Experimentelle Physik

Measurement of the Charged-Particle Multiplicity in Proton–Proton Collisions with the ALICE Detector

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften im Fachbereich Physik
der Mathematisch-Naturwissenschaftlichen Fakultät
der Westfälischen Wilhelms-Universität Münster

vorgelegt von
Jan Fiete Große-Oetringhaus
aus Herdecke

— 2009 —

Dekan:	Prof. Dr. Johannes P. Wessels
Erster Gutachter:	Prof. Dr. Johannes P. Wessels
Zweiter Gutachter:	Dr. Karel Šafařík

Tag der Disputation:	17.04.09
Tag der Promotion:	17.04.09

For my parents, Conny and Hajo,
and for Mareike

"I am more sure of the conclusions than of any single argument which suggested them to me."
Richard P. Feynman, 1969 [Fey69]

Contents

Introduction

On the 10th of September 2008, at 10:25 CEST, the first beam circulated through the Large Hadron Collider (LHC) at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. An event in science that was followed by an unprecedented number of people around the Globe; more than 300 journalists and 30 TV stations were present on the CERN site, 500 TV companies picked up the live feed, and CERN's web page recorded 100 million hits during the day. The start of the LHC, the largest particle accelerator in the world, fascinated people world-wide across profession and age. Particle physics and science in general have proven to be of special interest for everybody. Science may be on the way to becoming a mainstream topic again.

The LHC with an energy that has never been achieved before in a particle accelerator experiment represents a milestone, not only of high-energy physics and the research therein, but also for human knowledge. Research that focuses on the properties and origin of matter addresses basic questions driven by mankind's curiosity since the very beginning. Questions about the World, the Universe, and especially about the Universe's birth, the Big Bang.

LHC's physics goals are ambitious, ranging from the verification of the Standard Model via the discovery of the Higgs boson, over extensions of the Standard Model, like supersymmetry and extra-dimensions as well as the study of CP violation, to the recreation of the state that existed microseconds after the Big Bang, the quark–gluon plasma. To achieve these physics goals a detailed understanding of LHC collisions has to be acquired which in turn requires a deep knowledge of the detectors used. In particular the study of rare signals and signatures needs the understanding of the bulk part of the collisions. This includes the understanding of multiplicity and momentum spectra, particle abundances, as well as correlations between these observables. These basic measurements in p+p collisions are required by all LHC experiments and also as a reference point for the study of heavy-ion collisions that will be performed at the LHC. Many of the signatures of the quark–gluon plasma reveal themselves through comparison of heavy-ion and proton collisions and make therefore a precise and profound understanding of p+p collisions essential.

The ALICE detector at the LHC is optimized for collisions of heavy ions with the aim of studying strongly-interacting matter, especially the quark–gluon plasma and the associated phase transition. In the context of the LHC experiments, ALICE also plays a special role in p+p collisions. Its sensitivity at very low transverse momentum p_T and excellent particle identification allow measurements to be performed that are not possible for the other LHC experiments but in turn contribute to understanding their results.

This thesis describes my Ph.D. work as part of the ALICE collaboration. My main contribution is the preparation of two analyses that can be performed with early data, the charged-particle pseudorapidity density $dN_{ch}/d\eta$ and the multiplicity distribution of charged particles. The analysis procedures are fundamentally different because the former produces an average while the latter obtains a distribution and requires the unfolding of the measured spectrum. The aim is to have a fully developed analysis before the start of data-taking. Thus ALICE will be able to produce a measured spectrum on a very short time scale once data-taking has started. These two measurements are planned to be among the first publications of the ALICE collaboration. The analyses have been developed with and tested extensively on simulated data. Care has been taken to divide the available data such that a realistic case of measured data versus data for corrections is made. The associated systematic uncertainties have been evaluated, aided by the use of two different event generators to allow for different input spectra.

Some of the technical aspects that were part of my Ph.D. work are also described in this thesis. I have set up the CERN Analysis Facility (CAF), a PROOF cluster that allows fast analysis for ALICE collaborators. It has also been used to produce the analysis results presented in this thesis. My work included extensions of the ALICE Offline software framework and contributions to the PROOF development. I developed the Shuttle framework, a system for automatic readout of conditions data from the different subdetectors of the experiment. This system has been successfully integrated in the experimental setup of ALICE and has been used during cosmic ray data-taking since 2007 and during the LHC startup. It is described in the appendix.

The thesis is divided into two parts. The first part (Chapter 1 to 4) introduces the theoretical framework, presents models and previous charged-particle multiplicity measurements, and describes the LHC and the ALICE detector. The second part (Chapter 5 to 7) starts off with practical aspects of the analysis like trigger, event and track selection, and describes the two analysis procedures including the systematic studies. It then concludes with a summary of event generator and model predictions for LHC's energy. The reader who is familiar with the LHC and the ALICE detector may start reading the second part directly.

In detail the outline is the following: Chapter 1 introduces the theoretical concepts that are important for the measurements. A synopsis of pseudorapidity density and multiplicity distribution measurements that have been performed prior to the start of the LHC at $\sqrt{s} = 6\,\text{GeV} - 1.8\,\text{TeV}$ is given in Chapter 2. The subsequent chapters describe the LHC (Chapter 3) and the ALICE experiment (Chapter 4). Chapter 5 introduces the event and track selection as well as the datasets used in this thesis. The measurement of the pseudorapidity density $dN_{ch}/d\eta$ and the associated systematic uncertainties are detailed in Chapter 6. Chapter 7 describes the measurement of the multiplicity distribution including the systematic studies. In the last chapter event generator and model predictions for $\sqrt{s} = 10\,\text{TeV}$ and $14\,\text{TeV}$ are summarized.

Chapter 1

Theoretical Framework

This chapter gives a brief introduction into the theoretical concepts that are relevant for the charged-particle multiplicity measurement. The chapter begins with general concepts about multiple-particle production in high-energy collisions. Subsequently, theoretically- and phenomenologically-based descriptions of the multiplicity distribution are discussed.

The Standard Model that is very successful in the description of the fundamental forces and the composition of matter is introduced. The quark–gluon plasma, its signatures and their relation to p+p measurements are detailed. The concepts of models based on the dual topological unitarization, i.e. the Quark–Gluon String Model and the Dual Parton Model are briefly outlined. A discussion of the approaches of two event generators, Pythia and Phojet, to describe high-energy collisions follows. Theoretical descriptions of the multiplicity distribution are outlined. These are applied to existing data in the subsequent Chapter 2.

Family	Quarks			Leptons		
	Name	Charge	Mass	Name	Charge	Mass
1	u	$2/3\ e$	$1.5 - 3.3$ MeV/c^2	e^-	$-e$	0.511 MeV/c^2
	d	$-1/3\ e$	$3.5 - 6.0$ MeV/c^2	ν_e	0	< 2 eV/c^2
2	c	$2/3\ e$	$1.27^{+0.07}_{-0.11}$ GeV/c^2	μ^-	$-e$	106 MeV/c^2
	s	$-1/3\ e$	104^{+26}_{-34} MeV/c^2	ν_μ	0	< 0.19 MeV/c^2
3	t	$2/3\ e$	171.2 ± 2.1 GeV/c^2	τ^-	$-e$	1.78 GeV/c^2
	b	$-1/3\ e$	$4.2^{+0.17}_{-0.07}$ GeV/c^2	ν_τ	0	< 18.2 MeV/c^2

Table 1.1: Constituents of matter in the Standard Model [Ams08].
The table shows the constituents of matter divided into quarks and leptons, each in three families with two members.

1.1 The Standard Model

The Standard Model describes the fundamental forces and the composition of matter. It is a gauge theory including the strong, weak, and electromagnetic force and the related interactions; gravity is not part of the Standard Model and not further addressed. Matter is constituted out of point-like particles which have a spin of 1/2 and are grouped into three families. Each family has two *quark* and two *lepton* members; these are listed in Table 1.1. Leptons are affected by the weak force and the charged ones in addition by the electromagnetic force. Quarks have a property called *color* playing the role of charge in the strong force. The color can take one out of three possible values (conventionally *red*, *green*, and *blue*). They do not appear freely, they are confined and appear in the form of *hadrons* that are colorless (also called *white*), i.e. in the corresponding SU(3)-algebra the colors of the constituent quarks sum to 0 (note that red + green + blue = white). Hadrons are grouped into *baryons* and *mesons*. Baryons consist of three quarks, qqq or $\bar{q}\bar{q}\bar{q}$ (e.g. the proton: uud). Mesons consist of two quarks, $q\bar{q}$ (e.g. the π^+: $u\bar{d}$). Quarks are affected by the strong, weak, and electromagnetic force.

The forces are mediated by the exchange of gauge bosons which are listed in Table 1.2 together with their relative coupling strengths. Part of the Standard Model is the strong force which is mediated by gluons that have a color charge, the theoretical framework is *Quantum Chromodynamics (QCD)*. The theory of *Quantum Electrodynamics (QED)* describes the electromagnetic force, mediated by the exchange of photons. The weak force governed by the exchange of $W^\pm$s and Z^0s is described by the *electroweak theory* that includes the electromagnetic force. It has thus four gauge bosons (γ, $W^\pm$, and Z^0)

Force	Strength	Gauge Boson(s)	Applies on
Strong force	1	8 Gluons (g)	Quarks, gluons
Electromagnetic force	$\sim 10^{-2}$	Photon (γ)	All charged particles
Weak force	$\sim 10^{-7}$	$W^{\pm}$, Z^0	Quarks, leptons
Gravitation	$\sim 10^{-39}$	Gravitons	All massive particles

Table 1.2: Fundamental forces [Per00].

All forces, except gravitation, are described by the Standard Model. Their strength is given relative to the strength of the strong force for two protons at a distance of about 0.5 fm. The gravitons, the gauge bosons of the gravitation, are postulated but have not been found yet.

being the first successful theory that treats some of the fundamental forces that are different at low energies on an equal footing.

The Standard Model has also, however, limitations that require extensions to keep the theory consistent. The most prominent example being that the masses of the electroweak gauge bosons evaluate to zero within the theory. Something that is clearly inconsistent with experiment. This situation can be resolved by an additional gauge boson added to the theory, the *Higgs* boson [Hig64]. The Higgs mechanism generates the masses for the $W^{\pm}$ and Z^0 while the γ remains massless. The puzzle about the existence of the Higgs boson may soon be resolved by the ATLAS and CMS experiments at the LHC.

1.2 The Quark–Gluon Plasma

Quark-based matter appears at 'low temperature' as confined hadrons that are quark-antiquark-pairs or compounds of three quarks or antiquarks. These quarks cannot be isolated by pulling them apart; the potential energy between the quarks increases with the distance. Once the energy is above the threshold for $q\bar{q}$-pair production, such a pair is created (see e.g. [Per00])[1]. However, at high temperature T or high baryochemical potential μ_B a phase transition is predicted to a state where quarks and gluons are not confined [Cab75]. In the so-called *quark–gluon plasma (QGP)* quarks and gluons are not bound in hadrons. Long-range interactions are screened. Quarks and gluons are only subject to short-range interactions. The coupling constant at short distances is small leading to just weak coupling between the quarks and gluons. Lattice QCD calculations predict the phase transition to the QGP for $\mu_B \approx 0$ at approximately 150 MeV [Aok06]

[1] Quark pair creation by string breaking is also discussed in Section 1.5.1.

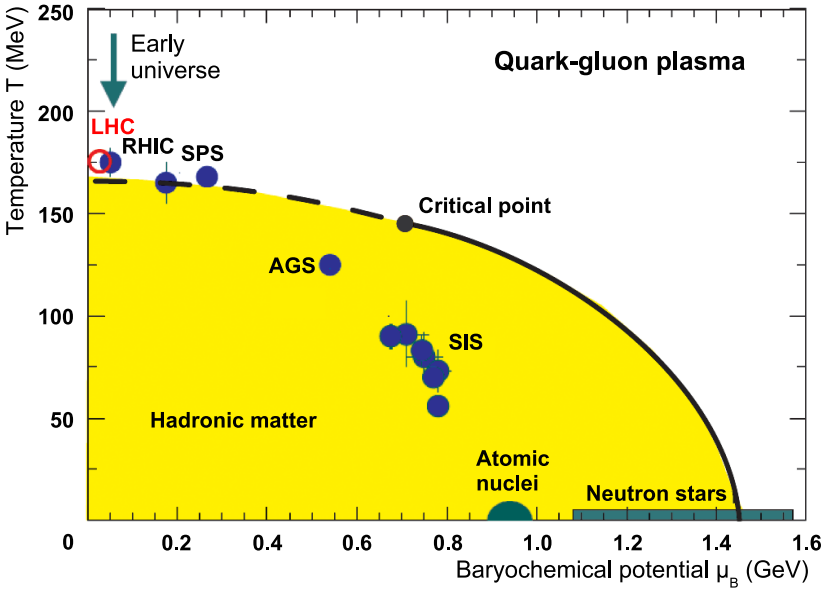

Figure 1.1: The phase diagram of hadronic matter.
Chemical freeze-out points measured by various experiments are also shown. The existence and position of the critical point as well as the nature of the phase transition are only indicative and still under intensive study and discussion.

to 190 MeV [Che06]. A simplified approach is the so-called *bag model* where quarks are considered massless in a finite area around them, called bag, and infinitely massive outside that bag. Quarks and gluons are confined due to a bag pressure that acts upon the bag. They get deconfined when their kinetic energy is larger than the bag pressure. Calculations based on the bag model yield a phase transition temperature of 144 MeV at $\mu_B = 0$ [Won94]. Figure 1.1 shows the phase diagram with the regions of hadronic matter and of the QGP. Also shown are the hadronization (chemical freeze-out) points measured by various experiments.

A QGP is expected to have existed shortly after the Big Bang. Currently it may exist in the very dense cores of neutron stars (see e.g. [Alf03]). To produce a QGP in an experiment, very high-energy densities are needed that can be achieved in ultra-relativistic heavy-ion collisions at the LHC. The spatial extension of the QGP phase in a heavy-ion collision is expected to be of the order of a few fm and the lifetime of the order of a few fm/c. The incoming particles may not overlap completely: it is distinguished between participating nucleons in the so-called *fireball* and spectator nucleons. While the quarks and gluons move apart the temperature decreases. When the system reaches the *chemical freeze-out*, the quarks and gluons form hadrons. At this stage the abundances of different particle species are fixed. Non-strongly interacting particles produced in the plasma like photons and leptons pass uninfluenced through this freeze-out. Particles

still interact elastically with each other until the system reaches the *thermal freeze-out* (also called *kinematic freeze-out*). At this point the composition and momentum spectra are essentially fixed, significant interactions no longer occur. These are the particles that are then measured in the experiment which provide only indirect evidence for the QGP. These signatures may already have been distorted by interactions during and after the freeze-out. In conclusion, a QGP phase cannot be detected directly in a high-energy physics experiment. Instead a set of signatures provides evidence; these are discussed in the following.

Many of the signatures require the comparison of collisions where a QGP is suspected to have formed and collisions where no QGP is expected. The collective effects in heavy-ion collisions need to be disentangled from effects already present in light collision systems like p+p or p+heavy-ion (where no QGP is expected to form). For this purpose the same observables in heavy and light collision systems are usually directly compared. A further possibility is the comparison of central and peripheral heavy-ion collisions. For example for particle yields so-called *nuclear modification factors* R_{AA} are calculated. These are the ratios of yields in heavy-ion collisions (N_{AA}) and in light collision systems like p+p (N_{pp}) normalized to the number of independent binary nucleon-nucleon collisions (N_{coll}):

$$R_{\mathrm{AA}}(p_T) = \frac{1}{N_{\mathrm{coll}}} \frac{d^3N_{\mathrm{AA}}/d\eta d^2p_T}{d^3N_{\mathrm{pp}}/d\eta d^2p_T}. \tag{1.1}$$

Depending on the compared systems an isospin correction has to be taken into account owing to the different mixture of protons and neutrons in the two systems. In the same manner, the factor R_{CP} denotes the ratio of central and peripheral collisions. However, it is model-dependent to estimate the number of participants in peripheral collisions that is needed for the above-mentioned normalization. Therefore, the systematic error in this comparison is larger than for the comparison to a light collision system. Generally, it is of big advantage to take both measurements in the same experiment thus under identical experimental conditions. Many systematic uncertainties that are present in the measurements cancel when data is used from the same experiment taken at the same energy.

The detailed understanding of the properties of a light collision system like p+p is therefore a precondition for a successful heavy-ion program. In the following paragraphs, signatures of the QGP are outlined with special focus on the need for p+p reference data.

Kinematic Probes. The behavior of the energy density ϵ, pressure p, and entropy density s as a function of the temperature T is studied with the aim of observing a behavior characteristic for the rapid change in the degrees of freedom, indicating a

phase transition. However, a first-order phase transition, which would result in a non-smooth dependence between the variables, is most likely not occurring at vanishing baryochemical potential and thus under the experimental conditions at the LHC.

With a few assumptions, T, s, and ϵ can be determined by measuring the average transverse momentum, the hadron rapidity[2] density, and the transverse energy, respectively [Har96]. However, the measurement of thermodynamical variables is affected by the late stages of the evolution of the system, i.e. after the freeze-out, and thus conclusions about the earlier stages are not straightforward.

Particle Yields and Ratios. Particle yields and ratios depend on the state of the system at the chemical freeze-out. Statistical models [Bra03] allow the calculation of hadron yields of an equilibrated system. In turn, the measured yields can be used to calculate the parameters of the equilibrated system, in particular the chemical freeze-out temperature and the baryochemical potential.

Data from the SPS (Super Proton Synchrotron) and RHIC (Relativistic Heavy-Ion Collider) suggest that approximate chemical equilibrium is achieved for u, d, and s quarks in their collisions [Bra96, Bra99, Let00, Bec03, Ada05].

Strangeness Enhancement. The ratio between produced s and u quarks does not show a significant $\sqrt{s}$-dependence in p+p collisions. Contrarily, the number of strange particles is enhanced in heavy-ion collisions. This is explained by the lower threshold energy of the production of strange particles in deconfined matter compared to hadronic matter: the effective masses of the quarks change from constituent masses in hadronic matter to bare masses in deconfined matter. As an example the associated production of strange particles and quarks can be compared: in hadronic matter the lightest strange particle is the kaon, thus the production of a kaon pair has a threshold energy of about 987 MeV. In the deconfined medium the production of an $s\bar{s}$-pair has a threshold of only about 140 – 260 MeV and the strange quark content is expected to reach equilibrium quickly [Raf82, Hei94]. Strangeness enhancement has been clearly observed in collisions at the SPS [And99, Ant02, Alt04] and at RHIC [Ada05].

Quarkonia Yields. Quarkonia (J/Ψ, Ψ', Υ, Υ', Υ'') are made of $c\bar{c}$- and $b\bar{b}$-pairs that are created in the initial phase of the collision. Color screening in the deconfined phase leads to melting of quarkonia states [Mat86, Sat90]. The freed c and b quarks are unlikely to recombine to quarkonia states during freeze-out if their concentrations are

[2]Kinematic variables that are commonly used in high-energy physics are defined in Appendix A.

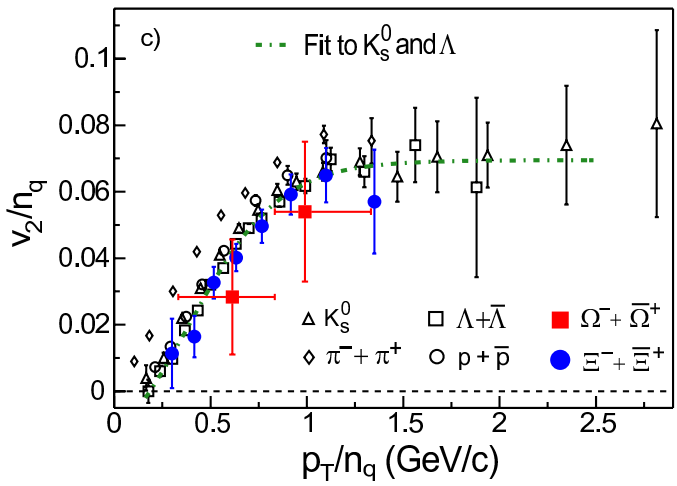

Figure 1.2: Constituent-parton scaling of elliptic flow.

Transverse momentum dependence of the event elliptic anisotropy parameter (v_2) for various hadron species (minimum-bias Au+Au collisions at $\sqrt{s_{NN}} = 200\,\text{GeV}$ measured by STAR). Both axes are scaled with the number of constituent valence quarks (n_q) in the given hadron indicating that the amount of flow depends only on the quark content and not on the hadron configuration. Figure taken from [Don04].

small. In this case, the quarkonia yield in the presence of a QGP compared to p+p collisions should be suppressed with a characteristic dependence on the size of the specific quarkonium. This has been observed at the SPS [Ram06]. Statistical hadronization assesses the probability for recombination into $c\bar{c}$- and $b\bar{b}$-pairs (so-called *quarkonia regeneration*) [And07]. At much higher energies c and b quarks are produced in large abundances and statistical hadronization may even lead to enhanced quarkonia production at freeze-out.

The formation of a $c\bar{c}$-bound state takes about 1 fm/c. Thus $c\bar{c}$-pairs with a large p_T may be able to escape the fireball before they separate resulting in a p_T-dependent suppression factor.

Flow. Radial flow is caused by the matter-density gradient between the center of the fireball and the boundary region. Additionally, elliptic flow occurs in non-central collisions due to the asymmetric, almond shaped, collision region which results in an anisotropic expansion. This effect is self-quenching in the sense that the expansion reduces the anisotropy and thus the elliptic flow. Therefore, the measurement of elliptic flow carries signatures of the earlier collision stages [Kol04]. The analysis of transverse momentum spectra at central rapidity with respect to the reaction plane in an event allows elliptic flow occurring in the collision to be analyzed.

RHIC results on elliptic flow agree with hydrodynamic calculations that assume an ideal relativistic fluid consisting of strongly-interacting matter with very short mean free paths that flows free of viscosity [Ada05]. Contrarily, calculations assuming a pure hadron gas (and no deconfined phase) underpredict the measured elliptic flow. The level of flow that is measured at RHIC and the fact that the measured elliptic flow of identified hadrons scales with the constituent quarks, shown in Figure 1.2, is at present one of the strongest arguments for the discovery of the QGP.

Identical Particle Interferometry. Two (or more) particle momentum correlations reveal information about the space–time dynamics of the collision. This procedure is analogous to *Hanbury-Brown and Twiss (HBT) interferometry* that has been successfully used in astrophysics to determine the angular diameter of stars [Han56]. In high-energy physics these correlations allow the size, lifetime, and flow patterns of the fireball at the moment when the hadronization occurs to be measured.

Jet Quenching. Partons traversing the dense medium scatter and predominantly lose energy by radiating gluons. This energy loss results in a suppression of high-p_T particles, the so-called jet quenching [Gyu90]. It can be observed via a p_T dependence of R_{AA} which requires, as mentioned earlier, the p+p measurement as reference. Model comparisons allow the gluon density in the medium to be extracted from the observed jet quenching.

The back-to-back correlation that can be usually observed in two-jet events due to momentum conservation in the hard parton–parton interaction is strongly influenced by the medium. This correlation is broadened; one jet may even be completely absorbed. Jet production rates measured in p+p collisions, again, provide an essential reference here.

Dilepton and Photon Spectra. Leptons and photons are produced throughout the entire evolution of the collision. However, leptons and photons produced in the earliest and hottest phase of the collision do not interact strongly with the fireball. Therefore, they are a probe of the phase at its highest temperature. In the measurement, both of them are dominated by large backgrounds from hadronic processes, for example from pions, kaons, ρ, and π^0. The yields of dileptons and photons are compared between heavy-ion and p+p collisions to extract signals from the early collision phase.

Dileptons are a signal of medium modifications of hadronic matter, e.g. of the mass of the ρ-meson (see next paragraph). Furthermore, dileptons from charm decay allow the

total charm yield to be accessed, which is important for the measurement of the total production cross-section of the J/Ψ.

Chiral-Symmetry Restoration. The Lagrangian of QCD implies approximate chiral symmetry. As a consequence the baryon number should be conserved for right-handed and left-handed quarks separately. In nature only the total baryon number is conserved thus chiral symmetry is broken. The symmetry breaking is twofold: to start with the symmetry is only approximate due to the finite, however small, bare quark masses that cause a so-called explicit symmetry breaking. Furthermore, the quarks acquire their constituent masses in the interaction with the QCD vacuum at low T which is a spontaneous breaking of the symmetry [Pok00]. It is predicted that the spontaneous breaking of chiral-symmetry is restored at temperatures prevailing in the QGP phase. As a consequence the position and width of the masses of the light vector mesons (ρ, ω, and ϕ) may change (see e.g. [Pis82]). Indications have been seen at the SPS [Ada03, Dam07].

Further discussion of the signatures of the QGP can be found in [Won94, Har96]. A synopsis of results from the SPS is in [Hei00]. The experimental evidence from RHIC is comprehensively discussed in [Ada05].

A central aspect of the ALICE research program is to analyze heavy-ion collisions to strengthen the evidence for the existence of the QGP, to study its properties as well as the phase transition between hadronic matter and the plasma. As outlined above many of the signatures require a solid p+p reference. Therefore, the measurement of p+p collisions is crucial for the study of the QGP and the phase transition.

It should be pointed out that the measurements of p+p collisions at $\sqrt{s} = 10\,\text{TeV}$ or 14 TeV cannot be directly used as reference for Pb+Pb collisions that will be performed at $\sqrt{s_{NN}}$ of 5.5 TeV. Instead the measurements at higher energies are used to interpolate to the energy in heavy-ion collisions. Ultimately, the measurement of p+p collisions at $\sqrt{s} = 5.5\,\text{TeV}$ is the preferred reference.

1.3 High-Energy Collisions

In a high-energy collision the two colliding particles have an energy much larger than their rest mass. At the moment of the collision the transferred momentum can be very small and essentially just change the configuration of the incoming particles. The

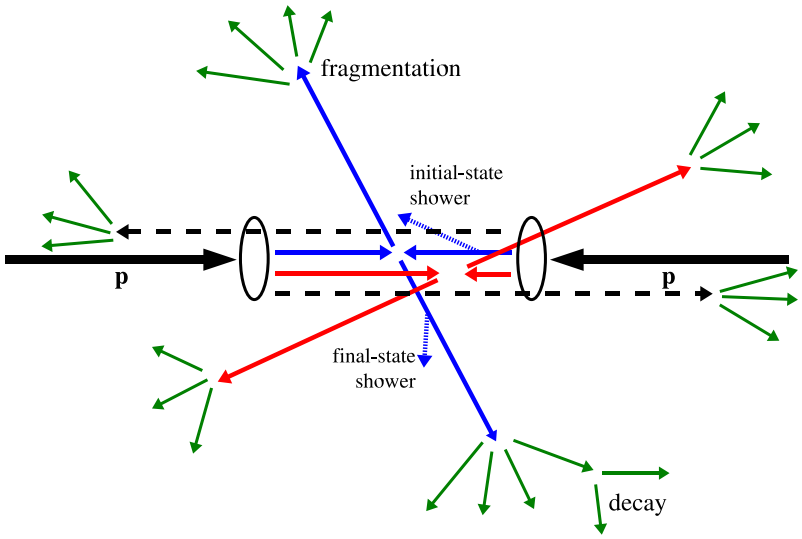

Figure 1.3: Schematic view of a high-energy collision.

transferred momentum can also be so large that the particles do not act as compound objects, instead the constituent partons participate in the collision.

Although the basic process that is to be studied might be simple in itself (e.g. a two photon decay of a Higgs particle, one of the channels the ATLAS and CMS experiments are looking for: $\mathrm{p}+\mathrm{p} \rightarrow H+X \rightarrow \gamma+\gamma+X$), the overall interaction (in the example: the 'X') is usually much more complex. Apart from the main parton interaction (of which there could be several), many other effects influence the process and the surrounding. These effects include initial- and final-state radiation, bremsstrahlung-type modifications, higher-order corrections that involve loop graphs, as well as confinement effects. The complex structure of a high-energy physics collision is illustrated in Figure 1.3. A typical event, e.g. originating from two protons, results therefore in tens to hundreds of final-state particles. The understanding of the underlying physics, i.e. multiple-particle production, is the common aim of many physics topics studied at the LHC. Specific parts of the interaction are understood very well and can be calculated precisely (e.g. parton scattering at large momentum transfer); the understanding of other effects is based on phenomenology (e.g. string fragmentation and decay) and gives rise to uncertainties. These various steps of a collision will be discussed in more detail in Section 1.5.1 using as an example the Pythia event generator.

The description of high-energy collisions within the parton model renders p+p collisions very similar to p+$\bar{\mathrm{p}}$ collisions. Therefore, in the following, p+p collisions also refer to p+$\bar{\mathrm{p}}$ collisions, unless otherwise stated.

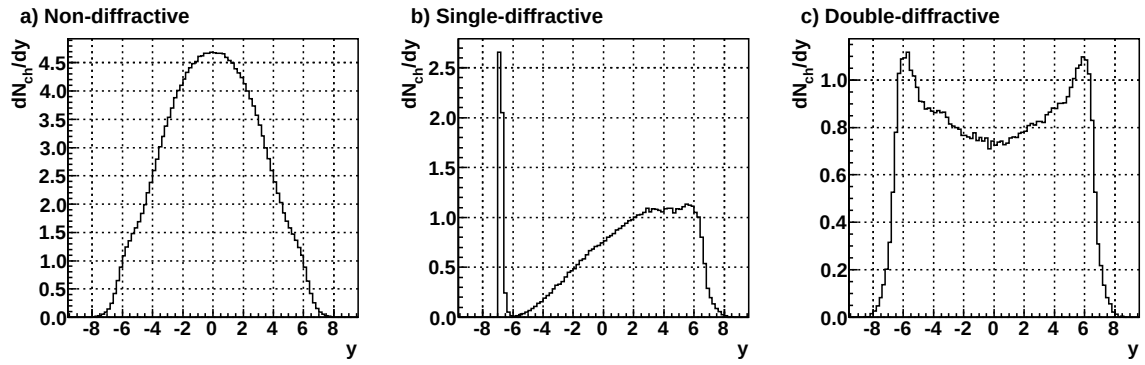

Figure 1.4: Rapidity distributions of different processes.
The rapidity y is chosen here to prevent a broadening of the structures that would appear if instead the pseudorapidity η is used. η is usually used experimentally due to its independence of the particle's mass. The distributions are obtained with Pythia at $\sqrt{s} = 900\,\mathrm{GeV}$.

1.3.1 Physics Processes

A common classification of inelastic p+p collisions is into *non-diffractive (ND)*, *single-diffractive (SD)*, and *double-diffractive (DD)* events. In diffractive events an incident particle is excited. A so-called diffractive system is created that carries the quantum numbers of the respective incoming particle (except possibly the spin). Subsequently it evolves and decays. A diffractive system is characterized by its mass M (also called *diffractive mass*). In single-diffractive events only one such system is created and the second particle remains intact; double-diffractive events feature two of them. Among the non-diffractive events are parton–parton interactions with medium to large momentum transfers of a few GeV/c.

The distribution of the particles in the final state are significantly different between these processes. The rapidity distributions of the different process types are shown in Figure 1.4 (900 GeV, Pythia). Non-diffractive collisions (left panel) have many particles in the central region, steeply falling to higher rapidities. In a single-diffractive collision only one of the beam particles breaks up and produces particles at high rapidities on one side. In the center panel only those single-diffractive collisions are shown where the particle going to positive y breaks up. The other incoming particle, nearly uninfluenced, is found at the rapidity of the beam. In a double-diffractive collision (right panel) both beam particles break up and produce particles at positive and negative high rapidities. A dip can be seen in the central region. The different scales of the three distributions should be noted. Integrating the histograms demonstrates that the

Process type	900 GeV	10 TeV	14 TeV
Non-diffractive (ND)	34.9 mb	52.1 mb	55.2 mb
Single-diffractive (SD)	11.7 mb	14.0 mb	14.3 mb
Double-diffractive (DD)	5.9 mb	9.3 mb	9.8 mb
Non single-diffractive (NSD)	40.8 mb	61.4 mb	65.0 mb
Inelastic (INEL)	52.5 mb	75.4 mb	79.3 mb

Table 1.3: Cross-sections at $\sqrt{s} = 900\,\text{GeV}$, 10 TeV, and 14 TeV (Pythia).

average total multiplicity is about a factor four higher in non-diffractive collisions than in diffractive collisions.

Integrated over many events these distributions look clearly different. Nevertheless, ALICE has only very limited capabilities to distinguish between the different process types on an event-by-event basis.[3] Thus results are presented for two combinations of the process types: for inelastic events which requires the lowest correction factors, as well as for *non single-diffractive (NSD)* events. The latter have been measured in many previous experiments because their triggering detectors used to be selective on this kind of events.

Table 1.3 shows the cross-sections of the different process types at $\sqrt{s} = 900\,\text{GeV}$, 10 TeV, and 14 TeV (Pythia). Non-diffractive collisions dominate compared to the diffractive processes.

1.4 The Quark–Gluon String Model and the Dual Parton Model

Most processes in high-energy hadronic collisions are of soft nature, the momentum transfer is small and the strong coupling constant accordingly too large to apply perturbative QCD for their description. The Quark–Gluon String Model (QGSM) [Kai03] and the Dual Parton Model (DPM) [Cap94] describe high-energy collisions by combining the topological expansion in QCD with Regge Field Theory (RFT, see e.g. [Col77]) making use of the parton structure of hadrons.

Hadron–hadron scattering amplitudes are calculated in the topological expansion by considering planar diagrams, which are associated with secondary-Reggeon exchange

[3] Trigger efficiencies for the different process types are discussed in Section 5.1.1.

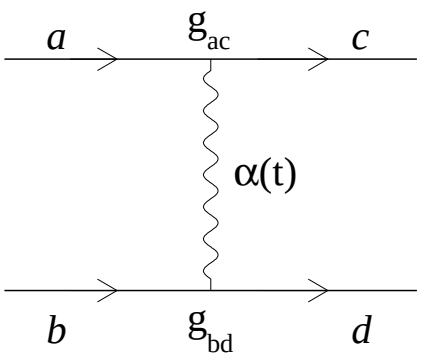

Figure 1.5: Feynman diagram representing a Regge-pole exchange.

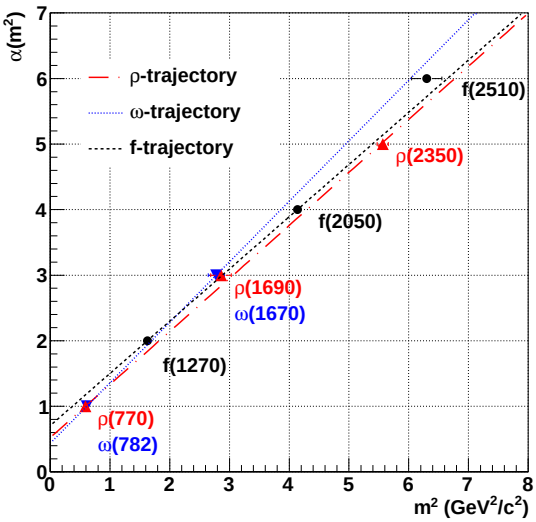

Figure 1.6: Chew-Frautschi plot showing the ρ-, ω- and f-trajectories.

in RFT, and cylinder-type diagrams, associated with Pomeron exchange in RFT (discussed below). The expansion parameter is $1/N$, where N refers to the number of colors or flavors. Therefore, the topological expansion is also called $1/N$ expansion. The expansion is dynamical in the sense that the speed of convergence depends on the kinematic region of the process under study and therefore, in general, all terms need to be taken into account.

RFT describes scattering amplitudes (in the s-channel) by singularities of the amplitude in the t channel.[4] The simplest singularity is the so-called *Regge-pole* which corresponds to the exchange of an object with 'spin' J that can be complex, the corresponding diagram is shown in Figure 1.5. J depends on the transferred momentum t and thus $J = \alpha(t)$ is defined as a so-called *Reggeon trajectory*. For values of t where $\alpha(t)$ is half-integer or integer, the object may correspond to a physical particle with mass m_J and a spin of $J = \alpha(m_J^2)$. RFT associates particles with the same trajectory that have identical quantum numbers but a difference in spin in units of 2. This can be visualized in a so-called *Chew-Frautschi plot* [Col77] showing the spin J as a function of the squared mass m_J^2. Figure 1.6 shows the ρ-, ω- and f-trajectories that are named after the particle family on the trajectory. These trajectories are parameterized in RFT by expressions of the linear form:

$$\alpha(t) = \alpha_0 + \alpha' t. \tag{1.2}$$

α_0 and α' are called *intercept* and *slope*, respectively.

[4] See Appendix A for the definition of the Mandelstam variables s and t.

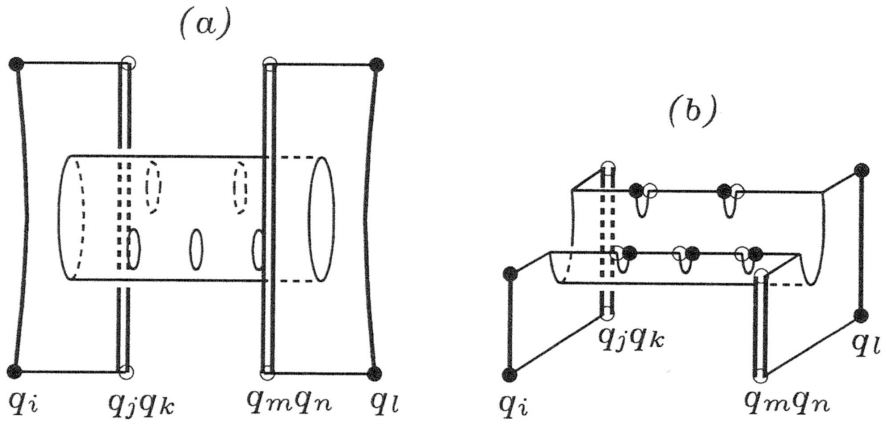

Figure 1.7: Feynman diagram for single Pomeron exchange.
The diagram is shown before (a) and after (b) cutting. For clarity, in (b) only one part after the cut is shown. q_i, q_j and q_k are the quarks of the first proton; q_l, q_m and q_n the quarks of the second. Figure taken from [Won94].

A Regge-pole exchange contributes to the scattering amplitude in the following form for large s:

$$A(s,t) \propto s^{\alpha(t)}. \tag{1.3}$$

The contribution from the Regge-pole to the total cross-section can be calculated using the optical theorem which for large s can be written as [Col77]:

$$\sigma_{\text{tot}} = \frac{1}{s}\Im(A(s,0)) \propto s^{\alpha(0)-1} = s^{\alpha_0-1}. \tag{1.4}$$

In this approximation σ_{tot} increases with s^{α_0-1}. However, for the shown Regge trajectories (Figure 1.6) α_0 is less than one, which is the case for all trajectories associated with physical particles [Col77]. To allow for the increase of the cross-section with $\sqrt{s}$ which is found experimentally, an object with $\alpha_0 > 1$ is needed, that has been named Pomeron. The Pomeron's nature in QCD is not fully clear [Kur76] and in the following it is associated with cylinder-type diagrams, shown in Figure 1.7a. Such an object, however, results in the fact that the cross-section increases following a power law. This is in contradiction with the Froissart bound [Col77] which states that for $s \to \infty$ the cross-section does not grow faster than $\ln^2 s$. This discrepancy is resolved by taking into account also multiple Pomeron exchanges (a procedure called *eikonalization*), the scattering amplitude is unitarized which yields the total cross-section proportional to $\ln^2 s$ [Ter86].

Cutting[5] the cylinder diagram associated with single Pomeron exchange, see Figure 1.7b, shows that the multiple-particle production is governed by two chains (or strings) of particles. One stretches between the quark (e.g. q_i) of the incoming particle and the diquark (e.g. $q_m q_n$) of the other incoming particle and vice versa. k cut-Pomerons thus produce $2k$ chains. The fragmentation of these chains results in multiple-particle production.

To calculate multiplicity spectra, the cross-section for k cut-Pomerons, $\sigma_k(\xi)$, as well as the distribution of particles produced by the chains, as a function of rapidity y, $f_k(\xi, y)$, and of multiplicity N, $W(\xi, N)$, are needed. ξ denotes the energy dependence that is usually expressed as $\xi = \ln(s/s_0)$ with the scale s_0 (typically set to the square of the mass of the proton to allow for a physical scale). From these pseudorapidity and multiplicity distributions can be obtained:

$$\frac{d\sigma}{dy}(\xi) = \sum_k \sigma_k(\xi)\ f_k(\xi, y), \tag{1.5}$$

$$\sigma(\xi, N) = \sum_k \sigma_k(\xi)\ W(\xi, N). \tag{1.6}$$

For the calculation of the multiplicity distributions, a distribution needs to be assumed for $W(\xi, N)$. Typically a Poisson distribution is chosen. Its parameter, the average number of particles $\langle N \rangle$, is determined from the rapidity density $d\sigma/dy$.

$$W(\xi, N) = W(\langle N \rangle(\xi), N), \qquad \langle N \rangle(\xi) = \frac{1}{\sigma_{\text{tot}}} \int \frac{d\sigma}{dy}(\xi) dy. \tag{1.7}$$

The probability for the production of a particle with rapidity y from the two chains of a single cut-Pomeron in a p+p collision, see Figure 1.7b, can be written as:

$$f_1(\xi, y) = \int dy_1 \int dy_2 w^{q_i}(\xi, y_1) w^{q_m q_n}(\xi, y_2) F(y_1 - y, y - y_2) + \text{second chain}. \tag{1.8}$$

A single chain has only a fraction of the total energy of the incoming proton: $w^{q_i}(\xi, y_1)$ is a structure function giving the probability to find a quark q_i with rapidity y_1 in the incoming proton. $w^{q_m q_n}(\xi, y_2)$ is defined analogously for the diquark. $F(y_1 - y, y - y_2)$ is the fragmentation function of the chain and depends on the rapidity difference of the produced particle and the rapidities of the quark and diquark (the ends of the chain). The term for the second chain in Eq. (1.8) is identical to the first term with q_i and $q_m q_n$ replaced by $q_j q_k$ and q_l, respectively.

[5] Unitarity allows the forward elastic scattering amplitude to be related with the interaction cross-section. An elastic reaction $i \to j$ takes place via various intermediate states $i \to n \to j$. 'Cutting' the process diagram results in $i \to n$ and $(j \to n)^*$ (the complex conjugate of $n \to j$).

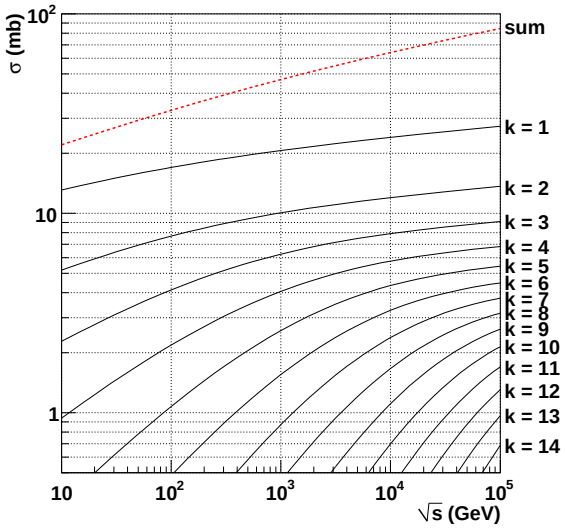

Figure 1.8: QGSM cross-sections for k cut-Pomerons.

The lines show the cross-sections for k cut-Pomerons. The top dashed line is the sum of the contributions (up to $k = 40$), that is the cross-section for ND collisions.

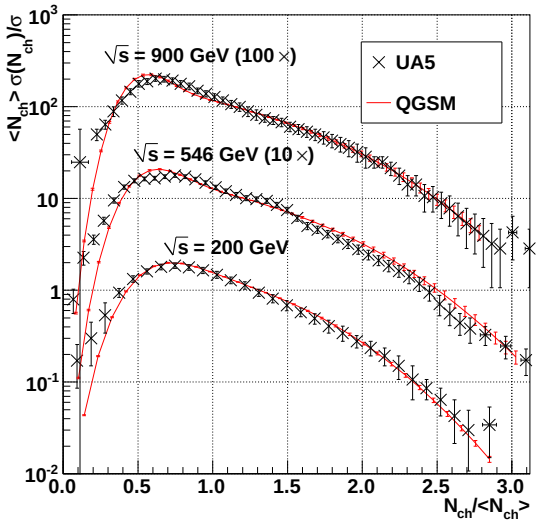

Figure 1.9: Comparison of QQGSM calculations with UA5 data.

Shown are distributions in full phase space; the different energies are scaled for visibility. Figure taken from [Pog08] (UA5 data from [Aln87, Ans89]).

In case of k cut-Pomerons, the chains of one cut-Pomeron stretch between valence quarks and diquarks, as before, the others stretch between sea quarks and diquarks which is expressed by:

$$f_k(\xi, y) = f_1^{\text{valence}}(\xi, y) + (k-1) f_1^{\text{sea}}(\xi, y) \tag{1.9}$$

where f_1^{valence} and f_1^{sea} are given by Eq. (1.8) with structure functions w corresponding to valence and sea quarks, respectively.

In the calculations of physical observables, QGSM and DPM deviate because they use different sets of diagrams as well as different parameterizations for the structure and fragmentation functions in Eq. (1.8). More details can be found in [Cap94, Kai03]. In the following some calculations using QGSM are presented.

For the calculation of the cross-section for k cut-Pomerons it is important to consider that k cut-Pomerons can be accompanied by an arbitrary number of uncut Pomerons. This cross-section can be derived with the Gribov-Regge calculus [Gri68] using the

AGK-cutting rules [Abr73]. Neglecting interactions between Pomerons in quasi-eikonal approximation it is found to be [Ter86]:

$$\sigma_k(\xi) = \frac{4\pi}{kC}(R_P^2 + \alpha'_P\xi)\left[1 - \exp(-z)\sum_{i=0}^{k-1}\frac{z^i}{i!}\right] \tag{1.10}$$

with

$$z = \frac{2C\gamma_p}{R_P^2 + \alpha'_P\xi}\exp\Delta\xi, \qquad C = 1.5. \tag{1.11}$$

$\Delta = \alpha_P(0) - 1$ and α'_P characterize the Pomeron trajectory (see Eq. (1.2)); γ_P and R_P^2 are parameters of the Pomeron residue. They are extracted from fits to data measured at ISR and Sp$\bar{\text{p}}$S [Ter86]. The values are:

$$\Delta = 0.12 \pm 0.01, \qquad \gamma_P = 2.14 \pm 0.03\,\text{GeV}^{-2}, \tag{1.12}$$

$$\alpha'_P = 0.22 \pm 0.02\,\text{GeV}^{-2}, \qquad R_P^2 = 3.30 \pm 0.02\,\text{GeV}^{-2}. \tag{1.13}$$

Figure 1.8 shows the energy dependence of cross-sections of k cut-Pomerons and the sum of the terms. The cross-section of terms with higher k falls steeply. At $\sqrt{s} = 50\,\text{GeV}$ the contribution of one cut-Pomeron dominates, while at center-of-mass energies around 1 TeV the mean is about 2. At $\sqrt{s} = 10\,\text{TeV}$ three cut-Pomerons are expected on average and the first nine terms contribute with at least 1 mb. Generally, in the energy range considered here the exchange of Pomerons dominates the multiple-particle production.

Calculated rapidity and multiplicity distributions successfully reproduce spectra measured by UA5[6], see e.g. [Kai99, Kai03]. Figure 1.9 shows the comparison of QGSM calculations with UA5 data. Depicted are multiplicity distributions in full phase space in KNO variables[7].

In summary, both models, QGSM and DPM, describe almost all available data starting from $\sqrt{s} \approx 10\,\text{GeV}$ (cross-sections and distributions) on soft hadronic high-energy interactions with only a few free parameters that are fixed by data.

1.5 Event Generators

Event generators provide simulated events that are as close as possible to real interactions as occur at the collision point. Naturally, this is limited by the present understanding of the underlying physics. Event generators combine perturbative solutions for well-understood areas and phenomenological approaches for other areas that can yet only be modeled. Generated events are used to obtain an understanding of the data and

[6]Previous experiments and their results are discussed in detail in Chapter 2.

[7]For the definition of KNO variables see Section 1.6.

signals that are to be expected, for preparing the analysis strategies and implementing the needed analysis code, as well as for estimating the needed corrections to obtain from the raw measured result the underlying *true* signal. In addition, results of event generators together with further simulation software are used to plan and optimize the detector design. Ultimately, although only to a limited extent, comparing results from event generators to events measured in an experiment allows the underlying physics to be understood. Event generators are also called *Monte Carlo (MC) event generators*, due to the fact that they make extensive use of random number generators: they 'roll the dice'.

Interactions in high-energy proton collisions can be described using a combined approach: perturbative QCD is very successful in describing parton interactions with large momentum transfer, so-called *hard interactions*. In the region of soft interactions the coupling constant α_s of the strong force approaches unity, thus the perturbative approach is not valid. Nevertheless, many of the particles produced in LHC collisions and especially in events triggered by minimum-bias triggers[8] originate from soft interactions. Furthermore, soft interactions are also present in an event that initially had a hard interaction. An approach to work around this problem is by combining perturbative QCD with a phenomenological approach that describes soft processes in the region where perturbation theory is not applicable.

The charged multiplicity, studied in this thesis, is very sensitive to the number of parton–parton scatterings. This is due to the fact that each parton interaction transfers energy from the collision system, i.e. in forward/backward longitudinal direction, to low-p_T particles in the central region. An event generator that aims at a correct description of the pseudorapidity and multiplicity distributions therefore has to estimate the correct amount and strength of partonic interactions [Mor07].

The following introduces two event generators, Pythia and Phojet, which have been used to estimate the corrections and systematic uncertainties for the measurements described in this thesis.

1.5.1 Pythia

Pythia [Sjo01] is an event generator that combines perturbative QCD and sophisticated, mostly phenomenologically motivated models. These are connected at a tunable cut-off parameter $p_{T,\mathrm{min}}$, where p_T is the momentum transfer in the hard interaction. In general, Pythia's approach results in many tunable parameters especially from phe-

[8] A minimum-bias trigger imposes the least possible bias on the triggered sample compared to 'all' available collisions. See Section 5.1 for more details.

nomenological models. These have a significant influence on the generated distributions. For example the $dN_{ch}/d\eta$ distribution is extremely sensitive to changes in the $p_{T,\min}$ parameter. A change from 1.8 GeV (the value used in ALICE's Pythia tune) to 1.7 GeV increases the $dN_{ch}/d\eta|_{\eta=0}$ by about 5% at $\sqrt{s} = 900\,\text{GeV}$. In turn $dN_{ch}/d\eta|_{\eta=0}$ is an important observable to tune especially the $p_{T,\min}$ parameter for a new energy regime.

The total p+p interaction cross-section used in Pythia is parameterized by:

$$\sigma_{\text{tot}}^{\text{pp}}(s) = 21.75 s^{0.0808} + 56.1 s^{-0.4525} \tag{1.14}$$

where the first term is arising from the aforementioned Pomeron exchange, the second from Reggeon exchange. The constants are found by fitting measured data [Don92].

The total cross-section is comprised of different partial cross-sections:

$$\sigma_{\text{tot}}^{\text{pp}} = \sigma_{\text{el}}^{\text{pp}} + \sigma_{\text{SD}}^{\text{pp}} + \sigma_{\text{DD}}^{\text{pp}} + \sigma_{\text{ND}}^{\text{pp}}. \tag{1.15}$$

The optical theorem is adopted to obtain the elastic contribution[9] $\sigma_{\text{el}}^{\text{pp}}$ and expressions calculated by Regge theory to determine the diffractive cross-sections $\sigma_{\text{SD}}^{\text{pp}}$ and $\sigma_{\text{DD}}^{\text{pp}}$. The remaining cross-section is $\sigma_{\text{ND}}^{\text{pp}}$. At LHC energies this is the largest contribution (see Table 1.3 on page 24). Pythia defines hard interactions as interactions with a momentum transfer larger than $p_{T,\min}$. Thus, by construction, all ND events are handled within the QCD expressions for hard interactions.

Due to the probabilistic nature of quantum mechanics, the first step in the event generation procedure is to randomize the process to be simulated. The selection is governed by the previously introduced cross-sections. The subsequent step depends on the kind of selected process. For the case of a hard interaction the procedure is the following:

- Two beams (specifically, two beam particles) move towards each other. Each of them consists of many partons (quarks, antiquarks, and gluons) whose distribution can be characterized by parton distribution functions (PDFs). A PDF $f_i(x, Q^2)$ describes the probability of finding a parton i with the momentum fraction x of the total momentum of the beam particle probed at a scale Q^2.[10] PDFs are parameterizations of experimental data. Several parameterizations exist; thus, the generated events also depend on the choice of the PDFs. The default setup for PDFs in Pythia is *CTEQ5L* [Lai00]. This parametrization results from a global fit to high-energy lepton–hadron and hadron–hadron collision measurements.

[9] Elastic processes are not of interest for ALICE since they cannot be measured by the experiment. They will not be further discussed.

[10] Probabilities of PDFs are given in the infinite momentum frame of the incoming particle, i.e. $E \approx |P| \gg m$.

- Partons from each of the beams may branch before the interaction (e.g. $q \to qg$), this is called initial-state shower (see also the description of final-state shower below).

- The hard interaction (e.g. $qg \to qg$ or $qg \to q\gamma$) occurs between two partons and produces outgoing partons. In this process short-lived resonances (e.g. Z^0) may appear whose decay has to be considered by the event generator. The partons that undergo the hard interaction contain only a fraction of the total beam energy. The remaining partons are called beam remnants (e.g. in a p+p collision: a u quark that took part in an interaction results in a ud diquark remnant). Possibly two further partons interact in the same collision (multiple-parton interaction).

  The choice of partons that interact with a particular momentum transfer is governed mainly[11] by the differential cross-section for $2 \to 2$ parton scattering, which in perturbative QCD is

$$\frac{d\sigma}{dp_T^2} = \sum_{i,j,k} \int dx_1 \int dx_2 \int d\hat{t} f_i(x_1, Q^2) f_j(x_2, Q^2) \frac{d\hat{\sigma}_{ij}^k}{d\hat{t}} \delta\left(p_T^2 - \frac{\hat{t}\hat{u}}{\hat{s}}\right). \tag{1.16}$$

  i and j sum over all partons available in the incoming particles. f_i and f_j are the PDFs of parton i and j, respectively. $\hat{t}$, $\hat{u}$, and $\hat{s}$ are the Mandelstam variables (see Appendix A) of the parton scattering, $d\hat{\sigma}_{ij}^k/d\hat{t}$ is the differential hard-scattering cross-section between partons i and j for the kth available subprocess between these partons. The argument Q^2 to f_i and f_j is p_T^2.

  The total interaction cross-section σ_{int} is found by integrating Eq. (1.16) starting from the mentioned parameter $p_{T,\text{min}}$. σ_{int} may become larger than σ_{ND}, which is interpreted as a multiple-parton interaction. Thus the average number of parton–parton interactions per event is directly given by the ratio of the cross-sections:

$$N_{\text{parton–parton}} = \frac{\sigma_{\text{int}}}{\sigma_{\text{ND}}}. \tag{1.17}$$

  The number of parton interactions in a given event is sampled. In the case that this procedure yields zero parton interactions, the case is treated in a special way: a very soft gluon is exchanged between the two incoming particles which changes the color configuration but essentially does not transfer momentum.

- Simultaneously to the hard process, other semi-hard processes can occur between the other partons.

[11] $2 \to N$ parton scattering with N different from 2 is also available in Pythia, but not further discussed here.

- After the interaction the outgoing partons may branch which is called final-state shower. This process becomes more important with larger energies and its realization has significant influence on the structure of jets. Pythia uses the so-called parton shower approach. Partons undergo a series of branchings, e.g. $q \to qg$, $q \to q\gamma$, and $g \to gg$. Each is described by a branching kernel $P(z)$ where z denotes the sharing of the energy and momentum between the two daughter products. These kernels are approximations of the matrix elements that describe the branching processes (in the leading-logarithmic picture used, they contain no terms in $\mathcal{O}(\alpha_s^2)$). Daughters can undergo further branching. The procedure is initiated with an energy matched to the hard interaction and terminates when the remaining energy is below a threshold of about 1 GeV.

- Strings span between the outgoing quarks and gluons that subsequently fragment to colorless hadrons due to QCD confinement. The string topologies are in principle found by decomposing the basic QCD cross-section into the definite color states. The fragmentation process (sometimes also called hadronization process) is not yet understood from first principles. Therefore, phenomenological approaches are used; Pythia implements the so-called Lund-model. An example for a fragmentation process in the Lund-model is of the type: string $\to$ hadron + remaining-string. The string is stretching e.g. between a q and a $\bar{q}$; while the quarks move apart the potential energy in the string increases due to confinement that holds them together. If the energy is large enough the string breaks producing an additional $\bar{q}'q'$-pair. The two resulting strings (q–$\bar{q}'$ and q'–$\bar{q}$) continue to fragment. If the energy of a pair is low enough, a meson is formed. In an analog way, a diquark–antidiquark-pair may be created when a string breaks (e.g. q–$\bar{q} \to q$–$q'q' + \bar{q}'\bar{q}'$–$\bar{q}$). Eventually, this results in the forming of baryons.

- Hadrons that are produced in the previous step may be instable and decay further. Therefore, Pythia contains lists of decay properties (branching ratios, decay products, life times) of relevant instable particles. Although a significant amount of experimental data on decay properties exists (see e.g. [Ams08]), the information about many particles is still incomplete, especially for charm and bottom mesons. This results in uncertainties in the event properties.

If a diffractive process was chosen for the event generation, a diffractive mass M is selected guided by the formulas that are used for the cross-section calculation. The diffractive system is then handled as a string (see above) that has the quantum numbers of the original hadron. Two different Pomeron couplings that may occur are used and mixed in equal proportions. The first coupling stretches the Pomeron between a quark

Parameter	Value	Comment
ISUB (active processes)	11–13, 28, 53, 68 92, 93 94 95	Hard QCD processes Single diffraction Double diffraction Low-p_T production
PARP(82)	1.8 GeV/c	$p_{T,\mathrm{min}}$
MSTP(82)	4	Complex scenario of multiple-parton interactions using the double Gaussian matter distribution in the hadron
PARP(84)	0.5	Setting for the double Gaussian matter distribution: the core radius is 50% of the hadronic radius

Table 1.4: Pythia parameters in the ALICE simulation. Only non-default values are listed.

and a diquark; the second between a quark, via a gluon, then to the diquark. Produced hadrons can decay further as in the case of hard interactions.

ALICE's Pythia tune

ALICE uses Pythia 6.2.14 with a tune that was developed by the ATLAS experiment, hence it is sometimes referred to as 'ATLAS tune' [Mor07]. The tune was obtained by changing the Pythia settings to match multiplicity distributions in full phase space and $dN_{ch}/d\eta$ distributions of NSD events at center-of-mass energies from 200 GeV to 1.8 TeV. Furthermore, $\langle N_{ch}\rangle$ and $\langle p_{T,\mathrm{sum}}\rangle$ in the underlying event[12] as a function of the p_T of the leading jet at $\sqrt{s} = 1.8$ TeV has been considered. Data from various experiments has been used: UA5 ($\sqrt{s} = 200$ GeV, 546 GeV, and 900 GeV), CDF ($\sqrt{s} =$ 1.8 TeV), and E735 ($\sqrt{s} = 1.8$ TeV). Details can be found in [Mor07].

The values that are different from the default Pythia settings are given in Table 1.4.

1.5.2 Phojet

The event generator Phojet [Eng95] is based on a two-component approach that describes high-energy collisions with a soft and a hard component. These are split in the

[12]The underlying event is in this context defined as the particles emitted transverse to the leading jet, i.e. $60° < |\Delta\varphi| < 120°$, $\Delta\varphi$ is the azimuthal angle from the leading jet.

calculation at a $p_{T,\text{cut-off}}$ parameter and their results are combined by a unitarization procedure. The ideas of the Dual Parton Model are employed for the soft component (see Section 1.4). The hard component is calculated by perturbative QCD like in Pythia.

Phojet calculates the total cross-section as well as the cross-sections for different processes using unitarized scattering amplitudes. These are derived using Regge arguments in the soft region and perturbative QCD in the hard region. Only considering first order graphs the cross-section can be written as:

$$\sigma_{\text{tot}} = \sigma_{\text{R}} + \sigma_{\text{P}_\text{s}} + \sigma_{\text{hard}} \tag{1.18}$$

with the cross-sections for Reggeon exchange σ_{R}, soft Pomeron exchange $\sigma_{\text{P}_\text{s}}$, and the hard component σ_{hard}. In the color-flow picture the hard cross-section is identified with the hard part of the single Pomeron exchange cross-section which allows a unified treatment. Multiple-Pomeron exchanges are taken into account to prevent deviation of the total cross-section (eikonalization, see also Section 1.4). For details see [Eng97].

Phojet samples the number of soft and hard interactions in an event governed by the relative contributions of the amplitudes:

$$\sigma(n_s, n_h, s) = \int d^2B \frac{(2\chi_S)^{n_s}}{n_s!} \frac{(2\chi_H)^{n_h}}{n_h!} \exp\left[-2(\chi_S + \chi_H)\right]. \tag{1.19}$$

$\chi_S(s, \vec{B})$ and $\chi_H(s, \vec{B})$ are eikonal functions in impact parameter representation ($\vec{B}$) that are a function of the amplitudes for the soft and hard part, respectively.

The soft interactions are then distributed among the various soft processes (single or multiple Pomeron exchange and diffraction). Due to this approach a diffractive and a hard interaction can be present in the same event. Also multiple-parton interactions are taken into account by several hard interactions.

In ND events the momentum transfer is sampled below the $p_{T,\text{cut-off}}$ from an exponential distribution, above $p_{T,\text{cut-off}}$ from the perturbative QCD cross-section. It is required that the transition between the soft and hard region is continuous which fixes the slope of the mentioned exponential distribution. In a diffractive event the diffractive mass and the momentum transfer are sampled from cross-sections derived with the triple-Pomeron approximation [Eng97]. The multiple-particle production is generated by simulating Pomeron-proton and Pomeron-Pomeron interaction with $\sqrt{s}$ equal to the sampled diffractive mass. Due to this treatment also hard interactions can occur between the Pomeron and the proton. This is different from the Pythia approach for diffractive events. The fragmentation process of the chains obtained by the cutting of the Pomerons as well as the hard scattered partons is treated within the Lund-model that is also in use by Pythia and has been described above.

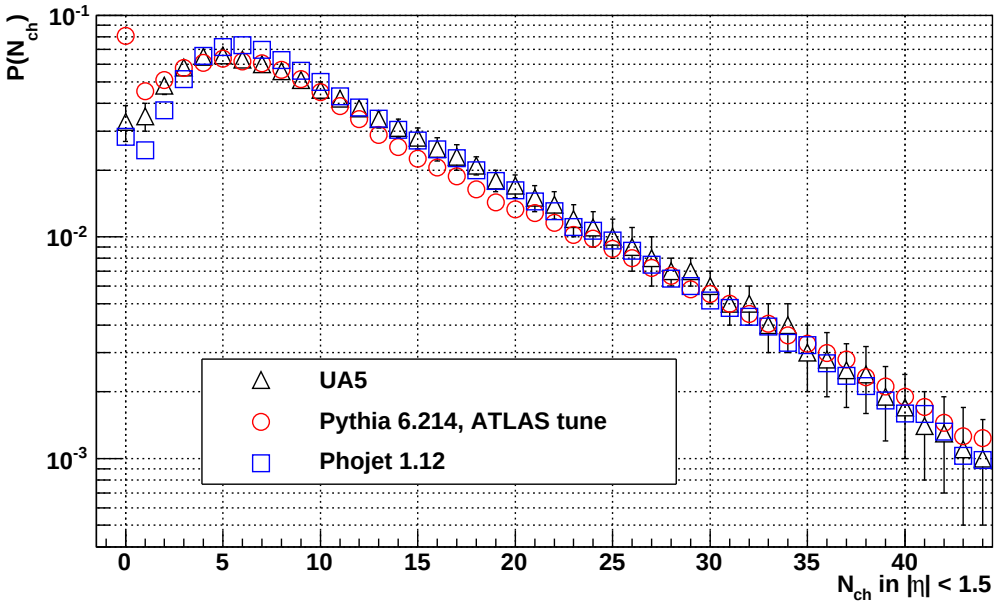

Figure 1.10: Comparison of Pythia and Phojet with UA5 data (data points from [Ans89]) for NSD events at $\sqrt{s} = 900\,\text{GeV}$.

Only a few parameters are adjustable in Phojet, most importantly the $p_{T,\text{cut-off}}$ and parameters that describe the couplings of the proton to the Pomeron and Reggeon. These are determined by fits to experimental data. These parameters cannot be changed individually without adjustment of the other parameters. The individual soft and hard cross-sections are dependent on $p_{T,\text{cut-off}}$, however, their sum is almost independent of $p_{T,\text{cut-off}}$ [Eng97]. Phojet includes a process called central diffraction ($\sigma \sim 1\,\text{mb}$), a process with double Pomeron exchange, that is not included in Pythia.

1.5.3 Comparison of Pythia and Phojet Predictions with UA5 Data

The simulation results of Pythia and Phojet have been verified by comparing their results to multiplicity measurements performed at $\sqrt{s} = 900\,\text{GeV}$ by UA5. Exemplarily, the comparison of the multiplicity distribution is shown in Figure 1.10.

Pythia agrees with UA5 except for the bin with 0 tracks in the region considered. The majority of events in this bin are of diffractive type. The χ^2/ndf is $34.3/45 = 0.76$ without the first bin ($85.9/46 = 1.87$ including the first bin).

Phojet agrees slightly better with UA5, but also exhibits discrepancies in the low-multiplicity bins (0 – 2). The deviation in the '0-bin' is, however, lower. The χ^2/ndf is $30.6/46 = 0.67$ (same χ^2/ndf without the first bin).

1.6 Charged-Particle Multiplicity

The charged-particle multiplicity, being one of the basic properties of high-energy collisions, has been studied by many previous experiments at various energies. Usually it was among the first publications, but has been studied again at a later stage, when higher statistics and a more thorough understanding of the detector was available.

Observables are the pseudorapidity density, i.e. the charged-particle multiplicity as a function of pseudorapidity $dN_{ch}/d\eta$ vs. η (or as a function of rapidity y) and the charged-particle multiplicity distribution $P(N_{ch})$ vs. N_{ch}, usually in a limited η-range, depending on the acceptance of the detector. Extrapolations to full phase space can be found in publications but these extrapolations are usually model-dependent due to the limited acceptance, especially of modern detectors. To study the scaling behavior $P(N_{ch})$ is expressed as $\langle N_{ch}\rangle P(z)$ vs. z with $z = N_{ch}/\langle N_{ch}\rangle$ (*KNO variables*, see Section 1.6.2). The energy dependence is studied by investigating the behavior of $dN_{ch}/d\eta$ at $\eta = 0$ as a function of $\sqrt{s}$ and the average multiplicity in full or limited phase space $\langle N_{ch}\rangle$ vs. $\sqrt{s}$.

The following gives an introduction to the theoretical concepts and models that describe the charged-particle multiplicity in p+p collisions. These partly fail to explain data at higher energies. In the subsequent chapter these models are applied to measurements that have been performed at energies between $\sqrt{s} = 6\,\text{GeV}$ and $1.8\,\text{TeV}$. The following concepts are introduced:

Feynman postulated in 1969 that the invariant cross-section can be written utilizing a scaling function that is independent of the collision energy, so-called *Feynman scaling*. As a consequence, he concluded that the mean total number of any kind of particle is proportional to $\ln\sqrt{s}$.

Based on Feynman scaling Koba, Nielsen, and Olesen derived theoretically in 1972 that multiplicity distributions should follow so-called *KNO scaling*. Deviations are observed for inelastic events at center-of-mass energies above about 30 GeV, for NSD events starting from 200 GeV.

In 1985, it was found by UA5 that the multiplicity distribution at $\sqrt{s} = 540\,\text{GeV}$ can be well described by a *Negative Binomial Distribution (NBD)*. Keeping a parameter of the NBD fixed implies KNO scaling (see discussion in Section 1.6.3).

Deviations of data from the NBD were discovered at $\sqrt{s} = 900\,\text{GeV}$ which led to a *two-component model* by Giovannini and Ugoccioni in 1999 who described the measured data by the combination of two NBDs, interpreting one as a soft and one as a semi-hard component. An alternative description interprets the results in favor of multiple-parton interactions which become more important at higher energies. The superposition of several interactions has influence on the multiplicity distribution and therefore explains the deviation from the scaling found at lower energies.

In case the underlying production process can be described by uncorrelated emission, i.e. the production of an additional particle is independent from the already produced particles, the multiplicity distribution is expected to be of Poissonian form. Any difference to this, indicates correlations between the produced particles. Forward–backward correlations have in fact been measured by UA5 in p+$\bar{\text{p}}$ collisions (e.g. [Alp83]) but are not further discussed here.

1.6.1 Feynman Scaling

Feynman derived that the mean total number of any kind of particle rises logarithmically with $\sqrt{s}$ [Fey69]. His argument is based on the fact that in two-body reactions exchanges of quantum numbers occur, e.g. an exchange of isospin. In the exchange, the current (here: of isospin) must reverse from the direction of one particle ($+z$) to the direction of the other ($-z$) with z being the beam axis. Currents are based on fields as sources, and these fields radiate during the current change, similar to the case of bremsstrahlung. Going to higher energies $W = \sqrt{s}/2$, the fields get narrower in z, leading to a distribution close to a δ-function in z. In turn the field's energy, dp_z, is evenly distributed, as can be found by Fourier transformation. The number of particles with a given energy E is thus distributed like:

$$\frac{dp_z}{E}. \tag{1.20}$$

Feynman extended this to the probability of finding a particle of kind i with mass m and transverse and longitudinal momentum p_T and p_z:

$$f_i(p_T, p_z/W)\frac{dp_z}{E}d^2p_T \tag{1.21}$$

with the energy of the particle:

$$E = \sqrt{m^2 + p_T^2 + p_z^2}. \tag{1.22}$$

$f_i(p_T, x = p_z/W)$ is a structure function and Feynman's hypothesis is that f_i is independent of W. This assumption is the *Feynman scaling* and f_i is called scaling function

or Feynman function. $x = p_z/W$ is the fraction of the longitudinal momentum of the particle p_z and the total energy of the incident particle W. x is now called *Feynman-x*.

In his paper [Fey69], Feynman concludes that the mean number of particles rises logarithmically, but does not give a mathematic proof. However, one can assess the asymptotic behavior by rewriting Eq. (1.21) in the form of the invariant cross-section[13]:

$$\frac{1}{\sigma} E \frac{d^3\sigma}{dp_z d^2 p_T} = f_i(p_T, x). \tag{1.23}$$

f_i factorizes approximately (found experimentally) and a normalization of g_i is chosen such that

$$\int f_i(p_T, x) d^2 p_T = f_i(x) \underbrace{\int g_i(p_T) d^2 p_T}_{=1} = f_i(x). \tag{1.24}$$

Integration of Eq. (1.23) and application of Eq. (1.24) yields:

$$\int \frac{1}{\sigma} E \frac{d^3\sigma}{dp_z d^2 p_T} \frac{d^3 p}{E} = \langle N \rangle = \int f_i(p_T, x) \frac{d^3 p}{E} = \int f_i(x) \frac{dpz}{\sqrt{W^2 x^2 + m_T^2}} \tag{1.25}$$

where on the left side the definition of the invariant cross-section is used with the average particle multiplicity $\langle N \rangle$, and for m_T an effective average-p_T is used.

Rewriting in x yields the expression used to prove Feynman's hypothesis:

$$\langle N \rangle = \int_{-1}^{1} f_i(x) \frac{dx}{\sqrt{x^2 + \frac{m_T^2}{W^2}}}. \tag{1.26}$$

The integral is symmetric because $f_i(x)$ is symmetric for collisions of identical particles. For other collision systems the integration can be performed separately for negative and positive x and yields the same result. Partial integration results in:

$$2 f_i(x) \ln \left(x + \sqrt{x^2 + \frac{m_T^2}{W^2}} \right) \Bigg|_0^1 - 2 \int_0^1 \frac{\partial f_i(x)}{\partial x} \ln \left(x + \sqrt{x^2 + \frac{m_T^2}{W^2}} \right) dx. \tag{1.27}$$

The first term is:

$$= 2 f_i(1) \ln \left(1 + \sqrt{1 + \frac{m_T^2}{W^2}} \right) - 2 f_i(0) \ln \frac{m_T}{W}. \tag{1.28}$$

[13] The definition of the Feynman function is different in some publications (e.g. [Loh05]), not considering the $1/\sigma$ term in Eq. (1.23). This approach, however, results in conclusions that are not confirmed by experiment. In detail compared to the results of the calculation presented in the following, the left sides of Eqs. (1.36) and (1.37) have to be multiplied by σ.

The asymptotic behavior can be assessed with

$$\lim_{W\to\infty} \ln\left(1+\sqrt{1+\frac{m_T^2}{W^2}}\right) = \ln 2 = const. \tag{1.29}$$

$$-2f_i(0)\ln\frac{m_T}{W} = 2f_i(0)\left(\ln W - \ln m_T\right) \tag{1.30}$$

$$\ln m_T = const. \qquad f_i(0) = const. \tag{1.31}$$

and the requirement that $f_i(0) > 0$. For small x Feynman assumes that $f_i(x) \to B$ with a small and positive B, but for the given argument it is sufficient that this is reached for $x = 0$.

In summary, for large W the first term of Eq. (1.27) is proportional to:

$$\ln W. \tag{1.32}$$

The second term of Eq. (1.27) converges for any W (also $W \to \infty$), which is shown in the following. $f_i(x)$ is finite and bounded due to energy conservation, the same is assumed for the derivative: $\partial f_i(x)/\partial x < A$ for $0 \le x \le 1$. The integral is split into two parts:

$$\int_0^1 \frac{\partial f_i(x)}{\partial x} \underbrace{\ln\left(x+\sqrt{x^2+\frac{m_T^2}{W^2}}\right)}_{=a(x)} dx$$

$$= -\int_0^\xi \frac{\partial f_i(x)}{\partial x}(-a(x))dx + \int_\xi^1 \frac{\partial f_i(x)}{\partial x}a(x)dx \tag{1.33}$$

with ξ defined by $a(\xi) = 0$. In this way both terms involving $a(x)$ are positive ($-a(x) \ge 0$ for $0 \le x \le \xi$ and $a(x) \ge 0$ for $\xi \le x \le 1$) and it can be estimated:

$$\int_0^\xi \frac{\partial f_i(x)}{\partial x}(-a(x))dx < \int_0^\xi A(-a(x))dx \tag{1.34}$$

and

$$\int_\xi^1 \frac{\partial f_i(x)}{\partial x}a(x)dx < \int_\xi^1 Aa(x)dx. \tag{1.35}$$

$\int a(x)dx$ exists, both integrals can be calculated. For $W \to \infty$, Eq. (1.33) evaluates to $\ln 2$. Thus, the second term has no influence on the asymptotic behavior.

In consequence, Feynman scaling implies that the average total multiplicity scales as

$$\langle N\rangle \propto \ln W \propto \ln\sqrt{s}. \tag{1.36}$$

Considering that the maximum reachable rapidity in a collisions increases also with $\ln\sqrt{s}$, and under the further assumption that the particles are evenly distributed in rapidity, it follows that:

$$\frac{dN}{dy} = const. \tag{1.37}$$

The same is obtained, however under less general assumptions, by directly writing Eq. (1.23) with Eq. (1.24) in the rapidity variable y ($Wdx = dp_z = Edy$). Feynman's assumption is that $f_i(x) = B = const.$ for small x. For the region of small x (where the bulk of particles is found) the following expression is obtained:

$$\frac{1}{\sigma}d\sigma = f_i(x)\frac{Wdx}{\sqrt{W^2x^2 + m_T^2}} \longrightarrow Bdy \tag{1.38}$$

which is the same as

$$\frac{1}{\sigma}\frac{d\sigma}{dy} = \frac{dN}{dy} = B = const. \tag{1.39}$$

The height of the rapidity distribution around mid-rapidity, the so-called plateau, is independent of $\sqrt{s}$. Equivalently, the pseudorapidity at mid-rapidity $dN/d\eta|_{\eta=0}$ is approximately constant when Feynman scaling applies. Here the transformation from y to η has to be taken into account. It depends on the average m_T which, however, is only weakly energy-dependent (a rough estimate of the change in the transformation factor is 1 – 2% from $\sqrt{s} = 100\,\text{GeV}$ to $1\,\text{TeV}$).

1.6.2 Koba–Nielsen–Olesen (KNO) Scaling

KNO scaling was suggested in 1972 by Koba, Nielsen, and Olesen [Kob72]. Their main assumption is Feynman scaling.

KNO scaling is derived by calculating the expression

$$\langle n(n-1)...(n-q-1)\rangle = \int f^{(q)}(x_1, p_{T,1}; ...; x_q, p_{T,q})\frac{dp_{z,1}}{E_1}dp_{T,1}^2...\frac{dp_{z,q}}{E_q}dp_{T,q}^2 \tag{1.40}$$

which is similar to Eq. (1.26) but for a q-dimensional Feynman scaling function $f^{(q)}$ (q particles with energy E_q, longitudinal momentum $p_{z,q}$, transverse momentum $p_{T,q}$, and Feynman-x x_q). Integration by parts is performed for all x_i and it is proven that the resulting function is uniquely defined by moments. This yields a polynomial in $\ln s$. With a substitution of the form $\langle n\rangle \propto \ln s$ the multiplicity distribution $P(n)$ is found to scale as

$$P(n) = \frac{1}{\langle n\rangle}\Psi(\frac{n}{\langle n\rangle}) + \mathcal{O}\left(\frac{1}{\langle n\rangle^2}\right) \tag{1.41}$$

where the first term results from the leading term in $\ln s$, that is $(\ln s)^q$. The second term contains all other terms in $\ln s$, i.e. $(\ln s)^{q'}$ for $q' < q$. $\Psi(z := n/\langle n\rangle)$ is a universal function and energy-independent. This means that multiplicity distributions at all energies fall onto one curve when plotted as a function of z. However, $\Psi(z)$ can be different depending on the type of reaction and the type of measured particles.

The moments c_q define $\Psi(z)$ uniquely[14] [Kob72]:

$$c_q = \int_0^\infty z^q \Psi(z) dz. \tag{1.42}$$

Substituting $z = n/\langle n\rangle$ results in

$$c_q = \langle n^q\rangle / \langle n\rangle^q. \tag{1.43}$$

Studying the moments of the distribution shows if the scaling hypothesis holds; in this case the moments are independent of energy. For example an experimentally accessible possibility is to calculate the standard deviation $\sigma^2 = \langle n^2\rangle - \langle n\rangle^2$; the relation $\sigma/\langle n\rangle = const.$ follows from Eq. (1.41) (if $\Psi(z)$ is not a δ-function, see [Kob72]).

1.6.3 Negative Binomial Distributions

The *Negative Binomial Distribution (NBD)* is defined as

$$P(n; p; k) = \left(\begin{array}{c} n+k-1 \\ n \end{array} \right) (1-p)^n \; p^k. \tag{1.44}$$

It gives the probability for n failures and $k-1$ successes in any order in the first $k+n-1$ trials and a success in the last, $k+n$th, trial of a Bernoulli experiment with a success probability p. The NBD is a Poisson distribution for $k^{-1} \to 0$ and a geometrical distribution for $k = 1$.

Multiplicity distributions have been found to follow NBDs with $p^{-1} = 1 + \langle n\rangle/k$, where $\langle n\rangle$ is the average multiplicity and k a parameter having influence on the shape [Aln85, Aln86a]:

$$P(n; \langle n\rangle; k) = \left(\begin{array}{c} n+k-1 \\ n \end{array} \right) \left(\frac{\langle n\rangle/k}{1+\langle n\rangle/k} \right)^n \frac{1}{(1+\langle n\rangle/k)^k}. \tag{1.45}$$

[14] Some authors (e.g. [Zaj86]) point out that the conclusion that the multiplicity distribution follows a universal function is only an approximation (neglecting the second term in Eq. (1.41)). The exact result is that the factorial moments $(\langle n(n-1)...(n-q-1)\rangle/\langle n\rangle^q)$ are required to be constant, not the reduced moments in Eq. (1.43).

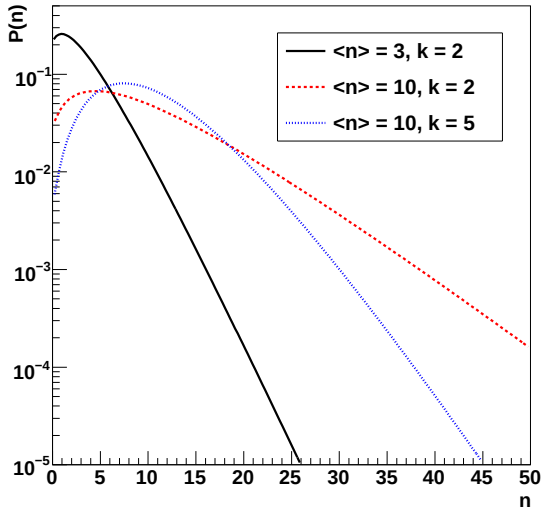

Figure 1.11: Examples of negative binomial distributions.

Figure 1.11 shows normalized NBDs with three different sets of parameters. $P(n;\langle n\rangle;k)$ follows KNO scaling if k is constant (energy-independent). Therefore, studying k as a function of $\sqrt{s}$, of multiplicity distributions that are described by NBDs, directly shows whether KNO scaling is fulfilled.

The physical motivation of a multiplicity distribution following a negative binomial shape has not been ultimately understood. One approach is to use the recurrence relation of collisions of multiplicities n and $n+1$ [Gio86]. The particles are in principle distinguishable, e.g. by their momenta, therefore it has to be taken into account that a collision of multiplicity $n+1$ can be related to $n+1$ collisions of multiplicity n (by removing any single one of the $n+1$ particles). Thus a $g(n)$ is defined by:

$$g(n) = \frac{(n+1)P(n+1)}{P(n)}. \tag{1.46}$$

Evaluating $g(n)$ for a Poisson distribution $P(n) = \lambda^n e^{-\lambda}/n!$, yields that $g(n) = \lambda = const$. The recurrence relation $g(n)$ is independent of n which means that the production of an additional particle is independent of the number of already present particles. This independent particle emission is expected for the Poisson distribution.

For NBDs, Eq. (1.46) can be written as

$$g(n) = a + bn \tag{1.47}$$

with

$$k = a/b \text{ and } \langle n\rangle = a/(1-b). \tag{1.48}$$

A model of partially stimulated emission identifies a in Eq. (1.47) with the production of particles which is independent of the already present particles and bn with emission that is enhanced by already present particles (Bose–Einstein interference). Following these rather simple assumptions results in two facts that are found experimentally: 1) k increases when the considered η-interval is enlarged; 2) k decreases with increasing $\sqrt{s}$ [Gio86].

The multiplicity distribution can be deduced as being of negative binomial shape within the so-called *clan model* [Gio86, Gio88]. It describes the underlying production by cascades of particles. In the clan model a particle can emit additional particles, e.g. by decay and fragmentation. A clan (or cluster) contains all particles that stem from the same ancestor or from its offspring. The ancestors are produced independently.

The production of ancestors, and thus clans, is governed by a Poisson distribution. The probability to produce N clans is given by $P(N, \langle N \rangle)$ with the parameter $\langle N \rangle$ being the average number of produced clans.

The probability to produce n_c particles in one clan $F_c(n_c)$ can be derived by defining that without particles there is no clan:

$$F_c(0) = 0 \tag{1.49}$$

and assuming that the production of an additional particle in a clan is proportional to the number of already existing particles with some probability p (see also Eq. (1.46)):

$$\frac{(n_c + 1)F_c(n_c + 1)}{F_c(n_c)} = pn_c. \tag{1.50}$$

By iteration, the following expression is obtained:

$$F_c(n_c) = F_c(1)\frac{p^{n_c - 1}}{n_c}. \tag{1.51}$$

The multiplicity distribution that takes into account the distribution of clans and the distribution of particles among the different clans is:

$$P(n) = \sum_{N=1}^{n} P(N, \langle N \rangle) \sum^{*} F_c(n_1)F_c(n_2)...F_c(n_N), \tag{1.52}$$

where $\sum^{*}$ runs over all combinations n_i for which $n = \sum_{i=1}^{N} n_i$ is valid. It can be shown that Eq. (1.52) is an NBD, identifying $\langle n \rangle = \langle N \rangle F_c(1)/(1 - p)$ and $k = \langle N \rangle F_c(1)/p$ [Gio86].

1.6.4 Two-Component Approaches

Combination of two NBDs

Multiplicity distributions measured by UA5 have been successfully fitted with a combination of two NBD-shaped components [Fug89]. A systematic investigation has been performed by Giovannini and Ugoccioni who interpret the two components as a soft and a semi-hard one [Gio99a]. These can be understood as events with and without minijets[15], respectively: the fraction of semi-hard events found corresponds to the fraction of events with minijets found by UA1. It is important that this approach combines two *classes of events*, not two *different particle-production mechanisms*. Therefore, no interference terms have to be considered and the final distribution is the sum of the two independent distributions.

In this approach, the multiplicity distribution depends on five parameters, that may all be $\sqrt{s}$-dependent:

$$P(n) = \alpha_{\text{soft}} \times P(n; \langle n \rangle_{\text{soft}}; k_{\text{soft}}) + (1 - \alpha_{\text{soft}}) \times P(n; \langle n \rangle_{\text{semi-hard}}; k_{\text{semi-hard}}). \quad (1.53)$$

The parameters and their dependence on $\sqrt{s}$ are found by fitting data from experimental measurements. The authors of [Gio99a] use data from UA5 taken at $\sqrt{s} = 200\,\text{GeV}$, 546 GeV, and 900 GeV in full phase space and yield ($\sqrt{s}$ in units of 1 GeV):

$$\langle n \rangle_{\text{soft}} = -5.54 + 4.72 \ln(\sqrt{s}), \quad (1.54)$$

$$\langle n \rangle_{\text{semi-hard}} \approx 2\langle n \rangle_{\text{soft}} \, [1 + 0.1 \ln^2(\sqrt{s})]. \quad (1.55)$$

Note that $\langle n \rangle$ is about two times larger in the semi-hard component than in the soft component. The second ($\ln^2$-dependent) term in Eq. (1.55) is suggested by data, but optional in the two-component approach. Furthermore, the fits show that the soft component follows KNO scaling ($k_{\text{soft}} \approx 7 = const.$), while the semi-hard component violates KNO scaling. Extrapolation of these fit results obtains predictions for higher energies which is discussed in Section 8.3.

A third, hard, component is proposed for large $\sqrt{s}$, where $\langle n \rangle$ is 3 – 10 times the average total multiplicity [Gio03]. This aspect still awaits experimental verification.

Interpretation in the Framework of Multiple-Parton Interactions

A different approach to identify a second component is by plotting the multiplicity distribution in KNO variables and subtracting the part of the distribution for which

[15]The authors of [Gio99a] use a definition from the UA1 collaboration: a minijet is a group of particles having a total transverse energy larger than 5 GeV.

KNO scaling holds [Ale98]. This is done by comparing the distribution to a KNO fit that is valid at ISR energies. Due to the large errors in the low-multiplicity bins, $\langle n \rangle$ cannot be satisfactorily determined. Therefore, it is found by using the empirical relation $\langle n \rangle \approx 1.25 n_{\max}$ which is inferred from the KNO fit at ISR energies. The authors find an interesting feature when the part that follows the KNO fit is subtracted and the remaining part plotted (not shown here). The remaining part does not follow KNO, its most probable value is 2, and its width is about $\sqrt{2}$ times the width of the KNO distribution. This procedure to identify the second component is similar to the one described in the previous section. The fact that the most probable value of the remaining part is 2 is reflected in the factor 2 in Eq. (1.55).

The authors conclude that the second part of the distribution is the result of two independent parton–parton interactions within the same collision. The cross-sections of the two contributions (σ_1, σ_2) can be calculated as a function of $\sqrt{s}$ (data between 200 GeV and 1.8 TeV is used). It is found that σ_1 is almost independent of $\sqrt{s}$, while σ_2 increases with $\sqrt{s}$. However, it is unclear if two parton–parton interactions in the same collision evolve independently to their final multiplicity due to final-state interactions.

The same reasoning and data is used in [Wal04] to identify a third component, three independent parton–parton interactions. In the framework of their calculations the authors extrapolate that the multiple-parton component (second and third component) starts to contribute to the measured distributions at an energy of $\sqrt{s} = 120$ GeV. A prediction based on this approach is discussed in Section 8.4.

Chapter 2

Multiplicity Measurements at Energies Below the LHC Energy

This chapter presents p+p($\bar{\text{p}}$) measurements that have been performed by experiments at hadron colliders, i.e. the ISR, Sp$\bar{\text{p}}$S, and Tevatron. The Intersecting Storage Rings (ISR), the very first hadron collider, was operating at CERN between 1971 and 1984. It collided p on p, $\bar{\text{p}}$, and α at a maximum center-of-mass energy of 63 GeV. The Super Proton Synchrotron (SPS) which has operated at CERN since 1976 has accelerated in its lifetime electrons, positrons, protons, anti-protons, and ions. After modification to a collider, it provided p on $\bar{\text{p}}$ collisions with a maximum $\sqrt{s}$ of 900 GeV, at that time it was called Sp$\bar{\text{p}}$S. The Tevatron at the Fermi National Accelerator Laboratory (FNAL) came into operation in 1983. It provides p+$\bar{\text{p}}$ collisions at energies up to $\sqrt{s} = 1.96$ TeV. In addition, results from bubble chamber experiments are included in the summary plots.

In the following, experiments at these accelerators that measured the charged-particle multiplicity are briefly introduced, their analysis methods and error treatments are discussed, and the results are presented. The validity of the theoretical descriptions given in the previous chapter (Section 1.6) are discussed. The chapter concludes with a summary that shows the dependence of the multiplicity on the collision energy. Note that the $dN_{ch}/d\eta$ measurements that can all be presented in the same figure are shown only in the summary section to avoid redundant plots.

The detectors described in this chapter have full azimuthal coverage unless otherwise stated.

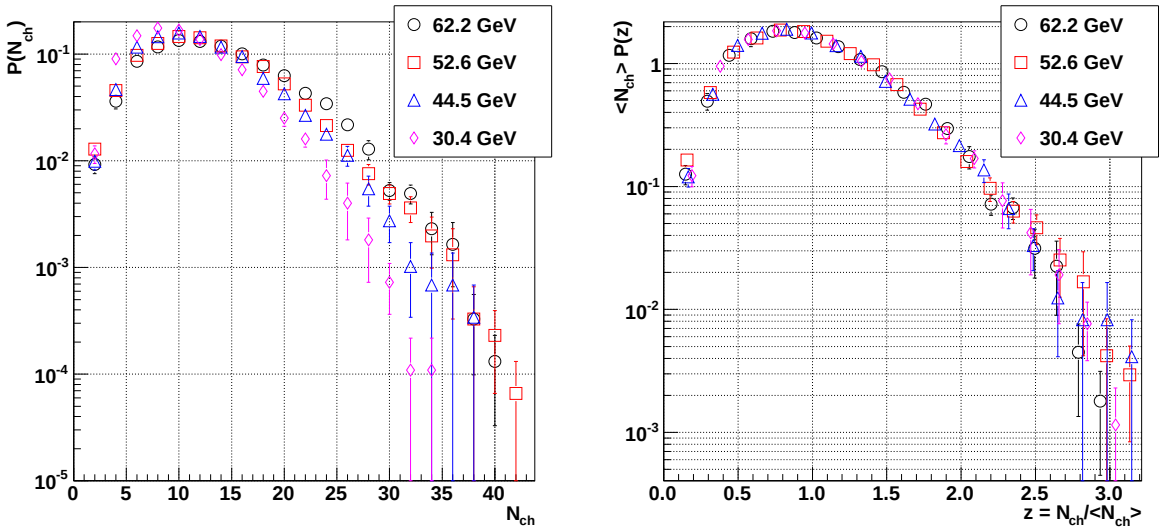

Figure 2.1: KNO scaling at ISR energies.
The figure shows normalized multiplicity distributions for NSD events in full phase space vs. multiplicity (left panel) and using KNO variables (right panel). The data was measured by the Split Field Magnet Detector [Bre84].

2.1 ISR

The **Split Field Magnet detector** at the ISR consisted of two symmetric sections on either side of the collision point. The two sections had opposite magnetic fields and their main sensitive elements were 40 multi-wire proportional chambers. The intersection region was completely surrounded giving the detector large phase space acceptance. More information about the setup can be found in [Del77].

The detector measured the multiplicity distribution for NSD and inelastic p+p events at $\sqrt{s} = 30.4$, 44.5, 52.6, and 62.2 GeV [Bre84]. The trigger required a coincidence of at least three chambers pointing to the same direction. It accepted about 95% of the inelastic events. Between 26 000 and 60 000 events were collected for each of the energies and corrected for decays of strange and neutral particles. The SD component was removed from the sample by means of its topology: events are considered SD when in one of the hemispheres no track or only one track carrying 80% of the incident proton's energy is found. The multiplicity spectrum was unfolded using χ^2-minimization. It is not mentioned in [Bre84] that a regularization procedure was used, which is surprising because χ^2-minimization without regularization usually fails to produce a unique nor a correct solution (χ^2-minimization and regularization will be discussed in detail in Section 7.2.2).

Figure 2.1 shows multiplicity distributions in full phase space for NSD events, the errors combine statistical and systematic uncertainties. The systematic errors include the error that arises from the corrections and in the low-multiplicity region from the subtraction of elastic events. By analyzing the moments of the distribution it is shown that KNO scaling holds for the NSD event sample (see also the right panel of Figure 2.1), but is broken for the inelastic event sample [Bre84].

A detector based on streamer chambers [Tho77] at the ISR with a limited acceptance of $|\eta| < 3.5$ measured pseudorapidity and multiplicity distributions for inelastic events at center-of-mass energies of 23.6 – 62.8 GeV. Between 2 300 and 5 900 events were measured for each energy. In the analysis corrections for the acceptance, the low-momentum cut-off (about 45 MeV/c), and secondary particles (called secondaries in the following) due to interactions with the material are taken into account. Contributions from γ-conversions, π^0, and strange-particle decays are subtracted. Results are included in Figures 2.8 (page 57) and 2.9. It was found that in a limited region of $|\eta| < 1.5$, KNO scaling is confirmed. The authors of [Tho77] extrapolate their result to full phase space and analyze the moments of the distribution. They find that KNO scaling is broken for inelastic events, consistent with the result presented above.

The UA5 (Underground Area 5) experiment, designed for the Sp$\bar{\text{p}}$S and therefore described in the subsequent section, has operated also at the ISR. A comparison of data taken in p+p and p+$\bar{\text{p}}$ collisions at $\sqrt{s} = 53$ GeV was made [Alp82]. 3 600 p+p events and 4 000 p+$\bar{\text{p}}$ events were used. The analysis corrects for effects of decays, γ-conversions, and secondaries. Trigger and vertex finding efficiencies as well as acceptance effects have been evaluated with a MC simulation that was tuned to reproduce ISR data. The $dN_{ch}/d\eta$ distribution was measured for both collision systems and compared. The comparison was done using the uncorrected data and only for events with at least two tracks. In this way the authors attempted to achieve lower systematic errors on the result. A ratio of 1.015 ± 0.012 (p+$\bar{\text{p}}$ vs. p+p) has been concluded. Furthermore, the multiplicity distributions were compared. It is concluded that these distributions agree within errors. The authors summarize that the differences between p+p and p+$\bar{\text{p}}$ collisions are less than 2%.

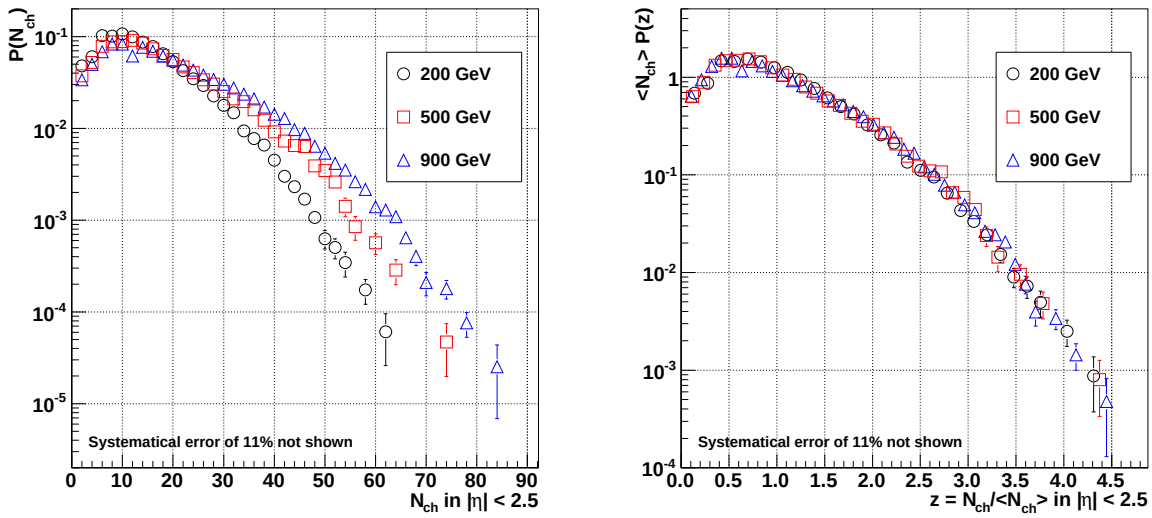

Figure 2.2: Normalized multiplicity distributions by UA1 in $|\eta| < 2.5$.
The distributions of NSD events measured at $\sqrt{s} = 200$, 500, and 900 GeV are shown vs. multiplicity (left panel) and in KNO variables (right panel) [Alb90].

2.2 SppS

The UA1 (Underground Area 1) experiment [Ala78] that operated at the SppS collider consisted of a tracking detector covering $|\eta| < 3.5$ and a calorimeter covering $|\eta| < 6$. The minimum-bias trigger was based on a set of hodoscopes that required at least one particle on both sides of the detector. This trigger accepted $96\% \pm 2\%$ of the ND events [Ast85]. Results from UA1 together with results from UA2 showed the existence of the gauge bosons of the weak force, the $W^{\pm}$ and Z^0, which led to the Nobel Prize in physics for Carlo Rubbia and Simon van der Meer.

Figure 2.2 shows the multiplicity distribution for NSD events in the interval $|\eta| < 2.5$ measured by UA1 at $\sqrt{s} = 200$, 500, and 900 GeV [Alb90]. 188 000 events were used, out of which 34% were recorded at the highest energy. The SppS was operated in a pulsed mode where data was taken during the energy ramp from 200 GeV to 900 GeV and vice versa. Therefore the data at 500 GeV is in fact taken in an energy range from 440 GeV to 560 GeV. The result is corrected for the acceptance and contributions from γ-conversions, strange-particle decays, and secondaries originating from interactions with the beam pipe. Only tracks with a p_T larger than 150 MeV/c are considered for the analysis to reduce the contamination by secondaries. Although not explicitly mentioned in the publication, it is assumed for this thesis that the low-momentum cut-off correction is part of the acceptance correction. UA1 quotes the overall system-

atic error to be 15%: contributions are strange-particle decays, photon conversions and secondary interactions (3%), as well as the uncertainty in the acceptance (4%). Other contributions arise from the selection criteria and uncertainties in the luminosity measurement (10%).[1] The distributions for $|\eta| < 2.5$ are consistent with KNO scaling up to $\sqrt{s} = 900\,\text{GeV}$ (see the right panel of Figure 2.2).

UA1 measured the $dN_{ch}/d\eta$ distribution at $\sqrt{s} = 540\,\text{GeV}$ [Arn83]. The results are included in Figure 2.8. The analysis used 8 000 events that have been taken without magnetic field which reduced the amount of particles lost at low-momenta to about 1%. The data has been corrected for acceptance and secondaries, as described in the previous paragraph. The systematic error of the applied corrections is estimated by the authors to be 5% without enumerating the contributions.

The UA5 (Underground Area 5) experiment [Rus81] at the Spp̄S consisted of two streamer chambers, one above and one below the beam pipe, with an acceptance of 95% for particles inside $|\eta| < 3$. The trigger was provided by scintillating-counter hodoscopes on either side (coverage of $2 < |\eta| < 5.6$); it has been evaluated to accept 95% (91%) of NSD events at $\sqrt{s} = 900\ (200)\,\text{GeV}$. At the same time almost all elastic and SD events were rejected [Ans89]. Naturally, the trigger efficiency was depending on the multiplicity and is only 30% for low multiplicities [Aln85]. Upon a trigger the chambers were photographed from both sides and the pictures were subsequently evaluated.

Figure 2.8 includes the $dN_{ch}/d\eta$ distribution at $\sqrt{s} = 200$ and $900\,\text{GeV}$ for NSD events measured by UA5 [Aln86b, Aln87]. 3500 (2100) events have been used for the analysis at 900 (200) GeV. It should be noted that the corrections are based on a MC simulation that has been tuned to reproduce data measured at $\sqrt{s} = 546\,\text{GeV}$. The results of the simulation were parameterized and scaled to $\sqrt{s} = 200$ and $900\,\text{GeV}$ in order to estimate the corrections for acceptance and contamination by secondaries. Unfortunately, the authors only mention statistical errors explicitly and therefore only these are included in the figure.

Measurements of the multiplicity distribution have been presented in [Aln84, Aln85, Aln86a, Ans89][2]. Figure 2.3 shows the multiplicity distribution at $\sqrt{s} = 200$, 540, and $900\,\text{GeV}$ in $|\eta| < 1.5$ for NSD events measured with UA5. The publications also present distributions in other η-ranges and extrapolated to full phase space. The analysis used

[1] The luminosity measurement uncertainty only applies to the cross-section measurement, not to the normalized distribution. The uncertainty due to the selection criteria is not quoted. Therefore, assuming that the systematic uncertainties were summed in quadrature, this uncertainty is 10% and the overall systematic error without the uncertainty on the luminosity is 11% which is the value applicable to Figure 2.2.

[2] [Ans89] partially revised the method to obtain the distribution and thus the results. Therefore, the results from [Ans89] are used instead of [Aln86a].

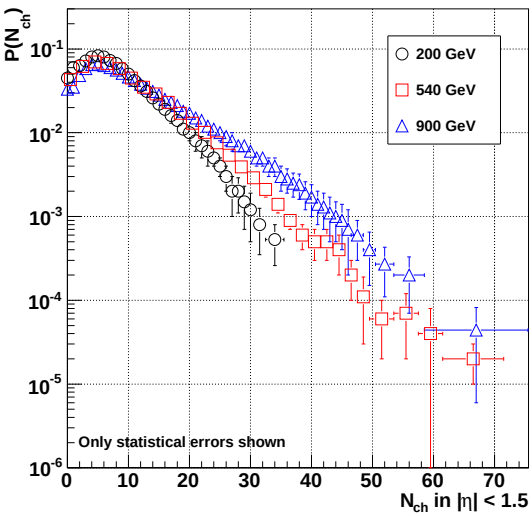

Figure 2.3: Normalized multiplicity distributions of NSD events in $|\eta| < 1.5$ at $\sqrt{s} = 200$, 540, and 900 GeV by UA5 (data from [Aln85, Ans89]).

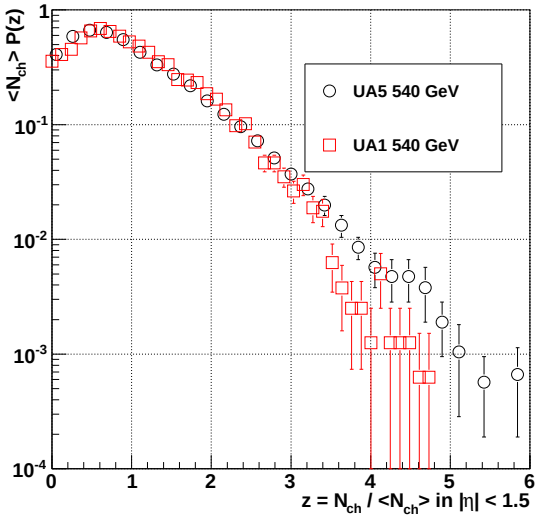

Figure 2.4: Multiplicity distribution of NSD events measured by UA1 and UA5 in $|\eta| < 1.5$ shown in KNO variables [Arn83, Aln85].

4 000, 7 000, and 9 000 events at 200, 540, and 900 GeV, respectively. In all cases the unfolding of the measured spectrum was performed by minimizing a χ^2-function. For the case of $\sqrt{s} = 540$ GeV [Aln85] it was required that the resulting function is an NBD which is regarded as a strong constraint. This has to be taken into account when interpreting the result at 540 GeV. The distributions at 200 and 900 GeV were unfolded using the maximum-entropy method [Ans89] which is considered to be a less restrictive assumption. The assessment of the systematic errors is not very comprehensive and concludes an uncertainty of about 2%.

UA5 reports the breaking of KNO scaling for $\sqrt{s} = 200 - 900$ GeV in the NSD sample for the extrapolation to full phase space. k is found to follow $k^{-1} \propto \ln s$ but is supposed to be independent of $\sqrt{s}$ for exact KNO scaling [Aln86a]. Furthermore, only approximate scaling at $\sqrt{s} = 540$ GeV in $|\eta| < 1.5$ is seen [Aln84]. This observation has been questioned by UA1 [Alb90]. A direct comparison between UA1 and UA5 at $\sqrt{s} = 540$ GeV in limited regions and in KNO variables shows that the two experiments agree in the interval $|\eta| < 0.5$ (both confirm KNO scaling) and disagree in the interval $|\eta| < 1.5$, but the violation of KNO scaling in the UA5 data is only due to an excess of events with $z > 3.5$, i.e. events that have more then 3.5 times the average multiplicity. This comparison has been performed in [Alb90] and is shown for $|\eta| < 1.5$ in Figure 2.4. Although not explicitly mentioned by the authors, by studying the mo-

ments of the distribution in [Ans89] it can be seen that KNO scaling is confirmed in a limited region of $|\eta| < 0.5$ at $\sqrt{s} = 900\,\text{GeV}$.

The multiplicity distribution is described by an NBD at $\sqrt{s} = 200$ and $540\,\text{GeV}$ in full phase space as well as in different η-ranges. This behavior is not continued for $\sqrt{s} = 900\,\text{GeV}$ [Ans89]. Figure 2.5 shows the multiplicity distribution together with an NBD fit in increasing pseudorapidity ranges at 900 GeV (top left panel). The respective normalized residuals are also shown (top right panel). The NBD fit works very well for the interval $|\eta| < 0.5$, but gets worse for larger η-ranges; it is more and more pronounced that the region around the most probable multiplicity is not reproduced. The structure found around the peak gave rise to the two-component approach, discussed previously, that suggests to fit the data with a combination of two NBDs. The bottom left panel of Figure 2.5 shows this fit, Eq. (1.53), and normalized residuals (bottom right panel) to the same data which yields good fit results for all pseudorapidity ranges.

A Forward Silicon Micro-Vertex detector that was tested in the context of a proposed hadronic B-physics experiment (P238) measured the $dN_{ch}/d\eta$ distribution in forward rapidities at $\sqrt{s} = 630\,\text{GeV}$ [Har97]. The detector consisted of six planes with two silicon micro-strip detectors each (one for the measurement of each of the coordinates x and y). The overall detector size was 4.5 cm x 4.5 cm x 20 cm and allowed a measurement of charged particles in $1.5 < |\eta| < 5.5$ due to the varying vertex position.

5 million events were recorded with a scintillator counter trigger located $\pm 3\,\text{m}$ from the interaction region. A coincidence between both sides was required. Only events for which a vertex position was found have been considered for the analysis. The sample is corrected for tracks from secondaries (2%) and SD events (0.5%). Acceptance and resolution effects are corrected by using MC simulations tuned to UA5 data. Their magnitude as well as the magnitude of the trigger- and vertex-efficiency correction are not mentioned. A normalization error of 5% dominates the systematic error that stems from inconsistent results when only the x or y tracking information is used compared to when both of them are used. Other effects such as detector efficiency, misalignment, and the SD cross-section are considered by the authors to not significantly contribute to the systematic uncertainty. Figure 2.8 includes the measured $dN_{ch}/d\eta$ distribution.

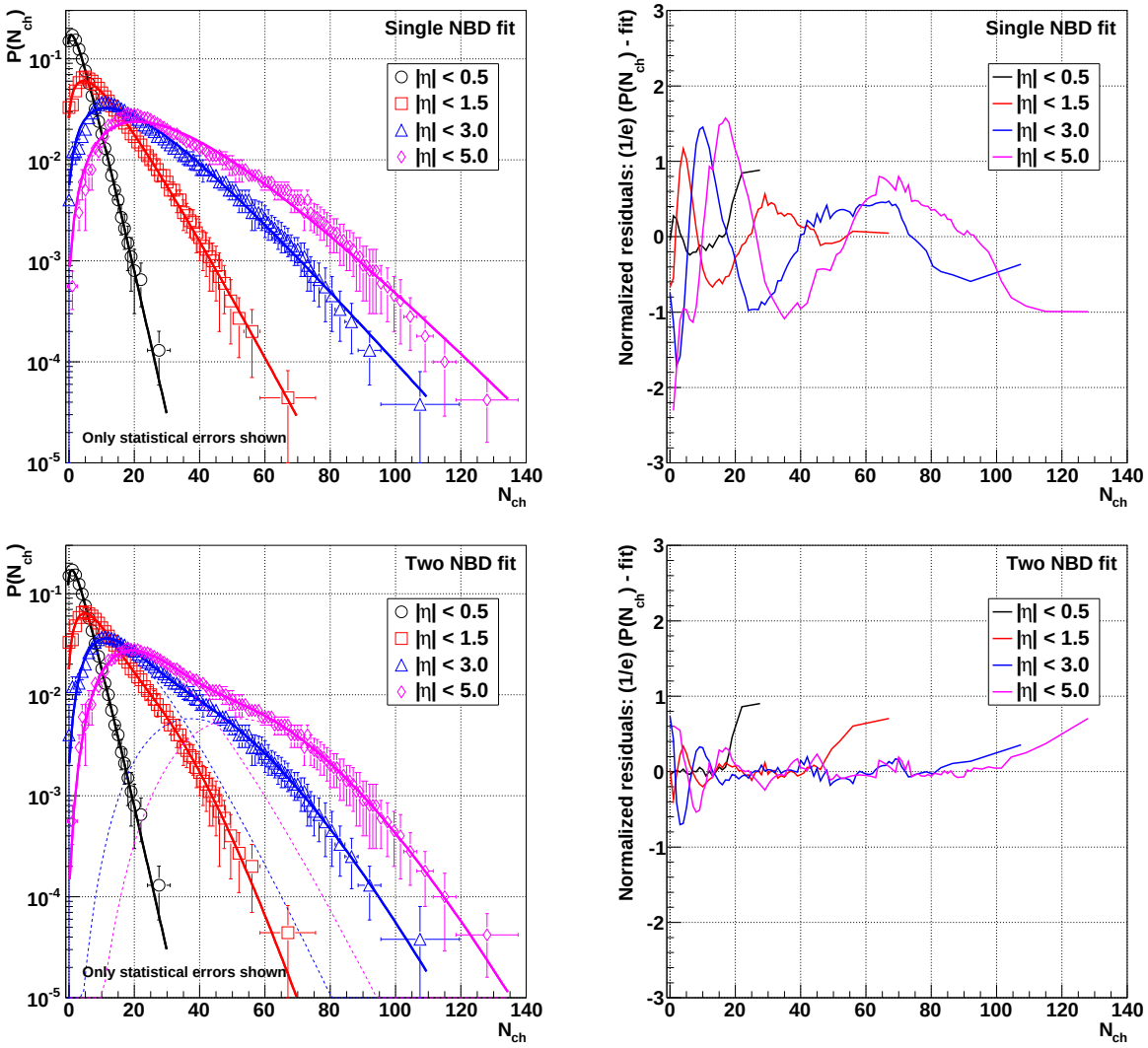

Figure 2.5: Multiplicity distributions of NSD events at $\sqrt{s} = 900\,\text{GeV}$. Normalized multiplicity distributions in various rapidity intervals are shown fitted with single NBDs (top left panel) or a combination of two NBDs (bottom left panel). The two contributing NBDs (dashed lines) are shown exemplarily for $|\eta| < 3.0$ and 5.0. The right panels show the normalized residuals with respect to the corresponding fits defined by $(1/e)(P(N_{ch}) - \text{fit})$ with e being the error on $P(N_{ch})$. These are smoothed over four data points to reduce fluctuations. The data has been measured by UA5 [Ans89].

2.3 Tevatron

The CDF (Collider Detector at Fermilab) experiment [Abe88], a detector at the Tevatron collider, consists of eight time-projection chambers (called VTPC for vertex time-projection chamber) that cover the central rapidity region in $|\eta| < 3.5$, a central tracking chamber covering $|\eta| < 1.0$, and beam–beam counters on either side of the detector as trigger at higher rapidities ($3.2 < |\eta| < 5.9$). The detector is contained in a 1.5 T magnetic field. The detector was upgraded after years of operation replacing the VTPCs by a silicon detector.

Figure 2.8 includes the $dN_{ch}/d\eta$ distribution at $\sqrt{s} = 630\,\text{GeV}$ and 1.8 TeV measured by CDF with their VTPCs [Abe90]. Unfortunately, the authors do not mention if their corrections correspond to NSD or inelastic events. However, the trigger configuration requires a hit on both sides. This points to the fact that the trigger is insensitive to the majority of SD events. Furthermore, the authors compare their measurement to NSD data from UA5 which confirms that the CDF data is for NSD events. 2 800 (21 000) events have been used for the analysis at 630 (1 800) GeV. Only events with at least 4 tracks are considered to reduce the beam-gas background. The authors stated that they "do not correct for events missed by the trigger or selection procedure" and estimated that the selection procedure misses $13\% \pm 6\%$ of the events. This is surprising because the normalization for $dN_{ch}/d\eta$ would be significantly distorted if this correction was not applied. This is not the case shown in the comparison to UA5 data. Tracks with $p_T < 50\,\text{MeV}/c$ are not found due to the magnetic field and a correction of $3\% \pm 2\%$ is applied to account for this loss. Contamination by photon conversions and secondaries from hadrons are estimated to be less than a few percent. A systematic error assessment is made; the error is dominated by uncertainties in the tracking efficiency and ranges from 3% (at $\eta = 0$) to 15% (at $|\eta| = 3.25$).

CDF measured the multiplicity distribution in various η-intervals for NSD events at $\sqrt{s} = 1.8\,\text{TeV}$ [Rim93]. Figure 2.6 shows the distribution for two pseudorapidity intervals. No errors can be shown as the data points were extracted from a plot with a very poor resolution. The publication does not mention the number of events used in the analysis. A systematic-error assessment is reported to be ongoing, but has not yet been published. The result has been corrected for γ-conversions, neutral decays, and secondaries. It is unclear if an unfolding method was used. The authors find that the data can be described by an NBD in the most central region, $|\eta| < 0.5$, but not in other pseudorapidity windows. A weak KNO scaling violation is observed but the authors suggest waiting for the detailed assessment of the systematic errors to make a final conclusion. The data can be fitted well by the combination of two NBDs which is shown superimposed in Figure 2.6.

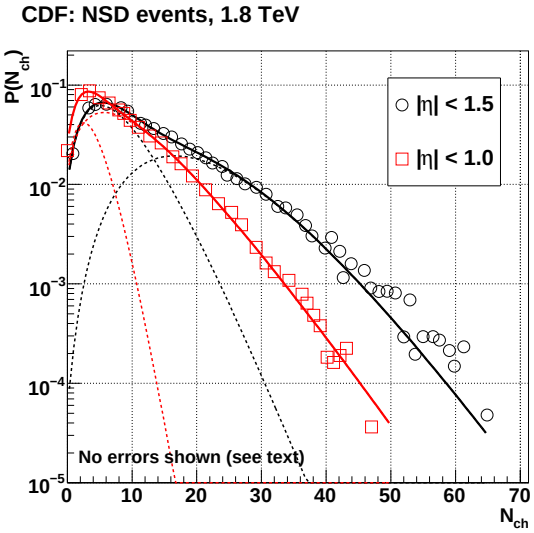

Figure 2.6: Multiplicity distributions by CDF at $\sqrt{s} = 1.8\,\text{TeV}$.

The figure shows multiplicity distribution in $|\eta| < 1.0$ (red squares) and $|\eta| < 1.5$ (black circles) [Rim93]. The data is fitted with the combination of two NBDs, the contributing NBDs are also shown (dashed lines).

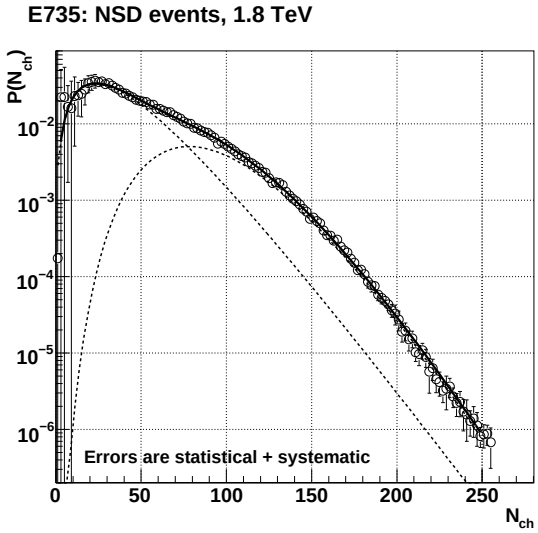

Figure 2.7: Multiplicity distributions by E735 at $\sqrt{s} = 1.8\,\text{TeV}$.

The figure shows multiplicity distribution of NSD events in full phase space [Ale98]. The data is fitted with the combination of two NBDs, the contributing NBDs are also shown (dashed lines).

The E735 experiment [Lin92] at the Tevatron collider measured the multiplicity distribution at energies up to $\sqrt{s} = 1.8\,\text{TeV}$. The experiment combined a multiplicity hodoscope covering $|\eta| < 3.25$, two trigger hodoscopes in the pseudorapidity interval $3.9 < |\eta| < 4.5$, as well as a TOF system and a magnetic spectrometer covering a smaller region of phase space.

Figure 2.7 shows the multiplicity distribution of NSD events in full phase space at $\sqrt{s} = 1.8\,\text{TeV}$ [Ale98]. Like before, the distribution is fitted with the combination of two NBDs. The extrapolation to full phase space has been done by the authors based on Pythia simulations. They provide no further information about the statistics used, the corrections, and in particular the question whether an unfolding was used. This has to be taken into account when the result is interpreted.

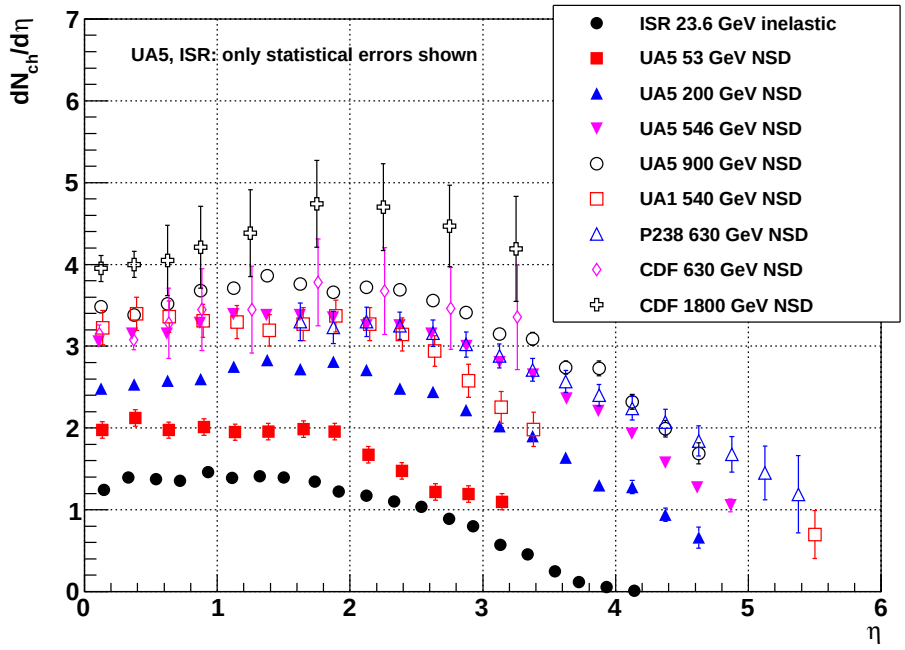

Figure 2.8: $dN_{ch}/d\eta$ at different $\sqrt{s}$.

Data points from [Tho77, Arn83, Aln85, Ans89, Abe90, Har97].

2.4 Summary and Critical Assessment

Figure 2.8 shows $dN_{ch}/d\eta$ at energies ranging over about two orders of magnitudes, from ISR ($\sqrt{s} = 23.6\,\text{GeV}$) to CDF ($\sqrt{s} = 1.8\,\text{TeV}$). Increasing the energy shows an increase in multiplicity. The multiplicity of the central plateau increases together with the variance of the distribution. Note that the data points at the lowest energy are for inelastic events, the other data points refer to NSD events.

Figure 2.9 shows $dN_{ch}/d\eta|_{\eta=0}$ as a function of $\sqrt{s}$. Filled symbols are data for inelastic events; open symbols for NSD events. $dN_{ch}/d\eta|_{\eta=0}$ increases with increasing $\sqrt{s}$. Two fits are shown for the NSD data[3]: a fit with $a + b \ln s$ (solid black line) and $a + b \ln s + c \ln^2 s$ (dashed red line). The $\ln s$ dependence was used to describe the data at center-of-mass energies below 1 TeV. Data at a higher energy from CDF showed that the fit is no longer satisfactory [Abe90]. The additional $\ln^2 s$ term yields a much better result and shows that the multiplicity increases faster than $\ln s$. The functional fits are

[3]Due to the fact that different published values include different errors, e.g. no systematic errors for the UA5 data, the errors are not used for the fit.

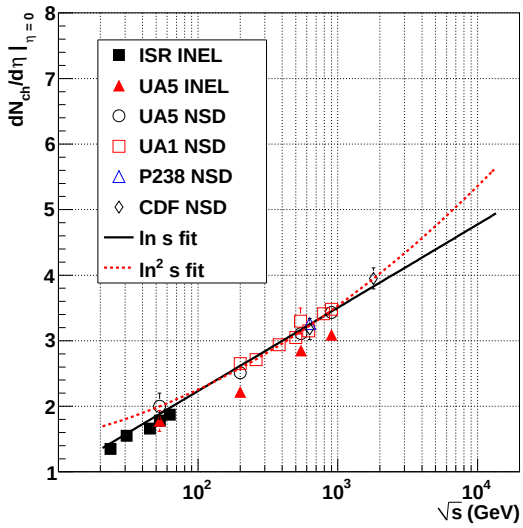

Figure 2.9: $dN_{ch}/d\eta|_{\eta=0}$ as a function of $\sqrt{s}$. Data points from [Tho77, Aln85, Ans89, Abe90, Alb90].

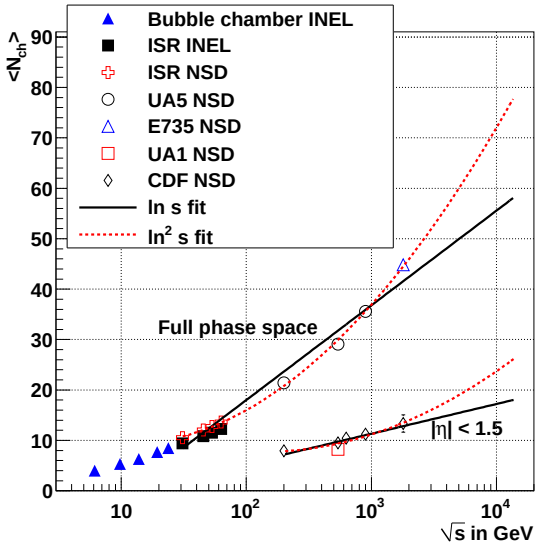

Figure 2.10: $\langle N_{ch}\rangle$ vs. $\sqrt{s}$ shown in full phase space and $|\eta| < 1.5$. Data points from [Sla72, Whi74, Arn83, Aln84, Bre84, Aln85, Ans89, Ale98, Rim93].

extrapolated up to the nominal LHC energy, $\sqrt{s} = 14\,\text{TeV}$. The increase of $dN_{ch}/d\eta|_{\eta=0}$, the plateau of the pseudorapidity distribution, is violating Feynman scaling.

Figure 2.10 shows the average multiplicity $\langle N_{ch}\rangle$ as a function of $\sqrt{s}$. Data is shown for full phase space and for a limited rapidity range of $|\eta| < 1.5$. In publications two different approaches are found to obtain average values in a limited η-range. The first uses a normalization to all events having at least one track in the considered phase space. The second approach uses a normalization to the total considered cross-section (inelastic or NSD) including events without any particle in the considered range (data shown here). While the latter is the more evident physical observable, the former is not dependent on the efficiency to measure the total cross-section. Thus the former is less dependent on model assumptions used in the evaluation of the trigger efficiency. Data from bubble chambers at low $\sqrt{s}$ is included in Figure 2.10: from the Mirabelle chamber at Serpukhov, Russia [Sla72] and from several bubble chambers at FNAL [Whi74]. Owing to their design, bubble chambers see the full collision, i.e. have full phase space acceptance. Both sets of NSD data are fitted, as before, ln and $\ln^2$-dependent. For full phase space the logarithmic dependence does not reproduce the data and is only shown to demonstrate the violation of Feynman scaling; the $\ln^2$ dependence fits the data well. For limited phase space the fit containing the $\ln^2$-term is preferred, but deviations from both fits are visible. The functional fits are extrapolated to $\sqrt{s} = 14\,\text{TeV}$.

In summary, for NSD events, KNO scaling has been observed at the ISR from $\sqrt{s} = 30.4\,\text{GeV}$ to $62.2\,\text{GeV}$. This scaling has been reported to fail starting at about $200\,\text{GeV}$ up to the highest measured energy of $1.8\,\text{TeV}$ in full phase space (UA5, E735). In the central region KNO scaling holds up to $900\,\text{GeV}$ (UA1 in $|\eta| < 2.5$ and UA5 in $|\eta| < 0.5$), but UA5 notices a departure from KNO scaling already at $540\,\text{GeV}$ in $|\eta| < 1.5$. UA1 and UA5 yield contradictory results with regard to this point. Measurements in limited phase space of CDF do not allow for a final conclusion. However, it is interesting to mention a study of CDF at $1.8\,\text{TeV}$ using only tracks with a p_T above $0.4\,\text{GeV}/c$ [Aco02]. Here, a weak KNO scaling violation is reported in $|\eta| < 1.0$. Furthermore, when they divide their data sample into two parts, they can confirm KNO scaling for the soft part of their events and at the same time rule it out for the hard part. In [Aco02] soft events are defined as events without clusters of tracks with a total transverse energy above $1.1\,\text{GeV}$, regarded as jets. Two further interesting features are observed together with the onset of KNO scaling violations [Alb90]: the average transverse momentum that was about $360\,\text{MeV}/c$ at ISR energies starts to increase. Furthermore, a $\sqrt{s}$-dependent correlation between the average-p_T and the multiplicity is discovered. Both observations point to the fact that the influence of hard scattering becomes important at these energies.

For inelastic events, KNO scaling has been observed in $|\eta| < 1.5$ for 23.6 – $62.8\,\text{GeV}$ (ISR). For full phase space it has not been found from $\sqrt{s} = 30.4\,\text{GeV}$. However, inelastic events taken at bubble-chamber experiments have been reported to follow KNO scaling at low $\sqrt{s}$ (10 – $24\,\text{GeV}$) [Sla72] which is not further discussed because this energy regime is not considered important for the measurements at the LHC.

For NSD events, the multiplicity distribution can be described by an NBD up to $\sqrt{s} = 540\,\text{GeV}$ in full phase space (ISR, UA5) and in the central region (UA5). The NBD succeeds to describe data up to $1.8\,\text{TeV}$ in $|\eta| < 0.5$ (UA5, CDF), but fails for larger η-intervals and full phase space (UA5, CDF). NBDs also reproduce multiplicity distributions of e^+e^- collisions (see e.g. [Bra89] for data in $\sqrt{s} = 14 - 43.6\,\text{GeV}$). However, ALEPH (Apparatus for LEP Physics) reported that already at $\sqrt{s} = 91.2\,\text{GeV}$ the charged-particle multiplicity distribution cannot be described by a single NBD for both, full phase space and restricted rapidity intervals (smallest analyzed region: $|y| < 0.5$) [Bus95].

The fit following the two-component model with two NBDs succeeds in full phase space and in the central region for NSD events at all mentioned energies underlining the likely importance of multiple-parton dynamics of the higher energies.

Chapter 3

The Large Hadron Collider

The Large Hadron Collider (LHC) [Pet95, LHC04] at CERN is the biggest particle accelerator world-wide. First discussions that led to the project started in 1984. The LHC project was approved in 1994 and construction work in the underground tunnel started in 2001 after dismantling of the LEP collider[1]. LEP was previously built into the tunnel which is located under the Swiss-French border area close to Geneva at a depth of 50 to 175 m. The LHC has a circumference of 27 km. Its largest achievable acceleration energies are 7 TeV for protons and 2.76 TeV per nucleon for lead ions, therefore providing collisions at $\sqrt{s} = 14\,\text{TeV}$ and $\sqrt{s_{NN}} = 5.5\,\text{TeV}$, respectively. These are the largest energies that have ever been accessible in particle collision experiments.

3.1 Design

The LHC is a synchrotron that accelerates two counter-rotating beams in separate beam pipes. In each of them bunches of particles travel many times around the accelerator ring before the collision energy is reached. The accelerator has to bend the beams around the ring, keep the bunches focused and accelerate them to their collision energy. Finally, the spatial dimension of the bunches has to be minimized to provide a high number of collisions per time interval at the collision points, i.e. a high luminosity[2]. A combination of magnetic and electric field components performs the mentioned tasks.

[1] The Large Electron-Positron (LEP) Collider operated in the years 1989 to 2000 with a maximum center-of-mass energy of 209 GeV.

[2] For a particle accelerator experiment, the luminosity is defined by: $\mathcal{L} = fnN^2/A$ with n bunches in both beams, N particles per bunch, cross-sectional area A of the beams that overlap completely, and revolution frequency f. The frequency of interactions (or in general of a given process) can be calculated from the corresponding cross-section σ and the luminosity: $dN/dt = \mathcal{L}\sigma$.

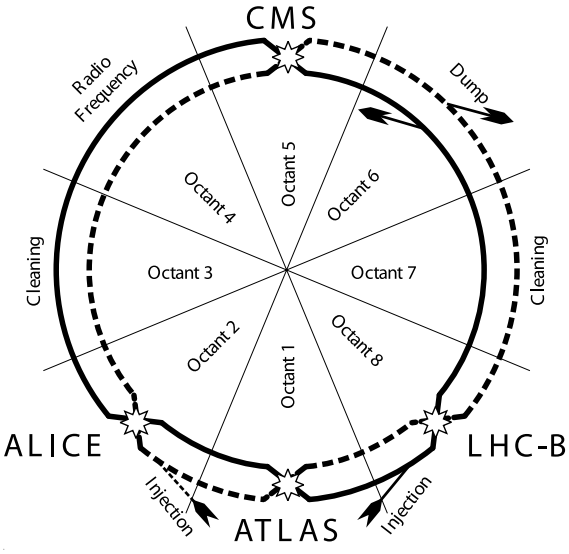

Figure 3.1: Schematic view of the LHC.

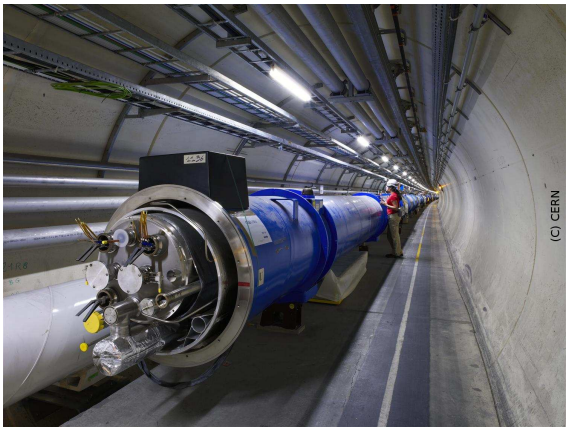

Figure 3.2: Photograph taken inside the LHC tunnel. Visible are several of the main LHC dipoles (blue); the two (not yet connected) beam pipes can be seen in the front.

The layout of the LHC is shown in Figure 3.1. It is segmented into octants, each has a straight section in its center, referred to as *points*. The arcs are called Sector xy where x and y are the numbers of the corresponding octants in clock-wise order, e.g. Sector 34. Four of the straight sections contain the experiments (points 1, 2, 5, and 8) which are the only locations where the beams cross. Particles are injected before point 2 and 8. The radio-frequency (RF) system that accelerates the particles is located at point 4; the beam dumping system is located at point 6. At point 3 and 7 collimation systems are placed that 'clean' the beam by removing particles that have either a too large spatial distance to their bunch (particles in the so-called beam-halo) or are too fast or too slow, thus separated in momentum-space. The cleaning prevents particles from being lost in an uncontrolled fashion within the accelerator.

The main components of the machine are 1 232 dipoles that bend the beam trajectories. An LHC dipole has a length of 14.3 m and contains superconducting magnets which operate at a temperature of 1.9 K, notably 0.8 K lower than the background temperature of the Universe. Powered by a maximum current of 11.7 kA the dipoles provide a magnetic field from 0.535 T during injection (beam energy of 450 GeV) to 8.33 T during nominal collisions (beam energy of 7 TeV). Several dipoles are shown in the photograph in Figure 3.2. Eight *RF cavities* per beam deliver radio-frequency power to accelerate the beams, keep the bunches of particles well-localized and compensate for energy loss due to synchrotron radiation. The cavities produce a field of 5.5 MV/m.

Despite the high luminosity reached, only a very small fraction of the particles of two bunches collides in a single bunch crossing. The others leave the interaction region essentially uninfluenced, are defocused, and continue to circulate in the accelerator.

Injection of bunches into the LHC is preceded by acceleration in the LINAC2, PS booster, PS, and SPS accelerators. The acceleration sequence is slightly different for heavy ions, bunches pass the LINAC3, LEIR, PS, and SPS accelerators (more information can be found in [LHC04, Chapter I-21]). Several injections to the LHC are needed until all bunches of both beams are filled. The design parameters foresee nominal operation, where each beam is filled with 2 808 bunches each consisting of 1.15×10^{11} protons. Bunches have a r.m.s.[3] length between 11.24 cm at injection and 7.55 cm at collision. They are separated by 25 ns.[4] LHC's design luminosity is $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for protons and $10^{27}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for Pb ions. However, the LHC will deliver a significantly lower luminosity to the ALICE experiment during proton collisions (about $3 \times 10^{30}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$) by means of defocusing or displacing the beams. At nominal luminosity about 2.4×10^9 p+p collisions are estimated to occur per second in the LHC which corresponds to about 2×10^{11} produced particles per second. These are recorded by six experiments that operate at the LHC:

ALICE (A Large Ion Collider Experiment) [ALI95] is a dedicated heavy-ion experiment designed to study strongly-interacting matter. It explores the phase transition to the quark–gluon plasma, its phase diagram, and its properties. Furthermore, ALICE will also study collisions of protons, on the one hand as a baseline for heavy-ion measurements and on the other hand it contributes to topics involving its supreme particle identification capabilities and its acceptance at very low transverse momenta. This thesis is based on the ALICE experiment described in detail in Chapter 4.

ATLAS (A Toroidal LHC ApparatuS) [ATL94] and **CMS (Compact Muon Solenoid)** [CMS94] are general-purpose proton–proton detectors that are built to cover the widest possible range of physics at the LHC. Specific topics are the search for the Higgs boson and physics beyond the Standard Model, e.g. new heavy particles postulated by supersymmetric extensions (SUSY) of the Standard Model and evidence of extra dimensions.

LHCb (The Large Hadron Collider beauty experiment) [LHC98] studies CP-symmetry violation processes in heavy b-quark systems.

[3] R.m.s. = Root mean square.

[4] For the acceleration of heavy ions the machine contains 592 bunches with 7×10^7 Pb ions each. However, recent discussions indicate possible changes of these values.

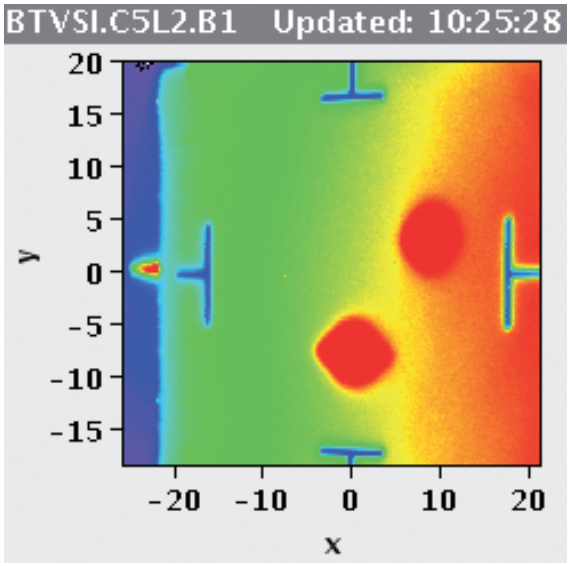

Figure 3.3: The first turn of a beam in the LHC (10.09.2008, 10:25). The two red spots show the beam at injection and after the first turn.

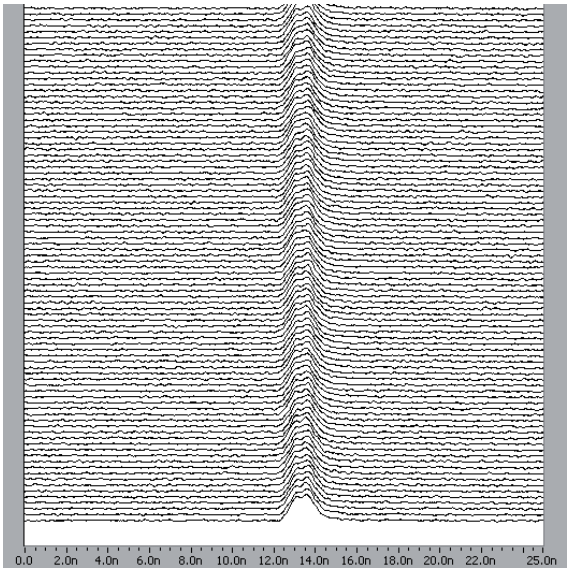

Figure 3.4: Captured beam with the RF (11.09.2008, 22:43). Each line is one turn of one bunch in the LHC.

LHCf (Large Hadron Collider forward experiment) [LHC05] measures forward particles created during LHC collisions to provide further understanding of high-energy cosmic rays. The detector is placed close to the ATLAS experiment.

TOTEM (TOTal Elastic and diffractive cross-section Measurement) [TOT99] measures the total cross-section, elastic scattering, and diffractive processes. The detector is located close to the CMS experiment.

3.2 Startup and Status

LHC started on the 10th September 2008 with great success. In less than an hour after the first injection the first beam had been sent successfully around the entire ring. Figure 3.3 shows a beam monitor of the first bunch that passed through the entire LHC. During the same day the second beam in the opposite direction successfully passed through the ring. In a few days commissioning made spectacular progress. The RF captured the beam successfully soon after and a stable circulating beam was achieved on the 12th of September. The monitoring of one of the first RF-captured bunches circulating can be seen in Figure 3.4.

Unfortunately, a transformer failure in point 8 stalled the commissioning for about a week. On the 19th September when LHC was basically ready for collisions at $\sqrt{s} = 900\,\text{GeV}$, an accident occurred during the 10 TeV magnet commissioning without beam in Sector 34, the last sector that was commissioned to this energy [Leb08]. To repair the damage that occurred in the machine the sector had to be warmed up, which delays LHC operations at least until late Summer 2009 (status of January 2009).

3.3 Collision Parameters

The startup scenario for first collision [LHC08] foresees: collisions at $\sqrt{s} = 900\,\text{GeV}$ with four bunches per beam at a bunch intensity of 4×10^{10} particles and a luminosity of $6.6 \times 10^{27}\,\text{cm}^{-2}\,\text{s}^{-1}$. Similarly, the first collisions at $\sqrt{s} = 10\,\text{TeV}$ are planned with 12 bunches per beam at an intensity of 3×10^{10} particles per bunch, yielding a luminosity of $1.7 \times 10^{29}\,\text{cm}^{-2}\,\text{s}^{-1}$. These numbers correspond to a probability of a collision per bunch crossing of 7.3×10^{-3} and 9×10^{-2} for $\sqrt{s} = 900\,\text{GeV}$ and 10 TeV, respectively. The probability of having a second or more collisions in a bunch crossing in which a collision occurred is 0.37% and 4.9% for the two energies, respectively. Clearly, the conditions under which first measurements will be performed can only be estimated at the present stage and thus these scenarios are only tentative.

Nominal collisions are most likely to be reached initially at $\sqrt{s} = 10\,\text{TeV}$ and with the following parameters: a luminosity of $3 \times 10^{30}\,\text{cm}^{-2}\,\text{s}^{-1}$ in ALICE with 2808 bunches per beam with an intensity of 1.15×10^{11} protons per bunch.[5] Due to the increased number of bunches the probability for a collision per bunch crossing is lower than before: 7.1×10^{-3}. The probability of having a second or more collisions in a bunch crossing in which a collision occurred is then 0.36%.

Numbers that depend on collision parameters use these scenarios, referred to as startup scenario and nominal running conditions.

[5]The change of other beam parameters between the startup scenario and nominal collisions results in a smaller increase in luminosity than the one that may be assumed from the number of bunches and their intensity.

Chapter 4

The ALICE Detector

ALICE (A Large Ion Collider Experiment) [ALI95] is a general-purpose particle detector designed to study heavy-ion collisions. It has been optimized for the very high-multiplicity environment that is created in central heavy-ion collisions. The design was developed for $dN_{ch}/d\eta = 4\,000$, but tested up to $dN_{ch}/d\eta = 8\,000$.[1] ALICE is built and operated by a collaboration of more than 1 000 members from about 30 countries.

The detector's unique features are the tracking and particle identification over a large range of momenta, from tens of MeV/c to over 100 GeV/c, therefore accessing physics topics starting from soft to jet physics and high-p_T particle production. The detector consists of a central barrel ($|\eta| < 0.9$) contained in a magnetic field of 0.5 T and optimized for the detection of hadrons, electrons, and photons; a muon spectrometer at forward rapidities; as well as additional forward and trigger detectors. Figure 4.1 shows a schematic view of the detector. Note that the ZDC (detector abbreviations defined in the following sections) is shown only on the C side and at a position much closer to the experiment than installed. The positions of FMD, T0, and V0 are only indicative, see the following sections for their exact positions. Figure 4.2 presents the acceptance in η of the various subdetectors. A picture of the detector in the cavern is shown in Figure 4.3.

This chapter will introduce the various subdetectors of ALICE with a special focus on the subdetectors that are used for the analysis described in this thesis. In the discussion of the subdetectors the variables introduced in Appendix A and the ALICE coordinate system (see Appendix B) are used.

[1] Results from RHIC indicate that the expected multiplicity at LHC energies is lower than assumed during the design stage, $dN_{ch}/d\eta = 1\,500 - 4\,000$ [Aam08].

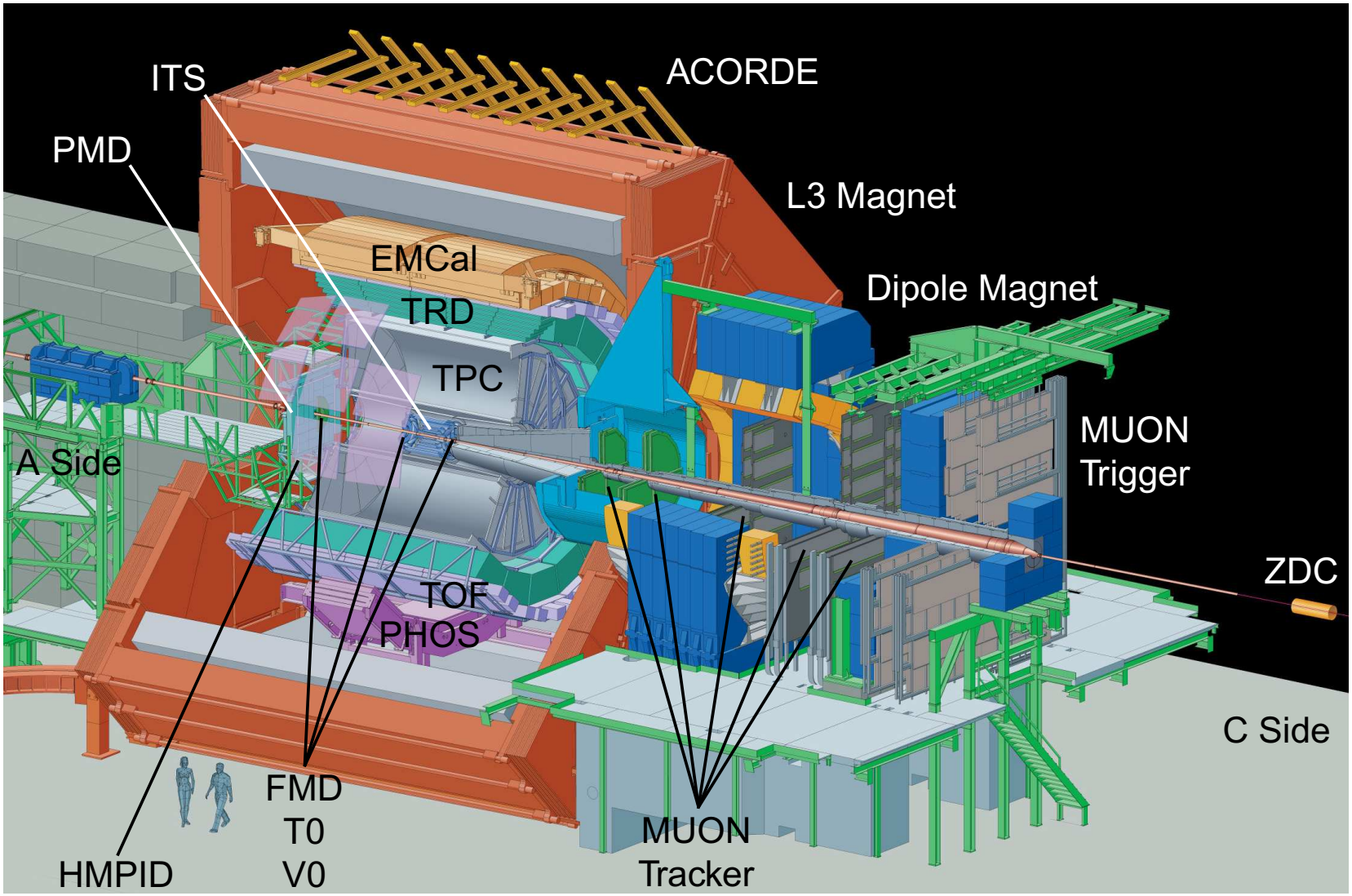

Figure 4.1: Schematic view of the ALICE detector.

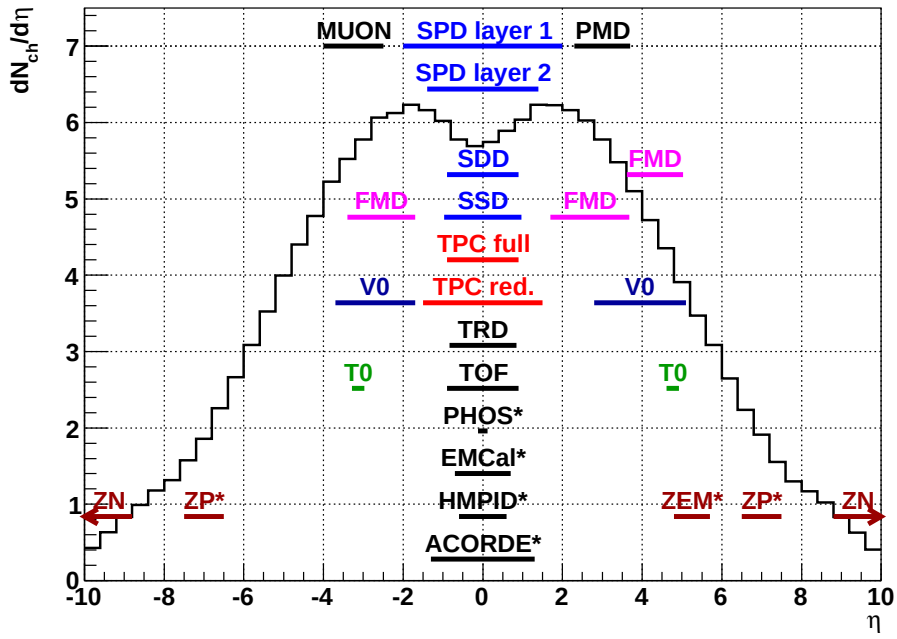

Figure 4.2: ALICE acceptance.

The figure shows the pseudorapidity acceptance of the subdetectors with an overlayed $dN_{ch}/d\eta$ prediction for p+p collisions by Pythia. The subdetectors have full coverage in azimuth except for the ones marked with an asterisk. SPD, SDD, and SSD are the subsystems of the ITS (see Section 4.1.1). Two ranges are given for the TPC, depending on the conditions imposed on the track length (full and reduced; see Section 4.1.2). ZN, ZP, and ZEM are the different parts of the ZDC (see Section 4.2.5).

4.1 The Central Barrel

A set of detectors covers the central region of ALICE. These are, in order of increasing radii: the Inner Tracking System (ITS), the Time-Projection Chamber (TPC), the Transition-Radiation Detector (TRD), and the Time-Of-Flight (TOF). These four detectors cover the central region ($|\eta| < 0.9$) and partly a larger acceptance which will be mentioned in the following. Their tasks are tracking and particle identification in the very high-multiplicity environment.

Additional detectors are located centrally but cover a significantly smaller region of phase space than the previously mentioned central region. These are the Photon Spectrometer (PHOS), the Electromagnetic Calorimeter (EMCal), the High-Momentum Particle Identification Detector (HMPID), and the ALICE Cosmic Ray Detector (ACORDE).

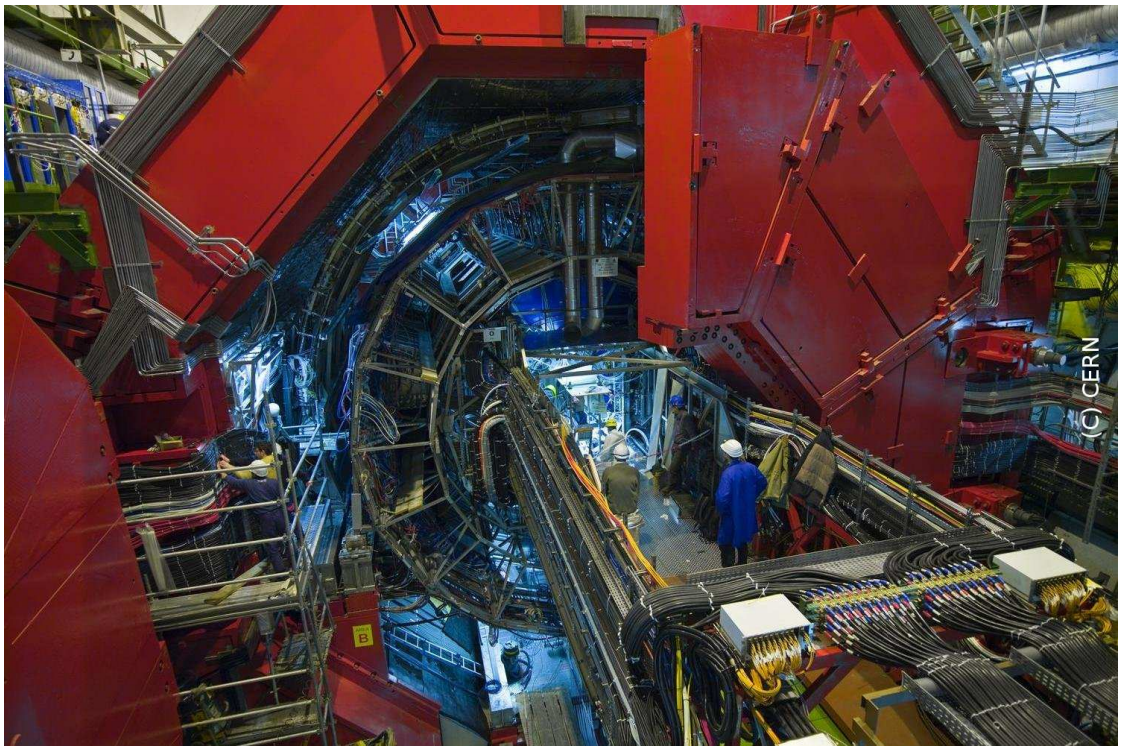

Figure 4.3: View of the ALICE detector with open L3 magnet taken in early 2008.

The central barrel is contained in the *L3 magnet* inherited from the LEP experiment L3. It has an inner length of 12.1 m and a radius of 5.75 m. This non-superconducting magnet was first operated in 1988.

4.1.1 The Inner Tracking System (ITS)

The ITS, shown in Figure 4.4, consists of six layers of silicon detectors with radii from 3.9 cm to 43 cm. The two innermost layers have an extended coverage to provide a measurement of the charged-particle multiplicity which, together with the FMD (see Section 4.2.2), results in a continuous coverage in η. The tasks of the ITS are the reconstruction of the primary vertex of the collision as well as the reconstruction of secondary vertices of heavy-quark decays (B and D mesons) and hyperons with a resolution better than 100 μm in transverse direction. The ITS contributes to the particle identification through the measurement of the specific energy loss (dE/dx) and to the tracking. It can be used to perform stand-alone tracking for low-momentum particles that do not reach the TPC (see subsequent section). The p_T cut-off at nominal field for the two innermost layers is about 35 MeV/c. These two layers have to sustain

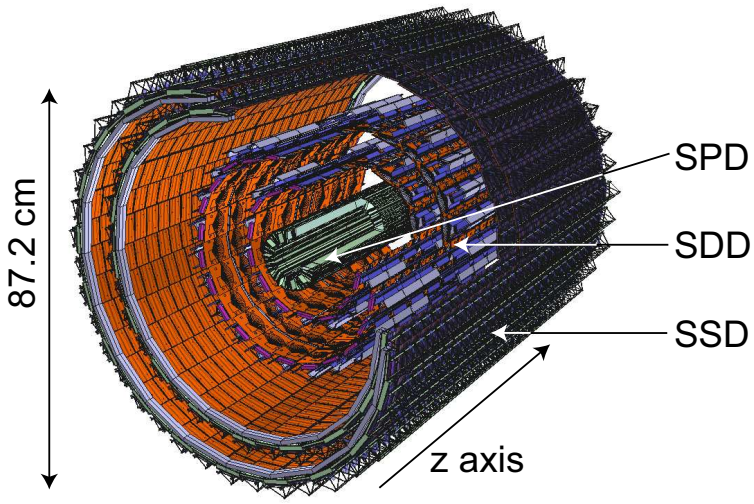

Figure 4.4: Schematic view of the ITS.

a very high rate of up to 50 particles per cm^2. For the outer layers the rate is below one particle per cm^2. The material budget is low, the total thickness in terms of radiation length X/X_0 is less than 8%; each layer contributes with about 1%, the remaining material is thermal shielding and support structures.

The two innermost layers, called Silicon Pixel Detector (SPD), are based on hybrid silicon pixels which consist of silicon detector diodes with a thickness of 200 μm. The first and the second layer are placed at 3.9 cm and 7.6 cm with an acceptance of $|\eta| < 2.0$ and $|\eta| < 1.4$ (for the nominal interaction point), respectively. In total 9.8 million channels are read out in a binary fashion, thus no energy-loss information is available and therefore the SPD does not contribute to particle identification. The information from the SPD alone is sufficient to form so-called tracklets and thus to measure the charged-particle multiplicity. The tracklet reconstruction is explained in detail in Section 4.6.3.

The SPD can be used as L0 trigger[2]. For this purpose each of the 1 200 readout chips provides a so-called *fast OR* signal indicating that at least one pixel of a given chip produced a signal. These signals are sent to an FPGA[3] which is able to implement boolean logic functions upon them. This allows trigger possibilities that range from a minimum-bias trigger to very complex trigger patterns. An example is to require a

[2]Triggers in ALICE are divided into levels (L0, L1, L2) depending at what latency after the collision they are issued. The ALICE trigger system is discussed in detail in Section 4.5.

[3]A Field-Programmable Gate Array (FPGA) is a microchip that performs predefined calculations based on a programmable logic. Thus the SPD FPGA is adaptable to allow for different trigger patterns.

certain number of two-chip combinations which resemble tracks pointing to the same vertex. This allows interactions to be triggered with an improved background rejection. The SPD integrates the trigger signal over 100 ns corresponding to four bunch crossings in nominal p+p runs[4], therefore the bunch crossing that caused the trigger needs to be identified with another detector, e.g. the V0 (discussed below).

The third and fourth layer, called Silicon Drift Detector (SDD), consist of a 300 μm thick layer of homogeneous high-resistivity silicon. Contrarily to the first two inner layers, the readout is analog, therefore energy-loss information for particle identification is provided. The SDD has 133 000 channels. The detector employs a drift time measurement resulting in a similar granularity as the SPD and SSD.

The two outermost layers, called Silicon Strip Detector (SSD), consist of sensors equipped on both sides with silicon micro-strips. These are arranged under a stereo angle of 35 mrad allowing for a two-dimensional measurement of the track position together with an energy-loss measurement for particle identification. The SSD has about 2.6 million channels.

4.1.2 The Time-Projection Chamber (TPC)

The TPC is the main tracking device of the ALICE detector and is located between radii of 0.85 m and 2.5 m (sensitive volume); it has a length of 5 m. It provides, in combination with the other central barrel detectors, the measurement of charged-particles, i.e. their momentum, particle identification, and production vertex. The TPC is able to track particles in $|\eta| < 0.9$ for full radial length[5] and up to $|\eta| < 1.5$ for 1/3 radial length (with reduced or no matching with the other detectors). Particles with a p_T from about 200 MeV/c (at nominal field) up to 100 GeV/c can be measured. The momentum resolution of the tracks is better than 2.5% for tracks with a momentum below 4 GeV/c. The material budget of the ITS and TPC is on average less than 11% of a radiation length.

Figure 4.5 shows a schematic picture of the TPC: it is a gas detector with a volume of 90 m^3 (the biggest TPC in the World), filled with a $Ne/CO_2/N_2$ gas mixture. A drift field of 100 kV stretches between the central electrode (at $z = 0$) and the two readout planes at $z = \pm 2.5$ m. A maximum of 160 clusters can be measured for a typical track which allows up to 20 000 tracks in one event to be reconstructed and identified. The

[4]Data-taking is organized in *runs*. A run is the collection of the data taken for a few hours under the same experimental conditions. Runs are marked with an incremental number.

[5]A track without full radial length traverses the TPC in a way that it leaves the TPC before reaching the outer radius; therefore it produces less tracking information than a track with full radial length.

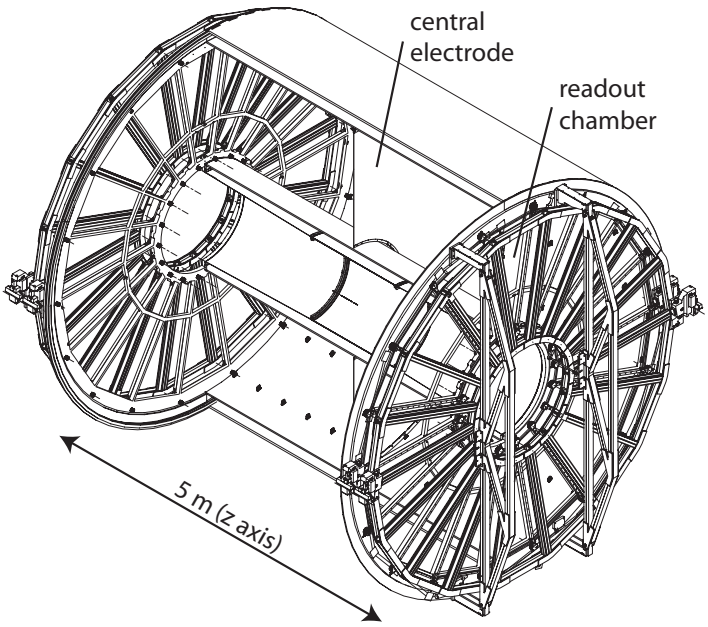

Figure 4.5: Schematic view of the TPC (adapted from [Aam08]).

readout plane consists of multi-wire proportional chambers that are mounted on the end-caps of both sides of the TPC. Electrons arriving from the drift volume have to pass a gating grid that is only opened upon a L1 trigger for the drift time interval (about 90 μs). The readout comprises about 560 000 channels. It is segmented into 18 sectors on each side (see Figure 4.5) and is insensitive at the sector boundaries, which results in a total insensitive area of about 10% [Aam08]. However, the reduction of the tracking efficiency is less than these 10% because the magnetic field bends tracks out of the insensitive region. This argument does not of course apply to high-p_T tracks.

The TPC is, due to its drift time of about 90 μs, the slowest detector in ALICE. This has to be taken into account for the trigger: once an event is accepted, usually no other event is measured with the TPC within the next 90 μs. A significant amount of pile-up[6] is expected: at a luminosity of $3 \times 10^{30}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ about 30 p+p interactions are detected together with the triggered event. Still the total occupancy is much lower than for Pb+Pb collisions; the tracks from pile-up events can be eliminated during the reconstruction because those point to different vertices than the vertex of the triggered event.

The large number of measured clusters allows the specific energy loss (dE/dx) of traversing particles to be calculated without being affected by the tails of the energy-

[6]Pile-up refers to the situation where more than one collision occurs during the readout time of a detector. These collisions can be in the same or in different bunch crossings.

loss (Landau) distribution. The *truncated mean* method is applied which uses only part of the clusters by skipping a fraction of clusters that have the largest values (and are thus likely to be from the tail of the distribution). The measured total energy loss is therefore nearly distributed like a Gaussian. The TPC aims at an energy resolution of 5.5% for tracks that have more than 140 clusters at low particle densities [ALI00]. Preliminary studies using data taken during the commissioning with cosmic rays show that this is achieved [Kal08]. The TPC provides a 3σ π/K- and K/p-separation in the region of $p_T \lesssim 1\,\mathrm{GeV}/c$, π/K-separation for $p_T \lesssim 0.5\,\mathrm{GeV}/c$, as well as good electron-pion separation up to a few GeV/c.

4.1.3 The Transition-Radiation Detector (TRD)

The TRD's task is to distinguish electrons from pions, especially at higher momenta above $1\,\mathrm{GeV}/c$. Furthermore, it contributes to the tracking of particles and acts as a trigger on high-momentum electrons. The detector is based on transition radiation (TR) which are photons with, in this case, wavelengths in the region of soft X-rays. TR occurs when a charged particle propagates through boundaries between media that have different dielectric constants. The probability for the creation of such a photon is linearly dependent on the particle's Lorentz factor γ; e.g. for particles with $p = 1\,\mathrm{GeV}/c$: $\gamma(e^{\pm})/\gamma(\pi^{\pm}) \approx 2\,000/7$. However, the overall probability to create TR at one media boundary is still low, thus many layers of media boundaries are used, so that on average more than one detectable X-ray photon is produced for particles with $\gamma > 1\,000$ [Aam08]. The achieved pion rejection is better than 100 for particles above $1\,\mathrm{GeV}/c$ at an electron efficiency of 90% [And04, Adl05, Wil09].[7]

The detector is located at radii from 2.9 m to 3.7 m. It is segmented into 18 sectors where each consists of six layers. Figure 4.6 shows one layer that combines a radiator, a drift chamber, and readout electronics. The radiator facilitates the production of TR. In the Xe/CO_2 gas mixture, TR is converted at the beginning of the drift region into an electron cluster which is subsequently detected. Average drift spectra for electrons and pions are shown in Figure 4.7. The readout electronics features 1.18 million channels.

A built-in tracklet processor combines the information from the six layers to form tracklets: these are used to identify high-momentum electrons which in turn provide a L1 trigger. Such a trigger is for example useful to increase the yield of Υs and high-p_T J/Ψs. The TRD needs a so-called pretrigger to wake up its electronics that is usually in standby to reduce the power consumption and thus heat production.

[7] A pion rejection of 100 at an electron efficiency of 90% means that while electrons are positively identified with 90%, only 1 pion among 100 is falsely identified as an electron.

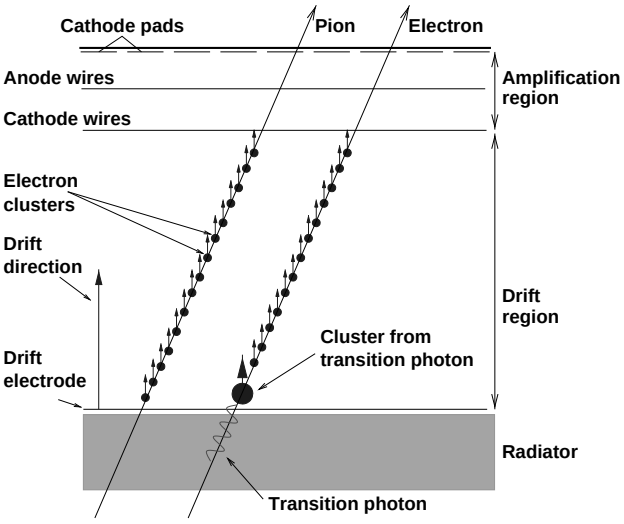

Figure 4.6: Schematic view of one TRD layer together with the clusters produced by an electron and a pion track.

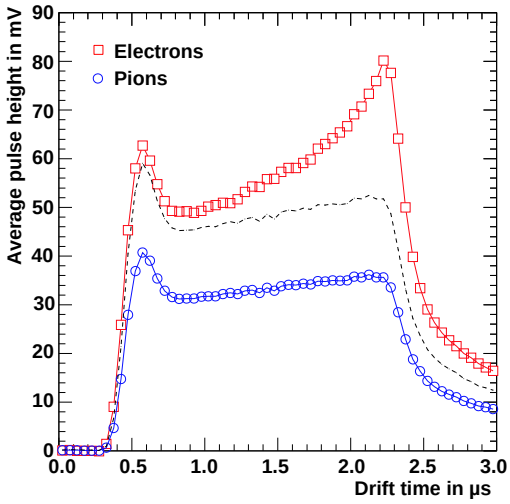

Figure 4.7: Average drift spectra of electrons (red squares) and pions (blue circles) in the TRD. The line indicates the drift spectrum of electrons without TR.

4.1.4 The Time-Of-Flight Detector (TOF)

The TOF detector's main task is to identify protons, kaons, and pions by measuring the time between the collision and the arrival of the particles in the TOF. The K/p separation up to 4 GeV/c and the π/K separation up to 2.5 GeV/c are better than 3σ. The TOF system provides the above-mentioned pretrigger signal to the TRD and an L0 trigger for ultra-peripheral collisions.

The detector consists of 18 sectors and is located at a radius of 3.8 m. The 140 m^2 large active area is a high-resolution array of so-called multigap resistive plate chambers. These are stacks of very thin structures (250 μm) featuring a high and uniform electric field and a $C_2H_2F_4$/i-C_4H_{10}/SF_6 gas mixture so that any traversing particle immediately triggers an avalanche. The setup achieves a very good time resolution of about 40 ps. Combined with other uncertainties, e.g. the uncertainty to determine the exact time of the interaction, the time of flight measurement for single particles has an overall resolution of better than 100 ps [Aam08]. The TOF detector has about 160 000 channels.

4.1.5 The Photon Spectrometer (PHOS)

The PHOS is a high-granularity calorimeter measuring photons. It allows for example the measurement of π^0 and η via their decay photons. For this purpose photons have to be discriminated against charged hadrons and neutrons which is partly performed by topological shower analysis. It features an excellent energy resolution, for example for 1 GeV-photons, σ_E/E is about 4% [ALI99b]. The PHOS can provide a L0 and L1 trigger.

The detector consists of an electromagnetic calorimeter of dense scintillating crystals (about $20X_0$) and detection cells made of lead-tungstate crystal ($PbWO_4$). It is located at a radius of 4.6 m and covers about 3.7% of phase space in the central region. A set of multi-wire proportional chambers in front of PHOS is used to reject charged particles, this part of the detector is called Charged-Particle Veto (CPV).

4.1.6 The ElectroMagnetic Calorimeter (EMCal)

The EMCal is a Pb-scintillator sampling calorimeter that measures photons, π^0, and η via their decay photons like the PHOS detector. It is, however, larger than PHOS with an acceptance of about 23% of phase space of the central region, but offers lower granularity and resolution. The detector is located approximately opposite to PHOS. It can provide a L0 and L1 trigger based on sums of deposited energy (towers) in sliding regions of the detector.

The EMCal has been added in a late stage to the experiment's design and therefore its construction only started in 2008.

4.1.7 The High-Momentum Particle Identification Detector (HMPID)

The HMPID is a proximity focusing Ring-Imaging Cherenkov (RICH) detector for particle identification of high-momentum hadrons. It extends ALICE's capability of π/K- and K/p-separation to 3 and 5 GeV/c, respectively, and therefore allows the inclusive measurement of charged particles within $1-5$ GeV/c. The detector's acceptance covers about 5% of the central region phase space. The detector consists of 10 m^2 of active CsI photocathode area which represents the largest scale application of a RICH.

4.1.8 The ALICE Cosmic Ray Detector (ACORDE)

ACORDE consists of 60 large scintillators that are used as L0 trigger on cosmic rays. The detection of single atmospheric muons and multi-muon events allows high-energy cosmic rays to be studied, which can provide insight into the energy region of the knee in the cosmic-ray spectrum[8]. Furthermore, cosmic-ray events are used for calibration and alignment. ACORDE has been used during the detector commissioning in 2007 and 2008. The rate of muons reaching the ALICE detector is about $4.5\,\mathrm{Hz/m^2}$. ACORDE's scintillators have been used before by the DELPHI (Detector with Lepton, Photon, and Hadron Identification) experiment.

4.2 Forward Detectors

4.2.1 The Photon Multiplicity Detector (PMD)

The PMD measures the multiplicity distribution of photons (e.g. decay products from π^0 and η) in the forward region ($2.3 < \eta < 3.7$, full azimuth). It consists of two gas proportional chambers. Between these a lead converter is located. The plane in front of the converter is used as a veto for charged particles while the information from the second plane is used to identify photons. The detector is positioned at a 3.64 m distance from the nominal interaction point. The PMD cannot be used as a trigger because of its slow readout.

4.2.2 The Forward Multiplicity Detector (FMD)

The FMD measures the charged-particle multiplicity over a large fraction of phase space, $-3.4 < \eta < -1.7$ and $1.7 < \eta < 5.0$, both in full azimuth. The detector is composed of silicon strips located in five rings at $z = 3.2\,\mathrm{m}$, $0.83\,\mathrm{m}$, $0.75\,\mathrm{m}$, $-0.63\,\mathrm{m}$ and $-0.75\,\mathrm{m}$. Due to its slow readout ($> 1.2\,\mu\mathrm{s}$) it cannot be used as a trigger.

4.2.3 The V0 detector

The information from the V0 detector is used as minimum-bias trigger, to reject beam-gas events, and to provide a pretrigger to the TRD. It consists of two arrays of seg-

[8]The flux of cosmic rays as a function of the cosmic-ray energy shows a power-law behavior. The slope changes between $10^{15} - 10^{16}\,\mathrm{eV}$, which, due to the shape in a double-logarithmic scale, is called *knee*.

mented scintillator counters that are located at $z = 3.4\,\mathrm{m}$ ($2.8 < \eta < 5.1$) and $-0.9\,\mathrm{m}$ ($-3.7 < \eta < -1.7$). The time resolution is about 1 ns [Car04] which allows beam-gas events that occurred outside of the nominal interaction region to be identified (discussed in Section 5.1).

4.2.4 The T0 detector

The T0 ('time 0') detector measures the collision time with a precision of 25 ps. This information is used as a time reference for the TOF detector and to determine the vertex position with a precision of about 1.5 cm. If the vertex position is inside a window where interactions are expected an L0 trigger is issued. A vertex position outside the region where collisions should appear is used as a beam-gas rejection signal. Furthermore, the T0 detector can also send a pretrigger to the TRD.

The detector consists of two units that each comprises twelve Cherenkov counters with quartz radiators. The units are located around the beam pipe at a distance of 3.75 m (positive z) and 0.73 m (negative z) from the nominal interaction point.

4.2.5 The Zero-Degree Calorimeter (ZDC)

The ZDC provides an estimate of the impact parameter of heavy-ion collisions by the measurement of the number of spectator nucleons which is related to the energy carried forward, i.e. in beam direction. The detector is located on both sides of the detector, at a distance of 116 m from the nominal interaction point. The measurement is performed by two calorimeters, one for neutrons (called ZN, $|\eta| < 8.8$) and one for protons (called ZP, $6.5 < |\eta| < 7.5$). At this distance from the interaction point neutrons and protons are separated by the magnets in the beam line. When they are not in use, the calorimeters are moved out of the beam line by a lifting platform to reduce their exposure to ionizing radiation. The measurement is complemented by an electromagnetic calorimeter (called ZEM, $4.8 < \eta < 5.7$) which measures the total forward energy at $z = 7.25\,\mathrm{m}$. This allows the distinction of central and very peripheral heavy-ion events: both deposit low energy in the forward ZDCs. In a central collision only a few spectators are emerging in forward direction; in peripheral collisions big fragments are produced that do not reach the ZDCs. The ZDC can provide a L1 trigger.

4.3 The MUON Spectrometer

The task of the MUON spectrometer is to measure the complete spectrum of quarkonia (J/Ψ, Ψ', Υ, Υ', Υ'') with a mass resolution that is good enough to separate these

states as well as the ϕ meson. The separation of the Υ states requires a resolution of $100\,\mathrm{MeV}/c^2$ in the $10\,\mathrm{GeV}/c^2$ invariant mass region. Furthermore, the production of open charm and beauty can be studied.

The spectrometer is located on the C side of the ALICE experiment. It accepts particles in $-4 < \eta < -2.5$ and has full azimuthal coverage for muons with $p > 4\,\mathrm{GeV}/c$. This cut-off is due to the fact that to reach the spectrometer muons first have to pass through the front absorber made of carbon, concrete, and steel. Successively they are measured by five tracking stations with two planes each made of very thin, high-granularity, cathode strip tracking stations. A dipole magnet with an integrated magnetic field of 3 Tm is located outside of the L3 magnet to allow the muons' momenta to be reconstructed. Two tracking stations are located in front of the dipole magnet. One tracking station is in its center; two are positioned behind the magnet. An iron wall of 1.2 m acts as a further muon filter after which two trigger stations with two planes each of resistive plate chambers are located. The whole spectrometer is shielded by means of a dense absorber tube against particles emerging from the beam pipe.

4.4 The Data Acquisition (DAQ)

The tasks of the ALICE DAQ system are the assembly of event fragments from individual subdetectors into complete events (event building) as well as buffering and export of assembled events to permanent storage. The DAQ is designed to process a data rate of up to 1.25 GB/s in heavy-ion runs. Event building is done in two steps. Data from the subdetectors is received by *Detector Data Links* (DDLs) on *Local Data Concentrators* (LDCs). The LDCs assemble the data into sub-events that are then shipped to *Global Data Collectors* (GDCs). A GDC receives all sub-events from a given event and assembles them into a complete event. Subsequently, these events are stored on a system called *Transient Data Storage* (TDS) that provides at present 45 TB of data storage. The export of the data and further processing is described below in Section 4.6.1. The DAQ has at present 83 LDCs and 43 GDCs while the fully equipped DAQ setup will comprise 200 LDCs and 60 GDCs [Cha08].

ALICE can simultaneously take data in several *partitions*, where each partition consists of a set of subdetectors. Obviously a given subdetector can only be active in one partition at a time. The active subdetectors in a given partition are grouped into clusters for which triggers can be defined. Therefore, upon a trigger only a subset of the whole partition may be read out. Furthermore, a triggering detector does not have to be necessarily part of the partition.

4.5 The ALICE Trigger System

ALICE has a two-layer trigger architecture [ALI04]. The low-level trigger is a hardware trigger called Central Trigger Processor (CTP). The High-Level Trigger (HLT) is implemented as a pure software trigger. The CTP combines inputs from different trigger sources, i.e. the various subdetectors. These inputs are single signals like a hit in the detector. At most, inputs can be fast calculations that are performed in the subdetectors. An example is the above-mentioned tracklet processor in the TRD. The HLT allows the implementation of sophisticated logic for the triggering. In contrast to the CTP, which governs the readout of the subdetectors, the HLT receives a copy of the data read out from the subdetectors and processes it.

4.5.1 The Central Trigger Processor (CTP)

The hardware trigger combines the trigger signals of the various subdetectors to decide if an event is accepted which means that it is read out and written to disk. Several trigger levels reduce the event rate depending on the input signals. The first level, called L0, is delivered after 1.2 μs, the second, called L1, after 6.5 μs. The final trigger, L2, is delivered after 100 μs, upon completion of the drift time in the TPC. Only after an L2 trigger the event is finally stored. Another task of the hardware trigger is to issue a pretrigger to wake up the TRD electronics which is needed in less than 900 ns after the interaction.

A past-future protection ensures that events are not superimposed by too many pile-up collisions. The readout times of the different detectors vary significantly, therefore the window in which pile-up is recognized depends on the detectors that are part of the current partition as well as on the collision system. For example in Pb+Pb collisions a reasonable condition for partitions that contain the TPC is the following: in a window of $\pm 90\mu$s (the TPC drift time) around the collision time of the event a maximum of four additional peripheral events and no additional semi-central event is allowed [Aam08]. Different conditions are applied for p+p collision where pile-up is always present due to the higher luminosity. However, in this case more pile-up is acceptable due to the much lower particle densities.

The trigger logic acts upon numerous inputs: up to 24 L0, 24 L1, and 12 L2 input signals. Out of these inputs up to 50 trigger classes can be defined. However, not all the inputs can be connected in an arbitrary way, for more information see [ALI04].

The rates of different trigger classes are very different. By definition minimum-bias triggers have the highest rate, other triggers that look for rare signals have much lower

rates. Therefore, downscaling factors can be applied to the trigger classes individually, i.e. only every nth event fulfilling the trigger condition is read out. The total recording rate is limited by the maximum bandwidth of data that can be recorded to disk and tape. To prevent losing precious events due to the fact that no space is available on the temporary memory and disk buffers in a moment where a trigger that looks for a rare signal occurs, the trigger system implements an event prioritization scheme. Therefore, trigger classes are grouped into common triggers and rare triggers. In the case that the utilization of the temporary storage is above a certain value (high-water mark) only rare triggers are accepted; as soon as the utilization drops below a given low-water mark all triggers are accepted again. This scheme significantly increases the acceptance of rare events.

The minimum-bias triggers available in ALICE and their efficiency are discussed in Section 5.1.

4.5.2 The High-Level Trigger (HLT)

ALICE's software trigger, called HLT, is a farm of multiprocessor computers. The aim is about 1 000 PCs processing the data in parallel allowing an online analysis of the events. A trigger decision is derived from much more complete information than is available for the hardware trigger. Therefore, it allows for more sophisticated triggers. Examples include triggers on high-energy jets or on muon pairs. Furthermore, the HLT can significantly reduce the event size by selecting regions of interest (partial readout of subdetectors) and by further compression of the data.

The HLT receives a copy of the raw data and performs per detector reconstruction, partly aided by hardware coprocessors. Subsequently, the trigger decision is based on the global reconstructed event. In the same step a region of interest can be selected. In the last optional step, if the trigger decision is positive, the data is compressed. The trigger decision, partial readout information, compressed data, and the reconstruction output is sent to LDCs and subsequently processed by the DAQ. In terms of the overall DAQ architecture, data sent by HLT is treated like stemming from a subdetector.

4.6 The ALICE Offline Software Framework

The data production of the LHC experiments (about 10 – 15 PB per year) is at a new scale compared to any previous experiment. In ALICE, an average Pb+Pb event will have a size of about 13.75 MB; on average a p+p event is about 1.1 MB. For a standard running year, of the order of 10^9 p+p events and 10^8 Pb+Pb events are

expected yielding a total raw data volume of 2.5 PB. The data taken with cosmics in 2008 amounts to about 300 TB. Two thirds were taken in so-called global runs with several participating subdetectors – a situation similar to real data-taking. The average size of the reconstruction output is 3 MB for a Pb+Pb event and 40 kB for a p+p event. This only includes high-level information needed for user analysis. Examples are the event-vertex position, reconstructed track parameters, and PID information (see also Section 4.6.3 below). The overall dataflow is discussed in detail in the subsequent Section 4.6.1.

The required computing resources for the reconstruction and analysis of the raw data as well as the production of simulated events needed for the understanding of the data exceed the computing power of single institutes and even centers like CERN. Therefore, institutes that are part of the collaboration also provide storage and computing resources. At present 80 centers contribute to ALICE's computing resources. Distribution of the data for reconstruction and analysis cannot be performed manually and this led to the need for an automated system. The concept of Grid [Fos04] was identified as a solution. ALICE uses the *ALICE Environment* (AliEn) system as a user interface to connect to a Grid composed of ALICE-specific services that are part of the AliEn framework and basic services of the Grid middleware installed at the different sites. AliEn is briefly described in Section 4.6.2; more information can be found in [Bag08].

A dedicated framework called AliRoot enables simulation and reconstruction of ALICE events to be performed. It is also the basis for any analysis performed on the data. The AliRoot framework is described in Section 4.6.3. ALICE offers to its users a system called CERN Analysis Facility (CAF) which enables the parallel use of a computing cluster to perform analysis. The CAF system is introduced in Section 4.6.4.

4.6.1 Dataflow

The raw data taken by the subdetectors has to be processed before it is available in the form of reconstructed events for further analysis. This happens in several stages and is illustrated in Figure 4.8. Data originating from the subdetectors (denoted by 1 in Figure 4.8) is processed by LDCs, global events are built by GDCs (2); see Section 4.4 for details. The so-called *publish agent* registers the assembled events into the AliEn system (3) and ships them to the CERN computing center where they are stored first on disks (4) and then permanently on tapes (5) by the *CASTOR* system [Dur04].

During data-taking the subdetectors also produce *conditions data* that is relevant for the calibration of individual detector signals. Conditions data provides information about the detector status and environmental variables during data-taking. Examples

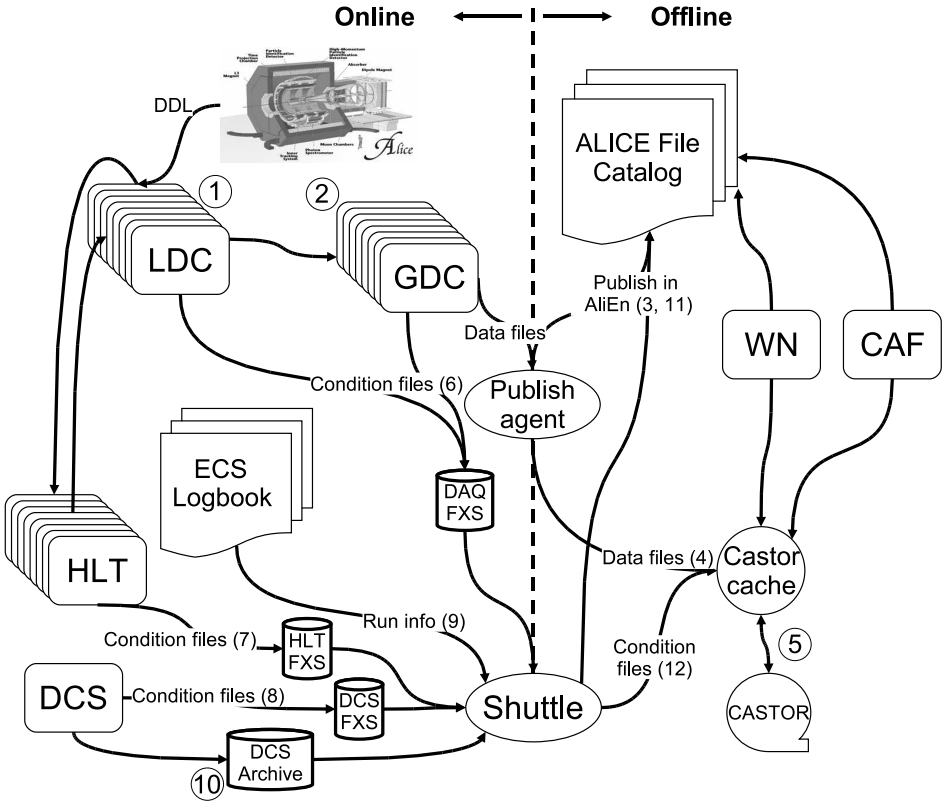

Figure 4.8: Global view of ALICE's data flow (Figure adapted from [Aam08]).

---

are inactive and noisy channel maps, distributions that describe the response of a channel, temperatures and pressure in a detector, and detector configuration. Many of the conditions data could in principle be calculated from the raw data and extracted offline after data-taking. However, such an approach would require an additional pass over the raw data before the reconstruction which is not feasible due to the limited computing resources. Therefore, conditions data is already extracted during data-taking.

Conditions data is produced by special programs that process the raw data stream and extract the needed values. These programs work in the realm of DAQ, DCS (Detector Control System), and HLT and store their output on so-called *File eXchange Servers* (FXS) (6-8 in Figure 4.8). A dedicated program called *Shuttle* collects these outputs and makes them available to the reconstruction. Furthermore, it retrieves information about the run from the ECS logbook (9) and collects continuously monitored values that are written by DCS into the *DCS Archive* (10). After processing the data, the Shuttle registers the produced condition files in AliEn (11) and stores the data in CASTOR (12). The Shuttle framework, which was developed as part of the thesis work, is described in Appendix D.

With the registration of the raw and conditions data the transition from the *online* to the *offline* world has taken place. Online denotes all actions and programs that have to run in real time. Offline processing is the subsequent step, like for example event reconstruction, which is executed on worker nodes (WN) of Grid sites located around the Globe.

4.6.2 The AliEn Framework

The Grid paradigm implies the unification of resources of distributed computing centers, in particular computing power and storage, to provide them to users all over the World. It allows computing centers to offer their resources to a wider community. This allows resources in large collaborations to be shared.

Software that implements the Grid concept is called Grid middleware. ALICE has developed a Grid middleware called AliEn [Bag08] since 2001. An ALICE user employs AliEn to connect to the ALICE Grid which is composed of a combination of general services that are provided by many Grid middleware solutions and ALICE-specific services provided by AliEn. Part of the ALICE Grid is a global file catalog that is a directory of files in storage elements distributed over the Globe, automatic matching of jobs for execution to a suitable location in one of the connected sites, a shell-like user interface, and API[9] services for the ROOT framework [Bru97].

Currently the ALICE Grid consists of about 80 sites located in 21 countries. The system has been tested extensively with up to 10 000 jobs running concurrently over several weeks. The simulated data used in this thesis has been produced at these sites. Figure 4.9 shows a map of ALICE's Grid sites.

4.6.3 The AliRoot Framework

AliRoot [Aam08, ALI09] is the offline framework for simulation, alignment, calibration, reconstruction, visualization, quality assurance, and analysis of experimental and simulated data. It is based on the ROOT framework. Most of the code is written in C++ with some parts in Fortran that are wrapped inside C++ code.

The AliRoot development started in 1998 and it has been extensively used for the optimization of the experiment's design. It has been used for large-scale productions, so-called Physics Data Challenges (PDCs), where millions of events are produced. These have been used to estimate the physics performance of ALICE (see [Car04, Ale06]).

[9] An Application Programming Interface (API) of a program is a set of publicly available functions that can be used to access its functions.

Figure 4.9: ALICE Grid sites.

The figure shows the computing centers that contribute to the ALICE Grid. Most of them are located in Europe, however, some are in other continents which can be seen in the inset showing the world map in the bottom right corner. Figure taken from [ALI08].

Such events are also used to develop analysis procedures and to estimate the associated systematic errors, as is performed in this thesis. Finally, AliRoot is used to reconstruct events that occurred in the detector.

For event simulation the framework provides the following functionality:

- **Event generation**. A collision is simulated by an event generator that is interfaced with AliRoot (e.g. Pythia [Sjo01], Phojet [Eng95], or HIJING [Gyu94]); this step produces the *kinematics tree* containing the full information about the generated particles (type, momentum, charge, production process, originating particle, and decay products).

- **Transport**. The particles are propagated through the detector material which is modeled as realistically as possible. In this process, particles can interact with matter, decay, and create additional particles. Naturally, these particles have to be propagated through the detector as well. The total number of particles after the transport is significantly larger than the number of particles created in the initial generation step. During this process all interactions of particles with sensitive detector parts are recorded as *hits* that contain the position, time, and energy deposit of the respective interaction. Furthermore, *track references* are stored that can be used to follow a track's trajectory, mainly needed for the debugging of the reconstruction algorithms. Programs that perform the transport and are interfaced with AliRoot are Geant3 [Bru78], Geant4 [Ago03], and Fluka [Fas03].

- **Digitization**. If a particle produced a signal in a sensitive part (hit), the corresponding digital output of the detector is stored as a *summable digit* taking into account the detector's response function. Possible noise is then added to the summable digit and it is stored as a *digit*. Summable digits allow events to be merged without duplication of noise. In the last step, the data is stored in the specific hardware format of the detector (raw data).

At this stage the raw data corresponds to the signals that would be produced by an interaction of the same kind within the detector. The subsequent reconstruction is identical, both for simulated as well as real events. It consists of the following steps:

- **Cluster finding**. Particles that interact with the detector usually leave a signal in several adjacent detecting elements or in several time bins of the detector. In this step these signals are combined to form *clusters*. This allows the exact position or time of the traversing particle to be determined and reduces the effect of random noise. Overlapping signals from several particles in a single cluster are unfolded. This step is performed for each subdetector where due to the different nature of the subdetectors the implementations vary significantly.

- **Track reconstruction**. The clusters are combined to form tracks that allow the track curvature and energy loss to be calculated with the aim of determining their momentum and particle type. The tracking is a global task as well as an individual procedure per detector. The global central barrel tracking starts from track seeds in the TPC which are found by combining information from a few outermost pad rows under the assumption that the track originated from the primary vertex. Tracks are then followed inwards using a procedure called the Kalman filter [Bil89]: in each step the track, i.e. the track parameters and the covariance matrix, is propagated to the next pad row. The covariance matrix is

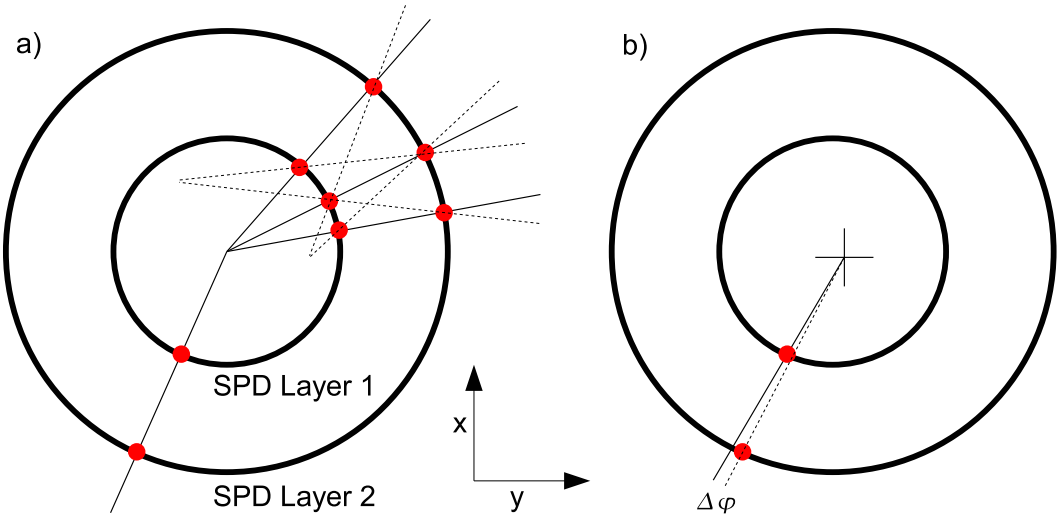

Figure 4.10: SPD tracklet finding.
The left panel shows the primary-vertex finding. Straight lines are formed through all two-cluster combinations (not all shown), the vertex is found where most intersect. In the right panel the subsequent tracklet finding is shown, combinations are valid where the vertex is in line with a cluster in the first and in the second layer. The difference in φ between the vertex and the first and second cluster has to be below a threshold and is a quality parameter of the tracklet.

updated adding a noise term that represents the information loss by stochastic processes such as multiple scattering and energy-loss fluctuations. If a cluster is found that fits to the track, it is added to the track, updating its parameters and the covariance matrix. Afterwards the same procedure is repeated by starting the seeding closer to the collision point. In a final step all clusters already associated to tracks are removed and the procedure is repeated without requiring that the seeds point to the primary vertex. The result, the so-called *TPC-only tracks* to which only TPC information contributed, is saved in the reconstruction output. Subsequently, these tracks are complemented with information from the ITS, TRD, and TOF as well as HMPID and the CPV of PHOS if the track is in their acceptance which produces so-called *global tracks*. Tracks can also be formed out of information from the ITS only. Tracks are represented by the parameters y, z, $\sin\varphi$, $\tan\lambda$, and $1/p_T$, see Appendix B for more details.

Among the track finding in single detectors is the SPD tracklet finding, illustrated in Figure 4.10. The event vertex as well as the tracklets are reconstructed by forming straight lines out of a cluster in each of the two SPD layers. The event vertex is reconstructed where most of these lines intersect. Lines that point to the vertex are identified as tracklets. A tracklet is represented by η, φ, $\Delta\varphi$, and

the primary-vertex position because tracklets originate by construction from the vertex. The quality parameter $\Delta\varphi$ is defined by $\Delta\varphi = \varphi_1 - \varphi_2$ where φ_1 (φ_2) is the azimuthal angle between the event vertex position and the cluster in the first (second) layer.

More information about the track finding can be found in [Ale06].

- **Primary-vertex reconstruction**. Various information is used to find the primary-vertex position of the interaction. Examples of information, each of which is sufficient to produce a vertex position, are clusters in the SPD, tracks in the TPC, and global tracks. When a vertex position is found the tracks are constrained to it: the vertex position is used as an additional point to estimate the track parameters. The TPC-only tracks are constrained with the vertex position found with TPC-only tracks while the global tracks are constrained with the vertex position found with global tracks. Of course this constraint is only used for tracks that actually pass in vicinity of the vertex.

- **Secondary-vertex reconstruction**. Tracks are combined to find secondary vertices in order to reconstruct decayed particles like $\Lambda^0 \rightarrow p\pi$ and photon conversions. For this purpose, opposite-sign tracks that originate sufficiently far away from the primary vertex are combined. If the closest approach and the topology of the two tracks is consistent with a decay, the pair is accepted as a potential secondary vertex.

The output of the reconstruction is called Event-Summary Data (ESD) which contains only high-level information such as the position of the event vertex, parameters of reconstructed charged particles together with their PID information, positions of secondary-vertex candidates, parameters of particles reconstructed in the calorimeters, and integrated signals of some subdetectors. This data is further reduced to Analysis-Object Data (AOD) format. These smaller-sized objects contain only information needed for the analysis. Therefore, the transformation procedure may already contain a part of the analysis algorithm, for example track selection. Several AODs, focusing on different physics studies, can be created for a given event.

4.6.4 The CERN Analysis Facility (CAF)

The processing of large samples of data is performed on the Grid, utilizing the previously introduced AliEn framework. To allow fast processing of medium-sized data samples, a system called CERN Analysis Facility (CAF) was set up as part of the thesis work. Contrary to the batch-type approach of the Grid, it allows interactive

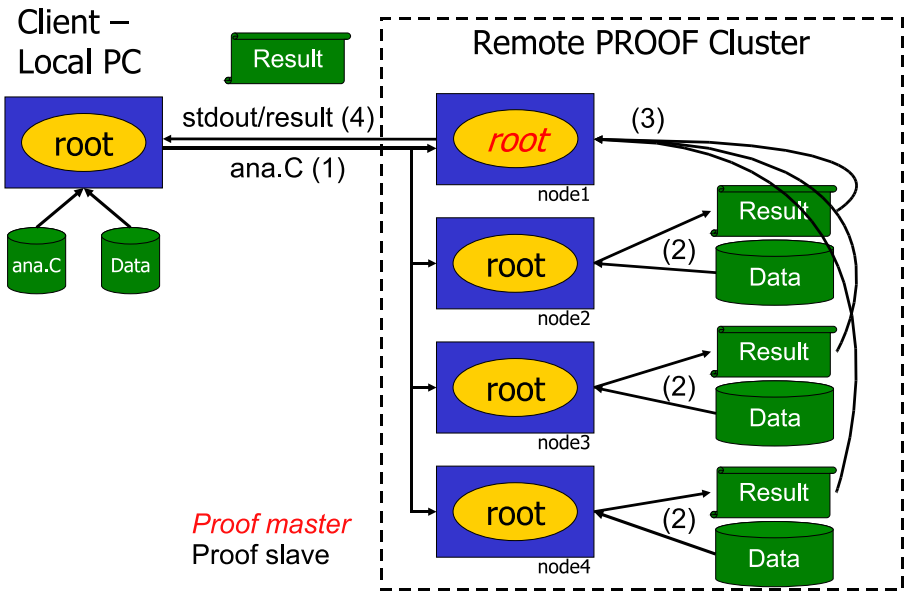

Figure 4.11: Schema of the PROOF system.

processing, thus the execution time is minimized and many development cycles are possible in a short period of time. The system's main tasks are prompt analysis of p+p data, pilot analysis of Pb+Pb data, fast event reconstruction, and above all calibration and alignment. A fraction of the total data recorded by the experiment as well as some simulated data will be available on the CAF. After successful prototyping, the analysis code can be sent as a Grid job to subsequently process larger sets of data.

The Parallel ROOT Facility (PROOF) [Bal03] enables interactive parallel data processing on a computing cluster. It is part of the ROOT framework. The system is particularly suited to process events produced by high-energy physics experiments: events can be processed in an arbitrary order and results obtained in parallel can be summed up after processing (event-based parallelism). Figure 4.11 shows a schematic view of the system. A user running a ROOT session on a client connects to a *PROOF master* node which in turn opens a ROOT session on each *PROOF worker* node. The user sends a query that consists of the analysis code and the name of a dataset known to the system that is to be processed (step 1). The master node assigns data fragments to each worker node which are then processed (step 2). The data is assigned such that data local to the worker node is processed first, then non-local data, if remaining. After processing, the results are merged on the master node (step 3), and returned to the user (step 4).

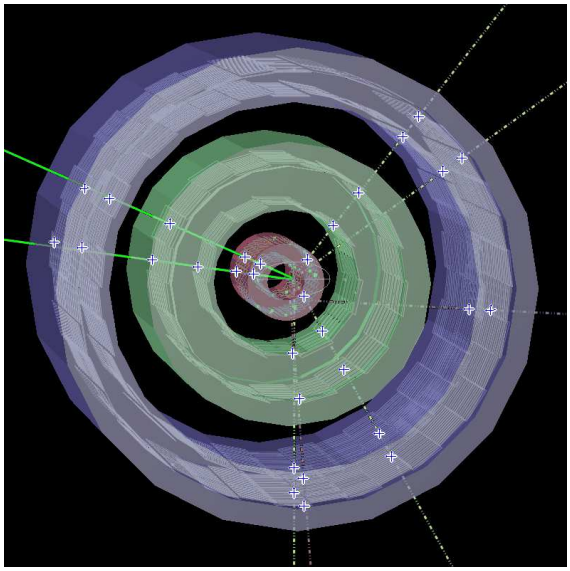

Figure 4.12: Interaction of a beam stray particle with the SPD detector.
Real data event recorded on the 11.09.2008 (run 58338).

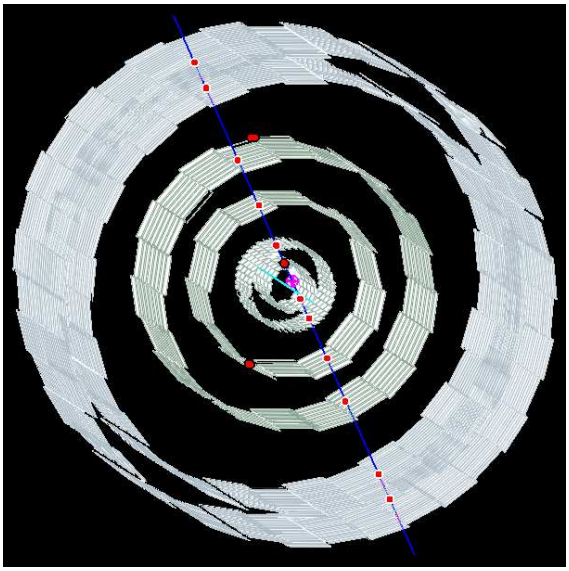

Figure 4.13: Cosmic-ray track used for alignment in the ITS.
Real data event recorded on the 20.09.2008 (run 60305).

The CAF has been available to ALICE users since May 2006. The PROOF system has been installed and an automatic staging solution was developed. It stages files residing on storage elements connected to the AliEn Grid. At present (January 2009) the system has 120 CPU cores and 32.5 TB of disk space local to the cluster. Since its introduction hundreds of users have exercised the system and many of them have been trained in regular tutorials that take place every 1 – 2 months. The system has been well received and an increasing number of users (at present about 50) use the system regularly. More information about the technical concept, the staging system, and monitoring can be found in [Gro08].

4.7 ALICE Startup Configuration

In 2008, ALICE took cosmic-ray data and was ready to detect first collisions. At this stage most subdetectors had finished their complete installation except for the TRD (4 out of 18 supermodules were installed), PHOS (1 out of 5 PHOS modules was installed without CPV), PMD (25% were installed), and EMCal, whose construction has only started in 2008.

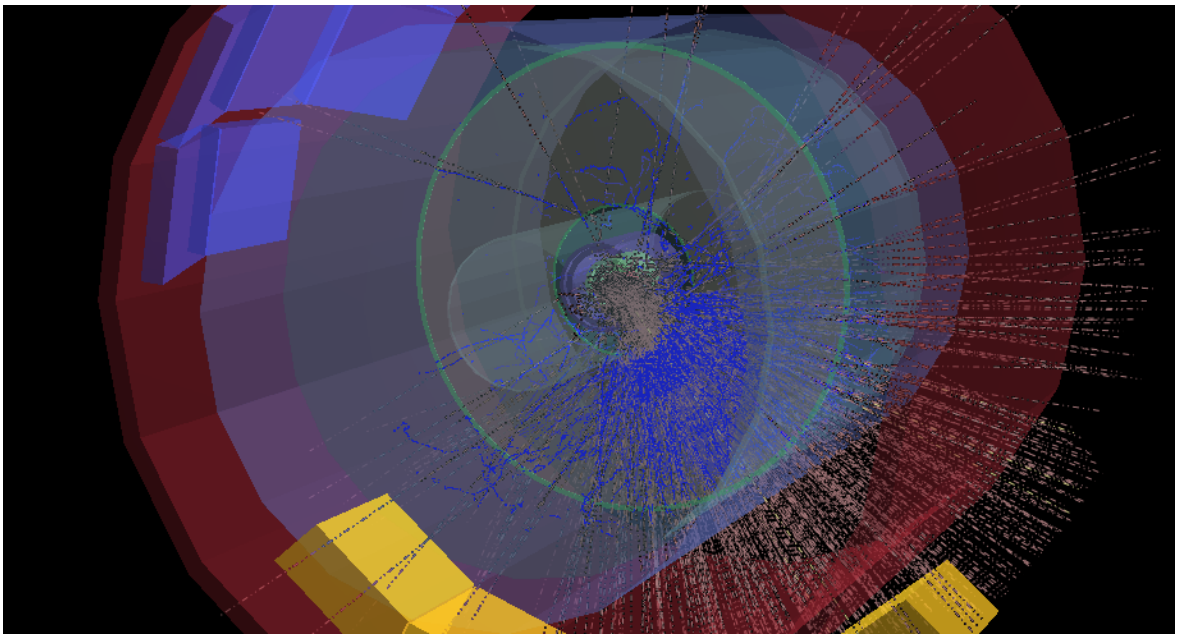

Figure 4.14: Particle shower in the TPC.

Real data event recorded on the 02.09.2008 (run 55604).

In 2008, millions of cosmic-ray events and some events during the circulating LHC beams were taken. Figure 4.12 shows an interaction of a stray particle of a circulating LHC beam with the first layer of the SPD. Figure 4.13 and Figure 4.14 show cosmic-ray events in the ITS and TPC, respectively. The shower in the TPC was caused by a high-energy cosmic-ray interaction in the muon absorber.

Further modules are expected to be installed before data-taking starts in 2009. Details about the progress and the current planning can be found in [Tau09].

4.7.1 Alignment Status

The subdetectors that each consist of many components are not located exactly at their nominal positions due to the limited precision during mounting and due to deformations caused by other components. A process called survey determines the detector positions with fiducial marks that have been added to many detector components at well defined places. Digital images are taken from various angles of the setup and the exact positions are calculated. This method achieves a precision of 1 mm when it is performed in the ALICE pit, and somewhat better for measurements done in the lab while assembling a detector. Further alignment has been performed using events that contain tracks

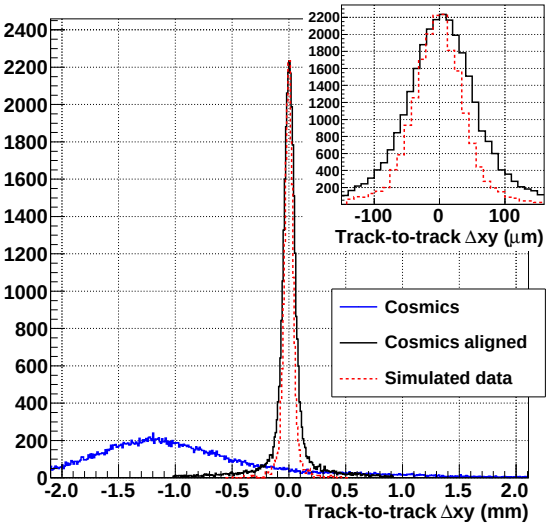

Figure 4.15: SPD alignment with cosmic-ray tracks.
The figure shows the track-to-track distance (see text) between cosmic-ray tracks reconstructed in the lower and upper half of the SPD. The distribution before (blue solid) and after (black solid) alignment is shown, as well as the distribution from simulated data without misalignment (red dashed). The inset in the top right shows a zoom in the central region. The simulated data is scaled to the same maximum value as the aligned distribution. Data points from [Bom09b].

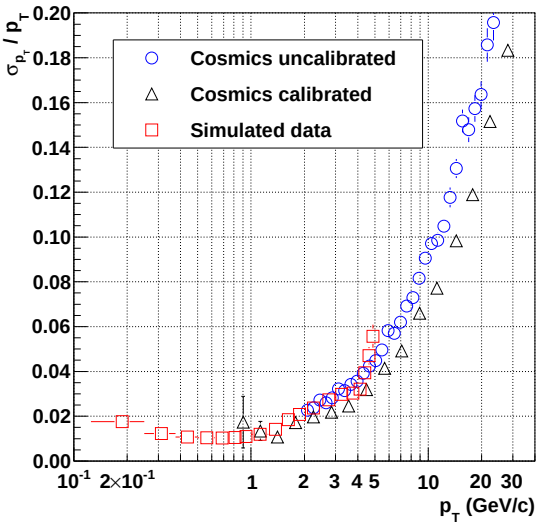

Figure 4.16: p_T resolution of the TPC.
The figure shows the p_T resolution determined from cosmic-ray tracks before (blue circles) and after (black triangles) preliminary calibration and alignment. Additionally, the resolution from the simulated data used in this thesis is shown (red squares). Cosmic-ray data points from [Iva08].

produced by cosmic rays. For early measurements no other information for alignment is available.

SPD

The SPD has been aligned with about 55 000 cosmic muons using the Millepede software [Blo02a] which performs a global χ^2-minimization of the residuals of a large number of tracks [Bom09b]. It is performed in a hierarchical way, starting with sectors, then half-staves, modules, and finally aligning the whole SPD barrel with respect to the next

layer of the SSD.[10] The same cosmic-ray tracks allow the quality of the alignment to be checked because each cosmic-ray track is reconstructed twice, once in the upper half of the detector and once in the lower half. Both tracks appear to originate from the center of the detector for the reconstruction software. The track parameters of these two tracks can be compared, in particular the track-to-track distance, i.e. Δxy in the direction normal to the tracks at $y = 0$. Figure 4.15 shows the Δxy distribution of cosmic-ray tracks before and after the alignment procedure, as well as that of simulated data with ideal geometry. The corresponding resolutions are $52\,\mu$m and $43\,\mu$m for cosmic-ray data after alignment and simulated data, respectively. The difference indicates the effect of the residual misalignment. In z-direction, the residual misalignment has less effect because the expected spatial resolution is anyway much lower (about $100\,\mu$m [Car04]). A further possibility is to compare the positions of clusters in areas where sensitive areas in the same layer overlap. This yields the spatial resolution in $r\varphi$-direction of clusters to be about $14\,\mu$m compared to $11\,\mu$m in simulations with the ideal geometry. A residual misalignment for clusters of about $8\,\mu$m can be concluded. The obtained resolution is about 25% higher than the theoretical achievable value. About 85% of the SPD are aligned, missing are sectors in positions that are not favored by cosmic-ray tracks, i.e. in the region $\varphi \approx 0$ and $\varphi \approx \pi$.

TPC

Preliminary calibration and alignment of the TPC has been performed with laser tracks [Iva08]. The resolutions are extracted from cosmic-ray tracks in the same way as it is done for the SPD. The reconstructed tracks in the upper and lower half of the TPC are compared. This allows the residual misalignment and in particular the p_T and dE/dx resolutions to be extracted. The resulting track-to-track resolution is smaller than 0.1 cm in z and smaller than 0.25 cm in xy (rφ) direction. The p_T resolution at $1\,\mathrm{GeV}/c$ is 1 – 2% and 6 – 7% at $10\,\mathrm{GeV}/c$. This is shown in Figure 4.16 before and after the preliminary calibration and alignment. Also shown is the p_T resolution extracted from the simulated data used in this thesis (see Section 5.5) which is close to, however below, the resolution from cosmic rays after calibration and alignment. The dE/dx resolution was evaluated to 5.7% after a first calibration with radioactive Krypton [Kal08].

[10] The SPD is structured in 10 sectors, each sector comprises 2 staves in the inner layer and 4 staves in the outer layer. Each stave consists of four modules (also called ladders).

Chapter 5

Event and Track Selection

This chapter describes the minimum-bias triggers available in ALICE and evaluates their efficiencies. It is discussed how a bunch-crossing trigger is utilized in early data-taking. Event and track selection criteria are introduced and evaluated. These are used in the multiplicity measurements discussed in the subsequent two chapters. The specific simulated datasets used in this and the subsequent chapters are described in the last section of this chapter.

5.1 Minimum-Bias Triggers

ALICE's trigger system has been introduced in Section 4.5. Various triggers can be configured simultaneously and more complex trigger patterns can be implemented in the HLT. Minimum-bias triggers are designed to trigger on all inelastic interactions occurring in the detector, even when the momentum transfer between the incoming particles is small or when only very few final-state particles are produced. These triggers should impose the least possible bias on the triggered sample compared to all inelastic collisions, hence their name. Thus minimum-bias triggers are the right choice for analyses discussed in this thesis that produce distributions considering all inelastic collisions.

In ALICE, information from the V0 detector (see Section 4.2.3) and the SPD (see Section 4.1.1) are combined to form a set of minimum-bias triggers:

- `MB1` = (`V0_OR` or `SPD_OR`) and not `V0_BG`;
- `MB2` = `V0_OR` and `SPD_OR` and not `V0_BG`;

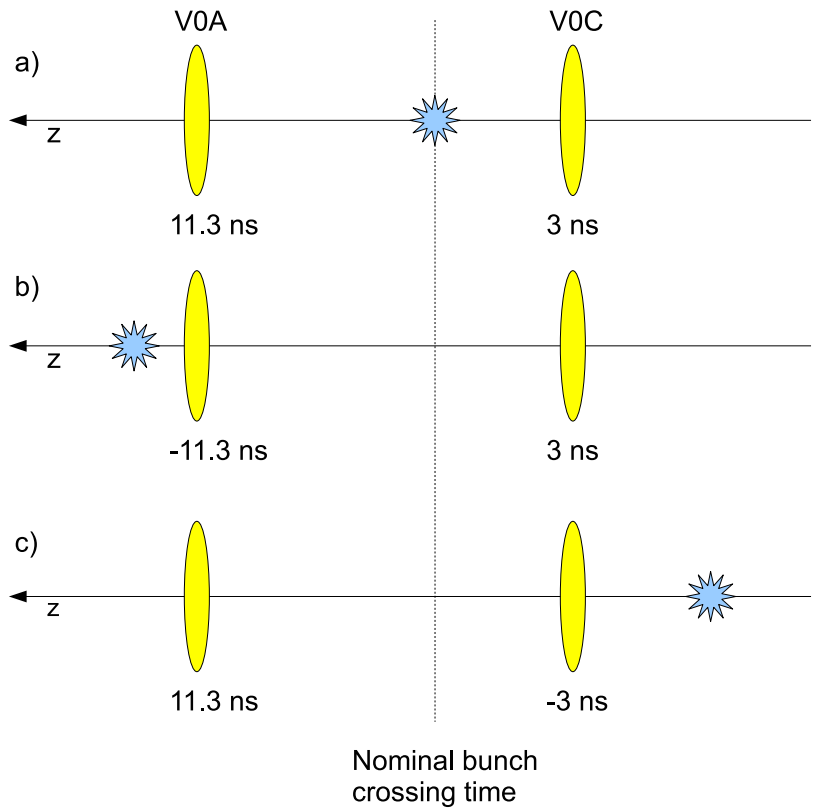

Figure 5.1: Beam-gas detection in the V0.
The time resolution of the V0 allows beam-gas and beam-halo events that occur outside of the detector to be identified: the arrival times of the first particle on both sides of the V0 are shown with respect to the nominal bunch-crossing time for: a) a normal interaction; b) and c) beam-gas events on either side of the V0.

- `MB3` = `V0_AND` and `SPD_OR` and not `V0_BG`,

where:

- `V0_OR` requires a signal in either of the two V0 sides;
- `V0_AND` requires signals on both sides of the V0;
- `V0_BG` indicates that a beam-gas or beam-halo collision (defined below) was detected by the V0 which utilizes the timing of the collision (see Figure 5.1);
- `SPD_OR` requires at least one chip that measured a signal in the SPD, i.e. the first two layers of the ITS. See the definition of the *fast OR* trigger in Section 4.1.1.

Process type	Pythia $\sqrt{s} = 900\,\text{GeV}$	Pythia $\sqrt{s} = 10\,\text{TeV}$
Non-diffractive (ND)	100.0	100.0
Single-diffractive (SD)	76.6	71.6
Double-diffractive (DD)	91.6	86.2
Non single-diffractive (NSD)	98.7	97.9
Inelastic (INEL)	93.8	93.0
Process type	**Phojet $\sqrt{s} = 900\,\text{GeV}$**	**Phojet $\sqrt{s} = 10\,\text{TeV}$**
Non-diffractive (ND)	100.0	100.0
Single-diffractive (SD)	85.7	78.0
Double-diffractive (DD)	98.0	93.9
Non single-diffractive (NSD)	99.8	99.5
Inelastic (INEL)	97.1	96.6

Table 5.1: `MB1` trigger efficiency in percent.

5.1.1 Trigger Efficiency

Table 5.1 shows the trigger efficiencies of the `MB1` trigger for the different process types and for NSD and inelastic events at $\sqrt{s} = 900\,\text{GeV}$ and $\sqrt{s} = 10\,\text{TeV}$ derived from the detailed detector simulation. Values obtained using the Pythia event generator and the Phojet event generator are shown. Trigger efficiencies for diffractive events are generally higher (up to 10%) for Phojet than for Pythia due to different assumptions of the kinematics of diffractive events. The diffractive trigger efficiencies reduce towards higher $\sqrt{s}$.

Table 5.2 compares the efficiencies of the three minimum-bias triggers at a fixed $\sqrt{s} = 10\,\text{TeV}$ using the Pythia event generator. Also shown are the trigger efficiencies for beam-gas and beam-halo events [Con05]. Note that these values are for $\sqrt{s} = 14\,\text{TeV}$, but should be very close to the ones at 10 TeV. These trigger efficiencies denote the percentage of beam-gas (beam-halo) events that are not identified as beam-gas (beam-halo) events and cause the given trigger.

Beam-gas collisions are collisions between particles of the beam and molecules or atoms of residual gas in the vicinity of the detector. In the analysis performed in [Con05] they are defined as collisions that occur in $|\text{vtx-}z| < 20\,\text{m}$, where vtx-$z$ is the distance in

	Collision type						
Trigger	ND	SD	DD	NSD	INEL	Beam-gas	Beam-halo
`MB1`	100.0	71.6	86.2	97.9	93.0	7.7	2.3
`MB2`	99.2	57.5	66.3	94.2	87.4	2.0	0.3
`MB3`	99.1	53.7	61.5	93.4	86.0	< 0.01	< 0.01

Table 5.2: Trigger efficiency in percent at $\sqrt{s} = 10\,\text{TeV}$ (Pythia).

the z-direction (beam-line) from the nominal interaction point. Beam-halo events are collisions between beam particles and molecules or atoms of residual gas that occur outside this region. Their collision products usually do not directly reach the detector, they are instead transported with the beam (in the beam-halo, hence the name) towards the detector and may cause a trigger. The mentioned study considers beam-halo events that occur between vtx-$z = 30\,\text{m}$ and vtx-$z = -50\,\text{m}$, but outside $|\text{vtx-}z| < 20\,\text{m}$ (the asymmetry is due to technical reasons).

Generally, the beam-gas and beam-halo rates are much lower than the collision rates. For the startup scenario, described in Section 3.3, the expected raw beam-gas rate is negligible, the beam-halo rate is estimated to about 5.4 Hz (at $\sqrt{s} = 900\,\text{GeV}$) and 12 Hz (at $\sqrt{s} = 10\,\text{TeV}$).[1] This has to be compared to the raw collision rates of about 330 Hz ($\sqrt{s} = 900\,\text{GeV}$) and 13 kHz ($\sqrt{s} = 10\,\text{TeV}$). Together with the trigger efficiencies the ratio beam-halo over proton–proton collision is about 1/2 500 (at $\sqrt{s} = 900\,\text{GeV}$) and approximately 1/42 000 (at $\sqrt{s} = 10\,\text{TeV}$) for the `MB1` trigger. For nominal running conditions, the beam-gas rate is about 820 Hz and the beam-halo rate is 55 kHz. Compared to the raw collision rate including the trigger efficiencies, there is a 3×10^{-4} probability that a `MB1`-triggered event is a beam-gas collision, and a 6×10^{-3} probability that it is a beam-halo collision. Although these numbers are low, it is important to recall that they depend crucially on the assumptions of residual-gas density and trigger efficiency for such events. Therefore, an assessment from measured data is needed which is discussed in the following section. The induced systematic uncertainties are discussed in Sections 6.3.5 and 7.4.7.

[1] Residual gas estimates are taken from [Ros04, Fig. 4] which are for the startup scenario $2 \times 10^{10}\,\text{H}_2$ equivalent/m^3 in the interaction region (beam-gas) and $4 \times 10^{12}\,\text{H}_2$ equivalent/m^3 outside the interaction region (beam-halo). For nominal running conditions, these are $3 \times 10^{11}\,\text{H}_2$ equivalent/m^3 and $2 \times 10^{13}\,\text{H}_2$ equivalent/m^3, respectively. The report only provides estimates for ATLAS and CMS, numbers are assumed to be equivalent for ALICE.

In general, the trigger with the lowest bias (MB1) is preferred for the study of inelastic events. However, possible beam-gas and beam-halo background might motivate the use of a trigger with a better background rejection, which are triggers that require additional coincidences, like MB2 and MB3. Furthermore, it is very useful to perform an analysis using data collected with different trigger selections. Together with the corresponding corrections this should yield the same result. This increases confidence in the trigger efficiencies that are purely calculated from MC. A detailed evaluation of further minimum-bias trigger and background rejection capabilities can be found in [Con05]. The MB1 trigger is used in the following chapters.

5.1.2 Bunch-Crossing Trigger

For first data the discussed minimum-bias triggers are not used directly. Instead the detector is read out upon each bunch crossing. As a consequence, most of the collected events are without p+p interaction. In such events no collisions occur in the bunch crossing since the probability for an interaction in a bunch crossing is about 1 – 10%. Nevertheless, the trigger information as it would be normally used (*trigger bits*) is recorded. Therefore, this method allows the proper functioning of the trigger to be validated by comparing the trigger bits with the recorded data in the subdetectors.

An additional trigger on single bunches, i.e. a passing bunch from one beam without the counterpart in the other beam, allows the amount of beam-gas and beam-halo events that pass the trigger condition to be measured. This is important to estimate the contamination caused by these events and to verify the estimated rates mentioned in the previous section.

When the bunch-crossing trigger is used, the minimum-bias triggers introduced above are adopted during the data analysis in an offline way. The trigger bits are not used directly, instead the recorded information in the subdetectors is utilized to determine whether the event had given rise to a trigger or not. For example for MB1 it is sufficient that either a signal is present in the SPD or in one of the sides of the V0 detector. In the following the usage of *triggered* means either the direct use of the trigger or the offline way.

5.2 Event Selection

Not all events are used for analysis. Thus, for simulated events as well as real data an event selection needs to be applied. These events must be selected by the chosen trigger for the analysis.[2] In case of real data the trigger selection for the analysis can be the same as that used in the hardware trigger during data acquisition and then this step is not needed.

Furthermore, to be able to consider the tracks of an event, the vertex-reconstruction algorithm must have determined the primary-vertex position of the collision. The probability to reconstruct the primary-vertex of an `MB1`-triggered event is 92.5% with the SPD and 69.4% using only the TPC (for events inside $|\text{vtx-}z| < 10\,\text{cm}$ at $\sqrt{s} = 10\,\text{TeV}$).

5.3 Primary-Particle Definition

The analyses discussed in the subsequent two chapters yield distributions of primary charged particles defined by the following:

Primary particles are all particles produced in the collision, including products of strong and electromagnetic decays as well as weak decays of charmed and beauty particles, but excluding feed-down products from strange weak decays and other secondary particles. These are for example γ-conversions and products from secondary hadronic interactions with the detector material. In the simulation these are the final-state particles created by the event generator, which are then propagated (and decayed) in the subsequent detector simulation.

Non-primary particles including decay products are referred to by *secondary particles* in the following.

5.4 Tracklet and Track Selection

Two goals are achieved by the tracklet and track selection: it is assured that the tracklet and track quality is good, which means that the reconstructed properties (e.g. momentum and distance from the primary vertex) are close to the real values. Furthermore, certain cuts select tracklets and tracks from primary particles and suppress those from

[2] For example data has been taken with the `MB1` trigger but the analysis requires events triggered by the `MB2` condition.

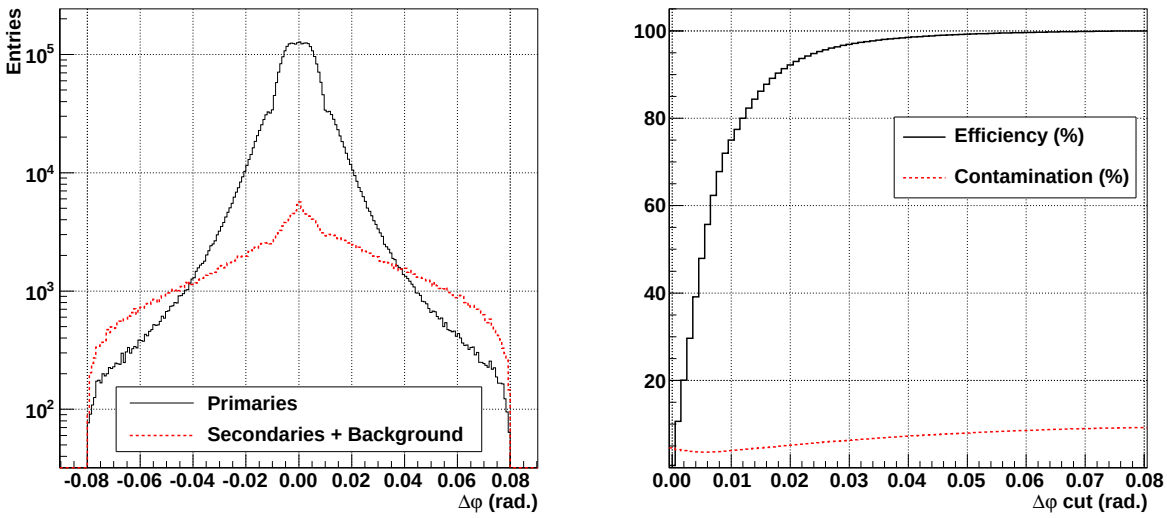

Figure 5.2: $\Delta\varphi$ distribution and cut of SPD tracklets.
The left panel shows the $\Delta\varphi$ distribution of tracklets subdivided into primaries and secondaries including background. An irregularity is seen around ± 0.01 rad. It is probably due to a geometrical effect but an explanation still needs to be found. However, it is not relevant for the efficiency and contamination because the cut is applied at a much larger value than 0.01 rad. The right panel shows the efficiency and contamination of a $\Delta\varphi$ cut.

secondary particles, which in the following sections will be called *primaries* and *secondaries*, respectively. The aim is a low contamination from secondaries while retaining a high efficiency for primaries.

5.4.1 SPD-Tracklet Selection

Only a few cuts are useful for SPD tracklets. As quality parameter the tracklet's $\Delta\varphi$ is used (see Section 4.6.3). Already the nominal magnetic field causes $\Delta\varphi$ to be non-zero; for example by about 14 mrad for a particle with $p_T = 200\,\mathrm{MeV}/c$. The left panel of Figure 5.2 shows the $\Delta\varphi$ distribution of SPD tracklets. Separately shown are primaries and secondaries including combinatorial background. Note that a tracklet is also counted as originating from a primary particle when the cluster in the outer layer stems from the primary's daughter particle. The total sample comprises about 91% tracklets from primaries, 6% from secondaries, and 3% from combinatorial background. The right panel shows the efficiency and contamination when a cut $|\Delta\varphi| < \Delta\varphi_{\mathrm{cut}}$ is applied.

The cut $|\Delta\varphi| < 0.05\,\text{rad}$ is chosen in such a way that 99% of the reconstructed tracklets from primary particles are selected. The contamination by secondary particles and combinatorial background is about 8%. Tracklets are already reconstructed with the event vertex taken into account, therefore no further constraint is used to ensure that tracklets originate from the vertex.

5.4.2 TPC-Track Selection

The reconstruction, described in Section 4.6.3, produces two sets of track parameters to which the TPC contributes: the first are the TPC-only tracks for which only TPC information is used. The second are global tracks to which in addition information from the ITS, TRD, and TOF contribute as well as other detectors provided the tracks are within their acceptance range. Therefore, these global tracks make use of more tracking information and their reconstruction is more precise. However, global track finding requires proper alignment between the different detectors and an understanding of the interplay between the information from the different detectors during the track finding. Therefore, for early measurements TPC-only tracks are used and are discussed in the following. It should be noted that the track-parameter cuts discussed can also be applied to global tracks, with adjustments to specific cut values.

Quality cuts can be applied to the number of clusters that were used for the reconstruction of the given track and the χ^2 per cluster, which determines the quality of the fit between the track and the contributing clusters. Furthermore, constraints can be placed on the five diagonal elements in the track-parameter covariance matrix[3]: the resolutions σ_y^2, σ_z^2, $\sigma^2_{\sin\varphi}$, $\sigma^2_{\tan\lambda}$, and σ^2_{1/p_T}.

A charged-particle decay inside the tracking volume can produce a kink on the track's trajectory, for example the decay $K^+ \rightarrow \mu^+\nu_\mu$. Due to the fact that the neutrino is not tracked, the kaon's trajectory appears changed, hence the name kink. In the reconstruction the K^+ and the μ^+ are found as separate tracks that obviously only correspond to one primary particle. The reconstruction identifies that the two tracks are related and flags the first as the *kink mother*, the second as the *kink daughter*.

To assure that the tracks originate from the primary vertex (in the following just called *vertex*), a cut on the distance between the vertex and the track is applied. For this purpose the closest point of the track's trajectory to the vertex is determined (Distance of Closest Approach – DCA). Either a cut on the absolute DCA (*absolute*

[3]See Appendix B for the definition of the track parameters.

DCA cut) or on the DCA divided by its estimated resolution is applied (*normalized DCA cut*). The first approach is more resilient in the case that the vertex position and the track-parameter resolutions are imprecise. The latter is in principle better as it uses more of the measured information. The accuracy of the resolutions calculated in the reconstruction is not well-known in early data-taking, therefore the first approach is more extensively discussed here.

Furthermore, the absolute DCA cut can be applied separately in two dimensions:

$$\Delta r < d_r \text{ and } \Delta z < d_z \tag{5.1}$$

or in combination:

$$\left(\frac{\Delta r}{d_r}\right)^2 + \left(\frac{\Delta z}{d_z}\right)^2 < 1, \tag{5.2}$$

still allowing for different values d_r and d_z in the r and z directions, respectively (elliptic cut). The choice depends on the status of the calibration and alignment and possible correlations between the two values. For example an insufficient drift-time calibration in the TPC leads to an imprecise value in z. In this case a narrow r cut could be combined with a wider z cut.

The optimal cut values are obtained in several steps. Loose quality criteria are applied in the first step to ensure a certain track quality. A track must have at least 50 contributing clusters and a χ^2 per cluster of less than 3.5. Furthermore, it is not allowed to be flagged as a kink daughter. Two cuts are applied to the track parameter resolutions: σ_y^2 and σ_z^2 have to be both less than $9\,\mathrm{cm}^2$; this corresponds to a positioning error at the DCA to the vertex of less than 3 cm in both directions.

After applying these quality cuts which at $\sqrt{s} = 10\,\mathrm{TeV}$, compared to the initial sample, remove 3.8% of primaries and 18% of secondaries, the sample contains about 64% primaries and 36% secondaries. These cuts remove more secondaries than primaries due to the fact that secondaries compared to primaries have typically a lower momentum and a smaller path length in the TPC. The second step is to reduce the amount of secondaries; this is achieved by the previously mentioned absolute DCA cut. To find the optimal values for the DCA cut, the DCA distribution is studied for tracks inside $|\eta| < 1$.[4]

[4]Studying this distribution for tracks outside $|\eta| < 1$ yields significantly different results due to the fact that tracks outside this region traverse considerably more material. To find optimal values for tracks for the full accessible η-region, the cut values would need to be determined as a function of η.

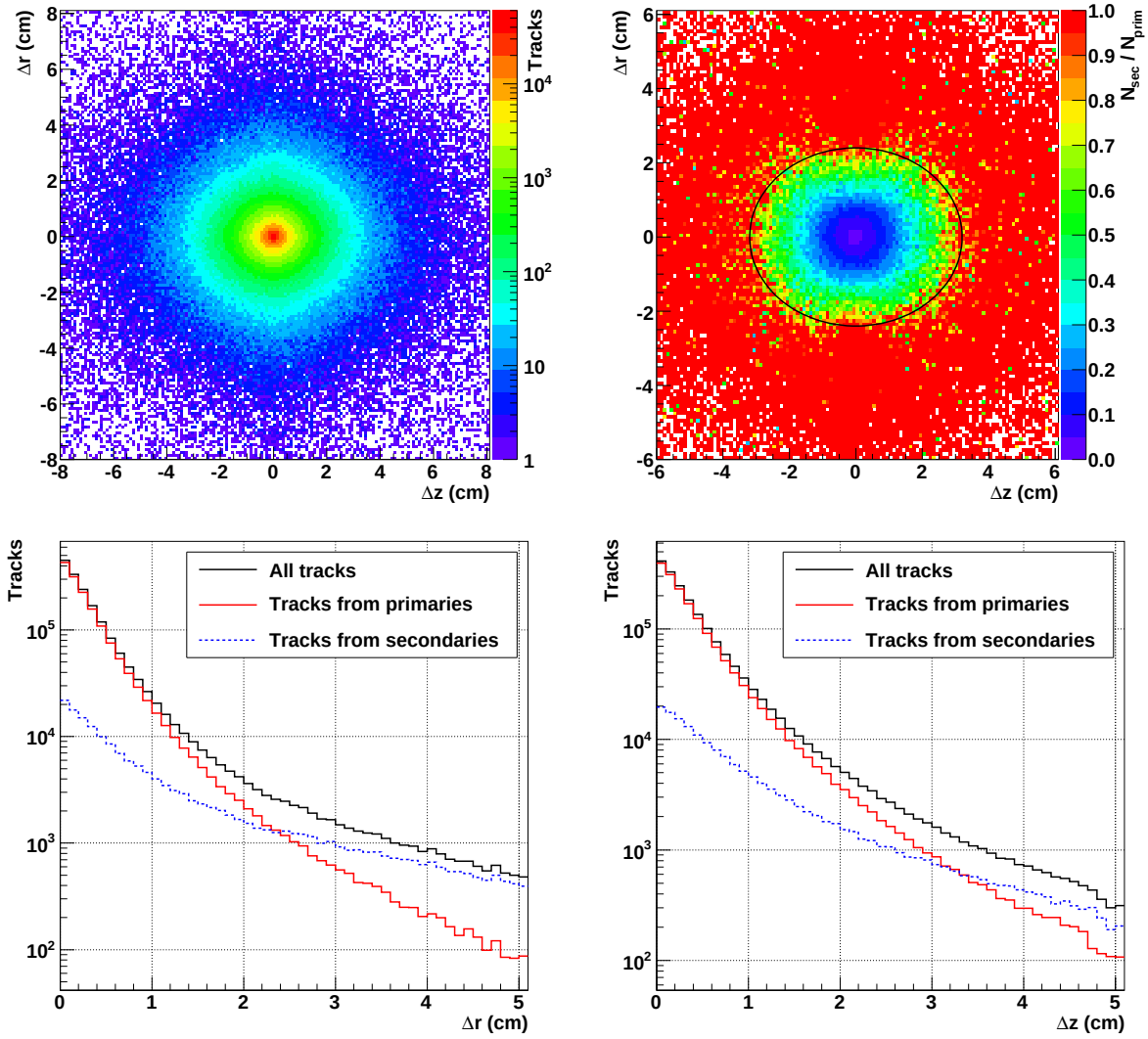

Figure 5.3: Absolute DCA track cut.

The DCA between the vertex and tracks that passed the loose quality-cuts (see text) is shown. The top left panel shows only the DCA of primary tracks in the Δr vs. Δz plane. The top right panel shows the number of secondary tracks divided by the number of primary tracks in the same plane. The ellipse indicates the applied cut (see text). In the bottom left and right panel the projections in Δr and Δz are shown (integrated within $\pm 0.5\,\mathrm{cm}$ of the other variable), respectively.

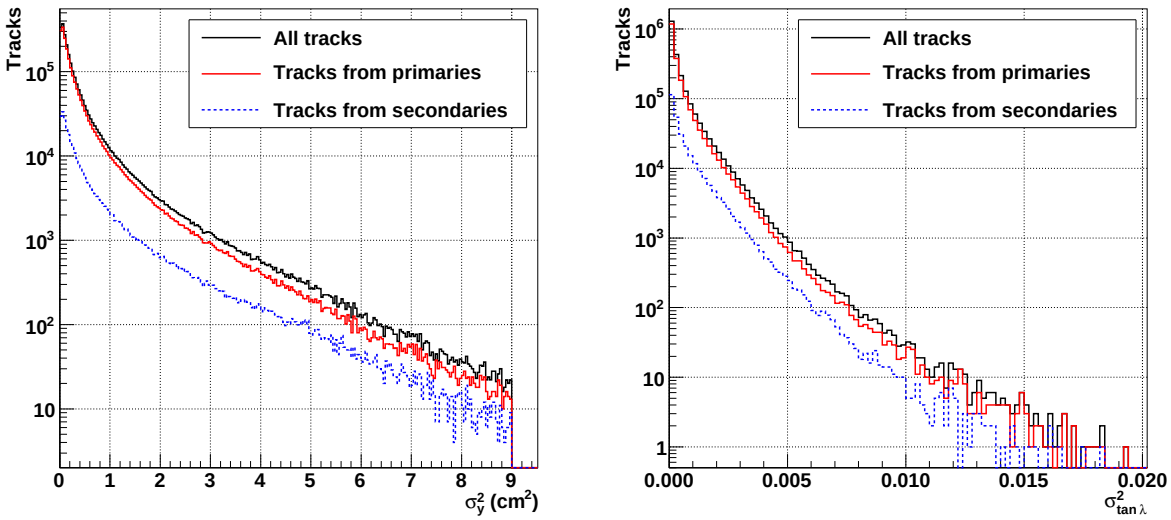

Figure 5.4: Track resolutions after track cuts.

The figure shows the σ_y^2 and $\sigma^2_{\tan\lambda}$ track resolutions after the quality and DCA track cuts. Results for primaries and secondaries are shown separately.

In the top left panel of Figure 5.3, the DCA of primary tracks is shown in the Δr vs. Δz plane. The top right panel contains the ratio between the number of secondaries and primaries. The scaling is set such that the maximum is 1 to indicate where the amount of secondaries exceeds the amount of primaries. In the bottom left panel, Δr is shown separately for tracks from primary and secondary particles in $|\Delta z| < 0.5\,\text{cm}$. The equivalent for Δz is shown in the bottom right panel. The yield of primaries and secondaries are equal at about 2.4 cm and 3.2 cm for Δr and Δz, respectively, which defines the chosen cut values and enclose well the area where the ratio is smaller than unity in the top right panel.

After applying the DCA cut in addition to the track-quality cuts, the sample contains 88% primaries and 12% secondaries, with 93% efficiency to select primaries. The question arises as to whether other cuts could be used to further improve the selection of primaries. Exemplarily, the distributions of σ_y^2 and $\sigma^2_{\tan\lambda}$ are shown in Figure 5.4. No further removal of secondaries can be achieved without the simultaneous removal of a significant amount of primaries. Furthermore, it can be seen that although no cut was applied to $\sigma^2_{\tan\lambda}$, the quality is quite good. More than 99% of the primary tracks are within $\sigma^2_{\tan\lambda} < 0.005$, which translates to a $\sigma_\eta < 0.07$ at η around 0.[5] This is smaller

[5]The given number is an upper limit including more than 99% of the tracks. Comparing the reconstructed values with the MC information, yields a distribution with a σ_η of about 0.005.

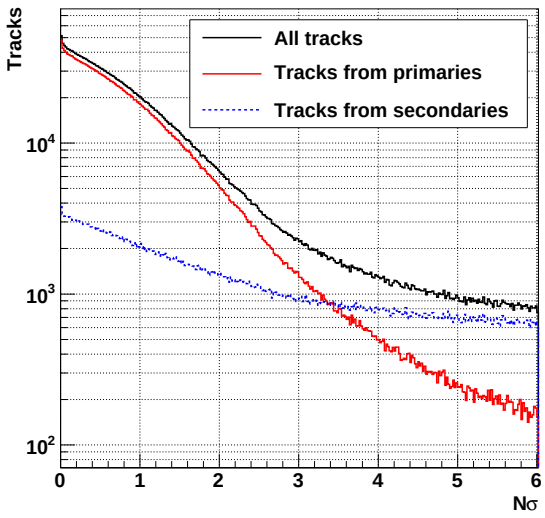

Figure 5.5: Normalized DCA cut. The figure shows the normalized DCA of the vertex and the track, see Eq. (5.4), separately for primaries and secondaries.

Quality cuts	$N_{\text{clusters}} > 50$ $\chi^2/\text{cluster} < 3.5$ $\sigma_y^2 < 9\,\text{cm}^2$ $\sigma_z^2 < 9\,\text{cm}^2$
Reconstruction flags	No kink daughter
Absolute DCA cut	$d_r = 2.4\,\text{cm}$ and $d_z = 3.2\,\text{cm}$
Normalized DCA cut	$N_\sigma < 4$

Table 5.3: Track cuts for TPC-only tracks. The DCA cuts are two separate possibilities (see text).

than the η-bin size used in the analysis performed in Chapter 6. Note that these values are for TPC-only tracks, the combination with other subdetectors is expected to significantly improve the quoted resolutions.

The previously mentioned normalized DCA cut takes into account the resolutions of the track parameters and the vertex position. It cuts on the normalized distance to the vertex defined by:

$$d_\sigma = \sqrt{\left(\frac{\Delta r}{\sigma_r^{\text{dca}}}\right)^2 + \left(\frac{\Delta z}{\sigma_z^{\text{dca}}}\right)^2}, \tag{5.3}$$

where $\Delta r/\sigma_r^{\text{dca}}$ and $\Delta z/\sigma_z^{\text{dca}}$ are the normalized distances in the transverse and longitudinal directions, respectively. σ_r^{dca} and σ_z^{dca} take into account the resolutions of the vertex position and the track parameters. The cut accepts a number of standard deviations (N_σ) of tracks, if they were distributed like a two-dimensional Gaussian and is thus also called N_σ-cut. To achieve the usual definition of N_σ with respect to a Gaussian (e.g. that a 1σ-cut includes 68% of all tracks), the following relation yields N_σ from d_σ:

$$N_\sigma = \sqrt{2}\,\text{erf}^{-1}(1 - \exp(-d_\sigma^2/2)) \tag{5.4}$$

(where erf^{-1} denotes the inverse error function). This formula is derived in Appendix C. Figure 5.5 shows the number of tracks, after the loose track-quality cuts, as function of N_σ. Tracks from primary and secondary particles are shown separately. Between a

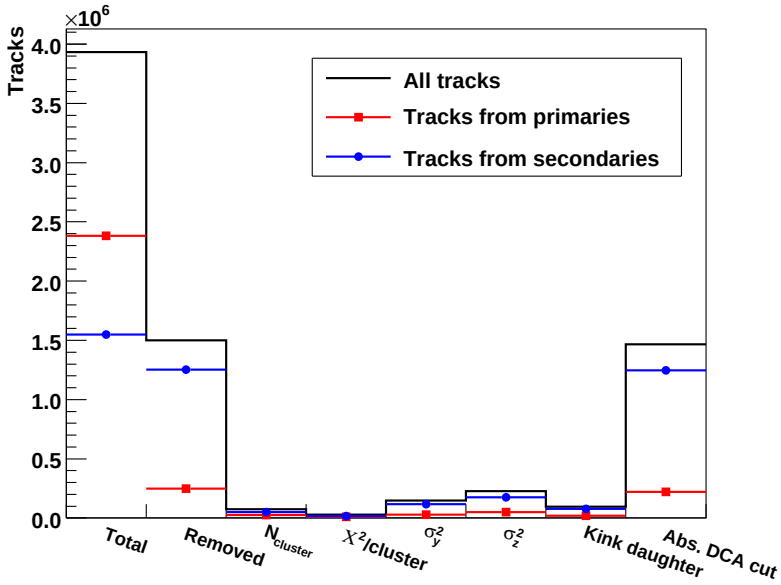

Figure 5.6: Track cut influence.

The plot summarizes the influence of the different track cuts. The DCA cut has the highest influence. Note that one track can fail several cuts and thus contribute to several bins.

3σ-cut and a 4σ-cut the number of tracks originating from primaries equals the number of tracks from secondaries.

The values for the different cuts used for the analysis in this thesis are summarized in Table 5.3. From the two options for the DCA cut, the absolute DCA cut is used. Figure 5.6 shows the influence of the different cuts on tracks from primary and secondary particles. The two first bins indicate the total number of tracks and the number of tracks that do not pass the track selection. The remaining bins show the number of tracks that did not pass the specified cuts. A single track can fail to pass several cuts and therefore can be counted more than once. The DCA cut is the most powerful and reduces the amount of secondaries significantly.

In summary, these track cuts remove about 38% of the total reconstructed tracks in $|\eta| < 1$. About 10% of the tracks from primary particles and 81% of the tracks stemming from secondary particles are removed. After the cuts the sample has a contamination with secondaries of about 12%; the efficiency to select primaries is 90%. Figure 5.7 shows the distribution of primaries and secondaries as a function of p_T after the cuts, as well as the efficiency and contamination.

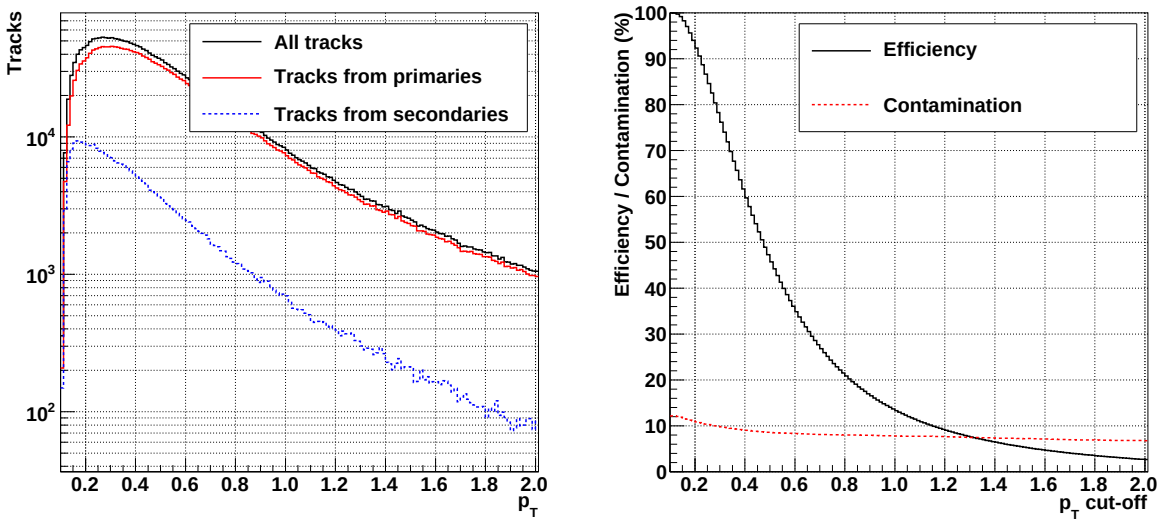

Figure 5.7: Track distributions after track cuts.

The number of primaries and secondaries in $|\eta| < 1.0$ after the track cuts is shown as a function of p_T (left panel). The right panel shows the efficiency and contamination when the p_T cut-off is set at the given p_T.

In the analysis, the values of the cuts have to be varied in order to assess the sensitivity to the chosen values. Different sets of cuts should result in the same results. Obviously the same set of cuts has to be used in the derivation of the corrections. Furthermore, the distribution of each variable that provides the basis for a cut has to be compared between the simulation and the real data. Incorrect estimates of resolutions (e.g. caused by wrongly estimated detector alignment and calibration) might result in distorted distributions of selection parameters and thus change the effect of the track selection cuts in an uncontrolled fashion. Systematic uncertainties associated with the track cuts are discussed in Sections 6.3.9 and 7.4.11.

5.5 Datasets Used in this Thesis

Simulated data was used to develop and evaluate the analyses in this thesis. This data has been produced in the Physics Data Challenge 2008 (PDC08). The following sets of p+p collision data have been used:

- LHC08c11: 270 000 Pythia events at $\sqrt{s} = 10\,\text{TeV}$ (AliRoot tag v4-14-Rev-04),
- LHC08c12: 310 000 Pythia events at $\sqrt{s} = 900\,\text{GeV}$ (AliRoot tag v4-14-Rev-04),

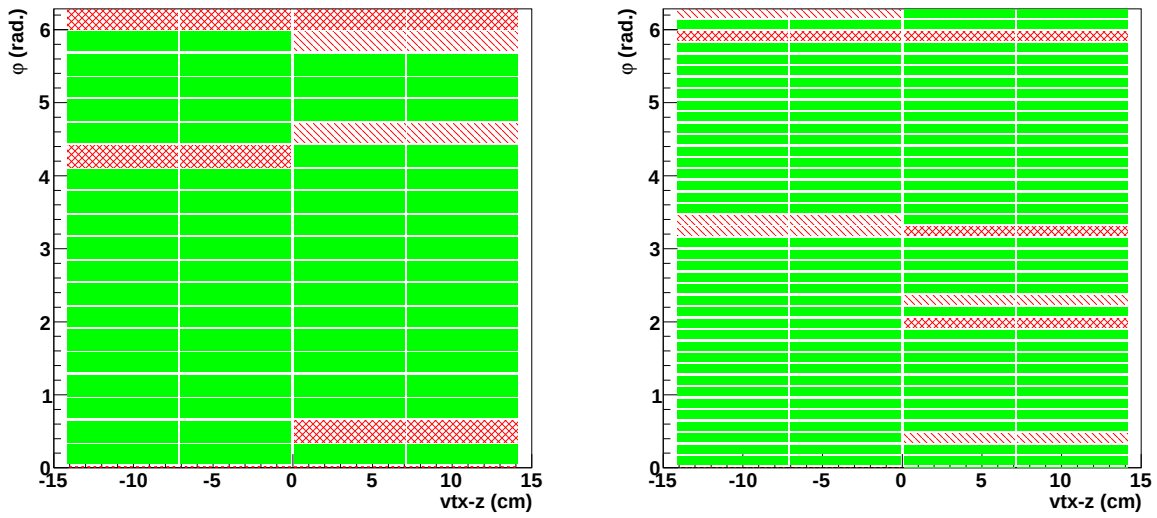

Figure 5.8: Inactive modules in the SPD.

The figure shows the distribution of active (solid green) and inactive (only LHC08c: single shaded red; LHC08c and LHC08e1: crossed shaded red) modules that were used in the simulation. The left panel shows the first layer, the right panel shows the second layer. Note that the spacing between the modules is increased for visibility.

- LHC08c15: 200 000 Phojet events at $\sqrt{s} = 10\,\text{TeV}$ (AliRoot tag v4-14-Rev-04),
- LHC08c16: 220 000 Phojet events at $\sqrt{s} = 900\,\text{GeV}$ (AliRoot tag v4-14-Rev-04),
- LHC08e1: 470 000 Pythia events at $\sqrt{s} = 10\,\text{TeV}$ (AliRoot tag v4-15-Rev-06).

The data has been produced at the nominal magnetic field ($B = 0.5\,\text{T}$). The respective AliRoot version tags are listed for completeness since the simulation environment is continuously evolving. AliRoot is built with ROOT tag v5-21-01-alice, Geant3 tag v1-9-6, Pythia 6.2.14, and Phojet version 1.12. For Pythia the 'ATLAS tune', see Section 1.5.1, was used. The conditions data in use reflects the status of the installed hardware and alignment as of August 2008.

In the SPD, the calibration status is different between the LHC08c and LHC08e productions because additional modules were found to have cooling problems during commissioning. In the LHC08c productions, 16 modules (out of 240 modules) have been marked inactive, while there are 30 modules marked inactive in LHC08e1. Figure 5.8 shows the distribution of these modules in the SPD layers. In the first production, half of these are in the first layer (8 out of 80, i.e. 10%) and the other half in the second layer (8 out of 160, i.e. 5%) and they do not overlap. Each tracklet needs a signal in

both layers, therefore these inactive areas result in a reduced tracking efficiency for tracklets by about 15%. In the second production, these numbers amount to 15% and 11.3% in the first and second layer, respectively, and cause a reduction of 22.5% in the tracking efficiency (some of the inactive modules overlap).

All TPC readout chambers were marked active in the simulated data, which results in an expected insensitive area of about 10% (by design; see also Section 4.1.2).

The expected distribution of the z-position of the collision vertex has a r.m.s. width of 3.8 cm at $\sqrt{s} = 10\,\text{TeV}$ and of 7.4 cm at $\sqrt{s} = 900\,\text{GeV}$ and these values are used in the simulation. Therefore, more events are found at larger $|z|$ at 900 GeV than at 10 TeV for two samples of the same size.

The data is stored in the AliEn Grid file catalog in the directories:

```
/alice/sim/PDC_08a/LHC08c11
/alice/sim/PDC_08a/LHC08c12
/alice/sim/PDC_08/LHC08c15
/alice/sim/PDC_08/LHC08c16
/alice/sim/PDC_08/LHC08e1
```

Figures and results in the following discussions correspond to Pythia at $\sqrt{s} = 10\,\text{TeV}$ unless otherwise indicated. Numbers and systematic uncertainties are primarily given for $\sqrt{s} = 10\,\text{TeV}$. In Chapter 6 the data with the tag LHC08c11 has been used. Chapter 7 makes use of data with the tag LHC08e1 that became available later and consists of more events. Comparisons with the Phojet data are made in the context of the systematic studies.

To see the effect of statistical fluctuations in the evaluations of the analyses, the available simulated data is split into two parts. A fraction of the events is taken that represents the data measured with the experiment, called *analysis input sample*. The remaining events are used to derive the correction factors, called *correction input sample*. Effects of statistical fluctuations cannot be seen when an identical sample is used for analysis and corrections.

Chapter 6

Pseudorapidity-Density Measurement

This chapter describes the measurement of the pseudorapidity density of primary charged particles $dN_{ch}/d\eta$. The analysis input, extracted from the reconstruction output, has to be corrected for various detector effects such as tracking efficiency, vertex reconstruction and trigger efficiency, as well as physical effects, e.g. secondaries originating from decays and γ-conversions. These corrections are derived in this chapter utilizing events produced by the detailed detector simulation and reconstruction AliRoot. In addition, the systematic effects that arise during this measurement are studied.

The analysis is performed using data from the Silicon Pixel Detector (SPD) and the Time-Projection Chamber (TPC). This allows the comparison between the results of two independent detector systems. However, due to the increased acceptance of the SPD, especially at low p_T, the result using the data from the SPD has intrinsically smaller statistical and systematic errors.

In Section 6.1 the procedure used to obtain the $dN_{ch}/d\eta$ distribution is presented. The corrections are described in detail in Section 6.2 and a study of the systematic uncertainties is presented in Section 6.3. Section 6.4 describes the steps that need to be followed in order to correct the data measured by the detector including the necessary verifications and checks.

The employed simulated dataset (LHC08c11, see Section 5.5) is split into 50 000 events as analysis input sample and 220 000 events to derive the correction factors.

6.1 Procedure Overview

The reconstruction has been described in Section 4.6.3. For this analysis, tracklets reconstructed with information from the SPD and tracks reconstructed with TPC information are used. The event and track selection that is applied was described in the preceding Chapter 5.

The goal of the measurement is to determine the number of primary charged particles per unit of pseudorapidity per collision. This is obtained by counting the number of tracks and the number of events and applying three corrections.

The first correction takes into account the difference between the number of measured tracks and the number of primary charged particles. This difference is caused by the limited tracking efficiency, admixtures by secondaries, and decay of primary particles. The corresponding correction is called **track-to-particle correction**.

The second correction considers the bias that is imposed by the vertex reconstruction on the triggered event sample. This bias is caused by the fact that due to the specific event properties it may not be possible to reconstruct the primary vertex position (in the following referred to as *vertex position*). This correction is named **vertex-reconstruction correction**.

The third correction takes into account the bias imposed by the trigger used to acquire the event sample. This correction is called **trigger-bias correction** and contains different numerical factors depending on whether the goal of the analysis is the $dN_{ch}/d\eta$ distribution for inelastic (in the following called *inelastic* **trigger-bias correction**) or NSD collisions (in the following called *NSD* **trigger-bias correction**). Note that this correction in particular is model-dependent because no properties of not triggered (and thus not measured) events can be deduced from the measured data.

Applying the **track-to-particle correction** alone, with either the **vertex-reconstruction correction** alone, or with both the **vertex-reconstruction correction** and the **trigger bias correction**, results in three different $dN_{ch}/d\eta$ distributions. Each of these represents a valid measurement, albeit for a different event sample. This is illustrated in Figure 6.1. Which event sample is measured depends on the corrections applied:

- applying only the **track-to-particle correction** leads to the $dN_{ch}/d\eta$ distribution of events that are triggered and have a reconstructed vertex (left panel of Figure 6.1);

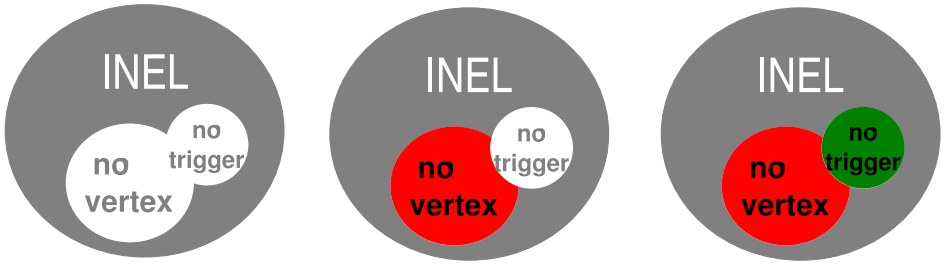

Figure 6.1: Event classes in the $dN_{ch}/d\eta$ analysis.
The full sample of collisions can be divided into sub-samples: 1) collisions that give rise to a trigger and where the vertex position is reconstructed (left panel), 2) collisions that give rise to a trigger and where the vertex position could not be reconstructed (center panel) and 3) all collisions, including those that do not give rise to a trigger and where the vertex position evidently cannot be reconstructed (right panel).

- applying also the **vertex-reconstruction correction** results in the $dN_{ch}/d\eta$ distribution for triggered events; this is commonly referred to as minimum-bias event sample (center panel of Figure 6.1);
- applying in addition the **trigger-bias correction** yields the $dN_{ch}/d\eta$ distribution for all considered collisions, i.e. inelastic or NSD events (right panel of Figure 6.1).

The **track-to-particle correction** is applied at the track level. The **vertex-reconstruction correction** as well as the **trigger-bias correction** are applied at the track and event level. The track-level corrections are determined as a function of η and the z-position of the collision vertex (abbreviated: vtx-z). The corrections for the measurement with the TPC when the magnetic field is turned on are determined also in a third dimension, the particles' p_T. In the following only the case with magnetic field is discussed.

The event level corrections are determined as a function of vtx-z and the number of accepted tracklets (SPD) or tracks (TPC) in the event (in the following referred to as n).

Figure 6.2 shows the procedure on track level in a simplified form as flowchart. Depicted are the different steps in the simulation, reconstruction, and correction procedure. Their influence on the number of particles is given.

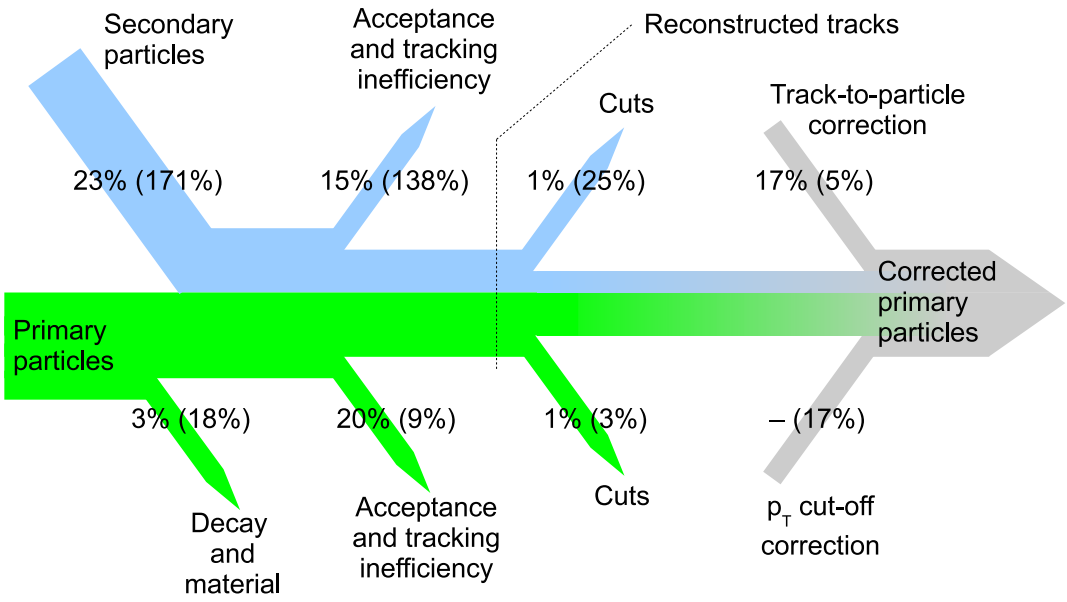

Figure 6.2: Flowchart of the analysis procedure.
The figure shows the evolution of the number of particles during propagation through the detector as well as during the correction procedure. The approximate percentages of particle loss and increase are given relative to the *initial* number of primary particles. The first value denotes the value for the SPD analysis, the second for the TPC analysis. Only particles in $|\eta| < 1$ of events that occurred in $|\text{vtx-}z| < 6\,\text{cm}$ are considered. Only secondaries with a p_T above $50\,\text{MeV}/c$ that are created within the tracking volume of the corresponding subdetector are shown. A combinatorial-background contribution of about 3% for the SPD measurement is included for simplicity in the number of secondaries (see Section 5.4.1). The acceptance and tracking inefficiency contains the p_T cut-off at $200\,\text{MeV}/c$ for the TPC. Therefore, the given values cannot be directly compared to the numbers given in Section 5.4.2 where all tracks in $|\eta| < 1$ are considered. The p_T cut-off correction is only applied in the TPC measurement.

6.2 Corrections

This section describes in detail the different corrections which are applied and gives estimations of their magnitude.

6.2.1 Correction Procedure

The procedure used to obtain the $dN_{ch}/d\eta$ distribution from the data adapting the corrections is summarized in the following paragraphs. An analytical description is given further below. In the analysis, two histograms are created:

1. each accepted track of each of the accepted events is recorded in an η vs. vtx-z (for the TPC with magnetic field: η vs. vtx-z vs. p_T) histogram; the entry in the histogram is weighted with the corresponding values of the corrections (discussed below);

2. a vtx-z vs. n histogram counts the number of accepted events, which is needed to normalize to the total number of events; the entries in this histogram are also weighted with the values of the corresponding corrections.

In the analysis procedure these two histograms are produced for each of the three event classes (see Figure 6.1), which means that all or only a subset of the corrections are applied, depending on the event class. After filling the histograms, the $dN_{ch}/d\eta$ distribution is calculated. A vtx-z range is chosen which is η-dependent due to the varying acceptance window at different vertex positions. The vtx-z and p_T variables are then integrated. Each η-bin is weighted with the total number of events within the same acceptance window. The number of events is calculated from the vertex position distribution histogram, where the multiplicity is integrated using the same vtx-z range. A special correction is applied to account for triggered events without reconstructed vertex; this is explained in the following discussion. Optionally, for the TPC measurement with active magnetic field, the effect of the p_T cut-off is corrected.

The mathematical description of the procedure outlined above is given in the following:

Track level

For clarity vtx-z is written as z in the mathematical description. In the following the three parameters η, z, p_T are used in all formulas. For the SPD measurement and the TPC measurement without magnetic field, p_T is not defined for obvious reasons and has to be omitted in all corresponding formulas.

Tracks are weighted by the **track-to-particle correction** $C_{\mathrm{trk}}(\eta, z, p_T)$, by the track-level **vertex-reconstruction correction** $C_{\mathrm{vtx}}(\eta, z, p_T)$, and the track-level **trigger-bias correction** $C_{\mathrm{trig}}(\eta, z, p_T)$ in order to obtain the number of particles.

To define these, the following functions are used:

$$G_{\mathrm{eventclass}}(\eta, z, p_T) \tag{6.1}$$

is the number of **generated particles** in the bin η, z, p_T, and

$$M_{\text{eventclass}}(\eta, z, p_T) \tag{6.2}$$

is the number of **reconstructed tracks** in the bin η, z, p_T in events from a given event class. The reconstructed tracks are associated with the primary MC particle that 'caused' the reconstructed track. In the case of successful association, η, z, p_T are the values of the MC particle, not the reconstructed values. The consequence of this treatment is that a reconstructed secondary particle is counted in the bin of the corresponding primary. Therefore, the correction includes the correction for secondaries. The event class can be: all events (abbreviated *all*), where all events stands for inelastic or NSD events; triggered events (abbreviated *trig*); and triggered events with reconstructed vertex (abbreviated *trigvtx*). In addition,

$$B(\eta, z, p_T) \tag{6.3}$$

denotes the number of **reconstructed tracks** in η, z, p_T that cannot be clearly assigned to a MC particle (**background**). The values η, z, p_T are the reconstructed ones, of course. In the case of the SPD, these are typically combinations of two clusters originating from different particles. In practice, no background appears in the TPC, all tracks in the TPC can be uniquely assigned to a MC particle due to the large number of clusters.

With the given definitions, the track-level corrections are:

$$C_{\text{trk}}(\eta, z, p_T) = \frac{G_{\text{trigvtx}}(\eta, z, p_T)}{M_{\text{trigvtx}}(\eta, z, p_T) + B(\eta, z, p_T)} , \tag{6.4}$$

$$C_{\text{vtx}}(\eta, z, p_T) = \frac{G_{\text{trig}}(\eta, z, p_T)}{G_{\text{trigvtx}}(\eta, z, p_T)} , \tag{6.5}$$

$$C_{\text{trig}}(\eta, z, p_T) = \frac{G_{\text{all}}(\eta, z, p_T)}{G_{\text{trig}}(\eta, z, p_T)} . \tag{6.6}$$

This approach neglects distortions due to resolution and binning. This is justified since the bin sizes are chosen larger than the resolution. The effect of secondaries is included as mentioned previously.

Upper limits for the expected resolutions are $\sigma_\eta = 0.005$ and $\sigma_z = 0.12\,\text{cm}$ for the SPD. For the TPC the estimated numbers are $\sigma_\eta = 0.01, \sigma_z = 0.25\,\text{cm}$, and the p_T resolution is around a few % in the p_T-region considered in this analysis (see also Section 4.7.1). Results from cosmic-ray data indicate that the resolutions for these quantities are even somewhat better. The given σ_z resolutions are for low-multiplicity events, for high-multiplicity events values of $\sigma_z = 80\,\mu\text{m}$ (SPD) and $\sigma_z = 0.15\,\text{cm}$ (TPC) are obtained.

The correction tables are stored in histograms where the binning can be adjusted. The choice of the binning should reflect the resolution in the specific variable, as well as the

overall available statistics. For practical reasons non-equidistant bins are used. For the statistics expected for first measurements, a reasonable η-bin width is 0.1. The z-bin width is 2 cm in $|z| < 10$ cm and 5 cm for $|z| > 10$ cm. The smallest p_T-bin width (at low p_T where the correction changes rapidly) is 25 MeV/c; the multiplicity-bin width is 1 at low multiplicity. Both get larger towards larger momenta and multiplicities.

p_T cut-off

The number of particles missed due to the p_T cut-off are determined by way of MC simulations and are applied as an η-dependent factor. The correction is determined as the fraction between the total number of generated particles and the number of generated particles above the p_T cut-off:

$$C_{\text{cut-off}}^{p_T,\text{min}}(\eta) = \frac{\int\int G_{\text{all}}(\eta, z, p_T) dp_T dz}{\int\int_{p_{T,\text{min}}}^{\infty} G_{\text{all}}(\eta, z, p_T) dp_T dz}. \tag{6.7}$$

This correction is only applied for the measurement with the TPC and active magnetic field.

Event level

In order to obtain the number of collisions, events are weighted by the event-level **vertex-reconstruction correction** $\tilde{C}_{\text{vtx}}(z, n)$ and the event-level **trigger-bias correction** $\tilde{C}_{\text{trig}}(z, n)$.

With

$$E_{\text{eventclass}}(z, n) \tag{6.8}$$

as the number of **events** in the bin z, n in the given event class, they are defined by:

$$\tilde{C}_{\text{vtx}}(z, n) = \frac{E_{\text{trig}}(z, n)}{E_{\text{trigvtx}}(z, n)}, \tag{6.9}$$

$$\tilde{C}_{\text{trig}}(z, n) = \frac{E_{\text{all}}(z, n)}{E_{\text{trig}}(z, n)}. \tag{6.10}$$

Distortions due to resolution and binning are neglected, following the same argument as before. Clearly, the resolution is worse than the bin size in the case of multiplicity, which will be discussed in detail in Section 7.2.1, where this fact becomes important for the measurement of the multiplicity distribution. However, for the measurement of $dN_{ch}/d\eta$, the *measured* multiplicity is consistently used for n. This variable is later integrated and therefore the binning has no influence on the final result.

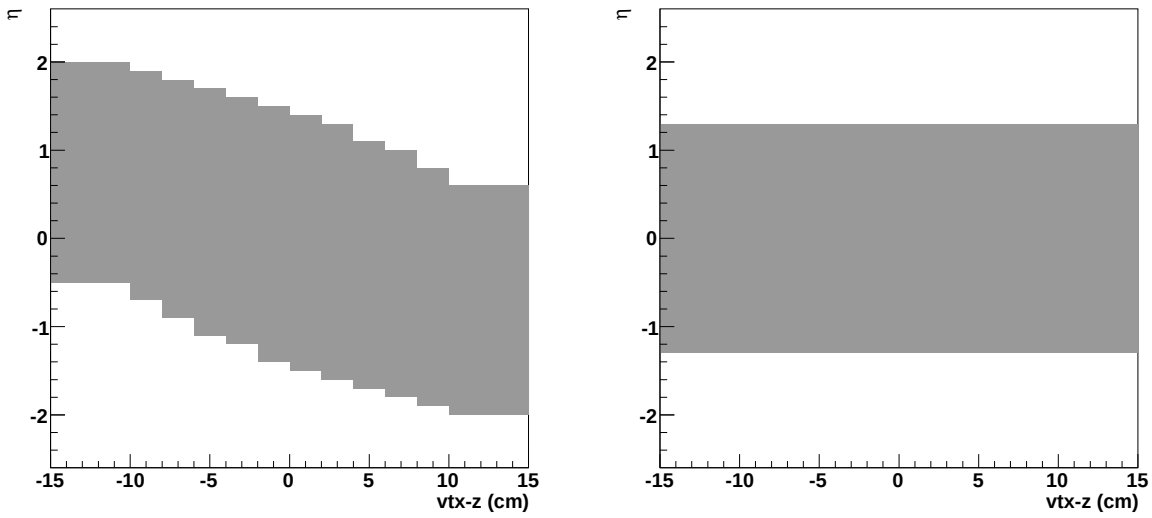

Figure 6.3: Acceptance maps.

The figure shows the acceptance maps of the SPD (left panel) and TPC (right panel) in the η vs. vtx-z plane. The acceptance in η changes depending on the vtx-z position for the SPD. This is not the case for the TPC.

Acceptance

The SPD as well as the TPC do not cover the entire pseudorapidity range η. For a given η, particles from collisions from within a certain z-vertex range reach the detector. Therefore, for each η only that particular z range should be considered. This has to be done at the level of the tracks, which is 'intrinsic' because without acceptance there are no tracks, and at the level of events, which is 'not intrinsic' and thus very important to obtain the correct number of events for the normalization. The z range for a given η, which spans from $z_{\mathrm{min}}(\eta)$ to $z_{\mathrm{max}}(\eta)$, is determined by requiring the correction factor $C_{\mathrm{trk}}(\eta, z, p_T)$ to be smaller than a certain limit C_{limit} (e.g. a limit of 5 means that at least about 20% of the primary particles are found in that specific bin). For the purpose of this definition p_T is integrated considering only the region that is used in the analysis ($p_T > p_{T,\mathrm{min}}$):

$$z_{\mathrm{min}}(\eta) = \min\left\{ z \;:\; \frac{\int_{p_{T,\mathrm{min}}}^{\infty} G_{\mathrm{trigvtx}}(\eta, z, p_T) dp_T}{\int_{p_{T,\mathrm{min}}}^{\infty} M_{\mathrm{trigvtx}}(\eta, z, p_T) + B(\eta, z, p_T) dp_T} < C_{\mathrm{limit}} \right\}, \tag{6.11}$$

$$z_{\mathrm{max}}(\eta) = \max\left\{ z \;:\; \frac{\int_{p_{T,\mathrm{min}}}^{\infty} G_{\mathrm{trigvtx}}(\eta, z, p_T) dp_T}{\int_{p_{T,\mathrm{min}}}^{\infty} M_{\mathrm{trigvtx}}(\eta, z, p_T) + B(\eta, z, p_T) dp_T} < C_{\mathrm{limit}} \right\}. \tag{6.12}$$

The resulting acceptance maps for $C_{\mathrm{limit}} = 5$ are shown in Figure 6.3. For the TPC, the acceptance does not depend on the event vertex. However, due to the fact that

the amount of material that needs to be traversed depends on the polar angle ϑ of the particle, a dependence of the correction factors on the event vertex position is still obtained (see Figure 6.6 on page 124).

Application of the corrections to the data

In this section values purely derived from the measured data are marked with an asterisk (*). The following quantities are measured: the number of tracks $M^*_{\text{trigvtx}}(\eta, z, p_T)$, the number of events $E^*_{\text{trigvtx}}(z, n)$, and the number of triggered events:

$$E^*_{\text{trig}}(n) = \int E^*_{\text{trig}}(z, n)dz. \tag{6.13}$$

$E^*_{\text{trig}}(z, n)$ is not available for triggered events without a reconstructed vertex because z is not known. Thus only the integral can be measured. η, z, p_T, and n are the measured quantities, of course.

The corrected number of particles P is calculated by:

$$P(\eta, z, p_T) = M^*_{\text{trigvtx}}(\eta, z, p_T) \times C_{\text{trk}}(\eta, z, p_T) \times C_{\text{vtx}}(\eta, z, p_T) \times C_{\text{trig}}(\eta, z, p_T). \tag{6.14}$$

The corrected number of interactions I (collisions) is calculated by:

$$I(z, n) = E^*_{\text{trigvtx}}(z, n) \times \tilde{C}_{\text{vtx}}(z, n) \times \tilde{C}_{\text{trig}}(z, n) \text{ for } n > 0. \tag{6.15}$$

$E^*_{\text{trigvtx}}(z, 0)$ is not defined because a reconstructed vertex requires at least one tracklet (SPD) or one accepted track (TPC).[1] By definition for $n = 0$, the event does not have a reconstructed vertex but may still be triggered. These $E^*_{\text{trig}}(0)$ triggered events need to be distributed among the different bins in z. The only information available is the vertex distribution found in events with a reconstructed vertex, and the distribution of the events has to be based on this quantity by defining:

$$\alpha^*(z) = \frac{\sum_{n=1}^{\infty} E^*_{\text{trigvtx}}(z, n)}{\sum_{n=1}^{\infty} \int E^*_{\text{trigvtx}}(z, n)dz}. \tag{6.16}$$

The z-vertex distribution of all collisions is independent of multiplicity. However, this is not the case for triggered events as well as triggered events without a reconstructed vertex. These subsets may be biased due to the geometrical acceptance of the chosen trigger detectors. This is taken into account by the introduction of $F(z)$ which is

[1] For the combined ITS and TPC measurement, which is not discussed here, the situation is different: the SPD reconstructs the vertex position and the tracks are reconstructed with information from the TPC and ITS. Thus an SPD vertex position may be reconstructed even without tracks in the TPC.

the relation between the z-vertex distribution of triggered events with and without a reconstructed vertex:

$$F(z) = \frac{E_{\text{trig}}(z,0) / \int E_{\text{trig}}(z,0) dz}{\sum_{n=1}^{\infty} E_{\text{trigvtx}}(z,n) / \sum_{n=1}^{\infty} \int E_{\text{trigvtx}}(z,n) dz}. \tag{6.17}$$

In practice, the influence of $F(z)$ is less than 10%.

Eqs. (6.16) and (6.17) allow $I(z,n)$ for $n = 0$ to be obtained:

$$I(z,0) = E^*_{\text{trig}}(0) \times \alpha^*(z) \times F(z) \times \tilde{C}_{\text{trig}}(z,0). \tag{6.18}$$

Finally, $dN_{ch}/d\eta$ is calculated:

$$\left.\frac{dN_{ch}}{d\eta}\right|_{\eta=\eta'} = \frac{\int_{z_{\min}(\eta')}^{z_{\max}(\eta')} \int_{p_{T,\min}}^{\infty} P(\eta', z, p_T) dp_T dz}{\int_{z_{\min}(\eta')}^{z_{\max}(\eta')} \sum_n I(z,n) dz} \times C_{\text{cut-off}}^{p_T,\min}(\eta'). \tag{6.19}$$

In practice, the distribution still needs to be normalized by the inverse width of the bins in the final histogram to obtain the differential distribution $dN_{ch}/d\eta$.

Replacing the measured quantities by the values obtained from the simulation (e.g. E^* by E) shows that Eq. (6.19) is exact. The given formula includes all corrections and corrects to the $dN_{ch}/d\eta$ for the event class of 'all' events (right panel in Figure 6.1). For the other event classes the corresponding formulas are similar, skipping the **trigger-bias correction** for the event class of the triggered events; and skipping the **trigger-bias** and **vertex-reconstruction corrections** for the event class of triggered events with reconstructed vertex. Therefore, Eqs. (6.14), (6.15), and (6.18) change correspondingly (leaving out some of the corrections) for triggered events:

$$\begin{aligned}
P_{\text{trig}}(\eta, z, p_T) &= M^*_{\text{trigvtx}}(\eta, z, p_T) \times C_{\text{trk}}(\eta, z, p_T) \times C_{\text{vtx}}(\eta, z, p_T), && (6.20)\\
I_{\text{trig}}(z,n) &= E^*_{\text{trigvtx}}(z,n) \times \tilde{C}_{\text{vtx}}(z,n) \text{ for } n > 0, && (6.21)\\
I_{\text{trig}}(z,0) &= E^*_{\text{trig}}(0) \times \alpha^*(z) \times F(z), && (6.22)
\end{aligned}$$

and for triggered events with reconstructed vertex:

$$\begin{aligned}
P_{\text{trigvtx}}(\eta, z, p_T) &= M^*_{\text{trigvtx}}(\eta, z, p_T) \times C_{\text{trk}}(\eta, z, p_T), && (6.23)\\
I_{\text{trigvtx}}(z,n) &= E^*_{\text{trigvtx}}(z,n) \text{ for } n > 0, && (6.24)\\
I_{\text{trigvtx}}(z,0) &= 0. && (6.25)
\end{aligned}$$

$P_{\text{trig/trigvtx}}$ and $I_{\text{trig/trigvtx}}$ are used to calculate $dN_{ch}/d\eta$ analog to Eq. (6.19).

Results and verification

Corrected $dN_{ch}/d\eta$ distributions are shown in Figure 6.4. The left panel shows the corrected result based only on the reconstruction output, that will become also available

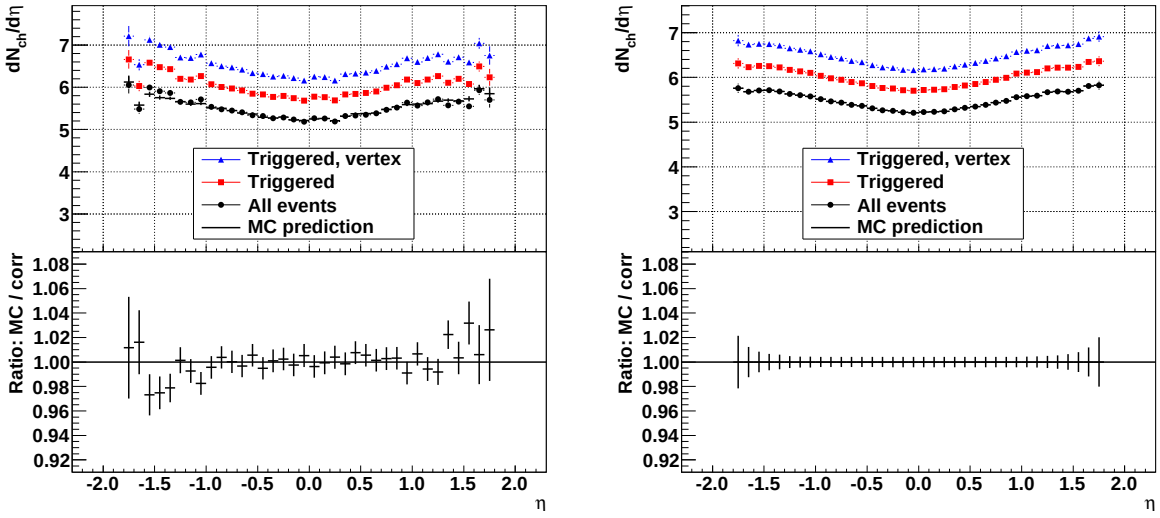

Figure 6.4: Verification of the analysis method.

The figure illustrates the result of the analysis using the SPD. The result in the left panel is based on the reconstruction of 50 000 events (Pythia at $\sqrt{s} = 10\,\text{TeV}$). The right panel shows the verification of the method that makes use of MC information and a different event sample (see text). In the upper part of each panel three different $dN_{ch}/d\eta$ distributions are shown: they are based on 1) events that are triggered and have a reconstructed vertex position (blue triangles), 2) triggered events (red squares), and 3) inelastic events (black circles). In the lower part the ratio of the MC input over the analysis result for the inelastic event sample is shown.

with real data. The small deviations in the ratio are attributed to resolution effects. The plot shows the $dN_{ch}/d\eta$ for the three event samples, triggered events with reconstructed vertex, triggered events, and inelastic events. The MC input distribution for inelastic collisions and its ratio to the corrected result are also shown.

The method is verified with the following analysis procedure: the same event sample is used as analysis input sample and to derive the correction factors. In addition, the measured quantities M^*_{trigvtx} and E^*_{trigvtx} are determined neglecting resolution effects. This is performed by using the MC vertex position instead of the reconstructed vertex position and the MC particle parameters for all reconstructed tracks or tracklets. Furthermore, background is neglected. The result of the verification can be seen in the right panel of Figure 6.4. The ratio is unity and shows that the corrected result corresponds exactly to the MC input.

In the following sections the corrections introduced in this section are described in detail.

6.2.2 Track-to-Particle Correction

The number of reconstructed tracks differs from the number of primary charged particles. This is due to a number of different effects: e.g. the acceptance of the detector, the detector and reconstruction efficiency, and secondaries.

A further contribution, tracks from background sources (cosmic rays or beam-gas events), is neglected since these tracks have a very small probability to point back to the vertex and they are therefore rejected by the cut requiring association with the primary vertex. The systematic uncertainty arising from beam-gas events that pass the trigger are nevertheless discussed in Section 6.3.5.

The **track-to-particle correction** takes all these effects into account and is calculated using the ratio between the number of primary charged particles and the number of selected tracks after the detailed detector simulation and reconstruction. The **track-to-particle correction** uses the event sample of triggered events where the vertex position has been reconstructed.

The **track-to-particle correction** depends in principle on many variables: examples are p_T, η, φ, particle species, charge, vtx-z, and the multiplicity of the event. In this analysis, the correction is determined as a function of η and vtx-z for the SPD and as a function of η, vtx-z, and p_T for the TPC and is integrated over the remaining variables. This integration can of course be subject to systematic errors (i.e. in the case when the event generator does not describe the data well). In most cases, however, these effects are expected to be negligible. The integration over φ does not impose any systematic effects since the collisions are on average azimuthally symmetric. On average the collisions have the same number of positive and negative particles. Therefore differences in the **track-to-particle correction** between positive and negative particles e.g. due to the different absorption cross-sections for protons and antiprotons cancel out. However, the charge dependence could be easily introduced in the corrections for the TPC by assigning a negative p_T for negatively charged particles.

The detector occupancy of the detector in p+p collisions is very low, which means that there are negligible saturation effects even in high-multiplicity p+p events. Thus, the integration over multiplicities does not impose systematic effects in p+p collisions. On the contrary, this may not be applicable for heavy-ion measurements where saturation effects get significant.

The integration over particle species is not as easy to justify, since the correction will differ significantly for the different particle species at low p_T (see Figure 6.11 on page 129). This has less effect for the measurement with the SPD because only very few particles are below the p_T cut-off, but becomes important for the measurement with

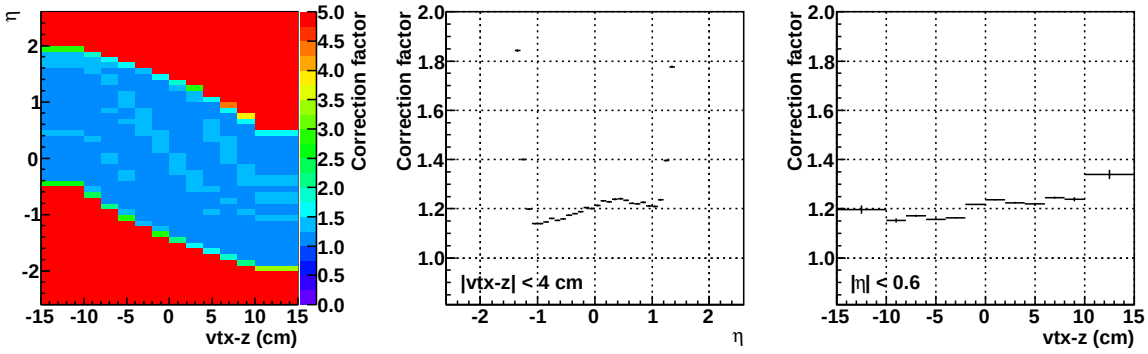

Figure 6.5: **Track-to-particle correction** for the SPD.
The correction is shown in the η vs. vtx-z plane (left panel) and as projections on the η (center panel) and the vtx-z axis (right panel).

the TPC. For the TPC the corrections can in principle be applied independently for different particle species which, however, requires particle identification and is therefore not suitable for the analysis of first data. Instead the default particle composition given by the event generator is used and systematic uncertainties arising from the uncertainty in the particle composition are estimated (see Section 6.3.2).

SPD

Figure 6.5 shows the **track-to-particle correction** for the SPD in the two-dimensional plane, together with projections on the η and vtx-z axis. The projections are for visualization purposes only and are not used in the analysis. The projections only consider a limited range in the integrated variable to prevent that large factors outside of the acceptance region influence the shown factors.

The correction factor is between 1.1 and 1.25 and increases towards the edges. The vertex position influences the acceptance in η, which can be seen in the left panel. The shape in the center and right panels is caused by the inactive modules in the SPD (see Section 5.5): the increased correction factor for positive vtx-z is explained by the fact that the inactive areas are predominantly on that side. The same reason causes the η-dependence of the correction factor (the positive η side corresponds to the positive vtx-z side).

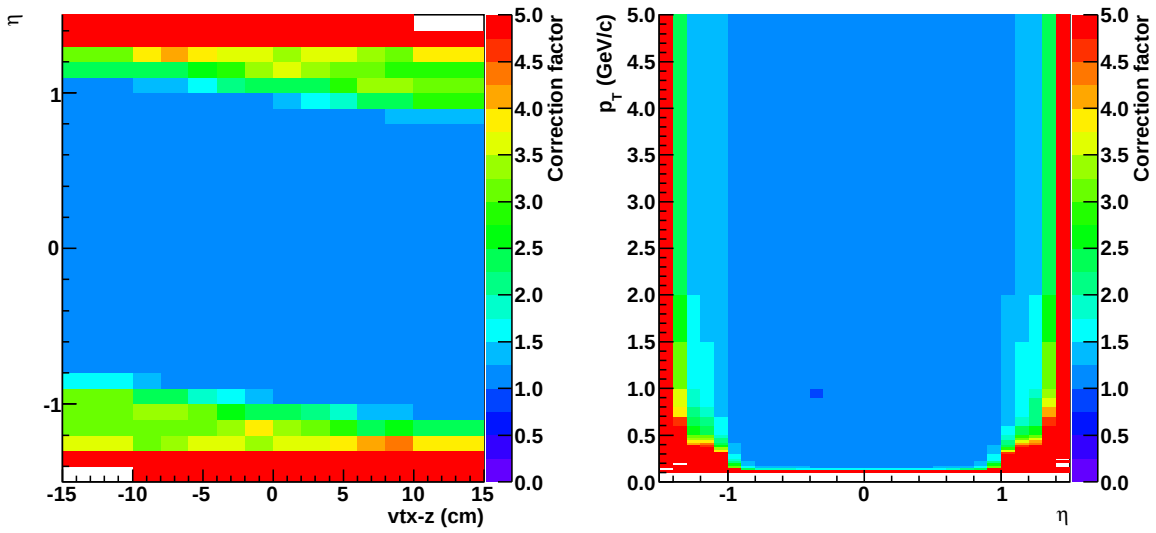

Figure 6.6: **Track-to-particle correction** for the TPC.
The correction is determined in three dimensions as a function of η, vtx-z, and p_T. For visibility, projections to the η vs. vtx-z (left panel) and p_T vs. η (right panel) planes are shown.

TPC

Figure 6.6 shows projections of the **track-to-particle correction** for the TPC. In the left panel the dependence on the z-vertex position can be seen. The effect is much smaller compared to the SPD case owing to the larger distance of the detector from the vertex. The dependence is due to the additional material that needs to be traversed depending on the vtx-z. The right panel shows that the correction flattens at high-p_T where absorption and decay no longer play a significant role. The TPC measures high-p_T particles up to $|\eta| < 1.4$, i.e. p_T above 750 MeV/c (right panel). However, this is not the case for low-momentum particles. Therefore, with the TPC the $dN_{ch}/d\eta$ distribution can only be determined in the region $|\eta| < 1.0$. The projection to the p_T vs. vtx-z plane (plot not shown) does not show any significant correlations.

6.2.3 Vertex-Reconstruction Correction

The **vertex-reconstruction correction** takes into account the bias introduced by events that are not counted because their vertex position was not reconstructed by the vertex-reconstruction algorithm. The correction is determined at the track level and the event level. Whether or not the vertex position is reconstructed is an event

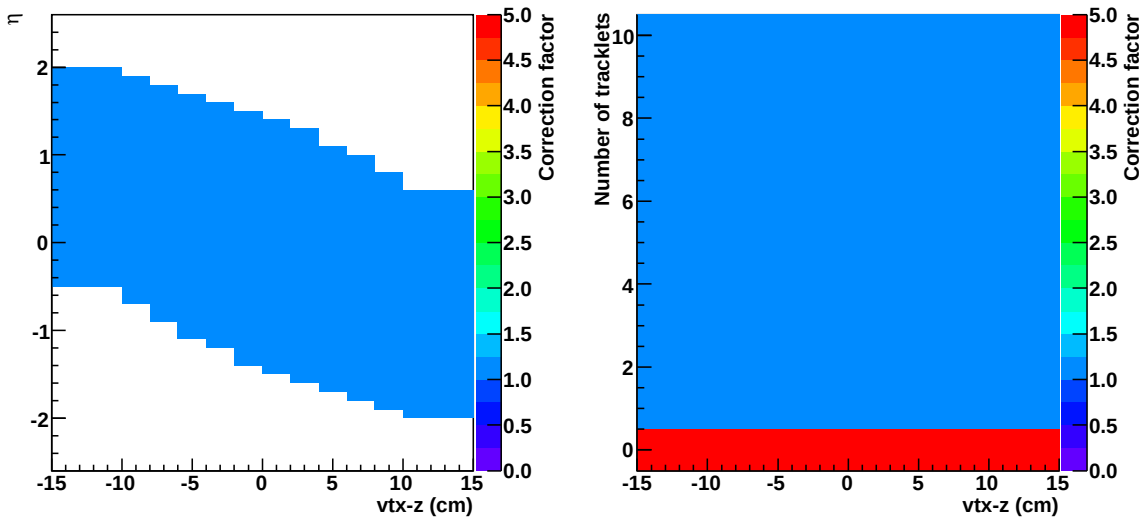

Figure 6.7: **Vertex-reconstruction correction** for the SPD at the track level (left panel) and event level (right panel).

property. However, the vertex requirement can bias the kinematics of the events, which also requires a track-level correction.

By design, this correction is dependent on the MC simulation. However, when real data is available it will be possible to compare some of the properties of triggered events without a reconstructed vertex to the corresponding events in the simulation. This will give an indication for how well the MC describes these events and may allow the minimization of the systematic uncertainty.

SPD

Figure 6.7 shows the **vertex-reconstruction correction** at the track level and the event level. The event-level correction is unity as soon as the number of tracklets is at least one. This is due to the fact that the vertex position determination is very similar to the tracklet finding procedure. Without tracklets the correction factor is very high and correspondingly the vertex reconstruction efficiency is very low. These bins with a multiplicity of 0 (called 0-bins in the following), however, are not used in the correction procedure (see Eqs. (6.15) and (6.18)). The track-level correction is negligible. Plots that contain values of mostly unity, like this one, will be omitted in the following and will simply be verbally described in the text.

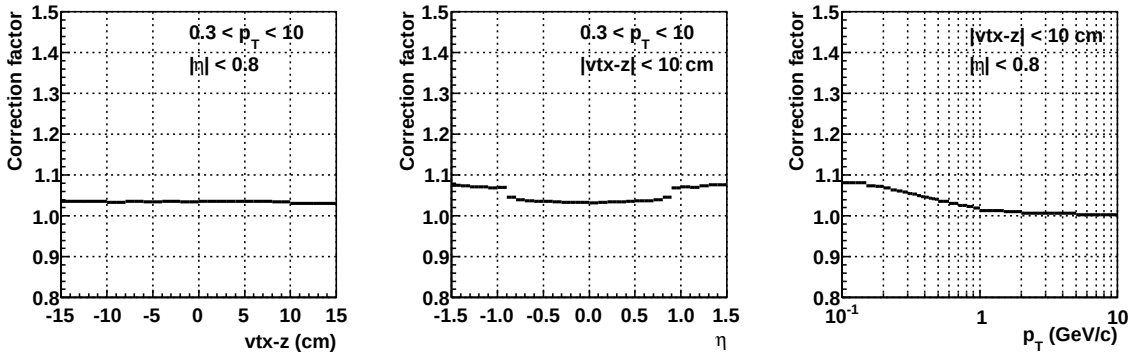

Figure 6.8: The figure shows projections of the track-level **vertex-reconstruction correction** to the vtx-z axis (left panel), the η axis (center panel), and the p_T axis (right panel) for the TPC.

TPC

The event-level **vertex-reconstruction correction** (plot not shown) is, like in the case of the SPD, deviating from unity only in the 0-bins, due to the fact that no tracks can be accepted without finding a vertex position in the event.

However, the vertex reconstruction imposes a bias on the event kinematics. Events in a certain kinematical region have a higher probability to have a reconstructed vertex than others. This is reflected in the track-level correction shown in Figure 6.8 as projections to their three axes. No dependence on the vertex position, but a kinematic bias with respect to η and p_T can be seen. Particles in the central region can contribute to the vertex reconstruction; particles with a high $|\eta|$ cannot contribute to it (because their tracks are not measured). Therefore, for events containing such tracks the efficiency to find the vertex position is lower than for events with particles in the central region. Events containing tracks with large p_T have a higher chance for the reconstruction of the vertex position than events containing low-p_T tracks.

6.2.4 Trigger-Bias Correction

The **trigger-bias correction** takes into account the bias arising from the difference between the triggered event sample and the collision sample of interest; this could be the sample of inelastic or NSD collisions. Both corrections are applied to the data, yielding the $dN_{ch}/d\eta$ for the two different collision types (INEL and NSD). These two corrections act in two different directions: the correction to the inelastic sample has to correct for events that have not been seen by the trigger. The correction to the NSD

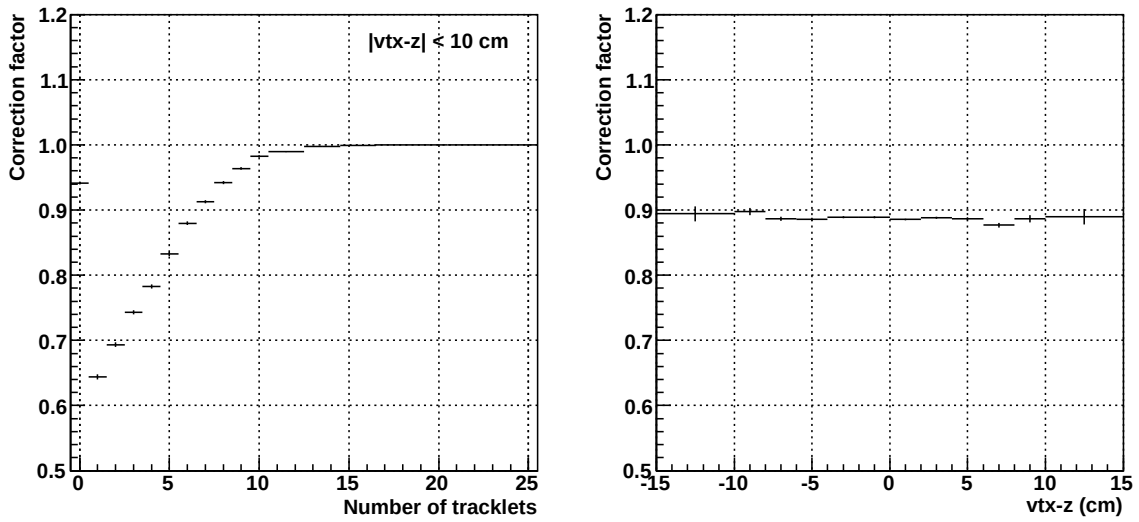

Figure 6.9: The figure shows the event-level NSD **trigger-bias correction** as a function of multiplicity (left panel) and vertex position (right panel) for the SPD.

event sample has to correct for unseen events as well as to remove SD events. Therefore, on average the first correction is above unity, the latter below unity.

The correction is based on the MC simulation and cannot be cross-checked with real data; this makes it purely model-dependent. However, there are means to constrain the contributions of the different event classes (ND, SD, and DD); this is explained in the systematic uncertainty assessment in Section 6.3.1.

SPD

The inelastic **trigger-bias correction** (plot not shown) is only different from unity in the 0-bins (where it is approximately 2.2). No bias on the kinematics due to the trigger has been observed.

Figure 6.9 shows the NSD **trigger-bias correction**: the event-level correction is shown as a function of the multiplicity and the vertex position. The correction to the NSD sample requires the removal of SD events which results in a correction factor lower than unity for a multiplicity of up to 15. No SD events are found with a higher multiplicity (in the Pythia simulation). The track-level correction (plot not shown) does not show any dependence on η or the vtx-z. Its average value is 0.95: the average multiplicity in the triggered sample is higher than in the NSD sample, which is corrected in this step.

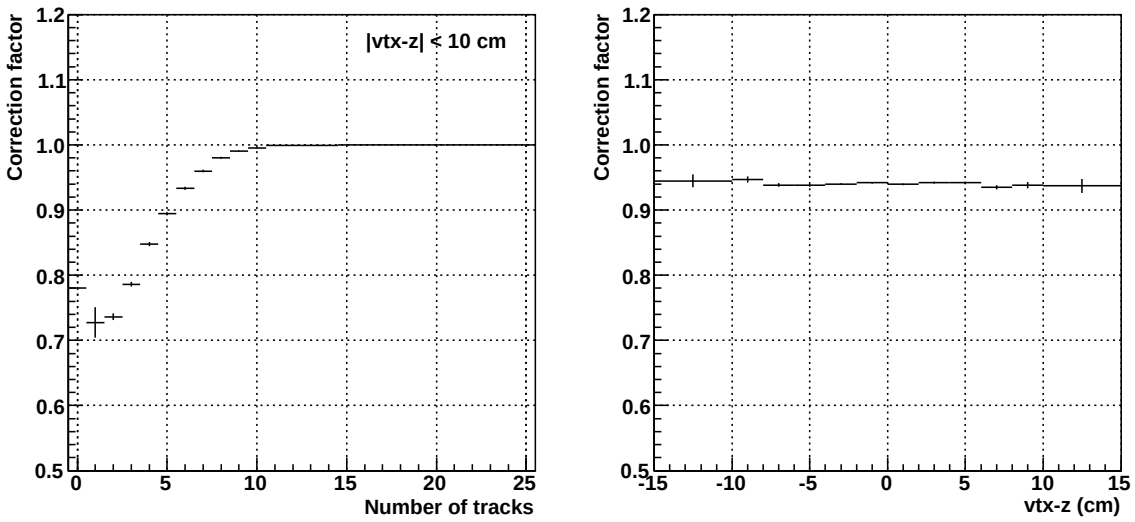

Figure 6.10: The figure shows projections of the event-level NSD **trigger-bias correction** as a function of multiplicity (left panel) and vertex position (right panel) for the TPC.

TPC

Similar to the SPD case, the inelastic **trigger-bias correction** (plot not shown) is only different from unity in the 0-bins (where it is approximately 1.3)[2]. No bias on the kinematics due to the trigger has been observed.

Figure 6.10 shows projections of the event-level NSD **trigger-bias correction**. The correction removes SD events which leads to a correction factor smaller than unity. No dependence on the vertex position can be seen.

6.2.5 Low-Momentum Cut-Off Correction

Particles below a certain p_T are lost or measured with very low efficiency. This is on the one hand due to the magnetic field. On the other hand it is due to multiple scattering which is proportional to $1/\beta p$ and to energy losses by ionization proportional to $1/\beta^2$ [Ams08]. This gives rise to the so called p_T cut-off for which a correction is needed.

[2]This value is smaller than the value for the SPD. The **trigger-bias correction** at a (reconstructed) multiplicity of 0 determines how likely it is for an event that has no reconstructed tracklets or tracks to be triggered. The TPC has a smaller acceptance, therefore this kind of events has a larger probability to be triggered than events without tracklets in the SPD, hence the correction factor is smaller.

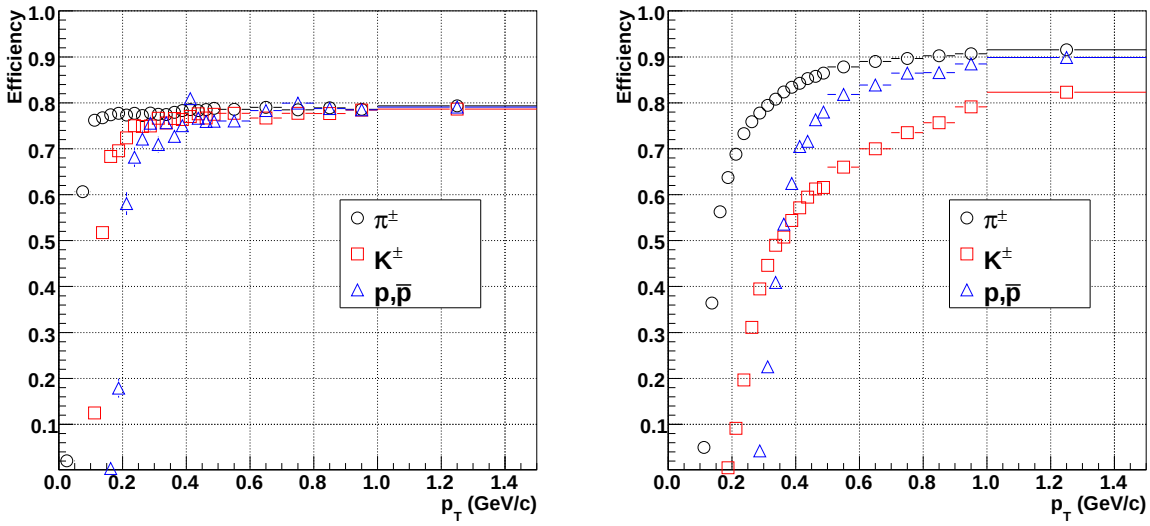

Figure 6.11: Physical tracking efficiency.

The figure shows the physical efficiency to track primary particles in the SPD (left panel) and TPC (right panel) as a function of p_T. The efficiency is shown separately for pions, kaons, and protons (Pythia at $\sqrt{s} = 10\,\text{TeV}$).

Naturally, the cut-off is different if the measurement is performed with the SPD or the TPC. The outer layer of the SPD is located at a radius of 7.6 cm and the outer edge of the TPC is found at $r = 278\,\text{cm}$. However, tracks are accepted already when they have 50 clusters, corresponding[3] to roughly $r = 123\,\text{cm}$. The nominal magnetic field deflects tracks in the SPD with a p_T of 35 (55) MeV/c in such a way that they are not accepted by the $\Delta\varphi$-cut of 0.08 (0.05) rad. For the TPC, tracks spiral at a $p_T \approx 92\,\text{MeV}/c$ such that they cannot reach 50 or more clusters. The rather small difference (compared to the difference in radii) is explained by the fact that in the TPC tracks can be reconstructed even with a quite large curvature, while the reconstruction in the SPD requires nearly a straight line, however, on a short distance.

Figure 6.11 shows the efficiency to measure particles as a function of p_T for the SPD and the TPC in the acceptance region.[4] Due to their different distances from the beam-line, the behavior is quite different. The maximum reached at large p_T is governed by the fraction of inactive areas (TPC: about 10% and SPD: about 15%, see Section 5.5) and reconstruction inefficiencies. In both cases, it can be seen that the efficiency, espe-

[3] Approximating a straight line and considering that the pads in the inner chamber of the TPC are smaller.

[4] Compared to the **track-to-particle correction** the efficiency is a few percent lower than the inverse of the correction factor (Figure 6.5 and 6.6). The reason is that secondaries and particles that are found several times are excluded from the efficiency.

Detector	Radius	Eff. p_T cut-off (GeV/c)			Particles below cut-off (%)			
	(cm)	$\pi^{\pm}$	$K^{\pm}$	p,$\bar{\text{p}}$	$\pi^{\pm}$	$K^{\pm}$	p,$\bar{\text{p}}$	Total
SPD	7	0.05	0.13	0.20	0.6	1.7	2.7	0.7
TPC	≈ 123	0.15	0.35	0.35	10.4	22.5	12.3	11.6

Table 6.1: Low-momentum cut-off.
The table shows the effective p_T cut-off (see text) and the yield of particles subdivided into different species below the cut-off for the SPD and TPC (Pythia at $\sqrt{s} = 10\,\text{TeV}$).

cially for kaons and protons, degrades at a higher p_T than the previously stated values corresponding to the magnetic field. Kaons and protons have a higher mass than pions and thus a larger $1/\beta$ ($1/\beta^2$)-factor at the same momentum which gives rise to more multiple scattering and a larger energy loss. Furthermore, decays reduce the total kaon yield. Table 6.1 contains the momentum values where the efficiency drops below 50% for the different particle species (this is referred to as the *effective p_T cut-off*). Also given is the percentage of particles below this cut-off.

For the SPD, the amount of particles below their effective p_T cut-off is about 0.7%. This small effect is included in the **track-to-particle correction** and a dedicated p_T cut-off correction is not needed.

For the TPC, the amount of particles lost is much higher: approximately 11.6% of the particles are below their effective p_T cut-off. In the analysis, a p_T cut-off is applied with a $p_{T,\text{min}}$ between $0.15\,\text{GeV}/c$ and $0.25\,\text{GeV}/c$. Each $p_{T,\text{min}}$ value requires a corresponding correction factor. Several values should be used together with the corresponding correction to check the stability of the result.

The p_T cut-off correction is calculated as a function of η by dividing the total number of primary particles by the number of primary particles above the p_T cut-off. The correction factors for a $p_{T,\text{min}}$ of $0.15\,\text{GeV}/c$, $0.20\,\text{GeV}/c$, and $0.25\,\text{GeV}/c$ can be seen in Figure 6.12. The shape stems from the transformation from y to η which depends on p_T. The correction factor as a function of y is flat. The correction at $0.2\,\text{GeV}/c$ is $1.2 - 1.25$ depending on η. The systematic uncertainty introduced by this correction is discussed in Section 6.3.3.

6.2.6 Estimation of the Required Simulated Data

The limited statistics of simulated events used to obtain the correction factors gives rise to a statistical uncertainty. The number of simulated events should be sufficiently

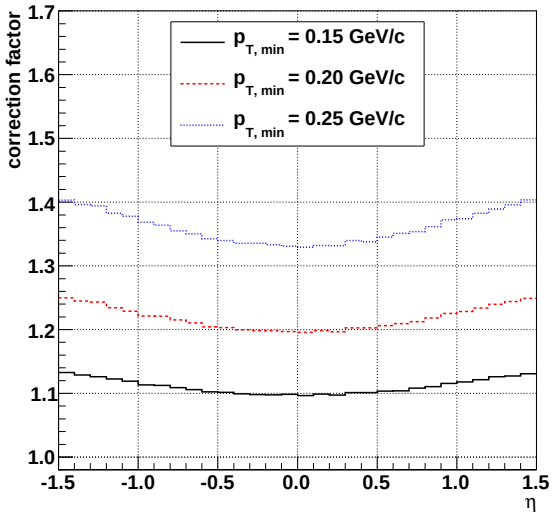

Figure 6.12: Low-momentum cut-off correction.
The correction is shown as a function of η for inelastic collisions. The p_T cut-off at 0.2 GeV/c excludes about 20% of the primary charged particles.

high such that this uncertainty is much smaller than other contributions to the uncertainty on the final distribution like the statistical error on the measured data and the systematic uncertainties. This section shows how the statistical errors are calculated. The errors are determined for the available statistics of 270 000 events and an estimate for the required statistics is given.

The variance of a binomial-distributed variable $a(r)$ (the probability for r successes out of N trials of a Bernoulli-experiment with success-probability p) is

$$Np(1-p). \tag{6.26}$$

Correspondingly, the variance of $a(r)/N$ is

$$\frac{p(1-p)}{N}. \tag{6.27}$$

Consider the calculation of an efficiency E: M out of N events fulfil a given criterion, the probability to find an event with the given criterion is

$$E = \frac{M}{N}. \tag{6.28}$$

M follows a binomial distribution (N trials with probability p which is the 'true' probability to find the criterion; in the limit of large N: $p \approx M/N$). Thus Eq. (6.27) gives the uncertainty on E:

$$\sigma_B(M,N) = \sqrt{\frac{\frac{M}{N}(1-\frac{M}{N})}{N}} \tag{6.29}$$

which are *binomial errors*. Note that such an error estimation is only valid in the case without background and duplicated tracks[5] (that occur e.g. during the tracking). In this analysis the binomial-error calculation applies to all corrections except:

- the track-level **track-to-particle correction** because of secondaries and duplicated tracks, and
- the NSD **trigger-bias correction** because here the SD events need to be subtracted. Thus the number of triggered events does not follow a binomial distribution with respect to all NSD events.

The following considerations allow the calculation of the errors on the **track-to-particle correction** (Eq. (6.4)):

$$C_{\mathrm{trk}}(\eta, z, p_T) = \frac{G_{\mathrm{trigvtx}}(\eta, z, p_T)}{M_{\mathrm{trigvtx}}(\eta, z, p_T) + B(\eta, z, p_T)} \tag{6.30}$$

which for simplicity is now written as $C = G/(M + B)$. For practical reasons the relative error of C^{-1} is studied, which is identical to the *relative* error of C. It can be written as

$$C^{-1} = \frac{M + B}{G} = \frac{M_1}{G} + \frac{M_2 + B}{G}. \tag{6.31}$$

M_1 refers to tracks of primary particles where even in the case of a duplicated reconstructed track only one is counted. M_2 contains tracks from secondaries and the duplicates that have not been considered for M_1. B is the background, i.e. tracks not associated to MC particles such as those defined at the beginning of Section 6.2.1.

M_1 is binomially distributed with $p = M_1/G$ as the probability to track a single particle, i.e. the tracking efficiency. The error on the first term of Eq. (6.31) is thus to be calculated following Eq. (6.29). Although M_2 and B are both somehow correlated with G (many primaries will also cause many secondaries) the error calculation is not *a priori* clear. For the second term the error is thus determined based on the statistical error of the numerator: $\sqrt{M_2 + B}/G$, which can be seen as an upper limit. Note that basing the calculation upon the denominator ($\sqrt{G}$) results in a larger absolute error, but in a smaller *relative* error, because in practice $M_2 + B < G$.

For the available data sample, the left panel of Figure 6.13 shows the relative errors on the first term of Eq. (6.31). It can be seen that in the central region ($|\text{vtx-}z| < 4\,\text{cm}$) the error is below 0.4%. It increases towards the edges, but is still around 1% at $\pm 10\,\text{cm}$. The increase is due to the z-vertex distribution in the simulated events. It can

[5] Under certain conditions a track can be reconstructed twice, e.g. in the TPC a particle can suffer a large energy loss and continue its path slightly altered. This can be interpreted by the reconstruction as two separate tracks.

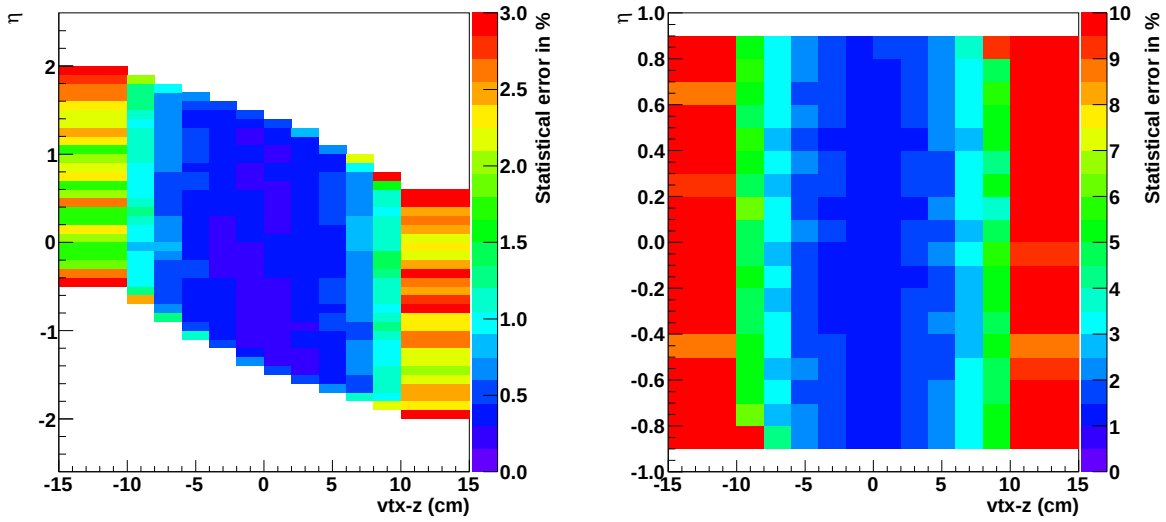

Figure 6.13: Statistical error of the **track-to-particle correction**. Shown is the relative error on the first term of Eq. (6.31), which is larger than the error on the second term. The left panel shows the result for the SPD and the right panel for the TPC in the slice $0.2\,\mathrm{GeV}/c < p_T < 0.3\,\mathrm{GeV}/c$.

be reduced by simulating a flat vertex distribution. The error on the second term of Eq. (6.31) is smaller having qualitatively a similar behavior.

Correspondingly for the TPC, Figure 6.13 (right panel) shows the relative errors on the first term of Eq. (6.31). Due to the fact that this correction is determined in three dimensions (also p_T), the available statistics in each bin is lower and the relative error is consequently larger. In the figure the slice $0.2\,\mathrm{GeV}/c < p_T < 0.3\,\mathrm{GeV}/c$ is shown. The relative error does not increase for increasing p_T because the bin size is larger at higher p_T. Data at $p_T < 0.2\,\mathrm{GeV}/c$ (p_T cut-off) is not considered. Therefore, the shown figure represents the largest relative error present in the correction. It is about 1.5% in the central region of vtx-z, but increases towards the edges and is relatively large $(3-6\%)$ for vertex positions around $\pm 8\,\mathrm{cm}$ with respect to the nominal vertex position. The error of the second term of Eq. (6.31) is smaller.

The error on the NSD **trigger-bias correction** can be calculated in a similar fashion. The event-level correction, Eq. (6.10), can be written as:

$$\tilde{C}_{\mathrm{trig}}^{-1} = \frac{E_{\mathrm{trig}}}{E_{\mathrm{NSD}}} = \frac{E_{\mathrm{trig,NSD}}}{E_{\mathrm{NSD}}} + \frac{E_{\mathrm{trig,SD}}}{E_{\mathrm{NSD}}} \tag{6.32}$$

where $E_{\mathrm{trig,NSD}}$ and $E_{\mathrm{trig,SD}}$ are the triggered NSD and SD events, respectively. $E_{\mathrm{trig,NSD}}$ is distributed binomially, thus the error can be calculated following Eq. (6.29). The

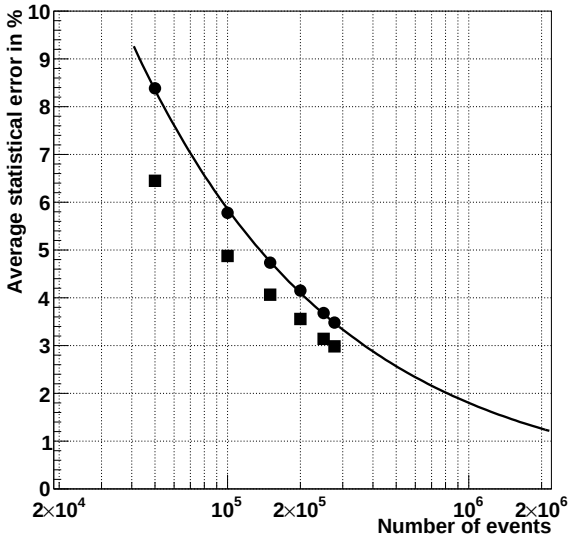

Figure 6.14: Statistical error as a function of the size of the event sample. Relative statistical error in the region 6 cm $<$ |vtx-z| $<$ 10 cm (low statistics region in the plot above) of the **track-to-particle correction** of the TPC as a function of the number of events. Shown are the errors on the first (circles) and the second term (squares) of Eq. (6.31). The larger errors (first term) are fitted and extrapolated (line).

error of the second term is calculated to be $\sqrt{E_{\mathrm{trig,SD}}}/E_{\mathrm{NSD}}$. The error on the track-level correction can be calculated in the same way.

For the SPD, the error on the event-level correction is about 0.5% for vertices near the nominal interaction point and increases, following the trend described above, to $1-2\%$ at |vtx-z| $\approx$ 10 cm. At the track-level the error is below 0.5%. For the TPC, the error at the event-level correction is similar. At the track level the error is throughout below 1%.

Other non-negligible uncertainties are found only in the statistical uncertainty on the inelastic **trigger-bias correction**. The error on the 0-bins (the other bins are unity) is 1% in the center and a few percent towards large |vtx-z| (for both SPD and TPC).

In summary, the statistical errors are all below 1% for the SPD in the central region. Larger values are found at larger |vtx-z|. For the TPC the errors are slightly larger; they also increase towards larger |vtx-z|. Errors of 1% or less are considered negligible compared to the systematic uncertainties that will be estimated in the following section. This has to be achieved for all corrections in all considered regions. The uncertainties increase towards larger |vtx-z| which is due to the vertex distribution in the simulated data. One solution is to simulate events with a flat vertex distribution. Alternatively,

an increased number of events would be needed if the vertex distribution is required to match the data. The largest uncertainty arises from the **track-to-particle correction** in the TPC. Figure 6.14 shows the average relative error in the region with the largest errors as a function of the number of events. From extrapolating it can be seen that with about 1 – 2 million simulated events, a statistical uncertainty close to 1% even at the edges can be achieved. For the SPD, 500 000 events are sufficient. Another option is to exclude events with larger |vtx-z| from the measurement. Such events constitute only a small fraction[6] of the total number of events taken. Furthermore, in the case of the measurement with the TPC, events with vertices far from the nominal interaction point do not increase the accessible η-range.

A value of 1% is estimated as an upper limit for the systematic uncertainty due to the corrections.

6.3 Systematic Uncertainties

In this section the systematic uncertainties are discussed and evaluated. Although systematic uncertainties might be correlated they are studied independently to estimate their size. Dependencies between different uncertainties are mentioned where appropriate.

For the evaluation of a given systematic effect the following workflow is used. It is recalled that the input sample is divided into two parts: the correction input sample is used to create the corrections and the analysis input sample is used as input data for the analysis. Any change whose effect is to be studied (e.g. a change of cross-sections) is applied to the correction input sample. Subsequently, the correction is re-extracted from that sample. At this stage two corrections exist (one from the original input sample and one from the modified input sample). The analysis is now performed twice on the analysis input sample, once for each correction. The ratio is calculated between the two resulting $dN_{ch}/d\eta$ spectra and allows the systematic effect to be estimated. If applicable this is performed at different stages of the correction procedure.

In certain cases a different approach was used; this will be described together with its motivation in the corresponding section.

The following systematic effects are studied for both analysis methods, using data from the SPD and the TPC:

- the uncertainty in the cross-sections of the collision processes in the event generator;

[6] At $\sqrt{s} = 10\,\text{TeV}$ only about 10% of the events are found in $6\,\text{cm} < |\text{vtx-}z| < 10\,\text{cm}$.

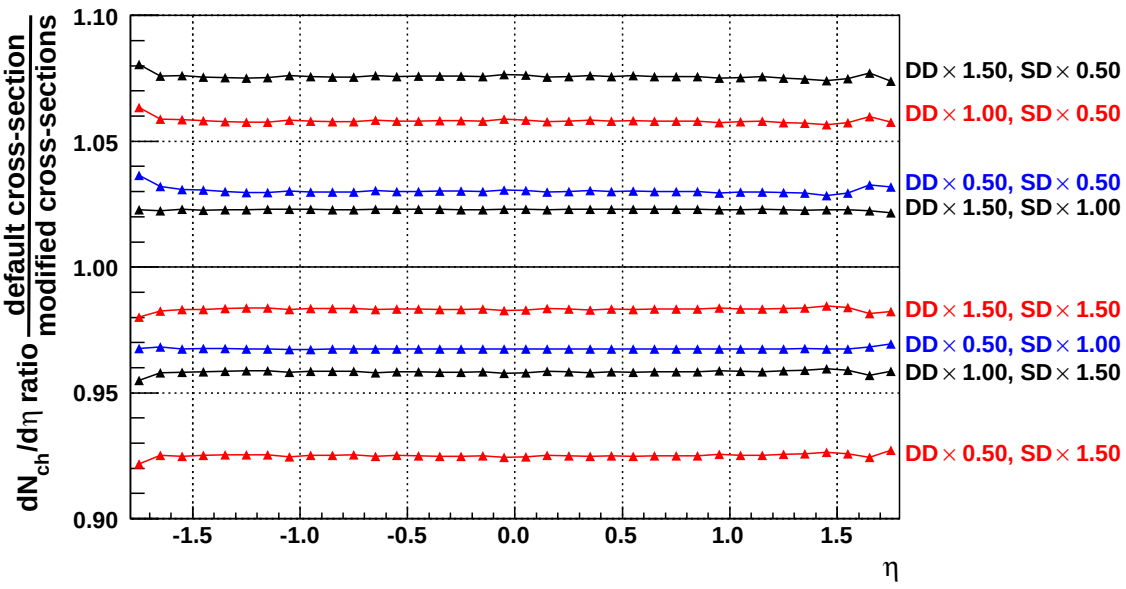

Figure 6.15: Systematic uncertainty due to unknown cross-sections (SPD, NSD). The figure shows the ratios between the $dN_{ch}/d\eta$ distribution calculated from the standard corrections and the distributions calculated from the corrections estimated with changed relative cross-sections. The plot is for the NSD event sample using data from the SPD. The changes in the cross-sections are indicated on the right of the plot. The systematic uncertainty is about 8%. The lines are drawn only to guide the eye and statistical errors are omitted.

- effects due to an incorrect assumption concerning the particle species abundances in the event generator;
- the effect due to the uncertainty in the p_T spectrum below the p_T cut-off;
- further effects due to different assumptions in the event generator by comparing Pythia and Phojet;
- the effect of beam-gas, beam-halo and pile-up events;
- the influence of an incorrect estimation of the material budget in the simulation software;
- the effect of remaining uncertainties in the alignment of the subdetectors;
- effects due to the tracklet and track selection cuts.

	SPD	TPC
INEL	2%	2%
NSD	8%	8%

Table 6.2: Systematic uncertainty due to unknown cross-sections.

6.3.1 Cross-sections of Physics Processes

The trigger efficiency and the vertex-reconstruction efficiency (for triggered collisions) are different for different processes (ND, SD, and DD). The corrections will therefore depend on the relative cross-sections of these processes with respect to each other. The absolute values have no influence because the $dN_{ch}/d\eta$ distribution is normalized to the number of events. The values predicted by Pythia have been given in Table 1.3 (page 24). In order to study the effect of a change in the relative cross-sections the corrections have been calculated by changing the diffractive cross-sections to 50% and 150% of the Pythia values, i.e. $\sigma_{\mathrm{SD}} = 7{-}28\,\mathrm{mb}$ and $\sigma_{\mathrm{DD}} = 4.6{-}18.6\,\mathrm{mb}$ at $\sqrt{s} = 10\,\mathrm{TeV}$.

A study was done for the **vertex-reconstruction** and the **trigger-bias correction** together and for each of them independently. The effect is strongest when both corrections are calculated using the modified cross-sections and these are the numbers presented here. Figure 6.15 shows exemplarily the effect of the changes in the relative cross-sections on the $dN_{ch}/d\eta$ distribution for the NSD event sample (SPD). It can be concluded that changing the relative cross-sections by ±50% changes the result of the analysis by about 8%. The effect is linear, i.e. changing the values by ±25% changes the effect to about 4%. For the correction to the inelastic event sample, the effect is only 2% for a change by ±50%. All cases are summarized in Table 6.2.

Note that the effect on the correction for inelastic events depends only on the difference in the response of the detector to the different processes. For example, if the trigger efficiencies for all processes were the same, no effect on the correction for inelastic events would be seen. It is expected that the values for SPD and TPC are similar because the same trigger is used. However, the different kinematics of the different process types might have an influence on the vertex reconstruction. Therefore, it is necessary to evaluate both cases.

The relative cross-sections can be constrained using information from several triggers, a method that was also used by UA5 [Aln86]. The procedure uses the number of triggered

events N^i_{trig} of several MB triggers and different sets of trigger efficiencies (ϵ^i_{SD}, ϵ^i_{DD}, and ϵ^i_{ND}) derived from the detector simulation. Each trigger i contributes an equation:

$$N^i_{\text{trig}} = N_{\text{SD}}\epsilon^i_{\text{SD}} + N_{\text{DD}}\epsilon^i_{\text{DD}} + N_{\text{ND}}\epsilon^i_{\text{ND}}. \tag{6.33}$$

Solving the system of equations (at least three triggers are needed) yields the number of collisions per process type: N_{SD}, N_{DD}, and N_{ND}. These numbers are derived for a given set of trigger efficiencies, i.e. one event generator. Comparing the results using efficiencies from different event generators (e.g. Pythia and Phojet) constrains the relative cross-sections. More details can be found in [Bom09a].

Once the cross-sections at LHC energies are measured, this uncertainty will decrease significantly. The TOTEM experiment expects to measure the ND, SD, and DD cross-sections with a precision of 0.06 mb, 0.6 mb, and 0.1 mb, respectively [TOT04], which is significantly smaller than the uncertainties considered here.

6.3.2 Particle Composition

The relative abundance of different particle species have an influence on the corrections. It is assumed that well-calibrated particle identification in ALICE is only available after the very first analysis, thus the corrections rely on the abundances given by the event generator. This may introduce a systematic error on the result. The magnitude of this systematic error can be studied by changing the particle abundances in the generator.

Pions, kaons, and protons (and anti-protons) make up more than 98% of all particles in Pythia events. Therefore, only these have been considered for this study. The reconstruction efficiency differs between the particle species which was shown in Figure 6.11 (page 129) as a function of p_T. Correction factors were determined from events with modified relative pion, kaon, and proton abundances. The number of pions was kept constant, while the number of kaons or protons was increased or reduced by 50%. Similar to the cross-sections case, only the relative abundances have influence on the final result.

For this study the correction was created in a two-step process. First, the number of generated particles and measured tracks are determined for each of the particle species (π, K, and p). A measured track is considered to belong to a given particle species if it is **a)** a track of a primary particle of this species or **b)** a secondary created (e.g. decay and hadronic interaction) by a primary particle of this species. The second step combines these numbers in which some or all of the particle species are enhanced or reduced. Combining the corrections using a factor of unity for all species should result in exactly the same correction as that which is obtained by the normal method. This has been verified.

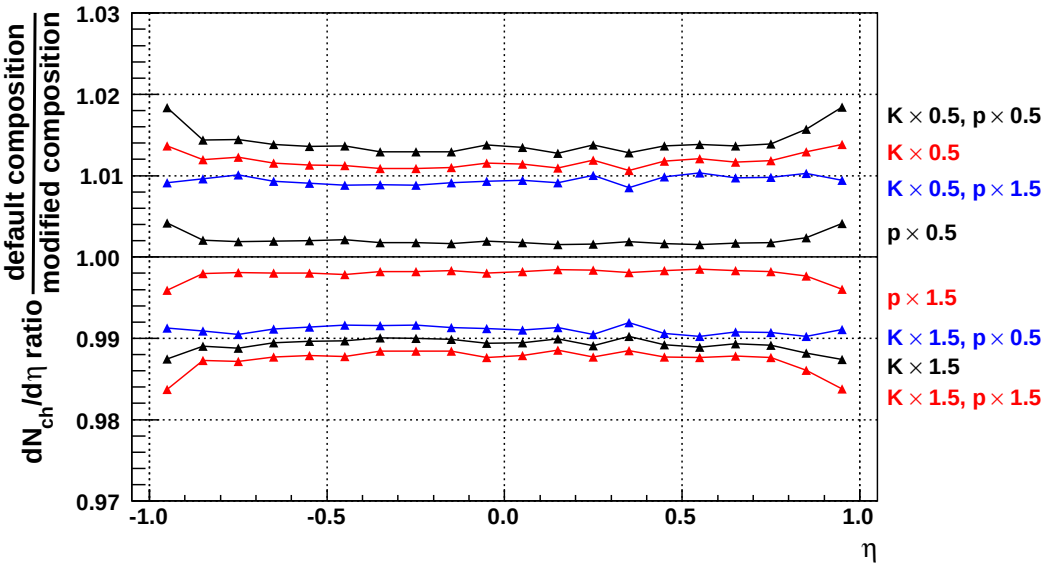

Figure 6.16: Systematic uncertainty related to assumptions of the particle yields (TPC). The figure shows the effect of significantly changing the particle composition. The ratios between the $dN_{ch}/d\eta$ obtained using the standard (Pythia) composition and particle compositions where the kaons and/or protons are changed by ±50% indicate a systematic effect of about 1.5%. The lines are drawn only to guide the eye and statistical errors are omitted.

Various corrections have been created by increasing or reducing the amount of kaons, protons, or both by 50%. The analysis was performed with these different corrections and the ratios between the results were calculated. For the measurement with the SPD, the effect is below 0.1% for all cases. For the TPC, the result is shown in Figure 6.16.

It is concluded that even with the large uncertainty of 50% in the relative yields, the resulting uncertainty on the measurement for the SPD is negligible. The uncertainty on the measurement with the TPC is about 1.5%.

6.3.3 p_T Spectrum

The p_T cut-off correction is only applied in the measurement with the TPC. Nevertheless, an uncertainty due to the p_T cut-off is also present for the SPD measurement. This is further discussed at the end of the section.

The low-momentum cut-off correction, introduced in Section 6.2.5, is model-dependent because the shape of the p_T distribution below the cut-off is unknown. To evaluate the systematic uncertainty imposed by this correction, the p_T cut-off correction was

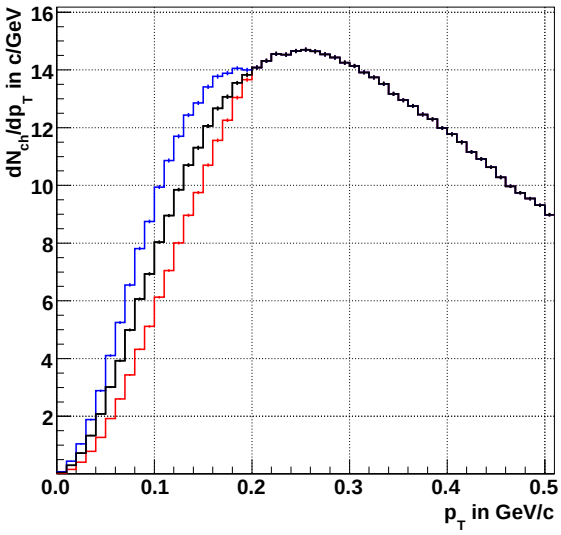

Figure 6.17: Uncertainty due to the p_T cut-off correction.
The figure shows the p_T spectrum in $|\eta| < 1$ predicted by Pythia as well as two modified spectra. These are obtained by changing the predicted spectrum from 0% at $p_T = 0.2\,\mathrm{GeV}/c$ to ±50% at $p_T = 0\,\mathrm{GeV}/c$. These spectra are used to determine the systematic uncertainty imposed by a misestimation of the shape of the p_T distribution to less than 3% for a p_T cut-off at $200\,\mathrm{MeV}/c$.

created from p_T spectra with different shapes below the p_T cut-off. Figure 6.17 shows the p_T spectrum predicted by Pythia. Furthermore, two other p_T spectra are shown that were obtained by changing the spectrum by a percentage increasing linearly from 0% at the p_T cut-off to ±50% at $0\,\mathrm{GeV}/c$. The gradual increase is motivated by the fact that the shape of the spectrum can be measured at (and above) the p_T cut-off, but remains unknown at lower p_T. The result of an extrapolation is therefore expected to be better close to the p_T cut-off than at lower p_T. The change of the spectrum applied here is much larger than the difference in the predictions of the two event generators Pythia and Phojet.

Comparing the p_T cut-off correction factor from the modified spectra to that from the Pythia spectrum indicates the systematic uncertainty on the p_T cut-off correction. It evaluates to less than 3% for the cut-off at $p_T = 0.2\,\mathrm{GeV}/c$. The effect depends on the p_T cut-off. For example, for $p_{T,\mathrm{min}} = 0.15\,\mathrm{GeV}/c$ the effect is 1.5%, for $p_{T,\mathrm{min}} = 0.25\,\mathrm{GeV}/c$ it is about 4.5%. A lower cut-off increases the contamination by secondaries and the uncertainty on the other corrections because the tracking efficiency reduces at low p_T. In practice, several cut-off values should be used and the analysis results should be compared.

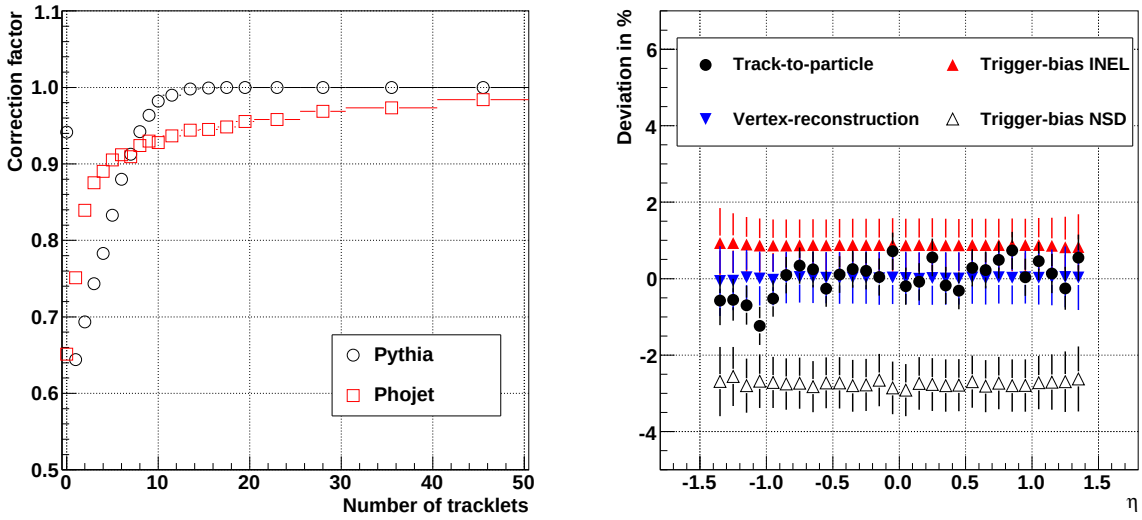

Figure 6.18: Systematic uncertainty due to differences in the event generators (SPD). The left panel compares the NSD **trigger-bias correction** derived with Pythia (black circles) and with Phojet (red squares). The right panel shows the effect on a Phojet sample when applying corrections derived from Pythia after each step of the corrections.

For the measurement with the SPD the corrections are integrated over p_T as explained previously. Thus these corrections rely on the correctness of the p_T distribution used in the simulation software. It is feasible to verify from first data that the measured p_T distribution is close to the one used in the simulation software to a certain extent. However, an uncertainty remains below the p_T cut-off where the p_T spectrum is not measured. To study this effect, the change of spectrum shown in Figure 6.17 is applied to the simulated data used to extract the corrections for the SPD. The overall correction factor changes by about 0.5%, which is correspondingly the associated systematic error owing to the uncertainty in the shape of the spectrum at low p_T.

6.3.4 Event-Generator Assumptions

As an estimate of the uncertainty due to different assumptions and predictions of the event generators, correction factors are determined also from events obtained with Phojet.[7] In this study a sample obtained with Phojet is corrected with the correction factors determined with Pythia. The multiplicity distributions of Pythia and Phojet are significantly different at $\sqrt{s} = 10\,\text{TeV}$, see Figure 8.5 (page 199): the probability

[7] Differences between simulated events by Pythia and by Phojet will be partly discussed in Section 8.2.

Correction	Uncertainty in %	
	SPD	TPC
Track-to-particle	≈ 0	-0.5
p_T cut-off	—	-1.0
Vertex-reconstruction	0	$+0.5$
Trigger-bias (INEL)	$+1.0$	$+2.0$
Trigger-bias (NSD)	-3.0	-3.0
Total (INEL)	$+1.0$	$+1.0$
Total (NSD)	-3.0	-4.0

Table 6.3: Integrated difference between Pythia and Phojet correction factors.

of high-multiplicity events is larger in Pythia than in Phojet. Therefore, in practice a Phojet sample can be corrected with Pythia, but a Pythia sample cannot be corrected with Phojet. Naturally, for measured data it has to be verified that the maximum reached multiplicity in the events used to derive the correction factors is sufficient.

The left panel of Figure 6.18 shows exemplarily the NSD **trigger-bias correction** determined with Pythia and Phojet. This difference, owing to the different kinematics of diffractive events of the two event generators, is the largest observed among all the corrections. The right panel shows a Phojet sample corrected with Pythia for the SPD measurement. Shown is the influence on each of the different correction steps. Table 6.3 summarizes the influence of each correction step for both the SPD and TPC and gives the total contributions after all corrections for the inelastic and the NSD sample. Some of the effects cancel, such that the total effect is smaller than the sum of the single effects. The influence of the **vertex-reconstruction correction** for the TPC is caused by the bias on the kinematics imposed by the vertex-reconstruction requirement, see Section 6.2.3. The largest difference is due to the **trigger-bias corrections**. The opposite effect occurs on the correction to inelastic and NSD events.

For the SPD, a systematic uncertainty of 1% (3%) for inelastic (NSD) events is concluded. For the TPC, some of the deviations cancel in this specific case, but this cannot be generally assumed. Therefore, for the TPC, a total systematic uncertainty of 4% (5%) is estimated for inelastic (NSD) events at $\sqrt{s} = 10\,\text{TeV}$. It is interesting to mention that for $\sqrt{s} = 900\,\text{GeV}$ the uncertainty is not significantly lower. Although, both event generators reproduce measured distributions at $\sqrt{s} = 900\,\text{GeV}$, many other properties differ in the generated events. Note that this uncertainty includes the effect of changes in the cross-sections, the particle composition, and the p_T spectrum which have been

evaluated separately in the previous three sections. Therefore, this uncertainty cannot be added to the uncertainties derived for the individual effects that were studied.

6.3.5 Beam-Gas and Beam-Halo Events

A systematic error on the measurement might be introduced by the presence of beam-gas and beam-halo events. In a typical beam-gas or beam-halo event only a few tracks are in the acceptance of the tracking detectors. Usually the vertex position of these events cannot be reconstructed, thus these events do not lead to the reconstruction of additional tracks. However, these events may cause a trigger and thus influence the overall normalization for the $dN_{ch}/d\eta$ measurement.

The probability that a triggered event is a beam-gas or beam-halo collision was estimated in Section 5.1.1. For the LHC startup scenario it should be less than 4×10^{-4}, which is negligible. Under nominal running conditions, i.e. with all bunches filled and at high luminosity, the probability is about 0.6%, which can be directly interpreted as systematic uncertainty on the normalization and therefore on the measured $dN_{ch}/d\eta$ distribution.

Apart from this estimation, the assessment of the presence and the rate of beam-gas and beam-halo collisions has to be performed from events taken with a trigger on single bunches that pass the detector. This has been discussed in Section 5.1.2.

6.3.6 Pile-Up Events

Pile-up events that occur within the SPD integration time (100 ns) or the opening time of the TPC gating grid (90 μs) might give rise to systematic uncertainties. Collisions that occur in different bunch crossings can be resolved using the V0 detector. The probability for pile-up within the same bunch crossing for the startup scenario was given in Section 3.3 to be 0.37% ($\sqrt{s} = 900\,\text{GeV}$) and 4.9% ($\sqrt{s} = 10\,\text{TeV}$); for nominal running conditions it amounts to 0.36%. Such events can be identified due to the fact that simultaneous collisions have different vertices. Events with more than one reconstructed vertex can be skipped in the analysis. Alternatively, only tracklets and tracks originating from one of the vertices are considered. Thus the uncertainty stems only from events where two collisions have the same vertex within the bounds of the vertex-reconstruction resolution. The probability of such events can be estimated with

a two-dimensional Gaussian distribution taking into account the expected variance in vtx-z (σ) and the vertex resolution (d):

$$\int_{-\infty}^{\infty}\int_{z_1-d}^{z_1+d} G_\sigma(z_1)G_\sigma(z_2)dz_2dz_1. \tag{6.34}$$

An upper limit for the resolution of the vertex reconstruction for low multiplicity events is $d = 0.12\,\text{cm}$ in the case of the SPD and $d = 0.25\,\text{cm}$ in the case of the TPC. The variance σ at $\sqrt{s} = 10\,\text{TeV}$ is expected to be 3.8 cm. With these numbers, the percentage of overlapping vertices is 1.2% and 3.7% for the SPD and TPC, respectively. This probability needs to be multiplied with the previously given probability that a triggered event contains more than one collision. Other means of separating two collisions might allow for further reduction of this contamination.

In summary, the percentage of unresolved pile-up events in the same bunch crossing is larger during startup than for nominal running conditions. Quantitatively, they are below 0.1% (SPD) and below 0.2% (TPC) and are thus negligible for the $dN_{ch}/d\eta$ measurement.

6.3.7 Material Budget

Secondaries are part of the sample used for the analysis despite the track cuts. This is corrected by the **track-to-particle correction**. Therefore, any error in the total number of secondaries produced in the detector simulation leads to an incorrect result.

Secondaries are decay products and particles created in interactions with the detector material. Uncertainties arising from decay products stem from a misestimation of the particle composition in the collision, discussed before in Section 6.3.2. The material is modeled very carefully in the software following the engineering drawings of the detector with the aim to describe the material budget in the tracking volume as close as 5% compared to reality. Methods exist to derive the material budget from the data, e.g. by reconstructing γ-conversions, but results of these studies are not available shortly after the first run.

To assess the effect caused by an incorrect estimation of the detector material, the material budget is varied by 10%. For this study the material budget is changed in the simulation step, but not in the reconstruction step. One option would be to change the dimensions of detector components. However, this would lead to overlapping detector volumes, which would result in technical difficulties in the transport software. Therefore, an alternative approach is chosen: the density of all material types is changed by $\pm 10\%$ which changes the material budget without producing overlapping detector parts.

Accepted Tracks	− 10% material		+ 10% material	
	SPD	TPC	SPD	TPC
Primaries	no change	$(+0.9 \pm 0.4)\%$	no change	$(-0.9 \pm 0.4)\%$
Secondaries	no change	$(-5.0 \pm 1.2)\%$	no change	$(+2.0 \pm 1.2)\%$

Table 6.4: Systematic uncertainty due to uncertainties in the material budget.

A production of 20 000 events was performed for this study. No effect is seen on the number of measured tracklets in the SPD. This can be explained by the very low material budget of the pixel layers ($X/X_0 \approx 1\%$ per layer). However, the number of reconstructed TPC tracks changes due to the change of the total material budget. Table 6.4 shows the change in the number of primary and secondary tracks. For the case of the TPC, only tracks above a p_T of $0.2\,\mathrm{GeV}/c$ that pass the track cuts are considered.

As expected the number of primaries reduces when increasing the material budget, while at the same time the number of secondaries increases. The total effect on the corrected $dN_{ch}/d\eta$ distribution is less than 1%.

6.3.8 Misalignment

The geometry modeled in the simulation framework corresponds to the ideal detector configuration. In reality this *ideal geometry* is changed, e.g. by uncertainties in the positioning of the different detector elements, production tolerances, displacements caused by deformations of support structures due to the weight of components, and due to magnetic-field forces.

Survey and alignment (see Section 4.7.1) allows the estimation of displacements between the ideal geometry and the installed geometry. The reconstruction framework takes these displacements into account. Global shifts of subdetectors measured by survey are included in the ideal geometry[8]. Alignment was performed using cosmic-ray data recorded in 2008. The results from the alignment procedures are applied and produce the so-called *realigned geometry*.

[8]These are quite significant. For example the ITS is shifted by 2.9 mm in the z direction.

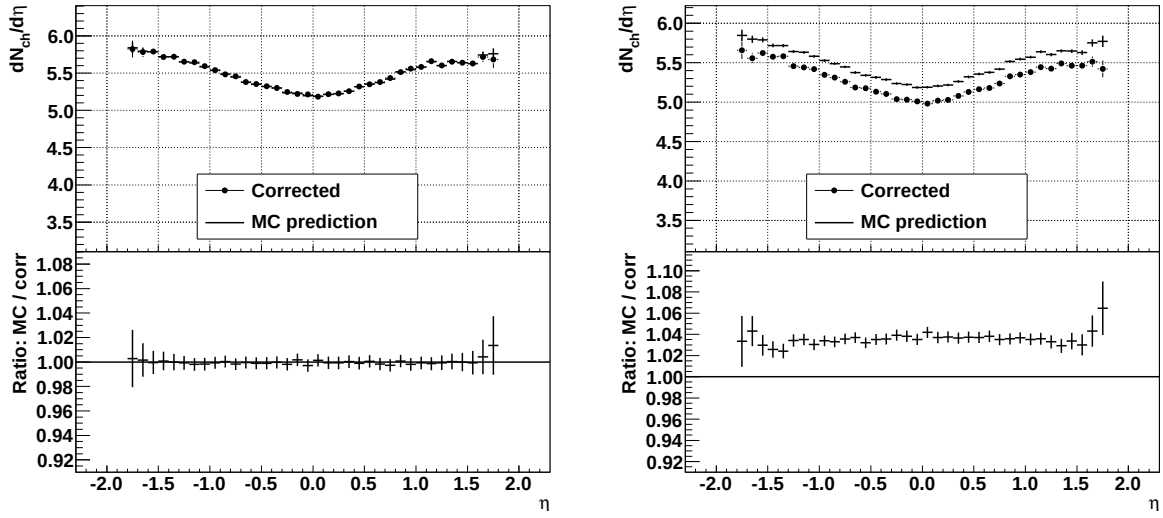

Figure 6.19: Systematic uncertainty without realignment.
An event sample was reconstructed twice, once with ideal geometry (left panel) and once with realigned geometry (right panel). Subsequently both have been corrected with corrections created from the sample produced with ideal geometry. The upper part presents the corrected result and the MC input for inelastic events. The lower part shows the ratio between the two.

SPD

The residual misalignment, the remaining uncertainty after alignment procedures, is at a level that has been shown to have no effect on this analysis. The influence on the $dN_{ch}/d\eta$ distribution is at the level of 10^{-3}, i.e. much smaller than other uncertainties.

However, a question that always remains is to which extent the alignment procedures have produced the correct results. Therefore, especially for first data, it is interesting to study the effect of the change in the geometry due to the realignment on the final analysis result. Detector simulations have been performed that use a different alignment in the simulation and reconstruction step. In the simulation step the ideal geometry was used. The reconstruction was performed twice, once with the same ideal geometry and once with the realigned geometry. In principle, the different geometries should be applied in the simulation step, but the changed geometry may produce the above-mentioned overlapping volumes that are problematic for the transport code. Furthermore, simulation takes much longer than reconstruction, thus this approach needs less computing time. The case of using a combination of ideal and realigned geometry resembles the case where the data from the detector is reconstructed without applying any alignment procedures.

Both reconstruction outputs are corrected with correction factors determined using the ideal geometry. Figure 6.19 shows the corrected distribution of the ideal geometry (left panel) and the realignment geometry (right panel) for the SPD. The distribution is about 3.5% lower than the input distribution, owing to reduced tracking efficiency caused by the misalignment.

Data measured in the experiment should be reconstructed with both the ideal geometry and the realigned geometry. This allows the effect of the performed alignment to be understood. Ultimately, the result using the realigned geometry is of course more exact. The difference obtained is not a systematic uncertainty that has to be applied to the measurement. However, it should be mentioned together with the result.

TPC

The TPC has less alignable objects, but a larger number of calibration constants that need to be evaluated, like drift velocities and gain factors. Furthermore, the residual imperfections concerning the parallel alignment of the electric and magnetic field causes a shift of the drift electrons ($E \times B$-effect). The TPC has been stably operated over long periods of time while taking cosmic-ray data and has shown already remarkable performance with respect to momentum and dE/dx resolutions, see Section 4.7.1. For this analysis it is sufficient that a track is reconstructed within the loose resolution requirements mentioned previously. It can be shown with the already recorded cosmic-ray data that this is the case. No significant effect is therefore expected on the $dN_{ch}/d\eta$ distribution due to the residual misalignment.

6.3.9 Tracklet and Track Selection

SPD

The selection of tracklets, described in Section 5.4.1, depends on the parameter $\Delta\varphi$. Uncertainties in $\Delta\varphi$ may arise from the residual misalignment. The spatial resolution of clusters is estimated from cosmic-ray data to be about 14 μm in $r\varphi$-direction, see Section 4.7.1. In fact the additional uncertainty due to the residual misalignment is expected to be only 8 μm. However, as an upper limit for the spatial resolution 14 μm is used; this is translated into an uncertainty in $\Delta\varphi$ of 0.36 mrad and 0.18 mrad for the first and second layer, respectively. Taking the sum of these two values provides a conservative estimate for the change in the number of accepted particles when moving the cut of $|\Delta\varphi| < 50$ mrad by ± 0.54 mrad. It is evaluated to be less than 10^{-3}. In fact a change of 1% in the number of accepted tracklets only occurs when the spatial

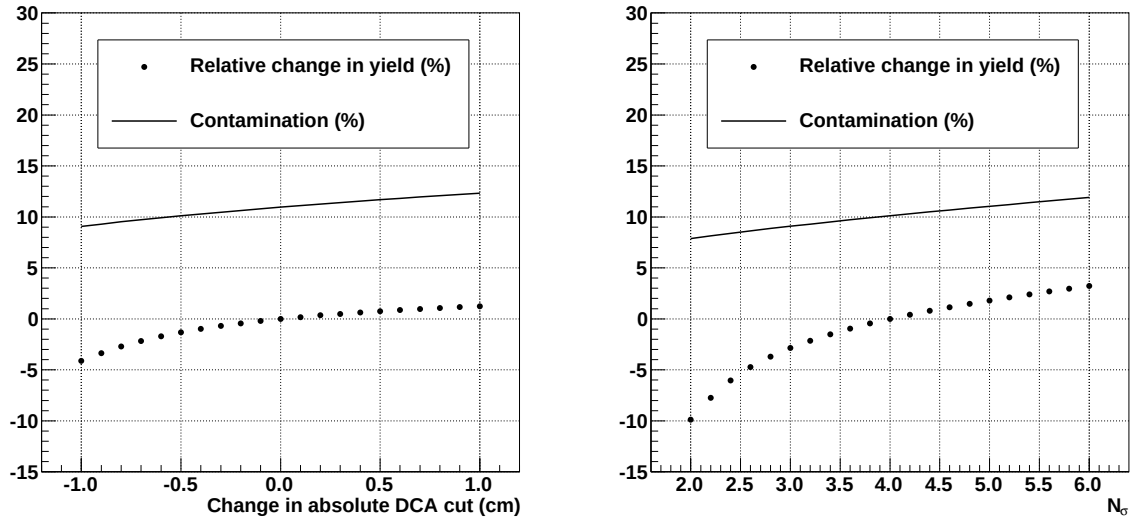

Figure 6.20: Uncertainty due to DCA cuts.
Shown is the relative yield of accepted tracks and contamination by secondaries when the cuts are varied from their default values for the absolute DCA cut (left panel) and the normalized DCA cut (right panel). Only tracks in $|\eta| < 1$ and with $p_T > 0.2\,\mathrm{GeV}/c$ are considered.

resolution is as poor as about 180 μm. The systematic effect on the $dN_{ch}/d\eta$ distribution from the tracklet selection is therefore negligible.

TPC

It was shown that most tracks are removed by the cut requiring the track to originate from the vertex of the interaction. Two possibilities were introduced, the absolute DCA cut and the normalized DCA cut. Although the first was mainly used, in this section it will be shown how the sensitivity of the measurement on the actual value used in both these cuts evolves.

In Figure 6.20 the change in the number of accepted tracks and the contamination from secondaries is shown as a function of the cut value. For the absolute DCA cut (left panel) the change is relative to the default cut values given in Section 5.4.2, $d_r = 2.4\,\mathrm{cm}$ and $d_z = 3.2\,\mathrm{cm}$. d_r and d_z are changed by the same value. In both cases, lowering the cut values reduces the contamination but increases the slope in the relative yield and therefore increases the effect of an uncertainty on the cut value.

The actual uncertainty on these values is difficult to estimate without real data. A scale for the uncertainty of the absolute DCA cut is the vertex resolution $\sigma \approx 0.25\,\mathrm{cm}$,

Uncertainty	SPD	TPC
Relative cross-sections (INEL)	2%	2%
Relative cross-sections (NSD)	8%	8%
Particle composition	negl.	1.5%
p_T spectrum	0.5%	3%
Event-generator assumptions (INEL)	1%	4%
Event-generator assumptions (NSD)	3%	5%
Beam-gas events (startup)*	negl.	negl.
Beam-gas events (nominal)*	0.6%	0.6%
Pile-up events	$< 0.1\%$	$< 0.2\%$
Material budget	negl.	$< 1\%$
Misalignment*	negl.	negl.
Track / tracklet selection cuts*	negl.	$< 1\%$
Corrections (stat. uncertainty)	1%	1%
Total (INEL)‡	2.3%	4.3%
Total (NSD)‡	8.1%	8.8%

Table 6.5: Summary of the various systematic uncertainties.

The uncertainties marked with an asterisk (*) can be better estimated with measured data.

‡ The sum in quadrature does not include the uncertainty due to the event-generator assumptions because it is mostly included in the cross-section, the particle-composition, and the p_T spectrum uncertainties (see Section 6.3.4). The beam-gas uncertainty for the startup has been used for the total.

which results in a change in accepted tracks of about 0.5%. For the normalized DCA cut no clear estimate can be done. An uncertainty of 1σ results in a change of the measured yield of $2-3\%$. However, the value of 1σ is completely arbitrary as it is not yet known how well the resolutions will be estimated with real data. Therefore, this cut is not in use for first data. As previously mentioned, in practice, several combinations of cut values need to be evaluated and the invariance of the analysis result needs to be verified.

6.3.10 Summary of the Systematic Uncertainties

The estimated systematic uncertainties are summarized in Table 6.5. The procedure to estimate the resulting total systematic error is not straightforward because of correlations among the different contributions. Certain effects cancel one another, e.g. lowering

the p_T cut-off increases the contamination by secondaries but decreases the uncertainty due to the p_T cut-off correction. The event-generator assumption uncertainty is mostly included in the uncertainty on the relative cross-sections, the particle composition, and the p_T spectrum. Other correlations are assumed to be small. Summing the uncertainties for the startup scenario in quadrature (without the uncertainty originating from differences in the assumptions of the event generators) yields a total systematic error of 2.3% for inelastic and 8.1% for NSD events for the SPD measurement. The result for the TPC measurement is 4.3% (inelastic) and 8.8% (NSD). In general, for NSD events the largest contribution is the uncertainty on the relative cross-sections. Therefore, the total uncertainty will significantly reduce once the cross-sections are constrained or measured at the LHC.

6.4 Towards the Corrected $dN_{ch}/d\eta$ Distribution

This section describes the steps required to obtain the corrected $dN_{ch}/d\eta$ distribution from the measured data. Apart from applying the corrections that have been outlined in this chapter, various verifications and checks have to be made using the measured data.

6.4.1 Event and Track Quality

Events taken when single bunches pass the detector should be used to assess the amount of beam-gas and beam-halo collisions. The rate of such events compared to events taken upon bunch crossings should correspond to the expected rate. This depends of course on the LHC running conditions (luminosity and beam intensity) and the quality of the vacuum in the beam pipe. The luminosity is most likely not known precisely at the beginning of data-taking. When the measured beam-gas and beam-halo trigger rates are too large this would be an indication for a different trigger sensitivity to such events.

The following quantities that judge the event and track quality should be compared between measured and simulated data. They are expected to be consistent with each other:

- track quality parameter distributions, in particular the values that are used in the cuts; among these the DCA distribution is of special importance because the DCA cut has the largest influence;

- the resolution of the vertex reconstruction and, in addition, the real vtx-z distribution can be compared to the simulation, however, due to the fact that the corrections are determined as function of the vertex position, the results should not depend on the assumed vertex distribution;
- the ratio of triggered events with and without reconstructed vertex; here, deviations between the result from measured data and simulated events indicate a different vertex-reconstruction efficiency;
- the p_T spectrum; this is relevant for the measurement with the SPD, because the corrections are integrated over p_T, however, it is sufficient if the p_T spectra are roughly similar; for the TPC, the corrections are determined as a function of p_T and are therefore less sensitive to an incorrect p_T spectrum in the simulation;
- properties of triggered events without reconstructed vertex (e.g. hit distributions etc.).

In case distributions differ significantly between measured and simulated data the reasons need to be understood. It may be necessary to adapt the simulation to correctly reproduce the experimental conditions. As a consequence some of the above-mentioned corrections may need to be reevaluated.

6.4.2 Pseudorapidity Distribution

The distribution should be obtained using different sets of correction maps (e.g. Pythia and Phojet), different vtx-z ranges, different values for the p_T cut-off (only TPC), and different tracklet and track cuts. The extracted $dN_{ch}/d\eta$ distribution should be robust against these changes for all event classes, i.e. triggered events with reconstructed vertex, triggered events, inelastic and NSD events, as well as before and after the p_T cut-off correction. The results from the measurement using the SPD and the TPC should lead to the same result in the overlapping η-region.

6.5 Summary

An analysis method for the pseudorapidity density of primary charged particles $dN_{ch}/d\eta$ for p+p collisions was developed. The procedure takes into account and corrects for detector and reconstruction effects, namely: the trigger bias, the vertex-reconstruction efficiency and effects due to acceptance and tracking efficiency. The

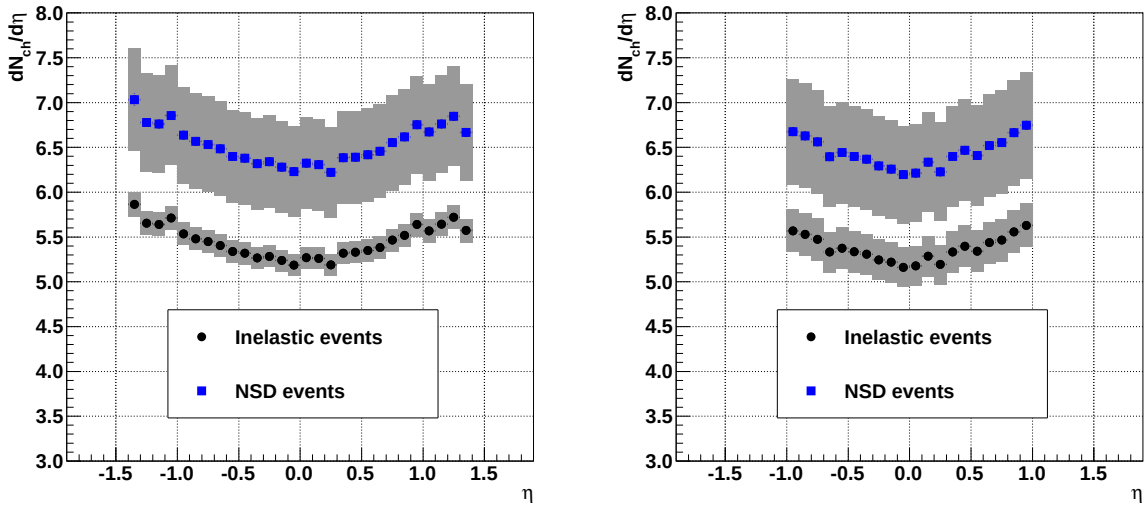

Figure 6.21: Final $dN_{ch}/d\eta$ spectrum.

The figure shows the corrected result based on 50 000 simulated events for the SPD (left panel) and the TPC (right panel). The grey bars indicate the estimated systematic uncertainty. With the given statistics for the SPD, the statistical errors on the measurement (see Figure 6.4) and on the corrections (see Section 6.2.6) are too large for a measurements outside $|\eta| \lesssim 1.4$ (at large $|\eta|$ only few events that have large |vtx-z| contribute). Therefore, a smaller η-range than in the preceding figures is shown. A larger number of events will allow to access $|\eta| \lesssim 2$ with the SPD.

method has been checked and verified extensively using simulated data at various energies. The procedure can be applied to data taken with and without magnetic field. The method has also been tested on beam-gas interactions taken during the LHC startup.

Uncertainties stemming from assumptions on the characteristics of p+p collisions and on the detector response have been evaluated. The study gives an extensive and as complete as possible description of all systematic uncertainties. Uncertainties whose determination requires measured data have also been discussed.

Figure 6.21 shows a corrected $dN_{ch}/d\eta$ spectrum with the SPD (left panel) and the TPC (right panel). Results for inelastic and NSD events are shown. For both event classes 50 000 simulated Pythia events were used. Assuming a collision rate of only 1 Hz (startup scenario with bunch-crossing trigger) 50 000 events can be collected within less than a day of data-taking time. Therefore, such a spectrum can be measured within the first few days after the start of data-taking.

Chapter 7

Multiplicity Distribution Measurement

This chapter describes the procedure to measure the charged-particle multiplicity distribution. It focuses on the measurement in the central barrel using the SPD. Furthermore, the measurement procedure using information from the TPC is outlined.

The following sections discuss in detail the procedure to measure the raw spectrum, the corrections that need to be applied, as well as the systematic uncertainties that arise during this measurement. The systematic uncertainties are given for the measurement of the SPD. It is straightforward to extract the uncertainties for the measurement with the TPC.

The outline of the chapter is the following: Section 7.1 introduces the general correction procedure. In Section 7.2 the unfolding and correction methods required in this analysis are described. The subsequent Section 7.3 discusses the details of the methods and their evaluation. The systematic uncertainties are evaluated in Section 7.4. The last section describes the steps needed to obtain the distribution from measured data.

7.1 Procedure Overview

The multiplicity distribution characterizes the multiplicity fluctuation for many events. Therefore, the full multiplicity in the considered region has to be measured for each event entering in the distribution. In contrast to the $dN_{ch}/d\eta$ analysis, where an *average* value is determined, events with different z-vertex positions (and thus with a different acceptance in η) cannot be used to extend the accessible η-range. Therefore, the desired range in η defines the usable range in vtx-z. This can be seen in Figure 6.3

	SPD			TPC
		Included events at $\sqrt{s}$ of		
η-region	vtx-z range	900 GeV	10 TeV	vtx-z range
$\lvert\eta\rvert < 0.5$	$\lvert$vtx-$z\rvert < 15$ cm	97%	100%	$\lvert$vtx-$z\rvert < 15$ cm
$\lvert\eta\rvert < 1.0$	$\lvert$vtx-$z\rvert < 6$ cm	59%	89%	$\lvert$vtx-$z\rvert < 15$ cm
$\lvert\eta\rvert < 1.4$	$\lvert$vtx-$z\rvert \lesssim 2$ cm	22%	40%	—

Table 7.1: Accessible η-regions and corresponding vtx-z ranges.

(page 118). Table 7.1 indicates the allowed vtx-z range for different η-ranges. The maximum vtx-z range given is ± 15 cm. Also given are the fractions of events included in the corresponding vtx-z ranges for $\sqrt{s} = 900$ GeV and 10 TeV for the SPD.

In principle the TPC allows a measurement in $|\eta| < 1.3$. However, for $|\eta| > 1$ the efficiency drops significantly and only particles with large p_T are measured (see Figure 6.6 on page 124).

Ideally the largest possible region is chosen for the analysis (and thus the smallest vtx-z range for the SPD). However, this reduces the number of events that can be used for the analysis. The plots in this chapter consider the case of $|\eta| < 1$.

Reconstructed tracklets (SPD) and tracks (TPC) are used for the analysis. The reconstruction procedure has been described in Section 4.6.3. Events and tracks have to fulfill certain criteria which have been explained in Sections 5.2 and 5.4. The tracklets or tracks are counted for each event that occurred in the chosen vtx-z range. This step results in a raw measured multiplicity spectrum. The correction of this spectrum is not straightforward due to the fact that events with different true multiplicities contribute to the same measured multiplicity. The measured spectrum has to be unfolded in order to obtain the true multiplicity spectrum. This problem and its solution will be discussed in detail in the following sections. The unfolded spectrum is the multiplicity distribution of primary particles for the events that have been triggered and have a reconstructed vertex.

Subsequently, this spectrum needs to be corrected for the bias introduced by the vertex reconstruction as well as the trigger. The spectrum of triggered events is obtained after correcting for the vertex-reconstruction efficiency. Finally, the trigger-bias correction results in the spectrum for inelastic or NSD events. The vertex-reconstruction bias as well as the trigger-bias correction have been described in detail in the previous chapter and are only briefly covered here. Contrary to the $dN_{ch}/d\eta$ analysis, these corrections

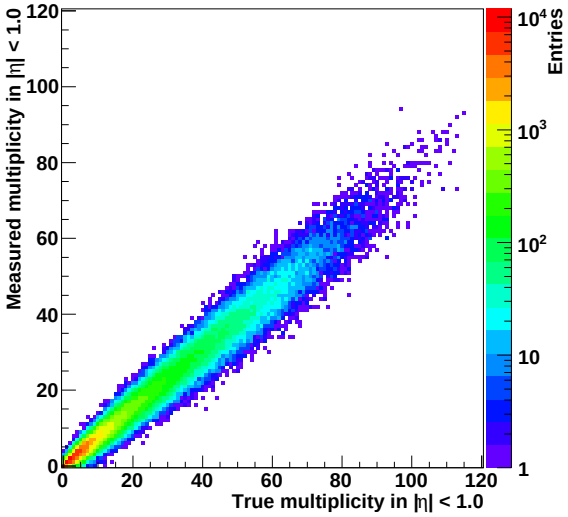

Figure 7.1: Detector response.
The figure shows the response matrix of the SPD. The number of tracklets vs. the number of generated primary particles in $|\eta| < 1$ is depicted.

are determined as a function of the true multiplicity since they are applied after the unfolding step.

Distributions are given before the unfolding step as a function of measured multiplicity (in *measured variables*). After the unfolding step they are given as a function of *unfolded* multiplicity (in *unfolded variables*) which is equivalent to the true multiplicity. Following this terminology a MC input distribution can be given in true or in unfolded variables.

To evaluate the method and to assess the systematic uncertainties the simulated data sample LHC08e1 has been used (see Section 5.5). 200 000 events are taken as the analysis input sample; the remainder is used to calculate the corrections (270 000 events).

7.2 Corrections

7.2.1 Detector Response

The response of the detector can be described by a matrix R. The matrix element R_{mt} gives the conditional probability that a collision with a true multiplicity t is measured as an event with the multiplicity m. The response matrix is created using the detailed detector simulation for a certain η and vtx-z range. An example is shown for $|\eta| < 1$

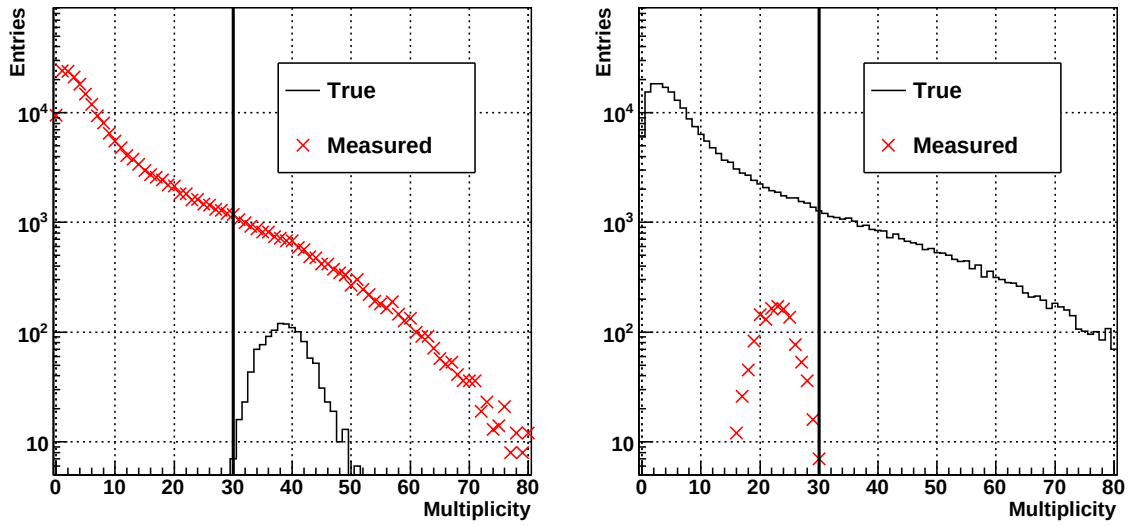

Figure 7.2: The need for unfolding.
The left panel shows the measured spectrum superimposed with the true distribution that caused the entries in one single measured bin (exemplarily at multiplicity 30 indicated by the line). Clearly the shape of this true distribution depends on the shape of the multiplicity distribution given by the model used (A suggestive example is if the true spectrum stopped at a multiplicity of 40: the true distribution that contributed to the measured multiplicity of 30 would clearly be different, still events at a multiplicity of 30 would be measured). Inversely, in the right panel, the true distribution is shown superimposed with the measured distribution caused by events with the true multiplicity 30 (exemplarily). The shape of this measured distribution depends only on the detector simulation, i.e. the transport code and reconstruction, and **not** on the multiplicity distribution given by the model (only events with multiplicity 30 contribute to the shown measured distribution).

and $|\text{vtx-}z| < 6\,\text{cm}$ in Figure 7.1. The average measured multiplicity is about 0.75 times the true multiplicity; this is due to the detector efficiency that is mainly affected by the inactive modules in the SPD (see Section 5.5). Furthermore, the limited resolution can be clearly seen: events from a given true multiplicity are spread over several measured multiplicity bins. The statistics becomes poor around a true multiplicity of 80 (less than 5 entries per bin), thus in the studied example meaningful results are only expected for multiplicities up to 80.

Given a true spectrum T, the measured spectrum M can be calculated by:

$$M = RT. \tag{7.1}$$

The aim of the analysis is to infer T from M. Simple weighting, i.e. assuming that a measured multiplicity m is caused 'mostly' by a true multiplicity t, would not be correct. Analogous, adding for each measured multiplicity the corresponding row of the response matrix to the true distribution is incorrect. This is model-dependent and thus in principle not possible. On the other hand the measured spectrum which is the result of a given true multiplicity is only determined by the detector simulation and is model-independent. This is illustrated in Figure 7.2.

Given a measured spectrum, the true spectrum is formally calculated as follows:

$$T = R^{-1}M. \tag{7.2}$$

R^{-1} cannot be calculated in all cases, because R may be singular; e.g. if two true multiplicities result with equal probabilities in two measured multiplicities. This can in principle be solved by choosing a more appropriate binning. But even if R can be inverted, the result obtained by Eq. (7.2) contains usually severe oscillations (due to statistical fluctuation caused by the limited statistics of events used to create the response matrix). This can be illustrated with the following example [Blo84]: a square response matrix is assumed to describe the detector:

$$R = \begin{pmatrix} 0.75 & 0.25 & 0 & & \cdots \\ 0.25 & 0.50 & 0.25 & 0 & \\ 0 & 0.25 & 0.50 & 0.25 & \\ & 0 & 0.25 & 0.50 & \\ \vdots & & & & \ddots \end{pmatrix}. \tag{7.3}$$

A true distribution T is assumed, and the expected measured distribution M is calculated with Eq. (7.1). The distribution M is used to generate a sample of 10 000 measurements: $\tilde{M}$. Using Eq. (7.2) the corresponding true distribution $\tilde{T}$ is calculated. Figure 7.3 shows these four distributions. Although the resolution effect on the shape of the measured distribution (left histogram) is very small, the unfolded solution (right histogram) suffers from large non-physical fluctuations. Clearly, this is not the spectrum that corresponds to the true one.

The information that is lost due to the resolution cannot be recovered in principle. However, constraining the result with *a priori* knowledge about the smoothness of the function allows the recovery of the true distribution. This is discussed in detail in the following sections, which present two unfolding methods to tackle this problem. The first method leads to the true spectrum by minimizing a χ^2-function; the second is an iterative method based on Bayes' theorem.

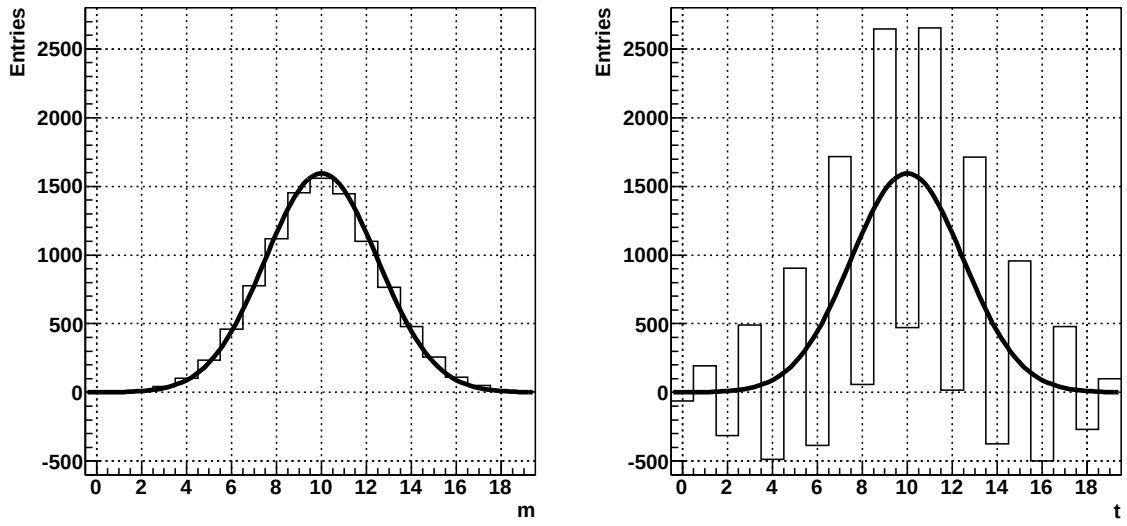

Figure 7.3: Illustration of the problem with simple matrix inversion.

The left panel shows a sample of the measured distribution $\tilde{M}$ with 10 000 entries (histogram). Using Eq. (7.2) the corresponding true distribution $\tilde{T}$ is calculated, which is shown in the right panel (histogram). The overlaid function is the true shape T. Although the resolution effect on the shape of the measured distribution (left) is very small, the solution obtained by matrix inversion suffers from large fluctuations.

7.2.2 Unfolding by χ^2-Minimization

An approach to unfold the measured multiplicity distribution is the minimization of a χ^2-function. Using the response matrix, this function gives a measure of how well an estimated unfolded spectrum describes the measured spectrum. A minimization program is used to find the unfolded spectrum that minimizes the χ^2-function. With e denoting the error on the measurement M, and U the guessed spectrum, a suitable χ^2-function is:

$$\hat{\chi}^2(U) = \sum_m \left(\frac{M_m - \sum_t R_{mt} U_t}{e_m} \right)^2 . \tag{7.4}$$

Eq. (7.4) with Eq. (7.1) results in $\hat{\chi}^2(T) = 0$, as required.

This method is a numerical approach to Eq. (7.2) using the inverse of the response matrix. Therefore, it is not surprising that the previously mentioned fluctuations exist also in the solution found by χ^2-minimization. In fact, the number of events is always finite and thus no solution U satisfies $\hat{\chi}^2(U) = 0$ exactly. As a consequence not only the true spectrum T minimizes this function. Many other, mostly fluctuating, solutions exist and it is not straightforward to find the 'correct' spectrum. An example of a

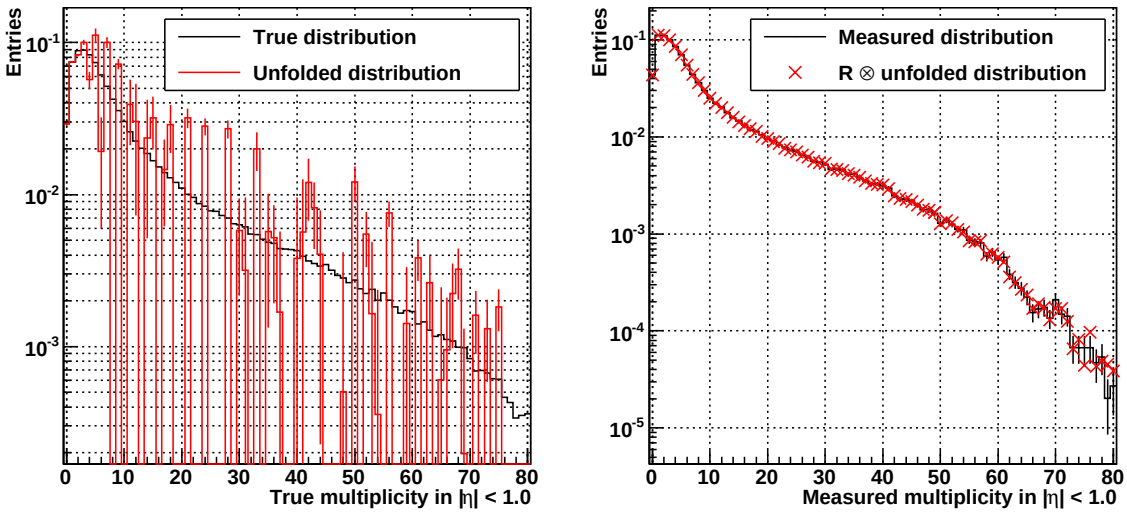

Figure 7.4: Result of an unfolding minimizing Eq. (7.4).

The left panel shows the unfolded spectrum (red) and the true spectrum (black), the right panel shows the measured spectrum (black histogram) and the response matrix multiplied with the unfolded spectrum (red crosses). The latter corresponds to the term $\sum_t R_{mt} U_t$ in Eq. (7.4). No regularization is used which results in strong fluctuations in the unfolded spectrum. The unfolded spectrum, which is clearly not the correct solution, still minimizes the χ^2-function as required.

fluctuating solution, that indeed minimizes the χ^2-function can be seen in Figure 7.4. The fact that causes such solutions to appear valid is that the bin size used in the response matrix is smaller than the resolution of the detector: events with a given true multiplicity t are spread (or smeared) over a range of multiplicities in the observed distribution (see Figure 7.2). A solution is to add a constraint to the χ^2-function that favors a certain shape of the unfolded spectrum:

$$\chi^2(U) = \hat{\chi}^2(U) + \beta P(U). \tag{7.5}$$

$P(U)$ is called regularization term. It depends only on the unfolded spectrum U (and not on R and M). β determines the weight that is given to the regularization with respect to the first term that governs the agreement with the measurement. A higher β will lead to an increased $\hat{\chi}^2$. Its optimal value needs to be evaluated, but generally it can already be remarked that a reasonable value of β adjusts the two terms in Eq. (7.5) in such a way that the introduced bias is negligible compared to the statistical error of the measurement. $\beta P(U)$ is also called *penalty term*. There are means to verify that the influence of the regularization term is not dominant, which will be outlined in Section 7.3. Many possibilities exist for the choice of the regularization: these range

from just requiring a smooth function to preferring a certain shape of the distribution. Generally, no specific functions should be used, otherwise the result is likely to look very similar to what has been required.

A set of different regularizations has been applied and evaluated:

$$P(U) = \sum_t \left(\frac{U'_t}{U_t}\right)^2 = \sum_t \left(\frac{U_t - U_{t-1}}{U_t}\right)^2, \tag{7.6}$$

$$P(U) = \sum_t \left(\frac{U''_t}{U_t}\right)^2 = \sum_t \left(\frac{U_{t-1} - 2U_t + U_{t+1}}{U_t}\right)^2, \tag{7.7}$$

$$\begin{aligned} P(U) &= P(\hat{U} := \ln U) = \sum_t \left(\frac{\hat{U}''_t}{\hat{U}_t}\right)^2 \\ &= \sum_t \left(\frac{\ln U_{t-1} - 2\ln U_t + \ln U_{t+1}}{\ln U_t}\right)^2, \end{aligned} \tag{7.8}$$

$$P(U) = P(\hat{U} := \frac{U}{\sum_t U_t}) = \sum_t \hat{U}_t \ \ln \frac{\hat{U}_t}{\epsilon_t}. \tag{7.9}$$

The use of Eq. (7.6) favors a constant function. A linear function is preferred by applying Eq. (7.7), which is also called *least curvature*. Eq. (7.8) is motivated by the fact that the multiplicity distribution has an approximately exponential shape at higher multiplicities. The denominator is introduced in Eqs. (7.6 – 7.8) to ensure equal weights along the steeply falling spectrum. Eq. (7.9) is the method of reduced cross-entropy [Sch94] allowing the consideration of an *a priori* distribution ϵ that describes the shape of the spectrum. This method was used e.g. by ALEPH to unfold multiplicity distributions in restricted rapidity intervals [Bus95]. If $\epsilon \equiv 1$ is used, Eq. (7.9) becomes the method of reduced entropy. In Section 7.3.2 the regularizations (7.6 – 7.9) and the influence of the weight parameter β are evaluated.

An unfolded distribution using the MINUIT [Jam75] minimization program is shown in Figure 7.5. The method reproduces the multiplicity distribution for the event class of triggered events that have a reconstructed vertex. To obtain the distribution for the triggered sample, inelastic or NSD events, further corrections, given by the inverse efficiency of the vertex reconstruction (or the trigger) as a function of the true multiplicity, have to be applied to the unfolded spectrum. This is discussed in Section 7.2.4.

Fit with Predefined Functions

The number of free parameters in a χ^2-minimization can be significantly reduced by using a parametrization. It can be chosen following a model prediction or previous

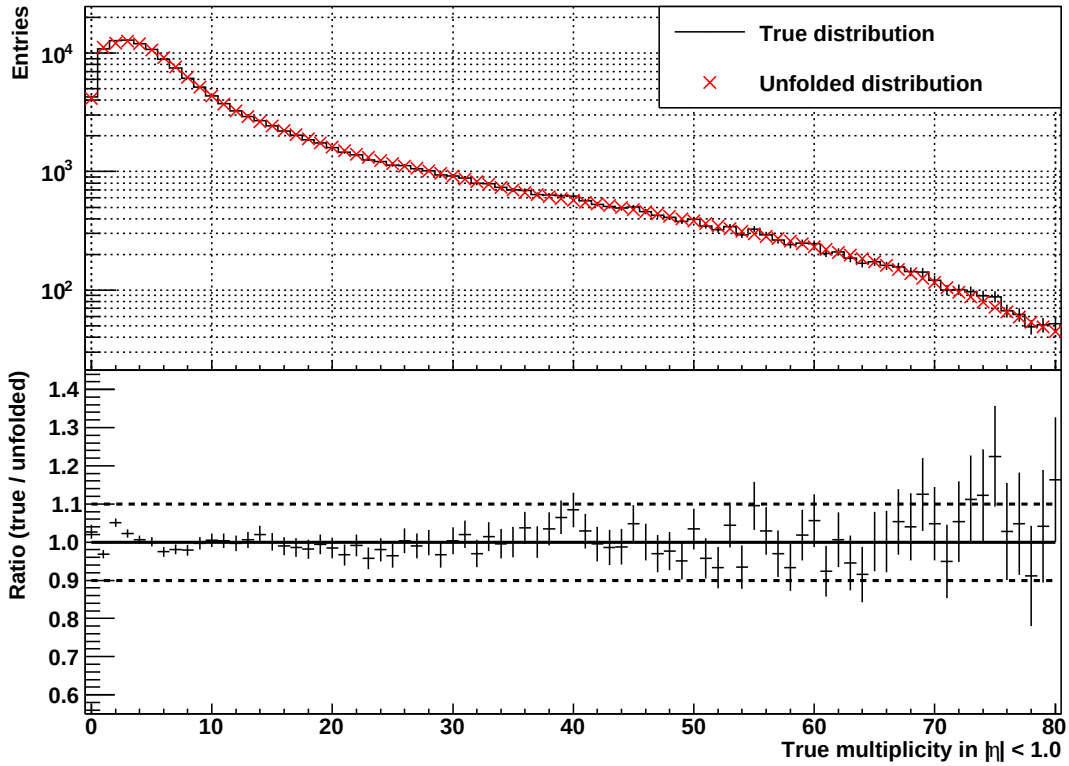

Figure 7.5: Unfolded distribution using χ^2-minimization with regularization. The top panel shows the unfolded distribution (red crosses) superimposed with the true distribution (black histogram). In the bottom panel the ratio between the two is shown. The band indicated by the dashed lines shows ±10% deviation. The oscillations that can be seen will be discussed in Section 7.4.2. Eq. (7.5) with regularization (7.6) and $\beta = 10^5$ was used (see text).

measurements. In any case significant *a priori* knowledge is assumed. Thus such an approach can only be used to *verify* a model prediction, *not to infer* potentially new properties of collisions at a new energy.

7.2.3 Bayesian Unfolding

An alternative approach to unfold the measured distribution is based on Bayes' theorem. The probability of an event A conditional on another event B is generally different

from the probability of B conditional on A. Bayes' theorem describes the definite relationship between these two conditional probabilities:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}. \tag{7.10}$$

$P(A)$ and $P(B)$ are the prior probabilities for the event A and B, respectively. $P(B|A)$ is the probability of event B under the condition that A is true. Given these three quantities, $P(A|B)$ (the probability of event A under the condition that B is true) can be inferred.

A is identified with a collision in the detector with a given true multiplicity and B with a measured event with a given measured multiplicity. The conditional probability $P(A|B)$ (also known as *smearing matrix* in the literature) can then be determined using the knowledge of the detector response matrix $P(B|A)$. However, both prior probabilities $P(A)$ and $P(B)$ need to be known in order to proceed. $P(B)$ is the measured spectrum and $P(A)$ is the true distribution which is the distribution that is to be obtained. This contradiction is solved by using an iterative method proposed in [Ago95, Ago99], which is outlined in the following.

Using the nomenclature of the previous section and using Eq. (7.1), Bayes' theorem is expressed as[1]:

$$\tilde{R}_{tm} = \frac{R_{mt} \cdot P_t}{\sum\limits_{t'} R_{mt'} P_{t'}}. \tag{7.11}$$

P_t is the *a priori* distribution of the true spectrum. In the case of complete ignorance it can be set to a flat distribution. In the present analysis the measured spectrum has been used as an *a priori* distribution. Other choices are discussed in Section 7.3.5. Having obtained $\tilde{R}_{tm}$, the measured spectrum allows the determination of the (not yet normalized) unfolded spectrum U_t:

$$U_t = \sum_m \tilde{R}_{tm} M_m. \tag{7.12}$$

U_t is equal to P_t, if P_t is the true spectrum; otherwise it is between P_t and the true spectrum [Ago95]. For the next iteration, U_t is used as the new *a priori* probability P_t. Optionally a smoothing can be applied at this stage reducing the influence of high-frequency fluctuations:

$$\hat{U}_t = (1 - \alpha) \cdot U_t + \alpha \cdot \frac{1}{3}(U_{t-1} + U_t + U_{t+1}). \tag{7.13}$$

α defines the weight of the smoothing ($\alpha = 0$ results in $\hat{U}_t \equiv U_t$) and its optimal value needs to be evaluated; this is discussed in Section 7.3.3. $\hat{U}_t$ is then used as the new

[1]Note, that Eq. (7.11) calculates the matrix element tm of $\tilde{R}_{tm}$. The right-hand side does not contain any sums except the explicitly mentioned one.

a priori probability P_t. Note that the smoothing is only applied to the distribution used as the a-priory probability for the next iteration. It is not applied to the (final) unfolded distribution. These formulas yield the multiplicity distribution without efficiency correction, which is the distribution for the event sample of triggered events with a reconstructed vertex.

Bayesian unfolding can also consider an efficiency ϵ_t which is the detector efficiency to detect an event with a given true multiplicity. In this analysis ϵ_t is the vertex-reconstruction and trigger efficiency. If the true spectrum for the triggered event sample was to be calculated, ϵ_t would contain only the vertex-reconstruction efficiency. In the case of the calculation for the inelastic and NSD event samples, ϵ_t contains also the respective trigger efficiency. The introduction of the efficiency, requires also that the efficiency is taken into account in the response matrix. A given event with a true multiplicity may not contribute to the response matrix, i.e. $\sum_m R^*_{tm} = \epsilon_t$.

With

$$U^*_t = \frac{U_t}{\epsilon_t} \qquad P^*_t = \frac{P_t}{\epsilon_t} \qquad R^*_{mt} = \epsilon_t R_{mt} \tag{7.14}$$

Eqs. (7.11) and (7.12) can be rewritten as:

$$\tilde{R}_{tm} = \frac{R^*_{mt} \cdot P^*_t}{\sum\limits_{t'} R^*_{mt'} P^*_{t'}} \qquad U^*_t = \frac{1}{\epsilon_t} \sum_m \tilde{R}_{tm} M_m. \tag{7.15}$$

These formulas produce the multiplicity distribution with efficiency correction. U^*_t and P^*_t are the unfolded distributions that have been efficiency-corrected. Note that $\tilde{R}_{tm}$ in Eq. (7.15) is identical to the one in Eq. (7.11). Thus the iteration steps are identical, i.e. the speed of convergence is identical. This has the advantage that it is sufficient to evaluate a single case. It is not relevant if the efficiency is already applied during the iteration procedure or afterwards. The only difference is that the smoothing is applied to U_t. In the other case it is applied to U^*_t, which can be shown to be not relevant for the evaluation.

The number of iterations is a free parameter. It can be fixed *a priori* or the method is terminated by convergence of the unfolded distribution. For this purpose a χ^2-test is performed between U_t and P_t: if the result drops below a threshold (in use here: 10^{-6} times the number of bins) the iteration procedure is stopped. However, limiting the number of iterations provides an implicit regularization [Blo02b]. Therefore, it is interesting to study the effect of using different numbers of iterations even though the convergence defines the number of iterations. This is performed in Section 7.3.3. An unfolded spectrum compared to the true distribution (MC input) can be seen in Figure 7.6.

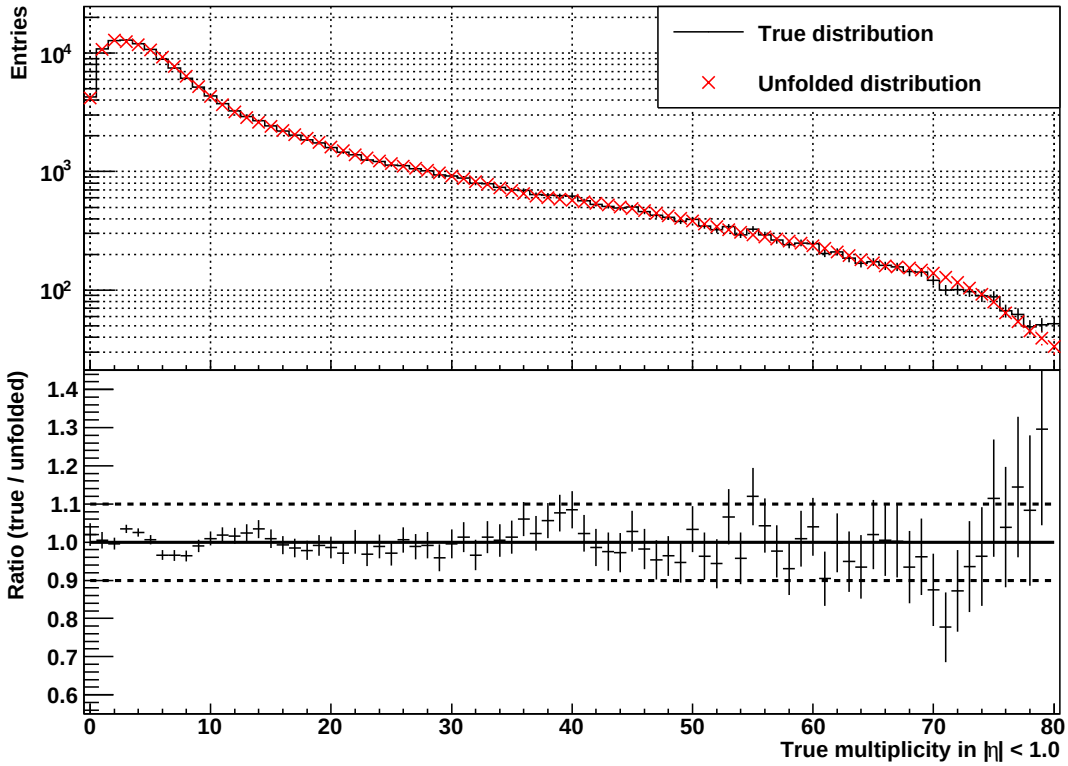

Figure 7.6: Example for Bayesian unfolding.
In the top panel the unfolded distribution (red crosses) is shown, superimposed with the true distribution (black histogram). The bottom panel shows their ratio. The band indicated by the dashed lines shows ±10% deviation. The unfolding used 10 iterations with $\alpha = 1$.

The calculation of the uncertainty on the result, i.e. the covariance matrix, is described in [Ago95]. Unfortunately it is an $\mathcal{O}(N^7)$ problem, where N is the number of bins in the unfolded spectrum. The calculation cannot be performed in practice, even on fast computers. Thus the uncertainty is calculated in a different way by randomizing the measured sample. This is discussed in Section 7.4.2.

In the case of convergence, χ^2-minimization and Bayesian unfolding yield the same result. With T_t being the exact solution of the χ^2-minimization, i.e. $\hat{\chi}^2(T_t) = 0$, and Eqs. (7.1), (7.11), and (7.12) ($P_t := T_t$), it can be obtained:

$$U_t = \sum_m M_m \frac{R_{mt} \cdot T_t}{\sum_{t'} R_{mt'} T_{t'}} = \sum_m R_{mt} \cdot T_t = T_t. \tag{7.16}$$

$U_t = T_t$ is nothing more than the convergence criterion of the Bayesian unfolding.

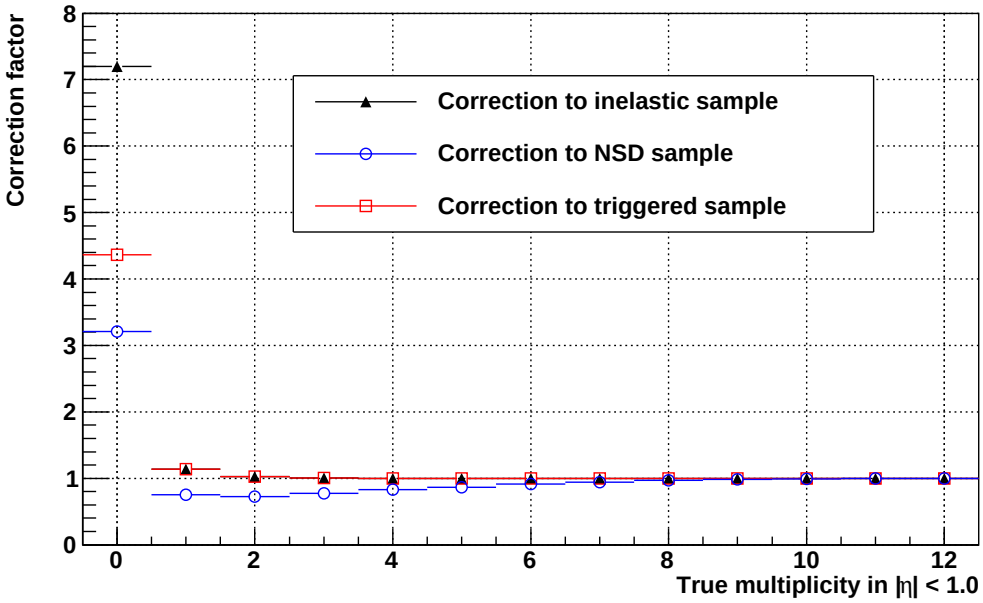

Figure 7.7: Trigger-bias and vertex-reconstruction corrections.
Applying only the vertex-reconstruction efficiency correction (red squares) leads to the result for the triggered sample. The combination of both corrections is shown correcting to the inelastic event sample (black triangles) and to the NSD event sample (blue circles). The values for the latter correction are below 1 because the SD events have to be subtracted. The values in the 0-bin are large which results in a large associated error. The efficiency, i.e. the inverse of the correction factor, for different process types is shown in Figure 7.19 (page 182).

7.2.4 Trigger-Bias and Vertex-Reconstruction Correction

Depending on the desired event class, a different response matrix has to be used in the unfolding methods. Here a response matrix is used for triggered events that have a reconstructed vertex. The correction for vertex-reconstruction efficiency and trigger efficiency is done in a subsequent step. Correcting only for the vertex-reconstruction efficiency obtains the spectrum for the triggered sample. Depending on the trigger efficiency correction, the result is the spectrum for inelastic or NSD events.

Figure 7.7 shows the correction for the vertex-reconstruction efficiency alone and its combination with the correction for the trigger efficiency. The corrections to the triggered and inelastic event sample differ from unity only in the low-multiplicity region below 4. The correction to the NSD event sample subtracts the SD events, therefore its values are below 1. The correction in the bin with zero multiplicity (called *0-bin* in the following) is very large, which raises the question of the meaningfulness of the result in

this bin. The correction factors cannot be directly compared to those presented in the previous chapter because here they are expressed as a function of the true multiplicity, not as a function of the measured multiplicity.

7.3 Evaluation of the Unfolding Methods

It has to be shown that each methods' unfolded spectrum reproduces the true distribution. The multiplicity distribution at large multiplicities is of exponential-type that ranges over several orders of magnitude. The distribution has to be reproduced well in all areas including where the slope changes and in the region of limited statistics. For this purpose simulated events are used, where the input distribution is known. A different set of events is used than the one used to generate the response matrix to allow for statistical independence of the samples. However, even these events follow the same MC distribution. Hence they do not demonstrate that the method also works when the true distribution has a different shape. To show the latter, a different MC generator as well as arbitrary input distributions have been used. One obtains the 'measured distribution' (that would be caused by an arbitrary input distribution) by a random sample governed by the function that results from the multiplication of the input distribution and the response matrix, i.e. Eq. (7.1).

7.3.1 Performance Measure

As a first indication of the correctness, the unfolded distribution can be compared to the input distribution 'by eye'. However, in order to find optimal values for the free parameters, a measure of the performance for the comparison between the unfolded distribution and the input distribution needs to be defined.

The quality of a result can be described by the difference between the input and the unfolded distribution, which is a function of the multiplicity. The performance measure can be defined as

$$Q_1 = \frac{1}{t'' - t' + 1} \sum_{t=t'}^{t''} \left| \frac{T_t - U_t}{e_t} \right| \tag{7.17}$$

with the unfolded distribution U and the input distribution T (with error e). The sum is not in quadrature to not overvalue the effect of deviations in single bins. Q_1 is normalized to allow the comparison of the performance in two regions with a different number of bins. In other words Q_1 is the average of the absolute residuals of the two distributions within a certain multiplicity region.

Region	$t' - t''$
Peak	1 – 10
Constant slope	20 – 65
Low statistics	70 – 80

Table 7.2: Performance measurement regions.

The performance of the unfolding methods in different regions of the distribution depends on its shape. This is mainly due to the regularization which prefers a certain type of function and the available number of events. Three regions of interest are chosen which are characteristic of the shape of multiplicity distributions.

- *Peak*: a region where the slope of the function changes rapidly.
- *Constant slope*: a region where the (exponential) slope does not change significantly.
- *Low statistics*: a region where the number of events is low (in the true distribution around 50 – 100 events per bin).

The positions of these regions depend on the number of events used and the η-range that is considered. For the statistics used in this study (200 000 events), the values in $|\eta| < 1$ (see Figure 7.5) are given in Table 7.2. In the following, results are obtained for each of the regions and overall optimal values are chosen for the free parameters. It has been verified that the optimal parameters depend only weakly on the considered η-region. The qualitative behavior and the optimal choice are similar if the SPD or the TPC is used.

Furthermore, it is interesting to evaluate the residuals between the measured distribution and the unfolded distribution convoluted with the response matrix. These are defined by:

$$\frac{M_m - \sum_t R_{mt} U_t}{e_m} \tag{7.18}$$

for the measured distribution M with error e and unfolded distribution U. If the only reason for the residuals to be non-zero is the statistical uncertainty, their distribution should be a Gaussian function with a width of 1. Figure 7.8 shows the two distributions and the residuals exemplarily for Bayesian unfolding. Their squared sum corresponds

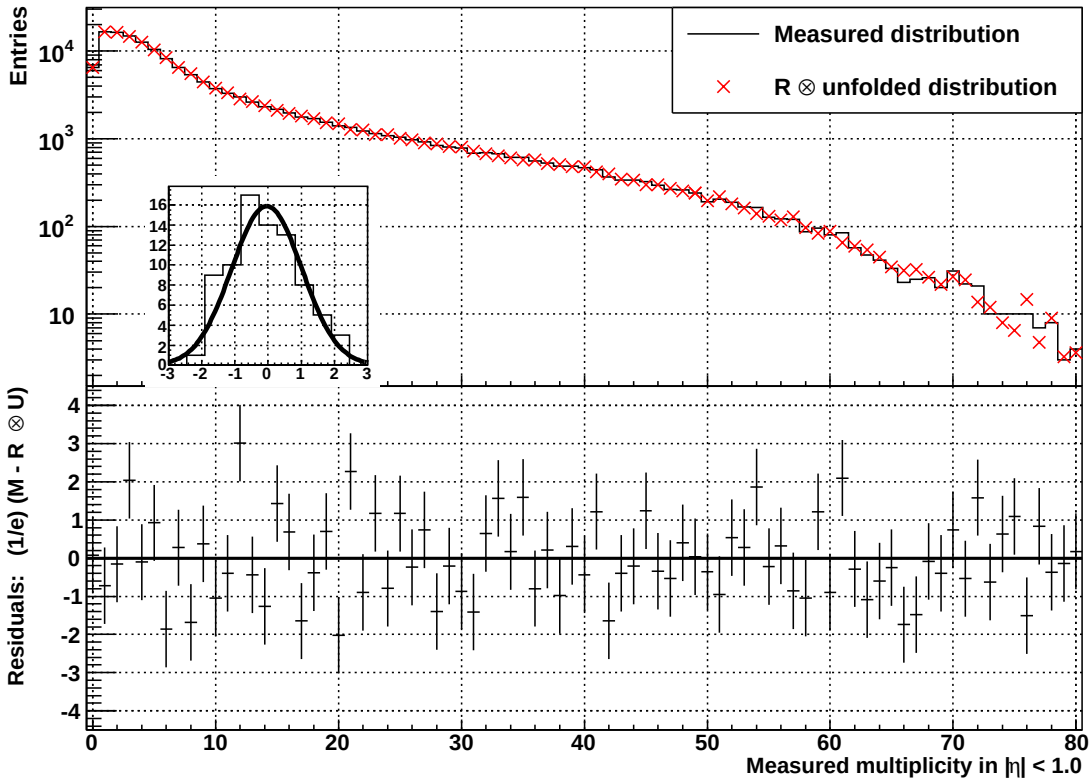

Figure 7.8: Residuals of the unfolding procedure.

The upper plot shows the measured spectrum (black histogram). The unfolded distribution, which has been convoluted with the response matrix (red crosses), is superimposed. The errors are suppressed for visibility. The bottom plot shows the residuals defined by Eq. (7.18) with error bars of 1. The small insert shows the distribution of the residuals fitted with a Gaussian ($\sigma \approx 1.05$).

to the part of the χ^2-function that describes the difference to the measured spectrum (identical to Eq. (7.4)) and defines a second performance criterion:

$$Q_2 = \sum_m \left(\frac{M_m - \sum_t R_{mt} U_t}{e_m} \right)^2 . \tag{7.19}$$

In the χ^2-minimization this term describes the influence of the regularization term.[2] If Q_2 is of the order of the number of degrees of freedom (the number of bins in the measured spectrum), the influence of the regularization does not exceed the statistical uncertainty (see also [Blo84]). The residuals can also be calculated with real data and give an indication of whether the unfolding procedure was successful.

[2]Without regularization Q_2 is identical to 0 (neglecting statistical effects).

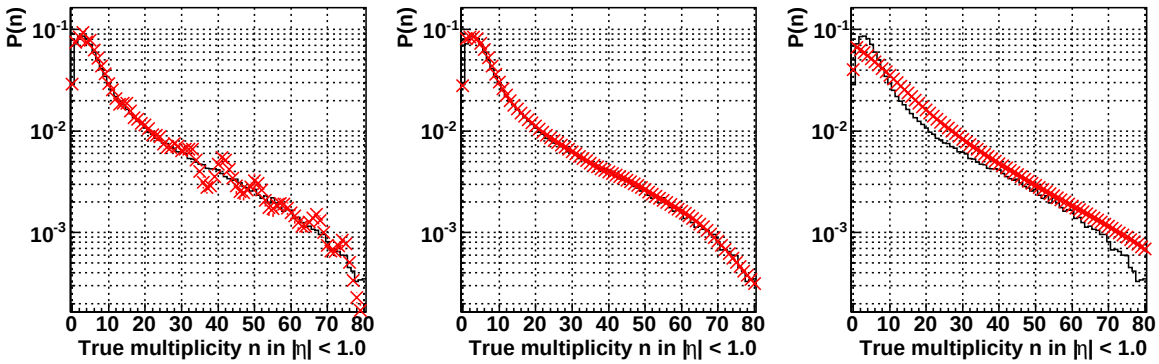

Figure 7.9: Illustration of the influence of the regularization.
Three unfolded distributions are shown which have been unfolded using a different weight parameter β for the regularization with Eq. (7.7). A small value of $\beta = 10$ (left panel), a medium value of $\beta = 10^4$ (center panel) and a large value of $\beta = 10^7$ (right panel). In each panel the input distribution (black histogram) is superimposed with the result obtained by unfolding (red crosses).

7.3.2 χ^2-Minimization

Several regularizations (7.6 – 7.9) are available for χ^2-minimization whose influence can be adjusted with the weight parameter β. Figure 7.9 illustrates the influence of the regularization. Three unfolded distributions which have been unfolded with a different weight parameter β are shown. The value used in the left panel is too low, hence the unfolded distribution fluctuates. The result in the right panel uses a too high β, thus the result does not reproduce the input spectrum. The result in the center panel uses an optimal value of β.

The χ^2-minimization was performed with each regularization and a broad range of values for β. Figure 7.10 summarizes the results by plotting Q_1 in the three regions and Q_2. Regularization (7.9) is not shown because its results have been unsatisfactory, neither without using an *a priori* distribution ϵ, nor with the measured distribution as ϵ. Other assumptions for ϵ would imply that an unjustified constraint is imposed on the unfolded distribution. Qualitatively the following can be observed:

- **Peak (top left panel, region 1)**: this region has the largest Q_1 (best values: 1 – 2) due to the higher influence of the regularization when the slope changes. Compared to the other regions, the different regularizations have their minima at significant different β. Regularization (7.8) (log) does not achieve a pronounced minimum in the considered region.

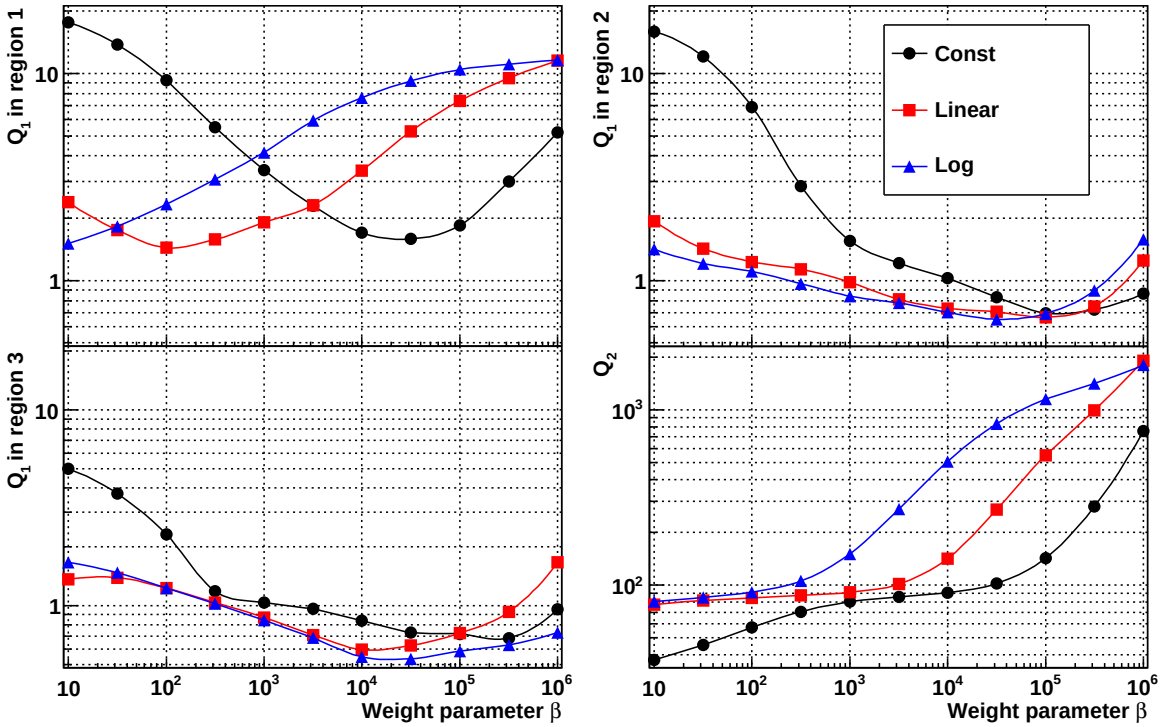

Figure 7.10: Performance of the χ^2-minimization.
The figure shows Q_1 for the regularizations (7.6) (const), (7.7) (linear), and (7.8) (log) and a range of values for β. The different panels show the result in the various regions: peak (top left panel, region 1), constant slope (top right panel, region 2), and low statistics (bottom left panel, region 3). The bottom right panel shows the sum of the squared residuals Q_2. The lines are drawn only to guide the eye. To present the results in the same plot, β in regularizations (7.6) and (7.8) has been multiplied by a factor 10^{-2}.

- **Constant slope (top right panel, region 2)**: the smallest Q_1 values obtained are significantly lower than for the first region. All regularizations reach their minima at $\beta = 10^4 - 10^5$. Around the minimum the value of Q_1 is not very sensitive to changes of β.

- **Low statistics (bottom left panel, region 3)**: the lowest values of Q_1 are achieved in this region. Like the previous region the value of Q_1 is not very sensitive to changes of β around the minima (also at $\beta = 10^4 - 10^5$). In this region significantly better results are achieved than in the case of Bayesian unfolding (discussed in the subsequent section).

- **Residuals (bottom right panel)**: as expected, the residuals increase with increasing weight factor. Each curve shows a rapid increase at a certain weight

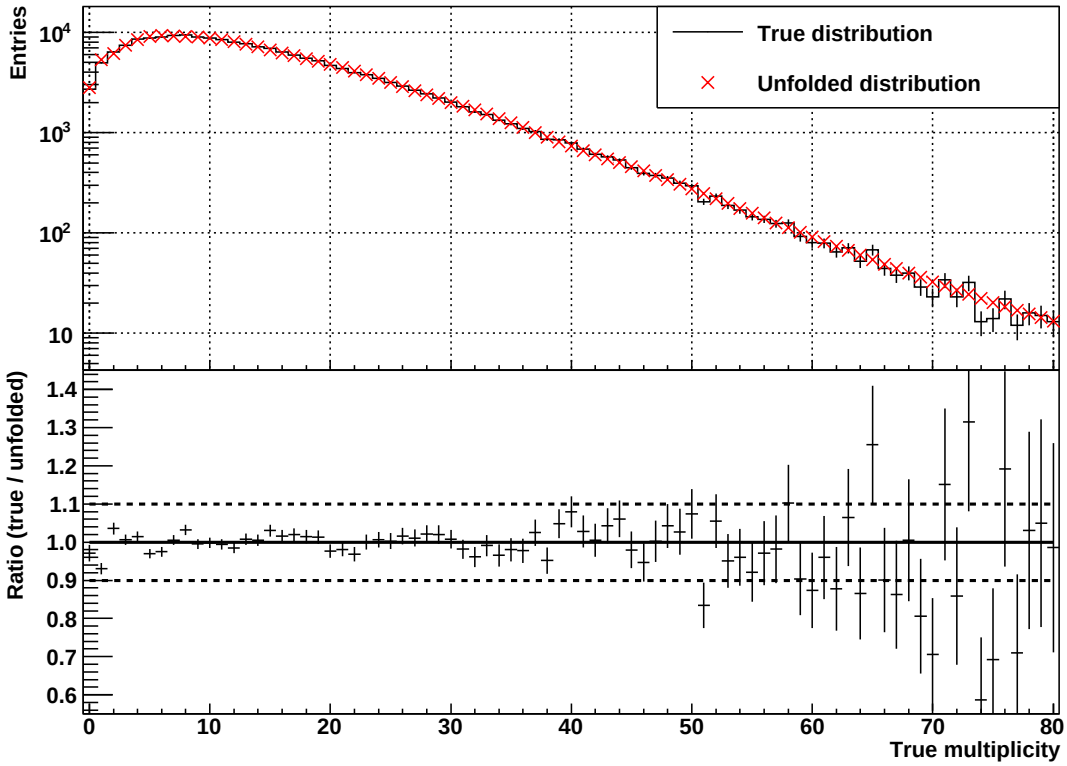

Figure 7.11: Verification of χ^2-minimization.
A negative binomial distribution ($\langle n \rangle = 15, k = 2$) has been used as input distribution. The response matrix created from a different input distribution (the one predicted by Pythia) was successfully used to unfold the spectrum using χ^2-minimization. The result is unsatisfactory starting from a multiplicity of about 70; this is explained by the low statistics in the input spectrum.

parameter. The residuals exceed twice the number of degrees of freedom (80 in this study) at $\beta = 10^5$, 10^4, and 10^3 for Eqs. (7.6), (7.7), and (7.8), respectively.

Proposed combinations that yield good results are Eq. (7.6) with $\beta = 10^5$ and Eq. (7.7) with $\beta = 10^3$. In the following sections regularization (7.6) is used with $\beta = 10^5$. For the unfolding of measured data several combinations should be used.

The validity of χ^2-minimization has been verified with different input spectra, including negative binomial distributions with different sets of parameters. An example is shown in Figure 7.11. In this example the number of events at multiplicities of 60 – 80 is less than in the Pythia example studied before, therefore fluctuations in this region are larger.

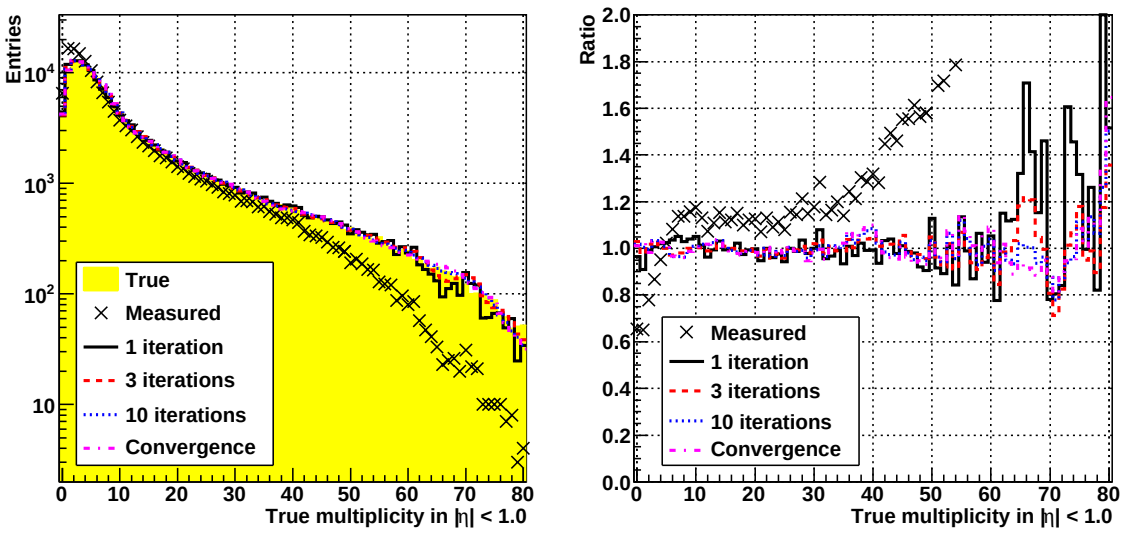

Figure 7.12: Convergence of the Bayesian unfolding.
The left panel shows the true distribution (filled histogram), the measured distribution (crosses), and unfolded distributions after 1, 3, and 10 iterations and after convergence (histograms). The right panel shows the ratio between the true distribution and the unfolded distributions. Also shown in the right panel is the ratio between the true and the measured distribution. $\alpha = 1$ has been used for the smoothing in the Bayesian unfolding.

7.3.3 Bayesian Unfolding

Bayesian unfolding converges quickly. Figure 7.12 shows the true distribution and unfolded distributions after 1, 3, and 10 iterations as well as after convergence of the method (30 – 40 iterations). Although the measured distribution is quite different from the true distribution, the result obtained by unfolding is close to the true distribution already after the first iteration. However, fluctuations can be seen especially at higher multiplicities. These reduce with further iterations. No significant difference can be seen between the result after 10 iterations and after convergence.

The free parameters of the method are evaluated in the following: these are the weight of the smoothing α (see Eq. (7.13)) and the number of iterations. α has been evaluated in its full range from 0 to 1 in steps of 0.2. The number of iterations is evaluated, starting from 2 iterations until the point at which the method terminates by convergence (which is in practice 30 – 40 iterations). For each of the cases the spectrum has been unfolded and the performance measures Q_1 in the three regions and Q_2 have been extracted. This is presented in Figure 7.13. Qualitatively the following can be remarked:

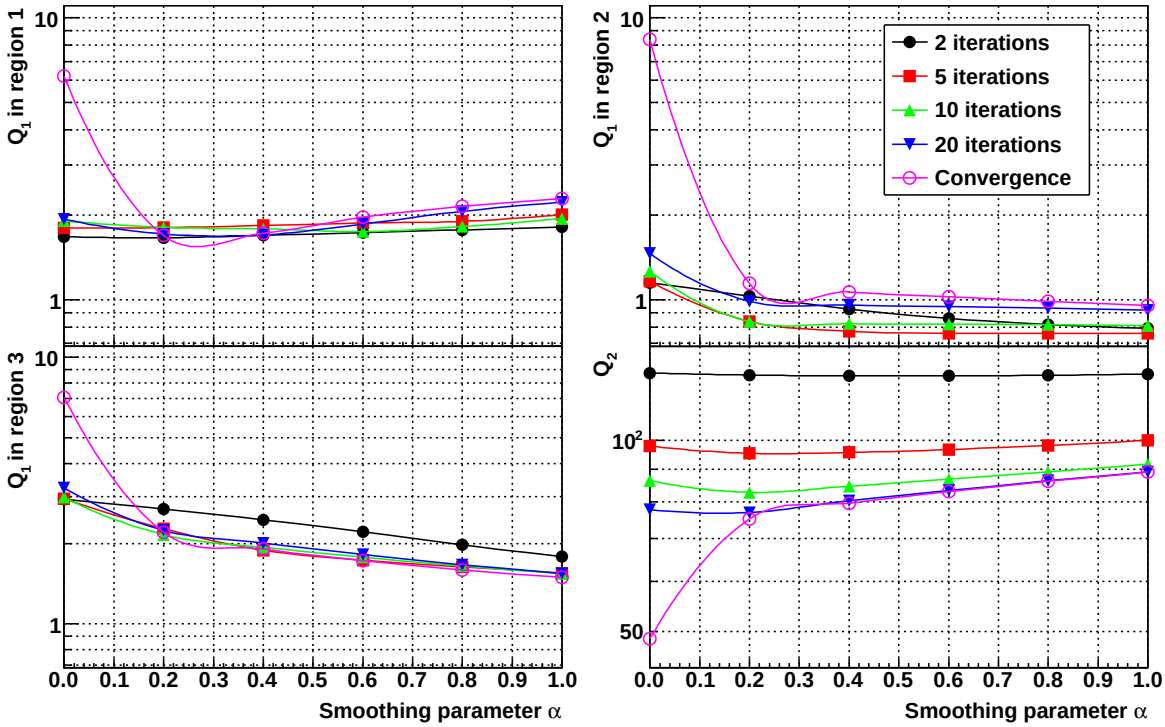

Figure 7.13: Performance of Bayesian unfolding.
The performance measure Q_1 for the full range of the smoothing parameter as well as a different number of iterations is shown. The panels show the results in the different regions: peak (top left panel, region 1), constant slope (top right panel, region 2), and low statistics (bottom left panel, region 3). The bottom right panel shows the sum of the squared residuals Q_2. The lines are drawn only to guide the eye.

- **Peak (top left panel, region 1)**: iterating a few times gives the best result. The smoothing has only a little influence for a low number of iterations but the results become slightly worse with increasing α. This is explained by the fact that the smoothing distorts a spectrum that has a fast changing slope. For a large number of iterations, light smoothing ($\alpha \approx 0.2$) improves the result; further smoothing has the opposite effect.

- **Constant slope (top right panel, region 2)**: applying light smoothing ($\alpha \approx 0.2$) improves the result. Only minor improvement can be seen by further increasing α. 5 to 10 iterations produce the best result. It becomes worse with more than 10 iterations.

- **Low statistics (bottom left panel, region 3)**: the influence of the smoothing is significant. Full smoothing ($\alpha = 1$) produces the best result. Two iterations

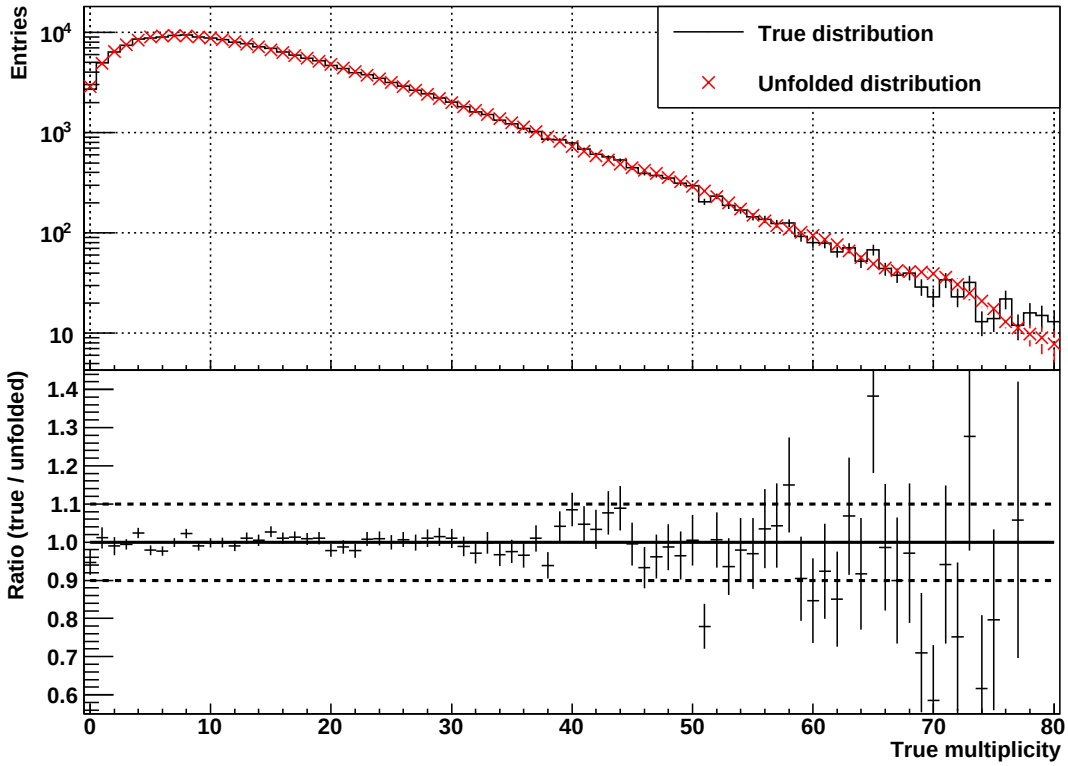

Figure 7.14: Verification of Bayesian unfolding.

A negative binomial distribution ($\langle n \rangle = 15$, $k = 2$) has been used as the input distribution. The response matrix created from a different input distribution (the one predicted by Pythia) was successfully used to unfold the spectrum applying Bayesian unfolding. As before, the result is unsatisfactory starting from a multiplicity of about 70; this is explained by the low statistics in the input spectrum.

are insufficient; no other strong dependence on the number of iterations can be seen (except for the convergence result, see below).

- **Residuals (bottom right panel)**: a higher number of iterations reduces the sum of the residuals. As expected the smoothing increases the residuals slightly. Still they remain of the order of the number of degrees of freedom (80 in this study).

In general, the result after convergence is unsatisfactory without smoothing. Q_1 in all regions is large and the residuals are very small. This is expected because for a large number of iterations the obtained result is close to the result of χ^2-minimization, which without regularization produces large fluctuations (see Figure 7.4 on page 159).

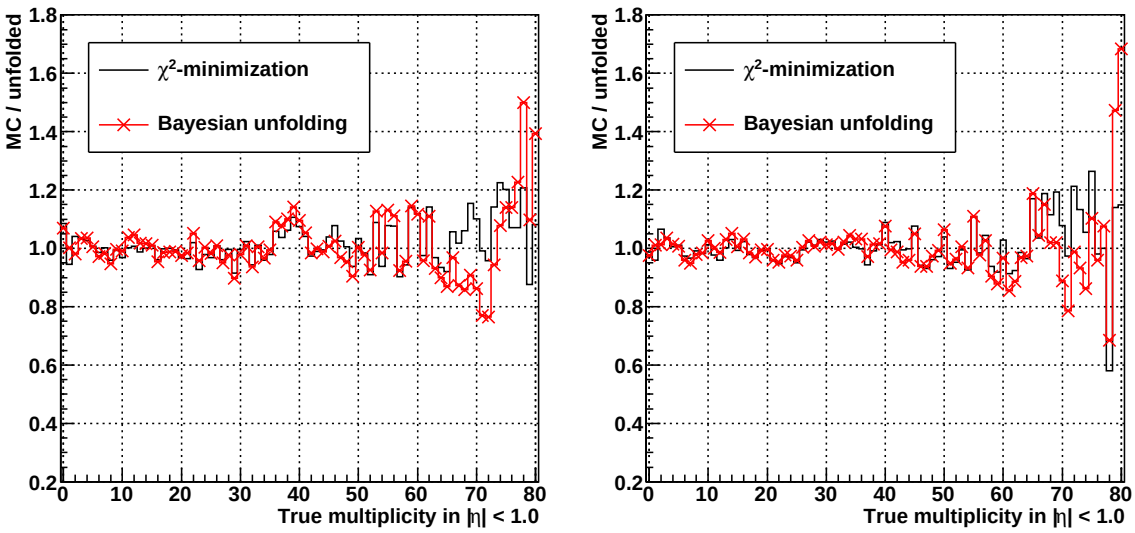

Figure 7.15: Comparison of χ^2-minimization and Bayesian unfolding.
The figure shows unfolded distributions for two different event samples (100 000 events each). Each plot shows the ratio of the unfolded distribution to the input distribution for χ^2-minimization (black histogram) and Bayesian unfolding (red crosses). The errors are suppressed for visibility. The fluctuations, caused by the limited statistics, agree mostly with each other for both methods.

For the following studies, 10 iterations with a smoothing parameter of $\alpha = 1$ is considered optimal. Clearly, a measured distribution needs to be unfolded with several parameter combinations to study the stability of the result.

Like χ^2-minimization, the validity of Bayesian unfolding has been verified with different input spectra. An example is shown in Figure 7.14.

7.3.4 Comparison of χ^2-Minimization and Bayesian Unfolding

Naturally, the two different methods should produce results that agree within statistical errors. Figure 7.15 shows unfolded distributions of two different event samples of 100 000 events each. Each of the results fluctuates due to the limited statistics. Naturally, each sample fluctuates in a different way. However, the fluctuations of the results of both methods agree mostly with each other, apart from the low-statistics region where the error associated with the unfolding method is larger (see Section 7.4.2). Such a comparison can be performed with measured data as a cross-check that the unfolding works successfully on the measured data.

	χ^2-minimization			Bayesian unfolding		
	Best values					
Region	Q_1	Regularization	β	Q_1	Iterations	α
Peak	1.44	Eq. (7.7)	10^2	1.66	2	0.2
Constant slope	0.65	Eq. (7.8)	3.2×10^4	0.78	5	0.6 – 1.0
Low statistics	0.53	Eq. (7.8)	3.2×10^4	1.49	convergence	1.0
	Chosen parameter set					
	Eq. (7.6) with $\beta = 10^5$			10 iterations with $\alpha = 1$		
Peak	$Q_1 = 1.81$			$Q_1 = 1.94$		
Constant slope	$Q_1 = 0.70$			$Q_1 = 0.81$		
Low statistics	$Q_1 = 0.71$			$Q_1 = 1.55$		
Residuals	$Q_2 = 140$			$Q_2 = 92$		

Table 7.3: Summary of the performance of the unfolding methods.
The table shows the best performance obtained in the three regions for both methods. Furthermore, the performance of the chosen parameter set is presented.

In the previous sections the performance in three regions that are characteristic for the shape of the distribution have been evaluated. Table 7.3 summarizes the best results. Note that the results for the different regions do not necessarily correspond to the same set of parameters. Therefore, the results for the chosen parameter set (given in the lower half of the table) are slightly worse than the optimal ones. It can be concluded that χ^2-minimization performs slightly better, especially in the region with low statistics.

7.3.5 Sensitivity to Initial Conditions

Both χ^2-minimization as well as Bayesian unfolding, start from an initial distribution. Usually the measured spectrum is used, as it is the only information available. However, the results should be consistent regardless of which initial distribution is chosen.

This has been verified by applying each method on the same measured spectrum using the same response matrix while changing only the initial conditions. In total six different initial distribution have been tried: three different measured distributions, the true distribution (predicted by Pythia), a negative binomial distribution, and a flat distribution were used. The unfolded distributions that are produced using different initial conditions are compared in Figure 7.16: the ratios between the unfolded distributions and the input distribution are shown. It can be seen that the initial conditions have

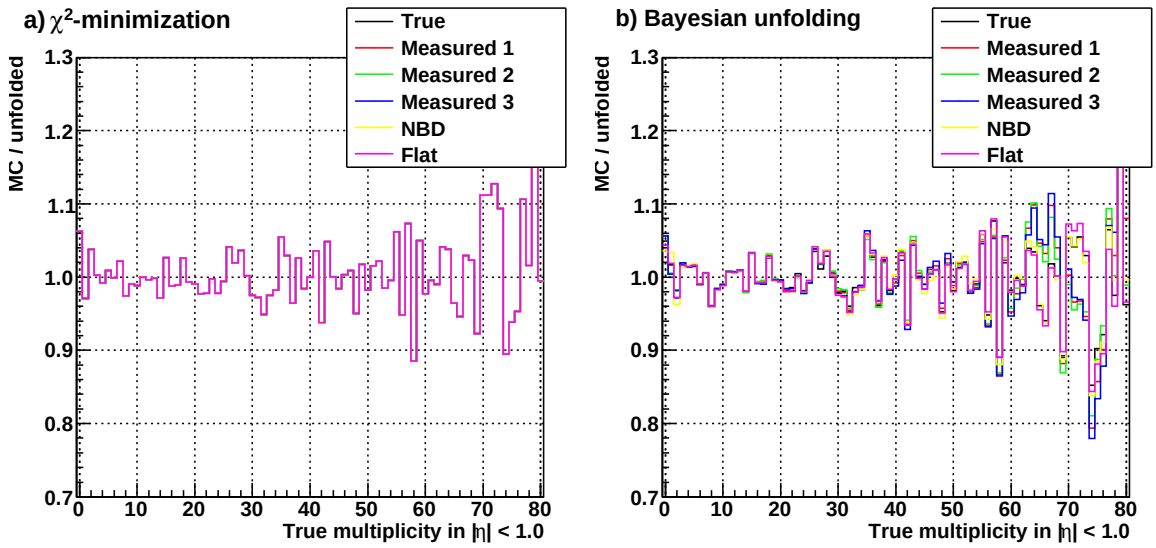

Figure 7.16: Effect of different initial conditions on the unfolded distribution. Shown are the ratios between the input distribution and the unfolded distribution. The left panel shows χ^2-minimization (all distributions overlap), the right panel Bayesian unfolding. The different results have been produced by only changing the initial conditions. The following initial conditions were used: three different measured distributions, the true distribution (MC input), a negative binomial distribution, and a flat distribution. Errors are omitted for visibility.

only little influence on the unfolded distribution. Slight deviations appear for Bayesian unfolding above a multiplicity of 60.

7.3.6 Conclusions

Two unfolding methods to reconstruct the multiplicity distribution have been introduced and evaluated: the χ^2-minimization and Bayesian unfolding. Both show good performance and their results agree within errors. To unfold real data both methods should be used and their results compared in order to increase the confidence in the unfolded spectrum.

7.4 Systematic Uncertainties

The unfolding methods that have been presented allow a measured spectrum to be unfolded based on a response matrix obtained from simulated events. It was shown that

the methods work independently of the shape of the multiplicity spectrum. Although the MC generator used to create these simulated events assumes a certain shape, this shape does not constrain the unfolded spectrum in any way. However, the procedure might be sensitive to other characteristics of the events and thus to assumptions made in the MC generator. Furthermore, effects like misalignment have an impact on the reconstruction and thus on the response matrix. This section will evaluate the systematic uncertainty of the multiplicity measurement.

The sources for systematic uncertainties that are discussed are similar to the ones discussed for the $dN_{ch}/d\eta$ measurement. In detail these are:

- the uncertainty of the unfolding method;
- the uncertainty in the cross-sections of the collision processes in the event generator;
- effects due to wrong assumptions of the particle-species abundances in the event generator;
- uncertainties due to the unknown p_T spectrum below the p_T cut-off;
- further effects due to incorrect assumptions in the event generator by comparing Pythia and Phojet;
- the effect of beam-gas, beam-halo, and pile-up events;
- the influence of a wrong estimation of the material budget in the simulation software;
- the effect of the uncertainty in the alignment of the detector;
- effects due to the tracklet selection cuts.

7.4.1 Characterization of a Systematic Uncertainty

The result of this measurement, the multiplicity distribution, is a probability distribution. Compared to the $dN_{ch}/d\eta$ distribution, a systematic effect will neither increase nor decrease the whole spectrum by a certain value. Instead the spectrum may be shifted or distorted. It is always (by construction) normalized to 1. Consequently a systematic effect can only be given as a function of multiplicity.

An artificial example can be seen in Figure 7.17. The figure shows the unfolded distribution of an event sample that was simulated with 5% lower tracking efficiency in

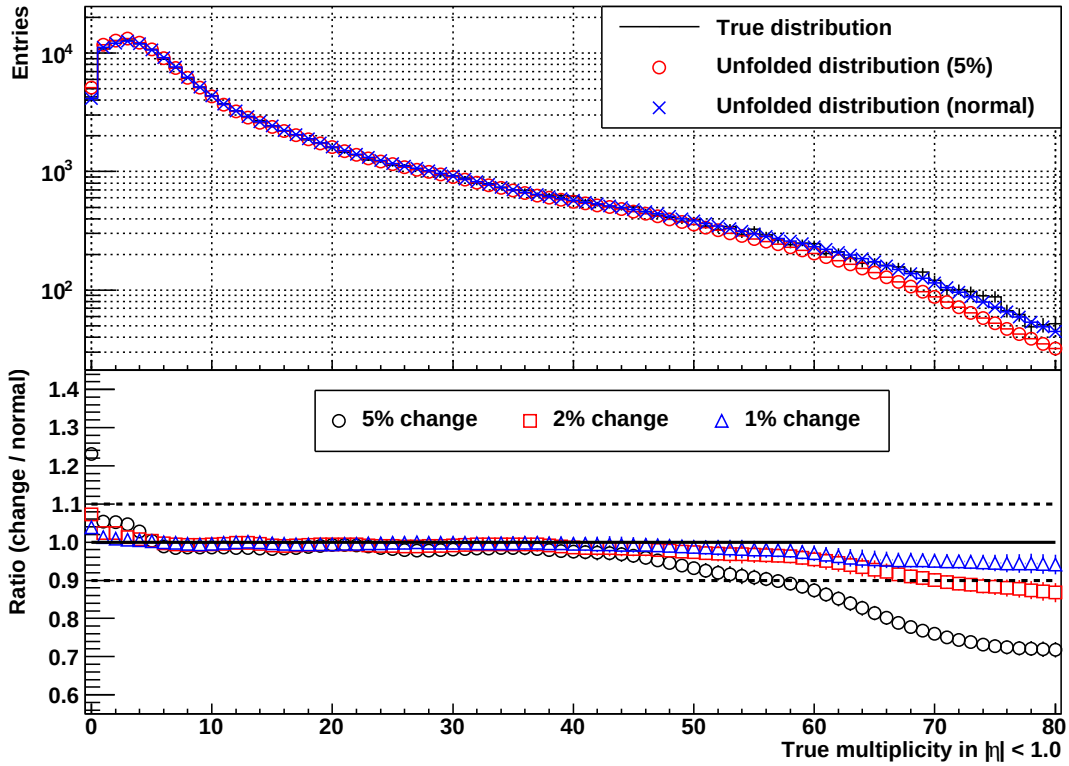

Figure 7.17: Illustration of a systematic uncertainty (artificial example).
The figure illustrates a systematic uncertainty by exaggerating an effect. It shows the unfolded distribution of an event sample that was simulated with 5% lower efficiency in the SPD (red circles). Superimposed is the unfolded distribution of an event sample simulated with unaltered efficiency (blue crosses), and the true distribution (MC input). The lower part of the figure shows the ratio of the two unfolded distributions which indicates the systematic uncertainty. Additionally, the ratios for the same case but simulated with 1% and 2% lower efficiency are shown.

the SPD (by removing 5% randomly selected reconstructed tracklets). Superimposed is the unfolded distribution of an event sample simulated with unaltered efficiency and the true distribution (MC input). The response matrix used was created from events simulated with unaltered efficiency as well. This example illustrates the effect on the unfolded distribution when the detector efficiency is wrongly estimated. However, a misestimation by 5% was chosen here for illustration only, less than 1% is expected in reality. The lower part of the figure shows the ratio of these two unfolded distributions. The difference from unity characterizes the systematic effect. As expected the spectrum is shifted to lower multiplicities. In addition, the effect of an event sample simulated with 1% and 2% lower efficiency is shown (only ratios, lower part of the

figure). Clearly, the magnitude of the effect at a given multiplicity depends on the shape of the unfolded distribution. Therefore, instead of applying the concluded uncertainties as they are found in this section to an unfolded measurement, measured data should be unfolded with several response matrices. These matrices should represent the different systematic uncertainties studied here. The unfolded distributions have to be subsequently compared and their difference will be used to quantify the systematic uncertainty for the actual measurement performed.

7.4.2 Uncertainty of the Unfolding Methods

A systematic uncertainty is associated with the unfolding method itself. This uncertainty is coupled with the statistical uncertainties of the measurement and of the response matrix. It describes how the statistical uncertainties propagate through the method. Thus the uncertainty derived in this section is linked to the number of events used as measurement and the number of events used to create the response matrix.

To find this uncertainty the sample is unfolded about 100 times, each time the entries of the measured spectrum, of the response matrix, or of both are randomized. Each bin is filled with a value sampled following a Poisson distribution using the value of that specific bin in the initial ('unrandomized') spectrum or matrix as its mean value. For each unfolded distribution the deviation from the unfolded distribution using the initial spectrum and matrix is determined. The standard deviation of the distribution per bin is used as uncertainty for that bin.

The measured value (that is used as mean for the sampling) is already randomized by nature and therefore the obtained error by this method is larger than the true one (approximately by a factor $\sqrt{2}$). The factor $\sqrt{2}$ is easily checked in an even more simple scenario by creating random samples from a parametrization, fitting the sample, and extracting the widths of the distributions of the fitted parameters. If the randomization is done twice cumulatively, i.e. using the second time the already randomized value as mean, the resulting dispersions are a factor $\sqrt{2}$ larger. The same factor is obtained if in the outlined procedure in each bin the value is randomized twice.

Figure 7.18 shows the uncertainty of the measurement (divided by $\sqrt{2}$) for the statistics used in this study (response matrix: 270 000 events; measured sample: 200 000 events). The uncertainty is about 1 – 2% for low multiplicities for both methods. It increases to 2% for χ^2-minimization and 4% for Bayesian unfolding at a multiplicity of 60. At a multiplicity of 80, χ^2-minimization reaches about 3% and Bayesian unfolding about 12%. The uncertainty is dominated by the variation of the measured spectrum. The variation of the response matrix gives a significantly smaller contribution. The uncer-

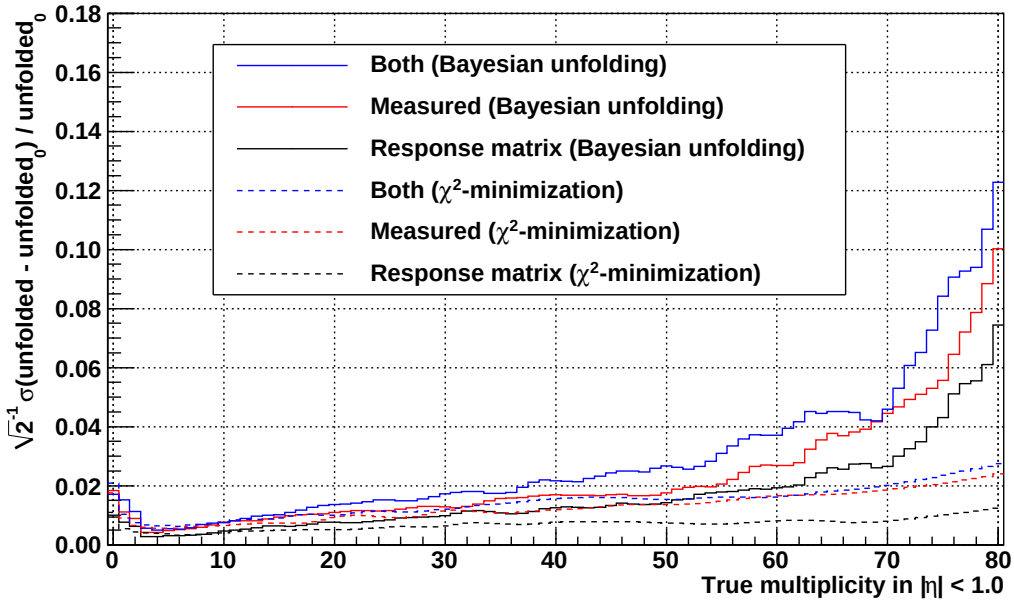

Figure 7.18: Uncertainty of the unfolding methods.

The figure shows the uncertainty due to the unfolding procedure: Bayesian unfolding (solid lines) and χ^2-minimization (dashed lines). The uncertainty that arises due to the limited statistics of the response matrix (black, bottom line), the measurement (red) and the combination of the two (blue, top line) are shown for both procedures. The uncertainty using χ^2-minimization is lower except for low multiplicities where the error is similar. For both methods the overall error is dominated by the statistical error on the measured spectrum.

tainty found by randomizing the measured spectrum can be interpreted as statistical uncertainty because it is only caused by the statistical uncertainty on the measured spectrum. However, the combined uncertainty (response matrix and measured spectrum) is used in this analysis and therefore handled as systematic uncertainty. The outlined procedure can also be applied to measured data.

A major cause for this uncertainty is the resolution of the detector which results in oscillations that can be seen in the unfolded distribution (e.g. Figures 7.5, 7.6, and 7.16). Tests have been performed with artificial response matrices that show that the width of the oscillations corresponds to the resolution of the detector. The resolution of the detector is multiplicity-dependent, which results in changes in the width of the oscillations. They have to be accepted as an artifact of the method, however, their magnitude is low and this is already included in the systematic uncertainty derived here.

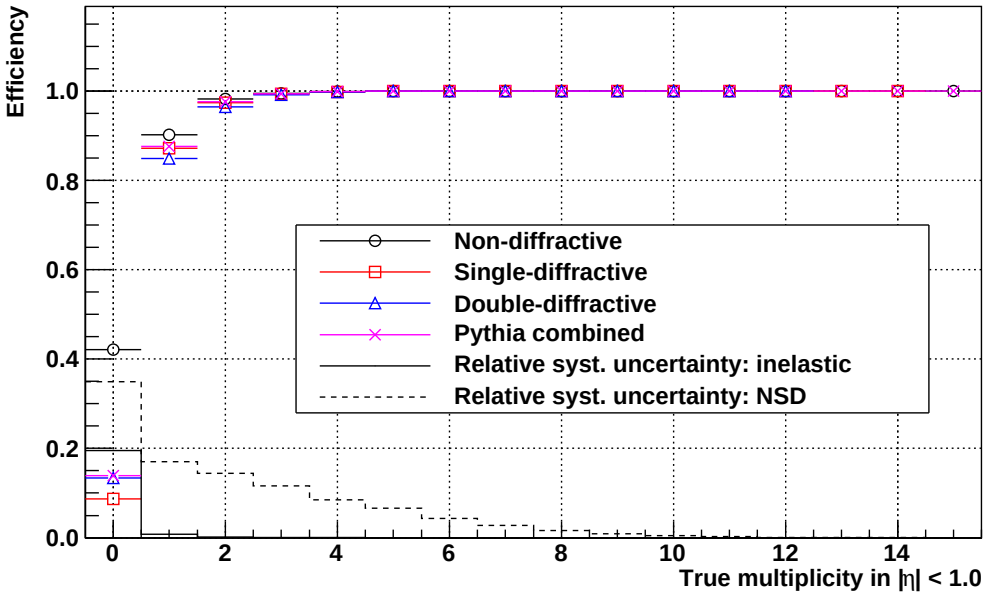

Figure 7.19: Systematic effect of the unknown cross-sections.
The figure shows the combined trigger and vertex-reconstruction efficiency for different process types and the combined value (Pythia) which corresponds to the correction to the inelastic event sample. The relative systematic uncertainties on the process types (see text) are shown as histograms (solid histogram: inelastic event sample; dashed histogram: NSD event sample).

7.4.3 Cross-sections of Physics Processes

Measured events originate from different physical processes. The Pythia simulation, used to generate events to build the corrections, assumes certain cross-sections for these process types. This section evaluates the uncertainty on the applied corrections due to the uncertainty on the cross-sections. In order to study the effect of a change in the relative cross-sections, the corrections have been calculated by changing the diffractive cross-sections to 50% and 150% of the Pythia values, similar to the procedure described in Section 6.3.1.

Several response matrices have been built, changing the mentioned cross-sections. The effect on the unfolded distribution is negligible (plot not shown). This is expected because the change of the cross-section basically reflects a change in the multiplicity spectrum, to which the method is insensitive. However, an effect can be seen on the

correction of the trigger and vertex-reconstruction efficiency. This correction is applied after the unfolding step[3] and its numerical factors depend on the process type.

Figure 7.19 shows the combined trigger and vertex-reconstruction efficiency for ND, SD, and DD events. Also shown is the combination of these using the cross-sections used by Pythia. Several other combinations have been tested by changing the cross-sections as mentioned before. The maximum deviation between these distributions and the default suggested by Pythia is calculated for each multiplicity. These values are shown as relative errors as histograms in Figure 7.19. The solid histogram shows the errors for the inelastic event sample and the dashed histogram for the NSD event sample. For the inelastic event sample the error is small except for the 0-bin where it is 20%. This is due to the fact that the trigger efficiency for this bin depends largely on the behavior outside the considered region. At a multiplicity of 1 the error is about 1% and it is below 1% for higher multiplicities. The uncertainty for the correction to the NSD event sample is larger due to the fact that the SD events have to be subtracted. It is about 35% in the 0-bin and drops progressively below 1% at a multiplicity of 10. The values of these histograms are the systematic uncertainties due to the uncertainty on the relative cross-sections of the different process types.

7.4.4 Particle Composition

The simulated events used to obtain the correction factors and the response matrix are produced under the assumption of certain ratios between the different particle species. These ratios will be measured at the LHC, but this might only be performed after the very first analysis. A study similar to the one for the $dN_{ch}/d\eta$ measurement (Section 6.3.2) has been performed to estimate the uncertainty introduced by the incomplete knowledge of the true particle ratios.

For this purpose several modified response matrices have been created altering the ratios between the particle types. Pions, kaons, and (anti-)protons, that compose 98% of all particles in Pythia events, have been taken into consideration. The number of pions was kept fixed, while the number of kaons, of protons, or of both was enhanced or reduced by 50%. Technically, for each generated particle or measured tracklet, the enhanced or reduced number was counted, to then be filled into the response matrix. Eight response matrices have been created by enhancing or reducing the yields of kaons and (anti-)protons. These changes alter the overall efficiency due to the fact that the

[3]In the case of Bayesian unfolding the trigger and vertex-reconstruction correction can also be applied during unfolding. However, it can also be applied afterwards and, as previously mentioned, the results are consistent.

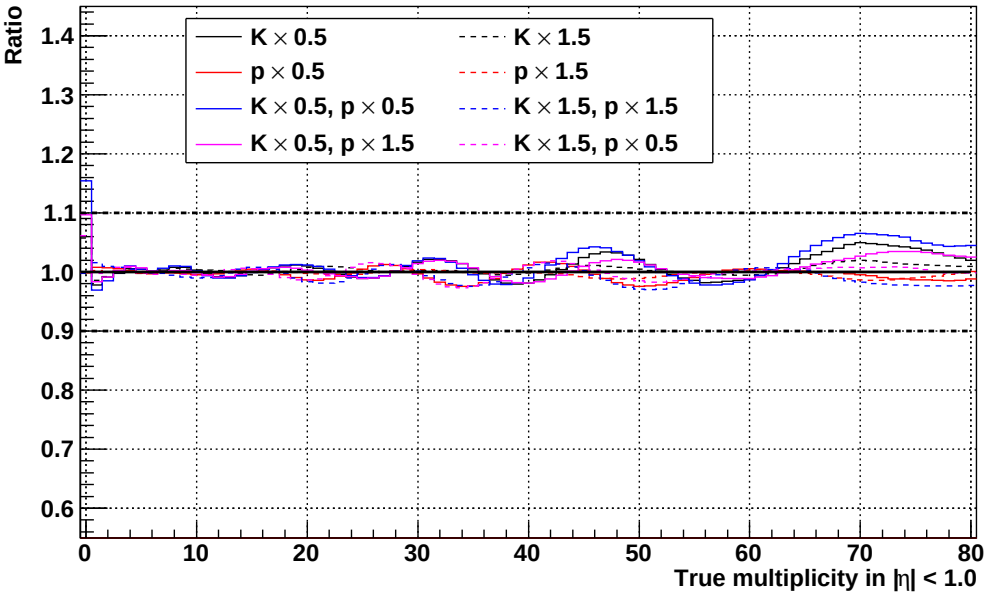

Figure 7.20: Systematic effect of changed particle ratios.
Plotted is the ratio between the unfolded spectrum obtained with an unaltered response matrix and unfolded distributions using modified response matrices. The ratios between pions, kaons, and protons have been enhanced or reduced by up to 50%.

efficiency to measure a particle depends on its species at low p_T (see Figure 6.11 on page 129). Each response matrix was used for unfolding the same measured spectrum.

Figure 7.20 shows the ratio between the unfolded distribution using the unmodified response matrix and the other unfolded distributions. The uncertainty is rather small, apart from the lowest multiplicity bin. It increases up to 6% at a multiplicity of 80. The oscillation is an effect of the previously described detector resolution.

7.4.5 p_T Spectrum

The MC simulation assumes a certain shape of the p_T spectrum which might be different from the real one. The response matrix is integrated over p_T, which may give rise to a systematic error. ALICE will measure the p_T spectrum, except the region of low p_T (cut-off at about 200 MeV/c, see Section 6.2.5). In the case the multiplicity distribution is measured before the p_T spectrum, some constraints for the p_T spectrum are still available from first data. A few percent of change of the spectrum do not influence the unfolded distribution. Therefore, it remains to study the effect of an uncertainty at low p_T.

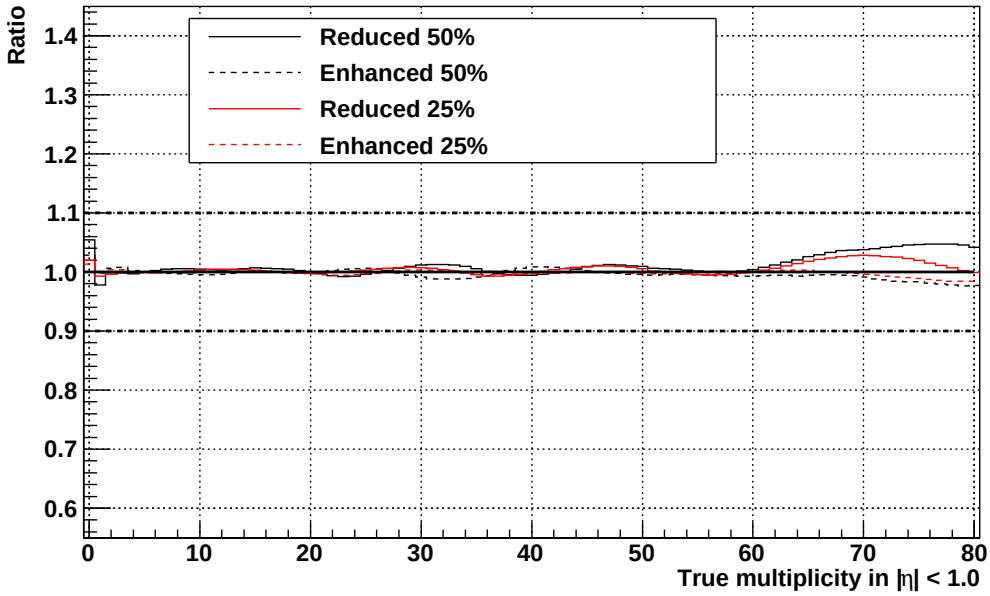

Figure 7.21: Systematic effect of the uncertainty in the p_T spectrum.
The figure shows the ratio between the unfolded spectrum obtained with an unaltered response matrix and unfolded distributions using modified response matrices. The modified response matrices have been obtained by changing the p_T spectrum (see text).

Similar to the procedure described in Section 6.3.3, the spectrum is changed below the p_T cut-off of 0.2 GeV/c by up to 50%, see Figure 6.17 (page 140). Response matrices are produced applying the modified p_T spectra. Because only integer particle numbers are meaningful, this is done in the following way: during the analysis each primary particle as well as each reconstructed tracklet is counted up to two times depending on its p_T and the respective change in the p_T spectrum. For example when the factor between the default and modified p_T spectrum is 0.9 for a given p_T, a random number generator is used to decide whether the particle is counted or not. The generator is configured in a way that it accepts 90% of the particles and rejects 10%. In case of a factor greater than 1, the random number generator decides if a particle is counted once or twice.

A measured distribution obtained with the unmodified p_T distribution is unfolded using an unmodified response matrix and the modified response matrices. Figure 7.21 shows the ratios between these unfolded distributions. The resulting effect is below 1% except for the 0-bin (6%) and increases starting from a multiplicity of 60 to about 6% at a multiplicity of 80.

7.4.6 Event-Generator Assumptions

The effect of the usage of a different event generator on the analysis result is studied by comparing Pythia and Phojet.[4] These have a significantly different multiplicity distribution, as shown in Figure 8.5 (page 199). Although the number of events is similar, the maximum multiplicity obtained at a given statistics is much lower for Phojet. The unfolding is insensitive to the multiplicity distribution in the simulated events used for the response matrix. However, the highest multiplicity in the measured distribution has also to be reached in the response matrix. Therefore, the raw measured distribution has to be compared to the distribution of the simulated events before the unfolding is applied. In the example studied here, the consequence is that a Phojet sample can be unfolded with a response matrix obtained from Pythia events but not the other way round (unless a significantly larger number of Phojet events was to be simulated). Figure 7.22 shows a Phojet sample unfolded with a Pythia and a Phojet response matrix. The comparison between the two unfolded distributions shows the typical fluctuations. The effect is at a level of a few percent, but it increases for multiplicities of $0 - 2$ and also above 40 to about 10%. These numbers cannot be directly compared to the other uncertainties due to the different input distribution used. This illustrates clearly that the final systematic uncertainties have to be extracted by applying different response matrices on the measured distribution.

7.4.7 Beam-Gas and Beam-Halo Events

The contamination by beam-gas and beam-halo events has been discussed in Section 5.1.1. For the startup scenario it has been estimated to be negligible. Under nominal running conditions the probability that a triggered event is a beam-gas or beam-halo collision is estimated to be 0.6%. Such events usually do not have a reconstructed vertex and are therefore not considered in this analysis. Therefore no systematic effect on the multiplicity distribution is expected.

However, if they were to be considered (e.g. through a modification of the analysis method) the mentioned 0.6% that have neither reconstructed vertex nor tracklets have to be seen in relation to the amount of events that have no primary particle in the considered η-range. For example for $|\eta| < 1$ the probability is 11% (Pythia), thus the effect of beam-gas would be about 5% in the first bin.

[4] The datasets LHC08c11 and LHC08c15 have been used for this study (see Section 5.5).

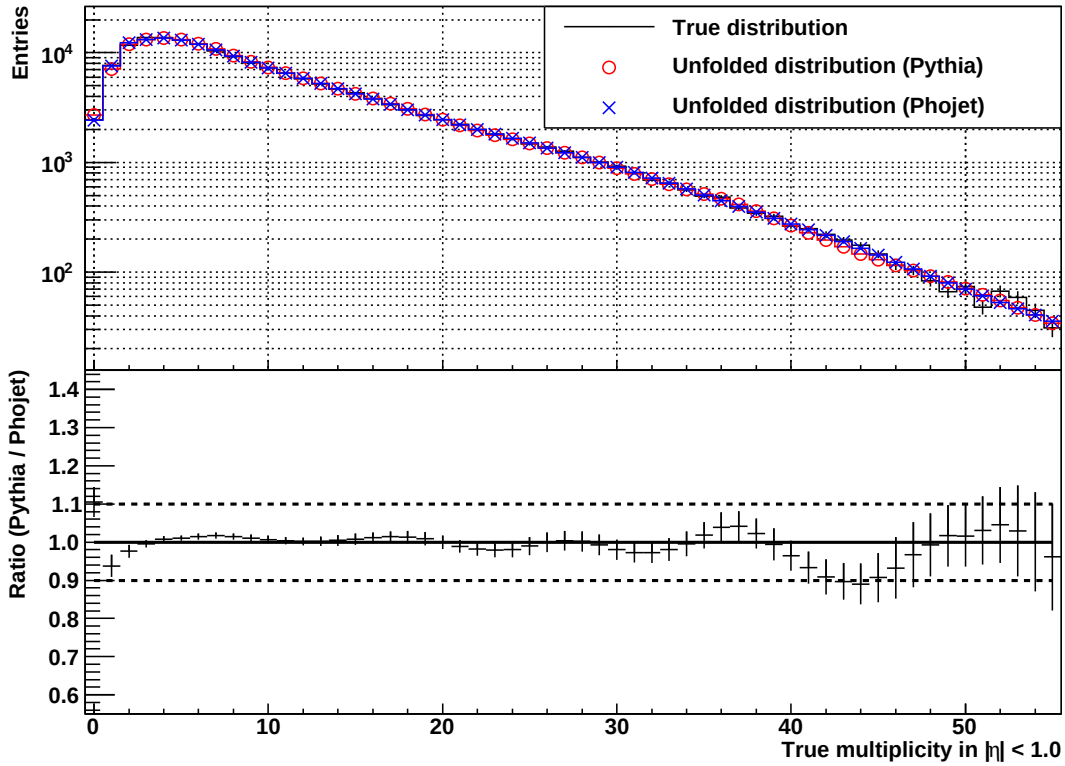

Figure 7.22: Effect of event-generator assumptions.

The figure shows a Phojet sample unfolded with a Pythia response matrix (red circles) and with a Phojet response matrix (blue crosses). In the lower panel the ratio between the two is shown. Compared to the previous plots, a different x-axis range is shown because the multiplicity distributions of Pythia and Phojet differ: the number of events that is reached with Pythia at a multiplicity of 80, is reached with Phojet already at a multiplicity of 55.

7.4.8 Pile-Up Events

The rate of unresolved pile-up events was estimated in Section 6.3.6 and it was found to be negligible for the $dN_{ch}/d\eta$ measurement. However, for the multiplicity distribution pile-up might influence the high-multiplicity region because two events with small multiplicity might appear as an event with higher multiplicity. The probabilities for unresolved pile-up are 4.4×10^{-5} (startup scenario at $\sqrt{s} = 900\,\text{GeV}$), 5.8×10^{-4} (startup scenario at $\sqrt{s} = 10\,\text{TeV}$), and 4.3×10^{-5} (nominal running conditions at $\sqrt{s} = 10\,\text{TeV}$).

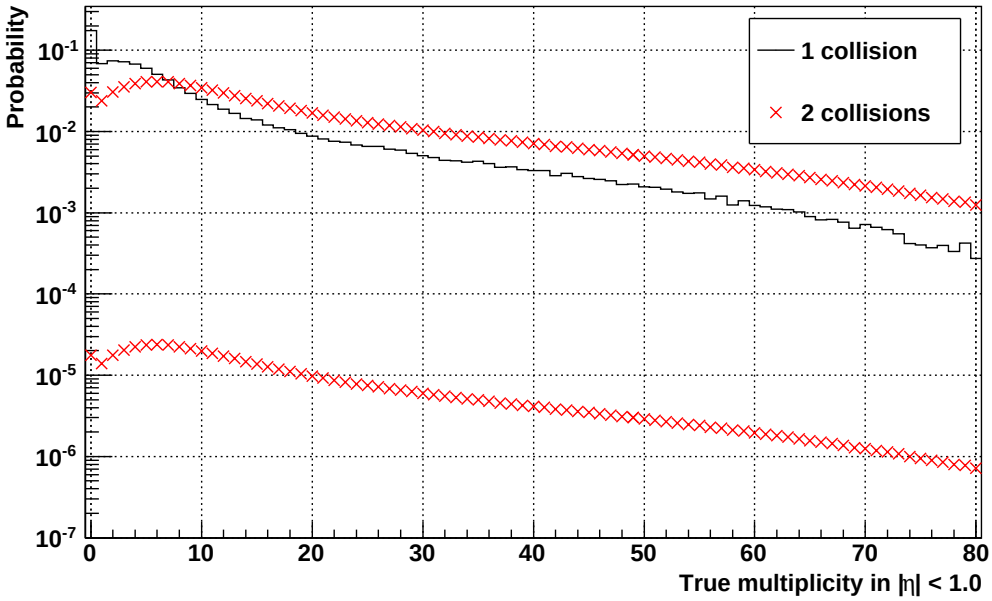

Figure 7.23: Pile-up assessment.

The figure shows the multiplicity distribution by Pythia for a single collision (black histogram) and the distribution of unresolved pile-up, thus two overlapping collisions (red crosses): the upper one is normalized to 1, the lower one is normalized by the probability that unresolved pile-up occurs.

Figure 7.23 estimates the influence for the case with the largest probability (startup scenario at $\sqrt{s} = 10\,\text{TeV}$). Shown is the multiplicity distribution in $|\eta| < 1$ for a single collision and for two overlapping collisions (this is calculated by convoluting the multiplicity distribution with itself). By scaling the convoluted distribution with the probability of unresolved pile-up events it can be seen that the influence increases for larger multiplicities, but is still below 0.3% at a multiplicity of 80. For the other cases (startup scenario at $\sqrt{s} = 900\,\text{GeV}$ and nominal running conditions), the uncertainty is about an order of magnitude lower. This uncertainty might become more significant if very large multiplicities are accessed.

7.4.9 Material Budget

A study of the effect of changes of the detectors' material budget was described in Section 6.3.7. It showed that a change of the material budget of ±10% has no effect on the number of tracklets measured with the SPD. Therefore the systematic effect by misestimation of the material budget is negligible.

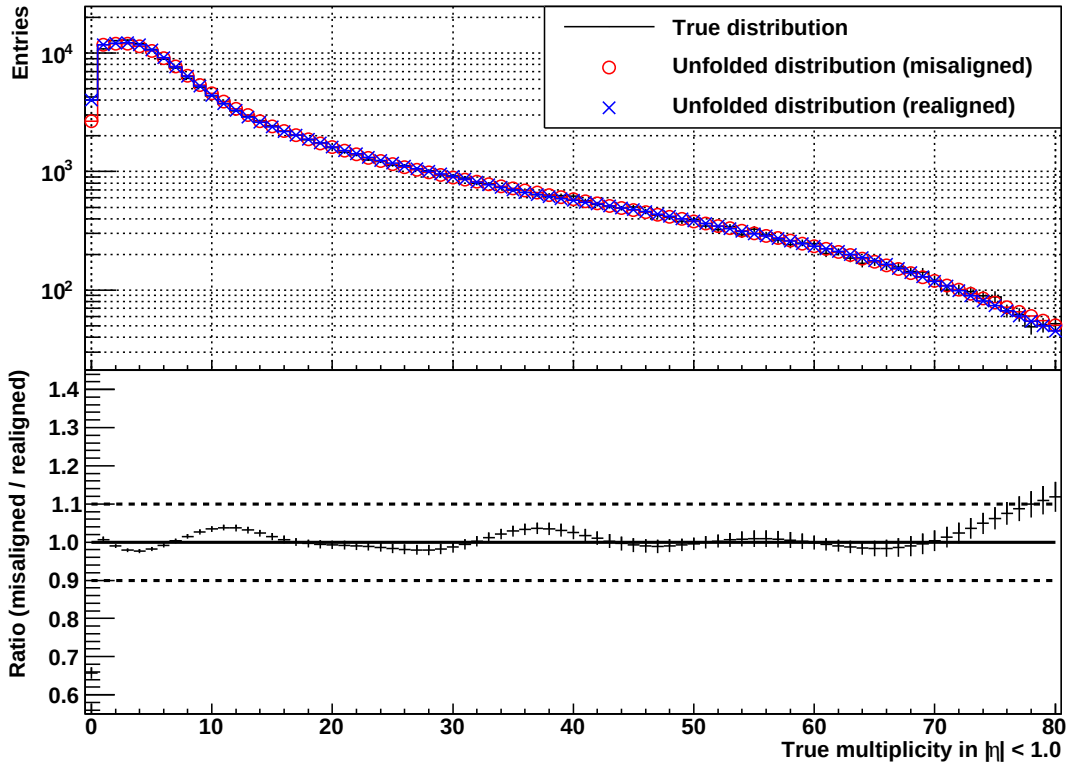

Figure 7.24: Systematic effect of misalignment.
A sample reconstructed without realignment is unfolded with a response matrix obtained from a sample reconstructed with (blue crosses) and without realignment (red circles). The lower panel shows the ratio between the two.

7.4.10 Misalignment

The alignment situation and the residual misalignment has been described in Section 4.7.1. Similar to the $dN_{ch}/d\eta$ measurement, no effect from the residual misalignment, i.e. after realignment procedures have been applied, is expected. This was discussed in Section 6.3.8. However, as in the previous chapter, the effect without realignment is studied. Therefore, a sample that has been simulated with the ideal geometry but reconstructed with the realigned geometry, is unfolded with a response matrix that used the ideal geometry also in the reconstruction step. Figure 7.24 shows the unfolded spectrum superimposed with the input distribution. The deviation is small and increases towards large multiplicities (about 10%); in the 0-bin it is about 35%. This effect only needs to be added to the systematic uncertainty if the reconstruction of measured data is performed without realignment; this is not expected for first data.

Uncertainty	$N = 0$	$1 \leq N \leq 80$
χ^2-minimization*	1.5%	1 – 3%
Bayesian unfolding*	2%	1 – 12%
Relative cross-sections (INEL)	20%	< 1%
Relative cross-sections (NSD)	35%	1 – 17%
Particle composition	16%	1 – 6%
p_T cut-off	8%	1 – 6%
Event generator assumptions	10%	1 – 10%
Beam-gas	negligible	
Pile-up	0	< 0.3%†
Material budget	negligible	
Misalignment	negligible	
Tracklet selection	negligible	
Total (INEL)‡	26%	2 – 9%
Total (NSD)‡	39%	2 – 18%

Table 7.4: Summary of the various systematic uncertainties.
The uncertainties are determined as a function of multiplicity which is represented by the spread given. The 0-bin is listed separately due to the larger uncertainties. For details see the corresponding sections and Figure 7.25. The uncertainties marked with an asterisk depend on the number of events measured and simulated for the response matrix.
† Only for the startup scenario at $\sqrt{s} = 10\,\text{TeV}$, otherwise negligible. See Section 7.4.8.
‡ The total does not include the event generator assumptions that are mostly included in other uncertainties.

7.4.11 Tracklet Selection

Uncertainties in the tracklet selection are similar to uncertainties in the efficiency. It has been discussed in Section 6.3.9 that an uncertainty on the number of accepted tracklets of less than 10^{-3} is expected. Therefore, the effect on the multiplicity distribution is negligible.

7.4.12 Summary of the Systematic Uncertainties

The systematic uncertainties that have been evaluated in this section are summarized in Table 7.4. The minimum and maximum values in the multiplicity region up to 80 are quoted. It is more illustrative to show the uncertainty as a function of the

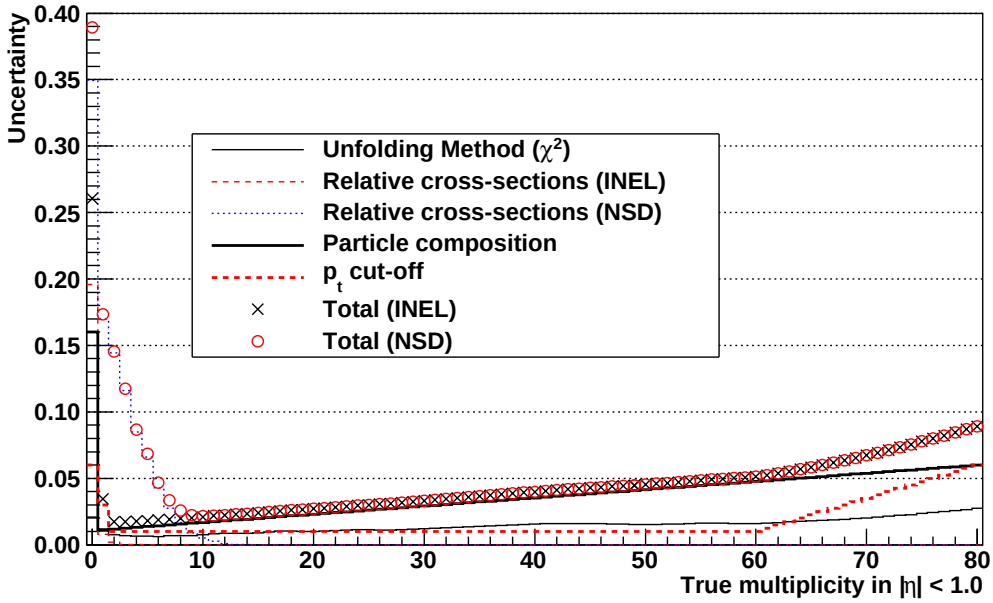

Figure 7.25: Summary of the various systematic uncertainties.
Summary of the uncertainties that need to be applied to an unfolded distribution using χ^2-minimization. In the cases of significant oscillations due to the unfolding, upper limits are used instead of the particular value at a given multiplicity. The uncertainty due to the event-generator assumptions cannot be displayed in this plot because it has been obtained with a different multiplicity axis. The small uncertainty due to pile-up is not shown. The total uncertainty (summed quadratically) is shown for the inelastic (black crosses) and NSD event samples (red circles).

multiplicity which can be seen in Figure 7.25. Also shown is the resulting uncertainty when the considered uncertainties are summed quadratically. Some of the uncertainties will reduce significantly when other measurements are performed, like the measurement of the cross-sections and the particle abundances.

7.5 Towards the Corrected Multiplicity Distribution

The steps that have to be taken to obtain the corrected spectrum from measured data have been discussed in detail in Section 6.4 for the $dN_{ch}/d\eta$ measurement. These steps apply also to the multiplicity distribution. However, the multiplicity distribution is less sensitive to beam-gas events (no normalization to the number of events is applied). Additionally, it has to be verified that the response matrix is populated sufficiently for

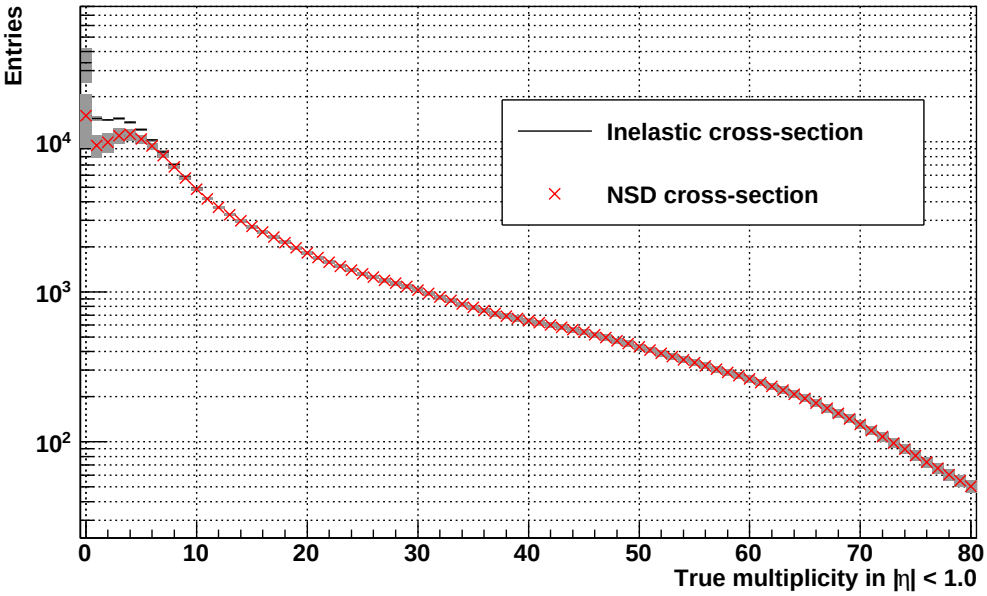

Figure 7.26: Final multiplicity distribution.

The figure shows a corrected multiplicity distribution of inelastic events (black line) and NSD events (red crosses) measured with the SPD. χ^2-minimization has been used to unfold the measured distribution. The gray band indicates the total systematic error based on the uncertainties discussed in this chapter.

all measured multiplicities. It is recalled that most if the systematic uncertainties have to be obtained by unfolding the measured distribution with different response matrices.

7.6 Summary

The method and the necessary corrections to produce a multiplicity distribution with p+p data taken by ALICE have been described. Two methods to correct the measured multiplicity distribution have been introduced and evaluated, including their systematic uncertainties. Figure 7.26 shows a corrected multiplicity distribution (inelastic and NSD event samples) with statistical and systematic errors for the case of the measurement with the SPD. Such a spectrum can be obtained with 200 000 measured events that can be collected even with the bunch-crossing trigger with about two days of data-taking.

Chapter 8

Predictions for LHC Energies

This chapter summarizes existing predictions for pseudorapidity density and multiplicity distributions of charged particles for LHC energies. Predictions based on QGSM calculations and results obtained with the event generators Pythia and Phojet are discussed and given for $\sqrt{s} = 10\,\text{TeV}$ and $14\,\text{TeV}$ (where available). Furthermore, three phenomenological approaches are mentioned: the first combines two negative binomial distributions (NBDs), the second is based on a multiple-parton interpretation, and the third is an approach that scales distributions based on observed trends at lower energies. The functional fits that have been applied to data at lower energies in Chapter 2 are also given. Some predictions are not shown in their respective sections to avoid redundant plots. These are included in the summary plots shown at the end of the chapter.

8.1 Quark–Gluon String Model

The Quark–Gluon String Model (QGSM), introduced in Section 1.4, allows the calculation of predictions for multiplicity and rapidity distributions. It is not straightforward to derive multiplicity distributions in limited pseudorapidity intervals and pseudorapidity distributions because QGSM does not predict the p_T distribution.

The translation from dN_{ch}/dy to $dN_{ch}/d\eta$ is done by [Won94]:

$$\frac{d^3N_{ch}}{d\eta d^2p_T} = \sqrt{1 - \frac{m_0^2}{m_T^2 \cosh^2 y}}\ \frac{d^3N_{ch}}{dyd^2p_T}. \tag{8.1}$$

Thus for large y (η), $dN_{ch}/d\eta \approx dN_{ch}/dy$. For an exact calculation, the dN_{ch}/dy per particle species (for m_0 and m_T) and the p_T distribution need to be known. These

translation factors are significant. For example Pythia and Phojet simulations show that the region $|\eta| < 0.5$ does not include 18% of the particles that are in $|y| < 0.5$ due to the fact that particles at low p_T may have a small rapidity y but still a large pseudorapidity η (at $\sqrt{s} = 10\,\text{TeV}$). In order to translate a dN_{ch}/dy prediction which is integrated over all particle species [Kai99], the conversion factor as a function of y is derived from simulated Phojet events.[1] The prediction before and after the transformation is shown in the left panel of Figure 8.3 (page 198). The transformed prediction as a function of η should only be considered as an approximation.

The same procedure could in principle be applied to multiplicity distributions. However, correlations between $\langle p_T \rangle$ and the multiplicity, which were already measured at lower energies, cannot easily be included within this procedure.

A prediction for the multiplicity distribution in full phase space [Pog08] is included in Figure 8.4 (page 199). Further predictions up to $\sqrt{s} = 100\,\text{TeV}$ can be found in [Kai99].

8.2 Pythia and Phojet

The event generators Pythia and Phojet were used to produce simulated data for this thesis. The data was used to evaluate the analysis procedures and to extract systematic uncertainties. Table 8.1 shows differences between Pythia with the tune mentioned in Section 1.5.1 and Phojet at $\sqrt{s} = 10\,\text{TeV}$ in relevant observables for this thesis. The Phojet events have a 24% higher ND component than the Pythia events. The largest difference is in the DD cross-sections (42%). The multiplicities predicted by the Phojet simulation are generally lower than those of the Pythia simulation; this is more significant for NSD events where they are 20% lower. This effect cannot be attributed to changes in the contributions of the different processes. ND events usually have a higher multiplicity and therefore their increase should also increase the multiplicity. In fact the opposite occurs, therefore the change in multiplicity is due to the different kinematics of the two generators.

Although both generators reproduce results at $\sqrt{s} = 900\,\text{GeV}$ reasonably well (Figure 1.10 on page 36), their predictions differ significantly at higher energies. Predictions for $dN_{ch}/d\eta$ and multiplicity distributions are included in Figures 8.3, 8.4, and 8.5. The first multiplicity measurements (even with the errors that are associated with early analyses) will already provide an indication as to which (if either) generator will reproduce the data.

[1] The result agrees within 1% if Pythia events are used.

Observable	Pythia	Phojet	Difference
σ_{ND}	52.1 mb	64.8 mb	24%
σ_{SD}	14.0 mb	10.9 mb	−22%
σ_{DD}	9.3 mb	5.4 mb	−42%
σ_{NSD}	61.4 mb	70.2 mb	14%
σ_{INEL}	75.4 mb	81.1 mb	8%
Inelastic events			
$\langle N_{ch}\rangle$	67.3	59.9	−11%
$dN_{ch}/d\eta\|_{\eta=0}$	5.21	4.54	−13%
$dN_{ch}/d\eta\|_{\|\eta\|<1}$	5.34	4.65	−13%
NSD events			
$\langle N_{ch}\rangle$	79.2	64.9	−18%
$dN_{ch}/d\eta\|_{\eta=0}$	6.24	4.97	−20%
$dN_{ch}/d\eta\|_{\|\eta\|<1}$	6.40	5.09	−20%

Table 8.1: Pythia and Phojet at $\sqrt{s} = 10\,\mathrm{TeV}$.

The table shows predicted cross-sections for the different physics processes and multiplicities in full and limited phase space. The difference is given as relative difference of the Phojet value to the Pythia value.

8.3 Two-Component Approach with NBDs

A two-component approach that uses two negative binomial distributions was introduced in Section 1.6.4. It is recalled that the multiplicity distribution is described by:

$$P(n) = \alpha_{\mathrm{soft}} \times P(n; \langle n\rangle_{\mathrm{soft}}; k_{\mathrm{soft}}) + (1 - \alpha_{\mathrm{soft}}) \times P(n; \langle n\rangle_{\mathrm{semi\text{-}hard}}; k_{\mathrm{semi\text{-}hard}}). \quad (8.2)$$

The parameters are fitted to measured data and extrapolated to provide a prediction for higher energies [Gio99a]. Figure 8.1 shows the fits and extrapolations of $\langle n\rangle_{\mathrm{soft}}$, $\langle n\rangle_{\mathrm{semi\text{-}hard}}$, and $\langle n\rangle_{\mathrm{total}}$ as a function of $\sqrt{s}$. α_{soft} can be calculated and falls with increasing $\sqrt{s}$, thus reducing the fraction of the soft component. While it is about 72% at $\sqrt{s} = 900\,\mathrm{GeV}$, it decreases to about 30% at 14 TeV.

For $k_{\mathrm{semi\text{-}hard}}$ three scenarios are proposed in [Gio99a]:

1. $k_{\mathrm{semi\text{-}hard}}$ does not change towards larger energies; this is equivalent to assuming KNO scaling above $\sqrt{s} = 900\,\mathrm{GeV}$. $k_{\mathrm{semi\text{-}hard}}$ stays 13, the value at 900 GeV;

2. the data fit is used to extrapolate $k_{\mathrm{semi\text{-}hard}}$; the authors use the fit ($\sqrt{s}$ in units of 1 GeV):

$$k_{\mathrm{total}}^{-1} = -0.082 + 0.0512 \ln(\sqrt{s}) \quad (8.3)$$

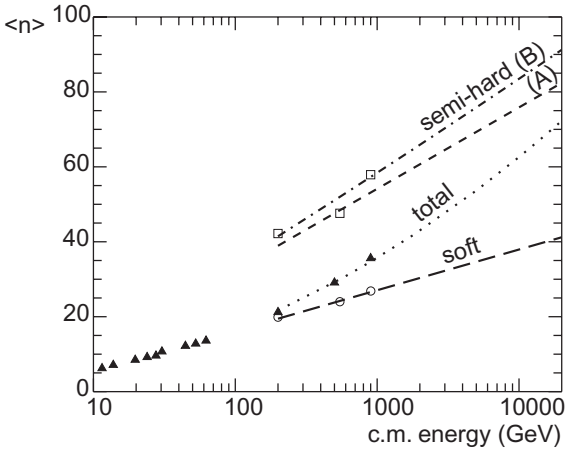

Figure 8.1: The two-component approach with NBDs.

The lines show the fits to the data and the extrapolations: $\langle n\rangle_{\text{total}}$ (dotted), $\langle n\rangle_{\text{soft}}$ (dashed), $\langle n\rangle_{\text{semi-hard}}$ without (A, short-dashed) and with (B, dot-dashed) the second term in Eq. (1.55). The points are data from ISR and Sp$\bar{\text{p}}$S. The figure is taken from [Gio99a].

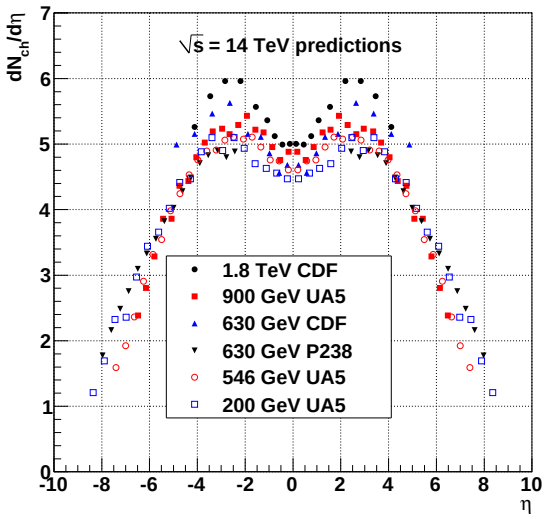

Figure 8.2: Phenomenological prediction for the $dN_{ch}/d\eta$ distribution.

All curves are predictions for $\sqrt{s} = 14\,\text{TeV}$ but based on the scaling of different datasets at different energies (given in the legend). The predictions are from [Bus08].

and relate it to k_{soft} and $k_{\text{semi-hard}}$ by summing the variances of the two components (the NBD's variance is $\sigma^2 = \langle n\rangle + \langle n\rangle^2/k$);

3. $k_{\text{semi-hard}}$ follows perturbative QCD in leading order (assuming that the calculated width of the distribution is also valid in the non-perturbative regime); remaining constants are fitted with the data, which results in ($\sqrt{s}$ in units of 1 GeV):

$$k_{\text{semi-hard}} = 0.38 - 0.42/\sqrt{\ln(\sqrt{s}/10)}. \tag{8.4}$$

These approaches lead to rather different predicted distributions at $\sqrt{s} = 14\,\text{TeV}$, that are included in Figure 8.4 (labeled '2NBD scenarios 1 – 3').

The comparison of this model to data at $\sqrt{s} = 1.8\,\text{TeV}$ from the E735 experiment shows that the second scenario is favored, but there are still clear deviations at higher multiplicities [Gio99a].

By fitting the parameters in limited pseudorapidity intervals the model allows the multiplicity distribution in limited intervals to be predicted [Gio99b]. Predictions for $\sqrt{s} = 10\,\text{TeV}$ in $|\eta| < 1$ are included in Figure 8.5.

8.4 Multiple-Parton Interaction Interpretation

The interpretation of the multiplicity distribution in the framework of multiple-parton interactions, discussed in Section 1.6.4, assumes that the distribution can be decomposed into contributions from one, two, and more parton collisions. In practice this is done for up to three parton collisions. Extrapolation of the cross-sections of the different contributions produces a prediction for higher energies [Wal04]. The predicted multiplicity distribution at $\sqrt{s} = 14\,\text{TeV}$ in full phase space is included in Figure 8.4. Furthermore, the cross-sections $\sigma_{\text{INEL}} = 77.4 \pm 5\,\text{mb}$ and $\sigma_{\text{NSD}} = 64.2 \pm 6\,\text{mb}$ are predicted for $\sqrt{s} = 14\,\text{TeV}$.

8.5 Trends in Multiple-Particle Production

A phenomenological approach is based on trends at lower energies [Bus08]. In particular, it is observed that the width and height of the pseudorapidity distribution grow with $\ln s$ (approximately, see Figure 2.9 on page 58). At the same time the particle density has to go to 0 at beam rapidity and the distribution is trapezoidal-shaped. Consequently, the shape of the distribution is independent of energy, the dependence of the width and height on $\ln s$ allows the extrapolation of *any* distribution to another energy by scaling the vertical and horizontal axes accordingly. This is shown in Figure 8.2 for NSD events at $\sqrt{s} = 14\,\text{TeV}$. The different curves are based on $dN_{ch}/d\eta$ distributions measured at $\sqrt{s} = 0.2 - 1.8\,\text{TeV}$, each scaled to 14 TeV. The resulting spread in the expectation is of the order of 20%. The range of the prediction is included in Figure 8.3 for comparison.

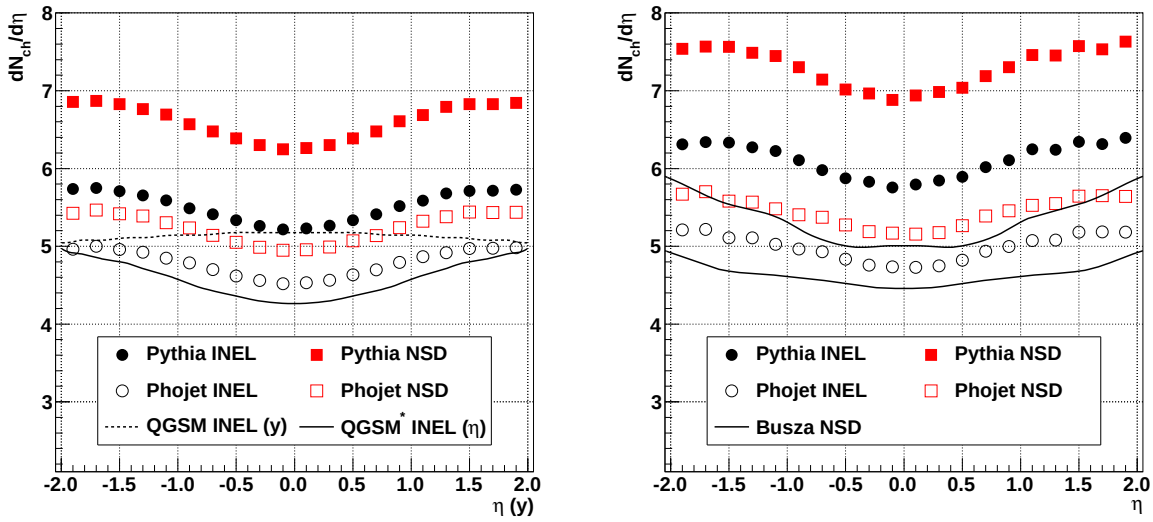

Figure 8.3: Pseudorapidity distribution predictions.

Distributions are given at $\sqrt{s} = 10\,\text{TeV}$ (left panel) and 14 TeV (right panel). The QGSM prediction [Kai99] is shown as a function of y and transformed to η (see Section 8.1). The two curves in the right panel are the upper and lower limits of the prediction by [Bus08], see Figure 8.2.

8.6 Summary

Figure 8.3 shows predictions for the pseudorapidity distributions at $\sqrt{s} = 10\,\text{TeV}$ (left panel) and 14 TeV (right panel). QGSM calculations (Section 8.1), Pythia and Phojet simulation results (Section 8.2), and the phenomenological approach which scales distributions from lower energies (Section 8.5) are shown. The QGSM result agrees with Phojet at large rapidities but not at mid-rapidity. This might be due to the conversion from y to η described in Section 8.1. The event-generator predictions differ by 13% and 20% for the inelastic and NSD event samples, respectively. This difference is 2 – 4 times larger than the systematic uncertainties derived in Chapter 6. The first measurement of the pseudorapidity distribution will therefore allow an assessment which event generator will provide the better description of the data.

Figure 8.4 summarizes predictions of the multiplicity distribution for NSD events in full phase space at $\sqrt{s} = 14\,\text{TeV}$. Results from Pythia, Phojet, QGSM, three scenarios of the two NBD approach, and the multiple-parton interpretation are shown. The distribution in full phase space cannot be measured with ALICE but allows the predictions mentioned in this chapter to be compared.

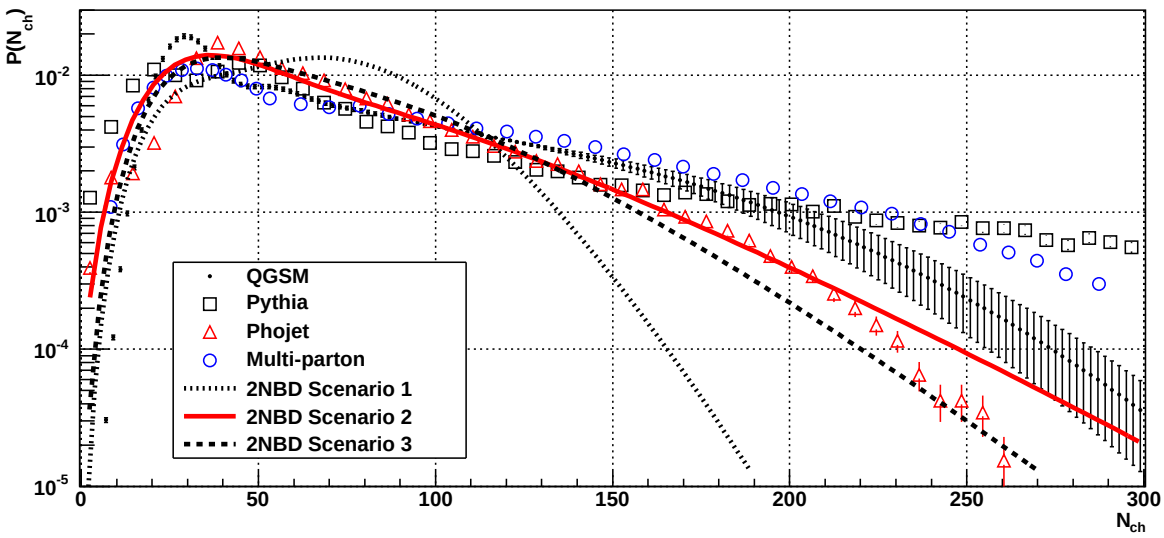

Figure 8.4: Multiplicity distribution predictions in full phase space, $\sqrt{s} = 14\,\text{TeV}$, NSD. The two-component model predictions using the two NBD parameterizations are from [Gio99a]. The QGSM values are from [Pog08]. The prediction following the multiple-parton interpretation is taken from [Wal04].

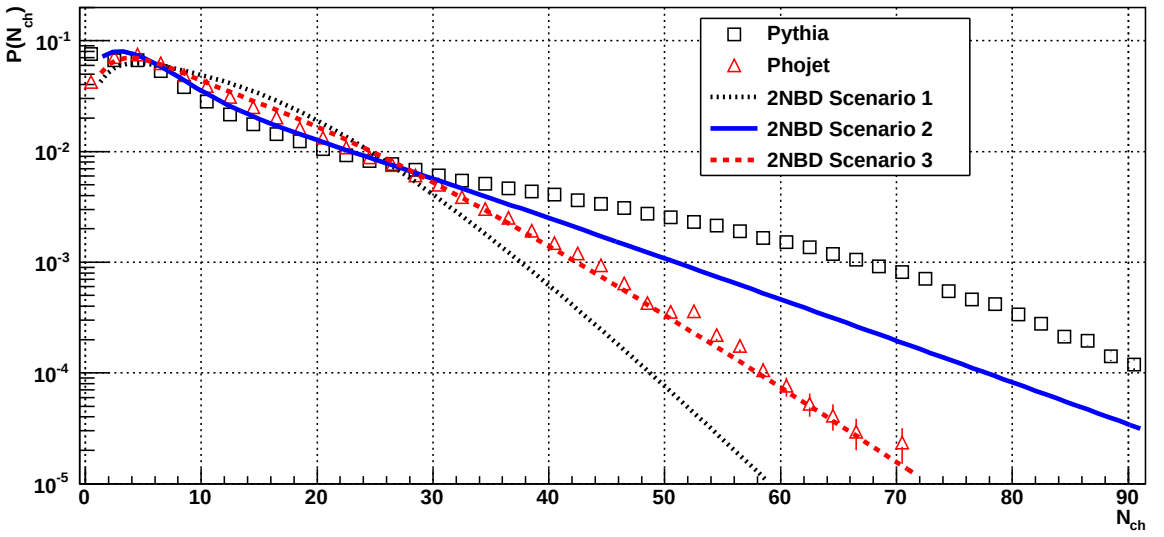

Figure 8.5: Multiplicity distribution predictions in $|\eta| < 1$ at $\sqrt{s} = 10\,\text{TeV}$ (NSD). The two-component model predictions using the two NBD parameterizations are from [Gio99b].

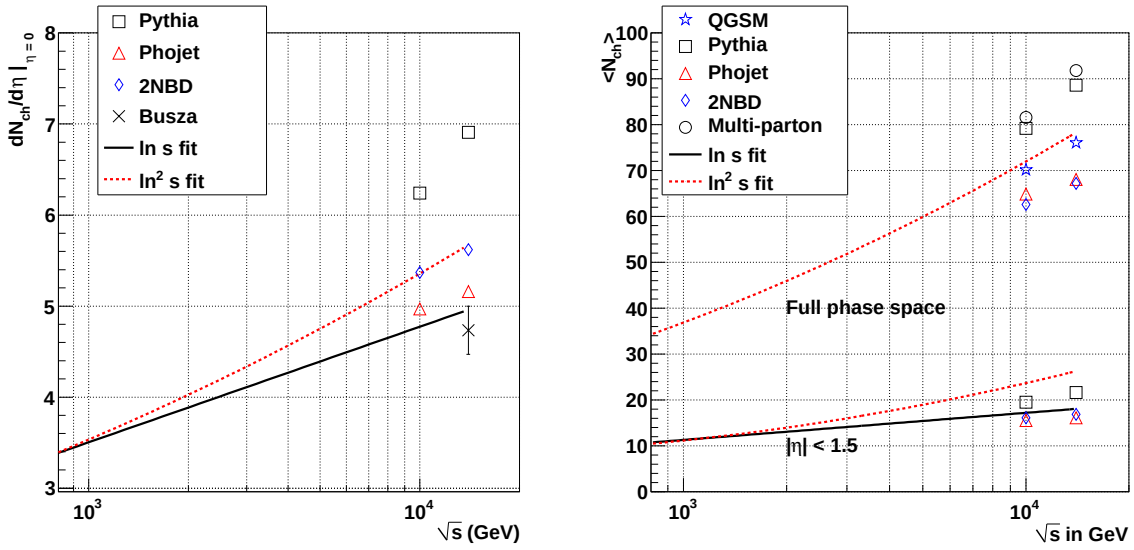

Figure 8.6: Predictions for $dN_{ch}/d\eta|_{\eta=0}$ and $\langle N_{ch}\rangle$.

Predictions for NSD events at $\sqrt{s} = 10\,\text{TeV}$ and $14\,\text{TeV}$ are shown along with the extrapolations using data from previous experiments (see Figures 2.9 and 2.10 on page 58). The left panel shows $dN_{ch}/d\eta|_{\eta=0}$ and the right panel $\langle N_{ch}\rangle$ in full phase space and $|\eta| < 1.5$ [Kai99, Gio99a, Gio99b, Wal04, Bus08, Pog08].

Figure 8.5 shows predictions for the distribution accessible to ALICE in a limited range of $|\eta| < 1$ for NSD events at $\sqrt{s} = 10\,\text{TeV}$. Results using Pythia and Phojet, as well as the three scenarios of the two NBD parametrization are shown. The distributions differ by about a factor 2 at multiplicities of 10 to 20 and by more than an order of magnitude starting from a multiplicity around 45. The systematic uncertainty deduced in Chapter 7 associated with the first measurement is less than 18% for NSD events for non-zero multiplicities. Thus the first measurement will be able to clarify which prediction matches the data best or to conclude that no prediction properly reproduces the data.

Figure 8.6 presents $dN_{ch}/d\eta|_{\eta=0}$ (left panel) and $\langle N_{ch}\rangle$ (right panel) as a function of $\sqrt{s}$ together with extrapolations of the functional fits found in Chapter 2. Only the fits that reproduce the data are shown. The predictions differ up to 30% at mid-rapidity and about 15% for the average multiplicities in full phase space.

Summary

In 2008, the Large Hadron Collider at CERN commenced operation. For the duration of its operation, it will provide proton and heavy-ion collisions at energies much larger than those achieved by previous accelerators. An outstanding and ambitious physics program is at its beginning and will advance human knowledge in the field of particle physics and beyond.

The physics programs of all LHC experiments begin with basic measurements to characterize collisions at LHC energies. For this purpose multiplicity and momentum distributions, particle abundances, and correlations are measured. Subsequently, the obtained basic understanding allows the analysis of rare signals and new physics: the centerpieces of the physics program.

This thesis has introduced the theoretical framework to describe multiple-particle production. The functioning of two event generators, Pythia and Phojet, as well as theoretical descriptions of the charged-particle multiplicity have been discussed. A summary of pseudorapidity-density ($dN_{ch}/d\eta$) and multiplicity-distribution measurements of charged particles has been presented. Existing results have been shown in an energy range of $\sqrt{s} = 6\,\text{GeV}$ to $1.8\,\text{TeV}$ from bubble chamber experiments and detectors at the ISR, Sp$\bar{\text{p}}$S, and Tevatron. The validity of the introduced models was reviewed and the behavior as function of $\sqrt{s}$ was discussed. Analysis procedures for two basic measurements with ALICE, the pseudorapidity density and the multiplicity distribution of charged particles, have been developed. The former allows corrections on a bin-by-bin basis, while the latter requires unfolding of the measured distribution. The procedures have been developed for two independent subdetectors of ALICE, the Silicon Pixel Detector (SPD) and the Time-Projection Chamber (TPC). This allows the comparison of the analysis result in the overlapping regions as an independent cross-check of the measured distribution. Their implementation successfully reproduces different assumed spectra. The procedures have been extensively tested on simulated data using two different event generators, Pythia and Phojet. A comprehensive list of systematic uncertainties was evaluated. Some of these uncertainties still require measured data to verify or extract their magnitude.

To obtain the pseudorapidity density the analysis corrects the measured data for tracking efficiency, acceptance, secondaries, and the low-momentum cut-off as well as the bias introduced by the vertex reconstruction and the trigger. Systematic uncertainties related to the incomplete knowledge about the physics in the new energy regime (cross-sections, particle abundances, and spectra) have been studied. Furthermore, experimental effects related to LHC conditions and effects of the ALICE configuration, such as beam-gas and pile-up events, track selection cuts, as well as uncertainties in

the material budget and the detector alignment were quantitatively evaluated. The total uncertainty, as it is expected before the start of data-taking, is for the $dN_{ch}/d\eta$ distribution of inelastic events 2.3% (4.3%) for the measurement with the SPD (TPC). For NSD events the uncertainty is 8.1% and 8.8% for the SPD and TPC, respectively.

Two methods have been introduced to unfold the measured multiplicity distribution: one based on a χ^2-minimization procedure, the other based on Bayes' theorem. Both reproduce the spectra successfully, independently of the assumed multiplicity distribution. This permits the unfolded distribution to be cross-checked. Furthermore, the multiplicity distribution is corrected for the bias introduced by the vertex reconstruction and the trigger. The previously given systematic uncertainties were also evaluated for the multiplicity distribution measurement with the SPD and amount for non-zero multiplicities to 2 – 9% (2 – 18%) for inelastic (NSD) events.

A road map has been presented for the measurement of the $dN_{ch}/d\eta$ and multiplicity distributions. It contains the steps needed to obtain the corrected distributions from measured data. This includes the verification of the trigger, the comparison of properties of simulated and measured data, and the production of various corrected distributions under different conditions that should produce consistent results.

Model, event generator, and phenomenological predictions for pseudorapidity density and multiplicity distributions were summarized for LHC energies. These are significantly different; the $dN_{ch}/d\eta$ distributions differ by about 13 – 20% and the multiplicity distributions are different by a factor of 2 up to more than an order of magnitude. Therefore the first measurements outlined in this thesis (with their associated systematic uncertainties) will already be able to differentiate between them.

In conclusion, ALICE being ready for data-taking, is also ready for first physics results. Two important measurements with LHC collisions can be performed on a very short time scale of days. The presented analysis procedures developed within this thesis including the already performed systematic studies will pave the way for a speedy publication of the results. Furthermore, the introduced analysis concepts can be used as a general recipe for studies in the field of minimum-bias physics in ALICE and beyond.

Zusammenfassung

Im Jahr 2008 nahm der Large Hadron Collider (großer Hadronenbeschleuniger) am CERN seinen Betrieb auf. Während der Laufzeit werden Protonen- und Schwerionenkollisionen, bei höheren Energien als je zuvor an Teilchenbeschleunigern erreicht, für Experimente zur Verfügung gestellt. Ein außergewöhnliches und ehrgeiziges Forschungsprogramm steht in den Startlöchern, welches das menschliche Wissen im Bereich der Teilchenphysik und darüber hinaus erweitern wird.

Die Forschungsprogramme aller Experimente am LHC beginnen mit grundlegenden Messungen zur Charakterisierung von Kollisionen bei LHC-Energien. Zu diesem Zweck werden Multiplizitäts- und Impulsverteilungen, Teilchenhäufigkeiten und Korrelationen gemessen. Anschließend erlaubt das gewonnene grundlegende Verständnis die Analyse von seltenen Signalen und Physik jenseits des Standardmodells, welche im Mittelpunkt der Forschungsprogramme steht.

Die vorliegende Arbeit hat theoretische Grundlagen zur Beschreibung von Mehrteilchenproduktion vorgestellt. Die Funktionsweise von zwei Ereignisgeneratoren, Pythia und Phojet, sowie theoretische Beschreibungen der Multiplizitätsverteilungen von geladenen Teilchen wurden diskutiert.

Umfassende Messergebnisse von früheren Experimenten zu Pseudorapiditätsdichte ($dN_{ch}/d\eta$) und Multiplizitätsverteilung von geladenen Teilchen wurden vorgestellt. Dies beinhaltete Ergebnisse von Blasenkammerexperimenten sowie Detektoren am ISR, Sp$\bar{\text{p}}$S und Tevatron bei Schwerpunktsenergien von $\sqrt{s} = 6\,\text{GeV}$ bis $1.8\,\text{TeV}$. Die Gültigkeit der vorgestellten Modelle wurde überprüft und das Verhalten der Verteilungen als Funktion der Schwerpunktsenergie diskutiert. Analyseverfahren für die Messung der Pseudorapiditätsdichte sowie der Multiplizitätsverteilung von geladenen Teilchen mit ALICE wurden entwickelt. Das Erstere erlaubt Korrekturen auf einer 'bin-für-bin' Basis, während das Letztere die Entfaltung der gemessenen Verteilung erfordert. Die Verfahren sind für zwei unabhängige Subdetektoren von ALICE, für den Silicon Pixel Detector (Silikonpixeldetektor, SPD) sowie die Time-Projection Chamber (Zeitprojektionskammer, TPC) vorgesehen. Dies ermöglicht den Vergleich der Analyseergebnisse in dem sich überlappenden Messbereich (in η) als unabhängige Überprüfung der gemessenen Verteilung. Die Verfahren wurden ausführlich an simulierten Daten zweier verschiedener Ereignisgeneratoren, Pythia und Phojet, getestet und reproduzieren erfolgreich verschiedene angenommene Eingangsspektren. Eine umfassende Liste von systematischen Unsicherheiten wurde evaluiert, wobei jedoch einige erst nach Beginn der Datennahme zusammen mit gemessenen Daten abschließend beurteilt werden können.

Für die Messung der Pseudorapiditätsdichte wird der Einfluss von Trackingeffizienz, Akzeptanz, Sekundärteilchen und der Unempfindlichkeit bei niedrigen Transversalimpulsen sowie von Vertexrekonstruktion und Trigger auf die gemessene Verteilung korrigiert. Systematische Unsicherheiten, verursacht durch die unvollständige Kenntnis von Kollisionen bei der neuen Energie (Wirkungsquerschnitte, Teilchenhäufigkeiten und Spektren), wurden untersucht. Des Weiteren wurde der Einfluss von experimentellen Effekten durch den Beschleuniger und den spezifischen Aufbau des ALICE-Experiments, wie z.B. Kollisionen zwischen Teilchenstrahl und Restgas, überlappenden Ereignissen, Auswahl der rekonstruierten Spuren sowie Unsicherheiten in Bezug auf die Materialdichte und der Detektorjustierung quantitativ ausgewertet. Die gesamte systematische Unsicherheit, abgeschätzt vor Beginn der Datennahme, beträgt für die $dN_{ch}/d\eta$-Verteilung der inelastischen Ereignisse 2.3% (4.3%) bei der Messung mit dem SPD (TPC). Für NSD[2] Ereignisse ergibt sich eine Unsicherheit von 8.1% bei der Messung mit dem SPD bzw. 8.8% mit der TPC.

Zwei Verfahren zur Entfaltung der Multiplizitätsverteilung wurden vorgestellt: Eine auf der Grundlage der Minimierung einer χ^2-Funktion, die andere auf der Basis des Theorems von Bayes. Beide Methoden reproduzieren Spektren erfolgreich unabhängig von der gewählten Eingangsverteilung. Dies erlaubt den Vergleich der Ergebnisse der beiden Methoden. Darüber hinaus werden Effekte der Vertexrekonstruktion und des Triggers auf die Multiplizitätsverteilung korrigiert. Die oben genannten systematischen Unsicherheiten wurden auch in Bezug auf die Multiplizitätsverteilung für die Messung mit dem SPD evaluiert. Für Multiplizitäten größer als 0 ergeben sich Unsicherheiten von 2 – 9% für die Messung von inelastischen Ereignissen bzw. 2 – 18% für die Messung von NSD Ereignissen.

Ein 'Fahrplan' für die Messung der $dN_{ch}/d\eta$- sowie der Multiplizitätsverteilung wurde vorgestellt. Dieser beinhaltet die notwendigen Schritte um die korrigierten Verteilungen aus den gemessenen Daten zu extrahieren. Im Detail umfassen diese die Prüfung des Triggers, den Vergleich von Eigenschaften von simulierten und gemessenen Daten und die Berechnung von verschiedenen korrigierten Verteilungen bei unterschiedlichen Bedingungen, die konsistente Ergebnisse liefern sollten.

Vorhersagen von Modellen, Ereignisgeneratoren und phänomenologischen Beschreibungen für Pseudorapiditätsdichte und Multiplizitätsverteilung bei LHC-Energien wurden zusammengefasst. Diese weichen signifikant voneinander ab: Die $dN_{ch}/d\eta$-Verteilungen unterscheiden sich um ca. 13 – 20% und die Multiplizitätsverteilungen um einen Faktor 2 bis zu mehr als einer Größenordnung. Daher werden bereits die ersten Messungen, selbst mit den damit verbundenen systematischen Unsicherheiten, in der Lage sein zwischen den Vorhersagen zu differenzieren.

[2]NSD = Non single-diffractive = Inelastische, jedoch nicht einfach-diffraktive, Kollisionen.

Zusammenfassend lässt sich sagen, dass ALICE zu diesem Zeitpunkt nicht nur für die Datennahme sondern auch für die ersten Messungen bereit ist. Zwei wichtige Verteilungen von LHC-Kollisionen können in einem kurzen Zeitraum von Tagen gemessen werden. Die in dieser Arbeit entwickelten Analysen einschließlich der bereits durchgeführten systematischen Studien ebnen den Weg für eine rasche Veröffentlichung der Ergebnisse. Des Weiteren können die Konzepte der vorliegenden Arbeit als allgemeine Anleitung für Studien im Bereich der Minimum-Bias-Physik in ALICE und darüber hinaus verwendet werden.

Appendix A

Kinematic Variables

A particle, with the energy E, the rest mass m_0, and the momentum $\vec{p}$, is described by its *four-momentum*[1]

$$P = (E, \vec{p}) = (E, p_x, p_y, p_z). \tag{A.1}$$

So-called *Mandelstam variables* are used to describe $2 \rightarrow 2$ particle reactions. With P_1 and P_2 (P_3 and P_4) being the four-momenta of the incoming (outgoing) particles, the following variables are defined:

$$s = (P_1 + P_2)^2 = (P_3 + P_4)^2, \tag{A.2}$$

$$t = (P_1 - P_3)^2 = (P_2 - P_4)^2, \tag{A.3}$$

$$u = (P_1 - P_4)^2 = (P_2 - P_3)^2. \tag{A.4}$$

$\sqrt{s}$ is the energy of a collision in the center-of-mass of the colliding particles. $\sqrt{t}$ is the momentum transfer in the reaction. It can be shown that

$$s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2. \tag{A.5}$$

Thus $s + t + u = 0$ for massless particles.

In the case of ion collisions, the energy of the collision is given per nucleon pair and denoted with $\sqrt{s_{NN}}$.

The momentum of a particle is divided into its *longitudinal momentum* p_l and *transverse momentum* p_T. Using the coordinate system of ALICE (Appendix B) it is written as (ϑ is the polar angle):

$$p = |\vec{p}| = \sqrt{p_l^2 + p_T^2}, \tag{A.6}$$

$$p_l \overset{\text{(ALICE)}}{=} p \cos\vartheta = p_z, \tag{A.7}$$

$$p_T \overset{\text{(ALICE)}}{=} p \sin\vartheta = \sqrt{p_x^2 + p_y^2}. \tag{A.8}$$

[1] The commonly adopted convention of $\hbar = c = 1$ is used.

Furthermore, the transverse mass is defined by:

$$m_T^2 = m_0^2 + p_T^2. \tag{A.9}$$

The transverse momentum is invariant under Lorentz transformations. On the contrary, the longitudinal momentum is not invariant. The *rapidity* y is defined by:

$$y = \frac{1}{2} \ln \left(\frac{E + p_l}{E - p_l} \right). \tag{A.10}$$

The rapidity y is additive under Lorentz transformation. However, the determination of the rapidity is complicated because E cannot be easily measured without determining the particle type. Thus the *pseudorapidity* η is used for the case of $E \gg m_0$:

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_l}{p - p_l} \right) \overset{\text{(ALICE)}}{=} - \ln \tan \frac{\vartheta}{2}. \tag{A.11}$$

For ultra-relativistic particles the rapidity y approaches the pseudorapidity η.

The following expressions relate the speed of a particle v with the speed of light:

$$\beta = \frac{v}{c}, \tag{A.12}$$

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}}. \tag{A.13}$$

Appendix B

The ALICE Coordinate System

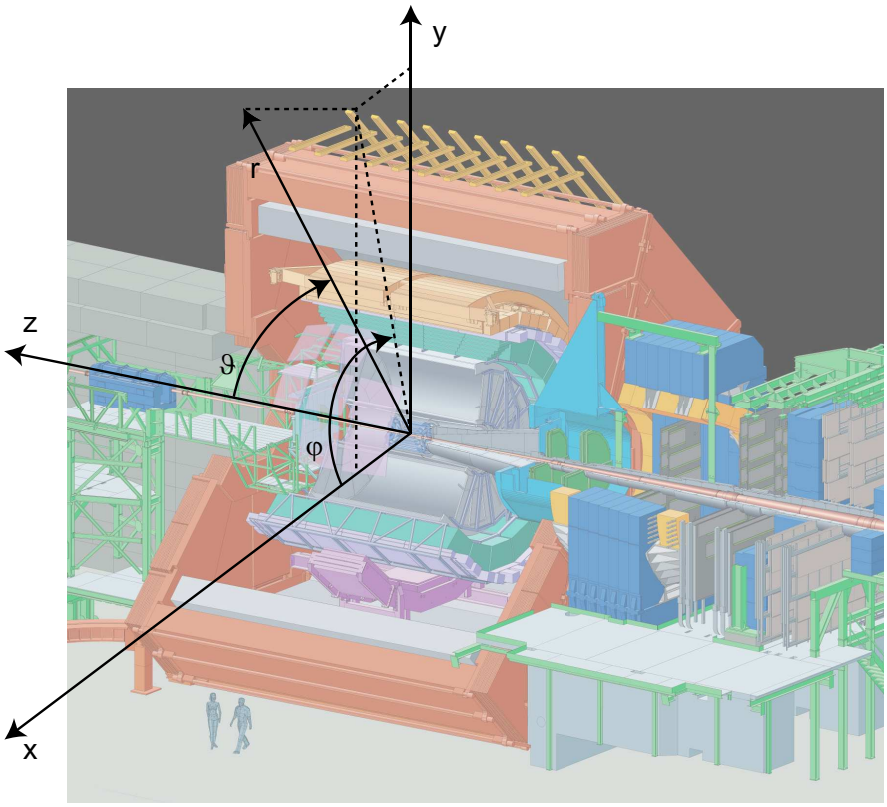

Figure B.1: The ALICE coordinate system.

The coordinate system of ALICE [ALI03], shown in Figure B.1, defines:

- **the point of origin** $x = y = z = 0$ at the nominal interaction point;

- **the *x*-axis** perpendicular to the mean local beam direction, aligned with the local horizontal plane of the LHC and pointing to the accelerator center;
- **the *y*-axis** perpendicular to the x-axis and the mean local beam direction, pointing upward;
- **the *z*-axis** parallel to the mean local beam direction. An observer looking to positive z has the accelerator center on the left. The muon arm is at negative z;
- **the polar angle** ϑ which increases from z ($\vartheta = 0$) passing the x–y plane ($\vartheta = \pi/2$) to $-z$ ($\vartheta = \pi$). At $\vartheta = \pi/2$ the rapidity y is 0; this is called *mid-rapidity*;
- **the azimuthal angle** φ which increases clockwise from x ($\varphi = 0$) passing y ($\varphi = \pi/2$) to x ($\varphi = 2\pi$) with the observer standing at negative z and looking towards the point of origin.

AliRoot characterizes tracks with five parameters y, z, $\sin\varphi$, $\tan\lambda$, and $1/p_T$ with $\lambda = \pi/2 - \vartheta$. These are defined at a position x in a local coordinate system of the track which is rotated by an angle α (in the x–y plane, i.e. the φ direction) with respect to the global coordinate system. While five parameters provide a complete spatial constraint of a track, the covariance matrix (the precision of the track parameters) depends on the current position of the track. Usually, tracks originating from the vertex are *propagated* to the DCA to the vertex. Track cuts, for example, are then applied using values of the covariance matrix at that point. The rotation by α which is in the same direction as φ is used in the track reconstruction. It allows for example all 18 TPC sectors to be treated in the same way by increasing α in steps of $\pi/9$.

Appendix C

Normalized DCA Cut (Nσ-cut)

In Section 5.4.2 a normalized DCA cut is used that considers the resolutions of the distance to the vertex. The cut is such that a certain percentage of primary particles is accepted (particles within a certain number of standard deviations). To achieve the usual meaning with respect to a Gaussian distribution, i.e. that a 1σ-cut accepts 68% of the primary particles, Eq. (5.4) is used. This formula is derived in this section. For this purpose it is assumed that the distribution in the transverse and longitudinal direction is Gaussian and that no correlation exists between their resolutions.

The error function defined by:

$$\mathrm{erf}(x) \;=\; \frac{2}{\pi}\int_0^x \exp(-t^2)dt \tag{C.1}$$

is used.

In the first step the included fraction (in $n\sigma$) of an one-dimensional Gaussian distribution is calculated:

$$\begin{aligned}&\int_{-n\sigma}^{n\sigma}\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma^2}\right)dx\\&=2\int_0^n\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right)dt=\mathrm{erf}\left(\frac{n}{\sqrt{2}}\right).\end{aligned} \tag{C.2}$$

For a two-dimensional Gaussian, the integral runs over $\sqrt{(x/\sigma_1)^2+(y/\sigma_2)^2}\le n$. After the transformation $t=x/\sigma_1$ and $s=y/\sigma_2$

$$\begin{aligned}&\int\frac{1}{\sigma_1\sqrt{2\pi}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)\frac{1}{\sigma_2\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma_2^2}\right)dxdy\\&=\int\frac{1}{2\pi}\exp\left(-\frac{t^2}{2}\right)\exp\left(-\frac{s^2}{2}\right)dtds\end{aligned} \tag{C.3}$$

the integral runs over $\sqrt{t^2+s^2} \leq n$ and can be expressed in polar coordinates ($r^2 = t^2+s^2$). It can be solved by substituting $a = r^2$:

$$\begin{aligned} &= \int_0^{2\pi} \int_0^n \frac{1}{2\pi} \exp\left(-\frac{r^2}{2}\right) r dr d\varphi \\ &= \int_0^n \exp\left(-\frac{r^2}{2}\right) r dr \\ &= 1 - \exp\left(-\frac{n^2}{2}\right). \end{aligned} \tag{C.4}$$

The requirement of the cut is that within a given number of σ the same part of the distribution is included regardless if one Gaussian or a multiplication of two Gaussians is considered. Thus Eqs. (C.2) and (C.4) are combined:

$$\operatorname{erf}\left(\frac{n}{\sqrt{2}}\right) = 1 - \exp\left(-\frac{m^2}{2}\right) \tag{C.5}$$

which leads to

$$m(n) = \sqrt{-2 \ln\left[1 - \operatorname{erf}\left(\frac{n}{\sqrt{2}}\right)\right]} \tag{C.6}$$

and respectively

$$n(m) = \sqrt{2}\, \operatorname{erf}^{-1}\left[1 - \exp\left(-\frac{m^2}{2}\right)\right]. \tag{C.7}$$

The latter is Eq. (5.4) that was to be demonstrated. In written words: if a track has a normalized distance to the vertex of m, it should be included if the $n(m)$ is smaller than the required number of σ.

Appendix D

The Shuttle Framework

During data-taking various subdetectors interact with several online systems (CTP, DAQ, DCS, ECS[1] and HLT). Data is read out by DAQ as raw data streams produced by the subdetectors. At the same time they also produce *conditions data* which is information about the detector status and environmental variables. Most of the conditions data could in principle be calculated from the raw data and extracted offline after data-taking. However, such an approach would require an additional pass over the raw data before the reconstruction which is not possible due to the limited computing resources. Therefore, conditions data is already extracted during data-taking. More details regarding the dataflow in ALICE can be found in Section 4.6.1.

A method is required that reads the conditions data produced by various subdetector algorithms that run in the online systems in a coordinated way. The possibility that each subdetector publishes the produced conditions data by itself is considered too complicated and not manageable. An additional technical issue is that the machines in the online systems producing the data are protected by a firewall from the public CERN network and the internet.

The Shuttle framework that is described in the following performs the following tasks:

- copying of data in any format produced by the online systems DAQ, DCS and HLT for each subdetector;
- preprocessing of the data, e.g. consolidation, fitting;
- reformatting to ROOT format;
- storing in the Grid Offline Conditions DataBase (OCDB);
- indicating that a given run has been processed, which is a precondition to starting the reconstruction.

The Shuttle development started in 2005 as a summer student project and has been continued by the 'ALICE offline group'. It is part of AliRoot.

This chapter describes the structure and implementation of the Shuttle framework. Furthermore, it instructs how to create and test the software class that performs the processing of the data for a given subdetector, the so-called preprocessor. A more extensive description of the Shuttle framework can be found in [Col08].

The Shuttle framework has been running successfully since the first commissioning exercise that started in December 2007. It was operating during the LHC startup in September 2008 and has processed, in total, more than 50 000 runs.

[1] ECS = Experiment Control System.

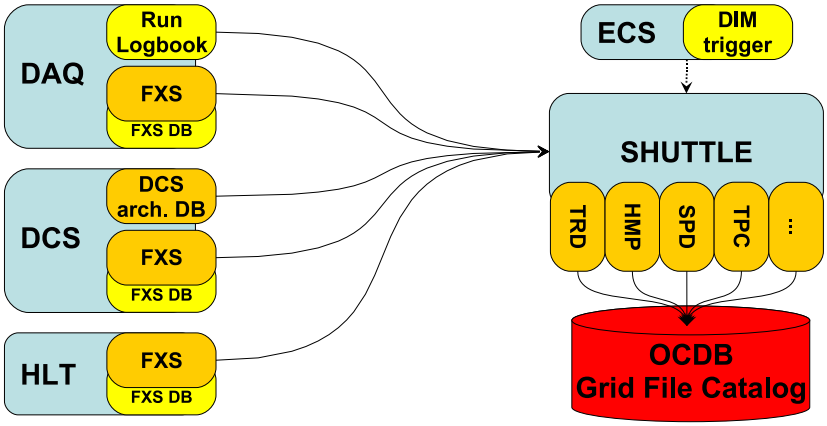

Figure D.1: Schema of the Shuttle framework.

D.1 Structure

The implementation of the Shuttle system involved the ALICE core offline team, the experts of the online systems (DAQ, DCS, ECS and HLT) and the software experts of the subdetectors. Its aim is to provide a common infrastructure to gather the conditions data from the experiment and store it in the OCDB. A general schema of the Shuttle framework and the connections among the main components is presented in Figure D.1.

The core of the Shuttle system has been implemented in AliRoot, ALICE's offline framework for simulation, reconstruction and analysis. Its main features are summarized hereafter; these are discussed in detail in Section D.1.2:

- the Shuttle is triggered upon any end-of-run (EOR) by the ECS system; furthermore, it can be operated in a self-triggered mode;
- it accesses the ECS logbook to read the run parameters (start and end time, run type, etc.) and the list of active subdetectors;
- it accesses the online systems and retrieves the conditions data produced during the run. According to the technique implemented to retrieve them, the experimental conditions data can be divided in two subsets:
  - parameters monitored continuously and archived in the DCS archive are retrieved by means of a dedicated communication protocol;
  - parameters created during data acquisition by the subdetectors' 'detector algorithms' (DAs) are stored to files which are moved to a so-called File eXchange Server (FXS), to which the Shuttle has read access. Each online system provides its own FXS;
- it provides a base class for the implementation of the subdetector-specific code (the 'preprocessors') for the treatment of the conditions data before final storage to the OCDB;
- it accesses the OCDB to store the conditions data and, if needed by the detector preprocessors, to read previously stored conditions data.

Figure D.2 shows the sequence diagram of the Shuttle system. At the end of each run the following actions are performed:

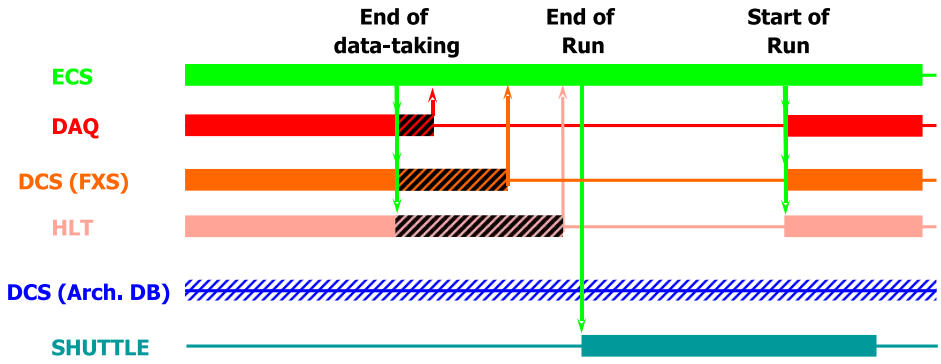

Figure D.2: Sequence diagram.

- at the end of data-taking the ECS informs the other online systems (DAQ, DCS and HLT) that data-taking has stopped. This information is then passed by each online system to the corresponding detector algorithms (DAs);
- the DAs finalize the conditions data observed in the run, store them as files and copy them to the corresponding FXS;
- once all the DAs of a given online system finish processing their data, the online system indicates its readiness back to ECS (i.e. the online system's state machine enters the state *ready for data-taking*);
- once all the online systems are ready, ECS sends an EOR signal to the Shuttle, which performs the following tasks per subdetector:
  - querying the monitored data stored in the DCS archive;
  - running the preprocessors. The Shuttle runs them sequentially, however the different preprocessors are fully independent of each other, which in principle allows parallel processing;
  - retrieving the data from the FXSs requested by the detector preprocessors;
  - storing the conditions data produced by the preprocessors to the OCDB.

  The Shuttle framework monitors the resource consumption of the preprocessors and aborts them if they exceed critical values or if they time out. In the event, of a failure of a preprocessor, it is restarted at a later stage until a certain number of retries is exceeded.

It is important to note that the Shuttle does not interfere with data-taking: as shown in the sequence diagram (Figure D.2), a new run can be started before the processing of the Shuttle finishes. Therefore, the Shuttle does not delay data-taking under any circumstances.

D.1.1 The Shuttle Core Package

A class diagram of the Shuttle is presented in Figure D.3. The steering part of the Shuttle is the class `AliShuttle`. It implements the (pure virtual) interface `AliShuttleInterface`, where the base functions which handle the communication with the detector preprocessors, the online systems and the OCDB are declared. A second implementation of `AliShuttleInterface` is the class `AliTestShuttle` which steers the 'lightweight' *TestShuttle* package. This package allows preprocessors to be

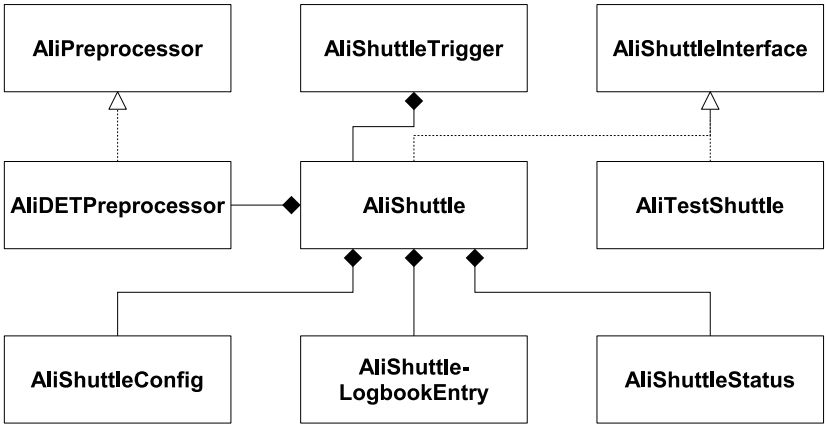

Figure D.3: Class diagram of the Shuttle framework.

tested without the need of external systems that provide data, and is used by the subdetector experts to test their preprocessors in the development phase. It is described in detail in Section D.2.

The `AliShuttleConfig` class contains the Shuttle configuration which is read from an LDAP server. For more details refer to Section D.1.5. The `AliShuttle` class holds an `AliShuttleConfig` instance.

`AliShuttleTrigger` provides the interface to the ECS trigger. It contains an instance of `DATENotifier` which implements the DIM notification service. `AliShuttleTrigger` triggers the collection of the new (unprocessed) runs at each EOR notification from ECS. In the case of timeout waiting for the EOR signal, the Shuttle can run the processing in self-triggered mode; the timeout value is set in the Shuttle configuration. Besides listening to the ECS signals, `AliShuttleTrigger` allows the Shuttle to be run manually to collect the data for a specific run or for all unprocessed runs.

`AliShuttleLogbookEntry` contains the run parameters read from the ECS run logbook for a given run and the processing status of the subdetector preprocessors read from the Shuttle logbook (see Section D.1.2).

`AliShuttleStatus` is used by the Shuttle to keep the status of each preprocessor while iterating over several runs. This information is used for error recovery (see Section D.1.3).

`AliPreprocessor` is the base class of the specific subdetector preprocessor classes, whose implementation is the responsibility of the subdetector experts. The class is named `AliDETPreprocessor` where `DET` is replaced by the subdetector name (see Section D.1.2).

The Shuttle makes use of the AliRoot OCDB access framework to store the conditions data in the OCDB. The framework provides a transparent API to store the data in the format of ROOT files on a local disk or on the Grid.

The client side of the DCS archive communication protocol, implemented in cooperation with the DCS team, is also part of the Shuttle package. It will be described in detail in Section D.1.2.

D.1.2 Basic Components

ECS Logbook

The ECS framework writes the relevant information about data acquisition in a database called 'logbook'. The Shuttle uses the logbook to determine which runs have to be processed. Two tables of this database are accessed by the Shuttle:

- the **run logbook** holds general information about the run. Among its table fields, the most relevant for Shuttle operations are the following:
  - `run`: the run number;
  - `DAQ_time_start`: the time when the run started;
  - `DAQ_time_end`: the time when the run ended;
  - `run_type`: a label identifying the type of run (e.g. 'physics', 'calibration');
- the **shuttle logbook** contains the 'global' processing status of each run as well as the processing status of each of the subdetectors that participated in the run. The ECS fills the table at the EOR, and the Shuttle updates it during processing. The table fields are:
  - `run`: the run number;
  - `shuttle_done`: a 'done' flag that indicates if the Shuttle has processed this run. It is set when all subdetectors have been processed. In this case the status of all preprocessors is either 'INACTIVE', 'DONE' or 'FAILED';
  - `update_time`: automatically set by the database to the time of the last update of the information of this run;
  - a set of 20 '**subdetector**' fields (the 18 ALICE subdetectors plus HLT and GRP), written in the three-letter convention used by the online systems [Bet03]. The Shuttle uses this information to determine which subdetectors participated in the run. The Shuttle updates this information when a given subdetector has been processed and is in a final state which can be either 'DONE' or 'FAILED'. The possible statuses are:
    - INACTIVE: the subdetector was inactive during the run (set by ECS);
    - UNPROCESSED: the subdetector was active during the run and it must be processed (set by ECS);
    - DONE: the Shuttle processed the subdetector successfully (set by the Shuttle);
    - FAILED: the Shuttle failed processing the subdetector (set by the Shuttle);
- the **trigger configuration logbook** contains the CTP configuration.

The information of the first two tables is made available in the Shuttle code through the class `AliShuttleLogbookEntry`. The Shuttle keeps an array of `AliShuttleLogbookEntry` objects, one for each unprocessed run found in the logbook.

Detector Algorithms

A detector algorithm (DA) is a program that runs in one of the online systems and produces conditions data. It publishes the collected data in the FXSs of the online systems which are accessed by the Shuttle. There are no other means of communication between DAs and the Shuttle.

More information about the DAQ DA framework can be found at [DA09]. Unfortunately, at present, there is no public documentation available about the DA frameworks in DCS and HLT.

File Exchange Servers

A file exchange server (FXS) is used as a temporary storage for data produced in a run that is to be picked up by the Shuttle. It is the 'data link' between the DAs and the Shuttle. Each of the three online systems DAQ, DCS and HLT provides a FXS. A FXS consists of a database that contains information about the available data and a storage solution. Data is stored in files, each file being identified by:

Field name	Description
run	Run number
detector	Subdetector name
fileId	File ID
DAQ: DAQsource DCS: – HLT: DDLnumbers	Source (usually machine that produced the file)
filePath	File path on the FXS
time_created	File creation time
time_processed	Processing time by Shuttle
time_deleted	File deletion time
size	File size
fileChecksum	File checksum

Table D.1: FXS MySQL database description.

- a run number (the run in which it was produced);
- the subdetector that produced the file;
- the source, i.e. the producing entity (optional, not used by DCS);
- the file ID, i.e. the identifying name of the file. This ID is set in the DA and is subsequently used by the preprocessor to find the file.

The combination of these four values (run number, subdetector, source, file ID) has to be unique; in a given run and for a given subdetector only one file per file ID and source exists. This is used to uniquely identify the file.

The online systems provide framework functions for DAs running in their sphere to write to the FXS. The Shuttle communicates with the FXS, retrieves the available files and provides them to the preprocessors.

Each online system implements its own FXS and associated database as well as the software for creation, handling and transfer of the conditions parameter files (produced by detector-specific code). In all the implementations a MySQL database is used to store the information about the available files. Table D.1 shows the description of the fields in each system's MySQL database. Successfully retrieved and processed files are flagged by updating the 'time_processed' field. This allows the FXS internal cleanup procedures to operate.

The Shuttle accesses the MySQL database via the ROOT class `TMySQL`; the actual transfer from the FXS to the Shuttle machine is performed with secure copy (scp). To assess transfer consistency, the Shuttle compares the retrieved file size and checksum with the numbers published in the database.

DCS Archive and the AMANDA Protocol

Certain conditions parameters (e.g. device temperatures and gas pressures) are monitored and archived continuously and asynchronously with respect to data acquisition by the DCS system. This data is stored in an Oracle database using the PVSS framework [ALI04]. Each monitored value is identified by a datapoint (DP) name (optionally also an alias name). Each value is associated with a timestamp that contains the exact moment of time when it was stored.

A server-client communication protocol was developed in collaboration with the DCS group with the goal to make these parameters available to the Shuttle. The protocol is described in [Col08, Appendix A]. It describes the communication between a service called AMANDA (that is running in DCS) and the Shuttle. The AMANDA service retrieves conditions data from the archive database and passes it to the Shuttle.

The implementation of the AMANDA service was provided by the DCS group. Details and download instructions are available at [AMA09]. The client has been developed by the offline group. The AMANDA client can be found in the AliRoot SVN in `SHUTTLE/DCSClient`. The folder contains a stand-alone Makefile and can be used without a full build of AliRoot.

The class `AliDCSClient` is used to query the AMANDA server. Its **host** and **port** are given in the constructor, as well as the **timeout** for the request (in ms) and the number of **retries** in case of a failure. The protocol allows several values to be queried in one request; the maximum number of values in one request is indicated by the parameter **multiSplit**:

```
AliDCSClient::AliDCSClient(const char* host, Int_t port,
UInt_t timeout = 5000, Int_t retries = 5, Int_t multiSplit = 100).
```

A single datapoint or alias can be queried with the functions `GetDPValues` or `GetAliasValues`, respectively:

```
Int_t GetDPValues(const char* dpName, UInt_t startTime,
UInt_t endTime, TObjArray* result),
```

```
Int_t GetAliasValues(const char* alias, UInt_t startTime,
UInt_t endTime, TObjArray* result).
```

Its parameters are the datapoint (**dpName**) or alias name (**alias**), respectively, as well as the start (**startTime**) and end time (**endTime**) of the query (as UNIX timestamps) and a pointer to a `TObjArray` (**result**) in which the result is stored. The **result** is filled with an `AliDCSValue` (see below) per retrieved value which have to be deleted by the user. The return value is the number of values retrieved or negative in case of an error.

The class `AliDCSValue` contains a single retrieved value. The function `GetTimeStamp` returns the point in time when the value was stored; `GetType` returns the type of the data; `GetBool`, `GetChar`, `GetInt`, `GetUInt` and `GetFloat` return the data value in the format `Bool_t`, `Char_t`, `Int_t`, `UInt_t` and `Float_t`, respectively.

As previously mentioned, several datapoints can be queried at the same time using the functions `GetDPValues` or `GetAliasValues`, respectively, with a different prototype:

```
TMap* GetDPValues(const TSeqCollection* dpList, UInt_t startTime,
UInt_t endTime, Int_t startIndex = 0, Int_t endIndex = -1),
```

```
TMap* GetAliasValues(const TSeqCollection* aliasList, UInt_t startTime,
UInt_t endTime, Int_t startIndex = 0, Int_t endIndex = -1).
```

The parameters are a `TSeqCollection` of the datapoint (**dpList**) or alias names (**aliasList**) that are to be queried. The list is queried from **startIndex** to **endIndex**, by default the whole list is queried. If the list has more entries than the number of values that can be queried in one request (see the previously mentioned **multiSplit** parameter), the request is subdivided into bunches, where each bunch respects the maximum request size. Furthermore, the start (**startTime**) and end time (**endTime**) of the query have to be passed. The function returns a `TMap` that has to be deleted by the user. The keys of the `TMap` are the datapoint or alias names, the values are `TObjArrays` that contain the data for each datapoint or alias name (as in the case of a single query).

OCDB framework

The Offline Conditions Database (OCDB) is the location where the experimental conditions data is stored. It is not a relational database but a set of entries in the AliEn file catalog that point to physical entities (ROOT files stored in various storage elements of the Grid) containing the conditions data.

The organization of the database is handled by the OCDB access framework, a package included in AliRoot. The OCDB design follows the following principles:

- conditions data stored in ROOT `TObjects` that are stored in ROOT files;
- calibration and alignment objects are run-dependent objects;
- the database is of write-once-read-many (WORM) type. Once an object is stored, it is never removed. However, an object with higher version number can be added (automatic version control of the stored objects);
- the objects in the OCDB are univocally identified by:
  - a (logical) path name (path of the file in the AliEn file catalog);
  - a validity expressed as a run range;
  - a main (Grid) version number;
  - a subversion number, only for locally stored objects.

The full OCDB is stored under a 'base folder', i.e. the logical path of the folder under which the conditions parameter files are stored. The 'official' base folder of the OCDB during data-taking is the following:

`/alice/data/<year>/<LHCPeriod>/OCDB/`.

The Shuttle composes the path to the OCDB by reading the LHC period field from the ECS logbook and determining the year through the time when the run started.

Besides in the OCDB, the Shuttle can store data in a second OCDB-like database, called *Reference DB*. The difference with respect to the OCDB is that reference data is not used for offline raw data reconstruction: the Reference DB is the place to store anything that may be useful for 'manual' debugging of the physics data. For these reasons data that is stored in the OCDB is replicated to the various Grid sites where reconstruction takes place so that the condition files are easily accessible when the reconstruction is performed. On the contrary the Reference DB is not replicated.

The Reference DB base folder path is the following:

`/alice/data/<year>/<LHCPeriod>/Reference/`.

It is also possible to select user-defined OCDB and Reference base folders, by using the static functions in `AliShuttle`:

```
SetMainCDB(const char* uri)
```

and

```
SetMainRefStorage(const char* uri).
```

However, this is not allowed in preprocessors running in the Shuttle.

Preprocessor

The preprocessor contains the specific code of the subdetector that handles the processing of the conditions data. It allows the subdetector experts to query the conditions data, reformat it into ROOT format if needed and store it in the OCDB.

The subdetector preprocessor implements the `AliPreprocessor` class. Preprocessors are registered at run time to the manager class `AliShuttle`. Note that the subdetectors must also be registered in the Shuttle configuration in order to be run by the program.

Each subdetector must implement at most one preprocessor. The preprocessors are distinguished by the name which they declare in their constructor to the base class `AliPreprocessor`. The name follows the three-letter online convention [Bet03]. Besides the 18 subdetectors of the ALICE experiment, two further preprocessors exist: one to retrieve HLT specific parameters and another for data not specific to a particular subdetector but to the whole experiment, called GRP (Global Run Parameters) preprocessor.

A subdetector preprocessor implements the following methods:

- `void Initialize(Int_t run, UInt_t startTime, UInt_t endTime)`:
  `Initialize` is called by the Shuttle before the actual processing. It may be implemented to initialize the preprocessor (e.g. for histogram booking and array initialization). The current run as well as the start (**startTime**) and end time (**endTime**) of the **run** are passed to the function. These values are also available by accessing the member variables `fRun`, `fStartTime`, and `fEndTime` of `AliPreprocessor`;
- `Bool_t ProcessDCS()`:
  `ProcessDCS` is called by the Shuttle before the query to the DCS archive. It may be implemented to skip DCS data retrieval in case it is not needed for particular run types (e.g. calibration runs). By default it returns kTRUE, i.e. the DCS values are queried;
- `UInt_t Process(TMap* dcsMap)`:
  The `Process` function contains the actual preprocessor algorithm. It is a virtual function in the base class, thus it must be implemented. The parameter **dcsMap** contains the DCS archive data that is queried by the Shuttle before the execution of the function. This function performs the preprocessor tasks (retrieval of data from the FXS, data formatting, storage). The function must return 0 in case of success and a non-zero positive number otherwise.

Besides these functions, the `AliPreprocessor` class provides a set of helper functions for retrieval and storage of the data. Most of these functions are delegated to the class `AliShuttle`:

- `void AddRunType(const char* runType)`:
  adds the run type **runType** to the list of run types that the preprocessor wants to process. The Shuttle will call the preprocessor only for the run types that are given in this list. This function is only allowed to be used in the constructor of the preprocessor;
- `TList* GetFileSources(Int_t system, const char* id)`:
  accesses the FXS of the given system and retrieves the list of available online sources that have stored files. The source names in the `TList` are wrapped into `TObjString`s. The parameters are the online **system** where to connect to (0=DAQ, 1=DCS, 2=HLT, constants defined in `AliShuttleInterface`) and the ID (**id**) of the file that is being searched. With this information the Shuttle queries the database of the FXS. The current detector and run number (further two parameters that are needed to find a file on a FXS) do not have to be specified because they are known by the Shuttle that executes the preprocessor. It is important to consider the two possible return values of this function:
  - 0 (**null pointer**) in case of failure connecting the FXS database;
  - a **valid but empty list** in case the query to the database gave no results, i.e. if no file is registered with the given file ID.

  The action to be taken by the preprocessor in case of an empty list depends on the behavior expected by the correspondent online DA: if it is expected to produce files in every run, then the

preprocessor should return an error if no file is found in the FXS; if on the contrary the DA may or may not produce files (for instance in case of insufficient statistics), then the preprocessor should not fail and continue the processing. The returned `TList` has to be deleted by the user.

Note that this function is only needed for the retrieval of files from DAQ and HLT. The DCS system does not allow different sources of files, here the function `GetFile` is used directly;

- `const char* GetFile(Int_t system, const char* id, const char* source)`:
  retrieves a file from the FXS. The parameters are the online **system** which to connect to (as in `GetFileSource`), the ID (**id**) of the file that is being searched, and its **source**. This function is used together with `GetFileSources`, which provides the source names. When this function is called, the Shuttle first queries the FXS database to identify the file in the FXS, then it copies it to the local disk. If the copy is successful it returns the local file name to the preprocessor, otherwise it returns a null pointer;

- `const char* GetRunParameter(const char* param)`:
  queries the run logbook for the given parameter (**param**), e.g. the trigger mask. If the parameter is not found in the logbook a null pointer is returned;

- `AliCDBEntry* GetFromOCDB(const char* pathLevel2, const char* pathLevel3)`:
  retrieves the most recent object '`DET/pathLevel2/pathLevel3`' from the current run from the OCDB where DET is the 'offline' name of the subdetector (e.g. HMP is 'translated' to HMPID). The retrieved object is returned and has to be deleted by the user. In case of failure the function returns a null pointer;

- `const char* GetRunType()`:
  returns the type of the current run, read from the run logbook. If the subdetector implements different DAs in different run types (e.g. 'physics', 'calibration'), the preprocessor can use this function to determine which kind of processing has to be performed;

- `Bool_t GetHLTStatus()`:
  indicates if HLT was participating in the run and therefore if the DAs running in HLT have produced data. The function returns kFALSE if HLT was running in mode A, i.e. HLT was not active in the current run. It returns kTRUE if HLT was running in other modes;

- `Bool_t Store(const char* pathLevel2, const char* pathLevel3, TObject* object, AliCDBMetaData* metaData, Int_t validityStart, Bool_t validityInfinite)`:
  stores the **object** with its OCDB metadata (**metaData**) into the OCDB. The object's path is '`DET/pathLevel2/pathLevel3`', where DET is the 'offline' name of the subdetector. The other two parameters are:

  - **validityStart**: sets the start of the object's run validity range to **currentRun - validityStart**;
  - **validityInfinite**: sets the end of the object's run validity range to infinity (999999999). If the conditions parameter shall be used to reconstruct the current and the 'next' runs (e.g. until the next calibration run), this flag must be set to kTRUE.

  This function stores the conditions object temporarily on the local disk. If storage is successful the function returns kTRUE. The actual transfer to the OCDB is performed at the end after the preprocessor finished, and only in the case of successful processing (see Section D.1.3);

- `Bool_t StoreReferenceData(const char* pathLevel2, const char* pathLevel3, TObject* object, AliCDBMetaData* metaData`:
  stores the **object** with its OCDB metadata (**metaData**) into the Reference DB. The parameters and behavior of this function are the same as in the previously described function `Store`. The only difference is the absence of the **validityStart** and **validityInfinite** parameters: objects stored in the reference folder can only be valid for the current run;

- `Bool_t StoreReferenceFile(const char* localFileName, const char* gridFileName)`: is used to copy non-ROOT files to the Grid or if the file cannot be stored using the OCDB access framework (e.g. due to its size). The file given by **localFileName** is copied into the following path: `<baseGridReferenceFolder>/DET/<runNumber>_<gridFileName>`, where DET is the subdetectors 'offline' name and `runNumber` is the current run number. The function returns kTRUE if the file was copied successfully. Like the previous two functions, the file is first stored locally and transferred after the preprocessor has finished;
- `void Log(const char* message)`: writes a **message** in the subdetector's log file (which is available from the monitoring page).

Two further functions, needed by the GRP preprocessor to manage the transfer of the merged raw tag files to the Grid, have been added to `AliPreprocessor`. Only the GRP preprocessor can use them:

- `TList* GetFileIDs(Int_t system, const char* source)`: is the counterpart of `GetFileSources`: it retrieves the list of file IDs saved by a given **source** on the given **system**. It is needed in case the source may create a variable number of files with different file IDs;
- `Bool_t StoreRunMetadataFile(const char* localFileName, const char* gridFileName)`: is used to store the merged raw data tags given by **localFileName** to the raw data folder: `/alice/data/<year>/<lhcPeriod>/<run>/raw/gridFileName`, where the year and LHC period are read from the run logbook;
- `const char* GetTriggerConfiguration()`: returns the trigger configuration that is stored in the DAQ logbook.

D.1.3 Shuttle Status and Error Recovering

The Shuttle stores the processing status for each run and each subdetector's preprocessor. This information is used for error recovering. Figure D.4 shows the different statuses together with the possible transitions between them. The following statuses and transitions exist for the processing of one subdetector for one run:

1. the Shuttle starts processing the subdetector (status: `Started`). First, it is determined if the preprocessor requires the processing of this run. This is done by verifying if the run type of the current run is part of the list of run types defined in the preprocessor. If the processing for this run is not required the status `Skipped` is set for the current subdetector and the processing is finished;
2. the program forks[2]. The preprocessor is run in the child process while the parent monitors the child's progress. This method assures that the parent Shuttle process cannot crash by malfunctioning preprocessor code. Furthermore, possible memory leaks in the child preprocessor do not affect the parent process;
3. if the preprocessor requires data from the DCS archive, the child process retrieves it (status: `DCS Started`). In the event of a failure to retrieve the DCS archive data, the Shuttle sets the `DCS Error` status for the current subdetector and the processing is finished;
4. the Shuttle calls the subdetector's preprocessor (status: `Preprocessor Started`). The possible exit states of the preprocessor are:

[2] A running program that forks is duplicated completely by the operating system. After forking two identical programs run (called parent and child) that usually branch to different tasks.

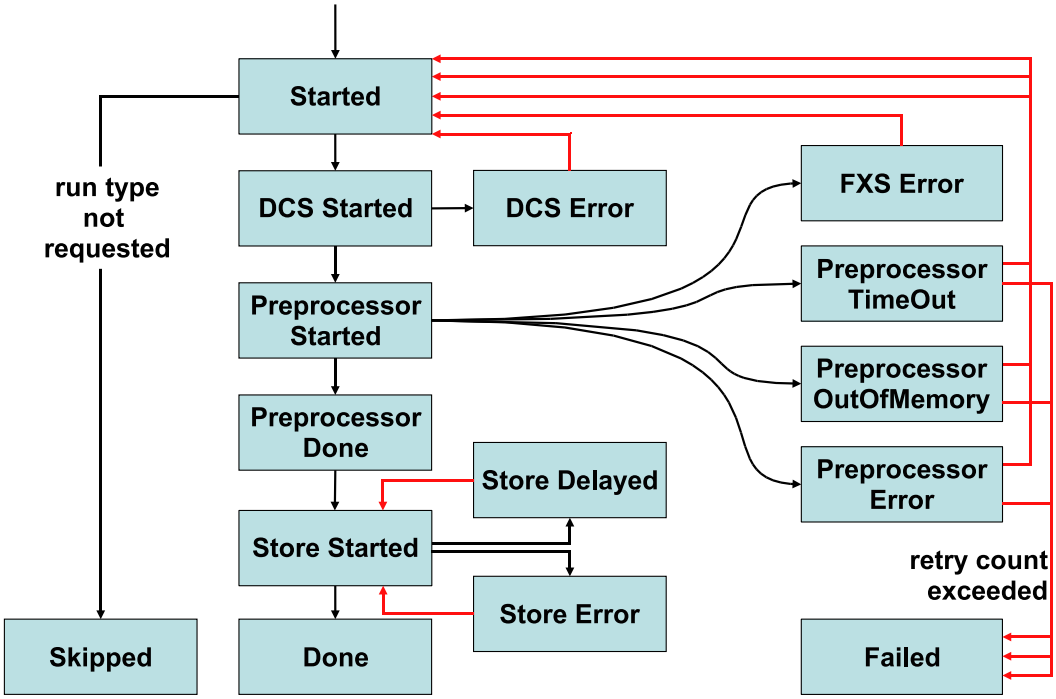

Figure D.4: Preprocessor statuses during the processing.

- `FXS Error`: during the processing a connection to one of the FXSs failed. Therefore, the run is to be reprocessed at a later stage;
- `Preprocessor Error`: the preprocessor failed to process the data for the current run and returned an error code;
- `Preprocessor TimeOut`: the preprocessor exceeded the allowed processing time (timeout set in the configuration);
- `Preprocessor OutOfMemory`: the preprocessor exceeded the allowed memory usage (memory limit set in the configuration);
- `Preprocessor Done`: the processing ended successfully. Conditions data has been stored on the local disc and is ready to be transferred to the Grid.

In case of one of the mentioned error statuses (`FXS Error`, `Preprocessor Error`, `Preprocessor TimeOut` and `Preprocessor OutOfMemory`) or if the child process terminates abnormally (e.g. segmentation violation) the processing is finished;

5. the Shuttle stores the produced data in the OCDB (status: `Store Started`). The following error states may occur:
   - if conditions data is supposed to be stored in the OCDB with infinite validity it is required that all previous runs (i.e. runs with a smaller run number) have been processed for this detector already. If this is not the case the storing of these objects is delayed, the status `Store Delayed` is set and the processing is finished;
   - in case of a failure transferring the conditions data to the OCDB, the status `Store Error` is set and the processing is finished;

6. if storage to the OCDB is successful, the Shuttle sets the `Done` status for the current subdetector and updates the ECS Shuttle logbook.

Upon failure of a preprocessor, its status in the ECS Shuttle logbook remains 'UNPROCESSED'. In the next iteration over the same run the program reads the previous processing exit status. The subsequent action depends on the previous exit status:

- if the status is `DCS Error` or `FXS Error`, the Shuttle restarts the processing without increasing the 'retry number'. In other words, the Shuttle tries to query the data from the DCS archive and the FXSs until the retrieval is successful;
- if the status is one of the possible preprocessor error statuses (`Preprocessor Error`, `Preprocessor TimeOut`, `Preprocessor OutOfMemory`) or `Preprocessor Started`, which means that the preprocessor crashed in the previous iteration, the Shuttle checks the retry number, which is saved together with the exit status. If the number of the allowed retries is exceeded (this parameter is set in configuration) the program declares the processing failed and updates the status and the Shuttle logbook. If the limit is not yet reached, it increases the retry count by one and restarts the processing;
- if the status is `Store Delayed` or `Store Error`, the Shuttle retries the transfer of the conditions data (still available on the local disc) to OCDB, without increasing the retry count.

The processing status and retry count are contained in instances of the class `AliShuttleStatus`. The Shuttle registers the status on the local disc each time it changes. For these operations the program uses the AliRoot OCDB access classes in 'local' mode.

D.1.4 MonALISA Monitoring

The processing status of the Shuttle is monitored using MonALISA [Leg04]. All status changes are sent to a MonALISA service and are visualized by the ALICE MonALISA repository. The monitoring is accessible at

`http://pcalimonitor.cern.ch/show?page=shuttle.jsp&instance=PROD`

The processing status and the history of statuses can be seen for each run and subdetector. Furthermore, the output of the processing (log file) can be accessed, separately by subdetector and run.

Figure D.5 shows an excerpt from the monitoring page. For each run that has been processed by the Shuttle the following information is available: the run type, the period in which the Shuttle has processed the run, the overall Shuttle processing status and the processing status per subdetector[3] which participated in the run (e.g. 'Done (1) h'). The number next to the status of the subdetector indicates the number of retries that have been performed. Clicking on the status opens a window with the log output for the given run and subdetector. Clicking on the 'h' (short for 'history') next to the status opens a window that shows the sequence of status changes for the given run and subdetector (for an example see Figure D.6).

The Shuttle status is linked with the general log file of the processing for the given run. The Shuttle column header is linked with the general (run–independent) Shuttle log file. A click on one of the headers of the subdetector (e.g. TPC) shows only the runs where the given subdetector has participated. A click on the run number (first column) shows the processing statuses as a function of time for the given run (for an example see Figure D.7). This plot can be used to evaluate the processing time spent on each subdetector and each action.

MonALISA also monitors whether the Shuttle is running, and sends alert mails to the people in charge of the service if the Shuttle is not running for more than 15 minutes.

[3] For visibility not all subdetector columns are shown in Figure D.5.

SHUTTLE statistics (current status: **OFFLINE**, processing run: **26059**, unprocessed runs: **?**)

Run#	Run type	First seen	Last seen	SHUTTLE	ACO	EMC	FMD	GRP	HLT	HMP	MCH	MTR
			Last day									
26059	INJECTOR ()	yesterday 16:37	yesterday 16:38	Done h				Done (1) h				
26057	DAQ ()	yesterday 16:32	yesterday 16:37	Done h				Done (1) h				Done (1) h
26056	PULSER ()	yesterday 16:21	yesterday 16:22	Done h				Done (1) h				
26055	PEDESTAL ()	yesterday 16:20	yesterday 16:21	Done h				Done (1) h				
26054	STANDALONE ()	yesterday 16:20	yesterday 16:20	Done h				Done (1) h				
26053	PEDESTAL ()	yesterday 16:19	yesterday 16:20	Done h				Done (1) h				
26052	STANDALONE ()	yesterday 16:19	yesterday 16:32	Done h				Done (1) h				
26051	PEDESTAL ()	yesterday 16:18	yesterday 16:19	Done h				Done (1) h				
26050	STANDALONE ()	yesterday 16:18	yesterday 16:18	Done h				Done (1) h				
26049	DAQ ()	yesterday 16:17	yesterday 16:18	Done h				Done (1) h				Done (1) h
26048	DAQ ()	yesterday 16:06	yesterday 16:07	Done h				Done (1) h				Done (1) h
26047	DAQ ()	yesterday 15:51	yesterday 15:52	Done h				Done (1) h				Done (1) h
26046	DAQ ()	yesterday 15:31	yesterday 15:31	Done h				Done (1) h				Done (1) h
26045	DAQ ()	yesterday 15:21	yesterday 15:21	Done h				Done (1) h				Done (1) h
26042	PHYSICS ()	yesterday 16:05	yesterday 16:32	Done h	Failed (2) h			Done (1) h				
26012	PHYSICS ()	yesterday 16:02	yesterday 16:32	Done h				Done (1) h				

Figure D.5: The main Shuttle monitoring page.

Operational log for MTR, run# 26047

Event time	Status	Count
09.03.2008 15:52	Done	1
09.03.2008 15:52	StoreStarted	1
09.03.2008 15:52	PPDone	1
09.03.2008 15:52	PPStarted	1
09.03.2008 15:51	Started	1

Figure D.6: Status changes of a given subdetector in a given run.

D.1.5 Configuration

The Shuttle reads configuration parameters from an LDAP database. The configuration consists of four LDAP classes:

- the global configuration which contains the main Shuttle running parameters, the ECS logbook address and the MonALISA monitoring page settings (see Table D.2);
- the system configuration which contains the FXS-specific information (FXS plus database addresses - see Table D.3);
- the detector configuration which contains the detector preprocessor base parameters (detector name, responsibles' email addresses – see Table D.4). The Shuttle processes only the subdetectors which are defined in the configuration. This information has to be provided by the subdetector;

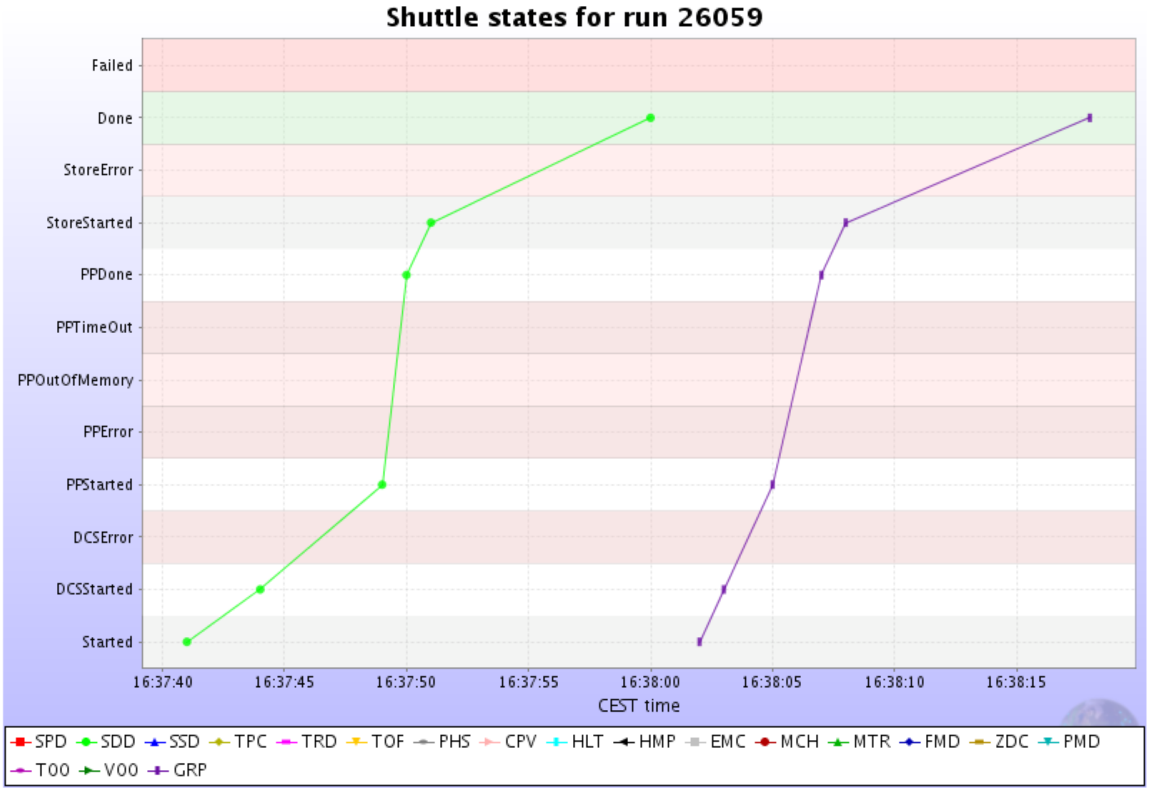

Figure D.7: Processing statuses as a function of time for a given run.

- the DCS configuration which contains the detector-specific settings to access the AMANDA server(s) and the list of DCS data points to be retrieved – see Table D.5.

The Tables D.2 - D.5 contain the name of the configuration fields, a short description, if they allow only a single or multiple values and if they are mandatory. The configuration definition is written in the file `shuttle_schema`, kept in the AliRoot SVN repository in the folder `SHUTTLE/schema` for traceability.

The LDAP server which holds the Shuttle configuration is running on `pcalishuttle01.cern.ch`. Two different configurations have been created, for the 'test setup' and in 'production' mode respectively. The two sets are identified by their 'distinguished name' (dn):

- test mode dn: `o=shuttle,dc=cern,dc=ch`;
- production mode dn: `o=shuttle_prod,dc=cern,dc=ch`.

The two configuration sets are stored in the AliRoot SVN repository, in the subfolders `SHUTTLE/schema` (test configuration) and `SHUTTLE/schema_prod` (production configuration). For each subdetector a file `DET.ldif` contains its configuration, with DET being the 3 letter online name.

Furthermore, a file is used to store passwords needed to access the database servers. The location of the file is given in the global configuration. The file syntax is described in the following paragraph.

Global configuration dn: name=globalConfig,<base-dn>				
n	Name	Description	Single value	Required
1	passwdFilePath	Path to a local file containing database access passwords	*	*
2	daqLbHost	DAQ logbook host	*	*
3	daqLbUser	DAQ logbook user	*	*
4	daqLbDB	DAQ logbook database name	*	*
5	daqLbTable	Run table name	*	*
6	shuttleLbTable	Shuttle's table name	*	*
7	runTypeLbTable	Run type table name	*	*
8	ppMaxRetries	Number of retries before Shuttle declares processing failed	*	*
9	ppTimeOut	Preprocessor timeout (in seconds)	*	*
10	ppMaxMem	Maximum memory consumption allowed during processing (in bytes)	*	*
11	dcsTimeOut	Timeout for DCS AMANDA queries (in seconds)	*	*
12	nDCSretries	Number of retries of DCS AMANDA queries	*	*
13	monitorHost	Monitoring server host	*	*
14	monitorTable	Monitoring server identifier	*	*
15	daqLbPort	DAQ logbook port	*	
16	triggerWait	Timeout before self-triggered processing (in seconds)	*	
17	mode	Running mode (test/prod)	*	
18	keepDCSMap	Keep DCS map flag (1/0)	*	
19	keepTempFolder	Keep temp folder flag (1/0)	*	
20	shuttleAdmin	Shuttle administrator(s) email address(es)		
21	amandaAdmin	AMANDA server administrator(s) email address(es)		
22	sendMail	Send mail flag (1/0)	*	

Table D.2: Global configuration schema.

Password File

The passwords needed to access the databases are not stored in LDAP. They are kept in a file on the Shuttle machine. This allows the content of the LDAP configuration to be published without exposing the passwords.

The syntax of each line of the file is:

```
key password
```

where `key` is one out of

- DAQ_LB: the given password is needed to access the DAQ logbook;
- DAQ_DB: the given password is needed to access the DAQ FXS database;
- DCS_DB: the given password is needed to access the DCS FXS database;
- HLT_DB: the given password is needed to access the HLT FXS database;

System configuration dn: system=DAQ/DCS/HLT,<base-dn>				
n	Name	Description	Single value	Required
1	system	System name (DAQ, DCS, HLT)	*	*
2	dbHost	FXS MySQL DB host	*	*
3	dbUser	FXS MySQL DB user	*	*
4	dbName	FXS MySQL DB name	*	*
5	dbTable	FXS MySQL DB table	*	*
6	fxsHost	FXS host	*	*
7	fxsUser	FXS user	*	*
8	dbPort	FXS MySQL DB port	*	
9	fxsPort	FXS port	*	
10	fxsAdmin	FXS administrator(s) email address(es)		

Table D.3: FXS configuration schema.

Detector configuration dn: det=DET,<base-dn>				
n	Name	Description	Single value	Required
1	det	Detector name (3-letter convention [Bet03])	*	*
2	strictRunOrder	Run-ordered processing flag (1/0)	*	*
3	responsible	Responsible(s) email address(es)		

Table D.4: Detector configuration schema.

and `password` is the password in plain text.

The file is typically stored in `$HOME/.shuttle` and should not be world-readable.

Accessing the Shuttle Configuration

The class `AliShuttleConfig` contains the Shuttle configuration. Upon creation the configuration is read from the LDAP server. The passwords are read from the file indicated in the configuration. The LDAP server and the base DN of the desired configuration set is specified in the constructor, e.g.:

```
AliShuttleConfig config
("pcalishuttle01.cern.ch", 389, "", "", "o=shuttle_prod,dc=cern,dc=ch").
```

D.2 The `TestShuttle` Package

A stand-alone *TestShuttle* package was created in order to allow the subdetector experts to implement and test their preprocessors without needing to access the full Shuttle system. Unlike the real Shuttle, the TestShuttle does not access the online systems to fetch the data queried by the preprocessor. Instead it returns local data and files that can be configured by the user which is explained in the following.

DCS configuration dn: dcsHost=<host>,det=DET,<base-dn>				
n	Name	Description	Single value	Required
1	dcsHost	AMANDA server host	*	*
2	dcsPort	AMANDA server port	*	*
3	dcsAlias	alias name		
4	dcsDP	data point name		
5	multiSplit	Number of data points to be queried in parallel	*	

Table D.5: DCS configuration schema.

The TestShuttle package is part of the Shuttle framework in AliRoot (folder: `SHUTTLE/TestShuttle`). It is a stand-alone package that does not need external dependencies to be built (except AliRoot).

The manager of the TestShuttle package is the class `AliTestShuttle`. It implements `AliShuttleInterface` like the 'real' Shuttle. Besides the interface functions it implements a set of input functions to set up the parameters to be returned to the preprocessor:

- `void SetDCSInput(TMap* dcsMap)`:
  sets a map of DCS archive-like data. The TestShuttle will pass it to the preprocessor as parameter of the `Process` function. The data must be of the same format as that which the real Shuttle provides[4];
- `void AddInputFile(Int_t system, const char* detector, const char* id, const char* source, const char* fileName)`:
  sets the input for the functions managing data retrieval from the FXS. The list of file sources and the file name declared by this function will be returned by `GetFileSources` and `GetFile`, respectively;
- `void SetInputRunType(const char* runType)`:
  sets the run type, which is returned by the function `GetRunType()`;
- `void AddInputRunParameter(const char* key, const char* value)`:
  sets the list of ECS logbook-like parameters to be retrieved with the `GetRunParameter` function which will return the 'value' corresponding to the input 'key';
- `Bool_t AddInputCDBEntry(AliCDBEntry* entry)`:
  sets an OCDB entry to be returned by `GetFromOCDB`. The entry is stored locally for this purpose;
- `SetInputHLTStatus(Bool_t status)`:
  sets the HLT status which is returned to the preprocessor by `GetHLTStatus`.

The run number and the run start and stop time are set in the `AliTestShuttle` constructor. The TestShuttle processing is started by the `Process` function.

Besides the manager class, the TestShuttle package provides an example of a preprocessor implementation (`AliTestPreprocessor`) and an example of an analyzer and container class for DCS archive data (`AliTestDataDCS`). Finally, the macro `TestPreprocessor.C` can be run as a 'tutorial' for the package.

[4] A `TMap` filled with alias / datapoint names as keys and `TObjArrays` of `AliDCSValues` as values. For details see the definition of the AMANDA protocol in Section D.1.2.

Bibliography

[Aam08] K. Aamodt *et al.* [ALICE Collaboration], "The ALICE Experiment at the CERN LHC", JINST **3** (2008) S08002.

[Abe88] F. Abe *et al.* [CDF Collaboration], "The CDF Detector: an Overview", Nucl. Instrum. Meth. A **271** (1988) 387.

[Abe90] F. Abe *et al.* [CDF Collaboration], "Pseudorapidity Distributions of Charged Particles Produced in p+$\bar{\text{p}}$ Interactions at $\sqrt{s}$ = 630 GeV and 1800 GeV", Phys. Rev. D **41** (1990) 2330.

[Abr73] V. A. Abramovsky, V. N. Gribov and O. V. Kancheli, "Character of Inclusive Spectra and Fluctuations Produced in Inelastic Processes by Multi - Pomeron Exchange", Yad. Fiz. **18** (1973) 595 [Sov. J. Nucl. Phys. **18** (1974) 308].

[Aco02] D. E. Acosta *et al.* [CDF Collaboration], "Soft and Hard Interactions in p+$\bar{\text{p}}$ Collisions at $\sqrt{s}$ = 1800 GeV and 630 GeV", Phys. Rev. D **65** (2002) 072005.

[Ada03] D. Adamova *et al.* [CERES/NA45 Collaboration], "Enhanced production of low-mass electron pairs in 40-A-GeV Pb Au collisions at the CERN SPS", Phys. Rev. Lett. **91**, 042301 (2003).

[Ada05] J. Adams *et al.* [STAR Collaboration], "Experimental and theoretical challenges in the search for the quark–gluon plasma: the STAR collaboration's critical assessment of the evidence from RHIC collisions", Nucl. Phys. A **757** (2005) 102.

K. Adcox *et al.* [PHENIX Collaboration], "Formation of dense partonic matter in relativistic nucleus–nucleus collisions at RHIC: Experimental evaluation by the PHENIX collaboration", Nucl. Phys. A **757** (2005) 184.

B. B. Back *et al.*, "The PHOBOS perspective on discoveries at RHIC", Nucl. Phys. A **757** (2005) 28.

I. Arsene *et al.* [BRAHMS Collaboration], "Quark–Gluon Plasma and Color Glass Condensate at RHIC? The perspective from the BRAHMS experiment", Nucl. Phys. A **757** (2005) 1.

[Adl05] C. Adler *et al.* [ALICE Collaboration], "Electron / Pion Identification with ALICE TRD Prototypes using a Neural Network Algorithm", Nucl. Instrum. Meth. A **552** (2005) 364.

[Ago95] G. D'Agostini, "A Multidimensional unfolding method based on Bayes' theorem", Nucl. Instrum. Meth. A **362** (1995) 487.

[Ago99] G. D'Agostini, "Bayesian reasoning in high-energy physics: Principles and applications", CERN Yellow Report, CERN-YELLOW-99-03 (1999).

[Ago03] S. Agostinelli *et al.* [GEANT4 Collaboration], "GEANT4: a Simulation Toolkit", Nucl. Instrum. Meth. A **506** (2003) 250.

[Ala78] A. Alan *et al.* [UA1 Collaboration], "A 4π solid angle detector for the SPS used as a proton-antiproton collider at a centre of mass energy of 540 GeV", CERN-SPSC-78-06 (1978), *http://cdsweb.cern.ch/record/319371.*

[Alb90] C. Albajar *et al.* [UA1 Collaboration], "A Study of the General Characteristics of p+$\bar{\text{p}}$ Collisions at $\sqrt{s}$ = 0.2 TeV to 0.9 TeV", Nucl. Phys. B **335** (1990) 261.

[Ale98] T. Alexopoulos *et al.*, "The role of double parton collisions in soft hadron interactions", Phys. Lett. B **435** (1998) 453.

[Ale06] B. Alessandro *et al.* [ALICE Collaboration], "ALICE: Physics Performance Report, Volume II", J. Phys. G **32** (2006) 1295.

[Alf03] M. G. Alford, "QCD at high density / temperature", Nucl. Phys. Proc. Suppl. **117** (2003) 65.

[ALI95] ALICE Collaboration, "ALICE Technical Proposal for A Large Ion Collider Experiment at the CERN LHC", CERN/LHCC 95-71 (1995).

[ALI99b] ALICE Collaboration, "ALICE Technical Design Report of the Photon Spectrometer (PHOS)", CERN/LHCC 99-4 (1999), *https://edms.cern.ch/document/398934/1*

[ALI00] ALICE Collaboration, "ALICE Technical Design Report of the Time-Projection Chamber", CERN/LHCC 2000-001 (2000), *https://edms.cern.ch/document/398930/1.*

[ALI03] ALICE Collaboration, "Definition of the ALICE Coordinate System and Basic Rules for Sub-detector Components Numbering", ALICE-INT-2003-038 (2003), *https://edms.cern.ch/document/406391/2*.

[ALI04] ALICE Collaboration, "ALICE Technical Design Report of the Trigger, Data Acquisition, High-Level Trigger, Control System", CERN/LHCC 2003/062 (2004), *https://edms.cern.ch/document/456354/2*.

[ALI08] From ALICE Monitoring with MonALISA [Leg04], *http://pcalimonitor.cern.ch* (2008).

[ALI09] *http://aliceinfo.cern.ch/Offline*.

[Aln84] G. J. Alner *et al.* [UA5 Collaboration], "Scaling Violation Favoring High-Multiplicity Events at 540 GeV CMS Energy", Phys. Lett. B **138** (1984) 304.

[Aln85] G. J. Alner *et al.* [UA5 Collaboration], "An Investigation of Multiplicity Distributions in Different Pseudorapidity Intervals in $\bar{\mathrm{p}}$+p Reactions at a CMS Energy of 540 GeV", Phys. Lett. B **160** (1985) 193.

[Aln86] G. J. Alner *et al.* [UA5 Collaboration], "p+$\bar{\mathrm{p}}$ Cross-Sections at 200 GeV and 900 GeV Center-of-Mass Energy", Z. Phys. C **32** (1986) 153.

[Aln86a] G. J. Alner *et al.* [UA5 Collaboration], "Scaling Violations in Multiplicity Distributions at 200 GeV and 900 GeV", Phys. Lett. B **167** (1986) 476.

[Aln86b] G. J. Alner *et al.* [UA5 Collaboration], "Scaling of Pseudorapidity Distributions at $\sqrt{s}$ up to 0.9 TeV", Z. Phys. C **33** (1986) 1.

[Aln87] G. J. Alner *et al.* [UA5 Collaboration], "UA5: a general study of proton-antiproton physics at $\sqrt{s} = 546$ GeV", Phys. Rept. **154** (1987) 247.

[Alp82] K. Alpgard *et al.* [UA5 Collaboration], "Comparison of p+$\bar{\mathrm{p}}$ and p+p Interactions at $\sqrt{s} = 53\,\mathrm{GeV}$", Phys. Lett. B **112** (1982) 183.

[Alp83] K. Alpgard *et al.* [UA5 Collaboration], "Forward-Backward Multiplicity Correlations in p+$\bar{\mathrm{p}}$ Collisions at 540 GeV", Phys. Lett. B **123** (1983) 361.

[Alt04] C. Alt *et al.* [The NA49 Collaboration], "Strangeness from 20-A-GeV to 158-A-GeV", J. Phys. G **30** (2004) S119.

[AMA09] *http://alice-project-dcs-amandaserver.web.cern.ch/alice-project-dcs-amandaserver*.

[Ams08] C. Amsler *et al.* [Particle Data Group], "Review of particle physics", Phys. Lett. B **667** (2008) 1.

[And99] E. Andersen *et al.* [WA97 Collaboration], "Strangeness enhancement at mid-rapidity in Pb+Pb collisions at 158-A-GeV/c", Phys. Lett. B **449** (1999) 401.

[And04] A. Andronic *et al.* [ALICE TRD Collaboration], "Electron identification performance with ALICE TRD prototypes", Nucl. Instrum. Meth. A **522** (2004) 40.

[And07] A. Andronic, P. Braun-Munzinger, K. Redlich and J. Stachel, "Statistical hadronization of heavy quarks in ultra-relativistic nucleus nucleus collisions", Nucl. Phys. A **789** (2007) 334.

[Ans89] R. E. Ansorge *et al.* [UA5 Collaboration], "Charged-Particle Multiplicity Distributions at 200 GeV and 900 GeV Center-of-Mass Energy", Z. Phys. C **43** (1989) 357.

[Ant02] F. Antinori *et al.*, "First results on strange baryon production from the NA57 experiment", Nucl. Phys. A **698**, 118 (2002).

[Aok06] Y. Aoki, Z. Fodor, S. D. Katz and K. K. Szabo, "The QCD transition temperature: Results with physical masses in the continuum limit", Phys. Lett. B **643** (2006) 46.

[Arn83] G. Arnison *et al.* [UA1 Collaboration], "Charged-Particle Multiplicity Distributions in Proton Anti-Proton Collisions at 540 GeV Center-of-Mass Energy", Phys. Lett. B **123** (1983) 108.

[Ast85] A. Astbury *et al.*, "The UA1 Calorimeter Trigger", Nucl. Instrum. Meth. A **238** (1985) 288.

[ATL94] ATLAS Collaboration, "ATLAS Technical Proposal", CERN/LHCC 94-43 (1994).

[Bag08] S. Bagnasco *et al.*, "AliEn: ALICE environment on the GRID", J. Phys.: Conf. Ser. 119 Volume 119 (2008) 062012.

[Bal03] M. Ballintijn *et al.*, "The PROOF distributed parallel analysis framework based on ROOT", *Conference Proceedings of Computing in High-Energy and Nuclear Physics (CHEP 03), La Jolla, California*, [arXiv:physics/0306110] (2003).

[Bec03] F. Becattini, M. Gazdzicki, A. Keranen, J. Manninen and R. Stock, "Study of chemical equilibrium in nucleus–nucleus collisions at AGS and SPS energies", Phys. Rev. C **69** (2004) 024905.

[Bet03] L. Betev *et al.*, "Naming and Numbering Convention for ALICE detector part identification – Generic Scheme", ALICE-INT-2003-039 (2003).

[Bil89] P. Billoir, "Progressive track recognition with a Kalman like fitting procedure", Comput. Phys. Commun. **57** (1989) 390.

[Blo84] V. Blobel, "Unfolding Methods in High-Energy Physics Experiments", DESY 84/118 (1984).

[Blo02a] V. Blobel and C. Kleinwort, "A new method for the high-precision alignment of track detectors", published in Durham, Advanced statistical techniques in particle physics, arXiv:hep-ex/0208021 (2002).
See also *http://www.desy.de/~blobel/mptalks.html*.

[Blo02b] V. Blobel, "An unfolding method for high-energy physics experiments", arXiv:hep-ex/0208022 (2002).

[Bom09a] M. Bombara, C. Lazzeroni, R. Lietava, Z.L. Matthews, S. Navin and O. Villalobos Baillie, "Trigger Corrections for the ALICE First Physics Programme", ALICE internal note, to be published in 2009.

[Bom09b] C. Bombonati, A. Dainese, M. Lunardon, S. Moretto, A. Rossi, "ALICE SPD alignment via Millepede with cosmic tracks", ALICE internal note, to be published in 2009.

[Bra89] W. Braunschweig *et al.* [TASSO Collaboration], "Charged Multiplicity Distributions and Correlations in e+ e- Annihilation at PETRA Energies", Z. Phys. C **45** (1989) 193.

[Bra96] P. Braun-Munzinger, J. Stachel, J. P. Wessels and N. Xu, "Thermal and hadrochemical equilibration in nucleus-nucleus collisions at the SPS", Phys. Lett. B **365** (1996) 1.

[Bra99] P. Braun-Munzinger, I. Heppe and J. Stachel, "Chemical equilibration in Pb + Pb collisions at the SPS", Phys. Lett. B **465** (1999) 15.

[Bra03] P. Braun-Munzinger, K. Redlich and J. Stachel, "Particle production in heavy-ion collisions", in Hwa, R.C. and X.-N. Wang (eds.), "Quark–gluon plasma 3", World Scientific Publishing (2004), arXiv:nucl-th/0304013.

[Bre84] A. Breakstone *et al.* [Ames-Bologna-CERN-Dortmund-Heidelberg-Warsaw Collaboration], "Charged Multiplicity Distribution in p+p Interactions at ISR Energies", Phys. Rev. D **30** (1984) 528.

[Bru78] R. Brun, R. Hagelberg, M. Hansroul and J. C. Lassalle, "Geant: Simulation Program for Particle Physics Experiments. User Guide and Reference Manual", CERN-DD-78-2-REV (1978).

[Bru97] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework", Nucl. Instrum. Meth. A **389** (1997) 81.

[Bus95] D. Buskulic *et al.* [ALEPH Collaboration], "Measurements of the charged-particle multiplicity distribution in restricted rapidity intervals", Z. Phys. C **69** (1995) 15.

[Bus08] W. Busza, "Trends in multiparticle production and some 'predictions' for p+p and Pb+Pb collisions at LHC", J. Phys. G **35** (2008) 044040.

[Cab75] N. Cabibbo and G. Parisi, "Exponential Hadronic Spectrum and Quark Liberation", Phys. Lett. B **59** (1975) 67.

[Cap94] A. Capella, U. Sukhatme, C. I. Tan and J. Tran Thanh Van, "Dual parton model", Phys. Rept. **236** (1994) 225.

[Car04] F. Carminati *et al.* [ALICE Collaboration], "ALICE: Physics Performance Report, Volume I", J. Phys. G **30** (2004) 1517.

[Cha08] S. Chapeland, private communication (2008).

[Che06] M. Cheng *et al.*, "The transition temperature in QCD", Phys. Rev. D **74** (2006) 054507.

[CMS94] CMS Collaboration, "The Compact Muon Solenoid Technical Proposal", CERN/LHCC 94-38 (1994).

[Col77] P. D. B. Collins, "An Introduction to Regge Theory & High-Energy Physics", Cambridge University Press (1977).

[Col08] A. Colla and J. F. Grosse-Oetringhaus, "The Shuttle Framework - A system for automatic readout and processing of conditions data", ALICE-INT-2008-011 (2008), *https://edms.cern.ch/document/924807/1.*

[Con05] J. Conrad, J.G. Contreras, and C.E. Jørgensen, "Minimum-Bias Triggers in p+p Collisions with the VZERO and Silicon Pixel Detector", ALICE

internal note, ALICE-INT-2005-025 (2005),
https://edms.cern.ch/document/664986/1.

[DA09] *http://ph-dep-aid.web.cern.ch/ph-dep-aid/DA-framework.*

[Dam07] S. Damjanovic *et al.* [NA60 Collaboration], "NA60 results on the ρ spectral function in In-In collisions", Nucl. Phys. A **783** (2007) 327.

[Del77] M. Della Negra *et al.* [CERN-College de France-Heidelberg-Karlsruhe Collaboration], "Observation of Jet Structure in High-p_T Events at the ISR and the Importance of Parton Transverse Momentum", Nucl. Phys. B **127** (1977) 1.

[Don92] A. Donnachie and P. V. Landshoff, "Total cross-sections", Phys. Lett. B **296** (1992) 227.

[Don04] X. Dong, S. Esumi, P. Sorensen, N. Xu and Z. Xu, "Resonance decay effects on anisotropy parameters", Phys. Lett. B **597** (2004) 328.

[Dur04] J. Durand, O. Barring, S. Ponce and B. Couturier, "CASTOR: Operational issues and new developments", *in Interlaken, Computing in high-energy physics and nuclear physics, 1221-1226* (2004).

[Eng95] R. Engel, "Photoproduction within the two-component dual parton model. 1. Amplitudes and cross-sections", Z. Phys. C **66** (1995) 203.
R. Engel and J. Ranft, "Hadronic photon-photon interactions at high energies", Phys. Rev. D **54** (1996) 4244.

[Eng97] R. Engel, "Hadronic Interactions of Photons at High Energies", Ph.D. Thesis, University of Siegen (1997).

[Fas03] A. Fasso *et al.*, "The physics models of FLUKA: Status and recent development", *Conference Proceedings of Computing in High-Energy and Nuclear Physics (CHEP 03), La Jolla, California*, arXiv:hep-ph/0306267 (2003).

[Fey69] R. P. Feynman, "Very high-energy collisions of hadrons", Phys. Rev. Lett. **23** (1969) 1415.

[Fos04] I. Foster and C. Kesselmann, "The Grid – Blueprint for a New Computing Infrastructure", 2nd edition, Morgan Kaufmann Publishers (2004).

[Fug89] C. Fuglesang, "UA5 Multiplicity Distributions and Fits of Various Functions", published in La Thuile Multipart.Dyn.1989:193-210 (1989).

[Gio86] A. Giovannini and L. Van Hove, "Negative binomial multiplicity distributions in high-energy hadron collisions", Z. Phys. C **30** (1986) 391.

[Gio88] A. Giovannini and L. Van Hove, "Negative binomial properties and clan structure in multiplicity distributions", Acta Phys. Polon. B **19** (1988) 495.

[Gio99a] A. Giovannini and R. Ugoccioni, "Possible scenarios for soft and semi-hard components structure in central hadron hadron collisions in the TeV region", Phys. Rev. D **59** (1999) 094020 [Erratum-ibid. D **69** (2004) 059903].

[Gio99b] A. Giovannini and R. Ugoccioni, "Possible scenarios for soft and semi-hard components structure in central hadron hadron collisions in the TeV region: pseudo-rapidity intervals", Phys. Rev. D **60** (1999) 074027.

[Gio03] A. Giovannini and R. Ugoccioni, "On signals of new physics in global event properties in p+p collisions in the TeV energy domain", Phys. Rev. D **68** (2003) 034009.

[Gri68] V. N. Gribov, "A Reggeon Diagram Technique", Sov. Phys. JETP **26** (1968) 414 [Zh. Eksp. Teor. Fiz. **53** (1967) 654].

[Gro08] J.F. Grosse-Oetringhaus, "The CERN Analysis Facility – A PROOF Cluster for Day-One Physics Analysis", J. Phys.: Conf. Ser. 119 Volume 119 (2008) 072017.

[Gyu90] M. Gyulassy and M. Plumer, "Jet Quenching in Dense Matter", Phys. Lett. B **243** (1990) 432.

[Gyu94] M. Gyulassy and X. N. Wang, "HIJING 1.0: A Monte Carlo program for parton and particle production in high-energy hadronic and nuclear collisions", Comput. Phys. Commun. **83**, 307 (1994).

[Han56] R. Hanbury Brown and R. Q. Twiss, "A test of a new type of stellar interferometer on Sirius", Nature **178** (1956) 1046.

[Har96] J. W. Harris, B. Müller, "The search for the quark–gluon plasma", Ann. Rev. Nucl. Part. Sci. **46** (1996) 71.

[Har97] R. Harr *et al.*, "Pseudorapidity distribution of charged particles in $\bar{\mathrm{p}}$+p collisions at $\sqrt{s} = 630$ GeV", Phys. Lett. B **401** (1997) 176.

[Hei94] U. W. Heinz, "Strangeness production and chemical equilibration in relativistic nuclear collisions", Nucl. Phys. A **566** (1994) 205C.

[Hei00] U. W. Heinz and M. Jacob, "Evidence for a new state of matter: An assessment of the results from the CERN lead beam programme", arXiv:nucl-th/0002042 (2000).

[Hig64] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", Phys. Rev. Lett. **13** (1964) 508.

F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", Phys. Rev. Lett. **13** (1964) 321.

G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, "Global Conservation Laws and Massless Particles", Phys. Rev. Lett. **13** (1964) 585.

[Iva08] M. Ivanov, GSI Darmstadt, A. Kisiel, Ohio State University, private communication (2008).

[Jam75] F. James and M. Roos, "MINUIT: a System for Function Minimization and Analysis of the Parameter Errors and Correlations", Comput. Phys. Commun. **10** (1975) 343.

[Kai99] A. B. Kaidalov, "Soft interactions of hadrons in QCD", Surveys in High-Energy Physics, 13:4, 265 - 330, 1999.

[Kai03] A. B. Kaidalov, "High-energy hadronic interactions (20 years of the quark–gluon string model)", Phys. Atom. Nucl. **66** (2003) 1994.

[Kal08] A. Kalweit, private communication (2008).

[Kob72] Z. Koba, H. B. Nielsen and P. Olesen, "Scaling of Multiplicity Distributions in High-Energy Hadron Collisions", Nucl. Phys. B **40** (1972) 317.

[Kol04] P. F. Kolb and U. W. Heinz, "Hydrodynamic description of ultrarelativistic heavy-ion collisions", in Hwa, R.C. and X.-N. Wang (eds.), "Quark–gluon plasma 3", World Scientific Publishing (2004, arXiv:nucl-th/0305084.

[Kur76] E. A. Kuraev, L. N. Lipatov and V. S. Fadin, "Multi - Reggeon Processes in the Yang-Mills Theory", Sov. Phys. JETP **44** (1976) 443 [Zh. Eksp. Teor. Fiz. **71** (1976) 840].

I. I. Balitsky and L. N. Lipatov, "The Pomeranchuk Singularity in Quantum Chromodynamics", Sov. J. Nucl. Phys. **28** (1978) 822 [Yad. Fiz. **28** (1978) 1597].

L. N. Lipatov, "The Bare Pomeron in Quantum Chromodynamics", Sov. Phys. JETP **63** (1986) 904 [Zh. Eksp. Teor. Fiz. **90** (1986) 1536].

Yu. A. Simonov, "Glueball Regge Trajectories and the Pomeron", Phys. Lett. B **249** (1990) 514.

[Lai00] H. L. Lai *et al.* [CTEQ Collaboration], "Global QCD analysis of parton structure of the nucleon: CTEQ5 parton distributions", Eur. Phys. J. C **12** (2000) 375.

[Leb08] P. Lebrun, "Interim Summary Report on the Analysis of the 19 September 2008 Incident at the LHC", CERN EDMS document no. 973073, ver. 1, *https://edms.cern.ch/document/973073/1* (2008).

[Leg04] I. C. Legrand, H. B. Newman, R. Voicu, C. Cirstoiu, C. Grigoras, M. Toarta and C. Dobre, "MonALISA: An agent based, dynamic service system to monitor, control and optimize Grid based applications", *Conference Proceedings of Computing in High-Energy and Nuclear Physics (CHEP), Interlaken, Switzerland (2004).*

[Let00] J. Letessier and J. Rafelski, "Observing quark-gluon plasma with strange hadrons", Int. J. Mod. Phys. E **9** (2000) 107.

[LHC98] LHCb Collaboration, "LHCb Technical Proposal", CERN/LHCC 98-4 (1998).

[LHC04] LHC Design Report Volume I+II+III, CERN-2004-003-V-1, CERN-2004-003-V-2, CERN-2004-003-V-3 (2004), *http://ab-div.web.cern.ch/ab-div/Publications/LHC-DesignReport.html.*

[LHC05] LHCf Collaboration, "Technical Proposal for the CERN LHCf Experiment", CERN/LHCC 2005-032 (2005).

[LHC08] $\sqrt{s} = 900\,\mathrm{GeV}$: *http://lhc-commissioning.web.cern.ch/lhc-commissioning/phases/450-unsqueezed-collisions.htm*
$\sqrt{s} = 10\,\mathrm{TeV}$: *http://lhc-commissioning.web.cern.ch/lhc-commissioning/phases/pilot-physics-run.htm* (2008).

[Lin92] C. S. Lindsey *et al.* [E735 Collaboration], "Results from E735 at the Tevatron p+$\bar{\mathrm{p}}$ collider with $\sqrt{s} = 1.8$ TeV", Nucl. Phys. A **544** (1992) 343.

[Loh05] E. Lohrmann, "Hochenergiephysik", 5. edition, Vieweg+Teubner Verlag (2005)

[Mat86] T. Matsui and H. Satz, "J/Ψ Suppression by Quark–Gluon Plasma Formation", Phys. Lett. B **178** (1986) 416.

[Mor07] A. Moraes, C. Buttar and I. Dawson, "Prediction for Minimum Bias and the Underlying Event at LHC Energies", Eur. Phys. J. C **50** (2007) 435.

[Per00] D. H. Perkins, "Introduction to High-Energy Physics", 4th edition, Cambride University Press (2000).

[Pet95] T. S. Pettersson (ed.), P. Lefèvre (ed.), "The Large Hadron Collider : conceptual design", CERN-AC-95-05 LHC (1995).

[Pis82] R. D. Pisarski, "Phenomenology of the Chiral Phase Transition", Phys. Lett. B **110** (1982) 155.

[Pog08] M. Poghosyan, private communication (2008).

[Pok00] S. Pokorski, "Gauge Field Theories", Cambridge University Press, 2nd edition (2000).

[Raf82] J. Rafelski and B. Muller, "Strangeness Production in the Quark–Gluon Plasma", Phys. Rev. Lett. **48** (1982) 1066 [Erratum-ibid. **56** (1986) 2334].

[Ram06] L. Ramello *et al.* [NA50 Collaboration], "Quarkonium production and suppression in Pb + Pb and p + A collisions at SPS energies", Nucl. Phys. A **774** (2006) 59.

[Rim93] F. Rimondi *et al.* [CDF Collaboration], "Multiplicity distributions in p+$\bar{\text{p}}$ interactions at $\sqrt{s}$ = 1800 GeV", published in Aspen Multipart. Dyn. 1993:0400-404 (1993).

[Ros04] A. Rossi, "Residual Gas Density Estimations in the LHC Insertion Regions IR1 and IR5 and the Experimental Regions of ATLAS and CMS for Different Beam Operations", LHC Project Report 783 (2004), *http://cdsweb.cern.ch/record/794411.*

[Rus81] J. G. Rushbrooke *et al.* [Bonn-Brussels-Cambridge-CERN-Stockholm Collaboration], "The UA5 Streamer Chamber Experiment at the Sp$\bar{\text{p}}$S Collider", Phys. Scripta **23** (1981) 642.

[Sat90] H. Satz, "Color Screening and Quark Deconfinement in Nuclear Collisions", Adv. Ser. Direct. High-Energy Phys. **6** (1990) 593.

[Sch94] M. Schmelling, "The method of reduced cross entropy: a general approach to unfold probability distributions", Nucl. Instrum. Meth. A **340** (1994) 400.

[Sjo01] T. Sjostrand *et al.*, "High-energy-physics event generation with PYTHIA 6.1", Comput. Phys. Commun. 135 (2001) 238.
T. Sjostrand, L. Lonnblad and S. Mrenna, "PYTHIA 6.2: Physics and manual", arXiv:hep-ph/0108264 (2001).
Version in use by ALICE: V6.214 with "ATLAS tune" [Mor07].

[Sla72] P. Slattery, "Evidence for the Onset of Semiinclusive Scaling in Proton Proton Collisions in the 50 GeV/c - 300 GeV/c Momentum Range", Phys. Rev. Lett. **29** (1972) 1624.

[Tau09] A. Tauro and S. Evrard, "ALICE Shutdown 2008-2009 Coordination", *https://edms.cern.ch/document/946915/1*.

[Ter86] K. A. Ter-Martirosian, "Froissart Type Rise of Cross-Sections and Predictions for Spectra and Multiplicities of Hadrons at Future Accelerators", Sov. J. Nucl. Phys. **44** (1986) 817 [Yad. Fiz. **44** (1986) 1257].

[Tho77] W. Thome *et al.* [Aachen-CERN-Heidelberg-Munich Collaboration], "Charged-Particle Multiplicity Distributions in p+p Collisions at ISR Energies", Nucl. Phys. B **129** (1977) 365.

[TOT04] TOTEM Collaboration, "TOTEM Technical Design Report", CERN/LHCC 04-002 (2004).

[TOT99] TOTEM Collaboration, "TOTEM Technical Proposal", CERN/LHCC 99-7 (1999).

[Wal04] W. D. Walker, "Multiparton Interactions and Hadron Structure", Phys. Rev. D **69** (2004) 034007.

[Whi74] J. Whitmore, "Experimental Results on Strong Interactions in the NAL Hydrogen Bubble Chamber", Phys. Rept. **10** (1974) 273.

[Wil09] A. Wilk, "Particle Identification using a Neural Network Algorithm in the ALICE TRD", Ph.D. Thesis, University of Münster, to be published in 2009.

[Won94] C.-Y. Wong, "Introduction to High-Energy Heavy-Ion Collisions", World Scientific Publishing (1994).

[Zaj86] W. A. Zajc, "KNO scaling isn't what it used to be", Phys. Lett. B **175** (1986) 219.

List of Acronyms

ACORDE	ALICE Cosmic Ray Detector
AGK cutting rules	Abramovsky-Gribov-Kancheli cutting rules
AGS	Alternating Gradient Synchrotron
ALEPH	Apparatus for LEP Physics
ALICE	A Large Ion Collider Experiment
AliEn	ALICE Environment
AOD	Analysis-Object Data
API	Application Programming Interface
ATLAS	A Toroidal LHC Apparatus
CAF	CERN Analysis Facility
CDF	Collider Detector at Fermilab
CERN	Conseil Européen de la Recherche Nucléaire (European Organization for Nuclear Research)
CMS	Compact Muon Solenoid
CTP	Central Trigger Processor
DA	Detector Algorithm
DAQ	Data Acquisition
DCS	Detector Control System
DD	Double-Diffractive
DDL	Detector Data Link
DP	Datapoint
DPM	Dual Parton Model
DELPHI	Detector with Lepton, Photon, and Hadron Identification
E735	Experiment 735
ECS	Experiment Control System
EMCal	Electromagnetic Calorimeter
ESD	Event-Summary Data
Fermilab (FNAL)	Fermi National Accelerator Laboratory
FMD	Forward Multiplicity Detector
FPGA	Field-Programmable Gate Array
FXS	File Exchange Server
GDC	Global Data Collector
GRP	Global Run Parameters
HIJING	Heavy-Ion Jet Interaction Generator
HLT	High-Level Trigger
HMPID	High-Momentum Particle Identification Detector
ISR	Intersecting Storage Rings
ITS	Inner Tracking System
KNO scaling	Koba–Nielsen–Olesen scaling
LDAP	Lightweight Directory Access Protocol
LDC	Local Data Concentrator

LEIR	Low Energy Ion Ring
LEP	Large Electron-Positron Collider
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty (experiment)
LHCf	Large Hadron Collider forward (experiment)
LINAC3	Linear Accelerator 3
MB	Minimum Bias
MC	Monte Carlo
NBD	Negative Binomial Distribution
ND	Non-Diffractive
ndf	Number of Degrees of Freedom
NSD	Non Single-Diffractive
OCDB	Offline Condition Database
PDC	Physics Data Challenge
PHOS	Photon Spectrometer
PMD	Photon Multiplicity Detector
PS	Proton Synchrotron
PVSS	Prozessvisualisierungs- und -steuerungssystem
QCD	Quantum Chromodynamics
QED	Quantum Electrodynamics
QGP	Quark–Gluon Plasma
QGSM	Quark–Gluon String Model
RFT	Regge Field Theory
RHIC	Relativistic Heavy-Ion Collider
RICH	Ring-Imaging Cherenkov (detector)
R.m.s.	Root mean square
SD	Single-Diffractive
SIS	Schwerionen-Synchrotron
SPS	Super Proton Synchrotron
SUSY	Supersymmetry
SVN	Subversion
TDS	Transient Data Storage
TOF	Time-Of-Flight (detector)
TOTEM	Total Elastic and Diffractive Cross-section Measurement
TPC	Time-Projection Chamber
TR	Transition Radiation
TRD	Transition-Radiation Detector
UA1	Underground Area 1
UA5	Underground Area 5
VTPC	Vertex Time-Projection Chamber
WN	Worker Node
WORM	Write-Once-Read-Many
ZDC	Zero-Degree Calorimeter

List of Figures

List of Tables

Acknowledgements

I would like to thank Johannes P. Wessels for an interesting and challenging Ph.D. thesis topic and for the possibility to perform my thesis work at CERN. I thank Predrag Buncic and Federico Carminati for supporting my doctoral student application.

Special thanks to the members of the ALICE Offline group for an entertaining, ambitious, and personal working atmosphere. I acknowledge Federico Carminati and his amazing skill to lead a team, allowing room for personal development and initiative while retaining success and discipline.

I enjoyed the good working atmosphere in the group of Johannes P. Wessels in Münster and in the first-physics working group of Jean-Pierre Revol and Karel Šafařík.

I appreciated the successful and enjoyable teamwork with Claus Ekman Jørgensen; in particular I thank for his open way of involving me in the first physics topics.

I enjoyed working with Alberto Colla, Raffaele Grosso, and Chiara Zampolli on the Shuttle, and with Gerardo Ganis and Marco Meoni on CAF and PROOF. I thank them for efficient, honest, and enjoyable teamwork and for being always capable to accept my direct way.

At this point, I would like to thank people that provided particular support during the writing of this thesis:

Michelle Connor and Winnie Wong for their linguistic expertise and for hours and hours of proofreading;

Federico Antinori, Cvetan Cheshkov, Andrea Dainese, Costin Grigoras, Peter Hristov, Adam Jacholkowski, Adam Kisiel, Christian Klein-Bösing, Roman Lietava, Mike Lisa, Andreas Morsch, Jean-Pierre Revol, Klaus Reygers, Karel Šafařík, and Jürgen Schukraft for fruitful discussions, that were of particular importance for this thesis;

Martin Poghosyan for his support in trying to understand the mysteries of QGSM;

Latchezar Betev, Yiota Foka, Michael Grevenstette, Andreas Joachim Peters, Suitbert Ramberger, Mareike Rosin, Matevz Tadel, and Chiara Zampolli for providing corrections and suggestions;

and Mareike for countless encouraging words.