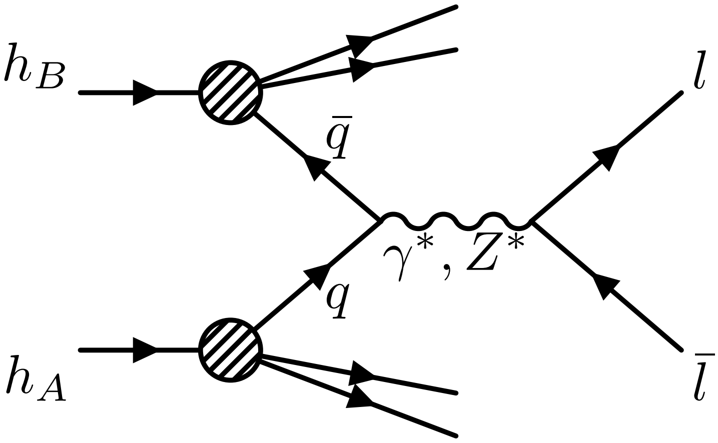


Figure 1.1: Drell-Yan process: a quark of one hadron and an antiquark of another hadron annihilate and create a pair of oppositely-charged leptons. The production of such lepton pairs proceeds through the exchange of a virtual photon or Z-boson.

measured in deep-inelastic scattering processes:

$$\sigma_{AB} = \int \mathrm{d}x_q \mathrm{d}x_{\bar{q}} f_{q/A}(x_q) f_{\bar{q}/B}(x_{\bar{q}}) \hat{\sigma}(q\bar{q} \to \mu^+\mu^-) \; . \tag{1.1}$$

This decomposition of the hadronic cross section is valid in the asymptotic *scaling* limit, i. e. with the center-of-mass energy $\sqrt{s}$ growing to infinity with a fixed ratio $M/\sqrt{s}$ of muon pair invariant mass M and center-of-mass energy $\sqrt{s}$.

Perturbative corrections for collinear gluon emission contribute large logarithmic terms that also arise in the calculation of structure functions in deep-inelastic scattering. The corrections can thus be absorbed in the definition of the parton density functions leading to logarithmical scaling violations. This key observation leads to the formulation of *factorization theorems* for generic hard-scattering processes [30]. Including leading logarithmic corrections, the hadronic cross section for Drell-Yan muon pair production now reads

$$\sigma_{AB} = \int \mathrm{d}x_q \mathrm{d}x_{\bar{q}} f_{q/A}(x_q, Q^2) f_{\bar{q}/B}(x_{\bar{q}}, Q^2) \hat{\sigma}(q\bar{q} \to \mu^+\mu^-) \; , \tag{1.2}$$

with Q^2 being a large momentum scale suitable to characterize the hard scattering processes, e. g. the muon pair invariant mass.

Perturbative $\mathcal{O}(\alpha_s^n)$ corrections to the leading-logarithmic cross section 1.2 are process-dependent:

$$\sigma_{AB} = \int \mathrm{d}x_q \mathrm{d}x_{\bar{q}} f_{q/A}(x_q, \mu_F^2) f_{\bar{q}/B}(x_{\bar{q}}, \mu_F^2) \left\{ \hat{\sigma}_0 + \alpha_s(\mu_R^2)\hat{\sigma}_1 + \ldots \right\}_{q\bar{q}\to\mu^+\mu^-} \quad . \tag{1.3}$$

Two scales enter the formula: The *factorization scale* μ_F separates long- and short-distance physics. The strong coupling α_s is evaluated at the *renormalization scale* μ_R. The scale dependence of all coefficients is compensated by the scale dependence of the parton distributions when calculating the cross section to all orders in perturbation theory. If higher-order corrections have not been calculated, the cross section prediction depends on the specific choices of μ_R and μ_F. Typically, the scales are chosen at the order of magnitude of momentum scales of the hard process. Often, both scales are even assumed to be equal. E. g. the standard scale choice for the Drell-Yan process is $\mu_F = \mu_R = M$, the muon pair invariant mass [28]. Similarly, appropriate scale choices to predict the production cross sections for Z-bosons, top quarks and large E_T jets are the invariant masses of Z-boson and top quark or the jet transverse energy [28].

For $\log(\mu) \gg \log(1/x)$, parton distributions used in hard-scattering cross-section calculations satisfy the *DGLAP equations* [31, 32, 33, 34] prescribing how the parton distributions evolve with Q^2:

$$\begin{aligned} \frac{\partial q_i(x,\mu^2)}{\partial \log \mu^2} &= \frac{\alpha_s}{2\pi} \int_x^1 \frac{\mathrm{d}z}{z} \left\{ P_{q_i q_j}(z,\alpha_s) q_j(\frac{x}{z},\mu^2) + P_{q_i g}(z,\alpha_s) g(\frac{x}{z},\mu^2) \right\} \quad , \\ \frac{\partial g(x,\mu^2)}{\partial \log \mu^2} &= \frac{\alpha_s}{2\pi} \int_x^1 \frac{\mathrm{d}z}{z} \left\{ P_{g q_j}(z,\alpha_s) q_j(\frac{x}{z},\mu^2) + P_{gg}(z,\alpha_s) g(\frac{x}{z},\mu^2) \right\} \quad . \end{aligned} \tag{1.4}$$

The *splitting functions* $P_{p'p}(z, \alpha_s)$ represent the probability that a parton of type p radiates a quark or gluon and becomes a parton of type p' carrying fraction z of the momentum of parton p. Perturbative expansions for the splitting functions exist.

The above procedure has been employed to calculate Standard Model cross sections and event rates at $p\bar{p}$ and pp colliders at next-to-leading order in perturbation theory. Results are shown in Fig. 1.2.

The dependence on Q^2 of the parton distributions can be determined via the DGLAP equations; the x dependence on the other hand is obtained from fitting experimental data (from deep-inelastic and other hard-scattering processes). Given a final state with invariant mass M and rapidity y and assuming the factorization scale Q^2 to be equal to M, the accessible (x, Q^2)-region is shown in Fig. 1.3. The center-of-mass energy is the LHC

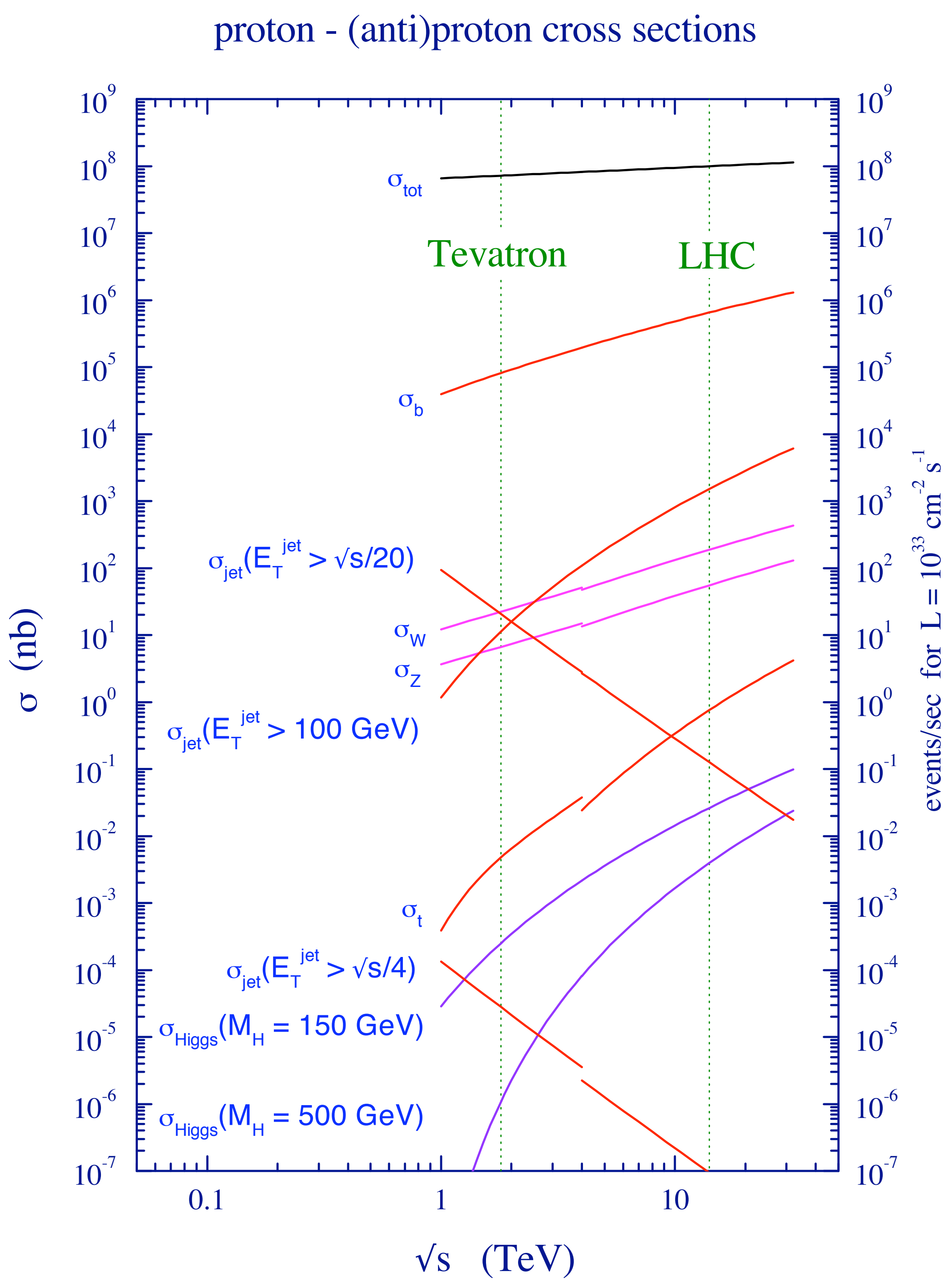


Figure 1.2: Standard model cross sections at the Tevatron and LHC colliders, calculated at next-to-leading order in perturbation theory with the parton distribution parametrization MRS98. (Lines are split at the transition from $p\bar{p}$ to pp cross sections.) Figure taken from [35].

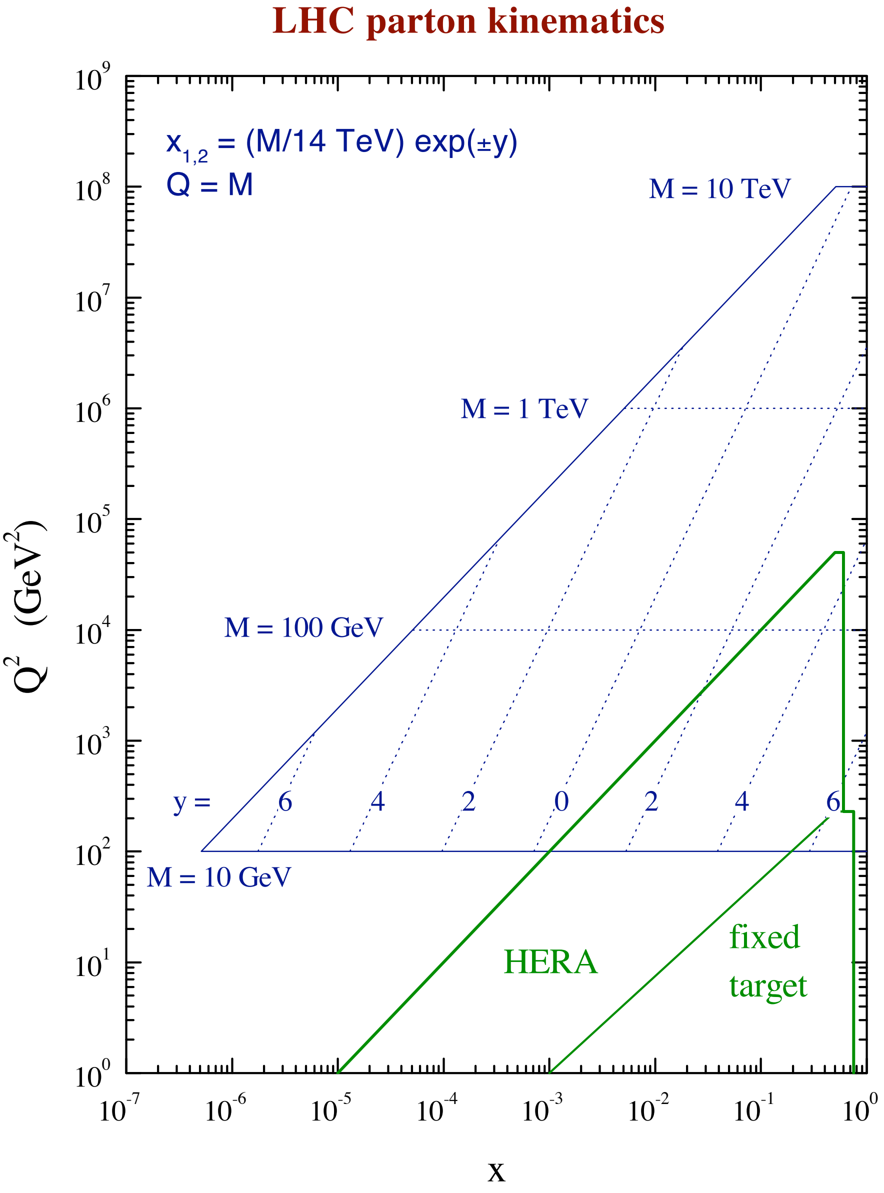


Figure 1.3: (x, Q^2)-plane shown together with kinematical variables corresponding to a final state of invariant mass M and rapidity y. Green lines indicate the range of former colliders. Figure taken from [35].

collision energy $\sqrt{s} = 14$ TeV and any given rapidity corresponds to one line for x_1 and one line for x_2. For instance, consider the production of a final state of invariant mass $M = 100$ GeV and rapidity $y = 2$. This proceeds via the scattering of two partons with momentum fractions $x_1 = 0.05$ and $x_2 = 0.001$ at virtuality $Q^2 = 10^4$ GeV^2.

1.2 Partonic cross sections

Calculating the hard process to lowest order (LO) in perturbation theory gives a simple expression of an observable under study. Such an approximation includes calculating the squared matrix element and subsequently integrating the squared matrix element over the available phase space. Often, this integration has to be performed numerically. Also, the matrix element may be divergent in some regions of phase space that hence have to be avoided.

Next-to-leading order calculations

In general, broad features of a given process are described by lowest-order calculations. As lowest-order calculations typically exhibit a strong dependence on the renormalization and factorization scales used, this approximation is often not sufficient to compare with experimental data. Furthermore, possible large logarithms may need to be resummed and additional partonic subprocesses may contribute. Consequently, next-to-leading order (NLO) calculations are mandatory ingredients of experimental analyses at the LHC.

Virtual and real radiation

Calculating QCD observables to next-to-leading order in perturbation theory involves studying all Feynman diagrams that add an extra factor α_s to the original diagram: Further quarks and gluons are included in the lowest-order diagrams. The new contributions can be *virtual*, corresponding to internal lines, or *real*, corresponding to external lines.

Virtual corrections contain infrared divergences that cancel against infrared singularities in real emissions. This is true at all orders in perturbation theory irrespective of the number of final-state particles [36, 37, 38]. Real corrections are divergent for unresolved gluons, i. e. gluons collinear to one of the quarks or soft gluons. A common method to render the calculation finite is *dimensional regularization*: Calculations are performed in $4-2\epsilon$-dimensional space ($\epsilon < 0$) and singularities in four dimensions turn into single and double poles in ϵ. The limit $\epsilon \to 0$ is taken after the divergences have cancelled.

Including real emission diagrams in a next-to-leading order calculation can account for an extended range of predictions compared to the lowest-order calculation. For instance, the gauge boson's transverse momentum in Drell-Yan processes can only be non-zero if the gauge boson is balanced by one or several partons emitted at next to leading order.

Scale dependence

When calculated to all orders in perturbation theory, physical observables must not depend on the renormalization (μ_R) and factorization (μ_F) scale assumed. A generic feature of higher order calculations is that observables calculated to order α_s^n depend on μ_F or μ_R up to order α_s^{n+1} [28]. Popular scale choices are e. g. scales with small NLO corrections to the LO cross section. Alternatively, typical momenta, energies or masses specific to the process under study are used, e. g. the W mass in the case of W production. The range of predictions corresponding to different scale choices is usually attributed to *theoretical uncertainties*.

The NLO K-factor

The ratio of NLO to LO cross section (inclusive or differential) is referred to as the *K-factor* for a given process. The K-factor summarizes the order of NLO corrections to the leading-order cross section and depends in general on the phase space region under study: The ratio between NLO and LO prediction can change significantly when studying e. g. the total cross section on the one hand and stringently selected events on the other hand. Also note that the K-factor is different when evaluated at different renormalization and factorization scales. Typically, LO parton density parametrizations are used to evaluate LO cross sections; NLO parton density parametrizations are used to evaluate NLO cross sections. These differences in the treatment of LO and NLO observables lead to the K-factor possibly being less than, equal to or greater than one.

All-order approaches: Parton Showers

Parton showers are one example of an *all-order* approach to describe high-energy collisions. They are a common tool used in many physics analysis e. g. via the implementation in the programs Pythia [39], Herwig [40, 41], and Sherpa [42]. Using the DGLAP formalism, these parton showers relate few partons from a hard interaction to partons at virtualities of the order of Λ_{QCD}. At soft scales, non-perturbative models are employed to hadronize the partonic final state. The probability to evolve from higher to lower scales without emitting a gluon is given by the *Sudakov form factor*.

The parton showering process consists of generating an evolution variable t, along with the kinematics and flavors of the emitted partons. Different showers employ different evolution variables t: Sherpa and older versions of Pythia use the parent parton's virtuality, Herwig uses $t = E^2(1 - \cos\theta)$, with E the parent parton's energy and θ the opening angle between both partons. Current versions of Pythia evolve the square of the transverse momentum of both partons in the splitting.

The Sudakov form factor corresponds to a resummation of soft and collinear gluon effects: Predictions become well-defined. Yet, non-singular contributions from large energy, wide angle gluon emission are not taken into account.

Merging parton showers and fixed order calculations

Parton showers excellently describe phase space regions dominated by soft and collinear gluon emission, but ignore large energy, wide angle gluon emission. Matrix element calculations on the other hand describe high-energetic partons emitted under large angles, including interference effects between amplitudes that yield the same external partons. However, matrix element calculations do not resolve soft and collinear gluon emissions. A combination of parton showers and matrix element calculations that makes best use of each method's strengths is thus desirable.

Note however that a transition region in phase space exists for that parton showers and matrix element calculations both return non-zero predictions: Double-counting is avoided by following a *merging prescription* such as e. g. the CKKW technique [43]. The CKKW technique employs matrix elements to calculate parton branchings at large angles or at large energies (or both) and employs parton showers to calculate small angle, low energy emissions: A "resolution scale" parameter (typically a virtuality or energy) divides the phase space for parton emission into matrix-element-dominated and parton-shower-dominated regions.

Parton-level calculations are typically compared with particle measurements by utilizing *jet algorithms*. Jet algorithms combine collimated beams of particles into clusters called *jets*. Jet algorithms should ideally yield the same jet objects, regardless of whether they were run on parton level or on particle level. A jet algorithm consists of two essential ingredients: a prescription to decide which input objects belong to a given jet (clustering scheme) and a prescription to combine selected objects into one four-vector (recombination scheme). Typical variables to associate objects to a jet are the input object's transverse momentum ($k_\perp$*-type algorithms* [44, 45]) and angle (*cone-type algorithms* [46]) relative to the jet axis.

1.3 Parton distribution functions

The previous section described methods to obtain partonic cross sections. These need to be convoluted with *parton distribution functions* to calculate production cross sections at hadron colliders. Parton distribution functions parametrize the momentum fraction x of a parton in a proton. They are obtained from global fits to deep-inelastic scattering data, Drell-Yan production and jet measurements. The most widely used sets of parton distribution functions are provided by the CTEQ [47] and MSTW [48] (formerly MRST) groups and by the HERA collaborations. Measurements at the same x but different Q^2 are compared by evolving the extracted parton distributions with the DGLAP equations. As mentioned in Sec. 1.1, the DGLAP equations are only applicable in the range $\log(\mu) \gg \log(1/x)$. Thus, at low x and Q^2, a different prescription such as the BFKL equations [49, 50, 51, 52] has to be used. However, no evidence for BFKL physics has been observed in experiment so far. Thus, the exclusive use of DGLAP equations for global fits to the available data seems justified.

Fig. 1.4 shows the CTEQ6.1M parton density parametrization for up-, down- and anti-up-quarks and for gluons evaluated at a virtuality $Q^2 = 100$ GeV^2. Note that the gluon distribution has been scaled down by a factor of 10. In fact, for momentum fractions less than $x = 0.01$, gluons are the dominant parton species in the proton. Valence quarks dominate at higher momentum fractions. Measurements at HERA delivered crucial input to constraining the gluon density at small momentum fractions [53, 54, 55, 56].

1.4 The underlying event in proton-proton interactions

Much of the complexity involved in describing proton-proton scatters derives from the composite nature of hadrons. Section 1.1 outlined how hadronic cross sections can be

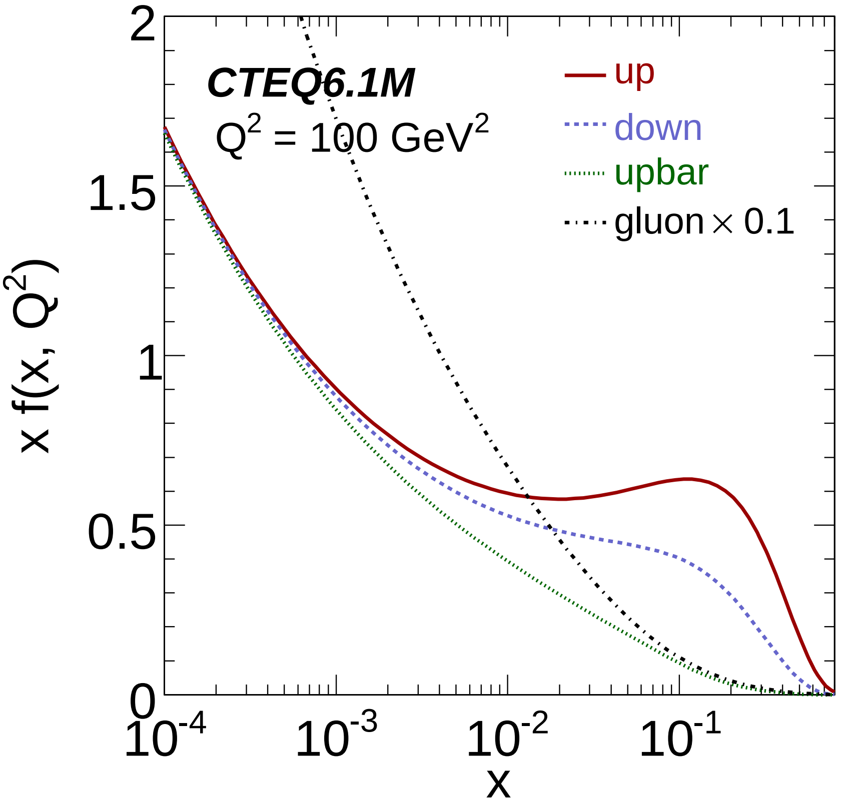


Figure 1.4: CTEQ6.1 parton distribution function parametrizations at $Q^2 = 100$ GeV^2 [47]. The gluon distribution has been scaled down by a factor of 10.

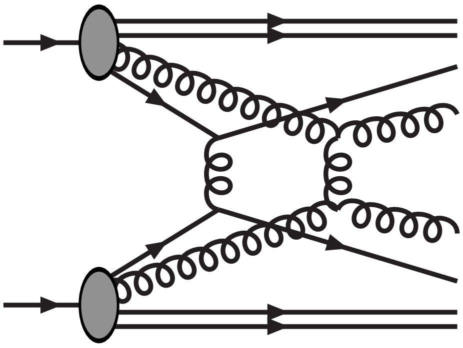


Figure 1.5: Schematic illustration of an event with two 2 → 2 perturbative interactions as simulated by an event generator. Figure taken from [57].

calculated with the help of the QCD factorization theorem: Parton-level cross sections are convoluted with parton distribution functions (determined from experimental data). However, in order to compare theory predictions with experimental data, additional steps are necessary. The set of final-state partons is transformed into a set of hadrons in the *fragmentation* step of the calculation; parton shower models simulate *gluon radiation in initial and final state*; extraction of partons undergoing hard scatters leaves behind a beam remnant carrying non-zero color quantum numbers; eventually, given the rise of parton densities towards low x, nothing prevents *additional parton-parton scatters* from taking place as depicted in Fig. 1.5.

Detailed prescriptions on how to calculate the hard scatter based on the QCD factorization theorem have been tested to immense precision. Hadronization is not yet understood from first principles - popular hadronization models are the *cluster fragmentation model* [58], implemented in the HERWIG program, and the *string fragmentation model* [59, 60], implemented in the PYTHIA program. These models have been tuned to describe data taken at the LEP and HERA colliders and their extrapolation to LHC energies are expected to be reliable. Parton shower models have been discussed in Sec. 1.2. Chap. 2 describes two multiple interaction models in use. These models have been tuned to Tevatron data. A significant uncertainty of LHC predictions stems from the unreliable extrapolation of these tunes to LHC energies.

The luminosity independent component of the pp scattered particle flow, that does not originate from the hard parton scatter, is referred to as the *underlying event*. Such a definition is likely to contain some ambiguity: In general, it is not possible to identify the "true" origin of a particle, since a hadronic final-state particle may have been the result of hadronizing a number partons that all have arisen from different sources. Alternatively, the underlying event in hadron-hadron collisions is associated with all particles produced by the hadron remnants "after" the hard scatter, where the notion of time-ordering is also poorly justified. It has become common practice to identify the underlying event depending on the process type under study: The underlying event in Drell-Yan muon pair production would simply be all event activity apart from the two final-state muons. By contrast, the underlying event definition in jet events is more complicated: Different approaches exist to associate particles in jet events either with the hard process or with the underlying event (cf. Sec. 5.1).

The underlying event in hadron-hadron collisions is only weakly constrained by existing data and comprises many aspects not describable by perturbative QCD. Traditionally, measurements of minimum-bias events were used to parametrize the underlying event [61]. Such a description eliminates by construction all interplay between hard process and underlying event. Correlations and fluctuations apparent in jet profiles [62, 63], jet pedestals [64, 65] and shifts in jet energies are not accounted for. In this context, perturbative QCD proved fruitful to describe some of these features. Measurements with the CDF experiment at the Tevatron collider characterized the underlying event by studying charged particle activity perpendicular to the leading jet. Only models including a description of several distinct parton-parton interactions seem to be able to describe the data [66].

Fig. 1.6 illustrates the complex variety of processes taking place in a single proton-proton collision. The hard process produces a W^+ boson that subsequently decays into a charm-quark and an antistrange-quark. The recoiled down-quark scatters on an initial state quark or antiquark from the opposite beam (top panel). Initial- and final-state radiation is added to the involved quarks and gluons (middle panel). On top of the hard process, a $gg \to gg$ scatter takes place with its initial- and final-state radiation (bottom panel). The hadronization step is not shown.

Every physics study at the LHC must thoroughly understand all features of "ordinary" QCD processes, i. e. minimum-bias events and jet production at LHC energies. Selection cuts are typically optimized in order to efficiently select interesting final states and suppress ordinary QCD events. Understanding common QCD processes involves more than to investigate the hard scatter: The *underlying event* - the soft component of a collision, adding further particles to the overall event activity - must be measured and understood.

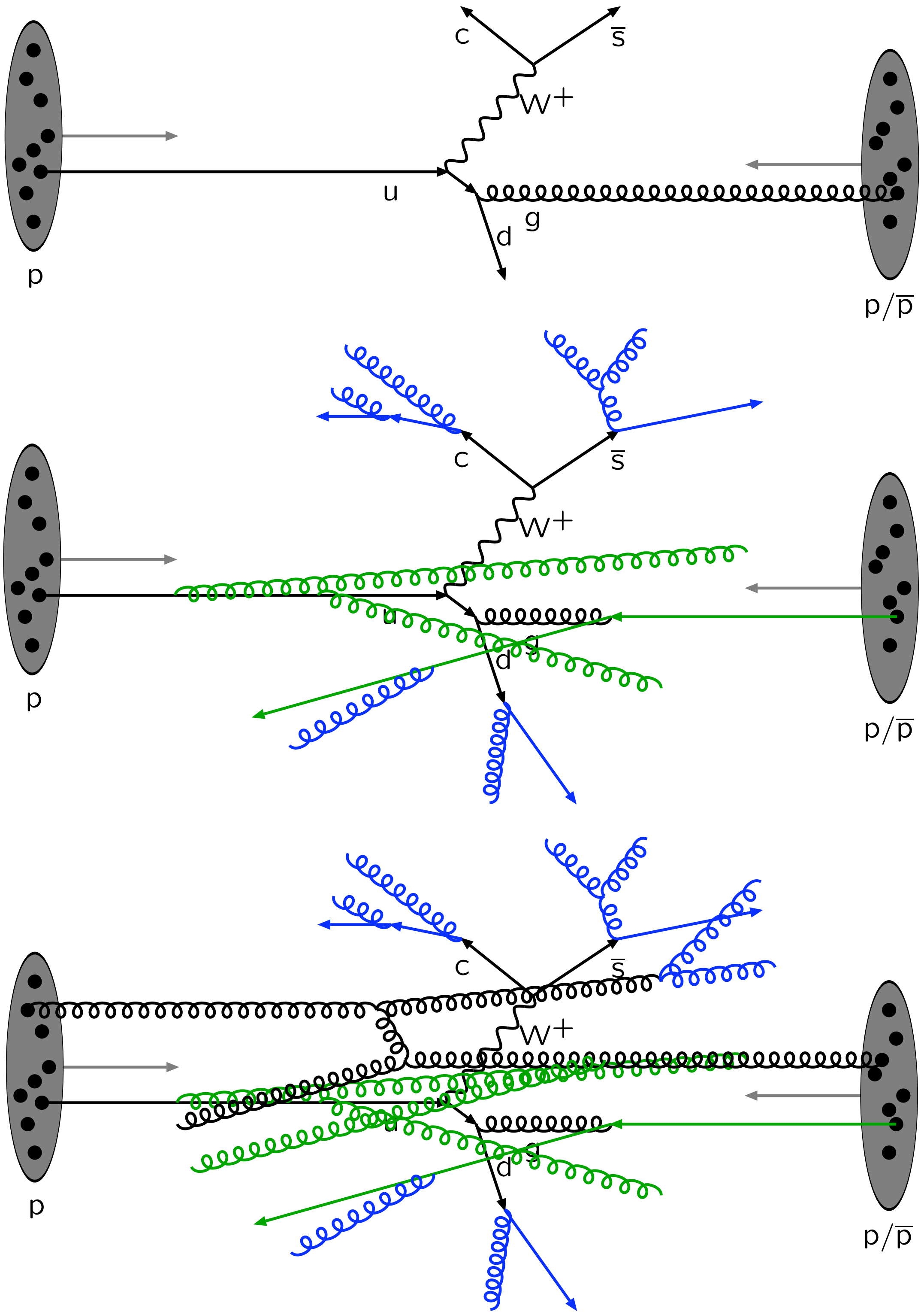


Figure 1.6: Sketch of a proton-proton collision at high energies: hard process (top), hard process with initial- and final-state radiation (center), hard process and second scatter, both with initial- and final-state radiation (bottom). Figures taken from [67].

Chapter 2

Multiple Parton-Parton Interactions in Proton-Proton Collisions

The least-well understood part in describing proton-proton scatters remains the treatment of beam remnants and multiple interactions. This chapter summarizes the main concepts that are used to simulate multiple parton-parton interactions in simulations of proton-proton collisions. Basic features of the perturbative description of multiple scatterings are given in Section 2.1. The remainder of the chapter explains the two most-widely used underlying event models, as implemented in the event generators PYTHIA (Section 2.2) and HERWIG (Section 2.3).

2.1 Basic Concepts of Modeling Multiple Parton-Parton Interactions

The key idea of the currently employed multiple interaction models is to assume QCD factorization (Sec. 1.1) to work not only for the hard process but also for additional scatters. Thus, the hadronic cross section for the second interaction remains of the form "partonic cross section convoluted with parton density". The basic concepts of this approach were initially formulated more than twenty years ago [68].

Interactions of the types $qq' \rightarrow qq'$, $q\bar{q} \rightarrow q'\bar{q}'$, $q\bar{q} \rightarrow gg$, $qg \rightarrow qg$, $gg \rightarrow gg$ and $gg \rightarrow q\bar{q}$ shall in the following be denoted as *QCD hard* $2 \rightarrow 2$ *processes*. The differential cross

Table 2.1: Differential cross sections for parton-parton scatters involving quarks, antiquarks and gluons as a function of the Mandelstam variables explained in the text (formulae taken from [69, 70]).

Process	Partonic cross section $\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\hat{t}}$
$qq' \rightarrow qq'$	$\frac{4}{9}\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}$
$qq \rightarrow qq$	$\frac{4}{9}\left(\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}+\frac{\hat{s}^2+\hat{t}^2}{\hat{u}^2}\right)-\frac{8}{27}\frac{\hat{s}^2}{\hat{u}\hat{t}}$
$q\bar{q} \rightarrow q'\bar{q}'$	$\frac{4}{9}\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}$
$q\bar{q} \rightarrow q\bar{q}$	$\frac{4}{9}\left(\frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}+\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}\right)-\frac{8}{27}\frac{\hat{u}^2}{\hat{s}\hat{t}}$
$q\bar{q} \rightarrow gg$	$\frac{32}{27}\frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}-\frac{8}{3}\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}$
$gg \rightarrow q\bar{q}$	$\frac{1}{6}\frac{\hat{t}^2+\hat{u}^2}{\hat{t}\hat{u}}-\frac{3}{8}\frac{\hat{t}^2+\hat{u}^2}{\hat{s}^2}$
$gq \rightarrow gq$	$-\frac{4}{9}\frac{\hat{s}^2+\hat{u}^2}{\hat{s}\hat{u}}+\frac{\hat{u}^2+\hat{s}^2}{\hat{t}^2}$
$gg \rightarrow gg$	$\frac{9}{2}\left(3-\frac{\hat{t}\hat{u}}{\hat{s}^2}-\frac{\hat{s}\hat{u}}{\hat{t}^2}-\frac{\hat{s}\hat{t}}{\hat{u}^2}\right)$

section for these QCD hard 2 → 2 processes with respect to the scale $p_\perp^2$ is (cf. Eq. 1.2):

$$\frac{\mathrm{d}\sigma_{\mathrm{int}}}{\mathrm{d}p_\perp^2} = \sum_{i,j,k,l} \int \mathrm{d}x_1 \int \mathrm{d}x_2 \int \mathrm{d}\hat{t} f_i(x_1, Q^2) f_j(x_2, Q^2) \frac{\mathrm{d}\hat{\sigma}_{ij\rightarrow kl}}{\mathrm{d}\hat{t}} \delta\left(p_\perp^2 - \frac{\hat{t}\hat{u}}{\hat{s}}\right) , \tag{2.1}$$

where x_1, x_2 are the partons' momentum fractions and $\hat{s}, \hat{t}, \hat{u}$ are the *Mandelstam* variables

$$\hat{s} = (p_1 + p_2)^2 = (p_3 + p_4)^2 = x_1 x_2 s , \tag{2.2}$$

$$\hat{t} = (p_1 - p_3)^2 = (p_2 - p_4)^2 , \tag{2.3}$$

$$\hat{u} = (p_1 - p_4)^2 = (p_2 - p_3)^2 , \tag{2.4}$$

with p_1, p_2 being the four-momenta of the incoming partons and p_3, p_4 being the four-momenta of the outgoing partons. In the following, it is assumed that the factorization scale is equal to the scatter's transverse momentum scale $p_\perp$, i. e. $\mu_F = p_\perp^2$.

Differential cross section formulae for parton-parton scatters are listed in Table 2.1. At small scattering angles, i. e. for $t \rightarrow 0$, the t-channel gluon exchange processes $qq' \rightarrow qq'$, $qg \rightarrow qg$ and $gg \rightarrow gg$ dominate the full matrix element [70]. For scatterings that are soft relative to $\hat{s}$, $|\hat{t}| \ll \hat{s}$, $|\hat{t}|$ can be approximated by $p_\perp^2$, as $p_\perp^2 = \hat{t}\hat{u}/\hat{s} \approx |\hat{t}|$. In this limit, the only difference between quark and gluon interactions are different color factors - in fact, the relative contributions from gg-, qg- and qq-scatters are

$$\hat{\sigma}_{gg} : \hat{\sigma}_{qg} : \hat{\sigma}_{qq} = \frac{9}{4} : 1 : \frac{4}{9} \ . \tag{2.5}$$

Hence 2.1 can be approximated by

$$\frac{\mathrm{d}\sigma_{\mathrm{int}}}{\mathrm{d}p_\perp^2} \approx \iint \frac{\mathrm{d}x_1}{x_1}\frac{\mathrm{d}x_2}{x_2} F(x_1, p_\perp^2) F(x_2, p_\perp^2) \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_\perp^2} \ , \tag{2.6}$$

where

$$\begin{aligned} \frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}p_\perp^2} &= \frac{8\pi\alpha_s^2(p_\perp^2)}{9p_\perp^4} \ , \\ F(x, Q^2) &= \sum_q \left(xq(x, Q^2) + x\bar{q}(x, Q^2) \right) + \frac{9}{4} xg(x, Q^2) \ . \end{aligned} \tag{2.7}$$

The integrated cross section thus can be expressed as a function of a minimal transverse momentum scale $p_{\perp\mathrm{min}}$ (for a constant α_s and neglecting the x integrations):

$$\sigma_{\mathrm{int}}(p_{\perp\mathrm{min}}) = \int_{p^2_{\perp\mathrm{min}}}^{(\sqrt{s}/2)^2} \frac{\mathrm{d}\sigma}{\mathrm{d}p_\perp^2} \mathrm{d}p_\perp^2 \propto \frac{1}{p^2_{\perp\mathrm{min}}} \xrightarrow{p_{\perp\mathrm{min}} \to 0} \infty \ . \tag{2.8}$$

In the limit $p_{\perp\mathrm{min}} \to 0$, the integrated cross section above a minimal $p_\perp$ diverges and will exceed the *total* proton-proton cross section [71] for $p_{\perp\mathrm{min}}$ of the order of a few GeV [57].

Two aspects can explain this phenomenon: First, σ_{int} counts *interactions*, but σ_{tot} counts *events*. An event containing two interactions counts twice in the interaction cross section but only once in the total cross section. Hence, the ratio between interaction cross section and total cross section is a measure of the average number of interactions per event which is perfectly allowed to be larger than unity:

$$\frac{\sigma_{\mathrm{int}}(p_{\perp\mathrm{min}})}{\sigma_{\mathrm{tot}}} = \langle n \rangle \left(p_{\perp\mathrm{min}} \right) \ . \tag{2.9}$$

Second, for decreasing transverse momenta, the average number of interactions is dampened by (nonperturbative) *color screening effects*. A decreasing transverse momentum of the exchanged gluon translates into an increasing transverse wavelength. If the transverse wavelength gets sufficiently large, individual color charges of the incoming parton bunch might not be resolved anymore such that the effective coupling decreases until the effective cutoff $p_{\perp\mathrm{min}}$ is reached. One choice of implementation of color screening is to suppress any scatter below the scale $p_{\perp\mathrm{min}}$. Alternatively, a better choice might be a dampened version [72], namely

$$p_{\perp \min}^2 \rightarrow p_{\perp 0}^2 + p_\perp^2 \ , \tag{2.10}$$

where $p_{\perp 0}$ has to be tuned to data. With this choice, the jet cross section, which is divergent like $\alpha_s^2(p_\perp^2)/p_\perp^4$ is smoothly regularized. Note that $p_{\perp \min}$ and $p_{\perp 0}$ do not have to be independent of energy as higher energies can probe parton densities at smaller values of Bjorken-x. Since parton densities grow with decreasing x-values, partons effectively become "denser packed" meaning that color charges are compensated on even shorter distances.

As we will see in the following, the modeling of multiple interactions when simulating proton-proton collisions makes use of several adjustable parameters. The set of parameters that describes a particular final state in proton-proton or proton-antiproton collisions is referred to as an *underlying event tune*. Underlying-event tunes differ in their choices of $p_{\perp \min}$ or $p_{\perp 0}$ and their energy dependence, impact-parameter dependence, correlations in momentum, flavor, and color, and in the choice of parton density parametrisations. For instance, Pythia Tune A [73] describes a large set of CDF minimum bias and jet data [66, 74]. This tune chooses a dampened cutoff for multiple interactions (cf. Eq. 2.10) $p_{\perp 0} = 2.0$ GeV at the reference energy 1.8 TeV with an *energy rescaling* proportional to $(\sqrt{s})^{1/4}$, i. e. at different collision energies, $p_{\perp 0}$ is calculated according to

$$p_{\perp 0}(\sqrt{s}) = 2.0 \text{ GeV} \times \left(\frac{\sqrt{s}}{1.8 \text{ TeV}} \right)^{1/4} \ , \tag{2.11}$$

corresponding to $p_{\perp 0} = 1.68$ GeV at $\sqrt{s} = 900$ GeV and to $p_{\perp 0} = 3.34$ GeV at $\sqrt{s} = 14$ TeV. The subsequent section summarizes relevant multiple interaction model ingredients for Pythia. The specific parameter choices for underlying-event tune A will be given as examples.

2.2 Simulation with Pythia 8

This section describes the multiple interactions physics scenario in Pythia 8 [39] based on earlier developments [68] implemented in versions 6.x of the Pythia program. Recent progress included a more careful study of flavor and color correlations, junction topologies and the relationship to beam remnants [57], and interleaving with initial- and final-state radiation (making use of transverse-momentum-ordered initial- and final-state showers [75]).

An unsolved issue is how the colors of all parton-parton scatters are correlated. For sure, there is a correlation coming from the color-singlet nature of the incoming beams, but in addition final-state color rearrangements may change the picture. Indeed such extra effects appear necessary to describe data, e. g. the rise of average track p_T with track multiplicity observed with the CDF experiment [76]. Within PYTHIA, a simple implementation of color rearrangement is found as part of the beam remnants description.

The number of parton-parton scatters in a proton-proton interaction is expected to be largest for a head-on collision of both hadrons; peripheral collisions are likely to only produce few parton-parton scatters (if any at all) in addition to the primary hard scatter. This feature is explored in more detail in the following.

Matter Overlap between Incoming Hadrons

PYTHIA characterizes different collisions by different impact parameters b. A small impact parameter is equivalent to a large overlap of the two protons' wave-functions, leading to a larger probability for additional parton-parton interactions. Likewise, a small proton-proton overlap (corresponding to a large impact parameter) decreases the probability for further scatterings to happen.

If one wants to quantify the effect of overlapping protons, some assumptions on the matter distribution ρ must be made. From this matter distribution, the spatial overlap of the two protons' wave-functions can be calculated: The time-integrated matter wave-function overlap $\mathcal{O}(b)$ of two hadrons with matter distributions ρ colliding with impact parameter b is

$$\mathcal{O}(b) \propto \int \mathrm{d}t \int \mathrm{d}^3x \rho(x,y,z)\rho(x+b,y,z+t) \ . \tag{2.12}$$

The default choice in PYTHIA is to distribute hadronic matter inside the proton according to a spherically-symmetric double-Gaussian:

$$\rho(r) \propto \frac{1-\beta}{a_1^3}\exp\left(-\frac{r^2}{a_1^2}\right) + \frac{\beta}{a_2^3}\exp\left(-\frac{r^2}{a_2^2}\right) \ , \tag{2.13}$$

with a_2 being the radius of a small core region of the matter distribution containing a fraction β of the total hadronic matter constituting the hadron of radius a_1. Within Pythia Tune A, these parameters are chosen to be $\beta = 0.5$ and $a_2/a_1 = 0.4$. The resulting wave-function overlap $\mathcal{O}(b)$ follows from Eq. 2.12:

$$\mathcal{O}(b) \propto \frac{(1-\beta)^2}{2a_1^2} \exp\left\{-\frac{b^2}{2a_1^2}\right\} + \frac{2\beta(1-\beta)}{a_1^2+a_2^2} \exp\left\{-\frac{b^2}{a_1^2+a_2^2}\right\} + \frac{\beta^2}{2a_2^2} \exp\left\{-\frac{b^2}{2a_2^2}\right\} . \tag{2.14}$$

The larger the overlap $\mathcal{O}(b)$, the more probable are parton-parton scatters between the incoming protons. In the following, it is assumed that the number of interactions $\tilde{n}$ for a given impact parameter b is distributed according to a Poissonian

$$\mathcal{P}(\tilde{n}) = \langle\tilde{n}\rangle^{\tilde{n}} \frac{\exp\left(-\langle\tilde{n}\rangle\right)}{\tilde{n}!} . \tag{2.15}$$

Note however that with the matter distribution $\rho(r)$ extending to infinity, this procedure possibly predicts events at very large impact parameters. Excluding these events from the simulation corresponds to asking for at least one semi-hard interaction per event. As a consequence, the distribution of number of interactions for a given impact parameter b is narrower than a Poissonian distribution. The subsequent section describes how Pythia models possible correlations in momentum and flavor between these interactions.

Parton Densities and Treatment of Momentum- and Flavor-Correlations

The major difficulty in simulating multiple partonic interactions is that several partons per proton participate in hard scatters. Therefore, possible correlations must be considered. In fact, modeling multiple partonic interactions requires multi-parton densities, i. e. joint probabilities to find n partons of flavors $\{f_i\}$ with proton momentum fractions $\{x_i\}$ when probed by interactions at scales $\{Q_i\}$. Unfortunately, no experimental information on multi-parton densities is available. As a consequence, the Pythia approach is to rescale one-parton distributions, as explained in the following.

Interactions are generated in an ordered sequence, starting with the largest $p_\perp$, continuing with the next-to-largest, all the way down to the smallest-allowed-$p_\perp$ interaction, where $p_{\perp\mathrm{min}}$ limits the $p_\perp$ evolution from below. For the primary hard scatter, smaller $p_\perp$ scales integrate out of the unknown fully correlated parton densities. Therefore, this interaction can be described by standard one-parton distributions. For further interactions, again, smaller $p_\perp$ scales are integrated out. However, correlations with harder scatters must be taken into account and are not integrated out. Pythia attempts to adapt a given parton distribution to the set of partons extracted from the hadron by modifying the standard parton density functions. These are the starting point in constructing parton distributions for the remnant hadronic object after one or several interactions have occurred. Parton densities are changed such that they still respect conservation of momentum and flavor.

The first step to achieve momentum conservation is to account for the momentum fraction x_i removed by interaction i from the hadron remnant. This can be implemented by evaluating parton densities not at x_i but at the rescaled value

$$x'_i = \frac{x_i}{1 - \sum_{j=1}^{i-1} x_j} , \tag{2.16}$$

which by construction ensures that not more than the energy available in the incoming beam is scattered.

Additional correlations enter by aiming for flavor conservation at all stages of event simulation. If for example a valence quark is scattered from an incoming hadron, the number of valence quarks in that hadron must be reduced. Thus, after n interactions, the corresponding valence quark distribution $q_{fn}(x, Q^2)$ of quark flavor f is given by

$$q_{fn}(x, Q^2) = \frac{N_{fn}}{N_f} \frac{1}{1 - \sum_{i=1}^{n} x_i} q_f \left(\frac{x}{\sum_{i=1}^{n} x_i}, Q^2 \right) , \tag{2.17}$$

where N_f is the original number of valence quarks in the hadron ($N_u = 2, N_d = 1$ for protons), q_f is the standard one-parton distribution and N_{fn} is the number of q_f valence quarks remaining after n scatters.

A sea quark kicked out of the proton leaves behind its antisea quark partner in the proton remnant (*companion quark*). Within the framework of perturbative QCD, sea quark q_s and companion quark q_c come from a gluon branching, such that the companion quark's momentum fraction x can be described with the help of the respective DGLAP splitting kernel $P_{q \to q_s q_c}$ (cf. Eq. 1.4):

$$q_c(x, x_s) \propto \int_0^1 g(y) P_{q \to q_s q_c}(z) \delta(x_s - zy) \mathrm{d}z , \tag{2.18}$$

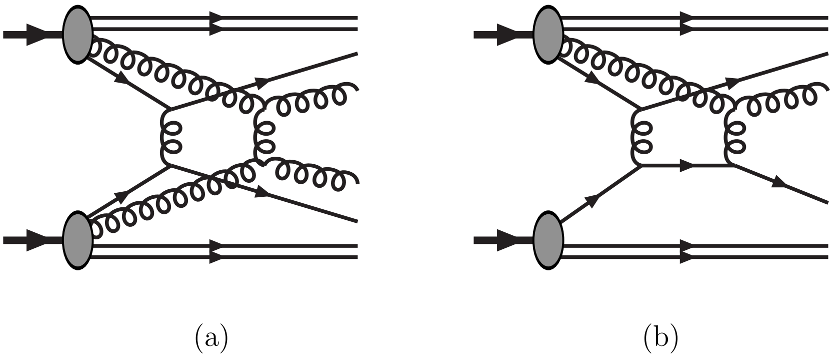

(a) (b)

Figure 2.1: Two distinct 2 → 2 scatters (a) and a 2 → 2 scatter with one outgoing parton entering a subsequent scatter (b) (rescattering). Figure taken from [77].

where y is the momentum fraction of the gluon g before splitting, and x_s is the momentum fraction of the sea quark. The normalization constant can be determined by imposing momentum sum rules.

Parton Rescattering

Parton-parton interactions subsequent to the hard interaction do not have to proceed between two partons different from the ones scattered before. Additional parton-parton interactions may also arise when a parton scatters more than once against partons from the other beam (*rescattering*). This concept is illustrated in Fig. 2.1: In the case of two distinct 2 → 2 scatters (a), correlations between both scatters only enter through the modified parton densities after extraction of the partons for previous interactions. In a rescattering process, both interactions are intimately coupled and non-trivial correlations arise.

Assuming factorised two-parton distributions, the overall influence of rescatters in proton-proton interactions was estimated to be small with respect to the influence of distinct 2 → 2 scatters [78]. Nevertheless, observables sensitive to multiple interaction effects may exhibit detectable changes after including rescattering in the multiple interaction framework: Beyond changes in multiplicity and transverse momenta, rescatters may affect properties of three-jet topologies in regions of phase space where initial- and final-state radiation are less dominant. Furthermore, a rescattered parton is likely to pick up additional transverse momentum, i. e. the average transverse momentum per event is expected to increase due to parton rescattering.

The key idea in simulating parton rescatters is to treat scattered parton and hadron remnant together as an incoming hadron with parton density

$$f(x,Q^2) \to \underbrace{f_{\mathrm{rescaled}}(x,Q^2)}_{\text{hadron remnant}} + \underbrace{\sum_n \delta(x-x_n)}_{\text{scattered parton(s)}} \quad . \tag{2.19}$$

After scattering, a parton is kept at a fixed momentum fraction x_n; the hadron remnant is still described by a continuous momentum density, albeit rescaled according to the extracted momenta, i. e. such that momentum conservation is ensured:

$$\int_0^1 \left[f_{\mathrm{rescaled}}(x,Q^2) + \sum_n \delta(x-x_n) \right] \mathrm{d}x = 1 \ . \tag{2.20}$$

In general, no well-defined prescription exists on how to ascribe a scattered parton to a particular hadron. Within Pythia's rescattering framework, scattered partons are assigned to hadrons based on their rapidity. The possibility of modeling parton rescatters is included in Pythia versions 8.114 and onwards.

Interleaving of Multiple Interactions and Parton Showers

We have described how multiple interactions are simulated in Pythia, following a decreasing $p_\perp$ evolution. For the i'th interaction, only correlations in momentum and flavor with respect to the $i-1$ preceding ones are taken into account. However, there are further possible correlations that have not yet been mentioned: Inital-state showers of one interaction may influence interactions at softer $p_\perp$ scales. Pythia thus evolves initial-state radiation (ISR), final-state radiation (FSR) and multiple interactions (MI) in one common decreasing $p_\perp$ sequence. The probability distribution for any kind of interaction at scale $p_\perp$ is composed from the various contributions:

$$\begin{aligned} \frac{\mathrm{d}\mathcal{P}}{\mathrm{d}p_\perp} = & \left(\frac{\mathrm{d}\mathcal{P}_{\mathrm{MI}}}{\mathrm{d}p_\perp} + \sum \frac{\mathrm{d}\mathcal{P}_{\mathrm{ISR}}}{\mathrm{d}p_\perp} + \sum \frac{\mathrm{d}\mathcal{P}_{\mathrm{FSR}}}{\mathrm{d}p_\perp} \right) \\ & \times \exp \left(- \int_{p_\perp}^{p_{\perp i-1}} \left(\frac{\mathrm{d}\mathcal{P}_{\mathrm{MI}}}{\mathrm{d}p'_\perp} + \sum \frac{\mathrm{d}\mathcal{P}_{\mathrm{ISR}}}{\mathrm{d}p'_\perp} + \sum \frac{\mathrm{d}\mathcal{P}_{\mathrm{FSR}}}{\mathrm{d}p'_\perp} + \right) \mathrm{d}p'_\perp \right) \quad , \end{aligned} \tag{2.21}$$

where the contribution from multiple interactions $\mathcal{P}_{\mathrm{MI}}$ is symbolic for separate $2 \to 2$ scatters as well as for parton rescatters.

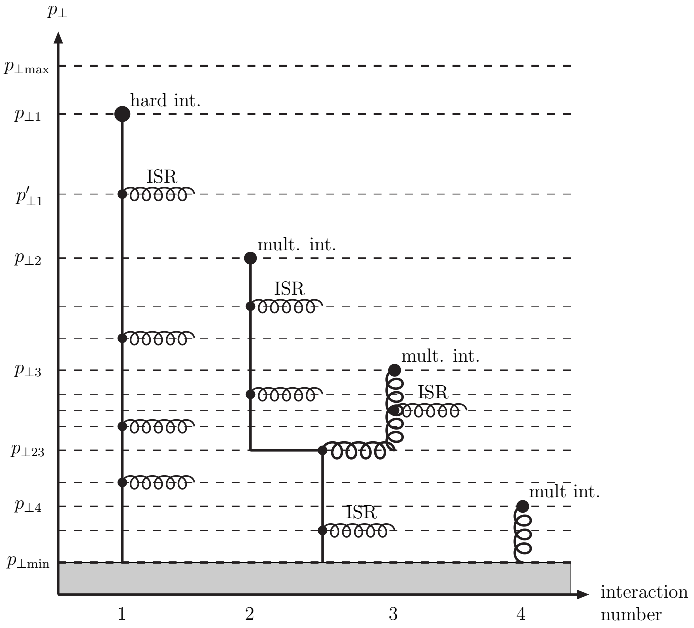


Figure 2.2: Schematic representation of the interleaved evolution of parton shower initiators in a hadron collision with 4 interactions, as modeled by Pythia. Figure taken from [75].

Figure 2.2 illustrates the interleaving of multiple interactions and parton showers. Four parton-parton scatters are shown together with their initial-state radiation by depicting the structure of one of the incoming hadrons. The downward $p_{\perp}$ evolution corresponds to reading the figure from top to bottom. An illustration of the full interaction would include a similar figure for the opposite hadron, with both figures joined at the scatters (full black circles in Fig. 2.2). The parton entering the first scatter at $p_{\perp} = p_{\perp 1}$ is further associated with four initial-state radiations, the first one happening at $p_{\perp} = p'_{\perp}$. Additional scatters occur at $p_{\perp} = p_{\perp 2}$, $p_{\perp} = p_{\perp 3}$ and $p_{\perp} = p_{\perp 4}$. Partons 2 and 3 originate from the same mother parton, but enter separate scatters with partons from the opposite hadron.

Summary: Multiple Interactions in Pythia 8

Pythia 8 simulates multiple parton-parton scatters in non-diffractive events. Main features of the multiple interaction framework are $p_{\perp}$-ordering and interleaving, small-$p_{\perp}$-dampening of perturbative QCD cross sections, variable impact parameters, and rescaling of parton density distributions. The model is currently being expanded to include the simulation of parton rescattering.

2.3 Simulation with Herwig++

The HERWIG++ event generator [41] includes a simulation of multiple parton-parton scatters within the same proton-proton collision, similar to the PYTHIA approach described in the previous section. HERWIG++'s multiple interaction model [79] is derived from an older model [80] originally developed in the context of photoproduction at HERA. The model extension to simulating multiple soft interactions in minimum-bias events is foreseen [81]. As discussed in Sec. 2.1, the total pp cross section violates unitarity if the scale $p_\perp^2$ of QCD 2 → 2 processes is allowed to become infinitely small. HERWIG++ imposes a sharp cutoff $p_{\perp\text{min}}$ rather than a smooth dampening of small-$p_\perp^2$-processes (cf. Eq. 2.10).

When modeling the overlap of two protons of finite extension, HERWIG++ makes use of the *eikonal picture* or optical model. The eikonal model was developed in the context of describing the propagation of waves through a nucleus of finite extension. This idea is applied to the propagation of two hadrons through each other.

Eikonal Model for the Simulation of Multiple Scatters

Within the eikonal model, partons are assumed to undergo independent scatterings for a given impact parameter. In contrast to PYTHIA's double-Gaussian ansatz (Eq. 2.13), HERWIG++ models the impact parameter dependence of partons in the proton by an electromagnetic form factor

$$\rho(\vec{r}) = \frac{1}{2\pi} \int \mathrm{d}^2\vec{k} \; \frac{\exp\left(i\vec{k}\cdot\vec{r}\right)}{(1+k^2 r_{\text{hadron}}^2)^2} \quad , \tag{2.22}$$

where r_{hadron} refers to the hadron radius. Elastic ep scattering experiments have determined the charge distribution inside a hadron - a distribution that is expected to be similar, but not necessarily identical to the matter distribution: As a consequence, r_{hadron} is treated as a free parameter within HERWIG++. The resulting matter overlap distribution is a modified Bessel function of the third kind:

$$\mathcal{O}(b) = \frac{\mu^2}{96\pi}(\mu b)^3 K_3(\mu b) \ , \tag{2.23}$$

where $\mu \equiv 1/r_{\text{hadron}}$ is the inverse hadron radius.

As Herwig++'s multiple interaction model assumes no correlations between different scatters, the number of scatters at fixed impact parameter follows a Poissonian distribution and the cross section σ_n for exactly n scatters with individual inclusive cross sections σ_{inc} becomes

$$\sigma_n(\sigma_{\text{inc}}) = \int \mathrm{d}^2 b \frac{(\mathcal{O}(b)\sigma_{\text{inc}})^n}{n!} \exp\{-\mathcal{O}(b)\sigma_{\text{inc}}\} \tag{2.24}$$

Together with the inelastic cross section $\sigma_{\text{inelastic}}$, the probability of n scatters, provided there is at least one scatter, follows:

$$\mathcal{P}_{n\geq 1}(\sigma_{\text{inc}}) = \frac{\sigma_n(\sigma_{\text{inc}})}{\sigma_{\text{inelastic}}(\sigma_{\text{inc}})} \tag{2.25}$$

Note that from this equation, only additional scatters of the same type as the primary hard process can be simulated. For an extension to different scattering types, see [79].

Initial- and Final-State Radiation of Hard Process and Additional Scatters

This subsection outlines the procedure of how to connect multiple scatterings to the parton shower and hadronisation steps. Herwig++ does not evolve parton showers and multiple interactions in an interleaved fashion as Pythia does. Instead, the first sequence that is performed during event generation deals entirely with the hard process: Incoming partons are evolved backward to larger x and smaller Q^2. Initial- and final-state showering is simulated via the coherent branching algorithm [82], a directly angular-ordered parton shower with soft gluon coherence.

Next, additional scatters (as many as predicted by Eq. 2.25) are generated with standard matrix elements and standard one-parton densities. This means that in contrast to Pythia, Herwig++ calculates additional scatters on the same footing as the hard scatter. After the required number of additional scatters has been generated, final-state showers are attached. Only for initial-state showers, parton densities are modified according to a parton extraction scheme, that is explained below.

The final event generation step is to combine partons into observable hadrons. In Herwig++, this is carried out according to the cluster fragmentation model [58]. This algorithm however only works with (anti)quarks or (anti)diquarks to begin with. Final-state gluons are hence forced to non-perturbatively decay into light $q\bar{q}$ pairs. Initial-state

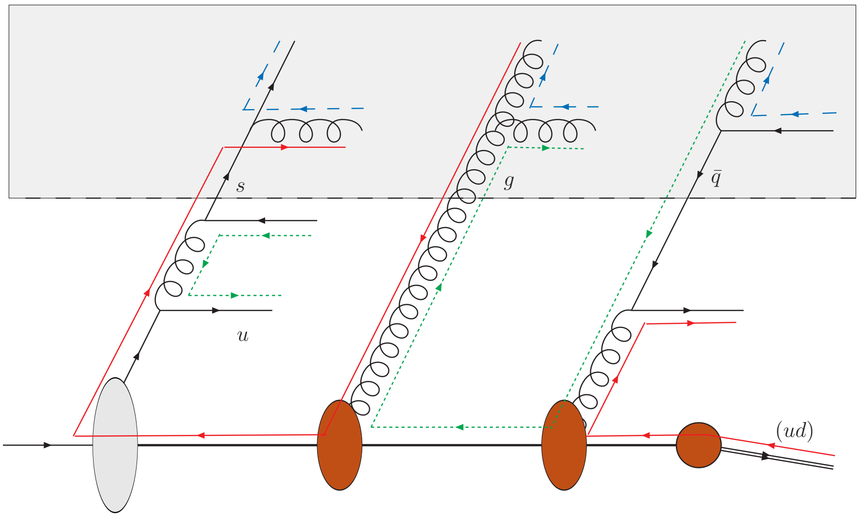


Figure 2.3: Scheme of how the forced splittings and color connections are implemented in HERWIG++. Splittings in the (shaded) box area stem from the hard scatters and the initial-state parton shower. The final splittings at the bottom are non-perturbative. Forced splittings might be necessary to prepare the parton input to the cluster hadronisation algorithm. Figure taken from [79].

gluons however are subject to the parton extraction model described in the subsequent subsection.

Parton Extraction from the Beam Remnant

As stated before, the cluster fragmentation algorithm cannot handle gluons as input particles. For the primary hard scatter, this is achieved by terminating the backward evolution on a valence quark. In this case, the proton remnant is a diquark that is for instance color-connected to a final-state parton emitted by the initial-state shower. Alternatively, the diquark is color-connected to a jet produced in the hard process.

If subsequent scatters were to be treated in the same way, another valence quark would be forced out of the already saturated proton. Instead, the HERWIG++ approach is to terminate the backward evolution on a gluon. As above, this is enforced by additional backward evolution steps, if necessary. For the backward evolution of additional scatters, parton densities without valence contributions are used. No further modifications or rescalings of parton densities are performed. This implies that energy-momentum conservation may be violated; if so, the generation is vetoed and redone.

Once all additional scatters have been backward-evolved to gluons color-connections to initial quark lines are attached in random order.

Figure 2.3 illustrates the concept of parton extraction in the HERWIG++ shower evolution. The (shaded) box in the upper part encloses partons from a perturbative parton shower. Below the box, the order in which partons are extracted from the proton beam is illustrated (from right to left): First, the hard scatter extracts a valence quark (rightmost circle). In additional scatterings, forced splittings ensure a desirable flavor structure of the proton: For example, scatter two (rightmost ellipse) involves an antiquark that is backward-evolved to a gluon.

Summary: Multiple Interactions in Herwig++

HERWIG++ simulates multiple parton-parton scatters in non-diffractive events. Multiple scatters are not ordered and not interleaved with parton showering. At small transverse momenta $p_\perp$, no dampening but a sharp cutoff on additional interactions is imposed. Matter distribution inside the proton follows the electromagnetic form factor, where the hadron radius is kept as free parameter. Parton densities are not modified except from excluding valence contributions. Possible violations of energy-momentum conservation are vetoed. Color-connections are included for all parton-parton scatters.

Chapter 3

The Compact Muon Solenoid Experiment at the Large Hadron Collider

The present chapter describes the *L*arge *H*adron *C*ollider (LHC) at CERN (Sec. 3.1) and one of its general-purpose experiments, the *C*ompact *M*uon *S*olenoid (CMS) detector (Sec. 3.2).

3.1 The Large Hadron Collider (LHC) at CERN

The Large Hadron Collider (LHC) [83] at CERN is designed to collide proton beams with center-of-mass energies of 14 TeV and a luminosity of 10^{34} cm^{-2}s^{-1}. It is a two-ring superconducting hadron accelerator and collider in a 26.7 km long tunnel of eight straight sections and eight arcs between 45 m and 170 m below the surface. As the LHC is a proton-proton collider, both beams cannot share the same phase space in a single ring, but two rings with counter-rotating beams are necessary. The LHC can also collide heavy ions with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27} cm^{-2}s^{-1}.

The present section describes the machine's performance goals (Sec. 3.1.2) and limitations (Sec. 3.1.3) as well as the general layout (Sec. 3.1.1).

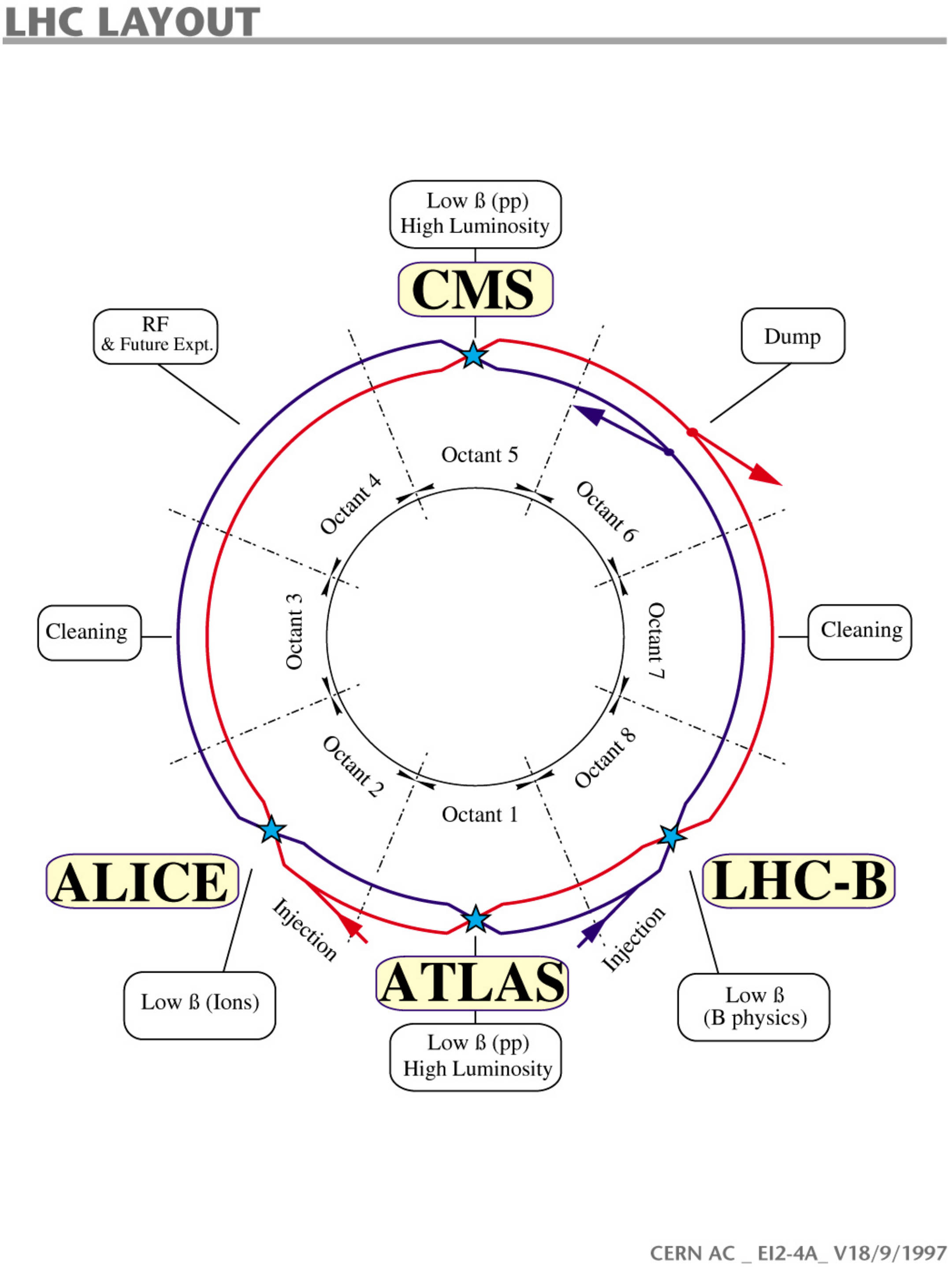


Figure 3.1: Schematic layout of the Large Hadron Collider (LHC) [83]. The four main experiments are LHC-B, which explores b-physics in proton-proton collisions, ALICE, which investigates heavy-ion physics in lead-lead collisions, and the two general-purpose experiments ATLAS and CMS.

3.1.1 Lattice layout

An overview of the LHC layout with its four main experiments is shown in Fig. 3.1. The experiments are:

- LHC-B, exploring b-physics in proton-proton collisions at a peak luminosity of $\mathcal{L} = 10^{32}$ cm^{-2}s^{-1},
- ALICE, investigating heavy-ion physics in lead-lead collisions at $\mathcal{L} = 10^{27}$ cm^{-2}s^{-1},
- and the two general-purpose experiments ATLAS and CMS, both exploring proton-proton collisions at $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}.

The LHC lattice layout follows the LEP tunnel geometry: The LHC is divided into eight arcs and eight straight sections of 528 m length. The two high-luminosity experiments, ATLAS and CMS, are located opposite to each other, at points 1 and 5. Point 2 houses the ALICE and point 8 houses the LHC-B experiment. The *beam dump insertion* is located at Point 6: Horizontally-deflecting fast-pulsed magnets combined with vertically-deflecting double steel septum magnets extract the beams vertically.

23 regular arc cells of 106.9 m length build up the arcs of the LHC lattice. Each arc cell is subdivided into two half cells, each containing a cold mass, a short straight section and three dipole magnets. The design of an LHC arc cell has been chosen such as to optimize the maximum integrated dipole field along the arc with the smallest attainable number of magnet interconnections and minimum beam envelopes.

3.1.2 Performance goals

In the past, a number of hadron accelerators, e. g. $Sp\bar{p}S$ or Tevatron, collided particles with their antiparticles. Meanwhile, with present-day technology, anti-proton beams cannot be made intense enough to reach the desired peak luminosities of the LHC. Therefore, it was decided to build the LHC as a proton-proton collider instead, even though two counter-rotating proton beams require opposite magnetic dipole fields in both rings. As a consequence, the common particle-antiparticle collider configuration of one vacuum and one magnet system for both beams is excluded. The LHC is therefore designed with separate vacuum chambers and magnetic fields in the main arcs. The two vacuum systems are only linked at the interaction regions in 130 m long common beam pipes. The magnetic fields must be strong enough to maintain the protons' circular paths. The design center-of-mass energy of 14 TeV thus implies a peak dipole field of

8.33 T and the use of superconducting magnets in the LHC. Taking into account the large number of bunches per beam (2808) and the nominal bunch spacing of 25 ns, on average between 25 and 40 *parasitic* collisions may occur per interaction region, depending on the exact value of the total proton-proton cross section at the LHC [84]. These parasitic collisions (pile-up) are suppressed with the help of dedicated crossing angle orbit bumps that separate the beams left and right from the interaction point.

In the following, we list some of the most common variables used in accelerator physics. The number of events $N(\text{event})$ per second generated in collisions with event cross section $\sigma(\text{event})$ and machine luminosity $\mathcal{L}$ is

$$N(\text{event})/\text{s} = \mathcal{L}\ \sigma(\text{event})\ . \tag{3.1}$$

The machine luminosity is a function of the beam parameters. For Gaussian-distributed beams with N_b particles per bunch, n_b bunches per beam, revolution frequency f_{rev}, relativistic gamma factor γ_r, normalized transverse beam emittance ϵ_n and beta function at the collision point β^*:

$$\mathcal{L} = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi\epsilon_n\beta^*}\ F\ . \tag{3.2}$$

A non-zero crossing angle at the interaction point reduces the luminosity, reflected by the geometric luminosity reduction factor F:

$$F = \left(1 + \left(\frac{\theta_c\sigma_z}{2\sigma^*}\right)^2\right)^{-1/2}\ , \tag{3.3}$$

with

- the full crossing angle at the interaction point θ_c,
- the root-mean-square of the bunch length σ_z and
- the transverse root-mean-square of the beam size at the interaction point $\sigma^* = \sqrt{\beta^*\epsilon_n}$.

The subsequent section discusses limitations to the LHC performance. Of particular interest are effects that influence the accelerator luminosity.

3.1.3 Performance limitations

The design performance of the LHC is affected by a number of effects. These effects are summarized in the following.

Nonlinear interactions between particles from both beams (*nonlinear beam-beam interactions*) limit the maximum particle density per bunch N_b. The linear tune shift ξ characterizes beam-beam interactions:

$$\xi = \frac{N_b r_p}{4\pi\epsilon_n} \ , \tag{3.4}$$

with r_p the classical proton radius, $r_p = e^2/(4\pi\epsilon_0 m_p c^2)$ and ϵ_n the normalized transverse beam emittance. An upper limit on the total linear tune shift summed over all interaction points of $\xi_{\mathrm{total}} < 0.015$ has been determined from experience with existing hadron colliders. Three experiments with head-on proton-proton collisions thus imply a linear tune shift per interaction point of not more than 0.005.

Another limitation of the LHC performance is based on geometry choices. The beam screen dimensions of 2×17.3 mm height and 2×22 mm total width determine the *geometrical aperture* of the LHC arcs. A peak nominal beam size of 1.2 mm follows from choosing a minimum aperture of 10 RMS beam sizes and allowing for imperfect machine, magnet alignment and geometry. The maximal value of the β-function in the LHC arcs is 180 m, consequently the transverse beam emittance must not exceed $\epsilon_n = 3.75$ µm. Together with the limit on the beam-beam tune shift, the LHC mechanical aperture limits the maximum bunch intensity to $N_b = 1.15 \times 10^{11}$ protons ber bunch. An additional aperture introduced by the triplet magnets limits the minimum attainable β^* value at the interaction points and the maximum attainable crossing angle orbit bump in the experimental interaction region.

The peak beam energy is a function of the integrated magnetic dipole field around the storage ring. Thus, the peak dipole field in the LHC storage ring limits the maximum beam energy. Design parameter is a field of 8.33 T, corresponding to a beam energy of 7 TeV. The actual field is however determined by *heat load* and temperature margins in the cryo-magnets (*quenching*), i. e. by possible beam losses during operation. One possible source of heat load on the cryogenic system might be the absorption of synchrotron radiation. Additional heat load is coming from luminosity-induced losses, impedance-induced losses (resistive wall effect), and electron cloud bombardment.

During nominal LHC operation, a 0.584 A total beam current is flowing corresponding to 362 MJ stored energy in the circulating beams; add this energy to 600 MJ electromagnetic energy to obtain a total stored energy exceeding 1 GJ. The *beam dumping* and magnet systems have to absorb this energy at the end of each run (or in case of an emergency) limiting the beam energy and intensity further.

Next, we will consider effects influencing the beam stability. The beam stability depends to a great extent on a high field quality. Typically, persistent currents in superconducting magnets decay and *snap back* at the beginning of the ramp. The accuracy of the field quality correction can be estimated assuming fixed limits for beam losses. *Collective beam instabilities* are caused by electromagnetic interactions of the charged beam particles with each other and with the vacuum system's conducting boundaries. Being a function of the vacuum system surface properties and geometry, collective beam instabilities are proportional to the beam currents, thereby limiting the beam intensity.

These degrading beam intensities and emittances cause the LHC luminosity to decay in the course of a physics run, mostly due to beam loss from collisions. The initial decay time τ_{nuclear} is the ratio between initial beam intensity $N_{\text{tot, 0}}$ and the product of initial luminosity $\mathcal{L}$, total cross section σ_{tot}, and the number of interaction points k:

$$\tau_{\text{nuclear}} = \frac{N_{\text{tot, 0}}}{\mathcal{L}\ \sigma_{\text{tot}}\ k}\ . \tag{3.5}$$

Thus, for LHC design parameters ($\mathcal{L} = 10^{34}\ \text{cm}^{-2}\text{s}^{-1}$, $k = 2$) and for a total cross section $\sigma_{\text{tot}} = 100\ \text{mb} = 10^{-25}\ \text{cm}^2$ at 14 TeV, the luminosity decreases to $1/e$ of the initial luminosity within 29 hours due to collision losses.

Further effects to consider are beam-gas interactions and nonlinear beam-beam interactions, as well as radio-frequency noise and intrabeam scattering; these effects lead to a slow *emittance blow-up* of the beam. Beam-beam interactions and radio-frequency noise are in turn compensated by synchrotron radiation damping. The net luminosity lifetime is eventually estimated from intrabeam, rest-gas and collision-loss contributions:

$$\frac{1}{\tau_{\mathcal{L}}} = \frac{1}{\tau_{\text{IBS}}} + \frac{1}{\tau_{\text{rest gas}}} + \frac{1}{\tau_{\text{nuclear, }1/e}}\ . \tag{3.6}$$

An overall estimate on the luminosity lifetime with time constants of 80 hours for intrabeam scatterings and 100 hours for rest-gas scatterings, yields:

$$\tau_{\mathcal{L}} = 15\ \text{hours}\ . \tag{3.7}$$

An entire cycle from filling the PS and SPS synchrotrons, ramping the beam energy in the LHC from 450 GeV to 7 TeV and ramping the magnets down to 450 GeV ending with a final check of all main systems, takes a minimum of 70 minutes. Practical experience at HERA shows that the average turnaround time is 6 times this minimum attainable time, thus leading to an estimate of the average LHC turnaround time of 7 hours.

The *integrated luminosity* for a run of duration T_{run} is:

$$\mathcal{L}_{\mathrm{int}} = \mathcal{L}_0 \tau_{\mathcal{L}} \left[1 - \exp\left\{-T_{\mathrm{run}}/\tau_{\mathcal{L}}\right\}\right] \ . \tag{3.8}$$

With 200 days of operation per year and a luminosity lifetime of 15 hours, as derived in Eq. 3.7, the maximum total integrated luminosity per year ranges between 80 fb^{-1} for machine turnaround times of 7 hours and 120 fb^{-1} for machine turnaround times of 70 minutes.

3.2 Compact Muon Solenoid (CMS)

This section describes the Compact Muon Solenoid (CMS) detector [85] operating at Experiment Point 5 of the Large Hadron Collider (described in Sec. 3.1). The design of CMS is driven by the aim to study proton-proton collisions at center-of-mass energies of 14 TeV and at luminosities of up to 10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. Its main features are a superconducting solenoidal magnet enclosing an all-silicon inner tracking system, a scintillating-crystals electromagnetic calorimeter made from lead-tungstate, and a sampling hadron calorimeter from brass-scintillator. Muon detectors are placed within the iron magnetic-flux return-yoke. A hermetic coverage up to high pseudorapidity values is achieved with the use of forward sampling calorimeters. The CMS detector weighs 12500 tons, is 21.6 m long and 14.6 m high. A perspective view of the CMS experiment is depicted in Fig. 3.2. Its components are described in more detail in the subsequent sections.

CMS adopts a coordinate system with the origin at the nominal interaction point, the y-axis pointing vertically upward and the x-axis pointing toward the LHC center. Azimuthal angles are measured with respect to the x-axis in the xy-plane, with radial coordinate r. Polar angles are measured with respect to the z-axis and often represented by the pseudorapidity $\eta = -\ln\tan(\theta/2)$.

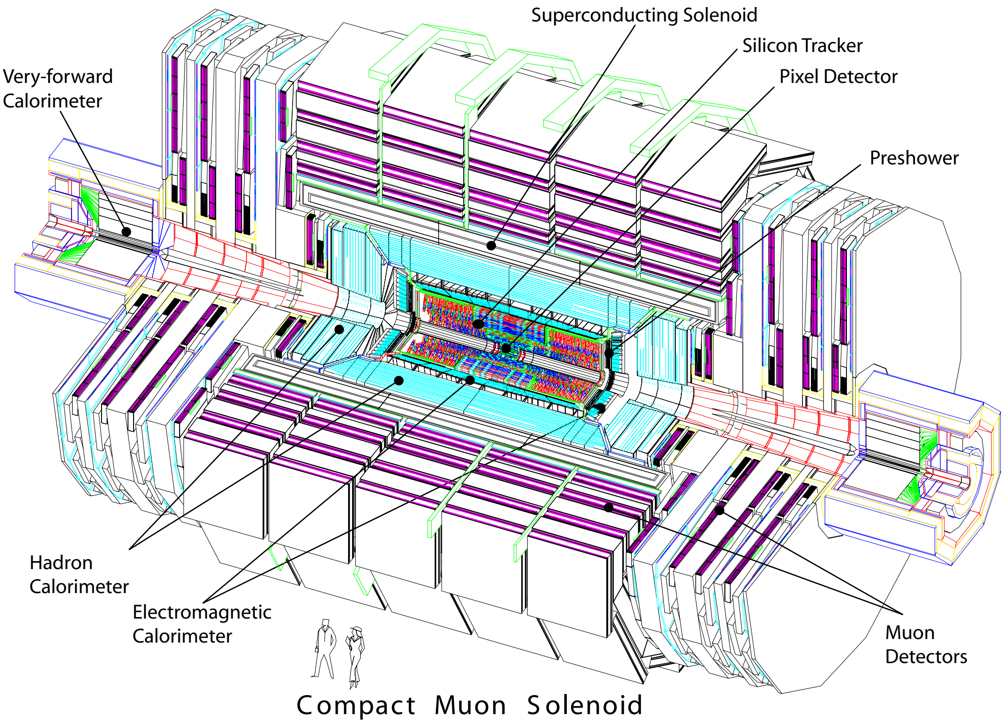


Figure 3.2: A perspective view of the 21.6 m long and 14.6 m high CMS detector [85].

With a total proton-proton cross section of 100 mb at 14 TeV center-of-mass energy, CMS will observe on average 10^9 inelastic collisions per second. These have to be reduced by the CMS trigger system to 100 events per second for storage and offline analysis; a difficult task in view of the short time between two LHC bunch spacings: Every 25 ns, on average 20 inelastic collisions produce approximately 1000 charged particles that influence the event measurement. Most of the time, these collisions will be low-p_T and low-multiplicity processes. Therefore, trigger systems attempt to efficiently select "interesting" events, where at least one hard proton-proton scatter occurs. For these triggered events, all particle signals will be read out, regardless of whether the particles stem from the scatter of interest or from additional interactions in the same bunch-crossing (*pile-up*). Pile-up affects the detector performance, because it increases the detector occupancy. However, with a high detector granularity and a fast time resolution, the impact of pile-up can be reduced. This has the consequence, that the number of detector electronic channels to be read out is of the order of millions. What's more, these channels need to be synchronized to a high degree. Besides these issues with acquiring the data, the large particle flux imposes a high level of radiation, so detectors and front-end electronics need to be radiation-hard.

The broad LHC physics program requires four main design choices:

- Identify *muons* with high efficiency and good momentum resolution over a wide momentum range and with a large pseudorapidity coverage. Resolve dimuon masses to a level of 1% at $m_{\mu\mu} = 100$ GeV/c^2 and determine muon charges up to momenta of 1 TeV/c.
- Reconstruct *charged particles* in the inner tracker with high efficiencies and good momentum resolution. Trigger efficiently on and tag offline τ *leptons and* b *jets.*
- Resolve *electromagnetic energy, diphoton and dielectron masses* to excellent levels (1% at 100 GeV/c^2). Cover a large pseudorapidity range, efficiently reject π^0 mesons and isolate photons and leptons at high luminosities.
- Resolve *missing transverse energy and dijet masses* to good levels.

All of these requirements are met by the CMS detector. We will describe the various subdetectors in the subsequent sections.

3.2.1 Superconducting Magnet

The CMS superconducting solenoid magnet provides a strong magnetic field of 3.8 T. In this field, charged particles are bent such that their transverse momenta can be measured with great precision. The CMS magnet is 13 m long and has an inner diameter of 6 m. The return field saturates 1.5 m of iron in an iron yoke of mass 10 000 tons. At full current, 2.6 GJ energy are stored in the magnet. This leads to a previously unreached value in mechanical deformation of the magnet structure (0.15%). The deformation is attained during energising, due to the large ratio between stored energy and cold mass of 11.6 KJ/kg.

3.2.2 Inner Tracking System

Charged particle trajectories and their production vertices are measured by the CMS inner tracking system. The CMS tracker is located around the interaction point with a length of 5.8 m and a radius of 1.25 m. A magnetic field of $B = 3.8$ T bends charged particle trajectories in order to measure the particle's momentum and charge from hit patterns in the all-silicon tracking detector, see Fig. 3.3.

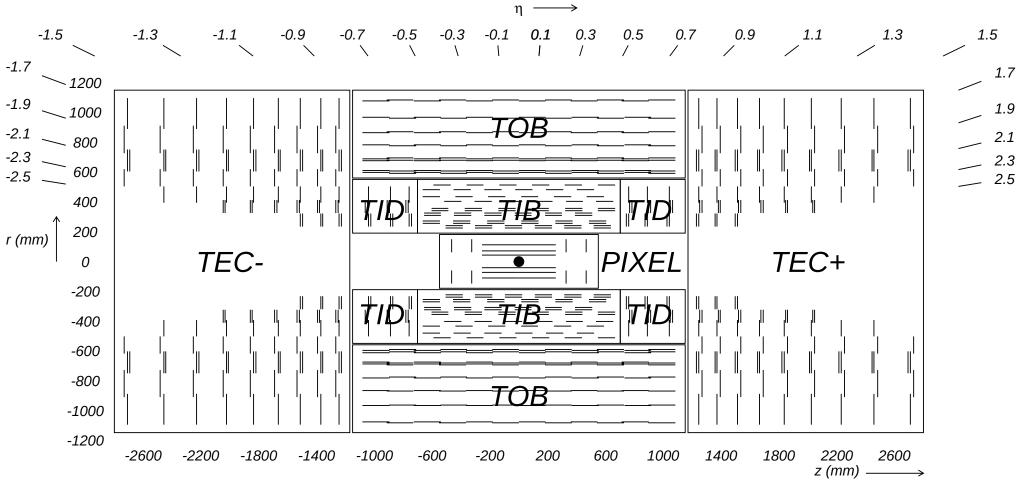


Figure 3.3: Longitudinal sketch of the CMS central tracking system [85]. Each line represents a detector module. Double lines indicate back-to-back modules delivering stereo hits.

The tracker design is a compromise between requiring high granularity and fast response on the one hand and suppressing multiple scattering, bremsstrahlung, photon conversion, and nuclear interactions of particles on the other hand. High granularity and fast response are achieved with on-detector electronics with efficient cooling, leading to an increase in the amount of material. However, the fewer material the less likely secondary interactions are taking place. Therefore, from the second requirement, a decrease in the amount of material would be preferred. On top of these conflicting considerations, radiation hardness of the detector components have to be taken into account: Highly intense particle fluxes severely damage the inner tracking system during the expected lifetime of 10 years. An all-silicon design of the CMS inner tracking system has been chosen to meet the requirements of high granularity, fast speed, and radiation hardness.

As illustrated in Fig. 3.3, three pixel layers are placed in the barrel at radii of 4.4 cm, 7.3 cm, and 10.2 cm followed by ten strip layers in the barrel up to radii of 1.1 m. Endcaps with silicon modules perpendicular to the beam axis complement the inner tracking system: Two pixel disks and three inner plus nine outer strip disks on each side cover pseudorapidities up to $|\eta| = 2.5$.

The CMS inner tracking system is designed to robustly, efficiently, and precisely reconstruct charged particles with transverse momenta above 1 GeV/c emerging into the

Table 3.1: Hit rate densities at LHC design luminosity in three example distances to the nominal vertex.

Distance to nominal vertex	Hit rate density
4 cm	1 000 000 Hz/mm^2
22 cm	60 000 Hz/mm^2
115 cm	3 000 Hz/mm^2

pseudorapidity region $|\eta| < 2.5$. Efficient heavy flavor identification relies on precise measurements of secondary vertices and impact parameters. Electrons are measured by the inner tracking system and ECAL together; muon measurements make use of the tracker and muon systems. τ-leptons can be identified in one-prong or three-prong decays. Finally, albeit not used at the first trigger level, track measurements are an important input to the high-level trigger algorithms.

Hit rate densities in three example distances to the nominal vertex at the LHC design luminosity are listed in table 3.1. The hit rate density does not fall like $1/r^2$ because of secondary interactions and backscattering from the calorimeter material. A desired occupancy of less than 1% dictates the use of pixel detectors in the innermost tracker up to radii of 10 cm. The pixel sensors in use at CMS have lengths of 100 µm transverse to the beam direction and 150 µm parallel to the beam direction, corresponding to an average occupancy of 10^{-4} per pixel and bunch crossing. 320 µm thick silicon micro-strip detectors with cell dimensions 10 cm × 80 µm are placed at intermediate radii $20\text{ cm} < r < 55\text{ cm}$. Strip occupancies range between 2% and 3% per bunch crossing. Strip cell dimensions are further increased to 25 cm × 180 µm in the outer region $55\text{ cm} < r < 110\text{ cm}$ with average occupancies up to 1%. The increased strip length reduces the number of read-out channels but increases the strip capacitance as well, hence a larger strip thickness of 500 µm has been chosen for these sensors.

Radiation damage is mostly caused by three effects that had to be accounted for when designing the inner tracking system:

Surface damage Silicon oxide layers may trap holes, thus changing the space charge configuration and possibly the *m*etal *o*xide *s*emiconductor (*MOS*) characteristics of front-end chips.

Bulk damage *N*on-*i*onizing *e*nergy *l*oss (*NIEL*) can modify the silicon crystal lattice, increasing the leakage current, changing the doping from *n*- to *p*-type and creating

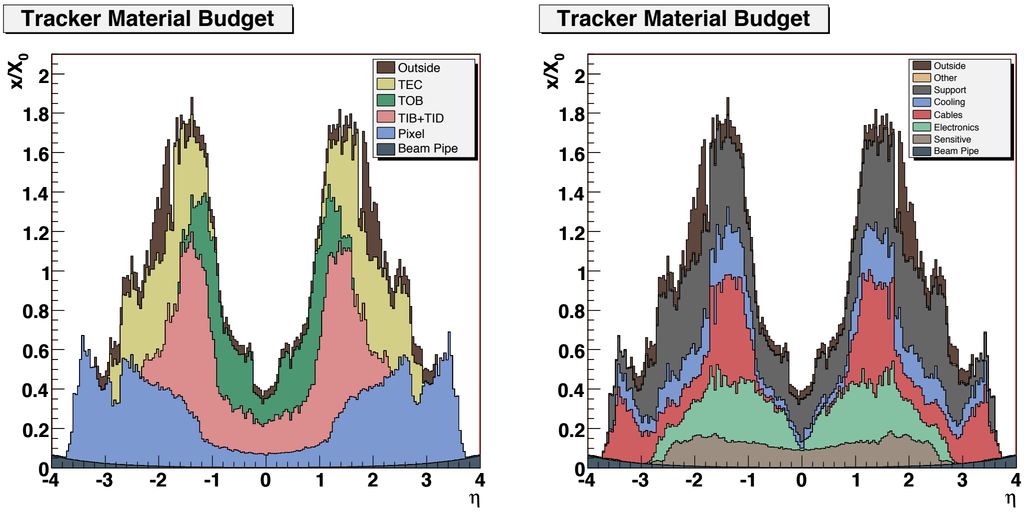


Figure 3.4: CMS tracker material budget in units of radiation lengths X_0 shown as a function of detector pseudorapidity [85]. Contributions from subdetectors (left) and material purpose (right) displayed.

further trapping centers. Bulk damages reduce the signal by about 10% after 10 years of LHC running.

Transient phenomena Ionizing particles can generate charges in the electronics, thus possibly changing the state of memory cells and corrupting the event read-out (*single event upset*).

Read-out chips have been produced following radiation-hard standard 0.25-µm-CMOS-technology; consequently, they are not the limiting factor in determining the lifetime of the CMS tracker. Bias voltages of up to 500 V guarantee over-depletion of the Silicon sensors to collect charges efficiently. The p-on-n type silicon micro-strip sensors have to be run in an over-depleted state, but the pixel detector has been designed as n^+ pixel on n-substrate specifically to enable running in an under-depleted state and thereby ensuring more than ten years lifetime for the outermost pixel layer.

The tracker layout as depicted in Fig. 3.3 guarantees at least three space point measurements per charged particle trajectory in the silicon pixel tracker as well as 9 hits in the silicon strip tracker.[1] The material budget of the CMS inner tracking system, depicted in Fig. 3.4, grows from 0.4 X_0 for tracks going perpendicular to the beam axis ($|\eta| = 0$) to 1.8 X_0 at $|\eta| = 1.4$. At $|\eta| = 2.5$, the material budget has decreased

---

[1] In general, the number of crossed layers, and hence the number of hits, depends on the particle's pseudorapidity and transverse momentum as well as on possible multiple scatterings within the tracker material (cf. Sec. 6.1.1).

to 1.0 X_0. Note that the sensitive material budget contributes only approximately 0.15 X_0 throughout the entire tracker acceptance. The two largest contributions to the tracker material budget come from mechanical support structures and cabling, followed by contributions from cooling and electronics.

The CMS inner tracking system resolves transverse momenta of $p_T = 100$ GeV/c to $1 - 2\%$ precision up to pseudorapidities of $|\eta| = 1.6$. The dominant contribution to the momentum resolution of low momenta comes from multiple scattering. Transverse and longitudinal impact parameters for low momentum tracks are heavily influenced by multiple scattering as well. Single muons are reconstructed with about 99% efficiency, in spite of a small performance drop for $\eta = 0$ due to gaps between the pixel ladders at $z = 0$. Large pseudorapidities are only partly covered by the pixel forward disks, resulting in a decreasing efficiency in those regions. Multiple interactions are more probable for pions (and hadrons in general), hence the overall tracking efficiency is smaller than the tracking efficiency for muons. A detailed performance study of the CMS inner tracker with respect to reconstructing charged particles used in the underlying-event analysis is presented in Sec. 6.1.

3.2.3 Track Reconstruction with the CMS Tracker

A reconstructed track is the result of a *pattern recognition* step and of a *track fit*. Pattern recognition refers to an algorithm that decides which measurements to use; the track fit refers to adjusting the track model parameters to the selected measurements. With the introduction of advanced fitting algorithms this distinction became less clear - in many cases pattern recognition and track fit can be handled in a single step. A charged particle in a magnetic field moves along a helix trajectory, which can be described by five parameters. Locally, i. e. on or local to a plane (typically the measurement plane of a detector), a charged-particle trajectory is parametrized by the following parameters:

q/p The particle charge q (plus or minus one) divided by the magnitude of the particle's momentum p,

$\mathrm{d}x/\mathrm{d}z$ the direction tangent in the local (x, z) plane,

$\mathrm{d}y/\mathrm{d}z$ the direction tangent in the local (y, z) plane,

x the particle's local x coordinate,

y the particle's local y coordinate,

$\mathrm{sgn}(z)$ the sign of the local z-component of the momentum.

Globally, e. g. in the CMS coordinate system, trajectories are parametrized by the points (x, y, z) and momenta (p_x, p_y, p_z) along the trajectories. Physics analyses are often concerned with the quality of the reconstruction of a given track. The track quality is parametrized by the total χ^2 of the track fit, by the number of degrees of freedom of the fit, by the number of measurements $n(\text{hits})$ used in the fit, and by the number of gaps in the sequence of measurements $n(\text{lost hits})$. The impact parameter of a track is characterized by the transverse impact parameter d_0 and longitudinal impact parameter Δz.

At high luminosity, each bunch-crossing will on average result in more than 20 inelastic proton-proton scatters[2] producing more than one thousand reconstructible tracks (cf. Sec. 3.1). The resulting track densities amount to up to 10 tracks per cm^2 per bunch crossing in 2 cm radial distance from the interaction point. The transverse momentum resolution needed to reconstruct narrow resonances is between 1 and 2% at track momenta of 100 GeV/c.

In the following, we summarize the algorithms in use to reconstruct tracks emerging from the interaction point [86, 87]. Standard reconstruction is carried out with a Kalman filter algorithm [88]. Alternative reconstructors for special applications are implemented as well. Reconstruction of tracks from cosmic rays is discussed elsewhere [89].

Standard Track Reconstruction: Combinatorial Track Finder

The standard track reconstruction algorithm employed in CMS measurements, the *Combinatorial Track Finder*, consists of four steps: seed generation, trajectory building, trajectory cleaning, and trajectory smoothing. These steps are explained in the following. Due to the CMS tracker's high material budget, 20% of all 1 GeV pions are stopped before reaching the strip tracker layers [87]. Thus, tracks with hits in inner layers, i. e. tracks with pixel hits, are implicitly favored.

Seed generation refers to initializing the trajectory with the first hits and a first estimate of the helix parameters. Next, trajectory building is performed in the following way: The trajectory candidate is grown layer by layer, starting from the seed and proceeding towards outside layers. A Kalman filter algorithm decides which hit measurements are compatible with the trajectory candidate. Detector inefficiencies might

---

[2] Note that here we refer to *inelastic* scatters, i. e. neglecting single- and double-diffractive interactions as well as elastic scatterings.

cause missing hit measurements (lost hits). Standard reconstruction allows trajectory candidates with these invalid hits, but discards the candidate if a maximum number is exceeded. In addition, the total number of candidates that is kept after each iteration step is limited. Hit assignment ambiguities are resolved in the subsequent trajectory cleaning step: Each seed often results in trajectory candidates composed to large parts of the same hits. A subset of compatible candidates is chosen based on the number of shared hits and the quality of tracks. Once the trajectory candidates have been propagated to the outer surface of the tracker, trajectory building is complemented with a trajectory smoothing step. Trajectory smoothing is carried out by employing a Kalman filter fit of the trajectory towards the vertex. At this point, optimal estimates at each surface have been obtained. In the standard reconstruction, the outward fit is repeated to remove possible biases from seeding (final track fit).

Single muon track reconstruction efficiencies of the combinatorial Kalman filter are better than 98% up to pseudorapidities of $|\eta| = 2.2$ [87]. The track reconstruction efficiency for pions is lower, as pions undergo nuclear interactions in the tracker material. The transverse momentum resolution for muons of $p_T = 100$ GeV/c is $1-2\%$ up to pseudorapidities of $|\eta| = 1.75$ [87]. The impact parameter resolution at high momenta is dominated by the resolution of the first hit in the pixel detector. At low transverse momenta, multiple scattering degrades the track parameter resolution.

The Combinatorial Track Finder is the default algorithm for most tracking applications in CMS analyses, including studies of the underlying event in charged-jet topologies (cf. Chapters 6 and 7). This algorithm has been found to perform excellently even under high luminosity conditions or in heavy ion collisions [90].

Alternative Track Finding Algorithms

Several alternative track reconstructors have been developed for special applications. These algorithms are assumed to be more apt to reconstruct tracks in dense environments or tracks suffering from a lot of material effects. We will summarize the main features of three of these reconstructors in the following.

The *Deterministic Annealing Filter* and *Multi-Track Filter* are special filters developed for environments with high track densities, e. g. high E_T b jets or τ jets. By deferring the final hit assignment to the final track fit, these account for misassociated hits and for hits contaminated by nearby tracks.

In addition, measurement errors are often non-Gaussian and material effects, such as energy loss or multiple scattering, tend to have have long tails (a non-negligible number of particles loses a large amount of energy or scatters many times or both). Also, neighboring tracks, electronic noise and δ-electrons degrade the hit resolution and introduce wrong assignments of hits to tracks. An algorithm that has been designed specifically to address these issues is the *Gaussian-sum filter* (GSF) [91]. This algorithm is a weighted sum of several Kalman filters. It models distributions with mixtures of multivariate Gaussian probability density functions, where the main component describes the core and additional components describe the tail of the distribution to be modeled. An example application of the Gaussian-sum filter is the track reconstruction of electrons.

3.2.4 Alignment Strategy for the CMS Tracker

With a hit resolution of $10-50$ µm, the 15 000 independent silicon sensors comprising the CMS tracker must be *aligned* with great precision, i. e. absolute position and orientation of individual detector elements with respect to the global CMS coordinate system must be known with great accuracy.

CMS will employ a three-step alignment procedure [92]. First, position and orientation of tracking devices were measured during assembly of the CMS tracker, e. g. using photogrammetry. Additionally, the positions of detector modules in the Tracker Inner Barrel (TIB), the Tracker Outer Barrel (TOB), and the Tracker Endcap Disks (TEC) are monitored during operation with the help of infrared laser beams (*Laser Alignment System*, LAS). Eventually, in order to achieve optimal alignment accuracy, track-based alignment procedures are employed.

Misalignment of the CMS tracker can be simulated by displacing and rotating tracker parts at the reconstruction level, i. e. after detector simulation [93]. Sensor hits are simulated in accord with the ideal tracker geometry; modules, rods, layers or half-barrels are displaced and rotated afterwards.

Track-based Alignment

Track-based alignment of the CMS tracker is carried out using complementary sets of track data. Cosmic ray muons and beam halo muons traverse the entire tracker and are used to cross-align opposite detector parts. Mass constraints on the invariant mass of muon pairs from Z boson or J/ψ meson decays help to correct the absolute momentum scale. In a similar manner, vertex constraints on tracks as well as single isolated tracks in minimum-bias events are utilised. At least six numbers (three space coordinates and three rotation angles) are necessary to specify position and orientation of each of the 15 000 silicon sensors in the CMS tracker. The resulting $\mathcal{O}(10^5)$ unknowns comprise a covariance matrix of $\mathcal{O}(10^{10})$ entries. The three track-based alignment algorithms employed by CMS are summarized in the following. Common feature is the minimization of some mathematical expression (objective function) built from normalized unbiased track hit residuals.

The track fit can be used to predict the impact point on a given silicon sensor. However, the actual track hit may differ from the predicted impact point due to misalignment. Aim of the *Hits-and-Impact-Points Algorithm* [94, 95] is to iteratively minimize the sum of residuals for a given module, rod, layer or half-barrel. The *Kalman filter* for track based alignment [96] includes alignment parameters in the track fit (cf. Sec. 3.2.3). In order to avoid too complex calculations, only "near-by" modules are taken into account. The *Millepede Algorithm* [97, 98] first linearizes and then minimizes the objective function. Track and alignment parameters are determined simultaneously, rather than being determined iteratively. The algorithm reduces the problem of finding $\mathcal{O}(10^{10})$ unknown covariance matrix entries to solving $\mathcal{O}(10^5)$ linear equations.

3.2.5 Electromagnetic Calorimeter

The CMS electromagnetic calorimeter (depicted in Fig. 3.5) consists of 61 200 lead tungstate ($PbWO_4$) crystals in the barrel part and 7 324 crystals in each of the endcap parts. The endcap crystals are sitting behind preshower detectors. Signals are read out with avalanche photodiodes (barrel) or vacuum phototriodes (endcaps). High density crystals allow for a fast, highly granular and radiation-resistant calorimeter. Good energy resolution enhances the mass resolution of two photons from the decay of a Standard-Model-Higgs boson.

Lead tungstate is an extremely dense (mass density 8.28 g/cm^3) material with short radiation length ($X_0 = 0.89$ cm) and small Molière radius (2.2 cm), the characteristic material constant giving the scale of the transverse dimension of the fully contained electromagnetic showers. In consequence, the CMS ECAL can be built compact and with fine granularity leading to a better shower position resolution, and better shower separation due to less shower overlaps. A fast scintillation decay time of the used crystals

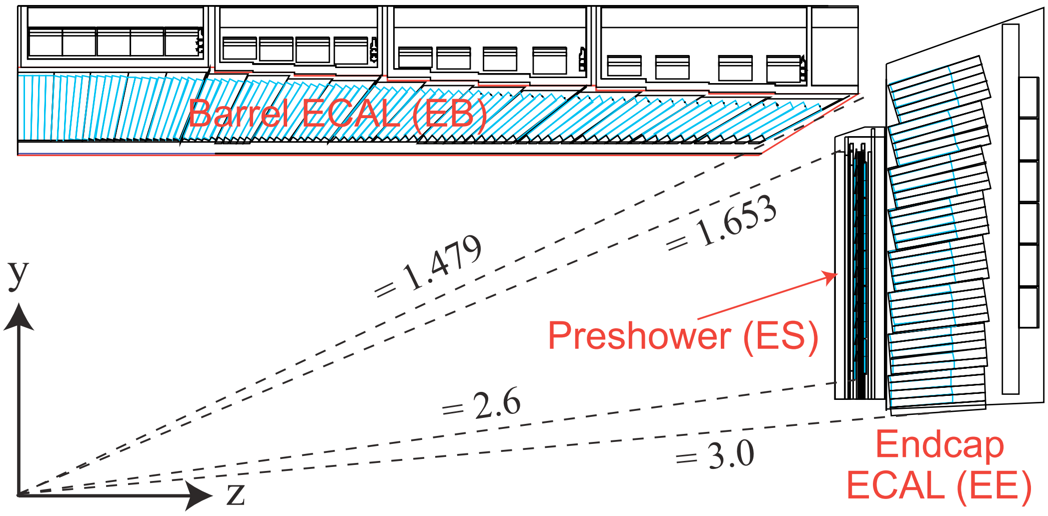


Figure 3.5: Longitudinal sketch of the CMS electromagnetic barrel (EB) and endcap (EE) calorimeters along with the preshower detector (ES) [92]. The layout of one quarter of the full subdetector is shown.

leads to 80% of the light being emitted in 25 ns, although the output is relatively small. In addition, the light yield depends on the temperature: 4.5 photoelectrons per MeV are collected at 18°C.

Ionizing radiation traversing the lead tungstate crystals will, in the course of the experiment, create absorption bands by forming color centers caused by oxygen vacancies and lattice impurities. Practically, this means a wavelength-dependent loss of light transmission. The injection of laser light helps to monitor the optical transparency. Radiation damage can thus be detected and corrected for. Damage and recovery at 18°C will eventually converge to a dose-rate dependent equilibrium.

The CMS *ECAL barrel* part covers pseudorapidities $|\eta| < 1.479$. The total of 61 200 crystals can be decomposed into 360 cells in azimuth and 2×85 cells in pseudorapidity with a front face of of 0.0174×0.0174 in (η, ϕ) positioned at 1.29 m from the beam axis. Barrel crystals are 25.8 radiation lengths X_0 long. Electromagnetic particles emerging into the pseudorapidity range $1.479 < |\eta| < 3.0$ are detected by the *ECAL endcaps* located at 315.4 cm distance to the nominal interaction point. Each endcap half (*dee*) holds 3 662 crystals arranged in a rectangular (x, y) grid and point at a focus 1.3 m beyond the nominal vertex. The endcap crystal front faces are square-shaped with 28.62-mm-long edges. The crystal depth is 24.7 radiation lengths X_0.

An additional detector, the CMS *Preshower* detector, identifies neutral pions with pseudorapidities $1.653 < |\eta| < 2.6$. Furthermore, it identifies electrons against minimum ionizing particles and with its high granularity helps to improve the electron and photon position resolution. Electromagnetic showers caused by electrons or photons initiate from lead radiators and are measured by silicon strip sensors sandwiched in-between two radiators. The sensors each cover active areas of 61×61 mm^2 divided into 32 strips of 1.9 mm pitch.

Energies below 500 GeV are resolved following a parametrization with dominant contributions from a stochastic term S, a noise term N, and a constant term C [85]:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \; . \tag{3.9}$$

Above 500 GeV, shower leakage from the rear of the calorimeter becomes significant and this simple parametrizaton of the energy resolution σ is not valid anymore.

The three contributions to the calorimeter energy resolution can be further broken down as follows:

Stochastic term The stochastic term is mainly determined from three contributions: The lateral shower containment fluctuates event-by-event ($1.5 - 2\%$ contribution). The number of released photoelectrons varies (2.1% contribution). The energy deposition in the preshower may differ from the silicon measurement ($1/E^{0.75}$ contribution).

Constant term Light collection is not uniform along the crystal ($< 0.3\%$). Miscalibration limits the optimal attainable energy resolution. Rear leakage influences the constant term only lightly: Charged particles leaking from the crystal rear side may directly trigger the avalanche photodiodes, but this effect has been determined to be small even at high electron energies.

Noise term Three dominant sources of noise can be identified: Noise from electronics, noise from digitization, or noise from pile-up. Electronics and digitization noise amount to 40 MeV per channel in the barrel and a transverse energy equivalent of 50 MeV in the endcaps. At low luminosities $\mathcal{L} = 2 \times 10^{33}$ cm^{-2}s^{-1}, the pile-up contribution to noise is small.

The above estimates were confirmed in test measurements with electron beams of energies $20 - 250$ GeV/c directed on a fully equipped barrel supermodule [85]. Typical

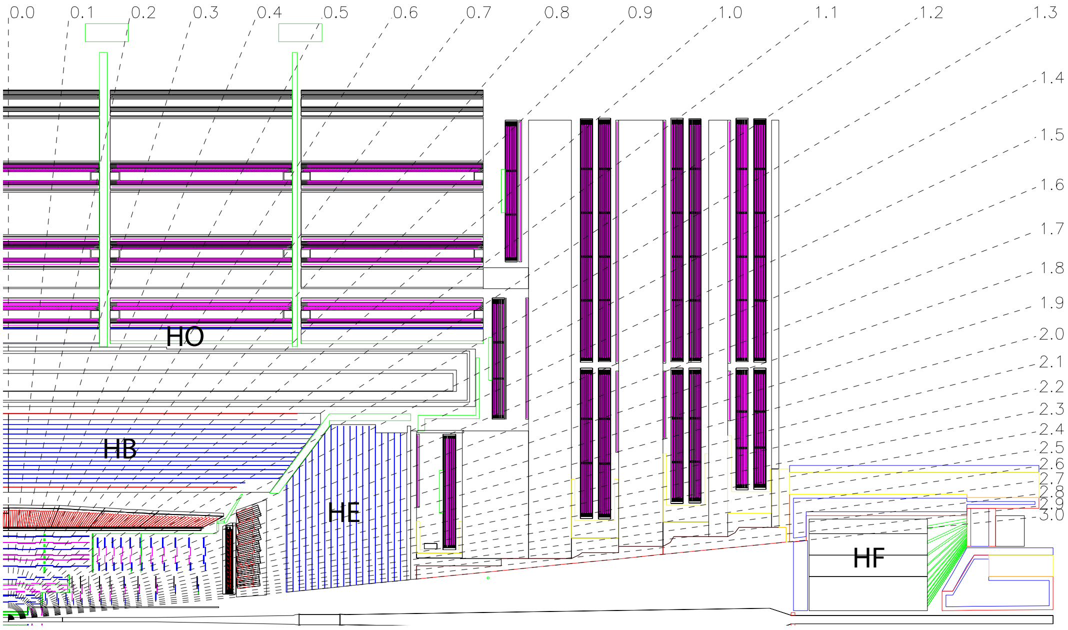


Figure 3.6: Longitudinal sketch of the CMS hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [85].

energy resolutions were found to be

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.3\%)^2 \;, \tag{3.10}$$

in agreement with the expected contributions.

3.2.6 Hadron Calorimeter

The CMS hadron calorimeter (HCAL) detects hadron jets and is an important component to measure missing transverse energy caused by neutrinos or exotic particles. A longitudinal schema of the CMS detector with the HCAL location is shown in Fig. 3.6 with dashed lines corresponding to fixed pseudorapidities. The hadron barrel calorimeter is located between the outer bound of the ECAL (at 1.77 m radial distance to the beam axis) and the inner limit of the magnet (at 2.95 m radial distance to the beam axis) constraining the total amount of material to absorb the hadronic shower. An additional hadron calorimeter placed outside of the magnet is supposed to absorb energy leaking beyond the magnet (*tail catcher*). The different HCAL subdetector parts are described in the following.

The pseudorapidity range $|\eta| < 1.3$ is covered by the *HCAL barrel* (HB), consisting of 36 identical azimuthal wedges that form two half-barrels. Individual wedges are further segmented into four sectors in azimuth, resulting in 72 towers in total, that are arranged in a projective (η, ϕ) geometry. Each wedge is constructed from flat brass (70% copper, 30% zinc) absorber plates aligned parallel to the beam axis. The absorber has a radiation length of $X_0 = 1.49$ cm and interaction length of $\lambda_I = 16.42$ cm. Readout is done by plastic scintillators, divided into 16 sectors in pseudorapidity, that have an (η, ϕ) segmentation of 0.087×0.087. At $\eta = 0$, the absorber is 5.82 interaction lengths thick, growing to 10.6 interaction lengths at $|\eta| = 1.3$. This is to be added to 1.1 interaction lengths from the electromagnetic calorimeter. 70 000 tiles together with wavelength shifting fibres read out light information in the CMS HCAL barrel part.

Jets emerging into the pseudorapidity regions $1.3 < |\eta| < 3$ are measured by the *hadron calorimeter endcaps* (HE). Typically, 34% of all produced particles fall into this pseudorapidity range [85]. Together with the high design luminosity of the LHC, this leads to high signal rates and radiation levels in the detector. The HCAL endcaps are located in the end parts of the CMS solenoid magnet and thus must not contain any magnetic material. They are built from a special type of cartridge brass (C26000) and attached to the muon endcap yokes. The absorber in the HCAL endcaps is designed such as to have a minimum of uncovered space between the HCAL barrel and endcap parts. The jet energy resolution in the endcaps is dominated by pile-up, magnetic field, and hadronization uncertainties. Single-particle energy resolution was thus not driving design considerations. The endcap region of HCAL and ECAL together are 10 interaction lengths long. Wavelength shifting fibres collect the scintillation light. Calorimeter cell sizes are $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ up to $|\eta| = 1.6$ and $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ in the forward region ($1.6 < |\eta| < 3$). The collected scintillation light is detected by multipixel hybrid photodiodes; these are only weakly sensitive to magnetic fields and exhibit a large dynamic range. Due to the high-radiation levels expected to hit the HCAL endcaps, the longitudinal segmentation has been designed accordingly: Each of the towers next to the beam line is divided and read out at three positions along its depth. Towers at larger distances to the beam are divided and read out at two positions along their depth.

The *hadron outer calorimeter* (HO) is placed in the central region ($|\eta| < 1.3$) outside of the solenoid. It catches the tails of hadron showers that are not fully contained within the ECAL and HCAL barrel part. The solenoid, being $1.4/\sin\theta$ interaction lengths thick, is an additional absorber for the outer HCAL (HO). Size and position adopt the HCAL barrel layout with cells of 0.087×0.087 in (η, ϕ) space. Wavelength-shifting fibres (multi-clad Y11 Kuraray) collect scintillation light. Testbeam measurements with 200 GeV pions showed that leakage losses are largest in the central region but can be recovered by the HO. Such a recovery is important as the cross section for events with at least one particle with transverse energy above 500 GeV is of the order of several picobarns. Without HO, these events would exhibit sizable amounts of "fake" missing transverse energy.

The hadron forward (HF) calorimeter (at 11.2 m longitudinal distance from the nominal vertex) covers pseudorapidities in the range $3 < |\eta| < 5.2$. This subdetector must deal with the highest particle flux of all HCAL parts: Each proton-proton collision deposits on average 760 GeV into the two forward calorimeters and 100 GeV into the other detector components. For the detector parts at $|\eta| = 5$, an integrated luminosity of 5×10^5 pb^{-1} corresponds to a radiation dose of 10 MGy and to a charged hadron rate above 10^{11} per cm^2 and second. The HF design was thus mainly driven by the need for radiation-hard active material: Quartz fibres (fused-silica core and polymer hard-cladding) are sufficiently radiation-hard to be used. Charged particles are detected if their energy surpasses the quartz fibres' Cherenkov threshold (190 keV for electrons). The HF thus mostly measures the electromagnetic shower component [99]. In pessimistic scenarios, optical transmission will have decreased to half of its original value after having received an accumulated dose of 10 MGy. Both HF parts are located 11.2 m from the nominal interaction point and segmented into towers with cell dimensions 0.175×0.175 in (η, ϕ).

3.2.7 Jet Reconstruction at CMS

In this subsection, we will summarize the basic concepts employed to reconstruct jets in the CMS calorimeters, where *calorimeter towers* are subject to *jet algorithms*. In a later section, we will present a study of jets reconstructed from charged particles (cf. Sec. 6.2). Other analyses use calorimeter energies together with track momenta or individually reconstructed particles as input to the jet algorithm.

Tower Definition and Thresholds

Calorimeter towers combine the measurements of ECAL and HCAL in appropriate (η, ϕ) bins. Due to the much finer granularity of the ECAL, tower sizes follow the HCAL segmentation of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ in the barrel region (and correspondingly larger in the endcap and forward regions). Towers thus follow a pattern in (η, ϕ) space, that is projective to the nominal interaction point. Tower energies are obtained from summing all contributing cells with energies above pre-configured thresholds, typically of the order of $E_T = 0.5$ GeV. In total, 4 176 towers are clustered by a number of jet algorithms, where towers are treated as massless particles with energies equal to the tower energy and emerging from the nominal interaction point to the center of the tower.

Jet Algorithms

We will describe examples for two widely-used classes of jet definition at hadron colliders: the $k_\perp$ algorithm [44,45], an example for sequential recombination jet algorithms, and the SISCone algorithm [46], an example for cone jet algorithms.

The $k_\perp$ *algorithm* starts with "protojets", such as calorimeter towers, tracks, or generator particles. The algorithm is characterized by a parameter D and identifies jets in a recursive procedure:

1. One defines $d_i = E_{TI}^2 D^2$ for each protojet and $d_{ij} = \min\left(E_{Ti}^2, E_{Tj}^2\right) \times ((\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2)$ for each pair of protojets.
2. The smallest of all $\{d_i\}$ and $\{d_{ij}\}$ is labelled $d_{\min}$.
3. If $d_{\min}$ is a d_{ij}, protojets i and j are merged into a new protojet by directly adding the protojet four-momenta (longitudinal E-scheme recombination).
4. Otherwise, i. e. if $d_{\min}$ is a d_i, protojet i is removed from the list of protojets and added to the list of jets.
5. This procedure is continued until no protojets are left.

The *SISCone algorithm* (*S*eedless *I*nfrared-*S*afe *C*one algorithm) is an infrared- and collinear-safe cone jet definition. It starts by identifying all stable circles (or cones) in the (η, ϕ) plane. A cone is stable if the direction of the four-momentum vector sum inside the cone coincides with the cone center. Ambiguities due to overlapping stable cones are resolved with a Tevatron run II type [100] split-merge prescription based on the transverse momentum shared by the overlapping jets.

3.2.8 Plans for Jet Energy Corrections at CMS

CMS has developed a factorized multi-level approach to relate measured calorimeter jet energies to final-state-particle jet energies. Note that any correction beyond particle level is ambiguous: Different hadronization models or different underlying event descriptions (or both) result in a range of possible predictions. We will therefore not describe such approaches here.

Calorimeter cells have been calibrated using testbeam data. The approaches described in the following calibrate calorimeter cells in-situ, i. e. without requiring dedicated runs or measurements. Typically, jet energy corrections are determined in $2 \to 2$ processes, where both final-state-objects must conserve transverse momentum (p_T-balance). p_T-balance-techniques are employed in dijet events, $\gamma+$jet-events and $Z+$jet-events. A schematic equation to determine the corrected jet energy reads:

$$\begin{aligned} \text{Corrected Jet Energy} = (&\text{Measured Jet Energy} - \text{Offset}) \\ &\times \text{Relative } \eta \text{ Correction} \\ &\times \text{Absolute } p_T \text{ Correction} \\ &[\times \text{EMF Correction}] \\ &[\times \text{Flavor Correction}] \; . \end{aligned} \tag{3.11}$$

Squared brackets indicate optional corrections. All terms are explained in more detail in the following.

Offset Correction

Offsets to the energies measured in the CMS calorimeters are introduced from electronics noise and pile-up. The *offset correction* subtracts average energies from the measured jet energy: An area is attributed to a given jet [101, 102]; next, the correction method subtracts the corresponding sum of average cell energies measured in zero-bias events. The offset correction grows roughly proportionally with the number of pp-scatters per bunch crossing which is a function of the instantaneous luminosity.

Relative Correction: Pseudorapidity Dependence

The jet energy response measured with CMS is not uniform in pseudorapidity. The relative η correction accounts for this effect and flattens the jet energy response with respect to pseudorapidity. Corrections are determined from dijet events either by employing the event's Monte Carlo truth or by imposing p_T-balance: Namely, the jet energy response in a given pseudorapidity bin is compared with the jet energy response in a control region.

Absolute Correction: Transverse Momentum Dependence

Corrections depending on the jet transverse momentum are established by comparing jet transverse momenta either to the Monte Carlo truth or to the transverse momenta of objects measured in other detector parts, e. g. photons or Z-bosons. p_T-balance techniques are employed in analyzing $\gamma + \text{jet}$- and $Z + \text{jet}$-events.

Optional Correction with respect to Electromagnetic Energy Fraction

Ideally, the jet energy response should be independent on the jet's relative content of electrons and photons. However, most calorimeters (including the CMS ECAL and HCAL) are *non-compensating*, i. e. electrons and hadrons are met with different energy responses by the detector. The fraction of electromagnetic energy (EMF) contained in a jet is correlated with the ratio of detector response for electrons and hadrons. Typical improvements in jet energy resolution when applying EMF corrections are of the order of $5 - 10\%$ [103].

Optional Flavor Correction

Flavor differences in jet fragmentation of light-quark- and gluon-induced jets and flavor variations in semileptonic decays of heavy quarks potentially lead to different jet energies. It should be noted however that strictly speaking, corrections depend on the assumed process type and on the derived jet flavor. To illustrate the impact of different parton flavors, compare jets from heavy quarks with jets from light quarks: b jets typically produce more charged hadrons than light quark jets do; yet, charged hadrons from b jets typically have on average less momentum than charged hadrons from light quark jets have. In addition, 22% of B-hadrons decay semileptonically, producing neutrinos;

these do not contribute to the jet energy measurement. In summary, the charged hadron energy fraction of b jets is smaller than the corresponding fraction for light quark jets. Flavor corrections are determined from MC truth, $t\bar{t}$-events, γb-events, $b\bar{b}Z$-events and $Z \rightarrow b\bar{b}$-events.

3.2.9 Forward Detectors

Two additional detector components cover the very forward region of the CMS experiment: *CASTOR* (Centauro And Strange Object Research) and *ZDC* (Zero Degree Calorimeter). The Castor detector is installed 14.38 m away from the nominal vertex, at pseudorapidities $5.2 < |\eta| < 6.6$; ZDC measures energies of neutral particles with pseudorapidities above $|\eta| \geq 8.3$ in 140 m distance to the interaction region.

The need for radiation hardness (as discussed in the context of the hadron forward calorimeter) holds for CASTOR and ZDC. Both detectors are tungsten-quartz Cherenkov calorimeters and are 10 (CASTOR) and 7.5 (ZDC) interaction lengths deep.

3.2.10 Muon System

The study of rare signatures involving muons, e. g. the decay of the Standard-Model-Higgs-boson into two Z bosons decaying further into 4 muons, $H \rightarrow ZZ \rightarrow \mu\mu\mu\mu$, is made difficult by the overwhelming background rate expected at the LHC. The CMS muon system identifies, measures the momentum of and triggers on muons over the entire kinematic range of the LHC. Muons are identified by three types of gaseous tracking chambers: *Drift tubes* (DT) in the barrel region ($|\eta| < 1.2$), *cathode strip chambers* (CSC) in the endcap region ($0.9 < |\eta| < 2.4$) and *resistive plate chambers* (RPC) in barrel and endcap regions ($|\eta| < 1.6$). An optical alignment system monitors the required precision of $75 - 200$ µm in the endcaps and $150 - 350$ µm in the barrel.

The four layers of drift tubes contain a gas mixture of 85% Ar and 15% CO_2 and are designed such that the drift time does not exceed 380 ns. Albeit large compared to the bunch crossing time of 25 ns, this time is sufficiently small because the expected muon rate in the barrel parts is low. Tubes were preferred to multiwire chambers to protect against broken wires and to suppress cross-talk between wires. The (r, ϕ)-resolution strived for is 100 µm that is of the order of the track uncertainty due to multiple scattering at $p_T = 200$ GeV/c.

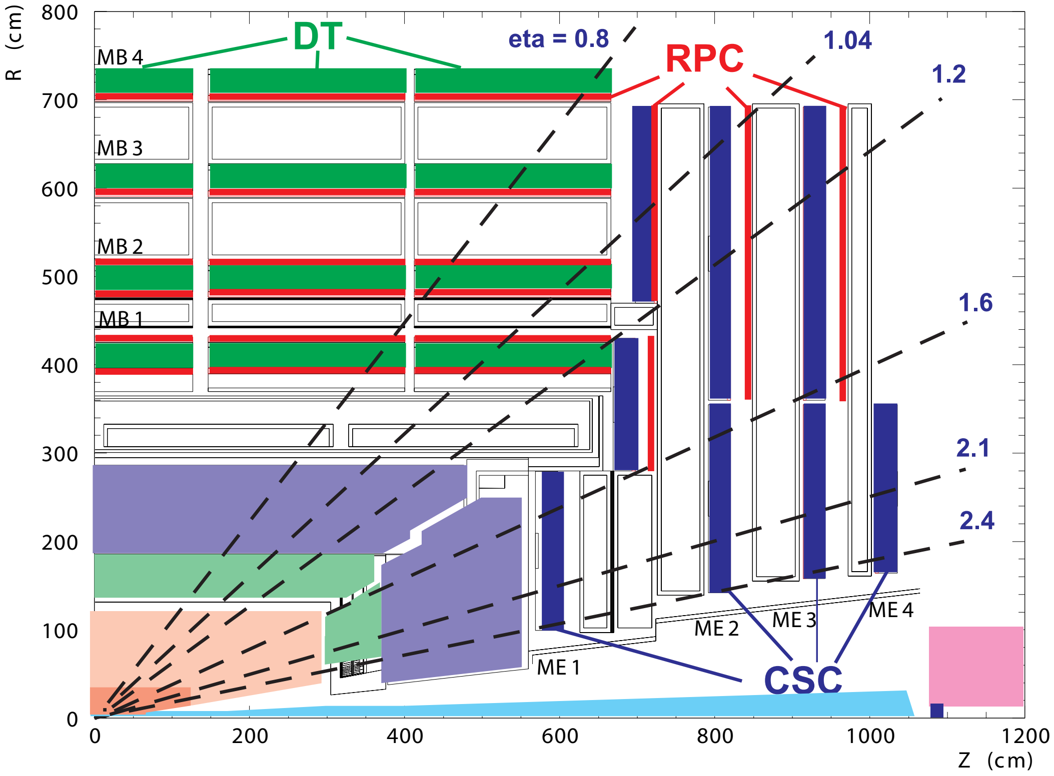


Figure 3.7: Longitudinal quarter view of CMS muon system layout for initial low luminosity running [92]. The RPC system is limited to $|\eta| < 1.6$ in the endcap, and for the CSC system only the inner ring of the ME4 chambers have been deployed.

468 trapezoidal cathode strip chambers (CSC), multiwire proportional chambers, make up the CMS detector's endcap muon system at LHC startup. The CSCs cover the pseudorapidity region $0.9 < |\eta| < 2.4$ in 10° or 20° segments in azimuth. The region $0.9 < |\eta| < 1.2$ is instrumented with both drift tubes and cathode strip chambers; muons emerging into the region $1.2 < |\eta| < 2.4$ cross only CSCs. At the first trigger step, the cathode strip chambers resolve the (r, ϕ) coordinate to 2 mm precision, but the precision is improved offline to values between 75 µm and 150 µm.

Resistive plate chambers (RPC) are installed in the barrel as well as in the endcaps to trigger on muons and associate the muon candidate with the respective bunch crossing. The CMS RPCs, gaseous parallel-plate detectors, operate in avalanche mode with two 2 mm-wide gaps sharing a common pickup read-out strip. The barrel region is instrumented with six RPC layers; three layers are installed in the endcap region (four at full luminosity).

3.2.11 Trigger

The LHC will produce an enormous amount of experimental data: Proton bunches cross at a frequency of 40 MHz each corresponding to several collisions producing particles impinging on the CMS subdetectors. The resulting rate is much too high to store every event on tape - the two-step CMS trigger system thus reduces the event rate to a value acceptable for further processing.

The Level-1 Trigger architecture is depicted in Fig. 3.8: Custom-designed and programmable electronics reduce the 40 MHz event rate to a L1 output rate of maximally 100 kHz. High resolution data are stored in pipelines on the front-end electronics and low resolution data are input to the L1 decision on keeping the event.

The High-Level Trigger (HLT) can analyze the entire high-resolution data and runs highly complex algorithms if necessary. Being a software trigger, the HLT is flexible with respect to possible changes in algorithms over the course of luminosity upgrades or adaptations to the possible discovery of new physics. L1 and HLT together are able to reduce the total event rate by a factor of 10^6 to an average output rate of 40 Hz.

As can be seen from Fig. 3.8, the Level-1 trigger decision is based on calorimeter towers and muon chamber information only, no track triggers are employed at the first trigger step. Three layers of increasing complexity build up the Level-1 trigger: local (e. g. calorimeter towers, track segments in muon chambers), regional (a combination of

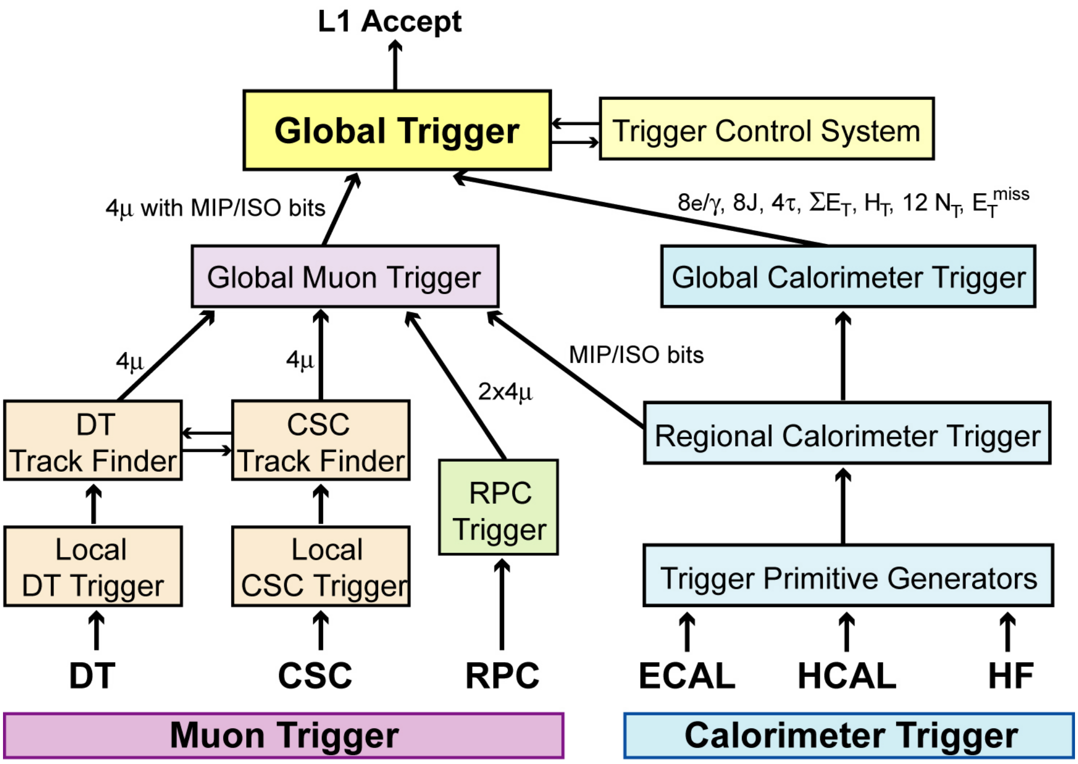


Figure 3.8: Architecture of the CMS Level-1 Trigger [85].

the former) and global components. The Global Calorimeter and Global Muon Trigger information is evaluated by the Global Trigger which either passes the event to the High-Level Trigger step or rejects the event altogether.

The calorimeter trigger employs trigger towers of areas 0.087×0.087 in (η, ϕ) up to $|\eta| = 1.74$; trigger towers in the forward and backward region have larger areas. From trigger towers, electron and photon candidates are generated by the regional calorimeter trigger. This trigger also determines the transverse energy sum, τ candidates, and information on minimum-ionizing particle and isolation properties. All calorimeter trigger objects are finally ranked by the *Global Calorimeter Trigger* calculating jets, total transverse energy and scalar E_T sum of jets above a configurable threshold. Also, the highest-ranked electron and photon candidates together with their isolation properties are identified.

All three muon system components contribute to gather trigger information: Track segments in azimuth and hit patterns in pseudorapidity are identified by the drift tubes; the cathode strip chambers provide three-dimensional track segments. Both trigger primitives are subsequently evaluated by the Regional Muon Trigger; this trigger performs a rough track reconstruction. In parallel, resistive plate chambers reconstruct track candidates based on RPC hit patterns only. The three subdetectors' trigger information is combined and evaluated by the *Global Muon Trigger*.

Trigger objects created by the Global Calorimeter and the Global Muon Triggers are passed to the *Global Trigger* which decides whether to continue processing the event. Input objects are electrons, photons, muons, hadronic jets, and τ jets. Other input can be total and missing transverse energies, scalar E_T sum of jets above a configurable threshold and jet multiplicities. Should the Global Trigger accept an event for further processing, information stored in the pipelines is sent to the HLT farm; there, the high-granularity information of the entire event is processed. There is an estimated total detector deadtime of less than 1% (at 100 kHz L1 output rate) during the time that the pipelines are emptied.

3.2.12 Data Acquisition

Together with the trigger system, the CMS Data Acquisition (DAQ) system collects and analyzes electronic signals from the CMS detector front-ends at the LHC bunch crossing frequency of 40 MHz, reduced to 100 kHz after having passed the Level-1 trigger. The CMS DAQ will have to read out 1 MB of zero-suppressed data per bunch crossing at the nominal LHC luminosity of 10^{34} cm^{-2}s^{-1}. With the L1 output rate of 100 kHz, the data flow amounts to 100 GB/s from 650 sources; these data are passed to the High-Level Trigger software farm.

3.2.13 Computing

All data has to be stored for the lifetime of the experiment. In particular, data analysis of all running periods has to be supported at all times. Consequently, transfer, storage and analysis of large data sets is foreseen by the CMS Computing Model. Real-time (raw) detector data passes pattern recognition, filtering and data reduction steps. Physics analysis activities are carried out regardless of the physical location of the physicist at several computing centers distributed throughout the world. Furthermore, theory predictions have to be simulated and distributed on large scales, so they can be compared to experimental data. Run conditions, calibration information, and additional non-event data are made accessible via databases.

The CMS offline computing model comprises a fully distributed system throughout worldwide computing centers. Data storage, processing, and analysis will mainly happen outside CERN at designated sites. The Worldwide LHC Computing Grid (WLCG) provides the necessary services to the LHC experiments to maintain such an infrastructure. The CMS computing model's key components shall be described in the following.

Event processing, filtering and analysis are applications common to offline and online computing. The aim of the CMS *application framework* is to provide well-defined interfaces that decouple physics analyses from event I/O, user interfaces and other constraints. This is accomplished by interfacing software modules to one standard application. All software modules may filter, analyze, or modify the central entity of the CMS data model, the *Event*. Communication between modules also happens exclusively through the Event. A user who wishes to process data (real or simulated) will have to define a sequence of modules to be executed. The application framework hides complexity such as accessing global services through database interfaces, etc.

Several *event formats* with well-defined data content exist that are typical for different processing steps of a data sample:

RAW: This type of event content consists of the entire detector information, including trigger decisions and metadata, for instance on the run conditions. According to the trigger signature, RAW data is classified into distinct *primary datasets*.

RECO: Reconstructed events are obtained after RAW data has been passed through pattern recognition and compression algorithm steps. Thus, in this event content, high-level physics objects are contained, including all basic detector input they were reconstructed from.

AOD: The *A*nalysis *O*bject *D*ata event content consists of high-level physics objects together with information needed to refit the kinematics. AOD data is obtained from RECO data by means of filtering.

In addition, event reconstruction relies on information about non-event data, e. g. construction data, equipment management data, configuration data, and conditions data. Conditions data includes information on alignment and calibration constants as well as information on the detector status.

The CMS computing needs exceed all presently available resources of a single site. Computing resources provided by collaborating institutes worldwide are thus integrated into a common hierarchical architecture of *Tiered* centers: One Tier-0 center at CERN,

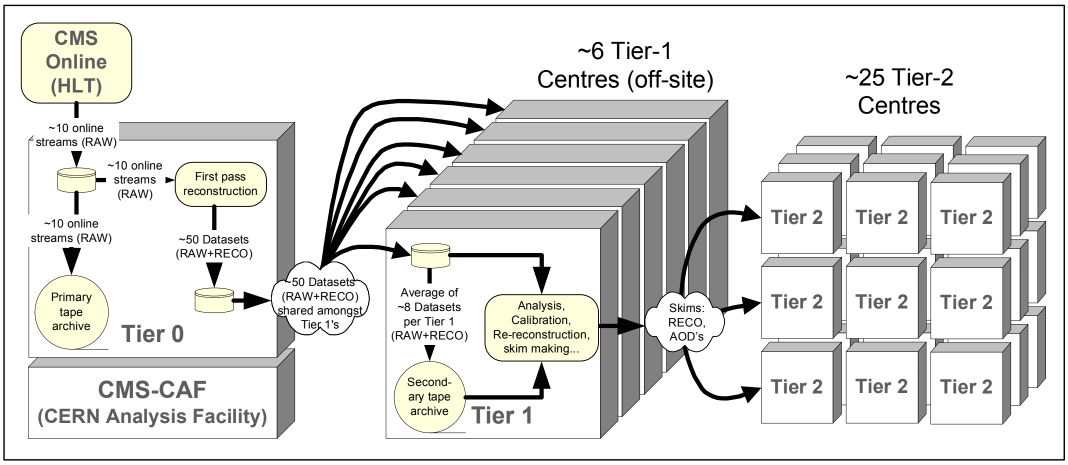


Figure 3.9: Dataflow between CMS Computing Centres [85].

six Tier-1 centers at large computing facilities and approximately[3] 25 Tier-2 centers at medium to large computing facilities. The dataflow between tiers is sketched in Fig. 3.9 and described in the following.

The CERN *Tier-0* center records experimental data at a mass storage system, promptly reconstructs RAW data into first-pass RECO datasets, and exports a copy of the RAW data to *Tier-1* centers for re-reconstruction. The Tier-1 centers are operated around the clock and provide large computing farms, dedicated mass storage, and fast network links to each other and to the Tier-0. RAW data are stored permanently to offer reconstruction whenever necessary, for instance if improved alignment and calibration constants are available. In addition, skimming and data-intensive central analysis applications run on the Tier-1 with the output being transferred to associated *Tier-2* centers. These centers support the bulk of user analysis and thus have to grant access to large computing farms to all CMS users with a valid grid certificate. Offline calibration and alignment studies as well as detector analyses are typically performed at the Tier-2 centers as well. Third, Tier-2s produce large amounts of simulated data that are transferred to the associated Tier-1 center for storage. In order to provide fast feedback on key analyses, the CERN Analysis Facility (CAF) provides large CPU resources together with fast access to the full data.

The CMS-specific *grid computing services* operate using gLite [104] and OSG [105] middleware tools that interface storage systems and CPU resources in a standardized fashion. The Worldwide LHC Computing Grid provides an infrastructure for job sub-

---

[3] Some Tier-2s are virtual centers, i. e. they are a confederation of two or more physical computing centers, that individually do not have enough resources to constitute a full Tier-2.

missions and remote data access. It does so, ensuring robust security and accounting. Nonetheless, some CMS-specific software is needed on sites. The *Dataset Bookeeping System* catalogues existing event data. Existing event data is mapped to its physical location or locations (if more than one copy of a dataset exists) with the help of the *Data Location Service*. Each site employs a *Local File Catalogue* translating a logical file name to the actual path in the respective file system. CMS-specific tools for large-scale data transfers schedule, monitor and verify the data transfer. Large-scale data processing is performed with specialized tools, comprised of automated job managers. Since remote analysis introduces additional layers of complexity to a physics analysis, dedicated tools have been developed to facilitate job submission, monitoring, and retrieval [106].

Chapter 4

Simulated Samples and Triggers

This analysis makes use of several fully simulated data samples that were produced by the CMS collaboration during two major computing challenges in 2008: The Computing, Software, and Analysis challenge CSA08 [107], carried out in May 2008, tested the full scope of offline data processing and analysis in a manner as close as possible to LHC startup conditions. With the produced samples, the impact of simulated detector misalignment and miscalibration on the reconstruction level results is evaluated. As part of a second production effort in 2008, fully simulated samples were produced with updated detector conditions, such as a new geometry and a new magnetic field (SUMMER08 [108]). Also, different underlying event models were processed by the full detector simulation. In this chapter, we summarize the parameters and software versions used for the production of the relevant Monte Carlo samples.

4.1 CSA08: 2008 Computing, Software, and Analysis Challenge

CSA08 consisted of a collection of CMS exercises intended to test the full scope of data handling and analysis activities that is needed for LHC data-taking operations in 2009. The challenge was structured into a part overlapping with the Common Computing Readiness Challenge (CCRC08) [109], and other activities taking place at a later time in case LHC data-taking was not to commence as the machine schedule at the time foresaw. CSA08 was the first full-scale offline challenge with large statistics that was carried out as close as possible to the conditions expected at LHC startup. The data samples simulated during CSA08 constitute a precious testing ground of how well the

Table 4.1: LHC commissioning scenarios assumed during CSA08. Integrated luminosities correspond to six days of running at 100% efficiency.

Scenario	Bunch Pattern	Instant. Luminosity $\mathcal{L}$	$\int \mathcal{L}dt$
S43	43×43	2×10^{30} cm^{-2}s^{-1}	1 pb^{-1}
S156	156×156	2×10^{31} cm^{-2}s^{-1}	10 pb^{-1}

underlying event activity can be measured with early LHC data. Two representative LHC commissioning scenarios, S43 and S156, were studied for CSA08, see Tab. 4.1. Fully simulated data samples were created for both settings. Running the LHC for six days at 100% efficiency corresponds to integrated luminosities of 1 pb^{-1} for scenario S43 and 10 pb^{-1} for scenario S156. Both scenarios are averages over several machine configurations.

Monte Carlo samples were processed through the full detector simulation and event reconstruction chain. Sample production and analysis was carried out using versions CMSSW_2_0_X of the CMS software. Proton-proton collisions at $\sqrt{s} = 10$ TeV in a detector with a 4 T magnetic field were simulated. Detector readout was simulated in zero-suppression-mode. Pile-up was not simulated, since in neither S43 nor S156, the instantaneous luminosity is large enough for a significant number of pile-up events to be produced. One consequence is that the CMS tracker is operated in "peak mode" [85]. The event types and processes interesting for the underlying event analysis are listed in Tab. 4.2. All of these samples were produced with PYTHIA version 6.409 and underlying event tune D6T [110] - a set of model parameters that leads to a successful description of CDF data on the underlying event in charged-jet topologies. Parameters of D6T are listed in Tab. 4.3.

The samples were simulated with phase space thresholds that were guided by the CMS high-level trigger thresholds: Lower thresholds on $\hat{p}_\perp$, the transverse momentum of partons after a $2 \rightarrow 2$ scatter, were chosen such that respective high-level triggers were fully efficient, see Tab. 4.2. No upper limits on $\hat{p}_\perp$ were imposed. As a consequence, when combining results obtained from several samples, care has to be taken to avoid double-counting: We choose to discard any events from a given sample with a $\hat{p}_\perp$ larger than the lower threshold of the subsequent sample.

Table 4.2: Monte Carlo sample input to the underlying event analysis. Samples were produced during Csa08 with Pythia version 6.409 and underlying event tune D6T.

Sample Name	$\hat{p}_\perp$ threshold	Scenario	Events	Cross Section
MinBias	None	S43, S156	25 000 000	75.3 mb
JetET20	30 GeV/c	S43, S156	4 000 000	101.6 µb
JetET30	45 GeV/c	S43, S156	4 000 000	21.6 µb
JetET50	75 GeV/c	S43, S156	4 000 000	2.5 µb
JetET80	120 GeV/c	S43, S156	4 000 000	323.7 nb
JetET110	160 GeV/c	S43, S156	4 000 000	88.7 nb
JetET150	220 GeV/c	S156	4 000 000	17.1 nb

Table 4.3: Relevant parameters, their purpose [39] and values for Pythia 6.4 and underlying event tune D6T [110]. With the selected energy extrapolation, the $p_\perp$ cutoff parameter for additional interactions amounts to $p_{\perp 0}(10\text{ TeV}) = 2.39$ GeV/c at collision energies of $\sqrt{s} = 10$ TeV.

Parameter	Purpose	Tune D6T
Mstp(51)	PDF set	Cteq6l
Mstp(81)	Choose "old" MI model	1
Parp(82)	Cutoff for multiparton interactions $p_{\perp 0}$ (GeV/c)	1.8387
Mstp(82)	Choose double-Gaussian matter distribution	4
Parp(83)	Radius of double-Gaussian core	0.5
Parp(84)	Hadronic matter fraction in double-Gaussian core	0.4
Parp(85)	Gluon production in MI	1.0
Parp(86)	Gluon production in MI (allow closed gluon loops)	1.0
Parp(89)	Reference energy scale (TeV)	1.96
Parp(90)	Energy rescaling power	0.16
Parp(62)	Space-like shower cutoff Q	1.25
Parp(64)	Space-like shower scale factor	0.2
Parp(67)	Maximum parton virtuality	2.5
Mstp(91)	Choose Gaussian for primordial k_T distribution	1
Parp(91)	Gaussian width of primordial k_T	2.1
Parp(93)	Upper cutoff for primordial k_T	15.0

Table 4.4: CSA08 Trigger paths, E_T(calorimeter jet) thresholds and prescale factors for $\mathcal{L} = 2 \times 10^{30}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ (S43) (upper half of table) and $\mathcal{L} = 2 \times 10^{31}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$ (S156) (lower half of table). The totally available L1 and HLT bandwidths amount to 12 kHz and 150 Hz. The expected HLT rates of all considered triggers are 26.5 Hz (S43) and 15.5 Hz (S156) [111].

	Path Name	Level-1 Condition	Threshold	L1 Prescale	HLT Prescale
S43	HLTZeroBias	L1_ZeroBias	None	300 000	1
	HLT1jet30	L1_SingleJet15	30 GeV	10	20
	HLT1jet50	L1_SingleJet30	50 GeV	1	10
	HLT1jet80	L1_SingleJet50	80 GeV	1	5
	HLT1jet110	L1_SingleJet70	110 GeV	1	1
	HLT1jet180	L1_SingleJet70	180 GeV	1	1
S156	HLTZeroBias	L1_ZeroBias	None	1 000 000	1
	HLT1jet30	L1_SingleJet15	30 GeV	1 000	10
	HLT1jet50	L1_SingleJet30	50 GeV	100	10
	HLT1jet80	L1_SingleJet50	80 GeV	1	50
	HLT1jet110	L1_SingleJet70	110 GeV	1	10
	HLT1jet180	L1_SingleJet70	180 GeV	1	1

After the first reconstruction pass (*prompt reconstruction*), data samples were transferred to the Tier-1 centers (cf. Fig. 3.9 for an overview of the dataflow between CMS computing centers). In the meantime, alignment and calibration constants were determined at the CERN Analysis Facility and copied to the Tier-1 centers as well. With improved constants, data samples were re-reconstructed and transferred to Tier-2 centers. It was at these Tier-2 centers, that physics analyses, e. g. the underlying event exercise, were carried out. The high-level trigger paths considered in the underlying event analysis during CSA08 are listed in Tab. 4.4. The analysis uses one zero-bias trigger and several single-jet triggers. Level-1 conditions and transverse energy thresholds are shown together with expected rates for the two scenarios under study during CSA08. Note that the rate estimates include prescale factors for the various paths at the Level-1 or at the high-level trigger step.

Table 4.5: Csa08: Fraction of accepted sample events per HLT stream. Numbers shown for S156 scenario. Only if at least 10% of all events in a given sample are accepted by a high-level trigger, is the sample included into the trigger stream composition.

	HLTZeroBias	HLT1jet30	HLT1jet50	HLT1jet80	HLT1jet110	HLT1jet180
MinBias	1	< 0.01	< 0.01	< 0.01	< 0.01	0
JetET20	1	0.81	0.17	< 0.01	< 0.01	0
JetET30	1	0.98	0.69	0.11	< 0.01	< 0.01
JetET50	1	0.99	0.98	0.73	0.24	< 0.01
JetET80	1	0.99	0.99	0.98	0.87	0.11
JetET110	1	0.99	0.99	0.99	0.99	0.67

Csa08: **Composition of Trigger Streams**

The simulated Monte Carlo samples are turned into *trigger streams*. A trigger stream is a collection of all events that are accepted by a given trigger. This mapping of events is done by first combining all samples weighted by their respective cross sections. For instance, MinBias events enter a given HLT spectrum with a relative weight with respect to JetET20 events of

$$\frac{\sigma(\text{MinBias})}{N(\text{Events})(\text{MinBias})} \Big/ \frac{\sigma(\text{JetET20})}{N(\text{Events})(\text{JetET20})} \approx 100 \; . \tag{4.1}$$

This implies that to some degree, spectra might be distorted by limited available Monte Carlo statistics: E. g. only six out of 25 million events in the entire Csa08 MinBias sample satisfy HLT1jet110. Therefore, samples are only included in a given HLT stream if at least 10% of the sample events contribute. Next, all events satisfying a given high-level trigger are combined into common trigger streams. These correspond to the actually measured data streams. Table 4.5 summarizes the fraction of events accepted by the high-level triggers under study for each of the Csa08 Monte Carlo samples: All JetET20 events are accepted by HLTZeroBias, 81% of JetET20 events are accepted by HLT1jet30, 17% of JetET20 events are accepted by HLT1jet50, etc. The numbers correspond to scenario S156, but results for S43 are similar. In particular, the sample subsets considered for a given trigger stream are the same for S43 and S156.

4.2 SUMMER08: Full Simulation Standard-Model Production at $\sqrt{s} = 10$ TeV

During the SUMMER08 [108] production effort, Monte Carlo samples were processed through the full detector simulation and event reconstruction chain. In contrast to the CSA08 effort, no misalignment or miscalibration of the detector was simulated, but ideal conditions were assumed. Sample production and analysis was carried out using versions CMSSW_2_1_X of the CMS software. Most notably, this software cycle includes a backward incompatible change of the tracker format. In general, CMSSW_2_1_X is understood as a software milestone in implementing all essential functionalities foreseen for first data-taking: the final detector geometry, an updated magnetic field map, data quality monitoring, and certification and common analysis tools. In addition, the adopted version 9.1p2 of GEANT4 [112, 113], a program to simulate the passage of particles through matter, includes a number of new physics models.

Proton-proton collisions at $\sqrt{s} = 10$ TeV in a detector with 3.8 T magnetic field were simulated. No pile-up events were included in the event simulation. Reconstruction was carried out at Tier-1 centers before copying the reconstructed data samples to the Tier-2 centers, where physics analyses were performed. Those samples that were analysed as part of the underlying event exercise are listed in Tab. 4.6. The studied samples were produced with either PYTHIA version 6.416 and underlying event tune D6T, or with HERWIG++ version 2.2.0. Note that during the SUMMER08 exercise, a minimum-bias sample was only produced with PYTHIA 6.416, since HERWIG++ 2.2.0 does not include[1] the possibility to simulate minimum-bias events. In order to compare both event generators in similar phase space regions, we choose to not include the SUMMER08 PYTHIA minimum-bias sample in the stream under study.

Parameters of D6T have already been listed in Tab. 4.3; default parameters of the HERWIG++ underlying event tune are listed in Tab. 4.7. Fig. 4.1 illustrates different

---

[1] The simulation of minimum-bias events with HERWIG++ has only been implemented recently, with version 2.3.0 [114].

Table 4.6: Monte Carlo samples of interest to the underlying event analysis produced during Summer08 [108]. QCDDiJetPtXXtoYY samples created with Pythia version 6.416 Tune D6T, HerwigQCDPtNN samples created with Herwig++ version 2.2.0.

Sample Name	$\hat{p}_\perp$ interval	Events	Cross Section
QCDDiJetPt15to20	15 – 20 GeV/c	100 000	949.4 µb
QCDDiJetPt20to30	20 – 30 GeV/c	100 000	401.0 µb
QCDDiJetPt30to50	30 – 50 GeV/c	100 000	94.7 µb
QCDDiJetPt50to80	50 – 80 GeV/c	100 000	12.2 µb
QCDDiJetPt80to120	80 – 120 GeV/c	50 000	1.6 µb
QCDDiJetPt120to170	120 – 170 GeV/c	50 000	256.0 nb
QCDDiJetPt170to230	170 – 230 GeV/c	50 000	48.3 nb
QCDDiJetPt230to300	230 – 300 GeV/c	50 000	10.6 nb
QCDDiJetPt300to380	300 – 380 GeV/c	50 000	2.6 nb
QCDDiJetPt380to470	380 – 470 GeV/c	50 000	722.1 pb
HerwigQCDPt15	$\hat{p}_\perp > 15$ GeV/c	750 000	1.5 mb
HerwigQCDPt30	$\hat{p}_\perp > 30$ GeV/c	600 000	111.7 µb
HerwigQCDPt80	$\hat{p}_\perp > 80$ GeV/c	500 000	2.0 µb
HerwigQCDPt170	$\hat{p}_\perp > 170$ GeV/c	400 000	68.0 nb
HerwigQCDPt300	$\hat{p}_\perp > 300$ GeV/c	300 000	4.1 nb
HerwigQCDPt470	$\hat{p}_\perp > 470$ GeV/c	200 000	354.3 pb

Table 4.7: Relevant parameters and values for Herwig++ 2.2.0 underlying event tune [79]. The hadronic matter distribution is chosen proportional to the electromagnetic form factor. Although the tune has been obtained from fits to CDF data, the $p_\perp$ cutoff for additional interactions is kept unchanged for the simulation of pp collisions at center-of-mass energies of $\sqrt{s} = 10$ TeV.

Parameter	Herwig++ UE Tune
PDF set	Mrst2001Lo
Cutoff for multiparton interactions $p_{\perp\mathrm{min}}$ (GeV/c)	3.1
Inverse hadron radius squared μ^2 (GeV^2)	1.8

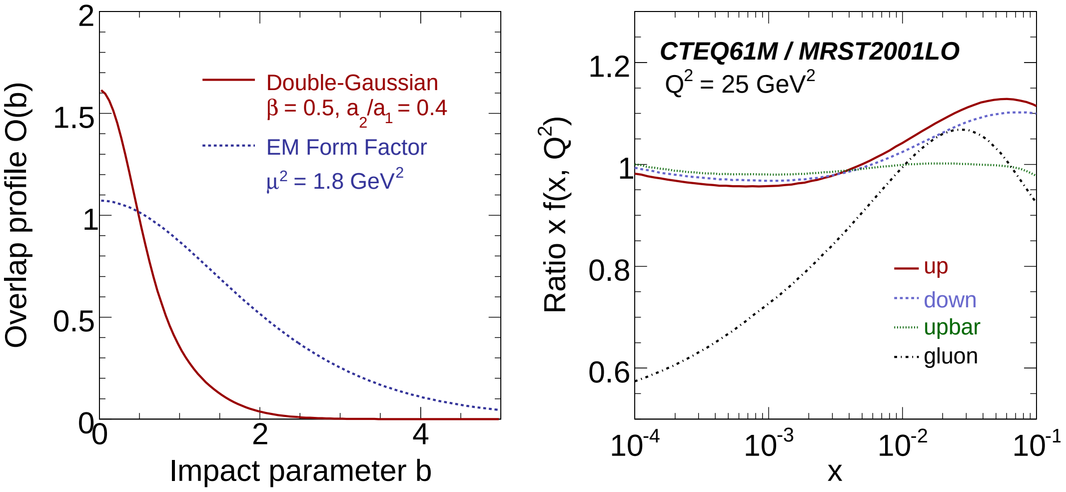


Figure 4.1: Resulting overlap profiles (left) and ratios of parton distribution sets (right) chosen for underlying event tunes Pythia D6T (double-Gaussian, Cteq6.1m) and Herwig++ default (electromagnetic form factor, Mrst2001lo). Overlap profiles normalized to equal integrals in the integration range $0 < b < 1$. PDF ratios shown for virtuality $Q^2 = 25$ GeV^2.

choices of the underlying event model tunes for Pythia and Herwig. The left panel shows matter overlap profiles for the matter distributions chosen: a double-Gaussian (D6T) matter distribution and a matter distribution following the electromagnetic form factor (Herwig++). Both distributions are normalized such that the integral in the range $0 < b < 1$ equals unity. In the right panel, ratios between the parton density distributions selected by each tune (Cteq6.1m for D6T, Mrst2001lo for Herwig++) are shown at virtualities $Q^2 = 25$ GeV^2 for u-, d-, and $\bar{u}$-quarks as well as for gluons. Most notably, the Cteq6.1m gluon distribution is only 40% of the corresponding Mrst2001lo parametrization for $x < 10^{-3}$.

Table 4.8 lists two trigger menus that have been updated with respect to the Csa08 trigger menus. The 8E29 menu is designed for startup detector misalignment and miscalibration and for startup luminosities [115]. It consists of triggers with the loosest possible requirements on thresholds, isolation, pixel-matching, etc., and includes a number of commissioning triggers. 8E29 makes use of no jet energy corrections. This menu should be considered as a candidate menu to be used for the first real collisions. The 1E31 menu is designed for a Monte Carlo exercise in late 2009 [116]. It contains thresholds, isolation, pixel-matching that have been tuned on Summer08 data samples. Monte-Carlo-driven, pseudorapidity-dependent jet energy corrections are applied. If the CMS

Table 4.8: High-Level Trigger paths for core menus: E_T(calorimeter jet) thresholds and prescale factors for $\mathcal{L} = 8 \times 10^{29}$ cm^{-2}s^{-1} (8E29, upper half of table) [115] and $\mathcal{L} = 1 \times 10^{31}$ cm^{-2}s^{-1} (1E31, lower half of table) [116]. At $\mathcal{L} = 8 \times 10^{29}$ cm^{-2}s^{-1}, jet energy thresholds are applied to uncorrected jets. At $\mathcal{L} = 1 \times 10^{31}$ cm^{-2}s^{-1}, jet energy thresholds are applied to corrected jets.

	Path Name	Level-1 Condition	Threshold	L1 Prescale	HLT Prescale
8E29	HLT_ZeroBias	L1_ZeroBias	None	5 000	5
	HLT_Jet15U	L1_SingleJet6U	15 GeV	25	1
	HLT_Jet30U	L1_SingleJet20U	30 GeV	1	5
	HLT_Jet50U	L1_SingleJet30U	50 GeV	1	1
1E31	HLT_ZeroBias	L1_ZeroBias	None	50 000	5
	HLT_Jet30	L1_SingleJet15	30 GeV	500	5
	HLT_Jet50	L1_SingleJet30	50 GeV	50	1
	HLT_Jet80	L1_SingleJet50	80 GeV	5	2
	HLT_Jet110	L1_SingleJet70	110 GeV	1	1
	HLT_Jet180	L1_SingleJet70	180 GeV	1	1

detector simulation described real data perfectly, 1E31 would be the apt trigger menu. This menu is also employed when carrying out the study of the underlying event with SUMMER08 samples.

SUMMER08: **Composition of Trigger Streams**

Trigger streams are composed in the same way as described for the CSA08 samples: Samples are combined according to their cross sections; next, events that satisfy a given high-level trigger are combined into a common trigger stream. If a sample would contribute less than 10% of its events to a trigger stream, this particular sample is ignored for the trigger stream in question. Event fractions accepted by the high-level triggers under study for each of the investigated SUMMER08 Monte Carlo samples are listed in Table 4.9.

Table 4.9: SUMMER08: Fraction of accepted sample events per HLT stream. Only if at least 10% of all events in a given sample are accepted by a high-level trigger, is the sample included into the trigger stream composition.

	HLT_ZeroBias	HLT_Jet30	HLT_Jet50	HLT_Jet80	HLT_Jet110	HLT_Jet180
QCDDiJetPt15to20	1	0.15	< 0.01	0	0	0
QCDDiJetPt20to30	1	0.38	0.02	< 0.01	0	0
QCDDiJetPt30to50	1	0.79	0.20	< 0.01	< 0.01	0
QCDDiJetPt50to80	1	0.99	0.79	0.16	0.01	< 0.01
QCDDiJetPt80to120	1	0.99	0.99	0.78	0.27	< 0.01
QCDDiJetPt120to170	1	1	0.99	0.99	0.87	0.10
QCDDiJetPt170to230	1	1	0.99	0.99	0.99	0.67
QCDDiJetPt230to300	1	1	1	0.99	0.99	0.96
QCDDiJetPt300to380	1	1	1	1	0.99	0.99
QCDDiJetPt380to470	1	1	1	0.99	0.99	0.99
HerwigQCDPt15	1	0.14	< 0.01	< 0.01	< 0.01	< 0.01
HerwigQCDPt30	1	0.78	0.23	0.02	< 0.01	0
HerwigQCDPt80	1	0.99	0.99	0.79	0.32	0.01
HerwigQCDPt170	1	1	1	0.99	0.99	0.70
HerwigQCDPt300	1	1	1	1	0.99	0.99
HerwigQCDPt470	1	1	1	1	1	0.99

Chapter 5

Observables to Study the Underlying Event and Multiple Interactions

As pointed out in Section 1.4, the underlying event in pp collisions is a convolution of several contributions: initial- and final-state radiation, multiple interactions, and beam remnants and their interactions. Unfortunately, most underlying event observables only probe the sum of these effects. Although multiple parton-parton scatters are believed to play a major role in simulating the underlying event, to exactly quantify their contribution to the underlying event is a nontrivial task. This chapter introduces typical observables for the underlying-event characterization and discusses possibilities to isolate effects from multiple interactions. It is meant to motivate further studies by investigating which observables depend strongest on variations of underlying event model parameters. Experimental issues are postponed to Chapters 6 and 7.

The underlying event in charged-jet events, where *charged-jet* refers to a jet clustered from charged particles, can be characterized by a *topological* or by a *kinematical* selection (Sec. 5.1). Characterizing the underlying event with event topologies corresponds to associating the underlying event activity with the region transverse in azimuth with respect to the leading (i. e. with largest transverse momentum) jet. Alternatively, the underlying event can be studied with event kinematics by using the event-by-event median of the ratio of jet p_T to active jet area. As we will see, in addition to being sensitive to the underlying event activity, these observables also probe the impact of multiple interactions.

In the past, several experiments searched for direct evidence for multiple parton interactions. One of these analyses, the search for two hard scatters (*double-parton-*

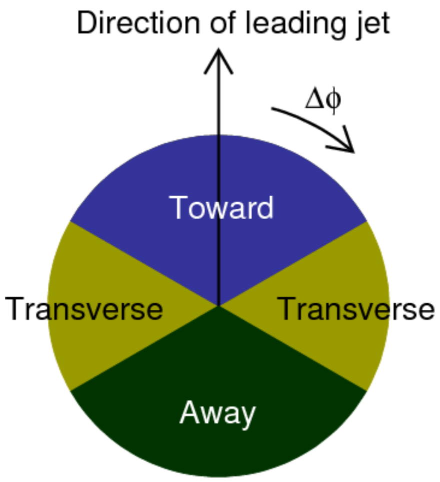


Figure 5.1: Sketch of toward, transverse, and away regions. The transverse region in charged-jet topologies is assumed to receive only few contributions from the primary scattering and is thus used to measure the underlying event in proton-proton collisions.

scattering) in final states with one photon and three jets ($\gamma + 3$ *jet events*) and its feasibility at LHC energies, is discussed in Sec. 5.2.

5.1 Charged-Jet Topologies and the Underlying Event

Significant progress in the understanding of the underlying event in hadron-hadron collisions has been achieved by measuring charged-jet topologies with the CDF detector at the Tevatron collider [66]. Fig. 5.1 illustrates the analysis idea: A jet algorithm clusters charged particles to *charged-jets*[1]. The direction of the jet with largest transverse momentum is used to separate the plane perpendicular to the beam direction ((r, ϕ) plane) in three regions:

- The *toward* region includes all charged particles that are close in azimuth to the leading charged-particle jet: $|\Delta\phi| < 60°$, where $\Delta\phi$ corresponds to the relative azimuth between charged particle and jet.
- The *away* region includes charged particles in the direction opposite to the leading charged-particle jet: $120° \leq |\Delta\phi| \leq 180°$.
- Both regions are separated by the *transverse* region: $60° \leq |\Delta\phi| < 120°$.

---

[1] When analyzing early LHC data, charged-jets are preferable to calorimetric jets, because it is expected that the CMS tracking system can be calibrated and aligned on a shorter time-scale than the electromagnetic and hadronic calorimeters.

Table 5.1: Physics object selections to study the underlying event in charged-particle jet topologies.

Charged particles	Status	Stable ($c\tau > 10$ mm)
	Transverse momentum	$p_T > 0.9$ GeV/c
	Pseudorapidity	$\|\eta\| < 2$
Charged-particle input	Status	Stable ($c\tau > 10$ mm)
	Transverse momentum	$p_T > 0.9$ GeV/c
	Pseudorapidity	$\|\eta\| < 2.5$
Charged-particle jets	Jet algorithm	SISCone ($R = 0.5$)
	Input objects	Charged-particle input
	Pseudorapidity	$\|\eta\| < 2$

The toward and away regions contain a significant part of the hard interaction. In fact, in a dijet event, the leading and next-to-leading jet are expected to emerge into these regions. On the other hand, activity besides the hard process, in particular additional parton-parton interactions, might be characterized by the activity in the region transverse to the leading jet [66] which is assumed to receive only small contributions from the primary scattering. One attempt to test this assumption is described in the following.

Table 5.1 summarizes the physics object selections imposed on charged particles and charged-particle jets. The average charged-particle activity is shown in Fig. 5.2 for the PYTHIA event generator (version 6.409, underlying event tune D6T, cf. Sec. 4.1) for proton-proton collisions at center-of-mass energy $\sqrt{s} = 10$ TeV. The left panel shows the average charged-particle multiplicity N; the right panel shows the average charged-particle scalar transverse momentum sum $\sum p_T$ (p_T sum) - both as a function of azimuth relative to the leading charged-jet. Event topologies are shown for three different scattering scales: minimum-bias events, jet production with parton transverse momenta above $\hat{p}_\perp > 45$ GeV/c, and jet production with parton transverse momenta above $\hat{p}_\perp > 160$ GeV/c for the hardest partonic process. The transverse region $60° < |\Delta\phi| < 120°$ is indicated by a yellow-shaded region. As stated before, the toward ($0° < |\Delta\phi| < 60°$) and away ($60° < |\Delta\phi| < 120°$) regions receive the dominant part of particle multiplicity and the p_T sum. Since the next-to-leading jet is not always exactly back-to-back with the leading jet, the charged-particle activity is less pronounced at $|\Delta\phi| \approx 180°$ than at $|\Delta\phi| \approx 0°$.

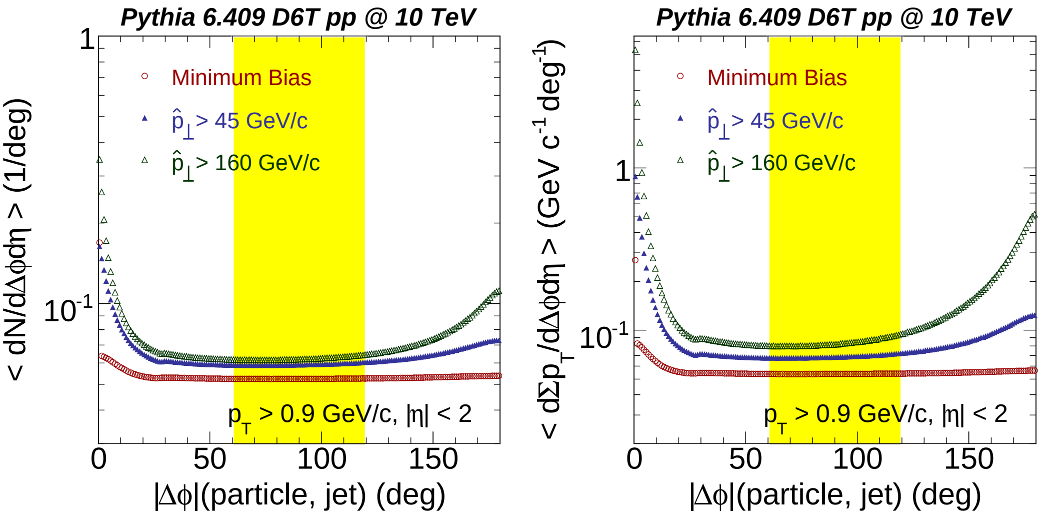


Figure 5.2: Pythia 6.409, underlying event tune D6T, $\sqrt{s}$ = 10 TeV: Average charged-particle multiplicity (left) and scalar p_T sum (right) prediction shown with respect to azimuthal distance to leading charged-particle jet. The yellow shaded region indicates the region transverse to the leading jet. Charged particles have $p_T >$ 0.9 GeV/c and $|\eta| < 2$. Predictions are shown for three event scales (minimum bias, $\hat{p}_\perp >$ 45 GeV/c, and $\hat{p}_\perp >$ 160 GeV/c).

For the transverse region, one may note that the charged-particle activity is not zero but rises for harder scatters. Part of the rise can be attributed to an increase in the amount of final-state radiation, since the distribution of the away region activity becomes broader for increasing scattering scales. However, the charged-particle activity also rises in that part of the transverse region that is closer to the leading jet ($|\Delta\phi| < 90°$). Therefore, final-state radiation alone cannot account for the transverse-region activity increase with growing scattering scale. A natural explanation for the rise would be given by attributing the activity in the transverse region to multiple interactions. As pointed out in Sec. 2.2, the number of parton-parton scatters in a proton-proton interaction is expected to be largest for a head-on collision of both hadrons; peripheral collisions are likely to produce only few parton-parton scatters in addition to the primary hard scatter. Thus, by associating harder collisions with increasingly head-on collisions, an increase in the particle activity coming from multiple interactions is to be expected. In addition, this would also explain why the charged-particle activity saturates when going to harder event scales: Once the hadrons collide fully head-on, harder collisions do not change the hadronic matter overlap further. Therefore, once this collision scale is reached, harder collisions do not change the probability for additional interactions to take place.

In order to get a quantitative estimate on the underlying event particle flow, we assume that the charged-particle activity in the transverse region originates solely from the underlying event. From Fig. 5.2, this corresponds to a contribution of nine to twelve charged particles (left) and $9 - 16$ GeV/c transverse momentum flow (right) per unit pseudorapidity, integrated over the whole azimuth.

A more detailed study of how the charged-particle activities in the toward, transverse, and away regions evolve with the event scale is shown in Fig. 5.3. Average charged-particle multiplicities (left) and average scalar transverse momentum sums (right) are shown as functions of the leading charged-jet's transverse momentum p_T(chg gen jet), which is a measure for the scatter's hardness. In all three regions, the rise of activity with respect to the scattering scale is confirmed, however, the transverse region multiplicity and the p_T sum saturate above scales of p_T(chg gen jet) > 100 GeV/c. There are interesting structures in the p_T dependence of the charged-particle multiplicity: All regions exhibit a strong rise at low p_T up to p_T(chg gen jet) ≈ 25 GeV/c. Up to p_T(chg gen jet) ≈ 40 GeV/c, the multiplicity slightly decreases until it starts to rise again. We attribute this behavior to the dominant impact of minimum-bias events in this kinematic region. These are low-multiplicity and low-p_T events, where a charged particle with slightly more transverse momentum than the other particles is identified as a "jet". Especially interesting is the comparison of toward and away multiplicity. Both curves rise at different speeds with respect to the jet transverse momentum. The toward multiplicity is larger than the away multiplicity at lower transverse momenta. However, at p_T(chg gen jet) ≈ 140 GeV/c, toward and away multiplicity are at the same level, with the away multiplicity being larger than the toward multiplicity at larger transverse momenta. At the same time, the toward p_T sum exceeds the away p_T sum and in addition rises faster at all scales. Note however that this is likely due to a bias of the selection of toward region (by definition $\Sigma p_T(\text{toward}) > \Sigma p_T(\text{away})$).

5.1.1 Multiple Interactions and the Transverse Region

We now turn to probing the sensitivity of charged-jet topologies to some of the key parameters of the PYTHIA 8 multiple interaction model, which has been introduced in Sec. 2.2. Note that a tune of PYTHIA 8 to Tevatron data is not yet available, hence it is not expected that predictions from PYTHIA 8 fully agree with predictions from tuned PYTHIA 6. Due to the recent enhancements of the multiple interaction framework of

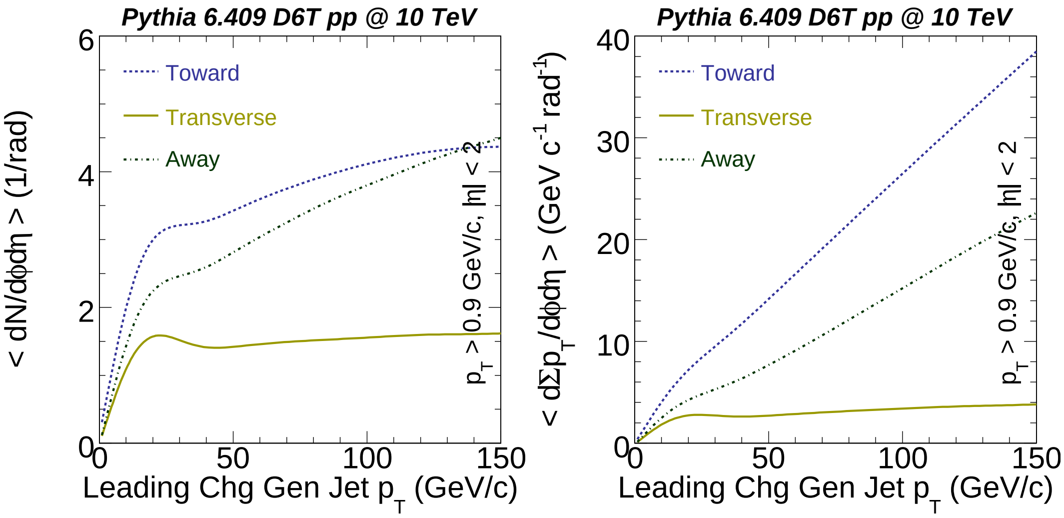


Figure 5.3: PYTHIA 6.409, underlying event tune D6T, $\sqrt{s} = 10$ TeV. Left: Average charged-particle multiplicity as a function of leading charged-particle jet p_T. Right: Average charged-particle p_T sum as a function of leading charged-particle jet p_T. Hadron-level predictions in toward (dashed line), transverse (solid line), and away (dash-dotted line) regions are shown. Note that for the toward region, the different scales for p_T sum and p_T(jet) are explained by the normalization to $|\eta| < 2$.

PYTHIA it was preferred to give an overview on how the parameters of this new model influence typical underlying event distributions.

With the program's version 8.108, QCD $2 \rightarrow 2$ events[2] have been generated for the kinematic region 15 GeV/c < $\hat{p}_T$ < 470 GeV/c. Thus, multiplicities and p_T sums are shown for p_T(chg gen jet) > 20 GeV/c. As noted above, here we are concerned with assessing the impact of model parameters on typical underlying event distributions. Beam particles are protons and unless otherwise noted, beam energies are set to the design LHC values, i. e. 7 TeV. As summarized in Table 5.1, stable charged particles with transverse momenta above $p_T = 0.9$ GeV/c and pseudorapidities $|\eta| < 2.5$ are clustered into charged-particle jets with a *SISCone* jet algorithm [46] with cone radius $R = 0.5$ and longitudinal energy recombination. Charged particle activities (where "charged particle" refers to a stable charged particle with transverse momentum $p_T > 0.9$ GeV/c and pseudorapidity $|\eta| < 2$) in the regions towards, transverse and away with respect to the leading charged-jet (with p_T(jet) > 20 GeV/c and $|\eta(\text{jet})| < 2$) are studied.

---

[2] Technical problems prevented the simulation of minimum-bias samples. In addition, the inclusion of minimum-bias events leads to a peculiar structure of particle multiplicity and p_T sum in the region of small charged-jet transverse momenta (cf. Fig. 5.3).

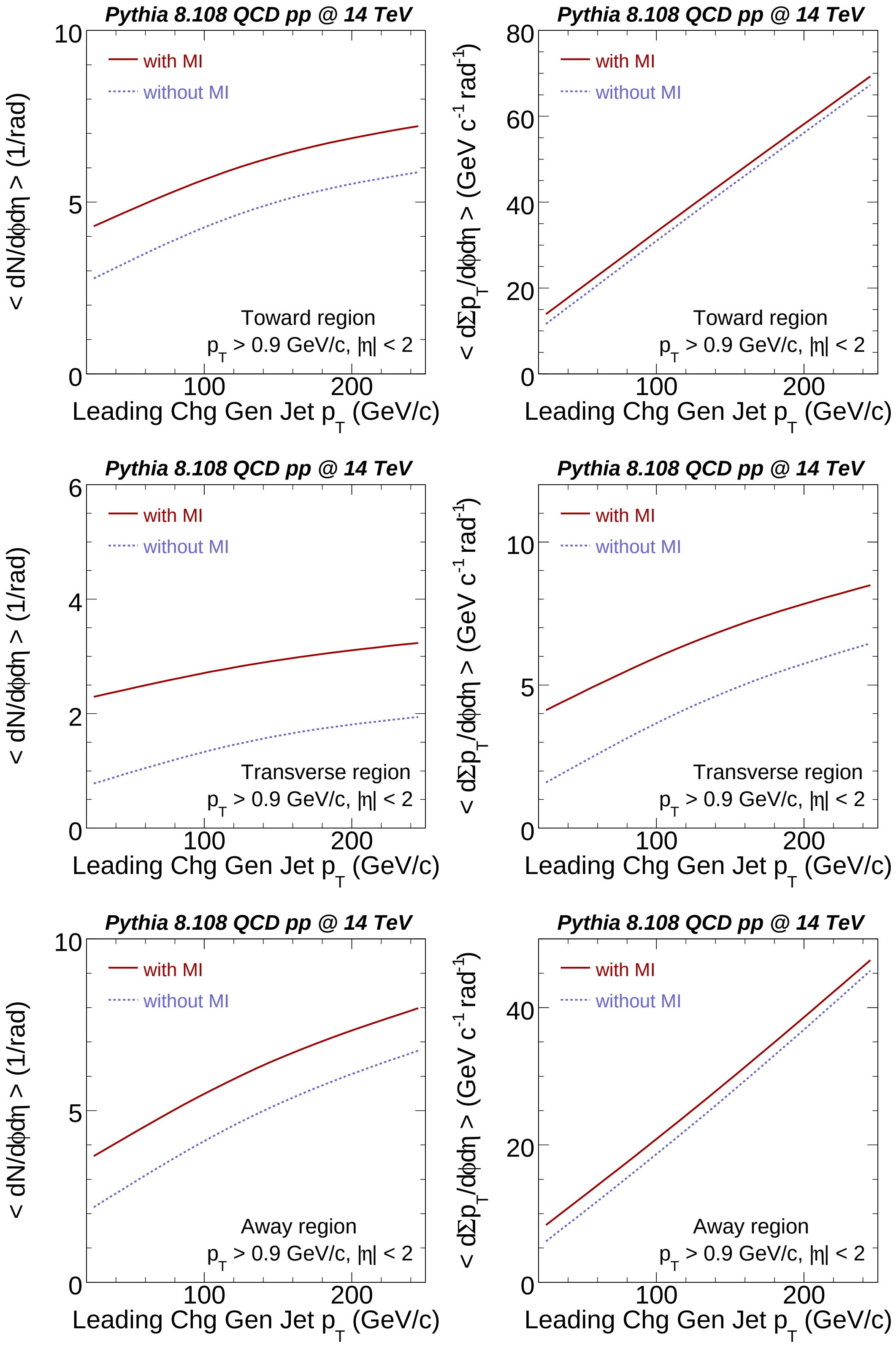


Figure 5.4: Pythia 8.108, $\sqrt{s} = 14$ TeV: Predictions on charged particle multiplicity (left) and p_T sum (right) as a function of leading charged-particle jet p_T in three azimuthal regions (from top to bottom: toward, transverse, away region). Effect of missing multiple interactions (MI scenario) shown.

As introduced earlier, typical observables to study the event activity are the charged particle multiplicity N and the scalar sum of charged particle transverse momenta $\sum p_T$. Figure 5.4 illustrates PYTHIA's prediction on N (left column panels) and on $\sum p_T$ (right column panels) in the toward (upper row panels), transverse (middle row panels) and away regions (lower row panels) as a function of the transverse momentum of the leading charged-jet. The default simulation settings (with multiple interactions, red solid line) are compared to a simulation where additional interactions have been explicitly switched off (without multiple interactions, dashed blue line). As described earlier, the event activity grows with the leading charged-jet's transverse momentum. Furthermore, in all three regions, including multiple interactions increases the event activity: Multiple interactions contribute everywhere, not just in one particular region. In the toward and away regions, the relative increase in multiplicity when switching on multiple interactions is larger than the relative increase in the p_T sum: Charged particles from additional interactions make up a small fraction of the p_T sum in these regions, since the hard scattering (with its high-momentum tracks) dominates there. The impact of multiple interactions being present or not is much more pronounced in the transverse region than in the toward and away regions: The predicted activity grows by a factor of two when switching on multiple interactions. Measuring the event activity in the transverse region is thus a well-suited observable to characterize multiple parton-parton interactions in proton-proton collisions. For this reason, we only show transverse region multiplicities and p_T sums in the following.

Especially interesting is the sensitivity of the transverse region activity to one of the key parameters of PYTHIA's multiple interaction model: the transverse momentum cutoff parameter $p_{\perp 0}$, governing the amount of multiple interactions in a proton-proton collision (cf. Sec. 2.1). The cross section for additional interactions at lower scales is dampened according to Eq. 2.10. Lowering $p_{\perp 0}$ leads to the simulation of further soft interactions and thus further soft activity. Raising $p_{\perp 0}$ suppresses the simulation of softer interactions and leads to less event activity (Fig. 5.5): Average multiplicity (left) and average transverse momentum sum (right) predictions for the transverse region are shown with respect to the leading charged-jet p_T. PYTHIA's default cutoff $p_{\perp 0} = 2.15$ GeV/c is varied by $\pm 10\%$ leading to a 20% change in the charged particle multiplicity and the charged particle p_T sum. As expected, lowering $p_{\perp 0}$, i. e. allowing interactions at softer scales, increases the activity in the transverse region and raising $p_{\perp 0}$, i. e. suppressing some interactions that were previously allowed, decreases the activity.

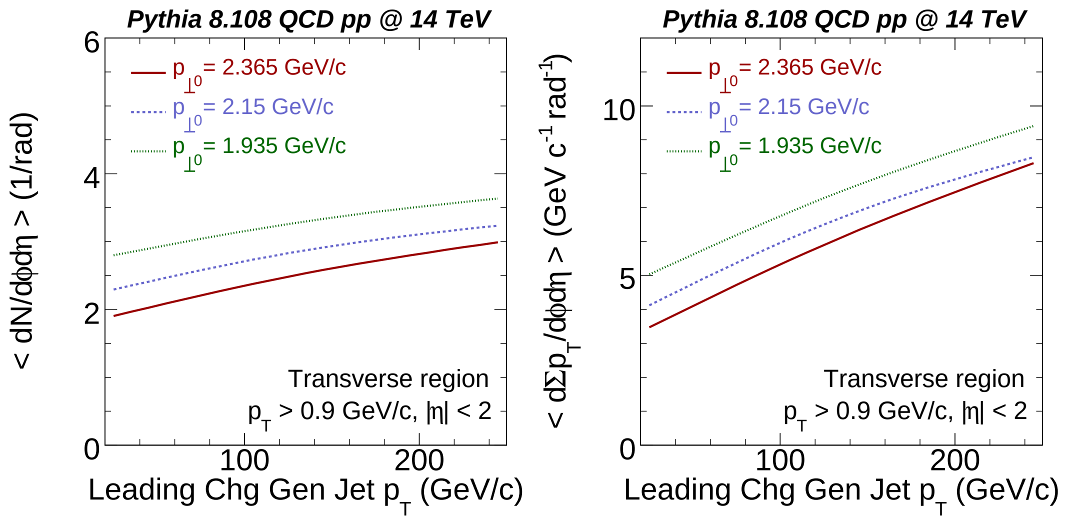


Figure 5.5: PYTHIA 8.108, $\sqrt{s} = 14$ TeV, transverse region in charged-particle jet topologies. Left: Average charged particle multiplicity as a function of leading charged-particle jet p_T. Right: Average charged-particle p_T sum as a function of leading charged-particle jet p_T. The transverse momentum threshold $p_{\perp 0}$ for additional interactions has been varied by 10% with respect to the program's default value $p_{\perp 0} = 2.15$ GeV/c.

For soft parton-parton scatterings ($|\hat{t}| \ll \hat{s}$), the following terms contribute predominantly to the partonic cross section (cf. Tab. 2.1):

- Scatterings of the types $qq' \to qq'$ and $qg \to qg$: $\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\hat{t}} \propto \frac{\hat{s}^2+\hat{u}^2}{\hat{t}^2}$,
- Scatterings of the type $gg \to gg$: $\frac{\mathrm{d}\hat{\sigma}}{\mathrm{d}\hat{t}} \propto \frac{\hat{s}\hat{u}}{\hat{t}^2}$,

where $\hat{s}$, $\hat{t}$, $\hat{u}$ are the Mandelstam variables for the partonic processes. The partonic cross sections are dimensionless; thus, the proton-proton QCD $2 \to 2$ cross sections at different center-of-mass energies only depend on the parton density functions. For larger center-of-mass energies, smaller proton momentum fractions x suffice to produce detectable particles and jets. As parton densities rise towards smaller x, the QCD $2 \to 2$ cross sections rise with increasing center-of-mass energy (Fig. 1.2). Consequently, the underlying event activity is expected to grow with the center-of-mass energy.

Activity predictions in the transverse region for center-of-mass energies of 900 GeV, 10 TeV and 14 TeV are compared in Fig. 5.6 for a fixed value of $p_{\perp 0} = 2.15$ GeV/c. Charged particle multiplicities as well as charged particle p_T sums increase with the center-of-mass energy of the proton-proton collision. At $\sqrt{s} = 900$ GeV, transverse region activities drop as a function of the jet transverse momentum, because charged-jets with $p_T(\text{chg gen jet}) > 90$ GeV/c correspond to substantial proton momentum frac-

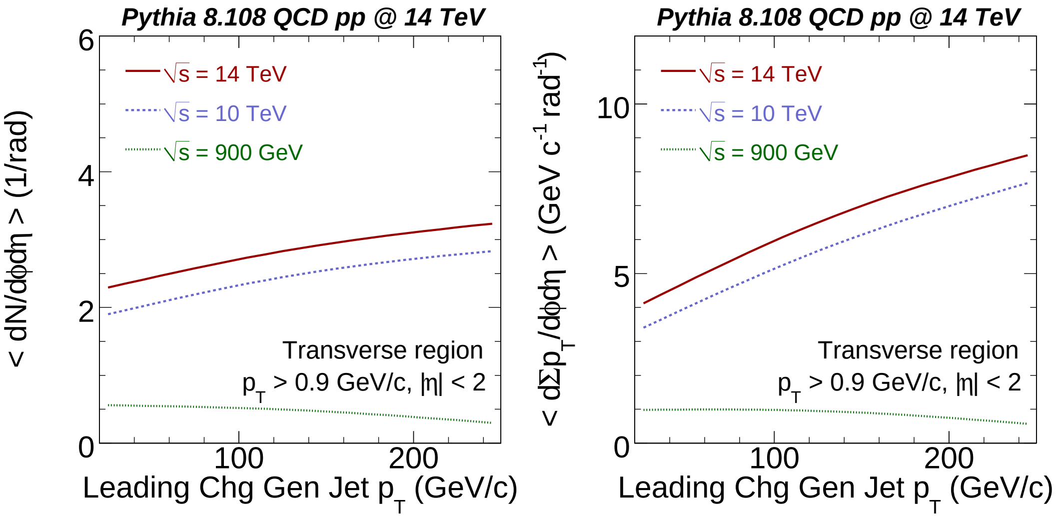


Figure 5.6: Pythia 8.108, transverse region in charged-particle jet topologies. Left: Average charged-particle multiplicity as a function of leading charged-particle jet p_T. Right: Average charged-particle p_T sum as a function of leading charged-particle jet p_T. Hadron-level predictions shown for center-of-mass energies $\sqrt{s} = 900$ GeV, 10 TeV, and 14 TeV.

tions involved in the scatter: A naive estimate requires two partons with momentum fractions $x > 0.1$ to produce two jets with transverse momenta $p_T(\text{jet}) > 90$ GeV/c. Accordingly, a significant fraction of proton momentum will enter the hard scatter, leaving less momentum to be used up by additional interactions, and thus effectively suppressing the transverse region activity. Transverse region activities for $\sqrt{s} = 10$ TeV and $\sqrt{s} = 14$ TeV rise at comparable speeds with increasing $p_T(\text{jet})$, with activities at $\sqrt{s} = 10$ TeV being typically smaller by 0.5 charged particles and 1 GeV/c per radian than activities at $\sqrt{s} = 14$ TeV are.

The transverse region activity thus behaves as expected under a variation of key parameters of multiple interaction models. We propose to measure charged-particle-activities in the transverse region at the LHC and tune available multiple interaction models on the resulting data. For a discussion of experimental issues, in particular of various alignment scenarios, and for a comparison of two different model predictions, see Chap. 7.

5.1.2 Underlying Event Characterization with Charged-Jet Areas

After having studied the charged-particle multiplicity and p_T sum in three azimuthal regions of charged-jet events, we now turn to a study of the underlying event with active jet areas in charged-jet events. When we studied the underlying event by investigating the transverse region, we attributed event constituents to the underlying event based on their *position* in (r, ϕ) space. Such a selection is to be contrasted with a more *dynamical* selection where event constituents might be attributed to the underlying event based on their *characteristics*, namely using active jet areas [101] as discussed originally in the context of pile-up subtraction [102].

The basic idea is the following: There are several ways the underlying event can modify the jet structure of an event. Particles from the underlying event might be clustered into an existing jet, modifying the jet's transverse momentum. Likewise, new jets might be formed from underlying event particles only. Finally, the subset of particles from the hard process that is clustered to a given jet might vary depending on the level of underlying-event activity. The *jet area* concept allows to study the first of these modifications, namely how the jet transverse momentum is affected by contributions from the underlying event. However, one drawback of this method is that it assumes particles from the underlying event to be distributed uniformly in (η, ϕ) space, thus probing the part of the underlying event that is uncorrelated with the hard process. Given this assumption, it is reasonable to suppose that the probability for a jet to include underlying event particles depends on the jet's "area coverage" in (η, ϕ). Albeit intuitively clear, care must be taken defining the area of a jet, as it is *not* the area covered by its constituents. Such a definition would neither be well-defined, since pointlike objects (the constituent positions) in a two-dimensional space (the covered detector region in (η, ϕ)) cannot be combined to an area, nor would this definition resolve situations with jets sharing an irregular boundary or jets with mutual overlap.

Two main definitions of jet areas have been developed: *passive* and *active* areas [101]. Passive areas are intended to characterize the jet's sensitivity to additional pointlike contributions to the overall event particle flow. The passive area of a jet J is defined with the help of a single infinitely soft "ghost" particle g that is moved over the (η, ϕ) plane:

$$a(J) \equiv \int \mathrm{d}\phi \mathrm{d}\eta f(g(\eta, \phi), J) \quad \text{where} \quad f(g, J) = \begin{cases} 1 & g \in J \\ 0 & g \notin J \end{cases} . \tag{5.1}$$

On the other hand, an appropriate description of the jet's sensitivity to the diffuse part of pile-up and underlying event is expected from active areas, which are described in the following. If jets are clustered by an infrared and collinear-safe algorithm, their momenta do not change when populating the (η, ϕ) region with infinitely soft particles. Given a set of infinitely soft particles $\{g_i\}$, the number $\mathcal{N}_g(J)$ of these ghost particles that is clustered to a given jet J is then used to determine the jet's active area A:

$$A(J|\{g_i\}) \equiv \frac{\mathcal{N}_g(J)}{\nu_g} , \tag{5.2}$$

where ν_g is the number of ghosts per unit area in (η, ϕ). Fig. 5.7 illustrates the active areas of two one-particle jets of $E_T = 10$ GeV each, obtained with the $k_\perp$ algorithm and radius 0.4. Four example configurations are shown that differ in the (η, ϕ) distance between the two jet axes between $\Delta R = \pi$ and $\Delta R = 0.05\pi$. In addition, the active area values for $\nu_g = 100/\mathrm{rad}$ are indicated together with a circle in (η, ϕ) of radius 0.4 around the jet positions. One can see that the shape and the value of the active areas change when going from panel to panel until both particles are clustered into a single jet in the last configuration (lower right panel). These fluctuations are a consequence of the specific choice of ghost particles $\{g_i\}$. The number of ghosts per unit area ν_g and the exact positions of the ghosts influence the outcome of the area calculation. A unique answer for the active area of a given jet is therefore only obtained after averaging over many different sets of ghosts with an increasing number of ghosts per unit area[3]:

$$A(J) \equiv \lim_{\nu_g \to \infty} \langle A(J|\{g_i\}) \rangle_g . \tag{5.3}$$

The active area concept is an attempt to quantify the average sensitivity of a jet to a uniform soft particle flow, as it could come from minimum-bias pile-up or from the soft part of the underlying event. For this reason, we employ active jet areas to study the underlying event in charged-jet topologies. Jet areas are calculated with a ghost density of $\nu_g = 100$ per unit area and averaging over five different sets of ghost particles, where ghosts have transverse momenta of the order $\sim 10^{-100}$ GeV/c. All area calculations are carried out using the FastJet package [117].

Adding one-hundred ghost particles per unit area to every event obviously increases the runtime of the jet clustering algorithm: Note that 100 ghosts per unit rapidity and radiant amount to approximately 3100 ghosts in the pseudorapidity range $|\eta| < 2.5$. This

---

[3] Note that the ghost transverse momentum density is to be kept infinitesimal when increasing the number of ghosts per unit area ν_g.

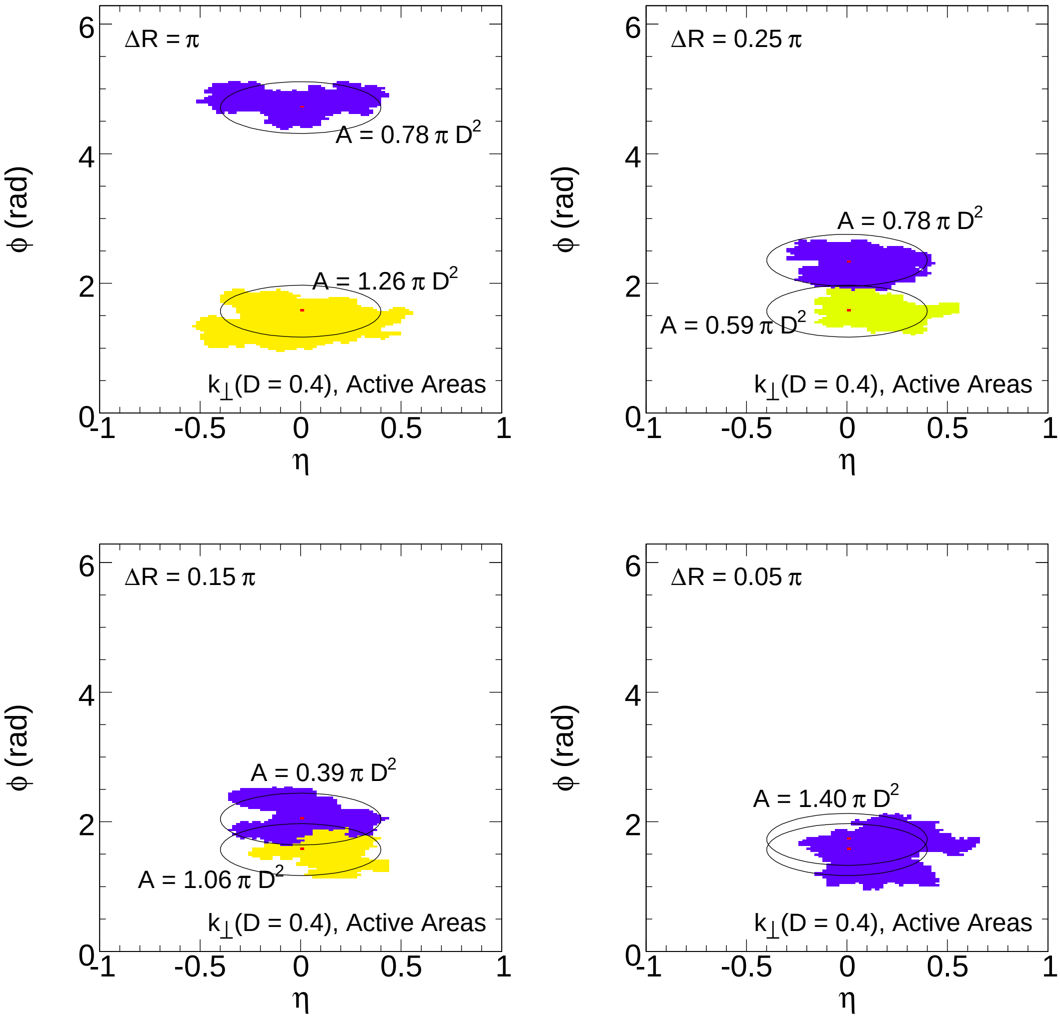


Figure 5.7: Illustration of the active area of one-particle jets determined with the $k_\perp$ algorithm and radius $D = 0.4$ for $\nu_g = 100/\text{rad}$. Shaded areas correspond to (η, ϕ) positions of ghost particles added to a previously existing jet. (Jets from ghost particles are not shown.) Active area values are indicated together with circles of radius 0.4 around the jet positions. The distance in (η, ϕ) space, $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, between the jet axes varies between π and 0.05π.

influences the decision of which jet algorithm to use. The runtimes of the $k_\perp$ algorithm and of the SISCone algorithm depend differently on the number of input particles N: The $k_\perp$ algorithm scales like $N \ln N$, but the SISCone algorithm scales like $N^2 \ln N$ [46]: Consequently, the amount of time needed by SISCone to cluster the above mentioned set of ghost particles is approximately 3100 times larger than the corresponding time needed by $k_\perp$. When trying to characterize the underlying event with jet areas, we thus prefer to cluster jets with the $k_\perp$ algorithm, due to its much faster runtime. Jets are reconstructed from stable charged particles with $p_T > 900$ MeV/c and $|\eta| < 2.5$. The $k_\perp$ jet algorithm is run with distance parameter $D = 0.4$ and longitudinal energy recombination.

As mentioned above, the active area of a jet is expected to describe the jet's average sensitivity to a uniform soft particle flow. We will now use this feature of active areas to identify the underlying event activity based on the event dynamics rather than based on particle positions in pseudorapidity-azimuth space. Assuming that the underlying event adds a constant amount of energy flow per unit area in (η, ϕ), jets can be classified according to the ratio $p_T(\text{jet})/A(\text{jet})$: This ratio is expected to represent the transverse momentum flow per unit area added by the underlying event provided that the jet contains predominantly particles from the underlying event. Jets from the hard process represent the event's dominant transverse momentum flow and are assumed to have a larger ratio p_T/A than jets from underlying event particles. If the transverse momentum threshold is set low enough, we expect a large number of jets consisting exclusively of underlying-event particles. The event would then exhibit few jets from the hard process each with a large ratio jet p_T/A and many jets from the underlying event with a small ratio jet p_T/A.

Clearly this argument relies on enough jets being reconstructed. The p_T threshold must not be so high that only the leading jets are reconstructed. We include charged-jets with transverse momenta $p_T > 1$ GeV/c into the analysis, a threshold that guarantees on average more than six jets per event (cf. a more detailed discussion in Sec. 7.4). Note that the common notion of a jet as being a highly-collimated flow of particles is not applicable for such a low threshold, since most "jets" selected in this way will only consist of single particles. The area of such a single-particle jet then probes the activity close to the particle. In this sense, the ratio between jet transverse momentum and jet active area is still expected to probe the transverse momentum flow per unit area. Predictions on the distribution of active jet areas and of ratios of jet p_T and jet area from Pythia 8.108 are shown in Fig. 5.8. Leading-jet distributions are compared with distributions for all jets. Arrows indicate average active areas for one-particle jets and ghost jets as they have been calculated numerically [101]:

$$\begin{aligned} A_{k_\perp}(\text{one-particle jet}) &\approx 0.812\pi D^2 \;, \\ A_{k_\perp}(\text{ghost jet}) &\approx 0.554\pi D^2 \;. \end{aligned} \tag{5.4}$$

The most probable area of all jets (dashed line, right panel) is of the order of $A = 0.4$ rad $\approx 0.8\pi D^2$, i. e. slightly smaller than the average area of one-particle jets. Above the most probable leading-jet area $A > 0.7$ rad $\approx 1.4\pi D^2$, the area distribution of all jets is dominated by contributions from the leading jet as is apparent from the developing "shoulder" in the distribution. The distribution of the jet p_T/A (right panel) exhibits even stronger differences between the leading jet (solid line) and all jets (dashed line). With a lower threshold of $p_T(\text{jet}) > 1$ GeV/c and the most probable jet area of $A = 0.4$ rad, the jet p_T/A of all jets (dashed line) is expected to be suppressed below $p_T/A \approx 1$ GeV/c/0.4 rad $= 2.5$ GeV/c rad^{-1}. Indeed, the lower jet p_T threshold translates into such a behavior of the distribution. In the leading-jet p_T/A distribution (solid line), phase space cuts of the simulation become visible: QCD $2 \to 2$ events have been simulated with a lower threshold $\hat{p}_T > 15$ GeV/c. Following the reasoning above, the leading jet p_T/A distribution is expected to be suppressed below $p_T/A \approx 15$ GeV/c/0.7 rad $= 21.4$ GeV/c rad^{-1} which is consistent with the observed distribution, given that the leading *charged*-jet does not include neutral particle flow. We next attempt to summarize the jets' transverse momenta and areas into a single event-by-event quantity that will then be used to characterize the underlying event activity.

Fig. 5.9-left shows Pythia's prediction on the event-by-event median ρ of the ratio of jet transverse momentum to jet active area:

$$\rho \equiv \mu_{1/2}\left(\left\{\frac{p_{Ti}}{A_i}\right\}_{p_{Ti}>1\ \text{GeV}/c}\right) \;, \tag{5.5}$$

where p_{Ti} refers to the transverse momentum of jet i and A_i refers to the active area of jet i ($\mu_{1/2}$ symbolizes the median operator). The event-by-event median ρ, i. e. the number separating the upper half of the p_T/area distribution from the lower half, is preferable to the event average of $p_T(\text{jet})/\text{area}(\text{jet})$ because it suppresses the influence of the leading jets: As long as the number of jets is large enough, the median ratio will give information on the energy flow per unit area from a diffuse, pile-up-like particle

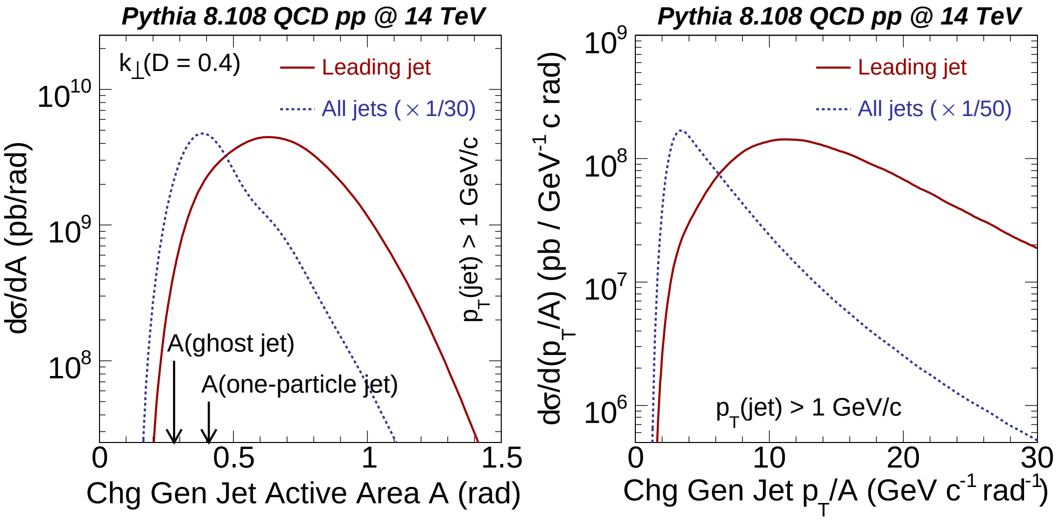


Figure 5.8: Pythia 8.108, $\sqrt{s} = 14$ TeV. Differential cross section prediction as a function of charged-particle jet active area (left) and of charged-particle jet p_T/A (right). Charged-particle jets are reconstructed with the $k_\perp$ jet algorithm with parameter $D = 0.4$. Distributions for all jets are scaled to ease comparison with the distributions for the leading jet. Arrows indicate active areas for one-particle jets and ghost jets as calculated in [101].

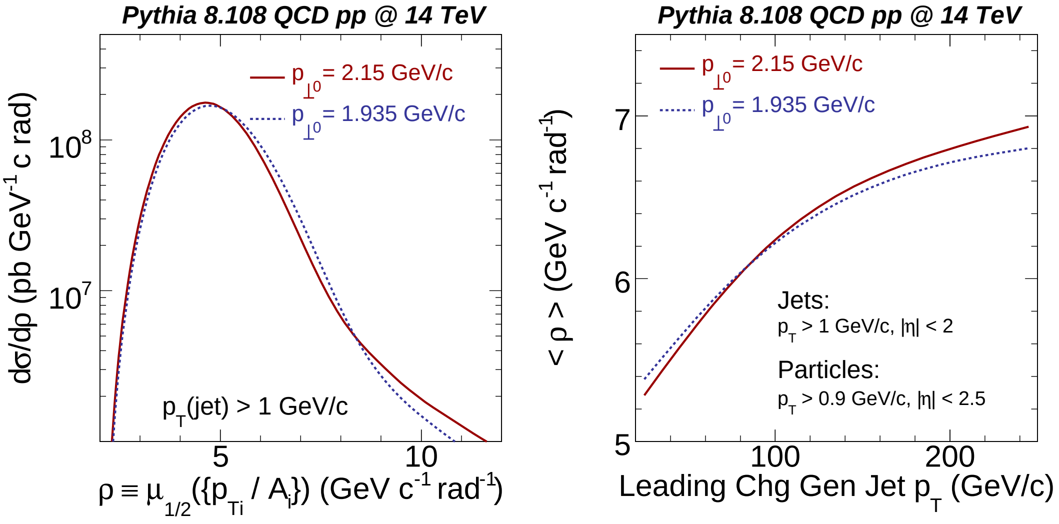


Figure 5.9: Pythia 8.108, $\sqrt{s} = 14$ TeV. Differential cross section prediction as a function of event-by-event median $\rho \equiv \mu_{1/2}(\{p_{Ti}/A_i\})$ (left) and average median as a function of leading charged-jet p_T (right). Predictions shown for two multiple interaction $p_\perp$ thresholds. Note the suppressed origin.

flow. Predictions are shown for two values of $p_{\perp 0}$. Both predictions differ in the tail of the distribution but agree fairly well in the most probable value. Differences are more pronounced when studying the change of ρ with the p_T of the leading charged-jet, Fig. 5.9-right. A higher threshold on the scale of additional interactions causes those interactions that are produced to be harder. Consequently, jets clustered from particles produced in a second or third interaction will have a larger ratio p_T/area and ρ will be larger. ρ is thus a measure of the hardness of additional jet activity normalized to the jet's susceptibility to a constant isotropic particle flow. As such it is an interesting complement to the underlying event characterization provided by the transverse region activity. Sec. 7.4 compares jet area predictions from HERWIG and PYTHIA after having passed the generated events through the full detector simulation and event reconstruction of CMS.

5.1.3 Conclusions

In this section, we have studied charged-jet events and described methods to identify the underlying-event contribution in this type of event. The transverse region is shown to be sensitive to parameters of multiple interaction models. Parameter tunes will be possible by studying charged-particle multiplicities and scalar sums of charged-particle transverse momenta. Another concept to characterize the underlying event contribution in charged-jet events is the use of jet areas. Assuming that the underlying-event contribution to a given jet depends on the jet size in (η, ϕ) space, the underlying event is associated with the event-by-event median transverse momentum flow per unit area. The transverse region is best studied in dijet topologies, but the jet area method is applicable in topologies with an arbitrary number of jets.

5.2 Double-Parton-Scattering in Final States with one Photon and three Jets

In the previous section, we studied the underlying event by investigating charged-jet topologies and kinematics. The observables under investigation are found to depend on multiple interaction parameters. And indeed, including multiple parton-parton scatters in the underlying event model has been shown to significantly improve the description of multiplicity and momentum flow in charged-jet events measured with the CDF experiment [66, 74]. However, this is not a formal proof of the existence of multiple parton-parton scatters. A big step towards such a proof would come from identifying multiple hard interactions in a single proton-proton collision. For instance, observing an excess of balanced pairs of jets where pairs are uncorrelated with respect to each other would support the hypothesis of multiple interactions taking place.

The production of four high-p_T jets is thus the most prominent process to directly study the impact of multiple interactions: Two independent scatters in the same pp or $p\bar{p}$ collision (*double-parton-scattering, DPS*), each producing two jets. Such a signature has been searched for by the AFS experiment [118] at the CERN ISR, by the UA2 experiment [119] at the CERN S$\bar{p}p$S and most recently by the CDF experiment [120] at the Fermilab Tevatron.

Searches for double-parton-scattering in four-jet events at hadron colliders face significant backgrounds from other sources of jet production, in particular from *QCD Bremsstrahlung*. Typical thresholds employed in jet triggers bias the event sample toward hard scatterings. However, a high-p_T parton is more likely to radiate additional partons, thus producing further jets. Thus, the relative fraction of jets from final-state showers above a given threshold is enlarged in jet trigger streams which is an unwanted bias. On the other hand, looking for four jets in a minimum-bias stream will yield little statistics. In a novel approach to detect double-parton-scattering, the CDF collaboration therefore studied final states with one photon and three jets [121] looking for pairwise balanced photon-jet and dijet combinations. The data sample was selected with an inclusive photon trigger, thereby avoiding a bias on the jet energy. The superior energy resolution of photons with respect to jets purifies the identification of E_T balanced pairs. CDF found an excess in pairs that are uncorrelated in azimuth with respect to the predictions from models without several hard parton scatters per proton-proton scatter. CDF interpreted this result as an observation of double-parton-scatters.

Analyses trying to identify two hard scatters in multi-jet events typically rely on methods to overcome combinatorics as there are three possible ways to group four objects into two pairs: Combinations are commonly selected pairwisely balanced in azimuth and energy. Alternatively, a final state without the need for p_T balancing is of great interest to searches for two hard scatters. One example of such a final state, that would not need p_T balancing, are events events with two b jets together with two additional jets [122]. In this case, one pair would be composed of the two b jets, and one pair would be composed of the two additional jets.

Options to generate double-parton-scatters help to identify background to topologies with negligible Standard Model cross sections. One example process is the production of two like-sign W bosons. These can only be produced through higher-order processes within the Standard Model. The predicted rate of double-parton-scatters at the LHC, where each scatter produces one W boson was found to be of the same size as higher-order Standard Model processes producing the same final state [123].

Cross section Estimation

If QCD factorization is applicable to double-parton-scattering, the corresponding inclusive cross section can be expanded as [124, 125, 126, 127, 128, 129]:

$$\sigma_{\mathrm{DPS}} = \int_{p_{\perp\mathrm{min}}} \mathrm{d}\vec{b}\mathrm{d}x_A\mathrm{d}x_B\mathrm{d}x'_A\mathrm{d}x'_B\mathcal{D}(x_A, x'_A, \vec{b})\mathcal{D}(x_B, x'_B, \vec{b})\hat{\sigma}_1(x_A, x_B)\hat{\sigma}_2(x'_A, x'_B) \ . \quad (5.6)$$

The partonic cross sections $\hat{\sigma}_1(x_A, x_B), \hat{\sigma}_2(x'_A, x'_B)$ (where indices $1, 2$ reflect possibly different subprocesses) are convoluted with two-parton density functions $\mathcal{D}(x_A, x'_A, \vec{b})$, $\mathcal{D}(x_B, x'_B, \vec{b})$ evaluated at the parton momentum fractions x_A, x'_A, x_B, x'_B. A transverse distance scale $\vec{b}$ represents the amount of spatial overlap between the hadrons' wavefunctions: The two-parton distribution functions refer to conditional probabilities $\mathcal{D}(x, x', \vec{b})$ to find one parton with momentum fraction x and one parton with momentum fraction x' separated by a distance $\vec{b}$ in transverse space in the same proton. The cross section integral is evaluated for scatters above a lower threshold $p_{\perp\mathrm{min}}$, i. e. for scatters producing hard enough jets. In the simplest approach, the two-parton densities $\mathcal{D}$ are assumed to factorize, i. e. they are uncorrelated:

$$\mathcal{D}(x, x', \vec{b}) = f_{\mathrm{eff}}(x) f_{\mathrm{eff}}(x') F(\vec{b}) \ , \quad (5.7)$$

with f_{eff} being an effective parton density function and $F(\vec{b})$ a matter overlap density. If this assumption holds, the cross section for double-parton-scattering is simply proportional to the product of the respective cross sections for single scatters:

$$\sigma_{\mathrm{DPS}} = \sigma_1\sigma_2 \int F^2(\vec{b})\mathrm{d}^2 b = \frac{\sigma_1\sigma_2}{\sigma_{\mathrm{eff}}} \ , \quad (5.8)$$

where the dependence on the transverse impact parameter $\vec{b}$ is absorbed in the *effective cross section* σ_{eff}. The CDF study of final states with one photon and three jets allowed the extraction of this effective cross section at the Tevatron [121]:

$$\sigma_{\text{eff}}(\text{CDF}) = \big(14.5 \ \pm 1.7 \ (\text{stat.}) \ ^{+1.7}_{-2.3} \ (\text{syst.})\big) \ \text{mb} \ . \tag{5.9}$$

One cross-check of the assumption that two-parton densities factorize is to test a possible x dependence of the effective cross section (cf. Sec. 5.2.3). If it is valid to assume uncorrelated two-parton densities, no dependence should be observed. This was confirmed by the CDF measurement: σ_{eff} was found to be independent of x [121].

A confirmation of the CDF observations at the LHC would be a fundamental step towards the characterization of multiple parton scatterings in proton-proton collisions. To this date, no studies exist on suitable thresholds and model predictions to detect double-parton-scattering in $\gamma + 3$ jet final states at the LHC. We aim to investigate these final states when produced in proton-proton collisions at center-of-mass energies of $\sqrt{s} = 14$ TeV. Important questions are whether it is possible to unambiguously differentiate between double-parton-scatters and other sources of similar topologies, and whether one can suggest observables suitable for the study of double-parton-scatters. In addition, it is interesting to investigate which jet thresholds are experimentally feasible but still allow a substantial amount of signal events.

Given the absence of feasibility studies for the LHC, we will first study model predictions and kinematics at generator level and describe these studies in the subsequent sections. Prompt-photon events, i. e. events of the type $pp \to \gamma + X$ are simulated with several generator settings. These events are subsequently analysed with selection cuts inspired by the CDF analysis, taking into account the different detector acceptances of CDF and CMS, and the different collision kinematics at the Tevatron and at the LHC. Experimental issues are postponed to a later study.

5.2.1 Simulation of Double-Parton-Scattering

CDF studies of the process $p\bar{p} \to \gamma + 3$ jets support the assumption of factorized two-parton densities according to equation 5.7 [129]. The cross section for double-parton-scatterings then only depends on the (known) single-parton-scattering cross sections and a parametrization of the protons' matter overlap. This section explains the meaning of the effective cross section in the context of Pythia's multiple interaction framework.

Pythia evaluates the double-parton-scattering cross section as follows: If the probability for an event to occur at any given time is independent of what happens at other times, the probability for n events to happen in a finite time interval is distributed according to a Poissonian:

$$\mathcal{P}_n = \langle n \rangle^n \frac{\exp(-\langle n \rangle)}{n!} \ , \tag{5.10}$$

with $\langle n \rangle$ being the average number of events. For small $\langle n \rangle$, one can approximate

$$\exp(-\langle n \rangle) \xrightarrow{\langle n \rangle \text{ small}} 1 \ \Rightarrow \ \mathcal{P}_1 = \langle n \rangle \ \text{ and } \mathcal{P}_2 = \frac{\langle n \rangle^2}{2} \ . \tag{5.11}$$

With different event types a and b (for example, γ jet production and dijet-production), where $\langle n \rangle = \langle n_a \rangle + \langle n_b \rangle$, the probability for having one event of type a and one event of type b is

$$\mathcal{P}_{2ab} = \mathcal{P}_{1a} \mathcal{P}_{2b} \ . \tag{5.12}$$

This applies to high-energy events where we assume interactions to occur at different $p_\perp$ values independently of each other inside inelastic nondiffractive events. Here,

$$\mathcal{P}_n \propto \frac{\sigma_n}{\sigma_{\text{nondiffractive}}} \ . \tag{5.13}$$

According to Eq. 2.8, as long as $p_{\perp\text{min}}$ is chosen small enough, it is correct to assume the total hard cross section above $p_{\perp\text{min}}$ to be much smaller than $\sigma_{\text{nondiffractive}}$ (a minimal cutoff of $p_{\perp\text{min}} = 20$ GeV/c at the LHC should suffice). Then, $\exp(-\langle n \rangle) \approx 1$ holds. Hence, the cross section for double-parton-scattering in photon-three-jet final states is:

$$\sigma_{(\gamma j)(jj)} = \langle f_{\text{impact}} \rangle \frac{\sigma(pp \to \gamma j + X) \ \sigma(pp \to jj + X)}{\sigma_{\text{nondiffractive}}} \ , \tag{5.14}$$

where the average *enhancement/depletion factor* $\langle f_{\text{impact}} \rangle$ accounts for different impact parameters: Central collisions will likely exhibit more average activity than peripheral ones will, resulting in a *trigger bias* towards central collisions. Selecting events with a hard process favours events at small impact parameter. In the previous section, we mentioned an alternative representation of $\sigma_{(\gamma j)(jj)}$ by means of the *effective* cross section

σ_{eff}:

$$\sigma_{(\gamma j)(jj)} = \frac{\sigma(pp \to \gamma j + X)\ \sigma(pp \to jj + X)}{\sigma_{\text{eff}}}\ . \tag{5.15}$$

Comparing Eqs. 5.14 and 5.15, it is apparent that in the context of PYTHIA's multiple interaction framework, the effective cross section is a measure of the enhancement factor, which in turn is sensitive to the spatial parton distribution in the proton through the matter overlap profile.

Double-parton-scatters can be directly simulated with either PYTHIA or HERWIG if first and second hard processes along with respective phase space cuts are specified. It should be made clear that no formalism different from the frameworks described in Sections 2.2 and 2.3 is invoked. The specification of double-parton-scatters is merely a matter of making the event generation more efficient: Events where the multiple interaction framework simulates unwanted multiple interactions (unwanted, because the second hard process does not fulfill the required selections on $\hat{p}_\perp$, etc.) are discarded. This will yield the same result as generating events with "normal" settings, i. e. including multiple interactions, but later on selecting only events with two hard interactions - a quite inefficient method.

Here, we aim to assess observables that are assumed well-suited to disentangle contributions to $\gamma + 3$ jet topologies from QCD Bremsstrahlung and multiple interactions. Since the jets in $\gamma + 3$ jet topologies can come from initial-state radiation, final-state radiation, or from multiple interactions, we try to identify how one can select events with large relative components of each of these sources of additional jets. Thus, in order to avoid any potential bias on the produced jets, prompt photon processes $pp \to \gamma + X$ are studied. Typically, the produced photon is balanced in transverse energy and azimuth by a jet. In addition, further jets might be produced through initial-state or final-state radiation, or through multiple interactions. By investigating p_T balancing variables, the hope is to separate those events with additional jets from QCD Bremsstrahlung from those events with additional jets from multiple interactions.

Generation of Test Samples with Herwig++ and Pythia 8

Prompt-photon events were simulated with PYTHIA version 8.108 and HERWIG++ version 2.2.0 including the processes (cf. Fig. 5.10 for the lowest-order Feynman graphs) $qg \to q\gamma$, $q\bar{q} \to g\gamma$, $gg \to g\gamma$ (box graph), $q\bar{q} \to \gamma\gamma$, $gg \to \gamma\gamma$ (box graph) with more

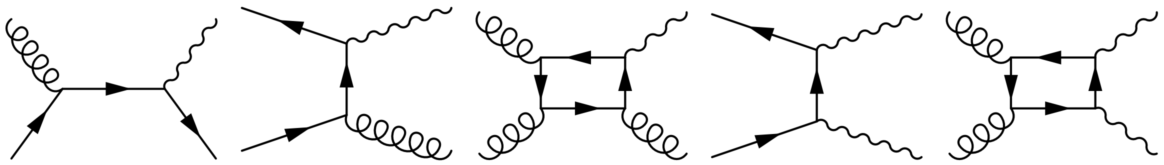


Figure 5.10: Lowest-order Feynman graphs for the generated prompt photon subprocesses, from left to right: $qg \to q\gamma$, $q\bar{q} \to g\gamma$, $gg \to g\gamma$ (box graph), $q\bar{q} \to \gamma\gamma$, $gg \to \gamma\gamma$ (box graph). Within the phase space region under study, the first two subprocess types constitute 99% of all events.

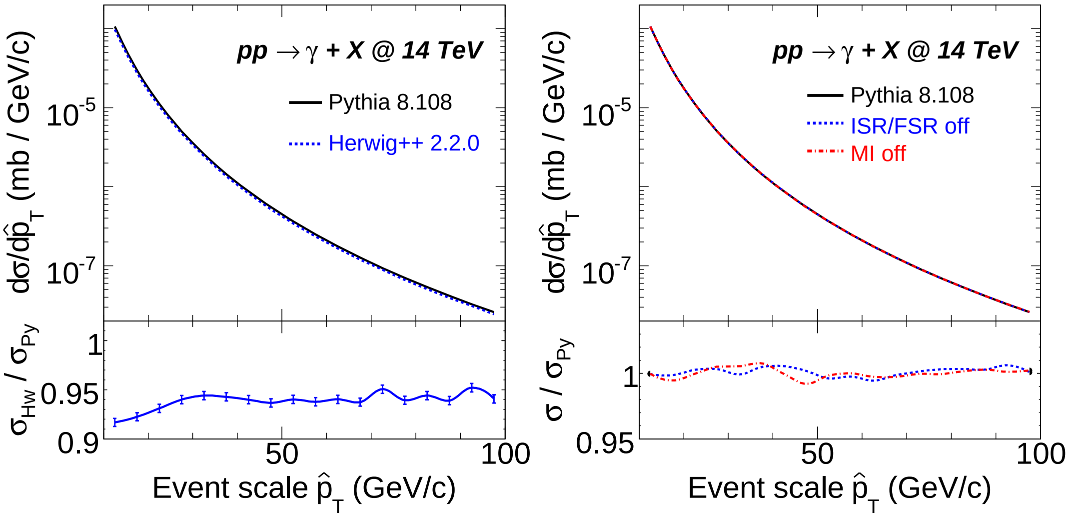


Figure 5.11: Differential cross section predictions for prompt photon production as a function of event scale $\hat{p}_T$. Herwig predicts cross sections at the level of 90% to 95% of the Pythia prediction (left panel). Various shower and multiple interaction settings of Pythia agree within 1% (right panel).

than 99% of events being of one of the first two subprocess types. Samples of 100 000 events each were simulated in $\hat{p}_T$ bins of 5 GeV/c, starting at $\hat{p}_T = 10$ GeV/c and going up to 100 GeV/c. Relative contributions of the various subprocesses change slightly with the hard event scale. For 10 GeV/$c < \hat{p}_T <$ 15 GeV/c, the sample consists of 94% $qg \to q\gamma$ events and 5% $q\bar{q} \to g\gamma$ events. For 95 GeV/$c < \hat{p}_T <$ 100 GeV/c, the sample consists of 90% $qg \to q\gamma$ events and 9% $q\bar{q} \to g\gamma$ events. All samples were normalized to the total prompt photon production cross section.

Fig. 5.11 compares cross section predictions of Pythia[4] and Herwig (left) and three different simulation settings of Pythia (right): Apart from having all default switches

---

[4] For the remainder of the chapter, Pythia refers to Pythia version 8.108 and Herwig refers to Herwig++ version 2.2.0.

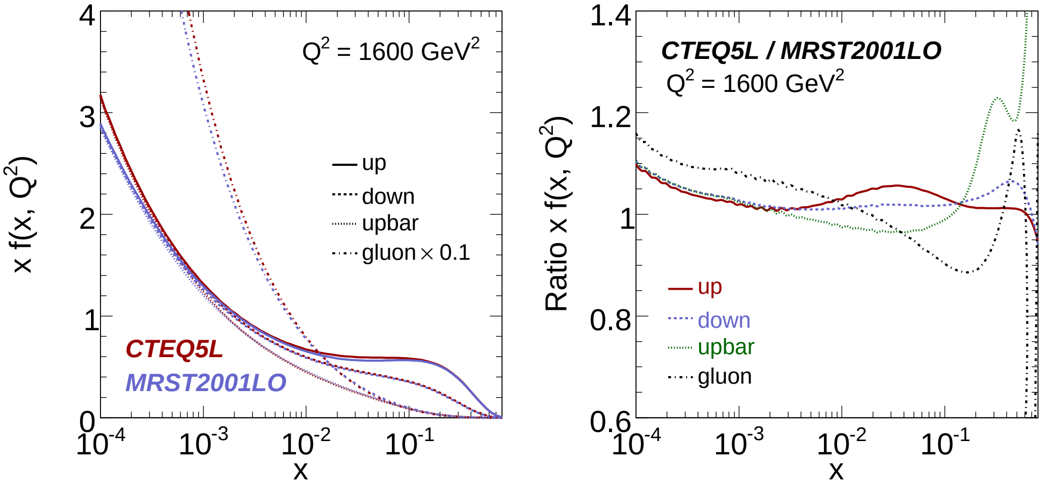


Figure 5.12: Left: Parton density distributions at $Q^2 = 40^2$ GeV^2 used per default by Pythia 8.108 (Cteq5l) and Herwig++ 2.2.0 (Mrst2001lo). Right: Ratio of default parton densities.

on (*Default* scenario) with jets being competitively produced from initial- and final-state-radiation or multiple interactions, two specific settings were simulated: Events without multiple interactions (*Shower* scenario) and events without initial- or final-state radiation (*MI* scenario) are expected to exhibit differences in kinematics and event topology. Switching parton showering and multiple interactions on and off does not affect the hard process itself therefore leaving the cross section prediction unchanged. Over the entire considered range, Pythia's cross section predictions are larger by approximately 5% than Herwig's predictions. This is due to a different default choice of parton density functions: Pythia uses the Cteq5l set; Herwig's default is Mrst2001lo. Both densities are depicted for a momentum transfer squared of $Q^2 = 40^2$ GeV^2 and Bjorken variable x ranging from 10^{-4} to 0.8 in Fig. 5.12 together with the respective ratios. At the selected scale, gluons are by far the dominant parton species below momentum fractions $x = 10^{-2}$. The Cteq5l gluon distribution exceeds the one from Mrst2001lo by up to 15% for $x < 10^{-3}$. Other Cteq5l distributions are larger than Mrst2001lo's parametrization by up to 10%. Given that prompt-photon production is dominated by quark-gluon scattering $qg \to q\gamma$ (see above), the different parton densities, in particular the differences in gluon distributions at small x, explain the different cross section predictions from Pythia and Herwig.

Table 5.2: CDF selection of photon-three-jet events [121] together with a suggested extrapolation to LHC energies.

	CDF	LHC extrapolation
Photon	$\|\eta\| \leq 1.1$	$\|\eta\| \leq 2.5$
	$E_T \geq 16$ GeV	$E_T \geq 50$ GeV
Jets	Cone $R = 0.7$	$k_\perp$ $D = 0.4$
	$\|\eta\| \leq 4.2$	$\|\eta\| \leq 5$
	$E_T \geq 5$ GeV	$E_T \geq 20$ GeV
	$E_{T4} < 5$ GeV	$E_{T4} < 10$ GeV
	$E_{T2}, E_{T3} < 7$ GeV	$E_{T2}, E_{T3} < 30$ GeV

5.2.2 Event Selection and Background Discrimination

Stable particles (except neutrinos) are clustered into jets using a longitudinally invariant $k_\perp$ algorithm with parameter $D = 0.4$ [44, 45]. Table 5.2 summarizes the kinematic selection on photon and jets as imposed by CDF [121] together with a suggested extrapolation of these cuts to LHC energies. The suggested thresholds follow the CMS detector acceptance (cf. Chap. 3), but should merely be seen as a first approximation to a final event selection. The threshold choices are motivated in the following.

The polar acceptances of the CMS electromagnetic and hadronic calorimeters are reflected in pseudorapidity cuts of $|\eta(\gamma)| \leq 2.5$ and $|\eta(\text{jet})| \leq 5$. Photon transverse energies are required to be above $E_T(\text{photon}) > 50$ GeV, jet transverse energies have to be above $E_T(\text{jet}) > 20$ GeV, in order to ensure a sufficient purity in reconstruction [92]. Three PYTHIA settings are studied:

Default: PYTHIA is used "out-of-the-box". Parton showers and multiple interactions are included in the event selection.

MI: The simulation of parton showers is switched off. Additional jets are produced exclusively by the multiple interaction framework.

Shower: Multiple interactions are switched off. Additional jets come from initial- or final-state radiation.

In the following, all comparisons between PYTHIA and HERWIG are carried out using PYTHIA's *Default* settings and HERWIG with its default underlying event tune. Specif-

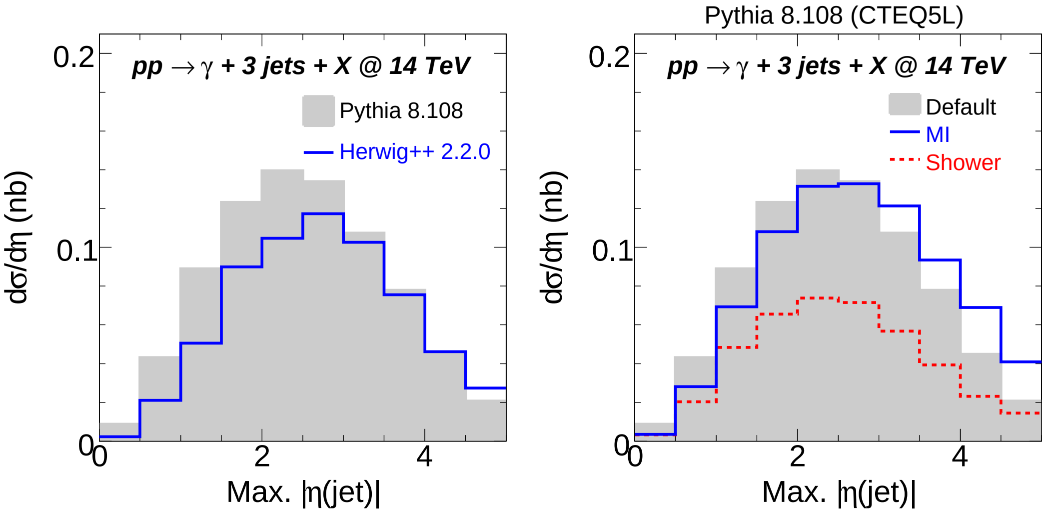


Figure 5.13: Distribution of maximal jet pseudorapidity for the three jets in photon-three-jet combinations. Comparison between Pythia and Herwig prediction (left) and between three Pythia settings (right).

ically, the simulations of multiple interactions and parton showers are switched on. We next study the kinematics of the three hardest jets in prompt photon events.

Pseudorapidity and transverse momenta of the three hardest jets are evaluated in Figures 5.13 and 5.14. Events with one photon and at least three jets are selected. For the three selected jets, the largest absolute jet pseudorapidity (Fig. 5.13) and the smallest jet transverse momentum (Fig. 5.14) are shown. As a general trend, Herwig predicts less prompt photon events with such a final state. Pythia predicts jets to be slightly more central than Herwig (Fig. 5.13-left). If jets only come from multiple interactions, they typically cover a larger pseudorapidity range (Fig. 5.13-right). A restriction to the central detector thus decreases the relative contribution of multiple interactions to the 4-body-system. Note that this is mostly an artifact of the selection in Table 5.2: Since the photon is required to be in the central region, its balancing jet often is in the central region as well. If this jet radiates further jets, they are also more likely to be in the central region. On the other hand, if jets 2 and 3 really come from a second hard scatter, one would not expect any correlation in pseudorapidity of these jets with the photon. The reader may be confused about the cross section prediction for *MI* events being larger than the cross section prediction for *Default* events for pseudorapidities $|\eta| > 3$. If we were to show only the particular subsample of *Default* events where additional jets come from multiple interactions, the cross section prediction for the *MI* scenario would by construction be less or equal to the *Default* scenario prediction. However,

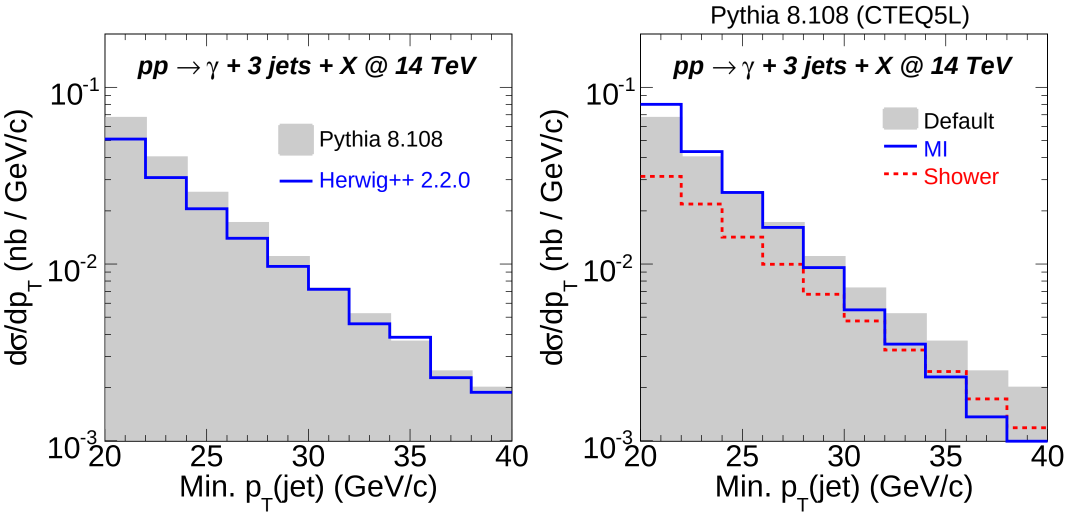


Figure 5.14: Distribution of minimal jet transverse momentum for the three jets in photon-three-jet combinations, i. e. p_T(jet 3). Comparison between Pythia and Herwig prediction (left) and between three Pythia settings (right).

the distributions have been obtained after forcing additional jets to come from multiple interactions - thereby distorting the spectra to the most extreme case. Additional jets in *MI* events are expected to be produced in pairs, therefore already the requirement of at least three jets might suppress *Default* events and *Shower* events stronger than it suppresses *MI* events. In addition, those events that do contain one photon and at least three jets might have different kinematics.

Figure 5.14 shows the distribution of the p_T of the jet with lowest transverse momentum of the three hardest jets. The distribution is shown for events with at least three jets and one photon. The differential cross section predictions are falling from 50 pb at p_T(jet 3) $= 20$ GeV/c to 2 pb at p_T(jet 3) $= 40$ GeV/c. As before, Herwig predicts in general less events with three jets than Pythia does (Fig. 5.14-left). Jets from multiple interactions are softer in p_T than jets from initial state radiation: A balance has to be found between selecting a jet p_T threshold where jet reconstruction is of sufficient quality and a p_T threshold that still allows multiple interactions to contribute significantly to the final state. Requiring a transverse jet energy of 20 GeV (cf. Table 5.2) is at the absolute limit of jet reconstruction in the CMS calorimeter, a lower threshold is experimentally not feasible. Since on the other hand, multiple interactions are more likely to produce soft jets, an upper cut of 30 GeV on the transverse jet energy of jets 2 and 3 is imposed. This limit suppresses contributions from events where additional jets are produced through parton showers.

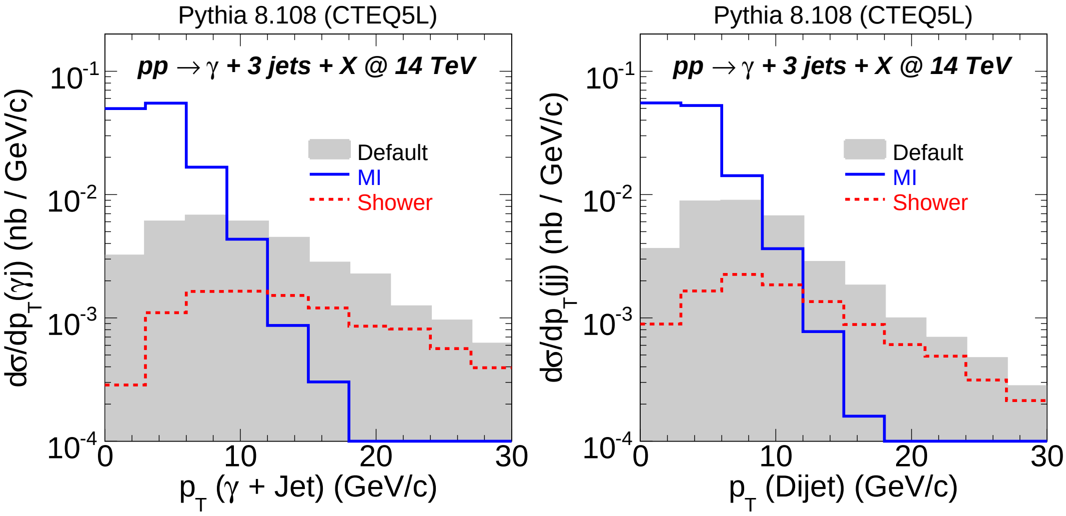


Figure 5.15: PYTHIA 8.108: Transverse momentum spectra for photon+jet (left) and dijet (right) systems, selected such that the variable defined in Eq. 5.16 is minimized.

We now turn to the question of how the photon and the three jets are combined into pairs. The photon must not be contained in one of the three leading jets, so a lower threshold of 0.2 is imposed on the distance between the photon and each jet in (η, ϕ) space, $\Delta R = \sqrt{\eta^2 + \phi^2}$. If the jet algorithm should yield more than three jets, these additional jets are restricted to have a transverse energy below 10 GeV in order to ensure a clean final state with exactly three well-defined hard jets. One photon-jet and one dijet pair are selected from the four final-state objects by optimizing the pairwise transverse momentum balance. Based on the assumption that pairs are back-to-back in double-parton-scattering events, p_T balance is strived for by choosing the combination that minimizes the pairwise vector sum of transverse momenta [119, 121]:

$$\frac{|\vec{p}_{T\gamma} + \vec{p}_{Ti}|^2}{|\vec{p}_{T\gamma}| + |\vec{p}_{Ti}|} + \frac{|\vec{p}_{Tj} + \vec{p}_{Tk}|^2}{|\vec{p}_{Tj}| + |\vec{p}_{Tk}|} \; . \tag{5.16}$$

In the following, we will refer to jet 1 as the jet combined with the photon - regardless of whether this jet is the hardest jet in the event or not.

Fig. 5.15 shows the resulting transverse momentum distribution of the photon-jet and dijet pairs. By construction of variable 5.16, combinations with small pairwise transverse momenta are selected. Both pairs' transverse momenta are more likely to vanish if additional jets are guaranteed to come from multiple interactions (*MI* scenario). With additional jets coming from initial- and final-state radiation, pairs are more likely to be slightly imbalanced in transverse momentum: In a double-Bremsstrahlung process (de-

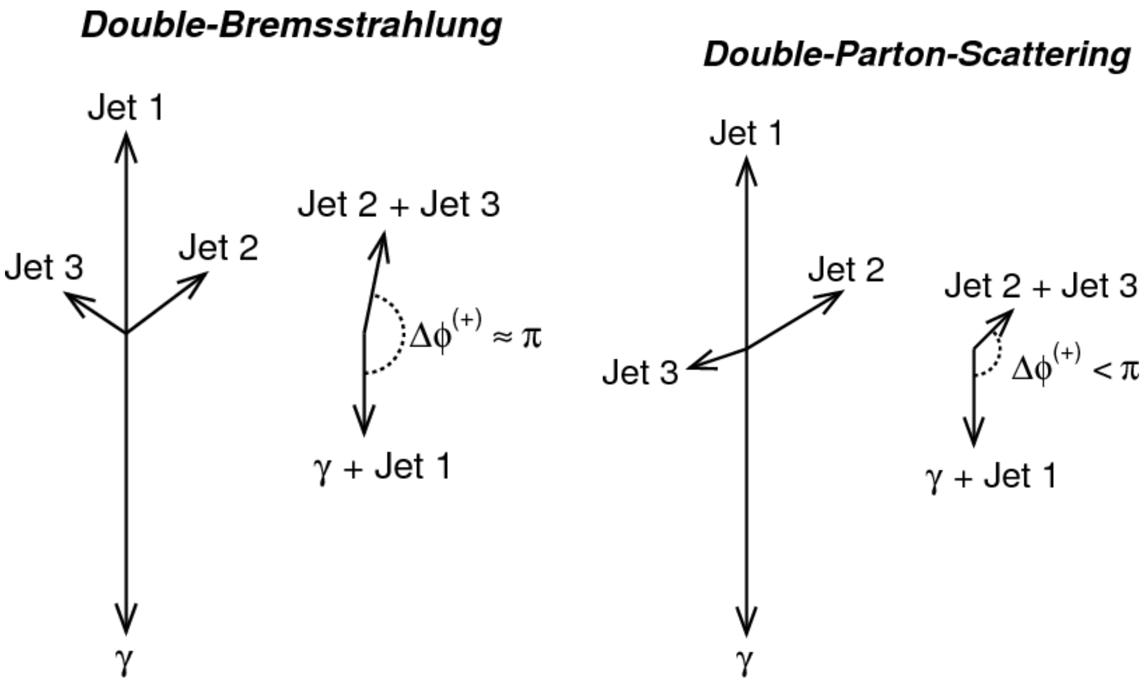


Figure 5.16: CDF definition of azimuthal angle between pairs, together with typical configurations of double-Bremsstrahlung (left) and double-parton-scattering events (right).

picted in Fig. 5.16-left), the parton initially balancing the photon will radiate twice to produce two further jets. At parton level, the photon-parton pair's transverse momentum corresponds to the transverse momentum given away by the parton to the two radiated partons. Most likely, the additional jets in double-Bremsstrahlung events will not be back-to-back in contrast to the dijet pair in double-parton-scattering events. This observation will be used to enrich the event sample with potential double-parton-scattering events (Fig. 5.16-right): Photon+3-jet combinations in DPS events are pairwise balanced in azimuth and transverse momentum and should be uncorrelated in angle and scattering hardness. To the contrary, in double-Bremsstrahlung events, γ+3-jet combinations will typically feature a photon balanced by three jets all emerging into the same hemisphere.

Correlations in azimuth (left) and transverse energy (right) of the photon-jet pair are depicted in Fig. 5.17. Photon and jet 1 are well balanced in azimuth and have approximately equal and opposite transverse energies if initial- and final-state radiation are switched off (*MI* scenario), since this setting suppresses further radiation of the parton balancing the photon. If however, the parton is allowed to further radiate (*Default* and *Shower* scenarios), photon and jet 1 are less well balanced in azimuth and the

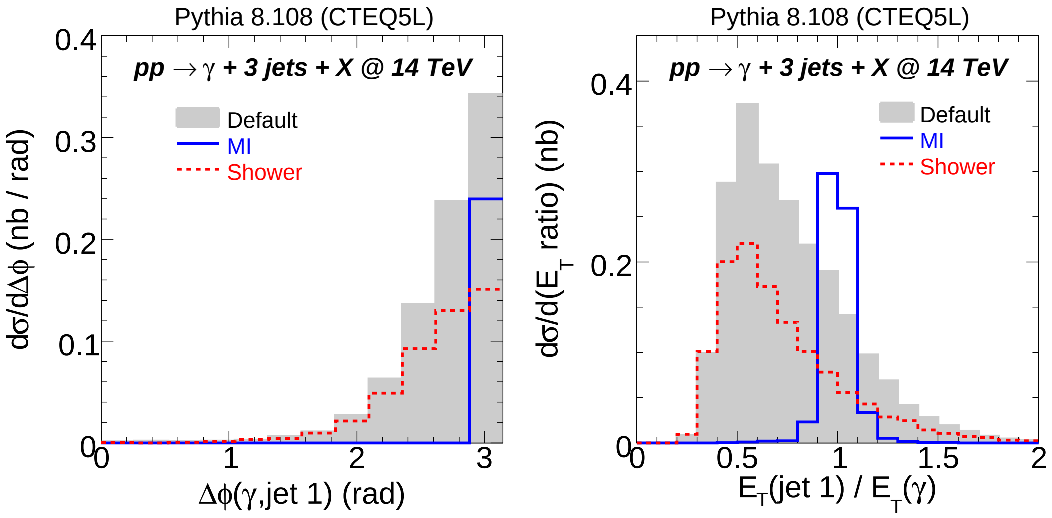


Figure 5.17: Photon-jet pair in photon-three-jet events: Differential cross section predictions from Pythia shown for relative azimuth (left) and transverse energy ratio (right).

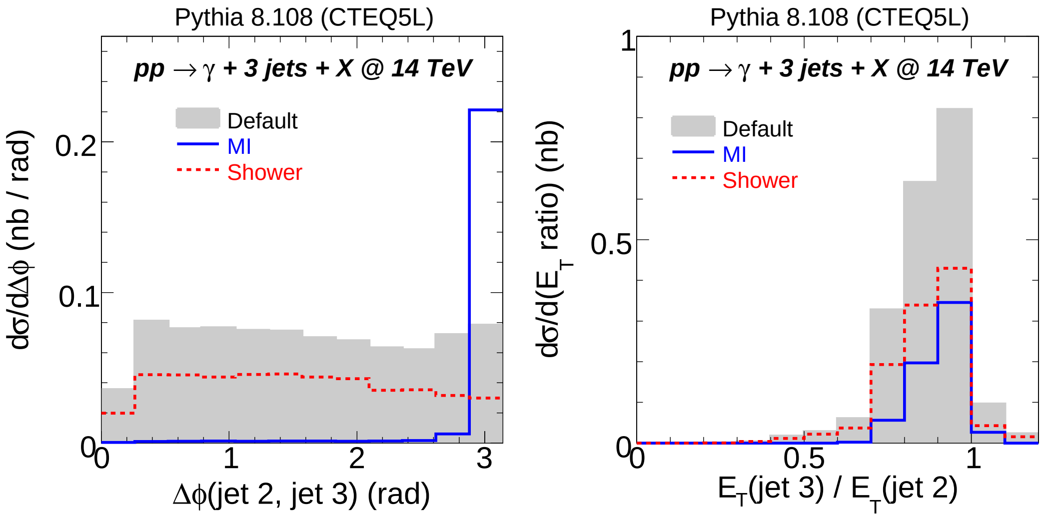


Figure 5.18: Dijet pair in photon-three-jet events: Differential cross section predictions from Pythia shown for relative azimuth (left) and transverse energy ratio (right).

balancing jet typically has only half the photon's transverse energy. Again, this at first surprisingly large contribution of large angle radiation can be attributed the special event selection of this analysis. Since the distributions are only for events with one photon and three jets, an implicit bias suppresses well-balanced photon-jet pairs, where the parton does not radiate further.

Correlations in azimuth (left) and transverse energy (right) of the dijet pair are depicted in Fig. 5.18. Again, if initial- and final-state radiation is switched off (*MI* scenario), the pair is perfectly balanced in azimuth and both jets have approximately equal and opposite transverse energies, since further radiation is suppressed after the scatter has taken place. If however, additional jets are produced by initial- or final-state radiation (*Shower* scenario), jets 2 and 3 are fairly uncorrelated in azimuth and have slightly different transverse energies. Obviously, the jet algorithm will prevent jets from being too close in (η, ϕ) space, because otherwise, both jets are merged into one jet. Jets nearby in azimuth, i. e. jet pairs with small $\Delta\phi(\text{jet } 2, \text{jet } 2)$, are thus forced to be separated in pseudorapidity. This additional requirement on jet pairs with small $\Delta\phi$ reduces the number of entries in the first bin of Fig. 5.18-left.

In DPS events, both scatterings are supposed to be uncorrelated in scale and direction. To test this assumption, AFS and CDF investigated azimuthal correlations between pairs. Both chose to study the azimuthal difference, $\Delta\phi^{(\pm)}$, between p_T vectors representing each of the pairs. AFS constructed said p_T vectors from the vector differences between the two objects to obtain $\Delta\phi^{(-)}$, but CDF constructed the pairs' p_T from the vector sum, obtaining $\Delta\phi^{(+)}$ (cf. Fig. 5.16):

$$\Delta\phi^{(-)}(\text{AFS}) = \angle\left(\vec{p}_T^{\,\gamma} - \vec{p}_T^{\,\text{jet 1}}, \vec{p}_T^{\,\text{jet 2}} - \vec{p}_T^{\,\text{jet 3}}\right) \tag{5.17}$$

$$\Delta\phi^{(+)}(\text{CDF}) = \angle\left(\vec{p}_T^{\,\gamma} + \vec{p}_T^{\,\text{jet 1}}, \vec{p}_T^{\,\text{jet 2}} + \vec{p}_T^{\,\text{jet 3}}\right) \tag{5.18}$$

As the pair's p_T must not be zero in order to compare its direction with the other pair's p_T, both methods fail for specific configurations: The AFS method fails for objects with equal transverse momentum vectors (yet, this case is suppressed, if not excluded, by the selection Eq. 5.16), and the CDF method fails for perfectly p_T-balanced objects. This means in particular, that $\Delta\phi^{(+)}$ is undefined for ideal DPS events. On the other hand, $\Delta\phi^{(+)}$ is well defined in double-Bremsstrahlung events. Therefore, studying $\Delta\phi^{(+)}$ might give insight as to how much γ+3-jet events deviate from the spectrum predicted by DBS only. Fig. 5.16 depicts two typical configurations for double-Bremsstrahlung (Fig. 5.16, left) and double-parton-scattering (Fig. 5.16, right). Additional jets produced in double-Bremsstrahlung typically point away from the photon and surround the jet balancing the photon. Expectations for the above described variables are therefore $\Delta\phi^{(-)}(\text{AFS}) \approx \pi/2$ and $\Delta\phi^{(+)}(\text{CDF}) \approx \pi$ if additional jets come from double-Bremsstrahlung. Otherwise, i. e. if additional jets come from multiple interactions, both variables should be distributed uniformly.

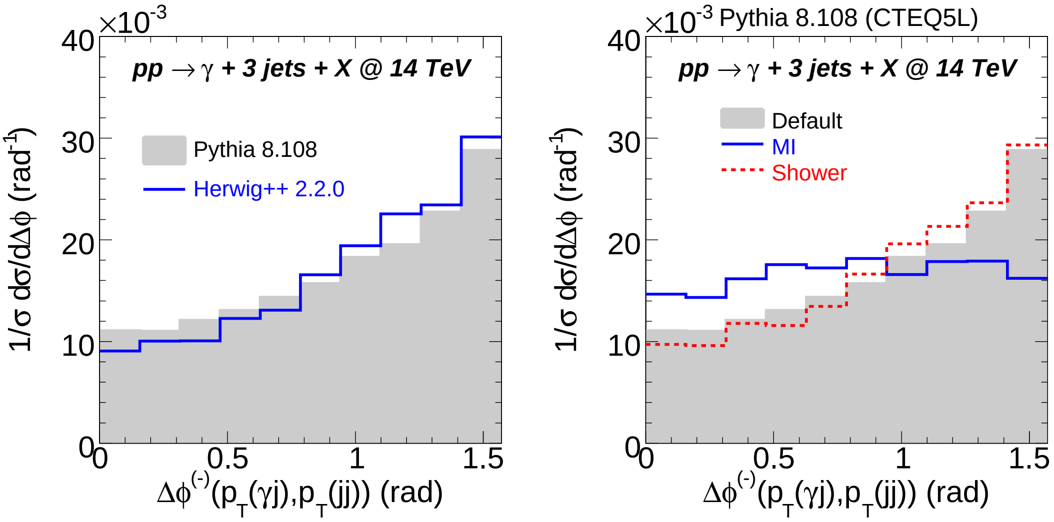


Figure 5.19: Differential cross section shape as a function of $\Delta\phi^{(-)}$(AFS) (Eq. 5.17). Predictions from Pythia (*Default* scenario) and Herwig (left panel) and from three different Pythia settings (right panel) shown.

Differential cross section shape predictions for the variable suggested by AFS, $\Delta\phi^{(-)}$, are shown in Fig. 5.19. Herwig and Pythia predict similar cross section shapes for the default settings which include multiple interactions and showering (Fig. 5.19-left). With multiple interactions switched off, $\Delta\phi^{(-)}$ is indeed most likely to be $\Delta\phi^{(-)} \approx \pi/2$. However, the correlation is weak with a factor of 3 between first bin and last bin, i. e. between events with both pairs being aligned in azimuth and events being orthogonal in azimuth. In fact, the difference between Pythia's *Default* and *Shower* scenarios is not significant within the available statistics (Fig. 5.19-right). Yet, both pairs are more or less uncorrelated if additional jets come from multiple interactions (*MI* scenario, Fig. 5.19-right).

Differential cross section shape predictions for the variable suggested by CDF, $\Delta\phi^{(+)}$, are shown in Fig. 5.20. Differences between Herwig and Pythia are especially pronounced for small $\Delta\phi^{(+)}$, corresponding to the photon-jet pair and the dijet pair both pointing in the same direction in azimuth (Fig. 5.20-left). Pythia predicts a larger fraction of uncorrelated pairs than Herwig does. Strong differences can also be seen when comparing Pythia's different simulation scenarios with each other (Fig. 5.20-right). As noted before, jets from initial- or final-state showers dominantly point away from the photon and combinations with small $\Delta\phi^{(+)}$ are largely suppressed. However, if additional jets come from multiple interactions (*MI* scenario), the dijet pair can have any orientation with respect to the photon-jet pair, thus the predicted distribution is

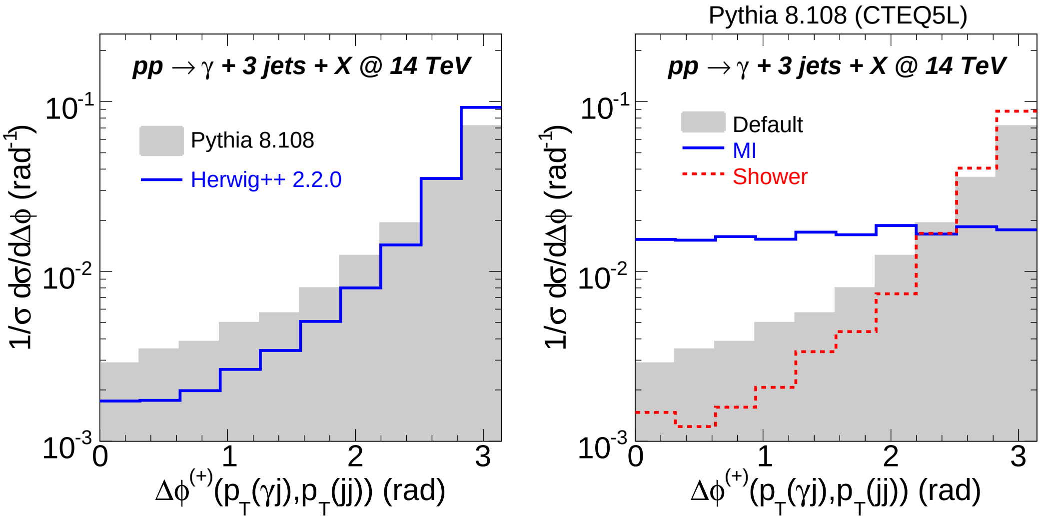


Figure 5.20: Differential cross section shape as a function of $\Delta\phi^{(+)}$(CDF) (Eq. 5.18). Predictions from Pythia (*Default* scenario) and Herwig (left panel) and from three different Pythia settings (right panel) shown. Note the logarithmic scale.

approximately flat. This large difference between the simulation scenarios makes $\Delta\phi^{(+)}$ an promising observable to search for double-parton-scattering.

We have noted earlier that although all cross section predictions on prompt photon production agree for the generator settings under study, specific topologies, for instance the predicted rate of additional jets, vary strongly. To give an example observable, we study predictions on the differential cross section with respect to the photon's transverse energy in the presence of three jets; the dijet pair is required to be balanced in azimuth, i. e. fulfilling

$$\cos\Delta\phi(\text{jet }2, \text{jet }3) \leq -0.9\ , \tag{5.19}$$

still imposing the selections listed in Table 5.2. Such a selection enriches the event sample with topologies typical for those originating from double-parton-scatters.

Fig. 5.21 shows cross section predictions for transverse photon energies in the range 50 GeV $< E_T(\gamma) <$ 90 GeV. Cross section predictions from Herwig are significantly smaller than predictions from Pythia (Fig. 5.21-left). Differences are even more pronounced when comparing $E_T(\gamma)$ spectra with jets from initial- or final-state radiation (*Shower* scenario) with $E_T(\gamma)$ spectra with jets from multiple interactions (*MI* scenario, Fig. 5.21-right). Throughout the investigated transverse energy range, approximately five times as many events from multiple interaction processes are predicted than events

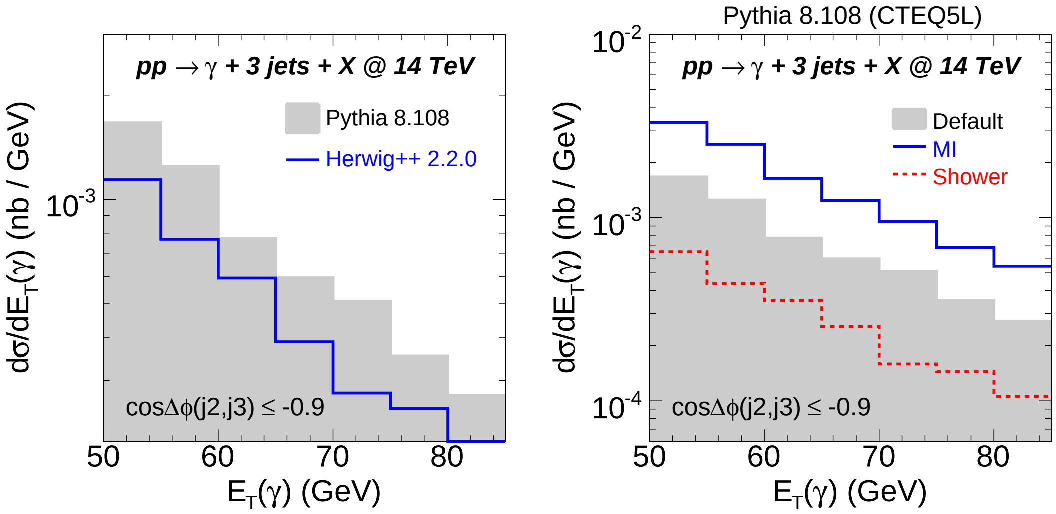


Figure 5.21: Differential cross section predictions with respect to $E_T(\gamma)$ for balanced dijet pairs. Predictions from PYTHIA and HERWIG (left) and from three PYTHIA settings shown. Note the logarithmic scale.

with additional jets from initial- or final-state radiation. Up to this point, we have seen that the study of prompt photon events with additional jets might give insight into a number of jet production mechanisms, in particular the rate of double-parton scatters. In the subsequent section, we will turn to the question of whether back-to-back pairs in γ+3-jet events exhibit kinematical correlations.

5.2.3 Bjorken-x Dependence and x Correlations

The CDF collaboration found the effective cross section to be independent of the scattering kinematics [121], hinting that the position and momentum densities of partons indeed factorize. A dependence of the parton spatial density on the proton momentum fraction x carried by the struck parton is for instance present in *hot core* models of the proton. A hot-core proton concentrates high-x partons in its center. When probed at higher x, the proton's effective size shrinks such that the matter overlap depends on the kinematics. As has been discussed earlier, a dependence of the effective cross section σ_{eff} (cf. the introduction to Sec. 5.2) would also undermine the validity of factorized two-parton densities.

In order to study a possible x dependence in double-parton-scattering events, the PYTHIA samples are enriched with "signal topologies" by requiring azimuthal balancing

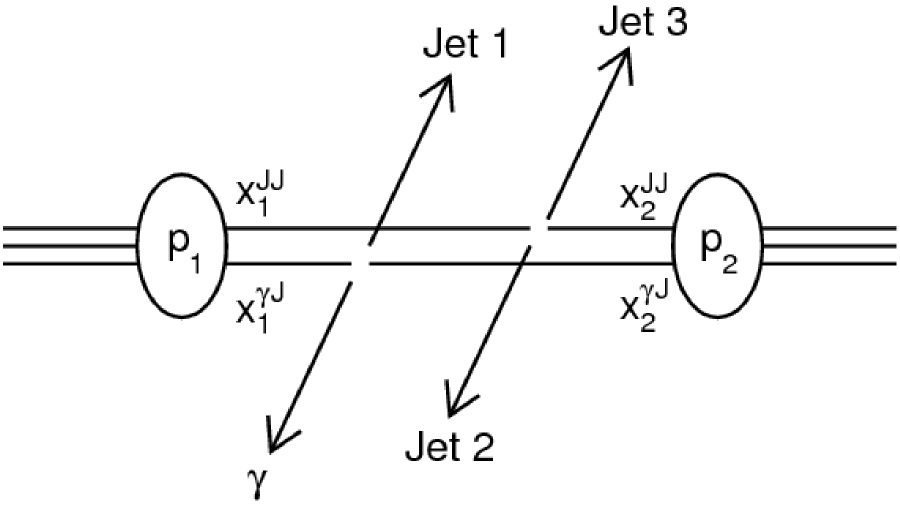


Figure 5.22: Illustration of parton momentum fractions: Four partons contribute to double-parton-scattering. The figure illustrates the extraction of the respective protons' momentum fractions x.

between the pairs:

$$\cos\Delta\phi(\gamma, \text{jet } 1) \leq -0.9 \ , \tag{5.20}$$

$$\cos\Delta\phi(\text{jet } 2, \text{jet } 3) \leq -0.9 \ . \tag{5.21}$$

Double-parton-scattering events have four partons in the initial state and hence four values of Bjorken-x. From each of the two protons p_1 and p_2, two partons enter hard scatters with momentum fractions $x_1^{\gamma J}$, x_1^{JJ}, $x_2^{\gamma J}$, and x_2^{JJ} (note that the momentum fractions were called x_A, x_B, x'_A, and x'_B in the general discussion earlier). The partons producing the γ-jet pair carry momentum fractions $x_{1,2}^{\gamma\mathrm{J}}$, and the partons producing the dijet pair carry momentum fractions $x_{1,2}^{\mathrm{JJ}}$. These momentum fractions are extracted from the photon's and jets' transverse momenta and pseudorapidities [130, 131] (cf. Fig. 5.22):

$$x_{1,2}^{\gamma\mathrm{J}} \equiv \frac{p_T^{\gamma}}{\sqrt{s}} \left(e^{\pm\eta_\gamma} + e^{\pm\eta_J}\right)$$

$$x_{1,2}^{\mathrm{JJ}} \equiv \frac{E_{T2} + E_{T3}}{2\sqrt{s}} \left(e^{\pm\eta_2} + e^{\pm\eta_3}\right) \tag{5.22}$$

In this way, it is possible to reconstruct the momentum fractions entering a double hard scatter that subsequently produces a photon-jet pair and a dijet pair. As a cross check of the method, in Fig. 5.23, the thus-obtained momentum fractions from proton 1 (left) and proton 2 (right) entering the hard prompt photon process are compared with the

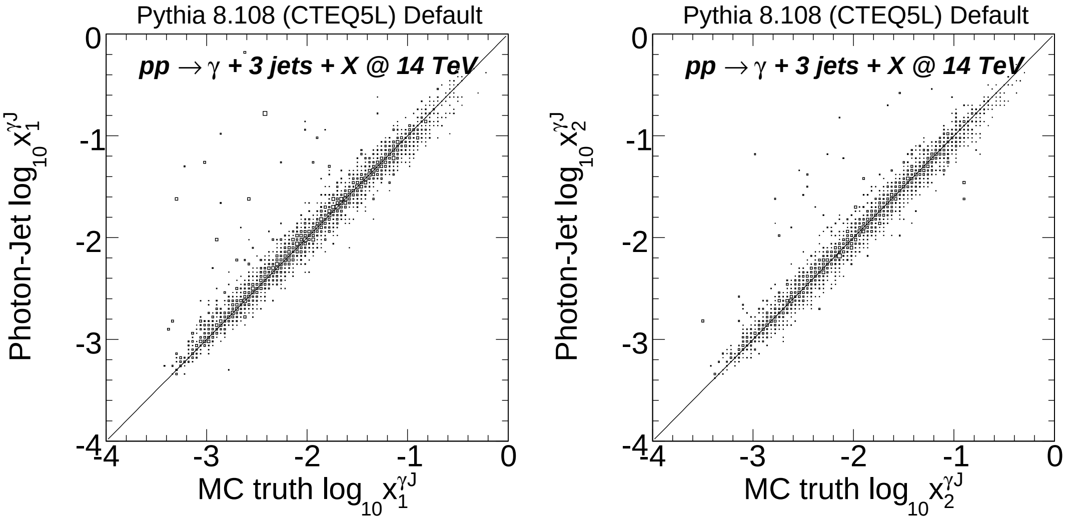


Figure 5.23: Proton momentum fractions entering the prompt photon production process determined with the photon-jet pair (Eq. 5.22) shown as a function of Monte Carlo truth. Correlation predictions shown for proton 1 (left) and proton 2 (right).

generator truth. The simulated parton momenta are successfully reconstructed with the described kinematic variables. Note that $x_{1,2}^{\mathrm{JJ}}$ represent parton momentum fractions only if the dijet pair has been produced by a second hard scatter. If on the other hand double-Bremsstrahlung produces additional jets, each proton contributes only one parton to the hard process and potential correlations between x_i^{JJ} and $x_i^{\gamma\ \mathrm{J}}$ only reflect that three hard jets are required by the event selection.

Next, we would like to correlate parton momenta from the same proton entering photon-jet or dijet production processes. Of particular interest is the question whether Pythia picks the momentum of the parton in the second hard process such that the momentum choice is influenced by the first parton. When studying correlations in the respective proton momentum fractions, partons from the same proton must be compared with each other: $x_1^{\gamma\mathrm{J}}$ is compared with x_1^{JJ} and $x_2^{\gamma\mathrm{J}}$ is compared with x_2^{JJ}. For each event, two combinations are thus filled into the correlation histogram.

Differential cross section predictions with respect to the partons' Bjorken-x entering photon-jet production (left) and dijet production (right) are shown in Fig. 5.24. The lower threshold on the photon's transverse energy restricts parton momenta entering the prompt photon production process to be above $x \approx 4 \cdot 10^{-4}$. As discussed above, $x_{1,2}^{\mathrm{JJ}}$ represent parton momentum fractions only for the *MI* scenario. Other scenarios are shown for completion. Given the lower thresholds on jet transverse energies, x^{JJ}

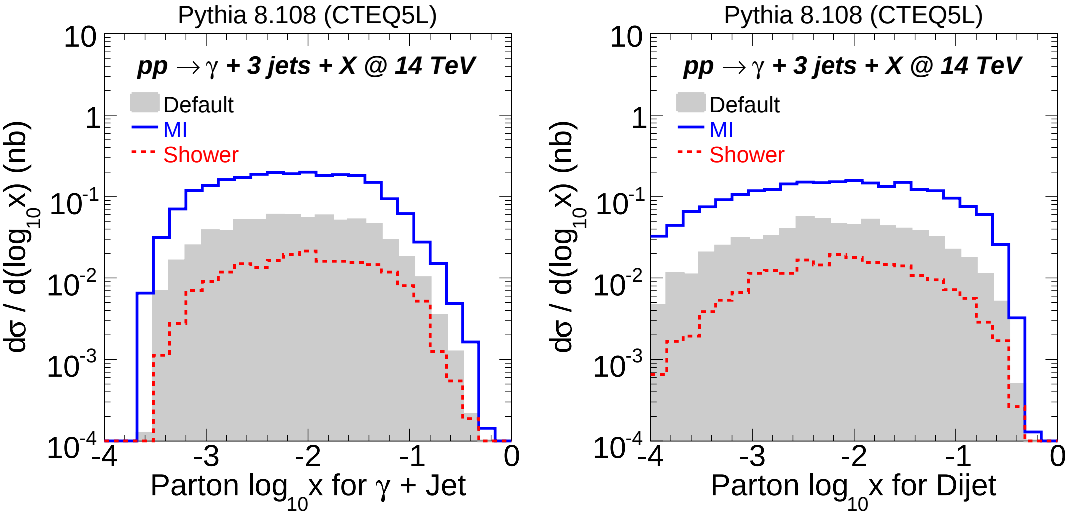


Figure 5.24: Differential cross section predictions with respect to Bjorken-x for partons entering photon-jet production (left) and dijet production (right) shown. Predictions from Pythia for three generator settings.

(right panel) is distributed at smaller values than $x^{\gamma\ J}$ is (left panel). From Fig. 5.12, it can be seen that gluons are the dominant parton species in this kinematic range. Given the differences between default parton densities chosen by Pythia and Herwig, a difference in 20% (and possibly more) between both programs might be fully attributed to uncertainties of parton densities. A future study will need to assess both generator's predictions with the same parton distributions.

Fig. 5.25 illustrates the correlation between momentum fractions entering dijet and photon-jet production. Distributions are shown for Pythia *Default* settings (left panel) and Pythia *MI* settings (right panel). Dijet pairs from double-Bremsstrahlung lead to a slightly positive correlation for the *Default* scenario. In contrast, no correlations between momentum fractions entering separate scatters (*MI* scenario, right panel) are observed. If Pythia had simulated the matter overlap using the hot core model mentioned earlier, a positive correlation would have been observed and probing at high x would have led to an enhanced probability to find another parton at high x.

In this subsection, we have analyzed events with back-to-back photon-jet and dijet pairs to investigate momentum correlations between partons entering double high-p_T scatters. A measurement of these correlations might yield insight on whether a model analogous to the hot-core model is needed to describe the spatial distributions of partons

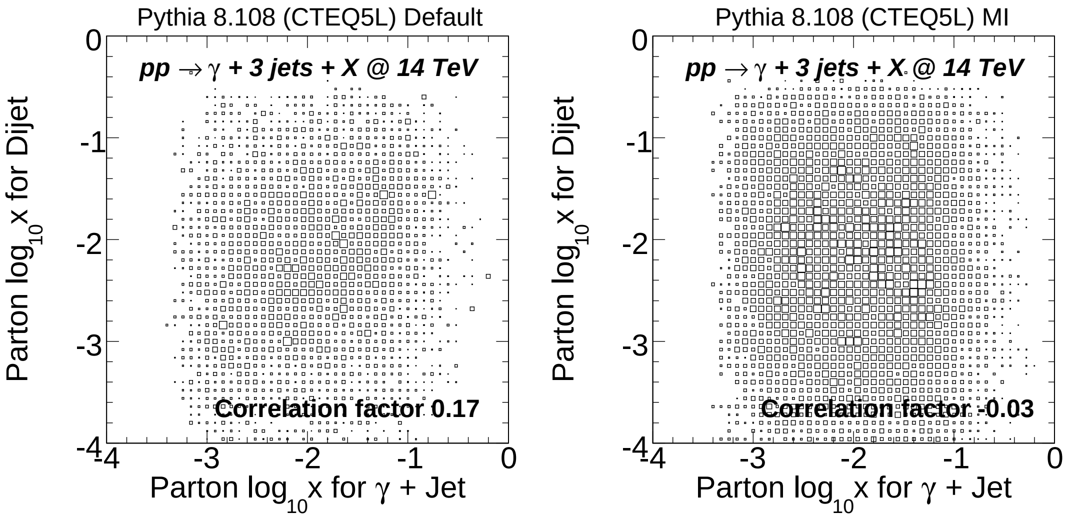


Figure 5.25: Correlation predictions from Pythia on extracted parton momentum fractions for *Default* (left) and *MI* (multiple interactions, right) settings shown. Note the double-logarithmic scale.

in the proton. Correlations like in the hot-core model are not included in the multiple-interaction models of Pythia and Herwig++.

5.2.4 Conclusions

In this section, we have studied various approaches to identifying double-parton-scatters in proton-proton interactions. Studies are performed on a final state composed of one photon and three jets, along the lines of a previous study by the CDF collaboration [121]. Different predictions from Herwig and Pythia can in part be attributed to different default choices of parton densities in both programs. However, in some observables, both models yield clearly different differential predictions, most notably with respect to the $\Delta\phi^{(+)}$ variable put forward by CDF. It should be noted, however, that the imposed selection cuts were only a first approximation to an extrapolation to the LHC. More studies will be needed to find the optimal selection cuts and to assess their experimental feasibility.

The one-dimensional variables under study try to describe correlations in four-object final states. This is likely to be a too simplistic approach and higher-dimensional observables might perform better to extract a double-parton-scattering signal at the LHC.

Chapter 6

Performance of Physics Object Reconstruction

Observables to study the underlying event and multiple interactions in charged-jet topologies are discussed in Chap. 5. Input objects to construct these observables are tracks and charged-jets, i. e. jets from tracks. Clearly, dominant experimental uncertainties are thus inherently based on the performance of the CMS central tracking detectors.

In this chapter, we study the reconstruction quality of those tracks that will be selected for the underlying event analysis. Charged-jets are reconstructed from these selected tracks and the properties of these jets are investigated: Comparisons with hadron-level predictions as well as with calorimeter jets are shown. The chapter concludes with a study of further charged-jet properties such as jet size, shape, and fragmentation.

6.1 Performance of Track Reconstruction

The reconstruction of charged particles in the CMS inner tracking detectors has been discussed in Sec. 3.2.3. Imperfections of the track reconstruction procedure can have several effects: A charged particle might not be reconstructed and identified as a track, for instance because its hit measurements are assigned to another track. Another consequence of imperfect tracking is the reconstruction of *fake tracks*, i. e. hit patterns that were in fact not caused by a single charged particle but by several, by noise, or by misidentified multiple scatterings in the detector material. Failure to detect a charged particle results in less tracks per event, whereas fake tracks increase the number of tracks per event. Ideally, an analysis using tracks (such as the underlying event analysis) should strive for low fake rates and high track reconstruction efficiencies. Those characteristics are studied in the present section on a fully simulated data sample, HerwigQCDPt80, produced during the SUMMER08 effort (cf. Sec. 4.2 and Tab. 4.6).

6.1.1 Optimal Filtering of Fake Tracks

Any track-based analysis should attempt to suppress fake tracks; at the same time, it should attempt to retain a high reconstruction efficiency of true tracks. In order to identify and study fake tracks, one has to choose an association method between reconstructed tracks and simulated tracks. The standard approach within CMS is the following [132]: Reconstructed tracks are associated to simulated tracks if at least 50% of the reconstructed track hits are shared with the simulated track. If only few hits are reconstructed, the robustness of this association is improved by requiring reconstructed and simulated track transverse momenta to not deviate by more than 20% from the simulated track's p_T:

$$\left| \frac{p_T(\text{reconstructed track}) - p_T(\text{simulated track})}{p_T(\text{simulated track})} \right| < 20\% \; . \tag{6.1}$$

According to this definition, tracks that are not associated to any simulated track are called *fake tracks*. The standard CMS combinatorial track finder (cf. Sec. 3.2.3) will reconstruct a significant fraction of fake tracks in an LHC event of average particle density. Suppressing tracks with a poor quality of the final fit and a poor compatibility with the event interaction vertex is expected to reduce the fake rate. However, quality cuts must be adapted to track pseudorapity, transverse momentum, and number of hits to retain the best possible efficiency [132]. We will discuss the optimal track selection for the underlying event analysis in the following.

In Fig. 6.1, we study how the tracking fake rate depends on the number of hit measurements. The shown distributions are for tracks in the central region with pseudorapidity $|\eta| < 0.3$ and with a transverse momentum $p_T > 1$ GeV/c and 10 GeV/c. For one, this selection ensures that tracks have enough transverse momentum to cross all silicon layers of the tracker and do not get curled up inside (*loopers*). In addition, the selected pseudorapidity prevents tracks from crossing disk or endcap modules, thus, the number of hits should be the same for all selected tracks. The tracker geometry suggests 17 possible hit measurements from the 3 pixel and 10 strip layers, of which 2 layers in the tracker inner barrel and 2 layers in the tracker outer barrel carry double-sided

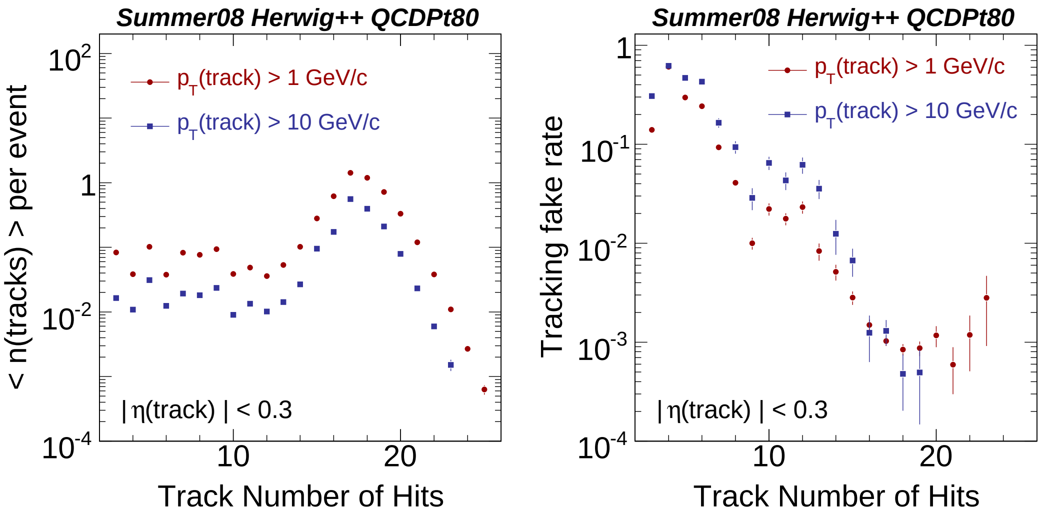


Figure 6.1: Average number of tracks found per event as a function of the number of hit measurements per track (left) and relative fake rate versus number of hit measurements (right) for tracks in the central region ($|\eta| < 0.3$). The tracker geometry suggests 17 crossed silicon detectors.

modules. With the above selection, each track should thus cross 17 silicon detectors, but since the tracker has overlaps between layers, 17 will not be a sharp edge in the distribution of the number of track hit measurements, as can be seen from Fig. 6.1 left. Tracks with large curvature are more likely to cross these overlaps (the “cross section” of the overlap region increases). To cross-check this assumption, hit distributions for two p_T-thresholds are compared: The maximum number of hits per track decreases, but there is still no sharp cutoff above 17 hit measurements.

The right panel of Fig. 6.1 shows the relative rate of fake tracks, i. e. the average fraction of reconstructed tracks per event that cannot be associated to a simulated particle, versus the number of track hits. Tracks with 14 or more hit measurements have an ab-initio low fake rate well below 1%. As a consequence, the number of hit measurements is used to weight the respective track selection cuts: Loose selections will be applied to tracks with many hits since those tracks are less likely to be fakes; tracks with few hits have to pass more stringent criteria to suppress possible fake tracks.

We adopt track selections optimized by the CMS tracking group [132] such that fake tracks are suppressed and the tracking efficiency retained: Track quality cuts are adapted to the number of hit measurements, transverse momentum p_T, and pseudorapidity η. Table 6.1 summarizes the formulas applied to select tracks, where δd_0 is the measured

Table 6.1: Track quality cuts. The selection cuts depend on the number of hit measurements following Fig. 6.1. Parameters $\{\alpha_i\}$ have been optimized to suppress fake tracks while retaining the tracking efficiency [132].

χ^2/ν (chi-square per degree of freedom)	$\chi^2/\nu < \alpha_0 \; n(hits)$
d_0 to the beam spot	$\lvert d_0 \rvert < (\alpha_1 \; n(hits))^4 \sigma_{d_0}(p_T)$
Δz to position of closest HLT primary vertex	$\lvert \Delta z \rvert < (\alpha_2 \; n(hits))^4 \sigma_{z_0}(p_T, \eta)$
transverse compatibility with beam spot	$\lvert d_0 \rvert / \delta d_0 < (\alpha_3 \; n(hits))^4$
longitudinal compatibility with closest HLT vertex	$\lvert \Delta z \rvert / \delta z_0 < (\alpha_4 \; n(hits))^4$

error on the transverse impact parameter d_0 and δz_0 is the measured error on the z-vertex position (for a discussion of common track parameters see Sec. 3.2.3).

For *tight* and *loose* tracks, the track transverse impact parameter resolution is parametrized following the functional form

$$\sigma_{d_0} = 30 \text{ µm} + \frac{100 \text{ µm}}{p_T \text{ (GeV}/c)} \; . \tag{6.2}$$

In addition, it is assumed that the resolution in z_0 follows the transverse impact parameter resolution, weighted by a factor accounting for different polar angles:

$$\sigma_{z_0} = \frac{\sigma_{d_0}}{\sin\theta} \; . \tag{6.3}$$

The tight selection can be refined further by selecting tracks qualified as *high purity*. For these, the resolution parametrization reads

$$\sigma_{d_0} = \sigma_{z_0} \sin\theta = 30 \text{ µm} + \frac{10 \text{ µm}}{p_T \text{ (GeV}/c)} \; . \tag{6.4}$$

Table 6.2 lists resulting track quality cuts for three example tracks with pseudorapidity $|\eta| = 0.8$. Optimal choices for $\{\alpha_i\}$ [132] are given in the first column. Tight selection and high-purity selection only differ in stricter impact parameter cuts for the latter due to different resolution parametrizations (cf. Eq. 6.2 and 6.4). The *loose* cut in the transverse impact parameter d_0 for tracks with $p_T = 3$ GeV/c, $\eta = 0.8$, and n(hits) $= 15$ has no effect, since tracks with $|d_0| = 28$ cm would originate well outside the pixel detector.

Table 6.2: Track quality selections for three example tracks with pseudorapidity $\eta = 0.8$, following the parametrization in Table 6.1 and Equations 6.2 and 6.4. The first column lists parameter choices $\{\alpha_i\}$ corresponding to track quality labels *loose*, *tight* and *high purity*.

		$p_T = 0.7$ GeV/c, n(hits) $= 5$	$p_T = 3$ GeV/c, n(hits) $= 7$	$p_T = 3$ GeV/c, n(hits) $= 15$
loose				
$\alpha_0 = 2.0$	χ^2/ν	10	14	30
$\alpha_1 = 0.55$	$\lvert d_0 \rvert$	11 mm	13 mm	28 cm
$\alpha_2 = 0.65$	$\lvert \Delta z \rvert$	22 mm	26 mm	54 mm
$\alpha_3 = 0.55$	$\lvert d_0 \rvert / \delta d_0$	57	220	4600
$\alpha_4 = 0.45$	$\lvert \Delta z \rvert / \delta z_0$	25	100	2100
tight				
$\alpha_0 = 0.9$	χ^2/ν	4.5	6.3	13.5
$\alpha_1 = 0.3$	$\lvert d_0 \rvert$	750 μm	875 μm	18 mm
$\alpha_2 = 0.35$	$\lvert \Delta z \rvert$	1.9 mm	2.2 mm	45 mm
$\alpha_3 = 0.4$	$\lvert d_0 \rvert / \delta d_0$	16	61	1300
$\alpha_4 = 0.4$	$\lvert \Delta z \rvert / \delta z_0$	16	61	1300
high purity				
$\alpha_0 = 0.9$	χ^2/ν	4.5	6.3	13.5
$\alpha_1 = 0.3$	$\lvert d_0 \rvert$	224 μm	648 μm	14 mm
$\alpha_2 = 0.35$	$\lvert \Delta z \rvert$	554 μm	1.6 mm	34 mm
$\alpha_3 = 0.4$	$\lvert d_0 \rvert / \delta d_0$	16	61	1300
$\alpha_4 = 0.4$	$\lvert \Delta z \rvert / \delta z_0$	16	61	1300

In the following, we study how well charged particles produced with pseudorapidities within the tracker coverage ($|\eta| < 2.5$) and transverse momenta p_T above 0.9 GeV/c can be reconstructed and qualified as high purity, tight, or loose tracks. Fake rates are estimated along with transverse momentum resolutions.

6.1.2 Tracking Performance Studies

Tracking performance is studied by evaluating tracking efficiency, fake rate, and transverse momentum resolution versus track pseudorapidity and versus track transverse momentum on the fully simulated HerwigQCDPt80 sample (cf. Sec. 4.2 and Tab. 4.6). Simulated and reconstructed tracks must fulfill the following selections:

- Tracks must be hard enough to cross all tracking layers, i. e. loopers are excluded (see below): $p_T(\text{track}) > 0.9$ GeV/c.
- Tracks must emerge into the geometric tracker acceptance: $|\eta(\text{track})| < 2.5$.

In addition to these preselections, reconstructed tracks must have one of the quality labels *loose*, *tight*, or *high purity*. In the context of this analysis, we define *tracking efficiency* as the probability that a simulated track is succesfully reconstructed as a track with quality label loose, tight, or high purity. Since there are no further requirements on the simulated tracks, for instance a minimal number of hits, this efficiency corresponds to a global efficiency to successfully reconstruct a charged particle. *Successful reconstruction* means that the simulated and the reconstructed track are associated to each other (in the context of Sec. 6.1.1) and that the simulated track shares more hits with this particular track than it does with any other reconstructed track. With such a definition of "successful track reconstruction", *track loopers*, i. e. soft tracks curling up inside the tracker, will typically have significantly more simulated hits than reconstructed hits have, because the default tracking algorithm only propagates tracks from inner to outer layers, but not from outer to inner layers. In a magnetic field of $B = 3.8$ T, tracks that will not reach the outer layers of the CMS central tracker have transverse momenta below $p_T(\text{track}) < 0.7$ GeV/c.

The *rate of fake tracks* is the average fraction of reconstructed tracks per event that cannot be associated to a simulated particle. The *transverse momentum resolution* is the standard deviation of the quantity $\delta p_T / p_T$, where $\delta p_T = |p_T(\text{measured}) - p_T(\text{true})|$.

The global tracking efficiency is depicted in Fig. 6.2 as a function of pseudorapidity (left panel) and as a function of transverse momentum (right panel). Efficiencies for

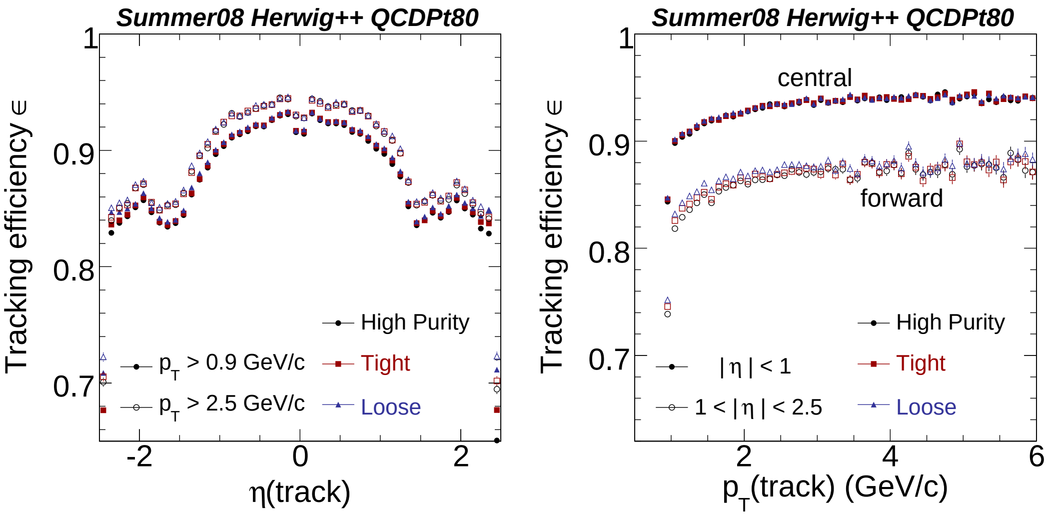


Figure 6.2: Global tracking efficiency distributions to reconstruct charged particles as high purity, tight or loose tracks. The efficiency is shown for p_T thresholds 0.9 GeV/c and 2.5 GeV/c as a function of pseudorapidity (left) and for $|\eta| < 1$ ("central") and $1 < |\eta| < 2.5$ ("forward") as a function of transverse momentum (right). Note that the global tracking efficiency is not independent of p_T in the region 0.9 GeV/$c < p_T$(track) < 1.5 GeV/c (cf. Sec 7.3.2).

two track p_T thresholds (open and closed symbols) are compared with each other in the left panel; two track pseudorapidity regions (open and closed symbols) are compared with each other in the right panel. Distributions are displayed for tracks classified *loose* (triangles), *tight* (squares), or *high purity* (circles) according to the above described criteria. It can be seen that the efficiency for central tacks with $p_T > 2.5$ GeV/c approaches 95%. In contrast to that, global tracking efficiencies in the forward regions are below 90% even for tracks with $p_T > 2.5$ GeV/c. A small drop in efficiency for η(track) ≈ 0 can be attributed to non-overlapping modules in this region. Comparing the efficiency distributions with the distribution of material budget in the inner tracking system (Fig. 3.4), one is let to conclude that the dominant source of efficiency loss is multiple scattering and secondary hadronic interactions in the tracker material. The global tracking efficiencies exhibit no dependence on the track quality label. Thus, no effect on the tracking efficiency as a function of pseudorapidity or transverse momentum is observed when tightening the track quality selection criteria.

The rate of fake tracks is displayed in Fig. 6.3 as a function of pseudorapidity for two p_T thresholds (left panel) and as a function of transverse momentum for two η bins (right panel). In general, fake rates for loose track selections are significantly larger than fake rates for tight or high-purity selections, which barely differ over the entire

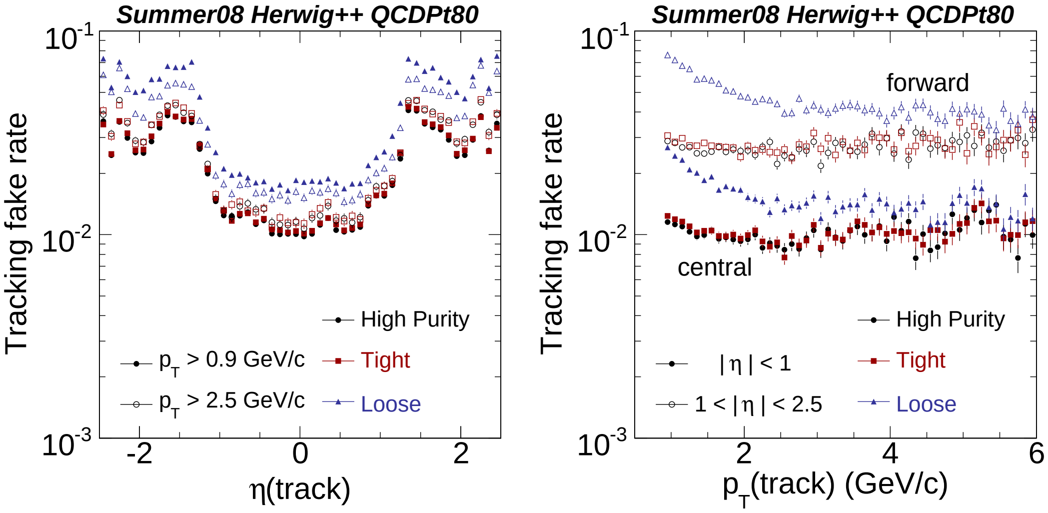


Figure 6.3: Fraction of fake tracks as a function of pseudorapidity for p_T thresholds 0.9 GeV/c and 2.5 GeV/c (left panel) and as a function of transverse momentum for $|\eta| < 1$ ("central") and $1 < |\eta| < 2.5$ ("forward") (right panel). Distributions are displayed for tracks with quality labels loose, tight, or high purity.

(η, p_T) region under study. For the tight track selections, the fake rate is of the order of 1% for $|\eta| < 1$ and increases to approximately 4% in the pseudorapidity regions with higher material budget ($1 < |\eta| < 2$). Tracking fake rates as a function of transverse momentum for $|\eta| < 1$ and $1 < |\eta| < 2.5$ are shown in Fig. 6.3-right. For transverse momenta $p_T < 5$ GeV/c, the set of loose tracks is contaminated by approximately twice as many fakes than the sets of tight or high purity tracks are. For higher transverse momenta, fake rates for all three track quality selections are comparable. In summary, the tracking fake rate is suppressed when moving from loose to tight or high-purity selections.

We now turn to assessing the p_T resolution of tracks. Fig. 6.4 shows the track transverse momentum resolution (obtained by a Gaussian fit) as a function of pseudorapidity for $p_T > 0.9$ GeV/c and $p_T > 2.5$ GeV/c (left) and as a function of transverse momentum for $|\eta| < 1$ and $1 < |\eta| < 2.5$ (right). All three track quality classes exhibit comparable resolutions ranging from less than 1% in the central region and 2 GeV/c $< p_T <$ 4 GeV/c to $\sim$2% for tracks with pseudorapidities $|\eta| > 2$. For small transverse momenta $p_T < 1$ GeV/c, the track p_T resolution falls with increasing transverse momentum due to an enhanced probability for multiple scattering. Above $p_T > 5$ GeV/c for central tracks and above $p_T > 2.5$ GeV/c for forward tracks, the track p_T resolution rises with

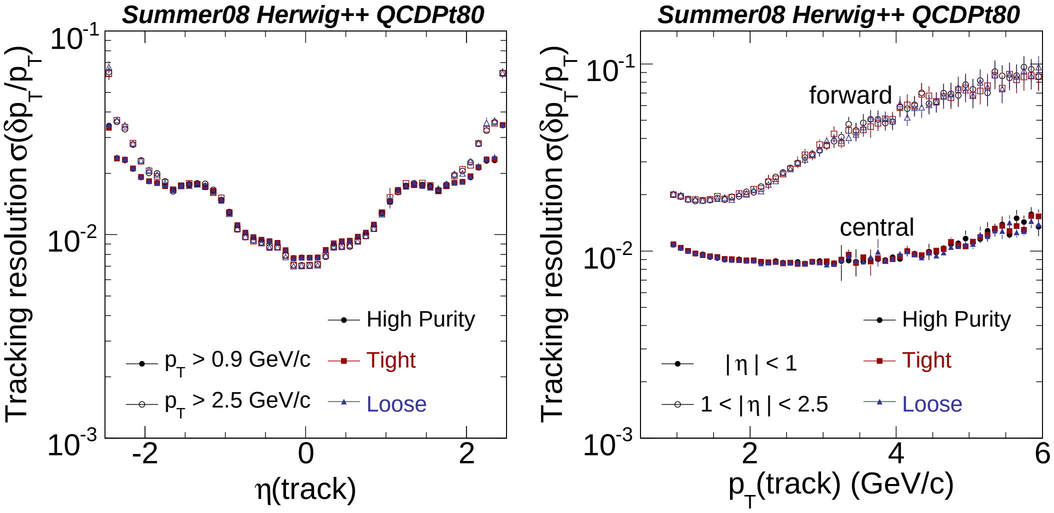


Figure 6.4: Track transverse momentum resolution depicted as a function of pseudorapidity for $p_T > 0.9$ GeV/c and $p_T > 2.5$ GeV/c (left) and as a function of transverse momentum for $|\eta| < 1$ and $1 < |\eta| < 2.5$ (right). Tracks are classified high purity, tight, or loose.

increasing transverse momentum, as expected for measurements dominated by the single hit resolution.

6.1.3 Conclusions

In this section, we have studied methods that suppress fake tracks and, at the same time, maintain the track reconstruction efficiency as high as possible. Track quality selections have been described, that are tightened or loosened depending on the number of hits, p_T, and η. Tracking efficiencies and transverse momentum resolutions have been studied with a fully simulated Monte Carlo sample from the SUMMER08 production and are found to be the same regardless of the particular set of track quality selections. Efficiencies are above 90% in the central region $|\eta| < 1$ and decrease in the forward direction due to an increased amount of material. Transverse momentum resolutions are typically of the order of a few percent. On the other hand, the rate of fake tracks when imposing loose track quality criteria is significantly larger compared to the rate of fake tracks when imposing tight or high-purity selections. Thus, as expected, the presented track quality selection suppresses fake tracks and retains the tracking efficiency. Both "tighter" quality labels, tight and high purity, qualify to be used for underlying event studies - the current studies do not exhibit different tracking performances for the two.

We choose to carry out the underlying event analysis with tracks fulfilling high-purity quality track criteria. Future studies will have to assess possible differences to charged-jet topologies identified with tracks fulfilling only tight quality criteria.

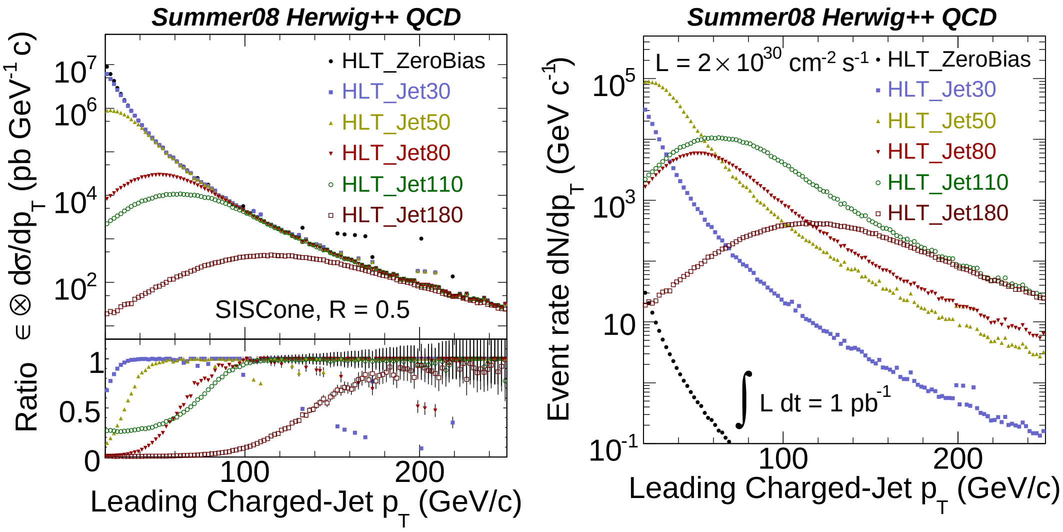
It should be noted that there are alternative methods to study the tracking performance that do not rely on simulated data. Cosmic muons can be used to determine the global reconstruction efficiency, by tagging muons that have been identified in the muon detectors and by subsequently probing whether a corresponding track has been found in the central tracker. In a similar manner, tracks from strip hits can be compared to tracks from pixel hits. From collision data, resonant decays such as e. g. $J/\psi \rightarrow \mu\mu$ or $B \rightarrow D\pi$ will be used by imposing constraints on the invariant mass of track combinations. All these methods will eventually be cross-checked with the performance studies that were carried out using simulation data in order to obtain the most reliable estimate of the tracking performance.

6.2 Characteristics of Charged-Jets

In this section, we turn to studying the properties of *charged-jets*, i. e. jets reconstructed from tracks. High purity tracks (as described in the previous section) are clustered into charged-jets by a SISCone jet algorithm [46] with radius $R = 0.5$ and longitudinal energy recombination scheme (cf. Sec. 3.2.7 for a description of the SISCone algorithm). After investigating the distribution of leading charged-jet transverse momenta and pseudorapidities for various high-level trigger streams, we address the question of how well charged-jets represent the particle flow from the hard scatter. We conclude with studies of typical jet properties (shape, size, and fragmentation) using charged-jets.

6.2.1 Jet Trigger Efficiency

Of particular interest to the underlying event analysis is the question, how charged-particle activities evolve with increasing event scales p_T(charged-jet), as discussed in Section 5.1. Thus, a thorough understanding of the charged-jet and its properties is of vital importance to the underlying event analysis. In order to collect a significant number of events with high-p_T jets, the analysis makes use of several high-level triggers, as listed in Table 4.8. These typically require a calorimetric jet with transverse energy above a configurable threshold in the event. As underlying event activities are studied



Summer08 Herwig++ QCD

Summer08 Herwig++ QCD

Figure 6.5: Left: Uncorrected differential cross section of charged-jet production as a function of transverse momentum, shown for six high-level trigger streams. The lower panel shows ratios between subsequent trigger streams. Right: Expected event rates for $\int \mathcal{L}dt = 1\ \text{pb}^{-1}$ of collected data for various high-level trigger streams. Prescales correspond to an instantaneous luminosity of $\mathcal{L} = 2 \times 10^{30}\ \text{cm}^{-2}\text{s}^{-1}$ (cf. Tab. 4.4). Charged-jets are reconstructed from high purity tracks with $p_T > 0.9$ GeV/c and $|\eta| < 2.5$.

Table 6.3: High-Level Trigger streams under study together with p_T(charged-jet)-ranges as extracted from the HERWIG++ samples produced during SUMMER08, see Fig. 6.5.

Name	p_T(charged-jet)-range
HLT_ZeroBias	$20 - 32$ GeV/c
HLT_Jet30	$32 - 48$ GeV/c
HLT_Jet50	$48 - 88$ GeV/c
HLT_Jet80	$88 - 102$ GeV/c
HLT_Jet110	$102 - 250$ GeV/c
HLT_Jet180	above 250 GeV/c

as a function of the charged-jet transverse momentum, we aim for a correspondence between HLT stream and p_T(charged-jet). This will allow us to make optimal use of the collected event statistics by identifying p_T ranges where a given trigger stream is used solely. The two-level trigger layout of CMS has been described in Sec. 3.2.11. All studies in this section have been performed with SUMMER08 Monte Carlo samples (cf. Sec. 4.2).

Fig. 6.5 shows differential cross sections (left panel) and event rates for an integrated luminosity of 1 pb^{-1} (right panel) at $\sqrt{s} = 10$ TeV. Leading charged-jet distributions are displayed for six high-level trigger streams. Trigger prescale factors for an instantaneous luminosity of $\mathcal{L} = 2 \times 10^{30}$ $\text{cm}^{-2}\text{s}^{-1}$ have been taken into account to estimate event rates (cf. Table 4.4). Large prescale factors on the low-threshold triggers will strongly suppress the recorded number of events from these trigger streams. As a consequence, the number of high-p_T events in the HLT_ZeroBias stream will be much smaller than the number of high-p_T events in e. g. the HLT_Jet110 stream. In the following, we will describe how we attempt to make optimal use of the recorded statistics. Given the prescale factors for high-threshold triggers, we aim to use all trigger streams as efficiently as possible for the purpose of the charged-jet analysis. We will therefore determine the value of p_T(charged-jet) at which a specific trigger becomes fully efficient. As soon as the cross section for a given trigger stream rises above 95% of the trigger stream with the next-lower threshold, the new trigger stream is used. The resulting p_T(jet)-bins corresponding to each trigger are given in Table 6.3: For instance, HLT_Jet50 is used to construct spectra for 48 GeV/c $<$ p_T(charged-jet) $<$ 88 GeV/c. Note that the jet trigger thresholds are applied to calorimeter jets, whereas the lower limit of a specific p_T-range corresponds to how fast a given trigger becomes fully efficient as a function of charged-jet transverse momentum. A charged-jet of $p_T = 48$ GeV/c corresponds to a calorimeter jet of significantly larger transverse energy, since the charged-jet does not include the particle flow of neutrals. In the following, all distributions shown as a function of p_T(charged-jet) are composed from the six HLT streams in their respective p_T-range. So for example, distributions shown in the range 48 GeV/c $<$ p_T(charged-jet) $<$ 88 GeV/c are constructed from events accepted by HLT_Jet50.

6.2.2 Charged-Jet Validation

Relying exclusively on track information, charged-jets are a well-suited tool to study the underlying event in early LHC runs, where the calibration of the CMS hadron calorimeter is poorly known. It remains to be shown that the hard process is sufficiently well correlated with the leading charged-jet, although charged-jets only catch the charged particle flow and ignore the neutral component. In particular, we study correlations of the leading charged-jet (reconstructed from tracks) with the leading charged-particle-jet (reconstructed from stable charged particles at hadron level) to control the reconstruction performance of charged-jets with respect to the Monte Carlo truth. Another object

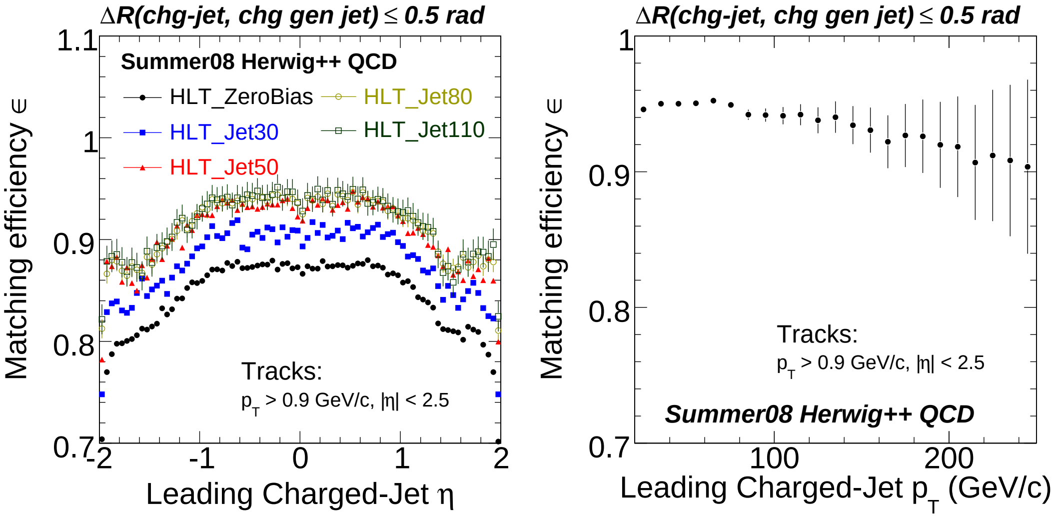


Figure 6.6: Matching efficiency between jet from charged particles at hadron level (*chg. gen. jet*) and charged-jet shown with respect to pseudorapidity (left) and transverse momentum of the charged-particle jet (right). Statistical errors for $\int \mathcal{L}\mathrm{d}t = 1\ \mathrm{pb}^{-1}$ shown.

of interest is the leading calorimeter jet. Comparing charged-jets with calorimeter jets can also be done with data alone and thus provides an important cross-check of the validity of the use of charged-jets. Correlations are searched for in (η, ϕ) space and in transverse momenta.

Validation with respect to Generator Truth

Fig. 6.6 shows the matching efficiency of the leading charged-jet and the leading charged-particle jet. We define "matching efficiency" as the ratio of the number of matched jets over the number of all jets in a given η or p_T bin. Jets at reconstruction and particle level are considered to be matched if their distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ in (η, ϕ) space is less than 0.5 rad. The efficiency is depicted as a function of pseudorapidity (left) and of transverse momentum (right) of the leading charged-particle jet. The charged-jet matching efficiency as a function of pseudorapidity behaves similar to the tracking efficiency (cf. Fig. 6.2): A flat curve in the range $|\eta| < 1$, a drop in efficiency at $\eta = \pm 1.5$ and a rise for $|\eta| > 1.5$. Above $p_T(\mathrm{jet}) = 20$ GeV/c the matching efficiency in $p_T(\mathrm{jet})$ reaches its maximum value to drop again to 90% at 200 GeV/c.

The (η, ϕ) distance between leading charged-jet and charged-particle jet is almost always less than $0.1\ \pi = 0.31$ rad (Fig. 6.7 left) - well within the cone radius of 0.5 (also in-

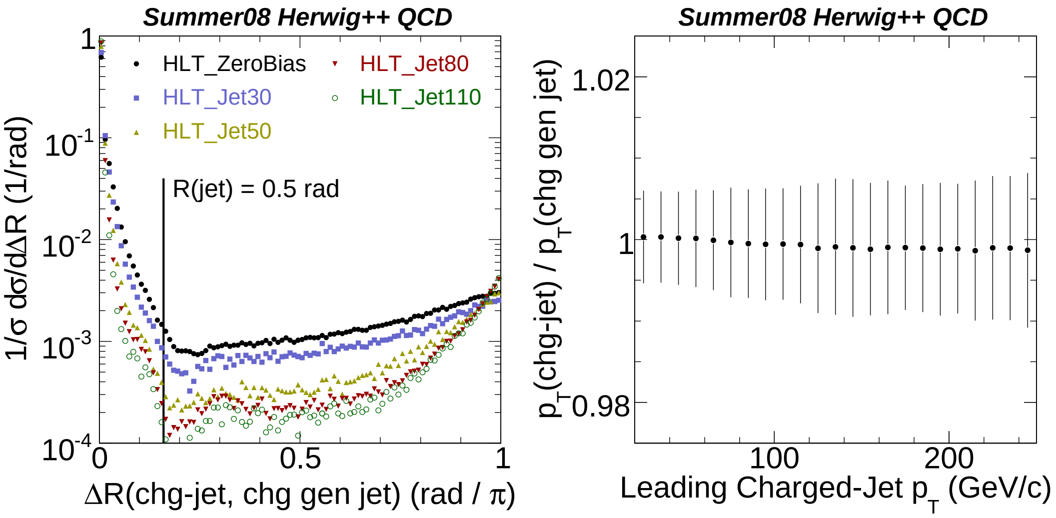


Figure 6.7: Left: Distance in (η, ϕ) space between leading charged-particle jet at hadron level (*chg. gen. jet*) and leading charged-jet. Right: Transverse momentum response shown as a function of charged-jet transverse momentum. Width of Gaussian fit to profile distributions shown.

dicated in the figure). Distributions correspond to the charged-jet transverse momentum ranges for each high-level trigger described in Tab. 6.3. Dijet-like events where the leading charged-jet is associated to the next-to-leading charged-particle-jet are characterized by $\Delta\phi(\text{charged-jet}, \text{chg gen jet}) \approx \pi$. This merely leads to the toward and away regions being flipped but the transverse region remains unchanged: Associating the underlying event with the activity in the transverse region is thus still appropriate. Events with problematic associations are of the type $\pi/3 < |\Delta\phi(\text{charged-jet}, \text{chg gen jet})| < 2\pi/3$. However, these events occur with a low frequency: When integrated over the ΔR region in question, the event rate in question amounts to $10^{-3} \times \pi/3 \approx 0.1\%$.

The right panel of Fig. 6.7 illustrates the excellent p_T(jet)-resolution. The figure shows the most probable ratio between the transverse momenta of charged-jet and charged-particle-jet. The two-dimensional distribution of $p_T(\text{charged-jet})/p_T(\text{chg gen jet})$ versus $p_T(\text{charged-jet})$ has been fitted with a Gaussian for each 10 GeV/c p_T-slice. The error bars correspond to the Gaussian width: The charged-jet's transverse momentum is precise to the level of 1%. The excellent reconstruction performance for charged-jets seems to contradict the results on tracking performance obtained in the previous section (cf. Fig. 6.2). However, the charged-jet performance has been estimated with low requirements on the association of true and reconstructed objects: Jets had to be close to each other in (η, ϕ) (Fig. 6.6) or simply be the ones with largest p_T (Fig. 6.7).

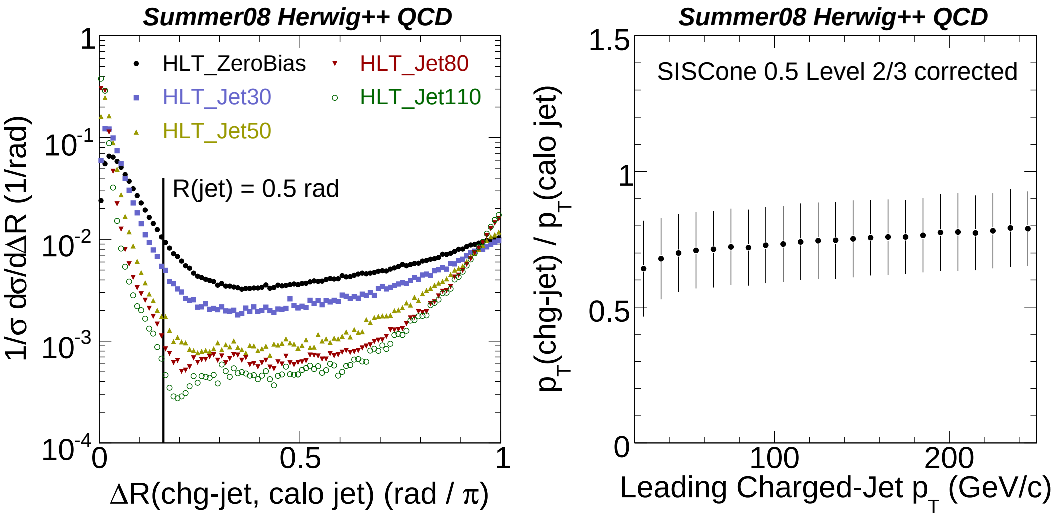


Figure 6.8: Left: Distance in (η, ϕ) space between calorimeter jet (*calo. jet*) and charged-jet (left). Right: Transverse momentum response shown as a function of charged-jet transverse momentum. Mean and width of Gaussian fit to profile distributions shown. Calorimeter jets are corrected in azimuthal balance and absolute scale.

Validation with respect to Calorimeter Jet

Charged-jets are compared to calorimeter jets in Fig. 6.8. Only the mandatory calorimeter jet energy corrections (offset corrections, relative η corrections, and absolute p_T corrections) have been applied (cf. Sec. 3.2.7).

We study the jet-jet distance ΔR in the pseudorapidity-azimuth plane (left) as well as the most probable transverse momentum ratio (right). Distributions correspond to the charged-jet transverse momentum ranges for each high-level trigger described in Tab. 6.3. Jet-jet directions agree nicely for the considered p_T-ranges, in fact the higher the jet threshold, the better charged-jet and calorimeter jet agree in pseudorapidity and azimuth. In most events, the calorimeter jet direction is well within the charged-jet cone radius of 0.5 (also indicated in the figure). As mentioned above, events with problematic associations are of the type $\pi/3 < |\Delta\phi(\text{charged-jet}, \text{calo jet})| < 2\pi/3$. These events are more frequent when comparing charged-jets to calorimeter jets than when comparing charged-jets to charged-particle jets. In any case, events with mismatched leading charged-jet and leading calorimeter jet occur with a frequency of less than one per-mille for events satisfying HLT_Jet50.

The transverse momentum response, i. e. the most probable ratio between leading charged-jet p_T and leading calorimeter jet p_T, is independent of p_T(charged-jet) within uncertainties that as before refer to the widths of Gaussian fits to the profile distributions along the p_T(charged-jet) axis. On average, the leading charged-jet carries 60% of the leading calorimeter jet p_T. This is consistent with the naive assumption, that only two thirds of all produced particles (the ones with charge ± 1) are captured by the charged-jet, whereas calorimeter jets include the entire particle flow emerging into their cone.

6.2.3 Charged-Jet Properties

We investigate further properties of charged-jets. The typical shape of the leading charged-jet can be inferred from Fig. 6.9. Average track multiplicity (upper row) and scalar p_T-sum (lower row) versus the radial distance between jet constituent and jet direction are both shown for six high-level trigger streams. Predictions from Herwig++ (closed symbols) and Pythia (open symbols) are depicted. In particular for the high-threshold streams (HLT_Jet80, HLT_Jet110) the main track flow is contained within 0.2 radians distance to the leading charged-jet. For high-threshold trigger streams, jets predicted by Herwig++ are slightly broader than jets predicted by Pythia.

Next, the jet shape dependence on the leading charged-jet transverse momentum is investigated. Fig. 6.10 displays the average radius of a cone around the charged-jet direction in (η, ϕ) space, such that the cone contains 80% of the jet particle flow. Predictions from Herwig++ (closed circles) and Pythia (open circles) are shown. 80% of all jet constituents are contained within 0.2 to 0.3 radians with respect to the charged-jet direction (left panel). The scalar track p_T sum reaches 80% of the total sum at radii of 0.2 rad for soft jets p_T(charged-jet) < 50 GeV/c and 0.1 rad for hard jets (right panel). There is thus no danger that constituents from the leading charged-jet leak into the region transverse to the leading jet in azimuth: This region contains tracks hardly depending on the leading charged-jet. Both distributions suggest that in the soft region, jets predicted by Pythia are more collimated than jets predicted by Herwig, consistent with Figures 6.9 and 6.9. Both predictions agree within statistical limits above p_T(jet) $= 100$ GeV/c. However, for lower p_T, reconstruction-level quantities systematically undershoot hadron-level predictions by approximately 10%. This behavior alludes to wide-angle emissions being less well reconstructed or less likely clustered to the charged-jet at reconstruction level or both. It could also be due to a trigger bias, since the calorimeter triggers preferentially on jets with few high-energetic particles and thus biases the jet shape. As a consequence, the jet would be more collimated at reconstruction level than it would be at hadron level.

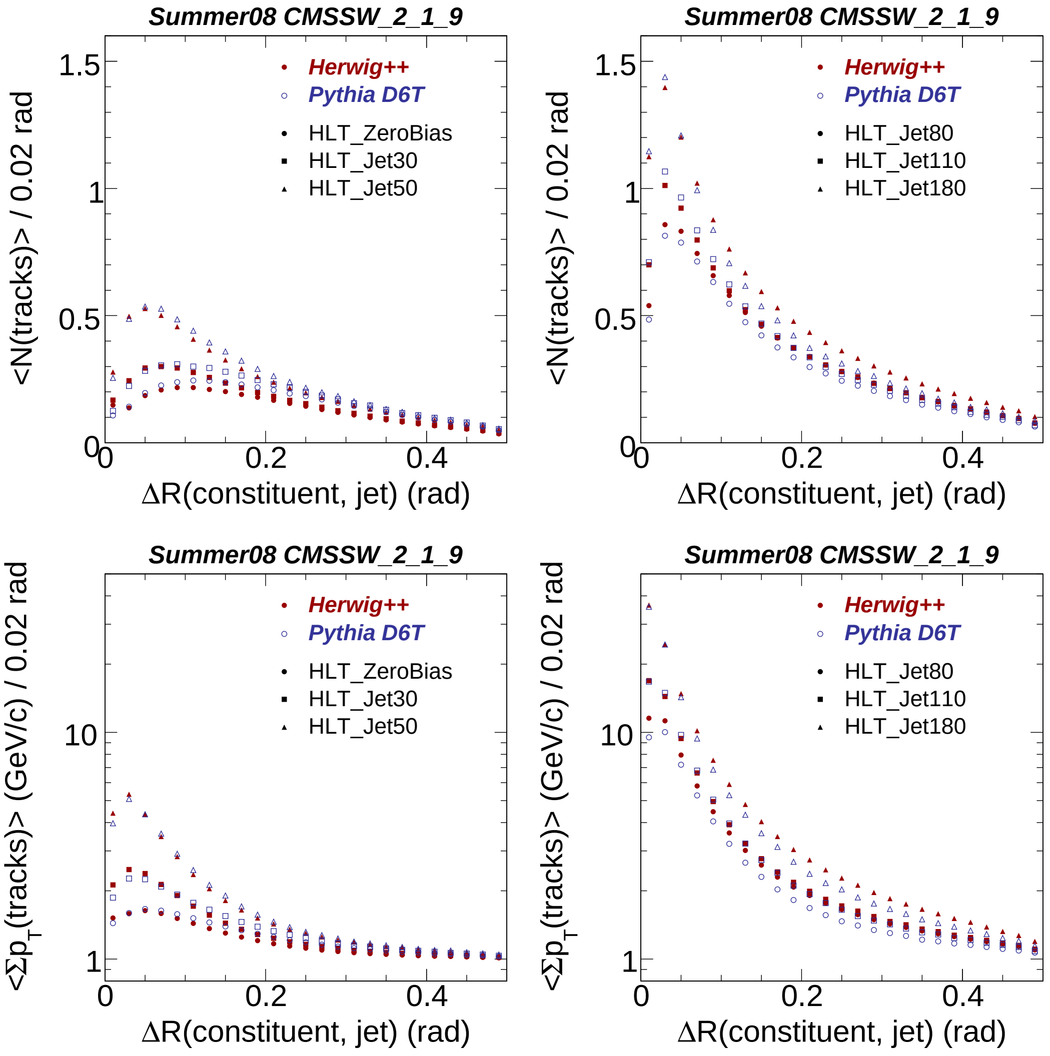


Figure 6.9: Summer08: Leading charged-jet shapes for low-threshold (left column) and high-threshold (right column) triggers. Upper row: Average track multiplicity in the radial distance R with respect to the jet direction. Lower row: Average scalar track p_T sum in the radial distance R with respect to the jet direction. High-purity tracks with $p_T > 0.9$ GeV/c and $|\eta| < 2.5$ have been selected.

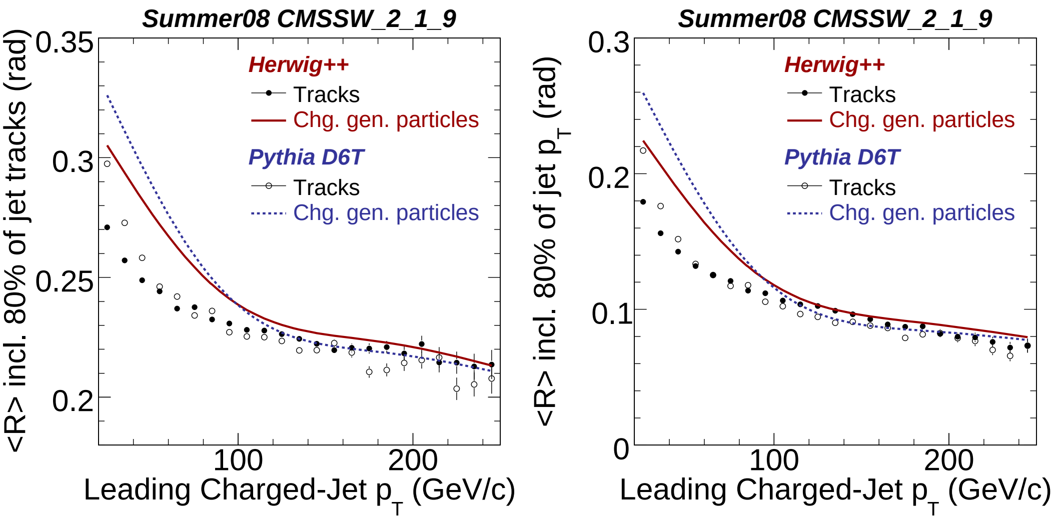


Figure 6.10: SUMMER08: Average radius of cone around the charged-jet direction containing 80% of the charged-jet's constituents (left) or 80% of the constituents' scalar p_T-sum (right) shown as a function of the leading charged-jet p_T.

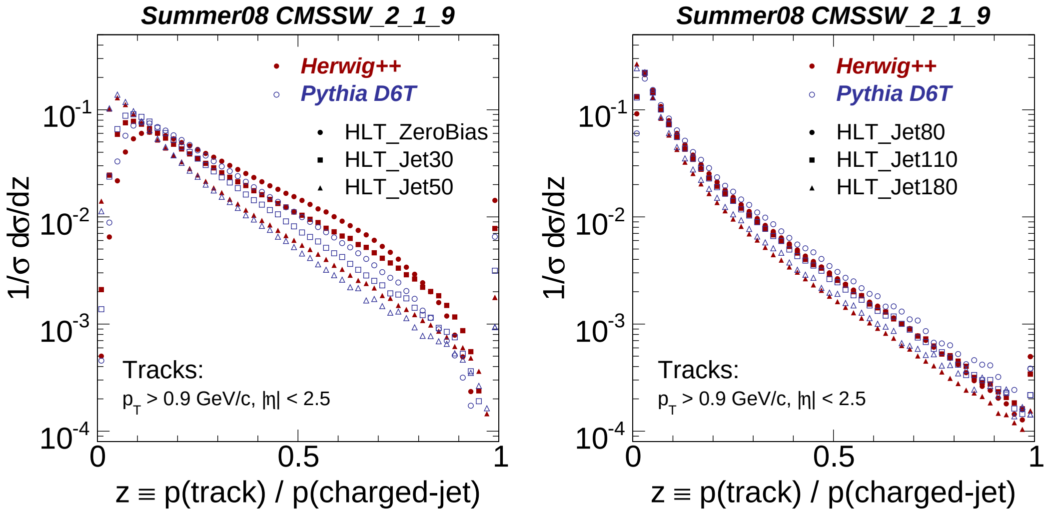


Figure 6.11: SUMMER08 Charged-jet momentum fragmentation: Normalized distributions of the momentum fractions that leading charged-jet constituents carry with respect to the leading charged-jet. Distributions are shown for low-threshold triggers (left) and high-threshold triggers (right).

We now turn to the question of how the jet momentum is shared between its constituents. In particular, we study the charged-jet's momentum fragmentation. Fig. 6.11 shows the leading charged-jet constituents' momentum fractions for the high-level trigger streams under study: For each jet constituent, the ratio z between constituent momentum and jet momentum is filled in the histogram. Distributions are normalized to the cross section, i. e. only shapes are compared in Fig. 6.11. Predictions from HERWIG++ (closed circles) and PYTHIA (open circles) are shown. Note the effect of the lower track p_T threshold on the distribution of the fragmentation variable z: For small z, the track p_T threshold causes the distribution to break down, as is especially apparent for the low-threshold trigger streams. Jets consisting of only one track are represented by the $z = 1$-bin. The relative frequency of one-track-only charged-jets is highest for the low-threshold-streams, but also in the high-threshold streams, one-track-only jets can be found. Overall, the jet fragmentation spectrum falls off steeply with increasing momentum fraction. Tracks are most likely to carry only a small fraction of the overall jet momentum. It is interesting to note that the model differences change sign when going from low-threshold triggers, HLT_ZeroBias, HLT_Jet30, HLT_Jet50 (left panel), to high-threshold triggers, HLT_Jet80, HLT_Jet110, HLT_Jet180 (right panel): HERWIG++ predicts soft jets (i. e. leading jets in the low-threshold streams) with harder fragmentation than soft jets predicted by PYTHIA. The situation is reversed for hard jets (i. e. leading jets in the high-threshold streams) where PYTHIA predicts a harder fragmentation than HERWIG++ does. Both models thus predict different charged-particle compositions of jets.

6.2.4 Conclusions

We have studied charged-jets, i. e. clusters of reconstructed tracks. Apart from their excellent reconstruction performance, charged-jets are strongly correlated with the hard interaction. Charged-jets are thus well-suited for physics studies, especially to study the underlying event in a track-based measurement. Furthermore, charged-jet constituents are very well contained in the *toward* region (cf. Sec. 5.1), thus hardly any contamination of the transverse region with constituents from the leading charged-jet is to be expected. This is different to calorimeter measurements where the large magnetic field leads to curling tracks and calorimeter energies spread on a larger part of (η, ϕ) space. To conclude, charged-jets are expected to be understood well enough to be used in a measurement of the underlying event in proton-proton collisions. We will describe such a measurement in the next chapter.

Chapter 7

Studying the Underlying Event with Charged-Jet Topologies

We now have all necessary ingredients at hand to study the underlying event in charged-jet topologies. Phenomenological models to simulate multiple parton interactions in proton-proton collisions are described in Chapter 2. In Chapter 5, we discuss the relevance of multiple interactions to describe the underlying event. The key component to measure the underlying event in charged-jet topologies is the CMS central tracking system (Sec. 3.2.2). Simulated data samples produced in CMS-wide computing exercises are described in Chapter 4. Track selection cuts that suppress fake tracks but maintain the tracking efficiency are suggested in Section 6.1. At last, in Section 6.2, we study characteristics of jets that are clustered from tracks, the *charged-jets*.

This chapter completes the preparation for an early measurement of the underlying event at the LHC by discussing experimental issues. These may arise when measuring the observables that have been introduced in Sec. 5.1:

- The impact of misaligned tracking detectors is studied with PYTHIA[1] samples produced during CSA08, scenarios S43 and S156 (Sec. 4.1).
- Different underlying event models may result in different experimental uncertainties, for instance if one model predicts more “difficult-to-reconstruct” particles than another model does. Model differences after full detector simulation and event reconstruction are explored with HERWIG and PYTHIA samples from the SUMMER08 effort (Sec. 4.2).

---

[1] In the following HERWIG refers to HERWIG++ version 2.2.0 with its default underlying event parametrization; PYTHIA refers to PYTHIA version 6.4 with its underlying event tune D6T. For subversion differences between CSA08 and SUMMER08 refer to Chap. 4.

Based on the observations made in Chap. 5, it is planned to measure the following observables[2] which are sensitive to the underlying event activity in proton-proton collisions:

1. Average track transverse momentum as a function of the track multiplicity for events accepted by a zero-bias high-level trigger (Figures 7.1 and 7.2) [76],
2. Transverse momenta of tracks emerging into the region *transverse* to the leading charged-jet (Figures 7.3 and 7.6) [66, 74],
3. Average track multiplicity and scalar p_T sum in the transverse region as a function of the leading jet p_T (Figures 7.4 and 7.7) [66, 74],
4. Density ratios for different track transverse momentum thresholds in the transverse region as a function of the leading charged-jet p_T (Fig. 7.10) [134],
5. Average event-by-event median of the ratio of jet transverse momenta and jet areas as a function of the leading charged-jet p_T (Fig. 7.15). The jet area concept [101, 102] is summarized in Section 5.1.

Hadron-level studies of variables (3) and (5) with predictions from PYTHIA 8.108 have been presented in Section 5.1, where the impact of multiple interaction model parameters was also discussed.

7.1 High-Level Triggers and Physics Object Selection

The analyzed Monte Carlo samples have been simulated during two CMS-wide production efforts: CSA08 (Tab. 4.2) and SUMMER08 (Tab. 4.6). Events are selected by one zero-bias and five jet triggers asking for a calorimeter jet with p_T above 30 GeV/c, 50 GeV/c, 80 GeV/c, 110 GeV/c, and 180 GeV/c (Tables 4.4 and 4.8).

Unless otherwise specified, physics object selections are given by Table 7.1: Tracks must be classified as *high purity* (Sec. 6.1.1) and are selected in the central region ($|\eta| < 2$) with transverse momenta above 0.9 GeV/c. Charged-jets are obtained in the following way: High-purity tracks in the full tracker acceptance ($|\eta| < 2.5$) with transverse momenta above 0.9 GeV/c are clustered by the SISCone jet algorithm with parameter $R = 0.5$ and longitudinal energy recombination scheme (Sec. 3.2.7). Charged-jets must be contained in the region $|\eta| < 2$. Out of these, the jet with largest transverse

---

[2] Further variables of interest in analyzing early LHC runs are discussed elsewhere [133].

Table 7.1: Physics object selections to study the underlying event in charged-jet topologies. Track quality selections have been discussed in Sec. 6.1.1.

Tracks	Track quality	“high purity”
	Transverse momentum	$p_T > 0.9$ GeV/c
	Pseudorapidity	$\lvert\eta\rvert < 2$
Tracks for Chg.-Jet	Track quality	“high purity”
	Transverse momentum	$p_T > 0.9$ GeV/c
	Pseudorapidity	$\lvert\eta\rvert < 2.5$
Charged-Jets	Jet Algorithm	SISCone ($R = 0.5$, long. E recomb.)
	Input objects	Tracks for Chg.-Jet
	Pseudorapidity	$\lvert\eta\rvert < 2$

momentum is referred to as the “leading” charged-jet (Sec. 6.2). If variables are presented as a function of the leading charged-jet p_T, spectra will be composed piecewise following the procedure described in Section 6.2.1 (Tab. 6.3).

7.2 Track Kinematics in Zero-Bias Events: $\langle p_T\rangle$ (n_{chg})

Typical observables to study the underlying event in track topologies are functions of the track multiplicity, of the tracks’ transverse momenta or of both. We will thus start by investigating the distribution of the average track p_T as a function of the track multiplicity: First by comparing different alignment scenarios (Sec. 7.2.1), next by comparing detector-level predictions from Herwig and Pythia (Sec. 7.2.2). A recent CDF measurement found that the average track p_T in $p\bar{p}$ collisions rises with the track multiplicity [76]. An attempt to explain this phenomenon is made by models that include color correlations. According to these models, a large number of color connections in final states with many particles creates additional transverse momentum in the event, as forward partons are “pulled” into the detector acceptance [135]. Other models allow individual partons to scatter more than once (*rescattering*, Sec 2.2).

7.2.1 Csa08 Zero-Bias: Impact of Alignment Uncertainties

Figure 7.1 shows Csa08 hadron-level and uncorrected reconstruction-level predictions from Pythia. The average track p_T in events that have been accepted by HLTZeroBias

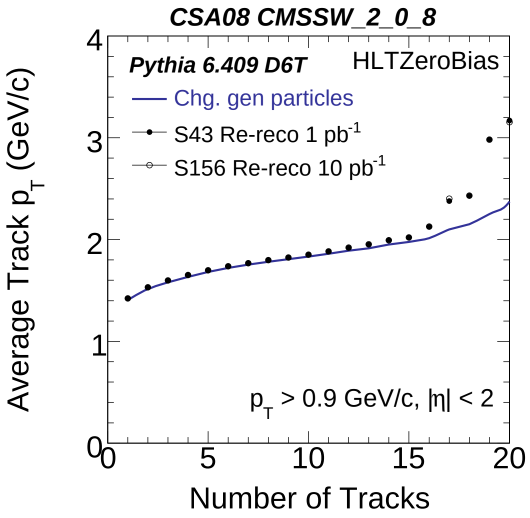


Figure 7.1: CSA08 HLTZeroBias: Average track transverse momentum as a function of track multiplicity shown for those events that have been accepted by HLTZeroBias. Hadron-level (solid line) and reconstruction-level (hollow/filled symbols) predictions shown.

is depicted as a function of the track multiplicity. Reconstruction-level predictions for alignment scenarios S43 and S156 are compared with each other. These scenarios simulate alignment conditions after having collected $\int \mathcal{L}dt = 1$ pb^{-1} and $\int \mathcal{L}dt = 10$ pb^{-1} of data (Sec. 4.1). PYTHIA predicts the average track transverse momentum to rise with the track multiplicity. As reconstruction-level predictions with S43 and S156 do not differ, the impact of misalignment on the observable measurement is expected to be small. Nonetheless, efficiency corrections will be needed to compare hadron-level predictions with track measurements. As has been shown in Sec. 6.1.2, low-p_T tracks are not measured as efficiently as high-p_T tracks (Fig. 6.2). This results in less tracks being measured and, at the same time, in a smaller scalar sum of track transverse momenta. For events with 15 particles and more, the measured average p_T typically exceeds the predicted average p_T, reflecting the smaller reconstruction efficiency for low-p_T tracks.

7.2.2 SUMMER08 Zero-Bias: Underlying-Event Model Predictions

Figure 7.2 shows SUMMER08 hadron-level and uncorrected reconstruction-level predictions from PYTHIA (hollow symbols) and HERWIG (filled symbols). The average track p_T in events that have been accepted by HLT_ZeroBias is shown as a function of the track multiplicity. Ideal detector conditions, in particular no misalignment, were assumed

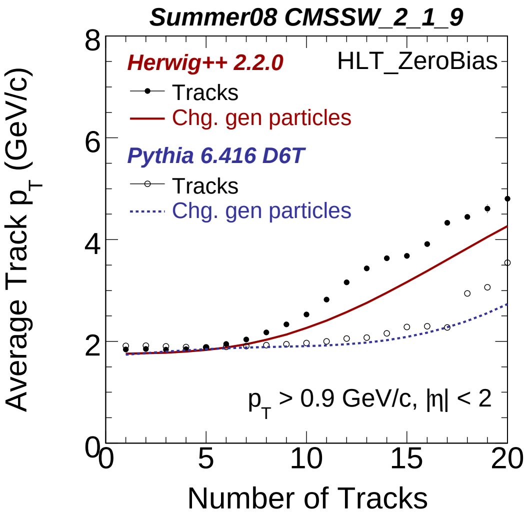


Figure 7.2: Summer08 HLT_ZeroBias: Average track transverse momentum as a function of track multiplicity shown for those events that have been accepted by HLT_ZeroBias. Hadron-level (solid/dashed lines) and reconstruction-level (hollow/filled symbols) predictions shown.

during simulation (Sec. 4.2). Note that the Summer08 samples under study do not include a minimum-bias sample (Sec. 4.2). This is reflected in the average track p_T in events with less than five tracks. Comparing Figure 7.1 with Figure 7.2, the Pythia predictions differ for low multiplicities: In Csa08 (where a minimum-bias sample has been included), the average track p_T rises even when going from one-track events to two-track events; yet, the distribution remains approximately constant up to seven to eight tracks in Summer08. Since all Summer08 samples under investigation have interaction scales above 15 GeV/c, asking for few tracks biases towards higher average transverse momenta.

The general trend for the average track p_T to rise with the number of tracks is predicted by both models. However, Herwig simulates a stronger rise of the average track p_T with the number of tracks than Pythia does. In events with more than seven tracks, Herwig's prediction on the average track p_T is larger than the prediction by Pythia. This observation is consistent with Herwig's prediction of the fragmentation spectrum of charged-jets (Fig. 6.11): Soft charged-jets from Herwig exhibit a harder fragmentation than soft charged-jets from Pythia, i. e. Herwig predicts charged-jet constituents that carry larger momentum fractions than the constituents predicted by Pythia. As in Figure 7.1, reconstruction-level predictions overshoot hadron-level predictions for a given number of tracks.

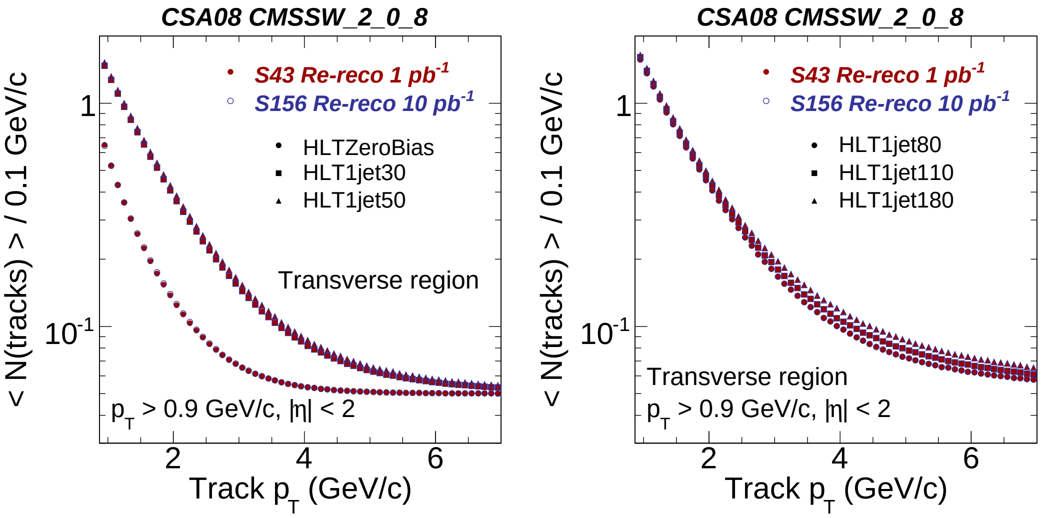


Figure 7.3: CSA08, transverse region in charged-jet topologies: Average track multiplicity in 0.1 GeV/c bins of track p_T for events accepted by low-threshold (left panel) and high-threshold (right panel) high-level triggers. Reconstruction-level predictions shown for alignment scenarios S43 (filled symbols) and S156 (hollow symbols).

7.3 Underlying-Event Studies with Tracks in the Transverse Region of Charged-Jet Events

The strategy to study the underlying event in the region transverse to the leading charged-jet (Fig. 5.1) has been described in Section 5.1. The transverse region is particularly sensitive to non-perturbative effects and event generator predictions must be tuned to data. We study the impact of the two example alignment scenarios from CSA08 (Sec. 7.3.1) on observables for the transverse-region as well as reconstruction-level predictions from HERWIG and PYTHIA from SUMMER08 (Sec. 7.3.2).

7.3.1 CSA08: Impact of Alignment Uncertainties

We start by investigating the track p_T distribution in the transverse region. Figure 7.3 shows CSA08 reconstruction-level predictions from PYTHIA, obtained with alignment conditions S43 (filled symbols) and S156 (hollow symbols). Distributions are depicted for low-threshold (left panel) and high-threshold (right panel) trigger streams. Each point corresponds to the average number of tracks per event in a 0.1 GeV/c bin of

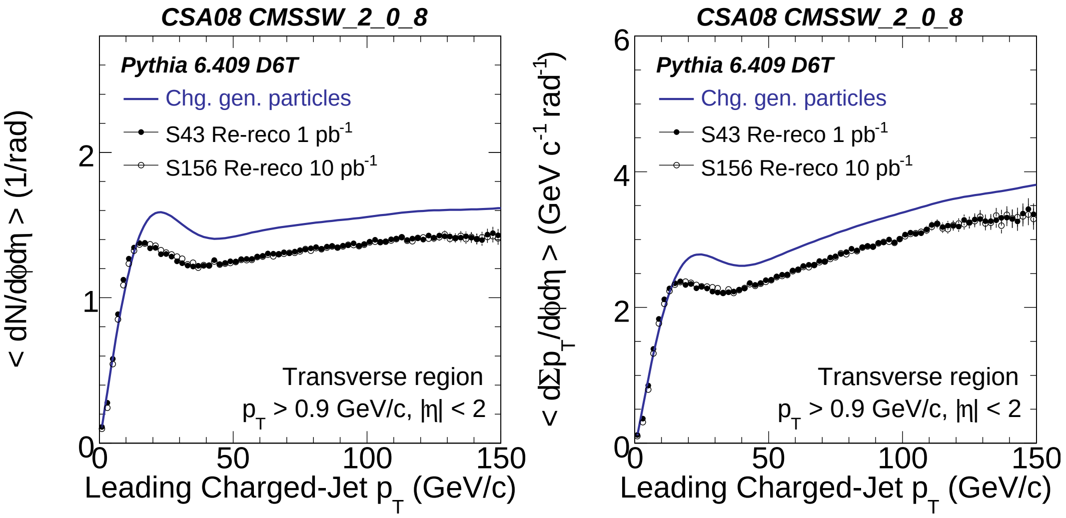


Figure 7.4: CSA08, transverse region in charged-jet topologies. Left: Average track multiplicity as a function of leading charged-jet p_T. Right: Average track scalar p_T sum as a function of leading charged-jet p_T. Hadron-level (solid line) distributions shown together with reconstruction-level distributions for alignment scenarios S43 (filled symbols) and S156 (hollow symbols).

p_T(track). The integral of each distribution is equal to the average number of tracks in the transverse region for a specific high-level trigger stream.

The p_T distribution of tracks in the transverse region becomes harder for increasing high-level trigger thresholds: High-p_T tracks in the transverse region are more likely if a hard interaction has occurred. Multiple parton interactions explain this behavior in a straightforward manner: Soft pp scatterings correspond to peripheral collisions with smaller spatial overlap of the proton wavefunctions and thus to a decreased likelihood for additional interactions to occur. Hard interactions correspond to more central collisions where the spatial parts of the proton wavefunctions have a significant overlap thus enhancing the probability for additional parton-parton interactions. The negligible difference between S43 and S156 suggests that the impact of alignment on the observable is small.

We have seen that typical track transverse momenta increase with increasing trigger thresholds (Fig. 7.3). We thus also expect the track multiplicity and the track p_T sum in the transverse region to grow with increasing event scales, where we represent the event scale by the leading charged-jet transverse momentum. Hadron-level charged-particle activities have been studied as a function of the leading charged-jet p_T in Section 5.1. Figure 7.4 displays CSA08 reconstruction-level predictions on the average track multi-

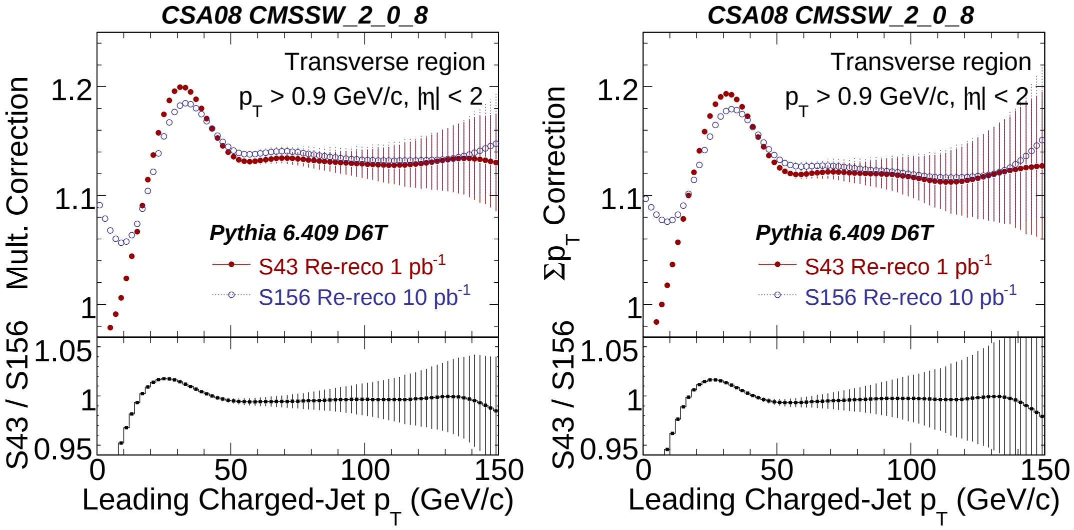


Figure 7.5: Csa08, transverse region in charged-jet topologies. Left: Correction factors for average track multiplicity as a function of leading charged-jet p_T (Eq. 7.2). Right: Correction factors for average track scalar p_T sum as a function of leading charged-jet p_T. Predictions shown for alignment scenarios S43 (filled symbols) and S156 (hollow symbols).

plicity (left panel) and on the average scalar track p_T sum (right panel) as a function of the leading charged-jet p_T, obtained with alignment scenarios S43 (filled symbols) and S156 (hollow symbols) together with hadron-level predictions (solid line). The distributions are composed following the procedure outlined in Sec. 6.2.1. Both activities exhibit a similar behavior: The activity increases with increasing jet p_T and then saturates, indicating the event scale at which the protons' spatial wave-functions fully overlap. The reconstructed track densities are not corrected for detector inefficiencies and tend to undershoot the generator predictions on charged particle densities. A possible correction procedure is addressed in the next paragraph.

Csa08: Impact of Misalignment on Track-Particle Correction

The underlying event activity as a function of the leading charged-jet p_T will be used to tune multiple interaction model parameters. This entails either that track quantities must be corrected to particle level, or that model predictions are processed by the full detector simulation and event reconstruction chain. Both approaches are possible; yet, we will only describe the first. In general, the number of reconstructed tracks is a sum of the number of successfully reconstructed charged particles and of the number of fake tracks. A similar statement holds for the scalar p_T sum:

$$\begin{aligned} N(\text{meas. tracks}) &= N(\text{rec. charged particles}) + N(\text{fake tracks}) \ , \\ \sum p_T(\text{meas. tracks}) &= \sum p_T(\text{rec. charged particles}) + \sum p_T(\text{fake tracks}) \ . \end{aligned} \tag{7.1}$$

The suggested bin-by-bin procedure corrects reconstruction-level observables to hadron-level observables in one go. Correction factors are calculated simply by taking the ratio of prediction and measurement in each bin of p_T(charged-jet), i. e. one divides a given model's hadron-level predictions by its reconstruction-level predictions (Fig. 7.4):

$$\begin{aligned} \text{Mult. Correction} &\equiv \frac{N(\text{chg. gen. particles})}{N(\text{tracks})} \ , \\ \sum p_T \ \text{Correction} &\equiv \frac{\sum p_T(\text{chg. gen. particles})}{\sum p_T(\text{tracks})} \ . \end{aligned} \tag{7.2}$$

The detailed determination of correction factors is as follows: We construct hadron-level predictions of a given observable. The obtained distribution is subject to a smoothing procedure [136]. After having passed the hadron-level prediction through the CMS detector simulation and event reconstruction, we construct reconstruction-level quantities. The ratio of hadron-level and reconstruction-level predictions for each bin of p_T(jet) is smoothed with the same procedure as above.

Figure 7.5 displays correction factors for track multiplicity (left panel) and scalar track p_T sum (right panel) as a function of the leading charged-jet p_T. Corrections have been obtained from CSA08 PYTHIA samples (Sec. 4.1). Ratios between corrections from alignment scenarios S43 and S156 are shown in the lower part of each panel. For p_T(charged-jet) > 50 GeV/c, both corrections agree within the available statistics. Typical correction factors are of the order of 1.12, meaning that bin entries of reconstruction-level predictions in Figure 7.4 typically need to be multiplied by a factor of 1.12 in order to account for imperfect reconstruction efficiencies. This is consistent with the overall reconstruction efficiencies for tracks with $p_T > 0.9$ GeV/c and $|\eta| < 2$ (Fig. 6.2).

7.3.2 SUMMER08: Reconstruction-Level Predictions

We now turn to comparing model predictions on the track p_T distribution in the transverse region. SUMMER08 reconstruction-level predictions from HERWIG (filled symbols) and PYTHIA (hollow symbols) are shown in Figure 7.6 for low-threshold (left panel)

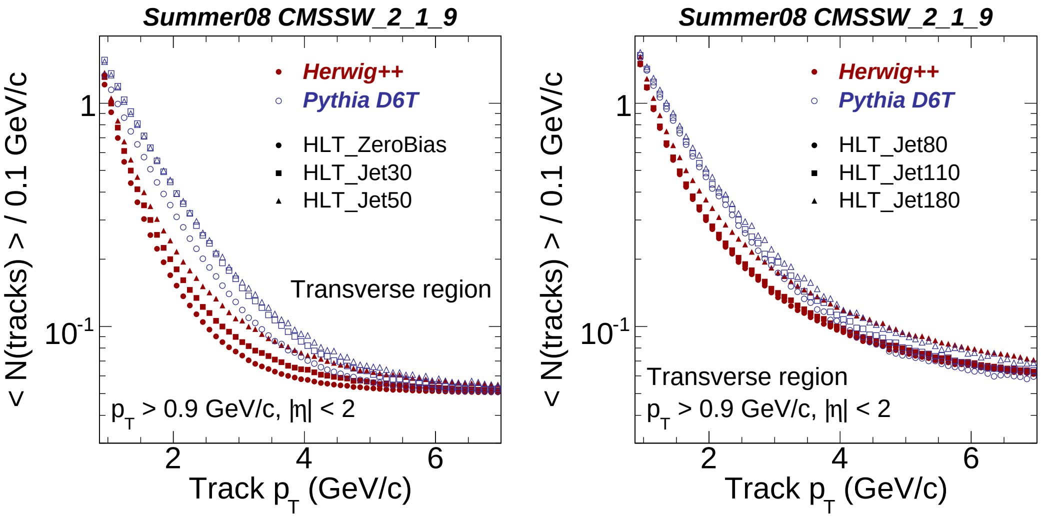


Figure 7.6: Summer08, transverse region in charged-jet topologies: Average track multiplicity in 0.1 GeV/c bins of track p_T for events accepted by low-threshold (left panel) and high-threshold (right panel) high-level triggers. Reconstruction-level predictions shown for underlying event tunes Herwig++ (filled symbols) and Pythia D6T (hollow symbols).

and high-threshold (right panel) trigger streams. Both models predict different track p_T distributions in the transverse region. For 1 GeV/c < p_T(track) < 5 GeV/c, Pythia predicts on average more tracks than Herwig does. Above p_T(track) = 5 GeV/c, model differences are less pronounced.

As discussed in Sec. 7.2, the Summer08 samples under study do not include a simulation of minimum-bias events. The resulting difference is particularly obvious when comparing the HLT_ZeroBias track p_T distribution in Figure 7.6-left to the corresponding distribution in Figure 7.3-left: The HLT_ZeroBias stream of the Summer08 event set contains on average more tracks with larger transverse momenta.

The underlying event activity as a function of the leading charged-jet transverse momentum, as obtained from Summer08 samples, is shown in Figure 7.7. Predictions from Pythia (hollow symbols, dashed line) and Herwig (filled symbols, solid line) on the track multiplicity (left panel) and scalar track p_T sum (right panel) are shown at both hadron level and at track level. Note that due to the smaller statistics available in the Summer08 samples, bin sizes are five times larger than bin sizes in the corresponding distributions (Fig. 7.4). Transverse region activities predicted by Pythia are significantly larger than predictions from Herwig. The tendency for Pythia to predict more tracks in the transverse region than Herwig does is consistent with the

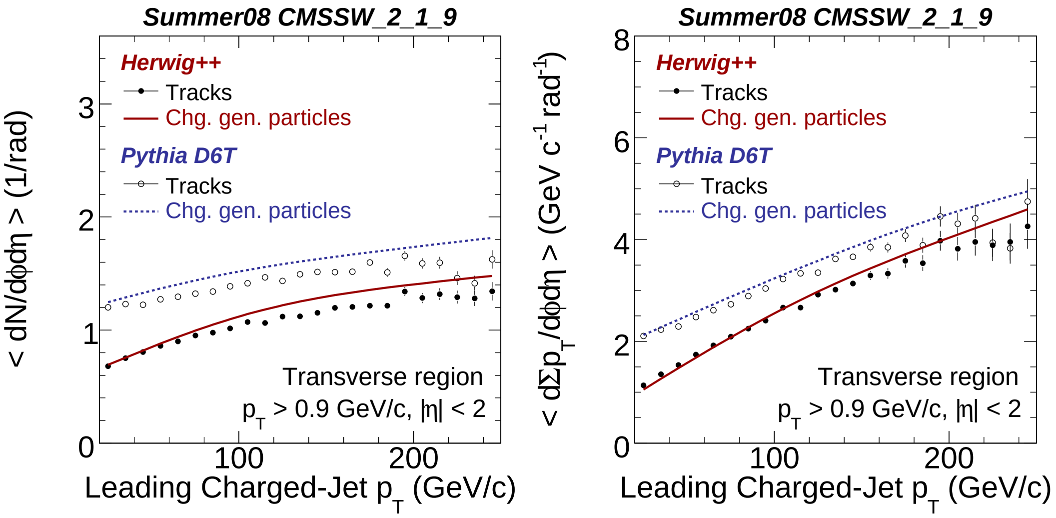


Figure 7.7: Summer08, transverse region in charged-jet topologies. Left: Average track multiplicity as a function of leading charged-jet p_T. Right: Average track scalar p_T sum as a function of leading charged-jet p_T. Hadron-level distributions shown together with reconstruction-level distributions for underlying event tunes Herwig++ and Pythia D6T.

predicted p_T distribution of tracks in the transverse region (Fig. 7.6). As discussed before, reconstruction-level observables must be corrected to account for inefficiencies in track reconstruction. A bin-by-bin correction, as explained in the previous section, is extracted from the Summer08 distributions. The resulting correction factors are shown in the subsequent section. Nonetheless, it is already apparent from Figure 7.7 that differences between reconstruction-level and hadron-level predictions from one model are smaller than differences between reconstruction-level predictions from Pythia and Herwig. Thus, already with a few pb^{-1} of collision data, first model tunes can be derived.

Summer08: Model Discrimination with MC-corrected Densities

Bin-by-bin correction factors for track multiplicities (left panel) and track scalar p_T sum (right panel) are displayed in Figure 7.8. Corrections have been determined following the procedure outlined in Sec. 7.3.1 using the predictions from Herwig and Pythia discussed in the previous section (Fig. 7.7). Strictly speaking, the determination of corrections for both models should be redundant, since reconstruction algorithms and detector conditions are the same when simulating the Pythia and Herwig samples. Thus, exactly equal detector performances are expected for both Monte Carlo samples.

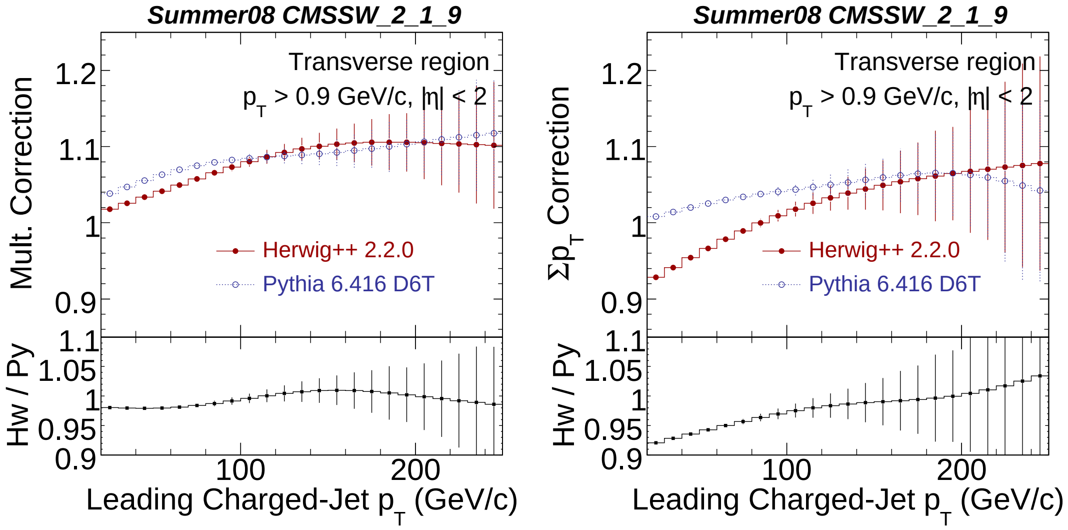


Figure 7.8: SUMMER08, transverse region in charged-jet topologies. Left: Correction factors for average track multiplicity as a function of leading charged-jet p_T (Eq. 7.2). Right: Correction factors for average track scalar p_T sum as a function of leading charged-jet p_T. Predictions shown for underlying event tunes HERWIG++ (filled symbols) and PYTHIA D6T (hollow symbols).

Yet, since the tracking efficiency is not uniform as a function of the track transverse momentum until $p_T \approx 3$ GeV/c (Fig. 6.2-right), different bin-by-bin correction factors are possible. As shown in Figures 7.2 and 7.6, both models predict different rates of soft tracks in the transverse region. It should be pointed out that a more sophisticated correction procedure should take non-uniform tracking efficiencies into account accordingly.

For p_T(charged-jet) > 100 GeV/c, both corrections agree within the available statistics. Ratios between the obtained corrections from HERWIG and PYTHIA are shown in the lower part of each panel. For p_T(charged-jet) < 100 GeV/c, HERWIG predicts smaller multiplicity corrections than PYTHIA. In this region, the $\sum p_T$ correction predicted by HERWIG falls below unity, i. e. the predicted transverse region track p_T sum is larger than the predicted transverse region charged-particle p_T sum. PYTHIA's correction predictions on $\sum p_T$ remain above unity. These model differences are an effect of the different track p_T predictions as discussed above. As can be seen in Figure 7.2, the average track p_T predicted by HERWIG is larger than the average track p_T predicted by PYTHIA. As low-p_T tracks are reconstructed less efficiently than high-p_T tracks, larger corrections must be applied to the PYTHIA samples. In other words, the track-particle correction factors obtained with PYTHIA are larger than the ones obtained with HERWIG. In the following, we will study an approach to absorb overall tracking inefficiencies by making use of ratio observables.

SUMMER08: **Model Discrimination with Density Ratios**

An approach that is assumed to suppress the impact of imperfect track reconstruction has been put forward recently [134]: Underlying-event activity predictions for the LHC with respect to the leading charged-jet p_T have been studied for different thresholds on the track transverse momenta, namely for 900 MeV/c and for 1.5 GeV/c. The specific choice of thresholds can be optimized in a later study. Note that higher p_T thresholds come at the price of a severe loss in statistics, as can be seen in Figure 7.6. Clearly, the underlying event activity decreases when considering less tracks per event (since the track threshold has been increased). Thus the ratio between quantities reconstructed from tracks with $p_T > 900$ MeV/c and quantities reconstructed from tracks with $p_T > 1.5$ GeV/c will be above one and will reflect the falloff of the p_T distribution of tracks in the transverse region (Fig. 7.6):

$$\begin{aligned}\text{Multiplicity Ratio} &\equiv \frac{\langle N(\text{tracks with } p_T > 0.9 \text{ GeV}/c)\rangle}{\langle N(\text{tracks with } p_T > 1.5 \text{ GeV}/c)\rangle} \ , \\ \sum p_T \text{ Ratio} &\equiv \frac{\langle \sum p_T(\text{tracks with } p_T > 0.9 \text{ GeV}/c)\rangle}{\langle \sum p_T(\text{tracks with } p_T > 1.5 \text{ GeV}/c)\rangle} \ . \end{aligned} \tag{7.3}$$

At the same time, we hope to absorb possible tracking inefficiencies in the ratio: As long as the tracking performance is uniform for the region 0.9 GeV/c $< p_T(\text{track}) <$ 1.5 GeV/c, density ratio predictions from generator particle quantities and from track quantities should agree with each other. However, as can be seen from Figure 6.2-right, the track reconstruction efficiency is not quite uniform in this range and we do not expect reconstruction-level and hadron-level predictions to fully coincide.

Predictions from HERWIG (filled circles, solid line) and PYTHIA (hollow circles, dashed line) on the average track multiplicity (left panel) and on the scalar track p_T sum (right panel) as a function of the leading charged-jet p_T are shown in Figure 7.9. As expected, less underlying event activity is predicted if the track p_T threshold is raised. Next, ratio distributions are obtained from the distributions shown in Figure 7.7 and Figure 7.9. The resulting multiplicity ratios (left) and p_T sum ratios (right) as a function of the leading charged-jet p_T are shown in Figure 7.10. Both models predict approximately half the number of particles and 50% less scalar p_T sum when raising the track

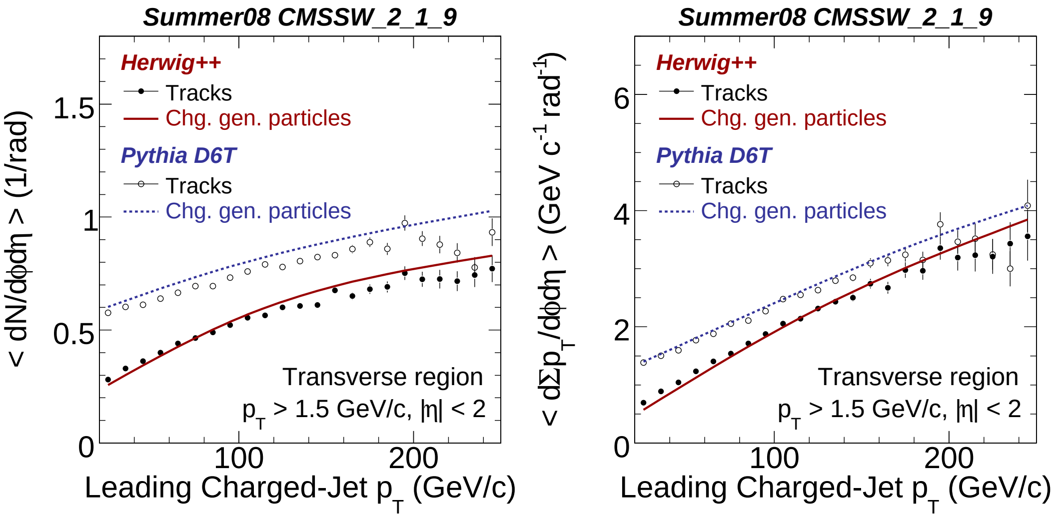


Figure 7.9: Summer08, transverse region in charged-jet topologies. Left: Average track multiplicity as a function of leading charged-jet p_T. Right: Average track scalar p_T sum as a function of leading charged-jet p_T. Hadron-level distributions shown together with reconstruction-level distributions for underlying event tunes Herwig++ and Pythia D6T. The lower threshold on track transverse momenta has been increased to $p_T > 1.5$ GeV/c with respect to predictions shown in Figure 7.7.

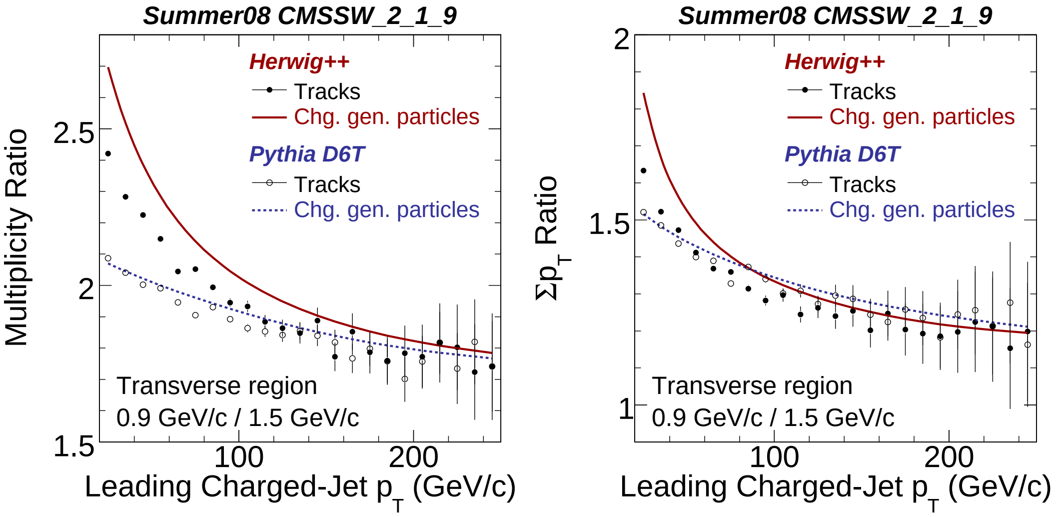


Figure 7.10: Summer08, transverse region in charged-jet topologies. Left: Multiplicity ratios for track p_T thresholds 0.9 GeV/c and 1.5 GeV/c as a function of leading charged-jet p_T (Eq. 7.3). Right: Scalar p_T sum ratios as a function of leading charged-jet p_T. Hadron-level distributions shown together with reconstruction-level distributions for underlying event tunes Herwig++ and Pythia D6T.

Table 7.2: Physics object selections to study the underlying event with jet areas. Track quality selections have been discussed in Sec. 6.1.1.

Tracks	Track quality	"high purity"
	Transverse momentum	$p_T > 0.9$ GeV/c
	Pseudorapidity	$\|\eta\| < 2$
Tracks for Chg.-Jet	Track quality	"high purity"
	Transverse momentum	$p_T > 0.9$ GeV/c
	Pseudorapidity	$\|\eta\| < 2.5$
Charged-Jets	Jet Algorithm	$k_\perp$ ($D = 0.4$, long. E recomb.)
	Input objects	Tracks for Chg.-Jet
	Pseudorapidity	$\|\eta\| < 2$

p_T threshold from 0.9 GeV/c to 1.5 GeV/c. As discussed above, reconstruction-level and hadron-level predictions do not coincide, as is especially visible for Herwig. Model differences are most pronounced for jet transverse momenta below 100 GeV/c. Above this value, Pythia and Herwig agree with each other in both density ratio distributions. More studies are necessary to judge the method's power to discriminate different underlying event models. Expected small systematic uncertainties make the measurement of density ratios an attractive approach when studying early collision data.

7.4 Studying the Underlying Event with Jet Areas

The jet area concept [101,102] is summarized in Section 5.1.2. Jet areas measure the susceptibility of a jet to contamination from a uniform, diffuse structure of particles added to the event. In the context of the underlying event in proton-proton collisions, jet areas have at least two possible applications. First, jet energy corrections for contributions from the underlying event are obtained by scaling the average underlying-event particle flow by the jet's active area. Second, jet areas qualify to study the diffuse underlying-event particle flow in multi-jet topologies. As discussed in Sec. 5.1.2, the event-by-event median of the ratio of jet p_T and jet active area A can be used to characterize the underlying-event activity. In particular, if a given jet is clustered exclusively from underlying-event particles, the jet's ratio p_T/A corresponds to the average underlying-event p_T flow per unit area for this jet.

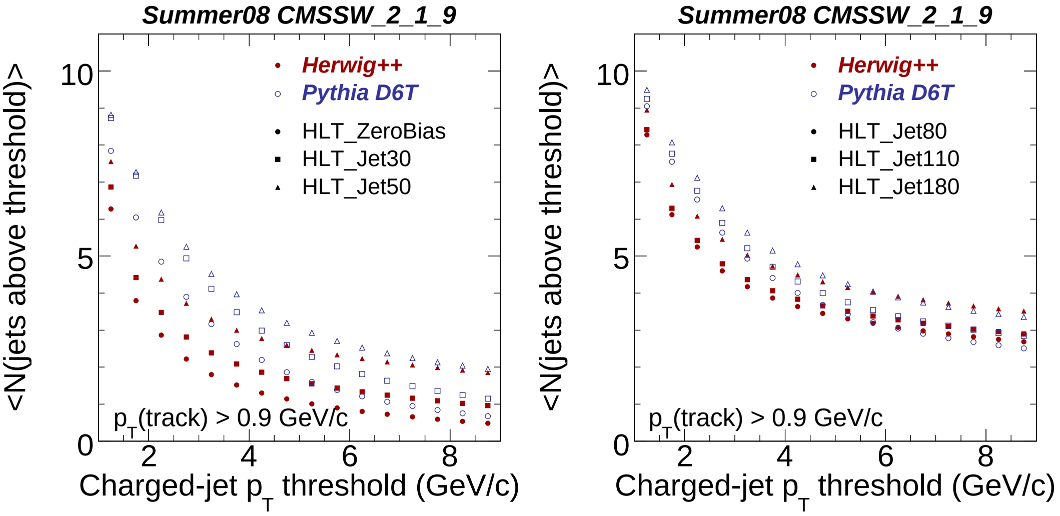


Figure 7.11: SUMMER08: Average number of charged $k_\perp$ jets as a function of the applied jet p_T threshold for events accepted by low-threshold (left panel) and high-threshold (right panel) high-level triggers. Reconstruction-level predictions shown for underlying event tunes HERWIG++ (filled symbols) and PYTHIA D6T (hollow symbols).

Active jet areas are determined for charged-jets clustered with the $k_\perp$ jet algorithm with parameter $D = 0.4$ (Tab. 7.2). This algorithm is preferred to the SISCone jet algorithm due to its faster runtime, as explained in Sec. 5.1.2. Average charged-jet multiplicities are displayed in Figure 7.11 as a function of the lower p_T threshold on the jet[3]. Predictions from HERWIG (filled symbols) and PYTHIA (hollow symbols) are shown for low-threshold (left panel) and high-threshold (right panel) trigger streams. For jet p_T thresholds of p_T(charged-jet) > 8 GeV/c and higher, the average jet multiplicity varies between zero and two. Only for the HLT_Jet180, the average jet multiplicity rises above three. Only jets from the hard scatter are hence reconstructed for these p_T thresholds. The lower the charged-jet p_T threshold, the more jets are reconstructed. Yet, the models' jet multiplicity predictions differ: PYTHIA predicts more jets than HERWIG does for any p_T threshold. For p_T(charged-jet) > 5 GeV/c, PYTHIA predicts on average approximately 1.5 charged-jets per event in the HLT_ZeroBias stream, but HERWIG predicts one charged-jet. In the HLT_Jet180 stream, both models predict on average four charged-jets with p_T(charged-jet) > 5 GeV/c. As discussed earlier, the number of jets must be large enough to ensure that the event-by-event median jet p_T divided by

---

[3] In the following, no corrections for trigger efficiencies have been applied.

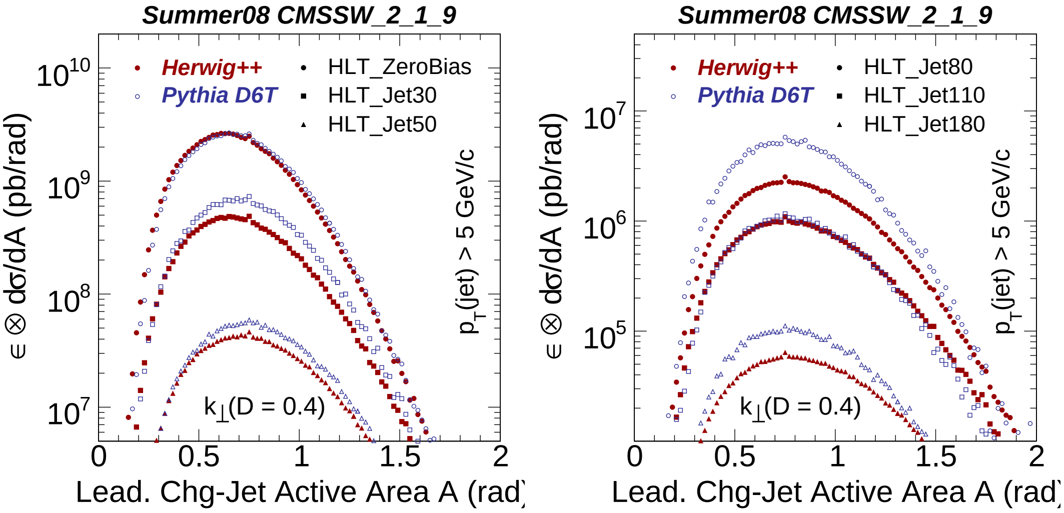


Figure 7.12: SUMMER08: Predictions on the distribution of the leading charged-jet area A from HERWIG (filled symbols) and PYTHIA (hollow symbols) for low-threshold triggers (left) and high-threshold triggers (right). Streams are weighted by cross section predictions, but not corrected for efficiency losses (indicated by the ϵ convolution). Charged $k_\perp$ jets with $p_T > 5$ GeV/c selected.

the jet's active area,

$$\rho \equiv \mu_{1/2}\left(\left\{\frac{p_{Ti}}{A_i}\right\}\right) , \tag{7.4}$$

is dominated by jets from underlying event particles. From the model predictions shown in Figure 7.11, it is apparent that the transverse momentum threshold has to be lowered down to values like $p_T > 1$ GeV/c in order to minimize bias from the hard scatter. With such a threshold, HERWIG and PYTHIA predict on average 6 and 8 charged-jets per event in the HLT_ZeroBias stream. Note that these are "jets" in the sense of the outcome of a jet algorithm. Even the leading charged-jet from time to time happens to consist of only one track (Fig. 6.11). Nonetheless, the areas of these one-track jets probe the event activity close to the track and are thus sensitive to the distribution of particle flow in the event.

Figure 7.12 compares predicted leading charged-jet active area distributions by PYTHIA (hollow symbols) with predictions by HERWIG (filled symbols). Jet areas are shown for low-threshold (left panel) and high-threshold (right panel) high-level trigger streams. Charged-jets have to pass a lower transverse momentum threshold of 5 GeV/c. Both models predict equal values for the most probable jet area irrespective of the high-level

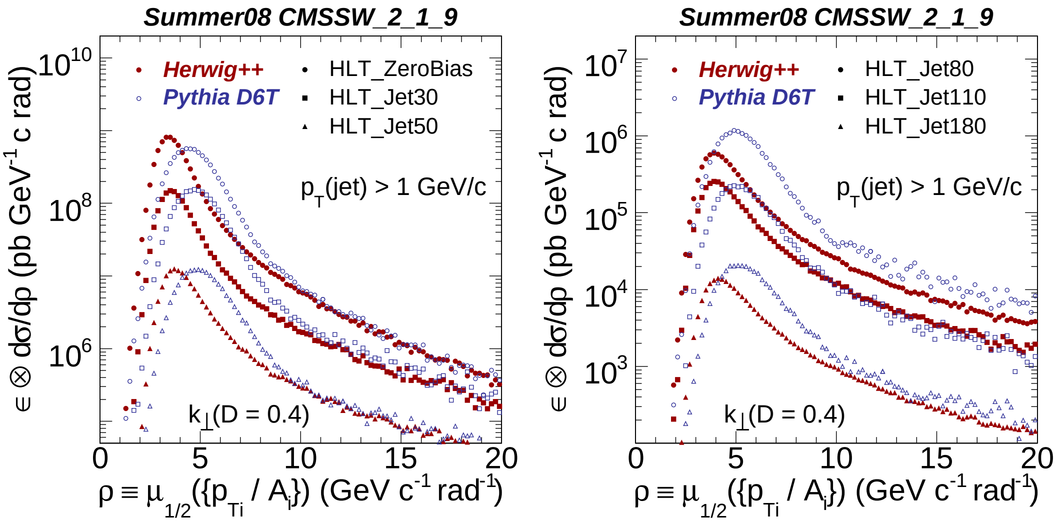


Figure 7.13: Summer08: Distribution of event-by-event median $\rho \equiv \mu_{1/2}(\{p_{Ti}/A_i\})$ built from charged-jets with transverse momenta p_{Ti} and active areas A_i. Streams are weighted by cross section predictions, but not corrected for efficiency losses (indicated by the ϵ convolution). Predictions from Herwig (filled symbols) and Pythia (hollow symbols) for six high-level trigger streams. Charged $k_\perp$ jets with $p_T > 1$ GeV/c selected.

trigger stream considered. Typical active areas of the leading charged-jet are of the order of $0.85 \approx 1.7\pi D^2$. Jet production rates differ vastly for both models under study: Pythia typically predicts more jets with $p_T > 5$ GeV/c than Herwig.

Figure 7.13 shows the distribution of the median ratio of jet p_T and jet active area A for low-threshold (left panel) and high-threshold (right panel) high-level trigger streams. Predictions are from Herwig (red filled symbols) and Pythia (blue hollow symbols). A slight shift of the most probable ρ to larger values with increasing high-level trigger threshold is observed. This is a consequence of increased jet transverse momenta and unchanged jet areas, as seen in Figure 7.12. It is interesting to note that Herwig and Pythia predict different distributions of ρ with Pythia being slightly above Herwig for all trigger streams. We interpret this behavior as Pythia predicting a larger underlying event transverse momentum flow per unit area than Herwig does - a conclusion in agreement with the observations on track multiplicity and scalar track p_T sum in the transverse region (Sec. 7.3).

The average leading charged-jet active area A as a function of the leading charged-jet p_T is shown in Figure 7.14. Predictions from Herwig (filled symbols) and Pythia (hollow symbols) are shown at hadron level and at reconstruction level. (Note the

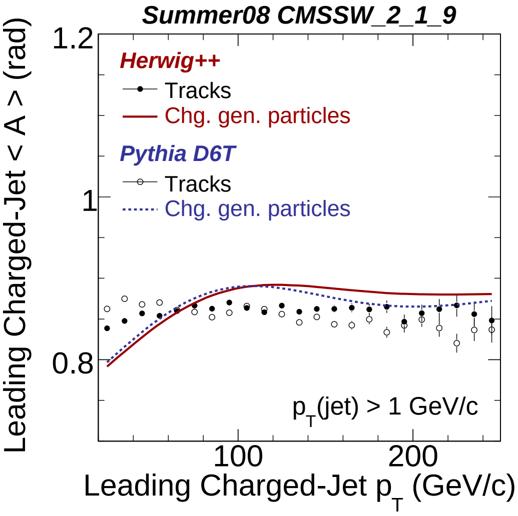


Figure 7.14: SUMMER08: Average leading charged-jet area A as a function of the leading charged-jet p_T. Predictions from HERWIG (filled symbols) and PYTHIA (hollow symbols) shown. Charged $k_\perp$ jets with $p_T > 1$ GeV/c selected.

suppressed origin.) Both model predictions agree within the available statistics. For p_T(charged-jet) > 60 GeV/c, both models predict the reconstruction-level predictions to undershoot the hadron-level predictions. Although the difference is below the 5% level, further studies of the origin of this effect are needed, because jet energy corrections for the underlying event and for pile-up typically subtract the average transverse momentum or transverse energy flow within the jet's area. Thus, a mis-measured jet area leads to a miscalibrated jet. At reconstruction level, the average leading charged-jet area shows no correlation with the leading charged-jet p_T. At hadron level, the average leading charged-particle jet area increases with leading charged-particle jet p_T in the region p_T(charged-jet) < 100 GeV/c. For higher charged-jet p_T, no correlation of $\langle A \rangle$ with p_T(charged-jet) is observed at hadron level .

We now turn to study how the median ratio of charged-jet p_T and charged-jet active area A evolves as a function of leading charged-jet p_T (Fig. 5.9). Predictions from HERWIG and PYTHIA, obtained with the SUMMER08 samples, are depicted in Figure 7.15. The origin is suppressed to exhibit model differences more clearly. The transverse momentum flow per unit area predicted by PYTHIA is larger than the one predicted by HERWIG, in agreement with Figure 7.13. If one interprets $\rho \equiv \mu_{1/2}(\{p_{Ti}/A_i\})$ as the diffuse p_T flow coming from the underlying event, PYTHIA predicts a higher underlying-event activity than HERWIG does - characterizing the underlying event with jet areas has thus lead to the same result that has been obtained with transverse-region topologies

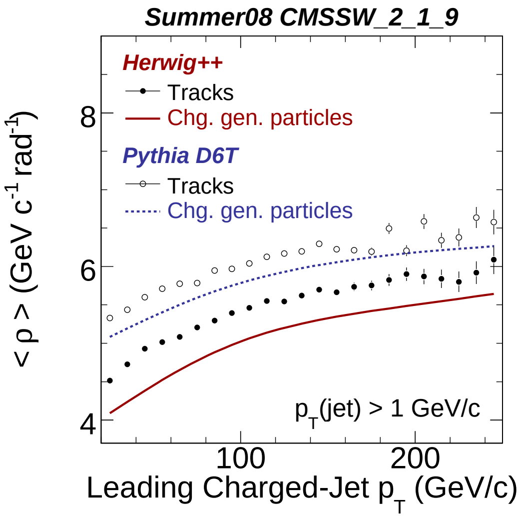


Figure 7.15: Summer08: Average event-by-event median ratio of jet p_T and jet area shown as a function of the leading charged-jet p_T. Predictions are from Herwig (filled symbols) and from Pythia (hollow symbols). Charged $k_\perp$ jets with $p_T >$ 1 GeV/c selected.

(Sec. 7.3.2). Differences between reconstruction-level and hadron-level predictions on $\langle\rho\rangle$ are of the same order than differences between predictions from Pythia and Herwig. More studies are necessary to assess the potential of jet areas to become an ingredient to underlying event tunes. In any case, the difference of predictions on the median ratio of jet p_T and jet active area is an appealing observation that justifies a measurement of jet areas in proton-proton collisions.

7.5 Conclusions

In this chapter, we have studied track-based observables that are assumed to characterize the underlying event in proton-proton collisions - either by identifying a region in (η, ϕ) space that is separated from the hard process or by studying multi-jet kinematics. Emphasis has been put on anticipating possible experimental issues when carrying out the measurement with early LHC data. Fully simulated samples have been analyzed with respect to potential influences from detector misalignment. Theoretical uncertainties have been investigated by analyzing fully simulated samples produced with the Herwig and Pythia event generator programs, which comprise different underlying event models.

With an integrated luminosity of $\int \mathcal{L}\mathrm{d}t = 1\ \mathrm{pb}^{-1}$, alignment uncertainties are under control and transverse-region activities can be measured sufficiently precise. Ratios of observables have been evaluated in the hope of absorbing tracking imperfections. Such a measurement is complementary to the measurement of absolute quantities and is expected to have small systematic uncertainties. Model uncertainties in reconstructing tracks back to particle level are small enough such that distinguishing different models will be possible by measuring multiplicity and scalar p_T sum of tracks in the transverse region.

Jet areas may help to characterize the underlying event, provided that a sufficient number of jets per event is reconstructed; the studies with charged-jets have shown that jet p_T thresholds have to be lowered to the experimental limit. The median ratio of jet p_T and jet active area probes the diffuse p_T flow coming from the underlying event. The study of this observable shows that PYTHIA predicts a larger p_T flow from the underlying event than HERWIG does. The models differ by smaller amounts than when studying the transverse region in charged-jet topologies. However, in high-multiplicity final states with many jets, such as hadronic top decays, the transverse region is undefined and jet areas may be the only method to determine the diffuse p_T flow from the underlying-event.

Chapter 8

Summary and Conclusions

Many reasons exist why the TeV-scale of particle physics should be investigated. The Large Hadron Collider collides protons on protons with beam energies of 7 TeV at a design luminosity of $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1} and meets all requirements to study new energy regions. All physics studies at the LHC rely on a thorough understanding of "ordinary" QCD processes, i. e. minimum-bias events and jet production at LHC energies. As a consequence, this involves to measure and to interpret the underlying event in proton-proton collisions - the soft component of a collision adding further particles to the overall event activity. A crucial step will be to determine the level of underlying event activity and to tune event generators as early as possible after LHC startup.

In the context of this work, a large number of new theoretical and analytical tools was successfully used:

- The new event generators PYTHIA 8 and HERWIG++ were tested and used to predict event topologies with one photon and three jets.
- HERWIG++ was successfully passed through the entire detector simulation and event reconstruction chain, and subsequently analysed.
- The new cone jet algorithm SISCone was successfully employed for the first time in a study of the underlying event in charged-jet topologies.
- The new concept of jet areas has been used to describe the diffuse particle flow from the underlying event.
- Last but not least, the complete CMS analysis chain was exercised making full use of the WLCG grid infrastructure.

This thesis described strategies to characterize the underlying event in proton-proton collisions measured with the Compact Muon Solenoid (CMS) detector at the LHC. In particular, feasibility studies were carried out for early collision data at center-of-mass energies of $\sqrt{s} = 10$ TeV and integrated luminosities of $\int \mathcal{L}dt = 1$ pb^{-1}. Current event generators model the underlying event by including contributions from multiple parton-parton scatters. The analysis explored observables that are sensitive to key parameters of multiple interaction parameters, such as the minimal $p_\perp$ cutoff on multiple interactions. Of particular interest are final states with jets clustered from tracks, referred to as charged-jets. In addition, hadron-level studies on final states with one photon and three jets were performed to assess the feasibility to carry out a search for double-parton-scattering at the LHC. Such a direct observation of multiple interactions at the LHC is deemed difficult with the investigated one-dimensional observables.

In describing design and performance of the LHC and the CMS detector, we put particular emphasis on the detection of charged particles with the CMS central tracking system. All detector-level studies were carried out with full-simulation samples of proton-proton collisions at center-of-mass energies of $\sqrt{s} = 10$ TeV. The CMS detector was simulated with a magnetic field of $B = 3.8$ T. Simulation was performed with two example early-alignment scenarios and with the ideal detector geometry. Charged particles with transverse momenta larger than $p_T = 0.9$ GeV/c within the tracker acceptance ($|\eta| < 2.5$) and a reconstructed vertex consistent with the primary event vertex are reconstructed with 90% efficiency; 2 to 3% of reconstructed tracks are fakes. Transverse momenta are resolved at the level of $\delta p_T / p_T = 1\%$.

A SISCone jet algorithm with radius $R = 0.5$ was employed to cluster selected tracks into charged-jets. Being less prone to imperfect alignment and calibration conditions, charged-jets are well-suited to identify the event scale and to determine the dominant direction of charged-particle flow. Charged-jets with pseudorapidities $|\eta| < 2$ have been studied in the transverse momentum range 20 GeV/c $< p_T$(charged-jet) $<$ 250 GeV/c. The matching efficiency to charged-particle-jets, which are reconstructed at hadron-level, is well above 90% in the considered (η, p_T) region. Less than 0.1% of leading charged-particle jets and less than 1% of leading calorimeter jets differ by more than 0.5 rad in (η, ϕ)-space from the direction of the leading charged-jet. The transverse momentum response of charged-jets with respect to charged-particle jets is consistent with unity to an uncertainty of less than 1%. The response with respect to calorimeter jets depends only weakly on p_T(charged-jet) and rises from p_T(charged-jet)/p_T(calorimeter jet) $=$ 60% at p_T(charged-jet) $=$ 20 GeV/c to p_T(charged-jet)/p_T(calorimeter jet) $=$ 80% at p_T(charged-jet) $=$ 250 GeV/c. Constituents of charged-jets are well contained in a cone around the jet axis of radius $\pi/3$ in (η, ϕ) space.

Track multiplicities and transverse momenta in minimum-bias events allow to explore sophisticated models of hadronization (including color reconnection effects) and parton rescattering. We investigated the average track transverse momentum as a function of the track multiplicity. Different alignment conditions do not affect the measurement but model predictions from HERWIG++ 2.2.0 and PYTHIA 6.4 differ significantly.

In order to identify the underlying event in charged-jet events, two approaches have been followed :

- the study of track multiplicities and transverse momenta in the azimuthal region transverse to the leading charged-jet
- and the study of median ratios of jet p_T and jet active area.

In the transverse region of charged-jet events, it was found that track multiplicities and track scalar p_T sums predicted by PYTHIA are larger than activities predicted by HERWIG. In order to tune underlying event models, track observables will need to be corrected to hadron-level. The adopted correction procedure builds bin-by-bin ratios of predictions at hadron-level and at track-level. The obtained correction factors from HERWIG and PYTHIA agree within statistics for p_T(charged-jet) $>$ 100 GeV/c, but exhibit small deviations at smaller transverse momenta. This is attributed to a non-uniform tracking performance in the relevant range of track transverse momenta. Track-level predictions obtained with different alignment scenarios do not differ with respect to the transverse region multiplicity and p_T sum. The impact of misalignment on determining the underlying event activity is thus negligible according to these scenarios. In addition, attempts were made to absorb tracking performances into ratio observables. Unfortunately, the p_T(track) region of uniform tracking efficiency starts at p_T values well above typical track transverse momenta. Thus, the method to absorb tracking performances into ratio observables is expected to work best in the high-momentum-tail of the p_T(track) distribution. In addition, model predictions on the studied ratio observables for the chosen track p_T thresholds did not differ strongly enough to allow for a discrimination of models.

The median ratio of charged-jet p_T and charged-jet active area has been put forward only recently as a possible observable to study the underlying event. It has not yet been measured at the Tevatron although it is an interesting alternative to characterizing the underlying event: Instead of employing topological arguments to identify particle flow not coming from the hard process, jets are classified as belonging to the hard interaction or to the underlying event based on their susceptibility to “catch” additional isotropic activity. Since the method relies on sufficiently many jets being reconstructed, jet p_T thresholds have to be reduced to small values. Including jets with transverse momenta above $p_T(\text{jet}) = 1$ GeV/c ensures on average six or more jets even in the low-bias trigger streams. Again, PYTHIA predicts larger values for the observable associated with the underlying event activity than HERWIG does; yet, PYTHIA's and HERWIG's predictions on the median ratio of charged-jet p_T and charged-jet active area differ by smaller amounts than PYTHIA's and HERWIG's predictions on transverse region activities in charged-jet topologies. Efficiency corrections for charged-jet areas were not studied and must be investigated before claiming real discriminative power of this observable. Nonetheless, the present work is the first experimental study of this exciting topic and future measurements of charged-jet areas will most likely be carried out in the way presented here.

To conclude, note that successful tunes of underlying event models to data involve a number of free parameters of the particular model, as well as a specific choice of parton density function. Specifically, it is to be kept in mind that tunes to one final state may yield different predictions on another final state. It is thus not without continuous dialogue between experimenters and theorists that a thorough understanding of the underlying event in proton-proton collisions at the LHC will be achieved.

We look forward to putting the newly developed tools to "real" use, once the first data arrives from the detectors at the LHC, quickly enabling us to characterize the underlying event in proton-proton collisions.

Bibliography

[1] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[2] A. Salam, Originally printed in *Svartholm: Elementary Particle Theory, Proceedings Of The Nobel Symposium Held 1968 At Lerum, Sweden*, Stockholm 1968, 367-377.

[3] S. L. Glashow, J. Iliopoulos and L. Maiani, Phys. Rev. **D2**, 1285 (1970).

[4] R. K. Ellis, W. J. Stirling and B. R. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8**, 1 (1996).

[5] P. W. Higgs, Phys. Lett. **12**, 132 (1964).

[6] P. W. Higgs, Phys. Rev. **145**, 1156 (1966).

[7] F. Englert and R. Brout, Phys. Rev. Lett. **13**, 321 (1964).

[8] G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, Phys. Rev. Lett. **13**, 585 (1964).

[9] Particle Data Group, C. Amsler *et al.*, Phys. Lett. **B667**, 1 (2008).

[10] G. Bertone, D. Hooper and J. Silk, Phys. Rept. **405**, 279 (2005), [hep-ph/0404175].

[11] E. Witten, Nucl. Phys. **B188**, 513 (1981).

[12] R. K. Kaul, Phys. Lett. **B109**, 19 (1982).

[13] R. K. Kaul, Pramana **19**, 183 (1982).

[14] S. Weinberg, Phys. Rev. **D19**, 1277 (1979).

[15] L. Susskind, Phys. Rev. **D20**, 2619 (1979).

[16] K. Lane, hep-ph/0202255.

[17] C. T. Hill and E. H. Simmons, Phys. Rept. **381**, 235 (2003), [hep-ph/0203079].

[18] R. Shrock, hep-ph/0703050.

[19] J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974).

[20] J. Wess and B. Zumino, Phys. Lett. **B49**, 52 (1974).

[21] H. P. Nilles and M. Nusbaumer, Phys. Lett. **B145**, 73 (1984).

[22] I. Antoniadis, Phys. Lett. **B246**, 377 (1990).

[23] N. Arkani-Hamed, S. Dimopoulos and G. R. Dvali, Phys. Lett. **B429**, 263 (1998), [hep-ph/9803315].

[24] L. Randall and R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999), [hep-ph/9905221].

[25] G. F. Giudice, R. Rattazzi and J. D. Wells, Nucl. Phys. **B544**, 3 (1999), [hep-ph/9811291].

[26] C. Csaki, M. L. Graesser and G. D. Kribs, Phys. Rev. **D63**, 065002 (2001), [hep-th/0008151].

[27] N. Arkani-Hamed, A. G. Cohen and H. Georgi, Phys. Lett. **B513**, 232 (2001), [hep-ph/0105239].

[28] J. M. Campbell, J. W. Huston and W. J. Stirling, Rept. Prog. Phys. **70**, 89 (2007), [hep-ph/0611148].

[29] S. D. Drell and T.-M. Yan, Ann. Phys. **66**, 578 (1971).

[30] J. C. Collins and D. E. Soper, Ann. Rev. Nucl. Part. Sci. **37**, 383 (1987).

[31] L. N. Lipatov, Sov. J. Nucl. Phys. **20**, 94 (1975).

[32] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15**, 438 (1972).

[33] G. Altarelli and G. Parisi, Nucl. Phys. **B126**, 298 (1977).

[34] Y. L. Dokshitzer, Sov. Phys. JETP **46**, 641 (1977).

[35] S. Catani, hep-ph/0005233.

[36] F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).

[37] T. Kinoshita, J. Math. Phys. **3**, 650 (1962).

[38] T. D. Lee and M. Nauenberg, Phys. Rev. **133**, B1549 (1964).

[39] T. Sjostrand, S. Mrenna and P. Skands, JHEP **05**, 026 (2006), [hep-ph/0603175].

[40] G. Corcella *et al.*, hep-ph/0210213.

[41] M. Bahr *et al.*, 0803.0883.

[42] T. Gleisberg *et al.*, JHEP **02**, 056 (2004), [hep-ph/0311263].

[43] S. Catani, F. Krauss, R. Kuhn and B. R. Webber, JHEP **11**, 063 (2001), [hep-ph/0109231].

[44] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, Nucl. Phys. **B406**, 187 (1993).

[45] S. D. Ellis and D. E. Soper, Phys. Rev. **D48**, 3160 (1993), [hep-ph/9305266].

[46] G. P. Salam and G. Soyez, JHEP **05**, 086 (2007), [0704.0292].

[47] D. Stump *et al.*, JHEP **10**, 046 (2003), [hep-ph/0303013].

[48] G. Watt, A. D. Martin, W. J. Stirling and R. S. Thorne, 0806.4890.

[49] V. S. Fadin, E. A. Kuraev and L. N. Lipatov, Phys. Lett. **B60**, 50 (1975).

[50] E. A. Kuraev, L. N. Lipatov and V. S. Fadin, Sov. Phys. JETP **44**, 443 (1976).

[51] E. A. Kuraev, L. N. Lipatov and V. S. Fadin, Sov. Phys. JETP **45**, 199 (1977).

[52] I. I. Balitsky and L. N. Lipatov, Sov. J. Nucl. Phys. **28**, 822 (1978).

[53] H1, C. Adloff *et al.*, Eur. Phys. J. **C21**, 33 (2001), [hep-ex/0012053].

[54] H1, C. Adloff *et al.*, Eur. Phys. J. **C30**, 1 (2003), [hep-ex/0304003].

[55] ZEUS, S. Chekanov *et al.*, Phys. Rev. **D67**, 012007 (2003), [hep-ex/0208023].

[56] ZEUS, S. Chekanov *et al.*, Eur. Phys. J. **C42**, 1 (2005), [hep-ph/0503274].

[57] T. Sjostrand and P. Z. Skands, JHEP **03**, 053 (2004), [hep-ph/0402078].

[58] B. R. Webber, Nucl. Phys. **B238**, 492 (1984).

[59] B. Andersson, G. Gustafson, G. Ingelman and T. Sjostrand, Phys. Rept. **97**, 31 (1983).

[60] T. Sjostrand, Nucl. Phys. **B248**, 469 (1984).

[61] UA5, G. J. Alner *et al.*, Nucl. Phys. **B291**, 445 (1987).

[62] UA5, R. E. Ansorge *et al.*, Z. Phys. **C37**, 191 (1988).

[63] E735, T. Alexopoulos *et al.*, Phys. Lett. **B353**, 155 (1995).

[64] UA1, C. Albajar *et al.*, Nucl. Phys. **B309**, 405 (1988).

[65] H1, S. Aid *et al.*, Z. Phys. **C70**, 17 (1996), [hep-ex/9511012].

[66] CDF, A. A. Affolder *et al.*, Phys. Rev. **D65**, 092002 (2002).

[67] Sjostrand, T., Monte Carlo Generators for the LHC, Academic Training Lectures, CERN, 2005.

[68] T. Sjostrand and M. van Zijl, Phys. Lett. **B188**, 149 (1987).

[69] B. L. Combridge, J. Kripfganz and J. Ranft, Phys. Lett. **B70**, 234 (1977).

[70] M. L. Mangano, Introduction to hadronic collisions: theoretical concepts and practical tools for the LHC, Lectures presented at Scuola Normale Superiore, Pisa, Italy, 2008.

[71] A. Donnachie and P. V. Landshoff, Phys. Lett. **B296**, 227 (1992), [hep-ph/9209205].

[72] Sjostrand, T., Minimum Bias and Underlying-Event Physics in Pythia, Monte Carlo School, Physics at the Terascale, DESY, 2008.

[73] CDF, R. Field and R. C. Group, hep-ph/0510198.

[74] CDF, D. E. Acosta *et al.*, Phys. Rev. **D70**, 072002 (2004), [hep-ex/0404004].

[75] T. Sjostrand and P. Z. Skands, Eur. Phys. J. **C39**, 129 (2005), [hep-ph/0408302].

[76] CDF, Measurement of inelastic $p\bar{p}$ inclusive cross sections at $\sqrt{s} = 1.96$ TeV, CDF note 9337.

[77] R. Corke, Multiple Interactions in Pythia 8, To appear in the Proceedings of the First International Workshop on Multiple Partonic Interactions at the LHC (MPI@LHC), Perugia, Italy, 2008.

[78] N. Paver and D. Treleani, Phys. Lett. **B146**, 252 (1984).

[79] M. Bahr, S. Gieseke and M. H. Seymour, JHEP **07**, 076 (2008), [0803.3633].

[80] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Z. Phys. **C72**, 637 (1996), [hep-ph/9601371].

[81] I. Borozan and M. H. Seymour, JHEP **09**, 015 (2002), [hep-ph/0207283].

[82] S. Gieseke, P. Stephens and B. Webber, JHEP **12**, 045 (2003), [hep-ph/0310083].

[83] e. . Evans, Lyndon and e. . Bryant, Philip, JINST **3**, S08001 (2008).

[84] M. Bahr, J. M. Butterworth and M. H. Seymour, 0806.2949.

[85] CMS, R. Adolphi *et al.*, JINST **3**, S08004 (2008).

[86] CMS, CERN-LHCC-98-06.

[87] T. Speer *et al.*, Nucl. Instrum. Meth. **A559**, 143 (2006).

[88] R. Fruhwirth, Nucl. Instrum. Meth. **A262**, 444 (1987).

[89] C. Noding, PoS **VERTEX2007**, 026 (2007).

[90] C. Roland, Nucl. Instrum. Meth. **A566**, 123 (2006).

[91] W. Adam, R. Fruhwirth, A. Strandlie and T. Todorov, ECONF **C0303241**, TULT009 (2003), [physics/0306087].

[92] CMS, G. L. Bayatian *et al.*, CERN-LHCC-2006-001.

[93] I. Belotelov *et al.*, CERN-CMS-NOTE-2006-008.

[94] V. Karimaki, A. Heikkinen, T. Lampen and T. Linden, physics/0306034.

[95] V. Karimaki, T. Lampen and F. P. Schilling, CERN-CMS-NOTE-2006-018.

[96] E. Widl, R. Fruhwirth and W. Adam, CERN-CMS-NOTE-2006-022.

[97] V. Blobel and C. Kleinwort, hep-ex/0208021.

[98] P. Schleper, G. Steinbruck and M. Stoye, CERN-CMS-NOTE-2006-011.

[99] N. Akchurin and R. Wigmans, Rev. Sci. Instrum. **74**, 2955 (2003).

[100] G. C. Blazey *et al.*, hep-ex/0005012.

[101] M. Cacciari, G. P. Salam and G. Soyez, JHEP **04**, 005 (2008), [0802.1188].

[102] M. Cacciari and G. P. Salam, Phys. Lett. **B659**, 119 (2008), [0707.1378].

[103] CMS, Plans for Jet Energy Corrections at CMS, AN-2007/055, 2008.

[104] P. Andreetto *et al.*, J. Phys. Conf. Ser. **119**, 062007 (2008).

[105] R. Pordes *et al.*, J. Phys. Conf. Ser. **119**, 052028 (2008).

[106] CMS, D. Spiga, Nucl. Phys. Proc. Suppl. **172**, 141 (2007).

[107] CMS, The 2008 CMS Computing, Software and Analysis Challenge, IN-2008/044, 2008.

[108] `https://twiki.cern.ch/twiki/bin/view/CMS/ProductionSummer2008`, Summer08: Full Simulation SM Production for Physics at 10 TeV.

[109] P. Mendez Lorenzo *et al.*, 3rd EGEE User Forum, Clermont-Ferrand, France, 11 - 14 Feb 2008, pp.136 [http://cdsweb.cern.ch/record/1123355].

[110] Field, R., Studying the Underlying Event at CDF and the LHC, Seminar presented at the University of California, Berkeley, 2009.

[111] `https://twiki.cern.ch/twiki/bin/view/CMS/TSG_27_V_08`, Trigger Menus for iCSA08: 2E30, 2E31 in 2_0_x, version 0.3 (May 2008).

[112] GEANT4, S. Agostinelli *et al.*, Nucl. Instrum. Meth. **A506**, 250 (2003).

[113] J. Allison *et al.*, IEEE Trans. Nucl. Sci. **53**, 270 (2006).

[114] M. Bahr *et al.*, 0812.0529.

[115] `https://twiki.cern.ch/twiki/bin/view/CMS/TSG_18_II_09_8E29`, High-Level Trigger paths for the 8E29 “core” menu, version 0.3 (February 2009).

[116] `https://twiki.cern.ch/twiki/bin/view/CMS/TSG_18_II_09_1E31`, High-Level Trigger paths for the 1E31 “core” menu, version 0.3 (February 2009).

[117] M. Cacciari and G. P. Salam, Phys. Lett. **B641**, 57 (2006), [hep-ph/0512210].

[118] Axial Field Spectrometer, T. Akesson *et al.*, Z. Phys. **C34**, 163 (1987).

[119] UA2, J. Alitti *et al.*, Phys. Lett. **B268**, 145 (1991).

[120] CDF, F. Abe *et al.*, Phys. Rev. **D47**, 4857 (1993).

[121] CDF, F. Abe *et al.*, Phys. Rev. **D56**, 3811 (1997).

[122] ATLAS, CERN-LHCC-99-15.

[123] M. Bahr, *Underlying Event Simulation in the* HERWIG++ *Event Generator*, PhD thesis, Universität Karlsruhe (TH), 2008.

[124] C. Goebel, F. Halzen and D. M. Scott, Phys. Rev. **D22**, 2789 (1980).

[125] N. Paver and D. Treleani, Nuovo Cim. **A70**, 215 (1982).

[126] B. Humpert, Phys. Lett. **B131**, 461 (1983).

[127] B. Humpert and R. Odorico, Phys. Lett. **B154**, 211 (1985).

[128] T. Sjostrand and M. van Zijl, Phys. Rev. **D36**, 2019 (1987).

[129] G. Calucci and D. Treleani, Nucl. Phys. Proc. Suppl. **71**, 392 (1999), [hep-ph/9711225].

[130] D0, H. Weerts, Presented at 9th Topical Workshop on Proton - Anti-proton Collider Physics, Tsukuba, Japan, 18-22 Oct 1993.

[131] W. T. Giele, E. W. N. Glover and D. A. Kosower, Phys. Rev. **D52**, 1486 (1995), [hep-ph/9412338].

[132] Azzurri, P. and Mangano, B., Optimal filtering of fake tracks, CMS IN-2008/017, 2008.

[133] P. Z. Skands, Prepared for 5th Les Houches Workshop on Physics at TeV Colliders, Les Houches, France, 11-29 Jun 2007.

[134] CMS, Measurement of the Underlying Event in Jet Topologies using Charged Particle and Momentum Densities, PAS-QCD-07-003, 2007.

[135] P. Skands and D. Wicke, Eur. Phys. J. **C52**, 133 (2007), [hep-ph/0703081].

[136] J. H. Friedman, Lectures presented at CERN School of Computing, Godoysund, Norway, Aug 11-24, 1974.

Acknowledgements

This dissertation would not have been possible without the support of many people inside and outside of the Institut für Experimentalphysik and CMS.

My Ph. D. project was carried out within the Hamburg University's CMS group, led by Peter Schleper and Robert Klanner. I am grateful to both for their supervision of my thesis work. Joachim Mnich and Johannes Haller kindly acted as referees for dissertation and disputation.

Many of the ideas presented in this work were influenced by discussions with Hartmut Stadie. Hartmut was not only an expert help in computing topics, but also a thorough reviewer of my physics analysis.

I sincerely enjoyed working within the CMS task force to study the underlying event in charged-jet topologies. Livio Fano, Paolo Bartalini, Rick Field, and Filippo Ambroglini provided a stimulating and fruitful atmosphere to carry out this exciting research line. I also would like to thank all members and convenors of the CMS QCD group whom I had the pleasure to work with.

During my four months in Lund, Torbjörn Sjöstrand took wonderful care of teaching me the concepts of Pythia's multiple interaction framework. My studies of double-parton scattering greatly profited from his expertise and insight.

My work for the CMS computing efforts focussed on the Tier-2 activities at DESY. Without the help and guidance of Michael Ernst, Birgit Lewendel, Christoph Wissing, and Yves Kemp, I could not possibly have helped to keep the CMS data transfers to and from DESY successful.

Special thanks are owed to all my colleagues at Hamburg University and DESY for contributing to an inspiring, friendly, and respectful environment. In particular, I am indebted to those colleagues involved in proof-reading the dissertation, and acting as questioners during my disputation: Hannes Jung, Judith Katzy, Katerina Lipka, Thomas Schörner-Sadenius, and Roger Wolf.

Financial support by the Marie Curie Fellowship program for Early Stage Research Training, by MCnet (contract number MRTN-CT-2006-035606) and by the German Federal Ministry of Education and Research (BMBF) is gratefully acknowledged.

“If the journey is interesting and educational,
why despair that the end station is not yet reached?”
— Torbjörn Sjöstrand