



Mean transverse momentum of primary charged particles produced in pp collisions at $\sqrt{s} = 5.02$ TeV measured with ALICE

MASTER'S THESIS

Mario KRÜGER
born on 24th February 1991

submitted on 1st September 2017

Institut für Kernphysik
Goethe-Universität Frankfurt am Main

Supervisor: Prof. Dr. Henner BÜSCHING

Eidesstattliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst, keine anderen als die angegebenen Hilfsmittel verwendet und sämtliche Stellen, die benutzten Werken im Wortlaut oder dem Sinne nach entnommen sind, mit Quellen- bzw. Herkunftsangaben kenntlich gemacht habe.

Frankfurt, den 1. September 2017

Mario Krüger

Contents

0	Introduction	3
1	Theoretical Background	5
1.1	The strong interaction	5
1.2	Particle production in proton-proton collisions	6
2	ALICE experiment	9
2.1	Inner Tracking System	11
2.2	Time Projection Chamber	12
2.3	Data acquisition	13
2.4	Primary vertex	13
2.5	Monte Carlo simulation	15
3	Analysis	17
3.1	Dataset	17
3.2	p_T spectra	18
3.3	$\langle p_T \rangle$ vs. N_{ch} - re-weighting method	20
3.4	$\langle p_T \rangle$ vs. N_{ch} - Bayesian unfolding method	28
3.5	Systematic uncertainties and results	41
4	Summary and outlook	49
5	Acknowledgements	51
6	References	52
A	Kinematic variables	55
B	Runlist	56

Chapter 0

Introduction

Scattering experiments have been essential in our current understanding of the constituents of matter, since the inelastic scattering of projectiles at a target can give access to otherwise hidden substructures. As the resolution of a scattering experiment is anti-proportional to the momentum p of the projectile, higher and higher collision energies provided access to a finer and finer picture of matter as we see it today.

With state-of-the art accelerators, particles can be brought to collision at very high energies. Experiments inspecting the collisions are powerful microscopes to look deeply into the substructure of matter. Particles produced in these collisions are created as ensembles of elementary building blocks, the so-called quarks and gluons. In ordinary matter these constituents are always bound in so-called hadrons. Yet, it is assumed, that at the beginning of the universe, shortly after the big bang, there was a phase where the quarks and gluons were able to move freely in a so-called Quark-Gluon Plasma. At high energy densities realized in heavy-ion collisions at the Large Hadron Collider at CERN, this state of matter can be reproduced in the laboratory and investigated with experiments like ALICE. Measurements of proton-proton collisions serve as a reference, where no such plasma is created.

The particle production mechanisms in these collisions can be studied by means of observables like the number of produced particles and their transverse-momentum spectra. In this work, the correlation between the mean values of the transverse-momentum spectra and the number of charged particles is studied for proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV. In the first chapter an introduction to the properties of the strong interaction and the particle production mechanisms in high-energy collisions is given. The second chapter discusses the experimental setup of ALICE. In the third chapter the analysis steps and results are presented. The work is concluded by a summary and outlook.

CHAPTER 0. INTRODUCTION

Chapter 1

Theoretical Background

In the first section of this chapter a short introduction to the properties of the strong force is given. The second section shortly discusses particle production in high energy collisions and the observables studied in this work.

1.1 The strong interaction

The strong interaction is the strongest of the four fundamental forces (among gravity, weak interaction and electromagnetism). Its theoretical description is provided by Quantum Chromodynamics (QCD), which is an integral part of the current standard model of particle physics. Besides protons and neutrons, many different particle species are found in high energy particle collisions. These particles consist of so-called partons, more precisely quarks and gluons (the mediators of the strong force). QCD describes the strong interactions between these partons. There are six different quark ‘flavours’, called up, down, strange, charm, top and bottom. The quarks as well as the gluons carry so-called colour charge as a quantum mechanical degree of freedom. These colours are denoted by red, green, blue and the respective anti-colours. No free quarks or gluons were observed yet. In ordinary matter, they are confined in so-called hadrons to form colour neutral objects. The so-called valence quarks define the quantum numbers of the hadron. There are two different types of hadrons: Baryons, which consist of three valence quarks and mesons, which consist of a quark and an anti-quark. The masses of these valence quarks do not add up to the total mass of the hadron. Its actual mass is mainly caused by a cloud of gluons and dynamically produced quark-anti-quark pairs, the so-called sea-quarks.

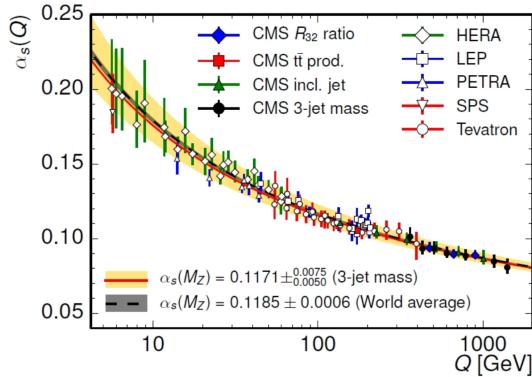


Figure 1.1: Coupling constant α_S of the strong interaction as a function of momentum transfer Q as measured by different experiments [KST⁺15].

The coupling strength of Quantum Chromodynamics has a strong dependence on the momentum transfer Q in the respective interaction:

$$\alpha_S = \frac{12\pi}{(33 - 2n_f) \cdot \ln(Q^2/\Lambda_{QCD}^2)} \quad (1.1.1)$$

Here Λ_{QCD} is the only free parameter in QCD and n_f number of quark flavours which can contribute at the given energy scale. This so-called running coupling constant α_S is shown as a function of Q in figure 1.1. For large momentum transfer respectively small distances, the interaction between the quarks is weak. This is called asymptotic freedom. Therefore, in heavy-ion collisions where the energy density is large, it is believed that the quarks can move quasi free. This state of matter is called a Quark Gluon Plasma (QGP). Its properties can be accessed indirectly through a variety of signatures, which are discussed in detail in [SSS10]. For more details on the strong interaction see [Gri87].

1.2 Particle production in proton-proton collisions

In particle accelerators like the Large Hadron Collider at CERN, protons can be brought to collision at very high centre-of-mass energies. If these protons scatter inelastically, new particles are produced, which can be measured by detectors surrounding the interaction point. Before the collision, the total momentum in the plane transverse to the beam axis is zero. Therefore, this momentum component of the particles produced in the collision is a direct consequence of the interactions in the scattering process.

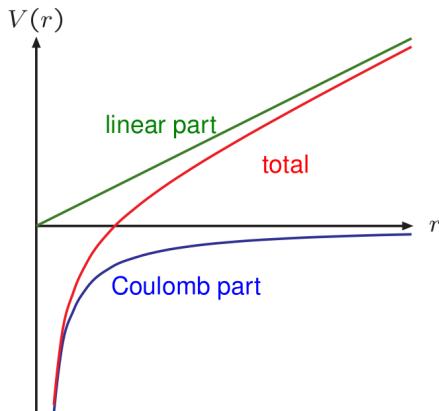


Figure 1.2: Quark-antiquark potential [Sjö09].

The hardness of a collision is characterized by the momentum transfer Q between the colliding protons. In soft collisions, with little momentum transfer, only few particles are produced, whereas in hard scattering events many particles with high transverse momenta emerge from the collision. For these hard collisions with a large momentum transfer Q , the coupling constant of the strong interaction α_s becomes small (as shown in figure 1.1). Therefore, in this regime the elementary cross sections of the scattering partons can be approximated using perturbative QCD (pQCD). Together with the so-called parton distribution functions, which describe the density distribution of the different partons in the proton, and a fragmentation function, this elementary cross section can be used to calculate the differential cross section of the overall particle production in a hard proton-proton collision. This approach is called factorization (for more details see [Lü10]).

In the description of the overall particle production observed in proton-proton collisions, these hard collisions are important for the high transverse momentum part of the spectrum. The particle production at low transverse momenta is dominated by soft processes. For soft collisions it is not possible to use the factorization described before, since the coupling constant is large and therefore the elementary cross sections can not be approximated using pQCD. Therefore, these soft processes can only be described by means of phenomenological models.

An example for such a phenomenological model is the so-called Lund string model [AGIS83], which is implemented in the PYTHIA event generator used in this work. The purpose of this model is to describe the process of hadronization by means

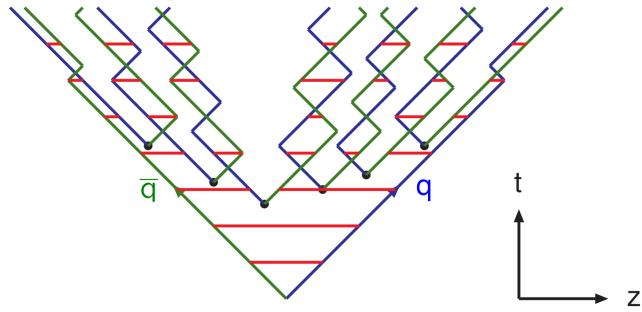


Figure 1.3: Illustration of fragmentation in the Lund string model [Sjö09].

of a fragmenting quark-antiquark pair. The potential between the quark and the antiquark is given by:

$$V(r) = -\frac{3}{4} \frac{\alpha_S(r) \hbar c}{r} + k \cdot r \quad (1.2.1)$$

It contains a Coulomb-like and a linear term. The shape of this potential is illustrated in figure 1.2. Due to the linear term in the potential, the energy contained in the field between the quark and the antiquark rises proportional to the string tension k with their distance. If this energy is large enough, a new quark-antiquark pair is created out of vacuum and the string breaks: $q - \bar{q} \rightarrow q - \bar{q}' + q' - \bar{q}$. This newly produced quark-antiquark pair can then continue to fragment further. This is illustrated in figure 1.3. The blue and the green lines indicate quarks, respectively antiquarks. The initial quark-antiquark pair moves apart in z direction. After some time t , the string, indicated by the red lines, breaks and a new quark-antiquark pair is created. This process repeats until the energy in the string is no longer large enough to create more quark-antiquark pairs. Analogously, diquark-antidiquark pairs can be created when the string breaks[GO09]: $q - \bar{q} \rightarrow q - q'q' + \bar{q}'\bar{q}' - \bar{q}$. This will eventually lead to the production of baryons. The number of final state hadrons, which can be produced via this fragmentation process depends on the energy in the initial quark-antiquark pair.

An important observable to confirm if the soft processes are modelled accurately in the event generator, is the relation between the number of particles produced in the collision and their respective transverse momentum spectrum. To visualize this relation, the transverse momentum spectrum is usually represented by its mean $\langle p_T \rangle$. In this work, $\langle p_T \rangle$ is studied for proton-proton collisions at a $\sqrt{s} = 5.02$ TeV.

Chapter 2

ALICE experiment

The experimental setup used in this work is located on the premises of the European Organization for Nuclear Research (CERN). ALICE (A Large Ion Collider Experiment) is one of the four major experiments at the world's largest ion storage ring, the so called Large Hadron Collider (LHC). Together with its preaccelerators the LHC can be found in the frontier area of France and Switzerland. It is located approximately 100 m underground and has a circumference of about 27 km. Inside the two LHC beam pipes, protons as well as lead ions, which are fully stripped of their electrons, can be accelerated up to nearly the speed of light. The counter-rotating beams are brought to collision at four designated crossing points, around one of which the ALICE experiment is built. ALICE is dedicated to investigate the properties of the Quark Gluon Plasma, an endeavour on which currently about 2800 [LHC] researchers from all over the world are collaborating.

To obtain as broad a picture as possible of the particle collisions occurring at the beam-crossing point, many different detector technologies are employed in ALICE. Its main feature is the ability to measure the three dimensional momenta of electrically charged particles even at high particle densities as they are found in lead-lead collisions. Additionally, some of the detectors provide means for identifying the different particle species. Their respective particle identification (PID) capability is advantageous for specific momentum ranges and particle species. In this regard, the various detector technologies complement one another. ALICE is also capable to measure the energies of neutral particles by means of calorimeters.

Figure 2.1 illustrates how the different detectors are arranged in the experiment. The main detector systems reside in a normally conductive 0.5 T solenoid, the so-called L3 magnet (named after the predecessor experiment), and are assembled in layers cylindrically around the interaction point. Because of their importance for

THE ALICE DETECTOR

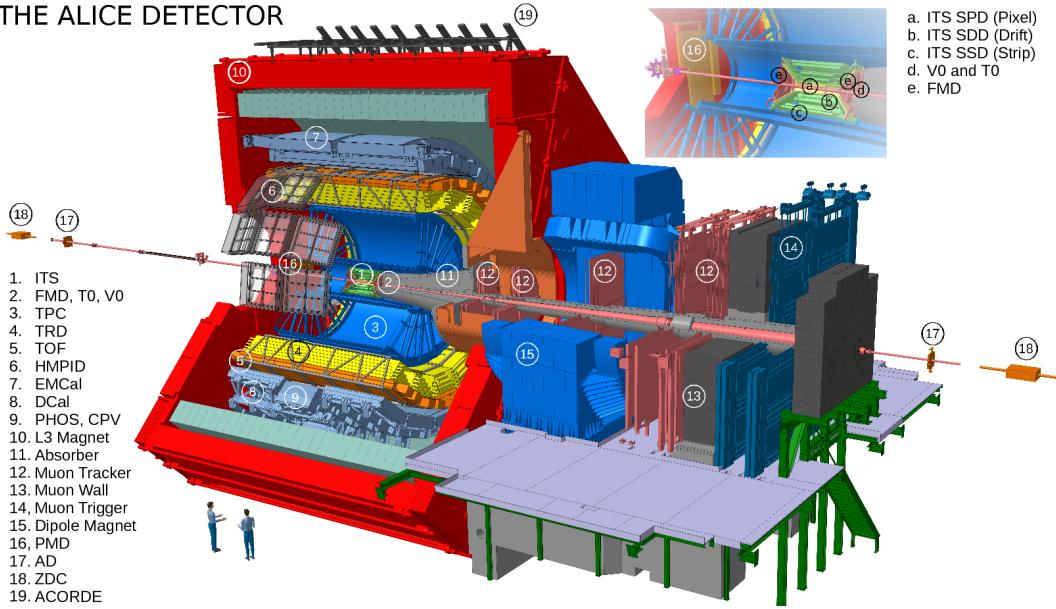


Figure 2.1: Illustration of the ALICE experiment [Ali14].

this work, the two innermost detectors of this onion-like arrangement, the Inner Tracking System (ITS) and the Time Projection Chamber (TPC) are discussed in more detail subsequently. The TPC is followed by the Transition Radiation Detector (TRD), which is used for particle identification, especially for distinguishing between electrons and pions. The TRD is surrounded by the Time of Flight Detector (TOF), which in combination with the T0 detector can be used to measure the time it takes for a particle to traverse the experiment. Together with the TPC momentum measurement, the particle velocity inferred from this time of flight can be used to identify the particle species. The TOF is followed by several detector systems, which in contrast to the afore mentioned detectors do not cover the full $\phi = 2\pi$ acceptance around the beam line. One of them is the High Momentum Particle Identification Detector (HMPID) based on a Ring Imaging Cherenkov Detector (RICH), which is capable of identifying charged particles with high momenta. Three electromagnetic calorimeters are placed at the top (Electromagnetic Calorimeter (EMCal)) respectively at the bottom (Photon Spectrometer (PHOS) and Dijet Calorimeter (DCal)) of the experiment. They provide the means to measure the energy and position of particles, which primarily interact via the electromagnetic force (e.g. photons) by stopping them and absorbing their energy. The Forward Multiplicity Detector (FMD) and the Photon Multiplicity Detector (PMD) are used to estimate the number of charged particles respectively photons in forward direction. The V0 detector is used for triggering and centrality determination in lead-lead collisions. Further along the beam line, the Zero Degree

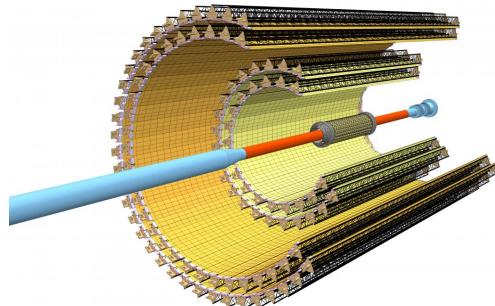


Figure 2.2: The Inner Tracking System, consisting of the three subsystems SPD (innermost), SDD and SSD (outermost) [Col14].

Calorimeters (ZDCs) can be found. They are used to estimate the centrality of lead-lead collisions by measuring remnants of the colliding nuclei. For the measurement of muons, a single-arm detector is installed in forward direction. In order to prevent all particles but the hardly interacting muons from reaching the detector, it is placed behind absorbers. The ALICE Cosmic Ray Detector (ACORDE) is located on top of the L3 magnet. It is used to trigger on cosmic particles, which can be used to calibrate the detectors.

2.1 Inner Tracking System

The Inner Tracking System (ITS) is the innermost detector in ALICE. It is assembled cylindrically around the beam pipe and consists of three subdetector systems, each consisting of two layers of silicon-based semiconductor detectors. These subdetector systems are the Silicon Pixel Detector (SPD), which is located close around the beam pipe, the Silicon Drift Detector (SDD) and the Silicon Strip Detector (SSD). Due to its proximity to the beam-crossing point, the ITS plays an important role in determining the position of the particle collision, the so-called primary vertex. Together with the Time Projection Chamber, it provides the space points for reconstructing tracks of the charged particles produced in the collision. These tracks are then used to obtain the transverse momenta of the corresponding particles (see next section). An illustration of the ITS is shown in figure 2.2. To provide a good spacial resolution of the tracks even for the high track densities close to the primary vertex, the SPD has a very high pixel granularity. In addition, the SDD and SSD are able to measure the energy loss of the traversing particles, which can be used for particle identification. A detailed technical description of the ITS can be found in [Col14].

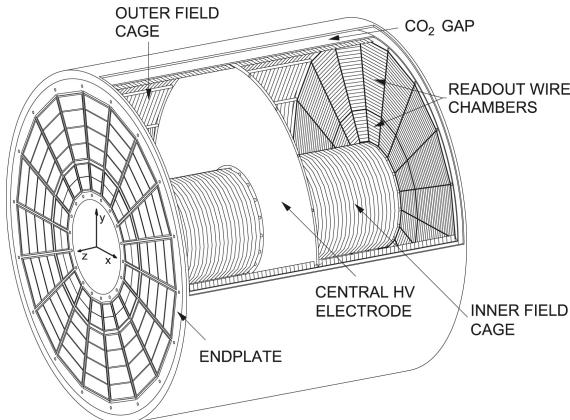


Figure 2.3: Illustration of the ALICE Time Projection Chamber [AA⁺¹⁰].

2.2 Time Projection Chamber

The Time Projection Chamber (TPC) consists of a large cylindrical barrel, which is installed around the ITS. It extends from a radial distance of $r_1 = 0.85$ m up to $r_2 = 2.47$ m and is 5 m long[A⁺⁰⁸]. As shown in figure 2.3, a central electrode divides the 88 m^3 volume filled with a gas mixture of Ar and CO₂ into two separate regions. A high voltage is applied between this central electrode and the two endplates of the barrel, resulting in an electric field of $E = 400 \text{ V/cm}$. The electric field is aligned parallel to the magnetic $B = 0.5 \text{ T}$ field of the L3 magnet. Charged particles produced in the collisions are bent in the magnetic field and ionize the gas contained in the TPC on their way through the detector. Due to the electric field, the liberated electrons are accelerated towards the endplates. Interactions with the gas molecules result in a constant drift velocity. At the endplates the electrons are measured using Multi-Wire Proportional Chambers (MWPCs) with a cathode pad readout. The spacial points obtained from this measurement give a two dimensional projection of the particle track. The third dimension is inferred from the drift time of the electrons. The curvature ρ of the tracks can be used to determine the magnetic rigidity:

$$\frac{p_T}{q} = B\rho \quad (2.2.1)$$

where p_T is the momentum component in the plane transverse to the beam axis. The majority of particles are charged with $q = \pm e$. Therefore, the transverse momentum can be calculated by $p_T = B\rho e$. Using the polar angle θ of the track, the total momentum $|\mathbf{p}|$ can be inferred from the definition of the transverse momentum $p_T = \sin(\theta) \cdot |\mathbf{p}|$. In addition to the momentum measurement, the TPC can be used

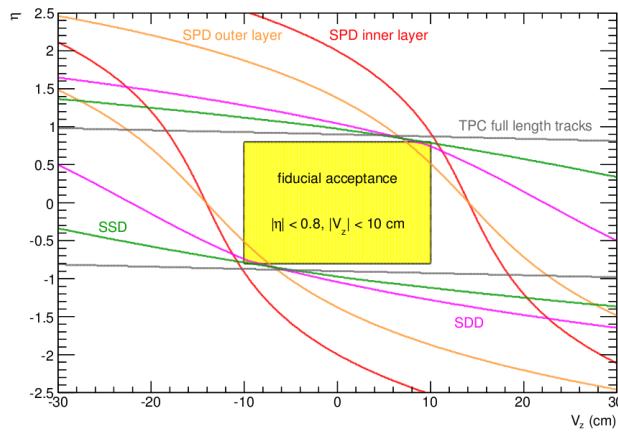


Figure 2.4: Detector acceptance of the TPC and the SPD, SDD and SSD subdetectors of the ITS depending on the position of the primary vertex in z direction [Kni14].

for identifying the particle species by means of their specific energy loss in the gas. A detailed technical description of the TPC can be found in [AA⁺10].

2.3 Data acquisition

For each beam crossing inside of ALICE, there is a certain probability for two particles to interact with each other, i.e. for a so-called event. In order to distinguish these events from other beam induced signals in the experiment, a set of detectors is required to respond in coincidence with the beam crossing. Only events fulfilling this so-called minimum-bias trigger condition are being recorded for further analysis. The basic unit of recorded data in ALICE is called a run. It corresponds to a time interval of continuous data taking (usually not more than a few hours) without changes in the detector configuration. Runs can be identified by a unique run number and are aggregated to so-called data taking periods. Those are usually defined by a certain physics setup of the LHC beam (e.g. collision system and centre-of-mass energy) and setting of the experiment (e.g. direction and magnitude of magnetic field).

2.4 Primary vertex

The primary vertex of an event is defined as the point of origin of all the primary particles produced in the collision. It is determined by the mean origin of the charged-particle tracks, which are constrained by space points from both the TPC

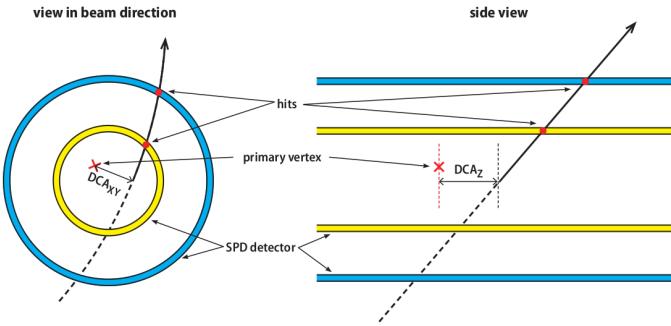


Figure 2.5: Illustration of the DCA in the plane transverse to the beam axis (left) and along the beam line (right) [Kni14].

and the ITS (so-called global tracks). Therefore, the accuracy of the vertex position depends on the number of tracks measured in the collision. The position of the primary vertex affects the spacial acceptance covered by the detectors and thereby changes the kinematic range of the particles recorded in the measurement. Figure 2.4 shows the detector acceptance in pseudorapidity (see appendix A) as a function of the position of the primary vertex along the beam direction (z) with respect to the nominal interaction point. The grey line indicates that for the TPC there is hardly any change of the detector acceptance for vertices within $|V_z| < 30$ cm. In contrast, the η window covered by the different subdetectors of the Inner Tracking System significantly shifts with the position of the interaction point. This is mainly caused by the limited longitudinal dimension of the ITS. The two layers of the SPD for example only have a length of $|l_z| = 14.1$ cm [A⁺08]. Therefore, the detectable emission angle with respect to the beam line is severely restricted for collisions located far from the middle of the detector. To ensure the pseudorapidity window of $|\eta| < 0.8$, which is used in this work, the vertex position must not be larger than $V_z \pm 10$ cm. The geometrical acceptance is also affected by the radial distance of the ITS subdetector to the collision vertex. This is shown in figure 2.4 for the outer (orange line) and inner (red line) SPD layers, which have the same longitudinal dimension, but are located at different radii.

The tracks of secondary particles usually do not point to the primary vertex. In ALICE, the closest distance of a track to the primary vertex position is called distance of closest approach (DCA). Figure 2.5 illustrates the geometrical meaning of the DCA. The left panel of the figure shows a profile of the two inner layers of the ITS in the plane transverse to the beam. Here, the distance between the interaction point and the track is called DCA_{xy} . The right panel of the figure shows a projec-

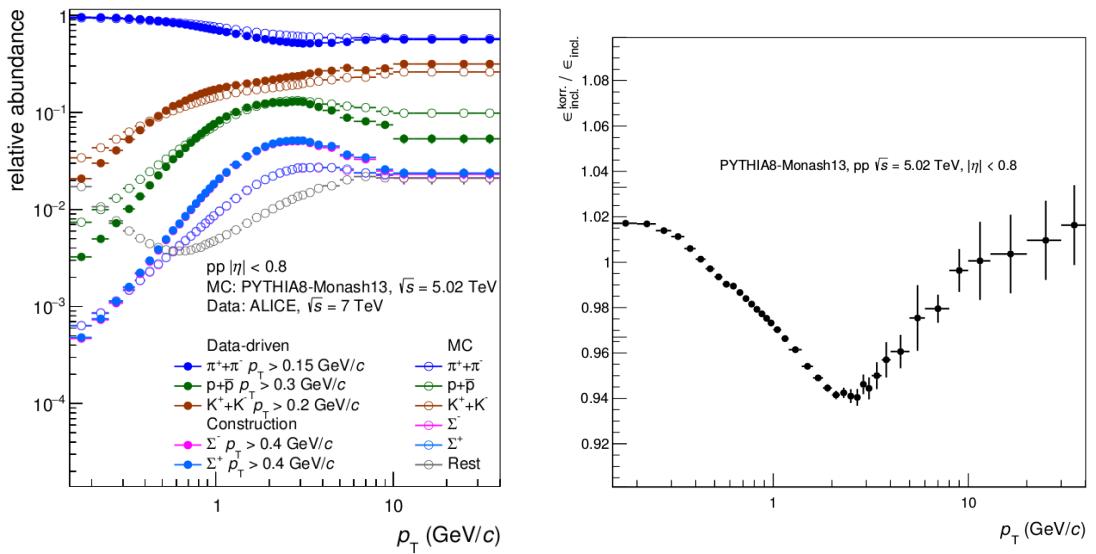


Figure 2.6: The left panel shows the relative abundances of the different particle species produced in pp collisions as a function of p_T for data and MC. In the right panel the ratio of the tracking efficiencies obtained with and without particle composition correction are shown [Huh17].

tion along the beam line, where the distance of closest approach is called DCA_z . In this work, restrictions on the DCA are used to suppress secondary particles in the measurement. These restrictions do not reject all secondary particles, leading to a remaining secondary contamination, which needs to be corrected for.

2.5 Monte Carlo simulation

Due to the limited acceptance and efficiency of the experiment, the raw data measured with ALICE needs to be corrected in order to obtain a physically meaningful image of the collision. To determine these corrections, a detailed knowledge of the geometries, materials and performance of the detector is indispensable. For that purpose, ALICE is modelled using the GEometry And Tracking (GEANT3 [BB⁺94]) software framework, which offers the possibility to simulate the detector response including all the interactions of a particle traversing the experiment. The virtual events used for this simulation are created by Monte Carlo (MC) event generators (in this analysis PYTHIA8 [SAC⁺15]), which emulate the pp collisions based on perturbative QCD in the hard scattering regime and phenomenological models in the soft scattering regime. The artificial measurement obtained from these simulations is then propagated through the same chain of reconstruction algorithms and

quality requirements as the real data.

Calculating corrections for the measured data based on MC simulations not only requires a proper understanding of the detector, but also depends on how accurate the simulated events describe the true physical interactions occurring in the collision. A known issue of the available event generators is their inability to properly reproduce the measured composition of particle species (pions, protons, kaons, sigmas, etc.). In the left panel of figure 2.6 the relative abundances of the different identified charged particles are shown as a function of p_T . The relative abundances of the measured data (filled markers) clearly differ from the ones in the Monte Carlo simulation (empty markers). Since the tracking capability of the TPC and the effect of the track quality constraints strongly depend on the specific particle properties (mass, lifetime, etc.), a bias in the particle composition also affects the overall efficiency. In order to take this into account, the tracking efficiency obtained from the simulation is re-weighted using the relative abundances derived from measurements of identified particles with ALICE. The resulting efficiency has a strong p_T dependence and can differ from the uncorrected one by up to six percent as shown in the right panel of figure 2.6. A detailed description of this data-driven procedure can be found in [Huh17].

Chapter 3

Analysis

The ALICE experiment is very well suited to study the charged-particle production in high-energy collisions. Its Time Projection Chamber and Inner Tracking System allow for a high precision measurement of the charged-particle transverse momentum spectra, which can be characterized by their mean and higher moments. This work aims to study the correlation between these moments and the corresponding number of particles in the collision. The true correlation between these two quantities is lost in the measurement and needs to be reconstructed by means of a Monte Carlo simulation.

In the first section of this chapter, the dataset and data quality constraints used in this work are described. Necessary corrections applied to the measured transverse momentum spectra are discussed in section two. Afterwards, two different approaches for reconstructing the correlation between the transverse momentum and the true number of particles in an event are presented in sections three (re-weighting of moments) and four (unfolding of spectra). Finally, the systematic uncertainties of the resulting moments are discussed and the two methods are compared in the last section of this chapter.

3.1 Dataset

This work is based on measurements of proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 5.02$ TeV from data recorded in the year 2015. A detailed list of the selected runs can be found in appendix B. The transverse momentum (p_T) spectra of primary charged particles are investigated in a kinematic range of $0.15 \text{ GeV}/c \leq p_T < 10 \text{ GeV}/c$ and pseudorapidity $|\eta| < 0.8$. To guarantee this η coverage for both, the ITS and TPC, the primary vertex position is restricted to $|V_z| < 10 \text{ cm}$ along the beam line. Each event must fulfil the minimum-bias trigger,

which requires hits in the two layers of the SPD and both of the forward scintillators V0A and V0C. In this work, the current ‘standard’ track quality constraints for charged-particle analyses in ALICE are used. They are described in detail in [Ali17] and will be mentioned here only for completeness. Each track is required to have:

- at least 2 hits in the ITS with at least one of them in either of the two innermost layers of the SPD.
- a distance of closest approach to the primary vertex of $DCA_z < 2$ cm and $DCA_{xy} < 7\sigma$ (where $\sigma = 26 + \frac{50}{(p_T / \text{GeV}/c)^{1.01}}$ is the standard deviation of the impact parameter resolution).
- a maximum deviation of the ITS and TPC tracking points from the global track fit of $\chi^2_{ITS}/\text{cluster} < 36$ and $\chi^2_{TPC}/\text{cluster} < 4$.
- a maximum deviation between the global track and a track constrained by the TPC space points and the vertex of $\chi^2_{TPC-ITS} < 36$.
- a minimum length of in the TPC (in cm) of $L = 130 - p_T^{-1.5}$ (with p_T in GeV/c), not including the pads lying within a dead zone of 3 cm away from the sector edges.
- a length measured in the TPC readout pad rows crossed by the track of more than $0.85 \cdot L$.
- a length measured in the TPC clusters (one cluster per pad row), which is larger than $0.7 \cdot L$.
- a ratio of crossed TPC pad rows to the number of findable TPC clusters, which is larger than 0.8.
- a fraction of TPC clusters shared with another track, which is less than 0.4.

3.2 p_T spectra

The baseline for further analysis in this work is the raw transverse momentum spectrum composed of the measured yield of global tracks fulfilling the quality conditions described in the previous section. This raw p_T distribution represents only a small sample of all the primary charged particles produced in the collisions, and it is biased by the limits of detector efficiency and acceptance as well as the residual contamination with secondary particles. Therefore, this measurement needs to be

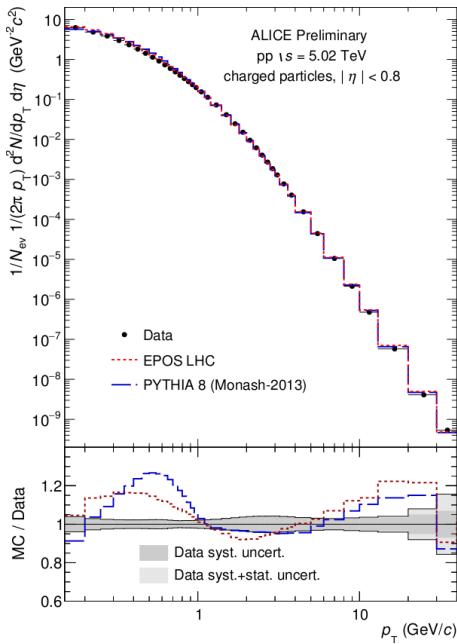


Figure 3.1: Invariant yield of primary charged particles measured in pp collisions at $\sqrt{s} = 5.02$ TeV in a pseudorapidity window of $|\eta| < 0.8$ with ALICE [Ali17].

corrected and properly normalized in order to obtain a physically meaningful observable which is independent of the detector setup. These corrections are done by means of a Monte Carlo simulation, which provides full knowledge of the particles produced in the virtual event as well as the ones reconstructed by the virtual detector. Such, a p_T -dependent detector efficiency can be calculated. This is done by dividing the number of reconstructed by the number of generated primary charged particles within the kinematic range of the measurement. By this definition the tracking efficiency implicitly also takes into account the geometrical acceptance of the experiment. Besides the tracking efficiency, the simulation also provides information about the magnitude of secondary contamination. The fraction of primaries in the reconstructed track sample obtained from the simulation serves as a correction factor for the measured particle yield. To correct for the inaccurate composition of particle species produced by the event generator, a data-driven particle composition correction is applied (for details see [Huh17]).

After applying these corrections to the raw yield, the actual yield of primary charged

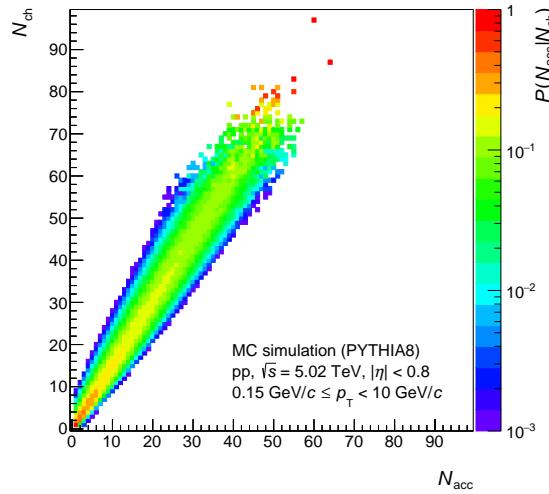


Figure 3.2: Detector response matrix for pp collisions at $\sqrt{s} = 5.02$ TeV.

particles produced in the given kinematic range is obtained. Differentiated in η and p_T as well as normalized, one obtains the Lorentz invariant yield:

$$\frac{d^3E}{dp^3} = \frac{1}{2\pi} \frac{1}{p_T} \frac{d^2N}{dp_T d\eta} \quad (3.2.1)$$

The resulting p_T spectrum for the dataset used in this work is shown in figure 3.1.

After introducing the quality cuts and corrections of the analysis, in the following chapters the multiplicity dependence of the spectra will be discussed. In the following the term ‘measured spectra’ refers to the measured and corrected p_T spectra. For the Monte Carlo results the term ‘reconstructed spectra’ will be used.

3.3 $\langle p_T \rangle$ vs. N_{ch} - re-weighting method

The number of charged particles (henceforth called charged particle multiplicity) produced in the collision increases with the momentum transfer between the two colliding particles. Therefore, the charged-particle multiplicity is a good reference quantity to investigate, how the p_T spectra change with the hardness of the collision. In order to visualize this relation, the shape of the spectra can be characterised by their moments, in particular their mean $\langle p_T \rangle$ and root mean square (RMS) $\sqrt{\langle p_T^2 \rangle}$. For a given p_T distribution the n -th moment is defined by

$$\langle p_T^n \rangle = \frac{\int_{p_T^{min}}^{p_T^{max}} p_T^n \frac{dN}{dp_T} dp_T}{\int_{p_T^{min}}^{p_T^{max}} \frac{dN}{dp_T} dp_T} \quad (3.3.1)$$

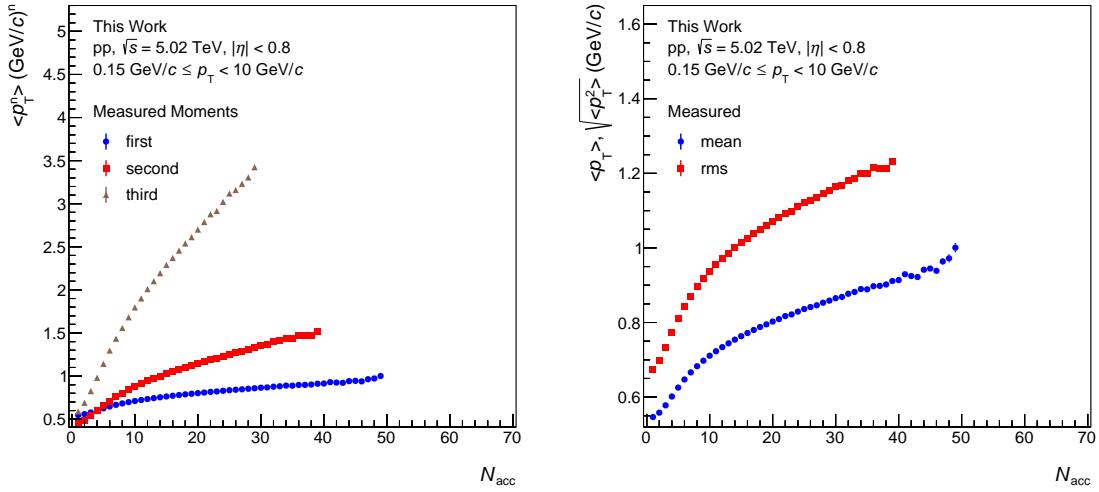


Figure 3.3: The left panel shows the first three moments of the measured p_T spectra as a function of N_{acc} with statistical uncertainties. In the right panel the mean and RMS are shown.

using the p_T differential particle yield dN/dp_T . In this work the integration boundaries $p_T^{min} = 0.15 \text{ GeV}/c$ and $p_T^{max} = 10 \text{ GeV}/c$ are used.

In this work the measured multiplicity N_{acc} is defined as the number of charged-particle tracks in an event, which are reconstructed by the detector and fulfil the quality requirements described in section 3.1. Therefore, N_{acc} is not equal to the true multiplicity N_{ch} , which in this work refers to the number of primary charged particles in an event, which are produced in the kinematic range $0.15 \text{ GeV}/c \leq p_T < 10 \text{ GeV}/c$ and $|\eta| < 0.8$. Many different true multiplicities have a certain probability to be measured as N_{acc} . Hence, a p_T spectrum assigned to N_{acc} contains a mixture of events with different true multiplicities N_{ch} . Based on the p_T spectra which are assigned to the different measured multiplicities N_{acc} , the correlation between N_{ch} and their corresponding transverse momentum spectra will be reconstructed.

From the PYTHIA Monte Carlo simulation used in this work, the conditional probability of an event with multiplicity N_{ch} to be measured with multiplicity N_{acc} is given by the elements $P(N_{acc}|N_{ch})$ of the detector response matrix, which is shown in figure 3.2. By definition, the elements of this response matrix fulfil the following normalization condition for each N_{ch} :

$$\sum_{N_{acc}} P(N_{acc}|N_{ch}) = 1 \quad (3.3.2)$$

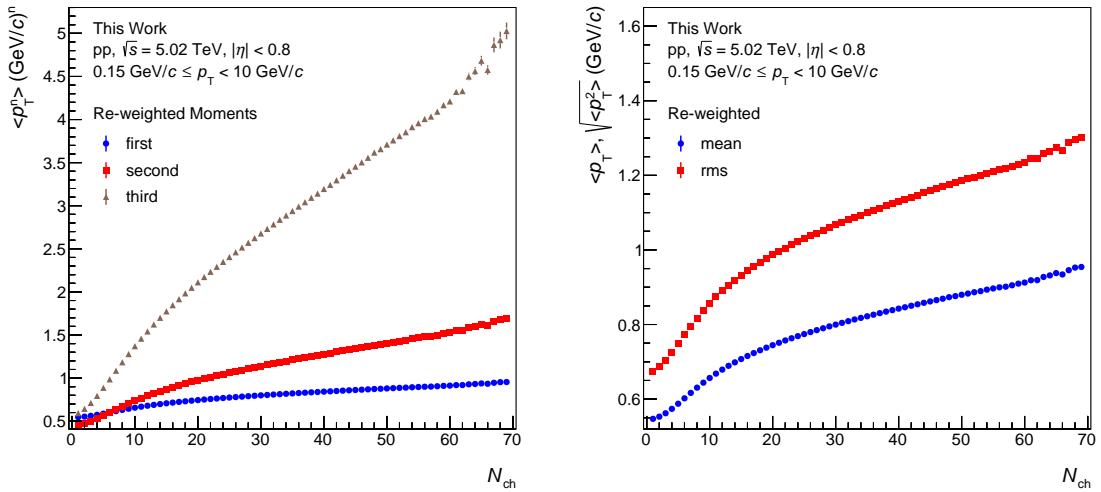


Figure 3.4: The left panel shows the first three re-weighted moments of the measured p_T spectra as a function of true multiplicity N_{ch} with statistical uncertainties. In the right panel the mean and RMS are shown.

These probabilities describe how likely it is for particles produced in an event with multiplicity N_{ch} to contribute to the p_T spectrum assigned to N_{acc} . Therefore, $P(N_{\text{acc}}|N_{\text{ch}})$ can be used as weights for reconstructing $\langle p_T^n \rangle$ as a function of N_{ch} from the moments $\langle p_T^n \rangle(N_{\text{acc}})$ obtained from the measured spectra:

$$\langle p_T^n \rangle(N_{\text{ch}}) = \sum_{N_{\text{acc}}} P(N_{\text{acc}}|N_{\text{ch}}) \cdot \langle p_T^n \rangle(N_{\text{acc}}) \quad (3.3.3)$$

Figure 3.3 shows the first three moments and the RMS of the measured p_T spectra as a function of N_{acc} . In figure 3.4 the resulting re-weighted moments $\langle p_T^n \rangle(N_{\text{ch}})$ are shown. Statistical uncertainties are propagated from the input spectra.

The re-weighted moments show a rise with multiplicity that is less steep than the one observed for $\langle p_T^n \rangle(N_{\text{acc}})$. At low multiplicities all the re-weighted moments and the RMS follow a characteristic S-shape and then develop into a linear slope at about $N_{\text{ch}} = 30$.

3.3.1 Extrapolation of the response matrix

The reach in multiplicity of this re-weighting method is strongly limited by the statistics of the Monte Carlo simulation. For $N_{\text{ch}} > 70$ the response matrix becomes very sparse (see figure 3.2) and therefore its entries do not represent the true response of the detector any more. For example, if only one event with multiplicity $N_{\text{ch}} = 92$ was created by the event generator and measured as $N_{\text{acc}} = 50$, the resulting $P(N_{\text{acc}}|N_{\text{ch}})$ would imply that the detector always reconstructs it in this way.

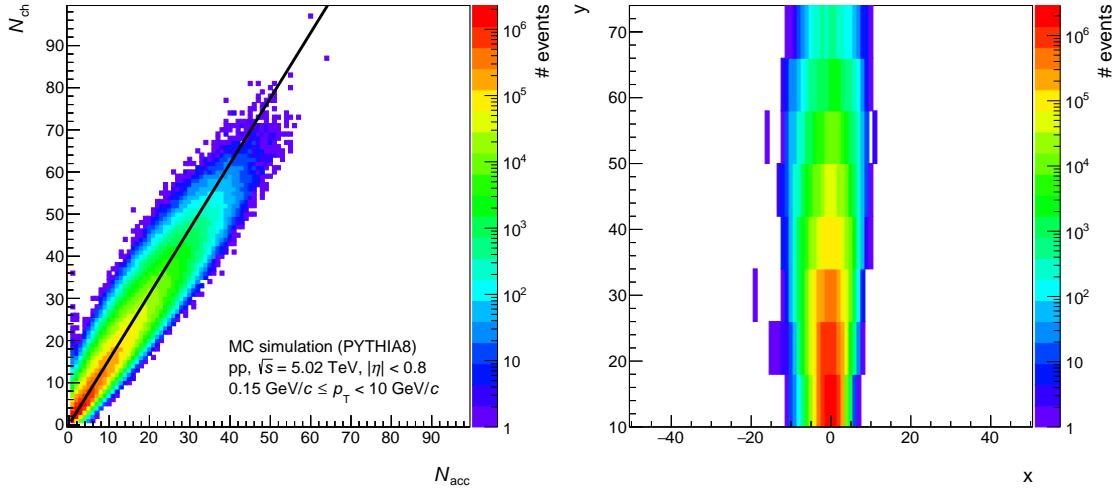


Figure 3.5: The original multiplicity correlation matrix (left) and the rotated and re-binned multiplicity correlation matrix (right), which is used as the basis for the Gaussian fits.

To overcome this statistical limitation, the response matrix can be extrapolated.

The response matrix shown in figure 3.2 represents the probabilities for a true multiplicity N_{ch} to be measured with multiplicity N_{acc} . It is obtained by normalizing the original multiplicity correlation matrix, which contains the number of events associated with a certain (N_{ch}, N_{acc}) -combination, according to equation 3.3.2. This original multiplicity correlation matrix is shown in the left panel of figure 3.5 and will be the basis for the extrapolation. The shape of the event distribution appears to follow a diagonal in the $N_{acc} - N_{ch}$ plane, which is indicated by the black line. Orthogonal to this diagonal the events seem to be distributed with a Gaussian shape.

In order to extrapolate this multiplicity correlation matrix, its entries are rotated in the $N_{acc} - N_{ch}$ plane until the former diagonal is located at the y-axis. The new frame of reference is now called $x - y$. For low N_{ch} , the multiplicity correlation matrix is not symmetrical around the diagonal, and for high N_{ch} it becomes sparse (compare figure 3.5 left). Therefore, the range between $10 < y < 74$ is chosen to be the basis for the extrapolation. It is divided into eight intervals of $\Delta y = 8$, as shown in the right panel of figure 3.5. Subsequently, each of these intervals is separately fitted with a Gaussian function:

$$g(x) = A \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (3.3.4)$$

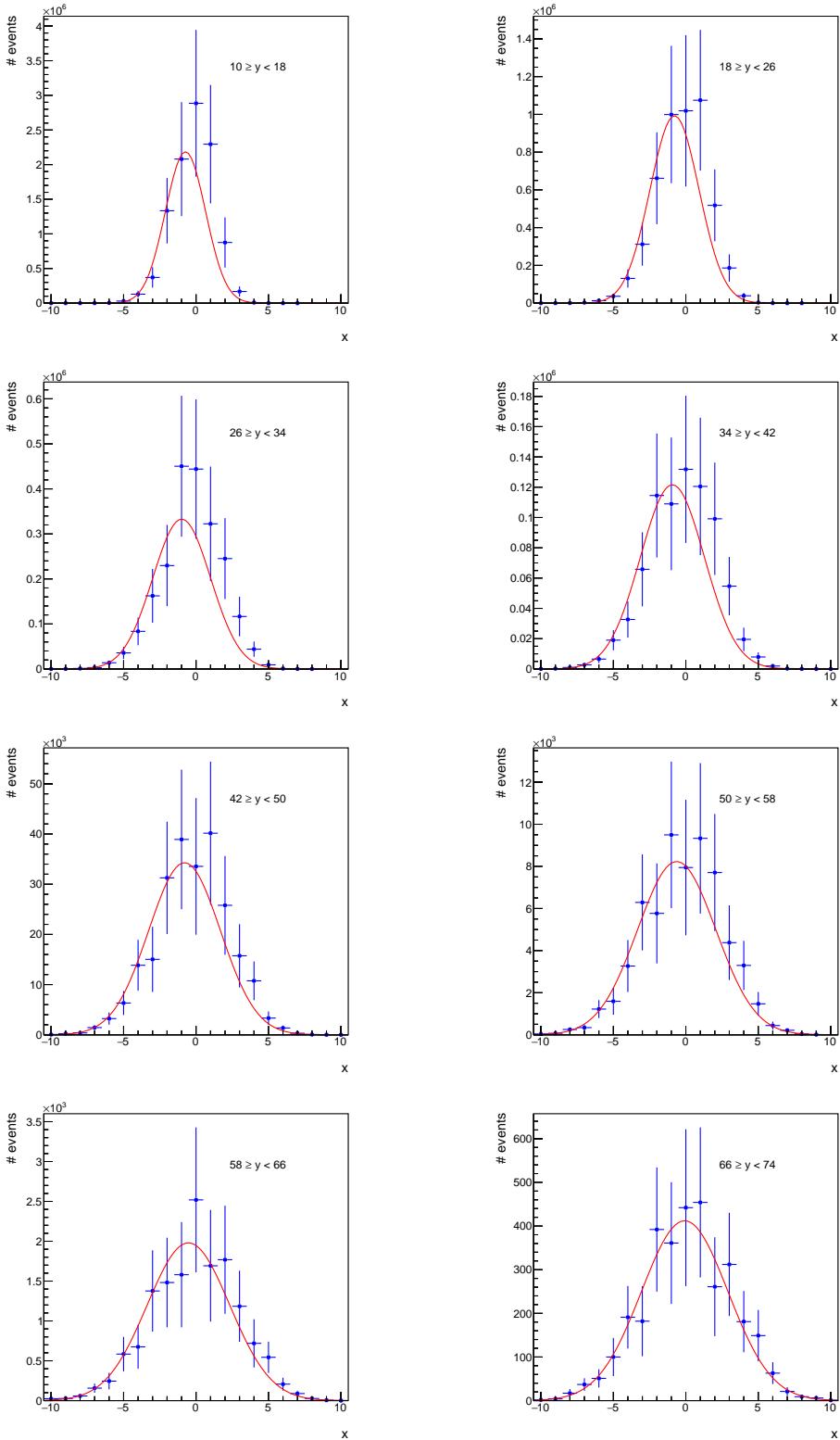


Figure 3.6: Gaussian fits to the different y intervals shown in the right panel of figure 3.5. The lowest interval is displayed in the upper left panel and the highest one in the bottom right panel.

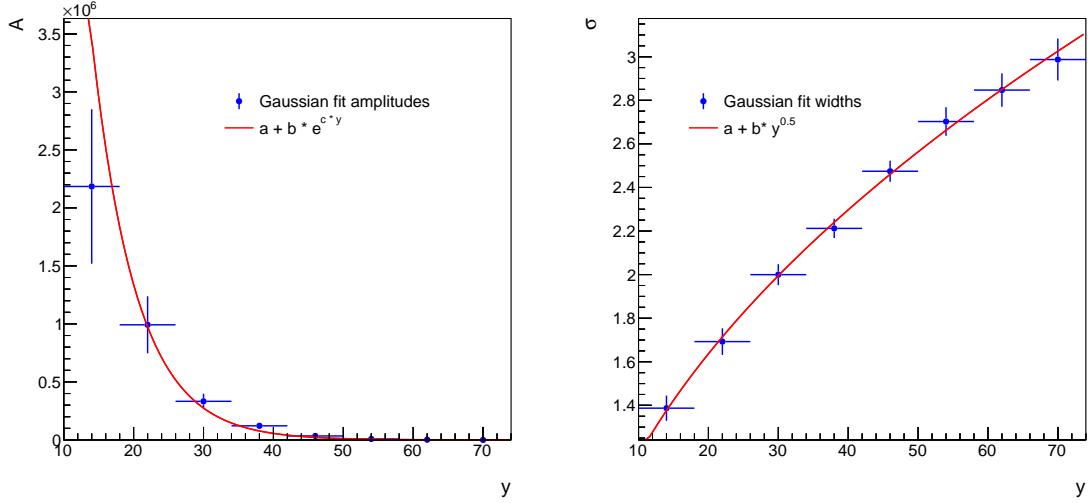


Figure 3.7: Parametrisation of the Gaussian fit parameters. The left panel shows the amplitude A , which is fitted with an exponential function. In the right panel the width σ is shown, fitted with a square root function.

The three fit parameters are the amplitude A , the mean μ and the width σ . In figure 3.6 the resulting fits are shown from the lowest y interval at the upper left panel to the highest one at the bottom right panel. For the extrapolation the parameters of these Gaussian fits are parametrized as a function of y . In figure 3.7 on the left panel the amplitude $A(y)$ of the Gaussian fits is shown together with an exponential fit. The right panel shows the width $\sigma(y)$, which is well described by a square root function. The mean value $\mu(y)$ of all the Gaussian fits fluctuates around zero and is fitted by a constant, which is not shown here.

These parametrizations of the Gaussian fit parameters are used to extrapolate the multiplicity correlation matrix in the $x - y$ plane. It is then rotated back into the $N_{acc} - N_{ch}$ frame of reference. Figure 3.8 shows the resulting multiplicity correlation matrix, which is smooth compared to the original one (figure 3.5 left). The starting point of the extrapolation is indicated by the black line. It is chosen in such a way that each row of the multiplicity correlation matrix consists either only of values from the original matrix or of values from the extrapolation. This way, the two do not get mixed when normalizing the different N_{ch} rows to obtain the detector response matrix.

Figure 3.9 shows the original (left panel) and the extrapolated (right panel) response matrices. The transition between the original matrix and the extrapolation

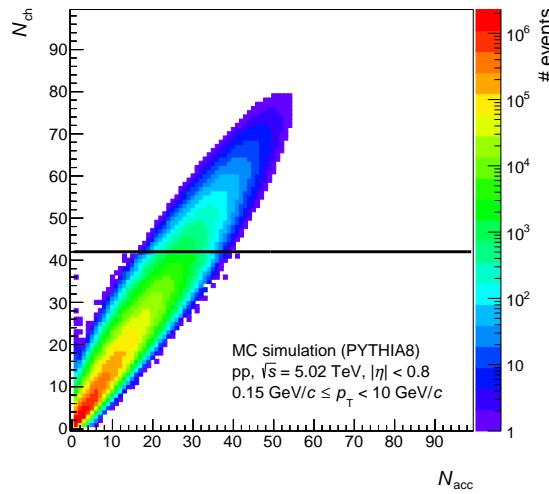


Figure 3.8: Extrapolated multiplicity correlation matrix. The extrapolation starts above the black line.

is quiet smooth and follows the trend suggested by the original response matrix. Figure 3.10 shows the mean and RMS of the measured spectra, re-weighted with the original response matrix on the left panel and with the extrapolated response matrix on the right panel. Beyond $N_{\text{ch}} = 60$ the extrapolation of the response matrix results in a smoothing of the fluctuating points. This is due to the fact that the weighting factors $P(N_{\text{acc}}|N_{\text{ch}})$ used in equation 3.3.3 now have more appropriate values. Yet, the re-weighting is still limited by the available moments $\langle p_T^n \rangle (N_{\text{acc}})$ from the measured spectra. These moments become more sparse and inaccurate for $N_{\text{acc}} > 60$ due to the limited number of events with such high multiplicities. This might explain, why the re-weighted mean and RMS do not precisely follow the previous linear trends for $N_{\text{ch}} > 70$.

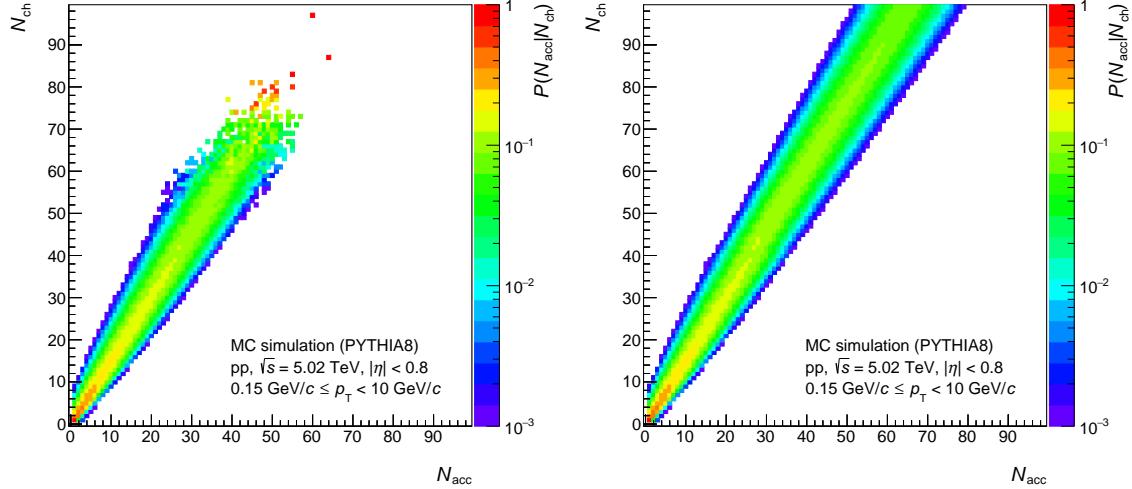
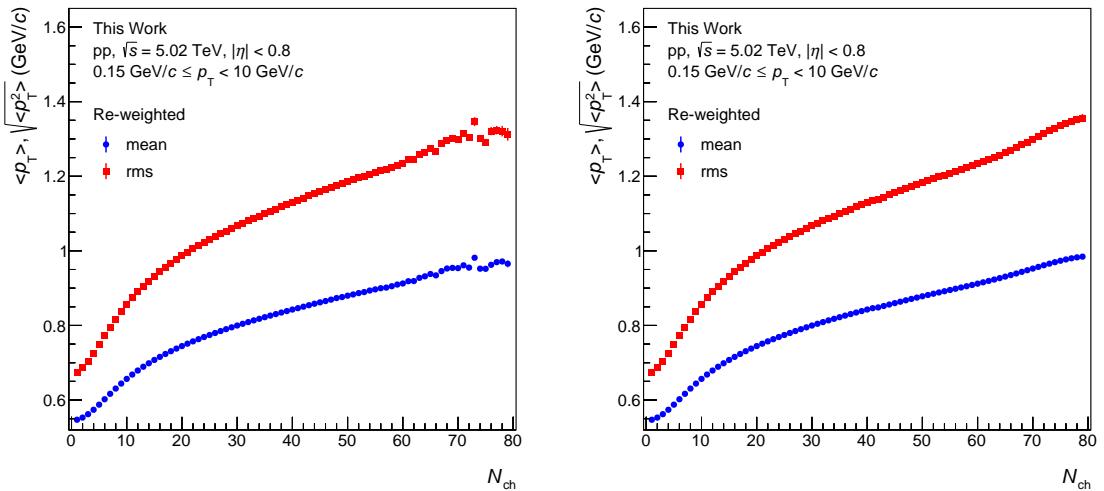


Figure 3.9: Response matrix before (left) and after (right) extrapolation.

Figure 3.10: Mean and RMS of the measured p_T spectra, re-weighted with the original (left) and the extrapolated (right) response matrix.

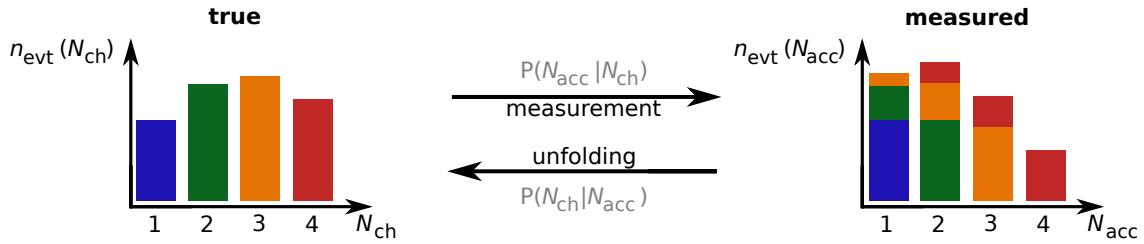


Figure 3.11: Illustration of the relations between the true and measured multiplicity distributions.

3.4 $\langle p_T \rangle$ vs. N_{ch} - Bayesian unfolding method

In this section, an alternative approach for reconstructing the correlation between the p_T spectra and the true event multiplicity N_{ch} is discussed. Instead of re-weighting the moments of the measured p_T spectra as described in the previous section, the particles contributing to the p_T spectrum of a measured N_{acc} are now directly redistributed to their corresponding true multiplicities N_{ch} . It is important to note, that just like in the previous section, the term measured spectra refers to the measured and corrected p_T spectra as a function of N_{acc} . The basic assumption for the following reasoning is that after applying the efficiency corrections described in section 3.2, each of these measured p_T spectra contains the true number of particles produced in the events, which contribute to the respective measured spectrum.

In order to perform this ‘resorting’ of particles, it is first necessary to know how the events contributing to the p_T spectrum assigned to N_{acc} are composed of events with true multiplicities N_{ch} . This will be discussed in the following. Subsequently, this knowledge about the composition of the events is used to decompose the measured p_T spectra and obtain the p_T spectra as a function of the true multiplicities N_{ch} .

3.4.1 Unfolding of events

The measured multiplicity distribution $n_{\text{evt}}(N_{\text{acc}})$ is the result of folding the true multiplicity distribution of the collisions $n_{\text{evt}}(N_{\text{ch}})$ with the response of the detector. This is illustrated in figure 3.11. To unfold the measured multiplicity distribution and again obtain $n_{\text{evt}}(N_{\text{ch}})$, it is necessary to know how $n_{\text{evt}}(N_{\text{acc}})$ is composed of events with the various true multiplicities N_{ch} . The fraction of events which are measured with multiplicity N_{acc} , but originally have a true multiplicity N_{ch} , is represented by the conditional probability $P(N_{\text{ch}}|N_{\text{acc}})$. This is not to be confused with the detector response $P(N_{\text{acc}}|N_{\text{ch}})$. In principle, $P(N_{\text{ch}}|N_{\text{acc}})$ could be obtained directly from the MC simulation just like the detector response. However,

this way it would strongly depend on how accurate the event generator describes the multiplicity distribution of the collisions. The detector response $P(N_{acc}|N_{ch})$ in contrast, in first approximation only depends on the GEANT reconstruction capability. Therefore, to be as independent of the event generator as possible, in this work $P(N_{ch}|N_{acc})$ is obtained by inverting the conditional probabilities $P(N_{acc}|N_{ch})$ from the response matrix by iteratively applying Bayes' theorem[BP63]. This procedure was developed by G.D'Agostini [D'A95] and is described in the following.

Suppose, $P(N_{ch})$ and $P(N_{acc})$ are the probability distributions for the occurrence of the true and measured multiplicities respectively and the probability for a certain (N_{ch}, N_{acc}) -pair is given by $P(N_{ch} \cap N_{acc})$. Then the corresponding conditional probabilities are defined as:

$$P(N_{acc}|N_{ch}) = \frac{P(N_{ch} \cap N_{acc})}{P(N_{ch})} \quad \text{and} \quad P(N_{ch}|N_{acc}) = \frac{P(N_{ch} \cap N_{acc})}{P(N_{acc})} \quad (3.4.1)$$

Therefore, Bayes' theorem can be expressed in the language of measured and true multiplicities as:

$$P(N_{ch}|N_{acc}) = \frac{P(N_{acc}|N_{ch}) \cdot P(N_{ch})}{P(N_{acc})} \quad (3.4.2)$$

Every event with a measured multiplicity N_{acc} has to originate from some physical multiplicity N_{ch} . If $P(N_{ch})$ is known, then the total probability to measure an event with a multiplicity N_{acc} can be constructed by summing over all true multiplicities in the following way:

$$P(N_{acc}) = \sum_{N_{ch}} P(N_{ch} \cap N_{acc}) = \sum_{N_{ch}} P(N_{acc}|N_{ch})P(N_{ch}) \quad (3.4.3)$$

Therefore, the elements $P(N_{ch}|N_{acc})$ of the so-called unfolding matrix can be expressed as:

$$P(N_{ch}|N_{acc}) = \frac{P(N_{acc}|N_{ch}) \cdot P(N_{ch})}{\sum_{N'_{ch}} P(N_{acc}|N'_{ch})P(N'_{ch})} \quad (3.4.4)$$

By this definition the unfolding matrix is normalized to unity when summing over the true multiplicities N_{ch} for a fixed N_{acc} :

$$\sum_{N_{ch}} P(N_{ch}|N_{acc}) = 1 \quad (3.4.5)$$

This is in contrast to the response matrix $P(N_{acc}|N_{ch})$, which is normalized to unity when summing over the measured multiplicities N_{acc} for a fixed N_{ch} (see equation 3.3.2).

Equation 3.4.4 represents the desired relation between the response matrix and the unfolding matrix. Unfortunately, in order to calculate $P(N_{ch}|N_{acc})$ it is necessary to know the actual probability distribution $P(N_{ch})$ of the true charged-particle multiplicities. This probability distribution is not accessible by the experiment. Therefore, an iterative procedure to approximate $P(N_{ch})$, and consequential $P(N_{ch}|N_{acc})$ will be described in the following.

Starting with an arbitrary choice for $P(N_{ch})$, henceforth called $\hat{P}(N_{ch})$, some first guess $\hat{P}(N_{ch}|N_{acc})$ for the unfolding matrix can be calculated by equation 3.4.4. Together with the number of events $n_{evt}(N_{acc})$, which were measured with multiplicity N_{acc} , this $\hat{P}(N_{ch}|N_{acc})$ can then be used to obtain the expected number of events $\hat{n}_{evt}(N_{ch})$ with a true multiplicity N_{ch} :

$$\hat{n}_{evt}(N_{ch}) = \sum_{N_{acc}} \hat{P}(N_{ch}|N_{acc}) n_{evt}(N_{acc}) \quad (3.4.6)$$

This equation represents the process of unfolding the true multiplicity distribution from the measured one. For each multiplicity N_{acc} a fraction $\hat{P}(N_{ch}|N_{acc})$ of the $n_{evt}(N_{acc})$ measured events is assigned to a true multiplicity N_{ch} . The resulting unfolded multiplicity distribution $\hat{n}_{evt}(N_{ch})$ can now be used to calculate a better estimate for the probability distribution of the true multiplicities:

$$\hat{P}(N_{ch}) = \frac{\hat{n}_{evt}(N_{ch})}{\sum_{N'_{ch}} \hat{n}_{evt}(N'_{ch})} \quad (3.4.7)$$

This probability distribution is closer to the actual $P(N_{ch})$ of the collisions than the initial guess, because the measured multiplicity distribution $n_{evt}(N_{acc})$, which is a direct effect of $P(N_{ch})$, contributes in the unfolding procedure (equation 3.4.6) and thereby constrains $\hat{P}(N_{ch})$. Therefore, using this $\hat{P}(N_{ch})$ as a better initial guess in equation 3.4.4 and starting the procedure over again will result in an even better estimate for $\hat{P}(N_{ch})$. This suggests to repeat this process several times. The closer $\hat{P}(N_{ch})$ is to $P(N_{ch})$, the better it is in agreement with the result from the previous iteration. This decreasing difference of $\hat{P}(N_{ch})$ between two consecutive steps of the iteration can be used as an indicator for when to stop the procedure. The speed of convergence depends on the initial probability distribution. It is important to note, that if the initial $\hat{P}(N_{ch})$ is set to zero for some N_{ch} , the unfolding procedure is unable to recover events with this multiplicity. In this work the initial guess for $P(N_{ch})$ is inferred from the true multiplicity distribution $n_{evt}(N_{ch})_{MC}$ generated by the Monte Carlo simulation.

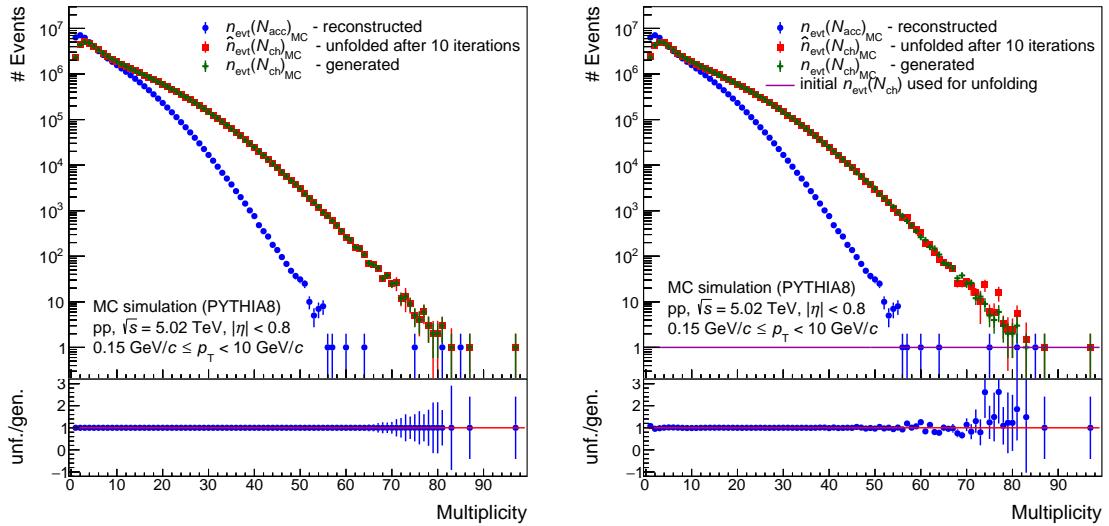


Figure 3.12: Reconstructed, unfolded (after ten iterations) and generated multiplicity distributions from the MC simulation. For the reconstructed multiplicity distribution the x-axis label 'Multiplicity' refers to N_{acc} and for the others it represents N_{ch} . In the left panel the unfolding procedure is started with the generated multiplicity distribution. In the right panel a flat distribution is used.

The Monte Carlo simulation is also predestined to verify if the unfolding procedure works as expected. Since the true multiplicity distribution $n_{\text{evt}}(N_{\text{ch}})_{\text{MC}}$ is known, it can be compared to the unfolded multiplicity distribution $\hat{n}_{\text{evt}}(N_{\text{ch}})_{\text{MC}}$ obtained by applying the unfolding procedure to the reconstructed multiplicity distribution $n_{\text{evt}}(N_{\text{acc}})_{\text{MC}}$. This is displayed in figure 3.12. The left panel shows the reconstructed (blue), unfolded (red) and generated (green) multiplicity distributions. The unfolding procedure is started with the generated multiplicity distribution just like it will be done in data. After an exemplary ten iterations, the resulting unfolded multiplicity distribution is (still) in perfect agreement with the generated one, as shown by the ratio in the bottom panel. In the right panel of figure 3.12, the unfolding procedure is started with a flat probability distribution for $\hat{P}(N_{\text{ch}})$. The resulting unfolded multiplicity distribution after ten iterations is also close to the generated one, but has not yet converged.

After explaining the unfolding procedure and verifying it with the MC simulation, it is now applied to the measured data. By studying the differences of the unfolded multiplicity distribution $\hat{n}_{\text{evt}}(N_{\text{ch}})$ between the consecutive steps of the procedure, ten iterations were found to be enough to assume convergence. Since the statisti-

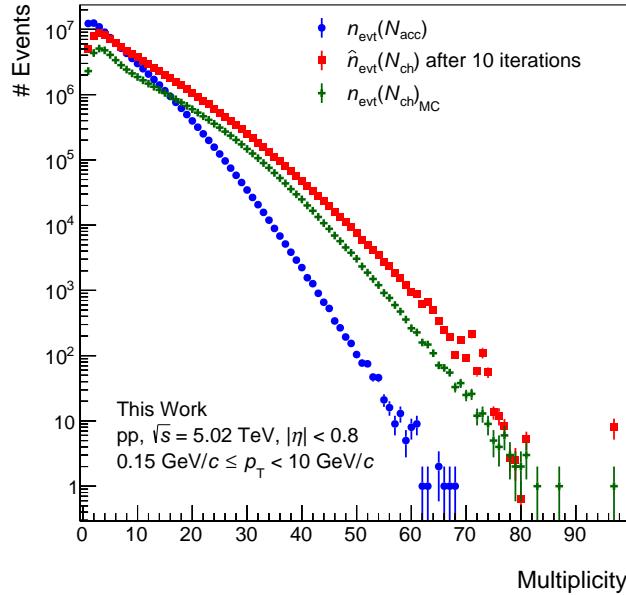


Figure 3.13: Measured, unfolded (after ten iterations) and MC truth multiplicity distributions in pp collisions at $\sqrt{s} = 5.02$ TeV. For the measured multiplicity distribution the x-axis label 'Multiplicity' refers to N_{acc} and for the others it represents N_{ch} .

cal uncertainty of the measured multiplicity distribution $n_{evt}(N_{acc})$ contributes to $\hat{n}_{evt}(N_{ch})$ in each iteration, it is not advisable to use too many iterations. A detailed calculation of this propagation of uncertainties can be found in [Ady11]. From now on it is assumed that after these ten iterations $\hat{n}_{evt}(N_{ch}) \approx n_{evt}(N_{ch})$ and consequential $\hat{P}(N_{ch}|N_{acc}) \approx P(N_{ch}|N_{acc})$.

Figure 3.13 shows the measured ($n_{evt}(N_{acc})$) and the unfolded ($\hat{n}_{evt}(N_{ch}) \approx n_{evt}(N_{ch})$) multiplicity distributions from data, as well as the true multiplicity distribution obtained from the MC simulation, which was used as the starting point for the unfolding procedure. The shapes of the simulated and the unfolded multiplicity distributions are similar, but not in perfect agreement. Noticeable is the apparent dip at $N_{ch} \approx 10$ in the MC truth distribution, which cannot be found in the unfolded distribution. This fact again highlights the importance of the initial reasoning to not obtain the unfolding matrix directly from the Monte Carlo simulation, but by means of a procedure which is relatively independent of the event generator.

The aim of all the previous calculations was to obtain the unfolding matrix, which is shown in figure 3.14. Its elements $P(N_{ch}|N_{acc})$ represent the fraction of

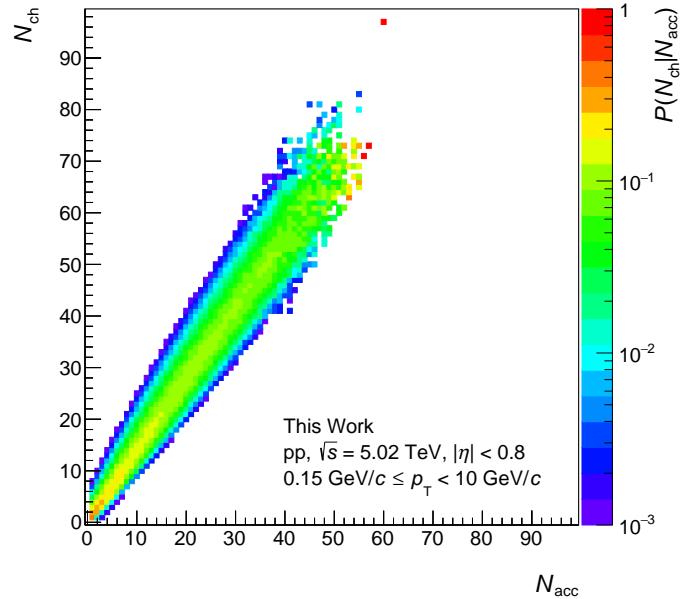


Figure 3.14: Unfolding matrix after ten iterations.

events measured with multiplicity N_{acc} , which originally have a true multiplicity N_{ch} . This gained understanding of the composition of the measured events will be used as the baseline for unfolding the measured p_T spectra, which is described in the following.

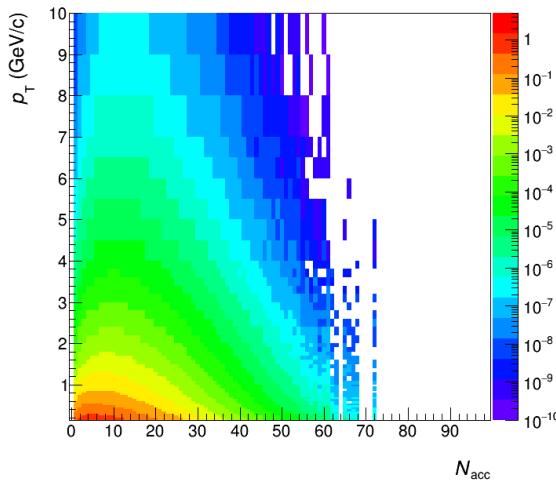


Figure 3.15: Measured p_T spectra in pp collisions at $\sqrt{s} = 5.02$ TeV as a function of N_{acc} .

3.4.2 Unfolding of spectra

Now, that the unfolding of the measured events is established, one can return to the original issue of unfolding the measured p_T spectra, which are shown in figure 3.15. The transverse momentum spectrum of events with a measured multiplicity N_{acc} consists of particles originating from events with many different true multiplicities N_{ch} . The p_T spectra of these contributing events have a characteristic shape depending on their respective multiplicity. Therefore, the shape of a measured p_T spectrum depends on its composition of the contributing N_{ch} spectra. A proper unfolding of the measured spectrum in principle needs to be done as a function of p_T . However, since the correlation between N_{ch} and N_{acc} given by the unfolding matrix is an event quantity, which does not contain any p_T dependence, this is not possible. Nevertheless, a measured p_T spectrum can still be unfolded under the assumption that the difference in the shapes of the N_{ch} dependent p_T spectra contributing to the measured spectrum is small.

To illustrate this, in figure 3.16 the results from an example MC simulation for events with $N_{acc} = 20$ are shown. The upper four panels show the average charged particle p_T distributions per event for true multiplicities $N_{ch} = 26, 27, 28$ and 29 . According to the unfolding matrix in figure 3.14, events with these multiplicities contribute to about 40 % of all events which lead to the measured multiplicity $N_{acc} = 20$.

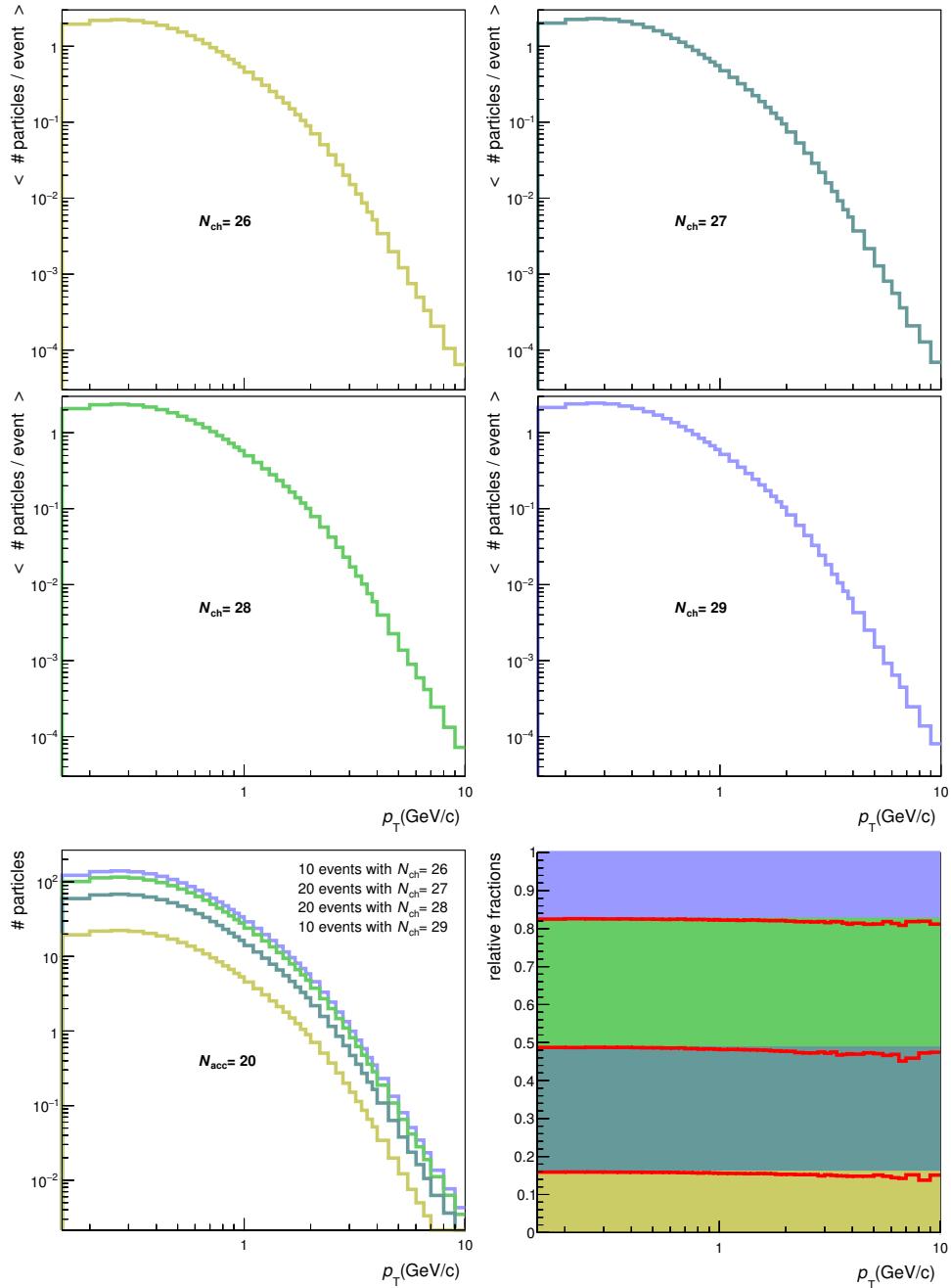


Figure 3.16: Toy simulation to illustrate the p_T -dependent relative composition of a measured p_T spectrum. The upper 4 panels show the mean number of particles per event as a function of p_T obtained from the MC simulation for the multiplicities $N_{ch} = 26, 27, 28, 29$, respectively. The bottom left panel shows a toy measurement for $N_{acc} = 20$ composed of these four N_{ch} multiplicities. The bottom right panel shows the resulting relative composition of the toy measurement with respect to the true multiplicities as a function of p_T (red lines) and the p_T -independent composition of the integral of particles (coloured areas).

In the example calculation 10 events with $N_{\text{ch}} = 26$, 20 events with $N_{\text{ch}} = 27$, 20 events with $N_{\text{ch}} = 28$ and 10 events with $N_{\text{ch}} = 29$ contribute to the p_T spectrum at $N_{\text{acc}} = 20$. This resulting transverse momentum spectrum, shown in the bottom left panel of figure 3.16, is the weighted sum of the average p_T spectra (per event) of the four different true multiplicities. The red lines in the bottom right panel of the figure show the relative composition of the measured p_T spectrum by the four weighted p_T spectra. These compositions show hardly any p_T dependence. In this example only four different true event multiplicities, which are relevant for the measurement at $N_{\text{acc}} = 20$, gives a similar result. This shows, that the ‘local’ similarity of the different N_{ch} -dependent p_T spectra allows for a p_T -independent unfolding of a measured p_T spectrum.

In addition to the red lines in the bottom right panel of figure 3.16, the p_T -independent fractions of the integrated number of measured particles with respect to the different contributing true multiplicities are indicated by the coloured areas. These fractions are in good agreement with the p_T -dependent relative compositions of the measured spectrum (red lines). Therefore, it can be concluded that these p_T -independent fractions can in good approximation be used to unfold the measured p_T spectra.

It should be emphasized, that the relative composition of the total number of particles assigned to a multiplicity N_{acc} is not equal to the corresponding relative composition of the number of measured events assigned to N_{acc} , which is given by the unfolding matrix $P(N_{\text{ch}}|N_{\text{acc}})$. This is shown in figure 3.17, where the upper part illustrates the unfolding of the events as discussed in the previous section, and the lower part illustrates the unfolding of the particles. The following reasoning aims to find out how the integral particles assigned to N_{acc} is composed of particles originating from events with true multiplicity N_{ch} .

Using equation 3.4.6, the total number of particles expected to originate from all events with a true multiplicity N_{ch} can be expressed by:

$$\hat{n}_{\text{par}}(N_{\text{ch}}) = N_{\text{ch}} \cdot \hat{n}_{\text{evt}}(N_{\text{ch}}) = N_{\text{ch}} \cdot \sum_{N_{\text{acc}}} P(N_{\text{ch}}|N_{\text{acc}}) n_{\text{evt}}(N_{\text{acc}}) \quad (3.4.8)$$

On the other hand, the number of particles measured in events at a certain N_{acc} is given by:

$$n_{\text{par}}(N_{\text{acc}}) = \langle N_{\text{ch}} \rangle(N_{\text{acc}}) \cdot n_{\text{evt}}(N_{\text{acc}}) \quad (3.4.9)$$

Here $\langle N_{\text{ch}} \rangle(N_{\text{acc}})$ denotes the average true number of particles contributed by each

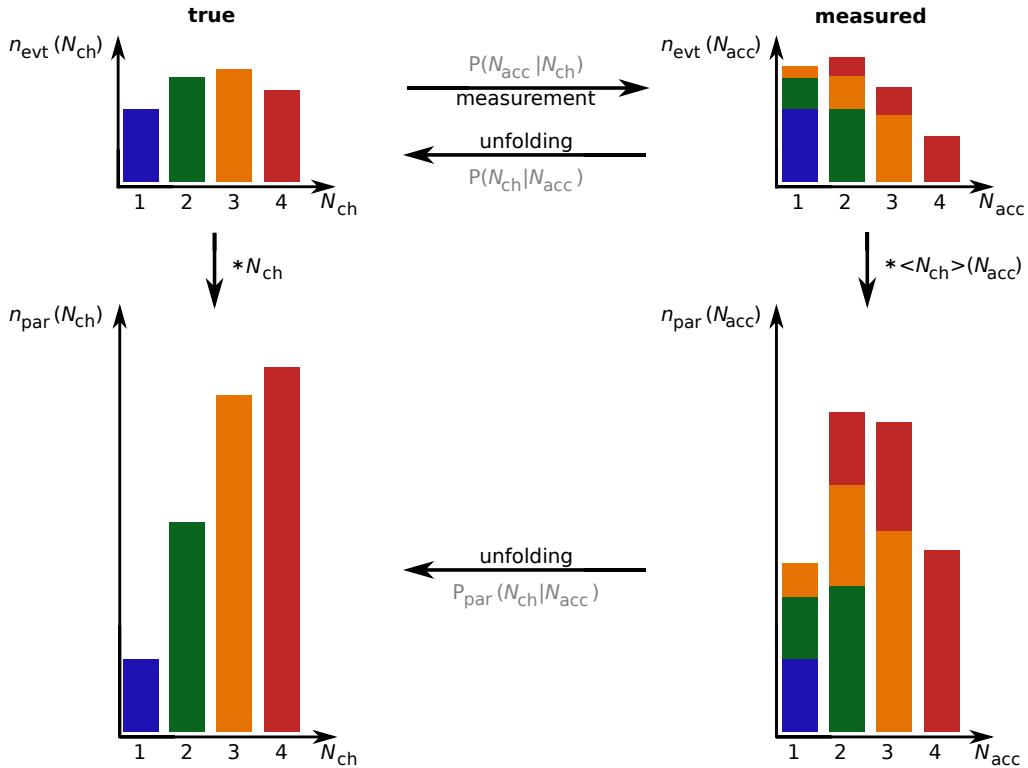


Figure 3.17: Illustration of the relation between event unfolding and particle unfolding.

of the $n_{\text{evt}}(N_{\text{acc}})$ events:

$$\begin{aligned} \langle N_{\text{ch}} \rangle(N_{\text{acc}}) &= \frac{1}{n_{\text{evt}}(N_{\text{acc}})} \sum_{N'_{\text{ch}}} N'_{\text{ch}} \cdot P(N'_{\text{ch}}|N_{\text{acc}}) n_{\text{evt}}(N_{\text{acc}}) \\ &= \sum_{N'_{\text{ch}}} N'_{\text{ch}} \cdot P(N'_{\text{ch}}|N_{\text{acc}}) \end{aligned} \quad (3.4.10)$$

Combining equations 3.4.8 and 3.4.10 yields:

$$\hat{n}_{\text{par}}(N_{\text{ch}}) = \sum_{N_{\text{acc}}} \frac{N_{\text{ch}}}{\langle N_{\text{ch}} \rangle(N_{\text{acc}})} P(N_{\text{ch}}|N_{\text{acc}}) n_{\text{par}}(N_{\text{acc}}) \quad (3.4.11)$$

This can be simplified by defining a particle(!) unfolding matrix as:

$$P_{\text{par}}(N_{\text{ch}}|N_{\text{acc}}) = \frac{N_{\text{ch}}}{\langle N_{\text{ch}} \rangle(N_{\text{acc}})} P(N_{\text{ch}}|N_{\text{acc}}) \quad (3.4.12)$$

The resulting unfolding relation for the number of measured particles

$$\hat{n}_{\text{par}}(N_{\text{ch}}) = \sum_{N_{\text{acc}}} P_{\text{par}}(N_{\text{ch}}|N_{\text{acc}}) n_{\text{par}}(N_{\text{acc}}) \quad (3.4.13)$$

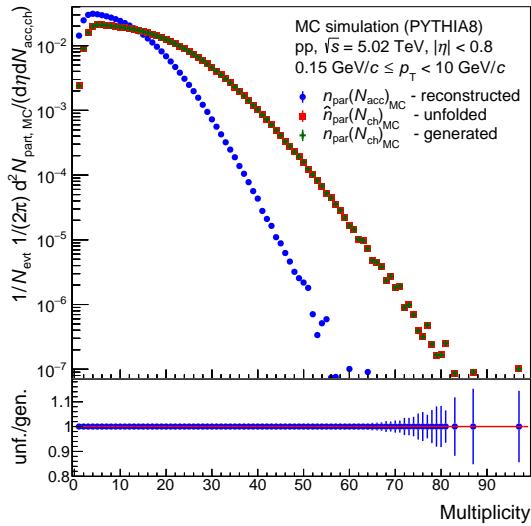


Figure 3.18: Reconstructed, unfolded and generated multiplicity distributions of the integral of particles from the MC simulation. For the reconstructed particle multiplicity distribution the x-axis label 'Multiplicity' refers to N_{acc} and for the others it represents N_{ch} .

is similar to the relation for the number of measured events given by equation 3.4.6. Analogously to the (event) unfolding matrix $P(N_{\text{ch}}|N_{\text{acc}})$, the particle unfolding matrix satisfies the following normalization condition for each N_{acc} :

$$\sum_{N_{\text{ch}}} P_{\text{par}}(N_{\text{ch}}|N_{\text{acc}}) = \sum_{N_{\text{ch}}} \frac{N_{\text{ch}}}{\langle N_{\text{ch}} \rangle (N_{\text{acc}})} P(N_{\text{ch}}|N_{\text{acc}}) = 1 \quad (3.4.14)$$

$P_{\text{par}}(N_{\text{ch}}|N_{\text{acc}})$ represents the probability that a particle(!) whose event was measured with a multiplicity N_{acc} actually comes from an event with true multiplicity N_{ch} . It should be emphasized, that this particle unfolding matrix contains the composition fractions of the measured particles, which can be used to unfold the measured p_T spectra.

Analogously to the event unfolding matrix, the particle unfolding matrix can be verified by applying it to the reconstructed multiplicity distribution of particles $n_{\text{par}}(N_{\text{acc}})_{\text{MC}}$ obtained from the Monte Carlo simulation and by comparing the unfolded result with the generated true multiplicity distribution of particles $n_{\text{par}}(N_{\text{ch}})_{\text{MC}}$. This is shown in figure 3.18, where $n_{\text{par}}(N_{\text{acc}})_{\text{MC}}$ is indicated by the blue, $n_{\text{par}}(N_{\text{ch}})_{\text{MC}}$ by the green and the unfolded multiplicity distribution of particles $\hat{n}_{\text{par}}(N_{\text{ch}})_{\text{MC}}$ by the red line. The bottom panel of the figure shows the ratio between the unfolded and the generated distribution. Both are in perfect agreement.

Furthermore, the unfolding procedure can be applied to the reconstructed p_T spec-

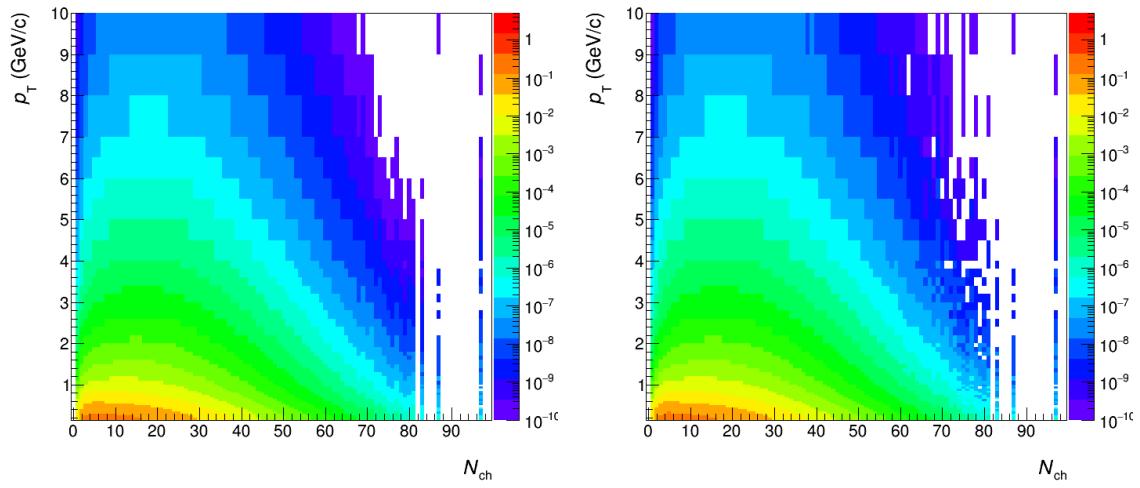


Figure 3.19: Unfolded (left) and generated (right) p_T spectra in pp collisions at $\sqrt{s} = 5.02$ TeV as a function of multiplicity obtained from the Monte Carlo simulation.

tra of the Monte Carlo simulation. Figure 3.19 shows the resulting unfolded p_T spectra on the left panel and for comparison the generated p_T spectra on the right panel. The spectra are similar, but are not in perfect agreement. This will be discussed quantitatively in section 3.5.2. The discrepancy is probably due to the applied approximation where the slight p_T dependence of the relative composition of the measured p_T spectra was ignored.

Finally, the unfolding procedure can be applied to the data. Figure 3.20 shows the measured as well as the unfolded p_T spectra obtained by the Bayesian unfolding method. The unfolding procedure is capable to extract the p_T spectra as a function of the true charged particle multiplicity up to $N_{\text{ch}} \approx 84$ from the measured p_T spectra, which only extend up to a measured multiplicity of about $N_{\text{acc}} \approx 70$. Based on these unfolded p_T spectra, their moments can be calculated as described in section 3.3. The resulting $\langle p_T^n \rangle(N_{\text{ch}})$ are shown in figure 3.21. These moments look very similar to the ones obtained with the re-weighting method. A quantitative comparison of the the results from the two methods is shown in section 3.5.4.

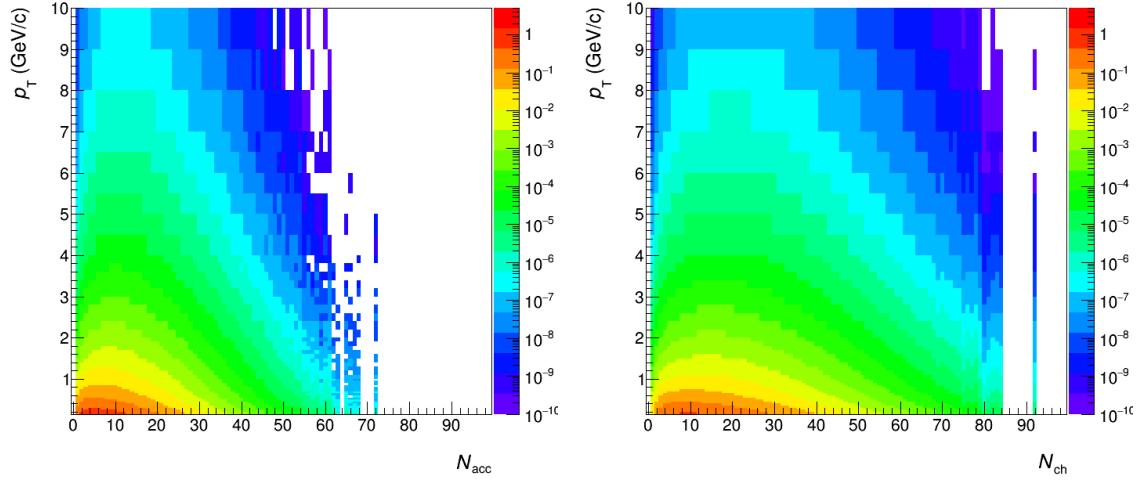


Figure 3.20: p_T spectra in pp collisions at $\sqrt{s} = 5.02$ TeV as a function of multiplicity before (left) and after (right) unfolding.

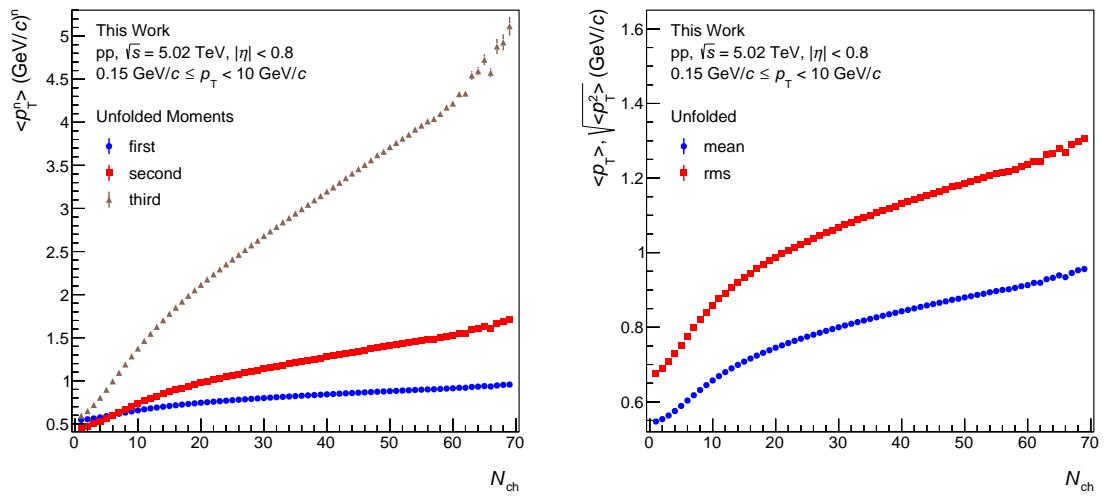


Figure 3.21: The left panel shows the first three moments of the unfolded p_T spectra as a function of true multiplicity N_{ch} with statistical uncertainties. In the right panel the mean and RMS are shown.

ID	Description	Nominal	Variations	
			lower	higher
1	max. DCA_z	2 cm	1 cm	5 cm
2	max. DCA_{xy}	7σ	4σ	10σ
3	max. ratio of crossed rows over findable clusters	0.8	0.7	0.9
4	max. fraction of shared TPC clusters	0.4	0.2	1
5	max. χ^2_{TPC} per TPC cluster	4	3	5
6	max. χ^2_{ITS} per ITS cluster	36	25	49
7	one hit in the SPD	required	not required	
8	geometric length (dead TPC area)	3 cm	2 cm	4 cm
9	geometric length (track length)	130	120	140
10	max. $\chi^2_{TPC-ITS}$	36	25	49

Table 3.1: Track selection criteria and their respective variations.

3.5 Systematic uncertainties and results

In order to obtain an estimate for the systematic uncertainties of the moments $\langle p_T^n \rangle (N_{ch})$ determined in this work, two different aspects are taken into account. On the one hand systematic effects of the choice of track quality constraints are studied. On the other hand the reliability of the re-weighting respectively unfolding method is tested by applying them to the reconstructed p_T spectra obtained from the Monte Carlo simulation.

3.5.1 Track selection

To study the systematic effects of the track selection on the moments $\langle p_T^n \rangle (N_{ch})$, their stability against a variation of the selection criteria is probed. The track quality requirements used in this work as well as their respective variations are listed in table 3.1. They follow the current ‘standard’ as used in [Ali17]. For each of these settings, the track reconstruction and $\langle p_T^n \rangle (N_{ch})$ determination is performed. If the virtual detector in the Monte Carlo simulation would perfectly describe the behaviour of ALICE, the efficiency correction would cancel out any effect of the track selection criteria. The residual effects on $\langle p_T^n \rangle (N_{ch})$, caused by an imperfection of this detector description, are therefore assigned as systematic uncertainties. Most of the track constraints used in this work (except for the requirement of one hit in the SPD) are varied to a lower and a higher value with respect to the nominal one.

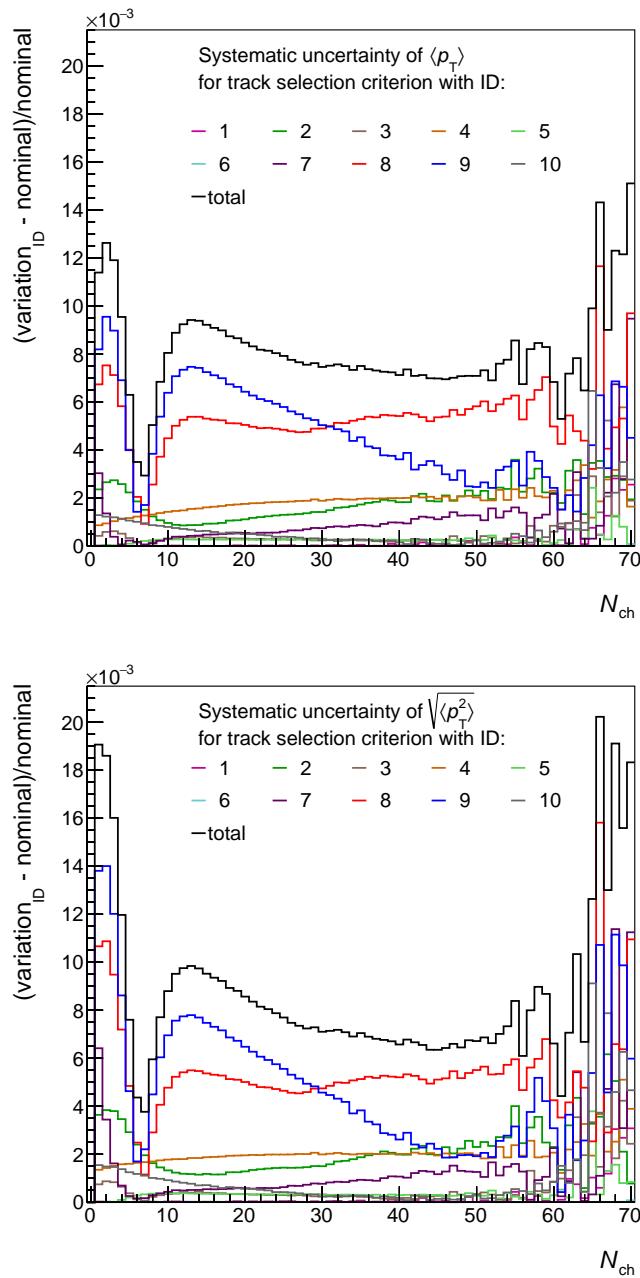


Figure 3.22: Systematic uncertainties of $\langle p_T \rangle(N_{\text{ch}})$ (upper panel) and $\sqrt{\langle p_T^2 \rangle}(N_{\text{ch}})$ (lower panel) obtained via Bayesian unfolding for the different track selection criteria described in table 3.1. For the criteria which are varied to a lower and a higher value, the maximum effect on $\langle p_T \rangle$ respectively $\sqrt{\langle p_T^2 \rangle}$ is taken into account for each multiplicity N_{ch} .

For these cases, the largest deviation from the nominal $\langle p_T^n \rangle (N_{\text{ch}})$ spectrum is used as symmetric systematic uncertainty. This is done separately for each multiplicity N_{ch} .

As an example, figure 3.22 shows the resulting systematic uncertainties of $\langle p_T \rangle (N_{\text{ch}})$ (upper panel) and $\sqrt{\langle p_T^2 \rangle} (N_{\text{ch}})$ (lower panel) obtained using the Bayesian unfolding method. The different track variations are labelled according to their ID defined in table 3.1. The black line indicates the total systematic uncertainty, which is calculated by adding the contributions from the different variations of the track selection criteria in quadrature. It reaches the highest values of about 1.5 % (mean) respectively 2.0 % (RMS) for multiplicities $N_{\text{ch}} \approx 70$. The two main contributions to the total systematic uncertainties represent variations of the required track length in the TPC (indicated in blue, ID:9 and red, ID:8).

3.5.2 Monte Carlo closure test

The systematic uncertainty of the re-weighting and the Bayesian unfolding method is studied by means of a so-called Monte Carlo closure test. In the Monte Carlo simulation the true p_T spectra produced by the event generator are known. Therefore, the corresponding moments $\langle p_T^n \rangle (N_{\text{ch}})_{\text{MC}}$ can be compared to the ones obtained by applying the two different methods to the reconstructed p_T spectra. Since this is a pure comparison within the Monte Carlo simulation, it is important not to apply the particle composition correction on the reconstructed spectra.

The purple markers in the upper panel of figure 3.23 show $\langle p_T \rangle (N_{\text{ch}})_{\text{MC}}$ as it is generated by the simulation. The blue markers represent $\langle p_T \rangle (N_{\text{acc}})_{\text{MC}}$ of the reconstructed p_T spectra. In red and green the re-weighted and unfolded $\langle p_T \rangle (N_{\text{ch}})_{\text{MC}}$ are displayed. Both methods have a similar effect on the input spectra (for details see next section). In the lower part of the upper panel, the ratios between the re-weighted respectively unfolded and the generated $\langle p_T \rangle (N_{\text{ch}})_{\text{MC}}$ are shown. These ratios follow a characteristic trend, which appears to be dominated by two different effects: For $N_{\text{ch}} < 10$ the ratios are above unity and show an excess, which has its maximum at low multiplicities $N_{\text{ch}} \approx 1$. At higher multiplicities $N_{\text{ch}} > 10$, the ratios are below unity and develop into a linear slope. For the systematic uncertainties, this trend is extrapolated in order to smooth the fluctuations of the ratio at high multiplicities $N_{\text{ch}} > 50$, which is caused by limited statistics in the generated spectra. This is indicated by the black line in the lower part of the upper panel.

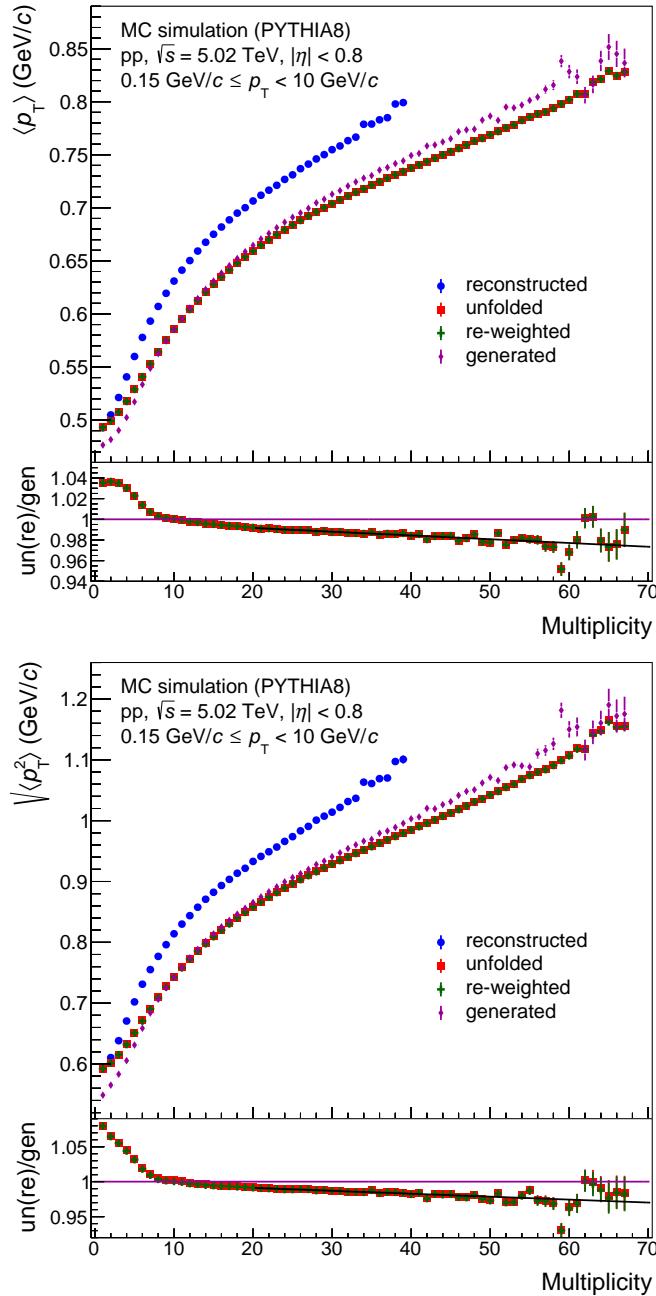


Figure 3.23: Monte Carlo closure test for $\langle p_T \rangle_{MC}$ (top panel) and $\sqrt{\langle p_T^2 \rangle}_{MC}$ (bottom panel). The x-axis label 'Multiplicity' refers to N_{acc} for the reconstructed mean respectively RMS and otherwise represents N_{ch} .

The maximum systematic uncertainties for the low multiplicity region is 4 % and for $N_{\text{ch}} > 10$ it is below 3 %. In the lower panel of figure 3.23 the same is shown for the RMS. The ratios between the re-weighted respectively unfolded $\sqrt{\langle p_T^2 \rangle}(N_{\text{ch}})_{\text{MC}}$ follow a similar trend to the one observed for $\langle p_T \rangle(N_{\text{ch}})_{\text{MC}}$. Here, for low multiplicities the deviation to the generated $\sqrt{\langle p_T^2 \rangle}(N_{\text{ch}})_{\text{MC}}$ is up to 6 % and for $N_{\text{ch}} > 10$ it is still below 3 %.

The second and third moments show the same behaviour (no picture). The maximum values of the ratios in the low respectively high multiplicity regions are: $\langle p_T^2 \rangle$: 17 % and 6 %; $\langle p_T^3 \rangle$: 40 % and 9 %. Those are significantly higher than the ones obtained for the mean and the RMS.

3.5.3 Total systematic uncertainty

The resulting total systematic uncertainty of the moments $\langle p_T^n \rangle(N_{\text{ch}})$ is calculated by adding the contributions from the track selection and the closure test in quadrature. Even though the systematic uncertainty from the Monte Carlo closure test has an asymmetric character, it is symmetrized in this work. Therefore, the total systematic uncertainty is a rather conservative estimate. In figure 3.24 the resulting $\langle p_T^n \rangle(N_{\text{ch}})$ obtained for the ALICE measurement in pp collisions at $\sqrt{s} = 5.02$ TeV by the re-weighting (upper panels) and the Bayesian unfolding method (lower panels) are shown with statistical and systematic uncertainties. The left panels show the first three moments and the right panels show the mean and RMS.

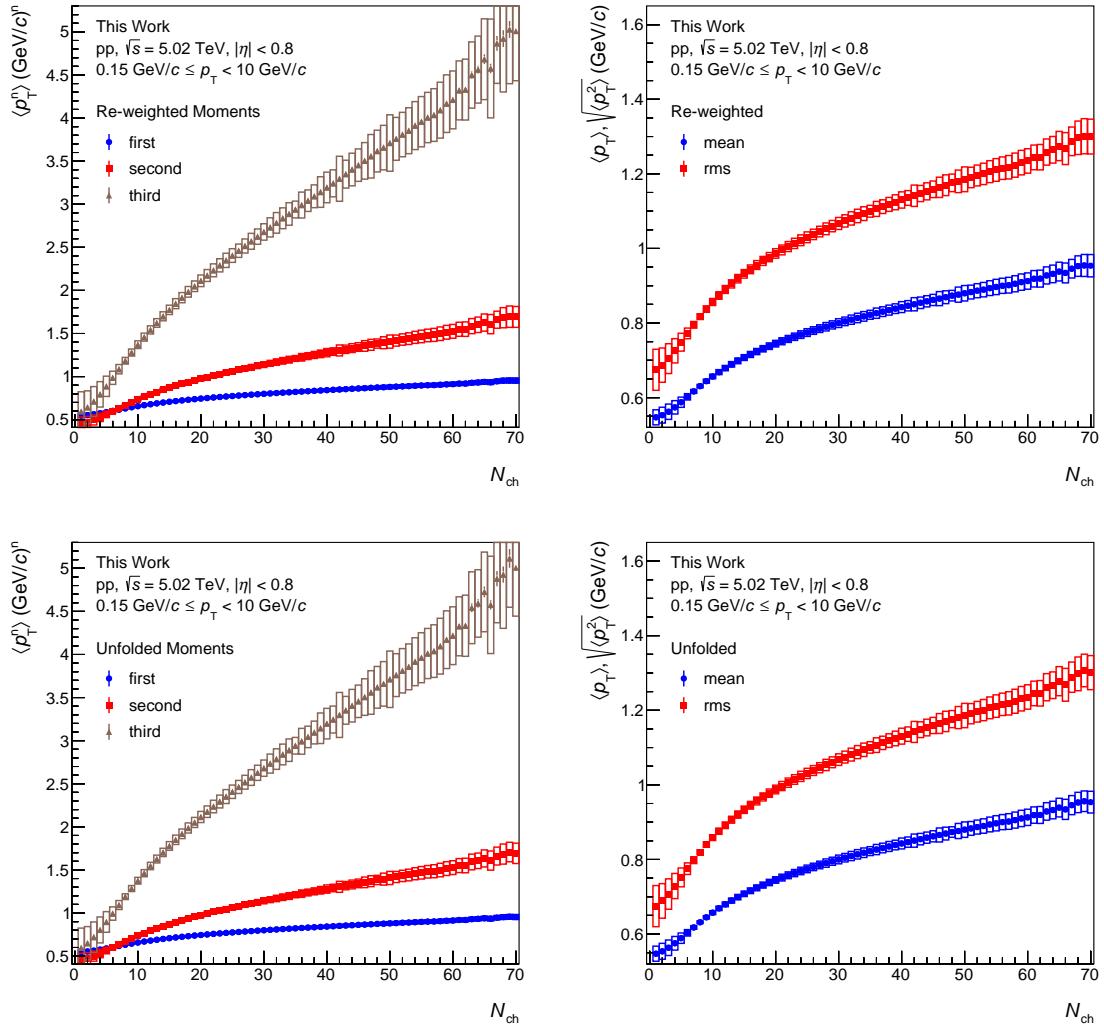


Figure 3.24: Re-weighted moments of the measured p_T spectra (upper panels) and moments of the unfolded p_T spectra (lower panels) as a function of true multiplicity N_{ch} with statistical and systematic uncertainties. The left panels show the first three moments and the right panels show the mean and RMS.

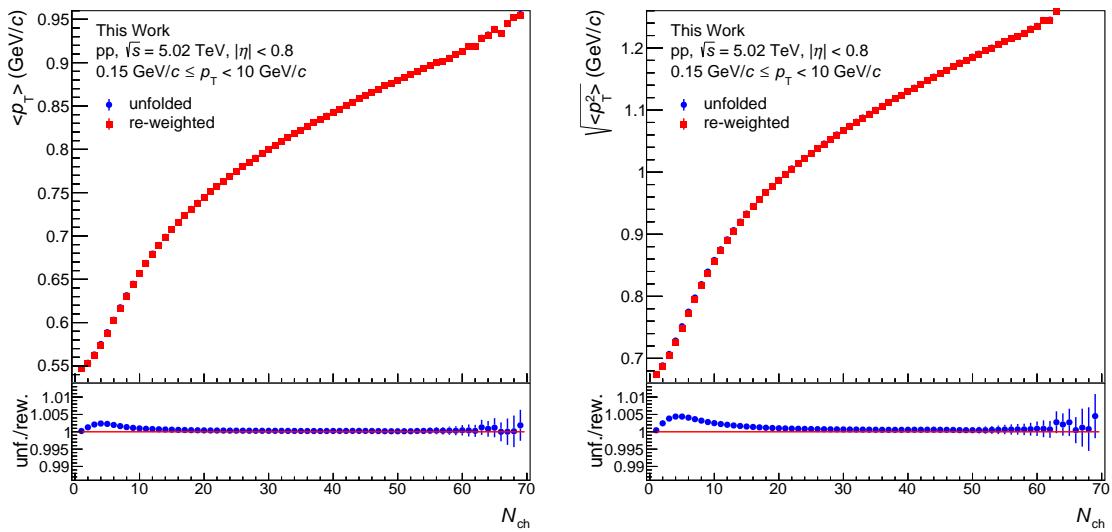


Figure 3.25: Comparison of re-weighting and Bayesian unfolding method for the mean (left) and the RMS (right).

3.5.4 Comparisons of the different methods

In figure 3.25, $\langle p_T \rangle(N_{\text{ch}})$ (left panel) and $\sqrt{\langle p_T^2 \rangle}(N_{\text{ch}})$ (right panel) are compared for the re-weighting and the Bayesian unfolding method. The bottom panel in both plots shows the ratio of the unfolded over the re-weighted results. Both methods are in good agreement, even though they are conceptionally very different. Only at low multiplicities, $N_{\text{ch}} \leq 10$ for the mean and $N_{\text{ch}} \leq 20$ for the RMS, the unfolding method yields a systematically higher value in the order of half a percent. This difference is less pronounced for the mean than for the RMS.

Chapter 4

Summary and outlook

In this work the mean transverse momentum of primary charged particles produced in proton-proton collisions at a centre-of-mass energy $\sqrt{s} = 5.02$ TeV is studied as a function of the event multiplicity.

The unidentified charged particles are measured within a kinematic range of $0.15 \text{ GeV}/c \leq p_T < 10 \text{ GeV}/c$ and $|\eta| < 0.8$ using the Inner Tracking System and the Time Projection Chamber of the ALICE experiment at the LHC. In this work, the event multiplicity is defined as the number of primary charged particles within in the afore mentioned kinematic range. Not all of these particles can be measured by the detector. Therefore, the raw particle yield needs to be corrected for these efficiency losses. This is done by means of a Monte Carlo simulation, which consists of an event generator (PYHTIA8) and a virtual model of the detector (GEANT3). Additionally, the true correlation between the transverse momentum spectra and the event multiplicity is lost in the measurement. In this work, two different approaches to reconstruct this correlation are presented. Both of them rely on a proper understanding of the detector response, which is obtained from the Monte Carlo simulation. At first, the conventional method used in prior ALICE publications (e.g. in [Ali13] and [Ali10]) is introduced and applied. It is based on re-weighting the moments of the measured p_T spectra. Secondly, an alternative procedure is applied, which aims to unfold the measured transverse momentum spectra using a Bayesian unfolding algorithm. The resulting mean and higher moments of these spectra are found to be in good agreement with the ones obtained with the re-weighting method. Only for low event multiplicities $N_{\text{ch}} \lesssim 10$ the unfolding method yields a slightly higher value in the order of 0.5 %.

As a next step of the analysis, the moments $\langle p_T^n \rangle (N_{\text{ch}})$ of the p_T spectra could be compared at different centre-of-mass energies. Up to now, data

CHAPTER 4. SUMMARY AND OUTLOOK

from pp collisions have been measured by ALICE at collision energies of $\sqrt{s} = 0.9, 2.76, 5.02, 7, 8,$ and 13 TeV, covering more than one order of magnitude in energy. Furthermore, $\langle p_T^n \rangle (N_{ch})$ could be obtained at the same centre-of-mass energy per nucleon pair $\sqrt{s_{NN}} = 5.02$ TeV for the three different collision systems of the LHC (pp, p-Pb and Pb-Pb).

Chapter 5

Acknowledgements

First and foremost I want to thank Prof. Dr. Henner Büsching for giving me the opportunity to work on this thesis and for the great support he gave me in the process. I am also grateful to Prof. Dr. Appelshäuser for reviewing this thesis.

Special thanks go to Patrick Huhn for his expert guidance and for always offering me advice throughout the progress of this work. He and my office mates Pascal Dillenseger and Stefan Heckel willingly answered all of my physics questions with great competence and patience. Without it, this thesis would not have been possible. This also applies to Anton Andronic and the whole spectra group, who always gave me useful advices when I got stuck with my analysis.

Furthermore I want to thank the rest of the Frankfurt High Energy Working Group, especially Joshua König, Matthias Kleiner, Rene Schach, Andrea Hornung, Hendrik Schulte and Lucas Westermann for all the support and fun times. The great office atmosphere and the positive everyday interactions helped me a lot to feel comfortable in this working environment.

Last but not least want to thank my family for making it possible for me to study physics and for always supporting me along the way. I am also grateful to my friends and fellow students Matthias Fix, Jan Teschabai-Oglu and Dominik Herzig for the great time we had during and after the physics courses. Thank You!

Chapter 6

References

- [A⁺08] Alice Collaboration et al. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08002, August 2008.
- [AA⁺10] J. Alme, Andres, et al. The ALICE TPC, a large 3-dimensional tracking device with fast readout for ultra-high multiplicity events. *Nuclear Instruments and Methods in Physics Research A*, 622:316–367, October 2010.
- [Ady11] Tim Adye. Unfolding algorithms and tests using RooUnfold. *Proceedings of the PHYSTAT Switzerland*, CERN-2011-006:313–318, 2011.
- [AGIS83] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. Parton fragmentation and string dynamics. *Physics Reports*, 97(2):31 – 145, 1983.
- [Ali10] Alice Collaboration. Transverse momentum spectra of charged particles in proton-proton collisions at $\sqrt{s} = 900 \text{ GeV}$ with ALICE at the LHC. *Physics Letters B*, 693:53–68, September 2010.
- [Ali13] Alice Collaboration. Multiplicity dependence of the average transverse momentum in pp, p-Pb, and Pb-Pb collisions at the LHC. *Physics Letters B*, 727:371–380, December 2013.
- [Ali14] Alice Collaboration. Performance of the ALICE experiment at the CERN LHC. *International Journal of Modern Physics A*, 29:1430044, Septem-

ber 2014.

- [Ali17] Alice Collaboration. Transverse momentum spectra and nuclear modification factors of charged particles measured with ALICE in pp, p–Pb and Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ and 5.02 TeV . *Paper draft and internal plots*, 2017.
- [BB⁺94] R. Brun, F. Bruyant, et al. GEANT detector description and simulation tool. Technical report, CERN Program Library long writeup W5013, 1994.
- [BP63] T. Bayes and R. Price. An Essay towards solving a Problem in the Doctrine of Chance. *Phil. Trans.*, 53:370–418, 1763.
- [Col14] Alice Collaboration. Technical Design Report for the Upgrade of the ALICE Inner Tracking System. *Journal of Physics G: Nuclear and Particle Physics*, 41(8):087002, 2014.
- [D'A95] G. D'Agostini. A multidimensional unfolding method based on Bayes' theorem. *Nuclear Instruments and Methods in Physics Research A*, 362:487–498, February 1995.
- [GO09] Jan Fiete Große-Oetringhaus. *Measurement of the Charged-Particle Multiplicity in Proton–Proton Collisions with the ALICE Detector*. PhD thesis, Wilhelms-University Münster, 2009.
- [Gri87] David Griffiths. *Introduction to Elementary Particles*. John Wiley & Sons, 1987.
- [Huh17] Patrick Huhn. Datenbasierte Effizienzkorrektur zum Nachweis geladener Teilchen im ALICE Experiment. Master's thesis, Goethe University Frankfurt, 2017.
- [Kni14] Michael Linus Knichel. *Transverse momentum distributions of primary charged particles in pp, p–Pb and Pb–Pb collisions measured with ALICE at the LHC*. PhD thesis, Technical University of Darmstadt, 2014.

- [KST⁺15] V. Khachatryan, A. M. Sirunyan, A Tumasyan, et al. Measurement of the inclusive 3-jet production differential cross section in proton-proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range. *European Physical Journal C*, 75:186, May 2015.
- [LHC] LHC website. <http://www.lhc-facts.ch>. (accessed on 1st December 2016).
- [Lü10] Philipp Lüttig. Mittlerer Transversalimpuls in Proton-Proton-Kollisionen bei $\sqrt{s} = 900 \text{ GeV}$ in ALICE. Master's thesis, Goethe University Frankfurt, 2010.
- [SAC⁺15] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, June 2015.
- [Sjö09] Torbjörn Sjöstrand. Old Ideas in Hadronization: The Lund String. Slides: <http://home.thep.lu.se/~torbjorn/talks/durham09.pdf>, 2009.
- [SSS10] S. Sarkar, H. Satz, and B. Sinha, editors. *The Physics of the Quark-Gluon Plasma*, volume 785 of *Lecture Notes in Physics*, Berlin Springer Verlag, 2010.

Appendix A

Kinematic variables

The momentum of particles produced in high-energy collisions in the plane transverse to the beam axis in terms of the polar emission angle θ and total momentum $|\mathbf{p}|$ is given by:

$$p_T = |\mathbf{p}| \cdot \sin(\theta) \quad (\text{A.1})$$

The momentum component along the beam axis is accordingly:

$$p_L = |\mathbf{p}| \cdot \cos(\theta) \quad (\text{A.2})$$

Since the longitudinal momentum is hard to measure, another important observable is the so-called rapidity y :

$$y = \frac{1}{2} \ln \left(\frac{E + p_L}{E - p_L} \right) \quad (\text{A.3})$$

Here, E is the energy of the particle. For ultra-relativistic particles the approximation $E \approx |\mathbf{p}|/c$ can be used to simplify this equation. The resulting quantity is called pseudorapidity:

$$\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_L}{|\mathbf{p}| - p_L} \right) \quad (\text{A.4})$$

In principle, the total momentum $|\mathbf{p}|$ is not known, but from geometrical considerations it can be expressed using the polar emission angle θ as:

$$\eta = -\ln [\tan(\theta/2)] \quad (\text{A.5})$$

Appendix B

Runlist

These are the specifications of the data sample used in this work:

Collision system:	proton - proton
Energy:	$\sqrt{s} = 5.02$ TeV
Data taking period:	LHC15n_pass4
Monte Carlo production:	LHC16k5a (PYTHIA8)
Runlist:	244340, 244343, 244351, 244355, 244359, 244364, 244377, 244411, 244416, 244418, 244421, 244453, 244456, 244480, 244481, 244482, 244483, 244484, 244531, 244540, 244542, 244617, 244618, 244619, 244626, 244627, 244628